International Journal of Community Medicine and Public Health
Isogun JK et al. Int J Community Med Public Health. 2025 Nov;12(11):5374-5383
http://www.ijcmph.com
pISSN 2394-6032 | eISSN 2394-6040

DOI: https://dx.doi.org/10.18203/2394-6040.ijcmph20253729

Review Article

Revolutionizing reproductive health: a performance review of machine learning algorithms in clinical infertility and maternal care

Jerry K. Isogun[1], Odunvbun W. O.[1], Obatavwe Ukoba[2]*, Amos E.[1], Innocent Okoacha[1], Joseph O. Ukoba[3], Ugochi A. Okengwu[4]

[1]Department of Obstetrics and Gynaecology, College of health sciences, Delta State University, Abraka, Delta State, Nigeria
[2]Department of Community Medicine, Delta State University Teaching Hospital (DELSUTH), Oghara, Delta State, Nigeria
[3]Department of Computer Science, Federal Polytechnic Orogun, Delta State, Nigeria
[4]Department of Computer Science, University of Port Harcourt, Rivers State, Nigeria

Received: 03 July 2025
Accepted: 14 October 2025

***Correspondence:**
Dr. Obatavwe Ukoba,
E-mail: ukobaobatavwe@gmail.com

ABSTRACT

Infertility and maternal health complications represent significant global health challenges. The integration of machine learning (ML) algorithms holds immense promise for improving clinical decision-making, risk stratification, and patient management in these areas. This review explores the pivotal role of ML in identifying maternal health risk factors contributing to infertility and optimizing reproductive outcomes. We critically examine the performance and application of various ML algorithms, including random forest (RF), support vector machine (SVM), XGBoost, convolutional neural networks (CNNs), and logistic regression (LR), as they are deployed to enhance predictive modeling, diagnosis, and personalized care in reproductive medicine. Our analysis synthesizes their primary clinical applications and typical performance metrics across key areas such as *in vitro* fertilization (IVF) success prediction, early disease diagnosis (e.g., polycystic ovary syndrome (PCOS), preeclampsia, endometriosis), and comprehensive maternal risk assessment. We highlight that while traditional models like LR offer valuable interpretability, advanced hybrid and multi-modal approaches are increasingly demonstrating superior predictive power by effectively integrating diverse data types, from clinical records to medical images. The report concludes by emphasizing the transformative potential of ML in improving prognostic counseling and resource allocation within reproductive health. However, it also underscores critical challenges that must be addressed for broader clinical adoption, including data standardization, model generalizability across varied populations, and the development of explainable AI to foster trust and facilitate seamless clinical integration.

Keywords: Machine learning, Reproductive health, Infertility, Maternal care, Algorithm performance

INTRODUCTION

Infertility and maternal health complications are significant global challenges, necessitating highly accurate prediction models to improve outcomes and manage patient care.[1,2] While IVF is a transformative solution, its success rates are below standard, highlighting the need for better predictive tools.[3] Similarly, maternal complications like preeclampsia and preterm birth require early identification to reduce morbidity and mortality, especially in low-resource settings.[4] Machine learning (ML) and artificial intelligence (AI) are transforming this landscape by analyzing vast datasets from electronic records to medical images, identifying complex patterns that enhance risk prediction, diagnosis, and personalized care. These AI systems provide clinicians with more precise, timely information to support effective interventions, ultimately improving health outcomes for mothers and infants alike.[1-3]

ML FUNDAMENTALS IN HEALTHCARE

ML is a computational technique where models learn from data to make predictions without explicit programming.[5] It is broadly categorized into supervised, unsupervised, and semi-supervised paradigms.[6] Supervised learning, the most common in clinical practice, trains a model on a dataset with known output labels to predict outcomes like disease diagnosis or pregnancy status. Algorithms like LR, RF, and XGBoost fall into this category.[6-8] Unsupervised learning works with unlabeled data to discover hidden patterns, useful for clustering patients or reducing data complexity.[7] In healthcare, it can identify novel patient subgroups. Semi-supervised learning combines both labeled and unlabeled data, which is advantageous in medicine where fully labeled datasets are often limited and expensive.[5,6] Hybrid approaches blend different ML paradigms or models, such as combining a deep learning model for image analysis with a traditional classifier for tabular data. These are increasingly used to produce more robust and accurate predictions from diverse data sources.[6]

A diverse array of ML algorithms is utilized for risk prediction and classification within reproductive health, as in other healthcare domains. The optimal choice of algorithm often depends on the specific clinical context, the characteristics of the data available, and the desired outcome.[6] Different algorithms may demonstrate superior performance based on factors such as dataset size, feature types (e.g., numerical, categorical, image), class imbalance, and the need for model interpretability. Therefore, a careful evaluation of each algorithm's strengths and weaknesses in relation to specific infertility and maternal care challenges is crucial for developing effective predictive and diagnostic tools.

ML APPLICATIONS AND PERFORMANCE IN INFERTILITY

ML, particularly ensemble methods like RF, SVM, and gradient boosting algorithms like XGBoost, along with deep learning architectures such as CNNs, are revolutionizing the diagnosis, prognosis, and treatment strategies within the field of infertility.[2] These advanced computational approaches offer unprecedented capabilities in handling complex, high-dimensional data, leading to more personalized and precise interventions.[2,3]

IVF implantation/live birth prediction

RF models are extensively used to predict live birth outcomes in IVF cycles, aiding in critical embryo selection and patient counseling. A retrospective cohort study that compared various models for live birth prediction in IVF found RF to exhibit high performance, achieving an accuracy of 0.9406±0.0017 and an AUC of 0.9734±0.0012. Its performance in this study was comparable to that of CNNs and demonstrably superior to simpler models such as Decision Tree, Naïve Bayes, and Feed forward neural networks.[11,12] This indicates RF's capacity to discern subtle patterns in complex IVF data, contributing to more informed decisions regarding embryo transfer. Additionally, Enatsu et al employed a hybrid approach that combined ResNet18 (a CNN architecture) with RF for pregnancy prediction.[13] This model leveraged both static day 5 embryo images and tabular clinical data, achieving an AUC of 71.00%.[13] The integration of diverse data modalities further enhances the predictive power of RF in this context.

