

Embryology

An interpretable artificial intelligence approach to differentiate between blastocysts with similar or same morphological grades

Hang Liu ^{1,†}, Longbin Chen ^{2,†}, Guanqiao Shan ¹, Chen Sun¹, Changfu Lu^{2,3,4,5}, Hongqing Liao^{6,7}, Shuoping Zhang ³, Shaonan Dong³, Xinxin Xu³, Qiuyun Yan³, Fei Gong^{2,3,4,5}, Zhuoran Zhang⁸, Changsheng Dai⁹, Wenyuan Chen¹⁰, Haocong Song¹, Lei Chen¹¹, Shanshan Wang ^{11,*}, Haixiang Sun ^{11,*}, Ge Lin ^{2,3,4,5,12,*}, Yu Sun ^{1,10,13,14,*}, and Yifan Gu ^{2,3,4,5,*}

¹Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada

²Institute of Reproductive and Stem Cell Engineering, Xiangya School of Basic Medical Science, Central South University, Changsha, Hunan Province, China

³Reproductive & Genetic Hospital of CITIC-Xiangya, Changsha, Hunan Province, China

⁴NHC Key Laboratory of Human Stem Cell and Reproductive Engineering, Xiangya School of Basic Medical Science, Central South University, Changsha, Hunan Province, China

⁵Clinical Research Center for Reproduction and Genetics in Hunan Province, Reproductive & Genetic Hospital of CITIC-Xiangya, Changsha, Hunan Province, China

⁶Hengyang Nanhua-Xinghui Reproductive Health Hospital, Hengyang, Hunan Province, China

⁷The Second Affiliated Hospital, Hengyang Medical School, University of South China, Hengyang, Hunan Province, China

⁸School of Science and Engineering, The Chinese University of Hong Kong-Shenzhen, Longgang District, Shenzhen, China

⁹Institute of Robotics and Intelligent Systems, Dalian University of Technology, Dalian, Liaoning Province, China

¹⁰Department of Computer Science, University of Toronto, Toronto, ON, Canada

¹¹Center for Reproductive Medicine and Obstetrics and Gynecology, Nanjing Drum Tower Hospital, Nanjing University Medical School, Nanjing, Jiangsu Province, China

¹²National Engineering and Research Center of Human Stem Cell, Changsha, Hunan Province, China

¹³Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada

¹⁴Institute of Biomedical Engineering, University of Toronto, Toronto, ON, Canada

*Correspondence address. Center for Reproductive Medicine and Obstetrics and Gynecology, Nanjing Drum Tower Hospital, Nanjing University Medical School, 321, Zhongshan Road, Nanjing, Jiangsu Province 210008, China. E-mail: wss_19860820@sina.com  <https://orcid.org/0000-0002-1281-2343> (S.W.); Center for Reproductive Medicine and Obstetrics and Gynecology, Nanjing Drum Tower Hospital, Nanjing University Medical School, 321, Zhongshan Road, Nanjing, Jiangsu Province 210008, China. E-mail: stevensunz@163.com  <https://orcid.org/0000-0002-1215-1792> (H.S.); Institute of Reproductive and Stem Cell Engineering, Xiangya School of Basic Medical Science, Central South University, 172, Tongzipo Road, Changsha, Hunan Province 410013, China. E-mail: linggf@hotmail.com  <https://orcid.org/0000-0002-3877-2546> (G.L.); Department of Mechanical and Industrial Engineering, University of Toronto, 5, King's College Road, Toronto, ON M5S 3G8, Canada. E-mail: sun@mie.utoronto.ca  <https://orcid.org/0000-0001-7895-0741> (Y.S.); Institute of Reproductive and Stem Cell Engineering, Xiangya School of Basic Medical Science, Central South University, 172, Tongzipo Road, Changsha, Hunan Province 410013, China. E-mail: evangoo@163.com  <https://orcid.org/0000-0001-7669-5046> (Y.G.)

†These authors share joint first authorship.

ABSTRACT

STUDY QUESTION: Can a quantitative method be developed to differentiate between blastocysts with similar or same inner cell mass (ICM) and trophectoderm (TE) grades, while also reflecting their potential for live birth?

SUMMARY ANSWER: We developed BlastScoringNet, an interpretable deep-learning model that quantifies blastocyst ICM and TE morphology with continuous scores, enabling finer differentiation between blastocysts with similar or same grades, with higher scores significantly correlating with higher live birth rates.

WHAT IS KNOWN ALREADY: While the Gardner grading system is widely used by embryologists worldwide, blastocysts having similar or same ICM and TE grades cause challenges for embryologists in decision-making. Furthermore, human assessment is subjective and inconsistent in predicting which blastocysts have higher potential to result in live birth.

STUDY DESIGN, SIZE, DURATION: The study design consists of three main steps. First, BlastScoringNet was developed using a grading dataset of 2760 blastocysts with majority-voted Gardner grades. Second, the model was applied to a live birth dataset of 15 228 blastocysts with known live birth outcomes to generate blastocyst scores. Finally, the correlation between these scores and live birth outcomes was assessed. The blastocysts were collected from patients who underwent IVF treatments between 2016 and 2018. For external application study, an additional grading dataset of 1455 blastocysts and a live birth dataset of 476 blastocysts were collected from patients who underwent IVF treatments between 2021 and 2023 at an external IVF institution.

PARTICIPANTS/MATERIALS, SETTING, METHODS: In this retrospective study, we developed BlastScoringNet, an interpretable deep-learning model which outputs expansion degree grade and continuous scores quantifying a blastocyst's ICM morphology and TE morphology, based on the Gardner grading system. The continuous ICM and TE scores were calculated by weighting each base

Received: November 8, 2024. Revised: February 7, 2025. Editorial decision: March 19, 2025.

© The Author(s) 2025. Published by Oxford University Press on behalf of European Society of Human Reproduction and Embryology. All rights reserved.

For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

grade's predicted probability and summing the predicted probabilities. To represent each blastocyst's overall potential for live birth, we combined the ICM and TE scores using their odds ratios (ORs) for live birth. We further assessed the correlation between live birth rates and the ICM score, TE score, and the OR-combined score (adjusted for expansion degree) by applying BlastScoringNet to blastocysts with known live birth outcomes. To test its generalizability, we also applied BlastScoringNet to an external IVF institution, accounting for variations in imaging conditions, live birth rates, and embryologists' experience levels.

MAIN RESULTS AND THE ROLE OF CHANCE: BlastScoringNet was developed using data from 2760 blastocysts with majority-voted grades for expansion degree, ICM, and TE. The model achieved mean area under the receiver operating characteristic curve values of 0.997 (SD 0.004) for expansion degree, 0.903 (SD 0.031) for ICM, and 0.943 (SD 0.040) for TE, based on predicted probabilities for each base grade. From these predicted probabilities, BlastScoringNet generated continuous ICM and TE scores, as well as expansion degree grades, for an additional 15 228 blastocysts with known live birth outcomes. Higher ICM and TE scores, along with their OR-combined scores, were significantly correlated with increased live birth rates ($P < 0.0001$). By fine-tuning, BlastScoringNet was applied to an external IVF institution, where higher OR-combined ICM and TE scores also significantly correlated with increased live birth rates ($P = 0.00078$), demonstrating consistent results across both institutions.

LIMITATIONS, REASONS FOR CAUTION: This study is limited by its retrospective nature. Further prospective randomized trials are required to confirm the clinical impact of BlastScoringNet in assisting embryologists in blastocyst selection.

WIDER IMPLICATIONS OF THE FINDINGS: BlastScoringNet provides an interpretable and quantitative method for evaluating blastocysts, aligned with the widely used Gardner grading system. Higher OR-combined ICM and TE scores, representing each blastocyst's overall potential for live birth, were significantly correlated with increased live birth rates. The model's demonstrated generalizability across two institutions further supports its clinical utility. These findings suggest that BlastScoringNet is a valuable tool for assisting embryologists in selecting blastocysts with the highest potential for live birth. The code and pre-trained models are publicly available to facilitate further research and widespread implementation.

