ELSEVIER

Computers in Biology and Medicine 38 (2008) 1177–1186

Computers in Biology and Medicine

www.intl.elsevierhealth.com/journals/cobm

Selection of human embryos for transfer by Bayesian classifiers

Dinora A. Morales[a,*], Endika Bengoetxea[a], Pedro Larrañaga[b]

[a]*Intelligent Systems Group, University of the Basque Country, Donostia-San Sebastián, Spain*
[b]*Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Madrid, Spain*

Received 22 August 2007; accepted 12 September 2008

Abstract

In this work we approach by Bayesian classifiers the selection of human embryos from images. This problem consists of choosing the embryos to be transferred in human-assisted reproduction treatments, which Bayesian classifiers address as a supervised classification problem. Different Bayesian classifiers capable of taking into account diverse dependencies between variables of this problem are tested in order to analyse their performance and validity for building a potential decision support system. The analysis by receiver operating characteristic (ROC) proves that the Bayesian classifiers presented in this paper are an appropriated and robust approach for this aim. From the Bayesian classifiers tested, the tree augmented naive Bayes, k-dependence Bayesian and naive Bayes classifiers showed to perform almost as well as the semi naive Bayes and selective naive Bayes classifiers.

Keywords: Bayesian classifiers; Supervised learning; Embryo image classification; In vitro fertilisation; Bayes methods

1. Introduction

Developments in medical technology have led to a number of methods designed for assisted human reproduction. Procedures such as artificial insemination, in vitro fertilisation (IVF) and embryo transfer increasingly provide new options in assisted reproductive technology. Assisted human reproduction deals with the social infertility problem. For various reasons, the number of embryos developed through the assisted reproductive techniques is usually greater than the number of the ones that can be transferred. This is the case for instance in Spain, where the law only permits three embryos to be transferred at most in each treatment. As another example, Italian law also permits up to three oocytes to be fertilised, although it is not allowed to freeze them for a possible future treatment.

Many studies have shown that morphological structures in the embryo can be used as bio-markers of embryonic quality [1–4], and that embryo selection based on morphology assessment is relevant to improve implantation and pregnancy rates [1,5]. Most of the existing scoring systems are based on combinations of several morphological parameters such as cleavage stage, embryonic fragmentation and blastomere uniformity [1,2,6]. In particular, the morphology of pronuclear oocyte has been linked with implantation and development to the blastocyst stage [7]. Embryologists select the embryos on the basis of subjective light microscopic morphological analysis.

Data mining techniques such as decision trees and the construction of predictive statistical models have been used previously on IVF to assist on the selection of the most promising oocyte for implantation. The literature contains a number of such studies with this goal. Saith et al. [8] apply decision trees to investigate the relationship of the features of the embryo, oocyte and follicle to the successful outcome of the embryo transfer. Trimarchi et al. [9] studied the models based on data mining techniques, in particular applying the C5.0 algorithm for inferring classification trees [10]. Jurisica et al. [11] presented an application of the TA3 system as an intelligent decision support system for IVF practitioners that, in some situations, can suggest possible treatments to improve the success rate. Patrizi et al. [12] presented a pattern recognition algorithm to select embryos from images, which tries to classify the objects given into a number of classes and to formulate from those

* Corresponding author. Tel.: +34 943 01 80 70; fax: +34 943 01 55 90.
E-mail address: dinora-morales@ehu.es (D.A. Morales).

0010-4825/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2008.09.002

a general rule. Manna et al. [13] compared the precision in the recognition of viable embryos by a group of experts to that of a machine recognition procedure.

The aim of this study is to apply for the first time Bayesian classifiers to information extracted from embryo images in order to predict their viability to succeed implantation on woman's uterus, as well as to overcome the unbalanced data set problem that is so common in clinical databases of infertility treatments. In other works in the literature Bayesian classifiers are applied for identifying the most relevant morphological and clinical variables that determine the outcome of IVF treatments [14], but the use of Bayesian classifiers to embryo images for classifying embryos according to the estimation of their implantation probability has not been analysed previously. The final objective is to develop a system able to do it automatically which could be used as a decision support system, based on non-invasive, precise and detailed analysis of human embryo morphology. This decision support system could at the same time be used as a training tool for novel embryologists, and therefore to help on improving success rates of infertility treatments.

The paper is organised as follows: the next section revises the problem of selecting embryos for transfer on infertility treatments as a supervised classification one. Section 3 describes digital acquisitions and provides a brief overview of the Bayesian classifiers used in this work. Section 4 shows experimental results and their interpretation. Finally in Section 5 conclusions and some future work lines are presented.

2. The problem in selection of human embryos

Infertility is defined as the state that normal couples, not using any form of contraceptive measures, fail to become pregnant over at least one year. Human-assisted reproduction methods like IVF through insemination of oocyte and sperm and the process called intracytoplasmic sperm injection (ICSI)—where the sperm is injected into oocyte—have been widely used to treat infertility. Oocyte and sperm are obtained separately in fertilisation treatment process. In order to obtain a sufficient number of oocyte ovulatory stimulants are used. These stimulants make pituitary to increase secretion of follicle stimulating hormones. Later, the first days after the fertilisation embryos are developed in an embryo culture medium under a controlled atmosphere. A few embryos, the ones deemed best by the clinician and embryologist regarding the likelihood of bringing forth a child, are chosen and transferred to the woman's uterus within 12–72 h from their formation.

An essential problem in human-ssisted reproduction is the selection of the embryos suitable to be transferred in a patient. Embryologists analyse the embryo by non-invasive techniques by inverted microscopy with Hoffman contrast. In usual practice the selection of human embryos—as well as their morphological characterisation—is based on the subjective judgement of the embryologist. Furthermore, the time is a limiting factor when analysing the embryo, which makes it difficult to investigate embryo morphology in detail for its transfer. That is why, the training of embryologists and the need to establish homologated criteria are essential to avoid as much as possible the subjectivity of the evaluation and selection of embryos. At the same time, the objectives of the task are more complex since we want to maximise implantation rates while limiting the incidence of multiple pregnancies.

