University of Sistan and Baluchestan

Journal of Epigenetics

Online ISSN: 2676 - 4350

Homepage: https://jep.usb.ac.ir

Iranian Genetics Society

Breast Cancer Detection Using Deep Multilayer Neural Networks

Mohammadmehdi Keikha[1,*], **Yahya Kord Tamandani**[2]

[1] *Corresponding author*, Department of Computer Science, University of Sistan and Baluchestan ,Zahedan, Iran. E-mail: keikha@cs.usb.ac.ir, Tel.: +989128396846.

[2] Department of Computer Science, University of Sistan and Baluchestan ,Zahedan, Iran. E-mail: yahya.kord@cs.usb.ac.ir

ARTICLE INFO

Article type:
Research Article

Article history:
Received: 23 February 2022
Revised: 09 March 2022
Accepted: 10 April 2022

Keywords:
Multitask learning,
Deep learning,
Convolutional Neural Network (CNN),
Breast cancer,
Machine Learning.

ABSTRACT

Breast cancer is the most common cancer among women and is the second leading cause of death. There is currently no efficient way to prevent breast cancer, but its detection in early stages can increase the patient's chances of being cured and surviving. Computer-aided diagnosis (CAD) systems, based on image processing techniques, can provide a more reliable interpretation of mammographic images to detect microcalcifications and have been able to identify and classify benign and malignant tumors. If we are dealing with a massive number of images, this system increases the ability and accuracy of detection. Also, in cases where the number of images is not large, CAD systems can significantly improve the image quality. In addition, a CAD system can identify suspicious areas to provide radiologists with a visual aid to interpret mammograms. Deep learning and convolutional neural networks have recently shown significant performance for visual applications. Convolutional neural networks have also been used efficiently to analyze medical images and diagnose mammograms. In this paper, a CAD system based on convolutional neural networks (Mask R-CNN) with multi-task learning to detect breast cancer and segment mammogram images is proposed. The Mask-RCNN technique is one of the strongest and most flexible deep grids ever designed for machine vision. In this article, multitask learning with the integration of two tasks of classification and segmentation is used to diagnose breast cancer. R-CNN convolution neural network is used to diagnose cancerous mass. This system consists of two main stages, including the production of pseudo-color image and segmentation-detection based on convolutional neural networks (R-CNN Mask). The INbreast dataset is employed for evaluation of the proposed method.

Introduction

Breast cancer is the most prevalent malignancy among women and is the second leading cause of death. There is currently no practical way to prevent breast cancer. Early detection of this cancer can increase the patient's chances for successful treatment. Cancer has been one of the greatest dangers to human life for many years. Cancer has been one of the leading causes of death in recent decades. According to the World Health Organization, 13% of the world's deaths in 2008 were due to cancer. Among the known cancers, breast cancer is the most common cancer among women. Statistics show that the newly diagnosed cases with this type of cancer are increasing in both developed and developing countries. In 2022, an estimated 287,850 new cases of invasive breast cancer are expected to be diagnosed in women in the U.S, and About 43,250 cases are expected to lose their lives. According to the World Health Organization, Breast cancer became the most common cancer

DOI: 10.22111/jep.2022.41712.1041

Publisher: University of Sistan and Baluchestan

How to Cite: Keikha, M., Kord Tamandani, Y. (2022). Breast Cancer Detection Using Deep Multilayer Neural Networks. Journal of Epigenetics, 3(1), 27-34. DOI: 10.22111/jep.2022.41712.1041

globally as of 2021, accounting for about 12% of all new annual cancer cases worldwide. There is currently no effective way to prevent breast cancer, but early detection can increase a patient's chances of being treated and surviving. International attempts and experiences show that this cancer can be well controlled by screening. Screening means evaluating seemingly healthy individuals to detect cases of disease (Di cea et al., 2020).

Screening is done in various ways, which include:

* Breast self-examination (self-examination on the seventh to the tenth day of menstruation in people over 20 years)

* Clinical examination (By a healthcare professional)

* Chest imaging.

