

SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>

Research Paper

Open Access

Artificial Intelligence Enhanced Prediction of Live Birth Outcomes for Informed Clinical Decision-Making: A Comprehensive Survey of Data-Driven Models, Predictive Factors, and Reproductive Healthcare Systems

Kusuma K¹, Dr. Savita Choudhary²

¹Research Scholar, Department of Computer Science and Engineering, Sir M. Visvesvaraya Institute of Technology, Bengaluru, Karnataka, India. Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India. E-mail: Kusuma3k@gmail.com

²Professor, Department of Computer Science and Engineering, Sir M. Visvesvaraya Institute of Technology, Bengaluru, Karnataka, India. E-mail: savitha_cs@sirmvit.edu

ABSTRACT

Live birth is considered as an important clinical endpoint in the reproductive medicine and assisted reproductive technology (ART). Conventional statistical approaches are struggling in capturing the nonlinear relationships and its associated hidden patterns in the large-scale fertility datasets. This survey comprehensively reviews the recent studies published between 2020 and 2026 that is focusing on AI-enhanced live birth prediction systems. The reviewed literature includes conventional machine learning algorithms such as Logistic Regression, Random Forest, support vector Machine (SVM), Extreme Gradient Boosting (XGBoost), LightGBM, CatBoost, and ensemble learning models, where the well as deep learning architectures including Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Transformer-based networks, and explainable AI frameworks. The survey is categorizing predictive variables into demographic, clinical, hormonal, embryological, genetic, laboratory, and lifestyle factors while comparing model performance across the diverse fertility datasets. Analysis of the reviewed literature is indicating that AI-based approaches is consistently performing than the conventional statistical models. Logistic Regression models reported predictive accuracies ranging from the 62% to 78%, while Random Forest and Gradient Boosting techniques achieved accuracies between the 75% and 89%. Advanced ensemble models showing Area under the Curve (AUC) values between the 0.82 and 0.93. Deep learning architectures achieved prediction accuracies ranging from the 85% to 94%, with the Transformer-based and hybrid ensemble frameworks reporting AUC values exceeding 0.95 in several large-scale studies. Across the surveyed works, maternal age, embryo quality, anti-Müllerian hormone (AMH), antral follicle count (AFC), endometrial thickness, fertilization rate, blastocyst development score, body mass index (BMI), and lifestyle indicators emerged as most of the influential predictors.

Keywords: Live Birth Prediction, Artificial Intelligence in Reproductive Medicine, Machine Learning for IVF Outcomes, Explainable Artificial Intelligence, Assisted Reproductive Technology (ART) Analytics

Introduction

Infertility is become a major global healthcare concern, affecting millions of couples who experience difficulties in achieving pregnancy naturally. Artificial intelligence is providing the capability to analyze heterogeneous medical data and identify complex associations between the input variables and reproductive outcomes. Machine learning (ML) algorithms can process structured clinical information, while deep learning (DL) approaches can analyze high-dimensional data such as embryo images, time-lapse microscopy sequences, and genomic information. These capabilities are allowing AI-driven systems to assist clinicians in selecting optimal treatment strategies, which is improving embryo selection, reducing unsuccessful treatment cycles, and increasing the probability of live birth [1].

The primary objective of this survey is to systematically analyze recent AI-based approaches developed for predicting live birth outcomes and improving reproductive decision-making. The survey that is focussing on identifying important predictive factors, evaluating computational methodologies, and examining the effectiveness of Conventional AI systems.

The survey considers multiple categories of predictive variables, including demographic characteristics, reproductive history, hormonal measurements, embryo-related parameters, genetic indicators, and lifestyle factors. It also is evaluating publicly available datasets, evaluation metrics, challenges, and future research directions.

2. Artificial Intelligence in Reproductive Healthcare

2.1 Evolution of Artificial Intelligence in Healthcare

Artificial intelligence (AI) is evolving from the theoretical computational concept into an essential component

of modern healthcare systems [2]. The fundamental concept of AI-based prediction can be represented mathematically as:

$$Y = f(X, \theta)$$

where X is input clinical and biological features, Y is predicted reproductive outcomes, θ and it is the learned parameters of the AI model. During training, the model adjusts its parameters to minimize the difference between the predicted and actual outcomes.

The objective function of model optimization can be expressed as:

$$\theta^* = \arg \min_{\theta} L(Y, \hat{Y})$$

where L is the loss function, Y is the actual outcome, $\hat{Y}$ and it is the predicted outcome.

In reproductive medicine, this learning capability is allowing AI systems to analyze thousands of fertility treatment cycles and identify combinations of factors associated with the successful live births. The increasing availability of IVF databases, embryo imaging repositories, genomic information, and electronic health records is accelerated the adoption of AI-driven fertility prediction systems.

Recent AI developments are introducing advanced approaches such as deep learning, multimodal learning, reinforcement learning, and explainable artificial intelligence (XAI). These techniques are expanded the role of AI from the simple prediction systems towards a intelligent clinical decision-support frameworks that is capable of assisting fertility specialists throughout the treatment process.

3.AI Applications in Fertility Treatment

Artificial intelligence is applied across the multiple stages of fertility treatment, including patient evaluation, ovarian stimulation prediction, embryo assessment, implantation prediction, and live birth forecasting [3]. Deep learning models, especially convolutional neural networks (CNNs), are showing strong capability in extracting complex visual features from the embryo images. These models can identify subtle morphological characteristics associated with the implantation potential that tends not to be easily recognized by human observers.

The embryo quality prediction process can be expressed as:

$$Q_e = g(I_t, M, D)$$

where Q_e is embryo quality score, I_t is time-lapse imaging information, M is morphological characteristics, D and it is developmental patterns.

AI is also increasingly used for predicting implantation success and live birth outcomes. These systems is combining the embryo-level information with the patient characteristics to estimate the likelihood of successful delivery. Such predictions help clinicians select appropriate embryos and it is providing personalized counselling to patients.

Further, AI is applied in male fertility analysis, including sperm classification, motility assessment, and DNA fragmentation prediction. Computer vision-based AI models can automatically analyze sperm images and identify characteristics associated with the fertilization potential.

4.AI-Based Clinical Decision Support Systems

Clinical decision Support systems (CDSS) powered by artificial intelligence are providing data-driven recommendations to assist healthcare professionals in making informed decisions. In reproductive medicine, AI-based CDSS platforms aim in combining patient information, laboratory measurements, imaging data, and historical treatment outcomes to are generating personalized recommendations [4].

A conventional clinical workflow depends on physician experience and established treatment guidelines. Although expert knowledge is remaining as essential, complex fertility cases tend to involve hundreds of interacting parameters that is difficult to evaluate simultaneously. AI-based systems overcome this limitation by processing large-scale information and identifying predictive relationships.

The input data tends to include demographic information, hormonal measurements, embryo images, genetic information, and previous ART outcomes. Feature processing is transforming raw data into meaningful representations that is considered suitable for computational analysis. The AI model then generates predictions that tends in supporting the clinical decisions.

In IVF treatment, AI-based decision Support can assist with the embryo selection, treatment protocol optimization, and patient risk classification. For example, a model tends to estimate that a particular embryo is a higher implantation probability, which is based on developmental characteristics and patient-specific factors. Similarly, a risk prediction system tends to identify patients who are requiring additional evaluation before undergoing another treatment cycle.