The aim of intelligent methods is to support the selection process and to choose from the available embryos the few of them which have a good quality and the greatest potential to implant in the uterus. Even if the final objective is that the embryo will later develop into a live child, on IVF treatments implantation is regarded by itself as a success since later pregnancy follow up is considered as a different gynaecological question.

In the present study we apply for the first time Bayesian classifiers to recognise potentially good embryos on the basis of a training sample set of images, in order to predict for each new embryo its outcome (if it will succeed on implanting or not). The supervised classification decision is based on the feature vector extracted from embryos morphology of a database of previous treatments and their class (outcome). Bayesian classifiers have demonstrated good precision in complex medical problems [15], which makes them likely to be applied on this new complex domain. In particular, the use of Bayesian classifiers able to consider higher degree dependencies is especially suited for obtaining better classification results in this type of complex problems in which the features are not conditionally independent given the class.

2.1. Selection of embryos as a supervised classification problem

In order to apply Bayesian classifiers, we regard the embryo selection as a supervised classification problem in which, having as a starting point an embryo image, the classifier has to provide an estimate on its potential or suitability to achieve successful implantation if it is transferred to the uterus.

In order to build the classifiers from a database of embryo images selected on IVF treatments—i.e. with known outcome—we propose to divide all images on two classes: images of embryos that managed to implant to class 1, and to class 0 otherwise. For each image a pattern vector of embryo features is defined by regarding each image as an array of intensities of grey-levels from which the frequency distribution of the grey tones for each image was obtained.

A comprehensive procedure to extract embryo characteristics from an image is defined in [12]. These authors propose to apply to the embryo image a procedure that takes into account the central moments or moment invariants which was first proposed in [16] to extract the shape and texture features of an image. These cental moments are invariant with respect to translation, rotation and scaling of the pattern represented in an image, and they can also be applied to binary or grey level images. In Biomedical research the central moments have been applied to obtain the characteristics of cell images [17] or medical images [18,19].

Given the image of an embryo, the marginal distribution of pixel values in the horizontal and vertical directions is determined and from each distribution the first six central moments

(polynomial functions of central tendency and spread) were calculated. This pixel-intensity profile indicates the homogeneity of the image—whether it has many dark and light spots and how they are distributed. The first moment was ignored, since it is zero. These values were used to form the pattern vector of each embryo that consists of eleven values: the ten central moments (extracted from its image) and the embryo class (1 or 0, i.e. implanted or not implanted). This data acquisition is the one proposed in the literature by Patrizi et al. [12].

More formally, the problem of predicting the outcome of a human embryo by means of supervised classification can be defined as follows. We denote by $\boldsymbol{x}=(x_1, \ldots, x_n)$ the vector of predictor variables—representing the characteristics of a concrete embryo—for the problem of classifying an embryo image into one of the two classes of variable C. The true class—i.e. the real outcome if we transfer this embryo—is denoted by c and it takes values in the domain set $\{0, 1\}$: Each pattern vector represents the embryo morphological features, assigned to class 1 if the embryo is suitable for transfer, and to class 0 otherwise. The supervised classification problem consists on creating a model that assigns any pattern vector $\boldsymbol{x}=(x_1, \ldots, x_n)$ to one of the two classes of variable C. Statistical classifiers such as Bayesian classifiers provide an estimate of $p(c|\boldsymbol{x})$, the probability that an embryo image with predictor vector $\boldsymbol{x}=(x_1, \ldots, x_n)$ belongs to class $c \in C$. We can regard the classifier as a $\gamma : (x_1, \ldots, x_n) \rightarrow \{0, 1\}$ function that assigns labels to observations. In supervised classification the objective is to build a classifier that minimises the total error cost by taking into account the joint probability distribution $p(x_1, \ldots, x_n, c) = p(c|x_1, \ldots, x_n)p(x_1, \ldots, x_n) = p(x_1, \ldots, x_n|c)p(c)$ which is unknown a priori. According to [20], for the particular case of a symmetric cost function, the total error cost is minimised when assigning to the example $\boldsymbol{x} = (x_1, \ldots, x_n)$ the class $\gamma(\boldsymbol{x})$ with the highest a posteriori probability:

$$\gamma(\boldsymbol{x}) = arg \max_c p(c|x_1, \ldots, x_n) \quad (1)$$

In the case of having different misclassification costs, the a posteriori probability is compared to a threshold t to predict the class of $\boldsymbol{x}$. This threshold is usually chosen to minimise the expected misclassification loss [21], resulting in the following classification rule: classify $\boldsymbol{x}$ to class 1 if and only if $p(C=1|\boldsymbol{x})$ is greater or equal than threshold t.

3. Material and methods

3.1. Selection of embryos and recording of digital images

The embryo images in our experiments as well as their characterisation by central moments are the ones obtained for the study presented by Patrizi et al. [12] in which the data and its validation are approached in a different way. The treatment of artificial insemination was conducted at the Genesis Center in Rome during the period January 1998 to December 2001, whose database included 275 cycles of ICSI for 195 patients [12]. Note that one of the limitations of our study is that we could not fully ensure complete independence of samples (i.e. some different embryo images correspond to a same couple).

For this study 249 embryos were selected and photographed. Embryo images at the 4-cell stage were taken 40–50 h after fertilisation and before transfer. Embryos were catalogued according Mils's score [22]. Fig. 1 shows examples of embryo images obtained following this procedure which were used as imported data on this study.

3.2. Bayesian classifiers

Bayesian classifiers have been chosen since they have already demonstrated a good precision in other complex medical problems [15], and also because these models are transparent and comprehensive for medical practitioners. Our concrete problem of predicting the viability to succeed implantation of an embryo is addressed as a supervised classification problem to allow the use of these classifiers. Bayesian classifiers are able to provide a concrete probability estimation of the capability of each embryo to be implanted. The final aim is to have a system able to support the selection of the most promising available embryos in order to choose the few of them which have good quality and the greatest potential for implantation based on the feature vector of embryo morphology obtained from an embryo image.

We present next some of the classifiers in the form of Bayesian networks [23] that have been proposed in the literature. Paradigms such as naive Bayes [24], selective naive Bayes [25], semi naive Bayes [26], tree augmented naive Bayes (TAN) [27] and the k-dependence Bayesian (kDB) classifier [28] are thought specifically for supervised classification problems.

