HJOG An Obstetrics and Gynecology International Journal

Review

HJOG 2026, 25 (2), 176-189. | DOI: 10.33574/HJOG.0626

Artificial Intelligence in Reproductive Medicine: Transforming Diagnosis, Treatment, and IVF Outcomes

Santosh Ramesh Achwani[1], Faique Tanvir Falke[2], C. Anchana Devi[3], C. Oviya[4], Harish Narayanan[5], Rajkumar Krishnan Vasanthi[6], Shamina S[7], Sovan Bagchi[8]

[1]Department of Family Medicine, Al Bateen Medical Center, Abu Dhabi Health Services Company (SEHA), UAE.
[2]Family Medicine Specialist, Burjeel Medical City, Abu Dhabi, UAE.
[3]PG & Research Department of Biotechnology, Women's Christian College, Chennai, Tamil Nadu, India.
[4]Annai Medical College, Pennalur, Sri perambudur, Chennai, Tamil Nadu, India.
[5]Department of Community Medicine, Sree Balaji Medical College and Hospital, Bharat Institute of Higher Education and Research, Chennai.
[6]Faculty of Health and Life Sciences, INTI International University, Nilai, Negeri Sembilan, Malaysia.
[7]Department of Biochemistry, RVS College of Arts and Science, Sulur, Coimbatore, Tamil Nadu, India.
[8]Department of Biomedical Sciences, College of Medicine, Gulf Medical University, UAE.

Correspondence
Dr. Sovan Bagchi, Professor, Department of Biomedical Sciences, College of Medicine, Gulf Medical University, UAE. Orcid id-0000-0003-3507-1944, e-mail: bagchisovan60@gmail.com

Abstract

Artificial Intelligence (AI) is rapidly transforming assisted reproductive technologies (ART) by enhancing diagnostic accuracy, treatment personalization, and clinical decision-making. AI-driven tools can predict embryo viability, optimize sperm selection in real time, and tailor ovarian stimulation protocols to individual patient characteristics. Emerging applications, such as convolutional neural networks for embryo imaging and machine learning models for sperm analysis, reduce subjectivity and improve the reproducibility of outcomes. Reinforcement learning and computer vision are also advancing stimulation regimens and laboratory automation, while robotic-assisted interventions benefit from greater precision. Despite these innovations, important challenges remain. Clinical adoption is limited by heterogeneity of training datasets, algorithmic bias, and a lack of transparency in model design. Ethical concerns—including data privacy, informed consent, and equitable access and potential health risks—further complicate integration into fertility care. Moreover, financial and infrastructural barriers restrict widespread implementation, particularly in resource-limited settings. This review highlights current computational approaches, evaluates their contributions to embryo and gamete assessment, ovarian stimulation, and surgical procedures, and discusses unresolved issues of validation and regulation. Ultimately, AI holds the potential to make ART more efficient,

precise, and patient-centred. Still, it's safe and equitable integration will require interdisciplinary collaboration, robust clinical validation, and adherence to ethical and regulatory standards.

Keywords: AI-based fertility prediction, AI in reproductive medicine, embryo and sperm selection using AI, machine learning in Assisted Reproductive Technology (ART), ovarian stimulation and AI optimization.

Introduction

Infertility affects an estimated 15% of couples globally, posing both medical and psychosocial challenges. It is commonly defined as the inability to achieve conception after one year of regular, unprotected intercourse [1]. Male factors, often related to sperm dysfunction, contribute to nearly half of infertility cases, while female infertility is frequently associated with ovulatory disorders, tubal damage, or diminished ovarian reserve [2]. Conventional treatments—ranging from ovulation induction and intrauterine insemination (IUI) to in vitro fertilization (IVF)—have improved outcomes but remain limited by high costs, variable success rates, and the inherent subjectivity of clinical decision-making. The growing volume of multimodal data in reproductive medicine—including electronic health records, hormone profiles, medical imaging, genomics, and time-lapse embryology—has created opportunities for computational approaches. Traditional statistical models, while valuable, often fail to capture the complex nonlinear relationships among clinical variables. AI, encompassing machine learning (ML), deep learning (DL), and NLP, offers a paradigm shift by enabling predictive modeling and automated decision support [3]. Early studies have demonstrated the potential of AI to enhance ART outcomes. Convolutional neural networks (CNNs) have been applied to embryo imaging, achieving higher predictive accuracy than manual assessment [4]. Supervised ML models such as random forests, support vector machines (SVM), and artificial neural networks (ANNs) have been employed to predict IVF success, sperm morphology, and fertilization potential [5]. Reinforcement learning techniques are under investigation for optimizing ovarian stimulation and surgical planning. However, challenges persist, including limited clinical validation, algorithmic bias, interpretability issues, and the need for standardized regulatory frameworks. This review, therefore, examines computational methods, clinical applications, and ethical implications of AI in ART, with an emphasis on nuanced challenges and algorithmic innovations that shape its translation into real-world practice.

