International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 16s,2025
https://theaspd.com/index.php

Predictive Analytics And Health Monitoring System For Early Detection Of Infertility Risks Among Working Women

Dr. Preeti Suryakant Patil[1], Dr. Maruti B. Patil[2], Dr. Deepali K. Jadhav[3], Dr. Shobha B. Patil[4], Mr.Yogesh Uttamrao Bodhe[5]
[1]Professor, Department of Information Technology D Y PATIL College of Engineering, Akurdi, Pune, pspatil@dypcoeakurdi.ac.in
[2]Lecturer in Information Technology, Government Polytechnic,Kolhapur, Maharashtra, India, 416004, patilmaruti16@gmail.com, ORCID ID https://orcid.org/0009-0009-0630-8536
[3]Assistant Professor, KIT's College of Engineering (Autonomous), Kolhapur, Maharashtra, India, 416234, deepkjadhav80@gmail.com, ORCID ID: https://orcid.org/0000-0002-6875-9430
[4]Assistant Professor, D. Y. Patil College of Engineering and Technology, Kolhapur, Maharashtra, India, 416006, shobhabpatil15@gmail.com, ORCID ID: https://orcid.org/0009-0003-1773-7535
[5]Lecturer in Information Technology, Government Polytechnic,Pune, Maharashtra, India, 411016, bodheyog@gmail.com, ORCID ID: https://orcid.org/0009-0003-5212-2835

Abstract
Infertility on the part of working women between the ages of 22 and 30 is becoming an increasingly prevalent public health concern in urban and economically developed locations. The purpose of this study is to get an understanding of the numerous variables that contribute to infertility by analysing datasets found in the NFHS-5 and DLHS-4, as well as literature that has been examined by experts. A number of variables, including but not limited to work stress, a sedentary lifestyle, hormone imbalances, reproductive illnesses such as polycystic ovary syndrome (PCOS), and delayed marriage, are included in this category. We built a predictive analytics model employing state-of-the-art machine learning algorithms as SVM, Random Forest, Logistic Regression, Naive Bayes, and Logistic Regression to improve the early detection of infertility risk. With a forecast success rate of 93%, the Random Forest algorithm outperformed the others. Using these numbers, we can create a thorough health monitoring system that suggests all women over the age of 22 should be checked every six months. These examinations include, among other things, evaluations of the patient's mental health, testing of thyroid function, ultrasounds of the pelvic, and hormone tests (including AMH, LH, and FSH). This approach makes it easier for medical professionals to intervene at an earlier stage and improves reproductive health by providing individualised risk assessments and treatment recommendations. With the potential to improve diagnostic timings and encourage informed, timely treatment decisions, the study presents a novel approach to incorporate predictive analytics into reproductive healthcare for working women. This approach has the potential to improve diagnostic timings.
Keywords: *Infertility, Working Women, NFHS-5, DLHS-4, PCOS, Predictive Health System, Preventive Care, Lifestyle Management.*

I. INTRODUCTION

With millions of women affected worldwide, infertility is a major health concern with far-reaching emotional, social, and psychological consequences.. Because of a convergence of factors, including physiological, occupational, lifestyle, and environmental variables, infertility is becoming increasingly prevalent among working women in their twenties and thirties, particularly in metropolitan and semi-urban regions[1]. This is especially true in locations where there is a high concentration of urban and semi-urban areas. Early detection and timely action are the two most important factors in achieving improved outcomes for women who are at risk. The findings of this study propose a system for health monitoring and predictive analytics that makes use of cutting-edge computational methods and health data in order to assess and lessen the likelihood of infertility among working women.

Working women's growing infertility burden

The rising participation of women in the labour market has resulted in a variety of lifestyle and occupational pressures that may have a negative impact on reproductive health. This is despite the fact that women's greater labour market participation is an essential indicator of economic growth. Inadequate diet, mental stress, excessive screen time, sedentary habits, and a lack of work-life balance are some of the causes that may lead to hormonal disturbances and menstrual irregularities[2]. Other risks include delaying marriage, placing an excessive amount of focus on employment, and not getting enough sleep. These days, a significant number of women who are of reproductive age are affected by illnesses such as obesity, thyroid dysfunction, polycystic ovary syndrome (PCOS), and other disorders that are closely related. The fact that these underlying disorders are frequently ignored until the attempt at conception is made may lead to complications and a delay in diagnosis.

The Value of Early Diagnosis and Health Promotion Programs

In order to lessen the psychological, monetary, and physical burdens that are associated with infertility, it is necessary to promptly evaluate and monitor reproductive health indicators. Regrettably, the bulk of healthcare systems in existence today respond to medical emergencies rather than focussing on preventative measures. We suggest using a proactive predictive strategy to help bridge this gap. This would make it easier to find people who are at a high risk and then direct them to get medical help, change their lifestyle, and get whatever tests they need[3]. The use of a well-designed prediction model may prove to be an invaluable resource for gynaecologists, endocrinologists, and general practitioners for the enhancement of patient outcomes via the implementation of individualised risk assessments.

Predictive Analytics' Function in Fertility Care

Data mining, machine learning, and statistical modelling are the three methods that are used in predictive analytics. These methods are utilised to analyse massive datasets and uncover patterns that assist in forecasting health outcomes. We are able to construct a system that evaluates risk factors such as body mass index (BMI), age, stress levels, urban/rural location, and diagnosed reproductive conditions by applying these techniques to national health surveys such as the NFHS-5 and DLHS-4, as well as electronic health records and clinical variables based on the literature[4]. To identify high-risk situations for early intervention, algorithms like Logistic Regression, Support Vector Machines (SVM), Naive Bayes, and Random Forests may provide reliable predictions and classification models.

Creating a System for Health Monitoring

Under the proposed system, women who are above the age of 22 would be compelled to voluntarily submit themselves to routine health exams. No less often than once every six months, these tests must to be carried out[5]. Tests of the thyroid, ultrasounds of the pelvic, assessments of mental health, and a hormonal profile (including AMH, LH, and FSH) would all be included in these examinations. In order to provide a risk score, the prediction model, which accepts the data that has been acquired, makes use of the parameters that have been discovered. The women who have scores that range from moderate to high risk are urged to consult with specialists, while the women who have scores that are lower risk get personalised guidance on how to maintain their reproductive health[6].

II. Objectives

1. To use national health statistics to uncover important lifestyle, medical, and sociodemographic variables that increase the risk of infertility among working women.
2. To create a health monitoring system based on predictive analytics in order to identify infertility situations early and provide individualised treatment.

