



Article

# Using Data Mining to Assist in Predicting Reproductive Outcomes Following Varicocele Embolization

Ana Paula Sousa <sup>1,2</sup>, Judith Santos-Pereira <sup>3</sup>, Maria José Freire <sup>2,4</sup>, Belmiro Parada <sup>2,4,5</sup>,  
Teresa Almeida-Santos <sup>1,2,5</sup>, Jorge Bernardino <sup>3,6</sup> and João Ramalho-Santos <sup>1,7,\*</sup>

- <sup>1</sup> Biology of Reproduction & Stem Cell Group, CNC-Center for Neuroscience and Cell Biology, CIBB, University of Coimbra, Azinhaga de Santa Comba, Celas, 3004-504 Coimbra, Portugal; anapauladesousa.apms@gmail.com (A.P.S.); anateresasantos.tas@gmail.com (T.A.-S.)
  - <sup>2</sup> Reproductive Medicine Unit, Centro Hospitalar e Universitário de Coimbra, Pr. Prof Mota Pinto, 3004-561 Coimbra, Portugal; mjfreire.urologia@gmail.com (M.J.F.); parada.belmiro@gmail.com (B.P.)
  - <sup>3</sup> Polytechnic of Coimbra, ISEC-Instituto Superior de Engenharia de Coimbra, R. Pedro Nunes, 3030-199 Coimbra, Portugal; santosj@hotmail.ca (J.S.-P.); jorge@isec.pt (J.B.)
  - <sup>4</sup> Urology Unit, Centro Hospitalar e Universitário de Coimbra, Pr. Prof Mota Pinto, 3004-561 Coimbra, Portugal
  - <sup>5</sup> Faculty of Medicine, University of Coimbra, Azinhaga de Santa Comba, Celas, 3000-548 Coimbra, Portugal
  - <sup>6</sup> Centre for Informatics and Systems, University of Coimbra, Polo II, Pinhal de Marrocos, 3030-290 Coimbra, Portugal
  - <sup>7</sup> Department of Life Sciences, University of Coimbra, Calçada Martim de Freitas, 3000-456 Coimbra, Portugal
- \* Correspondence: jramalho@uc.pt



**Citation:** Sousa, A.P.; Santos-Pereira, J.; Freire, M.J.; Parada, B.; Almeida-Santos, T.; Bernardino, J.; Ramalho-Santos, J. Using Data Mining to Assist in Predicting Reproductive Outcomes Following Varicocele Embolization. *J. Clin. Med.* **2021**, *10*, 3503. <https://doi.org/10.3390/jcm10163503>

Academic Editor: Gérard Tachdjian

Received: 7 July 2021  
Accepted: 29 July 2021  
Published: 9 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** We carried out a retrospective analysis of infertile couple data using several methodologies and data analysis techniques, including the application of a novel data mining approach for analyzing varicocele treatment outcomes. The aim of this work was to characterize embolized varicocele patients by ascertaining the improvement of some of their clinical features, predicting the success of treatment via pregnancy outcomes, and identifying data patterns that can contribute to both ongoing varicocele research and the more effective management of patients treated for varicocele. We retrospectively surveyed the data of 293 consenting couples undergoing infertility treatment with male varicocele embolization over a 10-year period, and sperm samples were collected before and at 3, 6, and 12 months after varicocele embolization treatment and analyzed with World Health Organization parameters—varicocele severity grades were assessed with medical assessment and scrotal ultrasound, patient personal information (e.g., age, lifestyle, and embolization complications) was collected with clinical inquiries, and varicocele embolization success was measured through pregnancy outcomes. Varicocele embolization significantly improved sperm concentration, motility, and morphology mean values, as well as sperm chromatin integrity. Following this study, we can predict that a male patient without a high varicocele severity grade (with grade I or II) has a 70.83% chance of conceiving after embolization treatment if his partners' age is between 24 and 33 with an accuracy of 70.59%. Furthermore, male patients successful in achieving pregnancy following embolization are mostly characterized by having a normal sperm progressive motility before treatment, a normal sperm concentration after treatment, a moderate to low varicocele severity grade, and not working in a putatively hazardous environment.

**Keywords:** varicocele; embolization; data mining; sperm parameters

## 1. Introduction

Varicocele is characterized by the dilation of the veins of the spermatic cord [1] in one (unilateral) or both (bilateral) testis. Studies estimate that varicocele is present in more than 35% of infertile couples [2]. The McGraw-Hill Concise Dictionary of Modern Medicine goes even further by stating that varicocele is present in 40% of males treated for infertility. Given that infertility affects an estimated 15% of couples globally [3] and male infertility

factors are responsible for 50% of infertility causes [2], the importance of assessing the data of patients with such a prevalent condition is clear.

It has been suggested that varicocele triggers several inflammatory pathways that negatively affect spermatogenesis [4], ultimately reflected in alterations in sperm quality [5]. The European Association of Urology (EUA) recommends that, in order to improve fertility rates, varicocele should be treated in infertile men with a clinical varicocele, abnormal semen parameters, and otherwise unexplained infertility in couples where the female partner has a good ovarian reserve [6]. For this purpose, several approaches have been suggested. One of the suggested treatments is the use of antioxidant supplementation to neutralize the oxidative stress caused by the activated inflammatory pathways. However, in addition to the existence of contradictory literature, the authors of a recent was concluded that this approach does not seem to improve either seminal parameters or pregnancy rates [7]. More effective results seem to be obtained when varicocele is treated by surgical varicocelectomy [6]. However, this is an invasive technique. Minimally invasive techniques for varicocele correction include the radiological embolization technique that introduces substances (such as coils, sclerosants, or glue) into the circulation to devitalize the enlarged veins [8]. It has been shown that embolization successfully corrects varicocele, whether with coils or glue, in an average of 92% of cases [9]. Furthermore, varicocele correction has an important impact on the treatment of infertility [10] since it has been found to lead to clinical reproductive outcome improvement of parameters, such as sperm quality and pregnancy rate, when comparing patients undergoing varicocele correction with those with the untreated condition [2,11,12].

