RESEARCH ARTICLE

Machine learning as a clinical decision support tool for diagnosing superficial peritoneal endometriosis in women with dysmenorrhea and acyclic pelvic pain

Letícia Luiza Alves Santos[1,2], Mateus Carvalho de Azevedo[1,2], Lia Keiko Shimamura[1,2], Antonio Alberto Nogueira[1], Francisco José Candido-dos-Reis[1,2], Eduardo Schor[3], Julio Cesar Rosa-e-Silva[1,2], Daniel Guimarães Tiezzi[1,2], Omero Benedicto Poli-Neto[1,2*]

[1]Department of Obstetrics and Gynecology, Ribeirão Preto Medical School of the University of São Paulo USP, Ribeirão Preto - SP, Brazil.
[2]Laboratory for Translational Data Science, CNPq (National Council for Scientific and Technological Development), Brazil.
[3]Department of Obstetrics and Gynecology, Federal University of São Paulo, São Paulo - SP, Brazil.

*e-mail: polineto@usp.br

ABSTRACT

Background: Superficial peritoneal endometriosis, despite being the most common type of lesion, presents the greatest challenge for non-invasive diagnosis, resulting in the majority being recognised surgically.

Objective: To evaluate the performance of machine learning in predicting superficial peritoneal endometriosis in women with chronic dysmenorrhoea and pelvic pain without abnormal ultrasound findings.

Design: Retrospective observational study.

Subjects: 298 women with severe dysmenorrhea and persistent acyclic pelvic pain after at least 6 months of hormonal treatment who underwent laparoscopy, with imaging examinations showing no significant abnormal findings.

Exposure: Data collected included clinical history, physical examination previously to the laparoscopy.

Main Outcome Measures: Augmented backward elimination was used as a procedure to obtain a baseline interpretable binomial logistic model. The performance of various machine learning models, including Random Forest, Light Gradient Boosting Machine, Extreme Gradient Boosting, Extremely Randomised Trees, Categorical Boosting, Adaptive Boosting, Support Vector, Multilayer Perceptron, Naive Bayes, Voting, and Stacking ensemble meta-classifiers, in predicting superficial peritoneal endometriosis. Feature importance was assessed using Shapley Additive Explanations (SHAP) values. **Results:** The presence of irregular menstrual cycle, irritable bowel syndrome, bladder pain syndrome, abdominal trigger point, and pelvic floor tenderness were independently associated with the diagnosis of superficial peritoneal endometriosis. SHAP values indicated that a history of pelvic inflammatory disease also suggested endometriosis. The soft voting classifier, which includes Extreme Gradient Boosting and Naive Bayes algorithms, demonstrated the highest recall (79.3%), while the Support Vector classifier achieved the best specificity (74.2%).

Conclusion: Irregular menstrual cycles, irritable bowel syndrome, bladder pain syndrome, abdominal trigger points, and pelvic floor tenderness are independent factors linked with intraoperative findings of superficial peritoneal endometriosis. Additional variables, such as a history of pelvic inflammatory disease, may further enhance preoperative diagnostic accuracy. Machine learning approaches show promise in predicting the disease through pre-operative clinical data in this population. This predictive capability can support personalised patient counselling and surgical decision-making.

Keywords: Chronic pelvic pain, Dysmenorrhoea, Endometriosis, Laparoscopy, Machine learning, Prediction.

Introduction

Endometriosis affects approximately 5-10% of women, with an annual incidence between 0.1-0.3%[1–4], affecting about 180 million women globally, predominantly those of reproductive age, though cases can be identified post-menopause[5]. The condition, associated with chronic inflammation, is characterised by endometrial-like tissue outside the uterine cavity, though its definition does not include symptoms or physiological changes[6]. Clinically, it is found in up to 7% of asymptomatic women undergoing tubal sterilisation, up to 50% of adolescents with severe dysmenorrhea, 5-24% of women with persistent non-cyclic pain, and 10-40% of women with infertility[7]. It significantly impacts women's lives[8] and the healthcare system, reducing quality of life[9], causing work productivity loss[10], and generating substantial economic costs, up to $50 billion annually in the USA[11,12].

Despite its clinical significance, diagnosing endometriosis still takes an average of 4 to 11 years, even in countries with universal healthcare systems[13]. This protracted delay is multifactorial, stemming from the reliance on laparoscopy as the gold standard for diagnosis, which is inherently invasive. Additionally, the condition is characterised by a diverse and often overlapping symptomatology, coupled with the lack of definitive, pathognomonic signs or symptoms[14]. Societal factors also play a critical role, including the normalisation of symptoms such as pelvic pain and dysmenorrhea, as well as pervasive gender-related stigma that can lead to the dismissal or underestimation of symptoms[15].

Endometriosis itself is highly heterogeneous, typically manifesting in three primary phenotypes: superficial peritoneal endometriosis (SPE), ovarian endometrioma, and deep infiltrating endometriosis, each with distinct pathophysiological and clinical characteristics[16]. Given this complexity, non-invasive diagnostic tools are preferable[17]. Imaging modalities, particularly transvaginal ultrasonography, are recommended as a first-line approach[18–20], and may even outperform surgery in select scenarios[21].

However, while certain authors have described suggestive signs for diagnosing SPE[22,23], this remains a substantial limitation of imaging modalities[24].

SPE is considered one of the most frequent forms of endometriotic lesions, potentially representing up to 80% of all identified cases[25]. Importantly, it has been independently associated with primary infertility and moderate to severe pain[26], underscoring the necessity for accurate diagnosis. Current guidelines often recommend symptom management before surgical intervention and advise laparoscopy for patients with negative imaging findings or when empirical treatment fails[18–20,27]. However, reliance on laparoscopy poses a challenge, as rates of negative surgical findings, where no visible lesions are identified, can exceed 50% in some cohorts[28–30].

Given its prevalence and clinical impact, improving pre-surgical diagnostic accuracy for SPE is critical to reducing the frequency of unnecessary laparoscopies, despite being a relatively safe procedure[31]. While the data supporting surgical excision or ablation of SPE remains limited[32], some evidence suggests that surgical intervention may provide benefits irrespective of findings[33]. This highlights the pressing need for more effective non-invasive diagnostic tools, specifically tailored to better identify SPE and improve patient outcomes.

Recent advances in artificial intelligence (AI) have greatly improved the interpretation of complex clinical data, offering significant promise in understanding chronic pain conditions[34]. A recent study evaluated an AI-powered mobile application designed to aid in the screening of endometriosis among patients presenting with chronic pelvic pain and/or unexplained infertility[35]. While the application demonstrated high sensitivity and positive predictive value, its specificity and negative predictive value were notably low. This limitation likely stems from the app's reliance on user-provided information, failing to address more complex diagnostic scenarios, particularly cases where ultrasonographic findings are unremarkable.

The number of published studies on endometriosis and AI has increased substantially in the last decade[36]. Current research has explored areas such as biomarkers, imaging analysis, and patient-reported outcomes. However, biomarkers have yet to demonstrate sufficient predictive value for endometriosis[37], and while imaging modalities exhibit high sensitivity in symptomatic populations, their utility diminishes in more nuanced presentations. Consequently, the greatest potential for AI may lie in interpreting patient-reported symptoms. Goldstein and Cohen (2023) developed a machine learning (ML) model for predicting endometriosis based on self-reported symptoms[38]. This study included data from individuals with and without endometriosis, collected via structured questionnaires. Although their model showed good accuracy, several limitations were noted. Chief among these were the reliance on self-reported diagnoses and the inherent heterogeneity within the non-endometriosis group, which likely included many healthy individuals. While such tools may facilitate self-diagnosis and expedite referral for patients with a high likelihood of endometriosis, their utility in clinical practice is limited. Specifically, they provide little additional insight into the likelihood of endometriosis in symptomatic populations, such as those experiencing chronic pelvic pain or dysmenorrhoea.

Given the challenges in predicting endometriosis in individuals with pelvic pain and no typical ultrasonographic findings, the lack of studies in the literature addressing this critical aspect, the ability of AI to effectively manage complex clinical data interactions, and the importance of thorough preoperative counselling, we hypothesise that machine learning (ML) could enhance the non-invasive diagnosis of superficial peritoneal endometriosis (SPE). Such advancements could enable personalised informed consent, improve the precision of interventions, and reduce unnecessary surgeries. Thus, the aim of this study was to evaluate the performance of ML in predicting an SPE diagnosis in women presenting with chronic dysmenorrhoea and pelvic pain, characterised by the absence of abnormal ultrasound findings and persistent symptoms despite hormonal therapy.

Materials and Methods

DESIGN
We conducted a retrospective observational study using an anonymized database of surgical requests performed at Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HCFMRP-USP) between 2010 and 2017. The study was approved by the Local Ethics Committee under number 193.015 and national registration 10863612.7.0000.5440. This study adhered to the TRIPOD-AI guidance, an extension of the Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis (TRIPOD) statement, specifically developed to enhance the reporting and appraisal of prediction model studies employing machine learning techniques[39].

SETTING
The HCFMRP-USP, located in Ribeirão Preto, São Paulo, Brazil, is the region's largest public hospital and serves as a tertiary care referral centre for a population of 4.5 million in northeastern São Paulo State. Affiliated with the University of São Paulo, it offers advanced surgical procedures across specialties, recognized nationally for excellence in gynaecology and endometriosis care.

PARTICIPANTS AND ELIGIBILITY
Among 476 patients with chronic pelvic pain and suspected endometriosis, 298 women aged 18-47 with severe dysmenorrhea and persistent acyclic pelvic pain after at least 6 months of hormonal treatment and no significant abnormal ultrasound or MRI findings were eligible. All underwent diagnostic conventional laparoscopy at HCFMRP-USP by the same expert team. Typical endometriotic lesions[40] or histopathological confirmation of atypical lesions observed during or immediately following laparoscopy served as diagnostic confirmation. The areas of adhesions without

obvious signs of endometriotic lesions were biopsied for histological confirmation. Women with a prior diagnosis of endometriosis, previous diagnostic surgeries, or evidence of bowel, bladder, or ureter involvement, as well as those with identified endometriomas or lesions deeper than 5 mm beneath the peritoneal surface during surgery, were not considered eligible for the study.

