Breast Cancer Images Classification using Convolution Neural Network

Mohammed Yahya Alzahrani

College of Computer Science and Information Technology, Albaha University, KSA

Abstract
One of the most prevalent disease among women that leads to death is breast cancer. It can be diagnosed by classifying tumors. There are two different types of tumors i.e: malignant and benign tumors. Physicians need a reliable diagnosis procedure to distinguish between these tumors. However, generally it is very difficult to distinguish tumors even by the experts. Thus, automation of diagnostic system is needed for diagnosing tumors. This paper attempts to improve the accuracy of breast cancer detection by utilizing deep learning convolutional neural network (CNN). Experiments are conducted using Wisconsin Diagnostic Breast Cancer (WDBC) dataset. Compared to existing techniques, the used of CNN shows a better result and achieves 99.66%% in term of accuracy.

Key words:
Breast cancer, deep learning, convolutional neural network, feature selection, WDBC dataset.

1. Introduction

One of the most prevalent disease among women that leads to death is breast cancer. According to [1] 1,688,780 new cancer cases and 600,920-cancer deaths occurred in the United States during 2017. The statistics show that among female cancer patients, 30% (the highest) are diagnosed as breast cancer that causing second highest death (14%).

Usually physicians diagnose breast cancer thru classifying tumors, whether they are malignant or benign types of tumors. Nevertheless, it is very difficult to distinguish tumors even by the experts. Therefore, physicians need a reliable diagnosis procedure and automation of diagnostic system to distinguish between these tumors.

There is a 30% chance that the cancer can be treated effectively if it is detected in its early stages. The treatment becomes more difficult in case of late detection of advanced-stage tumors [2,3]. The most popular techniques to detect breast cancer in early stages include surgical biopsy with almost reaches 100% correctness; Fine Needle Aspiration (FNA) using visual interpretation with the correctness level of 65% to 98% [4], and mammography with the correctness percentages of 63% to 97% [5]. Thus, the surgical biopsy is reliable; however, it is invasive and costly, while mammography and FNA with visual interpretation correctness varies widely.

Many researchers have attempted to apply machine learning algorithms for detecting survivability of cancers patients and the researchers proved that these algorithms work better in detecting cancer in the early stages. Borges [7] compares two machine learning techniques (Bayesian Networks and J48) to create two classifiers that differentiate benign from malignant breast lumps and uses Wisconsin Breast Cancer Diagnosis (WBCD) for the experiments. The author concludes that Bayesian Networks demonstrated a good performance when compared to the other algorithm, J48. (97.80% compare to 96.05% of accuracy). Furthermore, Gayathri et al. [6] summarize a survey on breast cancer diagnosis using various machine learning algorithms and methods that aim to improve the accuracy of predicting cancer. This paper discusses a diagnosis technique that uses the FNA with computational interpretation via deep learning and aims to detect breast cancer with a high level of accuracy and a low rate of false negatives.

The rest of the paper is organized as follows. Section 2 gives related work and theoretical background. Section 3 discusses the proposed method, including the proposed deep learning. Section 4 presents the experimental set up, results and discussion and lastly, Section 5 concludes the work.

2. Related Work and Theoretical Background

In the era of 1970s to the 1990s, researchers analyzed medical images with sequential application of low-level pixel processing such as edge and line detector filters, and region growing as well as mathematical modeling such as fitting lines, ellipses and circles to build compound rule-based systems that resolved particular tasks. The moment it was possible to digitize the images, researchers starting to develop automatic analysis systems.

Manuscript received August 5, 2023
Manuscript revised August 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.8.15

Supervised techniques became more and more popular in medical image analysis in the late 1990s. The techniques use training data to develop a system. Pattern recognition and machine learning approach are dominant in use by many successful commercial medical image analysis systems. Thus, the systems that completely designed by humans shifted to systems that are trained by computers using data where feature vectors are extracted from. The features extraction from the images is a crucial step in the design of such systems. Then computer algorithms select the best decision threshold in the high-dimensional feature space.