However, despite the existing evidence, the link between varicocele and male infertility remains controversial, as it is possible to find this condition in men with normal fertility. The same is true when the benefit of varicocele correction is analyzed [13]. Due to the heterogeneity of varicocele patterns and responses to its treatment, other advanced data analysis techniques, such as the use of data mining, may improve existing knowledge linked to varicocele and identify new insights regarding this infertility treatment. To the best of our knowledge, only traditional statistics has been used when assessing the efficacy of varicocele treatment in the context of human infertility.

Data mining is the process of discovering data patterns [14] where standard statistical exploratory data analysis procedures (traditional statistics) cannot offer useful insights [15]. In this field, traditional statistics are usually viewed as the primary data analysis technique, with data mining as a putatively relevant secondary technique due to its characteristics [16]. While the groundwork of both techniques is mathematics, data mining extends its range over topics such as machine learning, database systems, and visualization, which brings important gains over traditional statistics techniques [17]. The main advantages of data mining over traditional statistics methods are its capability to analyze different types of data (e.g., numbers, names, and severity degrees) and its ability to perform inductive analysis. This latter is fundamental in cases where researchers are trying to understand, for instance, the consequences of a treatment that are not fully known since many potential variables can hinder the formulation of a hypothesis to be validated or rejected [16].

In this context, through a retrospective analysis, the aim of this work was to characterize embolized varicocele patients by ascertaining the improvement of some of their clinical features, predicting the success of treatment via pregnancy outcomes, and identifying data patterns that can contribute to both ongoing varicocele research and the more effective management of patients treated for varicocele. To tackle these goals, we used statistical techniques (traditional approach) and then applied the most commonly used data mining techniques in healthcare. Though these data mining techniques have proven applicability in healthcare [17–20], they are rarely applied together in the medical treatment domain, as shown in the meta-analysis carried out by Tomar et al. [19]. To the best of our knowledge, such an advanced data analysis technique has never been applied to embolized varicocele patients.

## 2. Materials and Methods

All patients signed informed consent forms, and all human material was used in accordance with the appropriate ethical and internal review board (IRB) guidelines provided by the Centro Hospitalar e Universitário de Coimbra (ID number CHUC-115-12).

### 2.1. Description of Population

Data analysis was retrospectively carried out on a dataset of 293 infertile couples (i.e., couples that were unable to get pregnant after 1 year of regular intercourse) where male partners had undergone varicocele embolization between January 2007 and April 2016 in a Portuguese public hospital—the Reproductive Medicine Unit of the Centro Hospitalar e Universitário de Coimbra (CHUC)—with the aim of improving their chances of conceiving. Patients that did not perform a sperm analysis before the varicocele embolization and patients without partner were excluded from the analysis. Because a large percentage of the population achieved pregnancy after varicocele correction, not all female partners were thoroughly characterized.

The classification of varicocele was assessed by medical examination by one of the two physicians involved and/or by Doppler ultrasound, according to the European Association of Urology guidelines [6]:

- Subclinical: not palpable or visible at rest or during Valsalva maneuver, but can be shown by Doppler ultrasound.
- Grade 1: palpable during Valsalva maneuver.
- Grade 2: palpable at rest.
- Grade 3: visible and palpable at rest

### 2.2. Sperm Preparation and Analyses

Patients were evaluated through a routine semen analysis before varicocele treatment, as well as 3, 6, and 12 months after varicocele embolization. Semen samples were obtained by masturbation after 3–5 days of sexual abstinence and collected in the center, and routine seminal analysis was performed by one of the six available technicians according to the World Health Organization Guidelines [21]. Among other parameters (volume, pH, presence of leucocytes, and sperm agglutination), sperm parameters such as concentration (million sperm/mL), progressive motility (percentage of progressively motile sperm), and morphology (percentage of morphologically normal sperm) were assessed. Sperm chromatin status was also evaluated using the Diff-Quik like stain set (Dade Behring Inc., Newark, NJ, USA), as previously described [22], where the percentage of sperm with damaged chromatin (i.e., either decondensed or presenting fragmented DNA) was determined.

### 2.3. Follow-Up

Further data were obtained using follow-up inquiries and medical record assessments. The occurrence of complications after varicocele embolization and global patient satisfaction with the technique were evaluated. Complications were graded according to Clavien–Dindo classification [23]. The achievement of clinical pregnancy, both spontaneously or using assisted reproductive techniques (ART), was also assessed using follow-up inquiries and medical record assessments.

### 2.4. Statistics

Using the embolized varicocele patient sample data, we first applied statistical techniques to understand the collected data and ascertain the evolution of sperm parameter mean values, as well as chromatin integrity after embolization.

All patient features were described and explored statistically as suggested by Han et al. [14] to successfully prepare the data for the mining process. In fact, these authors stated that one should analyze the central tendency of data (i.e., compute the mean, median, and mode of each attribute), as well as its dispersion (i.e., compute the minimum value

(Min), the value of the first quartile ( $Q_1$ ), the value of the third quartile ( $Q_3$ ), the maximum value (Max), and the standard deviation (SD) of each patient feature). Hence, this study computed all these statistical measures and complemented the assessment by generating graphs such as time series, box plots, and histograms with the RapidMiner data mining platform and Excel software to better understand the collected attributes.

