International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

IVF Success Prediction using Machine Learning Techniques: A Comparative Study

Pooja Bagane[1], Saloni Kulshretha[2], Manasvi Mishra[3], Yashika Asrani[4], Vrinda Maheswari[5]

Submitted: 12/09/2023 **Revised:** 20/10/2023 **Accepted:** 06/11/2023

Abstract: IVF is a popularly used assisted reproductive therapy that aids couples who are having trouble getting pregnant naturally. Medical providers must be able to predict an IVF cycle's effectiveness to tailor their care and enhance outcomes. To forecast the likelihood of a successful IVF cycle, this study proposes a machine learning (ML) model based on the random forest approach.

A dataset of patient cycles was collected from the author of a research paper we came across while studying the topic. The dataset includes patient demographic and clinical variables, such as age, body mass index (BMI), semen test results, number of retrieved oocytes, and female and male infertility factors. After preprocessing and feature engineering, we used the random forest, support vector machine, gradient boosting, and logistic regression algorithms to build four classification models to predict IVF success and compared their results. The Gradient Boosting algorithm showed the highest accuracy of 87%, whereas the SVM model showed the least accuracy with 67%. The most important features for the prediction were age, number of retrieved oocytes, and embryo quality, consistent with previous studies. The research shows the potential of machine learning models for predicting IVF success, which can assist physicians in making wise choices and enhancing the results for patients. To confirm the generalizability of the approach, additional studies with larger datasets and more varied patient populations are required.

Keywords: *IVF, Machine Learning, Random Forest, Prediction, Assisted Reproductive Technology*

1. Introduction

The process of fertilising an egg with sperm outside the body in a laboratory setting is known as in vitro fertilisation (IVF), and it is one of the most common forms of assisted reproduction. The body mass index (BMI), ovarian reserve, patient age, and embryo quality all have a significant impact on how successful IVF is. Clinicians must accurately forecast IVF success in order to personalize patient care and enhance results.

The ability of machine learning to forecast the results of IVF cycles has shown considerable promise. Numerous studies have used patient demographic and clinical factors to predict IVF success using machine learning methods like logistic regression, support vector machines, and neural networks. However, no study has yet explored using a random forest algorithm for IVF prediction.

Random forest is an algorithm for ensemble learning which integrates multiple decision trees in order to improve the accuracy of predictions while minimising overfitting. It has been widely used in various domains, including medical diagnosis, financial prediction, and image recognition. We introduce a machine-learning model in this study that uses the random forest technique to forecast the likelihood that an IVF cycle will be successful.

[1,2,3,4,5] Department of Computer Science, Symbiosis Institute of Technology, (SIT) affiliated to Symbiosis International (Deemed University), Pune, India
ORCID ID : 0000-0001-9611-9601

** Corresponding Author Email: poojabagane@gmail.com*

2. Existing Methods

A frequent kind of ART, or assisted reproductive technology, used to help infertility couples or individuals conceive is in vitro fertilization (IVF). While IVF can be successful, there are many factors that can affect the likelihood of success. Researchers have investigated various patient- and treatment-related factors to help predict the chances of success with IVF. Additionally, predictive models have been developed to provide individualized estimates of success.

2.1. Medical Models:

For each patient undergoing IVF, various predictive algorithms have been devised to help determine their chances of success. These models incorporate a combination of patient-related and treatment-related factors to generate a personalized estimate of the chance of success. One commonly used model is the "CARE" (Clinical Assessment of the Reproductive Endocrinologist) model, which was developed using data from over 11,000 IVF cycles and incorporates factors such as age, ovarian reserve, and previous pregnancy history. Other models include the "Glasgow Prognostic Score" and the "ART Calculator." Several predictive models have been created to calculate the chances of IVF success.

Here are a few examples:

- ***Clinical Assessment of the Reproductive Endocrinologist (CARE) Model:*** The CARE model is one of the most used predictive models in IVF. It was developed using data from over 11,000 IVF cycles and incorporates factors such as age, ovarian reserve (as measured by AMH levels), and previous pregnancy history. The model generates a personalized estimate of the chance of success for each individual patient.
- ***Glasgow Prognostic Score (GPS):*** The GPS is another predictive model that uses a combination of patient-related and treatment-related factors to estimate the chance of success in IVF. The model considers factors such as age, BMI, AMH levels, and the number of embryos implanted. GPS has been shown to be effective in predicting IVF success in a variety of patient populations.
- ***ART Calculator:*** The ART Calculator is an online tool developed by the Society for Assisted Reproductive Technology (SART) that provides personalized estimates regarding IVF's likelihood of happening based on a patient's individual characteristics. The calculator considers factors such as age, BMI, AMH levels, and previous pregnancy history.
- ***Fertility Assessment and Consultation Tool (FACT):*** The FACT is a forecasting model created by academics at the University of California, San Francisco, that estimates the chance of success for both IVF and intrauterine insemination (IUI). The model considers factors such as age, BMI, and the number of previous cycles of ART.

It's important to note that while these predictive models can provide helpful information for patients and clinicians, they are not perfect and should be used in conjunction with clinical judgment and individualized counseling. IVF success rates can vary widely depending on individual factors, and no model can accurately predict outcomes with 100% certainty. Additionally, new factors may emerge over time that are not accounted for in current predictive models.

2.2. Machine learning Models:

Machine learning algorithms can be used to predict IVF success rates by analyzing large amounts to forecast the likelihood of IVF success by examining an assortment of data sources, including previous IVF cycles. These algorithms can identify patterns and relationships between various features and IVF outcomes, allowing them to make accurate predictions for individual patients.

