

Enhancing Predictive Accuracy in Embryo Implantation: The Bonna Algorithm and its Clinical Implications

Gilad Rave[1], Daniel E. Fordham[1], Alex M. Bronstein[2], and David H. Silver[1(✉)]

[1] Rhea Labs, Rhea Fertility, Singapore, Singapore
{gilad.rave,daniel.fordham,david.silver}@rhea-fertility.com
[2] Department of Computer Science, Technion-Israel Institute of Technology, Haifa, Israel
bron@cs.technion.ac.il

Abstract. In the context of in vitro fertilization (IVF), selecting embryos for transfer is critical in determining pregnancy outcomes, with implantation as the essential first milestone for a successful pregnancy. This study introduces the Bonna algorithm, an advanced deep-learning framework engineered to predict embryo implantation probabilities. The algorithm employs a sophisticated integration of machine-learning techniques, utilizing MobileNetV2 for pixel and context embedding, a custom Pix2Pix model for precise segmentation, and a Vision Transformer for additional depth in embedding. MobileNetV2 was chosen for its robust feature extraction capabilities, focusing on textures and edges. The custom Pix2Pix model is adapted for precise segmentation of significant biological features such as the zona pellucida and blastocyst cavity. The Vision Transformer adds a global perspective, capturing complex patterns not apparent in local image segments. Tested on a dataset of images of human blastocysts collected from Ukraine, Israel, and Spain, the Bonna algorithm was rigorously validated through 10-fold cross-validation to ensure its robustness and reliability. It demonstrates superior performance with a mean area under the receiver operating characteristic curve (AUC) of 0.754, significantly outperforming existing models. The study not only advances predictive accuracy in embryo selection but also highlights the algorithm's clinical applicability due to reliable confidence reporting.

Keywords: Embryo Implantation · Deep Learning · Predictive Modeling · Clinical Decision Support · Artificial Intelligence in Reproductive Medicine

1 Introduction

The accurate prediction of embryo implantation probability represents a central challenge in assisted reproductive technology (ART) [1]. As it is the first measurable outcome of success in in-vitro fertilization after embryo transfer, it

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
X. Xie et al. (Eds.): AIiH 2024, LNCS 14976, pp. 160–171, 2024.
https://doi.org/10.1007/978-3-031-67285-9_12

serves as a proxy measure for the effectiveness of embryo selection [2]. Existing methodologies, while diverse, often fall short in delivering the specificity, objectivity, and reliability required for effective clinical application [3,4]. These range from basic morphological evaluations to intricate predictive modeling that incorporates various biomarkers [5–8]. Techniques such as time-lapse imaging paired with morphokinetic algorithms, although promising, have either not been proven clinically [6] or consistently fail to achieve the necessary precision for clinical settings [5]. Furthermore, while preimplantation genetic testing (PGT) offers extensive data, it also introduces significant ethical and logistical challenges [9].

Addressing this, the Bonna algorithm utilizes a comprehensive suite of deep learning techniques, including the MobileNetV2 convolutional network [10] for efficient image processing, a custom Pix2Pix generative adversarial network [11] for precise embryonic feature segmentation, and Vision Transformers [12] for detailed image analysis. Not only does this amalgamation enhance the algorithm's capability to identify subtle patterns indicative of viability but it also significantly boosts predictive accuracy and reliability beyond traditional methods. Additionally, the Bonna algorithm employs methods to provide confidence scores that correlate with the model's uncertainty, further enhancing its clinical utility in real-life settings [13].

This study rigorously evaluates the Bonna algorithm across datasets from Israel, Spain, Ukraine, and Spain, ensuring robust validation of its capabilities. Recent studies [14] have similarly leveraged multi-center data, underscoring the need for adaptive algorithms due to variations in embryo quality and implantation outcomes. The design of the Bonna algorithm not only addresses these variations but also offers a scalable solution adaptable to diverse clinical environments by using segmentation module (Pix2Pix) to remove optical factors.