The chi-square and the ANOVA tests were used to assess the sperm parameter values for statistical inference, and appropriate post-hoc tests were used to identify which sperm parameter mean values were statistically significantly different. The Kruskal–Wallis and Mann–Whitney statistical tests were used to assess the statistical significance between the clustered patient’s mean feature values.  $p \leq 0.05$  was considered statistically significant.

### 2.5. Data Mining

Following the traditional statistical approach, we then applied the most commonly used data mining techniques in healthcare: classification, using a decision tree algorithm; clustering, using the k-means algorithm; and association, with the FP-growth algorithm. Though these data mining techniques have proven applicability in healthcare [17–20], they are rarely applied together in the medical treatment domain, as shown in the meta-analysis carried out by Tomar et al. [19].

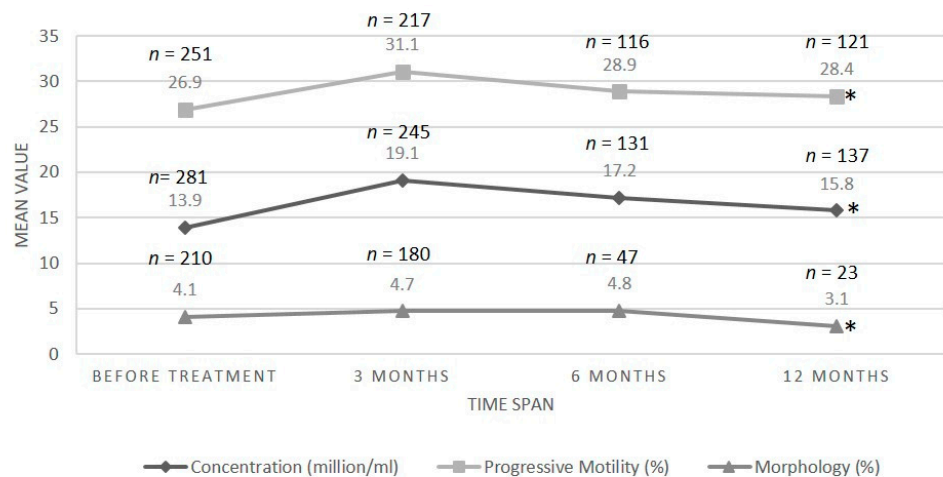
Data mining was achieved by following the cross-industry standard process for data mining project (CRISP-DM) methodology [24]. This involves a set of six phases that can be recursively executed and are briefly described as covering the following tasks:

- i Business understanding: determine the business objectives and data mining goals.
- ii Data understanding: collect, describe, and explore the data, as well as verify their quality.
- iii Data preparation: select, clean, construct, integrate, and format the previously understood data.
- iv Modeling: select data mining techniques, define the training/testing design, build the data mining models, and assess them with performance measures.
- v Evaluation: assess data models with respect to business objectives and domain expertise.
- vi Deployment: disclosure of results.

In the modeling phase, the selected data mining techniques were the most common in the healthcare domain (i.e., classification, clustering, and association), as previously described. Thus, with the use of the RapidMiner data mining platform, we applied the following algorithms: decision tree algorithm for the classification technique, k-means algorithm for the clustering technique, and FP-growth algorithm for the association technique.

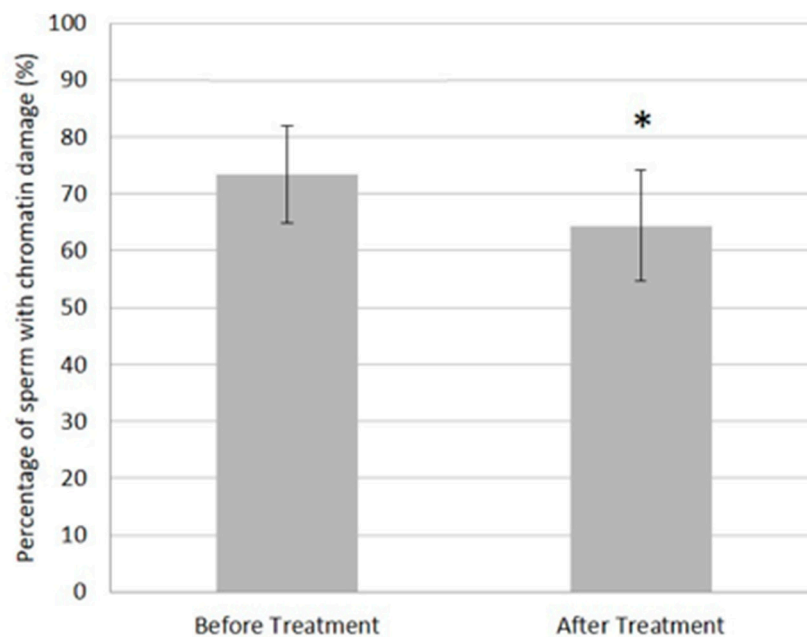
## 3. Results

Though there was significant inter-patient variability, semen analysis revealed that, before embolization, the population of patients with varicocele had values of sperm concentration ( $13.9 \pm 23.6$  million sperm/mL) and progressive motility ( $26.9 \pm 22.5\%$ ) below the WHO [21] reference values (Figure 1). Similarly, sperm chromatin integrity was also affected ( $73.4 \pm 8.5\%$  of sperm with damaged chromatin; Figure 2). More importantly, sperm parameter mean values (Figure 1) improved after embolization ( $p = 0.00006$ ,  $p = 0.003$ , and  $p = 0.05$  at 3 months after embolization for sperm concentration, motility, and morphology, respectively). However, sperm concentration was the only parameter that remained significantly improved (ANOVA  $p = 0.017$ ) a year after varicocele correction. Chromatin integrity (Figure 2) also improved after varicocele embolization ( $p = 0.043$ ), but it was still higher ( $64.4 \pm 9.7\%$ ) than the reference value for normality ( $<32\%$ ) that was previously described using the same methodology [22].



