Aykaç *et al.*
Reproductive Biology and Endocrinology (2024) 22:112
https://doi.org/10.1186/s12958-024-01268-w

Reproductive Biology
and Endocrinology

RESEARCH **Open Access**

The prediction of semen quality based on lifestyle behaviours by the machine learning based models

Aykut Aykaç[1*], Coşkun Kaya[1], Özer Çelik[2], Mehmet Erhan Aydın[1] and Mustafa Sungur[1]

Abstract

Purpose To find the machine learning (ML) method that has the highest accuracy in predicting the semen quality of men based on basic questionnaire data about lifestyle behavior.

Methods The medical records of men whose semen was analyzed for any reason were collected. Those who had data about their lifestyle behaviors were included in the study. All semen analyses of the men included were evaluated according to the WHO 2021 guideline. All semen analyses were categorized as normozoospermia, oligozoospermia, teratozoospermia, and asthenozoospermia. The Extra Trees Classifier, Average (AVG) Blender, Light Gradient Boosting Machine (LGBM) Classifier, eXtreme Gradient Boosting (XGB) Classifier, Logistic Regression, and Random Forest Classifier techniques were used as ML algorithms.

Results Seven hundred thirty-four men who met the inclusion criteria and had data about lifestyle behavior were included in the study. 356 men (48.5%) had abnormal semen results, 204 (27.7%) showed the presence of oligozoospermia, 193 (26.2%) asthenozoospermia, and 265 (36.1%) teratozoospermia according to the WHO 2021. The AVG Blender model had the highest accuracy and AUC for predicting normozoospermia and teratozoospermia. The Extra Trees Classifier and Random Forest Classifier models achieved the best performance for predicting oligozoospermia and asthenozoospermia, respectively.

Conclusion The ML models have the potential to predict semen quality based on lifestyles.

Keywords Lifestyle, Machine learning, Male reproductive health, Sperm quality

Introduction

Semen analysis is used to detect the fertility capacity of men in andrological practice and it is the first recommended test for an infertility diagnostic work-up [1]. It is recommended that the semen sample is given close to the laboratory [1] but giving a sample at the hospital could sometimes cause embarrassment and some men may hesitate to provide a sample to learn their fertility capacity. As a result, home semen analysis kits and smartphone-based semen analyzers have started to be used by men [2, 3] to minimize the embarrassment, but these methods have not been used worldwide and their accuracy and cost-effectiveness are have not been fully discussed [2, 3].

---

*Correspondence:
Aykut Aykaç
aykutdr@gmail.com
[1] Department of Urology, Eskisehir City Health Application and Research Center, Health Science University, Eskisehir, Turkey
[2] Department of Mathematics - Computer Science, Faculty of Science, Eskisehir Osmangazi University, Eskisehir, Turkey

According to our knowledge, there have not been any nomograms or predictors using conventional statistical methods to detect semen quality, but it is well known that lifestyle behaviors could affect semen quality [4]. With the increasing use of artificial intelligence in medicine, a limited number of studies have recently been published that aimed to predict semen quality based on lifestyle behaviors using machine learning algorithms [5–16].

In this study, we aimed to find the machine learning (ML) method that has the highest accuracy in predicting the semen quality of men based on basic questionnaire data about lifestyle behavior.

