Prediction of Unsuccessful Endometrial Ablation: Random Forest vs Logistic Regression

Kelly Yvonne Roger Stevens (kyr.stevens@gmail.com)
Catharina Ziekenhuis https://orcid.org/0000-0003-3512-9181

Liesbet Lagaert
Catharina Hospital: Catharina Ziekenhuis

Tom Bakkes
University of Technology Eindhoven: Technische Universiteit Eindhoven

Malou Evi Gelderblom
Catharina Hospital: Catharina Ziekenhuis

Saskia Houterman
Catharina Hospital: Catharina Ziekenhuis

Tanja Gijsen
Elkerliek Hospital: Elkerliek Ziekenhuis

Benedictus Schoot
Catharina Hospital: Catharina Ziekenhuis

Original Article

Keywords: Endometrial ablation, Machine Learning, Random Forest

Posted Date: December 22nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-124914/v1

Prediction of unsuccessful endometrial ablation: Random Forest vs Logistic Regression

Stevens, Kelly Yvonne Roger[1,5]; Lagaert, Liesbet; [1,5] Bakkes, Tom[2]; Gelderblom, Malou Evi[1]; Houterman, Saskia[3]; Gijsen, Tanja[4]; Schoot, Benedictus[1,5].

[1] *Department of Obstetrics and Gynaecology, Catharina Hospital, Eindhoven, the Netherlands*

[2] *Department of Department of Electrical Engineering, Biomedical Diagnostics lab, TU Eindhoven, the Netherlands*

[3] *Department of Education and Research, Catharina Hospital, Eindhoven, the Netherlands*

[4] *Department of Obstetrics and Gynaecology, Elkerliek Hospital, Helmond, the Netherlands*

[5] *Women's Clinic, Ghent University Hospital, Ghent, Belgium*

Email:

K.Y.R. Stevens: Kyr.stevens@gmail.com

L. Lagaert: Liesbet.lagaert@gmail.com

T Bakkes: t.h.g.f.bakkes@tue.nl

M.E. Gelderblom Malou@gelderblom.info

S. Houterman: Saskia.houterman@catharinaziekenhuis.nl

T. Gijsen tgijsen@elkerliek.nl

B.C. Schoot: Dick@schoot.com

Corresponding author contact information:

K.Y.R. Stevens, Department of Obstetrics and Gynaecology, Catharina Hospital,

Present address: Michelangelolaan 2, 5623 EJ Eindhoven, the Netherlands
Email: kyr.stevens@gmail.com
Telephone number: 0031620962469

ORCID: 0000-0003-3512-9181

L. Lagaert, Department of Obstetrics and Gynaecology, Catharina Hospital,

Present address: Michelangelolaan 2, 5623 EJ Eindhoven, the Netherlands

Email: Liesbet.lagaert@gmail.com

Telephone number: 0032494854176

Partly presented as abstract at:

- The 28th Annual International Congress of the European Society of Gynaecological Endoscopy, Thessaloniki, Greece, 2019, Oral (preliminary results)
- The 48th Global congress of the American Association of Gynaecologic Laparoscopists, Vancouver, Canada, 2019, Poster presentation (preliminary results)

Abstract

Background Five percent of premenopausal women experience abnormal uterine bleeding. Endometrial ablation (EA) is one of the treatment options for this common problem. However, this technique shows a decrease in patient satisfaction and treatment efficacy on the long term.

Study Objective: To develop a prediction model to predict surgical re-intervention (for example re-ablation or hysterectomy) within two years after endometrial ablation (EA) by using Machine Learning (ML). The performance of the developed prediction model was compared with a previously published multivariate logistic regression model (LR).

Design This retrospective cohort study, with a minimal follow up time of two years, included 446 pre-menopausal women (18+) that underwent an EA for complaints of heavy menstrual bleeding. The performance of the ML - and the LR model was compared using the area under the Receiving Operating Characteristic (ROC) curve.

Results: We found out that the ML model (AUC of 0.65 (95% CI 0.56-0.74)) is not superior compared to the LR model (AUC of 0.71 (95% CI 0.64-0.78)) in predicting the outcome of surgical re-intervention within two years after EA.

Conclusion Although Machine Learning techniques are gaining popularity in development of clinical prediction tools, this study shows that ML is not necessarily superior to the traditional statistical LR techniques. The performance of a prediction model is influenced by the sample size, the number of features of a dataset, hyperparameter tuning and the linearity of associations. Both techniques should be considered when developing a clinical prediction model.

Key words: Endometrial ablation, Machine Learning, Random Forest

Article

Introduction

Five percent of premenopausal women has complaints of abnormal uterine bleeding (1). Endometrial ablation (EA) is one of the treatment options for this common complaint. Due to the low costs and less invasive nature of this procedure (lower intra-operative complication risks, shorter recovery time, and lower post-operative morbidity), this form of treatment seems to be a less-invasive surgical treatment for menorrhagia compared to hysterectomy (2–6). However, long-term follow up shows a decrease in patient satisfaction and treatment efficacy. Due to permanent relief, the more invasive hysterectomy remains the most effective treatment of abnormal uterine bleeding (7–14).

According to literature, several factors prior to endometrial ablation appear to have an influence on the success-rate of this procedure. Younger age, complaints of dysmenorrhea, multiparity, a thicker pre-procedural endometrium, a duration of menstruation above seven days, presence of an intramural leiomyoma on transvaginal sonography, a history of sterilization or caesarean section, and a longer uterine depth are some of the possible negative influencing factors (1,2,8,9,11–18).

To optimize the counselling of patients with abnormal uterine bleeding, a prediction model based on the combined influence of the above-mentioned predictors could provide a better insight into the individual prognosis of endometrial ablation. In times of personalised medicine this can create better individual care leading to fewer re-interventions, lower healthcare costs and more patient satisfaction. With the use of a prediction model shared decision making can be optimized (19).

For this reason Stevens et al. (16) developed two multivariate prediction models to help counsel patients for failure of EA and for surgical re-intervention within two years after EA. The developed prediction models have a clinically acceptable c-index of 0.68 and 0.71 respectively. In addition, Stevens et al. is performing an external validation of these models, results of these data will follow.

In the field of gynaecology, many prediction models are developed using statistic multivariate logistic regression as a standard approach, these are based on a combination of various predictors that are significantly related to the outcome of interest. However, this method cannot automatically estimate the interconnection between predictors and in this way can overestimate the influence of an individual predictor (20,21).

We were also interested in other techniques of developing a prediction model. In recent years Machine Learning (ML) methods have been increasingly used in the development of clinical prediction models. ML is a scientific discipline that focuses on models that directly and automatically learn from data without using pre-identified statistical parameters and without assumption of a preconceived relationship between predictors and outcomes (20,22). A potential advantage of Machine Learning methods compared to the traditional statistical strategies is the possibility of capturing complex, nonlinear relationships in the data (23,24). ML algorithms use training data with well-defined input and output variables. This gives the opportunity to define a model with predictors which can be used for new and similar data. Compared to statistical logistic regression models, this can be done without a priori assumption of relevant variables (25).   We chose surgical re-intervention as most objective outcome measure to compare both prediction models in predicting unsuccessful endometrial ablation.

The aim of the study was to develop a Machine Learning model to predict the chance of surgical re-intervention (for example re-ablation or hysterectomy) within two years after EA. Furthermore, we compared the performance of the ML model with the prediction by the previously published multivariate logistic regression re-intervention model of Stevens et al (16).

