ARTTE
Applied Researches in Technics, Technologies and Education
Journal of the Faculty of Technics and Technologies, Trakia University
https://sites.google.com/a/trakia-uni.bg/artte/

URBAN ENVIRONMENTAL QUALITY ASSESSMENT BY SHAPE AND SPECTRAL INDICES OF MULBERRY LEAVES

Snejana Dineva, Zlatin Zlatev
Trakia University, Faculty of Technics and Technologies,
38 Graf Ignatiev str., 8602, Yambol, Bulgaria,
e-mail: snezhana.dineva@trakia-uni.bg, zlatin.zlatev@trakia-uni.bg

Abstract: *In this paper, an analysis of the potential use of the surface and geometric characteristics of mulberry leaves as parameters for environmental quality assessment is made. Methods have been used to reduce the amount of data of latent variables, linear and kernel variants of principal components. It has been found that a kernel variant of the principal components, combined with nonlinear separating functions of discriminant analysis and a method of support vector machines, are an appropriate methods for distinguishing the degree of air pollution from the mulberry leaf data. The results obtained could be used as preliminary baseline data for future evaluations and studies related to remote monitoring of urban air quality.*
Keywords: *Environmental quality, passive biomonitoring, mulberry leaves, principal components, latent variables, discriminant analysis, support vector machines, spectral indices, shape indices.*

1. INTRODUCTION

Environmental quality monitoring can be done by introducing biomonitoring using data from leaves of growing trees in urban environments whose characteristics are consistent with the air quality of their habitat [12,32]. Monitoring and analysis of the environment depending on changes in tissues and organs of plants is known as passive monitoring [3].
In order to realize such passive monitoring of the environment, it is necessary for the studied plants to be susceptible to these changes, changing the characteristics of their individual parts as a leaf, stem.
Suitable for passive monitoring of a small-scale studied plant in relation to the degree of air pollution in its habitats is mulberry [5,17]. The mulberry (*Morus* L.) is of economic importance in relation to the feeding of silkworms, cattle and goats. In some cases it is used as a park tree. In Bulgaria it is mainly found near major roads in the cities.
There are few studies on the qualitative indicators related to the surface texture of mulberry leaves, as well as the impact of the contaminated environment in the area of the plant's habitat.
The studies presented in the available literature [9,10,18] are related to the application of approaches to analyze different types of data due to the presence of environmental sensor networks and remote sensing tools. These technical tools determine the presence of harmful substances for humans. Different classification strategies are used to predict air quality. When solving tasks for the classification of areas with strong and less pollution in real time it is essential the time of classification. The use of classification methods, which require a long time to solve and are based on a large volume of computational procedures, is inappropriate for the tasks under consideration. When selecting an appropriate classifier, the main criteria to be considered are applicability, effectiveness and time for classification, with sufficiently high classification accuracy [21,25].

The main purpose of this study is to evaluate, through classifiers, the potential use of certain surface and geometric characteristics of mulberry leaves as parameters for assessing environmental quality.

2. MATERIAL AND METHODS

The mulberry leaf samples were taken from 4 zones, 2 of which had high traffic and 2 low road traffic. 25 leaves were used from each area taken from the sun-exposed side of the trees. The measurements were made in the town of Yambol, located in the southeastern part of Bulgaria.
The leaves are grouped into 2 groups - derived from less polluted zones **(LP)** and polluted zones **(P)**.
To determine the environmental parameters in the analyzed areas, an experimental set-up was used, developed at the Faculty of Technics and Technologies, Yambol, Bulgaria [33]. The measuring device consists of a sensor module and a microprocessor control system offering wireless communication.
By the system were measured: smoke gasses, ppm; particle matter PM>0,5 μm, μg/m^3; equivalent CO_2, (eCO_2), ppm; total volatile organic compounds, TVOC, ppb.
The measurements were made at a temperature of 22±3 ^oC and a relative humidity of 39±5 %RH.
The measured parameters of the environment in the habitats of mulberry are shown in Table 1. It has been found that in the high polluted areas the parameter values are significantly higher than in the less polluted zones.

Table 1. Environmental parameters

Zone	Smoke gasses, ppm	PM>0,5μm, μg/m^3	eCO_2, ppm	TVOC, ppb
P	0,57±0,1	79,32±0,6	641±37	35±2
LP	0,11±0,003	19,36±1,2	427±9	12±0,3

2.1. Experimental set-up

The experimental set-up used consists of a personal computer with software for receiving and processing images in the visible and near-infrared areas. The two video cameras used are mounted in an IP54 waterproof case. The shooting distance is adjusted by a movable stand. The leaves were captured with a video camera, 19,5 cm from the leaf to the camera. White LEDs with a maximum light intensity of 450 nm were used.

2.2. An algorithm for determining leaf shape indexes

A software algorithm has been developed in the Matlab environment. The size adjustment is made by measuring the caliper scale, to the nearest 0,05 mm. The correction factor obtained is 4,8 pix/mm. The *regionprops* function was used to determine the basic geometric dimensions of the leaves. This function determines the large and small axis, area and perimeter of the leaves. The geometric coefficients of these sheets are obtained by mathematical dependencies. All dependencies are calculated from the basic leaf sizes. D denotes the large axis and d represents the minor axis of the leaves, A - area, P - perimeter.

Ideal area, A_i

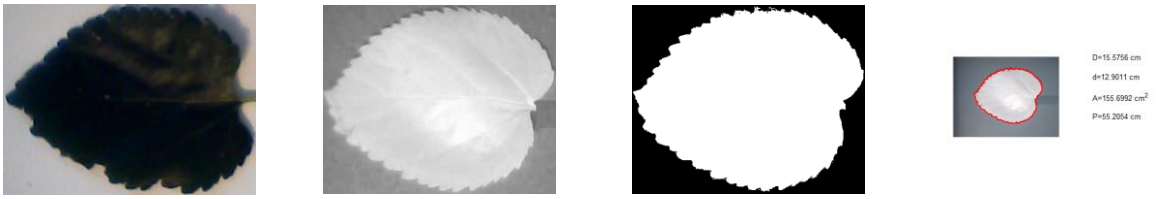
$$A_i = \frac{\pi dD}{4} \quad (1)$$

Area of bounding box, A_{mr}

$$A_{mr} = d.D \quad (2)$$

Coefficient of form, K_f	$K_f = \frac{P^2}{A}$	(3)
Eccentricity, K_1	$K_1 = \frac{D}{d}.100$	(4)
Ovality, c	$c = \frac{P^2}{4\pi A}$	(5)
Roundness, R	$R = \frac{1}{c}$	(6)
Area ratio, K_A	$K_A = \frac{A}{A_{ideal}}$	(7)
Area ratio, K_{AM}	$K_{AM} = \frac{A}{A_{mr}}$	(8)
K_2	$K_1 = \frac{D}{A}$	(9)
K_3	$K_2 = \frac{A}{D^3}$	(10)
K_4	$K_3 = \frac{\mathrm{A}}{\frac{D}{2}.\frac{d}{2}}$	(11)

Figure 1 shows the steps of the algorithm developed to determine the basic shape indices of a mulberry leaves. The digital image is loaded and converted from RGB to LMS model. It has been tested that the S component of the LMS model is suitable for separating the object (leaf) from the background.



a) loading an image | b) conversion to LMS and extraction of S component | c) conversion to black and white | d) determining of shape indices

Figure 1. Basic steps of an algorithm for determining shape indices of a mulberry leaf

The resulting image was converted to black and white with a binarization threshold of T=0,5. The basic dimensions of the leaf are determined using this *regionprops* function, and its basic geometric features are calculated using the mathematical dependencies mentioned above. The algorithm is run for each sample individually.

2.3. Obtaining spectral characteristics

The transformation of values from XYZ and LMS models into reflection spectra in VIS range, in the range 390-730nm, is done mathematically and the transformation is possible in both directions of equality [15]. Mathematical dependencies, with the possibility of converting in both directions of equality are:

$$X = \int_{380}^{780} A(\lambda)\bar{X}(\lambda)d\lambda; \quad Y = \int_{380}^{780} A(\lambda)\bar{Y}(\lambda)d\lambda; \quad Z = \int_{380}^{780} A(\lambda)\bar{Z}(\lambda)d\lambda \tag{12}$$

where A(λ) is a matrix for converting color to reflection spectra in the VIS range for the observer 2° and illumination D65 used. A correspondence matrix adapted by Mather was used [22]. The illumination data used to convert the VIS characteristics are in accordance with D65 (average daylight with UV component (6500K)).
The conversion function between RGB and XYZ model, in the range 380-780 nm can be represented as:

$$XYZ = RGB.M$$
$$M = \begin{bmatrix} 0{,}5767 & 0{,}2974 & 0{,}0270 \\ 0{,}1855 & 0{,}6273 & 0{,}0707 \\ 0{,}1882 & 0{,}0753 & 0{,}9911 \end{bmatrix} \tag{13}$$

where M is the transformation matrix under the specified conditions for observer 2° and illumination D65. The used matrices for converting (matching functions) of color components to spectrum are available in [8] for the VIS range.

