ORIGINAL ARTICLE

Deep Learning-Based Prediction of IVF Success: A Transformer Model Approach

Mahvash Zargar [1*] ; Seyed Masoud Rezaeijo [2*] , Mahin Najafian [1], Kobra Shojaei [1], Vahideh Yousefvand [1]

[1] *Department of Obstetrics and Gynecology, Fertility, Infertility and Perinatology Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran*

[2] *Department of Medical Physics, Faculty of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran*

*Corresponding Authors: Mahvash Zargar, Seyed Masoud Rezaeijo Received: 27 March 2025 / Accepted: 02 August 2025

Emails: zargar_m@ajums.ac.ir, masoudrezayi1398@gmail.com

Abstract

Purpose: Predicting the success of assisted reproductive technology (ART) remains a significant challenge due to the complex interplay of clinical, embryological, and demographic factors. This study aimed to develop and evaluate machine learning models, particularly deep learning-based approaches, to identify key predictors of ART success and improve outcome prediction accuracy.

Materials and Methods: A retrospective study was conducted on 500 infertile couples undergoing ART treatment between 2019 and 2024. A comprehensive dataset, including 84 clinical, embryological, and demographic variables, was analyzed. The key predictors included endometrial thickness, endometrial pattern, embryo transfer day, and hormonal markers (PRL, LH). Four machine learning models were implemented: Decision Tree, Random Forest, XGBoost, and a Transformer-Based Model. Data preprocessing involved feature selection, missing data handling, normalization, and oversampling techniques to address class imbalance. The models were trained and validated using k-fold cross-validation, and performance was assessed using accuracy, precision, recall, and F1 score.

Results: The Transformer-Based Model achieved the highest accuracy (99.7%), outperforming traditional machine learning models. This performance was validated using k-fold cross-validation and oversampling to mitigate overfitting and ensure generalizability. Endometrial pattern (r = 0.69) and endometrial thickness (r = 0.82) were the strongest predictors of ART success, emphasizing the dominant role of uterine factors. While female age and infertility duration had a weak negative correlation, male infertility factors and lifestyle variables (smoking, alcohol consumption) showed minimal predictive significance. Model-based feature importance confirmed uterine and embryological factors as the primary determinants of ART success, indicating a potential shift in treatment strategies toward optimizing endometrial receptivity and embryo transfer timing.

Conclusion: This study highlights the superiority of deep learning models in ART success prediction, with uterine factors emerging as the strongest predictors. Integrating AI-driven predictive models into clinical practice can enable personalized ART treatment, improved patient counseling, and optimized embryo transfer strategies, ultimately enhancing fertility outcomes. However, the findings are based on data from a single medical center, and further multi-center validation is needed to confirm the model's generalizability.

Keywords: Assisted Reproductive Technology; In Vitro Fertilization; Clinical Pregnancy Prediction; Endometrial Receptivity; Embryo Transfer Timing; Machine Learning; Deep Learning.

Copyright © 2026 Tehran University of Medical Sciences.
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (https://creativecommons.org/licenses/by-nc/4.0/). Noncommercial uses of the work are permitted, provided the original work is properly cited.
DOI: https://doi.org/10.18502/fbt.v13i2.21932

1. Introduction

Assisted Reproductive Technology (ART) refers to a set of medical interventions designed to assist individuals and couples facing infertility in achieving pregnancy [1,2]. Among these techniques, in vitro fertilization (IVF) is the most widely used and extensively studied method. IVF involves the retrieval of oocytes from the ovaries, fertilization with sperm in a controlled laboratory environment, and the subsequent transfer of viable embryos into the uterus. Since its introduction, IVF has undergone significant advancements, improving success rates through innovations in embryo culture, cryopreservation, and genetic screening [3–5]. However, despite these developments, IVF outcomes remain highly variable, with success rates influenced by multiple biological, clinical, and technical factors. Optimizing IVF success requires a thorough understanding of these determinants, prompting researchers to explore advanced analytical methods, including artificial intelligence and machine learning, to improve prediction accuracy and treatment outcomes [6–10].

The high financial burden of each IVF cycle, coupled with the emotional and psychological stress of repeated failures, places immense pressure on infertile couples. These challenges highlight the urgent need for more accurate predictive models that can assess the likelihood of IVF success before initiating treatment. By developing a reliable and efficient IVF outcome prediction model, clinicians can personalize treatment strategies, optimize embryo selection, and enhance the overall success rates of ART. A robust predictive framework can help reduce unnecessary procedures, lower financial costs, and improve patient counseling, ultimately leading to more effective infertility management and higher patient satisfaction [11,12]. These challenges highlight the urgent need for more accurate predictive models that can assess the likelihood of IVF success before initiating treatment. By developing a reliable and efficient IVF outcome prediction model, clinicians can personalize treatment strategies, optimize embryo selection, and enhance the overall success rates of ART [13–15].

Recent advances in artificial intelligence (AI), particularly in machine learning (ML), have shown promising potential in reproductive medicine by enabling more accurate and personalized predictions of IVF outcomes [10, 14, 16]. Despite these advances, predicting IVF success remains challenging due to the multifactorial and heterogeneous nature of infertility, has shown promising potential in reproductive medicine by enabling more accurate and personalized predictions of IVF outcomes [17–19]. ML algorithms such as decision trees, random forests, and deep learning networks can uncover complex relationships among physiological, embryological, and clinical variables that influence ART success [20, 21]. These AI-based approaches provide clinicians with refined tools for optimizing embryo selection, transfer timing, and overall treatment planning [22–26]. AI-based models have demonstrated improved predictive accuracy over traditional statistical methods in several studies, particularly in analyzing complex, high-dimensional ART datasets [10, 27].

Over the past decade, the application of ML techniques to IVF outcome prediction has evolved significantly, reflecting growing interdisciplinary collaboration between reproductive medicine and computational science. Dehghan *et al.* [28] conducted a comprehensive comparison of five prominent ML algorithms—Random Forest, Artificial Neural Network (ANN), Support Vector Machine (SVM), Recursive Partitioning and Regression Trees (RPART), and AdaBoost—in combination with genetic algorithm (GA) feature selection, for IVF success prediction. Their findings demonstrated that AdaBoost, when integrated with GA-based feature selection, achieved the highest accuracy (89.8%), while Random Forest with GA achieved 87.4%. Feature selection proved crucial, with ten key variables (including female age, AMH, endometrial thickness, sperm count, and embryo quality indicators) consistently identified as the most influential predictors. Similarly, Hassan *et al.* [29] utilized a hill climbing attribute selection algorithm and five ML models (MLP, SVM, C4.5, CART, and RF) to predict IVF outcomes, reporting that feature selection improved predictive accuracy and highlighting attributes such as age, antral follicle count, fertilization rate, and embryo transfer day as major determinants of pregnancy success. Yiğit *et al.* [30] compared nine classification algorithms in a cohort of 939 embryo transfers, finding that ensemble models such as Random Forest and Super Learner achieved the highest F1 scores (74% and 73%,

respectively) and accuracy (89%), with maternal age, transfer day, and oestradiol concentration emerging as significant predictors.

Comparative studies have further highlighted the advantages of advanced ML models over classic statistical approaches. Barnett-Itzhaki *et al.* [16] found that machine learning algorithms (SVM, NN) outperformed logistic regression in predicting key IVF outcomes, especially when clinical data was incorporated, achieving accuracies of up to 0.9 compared to 0.74 for logistic regression. Sadegh-Zadeh *et al.* [11] underscored the potential of ensemble learning paradigms—including Logit Boost and Random Forest—for enhancing prediction, with Logit Boost reaching an accuracy of 96.35% using extensive datasets and detailed patient profiling. Other recent works have explored the utility of convolutional neural networks (CNNs) and automated time-lapse imaging for single-embryo selection, reporting practical accuracies of up to 85% in supporting embryo selection and reducing observer variability [31]. Collectively, these studies demonstrate that advanced ML and feature selection techniques not only improve predictive performance but also identify novel determinants of IVF success, supporting more individualized and effective patient care. However, the literature also notes ongoing challenges regarding model generalizability, integration with clinical workflows, and the need for external validation in diverse patient populations.

Despite promising results, previous studies have faced limitations such as reliance on conventional ensemble algorithms, narrow variable selection, and limited generalizability due to suboptimal model complexity or data preprocessing techniques. These approaches often underrepresent critical feature interactions—particularly in embryological and uterine domains—and fail to achieve high precision necessary for clinical adoption. To address these limitations, our study introduces a Transformer-Based deep learning model capable of capturing nonlinear, high-dimensional relationships across 84 clinical, hormonal, demographic, and embryological features. This model not only achieves superior predictive performance but also shifts focus toward underexplored predictors such as endometrial pattern and thickness, thereby offering clinically actionable insights for personalized ART decision-making.

This study presents a novel approach to predicting the outcomes of ART and IVF cycles using advanced ML techniques. By systematically analyzing a wide range of patient-related, embryological, and clinical variables, this research provides new insights into the key determinants of ART success.

