

Optimized Hybrid Prognostics Using Hynetreg Model for Infertility Prediction

Kamal Upreti^{1*}, Jossy George¹, Sheela Hundekari²
and Mohammad Shabbir Alam³

¹Department of Computer Science, CHRIST (Deemed to be University),
Delhi NCR, Ghaziabad, Uttar Pradesh, India.

²Department of Computer Science and Engineering,
Pimpri Chinchwad University, Maval Talegaon, Pune, Maharashtra, India.

³Department of Computer Science, College of Computer Science and
Information Technology, Jazan University, Jazan, KSA.

*Corresponding Author E-mail: kamalupreti1989@gmail.com

<https://dx.doi.org/10.13005/bpj/3168>

(Received: 11 November 2024; accepted: 03 April 2025)

This paper develops an optimized hybrid approach to predict infertility with the HyNetReg Model. The HyNetReg Model combines deep feature extraction by using neural networks with logistic regression with regularization. It uses both hormonal and demographic information of 100 participants to clarify intricate interlinkages between demographic factors and salient hormonal levels, such as Luteinizing Hormone, Follicle Stimulating Hormone, Anti-Müllerian Hormone, and Prolactin, and the ability of these same factors to affect fertility outcomes. It applies heavy data pre-processing including normalization, missing values imputation, and class imbalance handling through oversampling techniques. A multi-layer neural network is utilized to extract features for the reduction of complex, non-linear interaction among the input variables. Then, regularized logistic regression is applied for classification on the same features. Performance evaluation metrics, including accuracy, precision, recall, F1-score, and ROC curve analysis, demonstrate the superiority of the HyNetReg Model over traditional logistic regression. The ROC curve was specifically utilized to assess the model's discrimination ability between infertile and fertile cases by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity). A higher Area Under the Curve indicated that the model effectively distinguished infertility risks based on hormonal and demographic features. The results indicate that the model can recover very slight interdependencies of hormones and influences of demographics, making it suitable for modeling multi-factorial determinants of infertility and holding significant implications for clinical decision-making.

Keywords: Luteinizing Hormone (LH); Follicle Stimulating Hormone (FSH); Anti-Müllerian Hormone (AMH); Alpha-Feto-Protein (AFP); Idiopathic Female Infertility (IFI); Blood Urea Nitrogen (BUN); multi-layer perceptron (MLP); poor ovarian response (POR); recurrent reproductive failure (RRF); total motile sperm count (TMSC).

Infertility is a condition that affects millions of couples worldwide, often leading to emotional, psychological, and social distress. It is best defined as the inability to conceive after 12

months of regular, unprotected sexual intercourse.¹ However, for women over the age of 35, medical guidelines suggest seeking medical evaluation after 6 months due to the natural decline in

fertility with age. Infertility can arise from male, female, or combined factors, and in many cases, the causes remain unexplained despite medical evaluation.²⁻⁴ The causes of infertility are complex and multifactorial, encompassing biological, environmental, and lifestyle-related factors. While certain factors—such as hormonal imbalances, genetic predispositions, and reproductive system disorders—are gender-specific, others, including age, obesity, smoking, and stress, can affect both partners.⁵ The complexity of these interrelated factors makes infertility prediction and diagnosis challenging.⁶⁻⁸ Recent advancements in medical informatics and machine learning (ML) have introduced novel approaches to understanding and predicting infertility. Machine learning techniques excel in capturing intricate, nonlinear relationships within large datasets, allowing for the identification of subtle patterns in hormonal and demographic variables that may contribute to infertility. Unlike traditional statistical models, ML methods can enhance clinical decision-making by improving predictive accuracy and providing data-driven insights for personalized fertility treatment.¹¹

To address the limitations of traditional logistic regression models, this study proposes HyNetReg, a hybrid machine learning approach that combines deep feature extraction using neural networks with logistic regression for classification. Neural networks are particularly effective at extracting meaningful features from complex datasets, allowing the model to learn intricate interactions between demographic and hormonal predictors of infertility. Applying logistic regression to these extracted features enhances model interpretability while improving classification performance. This study aims to bridge the existing gap in predictive modeling for infertility assessment by leveraging machine learning to provide clinicians with more reliable tools for early diagnosis and treatment planning. The proposed HyNetReg model offers a more robust and accurate approach compared to conventional statistical techniques, contributing to enhanced clinical decision-making in reproductive health. Demographic variables interact with the crucial hormonal levels, which include Luteinizing Hormone (LH), Follicle Stimulating Hormone (FSH), Anti-Müllerian Hormone (AMH), and Prolactin in determining infertility. This study

involves the use of machine learning in the analysis of reproductive health in an attempt to close the existing gap in predictive modeling for the insight of doctors to decide on the treatment of infertility and diagnosis appropriately. Despite machine learning's promising future, there is still a dearth of sufficient research on hybrid models that infer the reasons of infertility in the literature. Most of them either consist of traditional statistical models exclusively or machine learning in isolation without incorporating the power of the approach. This work fills the gap by showing that a combination of neural networks with logistic regression indeed achieves higher predictive accuracy, which is associated with better patient treatment outcomes in the field of reproductive health.

Literature Review

Millions are affected by infertility; World Health Organization estimates that between 60-80 million couples suffer from infertility, mainly in Sub-Saharan Africa.¹² Predictive capabilities of machine learning models in the prediction of outcomes in cases of infertility are analyzed in studies. A study applied naïve Bayes, decision trees, and multi-layer perceptron (MLP) models to predict the risk of infertility in Nigerian women, obtaining success rates as high as 74.4% with the decision tree and MLP compared to naïve Bayes.¹³ Another study aimed to predict in vitro fertilization (IVF) success rates using multiple machine learning models, including Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), C4.5 decision tree, Classification and Regression Tree (CART), and Random Forest (RF). The results indicated that feature selection significantly improved predictive accuracy, with Random Forest achieving the highest performance among the tested models. Feature selection of influential attributes resulted in a higher increase of accuracy in prediction, with a consideration of significant reduction.¹⁴ The third one was a retrospective analysis about comparing the models for the clinical pregnancy rate after treatments against infertility (IVF, ICSI, IUI) in which random forest achieved the highest accuracy. Key predictive factors were age, follicle-stimulating hormone, and endometrial thickness, especially in relation to pregnancy success and female age.¹⁵ Studies on infertility are focusing on developments in machine learning models to enhance the prediction of outcome regarding

