bioengineering

MDPI

Review

Artificial Intelligence in Embryo Selection: Current Approaches and Clinical Implications

Lucia Maresca [1,2], Antonio D'Amato [1,*], Camilla Coianiz [1], Alessandra Cavalieri [1], Erica Vella [1], Francesco Gebbia [1], Lorena Bori [3] and Marcos Meseguer [3]

[1] IVIRMA Global Research Alliance, IVI Roma, 00161 Rome, Italy; lucia.maresca@ivirma.com (L.M.); camilla.coianiz@ivirma.com (C.C.); alessandra.cavalieri@ivirma.com (A.C.); erica.vella@ivirma.com (E.V.); francesco.gebbia@ivirma.com (F.G.)
[2] Department of Medicine, University of Valencia, 46010 Valencia, Spain
[3] IVIRMA Global Research Alliance, IVI Valencia, IVI Foundation, Instituto de Investigación Sanitaria La Fe, 46026 Valencia, Spain; lorena.bori@ivirma.com (L.B.); marcos.meseguer@ivirma.com (M.M.)
* Correspondence: antonio.damato@ivirma.com

Abstract

Embryo selection remains one of the main unresolved challenges in in vitro fertilization, despite major advances in assisted reproductive technologies. Conventional assessment is still largely based on morphological evaluation, which is limited by subjectivity, static observation, and the difficulty of integrating heterogeneous clinical and biological data. In recent years, artificial intelligence has emerged as a decision-support tool in embryology, enabling the analysis of large datasets derived from embryo images, morphokinetic parameters, and clinical variables. This review summarizes current approaches to artificial intelligence in embryo selection, including models based on static images and time-lapse imaging data. Machine learning and deep learning techniques are discussed, including convolutional neural networks and spatiotemporal models. The evaluation of model performance is also examined, highlighting the clinical relevance of endpoints such as time to live birth compared with traditional outcome measures. Finally, ethical and clinical implications are considered, including issues related to transparency, responsibility, human oversight, and regulation. Artificial intelligence has the potential to improve embryo selection, although further validation and standardized implementation are needed before routine clinical use.

Keywords: artificial intelligence; embryo selection; in vitro fertilization; machine learning; deep learning; time-lapse imaging; morphokinetics

Check for updates

Academic Editor: Chenxi Huang

Received: 7 May 2026
Revised: 8 June 2026
Accepted: 10 June 2026
Published: 16 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction: The Bottleneck of Embryo Selection in IVF

Although advances in the field of assisted reproductive technologies are now considerable, the outcomes of in vitro fertilization are still not optimal. Improvements in ovarian stimulation protocols, in vitro fertilization procedures, embryo culture conditions, and the cryopreservation of gametes and embryos have certainly enhanced laboratory efficiency and expanded the therapeutic possibilities for success. Despite this, one of the main unresolved problems in clinical practice remains the ability to reliably identify the embryo with the highest developmental potential post-embryo transfer within a given cohort [1,2].

In many laboratories, embryo selection in routine ART cycles is still primarily based on conventional morphological assessment performed at predefined developmental time points. Within this framework, normal fertilization is confirmed by the presence of two pronuclei; during cleavage-stage development, embryos are evaluated according

to cleavage dynamics, blastomere number, and morphological features; and, at the blastocyst stage (day 5/6), assessment is based on the degree of expansion and on the quality of the inner cell mass and trophectoderm, according to established grading systems such as the Gardner classification. This approach remains widely used because it is well established, practical, and fully integrated into the daily workflow of embryology laboratories. However, it has well-recognized limitations. Embryo development is a continuous and dynamic process, whereas conventional morphological assessment relies on static observations performed at predefined time points. As a result, potentially informative developmental events may occur outside these observation windows and therefore remain undetected by static assessment [1,3–7]. Another crucial aspect is subjectivity. Morphological grading depends largely on the embryologist's interpretation of visual features, inevitably introducing a certain degree of variability [8]. When parameters such as cell symmetry, fragmentation, cleavage patterns, or blastocyst quality are assessed, this variability results in different evaluations. This, in turn, may lead to different embryo rankings, even when the same cohort is analyzed by experienced operators such as senior embryologists. All this does not represent merely a technical issue but rather a clinically relevant limitation, because embryo selection directly influences the choice of the best embryo to transfer and, consequently, treatment outcome [3,9].

The introduction of time-lapse systems has partially addressed these limitations. These systems allow continuous monitoring of embryos in culture and make it possible to obtain morphokinetic information that cannot be captured by conventional static assessment alone. This has expanded the amount of biological information available and has further highlighted the idea that the timing of embryo development may provide additional valuable indications of implantation potential. Different algorithms and scoring systems have been proposed, but their performance is not always reproducible across different laboratories, and their applicability may be influenced by culture conditions, annotation methods, technical variability, and patient-related factors [3,4,10].

Morphology, developmental kinetics, and clinical variables related to the patient or the treatment are all important sources of information, but they are often considered separately. This fragmentation represents one of the main reasons why embryo selection continues to be complex in clinical practice. Embryo competence and viability, in fact, do not depend on a single observable feature, and the clinical value of embryo ranking does not concern only the individual embryo but is placed within the broader context of treatment, including the characteristics of the entire embryo cohort [1,4,10,11].

For all these reasons, embryo selection still represents one of the main bottlenecks in IVF today. Although the currently available approaches are certainly useful in clinical practice, they remain limited by incomplete standardization, residual subjectivity, and a reduced ability to consistently integrate heterogeneous sources of biological and clinical information. It is precisely from these limitations that, in recent years, interest has grown in more objective, reproducible and integrative decision-support strategies [2,3].

Therefore, the aim of this narrative review is not only to summarize current artificial intelligence-based approaches for embryo assessment but also to critically examine their clinical relevance, methodological limitations, validation requirements, and implementation challenges. Particular attention is given to the distinction between algorithmic performance and true clinical utility, including whether AI-assisted embryo ranking can improve patient-centered outcomes such as live birth, cumulative treatment efficiency, and time to live birth.

A schematic overview of the transition from conventional embryo assessment to AI-assisted embryo ranking is shown in Figure 1.