Using algorithmic machine learning to predict IVF success entails several steps:

- ***Data gathering:*** Gathering information on IVF cycles is the initial approach, including patient characteristics, treatment protocols, and cycle outcomes. This data can come from electronic medical records, clinical databases, or research studies.
- ***Feature selection:*** Next, researchers must decide which features to include in the machine learning model. This involves selecting features that are known or suspected to be predictive of IVF success and features that may interact with each other to affect outcomes.
- ***Model training:*** Once the features are selected, the machine learning model must be trained using the data collected in step 1. This involves feeding the algorithm a set of labeled data (i.e., IVF cycles with known outcomes) and allowing it to learn the relationships between the features and outcomes.
- ***Model validation:*** After the model is trained, it must be validated to ensure accuracy and reliability. This is typically done by testing the model on a separate set of IVF cycles that were not used in the training process.
- ***Model deployment:*** Finally, if the model is found to be accurate and reliable, it can be deployed in clinical practice to help clinicians and patients make informed decisions about IVF treatment.

Forecasting models for IVF success rates have been created using artificial intelligence techniques. Following are a few instances of machine learning algorithms in use:

- ***Random Forest:*** Regression analysis can be performed using the well-liked machine learning method known as Random Forest. It functions by building several trees of decisions and combining their results to make predictions. Random Forest has been used to predict IVF success rates based on a variety of patient-related and treatment-related factors, including age, BMI, AMH levels, number of embryos transferred, and method of fertilization. It is an ensemble learning algorithm that can improve accuracy and reduce overfitting compared to individual decision trees. However, it can be computationally expensive and may not be as interpretable as other algorithms. In addition, random forest may not perform as well as other algorithms for small sample sizes or highly imbalanced data.
- ***Gradient Boosting:*** Gradient Boosting is another algorithm for artificial intelligence which is capable of being applied to regression analysis. It functions by building decision tree structures iteratively that are optimized to correct the errors of the previous trees. Gradient Boosting has been used to predict IVF success rates based on various factors, including age, AMH levels, and previous pregnancy history.

- ***Artificial Neural Networks (ANNs):*** This ML technique is built on the architecture and operation of the human nervous system, more specifically, the brain. ANNs have been used to predict IVF success rates based on various factors, including age, AMH levels, and the number of embryos transferred. ANNs can be particularly effective for handling complex, nonlinear relationships between variables. ANNs are a powerful class of algorithms that can learn complex relationships between features and outcomes. However, they can be computationally expensive and require large amounts of data to train effectively. In addition, neural networks can be difficult to interpret and may be sensitive to the choice of architecture and hyperparameters.

- ***Support Vector Machines (SVMs):*** SVMs, or Support Vector Machines, are primarily used for classification tasks, where the goal is to assign a given input to one of several categories or classes. While SVMs can be adapted for regression analysis, their primary use is in classification problems. SVMs operate by determining the best hyperplane for categorizing data points. SVMs have been used to predict IVF success rates based on various factors, including age, BMI, and AMH levels. SVM is a popular algorithm for binary classification tasks that can work well for highly dimensional feature spaces or small sample sizes. However, SVM can be sensitive to the choice of kernel function and regularization parameters and may not perform as well as other algorithms for non-linearly separable data. In addition, SVM may be less interpretable than some other algorithms.

- ***Logistic regression:*** Logistic regression is a simple and widely used algorithm for binary classification tasks, such as predicting IVF success or failure. It works by modeling the probability of success as a function of the input features. Logistic regression has been used successfully in several studies for IVF success prediction. Logistic regression is a relatively simple and interpretable algorithm that works well for binary classification tasks. However, it may not capture non-linear relationships between features and outcomes and may be sensitive to outliers or imbalanced data. In addition, logistic regression may not perform as well as more complex algorithms for highly dimensional feature spaces.

It's crucial to remember that even though machine learning algorithms can be great tools for forecasting IVF success rates, they depend on vast quantities of reliable data to function properly. Furthermore, machine learning models may be challenging to understand, which may limit their applicability in clinical settings. However, machine learning methods provide hope in this area and may be utilized to further develop and enhance models that predict IVF success rates. When predicting IVF success, machine learning algorithms provide several advantages over conventional statistical methods. They can consider non-linear correlations between predictors and outcomes as well as complicated interactions between features. They may also pick up new information as it comes out, which helps the model get better over time. Additionally, machine learning algorithms can help identify new predictors of IVF success that may not be apparent using traditional statistical methods.

3. Literature Survey

Artificial intelligence (AI) has become increasingly important in the field of in vitro fertilization (IVF), as it offers potential benefits for predicting and improving the success rates of fertility treatments. Several studies have explored the application of AI techniques in IVF prediction models and analysis. Here is a summary of the strengths and drawbacks of each study's findings:

1. "Artificial Intelligence a need in IVF" by Charalampos Siristatidis et al. (2021):

Strengths: The study recommends the use of a Learning Vector Quantizer (LVQ) classifier for IVF success prediction, based on previous data and the capacity to produce clusters.

Drawbacks: The authors only propose a theoretical model without implementing it.

2. "Predicting ongoing pregnancy chances after IVF and ICSI: a national prospective study" by A.M.E. Lintsen et al. (2007):

Strengths: The study analyzes success rates of pregnancy after IVF and ICSI, considering various variables such as age, subfertility, and pregnancy history. It provides insights into the impact of these factors on success rates.

Drawbacks: Missing data occurred in some prognostic variables, which required substitution. The study relies on retrospective data analysis and does not implement AI techniques.

3. "Data Mining Application on IVF Data for The Selection of Influential Parameters on Fertility" by M. Durairaj et al. (2013):

Strengths: The study proposes data mining techniques, including attribute selection algorithms, to identify influential parameters for IVF success prediction. It highlights the potential benefits for recommending IVF procedures to couples.