80 to 90 percent of breast cancers are diagnosed with routine physical examinations. Prior to the advent of mammography screening, approximately 2% of breast cancers were diagnosed by individuals. The size of these tumors, in such cases, is usually 2 to 5 cm. It is important to note that attempting screening by physical examination alone can reveal lesions when they are clinically palpable. It is important to note that screening alone by physical examination can detect lesions when they are clinically palpable. It should also be noted that there is a pre-clinical stage for at least two years in which the lesion is not yet palpable and this increases the risk of metastatic cancer. (Le et al, 2019). Mammography detects breast cancer before the lesions become palpable in the preclinical stage. Naturally, most of what looks abnormal on a mammogram are not malignant. Research shows that approximately 10% of mammograms that are interpreted abnormally require more diagnostic procedures, of which at least 20% will require open or needle biopsy for a definitive diagnosis (Le et al, 2019). Breast cancer diagnosis can also be done through different methods such as anthropometric data (Y.K.T, 2019).

Computer-aided detection (CAD) systems are a set of image processing, pattern recognition, and artificial intelligence technologies that can assist specialists in cancer diagnosis. In this regard, we can refer to the diagnosis of cancerous tumors in different areas such as skin, blood, uterus, breast, early diagnosis of epilepsy, analysis of brain signals in various diseases. Despite the different methods of breast imaging, mammography is still the most common and effective method of diagnosing breast cancer.

CAD systems help radiologists diagnose lesions such as calcifications and lumps from early mammography images. As mentioned earlier, malignant breast cancer is one of the leading causes of death in women and, if diagnosed quickly and appropriately the mortality rate can be significantly reduced. The most efficient method in diagnosing breast cancer is mammography. Unfortunately, mammography has poor performance in diagnosing cancer, especially in classifying benign and malignant masses. That is why many suspected mammograms lead to biopsy (tissue biopsy), while only 10 to 15% of women who undergo biopsy have malignant breast cancer, which can be costly, stressful, and uncomfortable for women. Be a reason not to pay attention to the necessary annual mammography examinations. CAD systems, based on image processing techniques, can provide a more reliable interpretation of mammographic images to detect microcalcifications and have been able to identify and classify benign and malignant tumors and thus solve many of these issues. If dealing with a massive number of images, this system increases the ability and accuracy of detection. Also, in cases where the number of images is not large, a CAD system can significantly enhance the image quality. In addition, a CAD system can identify suspicious areas for presentation to radiologists for a closer look. Therefore, this system can replace the second radiologist. Also, if a radiologist needs to test each mammogram two or more times, the CAD system can perform multiple tests. In this paper, a CAD system based on a convolutional neural network with multi-task learning is proposed to detect benign and malignant tumors in breast mammogram images. The convolution neural network (CNN) is a hierarchical neural network applied to two-dimensional images that integrates feature extraction and classification processes into a single, fully adaptive structure. This structure can automatically extract primary two-dimensional properties. It is also resistant to geometric and local distortions in input images and noise and disturbances. Deep learning and convolution neural networks have recently shown significant performance for visual applications. Convolutional neural networks have also been used successfully for medical image analysis and diagnosis in mammography (Karpathy, 2016). This paper uses a multi-task learning (MTL) approach to improve the performance of convolutional neural networks, which combines segmentation and categorization tasks. This neural network model based on multi-tasking learning is applied to the breast cancer dataset.

The main objectives of this article are:

* Diagnosis of benign and malignant masses of breast cancer

* Improving the performance of breast cancer diagnosis with the help of a multifunctional deep learning approach

* Simultaneous diagnosis and segmentation of breast cancer masses.

