

# Breast Tissue Classification based on Electrical Impedance Spectroscopy

LIU Chang

*Department of Communication Engineering, Xi'an University of Post & Telecommunication, Xi'an 710121, China*

CHANG Tiantian, LI Changxing

*Department of Mathematics, Xi'an University of Post & Telecommunication, Xi'an 710121, China*

**ABSTRACT:** The physiopathology state of human breast tissue can be reflected by the electrical impedance spectral characteristics. In this paper, breast tissue classification based on the electrical impedance spectral characteristics was proposed for the breast disease diagnosis in the early stage. 9 breast tissue characteristics were obtained by measuring the electrical impedance spectra of the breast tissue which is collected in vitro and fresh, and then the breast tissue can be classified by the support vector machine. The experimental results show that this method can classify the breast tissue effectively with accuracy in more than 80%. Particularly, the classification accuracy of breast cancer and fatty tissue are close to 100%.

**KEYWORD:** electrical impedance; breast tissue; support vector machine; spectral characteristic; spectroscopy

Breast disease is commonly happened in women, the incidence of the disease increases annually, the mortality rate is up to 40% above, and the disease tends to happened in younger person, it is serious threaten the modern women's health. Therefore, breast disease detection and diagnosis in the earlier stage has become as one of the most important issues in medicine. It was found that the electrical impedance spectral characteristics changes with the changes of breast tissue pathological physiological state[1].

Therefore, in this paper, the breast tissue state should be distinguished by the breast tissue spectra characteristics. That is to say, we can make a breast disease diagnosis for the patients. From the last century, the performance of the breast tissue measurement equipment improved with the development of measurement techniques, and the scholars found that the electrical impedance of the tissue is different under different measurement frequency, then, give birth to the concept of the impedance spectrum, and then the electrical impedance spectroscopy measurement technique developed[2].

From 1980s to 1990s, Jossinet and other scholars measures the electrical impedance spectrum of the breast tissue in different frequency range from 0.5kHz to 1MHz.

In the early study (1985), scholars measure the electrical impedance of the breast cancer tissues and normal adipose tissue, and the measurement results

show that[3]: cancer tissue resistance was significantly lower than normal adipose tissue while the operating frequency is 1000Hz. Since the breast tissue lesion can cause many types of disease, and the structure of these diseases are very complex, so the scholars take 64 patients, and divided them into six groups, and electrical impedance measurements were carried out. The 6 groups are: glandular tissue, connective tissue, adipose tissue, mastopathy, fibro-adenoma, and carcinoma. These six groups can be simply divided into two groups, the first three groups are the normal tissues, and others are lesion tissues. The first two of lesion tissues are benign disease, and the last one is the malignant disease. The scholars got the following results under the same experimental conditions, that is to say, the same measurement frequency range, measuring instruments and methods: adipose tissue and connective tissue have the highest impedance among the 6 groups, followed by carcinoma, the lowest one is fibro-adenoma, and the impedance of the glandular tissue and benign lesions have no significant difference.

The basic theories for classification of breast tissue were in the above. In this paper, the nonlinear classifier (SVM) was proposed for breast tissue classification via electrical impedance spectrum, and the experimental results show the effectiveness of the proposed method.

# 1 ELECTRICAL IMPEDANCE SPECTROSCOPY

Biological tissue contains large numbers of cells, the fluid of intracellular and extracellular can be regarded as an electrolyte, the tissue between cell membrane and cell membrane can be seen capacitance, so when we take DC or low frequency current is applied to biological tissue, the current bypass cells flowing through the extracellular fluid, the capacitance between the membrane and the membrane becomes smaller, when applied to increase the frequency of the biological tissue current, then part of the current will flow through the intracellular fluid across the cell membrane, so we can see that the biological tissue impedance greater at low frequency and high frequency impedance is small, and the characteristic of impedance change with frequency, this feature is called bioelectrical impedance spectrum [3].

In recent years, a series of tests was found for breast tissue electrical and dielectric measurements, in the frequency range 488Hz to 1MHz, the impedance and phase angle of the breast tissue has significant difference, so we preprocessing the data to obtain a set of electrical impedance characteristics in the initial various of cases, a significant difference was found with directly comparison, the data obtained from the literature can be seen carcinoma tissues and connective tissues are significantly different. In previous research, 4 impedance characteristics were used, such as IO, PA500, HFS, DA, in a recent study, we also used the new non-traditional features to complete better than the original report classification performance, these characteristics are AREA, A / DA, MAX IP, DR [4]. All of these features will be used in this paper. In this paper, To classify the 9 features of the 106(16 cases of glandular tissue, 14 cases of connective tissue, 22 cases of adipose tissue, 18 cases of mastopathy, 15 cases of fibro-adenomas, 21 cases of carcinoma) entity extraction.

