Andrology: Open Access
ISSN: 2167-0250
OPEN ACCESS Freely available online
Research Article

# Impact of Lifestyle and Mental Health on Sperm DNA Fragmentation Rate in Infertile Men and Construction of a Predictive Model: Based on Machine Learning Algorithms

**Ke Wang[1,2], Jing Hu[2], Ningxin Qin[1*], Jie Bai[3], You Zhang[1], Xin Huang[4], Yan Xu[2,3]**

[1]*Department of Maternal-Fetal Medicine and Gynecologic Oncology, Tongji University, Shanghai, China;* [2]*Department of International Medical Technology, Sanda University, Shanghai, China;* [3]*Department of Medicine, Tongji University, Shanghai, China;* [4]*Department of Medicine, Shanghai Jiao Tong University, Shanghai, China*

## ABSTRACT

**Objective:** This study aims to investigate the current status of sperm DNA Fragmentation rate (DFI) in infertile men, analyze the impact of lifestyle and mental health on it, and construct a predictive model for DFI using machine learning algorithms.

**Methods:** A cross-sectional survey was conducted from March, 2023 to August, 2023 involving 592 infertile male patients receiving Assisted Reproductive Technology (ART) treatment at the Reproductive Medicine Center of Tongji University Affiliated Obstetrics and Gynecology Hospital. Patients completed a general information questionnaire, the Self-Rating Anxiety Scale (SAS), the Self-Rating Depression Scale (SDS), and the Athens Insomnia Scale (AIS). DFI data were collected, and binary logistic regression was utilized to analyze independent risk factors affecting DFI quality. Predictive models for DFI quality were constructed using decision tree, random forest, and support vector machine algorithms based on machine learning, with 10-fold cross-validation used to search for optimal parameters. The models were evaluated using Receiver Operating Characteristic (ROC) curves, accuracy, precision, recall and F1 score to identify the best model and rank feature importance

**Results:** Among the surveyed infertile male patients, 434 (73.3%) had poor sperm DNA fragmentation rates. Binary logistic regression indicated that smoking status, alcohol consumption, regular exercise habits, depression, insomnia, and anxiety were independent risk factors for poor DFI ($p<0.05$). All three machine learning algorithms achieved an area under the curve (AUC) greater than 0.80, with the random forest-based predictive model demonstrating the best performance (AUC=0.95), making it the optimal model. Feature importance ranking from the optimal model revealed that insomnia was the most significant factor affecting DFI quality, followed by anxiety, alcohol consumption, depression, smoking and regular exercise.

**Conclusion:** In this study, the random forest predictive model exhibited the best performance and served as the optimal model. The high detection rate of abnormal DFI quality in infertile men is primarily influenced by lifestyle habits, negative emotions and insomnia. The establishment of the predictive model provides a convenient tool for male reproductive health professionals to assist clinical decision-making and adjust treatment plans. Healthcare providers should consider guiding patients to reduce smoking and alcohol intake, increase physical activity and improve negative emotions to enhance DFI quality in infertile men.

**Keywords:** Infertile men; Sperm DNA fragmentation rate; Machine learning; Predictive model; Risk factors

## INTRODUCTION

The incidence of infertility is increasingly becoming a concern among younger individuals. Numerous studies have revealed that the prevalence of male infertility is rising annually, with declining sperm quality identified as one of the direct causes. Surveys indicate that over 40 million individuals in China are affected by infertility, of which approximately 50% can be attributed to male factors [1]. For male patients with infertility, sperm DNA Fragmentation Index (DFI) has emerged as an important diagnostic reference in addition to conventional semen analysis [2,3].

DFI can assist in predicting pregnancy outcomes, evaluating embryo developmental potential, implantation rates, and the risk of genetic

**Correspondence to:** Ningxin Qin, Department of Maternal-Fetal Medicine and Gynecologic Oncology, Tongji University, Shanghai, China, E-mail: ningxinqin@51mch.com

**Received:** 22-Oct-2024, Manuscript No. ANO-24-34742; **Editor assigned:** 24-Oct-2024, PreQC No. ANO-24-34742 (PQ); **Reviewed:** 07-Nov-2024, QC No. ANO-24-34742; **Revised:** 14-Nov-2024, Manuscript No. ANO-24-34742 (R); **Published:** 21-Nov-2024, DOI: 10.35248/2167-0250.24.13.335

**Citation:** Wang K, Hu J, Qin N, Bai J, Zhang Y, Huang X, et al. (2024). Impact of Lifestyle and Mental Health on Sperm DNA Fragmentation Rate in Infertile Men and Construction of a Predictive Model: Based on Machine Learning Algorithms. Andrology. 13:335.