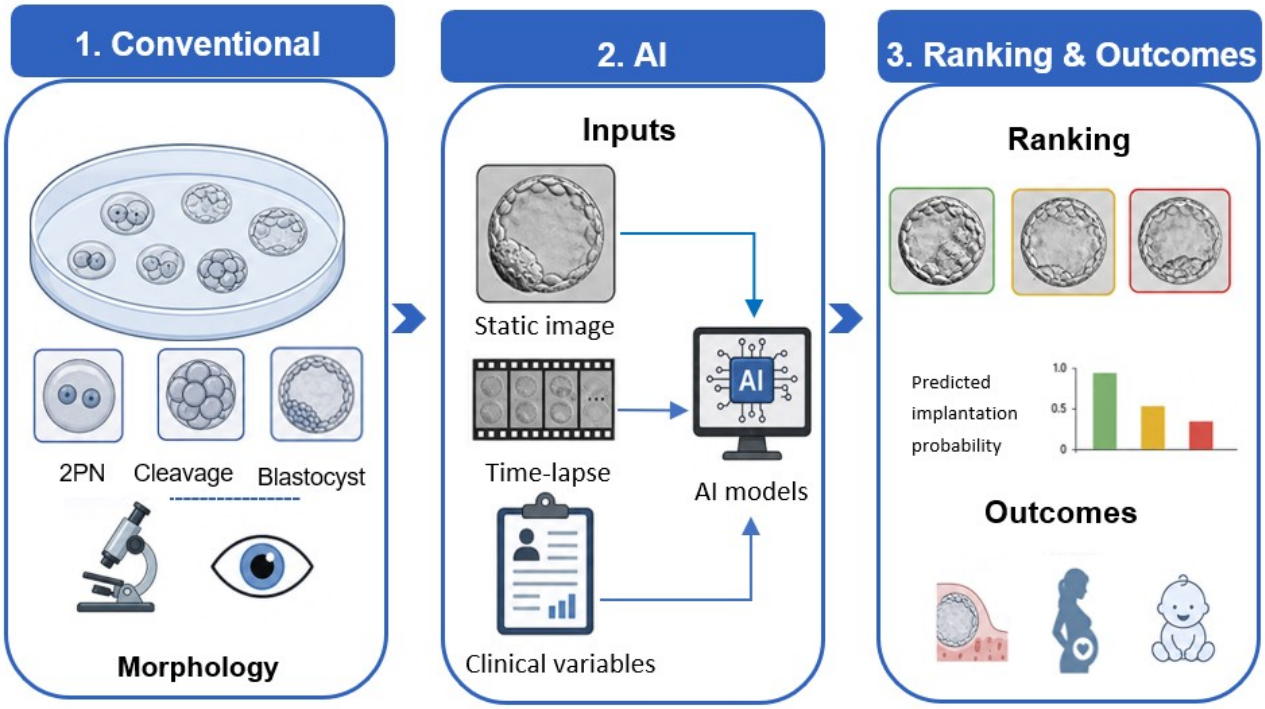


Figure 1. Schematic overview of embryo selection from conventional morphological assessment to AI-assisted ranking. Conventional evaluation is based on static embryo observations at key developmental stages, whereas AI-based approaches integrate static images, time-lapse information, and clinical variables to support embryo ranking and the prediction of clinically relevant outcomes.

2. Materials and Methods

This narrative review was conducted through a structured literature search of PubMed/MEDLINE, Web of Science, and Google Scholar for articles published in English up to April 2026. The search strategy was designed to identify studies addressing the application of artificial intelligence to embryo assessment, embryo ranking, and embryo selection in the context of assisted reproductive technology. Search terms included combinations of "artificial intelligence", "machine learning", "deep learning", "neural network", "convolutional neural network", "computer vision", "embryo selection", "embryo assessment", "embryo ranking", "embryo grading", "embryo viability", "blastocyst image", "static embryo image", "time-lapse imaging", "morphokinetics", "in vitro fertilization", "IVF", and "assisted reproductive technology".

Additional searches were performed using more specific terms related to clinically relevant outcomes and methodological aspects, including "embryo implantation", "implantation potential", "embryo ploidy", "euploidy", "aneuploidy", "fetal heartbeat", "clinical pregnancy", "live birth", "time to live birth", "model validation", "external validation", "calibration", "explainability", "black-box model", "clinical prediction model", and "AI ethics".

Reference lists of the selected articles and relevant reviews were manually screened to identify further studies of interest. Articles were included when they provided information relevant to the development, validation, interpretation, or clinical application of artificial intelligence models in embryo assessment and embryo selection. Particular attention was given to studies addressing model inputs, predicted outcomes, performance metrics, validation strategies, and clinical applicability. As this was a narrative review, the available evidence was summarized qualitatively, with the aim of highlighting current approaches, limitations, unresolved issues, and future perspectives in the use of artificial intelligence for embryo selection.

3. Why Artificial Intelligence Entered Embryo Selection

3.1. Artificial Intelligence in the Context of Embryology

In the field of reproductive medicine, artificial intelligence (AI) refers to computational approaches designed to analyze complex biomedical data and support clinical decision-making [12]. In IVF laboratories, AI can process large volumes of information derived from embryo images, morphokinetic parameters obtained from time-lapse systems, and patient clinical data [4,13,14]. By integrating these heterogeneous datasets, AI-based models aim to assist embryologists in predicting embryo viability, implantation potential, and treatment outcomes [13–15]. The application of AI in embryology seeks primarily to reduce the subjectivity associated with conventional embryo assessment. Traditional embryo evaluation relies largely on morphological criteria interpreted by embryologists, which may introduce inter- and intra-observer variability [5,16]. AI-driven approaches provide data-driven and standardized analyses, potentially improving the objectivity and reproducibility of embryo selection [13,14,16].

