# BAYESIAN SUPERVISED IMAGE CLASSIFICATION BASED ON A PAIRWISE COMPARISON METHOD

F. Calle-Alonso, J. P. Arias-Nicolás, C. J. Pérez and J. Martín
*Department of Mathematics, University of Extremadura, Cáceres, Spain*

Keywords: Bayesian regression, Interactive learning, Pairwise comparison, Pattern recognition, Relevance feedback, Supervised classification, Content-based image retrieval.

Abstract: In this work, a novel classification method is proposed. The method uses a Bayesian regression model in a pairwise comparison framework. As a result, we obtain an automatic classification tool that allows new cases to be classified without the interaction of the user. The differences with other classification methods, are the two innovative relevance feedback tools for an iterative classification process. The first one is the information obtained from user after validating the results of the automatic classification. The second difference is the continuous adaptive distribution of the model's parameters. It also has the advantage that can be used with problems with both a large number of characteristics and few number of elements. The method could be specially helpful for those professionals who have to make a decision based on images classification, such as doctors to determine the diagnosis of patients, meteorologists, traffic police to detect license plate, etc.

## 1 INTRODUCTION

Pattern recognition has become an active research area (Theodoridis and Koutroumbas, 2003). Its importance has increased in the last few years with the development of new Content-Based Image Retrieval (CBIR) methods, and it will be in the front sight until two fundamental problems are solved: how to best learn from users' query concepts, and how to measure human perceptual similarity.

The growth of information technologies and new computer tools have produced a huge increase of available information, specially of images and videos. Image classification methods are being used in many disciplines and they are considered important tools to help many professionals.

Some years ago, a very popular way to solve the problem of image classification was to label all the images and then classify them just by these keywords. However, there are two difficulties that make this method is not a good option to solve the problem. The first one is the fact that every image must be tagged, and a person is needed to perform this task. The second one is the human perception, which is clearly subjective, i.e.: the same image may be perceived with a different content by different people (Tversky, 1977). This is called perceptual or semantic gap.

In this work, a classification method to solve the problems exposed before is presented. With a mathematical classification rule, the use of labels is avoided. We propose an automatic method which has the advantage that it only should be partially supervised, so there is no need to have one person labeling all the images. The second difficulty (perceptual gap) appears specially in broad problems where images with general information produce scattered opinions from different users. It is solved if the method is only used in narrow problems (El-Naqa et al., 2004). Solving only specific problems is also an advantage because we overtake the troubles caused by broad domains, like unlimited and unpredictable variability (Smeulders et al., 2000). In narrow problems, we can find an objective expert that supervises the classification results applying always the same criterion. This allows to avoid the problem of the perceptual gap. Finally, the information obtained from both the expert and the automatic classification can be added to the next stage of classification to improve the results of the model.

The method can solve any classification or comparison problem. It is based on the pairwise comparison method proposed by (Arias-Nicolás et al., 2007b), which solved the preference aggregation problem by finding the optimal solution in a group decision making framework. This methodology has been adapted to support the inclusion of human interaction in the pattern recognition process by proposing a Bayesian

Calle-Alonso F., P. Arias-Nicolás J., J. Pérez C. and Martín J..
BAYESIAN SUPERVISED IMAGE CLASSIFICATION BASED ON A PAIRWISE COMPARISON METHOD.
DOI: 10.5220/0003876004670473
In *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods* (IATMLRP-2012), pages 467-473
ISBN: 978-989-8425-98-0

regression-based approach that uses image features instead of expected utilities. Moreover, it can be used for any other kind of classification just changing the data appropriately. We show later in the section 3 two different applications to illustrate the method.

## 2 BAYESIAN REGRESSION-BASED PAIRWISE COMPARISON METHOD

We propose a novel pairwise comparison method based on Bayesian regression to classify images. Firstly, image features are extracted and some pairs of images are compared to set them as similar or not. Then, the proposed model, with a weakly informative prior distribution, is carried out to obtain a classification rule. With this rule new images are classified. Next an expert supervise the results and this information is incorporated in the process to recalculate the classification rule. So, a continuous adaptive learning process is proposed until the method is able to classify new images in an automatic way with a high percentage of success. The complete process appears in the figure 1. The features extraction has been partially implemented on *Qatris IManager* (Arias-Nicolás et al., 2007a) and the parameter estimations for the Bayesian probit regression model has been obtained from *WinBUGS*.

