




Article

# Towards Automation in IVF: Pre-Clinical Validation of a Deep Learning-Based Embryo Grading System during PGT-A Cycles

Danilo Cimadomo <sup>1,\*</sup> , Viviana Chiappetta <sup>1</sup>, Federica Innocenti <sup>1</sup>, Gaia Saturno <sup>2</sup>, Marilena Taggi <sup>2</sup>, Anabella Marconetto <sup>3</sup>, Valentina Casciani <sup>1</sup>, Laura Albricci <sup>1</sup>, Roberta Maggiulli <sup>1</sup>, Giovanni Coticchio <sup>4</sup> , Aisling Ahlström <sup>5</sup>, Jørgen Berntsen <sup>6</sup> , Mark Larman <sup>7</sup>, Andrea Borini <sup>4</sup>, Alberto Vaiarelli <sup>1</sup>, Filippo Maria Ubaldi <sup>1</sup> and Laura Rienzi <sup>1,8</sup>

<sup>1</sup> Clinica Valle Giulia, GeneraLife IVF, Via De Notaris 2B, 00197 Rome, Italy

<sup>2</sup> Department of Biology and Biotechnology “Lazzaro Spallanzani”, University of Pavia, 27100 Pavia, Italy

<sup>3</sup> University Institute of Reproductive Medicine, National University of Cordoba, Cordoba 5187, Argentina

<sup>4</sup> 9.baby, GeneraLife IVF, 40125 Bologna, Italy

<sup>5</sup> Livio, GeneraLife IVF, 40229 Göteborg, Sweden

<sup>6</sup> Vitrolife A/S, 8260 Aarhus, Denmark

<sup>7</sup> Vitrolife Sweden AB, 421 32 Göteborg, Sweden

<sup>8</sup> Department of Biomolecular Sciences, University of Urbino “Carlo Bo”, 61029 Urbino, Italy

\* Correspondence: cimadomo@generalifeitalia.it or danilo.cimadomo@generalife.com

**Abstract:** Preimplantation genetic testing for aneuploidies (PGT-A) is arguably the most effective embryo selection strategy. Nevertheless, it requires greater workload, costs, and expertise. Therefore, a quest towards user-friendly, non-invasive strategies is ongoing. Although insufficient to replace PGT-A, embryo morphological evaluation is significantly associated with embryonic competence, but scarcely reproducible. Recently, artificial intelligence-powered analyses have been proposed to objectify and automate image evaluations. iDAScore v1.0 is a deep-learning model based on a 3D convolutional neural network trained on time-lapse videos from implanted and non-implanted blastocysts. It is a decision support system for ranking blastocysts without manual input. This retrospective, pre-clinical, external validation included 3604 blastocysts and 808 euploid transfers from 1232 cycles. All blastocysts were retrospectively assessed through the iDAScore v1.0; therefore, it did not influence embryologists’ decision-making process. iDAScore v1.0 was significantly associated with embryo morphology and competence, although AUCs for euploidy and live-birth prediction were 0.60 and 0.66, respectively, which is rather comparable to embryologists’ performance. Nevertheless, iDAScore v1.0 is objective and reproducible, while embryologists’ evaluations are not. In a retrospective simulation, iDAScore v1.0 would have ranked euploid blastocysts as top quality in 63% of cases with one or more euploid and aneuploid blastocysts, and it would have questioned embryologists’ ranking in 48% of cases with two or more euploid blastocysts and one or more live birth. Therefore, iDAScore v1.0 may objectify embryologists’ evaluations, but randomized controlled trials are required to assess its clinical value.

**Keywords:** IVF; PGT-A; artificial intelligence; automation; time-lapse microscopy; embryo selection; blastocyst; embryo assessment; euploidy; implantation



**Citation:** Cimadomo, D.; Chiappetta, V.; Innocenti, F.; Saturno, G.; Taggi, M.; Marconetto, A.; Casciani, V.; Albricci, L.; Maggiulli, R.; Coticchio, G.; et al. Towards Automation in IVF: Pre-Clinical Validation of a Deep Learning-Based Embryo Grading System during PGT-A Cycles. *J. Clin. Med.* **2023**, *12*, 1806. <https://doi.org/10.3390/jcm12051806>

Academic Editor: Francesco Fusi

Received: 22 December 2022

Revised: 13 February 2023

Accepted: 21 February 2023

Published: 23 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Embryo assessment and selection continues to be a major challenge in IVF, especially since IVF clinics started more commonly adopting a single embryo transfer (SET) policy [1]. In fact, embryologists worldwide strive to implement effective strategies to improve IVF efficiency (i.e., higher live birth rate (LBR) per transfer with less risks, efforts, and possibly costs) while preserving its efficacy (i.e., the cumulative live birth delivery rate (CLBdR) per cycle) [2]. Static embryo morphological assessment is still the predominant non-invasive embryo selection strategy used. It consists of several static microscopic observations at

fixed time points of preimplantation development focused on a few prognostic features [3]. At the blastocyst stage, the Gardner's score is the most applied grading system. It is a three-part scoring system based on the degree of blastocyst expansion, inner cell mass (ICM), and trophoctoderm (TE) morphology [4]. Blastocyst transfer elicits higher LBR per ET than cleavage stage ET with the same CLBdR per cycle and miscarriage rate per clinical pregnancy [5]. In addition, a significant correlation exists between blastocyst quality with euploidy and implantation in the context of both untested and euploid ET [6–8]. Regardless, static assessment suffers from several inherent limitations. Firstly, it is undermined by high subjectivity and both intra- and inter-operator variability [9–11]. Moreover, a few snapshots of embryo development cannot provide a complete evaluation of this complex and dynamic process and fail to capture abnormal events, such as abnormal fertilization and cleavage patterns, blastomere exclusion/extrusion, or spontaneous blastocyst collapse [12]. In fact, blastocysts classified as excellent/good quality are often aneuploid or fail to implant, just like blastocysts classified as poor quality (less than Gardner's BB grade) may actually be euploid and implant [13–17].

Implementation of preimplantation genetic testing for aneuploidies (PGT-A) allows the discrimination of chromosomally normal (euploid) from abnormal (aneuploid) embryos in a cohort of blastocysts produced during IVF through biopsy and analysis via comprehensive chromosome testing (CCT) technologies (e.g., q-PCR or NGS) of 5–10 cells from the TE [18]. In the hands of expert operators and well-equipped laboratories, TE biopsy does not negatively impact embryo viability [19,20]. Crucially, the transfer of euploid blastocysts in randomized controlled (RCT) or observational trials involves higher LBR per ET and lower miscarriage rate per clinical pregnancy with respect to untested blastocyst transfer [21,22]. In fact, when blastocysts affected by full-chromosome meiotic aneuploidies were transferred in blinded non-selection studies, they resulted in >98% embryo lethality rate with almost 90% of clinical pregnancies ending up in a miscarriage. Still, this technique requires extensive expertise and several euploid blastocysts fail to result in a live birth (LB), despite a predictive value on implantation as high as 65% [20].

More recently, the introduction of time-lapse technology (TLT) in IVF has allowed continuous monitoring of embryos, undisturbed culture, and precise reporting of developmental timings and abnormal cleavage patterns [12,23,24]. Nevertheless, the data about the true effectiveness of TLT for embryo selection purposes are controversial [25,26]. In fact, embryo morphodynamics is associated with, but cannot effectively predict, euploidy [27,28].

The latest development in this scenario is the combination of artificial intelligence (AI) with TLT. AI leverages computers and machines to mimic human problem-solving and decision-making capabilities. The definition of AI includes machine learning (ML) and deep learning (DL). ML is a data processing technology that can make predictions based on previously analyzed, structured, or labeled data. DL works as a set of neural networks, inspired by the human brain, to learn and detect features from large amounts of unlabeled data. The use of algorithms to guide human decisions would contribute greatly to achieving standardization in IVF, and thus, obtaining more consistent, comparable, and reproducible results, by preventing subjectivity in the evaluation process [29–31]. To that end, several systems have been developed lately that can assess individual embryos, segmenting and grading important developmental features, and generating a score (e.g., [32–34]). In theory, the AI-powered TLT assessment is a goldmine of information potentially useful for embryo selection purposes [23,27,30,35–38], but its clinical utility must be tested in properly designed studies and/or large real-life datasets. The software “intelligent data analysis (iDA) Score v1.0” (Vitrolife A/S), which can be directly integrated into an EmbryoScope+ incubator (Vitrolife A/S), is one of these tools. This software (a DL algorithm trained on hundreds of thousands of videos from implanted and non-implanted embryos) generates a score for each embryo from 1.0 to 9.9, which should be representative of its chance to implant.

We designed this study to assess the degree of concordance between iDAScore v1.0 with: (i) blastocyst morphological assessment carried out by senior embryologists according

to Gardner's criteria, and (ii) blastocysts' chromosomal constitution (euploidy, single aneuploidy, multiple aneuploidies). We also assessed how often iDAScore v1.0 would have ranked as top-quality euploid blastocysts among cohorts characterized by sibling euploid and aneuploid embryos. Lastly, we assessed in a retrospective simulation how often ranking for transfer of multiple euploid blastocysts by iDAScore v1.0 would have involved an earlier or later LB.

## 2. Material and Methods

### 2.1. Study Design

This is a retrospective study aimed at a pre-clinical validation of iDAScore v1.0 software in PGT-A cycles conducted between April 2013 and August 2022 at a private IVF clinic (Clinica Valle Giulia, GeneraLife IVF, Rome, Italy). Overall, we included 1232 PGT-A cycles with  $\geq 1$  biopsied blastocyst ( $N = 3604$  embryos) after undisturbed culture in EmbryoScope incubators (Vitrolife A/S, Aarhus, Denmark). All patients were included only once for their first PGT-A cycle conducted with their own fresh oocytes. Patients with an indication for PGT for structural rearrangements (PGT-SR) and PGT for monogenic conditions (PGT-M) were not included.

All videos were retrospectively assessed through the iDAScore v1.0 software to grade each blastocyst without influencing embryologists' clinical evaluations and decision-making process. iDAScore v1.0 was then assessed for its concordance with blastocyst quality as defined by senior embryologists according to the Gardner's criteria (ICM morphology, TE morphology, and overall blastocyst morphology), and day of biopsy defined according to the hpi ( $\leq 120$  hpi = day 5, 121–144 = day 6,  $>144$  = day 7) (Figure 1). Overall, 771 patients obtained  $\geq 1$  euploid blastocyst ( $N = 1443$  embryos) after TE biopsy-based chromosomal testing conducted via CCT technologies at an external genetic laboratory (Igenomix, Marostica, Italy). iDAScore v1.0 was also assessed for its association with blastocyst karyotype (Figure 1) categorized as euploid, single aneuploid (i.e., single monosomy or single trisomy), or complex aneuploid (i.e.,  $\geq 2$  aneuploidies). A sub-analysis of iDAScore v1.0 in these three groups was conducted within blastocyst quality categories as defined by the senior embryologists, as well as within day of biopsy categories according to the hpi. Finally, a receiver operating characteristic (ROC) curve analysis was conducted to assess the area under the curve (AUC) for the discrimination of euploidy based on the embryologists' assessment and based on iDAScore v1.0. Overall, 610 patients conducted  $\geq 1$  vitrified-warmed euploid blastocyst SET by the time of paper drafting ( $N = 808$  SETs). iDAScore v1.0 was finally assessed for its association with the outcome after euploid SET (i.e., either no LB or LB) (Figure 1). A sub-analysis of iDAScore v1.0 in these two groups was conducted within blastocyst quality categories as defined by the senior embryologists, as well as within day of biopsy categories according to the hpi. At last, a ROC curve analysis was conducted to assess the AUC for the discrimination of a LB after euploid SET based on embryologists' assessment and based on iDAScore v1.0.