STUDY FUNDING/COMPETING INTEREST(S): This work was supported by the Vector Institute and the Temerty Faculty of Medicine at the University of Toronto, Toronto, Ontario, Canada, via a Clinical AI Integration Grant, and the Natural Science Foundation of Hunan Province of China (2023JJ30714). The authors declare no competing interests.

TRIAL REGISTRATION NUMBER: N/A.

Keywords: blastocyst evaluation / blastocyst selection / embryo evaluation / live birth / interpretable / deep learning / artificial intelligence / in vitro fertilization / IVF

Introduction

Infertility affects one in six people globally, with IVF being the most used treatment (Adamson et al., 2023; World Health Organization, 2023). In IVF, oocytes are fertilized and cultured into multicellular blastocysts. Embryologists then evaluate and select the 'highest quality' blastocyst for transfer to the patient's uterus. For evaluating and selecting blastocysts, the Gardner and Schoolcraft morphological grading system, commonly known as the Gardner grading system, is the most used method in IVF laboratories worldwide (Gardner, 1999; Gardner and Schoolcraft, 1999; Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011). This system assesses three aspects of a blastocyst: expansion degree, inner cell mass (ICM) morphology, and trophectoderm (TE) morphology. The expansion degree has six grades (1–6) based on the volume of the blastocoel cavity and hatching status. For blastocysts with an expansion degree graded as 3–6, the ICM and TE are distinguishable. The ICM is graded as good, fair, or poor based on the number and compactness of cells, and the TE is graded as good, fair, or poor based on the number and cohesiveness of cells.

The Gardner grading system, despite its widespread use, has limited grading scale, i.e. only three grades (good, fair, and poor) for both ICM and TE. Consequently, when patients have multiple blastocysts at expansion degrees of 3 or higher, some of these blastocysts often receive similar or same ICM and TE grades. This similarity poses challenges for embryologists in determining which blastocyst to transfer. Additionally, manual grading of ICM and TE is subjective with intra- and inter-evaluator variability (Storr et al., 2017).

Deep-learning approaches have been developed for blastocyst evaluation. One approach involves using deep-learning models to directly predict IVF outcomes from blastocyst images. These IVF outcomes include blastocyst ploidy (Barnes et al., 2023), pregnancy (VerMilyea et al., 2020; Berntsen et al., 2022), and live birth

(Miyagi et al., 2019; Nagaya and Ukita, 2022; Liu et al., 2023b). Blastocysts can be ranked by the predicted probability of an outcome. Although these models achieved higher prediction accuracies than manual evaluation, their lack of interpretability raises growing concerns (Afnan et al., 2021; Lee et al., 2024). Specifically, these models output a numerical probability for an outcome but cannot provide explanations that embryologists can relate to their knowledge or practice, making it difficult for embryologists to trust or act on the predictions. This lack of interpretability also makes troubleshooting challenging. For instance, a recent clinical trial compared the clinical pregnancy rates of blastocysts selected by such a deep-learning model to those selected by manual evaluation (Illingworth et al., 2024). The model achieved a higher clinical pregnancy rate in fresh embryo transfer cycles (48.1% vs 44.5%, $P = 0.35$) but a significantly lower clinical pregnancy rate in frozen embryo transfer cycles (49.5% vs 61.3%, $P = 0.032$). Due to the model's lack of interpretability, the reason for this disparity remains unclear, making it difficult to troubleshoot and improve the model.

An alternative approach is to use deep-learning models to predict Gardner grades from blastocyst images (Kragh et al., 2019; Liu et al., 2023a), which is inherently interpretable to embryologists. By training on blastocyst images with majority-voted grades from multiple embryologists, these models have achieved comparable or higher grading accuracies than individual embryologists and do not suffer from intra- and inter-evaluator variability. However, these models cannot distinguish between blastocysts with similar or same ICM and TE grades.

This study introduces BlastScoringNet, an interpretable deep-learning-based model that processes blastocyst images to provide expansion degree grade and continuous scores that quantify a blastocyst's ICM morphology and TE morphology, based on the Gardner grading system. The continuous scores for ICM and TE effectively differentiate between blastocysts with similar or same ICM and TE grades. They also demonstrate a strong correlation

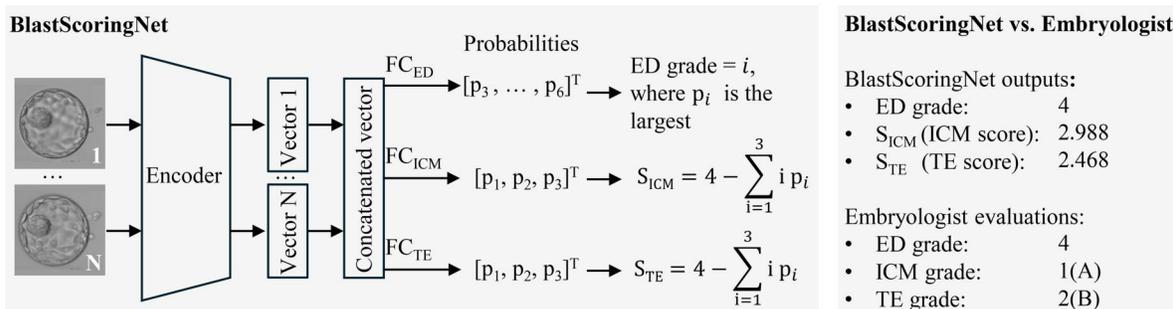


Figure 1. The architecture of BlastScoringNet and comparison with embryologist evaluations of a blastocyst. BlastScoringNet processes multi-focus images of a blastocyst to calculate expansion degree (ED) grade and continuous scores quantifying inner cell mass (ICM) morphology and trophoctoderm (TE) morphology, based on the Gardner grading system (Gardner, 1999; Gardner and Schoolcraft, 1999). A convolutional neural network serves as the encoder, converting the multi-focus images into numerical vectors that encode high-level semantic information for classification. These vectors are concatenated and passed through fully connected layers, followed by softmax layers to predict the probability for each grade in ED, ICM, and TE. The ED grade is determined as i , where p_i is the largest in the vector of probabilities for ED. BlastScoringNet classifies only blastocysts with ED grades of 3–6, as ICM and TE are distinguishable in these blastocysts. The Istanbul consensus workshop provides a numerical scale for the Gardner grading system, assigning a value of 1 for grade A (good), 2 for grade B (fair), and 3 for grade C (poor) for both ICM and TE (Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011). We applied this grading scale when calculating the ICM and TE scores. The ICM score (S_{ICM}) and TE score (S_{TE}) are calculated by summing the predicted probabilities, each weighted by its corresponding grade (1–3 for ICM and TE). These scores enable finer differentiation between blastocysts with similar or same ICM and TE grades. Refer to the ‘Data availability’ section for the GitHub repository containing the source code.

with live birth outcomes and enable more accurate live birth predictions than manual grading.

Materials and methods

Study design and datasets

The study design involves three steps: first, developing the proposed blastocyst scoring model, BlastScoringNet, using the grading dataset, which includes blastocysts with majority-voted Gardner grades; second, applying the model to the live birth dataset, which includes blastocysts with known live birth outcomes, to generate blastocyst scores; and finally, correlating these blastocyst scores with live birth outcomes and comparing this correlation to that between manual grading and live birth. Both the grading and live birth datasets were collected from two IVF institutions: the Reproductive and Genetic Hospital of CITIC-Xiangya (Institution A) for model development and validation, and the Nanjing Drum Tower Hospital (Institution B) to test the model’s generalizability.

