


ORIGINAL ARTICLE

Open Access

Prediction of unsuccessful endometrial ablation: random forest vs logistic regression



Kelly Yvonne Roger Stevens^{1,2*} , Liesbet Lagaert^{1,2*}, Tom Bakkes³, Malou Evi Gelderblom¹, Saskia Houterman⁴, Tanja Gijsen⁵ and Benedictus C. Schoot^{1,2}

Abstract

Background: Five percent of pre-menopausal 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 2 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 2 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 2 years after EA. Based on the ML model, dysmenorrhea and duration of menstruation have the highest impact on the surgical re-intervention rate.

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. Both techniques should be considered when developing a clinical prediction model. Both models can identify the clinical predictors to surgical re-intervention and contribute to the shared decision-making process in the clinical practice.

Keywords: Endometrial ablation, Machine learning, Random forest

* Correspondence: kyr.stevens@gmail.com; Liesbet.lagaert@gmail.com

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)

¹Department of Obstetrics and Gynaecology, Catharina Hospital, Michelangelolaan 2, 5623 EJ Eindhoven, The Netherlands

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

Five percent of pre-menopausal 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 7 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 clinical counselling of patients with abnormal uterine bleeding, a prediction model based on the combined influence of the abovementioned predictors could provide a better insight into the individual prognosis of endometrial ablation. In times of personalized 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 2 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]. 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 2 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, USA), and Thermablate® EAS (Idoman, Ireland). Recent publications have shown that these ablation techniques were equally effective [14, 25]. 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 or (suspicion of) endometrial malignancy or uterine cavity deformations (adenomyosis, anomalies, fibroids, or a polyp). Follow-up period after treatment was at least 2 years. This time-interval was chosen because previous literature stated that most re-interventions were done within 2 years. Follow-up ended on the day of hysterectomy, in case of death or on April 15, 2015 [9, 17, 18, 25–27].

Data were extracted from individual patient files by two researchers (K.S. and 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, 28, 29].

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 were used to develop the machine learning model. These were the pre-operative 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 [28, 29].

Multicollinearity and interaction between the significant variables in the model was tested. Bootstrap resampling was used for internal validation ($n = 5000$) [29, 30]. 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 can 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, 31]. 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 tree 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, 31, 32].

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 2 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. 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, 31, 33].

At each node split, a random subset of features (such as duration of menstruation and parity) is considered; this is done 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, 31–34].

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 [31, 35].

