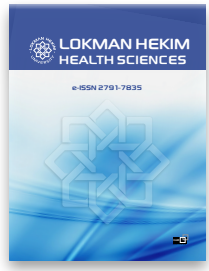
LOKMAN HEKIM HEALTH SCIENCES

DOI: 10.14744/lhhs.2026.91780
Lokman Hekim Health Sci 2026;6(2):290–302

SYSTEMATIC REVIEW / META ANALYSIS




lokmanhekimhs.com

# Artificial Intelligence in Unexplained Infertility: A Systematic Review of Machine Learning-Based Predictive Models

**Tuğba Tahta, Özge Aydoğan Aşir**

**Department of Midwifery, Ankara Medipol University Faculty of Health Sciences, Ankara, Türkiye**

## Abstract

**Introduction:** This systematic review evaluates the methodological rigor and risk of bias of machine learning-based predictive models developed to estimate the success of assisted reproductive technologies in cases of unexplained infertility, using the Prediction Model Risk of Bias Assessment Tool.

**Methods:** A systematic review was conducted to assess predictive modeling studies focused on unexplained infertility and based on machine learning, guided by the framework of the Prediction Model Risk of Bias Assessment Tool. After rigorously screening 912 records, only three studies met the inclusion criteria. While limited in number, these studies highlight emerging evidence in this underexplored area.

**Results:** The included studies applied supervised machine learning algorithms, such as Random Forest, Support Vector Machines, Partial Least Squares Discriminant Analysis, and neural networks, across various biomedical data types. Reported predictive performance varied by data modality: spectroscopy-based models demonstrated high classification accuracy, ranging from 92% to 100%, while a couple-based metabolic model with external validation achieved an accuracy of 73.8%. According to the PROBAST assessment, two studies were rated as low risk of bias, whereas one study exhibited an unclear risk, primarily due to limitations in external validation and analytical transparency.

**Discussion and Conclusion:** This systematic review demonstrates the potential of machine learning-based models to enhance clinical decision-making in the context of unexplained infertility.

**Keywords:** Artificial intelligence; Infertility; Machine learning

According to the World Health Organization (WHO), infertility is defined as the inability to achieve pregnancy after one year (or more) of regular, unprotected sexual intercourse. Unexplained infertility is characterized by the inability to conceive despite regular, unprotected intercourse in the absence of any proven reproductive pathology in either partner.[1–3] Currently, there is no consensus on the diagnostic tests required to identify unexplained infertility.[3]

Various assisted reproductive technologies (ART) are commonly used for couples diagnosed with unexplained infertility, with *in vitro* fertilization (IVF) being the most preferred treatment approach.[4–6] However, ART is typically only partially subsidized and remains an expensive treatment protocol, imposing a substantial financial burden on couples. Reported pregnancy rates after ART generally range between 30% and 40%, highlighting the challenges of individualizing prognosis, the need for detailed clinical

***Cite this article as:*** *Tahta T, Aydoğan Aşir Ö. Artificial Intelligence in Unexplained Infertility: A Systematic Review of Machine Learning-Based Predictive Models. Lokman Hekim Health Sci 2026;6(2):290–302.*

**Correspondence:** Tuğba Tahta, PhD. Ankara Medipol Üniversitesi, Sağlık Bilimleri Fakültesi, Ebelik Anabilim Dalı, Ankara, Türkiye
**E-mail:** tahtatugba@gmail.com **Submitted:** 09.12.2025 **Revised:** 20.12.2025 **Accepted:** 17.02.2026 **Available Online:** 21.05.2026

evaluations for each patient, and the cost-related barriers that add complexity to this issue.[7–9]

Artificial intelligence (AI), particularly its subfield of machine learning (ML), has demonstrated increasing efficacy in numerous studies aimed at improving clinical decision-support systems. ML algorithms are capable of identifying complex patterns and relationships within high-dimensional, heterogeneous biomedical data. In unexplained infertility—where the etiology is uncertain and data heterogeneity is high—ML offers a safe and robust alternative to traditional statistical approaches by providing enhanced predictive capability for IVF outcomes based on patient-specific characteristics. This allows clinicians to develop personalized treatment strategies and helps couples form more realistic expectations regarding pregnancy potential.[10–13]

A systematic review is needed to evaluate the quality, reliability, and risk of bias in existing ML models in this field. The Prediction Model Risk of Bias Assessment Tool (PROBAST) is a widely recognized framework for critically assessing predictive modeling studies.[14]

Despite the growing number of studies applying machine learning to reproductive medicine, a critical gap remains in the literature regarding unexplained infertility as a distinct clinical entity. Most ML-based infertility studies include heterogeneous populations, often combining multiple etiological subgroups without explicitly defining unexplained infertility or tailoring models to this diagnosis. As a result, the clinical validity and generalizability of these models for couples with unexplained infertility remain uncertain. Furthermore, existing reviews in this field primarily focus on algorithmic performance or technological advancements, while neglecting systematic evaluation of methodological rigor, risk of bias, and clinical applicability. To date, no systematic review has specifically assessed ML-based predictive models for unexplained infertility using a validated risk-of-bias framework. Addressing this gap is essential to determine whether current ML models provide reliable and clinically meaningful support for decision-making in this challenging and costly patient population.

The aim of this systematic review is to evaluate studies employing ML-based predictive models designed to forecast ART success in couples with unexplained infertility, using the PROBAST tool to assess methodological quality and provide insights for future research.

## Materials and Methods

This review was developed in accordance with a systematic review protocol, following the PRISMA-P 2015 checklist.[15]

The reliability and quality of the predictive models were assessed using PROBAST criteria,[14–16] ensuring transparent and systematic reporting of findings.

This systematic review was prospectively registered with PROSPERO (registration number: CRD420251037469). The reporting of this review was also conducted in line with the PRISMA-P 2015 statement for systematic reviews, particularly regarding the transparent reporting of the search strategy and study selection process.

### Research Question

In this study, the research question was structured based on the Population, Intervention, Comparison, Outcome, Study Design (PICOS) framework.[17]

The research question in this systematic review and/or meta-analysis was structured according to the PICOS framework:

- P (Population): The target population includes women, or heterosexual couples diagnosed with unexplained infertility.
- I (Intervention): The intervention involves the use of ML algorithms for predictive or diagnostic purposes related to infertility treatment outcomes.
- C (Comparison): The performance of ML models is compared to conventional diagnostic or predictive methods, or to other ML models when applicable.
- O (Outcomes): Primary outcomes include the prediction accuracy of pregnancy outcomes, accuracy of embryo selection, diagnostic accuracy, AUC, sensitivity, specificity, and other relevant performance metrics.
- S (Study Design): The included studies consist of randomized controlled trials, non-randomized quasi-experimental studies, correlational/observational studies, and descriptive studies, all of which were reported in a consistent manner.

The specific questions examined in this review are as follows:

- Research Question 1: What are the differences in terms of accuracy, sensitivity, and specificity between machine learning-based predictive models and traditional statistical methods in predicting pregnancy rates among individuals diagnosed with unexplained infertility?
- Research Question 2: What are the AUC values of different machine learning algorithms (e.g., Random Forest, SVM, Artificial Neural Networks, Deep Learning) in terms of embryo selection accuracy?

