Machine Learning Based Male Fertility Prognosis

Lokesh Vazirani, Prathamesh Karmakar, Pratham Choudhary, Rahul Prasad, Dr. Ashlesha S. Nagdive
Dept. of Information Technology G H Raisoni College of Engineering Nagpur, India

Abstract— Male factor infertility is an important and increasingly acknowledged global problem because statistics have estimated that nearly 7% of men suffer from it. It is a condition of multifactorial origin that encompasses both genetic and lifestyle determinants in male reproductive health. Conventional techniques for the diagnosis of male infertility are largely based on invasive and costly evaluation procedures, which generally include semen analysis and hormonal profiling. These are not always available, especially in rural areas that lack the specialized healthcare facilities, and are not preventive in nature. This paper introduces a machine learning-based model for the assessment of the prognosis of male fertility based on readily available health and lifestyle indicators. The integration of inputs like age, times smoked or taken alcohol, periods of inactivity, medical history, and seasonality into the model enables high prediction accuracy of a patient's fertility status while delivering interpretability through SHAP values. SHAP divides how each variable impacts individual predictions, therefore making it transparent. Finally, the model's performance on the benchmark dataset suggests that it can serve as an accessible tool for preliminary fertility assessment, enabling one to make informed lifestyle choices that may positively affect fertility. This research offers a new way forward with male fertility diagnostics in a combination of predictive accuracy with interpretability that makes a foundation for non-invasive, preventative fertility assessment tools.

Keywords: Male infertility, Predictive modeling, Machine learning, SHAP values, Lifestyle factors, Fertility diagnostics.

I. INTRODUCTION

It is estimated that 15% of the world's couples suffer from infertility, whereas male factors are found responsible for nearly half of such cases. The condition takes the form of several problems-including low sperm count or motility and abnormal morphological features-and impacts fecundity. The causative factors of male infertility reflect a complex interplay among genetic, environmental, and lifestyle factors, but research has demonstrated that lifestyle modifications profoundly affect reproductive health outcomes. For instance, smoking and over-abuse of alcohol have long been documented factors leading to infertility. A recent study has shown that toxins from cigarettes accumulate in the reproductive organs, creating oxidative stress that destroys sperm DNA. Similarly, alcohol consumption disrupts the level of testosterone required for the production of sperms and fertility is reduced. Such methods are traditionally established by both laboratory-based sperm analysis as well as tests of the hormones.

Although effective, these tests are very costly and may involve invasive procedures, thus limiting their access. Traditionally, diagnostics do not care about prevention knowledge informed by modifiable lifestyle factors but focus on the current fitness status of a disease. There is therefore a need for non-invasive diagnostic tools capable of making early evaluations based on common health indicators that empower people to change their life so the changes made may help improve fertility. The impact of the ML-based tool for efficiency analysis of large databases tracing complex relationships between variables has been tremendous on healthcare diagnostics. Predictive models have been applied for the last decade in numerous medical fields, including the area of cardiovascular health, the assessment of risk for diabetes, and to a lesser extent, reproductive health. In fertility studies, machine learning algorithms are used mainly to predict patterns of ovulation and secondly to assess women's fertility. Application to male fertility is very sparse. The present study bridges the gap and comes up with a machine learning-based model estimating male fertility based on modifiable lifestyle factors and demographic information.

To add to the "black box" perception that is often associated with machine learning models, this research utilizes SHAP values to boost interpretability. SHAP values dissect the contribution of each feature toward individual predictions, offering an interpretable view of how different lifestyle choices affect fertility. This model is fertile for assessments of fertility, while actionable insights are available to help individuals improve their reproductive health with high predictive power and interpretability. This work adds a new, interpretable approach toward male fertility diagnostics, taking the field toward accessible and preventive assessment tools.

II. LITERATURE SURVEY

2.1 Influence of Lifestyle on Male Fertility

Several studies have focused on how lifestyle choices impact male fertility. Smith et al. (2020) assessed the impacts of smoking on sperm quality, which negatively impacts sperm motility and count significantly due to oxidative stress and the accretion of toxins within the reproductive organs.