From Institution A, the grading dataset included 2760 blastocysts, each annotated by seven embryologists using the Gardner grading system, with majority-voted grades serving as labels to mitigate inter-observer variance. Additionally, a live birth dataset was collected from 15 228 blastocysts with live birth outcomes to assess the correlation between blastocyst scores and live birth. In both datasets, each blastocyst had two images captured under an inverted microscope (Axio Observer A1, Carl Zeiss, Germany; 20× objective lens, 200× image magnification), one focused on ICM and the other on TE. Information on live birth outcomes, Gardner grades, and IVF-related clinical features for these blastocysts was obtained from Institution A’s hospital information system (HIS).

From Institution B, the grading dataset included 1455 blastocysts with majority-voted Gardner grades from three embryologists, and the live birth dataset included an additional 476 blastocysts with live birth outcomes. The blastocysts in both datasets were cultured in a time-lapse incubator (CCM-iBIS, Astec, Japan; 10× objective lens, 320× image magnification), and for each blastocyst, five images were collected at different focal planes at 105 h (SD 3 h) post-fertilization (i.e. Day 5). Data on live birth outcomes, Gardner grades, and clinical features for these blastocysts were also obtained from Institution B’s HIS.

Blastocyst images from all the datasets were cropped to remove the blank background and focus on the blastocysts. The cropped images were then padded to a dimension of 500 × 500 pixels to facilitate model training, which requires images of uniform size. Since the time-lapse incubator at Institution B minimized light exposure during image capture, the blastocyst images were relatively dim. In addition to cropping and padding, a guided filter-based algorithm was used to enhance the brightness of the images (Shi et al., 2018).

Model design and training

BlastScoringNet is a classification model that analyzes blastocyst images to predict probabilities for each base grade in expansion degree, ICM, and TE (Fig. 1). During training, the model was optimized to maximize the probability of the target class while minimizing the probabilities of other classes.

Regarding the model architecture, we incorporated novel designs to fully utilize the 3D morphology captured by multi-focus images and adapt to varying numbers of these images. The model employs a convolutional neural network-based encoder to transform images into numerical vectors encoding high-level semantic information for classification. These vectors are concatenated and passed through fully connected layers and softmax layers to generate class probabilities. This design effectively integrates features from all multi-focus images, enhancing the model’s ability to capture comprehensive 3D morphological details. The shared encoder processes each image, and only the concatenation step adjusts to the number of inputs, enabling the model to seamlessly handle different numbers of multi-focus images.

Since training hyperparameters such as encoder architecture, batch size, and learning rate affect the performance of deep-learning models, to identify the optimal hyperparameters, we utilized Facebook Ax (version 0.2.2), an automatic hyperparameter-tuning tool. For training the model using the grading dataset, we selected a ResNet152 encoder (He et al., 2016), a batch size of 9, and an AdamW optimizer with a learning rate of 4.73e-5 and a weight decay of 0.51. The proposed model was implemented using PyTorch 2.1.0 and Python 3.8.10. Experiments were run on a workstation running CentOS 7.9 with an AMD Ryzen Threadripper 3970X CPU and NVIDIA RTX 6000 Ada GPUs.

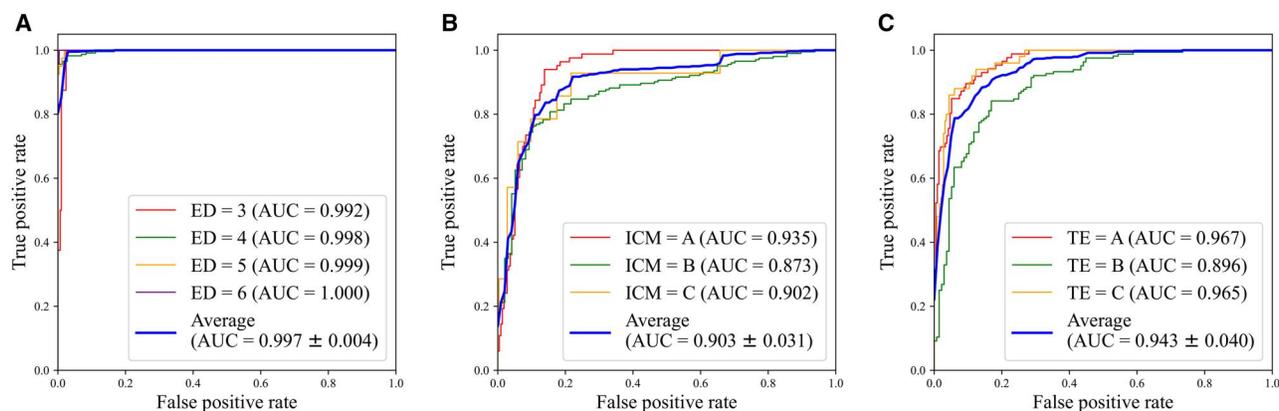


Figure 2. ROC curves for classifying ED (A), ICM (B), and TE (C) using the test dataset, which consists of 300 blastocysts. ROC, receiver operating characteristic; ED, expansion degree; ICM, inner cell mass; TE, trophoctoderm.

Calculation of ICM and TE scores from predicted probabilities

In a classification model, predicted probabilities quantify how closely an object's characteristics align with those typical of each class (Nasrabadi, 2007). A key innovation in our approach lies in how these probabilities are processed for ICM and TE. Existing methods use the highest probability to assign a single grade (Kragh et al., 2019; Liu et al., 2023a). In contrast, our model calculates continuous ICM and TE scores by summing the predicted probabilities, each weighted by its corresponding grade (1–3 for ICM and TE), allowing for finer differentiation between blastocysts with similar or same ICM and TE grades (Fig. 1).

For the ICM and TE scores, according to the Istanbul consensus on blastocyst grading, the three morphological grades—A (good), B (fair), and C (poor)—are assigned base grades of 1, 2, and 3, respectively (Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011). The final ICM or TE score is computed by summing these probability values, each weighted by their corresponding base grades. The calculated score ranges from 1 to 3, with 1 indicating the best quality and 3 indicating the poorest quality. To ensure that a higher score represents better morphological quality, the calculated score is subtracted from four for inverting the scale (Fig. 1).

Note that we did not calculate such a continuous score for the expansion degree. This is because the features that determine the expansion degree are relatively straightforward, allowing the model to classify it with high accuracy (Fig. 2). Consequently, the highest probability is nearly 1 and other probabilities are nearly 0, making the calculated continuous score nearly identical to the actual grade of the expansion degree.

Correlation with live birth

The BlastScoringNet model was subsequently applied to calculate continuous ICM and TE scores, as well as expansion degree grades, for additional blastocysts with known live birth outcomes. We assessed the correlation between ICM and TE scores and live birth outcomes. To represent each blastocyst's overall potential for live birth, we combined the ICM and TE scores using their odds ratios (ORs) for live birth. Since the ORs represent how much the odds of live birth increase with each unit increase in the ICM and TE scores, we calculated their combined effect by multiplying these ORs raised to the power of their respective scores. Specifically, the OR-combined score was calculated according to $(OR_{ICM})^{S_{ICM}} \times (OR_{TE})^{S_{TE}}$, where S_{ICM} and S_{TE} represent the ICM and TE scores, respectively, and OR_{ICM} and OR_{TE} are the

ORs for the live birth associated with each score. Live birth rates were analyzed across quintiles of these scores, and statistical significance was assessed using the Cochran–Armitage trend test. We also compared the predictive accuracy for live birth outcomes between BlastScoringNet and manual grading by developing logistic regression models based on each method separately. For these analyses, blastocysts with known live birth outcomes were stratified by transfer cycle type (frozen embryo transfer cycle or fresh cycle), maternal age (<35 years or ≥35 years), and blastocyst development day (Day 5 or Day 6).

