

CHARACTERIZATION OF BREAST LESIONS BY PROCESSING DIGITAL BREAST IMAGES

Submission Date: Sep 30, 2024, Accepted Date: Oct 05, 2024,

Published Date: Oct 10, 2024

Crossref Doi: <https://doi.org/10.37547/ijmspahr/Volume05Issue10-03>

Abdulumhssin Binhssan

Security Forces Hospital, Riyadh, Saudi Arabia

ABSTRACT

This Rendering to the World Health Organization, women in both developed and developing nations are most likely to develop breast cancer. This illness causes breast cells to grow and multiply out of control. According to research institutes and international organizations, there are various screening methods available based on age, and breast cancer can be cured if detected in time. The Breast Imaging Reporting and Data System (BIRADS) is a standardized system that is commonly used in these techniques to report results and findings. Results are sorted by BIRADS into six categories, numbered 0 through 6. Furthermore, mammography is the most widely utilized screening technique.

This study suggests using mammography data processing to identify breast lesions. Adaptive filters are used for image cropping and contrast enhancement during the pre-processing phase. The pectoral muscle is then segmented using segmentation techniques that consider morphological and area growth factors. The lesion is then divided into sections at the muscle and breast levels using the Discrete Wavelet Transform (DWT), which finds any micro calcifications. Furthermore, to distinguish between dense lesions and other kinds of lesions, an area cultivation approach combined with multiple thresholding techniques is employed. Lastly, the obtained segmentation is used to extract textural and morphological features.

When expert-segmented and automatically segmented images were compared, the Sorensen Decade similarity index was 0.73, indicating the effectiveness of the suggested method. Considering that the lesion area on a mammogram can only be roughly delineated by hand or automatically, this is a promising outcome.

KEYWORDS

Image processing, Mammography, Segmentation, BIRADS, Pectoral muscle.

INTRODUCTION

According to Phung et al. (2019), breast cancer is the most common cancer in women worldwide, and individual patients may have a wide range of disease forms. Generally speaking, "mutations" in the genes that control cell growth lead to cancer because they permit unchecked cell division and multiplication (Ranchod and Herndon, 2019). Breast cancer arises from uncontrollably developing cells within the breast, whether in the ducts, connective tissue, or the right or left lobe (CDC, n/a). The term "metastasis" refers to the

process by which it spreads to the outside through lymphatic and blood vessels. Histopathological classification is based on features seen under a light microscope of biopsy samples (American Cancer Society, 2019). These features can be broadly categorized as follows: Invasive carcinoma, where cancer cells grow outside of lobules or are driven to other parts of the breast tissue, and Carcinoma in situ, where cancer cells do not spread to other parts (Phung et al, 2019).

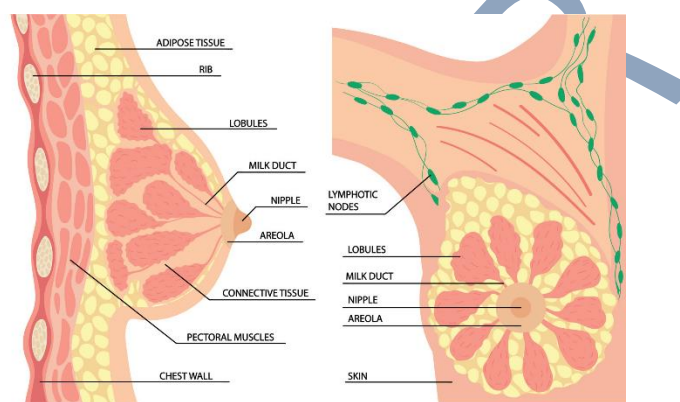


Figure 1: Frontal and transverse views of the breast (CDC, s/a).

As seen in Figure 1, lobes—little, round sac-like glands that produce milk—and milk ducts—passages that transport milk from the lobes to the nipple during feeding—make up breast tissue. intricate network.) in a pattern resembling grapes (Phung et al., 2019) (CDC, s/a).

The intricate network of fatty and fibrous tissue, lobules (small, round sac-like glands that produce milk), and ducts (channels that transport milk from the lobules to the nipple during lactation) that make up breast tissue is depicted in Figure 1 and is patterned after grapes (Phung et al, 2019) (CDC, n/a).

Every patient experiences breast cancer in a different way, and some do not exhibit any symptoms at all.