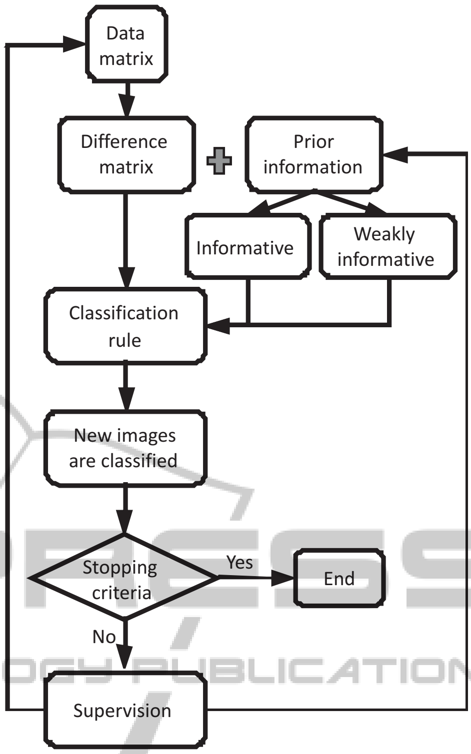

Figure 1: Complete process.

### 2.1 Feature Extraction

Numeric variables must be defined to represent the image features. Then, the objective will be to set a rule to classify the images in separated regions by considering their features (Jai et al., 2000) and (Schürmann, 1996). Three different kind of features are considered: color, texture and shape (Chorás, 2003).

- Color. We looked for a color model similar to the human perception. The one based on Hue, Saturation and Luminosity (HSL) seems to be suitable for this purpose (Smith, 1978). The 15 main colors that we use are (de la Escalera, 2001): white, grey, black, red, red-yellow, yellow, yellow-green, green, green-cyan, cyan, cyan-blue, blue, blue-pink, pink and pink red.
- Texture. Two methods are used to extract texture features. The first one is based on the gray level co-ocurrence matrix (Haralick and Shapiro, 1993). The second method detects linear texture primitives and uses the run length matrix (Galloway, 1975).
- Shape. This features are based on the methods proposed by (Belkasim et al., 1991). We use Hu's moments (first and second moments), centroid (center of gravity), angle of minimum inertia, area, perimeter, ratio of area and perimeter, and major and minor axis of fitted ellipse.

In order to extract these features *Qatris Imanager* software is used (Arias-Nicolás et al., 2007a) and (Arias-Nicolás and Calle-Alonso, 2010). It has been developed by SICUBO (spin-off of the University of Extremadura) with the cooperation of the research groups DIB (Decisión e Inferencia Bayesiana) and GIM (Grupo de Ingeniería de Medios) from the University of Extremadura. An example of the color features used is shown in the figure 2.

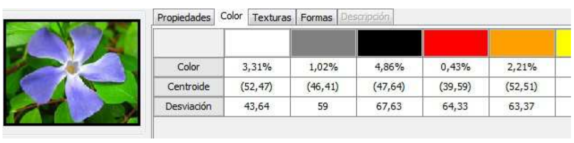

Figure 2: Color features.

### 2.2 Pairwise Comparison Method

The pairwise comparison method begins after the extraction of the features. We start with $r$ images and the values for each feature is saved in the data matrix $\mathbf{A}$, so we have $r$ features vectors $a_1,\dots,a_r$. This method compares every pair of images to obtain a matrix of differences $\Lambda$. This matrix is used later in the Bayesian regression.