Drawbacks: The dataset used is relatively small, containing records of only 250 patients and 27 different tests.

4. "Applications of Artificial Neural Network for IVF Data Analysis and Prediction" by Dr. M. Durairaj et al. (2013):

Strengths: The study explores the use of Artificial Neural Networks (ANNs), particularly Multilayer Perceptron (MLP), for IVF data analysis and prediction. It demonstrates the effectiveness of the ANN approach and reports a 73% success rate.

Drawbacks: The study focuses on a theoretical model without implementing it in practice.

5. "Making IVF more effective through the evolution of prediction models: is prognosis missing a piece of the puzzle?" by Mara Simopoulou et al. (2018):

Strengths: The study reviews prediction models in IVF and categorizes them based on variables incorporated. It discusses the prediction based on patient characteristics, providing valuable insights into prognosis.

Drawbacks: The study is theoretical and does not present new empirical findings.

6. "Prediction of implantation after blastocyst transfer in vitro fertilization: a machine learning perspective" by Celine Blank et al. (2019):

Strengths: The study combines medical and embryo factors using a machine learning perspective. It proposes the use of the Random Forest method for predicting blastocyst implantation, highlighting the potential of machine learning over traditional algorithms.

Drawbacks: The study only suggests the Random Forest method without implementing it, and the data was obtained from a single academic referral, potentially limiting generalizability.

7. "Application of Machine Learning and Artificial Intelligence techniques for IVF Analysis and Prediction" by Satya Kiranmai Tadepalli et al. (2019):

Strengths: The study highlights the significance of machine learning (ML) and AI techniques in ART therapies, based on survey findings. It recognizes ML and AI as the most reliable prediction methods.

Drawbacks: The study relies on survey data and does not present new empirical findings.

8. "Computational Prediction of Implantation outcome after embryo transfer" by Behnaz Raef et al. (2020):

Strengths: The study proposes a prediction model to aid researchers in selecting embryos with greater accuracy and reducing difficulties associated with ART procedures. It emphasizes the potential cost-saving benefits.

Drawbacks: The proposed model may not necessarily be generalizable to other clinics due to the reliance on data from a single source.

9. "Machine Learning Approach to Predict Clinical Pregnancy Potential in Women Undergoing IVF Program" by Nining Handayani et al. (2022):

Strengths: The study utilizes an extensive IVF registry to develop a prediction model, highlighting its potential in clinical practice.

Drawbacks: The primary drawback is the method of collecting hindsight information.

10. "Deep Learning Techniques for Automatic Classification and Analysis of Human in Vitro Fertilized (IVF) Embryos" by Prof. Sujata N Patil et al. (2018):

Strengths: The study presents a promising approach to analyzing and classifying human embryos using deep learning techniques. It has the potential to improve cryopreservation, embryo selection, and the selection process for viable embryos.

Drawbacks: The study relies on a small training set and requires further testing on a larger database.

In summary, these studies demonstrate the potential of AI and machine learning techniques in IVF prediction and analysis. However, many of them are theoretical or limited by small datasets. Further research and validation are needed to fully utilize the power of AI in improving IVF outcomes.

4. Proposed Methodology

In the proposed study, we seek to conduct a comparative analysis of machine learning models on the IVF dataset we have gathered, including the Random Forest Classifier Model, Support Vector Machine Model, Logistic Regression, and Gradient Boosting Classifier Model.

4.1. Data Collection

In-vitro fertilization as we are aware is a very sensitive topic, therefore, we made sure to get real-time data through reliable sources. We obtained our dataset from Dr P. Thamilselvan who has conducted prior research on the topic of In-vitro fertilization[4]. He along with his team had collected the data by surveying patients undergoing IVF Treatment , through a physical form.

4.2. Data Preprocessing

The dataset provided was a raw dataset, which needed preprocessing. We started with analyzing the dataset we were provided with. It consisted of 228 patient records along with 43 columns i.e attributes to be evaluated on. We further went on to drop the irrelevant columns like ‘Patient Name’, ‘Husband Name’ etc. so that our dataset doesn’t have unnecessary noise. Further, we understood the type of variables we are dealing with and the number of missing values in each column.

4.2.1. Handling Missing Values

We tried to understand the number of missing values in each column and further handled these missing values for each column individually. The attached results of the missing analysis are shown below.

	Missing_Number	Missing_Percent
GROSS AND MICROSCOPIC APPEARANCE	170	0.745614
BMI(F)	110	0.482456
FEAR AND NEGATIVE TREATMENT ATTITUDE	88	0.385965
STRAIN OF REPEATED TREATMENT	88	0.385965
UNCERTAINTY	88	0.385965
DIFFICULTY IN TOLEARATING NEGATIVE EMOTIONS FOR EXTENDED TIME	88	0.385965
PSYCHOLOGICAL AND EMOTIONAL FACTORS	88	0.385965
HORMONAL FACTOR	60	0.263158
EDUCATIONAL LEVEL OF THE WOMAN	60	0.263158
MEDICAL DISORDERS	58	0.254386
PREVIOUS SURGERY	58	0.254386
PRE-EXISTING SYMPTOMS OF DEPRESSION	58	0.254386
NO.OF EMBRYOS TRANSFERRED	25	0.109649
NO.OF OOCYTES RETRIEVED	25	0.109649

Fig.1: Amount of Missing values in each column

We undertook handling the missing values in three ways namely Dropping Irrelevant Columns, Imputation of Values, and Treating Missing Values as a new category. On the basis of the existing research, we dropped irrelevant columns like the 'Educational level of the woman', 'Difficulty in handling negative emotions' and 'Fear and negative treatment attitude'. Further on columns like 'Combined Factor', 'Previous Surgery', 'Psychological and Emotional Factors', 'Pre-Existing symptoms of depression', 'Medical Disorders' , 'Uncertainty' we treated the missing values and dashes (-) present as NO as we are aware if they were present they would have been marked as YES.