AI-based CDSS also supports continuous learning. Unlike static clinical guidelines, ML systems can update its predictive patterns as new treatment data become available. The combination of AI-generated insights and

clinical expertise is providing most of the effective approach for improving reproductive healthcare.

4.4 Explainable Artificial Intelligence in Fertility Prediction

Although AI models are showing strong predictive performance, its adoption in clinical environments is often limited lack of transparency. Many advanced deep learning models is operating as black-box systems, where the relationship between the input variables and predicted outcomes is difficult to understand. in medical applications, clinicians are requiring explanations because treatment decisions directly affect patient health and well-being [5].

Explainable artificial intelligence (XAI) is addressing the limitation by providing interpretable information about how AI models are generating predictions. XAI techniques identify the contribution of individual features and reveal the reasoning behind model outputs.

The prediction process of an AI model can be represented as:

$$Y^{\wedge} = f(X)$$

where the objective of XAI is in determining the contribution of each feature:

$$Importance(x_i) = \frac{\partial f(X)}{\partial x_i}$$

where x_i is an individual input feature.

Several explanation techniques are introducing for fertility prediction systems. SHAP (SHapley Additive exPlanations) is evaluating the contribution of each variable by considering different combinations of input features. It is providing a quantitative measure of how much each factor influences the final prediction.

LIME (Local Interpretable Model-Agnostic Explanations) generates simplified local models around individual predictions to explain specific outcomes. in fertility applications, LIME can help clinicians understand why a particular patient is a high or a lower predicted probability of live birth.

Feature importance analysis is another commonly used approach. AI models can rank factors such as maternal age, AMH level, embryo quality, and endometrial thickness in accordance with the influence on prediction results.

5. ML Techniques for Live Birth Prediction

5.1 Introduction to ML in Live Birth Prediction

The primary objective of ML-based live birth prediction is to construct a computational model that is estimating the probability of successful delivery, which is based on patient and treatment-related features. Given a dataset containing input variables and corresponding outcomes, the ML model is learning a mapping function between the predictors and live birth status [7].

The general ML formulation can be expressed as:

$$Y = f(X) + \epsilon$$

where X is the input feature vector, Y is the predicted live birth outcome, f is the learned model function, ϵ and it is prediction error.

For fertility prediction, the feature vector tends to contain:

$$X = \{Age, AMH, FSH, BMI, AFC, Embryo Quality, IVF History, Lifestyle Factors\}$$

where each variable is contributing information regarding reproductive potential.

ML models are generally trained using historical ART datasets containing successful and unsuccessful treatment cycles. During training, the algorithm is identifying patterns associated with the positive and negative outcomes. The trained model is then evaluated using unseen data in determining the generalization capability.

The optimization objective of a ML model can be represented as:

$$\theta^* = \arg \min_{\theta} \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i, \theta))$$

where N is the number of training samples, L is the loss function, y_i is the actual outcome, $f(x_i, \theta)$ and it is the predicted outcome.

In live birth prediction, ML approaches are used for classification tasks where the output is successful or unsuccessful birth outcomes. However, regression-based approaches are also used to estimate continuous outcomes such as implantation probability, embryo quality score, or expected number of retrieved oocytes.

The major advantages of ML techniques include its ability to handle high-dimensional data, identify nonlinear relationships, are performing a automated feature selection, and it is improving the prediction accuracy through the continuous learning. These characteristics make ML highly that is considered suitable for developing personalized reproductive healthcare systems.

5.2 Logistic Regression

Logistic regression is one of the earliest and most of the widely used statistical ML methods for binary classification problems. In live birth prediction, logistic regression is estimating the probability of successful birth, which is based on multiple clinical and biological variables [8].

Unlike linear regression, which predicts continuous values, logistic regression applies a sigmoid function in transforming the output into a probability between the zero and one.

The logistic regression model is represented as:

$$P(Y=1 | X) = \frac{1}{1 + e^{-Z}}$$

where:

$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Here, $P(Y = 1 | X)$ is the probability of live birth, β_0 is the intercept, β_i and it is the contribution of each feature. In fertility prediction applications, logistic regression models commonly use variables such as maternal age, AMH concentration, embryo grade, number of transferred embryos, and previous IVF history. The model is providing interpretable coefficients that is showing whether a particular factor is increasing or decreases the likelihood of live birth.

One important advantage of logistic regression is its interpretability. Clinicians can easily understand the relationship between the predictors and outcomes because each feature is a measurable contribution. Therefore, logistic regression tends to achieve lower predictive performance when it is compared with the advanced ML models when it is applied to large and heterogeneous fertility datasets. Nevertheless, logistic regression is remaining as an important baseline model for comparing the performance of more advanced AI techniques.

5.3 Decision Tree

Decision tree algorithms represent predictive decisions through the hierarchical structure consisting of nodes, branches, and leaves. Each internal node is a decision, which is based on a specific feature, while the final leaf node is the predicted outcome [9].

The decision process can be expressed as:

$$Decision = f(x_1, x_2, \dots, x_n)$$

where the algorithm selects important features to divide the dataset into groups with the similar outcomes.

In live birth prediction, a decision tree tends to first divide patients, which is based on maternal age, followed by embryo quality, hormone levels, or previous treatment history. For example, patients with the favorable embryo quality and adequate ovarian reserve tends to be classified into a higher probability group.

The splitting process is commonly, which is based on impurity measures such as entropy:

$$Entropy = - \sum_{i=1}^C p_i \log_2(p_i)$$

where p_i is the probability of each class.

Another commonly used criterion is the Gini index:

$$Gini = 1 - \sum_{i=1}^C p_i^2$$

Decision trees are providing clear visualization and interpretation, which is making them that is considered suitable for clinical decision-support applications. However, individual decision trees tend to suffer from the overfitting, especially when is trained on small fertility datasets.

An overly complex tree tends to memorize training data rather than learning general patterns. Therefore, ensemble approaches such as Random Forest and Gradient Boosting are developed in improving the stability and predictive performance.

5.4 Random Forest

Random Forest is an ensemble ML technique that is combining multiple decision trees in improving the prediction accuracy and it is reducing overfitting. Instead of relying on a single decision tree, Random Forest generates many independent trees using different subsets of data and features [10].

The final prediction is obtained through the aggregation:

$$Y^{\wedge} = Mode(T1, T2, \dots, Tn)$$

where Tn is individual decision trees.

For probability prediction, the output can be represented as:

$$P(Y) = \frac{1}{N} \sum_{i=1}^N P_i(Y)$$

where $P_i(Y)$ is the prediction probability from the each tree.

Random Forest is extensively applied in fertility prediction because it can handle nonlinear relationships and interactions among the multiple clinical variables. It can identify important predictors such as maternal age, AMH level, embryo quality, and treatment history.

A major advantage of Random Forest is its ability to estimate feature importance. The algorithm can rank variables in accordance with the contribution to prediction performance. This capability helps researchers identify critical reproductive factors.

However, Random Forest models are requiring large computational resources when hundreds or thousands of trees are used. in addition, although more interpretable than deep learning models, understanding the combined decision process of many trees can still be challenging.

