Journal of Universal Computer Science, vol. 26, no. 6 (2020), 734-746
submitted: 22/11/19, accepted: 20/4/20, appeared: 28/6/20 CC BY-ND 4.0

The Modified *Principal Component Analysis* Feature Extraction Method for the Task of Diagnosing Chronic Lymphocytic Leukemia Type B-CLL

Mariusz Topolski
(Department of Systems and Computer Networks
Faculty of Electronics, Wrocław University of Science and Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
mariusz.topolski@pwr.edu.pl)

Abstract: The vast majority of medical problems are characterised by the relatively high spatial dimensionality of the task, which becomes problematic for many classic pattern recognition algorithms due to the well-known phenomenon of the curse of dimensionality. This creates the need to develop methods of space reduction, divided into strategies for the selection and extraction of features. The most commonly used tool of the second group is the PCA, which, unlike selection methods, does not select a subset of the original set of features and performs its mathematical transformation into a less dimensional form. However, natural downside of this algorithm is the fact that class context is not present in supervised learning tasks. This work proposes a feature extraction algorithm using the approach of the PCA method, trying not only to reduce the feature space, but also trying to separate the class distributions in the available learning set. The problematic issue of the work was the creation of a method of feature extraction describing the prognosis for a chronic lymphocytic leukemia type B-CLL, which will be at least as good, or even better than when compared to other quality extractions. The purpose of the research was accomplished for binary and three-class cases in the event in which for verification of extraction quality, five algorithms of machine learning were applied. The obtained results were compared with the application of paired samples Wilcoxon test.

Key Words: *Principal Components Analysis*, data classification, recognition of patterns, lymphocytic leukemia type B-CLL.
Category: H.2, H.3.7, H.5.4

1 Introduction

The contemporary economy and relentless technological progress demands the continuous development of computer-aided task automatization. In the field of information technology, methods of machine learning play a very important role. Pattern recognition methods are applied to the problems of objects and images classification and to a variety of tasks related to computer networks. One of the many fields in which classification methods are applied is medicine. Frequently, we encounter a situation where a learning set has a huge number of features. The increase of data dimensionality does not improve classification accuracy, but it may lead to its deterioration.

It led to the situation, where fast development of both selection and extraction methods began. A main reason why increasing problem dimensionality does not improve classification is the so called *curse of dimensionality*. The notion was formulated by Richard Bellman [Bellman 2015]. The curse of dimensionality relates to a situation in which the correct classification of objects in the training set is almost impossible. The size of characteristics in a vector leads to a higher number of parameters, which in turn increases the complexity of a classifier. As a result of increasing feature dimensionality, the chance of *overfitting* increases and a decline of *generalization* values of a given classifier occurs. Increase of features number in a training set is proportional to the number of patterns necessary to build a convergent generalisation model. In the case of neural networks, a number of features needed for correct generalisation grow exponentially [Jimenez and Landgrebe 1999].

The task of dimensionality reduction is carried out by an extraction and selection of features. The task of feature selection is performed by selecting a subset or subsets of features, while extraction is used for i.e. linear combination of the given feature space. The second method is the main subject of the research conducted in the following paper.

2 Methods

2.1 Methods of feature extraction

Among various methods of data extraction, the most popular is the Principal Component Analysis (PCA). It is a projection method which focuses on mapping an original feature space with d dimensions and its transformation to a new $(s <= d)$ space. The purpose is to minimise a loss of information in a resulting projection [Alpaydin 2009].

The PCA method is an example of unsupervised learning methods, which do not require data labelling. Its main assumption is the feature-group separation, aiming to reduce mutual correlation and sorting according to a declining eigenvalue, and thus a declining variance. Eigenvectors are called principal components. Due to the various feature domains, they are subject to a standard normalisation at the beginning. The abovementioned eigenvectors create a matrix of co-variance in the task of extraction.