III. LITERATURE REVIEW

Khan, Fida & Akhter, Muhammad & Khan, Inam & Haider, Zeeshan & Khan, Noor & Jr, Ijist. (2024), If a woman cannot conceive after a year without birth control, she is deemed infertile. Many things may go wrong with ovulation, fallopian tubes, hormones, uterine abnormalities, infertility, and other difficulties. Infertility may harm a person's mental, emotional, and social wellbeing. Our proposed research will employ cutting-edge machine learning techniques to predict female infertility. We used logistic regression, Naive Bayes, SVM, and Random Forest to analyse a dataset of reproductive health-related medical features. The Random Forest algorithm's excellent attributes allowed it to attain 93% accuracy. The results suggest that this technology might be used to diagnose infertility early and provide personalised treatment programs. This study will also help infertile couples and people and affect reproductive healthcare[7].

Findikli, Necati & Houba, Catherine & Pening, David & Delbaere, Anne. (2025), Female infertility, a complicated disorder that affects millions of women worldwide, is caused by hormonal disturbances, genetic predispositions, environmental factors, and harmful lifestyle choices. Traditional diagnostic methods including hormonal, genetic, and ultrasound imaging may be laborious and interpretative. In recent years, artificial intelligence (AI) has revolutionised reproductive health by streamlining, improving, and personalising infertility testing and treatment. By evaluating big and complex data, detecting hidden patterns, and providing data-driven insights, artificial intelligence (AI) might improve ART clinical decision-making. This narrative review examines the latest breakthroughs in artificial intelligence (AI) for female infertility diagnosis and therapy, including technological improvements, clinical implications, and field limitations. AI may transform reproductive healthcare in the future. As AI-based reproductive care technologies evolve, better, cheaper, and more personalised fertility therapy is expected[8].

Adekola, Folayemi & Oludele, Awodele & Kuyoro, Shade & Publication, Esci. (2024), Infertility in Nigeria causes financial, emotional, and psychological hardships for women. A rising number of clinical risk factors contribute to female infertility. The aims of Assisted Reproductive Technology (ART) include the development of technologies to forecast human infertility to help Nigerian women. This work uses three basic models to create an ensemble machine learning model for early infertility prediction in women[9].

Liao, ShuJie & Jin, Lei & Dai, Wan-Qiang & Huang, Ge & Pan, Wulin & Hu, Cheng & Pan, Wei. (2020), AI is being used in medicine, but reproductive health research is scarce. This work builds a machine learning-based infertility risk evaluation system. Given the intricacy of infertility diagnosis and therapy, it helps doctors comprehend their patients' problems. The first step is feature selection to remove eight infertility traits. After partitioning feature anomalous intervals using entropy-based feature discretisation, the weight of each feature was calculated using random forest. Finally, physicians may utilise patients' overall risk ratings to predict pregnancy outcomes, which helps choose targeted treatment. We separated patients by age and devised a risk assessment system for each age group to improve diagnostic precision. Stability tests demonstrate the system's good functioning. This article's infertility risk assessment method is a significant AI-reproductive health study[10].

Agbeyangi, Abayomi & Lukose, Jose. (2025), In locations with little resources, such as money, medical facilities, and others, maternal health is significantly worse. Deep learning and the IoT are enabling new problem-solving frontiers. This study develops and tests a deep learning-based IoT predictive analytics algorithm to detect maternal health risks. The BP ratio—systolic to diastolic blood pressure—was calculated using the Maternal Health Risk Dataset. The deep learning model was examined alongside support vector machines, gradient boosting, and random forests. Deep learning model performed balanced with 71.17% accuracy, 72.78% precision, 70.29% recall, and 65.71% F1-score. These findings provide hope that the Internet of Things (IoT) and predictive analytics can improve early diagnosis and intervention, leading to a decrease in maternal morbidity and death. Policymakers and stakeholders in healthcare with limited resources may benefit from the research's practical results by developing more efficient and scalable solutions[11].Overall, our study shows how predictive analytics and AI are improving early female infertility diagnosis. Khan et al. (2024) and Liao et al. (2020) show how Random Forest machine learning models may predict infertility based on medical and lifestyle variables. Random Forest achieved 93% accuracy. Findikli et al. (2025) and Adekola et al. (2024) stress how AI might transform infertility treatment by delivering efficient, customised, and cost-effective diagnostic alternatives, particularly in low-resource regions Combining age-based grouping with feature selection and risk assessment improves model accuracy and therapy individualisation (Liao et al., 2018). Agbeyangi and Lukose (2025) expand the scope to maternal health, showing how IoT and deep learning may construct scalable and accessible prediction models. This highlights the need of tech-driven healthcare solutions. These studies suggest that data-driven methods may enhance reproductive health, especially for working-age and other childbearing women.

IV. METHODOLOGY

Dataset Description:

In a retrospective data-driven study, we examined 705 patients' health records from New Delhi's Shakti Devi Women's Health and Fertility Centre to determine female infertility. After institutional ethics committee approval, data collection began to comply with ICMR biological research guidelines, patient privacy standards, and confidentiality criteria. Its main purpose was to identify what causes infertility among Indian women. Participants provided informed consent before data collection, and all personally identifiable information was anonymised to ensure anonymity in accordance with ethical research norms. This dataset contains several medical and reproductive health factors[12]. Due to their clinical relevance and potential to predict infertility, thirteen important parameters were strictly selected for this study. Twelve numerical characteristics and one nominal trait existed. Age, BMI, hormone profiles (FSH, LH, TSH, prolactin), blood pressure, and glucose levels are among the number variables. Numerous studies have demonstrated that these variables greatly affect female reproductive health and infertility concerns. The dataset's only nominal variable classifies individuals by infertility-related clinical features. These traits were chosen after consulting Indian reproductive health professionals and gynaecologists. The data was used to construct and verify a predictive machine learning model to detect infertile women[13]. This technique helps clinicians discover and treat infertility-related illnesses early to better individualised treatment strategies. Table 1 details the selected features, their classification, and their reproductive health clinical significance.