By using an ensemble of several machine learning models [15,16], the algorithm capitalizes on their individual strengths in comprehensive image processing and segmentation. This approach provides a nuanced analysis of embryo viability, achieving state-of-the-art predictive performance in reproductive medicine. Ultimately, the Bonna algorithm aims to furnish clinicians with a reliable and personalized decision-support tool, designed to optimize IVF outcomes by selecting embryos with the highest potential for successful implantation.

2 Methodology

2.1 Dataset Collection and Preparation

The dataset for this retrospective cohort study was meticulously assembled from private clinics across three distinct geographical locations: Israel, Spain, Ukraine and France. This strategic selection captures a diverse embryological dataset, which is essential for generalizing model training. The data comprises time-lapse incubator (TLI) video files, specifically from the EmbryoScope and EmbryoScope+ (Vitrolife, Aarhus, Denmark) systems [17], with a resolution of 500 by 500 pixels. For this study, a single frame representing the blastocyst stage—Day 5 of incubation, was extracted from each TLI file at the point of

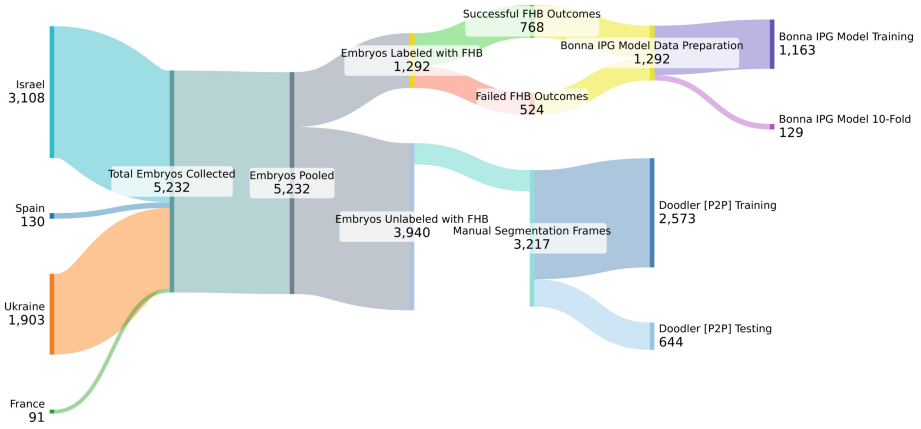


Fig. 1. Sankey diagram illustrating the comprehensive data flow of embryos from initial collection at various international sites through to the final stages of modeling. This includes steps such as categorization by Fetal Heartbeat (FHB) status, division into training and testing datasets, and processing through the Bonna IPG and DoodlerP2P (our custom Pix2Pix) models, highlighting the data preparation and utilization strategy employed.

embryo transfer. The dataset includes a total of 5,141 embryos derived from numerous patients who underwent ovarian stimulation cycles between 2012 and 2019. The demographic and embryological data specifics are as follows: Israel (3,108 embryos), Spain (130), Ukraine: (1,903), and France (91). This dataset was carefully stratified into categories based on the critical inclusion criterion of known implantation data (KID), defined here as the ultrasound evidence of a fetal heartbeat (FHB) [18] between 5 to 8 weeks post-transfer. This marker is pivotal not only for assessing embryo viability but also as a definitive test for implantation success, thereby indicating pregnancy [19].

2.2 Data Processing and Flow

Labeling and Stratification. A total of 1,292 embryos had Known Implantation Data (KID), with 741 embryos (57%) with positive Fetal Heartbeat (FHB) outcomes (successful implantation), and 551 embryos (43%) with negative Fetal Heartbeat (FHB) outcomes (failed implantation) (Fig. 1).

Demographics Summary of KID Embryos. Due to data de-identification, the exact number of patients in the study was unknown. Ethnicity, the reason for IVF, obstetric history, past medical history, or other medical classifiers were not available. 1,138 (88%) of the 1,292 embryos had associated maternal age at the time of embryo collection, ranging from 20 to 47 years (mean 34 years). These demographics were noted for contextual background and were not directly used in the model training or at inference time.