Material-methods

This retrospective-designed study was conducted after ethical approval was obtained (Eskisehir City Hospital, Non-Interventional Clinical Research Ethics Committee; Date: 16/02/2024; Decision Number: ESH/GOEK 2024/77). The medical records of the men whose semen had been analyzed for any reason between August 2021 and January 2023 at the Eskisehir City Hospital were collected. The exclusion criteria were: Aged < 18 or $^{>}$50 years, diagnosis of azoospermia, low semen volume (less than 1.5 mL), abnormal genetics, history of any type of testicular or genitourinary tract or pelvic surgery, recurrent or subclinical varicocele, cryptorchidism, small-sized testis (normal testicular volume is 12.5–19 cc), treated cancer, vascular problem, hematologic illness, systemic disease, genitourinary system infection, or hormonal problems. After the application of the exclusion criteria, the men who also had data about their lifestyle behavior on file were included in the study. This data included details of their Body Mass Index (BMI), smoking and alcohol consumption, coffee intake, physical activity, sauna usage, cell phone usage, and the wearing of tight-fitting underwear as described previously [17, 18]. To ensure strict selection, ex-alcoholics and ex-smokers, passive smokers, and those who only participate in the other lifestyle factors irregularly were also excluded from the study. The lifestyle factors were coded '1' if the BMI was ≥ 25, he smoked every day, drank any amount of alcohol, drank more than 3 cups of coffee a day, did not do any type of exercise regularly, regularly wore Tight underwear, went to a sauna-Turkish Bath regularly, or had a mobile phone ≥ 10 years during the 3-month window before semen collection. If a man who had a BMI of < 25, did not smoke, did not drink alcohol, did not drink more than 3 cups of coffee a day, exercised regularly, did not wear tight underwear, did not go to a sauna, or had used a cell phone < 10 years, the lifestyle factors were coded '0'. After collecting the data about the lifestyle behaviors, all semen analyses of the men included in the study were evaluated according to the WHO 2021 guideline. All semen analyses were categorized as normozoospermia (normal semen characteristic value) or abnormal. If oligozoospermia (sperm concentration $< 16 \times 10^6$/ml of semen) and/or asthenozoospermia (motility < 30% spermatozoa with progressive motility), and/or teratozoospermia (morphologically normal spermatozoa < 4%) [1] had been detected in a semen sample, these results were categorized abnormal. All results were then grouped as normozoospermia, oligozoospermia (sperm concentration $< 16 \times 10^6$/ml of semen), asthenozoospermia (motility < 30% spermatozoa with progressive motility), or teratozoospermia (morphologically normal spermatozoa < 4%). The 4 groups were analyzed separately by statistical methods and the ML algorithms were applied to each group.

The Shapiro–Wilk test was used to test the normality of data distribution. Continuous variables were expressed as mean ± standard deviation, median (minimum–maximum), and categorical variables were expressed as counts (percentages). Comparisons of normally distributed continuous variables between the materials/groups were performed using the student's t-test. Comparisons of non-normally distributed continuous variables between the groups were performed using the Mann–Whitney U Test. Comparisons of categorical variables between the groups were performed using the Yates Chi-Square test and the Monte Carlo Chi-Square test. A two-sided P value < 0.05 was considered statistically significant.

The study was designed according to the principles of ML. The Extra Trees Classifier, Average (AVG) Blender, Light Gradient Boosting Machine (LGBM) Classifier, eXtreme Gradient Boosting (XGB) Classifier, Logistic Regression, and Random Forest Classifier techniques were used as ML algorithms. 70% of the data was used for training and the remaining 30% for testing. In the tests conducted with these models, the model success rates were determined based on accuracy, sensitivity, and specificity values with confusion matrix metrics and the area under curve (AUC) graph in the receiver operating characteristic (ROC) curve analysis. A confusion matrix, which contains information on actual and predicted classifications performed by a classification system and the performance of such systems, is generally assessed using the data in the matrix. Independent variables that significantly affect each group's dependent variable were selected by the permutation feature importance method, which is based on a decrease in the model score when a single variable value is randomly shuffled (1).

Results

Seven hundred thirty-four men who met the inclusion criteria and had data about lifestyle behavior were included in the study. As seen in Table 1, 356 men

Table 1 Data of patients included in the study

	N **= 734**
Age (years) (median, min–max)	30 (16–60)
BMI (n, %)	
≥ 25 (Obesity)	421 (57.4)
Smoking (any amount) (n, %)	373 (50.8)
Alcohol intake (any amount) (n, %)	206 (28.1)
Coffee intake (> 3 cups/day) (n, %)	482 (65.7)
Do not exercise regularly (n, %)	612 (84.6)
Mobile phone ≥ 10 Years (n, %)	170 (23.2)
Tight-fitting underwear (regularly) (n, %)	170 (23.2)
Sauna (regularly) (n, %)	389 (53)
Sperm concentration (10^6/ml) (median, min–max)	35 (0.5–290)
Total sperm number (10^6) (median, min–max)	101.75 (8–867)
Progressive motility sperm rate (A + B) (%) (median, min–max)	45 (5–90)
Morphology (%) (median, min–max)	5 (1–17)
Spermiogram results (n, %)	
Normal	391 (51.5)
Anormal	356 (48.5)
Patients with oligozoospermia (n, %)	204 (27.8)
Patients with asthenozoospermia (n, %)	193 (26.3)
Patients with teratozoospermia (n, %)	265 (63.1)

BMI body mass index

(48.5%) had abnormal semen results, 204 (27.7%) showed the presence of oligozoospermia, 193 (26.2%) asthenozoospermia, and 265 (36.1%) teratozoospermia according to the WHO 2021. Smoking, regularly wearing tight underwear, and regularly going to a sauna / Turkish Bath were statistically significant between having normal and abnormal semen results ($p = 0.001$, 0.003 and 0.038, respectively). While asthenozoospermic males had a statistically significance difference in the age parameter ($p = 0.013$), teratozoospermic males had a statistically significance difference in age, smoking, and alcohol use parameters ($p = 0.025$, 0.001, and 0.034, respectively) (Table 2).