defects, while also serving as a more sensitive marker for predicting female pregnancy outcomes [4,5]. Lifestyle encompasses the activities and behavioral characteristics formed under specific social conditions and guided by value systems that meet individual needs, including aspects such as diet, housing, transportation, and leisure activities [6]. Increasing evidence suggests that lifestyle factors such as smoking, alcohol consumption, and exercise an influence semen quality; psychological stress is also a significant factor in declining sperm quality [7,8]. Currently, the application of artificial intelligence, particularly machine learning algorithms, in disease prediction and diagnosis is expanding within modern medical research. In reproductive medicine, the use of machine learning models is gradually infiltrating key aspects of the assisted reproduction process, including embryo development and implantation rate predictions. However, tools for predicting DFI quality in infertile males remain relatively scarce. Therefore, this study aims to investigate the current status of sperm DFI among infertile males, analyze the impacts of lifestyle and mental health on DFI in this population, and develop predictive models using three machine learning algorithms, evaluating each to identify the optimal model. The construction of these predictive models could provide a more scientific basis for clinical diagnosis and treatment while offering personalized, rational suggestions to improve sperm DFI, ultimately enhancing male reproductive capacity.

## MATERIAL AND METHODS

### Study subjects

A convenience sampling method was used to select male patients undergoing Intracytoplasmic Sperm Injection (ICSI) for assisted reproduction at the Reproductive Medicine Center of Tongji University Affiliated Obstetrics and Gynecology Hospital from March, 2024 to August, 2024.

**Inclusion criteria:** Diagnosis of male infertility according to the standards set forth in the "World Health Organization Laboratory Manual for the Examination and Processing of Human Semen" ($5^{th}$ edition), and acceptance of ICSI treatment; Normal male reproductive system and physical examination, with no history affecting sperm quality; No prior treatments or medications that could influence sperm quality before semen analysis; Voluntary participation in the study [9].

**Exclusion criteria:** Male patients with chromosomal abnormalities; Patients using donor sperm, surgical retrieval, or cryopreserved sperm for assisted reproduction; Patients with severe chronic diseases, tumors, or other significant health issues; Chromosomal or genetic abnormalities detected through pre-implantation genetic testing; Patients refusing to participate in the questionnaire survey.

Following the “Guidelines for the Diagnosis and Treatment of Male Infertility”, a total of 18 statistical variables were included in this study [10]. Based on Kendall's sample size estimation method, the required sample size was calculated as (18 × 10) × (100%+20%)=216, with a final inclusion of 592 cases [11]. This study adhered strictly to the eligibility criteria and complied with all relevant laws, regulations, and ethical principles, having received approval from the Ethics Committee of Tongji University Affiliated Obstetrics and Gynecology Hospital (Ethics Approval Number: KS2313).

### Survey instruments

**General demographic data:** A self-designed questionnaire was developed to collect information on patients' age, Body Mass Index (BMI), occupational status, residency type, educational level, duration of infertility, smoking history (greater than 20 cigarettes per day), alcohol consumption history (40°-68° liquor greater than 100 ml per day), cola consumption habit (greater than 500 ml per day), strong tea consumption habit (greater than 500 ml per day), coffee consumption habit (greater than 500 ml per day), average daily exercise duration, average daily sleep duration, exposure to high-risk work environments (high heat, radiation/radioactive exposure), and sauna usage habit (more than once per week).

**Self-Rating Depression Scale (SDS):** Patients self-evaluated their symptoms over the past week using a scale consisting of 20 items. The total raw score was calculated, and the standard score was derived as standard score=raw score × 1.25. According to Chinese normative data, a standard score of 53 indicates mild depression (53-62), moderate depression (63-72), and severe depression (greater than 72) [12].

**Self-Rating Anxiety Scale (SAS):** Similar to the SDS, patients self-evaluated their symptoms from the past week, with a total raw score calculated from 20 items and a standard score computed as standard score=raw score × 1.25. Based on Chinese normative data, a standard score of 50 indicates mild anxiety (50-59), moderate anxiety (60-69), and severe anxiety (70 and above) [13].

**Athens Insomnia Scale (AIS):** This scale assessed patients' conditions over the past month. The sum of the eight items yields a total score ranging from 0 to 24, where 0-3 indicates no sleep disturbances, 4-6 suggests suspected insomnia, and a total score greater than 6 is classified as insomnia [14].

**DFI measurement:** Sperm DNA Fragmentation Index (DFI) was assessed using sperm chromatin analysis. According to Evenson, a DFI between 0% and 15% indicates high reproductive potential, 16% to 29% indicates moderate reproductive potential, and a DFI of 30% or greater indicates low reproductive potential [15]. Therefore, in this study, a DFI greater than 15% was established as the threshold for assessing normality in male DFI.

