ARTICLE OPEN

Check for updates

Predicting sperm retrieval success in salvage testicular sperm extraction: a machine learning perspective

Berk Hazir[1], Ali Sahin[2], Mehmet Hamza Gultekin[3], Cevahir Ozer[4], Mehmet Vehbi Kayra[4], Eray Hasirci[5], Erman Ceyhan[5], Mesut Altan[6], Tufan Cicek[1], Yalcin Kizilkan[7], Mesut Berkan Duran[8], Muhammed Arif İbis[9], Gokhan Ceker[10], Cem Sah[11], Nebil Akdogan[12], İyimser Ure[13], Ali Can Albaz[14], Aykut Baser[15] and Murat Gül[2,16]✉

© The Author(s) 2026

This multicenter retrospective observational study aimed to identify predictive factors for successful secondary testicular sperm extraction (TESE) using advanced machine learning (ML) models. Data from 503 infertile men who underwent secondary TESE between 2021 and 2024 after a previous unsuccessful attempt were analyzed. Preoperative characteristics and laboratory findings were assessed using ten ML algorithms, including Xtreme Gradient Boosting, Random Forest, Gradient Boosting, Decision Tree, AdaBoost, Logistic Regression, Multi-Layer Perceptron, Support Vector Machine, K-Nearest Neighbors, and Naive Bayes. Model performance was evaluated using accuracy, sensitivity, specificity, predictive values, F1 score, Youden Index, and area under the ROC curve. Sperm retrieval succeeded in 211 patients (41.9%) and failed in 292 (58.1%). Mean infertility duration was 7.25 years. Xtreme Gradient Boosting consistently outperformed other algorithms across all performance metrics. Key predictors of successful secondary TESE included preoperative body mass index, TESE location, luteinizing hormone level, and semen volume. Testicular volumes and infertility duration also contributed significantly to prediction accuracy. Incorporating multiple clinical and laboratory parameters into ML-based predictive models can improve surgical planning and preoperative counseling. For men undergoing secondary TESE, these models may identify those with a very low chance of success, reducing unnecessary repeat surgical interventions.

IJIR: Your Sexual Medicine Journal; https://doi.org/10.1038/s41443-026-01291-y

INTRODUCTION

Non-obstructive azoospermia (NOA) is characterized by ejaculate volume in a normal range, whereas the absence of sperm at the semen analysis following centrifugation [1]. Approximately 1% of all men and 10% of all infertile men are presented with this clinical condition [2], accompanying the severe deficit of spermatogenesis mainly resulting from primary testicular dysfunction or hypothalamus-pituitary-gonadal (HPG) axis [3].

The standard surgical approach for achieving successful sperm retrieval is testicular sperm extraction (TESE) [4, 5]. With advances in knowledge and surgical techniques, microsurgical testicular sperm extraction (mTESE) has emerged as the more effective procedure, offering higher sperm retrieval rates and fewer complications by enabling surgeons to explore the entire testicular parenchyma and identify dilated tubules within the tissue [4, 6, 7]. Still, sperm can be successfully retrieved in up to 50% of patients during the initial mTESE procedure, enabling their use in intracytoplasmic sperm injection (ICSI) [8].

Salvage TESE procedures are frequently required following an unsuccessful TESE in patients seeking biological paternity. An initial TESE procedure may lead to testicular tissue injury and compromise the blood supply of the tunica albuginea, with potential consequences such as hematoma formation, fibrosis, or even testicular atrophy [9, 10]. As a result, the success rates of subsequent procedures may be slightly reduced [11]. Reported sperm retrieval rates (SRRs) following secondary TESE range from 12.5% to 74.7% [12–17]. Since secondary surgery may impose substantial economic and psychological burdens on couples desiring fertility, careful consideration is warranted when establishing the indication for salvage TESE. Additionally, knowledge of factors that increase or decrease the likelihood of a positive outcome following TESE may help physicians and patients determine the optimal timing for TESE.

The actual international guidelines from the European Association of Urology and American Association of Urology recommend microdissection TESE for men with non-obstructive azoospermia and emphasize individualized counseling regarding expected

[1]Department of Urology, Etlik City Hospital, Ministry of Health, Ankara, Türkiye. [2]Department of Urology, Selcuk University School of Medicine, Konya, Türkiye. [3]Department of Urology, Cerrahpasa Faculty of Medicine, Istanbul University, İstanbul, Türkiye. [4]Department of Urology, Baskent University Faculty of Medicine, Adana, Türkiye. [5]Department of Urology, Baskent University Faculty of Medicine, Ankara, Türkiye. [6]Department of Urology, Hacettepe University Faculty of Medicine, Ankara, Türkiye. [7]Department of Urology, Ankara City Hospital, Ministry of Health, Ankara, Türkiye. [8]Department of Urology, Pamukkale University Faculty of Medicine, Ankara, Türkiye. [9]Department of Urology, Ankara University Faculty of Medicine, Ankara, Türkiye. [10]Department of Urology, Basaksehir Cam and Sakura City Hospital, Istanbul, Türkiye. [11]Department of Urology, Adana Medline Hospital, Adana, Türkiye. [12]Department of Urology, Faculty of Medicine, Cukurova University, Adana, Türkiye. [13]Department of Urology, Faculty of Medicine, Osmangazi University, Eskisehir, Türkiye. [14]Department of Urology, Manisa Celal Bayar University Faculty of Medicine, Manisa, Türkiye. [15]Department of Urology, Medical Faculty, Bandırma Onyedi Eylul University, Balikesir, Türkiye. [16]Department of Andrology, Selçuk University School of Medicine, Konya, Türkiye. ✉email: murat.gul@selcuk.edu.tr

Received: 10 November 2025 Revised: 25 April 2026 Accepted: 7 May 2026
Published online: 20 May 2026

outcomes. However, no validated predictive tools are currently incorporated into these recommendations [3, 18].

Machine learning (ML) methods have become valuable for predictive modeling through the integration of large datasets, enabled by advances in computing power [19]. Several studies have applied machine learning to predict the likelihood of successful sperm retrieval in TESE surgeries [19, 20]. However, the rationale for predicting salvage TESE success differs from that of primary TESE. A prior unsuccessful surgical intervention may significantly alter the testicular microenvironment, as well as clinical and hormonal parameters, making secondary TESE a particularly complex prediction scenario [21]. Machine learning (ML) algorithms are better suited to handle such complex and non-linear relationships and may provide improved predictive performance and more accurate patient-specific risk stratification [22]. This, in turn, may support shared decision-making in this high-stakes clinical setting.

In this multi-institutional study, we aimed to investigate the utility of a machine learning (ML) model to predict the likelihood of successful or unsuccessful sperm retrieval in NOA patients undergoing secondary TESE. We hypothesized that ML models would provide more accurate predictions of secondary TESE outcomes after evaluating the clinical and demographic characteristics of the study cohort.

METHODS

Subjects

Between 2021 and 2024, a total of 503 men from couples suffering from infertility with a history of one unsuccessful TESE underwent a secondary TESE procedure. This multicenter dataset compiled by the Andrology Working Group of the Society of Urological Surgery (SUST) [23] in Turkey was retrospectively reviewed. The inclusion criterion was undergoing a second TESE following a previously unsuccessful attempt for NOA.

Exclusion criteria included detectable sperm in semen analysis; obstructive azoospermia (normal testicular histology); use of non-TESE retrieval methods; more than two prior TESE procedures; complete AZFa or AZFb deletions; and a history of undocumented fertility treatments. Patients with partial Y-chromosome microdeletions were not excluded.

