

Comparative Analysis of Logistic Regression and Machine Learning Algorithms for Predicting De Novo Stress Urinary Incontinence or Symptomatic Worsening Following Pelvic Floor Reconstruction: A Multicenter Study

Liyang He¹, Yin Chen², Qiang Ma¹, Feng Jiang¹, Huanqing Xu³, Biyun Sun¹, Yonghong Luo⁴

¹Department of Ultrasound, The First Affiliated Hospital of Wannan Medical College (Yijishan Hospital of Wannan Medical College), Wuhu, Anhui, People's Republic of China; ²Department of Ultrasound, Third People's Hospital, Yancheng, Jiangsu, People's Republic of China; ³School of Computer Science and Technology, Tianjin University, Tianjin, People's Republic of China; ⁴Department of Obstetrics and Gynecology, The First Affiliated Hospital of Wannan Medical College (Yijishan Hospital of Wannan Medical College), Wuhu, Anhui, People's Republic of China

Correspondence: Yonghong Luo, Department of Obstetrics and Gynecology, The First Affiliated Hospital of Wannan Medical College (Yijishan Hospital of Wannan Medical College), Wuhu, Anhui, People's Republic of China, Email 20111584@wnmc.edu.cn; Biyun Sun, Department of Ultrasound, The First Affiliated Hospital of Wannan Medical College (Yijishan Hospital of Wannan Medical College), Wuhu, Anhui, People's Republic of China, Email 15212275599@163.com

Background: Whether to perform concurrent anti-incontinence surgery (AIS) during pelvic floor reconstruction (PFR) remains controversial. Traditional logistic regression (LR) models for predicting postoperative stress urinary incontinence (SUI) are limited by their inability to capture complex non-linear relationships and high-order interactions among predictors. Machine learning (ML) algorithms offer advantages in modeling such complexities, but existing ML models often lack interpretability and external validation.

Objective: This study aims to develop and validate an interpretable ML model using multicenter data to predict de novo SUI or symptomatic worsening after PFR, comparing performance with traditional LR and seven ML algorithms.

Methods: Clinical and pelvic floor ultrasound data from 490 women who underwent PFR between Jan 2019 and Jan 2024 were retrospectively analyzed. Univariate and multivariate logistic regression analyses identified independent predictors and constructed a conventional model. Seven ML models were developed. Model performance was assessed using area under the curve (AUC), calibration, and decision curve analysis.

Results: A total of 590 patients were enrolled, of whom 155 (26.3%) developed de novo SUI or symptomatic worsening. The conventional logistic regression model identified seven independent predictors; while least absolute shrinkage and selection operator regression selected 11 optimal features. Among all models, the Multilayer Perceptron (MLP) demonstrated the best predictive performance, achieving an AUC of 0.922 (95% CI: 0.880–0.957) in internal validation and 0.893 (95% CI: 0.844–0.936) in external testing. DeLong test revealed significant AUC differences among models (all $P < 0.05$). SHapley Additive exPlanations (SHAP) analysis identified concurrent AIS, body mass index, urethral rotation angle, Valsalva retrovesical angle, and age as the most influential predictors.

Conclusion: The MLP model provides accurate and clinically meaningful predictions of postoperative SUI risk following PFR, demonstrating strong discrimination, good calibration, and potential clinical utility. This model may support personalized surgical planning and decision-making.

Keywords: pelvic floor reconstruction, de novo stress urinary incontinence, symptomatic worsening, machine learning, predictive modeling, multilayer perceptron

Introduction

Pelvic organ prolapse (POP) is a common condition that significantly impairs women's quality of life, with reported prevalence rates ranging from 9.6% to 20% in developing countries.¹ Defects of the anterior compartment account for up to 91.2% of POP cases,² and more than 50% of women with a uterus experience POP accompanied by related symptoms. Stress urinary incontinence (SUI)³ is defined as the involuntary leakage of urine during activities that increase intra-abdominal pressure, such as coughing, sneezing, or physical exertion. Pelvic Floor Reconstruction (PFR) is the standard surgical intervention for moderate-to-severe POP. Although PFR effectively restores anatomical support, it is associated with a 15–50% risk of worsening pre-existing SUI or inducing de novo SUI postoperatively.⁴ Accurate preoperative risk stratification for SUI and informed decision-making regarding concurrent anti-incontinence surgery (AIS) are therefore critical. Complications related to incontinence procedures, such as postoperative urinary retention, urinary tract infections, and sexual dysfunction, can substantially impair both physical and psychological well-being and increase healthcare utilization. Performing unnecessary AIS exposes patients to additional risks, such as urinary dysfunction, sling erosion, and mesh-related complications.⁵ Conversely, withholding AIS in patients at high risk of postoperative SUI may result in persistent symptoms, reduced quality of life, and the need for subsequent reoperation. Thus, optimal surgical planning requires a careful balance between these competing risks.

Previous studies have primarily investigated the predictive value of urodynamic findings, clinical characteristics, and epidemiological factors for postoperative SUI. For example, Oliveira et al⁶ and others⁷ identified diabetes mellitus, advanced-stage prolapse, occult SUI, older age, and reduced maximum urethral closure pressure as independent risk factors. However, these studies predominantly relied on traditional logistic regression (LR) models, which have limited ability to capture complex non-linear relationships and high-order interactions among variables.^{6,7} In contrast, machine learning (ML) algorithms, particularly ensemble methods that integrate multiple decision trees, such as Random Forest (RF), offer significant advantages in modeling complex, non-linear interactions in high-dimensional datasets.⁸ For example, Yu et al⁹ demonstrated that an RF model outperformed LR in predicting myofascial pelvic pain syndrome (MPPS). Similarly, Fu et al¹⁰ developed and validated an eXtreme Gradient Boosting (XGBoost) model to predict postoperative SUI across different surgical approaches, including mesh and non-mesh procedures, regardless of pre-operative continence status. Their model identified body mass index (BMI), age, pelvic organ prolapse quantification (POP-Q) grade, and surgical type as key predictors, achieving an area under the curve (AUC) of 0.714.¹⁰ Nonetheless, a major limitation of existing ML-based models is that they are typically developed and validated using single-center datasets, limiting their generalizability without rigorous external validation.

The multilayer perceptron (MLP) models, a classical feedforward artificial neural network, is particularly well suited for capturing complex nonlinear patterns through its multi-layered architecture of interconnected neurons. In clinical prediction tasks,¹¹ MLP models have demonstrated strong performance across diverse medical domains. For instance, Nistal-Nuño et al¹² developed an MLP-based model to predict outcomes in critically ill patients with respiratory tract tumors by integrating laboratory parameters, demographic data, organ support therapies, and other clinical variables. This model achieved an AUC of 0.885, markedly outperforming conventional prediction systems with AUCs below 0.74. Similarly, Lu et al¹³ applied an MLP model to identify urinary tract infections (UTIs) associated with renal calculi, achieving an AUC of 0.98. However, these studies lacked interpretability, which remains a major barrier to clinical adoption. Moreover, the application of MLP models to predict complications following female pelvic floor surgery has not yet been explored.

Against this background, the present study addresses these gaps through three key innovations: (1) the use of multicenter data to develop and compare eight predictive models, including a LR model; (2) the integration of traditional univariate and multivariate regression with Least Absolute Shrinkage and Selection Operator (LASSO) regression for feature selection; and (3) the application of SHapley Additive exPlanations (SHAP)¹⁴ to interpret the optimal model by quantifying feature contributions and enhancing clinical transparency. Our objective is to develop a robust and interpretable tool for preoperative SUI risk stratification, thereby supporting individualized decisions regarding concurrent AIS, reducing surgical complications, and improving patient outcomes.

Materials and Methods

Study Population and Design

This retrospective multicenter cohort study included 490 women who underwent PFR between January 2019 and January 2024. The study cohort was randomly divided into a training set ($n = 343$, 70%) and an internal validation set ($n = 147$) via stratified sampling based on the primary outcome (postoperative SUI status). To assess model generalizability, an independent external test cohort comprising 100 patients from a separate institution was included.

This study has been approved by the Ethics Committee of Wannan Medical University (Approval No.: 2024–207). All data were strictly de-identified, including the removal of names, medical record numbers, and contact information, in accordance with the principles of the Declaration of Helsinki.

Inclusion criteria were as follows: (1) POP-Q stage II–IV; (2) Availability of preoperative pelvic floor ultrasound (PFUS) with complete clinical data; (3) A minimum postoperative follow-up of 6 months with complete outcome data.

Exclusion criteria included: (1) Prior AIS; (2) Clinical records with more than 10% missing data.

Predictor Variables and Data Acquisition

PFUS Examination

PFUS was performed using SonoScape (China) and Mindray Resona 8s (China) systems, equipped with VE9-5 endocavitary 3D transducers (4–8 MHz) and D8-2U 3D transducers (5–9 MHz), respectively. All examinations were conducted with patients in the lithotomy position and a bladder volume of 50–100 mL. After lubrication, the transducer was placed on the perineum, and the insonation angle was adjusted to clearly visualize the pubic symphysis, bladder neck, urethra, and levator ani muscle (LAM). Two-dimensional (2D) images and dynamic cine loops were obtained at rest, during maximal Valsalva maneuver, and during voluntary pelvic floor muscle contraction.⁵

PFUS Parameters

Pelvic floor ultrasound parameters included postvoid residual urine volume (PVR), defined as the volume of urine remaining in the bladder after voiding. Bladder neck position (BNP) and retrovesical angle (RA) were measured both at rest (R_BNP, R_RA) and during the Valsalva maneuver (Va_BNP, Va_RA). Urethral tilt angle (UTA) was similarly recorded at rest (R_UTA) and during Valsalva (Va_UTA). Urethral rotation angle (UretRot) was calculated as the difference between the UTA during Valsalva and at rest ($Va_UTA - R_UTA$), while bladder neck descent (BND) was defined as the difference between the BNP during Valsalva and at rest ($Va_BNP - R_BNP$). Urethral funneling was considered present when the length of the urethral meatus exceeded 50% of the total urethral length during the Valsalva maneuver.

LAM Injury Assessment

LAM integrity was evaluated using three-dimensional tomographic ultrasound imaging (TUI)¹⁵ at the plane of minimal hiatal dimensions and at two cranial adjacent planes during voluntary contraction. Major LAM injury were defined as: (1) Bilateral loss of more than 50% of the internal pubococcygeus muscle; or (2) Complete loss of the muscle on at least one side.¹⁶

Clinical Parameters

The following clinical variables were collected: age, BMI, gravidity, parity, history of cesarean section, chronic obstructive pulmonary disease (COPD), menopausal status, diabetes mellitus, hypertension, POP-Q points (Aa/Ba), type of PFR procedure (colpocleisis, anterior colporrhaphy, sacrospinous ligament fixation, sacrocolpopexy, and vaginal mesh reconstruction) and performance of concurrent AIS.

Reliability Assessment

Pelvic floor ultrasound examinations were independently interpreted by two senior ultrasonographers (attending physicians with at least five years of experience in pelvic floor ultrasound diagnosis), who evaluated LAM defects on both static and dynamic images. To minimize interpretation bias, the assessors were blinded to all clinical information, including demographic characteristics, surgical history, and SUI-related symptoms.

Interobserver reliability for the classification of LAM defects was assessed by randomly selecting 20% of ultrasound images for repeat evaluation by both assessors. Agreement was quantified using Cohen's kappa coefficient, yielding a κ value of 0.89, which indicates excellent interobserver reliability ($\kappa > 0.80$ = excellent; 0.61–0.80 = good; 0.41–0.60 = moderate; $\kappa < 0.40$ = poor). Any discrepancies were resolved through consensus discussions with a third senior radiologist with more than 10 years of experience to ensure data accuracy.