3.2. Machine Learning and Deep Learning in Embryology

Machine learning (ML) is a core component of artificial intelligence and includes algorithms capable of identifying patterns in data without explicit programming [12]. In reproductive medicine, conventional ML approaches rely on predefined variables, such as morphokinetic parameters, morphological scores, or clinical features, which are selected a priori by clinicians or researchers [4,5]. Deep learning (DL), a subset of ML, is based on artificial neural networks with multiple hidden layers that automatically extract relevant features from raw data [12]. In embryology, DL models—particularly convolutional neural networks (CNNs)—have demonstrated strong performance in the analysis of embryo images and time-lapse recordings, enabling the detection of complex visual and temporal patterns associated with embryo development and often outperforming traditional approaches based on manually defined features [13,15,16]. From a methodological perspective, AI models can be classified as interpretable (white-box) or non-interpretable (black-box) systems [17]. Interpretable models, such as linear models and decision trees, provide insight into the contribution of individual variables to the final prediction, whereas DL models are generally considered black-box systems due to their limited transparency [17]. Despite their high predictive performance, the reduced interpretability of DL approaches may limit their clinical applicability, underscoring the need to balance accuracy and explainability in AI-based embryo selection tools [15,17].

4. Learning from a Single Image: AI Models Based on Static Embryo Assessment

In the time-lapse setting, some AI models rely on a single selected frame rather than on the full video sequence. These static-image approaches, most often applied at the blastocyst stage, extract morphological features to predict clinically relevant outcomes, primarily embryo implantation potential and chromosomal status (ploidy) [13,14,18].

Several studies have investigated whether deep learning models, particularly convolutional neural networks (CNNs), can infer clinically relevant outcomes from morphological information contained in a single embryo image. One of the earliest and most influential examples is the study by Khosravi et al., who trained a CNN on more than 10,000 blastocyst images and reported an accuracy of 96.94% in distinguishing high-quality from low-quality embryos, showing that automated image analysis can identify morphological patterns that are difficult to quantify objectively by visual inspection alone [13]. In a more clinically oriented application, VerMilyea et al. developed the Life Whisperer system, a CNN-based model designed to classify embryos as viable or non-viable from static images. In indepen-

dent blinded test sets, the model achieved a sensitivity of 70.1%, a specificity of 60.5%, and an overall accuracy of 64.3%, outperforming embryologist assessment in that setting [14]. More recently, the MAIA platform was prospectively tested in a routine clinical setting using static embryo images, further supporting the feasibility of image-based AI tools as embryologist-support systems [19].

Subsequent studies extended the scope of static-image AI beyond conventional morphology-based grading. Conversa et al. evaluated a deep learning model trained to predict fetal heartbeat after transfer of vitrified-warmed blastocysts from a single image, supporting the feasibility of extracting clinically relevant predictive information even in the absence of full developmental sequence analysis [20]. Chávez-Badiola et al. developed the Embryo Ranking Intelligent Classification Algorithm (ERICA), which used static blastocyst images to rank embryos according to predicted implantation potential and ploidy. In their study, ERICA ranked a euploid blastocyst first in 78.9% of cases and placed at least one euploid embryo within the top two positions in 94.7% of cases, outperforming both random classification and senior embryologists [18]. In addition, Bori, Meseguer, and colleagues explored a multimodal strategy combining blastocyst morphology with proteomic biomarkers. Their preliminary artificial intelligence models correctly classified embryos, leading or not leading to live birth, with accuracies of 100.0% for ANN1, 85.7% for ANN2, and 83.3% for ANN3, suggesting that integration of morphological and non-visual biological data may further improve embryo viability assessment [21].

Taken together, these studies show that the main value of static-image AI models lies not simply in automating conventional morphological evaluation but in extracting predictive information from embryo images for a range of clinically relevant outcomes, including embryo quality, viability, implantation-related potential, ploidy, fetal heartbeat, and live birth [13,14,18,20,21]. However, because these models rely on a single image, their performance remains intrinsically dependent on the timing and quality of image acquisition and does not account for the temporal dynamics of embryo development (Table 1).

Table 1. Main characteristics of the static-image artificial intelligence models included in this review, including input type, model architecture, predicted outcome, and reported performance metrics. AI, artificial intelligence; ANN, artificial neural network; CNN, convolutional neural network.

Study	Input	Model	Main Predicted Outcome	Reported Performance
Khosravi et al. [13]	Single blastocyst image	CNN	Embryo quality classification	Accuracy: 96.94%
VerMilyea et al. [14]	Static embryo image	CNN (Life Whisperer)	Viable vs. non-viable embryo classification	Sensitivity: 70.1%; Specificity: 60.5%; Accuracy: 64.3%
Conversa et al. [20]	Single image of vitrified-warmed blastocyst	Deep learning model	Fetal heartbeat prediction	Predictive ability reported; metric to specify from original article
Chávez-Badiola et al. [18]	Static blastocyst image	ERICA	Embryo ranking for implantation potential and ploidy	Euploid embryo ranked first: 78.9%; at least one euploid embryo in top two: 94.7%
Bori et al. [21]	Blastocyst morphology + proteomic profile	ANN	Live birth classification	ANN1: 100.0%; ANN2: 85.7%; ANN3: 83.3%

5. Learning from Embryo Dynamics: Time-Lapse and Video-Based AI

One of the main advantages of time-lapse imaging is the ability to extract morphokinetic parameters, which describe the timing of key developmental events. Pioneering

work by Meseguer et al. demonstrated that embryo implantation probability is closely linked to specific chronological milestones, identifying key predictive parameters such as the time of division to 5 cells (t5), the interval between the third and fourth cleavage (s2), and the duration of the second cell cycle (cc2). Furthermore, the observation of dysmorphic behaviors, such as abrupt division or multinucleation, provided a basis for multivariable selection models [4]. Building upon these clinical findings, Milewski et al. applied more advanced computational techniques, using artificial neural networks and principal component analysis (PCA) to quantify the predictive information within these morphokinetic datasets [22]. These approaches represent the first generation of AI-assisted embryo evaluation based on time-lapse imaging. Subsequent developments introduced deep learning models capable of analyzing entire time-lapse image sequences rather than predefined morphokinetic parameters. By processing large volumes of sequential embryo images, these models automatically learn spatial and temporal patterns associated with embryo competence.

More recent approaches have integrated advanced deep learning architectures, including convolutional neural networks combined with recurrent neural networks such as long short-term memory (LSTM) layers, to simultaneously analyze spatial and temporal information in embryo development. In this context, Paya et al. developed a spatiotemporal model specifically designed to classify embryo ploidy status using time-lapse videos from 10 to 115 h post-insemination. Their end-to-end approach utilized a CNN for spatial feature extraction and a bidirectional LSTM (or GRU) layer to analyze temporal dependencies, achieving a precision of 0.8205 for predicting euploidy and an overall accuracy of up to 0.7308 [23]. For instance, deep learning frameworks applied to time-lapse monitoring have achieved AUC values around 0.82 for predicting blastocyst formation and quality, highlighting the potential of temporal AI models to improve embryo selection [24].