fertility. One has built a model to stratify infertile and fertile couples using bioclinical signatures that achieves an accuracy of 74% in distinguishing between groups with anthropometric, metabolic, and oxidative factors that could be altered by lifestyle interventions.¹⁶ Another such research, carried out in Taiwan, investigated the major risk factors influencing male sperm count through the application of the machine learning algorithms of random forest and extreme gradient boosting. It identified sleep time, body fat, systolic blood pressure, and some novelties, such as alpha-fetoprotein (AFP), and blood urea nitrogen (BUN), as influential factors in male infertility.¹⁷ Another research used a deep learning convolutional neural network to forecast the male fertility rate from semen analysis, achieving accuracy levels of 81% with semen prediction and 86% with sperm concentration. The results show that there is a possibility of using deep learning in automated workflows of artificial insemination.¹⁸ It affects millions of people globally, especially in Sub-Saharan Africa, where the rate may be as high as 50% compared to the 20% rate found in the Eastern Mediterranean Region and 11% in the developed world.¹⁹ One study was conducted with the intention of developing a predictive model for the risk of infertility in Nigerian women by identifying and analyzing 14 classified risk factors across genital, endocrinal, developmental, and general categories. A study applied the MLP model using Waikato Environment for Knowledge Analysis (WEKA), an open-source machine learning software, identifying major factors influencing infertility. Another research targeted PCOS to predict using sophisticated machine learning techniques.²⁰ An optimized chi-squared (CS-PCOS) approach using an optimized GNB model also achieved high accuracy for early detection in PCOS and determined key factors such as prolactin, systolic blood pressure, and thyroid-stimulating hormone. But another study on idiopathic female infertility (IFI) used FTIR spectroscopy and machine learning for classification of follicular fluid from the IVF patients with an accuracy between 94% and 100%.²¹ The study has indicated that FTIR spectroscopy is indeed possible in distinguishing IFI patients from controls using the biochemical markers that are present in the follicular fluid, and

it can be used in the decision process regarding infertility treatment.²²

This is an important medical science in which the likelihood of a live birth can be predetermined for a successful in vitro fertilization so that the medical practitioners are in a better position to make their clinical decisions. A recent study has compared the performance between Gradient Boosting Regression GBR, Lasso Regression LaR, and Linear Regression LR models for predicting IVF success rates, and it found that the most accurate model to be 78% GBR.²³ A study tried to determine whether pre-intervention sperm parameters, particularly Total Motile Sperm Count (TMSC), can predict the success of varicocele repair, which determined that successful outcome may be more probable with TMSC improvement in cases of intermediate or high-grade.²⁴ In addition, A.I. has also been used to predict pregnancy outcome for patients with recurrent reproductive failure (RRF) by employing six different data panels. The autoantibodies panel had the highest prediction accuracy of live birth (AUC = 0.909).²⁵ Other reports of predicting POR have been by machine learning models using ANN and Random Forest; ANN showed the highest accuracy in pre-launch COS prediction.²⁶ Development of a prediction model for endometriosis using UK Biobank data using CatBoost algorithms with an excellent area under the ROC curve of 0.81 for predicting endometriosis.²⁷ For male infertility, a fuzzy logic model was developed, demonstrating effective decision-support by simulating risk factors using 4374 IF-THEN rules.²⁸ A study on diagnosis of polycystic ovary syndrome (PCOS) used machine learning algorithms whereby the highest testing accuracy and recall on PCOS prediction were realized by the Hybrid models of Random Forest and Logistic Regression (91%) and (90%), respectively.²⁹ Comparison of all the results were provide in Table 1.

MATERIALS AND METHODS

The proposed methodology as depicted in Fig. 1 aims at developing a hybrid model HyNetReg that combines the strengths of both neural networks and logistic regression to predict fertile factors. The following steps present the methodology used in this research.

Data Collection and Pre-processing

This study employed a dataset which is taken from publicly available website (Kaggle.com) containing hormonal and demographic data from 100 participants. Key demographic information was the women's age (AGE(W)) and their husbands' ages (AGE(H)). Also included as features in prediction were LH, FSH, AMH, and Prolactin - critical hormonal levels. It pre-processed the data and took care of missing values, normalized data, and resolved issues of class imbalance. Data analysis, pre-processing, and normalization were performed using Python libraries such as pandas, numpy, and scikit-learn. The neural network for feature extraction was implemented using TensorFlow and Keras, which were instrumental in building and training the model. Model evaluation, including metrics calculation and generating the confusion matrix, was conducted using scikit-learn.

Dealing with missing values: In the dataset, missing values were represented by 'nd'. They were standardized to NaN and imputed using mean values for each hormone. Records with missing data in critical columns such as 'final cause' and 'm/f/c' were removed to validate the dataset.

Normalization: To get uniform scaling of all the features, hormonal values were normalized to a uniformly consistent range, and hence this improved the efficiency of machine learning models.

Class imbalance handling: since the data set had an imbalance of different infertility outcomes, the class imbalance in the data set is balanced using the Random Over Sampler technique in the imblearn library. The way the technique works involves randomly duplicating instances of the minority class until class distributions are relatively balanced. That means that, for sure, the model would not bias toward the majority class.

Feature Extraction Using Neural Networks

For feature extraction, a neural network was applied to capture the non-linear relationships in the hormonal and demographic features. The network was built with several hidden layers so that different layers can learn to identify complex relationships between the input features. The network architecture consisted of the following:

Input Layer: This had input nodes

representing the features AGE(W), AGE(H), LH, FSH, AMH, and Prolactin. Two hidden layers consisted of 64 neurons each, and the ReLU (Rectified Linear Unit) activation function was used to introduce non-linearity, allowing for the model to learn complex feature interactions.

Output for Feature Extraction: The output from the second hidden layer becomes the extracted feature set; this was used as an input for the following logistic regression model.

The neural network model could thus be used to determine some of the key hormonal levels, namely luteinizing hormone, follicle stimulating hormone, prolactin, and anti-Müllerian hormone, in their effects on gynecological disorders. Specifically, anti-Müllerian hormone, known to be a marker of ovarian reserve, is highly significant in providing holistic information relating to reproductive health. Standardized hormones were presented to the neural network that filters out complex features arising from the subtle interplay of factors, which in turn is fed to logistic regression for the classification process. This hybrid approach has enhanced the applicability and accuracy of the model by helping the deep learning model to identify non-linear relationships and provide a robust method to predict gynecological conditions more effectively.

Logistic Regression Model

The features extracted from the neural network were further given to logistic regression for the final classification. Logistic regression was chosen because it is very interpretable and performed well as a classifier for only two outcomes. The model was fitted to the resampled version of the training set to obtain equal classes. The features were then fed into a logistic regression model for final classification. Logistic Regression was picked because it is very interpretable and effective to use in binary classification problems. The maximum iteration to ensure convergence was set during training at 1,000 iterations. The best value of the regularization parameter C was also determined from grid search, which, after finding the best trade-off between model complexity and generalization capability, cancelled any case of overfitting. The Hyperparameter tuning was carried out with the following parameters as shown in Table 2.

This surely mitigated class imbalance, leaving the logistic regression model capable of generalizing well on all classes. It was shown by evaluation metrics that the model was robust and accurate enough to be used for predictive tasks.

Model Training and Evaluation

The first thing was to split the dataset into a training set, which had 75% of the dataset, and a testing set comprising 25%. It allowed having adequate data for learning to the model and also maintaining an independent set for performance evaluation. We wanted to train and test datasets separately, ensuring that the model actually tested its predictive ability on data it had not seen in training. Therefore, we expected a good evaluation of generalization ability. The performance of the model is further analyzed by metrics like accuracy, precision, recall, and F1-score—all together would provide an all-inclusive view of its ability to predict the cause of infertility. These metrics would determine both the exactness with which the predictions were made and how well false positives and false negatives were balanced. To ascertain the strength of the HyNetReg model better, its performance was compared against a standalone logistic regression model based on the original feature set, which did not incorporate any neural network-based feature extraction. This comparison elicited added value in the usage of the hybrid approach in that prediction accuracy and robustness improved because of the infusion of non-linear feature extraction from the neural network.

