

Robust Breast Cancer Classification Using Wave Atom and Back Propagation Neural Networks

Mohammed Lubbad¹, Mohammed Alhanjouri¹ and Huda Alhalabi^{2*}

¹Department of Computer Engineering, Islamic University of Gaza (IUG), Gaza, 190403 Palestine

²National Advanced IPv6 Centre (NAv6), Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia

ABSTRACT

The breast cancer automatic diagnosis is a critical real world medical challenge. This study proposes a classifying cancer tumor method based on their gene expression signatures to specific diagnostic categories. The developed neural network model holds promise for patients, surgeons, and radiologists, providing them with information, which was only available using biopsy. This significantly reduces the number of pointless surgical procedures. This study utilizes Wave Atom Transform as feature extraction method, and Back Propagation Algorithm to classify cancer into pre-defined classes. The proposed model provides automatic detection with a high level of accuracy (90%).

Keywords: Back Propagation Network (BPN), breast cancer diagnosis, neural networks, wave atom transformation

INTRODUCTION

Cancer diagnosis and treatment research has become a vital issue for the scientific community. Cancer has become one of the main causes of death around the world. It is a complex clinical heterogeneous disease. Worldwide, breast cancer represents 10.4% of all cancer occurrences among women, making it the most common form of non-skin cancer in women, and the fifth most common cause of cancer death. Treatment includes chemotherapy,

hormonal therapy and radiation. The use of machine learning tools in medical diagnosis is increasing gradually. Classification and recognition systems have verified their effectiveness to help medical experts in diagnosing diseases. It is unnecessary to provide one specific algorithm to identify the disease, therefore using neural networks are perfect in recognizing diseases. The

ARTICLE INFO

Article history:

Received: 06 January 2018

Accepted: 27 March 2019

Published: 24 July 2019

E-mail addresses:

engmlubbad@gmail.com (Mohammed Lubbad)

mhanjouri@iugaza.edu.ps (Mohammed Alhanjouri)

Huda.elhalabi@gmail.com (Huda Alhalabi)

* Corresponding author

details of how to recognize the disease are not required, since neural networks are learned by examples. What is required is a set of examples representing all the disease variations. Many efforts have been taken in past to detect breast cancer using various algorithms and techniques (Bankman et al., 1992; Mazurowski et al., 2007).

Artificial Neural Network

An Artificial Neural Network (ANN) is an information processing paradigm, which is encouraged by how biological nervous systems process information. The novel structure of the information processing system is the key element of this paradigm. It consists of many highly interconnected processing elements (neurons) working in harmony to solve definite problems. ANNs, like people, learn by example. An ANN is constructed for a specific application, such as data classification or pattern recognition, through a learning process. Learning in biological systems includes changes to the synaptic connections between the neurons (Zurada, 1995). Many efforts have been taken in past using neural network in context of detecting cancer (Cheng et al., 1995; Fooladi et al., 2008; Tsapatsoulis et al., 1997).

Related Works

Numerous methods have been utilized to recognize and predict meaningful pattern for breast cancer diagnosis. Şahan et al. (2007) diagnosed breast cancer using hybrid machine learning method. Their approach integrated k-nearest neighbor algorithm with a fuzzy-artificial immune system. This method offered acceptable accuracy in Wisconsin Breast Cancer Dataset (WBCD). It could also be deployed for other breast cancer diagnosis problems. Ryu et al. (2007) proposed isotonic separation as a data classification method. The performances were compared against learning vector quantization, support vector machines, decision tree induction, and other techniques based on sufficient and insufficient breast cancer dataset. The isotonic separation was experimentally tested, and the results showed that it could be a practical tool for medical classification. Ubeyli (2007) presented a comprehensive view of automated diagnostic systems implementation for breast cancer detection. It compared the performances of combined neural network (CNN), multilayer perceptron neural network (MLPNN), support vector machine (SVM), probabilistic neural network (PNN) and recurrent neural network (RNN). Several hybrid systems and combinations deployed neural networks as a component. Nevertheless, because almost all the employed neural networks are conventional gradient descent BP ANN, the hybrid method still suffers from the neural networks drawbacks (Utomo et al., 2014).