The basic concept of many deep learning algorithms is as follows. Models (networks) consist of many layers that transform input data (e.g. images) to outputs (e.g. the present/absent of disease) while learning increasingly higher-level features. Then computers learn the features that optimally represent the data for the problem at hand. The most successful type of models for image analysis to date are convolutional neural networks (CNNs). CNNs contain many layers that transform their input with convolution filters of a small extent. Karbab et al. [8] introduces the use of deep learning CNN on the mobile malware detection. Instead of using images, the authors use malware signature as an input and the Convolution Layer acts as signatures/features extraction and then the other layers make a decision maliciousness and family of the malware.

Litjen et al. [9] review the major deep learning concepts pertinent to medical image analysis. According to [9] many researchers have carried out works on CNNs since the late seventies, such as works by Fukushima [10] and Lo et al. [11] have already applied the CNN to medical image analysis. The researchers saw their first successful real-world application in LeNet [12] for hand-written digit recognition. Even with these early accomplishments, the use of CNNs did not meet momentum until end of 2012. The turning point was the contribution of Krizhevsky et al. [13] that won with large margin the ImageNet challenge in December 2012. The authors proposed AlexNet, a CNN consists of feature maps of (96; 256; 384; 384; 256) kernels with pooling on the first, second, and fifth layers and kernel sizes are (11; 5; 3; 3; 3) respectively. At the end of the network, two fully connected layers containing 4096 units are added to the end of the network that gave rise to 60 million parameters. In following years, researchers make great progress in deep learning CNN development utilizing deeper architectures [14]. Deep convolutional networks reach the popularity to become the technique of choice.

2.1. The Deep Learning CNN

Convolutional neural networks (CNN) is a type of deep artificial neural networks (ANN) [15]. It is a feed-forward artificial neural network that can be considered as a composition of number of functions (1) [16].

$$g(x)=g_L(\ldots g_2(g_1(x;w_1);w_2)\ldots),w_L) \quad (1)$$

Each function g_l takes as input a data x_l and a parameter vector w_l and produces as output x_{l+1}. While the type and sequence of functions is usually handcrafted, the parameters $w=(w_1,\ldots,w_l)$ are learned from data in order to solve a classification or other target problem.

Data and functions in CNN have extra structure. The data $x_1, x_2, \ldots, x_n$ in general will be in 2D arrays. Every x_i will be an $MxNxC$ real array of MxN entries and K channels per entry. Thus, the first two dimensions of the array span space, whereas the last dimension spans channels. All data x_l are intermediate feature maps except $x=x_1$ as an actual input data to the network.

The functions g_l have a convolutional structure as well. g_l use an operator that is local and translation invariant to the input map x_l.

The first CNN is the regular linear convolution by a filter bank. A sample of single function relation as shown in (2).

$$g: RM\times N\times C\rightarrow RM'\times N'\times K',\ x\mapsto y \quad (2)$$

2.2. The General Architecture of CNN

This subsection provides the general architecture of CNN summarized from [17]. The general architecture of CNN consists of three layers: Convolution Layer, MaxPooling Layer and Fully Connected (FC) Layer.

Convolution Layer

This layer governs the output of neurons connected to local parts of the input by calculating the scalar product between their weights and the region connected to the input volume. An activation function called rectified linear unit (ReLU) is applied to the output of the activation produced by the previous layer. The ReLU function is defined as the positive part of its argument:

$$f(x) = x^+ = max(0,x) \quad (3)$$

MaxPooling Layer

This layer performs down sampling along the spatial dimensionality of the given input, which more reducing the number of parameters within that activation.

Fully Connected Layer

This layer will perform the same functions as in standard ANNs and try to yield class scores from the activations, for carrying out classification task. As to improve performance,

the activation function ReLu may be used between these layers.

Having done executing this modest method of transformation, CNNs are able to transform the original input layer by layer using convolutional and down sampling techniques to produce class scores for classification and regression purposes.

3. Methodology

3.1. The Proposed CNN

This section presents the CNN framework and its components used in the experiments. Figure 1 illustrates the overview of the approach of this work that adapted from the work in [8]. The CNN has a simple design and uses minimum as possible preprocessing to obtain the breast cancer information. The representation learning (feature extraction) and detection/attribution are based on the actual neural network. The CNN uses supervised machine learning, thus it is necessary to train the CNN. Then, the CNN is tested using different part of the dataset. The experiments use the same preprocessing procedure for both, the training and the testing phases to ensure the correctness of the detection results. The detection task needs only one neuron in the output layer because the network decides whether the image is benign or malignant tumor.