Challenges in reproductive medicine

Assisted reproductive technology (ART) has progressed remarkably since the birth of the first IVF child, yet significant challenges remain. Infertility affects over 187 million individuals worldwide, and despite advances in cryopreservation, genetic testing, and embryo culture systems, IVF success rates remain limited at 20–40% per cycle. These outcomes are compounded by high financial costs, prolonged treatment cycles, and unequal access to fertility services, creating emotional and physical burdens for patients [6]. The COVID-19 pandemic further magnified these challenges by delaying treatments and increasing psychological stress. A key barrier lies in embryo selection, where conventional morphology-based methods remain subjective and variable. Similarly, sperm analysis suffers from observer-dependent variability, while ovarian stimulation protocols often result in inconsistent responses and risks such as ovarian hyperstimulation syndrome (OHSS). These persistent limitations underscore the need for approaches that improve precision, reproducibility, and personalization in ART. Here, AI offers direct relevance. CNNs applied to time-lapse embryo imaging show promise in improving embryo viability prediction. AI-based sperm analysis platforms provide standardized and objective assessments, reducing bias in motility and morphology evaluation. Machine learning algorithms can also predict ovarian response, enabling individualized gonadotropin dosing to optimize stimulation while minimizing risks [7]. Thus, the very challenges that continue to limit reproductive medicine—subjectivity, variability, high cost, and inequity—are the same areas where AI technologies have begun to demonstrate transformative potential. Integrating these innovations responsibly may help bridge the gap between current limitations and more effective, equitable fertility care. Building on these advances, the application of AI is not confined to ovarian stimulation alone. Its integration with broader reproductive medicine has expanded into areas such as surgical precision, patient stratification, and large-scale data analysis. This progression naturally leads into the discussion of how diverse AI modalities are shaping the wider landscape of reproductive healthcare.

AI in RM

Extending beyond ovarian response prediction, AI's role in reproductive medicine encompasses a much wider spectrum of innovations. These include not only laboratory and diagnostic applications but also surgical assistance, patient monitoring, and knowledge extraction from vast clinical datasets. Reproductive medicine (RM) increasingly incorporates advanced technologies such as robotic surgery, ML, and natural language processing (NLP). ML primarily stratifies patients into clinically meaningful groups using structured data—such as genetic profiles and medical imaging—while NLP complements this by extracting insights from unstructured records, including electronic medical records (EMRs) [8–10]. However, successful implementation requires large, high-quality datasets; inadequate or heterogeneous data compromise predictive reliability and clinical translation [11]. For example, ML models predicting IVF outcomes have achieved accuracies of 59–68%, though these results remain difficult to generalize across patient populations, laboratory protocols, and reporting standards [12]. The studies faced limitations, including small, limited datasets, which restricted model training and accuracy, and the challenge of capturing all relevant prognostic factors, reflected in the maximum predictive accuracy of around 60%. Validation was primarily internal, using random data splits and cross-validation, but external validation with independent datasets was not reported, limiting assessments of generalizability. Consequently, the models' applicability across different populations and clinical settings remains uncertain, highlighting the need for larger, more diverse datasets and external validation to improve robustness and reliability for broader clinical use. Within RM, supervised learning has shown particular promise. CNNs enable advanced embryo image analysis, identifying subtle morphological cues often overlooked in conventional scoring. Similarly, algorithms such as SVM, random forests, and decision trees have been employed to forecast ART outcomes, offering improved predictive power compared to traditional clinical scoring [13]. These methods face persistent challenges, including "black box" interpretability, the risk of algorithmic bias, and limited reproducibility across clinical settings. Unsupervised learning approaches, though less commonly applied, are beginning to uncover latent infertility phenotypes and patient subgroups, suggesting opportunities for personalized treatment strategies. Reinforcement learning also holds theoretical potential in optimizing stimulation protocols, but its clinical use remains nascent. While AI applications in RM are rapidly advancing, their effective integration demands standardized data frameworks, greater model transparency, and interdisciplinary collaboration to balance innovation with clinical accountability [13]. Among these applications, one of the most clinically impactful areas where AI is beginning to demonstrate tangible value is in ART, particularly in predicting treatment outcomes and guiding personalized interventions.

AI in Female Infertility

Based on these methodological advances, the application of AI in IVF outcome prediction provides a clear example of how data-driven tools can move from theoretical potential to measurable clinical relevance. The probability of achieving pregnancy following IVF ranges widely (30–70%), influenced by maternal age, ovarian reserve, and treatment protocols. While conventional diagnostic workups—encompassing endocrinological profiling, imaging, and invasive assessments—remain the standard, recent advances in AI and ML offer opportunities to enhance decision-making in ART. Logistic regression, SVMs, decision trees, and random forests have been employed to predict outcomes using diverse clinical and biological parameters such as age, BMI, endometrial thickness, hormonal levels, and embryo quality [14]. However, a key limitation lies in model generalizability, as performance often declines when applied across heterogeneous populations. Embryo selection represents the most active frontier for AI integration. Traditional morphological grading (e.g., Gardner system, ASEBIR guidelines) is subjective and prone to inter-observer variability [15].

AI-driven morphokinetic analysis of time-lapse imaging attempts to overcome these limitations by detecting subtle developmental dynamics invisible to the human eye. Debates persist regarding the overfitting of models to specific datasets and the ethical implications of replacing embryologists with automated decision systems. The IMMATCH approach, combining HLA typing and peptide–MHC data, illustrates how AI may also reveal immunogenetic determinants of recurrent pregnancy loss (AUC 0.71, $p=0.0035$) [16]. These studies faced limitations such as small sample sizes and high HLA polymorphism, which complicate precise risk prediction. Validation was performed using leave-one-out cross-validation, ensuring internal validation within the dataset but lacking external validation. As a result, the models' generalizability to broader, diverse populations remains uncertain. Larger, independent cohorts are needed to confirm reliability and clinical applicability, highlighting the importance of external validation to assess how well the predictive models perform in different populations and real-world settings. Translation into routine practice requires rigorous validation in diverse cohorts. The widespread adoption of single embryo transfer (SET) policies further underscores the need for precise embryo prioritization. AI-assisted selection, when integrated with SET, shows promise in reducing multiple pregnancies while optimizing live birth rates [17]. However, unresolved challenges—including model transparency, reproducibility, and cost-effectiveness—remain barriers to clinical adoption. Future work must focus not only on algorithmic accuracy but also on harmonizing AI outputs with ethical, regulatory, and patient-centered considerations to ensure meaningful integration into fertility care. These challenges in model validation are not unique to immunogenetic predictions but extend across reproductive medicine, where small, homogeneous datasets often hinder algorithmic performance. A particularly relevant example is seen in male infertility, where variability in semen parameters further complicates the development of reliable AI-based diagnostic and predictive tools.