Wave Atom

Single channel denoising relies on applying single band low filter to speech, to filter noise and remove high frequencies. That process may cause damage to a significant data. To avoid this problem, many researches proposed using multi-band technique like wave atom, which is considered as a new part of wavelet family. When traditional wavelet transform goes through one phase to another, only the approximation will be decomposed. The wavelets packets' decomposition can be followed into the other not optimal sets. Thus, the optimality is linked to the decomposition maximum energy. Consequently, it is very important to search for the path resilient to the maximum energy within the different sub-bands. For numerical and image analysis, wave atom used as a multiscale transform. Some essential concepts can be described following (Demagnet & Ying, 2007).

The 2D Fourier transform can be defined as:

$$\hat{f}(w) = \int e^{-ixw} f(x) dx \quad (\text{Antoine \& Murenzi, 1996})$$

$$f(x) = \left(\frac{1}{2\pi}\right)^2 \int e^{ixw} \hat{f}(w) dw \quad (\text{Antoine \& Murenzi, 1996})$$

The indexes are determined by integer values associated to a point in the phase-space as follows: $x\mu = 2^{-jn}$, $w\mu = \pi 2^j m$, $C_1 2^j \leq \max_{i=1,2} |m_i| \leq C_2 2^j$. Li and Hagness (2001) proposed that two parameters were acceptable to index α . The index ($\alpha = 1$) if the decomposition was multi-scale, or ($\beta = 0$) if the decomposition was not multi-scale; and β specified if basic elements were poorly direction and localized ($\alpha = 1$) or, fully directed and extended ($\beta = 0$). In terms of α and β , the description might explain the connections between different recent analysis transforms. Directional and complex multi resolution analysis of wavelets correspond to $\alpha = \beta = 1$, for $\alpha = 1, \beta = 0$, curvelets $\alpha = \beta = 1/2$ and Gabor transform are defined for $\alpha = \beta = 0$. Wave atoms correspond to $\alpha = \beta = 1/2$. Wavelet classifications are shown in Figure 1. To present the wave atom, one should first study the 1D case. Practically, wave atoms are constructed from effectively chosen 1D wavelet packets tensor products. One dimensional family of real-valued wave packets is $\psi_{m,n}^j(x), j \geq 0, m \geq 0, n \in \mathbb{Z}$ centered in space around $x_{j,n} = 2^{-j}n$, centered in frequency around $\pm w_{j,m} = \pm \pi 2^j m$ with $C_1 2^j \leq m \leq C_2 2^j$. The one-dimensional form of the parabolic scaling informs that the support of $\psi_{m,n}^j(w)$ be of length $O(2^{2j})$, while $w_{j,m}(w) = O(2^{2j})$ (Peto & Peto, 1972). The desired consistent frequency tiling is shown in the bottom of Figure 2. Filter bank-based wavelet packets is considered in this study, as a proper definition of an orthonormal basis sustaining these localization properties. As illustrated in Figure 2, the wavelet packet tree that describes the dividing of the frequency axis in 1D, can be chosen to have depth j if the frequency is 2^{2j} . Figure 2 shows wave atoms represented by a wavelet packet tree. Please refer to Peto and Peto, (1972) for additional details. Villemoes wavelet packets on the positive frequency axis are shown in Figure 2. A frequency where a change of scale

occurs presented by the dot under the axis. The labels “LH”, “RH” indicate a left-handed, and right handed window, respectively (Mallat, 1999). In 20 domains, the existing structure can be adapted to suit some applications in numerical analysis or image processing: The orthobasis variant. Practically, instead of a tight frame, one may work with the original orthonormal basis $\varphi_{\mu}^+(x)$. Since $\varphi_{\mu}^+(x) = \varphi_{\mu}^1(x) + \varphi_{\mu}^2(x)$ each function $\varphi_{\mu}^+(x)$ move in two different directions, instead of one. This is called the orthobasis variant (Alhanjouri et al., 2013).

