Prediction of Infertility Treatment Outcomes Using Classification Trees

Anna Justyna Milewska[1], Dorota Jankowska[1], Urszula Cwalina[1], Dorota Citko[1], Teresa Więsak[2], Brian Acacio[3], Robert Milewski[1]

[1] Department of Statistics and Medical Informatics, Medical University of Bialystok, Poland
[2] Department of Gamete and Embryo Biology, Institute of Animal Reproduction and Food Research of Polish Academy of Sciences, Olsztyn, Poland
[3] Acacio Fertility Center, Laguna Niguel, California, USA

Abstract. Infertility is currently a common problem with causes that are often unexplained, which complicates treatment. In many cases, the use of ART methods provides the only possibility of getting pregnant. Analysis of this type of data is very complex. More and more often, data mining methods or artificial intelligence techniques are appropriate for solving such problems. In this study, classification trees were used for analysis. This resulted in obtaining a group of patients characterized most likely to get pregnant while using in vitro fertilization.

Introduction

Infertility is currently a common problem. It is present in 18–20% of couples in Poland. It is defined as the inability to become pregnant despite a year of sexual activity with an intercourse frequency of about 2–3 times per week, without using any contraceptive measures (Radwan et al., 2011). Often, the causes of infertility are unexplained, which complicates treatment. In many situations, the only way to get pregnant is to use methods of Assisted Reproductive Technologies (ART) or IVF. Unfortunately, these methods still do not have a sufficiently high efficacy. Analysis of data collected during the procedures of assisted reproduction allows one to search for the factors that have an impact on improving the effectiveness of treatment. Adequate interpretation of the results of the analysis of clinical, embryological, and other types of data may result in increased effectiveness of the above mentioned procedures. However, as shown in the literature, the use of conventional statistical methods does not yield satisfactory results in predicting

the treatment outcome. The use of data mining tools or artificial intelligence techniques may provide hoped-for results. These methods allow for the detection of relationships which, using basic analyses, remain invisible. They allow one to search huge databases to detect significant regularities. Principal Component Analysis helps to reduce the dimensionality of the database, so that it becomes easier to obtain knowledge from it (Milewska et al., 2014). Correspondence analysis and cluster analysis use solutions based on grouping cases (Milewska et al., 2012, 2013). Market basket analysis is based on the idea of discovering association rules found in the analyzed dataset (Milewska et al., 2011). Very promising results have also been achieved by the use of neural networks (Milewski et al., 2009, 2013).

There is a suspicion that the use of analyses based on decision trees may be an interesting approach to the problem of searching for factors that may improve the effectiveness of infertility treatment. These algorithms belong to the group of classification methods and, in this study, the purpose of their application was to select patients characterized most likely to get pregnant while using in vitro fertilization.

Decision Trees – Description of the Method

The method of data analysis which uses decision trees is based on recursive division of examined cases into separable subsets of similar properties. Its purpose can be both to predict and to explain the relationship between a dependent variable and each type of explanatory variable (Wieczorek et al., 2008). The main result of the decision tree is to obtain a tool for classifying cases which are not known to belong to considered or emerged groups. A model created using this method is also descriptive (Owczarek, 2014). It allows you to discover and describe the rules that guide the analyzed phenomenon. Decision trees can be used for both classification (when the examined variable is categorical) and regression tasks (when the analyzed variable is numerical). The subject of the research in this paper is to look for features that affect pregnancy or lack thereof. Therefore, only classification trees will be described in a more detailed way.

The idea of data analysis based on the use of decision trees has existed since the 1980s, when it appeared in a Breiman et al. publication (1984). Currently, this method continues to undergo rapid development. Therefore, a number of algorithms for the construction of trees have been prepared so far. The most popular of them are: CART (Classification and Regression Trees), CHAID (Chi-Square Automatic Interaction Detection), ID3 (Itera-

tive Dichotomiser 3) and C4.5 (successor of ID3), which belong to the CLS family (Concept Learning System) or QUEST (Quick, Unbiased, Efficient Statistical Trees) (Maimon et al., 2005; Rokach et al., 2015).