They don't exhibit any symptoms at all. Breast or nipple changes in size, shape, or feel, thickening inside the breast or underarm region, swelling, warmth, redness, or darkening of the breast, dimpling or creases in the skin, sores, or scales are some warning symptoms of breast cancer. among others, abrupt onset of nipple discharge and dermatitis (Susan G. Komen ORG, s/a) (Key et al, 2001).

Mammography, screening techniques, and a medical professional's physical examination of the breast are common bases for diagnosis (Saslow et al., 2004). To reach a definitive diagnosis in the event that the tests yield no results, the physician will perform an invasive tissue analysis, or biopsy (Taglicfacio et al., 2019). A descriptor lexicon, a suggested reporting structure, and a data collection framework are some of the



essential elements of the Breast Imaging Reporting and Data System (BIRADS) standard (Burnside et al, 2009). It has been said that it aids in reducing

misconceptions about how these various procedures should be interpreted. Table I provides a summary of the evaluation categories (D'Orsi et al., 2013).

Category	Evaluation
BIRADS 0	Incomplete: Needs further imaging evaluation.
BIRADS 1	Negative.
BIRADS 2	Benign.
BIRADS 3	Probably Benign.
BIRADS 4	4 A: Low suspicion for malignancy.
	4 B: Moderate suspicion for malignancy.
	4 C: High suspicion for malignancy.
BIRADS 5	Highly suggestive of malignancy.
BIRAD 6	Biopsy-proven malignancy.

Table I. BIRADS Classification (D'Orsi et al, 2013).

An estimated 9.6 million people worldwide lost their lives to cancer in 2018, with 2.09 million of those deaths being due to breast cancer (22% of all cancers, or 1 in 8 women) (WHO, 2018). According to estimates, women under the age of fifty account for one-third of these cancer cases. However, the percentage varies according to the socioeconomic status and geographic area (Romieu et al., 2019). According to the Colombia Alliance Against Cancer (s/a), this disease is the primary cause of cancer morbidity and mortality in Colombia

and the majority of Latin American and Caribbean countries, accounting for 22,174 women's deaths in the country over the previous ten years (Weisner et al., 2020).

According to several studies (Zielonke et al., 2020; Weingart et al., 2009), early detection of breast cancer is crucial for lowering mortality. However, in keeping with the same theme, it is crucial to consider the high rate of false positives that currently exist. Research has

revealed that the percentage of false positive mammography results is higher in individuals under 50 years old (355 per 100,000 versus 242 per 100,000 in those over 50 years old) (Wiesner et al, 2020). This can result in an increased risk of unnecessary treatment (biopsies and other procedures) (Román et al, 2011) (Lang et al, 2016) and/or significant psychological consequences (Aro et al, 2000), which can lead to an increased risk of developing breast cancer (Henderson et al.

al, 2015).

Computer-aided diagnosis (CAD), which was created in the 1980s and received FDA approval in 1998 (Kohli and Jha, 2017), has demonstrated promising outcomes in the processing and interpretation of medical images (Herinksen et al. 2018). The objective is to decrease supervision, thereby lowering the frequency of false positives or false negatives for medical professionals analyzing pictures (Katzen et al, 2018).

Thus, the goal of this study is to develop a method for characterizing digital mammography images that will enable accurate diagnosis in the future using BIRADS classification, matching or surpassing the accuracy of expert observers.

METHODOLOGY

The suggested methodology is broken down into five stages: segmentation, which is used to divide an image into distinct regions containing each pixel with similar attributes (Kamalakannan et al., 2015). This process divides the image into the pectoral muscle and the breast region, for example. The first stage involves selecting and organizing the image in BIRADS

categories. The second stage involves preprocessing, filtering, and augmentation of image contrast.

Finally, morphological and texture features are retrieved (González and Woods, 2007), showing variations in the external structure's shape and analyzing the distribution of intensities from the histogram of the masses discovered.

Experimental data

Categories—BIRADS 1, BIRADS 2-3, and BIRADS 4-5—are used to organize the 300 images. To obtain a reference image for assessing the algorithm's performance, 200 pieces representing the final two categories were manually segmented by experts using the online LabelBox tool.

Pre-treatment

Prior to processing mammography images, each picture was cropped to eliminate unnecessary details that were usually connected to the screening image (Figure 2-A). Furthermore, a breast orientation algorithm was applied, and if the orientation was accurate, the image was reversed to facilitate segmentation.