2.4. Determination of spectral indices

Spectral indices were used according to Cermakova et al. [6] and Atanassova et al. [2]. The indices are in the reflection spectra (R). The following formulas were used to calculate them:

Red Edge Index (REI) $$REI = \frac{R_{740}}{R_{720}} \tag{14}$$

Photochemical Transmittance Index (PTI) $$PTI = \frac{R_{530} - R_{570}}{R_{530} + R_{570}} \tag{15}$$

Carotenoid transmittance Index (CTI) $$CTI = \frac{1}{R_{510}} - \frac{1}{R_{550}} \tag{16}$$

Triangular Vegetation Index (TVI) $$TVI = 0{,}5(120(R_{750} - R_{550}) - 200(R_{670} - R_{550})) \tag{17}$$

Greennes Index (G) $$G = \frac{R_{550}}{R_{680}} \tag{18}$$

Normalized Excess Green Index (NExG) $$NExG = \frac{2R_{520} - R_{620} - R_{420}}{R_{520} + R_{620} + R_{420}} \tag{19}$$

Normalized green red difference index (NGRDI) $$NGRDI = \frac{R_{520} - R_{620}}{R_{520} + R_{620}} \tag{20}$$

Red-Green-Blue Vegetation Index (RGBVI) $$RGBVI = \frac{{R_{520}}^2 - R_{620}R_{420}}{{R_{520}}^2 + R_{620}R_{420}} \tag{21}$$

Green Leaf Index (GLI) $$GLI = \frac{2R_{520} - R_{620} - R_{420}}{2R_{520} + R_{620} + R_{420}} \tag{22}$$

Visible Atmospherically Resistant Index (VARI)
$$VARI = \frac{R_{520} - R_{620}}{R_{520} + R_{620} - R_{420}} \quad (23)$$

Excess Green Index (ExG)
$$ExG = 2R_{520} - R_{620} - R_{420} \quad (24)$$

2.5. Obtaining feature vectors

The following methods have been used to select features suitable for separating polluted areas and those with lower pollution by shape indices and spectral indices of mulberry leaves:
- ✓ Sub-feature method with comparable predictive power SFCPP [27];
- ✓ Features Selection Method by Adjacent Component Analysis, FSNCA [12];
- ✓ Method for ranking significant prediction parameters, RELIEFF [26];
- ✓ Method for selecting regression traits by neighbor component analysis, FSRNCA [14];

Vectors are composed of a total of 12 features, consisting of 4 features of the spectral indices of the adaxial part, 4 features of the abaxial patr of the leaves, and 4 indices of their shape.

2.6. Reducing the amount of data of feature vectors

To reduce the amount of data of feature vectors have been used [4,19,20,24]:
- ✓ Latent variables (LV) obtained by the method of partial least squares regression;
- ✓ Linear variant of principal components obtained by Principal Component Analysis (PCA) method;
- ✓ Kernel PCA method (kPCA). Two Simple and Polynomial kernel functions are used.

The use of a nonlinear variant of the PCA method is dictated by the fact that the spectral characteristics of mulberry leaves collected from areas with different degrees of contamination overlap.
Software tools described by Wang [30] and published on the Mathworks page as executable code [31] were used.
The PCA method was also used to determine the correspondence between the amount of data reduction method and the feature vectors. Principal components are calculated by rows and columns of the table with a general classification error and the results are presented graphically.

2.7. Classification methods

The Naïve Bayes classifier was used as the reference [1,16,28]. One of the classic algorithms in machine learning is the Naïve Bayes Classifier, which is based on the Bayes theorem for determining the posterior probability of an event occurring. The purpose of the classification is to determine to which class the object x belongs. Therefore, it is necessary to find the probability class of the object x, i.e. it is necessary for all classes to select the one that gives the maximum probability P(y=c|x).
Discriminant Analysis [11] is a multidimensional data analysis that is used when it is necessary to predict the values of a grouping variable. This is known as classification or pattern recognition.
The following separation (discriminant) functions were used in the discriminant analysis:
- ✓ Linear (L) - linear separation function, suitable for multivariate normal density data of each group, with a total covariance estimate;

- ✓ Quadratic (Q) - quadratic separator function (second degree), distributes data with multivariate normal density by calculating covariance and grouping them;
- ✓ Diagquadratic (DQ) - similar to the quadratic separating function but using the calculation of the diagonal of a covariance matrix (diagonal nonlinear separating function);
- ✓ Mahalanobis (M) - splits data into groups by Mahalonobis distance by determining covariance in the data.

The Support Vector Machines method (SVM) [7,29] uses a teacher-trained model and related algorithms to analyze the data used for classification. In a training sample, each element of the sample is associated with one of two categories, and the training algorithm builds a model in which the data is transformed into a new space. The model created presents the data in the new space in such a way that there is a separation between them.
SVM analysis uses the following separation functions:

- ✓ Linear (L) - linear separation function, suitable for multivariate normal density data of each group, with a total covariance estimate;
- ✓ Quadratic (Q) - quadratic separator function (second degree), distributes data with multivariate normal density by calculating covariance and grouping them;
- ✓ Polinomial - polynomial separating function;
- ✓ RBF - separating function defined by radial basis elements.

The performance of the classifiers used is estimated by a general classification error, which is described by the formula:

$$e = \frac{\sum_{i=1}^{n}(\sum_{k=1}^{n} y_{ik} - y_{ii})}{\sum_{i=1}^{n}\sum_{k=1}^{n} y_{ik}}.100,\% \quad (25)$$

where y_{ik} is the number of samples in class i classified by the classifier in class k; y_{ii}) number of correctly recognized samples; k = 1...n - number incorrectly assigned to a class i relative to the total number of samples; n - number of classes.
All data were processed at a level of significance α=0,05.

3. RESULTS AND DISCUSSION

Table 2 shows the selected signs for the classification of high and low pollution areas. The names of the vectors, the method of selection used, the features obtained from the spectral characteristics of the adaxial part of the leaves and those of the abaxial part of the leaves are presented. From the shape features are selected perimeter, area of bounding box, area, ideal area and some of the coefficients, that describe ratio of the different types of leaf areas. It can be seen from the spectral indices that those that influence the spectrum in the 400-550 nm range are selected. In the vectors of signs FV1, FV2, FV4, the same features for the adaxial and abaxial parts of the mulberry leaves are selected, while for the FV3 obtained by the RELIEFF method, the abaxial part is selected for G, instead of TVI, as for the adaxial part of the leaves.

Table 2. Selected feature vectors of mulberry leaves

Feature vector	Method	Spectral indices adaxial part	Spectral indices abaxial part	Shape indices
FV1	SFCPP	REI, PTI, TVI, G	REI, PTI, TVI, G	P, Amr, KA, KAM
FV2	FSNCA	CTI, TVI, G, VARI	CTI, TVI, G, VARI	A, Ai, Amr, Kv
FV3	RELIEFF	REI, ExG, PTI, TVI	REI, PTI, ExG, G	K1, d, K2, Kf
FV4	FSRNCA	CTI, TVI, G, VARI	CTI, TVI, G, VARI	A, P, Amr, Kv

Figure 2 shows reduced data from an FV1 feature vector by latent variables, linear and nonlinear variants of principal components. Using reduced data from this feature vector, separation is observed between the two leaf types collected from zones of high and low air pollution. In this case, even when using latent variables and linear variants of the principal components, the separability of the data is evident. The data obtained by these methods are close to the original data. For kernel variants of principal components, a noticeably better separation between polluted and less polluted areas is observed using a polynomial kernel, compared to all other methods.

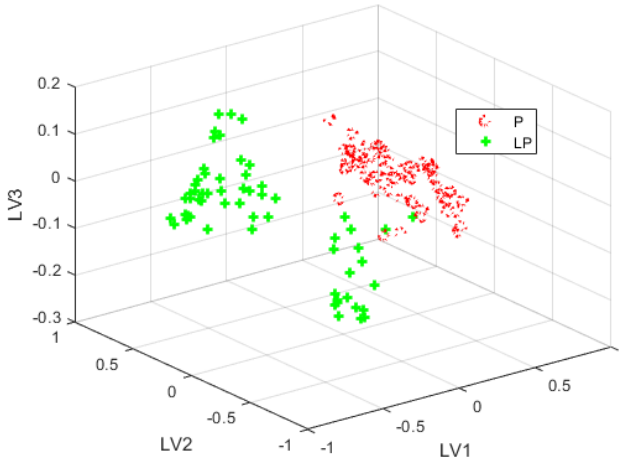


Figure 2 a) PLS

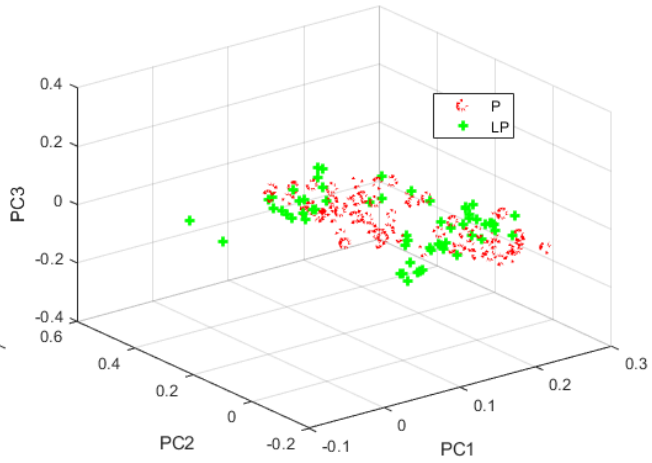


Figure 2 b) PCA

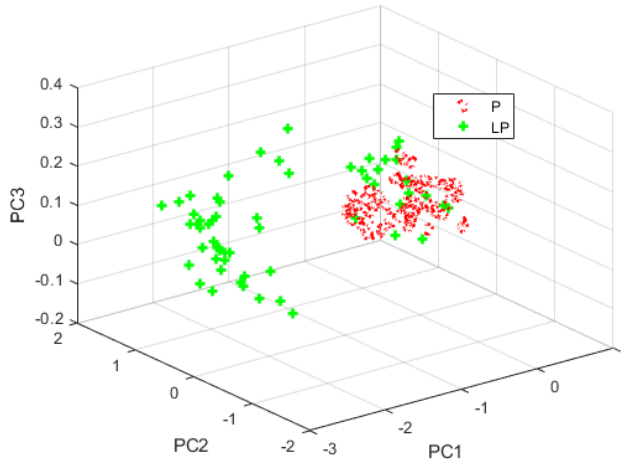


Figure 2 c) kPCA simple

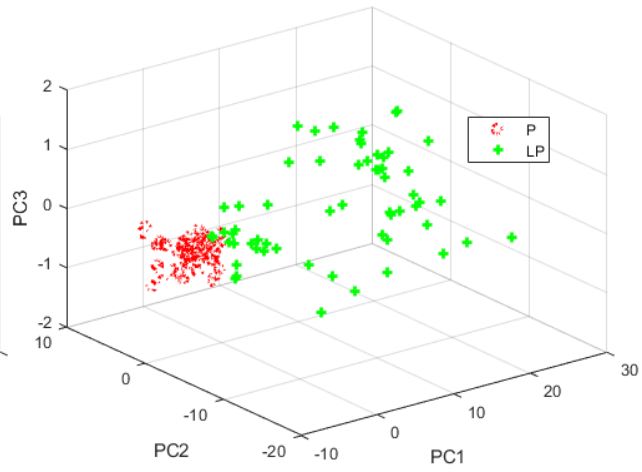


Figure 2 d) kPCA polynomial

Figure 2.
Reduced data of feature vector FV1

Figure 3 shows reduced data from an FV2 feature vector by latent variables, linear and nonlinear variants of principal components. No separation was observed between the two leaf types collected from zones of high and low air pollution using reduced data from this feature vector. Regardless of the reduction method used, data overlap is observed.