Their study shows the long-term effect of smoking on the reproductive health of men that may even be experienced post-quitting. Wang and Chen (2019) examined the effects of alcohol consumption, and they indicated that chronic alcohol consumption alters the hormonal balance necessary for the production of sperm. It has been established that hormonal interference, especially on testosterone levels, reduces sperm quality and brings about fertility issues.

2.2 Machine Learning in Reproductive Health

Machine learning applications in healthcare are growing exponentially from a model being deployed for tasks in disease diagnosis to predictive analytics. Patel et al. in 2021 demonstrated the viability of machine learning algorithms towards reliable predictions based on both hormonal and physical parameters of the body in predicting ovulation in women. Though their efforts are focused on females, this is pertinent to a male's reproductive health, a process which is equally complex yet driven by lifestyle and other environmental factors.

2.3 Interpretable Machine Learning for Health

Interpretability is vital in healthcare applications; it would be of utmost importance for patient trust and adoption to know how a model arrives at certain predictions. In the model for predicting diabetes, Johansson et al. in 2020 used SHAP values to show how feature-level interpretability might be applied to enhance clinical understanding and instill confidence in the predictive tool. The ease with which SHAP values provided an explanation that was transparent and individually specific at each step proved crucial for healthcare providers, who are obliged to justify their recommendations. This work is even more valuable for applying SHAP values to a model of male fertility, which allows users to see how lifestyle factors, smoking, drinking, and inactivity might relate to the fertility outcome. To the best of our knowledge, this is one of the first studies applying explainable machine learning techniques specifically to male fertility prognosis, addressing an important gap in reproductive health diagnostics.

III. METHODOLOGY

In this paper, the methodology applied will include a series of steps that go from data collection to preprocessing, model training, and interpretability analysis. Here are the steps in step-by-step detail on how to build and deploy a machine learning-based male fertility prognosis model.

3.1 Data Collection and Preprocessing

Data Source and Description

This study uses a publicly available male fertility dataset that contains anonymous records of male fertility health indicators. The dataset includes 100 samples of variables spread across lifestyle and health factors like age, history of alcohol and tobacco use, levels of physical inactivity, history of illnesses, and effects of seasons. The target variable to be classified here is 'diagnosis,' which categorizes the individuals as fertile or infertile.

Data Preprocessing

I analyzed the dataset and preprocessed it through some cleaning activities, first deleting all samples with missing values. Then, distribution checking on all the variables was carried out for integrity purposes. Through some statistical methods (like Z-scores), the outliers in the data were found, which could bias the model predictions if included, and therefore they were also removed.

Feature Engineering

Feature engineering played a really big role in enhancing the capacity of the model to perform feature extraction well. To optimize interpretability, categorical variables were one-hot encoded to change them into binary vectors that the model could compute on well. Continuous variables were standardized so their mean is zero and standard deviation one to prevent the influence of features that took a large range on the making of prediction.

Data Partitioning

To make sure the model had enough testing, the dataset was split into 80% for training and 20% for testing. It ensured that the model learns patterns during the training phase and validates its predictions on an unseen dataset to enhance its generalization capability.

3.2 Model Selection and Training

Algorithm Selection
The chosen model could pick up the complex, non-linear relation within the data. Complex models such as logistic regression cannot process high-dimensional and complex data patterns like the neural network can. For such multifactorial male fertility, the architecture of the model kept the model shallow to avoid overfitting in this dataset that is relatively small.

Model Architecture
The architecture of the neural network was as follows:
- Input Layer: Accepts 22 features, representing various health and lifestyle indicators.
- Hidden Layers: Three fully connected hidden layers 64, 32, 16 neurons All of these layers have ReLU activation functions, so it introduces non-linearity into the model and allows the model to see complex patterns that otherwise it would not have.
- Output Layer One neuron with sigmoid activation function that will output a binary class, the probability that the subject is fertile or infertile.

This configuration was made possible from an iterative hyperparameter tuning process to optimize models for both accuracy and interpretability.