- Research Question 3: Do machine learning-based predictive models developed for cases of unexplained infertility possess sufficient reliability to be used confidently in clinical decision-making processes?

## Inclusion Criteria

- Original research studies utilizing ML or AI algorithms.
- Studies specifically addressing unexplained infertility, IVF, embryo selection, and/or treatment success.
- Research conducted using clinical data (i.e., real patient and/or embryo data).
- Studies published in peer-reviewed journals from the year 2015 onwards.

## Exclusion Criteria

- Studies presenting only theoretical models (i.e., simulation-based studies without clinical data).
- Studies relying solely on traditional statistical methods without the use of ML/AI techniques.
- Animal experiments or *in vitro* cell culture studies.
- Case reports, review articles, editorials, or commentary papers.

## Literature Search / Search Strategy

A comprehensive literature search was conducted in the electronic databases PubMed, Web of Science, Scopus, Embase, Google Scholar, and the Cochrane Library. All databases were searched from January 1, 2010, to January 1, 2025, which was the date of the final search update. The search included articles published in Turkish and English within the last 15 years encompassing randomized controlled trials, non-randomized quasi-experimental studies, correlational/observational studies, and descriptive studies. Titles and abstracts of the identified articles were screened according to the inclusion criteria. Relevant studies were imported into an EndNote library for further selection and categorization. MeSH terms and free-text keywords were used to formulate the search strategy. The following terms and their combinations were employed: "unexplained infertility," "infertility," "machine learning," "predictive modeling," "IVF," "pregnancy outcomes," "embryo selection," "artificial intelligence in infertility," "fertility prediction," "supervised learning," "reproductive outcomes," "randomized controlled trials," "non-randomized quasi-experimental studies," "correlational/observational studies," and "descriptive studies."

For each database, a database-specific full Boolean search string combining controlled vocabulary (e.g., MeSH terms) and free-text keywords was developed and adapted to the syntax of that database (e.g., field tags, truncation symbols, and operators). A summary of the core search concepts is presented in Table 1, whereas the complete electronic search strategies (including full Boolean strings for each database, with all field tags and limits) are provided in Appendix 1, in accordance with PRISMA-P 2015 recommendations for transparent reporting of search strategies.

**Table 1.** Summary of core search concepts (keywords used in the electronic search)

| | |
|---|---|
| **Keywords** | Unexplained infertility |
| | Infertility |
| | Machine learning |
| | Predictive modeling |
| | IVF |
| | Pregnancy outcomes |
| | Embryo selection |
| | Randomized |
| | Randomized Controlled Trial |
| | Randomized Clinical Trial |

IVF: *In vitro* fertilisation.

## Study Selection

All records identified through database searches (n=912) were imported into EndNote for reference management and deduplication. After removing duplicates (n=101), 811 records were screened by title and abstract, and 785 were excluded, primarily due to keyword/indexing incompatibility with the review focus (n=570) or publication in other languages (n=215). The full texts of 26 reports were assessed for eligibility and all were successfully retrieved. Of these, 23 were excluded at the full-text stage for predefined reasons, most commonly because the study population was not explicitly restricted to unexplained infertility or because the unexplained infertility subgroup could not be isolated from mixed-etiology cohorts. Additional reasons for exclusion included different clinical conditions (e.g., PCOS/PCOD or recurrent pregnancy loss), outcomes not aligned with the review objectives, non-original article types, or absence of ML/AI-based predictive modeling. Ultimately, three studies met the inclusion criteria and were included in the synthesis (Fig. 1; Appendix 2).

Despite an extensive and systematic search across multiple databases, only three studies met the strict inclusion criteria of this review. This limited yield is primarily attributable to the narrow clinical definition of unexplained infertility, which is often inconsistently reported or entirely omitted in the

broader machine learning–based infertility literature. Most ML studies in reproductive medicine analyze infertility as a heterogeneous condition or combine multiple etiological subgroups without explicitly identifying unexplained infertility as a distinct diagnostic entity, rendering them unsuitable for disease-specific synthesis. In addition, many ML-based models focus on embryo selection, semen analysis, oocyte quality, or general IVF outcomes such as clinical pregnancy, rather than on the diagnosis or classification of unexplained infertility, which was the targeted focus of this review. Therefore, only studies that explicitly identified unexplained infertility as a primary diagnostic criterion and reported predictive modeling outcomes for this specific population were included, while studies that investigated infertility without distinguishing between diagnostic subgroups were excluded to maintain clinical and methodological coherence. Furthermore, the application of strict methodological criteria—limiting inclusion to peer-reviewed primary research involving human participants and explicit ML-based predictive modeling—further reduced the eligible pool in accordance with the PICOS framework. Consequently, only three primary research articles fulfilled all inclusion criteria and were eligible for full-text synthesis. This limited number of included studies reflects a genuine gap and the early stage of ML applications in the specific field of unexplained infertility rather than limitations of the search strategy, as illustrated in the PRISMA flow diagram.

In addition to peer-reviewed journal articles, gray literature sources—including conference proceedings, dissertations, preprints, and non-indexed reports—were also screened to minimize publication bias. Searches were performed in Google Scholar, ProQuest Dissertations & Theses, and institutional repositories using the same keyword combinations. However, none of the gray literature sources met the predefined inclusion criteria (i.e., machine learning-based predictive modeling specifically targeting unexplained infertility in human participants). Therefore, although gray literature was thoroughly reviewed, no additional eligible studies were identified.

In addition to this, a detailed list of studies included and excluded at the full-text screening stage, together with brief reasons for exclusion, is presented in Appendix 2 to further enhance the transparency and reproducibility of the study selection process.

Most of the excluded articles addressed infertility in a general context, did not specifically distinguish unexplained infertility as a diagnostic category, or did not utilize machine learning-based predictive modeling methods. The

