

RESEARCH

Open Access



A novel machine-learning framework based on early embryo morphokinetics identifies a feature signature associated with blastocyst development

S. Canosa^{1,2*†}, N. Licheri^{3†}, L. Bergandi⁴, G. Gennarelli^{1,2}, C. Paschero¹, M. Beccuti³, D. Cimadomo⁵, G. Coticchio⁶, L. Rienzi^{5,7}, C. Benedetto¹, F. Cordero^{3†} and A. Revelli^{1,8†}

Abstract

Background Artificial Intelligence entails the application of computer algorithms to the huge and heterogeneous amount of morphodynamic data produced by Time-Lapse Technology. In this context, Machine Learning (ML) methods were developed in order to assist embryologists with automatized and objective predictive models able to standardize human embryo assessment. In this study, we aimed at developing a novel ML-based strategy to identify relevant patterns associated with the prediction of blastocyst development stage on day 5.

Methods We retrospectively analysed the morphokinetics of 575 embryos obtained from 80 women who underwent IVF at our Unit. Embryo morphokinetics was registered using the Geri plus® time-lapse system. Overall, 30 clinical, morphological and morphokinetic variables related to women and embryos were recorded and combined. Some embryos reached the expanded blastocyst stage on day 5 (BL Group, $n = 210$), some others did not (nBL Group, $n = 365$).

Results The novel EmbryoMLSelection framework was developed following four-steps: Feature Selection, Rules Extraction, Rules Selection and Rules Evaluation. Six rules composed by a combination of 8 variables were finally selected, and provided a predictive power described by an AUC of 0.84 and an accuracy of 81%.

Conclusions We provided herein a new *feature-signature* able to identify with a high performance embryos with the best developmental competence to reach the expanded blastocyst stage on day 5. Clear and clinically relevant cut-offs were identified for each considered variable, providing an objective tool for early embryo developmental assessment.

Keywords Assisted reproduction, Blastocyst development, Artificial intelligence, Machine learning, Embryo developmental patterns

[†]S. Canosa and N. Licheri should be regarded as joint first authors.

[†]F. Cordero and A. Revelli should be regarded as joint last authors.

*Correspondence:

S. Canosa

s.canosa88@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

To date, an objective method to evaluate human embryo competence is still lacking, as conventional morphology is highly subjective, being performed using a wide range of grading systems affected by inter and intra-observer variability [1]. Indeed, consistency and reproducibility of embryo evaluation are still not guaranteed among different IVF laboratories [2]. In addition, standard grading only provides a momentary view of embryonic morphology at a specific time point, while developmental changes over multiple time points may provide a more robust judgement of embryo potential [3]. More than a decade ago, Time-Lapse Technology (TLT) has been introduced to perform a real-time, dynamic observation of pre-implantation embryo development. Morphokinetic events can be monitored at the exact time of occurrence, providing new insights into several steps of in vitro growth and, ultimately, suggesting objective data with potential clinical relevance [4]. TLT was claimed by some authors to improve embryo selection and IVF outcome [3–6]. However, its efficacy is still matter of debate [7–10], being mainly ascribed to unperturbed culture conditions (temperature, atmosphere), rather than to the improved identification of reliable morphokinetic biomarkers of embryo competence [8, 11]. Indeed, the real time observation of crucial events occurring during in vitro growth revealed a number of parameters that were associated to embryo development potential [4]. Looking forward, significant improvements of the technologies associated to TLT will drive a more detailed knowledge and understanding of the early developmental kinetics of human embryos [12].

Recently, Artificial Intelligence (AI) has rapidly developed in various fields, including human embryology [13, 14]. Artificial intelligence (AI) may be used as a tool to assist embryologists in daily activities (e.g. morphological selection of embryos to transfer or cryopreserve), as it is able to analyse a huge number of heterogeneous data, such as those relating to embryo development provided by TLT systems [15, 16]. Machine Learning (ML) algorithms allow the identification of the most relevant variables (also called features) included in a bulk set of data, and may be used to develop complex prediction models of embryo growth [17, 18]. Among the above-mentioned models, however, no one was able to identify clear and clinically relevant cut-offs for morphodynamic features to provide an objective tool to assist embryologists during embryo assessment.

In the present study, we aimed (i) at developing a novel ML framework using quantifiable values for both morphological and morphokinetic data of 575 in vitro-produced embryos, and (ii) at identifying which features of early embryo morphokinetics had key relevance in predicting the timely (day 5) growth to expanded blastocyst.

Material and methods

Patient cohorts

Training cohort

The study was carried out in accordance with the Declaration of Helsinki and was authorized as a retrospective observational study by the local Ethical Committee (authorization number: 0056908). A signed informed consent was obtained from all patients. The analysis included 575 embryos obtained in 80 IVF cycles performed at Physiopathology of Reproduction and IVF Unit of S. Anna Hospital (Turin, Italy) in 80 women (included only once) receiving the transfer in uterus of a single fresh blastocyst on day 5 between March 2018 and March 2020. We selected this patient population in order to have at least one viable blastocyst on day 5 to include into our analysis. These patients had mean age 35.3 ± 3.5 years (range 25–42), body mass index 24.2 ± 4.8 (range 18–25), ovarian reserve markers suggesting normal responsiveness to FSH stimulation (serum day 3 FSH < 12 IU/l, antral follicle count 8–18, anti-mullerian hormone 2.5–4 ng/ml) and were all stimulated using the GnRH-agonist “long” protocol. Patients with polycystic ovary syndrome (PCOS), ovarian endometriosis and/or unfavourable biomarkers leading to an expected poor/sub-optimal responsiveness to FSH [19] were excluded. Patients’ clinical characteristics and variables related to IVF cycle were recorded, including the total dose of exogenous gonadotropins, the number of retrieved COCs (cumulus-oocyte complexes), the ovarian sensitivity index ($OSI = \text{retrieved COCs} \times 1000 / \text{total gonadotropin dose}$) [20], the fertilization, cleavage, and viable blastocyst (day 5 + day 6) formation rate.

Validation cohort

The performance of the *selected* rules was tested by five classification algorithms on an independent validation cohort of 81 embryos obtained from other 10 patients receiving the transfer in uterus of a single fresh blastocyst on day 5 between February and March 2020. These patients had mean age 35.5 ± 3.1 years (range 25–42), body mass index 24.6 ± 2.6 (range 18–25), ovarian reserve markers suggesting normal responsiveness to FSH stimulation (serum day 3 FSH < 12 IU/l, antral follicle count 8–18, anti-mullerian hormone 2.5–4 ng/ml) and were all stimulated using the GnRH-agonist “long” protocol.

Controlled ovarian stimulation and IVF procedures

The gonadotropin-releasing hormone (GnRH)-agonist “long” protocol with recombinant FSH (Gonal-F®, Merck, Germany) at individually tailored daily dose (100–300 IU s.c.) was always used to carry out controlled ovarian stimulation (COS). However, the observed results are likely to be applicable also to GnRH-antagonist cycles, as a previous study demonstrated that the OSI is not significantly

affected by the type of pituitary suppression, being highly consistent when the same patient undergoes GnRH-agonist “long” protocol and then GnRH-antagonist protocol or vice-versa [21]. Circulating estradiol (E2) assessment and transvaginal ultrasound (US) examination were performed every second day from stimulation day 7 in order to monitor follicular growth. The FSH dose was adjusted accordingly. A single subcutaneous injection of 10,000 IU hCG (Gonasi HP, IBSA, Switzerland) was administered to trigger ovulation when at least two follicles reached 18 mm mean diameter, with appropriate E2 levels. US-guided oocyte retrieval (OPU) was performed 35–37 h after hCG trigger under local anaesthesia (paracervical block).