Bayesian classifiers need to be constructed using a score that measures the goodness of each configuration—i.e. classifier's structure showing relationships between all the variables. One of the main advantages of Bayesian classifiers is that their structure is an intuitive graphical representation, allowing to visualise the underlying probabilistic classification process and to provide a set of properties that can be appreciated by medical staff. Their graphical structure facilitates the interpretability and understanding by clinicians, reflecting probabilistic relationships between domain variables. In our concrete problem, the conditioned and marginal probabilities of the model can be of interest to embryologist who want to better understand the uncertainty of this medical domain.

A model hierarchy of increasing structural complexity can be established for Bayesian classifiers, where the naive Bayes is at the bottom and a general Bayesian network is at the top of this hierarchy. Fig. 2 illustrates examples of some Bayesian classifiers models due to dependencies that are used in this paper.

The *naive Bayes* classifier [24] applies the Bayes theorem to predict for each unseen instance $\boldsymbol{x}$, the class $c \in C$ for which it has a higher a posteriori probability. This a posteriori probability is computed as

$$p(c|\boldsymbol{x}) \propto p(c, \boldsymbol{x}) = p(c) \prod_{i=1}^{n} p(x_i|c) \quad (2)$$

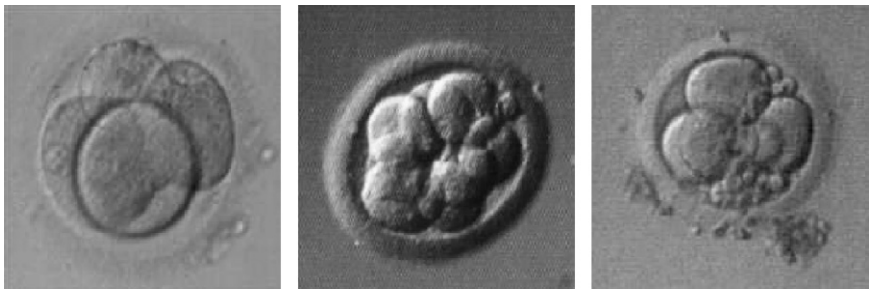

Fig. 1. Real embryo images of the database of Genesis Center (Rome, Italy), catalogued following the Mills's score [22] commonly accepted as an embryo quality cataloguing standard.

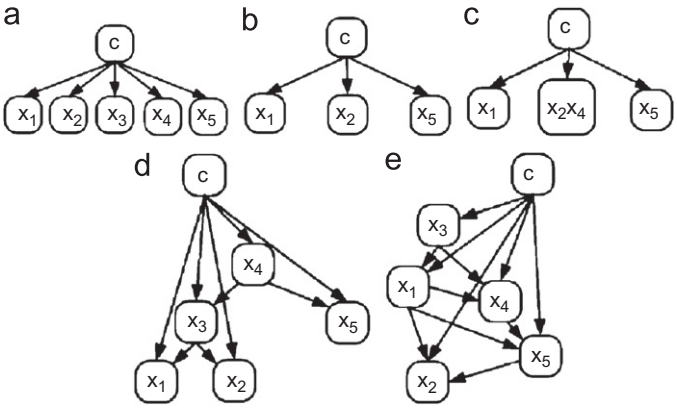

Fig. 2. Examples of the structure of different Bayesian classifiers for a concrete problem with five predictor variables. (a) Naive Bayes, (b) selective naive Bayes, (c) semi naive Bayes, (d) TAN, and (e) *k*DB.

where $p(x_i|c)$ represents the conditional probability of $X_i = x_i$ given that $C = c$ when all variables have discrete values. As a result, the naive Bayes classifier follows the following approach:

$$c^* = arg\max_c p(c)\prod_{i=1}^{n} p(x_i|c) \tag{3}$$

This paradigm has always the same structure: all predictor variables are included in the model. A naive Bayes classifier structure example is shown in Fig. 2(a) for a problem with five variables.

Despite the success of the naive Bayes classifier for some problems, in many real problem domains the predictive accuracy of learning algorithms is degraded by irrelevant predictor variables, where the information contribution is overlapped or repeated. The *selective naive Bayes* classifier [25,29] is able to detect irrelevant and redundant variables, even if similarly as naive Bayes it cannot notice dependencies between predictive variables.

The selective naive Bayes algorithm is the result of applying feature subset selection (FSS) techniques to the naive Bayes classifier. The final model of the selective naive Bayes approach contains some of the predictive variables, allowing the rest of predictor variables to be discarded and not be included in the Bayesian classifier.

In the classical literature the selective naive Bayes classifier is built following one of the following two standard ways: forwardly starting with an empty set of variables and adding them one by one, or backwardly by removing in each iteration one of the variables that is discarded. The forward sequential selection wrapper algorithm is one of the former possibilities, which stars with an empty set of variables. At each step the model adds the most accurate variable calculated by estimated accuracy [30] and stops when no improvement is obtained.

As an example of applying the selective naive Bayes classifier, if we consider the selective naive Bayes classifier illustrated in Fig. 2(b), the representation of an instance $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5)$, would be assigned to the class

$$c^* = arg\max_c p(c)p(x_1|c)p(x_2|c)p(x_5|c) \tag{4}$$

The selective naive Bayes algorithm is able to detect irrelevant and redundant variables, although no dependency between the variables present in the structure are taken into account. However, in most of real problems relationships between variables exist and need to be considered for a good classification performance. For this reason, other Bayesian classifiers that overcome the conditional independence assumption between variables have been developed.

In *semi naive Bayes* [31] a new kind of variable—a joint variable—is built via the cartesian product of a subset of variables. The variable is represented as a single node in the Bayesian network, allowing to surpass the assumption of conditional independence required in the literature [26] between the variables that are included in a same node. Each joint node represent a new variable that considers all the dependencies between the original variables that form it.