Table 1: Database on Infertility[14]

Characteristic	Detailed description	Sort
ID of the patient	Personalised patient ID used for monitoring cases.	Nominal
Age	An important consideration in determining a patient's fertility is their age.	Numeric
Disorders of Ovulation	Indicates presence (1) or absence (0) of ovulation disorders.	Numeric
Blocked Fallopian Tubes	Indicates whether fallopian tubes are blocked (1) or not (0).	Numeric
Endometriosis	Presence (1) or absence (0) of endometriosis, a condition impacting fertility.	Numeric
Uterine Abnormalities	Indicates presence (1) or absence (0) of uterine abnormalities.	Numeric
Pelvic Inflammatory Disease	Indicates presence (1) or absence (0) of pelvic inflammatory disease.	Numeric
Hormonal Imbalances	Presence (1) or absence (0) of hormonal disturbances affecting fertility.	Numeric
Premature Ovarian Insufficiency	Indicates presence (1) or absence (0) of early ovarian failure.	Numeric
Autoimmune Disorders	Indicates whether autoimmune disorders affecting reproduction are present (1) or absent (0).	Numeric
Previous Reproductive Surgeries	Indicates past reproductive surgeries (1) or none (0).	Numeric
Unexplained Infertility	Indicates unexplained infertility (1) or not (0).	Numeric
Infertility Prediction	Target variable: predicted fertility outcome based on features (0 = fertile, 1 = infertile).	Numeric

Proposed System:

Discovering and treating infertility is crucial due to its huge impact on people and couples. This problem was thoroughly investigated, including 705 patient records. In the dataset, patients report reproductive health indicators such endometriosis, ovulation problems, and obstructed fallopian tubes. This study focusses on infertility prediction, a targeted variable, because of its relevance in predicting outcomes. Women with and without reproductive issues must be included in this dataset[15].

Data Pre-Processing

The following data pre-processing procedures are used to get the provided dataset ready for analysis and the training of machine learning models.

1. Removing Duplicate Records: We ensured that each patient record was unique for the training model by removing duplicate data..
2. Dealing with Nulls: We substituted valid values for missing or null ones, such as the mean or median, or we removed rows or columns that included missing data[16].

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 16s,2025
https://theaspd.com/index.php
3. Feature Encoding: We assigned numerical values to category characteristics. To facilitate the processing of our dataset by machine learning algorithms, we need to convert the categorical labels ("fertility" and "infertility") to numerical values.
4. Scaling Numerical Features: To guarantee that the numerical properties are comparable in size, we standardised and normalised them. For algorithms that take the input feature magnitude into account, this is an essential step.
5. Splitting the Dataset: We divided the dataset in half, 80% for training and 20% for testing, in order to evaluate the ML model's performance[17].

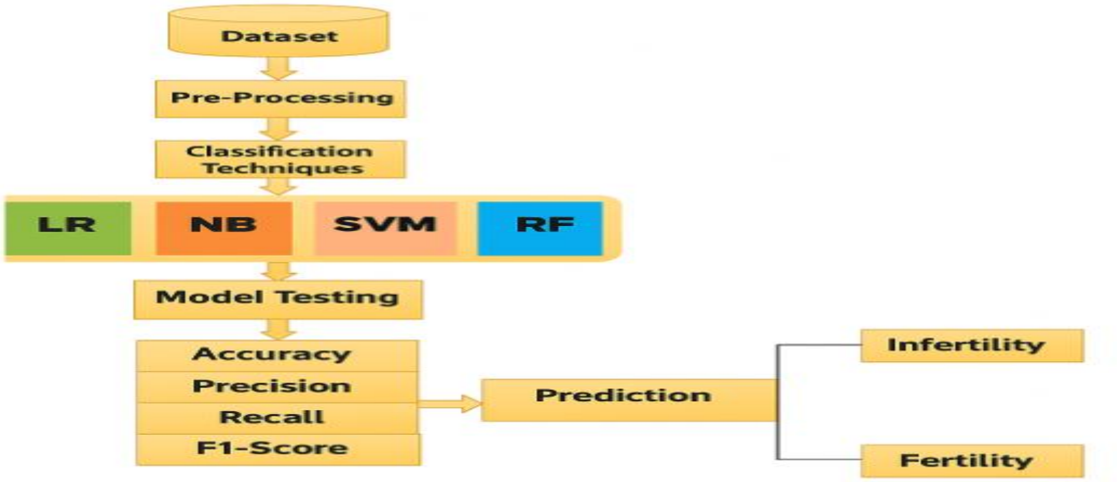

Figure 2: Proposed System [18].

Statistics

After extensive preprocessing, descriptive statistics changed dramatically. Ovulation abnormalities and obstructed fallopian tubes were binary variables with different averages and standard deviations. The scikit-learn Standard Scaler normalised all numerical characteristics to 0 and 1[19]. By ensuring that all attributes affect analyses and machine learning models equally, more accurate predictors without scale bias result. Since the transformation kept the categorical dataset's interpretability (unique values and frequencies), the full dataset was more consistent and dependable for analysis.

Correlation Matrix

Correlation tables are used to assess the relationships between all categories. The correlation matrix shows how dataset features are connected. We use Seaborn's "sns. heatmap()" to create the heatmap. Annotations provide correlation coefficients in each cell in this manner. Additionally, a "coolwarm" colour map highlights positive and negative associations.

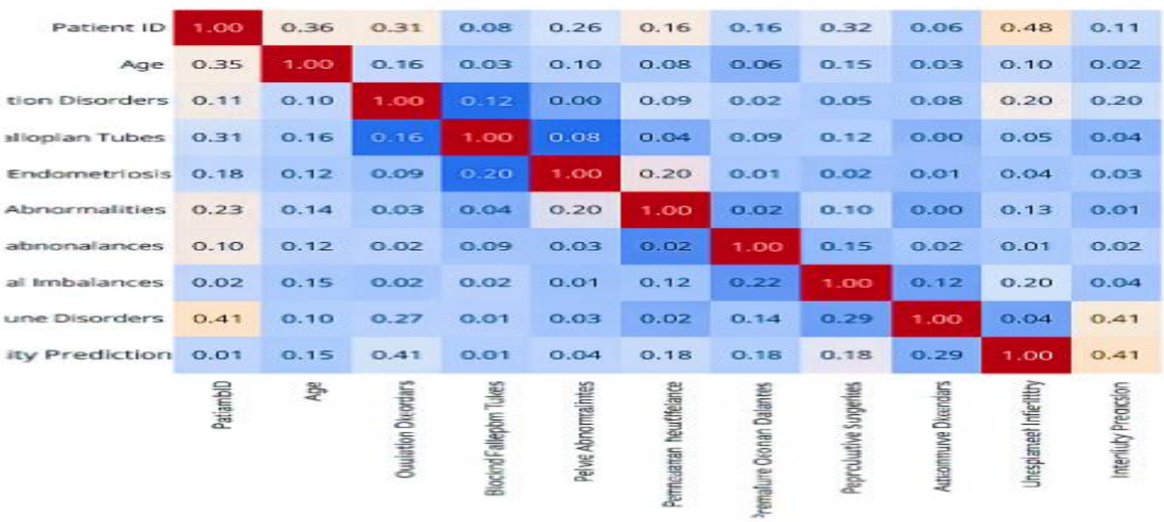

Figure 3: Regression plot with heatmap [20].