5.5 Support vector Machine (SVM)

Support Vector Machine is a supervised learning algorithm designed to classify data by finding an optimal decision boundary between the different classes. in live birth prediction, SVM separates successful and unsuccessful treatment outcomes, which is based on clinical and biological characteristics [11].

The decision function of SVM is:

$$f(x) = wTx + b$$

where w is the weight vector b and it is the bias term.

The objective of SVM is to maximize the margin between the classes:

$$Margin = \frac{2}{||w||}$$

A larger margin generally is improving classification robustness.

SVM can handle nonlinear relationships through the kernel functions. Common kernels include:

$$K(x, x_i) = e^{-\gamma ||x_i - x_j||^2}$$

where the radial basis function kernel maps input features into higher-dimensional space.

In fertility prediction, SVM is applied to classify IVF outcomes using hormonal measurements, embryo characteristics, and patient information. Its ability in performing a well with the limited datasets makes it attractive for medical applications where large labelled datasets tend not to always be available.

However, SVM performance depends strongly on parameter selection and feature scaling. in addition, interpreting the decision process is more difficult when it is compared with the simpler models.

5.6 Naïve Bayes

Naïve Bayes is a probabilistic ML algorithm, which is based on Bayes' theorem. It predicts outcomes by estimating the probability of a class, which is based on observed features.

The mathematical formulation is:

$$P(Y | X) = \frac{P(X | Y)P(Y)}{P(X)}$$

where:

$P(Y | X)$ is the probability of live birth given observed features,

$P(X | Y)$ is the likelihood of observing features,

$P(Y)$ is prior probability.

However, because reproductive variables often interact strongly with the each other, the independence assumption are limiting its ability to model complex biological relationships.

5.7K-Nearest Neighbors (KNN)

K-Nearest Neighbors is a distance-based ML algorithm that predicts outcomes by comparing new cases with the similar historical examples.

The distance between the samples is commonly calculated using Euclidean distance:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

The predicted outcome is determined by the majority class among the nearest neighbors.

In live birth prediction, KNN is identifying patients with the similar reproductive characteristics and is estimating the likelihood of success, which is based on previous outcomes. For example, a patient with the similar age, hormone profile, and embryo characteristics tend to are a comparable predicted outcome.

The simplicity of KNN makes it easy to implement, but its performance decreases when datasets contain many irrelevant variables. in addition, prediction is becoming computationally expensive for large fertility databases.

5.8Gradient Boosting Methods

Gradient boosting algorithms are become highly effective approaches for fertility prediction because it is combining multiple weak learners to create a powerful predictive model. Unlike Random Forest, which builds trees independently, gradient boosting creates trees sequentially, where each new model attempts to correct errors from the previous models.

The optimization process can be expressed as:

$$F_m(x) = F_{m-1}(x) + \eta h_m(x)$$

where $F_m(x)$ is the updated model, $h_m(x)$ is the new weak learner, η and it is the learning rate.

5.8.1XGBoost

Extreme Gradient Boosting (XGBoost) is an optimized gradient boosting algorithm widely used in medical prediction tasks. The objective function is:

$$Obj = \sum L(y_i, \hat{y}_i) + \sum \Omega(f_k)$$

where the first term is prediction error and the second term controls model complexity.

5.8.2LightGBM

LightGBM is that is considered suitable for fertility databases containing thousands of treatment cycles because it is reducing the computational requirements while it is maintaining high prediction accuracy.

5.8.3CatBoost

CatBoost is another gradient boosting algorithm designed to handle categorical variables effectively. in fertility prediction, categorical information such as infertility type, treatment protocol, and embryo grading categories can be directly incorporated.

CatBoost is reducing the preprocessing requirements and is improving prediction stability when dealing with the mixed clinical datasets.

5.9Ensemble Learning Approaches

Ensemble learning is combining multiple ML models to achieve improved prediction performance. Since different algorithms capture different aspects of reproductive data, combining them often produces more reliable predictions.

A general ensemble model can be represented as:

$$F(x) = \sum_{i=1}^n w_i f_i(x)$$

where $f_i(x)$ is individual models w_i and it is its weights.

Hybrid models combining Random Forest, XGBoost, SVM, and neural networks are showing improved performance in live birth prediction. Ensemble methods are reducing individual model limitations and it is improving the robustness against the variations in clinical datasets.

6.Deep Learning Techniques for Live Birth Prediction

6.1Introduction to Deep Learning in Reproductive Outcome Prediction

Deep learning (DL) is an advanced branch of artificial intelligence that uses multilayer neural network architectures to automatically learn complex patterns from the large-scale datasets. Unlike conventional ML approaches that often are requiring manual feature extraction, deep learning models can independently identify relevant features from the raw and high-dimensional data.

The increasing availability of fertility-related datasets, including electronic health records, IVF treatment databases, embryo time-lapse images, and genomic information, is accelerated the adoption of deep learning models. These models can be combining the heterogeneous information sources and identify hidden relationships that tends not to be captured using conventional statistical approaches.

The fundamental structure of a deep learning model is consisting of multiple interconnected layers. Each layer is transforming the input data into increasingly abstract representations. The output prediction is generated by combining these learned representations.

The mathematical representation of a neural network can be expressed as:

$$hl = f(Wlhl-1 + bl)$$

where hl is the output of layer l , Wl is the weight matrix, bl is the bias term, f and it is the activation function.

For live birth prediction, the deep learning model is learning a mapping between the reproductive characteristics and treatment outcomes:

$$P(LB) = F(X; \theta)$$

where $P(LB)$ is the probability of live birth, X is the input features, θ and it is the learned model parameters.

The optimization process is aiming to minimize the difference between the predicted and actual outcomes:

$$Loss = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

where y_i is the actual outcome, $\hat{y}_i$ and it is the predicted probability.

6.2 Artificial Neural Networks (ANN)

Artificial Neural Network (ANN) is one of the fundamental deep learning architectures inspired by the biological structure of the human brain [12]. The mathematical operation performed by an artificial neuron is:

$$z = \sum_{i=1}^n w_i x_i + b$$

where:

x_i is input features,

w_i is connection weights,

b is bias, and

z is the neuron output.

An activation function is transforming the output: $a = f(z)$, where the ReLU function is expressed as: $f(x) = \max(0, x)$. NN-based fertility prediction systems are applied to estimate IVF success probability using variables such as maternal age, AMH level, embryo quality, ovarian response, and previous treatment history. These models can capture nonlinear interactions between the reproductive factors, which is providing an advantage over conventional regression models.

For example, maternal age alone tends not to be determining the treatment success, but its interaction with the embryo quality, hormone levels, and ovarian reserve tends to be significantly influencing the final outcome. ANN models can learn these complex relationships automatically.

The major advantage of ANN is its flexibility and ability to model complicated biological systems. However, ANN models are requiring large datasets for effective training and tends to suffer from the overfitting when it is applied to limited fertility datasets. In addition, its internal decision-making process is difficult to interpret, which is encouraged the integration of explainable AI techniques.

6.3 Deep Neural Networks (DNN)

Deep Neural Networks (DNNs) extend conventional ANN architectures by incorporating multiple hidden layers. The increasing depth is allowing DNN models to learn hierarchical representations of complex data. In reproductive medicine, DNNs are increasingly used for predicting live birth outcomes because fertility-related variables often are showing complex dependencies that cannot be represented using shallow models [13].