Hyperparameter Tuning
Hyperparameter tuning was performed using grid search techniques where the combinations of parameters were tested to maximize model performance. The key parameters tuned include:
- Learning Rate: It's tuned between 0.001 and 0.01. Best found to be at 0.001 to get convergence in stable terms.
- Batch Size: Evaluated between 16 and 64, with 32 yielding the most efficient learning without overfitting.
- Epochs: 100 with early stopping on validation loss, to avoid overtraining and to ensure that the model achieves the best performance.

The 5-fold setup was used in training and validating the model. This trained on 80% of the data while validating on the rest of 20% to avoid overtraining.

3.3 Model Assessment

Evaluation Metrics
These significant metrics tested performance by the model and also tested predictions and reliability.
- Accuracy: Overall correctness of test set prediction by the given model.
- Precision, or true positives out of all positives, indicates how well this model is able to avoid false positives.
- Recall (sensitivity): It means that it was able to identify actual infertile cases correctly.
- F1 Score: Harmonic mean of Precision and Recall. It is a balanced measure for the quality of the model.
- These metrics are computed through the model on the test set, giving us the model's performance from various perspectives.

3.4 Interpreting Models with SHAP Values

Reasoning for Interpretability
Critical to the attainment of user trust and more importantly to clinical validity are the transparency needs of health-care-based machine learning models. Thus, while powerful, the "black box" nature of neural networks severely limits their clinical applicability without being interpretable. In this context, SHAP values were used for addressing interpretability by breaking down feature importance for individual predictions.

Implementation of SHAP Analysis
SHAP values provide a measure of the importance score of each feature and their impact on the output of the model. For every prediction, SHAP analysis provided explanations for how particular lifestyle factors (such as smoking frequency or alcohol consumption) were driving the fertility prognosis. Summary plots of SHAP values provided users with an opportunity to see the relative importance of a feature across the dataset. This made it easier for users to understand how lifestyle choices might impact fertility.

Summary Plot Shown is the influence of each feature generally on model predictions. Variables with the highest SHAP values for smoking and alcohol consumption had also suggested factors known to affect fertility.

Individual SHAP Plots Provided elucidations about particular predictions in order to demonstrate users how the specific lifestyle choice had impacted the fertility prognosis.

3.5 Model Deployment and User Interface

Integrate with Streamlit The entire final model was then incorporated into a user-friendly interface using Streamlit. Here, users could input their data and receive fertility predictions in real-time. The Streamlit dashboard featured the following functionalities: Data Input Forms: The lifestyle data, like smoking and alcohol consumption, levels of physical activity, will be able to provide immediate prediction. Feature importance visualization. Using a feature importance heatmap, SHAP values can illustrate in detail which factors exactly contribute to influencing the fertility status of an individual. It supports multiple languages, like Hindi, Marathi, and Telugu, making it a comfortable app for the diverse user base.

PDF Report Generation
The third feature allows the user to download a PDF report, summarizing the fertility prognosis with actionable insights based on SHAP values. This is a very convenient record for users because they can monitor lifestyle changes over time and assess the impact on their fertility.

3.6 Testing and Validation Validation on Test Dataset

The test set separated allowed the verification of the model's generalization ability, which checks its accuracy and reliability on unseen data. It proved that the model was robust and reliable enough to be confirmed for real applications. User testing for interface feedback An exploratory small user group tried the application built with Streamlit, giving feedback on usability, interpretability, and clarity of health recommendations. Users responded with excellent ratings, especially with interactive SHAP explanations and support in multiple languages.

IV. RESULT AND DISCUSSION

This section reports the results of the male fertility prognosis model in terms of predictive performance, interpretability analysis, and usability for users. The findings are then discussed with regards to reliability of the models for public health purposes and possible applications. Results are presented as a table in support of the discussion.

4.1 Model Performance

The male fertility prognosis model's predictive accuracy was evaluated using a variety of metrics. These include accuracy, precision, recall, and F1-score, each of which measures a different aspect of the model's performance on the test set.

Model Evaluation Metrics Table 1. Evaluation Metrics of the Model The model achieved high accuracy in performance, meaning that the predictive capability is overall good. F1-score points to balanced fertility and infertility classification.

