

Modification of the Principal Component Analysis Method Based on Feature Rotation by Class Centroids

Mariusz Topolski

(Department of Systems and Computer Networks,
Faculty of Information and Communication Technology, Wrocław University of Science and
Technology, Wrocław, Poland

 <https://orcid.org/0000-0002-5213-6845>, mariusz.topolski@pwr.edu.pl)

Marcin Beza

(Department of Systems and Computer Networks,
Faculty of Information and Communication Technology, Wrocław University of Science and
Technology, Wrocław, Poland

234968@student.pwr.edu.pl)

Abstract: Feature engineering is a branch of science that provides tools to support, for example, the preparation of feature spaces for a pattern recognition task. The present work focuses on the problem of feature extraction. The proposed model is based on the mechanisms of PCA principal component analysis. It fills a gap in the implementation of feature extraction by looking for spaces that best discriminate between classes. This was realized by rotating the features according to the centroids of the classes. In addition, a measure of their consistency was determined which allows precise estimation of the number of features for a particular component. Four experiments were conducted in this study. The first two were done on synthetic datasets, while the next two were conducted on ten real datasets. The synthetic data allowed to determine the characteristics depending on the percentage of informative features, the number of input features, the level of imbalance and the number of output components in the extraction task. The obtained results showed that the developed solution allows for a more precise extraction, thus increasing the quality of classification. Moreover, it was shown that the method based on class centroids allows to construct efficient ensembles of classifiers.

Keywords: Principal component analysis, Imbalanced data, Feature extraction

Categories: H.2, H.3.7, H.5.4

DOI: 10.3897/jucs.80743

1 Introduction

In recent years, machine learning has experienced a period of rapid and intensive development. Although the origins of this field of science and technology date back to the 1950s, it was only a few decades later that the computational power of computers became large enough to make it possible to apply its achievements in practice [Çelik 2018].

As information technology continues to grow, so does the demand for artificial intelligence systems that model intelligent behavior while minimizing the need for human intervention [Hamet and Tremblay 2018]. These characteristics of artificial intelligence are increasingly applicable in the area of automatic image reading and classification,

correlation with clinical findings, and use in image processing or reconstruction. Already, in nuclear medicine, artificial intelligence methods are being used in automated early stage disease image recognition and clinical decision making [Seifert et al. 2020]. Their ability to learn from large data sets makes them effective options in machine health monitoring systems (MHMS) [Zhao et al. 2019].

Artificial intelligence methods have a wide range of applications. They are used for automated recognition of early stage disease images and clinical decision making [Seifert et al. 2020], monitoring patients' health status [Zhao et al. 2019], facial recognition [Hazelwood et al. 2018], making predictions on future valuations of shares of listed companies [Sharma et al. 2017], detection and removal of spam in mailboxes [ZhuWei et al. 2017] and on social networking sites such as Twitter [Wu et al. 2017], optimizing the displayed results in search engines [Jain and Gupta], as well as in the development of autonomous cars [Stilgoe 2018], active learning approach to concept drift problem [Kurlej and Wozniak 2011], designing fusers on the basis of discriminants [Wozniak and Zmyslony 2010].

One of the key aspects in the topic of machine learning is the amount of data being processed. Nowadays, huge amounts of information are generated, which come from sources such as health care systems, public administration or social networks [Oussous et al. 2018], [Lensen et al. 2019]. The multidimensionality of this data can be a source of problems, which were referred to as 'The Curse of Dimensionality' by Richard Bellman [Bellman 1957]. The complexity of the problem increases as the number of variables (or dimensions) increases [Kuo and Sloan 2005], [Charles et al. 2019], [Han et al. 2017], [Crochepierre et al. 2020], [Cyganek et al. 2016]. In the context of machine learning, this enforces the use of feature selection and extraction methods to improve the ability of classifiers to generalize, while avoiding overfitting the learning data [Joachims 1998].

1.1 Feature selection and extraction

As technology advances, the need for newer and more efficient data processing methods is increasing. The techniques used to analyze original features and then extract and construct new features that are most relevant to machine learning algorithms are referred to as feature engineering. Transforming the feature space, for a particular problem, can improve its performance. This is a crucial but time-consuming process that requires expert knowledge in the field of data science [Khurana et al. 2016]. Using different arithmetic operations, it is possible to extract new features that can be better correlated with the target classes, which can make it easier to train the model [Nargesian et al. 2017].

In order to perform data classification (grouping, assigning specific classes), it is necessary to extract features specific to a particular set of samples. Careful engineering allows us to develop methods that will transform raw data, e.g., brightness levels of image pixels, into an appropriate representation or specific feature vector. In this way, the classifier will be able to find and extract patterns that will be used for further classification [Liang et al. 2018]. Due to the fact that some of the extracted features may be redundant and may result in lower quality classification, it may be necessary to apply one of the feature engineering methods which is feature selection. It is the process of obtaining a smaller subset from the original set of features, based on specific criteria [Lie and Jiawei 2017]. This removes irrelevant features that will not improve the quality of the classification but may significantly degrade it. By doing this, we can simplify the

input data and get rid of redundant information, which will result in faster and easier interpretation.

It will also reduce the learning time and reduce overfitting to the learning data. It is important to remember that selection should not reduce the quality of classification. In addition, classification is generally more accurate when the data pattern is simplified to include only the most important information [Cvetkovic et al. 2008]. There are different types of selection methods: filters (finding a measure that ranks features), wrappers (a subset of features is made using a specific classification model), and embedded methods (the selection method is built into the learning process of the classifier) [Chmielnicki 2012].

1.2 Motivation

There is a gap in the literature on feature extraction methods. It is due to the fact that there are many unsupervised extraction methods. By extracting features a reduction of their dimensionality is performed, by creating some subspaces. These are generalized to the case of whole sets. Therefore, there is a need to search for more precise extraction solutions allowing for better matching of feature subspaces, which will be better discriminating classes. Hence, the idea of using a modified principal component method, concentrating on the rotation of features, according to centroids of classes appeared. With this approach, it is possible to obtain sets of features that are more important in making certain decisions, i.e., class recognition. Therefore, the main objective set in the research is to create a modification of the PCA method by rotating the features according to the centroids of the classes. In this way, we answer the question "Will the construction of subspaces of features according to class centroids help to increase the quality of classification compared to other extraction methods, especially the PCA method?".

1.3 Main contributions

In this paper, a new feature extraction method is proposed whose main mechanism is based on principal component analysis. In order to better fit the feature space for the classification task, we focused on feature rotation according to class centroids. This approach is not known in the literature. The advantage of the developed method is to minimize the variance within classes, while maximizing it relative to the feature space in other classes. Moreover, a method for estimating the number of features and components have been proposed. Although various criteria for such estimation are known but e.g. selection of features in terms of eigenvalue may lead to too large or too small number of features. The main components of the paper are:

1. Mathematical description of the developed feature extraction method based on rotation by class centroids.
2. Mathematical description of the method for determining the number of features and components.
3. Conducting four experiments including:
 - (a) synthetic data for different number of components and informative features for balanced data for binary classification,
 - (b) synthetic data for the best variants of informative features according to the number of features in the input and the percentage rate of imbalance.

- (c) ten real-world sets of features with different numbers of features, patterns, and classes.
 - (d) ten real-world sets of features with different numbers of features, patterns, and classes.
4. ten real-world sets of features with different numbers of features, patterns, and classes.

2 Related Work

One of the most popular feature extraction algorithm is the principal component analysis (PCA). It allows to reducing the dimensions of a given feature space, in such a way that the sum of variances (feature variability) is the largest. The resulting features are linear combinations of the original features. [Lever et al. 2017].

The origins of this method date back to 1901, when Karl Pearson in his work [Pearson 1901] presented the idea of finding the best-fit hyperplane in a set of points in multi-dimensional Euclidean spaces. However, it was in the 1930' when Harold Hotelling gave this method the name still used today. He was looking for a method by which he could express multiple factors of learning ability (e.g., counting speed, reading speed) as a single generalized measure. Over time, many variations of the PCA method were developed, but they were still based on the main principle, as the transformation of a set of correlated attributes into a set of independent attributes [Daga et al. 2020]. The newly created features are called principal components. The transformation is performed in such a way that the first component has the highest possible variance, and each subsequent component, assuming it is orthogonal to the previous one, also has a high variance [Li and Liu 2019].

