A Survey of Pattern Recognition Applications in Cancer Diagnosis

Amir Atapour Abarghouei, Afshin Ghanizadeh, Saman Sinaie, and Siti Mariyam Shamsuddin
Soft Computing Research Group
Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia Skudai, Malaysia
aaamir4@siswa.utm.my
gafshin2@siswa.utm.my
ssaman2@siswa.utm.my
mariyam@utm.my

Abstract – **In this paper, some of the image processing and pattern recognition methods that have been used on medical images for cancer diagnosis are reviewed. Previous studies on Artificial Neural Networks, Genetic Programming, and Wavelet Analysis are described with their working process and advantages. The definition of each method is provided in this study, and the acknowledgement is granted for previous related research activities.**

Keywords- Pattern Recognition, Artificial Neural Network, Genetic Programming, Wavelet Analysis, Cancer Diagnosis.

I. INTRODUCTION

Cancer is the second leading cause of death for both men and women in the world, and is expected to become the leading cause of death in the next few decades. Therefore, more efficient cancer detection methods can save many lives. In recent years, the classical methods of cancer detection, which include interpretation of CT, PET, MRI and combination of these by physicians or trained technicians, Blood tests have become obsolete, due to their low accuracy. When referring to the accuracy of a medical test, a perfect test gives only true positives and true negatives, and the worst possible test would be the same as guessing. Unfortunately, many of the classical tests fall somewhere in between these two extremes, and this can increase the chance of misdiagnosis.

Cancer detection has become a significant area of research in the image processing and pattern recognition community. In order to further improve the efficiency and veracity of diagnoses and treatment, image processing and pattern recognition techniques have been widely applied to the analysis and recognition of cancer, evaluation of the effectiveness of treatment, and the prediction of the development of the cancer. The aim of this paper is to present a review on some of the useful image processing and pattern recognition approaches that have been applied on different types of medical images in this utmost important field. This paper only provides a perspective on the subject of cancer diagnosis through pattern recognition, and can be considered an introduction to a deeper research on the techniques presented.

The main purpose of pattern recognition in cancer diagnosis is to solve the pattern classification dilemma where a set of input features are used to determine if a patient has a particular disorder. A myriad of approaches, such as artificial neural networks, genetic algorithms, fuzzy sets, rough sets, wavelet filters and statistical transforms, have been employed in order to solve this problem. Optical Fourier transforms [1] and wavelet transforms are effective on micro calcification where the high spatial frequency in the Fourier spectrum allows for their identification, but may not be as accurate on masses where a lower spatial frequency exists similar to the surrounding tissues. Fuzzy logic [2], Bayesian networks [3], case-based systems [4] and artificial neural networks [5,6,7,8] have been utilized with different degrees of success. Artificial neural networks have demonstrated their capabilities in this area and have been widely adopted due to their generalization capabilities [9].

Generally, the systems, using the methods mentioned above, that can help in the process of diagnosis, namely computer aided diagnostic (CAD) systems, are used in the classification task where certain features (clinical findings) are used to assign a case to a particular pattern which represents a diagnosis. Therefore, these can improve the performance of the radiologist in terms of reducing the number of misdiagnosis and reducing the time taken to reach a diagnosis, which note the two most important criteria in developing a CAD system. Other

978-0-7695-3879-2/09 $26.00 © 2009 IEEE
DOI 10.1109/SoCPaR.2009.93

performance measures, such as computational complexity and operational load can be overlooked, if kept in an acceptable level.

The rest of the paper is organized as follows: different materials and methods used in cancer detection are addressed in section 2. The paper is concluded in section 3.

II. MATERIAL AND METHODS

Different methods must be used in the detection of different types of cancer. It is because of the different types of medical images used in the process of diagnosis. In this section, some of the more commonly used pattern recognition methods in cancer detection are briefly described.

A. Artificial Neural Network (ANN)

The Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way the brain processes information. The key to this paradigm is the novel architecture of the information processing system, consisting of a large number of neurons working together to solve problems. The advantages of neural network methods are claimed to be as follows:

- Ease of optimization, which results in flexible non-linear modeling of large datasets.
- Distributed processing and representation of information.
- Massive parallelism.
- Learning based adaptation.