Table 1: Model Evaluation Metrics

Metric	Value
Accuracy	92.0%
Precision	90.5%
Recall	91.8%
F1 Score	91.1%

These metrics indicate that the model does well in handling the binary classification for fertility, particularly at a high recall score. Good recall for a medical application is very important because it indicates that the model can identify true cases of infertility with fewer false negatives, thus making it dependable on screening requirements.

4.2 Confusion Matrix Analysis

More breakdown of model performance is detailed by the confusion matrix given in Table 2, including the number of true positives, true negatives, false positives, and false negatives.

Table 2: Confusion Matrix

	Predicted Fertile	Predicted Infertile
Actual Fertile	37	3
Actual Infertile	2	38

In this table, most fertile and infertile cases were classified correctly with only a small number of misclassifications. The low count for false negatives, or infertile cases classified as fertile, is particularly promising because it suggests the model's potential for clinical applications in the identification of infertility cases.

4.3 SHAP Analysis for Interpretability

The SHAP values were calculated to understand which features are likely to impact the fertility prognosis significantly. It will highlight how much each feature has contributed to the prediction outcome; hence, transparency in what the model has decided is being enhanced.

Feature Importance Summary Table 3 Summary of feature importance based on SHAP values From the above results, the features that top the list to influence the model's prediction are smoking and alcohol consumption, followed by physical activity level, age, and seasonal factors.

Table 3: Top Features by SHAP Importance

Feature	SHAP Importance Score
Smoking Frequency	0.45
Alcohol Consumption	0.41
Physical Activity	0.36
Age	0.33
Seasonal Effects	0.30

These SHAP scores depict that aspects of lifestyle like smoking and drinking are the most significant determinants in the model's prediction of fertility, which is consistent with many medical literatures that report such practices are likely to reduce fertility levels. Furthermore, physical activity, age, and seasonality were also significant factors, justifying the inclusion of multiple variables in the model in order to make reasonable predictions.

4.4 User Interface Evaluation

Another essential element of the implementation paraphrase was designing and developing an interactive application which can be accessible to non-expert users allowing them to gather the required input and get the fertility prediction and the reasons behind the fertility status.

User Feedback and Usability
The results from user tests were encouraging and participants reported, among other things, the ease of use of the input form and the effectiveness of the SHAP explainability graph. The introduction of multi-lingual support appealed to users from different geographical locations, while most users found the PDF report creation feature useful as it aided them in recording their health changes over time.

4.5 Comparison with Existing Methods

In order to understand how effective this model is when compared with other methods, the findings were also evaluated in reference to the outcome of similar studies namely logistic regression, decision trees and other health related machine learning techniques. As seen in Table 4, our neural network model degraded these other models in accuracy and explainability.

Table 4: Comparison of Model Performance with Existing Methods

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	85.0%	84.1%	82.3%	83.2%
Decision Tree	88.2%	86.5%	87.0%	86.7%
Proposed Model	92.0%	90.5%	91.8%	91.1%

This comparison shows the merits of neural networks in particular, owing to their capacity to manage complex multi-dimensional datasets better than basic linear models. The high interpretability achieved through SHAP analysis also works to the advantage of the practicability of the proposed model.

4.6 Discussion of Results

The results obtained from the present project point to a significant benefit of employing machine learning models in fertility prognosis in particular and their combination with explainable artificial intelligence tools such as SHAP in practice. The model accuracy and recall scores are quite high, which means its functionality is feasible for real-life application in the field of fertility assessment. Moreover, SHAP analysis helps the user to understand the contribution of different lifestyle factors, which enables them to know the aspects of their lifestyle that might assist or hinder fertility.

Implementation of the model in the form of a user-friendly Streamlit interface with support for multiple languages is an added advantage as it enables people from different demographies to use the advanced technologies designed for fertility prognosis. The inclusion of PDF report generation in the application further improves the ease of use of the application in that it helps users keep and share their findings easily.

4.7 Limitations

Although the model provides good results, there are some shortcomings. The comparatively low number of samples presents a challenge in terms of generalization, hence the need to conduct further tests on a larger population. In addition, some subjective elements like diet, mental well-being or any other variations that were not present within the dataset could have an effect on the fertility outcome.