### Survey method

After obtaining patient consent, researchers conducted a cross-sectional survey on the same day as patient registration for those who met the inclusion and exclusion criteria. Using standardized instructions, patients were asked to fill out the questionnaire based on their actual circumstances, provided with adequate time and a private environment. After ensuring there were no omissions, the completed questionnaires were collected. A total of 601 questionnaires were distributed, with 592 valid responses received, resulting in a response rate of 98.5%. DFI data were obtained from the semen analysis reports in the hospital's HIS system.

### Statistical methods

Data analysis was performed using SPSS version 26.0. Continuous variables were expressed as mean ± standard deviation ($\bar{x} \pm S$), and intergroup comparisons were made using the t-test. Categorical variables were presented as "n (%)", with $\chi^2$ tests used for comparisons between groups. Binary logistic regression analysis was conducted to identify independent risk factors affecting DFI quality in infertile males, with a p-value of <0.05 considered statistically significant.

### Prediction model construction and evaluation

Model construction and evaluation were conducted using Python 3.6. A stratified sampling method was used to randomly divide the dataset into a training set and a validation set in an 80:20 ratio, implementing 10-fold cross-validation. Significant independent risk factors identified through binary logistic regression were incorporated into three machine learning algorithms Decision Trees (DT), Random Forest (RF), and Support Vector Machines (SVM) to construct prediction models. The area under the Receiver Operating Characteristic (ROC) curve (AUC), accuracy, sensitivity, specificity, and F1 score were calculated to evaluate model stability and predictive performance. Additionally, SHAP analysis was utilized to interpret the model results and assess feature importance.

## RESULTS

### Current status of DFI in infertile males

In this study, DFI data were collected and organized for 592 infertile males. Among these patients, 158 (26.7%) had normal DFI quality, while 434 (73.3%) exhibited poor DFI quality.

### Comparison of clinical data related to DFI status in infertile males

Univariate analysis revealed statistically significant differences in the following factors: Age, residency type, alcohol consumption habits, smoking habits, regular exercise habits, depression status, anxiety status, and sleep quality ($p<0.05$). Detailed results can be found in Table 1.

**Table 1:** Univariate analysis of factors related to Sperm DNA Fragmentation Index in infertile males (n=592).

| Factor | Normal DFI (n=158) | Abnormal DFI (n=434) | $\chi^2$ | p |
|---|---|---|---|---|
| **Age** | | | 4.322 | 0.038* |
| Ages 26 to 40 | 123 (77.85%) | 300 (69.12%) | | |
| Ages 41 to 54 | 35 (22.15%) | 134 (30.88%) | | |
| **BMI** | | | 1.265 | 0.531 |
| <18.5 | 2 (1.27%) | 12 (2.76%) | | |
| 18.5˜24.9 | 98 (62.03%) | 258 (59.45%) | | |
| >24.9 | 58 (36.70%) | 164 (37.79%) | | |
| **Occupational status** | | | 3.247 | 0.072 |
| Non-employed | 18 (11.39%) | 76 (17.51%) | | |
| Employed | 140 (88.61%) | 358 (82.49%) | | |
| **Residency type** | | | 7.947 | 0.005* |
| Urban | 100 (63.29%) | 218 (50.23%) | | |
| Rural | 58 (36.71%) | 216 (49.77%) | | |
| **Educational level** | | | 0.676 | 0.713 |
| High school and below | 48 (30.38%) | 132 (30.41%) | | |
| Associate/Bachelor's degree | 88 (55.70%) | 230 (53.00%) | | |
| Master's degree and above | 22 (13.92%) | 72 (16.59%) | | |
| **Smoking history** | | | 8.043 | 0.005* |
| No | 34 (21.52%) | 146 (33.64%) | | |
| Yes | 124 (78.48%) | 288 (66.36%) | | |
| **Alcohol consumption history** | | | 75.181 | <0.001** |
| No | 68 (43.04%) | 48 (11.06%) | | |
| Yes | 90 (56.96%) | 386 (88.94%) | | |
| **Cola consumption habit** | | | 0.044 | 0.835 |
| No | 48 (30.38%) | 128 (29.49%) | | |
| Yes | 110 (69.62%) | 306 (70.51%) | | |
| **Tea consumption habit** | | | 0.716 | 0.397 |
| No | 58 (36.71%) | 176 (40.55%) | | |
| Yes | 100 (63.29%) | 258 (59.45%) | | |
| **Coffee consumption habit** | | | 1.556 | 0.212 |
| No | 106 (67.09%) | 314 (72.35%) | | |
| Yes | 52 (32.91%) | 120 (27.65%) | | |
| **Duration of infertility** | | | 1.986 | 0.159 |
| ≤5 Years | 120 (75.95%) | 304 (70.05%) | | |
| 5 Years | 38 (%) | 130 (29.95%) | | |
| **Average daily sleep duration (h)** | | | 5.84 | 0.12 |
| <5 | 26 (16.46%) | 70 (16.13%) | | |
| 5-7 | 82 (51.90%) | 232 (53.46%) | | |
| 7-9 | 44 (27.85%) | 128 (29.49%) | | |
| >9 | 6 (3.79%) | 4 (0.92%) | | |
| **Average daily exercise duration (h)** | | | 11.865 | 0.008* |
| Almost none | 50 (31.65%) | 196 (45.16%) | | |
| 0-0.5 | 78 (49.37%) | 182 (41.94%) | | |
| 0.5-1 | 28 (17.72%) | 46 (10.60%) | | |
| >1 | 2 (1.26%) | 10 (2.30%) | | |
| **Exposure to high-risk work environment** | | | 0.126 | 0.085 |

