CIA Open Access

Integrating Ovarian Reserve Biomarkers with Machine Learning for Gonadotoxicity Risk Prediction in Young Female Cancer Patients: A Scoping Review

***Chuhan Zhang*[1], *Muyu Liu*[1,2]**

[1] *Applied Biostatistics and Epidemiology, University of Southern California, CA, USA*

[1.2] *Center for Applied Statistics and School of Statistics, Renmin University of China, Beijing, China*

DOI: 10.63575/CIA.2026.40111

Abstract

Cancer treatment-related gonadotoxicity poses substantial threats to reproductive health in young female patients. Current risk stratification frameworks, including ASCO clinical practice guidelines and the Oncofertility Pediatric Initiative Network (O-PIN) classification, rely predominantly on population-level statistics and expert consensus, offering limited capacity for individualized prediction. This scoping review examines the emerging intersection of ovarian reserve biomarkers and machine learning (ML) methodologies for gonadotoxicity risk prediction in young female cancer patients. Through systematic analysis of published literature, publicly available datasets, and clinical decision support tools, this review identifies key biomarkers—including anti-Müllerian hormone (AMH), antral follicle count (AFC), and follicle-stimulating hormone (FSH)—that serve as critical input features for predictive models. ML approaches, particularly random forest classifiers, have demonstrated promising discriminative performance (area under the receiver operating characteristic curve up to 0.87) in predicting primary ovarian insufficiency following chemotherapy. This review further evaluates how artificial intelligence may facilitate personalized fertility preservation pathway selection and clinical workflow integration. Challenges including limited sample sizes, absence of external validation, data heterogeneity, and algorithmic fairness concerns are critically discussed. Recommendations for multicenter prospective registries and standardized evaluation frameworks are proposed to advance clinical translation of AI-assisted oncofertility decision-making.

Keywords: *oncofertility; machine learning; gonadotoxicity risk prediction; ovarian reserve biomarkers*

1. Introduction

1.1 Epidemiological Background: Cancer Incidence Among Women of Reproductive Age

The global incidence of cancer among adolescents and young adults (AYAs), defined as individuals aged 15 to 39 years, has demonstrated a persistent upward trajectory over the past three decades. According to the Global Burden of Disease (GBD) 2021 database, approximately 383,241 AYA women worldwide received a new diagnosis of female cancer in 2021, with the age-standardized incidence rate exhibiting an average annual percentage change (AAPC) of 0.34% (95% CI: 0.22%–0.46%) from 1990 to 2021. The Surveillance, Epidemiology, and End Results (SEER) Program, maintained by the United States National Cancer Institute, reports that the 5-year relative survival rate for AYA cancer patients has improved to approximately 86% as of 2024 data, representing a marked advancement from the 75% documented a decade earlier. This improvement in long-term survival has fundamentally shifted the clinical landscape: post-treatment quality of life, particularly reproductive health, has become a central concern. The American Society of Clinical Oncology (ASCO), in its 2018 clinical practice guideline on fertility preservation, formally recommended that health care providers discuss the possibility of infertility with all patients of reproductive age as early as possible before treatment initiation [1]. The updated 2025 ASCO guideline further reinforced this recommendation, expanding counseling directives to encompass all patients of reproductive potential regardless of demographic background or planned treatment regimen [2].

1.2 Gonadotoxicity of Cancer Treatments and the Imperative of Fertility Preservation

Cancer-directed therapies—including chemotherapy, radiotherapy, and surgery—exert dose-dependent and agent-specific gonadotoxic effects on ovarian function. Alkylating agents, particularly cyclophosphamide, and anthracyclines such as doxorubicin are among the most damaging chemotherapeutic classes. Chemotherapy-induced amenorrhea has been reported in 61% of patients younger than 40 years and in 95% of those older than 40 years following standard regimens. The Childhood Cancer Survivor Study (CCSS), a multi-institutional retrospective cohort encompassing 20,720 five-year survivors diagnosed before age 21, provides the most comprehensive epidemiological data on long-term reproductive outcomes. Among 5,149 eligible female CCSS participants aged 15 to 44 years, the relative risk (RR) for ever being pregnant was 0.81 (95% CI: 0.73–0.90; $P < .001$) compared with female siblings, with those receiving hypothalamic/pituitary radiation doses ≥30 Gy (RR: 0.61; 95% CI: 0.44–0.83) or ovarian radiation doses exceeding 5 Gy demonstrating substantially greater risk reduction [3]. A separate CCSS analysis documented that acute ovarian failure (AOF) occurred in 6.3% of eligible survivors, with abdominal-pelvic irradiation contributing to 75% of AOF cases and alkylating agent exposure serving as an independent risk factor [4].

1.3 Research Objectives and Scope of This Review

Established fertility preservation options—including embryo cryopreservation, oocyte cryopreservation, ovarian tissue cryopreservation (OTC), and gonadotropin-releasing hormone agonist (GnRHa) ovarian suppression—vary in appropriateness based on patient age, partner status, time constraints, and cancer treatment urgency. Selecting the optimal preservation strategy requires integrating multiple clinical variables, yet current decision-making predominantly relies on clinician experience and population-level risk categories. The emergence of machine learning (ML) as a tool for personalized risk prediction in oncology has opened new possibilities for individualized gonadotoxicity assessment. A landmark cohort study involving the CCSS and St. Jude Lifetime Cohort (SJLIFE) demonstrated that treatment exposure variables could predict acute ovarian failure with meaningful clinical accuracy [5]. This scoping review addresses three research questions: (1) What AI and ML methods have been applied to gonadotoxicity risk prediction in young female cancer patients? (2) Which ovarian reserve biomarkers and clinical variables are integrated into these predictive approaches? (3) What are the principal challenges and future directions for AI-assisted fertility preservation decision-making? Literature was identified through systematic searches of PubMed, Web of Science, and Scopus databases using combinations of terms including “oncofertility,” “machine learning,” “ovarian reserve,” “gonadotoxicity,” and “fertility preservation.”

