

O-203 Application of machine learning to predict aneuploidy and mosaicism in embryos from in vitro fertilization (IVF) cycles

J.A. Ortiz¹, R. Morales¹, B. Lledo¹, E. Garcia-Hernandez¹, A. Cascales¹, J.A. Vicente², J. González³, J. Ten⁴, A. Bernabeu⁵, J. Llacer⁵, R. Bernabeu⁵

¹Instituto Bernabeu, Biología Molecular y Genética, Alicante, Spain ;

²Universidad Nacional de Educación a Distancia UNED, Economía Aplicada y Estadística, Madrid, Spain ;

³Universidad Nacional de Educación a Distancia UNED, Economía de la Empresa y Contabilidad, Madrid, Spain ;

⁴Instituto Bernabeu, Embriología, Alicante, Spain ;

⁵Instituto Bernabeu, Medicina Reproductiva, Alicante, Spain

Study question: Is it possible to predict the likelihood of an IVF embryo being aneuploid and/or mosaic using a machine learning algorithm?

Summary answer: There are paternal, maternal, embryonic and IVF-cycle factors that are associated with embryonic chromosomal status that can be used as predictors in machine learning models.

What is known already: The factors associated with embryonic aneuploidy have been extensively studied. Mostly maternal age and to a lesser extent male factor and ovarian stimulation have been related to the occurrence of chromosomal alterations in the embryo.

On the other hand, the main factors that may increase the incidence of embryo mosaicism have not yet been established.

The models obtained using classical statistical methods to predict embryonic aneuploidy and mosaicism are not of high reliability. As an alternative to traditional methods, different machine and deep learning algorithms are being used to generate predictive models in different areas of medicine, including human reproduction.

Study design, size, duration: The study design is observational and retrospective. A total of 4654 embryos from 1558 PGT-A cycles were included (January-2017 to December-2020). The trophoctoderm biopsies on D5, D6 or D7 blastocysts were analysed by NGS. Embryos with $\leq 25\%$ aneuploid cells were considered euploid, between 25-50% were classified as mosaic and aneuploid with $>50\%$.

The variables of the PGT-A were recorded in a database from which predictive models of embryonic aneuploidy and mosaicism were developed.

Participants/materials, setting, methods: The main indications for PGT-A were advanced maternal age, abnormal sperm FISH and recurrent miscarriage or implantation failure. Embryo analysis were performed using Veriseq-NGS (Illumina).

The software used to carry out all the analysis was R (RStudio). The library used to implement the different algorithms was caret. In the machine learning models, 22 predictor variables were introduced, which can be classified into 4 categories: maternal, paternal, embryonic and those specific to the IVF cycle.

Main results and the role of chance: The different couple, embryo and stimulation cycle variables were recorded in a database (22 predictor variables). Two different predictive models were performed, one for aneuploidy and the other for mosaicism. The predictor variable was of multi-class type since it included the segmental and whole chromosome alteration categories.

The dataframe were first preprocessed and the different classes to be predicted were balanced. A 80% of the data were used for training the model and 20% were reserved for further testing. The classification algorithms applied include multinomial regression, neural networks, support vector machines, neighborhood-based methods, classification trees, gradient boosting, ensemble methods, Bayesian and discriminant analysis-based methods. The algorithms were optimized by minimizing the Log_Loss that measures accuracy but penalizing misclassifications.

The best predictive models were achieved with the XG-Boost and random forest algorithms. The AUC of the predictive model for aneuploidy was 80.8% (Log_Loss: 1.028) and for mosaicism 84.1% (Log_Loss: 0.929). The best predictor variables of the models were maternal age, embryo quality, day of biopsy and whether or not the couple had a history of pregnancies with chromosomalopathies. The male factor only played a relevant role in the mosaicism model but not in the aneuploidy model.

Limitations, reasons for caution: Although the predictive models obtained can be very useful to know the probabilities of achieving euploid embryos in an IVF cycle, increasing the sample size and including additional variables could improve the models and thus increase their predictive capacity.

Wider implications of the findings: Machine learning can be a very useful tool in reproductive medicine since it can allow the determination of factors associated with embryonic aneuploidies and mosaicism in order to establish a predictive model for both. To identify couples at risk of embryo aneuploidy/mosaicism could benefit them of the use of PGT-A.

Trial registration number: Not Applicable