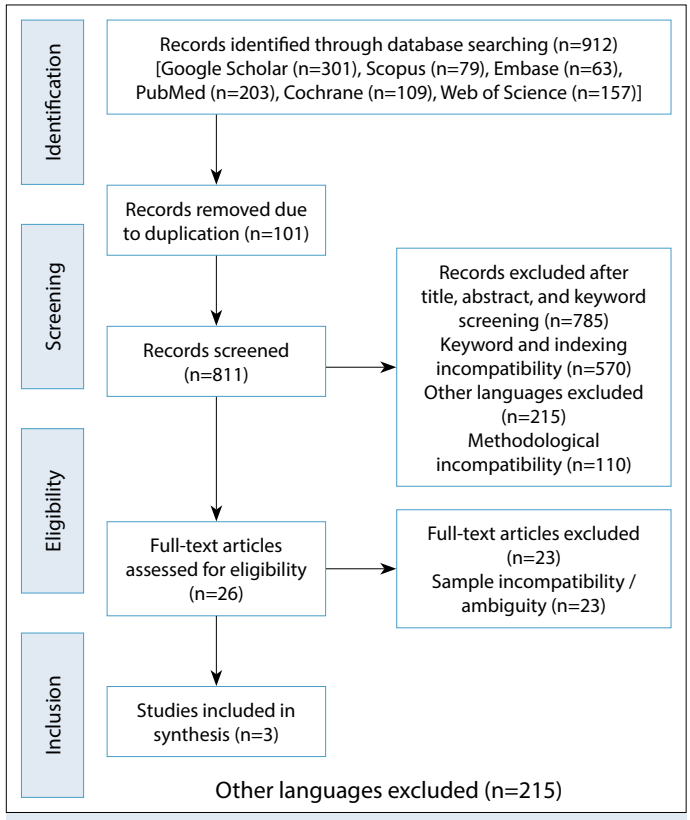



**Figure 1.** PRISMA flow diagram.

screening process was conducted independently by two reviewers (T.T., Ö.A.A.). Each reviewer assessed the titles, abstracts, and full texts according to predefined inclusion and exclusion criteria. Discrepancies between reviewers were resolved through discussion, and when necessary, a third reviewer was consulted to reach a consensus.

## Primary Outcomes

- Predictive accuracy of ML/AI models for clinical outcomes in unexplained infertility, particularly:
  - o Pregnancy prediction accuracy
  - o Embryo selection accuracy
  - o Diagnostic accuracy

Performance metrics reported in the included studies, such as:

- AUC
- Sensitivity (true positive rate)
- Specificity (true negative rate)
- Positive Predictive Value (PPV) and Negative Predictive Value (NPV)

Timing of outcome measurement: at the conclusion of the IVF treatment cycle or as defined by each study (e.g., clinical pregnancy, ongoing pregnancy, live birth).

**Table 2.** PROBAST risk of bias assessment – Depciuch et al.[18]

| Domain | Signaling questions | Assessment | Comments |
|---|---|---|---|
| Participants | Were appropriate data sources used? | Low risk | Participants were selected from an IVF center with a diagnosis of unexplained infertility. |
| | Were inclusion/exclusion criteria appropriate? | Low risk | Clear criteria focusing on unexplained infertility cases. |
| Predictors | Were predictors defined and assessed appropriately? | Low risk | Raman spectral features were systematically measured and analyzed. |
| | Were predictor assessments blinded to outcome data? | Unclear risk | The study does not specify blinding procedures. |
| Outcome | Was the outcome defined and determined appropriately? | Low risk | Outcomes were based on established oxidative stress markers and spectral analysis. |
| | Was the outcome determined without knowledge of predictor information? | Unclear risk | Blinding of outcome assessment is not detailed. |
| Analysis | Were appropriate statistical methods used? | Low risk | Utilized multiple machine learning algorithms with high classification accuracy. |
| | Was model overfitting avoided? | Low risk | High accuracy suggests robustness, but external validation is not mentioned. |
| | Were performance measures appropriate? | Low risk | Accuracy metrics were reported for model performance. |

PROBAST: Prediction Model Risk of Bias Assessment Tool.

Effect measure: Standardized reporting of model performance (e.g., AUC, accuracy, sensitivity, specificity), extracted as reported in each study for synthesis.

## Risk of Bias

The methodological quality and risk of bias of the included studies were evaluated using the PROBAST.[14] This tool is designed to assess the risk of bias and concerns regarding the applicability of studies that develop, validate, or update prediction models.

PROBAST consists of four key domains:

To provide greater methodological transparency and reproducibility, the PROBAST evaluation criteria were applied and reported in detail. PROBAST comprises 20 signaling questions organized into four core domains, each addressing potential sources of bias in prediction model studies:

- Participants (P): This domain assesses whether the study population is representative of the target population for the intended clinical use of the prediction model. Specific considerations include the data source (e.g., single-center vs. multicenter), recruitment strategy, eligibility criteria, and whether participant selection could introduce selection bias.
- Predictors (Pr): The predictors domain evaluates whether all candidate predictors were clearly defined, measured in a consistent and reproducible manner across participants, and assessed without knowledge of outcome status. Particular attention was paid to biological plausibility, measurement timing, and potential information leakage between predictors and outcomes.
- Outcome (O): This domain examines whether the outcome was explicitly defined using clinically accepted criteria, measured objectively, and determined independently of predictor information. The timing of outcome assessment and the presence or absence of blinding were also considered, as these factors may influence outcome misclassification.
- Analysis (A): The analysis domain focuses on the statistical rigor of model development and validation. This includes evaluation of sample size adequacy, handling of missing data, prevention of model overfitting, use of appropriate internal and/or external validation strategies, transparency of model specification, and the appropriateness of reported performance metrics (e.g., accuracy, sensitivity, specificity, AUC).

Each signaling question was independently rated by two reviewers as "Yes," "Probably Yes," "Probably No," "No," or "No Information," in accordance with PROBAST guidance. Domain-level risk of bias judgments (Low Risk, High Risk, or Unclear Risk) were derived by aggregating responses within each domain, with particular emphasis on the analysis domain due to its critical role in prediction model validity. In addition to risk of bias, applicability concerns were evaluated separately for the Participants, Predictors, and Outcomes domains to determine the clinical relevance of each study to the review question focusing on unexplained infertility. Any disagreements between reviewers were

**Table 3.** PROBAST Risk of Bias Assessment – Jakubczyk et al.[19]

| Domain | Signaling questions | Assessment | Comments |
|---|---|---|---|
| Participants | Were appropriate data sources used? | Low risk | Participants were recruited from a fertility clinic and included idiopathic infertility cases and fertile controls. |
| | Were inclusion/exclusion criteria appropriate? | Low risk | Inclusion and exclusion criteria were clearly described and appropriate for the research question. |
| Predictors | Were predictors defined and assessed appropriately? | Low risk | FTIR spectral features and gonadotrophin levels were predefined and biologically plausible predictors. |
| | Were predictor assessments blinded to outcome data? | Unclear risk | It is not explicitly stated whether spectral data analysis was blinded to infertility status, which may introduce bias. |
| Outcome | Was the outcome defined and determined appropriately? | Low risk | Idiopathic infertility diagnosis was made using standard clinical criteria. |
| | Was the outcome determined without knowledge of predictor information? | Low risk | Outcome classification was based on clinical records and unlikely to be influenced by predictor data. |
| Analysis | Were appropriate statistical methods used? | Some concerns | While multivariate and machine learning methods (PLS, SVM) were used, details on validation strategy were limited. |
| | Was model overfitting avoided? | Some concerns | External validation was not clearly reported; risk of overfitting exists due to potential model complexity. |
| | Were performance measures appropriate? | Low risk | Model performance was presented (sensitivity, specificity), though external validation was limited. |

PROBAST: Prediction Model Risk of Bias Assessment Tool.