Beyond the association studies, we also evaluated what would have been the impact of iDAScore v1.0 had it been used clinically to prioritize the blastocyst for transfer, either without or with the diagnostic information derived from aneuploidy testing. The first simulation was conducted in the 587 cycles ( $N = 587/1232$ , 47.6% of the cycles included) where sibling  $\geq 1$  euploid and  $\geq 1$  aneuploid blastocysts were diagnosed. Specifically, we calculated how often the embryologists and iDAScore v1.0 would have blindly graded a euploid or aneuploid blastocyst as top quality within each cohort (Figure 2A). The second simulation was carried out in 202 cycles conducted up to December 2021 ( $N = 202/1057$ , 19% of the cycles included) that, before we drafted this manuscript, had already achieved  $\geq 1$  LB from a cohort of  $\geq 2$  euploid blastocysts. Specifically, we calculated: (i) how often the embryologists and iDAScore v1.0 would have been equally effective in prioritizing the euploid blastocyst to transfer that, indeed, resulted in a LB (that is, embryologists and iDAScore v1.0 would have been equally effective), (ii) how often iDAScore v1.0 would have selected a euploid blastocyst for ET that resulted in a LB, but the embryologists transferred

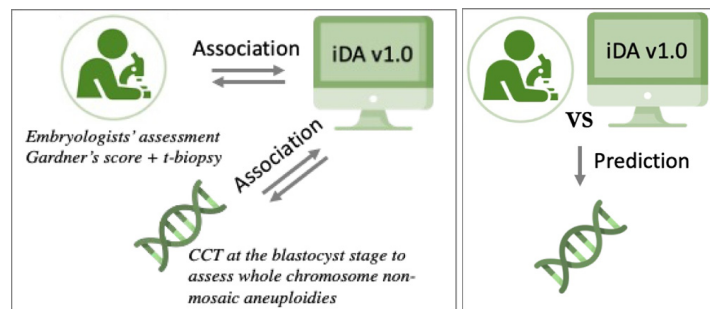
this embryo only after another one that did not result in a LB (that is, iDAScore v1.0 would have involved an earlier LB), and (iii) how often iDAScore v1.0 would have selected a euploid blastocyst that did not result in a LB and that was transferred by the embryologists only after another embryo which instead did result in a LB (that is, iDAScore v1.0 would have involved a later LB) (Figure 2B).

Period: April 2013 – August 2022

1232 cycles with  $\geq 1$  biopsied blastocyst after embryo culture  
in EmbryoScope incubators

- 330 cycles with 1 blastocyst
- 283 cycles with 2 blastocysts
- 207 cycles with 3 blastocysts
- 186 cycles with 4 blastocysts
- 108 cycles with 5 blastocysts
- 73 cycles with 6 blastocysts
- 17 cycles with 7 blastocysts
- 13 cycles with 8 blastocysts
- 10 cycles with 9 blastocysts
- 3 cycles with 10 blastocysts
- 2 cycles with 11 blastocysts

3604  
Blastocysts



771 cycles with  $\geq 1$  euploid blastocyst

- 394 cycles with 1 euploid blastocyst
- 208 cycles with 2 euploid blastocysts
- 97 cycles with 3 euploid blastocysts
- 40 cycles with 4 euploid blastocysts
- 19 cycles with 5 euploid blastocysts
- 8 cycles with 6 euploid blastocysts
- 4 cycles with 7 euploid blastocysts
- 1 cycle with 11 euploid blastocysts

1443  
Euploid blastocysts

610 cycles with  $\geq 1$  vitrified-warmed euploid SET

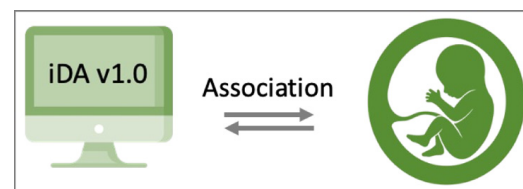
- 453 cycles with 1 euploid SET
- 122 cycles with 2 euploid SETs
- 29 cycles with 3 euploid SETs
- 6 cycles with 4 euploid SETs

808  
Euploid SETs

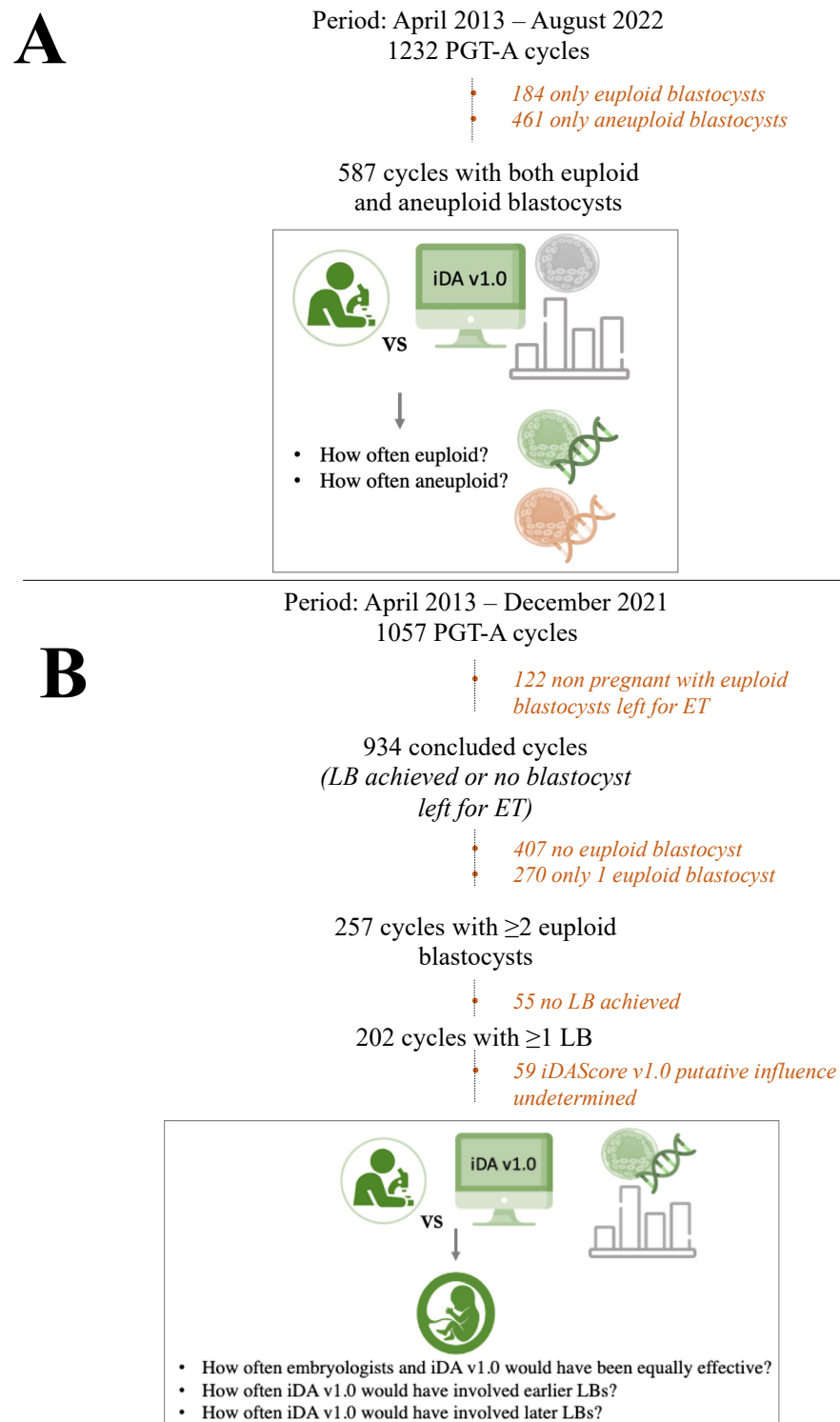
338 cycles with  $\geq 1$  LB

- 315 cycles with 1 LB
- 23 cycles with 2 LBs

361  
LBs



**Figure 1.** Association and prediction study workflow. T-biopsy, time of biopsy; CCT, comprehensive chromosome testing; iDA v1.0, Intelligent Data Analysis score version 1.0; SETs, single embryo transfers; LBs, live births.



**Figure 2.** Clinical utility study workflow. **(A)** Definition of top-quality blastocysts within each cohort according to the embryologists versus iDAScore v1.0: how often were they euploid and how often aneuploid? **(B)** Definition of top-quality euploid blastocysts within each cohort according to the embryologists versus iDAScore v1.0: how often would they have been equally effective? How often would iDAScore v1.0 have involved an earlier live birth (LB)? How often would iDAScore v1.0 involved a later LB? In both figures, the orange phrases summarize the excluded cycles with the reasons for exclusion. PGT-A, preimplantation genetic testing for aneuploidies; ET, embryo transfer.



## 2.2. IVF Protocols

Only the first IVF cycles in EmbryoScope incubators and PGT-A were included. Ovarian stimulation was conducted only with GnRH antagonist protocols and ovulation was triggered with either GnRH-agonist or hCG [39–41]. Oocytes were retrieved 35 h after trigger, and only ICSI was conducted as previously detailed [42]. Only continuous embryo culture was conducted in a continuous single culture medium (CSCM, Irvine Scientific, USA) with a refresh on day 5 in case of extended culture to day 6–7. Laser-assisted TE biopsy was conducted according to the “simultaneous zona pellucida (ZP) opening and biopsy method” [13,43]. This approach does not entail any ZP drilling at the cleavage stage (i.e., day 3 of preimplantation development), and the embryos are left undisturbed until their full expansion on day 5–7 [14]. Only CCT analyses were conducted [44–46] to identify non-mosaic full chromosome aneuploidies, and chromosome intermediate copy numbers (ICN) were reported as either euploid or aneuploid based on a 50% threshold according to the assessment of our reference genetic laboratory (Igenomix, Italy) [47,48]. Indeed, the report of putative mosaicism based on  $ICN < 50\%$  has been shown clinically ineffective in a recent blinded non-selection study [49]. Vitrification was conducted within 90 min from biopsy [43]. Only euploid blastocyst SETs were performed 2 h after warming in a following menstrual cycle. Endometrial preparation was managed with either a modified-natural cycle or through hormone replacement therapy [41]. All SETs from the same oocyte retrieval cycle were included.