A DNN is performing multiple nonlinear transformations:

$$h_1 = f(W_1X + b_1) \quad h_2 = f(W_2h_1 + b_2)$$

$$Y = f(W_nh_n + b_n)$$

where each hidden layer is extracting progressively higher-level information.

In IVF prediction, the first layers tend to capture basic relationships between the patient characteristics and treatment parameters, while deeper layers identify complex interactions associated with the successful embryo development and implantation.

DNN models are applied to large fertility datasets containing thousands of treatment cycles. These models can be combining the multiple feature categories, including:

- ☐ Clinical records
- ☐ Hormone measurements
- ☐ Embryo characteristics
- ☐ Treatment protocols
- ☐ Pregnancy outcomes

The ability of DNN models to automatically learn feature representations is reducing the dependence on manual feature engineering. This is particularly useful when dealing with the complex biological data where important predictors tend not to be immediately obvious.

However, DNN models are requiring careful optimization of parameters such as learning rate, network depth, and regularization methods. Insufficient training data or inappropriate parameter selection tends to result in poor generalization performance.

6.4 Convolutional Neural Networks (CNN)

CNNs are specialized deep learning architectures primarily designed for image processing and computer vision applications. However, these assessments tend to vary between the embryologists. CNN-based systems are providing automated and quantitative embryo evaluation by extracting visual features directly from the images [14].

The convolution operation in CNN can be expressed as:

$$Y(i, j) = \sum_m \sum_n X(i + m, j + n) K(m, n)$$

where:

X is the input image,

K is the convolution filter, and

Y is the extracted feature map.

CNN architectures typically consist of convolution layers, pooling layers, fully connected layers, and classification layers. Convolution layers identify local patterns, pooling layers are reducing dimensional complexity, and fully connected layers are generating final predictions.

In embryo assessment, CNN models analyze:

- ☐ Embryo morphology
- ☐ Cell structure
- ☐ Blastocyst expansion
- ☐ Developmental timing
- ☐ Implantation potential

The extracted image features are then combined with the clinical information to predict implantation probability or live birth outcome.

A multimodal CNN-based prediction model can be represented as:

$$P(LB) = CNN(I) + ANN(C)$$

where:

I is embryo images C and it is clinical information.

Recent studies are showing that CNN-based systems can achieve high accuracy in embryo grading and implantation prediction. These models are reducing observer variability and it is providing consistent evaluation.

6.5 Recurrent Neural Networks (RNN)

In reproductive medicine, many biological processes involve sequential changes, such as embryo development over time, hormonal fluctuations during treatment cycles, and repeated clinical observations [15]. The hidden state of an RNN can be represented as:

$$h_t = f(W_h h_{t-1} + W_x x_t + b)$$

where:

h_t is the current hidden state,

h_{t-1} is previous information, and

x_t is current input data.

RNN-based systems can analyze embryo development sequences captured through the time-lapse imaging. Instead of evaluating a single image, the model is learning developmental patterns across the multiple time points. For example, embryo division timing, cell transition patterns, and morphological changes are providing important information regarding implantation potential. Although RNNs are providing advantages for sequential analysis, it is suffering from the vanishing gradient problem when learning long-term dependencies. This limitation led to the development of advanced recurrent architectures such as LSTM and GRU.

6.6 Long Short-Term Memory Networks (LSTM)

Long Short-Term Memory networks are an advanced form of RNN designed to overcome limitations associated with the long-term dependency learning. LSTM networks use memory cells and gating mechanisms to selectively retain or discard information [16]. The LSTM architecture contains three primary gates: input gate, forget gate, and output gate.

The forget gate is represented as:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

The input gate is:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

The cell state update is:

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t$$

where C_t is the updated memory state.

6.7 Gated Recurrent Unit (GRU)

Gated Recurrent Unit (GRU) is another recurrent architecture designed in improving the sequential learning efficiency. when it is compared with the LSTM, GRU uses fewer gates and is a simpler structure, which is resulting in faster training [17].

The update gate is represented as:

$$z_t = \sigma(W_z[h_{t-1}, x_t])$$

The reset gate is:

$$r_t = \sigma(W_r[h_{t-1}, x_t])$$

GRU is combining memory control and hidden state updates in capturing important temporal information. In fertility prediction, GRU models are applied for sequential clinical data analysis, including hormone changes, treatment cycle information, and embryo development patterns. The advantage of GRU is reduced computational complexity when it is compared with the LSTM while it is maintaining comparable prediction performance. This makes GRU that is considered suitable for real-time clinical decision-support applications.

6.8 Transformer-Based Architectures

Transformer models represent one of most of the recent developments in deep learning. Unlike RNN-based models, transformers are relying on attention mechanisms in capturing relationships between the different parts of input data without sequential processing [18]. The attention mechanism is defined as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where:

Q is query,

K is key,

V is value, and

d_k is scaling factor.

Transformers are significant advantages in analyzing multimodal fertility data because it can identify relationships between the distant features. For example, a transformer model can simultaneously analyze maternal characteristics, embryo development information, and genetic indicators. Vision Transformers (ViTs) are introducing for embryo image analysis. Instead of processing images through the convolution operations, ViTs divide images into patches and apply attention mechanisms to identify important visual patterns.

Transformer-based fertility prediction systems are showing improved performance because of its ability to model complex interactions among the multiple data sources. However, it are requiring large datasets and significant computational resources.

6.9 Multimodal Deep Learning Frameworks

Multimodal deep learning is combining multiple data types within the unified predictive framework. in reproductive medicine, multimodal systems are combining the clinical records, embryo images, hormonal

information, and genetic data [19].

A multimodal model can be represented as:

$$P(LB) = F(C, E, G, H)$$

where:

C is clinical data,

E is embryo information,

G is genetic information, and

H is hormonal data.

These systems are providing a more comprehensive representation of reproductive status when it is compared with the single-source models.

7. Embryo Assessment and Computer Vision Systems

7.1 Introduction to AI-Based Embryo Assessment

Embryo assessment is one of most of the critical stages in assisted reproductive technology (ART) because embryo quality directly influences implantation probability and eventual live birth success. Traditionally, embryologists evaluate embryos, which is based on morphological characteristics observed through the microscopic examination. These characteristics include cell number, fragmentation level, symmetry, blastocyst expansion, inner cell mass quality, and trophectoderm appearance. Although expert embryologists are performing a reliable assessment, manual evaluation is involving the subjective variations between the observers.

The emergence of artificial intelligence (AI), especially computer vision and deep learning techniques, is introducing automated embryo assessment systems that is capable of analyzing complex visual patterns. These systems is extracting quantitative features from the embryo images and time-lapse videos that tends not to be easily identified through the conventional assessment methods [20]. By learning from the large collections of embryo images with the known clinical outcomes, AI models can predict embryo viability, implantation potential, and live birth probability.

The general objective of an AI-based embryo assessment system is to learn a relationship between the embryo characteristics and reproductive outcomes:

$$P(LB) = f(I, E, C)$$

where:

$P(LB)$ is the probability of live birth,

I is embryo imaging information,

E is embryo developmental characteristics, and

C is clinical parameters.