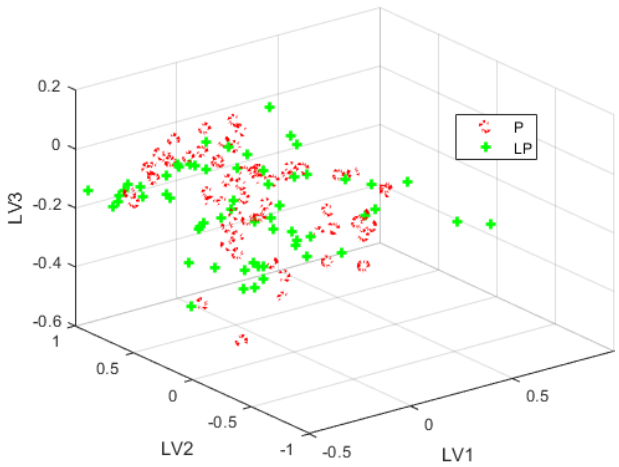


Figure 3 a) PLS

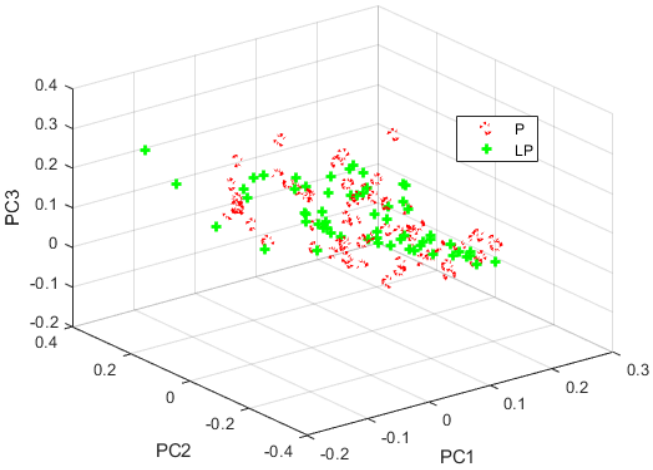


Figure 3 b) PCA

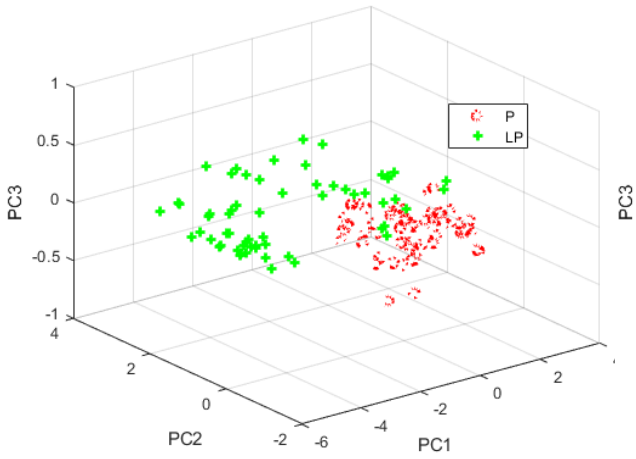


Figure 3 c) kPCA simple

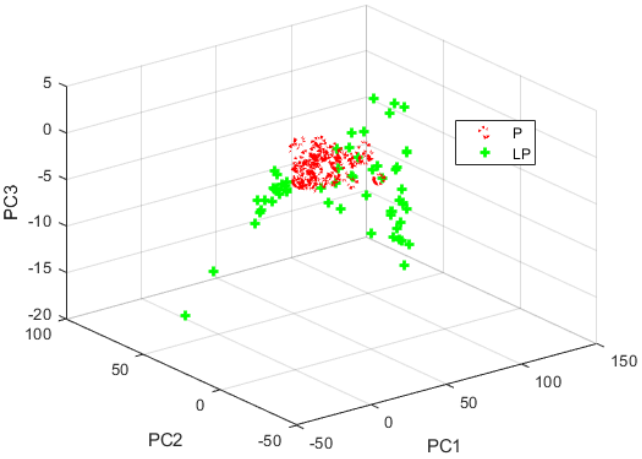


Figure 3 d) kPCA polynomial

Figure 3.
Reduced data of feature vector FV2

Figure 4 shows reduced data from an FV3 feature vector by latent variables, linear and nonlinear variants of principal components. No separation was observed between the two leaf types collected from zones of high and low air pollution using reduced data from this feature vector. Regardless of the reduction method used, data overlap is observed.

ARTTE
Applied Researches in Technics, Technologies and Education
Journal of the Faculty of Technics and Technologies, Trakia University
https://sites.google.com/a/trakia-uni.bg/artte/

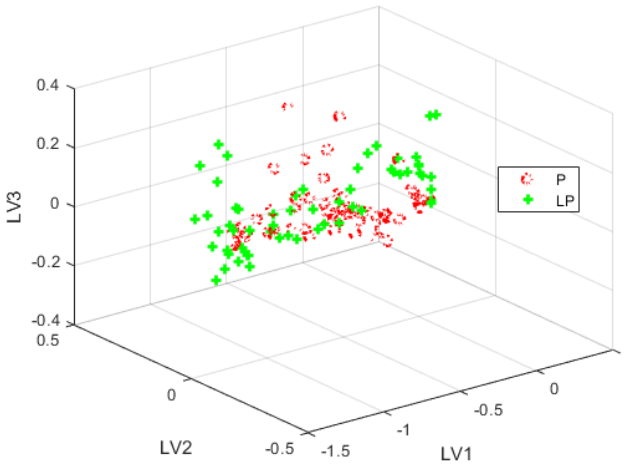


Figure 4 a) PLS

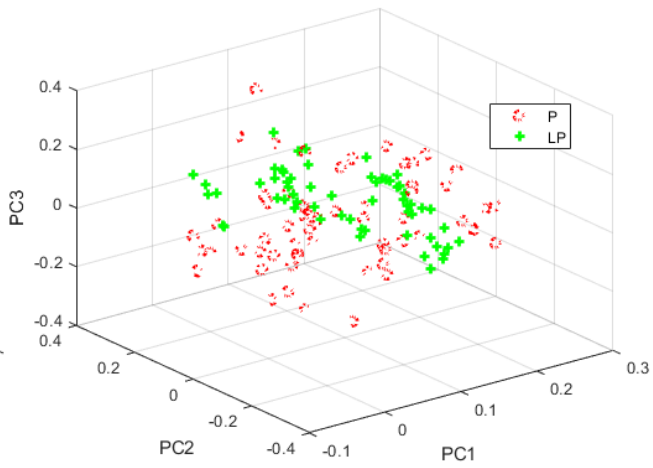


Figure 4 b) PCA

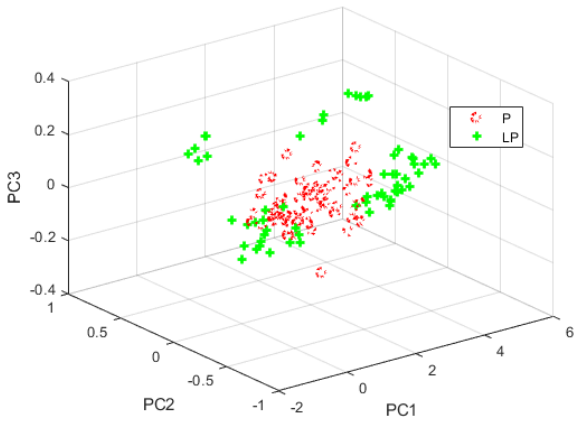


Figure 4 c) kPCA simple

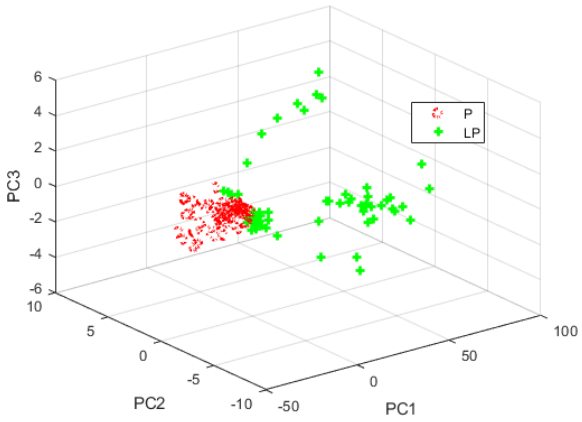


Figure 4 d) kPCA polynomial

Figure 4.
Reduced data of feature vector FV3

Figure 5 shows reduced data from an FV4 feature vector by latent variables, linear and nonlinear variants of principal components. No separation was observed between the two leaf types collected from zones of high and low air pollution using reduced data from this feature vector. Regardless of the reduction method used, data overlap is observed. In this case, the kernel variant of PCA using Simple kernel has a better separation between polluted and less polluted zones than other methods.
A classification was made with the Naïve Bayes classifier based on reduced data from the four feature vectors.