The columns 'No.of oocytes retrieved' and 'No.of embryos transferred' are numeric columns and have a fairly low missing percent. Here we opted to use imputation to handle the missing values. We understood the nature of the numeric values present and realized that 'No.of oocytes retrieved' column has a negatively skewed distribution therefore ,median imputation was used there and 'No.of embryos transferred' column has a normal distribution therefore ,mean imputation was used to handle the missing values. The remaining columns like 'Strain of repeated treatment', 'Hormonal Factors', 'Gross and Microscopic Appearance' were handled by mode imputation as they are categorical columns.

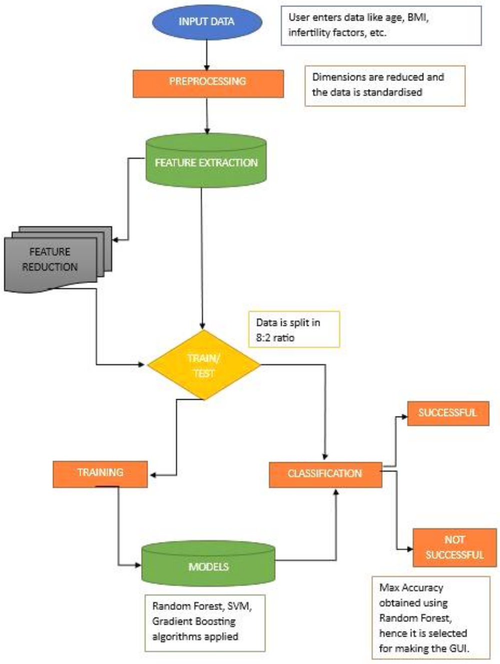

Fig.2 : Flowchart of the process

4.2.2. Data Standardization

While exploring the data we realized that the column 'Duration of Infertility (in years)' does not contain normalized data i.e it contained the responses in fractions, whole numbers and decimals. Thereby to use this column effectively we standardized the data by converting all the entries to float datatype.

4.2.3. Examination of Numerical Variables

The dataset provided to us had very few numerical columns namely 'Duration of Fertility(in years)', 'No.of oocytes retrieved' , 'No. of embryos transferred' and 'Stages of Endometriosis'. Thus, before converting the remaining categorical variables to numeric for model building, we evaluated these numeric values separately first. During this evaluation, we realized that there are many outliers present, which wasn't surprising because in the medical field, we are dealing with real time data of patients where the presence of outliers is unavoidable. Through our skewness and kurtosis analysis, we inferred that the 'No.of oocytes' column has a remarkable impact on our target variable i.e IVF Treatment.

Further we created a heatmap displaying the correlation of all our numeric columns. Through this heatmap we understood that :

- There is a strong positive correlation between "NO. OF OOCYTES RETRIEVED" and "NO. OF EMBRYOS TRANSFERRED", indicating that patients who had a higher number of eggs retrieved tend to have more embryos transferred.

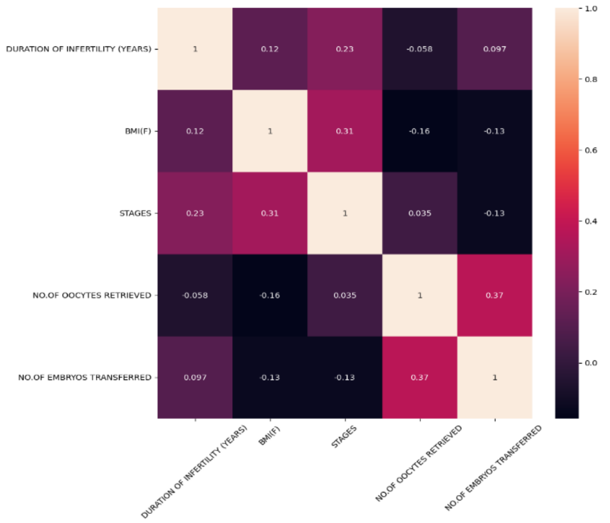

Fig 3: Heatmap of numeric variables

- There is a weak positive correlation between "STAGES" and "DURATION OF INFERTILITY (YEARS)", indicating that patients with a higher stage of infertility tend to have a slightly longer duration of infertility.
- There is a weak negative correlation between "BMI(F)" and "NO. OF OOCYTES RETRIEVED", indicating that patients with higher BMI values tend to have slightly fewer eggs retrieved.
- There is a weak negative correlation between "STAGES" and "NO. OF OOCYTES RETRIEVED", indicating that patients with a higher stage of infertility tend to have slightly fewer eggs retrieved.
- There is a weak negative correlation between "STAGES" and "NO. OF EMBRYOS TRANSFERRED", indicating that patients with a higher stage of infertility tend to have slightly fewer embryos transferred.
- There is a weak positive correlation between "NO. OF OOCYTES RETRIEVED" and "BMI(F)", indicating that patients who had more eggs retrieved tend to have a slightly higher BMI.