| | | | | |
|---|---|---|---|---|
| No | 144 (91.14%) | 411 (94.70%) | | |
| Yes | 14 (8.86%) | 23 (5.30%) | | |
| **Sauna usage habit** | | | 0.118 | 0.074 |
| No | 141 (89.24%) | 405 (93.32%) | | |
| Yes | 17 (10.76%) | 29 (6.68%) | | |
| **Anxiety status** | | | 131.342 | <0.001** |
| No anxiety | 140 (88.60%) | 154 (35.48%) | | |
| Mild anxiety | 14 (8.86%) | 186 (42.86%) | | |
| Moderate anxiety | 2 (1.27%) | 72 (16.59%) | | |
| Severe anxiety | 2 (1.27%) | 22 (5.07%) | | |
| **Depression status** | | | 90.026 | <0.001** |
| No depression | 108 (68.35%) | 112 (25.81%) | | |
| Mild depression | 42 (26.58%) | 276 (63.59%) | | |
| Moderate depression | 6 (3.80%) | 38 (8.76%) | | |
| Severe depression | 2 (1.27%) | 8 (1.84%) | | |
| **Sleep status** | | | 199.404 | <0.001** |
| No insomnia | 150 (94.94%) | 128 (29.49%) | | |
| Suspected insomnia | 4 (2.53%) | 220 (50.69%) | | |
| insomnia | 4 (2.53%) | 86 (19.82%) | | |

**Note:** p<0.05; p**<0.01

## Binary logistic regression analysis of factors affecting DFI in infertile males

The results of the binary logistic regression analysis indicated that smoking history, alcohol consumption history, average daily exercise duration, depression status, insomnia status, and anxiety status were independent risk factors affecting DFI quality in infertile males. The variable assignments were as follows: Age: 0=26 to 40 years, 1=41 to 54 years; Residency type: 0=Urban, 1=Rural; Smoking history: 0=No, 1=Yes; Alcohol consumption history: 0=No, 1=Yes; Average daily exercise duration: 0=Almost none, 1=0-0.5 h, 2=0.5-1 h, 3 ≥ 1 h; Depression status: 0=No depression, 1=Mild depression, 2=Moderate depression, 3=Severe depression; Anxiety status: 0=No anxiety, 1=Mild anxiety, 2=Moderate anxiety, 3=Severe anxiety; Insomnia status: 0=No insomnia, 1=Suspected insomnia, 2=Insomnia. Detailed results are shown in Table 2.

**Table 2:** Binary logistic regression analysis results of factors affecting DFI in infertile males (n=592).

| Variable | B | S.E. | Wald | p | OR | 95%CI | |
|---|---|---|---|---|---|---|---|
| | | | | | | Lower limit | Upper limit |
| Constant | 0.818 | 0.447 | 3.352 | 0.067 | 2.266 | | |
| Age (with 26-40 years as reference) | -0.487 | 0.327 | 2.211 | 0.137 | 0.615 | 0.323 | 1.168 |
| Residency type (with Urban as reference) | -0.097 | 0.295 | 0.108 | 0.743 | 0.908 | 0.509 | 1.619 |
| Smoking history (with no as reference) | 1.822 | 0.382 | 22.733 | <0.001** | 6.187 | 2.925 | 13.086 |
| Alcohol consumption history (with no as reference) | -2.083 | 0.396 | 27.7 | <0.001** | 0.125 | 0.057 | 0.271 |
| Average daily exercise duration (with almost none as reference) | | | 17.097 | <0.001** | | | |
| 0-0.5 h | 0.698 | 0.313 | 4.98 | 0.026* | 2.011 | 1.089 | 3.713 |
| 0.5-1 h | 2.2 | 0.543 | 16.443 | <0.001** | 9.029 | 3.117 | 26.156 |
| >1 h | 0.561 | 0.962 | 0.34 | 0.56 | 1.753 | 0.266 | 11.556 |
| Depression status (with no depression as reference) | | | 18.183 | <0.001** | | | |
| Mild depression | -1.198 | 0.333 | 12.981 | <0.001** | 0.302 | 0.157 | 0.579 |
| Moderate depression | 0.899 | 0.884 | 1.034 | 0.309 | 2.458 | 0.434 | 13.911 |
| Severe depression | 1.935 | 1.421 | 1.855 | 0.173 | 6.926 | 0.427 | 112.231 |
| Insomnia status (with no insomnia as reference) | | | 46.82 | <0.001** | | | |
| Suspected insomnia | -4.035 | 0.617 | 42.759 | <0.001** | 0.018 | 0.005 | 0.059 |
| Insomnia | -2.178 | 0.81 | 7.234 | 0.007* | 0.113 | 0.023 | 0.554 |
| Anxiety status (with no anxiety as reference) | | | 25.062 | <0.001** | | | |
| Mild anxiety | -1.8 | 0.383 | 22.147 | <0.001** | 0.165 | 0.078 | 0.35 |
| Moderate anxiety | -1.808 | 1.157 | 2.442 | 0.118 | 0.164 | 0.017 | 1.584 |
| Severe anxiety | -2.528 | 1.22 | 4.293 | 0.038* | 0.08 | 0.007 | 0.872 |

