Use of classifiers and recursive feature elimination to assess boar sperm viability

Lidia Sánchez-González[a,*], Laura Fernández-Robles[a], Manuel Castejón-Limas[a], Javier Alfonso-Cendón[a], Hilde Pérez[a], Hector Quintian[b], Emilio Corchado[b]

[a] *Departamento de Ingenierías Mecánica, Informática y Aeroespacial. Universidad de León, León, Spain*
[b] *Departamento de Informática y Automática, Universidad de Salamanca, Salamanca, Spain.*

Abstract

This paper extends previous work on the assessment of boar sperm cells in order to discriminate amongst intact or reacted acrosomes for fertility purposes. The aim of the study reported is twofold. On one hand to assess the quality of a different set of classifiers. On the other, to assess the feasibility of applying dimension reduction techniques in order to simplify the classification process. The supervised classification techniques used are Extremely Randomized Trees, Random Forest, Support Vector Machines and Gaussian Naive Bayes. The data sets used describe the local maximum gradient, the local mean gray levels and the local standard deviation along the inner contours of the sperm cells. The procedure to obtain these features is explained along as their mathematical nature. The first experiment reported uses each of the three data sets for performing a grid search with 50 fold cross validation in order to evaluate the scores of each classifier. The second experiment reported integrates the three previous datasets into a single one. After performing a recursive feature elimination stage to this data set the results show that only 5 features out of 840 suffice in order to provide satisfactory results according to veterinary experts.

Keywords: boar sperm cell, acrosome state, digital image processing, classification

1. Introduction

Insemination centers use computer based systems in order to estimate how fertile a sample of boar semen is by applying digital image processing [26, 8,

*Corresponding author
Email addresses: lidia.sanchez@unileon.es (Lidia Sánchez-González), l.fernandez@unileon.es (Laura Fernández-Robles), manuel.castejon@unileon.es (Manuel Castejón-Limas), javier.alfonso@unileon.es (Javier Alfonso-Cendón), hilde.perez@unileon.es (Hilde Pérez), hector.quintian@usal.es (Hector Quintian), es.corchado@usal.es (Emilio Corchado), (Emilio Corchado)

6]. Although there are several features that allows experts to make a fertility assessment, the acrosome status is the most significant [22]. A reacted boar sperm cell, which presents a lost acrosome due to the capacitation process, makes impossible to succeed in the fertilization of an oocyte [18] so detecting this status is crucial for insemination purposes [10]. Although it is possible to use staining techniques in order to analyze the acrosome status, it requires qualified staff, specialized devices [23] and even subjective visual estimation what yields an expensive proceduce [23, 17].

The assessment of the acrosome status using digital image processing has been developed in many works using different sets of descriptors. Some authors, González et al. [13], use Discrete Wavelet Transform to extract texture descriptors and classify acrosomes as intact or damaged with hit rates of 92.09%. Other works compute the gradient magnitude along the contour of the sperm cell and classify them using Learning Vector Quantization [20, 1]. Other works compare different texture descriptors as longitudinal and transversal profiles, Hu descriptors, Flusser descriptors and other statistical descriptors with results with a relevant error rate (30.58%) [3]. This error rate is reduced to a 23.20% when texture descriptors named N Concentric Squares is employed, Alegre et al. [4]. Texture features computed from the co-ocurrence matrix and combining the with the Discrete Wavelet Transform achieve an accuracy of 94.93%, Alegre et al. [5]. The hit rate is increased until a 97% when texture descriptors based on the Curvelet Transform are considered, Gonzalez-Castro et al. [15]. With the Hellinger distance the prior probability that minimizes the divergence between the test data distribution and validation distributions generated in a fully controlled way, Gonzalez-Castro et al. [14]. By using an early fusion of texture and contour descriptors and Support Vector Machines a 0.9913F-Score is obtained [11]. When n contours inside the boar sperm cell are considered and a set of features along those contours is computed, results show higher accuracy [24, 2, 25].

