

Classification of different types of beer according to their colour characteristics

Kr T Nikolova¹, R Gabrova², D Boyadzhiev³, E S Pisanova³, J Ruseva² and D Yanakiev^{3*}

¹Medical University, 55 Marin Drinov Blvd., Varna, Bulgaria

²University of Food Technologies, 26 Maritsa Blvd., 4000 Plovdiv, Bulgaria

³University of Plovdiv, 24 Tzar Assen Str., 4000 Plovdiv, Bulgaria

E-mail: kr.nikolova@abv.bg

Abstract. Twenty-two samples from different beers have been investigated in two colour systems - XYZ and SIELab - and have been characterised according to their colour parameters. The goals of the current study were to conduct correlation and discriminant analysis and to find the inner relation between the studied indices. K-means cluster has been used to compare and group the tested types of beer based on their similarity. To apply the K-Cluster analysis it is required that the number of clusters be determined in advance. The variant K = 4 was worked out. The first cluster unified all bright beers, the second one contained samples with fruits, the third one contained samples with addition of lemon, the fourth unified the samples of dark beers. By applying the discriminant analysis it is possible to help selections in the establishment of the type of beer. The proposed model correctly describes the types of beer on the Bulgarian market and it can be used for determining the affiliation of the beer which is not used in obtained model. One sample has been chosen from each cluster and the digital image has been obtained. It confirms the color parameters in the color system XYZ and SIELab. These facts can be used for elaboration for express estimation of beer by color.

1. Introduction

It is known that beer is the third most often consumed drink worldwide after water and tea and the first wide consumed alcoholic drink. Beer is a complex mixture consisting mainly of water and ethanol with about 0.5% of dissolved solids [1]. The analysis of beer is very important to evaluate its organoleptic characteristics, quality, nutritional aspects and safety. Healthy ingredients are identified [2] including antioxidants [3], mainly polyphenols [4], essential vitamins – particularly B vitamins [5-6] and minerals.

A wide variety of methods have been developed and optimized to characterize and quantify different types of beer. Beers are differentiated mainly according to their visual appearance and their fermentation process. The visual appearance of beer depends on both its color and its turbidity. This is not only because scattering reduces the transparency of beer, but also because the suspended particles themselves can contribute to light absorption [7].

Depending on the yeast used and the fermentation temperature, three main beer categories can be identified: top-fermented, bottom-fermented, and naturally-fermented. The third category beer is produced mainly in Belgium. In our investigation we use bottom-fermented beers, which are also called “Lager”, are produced by adding *saccharomycesuvarum* (or *pastorianus*) at cooler temperatures, namely between 7 and 12°C.



The objective of this paper was to analyze a database of colorimetric parameters obtained for different types of beer and using K-means cluster to compare and group the tested types of beer on the base of their similarity. The defined objective requires the solution of the following problems:

- Creation of database, including the different types of beer from Bulgarian hypermarkets;
- Determination the color parameters in two different colorimetric systems, SIE Lab and XYZ;
- Establishment of statistically significant differences between the indices of beer groups and development of suitable models describing their characteristics.

2. Materials and methods

2.1. Samples

The database includes 22 types of beer from 3 countries. Five Bulgarian beers are also used for estimation of the obtained model. The samples are purchased from the Bulgarian supermarkets. The information from the investigated samples is presented in table 1.

Table 1. The beer collection.

Country	Code	Class	Alcohol, %	Country	Code	Class	Alcohol, %
Bulgaria	1	Lager	4.5	Bulgaria	13		4.3
Bulgaria	2		4.95	Serbia	14		4.7
Bulgaria	3		4.3	Holland	15	Lager	5
Bulgaria	4		5	Bulgaria	16	Dark ale	6.5
Bulgaria	5		4.8	Bulgaria	17	fresh	2.3
Serbia	6		4.6	Bulgaria	18	fresh	2.3
Bulgaria	7	Lager	5	Bulgaria	19	lemon	1.8
Bulgaria	8		2.2	Serbia	20	fresh	4.5
Holland	9		4.8	Bulgaria	21		4.8
Bulgaria	10	Dark ale	5.2	Bulgaria	22		5
Bulgaria	11	Lager	4.5				
Bulgaria	12		5.2				

2.2. Used methods

Determining the color characteristics of the beers is made by using two independent methods:- spectrophotometer and digital imaging. The spectrophotometer Thermo Scientific Helios Omega UV-VIS is used in the visible region with the cuvette of thickness of 10 mm and capacity of 1.5 ml.