The integration of computer vision into fertility treatment is shifted embryo evaluation from the subjective visual grading towards a quantitative and reproducible analysis. These systems

are providing fertility specialists with the additional decision-support information, which is allowing more informed embryo selection and improved treatment planning.

7.2 Automated Embryo Grading Systems

Embryo grading is a standard procedure used by fertility specialists to evaluate developmental quality before embryo transfer [21]. The CNN feature extraction process can be represented as:

$$F_i = \text{Conv}(I, K_i)$$

where:

I is the embryo image,

K_i is convolution filters, and

F_i is extracted image features.

These extracted features are processed through the classification layers:

$$P(Q) = \text{Softmax}(WF + b)$$

where $P(Q)$ is embryo quality probability.

AI-based grading systems analyze characteristics including:

- ☐ Cell division patterns
- ☐ Embryo symmetry
- ☐ Fragmentation regions
- ☐ Blastocyst expansion
- ☐ Cell mass distribution

Unlike manual grading, AI systems are providing consistent evaluation across the different cases. it can also identify subtle morphological characteristics associated with the implantation potential.

Recent AI-based embryo grading frameworks are showing improved reproducibility when it is compared with the conventional scoring methods. These systems are particularly useful in high-volume fertility centers where thousands of embryos are requiring evaluation.

However, automated grading systems still are facing challenges. Embryo images vary significantly because of

differences in microscope settings, imaging resolution, culture conditions, and laboratory protocols. Therefore, models trained on one dataset tend not to always be performing effectively in another clinical environment.

7.3 Time-Lapse Imaging Systems

Time-lapse imaging technology is significantly improved embryo monitoring by capturing developmental changes throughout embryo culture [22]. Unlike conventional embryo evaluation, where embryos are periodically removed from the incubators for observation, time-lapse systems record uninterrupted developmental sequences without disturbing the culture environment.

The temporal information obtained from the time-lapse imaging is providing valuable insights into embryo viability. Important developmental events include:

- Time of first cell division
- Cleavage intervals
- Blastocyst formation timing
- Abnormal division patterns

AI models analyze these temporal patterns to identify embryos with the higher implantation potential.

The developmental sequence can be represented as:

$$E(t) = \{e1, e2, e3, \dots, en\}$$

where $E(t)$ is embryo development over time en and it is individual developmental stages.

Deep learning models such as recurrent neural networks (RNN), long short-term memory networks (LSTM), and transformers are that is considered suitable for analyzing time-dependent embryo information.

A temporal prediction model can be expressed as:

$$P(LB) = f(E(t1), E(t2), \dots, E(tn))$$

where the model is learning relationships between the developmental patterns and final reproductive outcomes.

Time-lapse AI systems are providing advantages because it is considering the embryo dynamics rather than evaluating a single developmental snapshot. For example, two embryos tend to appear similar at a specific time point, but its developmental trajectories tend to differ significantly. AI models can detect these differences, and it is improving the embryo selection accuracy.

7.4 CNN-Based Embryo Quality Assessment

Convolutional Neural Networks are become most of the widely used deep learning approach for embryo image analysis. CNNs are designed to automatically learn spatial features from the images through the convolution operations, which is making them highly effective for microscopic embryo analysis [23].

The CNN architecture generally is consisting of convolution layers, activation functions, pooling layers, and classification layers. The convolution operation is extracting local image patterns:

$$Y(i, j) = \sum_m \sum_n X(i + m, j + n) K(m, n)$$

where X is the input image K and it is the convolution kernel. CNN models are applied for:

- Embryo implantation prediction
- Blastocyst quality classification
- Morphological grading
- Pregnancy outcome prediction

The major advantage of CNN-based systems is its ability to eliminate manual feature engineering. Instead of researchers manually defining embryo characteristics, the network automatically discovers important visual representations.

However, CNN models are requiring large numbers of labelled embryo images for training. Since high-quality fertility datasets are difficult to obtain because of privacy concerns and limited clinical availability, data scarcity is remaining as a major challenge.

7.5 Vision Transformers for Embryo Analysis

Vision Transformers (ViTs) represent a recent advancement in computer vision that applies transformer architectures to image analysis. Unlike CNNs, which are relying on local convolution operations, transformers use attention mechanisms to identify relationships between the different image regions [24].

The self-attention mechanism is expressed as:

$$Attention(Q, K, V) = \frac{QK^T}{\sqrt{d_k}} \text{Softmax}(\text{---})V$$

Vision transformers divide embryo images into smaller patches and analyze relationships between the these patches. This approach is allowing the model in capturing global structural information that tends to be missed by conventional CNN approaches.

In embryo assessment, vision transformers can analyze:

- Overall embryo morphology
- Spatial relationships between the cells

- Developmental abnormalities
- Subtle visual indicators of viability

The advantage of transformer-based models is its ability in capturing long-range dependencies within the images. This capability is valuable because embryo quality is not determined by a single region but by the entire organization and development pattern.

However, transformer models typically are requiring larger datasets and higher computational resources when it is compared with the CNN-based approaches. Therefore, hybrid CNN-transformer architectures are increasingly being explored in combining local feature extraction with the global attention learning.

7.6 Image-Based Live Birth Prediction

Image-based live birth prediction is aiming to directly estimate the probability of successful delivery using embryo imaging information. Instead of predicting only embryo quality, these models are establishing a relationship between the embryo appearance and final clinical outcome [25].

The prediction framework can be represented as:

$$P(LB) = f(\text{ImageFeatures}, \text{ClinicalFeatures})$$

where image features are extracted from the embryo images and clinical features include patient-related variables.

Combining image data with the clinical information is improving prediction performance because live birth depends on both the embryo competence and maternal reproductive conditions.

8. Explainable and Trustworthy AI for Live Birth Prediction

8.1 Importance of Explainability in Fertility Care

Although AI models are showing high prediction performance, its clinical adoption depends strongly on transparency and interpretability. Many advanced deep learning models is operating as black-box systems where the internal reasoning process is difficult to understand. In reproductive healthcare, this limitation is particularly important because treatment decisions influence patient outcomes and are requiring clinical justification.

Explainable artificial intelligence (XAI) that is focussing on developing methods that reveal how AI models are generating predictions. Instead of providing only a probability score, explainable models identify the factors responsible for the prediction.

The prediction process can be expressed as:

$$Y = f(X)$$

where XAI attempts to determine:

$$\text{Contribution}(x_i) = \frac{\partial f(X)}{\partial x_i}$$

which is the influence of individual features.

In live birth prediction, explainability is allowing clinicians to understand whether predictions are influenced by maternal age, hormone levels, embryo quality, genetic factors, or previous treatment history.

8.2 SHAP-Based Interpretation

SHAP (SHapley Additive exPlanations) is one of most of the widely used explainable AI techniques for interpreting ML and deep learning models. It is, which is based on cooperative game theory, where each feature is considered a contributor to the final prediction.

The SHAP value for a feature can be represented as:

$$\phi_i = \sum_{S \subseteq F} \frac{|S|! (M - |S| - 1)!}{M!} |f(S \cup i) - f(S)|$$

where:

ϕ_i is the contribution of feature i ,

S is feature subsets, and

M is total number of features.

In fertility prediction systems, SHAP analysis can identify which variables contribute positively or negatively to live birth probability.

For example, a model tends to is determining the that embryo quality and AMH levels strongly increase predicted success, while advanced maternal age decreases probability.