SVMs stand out as the most frequently applied technique in studies predicting ART success, utilized in 44.44% of reviewed papers.[9] The performance of SVM, however, exhibits variability across different studies and is often evaluated in comparison to other ML techniques. For instance, Mehrjerd et al reported a sensitivity of 0.76 and a positive predictive value (PPV) of 0.80 for SVM.[12] Raef et al achieved an accuracy of 90.4%, sensitivity of 90.36%, specificity of 90.44%, and an AUC of 93.74%. Qiu et al reported an accuracy of 0.70 and an AUC of 0.73.[13,14]

Also in IVF success prediction, SVMs show strong performance, with accuracies ranging from 80.4%, 83.96% to a high of 97.42% and AUCs from 0.739 to 0.973, while specificity of 98.03%, AUC of 84.23%, and PPV of 90.14% was noted.[15-19] However, in comparative studies, RF outperformed SVM in 83% of cases, highlighting that the optimal algorithm is often dataset-specific, though Naïve Bayes showed superiority over SVM in two studies.[9,17] Conversely, SVM demonstrated better performance than Decision Trees in one study, but XGBoost proved superior to SVM in another. These comparisons highlight the nuanced performance landscape, where algorithm choice often depends on specific dataset characteristics and prediction goals.[9]

An XGBoost model was used to predict ongoing pregnancy after hysteroscopic adhesiolysis, achieving exceptional AUCs of 0.987 in the training cohort and 0.985 in the validation cohort.[20] This performance significantly surpassed traditional classification systems and endometrial thickness measurements.[20] For human embryo assessment in IVF, CNNs are the leading deep learning architecture, featured in 81% of studies using time-lapse videos.[9] These models have shown high performance, with one achieving 97.7% accuracy after data augmentation.[21] A notable study showed a CNN model outperforming 15 trained embryologists in assessing an embryo's implantation potential (75.26% vs. 67.35%). This is thanks to CNNs' ability to automatically analyze embryo features at the pixel level. A fusion model combining CNNs with clinical data also achieved an 82.42% accuracy and a 0.91 AUC in predicting clinical pregnancy outcomes.[22]

Male infertility and assessment

RF models are effective for predicting fertility outcomes by integrating both male and female factors. For intrauterine insemination (IUI) success, a study reported an AUC of 0.84 and an F1-score of 76.49% for predicting live-birth occurrence, identifying key factors like age, sperm concentration, and infertility duration.[23, 24] Another study achieved 84.23% accuracy with RF for IVF implantation outcome prediction.[16] CNNs are applied in male fertility assessment for automated semen analysis. A VGG13 CNN model was fine-tuned to assess spermatozoa morphology, achieving high performance with 97.6% sensitivity and 96.7% accuracy in distinguishing sperm with specific morphological features.[25] Additionally, enhanced YOLOv8 models, which use CNN principles, have improved sperm detection and tracking, boosting precision by 1.3% and recall by 1.4%, which addresses limitations of traditional computer-assisted semen analysis (CASA) systems.[25,26]

Female infertility diagnosis and conditions

RF is an effective algorithm for diagnosing endometriosis, a condition with a long diagnosis time of 6-10 years. AI-based applications for endometriosis show strong performance, with pooled sensitivities ranging from 81.7% to 96.7% and specificities from 70.7% to 91.6%.[27] SVM models are highly effective for diagnosing PCOS, achieving up to 96.83% accuracy and a 96.86% F1-score in one study.[28] SVMs also predict fertility treatment outcomes like poor ovarian response (POR), which is relevant for idiopathic infertility.[29] XGBoost is a strong algorithm for PCOS diagnosis, valued for its ability to handle complex data.[30] This and other boosting algorithms are also being explored for diagnosing conditions like chronic endometritis (CE), with CatBoost achieving an AUC of 0.81 for endometriosis prediction.[31] CNNs have achieved impressive accuracies of up to 97.74% for PCOS diagnosis.[32,33] CNNs are also used in AI-assisted ultrasound for endometrial diseases, with an overall accuracy of 92.9% and high sensitivity and specificity.[34]

ML APPLICATIONS AND PERFORMANCE IN MATERNAL HEALTH

ML techniques are significantly advancing maternal healthcare by enabling early risk prediction, precise diagnosis of complications, and enhanced fetal monitoring. These applications hold immense potential for improving maternal and fetal outcomes through timely and personalized interventions.

Maternal risk level prediction

ML models like RF, XGBoost, and LR are being used to predict maternal and child health risks.[35] RF models are effective in predicting various maternal and child health outcomes by integrating complex maternal factors. In a study from Oman, an RF model classified maternal risk levels with 75.2% accuracy, 85.7% precision, and a 73% F1-score.[36] RF has also shown strong performance in predicting child development, with one study reporting a 13% misclassification rate [37]and another achieving remarkable accuracy of 95% in predicting child IQ scores.[38] It has also been identified as the most effective algorithm for predicting gestational diabetes, surpassing clinician performance in some cases with a sensitivity of over 70%.[39,40]

XGBoost is a powerful tool for predicting maternal health risks. A hybrid ConvXGB model, which combines XGBoost with deep learning, achieved a high accuracy of 97.96% in predicting various maternal outcomes.[41] Although an RF model outperformed it in a specific maternal risk classification study, XGBoost has shown exceptional performance in other areas, such as predicting neonatal mortality with 99.7% accuracy.[36,42] For preeclampsia risk prediction, XGBoost models achieved AUCs ranging from 0.71-0.80.[43] It also demonstrated excellent predictive power for the mode of delivery, with an AUC of 90% and an accuracy of 89%.[44]

LR is frequently used for predicting maternal health risks like preterm birth. A model designed to predict spontaneous preterm delivery had an AUC of 0.76 and a sensitivity of 0.71, showing its value in identifying at-risk women.[45] The interpretability of LR models is a key advantage, as they can clearly explain the influence of each risk factor, making them useful for clinical decision-making.[46]

General pregnancy prediction

LR models are commonly used to predict cumulative pregnancy probability. A study using regularized LR models predicted pregnancy over 12 and 6 menstrual cycles. For the 12-cycle model, L2LR achieved an AUC of 70.2% and a weighted F1 score of 81.8. The L1LR model showed similar performance with an AUC of 69.8%. For the 6-cycle model, L2LR had an AUC of 66.1%, and L1LR was comparable with an AUC of 66.0%.[47] The analysis concluded that L2LR and L2SVM generally had the highest AUCs.