**Note:** p*<0.05; p**<0.001; $R^2$=0.672

## Construction and evaluation of the prediction model

**Model construction:** All study subjects were randomly assigned into a training set (n=474) and a testing set (n=118) in an 80:20 ratio. Based on the results of the six variables identified through binary logistic regression, prediction models were constructed using three machine learning algorithms: DT, RF, and SVM. The training involved 10-fold cross-validation, and the performance of the training set was evaluated using the confusion matrix to calculate accuracy, sensitivity, specificity, AUC, and F1 score. The results indicated that the Random Forest model exhibited superior predictive performance. Refer to Figure1 and Table 3 for details.

**Model evaluation:** The trained models were applied to the testing set, resulting in AUC values of 0.95 for RF, 0.94 for SVM, and 0.94 for DT, as shown in Figure 2. The overall performance of each model is detailed in Table 4. Although the RF model did not perform as well as SVM in terms of recall, it demonstrated superior accuracy and precision compared to SVM. Therefore, based on comprehensive performance, the RF prediction model exhibited the best overall efficacy.

**Model interpretation:** The SHAP (SHapley Additive exPlanations) plots illustrate the importance ranking of risk factors affecting DFI quality in infertile males, with the factors ordered as follows: Insomnia, anxiety, alcohol consumption, depression, smoking and regular exercise. In the SHAP plots, deeper red indicates higher risk, while deeper blue indicates lower risk, presents the SHAP values for each feature, ranked from highest to lowest shown in Figures 3 and 4.

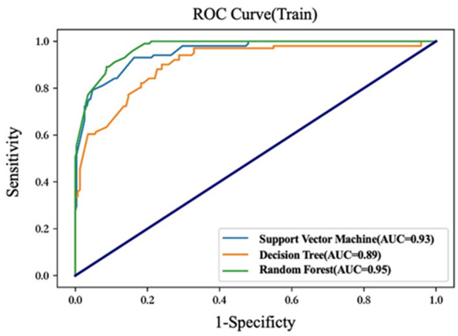


**Figure 1:** ROC curves of three models on the training set.

**Table 3:** Comparative performance of three models on the training set.

| | AUC | | Accuracy | | Precision | | Recall | | F1 Score | |
|---|---|---|---|---|---|---|---|---|---|---|
| Model | Mean | 95%CI | Mean | 95%CI | Mean | 95%CI | Mean | 95%CI | Mean | 95%CI |
| DT | 0.89 | 0.85-0.94 | 0.85 | 0.83-0.87 | 0.78 | 0.71-0.85 | 0.54 | 0.44-0.64 | 0.62 | 0.55-0.70 |
| RF | 0.95 | 0.93-0.96 | 0.89 | 0.87-0.91 | 0.85 | 0.81-0.90 | 0.7 | 0.61-0.78 | 0.76 | 0.70-0.82 |
| SVM | 0.93 | 0.90-0.96 | 0.9 | 0.87-0.92 | 0.83 | 0.80-0.86 | 0.74 | 0.64-0.83 | 0.77 | 0.71-0.83 |

