

Artificial Intelligence-Assisted Embryo Selection in In Vitro Fertilisation: A Systematic Review of Model Performance and Clinical Evidence

Emmanuel Emmanuel

emmatukwasi2@gmail.com

Babcock University <https://orcid.org/0009-0002-4349-7275>

Ekwueme Precious

Adeleke University <https://orcid.org/0009-0009-5423-4420>

Systematic Review

Keywords: Artificial intelligence, embryo selection, in vitro fertilisation, assisted reproductive technology, embryo assessment, machine learning, deep learning, fertility, reproductive medicine, clinical outcomes

Posted Date: September 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11182063/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Artificial intelligence (AI) is increasingly being investigated to support embryo assessment, ranking, and selection in in vitro fertilisation (IVF), but promising model performance does not necessarily translate into improved clinical outcomes. This systematic review evaluated the current evidence on the performance and clinical evaluation of AI-assisted embryo-selection methods in IVF, with emphasis on model performance, validation, and clinical outcomes. The review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement. PubMed/MEDLINE, Scopus, Web of Science Core Collection, and Embase were searched from inception to the final search date, supplemented by reference-list and citation searching. Eligible studies involved human IVF/ICSI embryos assessed, ranked, or selected using AI-based methods. Data on study characteristics, AI approaches and inputs, embryo stage, validation strategies, model performance, and clinical outcomes were extracted, and risk of bias was assessed using PROBAST + AI, RoB 2, or ROBINS-I, as appropriate. Because of substantial methodological and clinical heterogeneity, findings were synthesised descriptively. The evidence comprised retrospective and prospective model-development and validation studies, observational and comparative clinical studies, and randomised clinical evidence. AI approaches included conventional machine learning, deep learning, convolutional neural networks, multimodal models, interpretable AI, and foundation models, using static images, time-lapse imaging, morphokinetic variables, morphology, clinical variables, genetic information, or combinations of these inputs. Reported model performance was variable, with independent-test AUCs of approximately 0.67 and clinic-holdout AUCs of 0.60–0.75 reported in one study, while iDAScore v2.0 showed AUCs of approximately 0.621–0.707 across transfer-day evaluations. Clinical evidence remained more limited. In a multicentre randomized trial involving 1,066 patients, clinical pregnancy occurred in 46.5% of patients in the deep-learning group compared with 48.2% in the conventional morphology group, and the prespecified non-inferiority criterion was not demonstrated. Observational studies reported associations between AI-based embryo ranking and implantation or pregnancy outcomes, but their non-randomized designs limit causal interpretation. Overall, AI-assisted embryo selection has advanced substantially, particularly in model development, validation, and the use of deep learning and multimodal approaches; however, evidence of model performance remains stronger than evidence of clinical effectiveness. Further independent external validation and prospective clinical studies using standardized, clinically meaningful outcomes, particularly live birth, are needed before the clinical role of AI-assisted embryo selection can be established.

Introduction

Infertility is a major reproductive health problem affecting individuals and couples worldwide. A systematic review and meta-analysis of studies conducted between 1990 and 2021 estimated pooled lifetime and 12-month infertility prevalences of 17.5% and 12.6%, respectively (Cox et al., 2022). More recent estimates from the World Health Organization indicate that approximately one in six people of reproductive age experience infertility during their lifetime (World Health Organization, 2025). In vitro fertilisation (IVF) is an established treatment for infertility, although treatment outcomes are influenced by multiple patient and treatment-related factors (Ribeiro and Sousa, 2023). Within the IVF process, selecting the embryo with the greatest likelihood of successful implantation and subsequent live birth remains an important clinical challenge (Kragh and Karstoft, 2021).

Embryo selection has traditionally relied on morphological assessment, including blastomere characteristics during cleavage-stage development and blastocyst expansion, inner cell mass, and trophectoderm characteristics at the blastocyst stage. Although morphological assessment remains central to clinical embryology, its interpretation is subject to inter- and intra-observer variation and may not fully capture the biological characteristics associated with embryo developmental competence and reproductive potential (Kragh and Karstoft, 2021). Time-lapse imaging has expanded embryo assessment by providing continuous information on embryo morphology and developmental kinetics. However, clinical trials of time-lapse-based embryo assessment have not consistently demonstrated improved reproductive

outcomes compared with conventional approaches, indicating that the availability of additional imaging information alone does not necessarily translate into greater clinical effectiveness (Bhide et al., 2024).

Artificial intelligence (AI), particularly machine learning and deep learning, has consequently been investigated as a means of automating embryo assessment, ranking embryos within a treatment cohort, and identifying patterns associated with embryo competence and reproductive outcomes. AI models have been developed using static embryo images, time-lapse image sequences, morphokinetic features, and other embryo-related data. Recent studies have evaluated these models using large multicentre datasets and independent test cohorts, extending assessment beyond initial model development towards broader validation (Berntsen et al., 2022; Theilgaard Lassen et al., 2023; Gilboa et al., 2025).

However, reported model performance does not necessarily demonstrate clinical benefit. Measures such as area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, and specificity describe predictive or classification performance under particular evaluation conditions, but they do not establish whether AI-assisted embryo selection improves reproductive outcomes. Retrospective studies may also be affected by selection bias because outcomes are generally available for embryos that were transferred, while embryos not selected for transfer are not evaluated under the same conditions. This can bias comparisons between AI-based and conventional embryo-selection approaches (Kragh and Karstoft, 2021). Consequently, performance estimates derived from retrospective or internally validated datasets should not automatically be interpreted as evidence of clinical effectiveness.

Clinical evaluation therefore requires attention to how AI models perform beyond their original development settings and whether their use has been evaluated in clinical practice. Independent external validation can provide evidence regarding model generalisability across different clinical centres, patient populations, and data settings, while prospective clinical studies and randomised trials can provide evidence on the effects of AI-assisted selection on reproductive outcomes (Kragh and Karstoft, 2021). Recent research has begun to address these levels of evidence. Multicentre validation studies have evaluated AI performance across independent or previously unseen clinical datasets (Gilboa et al., 2025), while prospective clinical research has assessed AI-assisted embryo selection in comparison with conventional morphological assessment (Wang et al., 2023). In 2024, a multicentre randomised, double-blind non-inferiority trial directly compared deep-learning-assisted embryo selection with conventional morphology-based selection, with clinical pregnancy rate as the primary endpoint (Illingworth et al., 2024). More recent multicentre research has also examined the concordance between AI-based and conventional embryo selection, although prospective evidence remains limited relative to the growing body of retrospective model-development and validation studies (Wygocki et al., 2026).