This study was approved by the Ethics Committee of Selçuk University School of Medicine (Approval No: 2024/379). Written informed consent was obtained from all participants after they received detailed written and oral information about the study.

Study outline

Fasting blood samples were collected from all subjects between 8:00 a.m. and 10:00 a.m. for the analysis of metabolic markers and reproductive parameters. The diagnosis of NOA was established based on at least two semen analyses, in accordance with the criteria outlined in the sixth edition of the World Health Organization (WHO) manual [24]. Testicular volume was measured in milliliters and categorized by Prader orchidometry as normal (> 15 mL), hypoplastic (5–15 mL), or atrophic (< 5 mL). Varicocele was graded using the Dubin and Amelar system [25]. TESE was performed using conventional or microsurgical techniques [26], starting with one testis and proceeding to the other only if sperm retrieval failed. Patient data, preoperative labs, and second TESE outcomes were recorded using a standardized, pre-study template to ensure consistent reporting. TESE location was included as a procedural variable reflecting whether sperm retrieval was completed unilaterally or required bilateral exploration during the secondary intervention, and was categorized as unilateral right, unilateral left, or bilateral.

Histopathological findings from primary TESE were recorded according to descriptive pathological reports, including hypospermatogenesis, maturation arrest, Sertoli cell-only syndrome, Leydig cell hyperplasia, and focal/preserved spermatogenesis. Johnsen scoring [27] was not available due to the retrospective and multicenter nature of the dataset.

Focal/preserved spermatogenesis referred to seminiferous tubules demonstrating active germ cell maturation within an overall non-obstructive testicular environment and did not indicate normal global spermatogenesis as seen in obstructive azoospermia.

Development of machine learning models for predicting the success of secondary TESE procedures

Since the outcomes were defined as either "successful" or "unsuccessful," binary machine learning models were developed. The dataset was randomly split into 80% training and 20% testing subsets. Ten separate algorithms were employed for model development, namely Xtreme Gradient Boosting (XGB), Random Forest (RF), Gradient Boosting (GB), Decision Tree (DT), AdaBoost (AB), Logistic Regression (LR), Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (k-NN), and Naive Bayes (NB).

This study was conducted in accordance with the TRIPOD + AI and STREAM-URO guidelines [28, 29]. All algorithms were used with their default Scikit-learn configurations. Model performance was evaluated using a repeated stratified five-fold cross-validation, repeated 5 times, to ensure balanced class distribution and stable performance estimation. Final predictive performance was summarized across cross-validation folds and repetitions.

Evaluation of performance metrics for machine learning models

To comprehensively evaluate model performance, we calculated overall accuracy, sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV), F1 score, Youden Index and the area under the ROC curve. ROC curves were generated for each model based on the binary classification of TESE success versus failure. These metrics were computed for every cross-validation fold and repetition, and the mean and standard deviation were reported to ensure a robust evaluation.

SHapley additive explanations

Following the identification of the best-performing model for predicting secondary TESE success, a SHAP (SHapley Additive exPlanations) analysis was performed to identify the most influential features and to characterize the direction and magnitude of their contributions.

Statistical analysis

Statistical analyses were conducted using Python version 3.11.5, with the pandas (version 2.1.1) and SciPy (version 1.11.2) libraries. Numerical variables were presented as means, standard deviations, and confidence intervals, while categorical variables were summarized as frequencies and percentages. The Kolmogorov–Smirnov test was used to assess the normality of continuous variables. For normally distributed variables, between-group comparisons were performed using the independent samples t-test, whereas the Mann–Whitney U test was applied for non-normally distributed data. Missing data were assessed for all variables prior to analysis.

Missing data were assessed for all variables prior to analysis. The overall proportion of missing values was below 5%, with no evidence of a systematic pattern [30]. Therefore, a complete-case analysis was performed, and no imputation procedures were applied. Statistical significance was defined as $p < 0.05$.

RESULTS

Study population

A total of 503 patients who underwent a secondary TESE procedure were evaluated and categorized into two groups. The first group, consisting of patients with successful sperm retrieval, included 211 individuals (41.95%), while the second group, with unsuccessful retrieval outcomes, comprised 292 individuals (58.05%).

Patient characteristics

Table 1 presents the baseline characteristics of the study population. The mean age was 35.65 ± 6.28 years, with no statistically significant difference between patients with positive and negative sperm retrieval outcomes following the secondary TESE procedure.

Hormonal and semen parameters. Patients with positive sperm retrieval outcomes had significantly lower body mass index (BMI) ($p < 0.001$), semen volume ($p < 0.001$), and luteinizing hormone

Table 1. Baseline characteristics of the study population and differences in features according to sperm retrieval outcomes following secondary TESE procedures.

Features	Overall	Negative Retrieval	Positive Retrieval	*p* value
Number	503 (100)	211 (41.95)	292 (58.05)	–
Age	35.65 ± 6.28 [35.1–36.21]	35.26 ± 6.34 [34.39–36.12]	35.94 ± 6.23 [35.22–36.66]	0.111$^{\alpha}$
Infertility Duration	7.25 ± 4.83 [6.82–7.67]	7.52 ± 4.9 [6.85–8.18]	7.05 ± 4.78 [6.5–7.6]	0.217$^{\alpha}$
Klinefelter Syndrome	64 (12.72)	38 (7.55)	26 (5.17)	**0.002**
Y Chromosome Microdeletion	9 (1.79)	5 (0.99)	4 (0.8)	0.694
Y Chromosome Microdeletion Type				
No Microdeletion	494 (98.21)	206 (40.95)	288 (57.26)	0.694
AZFb	2 (0.4)	1 (0.2)	1 (0.2)	
AZFc	7 (1.39)	4 (0.8)	3 (0.6)	
Varicocelectomy Before Primary TESE	102 (20.28)	41 (8.15)	61 (12.13)	0.688
Cryptorchidism	38 (7.55)	14 (2.78)	24 (4.77)	0.507
Previous Scrotal Surgery	221 (43.94)	86 (17.1)	135 (26.84)	0.222
Previous Inguinal Hernia Repair	32 (6.36)	9 (1.79)	23 (4.57)	0.102
Orchitis History	7 (1.39)	0 (0)	7 (1.39)	**0.024**
UTI History	8 (1.59)	7 (1.39)	1 (0.2)	**0.008**
Solitary Testis	15 (2.98)	4 (0.8)	11 (2.19)	0.223
Family History of Infertility	26 (5.17)	18 (3.58)	8 (1.59)	**0.004**
GUTI Before Secondary TESE	10 (1.99)	8 (1.59)	2 (0.4)	**0.014**
Smoking Before Secondary TESE	145 (28.83)	73 (14.51)	72 (14.31)	**0.015**
Primary TESE Histopathology				
Hypospermatogenesis	24 (4.77)	7 (1.39)	17 (3.38)	**<0.001**
Maturation Arrest	83 (16.5)	48 (9.54)	35 (6.96)	
Sertoli Cell–Only Syndrome	70 (13.92)	41 (8.15)	29 (5.77)	
Leydig cell hyperplasia	5 (0.99)	3 (0.6)	2 (0.4)	
Focal/preserved spermatogenesis	321 (63.82)	112 (22.27)	209 (41.55)	
Complication After Primary TESE	4 (0.8)	2 (0.4)	2 (0.4)	0.743
Secondary TESE Types				
Microscopic	333 (66.2)	146 (29.03)	187 (37.18)	0.228
Conventional	170 (33.8)	65 (12.92)	105 (20.87)	
Variocelectomy Before Secondary TESE	77 (15.31)	41 (8.15)	36 (7.16)	**0.029**
Medical Treatment Before Secondary TESE				
No Medical Treatment	377 (74.95)	138 (27.44)	239 (47.51)	**<0.001**
Hormonal Therapy	66 (13.12)	44 (8.75)	22 (4.37)	
Estrogen Receptor Modulator	49 (9.74)	26 (5.17)	23 (4.57)	
Antioxidant	10 (1.99)	3 (0.6)	7 (1.39)	
Aromatase Inhibitor	1 (0.2)	0 (0)	1 (0.2)	
Secondary Tese Location				
Bilateral	352 (69.98)	196 (38.97)	156 (31.01)	**<0.001**
Right	76 (15.11)	5 (0.99)	71 (14.12)	
Left	75 (14.91)	10 (1.99)	65 (12.92)	
Chemotherapy Before Secondary TESE	4 (0.8)	2 (0.4)	2 (0.4)	0.743
BMI Before Secondary Tese	24.6 ± 1.56 [24.46–24.73]	24.88 ± 1.81 [24.64–25.13]	24.39 ± 1.32 [24.24–24.54]	**<0.001**$^{\alpha}$
Spermiogram Volume	2.68 ± 1.22 [2.57–2.79]	2.96 ± 1.16 [2.8–3.12]	2.48 ± 1.24 [2.34–2.62]	**<0.001**$^{\alpha}$
Right Testis Volume Before Secondary TESE	12.72 ± 5.61 [12.23–13.22]	11.4 ± 5.98 [10.58–12.21]	13.68 ± 5.13 [13.08–14.27]	**<0.001**$^{\alpha}$
Left Testis Volume Before Secondary TESE	12.79 ± 5.63 [12.29–13.28]	11.34 ± 5.93 [10.53–12.15]	13.83 ± 5.18 [13.23–14.43]	**<0.001**$^{\alpha}$

