Exploring the Role of Explainability in AI-Assisted Embryo Selection

Lucia URCELAY [*] [a], Daniel HINJOS[*] [a], Pablo MARTIN-TORRES[*] [a],
Marta GONZÁLEZ[*] [a], Marta MÉNDEZ [c], Salva CÍVICO [c],
Sergio ÁLVAREZ-NAPAGAO [a,b], Dario GARCIA-GASULLA [a,b]

[a] *Barcelona Supercomputing Center*
[b] *Universitat Politécnica de Catalunya - Barcelona TECH*
[c] *Hospital Clínic de Barcelona*

Abstract. In Vitro Fertilization is among the most widespread treatments for infertility. One of its main challenges is the evaluation and selection of embryo for implantation, a process with large inter- and intra-clinician variability. Deep learning based methods are gaining attention, but its opaque nature compromises their acceptance in the clinical context, where transparency in the decision making is key. In this paper we analyze the current work in the explainability of AI-assisted embryo analysis models, identifying the limitations. We also discuss how these models could be integrated in the clinical context as decision support systems, considering the needs of clinicians and patients. Finally, we propose guidelines for the sake of increasing interpretability and trustworthiness, pushing this technology forward towards established clinical practice.

Keywords. embryo selection, IVF, artificial intelligence, explainability, ethics

1. Introduction

Infertility is a common reproductive health problem that affects millions of people worldwide, causing social, psychological, physical and economic distress to the ones seeking to conceive [7]. In the coming years infertility rates are projected to grow due to environmental and lifestyle factors [18, 37]. In vitro fertilization (IVF) technology is used to overcome infertility, it involves the fertilization of an egg with sperm in the laboratory, followed by the transfer of the resulting embryos into the patient's uterus. The main challenge of IVF is the selection of the embryos that will be either selected for implantation, frozen (for later implantation) or discarded (if they exhibit undesirable features). This selection is to be performed during the early hours after embryo insemination, typically between three and five days after. During this time, embryos are monitored in time-lapse imaging incubators (TLI), facilitating uninterrupted embryo growth within stable culture conditions. This technology offers a dynamic perspective on in vitro embryonic development, augmenting the clinical effectiveness of IVF [29].

---

[*]Equal contributors

To assess quality, embryologists evaluate different morphological characteristics depending on the embryo development phase. Early development (days one to three) focus on cell number, symmetry and fragmentation rate, while embryos reaching bastocyst stage (day five) are further assessed by their expansion grade and the appearance of the inner cell mass (ICM) and the Zona Pellucida (ZP), as well as to the trophectoderm cells (TE) [30]. These morphological features represent the foundation of current development assessment guidelines such as Gardner's [15] or ASEBIR in Spain. These approaches are limited by the subjective assessment of embryologists, which causes inter and intra-observer variability. The success rates of IVF are restricted due to these limitations. According to [1], the global life birth rate via IVF in 2015 stood at 19.2% per oocyte retrieval with fresh embryo transfer and 24.8% with frozen embryo transfer.

Artificial Intelligence (AI), and specially Deep Learning (DL), due to its capacity for dealing with images, have recently been considered to assist in the embryo assessment and selection process. AI has the potential to facilitate and improve the process of embryo selection, increasing the implantation success rates, and reducing the chances of multiple pregnancies. AI can also mitigate inter and intra-observer variability, making results more reproducible and comparable [35]. Finally, AI can help reduce the financial, physical and emotional burden on patients, by optimising the treatment plan and minimising the need for repeated cycles of IVF.

Whenever AI is considered for an application which directly affects human life, trustworthiness must be assured. In this paper we review the current work on AI-assisted embryo selection with special focus on Explainable Artificial Intelligence (XAI), and how this is deployed. We analyze the limitations of current work and propose approaches that would yield AI systems that can provide better explanations for their recommendations. Moreover, we discuss key aspects in the integration of these systems in a clinical context where both clinicians and patients are involved in the decision-making process.