Related work

There are various imaging methods including mammography, MRI, ultrasound, and computed tomography. Mammography is the most common imaging method because it is inexpensive and can accurately determine the extent of the mass in a breast (Le et al, 2019). One of the symptoms of breast cancer is microcalcification deposits. These small deposits in mammographic images appear as small clear spots, usually single or clustered. A cluster can be associated with a malignant or benign complication. Due to the low resolution of mammographic images, the microcalcifications in these images may not be well visible and make it very difficult to analyze and evaluate the mammographic images. Research has shown that about 10 to 30 percent of the glands in a patient's breast are not detectable by a radiologist on mammographic images. CAD systems are a set of image processing, pattern recognition, and artificial intelligence technologies that can assist specialists in cancer diagnosis. CAD systems, based on image processing techniques, help radiologists diagnose lesions such as calcifications and lumps from early mammography images. The convolutional neural network (CNN) is a class of deep learning neural networks. CNNs represent a significant discovery and development in image recognition research. They are commonly used to study and interpret visual imagery in healthcare and other domains. (Mostavi et al, 2020) have demonstrated that accurate prediction of cancer type is essential for cancer treatment. Important marker genes can be extracted through a predictive model. In this research, several convolutional neural network models are proposed that receive unstructured gene expression inputs for classifying tumor and non-tumor samples as cancerous and normal cases. Numerous studies have attempted to use machine learning models for this purpose. However, no recent research has considered the effects of primary cancer tissue that can interfere with the diagnosis of cancer markers. (Samala et al, 2017) have shown that transfer learning in deep convolutional neural networks (DCNN) is a critical step in its application for medical imaging tasks. In this paper, DCNN multitasking transfer learning aims to transfer learned knowledge from non-medical images to medical diagnostic tasks through regulatory training and enhancement. (Le et al, 2019) have used a multi-training learning approach that combines segmentation and categorization to improve the performance of convolutional neural networks. The results have shown that joint training of classification and segmentation has provided better cooperation between tasks. (Di cea et al, 2020) have proposed a multi-training learning approach that trains the neural network for the two functions, object recognition, and image classification. The results show that this approach can assist radiologists in breast screening.

DATASET

Mammography images of INbreast database was collected from Centro Hospitalar de S. Joao [CHSJ], Breast center, Porto. INbreast database gathers data from Aug. 2008 to July 2010, which holds 115 cases with a total of 410 images. 90 cases were women having disease on both breasts. There are four types of breast diseases noted in the database, including Mass, Calcification, Asymmetries, and Distortions.

Table 1. INbreast database specifications.

Number of cases	115
Number of mammogram images	410
Number of masses	116
Pixel size (micrometers)	70
Contrast resolution (Bits)	14
Lesion size range (square millimeters)	[15,3689]

Methodology

4.1 Convolutional neural networks:

Biologically inspired algorithms such as convolutional neural networks (CNNs) are one of the best deep learning methods in the field of computer vision, as their results are comparable to the human ability to track and detect objects (Weinland, 2011). Before the age of deep learning, conventional object recognition techniques were based on hand-coded features that did not have the strength and stability required for different conditions and failed to detect and track objects by changing orientations (Venkatesan, 2018). CNNs have proven to be faster and more accurate algorithms. They are one of the most important deep learning methods, confirmed to be effective in issues such as image recognition in practice (Chauhan et al, 2016). CNN enhances the concept of perceptual space and shared weights, which not only greatly reduces training parameters but also reduces the complexity of the network model. The properties of each layer are created from the local area of the previous layer (perceptual area) by sharing the weight of the convolutional filter. These features make the network more suitable for learning and displaying image features compared to other neural networks.

4.2 Multi-task deep learning:

Collecting sufficient cancer samples is costly and time-consuming, leading to insufficient cancer tissue samples. Compared to single-task learning and transitional learning, multi-task learning can use cancer datasets to improve the performance of cancer classifications when the number of training samples is relatively insufficient. Recently, transfer learning and multi-task learning have been proposed that can solve this problem in computer vision and bioinformatics (Liao et al, 2019). Multi-task Learning (MTL) is close to transitional learning, it attempts to learn multiple datasets or different tasks simultaneously, even if the datasets are different. It requires that all tasks perform well as all datasets are important for the model (López-García et al, 2020).

4.3 Proposed method architecture:

Breast mammography cancer classification and segmentation are usually performed as separate tasks. In this article, multi-task learning is used to diagnose breast cancer and mammography by combining the two tasks of classification and segmentation. Recursive convolution neural network (R-CNN) is used to diagnose cancerous mass. In this paper, CAD detection system is used to detect mammographic mass and segmentation simultaneously.

The system consists of two main stages, first the production of pseudo-color image and then segmentation-detection based on the R-CNN, as shown in Fig.4. One of the innovations is the introduction of a pseudo-color mammogram (PCM), which creates quasi-mass patterns with color contrast to the background. The pseudo-color mammogram is created by adding two morphologically filtered mammograms to the grayscale mammogram in two adjacent image channels as shown in Figure 1. Multitasking with R-CNN is used to identify masses on PCMs. The proposed R-CNN deep learning quasi-color scheme provides a coherent solution for the diagnosis and segmentation of mammographic masses that does not require manual intervention. exists.