INDEPENDENT FEATURES

The independent variables included the duration (in months) and the intensity of symptoms, measured using a visual analogue scale (in millimetres), with the most severe symptoms selected from dysmenorrhoea, dyspareunia, dyschezia, and pelvic pain; frequency of symptoms (weekly); age at surgery; age of menarche; age of first sexual intercourse reported (coitarche); number of previous sexual partners; education level (elementary (< 9 years); high (≥ 9 years)); civil status (single, married, divorced, widow); paid job (any remunerated employment activity); active tobacco exposure (previous or current); alcohol misuse (frequent or regular consumption of more than an average of 2 units a day of alcoholic beverages); sedentarism: assessed by the short version of the International Physical Activity Questionnaire (IPAQ)[41], validated for use with Brazilian population[42]; body mass index (kg.m^{-2}) at the time of laparoscopy; irregular menstrual cycle (lower than 21 days or greater than 35 days in the last 3 months); heavy bleeding (defined as any of the following: frequent pad or tampon changes (every 1-2 hours), using two types of period products simultaneously, periods lasting over 7 days, passing blood clots larger than 2.5 cm, bleeding through clothes or bedding, and needing to avoid daily activities due to period bleeding); number of previous pregnancies; number of deliveries; number of previous miscarriages; number of previous C-sections; episiotomy in any previous delivery; previous delivery assisted by forceps; any previous abdominal surgery; any previous Pfannenstiel incision (C-section or not); infertility history: (attempting to conceive for more than 2 years without success having regular unprotected sex); report of any previous clinically confirmed pelvic inflammatory disease; chronic low back pain; report of previous clinically confirmed migraine; report of previous clinically confirmed depression; any kind of sexual violence previously suffered; any kind of emotional abuse or neglect previously suffered; any kind of physical violence previously suffered; premenstrual syndrome (symptoms such as bloating, breast tenderness, fatigue, and changes in sleep and eating habits usually associated with periods); insomnia (difficulty in falling asleep or staying asleep with negative consequences during the day); hypersomnia (sleeping for long hours at night and regularly falling asleep during the day); dyspareunia (painful intercourse that negatively impacts mental/physical health, body image, and relationships with partners); vaginismus (the vagina suddenly tightens up when trying to insert something into it); pain after sex (following sexual intercourse); pain after exercise (abdominal or pelvic pain following physical exercise); neuropathic pain component: assessed by the Brazilian version of the Douleur Neuropathique 4 (DN-4) Questionnaire[43], considering a score ≥ 4 as indicative of a neuropathic pain component; avoidance of activities (significant restriction of any physical activity imposed by symptoms); psychological distress (assessed by the Brazilian version[44] of the Self Reporting Questionnaire (SRQ-20), considering a score ≥ 7 as indicative of mental suffering; hypoactive sexual desire; insufficient vaginal (lack) lubrication; inability to achieve orgasm; functional diarrhoea according to Rome III criteria[45]; as well as functional constipation and irritable bowel syndrome; dyschezia; bladder pain syndrome (pelvic pain accompanied by at least one other irritative urinary symptom regularly present over the last 3 months (urinary urgency, urinary frequency, nocturia, dysuria, bladder discomfort, tenesmus)); abdominal trigger point (spots of extreme tenderness and hyperirritability in the muscles of the lower abdominal wall); abdominal

allodynia (pain due to a stimulus that does not normally provoke pain in the lower abdominal wall); vulvodynia (pain, burning, and discomfort in the vulva with no specific cause for at least the last three months); pelvic floor tenderness (painful discomfort during unidigital palpation performed to identify tenderness of the levator ani muscle bilaterally); pelvic congestion syndrome assessed by ultrasound screening with findings of dilated ovarian vein with reversed caudal flow, presence of varicocele, dilated arcuate veins crossing the uterine myometrium, polycystic changes of the ovary, and variable duplex waveform during the Valsalva's manoeuvre[46]; chronic non-steroidal anti-inflammatory drugs (NSAIDs) use (more than three times a week for more than the last three months); type of hormonal contraceptive (current usage in the last 3 months); antidepressants or gabapentinoids (current usage in the last 3 months).

EXPLORATORY ANALYSIS

Univariable analysis was conducted as part of the exploratory analysis. Continuous variables were summarised by mean, median, and interquartile range. Normality was assessed using the Kolmogorov-Smirnov test and visual graphics[47]. Categorical attributes were represented by absolute (N) and relative (%) frequencies. Depending on the data characteristics, Fisher's exact test or Pearson's Chi-squared test, and t-test or Wilcoxon rank-sum test were performed as appropriate. The Benjamini–Hochberg method was applied to adjust p-values for the false discovery rate (FDR).

MISSINGNESS PATTERN

The proportion of complete cases was 89.9% (n = 268/298). No variable exhibited a missingness rate exceeding 5%. Based on the similar pattern of missing data according to the outcome, we assumed the data were missing-at-random (MAR)[48], and multivariate imputation by chained equations was performed[49].

MULTICOLLINEARITY

Multicollinearity was assessed by computing the variance inflation factor (VIF), which measures the inflation of variance in a regression coefficient due to the correlation between that predictor and the others. Given the multiple categorical variables, the adjusted generalised variance inflation factor was used[50], with threshold values above 1.6 indicating a potential multicollinearity problem. Additionally, a correlation matrix[51] was considered to aid in feature selection. A panel of experts refined the list of variables to be included in the final model.

BASELINE LOGISTIC REGRESSION ANALYSIS

We employed a binomial logistic regression model with a logit link function. Augmented backward elimination, as implemented in the ABE package in R, was utilised to derive an interpretable baseline model, incorporating the Akaike Information Criterion (AIC) and a threshold of 0.05 for the relative change-in-estimate criterion[52]. This method begins with all potential predictors and iteratively removes the least significant variables, balancing both significance and change-in-estimate criteria. Compared to traditional backward elimination, it tends to select larger models and approximates the unselected model with minimal differences in the point estimates of the regression coefficients. We presented the results using crude and adjusted odds ratio (OR) and confidence interval. We checked for separation, and influential outliers (no issues were found). For assessing the goodness of fit and calibration of this model we used the omnibus test[53]. For checking discrimination, we used the Somers' D index[54], which is a measure of the strength and direction of association between two variables. A higher value indicates a stronger positive relationship, and it can be used to assess the accuracy of diagnostic tests.

MACHINE LEARNING CLASSIFICATION

Before modelling, the dataset was divided into training and test sets with an 80%-20% ratio. To avoid data leakage, all preprocessing steps, including missing data imputation, were first applied only to the training set. The same transformations were subsequently applied to the test set.

Logistic regression was employed as the baseline model to benchmark predictive performance on the training dataset. Although the primary outcome proportion was close to 50%, class weights were adjusted to account for potential imbalance. An initial exploratory analysis was conducted using additional machine learning algorithms, including Random Forest, Light Gradient Boosting Machine, Extreme Gradient Boosting, Extremely Randomised Trees, Categorical Boosting, Adaptive Boosting, Support Vector Machines, Multilayer Perceptron, and Naive Bayes. Each model was subjected to 100 simulations with random seeds to calculate confidence intervals, and ten-fold cross-validation was applied for internal validation.

Hyperparameters were optimised using both manual testing and Bayesian optimisation methods implemented in the Scikit-Optimize library (version 0.8.1). This method is based on probabilistic models. It iteratively updates the probability of each potential hyperparameter based on prior knowledge and new data, allowing for more efficient tuning of the model's parameters[55]. Predictions on the test dataset were then made to estimate performance measures.

Additionally, we incorporated hard (majority) and soft (weighted probability) voting by comparing all possible combinations of classifiers weighted by F1 score. Stacking ensemble meta-classifiers were also explored by combining all possible algorithm combinations[56]. We implemented these models using the Scikit-learn library (version 1.5.2).

THRESHOLD DETERMINATION AND OPTIMIZATION
After hyperparameter tuning, we optimised the decision threshold for converting predicted probabilities into binary outcomes. Rather than defaulting to the conventional 0.5 threshold, we explored thresholds ranging from 0 to 1 in small increments to maximise the F1 score. Additionally, we experimented with penalising false negatives to improve model sensitivity; however, this approach did not yield satisfactory results. Given the trade-off observed between specificity and accuracy, we ultimately retained the standard 0.5 threshold for final analyses.

In addition to optimising the threshold to maximise the F1 score, we also conducted a comparison of models at fixed levels of specificity (70%, 75%, and 80%). For each model, we identified the decision threshold that achieved the desired specificity by adjusting the threshold applied to predicted probabilities, with a tolerance of 1%. This approach ensured comparability between models while maintaining control over the ability to correctly identify negative cases.

MODEL PERFORMANCE
We proceeded with model validation on the test set and compared the performance through measures of F1-score (harmonic mean between recall and precision), precision, recall (sensitivity), specificity, balanced accuracy, and area under the receiver-operating characteristics (AUROC) curve (rate of true and false positives).

FEATURE IMPORTANCE
Shapley Additive Explanations values were utilised to standardise the identification and measurement of feature importance[57]. As our focus is on the positive class (endometriosis), we chose to order the best classifiers by F1 score.

SOFTWARE
We developed all applications in the Python language (version 3.9) and used Google Colaboratory platform, which provides a serverless Jupyter notebook environment for interactive development[58].

Results

Figure 1 displays the flowchart of the participant selection and inclusion process. A total of 48% (n = 144/298) had a diagnosis of superficial peritoneal endometriosis (SPE) confirmed at laparoscopy. Stage II disease was more common (69.4%, n = 100), followed by stage I (25.0%, n = 36) and stage III (5.6%, n = 8). The prevalence of intra-abdominal adhesions identified intraoperatively was similar between women with and without SPE (p = .998),

being 43.1% (n = 62/144) and 42.9% (n = 66/154), respectively. The clinical improvement rate at the end of three months was significantly better for women with endometriosis: 79.2% (n = 114/144) compared to 13.0% (n = 20/154) (p < .001). The full characterisation of the sample is presented as supplementary material (Supplementary Table S1).
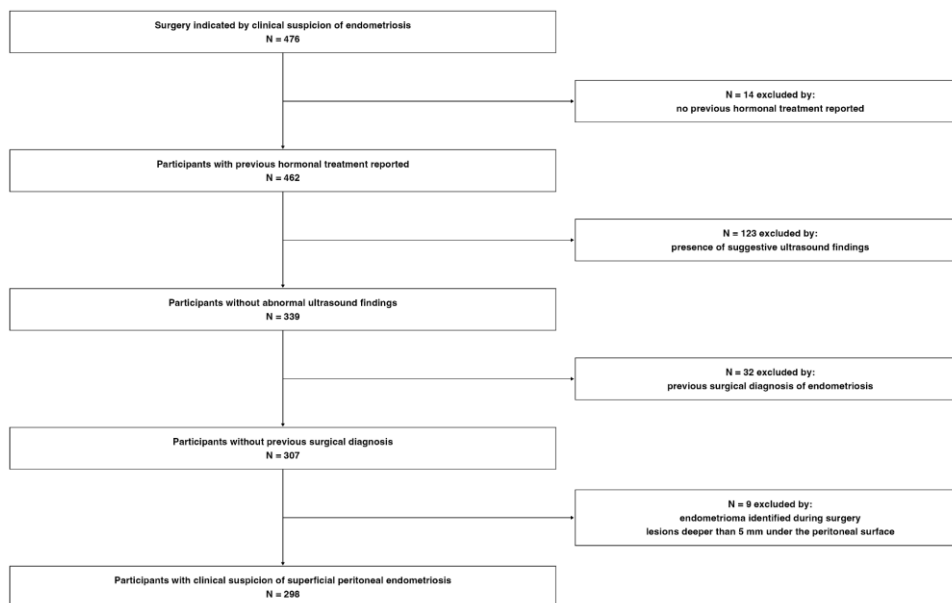

Figure 1: Flowchart illustrating the pathway followed for the selection of patient medical records to be included in the analysis.