The discussed method of statistical data analysis allows for an interesting and easy-to-interpret visualization of results in the form of a flowchart. The tree is a structure whose creation is started from the root node. The root poses at the top and it includes all cases from the training dataset. These items are distributed along the edges into subsamples – so-called child nodes. Then, these child nodes can be the parent nodes to the next divisions. A node without children is called a leaf. Each leaf receives the name of the class which appears most frequently in that particular leaf. In each internal node all predictors and their possible divisions are looked through. Then, the feature which will ensure a division of the analyzed data set into subsets that will be the most homogeneous in terms of the dependent variable is selected. Reasonable distribution can be ensured by choosing the measurement of degree of inhomogeneity or impurity in the node. The most common metrics are (Breiman et al., 1984; Kingsford et al., 2008; Łapczyński, 2009):

– Gini Index (applied in CART)

$$G = 1 - \sum_j p^2(j|t)$$

  t – number of observations in the node
  j – number of classes (number of variants of dependent variable)
  $p(j|t)$ – fraction of observations from the node that belong to j-th class

– Entropy (used by ID3, C4.5)

$$E = - \sum_j p(j|t) \log_2 p(j|t)$$

– Twoing criterion (it is used in the binary trees that allow you to create a model for dichotomous variables)

$$TC = \frac{(p_l p_r)}{4} \left(\sum_j |p(j|t_l) - p(j|t_r)| \right)^2$$

  t_l, t_r – number of observations in the left and right nodes, respectively
  p_l, p_r – probability that the observation is assigned to the left or right node, respectively

– The proportion of misclassifications

$$P = 1 - \max_j p(j|t_r)$$

Using one of the previously mentioned measures, the difference between impurity of the node and impurity of child nodes (which is computed as a weighted average of the levels of impurity in prepared nodes) should be determined. The obtained difference is called the improvement indicator or information gain. The variable for which the difference is maximal should be used to split the node.

Hierarchical divisions of the training dataset can be performed until leaves are obtained in which all cases belong to one class. That is the most natural stop criterion for development of a classification tree. However, this can lead to the preparation of very deep trees, and thus can be too complex. This can also cause overfitting of training examples. Some algorithms of building trees use a certain stop criteria. They may be associated with a predetermined minimum number of elements in a terminal node, the minimum frequency of classes in a leaf or pre-established depth of the tree. Construction of a tree can also be stopped when none of the features yield an improvement indicator or information gain higher than a predetermined level (Maimon et al., 2005; Rokach et al., 2015). Another way to reduce the size of the classification tree, which is permitted by some algorithms, is pruning the tree by deleting nodes. It involves a removal of the fragments of the tree that have minor importance for the quality of classification and do not provide additional information, and then collapsing internal nodes into leaves. It leads to reduction of predictive accuracy in a training dataset, but it can improve the quality of a classifier for new observations. The final effect of classification tree growth may be obtained by taking into account unequal costs of incorrect classification of some cases into the considered groups, which is acceptable in many algorithms (Owczarek, 2014).

Some of the tree construction algorithms provide the possibility of preparing a ranking of variables in each internal node (Łapczyński, 2009). For this purpose, predictors are ordered in accordance with the value of the improvement indicator and they are ranked on a scale of 0 to 1. The sum of all improvement indicators from all nodes for each of the independent variables is the basis for formation of a global ranking of predictors in a prepared tree.

Data analysis using decision trees allows for description of the relationships and rules by which the studied phenomenon is controlled. It gives rules which are clear, intuitive and relatively easy to interpret. Most algorithms created thus far do not make any demands on data. Variables measured at different scales can be taken into account. It is possible to analyze observations with missing values of some variables. This method also shows resistance to outliers.

Application of Decision Trees

Decision trees are a method used in many branches of science. They support the decision-making processes in various areas. This method is widely used, especially in image and video analysis. Decision trees are widely used in image detection in a variety areas. They can be used for recognizing three dimensional objects (Bullock et al., 1994; Spirkovska, 1993) and for high level vision (Kodratoff et al., 1994). Sometimes decision trees are the only method of analysis, but often they are used in conjunction with other statistical tools. In this way, the versatile advantages of this method are used more fully.

There are many examples of using decision trees in astronomy. For instance, identification of cosmic radiation in images taken by the Hubble Space Telescope is shown in Salzberg et al. (1995). Decision trees are also used to classify objects such as galaxies (Weir et al., 1995).

Another area in which the use of decision trees is well–known is molecular biology. Some works show how to use this tool in the analysis of human DNA. They are an effective method for identifying protein coding regions (Salzberg, 1995). In a publication by Shimozono et al. (1994), the idea for acquisition of knowledge from amino acid sequences was shown.

