ROC Based Evaluation and Comparison of Classifiers for IVF Implantation Prediction

Asli Uyar[1,⋆], Ayse Bener[1], H. Nadir Ciray[2], and Mustafa Bahceci[2]

[1] Bogazici University, Department of Computer Engineering, Istanbul, Turkey
{asli.uyar,bener}@boun.edu.tr
[2] German Hospital and Bahceci IVF Center, Istanbul, Turkey
{nadirc,mbahceci}@superonline.com

Abstract. Determination of the best performing classification method for a specific application domain is important for the applicability of machine learning systems. We have compared six classifiers for predicting implantation potentials of IVF embryos. We have constructed an embryo based dataset which represents an imbalanced distribution of positive and negative samples as in most of the medical datasets. Since it is shown that accuracy is not an appropriate measure for imbalanced class distributions, ROC analysis have been used for performance evaluation. Our experimental results reveal that Naive Bayes and Radial Basis Function methods produced significantly better performance with (0.739 $\pm$ 0.036) and (0.712 $\pm$ 0.036) area under the curve measures respectively.

1 Introduction

In-vitro fertilization (IVF) [1] is a common infertility treatment method during which female germ cells (oocytes) are inseminated by sperm under laboratory conditions. Fertilized oocytes are cultured between 2-6 days in special medical equipments and embryonic growth is observed and recorded by embryologists. Finally, selected embryo(s) are transferred into the woman's womb. Predicting implantation (i.e. attachment of the embryo to the inner layer of the womb) potentials of individual embryos may expedite and enhance expert judgement for two critical issues: 1) the decisions of number of embryos to be transferred and 2) selection of embryos with highest reproductive viabilities. Increasing the number of transference embryos may increase the pregnancy probability but also increase possible complications arising from multiple pregnancies. Therefore, in clinical practice, it is desired to transfer minimum number of embryos with the highest implantation potentials. In this study, implantation prediction is considered as a binary classification problem such that the embryos have been classified into two categories as implants and no-implants.

The recent literature presents applications of machine learning methods in IVF domain. Case-based reasoning system [2], neural networks [3], decision tree models [4] [5] [6] and Bayesian classification system [7] were utilized for prediction

⋆ Corresponding author.

P. Kostkova (Ed.): eHealth 2009, LNICST 27, pp. 108–111, 2010.

of IVF outcome. However, direct comparison of these studies is not possible due to variety of input features, training and testing strategies and performance measures. The specific aim of this paper is to determine the best classification algorithm for implantation prediction of IVF embryos.

2 IVF Dataset

We have constructed a dataset from a database including information on cycles performed at the German Hospital in Istanbul from January 2007 through August 2008. Dataset included 2453 records with 89% positive and 11% negative implantation outcomes. Each embryo was represented as a row vector including 18 features related to patient and embryo characteristics (Table 1).

Table 1. Selected dataset features for each embryo feature vector

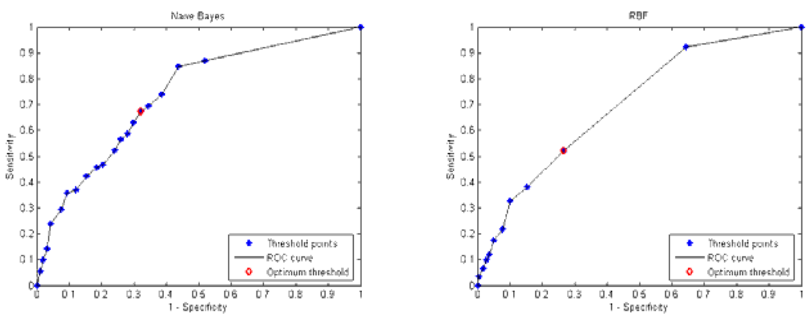
Dataset Features	**Data Type**
Patient Characteristics	
Woman age	Numerical
Primary or secondary infertility	Categorical
Infertility factor	Categorical
Treatment protocol	Categorical
Duration of stimulation	Numerical
Follicular stimulating hormone dosage	Numerical
Peak Estradiol level	Numerical
Endometrium thickness	Numerical
Sperm quality	Categorical
Embryo Related Data	
Early cleavage morphology	Categorical
Early cleavage time	Numerical
Transfer day	Categorical
Number of cells	Numerical
Nucleus characteristics	Categorical
Fragmentation	Categorical
Blastomeres	Categorical
Cytoplasm	Categorical
Thickness zona pellucida	Categorical

3 Experiments and Results

Six classification algorithms were applied: the Naive Bayes, k-Nearest Neighbor (kNN), Decision Tree (DT), Support Vector Machines (SVM), Multi Layer Perceptron (MLP) and Radial Basis Function Network (RBF) [8]. Two-thirds of the dataset was randomly selected for training a predictor model and the remaining one-third was utilized for testing. The random train set and test set partitioning was repeated 10 times avoiding sampling bias. The reported results were the mean values of these 10 experiments. A t-test with $P = 0.05$ was applied in order to determine the statistical significance of results.