Irregular menstrual cycle, irritable bowel syndrome, abdominal trigger point, pelvic floor tenderness, and bladder pain syndrome were associated with intraoperative diagnosis of superficial peritoneal endometriosis (SPE) according to the explainable logistic regression model. Otherwise, sleep disorder, lack of lubrication, pain after sex, previous episiotomy, and neuropathic components were associated with no SPE during laparoscopy. A forest plot including only significant risk factors shows in detail the magnitude of the odds (Figure 2). Supplementary material provides the crude and adjusted odds ratio and their 95% confidence intervals of all variables included in the final model (Supplementary Table S2). Sensitivity analysis considering complete cases is also presented as supplemental material (Supplementary Table S3). The findings were similar, with the addition that previous transverse incision in abdominal wall (Pfannenstiel incision) appear to confer an independent decrease in the odds of SPE by approximately 56%. The results of the Hosmer-Lemeshow test indicated that our model was well calibrated (Supplementary Figure S1), and the discrimination was satisfactory (pseudo-R^2 = .278, Somers' D = .678, p = .625).
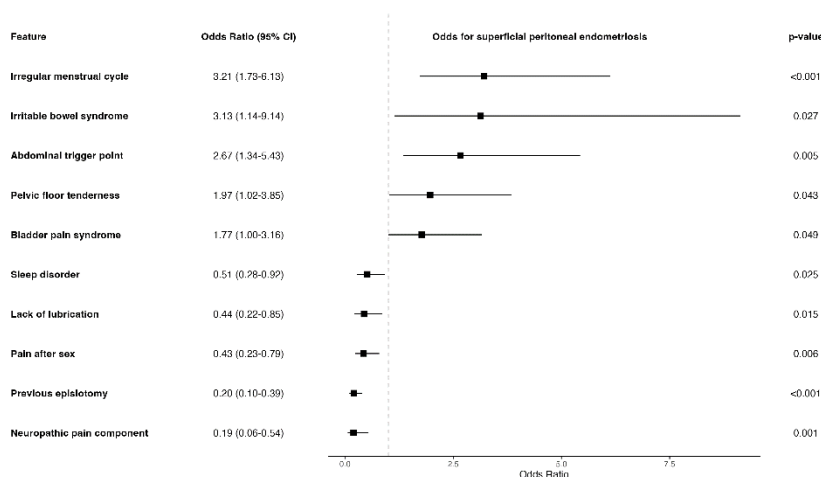

Figure 2: Forest plot displaying the adjusted odds ratios for variables independently associated with the diagnosis of superficial peritoneal endometriosis and the confidence interval of the estimate.

The performance of the models during training is available in the supplementary material (Supplementary Figure S2). In this phase, all models showed a general trend of outperforming logistic regression. The top-performing individual classifiers were the Support Vector Machine (SVM) and Naive Bayes, with the SVM reaching an F1 score of 63.0%, Recall of 58.6%, and Specificity of 79.3%, while Naive Bayes achieved an F1 score of 59.7%, Recall of 74.2%, but with a Specificity of only 19.4%. The SVM displayed good specificity, whereas Naive Bayes showed strong recall, albeit with limited specificity. The best-performing Soft Voting ensemble combined the Extreme Gradient Boosting and Naive Bayes models, attaining the highest F1 score of 64.8% and Recall of 79.3%, albeit with a specificity of only 38.7%. Table 1 presents the detailed performance metrics for each individual model and the highest-performing ensemble. At fixed specificity levels of 70% and 75%, the SVM emerged as the best classifier, achieving recall rates of 65.5% and 62.1%, respectively. All other classifiers showed recall values below 60% under these fixed specificity conditions (Supplementary Table S4).

Table 1. Performance measurements sorted by F1 Score

Classifiers	Performance measurements (%)					
	F1 Score	Recall	Specificity	Accuracy	Precision	AUC[1]
Soft Voting Ensemble[2]	64.8	79.3	38.7	59.0	54.8	61.7
Support Vector Machine	63.0	58.6	74.2	66.4	68.0	66.2
Naive Bayes	59.7	79.3	19.4	49.3	47.9	64.3
Multilayer Perceptron	59.6	58.6	64.5	61.6	60.7	66.2
Extremely Randomised Trees	56.1	55.2	61.3	58.2	57.1	65.1
Logistic Regression	55.6	51.7	67.7	59.7	60.0	67.7
Stacking Ensemble[3]	54.5	51.7	58.1	58.1	57.7	65.2
Categorical Boosting	54.5	51.7	64.5	58.1	57.7	62.6
Adaptive Boosting	54.2	55.2	54.8	55.0	53.3	60.3
Extreme Gradient Boosting	53.3	55.2	51.6	53.4	51.6	66.1
Light Gradient Boosting	53.3	55.2	51.6	53.4	51.6	64.5
Random Forest	49.1	44.8	64.5	54.7	54.2	59.0

[1]Area under the curve.
[2]Extreme Gradient Boosting plus Naive Bayes
[3]Categorical Boosting plus Naive Bayes with Logistic Regression as the final estimator

Figure 3 showcases the feature importance among the top classifiers, with logistic regression included for comparison. Violin plots display Shapley values (log-odds units) for (A) the soft voting classifier, (B) the support vector classifier, and (C) logistic regression, ordered by decreasing mean absolute values. These plots summarise feature importance, distribution, and variability, with wider sections indicating higher value density. Colours represent feature values: red for presence/high values and blue for absence/low values. The top ten predictive features for each model are highlighted, with previous pelvic inflammatory disease reports showing concentrated Shapley values favouring SPE prediction in the support vector and logistic regression classifiers, without significant outliers.
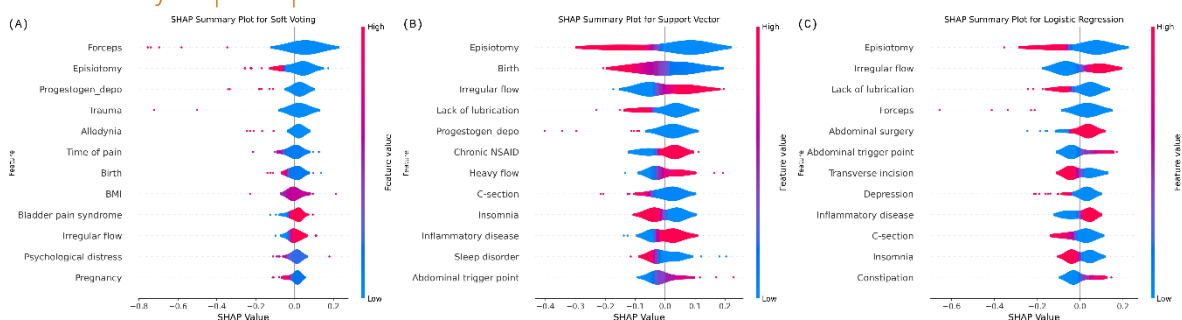
Figure 3: Violin plots of the Shapley values computed for each feature considering the best models (Soft Voting and Support Vector Machine) and Logistic Regression for comparison.

Supplementary Figure S3 illustrates how the feature set influences the probability of outcome for a random participant in the test dataset across each model. Waterfall plots provide explanations for individual predictions: starting from the model's expected output over the reference dataset, each row indicates how features contribute positively (red) or negatively (blue) to adjust the prediction for the specific case. Waterfall plots illustrate variable weighting by the Support Vector (top) and Soft Voting (bottom) models for correctly classifying participants with and without SPE. Arrows show how each feature's contribution shifts the model's initial outcome, highlighting personalised predictor combinations for each patient.

Discussion

Our results from the logistic regression model showed that irregular menstrual cycle, irritable bowel syndrome, bladder pain syndrome, abdominal trigger point, and pelvic floor tenderness are independently associated with superficial peritoneal endometriosis. Conversely, symptoms such as sleep disorder, lack of lubrication, pain after sex, previous episiotomy, and neuropathic components tend to suggest the absence of the disease, though not entirely ruling it out. Shapley values additionally highlight that a history of pelvic inflammatory disease may play a significant role in predicting SPE.

The literature linking endometriosis and alterations in menstrual bleeding patterns is relatively extensive. Although a clear explanation for this phenomenon is not entirely understood, various pathophysiological processes may be responsible for the observed changes in menstrual flow patterns, such as the endometrial immune and inflammatory environment[59], matrix metalloproteinases activity[60], and progesterone resistance[61]. Furthermore, there is a well-documented association between endometriosis and what we have termed here as "visceral syndromes"[62], including irritable bowel syndrome[63] and bladder pain syndrome[64], as well as "myofascial syndromes"[65], such as abdominal and pelvic floor tenderness[66]. These conditions are linked to alterations in the central processing of sensory input from the gut[67], bladder[68], and myofascial system[69], as well as endometriosis. Central sensitisation has been increasingly recognised as a crucial component in the pathogenesis of endometriosis-associated pain, occurring alongside peripheral nociceptive contributors on a continuum[70,71] and may be pivotal in post-surgical outcomes[72]. Additionally, specific features of peritoneal innervation in individuals with endometriosis, such as nociceptor signalling dysregulation, neurogenic and peripheral sensitisation, demonstrate that these central findings also have a corresponding local peripheral neurological component[73]. Our data suggest that the constellation of these symptoms, representing a potential convergent impairment of corresponding viscerotomes and myotomes, could be instrumental in predicting the intraoperative diagnosis of endometriosis. The association of endometriosis with pelvic inflammatory disease, although controversial, has also been reported from both a clinical[74] and transcriptomic perspective[75].