Outcome Definitions

SUI was diagnosed based on patient-reported symptoms of involuntary urine leakage during activities that increase abdominal pressure (eg, coughing or sneezing), findings from physical examination, and results of a 1-hour pad test to quantify urine loss.

Overt SUI was defined as involuntary urine leakage during increased abdominal pressure or a positive cough stress test (CST).¹⁷ Occult SUI (OSUI) was defined as the absence of subjective urinary leakage with a positive CST following manual reduction of prolapse.¹⁸ De novo SUI was defined as the absence of preoperative SUI with the presence of postoperative symptoms and/or a positive CST.¹⁹ Postoperative SUI status was evaluated at 6 months and classified as cured, improved, unchanged, or worsened.^{20,21} “Cured” was defined as resolution of preoperative overt SUI or a 1-hour pad test result of 0–5 g. “Improved” referred to symptom improvement or a reduction of at least 50% in pad test leakage. “Unchanged” indicated symptoms or pad test results comparable to baseline, while “worsened” was defined as symptom exacerbation or an increase of more than 50% in pad test leakage. Patients were further categorized into two outcome groups: the event group, consisting of patients with de novo SUI or worsened pre-existing SUI, and the non-event group, consisting of patients with cured, improved, or unchanged SUI.

Data Preprocessing

Variables with a missing rate exceeding 10% (including history of cesarean section and COPD) were excluded from further analysis. For the remaining variables, missing values were imputed using the SimpleImputer class from the scikit-learn library (version 0.24.2) in Python (version 3.7.1). Continuous variables were imputed using the median, while categorical variables were imputed using the most frequent value (mode). To address class imbalance in the training dataset, the Synthetic Minority Oversampling Technique (SMOTE) was applied. Oversampling was restricted to the training set to prevent data leakage and preserve the integrity of model evaluation.

Model Construction and Validation

As described in Study Population and Design, the 490 patients were stratified into a training set (70%, $n = 343$) and an internal validation set (30%, $n = 147$). An external validation cohort of 100 patients from an independent institution was used to assess model generalizability.

Traditional LR Model

Univariate analyses were first performed to screen potential predictors. Continuous variables were analyzed using Mann–Whitney U -test, while categorical variables were compared using chi-square test or Fisher's exact test, as appropriate. Variables with P value < 0.05 in univariate analysis were entered into a multivariate binary LR model. Forward stepwise regression based on the likelihood ratio test was applied to identify independent predictors, balancing model fit and parsimony. Entry and removal criteria were set at $P < 0.05$ and $P > 0.10$, respectively, consistent with commonly used variable selection strategies, including Akaike information criterion (AIC) and Bayesian information criterion (BIC).

ML Models

To reduce multicollinearity, pairwise feature correlations were assessed using a Spearman rank correlation matrix (Figure S1). When a correlation coefficient exceeded $|r| > 0.9$, one of the correlated variables was excluded based on clinical relevance to minimize redundancy and enhance model robustness. The remaining features were then entered into a LASSO regression model with 10-fold cross-validation. LASSO regression is a technique that automatically selects

the most important features by shrinking less relevant variable coefficients to zero, thereby simplifying the model and reducing overfitting. The optimal feature subset was selected using the λ_{1se} criterion, defined as the largest value of λ within one standard error of the minimum cross-validation error.

Seven ML models were constructed: Naive Bayes (NB), Gradient Boosting (GB), RF, MLP, XGBoost, Light Gradient Boosting Machine (LightGBM), and k-Nearest Neighbors (KNN). Hyperparameter optimization was conducted on the training set using RepeatedStratifiedKFold (10-fold, 10 repeats) in combination with GridSearchCV (scikit-learn), with balanced accuracy as the primary optimization metric to account for class imbalance. The final models were retrained on the full training set using the optimal hyperparameters. Detailed hyperparameter grids are provided in [Table S1](#).

Model Evaluation

The primary performance metric for all models was the AUC. Secondary performance metrics included accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1-score, which together provide a comprehensive assessment of diagnostic performance. Model calibration was evaluated using calibration plots and the Brier Score (BS), with lower BS values indicating better calibration. The clinical utility of each prediction model was assessed using decision curve analysis (DCA), which estimates net benefit across a range of threshold probabilities. Pairwise comparisons of AUCs between models were performed using DeLong's test. All performance estimates were reported with 95% confidence intervals (CIs), which were derived using non-parametric bootstrapping with 1000 resamples.

Integrated Discrimination Improvement Index (IDI) and Net Reclassification Index (NRI)

To quantify improvements in predictive performance, the integrated IDI and NRI were calculated using the multivariate logistic regression model. A baseline model was constructed using clinical variables alone, including age, BMI, history of vaginal delivery, POP-Q stage, and preoperative SUI status. A combined model was then developed by incorporating pelvic floor ultrasound parameters, such as urethral rotation angle and the retrovesical angle during the Valsalva maneuver, in addition to the clinical variables. The model-fitting strategy was kept consistent between the baseline and combined models to ensure comparability. The IDI and NRI were used to assess incremental improvements in overall discrimination and risk reclassification beyond conventional receiver operating characteristic (ROC) curve and AUC-based metrics.

Model Interpretability

Interpreting the decision-making mechanisms of ML models, particularly complex “black box” models, remains challenging. To address this issue, SHAP, a game theory-based interpretability method, was applied. SHAP quantifies the contribution of each input feature to a model's prediction by assigning Shapley values derived from cooperative game theory. This approach enables both global and local interpretability. At the global level, SHAP feature importance identifies the most influential predictors across the entire dataset, while at the local level, it explains individual predictions by decomposing model outputs into additive contributions from each feature. By making model behavior transparent and traceable, SHAP enhances clinical interpretability, promotes trust in model predictions, and facilitates practical implementation in clinical decision-making. SHAP analysis explains how each patient's characteristics, such as age or BMI, influence the model's risk prediction. It assigns a contribution value to each feature, where positive values increase risk and negative values decrease risk, making the model's decision process transparent.

Statistical Analysis

Continuous variables were presented as mean $\pm$ standard deviation (SD) if normally distributed, or as median with interquartile range (IQR) if non-normally distributed. Normality was assessed using the Shapiro–Wilk test, and homogeneity of variances was evaluated using Levene's test. For between-group comparisons of continuous variables, the independent samples *t*-test was applied to normally distributed data with homogeneous variances; Welch's *t*-test was used

when variances were heterogeneous; Mann–Whitney *U*-test was employed for non-normally distributed data. Categorical variables were expressed as frequencies and percentages. Group comparisons were performed using the Chi-square test when all expected cell counts were ≥ 5 ; otherwise, Fisher's exact test was applied. All statistical analyses were conducted using Python (version 3.7.1), with the *scipy.stats* package used for hypothesis testing and *pandas* for data manipulation. All tests were two-sided, and a *P* value < 0.05 was considered statistically significant.

Results

Clinical Baseline Characteristics

Between January 2019 and January 2024, a total of 658 PFR procedures were performed at Yijishan Hospital. After excluding patients who were lost to follow-up ($n = 35$), those with a prior history of PFR ($n = 23$), and those with a postoperative follow-up period of less than 6 months ($n = 110$), 490 patients were included in the final analysis. These patients were randomly assigned to a training set ($n = 343$, 70%) and an internal validation set ($n = 147$, 30%). An external validation cohort of 100 patients was recruited from Yancheng Hospital of Traditional Chinese Medicine (Figure 1). Baseline characteristics were generally well balanced between the training and internal validation cohorts, with the exception of duration of prolapse ($P < 0.001$), POP-Q point Aa ($P < 0.001$), postvoid residual urine volume ($P < 0.001$), presence of mixed incontinence ($P = 0.001$) and urethral internal orifice small funneling ($P = 0.003$) (Table 1). Among the total study population of 590 patients, 155 (26.3%) developed postoperative SUI, including 100 cases of de novo SUI and 55 cases of worsened pre-existing SUI.

Traditional LR Model

Univariate and multivariate LR analyses identified seven independent risk factors for postoperative de novo or aggravated SUI: BMI, concomitant AIS, Valsalva_retrovesical angle, urethral rotation angle, LAM Injury, diabetes mellitus, and sacrospinous ligament fixation (SSLF) (Table 2). These predictors were incorporated into a multivariate LR model and used to construct a nomogram for individualized risk estimation (Figure 2).

In the internal validation cohort, the model demonstrated strong discriminative performance, with an AUC of 0.919 (95% CI: 0.887–0.951), high sensitivity (0.921), and moderate specificity (0.780). However, model performance declined in the external validation set, yielding an AUC of 0.833 (95% CI: 0.765–0.901), suggesting potential overfitting to the training data. Although specificity improved to 0.817, sensitivity decreased to 0.711, resulting in a relatively low positive predictive value (PPV = 0.574; 95% CI: 0.482–0.666) (Table 3).

ML Models

Using LASSO regression, 11 key predictors were selected for ML model development. As shown in the LASSO coefficient path and cross-validation plots (Figure 3A and B), under the optimal penalty parameter (λ), the following variables retained non-zero regression coefficients at the optimal penalty parameter (λ): concurrent AIS, BMI, urethral rotation angle, Valsalva_retrovesical angle, age, LAM injury, POP-Q staging, rest_retrovesical angle, Ba point, diabetes mellitus and parity (Figure 3C). Notably, all imaging-related predictors identified by univariate and multivariate logistic regression were retained, with the exception of SSLF. Although SSLF had been identified as a protective factor in prior regression analyses, it was excluded from the final LASSO-selected feature set.

To investigate the exclusion of SSLF, variance inflation factor (VIF) analysis was performed to assess multicollinearity among candidate predictors. The results indicated a VIF value of 5.2 for SSLF and concurrent AIS, exceeding the predefined threshold of 5.0 and indicating moderate collinearity between these variables. All other predictors exhibited VIF values below 3.0, confirming the absence of significant collinearity. These findings suggest partial information redundancy between SSLF and AIS, which likely contributed to the exclusion of SSLF during LASSO feature selection.

In the internal validation set, all seven ML models demonstrated strong discriminative ability, with AUC values exceeding 0.870. This performance was largely maintained in the external test cohort, where all models achieved AUC values greater than 0.833, indicating good generalizability despite potential class imbalance. Among tree-based ensemble models (RF, Gradient Boosting, XGBoost, and LightGBM), several demonstrated superior performance in specific metrics.