2. Materials and Methods

Our study employed a systematic machine learning approach to predict in vitro fertilization outcomes, as illustrated in Figure 1. This framework integrates multi-dimensional patient data through a structured pipeline of data processing, model development, and performance evaluation. The methodology begins with comprehensive preprocessing of 84 variables across demographic, clinical, and embryological domains. Four distinct machine learning algorithms were systematically implemented and optimized: Decision Tree, Random Forest, Gradient Boosting (XGBoost), and a Transformer-based model. Each algorithm underwent rigorous hyperparameter tuning to maximize predictive accuracy. The framework culminates in a thorough evaluation using multiple performance metrics, enabling the identification of key success determinants for personalized ART treatment planning. This integrated approach represents a significant advancement in the application of artificial intelligence to reproductive medicine, offering clinicians a reliable tool for optimizing patient outcomes.

2.1. Study Design and Data Collection

2.1.1. Data Source and Patient Selection

This study was conducted at Imam Khomeini Hospital, affiliated with Ahvaz Jundishapur University of Medical Sciences, located in Ahvaz, Iran. Data were retrospectively collected from patients undergoing ART treatment between 2019 and 2024. Demographic, clinical, laboratory, and imaging data were extracted from electronic medical records and structured checklists completed by trained personnel. Hysterosalpingography (HSG) images were interpreted by an experienced reproductive specialist. All data were anonymized prior to analysis.

2.1.2. Inclusion and Exclusion Criteria

Eligible participants were infertile couples, defined by failure to conceive after one year of unprotected intercourse in women under 35 or after six months in women over 35. Exclusion criteria included: (1) use of ART for non-infertility reasons such as gender selection or multiple gestation; (2) use of donor gametes or surrogacy; (3) incomplete or unreliable medical records; and (4) refusal to provide informed consent. After applying these criteria, 500 cases were included in the final analysis.

2.2. Feature Selection and Data Preprocessing

2.2.1. Identification of Key Variables

In this study, a total of 84 variables were selected based on their potential impact on ART success. These variables were categorized into five main groups: (1) Demographic factors, including patient age, height, weight, and BMI; (2) Infertility history, such as primary or secondary infertility, infertility duration, and unexplained infertility; (3) Clinical and laboratory parameters, including hormone levels (FSH, LH, AMH, PRL, E2), metabolic conditions (diabetes, hypertension, thyroid disorders), and uterine abnormalities (fibroids, endometriosis, adenomyosis); (4) Lifestyle factors, such as smoking, alcohol consumption, and occupational hazards; and (5) Embryological characteristics, including the number and quality of retrieved oocytes, embryo morphology, day of embryo transfer, and freezing status. These variables were selected based on prior medical studies and expert consultation, ensuring that the dataset comprehensively captures the determinants of ART outcomes. After initial expert-driven selection based on clinical relevance, Recursive Feature Elimination(RFE) [32,33] using Random Forest and logistic regression models was applied to identify the most predictive variables. Correlated and redundant features were removed to enhance model stability and interpretability. A structured checklist includes demographic, clinical, and embryological factors (Table 1).

To reduce dimensionality and enhance model performance, feature selection was conducted in two stages. First, an expert panel preselected clinically relevant variables. Then, RFE using Random Forest and Logistic Regression was applied to identify the most predictive subset of features. This process reduced noise and improved generalizability.

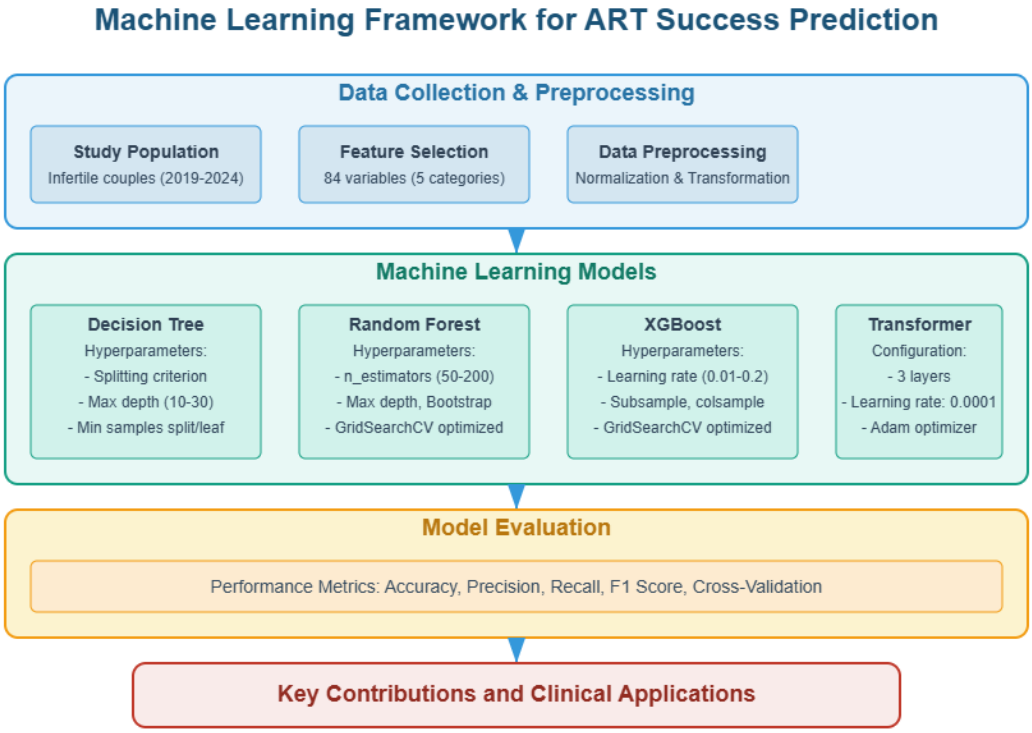


Figure 1. Comprehensive Machine Learning Framework for Predicting ART/IVF Success Outcomes

Table 1. Categorized List of Study Features with Descriptions

Category	Variables	Description
Demographic Variables	Female age, Male age, Female height, Male height, Female weight, Male weight, Female BMI, Male BMI, Female occupation, Male occupation, Blood type (female), Blood type (male)	Basic physiological and social characteristics of both partners
Infertility History	Infertility duration, Primary infertility (female/male), Secondary infertility (female/male), Unexplained infertility	Duration and type of infertility; unexplained causes
Female Reproductive Factors	Menstrual disorders, Severe pelvic adhesions, Endometriosis, AFC (total/right/left), HSG results, Hydrosalpinx, PCOS, Endometrioma, Uterine fibroids (intramural, subserosal, submucosal), Adenomyosis, Salpingitis	Gynecologic history, ovarian reserve, uterine anomalies, tubal disease
Male Factors	Semen analysis category, FSH (male), Testicular biopsy, TESE outcome, DFI, Varicocele surgery, Genital surgery, Karyotype (male)	Semen quality, hormonal markers, surgical history, genetic factors
Clinical & Laboratory Parameters	Female FSH, LH, PRL, AMH, E2 (baseline), Male FSH, Diabetes, Hypertension, Thyroid dysfunction, Anemia, Hepatitis, Lupus & Antiphospholipid Syndrome (female/male)	Hormonal and systemic health indicators for both partners
Lifestyle Factors	Smoking, Opioid use, Alcohol consumption, Vitamin D levels (female and male)	Behavioral and nutritional influences on fertility
COVID-19 Factors	COVID-19 history and vaccination status (female and male)	Impact of COVID-19 infection and immunization on ART outcomes
Embryological Factors	Number of retrieved oocytes, Number of transferred embryos, Embryo quality, Embryo freezing status, Embryo morphology, Day of embryo transfer	Lab-based embryo characteristics and transfer protocol
Outcome Variable	Final ART outcome (chemical pregnancy based on positive Beta HCG test)	Primary outcome variable representing ART success

2.2.2. Handling Missing Data and Outliers

To maintain data accuracy and reliability, rigorous preprocessing techniques were applied. Missing data were handled using a combination of imputation methods and exclusion criteria. For variables with minimal missing values (<5%), missing entries were replaced using mean or median imputation for continuous variables and mode imputation for categorical variables. However, cases with extensive missing data (>20% of key variables) were excluded to prevent bias. Outliers were detected using interquartile range (IQR) analysis and z-score thresholding. Clinically implausible values, such as extreme BMI values or unrealistic hormone levels, were further reviewed by domain experts and either corrected or removed. These preprocessing steps ensured that the dataset remained statistically robust and clinically meaningful.

A summary of missing data frequencies is presented in Table 2. Variables with more than 5% missing data included AMH (7.4%), PRL (6.9%), and embryo morphology (6.1%). In total, 28 patient records were excluded from the final dataset due to missing more than 20% of essential variables, resulting in 472 complete cases used for analysis.

Table 2. Missing Data Summary and Case Exclusions

Variable	Missing Rate (%)
AMH	7.4
PRL	6.9
Embryo Morphology	6.1
Vitamin D (Female)	5.8
Total Cases Excluded	28

2.2.3. Data Normalization and Transformation

To optimize machine learning performance and improve the stability of predictive models, data normalization and transformation techniques were implemented. Continuous variables (e.g., age, BMI,

hormone levels) were standardized using z-score normalization, where each value was rescaled based on the mean and standard deviation. This helped mitigate the effects of different scales among variables. Categorical variables (e.g., infertility type, embryo quality, smoking status) were encoded using one-hot encoding for nominal data and label encoding for ordinal categories. Additionally, skewed numerical distributions, such as hormone levels and embryo parameters, were transformed using log transformation to approximate a normal distribution, reducing the impact of extreme values on model predictions. These preprocessing steps ensured that the dataset was well-prepared for machine learning analysis, enhancing both interpretability and predictive accuracy.