Our study also identified factors associated with the absence of SPE. A history of episiotomy,

despite its potential to cause pelvic pain[76,77], is rarely linked to perineal scar endometriosis[78]. The presence of a neuropathic component in pain symptoms may also suggest the absence of SPE, acknowledging that neuropathic-like pain can occur in endometriosis but is typically associated with extensive lesions involving nervous tissue[79]. Vaginal dryness, often studied in postmenopausal women, lacks clear links to endometriosis in premenopausal individuals, potentially influenced by hormonal changes and psychological factors[80,81]. Similarly, sleep disturbances, commonly observed in endometriosis patients, may be exacerbated by pain but lack direct evidence linking them to the disease[82–84]. Therefore, these symptoms, when present without other supporting evidence of endometriosis beyond dysmenorrhea and pelvic pain, especially in the absence of imaging abnormalities, suggest the possibility of not finding identifiable lesions during surgical exploration.

Utilising a variety of machine learning algorithms for prediction offers the benefit of enhanced predictive capability and, in some cases, interpretability[56]. While logistic regression remains widely accepted due to its explainability, our findings revealed its discriminatory and predictive performance to be inferior compared to other models, achieving approximately 60% accuracy and precision, with a sensitivity of 51.7% and specificity of 67.7% for diagnosing the condition. Notably, each algorithm demonstrated distinct performance characteristics, and ensemble methods like Soft Voting led to overall performance gains. Specifically, combining two moderate-performing models within the Soft Voting framework yielded significant improvements in sensitivity by capturing nuanced aspects of data that were insufficient when analysed in isolation. Among individual classifiers, the SVM exhibited standout specificity, achieving 74.2%, and demonstrated superior recall compared to other models, reaching 65%. This balance contributed positively to its F1 score and highlighted its capacity to maintain a strong trade-off between

sensitivity and precision. The SVM's strong specificity is particularly beneficial in clinical and screening contexts, where minimising false positives is critical to reducing unnecessary interventions, patient anxiety, and the operational costs of evaluating individuals without the condition of interest. While the SVM's recall is not the highest possible, it remains superior relative to other models tested, striking a meaningful balance in performance for diagnostic applications.

The use of artificial intelligence in healthcare is rapidly expanding. Bendifallah et al.[85] highlighted ML's potential in diagnosing endometriosis, focusing on women with presumed diagnoses and higher disease stages detectable by imaging. In contrast, our study addresses the challenge of predicting endometriosis in symptomatic women without typical imaging findings and unresponsive to hormonal treatment. We emphasise the importance of surgical confirmation at this experimental stage for accurate diagnosis and scientific rigour.

Our method of evaluating feature importance using Shapley values provides valuable insights into the individualised predictive relevance of clinical factors, as illustrated in Figure 3. This approach reveals that the model identifies a unique set of predictive features for each participant, highlighting its ability to handle complex scenarios where linear relationships between predictor variables and outcomes are not straightforward, unlike logistic regression. This capability supports personalised reporting, aligning with the growing trend towards personalised medicine. It aids clinicians in understanding the intricate relationships between clinical variables and provides objective support to help patients make informed decisions about surgery.

Despite our study employing a contemporary and rigorous methodology, and utilising a well-documented database collected prospectively in a specialised reference hospital by a team of experts, it has limitations. Unfortunately, the database of individuals with these characteristics is small. One of the primary challenges is the reliance on large,

high-quality datasets for training ML algorithms. Initiatives to standardise the recording of patient history and characteristics, as well as the recording of surgical procedures and extent of disease, as proposed by the World Endometriosis Research Foundation with the Endometriosis Phenome and Biobanking Harmonisation Project[86,87], will undoubtedly help compile a relevant case series in the near future, also allowing for external validation with reliable international data.

The generalisability of model predictions to diverse populations remains a critical challenge, as clinical characteristics and symptom profiles can vary significantly across geographical regions, ethnicities, and patient demographics. While we conducted extensive internal validation, the model developed here lacks external validation, particularly in varied clinical settings. To address potential variability in predictors, we prioritised the use of nominal variables wherever feasible, especially for tools like IPAQ, DN-4, and SRQ-20. Although this approach sacrifices some granularity, it ensures these variables are considered at least nominally. However, the wide range of instruments and methodologies used to assess these factors introduces variability across studies and represents a potential source of bias in predictive models. Furthermore, while ML models hold great promise, they are susceptible to challenges such as overfitting and the need for continuous updates to accommodate evolving clinical data. Addressing these issues requires rigorous external validation in real-world settings to ensure the robustness and reliability of predictions across diverse patient populations.

In practical clinical settings, the application of ML models requires careful consideration of how clinicians can interpret and act on the predictions. While ML can provide valuable insights, the role of the clinician in validating and contextualising these predictions remains crucial. Clear guidance on how to integrate ML predictions with clinical expertise will be necessary to facilitate decision-making in areas such as patient counselling, treatment planning, and follow-up care. For instance, ML models can help clinicians identify high-risk patients for SPE or guide the choice of treatment options, but they should not replace the clinician's judgement. As such, the implementation of ML tools in clinical practice should always be accompanied by clear protocols for interpreting and acting on the predictions within the context of individual patient circumstances. Another significant limitation is the difficulty in establishing a secure temporal relationship between the disease and potential predictors in a retrospective study. Consequently, it is not possible to make any extrapolations about potential causalities. On the other hand, our study also generates intriguing additional hypotheses for future investigation. Upon examining Figure 3, an interesting observation emerges: we observe in image (B) that the use of depot progestogen (1=Progestogen _depo) negatively contributes to the identification of endometriosis. Conversely, in images (C) and (D), the absence of the medication (0=Progestogen _depo) positively contributes to the identification of the disease. The critical issue here is whether depot progestogen can significantly reduce endometriosis foci and hinder its macroscopic identification, given the potential presence of occult microscopic endometriosis undetectable by laparoscopy in apparently normal peritoneum[88]. Our study does not provide an answer to this question, nor does the current literature offer further explanations[89]. Nevertheless, it raises an important point.

In light of the Prediction model Risk Of Bias Assessment Tool (PROBAST), our model, while presenting a certain risk of bias, demonstrates low concerns regarding applicability. We view it as a proof-of-concept that encourages multicentre studies focusing on this population. Moreover, it represents an important step towards standardising measurement tools and predictor definitions, while respecting the sociocultural and geographical diversity inherent to global healthcare settings. By fostering collaboration and harmonisation, this work aims to pave the way for more generalisable and inclusive predictive models in clinical practice.

Conclusions

In conclusion, our study identifies irregular menstrual cycles, irritable bowel syndrome, bladder pain syndrome, abdominal trigger points, and pelvic floor tenderness as independent factors associated with intraoperative findings of SPE. Moreover, other variables, such as a history of pelvic inflammatory disease, may further aid in predicting preoperative diagnoses in a personalised manner. We propose that ML approaches show significant potential for predicting SPE by analysing pre-operative clinical data, particularly in patients presenting with severe dysmenorrhoea, acyclic pelvic pain unresponsive to hormonal treatment, and normal ultrasonographic results. This predictive tool could enhance patient counselling and inform surgical decision-making, fostering tailored and more effective care strategies. Regardless of laparoscopy outcomes, it remains crucial that informed consent comprehensively addresses the probabilities of disease presence and the potential short- and long-term implications of surgical interventions. This ensures that patients are thoroughly informed and empowered to make well-grounded decisions regarding their treatment options.

Conflict of Interest:

The authors have no conflicts of interest to declare.

Attestation statements:

Data regarding any of the subjects in the study has not been previously published.
Data will be made available to the editors of the journal pre and/or post publication for review or query upon request.

Data sharing statement:

The data that support the findings of this study are electronic health records from patients assisted on the HCFMRP-USP. The access to these data are protected by The General Data Protection Regulation, Law n° 13,709, dated August 14, 2018. Anonymous datasheet is available from the coordinator, [OBPN], upon reasonable request.

ORCID ID

1. Letícia Luiza Alves Santos, MD,
ORCID:0000-0002-2060-5205;
E-mail: leticialuiza06@gmail.com
2. Mateus Carvalho de Azevedo, MD,
ORCID:0000-0001-9962-6799;
E-mail: mateuscarvalhoazevedo@usp.br
3. Lia Keiko Shimamura, MD,
ORCID:0000-0002-1755-5067;
E-mail: keikoacademico1@gmail.com
4. Antonio Alberto Nogueira, MD, PhD,
ORCID: 0000-0003-0248-9463;
E-mail: aanoguei@fmrp.usp.br
5. Francisco José Candido-dos-Reis, MD, PhD,
ORCID: 0000-0001-5758-5917;
E-mail: fjcreis@fmrp.usp.br

6. Eduardo Schor, MD, PhD,
ORCID: 0000-0002-6141-320X;
E-mail: eduschor@gmail.com

7. Julio Cesar Rosa-e-Silva, MD, PhD,
ORCID:0000-0001-6930-1328;
E-mail: juliocrs@usp.br

8. Daniel Guimarães Tiezzi, MD, PhD,
ORCID:0000-0002-2660-0093;
E-mail: dtiezzi@usp.br

9. Omero Benedicto Poli-Neto, MD, PhD,
ORCID:0000-0002-3497-3367;
E-mail: polineto@usp.br

References:

1. Kristjansdottir A, Rafnsson V, Geirsson RT. Comprehensive evaluation of the incidence and prevalence of surgically diagnosed pelvic endometriosis in a complete population. *Acta Obstet Gynecol Scand.* 2023;102(10):1329-1337. doi:10.1111/aogs.14556

2. Christ JP, Yu O, Schulze-Rath R, Grafton J, Hansen K, Reed SD. Incidence, prevalence, and trends in endometriosis diagnosis: a United States population-based study from 2006 to 2015. *Am J Obstet Gynecol.* 2021;225(5):500.e1-500.e9. doi: 10.1016/j.ajog.2021.06.067

3. Sarria-Santamera A, Orazumbekova B, Terzic M, Issanov A, Chaowen C, Asúnsolo-del-Barco A. Systematic Review and Meta-Analysis of Incidence and Prevalence of Endometriosis. *Healthcare.* 2020;9(1):29. doi:10.3390/healthcare9010029

4. Eisenberg V, Weil C, Chodick G, Shalev V. Epidemiology of endometriosis: a large population-based database study from a healthcare provider with 2 million members. *BJOG Int J Obstet Gynaecol.* 2018;125(1):55-62. doi:10. 1111/1471-0528.14711

5. Zondervan KT, Becker CM, Koga K, Missmer SA, Taylor RN, Viganò P. Endometriosis. *Nat Rev Dis Primer.* 2018;4(1):9. doi:10.1038/s41572-018-0008-5