Methods:

This study used the same dataset as was used to develop the prediction models in the study from Stevens et al. , the full study protocol can be consulted there. (16)

This retrospective two-centred cohort study, performed in two non-university teaching hospitals in the Netherlands (Catharina Hospital, Eindhoven; Elkerliek Hospital Helmond), included 446 patients who have had an EA for complaints of abnormal uterine bleeding (16). Both hospitals used similar ablation techniques between 2004 and 2013, being Cavatherm® (Veldana Medical SA, Morges, Switzerland), Gynecare Thermachoice® (Ethicon, Sommerville, US) and Thermablate® EAS (Idoman, Ireland). Recent publications have shown that these ablation techniques were equally effective (14,26). Local medical ethical review boards approved the study. All patients gave informed consent.

Patients were identified in the Electronic Patient care System by using specified search terms related to endometrial ablation. Exclusion criteria were a postmenopausal status at time of EA; (suspicion of) endometrial malignancy or uterine cavity deformations (adenomyosis; anomalies; fibroids; or a polyp). Follow-up period after treatment was at least two years. This time-interval was chosen because previous literature stated that most re-interventions were done within two years. Follow-up ended on the day of hysterectomy, in case of death or on April 15, 2015 (9,17,18,26–28).

Data were extracted from individual patient files by two researchers (K.S. & D.M. (16)). Next, patients were asked to fill in a questionnaire regarding follow-up information. In case of non-response, patients were contacted by letter and ultimately by telephone by the authors of Stevens et al (16). The used questionnaire contained questions based on significant variables predicting surgical re-intervention after EA that were previously published (2,5,8,11–17, 31,32).

The entire dataset consists of 446 patients with different categorical and continuous variables. For the Machine Learning algorithms all features were extracted from the original dataset of Stevens et al. (16) A total of five pre-operative variables was used to develop the Machine Learning model. This were the preoperative variables that were significant predictors in the final multivariate re-intervention model of Stevens et al. (age, duration of menstruation, dysmenorrhea, parity and previous caesarean section) (16). The continuous data were not discretized into categories as was done in the development of the previously published logistic regression model(16).

Development of the Logistic regression model

Statistical analysis of the data was performed by using SPSS 21.0 for Windows (IBM Corp., Armonk, NY, USA).

To determine which variables were significant, univariable logistic regression analysis was used. The variables with a p-value <.10 were used in the multivariable analysis. This was followed by a backward stepwise manual selection process, progressively excluding the variable with the highest p-value (16).

As described by Steyerberg et al., the p-value of 0.10 was used to prevent a potential incorrect exclusion of a predictive factor. This would be far more detrimental for the test than missing a potential discriminating factor (31,32).

Multicollinearity and interaction between the significant variables in the model was tested. Bootstrap resampling was used for internal validation (n=5000) (32,33). To correct for over-optimism of the model, regression coefficients were multiplied by the calculated shrinkage factor. A detailed description of the development of the LR-model van be found in the study of Stevens et al. (16).

Development of the Machine Learning model (Random Forest model)

For the development of the Machine Learning model, we used a Random Forest (RF) technique. This is a Machine Learning method used for classification and regression, which operates by constructing a large ensemble of decision trees on training data (22,23,34). Each tree in the Random Forest is built using a bootstrap sample randomly drawn from a training dataset. This results in a reduction of variance and corrects for a single decision trees ability to overfit to a training set. Each tree in the forest gives an individual prediction on the outcome measure. For a classification problem (in this case, surgical re-intervention or no surgical re-intervention after EA) the final Random Forest model averages the prediction of all the trees in the forest (21,23,34,35).

Making the model, we first trained a RF model using the five following pre-operative predictors: age, duration of menstruation, dysmenorrhea, parity and previous caesarean section. These factors were associated with a higher probability of surgical re-intervention within two years after EA in the previously published multivariate logistic regression model (16).

As described above, a RF model is an ensemble of many decision tree models. When building decision trees, each tree in the forest uses random samples (patients) from the training set ("tree bagging"). Figure 1 shows an example of an individual decision tree in the Random Forest. The decision tree is a flowchart-like binary branch structure. At each 'node split' in the tree the data are divided in two, based on the value of variable of the decision node. If no more splits are possible a prediction will be calculated for the cases in the final leaf node (23,34,36).

At each node split a random subset of features (such as duration of menstruation and parity) is considered ("feature bagging"), this to avoid over-selection of strong predictive features, leading to similar splits in the trees. This finally leads to a robust model and prevents model overfitting (21,23,34–37).

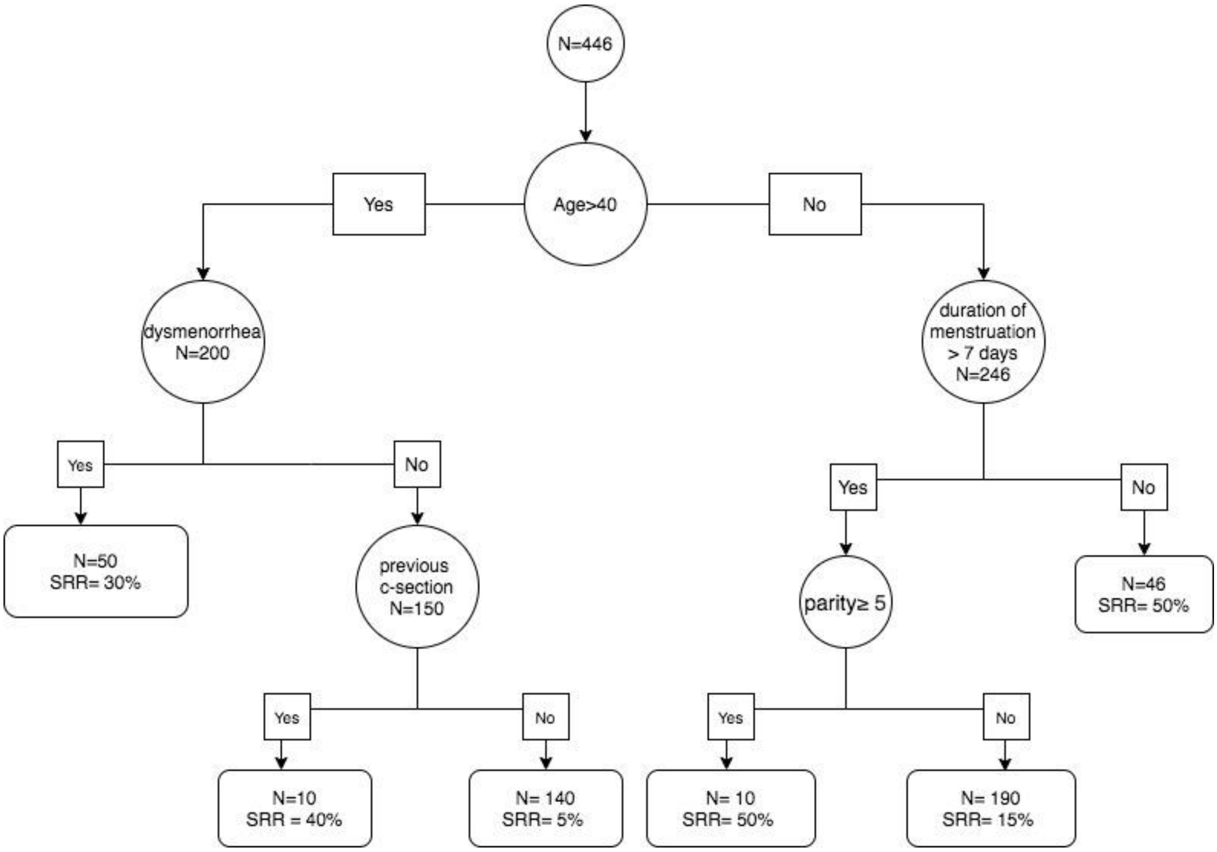

Figure 1. An illustration of a decision tree in the Random Forest model. The decision tree directs each case from the root node to the leaf nodes, resulting in a prediction.