Sperm concentration, motility, and morphology were assessed according to the World Health Organization guidelines [22]. Raw semen was assessed for both progressive and non-progressive motility. After density gradient centrifugation, sperm motility was expressed in terms of concentration of activated cells with progressive motility.

At OPU, the aspirated follicular fluids were immediately observed under a stereomicroscope, cumulus-oocyte complexes (COCs) were washed in buffered medium (Flushing medium, Cook Ltd., Ireland) and oocytes were inseminated within 4 h using either conventional IVF (34%, 27/80 cycles) or ICSI (66%, 53/80 cycles), according to semen quality. When performing ICSI, 2 h after OPU oocytes and cumulus cells (CCs) were separated from each COC by gently pipetting in a 40- μ l HEPES buffered medium containing 80 IU/ml hyaluronidase (Synvitro Hyadase, Origio Medicult, Denmark) [23]. Normal fertilization was confirmed when two pronuclei (2PN) and the extrusion of the second polar body were observed within 16–18 h after insemination.

Annotation of embryo morphokinetics

A total of 575 embryos were included into our analysis, obtained either with conventional IVF ($n=216$) or ICSI ($n=359$). Embryos were cultured in the Geri plus[®] TLS (Genea Biomedx, Australia), that is equipped with an integrated embryo monitoring system to observe one zygote/microwell, as previously described [24]. The dish format allowed the observation of each embryo individually at 11 different focal planes, even if all embryos shared a common 80- μ l medium drop. Up to day 3, embryos were cultured in pre-equilibrated Cleavage medium (Cook, Ireland) overlaid with mineral oil; then, a change of medium was performed, and the new medium (Blastocyst medium, Cook) was kept until the blastocyst stage. Bright-field images were captured by Geri plus[®] system every 5 min from the time of fertilization until day 5, when embryo transfer (ET), cryopreservation or discharge occurred.

Embryo morphological evaluation was first performed by one single senior embryologist on day 2 using the Integrated Morphology Cleavage Score (IMCS) [25], and then repeated on day 5 according to standardized criteria [1]. Notably, IMCS was based on the evidence of implantation and clinical pregnancy after double ET on day 2, and was incorporated into a complex prediction model for IVF outcome, recently shown to predict live birth with a remarkably good precision [26]. However, prospective application of that model proved that it was highly effective as a means of embryo selection to reduce twin implantation rates (from 28 to 2%) for both SET and DET [27]. Furthermore, all videos collected by Geri plus[®] were analysed by the same senior embryologist, and based on ESHRE recommendations [4] the following morphokinetic parameters (times) were manually annotated on all generated embryos, although affected by unavoidable intra-operator variability: pronuclear appearance (tPNa), pronuclear fading (tPNf), completion of cleavage to two, three, four, and eight cells (t2, t3, t4, and t8, respectively), time intervals tPNf-tPNa, t2-tPNf, t3-t2, t4-t3, t4-t2, and t8-t4. Day 5 blastocyst formation was assessed at the same time interval (116 ± 2 h post insemination) for all embryos. Embryos reaching the expanded blastocyst stage on day 5 (score 3 according to [1]) were included in the Blastocyst Group (BL Group, $n=210$), whereas those arrested or progressed to a stage earlier than expanded blastocyst were included in the Not-expanded Blastocyst Group (nBL Group, $n=365$), as previously described [28]. We decided to consider blastocyst expansion as the main criteria to identify BL or non-BL group, irrespectively of ICM and TE morphology, because it was recently confirmed that although embryologists may adopt the same grading scheme their agreement is limited, especially when the morphological quality is low [29–31].

Training dataset preparation

A total of 30 variables among those currently recorded during clinical routine were considered for each embryo and divided into three categories: (i) woman-related ($n=6$) describing patients clinical characteristics: age, BMI, day 3 FSH, AMH, antral follicle count (AFC), years of infertility; (ii) COS-related ($n=10$), describing the clinical characteristics and outcomes of ovarian stimulation: total exogenous FSH, peak E2, OSI, number of retrieved oocytes, number of mature oocytes, maturation rate, number of fertilized oocytes, fertilization rate, number of cleaved embryos, cleavage rate; (iii) embryo-related ($n=14$), describing embryo morphology and morphokinetics: insemination technique, IMCS score on day 2, tPNa, tPNf, t2, t3, t4, t8, tPNf-tPNa, t2-tPNf, t3-t2, t4-t3, t4-t2, t8-t4. Specifically, all the considered variables ($n=30$) were evaluated for each embryo obtained by ICSI ($n=359$), whereas for embryos obtained by

conventional insemination ($n=216$), tPNa and tPNf-tPNa were not annotated, as in this case normal fertilization was checked only after cumulus cells removal. Overall, this dataset composed the Training cohort used for ML framework development.

Development of the EmbryoMLSelection framework

The generation process of the novel EmbryoMLSelection framework was composed of four modules: Features Selection, Rules Extraction, Rules Selection and Rules Evaluation (Supplementary Fig. 1), in order to easily analyze embryo data and build a model able to distinguish between two groups of embryos: one characterized by a high likelihood of reaching the expanded blastocyst stage on day 5 (referred to as the BL group) and the other with a low chance (referred to as the nBL group). The initial component of EmbryoMLSelection is Feature Selection, which involves identifying the most informative features for the task while eliminating noisy, non-informative, irrelevant, and redundant features. The resulting list of features is then inputted into the subsequent step, called Rule Extraction, focusing on exploring rule-based models. Subsequently, a Rule Selection process is performed to eliminate unpredictable or redundant rules. In detail, each rule was evaluated in terms of the number of true and false positive/negative samples classified; these values were used to compute the Matthews Correlation Coefficient MCC, a measure of the quality of the binary classification prediction specifically for the imbalance dataset. MCC reflects how well the rules were able to distinguish between the class: specifically, MCC was used to identify those rules having a limited overlap in terms of both the features considered and the samples correctly classified, leading to an increase in the number of samples correctly classified. The MCC values must be interpreted as a correlation index (i.e. positive/negative values indicates direct/inverse relationship, while 0 implies random correlation). The Rules Evaluation phase is then implemented to assess the model's performance by measuring how accurately its predictions align with the observed data. A graphic formalism was adopted to visualize the functional dependencies among variables.

The machine learning methods devoted to the classification task are generally grouped into four overlapping categories: (i) the geometric models represent instances as points in a high-dimensional Euclidean space and exploit spatial concepts (e.g. distances, lines, planes) to make decisions. In this category, we have implemented two algorithms support vector machines (SVM) and k-nearest neighbors (kNN). (ii) The probabilistic models assume there is an underlying random process modeling the relationship between the data and the target variable (i.e. BL or nBL groups), and try to model it using probability distributions. Logistic regression estimates the

probability of a binary outcome by fitting a logistic function to the linear combination of input features, while Naive Bayes is based on Bayes' theorem with the naive assumption of feature independence. (iii) The rule-based models partition the instance space into instance space segments through a set of logical rules, using either a list or a tree-based structure. Decision trees are a popular rule-based model for classification and regression tasks; in this category, we have implemented the Random Forest algorithm. Finally, (iv) ensemble methods are related to a set of techniques wherein a target function is learned by training multiple individual learners and then consolidating their predictions. Gradient boosting and Ada Boosting are the two methods implemented in this category. The AUC and accuracy values reported in the paper are the highest values obtained from each of the five classification algorithms.