Models Comparison and Analysis

Accuracies, precisions, recalls, and F1-scores among the models were compared. It was anticipated that the HyNetReg model with neural network-based feature extraction will be a better performance because it captures more complex relationships within the hormonal and demographic factors contributing to infertility. It could find the non-linear interactions along with subtle patterns not noticeable for the traditional logistic regression model by exploiting the capability of neural network feature extraction. Results from both models were visualized using confusion matrices, Receiver Operating Characteristic (ROC) curves, and feature importance charts to give more profound analysis on how good the model is. Confusion matrices helped understand how correct

and incorrect predictions were distributed across different classes, whereas ROC curves showed the trade-off between sensitivity and specificity for each model. Feature importance charts highlighted the relative importance of different features, offering insights into which hormonal and demographic factors had the most substantial impact on infertility predictions.

RESULTS

Descriptive analysis

Descriptive statistics give a first good view of the data with respect to the infertility treatments in the dataset as shown in Fig.2. The mean age of women (AGE(W)) is 28.19 years, with a standard deviation of 4.65, indicating that most women are in their late twenties to early thirties, a critical age range for fertility considerations. The mean age of men (AGE(H)) is 32.79 years, slightly older than the women, with a standard deviation of 5.68, aligning with common fertility issue age ranges in men. Hormonal levels of LH and FSH have similar distributions, with means close to 6 and wide ranges (LH: 0 to 18.3, FSH: 0 to 27.6), reflecting varied reproductive health statuses among the subjects. The standard deviations for both LH and FSH are approximately 3.2, indicating significant variability. The modes for LH and FSH are 4.78 and 5.4, respectively, suggesting these are common values in the dataset. However, AMH and Prolactin values are marked as “nd” in the mode results, indicating a substantial amount of missing data or entries labelled as ‘not determined’.

The correlation matrix shows a very high positive correlation (0.96) between the ages of women (AGE(W)) and men (AGE(H)), in Fig.3. indicating that couples in our dataset tend to be of similar ages. LH and FSH show slight negative correlations with both AGE(W) and AGE(H) (ranging from -0.12 to -0.17), suggesting a minor decrease in these hormone levels as age increases, which could be related to changes in reproductive system efficiency. There is also some slight negative correlation between LH and FSH (-0.17), suggesting individual variability in hormonal balance that might be the result of PCOS, menopause, or other conditions. The correlations are preliminary in nature and will not be very strong, indicating that perhaps there

Table 1. Comparative Table of Infertility Prediction Studies

Ref.	Focus Area	Key Algorithms/Models	Accuracy
13	Infertility prediction using machine learning	Naïve Bayes, Decision Trees, Multi-Layer Perceptron (MLP)	75%
14	IVF treatment success prediction	MLP (78.2%), SVM (75.5%), C4.5 (72.3%), CART (74.6%), RF (82.1%)	Varies per model (Range: 70%–85%)
17	Prediction of clinical pregnancy rate	Random Forest	Highest in Random Forest 74%
18	Stratification of fertile/infertile couples	Machine Learning Model for Stratification	Varies with models
19	Male sperm count prediction	RF (Random Forest), Gradient Boosting, LASSO, Ridge Regression, XGBoost	(highest for GBR at 78%)
20	Men's fertility rate prediction	Convolutional Neural Network (CNN)	81% for semen prediction
21	Infertility risk prediction in Nigerian women	Multi-Layer Perceptron (MLP)	Highest accuracy achieved: 74.4%
22	Prediction of PCOS in women	Gaussian Naive Bayes (GNB)	Achieved: 91%
23	Characterization of follicular fluid in IVF treatment	PLS (Partial Least Squares), PCA (Principal Component Analysis)	94% – 100%
24	Prediction of IVF live birth occurrence	GBR (Gradient Boosting Regression), Lasso Regression, Linear Regression	(Depending on method)
25	Prediction of varicocele repair success	Brain Project Predictive Analysis (Hybrid AI Model integrating deep learning and statistical methods)	LR: 41%, LaR: 40%, GBR: 78%
26	Prediction of reproductive failure outcomes	Sparse Coding, A.I. Analysis	50% sensitivity, 82% specificity
27	Prediction of poor ovarian response (POR)	ANN, RF, Decision Trees, XGBoost, SVM	Up to 90%
28	Prediction model for endometriosis	Gradient Boosting, CatBoost, SHAP	Highest AUC result achieved with ANN
29	PCOS diagnosis using machine learning	Random Forest, Logistic Regression, Gradient Boosting	81% with CatBoost 91% with RFLR

are other variables not included in this analysis that might have a very significant impact on these relationships, thus pointing toward a greater need for multivariate analyses if one is to gain an understanding of the complex interactions affecting fertility.

The feature importance values obtained from the final logistic regression model indicated the importance of various predictors in determining infertility outcomes as shown in Fig.4. It had FSH, which ended with an importance score of 2.140604, as the most important feature. This was

Table 2. Detailed steps in resampling, splitting, and evaluating data models

Step	Description
Resampling Technique	Random Over Sampler from imblearn oversampling
Random State	42 (to ensure reproducibility)
Resampling Outcome	Balanced datasets Resampled and resampled
Data Split Ratio	75% training, 25% testing (using train_test_split from sklearnmodelselection)
Model	Logistic Regression (mixite=1000)
Evaluation Metrics	Classification report and confusion matrix



Fig. 1. Flowchart of proposed methodology

followed by LH, which had a score of 1.504992, and Prolactin, with 1.266443. These hormonal markers are very important to be used within the prediction of the model and are also relevant hormonal markers in reproductive health. The feature importance analysis indicates that AGE (W) (women’s age) had a relative importance score of 0.1607, suggesting a moderate influence on infertility prediction. In comparison, AMH (0.0578) and AGE (H) (husband’s age) (0.0508) exhibited lower importance scores. While these values are small, they suggest that AMH and

the male partner’s age have a minimal but non-negligible role in the model’s predictive capacity. The statistical significance of these results should be further evaluated to confirm their impact on infertility assessment.

