Submitted: 10 October, 2024 Accepted: 20 November, 2024 Published: 30 December, 2024 DOI:10.22514/jomh.2024.201

JOMH Journal of Men's Health

ORIGINAL RESEARCH

Developing a machine learning model for predicting varicocelectomy outcomes: a pilot study

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

[1]Department of Urology, Health Science University Eskisehir City Health Application and Research Center, 26000 Eskisehir, Turkey
[2]Department of Mathematics—Computer Science, Faculty of Science, Eskisehir Osmangazi University, 26000 Eskisehir, Turkey

***Correspondence**
coskun.kaya@sbu.edu.tr
(Coşkun Kaya)

Abstract

A further debatable issue in the treatment of varicocele is which men would benefit from a varicocelectomy. Despite the increasing interest in Machine Learning (ML) in urology, there have been limited studies on the detection and prediction of varicocelectomy using artificial intelligence. We aimed to develop a model to predict the improvement in semen parameters after varicocelectomy using ML.The data for male patients who had clinical varicocele, abnormal semen parameters (low sperm concentration, reduced total motile sperm count, decreased progressive motility, and/or poor sperm morphology) and had received a varicocelectomy were recorded retrospectively. Demographic, anthropometric variables, physical examination findings, hematological, radiological, and semen analysis parameters were evaluated. The patients were separated into two groups according to the improvement in total motile sperm count postoperatively as improvement (Group 1) and no improvement (Group 2). The Extra Trees Classifier, Light Gradient Boosting Machine Classifier, eXtreme Gradient Boosting Classifier, Logistic Regression, and Random Forest Classifier techniques were used as ML algorithms.41 males were included in the study. 31 (75.6%) and 10 (24.4%) patients were classified as Group 1 and 2, respectively. The Extra Trees Classifier algorithm was found to be the best ML technique for predictions, according to the accuracy rates (92.3%) with an Area Under Curve of 0.92. We have shown for the first time in the literature that basic laboratory and semen analysis findings can be used to select patients who will benefit from varicocelectomy with the use of five ML methods. ML models could be identified as a new prediction tool for selecting the patients who will benefit from varicocelectomy. More detailed ML studies will be needed a larger number of patients.

Keywords

Spermiogram; Varicocelectomy; Machine learning; Artificial intelligence

1. Introduction

Varicocele, defined as the abnormal dilation of the veins within the spermatic cord, is a common condition among men and is known to contribute to various andrological issues, including infertility and testicular discomfort [1]. It is thought to affect fertility through several mechanisms, but the exact mechanism has not yet been described. Another issue is whether all men with varicocele should receive the treatment. The guidelines suggest that a man with a subclinical varicocele or a man with a clinical varicocele and a normal semen analysis do not need a varicocelectomy [2]. A further debatable issue in the treatment of varicocele is which men would benefit from a varicocelectomy. Several studies, meta-analyses, and reviews have been carried out in the literature on the success and failure rate of this procedure [3–5]. Researchers have explored the use of preoperative laboratory and radiological findings as clinical predictors to estimate the likelihood of achieving normal semen analysis results following surgery [6–8].

Machine Learning (ML), a subset of artificial intelligence (AI), has begun to be used in medicine during the last decade, with applications showing great potential in assisting and improving care. ML is going to be more useful in surgical systems where large-scale data is a limited resource, such as in low- and middle-income countries [9]. Reinforcement learning in surgical decision-making, predictive analytics, and preoperative risk stratification are other areas where ML can be applied [10]. These new technologies have been in use in the field of urology for some time and the publication rate has been increasing recent years [11]. Despite the increasing interest in ML in urology, to the best of our knowledge, there have been only 2 studies on the detection and prediction of varicocelectomy using AI [12, 13].

In this study, we aimed to establish an ML model for predicting patients who will benefit from varicocelectomy using basic laboratory, scrotal Doppler ultrasound findings (SDU), and semen analysis findings.

2. Material method

2.1 Data source and study workflow

After obtaining ethical approval, the medical records of patients who had received a varicocelectomy due to both having clinical varicocele and impaired semen parameters at the Eskisehir City Hospital Department of Urology between January 2019 and July 2023 were evaluated retrospectively. The inclusion criteria were: aged ≥18 years, having had at least two semen analyses given at different times showing impaired semen quality in at least one of the following semen parameters: sperm concentration, total sperm count, total motility or morphology (where the recorded value was under the reference limits of World Health Organization (WHO 2021), a complete blood count (CBC), and a scrotal Doppler ultrasonography (SDU) performed preoperatively, and a semen analysis carried out 6 months after the varicocelectomy. Exclusion criteria were: azoospermia, where all semen parameters above the reference limits of WHO 2021 preoperatively, 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 mL), treated cancer, vascular problem, hematologic illness, systemic disease, genitourinary system infection or hormonal problems (Fig. 1).

The age, body mass index (BMI), reason for visiting the clinic, physical examination findings, semen analysis, CBC and SDU results of the patients were recorded. To standardize the varicocele examination and scrotal Doppler ultrasound, the protocol previously published by Foroughi was considered and implemented [14]. We utilized several components of the Complete Blood Count (CBC) as parameters to predict the outcomes of varicocelectomy. These included hemoglobin, neutrophils, lymphocytes, monocytes, platelets, red cell