The semi naive Bayes classifier requires an algorithm to build the Bayesian network structure. The forward sequential selection and joining (FSSJ) algorithm [26] is an example of techniques that can be applied for this purpose. This starts with an empty structure to which new nodes or new variables fused in existing nodes are added iteratively until non-improvement of the performance in terms of estimated accuracy is reached.

As an example, Fig. 2(c) shows a possible semi naive Bayes model that could have been induced using this approach. Under this classifier, the pattern $\boldsymbol{x}=(x_1, x_2, x_3, x_4, x_5)$ will be assigned

to the following class:

$$c^* = arg\max_c p(c)p(x_1|c)p(x_2, x_4|c)p(x_5|c) \quad (5)$$

The *tree augmented naive Bayes* (TAN) [27] is another Bayesian network classifier that allows dependencies between variables. The main restriction on the dependencies that this Bayesian classifier can take into account is that each predictive variable can have a maximum of two parents: the class variable C and one of the other predictive variables $X_1, \ldots, X_n$.

In order to build the TAN Bayes classifier's structure we need previously to learn the dependencies between the different variables $X_1, \ldots, X_n$. This algorithm applies a score based on the information theory, and the weight of an arc (X_i, X_j) is defined by the mutual information measure conditioned to the class variable as

$$I(X_i, X_j|C) = \sum_c \sum_{x_i} \sum_{x_j} p(x_i, x_j, c) \log \frac{p(x_i, x_j|c)}{p(x_i|c)p(x_j|c)} \quad (6)$$

Using the mutual information of each predictive variable and the class—$I(X_i, C)$—and the conditional mutual information of each pair of domain variables given the class—$I(X_i, X_j|C)$—the algorithm builds a tree structure.

Once having computed these values, the Bayesian network structure of the TAN Bayesian classifier is built in two phases, the first of which starts from an undirected complete graph where the nodes correspond to the predictor variables $X_1, \ldots, X_n$. The algorithm assigns the weight $I(X_i, X_j|C)$ to the edge connecting variables X_i and X_j. The tree is constructed by assigning the edge with the highest conditional mutual information. This process is repeated adding higher score edges unless a possible addition forms a loop, in which case that edge is discarded and the next edge is analysed. The procedure ends when $n - 1$ branches have been added. Finally the undirected graph is transformed in a directed one by choosing a random variable as the root.

In a second phase the structure is augmented to the naive Bayes classifier by adding a node labelled as C, and adding one arc from C to each of the predictor variables X_i $(i = 1, \ldots, n)$. Fig. 2(d) shows an example of a TAN classifier structure that could be induced using this approach, where an instance $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$ would be assigned to the class

$$c^* = arg\max_c p(c)p(x_1|c, x_3)p(x_2|c, x_3)p(x_3|c, x_4)p(x_4|c) \times p(x_5|c, x_4) \quad (7)$$

The *k-dependence Bayesian classifier* (kDB) [28] tries to avoid the restriction of TAN structure which limits the number of parents that each predictive variable can have to a maximum of two (the class and another predictive variable). In this approach, every predictive variable is allowed to have up to k parents besides the class-node. The main characteristic of a kDB structure is the fact that it is the user who fixes the restrictive condition of the value of k which represents the maximum number of parents per variable.

The kDB structure is built using mutual information—$I(X_i, C)$—and conditional mutual information—$I(X_i, X_j|C)$—scores. The procedure starts uniquely with the class-node C in the structure. Each iteration the algorithm selects the node not included in the structure with highest $I(X_i, C)$, its corresponding arc from C to X_i is added, and the value $I(X_i, X_j|C)$ is computed for all the possible new arcs from the X_j nodes already inserted in the structure. All these arcs are ordered from the highest to lowest, from which the highest k nodes are added to the structure (or all of them if the structure contains so far k or less nodes excluding C). Fig. 2(e) shows an example of a kDB classifier structure induced using this approach.

As an example of the result of a kDB classifier structure that could be obtained applying this procedure, if we consider the selective naive Bayes classifier illustrated in Fig. 2(e) as the representation of an instance $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5)$, this would be assigned to the class

$$c^* = arg\max_c p(c)p(x_1|c, x_3)p(x_2|c, x_1, x_5)p(x_3|c) \times p(x_4|c, x_1, x_3)p(x_5|c, x_1, x_4) \quad (8)$$

The main difference between all these classifiers is the number of interdependencies between variables that they can take into account, since higher order classifiers are able to consider more interdependencies between the different variables for each problem—i.e. TAN is able to consider up to 2 conditional dependencies per variable while kDB can consider up to k. Therefore, in complex problems in which the features have relevant conditional dependencies—such as in our embryo selection problem in which morphological variables are conditionally dependent to each other given the class—the last classifiers are more likely to perform better.

3.3. Unbalanced data sets

The unbalanced data set issue raises from the fact of having an unbalanced class distribution, that is, when the original data set from which the classifier is to be built has different proportion of cases for each of the classes. This is the case for the embryo selection problem, since clinical data contains always many more embryos that did not manage to implant than examples of successful ones, making databases highly unbalanced. This affects directly the performance of classifiers since they usually tend to avoid the less represented class, sometimes to the extent of practically never considering it if the unbalance is high.

Many research on how to handle highly unbalanced data sets has been focused on modifying the class distribution of the training set. Since our interest is to validate Bayesian classifiers for their application to this particular problem, an alternative for evaluating the performance of the classifier under these conditions is to apply the receiver operating characteristic (ROC) analysis [32] to measure the cost-benefit ratio of diagnostic decision making. Furthermore, the ROC curve is proven to be a better evaluation measure than accuracy in problem domains with unbalanced class distributions [21,33,34].

3.4. Validation of classifiers

ROC curves are frequently used in biomedical informatics research to evaluate computational models for decision support,

diagnosis, and prognosis [34]. ROC curves were introduced in machine learning by Spackman [35]. ROC curves have properties that make them especially useful as a tool for evaluating and comparing supervised classification algorithms [36,37].

ROC curves are produced by varying the threshold t between 0 and 1. This threshold value t is usually chosen to minimise the expected misclassification loss. The ROC curve allows to view graphically the performance of a classifier by plotting the sensitivity—which in our case is the proportion of embryo images correctly classified as being suitable for implantation—versus 1-specificity—the proportion of good embryos images misclassified as no suitable. Different points in the curve correspond to different values of the threshold on the class probability model. Several different indices can be calculated in order to summarise the accuracy of a classifier.