Let $d_k$ be the difference function between images for the $k$-feature, with $k = 1,\dots,K$. For each pair of images $(a_i,a_j)$ the values for the independent variables are computed as:

$$\mathbf{x}_{a_i a_j} = (d_1(a_i,a_j), d_2(a_i,a_j), \dots, d_K(a_i,a_j)). \quad (1)$$

The response variable is defined as:

$$y_{a_i a_j} = \begin{cases} 0, & \text{If images belongs to the same class,} \\ 1, & \text{If images belongs to different classes.} \end{cases} \quad (2)$$

Now we can define the matrix of diferences $\Lambda$. It contains (for every pair of images) the value of the independent variable, the differences between the features and, if a regression model with constant is required, a row of 1's.

$$\Lambda = \begin{pmatrix} y_{a_1a_2} & y_{a_1a_3} & \dots & y_{a_{r-1}a_r} \\ 1 & 1 & \dots & 1 \\ d_1(a_1a_2) & d_1(a_1a_3) & \dots & d_1(a_{r-1}a_r) \\ d_2(a_1a_2) & d_2(a_1a_3) & \dots & d_2(a_{r-1}a_r) \\ \vdots & \vdots & \dots & \vdots \\ d_K(a_1a_2) & d_K(a_1a_3) & \dots & d_K(a_{r-1}a_r) \end{pmatrix}. \quad (3)$$

The Bayesian binary regression is applied with these data. The objective is classify new images in their respective classes. With this method any kind of classification problem could be solved. It doesn't matter the number of classes.

## 2.3 Bayesian Regression

Let $\pi$ and $\beta$ be defined as:

$$\pi_{a_i a_j} = F(\beta^T \mathbf{x}_{a_i a_j}), \quad (4)$$

$$\beta = (\beta_1, \beta_2, \dots, \beta_K)^T, \quad (5)$$

with $F(\cdot)$ being a cumulative distribution function, and $\beta$ the regression parameters vector. A binary regression model is considered, so the independent variable follows a Bernoulli distribution $y_{a_i a_j} \sim Bernoulli(\pi_{a_i a_j})$. In order to select the specific regression model, a model choice approach based on the DIC (Deviance Information Criterion, (Spiegelhalter et al., 2002)) has been performed. We have considered eight different regression models with symmetric and asymmetric links. The probit model is usually the one which obtains the lowest DIC and the best fit for the data. Therefore, the probit model (McCullagh and Nelder, 1989) is considered, i.e.:

$$\pi_{a_i a_j} = F(\beta^T \mathbf{x}_{a_i a_j}) = \Phi(\beta^T \mathbf{x}_{a_i a_j}), \quad (6)$$

being $\Phi$ the standard normal cumulative distribution function.

In the Bayesian regression model, the coefficients are random variables. Firstly, the method is applied with no prior information, so a weakly informative prior distribution for the parameters is used (Zellner and Rossi, 1994). Specifically, a normal prior distribution with mean equal to zero and high variances (to let the parameters vary in a large range) is considered.

Since our objective is to determine a measurement of discrepancy among images, and as the independent variables are non-negative, the predictor will be non-negative. Then, the parameters of the model should be non-negative. In order to achieve this goal, a truncated normal in the interval $[0,u]$, with mean $\mu$ and variance $\sigma^2$ can be considered. Then, the cumulative distribution function is:



$$F(x) = \begin{cases} 0 & x < 0, \\ \frac{\Phi(\sqrt{\sigma}(x-\mu)) - \Phi(\sqrt{\sigma}(-\mu))}{\Phi(\sqrt{\sigma}(u-\mu)) - \Phi(\sqrt{\sigma}(-\mu))} & 0 < x < u, \\ 1 & x > u. \end{cases} \quad (7)$$

In order to estimate the parameters, Markov Chain Monte Carlo (MCMC) methods are implemented by using the software *WinBUGS* (Chen et al., 2000). We have to simulate a long chain to achieve convergence. Then, we discart the first iterations and use the rest to estimate the regression parameters.

After the coefficients have been estimated, any new image $a_{new}$ could be classified. The method automatically obtains all the pairwise comparisons between $a_{new}$ and any other $a_i$. With these values we can estimate every $y_{a_{new}a_i}$ for $i$ in 1 to $r$ by discretizing:

$$\pi_{a_{new}a_i} = \Phi(\beta^T x_{a_{new}a_i}). \quad (8)$$

Finally, we use a decision criterion to assign a class to every image. The nearest neighbors method (Fix and Hodges, 1951) is used, so we find the $N$ lowest values of $\mathbf{y}$ and their corresponding images. We can assign the new image to the class with highest frequency among the images selected.