This work decides to use the proposed model due to its efficiency and ability to run the proposed model on resource-constrained devices.

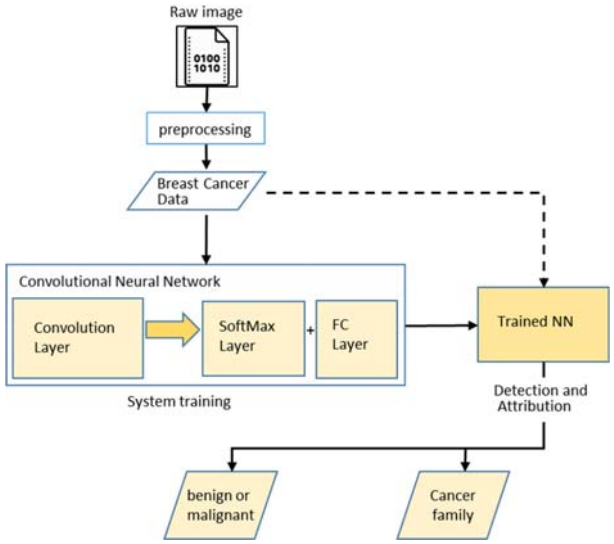

Fig. 1 Proposed beam former

As presented in Figure 2, the first layer is a convolution layer [18] with rectified linear unit (ReLU) activation function as presented in (3). Afterward, we use global max pool [18] and connect it to a fully connected layer. Notice that in addition to Dropout [19] used to prevent overfitting, this work also utilizes Bench normalization [19] to improve the results.

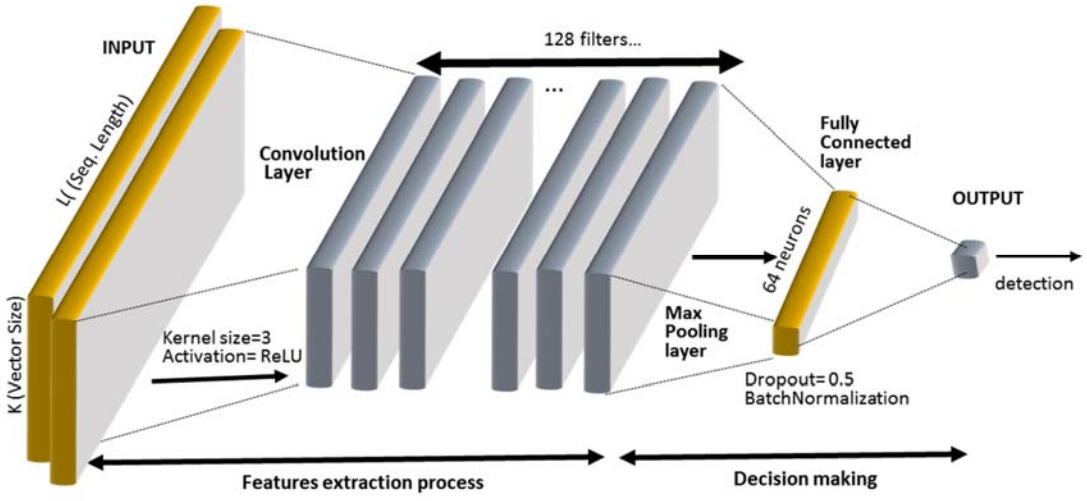

Fig. 1 The architecture of the proposed CNN.

Table 1. Attributes of the CNN.

Layer	Options	Activate function
Convolution	Filter:128, filter size=3	ReLU
MaxPooling	-	-
Fully Connected (FC)	#Neurons=64, Dropout=0.5	ReLU

3.2. The Wisconsin Diagnostic Breast Cancer dataset

The dataset used in this paper is publically available [20] and was created by Dr. William H. Wolberg, physician at the University of Wisconsin Hospital at Madison, Wisconsin, USA. To create the dataset, Wolberg [21] used uid samples, taken from patients with solid breast masses [21] (see Figure 3) and an easy-to-use graphical computer program called Xcyt [22], which is capable of perform the analysis of cytological features based on a digital scan. The program uses a curve-fitting algorithm, to compute ten features from each one of the cells in the sample.