2. Related Work

This work discusses the integration of explainability components into AI solutions for embryo analysis in IVF, as well as the particularities of their integration into a human-centric decision making process. Previous solutions proposed in the bibliography are reviewed within the main discussion of this work (see §3.1), as it contextualizes our contribution. In contrast, the related work presented in this section comprehends previous analysis and surveys conducted in the field, which have motivated this paper.

Among recent surveys on AI-assisted embryo selection in IVF, most restrict themselves to the technicalities of the machine learning problem itself. That is, what is the problem to solve, which data is used and how, and which are the top-of-the-line performance metrics achieved by the models built. A few discuss the biological aspects of the problem with particular depth [22], while others focus on the variety of tasks one can tackle (*e.g.*, embryo segmentation, quality grading, etc.) [20]. The most ambitious tasks, those of predicting clinical pregnancy or fetal heartbeat from five days old fertilized embryo images, are reviewed in [23], together with the ploidy status (number of chromosomes) prediction task. [26] includes the annotation of data (*e.g.*, development phase, cell counting) in its review, while the particularities of the time-lapse imaging are

the focus of [28]. On top of embryos, [11] also reviews AI applications for spermatozoa and oocyte analysis. [13] goes beyond DL methods, and also consider Bayesian Networks and SVMs in their study. Remarkably, very few works [2, 40, 34] study the ethical, socio-economic, legal and cultural (ELSEC) aspects in AI-assisted embryo selection.

None of the previously discussed works includes a review on the XAI methods used for embryo analysis, or discussion on practical and responsible deployment. This work aims at filling that gap, considering that both these issues need to be properly assessed before clinicians can integrate AI in their day-to-day practice.

3. The Role of Explainable Artificial Intelligence in Embryo Analysis

Embryo analysis post-fertilization lasts for a maximum of 7 days. During that time, clinicians assess the status and evolution of embryos, as these evolve under controlled conditions. Through continuous visual monitoring and tracking of their morphological features, experts eventually discard, freeze or select embryos for transfer. Current embryo incubators include periodic microscopy, with images being recorded at frequency in the order of minutes (*e.g.*, ten) at several focal planes.

AI models trained on time-lapse embryo images have shown promise in embryo analysis. Within the domain of images, explainable AI methods are often feature scorers, representing their output as saliency maps on the input image (see Figure 1). By nature, XAI methods for DL are approximations to the model's real behavior. It is worth noting that XAI is critical to ensure that the decision-making process of an algorithm is as transparent and understandable to clinicians and patients as possible. The lack of explainability can lead to a lack of trust in the technology, which ultimately hinders its adoption.

In this section we review the use of XAI on embryo images. Among the methods found in the literature, we can differentiate between those applied after the model is trained (post-hoc), and those where the explainability is part of the model architecture and design (intrinsic). Post-hoc methods include model-agnostic ones such as LIME [33] and KernelSHAP [27], that obtain the explanations by perturbing the inputs and observing the changes caused on the outputs; also, model-specific methods such as Grad-CAM [38] and Deep SHAP [27], that use the parameters of neural network models to obtain the saliency maps. Among intrinsic methods, we can find attention-based CNNs, where the explainability is obtained from attention layers. A different approach is the use of Speeded Up Robust Features (SURF) [6] to extract local visual features, in conjunction with Gaussian Mixture Models (GMM) to obtain a Fisher vector per image [21].
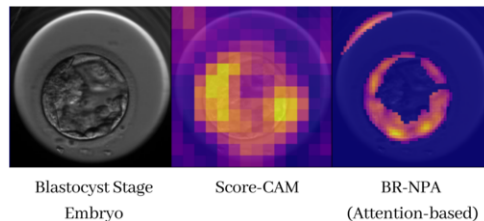

Blastocyst Stage Embryo   Score-CAM   BR-NPA (Attention-based)

Figure 1. Saliency maps for Score-CAM and BR-NPA XAI methods from [17].