4.2.4. Handling Categorical Variables

To be able to use all our features effectively in our model building and get an accurate prediction rate we are required to convert all the categorical variables to numeric variables. We handled this in two ways, namely Label Encoding and Mapping. For columns which had categorical variables like YES and NO mapping could be easily done by defining a mapping function where it consisted of a dictionary which assigned YES: 1 and NO: 0. This mapping function was performed on columns like ‘Previous Pregnancy’, ‘Previous Surgery’, ‘If yes, miscarriage caused’, ‘Medical Disorders’ and many more columns which contained only YES and NO values. Further columns like ‘IVF Treatment’ which contained various values like Success, Unsuccess and Running were mapped to 1, 0 and -1 respectively. ‘Gross and Microscopic Appearance’ column had categorical values like Clear, Denire Liquid, Turbid, and Thick gelatinous which were mapped to 0, 1, 2, 3, respectively. The columns ‘Age(F)’ and ‘Age(M)’ had the same categories ‘20-30’, ‘31-40’ and ‘>40’; these were mapped to 0, 1, 2, respectively.

4.3. Feature Engineering

4.3.1. Feature Selection

Feature selection techniques like univariate feature selection, feature selection on tree based algorithms, correlation and mutual information etc. For feature selection in our research, we primarily used four techniques: Univariate Feature Selection, Correlation Feature Selection, Mutual Information Feature Selection, Recursive Feature Elimination Method, and Feature Importance from Tree-based Model.

- Univariate Feature Selection : We evaluated the important feature using a Chi Squared. Out of our 43 features we narrowed down the top 20 features with the highest Chi Squared test. According to which ‘Gross and Microscopic Appearance’, ‘Unknown Factor’, ‘Male Factor only’, ‘Psychological and Emotional Factors’, ‘Strain of Repeated treatment’, ‘Severe Male Factor,’ ‘Duration of infertility(in years)’ have a major impact on the determination of our target variable IVF Treatment.
- Correlation Feature Selection: As the name suggests here we have evaluated it on the basis of correlation scores. According to which sperm vitality, sperm concentration, sperm morphology, psychological and emotional factors, strain of repeated treatment, semen ejaculate volume have a major impact on the determination of our target variable IVF treatment.
- Mutual Information Feature Selection : Here ‘Sperm concentration’, ‘Unknown factor’, ‘Duration of fertility(in years)’ , ‘Sperm morphology’, ‘BMI(f)’ , ‘Age(m)’ , and ,‘Male Factor’ have a major impact on the determination of our target variable IVF Treatment.
- Recursive Feature Elimination Method for feature selection: RFE is an iterative strategy that entails training a model (such as a decision tree, logistic regression, etc.) and eliminating the least significant feature(s) at each iteration until a predetermined amount of features are chosen. With this method, the features are prioritized according to importance, and the features at the top are chosen as the most pertinent. In light of this, we assessed the decision tree and logistic regression.

- Feature Importance from Tree-based Model for feature selection : Decision trees, random forests, and gradient boosting machines are examples of tree-based models that can provide feature relevance ratings depending on how often a feature is used to split the tree and how much it lowers impurity. These feature importance scores can be used to rank the features and determine which features are the most relevant.

We analyzed the top 20 important features from each method, and then we decided the most important features on the basis of common occurrences in all the models. After analyzing the top 20 important features from each feature selection technique we have selected the features which appear in 5 or 6 out of 7 of the methods. These are the Top 10 features :

1. Age(F)
2. Duration of infertility(years)
3. BMI(F)
4. Psychological and Emotional
5. Strain of repeated treatment
6. Gross and Microscopic Appearance
7. Sperm Motility
8. Sperm Vitality
9. No. Of Oocytes Retrieved
10. No. Of Embryos Transferred

4.3.2. Feature Extraction

In the feature extraction we used the Principle Component Analysis method to extract features. To evaluate how many principal components to create using the Scree plot. We created the scree plot for our dataset. From the Scree plot, we can infer that 4 principal components are enough to describe the dataset.

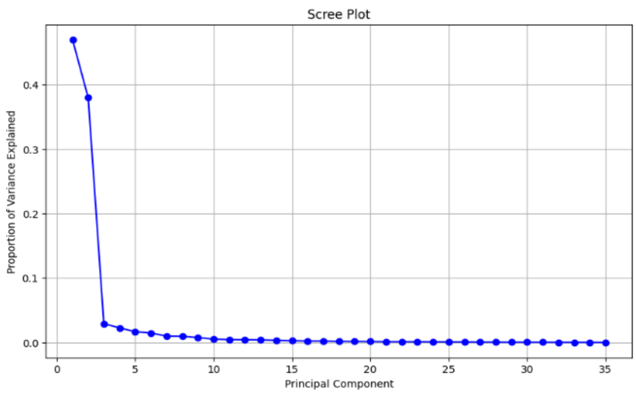

Fig 4: Scree Plot

After our inference from the scree plot we performed the principal component analysis on our dataset and plotted the results on a heatmap to improve the readability of the same.

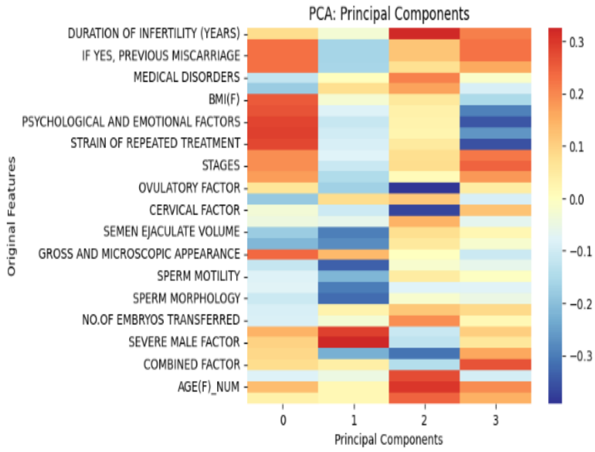

Fig 5: Heatmap of the PCA

The following are the inferences drawn from the Principle Component Analysis are below :

- The different components, such as PC1, PC2, PC3, and so on, are first arranged according to their explained variance ratio, which shows the percentage of the total variance in the data that each PC is responsible for explaining. As an illustration, we can observe from our PCA that PC1 accounts for 18% of the variance in our data and has a variance of 0.18.

