Anais do V Congresso Brasileiro de Eletromiografia e Cinesiologia e X Simpósio de Engenharia Biomédica - ISBN: 978-85-5722-065-2 - DOI: 10.29327/cobecseb.78931

PRINCIPAL COMPONENT ANALYSIS USAGE IN BIOMEDICAL ENGINEERING TO AID AT DIAGNOSING PATHOLOGIES

E.F. Esmanhoto*, W.L. Ripka* and L. Ulbricht*

* Federal University of Technology – Paraná, Curitiba, Brazil
e-mail: eduardo_esmanhoto@hotmail.com

Abstract – This paper aims to present a review of the many uses of the statistical method of Principal Component Analysis (PCA) in the Engineering Biomedical field, aimed specially in those where PCA was used as a tool to diagnose pathologies in the last 5 years. An exploratory study was made through the use of bibliometrics, narrowing down the initial search to a final portfolio of 26 papers, providing the latest and state-of-the-art researches on the desired field of study. It was found that PCA has been used in a wide spectrum of areas with significant results, and all around the world. The main use is to reduce the dimensionality of the data to a few principal variables which can explain most of the variance present in the original data. There were studies which reduced from 14 to 50 variables into 1 to 6 principal components, while retaining in average 80% of the variance, and others reduced from 51 to 140 variables into as low as 2 components, keeping 68% to 99% of the variance.
Keywords – PCA, Principal Component Analysis, Biomedical Engineering, Diagnose, Pathology.

Introduction

While Biomedical Engineering might be defined as an interdisciplinary branch of engineering, founded both in engineering and in the life sciences, when biomedical engineers work directed to a hospital or to a clinic, they are more properly called clinical engineers [1]. Some of Clinical Engineering applications regard equipment testing, assessing of their nature, accessibility and characteristics, aiding health professionals at investigating and diagnosing pathologies. In this sense, a variety of statistical analyses may be employed to further aid in such goal. Amongst these analyses there is the principal component analysis (PCA).

A principal component analysis is mainly concerned with explaining the variance of a set of variables through a few linear combinations of these variables [2]. Its general objectives are data reduction and interpretation. Simply stated, out of a set of p correlated variables, a new set of $q \leq p$ uncorrelated variables are provided, which are nothing but linear combinations of the firsts, while retaining great part of the variance of the data.

The use of PCA in biomedical engineering research is very wide. PCA is a multivariate statistical technique used for dimensionality reduction in many biomedical and clinical measurements by retaining those characteristics of the data that contribute the most to its variance [3]. In addition to that, PCA is used to identify the movement characteristics of various groups under various conditions [4].

Bearing that in mind, the main goal of this study is to evaluate and quantify the use of PCA in the biomedical engineering field, when diagnosing pathologies. How has PCA been employed in the past 5 years in state-of-the-art researches, directed to clinical and biomedical engineers on diagnosing pathologies? This study also seeks to understand how broadly it has been used in different areas, how well it fulfills its purposes, what applications for this analysis have been proposed, and, by answering these questions, bring to light to new possibilities.

Methods

In pursuing the most recent and valuable uses of the PCA in the biomedical area, there arose the need to carry out a bibliometric review of the literature on the theme, by gathering and organizing information from different articles, areas and the main uses of this statistical analysis. Bibliometrics can be defined as the application of mathematical or statistical methods on a group of bibliographic references [5]. The term bibliometrics, moreover, is used to quantify processes of written communication. Also, bibliometric indicators are employed to measure scientific production [6].

Bibliometrics, as an area of study of Information Technology, has an important role in the analyses of the scientific production of a country, since its indicators might portray the behavior and also the development of a field of study [7].

This research is characterized as an exploratory study. Certain steps had to be followed as required by the bibliometric technique, and are described in Figure 1.

The final portfolio comprises the twenty-six most recent studies on to the statistical analysis proposed. Throughout the filtering process, EndNote X6 (Thomas Reuters, USA) and Microsoft Excel 2016 were important tools utilized to organize the portfolio and also keep it clean after each step.