Several recent systematic reviews have addressed related but distinct aspects of AI in reproductive medicine and embryo selection. The 2023 systematic review by Salih et al. focused specifically on the performance of AI decision support compared with embryologists during embryo selection. A 2025 review by Roky et al. examined AI-assisted embryo selection more broadly, while Kaveh et al. (2025) considered AI applications in predicting and evaluating sperm and embryo quality in IVF. Other recent reviews have focused on specific clinical prediction tasks, including AI-based prediction of pregnancy outcomes (Mina et al., 2025) and live birth following IVF (Yadav et al., 2026). These reviews therefore differ in their objectives, scope, prediction targets, and approaches to evaluating AI systems. In addition, substantial heterogeneity remains across AI studies in terms of data modality, embryo developmental stage, target outcome, model architecture, study design, validation strategy, and reporting of performance measures (Kragh and Karstoft, 2021).

An updated and focused synthesis is therefore warranted to characterise the current evidence on AI-assisted embryo selection in IVF, integrating reported model performance with evidence from clinical evaluation. Such a synthesis can distinguish technical performance demonstrated under particular development or testing conditions from evidence

obtained through independent validation and clinical outcome assessment, while identifying limitations and evidence gaps relevant to the clinical application of AI-assisted embryo selection.

Rationale

The application of artificial intelligence (AI) to embryo selection has progressed substantially, with studies reporting the use of machine learning and deep learning for embryo assessment, ranking, and prediction of reproductive outcomes. However, the evidence base remains heterogeneous, with substantial differences in AI methods, data sources, embryo developmental stages, target outcomes, study designs, and approaches to model evaluation and validation (Kragh and Karstoft, 2021; Gilboa et al., 2025). The 2023 systematic review by Salih et al. specifically examined the performance of AI decision support compared with embryologists during embryo selection, but its search was conducted through January 2022 and its assessment was centred on comparative performance.

Since that review, additional clinical and multicentre evidence has emerged. A 2024 multicentre randomised trial evaluated deep-learning-assisted embryo selection against conventional morphology-based selection in 1,066 patients, while a 2025 multicentre study evaluated an AI embryo-ranking system using data from 10 IVF clinics across nine countries and included independent datasets from previously unseen clinics (Illingworth et al., 2024; Gilboa et al., 2025). More recent systematic reviews have also addressed related questions, including AI-assisted embryo selection more broadly, AI-based prediction of pregnancy outcomes, and AI-based live-birth prediction (Roky et al., 2025; Mina et al., 2025; Yadav et al., 2026).

Despite this expanding literature, the evidence remains difficult to interpret as a whole because technical model performance and clinical evidence are not necessarily equivalent. High reported performance on development or internal test datasets does not by itself establish generalisability or demonstrate improved reproductive outcomes. A focused synthesis of AI-assisted embryo selection that considers both reported model performance and the extent of clinical evaluation is therefore warranted.

Review Question

What is the current evidence on the performance and clinical evaluation of AI-assisted embryo-selection methods in in vitro fertilisation?

Aim

The aim of this systematic review was to evaluate the current evidence on the performance and clinical evaluation of AI-assisted embryo-selection methods in in vitro fertilisation.

Objectives

The objectives of this systematic review were to:

1. evaluate the reported performance of AI-assisted embryo-selection methods in in vitro fertilisation;
2. examine the validation and clinical evaluation of AI-assisted embryo-selection methods, including comparisons with conventional embryo-selection approaches where available; and
3. identify the principal limitations and evidence gaps affecting the clinical application of AI-assisted embryo selection.

Materials and Methods

Study Design

This study was conducted as a systematic review of the evidence on artificial intelligence (AI)-assisted embryo selection in in vitro fertilisation (IVF). The review focused on the reported performance of AI-assisted embryo-selection methods and the available clinical evidence evaluating their use in IVF. The review was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021).

Eligibility

Studies were eligible for inclusion if they investigated artificial intelligence-assisted assessment, ranking, or selection of human embryos generated through in vitro fertilisation (IVF) or intracytoplasmic sperm injection (ICSI). Eligible studies used AI methods to assess embryo quality, viability, ranking, or selection for transfer, or to generate information intended to inform an embryo-selection decision, and reported evidence relevant to model performance and/or clinical evaluation.

Studies were eligible when AI methods used static embryo images, time-lapse images, morphokinetic information, embryo morphology, clinical variables, or combinations of these data. Studies involving different embryo developmental stages, including cleavage-stage embryos and blastocysts, were considered eligible. Studies involving fresh or frozen embryo transfer, donor-oocyte cycles, or preimplantation genetic testing (PGT) were included when an AI-based embryo assessment, ranking, or selection component was present and was relevant to embryo selection.

Primary research studies were included regardless of study design, provided that sufficient information was available to evaluate the AI-assisted embryo-selection method or its clinical evaluation. Studies reporting model-performance measures such as area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, or other relevant measures were eligible, as were studies reporting clinical outcomes including implantation, clinical pregnancy, ongoing pregnancy, or live birth.

Studies were excluded if they focused exclusively on sperm or oocyte assessment, ovarian stimulation, fertilisation, or other assisted-reproduction processes without an embryo-selection component. Studies evaluating IVF outcomes without an AI-assisted embryo-assessment or selection component were also excluded. Non-human studies, reviews, systematic reviews, meta-analyses, editorials, commentaries, letters, and conference abstracts without sufficient primary data were excluded.

Information Sources

The literature search was conducted in four electronic bibliographic databases: PubMed/MEDLINE, Scopus, Web of Science Core Collection, and Embase. These databases were selected to provide complementary coverage of biomedical, reproductive medicine, multidisciplinary, and artificial intelligence literature relevant to artificial intelligence-assisted embryo assessment and selection in in vitro fertilisation (IVF). The search covered records from database inception to 15 June 2026, with no publication-date restriction.

Search Strategy

The search strategy combined controlled vocabulary and free-text terms representing three principal concepts: (1) embryo assessment, ranking, or selection; (2) artificial intelligence; and (3) in vitro fertilisation and assisted reproduction.

The embryo-selection concept included terms such as “embryo selection”, “embryo ranking”, “embryo assessment”, “embryo evaluation”, “embryo grading”, “embryo quality”, “embryo viability”, “embryo competence”, “blastocyst selection”,

“blastocyst assessment”, “blastocyst grading”, “embryo implantation”, “implantation potential”, and “implantation prediction”.

The artificial intelligence concept included terms such as “artificial intelligence”, “machine learning”, “deep learning”, “neural network”, “deep neural network”, “convolutional neural network”, “computer vision”, “automated embryo assessment”, “automated embryo selection”, and “automated embryo ranking”.

The IVF and assisted-reproduction concept included terms such as “in vitro fertilisation”, “in vitro fertilisation”, IVF, “intracytoplasmic sperm injection”, ICSI, “assisted reproductive technology”, and “assisted reproduction”.

Terms within each concept were combined using the Boolean operator OR, while the three concept groups were combined using AND. Controlled vocabulary terms, including relevant Medical Subject Headings (MeSH) and Emtree terms, were incorporated where applicable. The search strategy was adapted to the indexing, controlled vocabulary, and syntax of each database. The complete search strategies for PubMed/MEDLINE, Scopus, Web of Science Core Collection, and Embase were developed using database-specific search terms and Boolean operators

Study Selection

Records identified through the four database searches were imported into a reference-management system, and 115 duplicate records were removed. Titles and abstracts were screened against the predefined eligibility criteria, followed by full-text assessment of potentially eligible studies. Two reviewers independently assessed records at both stages, and disagreements were resolved through discussion and consensus.