Blastocyst morphology was graded by senior embryologists based on the Gardner’s scoring system [4]. Specifically, the ICM was graded “A” in case of a structure characterized by several strictly packed cells, “B” in case of a discernible structure with several but roughly packed cells, or “C” in case of a structure difficult to distinguish with few low-quality cells. Similarly, the TE was graded “A” in case of a well-organized cohesive epithelium with several cells, “B” in case of a loose epithelium with few cells, or “C” in case of very few and/or low-quality cells. Each blastocyst was graded in real time by two senior embryologists (Fleiss’ Kappa for ICM morphology assessment = 0.610, i.e., good agreement; Fleiss’ Kappa for TE morphology assessment = 0.806, i.e., excellent agreement). In case of disagreement, a third senior embryologist decided the grade. Our internal grading scheme clusters all “AA” blastocysts within the “excellent” quality category, “AB” and “BA” blastocysts within the “good” quality category, “BB”, “AC”, and “CA” within the “average” quality category, and “CC”, “BC”, and “CB” within the “poor” quality category [13]. Whenever  $\geq 2$  blastocysts were obtained in a cohort, the senior embryologists would identify the top-quality embryo based on: (i) its overall quality, (ii) the time of biopsy in hpi (the earlier, the better), (iii) the TE quality, and (iv) the expansion (the larger, the better).

## 2.3. iDAScore v1.0

The DL model iDAScore v1.0 is based on a 3D convolutional neural network [31,50]. The model was trained on a large data set from 18 clinics worldwide containing a total of 115,832 embryos. Of them, 14,644 embryos were transferred on day 5 or later, resulting in 4337 positive fetal heartbeats and 10,307 implantation failures. The input to the model is 128 images sampled one hour apart covering the time from 12 hpi to 140 hpi. No patient data (e.g., age) or morphokinetic parameters are used as input to the model. The model is intended to be used on all embryos without any pre-selection and as a decision support system where the final decision is made by the user. The software, which is an add-on to the existing EmbryoSuite software (Vitrolife A/S), generates for each embryo a score between 1.0 (lowest) and 9.9 (highest), which is meant to express its implantation potential. Clinica Valle Giulia (GeneralLife IVF) was not involved in training the model; therefore, this study should be considered an independent external pre-clinical validation.

## 2.4. Statistical Analysis

Continuous variables were reported as mean  $\pm$  SD and Shapiro–Wilk test was adopted to assess a Gaussian distribution of the data. Mann–Whitney U, Kruskal–Wallis, Student’s t-tests or ANOVA were adopted to define significant differences among each comparison. Fisher’s exact or chi-squared tests were instead adopted for categorical variables. Linear and logistic regressions were conducted to confirm significant associations. Putative confounders (relevant to patients, embryos, and cycle characteristics) were outlined through univariate analyses and eventually included to adjust the results in multivariate analyses. All statistical analyses were performed with the software SPSS (IBM, Armonk, NY, USA). Post-hoc statistical power analyses were conducted via G\*Power for all the main comparisons.

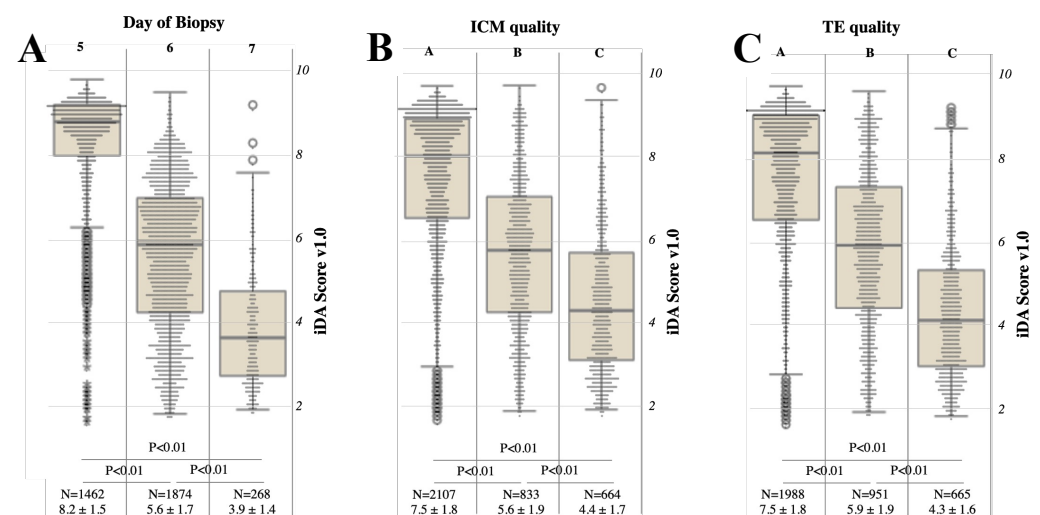
## 3. Results

### 3.1. The Patients Included Are Predominantly Poor Prognosis and of Advanced Maternal Age

The patients included in this study represent the average population of women undergoing IVF at our center, predominantly advanced maternal age ( $38.7 \pm 3.4$  years) and poor prognosis ( $2.9 \pm 1.8$  blastocysts biopsied, of which  $1.2 \pm 1.3$  euploids) (Supplementary Table S1).

### 3.2. A Generally Good Association Exists between the Conventional Parameters of Morphological Evaluation and iDAScore v1.0

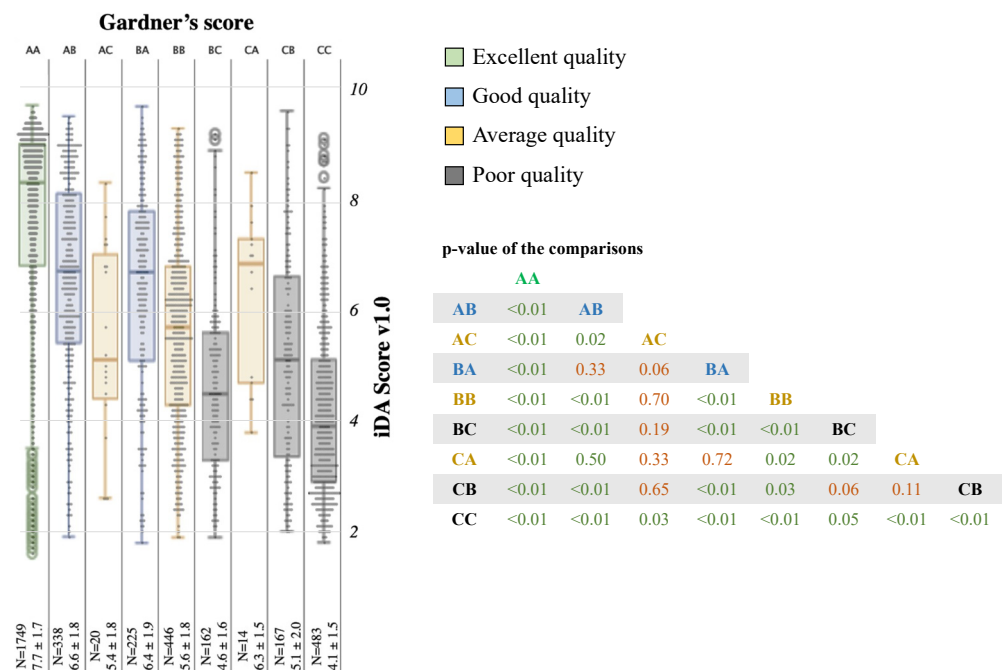
iDAScore v1.0 was significantly associated with the day of full blastocyst development (day 5 blastocysts,  $N = 1462$ ,  $8.2 \pm 1.5$  versus day 6 blastocysts,  $N = 1874$ ,  $5.6 \pm 1.7$  versus day 7 blastocysts,  $N = 268$ ,  $3.9 \pm 1.4$ ;  $p < 0.01$  and power = 99% for all comparisons; Figure 3A). The same data are presented in Supplementary Figure S1A as a dispersion plot that associates the time of biopsy of each embryo with its iDAScore v1.0. The linear regression analysis confirmed the significant association (unstandardized coefficient B:  $-0.092$ , 95% CI from  $-0.096$  to  $-0.089$ ,  $p < 0.01$ ). Nevertheless, although the mean and median values are significantly different across the groups, long tails were shown in both graphs around low iDAScore v1.0 values in the day 5 group, as much as around high iDAScore v1.0 in the day 7 group. Conversely, day 6 blastocysts show a more widespread distribution of the data.



**Figure 3.** iDAScore v1.0 is associated with the day of biopsy (A), the inner cell mass (ICM) quality (B), and the trophoctoderm (TE) quality (C). The day of biopsy is defined according to hours between insemination and achievement of a grade of blastocyst expansion compatible with a TE biopsy:  $\leq 120$  h post insemination (hpi) = day 5, 121–144 hpi = day 6,  $> 144$  hpi = day 7. ICM and TE quality were defined according to Gardner’s score as A, B, or C (from best to worst quality).

iDAScore v1.0 was also significantly associated with ICM quality (A grade,  $N = 2107$ ,  $7.5 \pm 1.8$  versus B grade,  $N = 833$ ,  $5.6 \pm 1.9$  versus C grade,  $N = 664$ ,  $4.4 \pm 1.7$ ;  $p < 0.01$  and power  $>99\%$  for all comparisons; Figure 3B). iDAScore v1.0 was significantly associated with TE quality as well (A grade,  $N = 1988$ ,  $7.5 \pm 1.8$  versus B grade,  $N = 951$ ,  $5.9 \pm 1.9$  versus C grade,  $N = 664$ ,  $4.3 \pm 1.6$ ;  $p < 0.01$  and power  $>99\%$  for all comparisons; Figure 3C). Additionally, in this analysis, long tails were shown in data distribution according to A and C grades, while B grade ICM/TE were associated with a more widespread distribution.

Conventionally, according to our internal grading method [13], AA blastocysts are considered of excellent quality; AB and BA of good quality; BB, AC, and CA of average quality; and CC, BC, and CB of poor quality. iDAScore v1.0 mirrors this clustering, as shown in Figure 4, except for a slight propensity to weigh the TE as more relevant than the ICM. In fact, in the “average quality” cluster, CA blastocysts ( $N = 14$ ,  $6.3 \pm 1.5$ ) showed iDAScore v1.0 higher than BB blastocysts ( $N = 446$ ,  $5.6 \pm 1.8$ ,  $p = 0.02$ ). In the “poor quality” (lower than BB) cluster, CC blastocysts ( $N = 483$ ,  $4.1 \pm 1.5$ ) resulted in an iDAScore v1.0 lower than both BC and CB ones ( $N = 162$ ,  $4.6 \pm 1.6$  and  $N = 167$ ,  $5.1 \pm 2.0$ , respectively;  $p = 0.05$  and  $p < 0.01$ , respectively; Figure 4). iDAScore v1.0 within each blastocyst morphology group also decreases according to the time of biopsy, with sharper decreases in the good and average quality groups and milder decreases in the excellent and poor quality ones (Supplementary Figure S1B).