## 2.4 Relevance Feedback

Now, we can classify new images using the method based on a weakly informative prior distribution, but the objective is to use the information achieved in the experiments to improve the results. The supervision of an expert is needed in the last stage of the method.

When some new images are classified automatically, they can be in the correct class or not because the system is not free of errors. To improve the classification and provide high quality results an expert may help supervising the new images that have just been evaluated. The information of the belonging class can be added to the data used in the method. First we make all the possible pairs between the initial images and the new ones and we add $y = 0$ if the two images compared are in the same class or $y = 1$ if they belong to a different class. The new pairs and their differences are incorporated to the data matrix and will be important in the next estimation of the model parameters.

However, there is more information we can provide from this results. The first time the parameter estimation (made with the training sample of images) is performed, a weakly informative prior distribution is used. Later, the posterior distribution is studied. The information obtained about this distribution is set as the new prior distribution for the next stage and the results should “learn” from the past experiment.

This interactive learning process let us to update the parameters $\beta$ of the model in a continuous way and both the new data supervised by the expert and the new information from the distribution will continue improving the results until the process has learnt enough to be applied in an completely automatic way.

# 3 ILLUSTRATIVE EXAMPLES

## 3.1 Sports Images

In order to illustrate the method we show a simple example of image classification. We have to classify sport images from TV into two classes: football or handball. To simplify and without loss of generality, in this example we will use only 15 color percentage independent variables. The objective is to classify correctly the images obtained from TV to know which sport is appearing at every moment. Two images that represent each class are shown in the figure 3.


Figure 3: Representative images of handball and football.

We start training the system with 20 images, 10 of each class. Let $a_1, a_2, \dots, a_{10}$ be football images and $b_1, b_2, \dots, b_{10}$ handball ones. By using *Qatris Imanager*, we extract the color features. Then, every pair of images is evaluated to obtain the distance between them (190 pairs). Also, the classification value is assigned to each class, either $y = 0$ when both images are from the same class, or $y = 1$ if one image is football and the other is handball. Finally a weakly informative prior distribution is considered for the regression parameters.

The simulation is performed by using *WinBUGS*. In order to achieve convergence 100,000 iterations are simulated, burning the first 90,000. With the last 10,000 iterations, we estimate the parameters $\beta$ for the Bayesian probit model. With these estimations, we can compute the value of $\hat{\mathbf{y}}$. These values can be compared with the real ones, because the classification of these images is well known. As it can be seen in Figure 4, all the estimations for training data are correct, because the real values are one hundred times $y = 0$ and ninety $y = 1$.

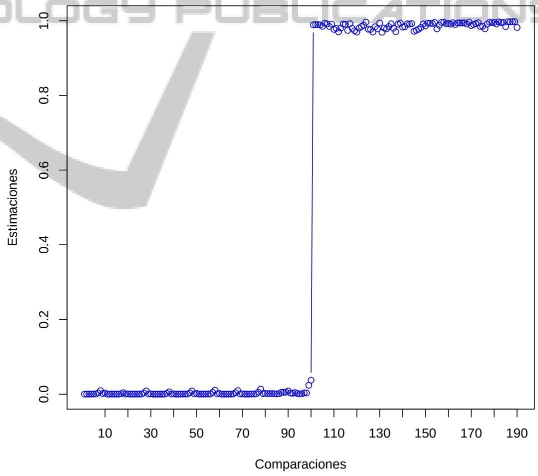


Figure 4: Values $\hat{\mathbf{y}}$ computed.

With the estimation of parameters of the model we can classify any new image. In this case we will choose 10 new images to be classified and, to test the method, we already know their classification. We will study five new football images $a_{11}, a_{12}, \dots, a_{15}$ and five handball images $b_{11}, b_{12}, \dots, b_{15}$.