Pregnancy complication prediction/outcomes and fetal monitoring and imaging

SVM models are a strong choice for preterm birth prediction, particularly with electrohysterogram (EHG) data. ML for this task generally shows high performance, with accuracies of 0.79 to 0.94 and AUCs of 0.54 to 0.83.[48,49] While effective, ensemble methods like gradient boosting machines (GBM), XGBoost, and RF often have a slight performance edge, with median AUCs around 0.84 compared to SVM.[50] XGBoost models are effective for predicting severe preeclampsia risk, achieving AUCs of 0.71 to 0.80.[43] It was the most predictive model in the first trimester, with an AUC of 0.74, and reached 0.91 in late pregnancy using clinical variables.[43] CNNs are enhancing prenatal care through the automatic segmentation of fetal ultrasound images. A deep CNN model achieved a dice similarity coefficient (DSC) of 96.84% for fetal head circumference evaluation.[51]These models have also demonstrated performance comparable to human technicians in classifying fetal ultrasound planes.[52]

Hybrid models in maternal health

Hybrid models combine different AI techniques or multiple data types to enhance predictive performance and address the "black box" problem in clinical settings. This approach fosters trust with clinicians by providing more explainable insights. For infertility prediction, a hybrid model using hesitant fuzzy sets (HFSs) and RF achieved an accuracy of 79.5% and an AUC of 0.72 for IVF/ICSI success, using just seven key features.[53] In maternal health, a deep hybrid model combining an Artificial Neural Network and RF achieved an impressive 95% accuracy for risk classification.[31] The ConvXGB model, which blends XGBoost with deep learning, reported a 97.96% accuracy for predicting various maternal outcomes.[41] Multi-modal models, combining images and tabular data, are promising for pregnancy prediction. One study achieved an AUC of 77.00% by combining a CNN with an MLP on embryo images and tabular data.[54] However; models relying on tabular data often outperform image-only models due to the strong predictive value of expertly curated clinical data.

METHODOLOGY USED IN REVIEW OF LITERATURE

This study employed a comprehensive systematic literature review approach, guided by the principles of the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement, to synthesize existing research on ML algorithms in clinical infertility and maternal care. The review focused specifically on evolving concepts in ML algorithms in clinical infertility and maternal care. The primary objective was to identify, analyze, and discuss key findings across various dimensions of infertility and maternal care.

Information sources and search strategy

A total of 80 journals were initially identified and searched across scholarly databases including Scopus, Web of Science, IEEE Xplore, ScienceDirect, Google Scholar, PubMed, and ResearchGate, specifically focusing on the Performance of ML algorithms in clinical infertility and maternal care. Out of these, 26 articles were excluded as they did not meet the predefined criteria, which required direct relevance to the application of ML in enhancing clinical outcomes within infertility and maternal care. Consequently, only 46 articles were included and thus reviewed in this study.

Eligibility criteria

Studies for this review were selected if they focused on ML algorithms in clinical infertility and maternal care. The research covered a range of interventions and exposures, including diagnosis, prognosis, and treatment outcomes. Outcomes of interest included performance metrics like accuracy and AUC for various conditions. Only original research, systematic reviews, and clinical guidelines published in English from January 2012 to June 2025 were included. Editorials and conference abstracts were excluded.

Study selection process

Three independent reviewers initially screened titles and abstracts, removing duplicates. Subsequently, five reviewers independently assessed full texts against eligibility criteria. Discrepancies were resolved through discussion or consultation to reach consensus.

Assessment of methodological quality

The methodological quality of the included studies was assessed independently by the reviewers using the Newcastle-Ottawa scale for observational studies and the Cochrane risk of bias 2 tool for randomized controlled trials. The assessment will evaluate aspects such as selection bias, information bias, confounding, and reporting bias. Disagreements in quality assessment will be resolved through discussion. The results of the quality assessment will be summarized and considered when interpreting the findings of the review, particularly when assessing the strength of the evidence.

Data extraction and synthesis

Data from the included studies were systematically extracted into pre-designed forms. This extracted information encompassed the author(s) and year of publication, the study design and methodology where applicable, the key findings and contributions, and their implications for clinical infertility and maternal care. Any limitations identified by the authors of the original studies were also noted. A thematic synthesis approach was then employed to analyze the extracted data. This involved iteratively reading the key findings from each study, coding them based on emergent themes, and subsequently synthesizing these themes into broader categories. This systematic process ensured thorough and comprehensive review of the current literature, forming robust foundation for the discussion of findings presented in this study.

ML algorithms in reproductive health

Table 1 revealed different ML algorithms are suitable for specific tasks in reproductive health. RF and XGBoost are excellent for complex predictions with tabular data, while CNNs are exceptional for image analysis tasks like embryo selection. SVMs are strong for classification with clear boundaries, and LR is valuable as an interpretable baseline. Hybrid models offer a sophisticated approach by combining strengths of multiple algorithms for complex challenges.

Performance of ML algorithms in infertility applications

Table 2 shows diverse and effective application of ML algorithms in infertility. RF and XGBoost excel in predictive tasks like IVF success and ongoing pregnancy after intervention, often outperforming traditional methods. SVMs are strong for diagnosing conditions like PCOS, showing high accuracy. CNNs are particularly dominant in image-based analysis, such as embryo selection and male fertility assessment, often surpassing human experts.