Table 1. continued

Features	Overall	Negative Retrieval	Positive Retrieval	*p* value
FSH Before Secondary TESE	15.21 ± 12.8 [14.08–16.33]	16.75 ± 14.02 [14.84–18.66]	14.09 ± 11.74 [12.73–15.45]	.062
LH Before Secondary TESE	8.27 ± 8.09 [7.56–8.98]	9.95 ± 10.8 [8.47–11.42]	7.06 ± 5.03 [6.47–7.64]	**<0.001**$^{\alpha}$
Testosterone Before Secondary TESE	210.27 ± 205.88 [192.15–228.4]	212.59 ± 230.95 [181.09–244.09]	208.6 ± 186.08 [187.05–230.16]	0.485$^{\alpha}$

α Mann-Whitney U-Test **β** t-Test.

Table 2. Performance metrics of the developed machine learning (ML) models.

ML Models	Accuracy	Sensitivity	Specificity	NPV	PPV	F1 Score	Youden Index	AUC
XGBoost	0.90 ± 0.03 [0.89–0.91]	0.92 ± 0.04 [0.90–0.93]	0.87 ± 0.07 [0.84–0.90]	0.89 ± 0.05 [0.87–0.91]	0.91 ± 0.04 [0.89–0.93]	0.91 ± 0.03 [0.90–0.92]	0.79 ± 0.07 [0.76–0.82]	0.96 ± 0.02 [0.95–0.97]
Random Forest	0.90 ± 0.03 [0.90–0.91]	0.92 ± 0.04 [0.92–0.94]	0.87 ± 0.05 [0.85–0.89]	0.89 ± 0.05 [0.88–0.92]	0.91 ± 0.03 [0.90–0.92]	0.91 ± 0.03 [0.91–0.93]	0.79 ± 0.06 [0.78–0.82]	0.96 ± 0.02 [0.95–0.97]
Gradient Boosting	0.89 ± 0.03 [0.88–0.91]	0.92 ± 0.04 [0.90–0.94]	0.86 ± 0.05 [0.84–0.88]	0.89 ± 0.06 [0.87–0.91]	0.90 ± 0.03 [0.89–0.92]	0.91 ± 0.03 [0.90–0.92]	0.78 ± 0.06 [0.76–0.80]	0.96 ± 0.02 [0.96–0.97]
Decision Tree	0.84 ± 0.03 [0.82–0.85]	0.87 ± 0.04 [0.85–0.89]	0.79 ± 0.07 [0.76–0.82]	0.82 ± 0.05 [0.79–0.84]	0.86 ± 0.04 [0.84–0.87]	0.86 ± 0.03 [0.85–0.87]	0.67 ± 0.07 [0.63–0.68]	0.83 ± 0.04 [0.81–0.84]
AdaBoost	0.83 ± 0.04 [0.82–0.85]	0.87 ± 0.05 [0.85–0.89]	0.77 ± 0.08 [0.75–0.81]	0.81 ± 0.06 [0.79–0.84]	0.84 ± 0.04 [0.83–0.87]	0.85 ± 0.03 [0.84–0.87]	0.64 ± 0.08 [0.62–0.69]	0.82 ± 0.04 [0.81–0.84]
Logistic Regression	0.75 ± 0.04 [0.73–0.76]	0.79 ± 0.04 [0.77–0.80]	0.69 ± 0.08 [0.65–0.72]	0.70 ± 0.04 [0.68–0.72]	0.78 ± 0.04 [0.76–0.80]	0.78 ± 0.03 [0.77–0.79]	0.47 ± 0.09 [0.44–0.51]	0.82 ± 0.04 [0.80–0.84]
MLP	0.74 ± 0.04 [0.73–0.76]	0.77 ± 0.06 [0.74–0.78]	0.70 ± 0.09 [0.70–0.75]	0.69 ± 0.05 [0.67–0.71]	0.78 ± 0.05 [0.78–0.81]	0.78 ± 0.04 [0.76–0.79]	0.47 ± 0.09 [0.45–0.52]	0.82 ± 0.03 [0.81–0.83]
SVM	0.74 ± 0.03 [0.72–0.75]	0.77 ± 0.06 [0.75–0.79]	0.69 ± 0.06 [0.67–0.71]	0.69 ± 0.05 [0.67–0.71]	0.78 ± 0.03 [0.76–0.79]	0.77 ± 0.03 [0.76–0.79]	0.46 ± 0.06 [0.43–0.49]	0.82 ± 0.04 [0.80–0.83]
k-NN	0.71 ± 0.04 [0.70–0.73]	0.79 ± 0.04 [0.78–0.81]	0.60 ± 0.06 [0.58–0.63]	0.68 ± 0.05 [0.66–0.70]	0.74 ± 0.03 [0.72–0.75]	0.76 ± 0.03 [0.75–0.78]	0.40 ± 0.08 [0.37–0.43]	0.77 ± 0.04 [0.76–0.79]
Naive Bayes	0.58 ± 0.07 [0.55–0.60]	0.40 ± 0.27 [0.29–0.51]	0.82 ± 0.27 [0.71–0.93]	0.51 ± 0.10 [0.46–0.55]	0.83 ± 0.13 [0.78–0.89]	0.48 ± 0.19 [0.40–0.55]	0.22 ± 0.11 [0.17–0.27]	0.77 ± 0.05 [0.75–0.79]

XGB Xtreme Gradient Boosting, *RF* Random Forest, *GB* Gradient Boosting, *DT* Decision Tree, *AB* AdaBoost, *LR* Logistic Regression, *MLP* Multi-Layer Perceptron, *SVM* Support Vector Machine, *k-NN* K-Nearest Neighbors, *NB* Naive Bayes, *NPV* Negative predictive value, *PPV* Positive predictive value, *AUC* Area under curve.

(LH) levels ($p < 0.001$), but significantly higher testicular volumes bilaterally ($p < 0.001$).