\*  $p < 0.05$  with the ANOVA test

**Figure 1.** Evolution of sperm parameter mean values according to patient follow-up time. Sperm concentration, progressive motility, and morphology were assessed according to WHO criteria [21].



**Figure 2.** Evolution of sperm chromatin integrity. Sperm chromatin integrity was monitored before and after varicocele embolization, as described in the Methods section. \*  $p = 0.043$ . Sperm chromatin integrity was assessed as previously described [22].  $n = 86$ .

Out of the 230 couples that we were able to determine a pregnancy outcome, 107 (i.e., 46.52%) got pregnant; of these, 54.21% (58/107) conceived with an ART procedure, 38.31% (41/107) conceived spontaneously. The remaining 7.48% (8/107) conceived both naturally and via ART, considering at least two different pregnancies. Interestingly, almost 80% of the pregnancies were obtained in the first year after embolization. During the time of the study, 78.50% (84/107) of couples had a single pregnancy, 17.75% (19/107) had two pregnancies, and 3.74% (4/107) had three pregnancies after the embolization treatment.

Importantly, patients that were able to conceive had a statistically significantly better sperm progressive motility before the embolization, with a mean difference of 6.9 (ANOVA  $p = 0.018$ ); better sperm morphology 3 months after the embolization, with a mean difference of 1.6 (ANOVA  $p = 0.004$ ); and increased sperm concentration 6 months after the embolization, with a mean difference of 8.2 (ANOVA  $p = 0.015$ ) (Table 1).

**Table 1.** Pregnancy outcome vs sperm parameter mean values through patient's follow up times.

Sperm Parameter		Pregnancy Outcome	<i>n</i>	Mean	Mean Difference	SD	
Concentration	Before treatment	Yes	107	14.5	−0.4	21.6	
		No	119	14.9		27.2	
	3 months	Yes	94	22.9	4.8	24.2	
		No	107	18.1		27.5	
	6 months **	Yes	50	22.9	8.2	29.5	
		No	65	14.7		20.6	
	12 months	Yes	60	18.8	4.3	19.5	
		No	69	14.5		18.2	
	Progressive Motility	Before treatment **	Yes	102	29.9	6.9	23.3
			No	107	23.0		20.8
		3 months	Yes	92	33.2	3.3	21.6
			No	92	29.9		21.1
6 months		Yes	49	33.5	6.2	25.1	
		No	57	27.3		25.6	
12 months		Yes	58	30.8	4.1	23.6	
		No	58	26.7		22.8	
Morphology		Before treatment	Yes	89	4.0	0.4	5.0
			No	87	3.6		3.5
		3 months ***	Yes	78	5.5	1.6	5.0
			No	74	3.9		3.4
	6 months	Yes	22	5.0	0.5	3.2	
		No	19	4.5		5.2	
	12 months ***	Yes	10	4.0	1.1	3.3	
		No	12	2.9		1.4	

\*\* ANOVA  $p < 0.05$ ; \*\*\* ANOVA  $p < 0.01$ .

The patient features that were significantly related with pregnancy outcome were: male partner age (ANOVA  $p = 0.018$ ), varicocele severity grade (chi-square  $p = 0.049$ ), sperm concentration at 6 months following embolization (ANOVA  $p = 0.015$ ), sperm progressive motility before treatment (ANOVA  $p = 0.018$ ), sperm morphology at 3 months (ANOVA  $p = 0.004$ ), sperm concentration normality at 3 months following embolization (chi-square  $p = 0.017$ ), sperm progressive motility normality before treatment (chi-square  $p = 0.027$ ), sperm progressive motility normality at 3 months following embolization (chi-square  $p = 0.022$ ), semen quality classification before treatment (chi-square  $p = 0.017$ ), semen quality classification at 3 months following embolization (chi-square  $p = 0.018$ ), and a potentially hazardous occupation when considering possible exposure to known toxicants (chi-square  $p = 0.023$ ).