Model Preparation and Training. The embryos were segregated into successful and unsuccessful groups based on FHB outcomes for targeted Bonna IPG (Implantation Probability Grade) model training. From the remaining 3,851 embryos that did not have FHB labeling, we selected 3,217 frames at random, from different development stages, for the manual segmentation training essential for the DoodleP2P model's initial training phase. This structured and detailed preparation of the dataset ensured that both the Bonna IPG and DoodleP2P models were trained on robust and highly representative data, thereby enhancing the reliability and accuracy of the implantation predictions.

2.3 Model Architecture and Implementation.

The Bonna algorithm is based on the power of multiple advanced deep-learning models to optimize the prediction of embryo implantation outcomes. This section outlines the specific architecture and implementation details.

Architecture. As schematically summarized in Fig. 2, Bonna includes three primary components, each chosen for its specific capabilities in image analysis and feature extraction:

MobileNetV2 [10]: Known for its efficient parameterization, MobileNetV2 was selected primarily for its robust capability in capturing contextual embeddings from embryo images. This model processes the input images to extract lightweight yet informative features that include textures and edges, which are critical for the initial stages of image analysis.

DoodleP2P (customized Pix2Pix) [11]: This generative adversarial network (GAN) is used for segmenting the embryo images. It isolates significant biological features such as the zona pellucida and blastocyst cavity, which are essential for assessing embryo quality. The output of the DoodleP2P model is a three-channel mask, with each channel representing different biological parts: the zona pellucida, the cells, and the blastocyst cavity. These channels provide a pseudo-coloring of the embryo, helping to highlight and differentiate critical structures.

Vision Transformer (ViT) [12]: The Vision Transformer complements the feature set by adding a global perspective to the analysis [20]. It processes the segmented images to capture complex patterns that are not readily apparent in local image segments, such as spatial relationships and integrative morphological features, which are important for a comprehensive evaluation.

Implementation. The implementation of the Bonna algorithm involved several important steps designed to maximize the predictive performance.

Pre-processing. comprised normalization of all input images to standard scales to reduce model sensitivity to light and color variations [21], and augmentation with rotation, scaling, and flipping employed to enrich the dataset and ensure that the model is robust to various orientations and sizes of embryos [22].

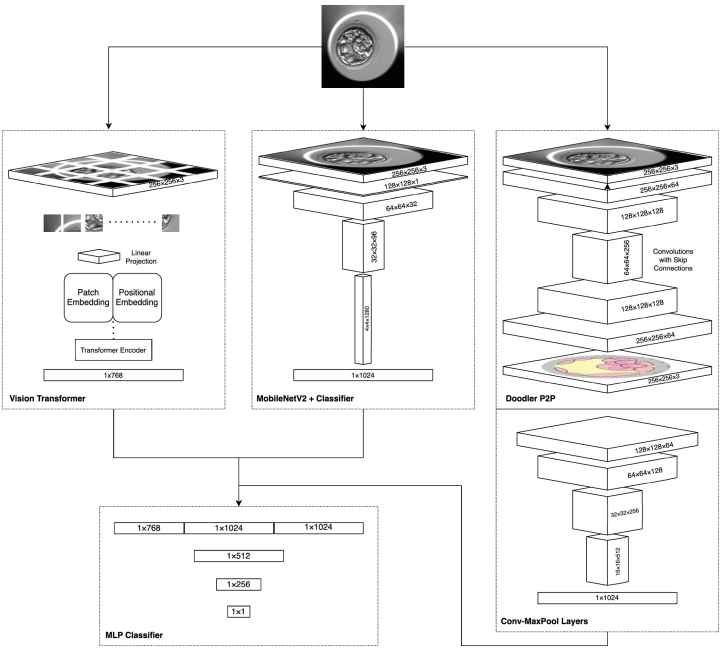


Fig. 2. Flowchart of the Bonna algorithm detailing the orchestrated integration of MobileNetV2 for feature extraction, DoodlerP2P for precise embryo segmentation, and Vision Transformer for enhanced pattern recognition. This schematic illustrates how these models synergistically contribute to the robust prediction of embryo implantation probabilities.