External application

Due to differences in blastocyst imaging conditions (e.g. appearance, magnification, and number of focal planes) and live birth rates across IVF institutions, a blastocyst evaluation model developed at one institution requires fine-tuning for optimal performance when applied externally (Afnan et al., 2021). Thus, this external application study involved fine-tuning BlastScoringNet for application at an external IVF site, Institution B, and validating its correlation with live birth outcomes. We collected two datasets from Institution B: a grading dataset comprising blastocysts with majority-voted Gardner grades and a live birth dataset comprising blastocysts with known live birth outcomes. At Institution B, blastocyst images were captured using a time-lapse incubator (CCM-iBIS, Astec, Japan), which differed in appearance, magnification and were captured at five focal planes per blastocyst instead of two, as in Institution A (Supplementary Figs S1 and S2). This setup was for testing BlastScoringNet's generalizability to these variations.

The fine-tuning process involved adapting BlastScoringNet to handle five multi-focus images of blastocysts, initializing the encoder with pre-trained weights from Institution A, and training the model on Institution B's dataset. To reduce the fine-tuning burden, we evaluated whether the predefined hyperparameters (e.g. learning rate, batch size) from Institution A could be directly applied to train the model at Institution B, and assessed the impact of using pre-trained weights on the model's classification accuracy. Additionally, we assessed the correlations between BlastScoringNet-calculated blastocyst scores and live birth outcomes.

Statistical analysis

The predictive accuracy of BlastScoringNet was assessed using receiver operating characteristic (ROC) curve analysis, with the AUC as a quantitative measure of performance. ROC analysis

was conducted to evaluate BlastScoringNet's ability to distinguish between blastocysts of different grades across expansion degree, ICM, and TE. Logistic regression models were used to predict live birth outcomes. In this study, all AUC values refer

Table 1. Distribution of blastocysts by expansion degree grade, ICM grade, and TE grade in the grading dataset from Institution A.

| Blastocyst grading parameters | Grades | Grading dataset (n=2760 blastocysts) |
|-------------------------------|----------|--------------------------------------|
| Expansion degree | 3 | 85 (3.08%) |
| | 4 | 2125 (76.99%) |
| | 5 | 350 (12.68%) |
| | 6 | 200 (7.25%) |
| ICM grade | A (good) | 862 (31.23%) |
| | B (fair) | 1779 (64.46%) |
| | C (poor) | 119 (4.31%) |
| TE grade | A (good) | 907 (32.86%) |
| | B (fair) | 1372 (49.71%) |
| | C (poor) | 481 (17.43%) |

Data are n (%). ICM, inner cell mass; TE, trophoctoderm.

specifically to the area under the ROC curve. The Cochran-Armitage trend test was performed to assess the monotonic increase in live birth rates across quintiles of ICM scores, TE scores, and their OR-combined scores. A P-value of less than 0.05 was considered statistically significant for all analyses.

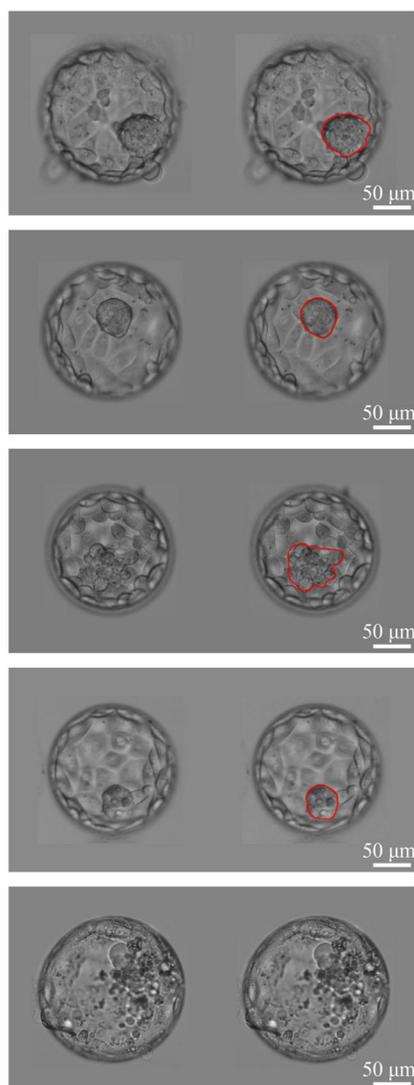
Ethical approval

This study was conducted in accordance with applicable guidelines and regulations, with ethical approval granted by the Ethics Committee at the Reproductive and Genetic Hospital of CITIC-Xiangya (approval number: LL-SC-2021-008) and the Ethics Committee at the Nanjing Drum Tower Hospital (approval number: 2023-171-01).

Results

AUCs of BlastScoringNet in classifying expansion degree, ICM, and TE

BlastScoringNet was developed using 2760 blastocysts, each assigned a majority-voted Gardner grade for expansion degree, ICM, and TE (Table 1). The blastocysts were randomly divided



- **ICM grade:** A (good)
- **ICM score:** 2.9982
- **Probability:** [0.9982, 0.0018, 0.0000]
- **Description:** ICM cells are tightly compacted, occupying a large size.

- **ICM grade:** A (good)
- **ICM score:** 2.4990
- **Probability:** [0.5033, 0.4924, 0.0043]
- **Description:** ICM cells are tightly compacted, occupying a reasonable size.

- **ICM grade:** B (fair)
- **ICM score:** 2.1681
- **Probability:** [0.1761, 0.8159, 0.0080]
- **Description:** ICM is flattened with cells that are loosely grouped together.

- **ICM grade:** B (fair)
- **ICM score:** 1.9748
- **Probability:** [0.0296, 0.9156, 0.0548]
- **Description:** ICM is small and made up of a few cells.

- **ICM grade:** C (poor)
- **ICM score:** 1.0441
- **Probability:** [0.0001, 0.0439, 0.9560]
- **Description:** There is no clearly identifiable ICM.

Figure 3. Example images of blastocysts with varying ICM morphological quality. Each image is accompanied by its BlastScoringNet-calculated ICM score, a vector showing the predicted probabilities for Gardner grades A, B, and C, and qualitative descriptions of the ICM morphology, based on the Gardner criteria. The region where the ICM is located is outlined in red. Refer to the 'Data availability' section for the GitHub repository containing the source code. ICM, inner cell mass. Scale bars show 50 μm.