**Table 4.** PROBAST Detailed Risk of Bias Table – Bachelot et al.[20]

| Domain | Signaling questions | Assessment | Comments |
|---|---|---|---|
| Participants | Were appropriate data sources used? | Low risk | Participants were recruited from a clinical fertility cohort; both infertile and fertile couples clearly defined. |
| | Were inclusion/exclusion criteria appropriate? | Low risk | Inclusion/exclusion criteria were explicitly reported; infertility defined as >12 months TTP. |
| Predictors | Were predictors defined and assessed appropriately? | Low risk | Anthropometric, metabolic, and antioxidative variables were well-described and clinically relevant. |
| | Were predictor assessments blinded to outcome data? | Low risk | Predictor measurement was independent of fertility status; outcome knowledge unlikely to bias assessment. |
| Outcome | Was the outcome defined and determined appropriately? | Low risk | Outcome (fertile vs. idiopathic infertility) was based on standard clinical definitions and timeline to pregnancy. |
| | Was the outcome determined without knowledge of predictor information? | Low risk | Outcomes were historical or prospectively collected; unlikely to be influenced by predictor values. |
| Analysis | Were appropriate statistical methods used? | Low risk | Random Forest and OPLS-DA were appropriate for classification; external validation was performed. |
| | Was model overfitting avoided? | Low risk | The model was trained on a development set and tested on a separate external validation set. |
| | Were performance measures appropriate? | Low risk | Accuracy was reported; variables were reduced to 13 features, and performance validated externally. |

PROBAST: Prediction Model Risk of Bias Assessment Tool.

resolved through discussion until consensus was reached. The detailed results of the PROBAST assessment are presented in Tables 2, 3, 4, with an overall summary provided in Table 5 and a narrative synthesis included in the Results section.[14–16]

## Synthesis Method

In this systematic review, statistical meta-analysis was not conducted due to the limited number of included primary studies (n=3). The primary reason for this decision was the lack of sufficient homogeneous data across studies to support

**Table 5.** Study characteristics and summary of PROBAST risk of bias assessment

| Study | Participants risk | Predictors risk | Outcome risk | Analysis risk | Overall PROBAST risk |
|---|---|---|---|---|---|
| Depciuch et al.[18] | Low risk | Low/unclear risk | Low/unclear risk | Low risk | Unclear risk |
| Jakubczyk et al.[19] | Low risk | Low/unclear risk | Low risk | Unclear risk | Unclear risk |
| Bachelot et al.[20] | Low risk | Low risk | Low risk | Low risk | Low risk |

A PROBAST scoring matrix was added to provide a visual summary of the risk-of-bias judgments across all four domains for each included study. The matrix illustrates domain-specific ratings (low, high, unclear) in a structured format, allowing readers to easily compare methodological strengths and limitations across studies. This matrix complements the narrative assessment and ensures transparency and reproducibility of the risk-of-bias evaluation.

meta-analysis. Instead, a narrative synthesis approach was adopted. The data extracted from each study were systematically organized in a summary table, considering factors such as the study's objectives, sample size and characteristics, machine learning algorithms used, targeted outcome variables (e.g., pregnancy prediction accuracy, embryo selection accuracy, diagnostic accuracy), and performance metrics (AUC, sensitivity, specificity, accuracy rate, etc.). Similarities and differences among the studies were comparatively analyzed under thematic headings.

Furthermore, to assess the methodological quality and potential bias risk of the predictive models developed in the studies, the PROBAST tool was utilized. PROBAST provides a systematic evaluation of the risk of bias across four domains (participants, predictors, outcomes, and analysis), offering insight into the reliability and generalizability of the models. This approach has made a significant contribution to deeply investigating the current state of machine learning applications for unexplained infertility and identifying gaps in the existing literature.[14]

## Results

A total of 3 studies meeting the inclusion and exclusion criteria were included in this systematic review. The methodological quality and risk of bias of the studies were assessed using the PROBAST criteria for risk of bias and applicability. Individual quality and risk assessments for each study, as well as the overall risk of bias summary table, are presented below.

### PROBAST Evaluation of Included Studies

This study demonstrated a low risk of bias across all PROBAST domains.[18] Participants were clearly defined, and the predictors—Raman spectral features and oxidative stress markers—were measured with robust methodology. A variety of machine learning algorithms with internal validation were applied, and outcome assessments were consistently conducted. Appropriate performance metrics were reported, supporting the model's predictive reliability. Overall, this study provides strong methodological rigor and offers reliable evidence for the diagnostic application of machine learning in the context of unexplained infertility.

Overall PROBAST Risk of Bias Judgment: Unclear Risk.

This study included well-characterized participants and clinically relevant predictors, such as FTIR spectra and gonadotrophin levels.[19] However, there are concerns regarding the analysis domain. Specifically, the absence of explicit external validation and the potential lack of blinding during predictor assessment introduce possible biases. Therefore, while the findings are promising, the risk of bias is considered moderate or unclear. Further external validation is recommended to enhance the generalizability and practical utility of the model.

Overall PROBAST Risk of Bias Judgment: Unclear Risk.

This study demonstrated a low risk of bias across all PROBAST domains.[20] The predictors were biologically meaningful, clinically relevant, and measured appropriately. Outcome definitions were clearly established and assessed independently of predictor data. A couple-based modeling approach was used, and external validation was performed with consistent performance metrics. These factors collectively support the methodological robustness and clinical relevance of the model for predicting unexplained infertility.

Overall PROBAST Risk of Bias Judgment: Low Risk.

The three studies included in this systematic review were critically analyzed using a classical systematic review table format. This approach allowed us to systematically compare the methodology, participants, and key findings of each study. The studies were selected based on their use of ML or AI algorithms to predict outcomes in unexplained infertility, specifically focusing on clinical data related to embryo selection, pregnancy outcomes, or diagnostic accuracy. Below is a summary of the studies in the systematic review table.

### Analysis of the Included Studies

This systematic review examined three original studies that explored the application of ML techniques in the diagnosis and classification of unexplained female infertility. These

**Table 6.** Systematic review table (classic format)

| Study | Design | Population | Sample size | Methods | Outcomes | Key findings |
|---|---|---|---|---|---|---|
| Depciuch et al.,[18] Türkiye | Retrospective observational study | Women diagnosed with unexplained infertility undergoing IVF | 128 follicular fluid samples (65 unexplained infertility, 63 controls) | Raman spectroscopy of follicular fluid; oxidative load measurements; multivariate analysis; six machine learning algorithms (Random Forest, C5.0 Decision Tree, k-NN, Deep Neural Networks, SVM, XGBoost) | Differences in Raman spectral bands associated with oxidative load, amide III, and lipids; classification accuracy of machine learning models | Significant spectral differences between groups; machine learning models achieved classification accuracy ranging from 92.04% to 99.17% |
| Jakubczyket al.,[19] Türkiye | Prospective observational study | Women diagnosed with idiopathic female infertility (IFI) undergoing IVF treatment at Academic Hospital IVF Center | 116 women: 58 with idiopathic female infertility and 58 fertile controls; follicular fluid samples collected during IVF procedures | Fourier-transform infrared (FTIR) spectroscopy combined with multivariate analysis and six machine learning algorithms (e.g., Random Forest, SVM) | Identification of FTIR spectral characteristics associated with ovarian reserve and reproductive hormone levels; differentiation between IFI and control groups using multivariate analysis and machine learning methods; classification accuracy ranging from 93.75% to 100% depending on the algorithm used. | FTIR spectra revealed significant differences in lipid and protein vibrations between groups; machine learning models achieved classification accuracies ranging from 93.75% to 100% depending on the algorithm used |
| Bachelot et al.,[20] France | Multicenter cross-sectional case–control study | Couples with primary idiopathic infertility (>12 months) and fertile couples with spontaneous conception (<12 months to pregnancy) | 197 couples (97 infertile, 100 fertile); development set: 136 couples; test set: 61 couples | Machine learning model (Random Forest) based on couple-level data. Model was trained on 136 couples and tested on 61 couples | Development of a couple-based machine learning model (OPLS-DA) to stratify infertile and fertile couples based on anthropometric, metabolic, and antioxidative parameters; refined model with 13 variables achieved 73.8% accuracy in external validation; couple-based approach outperformed models considering men and women separately. | The model identified 13 predictive variables and achieved 73.8% accuracy in classifying idiopathic infertility. Couple-based prediction outperformed individual-based models. |