Additional studies have further explored AI models using time-lapse datasets to predict clinically relevant outcomes. For example, Maekawa et al. developed an AI system trained on time-lapse images to predict both implantation potential and embryo euploidy, demonstrating that developmental dynamics may contain information related not only to embryo viability but also to chromosomal integrity [25]. Similarly, recent deep learning models analyzing time-lapse videos (such as those by Boucret et al.) have been trained on datasets of embryos with known implantation outcomes, enabling the identification of dynamic patterns associated with successful pregnancy [26].

Compared with approaches based solely on static images, video AI models offer several advantages. By capturing the continuous progression of embryo development, these systems provide information on cleavage dynamics, cell cycle synchronization, and developmental irregularities that cannot be inferred from a single image. As a result, dynamic models may detect subtle temporal patterns associated with embryo competence, potentially improving predictive performance and reducing subjectivity in embryo evaluation.

However, the use of time-lapse data also introduces technical and operational challenges. Time-lapse imaging generates extremely large datasets, requiring substantial computational resources for storage, processing, and model training. In addition, the development of reliable AI models depends on access to high-quality annotated video datasets, which remain limited in many clinical settings. Another important limitation is the dependence on specific time-lapse incubator systems, which may vary between laboratories and affect the generalizability of trained models. As highlighted in recent systematic reviews, these challenges regarding standardization and computational demand remain significant hurdles for the universal clinical implementation of automated grading systems [27].

The main characteristics of representative artificial intelligence models based on time-lapse imaging and embryo developmental dynamics are summarized in Table 2.

Table 2. Main characteristics of time-lapse and video-based artificial intelligence models included in this review.

Study	Input	Model	Main Predicted Outcome	Reported Performance
Milewski et al. [22]	Time-lapse morphokinetic parameters, fragmentation, multinucleation, blastomere evenness, and female age	ANN combined with PCA	Embryo implantation potential	AUC: 0.75; validation set AUC: 0.71
Liao et al. [24]	Time-lapse videos from the first three days of embryo development	Deep learning ensemble models integrating DenseNet201, LSTM, and gradient boosting classifiers	Blastocyst formation and usable blastocyst prediction	STEM: accuracy 78.2%, AUC 0.82; STEM+: accuracy 71.9%, AUC 0.79
Paya et al. [23]	Time-lapse videos from 10 to 115 h post-insemination	Spatiotemporal deep learning model combining CNN-based feature extraction with recurrent layers	Embryo ploidy status	Best model: precision 0.8205 for euploidy prediction; sensitivity 0.6957; specificity 0.7813; accuracy 0.7308
Boucret et al. [26]	Time-lapse videos of matched high-quality embryos with known implantation data	Self-supervised contrastive learning, Siamese neural network, and XGBoost prediction model	Implantation outcome among morphologically good-quality embryos	AUC: 0.64 without transfer-history information; AUC: 0.57 when using the previous transfer outcome from the same cycle
Maekawa et al. [25]	Time-lapse videos and maternal age	Spatiotemporal convolutional neural network	Clinical pregnancy and embryo ploidy status	Pregnancy model AUROC: 0.799 validation set, 0.717 internal test set, 0.746 external test set; ploidy model AUROC: 0.802 validation set, 0.738 internal test set, 0.759 external test set

6. How to Evaluate AI Models: Metrics and Clinical Relevance

The performance of AI models for embryo selection should be assessed using metrics that reflect both predictive accuracy and clinical applicability. The TRIPOD + AI guideline [28] outlines how the performance of clinical prediction models should be reported, emphasizing three main aspects: model discrimination, calibration, and clinical usefulness. Most studies assessing AI algorithms in IVF rely on conventional performance metrics that capture different aspects of model performance. The area under the receiver operating characteristic curve (AUC) measures discrimination, that is, the ability of a model to distinguish between embryos with and without the outcome of interest across all possible classification thresholds [28,29]. Accuracy reflects the overall proportion of correctly classified cases, whereas sensitivity and specificity indicate the proportions of positive and negative cases correctly identified, respectively [30,31]. Accordingly, AUC, accuracy, sensitivity, and specificity remain among the most frequently reported metrics in studies evaluating AI-based embryo selection models [14,32,33]. These measures primarily quantify the ability of a model to discriminate between embryos associated with different outcomes, typically positive implantation or live birth. Discrimination metrics such as the AUC are widely used

to assess the performance of AI models and to compare different algorithms or embryo ranking systems, and they remain among the most frequently reported indicators in the embryo selection literature [34]. Moreover, comparisons based on AUC can be misleading when models are evaluated across studies involving different embryo populations, outcome definitions, or label distributions, as these factors can substantially influence performance estimates and introduce bias in the interpretation of relative algorithm performance [35,36].

Nevertheless, in the clinical IVF setting, the key parameter is not to identify the absolute probability of success for an individual embryo across the entire population but rather to determine which embryo among those within the same cohort should be transferred first. In fact, one of the most significant challenges worldwide is the discontinuation of treatment by couples after one or more unsuccessful IVF attempts. This phenomenon, commonly referred to as treatment dropout, represents a major limitation in assisted reproduction because many couples interrupt the treatment pathway before completing the number of cycles required to achieve a high cumulative probability of live birth [37]. The causes of this phenomenon are complex and multifactorial, with psychological and socioeconomic factors being some of the most common ones reported in scientific literature [38].

For this reason, increasing attention has been directed toward analytical approaches that assess AI model performance within the same treatment cycle. In this context, the relevant question becomes whether an algorithm can correctly rank embryos belonging to the same cohort according to their probability of leading to a live birth. To address this issue, the concept of time to live birth has recently been proposed as a clinically relevant endpoint. Time to live birth (TTLB) is defined as the time or number of embryo transfer attempts required to achieve a live birth within a given treatment pathway when embryos are transferred according to a specific ranking or selection strategy. This endpoint provides a more accurate representation of how efficiently a selection method leads to the first live birth.