Ref.	Task	Architecture	XAI Method(s)	ClinicalEval
[10]	Embryo Quality Grading	DenseNet, ResNet	Grad-CAM++	No
[43]	Embryo Quality Grading	LWMA-Net	Grad-CAM	No
[21]	Embryo Quality Grading	SURF and GMM, ResNet-101, EfficientNet-B1	SURF, Grad-CAM	No
[4]	Embryo Segmentation	MASS-Net	Grad-CAM	No
[39]	Development Stage Identification	ResNet34, VGG16	Grad-CAM, SHAP, LIME	Yes
[32]	Embryo Quality Grading	ResNet50, VGG-16	Grad-CAM	No
[17]	Development Stage Identification	ResNet50, ABN	B-CNN, ABN, InterByParts, Grad-CAM++, RISE, Score-CAM, Ablation-CAM, AM	No
[12]	Fetal Heart Pregnancy, Live Birth	ResNet18, RandomForest	Grad-CAM, SHAP	No
[36]	Live Birth	ABN	ABN	Yes
[42]	Embryo Quality Grading	VGG-16	Grad-CAM	No
[41]	Embryo Quality Grading	Xception, Inception-ResNET-v2	CAM	No

Table 1. Summary table of studies that use XAI on embryo selection. The Clinical Evaluation column indicates whether the resulting saliency maps have been evaluated by experts or not.

3.1. Review on Current Work

Depending on the specific task, XAI is used and interpreted differently:

Embryo Quality Grading: the classification of embryos (*e.g.*, day 5 blastocysts), into classes for selecting the best for implantation. This challenge is the most addressed regarding XAI. Some studies apply Grad-CAM for visualization [32, 42] and observe that key regions that the model relies on seem to be consistent with clinical interpretation. [10] uses Grad-CAM++ for visualization, while statistically correlating their model outputs with the Gardner score. [41] directly use CAM saliency maps, and observe their model focus on known features like cellular fragmentation, blastomeres or vacuoles. A different approach compares Grad-CAM and SURF (computing a single Fisher Vector per image), finding the former more prone to erroneously focus on irrelevant areas [21].

Embryo Development Stage Identification: consists on the automated labelling of the embryo stage. Challenges include deformed cell shapes, poor visual features or similarities between embryos at different stages. [39] compares Grad-CAM and LIME classification activations, concluding that LIME explanations seem less consistent with biologically important regions. Authors also suggest SHAP could be useful to identify reasons for misclassification between adjacent cleavage stages.

Blastocyst Segmentation: four crucial morphological regions to segment: TE, ZP, ICM and BL (see §1). A multiscaling architecture that outputs segmentation masks for the four regions is introduced in [4], where authors apply Grad-CAM to different layers to show the evolution of activations. A U-Net has also been used to separate the background

from the blastocyst [21], using an ellipse on top of the segmentation mask to separate inner cell mass from trophectoderm.

Fetal Heart Pregnancy: is defined as the presence of fetal heartbeats in the uterus. Of course, the embryo implantation process does not guarantee a successful pregnancy, as it depends on many other factors (*e.g.*, age, progesterone levels, *etc.*). [12] uses Grad-CAM to detect morphological indicators, and SHAP to account for relevant metadata that influences the model's decision, concluding that embryo images are the best predictor for fetal heart beat, followed by age and pregnancy history.

Live Birth: prediction from time-lapse imaging, [36] uses an Attention Branch Network (ABN) [14]. Authors visualize embryo features of relevance for *Live Birth* by exploring the attention mechanism. Results indicated that no common visual features could be associated to the predicted outcome of live or non-live birth.

To the best of our knowledge, other very important challenges in this field, such as *Blastocyst Prediction* or *Ploidy Detection*, are not addressed in the context of explainability. This highlights the need for further research in these tasks.