PCA-based methods are encountered: *Sparse-PCA (SPCA)*[Deshpande and Montanari 2014], *Incremental-PCA (IPCA)*[Ross et al. 2008] and *kernel-PCA (kPCA)*[Schölkopf et al. 1997]. The SPCA method extends the original PCA method by selecting a few primary features rather than the entire set. It is shown that the SPCA method can maintain consistency even when the number of features is much larger than the number of samples in a given set [Johnstone and Lu 2009]. IPCA method is used for large datasets for which the memory complexity is too high.

For this purpose, the data is divided into smaller packets on which feature extraction is performed.

When the analyzed data cannot be linearly transformed, the projection of the data should be made into a nonlinear function, as in the kPCA method. This method, consists in mapping the space X into an extended feature space through some nonlinear mapping. The dimension of the newly created space, can be anything – either smaller or larger. The implementation of the nonlinear parameter transformation requires the *kernel trick* [Wang et al. 2016]. It relies on not having to coordinate the data in higher dimensions and plot hyperplanes on it.

The goal of the LDA (Linear discriminant analysis) method is to represent the original data matrix into a reduced dimensional matrix in terms of the number of features. In other words, it is a linear mapping of the input space into an output space of reduced size [Tharwat et al. 2016]. One of the most popular and widely used discriminative classifiers is Fisher's linear discrimination. This method was originally designed to separate two classes. Only later was it generalized, finding application also in the case of many classes and many features [Wen et al. 2018], [Leśkiewicz et al. 2016].

Linear discriminant analysis can be used in the processes of targeting advertisements to appropriate target groups. On the basis of questionnaires, it is possible to collect a lot of information and thus obtain a lot of features that can be characteristic for specific types of products. Another application can be speech recognition modules. On the basis of many features (frequency, intensity, tone) it is possible to detect the language in which words are spoken. Based on the analysis of features, it is possible to recognize specific words and to analyze them further, e.g. in terms of their emotional character.

Another known extraction method is Independent Component Analysis (ICA). The objective of the Independent Component Analysis method is to decompose multivariate signals into statistically independent signals [Hoyer and Hyvärinen 2000]. As a feature extraction method, it also focuses on dimensionality reduction, but is distinguished by generating components that will be as independent as possible.

The wavelet transform is a transformation that uses the scalar product of the signal under test with the kernel of the transformation. The result of the wavelet transform is the wavelet transform. This method of signal analysis offers more possibilities compared to the Fourier transform because it is not limited to the frequency domain only. It is based on irregular, non-periodic and asymmetric wavelet functions [Jaworek et al. 2001].

The advantage of the wavelet transform is its precision, which can be set arbitrarily. The wavelet transform can be used not only in the analysis of audio signals, but also digital images. In [Bobulski 2003] it is shown how a face recognition system can be built using it. Wavelets are also used to filter biomechanical signals and compress data [Jaworek et al. 2001], as well as for functional brain analysis, handwriting reading or speech recognition [Świebocka-Więk 2020].

Wrapper methods are often implemented using metaheuristics. These are general, high-level algorithms widely used in optimization problems. They are not related to any particular problem – they define a way of proceeding by operating on abstract concepts. Algorithms of this type are not guaranteed to find the globally best solution. They are used when the problem-specific way of constructing optimal solutions is not known and a brute force algorithm is not an option. It is necessary to construct an objective function that can evaluate the quality of each point in the solution space [Boussaïd et al. 2013]. Metaheuristics are thus one potential approach to this issue [Yusta 2009]. There is relatively high interest in applying metaheuristics to wrappers. Sharma and Kaur in their review [Sharma and Kaur 2020] mention a total of 100 different nature-inspired metaheuristics. Because of their characteristics, algorithms of this type can be extended, improved, and adapted to specific problems almost indefinitely. Many researchers take on this challenge resulting in many new methods that are variations or combinations of others already known. Marfaja and Mirjalili present a different approach to improving current performance [Marfaja and Mirjalili 2017]. The authors have developed a hybrid WOA-SA solution combining the features of whale optimization algorithm (WOA) with simulated annealing (SA). The former algorithm is responsible for global exploration while the latter performs local exploration. In comparison with the use of 18 data sets, WOA-SA proved more effective than genetic algorithm (GA), particle swarm optimization (PSO), and ant lion optimization (ALO) in classification accuracy. In most cases it also performed better in reducing the number of dimensions. Sayed et al. in 2019 presented a chaotic dragonfly algorithm (CDA) [Sayed et al. 2019]. There are also publications that focus on comparisons of effectiveness in concrete areas. Shukla et al. 2020 compared wrappers based on metaheuristics in feature selection from DNA microarrays [Shukla et al. 2020]. The authors tested GA, PSO, differential evolution (DE), and the ant colony algorithm (ACO).

Unlike feature selection, feature extraction uses different methods to extract new

features from the input data. Researchers have proposed many techniques in their work including multiresolution analysis (MRA) i.e. Wavelet (discrete wavelet transform) [Cvetkovic et al. 2008], Contourlet [Moayedi et al. 2010], Ridgelet [Ramos et al. 2012] and Curvelet [AlZubi et al. 2011].

3 Methods

In this section, the CCPCA (Centroid Class Principal Component Analysis) model is presented. This method rotates features not with respect to the center of the coordinate system determined for all patterns, but due to the centroids of the classes. In order to illustrate the method, it will be necessary to refer to classical PCA with an indication of the areas where there is a difference in the processing of the input data. Assume that we have a vector of n features: $X_j = [x_{j,1}, x_{j,2}, \dots, x_{j,n}]$, where j is the class. In addition, let us denote the point P_j being the centroid of the analyzed feature space for each j -th class. The coordinates of the centroid can be written in the form $C = (C_{x_1}, C_{x_2}, \dots, C_{x_j})$. The point of CCPCA is to rotate the axis coordinates to capture the content of the covariance information from its point of importance to the classes. In the first step, we standardize each feature using the z -scale for each i -th feature:

$$z_i = \frac{x_i - \mu_i}{\sigma_i} \quad (1)$$

The process of converting x_i values to the z_i scale causes a lossless shift of the entire feature space to the center of the coordinate system. In the PCA method, rotation occurs according to this central point. However, in the proposed CCPCA method, one should additionally, after normalizing (1) according to the centroids of each j -th class, displace all points to the center of the coordinate system by a vector $\bar{u}_j = [-C_{j,x_1} - C_{j,x_2}, \dots, -C_{j,x_n}]$, where n is the number of features. In the next step, we compute j covariance matrices which are numerical representations of the amount of information contained between n features.

$$A_j A_j^T = \begin{bmatrix} E[(x_1 - \bar{x}_1)(x_1 - \bar{x}_1)] \cdots E[(x_1 - \bar{x}_1)(x_n - \bar{x}_n)] \\ E[(x_2 - \bar{x}_2)(x_1 - \bar{x}_1)] \cdots E[(x_2 - \bar{x}_2)(x_n - \bar{x}_n)] \\ \vdots \quad \ddots \quad \vdots \\ E[(x_n - \bar{x}_n)(x_1 - \bar{x}_1)] \cdots E[(x_n - \bar{x}_n)(x_n - \bar{x}_n)] \end{bmatrix}. \quad (2)$$

Then, based on the matrix (2) we determine the matrix $B_j B_j^T$:

$$B_j B_j^T = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}, \quad (3)$$

The eigenvalue decomposition process can be represented as follows:

$$A_j A_j^T = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \times \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}, \quad (4)$$

where m is the number of samples, and n is the number of features.