studies utilized biochemical and spectroscopic analyses—primarily of follicular fluid—and applied a range of ML algorithms to identify potential biomarkers and enhance diagnostic precision.

In a retrospective, cross-sectional, observational study conducted,[18] a total of 128 follicular fluid samples were analyzed, including 65 from women with unexplained infertility and 63 from fertile controls. The samples were evaluated using Raman spectroscopy, and the spectral data were analyzed via multivariate statistical approaches and correlation tests to assess oxidative stress markers. Six machine learning algorithms were employed for classification purposes: Random Forest, C5.0 decision tree, k-nearest neighbors (k-NN), deep neural networks, support vector machines (SVM), and XGBoost. The results indicated significant differences in oxidative stress-related

Raman bands, particularly within the amide III and lipid regions, between the unexplained infertility group and the control group. Classification accuracy for the machine learning models ranged between 92.04% and 99.17%. This study demonstrates that Raman spectroscopy combined with machine learning approaches can be utilized to identify potential biomarkers in the diagnosis of unexplained infertility. The findings suggest that oxidative stress-associated molecular changes in follicular fluid may contribute to a better understanding of the etiology of infertility and aid in diagnostic strategies.

In this cross-sectional, analytical observational study conducted,[19] a total of 116 follicular fluid samples (58 from women with unexplained female infertility (IFI) and 58 from fertile controls) were analyzed. The study utilized Fourier Transform Infrared (FTIR) spectroscopy to assess biochemical differences in follicular fluid, in combination with gonadotrophin levels. Spectral data were processed using multivariate statistical techniques including Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA), along with machine learning algorithms such as Support Vector Machines (SVM) and Random Forest to classify the samples. The results revealed increased lipid absorption and decreased amide I and II bands in the IFI group, suggesting altered biochemical composition in the follicular environment. These alterations were associated with impaired reproductive potential. The applied machine learning models achieved classification accuracies up to 100%, demonstrating their diagnostic potential. This study highlights the promising role of FTIR spectroscopy combined with machine learning in identifying potential biomarkers for the diagnosis of unexplained infertility. The molecular alterations observed in follicular fluid may provide insight into biochemical mechanisms affecting oocyte quality and fertilization capacity.

In this prospective observational study conducted,[20] couples with primary unexplained infertility (failure to conceive after >12 months) were compared to fertile couples who achieved spontaneous conception within 12 months. A total of 197 couples were included (97 infertile and 100 fertile); 136 couples were used in the development dataset and 61 couples in the external test set. The researchers developed a machine learning model using Random Forest and Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) based on couple-level anthropometric, metabolic, and antioxidative parameters. A refined model incorporating 13 predictive variables achieved an external validation accuracy of 73.8%. Importantly, the couple-based predictive model significantly outperformed models that assessed men and women separately. This study demonstrates that integrating couple-level metabolic and biochemical data into machine learning models can effectively classify unexplained infertility. The findings support the clinical relevance of adopting a couple-based approach, rather than individual-based assessments, for improved diagnostic and predictive accuracy in infertility management.

### Types of Features and Predictors Used

The reviewed studies demonstrated variation according to data sources.

In spectroscopy-based studies,[18,19] predictor variables primarily included oxidative stress markers, Raman and FTIR spectral bands derived from follicular fluid, and gonadotropin levels—serving as biochemical and molecular indicators of oocyte quality and the follicular environment. In contrast, the study conducted[20] analyzed couple-level anthropometric data (e.g., BMI), metabolic markers (e.g., glucose and lipid levels), and antioxidant parameters. This systemic approach provided a broader and more holistic view of infertility-related physiology.

## Discussion

This systematic review examined three original studies applying ML techniques to the diagnosis and classification of unexplained female infertility. Among the three predefined research questions, the first and second were addressed through comparative evaluation of accuracy and AUC across studies, while the third—clinical reliability—was discussed in the context of the PROBAST assessment. However, due to the limited number of studies, some aspects could only be addressed narratively rather than quantitatively. The findings consistently indicate that integrating biochemical, metabolic, and spectroscopic data with ML approaches can significantly enhance diagnostic precision, potentially uncovering hidden pathophysiological mechanisms behind unexplained infertility. The selected studies demonstrate a strong topical alignment, all focusing on the application of machine learning methods to detect biomarkers or physiological indicators associated with unexplained infertility. They address critical gaps in understanding the etiology of unexplained infertility by employing advanced computational tools.[18–20]

The reviewed studies were published between 2021 and 2023. All three utilized observational study designs: two were cross-sectional analytical studies,[18,19] and one was

a prospective observational study.[20] The studies utilized a range of data types. Spectroscopy-based studies[18,19] focused on biochemical features extracted from follicular fluid samples, including Raman and FTIR spectral bands, oxidative stress markers, and gonadotrophin levels. In contrast,[20] employed anthropometric, metabolic, and antioxidative biomarkers at the couple level, providing a systemic physiological view. The primary objectives across studies were to enhance the diagnostic capabilities for unexplained infertility by applying ML methods to biomedical datasets, identify potential biomarkers, and improve understanding of the biological underpinnings of unexplained infertility. Notably,[20] extended the scope to consider couple-based features, emphasizing the dyadic nature of infertility. Although traditional supervised ML algorithms dominated, the inclusion of deep neural networks[18] represents a notable foray into deep learning approaches. However, there remains a limited application of unsupervised learning techniques across the reviewed studies. Future research should consider greater utilization of unsupervised and deep learning methods to discover latent structures in the data without predefined labels. While current studies predominantly focused on biochemical and metabolic predictors, the unexplored psychological or hormonal dimensions of unexplained infertility represent an important frontier.