Let λ_i be an eigenvalue p of this eigenvector, then the proportion of the total variance originating from r of first vectors, may be calculated by the equation:

$$\lambda = \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{i=1}^{r} \lambda_i + \sum_{r+1}^{p} \lambda_{r+1}} \tag{1}$$

There exists a strong connection between the correlation of initial variables and eigenvalue. The stronger the correlation, the bigger the eigenvectors, therefore the number of features that may be reduced increases. In a situation in which the input variables are weakly correlated, r will be similar to p and a loss of initial variance of a set of features increases [Alpaydin 2009]. A reduction of dimensions may be maximally equal to a number of samples [Ringnér 2008].

The PCA method, in order to separate principal components, uses linear transformation which represents a serious difficulty in a situation of nonlinear feature space. One may say that extraction from complex nonlinear feature spaces performed by the PCA method becomes impossible. The KPCA (*Kernel Principal Component Analysis*) deals with this problem; its essence being the application of various nonlinear functions of the kernel trick.

In the event of a nonlinear problem, the KPCA method does not require familiarity with transformation ϕ (but one is limited to a selection of an appropriate kernel function) [Schölkopf 2001].

The KPCA method was used in order to detect various defects in the monitoring of processes. It is the same as the PCA, but with a kernel performing nonlinear transformations. Regarding the problem of space of linear and nonlinear features, a genetic algorithm selecting features for further extraction was applied [Jiang and Yan 2018].

It is worth paying attention to another interesting modification of PCA, the FPCA. It uses the labels of classes in Karhunen-Loèvin theorem. This theorem is utilised in the PCA mainly in the transformation of images and data analysis. It assumes that a certain stochastic process is an infinite linear combination of orthogonal functions as similar to a series representation of Fourier's function [Mao 2005].

PCA is applied in medicine in the task of preparation of space of features for recognising various diseases. One such example is recognising cervical cancer, where in the task of classification, a Random Forest model was applied. The results obtained in 2018 were very promising [Abdoh et al 2018].

Teams from Poland also work on the application of the PCA in medicine. In this case, two linear methods PCA and LDA were applied in the task of recognising lung cancer based on microscopic photographs of tissues. Such a combination of PCA and LDA was also applied in the task of induction-emission matrices of selected groups of biological substances [Leskiewicz et al 2016].

The PCA method is becoming more and more popular. It is also used in detecting networks attacks [Xiao et al 2019]. It was also used in order to reduce dimensions before entering the data to the CNN (*Convolutional Neural Networks*). In the research, the results were compared with the results obtained by a reduction of features applied by the *Autoencoder* networks. The obtained results were better in the event of the PCA method application.

Due to great popularity and interest in the PCA method, various modifications have occurred. They aim at looking for new solutions in the extraction of features, to cope better with nonlinear complexity. One of these methods is

SuperPCA [Jiang et al 2018]. It is used in the classification problem related to *hyperspectral imagining* [Krawczyk et al 2014]. It was applied in the problem of classification based on hyperspectral imagery. This method is a merger of PCA and the segmentation algorithm. The applied method of super-pixelization. Then, for each area the PCA is applied. The advantage of this approach is retaining the spatial features of the image, which allows for improvement in classification quality.

In the task of tracing dynamics of data changes in i.e. data stream, the DiPCA (*Dynamic inner PCA*) method is applied, which is used for monitoring the course of the process, where its main purpose is to analyse the dynamically changing feature values [Dong and Qin 2018]. The assumption of this method is maximising co-variance between the presently indicated components, and the components indicated in previous patterns. The method separates a model of dynamic hidden variables. It is the hidden variables which are a set of input data for the PCA.

In the above section one of the most important unsupervised methods of extraction of features is presented [Zyblewski and Woźniak 2020]. In the event of supervised methods one can mention LDA. It is a linear discrimination analysis applicable in medical problem that is in the task of cancer recognition. For this purpose, a microscopic photograph of tissue was used [Kaznowska et al 2018]. Input data which was used here originated from computer tomography photographs. The reduced space of features was used in the task of classifying lung cancer with the application of (*ang. Optimal Deep Neural Network*). Application of the LDA contributed to a quality increase in the task of classification. Application of the LDA contributed to a quality increase in the task of classification. LDA, similar to PCA, got modified. One such method is RSLDA (*Robust Sparse Linear Discriminant Analysis*) [Wen et al 2018].