The study-selection process was presented using a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 flow diagram. Where multiple publications were identified from the same underlying study population or dataset, the reports were assessed for potential overlap and linked where appropriate to avoid double-counting of studies or participants.

Data Extraction

A standardised data-extraction form was developed to ensure consistent collection of information from the included studies. Two reviewers independently extracted the data, and discrepancies were resolved through discussion and consensus.

The following information was extracted: publication year, country, study design, clinical setting, number of centres, sample size, embryo developmental stage, type and source of input data, AI approach, model objective, and comparator where applicable. Model-development and validation characteristics were also recorded, including data-partitioning methods, internal or external validation, independent test datasets, and reported performance measures.

For studies reporting clinical evaluation, information was extracted on the clinical study design, embryo-selection approach, comparator, number of patients or treatment cycles, and reproductive outcomes, including implantation, clinical pregnancy, ongoing pregnancy, miscarriage, and live birth where reported. Performance measures such as area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and other relevant measures were extracted as reported. Results obtained from development, internal-validation, test, and external-validation datasets were recorded separately where this distinction was available.

Risk of Bias

Risk of bias was assessed independently by two reviewers, with disagreements resolved through discussion and consensus. The assessment tool was selected according to the design and purpose of each included study. Prediction-model and AI algorithm studies were assessed using PROBAST + AI, which evaluates risk of bias and applicability in

studies of prediction models developed or evaluated using regression or artificial intelligence methods (Moons et al., 2025).

Particular attention was given to participant selection, predictor and outcome measurement, data partitioning, handling of missing data, overfitting, sample size, model development, validation procedures, and the independence of evaluation datasets. Randomised clinical trials were assessed using the Cochrane Risk of Bias 2 (RoB 2) tool (Sterne et al., 2019). Non-randomised comparative clinical studies evaluating the effects of AI-assisted embryo selection were assessed using ROBINS-I (Sterne et al., 2016), where applicable.

Data Synthesis

The findings were synthesised descriptively because of substantial clinical and methodological heterogeneity among the included studies. Studies were grouped according to AI approach, embryo developmental stage, input modality, prediction or selection task, study design, validation strategy, and clinical evaluation.

Model-performance findings were summarised using the performance measures reported by individual studies, including AUC, accuracy, sensitivity, specificity, and other relevant measures. Performance estimates were interpreted according to the dataset on which they were obtained, with particular distinction between development, internal-validation, test, and independent external-validation datasets.

Clinical evidence was synthesised according to study design, comparator, and reproductive outcome. Where studies compared AI-assisted selection with conventional embryo selection, the reported clinical outcomes and performance measures were summarised without assuming that estimates from different studies were directly comparable. Findings from prospective and randomised clinical evaluations were considered separately from retrospective model-development and validation studies.

Because of differences in AI methods, input data, embryo developmental stages, target outcomes, validation procedures, and outcome definitions, statistical pooling was not undertaken. Results were presented in evidence tables and, where appropriate, graphical summaries were used to facilitate comparison across studies.

Results

Search and Study Selection

The literature searches identified 2,715 records across the searched sources. After removal of 115 duplicate records, the remaining 2,600 records were screened by title and abstract. Preliminary screening excluded 2,461 records that were clearly outside the predefined eligibility criteria, leaving 139 potentially relevant records for full-text assessment. Of these, 108 reports were excluded after full-text assessment, resulting in 31 studies being included in the systematic review.

The complete flow of records from database searching, full-text assessment, and final inclusion is presented in the PRISMA 2020 flow diagram. Records and reports were distinguished from individual studies where multiple publications related to the same underlying study.

Characteristics of the Evidence

The identified evidence was methodologically heterogeneous and included model-development studies, retrospective and prospective validation studies, observational clinical studies, comparative studies, and randomised clinical evidence. AI approaches included artificial neural networks, conventional machine learning, deep learning, convolutional neural networks, computer vision, interpretable AI, multimodal models, and foundation models. Input data included static embryo images, time-lapse image sequences, morphokinetic variables, morphology, clinical variables, and multimodal

combinations. The primary outcomes also varied considerably, including embryo quality, implantation potential, clinical pregnancy, live birth, embryo euploidy, and embryo ranking. The main characteristics of the included studies are summarised in Table 1.

Table 1
Characteristics of included studies

Study	Year	Study design / setting		Embryo stage	AI input		Primary objective
Cheredath et al.	2023	Retrospective model-development/validation study		Blastocyst	Embryo imaging		Embryo assessment
Danardono et al.	2023	Retrospective model-development study		Blastocyst	Embryo morphology		Predict embryo ploidy
Kato et al.	2023	Retrospective model-development/validation study		Blastocyst	Morphokinetic and morphological data		Predict blastocyst euploidy
Valera et al.	2023	Clinical validation study		Cleavage-stage embryos	Embryo images		Automated embryo classification
Zheng et al.	2023	Retrospective model-development study		Blastocyst	Embryo images		Quantitative blastocyst assessment
Cimadomo et al.	2023	Preclinical validation study		Blastocyst	Embryo imaging		Embryo grading during PGT-A
Barnes et al.	2023	Retrospective validation study		Blastocyst	Embryo images		Non-invasive prediction of ploidy
Zhu et al.	2023	External validation study		Day-3 embryos	Embryo images		Validate embryo-selection model
Illingworth et al.	2024	Multicentre randomised double-blind non-inferiority trial		Blastocyst	Time-lapse imaging/deep learning		Compare AI-assisted and conventional embryo selection
Sun et al.	2024	Retrospective/clinical evaluation study		Embryos	Embryo imaging and clinical information		Predict IVF embryo outcomes
Dissler et al.	2024	Clinical observational study		Blastocyst	Embryo imaging		AI-assisted ranking of sibling embryos
Papamentzelopoulou et al.	2024	Retrospective validation study		Blastocyst	Morphokinetic data		Predict live birth
Miyagi et al.	2024	Retrospective model-development/validation study		Blastocyst	3D embryo images and conventional parameters		Embryo assessment
Chavez-Badiola et al.	2024	Retrospective clinical evaluation study		Blastocyst	Static embryo images		Predict pregnancy loss
Li et al.	2024	Retrospective model-development/validation study		Blastocyst	Embryo images		Non-invasive chromosome screening
Luong et al.	2024	Retrospective model-development/validation study		Blastocyst	Embryo images		Explainable AI for ploidy prediction