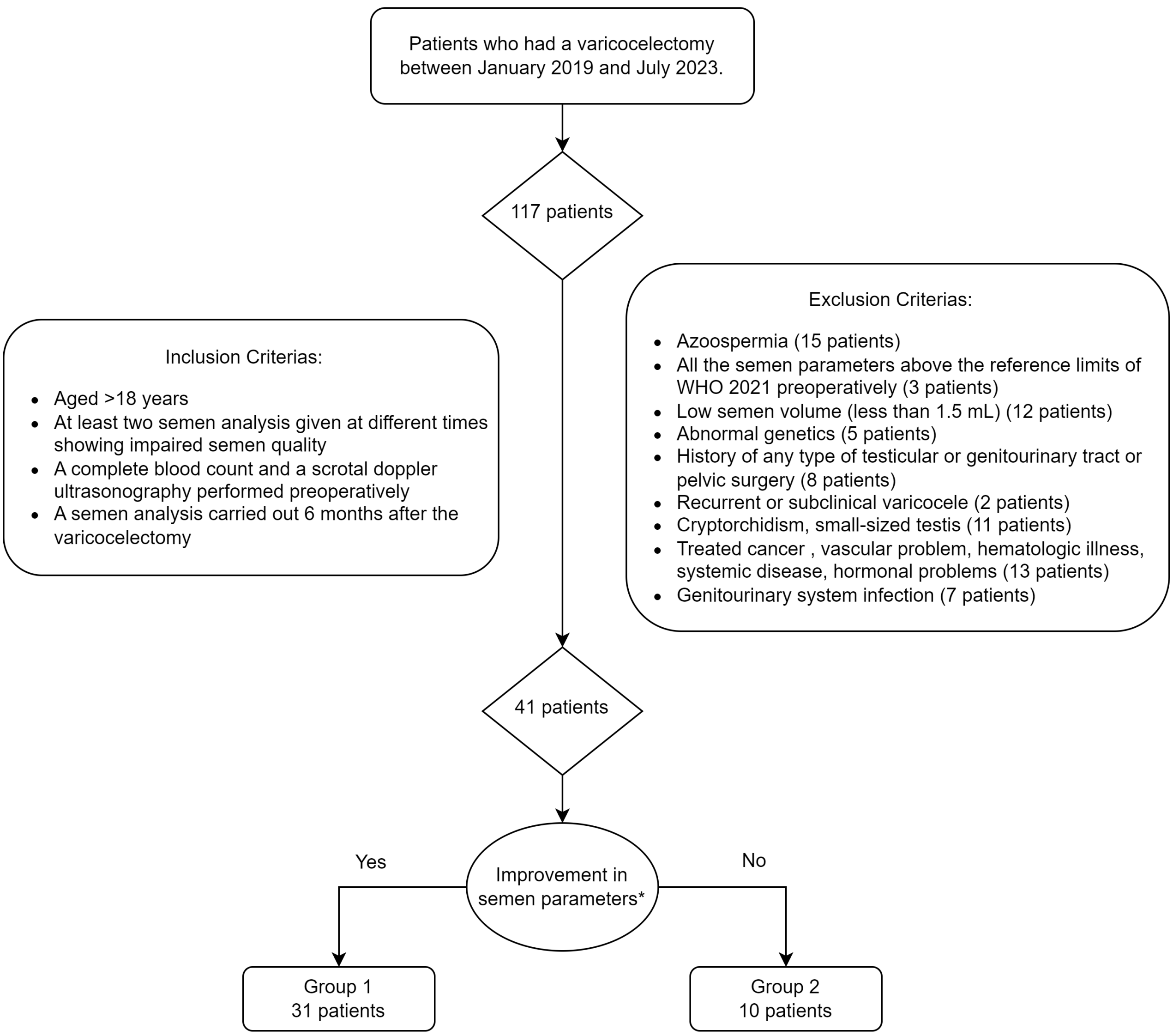

FIGURE 1. Flow chart of the study. WHO: World Health Organization. *A more than 50% increase in total progressive motile sperm count or 100% increase in a patient with a total progressive motile sperm count <5 million.

distribution width (RDW), mean platelet volume (MPV), plateletcrit (PCT), as well as the neutrophil-to-lymphocyte ratio (NLR), monocyte-to-lymphocyte ratio (MLR), and platelet-to-lymphocyte ratio (PLR). The testicular volume is automatically calculated by the ultrasound device using a formula based on the measurements of the length, width, and height of the testis. The device typically uses an ellipsoid formula to estimate the volume, where volume = (length $\times$ width $\times$ height $\times$ π)/6. Varicocele was graded according to the Dubin and Amelar classification (Grade 0: identifies subclinical varicocele (not clinically detectable but diagnosed by ultrasound), Grade 1: palpable only during the Valsalva maneuver, Grade 2: detectable even without the Valsalva maneuver, Grade 3: visible upon inspection) in the physical examination [15]. The largest vein diameter and the reversal blood flow of the pampiniform plexus of veins were measured and classified during the SDU, which was performed using a 14-megahertz probe (Siemens Acuson s1000, Germany) [16, 17]. In the evaluation of varicocele, the vein diameter is measured using Doppler ultrasound, typically during the Valsalva maneuver to enhance venous dilation. A varicocele is considered clinically significant if the diameter of the affected vein exceeds 2–3 mm. The vein measurement is taken at the level of the spermatic cord, usually at the point of maximum dilation, which is often located just above the testis or within the pampiniform plexus. These criteria are used to classify and grade varicocele severity, contributing to both diagnosis and treatment planning. To reduce bias, both clinical and SDU grading were independently performed by a single urologist and a single radiologist, each blinded to the other's assessments. Semen analyses within three months before surgery were performed according to World Health Organization (WHO) guidelines and all analyses, sperm concentration (10^6/mL), total sperm count (10^6), total motility (% motile), and morphology (% normal forms), were recorded. The total progressive motile sperm count (TMSC) ($\times 10^6$) was also obtained by multiplying the volume of the ejaculate by the sperm concentration, and by the proportion of the progressive motile sperm divided by 100 [18]. Given the timeline of data collection (from January 2018 to July 2023), we re-evaluated all sperm test results conducted prior to 2021 in accordance with the updated WHO 2021 semen analysis standards as the principles of the reproducible and reliable data based about the semen results and examination [19]. The subinguinal varicocelectomy procedures were performed with a standardized subinguinal varicocelectomy as previously described [20].

In our study, we initially evaluated 22 potential parameters to assess their contribution to predicting postoperative sperm improvement in patients undergoing varicocelectomy. Among these, five parameters were identified as important variables based on their relevance to the clinical outcomes observed in the model. These variables included lymphocyte, platelet, varicocele vein diameter, red cell distribution width (RDW), and hemoglobin levels. The selection of these parameters was made using the machine learning algorithm, which assessed their relationship to the TMSC improvement after surgery. It is important to note that these parameters were not arbitrarily chosen; they were identified because they reflect key physiological and clinical factors that can influence sperm quality and varicocele severity. For example, lymphocyte and platelet counts can be indicative of inflammatory processes, while RDW is often associated with oxidative stress, both of which have been shown to play a role in male fertility. In contrast, other semen parameters such as sperm concentration, total sperm count, total motility, and morphology, although commonly used in clinical practice to evaluate male fertility, were not selected as important features by the model. This finding highlights the potential for machine learning models to identify variables that may not align with conventional clinical thinking but could still provide valuable insights into predicting post-surgical sperm improvements. The absence of these variables as important predictors might reflect their limited ability to capture the underlying biological mechanisms that influence TMSC improvement in varicocele patients, as opposed to factors like inflammatory markers or vascular issues. As shown in Fig. 2, the model ultimately focused on preoperative TMSC as a key feature. This suggests that while conventional semen parameters are critical in clinical practice, machine learning models can help identify more nuanced predictors that might be more closely associated with the underlying mechanisms of varicocele-related infertility.

2.2 Data preprocessing

All 41 patients were regrouped according to the TMSC from the postoperative semen analyses as in D'Andrea *et al*. [21]. We used this parameter to group the patients as it has been shown to be a predictor parameter of spontaneous pregnancy rate [22], and for the outcome of intracytoplasmic sperm injection [23] in cases of male factor infertility. If a patient had more than a 50% increase in the TMSC or if a patient with a TMSC of $<$5 million had an increase of $>$100%, we defined this patient as having an improvement in semen parameters after surgery and determined these as Group 1. If a patient had less than a 50% increase in the TMSC or if a patient with a TMSC of $<$5 million but did not have an increase of $>$100%, we defined this patient as not having an improvement in semen parameters after surgery and these were classified as Group 2.