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

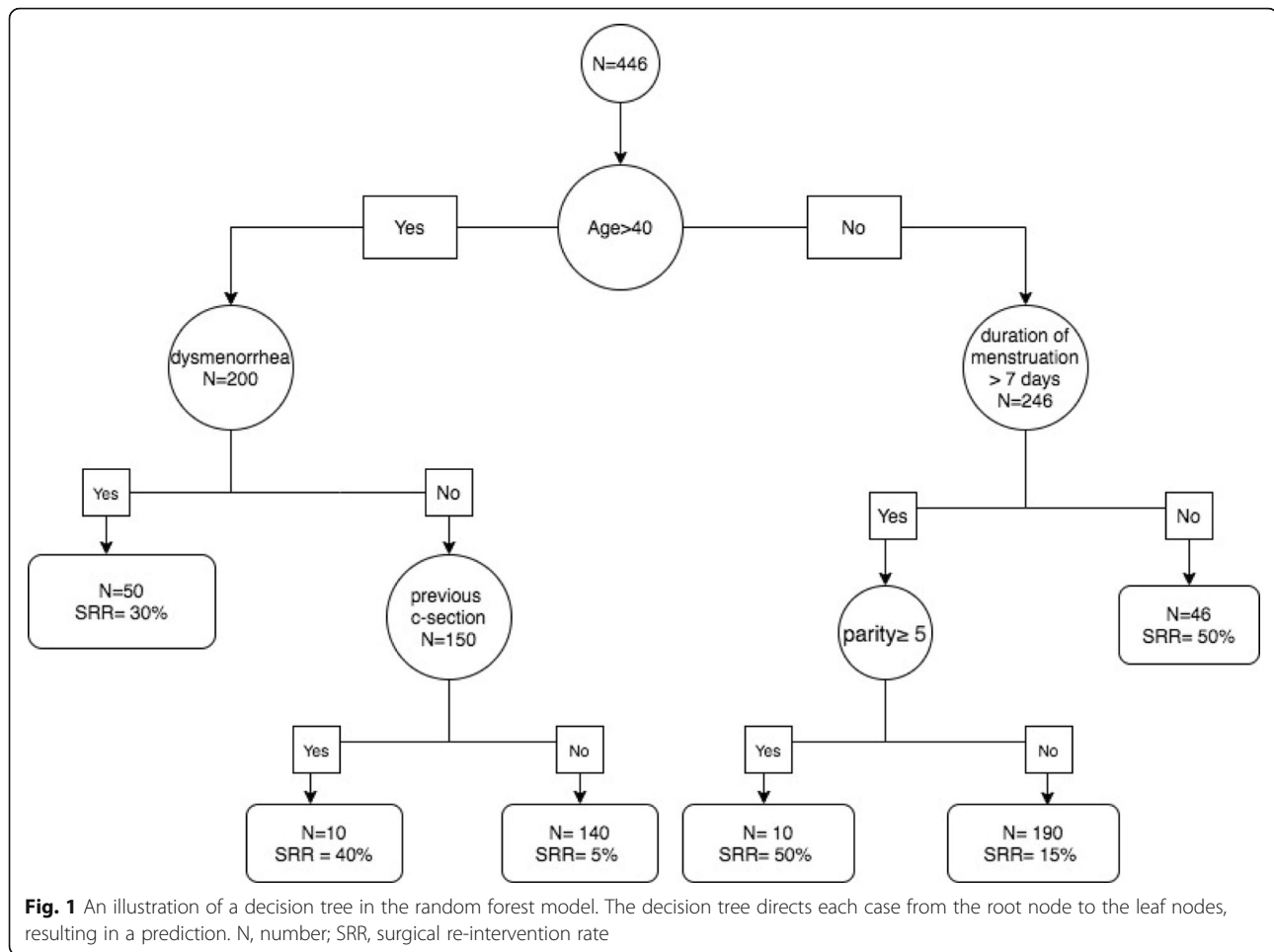
To predict the chance of surgical re-intervention within 2 years after EA, the model was initially trained and internally validated on the 446 cases. To make a good comparison between the RF and LR, the same validation technique was used. Therefore, a bootstrap resampling of 5000 was used. The performance measure area under the receiver operating curve (AUROC) was calculated.

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)) [36]. It was chosen to use the performance measure 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 the AUC statistic of the ML model. The AUC statistic is computed by randomly permutating (leaving out) the values of predictor x and comparing the resulting AUC to the not permuted AUC. Leaving out an important feature will result in a lower AUC of the ML model, while leaving out an unimportant feature will not change the AUC significantly [23, 35, 37].

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 2 years after EA.

