



ELSEVIER

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

A comparative study of PCA, SIMCA and Cole model for classification of bioimpedance spectroscopy measurements



Isar Nejadgholi*, Miodrag Bolic

School of Electrical Engineering and Computer Science, University of Ottawa, Canada

ARTICLE INFO

Article history:

Received 8 January 2015

Accepted 1 May 2015

Keywords:

Bioimpedance spectroscopy

Classification

Cole modeling

Principal Component Analysis

SIMCA

ABSTRACT

Due to safety and low cost of bioimpedance spectroscopy (BIS), classification of BIS can be potentially a preferred way of detecting changes in living tissues. However, for longitudinal datasets linear classifiers fail to classify conventional Cole parameters extracted from BIS measurements because of their high variability. In some applications, linear classification based on Principal Component Analysis (PCA) has shown more accurate results. Yet, these methods have not been established for BIS classification, since PCA features have neither been investigated in combination with other classifiers nor have been compared to conventional Cole features in benchmark classification tasks.

In this work, PCA and Cole features are compared in three synthesized benchmark classification tasks which are expected to be detected by BIS. These three tasks are classification of before and after geometry change, relative composition change and blood perfusion in a cylindrical organ. Our results show that in all tasks the features extracted by PCA are more discriminant than Cole parameters. Moreover, a pilot study was done on a longitudinal arm BIS dataset including eight subjects and three arm positions. The goal of the study was to compare different methods in arm position classification which includes all three synthesized changes mentioned above. Our comparative study on various classification methods shows that the best classification accuracy is obtained when PCA features are classified by a K-Nearest Neighbors (KNN) classifier. The results of this work suggest that PCA+KNN is a promising method to be considered for classification of BIS datasets that deal with subject and time variability.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Bioimpedance spectroscopy (BIS) is a safe, non-invasive and low cost method to measure electrical responses of living tissues to a low-level, alternating current at a range of frequencies. Single-frequency and multi-frequency bioimpedance have been measured and modeled in many research applications to explain how different sources of variation affect the response of a living tissue [1,2]. These applications have been diverse including impedance plethysmography and pulsatile blood flow [3,4], assessment of human body composition [5], hydration detection, characterization of fluid accumulation [6,7] and detection of electrical anomalies in neuromuscular diseases [8].

However, classification of measured bioimpedance spectra in order to make a diagnostic decision is a particular challenge [9,10]. The signal is recorded on the surface of a part of the body and reflects the internal phenomena in the body. Several decisions

regarding proper classifications of these signals need to be made including: (1) determining essential features or codes that encompass the information relevant to the diagnosis decision, (2) with a set of given features, determining kind of classifier to be trained to form the best decision boundaries in the feature space and to discriminate different classes of the recorded signal, (3) setting the required accuracy for a set of new recorded signals.

These are typical steps of the so-called classification task in the field of signal processing. In contrast to widespread used classification methods applied to ECG and EEG signals, standard classification methods for BIS data have not been established yet [11,12].

Cole parameters are the most common descriptive parameters extracted from bioimpedance measurements. In this method the measured spectrum is fitted onto an equivalent electrical circuit and the information of spectrum is summarized to four descriptive features based on similarities between the electrical circuits and the biological tissues. Cole parameters have been basically introduced and applied as explanatory features of bioimpedance measurements and have been known to have limitations [13]. Simplicity, popularity and explanatory power of these features make them the first choice and the standard way of descriptive

* Corresponding author

E-mail addresses: inejadgholi@gmail.com, inejadgh@uottawa.ca (I. Nejadgholi).

feature extraction, as well. However, researchers found that Cole parameters extracted from repetitive measurements on the same subject are too variable and thus not effective in achieving a classification task [14,15].

Principal Component Analysis (PCA) and Soft Independent Modeling of Class Analogy (SIMCA) are two techniques that have been reported before as successful feature extraction methods in BIS classification tasks [16–21]. PCA maps the whole data on its orthogonal and uncorrelated eigenvectors and extracts the most salient part of information on first eigenvectors as features. SIMCA implements the same decorrelation, but within each class separately and considers the distance between each data point and eigenvectors of each individual class as features. For example, in [16], PCA and linear classifiers were applied for skin type classification. First the measurements were classified to men and women based on PCA scores of capacitance. Then skin type classification was achieved for men and women separately, using PCA scores of impedance. In [17], SIMCA was applied on the magnitude of impedance to classify different skin types. In [18], PCA was applied on complex data considering each complex measurement as a set of two real values (real and imaginary) and the components were fed to a Linear Discriminate Analysis (LDA) classifier to classify different types of fish. In [19] SIMCA was applied on complex values consisting of magnitude and phase in order to classify contracted and relaxed muscles. In this work, the authors have found SIMCA to be more effective than PCA in achieving the classification task. In [14] the authors have used electrical features and decision trees to achieve a successful classification for diagnosing stroke in the brain.