To obtain a primary image of beer samples an optoelectronic system is used, containing a color CCD camera of the company Basler Vision Technologies (type Basler scA1390-17gc), a computer system with an interface to connect to the camera lighting emitting daylight (fluorescent lighting), consisting of 4 lamps with adjustable brightness; luminescent ring from Philips.

Four samples were chosen to obtain digital images – one for each group of light beer, dark beer, beer with lemon and beer with fruits. For this purpose a layer with thickness 10 mm is obtained.

Digital images of samples of beer in bmp format were obtained in a different luminance (from 940 to 1440 Lux at step 100). Images are photographed in the same conditions on standard cards of Color Checker Passport Photo, Company X-rite - card Classic Target, consisting of 24 colors with defined color characteristics and White Balance Target - 18% color gray. The deviation between the color components for each color of the cards, obtained by the image and those, specified in the passport are determined. In order to be able to obtain the best approximation of the color parameters of the beer samples to the real color parameters. A correction of the color characteristics of the images of samples is performed with the received deviations of color values of the reference colors. The correction shall be implemented in the Lab color model, which requires conversion of the RGB coordinates of the image in XYZ, and subsequent conversion into Lab coordinates.

2.3. Statistical analysis

The program “Statistica 10.0” has been used for the treatment of the data. The distribution was found to be Gaussian by the Kolmogorov-Smirnov criterion [8-9]. To establish statistically significant difference between the groups, the Scheffe criterion was applied for multiple comparisons [10]. Discriminatory analysis is used to model the group with *a priori* equal probabilities to fall into the groups [11-12]. The values for the Mahalanobis distances between the groups formed.

The K-means clustering has been applied on the base of all investigated parameters for comparing and grouping the beers [13-14]. This method classifies a given data set through a certain number of clusters fixed *a priori* [15-16]. Such an approach has been used in reference [17]. We determined the number of clusters to be 4. As a measure of similarity between samples squared Euclidean distance is used [18]. Also the cluster centers have been calculated after all objects have been assigned to a given cluster. The MATLAB™ 7.0 software has been used for determining the color characteristics of digital imagine.

3. Results and discussion

The database includes 22 types of beer from 3 different countries. For each of the samples studied, four independent measurements have been performed. The Scheffe criterion shows significant statistical differences in the studied samples. The presence of considerable difference in the color characteristics of beer provides the basis for a subsequent modeling of its origin. To model the beers by type (light, dark, fresh or with lemon) a step by step linear discriminatory analysis was used. A model with grouping parameter “beer type” was obtained and it includes the following parameters by the order of introduction into the model: color coordinate X, Z, x, chroma and hue angle. The classification of the probes according to the type is 100%.

Cluster analysis was applied with the method of K-means clusters. After some steps made with different number of clusters, most suitable was the grouping in 4 clusters. The samples, included in them had similar values of a set of the given parameters, which differentiate them from the samples in the other 3 clusters. Table 2 shows both the percentage distribution of the samples into the 4 clusters and the samples participation in each of the 4 clusters.

Table 2. Distribution of types beer in clusters.

Variants in clusters	Samples, number	Distribution of the samples of beer, %
1,2,3,4,5,6,7,9,11,12,13,14	12	55
8, 18, 20, 21	4	18
15,17, 19, 22	4	18
10, 16	2	9

The samples in the first cluster pose high lightness, greater than 92, and color parameter with a negative value. The percentage of samples is almost the same – around 18% - in the second and in the third clusters, which include 4 samples each. The beers in the second cluster have almost equal color coordinates x, y and b, have the same lightness. All of the samples are beers with fruits with exception of sample 21, which is unpasturised beer and its color parameters are similar to those of fresh beers.