- The coefficients of the original variables in PC1 represent the contributions of each variable to the first principal component (PC1). These coefficients, also known as loadings or eigenvectors, indicate the direction and magnitude of the relationship between the original variables and PC1.

The following are the inferences drawn from the resulting heatmap:

- The columns ‘Previous Pregnancies’, ‘ I’f yes, previous miscarriage’, ‘BMI(f)’, ‘Psychological and Emotional Factors’, ‘Strain of repeated treatment’, and ‘Gross and Microscopic Appearance’ contribute positively to PCI.

- The columns Sperm Concentration, Sperm Morphology, Sperm Vitality, Male Factor only, and Severe Male Factor contribute the most to PC2.

- Duration of infertility, Age(F), Female Factor, Ovulatory Factor, and Cervical Factor contribute the most to PC3.

- Combined Factor, Stages, and Strain of repeated treatment contributes most to PC4.

4.4. Model Selection

In our study, three models—the Random Forest Model, the Support Vector Model, and the Gradient Boosting Model—were assessed. We divided the dataset 80:20 into train and test for all of the model evaluations.

Specificity : Specificity, also known as true negative rate or selectivity, is the ability of a classification model to precisely identify the negative class instances (i.e., instances of the non-target class) among all the actual negative class cases in a dataset. It is calculated by dividing the total number of negative incidences that actually happened by the total number of accurate negative predictions. This is crucial in situations where identifying real negatives is crucial, such as in the medical industry. If the classification is binary, a greater value is preferred.

Formula : *Specificity = True Negatives / (True Negatives + False Positives)*

Sensitivity : Sensitivity, also known as True Positive Rate or Recall, is the ability of a classification model to accurately identify the positive class instances (i.e., instances of the target class) among all the real positive examples in a dataset. It is calculated as the ratio of actual positive events to all actual positive forecasts.

Formula: *Sensitivity = True Positives / (True Positives + False Negatives.*

4.4.1. Random Forest Classifier Model

The Random Forest Classifier is a machine learning technique that combines the forecasts of various decision trees to produce a classifier that is more reliable and accurate. For classification problems, where the objective is to predict the class or category of a given input based on its features or attributes, this approach has been widely employed in both academic research and commercial applications. We chose this model because our target variable will classify the IVF treatment results to either Yes or No. Thereby we have used this classification model.

In our specific model we have considered two hyperparameter tunings namely n_estimators and random_state.

- N_estimators : This parameter specifies the number of decision trees the ensemble model will be included in the ensemble model. We in the beginning of our research started with this parameter equal to 4, because as analyzed before the principle components were 4, but the model accuracy was low i.e only 0.76. Further we tried more changes to this and arrived at 100 where the accuracy was boosted to 0.86. Thereby we also inferred that as we increase the number of decision trees the accuracy is also bound to increase. It helps in reducing the overfitting of our model.

- Random_state : This parameter sets the random seed number for the random forest classifier. This ensures that the results of the model are reproducible i.e if we run the same model with the same random_state then we would get the same results with accuracy. We chose 42 because it is one of the most commonly used numbers, and usually considered a default number for random_state.

In this model we also evaluated the Specificity and the Sensitivity of the data.

- Specificity : 0.93, this implies that the model was able to classify 93% of the true negative values.

- Sensitivity : 0.75, this implies that the model was able to classify 75% of the positive values.

- Accuracy : 0.86

4.4.2. Support Vector Machine Model

For both classification and regression issues, Support Vector Machine (SVM), a well-liked and powerful machine learning approach, is used. A supervised learning algorithm chooses the ideal decision boundary (or hyperplane) in a high-dimensional feature space to categorise data points into different groups.The "maximum margin" principle states that SVM should find the hyperplane that most effectively divides the data points of different classes with the largest margin. The data points that are most closely related to the decision boundary and have the most impact on where the hyperplane is located are called support vectors, from where the word "Support Vector Machine" originates.

The maximisation principle is the foundation of the Support Vector Machine (SVM) method, which aims to locate the decision boundary (or hyperplane) in the feature space with the largest margin between different classes. This margin, or the separation between the decision boundary and the nearest data points of each class, is what SVM seeks to maximise.According to the maximum margin principle, a wider margin allows for better generalisation to unknown data points by supplying a larger zone of uncertainty around the decision border. This can enhance model performance by reducing the risk of overfitting and providing a more trustworthy decision boundary that is less likely to be affected by noise or outliers in the data.

In our SVM model we haven’t used any hyperparameter tuning in this phase of our project.

- Specificity: 1.00, this is a really good number as this represents that the model has correctly predicted all the negative cases.

- Sensitivity: 0.00, this isn’t a great number because it doesn’t predict any of the positive values correctly.

- Accuracy: 0.65, the accuracy of the model is pretty low

To improve the accuracy of this model, we introduced the Bagging Classifier, which is known as a bootstrap aggregator. It is used to create an ensemble of the SVM model. This entails using various subsets of the training data to train numerous instances of the base estimator (in this example, the SVM). In order to generate new training sets of the same size, these subsets are produced by randomly selecting the training data with replacement (bootstrap). The final prediction is then aggregated using a majority vote once each base estimator has been trained on one of these subsets.

We used the Bagging Classifier as it helps to reduce overfitting and increases the ability to generalize the new data.