In ROC analysis the full area under the ROC curve (AUC) is the most commonly used ROC index [38]. AUC is computed by non-parametric, parametric or re-sampling methods to evaluate and compare the performance of different classifiers. The AUC obtains some value between 0.5—associated to the diagonal of the square—and 1—corresponding to the curve that follows the top-left corner. In ROC curves, the accuracy of a classifier is estimated by the fraction of cases correctly classified and is determined by the appropriated chosen threshold. Recent years have seen an increase in the use of ROC graphs on the machine learning community. In addition to being a generally useful performance graphical method, they have properties that make them especially useful for domains with unbalanced data sets and unequal classification error costs [21,34]. These characteristics have become increasingly important as research into the areas of cost-sensitive learning and learning in the presence of unbalanced classes.

We evaluate the performance of Bayesian classifiers with ROC analysis in order to show their performance and to get the estimated accuracy of the classifier over seen instances. In our case, the areas under ROC curves were calculated by the trapezoidal rule [39] non-parametric method. This method is equivalent to the Mann–Whitney U statistics.

A technique to estimate the accuracy of classifiers is the k-fold stratified cross-validation [30]. In this method a data set S is partitioned into k folds such that each class is uniformly distributed among the k folds: As a result, the class distribution in each fold is similar to such of the original data set S, and each fold can be used as a test set. The test set serves the role of providing new data and the rest of the folds are used for training. This procedure is repeated k times. In our case, we applied 10-fold cross-validation to get the estimated accuracy of each Bayesian classifier.

4. Experimental results

For our study, we applied a database containing continuous variables where each case corresponds to morphological embryo features, and the categorical variable class. The Bayesian classifiers introduced previously are implemented to manage uniquely discrete data. Therefore, a pre-process step is applied to discretise the data, by means of the equal frequency algorithm [40] into three intervals. The Elvira software [41] is used in the implementation of the previously presented Bayesian classification models. In order to validate the Bayesian classification models a k-fold stratified cross-validation method is performed [30], and the *estimated accuracy* for each classifier is computed using this method.

Table 1
Results the best accuracy for selection of human embryo classification by Bayesian classifiers .

Classifier	Accuracy ± SD
Naive Bayes	0.8549 ± 0.1757
Selective naive Bayes	0.8333 ± 0.1610
Semi naive Bayes	0.7848 ± 0.0262
TAN	0.8917 ± 0.1580
*k*DB	0.8833 ± 0.1285
TAN-wrapper	0.9125 ± 0.1186
*k*DB-wrapper	0.8763 ± 0.1124

We applied a cost sensitive learning by varying the decision threshold values, in our case 1000 values have been taken into account for Bayesian learning. This process is similar to vary the prior probability of each class on the training set or the costs of errors on each class [42]. Cost sensitivity learning by varying the decision threshold allows the minimisation of the Bayes error rate, defined by the overlap between the two distributions of class values [43].

We performed our experiments with the same database used in the literature in [12], which contains the information of the six central moments extracted from a total of 249 embryo images. These are formed by a total of 35 successful—i.e. implantation was obtained—and 214 unsuccessful cases of embryos that were transferred.

We investigated the performance of different Bayesian classifier when applied to the problem of selecting the human embryo for transfer to the uterus. Table 1 shows the results after evaluating the estimated accuracy of seven Bayesian classifiers by stratified 10-fold cross validation. In TAN and *k*DB algorithms introduced in Section 3.2, the search of the best model is guided by the mutual information criterion. Alternatively to this, the algorithms TAN-wrapper and *k*DB-wrapper models [15] have been implemented, in which the search in the space of possible graph structures of TAN and *k*DB models, respectively, is guided by another score based on an estimated classification goodness measure. Our experiments show that the TAN-wrapper classifier obtained the best accuracy while semi naive Bayes resulted in the worst classifier.

Fig. 3 shows the ROC analysis for evaluation of the previously defined Bayesian classifiers. The experiment was repeated ten times to calculate by k-fold cross-validation the AUC at 95% of confidence. Each figure was plotted with one hundred points and shows ten error bars at 95% confidence intervals. Looking on Fig. 3 it must benoted that *TAN-wrapper*, *kDB-wrapper*, *TAN* and *k*DB classifiers showed the highest AUC. The *k*DB classifier reaches the best sensibility of 82% with a specificity of 80%. The *TAN-wrapper* classifier reaches higher