Feature extraction comprised MobileNetV2 and ViT features in tandem concatenated into a feature vector containing both local contextual as well as global morphological characteristics.

Segmentation and embedding relied on the DoodleP2P model to segment out critical areas within the embryo images [23]. The three-channel mask output by DoodleP2P, representing different biological parts, was used to train a custom convolutional neural network (CNN) that further refines the embeddings by focusing on these key features.

Classification: The final prediction was made using a multi-layer perceptron (MLP) classifier [24], which integrates the embeddings into its architecture to classify the likelihood of implantation. The classifier categorizes embryos into two primary outcomes: 'Likely to implant' and 'Unlikely to implant'. These categories are determined based on the presence or absence of a fetal heartbeat (FHB), which is the clinical endpoint of interest in this study. The MLP was trained using the standard cross-entropy loss, which measures the performance of the classification model whose output is a probability value between 0 and 1.

Inference Phase and Prediction Stability. The inference phase is essential for estimating the posterior probability of the Implantation Probability Grade (IPG) based on the images, which is critical to the Bonna algorithm's effectiveness. To enhance the robustness and reliability of predictions, the algorithm employs a "jitter" process involving multiple simulations and augmentations. This approach helps mitigate variability and stabilize the prediction outcomes.

At inference time, each embryo image first undergoes augmentation by 16 random transformations including rotations, translations, scaling, and flips. These augmentations accommodate different positional and morphological nuances of embryos as encountered in clinical settings [25].

Each of the 16 augmented images is then processed with 16 dropout simulations. This regularization technique mimics the effect of model averaging and helps stabilize the predictions by randomly omitting subsets of features or neurons [26].

This "jitter" process generates a total of 256 inputs per embryo, using which the predictive score distribution is obtained.

The standard deviation (STD) of the predicted score is used to assess the prediction confidence, quantified using the negative logarithm of the standard deviation, $-\log(\text{STD})$, with higher values denoting greater prediction confidence.

The empirical confidence intervals calculated from the score distribution are directly linked to the algorithm's performance, which we quantitatively assessed using the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) [27]. In the scenario of embryo selection, which fundamentally involves ranking each embryo in a cohort based on likelihood of success, the AUC is an invaluable metric. It provides a statistical measure of the model's ability to accurately rank embryos, estimating the probability that a viable embryo (positive instance) is scored higher than a non-viable one (negative instance). This capability to correctly prioritize embryos is critical for effective selection and significantly influences clinical outcomes. High AUC values indicate superior discrimination ability of the model, pivotal in clinical settings where the implications of false positives or negatives are substantial. Thus, by minimizing score variability and maximizing prediction confidence, the Bonna algorithm not only achieves high precision but also ensures dependable consistency across its predictions, reinforcing its suitability for high-stakes clinical decision-making.