The samples in the third cluster pose lemon, they are turbid with the hue angle around 76° and lightness between 54-60. In this group incorrectly is the state of sample 15, because its lightness (L) is in the interval 68-78 and hue angle is 78°. Dark beers which belong to the fourth cluster have very low value for L-between 40-50 -and the Hue angle is 65°-66°. The beer samples can be divided in groups by using only the hue angle. The first cluster was characterized with by hue angle between 90°-96°, the second cluster cannot be differentiated from the first according to this parameter. The third one has a hue angle of around 80°, the cluster with dark beer possesses the smallest value for hue angle at about 66°. The difference between the clusters is presented in figure 1.

For a better visualization of the results a subsequent canonical analysis was performed. On the basis of the first two canonical variables, the position of the separate samples for the model with included color parameters is presented in figure 2.

Thus the input set of correlating data was transformed into a new set with a smaller number of uncorrelated artificial variables, so-called „factors” or „principal canonical variables”, which explain the greatest part of the data variation. The reduction in the number of original variables was achieved by grouping the correlated variables in a common factor and the separation of uncorrelated ones in different factors.

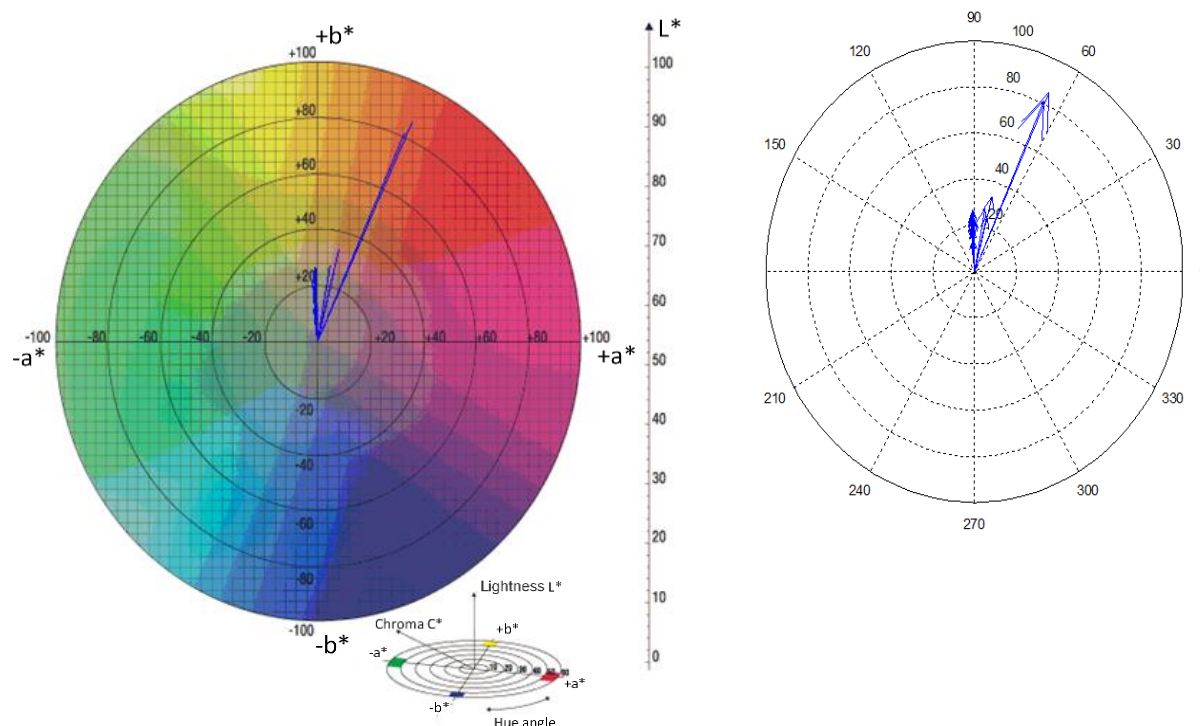


Figure 1. The dependence between color parameters a and b and hue angle for different clusters.