AI in Male Infertility

Building on these broader concerns of validation and generalizability, male infertility offers a representative case study. Here, AI has been increasingly explored to overcome the limitations of conventional diagnostic approaches, particularly in semen analysis and sperm selection for ART. Routine semen analysis remains the cornerstone of male infertility evaluation, yet its accuracy is constrained by inter-observer variability and biological heterogeneity. Computer-aided sperm analysis (CASA) offers standardization, but its clinical translation is limited by poor reproducibility in morphology assessment and the need for costly, specialized equipment [18]. ML and DL models have expanded beyond CASA, achieving >90% accuracy in predicting abnormal sperm concentration using demographic and lifestyle data; however, concerns remain about dataset bias, small cohort sizes, and limited generalizability [19]. CNNs, trained on video-based datasets such as VISEM, have shown promise in characterizing motility and detecting subtle morphological abnormalities. However, the interpretability of CNNs is debated, as "black-box" outputs challenge clinical trust and regulatory approval [20]. AI has also advanced sperm selection in intracytoplasmic sperm injection (ICSI). Systems capable of detecting viable non-motile sperm via nuclear integrity assessment offer non-invasive alternatives to staining, though questions persist regarding long-term safety and reproducibility [21]. Moreover, predictive clustering algorithms linking DNA fragmentation with IVF outcomes highlight AI's ability to uncover hidden biological patterns. Still, controversy exists over whether DNA fragmentation itself is a causative factor or a correlational marker of infertility [22]. The study's limitations include potential biases from manual training set classifications and reliance on CASA parameters, which may vary across different settings. Internal validation was conducted using repeated iterations and performance metrics like accuracy and precision, but there is limited mention of external validation on independent samples. Consequently, generalizability may be constrained by specific sample populations and CASA configurations. Further validation with diverse datasets is necessary to confirm the robustness and applicability of the CASAnova models across different laboratories and patient populations. Emerging algorithms integrate endocrine parameters, genetic data, and environmental exposures, achieving >95% predictive accuracy in experimental cohorts. However, challenges of overfitting, limited multicentre validation, and ethical considerations around data privacy remain unresolved. Moving forward, clinically meaningful AI in male infertility will require harmonized datasets, transparent algorithms, and prospective validation trials to shift from proof-of-concept tools toward trusted clinical decision support. While AI applications in male infertility highlight the potential of data-driven personalization, similar challenges extend to female reproductive medicine, particularly in optimizing ovarian stimulation protocols. The transition from sperm selection to ovarian response prediction underscores the broader role of AI in addressing variability and improving treatment safety across both male and female infertility domains.

AI in Ovarian Stimulation

On insights from AI-driven male infertility models, researchers have begun to leverage comparable predictive algorithms to tailor ovarian stimulation strategies, integrating hormonal, genetic, and clinical data for individualized care. This continuity highlights how AI, beyond sperm evaluation, can be harnessed to navigate the complexity of ovarian physiology, where heterogeneity poses similar diagnostic and therapeutic challenges. Ovarian stimulation is central to ART, yet its optimization remains challenging due to patient heterogeneity and the inherent unpredictability of ovarian response. Traditional gonadotropin regimens often expose women to supraphysiological doses of follicle-stimulating hormone (FSH), prolonging the “FSH window” and increasing the risk of OHSS. Recent ML approaches have advanced individualized dosing strategies by incorporating clinical parameters such as body weight, antral follicle count, and anti-Müllerian hormone (AMH) [23]. Dose–response prediction models, including gradient boosting and random forest algorithms, demonstrated that lower FSH dosing can achieve higher-quality blastocysts and reduced treatment costs. However, external validation across diverse populations is still limited. Controversy persists regarding the clinical utility of algorithm-based personalization: while some studies report improved live birth rates, others note only marginal benefits compared to conventional fixed protocols. Integration into clinical decision support systems (CDSS) offers dynamic, real-time adjustments, but challenges include algorithm transparency, clinician trust, and risk of overfitting in smaller datasets. AI-driven oocyte and embryo assessment tools highlight further innovation. CNN-based systems such as VIOLET™ and MAGENTA™ outperform embryologists in predicting blastulation and live birth potential, yet reproducibility across clinics remains uncertain [24]. Similarly, OsteraTest™, which applies transcriptomic profiling integrated via multiple ML models, achieves 86% accuracy for day-5 blastocyst prediction [25]. Despite promising results, issues of data standardization, bias from uneven training datasets, and the ethical implications of automated embryo selection demand scrutiny. Overall, AI holds transformative potential in tailoring ovarian stimulation and embryo selection, but its clinical adoption hinges on transparent validation, cross-population reproducibility, and integration within ethical, patient-centered frameworks. Building on embryo-focused applications, AI has also been increasingly employed in imaging-based assessments of ovarian and endometrial physiology. These approaches provide a bridge between predictive embryo models and non-invasive reproductive diagnostics, highlighting AI’s broader role in improving clinical decision-making across the entire IVF pipeline.