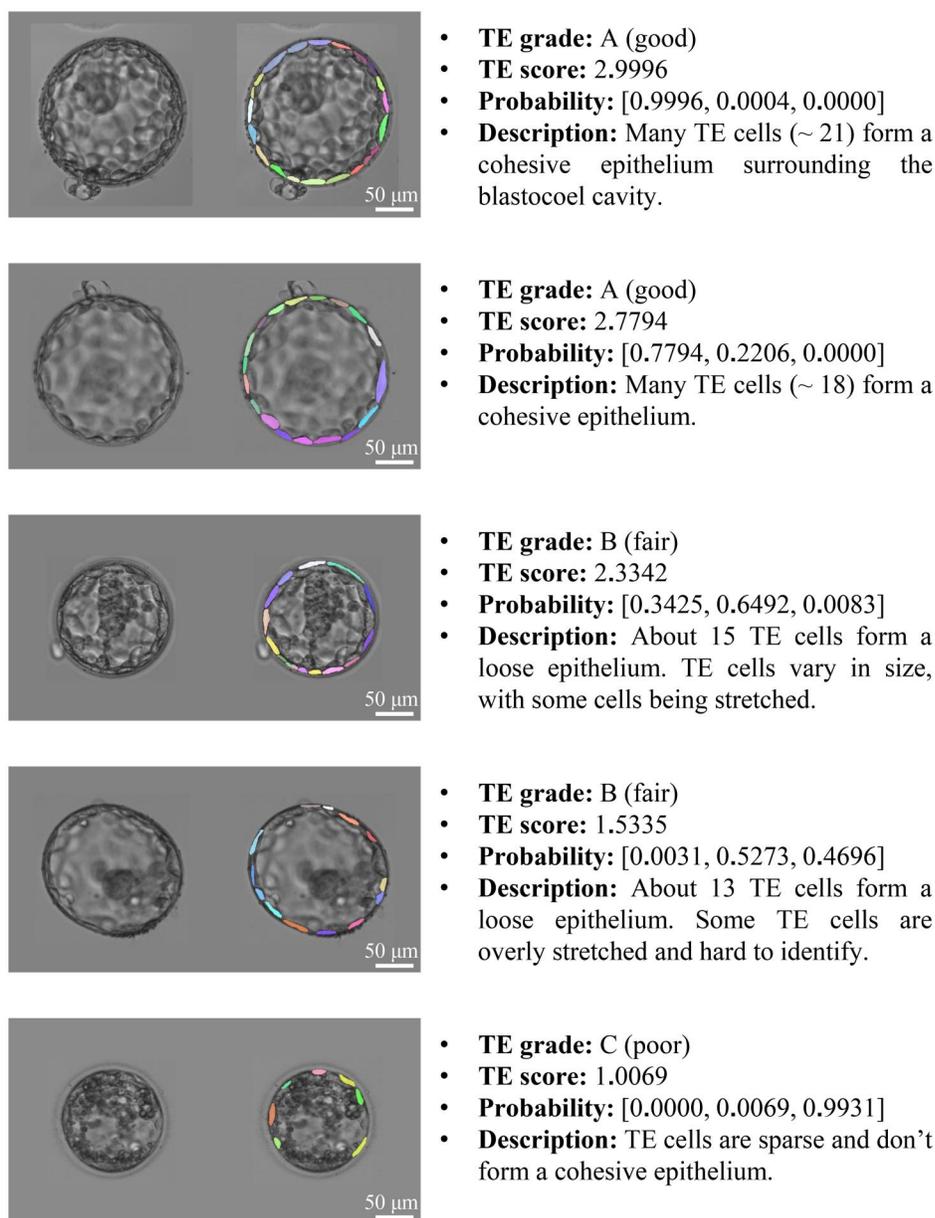


Figure 4. Example images of blastocysts with varying TE morphological quality. Each image is accompanied by its BlastScoringNet-calculated TE score, a vector showing the predicted probabilities for grades A, B, and C, and qualitative descriptions of the TE morphology, based on the Gardner criteria. The regions of individual TE cells that are visible in the outer layer surrounding the blastocoel cavity are highlighted in random colors. Refer to the 'Data availability' section for the GitHub repository containing the source code. TE, trophoctoderm. Scale bars show 50 µm.

into 2160 for training, 300 for validation, and 300 for testing. In classifying expansion degree, the model achieved a mean AUC of 0.997 (SD 0.004). Specifically, the AUCs were 0.992 for grade 3, 0.998 for grade 4, 0.999 for grade 5, and 1.000 for grade 6. These near-perfect AUC values are attributable to the distinct morphological features characterizing each expansion degree; for example, a thinned zona pellucida characterizes grade 4, and hemiating cells characterize grade 5. In classifying ICM, the model achieved a mean AUC of 0.903 (SD 0.031), with individual AUCs of 0.935 for grade A (good), 0.873 for grade B (fair), and 0.902 for grade C (poor). For TE classification, the model achieved a mean AUC of 0.943 (SD 0.040), with individual AUCs of 0.967 for grade A (good), 0.896 for grade B (fair), and 0.965 for grade C (poor) (Fig. 2).

ICM/TE scores and grades

Unlike existing works that only predict ICM and TE grades (Chen et al., 2019; Kragh et al., 2019; Liu et al., 2023b), continuous ICM

and TE scores were generated by summing the products of the predicted probabilities and their corresponding grade values (Fig. 1). These continuous scores enable finer differentiation between blastocysts with similar or same ICM and TE grades following the Gardner criteria (Figs 3 and 4). A higher ICM score indicates better morphological quality of the ICM, such as increased compactness (Fig. 3). Similarly, a higher TE score indicates better morphological quality of the TE, such as greater cell number and cohesiveness of the epithelium (Fig. 4).

Correlation between ICM and TE scores and live birth

To assess the correlation between ICM and TE scores and live birth, BlastScoringNet was then used to calculate ICM and TE scores, as well as the expansion degree grades, for blastocysts with known live birth outcomes. Expansion degree grades were included as a confounding factor in the analysis. A total of 15 228

Table 2. Patient demographic and treatment characteristics by blastocyst transfer cycles in the live birth dataset from Institution A.

| | FET cycle | Fresh cycle |
|---------------------------------------|---------------|-----------------------------|
| Number of blastocyst transfer cycles | 9734 | 2848 |
| Maternal age (years) | 32.20 (4.62) | 30.86 (4.46) |
| BMI (kg/m ²) | 21.88 (2.50) | 21.53 (2.51) |
| Infertility duration (years) | 3.76 (2.88) | 3.97 (2.89) |
| Infertility diagnosis | | |
| Tubal factor | 3621 (37.20%) | 1592 (55.90%) |
| Endometrial factor | 118 (1.21%) | 16 (0.56%) |
| Male factor | 475 (4.88%) | 184 (6.46%) |
| Combined factors | 5259 (54.03%) | 982 (34.48%) |
| Unexplained infertility | 261 (2.68%) | 74 (2.60%) |
| Antral follicle count in both ovaries | 22.46 (13.30) | 22.09 (12.77) |
| Retrieved oocyte number | 14.21 (6.22) | 12.80 (5.39) |
| Endometrium thickness (mm) | 11.87 (1.77) | 12.37 (2.19) |
| Number of transferred blastocysts* | 12 184 | 3044 |
| Maternal age <35 years | 8622 (70.76%) | 2439 (80.12%) |
| Maternal age ≥35 years | 3562 (29.24%) | 605 (19.88%) |
| Day 5 | 9456 (77.61%) | 3044 (100.00%) [†] |
| Day 6 | 2728 (22.39%) | 0 |
| Live birth | 3888 (39.94%) | 1388 (48.74%) |
| Maternal age <35 years | 3117 (45.62%) | 1185 (52.71%) |
| Maternal age ≥35 years | 771 (26.57%) | 203 (33.83%) |
| Day 5 | 3431 (44.56%) | 1388 (48.74%) [†] |
| Day 6 | 457 (22.47%) | 0 |

Data are n, n (%), or mean (SD).

* A maximum of two blastocysts were transferred per cycle. For cycles involving the transfer of two blastocysts, only outcomes of either two live births or non-live births were included. A minor confounder arises if two live births are monozygotic twins derived from a single blastocyst, in which case the second blastocyst would be misclassified as resulting in a live birth. However, the effect is minimal due to the low incidence of such cases (~0.1%).

[†] Only day-5 blastocysts were transferred in fresh cycles to align the embryo's developmental stage with the optimal implantation window. FET, frozen embryo transfer.

blastocysts with known live birth outcomes were included in the analysis (Table 2). Of these, 12 184 blastocysts originated from 9734 frozen embryo transfer (FET) cycles (mean maternal age: 32.20 years [SD 4.62]), while 3044 blastocysts came from 2848 fresh embryo transfer cycles (mean maternal age: 30.86 years [SD 4.46]). Among the blastocysts from FET cycles, 9456 (77.61%) were day-5 blastocysts, and 2728 (22.39%) were day-6 blastocysts. Note that all blastocysts from fresh cycles were day-5 blastocysts.