ARTTE
Applied Researches in Technics, Technologies and Education
Journal of the Faculty of Technics and Technologies, Trakia University
https://sites.google.com/a/trakia-uni.bg/artte/

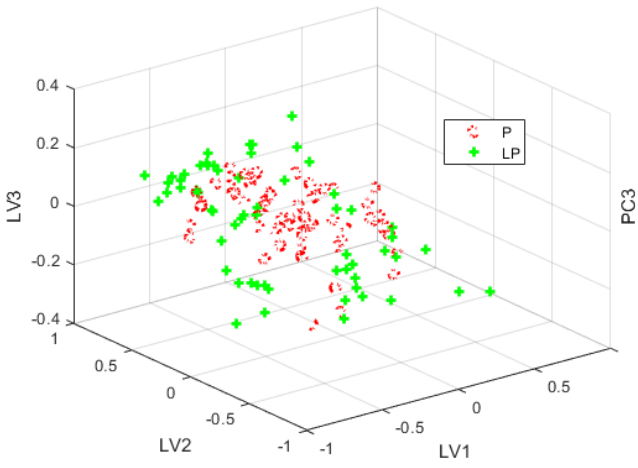


Figure 5 a) PLS

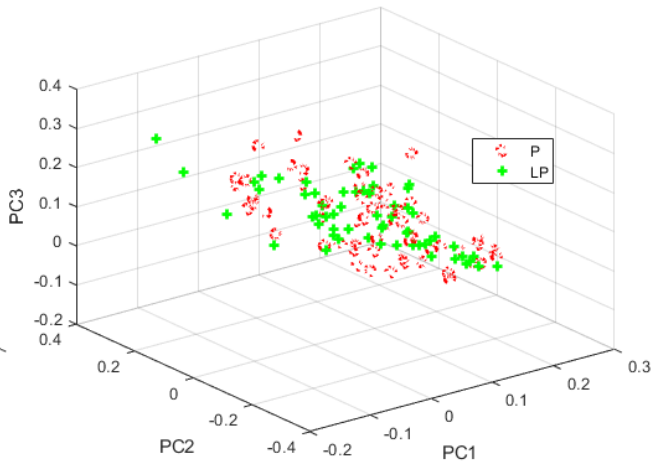


Figure 5 b) PCA

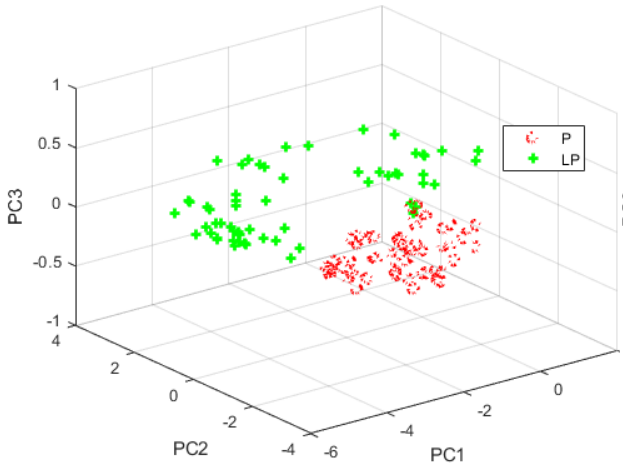


Figure 5 c) kPCA simple

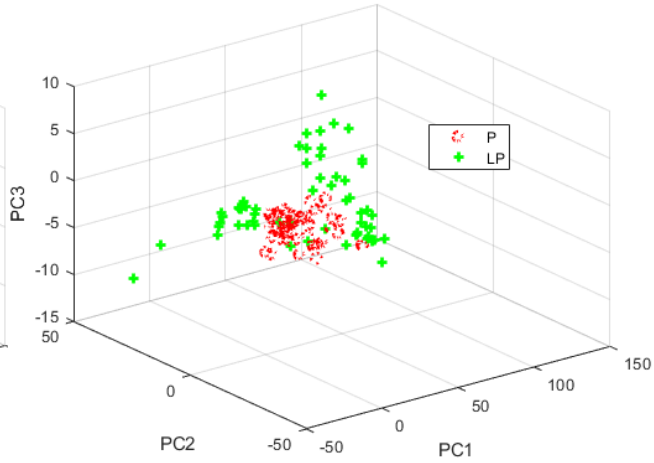


Figure 5 d) kPCA polynomial

Figure 5.
Reduced data of feature vector FV4

Figure 6 shows generally the results of a classification with a Naive Base classifier using a vector of FV1 traits. It can be seen that in all cases there is an overlap between the two classes (class 1 - P) and (class 2 - LP). The largest overlap was obtained using a linear PCA variant. In all other methods, the overlap is low in comparison to other methods used.

ARTTE
Applied Researches in Technics, Technologies and Education
Journal of the Faculty of Technics and Technologies, Trakia University
https://sites.google.com/a/trakia-uni.bg/artte/

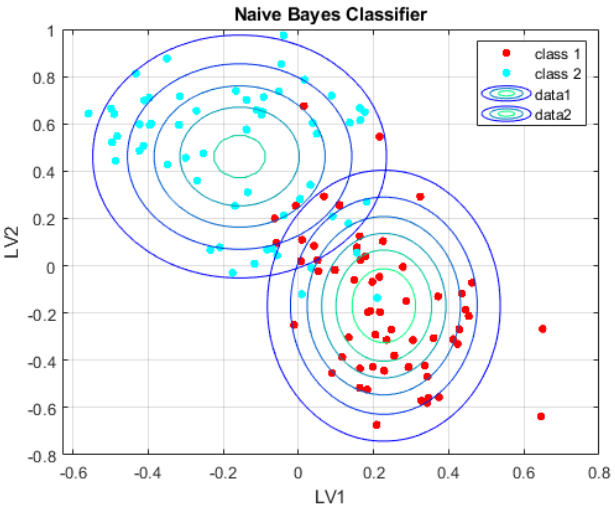


Figure 6 a) PLS

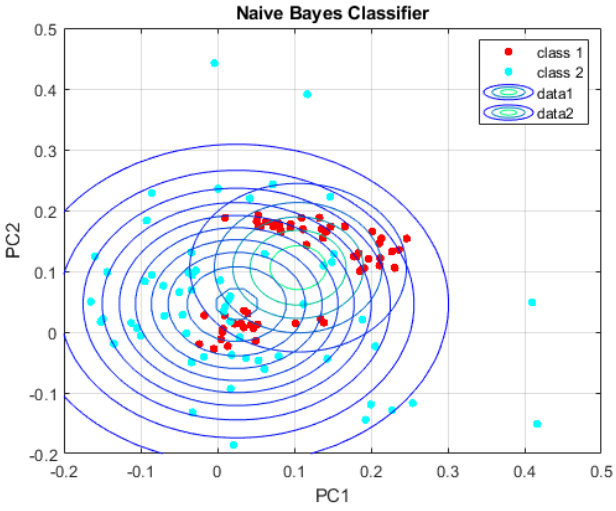


Figure 6 b) PCA

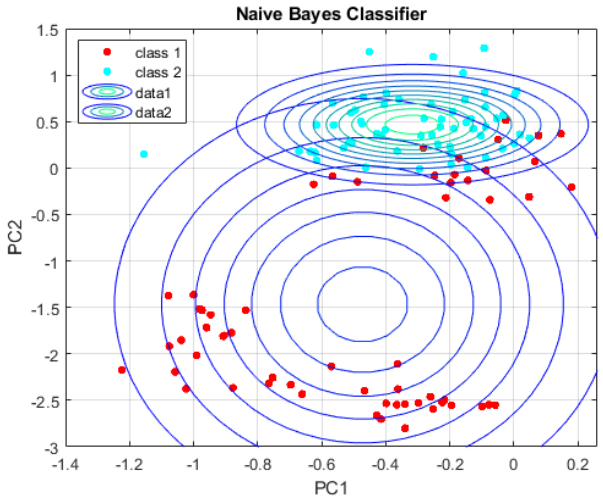


Figure 6 c) kPCA simple

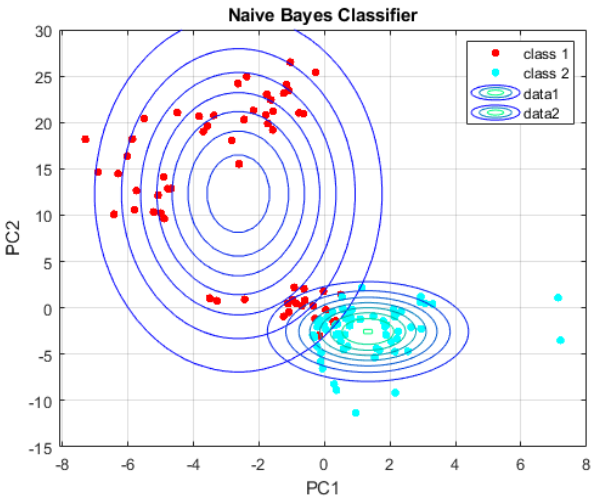


Figure 6 d) kPCA polynomial

Figure 6.
Classification with Naïve Bayes classifier by FV1 Reduced Data

Figure 7 shows generally the results of a classification with a Naïve Bayes classifier using an FV2 feature vector. It can be seen that in all cases there is an overlap between the two classes (class 1 - P) and (class 2 - LP). The largest overlap was obtained using latent variables and a linear variant of PCA. In the PCA method using kernel functions, the overlap is low.