Many other examples of the use of decision trees can be seen in various medical disciplines. They can support diagnostic processes in various ways, including through image recognition. There are several papers showing the use of classification trees in support of decision-making processes in cardiology (Falconer et al., 1994; Kokol et al., 1994; Long et al., 1993; Negassa et al., 2007). Some studies show the comparison of logistic regression with classification trees. One such study is *"Comparison of logistic regression model and classification tree: An application to postpartum depression data"* (Camdeviren et al., 2007). This paper demonstrates the use of both of these methods in the assessment of the risk of postpartum depression, depending on social and demographic factors. An example of the use of classification trees in gynecology can be found in a paper by Pouliakis et al. (2015). This tool is used to develop a model for triage and assessment of the risk factors for cervical intraepithelial neoplasia. Other areas of medicine in which decision trees can be useful are: gastroenterology (Judmaier et al., 1993), neurology (Takahashi et al., 2006), psychiatry (Mckenzie et al., 1993) and general medicine (Mello et al., 2006).

Application of Decision Trees to Select Patients Who Are Most Likely to Get Pregnant

The aim of the study described herein was to select patients who are most likely to get pregnant during in vitro fertilization. Classification trees that indicated variables with the best predictive properties and performed the classification of cases were applied. The Statistica Data Miner + QC 12.5 (StatSoft) program was used for the statistical analysis. To estimate differences between groups, a U Mann-Whitney test was used, then classification trees were applied. Next, the ROC curve was used to evaluate the predictive power of a given tree. The results were considered statistically significant at the level of $p < 0.05$.

The data used in the analysis included 610 cycles of the in vitro fertilization performed at a clinic in the United States. The variables were: the patient's age, the number and quality of cells obtained during stimulation (GV, M1, M2, M2* – stages of maturity – shown as a percentage), the number and quality of the resulting embryos (2PN, NEF 2PB, non2PN, clvd on day 3, ≥ 7 cell on day 3, blasts on day 5, blasts on day 6 – shown as a percentage), the number of transferred embryos (number ET), the day of transfer and information about the pregnancy.

In the analyzed group, biochemical pregnancy (confirmed with a pregnancy test) was obtained in 44.8% of cases (n = 273). Differences between the group of patients with achieved pregnancy and the group without pregnancy were statistically significant for the following variables: *female age (years), follicles (n), retrieved oocytes (n), ≥ 7 cell on day 3 (%), blasts on day 5 (%), blasts on day 6 (%), ET (day)* – all at p < 0,001.

Before the classification trees were applied, the data were randomly assigned to two groups, which did not differ significantly: "training" (n = 438) and "validation" (n = 172). The trees were built on the training group. Three attempts were made to define patients with the best parameters for determining pregnancy. In the first approach, a tree based on all the variables – that is, the information about the women's ages, oocytes, embryos and transfer – was constructed (Tree1 – Figure 1). In the second approach, only the data for the women's ages and the number and quality of oocytes obtained by the stimulation were used (Tree2 – Figure 2). And the third approach took into account the women's ages, number and quality of embryos, and information about the transfer (Tree3 – Figure 3).