AI in Gamete Assessment

AI has shown particular promise in refining imaging modalities central to female reproductive assessment. By extending predictive modeling into ultrasound-based evaluation, AI enables earlier, more standardized, and less subjective insights into ovarian reserve and endometrial receptivity. Ultrasound (US) remains a cornerstone in assessing ovarian reserve and endometrial receptivity, with indices such as follicle number, stromal blood flow, endometrial thickness, and curvature serving as predictors of implantation success and pregnancy. However, inter-observer variability has long been a limitation in conventional imaging. AI-based image analysis, particularly CNNs such as VGGNet-16, has demonstrated improved reproducibility in classifying endometrial lesions, reducing subjectivity in interpretation [26]. Despite these advances, challenges persist regarding dataset size, image quality, and generalizability across diverse populations. In hysteroscopic image analysis, AI-driven diagnostic support has shown promise, though regulatory and

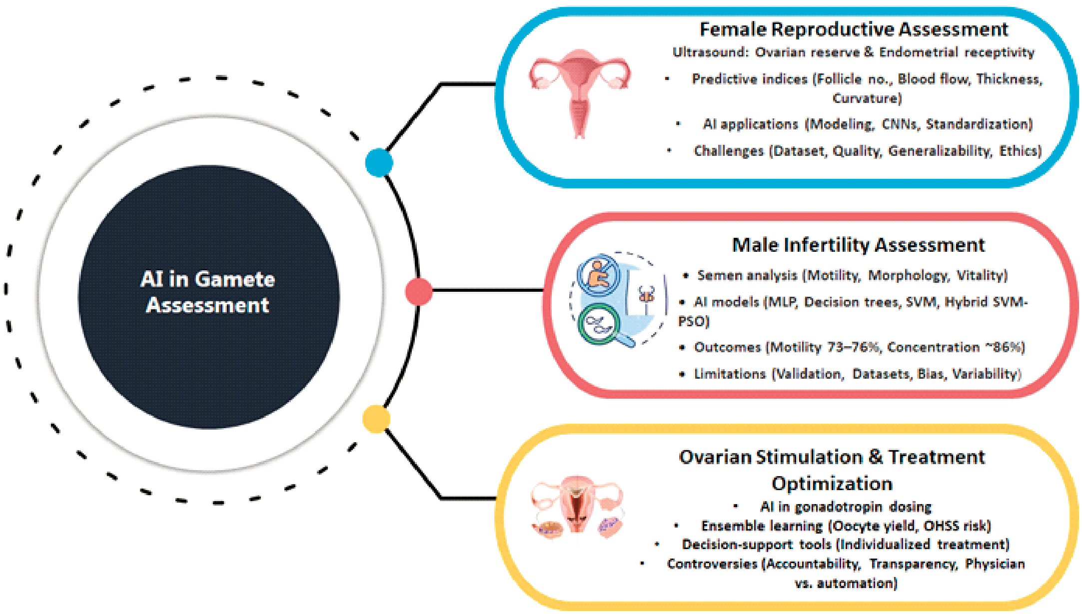

Figure 1. AI in Gamete Assessment.

ethical considerations limit real-world clinical integration. Male infertility, accounting for nearly half of cases, presents additional complexity. Traditional semen analysis—focusing on sperm motility, morphology, and vitality—provides limited predictive power. Recent AI frameworks, including multilayer perceptron, decision trees, SVM, and hybrid SVM–particle swarm optimization models, have achieved motility prediction accuracies of 73–76% and sperm concentration prediction of ~86% [27]. The studies primarily used internal validation through training and testing on subsets of their collected semen samples, limiting external validation and generalizability. They did not explicitly mention cross-validation with independent datasets, which could affect the robustness of the models. Consequently, the models' applicability to broader populations remains uncertain, as factors like demographic variability and different laboratory protocols were not addressed. The limited sample size and lack of external validation constrain the findings' generalizability, highlighting the need for further studies with diverse, larger datasets and external validation to confirm the models' effectiveness in varied clinical settings. Algorithmic bias arising from heterogeneous sample preparation and the underrepresentation of certain populations remains a concern. Advances in integrating biochemical markers (fructose, β-glucosidase, zinc) and lifestyle parameters into ANN-based models highlight the potential of multimodal approaches, though reproducibility across laboratories requires further validation. In ovarian stimulation, AI models now assist in optimizing gonadotropin dosing and trigger day selection, with ensemble learning strategies improving prediction of oocyte yield and ovarian hyperstimulation syndrome (OHSS) risk [28]. While these decision-support tools promise individualized treatment, controversies remain regarding clinical accountability, transparency of algorithmic decisions, and balancing automation with physician expertise. Moving forward, interpretability-focused AI and federated learning frameworks may address these gaps, fostering robust, equitable integration of AI into reproductive medicine. Beyond diagnostic decision-support, AI integration has also extended into interventional domains of reproductive medicine, where precision, consistency, and real-time adaptability are critical. One of the most visible areas of this shift is in surgical practice, where AI-enabled technologies complement minimally invasive and robotic-assisted procedures (Figure 1).