ARTTE
Applied Researches in Technics, Technologies and Education
Journal of the Faculty of Technics and Technologies, Trakia University
https://sites.google.com/a/trakia-uni.bg/artte/

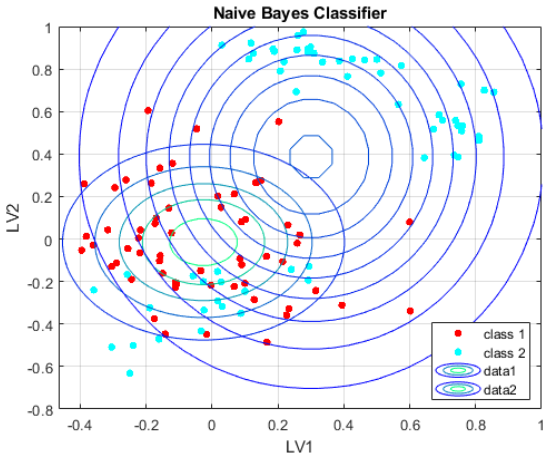


Figure 7 a) PLS

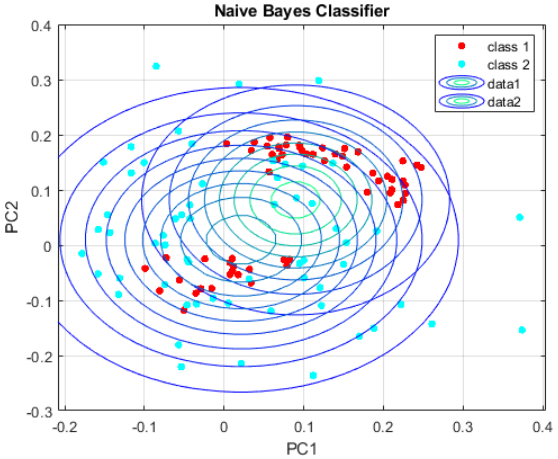


Figure 7 b) PCA

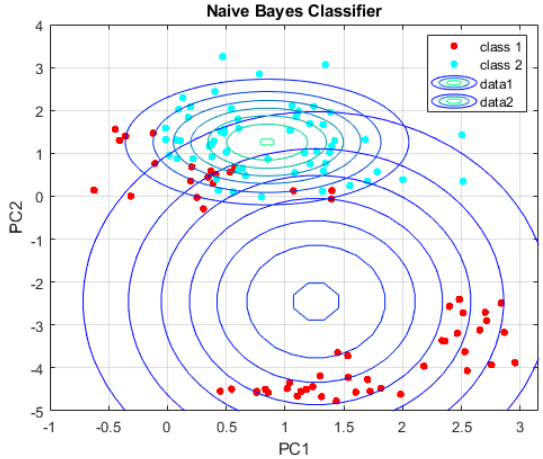


Figure 7 c) kPCA simple

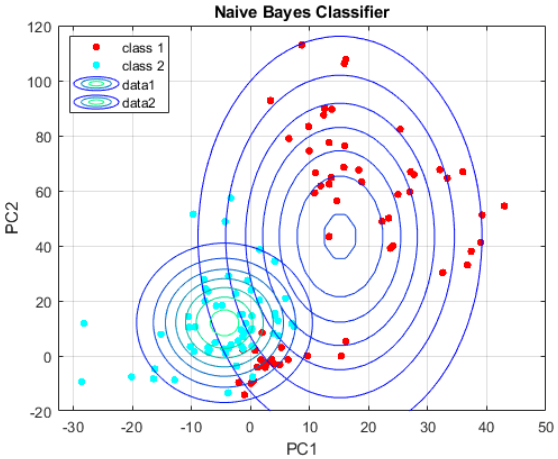


Figure 7 d) kPCA polynomial

Figure 7. Classification with Naïve Bayes classifier by FV2 Reduced Data

Figure 8 shows generally the results of a classification with a Naïve Bayes classifier using the FV3 feature vector. It can be seen that in all cases there is an overlap between the two classes (class 1 - P) and (class 2 - LP). The largest overlap was obtained using a linear PCA variant. In all other methods, the overlap is lower in comparison to other methods used.

ARTTE
Applied Researches in Technics, Technologies and Education
Journal of the Faculty of Technics and Technologies, Trakia University
https://sites.google.com/a/trakia-uni.bg/artte/

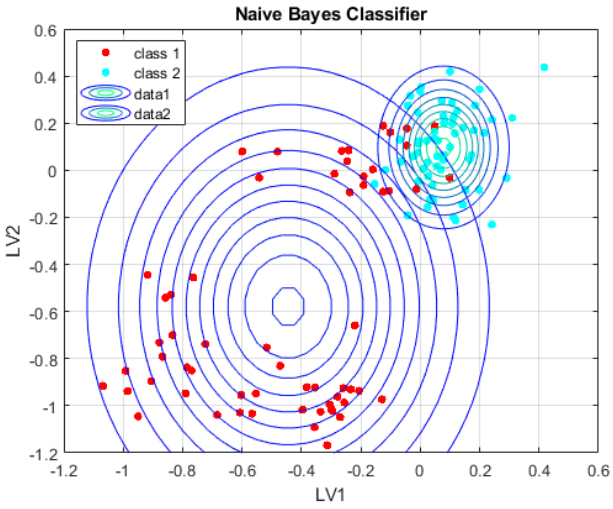


Figure 8 a) PLS

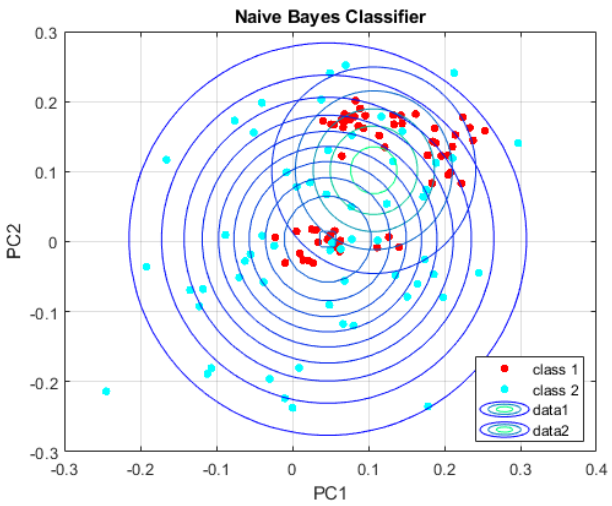


Figure 8 b) PCA

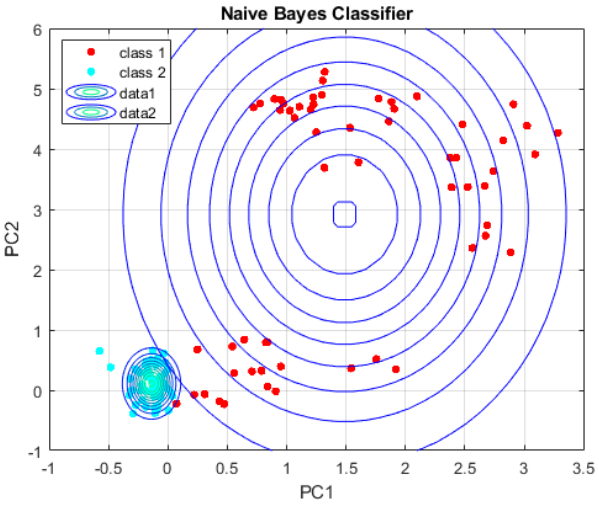


Figure 8 c) kPCA simple

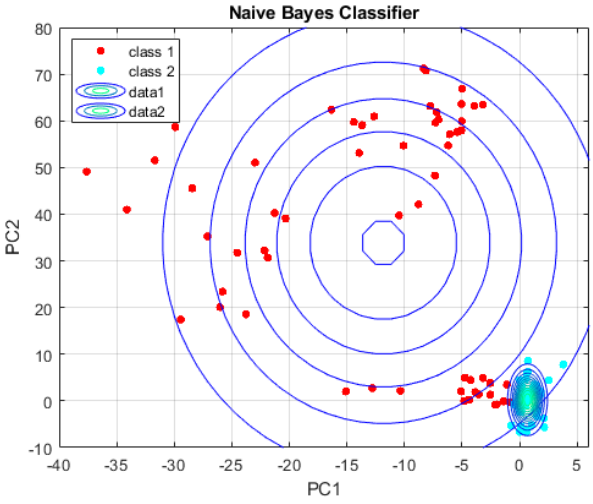


Figure 8 d) kPCA polynomial

Figure 8.
Classification with Naïve Bayes classifier by FV3 Reduced Data

Figure 9 shows generally the results of a classification with a Naïve Bayes classifier using an FV4 feature vector. It can be seen that in all cases there is an overlap between the two classes (class 1 - P) and (class 2 - LP). The largest overlap was obtained using latent variables and a linear variant of PCA. In the PCA method using kernel functions, the overlap is low in comparison to other methods used.

ARTTE
Applied Researches in Technics, Technologies and Education
Journal of the Faculty of Technics and Technologies, Trakia University
https://sites.google.com/a/trakia-uni.bg/artte/