This method was proposed due to certain limitations attached to standard LDA. RSLDA solves problems related to the interpretability of space of features, selection of an appropriate number of dimensions and sensitivity to outliers and extreme values of features. LDA modification consists in taking into consideration maximisation after the variance which was applied in PCA. In assessing certainty of features, the standard $\ell_{2,1}$ of the searched transformation matrix was used. The objective function also contains a parameter of error, which is minimised in the process of extraction.

The article presents a modified method of PCA analysis which is based on a rotation of factors against centroids of classes. It is an extension of the one proposed for a lymphocytic leukemia [Topolski and Topolska 2019].

2.2 CCPCA model of features extraction

The *Centroid Class Principal Component Analysis* (CCPCA) model is the author's modification of the *Principal Components Analysis* (PCA). The main assumption of the concept is the rotation of a normalised feature space of the objects, according to classes' centroids. The main goal of such an approach is the separation of such subspaces of features which will be oriented at classes. A set of observations X similarly as in the event of PCA is written with use of the matrix:

$$X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,b} \\ \vdots & \ddots & \vdots \\ x_{a,1} & \cdots & x_{a,b} \end{bmatrix}, \tag{2}$$

where: $x_{a,b}$ are consecutive patterns and b^{th} attribute of the dataset.

The subsequent step sets kernel K, for all elements $x_{a,b}$ which are a scalar product of the transformations $\phi(x_a)$ and $\phi(x_b)$, being the elements of the new, mapped space. In order to calculate the new feature space, we need to choose a proper transformation kernel:

$$K = \begin{bmatrix} k_{1,1} & \cdots & k_{1,b} \\ \vdots & \ddots & \vdots \\ k_{a,1} & \cdots & k_{a,b} \end{bmatrix}, \tag{3}$$

where: $k_{ab} = k(x_a, x_b)$.

The transformation kernel [Schölkopf 2001] may be defined through various functions. For the PCA method it will be a linear transformation $k(x_a, x_b) = (x_a x_b)$. The essence of the CCPCA, the fact that the shift $\phi(x_a)$ is carried out in the direction of the coordinates C - centroids of classes j.

$$C(c_1, c_2, ..., c_j) = \sum_{i=1}^{N} \min_{m=1,\dots,j} d(x_a, c_b), \tag{4}$$

where: j is the number of classes.

The finally normalised observations shifted to centroids of classes are set through:

$$\forall_j \quad \tilde{\phi}_j(x_a) = \phi(x_b) - \phi_j, \tag{5}$$

where: $\tilde{\phi}(x_a)$ is the projection of observation x_a in the transformed space, including the rotation axis shift in the direction of class-centroids, ϕ_j is the centre of the set of each of the classes.

$$\forall_{\phi_j(x_a)} cos_j^2(\alpha) = cos_j^2(\alpha) = diag(A) + sin^2(\alpha) \oplus sin_j^2(\alpha), \tag{6}$$

where: $diag(A)$ is a diagonal matrix and α is the rotation angle.

As a result of the rotation (Equation 6) we obtain for each of the classes j new matrices (Equation 3) $\tilde{K}$ which include scalar products shifted to the centroids (Equation 4) in the coordinate system in the spaces of features.

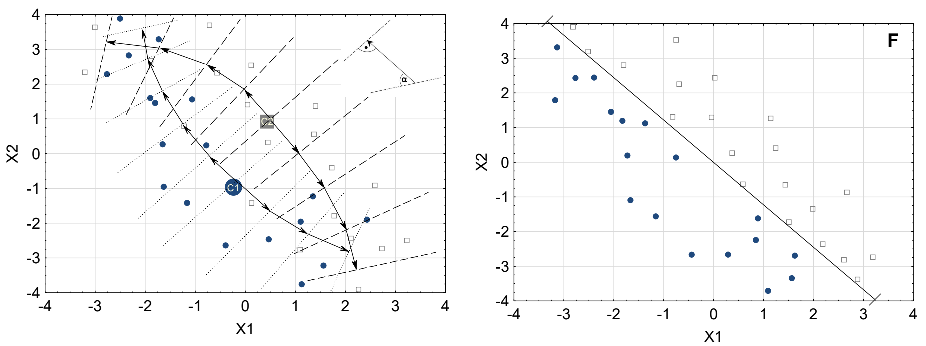
According to the PCA method we set eigenvectors v_i of the matrix $\tilde{K}$. Thus, the matrix of correlation is created as the product of $alpha$ and $\phi_j(x_a)$ (Equation 5):