2.3 Machine learning models

The study was structured based on the principles of ML. Several ML algorithms were employed, including the Extra Trees Classifier, Light Gradient Boosting Machine (LightGBM) Classifier, eXtreme Gradient Boosting (XGBoost) Classifier, Logistic Regression (LR) and Random Forest Classifier. These models were selected due to their widespread adoption, exceptional performance in predictive tasks, and extensive application in healthcare research [24].

The Extra Trees, Random Forest, LightGBM and XGBoost represent ensemble learning methods within ML. Ensemble methods amalgamate the predictions of multiple models to enhance performance and increase robustness. Two predominant ensemble techniques are bagging and boosting. Bagging, implemented in Extra Trees and Random Forest, minimizes variance by training multiple models on different subsets of data in parallel. In contrast, boosting, applied in LightGBM and XGBoost, builds models sequentially, with each new

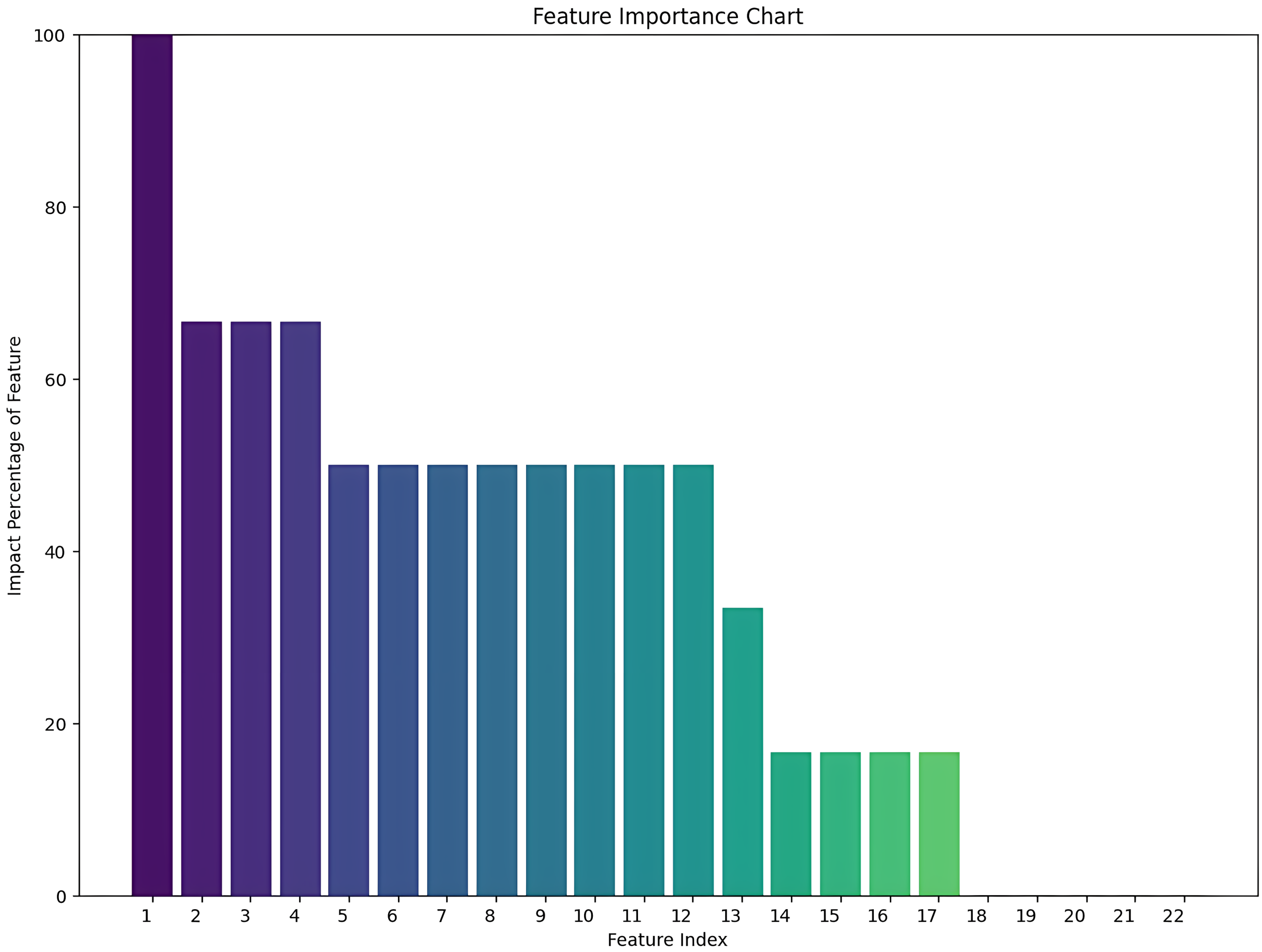

FIGURE 2. Impact factors of the variables. 1: Lymphocyte; 2: Platelet; 3: Varicocele Vein Diameter; 4: Red Cell Distribution Width; 5: Hemoglobin; 6: Monocyte; 7: Varicocele Grade; 8: Total Progressive Motile Sperm Count; 9: Complaint; 10: Neutrophil-to-Lymphocyte Ratio; 11: Platelet-to-Lymphocyte Ratio; 12: Age; 13: Plateletcrit; 14: Mean Platelet Volume; 15: Varicocele Side; 16: Neutrophil.

model correcting the errors of the previous one, thereby improving accuracy.

Logistic Regression (LR) is a supervised learning algorithm primarily used for binary classification. It models the relationship between input features and the probability of a specific outcome by fitting a logistic function (S-curve) to the data [25].

In our analysis, we considered 22 parameters as candidate explanatory variables to predict the outcomes of varicocelectomy. These parameters included hemoglobin, neutrophil count, platelet count, red cell distribution width (RDW), mean platelet volume (MPV), and several ratios such as the neutrophil-to-lymphocyte ratio (NLR), among others. As part of the preprocessing steps for the machine learning models, we applied scaling techniques to variables that did not follow a normal distribution. This is a standard practice in machine learning to ensure that all input variables are on a comparable scale and do not disproportionately affect the model due to differences in their distributions. Specifically, normalization and standardization methods were applied as needed to transform the data, allowing the model to perform more effectively and accurately. Such preprocessing techniques are essential to the performance and robustness of machine learning algorithms [26, 27].