The paper is organized as follows. Section 2.1 elaborates on the materials and methods. Section 2.1 describes the data set used in this study and provides details on how the features describing the images are calculated. Section 2.2 describes the experimental procedures and analysis performed on the features collected. Section 2.3 compares these results with previous research. Finally, conclusions are gathered in Section 3.

2. Material and Methods

2.1. Feature extraction

Different features that allows us to describe an acrosome status of a boar sperm cell are proposed in [24]. A subset of such features is used in [2] and [25] to assess the acrosome integrity. This paper uses the same dataset for the experiments. A bigger dataset that includes the considered samples is freely available (see [14]). This dataset is described in the following paragraphs. More detailed information about the image acquisition and feature extraction can be found in [2].

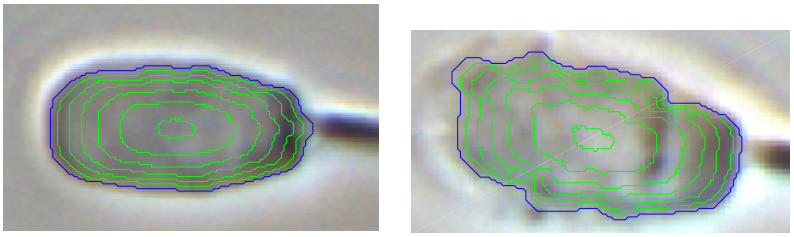

Figure 1: Images of the boundaries and inner contours for both acrosome-damaged (right) sperm cells and acrosome-intact (left) sperm cells.

First, sperm cells are segmented from digital images of semen samples taken by a camera connected to a microscope. After segmentation, the boundary of the sperm head image is obtained and 6 additional concentric contours ($c_i, i = 1, 2, ...6$) inside of the cell boundaries are computed. What defines the concentric contours is the Euclidean distance of each point of the boundary from the centroid (x_c, y_c), that is divided into 6 segments. So, each point at the boundary (x_p, y_p) produces 6 inner points (x_{inner}, y_{inner}) defined as the ones that are $d((x_p, y_p), (x_{inner}, y_{inner}))$ pixels far away from the point of the boundary (x_p, y_p). A logarithmic distance is used since the most of the information lays next to the border of the sperm cell. So, the closer the contours are, the closer the boundary is.

$$d((x_p, y_p), (x_{inner}, y_{inner})) = d((x, y), (x_c, y_c)) * \frac{c_i}{C+1} * (1 - 0.7 * \log_{10}(1 + C - c_i)) \quad (1)$$

being C the number of contours.

Some examples of the computed contours are shown in Fig. 1 for images of both types, with an intact acrosome and with an acrosome that has already reacted.

As it is explained in [2], different sets of texture features for a neighborhood of each point of the contours are computed:

- Local maximum gradient values: the maximum gradient value in a 5-by-5 neighborhood N after computing the scale-dependent gradient [21]:

$$v_i = \max_{(x,y)\in N} \left(\sqrt{\left(f * \frac{\partial g_\sigma}{\partial x} \right)^2 + \left(f * \frac{\partial g_\sigma}{\partial y} \right)^2} \right) \quad (2)$$

- Local mean of the gray level values: the mean of the gray levels of a pixel $I(x, y)$ in a 3-by-3 neighborhood N:

$$v_i = \frac{1}{9} \sum_{(x,y)\in N} I(x, y) \quad (3)$$

- Local standard deviation: the standard deviation of the gray levels in a n-sized neighborhood N (3-by-3), being μ_I the mean of the gray levels for that neighborhood: $\mu_I = \frac{1}{n}\sum_{(x,y)\in N} I(x,y)$:

$$v_i = \sqrt{\frac{1}{n-1}\sum_{(x,y)\in N}(I(x,y)-\mu_I)^2} \tag{4}$$

The features are computed from the point where tail starts along the contours clockwise. The obtained feature vector is interpolated to a constant size of 40 elements. So, each sperm cell produces a feature vector of 280 elements (40 features times 7 contours) describing the local maximum gradient values, another feature vector of 280 elements describing the local means of the gray level values and another feature vector of 280 elements describing the local standard deviation of the gray levels.