N= Number, SRR = Surgical Re-intervention Rate.

Following this process, the classification result of a RF model is produced by computing a large ensemble of those trees and averaging the prediction of each single decision tree on surgical re-intervention. Figure 2 shows a simplified example of the RF model. In practice, the decision trees and the resulting prediction model contain a large number of leaf nodes(34,38).

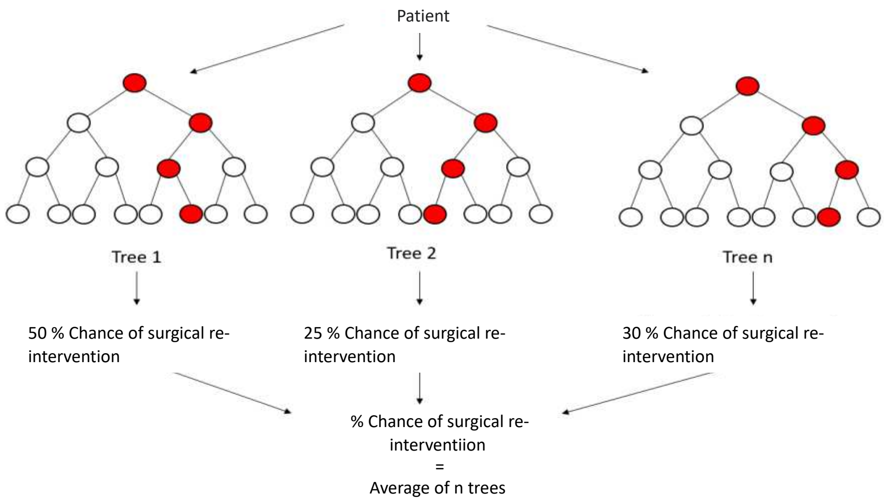

Figure 2: A simplified Random Forest model for the prediction of the surgical re-intervention .

The RF was trained in MATLAB (2018b) using the TreeBagger function in the Statistics and Machine Learning Toolbox.

We began running the RF module with default parameter values before starting to improve the RF`s performance by hyperparameter optimization. Default parameters are pre-set values for the hyperparameters on which the construction of the decision trees is based, for example 500 for ntree (number of trees in the forest) (34,35). Hyper-parameter optimization refers to the automatic optimization of the hyper-parameters of a ML model. Hyper-parameters are all the parameters of a model that are used to configure the model (e.g. minimum leaf size, number of splits, ntree and mtry, which are the number of features randomly selected as candidate feature at each split “feature bagging”).

To predict the chance of surgical re-intervention within two years after EA, the model was initially trained and internally validated on the 446 cases. To make a good comparison between de RF and LR the same validation technique was used. Therefore, a bootstrap resampling of 5000 was used to make training bags and test bags. The performance measure Area Under The Receiver Operating Curve (AUROC) was calculated on the test sets and averaged for the 5000 bootstrap samples.

Comparison of the prediction models

The performance of the models was tested and compared using the AUROC. Accuracy was not used as performance measure, since the database is unbalanced (ratio between re-intervention and no re-intervention 1:8 (53:446)) (43). It was chosen to use the performance measures (AUC) as used in the previous study of Stevens et al (16). In this way a good comparison can be made.

Predictors of surgical re-intervention: Variable importance measure (VIM)

To identify important predictors of surgical re-intervention we used two methods for analysis. First, a statistical univariate logistic regression analysis was applied to assess the importance of each variable. For each variable an odds ratio (OR) with a 95% confidence interval (CI) was calculated. Secondly, a permutation-based variable importance was used. This VIM is based on AUC statistic of the RF model. The AUC statistic is computed by randomly permutating the values of predictor x, and comparing the resulting AUC to the not permutated AUC. Leaving out an important feature will result in a lower AUC of the RF model, while leaving out an unimportant feature will not change the AUC significantly (23,38,41).

Results

Seven hundred sixty-two patients were identified retrospectively. Thirty-three patients were excluded, thirty did not meet the inclusion criteria and three underwent an incomplete endometrium ablation. The remaining 729 patients were contacted, resulting in a response-rate of 61% (N = 446). A total amount of 446 patients was available for analysis (16).

Fifty-three (11.9%) of these patients required a surgical re-intervention within two years after EA. Patients mean age during their EA was 43.8 years (SD ±5.5, range 20-55, missing values 0). The mean number of parity was 2.2 (SD ± 1.0, missing values 0). Sixty-one (13.7%) of the patients underwent a caesarean section. The mean number of previous caesarean section was 0.2 (SD ± 0.6, missing values 0)

Hundred sixty-nine (39.4%) of the patients had a menstruation period longer than seven days, the mean number of menstrual days was 9.4 (SD ± 6.0, missing values 17). Two hundred fifty-six (57.4%) of the patients had complaints of dysmenorrhea and four hundred thirty-four (97.3%) of the patients had complaints of abnormal uterine bleeding (16).

<u>Prediction models:</u>

<u>Logistic regression model</u>

Univariate analysis showed six significant predictors, multivariate analyses resulted in a logistic regression model consisting of five significant predictors: age (OR 0.95, 95% CI 0.90 – 1.00), duration of menstruation >7 days (OR 2.05, 95% CI 1.10 – 3.82), dysmenorrhea (OR 2.48, 95% CI 1.21 – 5.07), parity ≥5 (OR 7.63, 95% CI 1.51 – 38.46), and previous caesarean section (OR 2.21, 95% CI 1.05 – 4.64). The AUC of the final prediction model after correcting by the shrinkage factor was 0.71 (95% CI 0.64-0.78) (Figure 4).

The final model is described in the article of Stevens et al (16).

Random forest model

The Random Forest method resulted in a model which predicts the chance of re-intervention within two years after EA with an AUC of 0.63 (95% CI 0.54-0.71). An AUC of 0.65 (95% CI 0.56-0.74) was achieved after optimization of this model (Figure 4).

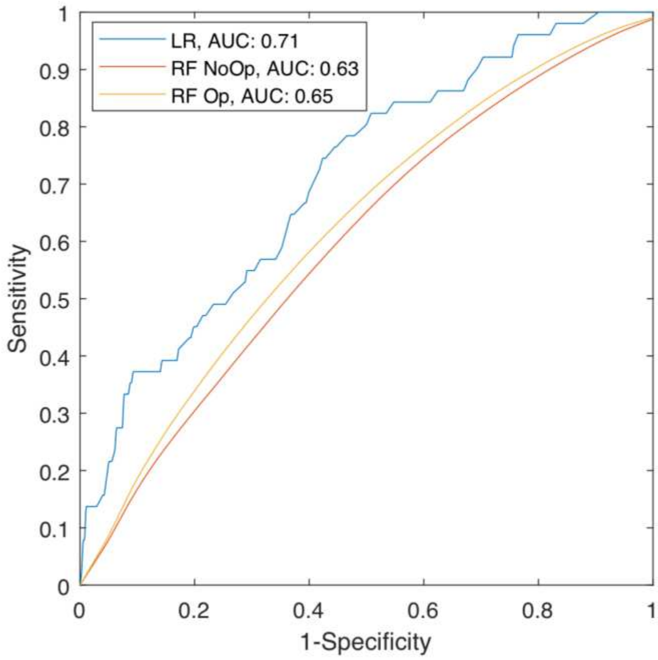

Fig4. ROC-curve of the logistic regression and Random Forest model. LR AUC 0.71 (95% CI 0.64-0.78). , NoOp AUC 0.63 (0.54-0.71), Op AUC: (0.56 – 0.74)

LR= logistic regression, RF= Random Forest, Op= after hyperparameter optimization, NoOp= before hyperparameter optimization

Predictors of surgical re-intervention: Variable importance

The AUC was used to quantify the importance of the predictor. For each RF model, the AUC was calculated on the test set. Then the same was done after permuting each predictive variable. By calculating the difference between the permuted and non-permuted AUC, the importance of each individual predictor can be quantified (Figure 5). The difference in AUC for the different predictors in the optimized model were in ascending order of importance: 0.005 for parity, 0.017 for previous caesarean section, 0.019 for age, 0.026 for dysmenorrhea and 0.051 for duration of menstruation.