3.2. Limitations and Suggestions

A limited amount of research comprehends the use of AI-assisted IVF in combination with XAI methods. A representative list can be found in Table 1. In many cases XAI is only considered through the illustration and minor discussion of a few saliency maps [4, 21, 41, 43]. Others take one more step and analyze the map activations in order to correlate them to morphological features [39, 42] or to the objective [12]. Few share the saliency maps with expert embryologists for evaluation [39, 36]. Meanwhile, several studies conclude that saliency maps should not be used as the sole source of explainability in high risk medical domains [5, 16].

Considering the features that clinicians base their decisions on, and the difficulty of interpreting saliency maps, we suggest the use of expert models. Different models targeting different morphokinetical features, such as fragmentation score, ICM and TE grades, ZP and Pronucleus characteristics, and phase transition times. The ensemble of all these experts should be a white box model (*e.g.*, GLM, Decision Tree) to gain an intermediate level of explainability on how each feature contributes to the objective.

In the papers reviewed, the XAI method used for interpretability is rarely baselined or justified. Two studies compare two explainability methods, [39] (LIME vs Grad-CAM) and [21] (Bag-of-visual-words-based approach vs Grad-CAM). Only one thorough comparative study was found [17], in which nine different XAI methods, including five post-hoc ones, were evaluated using seven faithfulness metrics. This is, as well, the only study in which explainability methods are assessed through quality metrics such as the Increase in Confidence or the Average Drop [9].

Another limitation in the field is privacy. None of the previous works makes the code (except for [4]) or the data (except for [17]) publicly available, making reproducibility impossible and replicability impractical. These flaws, in a task with a high degree of subjectivity (inter and intra-observer variability) induce a high risk of confirmation bias. This issue could be mitigated through a quantitative benchmark measuring the model's attention to certain embryo regions of interest across a relatively big amount of samples.

That would require an additional segmentation model such as [4], which isolates the TE, ZP, ICM, and BL regions. A good explanation from a good model should overlap with the key segmented areas, and could be assessed with metrics like IoU. This approach is limited to blastocyst images, and subject to the accuracy of the segmentation model.

4. Integrating Explainable Artificial Intelligence in Decision Support Systems

The performance of AI methods on top-of-the-line metrics (*e.g.*, precision, recall, *etc.*) is often the endpoint of most contributions in the embryo assessment field. However, transforming a quantitatively performing model into a successful practical deployment is no easy task, esp. when considering the high-stakes involved in the context of IVF. The question on how to effectively and safely introduce this technology into a clinical setting where the different actors, including clinicians and patients, are involved, remains unsolved. Next we aim at advancing towards a solution, filling a gap in current literature.

4.1. Opaque Models Create Responsibility Gaps

The use of opaque AI models raises ethical and legal accountability concerns [2], particularly when clinicians can not explain the decision-making process. This creates a "responsibility gap", and without established accountability mechanisms it is challenging to determine who should be held responsible for any potential harm. For example, in cases of sub-optimal embryo selection or injury due to model recommendations, the decision-making process must be explainable to patients seeking to understand what happened or, in more extreme cases, seeking compensation for the damage caused. As of today, distrust in AI applications in medicine also comes from doctors' fear of legal repercussions if something goes wrong due to unclear liability regimes [31]. If clinicians base their decision on these opaque AI models, the evaluation of the decision-making process and, consequently, the determination of who is responsible will be greatly hampered.

It is then essential that the explanations provided by the model are as clear and understandable as possible by the clinician. For embryo image analysis this motivates the use of saliency maps obtained through feature attribution methods, which provide reliable (*i.e.*, true to the underlying model behavior) but also accessible evidence. To further ground the produced explanations in the context of clinical embryo selection, we suggest the mapping of this evidence-based support to the established annotation and evaluation guidelines and practices of the field, *e.g.*, through the association of feature attributions with the already established morphokinetic markers (such as a part of an image showing inappropriate cellular fragmentation), so that evidence provided is given in a language appropriate for the clinician and the problem. Still, due to the limitations of these XAI methods, it is recommended to adhere to ethical guidelines in AI, such as the one proposed by the EU [3]. By following this framework, other key requirements such as data privacy and governance or human agency and oversight can be accomplished.