Unexplained infertility imposes a profound emotional toll that extends beyond medical diagnosis. Due to the absence of a clear explanation for fertility difficulties, couples often experience anxiety, psychological strain, and frustration, which negatively affect their quality of life.[2,3,21,22] In addition to psychological distress, infertility treatments are associated with a substantial financial burden, further compounding emotional strain.[21–25] In this context, minimizing diagnostic uncertainty and error through advanced AI-based decision-support systems is of critical importance, as inaccurate or delayed diagnoses may amplify both psychological and financial consequences for affected couples.[26,27] However, existing ML-based studies in unexplained infertility remain limited by methodological constraints, including small sample sizes and restricted generalizability.[14]

Although reducing the psychological and financial burden of unexplained infertility represents an important long-term objective of ML-based decision-support systems, current evidence suggests that existing models are still far from achieving this goal in a clinically meaningful manner. The reviewed studies primarily rely on biochemical, spectroscopic, or metabolic markers, while none incorporate patient-reported psychological distress, psychosocial variables, or cost-related outcomes. Moreover, the high diagnostic accuracy reported in spectroscopy-based studies (92–100%) is derived from small, single-center datasets without external validation, limiting their immediate applicability to routine clinical practice. Thus, while these early models may help shorten the diagnostic process and reduce uncertainty, their current level of development does not yet support direct translation into real-world clinical workflows. Future ML models should therefore be externally validated, integrated with broader biopsychosocial data, and evaluated prospectively to determine whether they can meaningfully improve patient experience and inform more cost-effective care pathways.

The comparative synthesis of the included studies demonstrates that machine learning models have strong potential in detecting biochemical and metabolic signatures associated with unexplained infertility; however, their current applicability is constrained by methodological and data-related factors. Among the algorithms used, spectroscopy-based datasets (Raman and FTIR) showed the highest predictive performance, with classification accuracies ranging from 92% to 100%, particularly when analyzed using Random Forest, SVM, and PLS-DA. These strong results are likely attributable to the high-dimensional molecular features contained in spectroscopic data, which are well suited to ML-based classification. In contrast, the couple-level metabolic dataset demonstrated more modest performance, achieving an externally validated accuracy of 73.8%. This suggests that systemic physiological markers provide clinically relevant but less discriminative signals than direct biochemical profiling of follicular fluid.

When examined through PROBAST, spectroscopy-based studies demonstrated a comparatively lower risk of bias in the predictor domain, as predictors were derived from standardized laboratory measurements. However, these studies showed higher risk of bias in the "participants" and "analysis" domains, primarily due to small sample sizes and the absence of external validation. Conversely, the couple-based study exhibited a lower risk of bias in the participant domain, supported by a larger dataset and the use of an independent test set. Nonetheless, it showed a higher risk of bias in the predictor domain, as metabolic and anthropometric variables may be influenced by unmeasured lifestyle or environmental confounders.

Taken together, three major barriers currently limit the translation of ML-based models for unexplained infertility into routine clinical practice:

(1) Small, single-center datasets that may inflate performance estimates and limit generalizability;

(2) Heterogeneity and lack of standardized outcome definitions, particularly regarding biochemical, clinical, and ongoing pregnancy; and

(3) Limited external and prospective validation, which is essential before clinical implementation.

A deeper methodological comparison reveals important differences in model development and validation strategies. Spectroscopy-based studies[18,19] relied on high-dimensional spectral data but did not employ robust external validation, increasing the risk of overfitting despite their high reported accuracies. In contrast, Bachelot et al.[20] incorporated clinically interpretable predictors and applied true external validation, resulting in lower accuracy but greater methodological robustness. Another key distinction relates to model transparency: interpretability was limited in spectroscopic models using PCA/PLS-DA combinations, whereas the couple-based model allowed clearer assessment of predictor contributions. These findings underscore the need for transparent modeling strategies, standardized biomarker definitions, and rigorous validation to support clinical readiness.

Ethical, governance, and regulatory considerations were seldom reported in the included studies. None of the studies provided detailed descriptions of data governance practices, privacy protections, or assessments of potential bias and fairness—elements that are increasingly recognized as important for the safe deployment of AI/ML-driven decision-support tools. Contemporary regulatory and multi-agency guidance for AI/ML-enabled medical technologies places strong emphasis on lifecycle documentation, risk-based evaluation, and transparency to intended users, including clear communication of a model's intended use, limitations, and appropriate human oversight. In this context, explainability and user-facing transparency are not uniformly framed as fixed "requirements" across all use cases, but they are increasingly highlighted as key enablers of trust, reproducibility, and safe integration into clinical workflows. Therefore, future studies in unexplained infertility should report governance safeguards more explicitly (e.g., privacy and access controls, dataset provenance, and bias monitoring), alongside transparent model reporting and validation, to support responsible clinical translation.

### Strengths and Limitations

This systematic review has several strengths. It is among the first to synthesize machine learning-based predictive models focusing on unexplained infertility outcomes. The application of the PROBAST tool for bias assessment is one of its strongest aspects. By addressing unexplained infertility—a critical but often overlooked issue in reproductive medicine—this review contributes valuable insights to the literature.

Nonetheless, several limitations exist. The small number of included studies restricts generalizability. There were notable differences across studies in sample sizes, data types, and ML algorithms used, which complicate comparisons. Future studies should employ multi-center datasets to enhance generalizability and reliability.

Finally, although biochemical and metabolic predictors were comprehensively addressed, psychological, environmental, and financial factors remain underrepresented in existing models.

Overall, the reviewed studies offer preliminary proof-of-concept evidence that ML-based models may capture biochemical and metabolic signatures relevant to unexplained infertility; however, the current literature remains methodologically heterogeneous and numerically limited, precluding firm clinical recommendations. Future research should prioritize multi-center datasets, standardized outcome definitions and reporting, and rigorous external (and preferably prospective) validation. In addition, incorporating broader clinical and biopsychosocial variables—including hormonal profiles and patient-reported measures—may improve model generalizability and support the development of clinically deployable AI tools for unexplained infertility.

## Conclusion

This systematic review highlights a major gap in the existing literature: despite the growing interest in artificial intelligence and machine learning in reproductive medicine, highly focused studies on unexplained infertility remain extremely limited. The sparseness of eligible studies underscores the need for more disease-specific, methodologically rigorous ML research to advance predictive modeling in this area.

Future studies incorporating ML or alternative AI techniques should aim to integrate the multidimensional factors involved in unexplained infertility for a more personalized approach. Additionally, the limited number of studies and the heterogeneity in datasets, algorithm types, and performance metrics prevented direct comparisons or meta-analysis.

**Ethics Committee Approval:** Not applicable. This study is a systematic review and meta-analysis based on previously published data and does not involve direct human or animal participation.

**Informed Consent:** As this study utilizes data from previously published studies, informed consent was obtained in the original investigations.

**Conflict of Interest:** None declared.

**Financial Disclosure:** The author declared that this study has received no financial support.

**Use of AI for Writing Assistance:** None declared.

**Authorship Contributions:** Concept: TT, ÖAA; Design: TT, ÖAA; Supervision: TT, ÖAA; Resource: TT, ÖAA; Materials: TT, ÖAA; Data collection and/or processing: TT, ÖAA; Analysis and/or interpretation: TT, ÖAA; Literature review: TT, ÖAA; Writing: TT, ÖAA; Critical review: TT, ÖAA.