2.3. Machine Learning Models and Implementation

2.3.1. Decision Tree Classifier

The Decision Tree Classifier is a simple yet effective model that partitions the dataset into decision nodes based on feature importance [34]. It was configured with different hyperparameters to optimize its performance. The splitting criterion was set to either Gini impurity or entropy, determining how nodes were divided. The maximum depth of the tree was tested with values None (unlimited), 10, 20, and 30, ensuring a balance between model complexity and overfitting. Additionally, min_samples_split were varied (2, 5, 10) to control the minimum number of samples required for a node to split, while min_samples_leaf (1, 2, 4) determined the minimum samples per leaf node to prevent overfitting. The random state was fixed at 42 for reproducibility. Unlike other models, the Decision Tree Classifier was implemented without GridSearchCV, as it primarily served as a baseline for comparison with more complex models.

2.3.2. Random Forest Classifier

The Random Forest Classifier [35], an ensemble learning method, was employed to enhance predictive accuracy by aggregating multiple decision trees. This model was optimized using GridSearchCV, ensuring systematic hyperparameter tuning. The number of trees (n_estimators) was varied across 50, 100, and 200, while the maximum tree depth was tested with values None, 10, 20, and 30, preventing overfitting by limiting node expansion. Additionally, the minimum samples required for node splitting (min_samples_split) and minimum samples per leaf node (min_samples_leaf) were tuned with values 2, 5, 10 and 1, 2, 4, respectively. The model also tested bootstrap sampling, which was either enabled (True) or disabled (False), determining whether the training data were sampled with replacement. A random state of 42 was used to ensure reproducibility. Given its ability to handle high-dimensional and imbalanced datasets, the Random Forest Classifier was expected to perform robustly across different ART outcome variables.

2.3.3. Gradient Boosting (XGBoost)

The XGBoost Classifier was implemented as a gradient boosting technique [35], optimizing decision trees sequentially to minimize errors iteratively. Hyperparameter tuning was conducted using GridSearchCV to enhance predictive accuracy. The number of estimators (n_estimators) was tested at 50, 100, and 200, while the maximum tree depth (max_depth) was varied across None, 10, 20, and 30 to balance model complexity. The learning rate (learning_rate) was adjusted with values 0.01, 0.1, and 0.2, controlling how much the model updated its weights during training. To prevent overfitting, subsample ratios (subsample) were tested at 0.8, 0.9, and 1.0, indicating the proportion of training data used per boosting iteration. Similarly, feature selection per tree (colsample_bytree) was varied at 0.8, 0.9, and 1.0, determining how many features were considered at each split. The random state was set to 42 for reproducibility. Given XGBoost's superior performance in handling non-linear interactions and imbalanced data, it was expected to perform effectively in predicting ART outcomes.

2.3.4. Transformer-Based Model

A Transformer-based deep learning model was implemented to capture complex feature interactions in ART prediction [36, 37]. This model consisted of three layers (num_layers = 3) and was trained using a batch size of 8 over 100 epochs to optimize weight updates effectively. The learning rate (lr) was set to

0.0001, ensuring gradual convergence without overshooting optimal values. The Cross Entropy Loss function was used to measure classification errors, making it suitable for multi-class categorization. The model was optimized using the Adam optimizer, which efficiently adjusted learning rates per parameter, incorporating weight decay regularization to mitigate overfitting. The Sigmoid activation function was applied to interpret output probabilities as ART success likelihoods. Model training was conducted using PyTorch and the Transformers library, leveraging Google Colab's cloud computing resources for computational efficiency. Given its ability to analyze sequential dependencies and complex feature interactions, the Transformer-based model aimed to provide superior predictive accuracy compared to traditional machine learning methods.

The Transformer model architecture [36, 38] consisted of 3 encoder layers, each with 4 multi-head attention heads, a feed-forward hidden layer of 128 dimensions, and an embedding size of 64. A batch size of 8 was used due to hardware constraints and to maintain stable gradient updates during training. The choice of three layers balanced complexity with our dataset size to reduce overfitting while capturing relevant feature interactions.

2.4. Model Training and Evaluation

2.4.1. Data Splitting (Training, Validation, and Testing Sets)

To ensure reliable model evaluation, the dataset was divided into three folds using k-fold cross-validation. This method allows each fold to be used as a test set while the remaining folds serve as the training set, enhancing the generalizability of the model. Stratified k-fold cross-validation was employed to maintain the same class distribution across all folds, preventing bias toward majority-class samples.

To further mitigate overfitting, dropout regularization (rate = 0.3) was implemented between layers in the Transformer model, in addition to weight decay. Despite these precautions, we acknowledge that the model was trained on data from a single institution and not validated on an external dataset. Future studies are needed to evaluate generalizability across different clinical settings.

2.4.2. Class Distribution and Oversampling

Given the inherent class imbalance in ART outcomes, oversampling was applied to the training sets to increase the representation of successful ART cycles. This technique involved synthetically replicating instances from the minority class, thereby balancing the dataset and preventing the model from being biased toward negative outcomes. Additionally, to mitigate overfitting, smoothing techniques were used to reduce the impact of noise and outlier data points. These techniques enhanced the model's ability to identify underlying patterns rather than memorizing specific instances, leading to improved generalization on unseen data. By combining oversampling and smoothing, the final model achieved a more stable and reliable performance across different data distributions.

To address class imbalance, we applied the Synthetic Minority Oversampling Technique (SMOTE) [39], which increased the number of successful ART cases from 117 to 234, achieving a near-balanced training dataset with 234 positive and 238 negative cases. This method was selected to avoid underfitting and ensure equitable learning across classes. To avoid data leakage and ensure robust model evaluation, SMOTE was applied separately within each training fold of the k-fold cross-validation procedure. The validation folds remained untouched, preserving the independence of the test data.

2.4.3. Hyperparameter Optimization

To enhance model performance, hyperparameter tuning was conducted using GridSearchCV and random search methods, allowing for systematic exploration of parameter combinations. For Decision Tree and Random Forest models, key hyperparameters such as maximum tree depth, minimum samples required for a split, minimum samples per leaf, and the number of estimators were optimized to balance model complexity and prevent overfitting. In the XGBoost model, critical parameters including learning rate, number of estimators, tree depth, subsampling rate, and feature selection ratio were fine-tuned to improve convergence speed and classification accuracy. Additionally, for the Transformer-based model, adjustments were made to the number of layers, batch size, learning rate, and weight decay to

optimize deep learning performance and enhance generalization. These optimizations ensured that each algorithm was efficiently trained, minimized overfitting, and achieved the highest possible predictive accuracy in ART outcome classification.

2.4.4. Performance Metrics

The predictive performance of the models was assessed using multiple evaluation metrics to ensure a comprehensive and reliable comparison. Accuracy was used as a general measure of the model's overall correctness in classifying ART outcomes. However, given the class imbalance in ART success prediction, additional metrics were considered. Precision quantified the proportion of correctly predicted positive cases out of all cases classified as positive, ensuring that the model did not generate excessive false positives. Recall (sensitivity) measured the model's ability to correctly identify actual positive cases, which is crucial in medical applications where missing a successful ART outcome could lead to suboptimal clinical decisions. To balance precision and recall, the F1 score, defined as the harmonic mean of these two metrics, was computed, providing a more robust measure of model performance. These metrics were calculated for each algorithm, enabling a comparative analysis to identify the most effective model for predicting ART success.

2.5. Statistical and Correlation Analysis

2.5.1. Correlation Between Clinical Factors and ART Success

To identify key predictors of ART success, correlation analysis was conducted between clinical variables and ART outcomes. Pearson's correlation coefficient was used for continuous variables (e.g., age, BMI, hormone levels, embryo characteristics), while Spearman's rank correlation was applied to categorical and ordinal variables (e.g., infertility type, previous surgeries, smoking status). The results indicated that female and male age had a strong negative correlation with ART success, while endometrial thickness, embryo morphology, and day of embryo transfer showed a positive correlation with successful implantation. Other commonly assumed predictors, such as BMI, smoking, alcohol consumption, and metabolic conditions, demonstrated no significant correlation with ART outcomes.

2.5.2. Statistical Tests for Feature Significance

To further evaluate the statistical significance of selected features, independent t-tests and Mann-Whitney U tests were used for continuous variables, comparing successful versus unsuccessful ART cycles. Chi-square tests were applied to categorical variables to determine their association with ART success. Additionally, logistic regression analysis was performed to estimate the predictive weight of each feature, identifying the most influential factors for ART outcomes. The statistical tests provided a robust foundation for feature selection and model refinement, ensuring that only clinically relevant variables were incorporated into the final predictive model.

3. Results

3.1. Descriptive Statistics

3.1.1. Demographic and Clinical Characteristics of Patients

A total of 500 infertile couples were included in this study. The mean age of women was 32.16 years (SD = 4.8), while the mean age of men was 36.38 years (SD = 5.3). The majority of women had a normal BMI, with a mean value of 26.69 kg/m² (SD = 3.4), and most men fell within the overweight category, with a mean BMI of 26.85 kg/m² (SD = 3.7). In terms of infertility history, primary infertility was more common than secondary infertility, with 62% of couples experiencing primary infertility and 38% experiencing secondary infertility. The mean duration of infertility was 4.2 years (SD = 2.1). Among clinical conditions, 22% of women had a history of endometriosis, 34% had PCOS, and 18% had uterine fibroids.