## 2 SUPPORT VECTOR MACHINE [5] (SVM)

### 2.1 SVM algorithm

Support vector machine (SVM) is a new data mining technology, mainly used it to solve classification and regression problems. The essence of the algorithm is linear non separable problem converting a mapping for the linearly separable problem by computing, the mapping is from a low to a high-dimensional feature space and even infinite dimensional feature space, which is a non-linear mapping, this mapping is usually a lot of computation to be performed, but support vector machine algorithm is to use the kernel function for the low-dimensional to high-dimensional linear

mapping, so the kernel function is the key of the support vector machine algorithm, as long as the choice of appropriate kernel function, without knowing the explicit expression of the nonlinear mapping can be calculated, since in the high dimensional space, a linear operation is performed, so computation is huge, at the same time avoiding the complexity of calculating by increase dimensional brought up.

### 2.2 Formula derivation

The primal objective function is given as follow:

$$\min_{\omega, b, \xi} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l \xi_i$$

$$s.t. \quad y_i (\langle \omega, g(x_i) \rangle + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, i=1, \dots, l$$

$l$  represent the entirety number of samples,  $x_i \in R^n$   $x_i \in R^n$  are samples,  $y \in \{-1, 1\}$  is the category of  $x_i, i=1, \dots, l$ ,  $K(x_i, x_j)$  is kernel function, it corresponds to the inner product of feature space, that is  $K(x_i, x_j) = \langle g(x_i), g(x_j) \rangle$ , transform of  $g: x \rightarrow z$  take the sample mapping from input space to feature space,  $\omega$  is a hyper-plane normal vector,  $b$  is hyper-plane's bias,  $x_i$  is a variables,  $C$  is a penalty factor[6].

The dual problem of the SVM is:

$$\min_{\alpha} \rho(\alpha) = \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

$$s.t. \quad \begin{cases} y^T \alpha = 0, \\ 0 \leq \alpha_i \leq C, i=1, \dots, l \end{cases}$$

$Q$  of Hessian matrix is a semi-definite matrix,  $Q_{ij} = y_i y_j \langle g(x_i), g(x_j) \rangle = y_i y_j K(x_i, x_j)$ ,  $e = (1, 1, \dots, 1)^T$ ;  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_l)^T$ ,  $\alpha_i$  is a corresponding Lagrange operator that used constraints inequality of  $y_i (\omega \cdot f(x_i) - b) \geq 1 - \xi_i$ .

## 3 THE EXPERIMENTAL RESULTS

### 3.1 The experimental data

This paper uses the data set is the electrical impedance spectral characteristics of breast tissue UCI database [7], the data set has a wide range of features. In this paper, firstly, we normalized the data, then we used principal component analysis [8] for dimensionality reduction, and finally classify the multi-class data.

### 3.2 Model evaluation criteria

**Accuracy:** In a binary classification problem, accuracy rate is all categories correct (including true positive and true negative) result the proportion of the total sample. It is a parameter of classification predictable and result is a percentage value.

**Precision:** In a binary classification problem, precision is true positives (including true and false positives) proportion of all positive results. Five randomized experiments to get the final accuracy average of about 25.6/32. This shows that there are eighty percent of the samples were correctly classified. The highest accuracy of classification can be achieved 27/32.

Precision = Total number of samples correctly classified samples / training set

**Confusion Matrix:** The column of confusion matrix represents the actual measured information, the value of each column is equal to the actual measured value corresponding to the data quantity of each category. This paper uses the confusion matrix for the final model assessment, from the confusion matrix can clearly see how much data is correctly classified, how much of the data is wrong, and we know what is wrong divided into categories, so this is a relatively intuitive model evaluation methods.

Confusion matrix: the row number of the matrix is  $i$  and column number of the matrix is  $j$ , when  $i=j$ ,  $i$  denotes the class is correctly classified, the value represents the number of correctly classified; When  $i \neq j$ , it means that the number of class  $i$  was wrong divided into the  $j$  class.

### 3.3 Experimental results

In this paper, 74 samples for training, and the remained 36 samples for testing, training and testing samples were selected randomly from six class attribute in proportion. The number of samples were used to predict, there are 5 cases of glandular tissue, 4 cases of connective tissue, 6 cases of adipose tissue, 6 cases of mastopathy, 5 cases of fibroadenomas, 6 cases of carcinoma. The results of this experiment are universal, because each experimental data are randomly selected.

Experimental parameters, SVM classifier type choice nu-svc model, selection RBF kernel function, set nu-svc parameter is 2, the gamma value of kernel function is 0.1.