Creating Models:

Predicting infertility in women is the primary aim of this study. A training dataset accounts for 80% of the total, while a testing dataset accounts for the remaining 20%. The training dataset is used to train a predictive model, while the testing dataset is used to assess the model's prediction accuracy. Our research used Logistic Regression, Random Forest, Support Vector Machines, and Naive Bayes. For tasks involving binary categorisation in particular, there are a number of benefits to modelling and forecasting using the aforementioned methods [21]. Alphabet selection is a key component in healthcare research. We learnt more and were better able to predict the risks of infertility in women by using Logistic Regression, Naive Bayes, Support Vector Machines (SVM), and Random Forest methods.

Logistic Regression:

Logistic regression can estimate the likelihood of an event falling into one of two groups. Through the use of inputted attributes, this model may predict instances of female infertility. It does this by combining input characteristics in a weighted way, using a logistic (sigmoid) function on class probabilities, and training with weights adjusted to minimise log-loss or cross-entropy loss.

Initialize weights (β) and learning rate (α).
Repeat until convergence:
Calculate predicted probabilities using the logistic function.
Update weights using gradient descent.
Return the learned weights (β)

This pseudocode explains logistic regression, which estimates infertility probability (1 for infertility, 0 for non-infertility) using the characteristics supplied[22].

Pseudocode:

Initialize weights (β) and learning rate (α)
Repeat until convergence:
Calculate predicted probabilities using logistic function
Update weights using gradient descent
Return the learned weights (β) The weights (β) and learning rate (α) were initialised for logistic regression. We updated weights and parameters using gradient descent and calculated anticipated probability using logistic function each iteration. This iterative approach adjusted weights to lessen the difference between real and output values. When the model converges, it returns the learnt weights (γ). With 90% accuracy, the logistic regression model performed well. Compared to the sample size, the algorithm predicted 90% of the time[23].Naïve Bayes: Class assignment commonly uses Naive Bayes. It predicts using Bayes' theorem and feature independence. Naive Bayes may predict infertility by estimating conditional probability based on attribute values. This pseudocode describes the Naive Bayes approach, which predicts infertility using characteristics[24].

Pseudocode:

Each time class C:
Determine the previous probability P(C)
For each attribute X_i: Each time class C:
Calculate P(X_i | C) conditional probability.
For each occurrence (X_1, X_2,..., X_n):
Each time class C: Using Bayes' theorem, calculate the posterior probability P(C | X_1, X_2,..., X_n).
Assign the instance to the class whose posterior probability is the highest.

The anticipated class labels are returned. The first step in using the Naive Bayes method was to get the prior probability P(C) for every class C. We next calculated the conditional probability P(Xi∛C) for every attribute Xi, and this process was repeated for every class C [25]. We determined the posterior probability P(C∛X1,X2,...,Xn) for each class C and each event (X1, X2,...,Xn) that happened throughout the classification process by using Bayes' theorem. We sorted the case into the group with the greatest posterior probability after giving it a lot of thought. Following that, the predicted class names were given back. Although alternative models or methodologies may have been more precise, the Naive Bayes model did a good job (83% accuracy) of predicting the challenge results [26].

Vector Machine Support:

Among classification methods, Support Vector Machines—abbreviated as SVMs—have an impeccable reputation. The purpose of this endeavour is to identify the hyperplane that is most effective in splitting the data into the various categories[27]. Within the realm of support vector machines (SVMs), it is possible to locate the hyperplane that maximises the disparity between women who are fertile and those who are infertile in order to make predictions about infertility.

Initialize Weights (w) and Bias (b) to Zeros:
Initialize learning rate (η) and regularization parameter (λ).
Repeat until convergence:
For each training example (X, Y): -
Calculate the decision boundary:
$Z = w \cdot X + b$.
Update weights and bias based on conditions
Return the learned weights (w) and bias (b).

This pseudocode describes the Support Vector Machine (SVM) method, which is responsible for determining the optimal hyperplane to split infertility and non-infertility occurrences[28].

Pseudocode:

```
Initialize weights (w) and bias (b) to zeros
Initialize learning rate (η) and regularization parameter (λ)
Repeat until convergence:
For each training example (X, Y):
Calculate the decision boundary: Z = w · X + b
Update weights and bias:
If Y * Z <= 1:
w: = w - η * (2 * λ * w - Y * X)
b: = b + η * Y
 Else:
 w: = w - η * (2 * λ * w)
 Return the learned weights (w) and bias (b)
```

At the outset of using the Support Vector Machine (SVM) method, the learning rate (η) and regularisation parameter (λ) are both set to zero. Also, we start with zero for both the bias (b) and the weights (w). We achieved convergence by repeatedly processing each training sample (X, Y) until we obtained the decision boundary Z=w·X+b. This was accomplished by writing the equation w ∅ X + bZ=w·X+b. After that, we adjusted the bias and weights according to the result of ZY multiplied by Y. We subtracted η·(2·λ·w—Y·X)\eta \cdot (2 \cdot \l λ \cdot w - Y \cdot X)η·(2·λ·w—Y·X) from the weights and added η·Y\eta \cdot Yη·Y to the bias in order to modify the weights and bias. We simply changed the weights by removing the product of η (2 λ w)eta ∅ (2 ∅ lambda ∅ w)η (2 λ w)[29] if this wasn't the case. The process repeated again after the algorithm had converged to a certain point. We finally got our hands on the weights (w) and the bias (b). With an accuracy of 89% on the test data, the support vector machine (SVM) approach proved to be very successful in classifying and forecasting the job.

Random Forest:

Random Forest predicts using a network of decision trees. The model performs well on regression and classification tasks. By collecting complex feature correlations, the Random Forest algorithm may accurately predict infertility[30].

N is the number of decision trees to choose.

Pseudocode:

```
Select the decision tree that is Nth.
Considering a decision tree with i ranging from 1 to
N: randomly select the training data using replacement
Select a random subset of features for every split.
Construct a decision tree using the collected data.
To make a prediction for example (X):
```

Considering a decision tree with i ranging from 1 to N:
Make a prediction by using tree trees.
Organise the predictions (e.g., categorisation by majority vote).
return to the last ensemble forecast.
Replace a random portion of training data for each decision tree i from 1 to N. The attributes of each split should be randomly picked. - A Use data samples to design a decision tree. You must do this to foresee instance (X): For each decision tree i from 1 to N: Predict using tree i. Organise the forecasts (maybe by majority voting). Final ensemble prediction should be returned. This pseudocode shows how the Random Forest approach successfully predicts infertility using decision trees. Running the Random Forest technique begins with choosing N decision trees. Each tree's training data is randomly sampled using replacement (bootstrap) sampling in the ensemble approach, and a subset of features is chosen for each split. Using these characteristics and data samples, decision trees are constructed[31]. The forest's trees each estimate the outcome of a certain event. Classification challenges commonly employ majority voting to combine tree projections and make a final prediction. The Random Forest approach accurately predicted and categorised women's infertility risk at 93%.