2.4 Performance metrics

The synthetic minority oversampling technique (SMOTE) was used to create a balanced data set. After the SMOTE process, the data set was balanced by adjusting the number of patients in Group 1 to the same number as in Group 2, which was low in this study. For ML, 80% of the data was used for training and the remaining 20% for testing. In the tests conducted with these models, the models' 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. Cross-validation is a statistical method used to estimate the ability of ML models that is widely used in applied ML to compare and select a model for a given predictive modeling problem. A confusion matrix 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.

2.5 Feature selection

Independent variables that significantly affect the increase in the TMSC 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. The Permutation Feature Importance Plot is given in Fig. 2.

ML algorithm tests were performed using Python software (Python Software Foundation, Wilmington, DE, USA. Python Language Reference, version 3.5).

3. Results

Of the 117 patients who had received a varicocelectomy operation at the Eskisehir City Hospital Department of Urology between January 2019 and July 2023 only 41 patients were included in the study after applying the inclusion and exclusion criteria (Fig. 1). Of the 41 patients, 31 (75.6%) and 10 (24.4%) were classified as Group 1 or 2, respectively, according to the status of improvement in their TMSC after surgery (Table 1).

The ML algorithms used, namely the Extra Trees Classifier, LGBM Classifier, XGB Classifier, Logistic Regression and

TABLE 1. The demographic and laboratory findings of the patients included the study.

The demographic and laboratory findings	N = 41	
Age (yr) (mean ± sd)	27.17 ± 6.28	
BMI (kg/m^2) (mean ± sd)	25.79 ± 3.68	
Complaint of Patients (n, %)		
Pain	11 (26.8%)	
Infertility	23 (56.1%)	
Scrotal Swelling	7 (17.1%)	
Varicocele Side (n, %)		
Right	0 (0.0%)	
Left	35 (85.4%)	
Bilaterally	6 (14.6%)	
Varicocele Grade (n, %)		
Grade 2	20 (48.8%)	
Grade 3	21 (51.2%)	
Varicocele Vein Diameter (median, min–max)	3.7 (2.4–7)	
	Preoperative	Postoperative
Sperm concentration (10^6/mL) (median, min–max)	10.5 (0.3–72)	20.4 (0.8–106)
Total sperm number (10^6/ejaculate) (median, min–max)	33.5 (1.65–309.6)	78.0 (2.50–372.4)
Total progressive motility (%) (mean ± sd)	22.46 ± 17.24	31.27 ± 17.06
Total progressive motile sperm count (×10^6) (median, min–max)	5.695 (0–107.1)	19.200 (0–167.8)
Sperm morphology (normal forms, %) (median, min–max)	3 (1–8)	4 (1–10)
Improvement in semen parameters (n, %)		
Yes*	31 (75.6%)	
No**	10 (24.4%)	
Hemoglobin (g/dL) (mean ± sd)	15.38 ± 1.01	
Neutrophil (10^3/μL) (median, min–max)	3.650 (1.97–14.90)	
Lymphocyte (10^3/μL) (mean ± sd)	2.45 ± 0.71	
Monocyte (10^3/μL) (mean ± sd)	0.57 ± 0.17	
Platelet (10^3/μL) (mean ± sd)	238.77 ± 61.73	
RDW (%) (median, min–max)	12.200 (1.80–18.30)	
MPV (fL) (mean ± sd)	9.68 ± 1.86	
PCT	0.22 ± 0.05	
NLR (median, min–max)	1.560 (0.82–8.56)	
MLR (median, min–max)	0.234 (0.11–0.57)	
PLR (mean ± sd)	102.81 ± 32.71	

BMI: body mass index; RDW: red cell distribution width; MPV: mean platelet volume; NLR: neutrophil-to-lymphocyte ratio; MLR: monocyte-to-lymphocyte ratio; PLR: platelet-to-lymphocyte ratio; PCT: plateletcrit; yr: year; sd standard deviation; min: minimum; max: maximum.

*If a patient had more than a 50% increase in the TMSC or if a patient with a TMSC of $<$5 million had an increase of $>$100%.

**If a patient had less than a 50% increase in the TMSC or if a patient with a TMSC of $<$5 million but did not have an increase of $>$100%.

Random Forest Classifier, had an accuracy of 0.923, 0.462, 0.846, 0.769, 0.846, respectively, in the prediction of a TMSC increase after the operation. The results of the algorithms used are summarized in Table 2 and Fig. 3. The Extra Trees Classifier algorithm was found to be the best ML technique for predictions according to the accuracy rates, ROC-AUC, sensitivity, and specificity (92.3%, 97.6%, 100% and 85.7%, respectively).

4. Discussion

A varicocele and its main treatment, varicocelectomy, is still a controversial topic in andrology [5]. It has been shown that varicocele does not affect semen quality in all men, and not all men with varicocele seek infertility treatment. Although the current underlying mechanisms that could lead to a decline in semen quality have not yet been defined, a varicocelectomy is recommended for males who have clinical varicocele and abnormal sperm parameters [28, 29]. The latest two meta-analyses from the Global Andrology Forum showed that a varicocelectomy had a positive effect on the recovery of the conventional semen parameters [3, 30]. In our study, 75.6% of the patients had an increase in TMSC after surgery, similar to the results reported in the literature [31].

It is not well known how a varicocelectomy improves semen quality and which men would benefit from this procedure. The most important question in varicocele is which men would be able to recover their sperm parameters after the operation, excluding those without non-obstructive azoospermia, those without cryptozoospermia, or those with normal semen analysis who have varicocele [32]. The answer to this question is important because if the urologist could predict the success rate of surgery with a high degree of certainty, unnecessary or ineffective procedures could be avoided, and preoperative counseling, which could be more effective, could be used. There have been studies to identify the predictors of the success of the operation in the literature [32]. Nomograms have been formulated using the possible indicators for predicting semen parameter improvement [33–36] and the spontaneous pregnancy rate [36]. However, only the nomogram by Samplaski [33] had an external validation study [37] and these nomograms have not been used in urology clinical practice. Although the literature suggests that varicocelectomy should only be performed on infertile patients with abnormal sperm quality [38], there is still a lack of evidence on the predictors of the improvement in sperm quality. One of the reasons could be that all of the studies in the literature have employed traditional statistical methods to show the prediction of the parameters and these predictive methods have limitations, such as assumption, overfitting *vs.* underfitting, inappropriate hypothesis testing, and not assessing the predictive power in the analysis, all of which could have significantly affected the success of the models. It is well known that AI has advantages in detecting the different patterns between parameters and making decisions more accurately than statistical methods [39].