The reason why most of researchers prefer to use Cole parameters and linear discriminant classifiers is the high explanatory power of these methods, which enables us to interpret our observations based on related biophysical phenomena. However, increasing number of applications of PCA-based feature extraction methods in BIS classification suggests that there is an emerging trend toward using signal processing methods in achieving classification of measured bioimpedance spectra. As mentioned in [22], in some applications of bioimpedance high classification performance compensates the need for an explanatory model. In previous works PCA has only been used in combination with linear classifiers. However, if the objective is to achieve high classification accuracy without the need of providing physical explanations, then using more powerful classifiers is a promising option. Also neither PCA nor SIMCA has been compared quantitatively in specific classification tasks, which is essential in order to establish new methods as standard techniques.

In this work, we first compare the above-mentioned methods in achieving three synthetic benchmark classification tasks. In each classification task we simulate two classes of data related to the tissue before and after a specific change that is known to affect the frequency response of the tissue. This change is expected to be detected by bioimpedance spectroscopy. These three changes are *change in relative composition of muscle and fat in the tissue*, *blood perfusion in a tissue consisting of muscle and fat* and *change of the tissue geometry*. For each classification task, an equivalent electrical circuit is used to synthesize a number of sample spectra related to before and after the change. PCA and Cole methods are then applied and compared in classifying these simulated datasets. Secondly, these methods are combined with four different types of classifiers (linear, quadratic, decision tree and K-nearest neighbors) and compared in classification of a set of experimentally measured bioimpedance data which in theory involves all three above-mentioned changes. This dataset includes longitudinal bioimpedance measurements of arm for eight subjects, in three different arm positions. Classification task involves classifying measured bioimpedance spectra to detect arm positions.

The rest of this paper is organized as follows. In Section 2, we focus on explaining feature extraction methods used in this paper which are Cole method, PCA and SIMCA. We also clarify what we mean by the term classification. In Section 3, we explain all the methods that we used in simulating and measuring BIS data, as well as classification tasks that we performed to compare the classification methods. Section 4 presents all the obtained results and compares the proposed methods quantitatively in classification tasks given in Section 3. Section 5 includes conclusion, discussion and suggestions for future work.

2. Methods

2.1. Cole feature extraction

Based on Cole model introduced in [23] impedance of a tissue can be written as

$$Z(f) = R_{\infty} + \frac{R_0 - R_{\infty}}{1 + j \left(\frac{f}{f_c}\right)^{\alpha}} \quad (1)$$

where, $Z(f)$ is the complex impedance at frequency f , R_0 and R_{∞} represent resistivity of tissue at zero and infinity frequency, respectively, f_c is the characteristic frequency related to the relaxation time of electrical dispersion and α is a constant added to this mathematical model to explain molecular interactions. $\alpha = 1$ represents the absence of molecular interaction and corresponds to an ideal capacitor in equivalent electrical circuit. As a whole, the Cole equation described in Eq. (1) is a mathematical expression for the impedance dispersion corresponding to bioimpedance data found experimentally. For extracting features based on the Cole model, a given measured spectrum is parametrized by fitting to this equation and four Cole parameters, $[\alpha, f_c, R_0, R_{\infty}]$ are obtained as representation of the measured spectra. It should be noted that Eq. (1) is equivalent to the equation introduced in [23]. In [23], ω and τ are used where $\omega = 2\pi f$ and $\tau = 1/f_c$. However, since f_c is the frequency corresponding to the epic of Nyquist plot (-imaginary versus real) of impedance and can be obtained from this plot as well as α , R_0 and R_{∞} , many researchers use the notation of Eq. (1) for simplicity. As mentioned above, these parameters can be interpreted based on similarities of electrical and biological tissues, and so are helpful in explaining BIS measurements. In this paper we will investigate whether they are also effective as extracted features for a classification task or not.

2.2. Principal component analysis

Principal Component Analysis assumes that there is a linear orthogonal basis for representation of a given data space and uses decorrelation techniques to find this basis [24]. In the context of bioimpedance, let N denote the number of frequencies in which the impedance is measured and M is the number of taken bioimpedance measurements. Having formed the $M \times N$ data matrix, X , we assume that the number of uncorrelated data variables, r , where $r = \text{Rank}(X)$, is less than the number of frequencies, N . PCA forms the covariance matrix of the mean-centered data matrix, $X - \text{mean}(X)$, which includes covariances between frequencies. This covariance matrix $C_X = [c_{ij}]$ can be written as

$$c_{ij} = \sum_{k=1}^M x_{ki}x_{kj}, \quad i, j = 1, 2, \dots, N \quad (2)$$

where x_{ij} is representing the element of mean-centered data matrix corresponding to measured impedance in i th observation

Download English Version:

<https://daneshyari.com/en/article/505218>

Download Persian Version:

<https://daneshyari.com/article/505218>

[Daneshyari.com](https://daneshyari.com)