Various sampling strategies have been applied in order to overcome the imbalanced data problem such as over-sampling and under-sampling. In a recent study, Maloof showed that Receiver Operating Characteristics (ROC) curves produced by sampling strategies are similar to those produced by varying the decision threshold [9]. ROC analysis provides reliable evaluation of classifier performance considering True Positive Rates (TPR) and False Positive Rates (FPR). All the values of TPR and FPR have been calculated by varying the decision thresholds in the range of [0:0.05:1]. The resulting set of (TPR (sensitivity), FPR (1-specificity)) pairs are plotted as a 2D ROC curve. The upper left point (0,1) on the ROC curve represents perfect classification. Therefore, the threshold value that gives the nearest point to (0,1) is accepted as the optimum decision threshold (t_{opt}).

We have used Weka machine learning tool [10] to perform classification. Among six methods, Naive Bayes and RBF were significantly better. The results were plotted as ROC curves, which appear in Figure 1 demonstrating the effect of threshold optimization. Naive Bayes classification with optimized threshold results in 67% sensitivity and 30% false alarm rate.



(a) ROC curve for Naive Bayes classification. AUC is 0.739 ± 0.036

(b) ROC curve for RBF classification. AUC is 0.712 ± 0.036

Fig. 1. ROC analysis representation of Naive Bayes and RBF classifiers for IVF dataset

4 Conclusions and Future Work

In this paper, we have presented a novel study predicting implantation potentials of individual IVF embryos. We have constructed an original dataset and performed predictions using six classification methods. Experimental results show that Naive Bayes and RBF produce significantly better performance for the implantation prediction problem. Model optimization and classifier comparison have been performed using ROC analysis.

To be applicable in clinical practice, the machine learning algorithms have been investigated in a comprehensive and comparative manner allowing reliable selection of best fitting algorithm. In machine learning applications, it is crucial to deal with biases arising from training-testing strategies. We have applied stratified cross validation in order to overcome the sampling bias. Statistical validity is investigated by conducting t-tests. Six models from important representatives of diverse algorithms (statistical classifiers, decision tree approaches, neural networks, support vector machines and nearest neighbor methods) have been applied. Since the data used in this study comes from a single source, it is crucial to consider the external validity of the presented results. The experiments need to be replicated on different IVF datasets, however, public dataset construction and data sharing has been a major research challenge in this domain.

Future work includes improving the information content of dataset by collecting new input features. Combination of classifiers should also be investigated to construct a more powerful decision support system for IVF domain.

References

1. Steptoe, P.C., Edwards, R.G.: Birth after re-implantation of a human embryo. Lancet 2, 366 (1978)
2. Jurisica, I., Mylopoulos, J., Glasgow, J., Shapiro, H., Casper, R.F.: Case-based reasoning in ivf: Prediction and knowledge mining. Artificial Intelligence in Medicine 12, 1–24 (1998)
3. Kaufmann, S.J., Eastauh, J.L., Snowden, S., Smye, S.W., Sharma, V.: The application of neural networks in predictingthe outcome of in-vitro fertilization. Human Reproduction 12, 1454–1457 (1997)
4. Saith, R., Srinivasan, A., Michie, D., Sargent, I.: Relationships between the developmental potential of human in-vitro fertilization embryos and features describing the embryo, oocyte and follicle. Human Reproduction Update 4(2), 121–134 (1998)
5. Passmore, L., Goodside, J., Hamel, L., Gonzalez, L., Silberstein, T., Trimarchi, J.: Assessing decision tree models for clinical in-vitro fertilization data. Technical report, Dept. of Computer Science and Statistics University of Rhode Island (2003)
6. Trimarchi, J.R., Goodside, J., Passmore, L., Silberstein, T., Hamel, L., Gonzalez, L.: Comparing data mining and logistic regression for predicting ivf outcome. Fertil. Steril (2003)
7. Morales, D.A., Bengoetxea, E., Larranaga, B., Garcia, M., Franco, Y., Fresnada, M., Merino, M.: Bayesian classification for the selection of in vitro human embryos using morphological and clinical data. Computer Methods and Programs in Biomedicine 90, 104–116 (2008)
8. Alpaydin, E.: Introduction to Machine Learning. MIT Press, Cambridge (2004)
9. Maloof, A.M.: Learning when data sets are imbalanced and when costs are unequal and unknown. In: Workshop on Learning from Imbalanced Data Sets (2003)
10. Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann, San Francisco (2005)