During this process, we obtain eigenvectors which are new dimensions of CCPCA in feature space and eigenvalues. Eigenvalues are the information content of the eigenvectors. The CCPCA method decomposes the larger feature space into its smaller dimension taking into account the class information, according to the singular values (SVD). Therefore, any matrix A with n features and m samples can be written in the form:

$$A_{j(m \times n)} = U_{j(m \times n)} S_{j(m \times n)} V_{j(n \times n)}^T, \tag{5}$$

where: U and V are orthogonal matrices with orthonormal eigenvalues from matrices AA^T and $A^T A$. Whereas S is a diagonal matrix with positive values.

Matrices U and V are orthogonal, which means that their cross product is 0. The orthonormal matrix means that: (a) matrices are orthogonal, (b) the determinant is 1.

It now remains to explain the author's extraction method. In the new space, each principal component as already mentioned creates an eigenvector, with eigenvalues for each feature:

$$CCPC_{j,k} = (\omega(x_1), \omega(x_2), \dots, \omega(x_n)), \tag{6}$$

where: k is the k -th component for j -th class.

In the figure 1 the principle of the CCPCA method compared to PCA is presented.

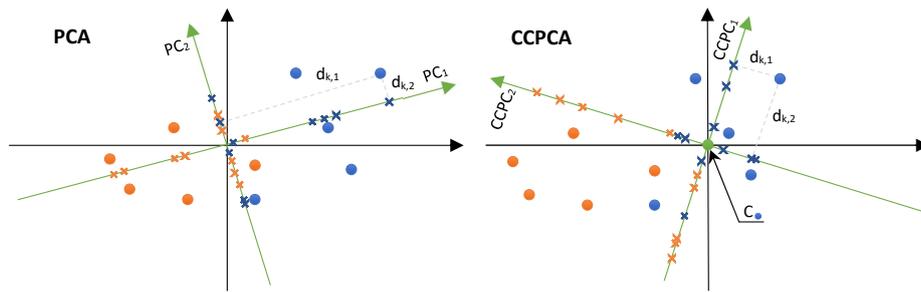


Figure 1: Example diagram showing the determination of CCPCA components by class centroids compared to PCA

Note that by rotating according to class centroids we get more components but also what we care about, i.e. better discrimination. In order to optimize the selection of an appropriate number of features for a given component, we should first sort the eigenvalues of equation (6) in descending order. Then starting with the two features with the largest eigenvalues from the matrix $A_j A_j^T$ (2) we calculate a measure of scale consistency:

$$\alpha = \frac{n}{n+1} \left(1 - \frac{\sum_{i=1}^n S_i^2}{S_{sc}^2} \right), \tag{7}$$

where: n - is the number of features, S_i^2 is the position value, and S_{sc}^2 is the scale value. The position value S_i^2 we can write with the formula:

$$S_i^2 = \sum_{i=1}^n E [(x_i - \bar{x}_i)(x_i - \bar{x}_i)]. \quad (8)$$

Whereas the scale value S_{sc}^2 we calculate as follows:

$$S_{sc}^2 = \sum_{k=1}^n \sum_{i=1}^n E [(x_k - \bar{x}_k)(x_i - \bar{x}_i)], \quad (9)$$

where: i are the columns of the matrix $A_j A_j^T$ (2), whereas k defines its rows.

In our case, for each component from the matrix (2) we determine the consistency (7) by adding features one by one until the maximum value is reached (7). If the maximum value $\alpha < 0.7$ then we extract the entire component. This means that the features in this component do not bring significant information to class discrimination by the feature vector considered in this component. Moreover, if two components extracted by the above method will contain the same feature subspace, all consistent components should be removed leaving only one.

For example, if for a given component for the features x_1 and x_5 the largest eigenvalues were obtained, the position values S_i^2 will be calculated as follows:

$$S_i^2 = E [(x_1 - \bar{x}_1)(x_1 - \bar{x}_1)] + E [(x_5 - \bar{x}_5)(x_5 - \bar{x}_5)]. \quad (10)$$

4 Experimental set-up

Four experiments were conducted in this study. The first experiment is concerned with evaluating the quality of classification as a function of the percentage of informative features and the extracted components. The second experiment determines the performance of classifiers for 5% and 10% informative features, depending on the number of features in the input for the extraction methods and the imbalance ratio. Both the above-mentioned experiments were conducted for binary classification. In Experiments 3 and 4, 10 sets of real data were used for testing. The selection of these sets is purposeful. The main criterion for data selection was class size, thus verifying how the proposed CCPCA method compared to other extraction methods is effective for multi-class cases. Real data gives a more complete view of the results. The last experiment was conducted for homogeneous ensembles of classifiers. The desirability of this study is due to the fact that rotating features by class centroids yields more feature subspaces than the PCA method. The last experiment aims to show whether the CCPCA method can generate good in the context of discrimination feature spaces for base classifiers working in ensembles.

Our research tries to find answers to four research questions:

1. What is the quality of classification after applying the author's method depending on the number of percentage of informative features in the dataset and the number of extracted components?
2. What is the quality of classification after applying the author's method depending on the number of features in the input and the imbalance ratio?
3. What is the quality of classification after the author's method for real data as a function of the number of features and classes?

4. What is the quality of classification after the application of author's method for real data depending on the number of features and classes for classification task of ensembles of classifiers based on the extracted feature spaces?

4.1 Used classifiers and extraction methods

Below all classifiers used in the experiments are listed. For each classifier, we provide hyperparameter sets that were considered for fine-tuning.

- **k-NN** - k Nearest Neighbours [Destercke 2012]
 - number of neighbors: 3, 5, 7
 - metrics: Minkowski, Euclidean, Manhattan
- **SVC** - Support Vector Classification/Support Vector Machine [Vladimir and Corinna 1995]
 - parameter C: 0.1, 1, 10, 100
 - kernel: linear, rbf, poly, sigmoid
 - gamma: scale, auto
- **CART** - Classification and Regression Trees [Li et al. 1984]
 - criterion: gini, entropy
 - splitter: best, random
 - maximum depth: 1, 2, 3, ..., 10
- **GNB** - Gaussian Naive Bayes - without parameters [Bayindir et al. 2017, Sopharak et al. 2008]
- **MLP** - Multi-layer perceptron [Arena et al. 1998]
 - number of hidden layers: 3,4,5,...,10
 - activation function: identity, logistic, entropy, SOS, Tanh, Linear, Softmax, Exponent
 - parameter alpha: 0.00001, 0.0001, 0.001, 0.01, 0.1
 - momentum: 0, 0.2, 0.4, 0.6, 0.8, 1

The extraction methods used in this study were PCA, KPCA, ICA and CCPCA.

4.2 Setup

For the first two experiments, we used synthetic data generated using the *make_classification* method from the *scikit-learn* [?] library. Using it, it is possible to create datasets for any multiclass problem. By modifying the parameters of the generator, it is possible to generate datasets of different complexity, so as to facilitate or complicate the classification process. The samples are generated according to the standard normal distribution. The most important parameters of the generator are:

- `n_samples` – number of samples,
- `n_features` – number of features (total),
- `n_informative` – number of informative features,
- `n_redundant` – number of redundant features (which are combinations of informative features),
- `n_repeated` – number of repeated features (duplicates among informative and redundant features),
- `weights` – the proportion of sample memberships to classes (imbalance),
- `flip_y` – percentage of samples with a misassigned class.

By default, without shuffling, the generated dataset contains informative, redundant and repeated features in sequence. The remaining features are random noise and carry no value.

In experiments using synthetic datasets, two-class datasets containing 1000 samples and 1000 features were generated. Among the features, there were only informative features and noise. Because of the equal distribution of the generated sets, 8 balanced sets were generated in the first experiment, containing 1%, 5%, 10%, 20%, 30%, 40%, 50%, 100% informative features, respectively. In this case, shuffling of informative attributes and noise was performed. In the second experiment, the generator parameter for the number of informative attributes was fixed and was set to 100 (10%). What was changed in this case, was the imbalance ratio (the number of samples per class). In the generator, the parameter `weights` is responsible for this. It took values $[x, 1 - x]$, for $x \in [0.01, 0.02, \dots, 0.5]$, for which x denoted the minority class proportion. In this case, for $x = 0.5$ the generated dataset was balanced and contained 500 samples of each class. For $x = 0.1$, the dataset contained 100 and 900 patterns of a given class, respectively.

