Journal of Engineering and Computer Sciences, Qassim University
Vol. 13, No. 1, pp. 31-43 (Jan. 2020 /Jumada Al-Awwal 1441H)

Classification for Breast Cancer Detection

Suhilah Alkhalifah[1]and Azza Alnssayan[2]
Computer Science Department College of Computer, Qassim University
[1]s.alkhalefah@qu.edu.sa
[2]a.alnssayan@qu.edu.sa

(Received 1/6/2020, accepted for publication 21/9/2020)

Abstract. Artificial intelligence is a branch of computer science capable of analyzing complex medical data. Their potential to exploit a meaningful relationship within a data set can be used in the diagnosis, treatment and predicting outcome in many clinical scenarios. This paper compares Artificial Neural Networks (ANN) applications like Mammogram, ultrasound, Thermal Imaging, MRI. It also shows the accuracy of Artificial Neural Networks (ANN) to detect breast cancer using the WEKA tool.

Keywords: Breast, Cancer, Artificial intelligence, and Artificial Neural Networks.

1. Introduction

Artificial intelligence (AI) has become widely used in medical applications in recent decades. It appears by medical devices that work with embedded AI algorithms. [1]

AI is a combination of algorithms that can learn attributes from data. Classification is one of the most important tasks of AI. Classification is a process of categorizing a given set of data into classes. [2]

Breast Cancer is one of the most types of cancers which infects women, that is because of a lot of reasons; for example, changes in lifestyle and hormonal disorders. The classification algorithms are one of the most used AI algorithms for breast cancer detection, and it uses the thermogram picture to detect benign and premalignant tumour [2]. Hence, information technology has created diverse computerized strategies for simple

and speedy diagnosis of breast cancer, out of which Artificial Neural Networks (ANN) is the most noteworthy strategy [3].

2. Methodology

2.1. Systematic literature review (SLR)

A systematic literature review (SLR) identifies, evaluates and critically interprets research relevant to a specific question. The systematic review has a clear methodological plan. The methodological steps, search strategy and research question are explicitly defined so that it can be replicated and reproduced by other researchers [4].

2.2. Research question

The general research question is: How to detect breast cancer using Artificial Neural Networks.

The search string used to identify primary studies formalized adoption the criterion P-I-C-O described in [5]. our research question has the following elements:

- Population: breast cancer.
- Intervention: Artificial Neural Networks.
- Outcomes: breast cancer detection.

2.3. Generating a search strategy

Our search strategy was searching by one search term: Artificial intelligence in breast cancer, with double quotation marks on "breast cancer". This search term used in Google Scholar database.

We did filter the search results based on publication year as we wanted to be as up to date as possible. Starting from 2020, then we read the first ten pages (nearly), then go to the previous year. Table 1 represents the number of hits for different years.

Table (1): The number of hits.

keyword	year	result	Read
Artificial intelligence in breast cancer	2020	11,500	14
	2019	14,000	13
	2018	11,700	16
	2017	8,640	14
	2016	7,710	16
	2015	7,390	19
	2014	7,070	22
	none	55,600	22

2.4. Study selection criteria

To identify the primary studies that are directly related to the research question. We used (Mendeley) to list exclusion criteria that:

- Old paper.
- Books/book chapters.

The study was followed by a number of steps. First, we collected papers by reading the titles and abstracts of papers. Then excluding papers not relevant to detect breast cancer using Artificial Neural Networks. For each paper, we classified it as being either relevant, irrelevant or not sure, based on the stated exclusion criteria. Each article was read by two of us. In case of disagreement among us, the paper marked as not sure. As a result of this step, we were left with 30 irrelevant and 18 relevant papers, as shown in Table 2.

The study was selected when it met both inclusion criteria and eliminated if it matched any of the exclusion criteria.

Table (2): Count of papers which is Irrelevant and relevant by our research question

Source	Irrelevant Paper	Relevant Paper
Scholar.Google	**30**	**18**

3. Data Synthesis

After we explored the previous studies on Artificial Neural Network and find it is the most popular AI technique in medicine. We come out with the different applications of ANN are used to detect breast cancer. These applications will explore in the first phase. Then, in the second phase, we will compare these applications.