2.2. Experimental results

This paper extends those analysis previously performed by [2] using for that purpose the same collection of sperm cells images. Local texture features along C concentric contours of the sperm cells have yielded the best hit rates. However, the set of features is quite big since is formed by 280 features for each image. Here we study the influence of each feature in order to identify the most relevant ones and keep the hit rate but reducing the set of features. Experiments use this data set, formed by 360 sperm cell images, 210 of them have an intact acrosome and 150 have a reacted acrosome. For each image three different data sets are obtained with the features formerly described: local maximum gradient, local mean gray levels and local standard deviation.

[2] performed a classification of the given dataset using k-Nearest-Neighbor classification, class conditional means and Relevance Learning Vector Quantization. The present paper extends that previous one in a twofold manner. Firstly, by using a different set of classifiers; these are Extremely Randomized Trees [12], Random Forest [7], Support Vector Machines [9] and Gaussian Naive Bayes [27]. Secondly, by comparing the results obtained on the separated classification of each feature set to those obtained by combining the three datasets into a single one and drastically reducing the number of features from 840 to 5 by performing a Recursive Feature Elimination process. After such feature reduction was performed, the classification procedures were repeated in order to assess whether the feature sets can be compressed. The experiments shared the same conditions across all datasets as they are described in what follows.

The experiments were run using a personal computer Intel i5, with 4Gb of RAM memory and in those affordable conditions they never took more than 10 minutes to finish. Python 3.6.2 was used as programming language to define the calculations. Specifically, the scikit-Learn library [19] provided a very convenient interface to the classifiers used in this paper. A grid search was performed in order to tune the different parameter of each classifier. The optimality of each of them was measured using cross validation with 50 folds. That number has been

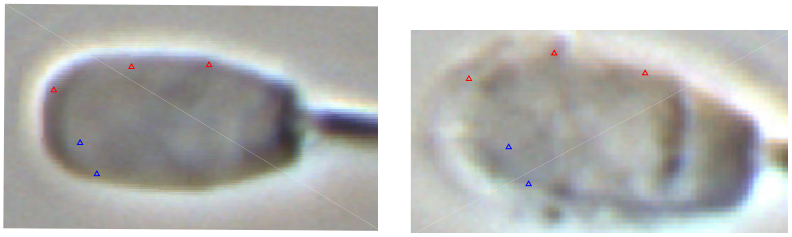

Figure 2: Locations of the most discriminant features as obtained from RFE for both intact (left) and reacted (right) boar sperm cells. With just 2 values of the local maximum gradient (blue) and 3 ones of the local standard deviation (red) the method keeps its performance.

chosen because it allows us to choose a higher number of samples for training keeping an accurate measurement of the error on the test set.

Extremely Randomized Trees (Extra Trees). Ensembles with 5, 10, 20, and 30 Extremely Randomized Trees classifiers were calculated. The rest of the parameters available through the scikit-Learn interface remained on their default values.

Random Forest. Ensembles with 5, 10, 20, and 30 Random Forest estimators were obtained. As before, the rest of the parameters available through the scikit-Learn interface remained on their default values.

Support Vector Machines. Both Support Vector Machines with linear and radial basis functions kernels were calculated. The Gamma parameter was allowed to vary between 0.001 and 0.0001 and the penalty parameter C of the error term between 1 and 10.

Gaussian Naive Bayes. The Gaussian Naive Bayes classifier was run using the sample frecuencies as probability priors. The rest of the parameters remained on their default values.

The first experiment provided good results as perceived by the veterinary experts. They are shown in Table 1. These results confirm that the local maximum gradient values and the local standard deviation values datasets provide the highest quality information in terms of discrimination amongst reacted and intact sperm cells.