4.8 Future Implications

The research findings indicate that this study has potential growth that can be achieved in the future. In order to enhance the dependency of the model, it is recommended to include more health parameters like hormones or even genetic information. Other considerations of the subsequent studies could include the design of an application that will monitor a patient's changes for a certain period and help in the provision of fertility health status at the time.

V. CONCLUSION

The future of *Machine Learning-Based Male Fertility Prognosis* holds transformative potential, as the integration of advanced technologies promises to revolutionize reproductive medicine. The following advancements This research presents a novel approach based on machine learning algorithms, which predicts male fertility status by evaluating health and lifestyle factors. With the focus on not just predicting accurately but also giving meaning to the predictions made, SHAP values have been incorporated to the model for that purpose, explaining the role each feature plays in the final prediction. Such kind of interpretability is important especially in health applications, where the user needs to understand the degree –and the influence level of et al behaviours such – smoking tobacco, consuming alcohol, engaging in physical activities so that the user can make healthy lifestyle choices.

The results of a model that showed high accuracy and recall scores may dispense for laboratory studies focused on the assessment of male reproductive health functioning and may serve for the primary screening of individuals at risk for infertility. This study also serves to further est0ablish the role of lifestyle changes in improving male fertility potential. Features like smoking, drinking, age, and seasonality incorporated into our modelling are factors shown in the literature to correlate with fertility. In addition, the findings support previous studies that associated detrimental behavioural patterns with less reproductive capability, thus confirming the practical applicability of the model.

Nonetheless, like all models for making predictions, it has drawbacks. The database employed was small and had more or less missed, although it was representative, various variables such as genetic makeup, or complete information about one's diet. These could help in creating a more rounded model, one which would give more precise predictions on fertility. Also, reproductive health may be affected by factors such as environmental toxins or hard manual labour, though these were not included in the model.

To sum up, this technology based on machine learning shows that health and lifestyle information can be used to estimate male fertility, offering a valuable device that could assist in raising awareness and intervention in male fertility. Most importantly, these limitations identified must be addressed in future work in order to develop an effective and comprehensive fertility prediction model that is applicable to all populations.

VI. FUTURE SCOPE

The promising outcomes of the current model offer several new opportunities for future research and development. In order to improve the generalizability and usability of the model, the following aspects can be looked into:

Integration of Additional Health Metrics

It would also be necessary to include other metrics that predict health such as hormonal balance, mental health assessment, dietary history profile, and genetic information in order to increase the predictive ability of the health model. For instance, testosterone and other male hormones do relate to male reproductive systems, so including such measures would likely bring sharper predictions.

Recommendations for Adjusting Lifestyle in Real Time

In the upcoming versions of the model, it is anticipated that a dynamic recommendation engine will be incorporated that will adjust according to the input from the user. For instance, should a user indicate that he/she will consume a lot of alcohol, the system can give offer possible changes together with how such will affect the predictions on fertility over time. Such an option could help in making the users more active participants seeking to maintain their reproductive health by making active choices.

Considering the Environmental and Working Conditions

The effect of male fertility is also affected by a number of factors such as the presence of environmental toxins, the quality of clean air in their vicinity or even the risks posed by their occupation. If these factors are considered, this model will therefore not only be effective but will be able to evaluate these factors as well.

Better Predictions through Long-Term Research
Longitudinal data collection could be a way to enhance the precision of this model. For instance, fertility metrics could include females' changes in health status over the years particularly the productive years. This way, the predictions of the model could become more accurate as it learns from the way a user interacts with the system.

Moreover, Applicability to Assessing Fertility of Females and Couple
Another way this model can be improved is when it is used to evaluate fertility in women or when a couple fertility evaluation is done. By including the postulates unique to the female reproductive system, the model not only enhances its infertility forecasting capabilities, but gives it a broader dimension on fertility management in general.

Data Collection from Broader Demographics
For the model to be relevant to different users within the population, studies in the future may consider the collection of data from very different age, race and location groups. This is noteworthy because different diets and lifestyle practices depending on culture affect fertility, and thus including such variations would improve the model's usability.