Tree 1 (Figure 1) included the following variables: *blasts on day 5 (%)* (importance = 0.49), *female age* (importance = 0.95), ≥ 7 *cell on day 3 (%)* (importance = 1.00), *retrieved oocytes* (n) (importance = 0.90), *M2 af-*
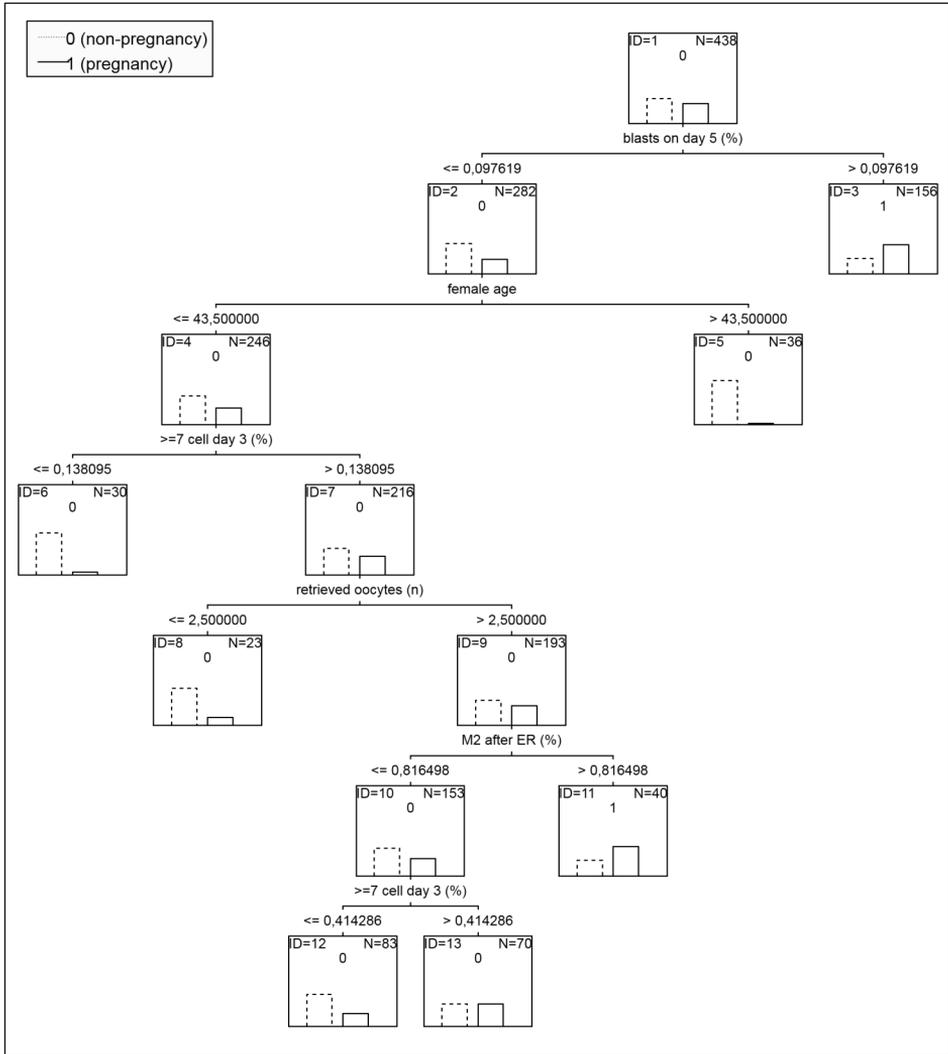
Figure 1. Classification tree constructed for all variables (the information about women's age, oocytes, embryos and transfer)

ter ER (%) (importance = 0.52). For tree 1, only two leaf nodes were characterized by a predominance of pregnancies. The first leaf node (ID = 3) concerned a group of 158 patients, of which the pregnancy percentage was 65.4%. This leaf was defined by the condition: variable *blasts on 5 day* above 9.8%. The second leaf node (ID = 11) included 40 patients, with 65.0% of cases resulting in pregnancy. These were the patients who met the following conditions: *blasts on day 5* were maximally 9.8%, *female age* maximally
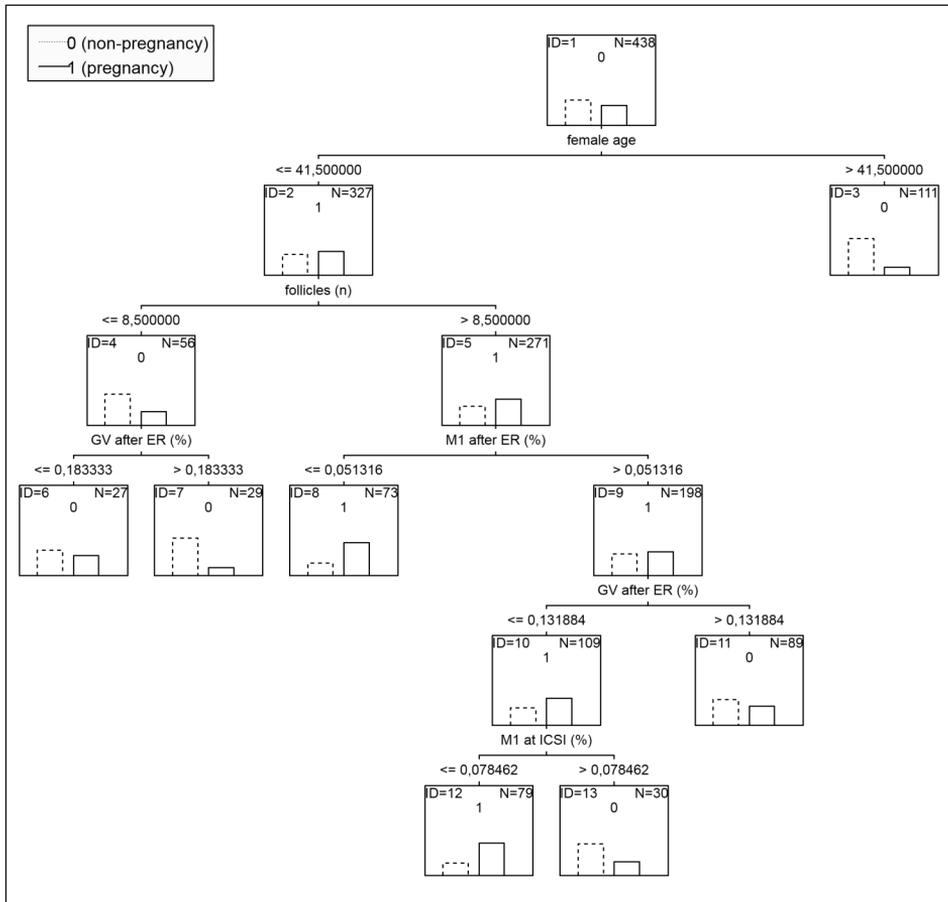
Figure 2. Classification tree constructed with variables relating to the women's ages and number and quality of oocytes obtained by the stimulation