Study	Year	Study design / setting		Embryo stage	AI input		Primary objective
Goswami et al.	2024	Preclinical/model-development study		Embryos	Label-free optical imaging		Assess embryo health
Gilboa et al.	2025	Multicentre validation study		Blastocyst	Embryo imaging		Development and multicentre validation of AI
Nicolielo et al.	2025	Prospective clinical study		Blastocyst	Embryo images		Clinical testing of MAIA embryo-selection tool
Boucret et al.	2025	Retrospective model-development/validation study		Blastocyst	Time-lapse imaging		Embryo selection
Liu et al.	2025	Retrospective validation study		Blastocyst	Embryo images		Differentiate similarly graded blastocysts
Bori et al.	2025	Retrospective prediction study		Blastocyst	Embryo ranking/deep learning		Predict time to live birth
Palmer et al.	2025	Clinical evaluation study		Euploid embryos	Embryo imaging		AI-guided embryo selection
Rajendran et al.	2025	Large-scale development study		Embryos/blastocysts	Time-lapse images		Develop a foundation model for IVF
Thirumalaraju et al.	2026	Retrospective validation/stability study		Blastocyst	Embryo imaging		Evaluate AI model stability and reliability
Rajendran et al.	2026	Trial-emulation/clinical validation study		Blastocyst	Time-lapse imaging		Evaluate clinical efficacy of a foundation model
Wygocki et al.	2026	Retrospective observational study		Blastocyst	Embryo images		Compare AI-based and conventional selection
Berntsen et al.	2026	Retrospective observational study		Blastocyst	AI embryo ranking		Evaluate embryologist–AI concordance
Giménez-Rodríguez et al.	2026	Retrospective AI/model-development study		Blastocyst	Embryo morphology/imaging		Estimate ploidy and pregnancy potential
Maekawa et al.	2026	Retrospective model-development/validation study		Blastocyst	Time-lapse imaging		Predict implantation potential and euploidy
Zhu et al.	2026	Retrospective clinical cohort		Embryos	AI-based embryo assessment		Evaluate non-invasive embryo selection

AI-Assisted Embryo-Selection Methods

Early studies used artificial neural networks and machine-learning models to combine embryo morphology, morphokinetic parameters, and clinical information. Subsequent studies increasingly adopted deep-learning approaches capable of extracting image features directly from static or time-lapse embryo images.

Time-lapse approaches represented a substantial proportion of the more recent evidence. Large multicentre datasets were used to develop and validate systems capable of ranking embryos according to implantation or live-birth potential. More recent approaches have incorporated multimodal information and foundation-model architectures trained using large image datasets.

Model Performance

Performance varied between models and validation populations. Berntsen et al. reported an independent-test AUC of approximately 0.67 for implantation-related embryo selection, with clinic-holdout AUC values ranging from approximately 0.60 to 0.75. Berntsen et al. (2026) subsequently evaluated iDAScore v2.0 using 181,428 embryos from 22 IVF clinics and reported AUC values ranging from 0.621 to 0.707 depending on transfer day.

Other studies demonstrated the ability of AI systems to predict embryo viability, implantation, live birth, or ploidy from static and time-lapse images. However, differences in outcome definition, dataset composition, embryo stage, imaging technology, and validation strategy limited direct comparison between reported performance estimates. Recent evidence also identified model stability as a separate consideration from discrimination. Thirumalaraju et al. (2026) found meaningful variation in embryo rankings generated by repeated deep-learning models, indicating that model stability should be considered alongside conventional performance metrics. Key characteristics of the AI approaches, input data, validation strategies, and reported model-performance measures are summarised in Table 2.

Table 2
AI approaches, inputs, validation, and model-performance evidence

Study	AI approach / model type	Main input	Prediction / selection task	Validation approach	Main performance finding
Berntsen et al.	Deep learning	Time-lapse sequences	Embryo selection	Independent test and clinic-holdout validation	Independent-test AUC approximately 0.67; clinic-holdout AUC approximately 0.60–0.75
Danardono et al.	Computer-assisted AI	Embryo morphology	Ploidy classification	Model development/validation	Demonstrated classification capability
Kato et al.	AI / machine learning	Morphokinetic and morphological data	Euploidy prediction	Model development/validation	Demonstrated predictive performance for blastocyst euploidy
Valera et al.	Automated AI classification	Embryo images	Cleavage-stage embryo classification	Clinical validation	Demonstrated feasibility of automated classification
Barnes et al.	AI/deep learning	Embryo images	Blastocyst ploidy	Validation	Demonstrated non-invasive ploidy prediction
Zhu et al.	AI model	Day-3 embryo images	Embryo selection	External validation	External validation supported model applicability
Papamentzelopoulou et al.	AI / machine learning	Morphokinetic annotation	Live-birth prediction	Validation	Demonstrated predictive value for live birth
Miyagi et al.	Dual AI system	3D blastocyst images + conventional variables	Embryo assessment	Model development /validation	Demonstrated potential benefit from combining imaging and conventional variables
Li et al.	Machine learning	Embryo imaging	Chromosome screening /grading	Validation	Demonstrated predictive classification of chromosome status
Luong et al.	Explainable AI	Embryo images	Ploidy prediction	Validation	Demonstrated ploidy prediction with interpretable features

Study	AI approach / model type	Main input	Prediction / selection task	Validation approach	Main performance finding
Goswami et al.	Machine learning	Label-free optical imaging	Embryo health assessment	Preclinical validation	Demonstrated feasibility of label-free embryo assessment
Gilboa et al.	AI/deep learning	Embryo imaging	Embryo evaluation	Multicentre validation	Demonstrated a methodological framework for reliable multicentre AI validation
Boucret et al.	Deep learning	Time-lapse imaging	Embryo selection	Model development/validation	Demonstrated predictive performance using time-lapse data
Liu et al.	Interpretable AI	Embryo images	Ranking similarly graded blastocysts	Validation	Demonstrated discrimination among similarly graded embryos
Rajendran et al.	Foundation model	Time-lapse images	IVF embryo assessment	Large-scale model development	Foundation-model approach trained using approximately 18 million time-lapse images
Thirumalaraju et al.	AI models	Embryo imaging	Embryo selection	Stability/reliability evaluation	Demonstrated that model performance can vary across conditions, highlighting stability as a validation issue
Giménez-Rodríguez et al.	AI	Embryo imaging/morphology	Ploidy and pregnancy potential	Model development/validation	Explored prediction beyond morphology alone
Maekawa et al.	AI/deep learning	Time-lapse imaging	Implantation potential and euploidy	Model development/validation	Demonstrated prediction of implantation potential and euploidy
Zhu et al.	AI-based predictive model	Embryo assessment	Non-invasive embryo selection	Clinical cohort evaluation	Reported accuracy 84.3%, sensitivity 82.1%, specificity 79.8%, and AUC 0.89

Clinical Evidence

Clinical evidence was less extensive than evidence of model development and validation.

The most prominent randomised evidence was provided by Illingworth et al., who conducted a multicentre randomized, double-blind non-inferiority trial involving 1,066 IVF patients across 14 clinics. Clinical pregnancy occurred in 46.5% of patients in the deep-learning group and 48.2% in the conventional morphology group. The prespecified non-inferiority criterion was not demonstrated.