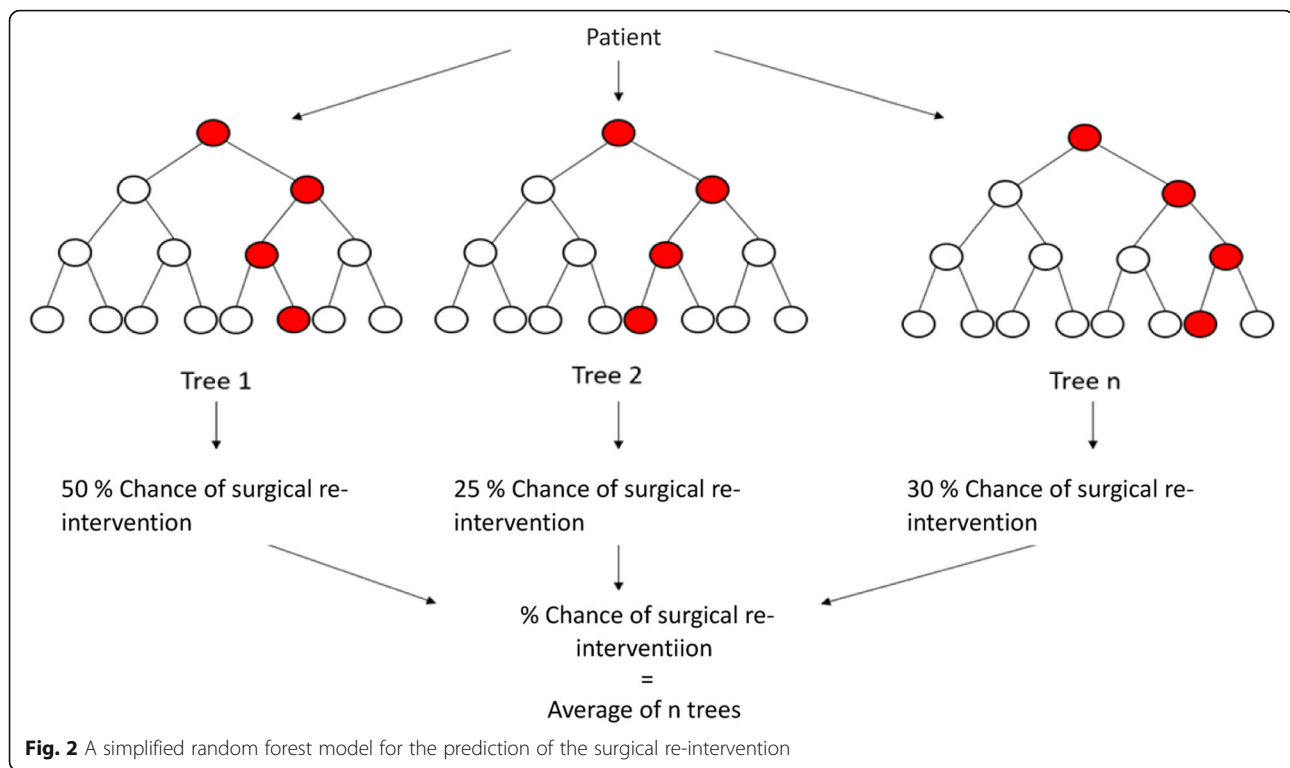
Patients' mean age during their EA was 43.8 years ($SD \pm 5.5$, range 20–55, missing values 0). The mean number of parity was 2.2 ($SD \pm 1.0$, missing values 0). Sixty-one (13.7%) of the patients underwent a caesarean section. The mean number of previous caesarean section was 0.2 ($SD \pm 0.6$, missing values 0)

Hundred sixty-nine (39.4%) of the patients had a menstruation period longer than 7 days, the mean number of menstrual days was 9.4 ($SD \pm 6.0$, 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].

Prediction models

Logistic regression model

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 \geq 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) (Fig. 3).

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 2 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 (Fig. 3).

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. The difference in AUC for the individual clinical predictors (permutation-based VIM) 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 (Fig. 4).