Among the six models, the AVG Blender model had the highest accuracy (61.2%) and AUC (58.4%) for predicting normozoospermia. The Extra Trees Classifier, Random Forest Classifier, and AVG Blender model achieved the best performance for predicting oligozoospermia, asthenozoospermia, and teratozoospermia with an accuracy of 75.5%, 69.6%, and 64.4%, respectively with an AUC of 80%, 74%, and 69.2%, respectively (Table 3 and Fig. 1). Age and smoking were the most significant featured factors for all-predictive models. Table 4 shows the confusion matrices of the algorithms, detailing the number of true positive, false positive, true negative, and false negative cases in the predicted results.

Discussion

Infertility is a major cause of stress for couples because the diagnosis and treatment are often thought to be very complex and the failure to find the cause of infertility makes the situation even worse. For this reason, healthcare professionals have started to use clinical tools to make accurate decisions for diagnosis and/or treatment to minimize the uncertainty for couples. With the increasing use of artificial intelligence in medicine, machine or deep learning-based tools have become increasingly used in reproductive medicine.

A semen analysis is the first laboratory method used for assessing male reproductive health. However, many men feel embarrassed about giving a semen sample, even though it is the most important tool. Attempts have been made to use artificial intelligence-based clinical decision-making tools to at least have an idea of semen quality instead of going to an andrology clinic. The ML-based algorithms used for this purpose have been used with questionnaire-based information about lifestyle behavior to predict semen results [5–16]. The authors of this study chose these factors to develop a prediction model because the separate and cumulative effects of lifestyle on semen quality have long been known [17–19]. GhoshRoy et al. reviewed the studies about lifestyle and environmental factor-based analyses for determining semen quality [20]. In this review, all the classifiers and performance parameters of these algorithms that have been used were summarized.

In our study, we collected the lifestyle behavior data from the questionnaire form that had been prepared previously [17]. All semen analyses were evaluated with the recently published WHO 2021 guideline. All the responses from the questionnaire forms and semen analyses were gathered as a dataset. After balancing and validating this dataset, we tried to find the ML model that could predict male fertility capacity with the highest accuracy and AUC. We found that AVG Blender had the highest performance in predicting normozoospermia and the LGBM Classifier could be used for predicting the oligozoospermic, teratozoospermic, or asthenozoospermic semen analysis results.

There are limited studies on the prediction of semen parameters based on modifiable lifestyle factors by AI methods. However, various non-validated questionnaire forms were used to collect the data. When these forms were analyzed, it could be seen that different parameters were identified as risky and/or modifiable lifestyle behavior without sufficient evidence. The second major limitation of these studies is that WHO guidelines for semen analysis have been updated. The reference limits of semen parameters were revised in

Table 2 Comparison of the data of the patients included in the study by grouping them according to spermiogram results, oligozoospermia, asthenozoospermia, and teratozoospermia conditions