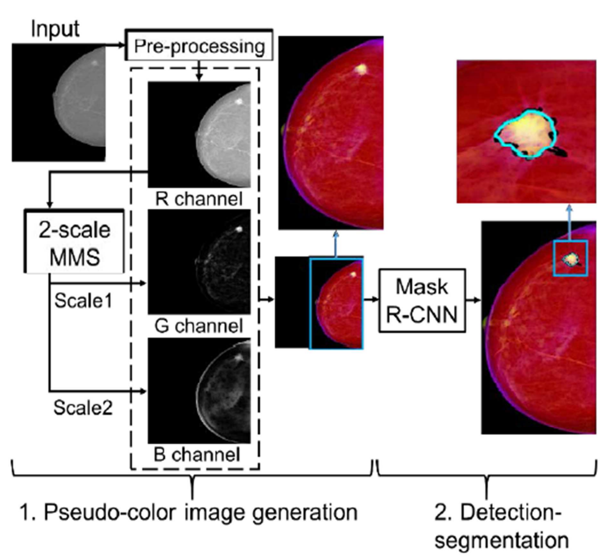

Fig. 1. Framework of the proposed method.

The proposed research method consists of following steps:

1. Pre-processing

2. Creating a pseudo-color image

3. Recognition and segmentation simultaneously with R-CNN Mask learning technique

The method converts gray mammograms to pseudo-colored mammograms using multidimensional morphological sifting (MMS). The MMS method can improve lesion-like patterns within a specified size range. Two scales are used in MMS and two output images are created.

The two images are appended to the gray mammogram to form an RGB pseudo-color image. Lesion-like patterns will show color contrast with the background in the new pseudo-color image, as shown in Fig. 2.

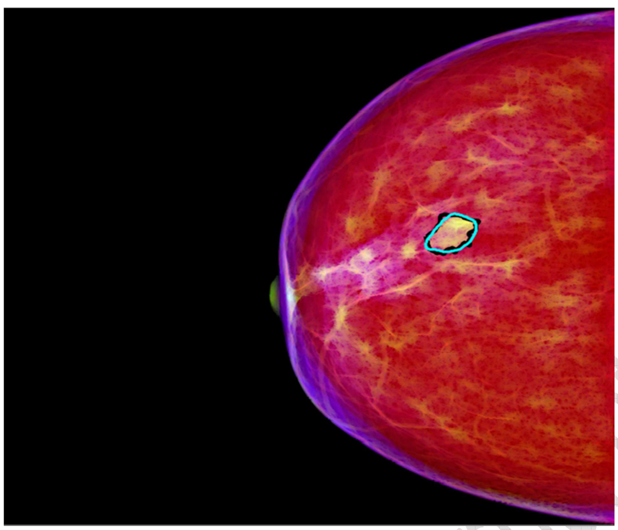

Fig. 2. Pseudo-color mammogram image to show the mass.

The black line in Figure 5 represents the mass margin and the greenish-blue line represents the segmentation created by the proposed method. Pigmented mammograms are then used as input for the R-CNN Mask Deep Learning technique. The R-CNN Deep Learning Mask technique is trained to perform the two tasks of detecting and segmenting breast masses simultaneously. Therefore, we expect the use of pseudo-color mammograms to improve diagnostic and segmentation performance using the R-CNN Mask technique compared to gray mammograms.

A color-coded mammogram consists of a gray mammogram (GR) in the red channel (R), an MMS output image of scale 1 in the green channel (G), and an output image of scale 2 in the blue channel (B) as shown in Fig. 3.

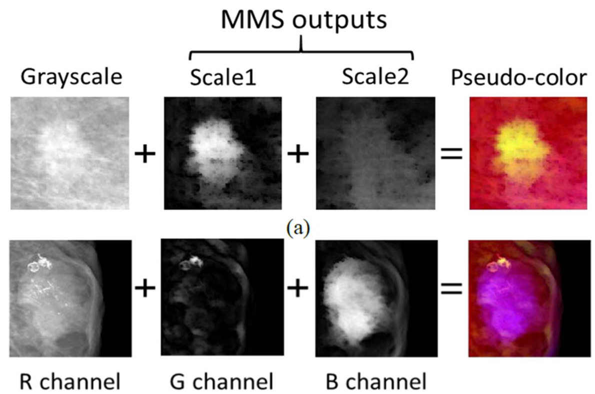

Fig. 3. Pseudo-color improvement for masses of different sizes. (a) Represents a small mass with a diameter of 6 mm; (b) a large diameter with a diameter of 27 mm.