The analysis workflow was implemented using Python 3 programming language [32]. detailed description of the modules and the methodology adopted is provided in the [supplemental Material and Methods](#). EmbryoMLSelection framework was registered in the Docker image, available at https://hub.docker.com/r/qbioturin/embryo_ml_workflow. The web page contains the list of requirements and the command line to run the code. An example of how to run the workflow using the Embryos input data and the complete list of the parameters is provided in order to ensure both functional and computational reproducibility of the experiments.

Statistical analysis

Principal Component Analysis (PCA) was executed using the scikit-learn library in Python. The heatmap, illustrating R coefficients derived from Pearson correlation between feature pairs, was generated using the Python Data Analysis Library called pandas. The arrangement of rows and columns is determined by grouping features based on their types. Prior to PCA and correlation analysis, feature values are normalized as z-scores.

Results

Correlation between the considered variables

Supplementary Table 1 summarizes the clinical characteristics of the 80 patients included into the Training dataset and the outcome of their IVF cycles. Overall, 575 embryos were obtained, of which 210 (36.5%) progressed to the expanded blastocyst stage on day 5 (BL group) whereas 365 (63.5%) did not (nBL Group) (Fig. 1A).

At first, all embryos were considered together for a preliminary analysis aimed at identifying a correlation cluster between the considered variables and embryo development where we applied the Principal Component Analysis (PCA) to generate a labelled scatter plot according to the clustering category of the embryos (BL or nBL).

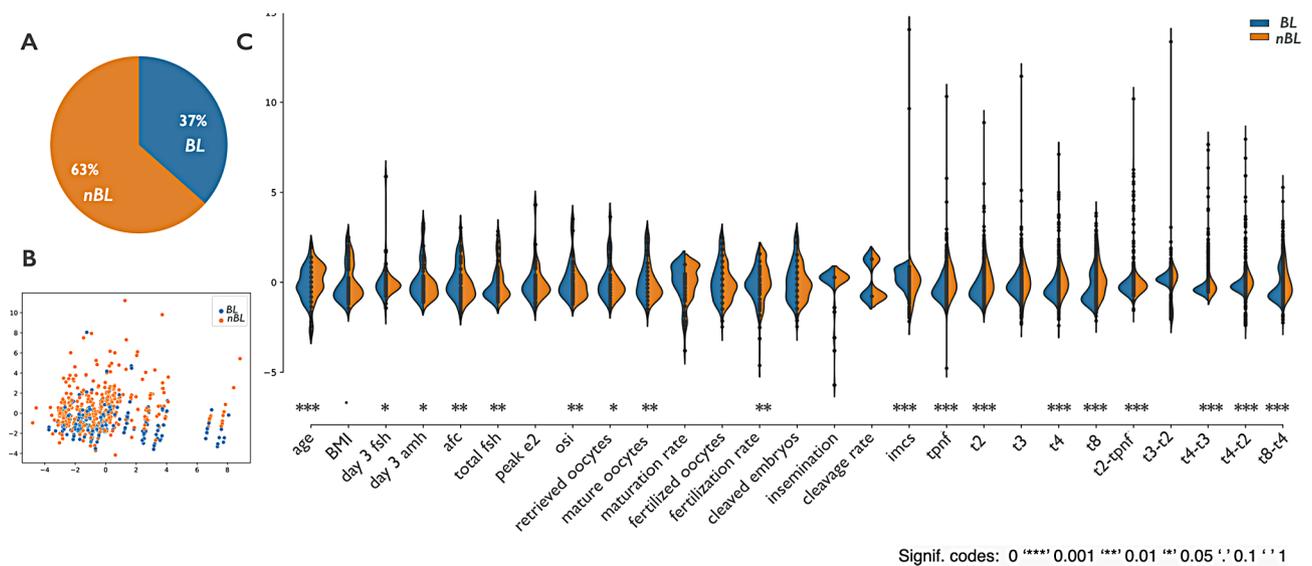


Fig. 1 (A) Pie chart showing the distribution of the 575 embryos: embryos progressed to the expanded blastocyst stage on day 5 (BL, blue) or not (nBL, orange). (B) Scatter plot obtained from a dimensionality reduction technique (principal component analysis) considering all 575 embryos and all variables. The color of the embryos identifies those grown to the expanded blastocyst stage on day 5 (BL, blue) or not (nBL, orange). (C) Split violin plots of the distribution of the z-score of the value of all features distinguishing embryos grown to the expanded blastocyst stage on day 5 (BL, blue) or not (nBL, orange). Each violin plot is divided in half allowing for a clear observation of differences in the distribution of values between the BL and nBL groups. The p-values expressed by a significance code are computed by a Kruskal-Wallis rank sum test

Although there was not a clear separation of the two embryo populations (Fig. 1B), different density curves could be noted after comparing the range of values of BL with those of nBL, as shown in Fig. 1C (see also Supplementary Table 2).

The inter-variables correlation analysis revealed two major clusters, the first between woman-related and COS-related variables, the second within embryo-related variables, as depicted in the heat map (Fig. 2). Interestingly, the insemination technique showed either weak or no correlation with the other variables (Fig. 2). In addition, a clear cluster was not detectable considering each set of variables to discriminate the BL and nBL embryo groups by scatter plots (Fig. 2), thus concluding that our preliminary analysis was not enough to identify predictive variables of embryo development.

Selection of rules associated with blastocyst development

To pursue our aim to define a specific feature signature, the first phase of EmbryoMLSelection framework was applied to identify the most predictive variables. This process exploits multiple selection algorithms (e.g. filter and embedded methods) to explore the ideal set composed by balanced cut-off between the number of features and the power of their association with embryo development. The selection strategy was performed on the Training cohort using the 70% ($n=403$) of the total number of embryos defined as training dataset. The identified set of features was tested on a classification task using stratified 10-fold cross validation repeated 100 times against

the test set composed of the remaining 30% ($n=172$) of the embryos of the Training cohort. A total number of 12 variables (OSI, Maturation rate, Fertilization rate, Cleavage rate, Age, Day 3 FSH, AFC, IMCS, tPNf, t4, t4-t3, t8-t4) composed a preliminary set providing the highest area under the curve (AUC=0.74 obtained by the gradient boosting classification algorithm). In a second phase, these variables were managed to define the set of rules (combination of variables) able to discriminate embryos of BL and nBL groups. The Rules extraction module was then applied to generate 71 rules from the training set. To explore the association of the *extracted* rules within each other and with blastocyst development, the Rule Selection module was applied to implement the Matthews Correlation Coefficient (MCC). MCC was used to remove the unproductive or redundant rules, leading to a new set of 23 selected rules. Figure 3A is the correlation graph obtained from the visualization module, where the rules were represented as nodes whose size is correlated with the relevance of the primary outcome (blastocyst development), while the edge between two nodes represented a correlation value for those rules higher than 0.8 (computed by Matthews Correlation Coefficient (MCC)). It is interesting to note that 8 out of 23 rules were isolated vertices (Fig. 3A, red nodes on the right) highlighting their low (i.e. minor than 0.8) functional dependency from the others rules but, at the same time, their high relevance with the primary outcome (the size of the node was generally large).