Robots and AI in surgery

AI-assisted decision-making in surgical innovations has increasingly embraced digital technologies to improve procedural outcomes. In this context, minimally invasive surgery (MIS) represents not only a technological leap in gynecology but also a platform where AI-driven analytics and robotic systems converge to enhance safety and precision. MIS has transformed gynecological practice by reducing morbidity, enhancing recovery, and improving cosmetic outcomes through high-resolution imaging and micro-instrumentation (Diana & Marescaux, 2015) [29]. The subsequent evolution—robotic-assisted surgery—has addressed several limitations of laparoscopy, including surgeon ergonomics, dexterity, and three-dimensional visualization (Lonnerfors, 2018) [30]. In gynaecology, robotic myomectomy has shown advantages such as reduced blood loss, shorter hospitalization, and faster convalescence [31,32]. However, the evidence remains inconsistent, with some studies reporting prolonged operative times and equivalent complication rates compared to conventional laparoscopy [33], raising debates regarding its cost-effectiveness and true incremental benefit. Beyond surgical ergonomics, the integration of AI-driven algorithms into robotic systems marks a paradigm shift. Machine vision and haptic feedback systems are under development to assist in differentiating tissue planes, predicting blood loss, and optimizing suture placement. Early adaptive algorithms have been piloted for automating certain repetitive steps, such as knot tying, which could reduce operator fatigue and improve standardization [34]. Controversies persist regarding the ethical implications of automation in procedures with direct fertility outcomes, such as adenomyomectomy and deep-infiltrating endometriosis resections, where long-term reproductive benefits remain poorly quantified. Recent innovations such as robotic laparoendoscopic single-site (LESS) surgery aim to minimize scarring. However, technical challenges—including instrument crowding, reduced triangulation, and limited tactile feedback—remain significant barriers (Park et al., 2019) [35]. Addressing these will require synergistic advances in AI-assisted navigation, real-time imaging analysis, and machine learning algorithms capable of intraoperative decision support. Thus, while robotics and AI hold transformative potential, current applications reflect a delicate balance between technological promise, clinical evidence, and unresolved controversies. The scope of AI in reproductive medicine extends into the diagnostic and prognostic domains. These advances are particularly relevant when addressing complex clinical scenarios where traditional evaluation provides limited guidance, such as unexplained or idiopathic infertility.

AI in Idiopathic Reproductive Disorders

Extending from the challenges encountered in surgical and intraoperative settings, the application of AI in infertility assessment underscores its broader relevance across the reproductive domain. This is particularly significant in cases of idiopathic infertility, where conventional diagnostic approaches frequently lack conclusive outcomes, thereby necessitating data-driven methodologies for more precise patient stratification. Idiopathic infertility, where no definitive medical cause is identifiable, remains one of the most perplexing challenges in reproductive medicine. While lifestyle and metabolic risk factors such as obesity and metabolic syndrome are known to impair outcomes in ART, predicting treatment success in idiopathic cases is particularly difficult. Recent efforts have explored AI-based models capable of stratifying patients by fertility potential. For example, a machine learning framework combining principal component analysis (PCA) and orthogonal partial least squares–discriminant analysis (OPLS-DA) was shown to differentiate between fertile and infertile couples using bioclinical characteristics [36]. Supplementary classifiers—including logistic regression, decision trees, SVMs, and k-nearest neighbours—were benchmarked in Python-based pipelines, highlighting variability in model interpretability and robustness. However, controversies persist. While SVMs and ensemble methods often outperform simpler classifiers, they are criticized for limited interpretability—a major obstacle in clinical adoption where transparency is essential [37]. Similarly, algorithmic performance is dataset-dependent; models trained on small or homogeneous cohorts risk poor generalizability across diverse populations. Another unresolved challenge is the absence of randomized controlled trials (RCTs) validating AI-guided decision-making in idiopathic infertility, raising questions about clinical readiness, beyond technical considerations, ethical and regulatory dimensions complicate integration. European regulatory frameworks emphasize explainable and human-centered AI, aligning with WHO guidance on safeguarding patient autonomy, informed consent, and privacy [38]. Striking a balance between maximizing predictive accuracy and ensuring algorithmic fairness remains a key debate. Future research must therefore not only refine predictive modelling but also address interoperability, transparency, and patient trust—ensuring AI augments, rather than replaces, reproductive decision-making. This need for fairness and transparency becomes even more pressing when evaluating AI tools that are being adopted in real-world clinical settings. The type of platform—whether developed internally within clinics or delivered as commercial solutions—directly influences the degree of oversight, adaptability, and trust that stakeholders can place in these systems.

Computational Methods in AI

In light of these considerations, the integration of AI into assisted reproduction is influenced not only by the accuracy of the algorithms but also by the technological delivery models through which they are deployed. The contrast between proprietary in-house systems and commercial platforms underscores broader discussions surrounding control, adaptability, and accountability in clinical implementation. IVF clinics increasingly rely on two broad categories of AI systems: proprietary in-house solutions and commercially available platforms. While smaller centers often prefer cloud-based technologies for their scalability, this introduces concerns regarding data privacy, interoperability, and standardization across heterogeneous clinical practices [39]. Algorithmic performance is highly contingent on the training dataset: CNNs have shown promise in embryo image analysis, yet their “black-box” nature raises concerns about interpretability and clinician trust [40]. Explainable AI (XAI) methods—such as saliency mapping and attention mechanisms are now being explored to address this issue, though they remain imperfect. Similarly, reinforcement learning has been applied in optimizing ovarian stimulation protocols, but questions remain about its reproducibility across populations with variable clinical characteristics [41]. Beyond performance metrics (e.g., AUC, MCC, F1-score), ethical and regulatory debates continue regarding algorithmic bias, particularly when maternal age or ethnicity disproportionately influence predictions [42]. Automated annotation pipelines may reduce subjectivity, yet their stability remains vulnerable to noise in imaging and temporal data streams. Integration of multimodal inputs—combining tabular records, imaging, genomics, and time-lapse videos—marks a major advancement, but poses validation challenges across diverse populations. Importantly, cost and infrastructure remain significant barriers, particularly in low-resource settings, where AI could otherwise provide the greatest benefit. Overall, AI in reproductive medicine must move beyond proof-of-concept accuracy claims toward robust, explainable, and ethically governed clinical tools. This requires longitudinal validation, transparent reporting of algorithm design, and stakeholder collaboration to ensure reliable, equitable, and scalable integration into IVF workflows. An equally important consideration is whether these AI models can sustain their reported performance when applied beyond controlled research settings. This naturally leads to the broader challenge of model generalizability and adaptability in diverse clinical contexts.