Procedure type and surgical history. Bilateral TESE during the secondary procedure was significantly more common among patients with negative sperm retrieval outcomes ($p < 0.001$).

The number of patients who underwent varicocelectomy prior to the primary TESE was similar in both groups. However, varicocelectomy performed before the secondary TESE was significantly more common among patients with negative sperm retrieval outcomes ($p = 0.029$).

Medical history and lifestyle factors. A significantly higher proportion of patients with negative sperm retrieval outcomes had a family history of infertility ($p = 0.004$), a history of genitourinary ($p = 0.014$) or urinary tract infections ($p = 0.008$), and a smoking history ($p = 0.015$). Conversely, a history of orchitis was more common among those with successful sperm retrieval ($p = 0.024$).

Genetic and histopathological findings. Klinefelter syndrome was associated more highly with negative retrieval outcomes ($p = 0.002$).

Negative sperm retrieval was also highly associated with classical spermatogenic histopathological patterns observed on primary TESE, including maturation arrest and Sertoli cell-only syndrome ($p < 0.001$). Leydig cell hyperplasia was recorded as an additional interstitial histopathological finding and was more frequently observed in patients with unsuccessful retrieval.

Performance of machine learning models

Accuracy. Accuracy, a commonly used metric for evaluating machine learning models, represents the proportion of correct predictions among all predictions. The models achieving the highest accuracy were XGB (0.90 ± 0.03; 95%CI [0.89–0.91]) and RF (0.90 ± 0.03; 95%CI [0.90–0.91]) (Table 2 and Fig. 1). Although XGB, RF, and GB (0.89 ± 0.03; 95% CI [0.88–0.91]) outperformed all other models, no significant differences were observed among these three (Supplementary Fig. 1A–C).

Sensitivity and specificity. Analysis of sensitivity revealed that three models achieved the highest performance (XGB: 0.92 ± 0.04; 95% CI [0.90–0.93], RF: 0.92 ± 0.04; 95% CI [0.92–0.94], and GB: 0.92 ± 0.04; 95% CI [0.90–0.94]) (Table 2 and Fig. 1). While these models demonstrated higher sensitivity than DT and AB, the differences did not reach statistical significance. Nevertheless, they showed statistically superior sensitivity compared with all other algorithms, excluding DT and AB (Supplementary Fig. 2A–C).

With respect to specificity, the highest values were obtained from XGB (0.87 ± 0.07; 95% CI [0.84–0.90]) and RF (0.87 ± 0.05; 95% CI [0.85–0.89]) (Table 2 and Fig. 1). Although both models exhibited higher specificity than DT, AB, and NB, these differences were not statistically significant. However, XGB and RF demonstrated significantly higher specificity than all remaining models (Supplementary Fig. 3A–C).

Negative and positive predictive values. NPV, defined as the proportion of true negatives among the sum of true negatives and

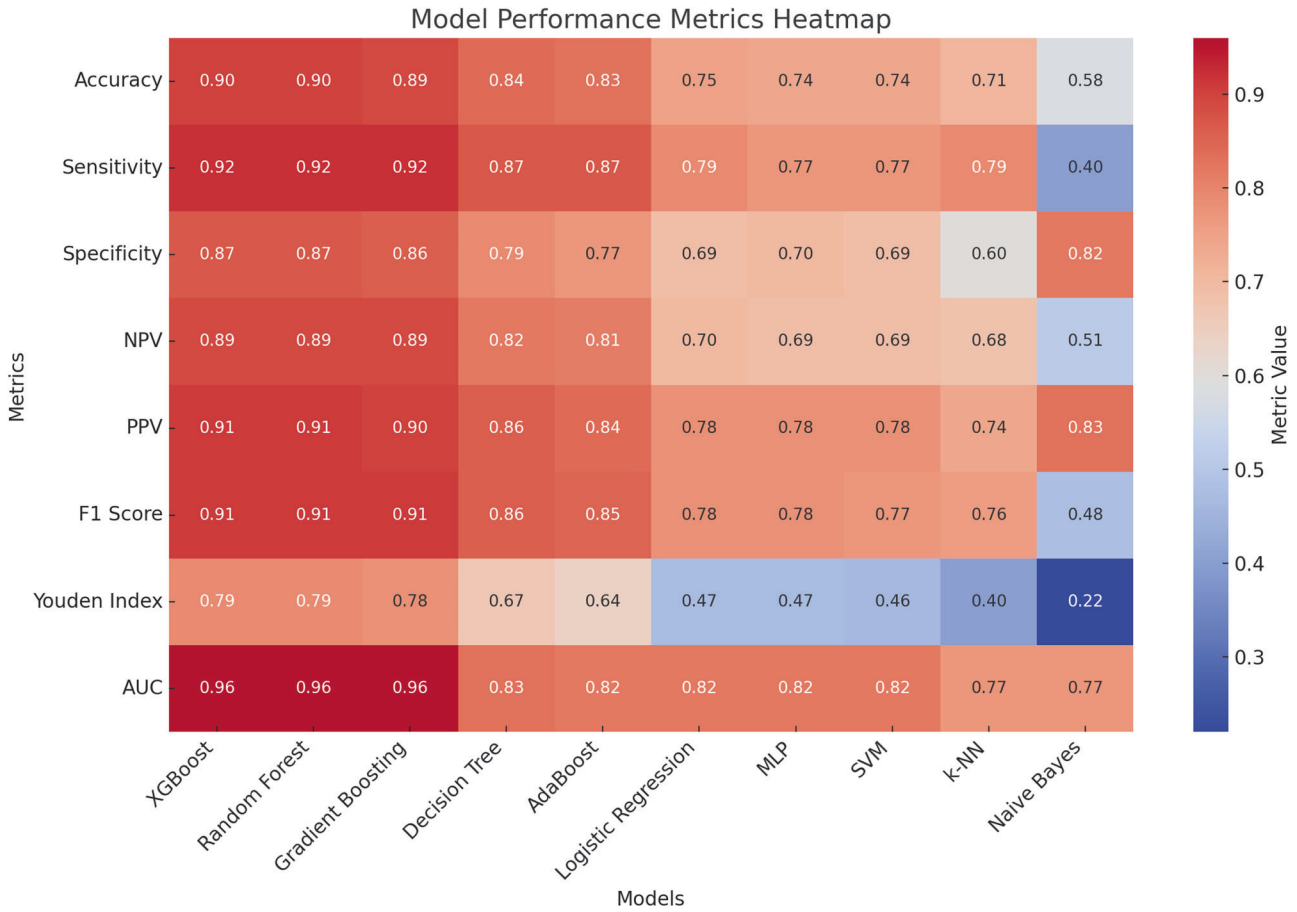


Fig. 1 Performance comparison of machine learning models predicting the success of secondary theses. This figure illustrates the performance metrics—accuracy, sensitivity, and specificity—of ten machine learning models (XGBoost, Random Forest, Gradient Boosting, Decision Tree, AdaBoost, Logistic Regression, MLP, SVM, k-NN, and Naive Bayes) applied to predict positive outcomes in patients undergoing secondary TESE procedures.

false negatives, was highest in XGB (0.89 ± 0.05; 95% CI [0.87–0.91]), RF (0.89 ± 0.05; 95% CI [0.88–0.92]), and GB (0.89 ± 0.06; 95% CI [0.87–0.91]) (Table 2 and Fig. 1). Although no significant differences were observed among these three models, each significantly outperformed all other algorithms (Supplementary Fig. 4A–C).