- Specificity: 1.00, remains the same
- Sensitivity: 0.06, increased but barely
- Accuracy: 0.67, increased accuracy by 0.2

4.4.3. Gradient Boosting Model

A frequently used ensemble learning method for classification and regression applications is gradient boosting. Unlike Random forest classifiers, it sequentially creates an ensemble of ineffective learners (usually decision trees), where each ineffective learner concentrates on fixing the mistakes committed by the prior learners. In this manner, the ensemble gradually raises the accuracy of its forecasts.

Gradient Boosting Classifier is a Gradient Boosting implementation for classification offered by the Python scikit-learn module. GradientBoosting requires more hyperparameter tuning when compared to Random Forest, but in this stage of the project, we haven't performed hyperparameter tuning.

The results of our model without any hyperparameter tuning is

- Specificity: 0.93; this shows that our model could predict 93% of the negative cases correctly
- Sensitivity: 0.75, this shows that our model could predict 75% of the positive cases correctly
- Accuracy: 0.87; this is the overall accuracy of the model.

4.4.4. Logistic Regression Model

Using binary values like 0 or 1, true or false, or positive or negative, logistic regression attempts to model the likelihood that an input example belongs to a certain class. The logistic function, also known as a sigmoid function, is used to model the input features, also known as independent variables, and estimate the probability of the binary outcome. The input features are translated by the logistic function to a probability value between 0 and 1.

The reason why we chose this model is because of the small size of our dataset. Logistic regression works better with smaller datasets when compared to other models like random forest and gradient boosting models.

Without any hyperparameter tuning, we have achieved the following

- Accuracy: 0.85, this accuracy without any hyperparameter tuning is very good.
- Specificity: 0.90, this signifies that the model was able to predict 90% of the negative test cases correctly
- Sensitivity: 0.75, this signifies that the model was able to predict 75% of the positive test cases properly

5. Future Work

In our work, we found that gradient boosting and random forest classifier models are suitable for predicting IVF success. During the initial testing phase, these models showed remarkable accuracy. We do, however, intend to do hyperparameter tuning to enhance the performance of these models. To improve accuracy, sensitivity, and specificity, the model's settings must be optimized.

Moreover, we aim to incorporate deep learning techniques to enhance the reliability of our IVF success prediction model. This includes implementing various deep learning algorithms such as artificial neural networks, convolutional neural networks, and recurrent neural networks. These advanced techniques can capture more complex relationships and patterns in the data, further improving the prediction accuracy.

In addition to these enhancements, we also intend to increase the sample size of our dataset by collecting more data from various fertility clinics. This will enable us to train our models on a more varied dataset and improve the generalizability of our findings. In addition, we will investigate several feature engineering methods to extract more useful features from the data and enhance the functionality of the models.

Furthermore, we will enhance the GUI by designing a more visually appealing and intuitive interface that makes it easy for users to navigate through the app and access the different features. We plan to incorporate interactive features such as sliders, drop-down menus, and checkboxes to make the app more engaging and allow users to customize their preferences.

Finally, we will conduct a comparative analysis of our models against other state-of-the-art methods used for predicting IVF success rates. This analysis will provide us with a better understanding of the strengths and limitations of our models and help us identify areas for future improvements.

Overall, our future work aims to enhance the accuracy, reliability, and generalizability of our IVF success prediction model using advanced machine learning techniques, increasing the sample size, optimizing feature engineering techniques, and conducting a comparative analysis.

6. Conclusion

In conclusion, our study demonstrates the potential of using machine learning models to predict the success rates of IVF treatments. We achieved high accuracy, sensitivity, and specificity during the initial testing phase by leveraging the random forest classifier and gradient-boosting classifier models.

However, our research also highlights the need for further improvements in the performance of the models. We identified incorporating deep learning algorithms, optimizing the models through hyperparameter tuning, and increasing the sample size of the dataset as areas where the models can be improved. We also demonstrated the importance of feature engineering and conducted a comparative analysis to improve the models' performance further.

The practical implications of our study for fertility clinics and patients seeking IVF treatments are significant. By using machine learning models to predict the success rates of IVF treatments, fertility clinics can provide more accurate counseling to patients, ultimately improving their chances of a successful outcome.

Our research contributes to the growing body of knowledge on applying machine learning techniques in healthcare. It highlights the potential of these techniques to improve patient outcomes and provides a roadmap for future research in this area.

In conclusion, our study provides valuable insights into using machine learning techniques for IVF success prediction. It demonstrates the potential of these techniques to improve the success rates of IVF treatment. By improving and refining these models, we can ultimately improve the quality of care for patients seeking IVF treatments

References

[1] Siristatidis, C., Pouliakis, A., Chrelias, C., & Kassanos, D. (2011). Artificial Intelligence in IVF: A Need. Systems Biology in Reproductive Medicine, 57(4), 179–185. https://doi.org/10.3109/19396368.2011.558607

[2] Lintsen, A. M. E., Eijkemans, M. J. C., Hunault, C. C., Bouwmans, C. A. M., Hakkaart, L., Habbema, J. D. F., & Braat, D. D. M. (2007). Predicting ongoing pregnancy chances after IVF and ICSI: a national prospective study. Human Reproduction, 22(9), 2455–2462. https://doi.org/10.1093/humrep/dem183

[3] Durairaj, M. & Nandha Kumar, R. (2013). Data mining application on IVF data for the selection of influential parameters on fertility. International Journal of Engineering and Advanced Technology (IJEAT), 2(6), 262-266. Retrieval Number: F2068082613/2013©BEIESP. https://www.ijeat.org/portfolio-item/f2068082613/