SHAP is providing both the global and individual explanations. Global explanations identify important predictors across the population, while individual explanations describe why a specific patient received a particular prediction.

8.3 LIME-Based Interpretation

Local Interpretable Model-Agnostic Explanations (LIME) is another popular explainability approach. Unlike SHAP, which considers entire feature contributions, LIME that is focussing on explaining individual predictions by creating a simplified local model around a specific case.

The process is involving the approximating the original complex model:

$$f(x) \approx g(x)$$

where:

$f(x)$ is the original AI model $g(x)$ and it is the interpretable local model.

In fertility prediction, LIME can explain why a patient is a lower predicted probability of live birth. For instance, it tends to be showing that reduced ovarian reserve, previous IVF failures, or poor embryo characteristics contributed to the prediction.

The advantage of LIME is its model-independent nature, which is allowing explanation of different ML architectures. However, because it that is focussing on local approximations, explanations tend to vary depending on the selected neighborhood.

8.4 Feature Importance Analysis

Feature importance analysis is identifying the relative contribution of individual variables within the AI prediction models. This approach is commonly used with the tree-based models such as Random Forest, XGBoost, and LightGBM.

The importance score can be represented as:

$$Importance_i = \Delta Error_i / TotalError$$

where $\Delta Error_i$ is the improvement obtained by including feature i .

In live birth prediction, feature importance analysis commonly is identifying variables such as maternal age, embryo quality, AMH concentration, fertilization rate, and endometrial thickness as influential predictors.

This information helps researchers understand reproductive mechanisms and supports clinical decision-making.

9. Publicly Available Datasets and Data Sources

9.1 Introduction to Fertility Datasets for AI-Based Prediction

In reproductive medicine, AI models are requiring large volumes of accurately labelled data containing patient characteristics, treatment information, embryo observations, and final clinical outcomes. Unlike many other medical domains where standardized datasets are widely available, fertility datasets are relatively limited because reproductive information contains highly sensitive personal and medical records.

A fertility prediction dataset generally is consisting of multiple categories of information collected during the ART process. These include demographic characteristics, reproductive history, hormonal measurements, laboratory parameters, embryo development information,

imaging data, genetic screening results, and pregnancy outcomes. The objective of an AI model is to learn the relationship between the input variables and the final live birth outcome.

The mathematical representation of a fertility dataset can be expressed as:

$$D = \{(X_i, Y_i)\}_N$$

where D is the complete dataset, X_i is the feature vector of the i th patient or treatment cycle, Y_i is the observed outcome, N and it is the total number of samples.

For live birth prediction, the feature vector can be represented as:

$$X_i = \{D_i, C_i, H_i, E_i, G_i, L_i\}$$

where D_i is demographic features, C_i is clinical characteristics, H_i is hormonal information, E_i is embryological information, G_i is genetic information, L_i and it is lifestyle-related variables. The availability of high-quality datasets determines whether AI models can achieve reliable generalization.

9.2 IVF and ART Registries

The advantage of ART registries is its large sample size and clinical relevance. Since the data are collected during real-world treatment procedures, AI models trained using these datasets can better represent practical fertility scenarios.

However, ART registry datasets also present several challenges. Different fertility centers tend to follow different stimulation protocols, embryo grading criteria, and laboratory procedures. These variations are introducing data heterogeneity that tends to be reducing model performance.

The generalization capability of a model trained using registry data can be represented as:

$$G = f(D_{train}, D_{test})$$

where D_{train} is the training dataset D_{test} and it is an independent clinical dataset.

A model is achieving strong generalization when performance is remaining as consistent across the different

patient populations and healthcare environments.

9.3 Fertility Clinic Databases

Individual fertility clinics maintain detailed databases containing patient-specific treatment information. These databases often include more comprehensive information when it is compared with the national registries because it capture detailed laboratory observations and treatment procedures.

Clinic-level datasets tend to contain embryo images, time-lapse videos, hormone profiles, medication details, and embryologist assessments. Such detailed information is allowing advanced AI applications, especially deep learning-based embryo evaluation.

For example, an embryo imaging dataset tends to contain:

$$E = \{I_1, I_2, \dots, I_n\}$$

where each I_n is an embryo image captured during different developmental stages.

AI models analyze these images to predict embryo viability and implantation potential.

The primary advantage of fertility clinic databases is its high-resolution information. However, its major limitation is limited population diversity. A dataset collected from the single clinic tends to represent specific demographic characteristics, treatment protocols, and laboratory practices. As a result, models developed using single-center data tends to are performing a poorly when it is applied to other institutions.

To overcome this limitation, researchers increasingly focus on multicenter datasets and collaborative learning approaches.

9.4 Electronic Health Records

Electronic Health Records (EHRs) are providing comprehensive patient information collected throughout healthcare interactions. in reproductive medicine, EHRs contain valuable clinical information including medical history, previous diagnoses, medications, laboratory tests, and pregnancy-related information.

10. Performance Evaluation Metrics

10.1 Introduction to Evaluation Metrics

Since fertility prediction is commonly formulated as a binary classification problem, models should distinguish between the successful and unsuccessful outcomes accurately.

An AI model produces four possible prediction outcomes:

	<i>Predicted Positive</i>	<i>Predicted Negative</i>
<i>Actual Positive</i>	<i>TP</i>	<i>FN</i>
<i>Actual Negative</i>	<i>FP</i>	<i>TN</i>

where TP is correctly predicted live birth cases, TN is correctly predicted unsuccessful outcomes, FP is incorrect positive predictions, and FN is missed successful outcomes.

Different evaluation metrics are providing different perspectives regarding model performance. Accuracy alone is insufficient because fertility datasets often contain class imbalance. Therefore, multiple evaluation measures are required.

10.2 Accuracy

Accuracy is the entire percentage of correctly classified cases.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy is commonly reported in fertility prediction studies because it is providing a general indication of model performance.

However, accuracy can be misleading when datasets are imbalanced. For example, a model tends to achieve high accuracy by predicting the majority class while performing poorly on successful live birth cases.

Therefore, accuracy should always be interpreted together with the other evaluation metrics.

10.3 Precision

Precision measures the proportion of correctly predicted positive outcomes among the all positive predictions.

$$\text{Precision} = \frac{TP}{TP + FP}$$

In live birth prediction, precision is indicating how reliable the model is when predicting successful treatment outcomes.

A high precision value means that patients predicted to achieve live birth are likely to actually experience successful outcomes.

Precision is particularly important when incorrect positive predictions tend to lead to unrealistic patient expectations or inappropriate treatment decisions.

10.4 Recall

Recall, also known as sensitivity, measures the ability of a model to identify actual successful live birth cases.

$$\text{Recall} = \frac{TP}{TP + FN}$$

A high recall value is indicating that the model successfully is identifying most of the patients who achieve successful outcomes.

In fertility prediction, recall is important because missing potential successful cases tend to be preventing clinicians from the identifying suitable treatment opportunities.

10.5 F1-Score

The F1-score is combining precision and recall into a single performance measurement.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The F1-score is particularly useful for imbalanced fertility datasets because it considers both the false positive and false negative errors.

A higher F1-score is indicating better balance between the identifying successful cases and avoiding incorrect predictions.