PPV, defined as the proportion of true positives among the sum of true positives and false positives, reached its highest values in XGB (0.91 ± 0.04; 95% CI [0.89–0.93]) and RF (0.91 ± 0.03; 95% CI [0.90–0.92]) (Table 2 and Fig. 1). Although these two models demonstrated higher PPV compared with GB (0.90 ± 0.03; 95% CI [0.89–0.92]), the difference was not statistically significant ($p = 1.0$). Nonetheless, both XGB and RF significantly outperformed all other models (Supplementary Fig. 5A–C).

F1 score. The F1 score, calculated as the harmonic mean of precision and recall, is widely applied to assess model performance in both binary and multi-class settings. In this study, the highest F1 scores were obtained from XGB (0.91 ± 0.03; 95% CI [0.90–0.92]), RF (0.91 ± 0.03; 95% CI [0.91–0.93]), and GB (0.91 ± 0.03; 95% CI [0.90–0.92]) (Table 2 and Fig. 1). XGB and GB exhibited significantly higher F1 scores compared with all models except RF, DT, and AB. The RF model, in turn, significantly outperformed all models other than XGB and GB (Supplementary Fig. 6A–C).

Youden index. The Youden Index, defined as sensitivity plus specificity minus one, provides a cumulative measure of these two metrics. Tools or biomarkers with a Youden Index above 0.5 are generally considered clinically useful. The highest values were observed in XGB (0.79 ± 0.07; 95% CI [0.76–0.82]) and RF (0.79 ± 0.06; 95% CI [0.78–0.82]) (Table 2 and Fig. 1). Although both models achieved higher Youden Index values, no significant differences were found when compared with GB ($p = 1.0$). Nevertheless, XGB and RF demonstrated statistically superior performance relative to all other models (Supplementary Fig. 7A–C).

Area under the curve (AUC). The AUC summarizes the ROC curve, with higher values indicating stronger discriminative ability. An AUC of 1.0 reflects perfect classification, whereas 0.5 corresponds to random performance. Examination of AUC values revealed the highest performance in XGB (0.96 ± 0.02; 95% CI [0.95–0.97]), RF (0.96 ± 0.02; 95% CI [0.95–0.97]), and GB (0.96 ± 0.02; 95% CI [0.96–0.97]) (Table 2 and Figs. 1 and 2). No significant differences were observed among these three models. However, each significantly outperformed all other algorithms in terms of AUC (Supplementary Fig. 8A–C).

Superiority of XGBoost in Predicting Secondary TESE Outcomes

Among the evaluated algorithms, XGBoost consistently achieved top-tier performance across all metrics, including accuracy, sensitivity, specificity, PPV, NPV, F1 score, Youden Index, and AUC. While Random Forest and Gradient Boosting also demonstrated strong results, XGBoost matched or exceeded their performance and significantly outperformed the remaining models. Its superior discriminative ability (AUC = 0.96 ± 0.02; 95% CI [0.95–0.97]), combined with balanced sensitivity and specificity and high predictive values, highlights its robustness and reliability for predicting secondary TESE outcomes.

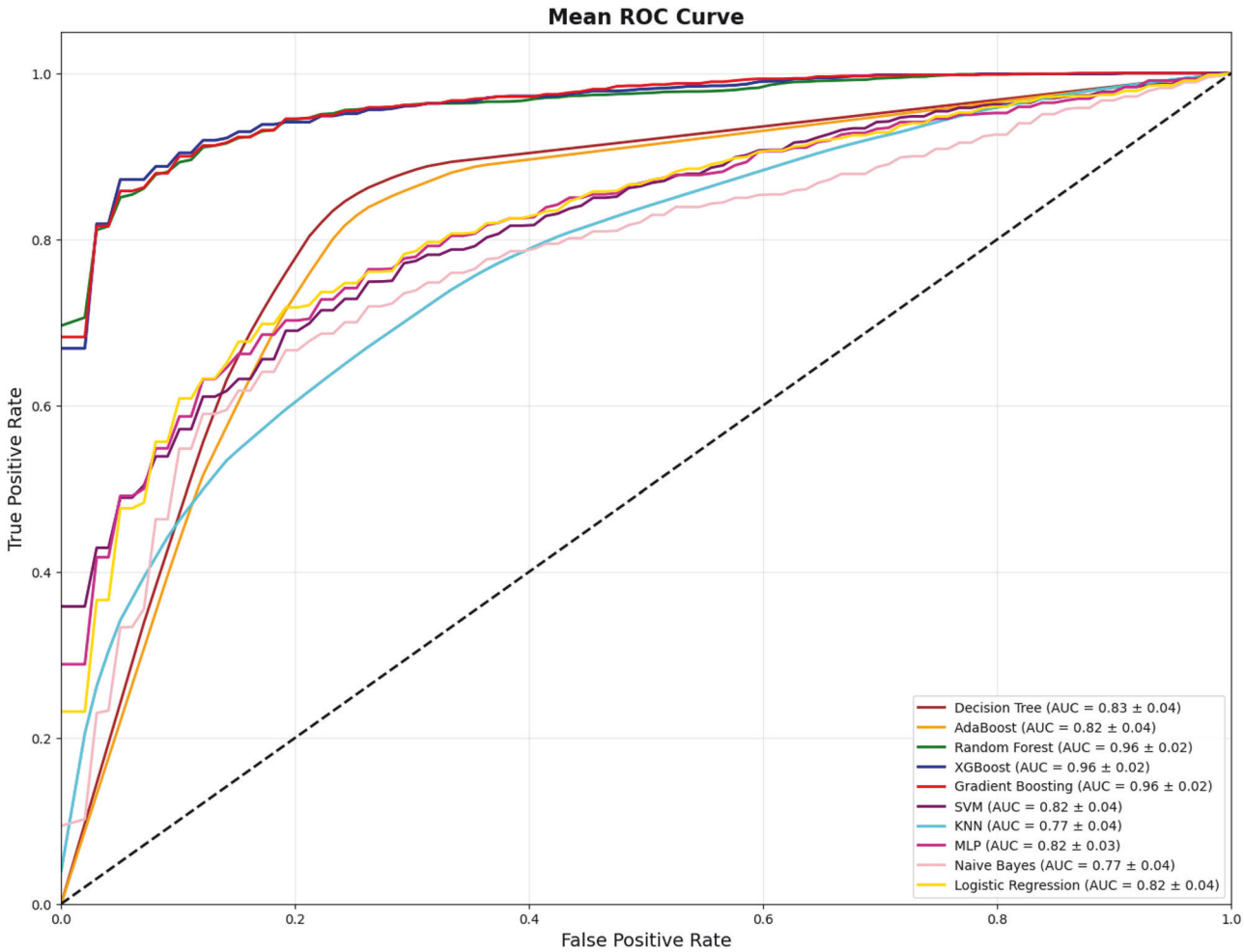


Fig. 2 Receiver operating characteristic (ROC) curve of the best-performing machine learning models for predicting the success of secondary TESE. This figure illustrates the receiver operating characteristic (ROC) curves of ten machine learning models (XGBoost, Random Forest, Gradient Boosting, Decision Tree, AdaBoost, Logistic Regression, MLP, SVM, k-NN, and Naive Bayes) developed to predict positive outcomes in patients undergoing secondary TESE. The curves demonstrate the discriminative performance of each model across different thresholds.

SHapley additive explanations results

The SHAP analysis revealed that several clinical and laboratory parameters had the strongest impact on predicting the success of secondary TESE procedures (Fig. 3). Among these, BMI prior to secondary TESE, TESE location, LH level before secondary TESE, and semen volume were the most influential predictors. Higher BMI values were generally associated with a negative contribution to model output, indicating a reduced likelihood of sperm retrieval. In contrast, certain TESE locations contributed positively, suggesting that surgical site selection may play a role in improving retrieval outcomes. Elevated LH levels and reduced semen volume were also among the key determinants influencing unsuccessful outcomes.

