Procedia Computer Science

Volume 29, 2014, Pages 2024–2033

ICCS 2014. 14th International Conference on Computational Science

Cachaça Classification Using Chemical Features and Computer Vision

Bruno Urbano Rodrigues[1], Ronaldo Martins da Costa[1]*, Rogério Lopes Salvini[1], Anderson da Silva Soares[1], Flávio Alves da Silva[1], Márcio Caliari[1], Karla Cristina Rodrigues Cardoso[1], and Tânia Isabel Monteiro Ribeiro[2]†

[1] Universidade Federal de Goiás, Goiânia, Goiás, Brasil
brunourb@gmail.com,ronaldocosta@inf.ufg.br,rogeriosalvini@inf.ufg.br, anderson@inf.ufg.br,flaviocamp@ufg.br,macaliari@ig.com.br,karlagropan@hotmail.com

[2] Instituto Politécnico de Bragança, Bragança, Portugal
tania.im.ribeiro@gmail.com

Abstract

Cachaça is a type of distilled drink from sugarcane with great economic importance. Its classification includes three types: aged, premium and extra premium. These three classifications are related to the aging time of the drink in wooden casks. Besides the aging time, it is important to know what the wood used in the barrel storage in order the properties of each drink are properly informed consumer. This paper shows a method for automatic recognition of the type of wood and the aging time using information from a computer vision system and chemical information. Two algorithms for pattern recognition are used: artificial neural networks and k-NN (*k-Nearest Neighbor*). In the case study, 144 cachaça samples were used. The results showed 97% accuracy for the problem of the aging time classification and 100% for the problem of woods classification.

Keywords: pattern recognition, drink analysis, computer vision

1 Introduction

Cachaça is the distilled drink most consumed among alcoholic beverages in Brazil. It is a special type of beverage produced from sugarcane (*Saccharum sp*) similar to rum. Its differential is the use of different types of wood in the aging process.

Aging consists of storing the cachaça in barrels or wooden casks for a certain time. This process produces changes in the chemical composition, aroma, flavor and color of the drink [2]. The legislation classifies the cachaça into three types: aged cachaça, premium cachaça and

*The authors thanks the research agencies CAPES and FAPEG for the support provided to this research.

†The authors thanks the research School of Agriculture, Polytechnic Institute of Bragança for the support provided to this research.

Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2014

extra premium cachaça. The difference among types is related to the quantity of storage and shelf life. A aged and premium cachaça have aging period of at least 1 year. Aged cachaça may have up to 50% of non-aged cachaça using blend process. The premium chacaça has 100% aged cachaça. The extra premium cachaça has a minimum aging period of at least 3 years and cannot contain non-aged cachaça.

As the most famous drinks in the world, whiskeys, brandies and even wines go for oak barrels. Cachaça is the only one that uses different woods for this process[2][8]. Each wood gives the drink a sensory analysis involving the measurement, interpretation and understanding of human responses to the properties perceived by the senses (taste - flavor, smell - aroma, vision - color). However, some woods only fortify the acidity of cachaça and do not interfere with their color or taste. The drink remains white and with its distinctive flavor even after properly fortified in contact with the wood[2]. Detailed knowledge of the chemical and sensory composition of cachaça, as well as the maturation time, constitute important factors in controlling beverage quality and evaluation of changes that may contribute to the improvement of production processes. This knowledge can contribute to the production process especially for small producers and artisan industries.

De Souza[6] uses gas chromatography - olfactometry - to separate and characterize the odors present in cachaça and rum, these two products of sugarcane were compared and the patterns identified from a descriptive sensory analysis. The disadvantage of this method is maintainability because it has high cost. [7] demonstrates the differentiation between cachaça and rum by using ionization mass spectrometry. The author used the principal component analysis (PCA), statistical approach in which data are represented by a subset of its eigenvectors, noting the type of wood (amburana -*Amburana cearensis* e jequitibá - *Cariniana legalis*). His work contributes to further studies can use this technique for the identification of artisanal and industrial cachaça as well as detection of adulteration by adding caramel and other substances such as dyes.