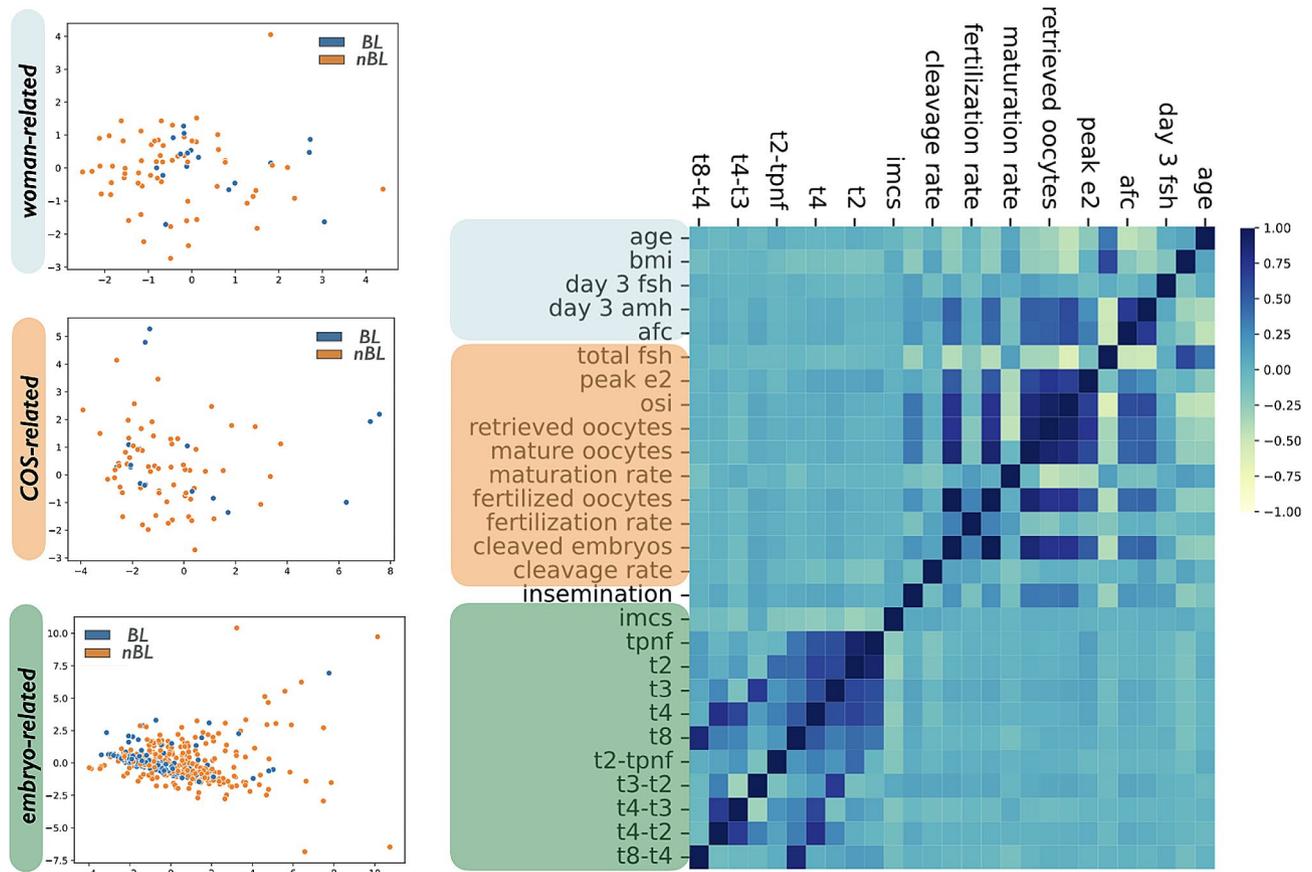


Fig. 2 Heatmap plot showing the Pearson correlation value among all variables (right panel of the figure). Scatter plots computed through a dimensionality reduction technique (principal component analysis) for each set of variables, independently (leftpanel)

We defined a *feature-signature* composed of 6 out of the 23 selected as the rule combination with the highest discrimination power. This evaluation was performed in terms of classification performance expressed as AUC value. Specifically, Fig. 3B shows the line plot reporting the AUC values (y-axis) computed by the seven classification algorithms for the classification of the expansion or not of the blastocyst, considering a given number of features (x-axis). The maximum value of the AUC (0.842) was reached by the AdaBoosting classifier considering 6 rules (yellow vertical ribbon); all coefficients computed for all the seven classifiers are reported in Supplementary Table 3. Figure 3C reported the 6 rules and their feature composition based on: one COS-related feature (OSI), two woman-related features (i.e. Age and AFC), and five embryo-related variables (i.e. tpnf, t4, t4-t3, t8-t4, imcs) (Fig. 3D).

Validation of the selected rules on an independent dataset

The performance of the 6 *selected* rules was tested by five classification algorithms on an independent validation cohort of 81 embryos obtained from other 10 patients with clinical characteristics comparable to the women

included in the study (Supplementary Table 1). We confirmed that the rules composing the *feature-signature* showed a predictive performance with AUC=0.842 and accuracy of 0.81. For the sake of convenience, the EmbryoMLSelection framework was also applied in a setting in which the variables were not previously selected. Specifically, all variables were used together in order to obtain at first 131 extracted rules, showing a high functional dependency highlighted by the high number of arcs (Supplementary Fig. 2A). From a subset of 30 selected rules, 17 were finally derived based on the AUC in order to choose the best combination of rules with the highest discrimination power (Supplementary Fig. 2B). However, in the validation cohort these rules reached lower values in terms of both AUC (0.79) and accuracy (0.70) with respect to the previously identified *feature signature*. Finally, when considering only the embryo-related variables, 146 rules were extracted with high functional dependency (Supplementary Fig. 3A). A subset of 30 selected rules allowed to identify, in the validation cohort, a final number of 21 rules reaching an AUC value=0.74 and an accuracy value=0.68 (Supplementary Fig. 3B). This performance confirmed that

