COMPARISON OF MACHINE LEARNING MODELS FOR THE PREDICTION OF LIVE BIRTH FOLLOWING IVF TREATMENT: AN ANALYSIS OF 463,669 CYCLES FROM A NATIONAL DATABASE

Business Insight
Corporation

Business Insight
Corporation

Session title: Session 53: Machine learning and artificial intelligence
Session type: Selected oral communications
Presentation number: O-197

I. Sfontouris[1], A. Patelakis[2], G. Panitsa[2], S. Theodoratos[2], N. Raine-Fenning[1].

1. University of Nottingham, Child Health- Obstetrics and Gynaecology, Nottingham, United Kingdom.

2. Business Insight Co, 20 Station Rd, Cambridge CB1 2JD, United Kingdom.

STUDY QUESTION:

What is the comparative ability of different machine learning models in predicting live birth following IVF based on the analysis of a large national database?

SUMMARY ANSWER:

Deep neural networks were associated with the highest accuracy and specificity for live birth prediction compared to other machine learning models.

WHAT IS KNOWN ALREADY:

The increasing wealth of IVF-treatment data has presented the opportunity, but also the challenge, to develop models for accurate and personalized outcome prediction facilitating optimal treatment design and patient counseling. Machine learning allows the construction of algorithms that can 'learn' from data and make predictions. It is a powerful way to analyze large and complex datasets, which may not be effectively interpreted by the use of conventional statistics. Machine learning has been applied in several fields of healthcare and may prove to be a useful tool in developing accurate personalized predictions in fertility treatment.

STUDY DESIGN, SIZE, DURATION:

This population-based cohort study used anonymous data obtained from the register of the Human Fertilization and Embryology Authority (HFEA), the ART statutory regulator in the UK. A total 463,669 fresh autologous IVF/ICSI, non-PGD cycles, with a full set of data, performed between 1991 and 2012 were analysed to predict live birth per cycle started.

PARTICIPANTS/MATERIALS, SETTING, METHODS:

A predictive model of live birth constructed using latest-technology deep neural networks (DNN) was compared with random forest (RF), decision trees (DT) and Naive Bayes (NB) machine learning models. Cycles were randomly divided into a training and testing set at a 80:20 ratio. The training set was used to develop the prediction model and the testing set to validate model performance. Comparisons were performed using McNemar's test.

MAIN RESULTS AND THE ROLE OF CHANCE:

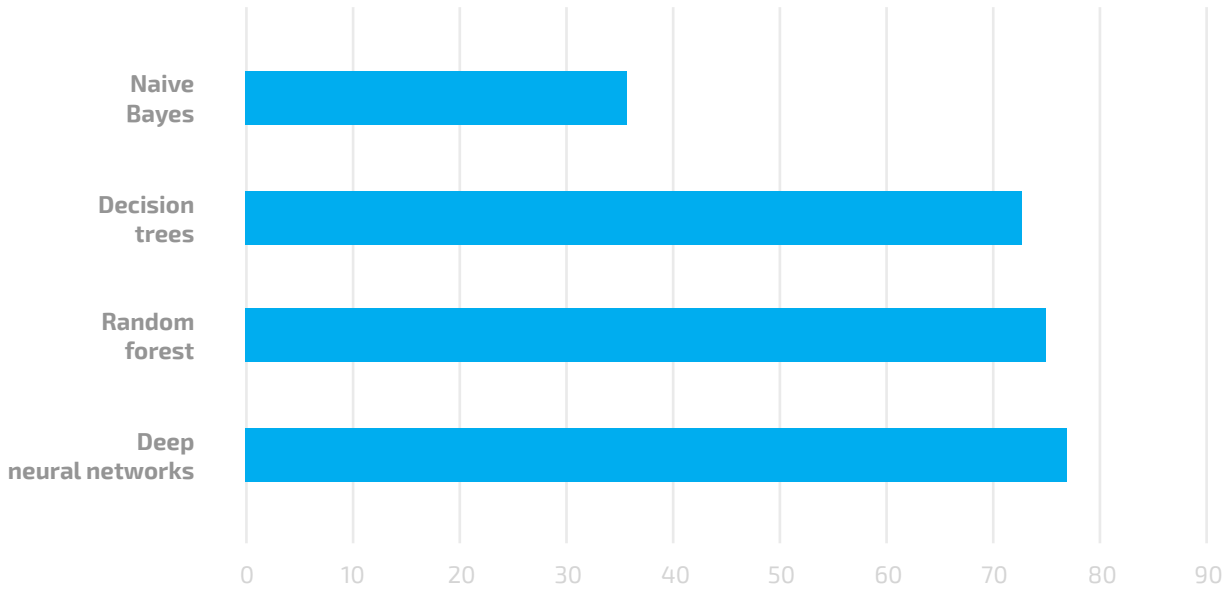
After exclusions, a total 463,669 cycles were included, of which 99,537 [(21.5% (95% CI: 21.4-21.6%)] resulted in a live birth. 370,935 and 92,734 cycles were assigned to the training and testing sets, respectively. Variables up to and including the day of embryo transfer were included in the analysis.

The DNN model was associated with significantly higher accuracy, specificity, positive predictive value (PPV), positive and negative likelihood ratios compared to other machine learning models. Conversely, sensitivity and negative predictive value (NPV) were lower in DNN compared to other models. All differences compared to DNN were statistically significant ($p<0.0001$).

Details of predictive parameters for all machine learning models are shown in Table 1. Values are expressed as percentages/ratios and 95% confidence intervals.

TABLE 1	Deep neural networks	Random forest	Decision trees	Naive Bayes
Accuracy %	76.83 76.59-77.08	74.94 74.66-75.22	72.90 72.61-73.18	35.91 35.61-36.22
Specificity %	94.86 94.72-95.01	91.16 90.95-91.36	87.07 86.82-87.31	18.30 18.02-18.58
Sensitivity %	11.09 10.71-11.49	16.16 15.65-16.67	21.53 20.96-22.11	99.76 99.68- 99.82
Positive Likelihood Ratio	2.16 2.07-2.26	1.83 1.76-1.90	1.67 1.61-1.72	1.22 1.22-1.23
Negative Likelihood Ratio	0.94 0.93-0.94	0.92 0.91-0.93	0.90 0.89-0.91	0.01 0.01-0.02
PPV %	37.20 36.16-38.25	33.52 32.65-34.40	31.48 30.79-32.19	25.20 25.13-25.27
NPV %	79.55 79.48-79.63	79.76 79.65-79.86	80.09 79.96-80.21	99.63 99.51-99.72

Method	*Deep neural networks*	*Random forest*	*Decision trees*	*Naive Bayes*
Accuracy %	**76.83**	**74.94**	**72.90**	**35.91**



LIMITATIONS, REASONS FOR CAUTION:

The anonymised HFEA database holds no information on the number of cycles performed per patient. Therefore, necessary adjustments and calculation of cumulative live birth were not possible. Furthermore, availability of features related to patient demographics, ovarian reserve, baseline and stimulation characteristics, would likely improve the models' predictive ability.

WIDER IMPLICATIONS OF THE FINDINGS:

The DNN model offered the highest specificity and accuracy, efficiently predicting cycles not leading to live birth. Prediction of unsuccessful treatment may be of value for counseling and patient assessment by clinics and funding bodies. Ongoing research using large datasets will attempt further improvements of the model's overall predictive performance.

Trial registration number:
Not applicable

Keywords:
machine learning artificial intelligence
live birth
ivf
fertility treatment

Business Insight Corporation Limited.

20 Station Rd,
Cambridge CB1 2JD,
United Kingdom.
info@businessinsightco.com