4.2. Building Trust through Explanations

The successful integration of AI-assisted embryo selection systems in the medical context heavily depends on the acceptance by the clinicians and, consequently, their trust in them.

According to [31], there are several levels of trust which fall along a spectrum, ranging from complete distrust to over-reliance on AI systems. Studies have shown that over-reliance on the suggestions can cause clinicians to take less initiative [25] and also to be more likely to accept incorrect diagnosis [19]; this is known as *automation bias* [24]. On the other side of the spectrum lie clinicians which do not trust an algorithm that they do not understand [8], a phenomenon known as *algorithmic aversion*. As a result, not only training and education programs are necessary to ensure that clinicians understand the capabilities and limitations of AI-assisted embryo selection systems, but also making sure that the explanation or the model are not having negative impacts on the clinician's decision making process.

It is recommended to avoid offering the discretized outcome of models (*e.g.*, this embryo is good) in favour of more detailed information, such as outcome probabilities (*e.g.*, this embryo is good with a probability of 55%). For further bias avoidance, evidence may be presented only in the form of explanations, with the outcome hidden (*e.g.*, this area shows relevant visual features for determining if the embryo is good or not), so that the clinician interprets and integrates the visual features unbiasedly *w.r.t.* to the model's final prediction. Unfortunately, this approach is not without flaws either. It can induce *confirmation bias* (*i.e.*, a clinician may only review the evidence that supports its own hypothesis), while preventing the methodology to exploit visual patterns not perceivable by humans, which may be a limiting factor in embryo analysis. The proposed solution is to provide all information available, outcome probabilities and saliency maps, and to extend that information to all relevant cases (*e.g.*, an area shows relevant visual features in favour of the embryo being good, an area shows features in favour of it being bad).

4.3. Patient, Clinician, AI

IVF is a highly personal procedure, with a large impact in the physical and mental health of individuals. Patients undergoing IVF have the right to know how decisions are being made about their own healthcare, and to maintain their decision-making power. To guarantee that, AI explanations should be also accessible to them; they cannot be left behind by the complexity of technology. Integrating explainability into AI-assisted embryo selection systems can not only target clinicians, it must also empower patients. By providing them with clear and easy to understand explanations of certain decisions, patients can be more involved and better equipped for the decision-making process regarding their own care. XAI results should be presented when relevant, and under the supervision of the clinician for its adequate interpretation and contextualization. Finally, it has to be mentioned that, since the use of these systems directly involves the processing of their data, the patient should be educated about how it will be processed and the use that will be made of it, in addition to asking for their informed consent.

5. Conclusions

We present an analysis of the presence of XAI for embryo selection on the literature. While there is a relative abundance of models for embryo selection and many have shown promising results when trained on retrospective data, the performance of these models in actual cases is yet unknown and their integration in the medical decision process remains

unclear. Due to the opaque nature of these models, explainability is crucial to ensure their applicability. However, although some studies tackle the issue of explainability, the interpretations offered are often insufficient or do not rely on medical experts for evaluation. We enumerate a set of guidelines and suggestions that could help increase the interpretability and trustworthiness of these systems, with the goal of advancing towards the successful and safe integration of this technology in clinical practice.

Acknowledgements

References

[1]   G. David Adamson et al. "International Committee for Monitoring Assisted Reproductive Technology: world report on assisted reproductive technology, 2011". In: *Fertility and Sterility* 110.6 (2018), pp. 1067–1080.

[2]   Michael Anis Mihdi Afnan et al. "Ethical Implementation of Artificial Intelligence to Select Embryos in In Vitro Fertilization". In: AIES '21. New York, NY, USA, 2021, pp. 316–326.

[3]   Pekka Ala-Pietilä et al. *Ethics Guidelines for Trustworthy AI*. en. Text. Dec. 2018.