43.5 years, ≥ 7 *cell on day 3* above 13.8%, *retrieved oocytes* at least 3 and *M2 after ER* over 81.6%.

Tree 2 (Figure 2) included the following variables: *female age* (importance = 0.91), *follicles* (n) (importance = 0.71), *GV after ER* (%) (importance = 0.86), *M1 after ER* (%) (importance = 0.61), *M1 at ICSI* (%) (importance = 1.00). Pregnancy dominated in 6 cases out of 13 leaves. The percentages of pregnancies were, respectively: leaf ID = 2 – 53.2%, leaf ID = 5 – 57.9%, leaf ID = 8 – 72.6%, leaf ID = 9 – 52.5%, leaf ID = 10 – 60.3%, leaf ID = 12 – 72.2%. The leaf of ID = 8 was characterized by the highest percentage of pregnancies. In this group of 73 women, pregnancy occurred in 72.6% of cases. These were patients who met the fol-
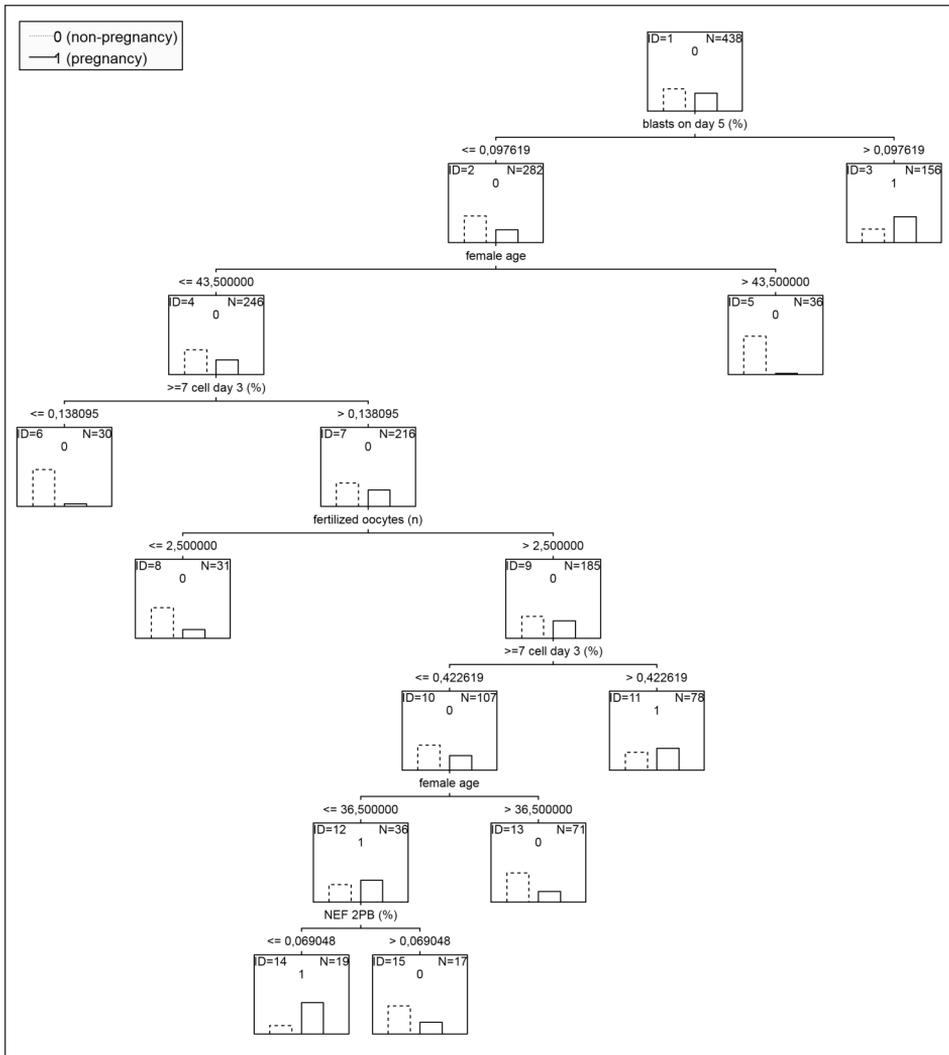
Figure 3. Classification tree constructed with variables relating to the women's ages, number and quality of embryos and information about the transfer