In Experiments 3 and 4, as mentioned earlier, different real-world datasets were used. Denoted by k – the number of patterns, n – the number of features, and j – the number of classes, these datasets have the following parameters:

- Multiple Sclerosis ($n=54; k=378; j=2$),
- Lymphocytic leukemia ($n=40; k=239; j=3$),
- Chemical risk ($n=63; k=40000; j=4$),
- Kidney disease ($n=54; k=1260; j=5$),
- Lung disease ($n=38; k=733; j=6$),

- Bone cancer ($n=67; k=420; j=7$),
- Epilepsy diagnosis ($n=390; k=391; j=10$),
- Brain tumors ($n=55; k=783; j=14$),
- Diangosis diabets ($n=84; k=1100; j=15$),
- Eye diseases ($n=51; k=540; j=17$),

From the classification point of view, it is important to divide the data set into training and testing sets. The proportion of data in train and test split is defined by stratified 5-fold cross-validation.

4.3 Evaluation of different feature extraction methods depending on the number of components and informative features

Experiment 1 was performed on the synthetic data for the balanced binary case. In Figure 2 the notation (1%, ..., 100%), was used for the extraction methods, which indicates the percentage of informative features in the generated dataset. The number of extracted components was assumed to range from 1 to 800. The notation "NO" indicates the case when no feature extraction was performed. The experiments yielded the Accuracy results shown in Figure 2.

The experimental results provide interesting conclusions. The best classification quality is found for 5% and 10% informative features in the whole dataset. Moreover, the best classification metrics are obtained when each extraction method is used with about 100 components. The use of PCA, KPCA, ICA and CCPCA in the extraction task gives better Accuracy than when it is not used. Comparing the developed CCPCA extraction method with other methods, it is observed that the base classifiers used have higher classification quality which is due to better preparation of feature space by rotating it according to class centroids. KNN proves to be the best classifier for synthetic data, followed by SVC and MLP.

4.4 Assessment of classification quality due to unbalance and number of informative features

The second experiment was carried out for two variants of the percentage of informative features in the synthetic dataset: 5% – Figure 3 and 10% – Figure 4. The choice of such parameters is dictated by the results obtained in Experiment 1, where the best Accuracy values were obtained for these two percentages of informative features. In Figures 3-4 the notation (10, ..., 800) is used for the extraction methods, which means the number of all features in the used dataset. This experiment is intended to verify the quality of extraction in situations with different degrees of data imbalance. The example considered is for the case of binary classification.

Analyzing the obtained results Figure 3 and Figure 4 we can see that for 10% of informative features the classification quality is higher than on 5%. With small datasets for both variants of informative features, the classification quality is close to the random classifier. In these cases, the classifiers receive few features, which may cause such low classification quality. After applying extraction for different levels of imbalance, better classification quality results were obtained than in the case of no extraction. We can also observe that the more input features the classification quality increases, which is caused

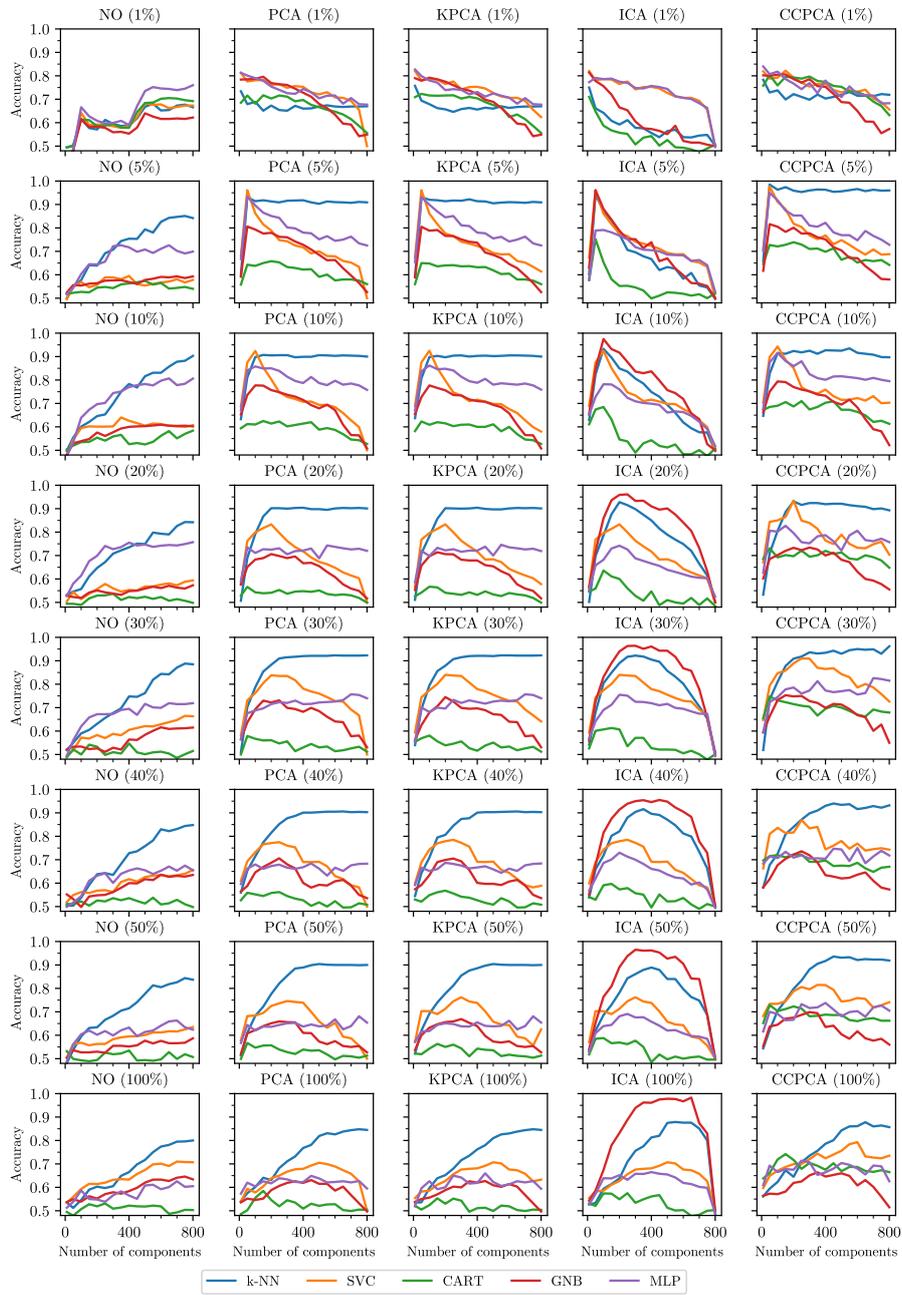


Figure 2: Dependence of classification quality on the number of components and the number of informative features

