Dimensionality reduction methods for Impedance Spectroscopy data of biological materials

R Cavalieri¹, P Bertemes-Filho¹
¹Dept. of Electrical Engineering, State University of Santa Catarina, Joinville - SC, Brazil.
E-mail: ricardo.cavalieri@edu.udesc.br

Abstract. Electrical impedance spectroscopy combined with Neural Networks can be a powerful combination to identify biological materials. This paper utilizes a data set containing two biological samples taken from different species and applies the most popular methods of dimensionality reduction. This is done in order to find out which method is able to minimize computational demand and maximize accuracy in the classification test. This paper proposes that the classic PCA method is the fastest and the most accurate under the configurations used.

1. Introduction
Electrical impedance spectroscopy (EIS) has been proven to be a useful tool in the analysis of the composition of several different kinds of materials, both organic and inorganic [1]. The extraction of electrical parameters that describe the observed tissue with accuracy through impedance analysis has been used for applications such as characterization of both animal [2] and plant [3] biological tissues.

Analysis is made through the injection of a known alternated current with frequencies normally ranging from 10 Hz to 10 MHz and measuring the voltage drop that happens for each of the frequency points observed [4]. One example of complex impedance obtained from the EIS technique applied to a biological tissue can be observed in Fig 1, where two separate plots describe the behaviour of the real $Re(\Omega)$ and imaginary $Im(\Omega)$ components of the complex impedance $Z$.

This data is then used to create a model for the tissue observed through a range of different methods such as the Cole function [5] or using an equivalent electrical circuit as a representation of the model, representing the cellular membranes as capacitors and the extracellular and intracellular material as resistors [6].

Such analysis requires that the electrical response of the system be simplified in order to be comprehensively represented in mathematical or circuital elements, which can imply that some of the properties of the biological tissue can be overlooked by the model created [7].

With the recent cost reduction of powerful processing units, the implementation of Artificial Neural Networks (ANNs) as a mean of interpretation of biological data (such as EIS) has become an alternative [8]. Its ability to represent complex non-linear behaviour [9] can be a powerful feature for the creation of more accurate biological tissue models.

In order to potentialize the use of Neural Networks for biological tissue characterization, it is important that data is properly prepared for training. Considering the high dimensionality of EIS analysis, different methods for dimensionality reduction can be applied to the process. These are especially useful when measurement is taken on a broad band of frequencies and require low-cost hardware or a short processing time [10].
Figure 1. Raw data obtained from an Impedance spectrometer connected to a potato tuber. Both components of the complex impedance for each frequency point are displayed in the image from 400Hz to 1MHz.

Figure 2. Example of dimensionality reduction. A helicoidal function is used to translate the three-dimensional space to a two-dimensional space. [10]

In essence, dimensionality reduction works as represented in Fig. 2, where a known function (in this case, representing a helicoidal behaviour) transforms the original data space $\chi$ into the component space $Z$ following the $\Phi_{extr}$ transformation. Conversely, the new space is translated back by utilizing the $\Phi_{gen}$ transformation.

This process can be implemented in a variety of methods, achieving better results depending on the characteristics of the original data [11]. Some of the more popular methods are Principal Component Analysis (PCA) [12], Locally Linear Embedding (LLE) [13], Metric Multidimensional Scaling (mMDS) [14], and Isomaps [15] Considering that the classification of biological tissues using EIS is a relatively new application, there is no consensus on which dimensionality reduction method would perform better applied to an electrical impedance spectroscopy dataset.
Each method will be benchmarked considering its ability to better represent the original data points while maximizing efficiency of the newly generated space using a well-defined competitive framework [16]. An exploration of each methods ability to translate the non-linear behaviour of the original dataset is to be conducted, ranking each technique explored according to its performance.

2. Methodology

The dataset used for evaluation of the dimensionality reduction methods consists of bioimpedance readings of two plant-based biological materials (Solanum tuberosum L. and Malus domestica), extracted with the Zurich HF12S Impedance Analyzer. Each sample is probed with 4 conductive needles of 50mm length and separated each by 80 mm. Measurements are repeated 50 times for each of the biological tissues analyzed. Full dataset is composed of a 100 individual impedance spectroscopy measurements. Frequency analyzed ranged from 400 Hz to 1 MHz, recording both absolute impedance values (|Z|) and the phase angle (θ) for 200 frequency points (increasing in logarithmic scale) [17]. Therefore, input data would consist of a matrix with 2 rows (representing absolute impedance and phase angle) and 200 columns (one for each frequency point where complex impedance is analyzed).

The configuration of the dataset can be better described in Table 1. It represents the data of one sample and contains the impedance and phase angle for each point of frequency.

|Freq.  | 400 Hz | 432 Hz | ... | 0.924 MHz | 1 MHz |
|-------|--------|--------|-----|-----------|-------|
| | 1409.67 Ω | 1314.58 Ω | ... | 587.31 Ω | 587.31 Ω |
|θ | 49.28º | 49.01º | ... | 2.47º | 0.74º |

Data is extracted from the original CSV file and prepared to the dimensionality reduction process via the pandas module for Python [18], using the DataFrame function for later analysis utilizing the original Neural Network, implemented via Python’s sci-kit learn Module. The prepared data set will be then divided into 10 subsets and used to populate the graphical representations employed in this paper.