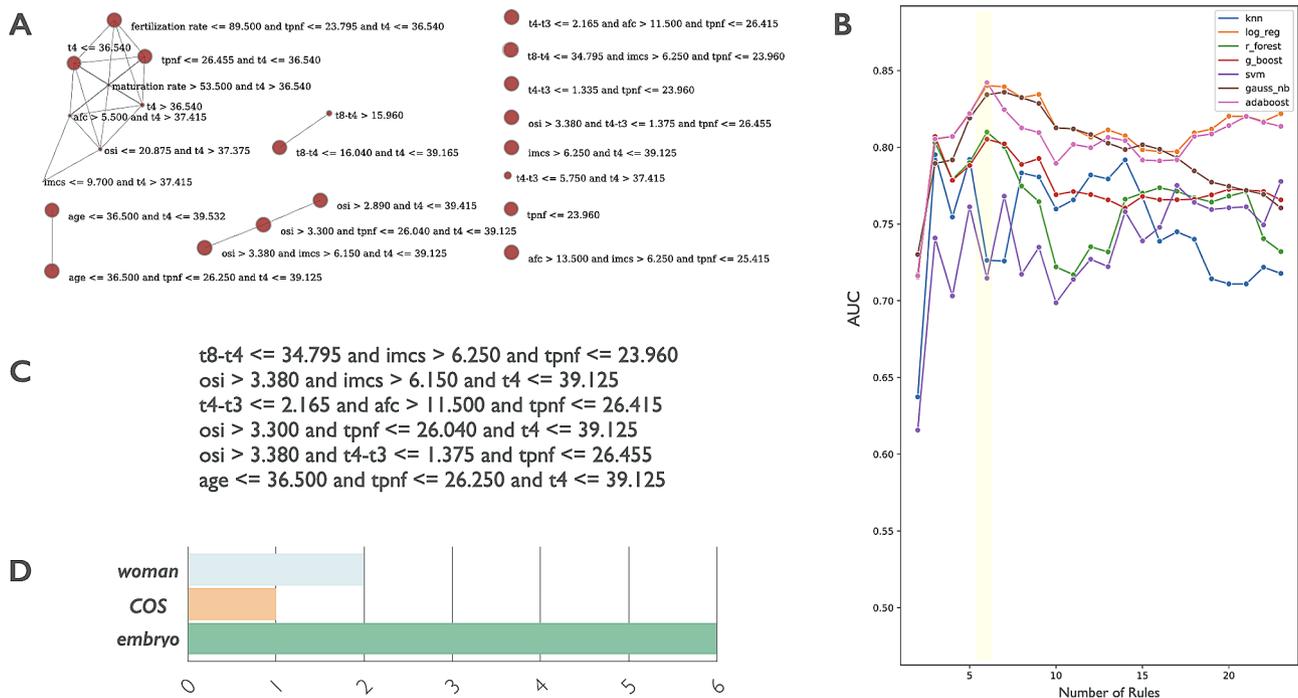


Fig. 3 (A) Correlation graph of the selected rules. Vertices represent the rules and arcs are reported only for a correlation value >0.8 (computed by Matthews Correlation Coefficient (MCC)). (B) Line plot representing the ability of different combinations of classifiers to classify the expanded or not expanded blastocyst stage. Each dot corresponds to the AUC computed using a different number of rules in input. (C) Set of 6 rules of the *feature-signature* with (D) their composition in terms of variable category

embryo-related variables alone are not enough to accurately describe blastocyst development, and that it is necessary to consider the overall set of variables while designing the framework.

Discussion

Artificial Intelligence is one of the most promising, objective methodologies aimed at standardizing embryo assessment in human IVF. An intriguing application of AI within the IVF laboratory is providing new knowledge on cellular profiles regulating embryo in vitro growth and embryo competence [14]. In particular, the use of explainable methods, as well as the rules-based models, makes the classification problem (i.e., the association of a particular object to a class considering a subset of features describing the object) understandable, transparent, and interpretable.

The introduction of TLT offers the possibility to obtain a vast bulk of data on the kinetic of human embryo growth, producing a much more detailed timeline of dynamic events as well as showing previously unrecognizable phenomena [33].

Exploring the ability of all features, clinical, TLT and patient’s characteristics to distinguish BL versus nBL embryos, become clear that dimensionality reduction methods, as well as Principal Component Analysis, ineffective in discerning between the embryos.

In the present study, we describe the performance of the novel EmbryoMLSelection framework in identifying a set of rules associated to a timely embryo development to the expanded blastocyst stage on day 5, called *feature-signature*. The rationale behind a two-step process of features selection, rule extraction and rule selection from a large number of variables/embryo ($n=30$) was the identification of a set of rules (from an initial number of 71 to the final 6) identifying an *feature-signature* composed of relevant features ($n=8$), describing cleavage stage embryos able to timely (within day 5) progress to the expanded blastocyst stage. As lower implantation and clinical pregnancy rates were reported in case of transfer of slow-growing blastocysts vs. fully expanded day 5 blastocysts [34, 35], probably due both to a poor embryo competence and to the loss of embryonic-endometrial synchrony [36], we considered the fully expanded blastocyst on day 5 as the optimal development stage, that confers the highest probability of embryo implantation. In fact, the number of days of blastocyst development represents the developmental potential of a blastocyst [37] and affects the outcome of transfer [38]. In addition, the developmental potential of Day 5, 6 and 7 blastocysts decreases gradually with the extension of culture time [39]. Therefore, the conventional practice in the laboratory is to select blastocysts for transfer, biopsy or cryopreservation, starting from expanded day 5 blastocysts.

In our dataset, 36.5% progressed to the expanded blastocyst stage on day 5 (BL group defined according to the score 3 provided by the Istanbul Consensus) whereas 63.5% did not (nBL Group). This data should be discussed in relation to the overall blastocyst formation rate of 53.8% observed in the enrolled patients. In fact, according to the Vienna Consensus [40], we considered also blastocysts with expansion score of 2 in the KPI “blastocyst formation rate”, excluded from the BL group of our study. For this reason, we believe we can exclude any selection bias from the patients population ensuring an adequately powered analysis during the framework development. Embryo selection models developed using morphokinetic parameters were previously shown to predict blastocyst development [41, 42]. Furthermore, the application of machine-learning technology provided an algorithm able to predict clinical pregnancy and live birth rate by analysing embryo morphokinetics [43]. Giscard d’Estaing [18] used a machine-learning system in order to build up a score for blastocyst formation with a prediction power having AUC=0.634. In another study, the prediction accuracy of embryo assessment performed by experienced embryologists with morphokinetic grading methods added to conventional static morphology was shown to range between 60% and 70%, with AUC=0.63–0.70 [44]. In a previous study, the efficacy of six in-house embryo-selection algorithms (ESAs) was investigated in a set of known implantation embryos [45]. Interestingly, although the primary endpoint considered and the nomenclature adopted were different, we both observed that tPNf, s2 (i.e. t4-t3) and cc3 (i.e. t8-t4) were associated to embryo developmental and reproductive competence. Since their results highlighted that ESAs are usually specific to the patient, treatment, and environment, we agree that currently available algorithms should be carefully validated before consider clinical applicability and, more importantly, they risk to lose their diagnostic value when externally applied. Herein, we provide evidence that the novel EmbryoMLSelection framework allowed to perform a more precise evaluation of embryo dynamic growth with a performance described by AUC=0.84 and accuracy of 81%. Notably, the Rules Selection step ensured such an increased performance providing a concomitant reduction of the rules and variables used ($n=6$ and 8, respectively). Importantly, the EmbryoMLSelection framework developed here was registered in the Docker image and therefore its application is globally accessible online. Of note, the rules associated with the ability of reaching the stage of expanded blastocyst on day 5 include early embryo-related variables, such as embryo morphological score on day 2, and some cytokinesis times occurring in the first three days of development (tPNf, t4, t4-t3, t8-t4). So far, only one study coupled TLT annotation with morphological embryo

assessment performed with the evidence-based score named IMCS [46]. According to our results, good quality embryos having static morphological score >6.0 on day 2 are more likely to reach the expanded blastocyst stage on day 5. In addition, the relevance of timings describing early embryo development is confirmed by previous studies reporting that a timely blastocyst development on day 5 can be predicted looking at the first three days of development [28, 47]. Moreover, morphokinetic data of cleavage stage embryos were found to be associated to both blastulation rate and blastocyst quality [48]. Indeed, embryos with quicker cleavage time from the 2-cells to the 8-cells stage have the highest potential to timely become blastocyst with good morphological score, and with the ability to expand and implant [49, 50]. In this context, the pivotal clinical significance of our framework would be to indicate on day 3 which embryos are more likely to develop into viable blastocysts, giving the potential advantage to select the most competent embryos on day 3 without the need to extend culture till day 5, thus saving time and resources.