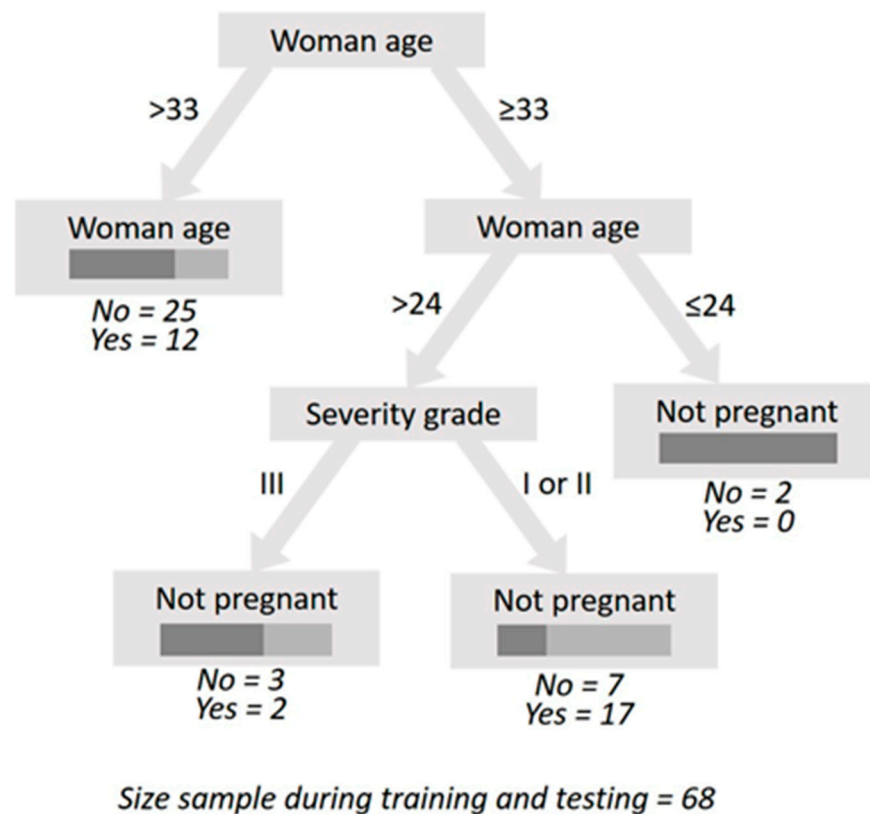
Varicocele embolization success, defined as whether the man's partner got pregnant, was able to be predicted with an accuracy of 70.59% by stating that a male patient without a high varicocele severity grade (with grade I or II) has a 70.83% chance (i.e.,  $(17 / (7 + 17)) = 70.83\%$ ) of conceiving after the embolization treatment if his partner is between the ages of 24 and 33 years old inclusively (Figure 3). The results in Figure 3 suggest that male patients with a female partner aged 24 or younger are still not able to conceive. This suggests that the cause for unsuccessful pregnancy following embolization in these

cases might involve female infertility, since this decision branch does not entail a male patient feature, as is the case when female partner age is above 33.

**Table 2.** Decision tree’s confusion matrix.

Actual	Predicted	
	Yes	No
Yes	7	1
No	3	6

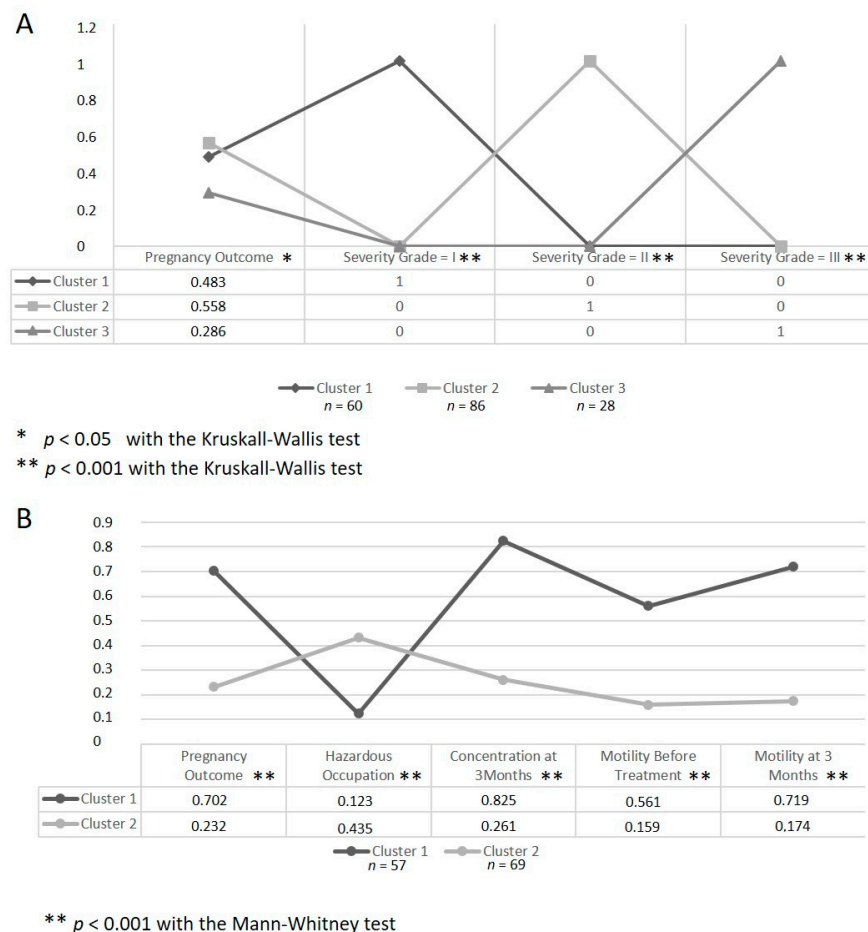
The target variable (pregnancy) has two values: positive (yes) or negative (no). The columns represent the actual values of the target variable. The rows represent the predicted values of the target variable.



**Figure 3.** Decision tree for the prediction of the varicocele embolization success. The test design entailed the sub-sampling of 85 patients: 48 patients for training; 20 patients for testing, and 17 patients for validating the decision tree. This computed decision tree was obtained with the following parameter values after exhaustive training/testing: sampling type during training/testing = linear; decision tree splitting criterion = accuracy; decision tree pruning = true; decision tree minimal size for split = 4; decision tree minimal gain = 0.1; decision tree minimal leaf size = 2; and decision tree maximal depth = 20. During its training/testing, the decision tree computed the following performance measures: accuracy = 80.00%; precision = 85.71%; recall = 66.67%; F-measure = 75.00%; and AUC = 0.717. During its validation, the performance measures were: accuracy = 70.59%; precision = 66.67%; recall = 75.00%; F-measure = 70.59%; and AUC = 0.750 with the confusion matrix shown in Table 2.