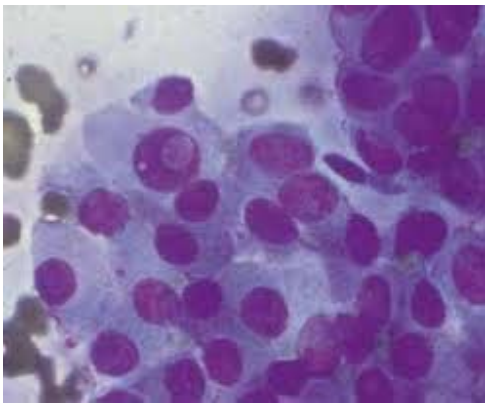

Fig. 3. A magnified image of a malignant breast fine needle aspirate (FNA).

The Wisconsin Diagnostic Breast Cancer (WDBC) dataset [20] consists of 569 breast masses with 357 benign and 212 malignant cases. In order to evaluate the size, shape, and texture of each cell nuclei, ten characteristics were derived and described as follows [22,23].

- Radius: length average of the radial line segments from the center of mass of the boundary to each of the boundary points.
- Perimeter: sum of the distances between consecutive boundary points.
- Area: count of the number of pixels on the interior of the boundary and adding one-half of the pixels on the perimeter, to correct for the error caused by digitization.
- Compactness: combination of the perimeter and area to give a measure of the compactness of the cell.
- Smoothness: quantity of the difference between the length of each radial line and the mean length of the two radial lines surrounding it.
- Concavity: the size of any indentations in the boundary of the cell nucleus.
- Concave points: similar to concavity, but counts only the number of boundary points lying on the concave regions of the boundary, not the magnitude of the concavities.
- Symmetry: the relative difference in length between pairs of line segments perpendicular to the major axis of the contour of the cell nucleus.
- Fractal dimension: The perimeter of the nucleus.
- Texture: the variance of the grayscale intensities in the component pixels.

Table 2 shows the ten characteristics of the cells and their types of values. These characteristics will used as feature parameters of the cells. For each image of the cells the mean, standard error, and the extreme (largest or "worst") value of each characteristic were calculated, which resulted in 30 features of 569 images, yielding a database of 569×30 entries.

Table 2. Feature parameters.

Features	Domain (value)
Radius	Numeric, 1-10
Perimeter	Numeric, 1-10
Area	Numeric, 1-10
Compactness	Numeric, 1-10
Smoothness	Numeric, 1-10
Concavity	Numeric, 1-10
Concave points	Numeric, 1-10
Symmetry	Numeric, 1-10
Fractal dimension	Numeric, 1-10
Texture	Numeric, 1-10
Class distribution	*Malignant: 212*
	Benign: 357
Number of instances	*569*

4. Experimental Set up, Results and Discussion

4.1. Hardware & Software Specifications

For experiments purpose, the proposed method is implemented on high-end server machine with the

following specifications. VPS 16 cores processor, 512GB RAM, and 3TB SSD storage using Java, Python Version 3.5.1, Keras and Tensorflow utilities/libraries [24].

4.2. Experimental Results

260 samples were randomly selected for training (150 benign masses and 110 malignant tumors), and the remaining 309 samples for test. Figure 4 shows training result. The graph shows that the accuracy during the training achieved 96% after 160 epochs. Figure 5 shows the accuracy of the test, which is 98.1% after 160 epochs. The results are considered good enough, due the fact that the losses during the training and during the testing are unavoidable.

