STUDIES IN LOGIC, GRAMMAR
AND RHETORIC 35 (48) 2013
DOI: 10.2478/slgr-2013-0033

Comparison of Artificial Neural Networks and Logistic Regression Analysis in Pregnancy Prediction Using the In Vitro Fertilization Treatment

Robert Milewski[1], Anna Justyna Milewska[1], Teresa Więsak[2], Allen Morgan[3]

[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] Shore Institute for Reproductive Medicine, Lakewood, NJ, USA

Abstract. Infertility is recognized as a major problem of modern society. Assisted Reproductive Technology (ART) is the one of many available treatment options to cure infertility. However, the efficiency of the ART treatment is still inadequate. Therefore, the procedure's quality is constantly improving and there is a need to determine statistical predictors as well as contributing factors to the successful treatment. There is a concern over the application of adequate statistical analysis to clinical data: should classic statistical methods be used or would it be more appropriate to apply advanced data mining technologies? By comparing two statistical models, Multivariable Logistic Regression analysis and Artificial Neural Network it has been demonstrated that Multivariable Logistic Regression analysis is more suitable for theoretical interest but the Artificial Neural Network method is more useful in clinical prediction.

Introduction

Based on available statistics, approximately 15% of Polish couples suffer from infertility. Some authors suggest the percentage is even 18–20% (Radwan, 2011). This value generally is in the range of 10–20% and reported differences depend on the data collection methods in different countries: in Denmark it is 11%, in France 16.4% and in the UK 17% (Radwan, 2011). In the United States, using the current duration approach, infertility among women 15–44 years old is 15.5% (Thoma et al., 2013).

Female factors such as: endometriosis, PCO or other ovulatory, uterine, or fallopian tube irregularities (Milewski et al., 2013) and male factors such as: oligoasthenospermia, asthenospermia, teratospermia, azoospermia, rare

ISBN 978–83–7431–392–6 ISSN 0860-150X

oligospermia and immunological factors (Radwan, 2011) contribute to infertility. Idiopathic infertility is a situation in which the clinical evaluation and laboratory tests are normal (within the range) but the couple is not able to conceive naturally.

The infertility treatment depends on the type of diagnosis, ART is one of those options. Many factors influence the efficacy of ART but the age of the female is the most important one because it affects the quality of oocytes. The efficiency of infertility treatment administered to women over 40 years old is only 10% (Milewski et al., 2008). It has been observed recently, that environmental factors contribute to both male and female infertility more and more. Other factors also influence the outcome of the treatment such as the uterine contraction level in women undergoing in vitro fertilization (Milewski, Pierzynski et al., 2012).

In spite of the constant improvement in techniques that enhance the efficacy of ART treatment, the pregnancy rate is still low and remains in the range of 40% (Milewski et al., 2013). To increase the likelihood of achieving pregnancy success, more than one embryo is transferred, subsequently causing multiple pregnancies, which is a major problem of infertility treatment. Therefore, a lot of studies across the world are focused on improving the single embryo transfer method without sacrificing the success of the ART treatment. Such a treatment also reduces the occurrence of multiple pregnancies. The key element of such an approach is to establish prognostic values that would allow the selection of an appropriate treatment protocol and method of treatment with the highest probability of a successful outcome.

It turns out that traditional methods of statistical analysis are inefficient in the precise determination of the reasons behind infertility and in providing effective predictors of treatment. Univariate analysis only determines the relationship between the analyzed factor and treatment outcome. Multivariate analyses (e.g. multivariate logistic regression) provide models that allow the prediction of pregnancy or lack of pregnancy with a higher level of accuracy. However, these methods include some restrictions that influence their effectiveness and limit their wide clinical application. Therefore, there is a necessity to apply new and more advanced statistical analyses such as data mining methods (Witten et al., 2011). The effort in this area is focused on determining which data mining method would be the best suited to analyze data derived from infertility treatment. Siristatidis et al. (2011) advocate for the usefulness of artificial intelligence methods in analyzing data concerning reproductive medicine. A lot of hope lies in the application of Artificial Neural Networks (ANNs), which so far gives excellent results in predicting negative outcomes of infertility treatment (Milewski et al., 2009). There are also some attempts to apply other methods, such as Basket Analysis (Milewska et al., 2011) and Correspondence Analysis (Milewska et al., 2012) into ART data analysis. It is possible that good statistical results could be produced by combining advanced data mining methods with the feature classification procedures (Milewski, Malinowski et al., 2011, 2012).