[4]   Muhammad Arsalan et al. "Human Blastocyst Components Detection Using Multiscale Aggregation Semantic Segmentation Network for Embryonic Analysis". In: *Biomedicines* 10.7 (2022), p. 1717.

[5]   Nishanth Arun et al. "Assessing the Trustworthiness of Saliency Maps for Localizing Abnormalities in Medical Imaging". In: *Radiology: Artificial Intelligence* 3.6 (2021), e200267.

[6]   Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. "SURF: Speeded Up Robust Features". In: *Computer Vision – ECCV 2006*. Ed. by Aleš Leonardis, Horst Bischof, and Axel Pinz. Springer Berlin Heidelberg, 2006, pp. 404–417.

[7]   Jacky Boivin et al. "International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care". In: *Human reproduction* 22.6 (2007), pp. 1506–1512.

[8]   Carrie J Cai et al. "Hello AI: uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making". In: *Proceedings of the ACM on Human-computer Interaction* 3.CSCW (2019), pp. 1–24.

[9]   Aditya Chattopadhay et al. "Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks". In: *2018 IEEE winter conference on applications of computer vision (WACV)*. IEEE. 2018, pp. 839–847.

[10] Sonya M. Diakiw et al. "An artificial intelligence model correlated with morphological and genetic features of blastocyst quality improves ranking of viable embryos". In: *Reproductive BioMedicine Online* 45.6 (2022), pp. 1105–1117.

[11] Irene Dimitriadis et al. "Artificial intelligence in the embryology laboratory: a review". In: *Reproductive BioMedicine Online* 44.3 (2022), pp. 435–448.

[12] Noritoshi Enatsu et al. "A novel system based on artificial intelligence for predicting blastocyst viability and visualizing the explanation". In: *Reproductive Medicine and Biology* 21.1 (2022), e12443.

[13] Eleonora Inácio Fernandez et al. "Artificial intelligence in the IVF laboratory: overview through the application of different types of algorithms for the classification of reproductive data". In: *Journal of Assisted Reproduction and Genetics* 37.10 (2020), pp. 2359–2376.

[14] Hiroshi Fukui et al. "Attention Branch Network: Learning of Attention Mechanism for Visual Explanation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019.

[15] D. K. Gardner and D. Sakkas. "Assessment of embryo viability: the ability to select a single embryo for transfer–a review". In: *Placenta* 24 Suppl B (2003), S5–12.

[16] Marzyeh Ghassemi, Luke Oakden-Rayner, and Andrew L Beam. "The false hope of current approaches to explainable artificial intelligence in health care". In: *The Lancet Digital Health* 3.11 (2021), e745–e750.

[17] Tristan Gomez, Thomas Fréour, and Harold Mouchère. *Comparison of attention models and post-hoc explanation methods for embryo stage identification: a case study*. 2022.

[18] A. C. Gore et al. "EDC-2: The Endocrine Society's Second Scientific Statement on Endocrine-Disrupting Chemicals". In: *Endocrine Reviews* 36.6 (2015), E1–E150.

[19] Yukinori Harada et al. "Effects of a differential diagnosis list of artificial intelligence on differential diagnoses by physicians: an exploratory analysis of data from a randomized controlled study". In: *International Journal of Environmental Research and Public Health* 18.11 (2021), p. 5562.

[20] Iza Sazanita Isa, Umi Kalsom Yusof, and Murizah Mohd Zain. "Image Processing Approach for Grading IVF Blastocyst: A State-of-the-Art Review and Future Perspective of Deep Learning-Based Models". In: *Applied Sciences* 13.2 (2023).

[21] Athanasios Kallipolitis et al. "Explainable computer vision analysis for embryo selection on blastocyst images". In: *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. 2022, pp. 1–4.

[22] Jihyun Kim, Jaewang Lee, and Jin Hyun Jun. "Non-invasive evaluation of embryo quality for the selection of transferable embryos in human in vitro fertilization-embryo transfer". In: *Clinical and Experimental Reproductive Medicine* 49.4 (2022), pp. 225–238.