Other relevant features included testicular volumes (left and right) and duration of infertility, both of which demonstrated substantial impact on model predictions. Lower testicular volumes and longer infertility duration were associated with reduced probability of sperm retrieval. Additionally, serum testosterone and FSH levels before secondary TESE showed moderate contributions, where lower testosterone and higher FSH tended to negatively influence outcomes.

DISCUSSION

XGBoost showed the strongest and most consistent predictive performance among the ML algorithms. SHAP analysis identified baseline hormonal profiles (LH, FSH, testosterone), testicular reserve indicators (testicular and semen volume), and clinical factors (BMI, infertility duration, TESE location) as key determinants of sperm retrieval success. The most influential predictors were BMI, TESE location, LH level, and semen volume. SHAP also provided interpretable insights, confirming known predictors of impaired spermatogenesis and emphasizing the importance of modifiable factors such as BMI.

In contrast to previously developed prediction models for SSR [20, 31, 32], our model achieved a very high level of predictive performance, which may be attributed to the ability of machine learning algorithms to capture complex, non-linear relationships. However, these results should be interpreted with caution, given the retrospective study design, potential center-related heterogeneity, and the absence of external validation.

TESE location demonstrated strong predictive importance in the SHAP analysis. However, it should not be considered a purely preoperative factor. Instead, it reflects decisions made during surgery, particularly the need to explore the opposite testis when sperm are not found initially. Therefore, its predictive value likely represents surgical complexity and the severity of the underlying testicular dysfunction rather than a direct biological effect. This point is important when evaluating the clinical usefulness of the model, especially for preoperative counseling.

Although FSH has been most widely studied [33, 34], increasing evidence suggests that LH and testosterone may also play important roles in predicting SRR. A recent study found that,

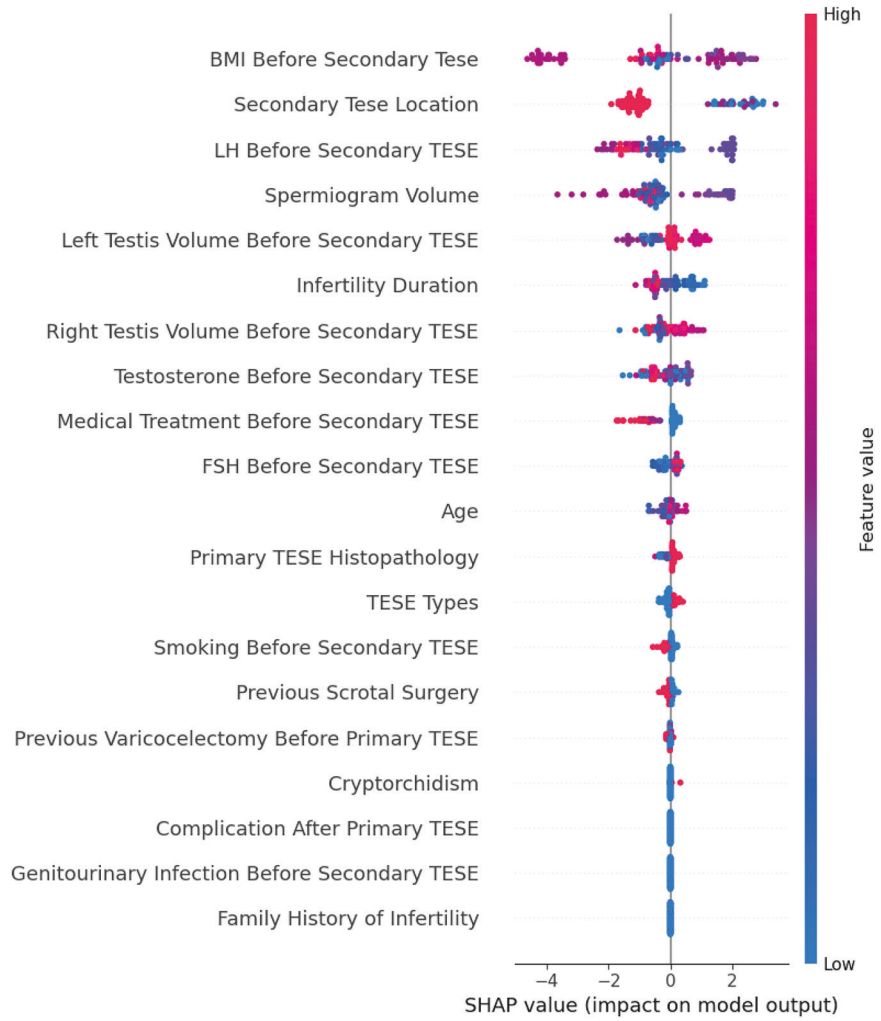


Fig. 3 SHapley Additive exPlanations (SHAP) results showing the clinical and laboratory parameters with the greatest impact on predicting the success of secondary TESE procedures. This figure illustrates the clinical and laboratory variables influencing the prediction of positive outcomes from secondary TESE procedures, ranked from the most to the least significant.

although repeat TESE cases exhibited higher LH and lower total testosterone compared to first-time cases, these hormonal differences did not influence sperm retrieval rates [21]. Another study of 79 NOA patients with a prior failed conventional TESE similarly reported no significant differences in LH or testosterone, regardless of retrieval outcome after salvage mTESE [35]. In contrast, Xu et al. observed in 52 NOA patients undergoing salvage mTESE that testosterone—but not LH or FSH—significantly influenced retrieval outcomes [32]. A large multicenter study of 3093 men undergoing TESE partly echoed and partly diverged from these findings, as its nomogram for sperm retrieval success identified both LH and testosterone, along with FSH, as promising predictors [36]. In contrast, in our cohort, including the largest population of patients undergoing repeated TESE, LH emerged as the strongest contributor, while testosterone demonstrated a moderate effect, underscoring the potential clinical relevance of these hormones. Such a finding likely reflects the central role of the hypothalamic–pituitary–gonadal axis in regulating spermatogenesis, where elevated LH could indicate impaired Leydig cell function and consequent intratesticular testosterone deficiency.

Smaller testicular volumes have been identified as predictive of azoospermia caused by spermatogenic dysfunction, or NOA [37]. Testicular volume has also been associated with the severity of testicular histopathology in men with NOA [34]. Several studies have demonstrated that larger testicular volumes may increase SRR in TESE procedures [8, 36, 38, 39], whereas others have found no significant difference between successful and unsuccessful retrieval after salvage mTESE [32, 35]. Moreover, patients with maturation arrest often present with normal testicular volumes but very low SRR [14, 15]. Therefore, in patients with NOA, larger testicular volume does not necessarily correspond to a higher probability of sperm retrieval. Some authors have suggested that combining testicular volume with histological subtype may serve as an independent predictor [31]. Our results are consistent with this view, as both variables were identified as predictors in our ML analysis.

Testicular histopathology has consistently been reported as one of the most reliable predictors of sperm retrieval success, particularly in salvage or repeat TESE settings. Previous studies have shown that men with hypospermatogenesis have significantly higher retrieval rates compared with those with maturation arrest or Sertoli cell-only syndrome [11]. In our cohort, classical spermatogenic patterns were strongly associated with secondary TESE outcomes, further supporting the central role of histology in prognostic stratification. Although histopathology did not emerge as the most influential variable in the SHAP ranking, its strong univariate association and well-established clinical relevance underscore its fundamental prognostic value.