The aim of this study is to compare logistic regression analysis (a classical statistical method that can predict effectiveness of infertility treatment) with the ANNs method that resembles the concept of the human brain in analytical ability.

Material and Methods

The data of 1995 infertility patients of the Shore Institute for Reproductive Medicine, Lakewood, NJ, USA, in the age range of 21–45 years old were analyzed. Pregnancy (defined according to a positive pregnancy test that is ≥ 5 IU HCG/ml on days 10–12 after embryo transfer) as a binary variable was the dependent feature in our analysis. Twenty-six various variables of patients' treatment were independent variables. Fourteen of these variables were quantitative and 12 were qualitative (Table 1). The quantitative variables were: age of the patients, number of oocytes retrieved and cultured, semen parameters and hormone levels. The qualitative variables were: diagnosis, type of treatment and stimulation protocol.

Table 1. List of independent variables (quantitative and qualitative)

Quantitative variables		Qualitative variables	
`Age`	Age of woman	`Insem_type`	1-ICSI, 0-classical IVF
`Nr_ET`	Number of transferred embryos	`Tubal`	Tubal factor
`Total_nr_eggs`	Total number of retrieved eggs	`Endometr`	Endometriosis
`Nr_MII_eggs`	Number of mature eggs	`Ovulatory`	Ovulatory factor
`Nr_eggs_ins`	Number of inseminated eggs	`MF`	Male factor
`Nr_2pn`	Number of fertilized eggs	`AMA`	Advanced maternal age
`Nr_cultured`	Number of cultured eggs	`PCOS`	Polycystic Ovary Syndrome
`Nr_clvd`	Number of cleavage embryos	`DOR`	Diminish Ovarian Reserve
`Vol_prewash`	Volume of the semen	`Idiopathic`	Idiopathic factor
`Ct_prewash`	Sperm concentration	`Down_regul`	1-Lupron, 0-Antagon
`Mot_prewash`	Sperm motility	`Lupron_dur`	Lupron duration
`Baseline_FSH`	FSH level on day 3	`HCG_dose`	HCG dose to induce ovulation (250, 5000, 7500, 10000 IU)
`Nr_stim_days`	Number of days on stimulation		
`E2_at_HCG`	E_2 level at HCG injection		

Univariate and multivariate logistic regression analyses were performed using software Stata/IC 12.1 (Stata Corp LP., College Station, TX, USA) to provide predictions for pregnancy occurrence. The data were also analyzed using Artificial Neural Networks technology with application of the software Statistica Data Miner + QC 10.0 (StatSoft, Tulsa, OK, USA). To determine the quality of obtained predictors, the Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) were analyzed (Hanley at al., 1982). Based on Hanley's algorithm, statistically significant differences between the predictors were obtained (Hanley et al., 1997). Statistical significance was determined at the $p < 0.05$ level.

The Logistic Regression Model

The logistic regression analysis is the most appropriate method to determine the relationship between the described above independent variables and pregnancy occurrence among the classical statistical methods. The univariate analysis provides a statistical significance level, an Odds Ratio with 95% Confidence Intervals and a Standard Error value for the probable effect of each independent variable on the dependent variable (pregnancy). Table 2 shows results of the univariate logistic regression analysis.

Table 2. Results of univariate logistic regression analysis

Pregnancy	Odds Ratio	Std. Error	p-value	95% Confidence Interval	
Age	0.950636	0.0107684	<0.001*	0.929763	0.971978
Nr_ET	1.047353	0.0395560	0.221	0.972625	1.127823
Total_nr_eggs	1.036945	0.0068559	<0.001*	1.023594	1.050470
Nr_MII_eggs	1.054609	0.0084598	<0.001*	1.038158	1.071321
Nr_eggs_ins	1.046144	0.0076691	<0.001*	1.031221	1.061284
Nr_2pn	1.066017	0.0104114	<0.001*	1.045805	1.086620
Nr_cultured	1.090955	0.0131257	<0.001*	1.065530	1.116987
Nr_clvd	1.097478	0.0135132	<0.001*	1.071310	1.124286
Vol_prewash	1.007906	0.0310475	0.798	0.948854	1.070632
Ct_prewash	0.999363	0.0011440	0.578	0.997124	1.001608
Mot_prewash	0.997700	0.0019475	0.238	0.993890	1.001524
Baseline_FSH	0.963693	0.0133326	0.008*	0.937912	0.990181
Nr_stim_days	0.983413	0.0267040	0.538	0.932443	1.037170
E2_at_hCG	1.000151	0.0000498	0.002*	1.000054	1.000249
Insem_type	0.733145	0.0684540	0.001*	0.610538	0.880373
Tubal	0.892669	0.1040469	0.330	0.710357	1.121771
Endometr	0.996447	0.1108244	0.974	0.801280	1.239152
Ovulatory	1.106040	0.1507395	0.460	0.846764	1.444704