Higher ICM scores, TE scores, and their OR-combined scores were significantly correlated with increased live birth rates across all stratification groups ($P < 0.05$), including the type of transfer cycle (FET cycle or fresh cycle), maternal age (<35 years or ≥35 years), and the day of blastocyst development (Day 5 or Day 6) (Table 3). Note that when considering only the ICM score or the TE score alone, there were instances where lower-ranked quintiles exhibited higher live birth rates. Specifically, in the stratification of fresh cycles with maternal age ≥35 years, the first quintile (Q1) for the ICM score had a live birth rate of 33.06%, which was higher than that of the second quintile (Q2) at 28.10% and the third quintile (Q3) at 32.23%. For the TE score in the stratification of FET cycles with maternal age ≥35 years and Day 6 blastocysts, the second quintile (Q2) had a live birth rate of 10.63%, slightly higher than the third quintile (Q3) at 10.19%. However, the OR-combined ICM and TE scores did not exhibit these inconsistencies, indicating that combining both scores provides a more comprehensive assessment of blastocyst quality and more accurately reflects their live birth potential. This finding aligns with clinical practice, where both ICM and TE are considered in the evaluation of a blastocyst. Additionally, these blastocyst scores demonstrated higher AUC values in predicting live birth outcomes compared to manual grades across all stratifications (Table 3).

External application study results

By fine-tuning the BlastScoringNet model using hyperparameters and pre-trained encoder weights from Institution A, we successfully

applied it to an external IVF institution. From Institution B, we collected 1455 blastocysts, each assigned a majority-voted Gardner grade for expansion degree, ICM, and TE (Supplementary Table S1). These blastocysts were randomly divided into 1309 for training and validation, and 146 for testing. Fine-tuning enabled BlastScoringNet to achieve similar classification accuracy using only 40% of the training and validation data (AUC 0.876) compared to a baseline model trained from scratch—without predefined hyperparameters or pre-trained weights—using 100% of the data (AUC 0.872). Moreover, when 100% of the training and validation data were used, the fine-tuned model achieved a higher AUC (0.915 vs 0.872) (Supplementary Fig. S3). Additionally, higher OR-combined ICM and TE scores consistently correlated with increased live birth rates ($P = 0.00078$), as assessed using 476 fresh day-5 blastocysts from Institution B (Supplementary Tables S2 and S3). The model's ability to generalize to multiple multi-focus images, rather than only two, streamlines the analysis of time-lapse incubator images by eliminating the need to manually select separate images focusing on the ICM and TE. Furthermore, multi-focus images provide a more comprehensive view of blastocyst morphology for improved scoring accuracy and enhanced live birth prediction (Supplementary Table S4).

Discussion

Previous studies reported quantitative approaches for evaluating blastocyst morphology. Filho *et al.* (2012) investigated automated grading of ICM and TE using traditional image-processing methods; however, the approach achieved limited accuracy on a small dataset. Ahlström *et al.* (2011, 2013) examined correlations between Gardner grades and live birth rates, but the method was not able to differentiate between blastocysts with the same grades. Ebner *et al.* (2016) investigated the correlation between live birth rates and ICM area and TE cell number, yet

Table 3. Correlation between BlastScoringNet outputs and live birth.

| | Live birth rate in FET cycles | | | | | | | | | | Live birth rate in fresh cycles | | | Live birth rate in all cycles (n = 15 228) | |
|----------------------------------|-------------------------------|------------------|----------------|------------------|------------------|------------------------|------------------|------------------|------------------|----------|---------------------------------|-----------------------------------|----------------------------------|--|----------------|
| | Maternal age <35 years | | | | | Maternal age ≥35 years | | | | | All (n = 12 184) | Maternal age <35 years (n = 2439) | Maternal age ≥35 years (n = 605) | | All (n = 3044) |
| | Day 5 (n = 6928) | Day 6 (n = 1694) | All (n = 8622) | Day 5 (n = 2528) | Day 6 (n = 1034) | All (n = 3562) | Day 5 (n = 9456) | Day 6 (n = 2728) | All (n = 12 184) | | | | | | |
| ICM score (S _{ICM}) | Q1 | 40.12% | 9.44% | 31.48% | 21.74% | 5.80% | 15.15% | 34.74% | 7.51% | 23.27% | 43.24% | 33.06% [¶] | 41.38% | 28.50% | |
| | Q2 | 45.56% | 17.70% | 38.52% | 23.76% | 8.70% | 18.26% | 38.97% | 14.68% | 32.95% | 47.34% | 28.10% | 42.53% | 33.40% | |
| | Q3 | 50.65% | 23.96% | 46.06% | 28.66% | 12.14% | 24.16% | 44.95% | 18.86% | 41.05% | 54.00% | 32.23% | 49.01% | 42.09% | |
| | Q4 | 52.92% | 30.68% | 49.77% | 32.87% | 19.81% | 29.21% | 47.70% | 26.61% | 46.98% | 59.22% | 34.71% | 55.67% | 46.44% | |
| | Q5 | 56.75% | 37.87% | 55.34% | 43.17% | 24.27% | 40.45% | 53.99% | 33.58% | 49.38% | 63.04% | 45.00% | 59.70% | 54.15% | |
| | P-value* | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | 0.0264 | <0.0001 | <0.0001 | |
| TE score (S _{TE}) | Q1 | 39.39% | 10.62% | 29.97% | 17.79% | 5.31% | 11.78% | 33.21% | 8.06% | 23.27% | 38.52% | 22.31% | 34.32% | 25.15% | |
| | Q2 | 44.98% | 17.40% | 39.04% | 23.76% | 10.63% [¶] | 18.82% | 38.18% | 15.41% | 32.95% | 48.16% | 26.45% | 44.66% | 35.44% | |
| | Q3 | 51.66% | 23.37% | 47.62% | 28.85% | 10.19% | 25.00% | 46.32% | 16.30% | 41.05% | 55.85% | 35.54% | 50.49% | 43.14% | |
| | Q4 | 54.44% | 32.74% | 51.33% | 38.81% | 17.39% | 34.55% | 50.61% | 28.81% | 46.98% | 61.27% | 39.67% | 57.14% | 49.23% | |
| | Q5 | 55.45% | 35.50% | 53.13% | 40.99% | 27.67% | 37.08% | 51.98% | 32.26% | 49.38% | 63.04% | 49.71% | 61.68% | 51.59% | |
| | P-value | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | |
| OR-combined score [†] | n | 3444 | 852 | 4302 | 1267 | 515 | 1785 | 4738 | 1357 | 6096 | 1227 | 302 | 1527 | 7582 | |
| | Q1 | 37.99% | 8.77% | 29.15% | 17.32% | 4.85% | 9.24% | 31.43% | 8.09% | 21.90% | 38.62% | 27.87% | 34.97% | 26.30% | |
| | Q2 | 43.52% | 20.00% | 38.14% | 27.27% | 6.80% | 19.05% | 36.85% | 13.65% | 31.75% | 46.12% | 33.33% | 44.26% | 35.75% | |
| | Q3 | 51.46% | 25.29% | 48.26% | 30.43% | 10.68% | 22.97% | 44.73% | 19.93% | 38.64% | 50.20% | 35.00% | 49.18% | 41.42% | |
| | Q4 | 56.04% | 30.00% | 51.28% | 37.15% | 19.42% | 31.65% | 51.11% | 27.31% | 48.15% | 63.67% | 45.00% | 54.75% | 51.25% | |
| | Q5 | 58.45% | 46.47% | 54.65% | 47.43% | 23.53% | 43.54% | 54.59% | 37.64% | 52.09% | 66.94% | 51.67% | 63.28% | 55.87% | |
| | P-value | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | |
| OR _{ICM} | | 1.0731 | 1.0933 | 1.4529 | 1.0555 | 3.8350 | 1.1175 | 1.0108 | 1.5952 | 1.0501 | 1.0178 | 1.4985 | 1.1149 | 1.1018 | |
| OR _{TE} | | 1.1217 | 1.1291 | 1.4367 | 1.0782 | 2.0836 | 1.1113 | 1.7115 | 1.0352 | 1.1163 | 1.0334 | 1.0717 | 1.0420 | 1.3362 | |
| AUC [‡] | | 0.5693 | 0.6512 | 0.6065 | 0.6098 | 0.6732 | 0.6493 | 0.5894 | 0.6436 | 0.6037 | 0.5968 | 0.6026 | 0.6005 | 0.6114 | |
| | | (0.0095) | (0.0193) | (0.0054) | (0.0146) | (0.0363) | (0.0130) | (0.0096) | (0.0231) | (0.0519) | (0.0125) | (0.0304) | (0.0135) | (0.0059) | |
| Manual grades of ED, ICM, and TE | AUC | 0.5485 | 0.5919 | 0.5874 | 0.5586 | 0.6318 | 0.6182 | 0.5740 | 0.6075 | 0.5621 | 0.5489 | 0.5634 | 0.5875 | 0.5777 | |
| | | (0.0452) | (0.0619) | (0.0098) | (0.0989) | (0.0275) | (0.0472) | (0.0220) | (0.0171) | (0.0847) | (0.0646) | (0.0446) | (0.0087) | (0.0492) | |