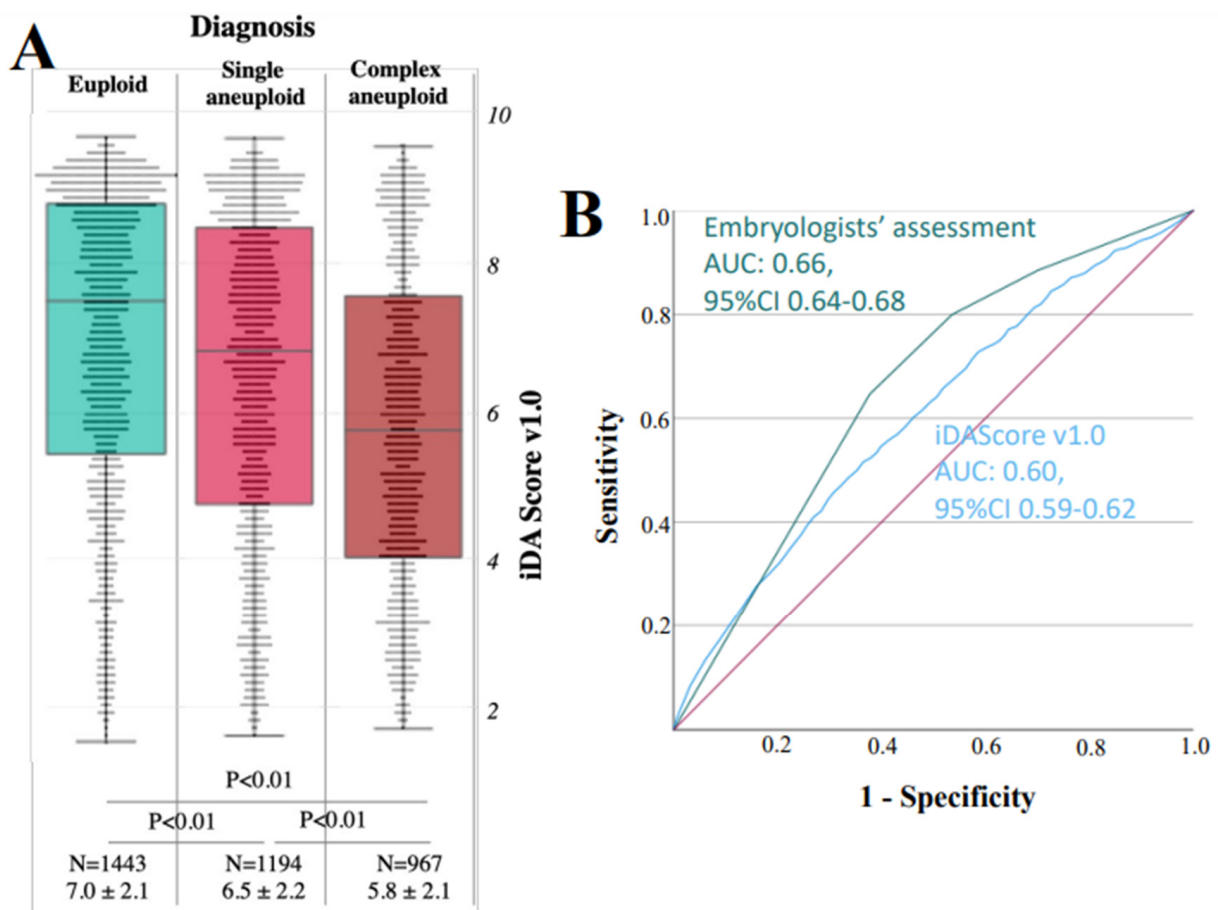
**Figure 4.** iDAScore v1.0 is associated with overall blastocyst quality. A, B, and C (from best to worst quality) outline inner cell mass (the first letter) and trophectoderm (the second letter) quality as defined by the embryologists according to the Gardner's score. Overall blastocyst quality is defined as excellent (AA, green), good (AB and BA, blue), average (AC, CA, and BB, gold), or poor (BC, CB, and CC, grey) according to Gardner's score adapted by Capalbo et al. [13]. The table summarizes the  $p$ -values of each sub-group comparison.

It is interesting that iDAScore v1.0 also slightly decreases according to maternal age at oocyte retrieval (unstandardized coefficient B:  $-0.036$ , 95% CI from  $-0.057$  to  $-0.015$ ,  $p < 0.01$ ; Supplementary Figure S2). Although clinically irrelevant, as this tool is intended to prioritize the blastocyst for transfer within a cohort of siblings (i.e., deriving from equally aged oocytes), this correlation supports a general association between advanced maternal age, poorer blastocyst morphology, and lower competence.



### 3.3. iDAScore v1.0 Is Significantly Associated with Euploidy, but the AUC Is 0.60

Euploid blastocysts showed significantly higher iDAScore v1.0 ( $N = 1443$ ,  $7.0 \pm 2.1$ ) than single ( $N = 1194$ ,  $6.5 \pm 2.2$ ,  $p < 0.01$  and power  $> 99\%$ ) and especially complex aneuploid embryos ( $N = 967$ ,  $5.8 \pm 2.1$ ,  $p < 0.01$  and power  $> 99\%$ ; Figure 5A). Indeed, a logistic regression analysis adjusted for maternal age confirmed an association (multivariate OR: 1.18, 95% CI 1.14–1.22,  $p < 0.01$ ) (Table 1). In addition, a good association was shown with the conventional parameters of embryo grading as reported in Table 2. The ROC curve analysis, in fact, highlighted an AUC of 0.66 (95% CI 0.64–0.68) for the discrimination between embryologists' assessment and euploidy, and a lower AUC of 0.60 (95% CI 0.59–0.62) for iDAScore v1.0 (Figure 5B). Of note, iDAScore v1.0 decreases rather uniformly according to the time of biopsy in the groups euploid, single, and complex aneuploid (Supplementary Figure S3A). There is no additional discrimination due to iDAScore v1.0 between euploid and aneuploid blastocysts within embryo quality groups as defined by the embryologists (Supplementary Figure S3B). Conversely, within the day of biopsy groups (5 and 6), significantly different iDAScore v1.0 were still observed between euploid and aneuploid embryos (Supplementary Figure S3C).



**Figure 5.** iDAScore v1.0 is associated with blastocysts' chromosomal constitution, but the AUC is 0.60. (A) Association between iDAScore v1.0 and blastocysts' chromosomal constitution clustered as euploid, single aneuploid, or complex aneuploid; (B) Receiver operating characteristic (ROC) curve analysis. The green curve represents the discrimination of embryologists' assessment upon euploidy with an area under the curve (AUC) of 0.66, while the blue curve represents the discrimination of iDAScore v1.0 upon euploidy with an AUC of 0.60.

**Table 1.** Logistic regressions for the association between iDAScore v1.0 with euploidy (adjusted for maternal age) and live birth (LB) after euploid single embryo transfer (SET).

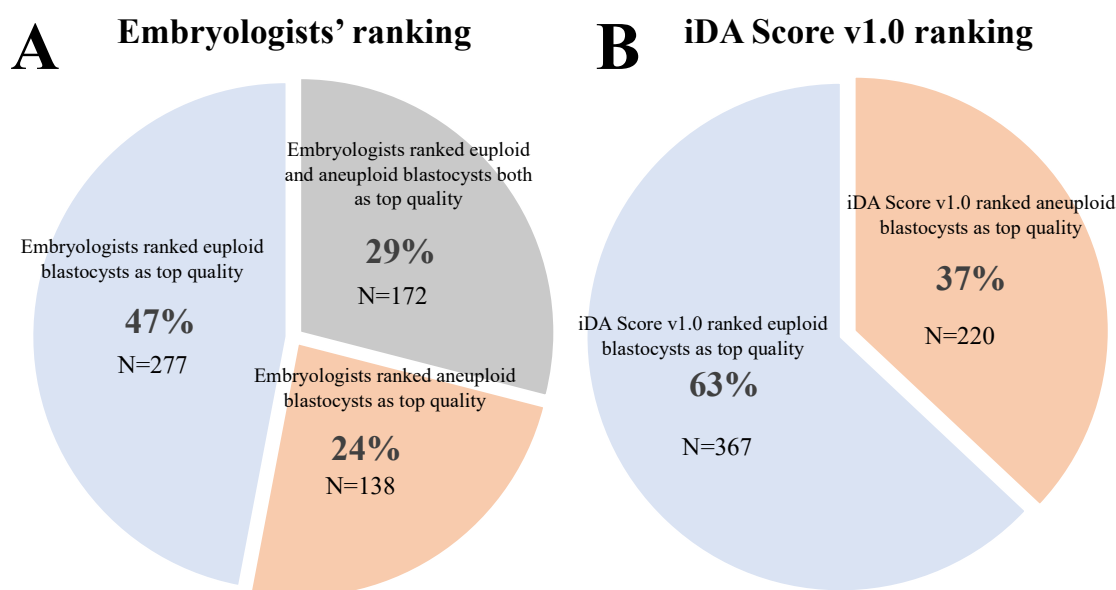
Outcome: euploidy	Univariate OR, 95% CI, p-value	Multivariate-OR, 95% CI, p-value
Maternal age	0.82, 95% CI 0.8–0.84, $p < 0.01$	0.82, 95% CI 0.8–0.84, $p < 0.01$
iDAScore v1.0	1.18, 95% CI 1.14–1.22, $p < 0.01$	1.18, 95% CI 1.14–1.22, $p < 0.01$
Outcome: LB per euploid SET	Univariate OR, 95% CI, p-value	-
iDAScore v1.0	1.30, 95% CI 1.2–1.4, $p < 0.01$	-

**Table 2.** Logistic regressions for the association between embryologists' assessment with euploidy (adjusted for maternal age) and live birth (LB) after euploid single embryo transfer (SET).