This means dysmenorrhea and duration of menstruation have the highest impact on the AUC of the RF model. (Figure 5)

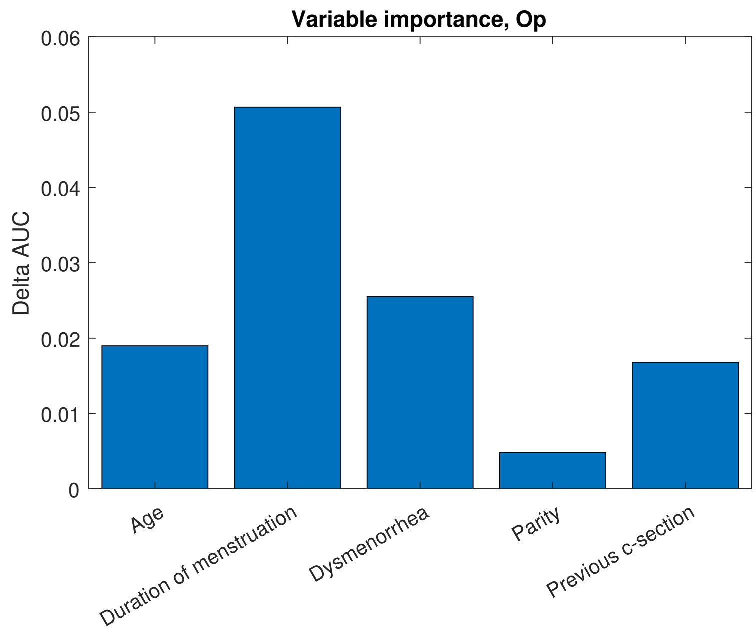

Figure. 5. Contribution of predictors of surgical re-intervention within 2 years after endometrial ablation, after hyperparameter optimization

Discussion

<u>Main findings</u>

In this study, a ML model was made using Random Forest technique to predict surgical re-intervention within two years after EA. Comparison of the predictive performance of the RF model with the existing logistic regression model of Stevens et al. was made (16).

The existing logistic regression model has a C-index of 0.71 (95% CI 0.64-0.78) (16). The ML model, developed in this study, shows a C-index of 0.65 (95% CI 0.56-0.74) after hyperparameter optimization. This shows that the LR prediction model developed by Stevens et al. (16) probably performs better in predicting surgical re-intervention within two years after EA than the newly developed RF model. However, this difference in performance is not statistically significant when we look at the confidence intervals.

In the LR model, high parity (≥5) is a predictive variable for surgical re-intervention. This can be related to the larger uterine cavity of grand multiparous women. However, when considering our RF model, parity has no large impact on the AUC. This is in line with previously reported studies that show no significant increased risk of treatment failure with increasing parity (1,15).

Previous caesarean section is also related to higher rates of surgical re-intervention which can be explained by irregularity of the uterine wall caused by the uterine scar (44). This can inhibit complete contact of the ablation device with the uterine wall, leading to residual active endometrium.

In our cohort, pre-operative dysmenorrhea is associated with a higher risk of surgical re-intervention. There is evidence that gynaecologic pathology causing this dysmenorrhea (adenomyosis and endometriosis) reduces the success of endometrial ablation (8,17,30,45,46). This can be explained by the fact that EA is not an appropriate treatment for these diseases due to the superficial effect of energy to the uterine wall of ablation. It could help to diagnose these diseases before performance of EA. However, sensitivity and specificity of the diagnostic tools for determining these diseases in the pre-operative setting are still low (47).

In line with previous studies, we found that younger age was associated with a higher risk of surgical re-intervention (7,9–13,29).

The duration of menstruation > 7 days is also a negative predictive factor for surgical re-intervention after EA. This may be caused by a thicker endometrium which is more difficult to completely remove by the device (7,10).

Interpretation in light of other evidence

There are several possible reasons to understand why the LR model probably performs better compared to the ML model.

Firstly, ML tends to work better for variables with strong predictive power (20,48). We observed that most of the candidate predictors in this model have low predictive power. The variables parity, age and previous c-section show low predictive power. The difference in area under the curve for these predictors that was produced using the permutation based variable importance was <0.02. There are different reasons to explain that this specific dataset, and its separate and combined predictors appeared to have a low predictive power. On one hand, the outcome can be unpredictable, meaning these candidate predictors have little influence on the outcome measure. On the other hand, the dataset can be too small to identify the predictive power of a candidate predictor. A larger dataset could possibly identify more predictors (20,48).

Secondly, some studies demonstrate that ML is performing better when a larger set of potential predictors are used in the prediction model. There seems to be an influence of the number of predictors (p) and the ratio of p:n (sample size). RF tends to perform better for increasing p and p:n. (20,24,49,50) In our study, to limit potential bias, the five identical predictors as published before (16) were considered for the LR and RF algorithms. We did this to allow a fair comparison between the two models, probably in disadvantage of the RF model (20,24,49,50).

Another possible reason for a lower AUC of the RF model is the necessity of big datasets to reach an optimal performance. A dataset with 446 participants might be too small for ML to make robust conclusions. For LR however, this number of patients can be enough to develop a prediction model.

Besides that, we didn`t discretize the continuous variables in the RF model, we found some literature suggesting discretization of variables can improve the classification performance (51).

Finally, we can also consider that for this clinical problem a logistic approach is better than a RF model for modelling the relationship between surgical re-intervention and the explanatory variables.

Probably the previously mentioned complex, nonlinear relationships that a ML approach can better capture are not present in this dataset.

Strengths and limitations

The predictors obtained by univariate and multivariate logistic regression are in accordance with the existing literature (1,8,10–15,17,51). However, when we compare the variable importance between the OR (LR) and the difference in AUC (ML) of each variable, we identify a different ranking in variable importance.

The difference in ranking of variable importance is a limitation of the study because there is no proper way to compare the importance of each predictor on surgical re-intervention between the RF and LR model. For the LR model the OR is defined for each predictor X as the odds of a surgical re-intervention in participants having predictor X over participants not having predictor X (Beta). While for the RF model the variable importance is defined as the difference in AUC when predictor X is not permuted.

Dysmenorrhea (OR 2.48) and a parity>5 (OR 7.63) have the highest odds ratio in the multivariate LR analysis, while for the RF model the duration of menstruation and dysmenorrhea are the most important variables. We consider two possible reason for the difference in importance. The first reason is that for the LR model all continuous variables (except age) were discretized, while for the RF model continuous variables were handled. A second reason is that in the LR the predictors have different units, and these were not standardized. This means that a subjective assessment of variable importance cannot easily be made by simply comparing the raw sizes of the OR (21,23,34,48). This can be seen as a strength of our study since the difference in AUC for each predictor (permuted vs. not permuted) reflects the variable importance in a standardized way.