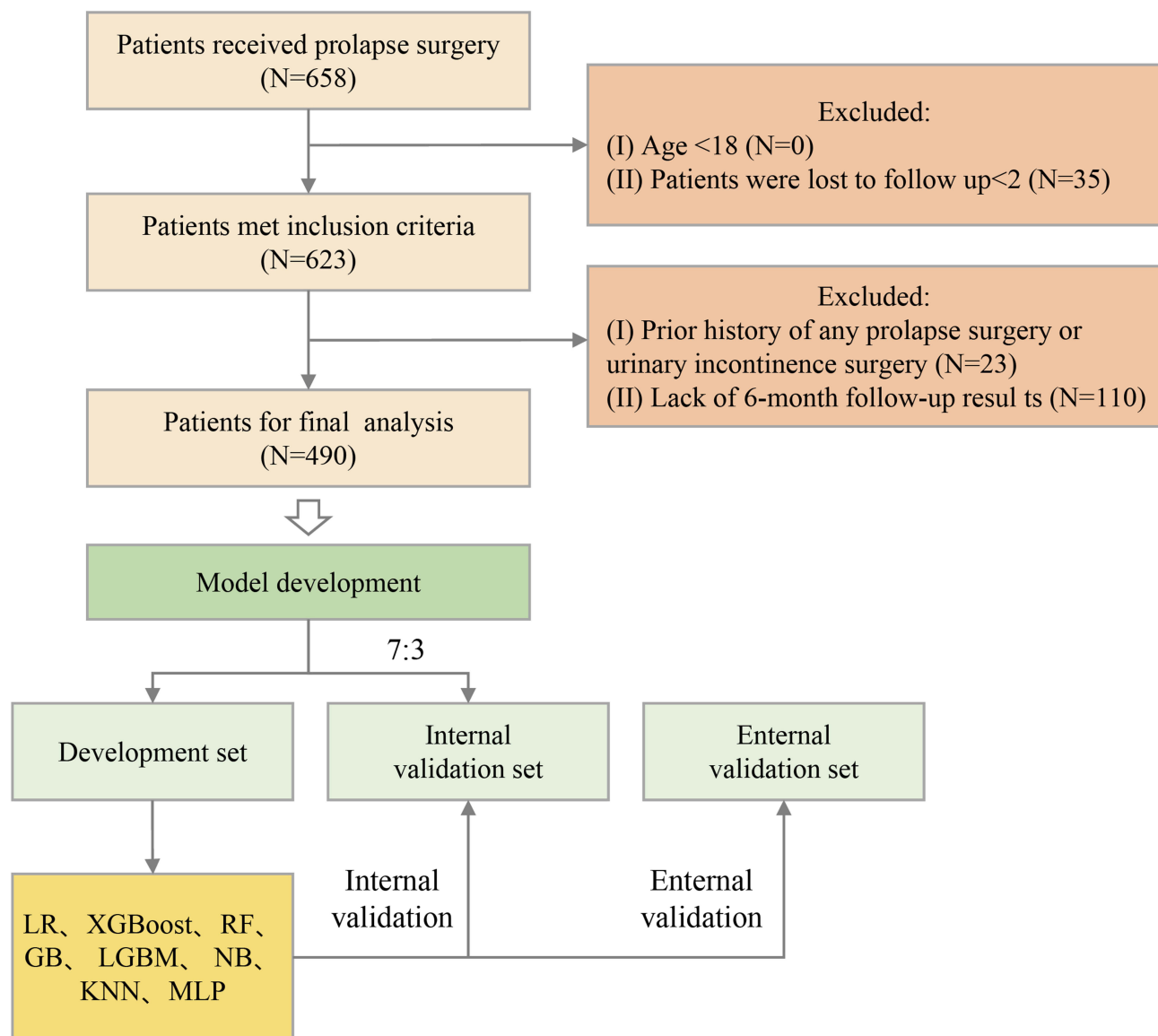


Figure 1 Illustrates the comprehensive process of prediction model validation and development, encompassing a range of machine learning and statistical models. This includes LR, NB, GB, RF, MLP, XGBoost, LGBM, and KNN. The flowchart emphasizes the importance of rigorous validation and development steps, as outlined in various prediction model studies, to ensure the robustness and applicability of the models across different domains.

Abbreviations: LR, Logistic Regression; NB, Naive Bayes; GB, Gradient Boosting Machine; RF, Random Forest; MLP, Multilayer Perceptron; XGBoost, Extreme Gradient Boosting; LGBM, Light Gradient Boosting Machine; KNN, k-Nearest Neighbors.

The RF model achieved the highest specificity (0.945) and positive predictive value (0.838) during internal validation; however, its sensitivity declined markedly in the external test cohort (from 0.816 to 0.500). In contrast, the MLP model demonstrated consistently strong and balanced performance across both validation cohorts. The MLP achieved an AUC of 0.922 in internal validation and maintained a high AUC of 0.893 in the external test set. Its accuracy (0.816), sensitivity (0.816), and specificity (0.817) remained stable and well balanced across datasets. Overall, the MLP model exhibited the most robust and consistent performance among all evaluated models (Figure 4A–C and Table 3). Pairwise comparisons of model performance using DeLong’s test revealed statistically significant differences in AUC among the prediction models in the external test cohort (all $P < 0.05$). Detailed results of these comparisons are provided in Table S2.

Table 1 Baseline Distribution Characteristics of All Indicators Across Three Datasets

Characteristic	Overall (n = 590)	Training Cohort (n = 343)	Internal Validation Cohort (n = 147)	External Validation Cohort (n = 100)	P-value
Age, mean (SD)	61.3 (8.1)	61.2 (8.1)	63.0 (8.0)	59.1 (7.9)	0.083
BMI, mean (SD)	22.3 (3.1)	21.8 (2.8)	22.1 (3.0)	24.1 (3.3)	0.204
Pelvic_Organ_Prolapse_grade, n (%)					0.507
II	263 (44.6)	148 (43.1)	67 (45.6)	48 (48.0)	
III	279 (47.3)	168 (49.0)	65 (44.2)	46 (46.0)	
IV	48 (8.1)	27 (7.9)	15 (10.2)	6 (6.0)	
Hysterectomy, n (%)					0.125
No	522 (88.5)	313 (91.3)	134 (91.2)	75 (75.0)	
Yes	68 (11.5)	30 (8.7)	13 (8.8)	25 (25.0)	
Number_of_pregnancies, n (%)					0.073
I	28 (4.7)	17 (5.0)	9 (6.1)	2 (2.0)	
2	215 (36.4)	127 (37.0)	47 (32.0)	41 (41.0)	
3	158 (26.8)	81 (23.6)	34 (23.1)	43 (43.0)	
≥ 4	189 (32.0)	118 (34.4)	57 (38.8)	14 (14.0)	
Parity, n (%)					0.105
I	199 (33.7)	108 (31.5)	51 (34.7)	40 (40.0)	
2	310 (52.5)	191 (55.7)	77 (52.4)	42 (42.0)	
≥ 3	81 (13.7)	44 (12.8)	19 (12.9)	17 (17.0)	
Years_of_prolapse, mean (SD)	4.6 (2.8)	4.9 (2.9)	4.6 (2.9)	3.3 (1.4)	< 0.001
Hypertension, n (%)					0.159
No	489 (82.9)	294 (85.7)	120 (81.6)	75 (75.0)	
Yes	101 (17.1)	49 (14.3)	27 (18.4)	25 (25.0)	
Diabetes_mellitus, n (%)					0.125
No	499 (84.6)	301 (87.8)	129 (87.8)	69 (69.0)	
Yes	91 (15.4)	42 (12.2)	18 (12.2)	31 (31.0)	
Aa, mean (SD)	-0.7 (1.4)	-0.5 (1.4)	-0.4 (1.3)	-1.7 (0.8)	< 0.001
Ba, mean (SD)	3.0 (1.5)	2.9 (1.5)	2.9 (1.6)	3.2 (1.2)	0.126
Mixed incontinence, n (%)					< 0.001
No	492 (83.4)	301 (87.8)	125 (85.0)	66 (66.0)	
Yes	98 (16.6)	42 (12.2)	22 (15.0)	34 (34.0)	
Residual urine volume, mean (SD)	35.7 (15.6)	33.9 (15.5)	36.0 (17.4)	41.5 (10.7)	< 0.001
Rest_ Bladder Neck, mean (SD)	22.6 (3.1)	22.0 (2.6)	22.4 (3.1)	25.1 (3.3)	0.001
Valsalva_ bladder neck, mean (SD)	-19.3 (8.4)	-19.3 (8.9)	-20.2 (8.6)	-18.0 (5.5)	0.128
Bladder neck mobility, mean (SD)	41.9 (8.4)	41.3 (8.9)	42.6 (8.3)	43.1 (6.3)	0.079
Rest_posterior vesical angle, mean (SD)	121.4 (17.5)	122.4 (18.5)	121.4 (16.8)	118.1 (14.2)	0.098
Valsalva_ posterior vesical angle, mean (SD)	134.4 (25.0)	133.0 (25.3)	131.5 (25.9)	143.5 (20.8)	0.211
Urethral rotation anglemean (SD)	76.7 (32.8)	76.0 (34.7)	76.0 (33.7)	80.2 (23.8)	0.513
Urethral internal orifice small funnel, n (%)					0.003
No	477 (80.8)	282 (82.2)	126 (85.7)	69 (69.0)	
Yes	113 (19.2)	61 (17.8)	21 (14.3)	31 (31.0)	
Levator ani muscle injury, n (%)					0.15
No	502 (85.1)	298 (86.9)	125 (85.0)	79 (79.0)	
Yes	88 (14.9)	45 (13.1)	22 (15.0)	21 (21.0)	0.35
Combined antidiuretic surgery, n (%)					
No	286 (48.5)	166 (48.4)	77 (52.4)	43 (43.0)	
Yes	304 (51.5)	177 (51.6)	70 (47.6)	57 (57.0)	
Reconstruction of pelvic floor surgical_procedure, n (%)					0.296
Vaginal Mesh Reconstruction	75 (12.7)	41 (12.0)	23 (15.6)	11 (11.0)	
Sacrosciatic ligament fixation	52 (8.8)	28 (8.2)	13 (8.8)	11 (11.0)	
Uterosacral ligament suspension	80 (13.6)	44 (12.8)	21 (14.3)	15 (15.0)	
Anterior vaginal wall repair	108 (18.3)	57 (16.6)	25 (17.0)	26 (26.0)	
Vaginal occlusion	275 (46.6)	173 (50.4)	65 (44.2)	37 (37.0)	

Abbreviation: BMI, Body Mass Index.

Table 2 Results of Univariate and Multivariate Analysis

Feature	Univariate		Multivariable	
	OR (95% CI)	P value	OR (95% CI)	P value
Age	1.007 (0.982–1.031)	0.607		
BMI	1.344 (1.234–1.464)	< 0.001	1.324 (1.181–1.483)	< 0.001
Aa	1.115 (0.962–1.294)	0.149		
Ba	1.113 (0.978–1.267)	0.105		
Residual urine volume	1.01 (0.998–1.023)	0.092		
Rest_Bladder_Neck	1.033 (0.961–1.112)	0.378		
Valsalva_bladder_neck	1.005 (0.982–1.028)	0.66		
Bladder neck mobility	0.998 (0.975–1.021)	0.869		
Rest_posterior_vesical_angle	1.006 (0.994–1.017)	0.321		
Number_of_pregnancies	1.114 (0.997–1.287)	0.235		
Parity	1.437 (1.039–1.995)	0.028	1.473 (0.909–2.386)	0.116
Valsalva_posteriorvesical_angle	0.966 (0.957–0.974)	< 0.001	0.971 (0.959–0.984)	< 0.001
Urethral_rotation_angle	1.023 (1.017–1.03)	< 0.001	1.022 (1.013–1.031)	< 0.001
Pelvic_Organ_Prolapse_grade				
II	0.479 (0.042–5.409)	0.552		
III	0.823 (0.073–9.226)	0.875		
IV	1.109 (0.093–13.263)	0.935		
Hysterectomy				
No	–	–		
Yes	1.641 (0.845–3.184)	0.143		
Hypertension				
No	–	–		
Yes	0.891 (0.502–1.579)	0.692		
Diabetes_mellitus				
No	–	–		
Yes	4.49 (2.57–7.846)	< 0.001	3.052 (1.33–7.005)	0.008
Mixed_incontinence				
No	–	–		
Yes	1.512 (0.858–2.664)	0.153		
Urethral_internal_orifice_small_funnel				
No	–	–		
Yes	1.006 (0.584–1.733)	0.981		
Levator_ani_muscle_injury				
No	–	–		
Yes	0.411 (0.197–0.856)	0.017	0.236 (0.083–0.672)	0.007
Combined_antidiuretic_surgery				
No	–	–		
Yes	0.036 (0.017–0.076)	< 0.001	0.045 (0.02–0.103)	< 0.001
Vaginal Mesh Reconstruction	–	–		
Sacrospinous ligament fixation	0.305 (0.104–0.894)	0.030	0.199 (0.045–0.868)	0.032
Uterosacral ligament suspension	0.603 (0.273–1.334)	0.212	0.827 (0.285–2.396)	0.726
Anterior vaginal wall repair	1.14 (0.567–2.293)	0.713	1.044 (0.382–2.854)	0.933
Vaginal occlusion	0.709 (0.387–1.298)	0.265	0.939 (0.409–2.155)	0.882

Abbreviations: BMI, Body Mass Index; OR, Odds Ratio; CI, Confidence Interval.