Outcome: euploidy	Univariate OR, 95% CI, p-value	Multivariate OR, 95% CI, p-value
Maternal age	0.82, 95% CI 0.8–0.84, $p < 0.01$	0.82, 95% CI 0.8–0.84, $p < 0.01$
Blastocyst quality:		
AA	-	-
AB, BA	0.57, 95% CI 0.47–0.69, $p < 0.01$	0.57, 95% CI 0.47–0.71, $p < 0.01$
BB, AC, CA	0.30, 95% CI 0.24–0.38, $p < 0.01$	0.32, 95% CI 0.25–0.40, $p < 0.01$
CC, BC, CA	0.23, 95% CI 0.19–0.27, $p < 0.01$	0.25, 95% CI 0.2–0.31, $p < 0.01$
Day of biopsy:		
5	-	-
6	0.62, 95% CI 0.54–0.72, $p < 0.01$	1.02, 95% CI 0.87–1.2, $p = 0.81$
7	0.34, 95% CI 0.25–0.45, $p < 0.01$	0.78, 95% CI 0.55–1.1, $p = 0.16$
Outcome: LB per euploid SET	Univariate OR, 95% CI, p-value	Multivariate OR, 95% CI, p-value
Blastocyst quality:		
AA	-	-
AB, BA	0.61, 95% CI 0.40–0.94, $p = 0.02$	0.72, 95% CI 0.46–1.11, $p = 0.14$
BB, AC, CA	0.39, 95% CI 0.22–0.70, $p < 0.01$	0.50, 95% CI 0.28–0.90, $p = 0.02$
CC, BC, CA	0.18, 95% CI 0.09–0.35, $p < 0.01$	0.24, 95% CI 0.12–0.47, $p < 0.01$
Day of biopsy:		
5	-	-
6	0.48, 95% CI 0.36–0.64, $p < 0.01$	0.59, 95% CI 0.44–0.81, $p < 0.01$
7	0.26, 95% CI 0.11–0.63, $p < 0.01$	0.47, 95% CI 0.19–1.18, $p = 0.11$

### 3.4. When Both Euploid and Aneuploid Embryos Were Diagnosed from the Same Cohort, iDAScore v1.0 Ranked the Euploid Blastocyst on Top in 63% of the Cases

In 587 cycles, both euploid and aneuploid blastocysts were diagnosed. According to the embryologists' assessment, the blastocysts ranked as top quality in their cohort of siblings were euploid in 47% of the cases and aneuploid in 24% of the cases. In the remaining 29% of the cases, euploid and aneuploid blastocysts were equally ranked as top quality (Figure 6A). According to iDAScore v1.0, in 63% and 37% of the cases, respectively, a euploid and an aneuploid blastocyst would have been ranked as top quality (Figure 6B). In the latter simulation, it is indeed unlikely that two or more blastocysts would have the same score. In fact, a 0.1 difference is sufficient to rank a blastocyst as better than another.



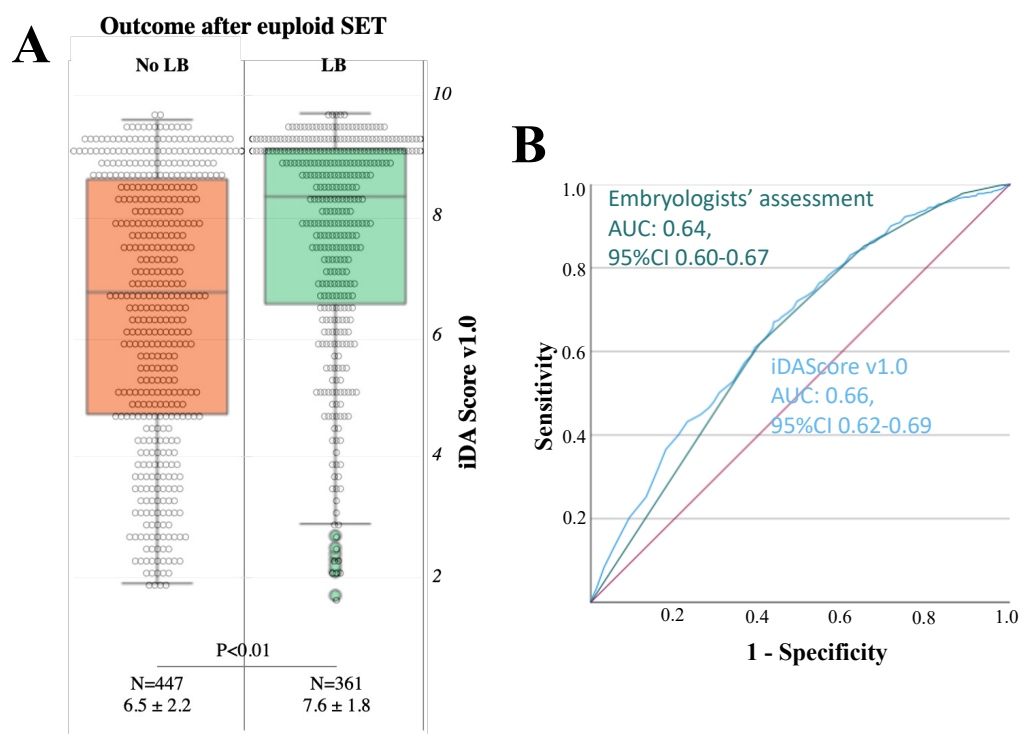
**Figure 6.** Association between the top-ranked blastocysts within each cohort according to: (A) the embryologists' ranking, and (B) iDA Score v1.0 ranking and their chromosomal constitution.

### 3.5. iDA Score v1.0 Is Significantly Associated with the Achievement of a LB after Euploid Blastocyst SET, with a 0.66 AUC

LB showed significantly higher iDA Score v1.0 ( $N = 361$ ,  $7.6 \pm 1.8$ ) than no LB ( $N = 447$ ,  $6.5 \pm 2.2$ ,  $p < 0.01$  and power  $> 99\%$ ; Figure 7A), and logistic regression analysis confirmed this association (OR: 1.3, 95% CI 1.2–1.4,  $p < 0.01$ ) (Table 1). Nevertheless, a good association was also shown with the conventional parameters of embryo grading and the day of biopsy, as described in Table 2. The ROC curve analyses, in fact, were almost superimposable: AUC 0.64 (95% CI 0.60–0.67) for the association between embryologists' assessment and euploidy, and AUC 0.66 (95% CI 0.62–0.69) for iDA Score v1.0 (Figure 7B). Interestingly, a larger reduction in iDA Score v1.0 was reported according to the time of biopsy in the group “no LB” with respect to the group “LB” (Supplementary Figure S4A). In addition, significantly higher iDA Score v1.0 characterized the blastocysts resulting in a LB versus the ones that did not also for both the sub-analyses within: (i) embryo morphology groups as outlined by the embryologists, except for the group “<BB” (Supplementary Figure S4B), and (ii) day of biopsy groups as outlined by the hpi, except for the group “day7” (Supplementary Figure S4C).

### 3.6. When at Least Two Euploid Blastocysts Were Available from the Same Cohort, the Embryologists Would Have Often Disagreed with iDA Score v1.0 on Their Ranking

In 202 cycles, at least two euploid blastocysts were available for transfer (the raw data are shown in Supplementary Table S2). In 52% of these cases, the embryologists and iDA Score v1.0 would have been equally effective since they would have either agreed on the blastocyst to prioritize for transfer, and that resulted in a LB, or they would have disagreed, but both would have been correct (Supplementary Figure S5; Supplementary Table S2). In 15% of the cases, iDA Score v1.0 would have identified the competent embryo earlier than the embryologists, while in 3% of the cases, iDA Score v1.0 would have identified the competent embryo later than the embryologists (Supplementary Figure S5; Supplementary Table S2). Nevertheless, this simulation is partially biased, because in 29% of the cases iDA Score v1.0 putative influence could not be assessed. Specifically, in discordant cases where the embryologists' choice for transfer resulted in a LB, but the highest ranked blastocysts according to iDA Score v1.0 had not been transferred, so their reproductive competence is unknown (Supplementary Figure S5; Supplementary Table S2). Consequently, the rate of equal and poorer effectiveness instances of iDA Score v1.0 in relation to the embryologists might be higher.



**Figure 7.** iDAScore v1.0 is associated with live birth (LB) after euploid blastocysts single embryo transfer (SET), but the AUC is 0.66. **(A)** Association between iDAScore v1.0 and a negative (no LB) or positive (LB) clinical outcome; **(B)** Receiver operating characteristic (ROC) curve analysis. The green curve represents the discrimination of embryologists' assessment upon a LB after euploid blastocyst SETs with an area under the curve (AUC) of 0.64, while the blue curve represents the discrimination of iDAScore v1.0 upon a LB after euploid blastocyst SETs with an AUC of 0.66.

#### 4. Discussion

AI and automation will strongly impact the future of IVF, meeting the needs for standardization and lower workload in the laboratories [11,32,51,52]. Nonetheless, AI-powered tools for embryo selection purposes requires further refinement, as most studies in this field show significant and recurrent limitations: (i) the nature of the training datasets is not representative of all clinical practices, (ii) the use of clinical pregnancy or fetal heartbeat as an endpoint rather than LB, (iii) the low sample sizes, and (iv) the lack of multicenter validation data [30,37]. In this study, we aimed at validating iDAScore v1.0 in ICSI cycles with TE biopsy, CCT analysis and vitrified-warmed euploid blastocyst SETs. We assessed: (i) its association with embryo morphology, day of development, euploidy, and LB, and (ii) its putative clinical utility in a retrospective simulation.

As previously reported by other studies, iDAScore v1.0 demonstrated a good correlation with the morphological parameters assigned by experienced embryologists to each blastocyst, either for ICM *per se* and TE *per se*, or for overall blastocyst morphology [31,53]. Notably, whenever the ICM and the TE of the same blastocyst were reported of a different quality (e.g., AC versus CA, BC versus CB), iDAScore v1.0 favored the latter. This trend inherently advocates a better predictivity of TE quality upon embryo implantation, as already suggested previously from other groups not using AI-powered tools [54–59]. Moreover, embryos achieving full blastocyst expansion on day 5 (<120 hpi) showed higher iDAScore v1.0 than embryos reaching that same stage on day 6 (121–144 hpi) or 7 (>144 hpi), also suggesting a better quality for the former, consistent with our previous analysis based on a different AI tool [14]. These data overall support the use of iDAScore v1.0 to objectify blastocyst morphological assessment within and between IVF clinics, as well as between professionals working at different laboratories, regardless of their experience, social, economic, clinical, and regulatory contexts, potentially influencing clinical choices [16,17,60].

Of note, the performance of iDAScore v1.0 in the evaluation of day 7 and/or blastocysts lower than BB might be suboptimal. In fact, the last frame analyzed by the software in its current version is at 140 hpi, thus it will not capture embryo development on day 7. In addition, poor-quality blastocysts are often deselected by embryologists worldwide, and thus are not sufficiently represented in the training data set. Future versions of iDAScore may benefit from datasets enriched in these populations of embryos and, so far, an early validation of v2.0 already shows significantly improved model performance in comparison to v1.0 with extended image analysis up to 148 hpi [61].