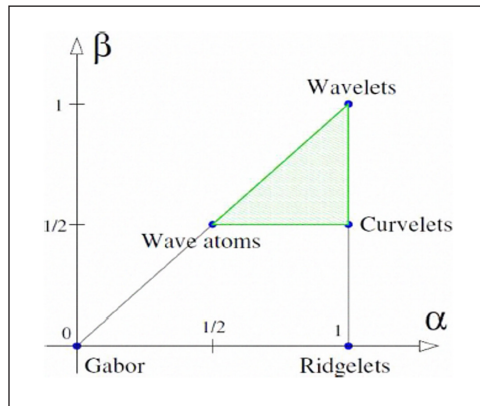


Figure 1. α and β Classification for Curvelet, Wavelets and Wave atoms

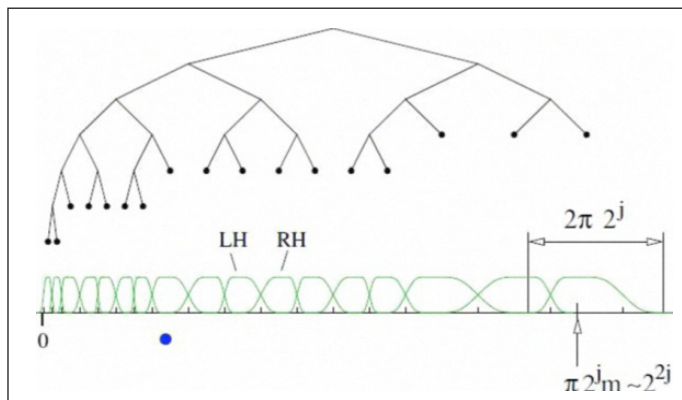


Figure 2. Wavelet packet tree corresponding to wave atoms

METHODS

Recently, the back-propagation architecture (BPN) is the most widespread, efficient, and easy to learn model for multi-layered and complex networks (Mishra et al., 2011).

THE BPN

The multi-layer perceptron or multi-layer feed-forward neural network, also called the BPN, is widely used and very common. The BPN depends on the supervised procedure,

i.e. the network figures data examples model with known outputs (Mumtaz et al., 2009). The BPN architecture is a layered feed forward neural network, in which the neurons - nonlinear elements- are organized in successive layers, and the information flows unidirectional, from input layer to output layer, within the hidden layer(s). Figure 3 shows a three layered feed forward neural network comprising of one hidden layer, one input layer, and one output layer.

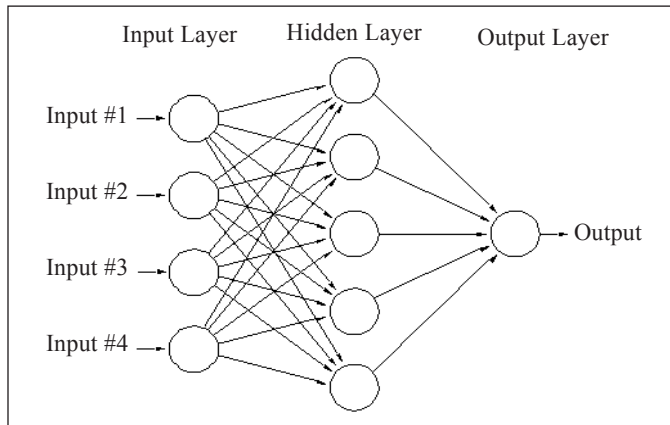


Figure 3. Feed Forward Neural Network

Learning in BPN

For a neural network to complete a task, the weights of each unit should be adjusted in such a way that the error between the actual output and the wanted output is eliminated (Rajasekaran & Pai, 2003). The process continues until the error becomes zero or it is reasonably low. To perform this process, the back-propagation algorithm is the most effective technique.

BPN Algorithm

Step 1. Normalize the outputs and inputs according to their maximum values. Assume there are 'l' inputs and 'n' outputs given by $\{ I \}_l$ and $\{ O \}_n$ respectively, for each training pair.

Step 2. Assume the number of neurons to lie between $L < m < 2 * L$, in the hidden layer.

Step 3. [V] denotes the weights of synapses connecting hidden and input neurons, and [W] denotes the synapses weights linking output and hidden neurons. Set small random values for the weights usually from -1 to 1.

α is defined as momentum coefficient. The value of α should be positive but less than 1 and threshold values can be taken as zero. η is called learning rate of the network.

$$[\Delta V]^0 = [\Delta W]^0 = [0]$$

Step 4. In the training data, set one set of inputs and outputs. Display the pattern to the input layer $\{I\}_I$ as inputs to the input layer. Evaluate the output of the input layer, by using linear activation function as

$$\{O\}_I = \{I\}_I$$

Step 5. Calculate the inputs to the hidden layer by multiplying corresponding synapses weights as

$$\{I\}_H = [V]^T \{O\}_I$$

Step 6. Evaluate the output using the sigmoidal function using the hidden layer units as

$$\{O\}_H = \left\{ \begin{array}{c} \cdot \\ \frac{1}{(1 + e^{-I_{Hi}})} \\ \cdot \end{array} \right\}$$