10.6 Receiver Operating Characteristic Curve and AUC

The Receiver Operating Characteristic (ROC) curve is evaluating model performance across the different classification thresholds. It plots sensitivity against the false positive rate.

The false positive rate is calculated as:

$$FPR = \frac{FP}{FP + TN}$$

The area under the ROC curve (AUC) is entire discrimination capability.

$$AUC = \int_0^1 TPR(FPR) dFPR$$

where TPR is true positive rate.

An AUC value close to one is indicating excellent predictive capability, while a value near 0.5 is indicating random prediction.

In live birth prediction studies, AUC is frequently used because it is evaluating model performance independent of a specific classification threshold.

10.7 Sensitivity and Specificity

Sensitivity measures the ability to correctly identify successful live birth cases:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

Specificity measures the ability to correctly identify unsuccessful cases:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Both metrics are important because fertility prediction is requiring accurate identification of both the high-probability and low-probability cases.

10.8 Calibration Metrics

Calibration is evaluating whether predicted probabilities correspond accurately to actual outcomes. A well-calibrated model should produce predictions that match observed frequencies.

For example, among the patients predicted with the 80% live birth probability, approximately 80% should achieve successful outcomes.

Calibration error can be represented as:

$$CE = \frac{1}{N} \sum_{i=1}^N |p_i - y_i|$$

where p_i is predicted probability and y_i is actual outcome.

Calibration is especially important in clinical decision Support because probability is estimating influence patient counselling and treatment planning.

10.9 Clinical Utility Measures

Beyond statistical performance, AI models should be showing clinical usefulness. Clinical utility measures evaluate whether implementing a prediction model is improving decision-making when it is compared with the Conventional approaches.

Decision curve analysis is commonly used to evaluate whether model predictions are providing meaningful benefits.

The net benefit can be expressed as:

$$NB = \frac{TP}{N} - \frac{FP}{N} \times \frac{p_t}{1 - p_t}$$

A clinically useful AI system should not only achieve high predictive accuracy but also be improving the treatment planning, are reducing unnecessary procedures, and Support personalized reproductive care.

11. Comparative Survey of Recent Studies (2020–2026)

11.1 Survey Methodology

The reviewed studies primarily focus on predicting IVF success, embryo implantation probability, pregnancy outcomes, and live birth rates. The survey considers studies that utilize demographic information, clinical parameters, hormonal measurements, embryo morphology, time-lapse imaging, genetic information, and multimodal reproductive datasets. The selected research works are analyzed, which is based on its AI methodology, dataset characteristics, prediction objectives, and reported performance metrics.

11.2 Study Selection Criteria

The selection of research studies was, which is based on methodological relevance, publication period, dataset availability, and contribution to AI-based live birth prediction. Studies focusing only on conventional fertility analysis without computational prediction methods were excluded because it is failing to contribute directly to AI-driven decision Support systems.

11.3 Comparison of Datasets

The performance of AI-based fertility prediction models depends significantly on dataset characteristics. Recent studies are using different data sources, including IVF clinic databases, ART registries, embryo imaging repositories, and electronic health records. The diversity of datasets directly is affecting the generalization ability of predictive models.

Clinical datasets generally contain demographic variables, hormonal measurements, infertility history, stimulation protocols, and treatment outcomes. Imaging datasets focus on embryo morphology and time-lapse development patterns. More advanced studies are combining multiple data modalities in improving the prediction reliability.

Table 1 is presenting a comparative analysis of datasets used in recent AI-based live birth prediction studies.

Study Period	Dataset Type	Data Characteristics	Number of Samples	Prediction Target
2020	IVF Clinical Dataset	Age, hormone levels, treatment history, embryo parameters	5,000+ cycles	Pregnancy and Live birth prediction
2021	ART Registry Dataset	Demographic, clinical, and laboratory variables	20,000+ cycles	IVF outcome classification
2022	Embryo Imaging Dataset	Time-lapse embryo images and morphology features	10,000+ embryos	Embryo viability prediction

2023	Multicenter IVF Dataset	Clinical and embryological parameters	30,000+ treatment cycles	Live birth prediction
2024	Multimodal Fertility Dataset	Clinical, imaging, hormonal, and genetic information	50,000+ records	Personalized ART outcome prediction
2025–2026	AI combined Fertility Database	Longitudinal patient records and embryo data	100,000+ observations	Intelligent decision support

The analysis is indicating that recent studies increasingly use larger and more diverse datasets. Multimodal datasets is showing improved prediction capability because it represent multiple biological aspects of reproduction rather than relying on individual variables.

11.4 Comparison of Predictive Features

Feature selection plays an essential role in AI-based fertility prediction because reproductive outcomes depend on complex interactions among the multiple factors. Recent studies are identified maternal age, AMH concentration, embryo quality, fertilization rate, endometrial thickness, and previous IVF history as frequently occurring predictive variables.

Table 2 is presenting commonly used predictive features identified in recent studies.

Feature Category	Important Predictive Variables	Influence on Live Birth Prediction
Demographic Factors	Maternal age, paternal age, BMI	Determines reproductive potential and pregnancy risk
Clinical Factors	Infertility duration, previous IVF attempts, reproductive disorders	is patient-specific fertility history
Hormonal Factors	AMH, FSH, LH, estradiol, progesterone	Indicates ovarian response and hormonal balance
Embryological Factors	Embryo grade, blastocyst quality, cell division timing	is embryo implantation capability
Genetic Factors	Chromosomal screening, embryo genetic profile	Determines embryo viability
Lifestyle Factors	Smoking, nutrition, physical activity	Influences reproductive health

The comparative analysis is showing that single-factor prediction models generally are providing limited accuracy because live birth is influenced by multiple interacting variables. Recent AI frameworks therefore emphasize multimodal feature integration.

11.5 Comparison of AI Models

Recent studies is showing significant improvement in predictive performance through the advanced AI algorithms. conventional models such as logistic regression tends to remain important because of its interpretability, but ensemble and deep learning methods generally achieve better prediction accuracy.

Table 3 is summarizing the performance comparison of different AI approaches reported in recent studies.

AI Technique	Primary Application	Accuracy Range (%)	AUC Range
Logistic Regression	Clinical risk prediction	62–78	0.70–0.82
Decision Tree	Treatment outcome classification	68–80	0.72–0.84
Random Forest	Feature-based IVF prediction	75–88	0.80–0.91
Support Vector Machine	Small and medium fertility datasets	72–86	0.78–0.89

XGBoost	ART outcome prediction	82-91	0.85-0.94
Artificial Neural Network	Patient-based prediction	80-90	0.86-0.93
CNN	Embryo image analysis	85-94	0.90-0.96
CNN-LSTM Hybrid	Time-lapse embryo analysis	88-95	0.92-0.97
Transformer Models	Multimodal fertility prediction	90-96	0.94-0.98
Decision Tree	Treatment outcome classification	68-80	0.72-0.84

The results is showing that deep learning and hybrid AI approaches achieve superior performance when it is compared with the conventional ML methods. However, improved accuracy is often associated with the higher computational requirements and reduced interpretability.

11.6 Comparison of Performance Metrics

Evaluation metrics used in recent studies vary depending on the prediction objective and dataset characteristics. Accuracy is commonly reported for balanced datasets, while AUC, sensitivity, specificity, and F1-score are preferred for clinically imbalanced fertility datasets.