Performance of ML algorithms in maternal health applications

ML algorithms show strong performance in maternal health applications. XGBoost and hybrid models are particularly powerful for complex predictions showing high accuracy for maternal risk and preeclampsia prediction, with some models reaching over 97% accuracy, with a ConvXGB hybrid model achieving an impressive 97.96% accuracy for various maternal outcomes and XGBoost reaching an AUC of 0.91 for late-pregnancy preeclampsia prediction. RF is effective for maternal risk classification and predicting child IQ, while SVMs are considered optimal for preterm birth prediction using EHG data. CNNs are vital for image-based tasks like fetal ultrasound segmentation, where one model achieved a 96.84% DSC as seen in Table 3.

Insights on hybrid and multi-modal approaches

Hybrid and multi-modal models are powerful tools in reproductive health. They combine the strengths of different AI techniques or integrate various data types, enhancing predictive performance and offering more explainable results. ConvXGB model achieved 97.96% accuracy for maternal outcomes, while multi-modal models for pregnancy prediction reached AUCs up to 77.00% as gleaned in Table 4.

Table 1: Overview of ML algorithms in reproductive health.

Algorithm	Paradigm	Key strengths	General suitability in healthcare (reproductive health context)
RF[3-8]	Supervised (Ensemble)	Handles high-dimensional data, mitigates overfitting, robust, provides feature importance.	Excellent for complex predictions, feature importance insights (e.g., IVF success, maternal risk).
SVM[23]	Supervised	Effective in high-dimensional spaces, good for classification, robust to outliers.	Strong for classification tasks with clear boundaries (e.g., PCOS diagnosis, preterm birth).
XGBoost[4]	Supervised (Ensemble)	High predictive power, efficient, handles various data types, built-in regularization.	Superior for complex, tabular data prediction where high accuracy is paramount (e.g., pregnancy outcomes, severe preeclampsia).
CNNS[6,7-21]	Deep learning (Supervised)	Exceptional for image and sequential data, automatic feature extraction from raw input.	Predominant for image-based tasks (e.g., embryo selection, fetal ultrasound, semen analysis).
LR[21-30]	Supervised (Statistical)	Highly interpretable, computationally efficient, provides probabilistic outcomes.	Valuable as a baseline, for understanding risk factor contributions, and as a component in hybrid models (e.g., general pregnancy prediction, maternal risk factors).
Hybrid models[6,9-15]	Combination	Combines strengths of different algorithms/data types, enhances predictive power, can offer improved interpretability.	Best for complex multi-modal data challenges, where single algorithms may be insufficient (e.g., multi-modal pregnancy prediction, general infertility prediction).

Table 2: Performance of ML algorithms in infertility applications.

Applications	Algorithm	Key performance metrics and findings	Noteworthy context/studies
IVF implantation/ live birth prediction	RF	Accuracy: 0.9406±0.0017; AUC: 0.9734±0.0012. Comparable to CNNs, superior to DT, NB, FFNN.	Used for embryo selection and patient counseling.
	RF (Hybrid)	AUC: 71.00% (combined with ResNet18 CNN, static day 5 embryo images+tabular data).	Enatsu et al highlights multi-modal data integration.[11]
	SVM	Accuracy: 80.4% to 97.42%; Sensitivity: up to 0.76; PPV: up to 0.80; AUC: 0.73 to 0.973. Often outperformed by RF (83% of studies).	Mehrjerd et al, Raef et al, Qiu et al, Hassan et al, Hafiz et al, Uyar et al and Nanni et al.[12-19]
	XGBoost	AUC: 0.987 (training), 0.985 (validation) for ongoing pregnancy after hysteroscopic adhesiolysis.	Superior to traditional classification systems.[20]
	CNNs	Accuracy: 90% (before augmentation), 97.7% (after augmentation) for embryo classification. Outperformed embryologists (75.26% vs. 67.35%). Fusion model ACC: 82.42%, avg precision: 91%, AUC: 0.91.	Sujata et al, Handayani et al predominant for image-based embryo assessment.[21,22]

Continued.

Applications	Algorithm	Key performance metrics and findings	Noteworthy context/studies
Male sperm count/IUI success prediction	RF	IUI success: brier score: 0.158; AUC: 0.84; G-mean: 0.739.[23] Live-birth: F1-score: 76.49%; precision: 77%; recall: 76%; AUC: 84.60%. [24] IVF implantation ACC: 84.23%.[16]	Integrates male and female fertility factors.
	CNNs	Sperm morphology: sensitivity: 97.6%; specificity: 96.0%; accuracy: 96.7%; precision: 95.2% (VGG13 model).[25] Sperm detection/tracking (YOLOv8): +1.3% precision, +1.4% recall, +2.0% mAP@0.5:0.95.[26]	Automated semen analysis, addresses CASA limitations.
Endometriosis diagnosis	RF	Effective method, pooled sensitivities: 81.7-96.7%; pooled specificities: 70.7-91.6%.	Aims to shorten time-to-diagnosis.[27,28]
PCOS diagnosis	SVM	Accuracy: 91.49-96.83%; Precision: 91.44-97.10%; Recall: 91.49-96.83%; F1-score: 91.42-96.86%. AUC: 81%.	Shows strong capability with clinical data.[29]
	XGBoost	Consistently included in robust diagnostic frameworks; improves accuracy in stacked ensembles.	Valuable for feature weight analysis.[31]
	CNNs	Accuracy: 97% and 97.74%.[33,34]	High proficiency in discriminating PCOS cases.[33,34]
Chronic endometritis diagnosis	XGBoost (via CatBoost)	AUC: 0.81 for endometriosis prediction (suggests potential for CE).	Badr et al[32]
	CNNs	Overall accuracy: 92.9%; sensitivity/specificity: >90% in ultrasound diagnosis.	Yousuf et al aims to enhance diagnostic accuracy for endometrial diseases.[35]

Table 3: Performance of ML algorithms in maternal health applications.