Particularly, in a recent publication, Bori et al. proposed a novel methodological approach to evaluate the TTLB and the clinical utility of embryo selection algorithms by considering all usable embryos within a treatment cycle, including those transferred and not transferred [2]. A multiple imputation approach by chained equations (MICE) was used to estimate the outcomes of non-transferred embryos [39] and applied to a large dataset. TTLB was then compared between AI-based embryo ranking and conventional morphological grading, showing a 6.1% shorter estimated TTLB with AI-guided selection [2].

Importantly, the distinction between algorithmic performance and clinical utility has also been highlighted by recent randomized evidence. In a randomized, double-blind non-inferiority trial comparing iDAScore-based deep learning embryo selection with standard morphology-based assessment, Illingworth et al. reported similar clinical pregnancy rates between groups, but non-inferiority of the deep learning strategy was not demonstrated [40].

7. Discussion

Artificial intelligence is increasingly being explored as a decision-support tool in embryo selection because it addresses some of the main limitations of conventional assessment, including subjectivity, inter-observer variability, and the difficulty of integrating heterogeneous sources of information. As discussed in the previous sections, AI models have been developed using both static embryo images and time-lapse data, with the aim of predicting clinically relevant outcomes such as embryo quality, implantation potential, ploidy, fetal heartbeat, and live birth [13,14,18,20–26]. These approaches differ in terms of input data, model architecture, and target outcomes, but collectively they show that embryo evaluation

can be enhanced by computational methods capable of identifying complex visual and temporal patterns beyond conventional manual assessment [4,13,22,23].

At the same time, the current literature also highlights important limitations. The performance of AI models is often reported using conventional discrimination and classification metrics, such as AUC, accuracy, sensitivity, and specificity, which are useful but do not necessarily reflect the true clinical value of a model in routine IVF practice [14,28–33]. In embryo selection, the clinically relevant question is not only whether a model performs well at the population level but also whether it improves ranking decisions within the same cohort and, ultimately, shortens the path to live birth. In this regard, recently proposed approaches such as time to live birth may provide a more clinically meaningful framework for evaluating AI-assisted embryo ranking strategies [2].

Another major issue is generalizability. AI models in reproductive medicine are often developed and tested in specific laboratory settings, using locally generated datasets, proprietary imaging systems, and center-specific annotation practices. This may limit reproducibility across clinics and reduce confidence in the transportability of model performance [27,35]. Furthermore, many currently available models remain insufficiently characterized in terms of external validation, calibration, and prospective clinical impact, all of which are essential for responsible implementation in real-world IVF settings [28,41,42].

The implementation of artificial intelligence in embryo selection requires consideration of several methodological, clinical, and ethical domains. Beyond predictive performance, future studies should address issues related to validation, reporting quality, clinical endpoints, interpretability, laboratory integration, and patient trust. These key aspects are summarized in Table 3 [28,41–51].

Table 3. Key challenges and recommended directions for the clinical implementation of artificial intelligence in embryo selection. AI, artificial intelligence; IVF, in vitro fertilization; TTLB, time to live birth.

Domain	Critical Issue	Recommended Direction
External validity	Many AI models are developed on retrospective or single-center datasets, limiting their generalizability across IVF laboratories	Validate models in independent, multicenter populations before routine clinical use [27,28,35]
Performance assessment	Reported performance often focuses on AUC, accuracy, sensitivity, and specificity, while calibration and clinical utility are less consistently addressed	Evaluate discrimination together with calibration, clinical usefulness, and transparent reporting standards [28,41,42,51]
Clinical endpoints	Outcomes such as embryo quality, implantation, ploidy, or clinical pregnancy may not fully reflect the patient-centered goal of treatment	Prioritize live birth and TTLB as clinically meaningful endpoints for embryo-ranking strategies [2,43]
Interpretability	Deep learning models may provide limited insight into how predictions are generated	Develop interpretable outputs that support embryologist judgment without replacing clinical responsibility [17,44,45]
Clinical and ethical governance	Issues related to accountability, fairness, patient communication, data protection, and regulation remain incompletely defined	Implement AI within transparent governance frameworks, preserving human oversight and patient trust [41–50]

Importantly, the limitations summarized in Table 3 indicate that the clinical value of AI in embryo selection cannot be inferred from predictive performance alone. A model

with good discrimination may still have limited clinical utility if it has not been externally validated, if its probability estimates are poorly calibrated, or if the outcome used for training does not correspond to a meaningful patient-centered endpoint [28]. This is particularly relevant in IVF, where embryo selection is not only a binary prediction task but also a ranking process within a specific embryo cohort [35]. For this reason, future studies should move beyond isolated performance metrics and evaluate whether AI-assisted ranking can improve the efficiency of treatment, reduce unnecessary embryo transfers, and shorten the time to live birth [2,43].

In this context, AI-assisted embryo selection should be approached with the same level of scrutiny applied to other reproductive medicine add-ons, requiring evidence of safety, effectiveness, and clinically meaningful benefit before widespread routine implementation [52].

Another important consideration is the transition from model development to clinical implementation. Many AI systems are developed under controlled research conditions, whereas routine IVF laboratories differ in terms of incubator platforms, culture protocols, image acquisition settings, embryo annotation practices, patient characteristics, and clinical policies [27]. These differences may substantially affect model performance and limit generalizability [35]. Therefore, before AI-based embryo selection tools are incorporated into routine practice, they should be evaluated in independent clinical settings and ideally within prospective implementation studies [28]. Such studies should assess not only predictive accuracy but also workflow integration, human–AI interaction, embryologist acceptance, patient communication, and the potential impact on clinical decision-making [41,42,51].