The last model representing results in this research is a logistic regression model, done with the Python “sklearn. linear model” library. This model was trained with the best parameters found from Grid Search: C=1, max_iter=500, solver=’saga’ to guarantee that it will perform at its best and that it converges. It was fitted using

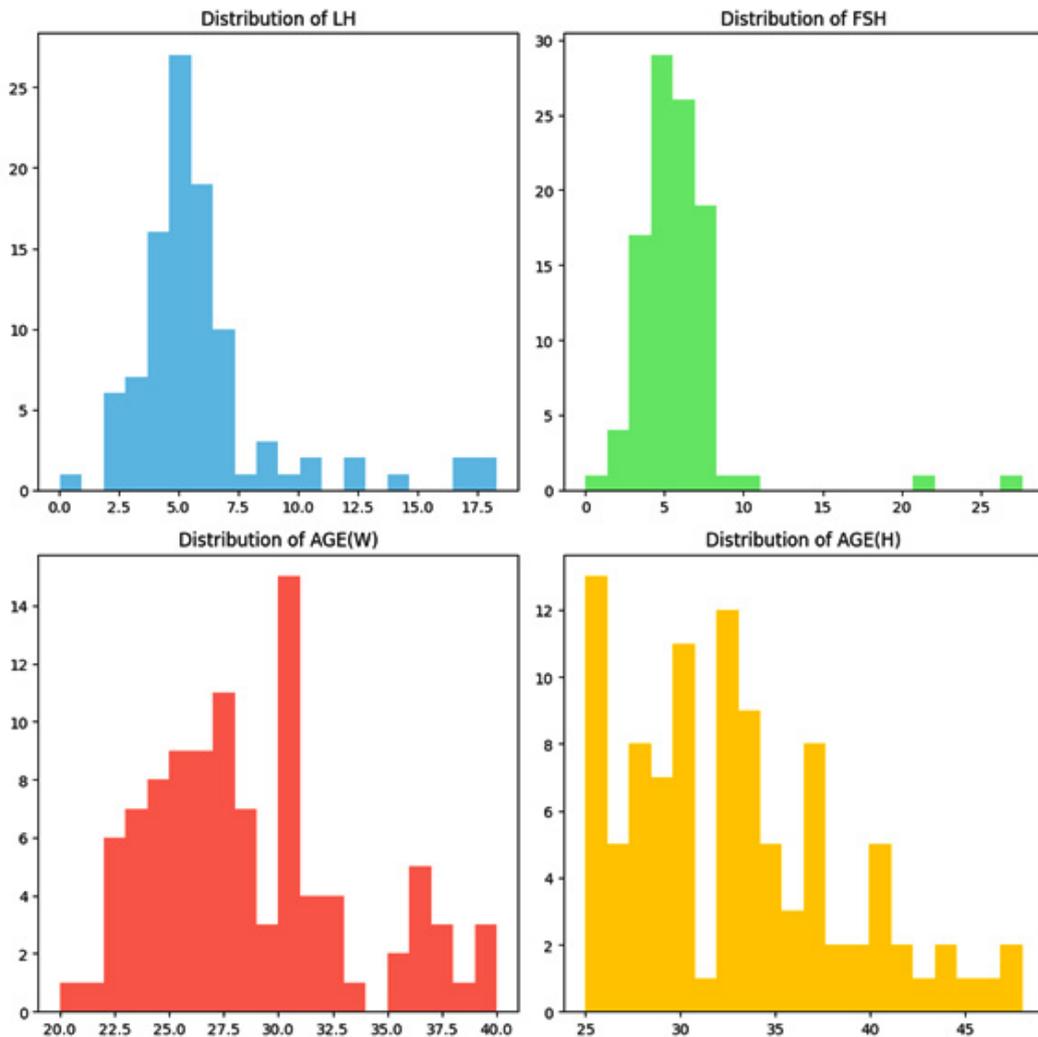


Fig. 2. Descriptive Statistics for Key Variables in the Infertility Dataset

a resampled dataset done by the SMOTEENN technique to handle class imbalance for better generalization across different classes. The ROC curve (Fig. 5) visualizes the model’s performance by illustrating the trade-off between sensitivity and specificity.

The pair plot illustrates the relationships between different levels of hormones, for example, LH, FSH, AMH, Thyroid, and various causes of infertility. Colors are used for different infertility causes, and this plot represents how these levels

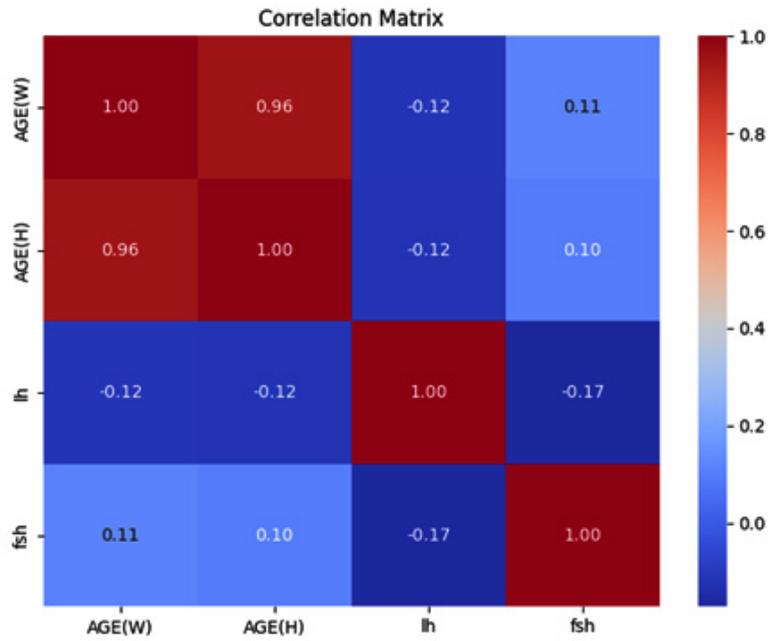


Fig. 3. Correlation Matrix of Key Variables in the Infertility Dataset

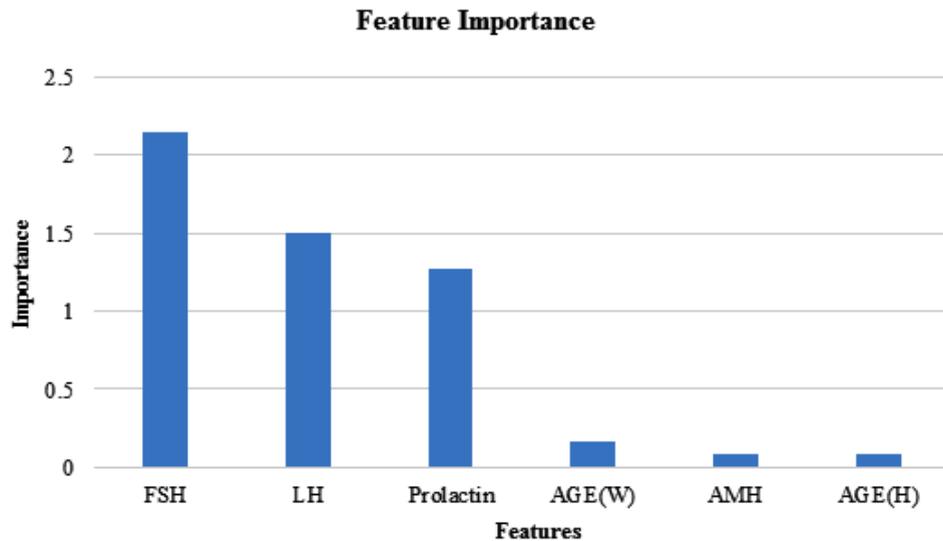


Fig. 4. Feature Importance Scores

of hormones interact and vary across different conditions. This plot helps in manifesting the trends/patterns within the dataset.