We used bootstrap resampling for internal validation (n=5000) in the LR and RF model. Using the same validation method limits potential bias.
Furthermore, the same predictors were considered for the LR and ML algorithms. This limits potential bias, but will limit the potential power of a RF technique as well.
Another important strength of this study is the use of all participants in evaluating the performance of the RF model. By using the test sets, there is no need for an independent validation dataset.

It could be seen as a limitation of this study that we did not perform an external validation in another cohort. However, we did not expect it to be significantly better in performance, since the internal validation of the RF did not perform better than the logistic regression model. In addition, an external validation for the logistic regression model is being performed at the time of this study.

Finally, we can state that ML models are in our experience not easily implemented in the clinical practice; since these are often not available in commonly used software packages in clinical practice. However, future structured data-registration is increasing, which makes it easier to create big datasets available for ML-programs.

Conclusion:

In conclusion we can state that for the prediction of surgical re-intervention within two years after EA, the logistic regression model gives a better prediction compared to the Machine Learning model. However, Machine Learning algorithms should always be considered as candidate prediction tool for classification or regression problems because of the possible advantages. So far there is no evidence for one single algorithm that outperforms the other in general use. Further research is needed for the evaluation of Machine Learning based predictive modelling.

Declarations

Disclosure of interests: There are no conflicts of interest to disclose.

Contribution to authorship

K.Y.R. Stevens:	Project development, Data collection/management, Data analysis, Manuscript writing/editing
L. Lagaert:	Project development, Data collection/management, Data analysis, Manuscript writing/editing
T. Bakkes:	Development of Random Forest Model (Machine learning)
M. Gelderblom:	Manuscript editing
S. Houterman:	Manuscript editing
T. Gijsen:	Data collection, Manuscript editing
B.C. Schoot:	Project development, Data collection, Manuscript editing

Details of ethics approval

All methods were carried out in accordance with relevant guidelines and regulations. The data collection was done in the first study (development of LR model) performed by Stevens et al (16). This study was approved by the local medical ethical review board of Catharina hospital and Elkerliek hospital. All patients gave informed consent. For this second study (using the same data), the ethical board in the Catharina hospital and in the Elkerliek hospital concluded this ethics approval was valid.

Availability of data and material:

The datasets generated and analysed during the current study are not publicly available due to privacy, but they are available from the corresponding author on a reasonable request.

Funding: None

Acknowledgements:

The authors want to thank the patients for completing the questionnaires and for consenting to participate in our study.

Consent for publication: Not applicable

Availability of data and materials:

The datasets generated and analyzed during the current study are not publicly available due to privacy, but they are available from the corresponding author on a reasonable request.

Literature references

1. Peeters JAH, Penninx JPM, Mol BW, Bongers MY. Prognostic factors for the success of endometrial ablation in the treatment of menorrhagia with special reference to previous cesarean section. Eur J Obstet Gynecol Reprod Biol [Internet]. 2013 Mar [cited 2018 Dec 3];167(1):100–3. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0301211512005301

2. Waddell G, Pelletier J, Desindes S, Anku-Bertholet C, Blouin S, Thibodeau D. Effect of Endometrial Ablation on Premenstrual Symptoms. J Minim Invasive Gynecol [Internet]. 2015 May [cited 2018 Dec 3];22(4):631–6. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1553465015000886

3. Laberge P, Leyland N, Murji A, Fortin C, Martyn P, Vilos G, et al. Endometrial Ablation in the Management of Abnormal Uterine Bleeding. J Obstet Gynaecol Canada. 2015;

4. Bouzari Z, Yazdani S, Azimi S, Delavar MA. Thermal balloon endometrial ablation in the treatment of heavy menstrual bleeding. Med Arch (Sarajevo, Bosnia Herzegovina). 2014;68(6):411–3.

5. Miller J, Troeger KA, Lenhart GM, Bonafede M, Basinski CM, Lukes AS. Cost effectiveness of endometrial ablation with the NovaSure® system versus other global ablation modalities and hysterectomy for treatment of abnormal uterine bleeding: US commercial and Medicaid payer perspectives. Int J Womens Health [Internet]. 2015 Jan [cited 2018 Dec 3];59. Available from: http://www.dovepress.com/cost-effectiveness-of-endometrial-ablation-with-the-novasurereg-system-peer-reviewed-article-IJWH

6. Angioni S, Pontis A, Nappi L, Sedda F, Sorrentino F, Litta P, et al. Endometrial ablation: First-vs. second-generation techniques. Minerva Ginecologica. 2016.

7. El-Nashar SA, Hopkins MR, Creedon DJ, St Sauver JL, Weaver AL, McGree ME, et al. Prediction of treatment outcomes after global endometrial ablation. Obstet Gynecol [Internet]. 2009 Jan [cited 2018 Dec 3];113(1):97–106. Available from: https://insights.ovid.com/crossref?an=00006250-200901000-00016

8. Wishall KM, Price J, Pereira N, Butts SM, Della Badia CR. Postablation risk factors for pain and subsequent hysterectomy. Obstet Gynecol [Internet]. 2014 Nov [cited 2018 Dec 3];124(5):904–10. Available from: https://insights.ovid.com/crossref?an=00006250-201411000-00007

9. Thomassee MS, Curlin H, Yunker A, Anderson TL. Predicting pelvic pain after endometrial ablation: which preoperative patient characteristics are associated? J Minim Invasive Gynecol [Internet]. 2013 Sep [cited 2018 Dec 3];20(5):642–7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1553465013001957

10. Bongers MY, Mol BWJ, Brölmann HAM. Prognostic factors for the success of thermal balloon ablation in the treatment of menorrhagia. Obstet Gynecol [Internet]. 2002 Jun [cited 2018 Dec 3];99(6):1060–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12052600

11. Longinotti MK, Jacobson GF, Hung Y-Y, Learman LA. Probability of hysterectomy after endometrial ablation. Obstet Gynecol [Internet]. 2008 Dec [cited 2018 Dec 3];112(6):1214–20. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00006250-200812000-00006

12. Shaamash AH, Sayed EH. Prediction of successful menorrhagia treatment after thermal balloon endometrial ablation. J Obstet Gynaecol Res [Internet]. 2004 Jun [cited 2018 Dec 3];30(3):210–6. Available from: http://doi.wiley.com/10.1111/j.1447-0756.2004.00189.x

13. Klebanoff J, Makai GE, Patel NR, Hoffman MK. Incidence and predictors of failed second-

generation endometrial ablation. Gynecol Surg [Internet]. 2017 Dec 15 [cited 2018 Dec 3];14(1):26. Available from: https://gynecolsurg.springeropen.com/articles/10.1186/s10397-017-1030-4

14. Louie M, Wright K, Siedhoff MT. The case against endometrial ablation for treatment of heavy menstrual bleeding. Curr Opin Obstet Gynecol [Internet]. 2018 Aug [cited 2018 Dec 3];30(4):287–92. Available from: http://insights.ovid.com/crossref?an=00001703-900000000-99318

15. Lybol C, van der Coelen S, Hamelink A, Bartelink LR, Nieboer TE. Predictors of Long-Term NovaSure Endometrial Ablation Failure. J Minim Invasive Gynecol. 2018;