Discussion

Main findings

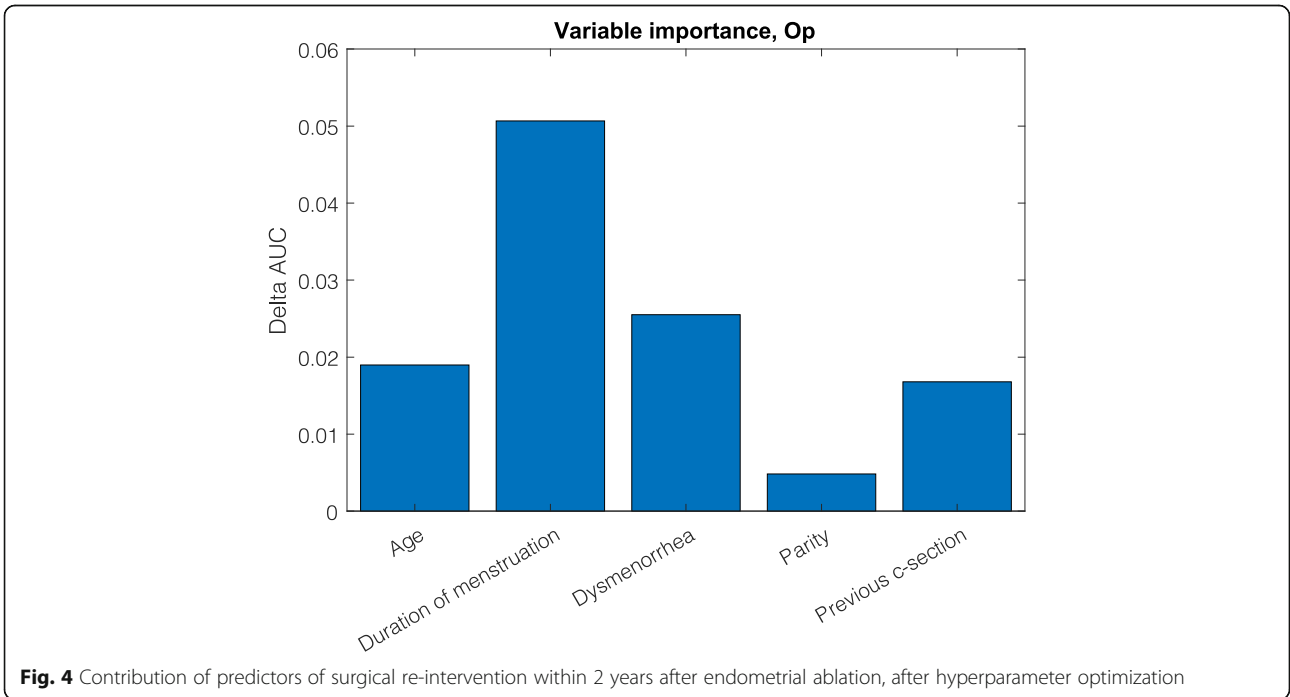
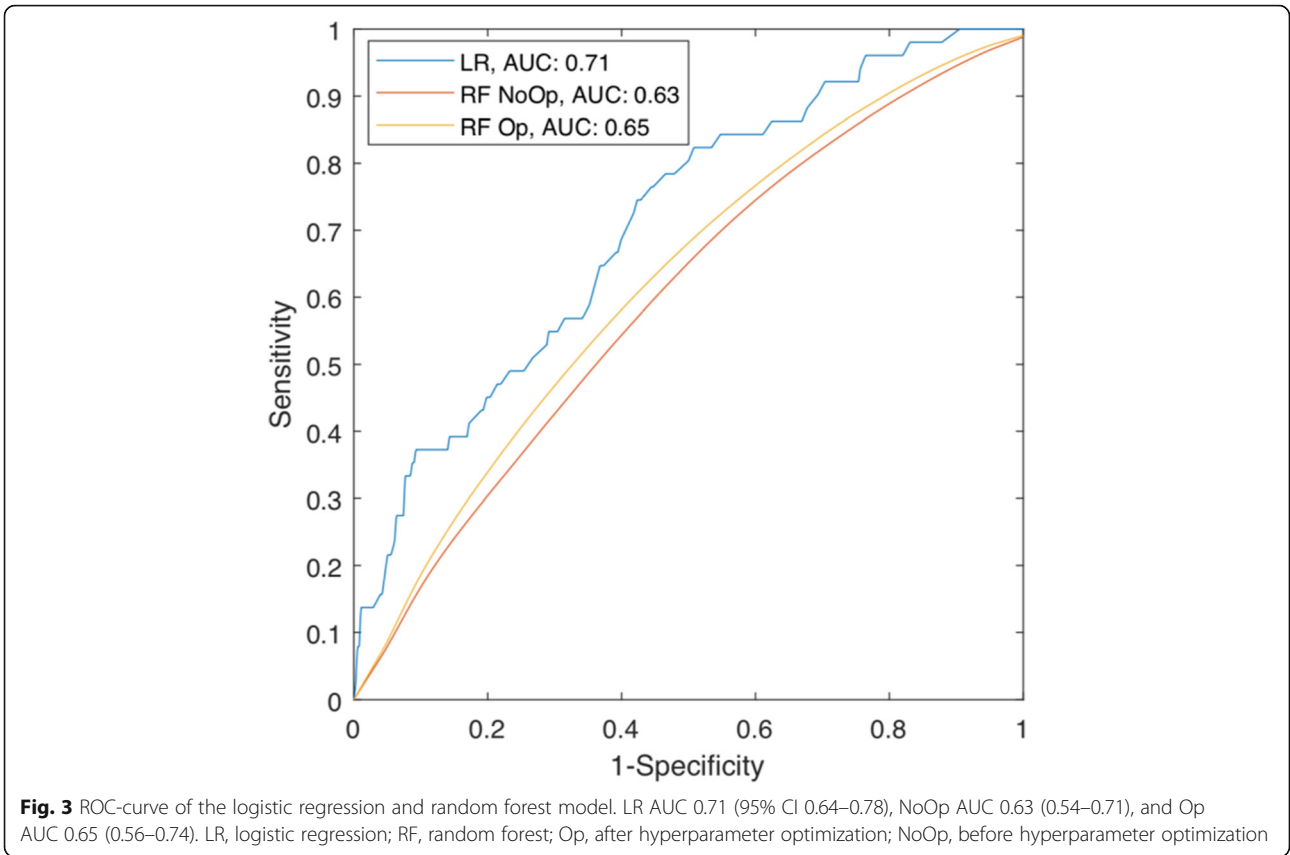
In this study, a ML model was made using random forest technique to predict surgical re-intervention within 2 years after EA. Comparison of the predictive performance of the RF model with the existing logistic regression model of Stevens et al. was made [16].