By applying the k-means algorithm to clusters of patients with similar and well-defined characteristics (namely the 174 patients where both pregnancy outcomes and severity grade were available (Figure 4A) and the 126 patients where simultaneous information on pregnancy outcomes, putative hazardous occupation, and the normality of the statistically significant sperm parameter values was available (Figure 4B)), we were able to identify these data patterns for successfully embolized patients (Figure 4). Please note that since the k-means algorithm groups patients that are similar to each other in

clusters and all patient features were normalized/mapped, the mean value of each cluster per each feature can be interpreted as the percentage of patients with a positive event. Indeed, in this analysis, the value of 1 was attributed to positive events and the value of 0 was attributed to negative events; hence, a positive pregnancy outcome, a specific severity grade, a patient working in a putative hazardous environment, and a patient having a sperm parameter value above the World Health Organization threshold all imply a value of 1 [21]. Specifically, regarding data pattern identification, for the 174 patients shown in Figure 4A, the algorithm selected three clusters showing that patients with a mild to medium varicocele severity grade are significantly (Kruskal–Wallis  $p = 0.041$ ) more prone to conceive than the ones with a high severity grade (i.e., severity grade I = 48.3% pregnancy and severity grade II = 55.8% pregnancy vs severity grade III = a clearly lower chance of 28.6% in terms of chances of conceiving). Furthermore, the data of 126 patients in Figure 4B show that patients with most successful embolization rarely work in putative hazardous occupations (12.3% vs. 43.5%) and are more likely to have a normal sperm concentration (82.5% vs. 26.1%) and normal progressive motility (71.9% vs. 17.4%) at 3 months after varicocele correction with a statistically significant difference (Mann–Whitney  $p = 0.0009$ ).



**Figure 4.** Data patterns related with the pregnancy outcome: (A) implication of the varicocele severity grade; (B) implication of a putative hazardous occupation and the normality of sperm parameter values.

Interestingly, although putative contact with hazardous environments seems to have an effect, other male patient lifestyle habits, such as drinking or smoking, or a previous disease or surgery did not seem to influence the success of treatment.

Furthermore, all couples that were able to spontaneously conceive had at least one of the sperm parameter categorized as normal or a previously diagnosed moderate varicocele



condition at 3 months following embolization. In fact, the FP-growth algorithm computed the following rules:

1. Severity grade = II  $\rightarrow$  spontaneous pregnancy = yes ( $p \leq 0.10$   $n = 107$ ) support = 0.252, confidence = 0.562, lift = 1.228, conviction = 1.239, chi-square = 3.82
2. Sperm concentration at 3 months following embolization  $\geq 15 \rightarrow$  spontaneous pregnancy = yes ( $p > 0.10$   $n = 107$ ) support = 0.262, confidence = 0.519, lift = 1.132, conviction = 1.126, chi-square = 1.61
3. Progressive motility at 3 months  $\geq 32 \rightarrow$  spontaneous pregnancy = yes ( $p \leq 0.05$   $n = 107$ ) support = 0.262, confidence = 0.560, lift = 1.223, conviction = 1.232, chi-square = 3.95
4. Morphology at 3 months  $\geq 4 \rightarrow$  spontaneous pregnancy = yes ( $p \leq 0.10$   $n = 107$ ) support = 0.224, confidence = 0.517, lift = 1.248, conviction = 1.265, chi-square = 3.58

However, based on the obtained chi-square values, unlike other sperm parameters, a normal sperm concentration at 3 months following embolization was not significantly related to spontaneous conception despite the lift measure indicating that both events are codependent (i.e., lift > 1).

Finally, regarding the feedback from the embolized patients, from the 292 responses obtained after varicocele embolization treatment, 270 patients (92.47%) had not experienced any complications. The complications reported in the 22 remaining patients were all classified as Grade I according to Clavien–Dindo classification [23], where 13 out of the 22 patients reported to have experienced some pain. Moreover, most patients (93.29%) reported that they would redo the procedure.

#### 4. Discussion

Varicocele has been pointed out as one of the main causes of male infertility. Though the exact mechanism that leads to sperm impairments is not totally known, classical sperm parameters are impaired and sperm chromatin integrity is compromised, and it was recently demonstrated that varicocele leads to oxidative stress-induced sperm DNA damage [25]. However, in an era that has seen a preponderance of ICSI (intracytoplasmic sperm injection) in terms of ART procedures, the benefit of varicocele correction has been questioned. In this study, we have seen that varicocele embolization improves all sperm parameter mean values, namely (and in the long run) sperm concentration. Accordingly, a recent analysis [9] showed that all its reviewed studies reported improved sperm parameters after varicocele embolization. Similarly, the benefit of varicocele repair in sperm DNA damage has also been reported [26].

This improvement has been suggested as important towards achieving pregnancy. After treatment, almost 47% of the patients obtained a pregnancy, either spontaneously or through ART, and 80% of them did so in the first year after embolization. Similarly, others have reported that varicocele repair increased pregnancy and live birth rates [12]. Reproductive outcomes are associated with not only sperm parameters found before treatment but also improvements obtained after embolization, as we demonstrated using data mining techniques. Furthermore, other parameters, such as the age of the female partner, varicocele grade, and a possible hazardous occupation (revealed by decision tree and/or the cluster analysis) seem to also have important roles. Though several studies have analyzed the association between varicocele/varicocele treatment and several relevant parameters, including sperm parameters [6], but their putative predictor values have never been assessed.