6. Audebert A, Bäckström T, Barlow DH, et al. Endometriosis 1991: a discussion document. *Hum Reprod Oxf Engl.* 1992;7(3):432-435. doi:10.1093/ oxfordjournals.humrep.a137665

7. Cramer DW, Missmer SA. The Epidemiology of Endometriosis. *Ann N Y Acad Sci.* 2002;955(1):11-22. doi:10.1111/j.1749-6632.2002.tb02761.x

8. Della Corte L, Di Filippo C, Gabrielli O, et al. The Burden of Endometriosis on Women's Lifespan: A Narrative Overview on Quality of Life and Psychosocial Wellbeing. *Int J Environ Res Public Health.* 2020;17(13):4683. doi:10.3390/ijerph17134683

9. Missmer SA, Tu FF, Agarwal SK, et al. Impact of Endometriosis on Life-Course Potential: A Narrative Review. *Int J Gen Med.* 2021;14:9-25. doi:10.2147/IJGM.S261139

10. Nnoaham KE, Hummelshoj L, Webster P, et al. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. *Fertil Steril.* 2011;96(2):366-373.e8. doi:10.1016/j.fertnstert.2011.05.090

11. Simoens S, Dunselman G, Dirksen C, et al. The burden of endometriosis: costs and quality of life of women with endometriosis and treated in referral centres. *Hum Reprod.* 2012;27(5):1292-1299. doi:10.1093/humrep/des073

12. Soliman AM, Taylor H, Bonafede M, Nelson JK, Castelli-Haley J. Incremental direct and indirect cost burden attributed to endometriosis surgeries in the United States. *Fertil Steril.* 2017;107(5):1181-1190.e2. doi:10.1016/j.fertnstert.2017.03.020

13. Soliman AM, Fuldeore M, Snabes MC. Factors Associated with Time to Endometriosis Diagnosis in the United States. *J Womens Health.* 2017; 26(7):788-797. doi:10.1089/jwh.2016.6003

14. Agarwal SK, Chapron C, Giudice LC, et al. Clinical diagnosis of endometriosis: a call to action. *Am J Obstet Gynecol.* 2019;220(4):354.e1-354.e12 . doi:10.1016/j.ajog.2018.12.039

15. Sims OT, Gupta J, Missmer SA, Aninye IO. Stigma and Endometriosis: A Brief Overview and Recommendations to Improve Psychosocial Well-Being and Diagnostic Delay. *Int J Environ Res Public Health.* 2021;18(15):8210. doi:10.3390/ ijerph18158210

16. Chapron C, Tosti C, Marcellin L, et al. Relationship between the magnetic resonance imaging appearance of adenomyosis and endometriosis phenotypes. *Hum Reprod.* 2017;32 (7):1393-1401. doi:10.1093/humrep/dex088

17. Avery JC, Deslandes A, Freger SM, et al. Noninvasive diagnostic imaging for endometriosis part 1: a systematic review of recent developments in ultrasound, combination imaging, and artificial intelligence. *Fertil Steril.* 2024;121(2):164-188. doi:10.1016/j.fertnstert.2023.12.008

18. Kuznetsov L, Dworzynski K, Davies M, Overton C, Guideline Committee. Diagnosis and management of endometriosis: summary of NICE guidance.

BMJ. 2017;358:j3935. doi:10.1136/bmj.j3935

19. Hirsch M, Begum M, Paniz É, Barker C, Davis C, Duffy J. Diagnosis and management of endometriosis: a systematic review of international and national guidelines. *BJOG Int J Obstet Gynaecol.* 2018;125(5):556-564. doi:10.1111/1471 -0528.14838

20. Becker CM, Bokor A, Heikinheimo O, et al. ESHRE guideline: endometriosis. *Hum Reprod Open.* 2022;2022(2):hoac009. doi:10.1093/hropen /hoac009

21. Goncalves MO, Siufi Neto J, Andres MP, Siufi D, de Mattos LA, Abrao MS. Systematic evaluation of endometriosis by transvaginal ultrasound can accurately replace diagnostic laparoscopy, mainly for deep and ovarian endometriosis. *Hum Reprod Oxf Engl.* 2021;36(6):1492-1500. doi:10.1093/ humrep/deab085

22. Leonardi M, Robledo KP, Espada M, Vanza K, Condous G. SonoPODography: A new diagnostic technique for visualizing superficial endometriosis. *Eur J Obstet Gynecol Reprod Biol.* 2020;254:124-131. doi:10.1016/j.ejogrb.2020.08.051

23. Pedrassani M, Guerriero S, Pascual MÁ, et al. Superficial Endometriosis at Ultrasound Examination— A Diagnostic Criteria Proposal. *Diagnostics.* 2023 ;13(11):1876. doi:10.3390/diagnostics13111876

24. Chen-Dixon K, Uzuner C, Mak J, Condous G. Effectiveness of ultrasound for endometriosis diagnosis. *Curr Opin Obstet Gynecol.* 2022;34(5) :324-331. doi:10.1097/GCO.0000000000000812

25. Horne AW, Daniels J, Hummelshoj L, Cox E, Cooper KG. Surgical removal of superficial peritoneal endometriosis for managing women with chronic pelvic pain: time for a rethink? *BJOG Int J Obstet Gynaecol.* 2019;126(12):1414-1416. doi:10.1111/1471-0528.15894

26. Reis FM, Santulli P, Marcellin L, Borghese B, Lafay-Pillet MC, Chapron C. Superficial Peritoneal Endometriosis: Clinical Characteristics of 203 Confirmed Cases and 1292 Endometriosis-Free Controls. *Reprod Sci.* 2020;27(1):309-315. doi:10. 1007/s43032-019-00028-1

27. Kalaitzopoulos DR, Samartzis N, Kolovos GN, et al. Treatment of endometriosis: a review with comparison of 8 guidelines. *BMC Womens Health.* 2021;21:397. doi:10.1186/s12905-021-01545-5

28. Jarrell J, Arendt-Nielsen L. Negative laparoscopy unveiled. *J Endometr Pelvic Pain Disord.* 2018; 10(1):18-21. doi:10.1177/2284026517749478

29. Tempest N, Efstathiou E, Petros Z, Hapangama DK. Laparoscopic Outcomes after Normal Clinical and Ultrasound Findings in Young Women with Chronic Pelvic Pain: A Cross-Sectional Study. *J Clin Med.* 2020;9(8). doi:10.3390/jcm 9082593

30. Steele LA, Mooney SS, Gilbee ES, Grover SR. When you see nothing at all: Outcomes following a negative laparoscopy. A systematic review. *Aust N Z J Obstet Gynaecol.* 2024;64(2):95-103. doi:10.1111/ajo.13749

31. Ahmad G, Gent D, Henderson D, O'Flynn H, Phillips K, Watson A. Laparoscopic entry techniques. *Cochrane Database Syst Rev.* 2015;8:C D006583. doi:10.1002/14651858.CD006583.pub4

32. Mackenzie SC, Stephen J, Williams L, et al. Effectiveness of laparoscopic removal of isolated superficial peritoneal endometriosis for the management of chronic pelvic pain in women (ESPriT2): protocol for a multi-centre randomised controlled trial. *Trials.* 2023;24(1):425. doi:10.1186 /s13063-023-07386-x

33. Mak J, Leonardi M, Condous G. 'Seeing is believing': arguing for diagnostic laparoscopy as a diagnostic test for endometriosis. *Reprod Fertil.* 2022;3(3):C23-C28. doi:10.1530/RAF-21-0117

34. Santana AN, de Santana CN, Montoya P. Chronic Pain Diagnosis Using Machine Learning, Questionnaires, and QST: A Sensitivity Experiment. *Diagnostics.* 2020;10(11):958. doi:10.3390/diagno stics10110958

35. Nezhat C, Armani E, Chen HCC, Najmi Z, Lindheim SR, Nezhat C. Use of the Free Endometriosis Risk Advisor App as a Non-Invasive Screening Test for Endometriosis in Patients with Chronic Pelvic Pain and/or Unexplained Infertility.

J Clin Med. 2023;12(16):5234. doi:10.3390/jcm 12165234

36. Dungate B, Tucker DR, Goodwin E, Yong PJ. Assessing the Utility of artificial intelligence in endometriosis: Promises and pitfalls. *Womens Health.* 2024;20:17455057241248121. doi:10.117 7/17455057241248121

37. Anastasiu CV, Moga MA, Elena Neculau A, et al. Biomarkers for the Noninvasive Diagnosis of Endometriosis: State of the Art and Future Perspectives. *Int J Mol Sci.* 2020;21(5):1750. doi:10.3390/ijms21051750

38. Goldstein A, Cohen S. Self-report symptom-based endometriosis prediction using machine learning. *Sci Rep.* 2023;13:5499. doi:10.1038/s41 598-023-32761-8

39. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open.* 2021; 11(7):e048008. doi:10.1136/bmjopen-2020-048008

40. Kennedy S, Bergqvist A, Chapron C, et al. ESHRE guideline for the diagnosis and treatment of endometriosis. *Hum Reprod Oxf Engl.* 2005; 20(10):2698-2704. doi:10.1093/humrep/dei135

41. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc.* 2003;35(8):1381-1395. doi:10.1249/01.MS S.0000078924.61453.FB

42. Peixoto SV, Mambrini JV de M, Firmo JOA, et al. Physical activity practice among older adults: results of the ELSI-Brazil. *Rev Saúde Pública.* 2018;52(Suppl 2):5s. doi:10.11606/S1518-8787.20 18052000605

43. Santos JG, Brito JO, de Andrade DC, et al. Translation to Portuguese and Validation of the Douleur Neuropathique 4 Questionnaire. *J Pain.* 2010;11(5):484-490. doi:10.1016/j.jpain.2009.09.014

44. Mari JJ, Williams P. A validity study of a psychiatric screening questionnaire (SRQ-20) in primary care in the city of Sao Paulo. *Br J Psychiatry J Ment Sci.* 1986;148:23-26.