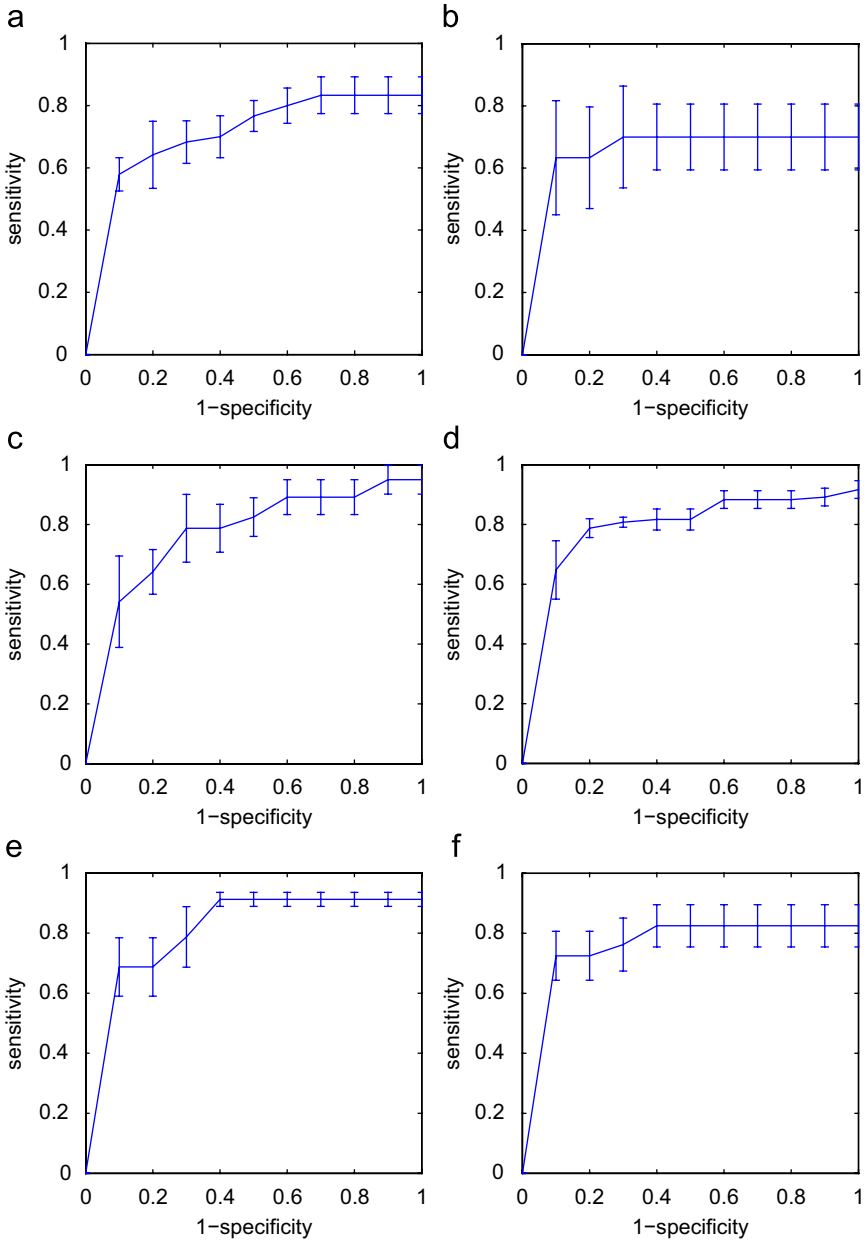

Fig. 3. Results on ROC curve for the different Bayesian classifiers at 95% confidence based on 10 runs. (a) Naive Bayes, (b) selective naive Bayes, (c) TAN, (d) *k*DB, (e) TAN-wrapper, and (f) *kDB*-wrapper.

sensibility of 92% with specificity 60%, and the classifier with highest specificity of 90% and the best sensitivity of 74% is *kDB-wrapper*.

The fact that the two Bayesian classifiers performing best are TAN and *k*DB is not a random. These two classifiers do not assume predictor variables to be conditionally independent to each other given the class variable, and are characterised by the fact that the resultant Bayesian classifier structure is able to represent conditional dependencies between variables. We can conclude that this information is relevant for providing a better classification result to predict the implantation outcome from data extracted out of an embryo image. Table 2 shows the areas under these curves and their confidence intervals at 95% confidence bound. The *TAN-wrapper*, *kDB*, *kDB-wrapper* and TAN have the highest AUC with 0.9994, 0.9991, 0.9819 and 0.9454 values, respectively. In the traditional Bayesian classifiers like *semi naive Bayes*, *selective naive Bayes* and *naive Bayes*, their AUCs takes values of 0.7008, 0.8483 and 0.8918, respectively.

Finally, in order to compare Bayesian classifiers to a baseline method that has been applied widely in the literature for the embryo selection problem, we performed the same experiment with logistic regression [44–46]. Using the default parameters of the WEKA workbench we applied a logistic regression model with a log-likelihood ridge estimator value of 1.0E−8 to the same data. The accuracy obtained following the same validation method as with Bayesian classifiers was 82.23%, and

Table 2
Areas under ROC curves (AUC) at 95% confidence for each of the different Bayesian classifiers when applied to the selection of human embryos.

Classifier	AUC ± SD
Naive Bayes	0.8918 ± 0.0718
Selective naive Bayes	0.8483 ± 0.0175
Semi naive Bayes	0.7008 ± 0.0643
TAN	0.9454 ± 0.0785
*k*DB	0.9991 ± 0.0525
TAN-wrapper	0.9994 ± 0.0589
*k*DB-wrapper	0.9819 ± 0.0264

with 0.535 of AUC. Since the results obtained by Bayesian classifiers improves these values—except from the accuracy of semi naive Bayes) once again the validity of our approach is confirmed.

5. Conclusions and future work

This paper constitutes the first Bayesian approach for the estimation of the implantation probability of embryos in artificial insemination treatments as from embryo images, and we investigated here the ability of Bayesian classifiers to predict the suitability of an embryo to succeed implantation if it is chosen for being transferred. The final objective is to support the selection of the most promising available embryos in order to choose the few of them which have good quality and the greatest potential for implantation. Bayesian classifiers could be used to build decision support systems for embryologists to address the problem of selection of embryos for transfer in human-assisted reproduction treatments. Here different Bayesian classifiers have been evaluated and compared by ROC analysis to overcome the unbalance data set problem.

Our results for Bayesian classifiers are positive enough as to allow considering them satisfactorily for this problem. From all the classifiers, the ones that performed best are TAN and *k*DB, the only two which are able to take into account conditional dependencies between variables to bring a conclusion, which evidences the complexity of the problem and the relevance to consider in order to be able to improve the classification accuracy.

If we compare our results to the ones obtained in the literature for the embryo selection problem, we obtain a good performance of Bayesian classifiers, in particular for TAN with wrapper feature selection with 91% of accuracy. The lowest accuracy with our Bayesian classifiers the 78% of the semi naive Bayes classifier, which is still comparable to the 82.33% obtained for the same data using logistic regression, and also to the 81.79% on the study in [12] with the original data set using a different validation method. Regarding other studies, Saith et al. [8] obtained a 74% of accuracy applying the second rule into the classification tree algorithm. Therefore, our work proves that Bayesian classifiers are a valid approach for the problem of classifying embryos for posterior transfer.

Additionally, we have studied here the expected and best performance of the Bayesian classifiers, considering the unbalance on the data set by analysing the ROC curves as well as the threshold for the best accuracy of the ROC curve solutions.

In our experiments, we have not yet considered other methods for extraction of characteristics from images. This has been left for future work.