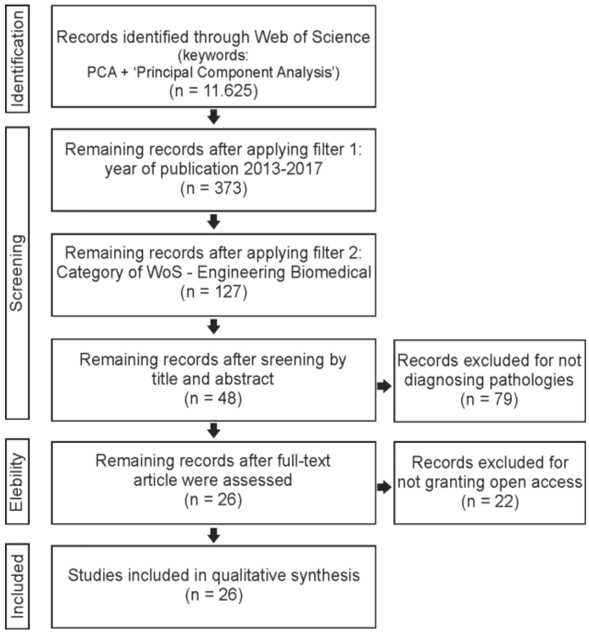

Figure 1: Articles selection to final portfolio.

Results and discussion

An important finding is that the amount of papers using PCA in diagnosing pathologies has been on a slight decrease through the years, as shown in Figure 2. This might indicate PCA's usage is declining, but as the final portfolio consists of only 26 papers, and the year of 2017 was computed only up to April, this may be a hasty conclusion.

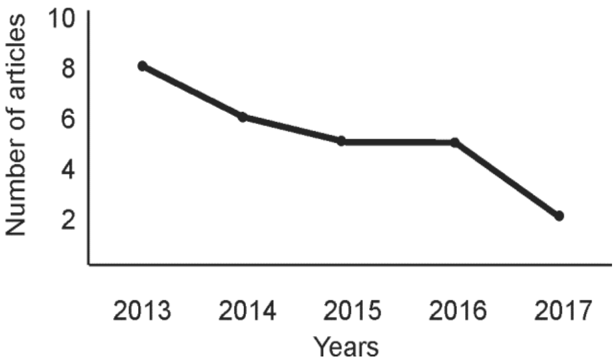

Figure 2: Number of articles published in 2013 – 2017.

Many times, an article refers to more than one area of study. In Figure 3 are presented the number of articles that not only belong to the field of Biomedical Engineer, but also take part in other areas. While only 4 papers were classified as only belonging to Biomedical Engineering, 22 papers crossed fields with as many as 11 different fields of study, such as Biophysics and Computer Science. It is highlighted, thus, how versatile PCA is.

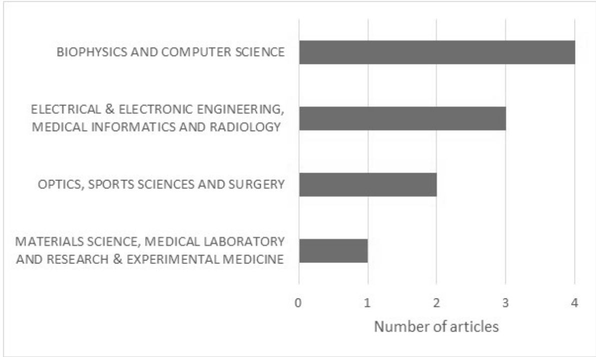

Figure 3: Number of articles per field of study.

Furthermore, Figure 4 shows to what biological systems the pathologies which these papers aimed in diagnosing belong. As one might see, the studies cover a wide spectrum of biological systems, yet again confirming the versatility of the PCA method.

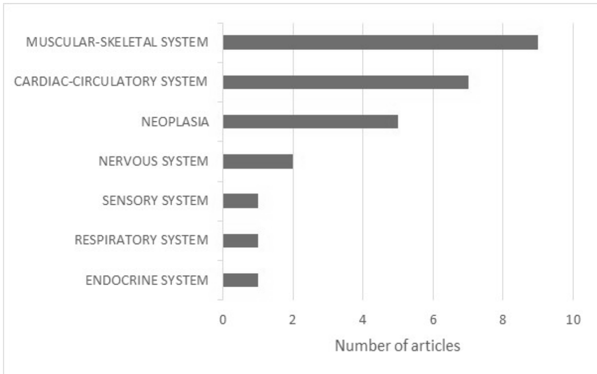

Figure 4: Number of articles per biological system.

In Figure 5 are shown the number of articles published per country, topped by Canada with 4 articles. The top 5 countries in number of articles are from 3 different continents, which may indicate that the usage of PCA seems to be quite spread rather than localized. There is also one Brazilian paper which used PCA to aid diagnosing sustained monomorphic ventricular tachycardia [8]. This data was taken directly from Web of Science analysis, which takes into account the fact that among these are some studies published in more than one country.