Other clinical studies reported associations between AI-assisted selection and reproductive outcomes. Wang et al. reported implantation rates of 80.87% with AI-assisted selection compared with 68.15% with conventional assessment in a prospective single-centre cohort, although the study was not randomised. Dissler et al. evaluated AI-assisted ranking of sibling embryos and reported potential improvement in first-cycle pregnancy selection, although part of the analysis involved counterfactual ranking rather than randomised clinical allocation. More recent studies have extended evaluation into prospective and multicentre settings. In the prospective MAIA study, the AI tool achieved an overall accuracy of 66.5% and an elective-transfer accuracy of 70.1%. Foundation-model approaches have also progressed from large-scale model development to multicentre clinical-efficacy evaluation. The principal clinical evaluations of AI-assisted embryo selection are summarised in Table 3.

Table 3
Clinical evaluation and translational evidence for AI-assisted embryo selection

Study	Clinical design	AI-assisted approach	Comparator / reference	Main clinical finding	Interpretation
Wang et al. 2023	Prospective single-centre cohort	Interpretable AI-assisted embryo selection	Conventional selection	Implantation was 80.87% with AI-assisted selection versus 68.15% with conventional selection	Association supports further prospective evaluation but does not establish causal benefit
Illingworth et al. 2024	Multicentre randomised double-blind non-inferiority trial; 1,066 patients, 14 clinics	Deep-learning embryo selection	Manual morphology-based selection	Clinical pregnancy: 46.5% versus 48.2%; non-inferiority was not demonstrated	Randomized evidence did not demonstrate non-inferiority of AI selection
Dissler et al. 2024	Clinical observational evaluation	AI-assisted ranking of sibling embryos	Conventional ranking/selection	Reported potential improvement in first-cycle pregnancy outcomes	Observational/counterfactual evidence requiring prospective confirmation
Nicolielo et al. 2025	Prospective clinical study; 200 single-embryo transfers	MAIA AI embryo-selection tool	Clinical embryo-selection process	Overall accuracy 66.5%; elective-transfer accuracy 70.1%	Demonstrates clinical feasibility but does not by itself establish improved live-birth outcomes
Gilboa et al. 2025	Multicentre validation	AI embryo evaluation framework	Multicentre/clinical assessment	Demonstrated multicentre evaluation of AI reliability	Supports external validation and methodological standardisation
Palmer et al. 2025	Clinical evaluation	AI-guided feedback for euploid embryo selection	Conventional embryologist assessment	Evaluated whether AI guidance could support selection among euploid embryos	Evidence concerns decision support rather than replacement of embryologists
Rajendran et al. 2026	Trial-emulation clinical evaluation	Foundation AI model for embryo selection	Conventional selection strategy	Evaluated clinical efficacy using trial-emulation methodology	Provides clinical-effectiveness evidence but is not equivalent to a randomized trial
Wygocki et al. 2026	Retrospective observational study	AI-based embryo ranking	Conventional embryo selection	Evaluated concordance between AI and	Demonstrates agreement/concordance rather than superiority

Study	Clinical design	AI-assisted approach	Comparator / reference	Main clinical finding	Interpretation
				conventional selection	
Berntsen et al. 2026	Retrospective observational study; 1,321 PGT-A cycles	AI embryo-ranking algorithm	Embryologist selection	Overall AI–embryologist concordance was 58.6%; fetal-heartbeat outcomes were also compared between concordant and discordant selections	Suggests substantial selection differences and supports evaluation of AI as an adjunct
Zhu et al. 2026	Retrospective clinical cohort; 300 IVF cycles, 742 embryos	AI-based non-invasive embryo selection	Clinical embryo assessment	Reported implantation 60.7%, clinical pregnancy 56.0%, ongoing pregnancy 51.3%, and live birth 48.7%; AI score was associated with clinical outcomes	Provides clinical predictive evidence but remains observational

External Validation and Generalisability

External validation was present in several studies, including evaluations across multiple IVF clinics. Nevertheless, validation strategies remained inconsistent across the evidence base.

The performance of AI-assisted embryo-selection systems may depend on patient populations, embryo characteristics, laboratory procedures, imaging platforms and outcome definitions. Consequently, performance obtained from a single development dataset cannot necessarily be assumed to generalise to other IVF settings. Recent studies have further emphasised that generalisability should be assessed not only through discrimination but also through measures such as ranking stability and agreement with embryologists.

Overall Evidence Pattern

Overall, the evidence demonstrates substantial progress in AI-assisted embryo assessment and selection. AI systems can provide automated embryo scores and rankings and can demonstrate clinically relevant associations with implantation, pregnancy, live birth and embryo characteristics.

However, evidence of model performance remains stronger than evidence of clinical effectiveness. The randomised evidence currently available has not demonstrated non-inferiority of deep-learning embryo selection to conventional morphology-based selection for clinical pregnancy.

The evidence therefore supports continued clinical evaluation of AI-assisted embryo selection while highlighting the importance of independent validation, prospective evaluation, standardised clinical outcomes, and assessment of model stability.

Discussion

This systematic review identified a rapidly developing body of evidence evaluating artificial intelligence-assisted embryo selection in in vitro fertilisation. The literature has progressed from early machine-learning and artificial-neural-network approaches towards deep-learning systems based on static and time-lapse imaging, multimodal models, interpretable AI, and foundation models trained on very large embryo-image datasets. A central finding was the difference between evidence of model performance and evidence of clinical effectiveness. Numerous studies demonstrated that AI systems can classify, score, or rank embryos and achieve measurable discrimination for implantation, pregnancy, live birth, embryo quality, or ploidy-related outcomes.

However, strong model performance does not by itself establish that AI-assisted embryo selection improves clinical IVF outcomes. This distinction is illustrated by the multicentre randomised, double-blind non-inferiority trial by Illingworth et al., which evaluated 1,066 patients across 14 IVF clinics. Clinical pregnancy occurred in 46.5% of patients in the deep-learning group compared with 48.2% in the conventional morphology group, and non-inferiority was not demonstrated. These findings illustrate the importance of evaluating AI-assisted embryo-selection systems using patient-level clinical outcomes, rather than relying solely on algorithmic discrimination measures.

The observational literature provides additional evidence that AI-derived embryo scores can be associated with implantation, pregnancy, live birth and embryo characteristics. However, such associations do not establish causality. Retrospective studies may be affected by patient selection, embryo availability, treatment decisions and other clinical factors.

Generalisability also remains an important consideration. Several models have been evaluated across multiple clinics, but the evidence remains heterogeneous with respect to patient populations, laboratory procedures, imaging platforms and outcome definitions. Consequently, performance observed in one setting cannot automatically be assumed to generalise to all IVF laboratories. Recent studies have also highlighted the importance of model stability. Different models trained on similar data can produce different embryo rankings, raising concerns about the reproducibility of AI-based selection decisions. Thus, AI-assisted embryo-selection systems should be evaluated not only according to how accurately they predict clinical outcomes, but also according to how consistently they generate embryo rankings.