Recent works use techniques of computer vision, neural networks, genetic algorithms and statistical methods for food classification. Wan[14] used the computer vision combined with artificial neural networks. A structural and microscopic approach of wines to be classified was used by analyzing the microstructure and texture, factor that influences the assignment of color to the sample. Starting from the idea that different wines have microstructural (microscopy) and micrograph (particles) changes, the study aimed at extracting common features to define a pattern. For such, neural networks were used for classifying samples. The presented results confirm that it is possible to classify the wine through its micrograph, allowing the use of the features in other contexts. Boisier[3] uses ΔE based on the CIELab color space in the samples and demonstrates the grouping according to the tones classified. The proposed goal was to represent the wines' colors with limited number of colors that were called *nuances*. The application of ΔE aimed at performing a comparison with the HVS model, observing the brightness, chromaticity and saturation, thus analyzing the color spectrum, sorting and grouping it according to tone. The results are encouraging since they permit a precise characterization and reproduction of wine color. The RGB color model is an additive color system consisting of Red (Red) Green (Green), and Blue (Blue). Additive colors are emitted or projected colors. The color is generated by mixing various greetings light wave, causing a color sensation when it reaches the eye. RGB formats, also known as true-color, use 8-bits per channel. The CIELab color model is a subtractive color system. CIELab describes the basic colors in three qualities: L * is lightness, a * and b * contain chroma information. L * is luminance, density measurement of the intensity of a reflecting light in a given direction. The information a* and b* refers to the amount of color [9].

Qiongshuai[12] in his analysis shows the gain from using genetic algorithms in lecture and classification of wines, combined with computer vision. Kruzlicova[10], evaluates data through an artificial neural network and comparative method use the analysis of variance (ANOVA). Cozzolino[5] proposed to investigate the relationship between sensory analysis, visibility (VIS) and infrared spectroscopy (NIR) to evaluate the sensory properties of commercial varieties of Australian wines by using the PCA (Principal Component Analysis).

The methods described in related articles do not address the variety of woods that can be used, as well as aggregate usage of chemical data values obtained by the colorimeter and there is no relation so far of digital photographs of samples using the RGB color model. Sometimes works use only the chemical data, other use only data from colorimeter (CIELab color model) and when using chemical data and data from the colorimeter, did not observe the RGB color model. The colorimeter is generally described as any instrument that characterizes color samples to get an objective measure of color characteristics. In turn, such equipment is available in research laboratories and industries. The relevant point is to make accessible this technology regardless of the producer. Observing the instrumental methods, cost, maintainability and handling are performed by a specialist. With computer methods results can be achieved optimizing time and resources.

Therefore, this paper proposes a method for classifying the aging process of cachaça in order to identify the wood and the aging time of a sample. Intersection of information obtained in the chemical analysis with that extracted from colorimeters will be performed, as well as data obtained by applying algorithms of images digital processing, digital photographs performed on samples of cachaça. It is used the technique of artificial neural networks to assess the influence of types of wood and the time that cachaça aging has in the color model obtained from digital photographs (RGB) and colorimeter (CIELab). Two techniques of pattern recognition will be used: neural networks and k-NN (*K-Nearest Neighbor*).

2 Materials and Methods

2.1 Samples

Cachaça samples with up to 36 months of in casks aging of amburana (*Amburana cearensis*), oak (*Quercus spp*) and nut (*Bertholletia excelsa H.B.K*) were evaluated. The aging time is described every 4 months (4-8-12-16-20-24-28-32-36). All samples evaluated are from 4 barrels of each timber. Thus, total number of samples to be analyzed are 36 samples per timber. Normative Instruction No. 13 of MAPA (Ministry of Agriculture, Livestock and Supply)[11] defines on the classification of Brazilian cachaça into three types: Aged cachaça, premium cachaça, extra-premium cachaça. All types have alcohol content between 38% and 48% by volume at 20°C. What differs is the type of storage time in cask wood. Aged cachaça has 50% of the sample stored in wooden cask for at least one year. The premium chacaça has in its entirety, aged in wooden cask for at least one year and finally the extra-premium cachaça, in its entirety, aged in wooden cask for a period not less than three years.