2. Ovarian Reserve Biomarkers and Current Gonadotoxicity Risk Assessment Frameworks

2.1 Key Ovarian Reserve Biomarkers in the Context of Oncofertility

Ovarian reserve biomarkers provide quantitative measures of remaining follicular pool size and reproductive potential. Anti-Müllerian hormone (AMH), produced by granulosa cells of primary, preantral, and small antral follicles, has emerged as the most sensitive and widely utilized marker. AMH levels peak at approximately 25 years of age and become undetectable before menopause, offering a continuous window for monitoring treatment-induced ovarian damage. Antral follicle count (AFC), assessed via transvaginal ultrasonography, provides a complementary morphological measure that correlates strongly with AMH. Follicle-stimulating hormone (FSH), inhibin B, and estradiol represent traditional endocrine markers; elevated FSH levels indicate diminished ovarian reserve, though these hormones demonstrate greater intra-cycle variability and lower sensitivity than AMH for detecting subclinical ovarian damage. A 2024 review published in npj Digital Medicine highlighted that automated, non-invasive methods to assess nuclear and cytoplasmic oocyte maturity would be particularly desirable for fertility preservation, as accurate prediction of oocyte quality from a pool of cryopreserved oocytes could guide decisions about the need for additional preservation cycles [6].

2.2 Established Risk Classification Systems: ASCO Guidelines, O-PIN, and CED Scoring

Multiple risk classification frameworks have been developed to stratify gonadotoxicity potential. The ASCO guidelines classify chemotherapy regimens into high, intermediate, and low gonadotoxicity risk categories based on historical cohort data. The Oncofertility Pediatric Initiative Network (O-PIN) risk stratification system, proposed in 2020, categorizes infertility risk into three tiers—minimally increased risk, significantly increased risk, and high level of increased risk—incorporating pubertal status, chemotherapy type and cumulative dose, and radiation exposure location and dose. The cyclophosphamide equivalent dose (CED) scoring system quantifies alkylating agent exposure by converting individual drug doses to a standardized cyclophosphamide-equivalent metric, enabling cross-regimen comparisons. A machine learning-based fertility risk calculator developed using multi-institutional data (N = 435) demonstrated that the current risk stratification system classified only 44% of patients who did not develop primary ovarian insufficiency (POI) into the correct low-risk category, compared to 100% accuracy achieved by the ML-based model for the same subgroup [7]. This performance gap underscores the granularity limitations inherent in categorical risk classification approaches.

2.3 Limitations of Traditional Risk Stratification Approaches

Traditional stratification systems share several structural constraints. These frameworks assign patients to broad risk categories (typically three to four tiers) rather than generating continuous, individualized probability estimates. Patient-specific variables—including baseline ovarian reserve (pre-treatment AMH and AFC), body mass index, and genetic predispositions such as BRCA1/2 mutation status—are insufficiently incorporated into existing classification rubrics. A mega-analysis of AMH levels in 608 female childhood cancer survivors demonstrated that AMH values correlated significantly with O-PIN risk level and that AMH levels returned to varying degrees over time, even after the highest-risk therapy, suggesting that static risk categories may not capture the dynamic nature of ovarian recovery trajectories [8]. The knowledge gap is particularly pronounced for newer treatment modalities: targeted therapies (tyrosine kinase inhibitors, monoclonal antibodies), immune checkpoint inhibitors, and cellular immunotherapy (CAR-T) lack systematic gonadotoxicity data. ASCO’s 2025 guideline explicitly acknowledged that effective risk stratification and counseling remain difficult to impossible for many novel agents. These limitations collectively establish the clinical rationale for ML-based approaches capable of integrating heterogeneous, multi-dimensional data into personalized gonadotoxicity risk assessments.

3. Machine Learning Applications in Gonadotoxicity Risk Prediction

3.1 Overview of ML Algorithms Applied in Oncofertility Research

The application of ML algorithms to oncofertility risk prediction represents an emerging research frontier, with the published literature documenting several distinct algorithmic approaches. Random forest (RF) classifiers have received the most extensive evaluation, owing to their capacity for handling non-linear feature interactions, robustness to overfitting in small-to-moderate datasets, and built-in feature importance estimation. Support vector machines (SVMs) have been employed for binary classification tasks (e.g., POI versus preserved ovarian function), leveraging kernel-based transformations to capture complex decision boundaries in high-dimensional feature spaces. Additional algorithms explored in the broader reproductive medicine literature include logistic regression, XGBoost, Bayesian network classifiers, and neural networks.