These methodological challenges are closely linked to ethical and professional considerations. In embryo selection, AI does not operate in a neutral technical space but within a high-stakes clinical context in which decisions directly influence treatment strategy and patient expectations. For this reason, predictive performance alone is not sufficient to justify clinical adoption. Transparency, explainability, accountability, fairness, and appropriate human oversight remain essential requirements for the responsible use of AI in IVF [41–50]. These requirements also extend to informed patient communication, since couples should be made aware of the role of AI in embryo assessment, the type of data used to generate predictions, and the limits of algorithmic output in the clinical decision-making process. In addition, the risk of automation bias should be carefully considered, as embryologists and clinicians may be inclined to over-rely on AI-generated rankings, particularly when model predictions appear highly accurate but remain only partially interpretable. Another important issue concerns fairness and dataset representativeness. Because many AI models are developed on retrospective, single-center, or laboratory-specific datasets, their performance may not be equally reliable across different patient populations, clinical settings, or laboratory conditions, with the consequent risk of embedding or amplifying pre-existing biases. Data governance also remains a central ethical concern, given that AI-assisted embryo selection may involve the large-scale use of sensitive reproductive, imaging, and potentially genetic data, requiring robust safeguards for privacy, traceability, and responsible data sharing. For these reasons, the clinical integration of AI in embryo selection should remain grounded in meaningful human oversight, transparent validation, and clearly defined professional responsibility so that algorithmic support strengthens rather than replaces expert clinical judgment. Importantly, current evidence does not support replacing embryologist judgment with algorithmic output; rather, AI should be integrated as a support tool embedded within expert clinical and laboratory practice [44,45,47].

Overall, the available evidence suggests that AI has considerable potential to improve embryo selection, particularly by enabling more standardized, data-driven, and scalable evaluation strategies. However, its current clinical role should still be considered support-

ive rather than definitive. Further progress will depend not only on technical refinement but also on robust validation, transparent reporting, and careful alignment between algorithmic development and the ethical, clinical, and organizational realities of reproductive medicine [2,28,41–50].

8. Future Directions

Future developments in this field will likely move beyond isolated prediction models toward a more integrated concept of the intelligent IVF laboratory. In this setting, AI could combine multiple sources of information generated throughout the laboratory workflow, including static and time-lapse imaging, morphokinetic annotation, embryology records, patient and cycle characteristics, laboratory environment data, and treatment outcomes. The integration of these heterogeneous data streams may allow the development of more comprehensive and clinically useful prediction models, better reflecting the complexity of embryo competence and treatment success [2,28].

In parallel, AI may also contribute to laboratory quality control. Potential applications include automated image annotation, consistency checks in embryo grading, monitoring of incubator and workflow parameters, detection of deviations in laboratory performance, and support for traceability and auditing processes. However, these developments should be pursued within clearly defined ethical and regulatory boundaries, with particular attention to data quality, bias control, transparency, human supervision, and patient trust [41–50].

In conclusion, future progress in AI-assisted embryo selection should not be defined solely by increasingly accurate algorithms but by the development of clinically meaningful, ethically sound, and well-governed systems that support both embryo selection and overall laboratory quality within routine IVF practice.

9. Conclusions

Artificial intelligence represents one of the most promising developments in embryo selection, offering the possibility of more objective, reproducible, and data-driven assessment. Current evidence suggests that AI models can extract clinically relevant information from static embryo images, time-lapse imaging, morphokinetic data, and multimodal datasets. However, the available literature also shows important limitations, including heterogeneous endpoints, limited external validation, variable reporting quality, and uncertainty regarding clinical implementation.

At present, AI should be considered a support tool rather than an autonomous decision-maker in embryo selection. Its clinical value will depend on whether future studies can demonstrate not only good predictive performance but also improved embryo ranking, better workflow integration, and meaningful benefits for patients. Robust validation, transparent reporting, ethical governance, and continued embryologist oversight will be essential before AI-based embryo selection can be fully integrated into routine IVF practice.

Author Contributions: Conceptualization, M.M. and L.B.; methodology, A.C.; investigation, C.C.; resources, L.B.; data curation, E.V.; writing—original draft preparation, L.M., A.D., A.C., E.V. and C.C.; writing—review and editing, L.B.; supervision, F.G. and M.M.; project administration, M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANN	Artificial Neural Network
ART	Assisted Reproductive Technologies
AUC	Area Under the Receiver Operating Characteristic Curve
CNN	Convolutional Neural Network
CONSORT AI	Consolidated Standards of Reporting Trials Artificial Intelligence extension
DL	Deep Learning
ERICA	Embryo Ranking Intelligent Classification Algorithm
GRU	Gated Recurrent Unit
IVF	In Vitro Fertilization
LSTM	Long Short-Term Memory
MICE	Multiple Imputation by Chained Equation
ML	Machine Learning
PCA	Principal Component Analysis
SPIRIT AI	Standard Protocol Items Recommendations for Interventional Trials Artificial Intelligence extension
TRIPOD AI	Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis Artificial Intelligence extension
TTLB	Time To Live Birth