However, despite having achieved interesting conclusions in our opinion, this study had some clear limitations. The number of patients that ended up being selected by the data mining algorithm was relatively small. However, after analyzing the literature only 30 clinical studies on varicocele embolization were found (reviewed in [9]). Reviewing the published information our dataset is actually larger than other related works in this field, which have analyzed, on average, 117 patients ( $\pm 102$  patients). In this study, because both spontaneous pregnancy and pregnancy after ART were included, we decided not to

exclude any female fertility factors, as ART overcomes almost all these factors. However, we are aware that this may have hindered the analysis of the results. Furthermore, as this was a retrospective study, no control group (e.g., placebo, antioxidant treatment, or varicocelelectomy) was included. Therefore, a larger prospective study should be conducted in order to confirm our findings.

## 5. Conclusions

The combined use of statistical and data mining techniques can contribute towards finding patterns associated with varicocele, since several parameters may be influencing the success of its treatment and thus muddle analyses. In fact, although several studies on this condition have been carried out, none of them have assessed a varicocele database with a non-hypothetic-deductive technique such as data mining in order to identify novel patient features that can positively influence the success of varicocele correction and, hence, contribute to the improvement of male infertility treatments. To our knowledge, this is the first such study.

We were able to identify interesting data patterns with different algorithms (k-means and FP-growth), which suggests that the success of varicocele embolization, evaluated through the achievement of a pregnancy/live birth, is positively influenced by the presence of a normal sperm progressive motility before treatment, a normal sperm concentration after treatment, a moderate to low varicocele severity grade, a young male patient age, and a male patient occupation that does not involve contact with putatively toxic environments or products.

Considering the results, and although ICSI is always a therapeutic possibility in cases of severe male infertility, it is important to treat the male factor in order to increase the chances of conceiving, either spontaneously or by ART, especially when treatment is possible with a simple, minimally invasive, and patient-friendly approach.

**Author Contributions:** A.P.S. obtained all semen parameter data, J.S.-P. and J.B. provided the data mining tools. B.P., M.J.F. and T.A.-S. provided all surgical and clinical parameters. A.P.S. and J.S.-P. ran analyses on the dataset. A.P.S., J.S.-P., T.A.-S. and J.R.-S. analyzed the data. A.P.S., J.S.-P. and J.R.-S. wrote the paper, which all authors approved. All authors have read and agreed to the published version of the manuscript.

**Funding:** Part of this work was financed by the European Regional Development Fund (ERDF) through the COMPETE 2020-Operational Programme for Competitiveness and Internationalisation and Portuguese national funds via FCT-Fundação para a Ciência e a Tecnologia under the project UID/NEU/04539/2019. Additional funding was provided by the ERDF through the Centro 2020 Regional Operational Programme: project CENTRO-01-0145-FEDER-000012-HealthyAging2020, the STEM@REST Project (CENTRO-01-0145-FEDER-028871), and also Merck, SA (CSD006003). The funding entities had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the Centro Hospitalar e Universitário de Coimbra (CHUC-115-12, date of approval: 2014/03/31).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** All available data are included in the manuscript.