The most rigorously evaluated ML model for chemotherapy-induced POI prediction was developed using pooled de-identified data from six published studies (N = 435 patients). Random forest binary classifiers trained on this composite dataset predicted individualized POI risk with an accuracy of 88% and an area under the receiver operating characteristic curve (AUC) of 0.87 (95% CI: 0.77–0.96; P < .001). Independent validation using a cross-sectional survey of 117 cancer survivors confirmed generalizability. A more recent multi-center effort focusing on breast cancer patients combined seven datasets from the United States, Hong Kong, France, Denmark, Italy, Belgium, and international trial groups. This ML model identified 20 variables significantly associated with the risk of developing chemotherapy-induced amenorrhea at 12 months, achieving an internal validation AUC of 0.820 (95% CI: 0.817–0.823) and an external validation AUC of 0.743 (95% CI: 0.666–0.818) [9].

Table 1. Summary of Machine Learning Algorithms Applied in Oncofertility and Reproductive Medicine Research. AUC = area under the receiver operating characteristic curve; POI = primary ovarian insufficiency; ART = assisted reproductive technology; MAE = mean absolute error.

Algorithm	Application	Sample Size	Performance (AUC)	Validation
Random Forest	POI prediction post-chemotherapy	N = 435 (6 studies)	0.87 (95% CI: 0.77–0.96)	Independent set (N=117)
Random Forest	Amenorrhea at 12 months (breast cancer)	Multi-center (7 datasets)	0.820 (internal); 0.743 (external)	External dataset
Support Vector Machine	Sperm/embryo classification in ART	Variable	0.82–0.91 (reported range)	Internal cross-validation
XGBoost	Oocyte yield prediction	62,354 cycles (9 studies)	MAE: 0.62–4.13	Internal (no external)
Logistic Regression	Gonadotoxicity risk stratification	Variable	Baseline comparator	Variable
Neural Network	Embryo quality assessment	Variable	Generally superior in imaging tasks	Internal (limited external)

Table 2. Key Input Features for ML-Based Gonadotoxicity Risk Prediction Models. AMH = anti-Müllerian hormone; AFC = antral follicle count; FSH = follicle-stimulating hormone; CED = cyclophosphamide equivalent dose; EHR = electronic health record.

Feature Category	Specific Variables	Data Source
Demographics	Age at diagnosis, age at treatment, BMI, race/ethnicity	Clinical records, EHR
Ovarian Reserve	Baseline AMH (ng/mL), AFC (count), FSH (mIU/mL), estradiol, inhibin B	Laboratory assays, ultrasound
Tumor Characteristics	Cancer type, histological subtype, disease stage, tumor location	Pathology reports, staging records
Treatment Exposure	Chemotherapy regimen, CED score (g/m²), radiation dose (Gy) and field, surgical extent	Treatment protocols, dosimetry records

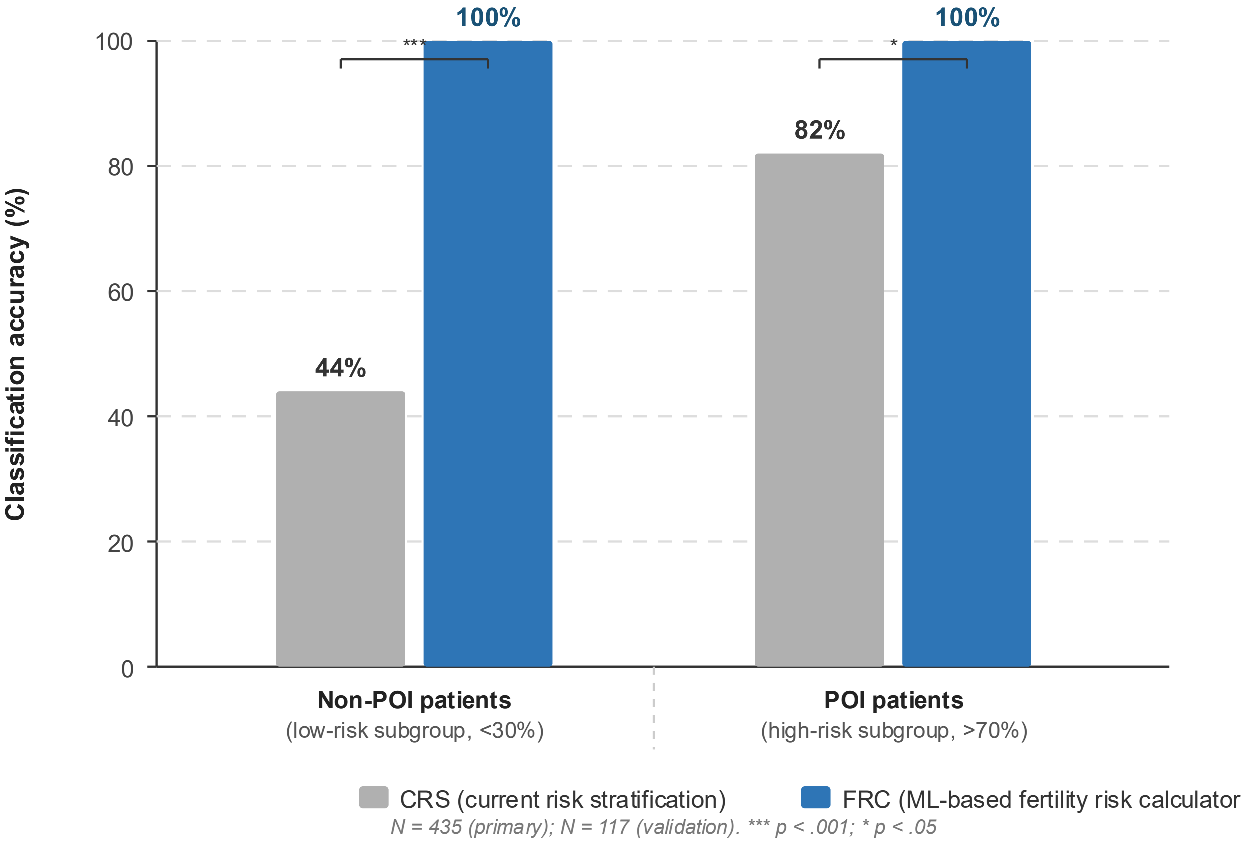
Feature Category	Specific Variables	Data Source
Genetic Factors	BRCA1/2 mutation status, other germline variants	Genetic testing records
Temporal Variables	Time since treatment completion, pubertal status at diagnosis	Clinical follow-up records