The physical and chemical analyses were performed in the laboratories for beverage technology and physicochemical analysis of the School of Agronomy, Federal University of Goiás. The determination of pH, density, real alcohol content at 20 °C, volatile, fixed and total acidity, dry extract, phenolic compounds, color and antioxidant activity were performed on times 0, 2, 4, 6, 8, 10 and 12, i.e. 2 on 2 months of storage for observing changes during the aging period. The analyses followed the following methodologies:

- pH (Features 4 and 5) - measured with digital potentiometer calibrated at 20°C;

Wood type	Aging time (months)	samples
amburana (*Amburana cearensis*),	4 - 8 - 12 - 16 - 20 - 24 -28 - 32 - 36	36
oak (*Quercus spp*)	4 - 8 - 12 - 16 - 20 - 24 -28 - 32 - 36	36
nut (*Bertholletia excelsa H.B.K*)	4 - 8 - 12 - 16 - 20 - 24 -28 - 32 - 36	36
Total		144

Table 1: Samples of cachaça analyzed for up to 36 months of aging time

- density (Features 1 and 2) - based on the relationship between the specific weight of water at 20°C using pycnometer or hydrostatic device based on the Archimedes' principle (in which one body immersed in a liquid is subjected to a vertical thrust of the liquid upward, equal to the weight of the displaced fluid);

- Real alcohol content at 20°C (Feature 7 to 13), volatile, fixed and total acidity and dry extract: were performed according to the Brazilian official methods of analysis for distilled drinks.

- Total phenolic compounds (Feature 3 to 14) : will be determined according to the official method of analysis of AOAC 952.03 (AOAC, 1997), derived from the standard-curve calibration with tannic acid with reading at 760nm absorbance.

- Color: will be determined in a ColorQuest II / Hunter Lab color spectrophotometer, adjusted for reflectance with specular included, using the blank No. C6299 of 03/96 and sample in bucket of clean glass 10mm-optical path with 1-inch field analysis. The configuration included illuminant D65 and angle of incidence of 10º. The readings were performed in the color universal system CIELab with turbidity (homogeneous dispersion of solids in solution) and without turbidity (clear sample). It will be conducted with the reading to determine the color-luminosity coordinates L, a* and b*. The color will also be assessed based on information from digital photographs that will be taken of all the 144 cachaça samples. The features that influence the color is: Feature 1, 2, 7, 8, 9, 10, 11, 12 and 13.

- total aldehydes (Features 6 to 14) and esters (Feature 3) and isoamyl (Feature 10), isobutyl (Feature 9) and n-propyl (Feature 8) higher alcohols: were determined in a gas chromatograph Shimadzu GC-17A equipped with automatic injection, automatic ionization detector, flame ionization detector and capillary column DB-VAX (30m x 0.25 mm x 0.25 mm). In determining the compounds concentration were performed the area method and calibration with external standards.

- Testing of antioxidant activity in vitro (Feature 3 to 14): were determined by the method described by Brand-Williams, Cuvelier, and Berset[4]. This method is based on the DPPH stable radical from the reaction medium by the action of antioxidants in the sample.

In the analysis is shown the attributes to be used in the model. The attributes are describeda in the table 2.

(a)

Chemical Features description	
Feature 1	Apparent Alcohol
Feature 2	Real Alcohol
Feature 3	Total Esters
Feature 4	Ethyl Acetate
Feature 5	Ethyl Lactate
Feature 6	Aldehydes
Feature 7	Total Alcohols

(b)

Chemical Features description	
Feature 8	n-propyl
Feature 9	Isobutanol
Feature 10	Isoamyl
Feature 11	1-Butanol
Feature 12	2-Butanol
Feature 13	Methyl Alcohol
Feature 14	Furfural

Table 2: Representation of Chemical features with their respective numbers and description

2.2 Computer vision system

Subsequently, samples were photographed by digital camera *Canon EOS REBEL XS* with setting ISO 100, aperture to 4.0mm and configured for RAW image that contains all of the image data as captured by the camera sensor format. The ambient light to photograph the samples was controlled by a device which allows the incidence of light in the opposite position to the lens of the camera. A special filter will prevent reflections in the liquid and will allow the capture of a digital image suitable for processing.