The image's edges are then enhanced and noise is eliminated using an adaptive filter, more precisely a diffuse contrast filter (Figure 2-B). Since this filter typically assumes near-absorptive pixel intensity values, each iteration causes a digital change in each pixel's intensity. Following that, it spreads to nearby pixels in accordance with the negative gradient. This indicates that a tiny amount of intensity from each central pixel is absorbed by its neighbors. Ultimately, the procedure comes to an end when the contrast agent is eliminated entirely (Gherghout et al, 2019).

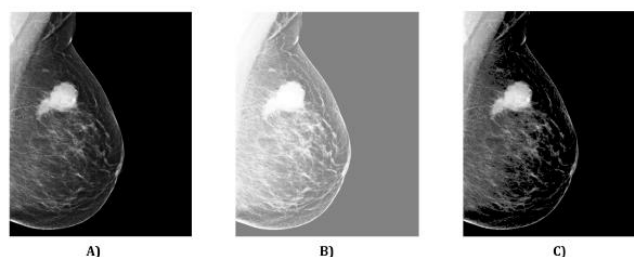


Figure 2: The second variation. A) The initial cropped picture. b) The diffusion contrast filter's outcome. c) Otsu method binary result.

Using Otsu's method, dualization is applied to the filtered image to increase the gray level contrast and achieve the goal of as little contrast within each histogram segment and as much contrast between segments, Menendez et al. (2016), in part. Ultimately, the original image is used for multiplication (Figure 2-C).

Segmentation

Pectoral muscles

The upper left corner of each image, which represents the location of the pectoral muscle, was used as the initial pixel in a region growth algorithm to compare the intensity of pixels within that region (Suárez et al., 2019). a). Following the application of morphological factors to refine the resulting mask (Figure 3-B), multi-threshold thresholding is used to determine whether or not a mass is associated with breast lymph nodes as figure 3-C.

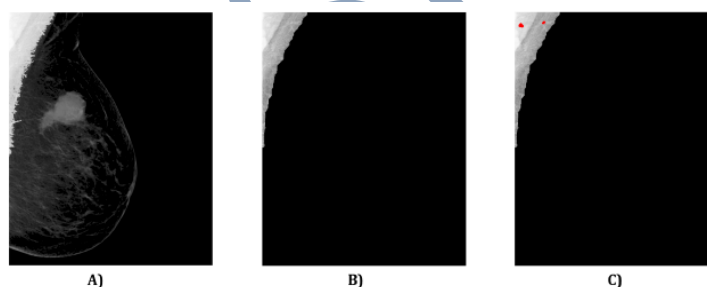


Figure 3: The third form. A) The result of applying area growth. b) Results of applying morphological factors. C) Multiple thresholding results.

Mammary gland area

Once again, region growth was employed, with a density level of 0.15 and 20% similarity seeded (Figure 3-A). Lastly, we refined the resulting mask using morphological operators (Figure 3-B). Daubesy with 8

coefficients (Db 8) was utilized in tandem with DWT to separate the image's high-frequency and low-frequency components. It was found that the approximate component was helpful in exposing intricate computations (Cruz et al., 2018).

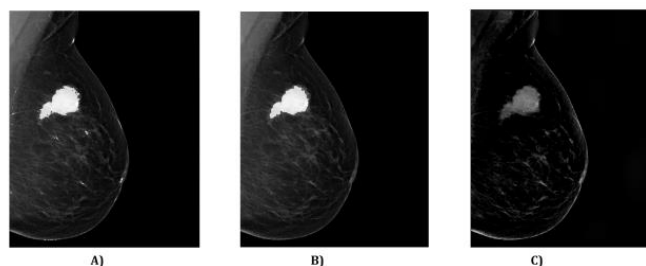


Figure 4: Results of applying region growing. b) Results of applying morphological factors. c) Results of removing the approximate component of the discrete wavelet transform.

Features Extraction

Once more, region growth was used, with 20% similarity seeded and a density level of 0.15 (Figure 4-A). Finally, we used morphological operators to refine the resulting mask (Figure 4-B). Together with DWT, Daubesy with 8 coefficients (Db 8) was used to separate the high-frequency and low-frequency components of the image. According to Cruz et al. (2018), the approximation component proved useful in revealing complex computations.