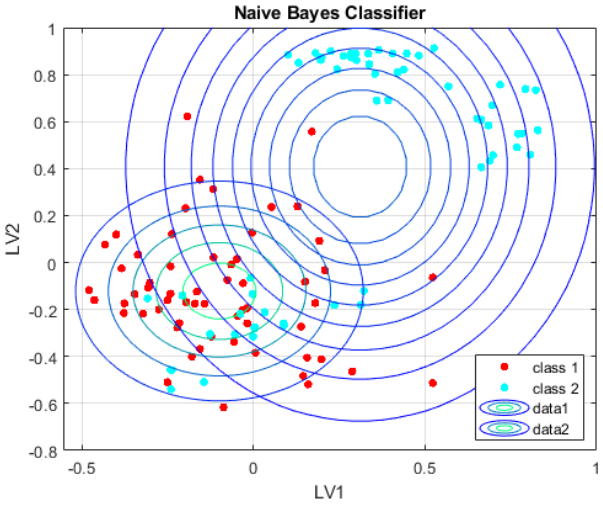


Figure 9 a) PLS

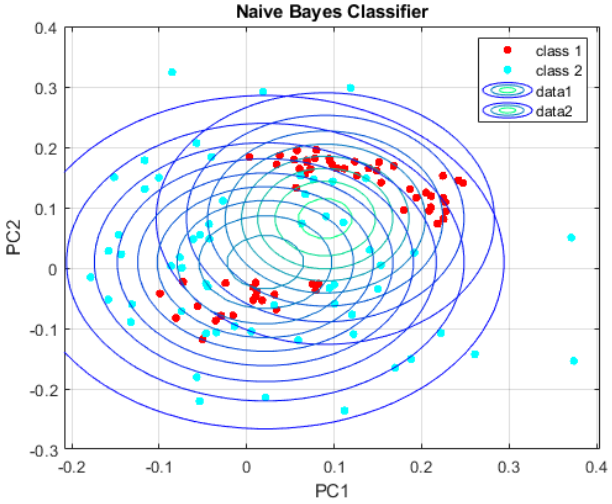


Figure 9 b) PCA

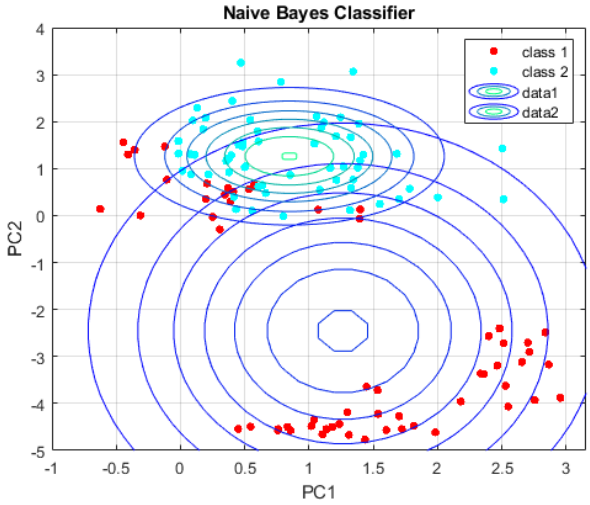


Figure 9 c) kPCA simple

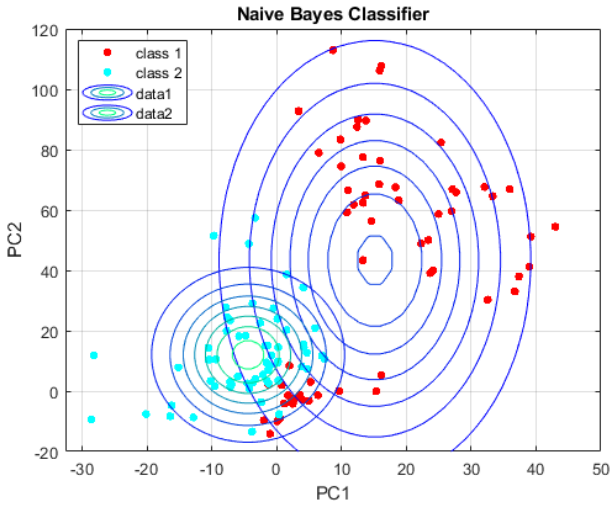


Figure 9 d) kPCA polynomial

Figure 9.
Classification with Naïve Bayes classifier by FV4 Reduced Data

Table 3 lists the values of the general classification error using the Naïve Bayes classifier. It can be seen that the highest values of this error are obtained using the linear variant of the principal components (e=23-32%). In terms of feature vectors, using FV3 results in a general classification error of up to 10%.

ARTTE
Applied Researches in Technics, Technologies and Education
Journal of the Faculty of Technics and Technologies, Trakia University
https://sites.google.com/a/trakia-uni.bg/artte/

Table 3. Classification results with a Naïve Bayes Classifier

Method / Feature vector	LV	PCA	kPCA Simple	kPCA Polynomial
FV1	14%	23%	14%	14%
FV2	22%	30%	19%	18%
FV3	10%	32%	4%	6%
FV4	20%	29%	17%	18%
LV-latent variable; kPCA-kernel principal component analysis; FV-feature vector				

Figure 10 shows the results of principal component analysis, which depicts the dependences between the reduction method and the feature vector obtained depending on the general classification error of the Naïve Bayes classifier. The results shown in the graph confirm the results obtained so far. The PCA method is farthest from all feature vectors, indicating that the data obtained with it is not suitable for classification. FV2 and FV4 have the same errors when using latent variables (LV) and are therefore spaced the same distance from this method. From the previous results, FV3 is also suitable for classification, after reducing the amount of its data by the kernel variant of the principal components. This vector is suitable regardless of the kernel function used.

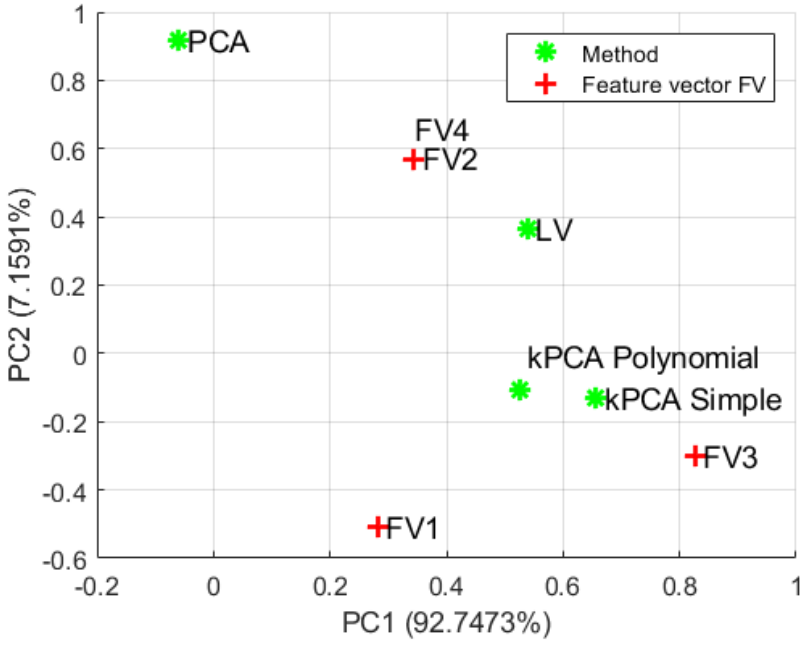


Figure 10. Principal component analysis of classification errors from Naïve Bayes Classifier

From the analyzes made, it has been found that suitable for classification of areas with high and low air pollution of the habitat of mulberry, by shape indices and spectral indices of leaves is suitable feature vector FV3, reduced by the methods of partial leas squires and kernel variant of the principal components. Discriminant analysis and Support vector machines method were used for classification.

Table 4 lists the results for a general classification error using discriminant analysis. It can be seen that classification errors below 10% are obtained using nonlinear separation functions. The lowest general classification error values are obtained from the reduced FV3 data by the kPCA method and the Simple kernel function.

Table 4. Discriminant analysis classification results

Discriminant function / Method	L	Q	DQ	M
LV	13%	9%	8%	11%
kPCA Simple	13%	3%	3%	5%
kPCA Polynomial	15%	5%	6%	13%
L-Linear; Q-quadratic; DQ-diagonal quadratic; M-Mahalanobis; LV-latent variable; kPCA-kernel principal component analysis				

Figure 11 shows the results of principal component analysis by which the correspondences between the data reduction method and the feature vector obtained depending on the general classification error in the Discriminant analysis are visualized. From the results of this analysis, it can be argued that a suitable FV3 vector reduced by the kernel variant of PCA is suitable for separating areas of high and lower air pollution by data from shape indices and spectral indices of mulberry leaves.

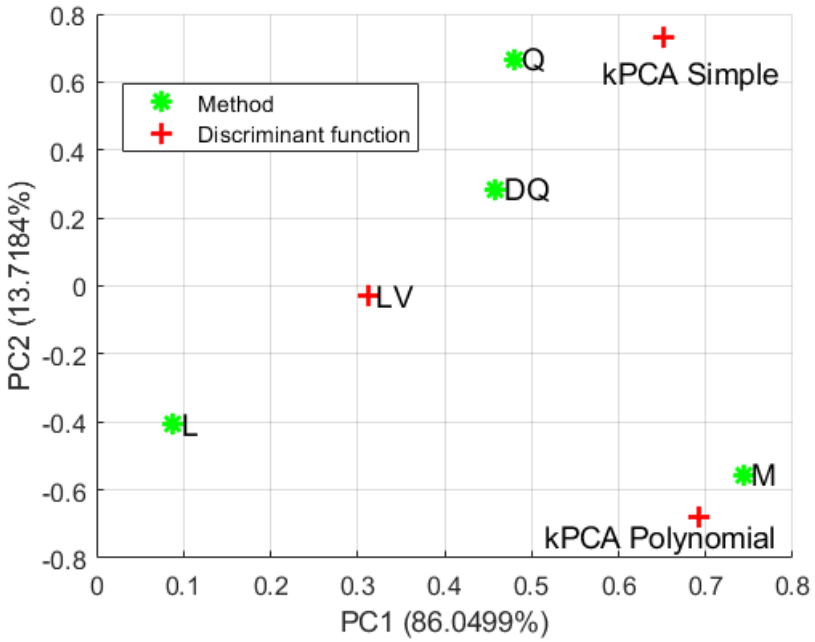


Figure 11.
Principal component analysis of classification errors from Discriminant analysis

Figure 12 shows the operation of the discriminant classifier with the reduced FV3 data using kPCA method and the Simple kernel function. It appears that the use of a linear discriminant function is inappropriate because most of the data in class P fall into class LP. This problem is solved by the quadratic discriminant function and the Mahalanobis function, in which the data of the two classes are separated more precisely.

ARTTE
Applied Researches in Technics, Technologies and Education
Journal of the Faculty of Technics and Technologies, Trakia University
https://sites.google.com/a/trakia-uni.bg/artte/