Additionally, 26% of men had a history of varicocele surgery, and 31% had abnormal semen analysis results. Regarding lifestyle factors, 19% of women and 27% of men reported smoking, while 12% of women and 21% of men reported alcohol consumption.

Statistical analysis confirmed significant associations between key clinical variables and ART success. Independent t-tests and Mann-Whitney U tests showed that endometrial thickness ($p < 0.001$) and endometrial pattern ($p = 0.002$) were significantly higher in successful ART cycles. Chi-square tests indicated no meaningful association between BMI ($p = 0.41$), smoking status ($p = 0.58$), or male infertility factors ($p = 0.67$) and ART success. Logistic regression analysis identified endometrial thickness (OR = 2.45, 95% CI: 1.88–3.12), embryo transfer day (OR = 1.72, 95% CI: 1.36–2.19), and baseline PRL levels (OR = 1.29, 95% CI: 1.10–1.52) as the most influential predictors, while female age (OR = 0.91, 95% CI: 0.86–0.97) and infertility duration (OR = 0.94, 95% CI: 0.89–0.99) had weaker effects. These findings reinforce the dominance of uterine receptivity and embryo-related factors in ART success and highlight their importance in predictive modeling and personalized treatment approaches.

Table 3 presents a summary of key demographic and clinical variables. Continuous variables are reported as means ± standard deviations (SD) along with ranges, while categorical variables are expressed as frequencies and percentages.

3.1.2. Distribution of ART Outcomes

The overall success rate of ART, defined by a positive Beta HCG test. A detailed analysis of success rates across different age groups revealed a gradual decline in ART success with increasing age, highlighting the impact of maternal and paternal age on reproductive outcomes. Among women, the highest success rate was observed in the 20–30-year age group (13.55%), which remained relatively stable in the 30–40-year group (13.23%). However, a sharp decline was noted in women aged 40–50 years, where the success rate dropped to 4.30%, indicating a significant reduction in fertility potential with advancing age. A similar trend was observed in men, where younger males (20–30 years) exhibited the highest success rate (14.15%), which decreased slightly in the 30–40-year age group (12.63%) and further declined to 11.11% in men aged 40–50 years. These findings underscore the negative impact of aging on ART success, particularly in women over 40, where the decline is more pronounced. The results emphasize the importance of early fertility intervention and personalized treatment strategies to optimize outcomes for couples undergoing ART. 3.2 Correlation Analysis and Feature Importance.

Table 3. Summary of Demographic and Clinical Characteristics

Variable	Mean ± SD (Range)	Frequency (%)
Female Age (years)	32.16 ± 4.8 (21–45)	-
Male Age (years)	36.38 ± 5.3 (24–52)	-
Female BMI (kg/m²)	26.69 ± 3.4 (18.2–34.6)	-
Male BMI (kg/m²)	26.85 ± 3.7 (19.1–35.3)	-
Infertility Duration (years)	4.2 ± 2.1 (1–12)	-
Primary Infertility	-	310 (62%)
Secondary Infertility	-	190 (38%)
Endometriosis (female)	-	110 (22%)
PCOS (female)	-	170 (34%)
Uterine Fibroids (female)	-	90 (18%)
Varicocele Surgery (male)	-	130 (26%)
Abnormal Semen Analysis (male)	-	155 (31%)
Smoking (female)	-	95 (19%)
Smoking (male)	-	135 (27%)
Alcohol Use (female)	-	60 (12%)
Alcohol Use (male)	-	105 (21%)

3.1.3. Correlation Between Clinical Factors and ART Success

A correlation analysis was conducted to evaluate the relationship between various clinical variables and ART success. Among female-related factors, maternal age demonstrated a negative correlation (-0.079) with ART success, indicating that older maternal age is associated with a lower probability of pregnancy. However, other commonly assumed predictors, including BMI, smoking, alcohol consumption, metabolic disorders (diabetes, hypertension, thyroid dysfunction), and infertility type (primary vs. secondary), showed no significant correlation with ART success. In contrast, endometrial factors emerged as the strongest predictors, with endometrial thickness ($r = 0.82$) and endometrial pattern ($r = 0.69$) showing a strong positive correlation with successful

implantation. These findings emphasize the crucial role of uterine receptivity in ART outcomes, suggesting that optimizing endometrial conditions may be more critical than other demographic or lifestyle factors.

Similarly, among male-related factors, paternal age exhibited a weak negative correlation (-0.065) with ART success, though its impact was less pronounced than maternal age. Other male factors, including semen analysis, history of infertility surgeries, smoking, alcohol consumption, and metabolic disorders, did not show a significant relationship with ART outcomes. Interestingly, primary male infertility had a weak positive correlation (0.051), suggesting a slight increase in ART success in these cases, though the effect size was negligible. Regarding medical and embryological variables, both embryo morphology ($r = 0.159$) and embryo transfer day ($r = 0.170$) had a positive correlation with ART success, highlighting the importance of embryo selection and optimal transfer timing. Conversely, cryopreservation of embryos exhibited a weak negative correlation (-0.042), suggesting a marginal advantage for fresh embryo transfers over frozen cycles. Notably, COVID-19 history, vaccination status, and autoimmune conditions (such as lupus and antiphospholipid syndrome) did not show any significant correlation with ART success, further reinforcing the dominance of uterine and embryological factors in determining treatment outcomes. For logistic regression, the model yielded an AUC-ROC of 0.84 and a Nagelkerke pseudo R^2 of 0.61, indicating strong discrimination and moderate explanatory power.

The reported success rates across age groups represent per-cycle chemical pregnancy rates, based on a single embryo transfer per treatment cycle. These rates are slightly lower than internationally reported averages (typically 20–35%) for IVF success. This discrepancy may be explained by characteristics of the study population, which includes patients with complex infertility etiologies and delayed access to ART services. Additionally, protocol variations and resource limitations inherent to a single-center setting may have influenced the outcomes. Cumulative success rates across multiple cycles were not assessed in this study but may yield higher overall pregnancy probabilities.

3.2. Key Predictors of ART Success

Feature importance analysis was conducted using Random Forest and XGBoost models, ranking variables based on their predictive contribution to ART success (Figures 2 and 3). The results revealed that uterine factors played the most significant role in determining ART outcomes. Among the top-ranking features, endometrial pattern emerged as the strongest predictor, emphasizing the importance of a receptive and well-structured endometrial lining for implantation. Similarly, endometrial thickness was identified as a critical determinant of successful implantation, aligning with clinical findings that a well-prepared uterine environment enhances pregnancy rates. Embryo transfer day also had a high predictive value, indicating that later-stage embryo transfers were associated with higher success rates, further reinforcing the importance of precise embryo transfer timing. Additionally, baseline prolactin (PRL) levels were found to be a key factor, highlighting the role of hormonal balance in ART success. Interestingly, while BMI showed no significant correlation in standard statistical analysis, it was identified as an influential predictor in some machine learning models, suggesting a potential nonlinear interaction with other physiological factors.

3.3. Model Performance Evaluation

3.3.1. Comparison of Machine Learning Models

The performance of the four machine learning models—Random Forest Classifier, Decision Tree Classifier, XGBoost Classifier, and Transformer-Based Model—was evaluated based on their ability to predict ART success. The results, summarized in Table 4, highlight the accuracy, precision, recall, and F1 score for each model.

The Transformer-Based Model achieved the highest overall performance, with an accuracy of 99.7% and an F1 score of 0.998, indicating exceptional predictive power. The Random Forest Classifier and XGBoost Classifier also performed well, with high accuracy and balanced precision-recall values, while the Decision Tree Classifier exhibited the weakest performance among the models. Transformer-Based Model: This model had the lowest False Negative rate, meaning it successfully identified