The figure confirms the stated hypothesis for the presence of significant differences between the separate types of beer. The analysis of the Mahalanobis distances between the four basic groups shows that the lager and light beers and the beers with fruits are close to each other and are relatively far from the dark ale beers. This is clearly seen from the figure 2, shown if we trace the projections of the clouds of the various sorts upon the first canonical variable which plays an important role in discrimination of the groups – the dark beers are projected on the positive, while all of the other samples are projected on the negative direction. For each of the sample groups of beers the Mahalanobis distances were calculated, and canonical analysis was simultaneously performed, because Mahalanobis distances provide an idea of the specific features of the examined group in the original space, whereas canonical representations are in a bidimensional space. They serve for a better visualization of the particular groups, since their canonic variables are linear combinations of the initial color indices. Mahalanobis distances represent individual groups, better than canonical variables, because they give an idea of the dynamics of the change in the distance between the group centroids.

On creating the discriminant functions for the sake of a better illustration of the results obtained, a canonical analysis is carried out consisting in the construction of new (latent) canonical variables. This is done according to the rule: each subsequent variable is chosen in a way as to contain together with the preceding variables a maximum of the existing information. In this way they are arranged so that

the first is the “most precious”, followed by the second “less precious” etc. The method of canonic correlation analysis is used to determine the coefficients in the linear combinations.

Greater differences were observed between the centroids of groups from beer with lemon and these from dark beer. The four samples are for the control of the adequacy of the created model for the description of the beer types. From the remaining five types of beer with a known type, the samples which are insufficient to form separate groups, two can be classified as light beers and two as the beer with fruits. The results from the classification according to the obtained models are shown in table 4. The investigated samples are classified 100% correctly.

3.1. Parameters of digital image

Determining the color coordinates of each beer sample is made using the MATLAB™ 7.0 software. The program includes: identifying the areas of each color map by using gradient algorithm, first image is binarized and filtered; determining the deviation between the color components for each color of the cards, obtained by the image and those, specified in the passport; correction of color characteristics of samples of beer with received deviations.

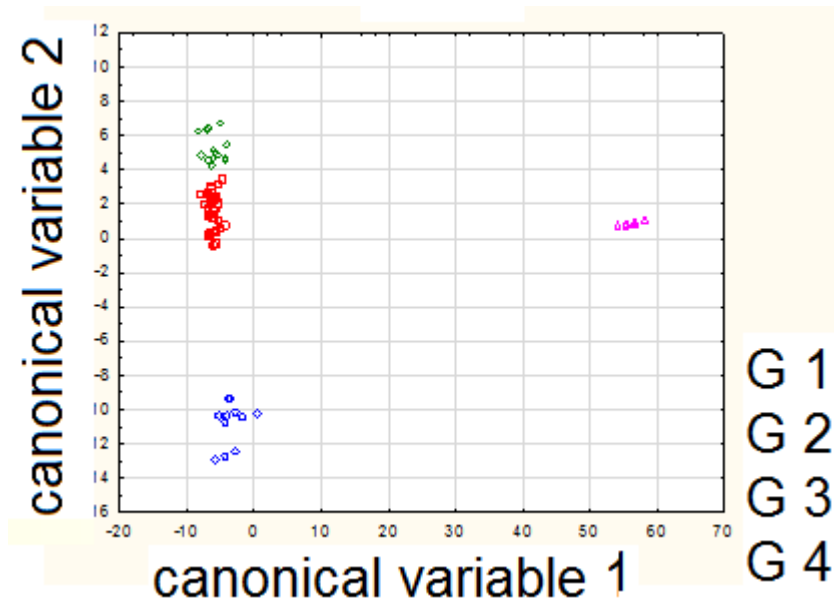


Figure 2. Disposition of the four types of beer in the plane of the first two canonical variables.

Table 3. Mahalanobic distances for different beer groups.

Groups	G 1	G 2	G 3	G 4
G 1	0.000	165.595	263.810	3673.606
G 2	165.595	0.000	34.273	3862.837
G 3	263.810	34.273	0.000	3879.297
G 4	3673.606	3862.837	3879.297	0.000

Table 4. Verification of the model for beer type with independent samples.