Anais do V Congresso Brasileiro de Eletromiografia e Cinesiologia e X Simpósio de Engenharia Biomédica - ISBN: 978-85-5722-065-2 - DOI: 10.29327/cobecseb.78931

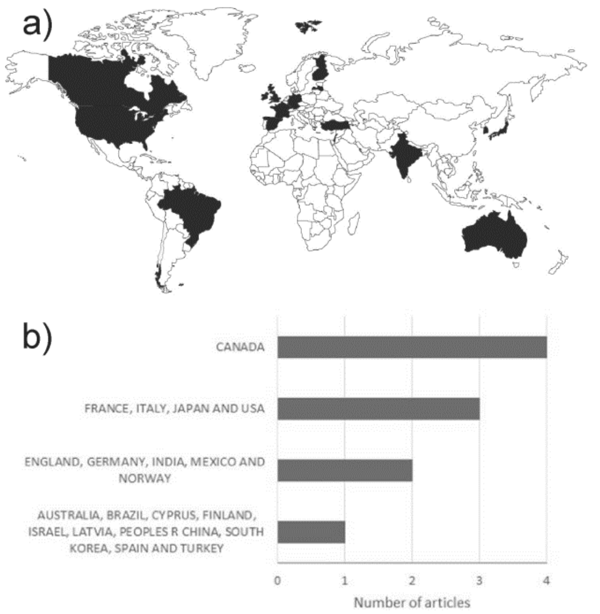

Figure 5: Distribution of articles around the world (a) and number of articles per country (b).

The most common objective for using PCA is that it reduces significantly the dimension of the data, maintaining a great portion of the variance explained. In order to evaluate the achievement of this goal in different studies, comparisons were made between the number of original variables, the number of principal components (PCs) selected, and how much of the total variance was explained by these PCs, presented in Table 1.

Although this data is highly relevant, some of the articles did not present it, making it more difficult not only to fully analyze the studies, but also to reproduce the results. Only 10 papers provided all this vital information. Out of these 10, the studies which had up to 50 initial variables reduced data dimension to only 1 to 6 principal components, while retaining in average 80% of data variance. In addition, the studies with 51 to 140 original variables showed reductions to as low as 2 components, and retained from 68% to 99% of the variance.

Apart from this most common usage for the PCA, that is, having few PCs to explain the maximum variance of the original data, 3 articles attract attention for proposing different approaches. In the paper written by Federolf, Boyer and Andriacchi [18], a linear combination with 3 specific principal components (namely PC1, PC8 and PC9, which accounted for 26,2% of the total variability) was the best vector to differentiate the two groups proposed by the study, healthy and knee-osteoarthritic gait. In the study of Kobayashi et al. [4], 23 PCs that represent 91% of the original 1818 variables are used, and it selects only one component, PC5, which was related to the risk of falling. At last, Zakeri et al. [19] propose the number of principal components, NPCs, for a given explained variance, as a quantity to indicate similarity between Seismocardiogram morphological changes. Once again it should be highlighted how versatile PCA presents itself.

Table 1: Original variables, principal components and explained variance per study.

Study	Area	Original variables	PCs	Explained Variance
Araki, T. [9]	Computer Science, Medical Informatics	56	14	99%
Brown, N., et al. [10]	Sport Sciences	14	1	57%
Hatfield, G. L. [11]	Sport Sciences	101	2	79,60%
Mezghani, N., et al. [12]	Biophysics	100	2	94%
Nasario, O. [8]	Medical Laboratory Technology	140	7	98,30%
Nazlibilek, S., et al. [13]	Research & Experimental Medicine	120	81	95%
Rachim, V. P., et al. [14]	Materials Science	15	5	95%
Schloemer, S. A. [15]	Biomedical Engineering	33	1 a 6	90%
Siuly, S. [16]	Computer Science, Medical Informatics	100	37	68,22%
Tarvainen, M. P., et al. [17]	Electrical & Electronic Engineering	32	2	75%

Conclusions

Principal Component Analysis presents itself as a promising tool in various fields of study. Studies were found in a number of areas, such as Biophysics, Computer Sciences and Radiology, aiming at pathologies from many different biological systems, such as Muscular-Skeletal and Cardio-Respiratory systems. Furthermore, its employment has been observed across the globe. Regarding the main reason for using PCA, which is reducing data variability, the studies which had from 14 to 50 original variables reduced data dimension to only 1 to 6 principal components, while retaining on average 80% of data variance. Also, studies with 51 to 140 original variables showed reductions to as low as 2 components, retaining as much as 99% of the variance.