Artificial neural network is formed from an input, middle (hidden), and output layer. The layers are interconnected by weighed connection lines. All the information is weighted and can increase or decrease the activation of the node. The simplest neural network model is represented by the perception, which consists of a linear combiner followed by an activation function. The summing node of the perceptron computes a linear combination of the inputs (weights). The result is applied to its activation function. Accordingly, the perceptron produces a certain output, depending on the type of the activation function. An example of a two-layer perceptron architecture is displayed in Fig. 1.

In order to understand what neural networks are applied on, we need to comprehend the method of imaging used. Endoscopic UltraSound (EUS) represents a high-resolution method of imaging of the Gastro Intestinal (GI) tract and the nearby organs. It requires the usage of a special video-endoscope coupled with a miniature ultrasound transducer located at the end of the endoscope [10]. Endoscopic UltraSound Elastography (EUSE) is a newly developed imaging procedure that characterizes the differences in the hardness and strain between diseased tissues and normal tissues. EUSE has been used in several studies for the differentiation of benign and malignant lymph nodes, with the sensitivity and accuracy, higher than conventional EUS methods. Classically, the images were analyzed by doctors to determine if a tumor is malignant or benign. However, this method exhibits a major disadvantage due to the subjective means in which clinician raters analyze a large range of color nuances, on the basis of which an objective decision regarding the type of the tumor is to be derived [11,12]. An image which is the product of the above-mentioned imaging procedure is given as an example in Fig. 2.

In order to apply the neural network methodology to differentiate between the images taken from the sample movies, which contain either benign or malignant tumors; the movies need to be digitalized. Since an EUSE sample movie (dynamic image) consists of a sequence of 125 frames (static images) displaying 255 colors, then, from a mathematical point of view, there must be a 125×255 matrix (*aij*) for each patient, each row corresponding to a certain frame of the sample movie and each column corresponding to a color [10].

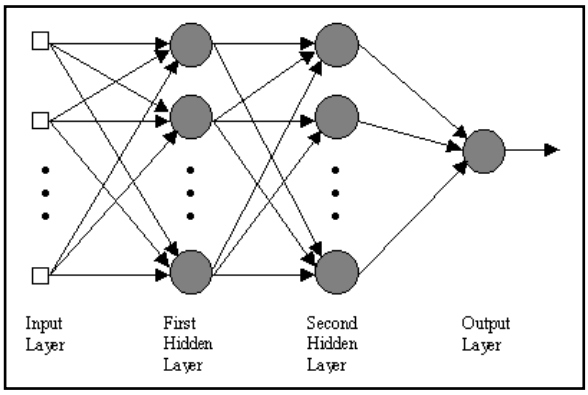

Figure 1. A two-layer perceptron architecture.

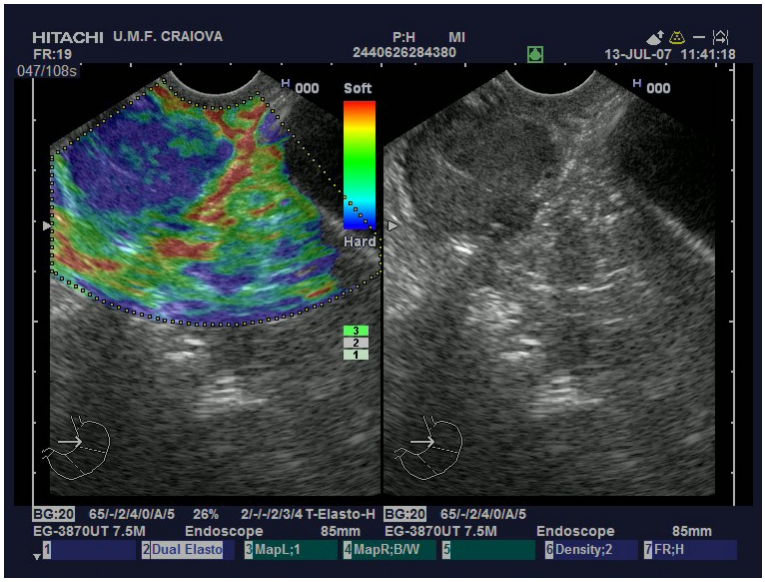

Figure 2. A frame from an EUSE sample movie, indicating a hard tumor (blue), and the soft surrounding tissues.