lowing conditions: *female age* maximally 41.5 years, *follicles* at least 9 and *M1 after ER* maximally 5%. A slightly lower pregnancy percentage (72.2%) occurred in 79 women defined, in the leaf of ID = 12, by the conditions: *female age* maximally 41.5 years, *follicles* at least 9, *M1 after ER* at least 5%, *GV after ER* maximally 13% and *M1 at ICSI* maximally 8%.

Tree 3 (Figure 3) included the variables: *blasts on day 5 (%)* (importance = 0.35), *female age* (importance = 1.00), *≥7 cell on day 3 (%)* (im-

portance = 0.61), *fertilized oocytes* (n) (importance = 0.71), *NEF 2PB (%)* (importance = 0.43). Pregnancy dominated in 4 cases out of 15 leaves: leaf ID = 3 – 65.4%, leaf ID = 11 – 55.1%, leaf ID = 12 – 55.6%, leaf ID = 14 – 78.9%. The highest percentage of pregnancies was obtained in the leaf of ID = 14, which included only 19 cases. In this group, patients were characterized by: *blasts on day 5* maximally 10%, *female age* below 44 years, *≥7 cell on day* maximally 14%, *fertilized oocytes* at least 3, then *≥7 cell on day* at most 42%, *female age* below 37 years and *NEF 2PB* not exceeding 7%.

The predictive properties of the trees have been verified using ROC curves and the calculated area under the curve. In all cases, the predictive power of the training group was higher than the predictive power of the validation group. For the training group, the predictive power was 0.75–0.76, while for the validation group, it was from 0.66 to 0.68 (Figure 4).
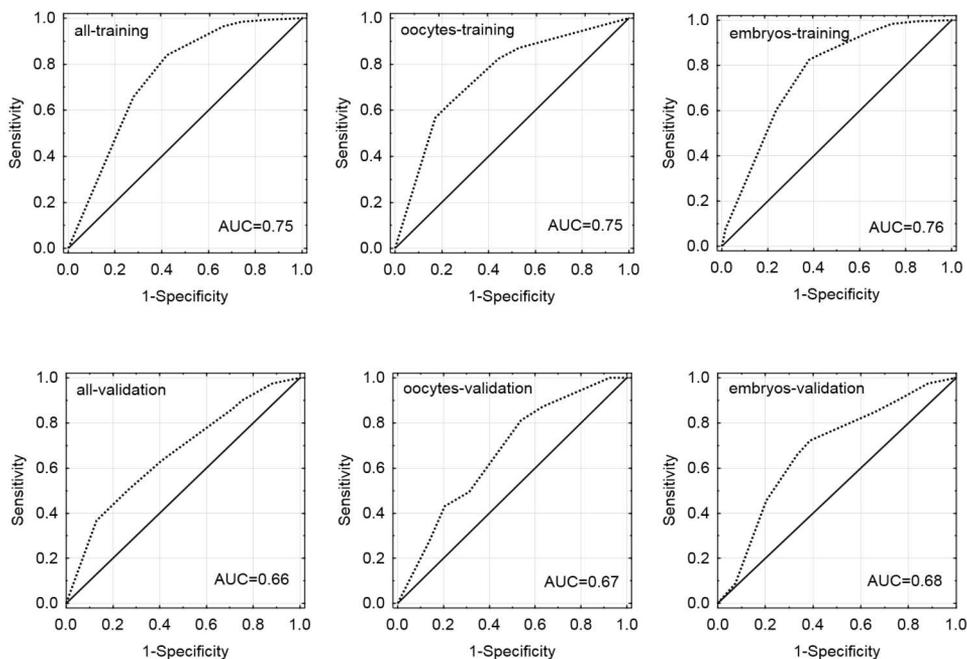
Figure 4. ROC curves of trees 1–3 for training and validation groups