In Figure 1, it shows the classification of the topics we come out from our primary studies. These topics are ANN, ANN importance, ANN fields, ANN applications and breast cancer.

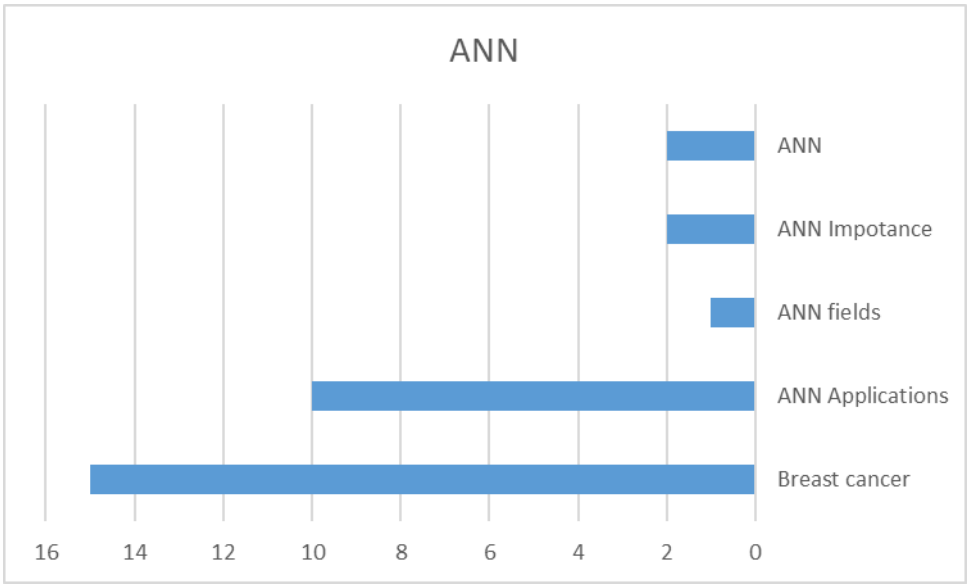

Fig. (1). A categorization of topics

Here we are summarizing the content of our papers: [6] "Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer": some properties of Mammogram, Ultrasound, Thermogram and Magnetic Resonance Imaging techniques. [1] "The use of artificial neural networks in decision support in cancer: A systematic review": Artificial Intelligence in medical. [7] "Artificial intelligence in medicine": ANN is the most popular AI technique in medicine. [3] "Artificial Neural Network: As Emerging Diagnostic Tool For Breast Cancer": Various new prognostic factors have identified using ANNs statistical models to discriminate benign and

malignant cancers. [8] "Artificial neural networks in medical diagnosis": ANNs use in medicine in: Cardiovascular diseases, Cancer and Diabetes. [9] "A Novel Approach for Breast Cancer Detection using Data Mining Techniques": what is breast cancer and type of tumors. [10] "Application of artificial neural networks to the analysis of dynamic MR imaging features of the breast": how the diagnosis to be done appropriately in MRI. [11] " Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks": disadvantage of Mammography strategy. [12] "Automated breast cancer detection and classification using ultrasound images: A survey": advantages of ultrasound. [2] "Infrared imaging technology for breast cancer detection – Current status, protocols and new directions" artificial intelligence algorithms in breast cancer detection. [13] "A comparative review of thermography as a breast cancer screening technique": some properties of thermography and Mammogram. [14] "The Effect of Thermography on Breast Cancer Detection": breast cancer detection techniques. [15] "MRI for breast cancer screening, diagnosis, and treatment": the advantage of MRI. [16] " Detection of Breast Cancer using Data Mining Tool (WEKA)": The importance of data mining for predict of breast cancer. [17] "Artificial Intelligence in the Interpretation of Breast Cancer on MRI": the advantage of MRI. [18] "Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art": The new release of mammography. [19] “Diagnosis of Breast Cancer using Decision Tree and Artificial Neural Network Algorithms”: The performance of the Artificial Neural Network for breast cancer prediction. [20] “Ultrasound for Breast Cancer Detection Globally: A Systematic Review and Meta-Analysis”: Importance of using Ultrasound early detection of breast cancer.