Step 7. Compute the inputs to the output layer by multiplying corresponding synapses weights as

$$\{I\}_O = [W]^T \{O\}_H$$

Step 8. Evaluate the network output using sigmoidal function as

$$\{O\}_O = \left\{ \begin{array}{c} \cdot \\ \frac{1}{(1 + e^{-I_{Oj}})} \\ \cdot \end{array} \right\}$$

Step 9. Calculate the difference between the desired output and network output as for the i th training set as

$$E^p = \frac{\sqrt{\sum (T_j - O_{Oj})^2}}{n}$$

Step 10. Find $\{d\}$ as

$$\{d\} = \left\{ \begin{array}{c} \cdot \\ (T_k - O_{Ok}) O_{Ok} (1 - O_{Ok}) \\ \cdot \end{array} \right\}$$

Step 11. Find $[Y]$ matrix as

$$[Y] = \{O\}_H \langle d \rangle$$

Step 12. Find $[\Delta W]^{t+1} = \alpha [\Delta W]^t + \eta[Y]$

Step 13. Find $\{e\} = [W] \{d\}$

$$\{d^*\} = \begin{pmatrix} \vdots \\ e_i(O_{Hi})(1 - O_{Hi}) \\ \vdots \end{pmatrix}$$

Find $[X]$ as : $[X] = \{O\} \langle d^* \rangle$

Step 14. Find $[\Delta V]^{t+1} = \alpha [V]^t + \eta[X]$

Step 15. Find

$$[V]^{t+1} = [V]^t + [\Delta V]^{t+1}$$

$$[W]^{t+1} = [W]^t + [\Delta W]^{t+1}$$

Step 16. Find error rate as $\text{Error rate} = \frac{\sum E_p}{nset}$

Step 17. Steps 4-16 will be repeated until the error rate convergence becomes less than the tolerance value.

In this study, the BPN algorithm performance is evaluated using Fine-Needle Aspiration (FNA) images dataset from different sources.

Breast Cancer Dataset

- Use FNA to evaluate Breast Lumps (Bukhari et al., 2011).
- Fine Needle Aspiration Cytology – HOLOGIC (McKee, 2017)
- Surgical Pathology Atlas - Image Databases (Surgical Pathology Atlas).

Figure 4 shows four different samples of Fine Needle Aspiration (FNA) images from our original dataset.

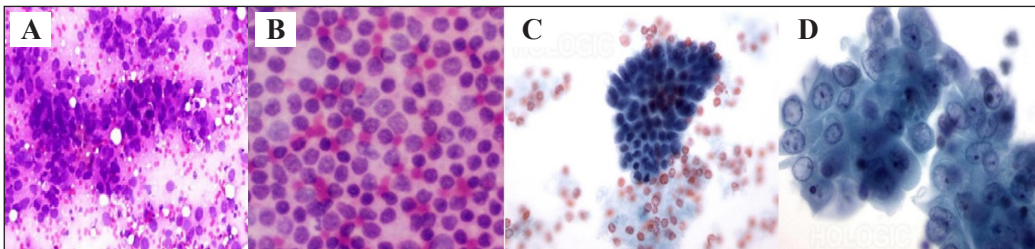


Figure 4. Fine Needle Aspiration (FNA) images from dataset of four different samples

Feature Extraction by Applying Wave Atom Transformation

Wave atom transform was applied to FNA images dataset shown in Figure 5.