$$v_i = \sum_{i=1}^{a} \alpha_i \phi_j (x_i) \tag{7}$$

Finally, groups in the matrix of eigenvectors V (Equation 7) allow us to set coordinates for any point x_a for the principal components in the feature space:

$$x_a^{CCPCA} = V \tilde{\phi_j} (x_a) \tag{8}$$

Figure 1 presents an example of CCPCA operation where C_1 and C_2 are centroids of two classes for dimensions $\phi(s_1)$ and $\phi(s_2)$. The rotation angle results from the Equation 6. An essential contribution of this method is that for each principal component – separated this way – observations may be assigned to the most strongly discriminating ones.



(a) The process of rotation and separation of principal components.

(b) Map of two principal components after separation and rotation.

Figure 1: Plot showing operation of the CCPCA method.

3 Experimental set-up

The purpose of the research was to develop a method for the extraction of features, including classes, which may contribute to improvement in quality of prognosis of lymphocytic leukemia type B-CLL. The problem examined by the research is very important as correct assessing the prognosis for leukemia may contribute to saving many human lives. As indicated in the previous chapter, many algorithms of extraction exist, however they are mainly unsupervised methods. There is a lack of studies which allow the performance of extraction including classes. It is particularly important in a situation when the type of decision depends on the various spaces of features and/or patterns. Appropriate selection of object patterns constitutes a primary database for classifiers here. In the clinical tests, in order to diagnose lymphocytic leukemia type B-CLL, the occurrence of monoclonal population of lymphocytes B in a patient's peripheral blood is determined. Even then first results of the blood test indicating the values of the lymphocytes population $>5000/\mu l$ and cardinality of cancer cells in the number >30 percent make it possible to determine the risk of the occurrence of leukemia [Kay et al 2002][Dmoszynska and Robak 2008][Hallek et al 1996].

Assessment of a prognosis of patients with B-CLL is very difficult. It is often related to the fact that we do not observe any symptoms of the disease. A correct diagnosis regarding a future patient, facilitates implementation of appropriate treatment methods, and thus increases the chance for survival [Monserrat et al 2003][Hamblin 2003][Rai and Chiorazzi 2003].

Selection of the samples for the research is not accidental. It is the response to a medical problem, which is related to a prognosis of the further growth of the disease. Research aimed at determination of these features, which will enable the improvement of accuracy in the classification of the patients with B-CLL, is ongoing . The most frequently indicated variables are: an indicator of the number of lymphoid cells in peripheral blood, a level of bone marrow destruction or time in which doubling of lymphocytes occurs [Brugiatelli et al 2003].

A set of data used in experiments covers 239 patterns with the diagnosis of an oncologist (doctor). People with B-CLL were subjected to homogeneous treatment by chemical therapy. During the research a set consisting of 40 features are taken into consideration in the prognosis of the state of patients' health: leucocytes, erythrocytes, haemoglobin, haematocrit, MCV, MCH, MCHC, platelets, lymphocytes, others (Eo, Bazo, Mono), Neutrophils, RDW-CV, CWKS, PDW, MPV, P-LCR, neutrophil granulocytes with rod-like kernel (nucleus), granulocytes with segmented kernel (nucleus), eosinophilic granulocytes, basophil granulocytes, CRP, 10 parameters of testing bone marrow, timodine infiltration, infiltration of bone marrow in cytologic evaluation, percentage of pro-lymphocytes in peripheral blood, concentration of 2-micro-globuline, time of doubling a number of lymphocytes, expression ZAP70 in leukemic cells, expression of CD38 in the leukemic cells and expression of CD23 in the leukemic cells. The patients were grouped into three classes $j =$ [0-good prognosis; 1-bad prognosis; 2 - very bad prognosis]. The number of classes is 114, 70 and 55 accordingly. In the experiments four methods of extraction were applied (PCA, KPCA, [Schölkopf 2001], LDA and CCPCA) and five base classifiers (k-NN, SVM, MLP, CART and FNB).