[23] Evangelos Maziotis Konstantinos Sfakianoudis. "Reporting on the Value of Artificial Intelligence in Predicting the Optimal Embryo for Transfer: A Systematic Review including Data Synthesis". In: *Biomedicines* 10.3 (2022), p. 697.

[24] John D Lee and Katrina A See. "Trust in automation: Designing for appropriate reliance". In: *Human factors* 46.1 (2004), pp. 50–80.

[25] Ariel Levy et al. "Assessing the impact of automated suggestions on decision making: Domain experts mediate model errors but take less initiative". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–13.

[26] Claudio Michael Louis et al. "Review of computer vision application in in vitro fertilization: the application of deep learning-based computer vision technology in the world of IVF". In: *Journal of Assisted Reproduction and Genetics* 38.7 (2021), pp. 1627–1639.

[27] Scott M. Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: NIPS'17 (2017), pp. 4768–4777.

[28] Kersti Lundin and Hannah Park. "Time-lapse technology for embryo culture and selection". In: *Upsala Journal of Medical Sciences* 125.2 (2020), pp. 77–84.

[29] Yasuyuki Mio and Kazuo Maeda. "Time-lapse cinematography of dynamic changes occurring during in vitro development of human embryos". In: *American journal of obstetrics and gynecology* 199.6 (2008), 660–e1.

[30] Nahid Nasiri and Poopak Eftekhari-Yazdi. "An overview of the available methods for morphological scoring of pre-implantation embryos in in vitro fertilization". In: *Cell Journal (Yakhteh)* 16.4 (2015), p. 392.

[31] Cecilia Panigutti et al. "Understanding the impact of explanations on advice-taking: a user study for AI-based clinical Decision Support Systems". In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–9.

[32] Elena Payá et al. "Automatic characterization of human embryos at day 4 post-insemination from time-lapse imaging using supervised contrastive learning and inductive transfer learning techniques". In: *Computer Methods and Programs in Biomedicine* 221 (2022).

[33] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. ""Why Should I Trust You?": Explaining the Predictions of Any Classifier". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 1135–1144.

[34] Vasilija Rolfes et al. "Artificial Intelligence in Reproductive Medicine – An Ethical Perspective". In: *Geburtshilfe und Frauenheilkunde* 83.1 (2023), pp. 106–115.

[35] Zev Rosenwaks. "Artificial intelligence in reproductive medicine: a fleeting concept or the wave of the future?" In: *Fertility and Sterility* 114.5 (2020), p. 905.

[36] Yuki Sawada et al. "Evaluation of artificial intelligence using time-lapse images of IVF embryos to predict live birth". In: *Reproductive Biomedicine Online* 43.5 (2021), pp. 843–852.

[37] Thalia R. Segal and Linda C. Giudice. "Before the beginning: environmental exposures and reproductive and obstetrical outcomes". In: *Fertility and Sterility* 112.4 (2019), pp. 613–621.

[38] Ramprasaath R. Selvaraju et al. "Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization". In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Oct. 2017.

[39] Akriti Sharma et al. "Explainable Artificial Intelligence for Human Embryo Cell Cleavage Stages Analysis". In: *Proceedings of the 3rd ACM Workshop on Intelligent Cross-Data Analysis and Retrieval*. 2022, pp. 1–8.

[40] Sivan Tamir. "Artificial intelligence in human reproduction: charting the ethical debate over AI in IVF". In: *AI and Ethics* (2022).

[41] Prudhvi Thirumalaraju et al. "Evaluation of deep convolutional neural networks in classifying human embryo images based on their morphological quality". In: *Heliyon* 7.2 (2021).

[42] Shanshan Wang et al. "A Deep Learning Framework Design for Automatic Blastocyst Evaluation With Multifocal Images". In: *IEEE Access* 9 (2021).

[43] Chongwei Wu et al. "LWMA-Net: Light-weighted morphology attention learning for human embryo grading". In: *Computers in Biology and Medicine* 151 (2022).