The emergence of foundation models represents a further stage in the development of AI-assisted embryo selection. Training on very large time-lapse image datasets may allow models to learn broader representations of embryo development. However, large training datasets do not eliminate the need for independent validation and prospective clinical evaluation. The current evidence therefore supports viewing AI primarily as a potential decision-support technology rather than automatically as a replacement for embryologists. Clinical implementation requires evidence that the use of the AI system changes clinical decision-making in a beneficial and reproducible way and ultimately improves outcomes that matter to patients.

Strengths and Limitations of the Evidence

The evidence base has several strengths, including increasing sample sizes, multicentre validation, diverse AI approaches, independent validation studies, prospective clinical testing and the emergence of randomised clinical evidence.

However, substantial heterogeneity limits direct comparison between studies. Studies differ in embryo stage, imaging modality, AI architecture, target outcome, patient population, laboratory setting and validation strategy. The predominance of retrospective evidence also limits causal interpretation.

Another limitation is that some studies evaluate surrogate or intermediate outcomes such as embryo grade, ranking agreement or predicted ploidy rather than live birth. Although these outcomes may be clinically relevant, they cannot automatically be considered equivalent to improved reproductive outcomes.

Implications for Future Research

Future research should place greater emphasis on independent external validation and prospective clinical evaluation. Studies should clearly distinguish model-development performance from clinical effectiveness and should report clinically meaningful outcomes, particularly live birth.

Future studies should also assess calibration, model stability, transportability across IVF centres, performance across relevant patient and embryo subgroups, and the influence of imaging and laboratory conditions.

Comparative studies should specify how AI influences the embryologist's final decision and should avoid treating agreement with an algorithm as equivalent to improved patient outcomes. Randomised or otherwise carefully controlled prospective evaluations will remain important for determining whether AI-assisted selection provides clinical benefit.

Conclusion

Artificial intelligence-assisted embryo selection has progressed substantially from early machine-learning approaches to sophisticated deep-learning, multimodal and foundation-model systems. Current evidence demonstrates that AI can analyse embryo images and related data and can provide clinically relevant embryo rankings and predictions. However, evidence for model performance is currently stronger than evidence for clinical effectiveness. The available randomised evidence has not demonstrated non-inferiority of deep-learning embryo selection to conventional morphology-based selection for clinical pregnancy, while much of the remaining clinical evidence is retrospective or observational.

Further independent, prospective and clinically outcome-focused evaluation is therefore required to determine the effectiveness, reliability and generalisability of AI-assisted embryo selection across routine IVF settings.

References

1. Barnes J et al (2023) Non-invasive artificial intelligence prediction of human blastocyst ploidy. *Lancet Digit Health*. 10.1016/S2589-7500(22)00213-8
2. Berntsen J et al (2026) Embryologist experience affects concordance with an artificial intelligence embryo ranking algorithm: benefit of artificial intelligence assistance. *F&S Rep* 7(3):188–194. 10.1016/j.xfre.2026.04.002
3. Berntsen J, Rimestad J, Theilgaard Lassen J, Tran D, Kragh MF (2022) Robust and generalizable embryo selection based on artificial intelligence and time-lapse image sequences. *PLoS ONE* 17(2):e0262661. 10.1371/journal.pone.0262661
4. Bhide P, Chan DY, Lanz D, Alqawasmeh O, Barry E, Baxter D, Gonzalez Carreras F, Choudhury Y, Cheong Y, Chung JPW, Collins B, Cong L, Doidge S, Heighway J, Patel D, Pardo MC, Rattos A, Wright A, Dodds J, Perez T, Khan KS, Thangaratinam S (2024) Clinical effectiveness and safety of time-lapse imaging systems for embryo incubation and selection in in-vitro fertilisation treatment (TILT): a multicentre, three-parallel-group, double-blind, randomised controlled trial. *Lancet* 404(10449):256–265. 10.1016/S0140-6736(24)00816-X

5. Bori L, Johansen MN, Berntsen J, Parner ET, Meseguer M (2025) Predicting time to live birth with deep learning embryo ranking: a novel multiple imputation approach. *Hum Reprod* 40(8):1516–1525. 10.1093/humrep/deaf102
6. Bormann CL, Kanakasabapathy MK, Yanez A, Miller D, Zaninovic N, Malmsten JE, Sakkas D, Shafiee H (2020) Performance of a deep learning based neural network in the selection of human blastocysts', *eLife*. 9:e55301. 10.7554/eLife.55301
7. Boucret L et al (2025) Deep-learning model for embryo selection using time-lapse imaging. *Sci Rep*. 10.1038/s41598-025-10531-y
8. Chavez-Badiola A et al (2024) Artificial intelligence embryo selection using static images to predict pregnancy loss. *Reprod Biomed Online* 49:103934. 10.1016/j.rbmo.2024.103934
9. Cheredath A, Uppangala S, Jijo CSA, Kumar ARVL, Joseph P, Kalthur DGANG, G. and, Adiga SK (2023) Combining machine learning with metabolomic and embryologic data improves embryo implantation prediction. *Reproductive Sci* 30(4):1006–1016. 10.1007/s43032-022-01071-1
10. Cimadomo D, Innocenti F, Taggi M, Saturno G, Campitiello MR, Guido M, Vaiarelli A, Ubaldi FM, Rienzi L (2023) Pre-clinical validation of a deep-learning based embryo grading system during preimplantation genetic testing for aneuploidy. *J Clin Med* 12(5):1806. 10.3390/jcm12051806
11. Cox CM, Thoma ME, Tchangalova N, Mburu G, Bornstein MJ, Johnson CL, Kiarie J (2022) 'Infertility prevalence and the methods of estimation from 1990 to 2021: a systematic review and meta-analysis', *Human Reproduction Open*, 2022(4), hoac051. 10.1093/hropen/hoac051
12. Danardono D et al (2023) Computer-assisted morphology assessment for embryo ploidy classification. *J Assist Reprod Genet*. 10.1016/j.xagr.2023.100209
13. Diakiw SM, Hall JMM, VerMilyea M, Lim AXY, Quangkananurug W, Chanchamroen S, Bankowski B, Stones R, Storr A, Miller A, Adaniya G, van Tol R, Hanson R, Aizpurua J, Giardini L, Johnston A, Van Nguyen T, Dakka MA, Perugini D, Perugini M (2022) An artificial intelligence model correlated with morphological and genetic features of blastocyst quality improves ranking of viable embryos. *Reprod Biomed Online* 45(6):1105–1117. 10.1016/j.rbmo.2022.07.018
14. Dissler N, Nogueira D, Keppi B, Sanguinet P, Ozanon C, Geoffroy-Siraudin C, Pollet-Villard X, Boussommier-Calleja A (2024) Artificial intelligence-powered assisted ranking of sibling embryos to increase first cycle pregnancy rate. *Reprod Biomed Online* 49(1):103887. 10.1016/j.rbmo.2024.103887
15. Gilboa D, Garg A, Shapiro M, Meseguer M, Amar Y, Lustgarten N, Desai N, Shavit T, Silva V, Papatheodorou A, Chatziparasidou A, Angras S, Lee JH, Thiel L, Curchoe CL, Tauber Y, Seidman DS (2025) Application of a methodological framework for the development and multicenter validation of reliable artificial intelligence in embryo evaluation. *Reproductive Biology Endocrinol* 23:16. 10.1186/s12958-025-01351-w
16. Giménez-Rodríguez A et al (2026) Bridging the gap between embryo euploidy, pregnancy potential and morphology using artificial intelligence for ploidy estimation. *Reprod Biomed Online*. 10.1016/j.rbmo.2025.105408
17. Goswami N, Winston N, Choi W, Lai NZE, Arcanjo RB, Chen X, Sobh N, Nowak RA, Anastasio MA, Popescu G (2024) EVATOM: an optical, label-free, machine learning assisted embryo health assessment tool. *Commun Biology* 7:268. 10.1038/s42003-024-05960-w
18. Illingworth PJ, Venetis C, Gardner DK et al (2024) Deep learning versus manual morphology-based embryo selection in IVF: a randomized, double-blind noninferiority trial. *Nat Med* 30:3114–3120. 10.1038/s41591-024-03166-5
19. Kato K et al (2023) Artificial intelligence/morphokinetic/morphological models for prediction of blastocyst euploidy. *Reprod Biomed Online*. 10.1016/j.rbmo.2022.09.010
20. Kaveh S, Ghaffari A, Sohrabei S (2025) Investigating artificial intelligence in predicting and evaluating sperm and embryo quality in the in vitro fertilization (IVF): a systematic review. *Discover Artif Intell* 5:183. 10.1007/s44163-025-00420-8