4 Experimental evaluation

Two experiments were performed for binary and three-class cases accordingly. Due to the influence of data imbalance on classification quality evaluation, *Balanced Accuracy Score* metric was applied. In each experiment the method of five-fold stratified cross-validation was applied. Each time in a learning set an extraction of variables was performed by four methods: PCA, KPCA, IDA and CCPCA.

4.1 Experiment 1 – Indicating quality of BAC classification for two classes

The purpose of the Experiment 1 is the comparison of four methods of the extraction of features, with the application of five algorithms of machine learning for classification. The binary case was assumed in this task. The obtained results were compared with the application of a non-parametric paired samples Wilcoxon test at the statistical significance $\alpha = 0.05$. The results of the experiment are presented in the Table 1.

Table 1: The results of the experiments for the binary case with application of *BAC-score* metrics. In the columns the extraction methods are presented, where NO means lack of extraction of an object's features.

CLASSIFIER	1 NO	2 PCA	3 KPCA	4 LDA	5 CCPCA
k-NN	0.721 –	0.779 1	0.783 1	0.815 1,2,3	0.831 *all*
SVC	0.733 –	0.784 1	0.787 1	0.824 1,2,3	0.846 *all*
MLP	0.741 –	0.789 1	0.792 1	0.832 1,2,3	0.853 *all*
CART	0.738 –	0.781 1	0.786 1	0.828 1,2,3	0.846 *all*
GNB	0.716 –	0.769 1	0.778 1	0.809 1,2,3	0.827 *all*

Very interesting results were obtained in Experiment 1. The first conclusion that comes to mind is that in the event of all the applied classifiers, comparable tendencies of the results for various methods of extraction of an object's features were obtained. For each algorithm, the statistically valid best quality of correct classifications was obtained after the application of the author's method namely CCPCA. The supervised LDA method allowed for obtaining significantly less correct classifications than CCPCA, but significantly better results than in other unsupervised methods. Interesting results were also obtained after application of the PCA and KPCA methods, where for each algorithm, comparable frequencies of correct classifications were obtained. Lack of differences between the above-mentioned methods may be explained by the fact that multi-dimensional space of features has a nature of hyperplane and application of Gauss' kernel in the case of KPCA, will not result in an increase in the accuracy of extractions. After the application of any method of extraction of an object's features, frequencies of correct classifications of each of the algorithms were better than in the event when the extraction was not performed. Moreover, the experiment's results indicated that the prognosis of B-CLL is most strongly and simultaneously dependent on 9 features: lymphocytes, timodine infiltration of lymphocytes, infiltration of bone marrow in cytologic evaluation, percentage of pro-lymphocytes in peripheral blood, concentration of 2-micro-blobuline, time for doubling the number of lymphocytes, expression AZP70 in leukemic cells, expression CD38 in leukemic cells and expression CD23 in leukemic cells. These 9 features were selected by the CCPCA algorithm. Other methods of extraction finally built spaces consisting of 10 to 15 features.

4.2 Experiment 2 – indicating quality of BAC classification for three classes

Experiment 2 is an extension of Experiment 1 for the three-class case. The purpose of this experiment was checking how the method of extraction of features CCPCA looks when compared to other methods for the case assuming more than two classes. For comparison of the validity of the results, non-parametric paired samples Wilcoxon test was applied. The results of Experiment 2 are presented in the Table 2.