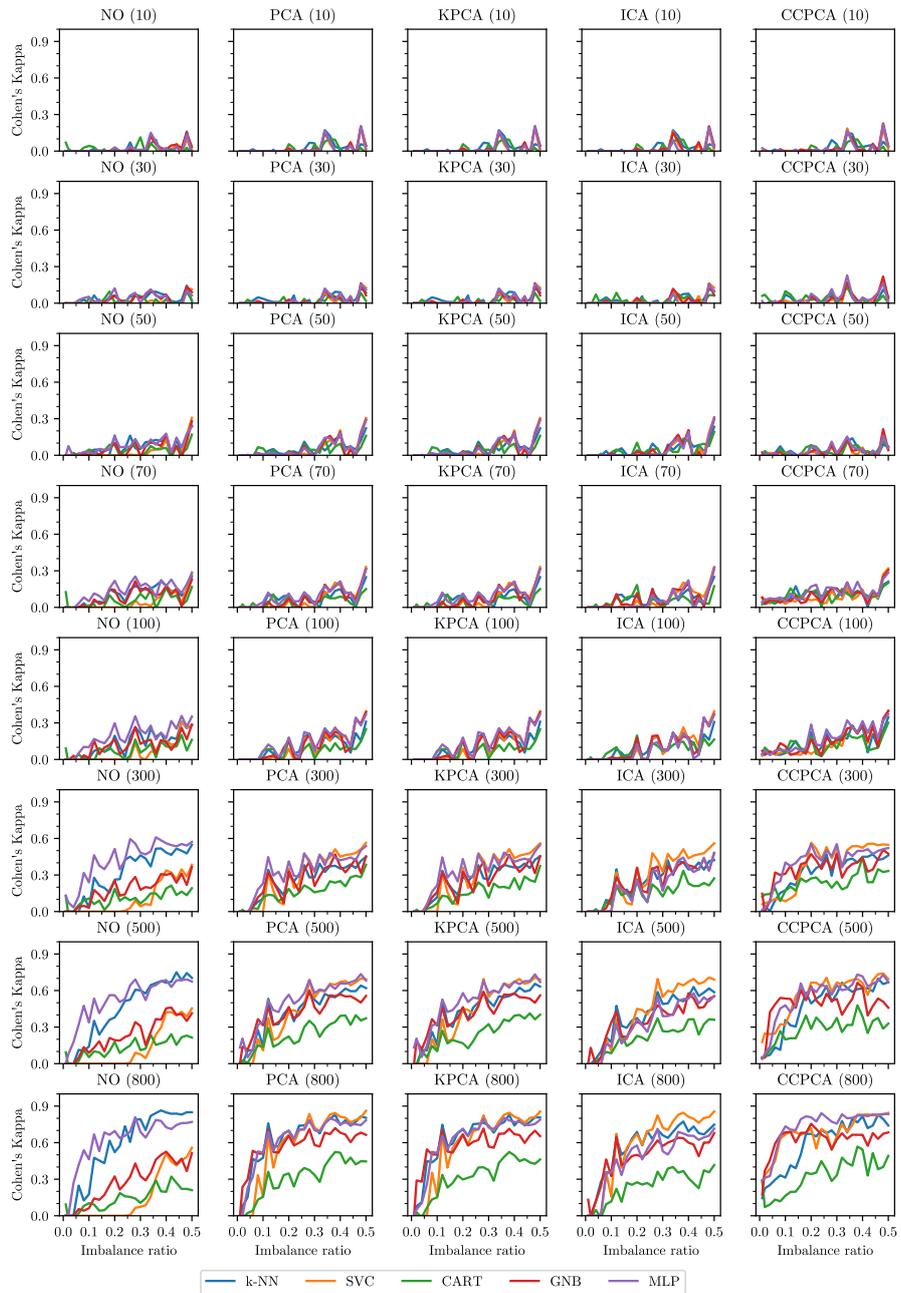


Figure 3: Dependence of classification quality on the number of features and class imbalance for 5% informative features

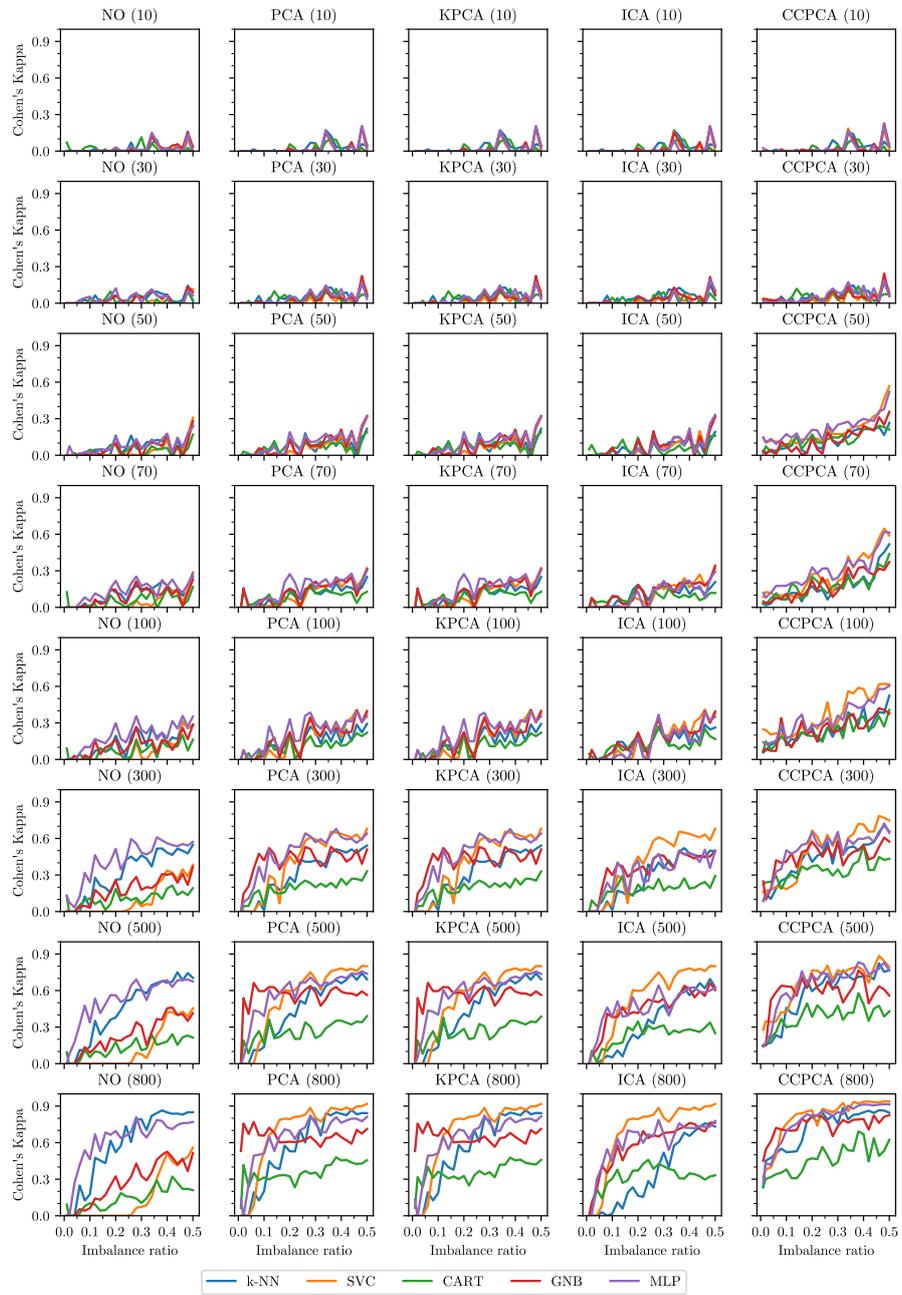


Figure 4: Dependence of classification quality on the number of features and class imbalance for 10% informative features

by better matching of the feature space. In addition, the CCPCA method achieves better classification quality on highly unbalanced datasets and reaches convergence faster, which is best illustrated by the case of 10% informative features. In this experiment as in the first one, KNN, SVC and MLP turn out to be the best performing classifiers with different dataset parameters and balancing.

4.5 Evaluation of the classification quality of extraction methods using real datasets

In the third experiment, ten real datasets are used. These are the cases for different number of features and classes which gives the variety of results obtained, especially for the multi-class case. Two best classifiers SVC and MLP were used in the experiment. These are the algorithms for which the best classification qualities were obtained on the ten datasets considered. The obtained results of the *BAC*-score metric are presented in Table 1.

For all real-world datasets, the classification performance after applied extraction is better than without extraction. Comparing each extraction method with the no extraction case, we can see an increase in *BAC*-score for: PCA (from 0.5% to 9.7%), KPCA (from 0.4% to 8.5%), ICA (from 0.7% to 8.6%), and CCPCA (from 1.4% to 11.3%). Thus, the CCPCA method due to rotation by class centroids produced statistically better classification qualities. The CCPCA method built a feature space on which the base classifiers compared to the other considered extraction methods obtained better *BAC*-score ranging from 0.7% to 6.4%. The largest differences were obtained on the Lymphocytic leukemia database.