The above Fig.6 is a general representation of a pair plot presented fully between levels of different hormones: LH, FSH, AMH and Thyroid against and correlates with these commonly known causes of infertility. Specifically, this type

of graph is useful in the multivariate relationships residing in the dataset of hormone variables by permuting across the classes of infertility. Varying colours are used for each plot, respectively, to be able to identify the patterns and trends, which indicate possible sources of reproductive health problems. The diagonal plots are used to indicate the distribution of each hormone, whereas the off-

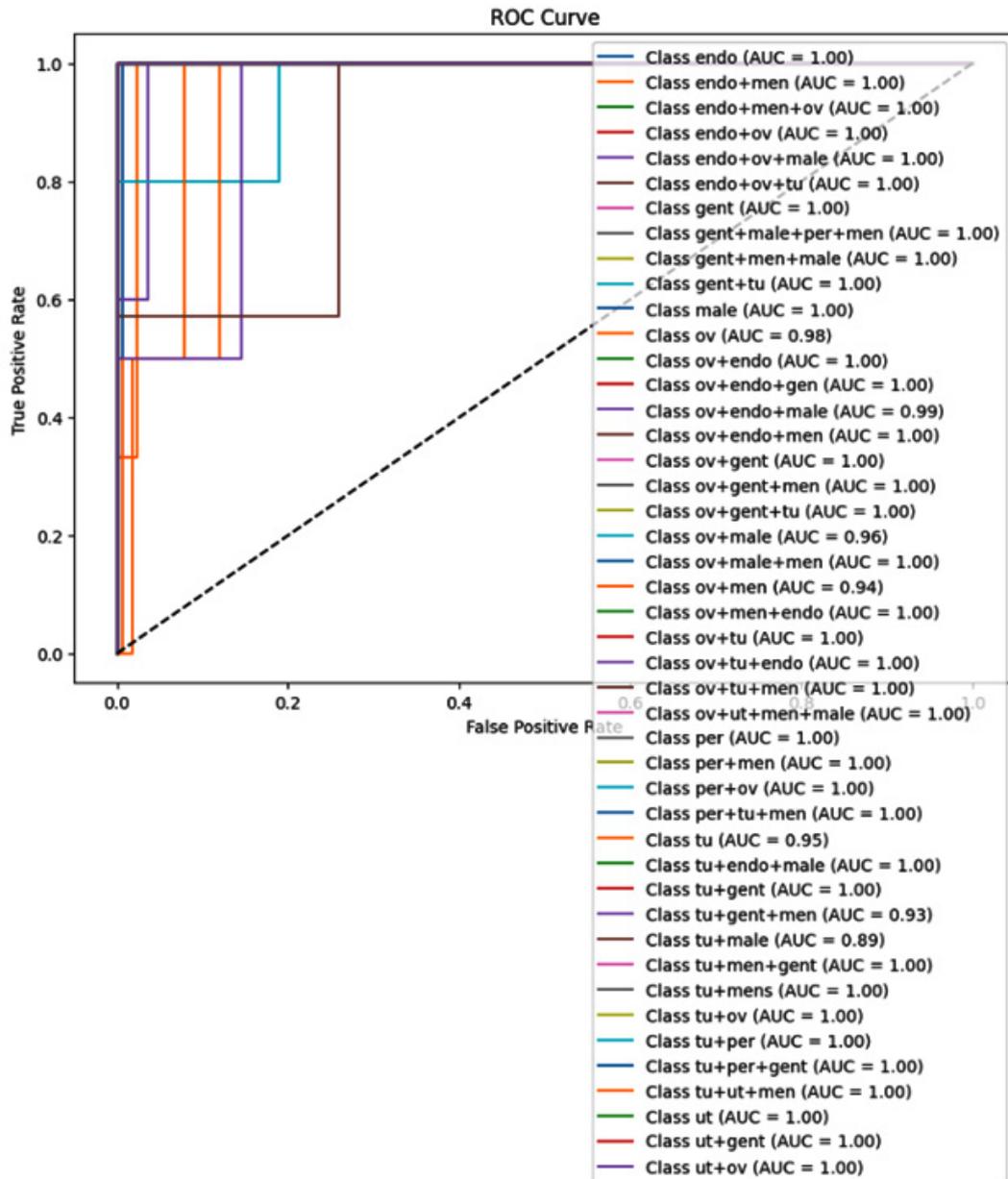


Fig. 5. Curve for the Multi-Class Classification Model in Predicting Causes of Infertility