Table 4 is providing a comparative summary of reported evaluation metrics.

Model Category	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
Statistical Models	62-78	65-80	60-76	63-78	0.70-0.82
ML Models	75-90	78-91	74-89	76-90	0.82-0.94
Deep Learning Models	85-95	87-96	85-95	86-95	0.90-0.97
Explainable Hybrid Models	88-96	90-97	88-96	89-96	0.92-0.98

The comparative results are showing that AI-based methods are improving the prediction capability. Explainable hybrid models is showing competitive performance while also providing transparency, which is making them more that is considered suitable for clinical deployment.

11.7 Key Findings from the Literature

The comparative survey of studies from the 2020 to 2026 is showing several important trends in AI-based live birth prediction. The first major observation is the increasing shift from the conventional statistical models towards a advanced ML and deep learning frameworks. While logistic regression models are providing interpretability, its ability in capturing complex reproductive relationships is limited.

Ensemble learning approaches such as Random Forest, XGBoost, LightGBM, and CatBoost are providing improved prediction performance because it effectively model nonlinear interactions among the fertility-related variables. Deep learning approaches is improving prediction accuracy by automatically extracting complex representations from the clinical and imaging data.

The survey also is indicating that embryo-related information is become increasingly important with the development of computer vision systems. CNN-based and transformer-based models are showing strong performance in embryo quality assessment and implantation prediction.

Another significant finding is the growing importance of explainable AI. Although deep learning models achieve high predictive accuracy, clinical acceptance is requiring transparent decision-making mechanisms. Integration of SHAP, LIME, and feature importance analysis is improving trust by identifying the contribution of individual predictive factors.

References

1. Vander Borgh, M., & Wyns, C. (2018). Fertility and infertility: Definition and epidemiology. *Clinical biochemistry*, 62, 2-10.
2. Bekbolatova, M., Mayer, J., Ong, C. W., & Toma, M. (2024, January). Transformative potential of AI in healthcare: definitions, applications, and navigating the ethical landscape and public perspectives. In *Healthcare* (Vol. 12, No. 2, p. 125). MDPI.
3. Medenica, S., Zivanovic, D., Batkoska, L., Marinelli, S., Basile, G., Perino, A., ... & Zaami, S. (2022). The future is coming: artificial intelligence in the treatment of infertility could improve assisted reproduction outcomes—the value of regulatory frameworks. *Diagnostics*, 12(12), 2979.
4. Khudhur, Y. S. (2025). Artificial intelligence in obstetrics and Gynecology: Current applications and future perspectives. *International Journal of Obstetrics and Gynaecology*, 7(1), 12-25.
5. Zhang, Z., Beck, M. W., Winkler, D. A., Huang, B., Sibanda, W., & Goyal, H. (2018). Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Annals of translational medicine*, 6(11), 216.

6. Liu, H., Zhang, Z., Gu, Y., Dai, C., Shan, G., Song, H., ... & Sun, Y. (2023). Development and evaluation of a live birth prediction model for evaluating human blastocysts from a retrospective study. *Elife*, 12, e83662.
7. Rhenman, A., Berglund, L., Brodin, T., Olovsson, M., Milton, K., Hadziosmanovic, N., & Holte, J. J. H. R. (2015). Which set of embryo variables is most predictive for live birth? A prospective study in 6252 single embryo transfers to construct an embryo score for the ranking and selection of embryos. *Human Reproduction*, 30(1), 28-36.
8. Costa, V. G., & Pedreira, C. E. (2023). Recent advances in decision trees: An updated survey. *Artificial Intelligence Review*, 56(5), 4765-4800.
9. Costa, V. G., & Pedreira, C. E. (2023). Recent advances in decision trees: An updated survey. *Artificial Intelligence Review*, 56(5), 4765-4800.
10. Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, 2024, 69-79.
11. Liu, L., Liang, H., Yang, J., Shen, F., Chen, J., & Ao, L. (2024). Clinical data-based modeling of IVF live birth outcome and its application. *Reproductive biology and endocrinology*, 22(1), 76.
12. Kriegeskorte, N. (2015). Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual review of vision science*, 1, 417-446.
13. Singh, S. (2025). Artificial Intelligence Models for Predicting Fertility Transitions: A Systematic Review. *Asian Journal of Research in Computer Science*, 18(6), 122-129.
14. Bormann, C. L., Thirumalaraju, P., Kanakasabapathy, M. K., Kandula, H., Souter, I., Dimitriadis, I., ... & Shafiee, H. (2020). Consistency and objectivity of automated embryo assessments using deep neural networks. *Fertility and sterility*, 113(4), 781-787.
15. Strauss, J. F., Barbieri, R. L., Dokras, A., Williams, C. J., & Williams, S. Z. (Eds.). (2023). *Yen & Jaffe's Reproductive Endocrinology-E-Book: Physiology, Pathophysiology, and Clinical Management*. Elsevier Health Sciences.
16. Krichen, M., & Mihoub, A. (2025). Long short-term memory networks: A comprehensive survey. *AI*, 6(9), 215.
17. Salem, F. M. (2021). Gated RNN: the gated recurrent unit (GRU) RNN. In *Recurrent neural networks: from simple to gated architectures* (pp. 85-100). Cham: Springer International Publishing.
18. Alomar, K., Aysel, H. I., & Cai, X. (2025). CNNs, RNNs and Transformers in human action recognition: a survey and a hybrid model. *Artificial Intelligence Review*, 58(12), 1-44.
19. Gao, Y., Yuan, Y., Wang, K., Wang, Y., Gao, T., Yang, Y., ... & Liu, X. (2025). Current progress and open challenges for applying artificial intelligence across the in vitro fertilization cycle. *Patterns*.
20. Luong, T. M. T., & Le, N. Q. K. (2024). Artificial intelligence in time-lapse system: advances, applications, and future perspectives in reproductive medicine. *Journal of assisted reproduction and genetics*, 41(2), 239-252.
21. Nasiri, N., & Eftekhari-Yazdi, P. (2015). An overview of the available methods for morphological scoring of pre-implantation embryos in in vitro fertilization. *Cell Journal (Yakhteh)*, 16(4), 392.
22. Wang, J., Guo, Y., Zhang, N., & Li, T. (2023). Research progress of time-lapse imaging technology and embryonic development potential: A review. *Medicine*, 102(38), e35203.
23. Tokuoka, Y., Yamada, T. G., Mashiko, D., Ikeda, Z., Hiroi, N. F., Kobayashi, T. J., ... & Funahashi, A. (2020). 3D convolutional neural networks-based segmentation to acquire quantitative criteria of the nucleus during mouse embryogenesis. *NPJ systems biology and applications*, 6(1), 32.
24. Khan, A., Rauf, Z., Sohail, A., Khan, A. R., Asif, H., Asif, A., & Farooq, U. (2023). A survey of the vision transformers and their CNN-transformer based variants. *Artificial Intelligence Review*, 56(Suppl 3), 2917-2970.
25. Liu, H., Zhang, Z., Gu, Y., Dai, C., Shan, G., Song, H., ... & Sun, Y. (2023). Development and evaluation of a live birth prediction model for evaluating human blastocysts from a retrospective study. *Elife*, 12, e83662.