References

1. Bori, L.; Paya, E.; Alegre, L.; Viloria, T.; Remohí, J.; Naranjo, V.; Meseguer, M. Novel and conventional embryo parameters as input data for artificial neural networks: An artificial intelligence model applied for embryo selection. *Fertil. Steril.* **2020**, *114*, 1232–1241. [CrossRef] [PubMed]
2. Bori, L.; Johansen, M.; Berntsen, J.; Thorlund Parner, E.; Meseguer, M. Predicting time to live birth with deep learning embryo ranking: A novel multiple imputation approach. *Hum. Reprod.* **2025**, *40*, 1516–1525. [CrossRef] [PubMed]
3. Bori, L.; Meseguer, F.; Valera, M.A.; Galan, A.; Remohí, J.; Meseguer, M. The higher the score, the better the clinical outcome: Retrospective evaluation of automatic embryo grading as a support tool for embryo selection in IVF laboratories. *Hum. Reprod.* **2022**, *37*, 1148–1160. [CrossRef]
4. Meseguer, M.; Herrero, J.; Tejera, A.; Hilligsøe, K.M.; Ramsing, N.B.; Remohí, J. The use of morphokinetics as a predictor of embryo implantation. *Hum. Reprod.* **2011**, *26*, 2658–2671. [CrossRef]
5. Alpha Scientists in Reproductive Medicine; ESHRE Special Interest Group of Embryology. The Istanbul consensus workshop on embryo assessment: Proceedings of an expert meeting. *Hum. Reprod.* **2011**, *26*, 1270–1283. [CrossRef] [PubMed]
6. Gardner, D.K.; Schoolcraft, W.B. In vitro culture of human blastocysts. In *Towards Reproductive Certainty: Infertility and Genetics Beyond*; Jansen, R., Mortimer, D., Eds.; Parthenon Press: Carnforth, UK, 1999; pp. 378–388.
7. Gardner, D.K.; Lane, M.; Stevens, J.; Schlenker, T.; Schoolcraft, W.B. Blastocyst score affects implantation and pregnancy outcome: Towards a single blastocyst transfer. *Fertil. Steril.* **2000**, *73*, 1155–1158. [CrossRef]
8. Martínez-Granados, L.; Serrano, M.; González-Utor, A.; Ortíz, N.; Badajoz, V.; Olaya, E.; Prados, N.; Boada, M.; Castilla, J.A.; Special Interest Group in Quality of ASEBIR (Spanish Society for the Study of Reproductive Biology). Inter-laboratory agreement on embryo classification and clinical decision: Conventional morphological assessment vs. time lapse. *PLoS ONE* **2017**, *12*, e0183328. [CrossRef]
9. Paternot, G.; Devroe, J.; Debrock, S.; D'Hooghe, T.M.; Spiessens, C. Intra- and interobserver analysis in the morphological assessment of early stage embryos during an IVF procedure: A multicentre study. *Reprod. Biol. Endocrinol.* **2011**, *9*, 127. [CrossRef]
10. Lundin, K.; Park, H. Time-lapse technology for embryo culture and selection. *Upsala J. Med. Sci.* **2020**, *125*, 77–84. [CrossRef]
11. Bori, L.; Toschi, M.; Esteve, R.; Delgado, A.; Pellicer, A.; Meseguer, M. External validation of a fully automated evaluation tool: A retrospective analysis of 68,471 scored embryos. *Fertil. Steril.* **2025**, *123*, 634–643. [CrossRef]
12. Russell, S.; Norvig, P. *Artificial Intelligence: A Modern Approach*; Pearson: London, UK, 2021.
13. Khosravi, P.; Kazemi, E.; Zhan, Q.; Malmsten, J.E.; Toschi, M.; Zisimopoulos, P.; Sigaras, A.; Lavery, S.; Cooper, L.A.D.; Hickman, C.; et al. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *npj Digit. Med.* **2019**, *2*, 21. [CrossRef]

14. VerMilyea, M.; Hall, J.M.; Diakiw, S.M.; Johnston, A.; Nguyen, T.; Perugini, D.; Miller, A.; Picou, A.; Murphy, A.P.; Perugini, M. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured during IVF. *Hum. Reprod.* **2020**, *35*, 770–784. [CrossRef] [PubMed]
15. Bormann, C.L.; Kanakasabapathy, M.K.; Thirumalaraju, P.; Gupta, R.; Pooniwala, R.; Kandula, H.; Hariton, E.; Souter, I.; Dimitriadis, I.; Ramirez, L.B.; et al. Performance of a deep learning-based neural network in the selection of human blastocysts for implantation. *eLife* **2020**, *9*, e55301. [CrossRef]
16. Kragh, M.F.; Rimestad, J.; Berntsen, J.; Karstoft, H. Automatic grading of human blastocysts from time-lapse imaging. *Comput. Biol. Med.* **2019**, *115*, 103494. [CrossRef]
17. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [CrossRef] [PubMed]
18. Chávez-Badiola, A.; Flores-Saiffe-Farías, A.; Mendizabal-Ruiz, G.; Drakeley, A.J.; Cohen, J. Embryo Ranking Intelligent Classification Algorithm (ERICA): Artificial intelligence clinical assistant predicting embryo ploidy and implantation. *Reprod. BioMed. Online* **2020**, *41*, 585–593. [CrossRef]
19. Nicolielo, M.; Jacobs, C.K.; Lourenço, B.; Maffeis, M.C.; Chéles, D.S.; Duarte, M.B.; Mendes, B.A.; Moraes, V.C.; Chehin, M.B.; Alegretti, J.R.; et al. MAIA platform for routine clinical testing: An artificial intelligence embryo selection tool developed to assist embryologists. *Sci. Rep.* **2025**, *15*, 32273. [CrossRef] [PubMed]
20. Conversa, L.; Bori, L.; Cobo, A.; Meseguer, M. Testing an artificial intelligence algorithm to predict fetal heartbeat of vitrified-warmed blastocysts from a single image: Predictive ability in different settings. *Hum. Reprod.* **2024**, *39*, 2240–2248. [CrossRef]
21. Bori, L.; Domingueaz, F.; Fernandez, E.I.; del Gallego, R.; Alegre, L.; Hickman, C.; Quinonero, A.; Gouveia Nogueira, M.F.; Rocha, J.C.; Meseguer, M. An artificial intelligence model based on the proteomic profile of euploid embryos and blastocyst morphology: A preliminary study. *Reprod. BioMed. Online* **2021**, *42*, 340–350. [CrossRef]
22. Milewski, R.; Kuczyńska, A.; Stankiewicz, B.; Kuczyński, W. How much information about embryo implantation potential is included in morphokinetic data? A prediction model based on artificial neural networks and principal component analysis. *Adv. Med. Sci.* **2017**, *62*, 152–160. [CrossRef] [PubMed]
23. Paya, E.; Pulgarin, C.; Bori, L.; Colomer, A.; Naranjo, V.; Meseguer, M. Deep learning system for classification of ploidy status using time-lapse videos. *FS Sci.* **2023**, *4*, 211–218. [CrossRef]
24. Liao, Q.; Zhang, Q.; Feng, X.; Huang, H.; Xu, H.; Tian, B.; Liu, J.; Yu, Q.; Guo, N.; Liu, Q.; et al. Development of deep learning algorithms for predicting blastocyst formation and quality by time-lapse monitoring. *Commun. Biol.* **2021**, *4*, 415. [CrossRef]
25. Maekawa, R.; Kiritani, T.; Abe, T.; Nakatsui, M.; Asai, Y.; Sugino, N. AI prediction models based on time-lapse imaging for good embryos with implantation potential and euploidy. *Sci. Rep.* **2026**, *16*, 9864. [CrossRef]
26. Boucret, L.; Chabrun, F.; Boguenet, M.; Reynier, P.; Bouet, P.-E.; May-Panloup, P. Deep-learning model for embryo selection using time-lapse imaging of matched high-quality embryos. *Sci. Rep.* **2025**, *15*, 28068. [CrossRef]
27. Berman, A.; Anteby, R.; Efros, O.; Klang, E.; Soffer, S. Deep learning for embryo evaluation using time-lapse imaging: A systematic review of diagnostic test accuracy. *Am. J. Obstet. Gynecol.* **2023**, *229*, 490–501.
28. Collins, G.S.; Moons, K.G.M.; Dhiman, P.; Riley, R.D.; Beam, A.L.; Van Calster, B.; Ghassemi, M.; Liu, X.; Reitsma, J.B.; van Smeden, M.; et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **2024**, *385*, e078378. [CrossRef]
29. Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36. [CrossRef]
30. Šimundić, A.M. Measures of diagnostic accuracy: Basic definitions. *EJIFCC* **2009**, *19*, 203–211.
31. Altman, D.G.; Bland, J.M. Diagnostic tests 1: Sensitivity and specificity. *BMJ* **1994**, *308*, 1552. [CrossRef]
32. Miyagi, Y.; Habara, T.; Hirata, R.; Hayashi, N. Feasibility of deep learning for predicting live birth from a blastocyst image in patients classified by age. *Reprod. Med. Biol.* **2019**, *18*, 190–203. [CrossRef]
33. Chavez-Badiola, A.; Flores-Saiffe Farias, A.; Mendizabal-Ruiz, G.; Garcia-Sanchez, R.; Drakeley, A.J.; Garcia-Sandoval, J.P. Predicting pregnancy test results after embryo transfer by image feature extraction and analysis using machine learning. *Sci. Rep.* **2020**, *10*, 4394. [CrossRef]
34. Salih, M.; Austin, C.; Warty, R.R.; Tiktin, C.; Rolnik, D.L.; Momeni, M.; Rezatofighi, H.; Reddy, S.; Smith, V.; Vollenhoven, B.; et al. Embryo selection through artificial intelligence versus embryologists: A systematic review. *Hum. Reprod. Open* **2023**, *2023*, hoad031. [CrossRef] [PubMed]
35. Kragh, M.F.; Karstoft, H. Embryo selection with artificial intelligence: How to evaluate and compare methods? *J. Assist. Reprod. Genet.* **2021**, *38*, 1675–1689. [CrossRef]
36. Beam, A.L.; Kohane, I.S. Big data and machine learning in health care. *JAMA* **2018**, *319*, 1317–1318. [CrossRef] [PubMed]
37. Domar, A.D.; Rooney, K.; Hacker, M.R.; Sakkas, D.; Dodge, L.E. Burden of care is the primary reason why insured women terminate in vitro fertilization treatment. *Fertil. Steril.* **2018**, *109*, 1121–1126. [CrossRef] [PubMed]