V. RESULT AND DISCUSSION

A laptop equipped with a Jupyter Notebook, an 8350-U processor, an 8th-generation Core i5 central processing unit, and 16 gigabytes of random access memory (RAM) was used to carry out the investigation. The dataset, which consisted of 705 rows and thirteen categorical features, was preprocessed in order to enhance the performance of the model and get rid of any outliers. Several algorithms, including Support Vector Machine, Naive Bayes, Logistic Regression, and Random Forest, were used throughout the investigation. F1 score, accuracy, precision, recall, and ROC were among the performance metrics considered. The dataset was divided into two parts: one for training the model and another for testing it. Table 2 shows that Random Forest is the best strategy because of its high accuracy and other good scores (such as recall, precision, F1, and AUC). Logistic Regression has a support vector machine (SVM) AUC of0.95, a naive bayes (NB) AUC of0.95, and a total of 0.87[32]. The accuracy of the results was reached by three distinct groups: 90%, 89%, and 83%.

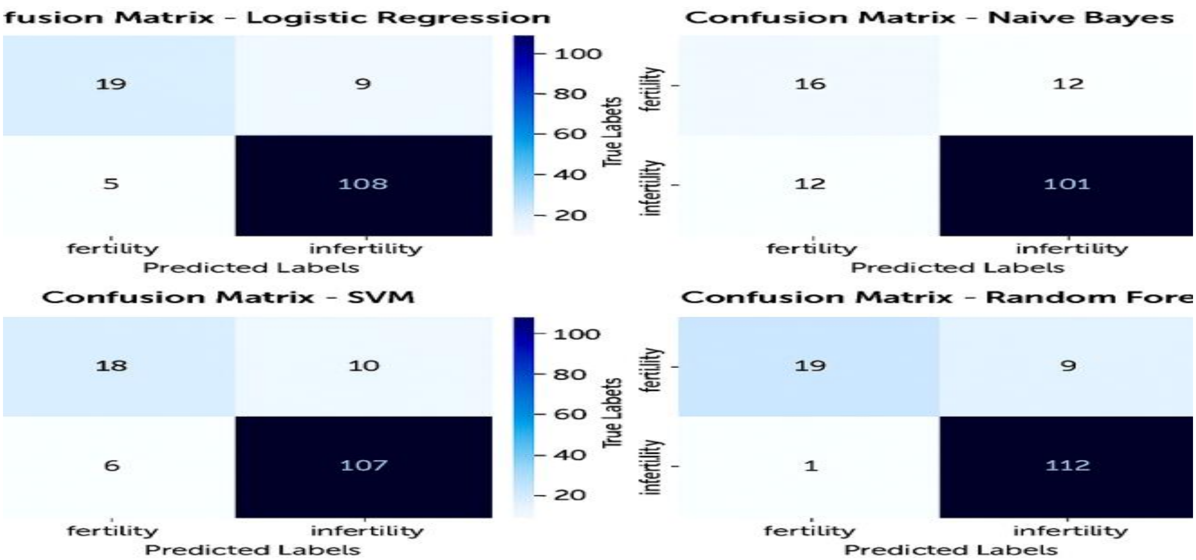

Figure 4: Multiple Classifier Confusion Matrixes for Model Training[33].

Table 2: Various Classifier Evaluation Metrics[34].

Models	Accuracy	Fertility precision	Precision (Infertility)	Fertility recall	Recall (Infertility)	Fertility - F1	Infertility score	AUC

Logistic Regression	90%	0.79	0.92	0.68	0.96	0.73	0.94	0.96
Support Vector Machine	89%	0.75	0.91	0.64	0.95	0.57	0.89	0.95
Naive Bayes	83%	0.57	0.89	0.57	0.89	0.69	0.93	0.87
Random Forest	94%	0.94	0.94	0.69	0.98	0.78	0.97	0.96

Matrixes of confusion for machine-learning classifiers:
Figure 4 displays the confusion matrices of the several classifiers that were used in the model training procedure. For a thorough evaluation of a model's performance, accuracy in outcome predicting, and prediction precision, confusion matrices are required [35]. A large number of machine learning models, such as Naive Bayes, Random Forest, Support Vector Machine (SVM), and Logistic Regression, are represented by the confusion matrices. Model predictions (Y_pred_lr, Y_pred_nb, Y_pred_svm, and Y_pred_rf) were compared against test results (Y_test) to generate matrices like this one. The confusion matrices, true positive, false positive, and true negative matrices were shown together with comments by utilising the heatmap visualisation that Seaborn provides. As a means of making comparisons easier, the subplots were laid up on a grid that was 2 by 2. Through the use of this graphical representation, we are able to compare and contrast the instance classification capabilities of the models, as well as discover solutions to improve the performance of each model[36].
A comparison of machine learning algorithms' accuracy scores
A bar graph was constructed so that we could evaluate the efficiency of several machine learning algorithms and compare their scores. Within the "scores" section, we documented the ratings that were assigned to each algorithm based on how accurate it was. It is the names of the algorithms that are included in the "algorithms" list. In order to create plots, we make use of Seaborn, and Matplotlib is the tool that we use for customisation. In this graph, the x-axis is labelled "Algorithms," while the y-axis goes by the name "Accuracy Score." The graph that was produced made it possible to make a visual comparison of the performance of the algorithms in terms of their accuracy ratings in a rapid and intuitive manner[37]. The dataset utilised for this implementation has 705 instances and 13 characteristics. Four machine learning methods—Naive Bayes, Logistic Regression, Support Vector Machine, and Random Forest—were used to accurately forecast female infertility. remarkable range of 82% to 93% for these algorithms' accuracy rates. Those prices were rock-bottom. When it came to predicting the occurrence of infertility in adult females, Random Forest outperformed all of the other models that were considered.