Since the natural input of a Neural Network (NN) is represented by vectors, a method of summarizing the matrix corresponding to an EUSE sample movie in a vector pattern is needed. Since a_{ij} represents the frequency of the color *j* in the *i*-th frame, then

$$a_j = \frac{1}{125}\sum_{i=1}^{125} a_{ij}$$

is the mean frequency of the color j in the sample movie. Consequently, the vector $(a_1, a_2,\ldots, a_{255})$ represents an average histogram summarizing the information provided by an EUSE sample movie.

During the classification process of the tumors types, a Multi-Layer Perceptron (MLP) or a Back propagation (BP) with two or more hidden layers is usually used. To evaluate the classification efficiency, two metrics can be computed [10]:

1. The training performance, which is the proportion of the cases which are correctly classified in the training process.
2. The testing performance, which is the proportion of the cases which are correctly classified in the testing process.

The testing performance provides the final check of the ANN classification efficiency. Hence, it is interpreted as a diagnosis accuracy using the ANN support.

Previous research in this area has been undertaken by various researchers. Wu et al. [13] used an ANN, which was trained using Back propagation Algorithm with ten hidden layers, to learn from 133 image instances. A single output node was trained to produce 1 for malignancy and 0 benign.

Another instance of BP was used by Floyd et al. [14] who used eight input parameters: mass size and margin, asymmetric density, architectural distortion, calcification number, morphology, density, and distribution. The questionable result of the experiments, done with BP, over 260 cases, was a classification accuracy of 50%.

Furundzic et al. [15] achieved the classification accuracy of 95% by presenting another BP with five hidden layers and 29 input features.

Gorunescu et al. [10] used only two layer MLP, and achieved a classification accuracy of 96% .

Although ANNs have a lot of applications in current cancer detection methods, but Hybrid systems, such as fuzzy-neural, fuzzy GAs, and neural GAs are mostly preferred, due to their much better conditions [16].

B. Genetic Programming

The broad adoption of Genetic Algorithms, where a population of individual parameters is encoded as streams of bits, was introduced by Holland [17], and has been used in many different areas. It has three major applications: intelligent search, optimization and machine learning. Genetic algorithm is a framework for solving any type of problem. The GA processes itself in a simple loop that always follows the same way of using the genetic operators and selection of objects for reproduction [16]. The key component of Genetic Algorithms is the emphasis on

crossover or recombination of individual solutions in a population.

Later, a new attempt to evolve computer programs began. This was an effort to create computer programs whose outputs were new computer programs, which solved problems. Rechenberg [18] created evolutionary strategies that started with a population of computer programs, and then by mutation and crossover created a new population. After that, by using a fitness measure, a test was performed to see whether the new program was better than its parent(s). Koza [19,20] proposed Genetic Programming (GP), which is the most widely used of such systems .

Some of the important features of GP include [21]:

- The ability to control the bias incorporated into feature selection.
- The ability to integrate data to produce mixed models of data.
- The ability to create human readable results.
- The ability to create unconstrained solutions.

One of the main uses of GP, which has many applications in cancer detection methods, including breast cancer, which is one of the main causes of death in women, is feature extraction. There are two phases in feature extraction [21]: first, the information needed for classification is extracted from data to an m-dimensional feature vector; second, an n-dimensional feature vector (n<m) is created from the parameter vector. The task of linear feature extraction measure transformation algorithm is to reduce the dimensionality of pattern observation space by finding a suitable linear subspace in which the class separability is optimally maintained [21].

To indicate the efficiency of GP, first four linear measures which are all independent feature extraction methods are briefly described [22].

1) Principal component analysis (PCA)

PCA is a well-established unsupervised method for feature extraction and dimensionality reduction in terms of a standardized linear transformation, which maximizes the variance of the transformed feature space. PCA does not require the information about datasets containing observations labeled by corresponding classes.

2) Fisher linear discriminant analysis (FLDA)

FLDA is the classical method for real-valued feature extraction using a linear transformation. FLDA is a supervised method and is designed optimally with its ability to maximize the ratio of between-class scatter and within-class scatter of projected features.

3) Alternative Fisher linear discriminant analysis (AFLDA)

AFLDA is a new measure to overcome the limitation of FLDA for binary classes, by replacing the original between-class scatter with a new scatter measure [22].