4. Breast Cancer

Cancer that develops from breast tissue is Breast Cancer. Breast Cancer caused when cancerous tumor emerges in the breast and when the tumor matures, it spread to more parts of the body. Tumors in the breast can be benign (not cancer) or malignant (cancer):

- Benign tumors: do not spread to different parts of the body and can be removed.
- Malignant tumors: maybe a danger to life [9].

To separate benign and malignant cancers was used ANN statistical models to identify different new prognostic variables and to present novel strategies [3].

5. Artificial Neural Networks (ANNs)

In medicine, ANN has been the most popular AI technique in the last two decades. ANNs are computational analytical instruments which are enlivened by the biological nervous system. They comprise of networks of very interconnected computer processors called 'neurons' that equipped for performing parallel calculations for data preparing and knowledge representation. Their capacity to gain from historical examples, analyses of the non-linear data, handle imprecise information and generalizes enabling application of the model to independent data has made them an exceptionally alluring explanatory apparatus in the field of medicine [7].

5.1 Overview of ANNs in Medical Diagnosis

ANNs models can be used in medicine in:

1. Cardiovascular diseases

ANNs have considerable potential in CVD determination because of the classification accuracies higher than 90% are achieved.

2. Cancer

ANNs have appeared to utilize different primary data, ranging from clinical parameters to biochemical values, and provide increased diagnostic accuracy for various kinds of cancers. To classify cancerous tumors, ANN is a powerful data mining technique that can be used [19].

3. Diabetes

The utilization of ANNs in the diagnosis of diabetes provides an improvement inexactness, sensitivity, and specificity in comparison with other strategies, thus contributing to improved clinical administration of diabetes mellitus [8, 10].

5.2 Applications of ANNs

5.2.1. Mammogram:

Mammography is one of the best strategies utilized in clinics for early detection of breast cancer. It has proven effective to lessen mortality as much as by 30%. The principle target of screening mammography is to early detect the malignant tumor and

remove it [6]. On the other hand, the analysis of mammographic pictures is a complex and cumbersome task which requires exceedingly specialized radiologists. The performance of the observer can be degraded by the huge resulting case-load, which is without a doubt caused by visual fatigue and other psychophysical mechanisms [11]. Also, it requires repeated exposure to radiation. Women with dense tissue have a high rate of false-positive results in terms of breast cancer [14]. Gabor wavelets and ANN utilized to classify normal and abnormal tissues which could increase the accuracy and save radiologist's time (Figure 2). Gabor wavelets changes have a useful attribute in image processing and computer vision. The outcome demonstrates that this combination of neural networks has good potential with 97% accuracy on complicated cases [6].

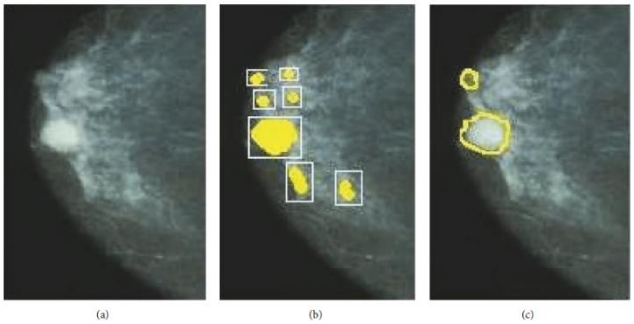

Fig. (2). Results (from (a)–(c)): original image, image after the first stage of NN processing, and image at the second stage of NN processing using Gabor wavelets as input for mammogram image

The new release of mammography is digital mammography (DM). In DM, the use of film was replaced with a digital x-ray detector. The important advantage for DM the simpler workflow. In terms of performance, DM has been shown to have improved clinical performance. The additional advantage is the ease with which the technology can be extended by developing more advanced image acquisition methods [18].