16. Stevens KYR, Meulenbroeks D, Houterman S, Gijsen T, Weyers S, Schoot BC. Prediction of unsuccessful endometrial ablation: a retrospective study. Gynecol Surg [Internet]. 2019;16(1):7. Available from: https://doi.org/10.1186/s10397-019-1060-1

17. Shavell VI, Diamond MP, Senter JP, Kruger ML, Johns DA. Hysterectomy subsequent to endometrial ablation. J Minim Invasive Gynecol [Internet]. 2012 Jul [cited 2018 Dec 3];19(4):459–64. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1553465012001161

18. Kreider SE, Starcher R, Hoppe J, Nelson K, Salas N. Endometrial ablation: is tubal ligation a risk factor for hysterectomy. J Minim Invasive Gynecol [Internet]. 2013 Sep [cited 2018 Dec 3];20(5):616–9. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1553465013001581

19. van Montfort P, Smits LJM, van Dooren IMA, Lemmens SMP, Zelis M, Zwaan IM, et al. Implementing a Preeclampsia Prediction Model in Obstetrics: Cutoff Determination and Health Care Professionals' Adherence. Med Decis Making. 2020 Jan;40(1):81–9.

20. Evangelia christodoulou, Jie MA, Collins GS, Steyerberg EW, Verbakel JY, van Calster B. A

systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;

21. Breiman L. Statistical Modeling: The Two Cultures. Stat Sci. 2001;

22. Deo RC. Machine learning in medicine. Circulation. 2015;

23. Couronné R, Probst P, Boulesteix AL. Random forest versus logistic regression: A large-scale benchmark experiment. BMC Bioinformatics. 2018;

24. Chen JH, Asch SM. Machine Learning and Prediction in Medicine — Beyond the Peak of Inflated Expectations. N Engl J Med. 2017;

25. Panesar SS, D’Souza RN, Yeh FC, Fernandez-Miranda JC. Machine Learning Versus Logistic Regression Methods for 2-Year Mortality Prognostication in a Small, Heterogeneous Glioma Database. World Neurosurg X. 2019;

26. Sambrook AM, Bain C, Parkin DE, Cooper KG. A randomised comparison of microwave endometrial ablation with transcervical resection of the endometrium: Follow up at a minimum of 10 years. BJOG An Int J Obstet Gynaecol. 2009;

27. Herman MC, Penninx JPM, Mol BW, Bongers MY. Ten-year follow-up of a randomized controlled trial comparing bipolar endometrial ablation with balloon ablation for heavy menstrual bleeding. Obstetrical and Gynecological Survey. 2014.

28. Penninx JPM, Herman MC, Mol BW, Bongers MY. Five-year follow-up after comparing bipolar endometrial ablation with hydrothermablation for menorrhagia. Obstet Gynecol [Internet]. 2011 Dec [cited 2018 Dec 3];118(6):1287–92. Available from: http://insights.ovid.com/crossref?an=00006250-201112000-00012

29. Bansi-Matharu L, Gurol-Urganci I, Mahmood T, Templeton A, van der Meulen J, Cromwell D. Rates of subsequent surgery following endometrial ablation among English women with

menorrhagia: population-based cohort study. BJOG An Int J Obstet Gynaecol [Internet]. 2013 Nov [cited 2018 Dec 3];120(12):1500–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23786246

30. Cramer MS, Klebanoff JS, Hoffman MK. Pain is an Independent Risk Factor for Failed Global Endometrial Ablation. J Minim Invasive Gynecol [Internet]. 2018 Sep [cited 2018 Dec 3];25(6):1018–23. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1553465018300591

31. Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;

32. Steyerberg EW, Eijkemans MJ, Habbema JD. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. J Clin Epidemiol [Internet]. 1999 Oct [cited 2018 Dec 3];52(10):935–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10513756

33. Steyerberg EW, Eijkemans MJ, Harrell FE, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med [Internet]. 2000 Apr 30 [cited 2018 Dec 3];19(8):1059–79. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10790680

34. Breiman L. Randomforest2001. Mach Learn. 2001;

35. Liu Y, Zhang Y, Liu D, Tan X, Tang X, Zhang F, et al. Prediction of ESRD in IgA nephropathy patients from an asian cohort: A random forest model. Kidney Blood Press Res. 2018;

36. Fawagreh K, Gaber MM, Elyan E. Random forests: From early developments to recent advancements. Syst Sci Control Eng. 2014;

37. Kaitlin ;, Smith T;, Sadler B. Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets. Recommended Citation Kirasich. 2018.

38. Gareth J, Daniela W, Trevor H, Rober T. An Introduction to Statistical Learning with Applications in R. Current medicinal chemistry. 2000.

39. Loh WY. Regression trees with unbiased variable selection and interaction detection . Stat Sin. 2002;

40. James G, Witten D, Hastie T, Tibshirani R. An introduction to Statistical Learning. Current medicinal chemistry. 2000.

41. Hastie TT. The Elements of Statistical Learning Second Edition. Math Intell. 2017;

42. Bergstra JAMESBERGSTRA J, Yoshua Bengio YOSHUABENGIO U. Random Search for HyperParameter Optimization. J Mach Learn Res. 2012;

43. Jeni LA, Cohn JF, De La Torre F. Facing imbalanced data - Recommendations for the use of performance metrics. In: Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013. 2013.

44. Bouzari Z, Yazdani S, Naeimi Rad M, Bijani A. Is thermal balloon ablation in women with previous cesarean delivery successful? TURKISH J Med Sci [Internet]. 2018 Apr 30 [cited 2018 Dec 3];48(2):266–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29714438

45. Riley KA, Davies MF, Harkins GJ. Characteristics of patients undergoing hysterectomy for failed endometrial ablation. J Soc Laparoendosc Surg. 2013;

46. Kalish GM, Patel MD, Gunn MLD, Dubinsky TJ. Computed Tomographic and Magnetic Resonance Features of Gynecologic Abnormalities in Women Presenting With Acute or Chronic Abdominal Pain. Ultrasound Q [Internet]. 2007 Sep [cited 2018 Dec 3];23(3):167–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17805165

47. Gordts S, Grimbizis G, Campo R. Symptoms and classification of uterine adenomyosis, including the place of hysteroscopy in diagnosis. Fertility and Sterility. 2018.

48. Ennis M, Hinton G, Naylor D, Revow M, Tibshirani R. A comparison of statistical learning methods on the GUSTO database. Stat Med. 1998;

49. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. New England Journal of Medicine. 2019.