Table 1 shown the correct rate is determined by the sum of correctly classified samples of the five random experiment results divided by the total number of samples. For example, in the random experiment, the sum of carcinoma training samples

is 30, the number of a single experiment samples is 6, from the table we can see that 30 samples were correctly classified, that is, all samples were correctly classified, and therefore the accuracy of classification close to 100%, the other five categories of classification accuracy can be calculated the same way.

Table 1 Accumulate five experimental resulting confusion matrix

	Total accuracy	1	2	3	4	5	6
1	100%	30	0	0	0	0	0
2	72%	0	18	6	1	0	0
3	60%	4	3	18	5	0	0
4	72%	1	2	4	18	0	0
5	80%	0	0	1	0	16	3
6	96.67%	0	0	0	0	1	29
Total	80.625%	35	23	29	24	17	32

(In the table 1 represent carcinoma, 2 represent fibroadenomas, 3 represent mastopathy, 4 represent glandular tissue, 5 represent connective tissue, 6 represent adipose tissue)

Confusion matrix were used to conduct overall model evaluation, we can visually see from the resulting matrix how many samples of which class were correctly classified, so this is a very good assessment criteria. From the confusion matrix, we can see that the overall classification results can reach more than 80%, carcinoma classification accuracy can reach 100%, at the same time classification accuracy of adipose tissue are more than 95%, if the data set has a large data modeling, classification, the classification of adipose tissue will be increased to nearly 100%.

Table2 shows the classification accuracy obtained the original literature[7], we can see this paper's classification accuracy significantly higher than the original literature[7], indicating classification performance of this paper's classifier is better than the classifier used in the original literature, because the original document using a linear classifier classification, while the paper, we use the nonlinear support vector machine classifier (SVM), through normalized the sample data in the database, PCA dimension reduction, projected coordinate system, at the last feature extraction. Such a series of data processing and adjusting the parameters of classifier, classification accuracy is indeed a certain degree of improvement, it can be seen from the experimental results. However, the classification accuracy of six classes can't be all 100%, indicating that there is still a problem in this article, further research is needed to resolve.

Table 2 The original document[7] classification accuracy

	Accuracy
Carcinoma	81.82%
Fibro-adenomas	66.67%
Mastopathy	16.67%
Glandular tissue	54.54%
Connective tissue	85.71%
Adipose tissue	90.91%
Total	66.37%

#### 4 SUMMARY

The author of the breast tissue were classified which is based on the electrical impedance spectral characteristics, hoping to effectively carry out normal tissue and tumor identification. First, the results of this experiment show that the breast tissue classification with high accuracy of electrical impedance spectroscopy characteristics, but in order to achieve better classification results, we need to adjust the classifier. Secondly, if the increase in the number of samples in the data set, we might get a better classification results. Finally, while increasing the dataset, we should also have a more detailed division of entities sample, which is due to different periods of cancer of the extracted feature data there is a big difference, will inevitably affect the results of the classification, which need further study.

#### ACKNOWLEDGEMENTS

Fund: Foundtion of Xi'an University of Posts & Telecommunications (NO: ZL2012-20), "Study of electrical impedance imaging algorithm", 01/2012~12/2013.

#### REFERENCES

- [1] Debnath Bhattacharyya, Samir Kumar Bandyopadhyay, Tai-hoon Kim. Diagnosis of breast cancer by tissue analysis. Chinese Journal of Cancer Research, 2013; 25(1):39-45.
- [2] Li Xin, Li Honghong, Li Cong, Li Changwu. An optimization algorithm of automatic diagnosis of breast tissues based on electrical impedance spectroscopy. Chinese Journal of Biomedical Engineering, 2013; 32(2): 154-160.
- [3] Liao Qimei, Dong Xiuzhen, Fu Feng. Study on dielectric properties of human breast tissues. International Journal of Biomedical Engineering, 2006; 29(4): 218-221.
- [4] Yanwei Miao, Jianlin Wu. Clinical utility of proton magnetic resonance spectroscopy in the diagnosis of breast tumors. Chinese-German Journal of Clinical Oncology, 2008; 7(6): 326-311.
- [5] Chen Xiao, Li Lei, Fan Xiaogang. Support vector machine for nonlinear multi-user detection. Journal of Xi'an University of Posts & Telecommunications, 2008; 13(1) 82-85.
- [6] Fang Hui, Wang Qian. Research of algorithm on support vectot machine. Journal of Clangchun Normal University (Natural Science), 2007; 26(3): 90-91.
- [7] Silva JE, Marques de S á JP, Jossinet J. Classification of breast tissue by electrical impedance spectroscopy. Med & Bio Eng & Computing, 2000; 38: 26-30.
- [8] Zheng Chunhong, Jiao Licheng & Li Yongzhao. Support vector classifier based on principal component analysis. Journal of Systems Engineering and Electronics, 2008; 19(1): 184-190.