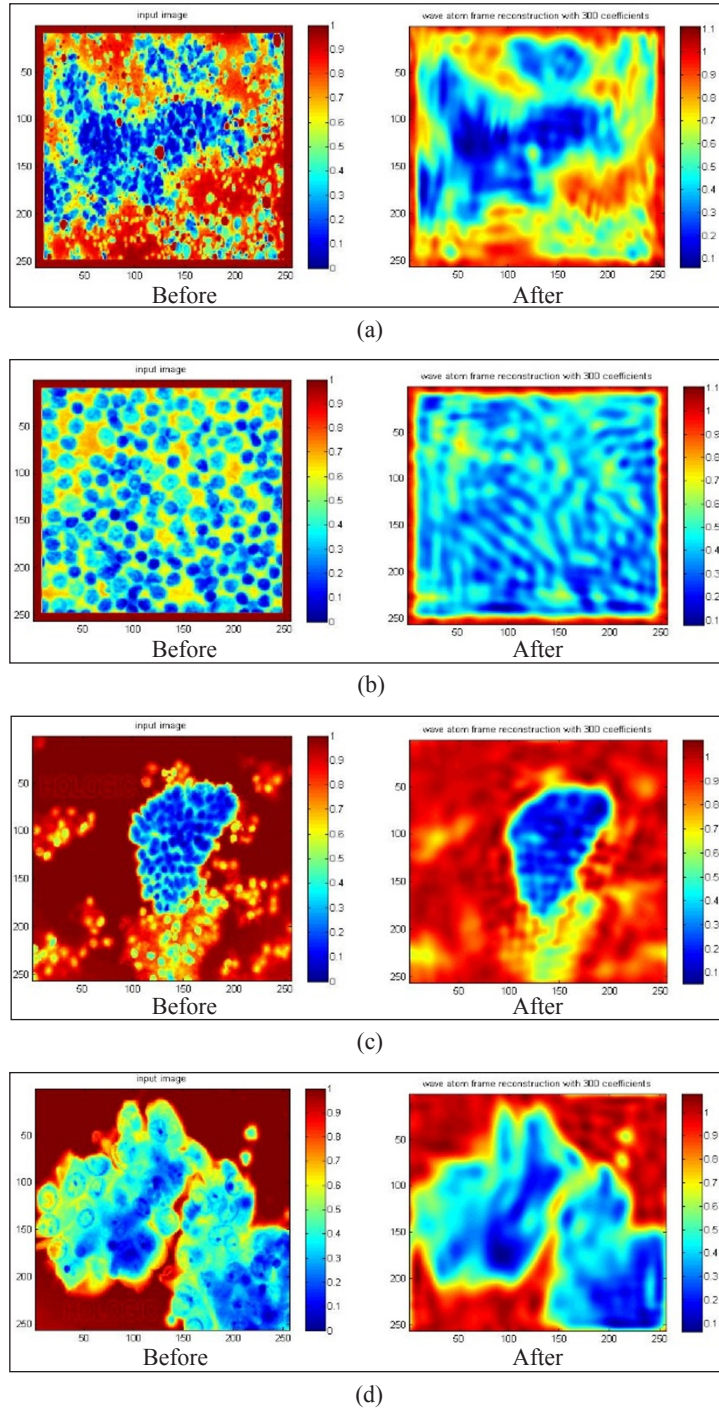


Figure 5. Wave Atom transformation in FNA images

Figure 5 shows four samples of FNA images before and after applying wave atom transformation, original in the left and transformed in the right of the figure.

Feature Selection by Gray Level Co-Occurrence Matrix (GLCM)

Using gray level co-occurrence matrix (GLCM), geometric and textural features were extracted from the Breast tissue and set as input to a BPN to classify breast tumor as cancerous or non-cancerous (Anand, 2010).

RESULTS AND DISCUSSION

Simulation and implementation had been done using MATLAB. Out of the total 60 samples, 20 samples were used for testing. Two methods were tested and compared, the BPN with Wave Atom and the BPN without Wave Atom. The estimated accuracy using the BPN algorithm with Wave atom reached 90%. While the accuracy for using BPN was only 75%. The results are summarized in the Table 1.

Table 1
Accuracy of the sample using Back Propagation Network

Type of neural network	Total Patterns taken for testing	Classification accuracy %	Training epochs
Back propagation Network with Wave Atom	20	90 % (approx.) (18 patterns correctly classified)	15000 (approx.)
Back propagation Network	20	75 % (approx.) (15 patterns correctly classified)	15000 (approx.)

CONCLUSION

The proposed neural network model with wave atom, provides automatic detection with a high amount of accuracy (90%). This percentage was earlier available only through biopsy, thus the proposed model significantly reduces the number of needless surgical procedures. Through this study, the classification accuracy rate is increased with 10% using Back Propagation Network & wave atom transformation for detection of Breast Cancer.

ACKNOWLEDGEMENT

Our sincere thanks to Prof. Abed El-Monem H. Lubbad, who offered all the infrastructure and facilities to carry out the research work.

REFERENCES

Alhanjouri, M., Lubbad, M. A., & Alkurdi, M. Z. (2013). Robust speaker identification using denoised wave atom and GMM. *International Journal of Computer Applications*, 67(5), 17-23.