45. Longstreth GF, Thompson WG, Chey WD, Houghton LA, Mearin F, Spiller RC. Functional Bowel Disorders. *Gastroenterology.* 2006;130(5): 1480-1491. doi:10.1053/j.gastro.2005.11.061

46. Park S, Lim J, Ko Y, et al. Diagnosis of Pelvic Congestion Syndrome Using Transabdominal and Transvaginal Sonography | AJR. *Am J Roentgenol.* 2004;182(3):683-688. doi:10.2214/ajr.182.3.1820683

47. Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. Descriptive Statistics and Normality Tests for Statistical Data. *Ann Card Anaesth.* 2019;22(1):67-72. doi:10.4103/aca.ACA_157_18

48. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol.* 2019;110:63-73. doi:10.1016/j.jclinepi .2019.02.016

49. Buuren SV, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw.* 2011;45(3). doi:10.18637/jss.v045.i03

50. Kim JH. Multicollinearity and misleading statistical results. *Korean J Anesthesiol.* 2019;72 (6):558-569. doi:10.4097/kja.19087

51. Baak M, Koopman R, Snoek H, Klous S. A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics. Published online March 9, 2019. doi:10.48550/ arXiv.1811.11440

52. Heinze G, Wallisch C, Dunkler D. Variable selection - A review and recommendations for the practicing statistician. *Biom J Biom Z.* 2018;60 (3):431-449. doi:10.1002/bimj.201700067

53. Nattino G, Pennell ML, Lemeshow S. Assessing the goodness of fit of logistic regression models in large samples: A modification of the Hosmer-Lemeshow test. *Biometrics.* 2020;76(2):54 9-560. doi:10.1111/biom.13249

54. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures.

Epidemiol Camb Mass. 2010;21(1):128-138. doi: 10.1097/EDE.0b013e3181c30fb2

55. Victoria AH, Maragatham G. Automatic tuning of hyperparameters using Bayesian optimization. *Evol Syst.* 2021;12(1):217-223. doi:10.1007/s12530 -020-09345-2

56. Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble Learning for Disease Prediction: A Review. *Healthc Basel Switz.* 2023;11(12):1808. doi:10.3390/healthcare11121808

57. Rodríguez-Pérez R, Bajorath J. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *J Comput Aided Mol Des.* 2020;34(10):1013-1026. doi:10.1007/s10822-020-00314-0

58. Bisong E. Google Colaboratory. In: Bisong E, ed. *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners.* Apress; 2019:59-64. doi:10.1007/978-1-4842-4470-8_7

59. Vallvé-Juanico J, Houshdaran S, Giudice LC. The endometrial immune environment of women with endometriosis. *Hum Reprod Update.* 2019; 25(5):565-592. doi:10.1093/humupd/dmz018

60. Ke J, Ye J, Li M, Zhu Z. The Role of Matrix Metalloproteinases in Endometriosis: A Potential Target. *Biomolecules.* 2021;11(11):1739. doi:10.33 90/biom11111739

61. Tang HC, Lin TC, Wu MH, Tsai SJ. Progesterone resistance in endometriosis: A pathophysiological perspective and potential treatment alternatives. *Reprod Med Biol.* 2024; 23(1):e12588. doi:10.1002/rmb2.12588

62. Hansen KE, Kesmodel US, Baldursson EB, Kold M, Forman A. Visceral syndrome in endometriosis patients. *Eur J Obstet Gynecol Reprod Biol.* 2014;179:198-203. doi:10.1016/j.ejo grb.2014.05.024

63. Nabi MY, Nauhria S, Reel M, et al. Endometriosis and irritable bowel syndrome: A systematic review and meta-analyses. *Front Med.* 2022;9:914356. doi:10.3389/fmed.2022.914356

64. Wu CC, Chung SD, Lin HC. Endometriosis increased the risk of bladder pain syndrome/interstitial cystitis: A population-based study. *Neurourol Urodyn.* 2018;37(4):1413-1418. doi:10.1002/nau.23462

65. Jarrell J. Endometriosis and abdominal myofascial pain in adults and adolescents. *Curr Pain Headache Rep.* 2011;15(5):368-376. doi:10. 1007/s11916-011-0218-y

66. Shafrir AL, Martel E, Missmer SA, et al. Pelvic floor, abdominal and uterine tenderness in relation to pressure pain sensitivity among women with endometriosis and chronic pelvic pain. *Eur J Obstet Gynecol Reprod Biol.* 2021;264:247-253. doi:10.1016/j.ejogrb.2021.07.029

67. Mayer EA, Ryu HJ, Bhatt RR. The neurobiology of irritable bowel syndrome. *Mol Psychiatry.* 2023;2 8(4):1451-1465. doi:10.1038/s41380-023-01972-w

68. Knox S, Offiah I, Hashim H. Evaluation of Central Sensitisation in Bladder Pain Syndrome: A Systematic Review. *Int Urogynecology J.* Published online May 7, 2024. doi:10.1007/s00192-024-05793-5

69. Stratton P, Khachikyan I, Sinaii N, Ortiz R, Shah J. Association of chronic pelvic pain and endometriosis with signs of sensitization and myofascial pain. *Obstet Gynecol.* 2015;125(3):719-728. doi:10.1097/AOG.0000000000000663

70. Coxon L, Horne AW, Vincent K. Pathophysiology of endometriosis-associated pain: a review of pelvic and central nervous system mechanisms. 2018;51:53-67. doi:10.1016/j.bpobg yn.2018.01.014

71. Morotti M, Vincent K, Becker CM. Mechanisms of pain in endometriosis. *Eur J Obstet Gynecol Reprod Biol.* 2017;209:8-13. doi:10.1016/j.ejogrb .2016.07.497

72. Orr NL, Huang AJ, Liu YD, et al. Association of Central Sensitization Inventory Scores With Pain Outcomes After Endometriosis Surgery. *JAMA Netw Open.* 2023;6(2):e230780. doi:10.1001/jama networkopen.2023.0780

73. Godin SK, Wagner J, Huang P, Bree D. The role of peripheral nerve signaling in endometriosis.

FASEB BioAdvances. 2021;3(10):802-813. doi:10.1096/fba.2021-00063

74. Tai FW, Chang C, Chiang JH, Lin WC, Wan L. Association of Pelvic Inflammatory Disease with Risk of Endometriosis: A Nationwide Cohort Study Involving 141,460 Individuals. *J Clin Med*. 2018;7(11):379. doi:10.3390/jcm7110379

75. Poli-Neto OB, Carlos D, Favaretto A, Rosa-e-Silva JC, Meola J, Tiezzi D. Eutopic endometrium from women with endometriosis and chlamydial endometritis share immunological cell types and DNA repair imbalance: A transcriptome meta-analytical perspective. *J Reprod Immunol*. 2021;145. doi:10.1016/j.jri.2021.103307

76. Quoc Huy NV, Phuc An LS, Phuong LS, Tam LM. Pelvic Floor and Sexual Dysfunction After Vaginal Birth With Episiotomy in Vietnamese Women. *Sex Med*. 2019;7(4):514-521. doi:10.1016/j.esxm.2019.09.002

77. Gün İ, Doğan B, Özdamar Ö. Long- and short-term complications of episiotomy. *Turk J Obstet Gynecol*. 2016;13(3):144-148. doi:10.4274/tjod.00087

78. Shanmuga Jayanthan S, Shashikala G, Arathi N. Perineal scar endometriosis. *Indian J Radiol Imaging*. 2019;29(4):457-461. doi:10.4103/ijri.IJRI_366_19

79. Coxon L, Wiech K, Vincent K. Is There a Neuropathic-Like Component to Endometriosis-Associated Pain? Results From a Large Cohort Questionnaire Study. *Front Pain Res Lausanne Switz*. 2021;2:743812. doi:10.3389/fpain.2021.743812

80. Goncharenko V, Bubnov R, Polivka J, et al. Vaginal dryness: individualised patient profiles, risks and mitigating measures. *EPMA J*. 2019;10(1):73-79. doi:10.1007/s13167-019-00164-3

81. Mark JKK, Samsudin S, Looi I, Yuen KH. Vaginal dryness: a review of current understanding and management strategies. *Climacteric J Int Menopause Soc*. 2024;27(3):236-244. doi:10.1080/13697137.2024.2306892

82. Chaichian S, Mehdizadehkashi A, Haghgoo A, et al. Sleep disorders in patients with endometriosis; a cross-sectional study. *BMC Womens Health*. 2024;24(1):340. doi:10.1186/s12905-024-03185-x

83. Ding W, Yang L, Shi E, et al. The endocannabinoid N-arachidonoyl dopamine is critical for hyperalgesia induced by chronic sleep disruption. *Nat Commun*. 2023;14(1):6696. doi:10.1038/s41467-023-42283-6

84. Irwin MR. Sleep and inflammation: partners in sickness and in health. *Nat Rev Immunol*. 2019;19(11):702-715. doi:10.1038/s41577-019-0190-z

85. Bendifallah S, Puchar A, Suisse S, et al. Machine learning algorithms as new screening approach for patients with endometriosis. *Sci Rep*. 2022;12(1):639. doi:10.1038/s41598-021-04637-2

86. Becker CM, Laufer MR, Stratton P, et al. World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: I. Surgical phenotype data collection in endometriosis research. *Fertil Steril*. 2014;102(5):1213-1222. doi:10.1016/j.fertnstert.2014.07.709

87. Vitonis AFAF, Vincent K, Rahmioglu N, et al. World Endometriosis Research Foundation Endometriosis Phenome and biobanking harmonization project: II. Clinical and covariate phenotype data collection in endometriosis research. *Fertil Steril*. 2014;102(5):1223-1232. doi:10.1016/j.fertnstert.2014.07.1244

88. Khan KN, Fujishita A, Kitajima M, Hiraki K, Nakashima M, Masuzaki H. Occult microscopic endometriosis: undetectable by laparoscopy in normal peritoneum. *Hum Reprod Oxf Engl*. 2014;29(3):462-472. doi:10.1093/humrep/det438

89. Brown J, Kives S, Akhtar M. Progestagens and anti-progestagens for pain associated with endometriosis. *Cochrane Database Syst Rev*. 2012;2012(3):CD002122.doi:10.1002/14651858.CD002122.pub2

Supplemental Tables

Supplementary Table S1. Participant characteristics.