38. Shen, Q.; Wang, B.; He, T.; Li, S.; Peng En Lei, J. Factors associated with discontinuation in fertility treatment: A systematic scoping review. *J. Assist. Reprod. Genet.* **2024**, *41*, 409–421. [PubMed]
39. Zhang, Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Ann. Transl. Med.* **2016**, *4*, 30.
40. Illingworth, P.J.; Venetis, C.; Gardner, D.K.; Nelson, S.M.; Berntsen, J.; Larman, M.G.; Agresta, F.; Ahitan, S.; Ahlstrom, 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.
41. Liu, X.; Cruz Rivera, S.; Moher, D.; Calvert, M.J.; Denniston, A.K.; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *Nat. Med.* **2020**, *26*, 1364–1374. [CrossRef] [PubMed]
42. Cruz Rivera, S.; Liu, X.; Chan, A.W.; Denniston, A.K.; Calvert, M.J.; SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. *BMJ* **2020**, *370*, m3210. [CrossRef] [PubMed]
43. Kieslinger, D.C.; Lambalk, C.B.; Vergouw, C.G. The inconvenient reality of AI-assisted embryo selection in IVF. *Nat. Med.* **2024**, *30*, 3059–3060. [CrossRef]
44. Afnan, M.A.M.; 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. [CrossRef] [PubMed]
45. Koplin, J.J.; Johnston, M.; Webb, A.N.S.; Whittaker, A.; Mills, C. Ethics of artificial intelligence in embryo assessment: Mapping the terrain. *Hum. Reprod.* **2025**, *40*, 179–185. [CrossRef]
46. Rolfes, V.; Bittner, U.; Gerhards, H.; Krüssel, J.S.; Fehm, T.; Ranisch, R.; Fangerau, H. Artificial intelligence in reproductive medicine: An ethical perspective. *Geburtshilfe Frauenheilkd.* **2023**, *83*, 106–115. [CrossRef]
47. Palmer, G.A.; Paredes, O.; Drakeley, A.; Chavez-Badiola, A.; Woolley, T.E.; Kaouri, K.; Cohen, J. Use and understanding of AI in the ART laboratory: An international survey. *Reprod. BioMed Online* **2025**, *50*, 104435. [CrossRef]
48. Cromack, S.C.; Lew, A.M.; Bazzetta, S.E.; Xu, S.; Walter, J.R. The perception of artificial intelligence and infertility care among patients undergoing fertility treatment. *J. Assist. Reprod. Genet.* **2025**, *42*, 855–863. [CrossRef]
49. Shevtsova, D.; Ahmed, A.; Boot, I.W.A.; Sanges, C.; Hudecek, M.; Jacobs, J.J.L.; Hort, S.; Vrijhoef, H.J.M. Trust in and acceptance of artificial intelligence applications in medicine: Mixed methods study. *JMIR Hum. Factors* **2024**, *11*, e47031. [CrossRef]
50. Pavlovic, Z.J.; Jiang, V.S.; Hariton, E. Current applications of artificial intelligence in assisted reproductive technologies through the perspective of a patient's journey. *Curr. Opin. Obstet. Gynecol.* **2024**, *36*, 211–217. [PubMed]
51. Vasey, B.; Nagendran, M.; Campbell, B.; Clifton, D.A.; Collins, G.S.; Denaxas, S.; Denniston, A.K.; Faes, L.; Geerts, B.; Ibrahim, M.; et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat. Med.* **2022**, *28*, 924–933. [CrossRef]
52. ESHRE Add-Ons Working Group. Good practice recommendations on add-ons in reproductive medicine. *Hum. Reprod.* **2023**, *38*, 2062–2104. [CrossRef]

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