5.2.2. Ultrasound:

Neural network (NN) additionally plays a role in ultrasound images in detecting breast cancer. It detects tumor by passing sound waves on the surface of the tissue [14]. Utilization of ultrasound can increase general cancer detection by 17% and lessen the number of unnecessary biopsies by 40%. Breast ultrasound (BUS) imaging is better than the mammography in the facts: (1) Since having no radiation. (2) Ultrasound is more sensitive than mammography for detecting abnormalities in dense breasts. (3) There is a high rate of false positives in mammography, which causes a lot

of unnecessary biopsies [12]. Overall, the accuracy of ultrasound found to depend on three factors: quality of the tools, the expertise of the physician in conducting the procedure and in interpreting the image, and the use of a multidisciplinary approach for breast cancer detection [13], particularly in low-resource settings where mammography is not available, the ultrasound may be an effective primary detection tool and triage method for breast lesions [20].

5.2.3. Thermal Imaging:

Thermogram is a promising cutting edge screening instrument as it can alert women of breast malignancy up to 10 years early. A few studies have used several types of ANNs to manipulate and classify IR images by taking the IR image as an input to the ANN [6]. Breast infrared thermography is a noninvasive procedure that does not involve compression of the breast tissue or exposure to radiation, and functions through an assessment of physiological function, through high-resolution temperature measurements of breast tissue. As with mammography, there are limitations regarding the technology's ability to detect abnormalities in breast tissue. Thermography is unable to localize a lesion or tumor since abnormalities found by infrared imaging do not define an area that can be surgically biopsied [13].

5.2.4. MRI:

Magnetic Resonance Imaging technique has been utilized broadly in medical examinations, particularly for cancer investigation for a few decades. For the diagnosis to made appropriately, breast region should be extracted from other surrounding areas and tissues utilizing image segmentation methods. In the field of breast MR imaging, recent publications are focusing on the application of artificial neural networks (ANNs) for image division and computer-aided lesion classification. ANN strategy frequently uses to build computer models for medical decision-making, and it also can extract new information from complex data sets by performing multifactorial analysis. [10]

In Figure 3, numerous neural networks models were used to help MRI for enhancing the detection and the classification of the breast tumors [6].

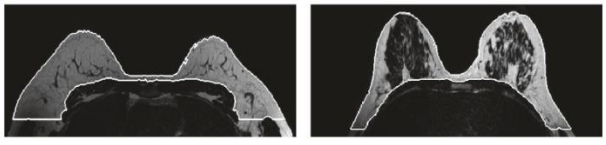

Fig. (3). Multistate CNN used to segment small fatty breast and medium dense breast for MRI image

Compared with mammography, MRI has a higher sensitivity for the detection of breast cancer and is not affected by breast density [15, 17]. Furthermore, continued development in radiomic MRI features may lead to earlier initiation of therapeutic plans, deeper understanding of tumor prognostics, and improved patient outcomes [17].

6. Results

There are so many Digital image processing tools for breast cancer detection like Mammogram, ultrasound, Thermal Imaging, MRI etc. Mammogram is the most widely used technique that used a screening tool, but it is not a perfect one since it is a complex and cumbersome task which requires highly specialized radiologists. Ultrasound is more convenient and safer than mammography since it has no radiation. Also, it is cheaper and faster than mammography, but its accuracy depends on the quality of the tool and physician knowledge in using the multidisciplinary approach. Thermogram has got more accuracy and sensitivity when combines with mammography but with a high false-positive rate and false-negative rate. Ultrasound, MRI is recommended for the positive mammogram.

Overall, Kennedy et al. [13] state that there is no single tool that provides excellent predictability over breast cancer detection; the only definitive diagnostic tool is a biopsy.