Texture properties

1st Order Texture Features

Among these characteristics are basic pixel features, which indicate pixel contrast, asymmetry, dominance, narrowness, and uniformity by describing the intensity distribution of pixels taken from the ROI histogram. They consist of the mean (the ROI's average intensity value), variance (the degree to which the intensity values are scattered from the mean), skewness (the distribution of pixel intensities around a center point), kurtosis (the degree to which the intensity distribution is flat), energy (the degree to which it is uniform), and entropy (the degree to which it is random) (Margaliot, 2008) (Osorio, 2015).

2nd Order Texture Features

When considering pairs of pixels, quadratic properties take into account 2D histograms to create a co-occurrence matrix (GLCM), where $CCdd\theta\theta(ii,jj)$ is the probability that a pixel has a gray level j at a distance d and direction θ from a pixel with gray level i . [Garghout and others, 2019]. Haralick and Shanmugam (1973) established global metrics, energy (degree of uniformity), entropy (degree of randomness), and variance (local dispersion), to characterize texture based on a co-occurrence matrix of directional means and standard deviations. According to Löfstedt et al. (2019), additional metrics include ROI gray levels, inertia, correlation (which quantifies the linear relationship between gray levels between pixels and the particular locations connected to each in the array), smoothing, inverse difference, inverse difference moment, and cluster orientation.

3rd Order Texture Features

A series of consecutive pixels in one direction and with the same gray level is called a gray level sequence. According to Osorio (2015), the amount of pixels in these sequences determines their length. The gray level matrix (GLRLM), a two-dimensional matrix where each position $PP(i, j | \theta)$ represents the number of times a sequence of length j with gray level i appears in a direction θ , is employed as the sequence length matrix within the third order characteristics (Osorio, 2015). Many characteristics, including short and long

sequence emphasis, non-uniformity of the gray levels, non-uniformity of the sequence length, low and high gray sequence emphasis, and sequence percentage, can be extracted from the interest once it has been calculated (Sayeed et al, 2011).

Morphological Features

Shape analysis divides an image into pixels inside the mass and pixels outside the mass with the purpose of extracting multiple quantitative metrics from the binary image (Gherghout et al, 2019). This document

makes use of the following morphological properties: area, which is the total number of pixels that comprise the ROI; circularity, which assesses the ROI's degree of circularity; eccentricity, which is the ratio of the two foci's distance to the major axis of the associated ellipse's length; equivalent diameter, which is the diameter of a circle with the same area of the ROI; perimeter, which is the total number of pixels that comprise the perimeter of the ROI; the length in pixels of the major and minor axes of the linked ellipse.

RESULTS

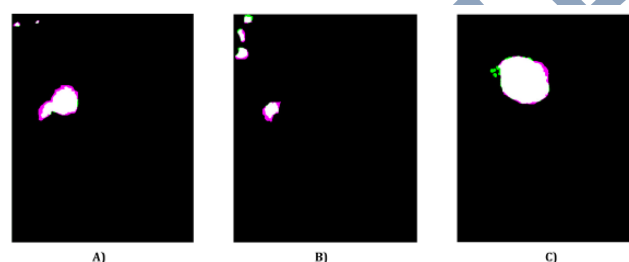


Figure5: Overlap between hashing performed by the proposed algorithm

200 segmented images were compared manually and automatically (Fig. 5 – A) in order to assess the effectiveness of the suggested method. The average Sørensen-Dice similarity index obtained was 0.73, with a standard deviation of 0.21. After that, the segmentation's overall area was compared, yielding a 33% average error and a 16% standard deviation (Fig. 5-B). Considering that experts point out that lesion areas in mammography images can only be roughly identified by hand or automatically, these results are encouraging.

Ultimately, an accuracy of 76.25% with a false positive range of 24% was obtained by segmenting blocks of the pectoral and breast muscle regions (Fig. 5-C). Given that the range of misdiagnoses is between 19 and 28%, this is a significant finding (Ventura-Alfaro, 2018).

CONCLUSIONS

The pectoral muscle region should always be identified separately when segmenting masses in mammograms. This is because, according to the BIRADS classification, masses in this region are not always malignant, making it one of the important pectoral muscle regions. The shift in contrast between the various tissues (breast region, pectoral muscle, tumor, adipose tissue) as a result of their density and radiolucency in each location is a distinctive feature of mammograms (Méndez et al., 2016).

Using threshold segmentation methods, region growing, and morphological factor-based techniques, significant results were obtained when segmenting masses in the pectoral muscle and breast regions.

It is suggested that more research be done by putting into practice a classifier of lesions from the various BIRADS categories, which can be effectively and automatically used for breast diagnosis.