In patients with historically low retrieval probability, such as those with Klinefelter syndrome or Sertoli cell-only syndrome, success rates are significantly reduced [40]. Nevertheless, sperm retrieval is not impossible, and selected patients may still benefit from repeat attempts. In current practice, salvage TESE is often considered in the context of individualized counseling, patient preference, and previous surgical findings. Therefore, predictive models may be particularly valuable in helping clinicians and patients balance the potential benefits and low likelihood of success when considering repeat procedures.

The relationship between male obesity and sperm retrieval success remains controversial. A retrospective study of 997 patients with NOA demonstrated that increased male BMI was associated with a lower likelihood of achieving clinical pregnancy after ICSI, even when sperm was successfully retrieved via mTESE [41]. However, Iwatsuki et al. questioned this association, reporting in their analysis of 172 azoospermic men that there was no significant difference in sperm retrieval success between patients with BMI < 25 kg/m^2 and those with BMI ≥ 25 kg/m^2 [42]. Similarly, Pavan-Jukic et al., in a study of 75 azoospermic men, also found no significant correlation between BMI and successful sperm retrieval [43]. In line with some previous reports, our findings demonstrated an association between higher BMI and lower TESE success. However, this relationship should be interpreted cautiously, as the retrospective design does not allow causal inference, and the direct effect of weight reduction on sperm retrieval outcomes remains uncertain.

This study has several strengths. It is the largest multicenter cohort to date examining predictors of secondary TESE in NOA, enhancing statistical power and generalizability. Including surgical cases from many centers strengthens the feasibility data of the surgical method and the multivariable predictivity of the pathology parameter, not being burdened by the potential bias (either negative or positive) of a single surgeon and a single embryologist. Multiple ML algorithms were applied, with XGBoost showing the best performance, and SHAP analysis providing interpretable insights. The models confirmed known predictors (FSH, testicular volume, histopathology) and identified modifiable factors such as BMI. Focusing on salvage TESE, a challenging subgroup, the findings offer valuable guidance for preoperative counseling and surgical decisions.

This study has several limitations. First, the retrospective design of the study may introduce selection bias. Second, surgical sperm retrieval technique (conventional vs micro-TESE) and the use of varicocelectomy prior to primary TESE were not standardized across centers, and were dependent on surgeon preference and institutional resources, which may have contributed to performance bias. Third, variability in histopathological assessment from the primary TESE across participating centers may have influenced the classification of spermatogenic patterns. In addition, the multicenter nature of the study introduces potential heterogeneity in surgical expertise, sperm processing protocols, and laboratory practices, which could affect sperm retrieval outcomes and model performance. Fourth, although the machine learning models demonstrated high predictive performance, external validation in independent cohorts is essential, particularly given the risk of overfitting inherent to ML approaches. Model performance observed in this study may not be directly generalizable to other populations or clinical settings without such validation. Finally, residual confounding from unmeasured factors, including environmental exposures, lifestyle variables, and genetic influences, cannot be excluded. Therefore, the findings should be interpreted as hypothesis-generating rather than definitive tools for clinical decision-making.

CONCLUSION

This multicenter cohort study used ML to assess the impact of traditional clinical and modifiable factors on secondary TESE outcomes in men with NOA. Integrating diverse patient data into predictive models may improve counseling, guide surgical decisions, and reduce unnecessary procedures. However, these models should currently be considered adjunctive decision-support tools rather than replacements for expert clinical judgment. Future prospective studies with external validation are needed to confirm and translate these findings into clinical practice.

Guarantor statement

M.G had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Attestation statements

Data regarding any of the subjects in the study have not been previously published unless specified. - Data will be made available to the editors of the journal pre and/or post-publication for review or query upon request. - TRIPOD + AI and STREAM-URO guidelines for this study design were followed

DATA AVAILABILITY

The de-identified dataset generated and analyzed during the current study is not publicly available due to patient confidentiality considerations and multicenter data governance, but may be made available from the corresponding author upon reasonable request.

REFERENCES

1. Cooper TG, Noonan E, von Eckardstein S, Auger J, Baker HW, Behre HM, et al. World Health Organization reference values for human semen characteristics. Hum Reprod Update. 2010;16:231–45.
2. Cocuzza M, Alvarenga C, Pagani R. The epidemiology and etiology of azoospermia. Clinics. 2013;68:15–26.
3. Salonia A, Boeri L, Capogrosso P, Corona G, Dinkelman-Smith M, Gül M, et al. EAU guidelines on sexual and reproductive health 2026. In: European Association of Urology guidelines. Arnhem: EAU Guidelines Office; 2026.
4. Bernie AM, Mata DA, Ramasamy R, Schlegel PN. Comparison of microdissection testicular sperm extraction, conventional testicular sperm extraction, and testicular sperm aspiration for nonobstructive azoospermia: a systematic review and meta-analysis. Fertil Steril. 2015;104:1099–103.e1-3.
5. Su LM, Palermo GD, Goldstein M, Veeck LL, Rosenwaks Z, Schlegel PN. Testicular sperm extraction with intracytoplasmic sperm injection for nonobstructive azoospermia: testicular histology can predict the success of sperm retrieval. J Urol. 1999;161:112–6.
6. Amer M, Ateyah A, Hany R, Zohdy W. Prospective comparative study between microsurgical and conventional testicular sperm extraction in non-obstructive azoospermia: follow-up by serial ultrasound examinations. Hum Reprod. 2000;15:653–6.
7. Ramasamy R, Yagan N, Schlegel PN. Structural and functional changes to the testis after conventional versus microdissection testicular sperm extraction. Urology. 2005;65:1190–4.
8. Corona G, Minhas S, Giwercman A, Bettocchi C, Dinkelman-Smit M, Dohle G, et al. Sperm recovery and ICSI outcomes in men with non-obstructive azoospermia: a systematic review and meta-analysis. Hum Reprod Update. 2019;25:733–57.
9. Schlegel PN, Su LM. Physiological consequences of testicular sperm extraction. Hum Reprod. 1997;12:1688–92.
10. Tash JA, Schlegel PN. Histologic effects of testicular sperm extraction on the testicle in men with nonobstructive azoospermia. Urology. 2001;57:334–7.
11. Boeri L, Bebi C, Dente D, Greco E, Turetti M, Capece M, et al. Outcomes and predictive factors of successful salvage microdissection testicular sperm extraction (mTESE) after failed classic TESE: results from a multicenter cross-sectional study. Int J Impot Res. 2022;34:795–9.
12. Vernaeve V, Verheyen G, Goossens A, Van Steirteghem A, Devroey P, Tournaye H. How successful is repeat testicular sperm extraction in patients with azoospermia? Hum Reprod. 2006;21:1551–4.
13. Yücel C, Budak S, Keskin MZ, Kisa E, Kozacioglu Z. Predictive factors of successful salvage microdissection testicular sperm extraction (mTESE) after failed mTESE in patients with non-obstructive azoospermia: Long-term experience at a single institute. Arch Ital Urol Androl. 2018;90:136–40.
14. Shiraishi K, Ohmi C, Shimabukuro T, Matsuyama H. Human chorionic gonadotrophin treatment prior to microdissection testicular sperm extraction in non-obstructive azoospermia. Hum Reprod. 2012;27:331–9.
15. Özman O, Tosun S, Bayazıt N, Cengiz S, Bakırcıoğlu ME. Efficacy of the second micro-testicular sperm extraction after failed first micro-testicular sperm extraction in men with nonobstructive azoospermia. Fertil Steril. 2021;115:915–21.