Model Calibration and Clinical Decision Curve Analysis

In the internal validation cohort, all models demonstrated good calibration, as indicated by low Brier scores ranging from 0.089 to 0.118. Calibration curves closely approximated the 45° reference line, indicating strong agreement between predicted and observed probabilities (Figure 5A and B). In the external validation cohort, BS increased

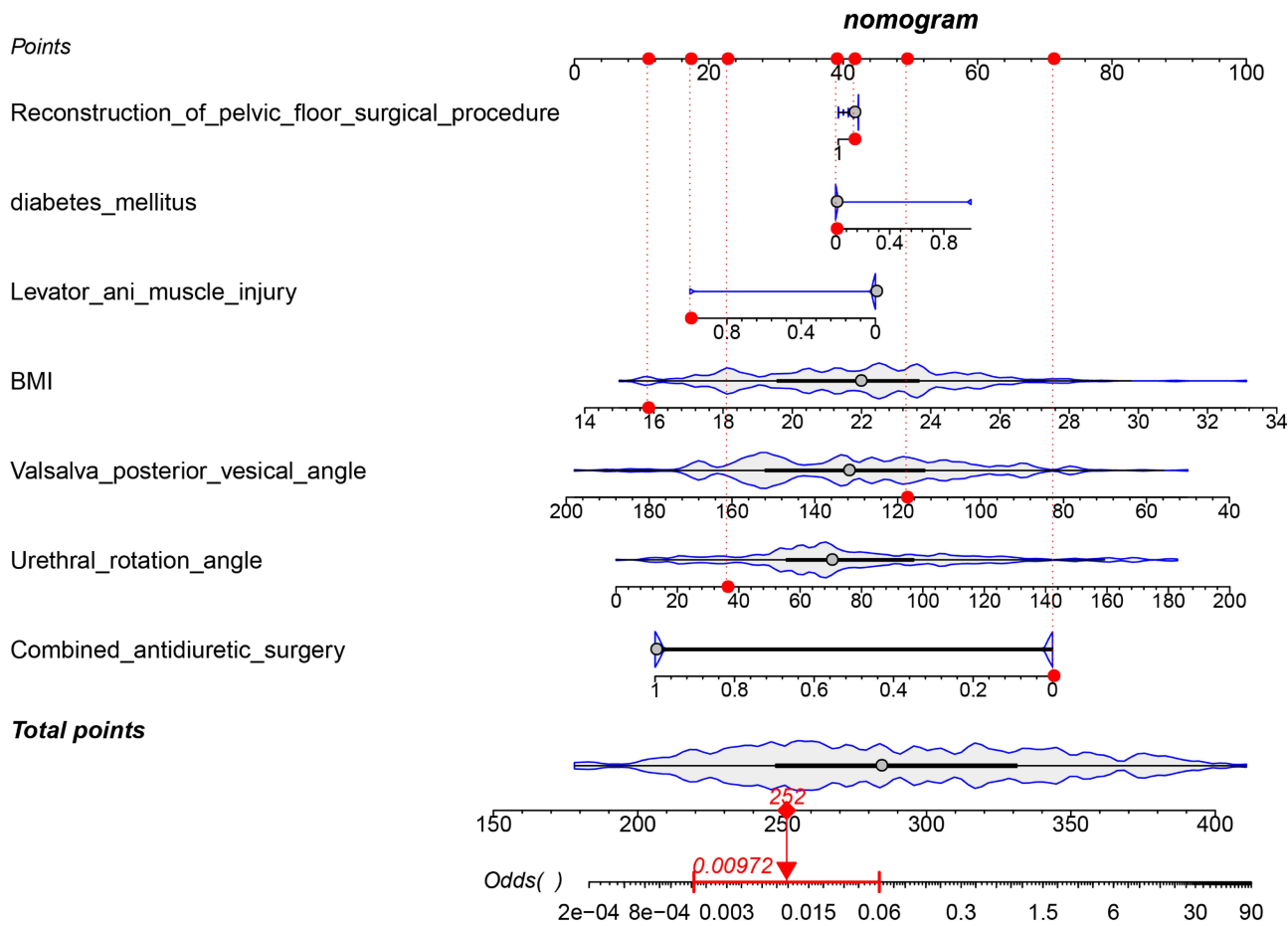


Figure 2 Illustrates the Nomogram for a traditional LR model, which incorporates seven independent variables known to influence the outcome.
Abbreviation: LR, Logistic Regression.

slightly (0.125–0.167), suggesting mild risk overestimation; nevertheless, overall calibration remained acceptable (Figure 5C).

In the internal validation cohort, the KNN model exhibited lower net benefit than the “no treatment” strategy curve within the threshold probability ranges of 0.7–0.8 and 0.9–1.0. In contrast, all other models maintained favorable net benefits across a wide range of threshold probabilities (0–0.9) (Figure 5D and E). In the external validation cohort, the net benefit of all models decreased relative to internal validation; however, they continued to demonstrate substantial clinical utility within threshold probabilities ranging from 0 to 0.7 (Figure 5F).

Comparison Between the New Model and the Baseline Model

On the external test cohort, the new model demonstrated significantly superior performance compared with the baseline model, as evidenced by higher NRI and IDI values (both $P < 0.05$). These findings indicate that the new model provides improved risk reclassification and enhanced overall discriminative ability beyond conventional clinical predictors alone. Detailed results are presented in the Table S3.

Model Interpretability: SHAP Analyses

Model interpretability was assessed using SHAP. Global interpretability was evaluated using Shap summary bar plots, beeswarm plots, and dependence plots, while local interpretability for individual predictions was examined using SHAP waterfall plots.

Table 3 Comprehensive Performance Metrics of Machine Learning Models on Internal Validation and External Test Sets

Dataset Indicator	Model	AUC	Accuracy	Sensitivity	Specificity	F1-score	PPV	NPV
Internal validation set	LR	0.919	0.816	0.921	0.780	0.722	0.593	0.966
	NB	0.916	0.796	0.868	0.771	0.688	0.569	0.944
	GB	0.926	0.850	0.789	0.872	0.732	0.682	0.922
	RF	0.921	0.912	0.816	0.945	0.827	0.838	0.936
	MLP	0.922	0.837	0.816	0.844	0.721	0.646	0.929
	XGB	0.924	0.857	0.842	0.862	0.753	0.681	0.940
	LGBM	0.924	0.844	0.868	0.835	0.742	0.647	0.948
	KNN	0.871	0.850	0.684	0.908	0.703	0.722	0.892
External test set	LR	0.833	0.789	0.711	0.817	0.635	0.574	0.890+
	NB	0.883	0.782	0.974	0.716	0.698	0.544	0.987
	GB	0.893	0.762	0.947	0.697	0.673	0.522	0.974
	RF	0.870	0.803	0.500	0.908	0.567	0.655	0.839
	MLP	0.893	0.816	0.816	0.817	0.697	0.608	0.927
	XGB	0.850	0.755	0.763	0.752	0.617	0.518	0.901
	LGBM	0.849	0.782	0.763	0.789	0.644	0.558	0.905
	KNN	0.846	0.796	0.447	0.917	0.531	0.654	0.826

Abbreviations: LR, Logistic Regression; NB, Naive Bayes; GB, Gradient Boosting Machine; RF, Random Forest; MLP, Multilayer Perceptron; XGB, Extreme Gradient Boosting; LGBM, Light Gradient Boosting Machine; KNN, k-Nearest Neighbors; PPV, Positive Predictive Value; NPV, Negative Predictive Value; AUC, Area Under the Curve.

The SHAP summary bar plot revealed that the five most influential predictors, ranked by mean absolute SHAP values, were concomitant AIS, BMI, urethral rotation angle, Valsalva_retrovesical angle, and age (Figure 6A). SHAP beeswarm plots further illustrated the direction and magnitude of each feature's effect on model predictions, showing that higher urethral rotation angle, elevated BMI, and advanced POP-Q stage were associated with an increased risk of de novo or aggravated SUI. In contrast, concomitant AIS, larger Valsalva retrovesical angle and older age were associated with reduced risk (Figure 6B). SHAP dependence plots provided a detailed visualization of how individual features influenced model output across the dataset. Positive SHAP values indicated increased predicted risk, whereas negative values indicated risk reduction. For example, BMI > 24 kg/m², urethral rotation angle > 60°, and POP-Q grade ≥ III were associated with a higher predicted risk of postoperative SUI (Figure 6C). Local interpretability was demonstrated using SHAP waterfall plots, which illustrated how individual feature contributions shifted the model output from the baseline prediction ($E[f(x)]$) to the final prediction for a given patient. In one representative low-risk case, Ba point and age contributed positively to risk prediction (+0.41 and +0.19, respectively), whereas concurrent AIS and the Valsalva retrovesical angle contributed negatively (−1.53 and −1.19, respectively). The baseline model output value for this patient was −1.413, which shifted to a final output value of −3.788, corresponding to a low-risk classification (Figure 6D). Conversely, in a representative high-risk case, the baseline output value of −1.517 increased to a final output value of 2.557 due to differing magnitudes and directions of feature contributions, resulting in a high-risk classification (Figure 6E).

Discussion

A recent comprehensive review of predictive models for SUI after POP surgery highlighted several methodological limitations, including small sample sizes, retrospective study designs, heterogeneous predictor selection, and limited external validation.²² In response, the present study developed and rigorously validated a transparent ML model using multicenter data to predict SUI progression following PFR surgery.