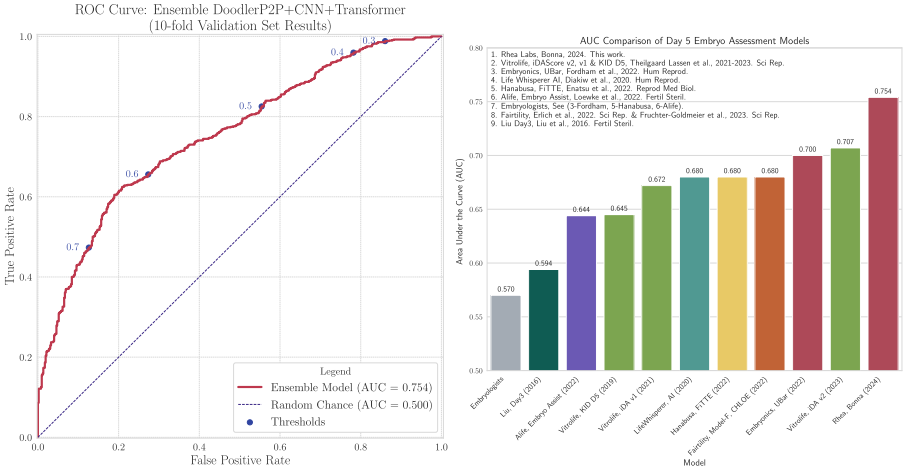
3 Results

This section summarizes the tools and methodologies employed to evaluate the performance of the Bonna algorithm, focusing on comparative analysis with existing models through Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) metrics.

3.1 Model Performance Evaluation

The performance of the Bonna algorithm was systematically evaluated against a range of existing models, with a particular focus on those utilized for the embryo

implantation prediction task. The evaluation criteria centered around the AUC metric, which provides an aggregate measure of performance across all possible classification thresholds. The ROC curve and AUC comparison chart highlight the effectiveness and superiority of our model.



(a) Bonna algorithm ROC curve (b) AUC score comparison

Fig. 3. (a) ROC plot depicting the sensitivity (true positive rate) and specificity (1−false positive rate) of the Bonna algorithm's predictions, providing insights into its diagnostic ability. (b) Mean AUC scores of the Bonna algorithm compared with those attained by contemporary embryo selection models [6,14,28–32], underscoring its superior predictive capabilities.

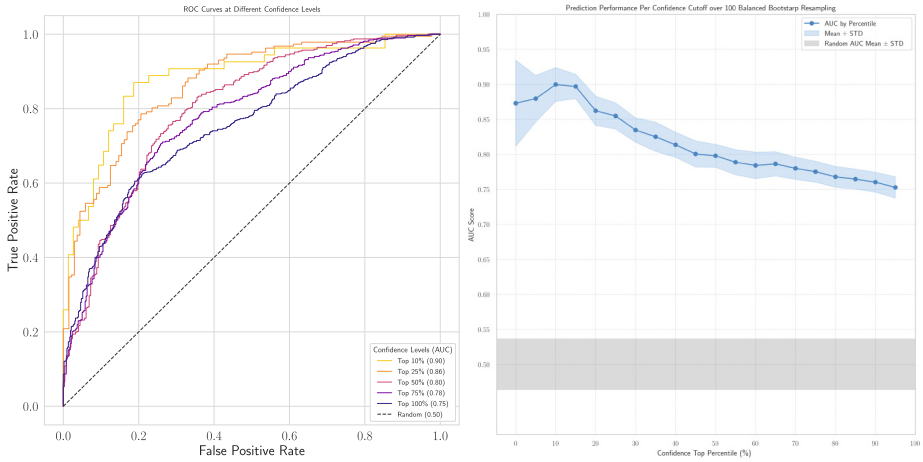
The ROC curve (Fig. 3) plots the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings.

The Area Under the ROC Curve (AUC) score provides a useful quantitative measure of the sensitivity-specificity tradeoff. An AUC of 1 represents perfect prediction, while an AUC of 0.5 suggests no discriminative power on a balanced test set. The Bonna algorithm's AUC of 0.754 reflects high competency in distinguishing between classes (successful vs. unsuccessful implantation), far exceeding the typical benchmarks set by existing models.

3.2 Confidence and Predictive Accuracy

A standout result from our study is the pronounced correlation between prediction confidence and AUC scores, which notably underscores the Bonna algorithm's precision. Our analysis involves setting various confidence cutoff percentiles and observing their impact on the AUC scores. The results are compelling, particularly at higher confidence levels.

The framework used in the Bonna algorithm allows us to set empirical confidence thresholds, which helps categorize predictions according to their likelihood of accuracy. By increasing the confidence cutoff percentile, we selectively evaluate predictions that the model deems to be more certain by binning the confidence score into quantile levels, as depicted in Fig. 4.