Conclusions

Decision trees applied for the purpose of this study led to the indication of more pregnancy predictors than can be obtained using tradi-

tional tests of significance. These indicators are: *female age, follicles* (n), *retrieved oocytes* (n), *M1 after ER* (%), *M2 after ER* (%), *GV after ER* (%), *M1 at ICSI* (%), *fertilized oocytes* (n), *≥7 cell on day 3* (%), *NEF 2PB* (%), *blasts on day 5* (%). Due to the use of decision trees, the characteristics of patients for whom the chance of pregnancy using the ICSI method is more than 72% were defined. Decision tree use can be very useful in predicting the results of infertility treatment. Although not the main subject of this analysis, created trees also provide the possibility of defining groups of women for whom the probability of pregnancy is very small.

R E F E R E N C E S

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees*. Wadsworth Publishing, CA.

Bullock, M. E., Wang, D. L., Fairchild, S. R., & Patterson, T. J. (1994, July). Automated training of 3–D morphology algorithm for object recognition. *Proceedings of SPIE – The International Society for Optical Engineering, 2234, Automatic Object Recognition IV*, 238.

Camdeviren, H. A., Yazici, A. C., Akkus, Z., Bugdayci, R., & Sungur, M. A. (2007). Comparison of logistic regression model and classification tree: An application to postpartum depression data. *Expert Systems with Applications*, 32(4), 987–994.

Falconer, J. A., Naughton, B. J., Dunlop, D. D., Roth, E. J., Strasser, D. C., & Sinacore, J. M. (1994). Predicting stroke inpatient rehabilitation outcome using a classification tree approach. *Archives of Physical Medicine and Rehabilitation*, 75(6), 619–625.

Judmaier, J., Meyersbach, P., Weiss, G., Wachter, H., & Reibnegger, G. (1993). The role of Neopterin in assessing disease activity in Crohn's disease: Classification and regression trees. *The American Journal of Gastroenterology*, 88(5), 706–711.

Kingsford, C. & Salzberg, S. L. (2008). What are decision trees? *Nature Biotechnology*, 26(9), 1011–1013.

Kodratoff, Y., & Moscatelli, S. (1994). Machine learning for object recognition and scene analysis. *International Journal of Pattern Recognition and Artificial Intelligence*, 8(1), 259–304.

Kokol, P., Mernik, M., Zavrsnik, J., & Kancler, K. (1994). Decision trees based on automatic learning and their use in cardiology. *Journal of Medical Systems*, 18(4), 201–206.

Łapczyński, M. (2009). Zmienne zastępcze i konkurencyjne w interpretacji drzew klasyfikacyjnych CART. *Zeszyty Naukowe Uniwersytet Ekonomiczny w Krakowie*, 800, 99–114.

Long, W. J., Griffith, J. L., Selker, H. P., & D'Agostino, R. B. (1993). A comparison of logistic regression to decision-tree induction in a medical domain. *Computers and Biomedical Research*, 26(1), 74–97.

Maimon, O., & Rokach, L. (2005). *Data Mining and Knowledge Discovery Handbook*. Springer.

Mckenzie, P. D., Mcgorry, D. P., Wallace, C. S., Low, L. H., Copolov, D. L., & Singh, B. S. (1993). Constructing a minimal diagnostic decision tree. *Methods of Information in Medicine*, 32(2), 161–166.

Mello, F. C. d. Q., Bastos, L. G. d. V., Soares, S. L. M., Rezende, V. M. C., Conde, M. B., Chaisson, R. E., Kritski, A. L., et al. (2006). Predicting smear negative pulmonary tuberculosis with classification trees and logistic regression: a cross-sectional study. *BMC Public Health*, 6:43. DOI:10.1186/1471-2458-6-43

Milewska, A. J., Górska, U., Jankowska, D., Milewski, R., & Wołczyński, S. (2011). The use of the basket analysis in a research of the process of hospitalization in the gynecological ward. *Studies in Logic, Grammar and Rhetoric. Logical, Statistical and Computer Methods in Medicine*, 25(38), 83–98.

Milewska, A. J., Jankowska, D., Citko, D., Więsak, T., Acacio, B., & Milewski, R. (2014). The use of principal component analysis and logistic regression in prediction of infertility treatment outcome. *Studies in Logic, Grammar and Rhetoric. Logical, Statistical and Computer Methods in Medicine*, 39(52), 7–23.