Other future work trends in this direction include the acquisition of new data that includes morphological embryo features from human embryo images and clinical data of the patient, since increasing the number of predictor variables that may improve the efficiency of Bayesian classifiers. Additional information provided by these information sources has been shown to be related to the outcome of IVF treatments in the literature [14]. On the other hand, the use of Bayesian classifiers for continuous data [47] is an alternative to improve the best solutions found on the analysis provided by ROC curves.

Conflict of Interest Statement

None declared.

Acknowledgements

The authors would also like to thank the invaluable contribution of Ph.D. G. Patrizi and M.D. C. Manna to this paper, for their provision of data and support regarding their previously published paper [12].

This work has been partially supported by the following grants: Etortek, Saiotek and Research Group Funding programs 2007-2012 (IT-242-07) (Basque Government), TIN2008-06815-C02-01, TIN2008-06815-C02-02 and Consolider Ingenio 2010-CSD2007-00018 projects (Spanish Ministry of Education and Science) and COMBIOMED network in computational biomedicine (Carlos III Health Institute).

References

[1] C. Giorgetti, P. Terriou, P. Auquier, E. Hans, J.L. Spach, J. Salzmann, R. Roulier, Embryo score to predict implantation after in-vitro fertilization: based on 957 single embryo transfers, Human Reprod. 10 (1995) 2427–2431.

[2] F. Puissant, M. Van Rysselberge, P. Barlow, J. Deweze, F. Leroy, Embryo scoring as a prognostic tool in IVF treatment, Human Reprod. 2 (1987) 705–708.

[3] E. Van Royen, K. Mangelschots, D. De Neubourg, M. Valkenburg, M. Van de Meerssche, G. Ryckaert, W. Eestermans, J. Gerris, Characterization of a top quality embryo, a step towards single-embryo transfer, Human Reprod. 14 (9) (1999) 2345–2349.

[4] A. Schulman, I. Ben-Num, Y. Gethler, H. Kaneti, M. Shilon, Y. Beyth, Relationship between embryo morphology and implantation rate after in vitro fertilization treatment in conception cycles, Fertility and Sterility 60 (1993) 123–126.

[5] G.A. Hill, M. Freeman, M.C. Bastias, B.J. Rogers, C.M. Herbert III, K.G. Osteens, A.C. Wentz, The influence of oocyte maturity and embryo quality on pregnancy rate in a program for in vitro fertilization-embryo transfer, Fertility and Sterility 52 (1989) 801–806.

[6] M. Erenus, C. Zouves, P. Rajamahendran, S. Leung, M. Fluker, V. Gomel, The effect of embryo quality on subsequent pregnancy rates after in vitro fertilization, Fertility and Sterility 56 (1991) 707–710.