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). This shows that the LR prediction model developed by Stevens et al. [16] probably performs better in predicting surgical re-intervention within 2 years after EA than the newly developed ML model. However, this difference in performance is not statistically significant when looking at the confidence intervals.

Explaining the significant factors in the model

In the LR model, high parity (\geq 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 ML model, parity does not have a 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 [38]. 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, 39–41]. 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. 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 [42].

In line with previous studies, we found that younger age was associated with a higher risk of surgical re-intervention [7, 9–13, 43].

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, 44]. 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. On the 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, 44].

Secondly, some studies demonstrate that ML is performing better when a larger set of potential predictors are used. There seems to be an influence of the number of predictors (p) and the ratio of $p:n$ (sample size). ML tends to perform better for increasing p and $p:n$ [20, 24, 45, 46]. In our study, to limit potential bias, the five identical predictors as published before [16] were considered for the LR and ML algorithms. We did this to allow a fair comparison between the two models, probably in disadvantage of the ML model [20, 24, 45, 46].

Another possible reason for a lower AUC of the ML 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.

Finally, we can also consider that for this clinical problem a logistic approach is better than a ML 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, 47]. However, when we compare the variable importance between the LR and 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 ML and LR model because of different calculation methods (OR for the LR model and difference in AUC for the ML model).