4) Modified Fisher linear discriminant analysis (MFLDA)

The limitation of within-class scatter of FLDA is that it cannot reveal the distribution of observations precisely. Although the computational demand of FLDA is low, it has one disadvantage when it is used as a measure of class separation. A large value of FLDA may be due to well-separated clusters. However, two overlapping classes with small values of variance may also result in large values of FLDA. Generally, the overlap is the main reason of the classification. Hence the higher ratio of between-class scatter and the within-class scatter may not help to improve the classification. In some cases, the classification performance may even drop. To overcome the limitation of FLDA and genuinely reflect the distribution of each pattern, a new within-class scatter that uses the distance between any two patterns belonging to the same class instead of the variance is developed [23].

GP, as a form of evolutionary algorithm and an extension of GA, has been proposed as the framework for the feature generation. GP derives the ability of feature selection from GA, but provides much larger feature space than GA to generate new features by nonlinearly combining terminators and operators. Some Fisher criterion measures are employed as the objective function of GP especially in order to evaluate the effectiveness of features and determine the features that can survive during the evolutionary process [22]. Various classifiers have previously been b\used in this field, some which that have been proven more efficient through the experiments of previous researchers are cited in this paper.

Guo et al. [22] has employed minimum distance classifier (MDC), Multi-layer Perceptron (MLP), and Support Vector Machine (SVM) as classifiers in order to evaluate the discriminating ability of features generated by the previously-described feature extraction methods, and GP.

MDC finds centers of classes, and measures distances between these centers and the test data. The distance is defined as a measure of similarity in the

way that the minimum distance stands for the maximum similarity.

MLP, which usually consists of one hidden layer with between 2 and 14 neurons, and one output layer when used as a classifier, is one of the most common classifiers because of its ability to learn and identify patterns in the source data.

SVM is, in fact, a supervised learning algorithm capable of solving classification problems [24, 25].

Guo et al. [22] has compared the result of classification, using one to five linear features extracted by PCA, FLDA, AFLDA, and MFLDA as the input to MDC. The results are shown in table 1.

Table 2, indicates a comparison of classification results for breast cancer diagnosis using different pattern recognition systems, which were explained previously. As it can be seen, the most accurate pattern recognition system is GP based on Modified Fisher criterion (MF-GP) with MDC with the accuracy of 98.94%.

C. Wavelet Analysis

As mentioned earlier breast cancer in women is one of the most common types of cancer, threatening many lives. The required medical image for the diagnosing process of breast cancer is mammogram, and is considered the most reliable method in early detection [26]. However, because of the high volume of images to be analyzed and lack of senior radiologists, the accuracy rate tends to decrease. Therefore, automatic reading of these digital medical images, with image processing and pattern recognition, is now highly desirable.

As previously-mentioned the main purpose is to differentiate benign and malignant masses. Mousa et al. [27] proposed a system based on wavelet analysis [28] of the image and the adaptive neuro-fuzzy interface system [29, 30] for creating the classifiers.

TABLE I. Comparison of classification accuracy *(%)* of PCA/MDC, FLDA/MDC, AFLDA/MDC and MFLDA/MDC, using WDBC data [22].

Algorithms	**Inputs**	**Best accuracy**(%)	**Average**(%)
PCA/MDC	2	88.76	88.62
FLDA/MDC	3	88.94	88.59
AFLDA/MDC	2	88.94	88.69
MFLDA/MDC	4	**89.29**	**89.01**

TABLE II. Comparison of classification accuracy *(%)* of stand-alone MLP and SVM, F-GP/MDC, AF-GP/MDC and MF-GP/MDC, using WDBC data [22].

Algorithms	**Inputs**	**Best accuracy (%)**	**Average (%)**	**Std (%)**
MLP	30	97.34	96.21	1.73
SVM	30	96.72	96.32	0.82
F-GP/MDC	1	98.77	97.40	1.60
AF-GP/MDC	1	98.42	97.36	1.39
MF-GP/MDC	1	**98.94**	**97.47**	**1.56**

A wavelet is a waveform of effectively limited duration that has an average value of zero [31]. Wavelet analysis is the breaking up of a signal into shifted and scaled versions of the original wavelet. The use of a fully scalable modulated window solves the single-cutting problem. The spectrum is calculated for the window, each time it is shifted. The same process is repeated with a shorter or longer window, for every new cycle [32].