50. Kononenko I. Machine learning for medical diagnosis: History, state of the art and perspective. Artif Intell Med. 2001;

51. Lustgarten JL, Gopalakrishnan V, Grover H, Visweswaran S. Improving classification performance with discretization on biomedical datasets. AMIA . Annu Symp proceedings AMIA Symp. 2008;

Figures

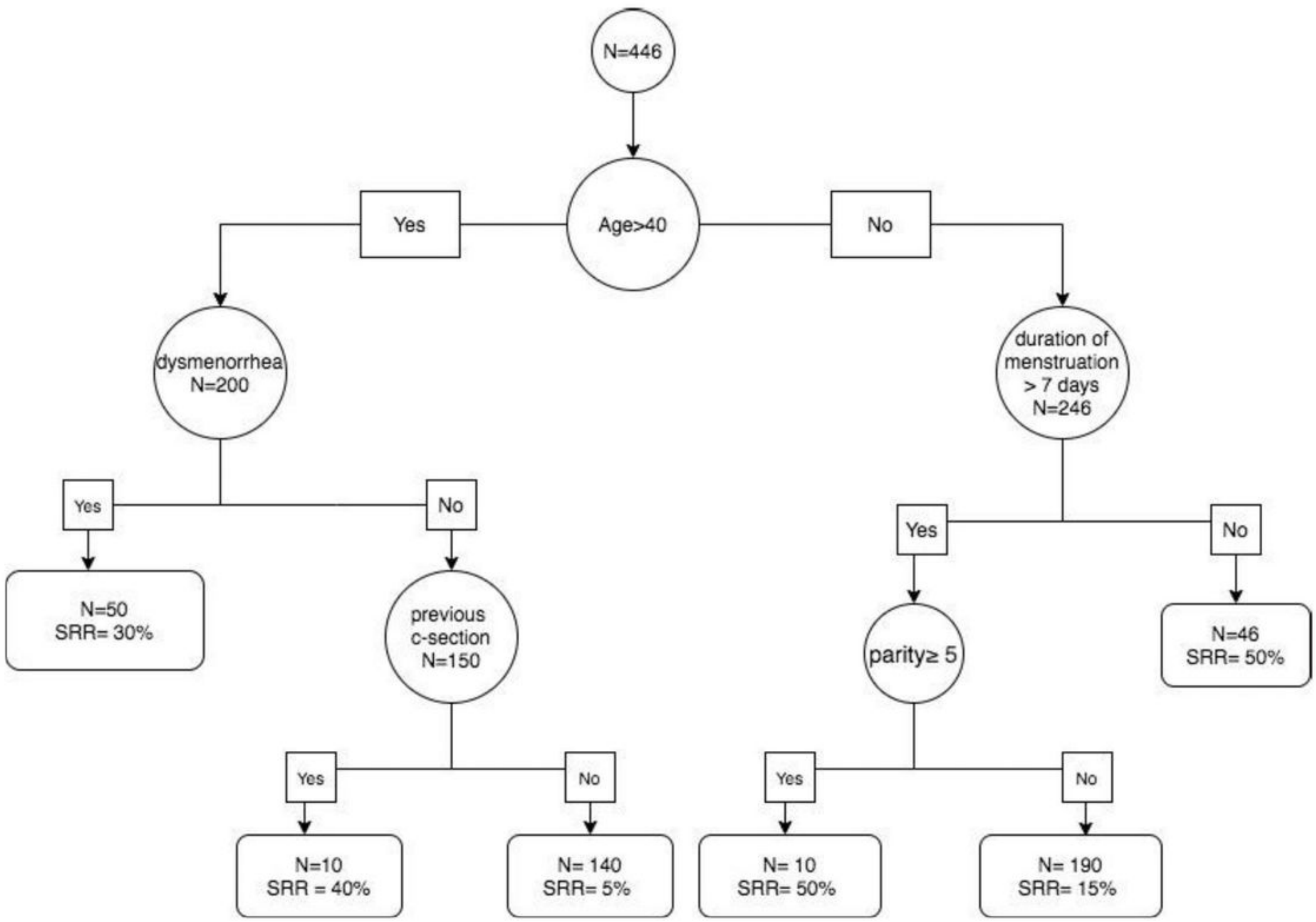

Figure 1

An illustration of a decision tree in the Random Forest model. The decision tree directs each case from the root node to the leaf nodes, resulting in a prediction. N= Number, SRR = Surgical Re-intervention Rate.

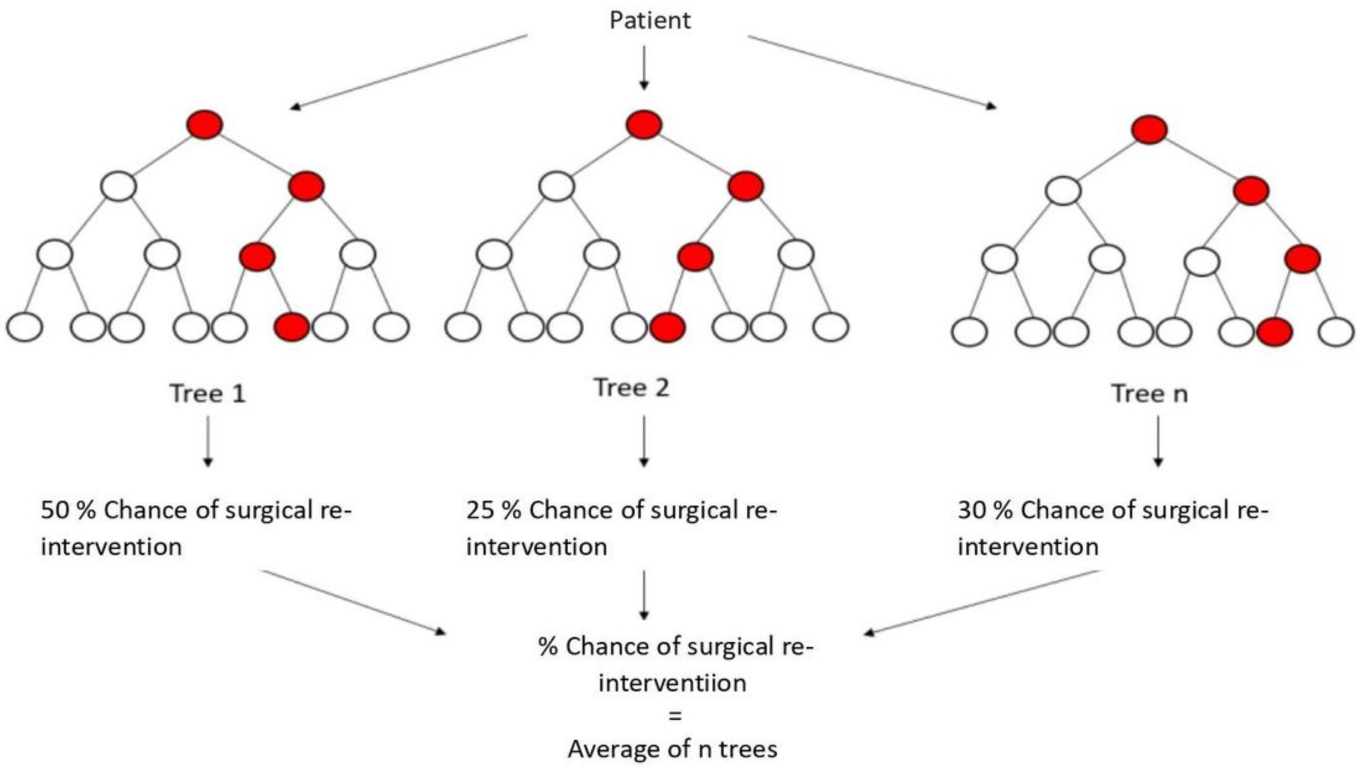

Figure 2

A simplified Random Forest model for the prediction of the surgical re-intervention .