Dysmenorrhea (OR 2.48) and a parity > 5 (OR 7.63) have the highest odds ratio in the multivariate LR analysis, while for the ML model the duration of menstruation and dysmenorrhea are the most important variables. We consider two possible reasons for the difference in importance. The first reason is that for the LR model, all continuous variables (except age) were discretized, while for the ML 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, 31, 44]. 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 ML 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 ML technique as well.

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 ML 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 mostly LR models are used in the clinical practice since ML models are not easily implemented in the clinical practice. These models 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 programmes. In this way, we can clinically benefit from the advantages of the ML models.

Conclusion

In conclusion, we can state that for the prediction of surgical re-intervention within 2 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 because of the possible clinical advantages. So far, there is no evidence for one single algorithm that outperforms the other in general use. Both the ML and LR model can identify the clinical predictors to surgical re-intervention and contribute to the shared decision-making process in the clinical practice. Based on our ML model, a longer duration of menstruation and the presence of dysmenorrhea are important predictive factors for surgical re-intervention.

Acknowledgements

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

Declarations

Authors' contributions

KYRS: Project development, data collection/management, data analysis, and manuscript writing/editing

LL: Project development, data collection/management, data analysis, and manuscript writing/editing

TB: Development of random forest model (machine learning)

MG: Manuscript editing

SH: Manuscript editing

TG: Data collection and manuscript editing

BCS: Project development, data collection, and manuscript editing

The authors read and approved the final manuscript.

Funding

None

Availability of data and materials

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.

Ethics approval and consent to participate

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.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Obstetrics and Gynaecology, Catharina Hospital, Michelangelolaan 2, 5623 EJ Eindhoven, The Netherlands. ²Women's Clinic, Ghent University Hospital, Corneel Heymanslaan 10, Ghent 9000, Belgium. ³Department of Department of Electrical Engineering, Biomedical Diagnostics Lab, TU Eindhoven, Eindhoven 5612 AZ, The Netherlands. ⁴Department of Education and Research, Catharina Hospital, Michelangelolaan 2, Eindhoven 5623 EJ, The Netherlands. ⁵Department of Obstetrics and Gynaecology, Elkerliek Hospital, Wesselmanlaan 25, Helmond 5707 HA, The Netherlands.

Received: 6 December 2020 Accepted: 6 June 2021

Published online: 29 June 2021

References

- Peeters JAH, Penninx JPM, Mol BW, Bongers MY (2013) 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* 167(1):100–103 [cited 2018 Dec 3]; Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0301211512005301>
- Waddell G, Pelletier J, Desindes S, Anku-Bertholet C, Blouin S, Thibodeau D (2015) Effect of endometrial ablation on premenstrual symptoms. *J Minim Invasive Gynecol* 22(4):631–636 [cited 2018 Dec 3]; Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1553465015000886>
- Laberge P, Leyland N, Murji A, Fortin C, Martyn P, Vilos G et al (2015) Endometrial ablation in the management of abnormal uterine bleeding. *J Obstet Gynaecol Can*
- Bouzari Z, Yazdani S, Azimi S, Delavar MA (2014) Thermal balloon endometrial ablation in the treatment of heavy menstrual bleeding. *Mediev Archaeol* 68(6):411–413
- 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 Women's Health*. 2015 59. [cited 2018 Dec 3]; Available from: <http://www.dovepress.com/cost-effectiveness-of-endometrial-ablation-with-the-novasurereg-system-peer-reviewed-article-IJWH>
- Angioni S, Pontis A, Nappi L, Sedda F, Sorrentino F, Litta P et al (2016) Endometrial ablation: first-vs. second-generation techniques. *Minerva Ginecol*
- El-Nashar SA, Hopkins MR, Creedon DJ, St Sauver JL, Weaver AL, McGree ME et al (2009) Prediction of treatment outcomes after global endometrial ablation. *Obstet Gynecol* 113(1):97–106 [cited 2018 Dec 3] Available from: <https://insights.ovid.com/crossref?an=00006250-200901000-00016>
- Wishall KM, Price J, Pereira N, Butts SM, Della Badia CR (2014) Postablation risk factors for pain and subsequent hysterectomy. *Obstet Gynecol* 124(5): 904–910 [cited 2018 Dec 3]; Available from: <https://insights.ovid.com/crossref?an=00006250-201411000-00007>
- Thomassee MS, Curlin H, Yunker A, Anderson TL (2013) Predicting pelvic pain after endometrial ablation: which preoperative patient characteristics are associated? *J Minim Invasive Gynecol* 20(5):642–647 [cited 2018 Dec 3]; Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1553465013001957>
- Bongers MY, Mol BWJ, HAM B (2002) Prognostic factors for the success of thermal balloon ablation in the treatment of menorrhagia. *Obstet Gynecol* 99(6):1060–1066 [cited 2018 Dec 3]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12052600>
- Longinotti MK, Jacobson GF, Hung Y-Y, Learman LA (2008) Probability of hysterectomy after endometrial ablation. *Obstet Gynecol* 112(6):1214–1220 [cited 2018 Dec 3]; Available from: http://content.wkhealth.com/linkback/openurl?sid=WKP_TLP:landingpage&an=00006250-200812000-00006
- Shamash AH, Sayed EH (2004) Prediction of successful menorrhagia treatment after thermal balloon endometrial ablation. *J Obstet Gynaecol Res* 30(3):210–216 [cited 2018 Dec 3]; Available from: <http://doi.wiley.com/10.1111/j.1447-0756.2004.00189.x>