All proposed methods of dimensionality reduction are implemented via the scikit-learn Python module [19].

Principal Component Analysis (PCA) is applied to the data for 400 features, utilizing the full singular value decomposition (SVD) via the standard LAPACK solver [20]. Number of kept components are set to 10 values which are selected in post processing.

Locally Linear Embedding (LLE) is trained for 10 components, considering 5 neighbor values. Eigenvalues solver is set to dense, which in turn uses the LAPACK solver. Random state term is not defined.

Metric Multidimensional Scalings (mMDS) is again applied for an output of 10 components. Verbosity is considered 0, to a maximum of 300 iterations. Convergence is declared after the Stress parameter reaches the value of 0.001.

Isomap configuration is similar to the one used with the LLE method. Neighboring values are set to 5, utilizing the Eigenvalues solver set to dense with the LAPACK solver. Shortest path finding method used is the Floyd-Warshall algorithm [21].

An overview of the specifics used for each method can be found in Table 2.

The resulting matrix obtained from each method will be graphically represented in a low-dimensional space for evaluation in order to check if the generated data has linear separability.

Method ranking will be based on the id estimation proposed in [16]. Considering that the objective of the original dataset was to achieve a binary classification, the transformed data is going to be tested against a simple classification neural network.

Neural network used for validation consists of a multi-layer Perceptron (MLP) implemented in Python with the scikit-learn module. Weight optimization solver used is L-BFGS with a Rectified Linear Unit


Table 2. General configuration of each applied Dimensionality Reduction method

| Method | Components | Neighbors | Solver          | Max. Iter. |
|--------|------------|-----------|-----------------|------------|
| PCA    | 10         | 15        | LAPACK          | 500        |
| LLE    | 10         | 5         | LAPACK (dense)  | 300        |
| mMDS   | 10         | 5         | LAPACK (dense)  | 300        |
| Isomap | 10         | 5         | LAPACK (dense)  | 300        |

(ReLU) as the activation function. Maximum iterations is set to 500 over 5 hidden layers. Learning rate (α) is set to 0.0001.

Considering that in the original publication of the dataset, the classification network utilizing the full dataset achieved 100% accuracy [17], successful dimensionality reduction is expected to achieve similar results.

3. Results and Discussions

In order to represent the process of dimensionality reduction graphically, every method implemented will be represented in a three dimensional space. This is done in order to visually identify the methods that can most successfully classify the biological materials.

First dimensionality reduction applied is PCA. As expected, the processing time required was the lowest among the methods tested, clocking in at 36.01 ms. Classification utilizing the PCA values obtained obtained a 98.6% score, similar to the network utilizing the full data set.

As can be observed in Fig 3, one type of biological material is well concentrated in the plot, while the other type of material can be found all over the grid. This suggests that the process was able to facilitate the neural network classification.

![Figure 3. Three dimensional representation of PCA applied to the data set.](image)

Locally Linear Embedding took almost twice the processing time as compared to the PCA process: 73.01 ms. LLE performed at the classification taking the score of 88.2%.

It is possible to observe in Fig 4 that the LLE fails to clearly separate in a three dimensional space the two types of biological materials. This could be one of several explanations as to why the system wasn’t able to beat a simple PCA method in a classification score.
The mMDS method is by far the most hardware demanding method utilized in this paper, its processing time clocked in at 882.19 ms. Even being the most processing intensive method in this paper, the mMDS had a comparatively bad score regarding the classification test via neural network, scoring 77.9%.

Considering the three dimensional representation in Fig. 5, it is possible to conclude that this particular method was not able to identify the samples biological features, as it appears to be no boundary between the two kinds of samples.
Isomaps ranked second among the methods tested in this paper, with a score of 89.8%.

It is possible to identify the concentration of one type of material in Fig. 6 to the center of the plot which indicates that this method is capable of identifying and separating the specific characteristics of the samples.

![Figure 6. Three dimensional representation of Isomaps applied to the data set](image)

A summary of the results obtained can be observed in Table 3 organized in order of lowest processing time and highest classification score.

| Method                    | Processing time [ms] | Classification score [%] |
|---------------------------|----------------------|--------------------------|
| PCA                       | 36.01                | 98.6                     |
| Isomap                    | 64.01                | 89.8                     |
| LLE                       | 73.01                | 88.2                     |
| mMDS                      | 882.19               | 77.9                     |
| No dimensionality reduction [17] | 223                  | 100                      |

The results shown here would be drastically different considering the behaviour of other types of data. For this reason, the comparison between methods of dimensionality reduction utilizing impedimetric data of biological materials can represent a substantial contribution to the developing area of computational methods applied to impedance spectroscopy data.

4. Conclusion
For this application, the best suited method for dimensionality reduction is PCA. In the tests performed, this method was shown to be the fastest, most accurate with classification and even provided the best three dimensional representation of the data set. This agrees with the consensus in the scientific community about using PCA for dimensionality reduction of signals extracted from biological materials.