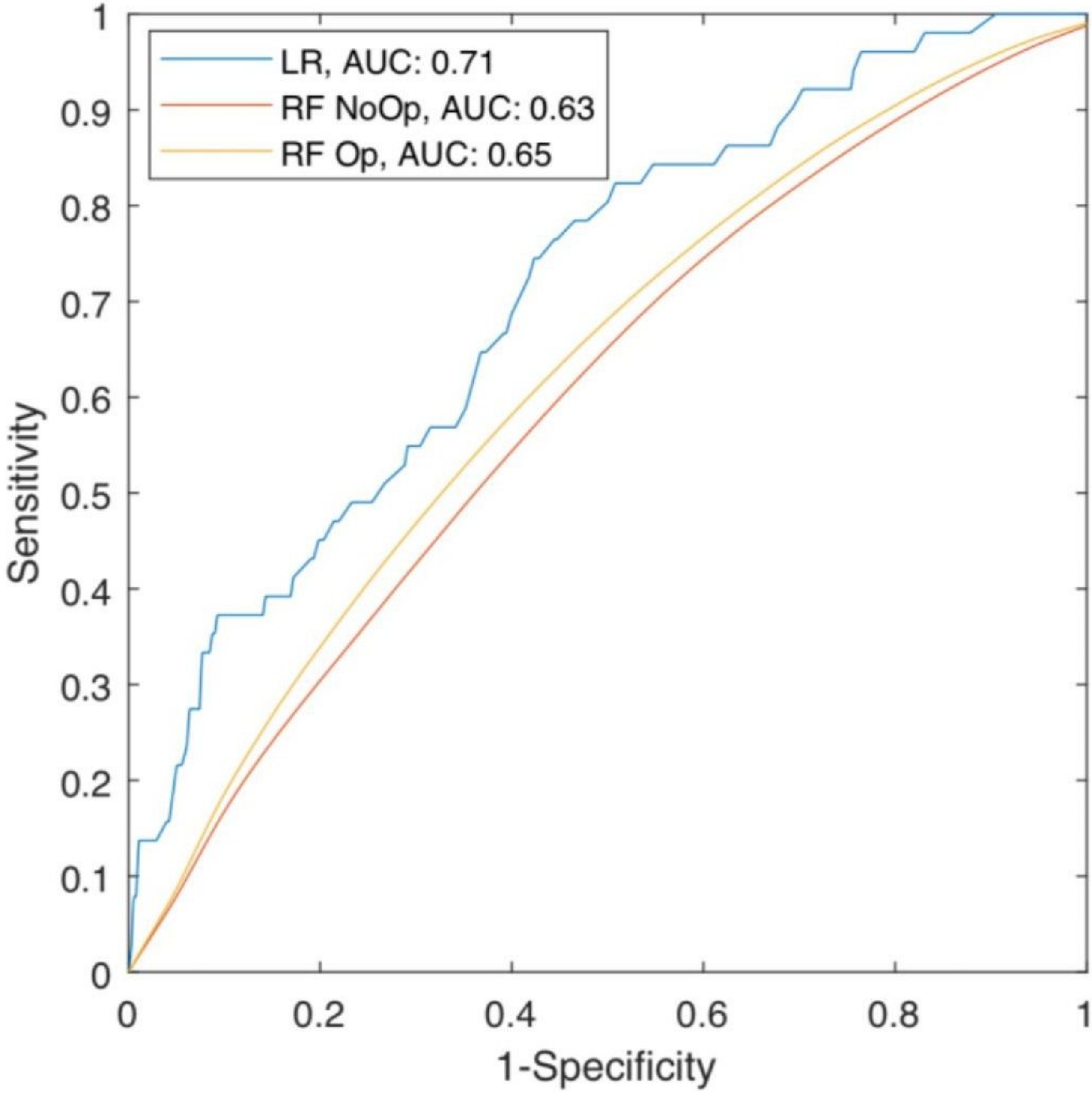

Figure 3

ROC-curve of the logistic regression and Random Forest model. LR AUC 0.71 (95% CI 0.64-0.78). , NoOp AUC 0.63 (0.54-0.71), Op AUC: (0.56 – 0.74) LR= logistic regression, RF= Random Forest, Op= after hyperparameter optimization, NoOp= before hyperparameter optimization

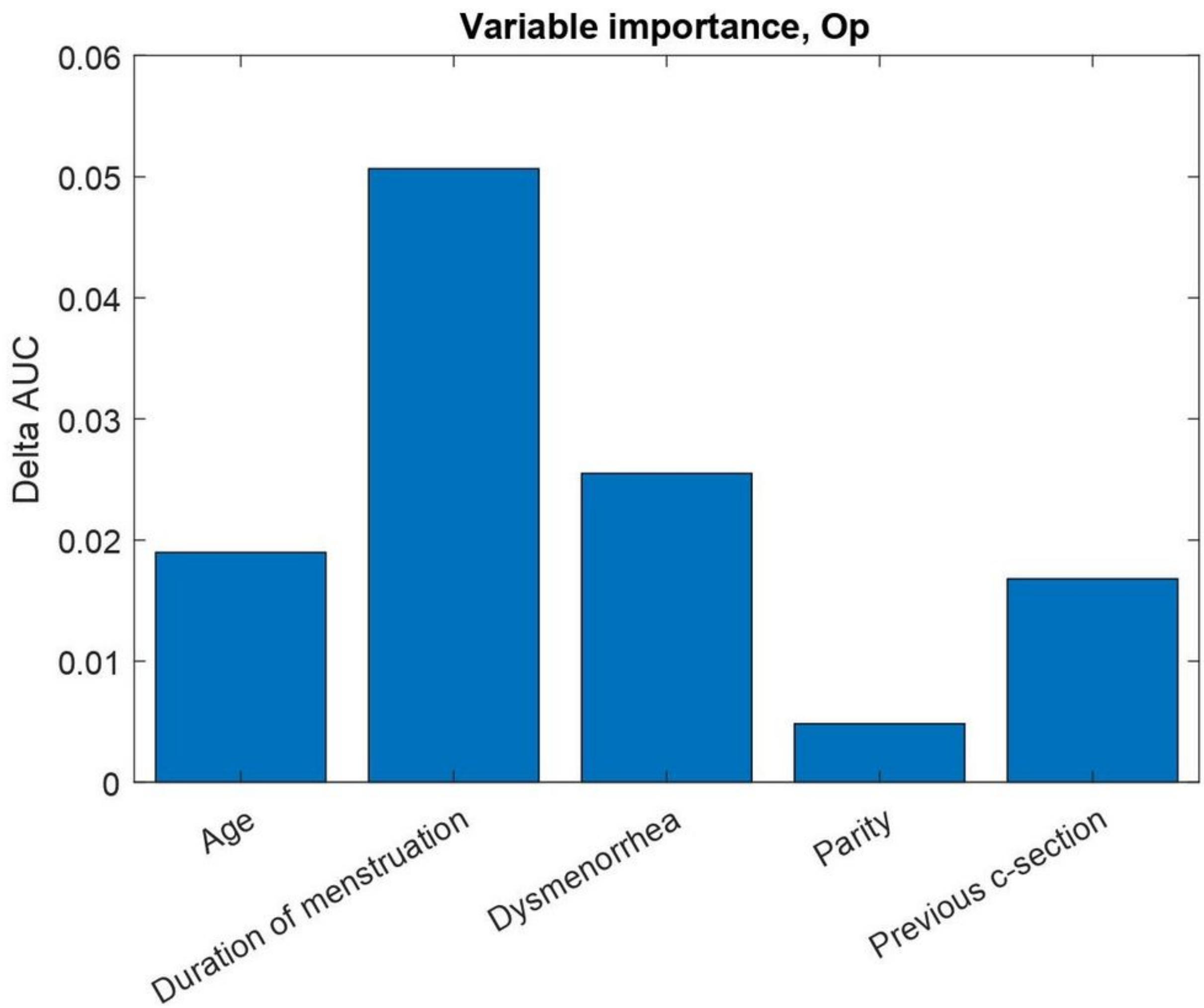

Figure 4

Contribution of predictors of surgical re-intervention within 2 years after endometrial ablation, after hyperparameter optimization