Pregnancy	Odds Ratio	Std. Error	*p*-value	95% Confidence Interval	
MF	1.136840	0.1037123	0.160	0.950704	1.359418
AMA	0.880833	0.1446499	0.440	0.638425	1.215283
PCOS	1.105390	0.3373659	0.743	0.607754	2.010494
DOR	0.810750	0.0949209	0.073	0.644511	1.019868
Idiopathic	0.854709	0.2195203	0.541	0.516653	1.413962
Down_regul	1.292339	0.1439821	0.021*	1.038825	1.607720
Lupron_dur	0.756088	0.0831845	0.011*	0.609429	0.938040
HCG_dose	0.999952	0.0000198	0.016*	0.999913	0.999991

The 13 independent variables statistically affected the pregnancy variable (see p value with the asterisk). Univariate analyses determine only statistical differences between variables but without the power of prediction. The analyzed variables were not able to efficiently demonstrate a relationship with the dependent variable. Therefore, multivariate analyses were applied to create a prediction model. The multivariate model was built by selecting variables with at least $p < 0.2$ from the univariate analysis. The $p < 0.2$ value was chosen to allow for variables with slightly diminished statistical power to enter into our model. Variables that were strongly correlated with each other were excluded. Finally, the 6 selected variables and the intercept were included into the model (Table 3).

Table 3. Multivariate logistic regression model

Pregnancy	Odds Ratio	Std. Error	*p*-value	95% Confidence Interval	
Age	0.958398	0.0113094	<0.000	0.936486	0.980822
Nr_clvd	1.085128	0.0137621	<0.000	1.058487	1.112439
Insem_type	0.690624	0.0760208	0.001	0.556603	0.856916
MF	1.355477	0.1446616	0.004	1.099635	1.670842
Down_regul	1.226959	0.1455581	0.085	0.972410	1.548143
Lupron_dur	0.630935	0.0739802	<0.000	0.501391	0.793948
Const	2.629308	1.1784360	0.031	1.092294	6.329123

This multivariate logistic regression model allowed us to establish probability of achieving pregnancy by the individual patient (`fit_pr`) and was the searching predictor. Figure 1 presents the ROC curve for the `fit_pr` and Table 4 contains the Area Under the Curve with the 95% Confidence Intervals and Standard Error. The cut-off point for the probability value produced by our model (`fit_pr` = 0.479) was established by applying the minimum sum of squared components method. The sensitivity and specificity were 65.4% and 56.8%, respectively. However, this model is not able to predict presence or lack of pregnancy with perfect accuracy because it misses around 35% of IVF cycles with true pregnancies and more than 43% of cycles with lack of pregnancy.

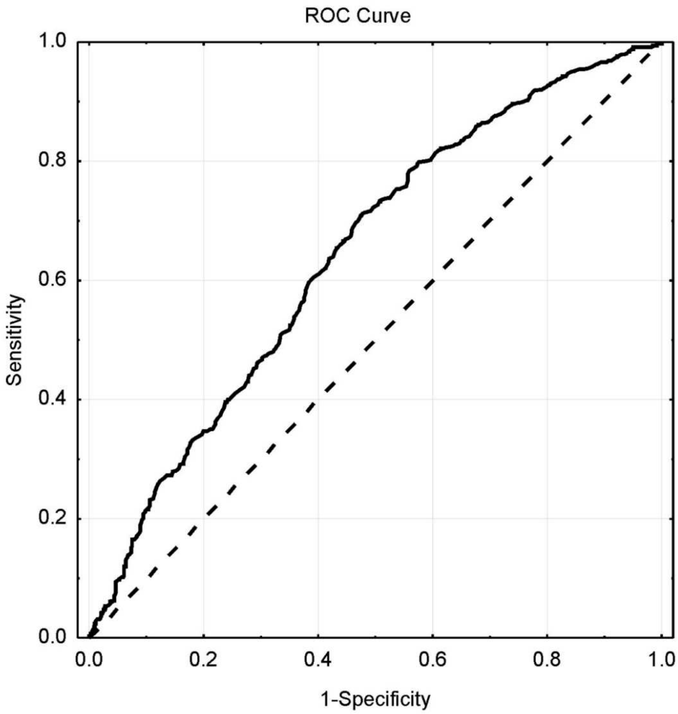

Figure 1. The ROC Curve for the Logistic Regression model

Table 4. Characteristics of the ROC Curve created for the Logistic Regression model