Milewska, A. J., Jankowska, D., Cwalina, U., Więsak, T., Morgan, A., & Milewski, R. (2013). Analyzing outcome of intrauterine insemination treatment by application of Cluster Analysis or Kohonen Neural Networks. *Studies in Logic, Grammar and Rhetoric. Logical, Statistical and Computer Methods in Medicine*, 35(48), 7–25.

Milewska, A. J., Jankowska, D., Górska, U., Milewski, R., & Wołczyński, S. (2012). Graphical representation of the relationships between qualitative variables concerning the process of hospitalization in the gynecological ward using correspondence analysis. *Studies in Logic, Grammar and Rhetoric. Logical, Statistical and Computer Methods in Medicine*, 29(42), 7–25.

Milewski, R., Jamiołkowski, J., Milewska, A. J., Domitrz, J., Szamatowicz, J., & Wołczyński, S. (2009). Prognosis of the IVF ICSI/ET procedure efficiency with the use of artificial neural networks among patients of the Department of Reproduction and Gynecological Endocrinology. *Ginekologia Polska*, 80(12), 900–906.

Milewski, R., Milewska, A. J., Więsak, T., & Morgan, A. (2013). Comparison of artificial neural networks and logistic regression analysis in pregnancy prediction using in the in vitro fertilization treatment Networks. *Studies in Logic, Grammar and Rhetoric. Logical, Statistical and Computer Methods in Medicine*, 35(48), 39–48.

Negassa, A., Monrad, E. S., Bang, J. Y., & Srinivas, V. S. (2007). Tree-structured Risk Stratification of In-hospital Mortality Following Percutaneous Coronary Intervention for Acute Myocardial Infarction: A Report from the New York State Percutaneous Coronary Intervention Database. *American Heart Journal*, 154(2), 322–329.

Owczarek, A. J. (2014). Classification trees in medicine. *Annales Academiae Medicae Silesiensis*, 68(6), 449–456.

Pouliakis, A., Karakitsou, E., Chrelias, C., Pappas, A., Panayiotides, I., Valasoulis, G., Kyrgiou, M., et al. (2015). The Application of Classification and Regression Trees for the Triage of Women for Referral to Colposcopy and the Estimation of Risk for Cervical Intraepithelial Neoplasia: A Study Based on 1625 Cases with Incomplete Data from Molecular Tests. *BioMed Research International*, 2015, 914740. DOI: 10.1155/2015/914740

Radwan, J., & Wołczyński, S. (Eds.). (2011). *Niepłodność i rozród wspomagany*. Poznań, Polska: Termedia Wydawnictwa Medyczne.

Rokach, L., & Maimon, O. (2015). *Data Mining with Decision Trees. Theory and Applications*. 2nd Edition. World Scientific. Singapore.

Salzberg, S. (1995). Locating protein coding regions in human DNA using a decision tree algorithm. *Journal of Computational Biology*, 2(3), 451–458.

Salzberg, S., Chandar, R., Ford, H., Murthy, S., & White, R. (1995). Decision Trees for Automated Identification of Cosmic-Ray Hits in Hubble Space Telescope Images. *Publications of the Astronomical Society of the Pacific*, 107(709), 279–288.

Shimozono, S., Shinohara, A., Shinohara, T., Miyano, S., Kuhara, S., & Arikawa, S. (1994). Knowledge acquisition from amino acid sequences by machine learning system BONSAI. *Transactions of the Information Processing Society of Japan*, 35(10), 2009–2018.

Spirkovska, L. (1993). Three dimensional object recognition using similar triangles and decision trees. *Pattern Recognition*, 26(5), 727–732.

Takahashi, O., Cook, E. F., Nakamura, T., Saito, J., Ikawa, F., & Fukui, T. (2006). Risk stratification for in-hospital mortality in spontaneous intracerebral haemorrhage: A Classification and Regression Tree Analysis. *QJM: An International Journal of Medicine*, 99(11), 743–750.

Weir, N., Fayyad, U. M., & Djorgovski, S. (1995). Automated Star/Galaxy Classification for Digitized Poss-II. *Astronomical Journal*, 109, 2401.

Wieczorek, T., & Świtała, P. (2008). Wykorzystanie algorytmu CART do automatycznego tworzenia bazy wiedzy systemu ekspertowego. In S. Kozielski, B. Małysiak, P. Kasprowski, & D. Mrozek (Eds.), *Bazy Danych: Rozwój metod i technologii*. WKŁ.