The ability of a urologist to predict whether sperm parameters will improve in a patient with varicocele before surgery and to counsel the patient accordingly provides several clinical benefits. Predicting the likelihood of improvement in sperm parameters allows urologists to guide patients toward the most appropriate treatment. Patients with a higher chance of benefiting from surgery can be recommended for the procedure, while those with lower chances may consider alternative treatments. Knowing whether sperm parameters will improve after surgery helps manage patient expectations. All these benefits ensure that patients are realistically informed about the potential outcomes of the surgery, and reduce the risk of disappointment as well as preventing unnecessary surgical interventions where success is unlikely. This not only protects patients from unnecessary risks but also promotes the more efficient use of healthcare resources. By providing information on whether surgery will improve sperm parameters, the urologist helps the patient make an informed decision. This transparency enhances patient satisfaction and trust in the treatment process. ML-based models can predict personalized varicocele surgery outcomes more accurately and efficiently, which is beneficial for both patients and physicians. ML models can analyze large datasets of patient information to predict the effects of varicocele surgery more accurately based on sperm parameters and they can assess individual risk factors and health conditions to offer personalized predictions about the potential benefits of surgery.

The articles about the use of AI in andrology have been reported, but the number of these articles is insufficient when compared to other medical fields [40, 41]. In andrology, several studies have leveraged unsupervised machine learning to uncover novel insights. For instance, unsupervised clustering algorithms like k-means or hierarchical clustering have been used to categorize sperm morphology or semen quality data, helping to identify distinct patient subgroups with similar characteristics, often leading to more personalized treatments [42, 43]. These techniques have been shown to identify un-

TABLE 2. Prognosis prediction results of different machine learning algorithms.

Model Name	ROC-AUC	Accuracy	Precision	Sensitivity	Specificity	Interval
XGB Classifier	0.857	0.846	1.000	0.667	1.000	0.650–1.000
Random Forest Classifier	0.929	0.846	0.750	1.000	0.714	0.650–1.000
Extra Trees Classifier	0.976	0.923	0.857	1.000	0.857	0.778–1.000
Logistic Regression	0.810	0.769	0.714	0.833	0.714	0.540–0.998
LGBM Classifier	0.500	0.462	0.462	1.000	0.000	0.191–0.733

LGBM: Light Gradient Boosting Machine; XGB: eXtreme Gradient Boosting; ROC-AUC: Receiver Operating Characteristic-Area Under Curve.

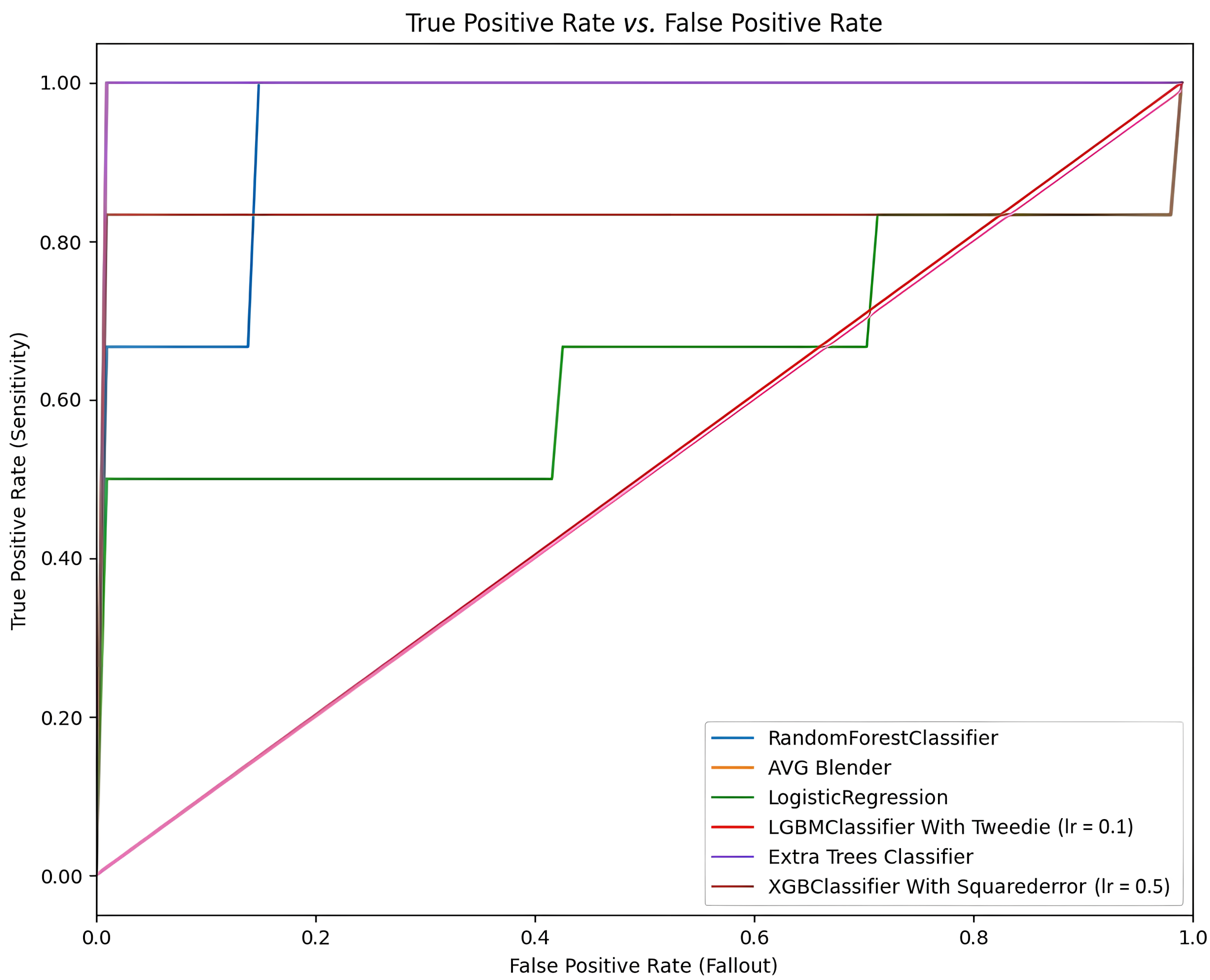

FIGURE 3. Receiver operating characteristic curve graphs. LGBM: Light Gradient Boosting Machine; XGB: eXtreme Gradient Boosting; AVG: Average.

derlying correlations in semen parameters and biomarkers that may not have been apparent using traditional methods.

There are only two studies about varicocele and AI, and only one study in the literature on an ML-based predictive model for the improvement of semen parameters after varicocelectomy, reported by Ory *et al*. [12]. The baseline follicle-stimulating hormone (FSH) and the testosterone level, together with other demographic characteristics and clinical findings, were determined in their dataset. Only the Random Forest model was used to predict an upgrade in sperm concentration, and it was found that this model accurately predicted an upgrade in 86.7% (AUC 0.72). The authors did not give any information as to why they did not choose other ML models for constructing a prediction model. The second report was on a deep learning model constructed to detect varicocele in the SDU and the researchers concluded that the efficacy of the proposed system had high accuracy for classified varicocele veins [13].