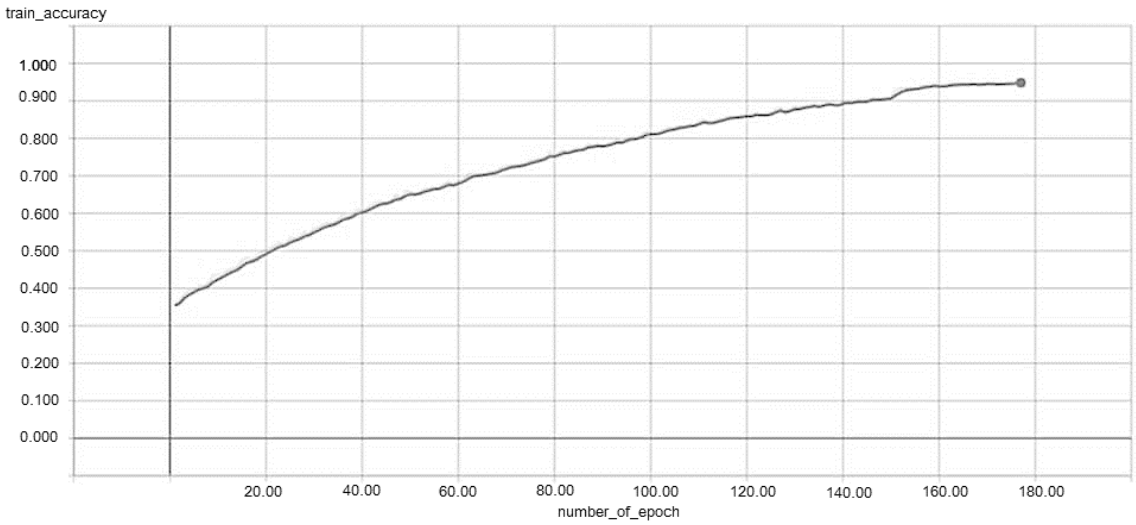

Fig. 4. Accuracy results during the training.

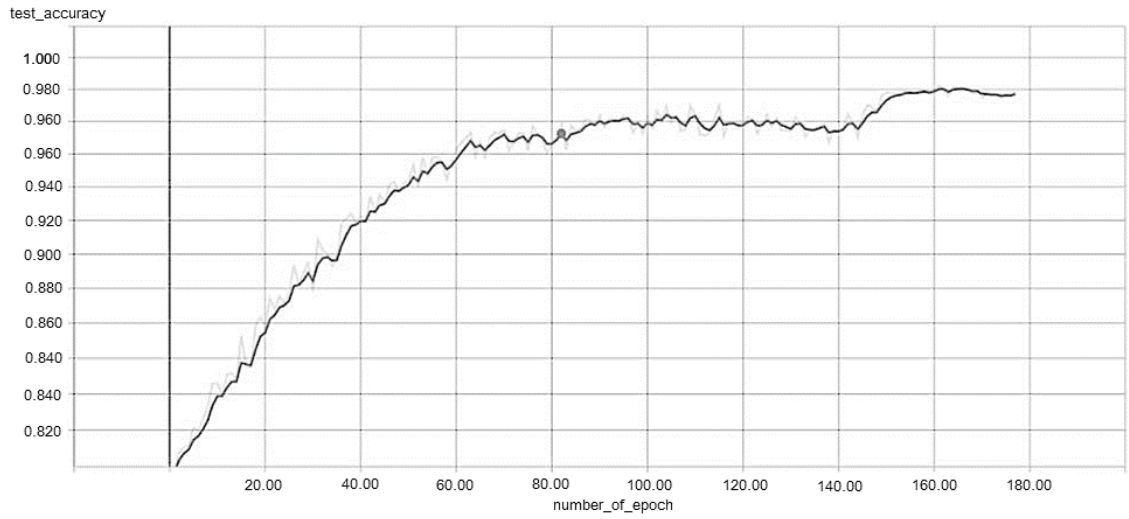

Fig. 5. Accuracy results during the testing.

3.3. Detection Performance

This section gives an evaluation of how effectively the proposed system can distinguish between malicious and benign images in terms of false positive rate. Table 3 depicts the metrics to measure the performance.

Table 3. Performance Metrics.

Metrics	Measurement
True positives (TP)	The number of successfully detected malignant
False negatives (FN)	The number of incorrectly classified malignant.
False positives (FP)	The number of incorrectly classified benign.
True negative (TN)	The number of successfully classified benign.
False positive rate (FPR)	FPR= FP/(FP+TN)
Accuracy (A)	A=(TP+TN)/(TP+TN+FP+FN)

The trained CNN performs the malignant detection on the WDBC dataset. The experiments are repeated 10 times. Table 4 shows the results.

Table 4. Experiment results on malignant tumors detection.