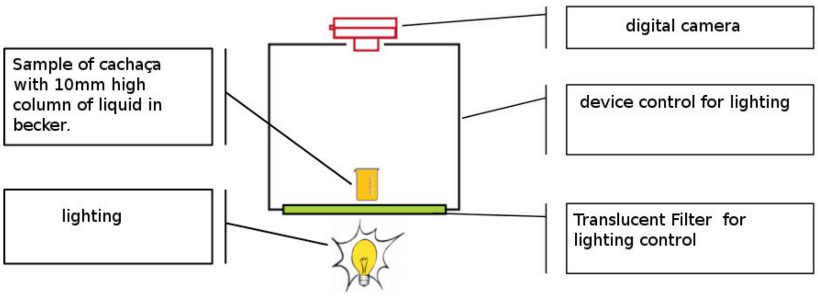


Figure 1: Computer Vision System

Figure 1 shows the project of the device designed to be used in this work, a technique inspired by Sun[13] in his work for bovine meat classification. The device measures were $50cm^2$, with translucent filter of $30cm^2$, opening digital camera $10cm^2$ radius. The purpose of the device is to control the environment of digital photography for better absorption of colors of the target object, in this context the cachaça, in order to observe a correlation between the color characteristics obtained by the colorimeter (model CIELab L* a* b* - Lightness, redness and yelowness) and the RGB model (Red, Green and Blue). Afterwards, assign white balance, a process for removal of unreal colors, so that making white objects that appear being white to our eyes. The color balance is previously made, both in photography and film, to digital photography. The color balance is related to neutrality and should not be confused with color balance that painters and designers often apply for matching colors.

The representation of the properties used in the model RGB and CIELab shown in the table:

As in chemical analysis, the properties of the CIELab and RGB color models have been named and separated for use in the classifier.

(a)

CIELab Features description	
Feature 15	Lightness
Feature 16	Redness
Feature 17	Yelowness

(b)

RGB Features description	
Feature 18	Color Red
Feature 19	Color Green
Feature 20	Color Blue

Table 3: Representation of CIELab features (a) and RGB features (b) with their respective numbers and description

2.3 Pattern Recognition Algorithms

In this work is proposed the use of two algorithms for pattern recognition: artificial neural network and k-NN. Both techniques use supervised learning type.

Artificial Neural Networks (ANN) are mathematical models for data analysis inspired in neuronal structures of the brain. It is a connectionist model, with great power to solve complex and non-linear problems, with application in several areas. A multilayer perceptron neural network (MLP) with 11 neurons in the hidden layer will be used. The training algorithm used was backpropagation.

Another method used is the k-NN (*k-Nearest Neighbor*). Lazy type supervised learning technique, introduced by Aha[1]. The general idea of this technique is to find the k closest labeled examples to that unlabeled; based on the labeling of the closest examples the decision on the class of unlabeled example is made. The size of k in this work is 1 using the Euclidean distance.

Due to the limited number of samples, the cross-validation technique was used to measure the accuracy of classifiers. In this technique, samples are divided into n mutually exclusive partitions. At each iteration a different partition is used to test the classifier and all the others n-1 partitions are used to train the classifier. The hit rate and error is the average of all rates calculated for the n iterations. In this work we used the n equal to 10.

3 Results and discussion

The pool of colorimeter information, chemical analysis and digital photographs at the entrance of the classifiers was carried out. Two pattern recognition algorithms were used: neural networks and k-NN (*k-Nearest Neighbor*).

In the first experiment is used as attributes only the chemical information, i.e. without using the information from colorimeter and RGB model for identification of the aging time and wood. The results are shown in Table 4.

Problem	Aging time	Wood type
hits(%)	94.44%	96.26%
errors(%)	5.56%	3.74%

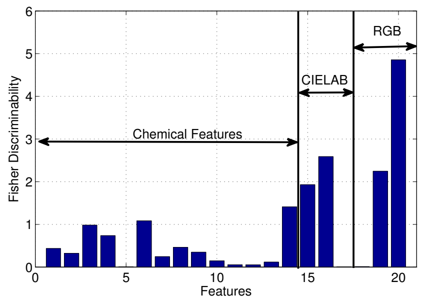
Table 4: Recognition results of chemical features in neural network.

Problem	Aging time	Wood type
hits(%)	83.33%	95.33%
errors(%)	16.67%	4.67%

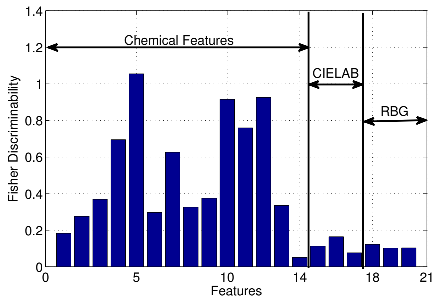
Table 5: Recognition results of chemical features in k-Nearest Neighbor.

According to the results, both classifiers considered achieved high success rate using the data of chemical analysis. The only caveat presented is that the result for the classification of the aging time using the k-NN obtained a relatively lower rate of accuracy (83.33%).