3.2 Multi-Dimensional Feature Integration for Personalized Risk Prediction

The predictive superiority of ML approaches over traditional risk classification derives principally from their capacity to integrate heterogeneous feature types and capture non-linear interactions among variables. In the fertility risk calculator study, the model incorporated patient age at diagnosis, menstrual status at baseline, specific chemotherapy agents and cumulative doses, radiation exposure parameters, and BMI, producing a continuous probability output rather than a categorical tier assignment. Shapley Additive Explanations (SHAP) analysis—a model-agnostic interpretability technique—has been applied to rank feature importance and provide clinicians with transparent rationale for individual predictions. A broad review of artificial intelligence in reproductive medicine documented that SVMs, random forests, and Bayesian networks represent the most frequently employed and empirically validated ML approaches across the reproductive health domain [10].

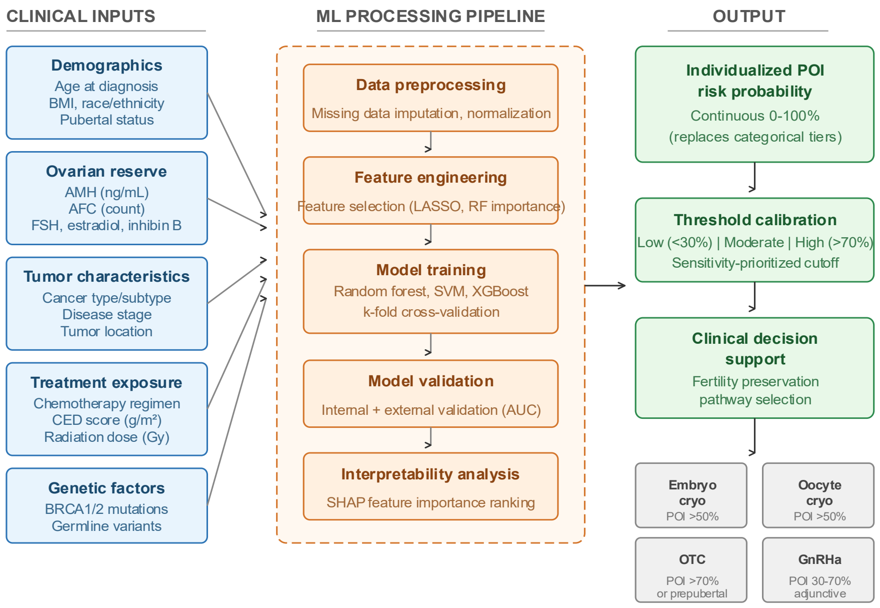
The multi-center breast cancer amenorrhea prediction study identified the 20 most influential variables through systematic feature selection, with age at chemotherapy initiation, baseline hormonal profiles, and alkylating agent cumulative dose consistently ranking among the highest-weighted predictors. The deliberate selection of a low classification threshold (0.20) in that study prioritized sensitivity (91.3% internal; 92.9% external) over precision, reflecting the clinical imperative that false negatives—patients incorrectly classified as likely to regain menses—could result in missed opportunities for timely fertility preservation. This threshold calibration illustrates how ML model design can be aligned with the asymmetric cost structure inherent in oncofertility decision-making, where the consequence of underestimating risk substantially exceeds that of overestimation.

Figure 1. Performance Comparison of ML-Based Fertility Risk Calculator Versus Traditional Gonadotoxicity Risk Stratification.



Bar chart depicting classification accuracy stratified by true outcome status. Among patients who did not develop POI, the ML-based fertility risk calculator (FRC) achieved 100% accuracy in the low-risk subgroup (<30% predicted probability), compared with 44% accuracy for the current risk stratification (CRS) system. Among patients who developed POI, FRC achieved 100% accuracy in the high-risk subgroup (>70% predicted probability), while CRS achieved 82%. Data derived from Benedict et al. (2022); N = 435 primary dataset, N = 117 independent validation set.

Figure 2. Conceptual Framework of Multi-Dimensional Feature Integration for Personalized Gonadotoxicity Risk Prediction.



AMH = anti-Müllerian hormone; AFC = antral follicle count; FSH = follicle-stimulating hormone; CED = cyclophosphamide equivalent dose; OTC = ovarian tissue cryopreservation; GnRHa = gonadotropin-releasing hormone agonist; SHAP = Shapley Additive Explanations

Schematic diagram illustrating the data pipeline from multi-source clinical inputs (demographics, ovarian reserve biomarkers, tumor characteristics, treatment exposure variables, and genetic factors) through ML processing (feature engineering, model training with cross-validation, SHAP-based interpretability analysis) to individualized risk probability output. The framework depicts how continuous probability estimates replace categorical risk tiers, enabling threshold-calibrated clinical decision support for fertility preservation pathway selection.