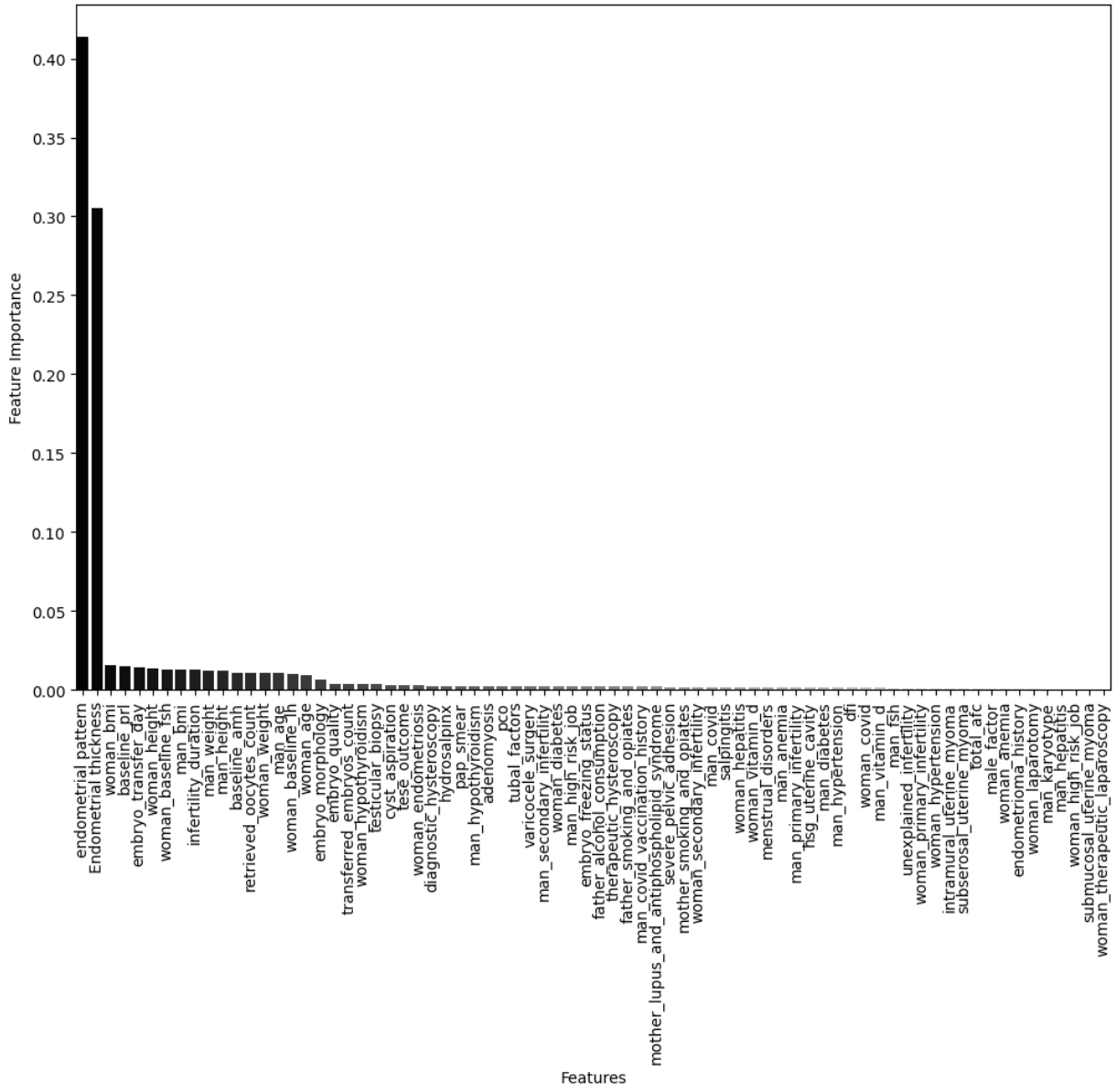


Figure 2. Feature Importance Analysis Using the Random Forest Model

the majority of positive cases (successful ART cycles) with minimal misclassification (Figure 4). A comparative analysis of the confusion matrices revealed that the Transformer-Based Model achieved the highest accuracy in predicting ART outcomes, with the lowest false negative rate—making it especially effective at identifying successful ART cases. In contrast, the Random Forest and XGBoost classifiers, while performing well overall, showed higher false positive rates and a tendency to overestimate success. The Decision Tree Classifier exhibited the weakest performance, with higher rates of misclassification for both positive and negative cases. Overall, these results demonstrate that deep learning methods, particularly transformer-based models, provide superior precision and reliability for ART success prediction compared to traditional machine learning approaches.

Figure 5 presents a t-SNE visualization illustrating the distribution of data points in the feature space across different machine learning models. This figure illustrates how each machine learning model differentiates between successful and unsuccessful ART cases in the high-dimensional feature space. The Transformer-Based Model shows clear separation between classes, indicating superior feature representation and predictive discrimination, while other models show greater overlap, reflecting lower classification precision

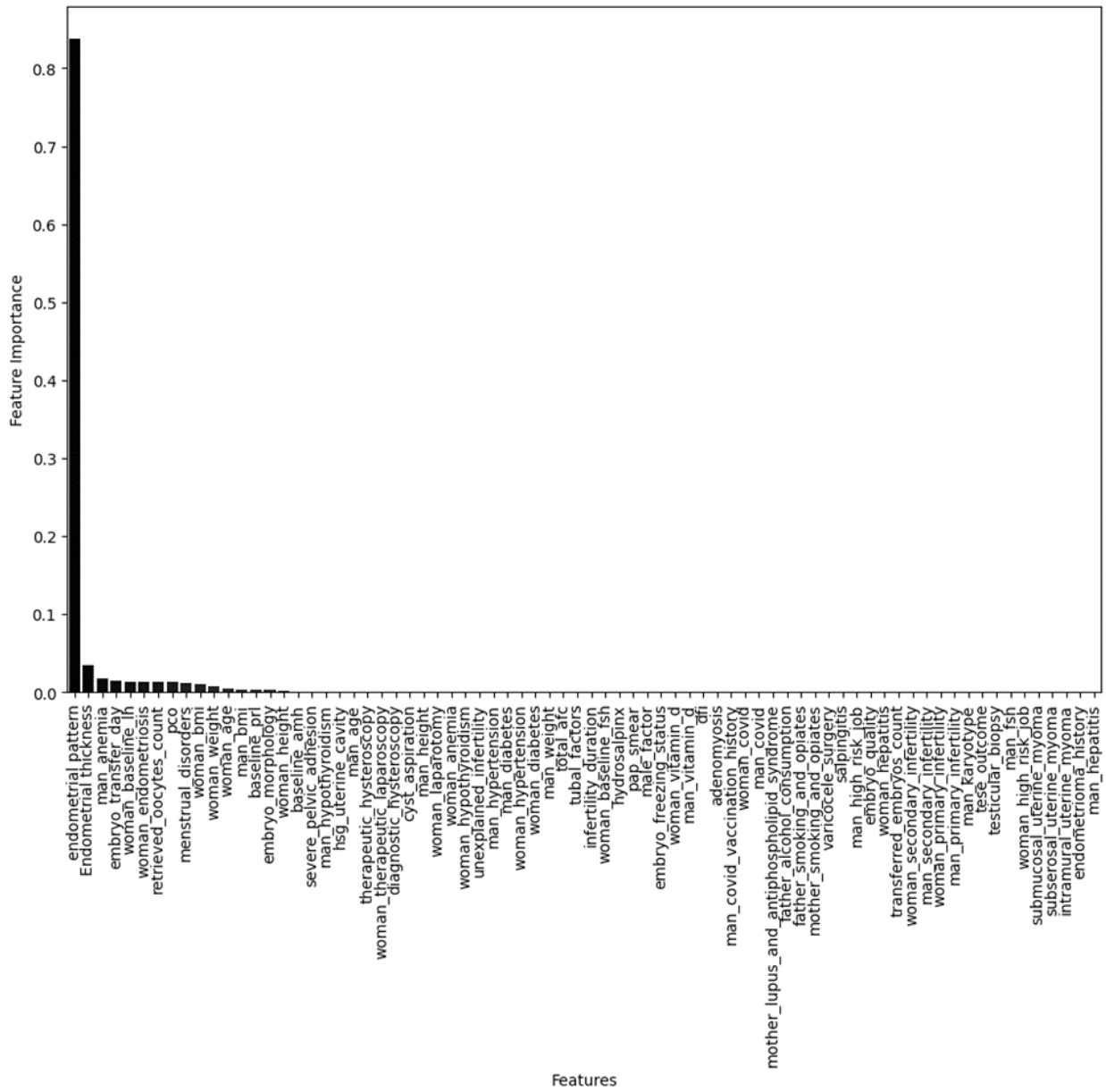


Figure 3. Feature Importance Analysis Using the XGBoost Model

Table 4. Model Performance Metrics

Model	Accuracy	Precision	Recall	F1 Score
RandomForest	0.9940	0.9966	0.9754	0.9856
DecisionTree	0.9940	0.9861	0.9859	0.9860
XGBoost	0.9940	0.9796	0.9930	0.9862
Transformer-Based	0.9970	1.0000	0.9970	0.9980

4. Discussion

This study aimed to develop and evaluate machine learning models for predicting ART success by analyzing a comprehensive set of clinical, embryological, and demographic factors. Our findings revealed that endometrial pattern and thickness were the strongest predictors of ART success, emphasizing the critical role of uterine receptivity in implantation outcomes. Additionally, embryo transfer day and baseline prolactin (PRL) levels were identified as influential factors, whereas BMI, male infertility parameters, and lifestyle factors (e.g., smoking, alcohol consumption) had minimal predictive value.