Exp. #	TP	FP	FN	TN	Accuracy
1	210	2	2	355	0.9894
2	212	0	2	355	0.9964
3	212	0	0	357	1.0000
4	211	1	2	355	0.9947
5	212	0	0	354	1.0000
6	211	1	0	357	0.9982
7	210	2	1	356	0.9947
8	212	0	1	356	0.9982
9	212	0	1	356	0.9964
10	212	0	1	356	0.9982
Total					9.9662

From Table 4, the average accuracy is calculated as follows. Accuracy = 9.9662/10*100%.= 99.66%. The author compares the accuracy of the work with other works that use same dataset [7]. Table 5 depicts the comparison results.

Table 5. Comparison of the Accuracy (based on [7]).

Algorithm	Accuracy	Reference & year
Back propagation	94.90%	[25]/1992
MSM-T	97.00%	[21]/1993
MSM-T+ Preprocessing	97.50%	[22]/1995
Fuzzy-Genetic	98.80%	[26]/1999
GRNN	97.80%	[27]/2004
Fuzzy+ KNN	99.14%	[28]/2006
Hybrid SVM	99.51%	[29]/2008
BP-MLP	99.28%	[30]/2011
This work	99.66%	

3.4. Discussion

The promising results in this work are because of the deep learning CNN that avoids the overfitting that contributes to the improvement of the results. As can be seen in Figure 4 and Figure 5, the accuracy during the training is higher than the accuracy during the testing. It indicates that the model is not under-fitting. The convolutional neural network used in this work contains 128 non-linear hidden layers (as illustrated in Figure 3) and this makes them very expressive models that can learn very complicated relationships between their inputs and outputs. Since the CNN in this work is a large network, it requires large amounts of training data and there is no enough data available to train different networks on different subsets of the data. With limited training data, however, many of these complicated relationships will be the result of sampling noise, so they will exist in the training set but not in real test data. This situation leads to overfitting and many methods have been developed for reducing it. This work uses Dropout technique [31].

Dropout is a technique that prevents overfitting and provides a way of approximately combining exponentially many different neural network architectures efficiently. The term dropout refers to dropping out units (hidden and visible) in a neural network. Dropping a unit out means temporarily removing it from the network, along with all its incoming and outgoing connections. This work drops randomly a neuron in the Hidden Layer with probability of 0.5.

5. Conclusion and Future Work

This work has shown the adaptation of CNN for assessing breast cancer dataset from Wisconsin University. The detection performance is relatively better than existing available methods. The CNN provides more accurate detection result, due to its convolution layer nature that filters the features more details.

As for future work, the author will consider feature ensemble learning based on Sparse Autoencoders and Softmax Regression for classification of Breast Cancer.

Acknowledgment

The author thanks to Scientific Research Deanship, Albaha University for the research funding under grant no.: 026//1439.

References

[1] Siegel R.L, Miller K.D, Jemal A, Cancer statistics, 2017,
DOI: https://doi.org/10.3322/caac.21387.

[2] Elmore J.G, Nakano C.Y, Koepsell T.D, Desnick L.M, Ransoho D.F, International variation in screening mammography interpretations in community-based programs. J Natl Cancer Inst, 2003;95(18):1384-1393.

[3] Veronesi U, Boyle P, Goldhirsch A, Orecchia R, Viale G, Breast cancer. Lancet, 2005; 365:17271741.

[4] Raimond W. M, Giard MD, Jo Hermans: The value of aspiration cytologic examination of the breast a statistical review of the medical literature. American Cancer Society, 1992;69(8):2104-2110.

[5] Elmore, J.G; Armstrong K, Lehman C.D, Fletcher S.W, Screening for breast cancer. The Journal of the American Medical Association, 2005;293(10):1245-56.

[6] Gayathri. B.M, Sumathi C.P, Santhanam T, Breast cancer diagnosis using machine learning algorithms –a survey, International Journal of Distributed and Parallel Systems (IJDPS), 2013; 4(3).
DOI : 10.5121/ijdps.2013.4309 105.

[7] Borges L.R. Analysis of the Wisconsin Breast Cancer Dataset and machine learning for breast cancer detection, XI Workshop de Visão Computacional, WVC´2015, São Carlos – SP – Brazil, 5-7 October, 2015:15-19.

[8] Karbab E.B, Debbabi M, Derhab A, Mouheb D. MalDozer: Automatic framework for android malware detection using deep learning, Proceedings of the Fifth Annual DFRWS Europe, Digital Investigation, 2018;24:S48-S59.
DOI: https://doi.org/10.1016/j.diin.2018.01.