In order to assess the quality of this information, a second experiment was run. By combining all the features from the three different datasets and then reducing the number of features of the dataset by an automatic method such as Recursive Feature Elimination [16] we obtained similar results (see Table 1). This is specially remarkable because of the drastic reduction from 840 features to 5. These 5 features are shown on the sperm cells in Fig 2. They correspond with 2 values of the local maximum gradient on the second and third contour (blue marks) and 3 values of the local standard deviation on the outter (first) contour (noted by red marks).

Table 1: Classifier results

		mean score	std deviation
Local maximum gradient dataset	Extra Trees	0.98	0.05
	Random Forest	0.97	0.05
	SVM	0.97	0.07
	Gaussian Naive Bayes	0.96	0.09
Local standard deviation dataset	Extra Trees	0.99	0.03
	Random Forest	0.98	0.05
	SVM	0.99	0.04
	Gaussian Naive Bayes	0.99	0.04
Local mean gray values dataset	Extra Trees	0.96	0.07
	Random Forest	0.97	0.06
	SVM	0.94	0.08
	Gaussian Naive Bayes	0.77	0.20
Recursive feature elimination	Extra Trees	0.98	0.05
	Random Forest	0.98	0.05
	SVM	0.96	0.07
	Gaussian Naive Bayes	0.97	0.07

2.3. Discussion

Table 1 shows that the best cross validation mean results are obtained using the features corresponding to the standard deviation and maximum value of the gradient. These results are similar to those obtained in [2] using a different set of classifiers. As far as the gray level features are concerned, these classifiers provide better cross validation mean results than those reported in [2] since the hit rate achieved is a 97% instead of the previous 95%. Nevertheless, this paper's most significant result is the assessment of the good results that can be obtained on a highly compressed vector with minimum losses.

From a statistical point of view, the interpretation of the results leads to even more optimistic conclussions. Lets consider the errors of the models fitted using cross validation on each of the different feature sets. Figure 3 shows the box and whiskers plots for these errors. The performance shown on the figure provides the following insight: it could be the case where the performance of some of the classifiers could be the same irrespectively of the feature set amongst those considered on this paper. To test this hypothesis an ANOVA and a Kruskal Wallis tests were carried out. Table 3 provides the p-values of these tests. According to these evidences, classifiers such as SVM and Gaussian Naive Bayes clearly display differences on perfomance accross feature sets while Extremely Randomized Trees and Random Forest enjoy statistically similar performances accross feature sets. This result strengthens those obtained on the second experiment reported on this paper and supports the idea of compressing the feature set from 804 features to just 5 without a significant loss of quality on the classification power.