Moreover, other clinical variables, such as age, AFC and OSI, were associated to the timely progression to the blastocysts stage.

Indeed, female age defined as advanced (AMA; >35 year) was extensively associated with a decline in oocyte yield, fertilization, and overall oocyte/embryo developmental competence, mainly due to an increased incidence of aneuploidies and a decreased mitochondrial activity [51, 52]. Studies reporting embryo morphokinetics from the fertilization to the pre-implantation period in women of AMA remain limited; in our dataset, 44 (55%) of cycles were performed in patients with AMA and a total of 303 (53%) embryos were included in our analysis suggesting a link between morphokinetic pattern and maternal age. Maternal age seems to have a relevant impact on the regulation of cell polarity during compaction, as well as on blastocoel cavity expansion, suggesting that AMA may affect embryo competence irrespective of the well-known consequences of oocyte meiotic errors [53]. On the other hand, AFC and OSI are markers of ovarian reserve and responsiveness to COS, and are associated not only with female age, but also with circulating AMH levels, oocyte yield and, ultimately, clinical pregnancy [54, 55].

Interestingly enough, the insemination technique (conventional IVF or ICSI) was not included as relevant variable in the selected rules, confirming previous evidence showing only minor morphokinetic differences between the two procedures [56].

Importantly, the rules proposed in this study are presented as a signature rather than a machine learning model. Despite the AdaBoosting algorithm had the highest AUC value in discriminating between BL and nBL and provided a model comprising rule-to-weight associations

which influence each rule's contribution to the final prediction, our focus was primarily to identify the rules rather than their weight.

The present study has the following limitations: (i) only couples undergoing single blastocyst transfer were considered in this study; (ii) the overall number of considered embryos was limited ($n=575$) but it constituted the entire time-lapse database available in our centre; (iii) embryo developmental timings were manually annotated, with unavoidable intra-operator variability [57]; (iv) only blastocyst expansion was considered irrespectively of ICM and TE morphology, but a timely blastocyst formation has a limited association coefficient with embryo ploidy and implantation chance [58]. As a consequence, we recognise that our analysis was performed on a restricted patient population with a limited sample size, thus making the current findings less generalizable in different clinical and laboratory settings.

Conclusions

To summarize, we identified specific rules composed of a combination of demographic, morphological and morphokinetic variables, that can represent a *feature signature* significantly associated with the likelihood to progress to the expanded blastocyst stage on day 5. We also identified for the first time clear and clinically relevant cut-offs for each considered variable, providing new insights on the critical range of values affecting embryo developmental competence.

In our opinion, it is likely that AI will be soon widely used in IVF labs, as it may be of high value in Lab automatization [59, 60]. Further studies are needed to prospectively validate our framework in a clinical setting, aimed at assessing whether it may provide a more objective, accurate, rapid and standardized tool for the assessment of embryo development.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13048-024-01376-6>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4

Acknowledgements

None.

Author contributions

S.C., N.L., F.C. and L.B. conceived the study; S.C., and C.P. were responsible for the time-lapse data acquisition and analysis; S.C., N.L. and F.C. performed the data analysis; S.C., N.L., F.C., L.B. and A.R. contributed to the validation, the interpretation of the results, the writing and editing of the manuscript; G.G., M.B., D.C., G.C., L.R. and C.B. contributed to the final interpretation of the data and editing of the manuscript. All authors gave their final approval.

Funding

University of Turin.

Data availability

The study data are available on reasonable request.

Declarations

Ethics approval and consent to participate

The study was carried out in accordance with the Declaration of Helsinki and was authorized as a retrospective observational study by the local Ethical Committee (authorization number: 0056908). A signed informed consent was obtained from all patients.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Gynecology and Obstetrics 1U, Physiopathology of Reproduction and IVF Unit, Department of Surgical Sciences, S. Anna Hospital, University of Turin, Turin, Italy

²IVIRMA Global Research Alliance, Livet, Turin, Italy

³Department of Computer Science, University di Turin, Turin, Italy

⁴Department of Oncology, University of Turin, Turin, Italy

⁵IVIRMA Global Research Alliance, Genera, Clinica Valle Giulia, Rome, Italy

⁶IVIRMA Global Research Alliance, 9.Baby, Bologna, Italy

⁷Department of Biomolecular Sciences, University of Urbino "Carlo Bo", Urbino, Italy

⁸Gynecology and Obstetrics 2U, Department of Surgical Sciences, S. Anna Hospital, University of Turin, Turin, Italy

Received: 25 August 2023 / Accepted: 16 February 2024

Published online: 15 March 2024

References

1. Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology. The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. *Hum Reprod Oxf Engl*. 2011;26:1270–83.
2. Storr A, Venetis CA, Cooke S, Kilani S, Ledger W. Inter-observer and intra-observer agreement between embryologists during selection of a single day 5 embryo for transfer: a multicenter study. *Hum Reprod Oxf Engl*. 2017;32:307–14.
3. Wong CC, Loewke KE, Bossert NL, Behr B, De Jonge CJ, Baer TM, et al. Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nat Biotechnol*. 2010;28:1115–21.
4. ESHRE Working group on Time-lapse technology, Apter S, Ebner T, Freour T, Guns Y, Kovacic B, et al. Good practice recommendations for the use of time-lapse technology†. *Hum Reprod Open*. 2020;2020:h0aa008.
5. Pribenszky C, Nilselid A-M, Montag M. Time-lapse culture with morphokinetic embryo selection improves pregnancy and live birth chances and reduces early pregnancy loss: a meta-analysis. *Reprod Biomed Online*. 2017;35:511–20.
6. Meseguer M, Herrero J, Tejera A, Hilligsøe KM, Ramsing NB, Remohí J. The use of morphokinetics as a predictor of embryo implantation. *Hum Reprod Oxf Engl*. 2011;26:2658–71.
7. Polanski LT, Coelho Neto MA, Nastri CO, Navarro PA, Ferriani RA, Raine-Fenning N, et al. Time-lapse embryo imaging for improving reproductive outcomes: systematic review and meta-analysis. *Ultrasound Obstet Gynecol*. 2014;44:394–401.
8. Armstrong S, Arroll N, Cree LM, Jordan V, Farquhar C. Time-lapse systems for embryo incubation and assessment in assisted reproduction. *Cochrane Database Syst Rev*. 2015;CD011320.
9. Racowsky C, Kovacs P, Martins WP. A critical appraisal of time-lapse imaging for embryo selection: where are we and where do we need to go? *J Assist Reprod Genet*. 2015;32:1025–30.

