A Novel Classification Approach -1 on Breast Tissue dataset

Dr.G.Ayyappan,
Associate Professor, Bharath Institute of Higher Education and Research,Chennai
ayyappangmca@gmail.com

K.SivaKumar,
SIPS Technologies, Chennai, India.

Abstract

Data Mining is the process of discovering a novel pattern. Now days it occupies all sectors like engineering and medical. Electrical impedance spectroscopy is a minimally invasive technique that has clear advantages for living tissue characterization owing to its low cost and eases of use. Dataset with electrical impedance measurements of freshly excised tissue samples from the breast. This research work presents how this novel pattern can be applied to breast tissue classification. The Results of classification obtained from a data set of 106 cases representing six classes of excised breast tissue show an overall classification method.

Keywords: BayesNet, Function, Lazy, Meta, Misc, Rules, and Trees

1. INTRODUCTION

Breast cancer is typically diagnosed using a tissue biopsy. This is a process by which a sample of suspect tissue is removed from the patient and sent away for histological and chemical analysis. This process typically takes 1 to 2 days, but some samples give inconclusive results and require the patient to have a second biopsy. If a faster method of tissue analysis or a method to determine if a tissue sample is a good candidate for further screening, the cost of screening and stress endured by the patient in awaiting results could be reduced.

Electrical Impedance Spectroscopy (EIS) is a method that can used to measure the complex impedance properties of a material. A voltage or current of known frequency and amplitude is applied across the system and the response current or voltage is recorded. The differences in phase and magnitude of the applied stimulus and response are used to determine the complex impedance at that frequency. This is repeated for a range of frequencies to determine how the complex impedance of the system changes with stimulus frequency. Results are often displayed in a nyquist plot, showing the resistive and reactive components of impedance at each frequency measured. Often these plots will have characteristic shapes; for example, figure 1 shows the characteristic curved nyquist plot for a simple RC circuit.

In this research work section 1 contains the introduction about this research work. Section 2 focuses the literature reviews of this research work. In Section 3 contains the Materials and methods of this breast tissue dataset. In Section 4 contains the results and discussions of this research work and finally, it focuses on conclusion of this research work.

2. RELATED WORKS

Electrical impedance techniques have long been used for tissue characterisation and in monitoring applications, of which impedocardiography is the best known (KUBICEK et al., 1970). These techniques have also enabled impedance mapping (TACHIBANA et al., 1970; HENDERSON and WEBSTER, 1978) and, more recently, dynamic imaging (BROWN et al., 1994).Specific impedance (also termed 'impedivity') is the AC equivalent of resistivity for DC current. The specific impedance of a tissue is determined by its electric and dielectric properties, which depend, among other things, on the cell concentration, membrane capacitance, electric conductivity in interstitial space and the intracellular medium (SCHWAN, 1959; FOSTER and SCHWAN, 1989). The easiness, low cost and minimum invasiveness are praised features of the impedance techniques.

In recent decades, electric and dielectric measurements have been carried out in breast tissue under a range of experimental conditions including in-vivo or ex-vivo measurements and using various measurement techniques (SuROWIEC et al., 1988; MORIMOTO et al., 1990; CAMPBELL and LAND, 1992; HEINITZ and MINET, 1995). In the 488Hz-1 MHz range, significant differences in the impedivity modulus and phase angle from among six groups of breast tissue were found by one of the authors (JoSSINET, 1998).

These findings suggest that electrical impedance spectroscopy (EIS) is potentially usable for the discrimination of breast tissue and especially for the detection of breast cancer. The present paper describes a method for classifying breast tissues based on EIS. These new results were attained using features derived from the Argand plot of the impedivity data collected in freshly excised tissue. The set of features used comprised those defined in a previous study (JosSrNET and LAVANDIER, 1998)

3. MATERIALS AND METHODS

In this section, the dataset borrowed from http://archive.ics.uci.edu/ml/datasets/breast+tissue. This dataset has asmultivariate type, it has 106 instances, Attribute characteristics are real and 10 attributes. The dataset can be used for predicting the classification of either the original 6 classes or of 4 classes by merging together the fibro-adenoma, mastopathy and glandular classes whose discrimination is not important