6.1 Prediction of Breast Cancer Using WEKA Tool

By using data mining, we can predict the occurrence of breast cancer most efficiently. There has to be the availability of precise and accurate data so that a model with accurate model helps the doctors to predict and diagnose cancer, whether it is benign or malignant at the early stage. This will save time for the physicians and improve their efficiency. [16]

Here we apply ANN technique on breast cancer datasets in WEKA tool. The data used is breastcancer.arff which located in WEKA tool, with a total number of instances = 286. The confusion matrix will show the accuracy of ANN.

Table (3): Performances of the accuracy of ANN with its Confusion Matrices

	Predicted (→)		TP (%)	FP (%)	ROC Area	*Precision%*
Actual (↓)	**no-recurrence-events (Negative)**	**recurrence-events (Positive)**				
no-recurrence-events (Negative)	TN= 150	FP= 51	0.746	0.588	0.623	0.623
recurrence-events (Positive)	FN= 50	TP= 35	0.412	0.254	0.623	0.623
Weighted Average			0.647	0.489	0.623	0.648

In Table 3, TP is the proportion of positive cases that correctly identified, and TN is the proportion of negative cases that correctly identified. FP and FN are also called type I error and type II error, respectively. TP (%) and FP (%) are the rates of TP and FP, respectively. ROC examines the performance of classifiers (Swets, 1988) as an additional way besides the confusion matrix. It examines with the false positive rate and the true positive rate. The higher the rate or ROC area is better the performance. Precision or positive predictive value is an assessor in the confusion matrix. The higher the precision is the better the option. "Weighted Average" is weighting the results based on the sample sizes for each class. ANN identified as the good classifier based on the percentage of precision (0.648%) and ROC area (0.623) on the criteria of "Weighted Average" approach.

7. Conclusion

The proficiency of artificial intelligent techniques explores in almost every field of medicine. The artificial neural network was the most commonly used analytical tool.

In this paper, we represent how Artificial intelligence can analyse complex medical data. We focused on using Artificial Neural Networks (ANNs) in medical, especially in breast cancer detecting and its applications. The result compares between these applications. Finally, we apply ANN technique on breast cancer datasets in WEKA tool to show the accuracy of ANN.

8. Acknowledgement

The authors would like to thank Dr. Mohammed Mahdi Hassan for these valuable comments and suggestions, which have greatly improved the quality of the paper.

9. References

[1] Paulo J. Lisboa, Azzam F.G. Taktak, "The use of artificial neural networks in decision support in cancer: A systematic review," Liverpool, UK, 2006.

[2] Satish G. Kandlikar , Isaac Perez-Raya , Pruthvik A. Raghupathi, Jose-Luis Gonzalez-Hernandez a , Donnette Dabydeen b , Lori Medeiros c , Pradyumna Phatak d, "Infrared imaging technology for breast cancer detection – Current status, protocols and new directions," in *ELSEVIER*, USA, 2017.

[3] Ishwinder K. Sandhu, M. Nair, H. Shukla and S.S. Sandhu, "Artificial Neural Network: As Emerging Diagnostic Tool For Breast Cancer," India, 2015.

[4] Zuhal Hamad, Naomie Salim, "Systematic Literature Review (SLR) Automation: A Systematic Literature Review," in *Sudan university of science and technology, university technology Malaysia.*, Malaysia, 2014.

[5] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in Software Engineering," in *Keele University and Durham University Joint Report, Tech. Rep. EBSE 2007-001*, Durham , 2007.

[6] M. M. Mehdy, P. Y. Ng, E. F. Shair, N. I. Md Saleh, and C. Gomes, " Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer," Selangor, Malaysia, 2017.

[7] AN Ramesh, C Kambhampati, JRT Monson, PJ Drew, "Artificial intelligence in medicine," in *Ann R Coll Surg Engl 2004; 86*, UK, 2004.

[8] Alberto López-Rodríguez, Eladia M Peña-Méndez, Petr Vaňhara and Josef Havel, "Artificial neural networks in medical diagnosis," in *Journal of Applied Biomedicine*, 2013.