10. Chen M, Wei S, Hu J, Yuan J, Liu F. Does time-lapse imaging have favorable results for embryo incubation and selection compared with conventional methods in clinical in vitro fertilization? A meta-analysis and systematic review of randomized controlled trials. *PLoS ONE*. 2017;12:e0178720.
11. Swain JE. Controversies in ART: considerations and risks for uninterrupted embryo culture. *Reprod Biomed Online*. 2019;39:19–26.
12. Zaninovic N, Irani M, Meseguer M. Assessment of embryo morphology and developmental dynamics by time-lapse microscopy: is there a relation to implantation and ploidy? *Fertil Steril*. 2017;108:722–9.
13. Wang R, Pan W, Jin L, Li Y, Geng Y, Gao C, et al. Artificial intelligence in reproductive medicine. *Reprod Camb Engl*. 2019;158:R139–54.
14. Zaninovic N, Rosenwaks Z. Artificial intelligence in human in vitro fertilization and embryology. *Fertil Steril*. 2020;114:914–20.
15. Manna C, Nanni L, Lumini A, Pappalardo S. Artificial intelligence techniques for embryo and oocyte classification. *Reprod Biomed Online*. 2013;26:42–9.
16. Dimitriadis I, Zaninovic N, Badiola AC, Bormann CL. Artificial intelligence in the embryology laboratory: a review. *Reprod Biomed Online*. 2022;44:435–48.
17. Raef B, Ferdousi R. A review of machine learning approaches in assisted reproductive technologies. *Acta Inf Med AIM J Soc Med Inf Bosnia Herzeg Cas Drustva Za Med Inf BiH*. 2019;27:205–11.
18. Giscard d'Estaing S, Labrune E, Forcellini M, Edel C, Salle B, Lornage J, et al. A machine learning system with reinforcement capacity for predicting the fate of an ART embryo. *Syst Biol Reprod Med*. 2021;67:64–78.
19. Revelli A, Rovei V, Dalmasso P, Gennarelli G, Racca C, Evangelista F, et al. Large randomized trial comparing transabdominal ultrasound-guided embryo transfer with a technique based on uterine length measurement before embryo transfer. *Ultrasound Obstet Gynecol off J Int Soc Ultrasound Obstet Gynecol*. 2016;48:289–95.
20. Huber M, Hadziosmanovic N, Berglund L, Holte J. Using the ovarian sensitivity index to define poor, normal, and high response after controlled ovarian hyperstimulation in the long gonadotropin-releasing hormone-agonist protocol: suggestions for a new principle to solve an old problem. *Fertil Steril*. 2013;100:1270–6.
21. Revelli A, Gennarelli G, Biasoni V, Chiadò A, Carosso A, Evangelista F, Paschero C, Filippini C, Benedetto C. The ovarian sensitivity Index (OSI) significantly correlates with Ovarian Reserve biomarkers, is more predictive of clinical pregnancy than the total number of oocytes, and is consistent in consecutive IVF cycles. *J Clin Med*. 2020;9(6):1914. <https://doi.org/10.3390/jcm9061914>.
22. World Health Organization. WHO laboratory manual for the examination and processing of human semen [Internet]. World Health Organization.; 2010. Available from: <https://apps.who.int/iris/handle/10665/44261>.
23. Canosa S, Adriaenssens T, Coucke W, Dalmasso P, Revelli A, Benedetto C, et al. Zona pellucida gene mRNA expression in human oocytes is related to oocyte maturity, zona inner layer retardance and fertilization competence. *Mol Hum Reprod*. 2017;23:292–303.
24. Canosa S, Paschero C, Carosso A, Leoncini S, Mercaldo N, Gennarelli G, et al. Effect of a combination of Myo-Inositol, Alpha-Lipoic Acid, and folic acid on oocyte morphology and embryo morphokinetics in non-PCOS Overweight/Obese patients undergoing IVF: a pilot, prospective, randomized study. *J Clin Med*. 2020;9:E2949.
25. Holte J, Berglund L, Milton K, Garelo C, Gennarelli G, Revelli A, et al. Construction of an evidence-based integrated morphology cleavage embryo score for implantation potential of embryos scored and transferred on day 2 after oocyte retrieval. *Hum Reprod Oxf Engl*. 2007;22:548–57.
26. Vaegter KK, Lakic TG, Olovsson M, Berglund L, Brodin T, Holte J. Which factors are most predictive for live birth after in vitro fertilization and intracytoplasmic sperm injection (IVF/ICSI) treatments? Analysis of 100 prospectively recorded variables in 8,400 IVF/ICSI single-embryo transfers. *Fertil Steril*. 2017;107:641–648e2.
27. Rhenman A, Berglund L, Brodin T, Olovsson M, Milton K, Hadziosmanovic N, Holte J. Which set of embryo variables is most predictive for live birth? A prospective study in 6252 single embryo transfers to construct an embryo score for the ranking and selection of embryos. *Hum Reprod*. 2015;30(1):28–36.
28. Canosa S, Bergandi L, Macri C, Charrier L, Paschero C, Carosso A, et al. Morphokinetic analysis of cleavage stage embryos and assessment of specific gene expression in cumulus cells independently predict human embryo development to expanded blastocyst: a preliminary study. *J Assist Reprod Genet*. 2020;37:1409–20.
29. Hammond ER, Foong AKM, Rosli N, Morbeck DE. Should we freeze it? Agreement on fate of borderline blastocysts is poor and does not improve with a modified blastocyst grading system. *Hum Reprod*. 2020;35:1045–53.
30. Cimadomo D, Sosa Fernandez L, Soscia D, Fabozzi G, Benini F, Cesana A, Dal Canto MB, Maggiulli R, Muzzi S, Scarica C, et al. Inter-centre reliability in embryo grading across several IVF clinics is limited: implications for embryo selection. *Reprod Biomed Online*. 2021;44:39–48.
31. Chiappetta V, Innocenti F, Coticchio G, Ahlström A, Albricci L, Badajoz V, Hebles M, Gallardo M, Benini F, Canosa S, Kumpošt J, Milton K, Montanino Oliva D, Maggiulli R, Rienzi L, Cimadomo D. Discard or not discard, that is the question: an international survey across 117 embryologists on the clinical management of borderline quality blastocysts. *Hum Reprod*. 2023;38(10):1901–9.
32. Van Rossum G, Drake FL. Python 3 reference Manual. Scotts Valley, CA: CreateSpace; 2009.
33. Ciray HN, Campbell A, Agerholm IE, Aguilar J, Chamayou S, Esbert M, et al. Proposed guidelines on the nomenclature and annotation of dynamic human embryo monitoring by a time-lapse user group. *Hum Reprod Oxf Engl*. 2014;29:2650–60.
34. Shapiro BS, Richter KS, Harris DC, Daneshmand ST. A comparison of day 5 and day 6 blastocyst transfers. *Fertil Steril*. 2001;75:1126–30.
35. Barrenetxea G, López de Larruzea A, Ganzabal T, Jiménez R, Carbonero K, Mandiola M. Blastocyst culture after repeated failure of cleavage-stage embryo transfers: a comparison of day 5 and day 6 transfers. *Fertil Steril*. 2005;83:49–53.
36. Fransiak JM, Forman EJ, Patounakis G, Hong KH, Werner MD, Upham KM, et al. Investigating the impact of the timing of blastulation on implantation: management of embryo-endometrial synchrony improves outcomes. *Hum Reprod Open*. 2018;2018:hoy022.
37. Hashimoto S, Amo A, Hama S, Ito K, Nakaoka Y, Morimoto Y. Growth retardation in human blastocysts increases the incidence of abnormal spindles and decreases implantation potential after vitrification. *Hum Reprod*. 2013;28(6):1528–35.
38. Yi HJ, Koo HS, Cha SH, Kim HO, Park CW, Song IO. Reproductive outcomes of retransferring retained embryos in blastocyst transfer cycles. *Clin Exp Reprod Med*. 2016;43(2):133–8.
39. Liu X, Lou H, Zhang J, Du M, Du Y, Wu S, Guan Y, Liu J. Clinical outcome analysis of frozen-thawed embryo transfer on Day 7. *Front Endocrinol (Lausanne)*. 2022;13:1082597.
40. ESHRE Special Interest Group of Embryology. Alpha scientists in Reproductive Medicine. The Vienna consensus: report of an expert meeting on the development of art laboratory performance indicators. *Hum Reprod Open*. 2017;2017(2):hox011.
41. Conaghan J, Chen AA, Willman SP, Ivani K, Chenette PE, Boostanfar R, et al. Improving embryo selection using a computer-automated time-lapse image analysis test plus day 3 morphology: results from a prospective multicenter trial. *Fertil Steril*. 2013;100:412–419e5.
42. Motato Y, de los Santos MJ, Escriba MJ, Ruiz BA, Remohí J, Meseguer M. Morphokinetic analysis and embryonic prediction for blastocyst formation through an integrated time-lapse system. *Fertil Steril*. 2016;105:376–384e9.
43. Yang L, Peavey M, Kaskar K, Chappell N, Zhu L, Devlin D, et al. Development of a dynamic machine learning algorithm to predict clinical pregnancy and live birth rate with embryo morphokinetics. *FS Rep*. 2022;3:116–23.
44. VerMilyea M, Hall JMM, Diakiv SM, Johnston A, Nguyen T, Perugini D, et al. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. *Hum Reprod Oxf Engl*. 2020;35:770–84.
45. Barrie A, Homburg R, McDowell G, Brown J, Kingsland C, Troup S. Examining the efficacy of six published time-lapse imaging embryo selection algorithms to predict implantation to demonstrate the need for the development of specific, in-house morphokinetic selection algorithms. *Fertil Steril*. 2017;107(3):613–21.
46. Revelli A, Canosa S, Carosso A, Filippini C, Paschero C, Gennarelli G, et al. Impact of the addition of early embryo viability Assessment to morphological evaluation on the accuracy of embryo selection on day 3 or day 5: a retrospective analysis. *J Ovarian Res*. 2019;12:73.
47. Scarica C, Cimadomo D, Dovere L, Giancani A, Stoppa M, Capalbo A, Ubaldi FM, Rienzi L, Canipari RJ. An integrated investigation of oocyte developmental competence: expression of key genes in human cumulus cells, morphokinetics of early divisions, blastulation, and euploidy. *Assist Reprod Genet*. 2019;36(5):875–87.
48. Kirkegaard K, Kesmodel US, Hindkjær JJ, Ingerslev HJ. Time-lapse parameters as predictors of blastocyst development and pregnancy outcome in embryos from good prognosis patients: a prospective cohort study. *Hum Reprod Oxf Engl*. 2013;28:2643–51.