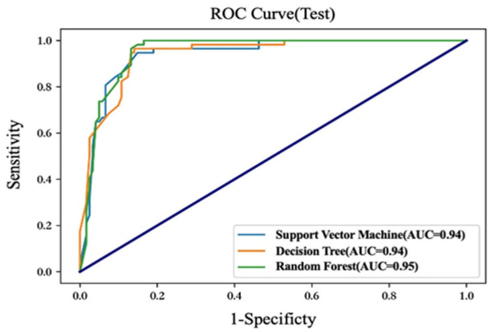


**Figure 2:** ROC curves of three models on the testing set.

**Table 4:** Comparative performance of three models on the testing set.

| | AUC | | Accuracy | | Precision | | Recall | | F1 Score | |
|---|---|---|---|---|---|---|---|---|---|---|
| Model | Mean | 95%CI | Mean | 95%CI | Mean | 95%CI | Mean | 95%CI | Mean | 95%CI |
| DT | 0.94 | 0.92-0.95 | 0.85 | 0.83-0.87 | 0.76 | 0.71-0.85 | 0.58 | 0.48-0.69 | 0.62 | 0.55-0.70 |
| RF | 0.95 | 0.94-0.96 | 0.9 | 0.88-0.92 | 0.85 | 0.80-0.91 | 0.72 | 0.64-0.78 | 0.77 | 0.71-0.83 |
| SVM | 0.94 | 0.93-0.96 | 0.89 | 0.87-0.91 | 0.83 | 0.82-0.84 | 0.73 | 0.64-0.83 | 0.77 | 0.71-0.83 |

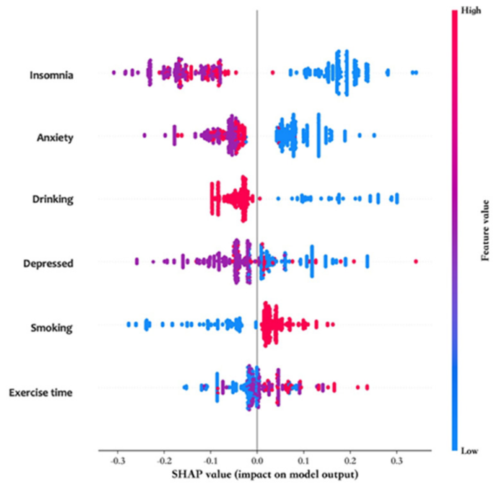


**Figure 3:** SHAP plot interpretation of the RF model.

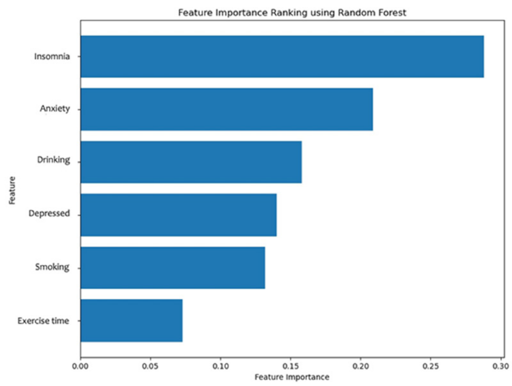


**Figure 4:** Variable importance ranking.

# DISCUSSION

DFI is an important tool for assessing sperm quality, directly reflecting the integrity of sperm DNA strands and serving as a critical indicator of sperm quality. Abnormal DFI can negatively impact male fertility, primarily linked to decreased sperm and embryo developmental capabilities [16]. Increasing evidence suggests that DFI has better predictive value for embryo development and pregnancy outcomes. DNA damage in sperm results from a combination of factors; in addition to irreversible factors such as urinary tract infections, testicular injuries and histories of radiotherapy or chemotherapy, it is also related to environmental, dietary, and stress-related lifestyle factors that can vary in their impact on sperm DNA integrity. In this study, the detection rate of abnormal DFI was 73.3%, slightly higher than previous studies [17]. This may be attributed to the focus on male infertility patients in our research, which allowed for clearer inclusion criteria and a more concentrated population, leading to a higher detection rate of low-quality DFI.

In recent years, machine learning has emerged as a significant data analysis method, proving effective in enhancing disease diagnosis and prediction, with increasing applications in the medical field [18]. This study used logistic regression to analyze variables, identifying six independent risk factors affecting DFI quality in infertile men. We constructed predictive models using DT, RF, and SVM algorithms, with the RF model demonstrating superior performance. Calibration curves indicated good consistency, and the RF algorithm predicts outcomes effectively by averaging predictions across multiple decision trees, which mitigates overfitting and enhances model accuracy and robustness. Additionally, the RF model evaluates the importance of each feature in predicting outcomes, providing clinical practitioners with a valuable tool for identifying DFI quality risks in infertile men.