[9] Vikas Chaurasia, Saurabh Pal , "A Novel Approach for Breast Cancer Detection using Data Mining Techniques," India , 2017.

[10] Botond K. Szabó, Maria Kristoffersen Wiberg, Beata Boné and Peter Aspelin, "Application of artificial neural networks to the analysis of dynamic MR imaging features of the breast," in *European Radiology* , August 2004.

[11] L. Bocchi, G. Coppini, J. Nori, G. Valli, "Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks," Italy, 2004.

[12] H.D. Chenga, Juan Shana, Wen Ju, Yanhui Guo, Ling Zhang, "Automated breast cancer detection and classification using ultrasound images: A survey", an," in *Department of Computer Science, Utah State University, Logan, UT 84322, and bSchool of Mathematics*, USA and bSchool of Mathematics and System Science, Shandong University, China,, 2010.

[13] Kennedy DA, Lee T, Seely D, "A comparative review of thermography as a breast cancer screening technique," in *Integrative cancer therapies*, 2009.

[14] PR Pavithra, KS Ravichandran, KR Sekar, R Manikandan, "The Effect of Thermography on Breast Cancer Detection," in *PR Pavithra, KS Ravichandran, KR Sekar, R Manikandan School of Computing, SASTRA University*, India, 2018.

[15] Monica Morrow, Janet Waters, Elizabeth Morris, "MRI for breast cancer screening, diagnosis, and treatment," New York, USA, 2011.

[16] Jyotismita Talukdar and Dr. Sanjib Kr. Kalita, "Detection of Breast Cancer using Data Mining Tool (WEKA)," in *International Journal of Scientific & Engineering Research*, November-2015 .

[17] Deepa Sheth, MD, and Maryellen L. Giger, PhD, "Artificial intelligence in the interpretation of breast cancer on MRI," *Journal of Magnetic Resonance Imaging* , July 2019.

[18] Ioannis Sechopoulos, Jonas Teuwen, and Ritse Mann, "Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art," *Seminars in Cancer Biology*, 9 June 2020.

[19] A. Higa, "Diagnosis of Breast Cancer using Decision Tree and Artificial Neural Network Algorithms," *International Journal of Computer Applications Technology and Research,* 2018.

[20] Rupali Sood, Anne F. Rositch, Delaram Shakoor Emily Ambinder,Kara-Lee Pool, Erica Pollack, Daniel J. Mollura, Lisa A. Mullen, and Susan C. Harvey,, "Ultrasound for Breast Cancer Detection Globally: A Systematic Review and Meta-Analysis," *American Society of Clinical Oncology,* 2019.

[21] Saira Charan, Muhammad Jaleed Khan, Khurram Khurshid., "Breast Cancer Detection in Mammograms using Convolutional Neural Network," in *IEEE International Conference on Mathematics and Engineering Technologies (iCoMET)*, March 2018.

Journal of Engineering and Computer Sciences, Qassim University
Vol. 13, No. 1, pp. 31-43 (Jan. 2020 /Jumada Al-Awwal 1441H)

خوارزمية التصنيف للكشف عن سرطان الثدي

سهيلة الخليفة[1] عزة النصيان[2]
كلية الحاسب جامعة القصيم
[1]s.alkhalefah@qu.edu.sa
[2]a.alnssayan@qu.edu.sa

(قدم للنشر في 2020/6/1 وقبل للنشر في 2020/9/21)

ملخص البحث. الذكاء الاصطناعي هو فرع من علوم الحاسب الآلي له القدرة على تحليل البيانات الطبية المعقدة. يستخدم الذكاء الاصطناعي لإيجاد علاقة ذات معنى لمجموعة بيانات في التشخيص والعلاج والتنبؤ بالنتائج في العديد من السيناريوهات السريرية. توضح هذه الورقة كيفية الكشف عن سرطان الثدي باستخدام الشبكات العصبية الاصطناعية (ANN) وتطبيقاتها وستكون النتيجة مقارنة بين هذه التطبيقات.