AUC	Std. Error	95% Confidence Interval	
0.642	0.015	0.613	0.671

The Artificial Neural Networks Model

The classical statistical analysis gives us a statistical relationship among the independent and dependent variables with the statistical power for each individual variable. In contrast, the Artificial Neural Networks method creates such a statistical model which will provide the best prognosis for the studied phenomena based on the entered values, without determining the influence of each variable (Tadeusiewicz, 1993). The ANNs method relies on processing data in such a way as the human brain does. The attractiveness of ANNs comes from their remarkable information processing characteristics pertinent mainly to nonlinearity, high level of parallelism, noise and fault tolerance, and learning and generalization capabilities (Basheer et al., 2000). Therefore, it is a useful method in providing prognostic values for the effectiveness of infertility treatment (Milewski et al., 2009; Siristatidis et al., 2011). There are many types of ANNs that are classified according to used structure, network parameters or selected learning algorithms (Osowski, 2009).

To create the best network, the algorithm was run 30 000 times with random startup parameters (type of network, error function, activation functions, an initial value of the generator) each time. Based on the quality of training, testing and validation, the following classifying network was chosen: three-layer perception with 40 neurons in the input layer, 6 neurons in the hidden layer and two neurons in the output layer (for 26 input variables and one output variable). Table 5 contains the parameters of the selected network.

Table 5. Characteristics of selected ANN

Network name	Quality (training)	Quality (validation)	Quality (testing)	Training algorithm	Error function	Activation (hidden)	Activation (output)
MLP 40-6-2	63.777	65.217	64.251	BFGS 21	Entropy	Logistic	Softmax

Based on the created network MLP 40-6-2 the probability of pregnancy for the each patient (`MLP_pr`) was set up. This parameter was used as a predictor for the analysis. Figure 2 shows the ROC curve established for the `MLP_pr`. The Area Under the Curve, along with the 95% Confidence Interval and Standard Error values, is presented in Table 6.

Table 6. Characteristics of the ROC Curve created for the Artificial Neural Networks model

AUC	Std. Error	95% Confidence Interval	
0.703	0.014	0.676	0.730

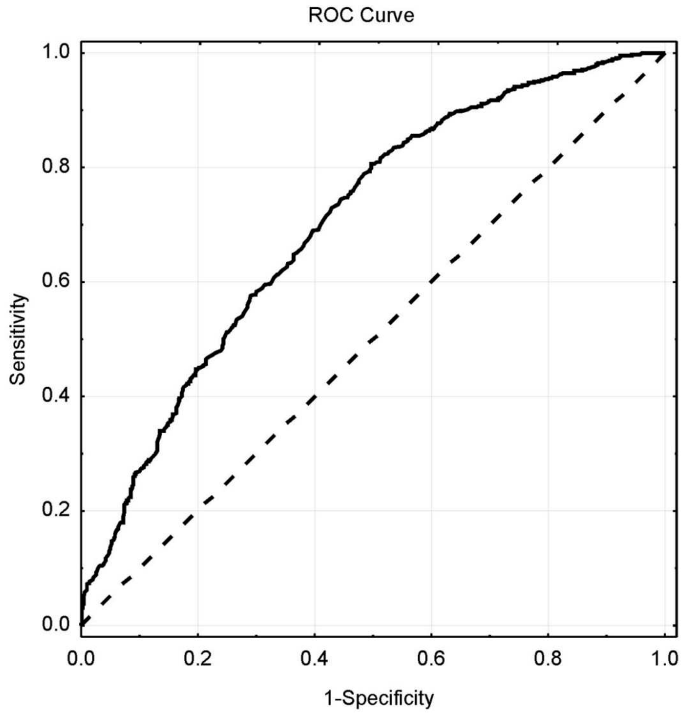

Figure 2. The ROC Curve for the Artificial Neural Networks model

The cut-off point for the probability value produced by the model (`MLP_pr` = 0.477) was established by using the minimum sum of squared components method. The sensitivity and specificity were 69.0% and 60.3%, respectively. It means that our model correctly predicts about 70% of pregnancies and misdiagnoses lack of pregnancy less than 40% of the time.