Table1: Dataset Description

S.No	Dataset content	Properties
1	Dataset characteristics	Multivariate
2	Attribute Characteristics	Real
3	Numbere of Instances	106
4	Number of attributes	10

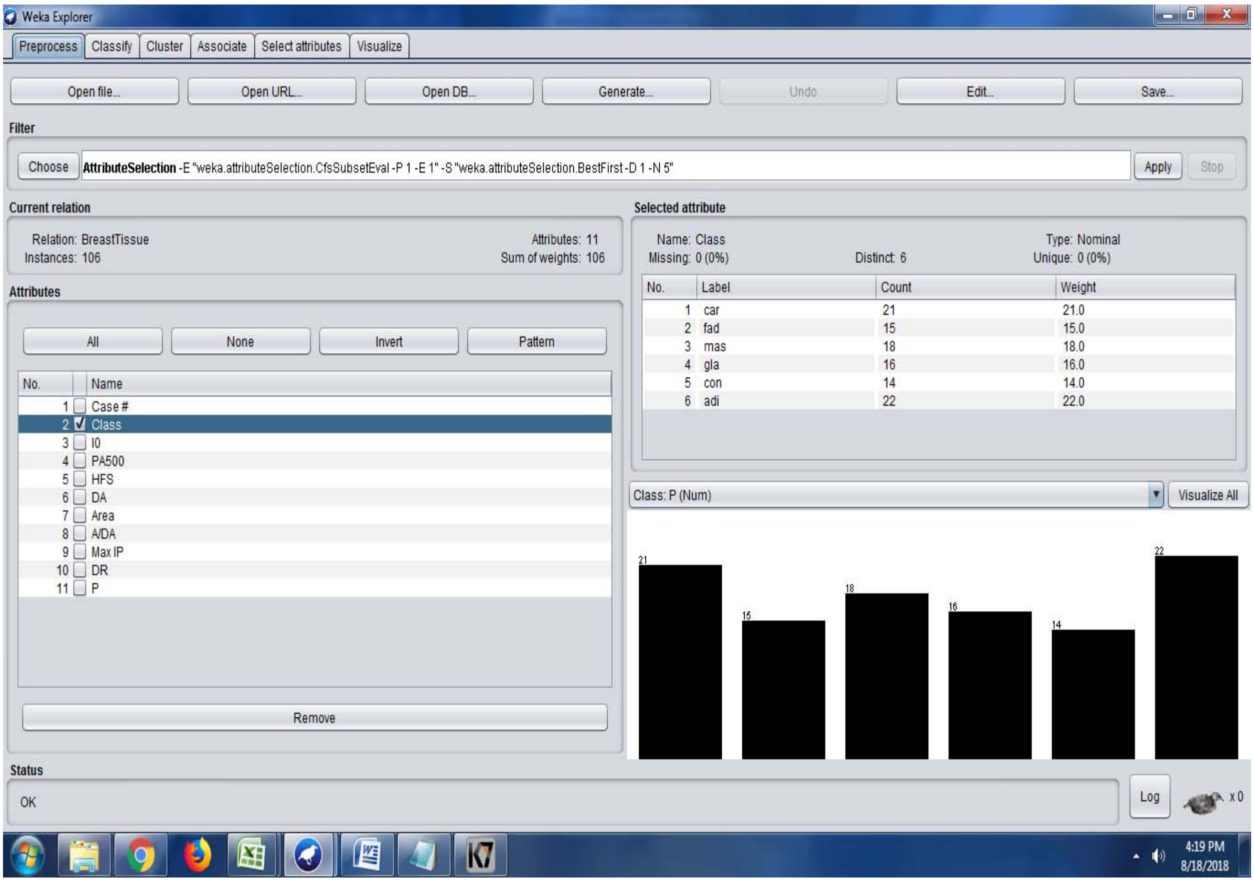


Fig 1:Class representation in weka

The above diagram represents the the dataset implemented and class representation in weka machine learning tool.