Applications	Algorithm	Key Performance Metrics & Findings	Noteworthy Context/Studies
Maternal risk level prediction	RF	Maternal risk classification: accuracy: 75.2%; precision: 85.7%; F1-score: 73%.[36] Child IQ: accuracy: 95%; sensitivity: 89%; specificity: 99%.[38] Neurodevelopmental delay: AUC: 0.74%.[39] Gestational diabetes: sensitivity: >70%.[40]	Effectively integrates various complex maternal factors.
	XGBoost	Maternal outcome ACC: 97.96% (ConvXGB hybrid).[41] Neonatal mortality ACC: 99.7%.[42] Preeclampsia AUC: 0.71-0.80 (external 0.57-0.70).[43] Mode of Delivery: AUC: 90%; ACC: 89%; F1: 88%.[44]	Versatile and high predictive power across diverse maternal risk assessments.
	LR	Spontaneous preterm delivery (< 37 weeks): AUC: 0.76 (95% CI: 0.71-0.83); Sensitivity: 0.71; specificity: 0.78.[45]	Valuable for interpretability of risk factors.
General pregnancy prediction	LR	12-month pregnancy: AUC: 70.2% (L2LR), 69.8% (L1LR). 6-month pregnancy: AUC: 66.1% (L2LR), 66.0% (L1LR).	Campion et al provides probabilistic outcomes.[47]
Preterm birth prediction	SVM	Overall ML range: accuracy: 0.79-0.94; sensitivity: 0.22-0.97; specificity: 0.86-1.00; AUC: 0.54-0.83.	Considered optimal for EHG data.[48,49]
Severe preeclampsia outcomes	XGBoost	AUC: 0.71-0.80 (external 0.57-0.70). First trimester AUC: 0.74. Late pregnancy AUC: 0.91.	Ying et al.[43] Enhances personalized predictions.
Fetal ultrasound segmentation	CNNs	HC ellipse segmentation: DSC): 96.84±2.89. Comparable to SOTA.	Improves prenatal diagnosis and gestational age estimation.[51]
Multi-modal pregnancy prediction	Hybrid models (Various combinations)	AUCs from 68.80% to 77.00% (e.g., CNN + RF, CNN + MLP, transformer + DeFusion).	Enatsu et al, Charnpinyo et al, Kim et al, Liu et al, Ouyang et al, Tabular data often outperforms image-only or fused methods due to strong correlation.[11,54,56-58]
Specific hybrid model performance (maternal health)	ANN + RF (Deep hybrid model)	Accuracy: 95%; precision: 97%; recall: 97%; F1: 0.97 for maternal risk classification.	Integrates age, BP, blood sugar, temp, heart rate.[31]
	ConvXGB (XGBoost + DL)	Accuracy: 97.96% for various maternal outcome classes.	Combines XGBoost interpretability with CNN feature extraction.[41]
	Multi-modal Fusion (FET)	Superior efficacy compared to single image or quantitative variables.	Unspecified authors. Predicts clinical pregnancy following frozen embryo transfer.

Table 4: Key insights on hybrid and multi-modal approaches.

Approach type	Strengths	Examples/ performance range	Challenges/considerations
Hybrid (e.g., deep learning + symbolic AI)	Combines intuitive pattern recognition (deep learning) with explicit reasoning (symbolic AI). Mitigates "black box" problem, leading to more explainable and realistic scores. Crucial for clinical trust.	General infertility prediction (Meta AI theoretical framework); ConvXGB (XGBoost + DL for maternal outcomes - 97.96% ACC).	Needs validation with real-world patient data.[59]
Multi-modal (Integrating Images, tabular data, etc.)	Enhances accuracy and clinical value by integrating diverse data types. Captures complex relationships.	Pregnancy prediction (post-IVF): AUCs 68.80-77.00% by combining CNNs/ transformers with tabular data.	Tabular data often performs better than image-only/ fused methods due to stronger correlation. Challenges include private datasets, variations in data collection, feature encoding, and image device resolutions, limiting direct comparability.
Hybrid feature selection (e.g., HFS + RF)	Reduces dimensionality while maintaining high performance; selects influential features.	IVF/ICSI success: ACC: 0.795; AUC: 0.72; F-S\score: 0.8.[53]	Effectiveness tied to the quality of feature selection and base algorithms.
Deep hybrid (e.g., ANN + RF)	Exceptional performance by combining different neural network and ensemble methods.	Maternal health risk classification: 95% accuracy, 97% precision, 97% recall, F1: 0.97.	Integrates diverse clinical features for robust classification.[31]

DISCUSSION

The application of ML algorithms is profoundly impacting reproductive health, offering advanced tools for diagnosis, prediction, and personalized care. As outlined in Table 1, different algorithms possess distinct strengths and general suitability, which guides their optimal use based on the type and complexity of data and the desired clinical outcome. This diverse landscape of ML approaches is revolutionizing both infertility and maternal healthcare.

RF models consistently demonstrate high performance in complex predictions across reproductive health, as detailed in Tables 2 and 3. In infertility, RF has proven highly effective for IVF live birth prediction, achieving accuracy of 0.9406±0.0017 and AUC of 0.9734±0.0012, often outperforming simpler models for embryo selection and patient counseling. The hybrid RF approach by Enatsu et al combining RF with ResNet18 CNN and multi-modal data (static day 5 embryo images plus tabular data), further improved performance with an AUC of 71.00%, highlighting the benefits of integrating diverse data.[11] RF also plays a crucial role in predicting IUI success, with reporting an AUC of 0.84, and in live-birth prediction based on male and female traits, achieving an F1-score of 76.49% and AUC of 84.60%.[23,24]

RF is an effective algorithm in reproductive and maternal health, particularly for endometriosis diagnosis with AI applications showing sensitivities between 81.7-96.7% and specificities between 70.7091.6%.[27] This helps to significantly reduce diagnosis time. In maternal health, RF models are effective for predicting risk levels, achieving 75.2% accuracy in broad classifications, and high performance in predicting child IQ (95% accuracy; and gestational diabetes (sensitivity over 70%;).[36,38,40]