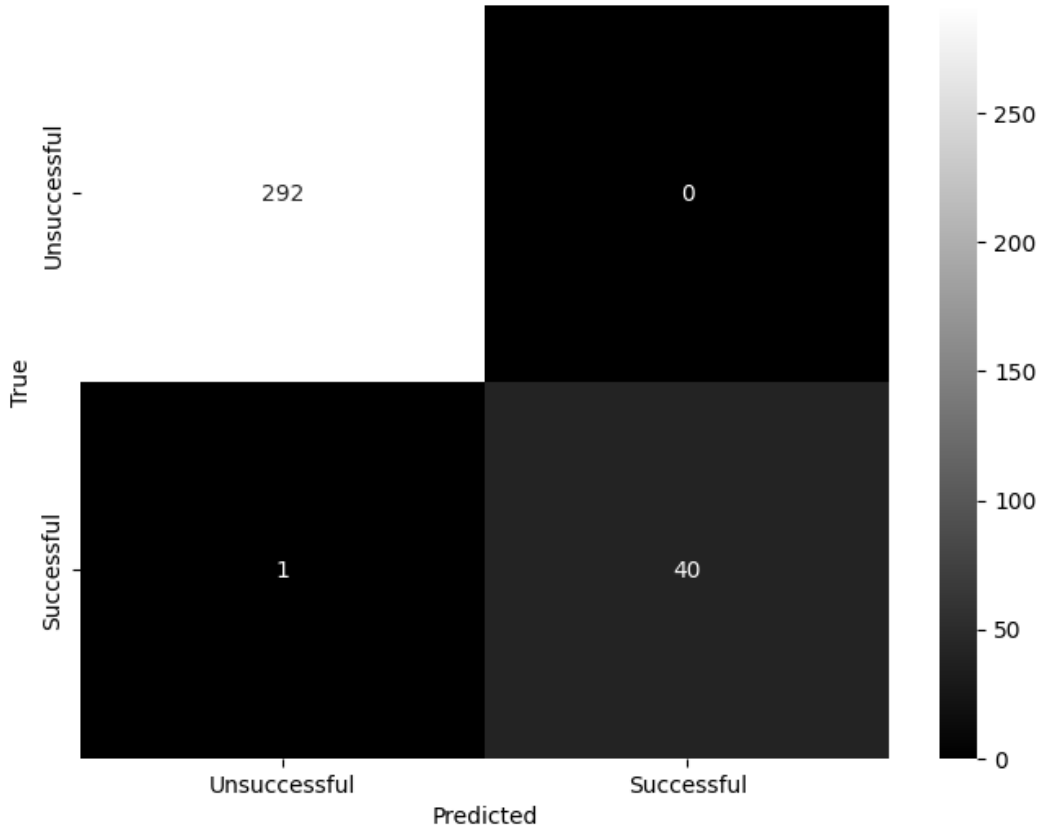


Figure 4. Confusion Matrix of the Transformer-Based Model for ART Success Prediction

Among the machine learning models tested, the Transformer-Based Model achieved the highest accuracy (99.7%) and the lowest false negative rate, making it the most effective tool for ART outcome prediction.

Our findings align with the study by Wang *et al.* [40], which used machine learning algorithms to predict clinical pregnancy in IVF and identified female age and duration of infertility as negative predictors, while the total number of frozen and transferred embryos had a positive effect on pregnancy outcomes. However, unlike Wang *et al.* [40], our study found that uterine factors (endometrial pattern and thickness) were more critical than embryo quantity, suggesting that implantation potential may be more dependent on uterine receptivity rather than the number of transferred embryos. These results indicate that future ART treatment strategies should focus more on optimizing the uterine environment rather than increasing the number of transferred embryos. Similarly, Dehghan *et al.* [28], who employed genetic algorithms (GA) for feature selection in IVF prediction, identified female age, AMH, endometrial thickness, and embryo quality as key determinants of ART success. While our study similarly found endometrial thickness to be a major predictive factor, we observed a lower impact of AMH and female age, potentially due to the stronger influence of uterine factors and hormonal markers in our model. Additionally, our Transformer-Based Model outperformed traditional ensemble models, whereas Dehghan *et al.* [28] found AdaBoost to be the most accurate predictive algorithm, suggesting that deep learning approaches may provide enhanced predictive power in complex ART datasets. In Dehghan *et al.*'s study, AdaBoost achieved an accuracy of 91.2% and an F1 score of 0.89 for IVF outcome prediction, while Peng *et al.* reported that their Random Forest model achieved an accuracy of 93.6% with an F1 score of 0.91. In contrast, our Transformer-Based Model outperformed both, with an accuracy of 99.7% and an F1 score of 0.998. This substantial improvement underscores the Transformer architecture's enhanced capacity to capture complex feature interactions and sequential dependencies,

positioning it as a next-generation tool for ART outcome prediction.

A detailed evaluation of the predictive models provided valuable insights into the key factors influencing ART success and their clinical relevance. Machine learning algorithms, particularly the Transformer-Based Model and XGBoost, identified uterine factors, embryo characteristics, and hormonal markers as the most significant predictors of ART outcomes. In contrast, demographic factors, lifestyle variables, and male infertility parameters played a minimal role in determining ART success. These findings underscore the importance of precisely assessing and optimizing the uterine environment to enhance implantation rates. Additionally, the models demonstrated that a personalized approach to ART treatment—considering individual hormonal and embryological profiles—could improve success rates. From a clinical perspective, the integration of machine learning-driven prediction models can aid in treatment customization, better patient counseling, and improved embryo selection strategies, ultimately leading to more effective fertility treatments and enhanced patient outcomes.

In contrast to the study by Mehrjerd *et al.* [41], which focused on sperm quality parameters (morphology, motility, and count) as predictors of clinical pregnancy in ART, our findings indicate that male factors had limited predictive power for IVF success. Their study, using ensemble machine learning models, found that sperm motility had a positive effect on ART success, particularly in cases with a lower number of retrieved oocytes. However, in our study, sperm-related variables, including semen

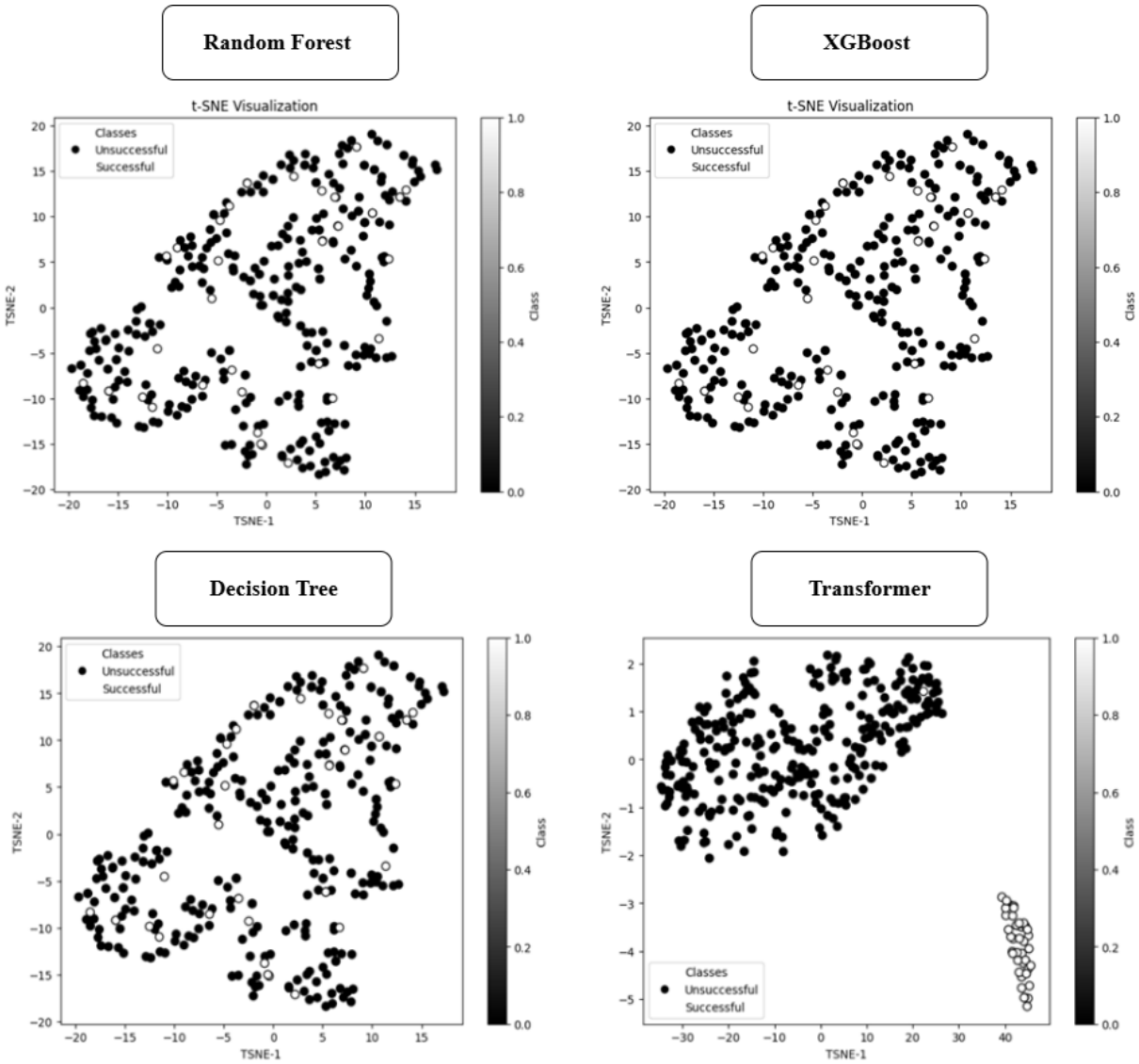


Figure 5. t-SNE Visualization of Feature Space Distribution Across Machine Learning Models

analysis results and varicocele surgery history, were not among the top predictors. These differences suggest that male factors may have a greater influence in IUI cycles or specific subgroups of IVF patients, whereas in standard IVF, uterine and embryonic factors dominate. Similarly, the study by Tian *et al.* [42], which used a Bayesian Network Model to predict failed fertilization in ART, found that infertility type, ART method, and the number of retrieved oocytes were direct predictors of fertilization failure. While their study emphasized male factors such as sperm count and motility, our findings suggest that fertilization success may not directly translate to ART success, as implantation is more dependent on endometrial and embryological factors. These differences highlight the complex and multi-step nature of ART outcomes, where different predictors may play a role at each stage of the reproductive process.