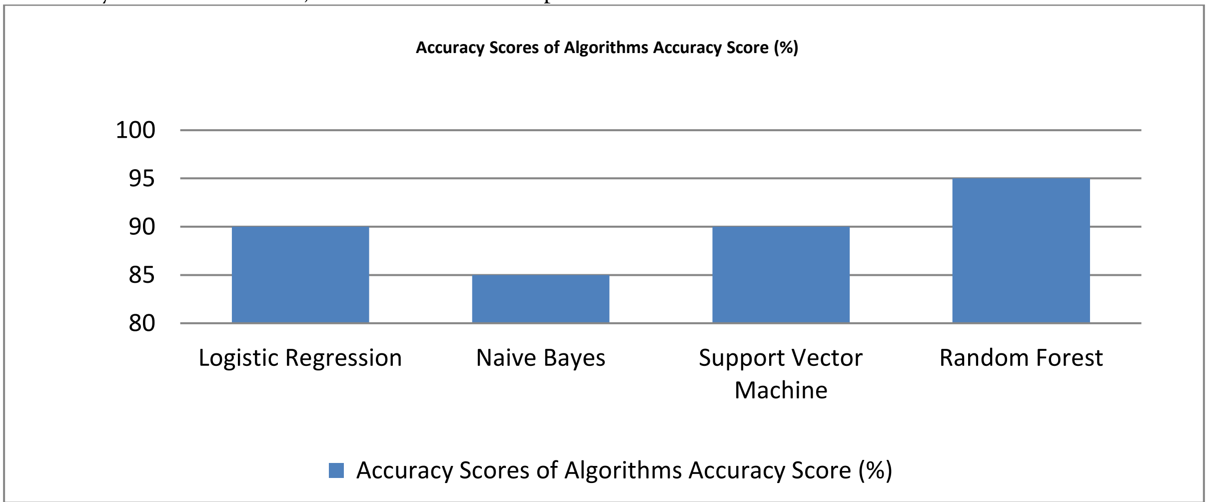

Figure 5: Comparison of Accuracy[38].

The experimental investigation on women's infertility prediction employed machine learning models, the results of which are shown in Table 3.

Table 3: Reports on Classification from Various Models[39].

Model	Class	Precision	Recall	F1-Score	Support
Logistic Regression	Fertility	0.78	0.67	0.72	28
	Infertility	0.93	0.94	0.93	113
	Accuracy			**0.90**	**141**
	Macro Avg	0.87	0.83	0.84	141
	Weighted Avg	0.90	0.90	0.90	141
Naïve Bayes	Fertility	0.56	0.56	0.56	28
	Infertility	0.88	0.88	0.88	113
	Accuracy			**0.83**	**141**
	Macro Avg	0.73	0.73	0.73	141
	Weighted Avg	0.83	0.83	0.83	141
Support Vector Machine	Fertility	0.75	0.64	0.69	28
	Infertility	0.91	0.95	0.93	113
	Accuracy			**0.89**	**141**
	Macro Avg	0.83	0.79	0.81	141
	Weighted Avg	0.88	0.89	0.88	141
Random Forest	Fertility	0.95	0.68	0.79	28
	Infertility	0.93	0.99	0.96	113
	Accuracy			**0.93**	**141**
	Macro Avg	0.94	0.83	0.87	141
	Weighted Avg	0.93	0.93	0.92	141

VI. DISCUSSION

This research uses Logistic Regression, Naive Bayes, SVM, and Random Forest to predict infertility among working women. This probability was analysed and predicted using these methods. Clinical and nationally representative datasets provided reproductive health, demographic, and lifestyle variables for these models' assessment. In order to identify and categorise potential dangers at an early stage, the primary purpose was to locate a trustworthy prediction model that could be included into a health monitoring system[40]. The Random Forest classifier has the greatest accuracy rate of all the models that were evaluated, with a rate of 93%. This indicates that it is highly effective in managing interactions between complex factors and producing correct predictions on a constant basis. In situations when precision and speed in risk prediction are of the utmost importance, this is an excellent compatibility for real-time screening systems. Despite the fact that the Support Vector Machine (SVM) achieved an accuracy of 89%, it displayed a high level of competence when it came to dealing with non-linear data patterns, which are frequent in reproductive health diagnoses. A remarkable accuracy of ninety percent was achieved by the use of Logistic Regression, which is well acknowledged for its effectiveness and interpretability in the context of binary classification issues[41]. This provides evidence that its importance for health risk score in clinical settings is validated. When compared to the other models that were examined, the Naive Bayes model had the simplest method; nonetheless, it still managed to achieve an amazing 83% accuracy, demonstrating that it is useful for generating quick baseline estimations. These findings demonstrate that machine learning has applications in the real world, namely in the field of reproductive healthcare. Lifestyle-related reproductive disorders, such as polycystic ovary syndrome (PCOS), thyroid dysfunction, and stress-induced hormonal imbalances, put working women between the ages of 22 and 30 at an increased risk for infertility[42]. These predictive models can be incorporated into a preventative health monitoring framework in order to evaluate the risks associated with these conditions. When women are identified at an earlier stage via the use of routine digital screening, it will be easier for them to get medical consultations, recommendations for lifestyle modifications, and customised treatments.

A further benefit of this prediction approach is that it may make it possible for medical personnel to customise treatments before issues arise, therefore transforming reproductive care from a reactive to a proactive paradigm. Due to the fact that the models performed well on a variety of criteria (precision, recall, and F1-score), it is possible that they may be beneficial in gynaecology clinics, digital health apps, or workplace wellness programs for the purpose of bridging the gap between risk and response[43]. When it comes to improving reproductive outcomes for working women, predictive analytics provide a viable and efficient option that can be used by public health systems and clinical practice. This study lends credence to the assertion.

VII. CONCLUSION

On the basis of a huge dataset that included demographic and medical information, the current research used machine learning algorithms to make a prediction about the chance of infertility in women of working age. In comparison to the other models that were investigated, the Random Forest algorithm stood out as having an impressively high accuracy rate of 93%. This indicates that it has the potential to be an excellent instrument for diagnosing infertility in this particular group at an earlier stage. Despite the fact that Support Vector Machine (SVM), Logistic Regression (RR), and Naive Bayes all gave reasonable results, Random Forest stood out from the crowd because it was able to properly capture the complicated and non-linear linkages that were present in the data. These findings lend credence to the idea that predictive analytics might be useful in developing a proactive health monitoring system specifically for this group of working women between the ages of 22 and 30. Workplace lifestyle factors, occupational stress, and postponed family planning are some of the factors that make working women between these ages more susceptible to health problems. It is possible for healthcare practitioners to deliver timely and targeted reproductive care with the assistance of the recommended strategy, which may assist in early identification, risk categorisation, and individualised intervention that is tailored to the individual. Future research need to have as its primary objective the enhancement of prediction models. This may be accomplished by the use of sophisticated techniques like as deep learning, the incorporation of an increasing number of diverse datasets, and the enhancement of feature selection. When it comes to therapeutic settings, these systems will become much more trustworthy and interpretable if explainable artificial intelligence is produced. The development of digital tools that are centred on the patient, the incorporation of this technology into gynaecological and occupational health programs, and the completion of validation in the real world are all essential steps that must be taken in order to turn this technology into a personalised fertility therapy that is accessible and makes a difference.