Table 2: Information about an Attribute

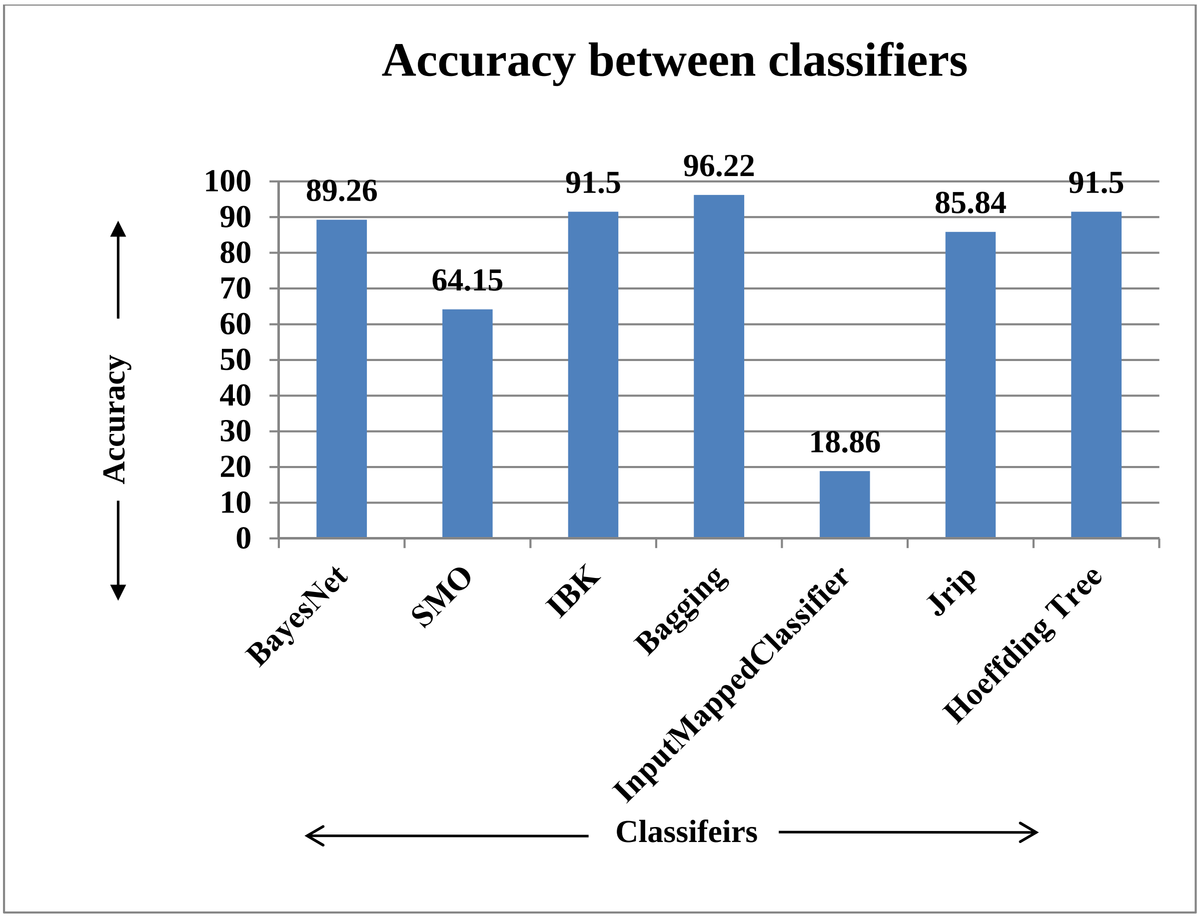
S.No	Acronym	Meaning
1	I0	Impedivity (ohm) at zero frequency **(**15.625, 31.25, 62.5, 125, 250, 500, 1000KHz.**)**
2	PA500	phase angle at 500 KHz
3	HFS	high-frequency slope of phase angle
4	DA	impedance distance between spectral ends
5	AREA	area under spectrum
6	A/DA	area normalized by DA
7	MAX IP	maximum of the spectrum
8	DR	distance between I0 and real part of the maximum frequency point
9	P	length of the spectral curve(real,imaginary)
10	Class	car(carcinoma), fad (fibro-adenoma), mas (mastopathy), gla (glandular), con (connective), adi (adipose).

Above mentioned dataset has utilized by weka tool. It has one of the leading machine learning tool. Here we implemented most leading classifier methods namely, BayesNet from Bayes, SMO from Function, IBK from Lazy, Bgging from Meta, InputMappedClassifier from Misc, JRip from Rules and Hoeffding Tree from Tree classifiers. Before classifying this dataset first preprocessed this data set supervised learning method and attribute selection method and then classified all the leading classifications. The training set tests 10 cross validation and produced several output.