First we form all possible pairs between the new images and the old ones to obtain the differences matrix. With this matrix and the estimation of $\beta$ we can compute the values of $\hat{\mathbf{y}}$. These values are arranged from lowest to highest and, following the nearest neighbor criterion, we study the ten images more similar to each new image. Two images are more similar as their corresponding value of $y$ gets closer to zero. For example, we can see in the table 1 the ten

Table 1: Ten most similar images to $a_{11}$ and $b_{11}$.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $a_{11}$ | $a_5$ | $a_2$ | $a_7$ | $a_4$ | $a_6$ | $a_1$ | $a_3$ | $a_8$ | $a_{10}$ | $a_9$ |
| $b_{11}$ | $b_{10}$ | $b_9$ | $b_2$ | $b_1$ | $b_3$ | $b_5$ | $b_4$ | $b_8$ | $b_6$ | $b_7$ |

most similar images to the new $a_{11}$ and $b_{11}$.

The ten images closest to $a_{11}$ are football images, and to $b_{11}$ are handball images, so we can classify them in those classes. By following the same process for all the new images, we obtain a perfect classification. All the new images are classified into their correct class and there is no need to continue improving the method with the supervision.

In other problems, there could be some mistakes. At that moment, the expert should start the supervision. When the new images are supervised and corrected (if there is any error), they are moved into the data matrix to be used in the next stages of the procedure, constructing the classification rule. This way to proceed is really valuable, since by providing the opinion of an expert, the model can be guided in the correct direction.

But there is still some information we can take advantage of. The estimation of $\beta$ gives us the opportunity to study its posterior distribution. As a result we estimate the distribution function for each parameter and incorporate them to the model as a prior distribution. For example, for $\beta_1$ we estimate a normal distribution with mean 2.0608 and variance 0.3389.

The system continues receiving images and in every stage it classifies and learns from the expert and from the posterior distribution of the previous stage. Actually, the expert only needs to supervise the results a few times until the method learns how to classify correctly, then it will do a completely automatic classification.

## 3.2 Breast Tissue Classification

Our method is not only useful for images with the features described before, we can use it to classify any kind of data. In order to prove it, a real experiment is evaluated. In this problem (Estrela da Silva et al., 2000), the authors try to classify 106 patients, with the results of a spectroscopy test, into 6 classes. This type of test to diagnose breast cancer has the great advantages to be a minimally invasive technique, very easy to use and also inexpensive. Significant differences of breast tissue were found by (Jossinet, 1998) using this technique, so breast cancer could be fairly detected with the Electrical Impedance Spectroscopy (EIS).

Breast tissue was sampled from 106 patients undergoing breast surgery. The nine features used are extracted from Argand plot. Four of them were previously defined and studied statistically (Jossinet and Lavandier, 1998):

- $I_0$: Impedivity at zero frequency.
- $PA_{500}$: Phase angle at 500kHz.
- $S_{HF}$: High frequency slope of phase angle (at 250, 500 and 1000 kHz).
- $D_A$: Impedance distance between spectral ends.

The other five features were newly presented in the article (Estrela da Silva et al., 2000):

- *Area*: Area under spectrum.
- $Area_{DA}$: Area normalised by $D_A$.
- $IP_{MAX}$: Maximum of the spectrum.
- $D_R$: Distance between $I_0$ and real part of the maximum frequency point.
- *Perim*: Length of spectral curve.

There are six tissue classes, including three normal and three pathological. We will try to discriminate all of them, but with special attention to carcinoma.

1. Connective tissue (normal).
2. Adipose tissue (normal).
3. Glandular tissue (normal).
4. Carcinoma (pathological).
5. Fibro-adenoma (pathological).
6. Mastopathy (pathological).

The authors found that it was difficult to solve the problem with six classes together. As a solution they selected a specific method to try to achieve their goal. It was a hierarchical approach (Swain and Hauska, 1977) using linear discriminant analysis. First, they obtained a 66.37% of general efficiency when classifying six classes (see the reference column in table 2).