16. Amer MK, Ahmed AR, Hamid AAA, GamalEl Din SF. Factors determining the sperm retrieval rate in fresh versus salvage micro-TESE: a comparative cohort study. Int Urol Nephrol. 2019;51:401–8.
17. Talas H, Yaman O, Aydos K. Outcome of repeated micro-surgical testicular sperm extraction in patients with non-obstructive azoospermia. Asian J Androl. 2007;9:668–73.
18. Brannigan RE, Hermanson L, Kaczmarek J, Kim SK, Kirkby E, Tanrikut C. Updates to Male Infertility: AUA/ASRM Guideline (2024). J Urol. 2024;212:789–99.
19. Bachelot G, Dhombres F, Sermondade N, Haj Hamid R, Berthaut I, Frydman V, et al. A machine learning approach for the prediction of testicular sperm extraction in nonobstructive azoospermia: algorithm development and validation study. J Med Internet Res. 2023;25:e44047.
20. Zeadna A, Khateeb N, Rokach L, Lior Y, Har-Vardi I, Harlev A, et al. Prediction of sperm extraction in non-obstructive azoospermia patients: a machine-learning perspective. Hum Reprod. 2020;35:1505–14.
21. Alriyalat S, Deameh MG, Farraj H, Alsmadi JK, Bani Irshid BA, Abu Rahmeh H, et al. Predictors of sperm retrieval success in first-time and repeated Micro-TESE for nonobstructive azoospermia. Future Sci OA. 2025;11:2511449.
22. Beaulieu-Jones BK, Yuan W, Brat GA, Beam AL, Weber G, Ruffin M, et al. Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians? NPJ Digit Med. 2021;4:62.
23. Çeker G, Deliktaş H, Gül Ü, Turunç T. The andrology working group of the society of urological surgery in turkiye: bridging academic gaps and advancing male reproductive or sexual health. J Urol Surg. 2025;12:101–107.
24. Organization WHO. WHO Laboratory Manual for the Examination and Processing of Human Semen. Who's Laboratory Manual for the Examination and Processing of human semen, 2021.
25. Dubin L, Amelar RD. Varicocelectomy as therapy in male infertility: a study of 504 cases. Fertil Steril. 1975;26:217–20.
26. Schlegel PN. Testicular sperm extraction: microdissection improves sperm yield with minimal tissue excision. Hum Reprod. 1999;14:131–5.
27. Johnsen SG. Testicular biopsy score count–a method for registration of spermatogenesis in human testes: normal values and results in 335 hypogonadal males. Hormones. 1970;1:2–25.
28. Kwong JC, McLoughlin LC, Haider M, Goldenberg MG, Erdman L, Rickard M, et al. Standardized reporting of machine learning applications in urology: the STREAM-URO framework. European Urology Focus. 2021;7:672–82.
29. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD +AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. Bmj. 2024;385:e078378.
30. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol. 2017;17:162.
31. Zhang F, Dai M, Yang X, Cheng Y, Ye L, Huang W, et al. Predictors of successful salvage microdissection testicular sperm extraction (mTESE) after failed initial TESE in patients with non-obstructive azoospermia: a systematic review and meta-analysis. Andrology. 2024;12:30–44.
32. Xu T, Peng L, Lin X, Li J, Xu W. Predictors for successful sperm retrieval of salvage microdissection testicular sperm extraction (TESE) following failed TESE in nonobstructive azoospermia patients. Andrologia. 2017;49 https://doi.org/10.1111/and.12642.
33. Martin-du-Pan RC, Bischof P. Increased follicle-stimulating hormone in infertile men. Is increased plasma FSH always due to damaged germinal epithelium? Hum Reprod. 1995;10:1940–5.
34. Kavoussi PK, Hudson K, Machen GL, Barsky M, Lebovic DI, Kavoussi SK. FSH levels and testicular volumes are associated with the severity of testicular histopathology in men with non-obstructive azoospermia. J Assist Reprod Genet. 2021;38:3015–8.
35. Caroppo E, Castiglioni F, Campagna C, Colpi EM, Piatti E, Gazzano G, et al. Intrasurgical parameters associated with successful sperm retrieval in patients with non-obstructive azoospermia undergoing salvage microdissection testicular sperm extraction. Andrology. 2021;9:1864–71.
36. Ceyhan E, Kayra MV, Gul Ates E, Kizilkan Y, Altan M, Yildirim O, et al. A nomogram predicting testicular sperm extraction success in men with non-obstructive azoospermia: a multi-center study. Urology. 2025;196:155–61.
37. Schoor RA, Elhanbly S, Niederberger CS, Ross LS. The role of testicular biopsy in the modern management of male infertility. J Urol. 2002;167:197–200.
38. Ntorkou A, Tsili AC, Goussia A, Astrakas LG, Maliakas V, Sofikitis N, et al. Testicular apparent diffusion coefficient and magnetization transfer ratio: can these MRI parameters be used to predict successful sperm retrieval in nonobstructive azoospermia? AJR Am J Roentgenol. 2019;213:610–8.
39. Cetinkaya M, Onem K, Zorba OU, Ozkara H, Alici B. Evaluation of microdissection testicular sperm extraction results in patients with non-obstructive azoospermia: independent predictive factors and best cutoff values for sperm retrieval. Urol J. 2015;12:2436–43.
40. Boeri L, Palmisano F, Preto M, Sibona M, Capogrosso P, Franceschelli A, et al. Sperm retrieval rates in non-mosaic Klinefelter patients undergoing testicular sperm extraction: What expectations do we have in the real-life setting? Andrology. 2020;8:680–7.
41. Ramasamy R, Bryson C, Reifsnyder JE, Neri Q, Palermo GD, Schlegel PN. Overweight men with nonobstructive azoospermia have worse pregnancy outcomes after microdissection testicular sperm extraction. Fertil Steril. 2013;99:372–6.
42. Iwatsuki S, Sasaki S, Taguchi K, Hamakawa T, Mizuno K, Okada A, et al. Effect of obesity on sperm retrieval outcome and reproductive hormone levels in Japanese azoospermic men with and without Klinefelter syndrome. Andrology. 2017;5:82–6.
43. Pavan-Jukic D, Starc A, Stubljar D, Jukic T. Obesity with high body mass index does not influence sperm retrieval in males with azoospermia. Med Sci Monit. 2020;26:e923060.

ACKNOWLEDGEMENTS

We would like to thank the Andrology Working Group of SUST for their endless support.

AUTHOR CONTRIBUTIONS

Conceptualization: BH, AŞ and MG. Methodology: BH, AŞ and MG Software: AŞ Formal analysis: AŞ. Investigation: BH, AŞ, MHG, CO, MVK, EH, EC, MA, TC, YK, MBD, MAİ, GC, CS, NA, İU, ACA, AB and MG. Resources: BH, AŞ, MHG, CO, MVK, EH, EC, MA, TC, YK, MBD, MAİ, GC, CS, NA, İU, ACA, AB and MG. Data curation: BH, AŞ, MHG, CO, MVK, EH, EC, MA, TC, YK, MBD., MAİ, GC, CS, NA, İU, ACA, AB and MG. Writing—original draft: BH and AŞ Writing—review and editing: MG Visualization: AŞ Supervision: MG Project administration: MG

FUNDING

Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

COMPETING INTERESTS

Murat Gul, a co-author of this study, serves as an Associate Editor of the International Journal of Impotence Research (IJIR). All other authors declare no competing interests.

ETHICS STATEMENT

The present study protocol was reviewed and approved by the institutional review board of Selçuk University (Reg. No. 2024/379). Informed consent was submitted by all subjects when they were enrolled. Relevant guidelines and regulations were followed throughout the study.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41443-026-01291-y.

Correspondence and requests for materials should be addressed to Murat Gül.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026