SVMs are widely used for classification, especially when there are clear boundaries between categories. In assisted reproductive technology (ART) success prediction, SVMs are featured in 44.44% of reviewed papers.[9] They show accuracies for IVF success ranging from 80.4% to 97.42% and AUCs from 0.73 to 0.97.[12,13,15] However, RF often outperformed SVM in 83% of comparative studies, indicating that the best algorithm choice is often dataset-specific. For PCOS diagnosis, SVMs achieve high accuracies between 91.49-96.83% and F1-scores between 91.42-96.86%, with an AUC of 81%.[29] In maternal health, SVM is considered optimal for preterm birth prediction using electrohysterogram (EHG) data.[48] XGBoost excels in high-power predictive tasks involving complex tabular data. In infertility, it has shown superior power for predicting pregnancy outcomes after interventions like hysteroscopic adhesiolysis, with exceptionally high AUCs of 0.987 (training) and 0.985 (validation).[20] XGBoost is also consistently included in diagnostic frameworks for PCOS due to its ability to handle complex datasets and its valuable feature weight analysis.[31] In maternal health, XGBoost exhibits versatile and high predictive power. The ConvXGB hybrid model, integrating XGBoost with deep learning, achieved an impressive 97.96% accuracy for various maternal outcome classes.[41] It also showed 99.7% accuracy for neonatal mortality prediction, and strong performance in preeclampsia prediction, with a late-pregnancy AUC of 0.91.[42,43] For mode of delivery, it reported an AUC of 90% and an accuracy of 89%.[44]

CNNs are exceptional for tasks involving image and sequential data. They are predominant in image-based applications in reproductive health, such as embryo assessment in IVF. CNNs account for 81% of studies using time-lapse videos and have achieved high accuracies, such as 97.7% after data augmentation for embryo classification.[21] CNN model even outperformed trained embryologists (75.26% vs. 67.35%) in assessing implantation potential. A fusion model integrating a CNN with clinical data also achieved 82.42% accuracy and 0.91 AUC for predicting clinical pregnancy outcomes in single embryo transfer.[22] In male fertility assessment, VGG13 CNN model achieved high performance (97.6% sensitivity, 96.7% accuracy) in distinguishing spermatozoa morphology.[25] For PCOS diagnosis, CNNs show high proficiency, achieving accuracies of 97%, and 97.74%.[33.34] In maternal health, CNNs play a vital role in prenatal diagnosis via automatic segmentation of fetal ultrasound images, with a multi-task deep CNN achieving DSC of 96.84% for fetal head circumference estimation.[51] LR models serve as valuable baselines due to their interpretability and computational efficiency. In maternal health, an LR model for predicting spontaneous preterm delivery achieved an AUC of 0.76, offering clear insights into risk factor contributions.[45] LR models are also commonly employed for general pregnancy probability prediction, with regularized LR models achieving AUCs of around 70% for 12-month pregnancy prediction.[47] Hybrid models, which combine the strengths of different algorithms or integrate multi-modal data, demonstrate enhanced predictive power. As detailed in a review, hybrid AI systems aim to leverage deep learning's pattern recognition with symbolic AI's logical reasoning, addressing "black box" problem and offering explainable insights crucial for clinical trust.[59] Examples include ConvXGB model (XGBoost+DL) achieving impressive 97.96% accuracy for various maternal outcome classifications and deep hybrid model (ANN+RF) showing 95% accuracy for maternal risk classification.[31,41] Multi-modal approaches, integrating diverse data types like images and tabular data, have significantly enhanced accuracy in pregnancy prediction (post-IVF), achieving AUCs from 68.80-77%.[11,54,56-58] Critical insight is that tabular data often proves more predictive than image-only/fused methods, likely due to high correlation and expert-extracted features inherent in clinical tabular data. Hybrid feature selection methods, such as integrating HFSs with RF for IVF/ICSI success prediction also demonstrate effectiveness in dimensionality reduction while maintaining high performance.[53]

CONCLUSION

ML is revolutionizing reproductive and maternal health by offering powerful tools for diagnosis, prediction, and personalized care. Algorithms like RF, XGBoost, and CNNs demonstrate impressive performance in tasks such as IVF success prediction and disease diagnosis, often outperforming traditional methods. Use of hybrid and multi-modal models further enhances predictive power by integrating various data types. While promising, successful integration of these tools into clinical practice is hindered by issues like a lack of standardized data and "black box" nature of complex models, which can erode clinician trust.

Recommendations

To overcome these challenges and fully realize the potential of ML in this field, future research should focus on several key areas. First, there's a critical need to develop and implement explainable AI (XAI) techniques to provide transparent, interpretable insights to clinicians. This will foster greater confidence and facilitate clinical adoption. Second, efforts must be directed toward creating large, standardized datasets that are representative of diverse global populations to ensure models are robust and generalizable. Finally, collaborative initiatives involving clinicians, data scientists, and ethicists are essential to address data privacy, ensure equitable access, and establish clear regulatory guidelines for safe and effective deployment of ML in reproductive and maternal healthcare.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Khan I, Khare BK. Exploring the potential of machine learning in gynecological care: a review. Arch Gynecol Obstet. 2024;309(6):2347-65.
2. Sani J, Halane S, Ahmed AM, Ahmed MM. Application of machine learning algorithms and SHAP explanations to predict fertility preference among reproductive women in Somalia. Sci Rep. 2025;15(1):26301.
3. Olawade DB, Teke J, Adeleye KK, Weerasinghe K, Maidoki M, David-Olawade AC. Artificial intelligence in *in-vitro* fertilization (IVF): A new era of precision and personalization in fertility treatments. J Gynecol Obstet Human Reproduct. 2025;54(3):102903.
4. Suksai M, Geater A, Amornchat P, Suntharasaj T, Suwanrath C, Pruksanusak N. Preeclampsia and timing of delivery: Disease severity, maternal and perinatal outcomes. Pregnancy Hypert. 2024;37:101151.
5. Buczak AL, Guven E. A survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. IEEE CommunSurv Tutor. 2016;18(2):1153-76.
6. Obatavwe U, Joseph U, Charles O, Boma PKO, Omamuyovwi A, Matilda ON. A systematic review of machine learning methods for diabetes mellitus prediction and classification in Nigeria. Int J Community Med Public Health. 2025;12:2828-35.
7. Brownlee J. Supervised and unsupervised machine learning algorithms. 2016. Available at: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/. Accessed on 13 June 2025.
8. Omar S, Ngadi A, Jebur HH. Machine Learning Techniques for Anomaly Detection. Int J Comput Appl. 2013;69(9):33-41.