When using wavelet analysis or many other techniques, three stages form the system: preprocessing, feature extraction and classification process. In the preprocessing stage, the quality of the image is improved in order to make the feature extraction phase more reliable. In the feature extraction stage, the features are extracted based on the wavelet decomposition process and are passed to classification stage, where the abnormality in digital mammograms are classified. According to Mousa et al. [27] globally processed image and the locally processed image are some of the best techniques used for classification.

III. CONCLUSION

Cancer is one of the leading causes of the death among men and women. Therefore accurate cancer detection methods can save many lives. Image processing and pattern recognition can play important roles in correct diagnosis of cancer. Different types of digital images are used for the detection of different types of cancer. Therefore, different techniques must be used to analyze these images. In this paper, three of the main techniques used in cancer detection were reviewed. Artificial Neural Networks are commonly used in Endoscopic Ultrasound images and because of their learning based adaption and ease of optimization, are very popular. Genetic Programming is widely used in feature classification of mainly breast cancer, and with the help of proper classifiers, highly accurate results can be achieved. Wavelet analysis is also a useful method, which is mostly applied on mammograms for breast cancer detection. It can also increase the accuracy of results, if being used with the proper classifiers. The techniques described in this paper are used on different types of medical imaging and data. The neural network technique described in this paper in commonly used the frames of CT scanned videos, while the wavelet analysis techniques is used on mammograms based on experiments done by the previous researchers. The GP is commonly used to extract features that allow pattern vectors belonging to different categories to distribute compactly and disjoint regions.

ACKNOWLEDGEMENT

This work is supported by University Teknologi Malaysia, Skudai Johor Bahru MALAYSIA and Ministry of Higher Education (MOHE) under Fundamental Research Grant Scheme (FRGS VOT 78182). Authors would like to thank *Soft Computing Research Group* (SCRG) for their moral support and incisive comments to improve this article.

REFERENCES

[1] C. Yelleswarapu, S. Kothapalli, and D. Rao, "Optical Fourier Techniques for Medical Image Processing and Phase Contrast Imaging", Optical Communications, 28, 2008, pp. 1876–1888.

[2] Y. Lee, and D. Tsai, "Computerized Classification of Microcalcifications on Mammograms using Fuzzy Logic and Genetic Algorithm", Proceedings of the SPIE on Medical Imaging: Image Processing, Vol. 5370, 2004, pp. 952–959.

[3] N. Ramirez, H. Acosta-Mesa, H. Carillo-Calvert, L. Nava-Fernandez, and R. Barrientos- Martinez, "Diagnosis of Breast Cancer using Bayesian Networks: A Case Study", Computers in Biology and Medicine, 37, 2007, pp. 1553–1564.

[4] G. Tourassi, B. Haarawood, S. Singh, J. Lo, and C. Floyd, "Evaluation of information-Theoretic Similarity Measures for Content Based Retrieval and Detection of Masses in Mammograms", Medical Physics, 34, 2007, pp.140–150.

[5] H.D. Cheng, X.J. Shi, R. Min, L.M. Ju, X.P. Cai, and H.N. Du, "Approaches for Automated Detection and Classification of Masses in Mammograms", Pattern Recognition, 39(4), 2006, pp. 646–668.

[6] K. Bovis, S. Singh, J. Fieldsend, and C. Pinder, "Identification of Masses in Digital Mammograms with MLP and RBF Nets", Proceedings of IEEE-IJCNN, 1, 2000, pp. 342–347.

[7] P. Lisboa, and A. Taktak, "The Use of Artificial Neural Networks in Decision Support in Cancer: A Systematic Review" , Neural Networks, 19, 2006, pp. 408–415.

[8] P. Lisboa, "A Review of Evidence of Health Benefit from Artificial Neural Networks in Medical Intervention", Neural Networks, 15, 2002, pp. 11–39.

[9] H. Georgiou, M. Mavrofarakis, N. Dimitropoulos, D. Cavouras, and S. Theodoridis, "Multi-scaled Morphological Features for The Characterization of Mammographic Masses using Statistical Classification Schemes", Artificial Intelligence in Medicine, 41, 2007, pp. 39–55.

[10] F. Gorunescu, M. Gorunescu, E. El-Darzi, and S. Gorunescu, "An Evolutionary Computational Approach to Probabilistic Neural Network with Application to Hepatic Cancer Diagnosis". CBMS, 2005, pp. 461-466.