In our study, the parameters used most often in previous reports were determined in the dataset, except for hormone levels. Varicocele is known to negatively impact hormonal production, particularly by altering testosterone levels and increasing the levels of luteinizing hormone (LH) and follicle-stimulating hormone (FSH). These hormonal imbalances are thought to contribute to impaired spermatogenesis and may influence male fertility outcomes. The pathophysiological effects of varicocele, including disrupted blood flow and testicular hyperthermia, could further exacerbate these hormonal imbalances, leading to suboptimal semen quality and motility, as reflected in the machine learning model's variable selections. While hormonal imbalances associated with varicocele are known to affect spermatogenesis and male fertility outcomes, the lack of consistent hormone data meant these parameters could not be integrated into the model.

Given the growing interest in inflammatory markers in reproductive health, CBC parameters could provide additional insights into varicocelectomy outcomes. Although these factors may have a limited direct association with spermatogenesis, they might still influence the outcome of varicocelectomy through their role in the inflammatory processes linked to varicocele pathology. This approach was supported by prior research that emphasizes the relevance of inflammatory markers in the context of male infertility [44]. Inflammation is a known contributor to male infertility, and CBC parameters, including neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR), have been suggested in previous studies as potential markers of altered spermatogenic function in varicocele patients. Elevated NLR and PLR, for instance, have been associated with impaired semen quality and may reflect an underlying inflammatory response that affects fertility [45]. In our study, the following variables were identified as important predictors of varicocelectomy outcomes: lymphocyte count, platelet count, varicocele vein diameter, red cell distribution width (RDW), and hemoglobin levels. While these parameters were selected by the machine learning model, it is essential to interpret their potential medical relevance in the context of male reproductive health. The common intersection between the parameters identified in the study lies in their potential to reflect systemic health conditions such as inflammation, oxidative stress, and vascular health, all of which are crucial in male reproductive function.

Our study represents the first comprehensive effort to employ five distinct machine learning algorithms to predict the improvement in TMSC after varicocelectomy. The Extra Trees Classifier emerged as the top-performing model, achieving an accuracy of 92.3% with an AUC of 0.976. We identified several key parameters, including age, BMI, vein diameter, TMSC, platelet count, and monocyte-to-lymphocyte ratio, as significant predictors. This groundbreaking analysis underscores the potential of basic laboratory tests and semen analysis, using machine learning, to effectively identify patients who are likely to benefit from varicocelectomy, marking a novel approach in the field. This study significantly contributes to the existing literature by being the first to apply multiple machine learning algorithms to predict TMSC improvement following varicocelectomy. By identifying key parameters such as vein diameter, platelet count, and monocyte-to-lymphocyte ratio, our work demonstrates that machine learning models can be used to predict post-surgical outcomes with high accuracy. This approach not only advances the precision of patient selection for varicocelectomy but also introduces a novel method for integrating basic laboratory and semen analysis data in clinical decision-making.

This study has some limitations. We acknowledge that a sample size of 41 patients is relatively small for statistical comparison, especially when developing a machine-learning model. However, we chose to proceed with this dataset because of the importance of exploring predictive methods for counseling the patients who will or will not benefit from the procedure before surgery. While this study aimed to focus on the impact of varicocelectomy on fertility outcomes, several exclusion criteria were applied to ensure a more homogeneous study population. Conditions such as abnormal genetics and low semen volume were excluded, as their effects on varicocele outcomes are not well-established and could confound the results. The exclusion of such conditions, though reducing sample size, was necessary to minimize potential biases and ensure that the fertility concerns were primarily related to varicocele. The lack of a clear medical consensus on the effectiveness of varicocelectomy in patients with abnormal genetic factors further justified their exclusion. This approach, however, limits the generalizability of the findings to a broader patient population. We acknowledge that these exclusions may reduce the diversity of the sample and could potentially limit the external validity of the results. Future studies could explore the impact of varicocelectomy in a more diverse cohort, including individuals with these excluded conditions. While our sample size may be limited, our aim was to demonstrate the feasibility of using ML in this context and to lay the groundwork for future studies with larger and more diverse datasets. In this study, we employed supervised machine learning techniques to predict varicocelectomy outcomes. As noted in prior research, machine learning methods generally require larger sample sizes compared to traditional statistical analyses to ensure reliable and generalizable results. This is due to the complex nature of machine learning models, which rely on high-dimensional feature spaces and require sufficient data to avoid overfitting and produce robust predictions [27]. As supervised machine learning methods typically demand sample sizes that can accommodate these intricacies, we recognize the importance of ensuring adequate data to strengthen the model's predictive power. Consequently, the sample size used in this study was based on practical considerations, with a focus on balancing statistical power with available data. While large sample sizes typically lead to more robust and generalizable results, studies like those by Topol [46] have demonstrated how AI models can still deliver high accuracy in clinical applications with relatively smaller datasets, provided the data is well-processed and features are appropriately selected. We also acknowledge that larger sample sizes may further improve model accuracy and reduce variance in predictions. Future studies with a larger cohort may provide more definitive conclusions on the robustness of these findings. ML models, especially when properly tuned and validated, are capable of extracting meaningful patterns from limited data, provided that the data is of high quality and the feature set is appropriately selected. We acknowledge that the number of individuals in both groups is relatively small compared to other studies using supervised machine learning. However, several factors contributed to this decision. First, our study focused on a specific population of patients undergoing varicocelectomy, which naturally limits the sample size. Additionally, while small sample sizes can be challenging, recent literature suggests that supervised machine learning models can still perform well with limited data, especially when appropriate methods such as data augmentation, cross-validation, and regularization techniques are employed [47, 48].

In our case, we utilized the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance and enhance the dataset. While the small sample size does pose challenges, our model demonstrated robust performance metrics, including high accuracy, sensitivity, and specificity, which are often associated with reliable predictive models [49]. Furthermore, recent studies indicate that machine learning models can still provide meaningful insights from small datasets if the model is well-regularized and cross-validation techniques are used effectively [50]. We believe these strategies, along with the model's strong validation results, support the robustness and reliability of our findings.

However, external larger validation data sets are needed to boost our model's accuracy. Lifestyle behaviors that affect a man's fertility (smoking, alcohol, caffeine consumption) were not recorded during the clinic visit, so unhealthy lifestyle behaviors could not be identified as exclusion criteria. While having at least one abnormal sperm parameter has been used to classify patients in the literature, we preferred to use the TMSC as a group criterion as described previously [6, 32].