13. Klebanoff J, Makai GE, Patel NR, Hoffman MK (2017) Incidence and predictors of failed second-generation endometrial ablation. *Gynecol Surg* 14(1):26 [cited 2018 Dec 3]; Available from: <https://gynecolsurg.springeropen.com/articles/10.1186/s10397-017-1030-4>
14. Louie M, Wright K, Siedhoff MT (2018) The case against endometrial ablation for treatment of heavy menstrual bleeding. *Curr Opin Obstet Gynecol*. 30(4):287–292 [cited 2018 Dec 3]; Available from: <http://insights.ovid.com/crossref?an=00001703-90000000-99318>
15. Lybol C, van der Coelen S, Hamelink A, Bartelink LR, Nieboer TE (2018) Predictors of long-term NovaSure endometrial ablation failure. *J Minim Invasive Gynecol*
16. Stevens KYR, Meulenbroeks D, Houterman S, Gijzen T, Weyers S, Schoot BC (2019) Prediction of unsuccessful endometrial ablation: a retrospective study. *Gynecol Surg* 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 (2012) Hysterectomy subsequent to endometrial ablation. *J Minim Invasive Gynecol* 19(4):459–464 [cited 2018 Dec 3]; Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1553465012001161>
18. Kreider SE, Starcher R, Hoppe J, Nelson K, Salas N (2013) Endometrial ablation: is tubal ligation a risk factor for hysterectomy. *J Minim Invasive Gynecol* 20(5):616–619 [cited 2018 Dec 3] 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 (2020) Implementing a preeclampsia prediction model in obstetrics: cutoff determination and health care professionals' adherence. *Med Decis Mak* 40(1):81–89
20. Christodoulou E, Jie MA, Collins GS, Steyerberg EW, Verbakel JY, van Calster B (2019) A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*
21. Breiman L (2001) Statistical modeling: the two cultures. *Stat Sci*
22. Deo RC (2015) Machine learning in medicine. *Circulation*.
23. Couronné R, Probst P, Boulesteix AL (2018) Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics*
24. Chen JH, Asch SM (2017) Machine learning and prediction in medicine — beyond the peak of inflated expectations. *N Engl J Med*
25. Sambrook AM, Bain C, Parkin DE, Cooper KG (2009) 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*
26. Herman MC, JPM P, Mol BW, Bongers MY (2014) Ten-year follow-up of a randomized controlled trial comparing bipolar endometrial ablation with balloon ablation for heavy menstrual bleeding. *Obstet Gynecol Surv*
27. Penninx JPM, Herman MC, Mol BW, Bongers MY (2011) Five-year follow-up after comparing bipolar endometrial ablation with hydrothermablation for menorrhagia. *Obstet Gynecol* 118(6):1287–1292 [cited 2018 Dec 3] Available from: <http://insights.ovid.com/crossref?an=00006250-201112000-00012>
28. Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF (2001) Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*
29. Steyerberg EW, Eijkemans MJ, Habbema JD (1999) Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *J Clin Epidemiol* 52(10):935–942 [cited 2018 Dec 3]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10513756>
30. Steyerberg EW, Eijkemans MJ, Harrell FE, Habbema JD (2000) Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med* 19(8):1059–1079 [cited 2018 Dec 3]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10790680>
31. Breiman L (2001) Randomforest 2001. *Mach Learn*
32. Liu Y, Zhang Y, Liu D, Tan X, Tang X, Zhang F et al (2018) Prediction of ESRD in IgA nephropathy patients from an asian cohort: a random forest model. *Kidney Blood Press Res*
33. Fawagreh K, Gaber MM, Elyan E (2014) Random forests: from early developments to recent advancements. *Syst Sci Control Eng*
34. Kaitlin, Smith T, Sadler B (2018) Random forest vs logistic regression: binary classification for heterogeneous datasets. *Recommended Citation Kirasich Gareth J, Daniela W, Trevor H, Rober T (2000) An introduction to statistical learning with applications in R. Curr Med Chem*
35. Jeni LA, Cohn JF, De La Torre F (2013) Facing imbalanced data - recommendations for the use of performance metrics. In: *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction. ACII 2013*
36. Hastie TT (2017) *The elements of statistical learning second edition. Math Intell*
37. Bouzari Z, Yazdani S, Naeimi Rad M, Bijani A (2018) Is thermal balloon ablation in women with previous cesarean delivery successful? *Turkish J Med Sci* 48(2):266–270 [cited 2018 Dec 3] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29714438>
38. Cramer MS, Klebanoff JS, Hoffman MK (2018) Pain is an independent risk factor for failed global endometrial ablation. *J Minim Invasive Gynecol* 25(6): 1018–1023 [cited 2018 Dec 3] Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1553465018300591>
39. Riley KA, Davies MF, Harkins GJ (2013) Characteristics of patients undergoing hysterectomy for failed endometrial ablation. *J Soc Laparoendosc Surg*
40. Kalish GM, Patel MD, Gunn MLD, Dubinsky TJ (2007) Computed tomographic and magnetic resonance features of gynecologic abnormalities in women presenting with acute or chronic abdominal pain. *Ultrasound Q* 23(3):167–175 [cited 2018 Dec 3]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17805165>
41. Gordts S, Grimbizis G, Campo R (2018) Symptoms and classification of uterine adenomyosis, including the place of hysteroscopy in diagnosis. *Fertil Steril*
42. Bansi-Matharu L, Guroi-Urganci I, Mahmood T, Templeton A, van der Meulen J, Cromwell D (2013) Rates of subsequent surgery following endometrial ablation among English women with menorrhagia: population-based cohort study. *BJOG An Int J Obstet Gynaecol* 120(12):1500–1507 [cited 2018 Dec 3] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23786246>
43. Ennis M, Hinton G, Naylor D, Revow M, Tibshirani R (1998) A comparison of statistical learning methods on the GUSTO database. *Stat Med*
44. Rajkomar A, Dean J (2019) Kohane I. Machine learning in medicine, New England Journal of Medicine
45. Kononenko I (2001) Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med*
46. Lustgarten JL, Gopalakrishnan V, Grover H, Visweswaran S (2008) Improving classification performance with discretization on biomedical datasets. *AMIA Annu Symp proceedings AMIA Symp*

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)