4.6 Evaluation of the classification quality of extraction methods using real datasets and ensembles of classifiers

In the last, fourth experiment, the *BAC*-score classification quality was evaluated similarly to the third experiment, except that one base classifier is assigned for each emerged component constituting the feature subspace. This results in a homogeneous ensemble of classifiers, where each classifier performs classification on a distinct component. Two separate ensembles of classifiers based on SVC and MLP algorithms were used. In case of no extraction, random partitioning of feature space was used. Majority voting method was used in this task. The results of the experiment are presented in the Table 2.

Similar to the third experiment, we can see an advantage in the *BAC*-score after extraction compared to before extraction. Comparing each extraction method with the no extraction case, we can see an increase in *BAC*-score for: PCA (from 1.42% to 9.48%), KPCA (from 0.15% to 8.33%), ICA (from 0.31% to 8.81%) and CCPCA (from 3.16% to 12.18%). Thus, the CCPCA method due to rotation by class centroids produced statistically better classification qualities for ensembles of classifiers. The CCPCA method built a feature space on which the base classifiers compared to the other extraction methods considered obtained *BAC*-score better ranging from 1.4% to 6.7%. It is interesting to compare the results for experiment 4 and 3. We can thus demonstrate the validity of building an ensemble of classifiers on the emerged components. By comparing the results of experiment 4 with experiment 3, we can see that there was an increase in the quality of *BAC*-score by building ensembles of homogeneous classifiers. After using the PCA method, the classifier ensembles achieved an improvement in *BAC*-score quality up to 1.2%, for KPCA up to 1.1%, for ICA up to 1.3% and the most pronounced increase of

Table 1: BAC-score for real data sets for different extraction methods.

Dataset	Algorithm	n/j	NO _a	PCA _b	KPCA _c	ICA _d	CCPCA _e
Multiple Sclerosis	SVC	54/2	0.781 –	0.798 a	0.81 a	0.806 a	0.823 abcd
	MLP	54/2	0.789 –	0.794 a	0.802 a	0.801 –	0.819 abcd
Lymphocytic leukemia	SVC	40/3	0.733 –	0.784 a	0.787 a	0.782 a	0.846 abcd
	MLP	40/3	0.741 –	0.789 a	0.792 a	0.789 a	0.853 abcd
Chemical risk	SVC	63/4	0.799 –	0.824 a	0.821 a	0.819 a	0.841 abcd
	MLP	63/4	0.784 –	0.833 a	0.822 a	0.823 a	0.844 acd
Kidney disease	SVC	54/5	0.834 –	0.891 acd	0.867 a	0.872 a	0.913 abcd
	MLP	54/5	0.831 –	0.882 acd	0.856 a	0.865 a	0.902 abcd
Lung disease	SVC	38/6	0.882 –	0.917 acd	0.897 a	0.902 a	0.931 abcd
	MLP	38/6	0.885 –	0.912 acd	0.892 –	0.901 a	0.925 abcd
Bone cancer	SVC	67/7	0.892 –	0.921 ad	0.915 a	0.908 a	0.934 abcd
	MLP	67/7	0.889 –	0.907 a	0.901 a	0.905 a	0.928 abcd
Epilepsy diagnosis	SVC	73/10	0.713 –	0.789 acd	0.767 a	0.771 a	0.803 abcd
	MLP	73/10	0.702 –	0.799 ac	0.787 a	0.788 a	0.813 abcd
Brain tumors	SVC	55/14	0.812 –	0.833 ac	0.819 a	0.824 a	0.857 abcd
	MLP	55/14	0.801 –	0.821 ac	0.807 a	0.820 ac	0.852 abcd
Diagnosing diabetes	SVC	84/15	0.845 –	0.873 a	0.865 a	0.864 a	0.891 abcd
	MLP	84/15	0.831 –	0.869 a	0.843 a	0.842 a	0.885 abcd
Eye diseases	SVC	51/17	0.721 –	0.734 a	0.731 a	0.736 a	0.753 abcd
	MLP	51/17	0.722 –	0.738 a	0.735 a	0.738 a	0.751 abcd

n – number of features
j – number of classes

BAC-score is observed for CCPCA up to 3.1%. This indicates that ensembles of base classifiers can be built on the spaces emerged by the new CCPCA method.

5 Lessons learned

Conducted research allowed to answer four research questions. Thus, new knowledge was obtained. According to the results obtained in the first experiment we can conclude

Table 2: BAC-score for different extraction methods and homogeneous ensembles

Dataset	Algorithm	n/j	NO _a	PCA _b	KPCA _c	ICA _d	CCPCA _e
Multiple Sclerosis	SVC	54/2	0.794 –	0.808 a	0.821 ab	0.818 a	0.835 abcd
	MLP	54/2	0.789 –	0.806 a	0.809 a	0.81 a	0.832 abcd
Lymphocytic leukemia	SVC	40/3	0.733 –	0.793 a	0.795 a	0.794 a	0.851 abcd
	MLP	40/3	0.742 –	0.799 a	0.8 a	0.797 a	0.863 abcd
Chemical risk	SVC	63/4	0.802 –	0.833 a	0.822 a	0.827 a	0.858 abcd
	MLP	63/4	0.796 –	0.837 a	0.832 a	0.833 a	0.851 abcd
Kidney disease	SVC	54/5	0.834 –	0.894 acd	0.872 a	0.876 a	0.921 abcd
	MLP	54/5	0.834 –	0.885 acd	0.859 a	0.875 a	0.918 abcd
Lung disease	SVC	38/6	0.885 –	0.928 acd	0.898 a	0.91 a	0.942 abcd
	MLP	38/6	0.892 –	0.921 acd	0.902 a	0.913 a	0.937 abcd
Bone cancer	SVC	67/7	0.902 –	0.927 a	0.923 a	0.916 a	0.941 abcd
	MLP	67/7	0.898 –	0.919 a	0.908 a	0.912 a	0.937 abcd
Epilepsy diagnosis	SVC	73/10	0.716 –	0.795 acd	0.776 a	0.783 a	0.814 abcd
	MLP	73/10	0.713 –	0.807 acd	0.796 a	0.801 a	0.831 abcd
Brain tumors	SVC	55/14	0.821 –	0.836 a	0.826 a	0.824 a	0.873 abcd
	MLP	55/14	0.811 –	0.83 ac	0.812 a	0.826 a	0.864 abcd
Diagnosing diabetes	SVC	84/15	0.851 –	0.884 ac	0.871 a	0.873 a	0.906 abcd
	MLP	84/15	0.838 –	0.875 acd	0.846 a	0.85 a	0.894 abcd
Eye diseases	SVC	51/17	0.723 –	0.745 a	0.739 a	0.737 a	0.779 abcd
	MLP	51/17	0.735 –	0.75 a	0.744 a	0.743 a	0.767 abcd

n – number of features
j – number of classes

that in comparison with other extraction methods i.e. PCA, KPCA and ICA the developed CCPCA method allows to increasing the quality of classification. Interesting results were obtained for datasets consisting of 5% and 10% informative features. As the extracted components increase, although the classification quality decreases, but for CCPCA this decrease is slower. By rotating the feature space according to the class centroids, we can increase the classification quality but also reduce the negative impact on the classification quality of sets containing large measurement noise.

The second experiment was also conducted on synthetic data. It allowed to answer the second research question. The first conclusion that emerges from the research is that for small data sets, the results of correct classifications are low. Using the CCPCA method, it was noticed that the quality of correct classifications increases faster or reacts due to the increase in the number of features in the input and to the imbalance. In case of highly imbalanced sets using the author's method it was possible to obtain higher classification quality compared to other feature extraction methods. It was also shown that in case of very high imbalance PCA, KPCA and ICA methods allow to obtain worse classification quality.

Third experiment helped answer the third research question. The experiment is based on ten real data sets. Thus, it was shown that statistically significantly ($p < 0.05$) for the cases of binary and multiclass classification, feature extraction performed by the author's method produced better quality correct classifications than the PCA, KPCA and ICA methods. Moreover, after using any extraction method, the classification quality is higher than when no extraction is used. In the experiments on real data, the best qualities of correct classifications were obtained for SVC and MLP methods, hence they are finally presented.