4. RESULTS AND DISCUSSIONS

In this section, the novel approach presents the below output. The Bayesnet classifier belongs to Bayes classifier. It has 89.26% accuracy level. SMO classifier belongs to functions and it has 64.15% accuracy level.IBK belongs to Lazy classification. It has 91.50%.The Bagging classifier belongs to Meta classification. It has 96.22% accuracy level. The InputMappedClassifier belongs to Misc.It has 18.86% accuracy level. The JRip classifier belongs to JRip. It has 85.84% accuracy level.The Hoeffding Tree classifier belongs to Tree classification.It has 91.50% accuracy level.

S.No	Base Classifier	Classification	Correctly Classified Instances	Accuracy	Time Taken to build the model (In Seconds)	Kappa Statistic
1	**Bayes**	**BayesNet**	**95**	**89.26%**	**0.05**	**0.8749**
2	**Function**	**SMO**	**68**	**64.15%**	**0.71**	**0.5632**
3	**Lazy**	**IBK**	**97**	**91.50%**	**0**	**0.8975**
4	**Meta**	**Bagging**	**102**	**96.22%**	**0.07**	**0.9545**
5	**Misc**	**InputMappedClassifier**	**20**	**18.86%**	**0**	**-0.0215**
6	**Rules**	**JRip**	**91**	**85.84**	**0.06**	**0.8283**
7	**Tree**	**Hoeffding Tree**	**97**	**91.50**	**0.07**	**0.8976**



The above diagram demonstrates the Bagging, IBK and Hoeffding Tree classifiers produce the highest accuracy lies in above 90%, Bayesnet and JRip classifiers have the accuracy level in between 80% to 90%. The SMO classifier has 64.15%. Finally InputMapped Classifier has lowest accuracy level.

5. CONCLUSION

In this paper focuses the novel approach of these classification methods. The Bagging Classifier belongs to Meta classification produces the highest accuracy level 96.22% and very lowest accuracy level is InputMapped Classifier has 18.86%. It belongs to Miscellaneous classifier. The accuracy of the remaining classifiers namely IBK , Hoeffding ,SMO, Bayesnet and JRip are appearing in between these highest and lowest classifier

References:

[1] Jossinet J (1996) Variability of impedivity in normal and pathological breast tissue. Med. & Biol. Eng. & Comput, 34: 346-350. Silva JE, Marques de Sá JP, Jossinet J (2000) Classification of Breast Tissue by Electrical Impedance Spectroscopy. Med & Bio Eng & Computing, 38:26-30.
[2] http://archive.ics.uci.edu/ml/datasets/breast+tissue
[3] http://homepages.cae.wisc.edu/~ece539/fall13/project/Nonte_rpt.pdf
[4] American Cancer Society.
[5] http://www.cancer.org/treatment/understandingyourdiagnosis/examsandtestdescriptions/testingbiopsyandcytologyspecimensforcancer/testing-biopsy-and-cytology-specimens-for-cancer-how-long-doestesting-take. Accessed 7 June. 2018.
[6] Lvovich, V. F. (2012). Impedance spectroscopy: applications to electrochemical and dielectricphenomena. John Wiley & Sons.
[7] Williams, J. C., Hippensteel, J. A., Dilgen, J., Shain, W., & Kipke, D. R. (2007). Compleximpedance spectroscopy for monitoring tissue responses to inserted neural implants. Journal of neuralengineering, 4(4), 410.
[8] Jossinet, J. (1996). Variability of impedivity in normal and pathological breast tissue. Medical andBiological Engineering and Computing, 34(5), 346-350.
[9] UCI Machine Learning Repository. [http://archive.ics.uci.edu/ml/datasets/Breast+Tissue]. Accessed 7 June. 2018.
[10] Daliri, M. R. (2013). Combining extreme learning machines using support vector machines forbreast tissue classification. Computer methods in biomechanics and biomedical engineering, (ahead-ofprint),
[11] [11]https://www.researchgate.net/publication/12489518_Classification_of_breast_tissue_by_electrical_impedance_spectroscopy
[12] [12]https://link.springer.com/article/10.1007/BF02344684