Although a large proportion of excellent quality blastocysts (AA) in a general population of advanced maternal age women are aneuploid ( $\approx 50\%$ ), while  $\approx 25\%$  of poor quality ones (lower than BB) might be euploid [15], a significant association exists between embryos' morphology and their chromosomal and reproductive competence [6–8]. Therefore, we tested iDAScore v1.0 in our dataset for its association with euploidy, to investigate whether this software may play a role in prioritizing euploid blastocysts for ET. Indeed, invasive PGT-A is still the only approach to reliably deselect blastocysts diagnosed with full chromosome aneuploidies and achieve a negative predictive value as high as 98% (i.e., lethality rate when aneuploid ETs were conducted in blinded non-selection studies) [62]. Yet, PGT-A is not universally applicable due to regulations, costs, and expertise; therefore, the long-lasting quest for non-invasive biomarkers of euploidy has also recently focused on AI-powered morphological and morphodynamic assessments [8,63–66]. Here, we report a significant association between iDAScore v1.0 and euploidy, even after results are clustered according to the day of full blastocyst expansion (day 5–7). The same result is not achieved for excellent, good, average, and poor blastocyst morphology clusters, as defined by conventional embryologists' assessment. This sub-analysis explains why the AUC for euploidy prediction derived from embryologist's evaluation performs better than that of iDAScore v1.0. It must be said, though, that the former approach is limited by its intrinsic subjectivity and limited generalizability [11] as well as by its poor ranking potential (i.e., based on three ICM morphology classes and three TE morphology classes in day 5, 6 or 7 after insemination). Conversely, the latter is more objective and reproducible, and leverages on a score difference as low as 0.1 to discriminate between embryos of slightly different quality. Indeed, among those cycles with at least one euploid and one aneuploid sibling blastocysts (about 50% of the cycles conducted during the study period), iDAScore v1.0 would have ranked euploid embryos as top quality in 63% of the cases versus 47% of the cases for embryologists' assessment. This latter estimate was indeed influenced by 29% of cases where euploid and aneuploid blastocysts were both tied as top quality according to embryologists' assessment.

Of note, iDAScore v1.0 was trained to predict implantation, not euploidy, and although (except for vital aneuploidies) a blastocyst must be euploid to result in a LB, as many as 50% of euploid blastocysts typically fail to implant [20–22]. It is therefore remarkable that iDAScore v1.0 showed a more evident association with LB in the context of euploid blastocyst SETs than with euploidy *per se*, consistent with recent evidence shown by a Japanese group for untested embryos [67]. It is promising that the AUC for the LB outcome mirrored the AUC resulting from embryologists' assessment. It is also reassuring that significantly higher iDAScore v1.0 output was observed for euploid blastocysts resulting in a LB versus reproductively incompetent ones among embryo quality and day of development clusters. These observations may suggest an additional application for iDAScore v1.0, namely it may be the swing vote among sibling euploid blastocysts to rank the embryos for ET. To this end, we conducted a second simulation in all the cycles where at least two euploid blastocysts and at least one LB were obtained (about 20% of the cycles included). This was aimed at understanding how often iDAScore v1.0 would have modified embryologists' choice, had it been used clinically. Notably, a different choice would have occurred in about 50% of the cases, suggesting a concrete influence of this tool, also in the context of PGT-A cycles. According to our data, iDAScore v1.0 would have involved delayed LB outcome in 3% of the cases, and an earlier one in 15% of the cases. However, this simulation may be



unbalanced in favor of the software over the operators, because in 59 cases (29%) iDAScore v1.0 putative influence could not be assessed. Specifically, in these cycles the embryologists transferred euploid blastocysts which resulted in a LB, but which were not graded as top quality by iDAScore v1.0, while top scoring blastocysts were not yet transferred at the time of drafting of this paper. On the contrary, the cases where iDAScore v1.0 would have outperformed the embryologists could all be computed, as the incorrect choice of the latter (i.e., no LB achieved) was always evident. Nevertheless, we chose to report these preliminary data here because they represent relevant (although only observational) evidence of the potential contribution of this tool for embryo selection purposes beyond euploidy. An RCT comparing embryologists with iDAScore performance per ET is certainly needed now to assess the true clinical contribution of this tool for embryo selection purposes.

## 5. Conclusions

Several embryo evaluation tools based on AI technologies have been proposed in IVF to date. For instance, the Spatial–Temporal Ensemble Model (STEM) and its upgrade, STEM+, promisingly reported to be able to predict blastocyst formation with high accuracy and AUC [68]. In a previous work from our group, we reported good consistency between an AI-powered software named CHLOE<sup>TM</sup> and blastocyst quality as defined by clinical embryologists [14]. IVY, a deep learning model producing a score between 0 and 1, was also tested for its prediction of the likelihood of blastocyst implantation and showed encouraging results [50]. Lately, then, some tools have been assessed for a putative prediction of euploidy. Embryo Ranking Intelligent Classification Algorithm (ERICA), for instance, outperformed clinical embryologists in ranking euploid blastocysts as top quality in their cohorts [33], and others such as Euploid Prediction Algorithm (EPA), STORK-A and iDAScore v1.0 itself showed good correlation with euploidy [65,66,69]. Nonetheless, in our view, present and future AI-powered tools should be aimed at supporting embryologists in prioritizing for (either untested or euploid) transfer the embryo(s) more likely to result in a live birth in their cohort, rather than at predicting euploidy (e.g., [70]). The accurate diagnosis of euploidy, as of today, still requires comprehensive chromosome testing technologies—with no report of mosaicism based on ICN—and invasive TE biopsy sampling approaches. Most importantly, although it is essential, euploidy is not sufficient to obtain a healthy baby, and the prediction of this latter outcome—and not of euploidy—should be the main aim of any embryo selection tool. This manuscript summarizes an independent, external, pre-clinical validation of iDAScore v1.0, one of the currently available AI-powered software programs for embryo selection. Within the limitations of retrospective design, our data support iDAScore as a promising tool to objectify embryo evaluation across embryologists and clinics, while preventing time-consuming and potentially biased morphokinetic manual annotations. The current v1.0 model performance upon euploidy and LB after euploid SETs is equivalent to embryologists' performance. Nonetheless, this can be considered a positive result for at least two reasons: (i) iDAScore was not trained to address these outcomes, and (ii) the AUC would be independent from embryologists' subjective assessment and increase objectivity and reproducibility. We also provided preliminary evidence of the current clinical utility of iDAScore v1.0, had it been used for embryo ranking purposes. We advocate for a prospective, possibly multicenter, study to confirm our data with an RCT design. A cost-effectiveness analysis is desirable as well, which should include information about lab workload with and without this tool.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jcm12051806/s1>. Figure S1: iDAScore v1.0 decreases with increasing time of biopsy in hours post insemination (hpi); Figure S2: iDAScore v1.0 decreases with increasing maternal age (with 95% CI); Figure S3: (A) iDAScore v1.0 decreases according to the time of biopsy (in hours post insemination, hpi) in the three clusters of blastocyst's chromosomal constitution (euploid, green; single aneuploid, light red; complex aneuploid, dark red). In the sub-group analyses, (B) iDAScore v1.0 were similar between euploid and aneuploid blastocysts across the 4 clusters of overall blastocyst quality as defined by the embryologists (AA [excellent quality blastocysts]; AB or BA [good quality blastocysts]; AC, CA, or BB [average quality blastocysts]; BC, BC, or CC [poor quality blastocysts]), while (C) it was still significantly different among blastocysts clustered as day 5 ( $\leq 120$  hpi) or day 6 (121–144 hpi) (for day 7 blastocysts,  $>144$  hpi, the  $p$ -value was  $>0.05$ ); Figure S4: (A) iDAScore v1.0 shows a sharper decrease according to the time of biopsy (in hours post insemination, hpi) among euploid blastocysts that did not result in a live birth (LB, orange) than among the ones that did result in a LB (light green). In the sub-group analyses, (B) iDAScore v1.0 were significantly different in the “no LB” and “LB” groups further clustered according to the overall blastocyst quality as defined by the embryologists (AA [excellent quality blastocysts]; AB or BA [good quality blastocysts]; AC, CA or BB [average quality blastocysts]; for BC, BC or CC [poor quality blastocysts]  $p$ -value was  $>0.05$ ). Similarly, (C) iDAScores v1.0 were significantly different in the “no LB” and “LB” groups further clustered according to the day of biopsy (day 5,  $\leq 120$  hpi; day 6, 121–144 hpi; for day 7 blastocysts,  $>144$  hpi, the  $p$ -value was  $>0.05$ ); Figure S5: Simulation of iDAScore v1.0 clinical utility had we use it in the context of IVF cycles with preimplantation genetic testing for aneuploidies (PGT-A); Table S1: Patients and cycles characteristics; Table S2: Raw data for the concordance between the operators and the iDA Score v1.0 in ranking euploid blastocysts for transfer. We included all cycles between April 2013 and December 2021 where at least 2 euploid embryos were obtained, at least one live birth was achieved, and the ranking outlined by the iDA Score v1.0 could be assessed as beneficial or not.

**Author Contributions:** D.C., J.B., M.L., A.V., F.M.U. and L.R. conceived and designed the study. D.C., V.C. (Viviana Chiappetta), F.I., L.A., G.S. and A.M. collected the data. D.C. and J.B. analyzed the data. D.C., V.C. (Viviana Chiappetta), F.I., G.S., M.T. and A.M. drafted the manuscript. L.A., R.M., V.C. (Valentina Casciani), G.C., A.A., M.L., A.V., F.M.U., A.B. and L.R. reviewed the manuscript and contributed to the discussion of the evidence. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Clinica Valle Giulia (CVG25052021; date: 25 May 2021).

**Informed Consent Statement:** Informed consent for the retrospective anonymous analysis of the data was obtained from all subjects involved in the study.

**Data Availability Statement:** Data are contained within the article and Supplementary Material.

**Acknowledgments:** The work performed by GS and MT is part of their Master Course in ‘Biology and Biotechnology of Reproduction: from research to the clinics’ of the University of Pavia.

**Conflicts of Interest:** J.B. and M.L. are employees and shareholders of Vitrolife. All other authors declare no conflict of interest related with this study.