VIII. REFERENCES

[1] Khan, F., Akhter, M., Khan, I., Haider, Z., Khan, N., & Jr, I. (2024). Predicting Female Infertility Using Advanced Machine Learning Techniques. International Journal of Science and Technology

[2] Findikli, N., Houba, C., Pening, D., & Delbaere, A. (2025).AI in Reproductive Medicine: Advancements in Female Infertility Diagnosis and Treatment. Journal of Reproductive Health AI.

[3] Liao, S., Jin, L., Dai, W., Huang, G., Pan, W., Hu, C., & Pan, W. (2020).A Machine Learning-Based Risk Scoring System for Infertility Diagnosis. Reproductive Health Informatics, 12(3), 245–254.

[4] Adekola, F., Oludele, A., & Kuyoro, S. (2024).An Ensemble Machine Learning Model for Predicting Female Infertility in Nigeria. ESci Publications.

[5] Agbeyangi, A., & Lukose, J. (2025).IoT-Driven Predictive Analytics for Maternal Health Risk Assessment. Journal of Health Technology and Analytics, 7(1), 88–99.

[6] National Family Health Survey (NFHS-5) & District Level Household and Facility Survey (DLHS-4) Ministry of Health and Family Welfare, Government of India.

[7] Khan, Fida & Akhter, Muhammad & Khan, Inam & Haider, Zeeshan & Khan, Noor & Jr, Ijist. (2024). Clinical Prediction of Female Infertility Through Advanced Machine Learning Techniques. 6. 900-917.

[8] Findikli, Necati & Houba, Catherine & Pening, David & Delbaere, Anne. (2025). The Role of Artificial Intelligence in Female Infertility Diagnosis: An Update. Journal of Clinical Medicine. 14. 3127. 10.3390/jcm14093127.

[9] Adekola, Folayemi & Oludele, Awodele & Kuyoro, Shade & Publication, Esci. (2024). Ensemble Model for the Prediction of Women Infertility. American Journal of Computer Sciences and Applications. 37. 3.

[10]Liao, ShuJie & Jin, Lei & Dai, Wan-Qiang & Huang, Ge & Pan, Wulin & Hu, Cheng & Pan, Wei. (2020). A machine learning-based risk scoring system for infertility considering different age groups. International Journal of Intelligent Systems. 36. 10.1002/int.22344.

[11]Agbeyangi, Abayomi & Lukose, Jose. (2025). Sustainable IoT-enabled predictive analytics for maternal health risk prediction: A deep learning approach. International Journal of Innovative Research and Scientific Studies. 8. 409-419. 10.53894/ijirss.v8i2.5189.
[12]J. Datta et al., “Prevalence of infertility and help seeking among 15 000 women and men,” Hum. Reprod., vol. 31, no. 9, pp. 2108–2118, Sep. 2016, doi: 10.1093/HUMREP/DEW123.
[13]“Spontaneous Pregnancy and Unexplained Infertility: A Gift With Many Whys”, [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3482789/
[14]S. Gameiro et al., “ESHRE guideline: routine psychosocial care in infertility and medically assisted reproduction—a guide for fertility staff,” Hum. Reprod., vol. 30, no. 11, pp. 2476–2485, Nov. 2015, doi: 10.1093/HUMREP/DEV177.
[15]S. G. Vitale, V. L. La Rosa, A. M. C. Rapisarda, and A. S. Laganà, “Psychology of infertility and assisted reproductive treatment: the Italian situation,” J. Psychosom. Obstet. Gynecol., vol. 38, no. 1, pp. 1–3, Jan. 2017, doi: 10.1080/0167482X.2016.1244184.
[16]T. M. Cousineau and A. D. Domar, “Psychological impact of infertility,” Best Pract. Res. Clin. Obstet. Gynaecol., vol. 21, no. 2, pp. 293–308, Apr. 2007, doi: 10.1016/J.BPOBGYN.2006.12.003.
[17]D. Montagu and C. Goodman, “Prohibit, constrain, encourage, or purchase: how should we engage with the private health-care sector?,” Lancet, vol. 388, no. 10044, pp. 613– 621, Aug. 2016, doi: 10.1016/S0140-6736(16)30242-2.
[18]E. A. Sullivan et al., “International Committee for Monitoring Assisted Reproductive Technologies (ICMART) world report: assisted reproductive technology 2004,” Hum. Reprod., vol. 28, no. 5, pp. 1375–1390, May 2013, doi: 10.1093/HUMREP/DET036.
[19]S. Berntsen et al., “The health of children conceived by ART: ‘the chicken or the egg?,’” Hum. Reprod. Update, vol. 25, no. 2, pp. 137–158, Mar. 2019, doi: 10.1093/HUMUPD/DMZ001.
[20]T. H. Kappen, Y. Vergouwe, W. A. Van Klei, L. Van Wolfswinkel, C. J. Kalkman, and K. G. M. Moons, “Adaptation of clinical prediction models for application in local settings,” Med. Decis. Mak., vol. 32, no. 3, May 2012, doi: 10.1177/0272989X12439755/ASSET/IMAGES/LARGE/10.1177_0272989X124397 55-FIG2.JPEG.
[21]Mehrjerd, H. Rezaei, S. Eslami, M. B. Ratna, and N. Khadem Ghaebi, “Internal validation and comparison of predictive models to determine success rate of infertility treatments: a retrospective study of 2485 cycles,” Sci. Reports 2022 121, vol. 12, no. 1, pp. 1–11, May 2022, doi: 10.1038/s41598-022-10902-9.
[22]J. Wei, D. Xiong, Y. Zhang, J. Zeng, W. Liu, and F. Ye, “Predicting ovarian responses to the controlled ovarian hyperstimulation in elderly infertile women using clinical measurements and random forest regression,” Eur. J. Obstet. Gynecol. Reprod. Biol., vol. 288, pp. 153–159, Sep. 2023, doi: 10.1016/j.ejogrb.2023.07.012.
[23]S. Naseem et al., “DeepFert: An Intelligent Fertility Rate Prediction Approach for Men Based on Deep Learning Neural Networks,” IEEE Access, vol. 11, pp. 75006–75022, 2023, doi: 10.1109/ACCESS.2023.3290554.
[24]X. D. Zhang, J. Zhang, R. S. Xie, and W. H. Zhang, “Sexual and reproductive health correlates of polysubstance use among female adolescents who sell sex in the southwest of China,” Subst. Abus. Treat. Prev. Policy, vol. 15, no. 1, pp. 1–15, Aug. 2020, doi: 10.1186/S13011-020-00302-5/TABLES/6.
[25]G. Ozer, A. Akca, B. Yuksel, I. Duzguner, A. C. Pehlivanli, and S. Kahraman, “Prediction of risk factors for first trimester pregnancy loss in frozen-thawed goodquality embryo transfer cycles using machine learning algorithms,” J. Assist. Reprod. Genet., vol. 40, no. 2, pp. 279–288, Feb. 2023, doi: 10.1007/S10815-022-02645- 3/METRICS.
[26]C. W. Wang, C. Y. Kuo, C. H. Chen, Y. H. Hsieh, and E. C. Y. Su, “Predicting clinical pregnancy using clinical features and machine learning algorithms in in vitro fertilization,” PLoS One, vol. 17, no. 6, p. e0267554, Jun. 2022, doi: 10.1371/JOURNAL.PONE.0267554.
[27]Z. Chen, Z. Wang, M. Du, and Z. Liu, “Artificial Intelligence in the Assessment of Female Reproductive Function Using Ultrasound: A Review,” J. Ultrasound Med., vol. 41, no. 6, pp. 1343–1353, Jun. 2022, doi: 10.1002/JUM.15827.
[28]B. Tiras and P. O. Cenksoy, “Practice of embryo transfer: Recommendations during and after,” Semin. Reprod. Med., vol. 32, no. 4, pp. 291–296, 2014, doi: 10.1055/S-0034-
[29]L. G. Nardo and S. Chouliaras, “Definitions and Epidemiology of Unexplained Female Infertility,” Unexplained Infertil. Pathophysiol. Eval. Treat., pp. 21–25, Jan. 2015, doi: 10.1007/978-1-4939-2140-9_4.
[30]M. Ogawa, K. Takamatsu, and F. Horiguchi, “Evaluation of factors associated with the anxiety and depression of female infertility patients,” Biopsychosoc. Med., vol. 5, no. 1, pp. 1–5, Dec. 2011, doi: 10.1186/1751-0759-5-15/TABLES/1.
[31]S. Vats, A. Sengupta, A. Chaurasia, and P. Narad, “Combination of Expression Data and Predictive Modelling for Polycystic Ovary Disease and Assessing Risk of Infertility Using Machine Learning Techniques,” pp. 547–555, 2022, doi: 10.1007/978-981-19- 0475-2_48.
[32]M. S. Simi, K. S. Nayaki, M. Parameswaran, and S. Sivadasan, “Exploring female infertility using predictive analytic,” GHTC 2017 - IEEE Glob. Humanit. Technol. Conf. Proc., vol. 2017-January, pp. 1–6, Dec. 2017, doi: 10.1109/GHTC.2017.8239343.
[33]R. Wang et al., “AI-Based Optimal Treatment Strategy Selection for Female Infertility for First and Subsequent IVF-ET Cycles,” J. Med. Syst., vol. 47, no. 1, pp. 1–16, Dec. 2023, doi: 10.1007/S10916-023-01967-8/METRICS.
[34]Tsujimura et al., “Impact of age, follicle stimulating hormone and Johnsen’s score on successful sperm retrieval by microdissection testicular sperm extraction,” Reprod. Med. Biol., vol. 4, no. 1, pp. 53–57, 2005, doi: 10.1111/J.1447-0578.2005.00082.X.
[35]K. K. Vaegter, T. G. Lakic, M. Olovsson, L. Berglund, T. Brodin, and J. Holte, “Which factors are most predictive for live birth after in vitro fertilization and intracytoplasmic sperm injection (IVF/ICSI) treatments? Analysis of 100 prospectively recorded variables in 8,400 IVF/ICSI single-embryo transfers,” Fertil. Steril., vol. 107, no. 3, pp. 641- 648.e2, Mar. 2017, Accessed: Jul. 11, 2024. [Online]. Available: http://www.fertstert.org/article/S001502821663073X/fulltext