Variables	All participants N = 298[1]	Endometriosis Yes N = 144 (48%)[1]	No N = 154 (52%)[1]	p-value[2]
Time of symptoms	79.2 [48.0] (6.0, 360.0)	72.7 [48.0] (6.0, 360.0)	85.2 [65.0] (7.0, 312.0)	0.097
Intensity of pain	68.4 [71.0] (30.0, 100.0)	69.4 [71.5] (30.0, 100.0)	67.5 [69.5] (30.0, 100.0)	0.402
Weekly symptoms	250 (83.9%)	112 (77.8%)	138 (89.6%)	0.005
Age at surgery	34.3 [34.0] (18.0, 47.0)	33.5 [34.0] (18.0, 47.0)	34.9 [35.5] (18.0, 47.0)	0.050
Age of menarche	12.3 [12.0] (8.0, 16.0)	12.2 [12.0] (9.0, 16.0)	12.3 [12.0] (8.0, 16.0)	0.487
Age of coitarche	17.4 [17.0] (12.0, 35.0)	17.6 [17.0] (12.0, 35.0)	17.2 [17.0] (12.0, 28.0)	0.323
Number of previous partners	2.4 [2.0] (1.0, 15.0)	2.5 [2.0] (1.0, 12.0)	2.3 [2.0] (1.0, 15.0)	0.349
Education level				0.003
Elementary	117 (39.3%)	44 (30.6%)	73 (47.4%)	
High	181 (60.7%)	100 (69.4%)	81 (52.6%)	
Civil status				0.227
Single	53 (17.8%)	32 (22.2%)	21 (13.6%)	
Married	216 (72.5%)	97 (67.4%)	119 (77.3%)	
Divorced	25 (8.4%)	13 (9.0%)	12 (7.8%)	
Widow	4 (1.3%)	2 (1.4%)	2 (1.3%)	
Paid job	205 (68.8%)	99 (68.8%)	106 (68.8%)	0.988
Tobacco exposure	38 (12.8%)	16 (11.1%)	22 (14.3%)	0.412
Alcohol misuse	19 (6.4%)	10 (6.9%)	9 (5.8%)	0.698
Sedentarism	10 (3.4%)	5 (3.5%)	5 (3.2%)	>0.999
Body mass index	26.7 [25.7] (16.8, 47.7)	26.2 [25.5] (16.8, 44.9)	27.3 [26.1] (18.4, 47.7)	0.069
Irregular menstrual cycle	111 (37.2%)	68 (47.2%)	43 (27.9%)	<0.001
Heavy bleeding	86 (28.9%)	48 (33.3%)	38 (24.7%)	0.099
Previous pregnancies	2.0 [2.0] (0.0, 13.0)	1.5 [1.0] (0.0, 8.0)	2.4 [2.0] (0.0, 13.0)	<0.001
Deliveries	1.6 [2.0] (0.0, 7.0)	1.2 [1.0] (0.0, 5.0)	2.0 [2.0] (0.0, 7.0)	<0.001
Miscarriages	0.4 [0.0] (0.0, 9.0)	0.4 [0.0] (0.0, 6.0)	0.4 [0.0] (0.0, 9.0)	0.961
Previous C-section	0.9 [0.5] (0.0, 4.0)	0.7 [0.0] (0.0, 3.0)	1.0 [1.0] (0.0, 4.0)	0.025
Previous episiotomy	101 (33.9%)	29 (20.1%)	72 (46.8%)	<0.001
Previous forceps	20 (6.7%)	2 (1.4%)	18 (11.7%)	<0.001
Previous abdominal surgery	215 (72.1%)	100 (69.4%)	115 (74.7%)	0.314
Previous transverse incision	151 (50.7%)	61 (42.4%)	90 (58.4%)	0.006
Infertility history	32 (10.7%)	24 (16.7%)	8 (5.2%)	0.001
Pelvic inflammatory disease	130 (43.6%)	69 (47.9%)	61 (39.6%)	0.148
Low back pain history	132 (44.3%)	61 (42.4%)	71 (46.1%)	0.516
Migraine diagnosis	120 (40.3%)	58 (40.3%)	62 (40.3%)	0.997
Depression diagnosis	65 (21.8%)	22 (15.3%)	43 (27.9%)	0.008
Previous sexual abuse	6 (2.0%)	2 (1.4%)	4 (2.6%)	0.685
Previous emotional abuse	48 (16.1%)	20 (13.9%)	28 (18.2%)	0.314
Previous physical violence	32 (10.7%)	12 (8.3%)	20 (13.0%)	0.195
Premenstrual syndrome	58 (19.5%)	31 (21.5%)	27 (17.5%)	0.384
Insomnia	128 (43.0%)	49 (34.0%)	79 (51.3%)	0.003
Hypersomnia	12 (4.0%)	7 (4.9%)	5 (3.2%)	0.479
Dyspareunia	203 (68.1%)	98 (68.1%)	105 (68.2%)	0.981

Variables	All participants N = 298[1]	Endometriosis Yes N = 144 (48%)[1]	No N = 154 (52%)[1]	p-value[2]
Vaginismus	40 (13.4%)	15 (10.4%)	25 (16.2%)	0.141
Pain after sex	198 (66.4%)	85 (59.0%)	113 (73.4%)	0.009
Pain after exercise	197 (66.1%)	89 (61.8%)	108 (70.1%)	0.129
Neuropathic pain component	28 (9.4%)	6 (4.2%)	22 (14.3%)	0.003
Avoidance of activities	126 (42.3%)	60 (41.7%)	66 (42.9%)	0.835
Psychological distress	7.6 [7.0] (0.0, 20.0)	7.0 [6.0] (0.0, 20.0)	8.1 [8.0] (0.0, 20.0)	0.039
Hypoactive desire	88 (29.5%)	37 (25.7%)	51 (33.1%)	0.160
Lack of lubrication	82 (27.5%)	32 (22.2%)	50 (32.5%)	0.048
Absence of orgasm	88 (29.5%)	37 (25.7%)	51 (33.1%)	0.160
Functional diarrhoea	12 (4.0%)	4 (2.8%)	8 (5.2%)	0.289
Functional constipation	115 (38.6%)	57 (39.6%)	58 (37.7%)	0.734
Irritable bowel syndrome	27 (9.1%)	17 (11.8%)	10 (6.5%)	0.110
Dyschesia	55 (18.5%)	32 (22.2%)	23 (14.9%)	0.105
Painful bladder syndrome	145 (48.7%)	78 (54.2%)	67 (43.5%)	0.066
Abdominal trigger point	74 (24.8%)	46 (31.9%)	28 (18.2%)	0.006
Abdominal allodynia	15 (5.0%)	5 (3.5%)	10 (6.5%)	0.233
Vulvodynia	14 (4.7%)	5 (3.5%)	9 (5.8%)	0.334
Pelvic floor tenderness	87 (29.2%)	48 (33.3%)	39 (25.3%)	0.129
Pelvic congestion syndrome	28 (9.4%)	8 (5.6%)	20 (13.0%)	0.028
Chronic NSAID use[3]	178 (59.7%)	95 (66.0%)	83 (53.9%)	0.034
Hormonal contraceptive				0.001
Progestogen_pill	99 (33.2%)	52 (36.1%)	47 (30.5%)	
Progestogen_depo	48 (16.1%)	12 (8.3%)	36 (23.4%)	
Progestogen_iud	14 (4.7%)	4 (2.8%)	10 (6.5%)	
Combined_pill	137 (46.0%)	76 (52.8%)	61 (39.6%)	
Antidepressants/gabapentinoid	70 (23.5%)	24 (16.7%)	46 (29.9%)	0.007

[1]Mean [Median] (0%, 100%); n (%)

[2]Wilcoxon rank sum test; Pearson's Chi-squared test; Welch Two Sample t-test; Fisher's exact test

[3]NSAID: non-steroidal anti-inflammatory drugs

Supplementary Table S2. Results of logistic regression model presenting variables independently associated with superficial peritoneal endometriosis diagnosis.

Characteristic	Univariate OR[1]	95% CI[1]	p-value	Multivariate OR[1]	95% CI[1]	p-value
Irregular menstrual cycle	2.31	1.43, 3.75	<0.001	3.21	1.73, 6.13	<0.001
Irritable bowel syndrome	1.93	0.87, 4.51	0.115	3.13	1.14, 9.14	0.027
Abdominal trigger point	2.11	1.24, 3.65	0.007	2.67	1.34, 5.43	0.005
Infertility history	3.65	1.65, 8.95	0.002	2.01	0.72, 6.11	0.19
Pelvic floor tenderness	1.47	0.89, 2.44	0.130	1.97	1.02, 3.85	0.043
Painful bladder syndrome	1.53	0.97, 2.43	0.066	1.77	1.00, 3.16	0.049
Chronic NSAID[2] use	1.66	1.04, 2.66	0.034	1.75	0.98, 3.14	0.057
Number of previous partners	1.06	0.94, 1.20	0.378	1.12	0.97, 1.30	0.13
Previous pregnancies	0.69	0.58, 0.81	<0.001	0.85	0.70, 1.03	0.10
Progestogen only	0.59	0.37, 0.93	0.023	0.80	0.43, 1.50	0.48
Previous transverse incision	0.52	0.33, 0.83	0.006	0.60	0.32, 1.09	0.094

Characteristic	Univariate			Multivariate		
	OR[1]	95% CI[1]	p-value	OR[1]	95% CI[1]	p-value
Depression diagnosis	0.47	0.26, 0.82	0.009	0.53	0.25, 1.11	0.093
Sleep disorder	0.53	0.33, 0.84	0.007	0.51	0.28, 0.92	0.025
Lack of lubrication	0.59	0.35, 0.99	0.049	0.44	0.22, 0.85	0.015
Pain after sex	0.52	0.32, 0.85	0.009	0.43	0.23, 0.79	0.006
Previous episiotomy	0.29	0.17, 0.48	<0.001	0.20	0.10, 0.39	<0.001
Neuropathic pain component	0.26	0.09, 0.63	0.005	0.19	0.06, 0.54	0.001

[1]OR = Odds Ratio, CI = Confidence Interval

[2]NSAID = non-steroidal anti-inflammatory drugs

Supplementary Table S3. Results of logistic regression model presenting variables independently associated with superficial peritoneal endometriosis diagnosis. (sensitivity analysis - complete cases)

Characteristic	Univariate			Multivariate		
	OR[1]	95% CI[1]	p-value	OR[1]	95% CI[1]	p-value
Irritable bowel syndrome	2.44	1.05, 6.19	0.046	6.50	2.17, 21.4	<0.001
Irregular menstrual cycle	2.19	1.32, 3.66	0.003	3.77	1.88, 7.87	<0.001
Pelvic floor tenderness	1.68	0.99, 2.86	0.056	2.92	1.42, 6.25	0.003
Abdominal trigger point	2.06	1.18, 3.64	0.012	2.76	1.33, 5.93	0.006
Painful bladder syndrome	1.66	1.03, 2.70	0.039	2.01	1.07, 3.83	0.029
Infertility history	3.68	1.65, 9.05	0.002	1.98	0.67, 6.28	0.22
Chronic NSAID[2] use	1.59	0.97, 2.61	0.066	1.72	0.92, 3.25	0.090
Coitarche	1.03	0.95, 1.11	0.481	0.92	0.82, 1.02	0.12
Progestogen only	0.59	0.36, 0.95	0.031	0.87	0.43, 1.73	0.68
Previous pregnancies	0.69	0.58, 0.82	<0.001	0.82	0.66, 1.00	0.051
Vaginism	0.62	0.29, 1.28	0.203	0.48	0.18, 1.24	0.13
Sleep disorder	0.51	0.31, 0.83	0.007	0.44	0.23, 0.82	0.009
Previous transverse incision	0.47	0.29, 0.76	0.002	0.44	0.22, 0.84	0.013
Pain after sex	0.47	0.28, 0.80	0.005	0.32	0.16, 0.65	0.001
Lack of lubrication	0.47	0.26, 0.84	0.011	0.30	0.14, 0.64	0.001
Previous episiotomy	0.29	0.17, 0.49	<0.001	0.14	0.06, 0.31	<0.001
Neuropathic pain	0.25	0.08, 0.65	0.008	0.13	0.03, 0.42	<0.001

[1]OR = Odds Ratio, CI = Confidence Interval

[2]NSAID = non-steroidal anti-inflammatory drugs

Supplementary Table S4. Performance metrics of various classifiers at fixed specificity levels for superficial peritoneal endometriosis diagnosis.