Comparison of Models and Conclusions

Comparing predictive power of the two studied models, the better results were obtained with the ANNs technologies. The results of the comparison of the AUC for the ROC curves of these two models are shown in Table 7. The statistically significant differences between the predictive powers of both models were at the $p < 0.0001$ level. Comparing the ROC curve in Figure 1 with the curve in Figure 2, it is evident that the shape of the curve for the ANNs is more convex then other one. The Area Under the Curve, sensitivity and specificity for the ANNs were higher (10%, 3.5% and 3.5%, respectively) than for the multivariate logistic regression model.

Table 7. Comparison of two predictors

AUC `MLP_pr`	0.7026
AUC `fit_pr`	0.6423
SD `MLP_pr`	0.0139
SD `fit_pr`	0.0147
AUC difference	-0.0603
SD AUC difference	0.0116
95% CI AUC difference '1	-0.0829
95% CI AUC difference '2	-0.0376
p-value	<0.0001

To obtain a predictive model for clinical treatment, the ANNs technologies are better suited than classical statistical analysis. However, they are not able to detect which variable and to what degree it influences the final results. In contrast, the logistic regression analyses allow for the selection of variables that affect treatment and their statistical power. Thus, for clinical prediction purposes, the ANNs technologies are better to apply but classical technology (in this case logistic regression analysis) is more appropriate for theoretical (scientific) purposes.

REFERENCES

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: Fundamentals, computing, design, and application. *Journal of Microbiological Methods*, 43(1), 3–31.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29–36.

Hanley, J. A., & Hajian-Tilaki, K. O. (1997). Sampling variability of nonparametric estimates of the areas under receiver operating characteristic curves: An update. *Academic Radiology*, 4(1), 49–58.

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

Milewski, R., Jamiolkowski, J., Milewska, A. J., Domitrz, J., Szamatowicz, J., & Wolczynski, 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., Malinowski, P., Milewska, A. J., Czerniecki, J., Ziniewicz, P., & Wolczynski, S. (2011). Nearest neighbor concept in the study of IVF ICSI/ET treatment effectiveness. *Studies in Logic, Grammar and Rhetoric. Logical, Statistical and Computer Methods in Medicine*, 25(38), 49–57.

Milewski, R., Malinowski, P., Milewska, A. J., Ziniewicz, P., Czerniecki, J., Pierzynski, P., & Wolczynski, S. (2012). Classification issue in the IVF ICSI/ET data analysis. *Studies in Logic, Grammar and Rhetoric. Logical, Statistical and Computer Methods in Medicine*, 29(42), 75–85.

Milewski, R., Milewska, A. J., Czerniecki, J., Lesniewska, M., & Wolczynski, S. (2013). Analysis of the demographic profile of patients treated for infertility using assisted reproductive techniques in 2005–2010. *Ginekologia Polska*, 84(7), 609–614.

Milewski, R., Milewska, A. J., Domitrz, J., & Wolczynski, S. (2008). In vitro fertilization ICSI/ET in women over 40. *Przeglad Menopauzalny*, 7(2), 85–90.

Milewski, R., Pierzynski, P., Milewska, A. J., Zbucka-Kretowska, M., & Wolczynski, S. (2012). Concept of system allowing non invasive detection of uterine contractions in women undergoing in vitro fertilization – embryo transfer treatment. *Studies in Logic, Grammar and Rhetoric. Logical, Statistical and Computer Methods in Medicine*, 29(42), 99–106.

Osowski, S. (2009). Artificial neural networks – basic network structures and learning algorithms. *Przeglad Elektrotechniczny*, 85(8), 1–8.

Radwan, J. (2011). Epidemiologia nieplodnosci. In J. Radwan, & S. Wolczynski (Eds.), *Nieplodnosc i rozrod wspomagany* (pp. 11–14). Poznan: Termedia.

Siristatidis, C., Pouliakis, A., Chrelias, C., & Kassanos, D. (2011). Artificial intelligence in IVF: A need. *Systems Biology in Reproductive Medicine*, 57(4), 179–185.

Tadeusiewicz, R. (1993). *Sieci neuronowe*. Warszawa: Akademicka Oficyna Wydawnicza RM.

Thoma, M. E., McLain, A. C., Louis, J. F., King, R. B., Trumble, A. C., Sundaram, R., & Buck Louis, G. M. (2013). Prevalence of infertility in the united states as estimated by the current duration approach and a traditional constructed approach. *Fertility and Sterility*, 99(5), 1324–1331e1.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.). San Francisco, CA, USA: Morgan Kaufmann.