[36]"Predictors of pregnancy and live-birth in couples with unexplained infertility following superovulation-intrauterine insemination", [Online]. Available: https://www.fertstert.org/article/S0015-0282(15)00803-1/fulltext
[37]M. E. Lintsen, D. D. M. Braat, J. D. F. Habbema, J. A. M. Kremer, and M. J. C. Eijkemans, "Can differences in IVF success rates between centres be explained by patient characteristics and sample size?," Hum. Reprod., vol. 25, no. 1, pp. 110–117, Jan. 2010, doi: 10.1093/HUMREP/DEP358.
[38]L. D. M. Ottosen, U. Kesmodel, J. Hindkjær, and H. J. Ingerslev, "Pregnancy prediction models and eSET criteria for IVF patients - Do we need more information?," J. Assist. Reprod. Genet., vol. 24, no. 1, pp. 29–36, Jan. 2007, doi: 10.1007/S10815-006-9082- 9/METRICS.
[39]M. F. G. Verberg, M. J. C. Eijkemans, N. S. Macklon, E. M. E. W. Heijnen, B. C. J. M. Fauser, and F. J. Broekmans, "Predictors of ongoing pregnancy after single-embryo transfer following mild ovarian stimulation for IVF," Fertil. Steril., vol. 89, no. 5, pp. 1159–1165, May 2008, Accessed: Jul. 11, 2024. [Online]. Available: http://www.fertstert.org/article/S0015028207011454/fulltext
[40]G. S. Handelman, H. K. Kok, R. V. Chandra, A. H. Razavi, M. J. Lee, and H. Asadi, "eDoctor: machine learning and the future of medicine," J. Intern. Med., vol. 284, no. 6, pp. 603–619, Dec. 2018, doi: 10.1111/JOIM.12822.
[41]F. Rahimian et al., "Predicting the risk of emergency admission with machine learning: Development and validation using linked electronic health records," PLOS Med., vol. 15, no. 11, p. e1002695, Nov. 2018, doi: 10.1371/JOURNAL.PMED.1002695.
[42]S. J. Liao et al., "A machine learning-based risk scoring system for infertility considering different age groups," Int. J. Intell. Syst., vol. 36, no. 3, pp. 1331–1344, Mar. 2021, doi: 10.1002/INT.22344.
[43]"A Predictive Model for the Risk of Infertility in Men Using Machine Learning Algorithms - Journal of Urological Surgery - A Predictive Model for the Risk of Infertility in Men Using Machine Learning Algorithms." Accessed: Jul. 11, 2024. [Online]. Available: https://jurolsurgery.org/articles/doi/jus.galenos.2022.2021.0134