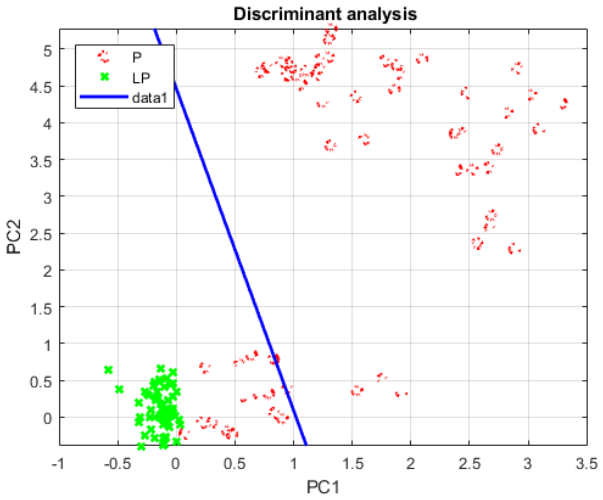


Figure 12 a) Linear

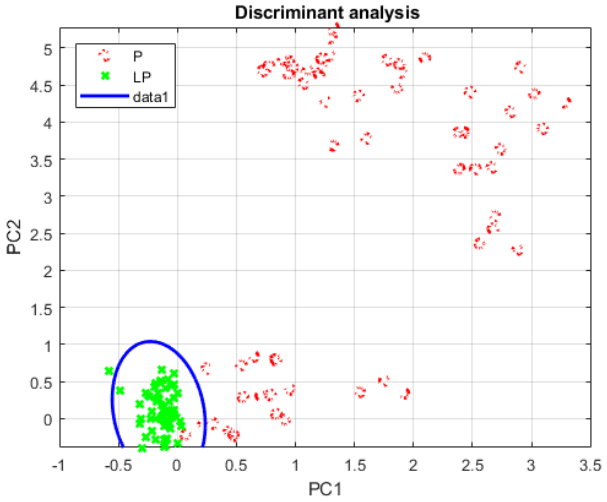


Figure 12 b) Quadratic

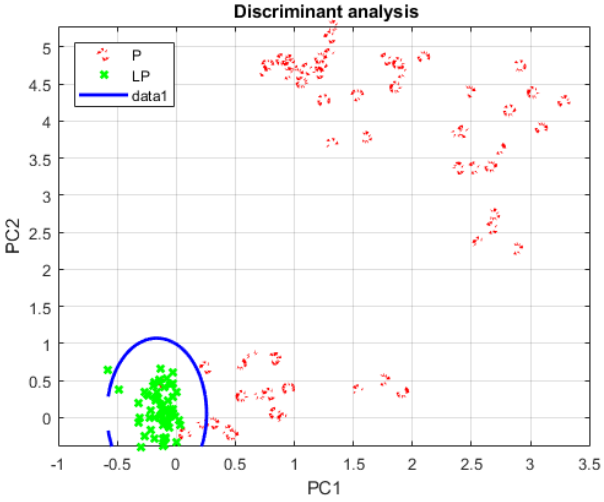


Figure 12 c) DiagQuadratic

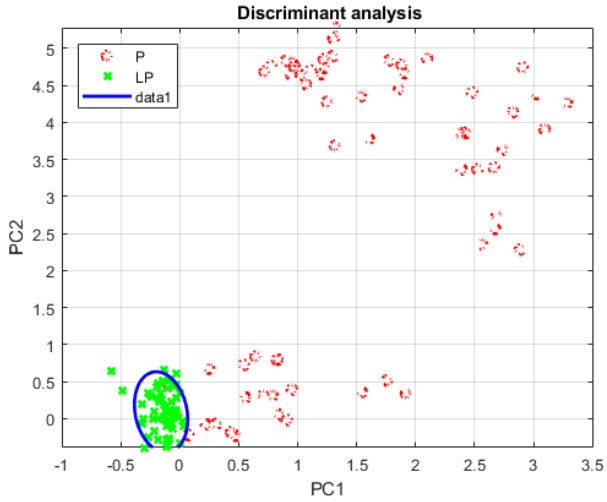


Figure 12 d) Mahalanobis

Figure 12.
Discriminant Classification by FV4 Reduced Data Using kPCA Simple Method

Table 5 lists the results for a general classification error using the Support Vector Machines method. It can be seen that using latent variables produces a total classification error of e=1%. In all other cases, the general classification error is e=0%, regardless of the method used to reduce the amount of FV3 data and the separating function of the classifier.

ARTTE
Applied Researches in Technics, Technologies and Education
Journal of the Faculty of Technics and Technologies, Trakia University
https://sites.google.com/a/trakia-uni.bg/artte/

Table 5. Results of the classification by Support vector machines

Separating function / Method	L	Q	Polynomial	RBF
LV	1%	1%	0%	1%
kPCA Simple	0%	0%	0%	0%
kPCA Polynomial	0%	0%	0%	0%

L-Linear; Q-quadratic, RBF-radial basis function; LV-latent variable; kPCA-kernel principal component analysis

Figure 13 shows the results of principal component analysis, by which the correspondences between the reduction method and the feature vector obtained depending on the general classification error in the Support Vector Machines method are visualized. From the results of this analysis, it can be argued that a suitable FV3 vector reduced by the kernel variant of PCA is suitable for separating areas of high and lower air pollution by data from the shape indices and spectral indices of mulberry leaves. Better results are obtained by Principal components of kPCA Simple, when combined with the polynomial separating function of the classifier in comparison to other methods used.

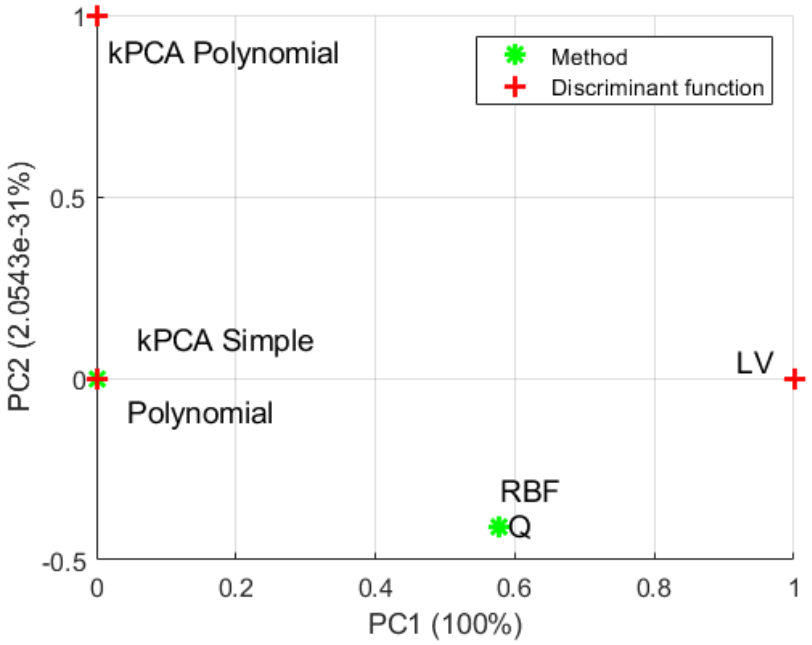


Figure 13.
Principal component analysis of classification errors from Support vector machines

Figure 14 shows the operation of an SVM classifier with reduced FV3 data, using the kPCA method and the Simple kernel function. It appears that the use of a linear separating function is inappropriate because most of the data in class P fall into class LP. This problem is solved by nonlinear separating functions, in which the data of the two classes are further separated.

ARTTE
Applied Researches in Technics, Technologies and Education
Journal of the Faculty of Technics and Technologies, Trakia University
https://sites.google.com/a/trakia-uni.bg/artte/

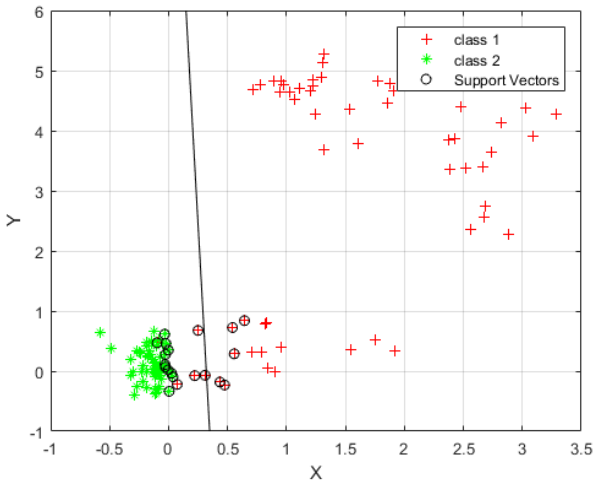


Figure 14 a) Linear

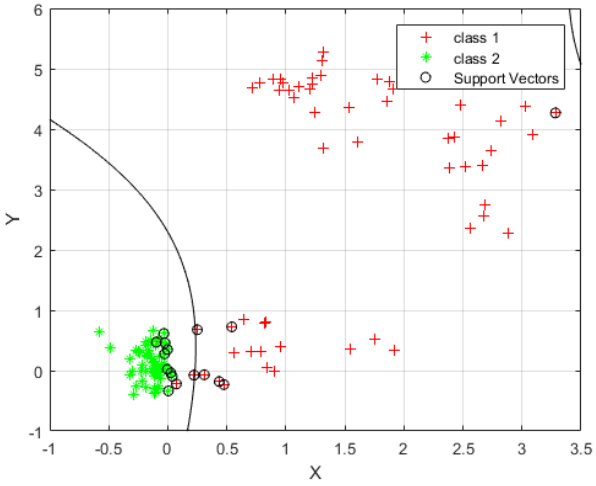


Figure 14 b) Quadratic

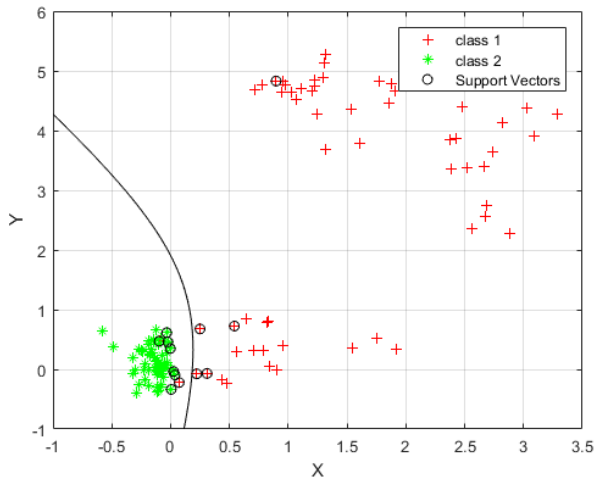


Figure 14 c) Polynomial

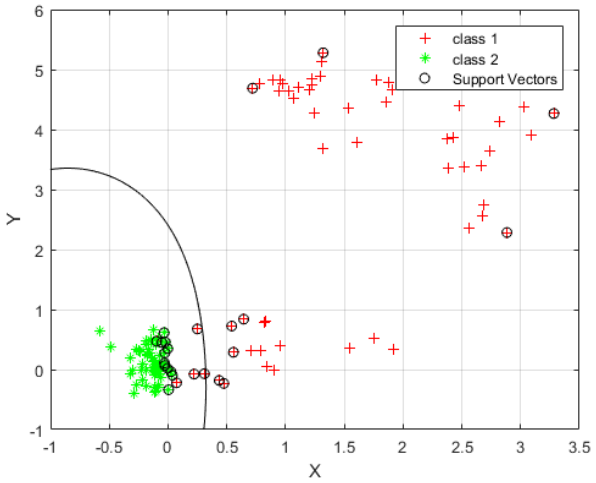


Figure 14 d) RBF

Figure 14. SVM classification based on FV4 reduced data, using method kPCA Simple

The results obtained from classification by discriminant analysis and the method of Support vector machines indicate that the use of latent variables and linear variants of principal components to reduce the amount of data from feature vectors containing spectral indices and mulberry leaf shape indices is not appropriate for solving the problem of separating areas with high and low air pollution.