	Spermiogram			**Oligozoospermia**			**Asthenozoospermia**			**Teratozoospermia**		
	Normozoospermia (*n* = 391)	**Anormal (*n* = 356)**	***p***	**No (*n* = 530)**	**Yes (*n* = 204)**	***p***	**No (*n* = 541)**	**Yes (*n* = 193)**	***p***	**No (*n* = 469)**	**Yes (*n* = 265)**	***p***
Age (years) (median. min–max)	30 (16–55)	31 (18–60)	0.408^{a}	30 (16–55)	31 (19–60)	0.383^{a}	30 (16–60)	31 (19–55)	**0.013^{a}**	30 (16–55)	31 (18–60)	**0.025^{a}**
BMI ≥ 25 (n, %)	223 (55.8)	198 (59.3)	0.335^{b}	304 (57.4)	117 (57.4)	0.999^{b}	303 (56)	118 (61.1)	0.216^{b}	262 (55.9)	159 (60)	0.276^{b}
Smoking (any amount) (n, %)	181 (45.3)	192 (57.5)	**0.001^{b}**	263 (49.6)	110 (53.9)	0.297^{b}	269 (49.7)	104 (53.9)	0.321^{b}	217 (46.3)	156 (58.9)	**0.001^{b}**
Alcohol intake (any amount) (n, %)	117 (29.3)	89 (26.6)	0.434^{b}	155 (29.2)	51 (25)	0.251^{b}	160 (29.6)	46 (23.8)	0.128^{b}	144 (30.7)	62 (23.4)	**0.034^{b}**
Coffee intake (> 3 cups/day) (n, %)	262 (65.5)	220 (65.9)	0.917^{b}	352 (66.4)	130 (63.7)	0.492^{b}	351 (64.9)	131 (67.9)	0.452^{b}	311 (66.3)	171 (64.5)	0.625^{b}
Do not exercise regularly (n, %)	331 (82.8)	290 (86.8)	0.128^{b}	446 (84.2)	175 (85.8)	0.583^{b}	452 (83.5)	169 (87.6)	0.184^{b}	391 (83.4)	230 (86.8)	0.217^{b}
Mobile phone ≥ 10 Years (n, %)	352 (88)	289 (86.5)	0.550^{b}	165 (87.7)	176 (86.3)	0.594^{b}	474 (87.6)	167 (86.5)	0.697^{b}	410 (87.4)	231 (87.2)	0.922^{b}
Tight-fitting underwear (regularly) (n, %)	76 (19)	94 (28.1)	**0.003^{b}**	116 (21.9)	54 (26.5)	0.187^{b}	121 (22.4)	49 (25.4)	0.393^{b}	99 (21.1)	71 (26.8)	0.080^{b}
Sauna (regularly) (n, %)	198 (49.5)	191 (57.2)	**0.038^{b}**	273 (51.5)	116 (56.9)	0.193^{b}	288 (53.2)	101 (52.3)	0.829^{b}	241 (51.4)	148 (55.8)	0.245^{b}

BMI Body mass index

[a] Mann Whitney U

[b] Chi-Square

Table 3 Prediction results of machine learning algorithms

Model Name	Anormal Sperm Results		Oligozoospermia		Asthenozoospermia		Teratozoospermia	
	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
XGB Classifier	0.535	0.537	0.752	0.726	0.716	0.659	0.641	0.628
AVG Blender	**0.584**	**0.612**	0.776	0.731	0.705	0.664	**0.692**	**0.644**
Extra Trees Classifier	0.472	0.5	**0.8**	**0.755**	0.727	0.668	0.649	0.617
LGBM Classifier	0.527	0.53	0.723	0.66	0.693	0.668	0.656	0.628
Random Forest Classifier	0.509	0.493	0.797	0.75	**0.74**	**0.696**	0.683	0.649
Logistic Regression	0.586	0.537	0.512	0.491	0.624	0.594	0.597	0.564

AVG Blender Average Blender, *LGBM* Light Gradient Boosting Machine, *XGB* eXtreme Gradient Boosting, *AUC* Area Under Curve