If a mass is relatively small and in the size range of scale 1, it will have higher intensity in the green (G) channel and more yellow appearance on the pseudo-colored mammogram, shown in Figure (6-a). On the other hand, if a mass is rather large and within the size range of scale 2, it will have higher intensity in the blue channel (B) and appear purple on the PCM, indicated in Figure (6-b). This quasi-color scheme can improve mass-like patterns by adding contrast between the mass and the background when combining three channels.

Result

In order to evaluate the performance of the proposed method, the INbreast database has been used. The database contains 410 images and includes various injuries such as masses, calcifications, asymmetries, and tissue damage, collected at the CHSJ Research Center in Porto under the supervision of the Hospital Ethics Committee and the National Committee for Data Protection from April 2008 to July 2010.The database contains 115 cases with 410 mammograms. Including a total of 116 breast cancer masses, with sizes ranging from 15 mm2 to 3689 mm2. The pixel size of mammograms is

70 microns, and the depth is 14 bits. 60% of datasets are used for training, 20% for validation, and 20% for experimental data.

A ROC curve, also known as a receiver operating characteristic curve, is a graph showing the performance of a classification model. Using the results obtained from the ROC curve, concepts such as cut-off point, sensitivity, and specificity of a diagnostic test can be obtained. The ROC diagram of the proposed method is shown in Fig. 4.

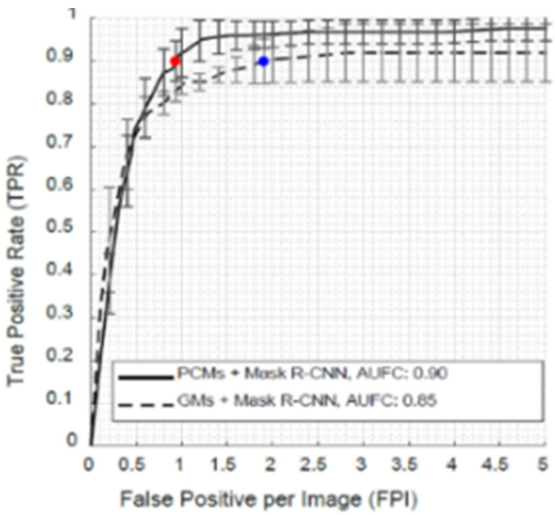

Fig. 4. The ROC diagram of the proposed method

The ROC curve of the PCM method with Mask R-CNN (He, k., et al 2017) is shown with the solid line, and the dashed line indicates the ROC curve of the GM method with Mask R-CNN. For the red dot on the PCM + Mask R-CNN curve, the sensitivity is 0.9, 1-specificity is 0.9, and the area under the curve (AUC) is 0.9. in the case of the blue dot on the GM + Mask R-CNN rock curve, the sensitivity is 0.9, 1-specificity is 0.98, and the AUC is 0.85. Therefore, the diagnostic power of the PCM + Mask R-CNN method is higher than GM + Mask R-CNN method. Dice similarity index (DSI) is one of the main criteria for evaluating the results of image segmentation. This value indicates how similar the results of the segmentation model are to the actual texture mask. The result of this DSI is a number between 0 and 1, value closer to 1, indicates there is more similarity between the two images. The DSI was 0.88 for the PCM + Mask C-NN method and 0.87 for GM + Mask R-CNN method. Performance comparison of the proposed methods with other methods are given in Table 2 and Fig. 5.

Table 2. Comparison of mass detection and segmentation performance of various methods..

Method	Sensitivity	1-specificity	DSI
PCM+Mark R-CNN	0.90	0.90	0.88
GM+Mask R-CNN	0.90	0.98	0.87
Min, H. et al (2019)	0.90	0.81	0.86
Dhungel, N. et al (2017)	0.90	0.13	0.85
Oliveira,H. S. et al (2019)	0.85	0.05	0.83
Li, Y. et al (2018)	0.90	0.88	-