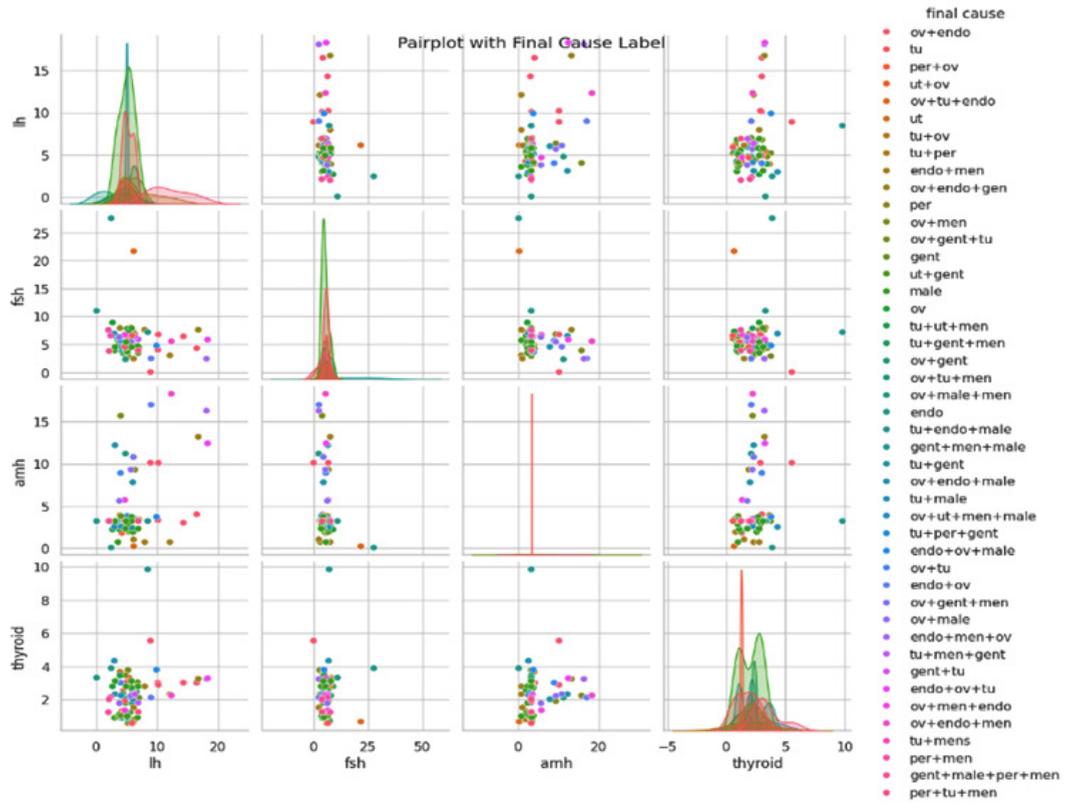


Fig. 6. Pair-plot of Hormonal Levels and Infertility Causes

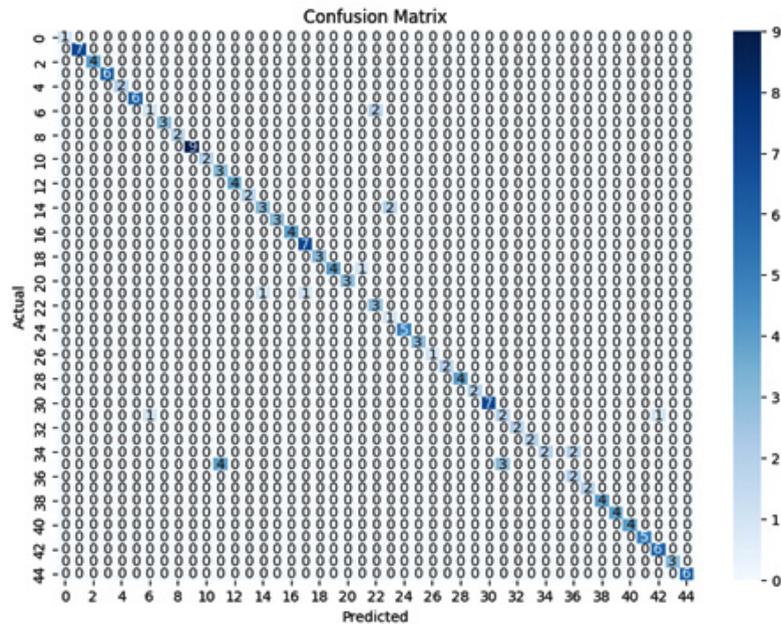


Fig. 7. Confusion Matrix Showing Actual vs. Predicted Classifications for Infertility Causes

diagonal plots help to display pairwise relationships in the data.

The Receiver Operating Characteristic curve, shows the performance of a classification model on multiple classes by plotting the True Positive Rate against the False Positive Rate for every class. In the above case, the diagonal line is the random classifier, and any curve above it performs better than random. Each colored line represents a different class, with the legend showing the class names and their corresponding

AUC (Area Under the Curve) values. Many classes, such as ‘endo’, ‘men+ov’, and ‘ov+tu+endo’, have an AUC of 1.00, indicating perfect classification. Other classes have AUC values slightly less than 1.00, like 0.98 or 0.95, indicating high but not perfect performance. The high AUC values across a wide range of classes demonstrate the model’s robustness and effectiveness in accurately classifying various categories within the dataset, particularly in predicting infertility causes based on hormonal and demographic data as shown in Fig.5.



Fig. 8. Performance Metrics of the Classification Model

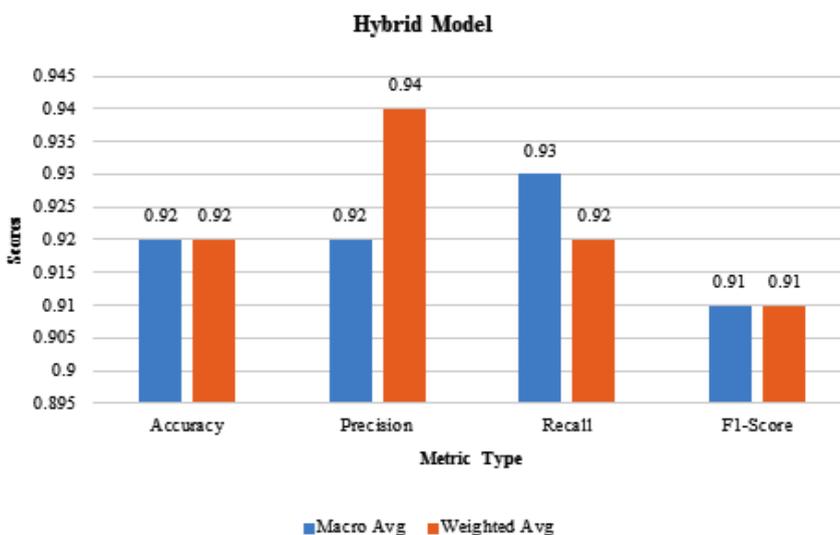


Fig. 9. Performance Metrics of the HyNetReg Classification Model

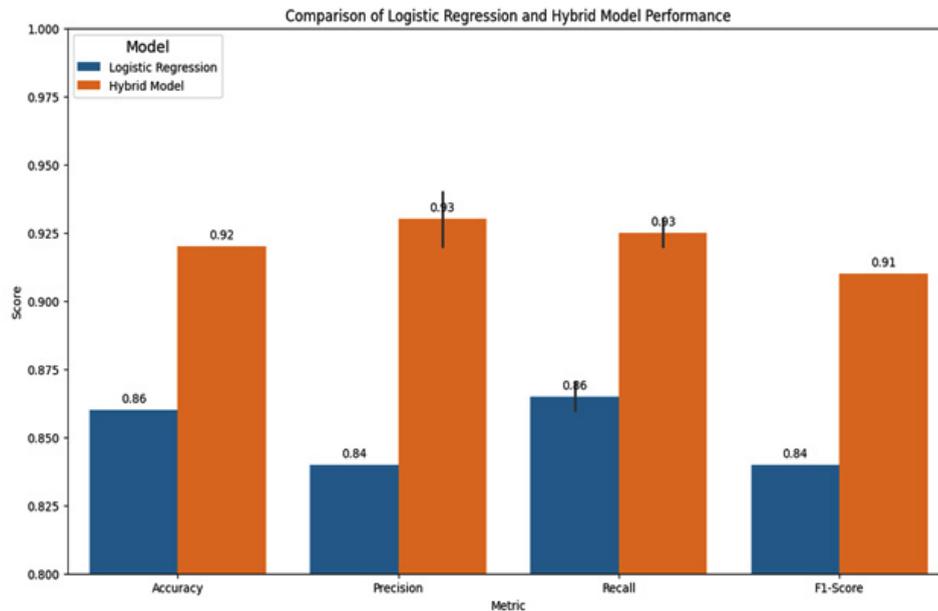


Fig. 10. Comparative Analysis of Performance Metrics both the Models

The confusion matrix displayed in Fig. 7 provides a detailed view of the performance of a multi-class classification model. Each cell in the matrix represents the count of actual versus predicted instances for each class. The diagonal elements indicate how many of each class were predicted correctly, while off-diagonal elements constitute misclassifications. High values for most classes along this diagonal indicate that most have very high true positive rates; that is, most actual classes are predicted correctly. The off-diagonals of some classes are very small, but they do include some misclassifications of the actual class that the model had. Color intensity helps in getting an idea of the concentration of correct and incorrect predictions: the darker the shade, the higher the count.

DISCUSSIONS

Performance Evaluation

This section presented the performance of a multi-class logistic regression model in predicting infertility causes across different classes. It showed in detail the average of precision, recall, and F1-score from all classes. These metrics provide a full view of how accurate the model is, how well

it identifies the relevant instances, and helps know if there is a balance between precision and recall.