Key Research Findings

In this multicenter retrospective cohort study of 590 patients who underwent PFR, 155 patients (26.3%) developed either de novo or worsened SUI postoperatively. This incidence aligns with the 15–50% range reported in previous studies,⁴

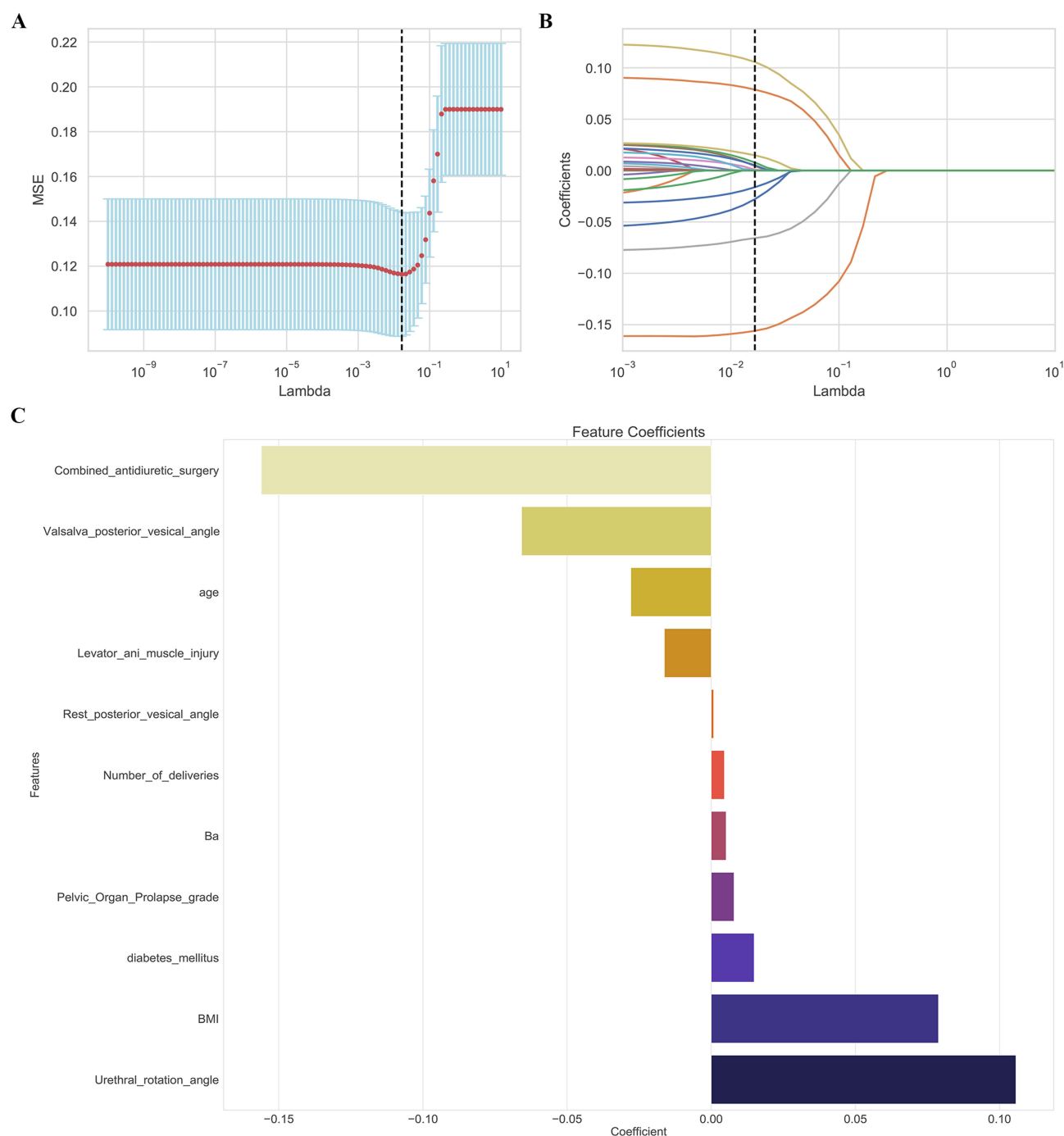


Figure 3 Illustrates the LASSO Regularization Path and optimal feature selection, with plot. **(A)** Showing how feature selection is achieved through LASSO regression. The vertical dashed lines indicate the optimal λ value, which is determined by the minimum criteria; **(B)** Coefficient profile plot. The vertical line signifies the $\log(\lambda)$ value at which Lasso regression identified the optimal set of 11 features with non-zero coefficients; **(C)** The bar plot illustrates the coefficients of the final 11 features selected by the LASSO model, indicating their relative importance and the direction of their impact on the outcome. LASSO regression, known for its feature selection capabilities, tends to shrink the coefficients of less important features to zero, resulting in a sparse model that emphasizes the most influential predictors.

Abbreviation: LASSO, Least Absolute Shrinkage and Selection Operator.

confirming the clinical prevalence of postoperative SUI and providing crucial baseline data for developing robust predictive models.

Using this cohort, we systematically developed and validated eight predictive models, including traditional LR and seven ML algorithms: NB, GB, RF, MLP, XGBoost, LGBM, and KNN. Feature selection was conducted via LASSO regression, which identified 11 key predictors. Internal validation and independent external testing demonstrated that

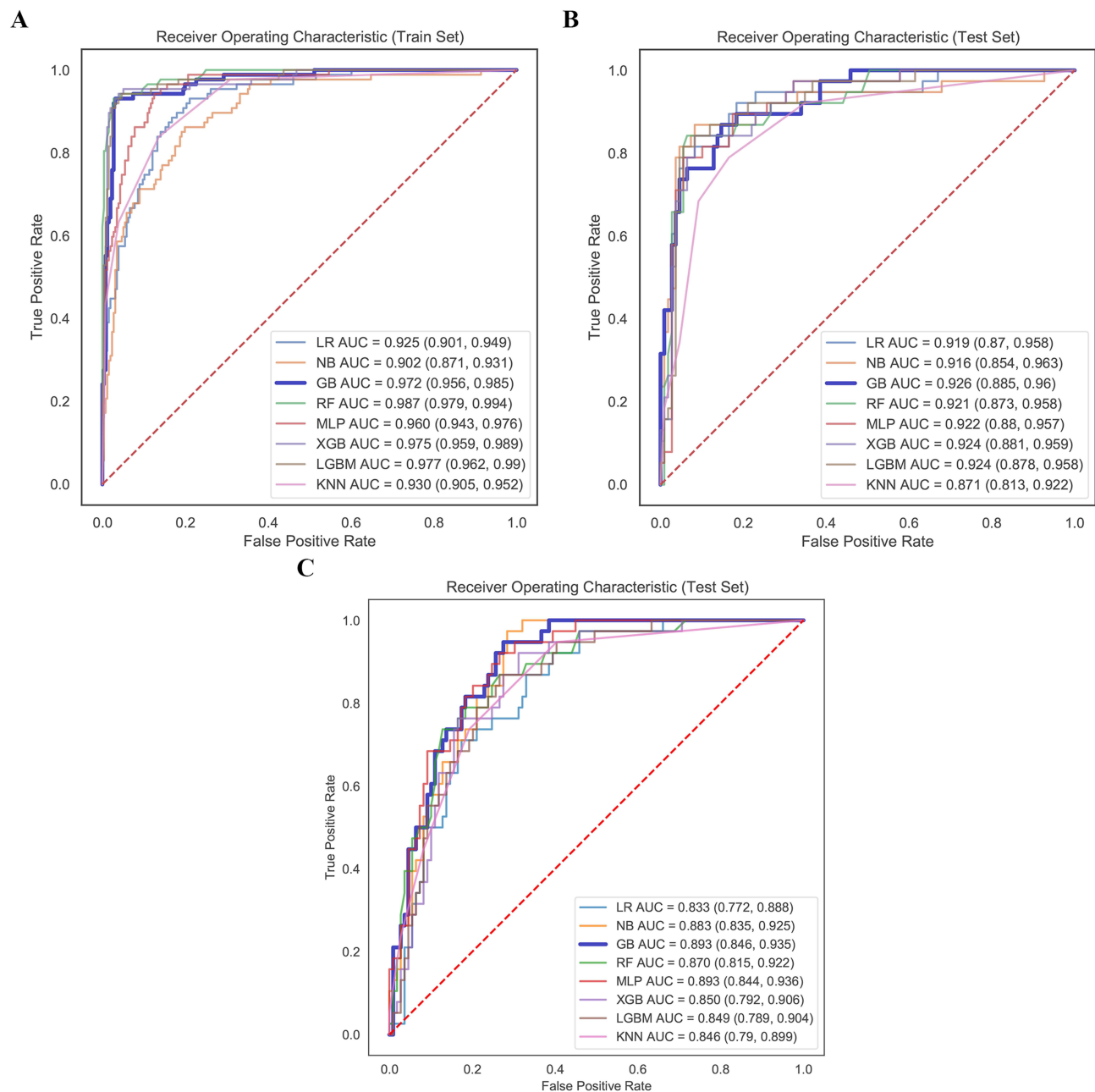


Figure 4 ROC curves of eight models for predicting worsening and de novo of SUI after PFR. **(A)** Training set; **(B)** Internal validation set; **(C)** External validation set. **Abbreviations:** LR, Logistic Regression; NB, Naive Bayes; GB, Gradient Boosting Machine; RF, Random Forest; MLP, Multilayer Perceptron; XGB, Extreme Gradient Boosting; LGBM, Light Gradient Boosting Machine; KNN, k-Nearest Neighbors; ROC, Receiver Operating Characteristic; SUI, Stress Urinary Incontinence; AUC, Area Under the Curve.

the MLP model outperformed all other models in terms of discrimination, calibration, and clinical utility. Specifically, the MLP achieved an AUC of 0.922 (95% CI: 0.880–0.957) in internal validation and 0.893 (95% CI: 0.844–0.936) in external testing, indicating high discriminative ability. Its accuracy (0.816), sensitivity (0.816), and specificity (0.817) were consistently high across datasets, highlighting robust predictive performance. In contrast, other models exhibited notable limitations. The RF model, for example, demonstrated a significant drop in sensitivity to 0.500 during external validation, missing approximately half of high-risk cases. GB and NB had suboptimal specificity (0.697 and 0.716, respectively), leading to excessive false positives and the potential for unnecessary AIS interventions. The traditional LR model performed well internally (AUC = 0.919) but showed reduced external validity (AUC = 0.833), likely due to