(a) ROC Curves for different confidence threshold percentiles

(b) AUC with Confidence Intervals

Fig. 4. Graph of AUC scores as a function of confidence percentile cutoffs for the Bonna algorithm. (a) Demonstrates exceptional predictive accuracy particularly at higher confidence levels, with AUC peaking at 0.9 at the 10th percentile. (b) The trend suggests that the higher the confidence percentile cutoff, the more reliable the predictions. The shaded areas are confidence intervals from bootstrapped sampling of the dataset 100 times.

This strong correlation between prediction confidence and AUC scores not only validates the Bonna algorithm's robust predictive capabilities but also enhances its clinical applicability. By implementing a model with high precision at elevated confidence levels, clinicians can make more informed decisions, potentially leading to improved IVF outcomes. This reliability is crucial for clinical settings where decision accuracy is paramount, and the Bonna algorithm stands out as a particularly dependable tool in such high-stakes environments.

4 Discussion and Conclusion

The Bonna algorithm has established a profound correlation between high confidence predictions and AUC scores, achieving a peak AUC of 0.9 at the 10th and 15th confidence percentiles. These findings are of considerable clinical relevance.

The algorithm's capability to deliver accurate predictions on embryo implantation based on high-confidence intervals and acknowledge cases in which the prediction is less accurate offers a reliable tool for clinicians practicing assisted reproductive technology (ART). This advancement facilitates enhanced personalization in treatment strategies, potentially optimizing patient outcomes and improving resource management.

Study Highlights. This study's methodology integrates a combination of deep learning techniques, including MobileNetV2 for efficient image processing, the customized DoodleP2P for precise feature segmentation, and a Vision Transformer for advanced pattern recognition. This integration achieves state-of-the-art predictive accuracy. The rigorous validation of the algorithm, involving 10-fold cross-validation and robust testing procedures such as image augmentation and dropout at the inference stage, further underscores the model's reliability and robustness.

Comparison to Competing Models. The performance of the Bonna algorithm was compared with several key models in the domain. Firstly, although not tested on identical datasets, the predictive performance of the Bonna algorithm demonstrates notable superiority over traditional embryologists' assessments and when compared to the iDAScore v2 [14], which is recognized as best-in-class for this task.

It is worth noting that although some methodologies claim very high AUC scores [5,33] of 0.93 & 0.98 respectively, they suffer from methodological flaws including non-transferred embryos as negatives or using human scores as ground-truth labels. These practices not only introduce selection bias but also inaccurately inflate performance metrics. Such high AUC scores are generally achieved in simpler tasks and do not reflect the complex nature of predicting successful embryo implantation, which also depends significantly on transfer skill and maternal factors.

Clinical Implications of the Confidence Score. Providing confidence scores significantly enhances the Bonna algorithm's reliability and confidence in clinical predictions of embryo implantation. This method uses the variability among 256 outcomes per embryo, derived from multiple dropout simulations and image augmentations, to calculate prediction certainty. This is quantified through the negative logarithm of the standard deviation $-log(\text{STD})$, where higher values denote greater confidence. The clinical utility of conformal prediction improves the precision and reliability of medical decisions by aligning the confidence measure with the Area Under the Curve (AUC).

Clinical Integration and Future Directions. The remarkable performance achieved by the Bonna algorithm allows its application in clinical settings, where its high predictive accuracy and certainty can significantly aid in informed

decision-making processes. This tool's integration into clinical practice promises to refine how decisions are made in ART, with a solid potential to enhance the efficacy of embryo selection and implantation strategies.

The Bonna algorithm is, however, limited by its reliance on retrospective data and the specificity of its current applications. Prospective studies are essential to establish its real-time effectiveness and adaptability across various IVF clinics worldwide and should include a broader demographic of patients. While expanding outcome measures to include live births could provide a more comprehensive success metric, this introduces complexity as such measures integrate broader, non-embryonic factors like maternal health, which are beyond the predictive scope of embryo imagery. Furthermore, the algorithm's broader acceptance and utility would benefit from improvements in explainability [34], helping clinicians understand decision processes. However, explainability can come at a cost [35].