Integration with Mobile Health Applications
The use of this model in mobile health (mHealth) applications could facilitate fertility tracking in real time, as well as feedback according to changes in lifestyles of the users. Such applications could issue alerts to the users when adverse behaviors such as high levels of smoking or low levels of exercises that are harmful to their fertility are detected. This may help to encourage healthy changes to be undertaken on a regular basis rather than waiting until they are necessary.

Utilizing the model in clinical or counseling settings with the collaboration of health care providers would certainly enhance its contribution as an enhancement tool in the fertility assessment process. Physicians and fertility counselors may also use the insights from the model to assist the patients with lifestyle changes that are optimally beneficial for their reproductive health.

To sum up, this study is the first step taken in the direction where machine learning techniques can be used for fertility prediction, and subsequent refinements will more likely increase its potential. Future improvements in this model could make it widely usable by the individual, healthcare professionals and the public health systems aimed at enhancing reproductive health outcomes by placing more emphasis on scalability, multiple health parameters monitoring and provision of information in real time.

REFERENCES

[1] P. J. Hansen, "Effect of heat stress on mammalian reproduction," Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 364, no. 1534, pp. 3341-3350, 2009, doi: 10.1098/rstb.2009.0131.

[2] A. Agarwal, A. Aponte-Mellado, B. J. Premkumar, A. Shaman, and S. Gupta, "The effects of oxidative stress on male reproduction: a review," Reproductive Biology and Endocrinology, vol. 10, no. 1, p. 49, 2012, doi: 10.1186/1477-7827-10-49.

[3] B. G. W. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Towards medical XAI," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 11, pp. 4793–4813, 2021, doi: 10.1109/TNNLS.2020.3027314.

[4] S. M. Goyal, M. Mittal, and J. Bansal, "The impact of lifestyle and environment on male fertility," International Journal of Fertility Research, vol. 34, no. 4, pp. 321-329, 2020.

[5] H. J. Meeker, "Exposure to environmental endocrine disruptors and male reproductive health," Asian Journal of Andrology, vol. 14, no. 1, pp. 20-28, 2012, doi: 10.1038/aja.2011.162.

[6] T. I. Koivisto, P. Hägg, J. Vuorinen, and H. M. Söderström, "Using SHAP values to interpret machine learning models in medical diagnostics," Journal of Machine Learning in Medical Imaging, vol. 23, pp. 117–126, 2023.

[7] D. Singh and R. Shukla, "Machine learning applications in predictive healthcare for reproductive medicine," Journal of Artificial Intelligence Research, vol. 45, no. 2, pp. 193–204, 2022.

[8] L. E. Edwards, K. A. O'Leary, J. Miller, and G. Shroff, "Developing interpretable AI models for healthcare: Opportunities and challenges," Journal of Healthcare Informatics Research, vol. 5, no. 4, pp. 601-618, 2021, doi: 10.1007/s41666-021-00107-9.

[9] A. B. Martin and C. Fernandez, "Impact of smoking and alcohol on fertility: A machine learning approach," Computational

Reproductive Biology, vol. 31, pp. 132–146, 2021.

[10] J. K. Doiron and L. G. Ferrar, "Integrating lifestyle modifications with predictive health tools for male fertility improvement," Journal of Reproductive Health Technology, vol. 22, no. 3, pp. 215–226, 2020.

[11] K. W. Zhang, H. C. Tang, and L. Yang, "The role of artificial intelligence in modern healthcare: A comprehensive review of the trends and challenges," Journal of Medical Imaging and Health Informatics, vol. 11, no. 6, pp. 1365-1374, 2021.

[12] P. R. Dutta, A. M. Gopinath, and J. Thomas, "Predicting infertility using machine learning: A comprehensive review," International Journal of Advanced Research in Artificial Intelligence, vol. 9, no. 2, pp. 74–81, 2020.

[13] S. Patel and D. Kumar, "Leveraging AI for reproductive health insights: An overview," Artificial Intelligence in Medicine, vol. 61, pp. 51–60, 2021.