The fourth research question concerned ensembles of homogeneous classifiers. The experiment was designed to verify how extracted feature subspaces affect classification quality. For this purpose, one classifier was embedded on each component representing a feature subspace. The results obtained confirmed the analysis obtained in Experiment 3. The CCPCA method proved to be the best discriminator of the feature space for the classifier ensemble recognition task. Moreover, an interesting result was obtained showing that using ensembles of classifiers one can build on the extracted subspaces models of classifiers which obtain better classification qualities than in the case of single classifier.

The obtained results allow us to draw a final conclusion that by rotating the feature space according to class centroids and using a measure of the number of features and components evaluation based on Cronbach's alpha coefficient, we can perform feature extraction increasing the classification quality of single and ensemble classifiers compared to other extraction methods i.e. PCA, KPCA or ICA. Moreover, as shown, this space can also be well constructed for imbalanced data.

6 Conclusions

The paper presents a proposal of modification of PCA method which assumes rotation of features according to determined centroids of classes. As a result of conducted research, a tool supporting pattern recognition and classification was obtained. The method was verified in four experiments, which were based on synthetic and real data. It was shown that, compared to PCA, ICA and KPCA methods, the CCPCA proposal allows to build subspaces which better discriminate classes, which is a desirable phenomenon in pattern classification. Moreover, it was shown that by performing rotation by classes, the constructed feature subspaces allow to obtain better classification results for strongly imbalanced data than when using unsupervised feature extraction methods. Ten real datasets with different number of features and classes were used in this study. This made it possible to compare the classification quality after CCPCA extraction for multi-class cases. Again, better classification quality results were obtained here. Testing on real data is an important part of the research because in practice there are various relationships between features which affect the quality of extraction. Since the CCPCA method builds

more components than PCA, which is a phenomenon that occurs when rotating by classes, 4 sets of classifiers were used in the experiment. Thus, it was shown that the developed solution can be the basis for building such classifiers, since they operate on components that are strongly associated with a class. This resulted in an additional improvement in the quality of classification compared to the use of single classifiers. An important element of the developed method is that it estimates the number of components based on the internal consistency coefficient. By maximizing the eigenvalue, a very good fit of the components to the construct of all features discriminating a class was obtained. The results obtained indicate the validity of extending this method. Similar approach can be applied to other methods i.e. kPCA, ICA, which are based on similar mechanisms of determining principal or independent components. In addition, tests will be carried out taking into account different levels of the consistency coefficient. The goal of the research has been achieved. A method modifying classical PCA has been created, showing the superiority of the CCPCA method in feature extraction for the classification task of single classifiers and their ensembles.

Finally, the following further research directions were defined:

1. The application of the k-means method to determine other clusters in the data and their centroids than those based solely on classes. This is of particular importance for different spatial structures of features.
2. Application of CCPCA method in the task of feature extraction for data streams and biomedical signal extraction.
3. Modification of the CCPCA method for the cases of imbalanced and multi-class data.

Acknowledgements

This work was supported by the Polish National Science Center under the grant No. 2017/27/B/ST6/01325.

References

- [AlZubi et al. 2011] AlZubi, S. Islam, N. Abbod, M. “Multiresolution analysis using wavelet, ridgelet, and curvelet transforms for medical image segmentation,” *International journal of biomed. imaging*, vol. 2011, 2011.
- [Arena et al. 1998] Arena, P. Fortuna, L. Muscato, G. Xibilia, MG. “Neural Netw. in Multidim. Domains: fundamentals and new trends in modelling and control”. Springer 1998.
- [Bayindir et al. 2017] Bayindir, R. Yesilbudak, M. Colak, M. Naci, G. “A novel application of naive bayes classifier in photovoltaic energy prediction,” 523–527, 12 2017.
- [Bellman 1957] Bellman, RE. “Dynamic programming” Princeton University Press, Princeton, 497–520, 1957.
- [Bobulski 2003] Bobulski, J. “Ekstrakcja cech twarzy za pomocą transformaty falkowej,” *Inform. Teoretyczna i Stosowana*, vol. 3, no. 4, 127–134, 2003.
- [Boussaïd et al. 2013] Boussaïd, I. Lepagnot, J. Siarry, P. “A survey on optimization metaheurist,” vol. 237, 82–117. 2013

- [Çelik 2018] Çelik Ö.: "A research on machine learning methods and its applications," *Journal of Educational Technology and Online Learning*, vol. 1, no. 3, 25-35, 2018.
- [Charles et al. 2019] Charles, V. Aparicio, J. Zhu, J. "The curse of dimensional. of decision making units: A simple approach to increase the discriminatory power of data envelopment analysis," *European Journal of Operational Research*, vol. 279, no. 3, 929 – 940, 2019.
- [Chmielnicki 2012] Chmielnicki, W. "Efektywne metody selekcji cech i rozwiązywania problemu wieloklasowego w nadzorowanej klasyfikacji danych" 44 – 52, 2012.
- [Crochepierre et al. 2020] Crochepierre, L. Boudjeloud-Assala, L. Barbesant, V. "Interpretable dimensionally-consistent feature extraction from electrical network sensors," *European Confer. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases ECML/P-KDD'20*, 2020.
- [Cvetkovic et al. 2008] Cvetkovic, D. Übeyli, E. D. Cosic, I. "Wavelet transf. feature extraction from human ppg, ecg, and eeg signal responses to elf pemf exposures: A pilot study" *Digital signal processing*, vol. 18, no. 5, 861–874, 2008.
- [Cyganek et al. 2016] Cyganek, B. Graña, M. Krawczyk, B. Kasprzak, A. Porwik, P. Walkowiak, K. Woźniak, M. "A survey of big data issues in electronic health record analysis," *Applied Artificial Intelligence*, vol. 30, no. 6, 497–520, 2016.
- [Daga et al. 2020] Daga, A. P. Fasana, A. Garibaldi, L. Marchesiello, S. "On the use of pca for diagnostics via novelty detection: interpretation, practical application notes and recommendation for use," in *PHM Society European Conference*, vol. 5, 2020.
- [Deshpande and Montanari 2014] Deshpande, Y. Montanari, A. "Information-theoretically optimal sparse pca," in *2014 IEEE International Symposium on Information Theory*, pp. 2197–2201, IEEE, 2014.
- [Destercke 2012] Destercke, S. "A k-nearest neighbours method based on imprecise probabilities," *Soft Comput* vol. 5. 833–844. 2012.
- [Hamet and Tremblay 2018] Hamet, P. Tremblay, J. "Artificial intelligence in medicine," *Metabolism*, vol. 69, 2017. *Insights Into the Future of Medicine: Technologies, Concepts and Integration*. 36 – 40, 2017.
- [Han et al. 2017] Han, J. Jentzen, A. Weinan, E. "Overcoming the curse of dimensionality: Solving high-dimensional partial differential equations using deep learning," *arXiv preprint arXiv:1707.02568*, 1–13, 2017.
- [Hazelwood et al. 2018] Hazelwood, K. Bird, S. Brooks, D. Chintala, S. Diril, U. Dzhulgakov, D. Fawzy, M. Jia, B. Jia, Y. Kalro, A. "Applied machine learning at facebook: A datacenter infrastructure perspective," in *2018 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, 620–629, IEEE, 2018.
- [Hoyer and Hyvärinen 2000] Hoyer P. Hyvärinen, A. "Independent component analysis applied to feature extraction from colour and stereo images," *Network: computation in neural systems*, vol. 11, no. 3, 191–210, 2000.
- [Jain and Gupta] Jain, A. K. Gupta, B. B. "Towards detection of phishing websites on client side using machine learning based approach," *Telecommunication Systems*, vol. 68, no. 4, 687–700, 2018.
- [Jaworek et al. 2001] Jaworek, K. Kownacki, C. Pauk, J. "Transformata falkowa-nowoczesne narzędzie do analizy sygnałów pomiarowych," *Zeszyty Naukowe Politechniki Białostockiej. Budowa i Eksploatacja Maszyn*, pp. 199–212, 01 2001
- [Lie and Jiawei 2017] Jie, C. Jiawei, L. "Feature selection in machine learning: A new perspective" 2017.
- [Joachims 1998] Joachims, T. "Text categorization with support vector machines: Learning with many relevant features," in *Machine Learning: ECML-98 (C. Nédellec and C. Rouveirol, eds.)*, (Berlin, Heidelberg), pp. 137–142, Springer Berlin Heidelberg, 1998.