Table 2: Classification of breast tissue in six classes.

| | 1st Iteration | 2nd Iteration | Reference |
|---|---|---|---|
| Carcinoma | 95.24 | 95.24 | 81.82 |
| Fibro-adenoma | 40.00 | 46.67 | 66.67 |
| Mastopathy | 83.33 | 83.33 | 16.67 |
| Glandular | 56.25 | 62.50 | 54.54 |
| Connective | 100 | 100 | 85.71 |
| Adipose | 100 | 100 | 90.91 |
| Average | 79.14 | 81.29 | 66.37 |

As the results were poor for six classes, they tried to classify different groups of classes, concluding that the most important was to discriminate carcinoma from fibro-adenoma + mastopathy + glandular tissue.

In this particular problem, they obtained a percentage higher than the previous one: 92.21% overall efficiency and 86.36% for carcinoma discrimination, as it is shown in the reference column of table 3.

With our method we try to solve both the six classes and the two classes problems. We use just the same data from the article mentioned to compare the results obtained. First we take the data matrix and we compute the differences between all the cases (as mentioned in section 2.2). The result is a new matrix with 5565 rows and 9 features variables. One extra variable is added to indicate if the cases compared in each row belongs to the same class or not.

In the first stage a weakly informative Gaussian distribution is used, because we don't have any initial information. With the estimations of the parameters for the probit model, we obtain a value from 0 to 1 indicating their similarity for every pair of case. We want to see if every case is well classified, so for each one we select the top ten similar cases and the most repeated class within this ten is the class assigned to the case studied (nearest neighbor algorithm).

As we propose an interactive learning algorithm, in this experiment we use the posterior distribution of the parameters from an stage to include it as the prior distribution for the next one. The method will be learning and adapting every time the process is repeated. The number of iterations is not fixed at the beginning, so we could compute it until the objective efficiency is achieved.

We start by classifying all the cases into six classes. One solution could be to classify first only between normal or pathological, and then try to identify the specific class. We do it with all the classes at the same time and then we can teach the system to learn from its errors. As Estrela da Silva et al. present in their paper, the most difficult classes to discriminate are fibro-adenoma and mastopathy. The results in table 2 show this fact. In the first iteration we obtain only a 40% correct classification. This is the lowest classification efficiency in the six classes. On the other hand we obtain a perfect classification of connective and adipose tissues and very high percentages for carcinoma (95.24%) and mastopathy (83.33%). The overall efficiency is higher than the one of reference: 79.14% facing 66.37%.

After this first iteration, we have executed the problem once, so the method will be improved with the information we have now. In the next iteration, the information about the parameters of the model is included as prior information and the results of right classified cases increase. We appreciate that the classes that already had a good classification skill remain the same but the two lowest values of classification improve their results. Fibro adenoma goes from 40% to 46.67%, and glandular tissue from 56.25% to 62.50%. With this percentage rising we obtain an overall efficiency of 81.29%, much higher than the 66.37% presented by the other authors. Also the most important class, carcinoma, achieve a 95.24% of correct classified cases against the 81.82% reached in the reference paper.

If we set a stopping criterion, the process will stop when it is reached. For example, if we want at least an overall 80% of efficiency and over 90% for carcinoma class we have achieved this with two iterations. This feedback has only been performed with the training data in order to compare the results with the ones from other authors (see (Jossinet, 1998), (Jossinet and Lavandier, 1998) and (Estrela da Silva et al., 2000)). If new cases appear, the system would have more information and the relevance feedback could be even greater than the one performed.

In order to conclude, we show the results in table 3 for a situation with two classes: carcinoma and fibro-adenoma + mastopathy + glandular. In the first iteration we achieve the 100% correct classifications, so we stop here and there is no need to improve the method with some new iterations. In comparison with the results obtained by Estrela da Silva et al., all the percentages are improved.

Table 3: Classification of breast tissue in two classes.

| | Carcinoma | Others | Percent correct | Reference |
|---|---|---|---|---|
| Carcinoma | 21 | 0 | 100 | 86.36 |
| Others | 0 | 49 | 100 | 94.54 |
| Total | 21 | 49 | 100 | 92.21 |

## 4 CONCLUSIONS

In this paper we have proposed a novel pairwise comparison method based on a Bayesian regression to classify automatically. By extracting color, texture and shape features from the digital images, fair results are obtained for some real pattern recognition problems. The method classifies any new cases and improve the results every time it is executed again, thanks to the relevance feedback. This interactive methodology increases the quality of the results with the supervision of an expert and the readjustment of the prior distribution for the parameters of the model.