References

[1] Bronzino J D Management of medical technology, a primal for clinical engineers (1992) Butterworth-Heinemann, Stoneham.

[2] Johnson R A, Wichern D W (2007) Applied multivariate statistical analysis Applied multivariate statistical analysis Springer Berlin Heidelberg, Berlin.

[3] De Panfilis S, Moroni C, Peccianti M, Chiru OM, Vashkevich V, et al. (2013) Multi-point accelerometric detection and principal component analysis of heart sounds. Physiological Measurement 34: L1-L9.

[4] Kobayashi Y, Hobara H, Matsushita S, Mochimaru M (2014) Key joint kinematic characteristics of the gait of fallers identified by principal component analysis. Journal of Biomechanics 47: 2424-2429.

[5] Rostaing H (1997) La Bibliométrie Et Est Techniques. Marseille: Centre de Recherche Rétrospective de Marseille: Tolouse: Sciences de la Société.

[6] Reveles AG, Takahashi RT (2007) Educação em saúde ao ostomizado: um estudo bibliométrico Rev Esc Enferm USP 41: 245-50.

[7] Araújo RF (2011) A bibliometria na pesquisa científica da pós-graduação brasileira de 1987 a 2007 Enc. Bibli: R. Eletr. Bibliotecon16: 51-70

[8] Nasario O, Benchimol-Barbosa PR, Nadal J (2014) Principal component analysis in high resolution electrocardiogram for risk stratification of sustained monomorphic ventricular tachycardia. Biomedical Signal Processing and Control 10: 275-280.

[9] Araki T, Ikeda N, Shukla D, Jain PK, Londhe ND, et al. (2016) PCA-based polling strategy in machine learning framework for coronary artery disease risk assessment in intravascular ultrasound: A link between carotid and coronary grayscale plaque morphology. Computer Methods and Programs in Biomedicine 128: 137-158.

[10] Brown N, Bichler S, Fiedler M, Alt W (2016) Fatigue detection in strength training using three-dimensional accelerometry and principal component analysis. Sports Biomechanics 15:139-150.

[11] Hatfield GL, Stanish WD, Hubley-Kozey CL (2015) Relationship between knee adduction moment patterns extracted using principal component analysis and discrete measures with different amplitude normalizations: Implications for knee osteoarthritis progression studies. Clinical Biomechanics 30:1146-1152.

[12] Mezghani N, Fuentes A, Gaudreault N, Mitiche A, Aissaoui R, et al. (2013) Identification Of Knee Frontal Plane Kinematic Patterns In Normal Gait By Principal Component Analysis. Journal of Mechanics in Medicine and Biology 13: 13.

[13] Nazlibilek S, Karacor D, Erturk KL, Sengul G, Ercan T, et al. (2015) White Blood Cells Classifications by SURF Image Matching, PCA and Dendrogram. Biomedical Research-India 26: 633-640.

[14] Rachim VP, Li G, Chung WY (2014) Sleep apnea classification using ECG-signal wavelet-PCA features. Bio-Medical Materials and Engineering 24: 2875-2882.

[15] Schloemer SA, Thompson JA, Silder A, Thelen DG, Siston RA (2017) Age-Related Differences in Gait Kinematics, Kinetics, and Muscle Function: A Principal Component Analysis. Annals of Biomedical Engineering 45: 695-710.

[16] Siuly S, Li Y (2015) Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification. Computer Methods and Programs in Biomedicine 119: 29-42.

[17] Tarvainen MP, Cornforth DJ, Jelinek HF, Ieee (2014) Principal Component Analysis of Heart Rate Variability Data in assessing Cardiac Autonomic Neuropathy. 2014 36th Annual International Conference of the Ieee Engineering in Medicine and Biology Society. New York: Ieee. pp. 6667-6670.

[18] Federolf PA, Boyer KA, Andriacchi TP (2013) Application of principal component analysis in clinical gait research: Identification of systematic differences between healthy and medial knee-osteoarthritic gait. Journal of Biomechanics 46: 2173-2178.

[19] Zakeri V, Tavakolian K, Arzanpour S, Zanetti JM, Dumont GA, et al. (2014) Preliminary Results on Quantification of Seismocardiogram Morphological Changes, Using Principal Component Analysis. 2014 36th Annual International Conference of the Ieee Engineering in Medicine and Biology Society. New York: Ieee. pp. 6096-6099.