Our findings are in agreement with Drapkina *et al.* [43], who compared logistic regression, decision tree algorithms, and Random Forest models for ART prediction. Their study found Random Forest to be the most accurate model, identifying embryonic arrest, the number of embryos of excellent and average quality, and hormonal markers (FSH, AMH) as key predictors. Similarly, we found that hormonal markers (particularly PRL and LH) and embryo transfer day were highly predictive, but our Transformer-Based Model significantly outperformed Random Forest and other traditional machine learning methods. This suggests that advanced AI techniques, such as transformers, may provide a higher degree of accuracy and robustness in ART prediction compared to traditional ML models.

Moreover, Peng *et al.* [44] conducted a large-scale ART outcome prediction study using random forest, extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), and logistic regression, finding that random forest and logistic regression achieved the best predictive performance. Their study identified maternal age, FSH, sperm motility, progesterone (P) on HCG day, estradiol (E2) on HCG day, and luteinizing hormone (LH) on HCG day as key predictors. While we also found hormonal markers (PRL, LH) to be significant predictors, our deep learning model demonstrated superior performance compared to random forest and logistic regression, further reinforcing the value of advanced AI techniques in ART success prediction.

Conversely, several variables demonstrated low predictive importance. Lifestyle factors such as smoking and alcohol consumption, history of autoimmune disorders (e.g., lupus, antiphospholipid syndrome), and COVID-19 infection or vaccination status had minimal impact on ART success. Similarly, male infertility factors, including semen quality, history of varicocele surgery, and testicular biopsy results, did not significantly contribute to the predictive models, indicating that uterine and embryological factors outweighed male-related variables in determining ART outcomes. Further refinement of feature importance using the XGBoost model confirmed that endometrial pattern and thickness remained the most influential variables. Notably, anemia in men unexpectedly ranked higher than anticipated, suggesting a potential link between male hemoglobin levels and ART success, which warrants further investigation. Additionally, baseline LH levels in women were found to influence ART outcomes, reinforcing the critical role of ovarian function in fertility treatment. These findings suggest that uterine factors play a far more significant role in ART success than demographic or lifestyle variables, providing clinicians with valuable insights to prioritize uterine environment optimization rather than solely addressing patient demographics or lifestyle modifications.

The unexpectedly high feature importance of male anemia may reflect indirect associations with hormonal imbalances or underlying systemic inflammation, which can affect spermatogenesis and overall reproductive health. Prior studies have linked low hemoglobin levels with altered testosterone regulation and reduced semen quality, potentially impacting ART outcomes (Ribas-Maynou *et al.* [7]; Graham *et al.* [4]). While the exact mechanism remains unclear, this finding warrants further investigation to determine whether correcting anemia could enhance male fertility potential in ART settings.

To support the correlation analysis, p-values and 95% confidence intervals were calculated for all reported coefficients. The correlation between endometrial thickness and ART success ($r = 0.82$, $p < 0.001$) and between endometrial pattern and ART success ($r = 0.69$, $p < 0.001$) were both statistically

significant, confirming the robustness of these associations.

4.1. Analytical Interpretation and Clinical Implications

Our findings indicate that uterine and embryological factors play a far greater role than previously assumed, challenging traditional notions that age, AMH, or male infertility parameters are the strongest predictors. While many previous studies have emphasized ovarian and sperm-related factors, our study highlights the dominance of endometrial receptivity in ART success. Unlike prior research that relied on logistic regression, Bayesian networks, or traditional ensemble models, our Transformer-Based Model demonstrated superior predictive accuracy, suggesting that deep learning approaches can improve ART prediction by capturing complex interactions between multiple variables. From a clinical perspective, these findings emphasize the need for a paradigm shift in ART treatment strategies, moving beyond traditional approaches that prioritize ovarian function and sperm parameters. Instead, personalized ART protocols should focus on optimizing uterine receptivity before embryo transfer, as our study highlights endometrial pattern and thickness as the strongest predictors of ART success. By tailoring treatment plans to enhance endometrial conditions, clinicians can maximize implantation potential and improve pregnancy outcomes.

Additionally, the integration of machine learning-based prediction models in ART practice can significantly enhance patient counseling and decision-making. By leveraging real-time clinical data, these models enable clinicians to provide individualized success probability estimates, allowing for more informed discussions with patients about their likelihood of achieving pregnancy. This level of precision can help set realistic expectations and guide patients toward the most effective treatment options. Traditional practices that emphasize higher embryo transfer numbers may not necessarily lead to better outcomes, whereas precisely timed transfers in an optimized uterine environment can significantly improve ART success rates. By adopting data-driven, personalized approaches, clinicians can enhance fertility treatment efficiency and ultimately improve live birth outcomes for ART patients.

Despite the impressive predictive performance of the Transformer-Based Model, the potential for overfitting cannot be overlooked, especially given the relatively small, single-center dataset used for model development. To ensure generalizability and clinical reliability, future research must incorporate external validation using independent datasets from other ART centers. Cross-center studies with diverse patient populations and standardized data collection protocols will be essential for adapting the model to varied clinical environments. Moreover, the deployment of AI models in real-world ART settings requires careful consideration of interoperability with electronic medical records, clinician trust, and institutional readiness. These challenges must be addressed before such models can be widely implemented in routine reproductive care.

Beyond predictive performance, the clinical integration of AI-based models in ART workflows requires overcoming several practical hurdles. High-performance models, such as the Transformer architecture used in this study, often demand significant computational resources and may be challenging to deploy in low-resource clinical settings. Additionally, variability in data formats, electronic medical record systems, and terminology across institutions necessitates robust data standardization and interoperability frameworks. Another key barrier is clinician training—effective use of AI tools depends on user trust, interpretability of predictions, and adequate education in AI-assisted decision-making. Addressing these barriers will be essential to facilitate the successful and ethical deployment of AI in reproductive medicine.

4.2. Strengths, Limitations, and Future Directions

A major strength of this study is the integration of deep learning techniques, particularly the Transformer-Based Model, which outperformed traditional machine learning models by achieving higher accuracy and lower misclassification rates. This advantage stems from the model's ability to capture complex, nonlinear interactions among multiple clinical factors, making it a highly effective tool for ART outcome prediction. Additionally, the study utilized a comprehensive dataset, incorporating clinical, embryological, and hormonal variables,

providing a holistic and multidimensional approach to ART success prediction. Unlike previous studies that focused primarily on demographic or single-factor analyses, our model considered a broad spectrum of physiological and embryological parameters, enhancing the reliability of its predictions.

Despite these strengths, certain limitations should be acknowledged. First, the study did not include genetic and immunological markers, which are emerging as potentially significant contributors to ART outcomes. The absence of these variables may have limited the model's ability to capture underlying genetic and immune-related influences on implantation and pregnancy success. Second, the dataset was derived from a single medical center, which, while ensuring consistency in clinical protocols, may limit generalizability across diverse patient populations. External validation using multi-center datasets is essential to confirm the model's robustness and applicability in broader clinical settings. Lastly, future research should focus on integrating genetic algorithms (GA), advanced ensemble learning techniques, and real-time clinical data to further refine predictive accuracy and enhance the clinical utility of AI-driven ART decision-support systems.

Future research should also explore the integration of high-dimensional biological data such as single-cell RNA sequencing (scRNA-seq) to assess gene expression profiles related to endometrial receptivity and embryo viability. Additionally, the development of multimodal AI models that combine clinical, hormonal, and imaging data (e.g., time-lapse embryo imaging or ultrasound-derived endometrial assessments) may offer a more holistic understanding of ART outcomes. These advanced models could significantly enhance the precision of predictions and enable real-time clinical decision-making tailored to individual patient profiles.

5. Conclusion

In conclusion, our study highlights the dominance of uterine and embryological factors over traditional demographic or lifestyle predictors in ART success. The Transformer-Based Model demonstrated superior predictive accuracy, outperforming Random Forest, XGBoost, and other traditional ML models. Comparisons with previous research suggest that deep learning techniques provide a significant advantage in ART outcome prediction, allowing for more precise, personalized, and data-driven fertility treatments. Future advancements in AI-driven reproductive medicine will likely refine these models further, enhancing ART success rates and potentially improving patient outcomes, pending validation in diverse clinical settings.

While the Transformer-Based Model demonstrated superior performance in predicting ART outcomes, the findings should be interpreted in light of certain limitations. The use of a single-center dataset may constrain generalizability, and the exclusion of genetic and immunological markers limits the model's scope. Future research should validate the model in multi-center cohorts and explore the integration of genomic and imaging data to enhance predictive precision and clinical applicability.

Acknowledgment

The study was reviewed and approved by the Ahvaz Jundishapur University of Medical Sciences (ethics code: IR.AJUMS.HGOLESTAN.REC.1403.110) . The subjects have given their written informed consent. The research complies with the guidelines for human studies and the research was conducted ethically in accordance with the World Medical Association Declaration of Helsinki.

References

1- René C, Landry I, de Montigny F. "Couples' experiences of pregnancy resulting from assisted reproductive technologies: A qualitative meta-synthesis. "*Int J Nurs Stud Adv*.;4:100059, (2022).

2- Kuhnt AK, Passet-Wittig J. Families formed through assisted reproductive technology: Causes, experiences, and consequences in an international context." Vol. 14, *Reproductive biomedicine & society online. Elsevier*; p. 289–96, (2022).

3- Fang Y, Wu R, Lee JM, Chan LHM, Chan KYJ- "Microfluidic in-vitro fertilization technologies: Transforming the future of human reproduction." *TrAC Trends Anal Chem*.;160:116959, (2023).