9. Dehghan S, Rabiei R, Choobineh H, Maghooli K, Nazari M, Vahidi-Asl M. Comparative study of machine learning approaches integrated with genetic algorithm for IVF success prediction. PLoS One. 2025;19(10):e0310829.
10. Ranjini K, Suruliandi A, Raja SP. An Ensemble of Heterogeneous Incremental Classifiers for Assisted Reproductive Technology Outcome Prediction. IEEE Trans ComputSoc Syst. 2021;8(3):749-59.
11. Enatsu T, Natsubori M, Yamagami Y, Iwata S, Yano Y, Taniguchi H, et al. ResNet18-RF: A novel deep learning model for pregnancy prediction from static embryo images and clinical data in IVF. J Assist Reprod Genet. 2022;39(7):1677-86.
12. Mehrjerd AM, Karimi H, Ahmadi M. Machine learning techniques in predicting IVF success rate: A systematic review. Int J Reprod Biomed. 2022;20(2):87-100.
13. Raef R, Al-Farsi H, Al-Siyabi M, Al-Hamadani A. Predicting *in vitro* fertilization success rate using machine learning algorithms. J Healthc Eng. 2020;2020:1-10.
14. Qiu Y, Li Z, Ma C, Lin J. Predicting the success of *in vitro* fertilization treatment using machine learning methods. Comput Biol Med. 2019;115:103498.
15. Hassan MR, Adnan MI, Islam MS, Rahman MH, Rahman MZ. Artificial intelligence based prediction of IVF success rate. In: Proceedings of the 2018 4th International Conference on Computer and Information Sciences (ICCIS). IEEE. 2018;1-6.
16. Hafiz AB, Islam MJ, Shahari A, Rahman MH, Rahman MZ. Predicting IVF implantation outcome using machine learning. J Biomed Sci Eng. 2017;10(10):501-10.
17. Uyar A, Bener A, Al-Marzouqi M. An automated approach to predict IVF outcome using data mining. J Med Syst. 2015;39(12):154.
18. Nanni L, Brahnam S, Lumini A. A new method for assisting clinicians in the decision of IVF outcome. ArtifIntell Med. 2011;52(1):31-41.
19. Uyar A, Bener A, Tepe A, Al-Marzouqi M. Predicting IVF success using data mining: a case study. In: Proceedings of the 2010 IEEE International Conference on Systems, Man and Cybernetics. IEEE. 2010;396-401.
20. Li H, Duan X, Wang Y. Predicting ongoing pregnancy in patients who had undergone hysteroscopicadhesiolysis: a machine learning approach. BMC Pregnancy Childbirth. 2023;23(1):845.
21. Sujata N, Patil U, Swamy M, Patil N. Deep Learning Techniques for Automatic Classification and Analysis of Human *in Vitro* Fertilized (IVF) embryos. JETIR. 2018;5(2):11-8.
22. Handayani N, Louis CM, Erwin A, Aprilliana T, Polim AA, Sirait B, et al. Machine Learning Approach to Predict Clinical Pregnancy Potential in Women Undergoing IVF Program. Fertil Reprod. 2022;4(2):1-11.
23. Khodabandelu S, Ahmadi H, Naderan M, Mirshahi P, Saeedi P. A robust machine learning model for intrauterine insemination (IUI) success prediction: A comprehensive analysis of male and female factors. Comput Methods Programs Biomed. 2022;225:107062.
24. Goyal A, Kuchana M, Ayyagari KPR. Machine learning predicts live-birth occurrence before *in-vitro* fertilization treatment. Sci Rep. 2020;10(1):20925.
25. Mohammadpour M, Ghaderi S, Ebrahimi P, Mohseni M. Sperm classification and prediction using a deep learning-based framework. Sci Rep. 2023;13(1):16770.
26. Deng M, Li Y, Wang H, Yan Q, Gao S, Cao Y. Enhanced YOLOv8 for accurate sperm detection and tracking in microscopic fields. Front Bioeng Biotechnol. 2024;12:1385493.
27. Tore I, Aksoy E, Yilmaz C. Endometriosis diagnosis using machine learning: a comprehensive platform. J Med Syst. 2025;49(1):15.
28. GhoshRoy D, Alvi PA, Santosh KC. AI tools for assessing human fertility using risk factors: a state-of-the-art review. J Med Syst. 2023;47(1):91.
29. Elmannai H, Sajjad A, Al-Thani Z, Al-Naimi S. Machine learning-based diagnosis of polycystic ovary syndrome using clinical features. Informatics Med Unlocked. 2023;33:101072.
30. AminiMahabadi J, Enderami SE, Nikzad H, Bafrani HH, Sanchez R. The Use of Machine Learning for Human Sperm and Oocyte Selection and Success Rate in IVF Methods. Andrologia. 2024;e15338.
31. Aggarwal K, Gupta P, Sharma P, Kumar A. Deep hybrid model for maternal health risk classification in pregnancy. In: Proceedings of the 2023 IEEE 12th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE. 2023.
32. Badr S, Tahri M, Maanan M, Kašpar J, Yousfi N. An intelligent decision-making system for embryo transfer in reproductive technology: a machine learning-based approach. Syst Biol Reprod Med. 2025;71(1):52-65.
33. Chouhan R, Singh SK, Thakur P. Polycystic Ovary Syndrome (PCOS) diagnosis using machine learning techniques: A comparative study. In: Proceedings of the 2023 3rd International Conference on Emerging Trends in Information Technology and Engineering (ICETITE). IEEE. 2023.
34. Bibi S, Javed U, Mahmood A, Khan Z, Khan Z, Zafar S, et al. An ensemble deep learning model for polycystic ovary syndrome (PCOS) detection. J Healthc Eng. 2022;2022:7472016.
35. Yousuf A, Khan H, Al-Siyabi T, Al-Rasbi M. Deep learning for diagnosis of chronic endometritis using transvaginal ultrasound images. In: Proceedings of the 2023 International Conference on Computer Science and Information Technology (CSIT). IEEE. 2023.
36. Al-Maskari MY, Al-Rasbi M, Al-Sabri M, Al-Mahrouqi H, Al-Ghafri M, Al-Adawi S, et al.