21. Khosravi P, Kazemi E, Zhan Q, Malmsten JE, Toschi M, Zisimopoulos P, Sigaras A, Lavery S, Cooper LAD, Hickman C, Meseguer M, Rosenwaks Z (2019) Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *npj Digit Med* 2:21. 10.1038/s41746-019-0096-y
22. Kragh MF, Karstoft H (2021) Embryo selection with artificial intelligence: how to evaluate and compare methods? *J Assist Reprod Genet* 38:1675–1689. 10.1007/s10815-021-02254-6
23. Lefebvre C, Glanville J, Briscoe S, Featherstone R, Littlewood A, Metzendorf M-I, Noel-Storr A, Paynter R, Rader T, Thomas J, Wieland LS (2025) 'Chapter 4: Searching for and selecting studies', in Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J. and Welch, V.A. (eds.) *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.5.1. Cochrane. doi/online resource available through Cochrane
24. Li Y et al (2024) Machine-learning guided non-invasive chromosome screening grading system. *Reproductive Biology Endocrinol*. 10.1186/s12958-024-01231-9
25. Liu Y et al (2025) Interpretable artificial intelligence to differentiate similarly graded blastocysts. *Hum Reprod*. 10.1093/humrep/deaf066
26. Luong M et al (2024) Explainable artificial intelligence for embryo ploidy prediction. *J Assist Reprod Genet*. 10.1007/s10815-024-03178-7
27. Maekawa M et al (2026) Artificial intelligence prediction models from time-lapse imaging for implantation potential and euploidy. *Sci Rep*. 10.1038/s41598-026-40917-5
28. Mina A, Younesi M, Doohandeh T, Darzi S, Ajabi Ardehjani N, Sheibani S, Hosseini-rad H, Valizadeh R (2025) Predicting pregnancy outcomes in IVF cycles: a systematic review and diagnostic meta-analysis of artificial intelligence in embryo assessment. *Contracept Reproductive Med* 10:59. 10.1186/s40834-025-00400-4
29. Miyagi M et al (2024) Dual artificial intelligence system using three-dimensional blastocyst images and conventional parameters. *Reproductive Biology Endocrinol*. 10.1002/rmb2.12612
30. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, Beam AL, Van Calster B, Celi LA, Denaxas S et al (2025) PROBAST + AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ* 388:e082505. 10.1136/bmj-2024-082505
31. Nicolielo M et al (2025) Clinical testing of an artificial intelligence embryo-selection tool (MAIA)', *Scientific Reports*. 10.1038/s41598-025-17755-y
32. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE et al (2021) 'The PRISMA 2020 statement: an updated guideline for reporting systematic reviews', *BMJ*, 372, n71. 10.1136/bmj.n71
33. Palmer R et al (2025) Artificial intelligence-guided feedback for selection of euploid embryos. *Reprod Biomed Online*. 10.1016/j.rbmo.2025.104990
34. Papamentzelopoulou A et al (2024) Artificial intelligence model versus manual morphokinetic annotation for live-birth prediction. *Reproductive Biology Endocrinol*. 10.1186/s12958-024-01198-7
35. Rajendran P et al (2025) A foundational model for in vitro fertilization trained on 18 million time-lapse images. *Nat Commun*. 10.1038/s41467-025-61116-2
36. Rajendran P et al (2026) Trial emulation for validating the clinical efficacy of a foundational AI model in embryo selection. *npj Digit Med* 9:489. 10.1038/s41746-026-02672-9
37. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, Koffel JB, Group PRISMA-S (2021) PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *J Med Libr Association* 109(2):174–200. 10.5195/jmla.2021.962
38. Ribeiro S, Sousa M (2023) In Vitro Fertilisation and Intracytoplasmic Sperm Injection predictive factors: a review of the effect of female age, ovarian reserve, male age, and male factor on IVF/ICSI treatment outcomes. *JBRA Assist*

