Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 120 (2017) 402–410

Procedia Computer Science

www.elsevier.com/locate/procedia

9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW 2017, 24-25 August 2017, Budapest, Hungary

Machine learning techniques for classification of breast tissue

Abdulkader Helwan[a]*, John Bush Idoko[b], Rahib H.Abiyev[b]

[a]Biomedical Engineering Department Near East University, POBOX:99138, Nicosia, North Cyprus, Mersin 10, Turkey
[b]Computer Engineering Department Near East University, POBOX:99138, Nicosia, North Cyprus, Mersin 10, Turkey

Abstract

This paper presents an automated classification of breast tissue using two machine learning techniques: Feedforward neural network using the backpropagation learning algorithm (BPNN) and radial basis function network (RBFN). The two neural network models are implored basically to identify the best model for breast tissue classification after an intense comparison of experimental results. An electrical impedance spectroscopy method was used for data acquisition while BPNN and RBFN were the models implored for the execution of the classification task. The approach implored in this paper is made out of the following steps; feature extraction, feature selection and classification steps. The features are obtained using the electrical impedance spectroscopy (EIS) at the feature extraction stage. These extracted features are impedance at zero frequency (I0), the high frequency slope of phase angle, the phase angle at 500KHz, the area under spectrum, the maximum of spectrum, the normalized area, the impedance distance between spectral ends, the distance between the impedivity at I0 and the real part of the maximum frequency point and the length of the spectral curve. Information theoretic criterion is the strategy used in the proposed algorithm for feature selection and classification phase that was executed using the BPNN and RBFN. The performance measure of the two algorithms is the accuracy of the BPNN and RBFN models. The RBFN outperforms the BPNN in terms of accuracy in classifying breast tissues, minimum square error reached, and time to learn as demonstrated in the experimental results.

Peer-review under responsibility of the scientific committee of the 9th International Conference on Theory and application of Soft Computing, Computing with Words and Perception.

Keywords: Breast tissue; electrical impedance spectroscopy; neural networks; radial basis function network classifier.

* Corresponding author. . Tel.: +905488789719 ; fax: +903922236622
E-mail address: abdulkadir.helwan@neu.edu.tr

1877-0509 © 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the 9th International Conference on Theory and application of Soft Computing, Computing with Words and Perception.
10.1016/j.procs.2017.11.256

1. Introduction

Electrical impedance spectroscopy remains the key screening tool for breast tissue classification as well as the detection of abnormalities, because it allows identification of tumor before being palpable. Vacek et al. (2002) demonstrated that the proportion of breast tumors that were detected in vermont by screening mammography increased from 2% between 1974−1984 to 36% between 1995−1999. However, of all lesions previously diagnosed as suspicious and sent for biopsy, approximately 25% were confirmed malignant lesions, and approximately 75% were diagnosed benign. This high false-positive rate is related to the difficulty of obtaining accurate diagnosis as depicted by Basset and Gold (1987). For this reason, computerized image analysis plays an essential role in improving issues with diagnosis. Computer-Aided Diagnosis (CAD) systems are composed of a set of tools to help radiologists diagnose and detect new cases. Ho and Lam (2003) have shown that the sensitivity of these systems has significantly decreased as the density of the breast increased while the specificity of the systems remained relatively constant. The dataset used in this paper was deduced from the operations of electrical impedance spectroscopy (EIS) and could be found at the UCI repository.

The electrical impedance procedures have for quite some time been utilized in classifying tissue as well as impedocardiography applications by Kubicek et al. (1970). These strategies have additionally empowered impedance mapping by Tachibana et al. (1970) and Henderson and Webster (1994) and recently, dynamic imaging by Brown et al. (1994). The AC equivalent of resistivity for DC current is known as impedivity/particular impedance. The electric and dielectric properties dictates impedivity of a tissue and this depend, in addition to other things such as; membrane capacitance, cell concentration, intracellular medium and the electric conductivity in interstitial space as mentioned by Gabriel et al. (1996). Some of the good features of impedance techniques includes; minimum invasiveness, easiness and low cost.

Initially in the 80s, estimations of electric and dielectric has been performed in breast tissue under a scope of test conditions incorporating into vivo/ex-vivo estimations as well as utilizing different methods of measurement given in Surowiec et al. (1988), Campbell and Land (1992), Heinitz and Minet (1995). In the 488Hz-1MHz range, Jossinet (1998) found critical contrasts in phase angle and impedivity modulus from among the six groups of breast tissue.

EIS is conceivably applicable in breast cancer detection and breast tissue separation as proposed in the above discoveries. Using EIS, this paper demonstrate a strategy for the classification of breast tissues. Feature set utilized in this paper is the same as those features defined by Jossinet and Lavandier (1998) and also extra features chose for their separation capacity. Twelve-point and seven-point spectra were used to choose the statistical hierarchical approach.

A non-invasive strategy used to measure the impedance of cells in a scope of frequencies from a surface of tissue is termed an electrical impedance spectroscopy. Changes that occur in the nature of tissues are as a result of changes in impedance. Along these lines, fitness level of the fundamental tissue can be demonstrated by the variation of impedances. The above ideology makes electric impedance spectroscopy an essential strategy for detecting/diagnosing irregularities, cancer and abnormalities particularly to diagnose women breast cancer as it was mentioned by Kerner et al. (2002) and Zheng et al. (2008).

NN and RBFN classifiers are applicable in virtually every situation in which a relationship between the predictor variables (independents, inputs) and predicted variables (dependents, outputs) exists, even when that relationship is very complex and not easy to articulate in the usual terms of correlations or differences between groups. The type of problem amenable to solution by a neural network is defined by the way they work and the way they are trained. NN and RBFN work by feeding in some input variables, and producing some output variables. They can therefore be used where there is some known information, and would like to infer some unknown information.

The main objective of this research work is to train both NN and RBFN to predict which group of six classes of freshly excised tissue the breast tissue belongs, when it is given other attributes as input. First thing needed to execute this task is to have a dataset. The dataset used in this experiment can be found at UCI repository under classification category. The name of the dataset is breast tissue database. The Dataset contains information about electrical impedance measurements in samples of freshly excised tissue from the breast. Several constraints were placed on the selection of these instances from a larger database. This database includes 106 instances. Each instance belongs to one class. Six classes of freshly excised tissue were studied using electrical impedance measurements: Carcinoma, fibro-adenoma, mastopathy, glandular, connective and adipose. The characteristics (input attributes) that