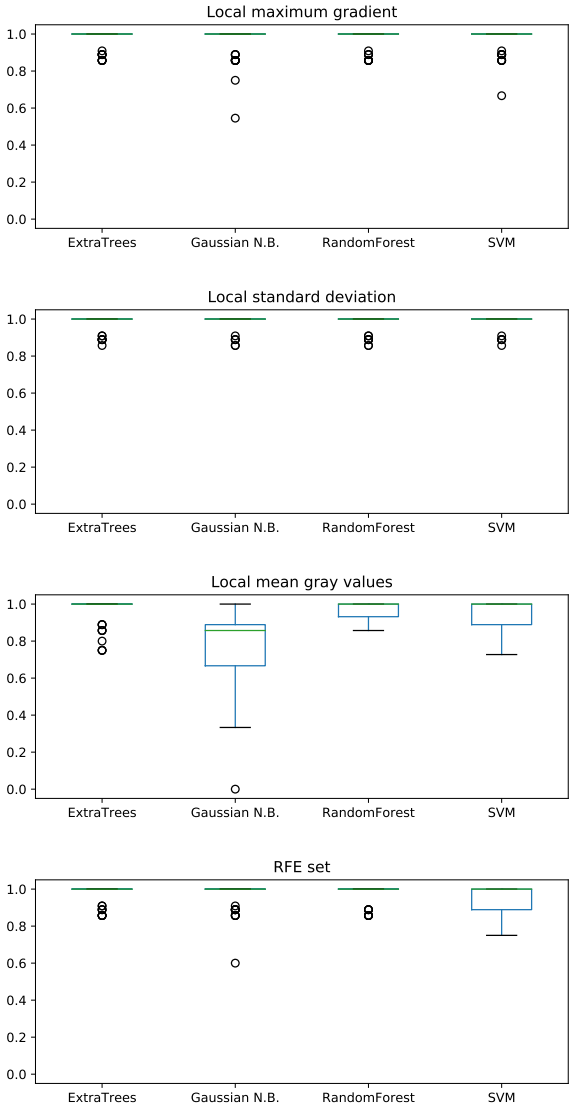

Figure 3: Errors obtained using cross validation on each feature set and technique. Recursive Feature Elimination feature set preserves the good performance in spite of being compressed from 840 to just 5 features.

	ExtraTrees	Gaussian N.B.	RandomForest	SVM
ANOVA	0.519	0.000	0.606	0.002
Kruskal Wallis	0.805	0.000	0.613	0.002

Table 2: p-values obtained on the ANOVA and Kruskal Wallis tests while comparing the performance of the classifiers across the different feature sets

3. Conclusions

In this paper we have tested a set of classifiers in order to assess the acrosome status of a given boar sperm cell so as to determine whether it has lost its acrosome. For that purpose both the individual datasets describing local maximum gradient levels, local mean gray levels and local standard deviation, and an aggregated one combining these three datasets were assessed. The results obtained confirm that it is possible to drastically reduce the number of features used with only marginal losses. Results show that, using just 5 features out of the 840 previously considered, provides statistically similar results on the error rate. That improvement implies reducing the computational cost of the system and greatly simplifies the analysis.

Acknowledgements

Funding: This work was supported by Spanish Ministerio de Economï¿œa, Industria y Competitividad [grant number DPI2016-79960-C3-2-P].

References

[1] Alegre, E., Biehl, M., Petkov, N., Sanchez, L., 2008. Automatic classification of the acrosome status of boar spermatozoa using digital image processing and lvq. Computers in Biology and Medicine 38 (4), 461–468.

[2] Alegre, E., Biehl, M., Petkov, N., Sanchez, L., 2013. Assessment of acrosome state in boar spermatozoa heads using n-contours descriptor and rlvq. Computers Methods and Programs in Biomedicine 111, 525–536.

[3] Alegre, E., Garcia-Olalla, O., Gonzalez-Castro, V., Joshi, S., 2011. Boar spermatozoa classification using longitudinal and transversal profiles (ltp) descriptor in digital images. In: 14th International Workshop on Combinatorial Image Analysis (IWCIA),Madrid, Spain.

[4] Alegre, E., Garcia-Ordas, M., Gonzalez-Castro, V., Karthikeyan, S., 2011. Vitality assessment of boar sperm using ncsr texture descriptor in digital images. In: Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA), La Palma de Gran Canaria, Spain.

[5] Alegre, E., Gonzalez-Castro, V., Alaiz-Rodriguez, R., Garcia-Ordas, M., 2012. Texture and moments-based classification of the acrosome integrity of boar spermatozoa images. Computer Methods and Programs in Biomedicine 108 (2), 873–881.

[6] Bijar, A., nalver Benavent, A. P., Mikaeili, M., Khayati, R., 2012. Fully automatic identification and discrimination of sperm parts in microscopic images of stained human semen smear. Journal of Biomedical Science and Engineering, 384–395.

[7] Breiman, L., 2001. Random forests. Machine Learning 45(1), 5–32.

[8] Chan, J., Krause, W., Bohring, C., 2002. Computer-assisted analysis of sperm morphology with the aid of lectin staining. Andrologia 34 (6), 379–383.