39. Roky MAB, Ray AS, Saad AM (2025) Advancing embryo selection in artificial intelligence-assisted reproductive technologies: a systematic review. *Artif Intell Health* 2(3):1–23. 10.36922/aih.7170
40. Salih M, Austin C, Warty RR, Tiktin C, Rolnik DL, Momeni M, Rezaatofghi H, Reddy S, Smith V, Vollenhoven B, Horta F (2023) 'Embryo selection through artificial intelligence versus embryologists: a systematic review', *Human Reproduction Open*, 2023(3), hoad031. 10.1093/hropen/hoad031
41. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I, Carpenter JR, Chan A-W, Churchill R, Deeks JJ, Hróbjartsson A, Kirkham J, Jüni P, Loke YK, Pigott TD, Ramsay CR, Regidor E, Rothstein HR, Sandhu L, Santaguida PL, Schünemann HJ, Shea B, Shrier I, Tilling K, Turner L, Valentine JC, Wager E, Higgins JPT (2016) 'ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions', *BMJ*, 355, i4919. 10.1136/bmj.i4919
42. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng H-Y, Corbett MS, Eldridge SM, Emberson JR, Hernán MA, Hopewell S, Hróbjartsson A, Junqueira DR, Jüni P, Kirkham JJ, Lasserson T, Li T, McAleenan A, Reeves BC, Shepperd S, Shrier I, Stewart LA, Tilling K, White IR, Whiting PF, Higgins JPT (2019) 'RoB 2: a revised tool for assessing risk of bias in randomised trials', *BMJ*, 366, l4898. 10.1136/bmj.l4898
43. Sun L, Li J, Zeng S, Luo Q, Miao H, Liang Y, Cheng L, Sun Z, Tai WH, Han Y, Yin Y, Wu K, Zhang K (2024) Artificial intelligence system for outcome evaluations of human in vitro fertilization-derived embryos. *Chin Med J* 137(16):1939–1949. 10.1097/CM9.0000000000003162
44. Theilgaard Lassen J, Fly Kragh M, Rimestad J, Nygård Johansen M, Berntsen J (2023) Development and validation of deep learning based embryo selection across multiple days of transfer. *Sci Rep* 13:4235. 10.1038/s41598-023-31136-3
45. Thirumalaraju P, Kanakasabapathy MK, Kandula H, Kandula T, Reddy Katkuri AV, Cipriano C, Malmsten JE, Zaninovic N, Bormann CL, Shafiee H (2026) Stability and reliability of artificial intelligence models in embryo selection for in vitro fertilization. *Fertil Steril* 125(2):277–286. 10.1016/j.fertnstert.2025.08.021
46. Valera R et al (2023) Clinical validation of automatic cleavage-stage embryo classification. *Hum Reprod* 38. 10.1093/humrep/dead058
47. VerMilyea M, Hall JMM, Diakiw SM, Johnston A, Nguyen T, Perugini D, Miller A, Perugini M (2020) Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images. *Hum Reprod* 35(4):770–784. 10.1093/humrep/deaa013
48. Wang S, Chen L, Sun H (2023) Interpretable artificial intelligence-assisted embryo selection improved single-blastocyst transfer outcomes: a prospective cohort study. *Reprod Biomed Online* 47(6):103371. 10.1016/j.rbmo.2023.103371
49. World Health Organization (2025) Infertility. World Health Organization, Geneva. 28 November 2025
50. Wygocki P, Gilewicz T, Pawlik P, Siennicki M, Brzozowski M, Kuśmierczyk-Kubiak J, Sankowska U, Milewski R, Chmielowska A, Kordalewska M, Różańska M, Kuczyński W, Wojtasik B, Sankowski P, Liebermann J (2026) Evaluating the concordance between AI-based and conventional embryo selection: implications for clinical decision-making. *Reprod Biomed Online* 52(4):105502. 10.1016/j.rbmo.2026.105502
51. Yadav R, Banerjee D and K., J.R (2026) AI-based live birth prediction in IVF cycles: a systematic review without meta-analysis of model performance and validation. *Middle East Fertility Soc J* 31:44. 10.1186/s43043-026-00334-0
52. Zheng Z, Wang Y, Ni N, Tong G, Cheng N, Yin P, Chen Y, Wu Y, Xie G, Yang T (2023) 'Deep Learning-Based Quantitative Blastocyst Assessment', *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2023, pp. 1–4. 10.1109/EMBC40787.2023.10340963
53. Zhu P, Bi X, Su D, Zhang X, Wu X (2026) Clinical evaluation and predictive model development of AI-based non-invasive embryo selection in in vitro fertilization patients. *Front Endocrinol* 17:1920306.

Figures

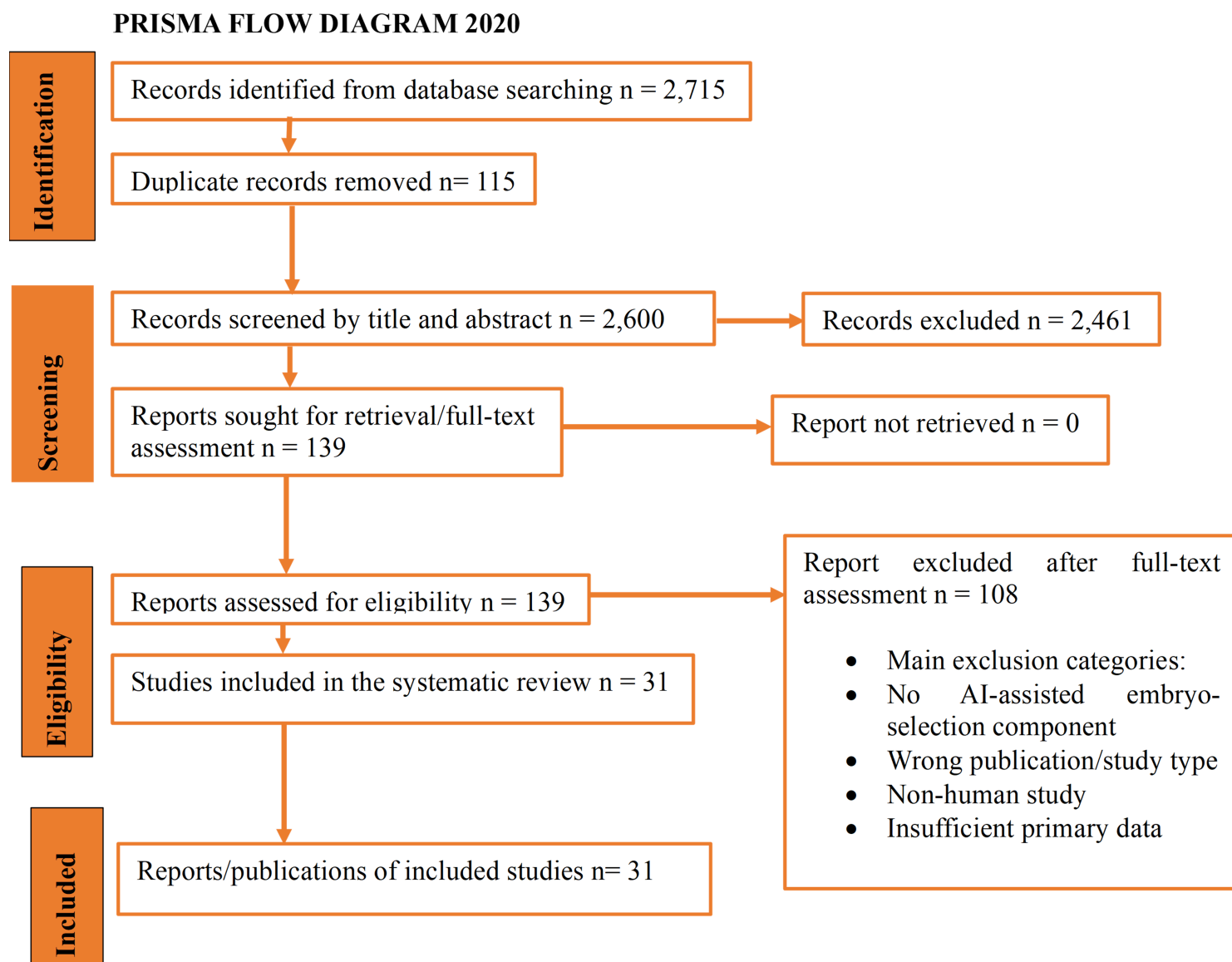


Figure 1. PRISMA 2020 flow diagram illustrating the identification, screening, eligibility assessment, and inclusion of studies on artificial intelligence-assisted embryo selection in in vitro fertilisation.

Figure 1

See image above for figure legend