Ethical Considerations

Building on this concern, the issue of generalizability becomes central to evaluating the true clinical utility of AI systems in reproductive medicine. A critical question that arises concerns the generalizability of these models—specifically, how effectively they can apply to novel datasets not represented in earlier repositories. Moreover, the mere application of sophisticated AI algorithms does not guarantee optimal solutions, as standardized guidelines for adjusting algorithmic parameters remain limited. Within the domain of medical ethics, patient safety must remain the paramount consideration in the development of AI. Accordingly, AI systems should be designed to demonstrate robustness, transparency, verifiability, and reliability, ensuring both clinical utility and ethical integrity. ML algorithms necessitate extensive datasets, yet regulatory frameworks governing their use exhibit significant variability across countries and regions, particularly regarding the protection of patient information. At present, the evidence base supporting the application of AI in RM remains limited, with a scarcity of randomized controlled trials. Emerging consensus suggests that AI models intended for RM must prioritize interpretability and rigorous oversight to ensure safe and effective clinical deployment [43].

Furthermore, technical challenges persist, including the lack of standardization and interoperability among medical information technology systems. Addressing these barriers is essential to accelerate the responsible integration and development of AI-driven solutions in European healthcare. The overarching objective is to position Europe as a global leader in the development and deployment of ethical, human-centered AI systems—specifically, technologies designed to augment and enhance human capabilities rather than supplant them. The WHO report provides recommendations grounded in fundamental ethical principles, emphasizing that individuals must retain authority over their personal healthcare decisions and systems. Central to these recommendations is the safeguarding of privacy and confidentiality, alongside the requirement that patients provide genuine informed consent consistent with rigorous data protection standards. Furthermore, it is imperative to examine the sociocultural and anthropological factors that influence human interaction with and trust in AI technologies, thereby fostering systems that are congruent with human behavior and societal norms.

Future Perspectives

The advancement of AI has been significantly accelerated by the availability of large, publicly accessible datasets, which provide high-quality training data for machine learning models. AI enables clinicians to collect, process, and analyze complex datasets in previously unattainable ways, thereby enhancing decision-making in ART, including the selection of optimal embryos and sperm. The integration of big data analytics offers the potential to derive robust, evidence-based insights from complex biological patterns. As these databases continue to expand and evolve, reproductive medicine is poised to experience substantial progress. Comprehensive data analysis facilitates the extraction of actionable knowledge, and the performance of AI systems can be further augmented when combined with complementary computational approaches, such as machine learning, provided that high-quality data collection, integration, and interpretation are ensured. The future of AI in reproductive medicine depends on moving from conceptual promise to clinically testable hypotheses supported by rigorous validation. For embryo selection, CNN models trained on standardized, multi-centre datasets may outperform embryologists in predicting implantation potential, which can be tested through randomized controlled trials comparing AI-assisted versus conventional selection. In male infertility, hybrid AI models (e.g., SVM with deep learning) could achieve higher accuracy in assessing sperm motility and morphology, requiring prospective validation across diverse laboratories and demographics. For ovarian stimulation, reinforcement learning–based dosing algorithms are hypothesized to reduce OHSS incidence without compromising oocyte yield, a claim testable through clinical trials of AI-guided versus physician-guided dosing. Idiopathic infertility may be approach by integrative big data combining genomics, imaging, and clinical records to identify novel biomarkers, which can be validated in longitudinal cohort studies with AI-driven feature extraction. In reproductive surgery, AI-augmented robotic systems are predicted to shorten operative time and reduce complication rates, a hypothesis suitable for comparative effectiveness studies against current robotic techniques. Collectively, these testable directions underscore the importance of building large-scale harmonized datasets, applying NLP and DL for knowledge extraction, and embedding AI-driven decision-support systems into clinical workflows in a safe, reproducible, and equitable manner.

Conclusion

AI is redefining the landscape of reproductive medicine by offering tools that enhance embryo selection, personalize ovarian stimulation, and improve surgical precision. Its integration into ART promises to reduce subjectivity, accelerate diagnosis, and deliver more individualized treatment strategies. However, the transition from experimental applications to clinical practice is far from straightforward. A central challenge lies in ensuring generalizability and fairness. Models trained on homogeneous datasets risk perpetuating bias when applied across diverse populations. The demand for explainable AI underscores the need to bridge algorithmic performance with clinical interpretability, thereby strengthening clinician and patient trust.