Experiment 2 reveals similar conclusions to experiment 1. In this case, the method of extraction CCPCA also turned out statistically best. Experiment 2 confirmed that in the event of recognising the data we deal with a hyperplane therefore the PCA and KPCA do not differ significantly. In Experiment 2, due to the three-class case, the frequencies of correct classifications of each of five algorithms decreased. The best quality of correct classifications was obtained through the neuron network of multi-layer perceptron (MLP) and the algorithm of the decision tree (CART). In both experiments the supervised methods supervised with regard to the data, in the problem of classification of lymphocytic leukemia type B-CLL, turned out to be more precise. The number of features after extraction of features similarly as in experiment 1 amounted to 9-15. It

Table 2: The results of the experiments for the binary case with application of *BAC-score* metrics. In the columns the extraction methods are presented, where NO means lack of extraction of an object's features.

CLASSIFIER	1 NO	2 PCA	3 KPCA	4 LDA	5 CCPCA
k-NN	0.712 –	0.765 1	0.766 1	0.802 1,2,3	0.819 *all*
SVC	0.723 –	0.767 1	0.778 1	0.816 1,2,3	0.825 *all*
MLP	0.733 –	0.778 1	0.779 1	0.818 1,2,3	0.843 *all*
CART	0.728 –	0.775 1	0.777 1	0.815 1,2,3	0.836 *all*
GNB	0.702 –	0.754 1	0.757 1	0.793 1,2,3	0.814 *all*

may be added, that despite the greater extraction by the CCPCA method, better classifications were obtained by all the algorithms.

5 Conclusions

The purpose of the research was to build a modified version of PCA algorithm. The modification aimed at rotating the space according to the centroids of classes. The proposed method creates such a sub-space F for which a percentage of the explained variance will be maximal. The method was created for the needs of diagnostics of chronic lymphocytic leukemia type B-CLL. In order to verify whether and how various methods of extraction of features contribute to an improvement in classification quality, five algorithms of machine learning i.e. k-NN, SVM, MLP, CART and GNB were applied. Other than the extraction method CCPCA, two unsupervised methods PCA and KPCA and one supervised LDA were compared.

The research was divided into two experiments for binary and three-class cases. The performed experiments delivered interesting conclusions. After applying the extraction of features, correct classifications for each of the algorithms were not significantly higher than in the event of concluding according to all features. In each case the author's method facilitated the obtainment of the best classifications, and the best algorithm of the classification turned out to be the neuron network type multilayer perceptron (MLP). The supervised LDA method built a set of features in a more precise manner than classic PCA and KPCA. It was indicated that the methods PCA and KPCA for the considered problem of diagnostics of B-CLL give statistically comparable solutions of sub-space of features. It is related to the fact that they create a hyperplane. The satisfactory outcome of the study was achieved also by the fact that it was possible to carry out the extraction of 40 test parameters and indicate the 9 of those which are strongest in prognosis regarding a patient's state of health or in the recurrence of a cancer disease. The obtained results shed light some onto the sense of developing this topic of research. The subsequent steps will involve testing the developed CCPCA method in the synthetic sets and high-dimensional problems of various complexity with various number of classes and their complexity.

Acknowledgments

This work was supported by the statutory funds of the Department of Systems and Computer Networks, Wroclaw University of Science and Technology.

References

[Abdoh et al 2018] Abdoh, S.F., Rizka, M.A. and Maghraby, F.A., 2018. Cervical cancer diagnosis using random forest classifier with SMOTE and feature reduction techniques. IEEE Access, 6, pp.59475-59485.

[Alpaydin 2009] Alpaydin, E., 2009. Introduction to machine learning. MIT press.

[Balakrishnama and Ganapathiraju 1998] Balakrishnama, S. and Ganapathiraju, A., 1998. Linear discriminant analysis-a brief tutorial. Institute for Signal and information Processing, 18, pp.1-8.

[Balasubramanian and Schwartz 2002] Balasubramanian, M. and Schwartz, E.L., 2002. The isomap algorithm and topological stability. Science, 295(5552), pp.7-7.

[Bellman 2015] Bellman, R.E., 2015. Adaptive control processes: a guided tour (Vol. 2045). Princeton university press.

[Bootou 2010] Bootou, L., 2010. Large-Scale Machine Learning with Stochastic Gradient Descent, Proceedings of COMPSTAT'2010 pp 177-186

[Brugiatelli et al 2003] Brugiatelli M, Mannina D, Neri S i wsp. Recent update of prognosis and staging of chronic lymphocytic leukemia. Hematol 2003;88 (suppl.10):30-31.