[9] Litjens G, Kooi T, Bejnordi B.E, Setio A.A.A, Ciompi F, Ghafoorian M, van der Laak J.A.W.M, van Ginneken B, S´anchez C.I, A Survey on deep learning in medical image analysis, Medical Image Analysis Journal, 2017;42:60–88,
DOI: https://doi.org/10.1016/j.media.2017.07.005

[10] Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 1980;36 (4), 193–202.

[11] Lo S.-C, Lou S.-L, Lin J.-S, Freedman, M.T, Chien, M.V, Mun, S. K. Artificial convolution neural network techniques and applications for lung nodule detection. IEEE Transactions on Medical Imaging, 1995;14:711–718.

[12] Lecun Y, Bottou L, Bengio Y, Ha ner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86, 1998;2278–2324.

[13] Krizhevsky A, Sutskever I., Hinton G. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. 2012;1097–1105.

[14] Russakovsky, O, Deng, J, Su, H, Krause, J., Satheesh, S., Ma, S, Huang, Z., Karpathy, A, Khosla, A, Bernstein, M., Berg, A C, Fei-Fei, L. ImageNet large scale visual recognition challenge. International Journal of Computer Vision, 2014;115 (3), 1–42.

[15] Goodfellow I, Bengio Y and Courville A. Deep Learning, MIT Press, 2016.

[16] Lecun Y, Bengio, Y Hinton G. Deep learning. Nature, 2015;521 (7553), 436–444.

[17] O'Shea K & Nash R. An Introduction to Convolutional Neural Networks. ArXiv e-prints, 2015.

[18] Kim Y. Convolutional neural networks for sentence classification, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25-29 October, 2014:1746–1751.

[19] Pennington J, Socher R. et al., GloVe: Global Vectors for Word Representation, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25-29 October, 2014: 1532–1543.

[20] University of Wisconsin-Madison. Machine Learning for Cancer Diagnosis and Prognosis. http://pages.cs.wisc.edu/ olvi/uwmp/cancer.html.

[21] Wolberg W.H, Street W.N, Heisey D.M, and Mangasarian O.L. Computer-derived nuclear features distinguish malignant from benign breast cytology, Human Pathology, 1995;26:792--796.

[22] Wolberg W.H, Street W.N, and Mangasarian O.L. Image analysis and machine learning applied to breast cancer diagnosis and prognosis. Analytical

and Quantitative Cytology and Histology, 1995;17(2);77-87, April 1995.

[23] Mu T. Breast cancer diagnosis from fine-needle aspiration using supervised compact hyperspheres and establishment of confidence of malignancy Tingting Mu, Asoke K. Nandi A.K. 16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008.

[24] Tensorflow - https://www.tensorflow.org (2017).

[25] Bennett K. P. and Mangasarian, O.L. Neural Network Training via Linear Programming. Advances in Optimization and Parallel Computing, Pardalos P.M.(Ed.),Elsevier Science Publishers B. V., 1992:56-67.

[26] Pena-Reyes C.A. and Sipper M. A Fuzzy-genetic approach to breast cancer diagnosis. Artificial Intelligence in Medicine, Elsevier, 1998;17(2):131-55.

[27] Kiyan T. and Yildirim Y. Breast cancer diagnosis using statistical neural networks. Journal of Electrical & Eletronics Engineering, Istanbul University, 2004;4(2):1149-1153.

[28] Sahan S, Polat K, Kodaz H. Gunes S. A new hybrid method based on fuzzy-artificial immune system and k-NN algorithm for breast cancer diagnosis, Computers in Biology and Medicine, 2001;37:415-423.

[29] Akay M.F. Support vector machines combined with feature selection for breast cancer diagnosis. Expert Systems with Applications, 2008;36:3240-3247.

[30] Paulin F, Santhakumaran A. Classification of Breast cancer by comparing Back propagation training algorithms, International Journal on Computer Science and Engineering (IJCSE), 2011;3: 327-332.

Mohammed Y. Alzahrani received the Master and PhD in Computer Science from Heriot-Watt University, UK in 2010 and 2015, respectively. He is currently the dean of College of Computer Science and Information Technology, Albaha University, KSA. His research interests include model checking and verification, intelligent healthcare systems, and information security.