Sort	Classified as
Zagorka	Light beer
Kamenitza	Light beer
Kamenitza fresh	Kamenitza fresh
Kamenitza lemon	Kamenitza fresh

3.1. Parameters of digital image

The resulting color coordinates of the samples based on the images correspond to those obtained by the spectrophotometer with the exception of experiments in illuminance 1440 Lux, which may be due to the strong glare. The maximum deviation of the Lab coordinates is 4.2%, disregarding the experiment at 1440 Lux. Minimum deviations are obtained in illuminance of 1140 Lux.

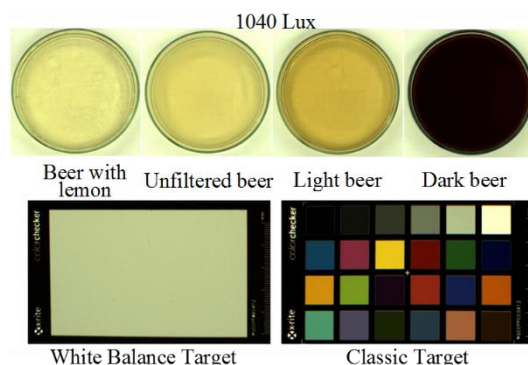


Figure 3. Digital image

4. Conclusion

The analysis of the data base gives the opportunity to characterize different types of beer by using the discriminant analysis. The models and the associated Mahalanobis distances enable the classification of unknown samples. By applying cluster analysis the obtained results give us the opportunity to point out the different types of beer, which have indices around the average for each group and which are similar. They have been featured in accordance with the examined indices. It was pointed out which of them had the biggest influence on their unification. The obtained color characteristics on the base of digital image correspond to these obtained by spectrophotometer with minimal deviation for 1140 Lux.

Acknowledgments

This study was supported by Project SP15-FFIT-004 of the University of Plovdiv (Bulgaria).

References

- [1] Cortacero-Ramirez S, Castro M H B, Segura-Carretero A, Cruces-Blanco and Fernandez-Gutierrez A 2003 *Trends in Analytical Chemistry* **22** 440
- [2] Bamforth C W 2002 *Nutrition Research* **22** 227
- [3] Zhao H, Chen W, Lu J and Zhao M 2010 *Food Chem.* **119** 1150
- [4] Piazzon A, Forte M and Nardini M 2010 *J. Agric. Food Chem.* **58** 10677
- [5] Pietercelie A, Allardin D and Van Nederveelde L 2011 *Cerevisia* **36** 41
- [6] Hucker B, Wakeling L and Vriesekoop F 2011 *J. Agric. Food Chem.* **59** 12278
- [7] Mignani A, Ciaccheria L, Mencaglia A, Ottevaere H, Baca E and Thienpont H 2012 *Sensors and Actuators B* **23** 141
- [8] Geoffrey J 1992 *Discriminant Analysis and Statistical Pattern Recognition* ed Mc. Lachlan (New Jersey, Canada: John Wiley & Sons. Inc) p 132
- [9] Smeyers J J 1998 *Handbook of Chemometrics and Qualimetrics Part A. Part B* ed Verbeke (Amsterdam, The Netherlands, Elsevier) p 145
- [10] Ivanov K 2001 *Statistical methods in social science* (Sofia, Bulgaria: New Bulgarian University) p 124
- [11] Bondar A and Statuha G 1976 *Experiment in chemical technologies* (Moscow, Russia) p 194
- [12] Lakin L 1990 *Biometry* (Moscow, Russia) p 105
- [13] Vundev D 2003 *Applied Statistic* (Sofia, Bulgaria: SU Kliment Ohridski) p 76
- [14] Duran B and Odelle P 1977 *Cluster analysis* (Moscow, Russia: Statistics) p 115
- [15] Hartigan J 1985 *Journal of Classification* **2** 63

- [16] Hartigan J A and Wong M A 1979 *Applied statistics* **28** 100
- [17] Milev M, Nikolova Kr, Ivanova Ir and Dobрева M 2015, *AIP Conf. Proc.* **1690** 020021
- [18] Ward J H 1963 *J. Amer. Stat. Assoc.* **58** 236