Data are n, n (%), or mean (SD), % data are live birth rate.

* The Cochran-Armitage trend test was applied to assess the monotonic increase in live birth rates across quintiles, with a P-value less than 0.05 considered statistically significant.

† OR represents odds ratio. The OR-combined score values were calculated as $(OR_{ICM})^{S_{ICM}} \times (OR_{TE})^{S_{TE}}$, with OR_{ICM} and OR_{TE} calculated from the logistic regression model predicting live birth using ICM score, TE score, and expansion degree grade, with the latter included to adjust the weight. Half of the blastocysts in each stratification were used to train the model to obtain weight, while the other half were used to calculate live birth rates based on quintiles of the OR-combined score values.

‡ AUC values were calculated from a 10-run process, during which logistic regression models were built to predict the live birth. In each run, the dataset in each stratification was split into two halves: one for training and the other for testing. Before each run, the dataset was shuffled with a different random state. The model with the best validation accuracy was then used to derive the ORs and calculate the OR-combined score values using the corresponding test dataset. The same training and testing datasets were also used to train logistic regression models using manual grades to predict live birth and calculate the corresponding AUC values.

¶ Values that belonged to a lower-ranked quintile but had a higher live birth rate were marked. Q1-Q5, the first through fifth quintiles; OR, odds ratio; FET, frozen embryo transfer; ICM, inner cell mass; TE, trophectoderm.

the metrics did not fully capture the complexity of ICM and TE morphology. In contrast, the continuous ICM and TE scores calculated by BlastScoringNet overcome the limited grading scale in the Gardner grading system and enable finer differentiation between blastocysts with similar or same grades. Higher ICM and TE scores, as well as their OR-combined scores, were significantly correlated with increased live birth rates, demonstrating the model's ability to rank blastocysts based on their live birth potential. Furthermore, the generalizability of the proposed model was confirmed through application in an external IVF institution. These results suggest that BlastScoringNet is a valuable and generalizable tool for assisting embryologists in selecting blastocysts with the highest likelihood of resulting in a successful birth.

One of BlastScoringNet's strengths in blastocyst selection is its interpretability. This arises from its ICM and TE scores, which quantify the morphological quality of the ICM and TE consistent with the Gardner criteria. These scores are combined using their ORs for the live birth to evaluate and rank blastocysts. In clinical applications, BlastScoringNet's evaluation and ranking process is understandable to embryologists, thereby avoiding well-recognized concerns associated with black-box models, such as reduced trust, challenges in real-time error-checking and troubleshooting, safety issues, and ambiguity regarding responsibility for blastocyst selection decisions (Afnan et al., 2021).

Another strength of BlastScoringNet is its ability to consistently rank blastocysts based on their live birth potential. To select the blastocyst with the highest live birth potential from a cohort, machine-learning models must evaluate and rank each blastocyst to identify the best one. This requires a model to provide consistent evaluations that differentiate the relative live birth potential of each blastocyst. However, previous live birth prediction models lacked the ability to perform this critical ranking function (Miyagi et al., 2019; Nagaya and Ukita, 2022; Liu et al., 2023b). In our work, we observed a statistically significant and consistent relationship between higher OR-combined ICM and TE scores and increased live birth rates, as confirmed by the Cochran–Armitage trend test ($P < 0.0001$, Table 3). This finding was further validated in different subgroups of blastocysts stratified by transfer cycle type (FET cycle or fresh cycle), maternal age (< 35 years or ≥ 35 years), and blastocyst development day (Day 5 or Day 6) ($P < 0.0001$, Table 3).

Finally, we validated the generalizability of BlastScoringNet by fine-tuning it for use in an external IVF institution, accounting for common variations that exist across IVF institutions worldwide, such as differences in blastocyst image conditions (e.g. appearance, magnification, and number of focal planes), IVF success rates, and embryologists' experience levels. Given these variations, blastocyst evaluation tools developed using one dataset should be fine-tuned to ensure optimal performance when applied to external institutions. For example, iDAScore, a deep-learning model developed by Vitrolife to predict pregnancy from blastocyst images using data from multiple clinics, exhibited large variations in accuracy when generalized to new clinics in their clinic hold-out test (e.g. AUC ranging from 0.60 to 0.75) (Bermtsen et al., 2022). To facilitate the fine-tuning process, we have made our hyperparameters (e.g. encoder architecture, batch size, and learning rate) and the pre-trained model publicly accessible. Providing these hyperparameters reduces the time needed to search for optimal hyperparameters on a new dataset. Additionally, the pre-trained model helps improve model accuracy compared to training from scratch, as confirmed in our external application study (Supplementary Fig. S3).

A few limitations should be considered. First, to apply BlastScoringNet externally, ~1000 blastocyst images with majority-voted grades are currently needed to achieve high classification accuracy (e.g. AUC > 0.90 , Supplementary Fig. S3). Constructing a comprehensive dataset of blastocyst images captured from a number of imaging tools that are commonly used in IVF centers (e.g. Carl Zeiss, Nikon, Olympus, EmbryoScope) could help mitigate the fine-tuning effort. Training BlastScoringNet on such a comprehensive dataset could improve its generalizability, ultimately reducing or eliminating the need for fine-tuning. Second, the live birth prediction accuracy of BlastScoringNet-calculated blastocyst scores has room for further improvement, with AUCs of 0.57–0.67 across different stratifications in the present study. These AUC values are similar to those by black-box models that directly predict live birth outcomes from blastocyst images (Miyagi et al., 2019; Nagaya and Ukita, 2022; Liu et al., 2023b). Besides the morphological quality of blastocysts revealed in static images, incorporating other parameters such as spontaneous blastocyst collapse (Bickendorf et al., 2023; Zhu et al., 2024) and timings to specific stages (e.g. reaching the 6-cell or morula stage) (Bamford et al., 2023; Ten et al., 2024) may further improve the prediction accuracy. Third, the correlation between BlastScoringNet-calculated blastocyst scores and live birth outcomes was validated in this study using datasets from two IVF institutions both in China. Further validation in more diverse ethnic and racial groups could be insightful. Finally, as this study was based on retrospective data, the effectiveness of BlastScoringNet in improving IVF outcomes requires confirmation through prospective randomized controlled trials, which is the next step in our work.