REFERENCES

1. American Cancer Society. (2019). Types of Breast Cancer. Consultado el 20 de mayo de 2021 en <https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/types-of-breast-cancer.html>
2. Aro, A., Pilvikki, S., Van der Ploeg, E. y Van der Kamp, L. J. T. (2000). False-positive findings in mammography screening induces short-term distress - breast cancer-specific concern prevails longer. *European Journal of Cancer*, Vol. 36, No. 9, pp. 1089–1097.
3. Bahl, M. (2014). Detecting Breast Cancers with Mammography: Will AI Succeed Where Traditional CAD Failed. *Radiology*, Vol. 290, No. 2, pp. 315–316.
4. Burnside, E. S., Sickles, E. A., Bassett, L. W., Rubin, D. L., Lee, C. H., Ikeda, D. M. y D'Orsi, C. J. (2009). The ACR BI-RADS® Experience: Learning From History. *Journal of the American College of Radiology*, Vol. 6, No. 12, pp. 851–860.
5. Centers for Disease Control and Prevention. U.S. Department of Health & Human Services. (s/a) Breast Cancer. Consultado el 20 de mayo de 2021 en https://www.cdc.gov/cancer/breast/basic_info/what-is-breast-cancer.htm
6. Contreras, W., Nora, P. y Ruiz, M. E. (2010). Procesamiento Digital de textura: Técnicas utilizadas en aplicaciones actuales de CBIR. *Revista de Investigación de Sistemas e Informática*. Vol. 7, No. 1, pp. 57-64.
7. D'Orsi, C.J., Sickles, E.A., Mendelson, et al. (2013). ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA, American College of Radiology.
8. Da Cruz, T. N., Da Cruz, T. M. y Pinheiro, W. (2018) Detection and classification of mammary lesions using artificial neural networks and morphological wavelets. *IEEE Latin America Transactions*, Vol. 16, No. 3, pp. 926–932.
9. Gherghout, Y., Tlili, Y. y Souici, L. (2019). Classification of breast mass in mammography using anisotropic diffusion filter by selecting and aggregating morphological and textural features. *Evolving Systems*. doi:10.1007/s12530-019-09270-z
10. González, R. C. y Woods, R. E. (2007) *Digital Image Processing*. Pearson Education International, pp. 976.
11. Henderson, L. M., Hubbard, R. A., Sprague, B. L. et al (2015). Increased Risk of Developing Breast Cancer after a False-Positive Screening Mammogram. *Cancer Epidemiology Biomarkers & Prevention*, Vol. 24, No. 12, pp. 1882–1889.
12. Henriksen, E. L., Carlsen, J. F., Vejborg, I. M. et al (2018). The efficacy of using computer-aided detection for detection of breast cancer in mammography screening: a systematic review. *Acta Radiologica*, Vol. 60, No. 1, pp. 13-18.
13. Kamalakannan, J., Krishna, P. V., Babu, M. R. y Mukeshbhai, K. D. (2015) Identification of abnormality from digital mammogram to detect breast cancer. *International Conference on Circuits, Power and Computing Technologies*, India, pp. 1-5.
14. Katzen, J. y Dodelzon, K. (2018). A review of computer aided detection in mammography. *Clinical Imaging*, Vol. 52, pp. 305-309.
15. Key, T. J., Verkasalo, P.K. y Banks, E. (2001). Epidemiology of Breast Cancer. *Lancet Oncol*, Vol. 2, No. 3, pp. 133-40.
16. Kohli, A. y Jha, S. (2017). Why CAD Failed in Mammography. *Journal of the American College of Radiology*, Vol. 15, No. 3, pp. 535–537.