[4] Durairaj, M., & Thamilselvan, P. (2013). Applications of Artificial Neural Network for IVF Data Analysis and Prediction. Journal of Engineering Computers & Applied Sciences. https://www.semanticscholar.org/paper/Applications-of-Artificial-Neural-Network-for-IVF-Durairaj-Thamilselvan/9232f53eb29a933be580eb050a9df8abda922981

[5] Vogiatzi, P., Pouliakis, A., & Siristatidis, C. (2019). An artificial neural network for the prediction of assisted reproduction outcome. Journal of assisted reproduction and genetics, 36(7), 1441–1448. https://doi.org/10.1007/s10815-019-01498-7

[6] Blank, C., Wildeboer, R. R., DeCroo, I., Tilleman, K., Weyers, B., de Sutter, P., Mischi, M., & Schoot, B. C. (2019). Prediction of implantation after blastocyst transfer in in vitro fertilization: a machine-learning perspective. Fertility and Sterility, 111(2), 318–326. https://doi.org/10.1016/j.fertnstert.2018.10.030

[7] Tadepalli, S. K., & Lakshmi, P. V. (2019). Application of Machine Learning and Artificial Intelligence Techniques for IVF Analysis and Prediction. International Journal of Big Data and Analytics in Healthcare, 4(2), 21–33. https://doi.org/10.4018/ijbdah.2019070102

[8] Raef, B., Maleki, M., & Ferdousi, R. (2020). Computational prediction of implantation outcome after embryo transfer. Health informatics journal, 26(3), 1810–1826. https://doi.org/10.1177/1460458219892138

[9] Handayani, N., Louis, C. M., Erwin, A., Aprilliana, T., Polim, A. A., Sirait, B. I., Boediono, A., & Sini, I. (2022). Machine Learning Approach to Predict Clinical Pregnancy Potential in Women Undergoing IVF Program. FERTILITY & REPRODUCTION, 4(2), 1–11. http://repository.uki.ac.id/8721/

[10] Sujata, N., Patil, Uday Wali, Swamy, M., & Patil, N. (2018). Deep Learning Techniques for Automatic Classification and Analysis of Human in Vitro Fertilized (IVF) embryos. JETIR1802014 Journal of Emerging Technologies and Innovative Research, 5. https://www.jetir.org/papers/JETIR1802014.pdf

[11] C. C. Aggarwal and C. K. Reddy, Data clustering: algorithms and applications. CRC Press, 2013.

[12] P. N. Tan, K. Steinbach, and V. Kumar, “Data mining cluster analysis: Basic concepts and algorithms,” 2006.

[13] T. M. Mitchell, “Machine learning. 1997,” Burr Ridge, IL: McGraw Hill, vol. 45, 1997

[14] K. Meena, M. Durairaj, and K. R. Subramanian, “Machine Learning Techniques of Artificial Neural Network Modeling to Predict Fertility Rate of Sperm from the Outcome of IVF Functional Tests”, International Journal of Computer Science and Applications.

[15] M. Durairaj, K. Meena, and S. Selvaraju, ―Applying a data mining approach of rough sets on spermatological data analysis as predictors of in-vitro fertility of bull semen‖, International Journal of Computer Mathematical Sciences and Applications, Serials Publications, ISSN: 0973-6786, Vol. 2(3), pp. 221-231, Dec 2008.

[16] Kaufmann, S.J., Eastaugh, J.L., Snowden, S., Smye, S.W. and Sharma, V. The Application of neural networks in predicting the outcome of in-vitro fertilization. Human Reproduction, (1997) vol.12 no. 7 pp. 1544-1457.

[17] Larsson, B. and Rodriguez-Martinez, H. Can we use in vitro fertilization tests to predictsemen fertility? Anim. Reprod. Sci. (2000) 60-61: 327-336.

[18] S.J.Kaufmann, J.L.Eastaugh, S.Snowden, S.W.Smye and V.Sharma (1997)‖The application of Neural Networks in predicting the outcome of in-vitro fertilization ―Human Reproduction vol.12 no. 7 pp.1454–1457.

[19] Kay Elder, & Brian Dale., 2000, ―In- Vitro Fertilization‖, Second Edition, United Kingdom at the University Press, Cambridge.

[20] Barnett-Itzhaki Z, Elbaz M, Butterman R, et al. Machine learning vs. classic statistics for the prediction of IVF outcomes. J Assist Reprod Genet. 2020;37(10):2405–12. https://doi.org/10.1007/s10815-020- 01908-1

[21] Passmore L, Goodside J, Hamel L, Gonzales L, Silberstein T, Trimarchi J. Assessing decision tree models for clinical in-vitro fertilization data. December 2003:1–15. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.7597&rep=rep1&type=pdf

[22] https://www.cdc.gov/art/ivf-success-estimator/

[23] https://www.arcfertility.com/sart-patient-predictor/

[24] https://machinelearningmastery.com/basic-data-cleaning-for-machine-learning/

[25] https://link.springer.com/article/10.1007/s00521-018-3693-9

[26] Shixia Liu, Xiting Wang, Mengchen Liu, Jun Zhu, Towards better analysis of machine learning models: A visual analytics perspective, Visual Informatics, Volume 1, Issue 1, 2017, Pages 48-56, ISSN 2468-502X, https://doi.org/10.1016/j.visinf.2017.01.006.

[27] Ms. Mohini Dadhe, Ms. Sneha Miskin. (2015). Optimized Wireless Stethoscope Using Butterworth Filter. International Journal of New Practices in Management and Engineering, 4(03), 01 - 05. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/37

[28] Raj, R., & Sahoo, D. S. S. . (2021). Detection of Botnet Using Deep Learning Architecture Using Chrome 23 Pattern with IOT. Research Journal of Computer Systems and Engineering, 2(2), 38:44. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/31