- Anand, S. V. (2010, October 7-9). Segmentation coupled textural feature classification for lung tumor prediction. In *2010 International Conference on Communication Control and Computing Technologies* (pp. 518-524). Ramanathapuram, India.
- Antoine, J. P., & Murenzi, R. (1996). Two-dimensional directional wavelets and the scale-angle representation. *Signal Processing*, *52*(3), 259-281.
- Bankman, I. N., Christens-Barry, W. A., Weinberg, I. N., Kim, D. W., Semmel, R. D., & Brody, W. R. (1992, June 14-17). An algorithm for early breast cancer detection in mammograms. In *1992 Proceedings Fifth Annual IEEE Symposium on Computer-Based Medical Systems* (pp. 362-369). Durham, NC, USA.
- Bukhari, M. H., Arshad, M., Jamal, S., Niazi, S., Bashir, S., & Bakhshi, I. M. (2011). Use of fine-needle aspiration in the evaluation of breast lumps. *Pathology research international*, *2011*, 1-10.
- Cheng, H. D., Chen, C. H., & Freimanis, R. I. (1995, October 23-26). A neural network for breast cancer detection using fuzzy entropy approach. In *1995 Proceedings International Conference on Image Processing* (Vol. 3, pp. 141-144). Washington, DC, USA.
- Demagnet, L., & Ying, L. (2007). Wave atoms and sparsity of oscillatory patterns. *Applied and Computational Harmonic Analysis*, *23*(3), 368-387.
- Fooladi, S., Khaloozadeh, H., & Behjati, F. (2008, July 12-15). Detection of increased level of chromosome breakage in peripheral blood of Iranian women with sporadic breast cancer using neural networks. In *Machine Learning and Cybernetics, 2008 International Conference on* (Vol. 3, pp. 1227-1231). Kunming, China.
- Li, X., & Hagness, S. C. (2001). A confocal microwave imaging algorithm for breast cancer detection. *IEEE Microwave and Wireless Components Letters*, *11*(3), 130-132.
- Mishra, B. K., Singh, S. K., & Bhala, S. (2011, February 25-26). Breast cancer diagnosis using back-propagation algorithm. In *Proceedings of the International Conference & Workshop on Emerging Trends in Technology* (pp. 470-473). Mumbai, Maharashtra, India.
- Mumtaz, K., Sheriff, S. A., & Duraiswamy, K. (2009, June 4-6). Evaluation of three neural network models using Wisconsin breast cancer database. In *2009 International Conference on Control, Automation, Communication and Energy Conservation* (pp. 1-7). Perundurai, Tamilnadu, India.
- Mazurowski, M. A., Habas, P. A., Zurada, J. M., & Tourassi, G. D. (2007, September 25-28). Case-base reduction for a computer assisted breast cancer detection system using genetic algorithms. In *2007 IEEE Congress on Evolutionary Computation* (pp. 600-605). Singapore.
- Mallat, S. (1999). *A Wavelet Tour of Signal Processing*. San Diego: Academic press.
- McKee G. T. (2017). *Fine Needle Aspiration Cytology*. Retrieved November 2, 2017, from <http://www.cytologystuff.com/study/section30ng.htm>
- Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society: Series A (General)*, *135*(2), 185-198.
- Rajasekaran, S., & Pai, G. V. (2003). *Neural Networks, Fuzzy Logic and Genetic Algorithm: Synthesis and Applications (With CD)*. New Delhi: PHI Learning Pvt. Ltd.

- Ryu, Y. U., Chandrasekaran, R., & Jacob, V. S. (2007). Breast cancer prediction using the isotonic separation technique. *European Journal of Operational Research*, 181(2), 842-854.
- Surgical Pathology Atlas. (2016). *Image Databases*. Retrieved January 1, 2017, from <http://www.surgicalpathologyatlas.com/glfusion/mediagallery/>
- Sahan, S., Polat, K., Kodaz, H., & Güneş, S. (2007). A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis. *Computers in Biology and Medicine*, 37(3), 415-423.
- Tsapatsoulis, N., Schnorrenberg, F., Pattichis, C., Kollias, S., & Schizas, C. (1997, July 2-4). A neural network assisted algorithm for automated detection of breast cancer nuclei. In *Proceedings of 13th International Conference on Digital Signal Processing* (pp. 269-272). Santorini, Greece.
- Übeyli, E. D. (2007). Implementing automated diagnostic systems for breast cancer detection. *Expert Systems with Applications*, 33(4), 1054-1062.
- Utomo, C. P., Kardiana, A., & Yuliwulandari, R. (2014). Breast cancer diagnosis using artificial neural networks with extreme learning techniques. *International Journal of Advanced Research in Artificial Intelligence*, 3(7), 10-14.
- Zurada J. M. (1995). *Introduction to Artificial Neural Systems*. Boston: PWS Publishing.