The HyNetReg model outperforms traditional logistic regression due to its ability to capture non-linear interactions in hormonal and demographic data. Unlike logistic regression, which assumes a linear relationship, HyNetReg leverages neural network-based feature extraction to enhance predictive accuracy by identifying complex patterns. Key predictors, such as FSH, LH, and Prolactin, gain more significance through this approach, while class imbalance handling using Random Over Sampler ensures better recall. Additionally, regularization techniques prevent overfitting, leading to better generalization. The higher AUC values in ROC analysis further confirm its superior classification performance, making it a more robust tool for infertility prediction.

The performance metrics of the classification model, given in Fig. 8, include the values for accuracy, precision, recall, and the F1 score, with both the macro and the weighted average. Both averages return an accuracy of 0.86, which means this model turns out to be correct 86% of the time. The precision is also steady at 0.84 for both averages, thus, the model is correct in classifying 84% of the positive instances. There

is a slight difference in recall, where macro average is 0.87 and weighted average is 0.86, indicating that there might be a small amount of variance in how well the model is capturing all positives across class imbalances. The F1-score, which is a balance between precision and recall, comes in at 0.84 for both, giving a good, robust model performance across different classes and metrics, not significantly affected by class frequency. It is in this regard that uniformity across both the macro and weighted measures may suggest that this model is well-tuned to offer balanced classification accuracy across the dataset's different class sizes or distributions.

The graph in Fig.9 plots some performance metrics for a classification model, distinguishing between macro and weighted averages. For both averages, accuracy comes to be 0.92, thereby showing the model is working fine at returning expected outcomes. Precision, the proportion of positive predictions that are genuinely so, is also consistent for both averages: 0.92 for the macro average, a shade higher, 0.94, for the weighted average; however, the latter seems to exaggerate precision for more prevalent classes. It is highest in the macro average, indicating a slight advantage in the identification of positive cases across all classes uniformly. The F1 score, balancing precision and recall, is 0.91 for both averaging methods, suggesting that it is a quite robust model performance that manages to balance precision and recall regardless of class distribution. These metrics all emphasize that a model performs well where predictive accuracy is strong and class variability is balanced.

The performance metrics of the logistic regression model versus the HyNetReg model along the four dimensions of accuracy, precision, recall, and F1-score are contrasted in Fig.10. More interestingly, all the metrics for the former are considerably better compared to those for the latter. It has high accuracy of 0.92 as against 0.86 by the logistic regression, underpinning better overall predictive performance. Precision and recall of the HyNetReg model are also much higher, at 0.93, against 0.84 and 0.86, respectively, for the logistic regression model. Trends similar to this can be seen in the F1-score, where the result for the HyNetReg model was 0.91 as compared to the logistic regression's 0.84, hence indicating a

much more balanced performance of the hybrid approach between precision and recall. These improvements in the HyNetReg model metrics likely came from incorporating neural network-based feature extraction, which would aid in modeling more complex patterns that logistic regression might miss, leading to improved accuracy and effectiveness in its classification.

While HyNetReg excels in predictive performance, logistic regression remains preferable for interpretability, computational efficiency, and smaller datasets. Logistic regression provides clear feature importance, making it useful for clinical decisions, while HyNetReg captures complex non-linear patterns but requires more data and computation. For simpler, linearly separable datasets, logistic regression may perform comparably with lower risk of overfitting. Thus, HyNetReg suits complex cases, while logistic regression remains valuable for transparent, efficient modeling.

CONCLUSION

It should be noted that the study showed the efficacy of using a HyNetReg machine learning model in infertility analysis not only for predictive accuracy but also for balanced performance measures against the traditionally used logistic regression model. Neural network-based feature extraction combined with logistic regression resulted in a hybrid model that performed much better than the logistic regression model on all metrics evaluated, realizing big improvements in accuracy to 0.92, precision up to 0.94, recall to the extent of 0.93, and an F1-score of 0.91. These improvements thus signify the need for advanced feature extraction methodologies that may capture hormone and demographic data in a more complex relationship, thus enriching the model's ability to make out subtle nuances associated with different infertility causes. The current study thus had potential in terms of advanced machine learning methodologies for enhancing diagnosis accuracies in medical science, more specifically in the challenging domain of reproductive health.

ACKNOWLEDGMENTS

The authors wish to express their gratitude

to the Centre for Research Projects (CRP) at CHRIST (Deemed to be University), Bangalore Central Campus, Bengaluru, India, for generously supporting this research with Seed Money for the academic year 2023-24 (Project Number CU: CRP: SMSS-2351).

Funding Sources

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest

The author(s) do not have any conflict of interest.

Data Availability Statement

This statement does not apply to this article.

Ethics Statement

This research did not involve human participants, animal subjects, or any material that requires ethical approval.

Informed Consent Statement

This study did not involve human participants, and therefore, informed consent was not required.

Clinical Trial Registration

This research does not involve any clinical trials

Permission to reproduce material from other sources

Not Applicable.

Author Contributions

Upreti K designed the research; George JP and Hundekari S performed the research; Alam MS contributed to data & sample collection; Upreti K and George JP contributed analytic tools and analyzed the data; Upreti K and Hundekari S wrote the paper.

REFERENCES

- Legese, N., Tura, A. K., Roba, K. T., & Demeke, H. The prevalence of infertility and factors associated with infertility in Ethiopia: Analysis of Ethiopian Demographic and Health Survey (EDHS). *PLoS ONE*, 2022;18(10):e0291912. <https://doi.org/10.1371/journal.pone.0291912>
- Zhao, X., Liu, Y., Zhang, A., Gao, B., Feng, Q., Huang, H., ... & Xu, D.. Logistic regression analyses of factors affecting fertility of intrauterine adhesions patients. *Annals of translational medicine*, 2020;8:4. doi: 10.21037/atm.2019.11.115
- Mazzilli, R., Rucci, C., Vaiarelli, A., Cimadomo, D., Ubaldi, F. M., Foresta, C., & Ferlin, A. Male factor infertility and assisted reproductive technologies: indications, minimum access criteria and outcomes. *Journal of Endocrinological Investigation*, 2023;46(6):1079-1085 <https://link.springer.com/article/10.1007/s40618-022-02000-4>.
- Hashemian, F., Maleki, N., & Zeinali, Y. From User Behavior to Subscription Sales: An Insight Into E-Book Platform Leveraging Customer Segmentation and A/B Testing. *Services Marketing Quarterly*, 2024;45(2):153–181. <https://doi.org/10.1080/15332969.2024.2313873>
- Huang, L., Shiao, W., & Lin, Y. What factors satisfy e-book store customers? Development of a model to evaluate e-book user behavior and satisfaction. *Internet Research*, 2017;27(3):563–585. <https://doi.org/10.1108/intr-05-2016-0142>
- Paffoni, A., Vitagliano, A., Corti, L., Somigliana, E., & Viganò, P. Intracytoplasmic sperm injection versus conventional in vitro insemination in couples with non-male infertility factor in the 'real-world' setting: analysis of the HFEA registry. *Journal of Translational Medicine*, 2024;22(1):687. <https://link.springer.com/article/10.1186/s12967-024-05515-x>
- Sheikhian, M., Loripoor, M., Ghorashi, Z., & Safdari-Dehcheshmeh, F. Relation between sexual function, perceived social support, and adherence to treatment with infertility factor in women: A cross-sectional study. *International Journal of Reproductive BioMedicine (IJRM)*. 2023. <https://doi.org/10.18502/ijrm.v21i7.13893>
- GhoshRoy, D., Alvi, P. A., & Santosh, K. C. AI tools for assessing human fertility using risk factors: A state-of-the-art review. *Journal of Medical Systems*, 2023;47(1):91. <https://link.springer.com/article/10.1007/s10916-023-01983-8>
- Shofiyah, S., & Mahmudy, W. F. Exploring Machine Learning Techniques for Male Infertility Prediction: A Review. In *Proceedings of the 8th International Conference on Sustainable Information Engineering and Technology* 2023;235-240. <https://dl.acm.org/doi/abs/10.1145/3626641.3627146>
- Raef, B., & Ferdousi, R. A review of machine learning approaches in assisted reproductive technologies. *Acta Informatica Medica*, 2019;27(3):205. doi: 10.5455/aim.2019.27.205-211
- Ranjini, K., Suruliandi, A., & Raja, S. P. Machine learning techniques for assisted reproductive technology: A review. *Journal of Circuits*,