3.3 Publicly Available Datasets and Data Infrastructure for Oncofertility ML Research

The development and validation of ML models for gonadotoxicity risk prediction depend critically on the availability of large, well-annotated, multi-institutional datasets. The CCSS remains the largest and most comprehensively characterized cohort resource, encompassing 20,720 five-year survivors with detailed treatment exposure annotation, longitudinal follow-up data, and an established biorepository. Public access data tables are available through the St. Jude Survivorship Portal (https://ccss.stjude.org/public-access-data.html), providing demographic, treatment, and outcome variables for external investigators. The SEER Program offers population-based cancer incidence and survival data spanning 1975 to 2022, with geographic, demographic, and tumor-specific stratification. The GBD 2021 database provides global estimates of disease burden across 204 countries, enabling international epidemiological contextualization. A review of AI readiness in assisted reproductive technologies emphasized that the performance of predictive tools requires ongoing surveillance and that implementation must proceed as an informed, cautious, and circumspect process [11].

Despite the existence of these resources, several data infrastructure challenges impede progress. Sample sizes for oncofertility-specific ML studies remain modest (N = 117 to 608 in the largest published analyses), constraining statistical power and generalizability. Multi-center data heterogeneity—arising from differences in AMH assay platforms, treatment protocol documentation standards, and outcome definitions—introduces systematic variability that complicates model aggregation. The absence of standardized ontologies for fertility-related outcomes (e.g., varying definitions of amenorrhea duration, AOF criteria, and POI diagnostic thresholds) further limits cross-study harmonization. Federated learning, which enables collaborative model training across decentralized institutional datasets without sharing raw patient-level data, has been proposed as a privacy-preserving solution to these challenges, though its application in oncofertility remains theoretical.

Table 3. Publicly Available Datasets Relevant to Oncofertility Machine Learning Research. CCSS = Childhood Cancer Survivor Study; SJLIFE = St. Jude Lifetime Cohort; SEER = Surveillance, Epidemiology, and End Results; GBD = Global Burden of Disease; IHME = Institute for Health Metrics and Evaluation; IARC = International Agency for Research on Cancer; NCI = National Cancer Institute.

Dataset	**Source/Sponsor**	**Population**	**Key Variables**	**Access**
CCSS	NCI (U24 CA55727) / St. Jude	20,720 survivors (<21 at diagnosis, 1970–1999)	Treatment exposures, fertility outcomes, AOF, hormonal data	Public access tables; proposal-based
SJLIFE	St. Jude Children’s Research Hospital	~5,000+ adult survivors with clinical assessments	Prospective clinical evaluations, biomarkers, organ function	Proposal-based collaboration
SEER	NCI Surveillance Research Program	Population-based (U.S.), 1975–2022	Incidence, survival, staging, demographics by age/sex/race	Research data agreement
GBD 2021	IHME	204 countries, 371 diseases, 1990–2021	Incidence, mortality, DALYs, prevalence by age/sex	Open access (https://ghdx.healthdata.org/)
GLOBOCAN	IARC / WHO	185+ countries	Cancer incidence and mortality by site, age, sex	Open access (https://gco.iarc.fr/)

4. AI-Assisted Fertility Preservation Decision Support

4.1 AI-Driven Pathway Optimization for Fertility Preservation Options

The selection of an appropriate fertility preservation strategy depends on multiple interacting factors, including patient age, partner status, available time before treatment initiation, cancer type, and the specific gonadotoxicity profile of the planned treatment regimen. Embryo cryopreservation, the most established technique, requires an available sperm source and a two-week ovarian stimulation cycle. Oocyte cryopreservation offers comparable outcomes without requiring partner involvement and has become the predominant option for unpartnered patients. Ovarian tissue cryopreservation (OTC) is the only option available for prepubertal patients or those requiring immediate treatment initiation without time for ovarian stimulation. GnRHa ovarian suppression may offer an adjunctive protective benefit during chemotherapy, though its role as a standalone preservation method remains debated. An emerging review of AI applications across the in vitro fertilization cycle noted that digital twins of future parents and embryos could allow clinicians to simulate treatment options virtually before making clinical decisions, representing a paradigm shift in how fertility care is delivered [12].

AI-driven pathway optimization could integrate individualized gonadotoxicity risk probability with patient-specific clinical parameters to generate ranked preservation recommendations. A patient predicted to have a high probability of POI (e.g., >70%) with limited time before chemotherapy initiation might receive a primary recommendation for OTC with concurrent GnRHa, whereas a patient with moderate risk (30–70%) and adequate time could be directed toward oocyte cryopreservation. This risk-stratified approach represents a departure from the current paradigm, in which preservation decisions are often driven by clinician preference and institutional availability rather than by quantitative, patient-specific risk estimates.

Table 4. Fertility Preservation Options with Proposed AI-Assisted Decision Criteria. POI = primary ovarian insufficiency; GnRHa = gonadotropin-releasing hormone agonist. Decision criteria represent conceptual thresholds derived from current evidence and are not yet validated in prospective clinical trials.