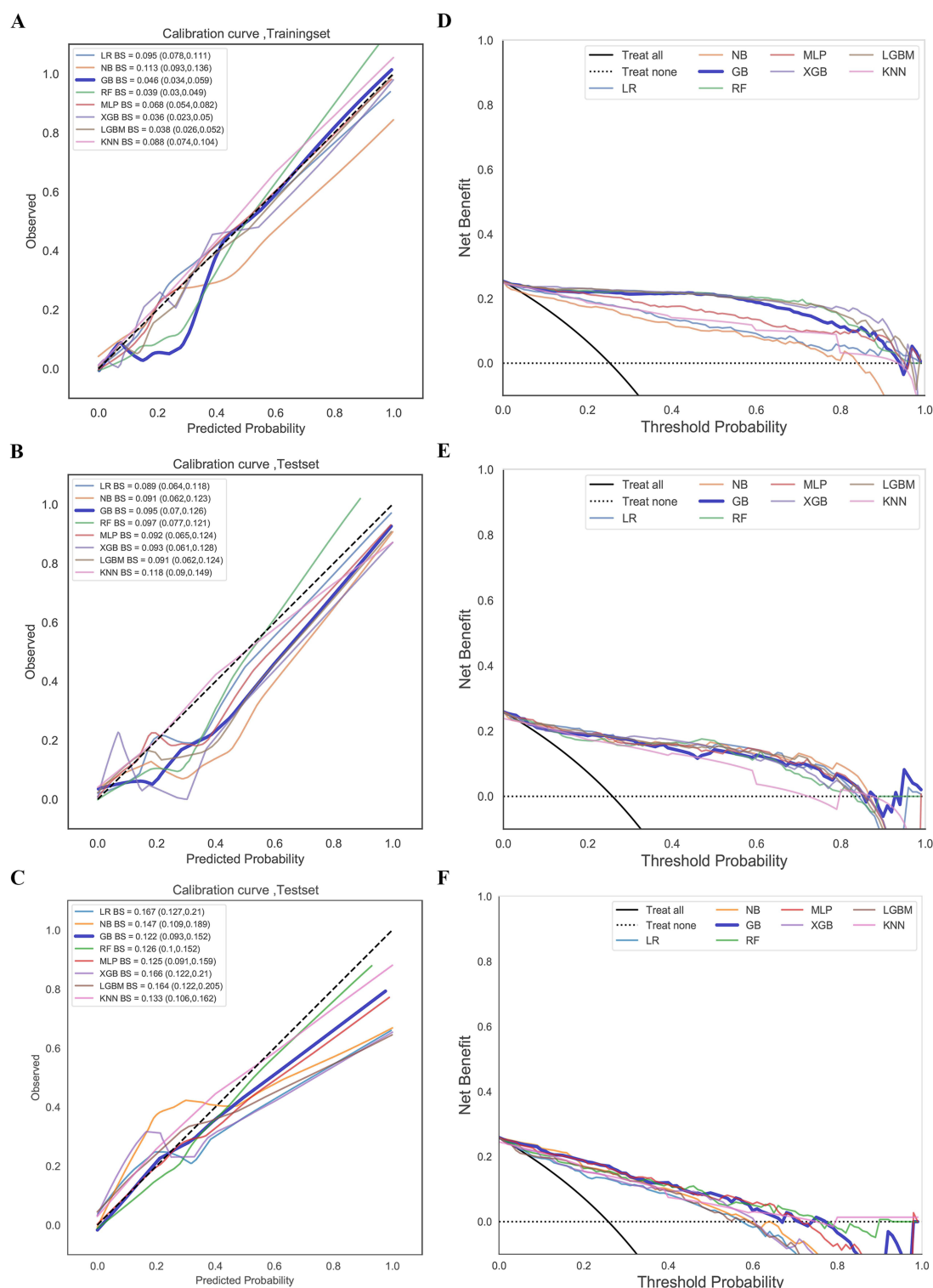


Figure 5 The eight machine learning models developed for predicting worsening and de novo of SUI following pelvic floor reconstruction were evaluated using calibration curves, brier scores, and clinical decision curves to assess their predictive accuracy and reliability. **(A)** Training set; **(B)** Internal validation set; **(C)** External validation set; **(D)** Training set; **(E)** Internal validation set; **(F)** External validation set. Calibration curves, which plot predicted probabilities against observed frequencies, were used to assess the alignment between model predictions and actual outcomes. A well-calibrated model would closely follow the diagonal line ($y = x$). The BS quantifies the overall magnitude of prediction error, with lower values indicating greater predictive accuracy. The Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement recommends reporting calibration curves, as they visually demonstrate how well estimated risks correspond to true event rates. The BS acts as a complementary, quantitative measure of calibration performance, with superior probability estimation reflected in lower values. Together, these tools offer a comprehensive evaluation of the model's predictive accuracy. DCA further quantified the net benefit of the models within a specified probability threshold range. **Abbreviations:** LR, Logistic Regression; NB, Naive Bayes; GB, Gradient Boosting Machine; RF, Random Forest; MLP, Multilayer Perceptron; XGB, Extreme Gradient Boosting; LGBM, Light Gradient Boosting Machine; KNN, k-Nearest Neighbors; SUI, stress urinary incontinence; BS, Brier score; DCA, Decision curve analysis.

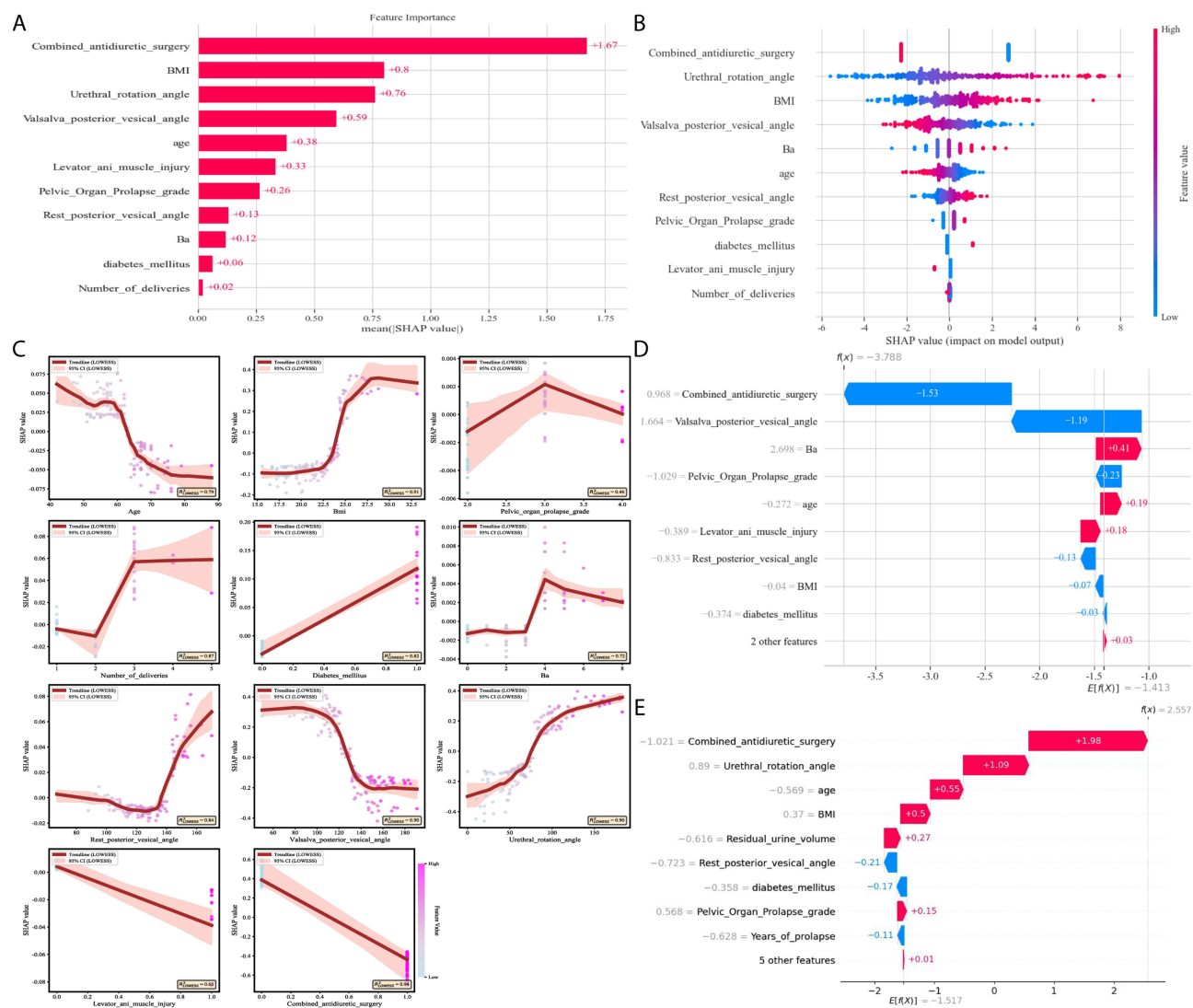


Figure 6 Interpretation of the optimal prediction model using SHAP for both global and local explainability. **(A)** SHAP Summary Bar Plot: This plot visually represents the overall importance of each feature in the model. Features are ranked by their mean absolute SHAP value, which quantifies their average impact on predictions across the entire dataset. The horizontal length of each bar indicates the magnitude of each feature's contribution to the model's decisions; **(B)** SHAP Beeswarm Plot: This visualization shows the distribution of SHAP values for each patient in the dataset. Each point represents an individual instance, with its position relative to the centerline indicating its effect: points to the right (positive SHAP values) increase the predicted risk of SUI progression, while points to the left (negative SHAP values) reduce it. The color of the points reflects the original feature value, with red representing high values and blue indicating low values, thus revealing patterns of risk (eg, high BMI, shown in red, is predominantly associated with increased risk); **(C)** SHAP Dependence Plots: These charts depict the nonlinear relationship between a specific feature's value (x-axis) and its corresponding impact on the prediction (SHAP value, y-axis). A LOESS-fitted trend line (in red) clarifies the overall effect: an upward trend indicates value ranges where the feature acts as a risk factor, while a downward trend suggests a protective effect. This offers an intuitive method for establishing clinical decision thresholds (eg, identifying a BMI > 22.5 kg/m² as a critical risk point). SHAP, SHapley Additive exPlanations; BMI, body mass index; SUI, stress urinary incontinence. **(D and E)** SHAP Waterfall Plot: This diagram (D) deconstructs the prediction for a single, representative patient, illustrating how the model arrives at its final output from the base value ($E[f(x)]$), the average prediction. Features are ordered by the size of their contribution. Red bars represent factors that increased the patient's predicted risk, while blue bars indicate protective factors that lowered it; Similarly, another patient (E) had a baseline model output value ($E[f(x)]$) of -1.517. When parameters such as combined antidiuretic surgery, urethral rotation angle, and age positively influenced the postoperative SUI outcome, the final output value was 2.557, indicating this patient was in the high-risk group.

its linear architecture and reliance on P-value-based variable selection. LR lacks the regularization and feature compression inherent to LASSO, making it more susceptible to overfitting specific training data patterns. By contrast, the MLP model, with a two-layer nonlinear neural architecture (hidden_layer_sizes = (100,50)), effectively captured complex interactions among predictors. L2 regularization (alpha = 0.0001) mitigated overfitting, and nested cross-validation with stratified sampling ensured balanced, high-performance outcomes across validation cohorts.²³ Beyond discrimination, the MLP model exhibited excellent calibration, with Brier scores of 0.092 (internal) and 0.125 (external). DCA further demonstrated substantial clinical net benefit across threshold probabilities of 0–0.8 (internal)

and 0–0.7 (external), outperforming both “intervene-all” and “intervene-none” strategies and highlighting its practical utility.

Feature selection employed a dual-path approach combining traditional regression analysis and LASSO regression. The traditional method identified independent predictors via univariate and multivariate LR, while LASSO used L1 regularization to optimize the feature subset. The two approaches yielded nearly overlapping feature sets, validating the robustness of predictor selection. Notably, sacrospinous ligament fixation (SSLF; OR = 0.20, $P = 0.03$), identified as a protective factor in multivariate LR, was excluded in the LASSO model due to moderate collinearity with AIS (VIF = 5.2). All other retained features had VIF values below 3. This exclusion reflects the core mechanism of L1 regularization: compressing the feature space by prioritizing variables with superior predictive power among highly correlated features.²⁴ Importantly, the exclusion of SSLF does not diminish its clinical relevance; rather, it indicates that its predictive value overlaps with AIS. Literature supports the protective effect of SSLF in reducing postoperative SUI,²⁵ and both this study and current clinical principles suggest that combining SSLF with other techniques (eg, Prolift mesh combined with TVT-O sling) can substantially improve postoperative SUI outcomes.

Clinical Significance of Key Predictive Factors (SHAP Analysis)

The strong predictive performance of the proposed model provides a solid foundation for clinical application. However, translating model outputs into actionable clinical decisions requires interpretability. SHAP analysis bridges this gap by elucidating the biological relevance of each predictor and identifying clinically meaningful risk thresholds. The SHAP summary bar plot identified concomitant AIS, BMI, urethral rotation angle, bladder angle after Valsalva maneuver, and age as the five most influential predictors.