**Peer-review:** Double blind peer-reviewed.

## References

1. Zegers-Hochschild F, Adamson GD, Dyer S, Racowsky C, de Mouzon J, Sokol R, et al. The International glossary on infertility and fertility care 2017. Fertil Steril 2017;108(3):393-406. [CrossRef]
2. Raperport C, Desai J, Qureshi D, Rustin E, Balaji A, Chronopoulou E, et al. The definition of unexplained infertility: A systematic review. BJOG 2024;131(7):880-97. [CrossRef]
3. Mahabbat NA, Khan TA, Elyas MF, Mahabbat AA, Oraif AM. Role of *in vitro* fertilization (ivf) in unexplained infertility management: a systematic review. Cureus 2024;16(10):e72527. [CrossRef]
4. Niederberger C, Pellicer A, Cohen J, Gardner DK, Palermo GD, O'Neill CL, et al. Forty years of IVF. Fertil Steril 2018;110(2):185-324.e5.
5. Griffin DK, Ogur C. Chromosomal analysis in IVF: just how useful is it? Reproduction. 2018;156(1):F29-F50. [CrossRef]
6. Mol BW, Tjon-Kon-Fat R, Kamphuis E, van Wely M. Unexplained infertility: Is it over-diagnosed and over-treated? Best Pract Res Clin Obstet Gynaecol 2018;53:20-9. [CrossRef]
7. Inhorn MC, Patrizio P. Infertility around the globe: new thinking on gender, reproductive technologies and global movements in the 21st century. Hum Reprod Update 2015;21(4):411-26. [CrossRef]
8. Jing X, Gu W, Zhang L, Miao R, Xu X, Wang M, et al. Coping strategies mediate the association between stigma and fertility quality of life in infertile women undergoing *in vitro* fertilization-embryo transfer. BMC Womens Health 2021;21(1):386. [CrossRef]
9. Shingshetty L, Maheshwari A, McLernon DJ, Bhattacharya S. Should we adopt a prognosis-based approach to unexplained infertility? Hum Reprod Open 2022;2022(4):hoac046. [CrossRef]
10. Dehghan S, Moghaddasi H, Rabiei R, Choobineh H, Maghooli K, Vahidi-Asl M. Machine learning in predicting infertility treatment success: A systematic literature review of techniques. J Educ Health Promot 2025;14:103. [CrossRef]
11. Islam MN, Mustafina SN, Mahmud T, Khan NI. Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda. BMC Pregnancy Childbirth 2022;22(1):348. [CrossRef]
12. Medenica S, Zivanovic D, Batkoska L, Marinelli S, Basile G, Perino A, et al. The future is coming: Artificial intelligence in the treatment of infertility could improve assisted reproduction outcomes-the value of regulatory frameworks. Diagnostics 2022;12:2979. [CrossRef]
13. Giaxi P, Vivilaki V, Sarella A, Harizopoulou V, Gourounti K. Artificial intelligence and machine learning: an updated systematic review of their role in obstetrics and midwifery. Cureus 2025;17(3):e80394. [CrossRef]
14. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170:51-58. [CrossRef]
15. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4(1):1. [CrossRef]
16. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med 2013;10(2):e1001381. [CrossRef]
17. Eldawlatly A, Alshehri H, Alqahtani A, Ahmad A, Al-Dammas F, Marzouk A. Appearance of population, intervention, comparison, and outcome as research question in the title of articles of three different anesthesia journals: a pilot study. Saudi J Anaesth 2018;12(2):283-6. [CrossRef]
18. Depciuch J, Paja W, Pancerz K, Uzun Ö, Bulut H, Tarhan N, et al. Analysis of follicular fluid and serum markers of oxidative stress in women with unexplained infertility by Raman and machine learning methods. J Raman Spectrosc 2023;54:501511. [CrossRef]
19. Jakubczyk P, Paja W, Pancerz K, Cebulski J, Depciuch J, Uzun Ö, et al. Determination of idiopathic female infertility from infrared spectra of follicle fluid combined with gonadotrophin levels, multivariate analysis and machine learning methods. Photodiagnosis Photodyn Ther 2022;38:102883. [CrossRef]
20. Bachelot G, Lévy R, Bachelot A, Faure C, Czernichow S, Dupont C, et al. Proof of concept and development of a couple-based machine learning model to stratify infertile patients with idiopathic infertility. Sci Rep 2021;11(1):24003. [CrossRef]
21. Nik Hazlina NH, Norhayati MN, Shaiful Bahari I, Nik Muhammad Arif NA. Worldwide prevalence, risk factors and psychological impact of infertility among women: a systematic review and meta-analysis. BMJ Open 2022;12(3):e057132. [CrossRef]
22. Greil AL, Slauson-Blevins K, McQuillan J. The experience of infertility: A review of recent literature. Sociol Health Illn 2010;32:140-62. [CrossRef]
23. Szkodziak F, Krzyżanowski J, Szkodziak P. Psychological

aspects of infertility. A systematic review. J Int Med Res 2020;48(6):300060520932403. [CrossRef]
24. Duffy JMN, Adamson GD, Benson E, Bhattacharya S, Bhattacharya S, Bofill M, et al. Top 10 priorities for future infertility research: an international consensus development study. Fertil Steril 2021;115(1):180-90. [CrossRef]
25. Woods BM, Bray LA, Campbell S, Holland A, Mrug S, Ladores S. A review of the psychometric properties and implications for the use of the fertility quality of life tool. Health Qual Life Outcomes 2023;21(1):45. [CrossRef]
26. Chambers GM, Hoang VP, Sullivan EA, Chapman MG, Ishihara O, Zegers-Hochschild F, et al. The impact of consumer affordability on access to assisted reproductive technologies and embryo transfer practices: an international analysis. Fertil Steril 2014;101(1):191-8.e4. [CrossRef]
27. Dyer SJ, Patel M. The economic impact of infertility on women in developing countries a systematic review. Facts Views Vis Obgyn 2012;4(2):102-9.