- [Johnstone and Lu 2009] Johnstone, I. M. Lu, A. Y. "On consistency and sparsity for principal components analysis in high dimensions," *Journal of the American Statistical Association*, vol. 104, no. 486, 682–693, 2009.
- [Khurana et al. 2016] Khurana, U. Nargesian, F. Samulowitz, H. Khalil, E. "Automating feature engineering" *Transformation*, vol. 10, no. 10, 10, 2016
- [Kuo and Sloan 2005] Kuo F. Y. Sloan, I. H. "Lifting the curse of dimensionality," *Notices of the AMS*, vol. 52, no. 11, 1320–1328, 2005.
- [Kurlej and Wozniak 2011] Kurlej B. Wozniak, M. "Active learning approach to concept drift problem" *Logic Journal of the IGPL*, vol. 20, 550–559, 02 2011.
- [Lensen et al. 2019] Lensen, A. Xue, B. Zhang, M. "Can genetic program. do manifold learning too?," 114–130, 2019.
- [Leśkiewicz et al. 2016] Leśkiewicz, M. Kaliszewski, M. Mierczyk, Z. Włodarski, M. "Compar. of principal component analysis and linear discriminant analysis applied to classification of excitation-emission matrices of the selected biological material," 2016.
- [Lever et al. 2017] Lever, J. Krzywinski, M. Altman, N. "Points of significance: PCA," 2017.
- [Li and Liu 2019] Li D. Liu, S. "Chapter 4 - water quality evaluation," *Water Quality Monitoring and Management* (D. Li and S. Liu, eds.), 113–159, Academic Press, 2019.
- [Li et al. 1984] LI, B. Friedman, J. Olshen, R. Stone, C. "Classification and Regression Trees (CART)", vol. 40. 09 1984.
- [Liang et al. 2018] Liang, H. Sun, X. Sun, Y. Gao, Y. "Correction to: Text feature extraction based on deep learning: a review," 2018
- [Mafarja and Mirjalili 2017] Mafarja, M. M. Mirjalili, S. "Hybrid whale optimization algorithm with simulated annealing for feature selection," vol. 260, 302–312. 2017.
- [Moayed et al. 2010] Moayed, F. Azimifar, Z. Boostani, R. "Contourlet-based mammography mass classification using the svm family," *Computers in biology and medicine*, vol. 40, no. 4, 373–383, 2010.
- [Nargesian et al. 2017] Nargesian, F. Samulowitz, H. Khurana, U. Khalil, E. B. Turaga, D. S. "Learning feature engineering for classification.," 2529–2535, 2017.
- [Oussous et al. 2018] Oussous, A. Benjelloun, F. Z. Lahcen, A. A. Belfkih, S. "Big data technolog.: A survey," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 4, 431–448, 2018.
- [Pearson 1901] Pearson, K. "Liii. on lines and planes of closest fit to systems of points in space" *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, vol. 2, no. 11, 559–572, 1901.
- [Ramos et al. 2012] Ramos, R. P. Nascimento, M. Z. Pereira, D. C. "Texture extraction: An evaluation of ridgelet, wavelet and co-occurrence based methods applied to mammograms," *Expert Systems with Applications*, vol. 39, no. 12, 11036–11047, 2012.
- [Ross et al. 2008] Ross, D. A. Lim, J. Lin, R.-S. Yang, M.-H. "Incremental learning for robust visual tracking," *International journal of computer vision*, vol. 77, no. 1, 125–141, 2008.
- [Sayed et al. 2019] Sayed, G. I. Tharwat, A. Hassanien, A. E. "Chaotic dragonfly algorithm: an improved metaheuristic algorithm for feature selection," vol. 49, no. 1, 188–205, 2019.
- [Schölkopf et al. 1997] Schölkopf, B. Smola, A. Müller, K. "Kernel principal compon. analysis," in *International conference on artificial neural networks*, pp. 583–588, Springer, 1997.
- [Seifert et al. 2020] Seifert, R. Weber, M. Kocakavuk, E. Rischpler, C. Kersting, D. "Ai and machine learning in nuclear medicine: Future perspectives," *Seminars in Nuclear Medicine*, 2020.

- [Sharma et al. 2017] Sharma, A. Bhuriya, D. Singh, U. "Survey of stock market prediction using machine learning approach," in 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), vol. 2, 506–509, IEEE, 2017.
- [Sharma and Kaur 2020] Sharma, M. Kaur, P. "A comprehensive analysis of nature-inspired metaheuristic techniques for feature selection problem" Archives of Computational Methods in Engineering, 2020.
- [Shukla et al. 2020] Shukla, A. K. Tripathi, D. Reddy, B. R. Chandramohan, D. "A study on metaheuristics approaches for gene selection in microarray data: algorithms, applications and open challenges," vol. 13, no. 3, 309–329. 2020.
- [Sopharak et al. 2008] Sopharak, A. Nwe, K. Moe, Y. Dailey, M. Uyyanonvara, B. "Automatic exudate detection with a naive bayes classifier," International Conference on Embedded Systems and Intelligent Technology, 139–142, 2008.
- [Stilgoe 2018] Stilgoe, J. "Machine learning, social learning and the governance of self-driving cars," Social studies of science, vol. 48, no. 1, 25–56, 2018.
- [Świebocka-Więk 2020] Świebocka-Więk, J. "Transformata falkowa", dostep: listopad 2020.
- [Tharwat et al. 2016] Tharwat, A. Gaber, T. Ibrahim, A. Hassanien, A. E. "Linear discriminant analysis: A detailed tutorial," AI communications, vol. 30, no. 2, 169–190, 2017.
- [Vladimir and Corinna 1995] Vladimir, V. Corinna, C. "Support-vector networks," Machine Learning, vol. 20, no. 3, 273–297, 1995.
- [Wang et al. 2016] Wang, S. Zhan, T. Chen, Y. Zhang, Y. Yang, M. Lu, H. Wang, H. Liu, B. Phillips, P. "Multiple sclerosis detection based on biorthogonal wavelet transform, rbf kernel principal component analysis, and logistic regression," IEEE Access, vol. 4, 567–7576, 2016
- [Wen et al. 2018] Wen, J. Fang, X. Cui, J. Fei, L. Yan, K. Chen, Y. Xu, Y. "Robust sparse linear discriminant analysis," IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 2, pp. 390–403, 2018.
- [Wozniak and Zmyslony 2010] Wozniak, M. Zmyslony, M. "Designing fusers on the basis of discriminants - evolutionary and neural methods of training," in Hybrid Artificial Intelligence Systems, 5th International Conference, HAIS 2010, San Sebastián, Spain, June 23–25, 2010. Proceedings, Part I (M. G. Romay, E. Corchado, and M. T. Garcia Sebastián, eds.), vol. 6076 of Lecture Notes in Computer Science, 590–597, Springer, 2010.
- [Wu et al. 2017] Wu, T. Liu, S. Zhang, J. Xiang, Y. "Twitter spam detection based on deep learning," in Proceedings of the australasian computer science week multiconference, pp. 1–8, 2017.
- [Yusta 2009] Yusta, S. C. "Different metaheuristic strategies to solve the feature selection problem," vol. 30, no. 5, 525–534, 2009.
- [Zhao et al. 2019] Zhao, R. Yan, R. Chen, Z. Mao, K. Wang, P. Gao, R. X. "Deep learning and its applications to machine health monitoring," Mechanical Systems and Signal Processing, vol. 115, 213 – 237, 2019.
- [ZhuWei et al. 2017] ZhiWei, M. Singh, M. M. Zaaba, Z. F. "Email spam detection: a method of metaclassifiers stacking," in The 6th international conference on computing and informatics, 750–757, 2017.