Machine Learning for predicting maternal mortality: A nationwide study from Oman. Sci Rep. 2023;13(1):17562.
37. Balaraman V. Predictive modeling of child development using maternal risk factors: A machine learning approach. Int J Comput Appl. 2016;139(7):1-5.
38. Bowe J. Predicting child IQ from maternal factors: A machine learning approach. J Child Psychol Psychiatry. 2022;63(4):423-35.
39. Yang Z. Machine learning for neurodevelopmental delay prediction in children based on maternal risk factors. J Child Neurol. 2024;39(1):1-8.
40. Rao R, Mishra S, Singh P, Kumar A. A machine learning approach to predict gestational diabetes using clinical data. In: Proceedings of the 2023 IEEE International Conference on Data Science and Computing (ICDSC). IEEE. 2023.
41. Al-Khazraji A, Al-Jabery A, Al-Ameri S. ConvXGB: A novel predictive model for maternal outcome prediction integrating XGBoost with deep learning. In: Proceedings of the 2023 International Conference on Computer Science and Information Technology (CSIT). IEEE. 2023.
42. Mapari M, Patil U, Khose K. Machine Learning-Based Box Models for Pregnancy Care and Maternal Mortality Reduction: A Literature Survey. In: Proceedings of the 2023 International Conference on Computer Science, Engineering and Applications (ICCSEA). IEEE. 2023.
43. Ying H, Chen Y, Wang M, Chen X, Ye S. Machine learning models for preeclampsia prediction: a systematic review. J Matern Fetal Neonatal Med. 2023;36(1):2191563.
44. Kim J, Cho H, Kim T, Cho Y, Lee M, Lee S, et al. Development of a machine learning model for predicting mode of delivery based on clinical and ultrasound features. Sci Rep. 2023;13(1):16448.
45. Moreira L, de Carvalho JL, Dinis-Ribeiro M, Pereira A, Cruz M, Rocha M. Predicting spontaneous preterm delivery using machine learning: a systematic review and meta-analysis. BJOG. 2018;125(9):1075-83.
46. Adeoye S, Akinbode A. Machine Learning Approaches for Maternal Health Risk Classification: A Comparative Study. Int J Innov EngSci Res. 2025;9(1):45-56.
47. Campion R, Bisson M, Boily C, Desrosiers M, Gagnon N, Gagnon E, et al. Prediction of cumulative probability of pregnancy over 12 menstrual cycles using regularized logistic regression. Fertil Steril. 2020;114(4):780-7.
48. Tarca AL, Cha EMS, Kim M, Erez O, Mazor M, Bhat G, et al. Machine learning prediction of preterm birth from electrohysterography and clinical data: a systematic review. Ultrasound Obstet Gynecol. 2021;58(5):668-79.
49. Mohammadi Far K, Bahar M, Farzad A, Bahmani B. Preterm birth prediction using machine learning algorithms based on electrohysterogram (EHG) signals: A systematic review and meta-analysis. Comput Methods Programs Biomed. 2022;225:107062.
50. Hu Q, Jiang J, Zhao R, Gao J. A comprehensive review of machine learning in preterm birth prediction. Int J Med Inform. 2020;143:104273.
51. Li D, Yu Z, Ma H, Guo W, Li K. Multi-Task Link-Net: A multi-task deep convolutional neural network for automatic segmentation and estimation of fetal head circumference ellipse from ultrasound images. Comput Methods Programs Biomed. 2023;231:107412.
52. Zhao R, Li Y, Guo W, Ma H, Li K. Automated classification of common maternal-fetal ultrasound planes using deep convolutional neural networks. Comput Methods Programs Biomed. 2022;225:107106.
53. Ghods D, Kadam M. Machine learning approaches for predicting maternal health outcomes: A systematic review. Health Informatics J. 2021;27(4):1460-73.
54. Liu W, Zhang C, Cui J, Wu X, Sun H. Deep learning for embryo implantation prediction: A multi-modal approach combining embryo images and clinical data. In: Proceedings of the 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE. 2022.
55. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
56. Charnpinyo C, Uematsu H, Choe J, Kim JW, Park HJ, Choi DH, et al. A hybrid deep learning model for embryo implantation prediction by integrating image and tabular data. Sci Rep. 2023;13(1):15389.
57. Kim S, Choi H, Lee Y, Kang S, Kim D, Lee H. Interpretable deep learning model for embryo viability prediction combining image segmentation and clinical data. J Assist Reprod Genet. 2024;41(1):215-26.
58. Ouyang Z, Yu Q, Li Z, Zhou C, Liu R, Zhang C, et al. Transformer and DeFusion for early embryo development prediction in IVF. In: Proceedings of the 2025 IEEE International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). IEEE. 2025.
59. Zou X, Sun M, Wang Y, Liang Y. A review on hybrid AI systems and their applications. J Parallel DistribComput. 2021;156:1-13.

Cite this article as: Isogun JK, Odunvbun WO, Ukoba O, Amos E, Okoacha I, Ukoba JO, et al. Revolutionizing reproductive health: a performance review of machine learning algorithms in clinical infertility and maternal care. Int J Community Med Public Health 2025;12:5374-83.