Besides chemical attributes, variables that make use of color information were measured using the CIELab and RGB color model. Figure 2 shows the Fisher discriminative capacity for chemical attributes, RGB and CIELab to the problem of classifying the type of wood and aging time. As one can see, the attributes related to color information have more discriminability to the problem of wood classification. The attributes 16 and 20 have the highest discriminability. For the problem of aging time classification is possible to note from the Figure 2(b) that the attributes of greatest discriminability are related to chemical data. Information related to computer vision system has low discriminability to the problem. It is noteworthy that the Fisher's discriminability considers the attribute of univariate analysis, thus the use of the most discriminative attributes does not imply a good classification model.



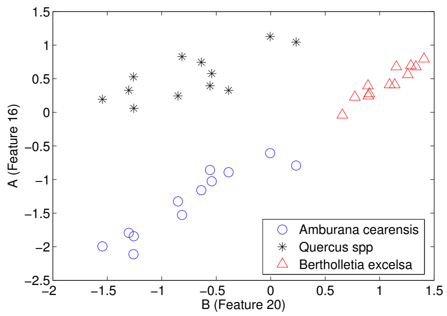
(a) Fisher discriminability of class wood type



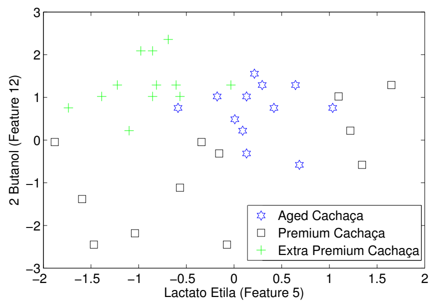
(b) Fisher discriminability of class aging time

Figure 2: The Fisher Discriminability in wood type and aging time problem, using chemical features, color model CIELab and RGB. In Wood type problem the CIELab and RBG features has major discriminability as showed in (a). However, in aging time problem the chemical features the chemical features has the major discriminability as showed in (b).

From the calculated discriminability, the two most discriminative variables for each problem considered for viewing a scatterplot of objects for the problems of wood and aging time classification were used. Figure 3(a) shows that the classification of the wood type is a simpler problem than the aging time classification observed in Figure 3(b).



(a)



(b)

Figure 3: Object disperison in wood type(a) and aging time(b) pattern recogntion problems.

Verified that the information may contain relevant details to the problems considered, an experiment was performed using only the attributes of RGB and CIELab in the classifiers considered without the use of chemical attributes. The results are shown in Tables 6 and 7.

Problem	Aging time	Wood type
hits(%)	52.78%	97.20%
errors(%)	47.22%	2.80%

Table 6: Recognition results of color model CIELab and RGB in neural network without chemical features.

Problem	Aging time	Wood type
hits(%)	44.44%	98.13%
errors(%)	55.56%	1.87%

Table 7: Recognition results of color model CIELab and RGB in k-NN without chemical features.

Satisfactory results for the problem of wood classification can be obtained by using only the attributes of the Computer Vision System (RGB and CIELab features) system. However, for the problem of aging time classification the results for both classifiers have a bad hit rate.

From the observation that the information related to colors may contain useful information about the classification problem considered in this paper, a new experiment was performed using the chemical attributes from the CIELab color model.

Problem	Aging time	Wood type
hits(%)	91.67%	100.00%
errors(%)	8.33%	0.00%

Table 8: Recognition results of chemical features using color model CIELab in neural network.

Problem	Aging time	Wood type
hits(%)	86.11%	97.19%
errors(%)	13.89%	2.81%

Table 9: Recognition results of chemical features using color model CIELab in k-NN.

From the results of Tables 8 and 9, one can see that the classifier achieved high success rate using chemical information associated with the CIELab attributes. Both classifiers considered achieved high success rate compared to the classification result of tables 4 and 5. There was improvement in the success rate for the problem of wood classification for both classifiers. In the problem of aging type classification, there was improvement only in the k-NN classifier.

In the third experiment were used chemical attributes, the attributes of the RGB and CIELab.

Problem	Aging time	Wood type
hits(%)	97.22%	100.00%
errors(%)	2.78%	0.00%

Table 10: Recognition results of chemical features using color model CIELab and RGB in neural network.

Problem	Aging time	Wood type
hits(%)	88.89%	100.00%
errors(%)	11.11%	0.00%

Table 11: Recognition results of chemical features using color model CIELab and RGB in k-NN.