**Acknowledgments:** We appreciate all the helpful discussion and assistance given by all members of the Biology of Reproduction & Stem Cell Group of the CNC-Center for Neuroscience and Cell Biology, University of Coimbra, Portugal.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Arif, C.; Kotoulas, K.; Georgellis, C.; Frigkas, K.; Bantis, A.; Patris, E. Two Case Reports of Varicocele Rupture during Sexual Intercourse and Review of the Literature. *Case Rep. Urol.* **2018**, *2018*, 4068174. [[CrossRef](#)] [[PubMed](#)]
2. Kirby, E.W.; Wiener, L.E.; Rajanahally, S.; Crowell, K.; Coward, R.M. Undergoing varicocele repair before assisted reproduction improves pregnancy rate and live birth rate in azoospermic and oligospermic men with a varicocele: A systematic review and meta-analysis. *Fertil. Steril.* **2016**, *106*, 1338–1343. [[CrossRef](#)] [[PubMed](#)]
3. Agarwal, A.; Mulgund, A.; Hamada, A.; Chyatte, M.R. A unique view on male infertility around the globe. *Reprod. Biol. Endocrinol.* **2015**, *13*, 37. [[CrossRef](#)] [[PubMed](#)]
4. Panner Selvam, M.; Agarwal, A.; Baskaran, S. Proteomic analysis of seminal plasma from bilateral varicocele patients indicates an oxidative state and increased inflammatory response. *Asian J. Androl.* **2019**, *21*, 544. [[PubMed](#)]
5. Majzoub, A.; Esteves, S.C.; Gosálvez, J.; Agarwal, A. Specialized sperm function tests in varicocele and the future of andrology laboratory. *Asian J. Androl.* **2016**, *18*, 205–212. [[CrossRef](#)] [[PubMed](#)]
6. Salonia, A.; Bettocchi, C.; Boeri, L.; Capogrosso, P.; Carvalho, J.; Cilesiz, N.C.; Cocci, A.; Corona, G.; Dimitropoulos, K.; Gül, M.; et al. European Association of Urology Guidelines on Sexual and Reproductive Health—2021 Update: Male Sexual Dysfunction. *Eur. Urol.* **2021**, S0302-2838(21)01813-3. [[CrossRef](#)]
7. Pyrgidis, N.; Sokolakis, I.; Palapelas, V.; Tishukov, M.; Mykoniatis, I.; Symeonidis, E.N.; Zachariou, A.; Kaltsas, A.; Sofikitis, N.; Hatzichristodoulou, G.; et al. The Effect of Antioxidant Supplementation on Operated or Non-Operated Varicocele-Associated Infertility: A Systematic Review and Meta-Analysis. *Antioxidants* **2021**, *10*, 1067. [[CrossRef](#)] [[PubMed](#)]
8. Stegman, J.K. *Medical Dictionary for the Health Professions and Nursing*; Lippincott Williams & Wilkins: Baltimore, MD, USA, 2012.
9. Makris, G.C.; Efthymiou, E.; Little, M.; Boardman, P.; Anthony, S.; Uberoi, R.; Tapping, C. Safety and effectiveness of the different types of embolic materials for the treatment of testicular varicoceles: A systematic review. *Br. J. Radiol.* **2018**, *91*, 20170445. [[CrossRef](#)] [[PubMed](#)]
10. Samplaski, M.K.; Lo, K.C.; Grober, E.D.; Zini, A.; Jarvi, K.A. Varicolectomy to “upgrade” semen quality to allow couples to use less invasive forms of assisted reproductive technology. *Fertil. Steril.* **2017**, *108*, 609–612. [[CrossRef](#)] [[PubMed](#)]
11. Hamada, A.; Esteves, S.C.; Agarwal, A. Insight into oxidative stress in varicocele-associated male infertility: Part 2. *Nat. Rev. Urol.* **2013**, *10*, 26–37. [[CrossRef](#)] [[PubMed](#)]
12. Birowo, P.; Tendi, W.; Widyahening, I.S.; Atmoko, W.; Rasyid, N. The benefits of varicocele repair for achieving pregnancy in male infertility: A systematic review and meta-analysis. *Heliyon* **2020**, *6*, e05439. [[CrossRef](#)] [[PubMed](#)]
13. Su, J.S.; Farber, N.J.; Vij, S.C. Pathophysiology and treatment options of varicocele: An overview. *Andrologia* **2021**, *53*, e13576. Available online: <https://onlinelibrary.wiley.com/doi/10.1111/and.13576> (accessed on 27 July 2021). [[CrossRef](#)] [[PubMed](#)]
14. Han, J.; Pei, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Elsevier: Amsterdam, The Netherlands; Waltham, MA, USA, 2011.
15. Hand, D.J.; Blunt, G.; Kelly, M.G.; Adams, N.M. Data Mining for Fun and Profit. *Stat. Sci.* **2000**, *15*, 111–126.
16. Hand, D.J. Data Mining: Statistics and More? *Am. Stat.* **1998**, *52*, 112–118.
17. Tekieh, M.H.; Raahemi, B. Importance of Data Mining in Healthcare: A Survey. In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Paris, France, 25–28 August 2015; Association for Computing Machinery: New York, NY, USA; 2015; pp. 1057–1062. [[CrossRef](#)]
18. Ahmad, P.; Qamar, S.; Rizvi, S.Q.A. Techniques of Data Mining In Healthcare: A Review. *Int. J. Comput. Appl.* **2015**, *120*, 38–50. [[CrossRef](#)]
19. Tomar, D.; Agarwal, S. A survey on Data Mining approaches for Healthcare. *Int. J. Bio-Sci. Bio-Technol.* **2013**, *5*, 241–266. [[CrossRef](#)]
20. Pereira, A.; Sousa, A.P.; Ramalho-Santos, J.; Bernardino, J. Open Source Data Mining Tools Evaluation using OSSpal Methodology. In Proceedings of the 15th International Joint Conference on e-Business and Telecommunications (ICETE 2018), Porto, Portugal, 26–28 July 2018; pp. 706–712.
21. World Health Organization. *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 5th ed.; World Health Organization: Geneva, Switzerland, 2010.
22. Sousa, A.P.M.; Tavares, R.S.; Velez de la Calle, J.F.; Figueiredo, H.; Almeida, V.; Almeida-Santos, T.; Ramalho-Santos, J. Dual use of Diff-Quik-like stains for the simultaneous evaluation of human sperm morphology and chromatin status. *Hum. Reprod. Oxf. Engl.* **2009**, *24*, 28–36. [[CrossRef](#)] [[PubMed](#)]
23. Dindo, D.; Demartines, N.; Clavien, P.-A. Classification of Surgical Complications: A New Proposal With Evaluation in a Cohort of 6336 Patients and Results of a Survey. *Ann. Surg.* **2004**, *240*, 205–213. [[CrossRef](#)] [[PubMed](#)]
24. Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C.; Wirth, R. *Step-by-Step Data Mining Guide*; SPSS Inc.: Chicago, IL, USA, 2000.
25. Jeremias, J.T.; Belardin, L.B.; Okada, F.K.; Antoniassi, M.P.; Fraietta, R.; Bertolla, R.P.; Intasqui, P. Oxidative origin of sperm DNA fragmentation in the adult varicocele. *Int. Braz. J. Urol. Off. J. Braz. Soc. Urol.* **2021**, *47*, 275–283. [[CrossRef](#)] [[PubMed](#)]
26. Roque, M.; Esteves, S.C. Effect of varicocele repair on sperm DNA fragmentation: A review. *Int. Urol. Nephrol.* **2018**, *50*, 583–603. [[CrossRef](#)]