Preservation Option	Time Required	Optimal Candidate Profile	AI-Derived Decision Criteria
Embryo cryopreservation	~2 weeks (ovarian stimulation)	Postpubertal, with partner, moderate-high risk, adequate time	POI probability >50%; time to treatment ≥14 days; partner available
Oocyte cryopreservation	~2 weeks (ovarian stimulation)	Postpubertal, any partner status, moderate-high risk, adequate time	POI probability >50%; time to treatment ≥14 days; no partner requirement

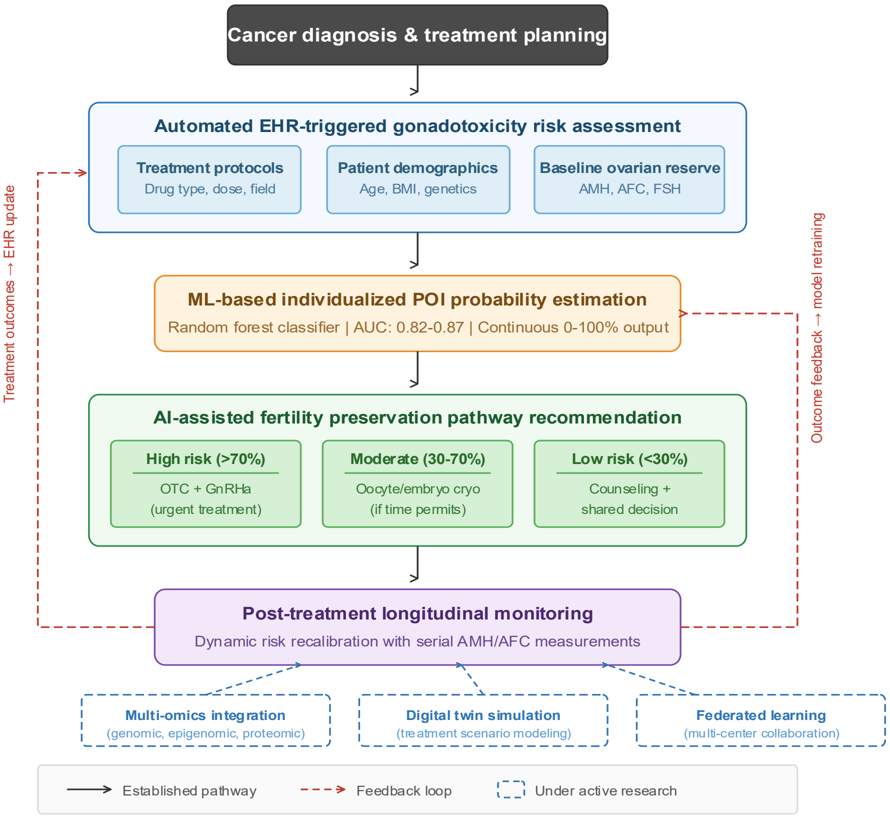
Preservation Option	Time Required	Optimal Candidate Profile	AI-Derived Decision Criteria
Ovarian tissue cryopreservation	1–2 days (surgical)	Pre- or postpubertal, high risk, urgent treatment	POI probability >70%; time to treatment <14 days; or prepubertal status
GnRHa ovarian suppression	Concurrent with treatment	Postpubertal, adjunctive use, any risk level	Supplementary to other methods; moderate POI probability (30–70%)
No active preservation	N/A	Minimal risk, patient preference	POI probability <20%; patient-informed shared decision

4.2 Integration of AI into Oncofertility Clinical Workflows

Translating ML-based risk prediction into routine clinical practice requires integration with existing oncology and reproductive medicine workflows. Electronic health record (EHR) systems offer a natural infrastructure for embedding AI-driven clinical decision support. Automated risk alerts triggered by treatment orders containing gonadotoxic agents could prompt oncologists to initiate fertility counseling and specialist referral at the point of care, addressing the well-documented gap in oncofertility referral rates. The feasibility of such integration has been demonstrated in other clinical domains, where EHR-embedded ML models have achieved successful deployment for sepsis prediction, readmission risk scoring, and clinical trial eligibility screening. A comprehensive review of fertility preservation and long-term gonadotoxicity monitoring emphasized that a multidisciplinary team—including oncologists, pediatricians, gynecologists, endocrinologists, and psychologists—should collaboratively decide on preservation options following the estimation of gonadotoxicity potential [13].

Beyond risk prediction, AI applications extend into the fertility preservation procedure itself. In the IVF laboratory, ML-based tools for oocyte quality assessment, personalized gonadotropin dosing, and embryo selection have progressed toward clinical deployment. For oncofertility patients undergoing urgent oocyte cryopreservation before chemotherapy, AI-assisted ovarian stimulation protocols could optimize the balance between maximizing oocyte yield and minimizing cycle duration. The development of web-based clinical predictive tools—such as the freely accessible fertility risk calculator—demonstrates the potential for democratizing access to personalized risk information beyond specialized oncofertility centers. AI may also serve a connective function within multidisciplinary care teams, synthesizing inputs from oncology, reproductive endocrinology, genetics, and psychosocial services into a unified patient management dashboard.

Figure 3. Proposed Architecture for AI-Integrated Oncofertility Clinical Decision Pathway.



Flow diagram depicting the integration of AI modules within the oncofertility clinical workflow. The pathway begins with cancer diagnosis and treatment planning, progressing through automated EHR-triggered gonadotoxicity risk assessment (incorporating treatment protocols, patient demographics, and baseline ovarian reserve biomarkers), ML-based individualized POI probability estimation, AI-assisted fertility preservation pathway recommendation, and post-treatment longitudinal monitoring with dynamic risk recalibration. Feedback loops connect treatment outcomes to model retraining, enabling continuous performance improvement. Dashed lines indicate components that remain under active research development.