Using the SHAP method, we ranked the importance of risk factors affecting DFI quality, identifying insomnia, anxiety, alcohol consumption, depression, smoking and exercise frequency as significant contributors. Poor lifestyle choices primarily increase oxidative stress levels within the body, generating free radicals that damage sperm cell membranes and DNA, resulting in decreased DFI quality. Furthermore, as age increases, the cumulative harmful effects of unhealthy lifestyle habits may further deteriorate DFI [19]. A prospective cohort study found that smoking negatively affects all semen parameters in infertile men, particularly DFI, potentially due to toxic substances in cigarettes impacting male reproductive cells [20]. Reports indicate that smoking adversely affects eight nicotinic acetylcholine receptor (n Ach R) subunits in human sperm, leading to semen damage [21]. Excessive reactive oxygen species can cause oxidative stress damage. Long-term heavy smoking increases reactive oxygen levels in seminal plasma, reducing sperm motility and increasing mortality, ultimately impacting sperm capacitation, fertilization capacity, and DFI [22]. Chronic heavy alcohol consumption can lead to gonadal toxicity and degeneration, which may cause testicular atrophy and adversely affect male fertility, as well as abnormalities in testosterone and gonadotropin levels, negatively impacting sperm activation, differentiation, and development, thus affecting DFI quality [23,24]. Additionally, physical activity correlates with DFI quality; in this study, infertile men who exercised between 0 to 1 h daily exhibited better DFI quality than those with minimal exercise. Research shows that long-term aerobic exercise (>12 weeks) can enhance the expression of Transient Receptor Potential Ankyrin 1 (TRPA1) in the epididymal epithelium of diet-induced obese rats, stimulating $Ca_2^+$ influx and promoting the epithelial secretion of $Cl^-$ and $HCO_3^-$, which creates a necessary microenvironment for sperm maturation, thereby improving sperm quality [25]. However, prolonged high-intensity exercise or sedentary behavior may diminish sperm quality and vitality [26]. Psychological issues can also disrupt the endocrine system, leading to hormonal imbalances and neurotransmitter metabolism irregularities, which can cause dysfunction in the hypothalamic-pituitary-adrenal and hypothalamic-pituitary-gonadal axes, potentially leading to sexual and reproductive dysfunction, thereby decreasing DFI quality [27]. Furthermore, sleep is critical for human reproductive health; insomnia affects immune, endocrine, cardiovascular, and nervous system functions, and can also impair male reproductive capabilities [28]. Irregular sleep patterns or insomnia are associated with detrimental effects on reproductive health, inhibiting testosterone secretion and adversely impacting semen quality, resulting in increased DFI [29].

## CONCLUSION

This study utilized machine learning algorithms to identify factors influencing poor DFI quality in male infertility patients, specifically insomnia, anxiety, alcohol consumption, depression, smoking, and exercise habits. Among the three predictive models constructed, the RF model demonstrated strong performance, validated with good predictive efficacy and robust interpretability. Clinically, it can serve as a valuable tool for assisting healthcare providers in assessments. Healthcare professionals may consider developing corresponding intervention strategies aimed at reducing smoking and alcohol consumption, increasing physical activity, and improving negative emotional states to enhance DFI quality in infertile male patients.

## LIMITATIONS

This study has limitations: It did not include external validation, which may hinder generalizability to other hospitals; it only included male infertility patients receiving Intracytoplasmic Sperm Injection (ICSI) treatment, excluding those undergoing Artificial Insemination by Husband (AIH) or *In Vitro* Fertilization (IVF), which may introduce selection bias; and it did not consider other potential factors affecting DFI quality, such as diet and air pollution. We will continue to conduct related investigations in future studies.

## FUNDING

National Natural Science Foundation of China Youth Project (82201882); Shanghai Natural Science Foundation Project (23ZR1450200); Project of Obstetrics and Gynecology Hospital Affiliated to Tongji University (2023HL20).