[11] M. Giovannini, L. Hookey, E. Borie, C. Pesenti, G. Monges, and J. Delpero, "Endoscopic Ultrasound Elastography: The First Step Towards Virtual Biopsy Preliminary Results in 49 Patients", Endoscopy, 38(4), 2006, pp. 344-348.

[12] A. Saftoiu, P. Vilmann, H. Hassan, and F. Gorunescu, "Analysis of Endoscopic Ultrasound Elastography used for Characterization and Differentiation of Benign and Malignant Lymph Nodes". Ultraschall in der Medizin-European Journal of Ultrasound, 27, 2006, pp. 535-42.

[13] Y. Wu, M. GIger, K. Doi, C. Vyborny, R. Schmidt, and C. Metz, "Artificial Neural Networks in Mammography: Application to Detection Making in the Diagnosis of the Breast Cancer". Radiology, 187, 1993, pp. 81-7.

[14] C. Floyd, J. Lo, A. Yun, D. Sullivan, and P. Kornguth, "Prediction of Breast Cancer Malignancy using an Artificial Neural Network", Cancer, 74(11), 1994, pp.2944-8

[15] D. Furundzic, M. Djordjevic, and A. Bekic. "Neural networks approach to early breast cancer detection". Syst Architect, 44(8), 1998, pp. 617-33.

[16] A. Yardimci, "Soft computing in medicine.", Applied Soft Computing, 9, 2009, pp. 1029–1043

[17] J. Holland, "Outline for a Logical Theory of Adaptive Systems". J ACM, 1962, pp. 279–314.

[18] I. Rechenberg, "Cybernetic Solution Path of an Experimental Problem", Royal Aircraft Establishment Library Translation No.1122, 1965.

[19] J. Koza, "Hierarchical Genetic Algorithms Operating on Populations of Computer Programs". In: Sridhar an NS, editor. Proceedings of the 11th International Joint Conference on Artificial Intelligence, 1989. pp. 768–74.

[20] M. Brameier, and W. Banzhaf, "A Comparison of Linear Genetic Programming and Neural Networks in Medical Data Mining" , IEEE Trans. Evol. Comput. 5 (1), 2001, pp. 17–26.

[21] W. Worzel, J. Yub, A. Almala, and A. Chinnaiyan, "Applications of Genetic Programming in Cancer Research", The International Journal of Biochemistry and Cell Biology, 41, 2009, pp. 405–413.

[22] H. Guo, and A. Nandi. "Breast Cancer Diagnosis using Genetic Programming Generated Feature", The Journal of Pattern Recognition, 39, 2006, pp. 980 – 987.

[23] S. Chen, and X. Yang, "Alternative Linear Discriminant Classifier", Pattern Recognition, 37, 2004, pp. 1545–1547.

[24] H.F. Li, T. Jiang, and K.S. Zhang, "Efficient and Robust Feature Extraction by Maximum Margin Criterion" , Seventeenth Annual Conference on Neural Information Processing Systems, Vancouver, Canada, 2003, pp. 97–104.

[25] B. Heisele, T. Serre, S. Prentice, and T. Poggio, "Hierarchical Classification and Feature Reduction for Fast Face Detection with Support Vector Machines", Pattern Recognition, 36, 2003, pp. 2007–2017.

[26] K. Arun, "Computer Vision Fuzzy-Neural Systems". Englewood Cliffs, Prentice-Hall, 2001.

[27] R. Mousa, Q. Munib, and A. Moussa. "Breast cancer Diagnosis System Based on Wavelet Analysis and Fuzzy-Neural.", Expert Systems with Applications, 28, 2005, pp. 713–723.

[28] J. Portilla, and E. Simoncelli, "A Parametric Texture Model based on Joint Statistics of Complex Wavelet Coefficients". International Journal of Computer Vision, 40(1), 2000, pp. 49–71.

[29] R. Jang, "ANFIS: Adaptive-network-based Fuzzy Inference System", IEEE Transactions on System, Man and Cybernetics, 23(6), 1993, pp. 181–198.

[30] R. Jang, Sun, C., and E. Mizutani, "Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence". Englewood Cliffs, Prentice-Hall. 1997.

[31] G. Kaiser, "A Friendly Guide to Wavelets" Boston, Birkhauser, 1994.

[32] S. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation." IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(4), 1989, pp. 357–381.