Anti-incontinence surgery is primarily performed to treat SUI, with common procedures including midurethral suspension (MUS) and lower urethral slings. These procedures improve urinary continence by reinforcing urethral support structure or increasing urethral closure pressure.²⁶ POP may obscure latent SUI symptoms; restoration of normal pelvic anatomy after prolapse repair can unmask underlying sphincter dysfunction. Combined surgical approaches simultaneously correct anatomical defects and enhance urethral support, thereby significantly reducing the risk of postoperative SUI.²⁷ Previous studies report that the incidence of de novo SUI after prolapse surgery alone ranges from 20% to 35%, whereas concomitant anti-incontinence procedures such as MUS or Burch colposuspension substantially reduce this risk.²⁸ Consistent with these findings, the SHAP beeswarm plot in the present study showed predominantly negative SHAP values for AIS, identifying it as the most influential protective factor against postoperative SUI.

Obesity contributes to sustained increases in intra-abdominal pressure, placing chronic mechanical stress on the pelvic floor musculature and fascial support, thereby impairing their functional integrity. This mechanism represents a shared pathophysiological basis for both POP and SUI in obese women.^{29,30} Prolonged mechanical loading may also result in neuromuscular injury, further exacerbating continence dysfunction.³¹ The SHAP dependence plot demonstrated a significant positive correlation between SHAP values and BMI > 24 kg/m², consistent with multivariate regression results (OR = 1.324, $P < 0.001$). This finding aligns with prior studies showing that elevated BMI, particularly above 24 kg/m², is an independent risk factor for adverse outcomes following PFR, including persistent or de novo urinary incontinence.^{6,32} These results emphasize the importance of perioperative weight management in patients undergoing PFR.

The urethral rotation angle reflects the integrity of pelvic floor support structures. An increased angle indicates excessive urethral mobility and compromised surrounding support, which may impair urinary control even after anatomical correction.³³ Notably, when the urethral rotation angle exceeds 60°, there is a risk of functional “over-correction”. Excessive elevation of the suspension point, inappropriate tension, or abnormal traction vectors during static conditions or Valsalva maneuvers, may induce non-physiological posterior or superior urethral displacement, resulting in excessive rotation or torsion. Such iatrogenic alterations can disrupt the urethral closure mechanism and contribute to postoperative SUI.^{34,35} In this study, SHAP values increased linearly with larger urethral rotation angles, supporting prior observations and suggesting that 60° may represent a clinically relevant intervention threshold.

The posterior bladder angle during the Valsalva maneuver is another important anatomical indicator. Previous studies have shown that Green type III bladder configuration maintains a complete but smaller posterior bladder angle compared with type II, and this reduced angle is strongly associated with occult urinary incontinence (OR = 10.9).³⁶ In the present study, SHAP dependence analysis identified a preoperative posterior bladder angle of 60–130° as a high-risk window for postoperative SUI. PFR surgery may partially mitigate this risk by relieving obstruction and modifying bladder neck or urethral positioning, thereby improving functional outcomes.³⁷

Age also emerged as a significant predictor, though its effect warrants nuanced interpretation. While earlier studies suggest that women aged ≥ 70 years are at increased risk of persistent or de novo SUI after prolapse surgery,³⁸ our findings indicate that patients aged 40–65 years represent a particularly vulnerable group. This apparent discrepancy may reflect differing mechanisms of continence dysfunction across life stages. Epidemiological studies show that SUI prevalence peaks between 40 and 59 years, coinciding with the peri-menopausal period characterized by declining estrogen levels and reduced connective tissue elasticity. A large Chinese study reported the highest SUI prevalence (28%) among women aged 50–59, consistent with global trends.³⁹ During this transitional phase, pelvic floor support structures experience diminishing functional reserve but have not yet reached the compensatory adaptations seen in advanced age. This “transitional vulnerability” may predispose patients to continence failure following surgical stress, explaining the elevated risk observed in this age group.

Comparison with Previous Studies

Previous predictive models for postoperative SUI are limited by methodological constraints, including reliance on linear assumptions, modest predictive accuracy, and restricted generalizability. Traditional LR models, in particular, may fail to fully capture the complex, nonlinear relationships underlying SUI risk. This limitation is evident in several published studies on de novo SUI prediction following POP repair. For instance, Liang et al⁴⁰ developed an LR model incorporating POP-Q scores, surgical variables, pelvic floor ultrasound findings, and urodynamic parameters, achieving a modest bootstrap-corrected AUC of 0.826 in internal validation. Similarly, He et al⁴¹ integrated clinical indicators (including a positive pressure urethral pressure test and Ba point) with pelvic floor ultrasound parameters (such as internal urethral orifice funneling and LAM injury). Their LR model demonstrated good discrimination, with AUCs of 0.885 in the training set and 0.823 in the internal validation set; however, the absence of independent external validation limits the robustness and clinical generalizability of these findings. Although ML models have been introduced to overcome the linear constraints of LR, their performance has remained variable. For example, Fu et al¹⁰ employed an XGBoost-based model and reported an external validation AUC of 0.704, which, while not exceptional, highlights the challenges of achieving stable generalization in real-world settings. Xu et al⁴² compared LR, support vector machine (SVM), C5.0 decision tree, and neural network models using combined clinical and pelvic floor ultrasound data, with the neural network model achieving the highest internal validation AUC (0.857). Nevertheless, most prior ML studies lacked independent external validation and did not incorporate interpretable frameworks, such as SHAP analysis, to clarify how individual predictors influence model decisions, an omission that limits clinical trust and applicability.

To address these limitations, the MLP model developed in this study introduces several key innovations. First, enhanced generalizability and robustness were achieved through a multicenter cohort design and the inclusion of an independent external validation set ($n = 100$). The dual-layer neural network architecture (hidden layer sizes = 100 and 50) effectively captured complex nonlinear interactions among predictors, while L2 regularization ($\alpha = 0.0001$) mitigated overfitting. As a result, the model demonstrated excellent discrimination, with an AUC of 0.922 in internal validation and a consistently high AUC of 0.893 in external validation, outperforming both traditional LR models and previously reported ML approaches. Second, this study implemented an innovative dual-path feature selection strategy that integrated traditional regression analysis with LASSO regularization. While univariate and multivariate regression identified seven independent risk factors, LASSO optimized the feature space by retaining 11 key predictors and excluding redundant variables, such as sacrospinous ligament fixation, due to its collinearity with concomitant AIS. This approach preserved clinically meaningful variables, such as BMI, a well-established risk factor, while improving model stability and parsimony. Third, and most importantly, this study advances interpretability through comprehensive SHAP analysis. By quantifying the direction and magnitude of each feature’s contribution to model predictions, SHAP

effectively addresses the “black box” limitation of ML models. This analysis identified clinically actionable thresholds, such as BMI > 24 kg/m² and urethral rotation angle > 60°, thereby transforming complex model outputs into transparent, decision-support tools suitable for real-world clinical application.

Limitations and Future Perspectives

Several limitations of this study should be acknowledged. First, the retrospective study design is inherently susceptible to selection bias and measurement bias. This limitation is further compounded by incomplete documentation of certain clinically relevant details, such as the specific intraoperative considerations that guided the decision to perform concomitant AIS. Second, the study population was predominantly drawn from a single geographic region, which may limit the representativeness and generalizability of the findings to broader or more diverse populations. Third, although standardized quality control procedures were implemented across centers, unavoidable heterogeneity remained in ultrasound equipment specifications, operator expertise, and surgical technique variations. These factors may have introduced inter-center variability that is difficult to fully control in retrospective multicenter analyses. Future research will focus on several directions. First, prospective, multicenter validation studies are warranted to further refine and validate the imaging feature system. Incorporating detailed intraoperative variables and surgeon-related factors may enhance model generalizability and clinical applicability. Second, advanced deep learning approaches, such as convolutional neural networks (CNNs), should be explored to enable direct analysis of raw pelvic floor ultrasound images or dynamic cine loops. The integration of multimodal data, including electronic medical record (EMR) text, urodynamic waveforms, and imaging features, may further improve predictive accuracy and support the development of more comprehensive and intelligent clinical decision-support systems.

Conclusion

In this study, we developed and validated a MLP model capable of accurately predicting the risk of de novo or worsened stress urinary incontinence following pelvic organ prolapse repair. The model demonstrated superior discriminative performance compared with traditional logistic regression and existing machine learning approaches, while also achieving excellent calibration and clinically meaningful net benefit. Specifically, the MLP model achieved an AUC of 0.922 in internal validation and 0.893 in external validation, substantially outperforming the traditional logistic regression model (AUC = 0.833 in external validation) and the previously reported XGBoost model (AUC = 0.714). The MLP model also demonstrated more balanced sensitivity and specificity (0.816 and 0.817, respectively) compared with the Random Forest model, which showed a marked sensitivity decline to 0.500 in external validation. Furthermore, the MLP model maintained excellent calibration (Brier score = 0.125 in external validation) and clinically meaningful net benefit across a wide threshold probability range (0–0.7), whereas the KNN model exhibited negative net benefit at higher thresholds. These findings highlight the MLP model’s superior robustness, generalizability, and clinical utility in predicting postoperative SUI risk. Importantly, SHAP-based interpretability analysis elucidated the biological relevance and clinically actionable thresholds of key predictors, including concomitant AIS, body mass index, urethral rotation angle, Valsalva retrovesical angle, and age, thereby enhancing transparency and supporting individualized clinical decision-making. Future work will focus on prospective validation and the development of dynamic, longitudinal risk prediction models incorporating postoperative follow-up data, with the ultimate goal of improving postoperative outcomes and quality of life for patients undergoing prolapse repair.

Data Sharing Statement

Data can be acquired upon reasonable request from the corresponding author.

Ethics Approval and Consent to Participate

This study has been approved by the Ethics Committee of Wannan Medical University (Approval No.: 2024-207), and the study was conducted in accordance with the Declaration of Helsinki. Given the retrospective nature of the study and the use of anonymized data, the requirement for informed consent was waived by the institutional review board.

Author Contributions

Liying He: Conceptualization, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Funding acquisition, Writing – original draft. Yin Chen: Investigation, Resources, Data curation, Validation, Writing – review & editing. Qiang Ma: Methodology, Validation, Formal analysis, Writing – review & editing. Feng Jiang: Supervision, Project administration, Validation, Writing – review & editing. Huanqing Xu: Methodology, Software, Formal analysis, Visualization, Writing – review & editing. Biyun Sun: Conceptualization, Supervision, Validation, Writing – review & editing. Yonghong Luo: Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing. All authors made a significant contribution to the work reported, whether in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising, or critically reviewing the article; gave final approval of the version to be published; agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

The research was funded by Young and Middle-aged Research Fund Project of Wannan Medical University (WK2023ZQN29) and Clinical and Translational Research Project of Anhui Province (202427b10020084).

Disclosure

The authors declare that they have no competing interests in this work.