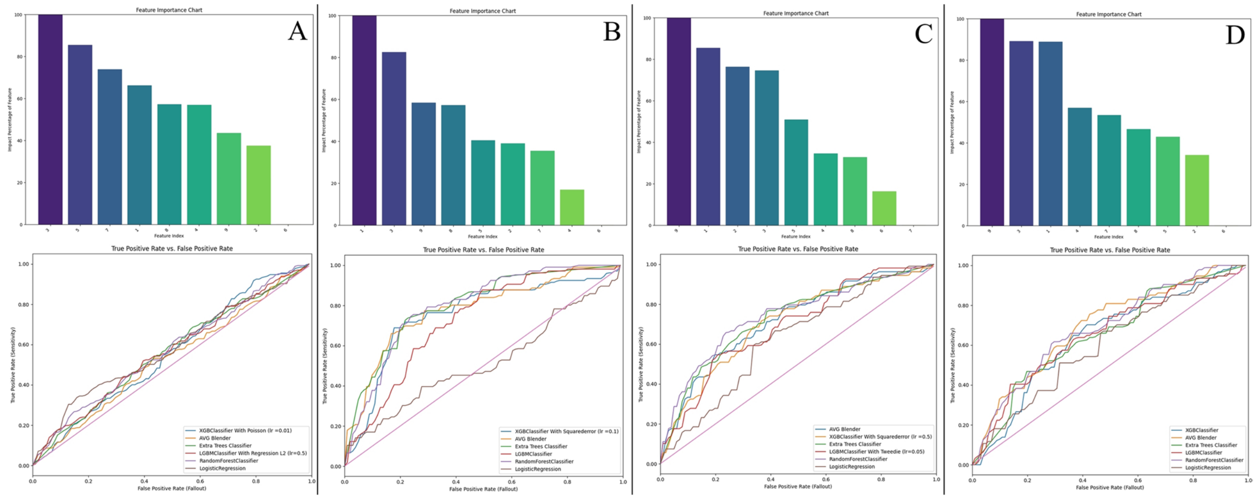

Fig. 1 Impact factors of the variables and the receiver operating characteristic (ROC) curve analysis of the study. **A** AVG Blender of the normozoospermia, **B** Extra Trees Classifier of the oligozoospermia, **C** Random Forest Classifier of the asthenozoospermia, **D** AVG Blender of the teratozoospermia (1: Age, 2: Body Mass Index, 3: Smoking (any amount), 4: Alcohol intake (any amount), 5: Coffee intake (> 3 cups/day), 6: Don't exercise regularly, 7: Mobile phone ≥ 10 Years, 8: Tight-fitting underwear (regularly), 9: Sauna (regularly)) (AVG Blender: Average Blender, LGBM: Light Gradient Boosting Machine)

Table 4 Confusion matrix showing the results of the classification processes

Machine Learning Algorithms	Anormal Sperm Results		Oligozoospermia		Asthenozoospermia		Teratozoospermia	
	AVG Blender		Extra Trees Classifier		Random Forest Classifier		LGBM Classifier	
	Yes	No	Yes	No	Yes	No	Yes	No
Yes	39[a]	28[b]	78[a]	28[b]	72[a]	36[b]	61[a]	33[b]
No	24[c]	43[d]	24[c]	82[d]	30[c]	79[d]	33[c]	61[d]

AVG Blender, Average Blender, *LGBM* Light Gradient Boosting Machine

[a] True positive, [b]False Negative, [c]False positive, [d]True negative

the 6th version of the WHO guidelines, therefore the results of these studies have become invalid and the algorithms recommended by these previous analyses should be reconducted. As with the previous reports in the literature, our outputs will become invalid and the algorithms will need to be re-run if the reference limits of semen parameters are updated by later versions of the WHO guidelines. Another limitation of this current study is that our data was obtained from a single center, and we did not validate the ML models with data from external infertility clinics. Another limitation of the prediction of semen quality based on lifestyle behaviors is that various AI methods have been employed and the reports lack information about the development of the models, the various parameters that have been used to find the highest performance to predict the semen quality, and that the health professionals related to infertility may not have the knowledge to fully understand the process and the results.

Conclusion

The ML models used in this study have the potential to predict semen quality based on lifestyles. Studies with larger training datasets obtained from standardized and validated questionnaire forms about lifestyle behavior should be designed and the AI methods should be developed with a wide range of performance parameters. Furthermore, extensive information should be reported about the construction methods of the models to enable clinicians and couples to easily understand the results.

Authors' contributions

AA and CK wrote the main manuscript text and MEA prepared figures 1. CK and ÖÇ designed of the work. CK and MEA interpretation of data. AA and MS have drafted the work or substantively revised it. All authors reviewed the manuscript.

Funding

The authors did not receive support from any organization for the submitted work. The authors have no relevant financial or non-financial interests to disclose.

Availability of data and materials

No datasets were generated or analysed during the current study.

Declarations

Competing interests

The authors declare no competing interests.