## REFERENCES

1. Bochen P, Yingpu S, Haixiang S, Cheng-yan D, Ping L, Canquan Z, et al. Expert consensus on the etiology and clinical diagnosis and treatment of asthenozoospermia. J Reprod Med. 2023;32:157-169.
2. Liu M, Liu P, Chang Y, Xu B, Wang N, Qin L, et al. Genome-wide DNA methylation profiles and small noncoding RNA signatures in sperm with a high DNA fragmentation index. J Assist Reprod Genet. 2022;39(10):2255-2274.
3. Xue Z, Yanfen Z, Guohai S. Study on the correlation between sperm DNA fragmentation index and routine semen parameters in males. Chin J Sex Sci. 2021;30(3):6-10.
4. Esteves SC, Zini A, Coward RM, Evenson DP, Gosalvez J, Lewis SE, et al. Sperm DNA fragmentation testing: Summary evidence and clinical practice recommendations. Andrologia. 2021;53(2):e13874.
5. Nicopoullos J, Vicens-Morton A, Lewis SE, Lee K, Larsen P, Ramsay J, et al. Novel use of COMET parameters of sperm DNA damage may increase its utility to diagnose male infertility and predict live births following both IVF and ICSI. Hum Reprod. 2019;34(10):1915-1923.
6. Keyes D, Patel NI, Correa KA. Overview of lifestyle medicine. Chin J Health Manag. 2022;16(8):588-592.
7. Szabo A, Vancsa S, Hegyi P, Varadi A, Forintos A, Filipov T, et al. Lifestyle-, environmental-, and additional health factors associated with an increased sperm DNA fragmentation: A systematic review and meta-analysis. Reprod Biol Endocrinol. 2023;21(1):5.
8. Bhongade MB, Prasad S, Jiloha RC, Ray PC, Mohapatra S, Koner BC. Effect of psychological stress on fertility hormones and seminal quality in male partners of infertile couples. Androl. 2015;47(3):336-342.
9. Organizzazione mondiale della sanita. WHO laboratory manual for the examination and processing of human semen. World Health Organization. 2010:5-97.
10. Cheng-liang X, Zheng L, Xue-jun S, Feng P. Consensus of Chinese experts on the assessment of male fertility. Chin J Androl. 2022;28(09):848-858.
11. Chen Y, Du H, Wei BH, Chang XN, Dong CM. Development and validation of risk-stratification delirium prediction model for critically ill patients: A prospective, observational, single-center study. Med. 2017;96(29):e7543.
12. Zung WW. A self-rating depression scale. Arch Gen Psychiatry. 1965;12(1):63-70.
13. Zung WW. A rating instrument for anxiety disorders. Psychosomatics: J Consult Liaison Psychiatry. 1971.
14. Soldatos CR, Dikeos DG, Paparrigopoulos TJ. Athens Insomnia Scale: Validation of an instrument based on ICD-10 criteria. J Psychosom Res. 2000;48(6):555-560.
15. Evenson D, Jost L. Sperm chromatin structure assay is useful for fertility assessment. Methods Cell Sci. 2000;22:169-189.
16. Zaren P, Alson S, Henic E, Bungum M, Giwercman A. Interaction between serum levels of Anti-Mullerian Hormone and the degree of sperm DNA fragmentation measured by sperm chromatin structure assay can be a predictor for the outcome of standard in vitro fertilization. Plos one. 2019;14(8):e0220909.
17. Yan Y, Song C, Shi S, Qin H, Hao N, Wang L. Comparative Analysis of Semen Quality and DFI in 2061 Infertile Men of Different Nationalities. Adv Clin Med. 2024;14(3):772-779.
18. Chen M, Qian D, Wang Y, An J, Meng K, Xu S, et al. Systematic review of machine learning applied to the secondary prevention of ischemic stroke. J Med Syst. 2024;48(1):8.
19. Nowicka-Bauer K, Nixon B. Molecular changes induced by oxidative stress that impair human sperm motility. Antioxid. 2020;9(2):134.
20. Yingjie YA, Jinfeng CA, Jianghou HO, Yunyan CH, Ming XI, Haiyun YA, et al. The Effect of Smoking on the Semen Quality in Male Infertile. Journal of Kunming Medical University. 2024;45(1):163-167.
21. Condorelli RA, La Vignera S, Duca Y, Zanghi GN, Calogero AE. Nicotine receptors as a possible marker for smoking-related sperm damage. Protein Pept Lett. 2018;25(5):451-454.
22. Agarwal A, Sharma RK, Nallella KP, Thomas Jr AJ, Alvarez JG, Sikka SC. Reactive oxygen species as an independent marker of male factor infertility. Fertil Steril. 2006;86(4):878-885.
23. Ramírez N, Estofan G, Tissera A, Molina R, Luque EM, Torres PJ, et al. Do aging, drinking, and having unhealthy weight have a synergistic impact on semen quality?. J Assist Reprod Genet. 2021;38:2985-2994.
24. Zhao Ruqing. Investigation of semen quality and related factors in fertile males in Panyu District, Guangzhou. Smart Health. 2022;8(29):17-20.
25. Gao DD, Ding N, Deng WJ, Li PL, Chen YL, Guo LM, et al. Aerobic exercises regulate the epididymal anion homeostasis of high-fat diet-induced obese rats through TRPA1-mediated $Cl^-$ and $HCO_3^-$ secretion. Biol Reprod. 2023;109(1):53-64.

26. Rong H, Lihong Z, Jingjing Z. Research progress on the impact of exercise on male sperm quality. Chin J Androl. 2024;38(04):128-132.

27. Yufeng M, Chaoxu L, Mei Q. Effects of male anxiety and stress on methylation of sperm imprinting genes. Chin J School Doctor. 2018;32(8):589-595.

28. Jensen TK, Andersson AM, Skakkebaek NE, Joensen UN, Jensen MB, Lassen TH, et al. Association of sleep disturbances with reduced semen quality: A cross-sectional study among 953 healthy young Danish men. Am J Epidemiol. 2013;177(10):1027-1037.

29. Qing C, Xianghong L, Shengyu F, Jinyan Z, Xiaoning L, Xiang X. Effect of sleep time on sperm routine parameters, sperm morphology and sperm integrity. Pract Prev Med. 2017;24(6):650-661