[Buja et al 2008] Buja, A., Swayne, D.F., Littman, M.L., Dean, N., Hofmann, H. and Chen, L., 2008. Data visualization with multidimensional scaling. Journal of Computational and Graphical Statistics, 17(2), pp.444-472.

[DeCoster 1998] DeCoster, J., 1998. Overview of factor analysis.

[Dmoszynska and Robak 2008] Dmoszynska A, Robak T. Podstawy hematologii. Wydawnictwo Czelej, Lublin 2008, wyd.2

[Dong and Qin 2018] Dong, Y. and Qin, S.J., 2018. A novel dynamic PCA algorithm for dynamic data modeling and process monitoring. Journal of Process Control, 67, pp.1-11.

[Duch 2006] Duch, W., 2006. Filter methods. In Feature Extraction (pp. 89-117). Springer, Berlin, Heidelberg.

[Fauvel et al 2006] Fauvel, M., Chanussot, J. and Benediktsson, J.A., 2006, June. Kernel principal component analysis for feature reduction in hyperspectrale images analysis. In Proceedings of the 7th Nordic Signal Processing Symposium-NORSIG 2006 (pp. 238-241). IEEE.

[Fort and Lambert-Lacroix 2014] Fort, G. and Lambert-Lacroix, S., 2004. Classification using partial least squares with penalized logistic regression. Bioinformatics, 21(7), pp.1104-1111.

[Guyon and Elisseeff 2006] Guyon, I. and Elisseeff, A., 2006. An introduction to feature extraction. In Feature extraction (pp. 1-25). Springer, Berlin, Heidelberg.

[Hallek et al 1996] Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G i wsp. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the In-ternational Workshop on Chronic Lymphocytic Leukemia (IW-CLL) updating the Na-tional Cancer Institute-Working Group (NCI-WG) 1996 guidelines. 2008;111:5446-5456.

[Hamblin 2003] Hamblin TJ. CLL, how many diseases? Hematol J 2003;4 (suppl.3):183-186.

[Hyvärinen and Oja 2000] Hyvärinen, A. and Oja, E., 2000. Independent component analysis: algorithms and applications. Neural networks, 13(4-5), pp.411-430.

[Jain and Shandliya 2013] Jain, P.M. and Shandliya, V.K., 2013. A survey paper on comparative study between principal component analysis (PCA) and exploratory factor analysis (EFA). International Journal of Computer Science and Applications, 6(2), pp.373-375.

[Jiang and Yan 2018] Jiang, Q. and Yan, X., 2018. Parallel PCA–KPCA for nonlinear process monitoring. Control Engineering Practice, 80, pp.17-25.

[Jiang et al 2018] Jiang, J., Ma, J., Chen, C., Wang, Z., Cai, Z. and Wang, L., 2018. SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing, 56(8), pp.4581-4593.

[Jimenez and Landgrebe 1999] Jimenez, L.O. and Landgrebe, D.A., 1999. Hyperspectral data analysis and supervised feature reduction via projection pursuit. IEEE Transactions on Geoscience and Remote Sensing, 37(6), pp.2653-2667.

[Kay et al 2002] Kay NE, Hamblin TJ, Jelinek DF i wsp. Chronic Lymphocytic Leukemia. American Society of Hematology. Hematology 2002;193-213.

[Kaznowska et al 2018] Kaznowska, E., Depciuch, J., Łach, K., Kołodziej, M., Koziorowska, A., Vongsvivut, J., Zawlik, I., Cholewa, M. and Cebulski, J., 2018. The classification of lung cancers and their degree of malignancy by FTIR, PCA-LDA analysis, and a physics-based computational model. Talanta, 186, pp.337-345.

[Kline 2014] Kline, P., 2014. An easy guide to factor analysis. Routledge.

[Kołodziej et al 2012] Kołodziej, M., Majkowski, A. and Rak, R.J., 2012. Linear discriminant analysis as EEG features reduction technique for brain-computer interfaces. Przeglad Elektrotechniczny, 88, pp.28-30.