Received: 21 March 2024 Accepted: 24 July 2024

References

1. WHO laboratory manual for the examination and processing of human semen. 6th ed. Geneva: World Health Organization; 2021. https://iris.who.int/bitstream/handle/10665/343208/9789240030787-eng.pdf?sequence=1.
2. Park MJ, Lim MY, Park HJ, Park NC. Accuracy comparison study of new smartphone-based semen analyzer versus laboratory sperm quality analyzer. Investig Clin Urol. 2021;62:672–80. https://doi.org/10.4111/icu.20210266.
3. Lai JD, Fantus RJ, Meza JA, Hudnall MT, Pham M, Brannigan RE, et al. Cost-effectiveness of early screening home semen analysis in couples attempting to conceive. Urology. 2022;170:104–10. https://doi.org/10.1016/j.urology.2022.06.053.
4. Greeson KW, Crow KMS, Edenfield RC, Easley CA. Inheritance of paternal lifestyles and exposures through sperm DNA methylation. Nat Rev Urol. 2023;20:356–70. https://doi.org/10.1038/s41585-022-00708-9.
5. Badura A, Marzec-Wroblewska U, Kaminski P, Lakota P, Ludwikowski G, Szymanski M, et al. Prediction of semen quality using artificial neural network. J Appl Biomed. 2019;17:167–74. https://doi.org/10.32725/jab.2019.015.
6. Huang HH, Hsieh SJ, Chen MS, Jhou MJ, Liu TC, Shen HL, et al. Machine learning predictive models for evaluating risk factors affecting sperm count: predictions based on health screening indicators. J Clin Med. 2023;12:12. https://doi.org/10.3390/jcm12031220.
7. Girela JL, Gil D, Johnsson M, Gomez-Torres MJ, De Juan J. Semen parameters can be predicted from environmental factors and lifestyle using artificial intelligence methods. Biol Reprod. 2013;88:99. https://doi.org/10.1095/biolreprod.112.104653.
8. Zhou M, Yao T, Li J, Hui H, Fan W, Guan Y, et al. Preliminary prediction of semen quality based on modifiable lifestyle factors by using the XGBoost algorithm. Front Med (Lausanne). 2022;9:811890. https://doi.org/10.3389/fmed.2022.811890.
9. GhoshRoy D, Alvi PA, Santosh KC. Unboxing industry-standard AI models for male fertility prediction with SHAP. Healthcare (Basel). 2023;11:11. https://doi.org/10.3390/healthcare11070929.
10. Gil D, Girela JL, De Juan J, Gomez-Torres MJ, Johnsson M. Predicting seminal quality with artificial intelligence methods. Expert Syst Appl. 2012;39:12564–73. https://doi.org/10.1016/j.eswa.2012.05.028.
11. Huang HH, Lu CJ, Jhou MJ, Liu TC, Yang CT, Hsieh SJ, et al. Using a decision tree algorithm predictive model for sperm count assessment and risk factors in health screening population. Risk Manag Healthc Policy. 2023;16:2469–78. https://doi.org/10.2147/RMHP.S433193.
12. Sahoo AJ, Kumar Y. Seminal quality prediction using data mining methods. Technol Health Care. 2014;22:531–45. https://doi.org/10.3233/THC-140816.
13. Simfukwe M, Kunda D, Chembe C. Comparing naive bayes method and artificial neural network for semen quality categorization. Int J Innov Sci Eng Technol. 2015;2:689–94.
14. Dash SR, Ray R. Predicting seminal quality and its dependence on life style factors through ensemble learning. Int J E-Health Med Commun (IJEHMC). 2020;11:78–95.
15. Yibre AM, Koçer B. Semen quality predictive model using feed forwarded neural network trained by learning-based artificial algae algorithm. Eng Sci Technol Int J. 2021;24:310–8.
16. GhoshRoy D, Alvi PA, Santosh K. Explainable AI to Predict Male Fertility Using Extreme Gradient Boosting Algorithm with SMOTE. Electronics. 2023;12:15. https://www.mdpi.com/about/announcements/784. https://www.mdpi.com/2079-9292/12/1/15.
17. Kaya C, Aykac A, Kaya Y, Tas M. The effect of modifiable lifestyle factors on semen quality. Rev Int Androl. 2020;18:151–8. https://doi.org/10.1016/j.androl.2019.09.001.
18. Wogatzky J, Wirleitner B, Stecher A, Vanderzwalmen P, Neyer A, Spitzer D, et al. The combination matters–distinct impact of lifestyle factors on sperm quality: a study on semen analysis of 1683 patients according to MSOME criteria. Reprod Biol Endocrinol. 2012;10:115. https://doi.org/10.1186/1477-7827-10-115.
19. Lyons HE, Gyawali P, Mathews N, Castleton P, Mutuku SM, McPherson NO. The influence of lifestyle and biological factors on semen variability. J Assist Reprod Genet. 2024. https://doi.org/10.1007/s10815-024-03030-y.
20. GhoshRoy D, Alvi PA, Santosh KC. AI tools for assessing human fertility using risk factors: a state-of-the-art review. J Med Syst. 2023;47:91. https://doi.org/10.1007/s10916-023-01983-8.

Publisher's Note