[7] L. Scott, Pronuclear scoring as a predictor of embryo development, Reprod. BioMed. Online 6 (2) (2003) 201–214.
[8] R. Saith, A. Srinivasan, D. Michie, I. Sargent, Relationships between the developmental potential of human in-vitro fertilization embryos and features describing the embryo, oocyte and follicle, Human Reprod. Update 4 (2) (1998) 121–134.
[9] J.R. Trimarchi, J. Goodside, L. Passmore, T. Silberstein, L. Hamel, L. Gonzalez, Comparing data mining and logistic regression for predicting IVF outcome, Fertility and Sterility 80 (3) (2003) 100.
[10] J.R. Quinlan, Induction of decision trees, Mach. Learn. 1 (1) (1986) 81–106.
[11] I. Jurisica, J. Mylopoulos, J. Glasgow, H. Shapiro, R. Casper, Case-based reasoning in IVF: prediction and knowledge mining, Artif. Intell. Med. 12 (1) (1998) 1–24.
[12] G. Patrizi, C. Manna, C. Moscatelli, L. Nieddu, Pattern recognition methods in human-assisted reproduction, Int. Trans. Oper. Res. 11 (4) (2004) 365–379.
[13] C. Manna, G. Patrizi, A. Rahman, H. Sallam, Experimental results on the recognition of embryos in human assisted reproduction, Reprod. BioMed. Online 8 (4) (2004) 460–469.
[14] D. Morales, E. Bengoetxea, P. Larrañaga, M. García, Y. Franco-Iriarte, M. Fresnada, M. Merino, Bayesian classification for the selection of in-vitro human embryos using morphological and clinical data, Comput. Meth. Programs Med. 90 (2) (2008) 104–116.
[15] R. Blanco, I. Inza, M. Merino, J. Quiroga, P. Larrañaga, Feature selection in Bayesian classifiers for the prognosis of survival of cirrhotic patients treated with TIPS, Biomed. Informatics 38 (2005) 376–388.
[16] M.K. Hu, Visual pattern recognition by moment invariants, IRE Trans. Inform. Theory 8 (1962) 179–187.
[17] F. Albregtsen, H. Schulerud, L. Yang, Texture classification of mouse liver cell nuclei using invariant moments of consistent regions, in: Computer Analysis of Images and Patterns, 2006, pp. 496–502.
[18] A. Ruggeri, S. Pajaro, Automatic recognition of cell layers in corneal confocal microscopy images, Comput. Meth. Programs Biomed. 68 (1) (2002) 25–35.
[19] J.F. Mangin, F. Poupon, E. Duchesnay, D. Rivière, A. Cachia, D.L. Collins, A.C. Evans, J. Régis, Brain morphometry using 3d moment invariants, Med. Image Anal. 8 (3) (2004) 187–196 (Medical Image Computing and Computer-Assisted Intervention—MICCAI 2003).
[20] R. Duda, P. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.
[21] M.A. Maloof, Learning when data sets are imbalanced and when costs are unequal and unknown, in: Workshop on Learning from Imbalanced Sets II, Washington, DC, 2003.
[22] C.L. Mills, Factors affecting embryological parameters and embryo selection for IVF–ET, in: P.R. Brinsden, P.A. Rainsbury (Eds.), A Textbook of In Vitro Fertilization and Assisted Reproduction, The Parthenon Group, 1992, pp. 187–204.
[23] J. Pearl, Probabilistic Reasoning in Intelligence Systems, Morgan Kaufmann, Los Altos, CA, 1988.
[24] M. Minsky, Steps toward artificial intelligence, Trans. Inst. Radio Eng. 49 (1961) 8–30.
[25] P. Langley, S. Sage, Induction of selective Bayesian classifiers, in: Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence, Seattle, WA, 1994, pp. 399–406.
[26] M. Pazzani, Searching for dependencies in Bayesian classifiers, in: D. Fisher, H.-J. Lenz (Eds.), Learning from Data: Artificial Intelligence and Statistics V, Springer-verlag, New York, NY, 1997, pp. 239–248.
[27] N. Friedman, D. Geiger, M. Goldsmidt, Bayesian network classifiers, Mach. Learn. 29 (2) (1997) 131–163.
[28] M. Sahami, Learning limited dependence Bayesian classifiers, in: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, 1996, pp. 335–338.
[29] R. Kohavi, G. John, Wrappers for feature subset selection, Artif. Intell. 97 (1–2) (1997) 273–324.
[30] M. Stone, Cross-validatory choice and assessment of statistical predictions, J. Roy. Statist. Soc. Ser. B 36 (1974) 111–147.
[31] I. Kononenko, Semi-naïve Bayesian classifiers, in: Proceedings of the 6th European Working Session on Learning, 1991, pp. 206–219.
[32] J.P. Egan, Signal Detection Theory and ROC Analysis, Academic Press, New York, 1975.
[33] T. Fawcett, ROC graphs: notes and practical considerations for researchers ⟨http://www.hpl.hp.com/personal/Tom_Fawcett/papers/ROC101.pdf⟩.
[34] T.A. Lasko, J.G. Bhagwat, K.H. Zou, L. Ohno-Machado, The use of receiver operating characteristic curves in biomedical informatics, J. Biomed. Informatics 38 (2005) 404–415.
[35] K.A. Spackman, Signal detection theory: valuable tools for evaluating inductive learning, in: Proceedings of the Sixth International Workshop on Machine Learning, Morgan Kaufmann, San Mateo, CA, 1989, pp. 160–163.
[36] F. Provost, T. Fawcett, R. Kohavi, The case against accuracy estimation for comparing induction algorithms, in: Proceedings of the Fifteenth International Conference on Machine Learning, Morgan Kaufmann, San Francisco, 1998, pp. 445–453.
[37] F. Provost, T. Fawcett, Robust classification for imprecise environments, Mach. Learn. 42 (2001) 203–231.
[38] J.A. Hanley, B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology 143 (1) (1982) 29–36.
[39] D. Bamber, The area above the ordinal dominance graph and the area below the receiver operating characteristic graph, J. Math. Psychol. 12 (1975) 387–415.
[40] K. Dougherty, R. Kohavi, M. Sahami, Supervised and unsupervised discretization of continuous features, in: Proceedings of the Twelfth International Conference on Machine Learning, 1995, pp. 194–202.
[41] Elvira Consortium, Elvira: an environment for creating and using probabilistic graphical models, in: Proceedings of the 1st European Workshop on Probabilistic Graphical Models, Cuenca, Spain, 2002, pp. 222–230.
[42] L. Breiman, J. Friedman, R. Olshen, C. Stone, Classification and Regression Trees, Chapman & Hall/CRC Press, London, Boca Raton, FL, 1984.
[43] R. Duda, P. Hart, D.G. Stork, Pattern Classification, Wiley Interscience, New York, 2001.
[44] M.F. Verberg, M.J. Eijkemans, N.S. Macklon, E.M. Heijnen, B.C. Fauser, F.J. Broekmans, Predictors of ongoing pregnancy after single-embryo transfer following mild ovarian stimulation for IVF, Fertility and Sterility 89 (5) (2007) 1159–1165.
[45] T. della Ragione, G. Verheyen, E.G. Papanikolaou, L. Van Landuyt, P. Devroey, A. Van Steirteghem, Developmental stage on day-5 and fragmentation rate on day-3 can influence the implantation potential of top-quality blastocysts in IVF cycles with single embryo transfer, Reprod. Biol. Endocrinol. 5 (2) (2007).
[46] A.M. Van Peperstraten, I.A. Kreuwel, R.P. Hermens, W.L. Nelen, P.A. Van Dop, R.P. Grol, J.A. Kremer, Determinants of the choice for single or double embryo transfer in twin prone couples, Acta Obst. Gynecol. Scandinavica 87 (2) (2008) 226–231.
[47] A. Pérez, P. Larrañaga, I. Inza, Supervised classification with conditional Gaussian networks: increasing the structure complexity from naive Bayes, Int. J. Approx. Reason. 43 (2006) 1–25.

Dinora A. Morales received her B.A. and M.Sc. degree in Computer Science from Technological Institute of Querétaro, México in 1999, and M.Sc. in Neurobiology from Institute of Neurobiology, National University of México in 2003. Since October, 2005 she is a member of Intelligent Systems Group of the University of the Basque Country where she has been working for her Ph.D. She is interested in learning from data within machine learning, artificial intelligence, and statistics. In particular, she is interested in learning probabilistic graphical models such as Bayesian networks. Actually her work concentrates in developing algorithms for supervised classification from human embryo selection in assisted reproduction techniques data.

Endika Bengoetxea received his B.Sc. in Computer Studies from the University of the Basque in 1994, his M.Sc. in Medical Imaging (Medical Physics) from the University of Aberdeen in 1999, and his Ph.D. in Signal and Image

Treatment from the French Ecole Nationale Supérieure des Télécommunications in 2002. He joined the University of the Basque Country in 1996 where he is currently professor of the Computer Engineering Faculty, and member of the research group Intelligent Systems Group. His research interests are on the application of Bayesian networks for optimisation or classification problems applied to biomedical problems.

Pedro Larrañaga received his M.Sc. degree in mathematics from the University of Valladolid, in Spain, in 1981, and his Ph.D. in computer science from the University of the Basque Country, in Spain, in 1995. He is professor at the Department of Artificial Intelligence of the Technical University of Madrid. His research topics include Bayesian networks and evolutionary computation with applications in medicine and bioinformatics.