4- Graham ME, Jelin A, Hoon Jr AH, Wilms Floet AM, Levey E, Graham EM. "Assisted reproductive technology:

Short-and long-term outcomes." *Dev Med Child Neurol.*;65(1):38–49, (2023).

5- Chiware TM, Vermeulen N, Blondeel K, Farquharson R, Kiarie J, Lundin K, *et al.* "IVF and other ART in low-and middle-income countries: a systematic landscape analysis." *Hum Reprod Update*.;27(2):213–28, (2021).

6- Bulletti FM, Sciorio R, Conforti A, De Luca R, Bulletti C, Palagiano A, *et al.* "Causes of embryo implantation failure: A systematic review and metaanalysis of procedures to increase embryo implantation potential." *Front Endocrinol (Lausanne)*.; 15:1429193, (2025).

7- Ribas-Maynou J, Yeste M, Becerra-Tomás N, Aston KI, James ER, Salas-Huetos A. "Clinical implications of sperm DNA damage in IVF and ICSI: updated systematic review and meta-analysis." *Biol Rev*.; 96(4):1284–300, (2021).

8- Amini P, Ramezanali F, Parchehbaf-Kashani M, Maroufizadeh S, Omani-Samani R, Ghaheri A. "Factors associated with in vitro fertilization live birth outcome: a comparison of different classification methods." *Int J Fertil Steril.*;15(2):128, (2021).

9- Siristatidis C, Stavros S, Drakeley A, Bettocchi S, Pouliakis A, Drakakis P, *et al.* "Omics and artificial intelligence to improve in vitro fertilization (IVF) success: a proposed protocol." *Diagnostics.*;11(5):743, (2021).

10- Hew Y, Kutuk D, Duzcu T, Ergun Y, Basar M. "Artificial Intelligence in IVF Laboratories: Elevating Outcomes Through Precision and Efficiency." *Biology (Basel)*.;13(12):988, (2024).

11- Sadegh-Zadeh SA, Khanjani S, Javanmardi S, Bayat B, Naderi Z, Hajiyavand AM. "Catalyzing IVF outcome prediction: exploring advanced machine learning paradigms for enhanced success rate prognostication." *Front Artif Intell*.;7:1392611, (2024).

12- Hanevik HI, Hessen DO. "IVF and human evolution." *Hum Reprod Update*. ;28(4):457–79, (2022).

13- Medenica S, Zivanovic D, Batkoska L, Marinelli S, Basile G, Perino A, *et al.* "The future is coming: artificial intelligence in the treatment of infertility could improve assisted reproduction outcomes—the value of regulatory frameworks." *Diagnostics*.;12(12):2979, (2022).

14- Cai J, Jiang X, Liu L, Liu Z, Chen J, Chen K, *et al.* "Pretreatment prediction for IVF outcomes: generalized applicable model or centre-specific model?" *Hum Reprod*.;39(2):364–73, (2024).

15- AlSaad R, Abd-Alrazaq A, Choucair F, Ahmed A, Aziz S, Sheikh J. "Harnessing artificial intelligence to predict ovarian stimulation outcomes in in vitro fertilization: scoping review." *J Med Internet Res*.;26:e53396, (2024).

16- Barnett-Itzhaki Z, Elbaz M, Butterman R, Amar D, Amitay M, Racowsky C, *et al.* Machine learning vs. classic statistics for the prediction of IVF outcomes. *J Assist Reprod Genet*.;37(10):2405–12, (2020).

17- Chavez Badiola A. "Artificial Intelligence (AI) for Embryo Ranking and its Use in Human Assisted Reproduction." *University of Kent*; (2024).

18- Liu K, Zhang Y, Martin C, Ma X, Shen B. "Translational bioinformatics for human reproductive biology research: examples, opportunities and challenges for a future reproductive medicine." *Int J Mol Sci*.;24(1):4, (2022).

19- Zhang Q, Liang X, Chen Z. "A review of artificial intelligence applications in in vitro fertilization." *J Assist Reprod Genet*.;42(1):3–14, (2025).

20- Hariton E, Pavlovic Z, Fanton M, Jiang VS. "Applications of artificial intelligence in ovarian stimulation: a tool for improving efficiency and outcomes." *Fertil Steril*.;120(1):8–16, (2023).

21- Khan HL, Khan S, Bhatti S, Abbas S. "Role of artificial intelligence in quality assurance in ART: a review." *Fertil Reprod.*;5(01):1–7, (2023).

22- Bulletti FM, Berrettini M, Sciorio R, Bulletti C. "Artificial intelligence algorithms for optimizing assisted reproductive technology programs: A systematic review." *Glob Transl Med.*;2(2):1–12, (2023).

23- Mapari SA, Shrivastava D, Bedi GN, Pradeep U, Gupta A, Kasat PR, *et al.* "Revolutionizing Reproduction: The Impact of Robotics and Artificial Intelligence (AI) in Assisted Reproductive Technology: A Comprehensive Review." *Cureus*.;16(6), (2024).

24- Wu YC, Su ECY, Hou JH, Lin CJ, Lin KB, Chen CH. "Artificial intelligence and assisted reproductive technology: A comprehensive systematic review." *Taiwan J Obstet Gynecol*.;64(1):11–26, (2025).

25- de Santiago I, Polanski L. "Data-driven medicine in the diagnosis and treatment of infertility." *J Clin Med*.;11(21):6426, (2022).

26- Abdullah KAL, Atazhanova T, Chavez-Badiola A, Shivhare SB. "Automation in ART: paving the way for the future of infertility treatment." *Reprod Sci*.;30(4):1006–16, (2023).

27- VerMilyea M, Hall JMM, Diakiw SM, Johnston A, Nguyen T, Perugini D, *et al.* "Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical

light microscopy during IVF." *Hum Reprod.*;35(4):770–84, (2020).

28- Dehghan S, Rabiei R, Choobineh H, Maghooli K, Nazari M, Vahidi-Asl M. "Comparative study of machine learning approaches integrated with genetic algorithm for IVF success prediction." *PLoS One*.;19(10):e0310829, (2024).

29- Hassan MR, Al-Insaif S, Hossain MI, Kamruzzaman J. "A machine learning approach for prediction of pregnancy outcome following IVF treatment." *Neural Comput Appl.*;32(7):2283–97, (2020).

30- Yiğit P, Bener A, Karabulut S. "Comparison of machine learning classification techniques to predict implantation success in an IVF treatment cycle." *Reprod Biomed Online*.;45(5):923–34, (2022).

31- Patil SN, Wali U V, Swamy MK. "Selection of single potential embryo to improve the success rate of implantation in IVF procedure using machine learning techniques." *In: 2019 International Conference on Communication and Signal Processing (ICCSP). IEEE*; p. 881–6, (2019).

32- Yan K, Zhang D. "Feature selection and analysis on correlated gas sensor data with recursive feature elimination." *Sensors Actuators B Chem*.; 212:353–63, (2015).

33- Chen X wen, Jeong JC. "Enhanced recursive feature elimination." I*n: Sixth international conference on machine learning and applications (ICMLA 2007). IEEE*;. p. 429–35, (2007).

34- Brijain M, Patel R, Kushik MR, Rana K. "A survey on decision tree algorithm for classification. "(2014);

35- Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. "Random forest: a classification and regression tool for compound classification and QSAR modeling." *J Chem Inf Comput Sci*.;43(6):1947–58, (2003).

36- Pereira GA, Hussain M. "A review of transformer-based models for computer vision tasks: Capturing global context and spatial relationships." *arXiv Prepr arXiv240815178*. (2024);

37- Zhang S, Fan R, Liu Y, Chen S, Liu Q, Zeng W. "Applications of transformer-based language models in bioinformatics: a survey." *Bioinforma Adv*.;3(1):vbad001, (2023).

38- Khan A, Rauf Z, Sohail A, Khan AR, Asif H, Asif A, *et al.* "A survey of the vision transformers and their CNN-transformer based variants." *Artif Intell Rev*.;56(Suppl 3):2917–70, (2023).

39- Chawla N V, Bowyer KW, Hall LO, Kegelmeyer WP. "SMOTE: synthetic minority over-sampling technique." *J Artif Intell Res.*;16:321–57, (2002).

40- Wang CW, Kuo CY, Chen CH, Hsieh YH, Su ECY. "Predicting clinical pregnancy using clinical features and machine learning algorithms in in vitro fertilization." *PLoS One*.;17(6):e0267554, (2022).

41- Mehrjerd A, Dehghani T, Jajroudi M, Eslami S, Rezaei H, Ghaebi NK. "Ensemble machine learning models for sperm quality evaluation concerning success rate of clinical pregnancy in assisted reproductive techniques." *Sci Rep*.;14(1):24283, (2024).

42- Tian T, Kong F, Yang R, Long X, Chen L, Li M, *et al.* "A Bayesian network model for prediction of low or failed fertilization in assisted reproductive technology based on a large clinical real-world data." *Reprod Biol Endocrinol.*;21(1):8, (2023).

43- Drapkina YS, Makarova NP, Vasiliev RA, Amelin V V, Kalinina EA. "Comparison of predictive models built with different machine learning techniques using the example of predicting the outcome of assisted reproductive technologies." *Obstet Gynecol.*;(2):97–105, (2024).

44- Peng J, Geng X, Zhao Y, Hou Z, Tian X, Liu X, *et al.* "Machine learning algorithms in constructing prediction models for assisted reproductive technology (ART) related live birth outcomes." *Sci Rep*.;14(1):32083, (2024).