- Systems and Computers*, 2020;29(11):2030010. <https://doi.org/10.1142/S021812662030010X>
12. Zhao, X., Liu, Y., Zhang, A., Gao, B., Feng, Q., Huang, H., ... & Xu, D. Logistic regression analyses of factors affecting fertility of intrauterine adhesions patients. *Annals of translational medicine*, 2020;8(4). doi: 10.21037/atm.2019.11.115
 13. Wang, L., Tang, Y., & Wang, Y. Predictors and incidence of depression and anxiety in women undergoing infertility treatment: A cross-sectional study. *PLoS one*, 2023;18(4):e0284414. <https://doi.org/10.1371/journal.pone.0284414>
 14. Balogun, J. A., Egejuru, N. C., & Idowu, P. A. Comparative analysis of predictive models for the likelihood of infertility in women using supervised machine learning techniques. *Computer Reviews Journal*, 2018;2(1):313-330. <https://d1wqtxts1xzle7.cloudfront.net/106597674>
 15. Hassan, M. R., Al-Insaif, S., Hossain, M. I., & Kamruzzaman, J. A machine learning approach for prediction of pregnancy outcome following IVF treatment. *Neural computing and applications*, 2020;32(7):2283-2297. <https://link.springer.com/article/10.1007/s00521-018-3693-9>
 16. Mehrjerd, A., Rezaei, H., Eslami, S., Ratna, M. B., & Khadem Ghaebi, N. Internal validation and comparison of predictive models to determine success rate of infertility treatments: a retrospective study of 2485 cycles. *Scientific reports*, 2022;12(1):7216. <https://www.nature.com/articles/s41598-022-10902-9>
 17. Bachelot, G., Lévy, R., Bachelot, A., Faure, C., Czernichow, S., Dupont, C., & Lamazière, A. Proof of concept and development of a couple-based machine learning model to stratify infertile patients with idiopathic infertility. *Scientific Reports*, 2021;11(1):24003. <https://www.nature.com/articles/s41598-021-03165-3>
 18. Huang, H. H., Hsieh, S. J., Chen, M. S., Jhou, M. J., Liu, T. C., Shen, H. L., ... & Lu, C. J. Machine learning predictive models for evaluating risk factors affecting sperm count: predictions based on health screening indicators. *Journal of Clinical Medicine*, 2023;12(3):1220. <https://www.mdpi.com/2077-0383/12/3/1220>
 19. Naseem, S., Mahmood, T., Saba, T., Alamri, F. S., Bahaj, S. A. O., Ateeq, H., & Farooq, U. DeepFert: An intelligent fertility rate prediction approach for men based on deep learning neural networks. *IEEE Access*, 2023;11:75006-75022. <https://ieeexplore.ieee.org/abstract/document/10168104>
 20. Idowu, P. A., Sarumi, S. O., & Balogun, J. A. A Prediction Model for the Likelihood of infertility in women. In *9TH International Conference on Information and Communications Technology (ICT) Applications, Ilorin, Kwara 2017*;78-88. <https://doi.org/10.1142/S021812662030010X>
 21. Nasim, S., Almutairi, M. S., Munir, K., Raza, A., & Younas, F. A novel approach for polycystic ovary syndrome prediction using machine learning in bioinformatics. *IEEE Access*, 2024;10:97610-97624. <https://ieeexplore.ieee.org/abstract/document/9885199>
 22. Jakubczyk, P., Paja, W., Pancercz, K., Cebulski, J., Depciuch, J., Uzun, Ö., ... & Guleken, Z. Determination of idiopathic female infertility from infrared spectra of follicle fluid combined with gonadotrophin levels, multivariate analysis and machine learning methods. *Photodiagnosis and Photodynamic Therapy*, 2022;38:102883. <https://www.sciencedirect.com/science/article/abs/pii/S1572100022001697>
 23. GS, G., NAYAK, S., & CHOLLI, N. Evaluation of Machine Learning Algorithms on the Prediction of Live Birth Occurrence. *International Journal of Pharmaceutical Research (09752366)*, 2021;13:2. <http://www.ijpronline.com/ViewArticleDetail.aspx?ID=21146>
 24. Crafa, A., Russo, M., Cannarella, R., Gül, M., Compagnone, M., Mongioi, L. M., ... & Calogero, A. E. Predictability of varicocele repair success: preliminary results of a machine learning-based approach. *Asian Journal of Andrology*, 2024;10-4103. https://journals.lww.com/ajandrology/fulltext/9900/predictability_of_varicocele_repair_success_218.aspx
 25. Huang, C., Xiang, Z., Zhang, Y., Tan, D. S., Yip, C. K., Liu, Z., ... & Tu, W. Using deep learning in a monocentric study to characterize maternal immune environment for predicting pregnancy outcomes in the recurrent reproductive failure patients. *Frontiers in Immunology*, 2021;12:642167. <https://doi.org/10.3389/fimmu.2021.642167>
 26. Yan, S., Jin, W., Ding, J., Yin, T., Zhang, Y., & Yang, J. Machine-intelligence for developing a potent signature to predict ovarian response to tailor assisted reproduction technology. *Ageing (Albany NY)*, 2021;13(13): 17137. <https://pubmed.ncbi.nlm.nih.gov/articles/PMC8312467/>
 27. Blass, I., Sahar, T., Shraibman, A., Ofer, D., Rappoport, N., & Linial, M. Revisiting the risk factors for endometriosis: a machine learning approach. *Journal of Personalized Medicine*, 2022;12(7):1114. <https://doi.org/10.3390/jpm12071114>

28. Oyegoke, T., Amoo, A., Balogun, J., Alo, T., & Idowu, P. A Predictive Model for the Risk of Infertility in Men using Fuzzy Logic. *International Journal of Computers*, 2019;4. <https://d1wqtxts1xzle7.cloudfront.net/75850734/006-00042019>
29. Bharati, S., Podder, P., & Mondal, M. R. H. Diagnosis of polycystic ovary syndrome using machine learning algorithms. In *2020 IEEE region 10 symposium (TENSYP)* 2020;1486-1489. IEEE. <https://ieeexplore.ieee.org/abstract/document/9230932>