The strengths of our study are that we used more ML algorithms, conducted comparisons between the algorithms, presented all accuracy rates of the models, and used a classifier confusion matrix of the best model in terms of prediction accuracy. We employed techniques to mitigate the risks associated with a small dataset, such as cross-validation, which helps ensure that the model's performance is not overly optimistic and that it generalizes unseen data well. Furthermore, our model demonstrated robust performance metrics that we believe validate its reliability even with the small sample size. Thus, to the best of our knowledge, this study is the first prediction model to use appropriate methods based on ML for the improvement of semen parameters after varicocelectomy.

5. Conclusions

In conclusion, we recommend emphasizing the key findings, particularly the significant improvement in TMSC following varicocelectomy as predicted by machine learning algorithms. Highlight the use of variables such as vein diameter, platelet count, and monocyte-to-lymphocyte ratio in predicting these outcomes. This reinforces the novelty of the study, showing how machine learning models can integrate basic clinical parameters to guide clinical decision-making. Furthermore, suggest potential future research areas where these models can be refined or expanded for broader clinical applications.

ML could provide a decision-making support system to decide the need for surgery, predict the success rate, complications, length of hospital stay, and re-admission rate to hospital after surgery. In this pilot study, we report promising preliminary results using ML to predict the success of varicocelectomy. This study's results suggest that prospective studies with larger, more diverse patient populations, and including additional clinical parameters, such as hormonal profiles and lifestyle risk factors, would provide a clear roadmap for building upon the current findings.

AVAILABILITY OF DATA AND MATERIALS

The data presented in this study are available on reasonable request from the corresponding author.

AUTHOR CONTRIBUTIONS

MEA and CK—designed the research study. MS and AA—provided help and advice on searching data and writing manuscript. ÖÇ and MEA—analyzed the data. MS and CK—wrote the manuscript. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Permission for this study was obtained from the Eskişehir City Hospital on (27 December 2023) (Decision number: ESH/GOEK 2023/6). Because the patients remained anonymous and the clinical data were deidentified, the requirement for informed consent was waived by Eskisehir City Hospital.

ACKNOWLEDGMENT

Not applicable.

FUNDING

This research received no external funding.

CONFLICT OF INTEREST

The authors declare no conflict of interest. Coşkun Kaya is serving as one of the Editorial Board members of this journal. We declare that Coşkun Kaya had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to VMA.

REFERENCES

[1] Ghanem MA, Adawi EA, Safan MA, Ghanem AA. Outcome of varicoceles scrotal sclerotherapy in infertile patients with recurrent varicocele. International Archives of Urology and Complications. 2018; 4: 037.

[2] European Association of Urology. EAU guidelines. EAU Guidelines Office: Arnhem, the Netherlands. 2023.

[3] Agarwal A, Cannarella R, Saleh R, Boitrelle F, Gül M, Toprak T, *et al*. Impact of varicocele repair on semen parameters in infertile men: a systematic review and meta-analysis. The World Journal of Men's Health. 2023; 41: 289–310.

[4] Cannarella R, Shah R, Hamoda TAA, Boitrelle F, Saleh R, Gul M, *et al*. Does varicocele repair improve conventional semen parameters? A meta-analytic study of before-after data. The World Journal of Men's Health. 2024; 42: 92–132.

[5] Persad E, O'Loughlin CA, Kaur S, Wagner G, Matyas N, Hassler-Di Fratta MR, *et al*. Surgical or radiological treatment for varicoceles in subfertile men. Cochrane Database of Systematic Reviews. 2021; 4: CD000479.

[6] Shomarufov AB, Bozhedomov VA, Sorokin NI, Matyukhov IP, Fozilov AA, Abbosov SA, Kamalov AA. Predictors of microsurgical varicocelectomy efficacy in male infertility treatment: critical assessment and systematization. Asian Journal of Andrology. 2023; 25: 21–28.

[7] Samplaski M, Jarvi K. Prognostic factors for a favorable outcome after varicocele repair in adolescents and adults. Asian Journal of Andrology. 2016; 18: 217–221.

[8] Ghanem MA, Adawi EA, Hakami NA, Ghanem AM, Ghanem HA. The predictive value of the platelet volume parameters in evaluation of varicocelectomy outcome in infertile patients. Andrologia. 2020; 52: e13574.

[9] Eyler Dang L, Hubbard A, Dissak-Delon FN, Chichom Mefire A, Juillard C. Right population, right resources, right algorithm: Using machine learning efficiently and effectively in surgical systems where data are a limited resource. Surgery. 2021; 170: 325–328.

[10] Loftus TJ, Vlaar APJ, Hung AJ, Bihorac A, Dennis BM, Juillard C, *et al*. Executive summary of the artificial intelligence in surgery series. Surgery. 2022; 171: 1435–1439.

[11] Abid R, Hussein AA, Guru KA. Artificial intelligence in urology: current status and future perspectives. Urologic Clinics of North America. 2024; 51: 117–130.

[12] Ory J, Tradewell MB, Blankstein U, Lima TF, Nackeeran S, Gonzalez DC, *et al*. Artificial intelligence based machine learning models predict sperm parameter upgrading after varicocele repair: a multi-institutional analysis. The World Journal of Men's Health. 2022; 40: 618–626.

[13] AlZoubi O, Abu Awad M, Abdalla AM, Samrraie L. Varicocele detection in ultrasound images using deep learning. Multimedia Tools and Applications. 2024; 83: 63617–63634.

[14] Foroughi AA, Yazdanpanah E, Nazeri M, Eghbali T, Arasteh P, Ariafar A. Clinical grading and color Doppler ultrasonography-based grading of varicocele: how compatible are the two grading systems? World Journal of Urology. 2019; 37: 1461–1465.

[15] Dubin L, Amelar RD. Varicocele size and results of varicocelectomy in selected subfertile men with varicocele. Fertility and Sterility. 1970; 21: 606–609.

[16] Sarteschi MF. Andrological ultrasound. 1st edn. Athena audiovisuals: Modena. 2003.

[17] Freeman S, Bertolotto M, Richenberg J, Belfield J, Dogra V, Huang DY, *et al*. Ultrasound evaluation of varicoceles: guidelines and recommendations of the European Society of Urogenital Radiology Scrotal and Penile Imaging Working Group (ESUR-SPIWG) for detection, classification, and grading. European Radiology. 2020; 30: 11–25.

[18] Ayala C, Steinberger E, Smith DP. The influence of semen analysis parameters on the fertility potential of infertile couples. Journal of Andrology. 1996; 17: 718–725.

[19] Björndahl L, Barratt CLR, Mortimer D, Agarwal A, Aitken RJ, Alvarez JG, *et al*. Standards in semen examination: publishing reproducible and reliable data based on high-quality methodology. Human Reproduction. 2022; 37: 2497–2502.

[20] Johnson D, Sandlow J. Treatment of varicoceles: techniques and outcomes. Fertility and Sterility. 2017; 108: 378–384.