4.3 Ethical Considerations and Patient-Centered Decision-Making

The deployment of AI-based tools in oncofertility decision-making raises several ethical considerations that warrant careful examination. Algorithmic fairness represents a primary concern: training datasets in the published literature are disproportionately composed of participants from high-income countries and predominantly White populations. The CCSS cohort, while large, reflects the demographic composition of participating U.S. institutions during the 1970–1999 diagnosis period. Racial and ethnic minority populations, who already face documented disparities in oncofertility referral rates, may receive less accurate predictions from models trained on non-representative data. Assessment of cancer treatment effects on ovarian reserve in diverse populations has demonstrated that alkylating agent dose score and pelvic radiation exposure were associated with impaired markers in a dose-dependent manner, yet the magnitude of these effects may vary across demographic groups [14].

The communication of AI-generated risk predictions to patients demands nuanced attention to uncertainty quantification. Presenting a continuous probability (e.g., "63% likelihood of POI") rather than a categorical label ("high risk") may enhance patient comprehension and engagement in shared decision-making, but it may also elevate anxiety, particularly in the emotionally fraught period surrounding cancer diagnosis. The absence of prospective validation for existing ML models means that all current predictions carry substantial epistemic uncertainty that must be transparently communicated. Regulatory pathways for AI-based clinical decision support tools in oncofertility remain incompletely defined, as these tools occupy an ambiguous space between informational resources and medical devices. The development of evaluation standards aligned with the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) and PROBAST (Prediction Model Risk of Bias Assessment Tool) frameworks is needed to ensure methodological rigor prior to clinical deployment.

5. Discussion and Future Directions

5.1 Current Challenges in AI-Assisted Oncofertility Decision-Making

The evidence reviewed in this paper reveals a field at an early but promising stage of development. ML-based gonadotoxicity risk prediction has demonstrated discriminative performance substantially exceeding that of traditional categorical risk stratification, with AUC values ranging from 0.74 to 0.87 in available studies. These results warrant cautious optimism rather than premature enthusiasm: all published models carry significant methodological limitations. Sample sizes remain modest by ML standards, with the largest composite dataset comprising 435 patients from six heterogeneous studies. External validation—the critical benchmark for assessing generalizability—has been performed in only two published models, and no prospective validation study has been completed. Reporting standards are inconsistent, with many studies lacking transparent documentation of data preprocessing pipelines, missing value handling strategies, and variable transformation procedures. The risk of bias assessment tools PROBAST and TRIPOD, now standard in clinical prediction model research, have been applied to fewer than half of the oncofertility ML studies reviewed. Clinicians should approach the current generation of models as investigational tools requiring independent validation before integration into clinical care pathways.

5.2 Emerging Technologies and Multi-Modal Data Integration

Several technological developments may catalyze progress in AI-assisted oncofertility care. Multi-omics data integration—combining genomic, epigenomic, transcriptomic, and proteomic profiles with clinical variables—represents a logical extension of current feature engineering approaches. Germline variants in DNA repair and ovarian function genes may modulate individual susceptibility to treatment-induced gonadotoxicity, providing mechanistic features that complement phenotypic biomarkers. The digital twin concept, in which patient-specific computational models simulate responses to alternative treatment scenarios, has gained traction in oncology precision medicine and could be adapted for fertility outcome prediction. Large language models (LLMs) present opportunities for patient education and counseling support, generating personalized informational content about gonadotoxicity risks and preservation options at accessible literacy levels. Federated learning architectures offer a technically viable framework for multi-institutional collaboration without centralized data pooling, addressing both privacy requirements and sample size limitations. Wearable devices and remote monitoring platforms may provide longitudinal data streams—including menstrual cycle tracking, hormonal fluctuation patterns, and physical activity metrics—that enrich the temporal dimension of predictive models.

5.3 Recommendations for Future Research and Clinical Implementation

This review identifies five priorities for advancing the clinical translation of AI-assisted oncofertility decision-making. The establishment of large-scale, multicenter, prospective oncofertility data registries with standardized variable definitions and outcome ascertainment protocols is the foundational requirement. These registries should incorporate pretreatment ovarian reserve biomarkers (AMH, AFC) as mandatory baseline assessments, enabling the development of models with clinically actionable input features. The creation of harmonized evaluation standards for oncofertility AI tools—adapted from TRIPOD, PROBAST, and related frameworks—would enable systematic quality assessment and cross-study comparison. Advocacy for the inclusion of AI-based risk assessment tools within ASCO, European Society for Medical Oncology (ESMO), and related clinical practice guidelines would facilitate evidence-based clinical adoption. Systematic investigation of gonadotoxicity profiles for targeted therapies, immune checkpoint inhibitors, and cellular immunotherapies is essential to populate the knowledge base upon which ML models depend. Active engagement of diverse patient populations in cohort enrollment and model validation is required to mitigate algorithmic fairness risks and ensure equitable predictive performance across demographic groups.

The integration of ovarian reserve biomarkers with machine learning for gonadotoxicity risk prediction represents a meaningful—if nascent—advancement in personalized oncofertility care. The evidence reviewed here supports the feasibility and potential clinical value of ML-based risk prediction, while underscoring the considerable distance that remains between research prototypes and validated clinical tools. Progress will depend on sustained, collaborative investment in data infrastructure, methodological rigor, and patient-centered evaluation.