[Krawczyk et al 2014] Krawczyk, B., Ksieniewicz, P., & Woźniak, M. (2014, June). Hyperspectral image analysis based on color channels and ensemble classifier. In International Conference on Hybrid Artificial Intelligence Systems (pp. 274-284). Springer, Cham.

[Kulchandani and Dangarwala 2014] Kulchandani, J. and Dangarwala, K.J., 2014. Blind Source Separation via Independent Component Analysis: Algorithms and Applications. International Journal of Computer Science and Information Technologies, 5(5), pp.6739-674.

[Leskiewicz et al 2016] Leśkiewicz, M. , Kaliszewski, M. , Mierczyk, Z. , Włodarski, M., 2016. Porównanie liniowych metod PCA (Principal Component Analysis) i LDA (Linear Discriminant Analysis) zastosowanych do klasyfikacji matryc wzbudzeniowo-emisyjnych wybranych grup substancji biologicznych, Biuletyn Wojskowej Akademii Technicznej 2016 Vol. 65, nr 1, 15-31

[Makeig et al 1996] Makeig, S., Bell, A.J., Jung, T.P. and Sejnowski, T.J., 1996. Independent component analysis of electroencephalographic data. In Advances in neural information processing systems (pp. 145-151).

[Mao 2005] Mao, K.Z., 2005. Identifying critical variables of principal components for unsupervised feature selection. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 35(2), pp.339-344.

[Monserrat et al 2003] Monserrat E, Gine E, Bosch F. Redefining prognostic elements in chronic lymphocytic leukemia. Hematol J 2003; 4 (suppl. 3): 180-182.

[Mwange et al 2014] Mwangi, B., Tian, T.S. and Soares, J.C., 2014. A review of feature reduction techniques in neuroimaging. Neuroinformatics, 12(2), pp.229-244.

[Rai and Chiorazzi 2003] Rai KR, Chiorazzi N. Determining the clinical course and outcome in chronic lympho-cytic leukemia. N Engl J Med 2003;348:1797-1799.

[Reunanen 2006] Reunanen, J., 2006. Search strategies. In Feature Extraction (pp. 119-136). Springer, Berlin, Heidelberg.

[Ringnér 2008] Ringnér, M., 2008. What is principal component analysis?. Nature biotechnology, 26(3), p.303.

[Saeys et al 2007] Saeys, Y., Inza, I. and Larrañaga, P., 2007. A review of feature selection techniques in bioinformatics. bioinformatics, 23(19), pp.2507-2517.

[Schölkopf 2001] Schölkopf, B., 2001. The kernel trick for distances. In Advances in neural information processing systems (pp. 301-307).

[Tenenbaum and Langford 2000] Tenenbaum, J.B., De Silva, V. and Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. science, 290(5500), pp.2319-2323.

[Topolski and Topolska 2019] Topolski, M. Topolska, K., 2019. Algorithm for Constructing a Classifier Team Using a Modified PCA (Principal Component Analysis) in the Task of Diagnosis of Acute Lymphocytic leukemia Type B-CLL, Hybrid Artificial Intelligent Systems Springer, pp.390-403.

[Ververidis and Kotropoulos 2005] Ververidis, D. and Kotropoulos, C., 2005, September. Sequential forward feature selection with low computational cost. In 2005 13th European Signal Processing Conference (pp. 1-4). IEEE.

[Wen et al 2018] Wen, J., Fang, X., Cui, J., Fei, L., Yan, K., Chen, Y. and Xu, Y., 2018. Robust sparse linear discriminant analysis. IEEE Transactions on Circuits and Systems for Video Technology, 29(2), pp.390-403.

[Xiao et al 2019] Xiao, Y., Xing, C., Zhang, T. and Zhao, Z., 2019. An intrusion detection model based on feature reduction and convolutional neural networks. IEEE Access, 7, pp.42210-42219.

[Zyblewski and Woźniak 2020] Zyblewski P, Woźniak M. Novel clustering-based pruning algorithms. Pattern Analysis and Applications. 2020 Feb 10:1-0.