17. Lång, K., Nergården, M., Andersson, I. et al. (2016). False positives in breast cancer screening with one-view breast tomosynthesis: An analysis of findings leading to recall, work-up and biopsy rates in the Malmö Breast Tomosynthesis Screening Trial. *Eur Radiol*, Vol. 26, No. 11, pp. 3899–3907.
18. Liga Colombiana contra el cáncer. (s/a) La Mamografía: "Un derecho de las mujeres para la detección temprana del cáncer de mama". Consultado el 20 de mayo de 2021 en <https://www.ligacancercolombia.org/wp-content/uploads/2020>
19. Löfstedt, T., Brynolfsson, P., Asklund, T., Nyholm, T. y Garpebring, A. (2019). Gray-level invariant Haralick texture features. *PLOS ONE*, Vol. 14, No. 2, pp. 1-18.
20. Margaliot M. (2008) Pattern recognition. (Theodoridis, S. and Koutroumbas, K.; 2006) [book reviews]. *IEEE Trans Neural Network*, Vol. 19, No.2, pp. 376
21. MathWorks (s/a) RegionProps. Consultado el 1 de junio de 2021 en <https://www.mathworks.com/help/images/ref/regionprops.html>
22. Menéndez, R., López-Potilla, B., Duarte, A. y Orea, I. (2016) Detección de masas en mamografía asistida por computadora. *Revista Cubana de Informática Médica*, Vol. 8, No. 3, pp. 582-594.
23. Organización Mundial de la Salud. (2018, Septiembre) Cancer. Consultado el 20 de Mayo de 2021 en <https://www.who.int/news-room/fact-sheets/detail/cancer>
24. Osorio, D. I. (2015) Procesamiento y análisis de mamografías para la detección de tumores. Escuela Técnica Superior de Ingeniería Informática.
25. Phung, M. T., Tin, S. y Elwood, J. M. (2019). Prognostic models for breast cancer: a systematic review. *BMC cancer*, Vol. 19, No. 1, pp. 230.
26. Ranchod, T. y Herndon, J. (2019) A Comprehensive Guide to Breast Cancer. Consultado el 20 de mayo de 2021 en <https://www.healthline.com/health/breast-cancer>
27. Reeder J.G. y Vogel V.G. (2012) "Breast cancer prevention". *Advances in Breast Cancer Management*, Second Edition. Springer.
28. Román, R., Sala, M., De La Vega, M. et al. (2011). Effect of false-positives and women's characteristics on long-term adherence to breast cancer screening. *Breast Cancer Res Treat*, Vol. 130, No. 2, pp. 543-552.
29. Romieu I., Biessy C., Torres-Mejía G., Ángeles-Llerenas A., Sánchez G.I., Borrero M., et al. (2019). Project profile: a multicenter study on breast cancer in young women in Latin America. *Salud Publica Mex*. Vol. 61, No. 5, pp. 601-608.
30. Sayeed, A., Bhattacharya, P., Mudur, S. P. y Krishnamurthy, S. (2011). Local relative GLRLM-based texture feature extraction for classifying ultrasound medical images. *Proceedings of the 24th Canadian Conference on Electrical and Computer Engineering, CCECE 2011, Canada*.
31. Saslow, D., Hannan, J., Osuch, J., Alciati, M.H., Baines, C., Barton, M., et al. (2004) Clinical breast examination: practical recommendations for optimizing performance and reporting. *CA: A Cancer Journal for Clinicians*, Vol. 54, No. 6, pp. 327–344.
32. Suárez, E.D., Pérez, M. y Orozco, R. Diseño de Sistema Automatizado para Detección de Anomalías en Imágenes Digitales de Mama. *Journal health med. sci*. Vol. 5, No. 4, pp. 229-243.
33. Susan G. Komen ORG. (s/a). Breast Cancer Information. Consultado el 20 de Mayo de 2021 en <https://www.komen.org/breast-cancer/>
34. Taglicfacio, A. S., Piana, M., Schenone, D, Lai, R., Massone, A. M. y Houssami, N. (2020). Overview of

- radiomics in breast cancer diagnosis and prognostication. *The Breast*, Vol. 49, pp. 74-80.
35. Ventura-Alfaro, C. (2018). Errores de medición en la interpretación mamográfica por radiólogos. *Revista de Salud Pública*, Vol. 20, No. 4, pp. 518-524.
36. Weingart S.N., Saadeh M.G., Simchowitz B, Gandhi T.K., Nekhlyudov L. et al. (2009) Process of care failures in breast cancer diagnosis. *Journal of General Internal Medicine*, Vol. 24, No. 6, pp. 702-709.
37. Wiesner, C., Diaz, S., Sánchez, O., Puerto, D., Bravo, L.E. y Murillo, R. (2020). Políticas basadas en la evidencia científica: El caso del control del cáncer de mama en Colombia. *Revista Colombiana de Cancerología*, Vol. 24, No. 3, pp.103-112.
38. Zielonke, N., Gini, A., Jansen, E. E. L., Anttila, A., Segnan, N., Ponti, A. y Prialux, J. (2020). Evidence for reducing cancer-specific mortality due to screening for breast cancer in Europe: A systematic review. *European Journal of Cancer*, Vol. 127, pp. 191-206.

IJMSPHR