References

[1]. Oktay, K., Harvey, B. E., Partridge, A. H., Quinn, G. P., Reinecke, J., Taylor, H. S., Wallace, W. H., Wang, E. T., & Loren, A. W. (2018). Fertility preservation in patients with cancer: ASCO clinical practice guideline update. Journal of Clinical Oncology, 36(19), 1994–2001. https://doi.org/10.1200/JCO.2018.78.1914

[2]. Su, H. I., Lacchetti, C., Letourneau, J., Partridge, A. H., Qamar, R., Quinn, G. P., Reinecke, J., Smith, J. F., Tesch, M., Wallace, W. H., Wang, E. T., & Loren, A. W. (2025). Fertility preservation in people with cancer: ASCO guideline update. Journal of Clinical Oncology, 43(12), 1488–1515. https://doi.org/10.1200/JCO-24-02782

[3]. Green, D. M., Kawashima, T., Stovall, M., Leisenring, W., Sklar, C. A., Mertens, A. C., Donaldson, S. S., Byrne, J., & Robison, L. L. (2009). Fertility of female survivors of childhood cancer: A report from the Childhood Cancer Survivor Study. Journal of Clinical Oncology, 27(16), 2677–2685. https://doi.org/10.1200/JCO.2008.20.1541

[4]. Green, D. M., Sklar, C. A., Boice, J. D., Jr., Mulvihill, J. J., Whitton, J. A., Stovall, M., & Yasui, Y. (2009). Ovarian failure and reproductive outcomes after childhood cancer treatment: Results from the Childhood Cancer Survivor Study. Journal of Clinical Oncology, 27(14), 2374–2381. https://doi.org/10.1200/JCO.2008.21.1839

[5]. Clark, R. A., Mostoufi-Moab, S., Yasui, Y., Vu, N. K., Sklar, C. A., Motan, T., Engel, M. E., Green, D. M., Leisenring, W. M., Robison, L. L., Armstrong, G. T., & Levine, J. E. (2020). Predicting acute ovarian failure in female survivors of childhood cancer: A cohort study in the Childhood Cancer Survivor Study (CCSS) and the St Jude Lifetime Cohort (SJLIFE). The Lancet Oncology, 21(3), 436–445. https://doi.org/10.1016/S1470-2045(19)30818-6

[6]. Hanassab, S., Abbara, A., Yeung, A. C., Jayasena, C. N., & Dhillo, W. S. (2024). The prospect of artificial intelligence to personalize assisted reproductive technology. npj Digital Medicine, 7, Article 55. https://doi.org/10.1038/s41746-024-01006-x

[7]. Benedict, C., Nieh, J. L., Guana, A. L., Gingold, J., & Hershlag, A. (2022). Development of a fertility risk calculator to predict individualized chance of ovarian failure after chemotherapy. Journal of Cancer Survivorship, 16, 1174–1183. https://doi.org/10.1007/s11764-021-01109-w

[8]. Maher, J. C. Y., Kumnick, A., Sinaii, N., Su, H. I., Cameron, K. E., George, S. A., Gracia, C., Meacham, L. R., & Gomez-Lobo, V. (2025). A mega-analysis of anti-Müllerian hormone levels in female childhood cancer survivors based on treatment risk, time since treatment, and pubertal status. Journal of Adolescent and Young Adult Oncology, 14(1). https://doi.org/10.1089/jayao.2024.0093

[9]. Song, L., Edib, Z., Aickelin, U., Akbarzadeh Khorshidi, H., Hamy, A.-S., Jayasinghe, Y., Hickey, M., Anderson, R. A., Lambertini, M., Condorelli, M., Demeestere, I., Ignatiadis, M., Pistilli, B., Su, H. I., Chang, S., Pang, P. C.-I., Reyal, F., Nelson, S. M., Sukumvanich, P., & Peate, M. (2025). Development of a machine learning model for predicting treatment-related amenorrhea in young women with breast cancer. Bioengineering, 12(11), Article

[10]. https://doi.org/10.3390/bioengineering12111171

[11]. Wang, R., Pan, W., Jin, L., Li, Y., Geng, Y., Gao, C., Chen, G., Wang, H., Ma, D., & Liao, S. (2019). Artificial intelligence in reproductive medicine. Reproduction, 158(4), R139–R154. https://doi.org/10.1530/REP-18-0523

[12]. Letterie, G. (2023). Artificial intelligence and assisted reproductive technologies: 2023. Ready for prime time? Or not. Fertility and Sterility, 120(1), 32–37. https://doi.org/10.1016/j.fertnstert.2023.05.146

[13]. Jiang, V. S., Bormann, C. L., & Sturmey, R. G. (2025). Current progress and open challenges for applying artificial intelligence across the in vitro fertilization cycle. Patterns, 6(9), Article

[14]. https://doi.org/10.1016/j.patter.2025.101128

[15]. Wołczyński, S., Łapucińska, K., & Grygoruk, C. (2021). Fertility preservation and long-term monitoring of gonadotoxicity in girls, adolescents and young adults undergoing cancer treatment. Cancers, 13(2), Article 202. https://doi.org/10.3390/cancers13020202

[16]. Gracia, C. R., Sammel, M. D., Freeman, E., Prewitt, M., Carlson, C., Ray, A., Vance, A., & Ginsberg, J. P. (2012). Assessing the impact of cancer therapies on ovarian reserve. Fertility and Sterility, 97(1), 134–140. https://doi.org/10.1016/j.fertnstert.2011.10.040