49. Dal Canto M, Coticchio G, Mignini Renzini M, De Ponti E, Novara PV, Brambillasca F, et al. Cleavage kinetics analysis of human embryos predicts development to blastocyst and implantation. *Reprod Biomed Online*. 2012;25:474–80.
50. Coticchio G, Ezoe K, Lagalla C, Zacà C, Borini A, Kato K. The destinies of human embryos reaching blastocyst stage between day 4 and Day 7 diverge as early as fertilization. *Hum Reprod*. 2023 Jun;28:dead136.
51. Cimadomo D, Fabozzi G, Vaiarelli A, Ubaldi N, Ubaldi FM, Rienzi L. Impact of maternal age on oocyte and embryo competence. *Front Endocrinol (Lausanne)*. 2018;9:327.
52. Ubaldi FM, Cimadomo D, Vaiarelli A, Fabozzi G, Venturella R, Maggiulli R, Mazzilli R, Ferrero S, Palagiano A, Rienzi L. Advanced maternal age in IVF: still a challenge? The Present and the future of its treatment. *Front Endocrinol (Lausanne)*. 2019;10:94.
53. Ezoe K, Miki T, Akaike H, Shimazaki K, Takahashi T, Tanimura Y, Amagai A, Sawado A, Mogi M, Kaneko S, Ueno S, Coticchio G, Cimadomo D, Borini A, Rienzi L, Kato K. Maternal age affects pronuclear and chromatin dynamics, morula compaction and cell polarity, and blastulation of human embryos. *Hum Reprod*. 2023;38(3):387–99.
54. Biasoni V, Patriarca A, Dalmaso P, Bertagna A, Manieri C, Benedetto C, et al. Ovarian sensitivity index is strongly related to circulating AMH and may be used to predict ovarian response to exogenous gonadotropins in IVF. *Reprod Biol Endocrinol RBE*. 2011;9:112.
55. Carosso AR, van Eekelen R, Revelli A, Canosa S, Mercaldo N, Benedetto C, et al. Women in Advanced Reproductive Age: are the follicular output rate, the follicle-oocyte index and the ovarian sensitivity index predictors of live birth in an IVF cycle? *J Clin Med*. 2022;11:859.
56. De Munck N, Bayram A, Elkhatib I, Abdala A, El-Damen A, Arnanz A, et al. Marginal differences in preimplantation morphokinetics between conventional IVF and ICSI in patients with preimplantation genetic testing for aneuploidy (PGT-A): a sibling oocyte study. *PLoS ONE*. 2022;17:e0267241.
57. Sundvall L, Ingerslev HJ, Breth Knudsen U, Kirkegaard K. Inter- and intra-observer variability of time-lapse annotations. *Hum Reprod Oxf Engl*. 2013;28:3215–21.
58. Cimadomo D, Rienzi L, Conforti A, Forman E, Canosa S, Innocenti F, Poli M, Hynes J, Gemmell L, Vaiarelli A, Alviggi C, Ubaldi FM, Capalbo A. Opening the black box: why do euploid blastocysts fail to implant? A systematic review and meta-analysis. *Hum Reprod Update*. 2023 May;16:dmad010.
59. Glatstein I, Chavez-Badiola A, Curchoe CL. New frontiers in embryo selection. *J Assist Reprod Genet*. 2023;40(2):223–34.
60. Cimadomo D, Chiappetta V, Innocenti F, Saturno G, Taggi M, Marconetto A, Casciani V, Albricci L, Maggiulli R, Coticchio G, Ahlström A, Berntsen J, Larman M, Borini A, Vaiarelli A, Ubaldi FM, Rienzi L. Towards automation in IVF: pre-clinical validation of a deep learning-based embryo grading system during PGT-A cycles. *J Clin Med*. 2023;12(5):1806.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.