References

1. Abebe D, Kure MA, Demssie EA, et al. One in ten ever-married women who visited health facilities for various reasons have pelvic organ prolapse in Harari regional state, Eastern Ethiopia. *BMC Womens Health*. 2022;22(1):223. doi:10.1186/s12905-022-01817-8
2. Pang H, Zhang L, Han S, et al. A nationwide population-based survey on the prevalence and risk factors of symptomatic pelvic organ prolapse in adult women in China - a pelvic organ prolapse quantification system-based study. *BJOG*. 2021;128(8):1313–1323. doi:10.1111/1471-0528.16675
3. Wong JWH, Ramm O. Urinary incontinence and pelvic organ prolapse. *Clin Obstet Gynecol*. 2021;64(2):314–320. doi:10.1097/GRF.0000000000000615
4. Moosavi SY, Samad-Soltani T, Hajebrabimi S, et al. Determining the risk factors and characteristics of de novo stress urinary incontinence in women undergoing pelvic organ prolapse surgery: a systematic review. *Turk J Urol*. 2020;46(6):427–435. doi:10.5152/tud.2020.20291
5. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1(5):206–215. doi:10.1038/s42256-019-0048-x
6. Oliveira IKB, da Silva Calisto SL, Ferreira CWS, et al. Occult urinary incontinence, diabetes, obesity, prolapse severity, and type of surgical repair as risk factors for de novo stress urinary incontinence in women undergoing surgical repair of pelvic organ prolapse: a systematic review and meta-analysis. *Neurourol Urodyn*. 2025;44(1):194–206. doi:10.1002/nau.25619
7. Wang X, Lu Y. Study on the factors affecting worsening or new-onset stress urinary incontinence after pelvic organ prolapse surgery [in Chinese]. *Chin J Obstet Gynecol*. 2021;22(2):139–142. doi:10.13390/j.issn.1672-1861.2021.02.009
8. Wang Q, Wang X, Jiang X, et al. Machine learning in female urinary incontinence: a scoping review. *Digit Health*. 2024;10:2055207.
9. Yu H, Zhao H, Liu D, et al. Prediction of myofascial pelvic pain syndrome based on random forest model. *Heliyon*. 2024;10(11):e31928. PMID: 38868063; PMCID: PMC11167342. doi:10.1016/j.heliyon.2024.e31928
10. Fu L, Huang G, Sun Z, Zhu L. Predicting the occurrence of stress urinary incontinence after prolapse surgery: a machine learning-based model. *Ann Transl Med*. 2023;11(6):251. PMID: 37082678; PMCID: PMC10113075. doi:10.21037/atm-22-3648
11. Wang Y, Rong J, Wei Z, et al. Predicting ICU mortality in heart failure patients based on blood tests and vital signs. *Front Cardiovasc Med*. 2025;12:1590367. PMID: 40636830; PMCID: PMC12238217. doi:10.3389/fevm.2025.1590367
12. Nistal-Nuño B. Outcome prediction for critical care patients with respiratory neoplasms using a multilayer perceptron neural network. *Einstein*. 2023;21:eAO71. PMID: 37729310; PMCID: PMC10501764. doi:10.31744/einstein_journal/2023AO0071
13. Lu J, Zhu K, Yang N, et al. Radiomics and clinical features for distinguishing kidney stone-associated urinary tract infection: a comprehensive analysis of machine learning classification. *Open Forum Infect Dis*. 2024;11(10):ofae581. PMID: 39435322; PMCID: PMC11493090. doi:10.1093/ofid/ofae581
14. Hou L, He K, Zhao J, Su K, Zhang C. Application of an interpretable machine learning model based on optimal feature selection for predicting triple-vessel coronary disease: a multicenter retrospective study. *PeerJ*. 2025;13:e20423. PMID: 41394415; PMCID: PMC12700116. doi:10.7717/peerj.20423
15. Jelovsek JE, Chagin K, Brubaker L, et al. A model for predicting the risk of de novo stress urinary incontinence in women undergoing pelvic organ prolapse surgery. *Obstet Gynecol*. 2014;123(2 Pt 1):279–287. doi:10.1097/AOG.0000000000000094
16. Duarte Thibault M, Chen L, Huebner M, et al. A comparison of MRI-based pelvic floor support measures between young and old women with prolapse. *Int Urogynecol J*. 2023;34(9):2081–2088. doi:10.1007/s00192-023-05505-5
17. Toptas KK, Goksu M, Karadeniz O, et al. Can preoperative urodynamic studies predict de novo stress urinary incontinence following advanced pelvic organ prolapse surgery? *Can Urol Assoc J*. 2025;19(5):E183–E188. PMID: 40209185; PMCID: PMC12091042. doi:10.5489/cuaj.8818
18. Rosato E, Vacca L, Lombisani A, et al. Is there a role for urodynamic investigation in the management of pelvic organ prolapse? *J Clin Med*. 2025;14(4):1163. PMID: 40004694; PMCID: PMC11856066. doi:10.3390/jcm14041163

19. Su M, Wang S, Liu X. Predicting postoperative stress urinary incontinence after prolapse surgery via machine learning and regression models: development and validation study. *JMIR Med Inform.* **2025**;13:e76021. PMID: 41183292; PMCID: PMC12582534. doi:10.2196/76021
20. Lin CF, Chen HY, Wu CT, et al. Outcomes of transobturator tape surgery for stress and mixed urinary incontinence: a 12-year retrospective review. *Diagnostics.* **2025**;15(20):2611. PMID: 41153283; PMCID: PMC12564240. doi:10.3390/diagnostics15202611
21. Wang XX, Xu Y, Song YF, et al. Efficacy of two modes of transvaginal temperature-controlled radiofrequency for female stress urinary incontinence. *Int Urogynecol J.* **2025**;36(9):1759–1766. PMID: 39945801. doi:10.1007/s00192-025-06065-6
22. Yu BJ, He HC, Wang L, et al. Risk prediction models for stress urinary incontinence after pelvic organ prolapse (POP) surgery: a systematic review and meta-analysis. *BMC Womens Health.* **2025**;25(1):55. PMID: 39923045. doi:10.1186/s12905-025-03584-8
23. Ahmed FR, Al-Yateem N, Nejadghaderi SA, et al. Harnessing machine learning for predicting successful weaning from mechanical ventilation: a systematic review. *Aust Crit Care.* **2025**;38(3):101203. PMID: 40058181. doi:10.1016/j.aucc.2025.101203
24. An B, Zhang B. Logistic regression with image covariates via the combination of L1 and Sobolev regularizations. *PLoS One.* **2020**;15(6):e0234975. PMID: 32589677; PMCID: PMC7319310. doi:10.1371/journal.pone.0234975
25. Başer E, Seçkin KD, Kadiroğullari P, et al. The effect of sacrospinous ligament fixation during vaginal hysterectomy on postoperative de novo stress incontinence occurrence: a prospective study with 2-year follow-up. *Turk J Med Sci.* **2020**;50(4):978–984. PMID: 32490650; PMCID: PMC7379432. doi:10.3906/sag-2005-117
26. Lee PJ, Kuo HC. Low detrusor contractility has a less favorable outcome of anti-incontinence surgery for women with stress urinary incontinence. *Int Urol Nephrol.* **2023**;55(11):2789–2798. PMID: 37505426. doi:10.1007/s11255-023-03725-8
27. Baessler K, Christmann-Schmid C, Maher C, et al. Surgery for women with pelvic organ prolapse with or without stress urinary incontinence. *Cochrane Database Syst Rev.* **2018**;8(8):CD013108. PMID: 30121956; PMCID: PMC6513383. doi:10.1002/14651858.CD013108
28. Matsuoaka PK, Castro RA, Baracat EC, et al. Occult urinary incontinence treatment: systematic review and meta-analysis-brazilian guidelines. *Rev Bras Ginecol Obstet.* **2019**;41(2):116–123. PMID: 30786309; PMCID: PMC10418705. doi:10.1055/s-0038-1676842
29. Lee UJ, Kerkhof MH, van Leijsen SA, et al. Obesity and pelvic organ prolapse. *Curr Opin Urol.* **2017**;27(5):428–434. PMID: 28700456. doi:10.1097/MOU.0000000000000428
30. Declémy A, Haddad R, Chesnel C, et al. Prevalence of comorbidities in multiple sclerosis patients with neurogenic bladder. *Prog Urol.* **2021**;31(12):732–738. PMID: 33941463. doi:10.1016/j.purol.2020.10.011
31. Fuselier A, Hanberry J, Margaret Lovin J, et al. Obesity and stress urinary incontinence: impact on pathophysiology and treatment. *Curr Urol Rep.* **2018**;19(1):10. PMID: 29468457. doi:10.1007/s11934-018-0762-7
32. Ringel NE, Lenger SM, High R, et al. Effects of obesity on urogynecologic prolapse surgery outcomes: a systematic review and meta-analysis. *Obstet Gynecol.* **2024**;143(4):539–549. PMID: 38330397. doi:10.1097/AOG.00000000000005525
33. Guo X, Ding C, Zhang S. 4D transperineal ultrasound for the diagnosis and classification of stress urinary incontinence in postmenopausal women. *J Coll Physicians Surg Pak.* **2023**;33(4):438–442. PMID: 37190718. doi:10.29271/jcpsp.2023.04.438
34. Bansal D, Kumar S, Jain S, Dogra PN. ‘Poles apart presentation’: diagnosis and management of iatrogenic posterior urethral false tract in cases of pelvic fracture urethral injury. *BMJ Case Rep.* **2019**;12(8):e231166. PMID: 31439562; PMCID: PMC6720746. doi:10.1136/ber-2019-231166
35. Sharabiany S, Strijk GJ, Blok RD, et al. Quantifying displacement of urogenital organs after abdominoperineal resection for rectal cancer. *Colorectal Dis.* **2021**;23(11):2923–2931. PMID: 34427972; PMCID: PMC9291959. doi:10.1111/codi.15885
36. Karjalainen PK, Gillor M, Dietz HP. Predictors of occult stress urinary incontinence. *Aust N Z J Obstet Gynaecol.* **2021**;61(2):263–269. PMID: 33368207. doi:10.1111/ajo.13290
37. Sato H, Abe H, Ikeda A, et al. Severity of cystocele and risk factors of postoperative stress urinary incontinence after laparoscopic sacrocolpopexy for pelvic organ prolapse. *Gynecol Minim Invasive Ther.* **2022**;11(1):28–35. PMID: 35310120; PMCID: PMC8926053. doi:10.4103/GMIT.GMIT_2_21
38. Joukhadar R, Radosa J, Paulus V, et al. Influence of patient’s age on the outcome of vaginal and laparoscopic procedures in urogynaecology. *Geburtshilfe Frauenheilkd.* **2019**;79(9):949–958. PMID: 31523095; PMCID: PMC6739203. doi:10.1055/a-0854-5916
39. Lee UJ, Feinstein L, Ward JB, et al. Prevalence of urinary incontinence among a nationally representative sample of women, 2005–2016: findings from the urologic diseases in america project. *J Urol.* **2021**;205(6):1718–1724. PMID: 33605795. doi:10.1097/JU.0000000000001634
40. Liang Y. Risk prediction of stress urinary incontinence progression after pelvic floor reconstruction in patients with pelvic organ prolapse. Southern Medical University; **2025**. doi: 10.27003/d.cnki.gojyu.2025.001190
41. Liying H, Feng J, Wei W, et al. Value of a line chart model based on pelvic floor ultrasound parameters in predicting progressive stress urinary incontinence after pelvic floor reconstruction. *J Clin Ultrasound Med.* **2023**;25(08):613–620. doi:10.16245/j.cnki.issn1008-6978.2023.08.001
42. Xu J. Development of a predictive model for stress urinary incontinence based on pelvic floor ultrasound and clinical data. Nantong University; **2023**. doi: 10.27255/d.cnki.gntyc.2023.000071

International Journal of Women’s Health

Publish your work in this journal

The International Journal of Women’s Health is an international, peer-reviewed open-access journal publishing original research, reports, editorials, reviews and commentaries on all aspects of women’s healthcare including gynecology, obstetrics, and breast cancer. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-womens-health-journal>

Dovepress
Taylor & Francis Group