## References

1. Ferraretti, A.; Nygren, K.; Andersen, A.N.; De Mouzon, J.; Kupka, M.; Calhaz-Jorge, C.; Wyns, C.; Gianaroli, L.; Goossens, V.; European IVF-Monitoring Consortium (EIM), for the European Society of Human Reproduction and Embryology (ESHRE); et al. Trends over 15 years in ART in Europe: An analysis of 6 million cycles†. *Hum. Reprod. Open* **2017**, 2017, hox012. [\[CrossRef\]](#)
2. Rienzi, L.; Cimadomo, D.; Vaiarelli, A.; Gennarelli, G.; Holte, J.; Livi, C.; Masip, M.A.; Uher, P.; Fabozzi, G.; Ubaldi, F.M. Measuring success in IVF is a complex multidisciplinary task: Time for a consensus? *Reprod. Biomed. Online* **2021**, 43, 775–778. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Alpha Scientists in Reproductive Medicine; ESHRE Special Interest Group of Embryology. The Istanbul consensus workshop on embryo assessment: Proceedings of an expert meeting. *Hum. Reprod.* **2011**, 26, 1270–1283. [\[CrossRef\]](#) [\[PubMed\]](#)

4. Gardner, D.K.; Schoolcraft, B. *In Vitro Culture of Human Blastocysts, toward Reproductive Certainty: Fertility and Genetics Beyond*; Parthenon Publishing London: London, UK, 1999; pp. 378–388.
5. Glujovsky, D.; Farquhar, C.; Retamar, A.M.Q.; Sedo, C.R.A.; Blake, D. Cleavage stage versus blastocyst stage embryo transfer in assisted reproductive technology. *Cochrane Database Syst. Rev.* **2016**, *5*, CD002118. [[CrossRef](#)] [[PubMed](#)]
6. Zhan, Q.; Sierra, E.; Malmsten, J.; Ye, Z.; Rosenwaks, Z.; Zaninovic, N. Blastocyst score, a blastocyst quality ranking tool, is a predictor of blastocyst ploidy and implantation potential. *F&S Rep.* **2020**, *1*, 133–141. [[CrossRef](#)]
7. Shear, M.A.; Vaughan, D.A.; Modest, A.M.; Seidler, E.A.; Leung, A.Q.; Hacker, M.R.; Sakkas, D.; Penzias, A.S. Blasts from the past: Is morphology useful in PGT-A tested and untested frozen embryo transfers? *Reprod. Biomed. Online* **2020**, *41*, 981–989. [[CrossRef](#)]
8. Bamford, T.; Barrie, A.; Montgomery, S.; Dhillon-Smith, R.; Campbell, A.; Easter, C.; Coomarasamy, A. Morphological and morphokinetic associations with aneuploidy: A systematic review and meta-analysis. *Hum. Reprod. Update* **2022**, *28*, 656–686. [[CrossRef](#)]
9. Storr, A.; Venetis, C.; 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.* **2017**, *32*, 307–314. [[CrossRef](#)]
10. Paternot, G.; Devroe, J.; Debrock, S.; D’Hooghe, T.M.; Spiessens, C. Intra- and inter-observer analysis in the morphological assessment of early-stage embryos. *Reprod. Biol. Endocrinol.* **2009**, *7*, 105. [[CrossRef](#)]
11. Cimadomo, D.; Sosa Fernandez, L.; Soscia, D.; Fabozzi, G.; Benini, F.; Cesana, A.; Dal Canto, M.B.; 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. [[CrossRef](#)]
12. Coticchio, G.; Barrie, A.; Lagalla, C.; Borini, A.; Fishel, S.; Griffin, D.; Campbell, A. Plasticity of the human preimplantation embryo: Developmental dogmas, variations on themes and self-correction. *Hum. Reprod. Update* **2021**, *27*, 848–865. [[CrossRef](#)]
13. Capalbo, A.; Rienzi, L.; Cimadomo, D.; Maggiulli, R.; Elliott, T.; Wright, G.; Nagy, Z.P.; Ubaldi, F.M. Correlation between standard blastocyst morphology, euploidy and implantation: An observational study in two centers involving 956 screened blastocysts. *Hum. Reprod.* **2014**, *29*, 1173–1181. [[CrossRef](#)] [[PubMed](#)]
14. Cimadomo, D.; Soscia, D.; Casciani, V.; Innocenti, F.; Trio, S.; Chiappetta, V.; Albricci, L.; Maggiulli, R.; Erlich, I.; Ben-Meir, A.; et al. How slow is too slow? A comprehensive portrait of Day 7 blastocysts and their clinical value standardized through artificial intelligence. *Hum. Reprod.* **2022**, *37*, 1134–1147. [[CrossRef](#)] [[PubMed](#)]
15. Cimadomo, D.; Soscia, D.; Vaiarelli, A.; Maggiulli, R.; Capalbo, A.; Ubaldi, F.M.; Rienzi, L. Looking past the appearance: A comprehensive description of the clinical contribution of poor-quality blastocysts to increase live birth rates during cycles with aneuploidy testing. *Hum. Reprod.* **2019**, *34*, 1206–1214. [[CrossRef](#)] [[PubMed](#)]
16. Kemper, J.M.; Liu, Y.; Afnan, M.; Hammond, E.R.; Morbeck, D.E.; Mol, B.W.J. Should we look for a low-grade threshold for blastocyst transfer? A scoping review. *Reprod. Biomed. Online* **2021**, *42*, 709–716. [[CrossRef](#)]
17. Morbeck, D.E. Blastocyst culture in the Era of PGS and FreezeAlls: Is a ‘C’ a failing grade? *Hum. Reprod. Open* **2017**, *2017*, hox017. [[CrossRef](#)]
18. Cimadomo, D.; Rienzi, L.; Capalbo, A.; Rubio, C.; Innocenti, F.; Garcia-Pascual, C.M.; Ubaldi, F.M.; Handyside, A. The dawn of the future: 30 years from the first biopsy of a human embryo. The detailed history of an ongoing revolution. *Hum. Reprod. Update* **2020**, *26*, 453–473. [[CrossRef](#)]
19. Scott, R.T., Jr.; Upham, K.M.; Forman, E.J.; Zhao, T.; Treff, N.R. Cleavage-stage biopsy significantly impairs human embryonic implantation potential while blastocyst biopsy does not: A randomized and paired clinical trial. *Fertil. Steril.* **2013**, *100*, 624–630. [[CrossRef](#)]
20. Tieg, A.W.; Tao, X.; Zhan, Y.; Whitehead, C.; Kim, J.; Hanson, B.; Osman, E.; Kim, T.J.; Patounakis, G.; Gutmann, J.; et al. A multicenter, prospective, blinded, nonselection study evaluating the predictive value of an aneuploid diagnosis using a targeted next-generation sequencing-based preimplantation genetic testing for aneuploidy assay and impact of biopsy. *Fertil. Steril.* **2020**, *115*, 627–637. [[CrossRef](#)]
21. Dahdouh, E.M.; Balayla, J.; Garcia-Velasco, J.A. Comprehensive chromosome screening improves embryo selection: A meta-analysis. *Fertil. Steril.* **2015**, *104*, 1503–1512. [[CrossRef](#)]
22. Chen, M.; Wei, S.; Hu, J.; Quan, S. Can Comprehensive Chromosome Screening Technology Improve IVF/ICSI Outcomes? A Meta-Analysis. *PLoS ONE* **2015**, *10*, e0140779. [[CrossRef](#)] [[PubMed](#)]
23. Apter, S.; Ebner, T.; Freour, T.; Guns, Y.; Kovacic, B.; Le Clef, N.; Marques, M.; Meseguer, M.; Montjean, D.; Sfontouris, I.; et al. Eshre Working group on Time-lapse technology: Good practice recommendations for the use of time-lapse technology. *Hum. Reprod. Open* **2020**, *2020*, hoaa008. [[CrossRef](#)] [[PubMed](#)]
24. Ciray, H.N.; Campbell, A.; Agerholm, I.E.; Aguilar, J.; Chamayou, S.; Esbert, M.; Sayed, S.; Time-Lapse User, G. Proposed guidelines on the nomenclature and annotation of dynamic human embryo monitoring by a time-lapse user group. *Hum. Reprod.* **2014**, *29*, 2650–2660. [[CrossRef](#)] [[PubMed](#)]
25. Armstrong, S.; Bhide, P.; Jordan, V.; Pacey, A.; Farquhar, C. Time-lapse systems for embryo incubation and assessment in assisted reproduction. *Cochrane Database Syst. Rev.* **2018**, *5*, CD011320. [[CrossRef](#)]
26. 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*, 511–520. [[CrossRef](#)]
27. 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*, 722–729. [[CrossRef](#)]