[21] D'Andrea S, Micillo A, Barbonetti A, Giordano AV, Carducci S, Mancini A, *et al*. Determination of spermatic vein reflux after varicocele repair helps to define the efficacy of treatment in improving sperm parameters of subfertile men. Journal of Endocrinological Investigation. 2017; 40: 1145–1153.

[22] Hamilton JA, Cissen M, Brandes M, Smeenk JM, de Bruin JP, Kremer JA, *et al*. Total motile sperm count: a better indicator for the severity of male factor infertility than the WHO sperm classification system. Human Reproduction. 2015; 30: 1110–1121.

[23] Borges E Jr. Total motile sperm count: a better way to rate the severity of male factor infertility? JBRA Assisted Reproduction. 2016; 20: 47–48.

[24] Bai X, Zhou Z, Luo Y, Yang H, Zhu H, Chen S, *et al*. Development and evaluation of a machine learning prediction model for small-for-gestational-age births in women exposed to radiation before pregnancy. Journal of Personalized Medicine. 2022; 12: 550.

[25] Yakar M, Etiz D, Metintas M, Ak G, Celik O. Prediction of radiation pneumonitis with machine learning in stage III lung cancer: a pilot study. Technology in Cancer Research & Treatment. 2021; 20: 15330338211016373.

[26] Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. 1st edn. Springer: New York, NY. 2009.

[27] Bishop CM, Nasrabadi NM. Pattern recognition and machine learning. 1st edn. Springer-Verlag: New York. 2006.

[28] Schlegel PN, Sigman M, Collura B, De Jonge CJ, Eisenberg ML, Lamb DJ, *et al*. Diagnosis and treatment of infertility in men: AUA/ASRM guideline part I. Fertility and Sterility. 2021; 115: 54–61.

[29] Minhas S, Bettocchi C, Boeri L, Capogrosso P, Carvalho J, Cilesiz NC, *et al*. European association of urology guidelines on male sexual and reproductive health: 2021 update on male infertility. European Urology. 2021; 80: 603–620.

[30] Abdel-Meguid TA. Predictors of sperm recovery and azoospermia relapse in men with nonobstructive azoospermia after varicocele repair. The Journal of Urology. 2012; 187: 222–226.

[31] Bozhedomov VA, Shomarufov AB, Bozhedomova GE, D OA, Kamalov DM, Sorokin NI, *et al*. Varicocele and reproductive function: pathozoospermia treatment (a prospective comparative study). Urologiia. 2021; 62–68. (In Russian)

[32] Crafa A, Cannarella R, Condorelli RA, Mongioì LM, Vignera S, Calogero AE. Predictive parameters of the efficacy of varicocele repair: a review. Asian Journal of Andrology. 2024; 26: 441–450.

[33] Samplaski MK, Yu C, Kattan MW, Lo KC, Grober ED, Zini A, *et al*. Nomograms for predicting changes in semen parameters in infertile men after varicocele repair. Fertility and Sterility. 2014; 102: 68–74.

[34] Liu X, Liu D, Pan C, Su H. Nomogram for predicting semen parameters improvement after microscopic varicocelectomy in infertile men with abnormal semen parameters. Journal of Personalized Medicine. 2022; 13: 11.

[35] Kandevani NY, Namdari F, Hamidi M, Dialameh H, Behzadi A. Developing a novel prediction model for the impact of varicocelectomy on postoperative fertility. European Journal of Translational Myology. 2022; 32: 10411.

[36] Liu L, Li J, Liu G, Pan C, Bai S, Zhan Y, *et al*. Nomogram for predicting spontaneous pregnancy after microscopic varicocelectomy in infertile men with normal hormone. BMC Pregnancy and Childbirth. 2022; 22: 791.

[37] Jang WS, Kim KH, Lim KT, Lee J, Heo JE, Kwon H, *et al*. External validation of the post-varicocele repair semen analysis nomogram to predict total motile sperm count: a multicenter study. Andrologia. 2020; 52: e13809.

[38] Schlegel PN, Sigman M, Collura B, De Jonge CJ, Eisenberg ML, Lamb DJ, *et al*. Diagnosis and treatment of infertility in men: AUA/ASRM guideline PART II. The Journal of Urology. 2021; 205: 44–51.

[39] Pimenov DY, Bustillo A, Wojciechowski S, Sharma VS, Gupta MK, Kuntoğlu M. Artificial intelligence systems for tool condition monitoring in machining: analysis and critical review. Journal of Intelligent Manufacturing. 2023; 34: 2079–2121.

[40] Ghayda RA, Cannarella R, Calogero AE, Shah R, Rambhatla A, Zohdy W, *et al*. Artificial intelligence in andrology: from semen analysis to image diagnostics. The World Journal of Men's Health. 2024; 42: 39–61.

[41] Calogero AE, Crafa A, Cannarella R, Saleh R, Shah R, Agarwal A. Artificial intelligence in andrology—fact or fiction: essential takeaway for busy clinicians. Asian Journal of Andrology. 2024; 26: 600–604.

[42] Peng T, Liao C, Ye X, Chen Z, Li X, Lan Y, *et al*. Machine learning-based clustering to identify the combined effect of the DNA fragmentation index and conventional semen parameters on *in vitro* fertilization outcomes. Reproductive Biology and Endocrinology. 2023; 21: 26.

[43] Tanaka T, Kojo K, Nagumo Y, Ikeda A, Shimizu T, Fujimoto S, *et al*. A new clustering model based on the seminal plasma/serum ratios of multiple trace element concentrations in male patients with subfertility. Reproductive Medicine and Biology. 2024; 23: e12584.

[44] Ates E, Ucar M, Keskin MZ, Gokce A. Preoperative neutrophil-to-lymphocyte ratio as a new prognostic predictor after microsurgical subinguinal varicocelectomy. Andrologia. 2019; 51: e13188.

[45] Noori Alavije H, Ahmadi-Hamedani M, Moslemi H. Evaluation of platelet indices and mean platelet volume to platelet count ratio in experimentally varicocele-induced adolescent and adult rats. Andrologia. 2022; 54: e14345.

[46] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine. 2019; 25: 44–56.

[47] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research. 2002; 16: 321–357.

[48] He H, Garcia EA. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering. 2009; 21: 1263–1284.

[49] Brownlee J. Machine learning mastery with Python: understand your data, create accurate models, and work projects end-to-end. Machine Learning Mastery: San Francisco. 2016.

[50] Khoshgoftaar TM, Hulse JV, Napolitano A. Comparing boosting and bagging techniques with noisy and imbalanced data. IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans. 2011; 41: 552–568.

How to cite this article: Coşkun Kaya, Mehmet Erhan Aydın, Özer Çelik, Aykut Aykaç, Mustafa Sungur. Developing a machine learning model for predicting varicocelectomy outcomes: a pilot study. Journal of Men's Health. 2024; 20(12): 65-74. doi: 10.22514/jomh.2024.201.