## ACKNOWLEDGEMENTS

This research was partially supported by projects

TIN2008-06796-C04-03 and MTM2011-28983-C03-02 from Ministerio de Ciencia e Innovación, and project GRU10100 from Junta de Extremadura.

## REFERENCES

Arias-Nicolás, J. P. and Calle-Alonso, F. (2010). A novel content-based image retrieval system based on bayesian logistic regression. In *18th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2010 (WSCG'2010) Poster proceedings*, pages 19–22, PLZEN, Czech Republic. UNION Agency, Science Press.

Arias-Nicolás, J. P., Calle-Alonso, F., and Horrillo-Sierra, I. (2007a). Qatris imanager, un sistema cbir basado en regresión logística. *Boletín de Estadística e Investigación Operativa*, 23(2):9–13.

Arias-Nicolás, J. P., Pérez, C. J., and Martín, J. (2007b). A logistic regression-based pairwise comparison method to aggregate preferences. *Group Decision and Negotiation*, 17(3):237–247.

Belkasim, S. O., Shridhar, M., and Ahmadi, M. (1991). Pattern recognition with moment invariants: A comparative study and new results. *Pattern Recognition*, 24(12):1117–1138.

Chen, M. H., Spiegelhalter, D., Thomas, A., and Best, N. (2000). *The BUGS Project*. MRC Biostatistics Unit, Cambridge, UK, http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml.

Chorás, R. S. (2003). Content-based retrieval using color, texture and shape information. *Lecture Notes in Computer Sciences*, 2905:619–626.

de la Escalera, A. (2001). *Visión por Computador, Fundamentos y Métodos*. Prentice-Hall.

El-Naqa, I., Yang, Y., Galatsanos, N. P., Nishikawa, R. M., and Wernick, M. N. (2004). A similarity learning approach to content-based image retrieval: application to digital mammography. *IEEE Transactions on Medical Imaging*, 23(10):1233–1244.

Estrela da Silva, J., Marques de Sá, J. P., and Jossinet, J. (2000). Classification of breast tissue by electrical impedance spectroscopy. *Medical and Biological Engineering and Computing*, 38:26–30.

Fix, E. and Hodges, J. L. (1951). Discriminatory analysis, nonparametric discrimination consistency properties. Technical Report 4, USAF School of Aviation Medicine.

Galloway, M. M. (1975). Texture analysis using grey level run lengths. *Computer Graphics and Image Processing*, 4:172–179.

Haralick, R. M. and Shapiro, L. G. (1993). *Computer and Robot Vision Vol.I*. Addison-Wesley.

Jai, A. K., Duin, R., and Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transaction on pattern and machine intelligence*, 22(1):4–37.

Jossinet, J. (1998). The impedivity of freshly excised human breast tisue. *Physiology Measures*, 19:61–75.

Jossinet, J. and Lavandier, B. (1998). The discrimination of excised cancerous breast tissue samples using impedance spectroscopy. *Bioelectrochemistry and bioenergetics*, 45:161–167.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Lineal Models*. Monographs on Statistics and Applied Probability. Chapman and Hall.

Schürmann, J. (1996). *Pattern Classification: A Unified View of Statistical and Neural Approaches*. Wiley and sons.

Smeulders, A., Worring, M., Santini, S., Gupta, A., and Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380.

Smith, A. R. (1978). Color gamut transform pairs. *SIGGRAPH'78 Proceedings of the 5th annual conference on Computer graphics and interactive techniques*, pages 12–19.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 64(4):583–639.

Swain, P. H. and Hauska, H. (1977). The decision tree classifier: design and potential. *IEEE Transaction on Geoscience and remote Sensing*, GE-15:142–147.

Theodoridis, S. and Koutroumbas, K. (2003). *Pattern Recognition*. Elsevier Academia, 4th edition.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4):327–352.

Zellner, A. and Rossi, P. (1994). Bayesian analysis of dichotomous quantal response models. *Journal of Econometrics*, 25(3):365–393.