Furthermore, financial and infrastructural barriers may restrict adoption in low-resource settings, amplifying existing disparities in access to fertility care. Looking forward, research should move beyond proof-of-concept studies to large-scale, longitudinal validations across varied demographics. Priorities include developing standardized evaluation metrics, refining multimodal integration approaches, and establishing clear regulatory and ethical frameworks. Interdisciplinary collaboration—uniting clinicians, data scientists, ethicists, and policymakers—will be essential to ensuring that AI systems are not only technically robust but also ethically responsible and globally accessible. With careful governance, AI has the potential to transition from a supplementary tool into a core enabler of precision reproductive medicine.

Conflict of Interest

Conflict of interest declared none.

References

1. Agarwal A, Baskaran S, Parekh N, Cho CL, Henkel R, Vij S, Arafa M, Selvam MK, Shah R. Male infertility. Lancet. 2020, 397, 319–333.
2. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. PeerJ. 2019;7:e7702. doi:10.7717/peerj.7702
3. Wang R, Pan W, Jin L, Li Y, Geng Y, Gao C, Chen G, Wang H, Ma D, Liao S. Artificial intelligence in reproductive medicine. Reproduction. 2019; 158(4): R139-R154. doi:10.1530/REP-18-0523
4. Cosenza DN, Korhonen L, Maltamo M, Packalen P, Strunk JL, Næsset E, Gobakken T, Soares P, Tomé M. Comparison of linear regression, k-nearest neighbour and random forest methods in airborne laser-scanning-based prediction of growing stock. Forestry. 2021;94(2):311-323. doi:10.1093/forestry/cpaa034
5. Guerrero MC, Parada JS, Espitia HE. EEG signal analysis using classification techniques: Logistic regression, artificial neural networks, support vector machines, and convolutional neural networks. Heliyon. 2021;7(6):e07258. doi:10.1016/j.heliyon.2021.e07258
6. Niederberger C, Pellicer A, Cohen J, Gardner DK, Palermo GD, O'Neill CL, Chow S, Rosenwaks Z, Cobo A, Swain JE, et al. Forty years of IVF. Fertil Steril. 2018;110(2):185.e5-324.e5. doi:10.1016/j.fertnstert.2018.06.005
7. Danilov GV, Shifrin MA, Kotik KV, Ishankulov TA, Orlov YN, Kulikov AS, et al. Artificial intelligence in neurosurgery: A systematic review using topic modeling. Part I: Major research areas. Sovrem Tekhnologii Med. 2021;12:106-112. doi:10.17691/stm2020.12.5.12
8. Darcy AM, Louie AK, Roberts LW. Machine learning and the profession of medicine. JAMA. 2016;315(6):551-552.doi:10.1001/jama.2015.18421
9. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc. 2011;18(5):544-551. doi:10.1136/amiajnl-2011-000464
10. Mehta N, Devarakonda MV. Machine learning, natural language programming, and electronic health records: the next step in the artificial intelligence journey? J Allergy Clin Immunol. 2018;141(6):2019.e1-2021.e1. doi:10.1016/j.jaci.2018.02.025
11. Bastanlar Y, Ozuysal M. Introduction to machine learning. Methods Mol Biol. 2014;1107:105-128. doi:10.1007/978-1-62703-748-8_7
12. Kaufmann SJ, Eastaugh JL, Snowden S, Smye SW, Sharma V. The application of neural networks in predicting the outcome of in-vitro fertilization. Hum Reprod. 1997;12(7):1454-1457. doi:10.1093/humrep/12.7.1454
13. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science. 2015;349(6245):255-260. doi:10.1126/science.aaa8415
14. Goyal A, Kuchana M, Ayyagari KPR. Machine learning predicts live-birth occurrence before in-vitro fertilization treatment. Sci Rep. 2020;10:20925. doi:10.1038/s41598-020-76928-z
15. Bori L, Paya E, Alegre L, Viloria TA, Remohi JA, Naranjo V, Meseguer M. Novel and conventional embryo parameters as input data for artificial neural networks: An artificial intelligence model applied for prediction of the implantation potential. Fertil Steril. 2020;114(6):1232-