The data obtained confirm and complement those reported by Jun et al. [17], who use spectral characteristics to predict pesticide content in mulberry leaves. A correct prediction of 87% accuracy requires a more sophisticated method of analysis, such as regression of the support vectors.

Direct use of spectral and shape indices is not appropriate. For this reason, in the present work, the separation between leaves collected from polluted and less polluted zones is obtained after applying a kernel variant of the principal components, combined with nonlinear separating functions of classifiers.

According to the results obtained here, it may be recommended to use complex methods of analysis to evaluate changes in mulberry leaves depending on the contamination of the habitat area of the plant, replacing laboratory measurements [23] with methods suitable for remote monitoring and mobile application creation.

4. CONCLUSION

The results of this study indicate that the surface and shape characteristics of mulberry leaves have the potential to be used as parameters for assessing urban environmental quality.
An approach for recognizing areas with high and low pollution based on surface and shape characteristics of mulberry leaves, based on extracted features and classification with three types of classifiers, has been adapted.
Suitable for distinguishing areas with high and low pollution has been found to be a feature vector containing data from the spectral indices REI, ExG, PTI, TVI for the adaxial part of the leaves and REI, PTI, ExG, G for their abaxial part. Of the features of the form, the vector includes K1, d, K2, Kf.
After comparative analyzes, it was found that the resulting feature vector can be used to distinguish between zones with high and low pollution, after applying a kernel variant of the principal components, combined with nonlinear separating functions of discriminant analysis and the method of reference vectors. In this variant, the total classification error is 0-3%.
The use of methods to reduce the amount of data of feature vectors by which values close to the original data are obtained is not appropriate. Proof of this is the results after applying the latent variable methods and the linear variant of the principal components. When used, the total classification error reaches 22-32%. These results improve and complement those reported in the available literature. They can be used to refine the approaches and methods used so far to passively determine the degree of contamination in the area of the mulberry habitat.
The results obtained could be used as preliminary baseline data for future evaluations and studies related to remote monitoring of urban air quality.

5. ACKNOWLEDGEMENTS

The work in this article is supported by the project 3.FTT/2018 "Evaluation of the ecological purity of food raw materials and products", headed by Snejana Dineva.

6. REFERENCES

[1] Arabadzieva-Kalcheva, N., Nikolov, N. (2017). Comparative analysis of the naive bayes classifier and sequential minimal optimization for classifying text in bulgarian in machine learning, Computer science and technologies, year XV, No.1, ISSN 1312-3335, pp. 97-105 (in Bulgarian).
[2] Atanassova, S., Nikolov, P., Valchev, N., Masheva, S., Yorgov, D. (2019). Early detection of powdery mildew (Podosphaera xanthii) on cucumber leaves based on visible and near-infrared spectroscopy. AIP Conference Proceedings, 2075, 160014-1-160014-5.
[3] Azadi, M., Doley, D. (2004). Biological indicators of air quality in Brisbane, Australia. International Journal of Environmental Science & Technology, Vol. 1, No. 1, pp. 59-68.
[4] Baycheva, S. (2016). Application of devices of measurement of colour in analysis of food products. Journal of Innovation and entrepreneurship, Vol. 4, No. 4, pp. 43-59.

[5] Bhosle, K., Musande, V. (2017). Stress monitoring of mulberry plants by finding rep using hyperspectral data. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-1/W1, 2017 ISPRS Hannover Workshop: HRIGI 17 – CMRT 17 – ISA 17 – EuroCOW 17, 6–9 June 2017, Hannover, Germany, pp. 383-386.
[6] Cermakova, I., Komarkova, J., Sedlak, P. (2019). Calculation of Visible Spectral Indices from UAV-Based Data: Small Water Bodies Monitoring. In 2019 14th Iberian Conference on Information Systems and Technologies CISTI, 1-5.
[7] Classify using support vector machine (SVM), [online]. Available at: https://www.mathworks.com/help/stats/svmclassify.html (Accessed: 12.05.2019).
[8] Colour & Vision Research laboratory and database, http://cvrl.ioo.ucl.ac.uk/ (available on 13.10.2018).
[9] Dimov, I., Petkova, N., Nakov, G., Taneva, I., Ivanov, I., Stamatovska, V. (2018). Improvement of antioxidant potential of wheat flours and breads by addition of medicinal plants. Ukrainian Food Journal, Vol. 7, No. 4, pp. 671-681.
[10] Dimov, M. (2012). Contemporary method for utilization of rubber wastes. Management and education, Vol. 8, No. 4, pp. 173-176.
[11] Discriminant analysis, [online]. Available at: https://www.mathworks.com/help/stats/classify.html (Accessed: 06.03.2019).
[12] El-Khatib, A., El-Shanawany, A., E., El-Amery, E. (2016). Urban tree leaf as bio-indicator for air pollution around superphosphate fertilizers plant, upper Egypt. Journal of Ecology of Health & Environment, Vol. 4, No. 2, pp. 95-101.
[13] FSNCA, Feature selection using neighborhood component analysis for classification, [online]. Available at: https://www.mathworks.com/help/stats/fscnca.html (Accessed: 27.08.2019).
[14] FSRNCA, Feature selection using neighborhood component analysis for regression, [online]. Available at: https://www.mathworks.com/help/stats/fsrnca.html (Accessed: 27.08.2019).
[15] Glassner, A. (1989). How to derive a spectrum from an RGB triplet, IEEE Computer Graphics and Applications, Vol. 9, No. 4 (July 1989), pp. 95-99.
[16] Gore, R., Deshpande, D. (2017). An approach for classification of health risks based on air quality levels. 1st International Conference on Intelligent Systems and Information Management (ICISIM), pp. 58-61.
[17] Jun, S., Shuying, J., Meixia, Z., Hanping, M., Xiaohong, W., Qinglin, L. (2016). Detection of pesticide residues in mulberey leaves using vis-nir hyperspectral imaging technology. Journal of Residuals Science & Technology, Vol. 13, No. 1, pp. S125-S131.
[18] Kang, G., Gao, J., Chiao, S., Lu, S., Xie, G. (2018). Air quality prediction: big data and machine learning approaches. International Journal of Environmental Science and Development, Vol. 9, No. 1, pp. 8-16.
[19] Kazlacheva, Z., Dineva, P. (2017). An Investigation of Pattern Making of Twisted Draperies. ARTTE Applied Researches in Technics, Technologies and Education, Vol. 5, No. 2, 2017, pp. 85-93.
[20] Kazlacheva Z., Ilieva J. (2018). An investigation of Design of Twist Knot Drape Clothes. The Aegean International Textile and Advanced Engineering Conference AITAE 2018, 5-7 September 2018, Lesvos, Greece, IOP Conf. Ser.: Mater. Sci. Eng. 459 (2018) 012079.
[21] Kirilova, E., Daskalov, P., Georgieva, Ts., Tzonev, R. (2013). Recognition and grading of sound and Fusarium damaged corn seeds of different varieties using prototype system based on machine vision. Information Communication and Control Systems and Technologies, Vol. 2, No. 1, pp.43-49.

[22] Mather, J. (2010). Spectral and XYZ Color Functions, [online]. Available at: www.mathworks.com (Accessed 14.06.2018).
[23] Mezghani, I., Zouari, M., Rouina, B., Abdallah, F. (2019). Mulberry leaves as a bioindicator of fluoride pollution in the vicinity of a phosphate fertilizer factory located in Sfax, Tunisia. Research report Fluoride, Vol. 52, No.4, pp. 537-545.
[24] Mladenov, M. (2015). Complex assessment of the quality of food products through visual images, spectrophotometric and image analysis hyper-spectral characteristics. Monograph. University publishing Center of Rousse University "A. Kanchev", Rousse, 2015. (in Bulgarian).
[25] Nakov, G., Jankuloska, V., Georgieva-Nikolova, M. (2019). Influence of food by-products addition on the spectral characteristics of bakery products. Innovation and entrepreneurship, Vol. 7, No. 3, pp.138-149.
[26] RELIEFF, Rank importance of predictors using ReliefF or RReliefF algorithm, [online]. Available at: https://www.mathworks.com/help/stats/relieff.html (Accessed: 27.08.2019).
[27] SFCPP, Select Subset of Features with Comparative Predictive Power, [online]. Available at: https://www.mathworks.com/help/stats/feature-selection.html, (Accessed: 26.08.2019).
[28] Titova T., Nachev, V., Damyanov, Ch. (2012). Non-destructive egg quality determination with intelligent classifiers. XI International SAUM Conference on Systems, Automatic Control and Measurements Niš, Serbia, November 14th-16th, 2012, pp. 451-454.
[29] Train support vector machine classifier, [online]. Available at: https://www.mathworks.com/help/stats/svmtrain.html (Accessed 03.04.2017).
[30] Wang, Q. (2011). Kernel principal component analysis and its applications in face recognition and active shape models, Pattern Recognition at RPI, Troy, NY, USA, 2011.
[31] Wang, Q., Kernel PCA and Pre-Image Reconstruction, https://www.mathworks.com/matlabcentral/fileexchange/39715-kernel-pca-and-pre-image-reconstruction (available on 13.10.2018).
[32] Zadeh, A., Veroustraete, F., Buytaert, J., Dirckx, J., Samson, R. (2013). Assessing urban habitat quality using spectral characteristics of Tilia leaves. Environmental Pollution, Vol. 178, 7-14.
[33] Zlatev, Z. (2019). Development of a system for monitoring air environment parameters. Innovation and entrepreneurship, Vol. 7, No. 1, pp. 36-48.