According to the results shown in Tables 10 and 11, the problem of wood type classification showed 100% accuracy for both classifiers considered. The result for the aging time classification showed improvement in the accuracy rate for the k-NN classifier (88.89%) and also improvement to the neural network (97.22%). Table 12 shows the neural network confusion matrix for the classification problem of aging time. The only classifier error was to indicate a sample of Premium cachaça and Extra Premium cachaça.

Aging time	Aged Cachaça	Premium Cachaça	Extra Premium Cachaça
Aged Cachaça	12	0	0
Premium Cachaça	0	11	1
Extra Premium Cachaça	0	0	12

Table 12: Confusion matrix of aging time generated by neural network.

4 Conclusion

This paper proposed the use of pattern recognition algorithms to identify the type of wood and aging time of cachaça samples. From the results it was observed that for the wood classification problem was possible to obtain classifiers with 100% accuracy. Still to this problem, it was found that the use of computer vision system only, without the use of chemical information is sufficient to identify the wood type with high accuracy rate. For the problem of aging time classification, the best result (97%) was obtained by a neural network using the chemical information and the information from the computer vision system.

References

[1] D. Aha and D. Kibler. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991.

[2] Francisco W. B. Aquino, Ronaldo F. Nascimento, Sueli Rodrigues, and Antônio Renato S. Casemiro. Determinação de marcadores de envelhecimento em cachaças. *Food Science and Technology (Campinas)*, 26:145 – 149, 03 2006.

[3] B. Boisier, A. Mansouri, P. Gouton, and P. Trollat. Wine color characterization and classification for nuances reproduction. In *Signal-Image Technology Internet-Based Systems (SITIS), 2009 Fifth International Conference on*, pages 93–98, 2009.

[4] W. Brand-Williams, M.E. Cuvelier, and C. Berset. Use of a free radical method to evaluate antioxidant activity. {*LWT*} - *Food Science and Technology*, 28(1):25 – 30, 1995.

[5] D. Cozzolino, G. Cowey, K.A. Lattey, P. Godden, W.U. Cynkar, R.G. Dambergs, L. Janik, and M. Gishen. Relationship between wine scores and visible-near-infrared spectra of australian red wines. *Analytical and Bioanalytical Chemistry*, 391(3):975–981, 2008.

[6] Maria D. C. A. de Souza, Pablo Vásquez, Nélida L. del Mastro, Terry E. Acree, and Edward H. Lavin. Characterization of cachaça and rum aroma. *Journal of Agricultural and Food Chemistry*, 54(2):485–488, 2006. PMID: 16417309.

[7] Patterson P. de Souza, Daniella V. Augusti, Rodrigo R. Catharino, Helmuth G. L. Siebald, Marcos N. Eberlin, and Rodinei Augusti. Differentiation of rum and brazilian artisan cachaça via electrospray ionization mass spectrometry fingerprinting. *Journal of Mass Spectrometry*, 42(10):1294–1299, 2007.

[8] J. B. Faria, D. W. Franco, and J. R. Piggott. The quality challenge: cachaça for export in the 21st century. In *Distilled spirits: tradition and innovation 2004*, pages 215–221. Nottingham University Press, Nottingham, UK, 2004.

[9] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[10] Dasa Kruzlicova, Jan Mocak, Branko Balla, Jan Petka, Marta Farkova, and Josef Havel. Classification of slovak white wines using artificial neural networks and discriminant techniques. *Food Chemistry*, 112(4):1046 – 1052, 2009.

[11] MAPA Ministério da Agricultura, Pecuária e Abastecimento. Instrução normativa n^o 13, de 29 de junho de 2005. [online], January 2014. `http://extranet.agricultura.gov.br/sislegis-consulta/servlet/VisualizarAnexo?id=14175`.

[12] Lv Qiongshuai and Wang Shiqing. A hybrid model of neural network and classification in wine. In *Computer Research and Development (ICCRD), 2011 3rd International Conference on*, volume 3, pages 58–61, 2011.

[13] X. Sun, H. J. Gong, F. Zhang, and K. J. Chen. A digital image method for measuring and analyzing color characteristics of various color scores of beef. In *Image and Signal Processing, 2009. CISP '09. 2nd International Congress on*, pages 1–6, 2009.

[14] Yi Wan, Xingbo Sun, and Rong Guo. Shape and structure features based chinese wine classification. In *Computational Intelligence and Natural Computing, 2009. CINC '09. International Conference on*, volume 2, pages 39–43, 2009.