Future Research. Future research should focus on refining prediction confidence techniques to enhance reliability and accuracy, further integrating the algorithm into clinical practice. Additionally, applying the principles and techniques developed for the Bonna algorithm to other areas of medical imaging could significantly expand its impact and utility across different medical fields. This expansion not only broadens the clinical applications but also fosters innovative approaches in predictive modeling and patient care optimization.

Competing Interests. The developed algorithm has potential for future commercial applications by an entity employing the authors.

References

1. Mastenbroek, S., et al.: Embryo selection in IVF. Hum. Reprod. **26**(5), 964–966 (2011). https://doi.org/10.1093/humrep/der050
2. Maheshwari, A., McLernon, D., Bhattacharya, S.: Cumulative live birth rate: time for a consensus? Hum. Reprod. **30**(12), 2703–2707 (2015). https://doi.org/10.1093/humrep/dev263
3. Van den Abbeel, E., et al.: Association between blastocyst morphology and outcome of single-blastocyst transfer. Reprod. Biomed. Online **27**(4), 353–361 (2013). https://doi.org/10.1016/j.rbmo.2013.07.006. Epub 2013 Jul 18. PMID: 23953585
4. 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. **26**(6), 1270–1283 (2011). https://doi.org/10.1093/humrep/der037
5. Khosravi, P., Kazemi, E., Zhan, Q., et al.: Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. NPJ Digit. Med. **2**, 21 (2019). https://doi.org/10.1038/s41746-019-0096-y
6. Fordham, D.E., et al.: Embryologist agreement when assessing blastocyst implantation probability: is data-driven prediction the solution to embryo assessment subjectivity? Hum. Reprod. **37**(10), 2275–2290 (2022). https://doi.org/10.1093/humrep/deac171

7. Gardner, D.K., et al.: Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer. Fertil. Steril. **73**(6), 1155–1158 (2000). https://doi.org/10.1016/S0015-0282(00)00518-5
8. Hernández-Vargas, P., Muñoz, M., Domínguez, F.: Identifying biomarkers for predicting successful embryo implantation: applying single to multi-omics to improve reproductive outcomes. Hum. Reprod. Update **26**(2), 264–301 (2020). https://doi.org/10.1093/humupd/dmz042
9. Munné, S., et al.: Preimplantation genetic testing for aneuploidy versus morphology as selection criteria for single frozen-thawed embryo transfer in good-prognosis patients: a multicenter randomized clinical trial. Fertil. Steril. **112**(6), 1071–1079.e7 (2019). https://doi.org/10.1016/j.fertnstert.2019.07.1346
10. Sandler, M., et al.: Mobilenetv2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
11. Isola, P., et al.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017). https://arxiv.org/abs/1611.07004
12. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. In: ICLR (2021). https://openreview.net/forum?id=YicbFdNTTy
13. Shafer, G., Vovk, V.: A tutorial on conformal prediction. J. Mach. Learn. Res. **9**, 371–421 (2008). http://jmlr.org/papers/v9/shafer08a.html
14. Theilgaard Lassen, J., et al.: Development and validation of deep learning based embryo selection across multiple days of transfer. Sci. Rep. **13**(1), 4235 (2023). https://doi.org/10.1038/s41598-023-31136-3
15. Dietterich, T.G.: Ensemble methods in machine learning. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 1–15. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-45014-9_1
16. Zhou, Z.H.: Ensemble Methods: Foundations and Algorithms. Chapman and Hall/CRC (2012). https://www.crcpress.com/Ensemble-Methods-Foundations-and-Algorithms/Zhou/p/book/9781439830031
17. Montag, M., Kajhøj, T.Q., Agerholm, I.E.: Description of time-lapse systems: embryoscope^{TM}. In: Meseguer, M. (ed.) Time-Lapse Microscopy in In-Vitro Fertilization. chapter 11, pp. 11–30. Cambridge University Press, Cambridge (2016)
18. Zegers-Hochschild, F., et al.: The international glossary on infertility and fertility care, 2017. Fertil. Steril. **108**(3), 393–406 (2017). https://doi.org/10.1016/j.fertnstert.2017.06.005
19. Sayed, S., et al.: Time-lapse imaging derived morphokinetic variables reveal association with implantation and live birth following in vitro fertilization: a retrospective study using data from transferred human embryos. PLoS ONE **15**(11), e0242377 (2020). https://doi.org/10.1371/journal.pone.0242377
20. Maurício, J., Domingues, I., Bernardino, J.: Comparing vision transformers and convolutional neural networks for image classification: a literature review. Appl. Sci. **13**, 5521 (2023). https://doi.org/10.3390/app13095521
21. Singh, D., Singh, B.: Investigating the impact of data normalization on classification performance. Appl. Soft Comput. **97**(Part B), 105524 (2020). https://doi.org/10.1016/j.asoc.2019.105524
22. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. J. Big Data **6**, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0