28. Rienzi, L.; Capalbo, A.; Stoppa, M.; Romano, S.; Maggiulli, R.; Albricci, L.; Scarica, C.; Farcomeni, A.; Vajta, G.; Ubaldi, F.M. No evidence of association between blastocyst aneuploidy and morphokinetic assessment in a selected population of poor-prognosis patients: A longitudinal cohort study. *Reprod. Biomed. Online* **2015**, *30*, 57–66. [\[CrossRef\]](#)
29. Swain, J.; VerMilyea, M.T.; Meseguer, M.; Ezcurra, D.; Fertility, A.I.F.G. AI in the treatment of fertility: Key considerations. *J. Assist. Reprod. Genet.* **2020**, *37*, 2817–2824. [\[CrossRef\]](#)
30. Riegler, M.A.; Stensen, M.H.; Wiczak, O.; Andersen, J.M.; Hicks, S.A.; Hammer, H.L.; Delbarre, E.; Halvorsen, P.; Yazidi, A.; Holst, N.; et al. Artificial intelligence in the fertility clinic: Status, pitfalls and possibilities. *Hum. Reprod.* **2021**, *36*, 2429–2442. [\[CrossRef\]](#)
31. Berntsen, J.; Rimestad, J.; Lassen, J.T.; Tran, D.; Kragh, M.F. Robust and generalizable embryo selection based on artificial intelligence and time-lapse image sequences. *PLoS ONE* **2022**, *17*, e0262661. [\[CrossRef\]](#)
32. Khosravi, P.; Kazemi, E.; Zhan, Q.; Malmsten, J.E.; Toschi, M.; Zisimopoulos, P.; Sigaras, A.; Lavery, S.; Cooper, L.A.D.; Hickman, C.; et al. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *NPJ Digit. Med.* **2019**, *2*, 21. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Chavez-Badiola, A.; Flores-Saiffe-Farias, A.; Mendizabal-Ruiz, G.; Drakeley, A.J.; Cohen, J. Embryo Ranking Intelligent Classification Algorithm (ERICA): Artificial intelligence clinical assistant predicting embryo ploidy and implantation. *Reprod. Biomed. Online* **2020**, *41*, 585–593. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Kragh, M.F.; Rimestad, J.; Berntsen, J.; Karstoft, H. Automatic grading of human blastocysts from time-lapse imaging. *Comput. Biol. Med.* **2019**, *115*, 103494. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Coticchio, G.; Ezoe, K.; Lagalla, C.; Shimazaki, K.; Ohata, K.; Ninomiya, M.; Wakabayashi, N.; Okimura, T.; Uchiyama, K.; Kato, K.; et al. Perturbations of morphogenesis at the compaction stage affect blastocyst implantation and live birth rates. *Hum. Reprod.* **2021**, *36*, 918–928. [\[CrossRef\]](#)
36. Coticchio, G.; Mignini Renzini, M.; Novara, P.V.; Lain, M.; De Ponti, E.; Turchi, D.; Fadini, R.; Dal Canto, M. Focused time-lapse analysis reveals novel aspects of human fertilization and suggests new parameters of embryo viability. *Hum. Reprod.* **2018**, *33*, 23–31. [\[CrossRef\]](#)
37. Kragh, M.F.; Karstoft, H. Embryo selection with artificial intelligence: How to evaluate and compare methods? *J Assist Reprod Genet* **2021**, *38*, 1675–1689. [\[CrossRef\]](#)
38. Reignier, A.; Lammers, J.; Barriere, P.; Freour, T. Can time-lapse parameters predict embryo ploidy? A systematic review. *Reprod. Biomed. Online* **2018**, *36*, 380–387. [\[CrossRef\]](#)
39. Rienzi, L.; Ubaldi, F.M.; Iacobelli, M.; Minasi, M.G.; Romano, S.; Ferrero, S.; Sapienza, F.; Baroni, E.; Litwicka, K.; Greco, E. Significance of metaphase II human oocyte morphology on ICSI outcome. *Fertil. Steril.* **2008**, *90*, 1692–1700. [\[CrossRef\]](#)
40. Ubaldi, F.M.; Capalbo, A.; Colamaria, S.; Ferrero, S.; Maggiulli, R.; Vajta, G.; Sapienza, F.; Cimadomo, D.; Giuliani, M.; Gravotta, E.; et al. Reduction of multiple pregnancies in the advanced maternal age population after implementation of an elective single embryo transfer policy coupled with enhanced embryo selection: Pre- and post-intervention study. *Hum. Reprod.* **2015**, *30*, 2097–2106. [\[CrossRef\]](#)
41. Cimadomo, D.; Capalbo, A.; Dovere, L.; Tacconi, L.; Soscia, D.; Giancani, A.; Scepi, E.; Maggiulli, R.; Vaiarelli, A.; Rienzi, L.; et al. Leave the past behind: Women's reproductive history shows no association with blastocysts' euploidy and limited association with live birth rates after euploid embryo transfers. *Hum. Reprod.* **2021**, *36*, 929–940. [\[CrossRef\]](#)
42. Maggiulli, R.; Cimadomo, D.; Fabozzi, G.; Papini, L.; Dovere, L.; Ubaldi, F.M.; Rienzi, L. The effect of ICSI-related procedural timings and operators on the outcome. *Hum. Reprod.* **2020**, *35*, 32–43. [\[CrossRef\]](#)
43. Maggiulli, R.; Giancani, A.; Cimadomo, D.; Ubaldi, F.M.; Rienzi, L. Human Blastocyst Biopsy and Vitrification. *J. Vis. Exp.* **2019**, *149*, e59625. [\[CrossRef\]](#)
44. Treff, N.R.; Tao, X.; Ferry, K.M.; Su, J.; Taylor, D.; Scott, R.T., Jr. Development and validation of an accurate quantitative real-time polymerase chain reaction-based assay for human blastocyst comprehensive chromosomal aneuploidy screening. *Fertil. Steril.* **2012**, *97*, 819–824. [\[CrossRef\]](#)
45. Garcia-Pascual, C.M.; Navarro-Sanchez, L.; Navarro, R.; Martinez, L.; Jimenez, J.; Rodrigo, L.; Simon, C.; Rubio, C. Optimized NGS Approach for Detection of Aneuploidies and Mosaicism in PGT-A and Imbalances in PGT-SR. *Genes* **2020**, *11*, 724. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Girardi, L.; Serdarogullari, M.; Patassini, C.; Poli, M.; Fabiani, M.; Caroselli, S.; Coban, O.; Findikli, N.; Boynukalin, F.K.; Bahceci, M.; et al. Incidence, Origin, and Predictive Model for the Detection and Clinical Management of Segmental Aneuploidies in Human Embryos. *Am. J. Hum. Genet.* **2020**, *106*, 525–534. [\[CrossRef\]](#)
47. Paulson, R.J.; Treff, N. Isn't it time to stop calling preimplantation embryos "mosaic"? *F&S Rep.* **2020**, *1*, 164–165. [\[CrossRef\]](#)
48. Forman, E.J. Demystifying "mosaic" outcomes. *Fertil. Steril.* **2019**, *111*, 253. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Capalbo, A.; Poli, M.; Rienzi, L.; Girardi, L.; Patassini, C.; Fabiani, M.; Cimadomo, D.; Benini, F.; Farcomeni, A.; Cuzzi, J.; et al. Mosaic human preimplantation embryos and their developmental potential in a prospective, non-selection clinical trial. *Am. J. Hum. Genet.* **2021**, *108*, 2238–2247. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Tran, D.; Cooke, S.; Illingworth, P.J.; Gardner, D.K. Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. *Hum. Reprod.* **2019**, *34*, 1011–1018. [\[CrossRef\]](#)



51. Alikani, M.; Go, K.J.; McCaffrey, C.; McCulloh, D.H. Comprehensive evaluation of contemporary assisted reproduction technology laboratory operations to determine staffing levels that promote patient safety and quality care. *Fertil. Steril.* **2014**, *102*, 1350–1356. [\[CrossRef\]](#)
52. Veiga, E.; Olmedo, C.; Sanchez, L.; Fernandez, M.; Mauri, A.; Ferrer, E.; Ortiz, N. Recalculating the staff required to run a modern assisted reproductive technology laboratory. *Hum. Reprod.* **2022**, *37*, 1774–1785. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Ezoe, K.; Shimazaki, K.; Miki, T.; Takahashi, T.; Tanimura, Y.; Amagai, A.; Sawado, A.; Akaike, H.; Mogi, M.; Kaneko, S.; et al. Association between a deep learning-based scoring system with morphokinetics and morphological alterations in human embryos. *Reprod. Biomed. Online* **2022**, *45*, 1124–1132. [\[CrossRef\]](#)
54. Ahlstrom, A.; Westin, C.; Wikland, M.; Hardarson, T. Prediction of live birth in frozen-thawed single blastocyst transfer cycles by pre-freeze and post-thaw morphology. *Hum. Reprod.* **2013**, *28*, 1199–1209. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Hill, M.J.; Richter, K.S.; Heitmann, R.J.; Graham, J.R.; Tucker, M.J.; DeCherney, A.H.; Browne, P.E.; Levens, E.D. Trophoctoderm grade predicts outcomes of single-blastocyst transfers. *Fertil. Steril.* **2013**, *99*, 1283–1289.e1281. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Chen, X.; Zhang, J.; Wu, X.; Cao, S.; Zhou, L.; Wang, Y.; Chen, X.; Lu, J.; Zhao, C.; Chen, M.; et al. Trophoctoderm morphology predicts outcomes of pregnancy in vitrified-warmed single-blastocyst transfer cycle in a Chinese population. *J. Assist. Reprod. Genet.* **2014**, *31*, 1475–1481. [\[CrossRef\]](#)
57. Thompson, S.M.; Onwubalili, N.; Brown, K.; Jindal, S.K.; McGovern, P.G. Blastocyst expansion score and trophoctoderm morphology strongly predict successful clinical pregnancy and live birth following elective single embryo blastocyst transfer (eSET): A national study. *J. Assist. Reprod. Genet.* **2013**, *30*, 1577–1581. [\[CrossRef\]](#)
58. Honnma, H.; Baba, T.; Sasaki, M.; Hashiba, Y.; Ohno, H.; Fukunaga, T.; Endo, T.; Saito, T.; Asada, Y. Trophoctoderm morphology significantly affects the rates of ongoing pregnancy and miscarriage in frozen-thawed single-blastocyst transfer cycle in vitro fertilization. *Fertil. Steril.* **2012**, *98*, 361–367. [\[CrossRef\]](#)
59. Ahlstrom, A.; Westin, C.; Reismer, E.; Wikland, M.; Hardarson, T. Trophoctoderm morphology: An important parameter for predicting live birth after single blastocyst transfer. *Hum. Reprod.* **2011**, *26*, 3289–3296. [\[CrossRef\]](#)
60. Hammond, E.R.; Foong, A.K.M.; Rosli, N.; Morbeck, D.E. 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*, 1045–1053. [\[CrossRef\]](#)
61. Lassen, J.T.; Kragh, M.F.; Rimestad, J.; Johansen, M.N.; Berntsen, J. Development and validation of deep learning based embryo selection across multiple days of transfer. *arXiv* **2022**. [\[CrossRef\]](#)
62. Capalbo, A.; Poli, M.; Jalas, C.; Forman, E.J.; Treff, N.R. On the reproductive capabilities of aneuploid human preimplantation embryos. *Am. J. Hum. Genet.* **2022**, *109*, 1572–1581. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Gazzo, E.; Pena, F.; Valdez, F.; Chung, A.; Bonomini, C.; Ascenzo, M.; Velit, M.; Escudero, E. The Kidscore(TM) D5 algorithm as an additional tool to morphological assessment and PGT-A in embryo selection: A time-lapse study. *JBRA Assist. Reprod.* **2020**, *24*, 55–60. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Diakiw, S.M.; Hall, J.M.M.; VerMilyea, M.; Lim, A.Y.X.; Quangkananurug, W.; Chanchamroen, S.; Bankowski, B.; Stones, R.; Storr, A.; Miller, A.; et al. An artificial intelligence model correlated with morphological and genetic features of blastocyst quality improves ranking of viable embryos. *Reprod. Biomed. Online* **2022**, *45*, 1105–1117. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Huang, B.; Tan, W.; Li, Z.; Jin, L. An artificial intelligence model (euploid prediction algorithm) can predict embryo ploidy status based on time-lapse data. *Reprod. Biol. Endocrinol.* **2021**, *19*, 185. [\[CrossRef\]](#)
66. Kato, K.; Ueno, S.; Berntsen, J.; Kragh, M.F.; Okimura, T.; Kuroda, T. Does embryo categorization by existing artificial intelligence, morphokinetic or morphological embryo selection models correlate with blastocyst euploidy rates? *Reprod. Biomed. Online* **2022**, *46*, 274–281. [\[CrossRef\]](#)
67. Ueno, S.; Berntsen, J.; Ito, M.; Okimura, T.; Kato, K. Correlation between an annotation-free embryo scoring system based on deep learning and live birth/neonatal outcomes after single vitrified-warmed blastocyst transfer: A single-centre, large-cohort retrospective study. *J. Assist. Reprod. Genet.* **2022**, *39*, 2089–2099. [\[CrossRef\]](#)
68. Liao, Q.; Zhang, Q.; Feng, X.; Huang, H.; Xu, H.; Tian, B.; Liu, J.; Yu, Q.; Guo, N.; Liu, Q.; et al. Development of deep learning algorithms for predicting blastocyst formation and quality by time-lapse monitoring. *Commun. Biol.* **2021**, *4*, 415. [\[CrossRef\]](#)
69. Barnes, J.; Brendel, M.; Gao, V.R.; Rajendran, S.; Kim, J.; Li, Q.; Malmsten, J.E.; Sierra, J.T.; Zisimopoulos, P.; Sigaras, A.; et al. A non-invasive artificial intelligence approach for the prediction of human blastocyst ploidy: A retrospective model development and validation study. *Lancet Digit. Health* **2023**, *5*, e28–e40. [\[CrossRef\]](#)
70. Rocafort, E.; Enciso, M.; Leza, A.; Sarasa, J.; Aizpurua, J. Euploid embryos selected by an automated time-lapse system have superior SET outcomes than selected solely by conventional morphology assessment. *J. Assist. Reprod. Genet.* **2018**, *35*, 1573–1583. [\[CrossRef\]](#)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.