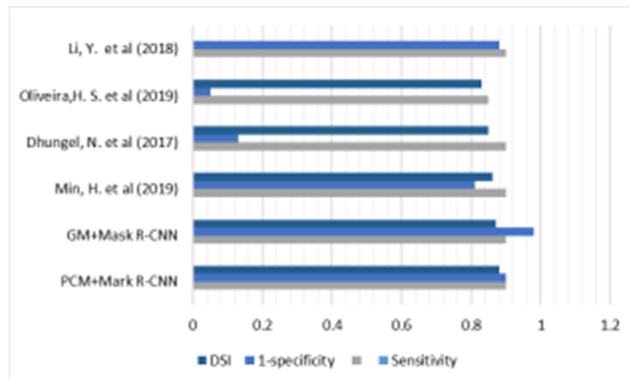

Fig. 5. Comparison of mass detection and segmentation performance of various methods

Conclusion

In this paper, a CAD system based on convolutional neural networks with multitask learning to detect and segment mammogram images is proposed. This study aimed to use features that are hard or cannot be visually assessed by a physician, but neural networks can be trained and respond to them to assist the physician in making the final decision. To improve the performance of convolutional neural networks, this paper uses the multitask learning approach, which combines segmentation and categorization tasks. The neural network model based on multitask learning is applied to the breast cancer dataset. Breast mammography cancer classification and segmentation are usually performed as separate tasks. In this article, multitasking is used to diagnose breast cancer and mammography by combining the two tasks of classification and segmentation. R-CNN convolution neural network is used to diagnose cancerous mass. Detection of mammographic mass and segmentation are done simultaneously.

Acknowledgements

The authors declare no conflict of interest.

References

Chauhan, R., Kaur, H., & Alam, M. A. (2010). Data clustering method for discovering clusters in spatial cancer databases. *International Journal of Computer Applications*, *10*(6), 9-14.

De Cea, M. V. S., Diedrich, K., Bakalo, R., Ness, L., & Richmond, D. (2020, October). Multi-task Learning for Detection and Classification of Cancer in Screening Mammography. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 241-250). Springer, Cham.

Dhungel, N., Carneiro, G., & Bradley, A. P. (2017). A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Medical image analysis*, *37*, 114-128.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

Karpathy, A. (2016). Cs231n convolutional neural networks for visual recognition. *Neural networks*, *1*(1).

Le, T. L. T., Thome, N., Bernard, S., Bismuth, V., & Patoureaux, F. (2019). Multitask classification and segmentation for cancer diagnosis in mammography. *arXiv preprint arXiv:1909.05397*.

Li, Y., Chen, H., Zhang, L., & Cheng, L. (2018, August). Mammographic mass detection based on convolution neural network. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 3850-3855). IEEE.

Liao, Q., Ding, Y., Jiang, Z. L., Wang, X., Zhang, C., & Zhang, Q. (2019). Multi-task deep convolutional neural network for cancer diagnosis. *Neurocomputing*, *348*, 66-73.

López-García, G., Jerez, J. M., Franco, L., & Veredas, F. J. (2020). Transfer learning with convolutional neural networks for cancer survival prediction using gene-expression data. *PloS one*, *15*(3), e0230536.

Min, H., Chandra, S. S., Crozier, S., & Bradley, A. P. (2019). Multi-scale sifting for mammographic mass detection and segmentation. *Biomedical Physics & Engineering Express*, *5*(2), 025022.

Mostavi, M., Chiu, Y. C., Huang, Y., & Chen, Y. (2020). Convolutional neural network models for cancer type prediction based on gene expression. *BMC Medical Genomics*, *13*, 1-13.

Oliveira, H. S., Teixeira, J. F., & Oliveira, H. P. (2019, September). Lightweight deep learning pipeline for detection, segmentation and classification of breast cancer anomalies. In *International Conference on Image Analysis and Processing* (pp. 707-715). Springer, Cham.

Samala, R. K., Chan, H. P., Hadjiiski, L. M., Helvie, M. A., Cha, K. H., & Richter, C. D. (2017). Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms. *Physics in Medicine & Biology*, *62*(23), 8894.

Venkatesan, R., & Ganesh, A. B. (2018). Deep recurrent neural networks based binaural speech segregation for the selection of closest target of interest. Multimedia Tools and Applications.

Weinland, D., Ronfard, R., & Boyer, E. (2011). A survey of vision-based methods for action representation, segmentation and recognition. *Computer vision and image understanding*, *115*(2), 224-241.

Yahya Kord Tamandani (2020) Early Detection of Breast Cancer in Women Using a Cost-effective Procedure. *journal of Epgenetics Volume 2, Issue 1, Pages 1-6*.