Even though the other methods achieved a considerably high classification score, it is important to remember that the original data set without any sort of dimensionality reduction applied to it, achieved
100% classification rate. Considering that, it is possible to conclude that some of the methods presented here will help to make neural network classification faster (and not more accurate).

Isomap and LLE achieved similar results both in classification score and processing time. Both reach a classification score close to 90%, but are considerably less accurate and requires more processing time when compared to PCA.

mMDS on the other hand, performed poorly not only on the classification score but also demanded over 24 times the processing time required for a PCA analysis. From this it is possible to determine that under the configurations implemented in this paper, mMDS is not suitable for the application. On top of that, the method underperformed the classification process utilizing the raw data gathered both in time and accuracy. In summary, the mMDS method not only did not reduce the computation time required to apply classification to the raw data but introduced error to the classification.

Future studies should utilize different configuration variables for each methods, trying to find out if they can perform better under different configurations.

Acknowledgments

Funding for this research was provided by FAPESC (Fundação de Amparo À Pesquisa e Inovação do Estado de Santa Catarina) and UDESC (Universidade do Estado de Santa Catarina).

References

[1] J. R. Macdonald et al., Impedance spectroscopy. Wiley New York ete, 1987, vol. 41.
[2] D. A. Dean, T. Ramanathan, D. Machado, and R. Sundararajan, “Electrical impedance spectroscopy study of biological tissues,” Journal of electrostatics, vol. 66, no. 3-4, pp. 165–177, 2008.
[3] L. Wu, Y. Ogawa, and A. Tagawa, “Electrical impedance spectroscopy analysis of eggplant pulp and effects of drying and freezing–thawing treatments on its impedance characteristics,” Journal of Food Engineering, vol. 87, no. 2, pp. 274–280, 2008.
[4] U. Pliquett and M. R. Prausnitz, “Electrical impedance spectroscopy for rapid and noninvasive analysis of skin electroporation,” in Electrochemotherapy, Electrogenetherapy, and Transdermal Drug Delivery. Springer, 2000, pp. 377–406.
[5] K. S. Cole, “Permeability and impermeability of cell membranes for ions,” in Cold Spring Harbor Symposia on Quantitative Biology, vol. 8. Cold Spring Harbor Laboratory Press, 1940, pp. 110–122.
[6] S. Grimes and Ø. G. Martinsen, “Biompedance,” Wiley encyclopedia of biomedical engineering, 2006.
[7] E. T. McAdams and J. Josseret, “Problems in equivalent circuit modelling of the electrical properties of biological tissues,” in Bioelectrochemistry and Bioenergetics, vol. 40, no. 2. Elsevier Science S.A., 1996, pp. 147–152.
[8] A. Miller, B. Blott et al., “Review of neural network applications in medical imaging and signal processing,” Medical and Biological Engineering and Computing, vol. 30, no. 5, pp. 449–464, 1992.
[9] S. Chen, S. Billings, and P. Grant, “Non-linear system identification using neural networks,” International journal of control, vol. 51, no. 6, pp. 1191–1214, 1990.
[10] M. Scholz, “Approaches to analyse and interpret biological profile data,” 2006.
[11] I. A. Lee and M. Verleysen, Nonlinear dimensionality reduction. Springer Science & Business Media, 2007.
[12] G. Ivosev, L. Burton, and R. Bonner, “Dimensionality reduction and visualization in principal component analysis,” Analytical chemistry, vol. 80, no. 13, pp. 4933–4944, 2008.
[13] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” science, vol. 290, no. 5500, pp. 2323–2326, 2000.
[14] H. Abdi, “Metric multidimensional scaling (mds): analyzing distance matrices,” Encyclopedia of measurement and statistics, pp. 1–13, 2007.
[15] A. Saxena, A. Gupta, and A. Mukerjee, “Non-linear dimensionality reduction by locally linear isomaps,” in International Conference on Neural Information Processing. Springer, 2004, pp. 1038–1043.
[16] P. Campadelli, E. Casiraghi, C. Ceruti, and A. Rozza, “Intrinsic dimension estimation: Relevant techniques and a benchmark framework,” Mathematical Problems in Engineering, vol. 2015, 2015.
[17] R. Cavalieri and P. Bertemes-Filho, “Plant tissue differentiation using electrical impedance spectroscopy with deep neural networks,” International Journal of Biosensors & Bioelectronics, vol. 6, no. 1, 2020. [Online]. Available: https://doi.org/10.15406/ijbse.2020.06.00182
[18] T. pandas development team, “pandas-dev/pandas: Pandas,” Feb. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3509134
[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[20] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. D. J. Dongarra, J. D. Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, LAPACK Users’ Guide, 3rd ed. Philadelphia, Pennsylvania, USA: SIAM, 1999.

[21] R. W. Floyd, “Algorithm 97: Shortest path,” Commun. ACM, vol. 5, no. 6, p. 345, Jun. 1962. [Online]. Available: https://doi.org/10.1145/367766.368168