23. Garcia-Garcia, A., et al.: A review on deep learning techniques applied to semantic segmentation. arXiv preprint arXiv:1704.06857 (2017). https://arxiv.org/abs/1704.06857
24. Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice Hall PTR (1994)
25. Börnfors, F., Klint, E.: Data Augmentation to Increase Multi-Site Robustness for Convolutional Neural Networks - A case study on MRI segmentation of target and organs at risk for prostate cancer (2019). ISSN 1404-6342. Student Paper
26. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: Proceedings of the 33rd International Conference on Machine Learning - vol. 48, pp. 1050–1059, New York, NY, USA (2016). JMLR.org
27. Hajian-Tilaki, K.: Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation. Caspian J. Int. Med. **4**(2), 627–635 (2013). PMID: 24009950; PMCID: PMC3755824
28. Liu, Y., et al.: Time-lapse deselection model for human day 3 in vitro fertilization embryos: the combination of qualitative and quantitative measures of embryo growth. Fertil. Steril. **105**(3), 656–662.e1 (2016). https://doi.org/10.1016/j.fertnstert.2015.11.003. Epub 2015 Nov 23
29. Loewke, K., et al.: Characterization of an artificial intelligence model for ranking static images of blastocyst stage embryos. Fertil. Steril. **117**(3), 528–535 (2022). https://doi.org/10.1016/j.fertnstert.2021.11.022
30. Enatsu, N., et al.: A novel system based on artificial intelligence for predicting blastocyst viability and visualizing the explanation. Reprod. Med. Biol. **21**(1), e12443 (2022). https://doi.org/10.1002/rmb2.12443
31. Erlich, I., et al.: Pseudo contrastive labeling for predicting IVF embryo developmental potential. Sci. Rep. **12**, 2488 (2022). https://doi.org/10.1038/s41598-022-06336-y
32. Diakiw, S.M., et al.: Development of an artificial intelligence model for predicting the likelihood of human embryo euploidy based on blastocyst images from multiple imaging systems during IVF. Hum. Reprod. **37**(8), 1746–1759 (2022). https://doi.org/10.1093/humrep/deac131
33. Tran, D., et al.: Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. Hum. Reprod. **34**(6), 1011–1018 (2019). https://doi.org/10.1093/humrep/dez064
34. Weiss, T., et al.: Interpretable deep-learning unveils structure-property relationships in polybenzenoid hydrocarbons. ChemRxiv (2022). This content is a preprint and has not been peer-reviewed
35. Bronstein, M.: The road to biology 2.0 will pass through black box data. Towards Data Science (2024). https://towardsdatascience.com/the-road-to-biology-2-0-will-pass-through-black-box-data-bbd00fabf959. Accessed 15 Apr 2024