Classifiers	Specificity											
	70%				75%				80%			
	Recall	Precision	F1	AUC[1]	Recall	Precision	F1	AUC[1]	Recall	Precision	F1	AUC[1]
Support Vector Machine	65.5	67.9	66.7	66.2	62.1	69.2	65.4	66.2	37.9	64.7	47.8	66.2
Multilayer Perceptron	55.2	64.0	59.3	66.2	48.3	63.6	54.9	66.2	48.3	70.0	57.1	66.2
Extremely Randomised Trees	48.3	60.9	53.8	65.1	–	–	–	–	41.4	66.7	51.1	65.1
Random Forest	44.8	59.1	51.0	59.0	44.8	61.9	52.0	59.0	41.4	66.7	51.1	59.0

| Classifiers | Specificity | | | | | | | | | | | |
| | 70% | | | | 75% | | | | 80% | | | |
	Recall	Precision	F1	AUC[1]	Recall	Precision	F1	AUC[1]	Recall	Precision	F1	AUC[1]
Extreme Gradient Boosting	44.8	59.1	51.0	66.1	44.8	61.9	52.0	66.1	44.8	68.4	54.2	66.1
Light Gradient Boosting	48.3	60.9	53.8	64.5	48.3	63.6	54.9	64.5	44.8	68.4	54.2	64.5
Logistic Regression	48.3	60.9	53.8	67.8	48.3	63.6	54.9	67.8	48.3	70.0	57.1	67.8
Categorical Boosting	41.4	57.1	48.0	62.6	41.4	60.0	49.0	62.6	41.4	66.7	51.1	62.6
Adaptive Boosting	–	–	–	–	41.4	60.0	49.0	60.3	41.4	66.7	51.1	60.3
Naive Bayes	–	–	–	–	–	–	–	–	–	–	–	–

[1]Area under the curve.

Supplemental Figures
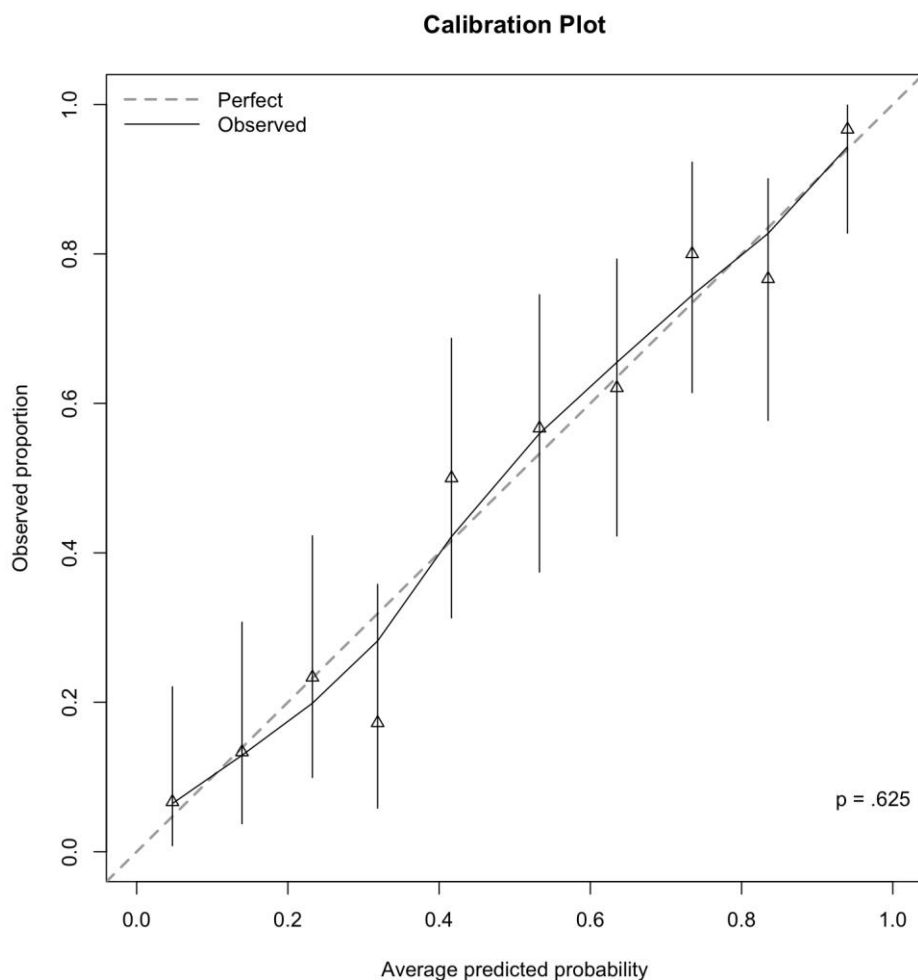
Supplementary Figure S1: Line plot showing the calibration of the binomial logistic regression model. The graph shows the agreement between predictions and observations in different deciles of the predicted values
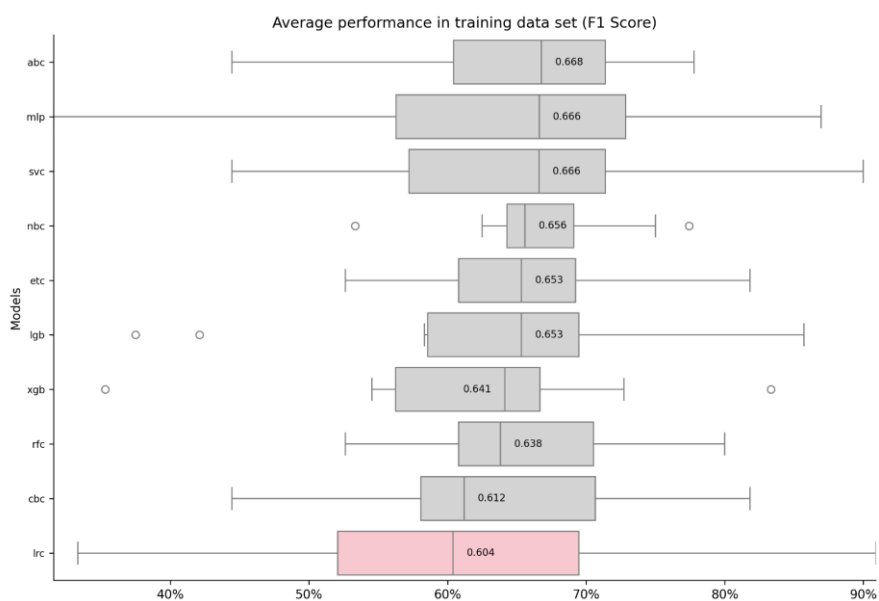
Supplementary Figure S2: Horizontal box plot showing the F1 score performance of models during training.
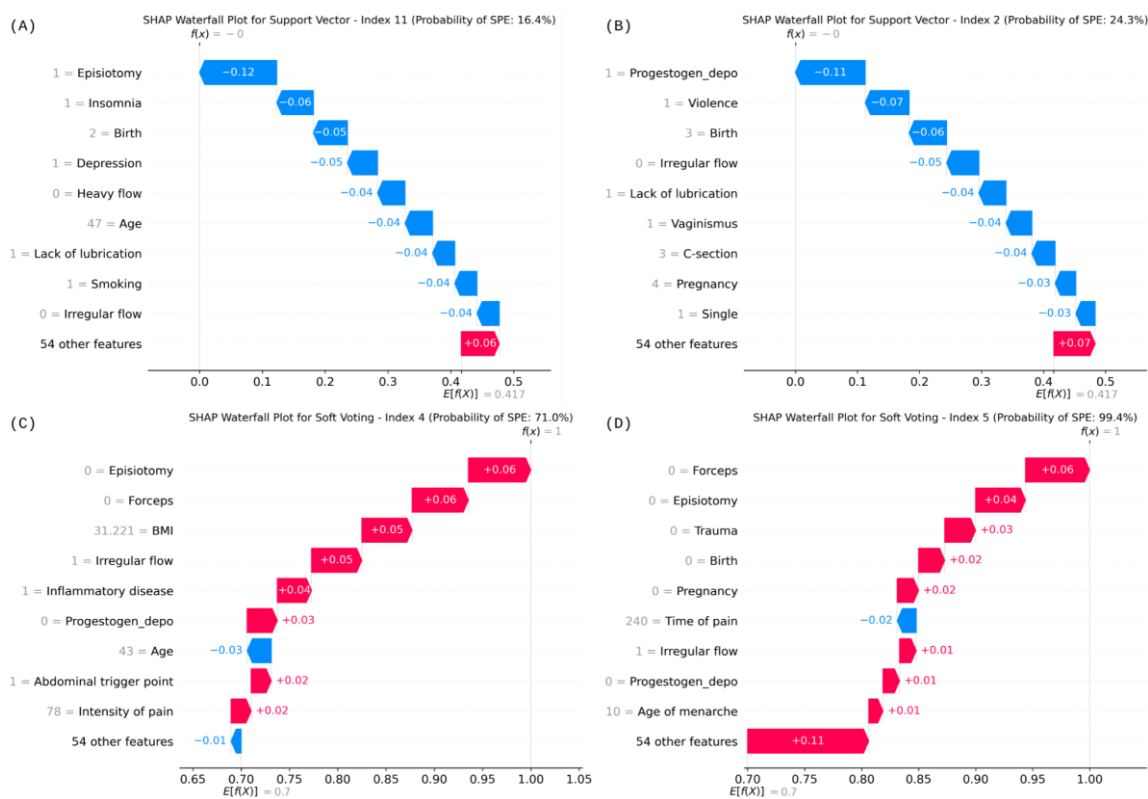
Supplementary Figure S3: Waterfall plot of the Shapley values on two different selected for individual and personalised interpretation.