**Appendix 1.** Full electronic search strategies for all databases

| Database | Date searched | Filters applied | Full Boolean search string | Records retrieved |
|---|---|---|---|---|
| PubMed | 15–21 April 2025 | 2010–2025; Humans; English/ Turkish | ("Infertility"[MeSH] OR "infertility"[tiab] OR "unexplained infertility"[tiab] OR "idiopathic infertility"[tiab]) AND ("Machine Learning"[MeSH] OR "machine learning"[tiab] OR "artificial intelligence"[MeSH] OR "artificial intelligence"[tiab] OR "deep learning"[tiab] OR "neural network*"[tiab]) AND ("Predictive model*"[tiab] OR predict*[tiab] OR model*[tiab] OR prediction[tiab] OR classification[tiab]) | n=203 |
| Embase | 15–21 April 2025 | 2010–2025; Humans; English/ Turkish | ('infertility'/exp OR infertilit* OR 'unexplained infertility' OR 'idiopathic infertility') AND ('machine learning'/exp OR 'machine learning' OR 'artificial intelligence'/exp OR 'artificial intelligence' OR 'deep learning' OR 'neural network*') AND (predict* OR 'predictive model*' OR model* OR classification OR algorithm*) | n=63 |
| Scopus | 15–21 April 2025 | 2010–2025; English/ Turkish | TITLE-ABS-KEY(infertilit* OR "unexplained infertility" OR "idiopathic infertility" ) AND TITLE-ABS-KEY( "machine learning" OR "artificial intelligence" OR "deep learning" OR "neural network*" ) AND TITLE-ABS-KEY( predict* OR "predictive model*" OR model* OR classification OR algorithm* ) | n=79 |
| Web of Science | 15–21 April 2025 | 2010–2025; English/ Turkish | TS=(infertilit* OR "unexplained infertility" OR "idiopathic infertility") AND TS=("machine learning" OR "artificial intelligence" OR "deep learning" OR "neural network*") AND TS=(predict* OR "predictive model*" OR model* OR classification OR algorithm*) | n=157 |
| Cochrane | 15–21 April 2025 | 2010–2025; English/ Turkish | ("infertility" OR "unexplained infertility" OR "idiopathic infertility") AND ("machine learning" OR "artificial intelligence" OR "deep learning") | n=109 |
| Google Scholar | 15–21 April 2025 | 2010–2025; First 200 results screened | ((Unexplained infertility [Title]) OR ((Infertility [Title]) AND ((Machine learning [Title]) AND ((Randomized [Title]) OR (Randomized Controlled Trial [Title]) OR (Randomized Clinical Trial [Title])))) | n=301 |

**Appendix 2.** Studies excluded at full-text screening and reasons for exclusion

| Study (author, year) | Full reference | Reason for exclusion |
|---|---|---|
| Liao et al., 2020 | Liao S, Pan W, Dai WQ, Jin L, Huang G, Wang R, et al. *JAMA Network Open*. 2020;3(11):e2023654. | Heterogeneous infertility population; not specific to unexplained infertility |
| Liao et al., 2021 | Liao S, Jin L, Dai WQ, Huang G, Pan W, Hu C, et al. *Int J Intelligent Systems*. 2021;36(3):1331–1344. | Heterogeneous infertility population; not specific to unexplained infertility |
| Balogun et al., 2018 | Balogun JA, Egejuru NC, Idowu PA. *Computer Reviews Journal*. 2018;2(1):313–330. | Wrong population; general infertility prediction |
| Zhang et al., 2024 | Zhang R, Zhou L, Hao X, Yang L, Ding L, Xing R, et al. *Metabolites*. 2024;14(9):492. | Wrong population; metabolic profiling in general infertility |
| Subha et al., 2024 | Subha R, Nayana BR, Sumalatha P. *Eng Appl Artif Intell*. 2024;127:107400. | Wrong population; general infertility detection |
| Khan et al., 2024 | Khan FM, Akhter MS, Khan IU, Haider ZA, Khan NH. *Int J Innovations Sci Technol*. 2024;6(2):943–960. | Wrong population; general infertility classification |
| Vats et al., 2022 | Vats S, Sengupta A, Chaurasia A, Narad P. In: *Innovations in Computational Intelligence and Computer Vision*. 2022:547–555. | Wrong condition; PCOS-focused study |
| Raef & Ferdousi, 2019 | Raef B, Ferdousi R. *Acta Informatica Medica*. 2019;27(3):205–210. | Review article; non-original research |
| Hassan et al., 2020 | Hassan MR, Al-Insaif S, Hossain MI, Kamruzzaman J. *Neural Comput Appl*. 2020;32(7):2283–2297. | Wrong population; IVF pregnancy outcome prediction |
| Mehrjerd et al., 2022 (Sci Rep) | Mehrjerd A, Rezaei H, Eslami S, Ratna MB, Khadem Ghaebi N. *Scientific Reports*. 2022;12:7216. | Wrong population; general infertility treatment outcomes |
| Jha et al., 2024 | Jha T, Sirisha M, Bhargavi MS. In: *2024 Int. Conf. for Women in Computing (InCoWoCo)*. IEEE; 2024:1–7. | Wrong condition; PCOS/PCOD diagnosis |
| Shofiyah & Mahmudy, 2023 | Shofiyah S, Mahmudy WF. *8th Int. Conf. Sustainable Information Engineering & Tech*. 2023:235–240. | Review article; non-original |
| Barnett-Itzhaki et al., 2020 | Barnett-Itzhaki Z, Elbaz M, Butterman R, Amar D, Amitay M, Racowsky C, et al. *J Assist Reprod Genet*. 2020;37(10):2405–2412. | Wrong population; general IVF population |
| Abdullah et al., 2023 | Abdullah KAL, Atazhanova T, Chavez-Badiola A, Shivhare SB. *Reproductive Sciences*. 2023;30(4):1006–1016. | Review/conceptual article |
| Allameh et al., 2021 | Allameh F, Fallah KM, Zadehmodarres S, Abedi AR, Eslami MJ, Hajian MR, et al. *Machine learning approaches to predict intrauterine insemination success rate*. 2021. | Wrong population; IUI prediction |
| Medenica et al., 2022 | Medenica S, Zivanovic D, Batkoska L, Marinelli S, Basile G, Perino A, et al. *Diagnostics*. 2022;12(12):2979. | Review article; regulatory/ethical focus |
| Wang et al., 2024 | Wang C, Johansson AL, Nyberg C, Pareek A, Almqvist C, Hernandez-Diaz S, Oberg AS. *Fertil Steril*. 2024;122(1):95–105. | Outcome not aligned; prediction of pregnancy-related complications |
| Mehrjerd et al., 2022 (Endometrial thickness) | Mehrjerd A, Rezaei H, Eslami S, Khadem Ghaebi N. In: *Challenges of Trustable AI and Added-Value on Health*. 2022:264–268. | Study used classical statistics only (no ML/AI) |
| Goyal et al., 2020 | Goyal A, Kuchana M, Ayyagari KPR. *Scientific Reports*. 2020;10:20925. | Wrong population; general IVF live-birth prediction |
| Dehghan et al., 2024 | Dehghan S, Rabiei R, Choobineh H, Maghooli K, Nazari M, Vahidi-Asl M. *PLoS One*. 2024;19(10):e0310829. | Wrong population; general IVF success prediction |
| Pouresmaeili et al., 2023 | Pouresmaeili F, Alidoost S, Azimirad M, Azizmohammad Looha M, Emami Meibodi A, Abedin-Do A, et al. *Mol Biol Rep*. 2023;50(11):8785–8797. | Wrong population; unexplained recurrent miscarriage (RPL), not unexplained infertility |
| Koshy & Anuradha, 2022 | Koshy S, Anuradha K. *Int J Eng Technol Manag Sci*. 2022;6:287. | Review article; not original ML/AI research |
| GhoshRoy et al., 2023 | GhoshRoy, D., Alvi, P. A., & Santosh, K. C. (2023). *Journal of Medical Systems*. 47(1), 91. | **Review / non-original research; Wrong population** |