1241. doi:10.1016/j.fertnstert.2020.08.023

16. Mora-Sánchez A, Aguilar-Salvador DI, Nowak I. Towards a gamete matching platform: Using immunogenetics and artificial intelligence to predict recurrent miscarriage. NPJ Digit Med. 2019;2:12. doi:10.1038/s41746-019-0089-x
17. Raef B, Maleki M, Ferdousi R. Computational prediction of implantation outcome after embryo transfer. Health Inform J. 2019;26(3):1810-1826. doi:10.1177/1460458219892138
18. Shahzad S, Ilyas M, Lali MI, Rauf HT, Kadry S, Nasr EA. Sperm abnormality detection using sequential deep neural network. Mathematics. 2023;11(3):515. doi:10.3390/math11030515
19. Saiffe Farías AF, et al. Single-sperm motility analysis during ICSI using an artificial intelligence sperm identification software (SID) and correlation with morphology. Fertil Steril. 2022;118(Suppl 1):e56-e57. doi:10.1016/j.fertnstert.2022.08.178
20. Jiang A, Jiaqi W, Zhao H, Zhang Z, Sun Y. Identifying viability of immotile sperm at one glance: Sperm viability classifier powered by deep learning. Fertil Steril. 2022;118(Suppl 1):e297-e298. doi:10.1016/j.fertnstert.2022.08.828
21. Wu DJ, Badamjav O, Reddy VV, Eisenberg M, Behr B. A preliminary study of sperm identification in microdissection testicular sperm extraction samples with deep convolutional neural networks. Asian J Androl. 2021;23(2):135-139. doi:10.4103/aja.aja_66_20
22. Goodson SG, White S, Stevans AM, Bhat S, Kao CY, Jaworski S, Marlowe TR, Kohlmeier M, McMillan L, Zeisel SH, et al. CASAnova: a multiclass support vector machine model for the classification of human sperm motility patterns. Biol Reprod. 2017;97(5):698-708. doi:10.1093/biolre/iox120
23. Fanton M, et al. An interpretable machine learning model for individualized gonadotropin starting dose selection during ovarian stimulation. Reprod Biomed Online. 2022;45(5):1012-1022. doi:10.1016/j.rbmo.2022.07.010
24. Nayot D, Meriano J, Casper R, Alex K. An oocyte assessment tool using machine learning; predicting blastocyst development based on a single image of an oocyte. Hum Reprod. 2020; 35(Suppl 1):129-130.
25. Hanassab S, Abbara A, Yeung AC, Voliotis M, Tsaneva-Atanasova K, Kelsey TW, Trew GH, Nelson SM, Heinis T, Dhillo WS. The prospect of artificial intelligence to personalize assisted reproductive technology. NPJ Digit Med. 2024; 7(1):55. doi:10.1038/s41746-024-01006-x
26. Zhang Y, Wang Z, Zhang J, Wang C, Wang Y, Chen H, Shan L, Huo J, Gu J, Ma X. Deep learning model for classifying endometrial lesions. J Transl Med. 2021;19(1):10. doi:10.1186/s12967-020-02660-x
27. Vickram AS, Kamini AR, Das R, Pathy MR, Parameswari R, Archana K, Sridharan TB. Validation of artificial neural network models for predicting biochemical markers associated with male infertility. Syst Biol Reprod Med. 2016;62(4):258-265. doi:10.1080/19396368.2016.1185654
28. Robertson I, Chmiel F, Cheong Y. Streamlining follicular monitoring during controlled ovarian stimulation: a data-driven approach to efficient IVF care in the new era of social distancing. Hum Reprod. 2021;36(1):99-106. doi:10.1093/humrep/deaa251
29. Diana M, Marescaux J. Robotic surgery. Br J Surg. 2015;102(2):e15-e28. doi:10.1002/bjs.9711
30. Lonnerfors C. Robot-assisted myomectomy. Best Pract Res Clin Obstet Gynaecol. 2018;46:113-119. doi:10.1016/j.bpobgyn.2017.09.005
31. Barakat EE, Bedaiwy MA, Zimberg S, Nutter B, Nosseir M, Falcone T. Robotic-assisted, laparoscopic, and abdominal myomectomy: a comparison of surgical outcomes. Obstet Gynecol. 2011;117(2 Pt 1):256-265. doi:10.1097/AOG.0b013e318207854f

32. Advincula AP, Xu X, Goudeau S, Ransom SB. Robot-assisted laparoscopic myomectomy versus abdominal myomectomy: a comparison of short-term surgical outcomes and immediate costs. J Minim Invasive Gynecol. 2007;14(6):698-705. doi:10.1016/j.jmig.2007.06.008
33. Gargiulo AR, Srouji SS, Missmer SA, Correia KF, Vellinga TT, Einarsson JI. Robot-assisted laparoscopic myomectomy compared with standard laparoscopic myomectomy. Obstet Gynecol. 2012;120(2 Pt 1):284-291. doi:10.1097/AOG.0b013e3182602c7d
34. Caillet M, Vandromme J, Rozenberg S, Paesmans M, Germay O, Degueldre M. Robotically assisted laparoscopic microsurgical tubal reanastomosis: a retrospective study. Fertil Steril. 2010;94(5):1844-1847. doi:10.1016/j.fertnstert.2009.10.028
35. Scheib SA, Fader AN. Gynecologic robotic laparoendoscopic single-site surgery: prospective analysis of feasibility, safety, and technique. Am J Obstet Gynecol. 2015;212(2):179.e1-179.e8. doi:10.1016/j.ajog.2014.07.057
36. Bachelot G, Lévy R, Bachelot A, Faure C, Czernichow S, Dupont C, Lamazière A, Aknin I, Cedrin-Durnerin I, Cens S, et al. Proof of concept and development of a couple-based machine learning model to stratify infertile patients with idiopathic infertility. Sci Rep. 2021;11:24003. doi:10.1038/s41598-021-03165-3
37. Chowdhury M, Sadek AW. Advantages and limitations of artificial intelligence. Artif Intell Appl Crit Transp Issues. 2012;6:360-375.
38. Hogan NR, Davidge EQ, Corabian G. On the ethics and practicalities of artificial intelligence, risk assessment, and race. J Am Acad Psychiatry Law. 2021;49(3):326-334.
39. Swain JE. Optimizing the culture environment in the IVF laboratory: impact of pH and buffer capacity on gamete and embryo quality. Reprod Biomed Online. 2010;21(1):6-16. doi:10.1016/j.rbmo.2010.03.012
40. Ravichandran N, Inaganti AC, Muppalaneni R, Nersu SR. AI-powered workflow optimization in IT service management: Enhancing efficiency and security. Artif Intell Mach Learn Rev. 2020;1(3):10-26.
41. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci. 2021;7:e623. doi:10.7717/peerj-cs.623
42. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics. 2020;21(1):6. doi:10.1186/s12864-019-6413-7
43. Jia X, Ren L, Cai J. Clinical implementation of AI technologies will require interpretable AI models. Med Phys. 2020;47(1):1-4. doi:10.1002/mp.13891

Received 4-9-2025
Revised 17-9-2025
Accepted 20-9-2025