[9] Cortes, C., Vapnik, V., 1995. Support vector networks. Machine learning 20(3), 273–297.

[10] Fazeli, A., Hage, W., Cheng, F.-P., Voorhout, W., Marks, A., Bevers, M., Colenbrander, B., 1997. Acrosome-intact boar spermatozoa initiate binding to the homologous zona pellucida in vitro. Biology of reproduction 56, 430–438.

[11] Garcia-Olalla, O., Alegre, E., Fernandez-Robles, L., Malm, P., Bengtsson, E., 2015. Acrosome integrity assessment of boar spermatozoa images using an early fusion of texture and contour descriptors. Computer methods and programs in biomedicine 120 (1), 49–64.

[12] Geurts, P., Ernst., D., Wehenkel, L., 2006. Extremely randomized trees. Machine Learning 63(1), 3–42.

[13] González, M., Alegre, E., Alaiz, R., Sánchez, L., 2007. Acrosome integrity classification of boar spermatozoon images using DWT and texture techniques. In: VipIMAGE - Computational Vision and Medical Image Processing. Algarve, Portugal, pp. 165–168.

[14] Gonzalez-Castro, V., Alaiz-Rodriguez, R., Alegre, E., 2013. Class distribution estimation based on the hellinger distance. Information Sciences 218, 146–164.

[15] Gonzalez-Castro, V., Alegre, E., Garcia-Olalla, O., Garcia-Ordas, D., Garcia-Ordas, M., Fernandez-Robles, L., 2012. Curvelet-based texture description to classify intact and damaged boar spermatozoa. In: International Conference on Image Analysis and Recognition, Aveiro.

[16] Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Machine Learning 46(1-3), 389–422.

[17] Neuwinger, J., Behre, H., Nieschlag, E., 1990. External quality control in the andrology laboratory: an experimental multicenter trial. Fertil. Steril. 54 (2), 308–314.

[18] Oliva-Hernandez, J., Corcuera, B., Perez-Gutierrez, J., 2005. Epidermal growth factor (egf) effects on boar sperm capacitation. Reproduction in Domestic Animals 40, 353.

[19] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

[20] Petkov, N., Alegre, E., Biehl, M., Sanchez, L., 2006. Lvq acrosome integrity assessment of boar sperm cells. In: Proceedings of Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applications (CompIMAGE).

[21] Petkov, N., Alegre, E., Biehl, M., Sanchez, L., 2006. LVQ acrosome integrity assessment of boar sperm cells. In: Proceedings of Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applications (CompIMAGE).

[22] Petrunkina, A., Petzoldt, R., Stahlberg, S., Pfeilsticker, J., Beyerbach, M., Bader, H., T´opfer-Petersen, E., 2001. Sperm-cell volumetric measurements as parameters in bull semen function evaluation: correlation with nonreturn rate. Andrologia 33, 360–367.

[23] Pinart, E., Bussalleu, E., Yeste, M., Briz, M., Sancho, S., Garcia-Gil, N., Badia, E., Bassols, J., Pruneda, A., Casas, I., Bonet, S., 2005. Assessment of the functional status of boar spermatozoa by multiple staining with fluorochromes. Reproduction in Domestic Animals 40, 356.

[24] Sanchez, L., 2007. Boar sperm cell classification using digital image processing. Ph.D. thesis, University of Leon, Spain.

[25] Sánchez, L., Quintián, H., Alfonso-Cendón, J., Pérez, H., Corchado, E., 2016. Use of support vector machines and neural networks to assess boar sperm viability. In: International Joint Conference SOCO'16-CISIS'16-ICEUTE'16 - San Sebastián, Spain, October 19th-21st, 2016, Proceedings. pp. 13–19.

[26] Verstegen, J., Iguer-Ouada, M., Onclin, K., 2002. Computer assisted semen analyzers in andrology research and veterinary practice. Theriogenology 57, 149–179.

[27] Zhang, H., 2004. The optimality of naive bayes. Proc. FLAIRS.