In conclusion, BlastScoringNet is an interpretable and generalizable tool that aligns with the established clinical standard. The strong correlation between BlastScoringNet-calculated blastocyst scores and live birth outcomes indicates its potential to assist embryologists in selecting blastocysts with higher live birth potential. Additionally, BlastScoringNet may serve as a valuable tool for assessing blastocyst morphology in relation to other IVF outcomes, such as blastocyst ploidy, gestational age, birth weight and sex, and perinatal outcomes.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The grading and live birth datasets analyzed in this study are not publicly available, owing to reasonable privacy and security concerns. Datasets from Institution A are available from the corresponding authors Y.S., G.L., and Y.F.G. on reasonable request. Datasets from Institution B are available from the corresponding authors Y.S., H.X.S., and S.S.W. on reasonable request. The source code and pre-trained model are publicly available on GitHub (<https://github.com/robotVisionHang/BlastScoringNet>).

Authors' roles

H.L., L.B.C., S.S.W., H.X.S., G.L., Y.S., and Y.F.G. conceptualized and designed the study. L.B.C., H.Q.L., C.F.L., S.P.Z., S.N.D., X.X.X., Q.Y.Y., F.G., G.L., and Y.F.G. provided the grading dataset and the live birth dataset from Institution A. L.C., S.S.W., and H.X.S. provided the grading dataset and the live birth dataset from Institution B. H.L., G.Q.S., C.S., and Y.S. developed the

BlastScoringNet model for Institutions A and B. W.Y.C. and H.C.S. contributed to model training and assessment. H.L., L.B.C., Y.S., and Y.F.G. performed the statistical analysis on live birth dataset from Institution A. H.L., L.C., Y.S., S.S.W., and H.X.S. performed the statistical analysis on live birth dataset from Institution B. Z.R.Z. and C.S.D. verified the raw data and findings. The manuscript was written primarily by H.L., Y.S., and Y.F.G., with support from all authors. S.S.W., H.X.S., G.L., Y.S., and Y.F.G. supervised this study. All authors verified and had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Funding

Vector Institute and Temerty Faculty of Medicine, University of Toronto (via a Clinical AI Integration Grant); Natural Science Foundation of Hunan Province of China (2023JJ30714).

Conflict of interest

None declared.

References

- Adamson GD, Zegers-Hochschild F, Dyer S. Global fertility care with assisted reproductive technology. *Fertil Steril* 2023;**120**:473–482.
- Afnan MAM, Liu Y, Conitzer V, Rudin C, Mishra A, Savulescu J, Afnan M. Interpretable, not black-box, artificial intelligence should be used for embryo selection. *Hum Reprod Open* 2021;**2021**:hoab040.
- Ahlström A, Westin C, Reismar E, Wikland M, Hardarson T. Trophectoderm morphology: an important parameter for predicting live birth after single blastocyst transfer. *Hum Reprod* 2011;**26**:3289–3296.
- Ahlström 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.
- 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* 2011;**26**:1270–1283.
- Bamford T, Smith R, Easter C, Dhillon-Smith R, Barrie A, Montgomery S, Campbell A, Coomarasamy A. Association between a morphokinetic ploidy prediction model risk score and miscarriage and live birth: a multicentre cohort study. *Fertil Steril* 2023;**120**:834–843.
- Barnes J, Brendel M, Gao VR, Rajendran S, Kim J, Li Q, Malmsten JE, Sierra JT, 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.
- Berntsen J, Rimestad J, Lassen JT, Tran D, Kragh MF. Robust and generalizable embryo selection based on artificial intelligence and time-lapse image sequences. *PLoS One* 2022;**17**:e0262661.
- Bickendorf K, Qi F, Peirce K, Natalwala J, Chapple V, Liu Y. Spontaneous collapse as a prognostic marker for human blastocysts: a systematic review and meta-analysis. *Hum Reprod* 2023;**38**:1891–1900.
- Chen T-J, Zheng W-L, Liu C-H, Huang I, Lai H-H, Liu M. Using deep learning with large dataset of microscope images to develop an automated embryo grading system. *Fertil Reprod* 2019;**1**:51–56.
- Ebner T, Tritscher K, Mayer RB, Oppelt P, Duba HC, Maurer M, Schappacher-Tilp G, Petek E, Shebl O. Quantitative and qualitative trophectoderm grading allows for prediction of live birth and gender. *J Assist Reprod Genet* 2016;**33**:49–57.
- Filho ES, Noble JA, Poli M, Griffiths T, Emerson G, Wells D. A method for semi-automatic grading of human blastocyst microscope images. *Hum Reprod* 2012;**27**:2641–2648.
- Gardner DK, Schoolcraft WB. In vitro culture of human blastocysts. In: Jansen R, Mortimer D (eds). *Toward Reproductive Certainty: Fertility and Genetics Beyond* 1999. London: Parthenon Publishing, 1999,378–388.
- Gardner DK, Schoolcraft WB. Culture and transfer of human blastocysts. *Curr Opin Obstet Gynecol* 1999;**11**:307–311.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. New York, NY, USA: Computer Vision Foundation, 2016, 770–778.
- Illingworth PJ, Venetis C, Gardner DK, Nelson SM, Berntsen J, Larman MG, Agresta F, Ahitan S, Ahlström A, Cattrall F et al. Deep learning versus manual morphology-based embryo selection in IVF: a randomized, double-blind noninferiority trial. *Nat Med* 2024;**30**:3114–3120.
- Kragh MF, Rimestad J, Berntsen J, Karstoft H. Automatic grading of human blastocysts from time-lapse imaging. *Comput Biol Med* 2019;**115**:103494.
- Lee T, Natalwala J, Chapple V, Liu Y. A brief history of artificial intelligence embryo selection: from black-box to glass-box. *Hum Reprod* 2024;**39**:285–292.
- Liu H, Li D, Dai C, Shan G, Zhang Z, Zhuang S, Lee C-W, Wong A, Yue C, Huang Z et al. Automated morphological grading of human blastocysts from multi-focus images. *IEEE Transactions on Automation Science and Engineering*. New York, NY, USA: IEEE, 2023a, 1–9.
- Liu H, Zhang Z, Gu Y, Dai C, Shan G, Song H, Li D, Chen W, Lin G, Sun Y. Development and evaluation of a live birth prediction model for evaluating human blastocysts from a retrospective study. *Elife* 2023b;**12**:e83662.
- Miyagi Y, Habara T, Hirata R, Hayashi N. Feasibility of artificial intelligence for predicting live birth without aneuploidy from a blastocyst image. *Reprod Med Biol* 2019;**18**:204–211.
- Nagaya M, Ukita N. Embryo grading with unreliable labels due to chromosome abnormalities by regularized PU learning with ranking. *IEEE Trans Med Imaging* 2022;**41**:320–331.
- Nasrabadi NM. Pattern recognition and machine learning. *J Electron Imaging* 2007;**16**:049901.
- Shi Z, Zhu MM, Guo B, Zhao M, Zhang C. Nighttime low illumination image enhancement with single image using bright/dark channel prior. *EURASIP J Image Video Process* 2018;**2018**:13.
- 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* 2017;**32**:307–314.
- Ten J, Tió M, Linares Á, Zepeda A, Pini P, Brualla A, Díaz N, Herreros M, Bernabeu A, Bernabeu R. The role of artificial intelligence (AI) in predicting live birth using embryo morphokinetic parameters. *Reprod BioMedicine Online* 2024;**48**:104047.
- VerMilyea M, Hall JMM, Diakiw SM, Johnston A, Nguyen T, Perugini D, Miller A, Picou A, Murphy AP, Perugini M. 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* 2020;**35**:770–784.
- World Health Organization. 2023. *Infertility prevalence estimates, 1990–2021*. Geneva, Switzerland: World Health Organization.
- Zhu J, Zou J, Wu L, Xiong S, Gao Y, Liu J, Huang G, Han W. Total duration of spontaneous blastocyst collapse during the expansion stage is an independent predictor of euploidy and live birth rates. *Reprod BioMedicine Online* 2024;**49**:103863.