Heliyon 10 (2024) e31928

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Prediction of myofascial pelvic pain syndrome based on random forest model

Hang Yu [a], Hongguo Zhao [a], Dongxia Liu [a], Yanhua Dong [a], Manman Nai [a], Yikun Song [a], Jiaxi Liu [a], Luwen Wang [a], Lei Li [a,**], Xinbin Li [b,*]

[a] The Department of Obstetrics and Gynecology, The Third Affiliated Hospital of Zhengzhou University, Zhengzhou, 450052, PR China
[b] School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, 710072, PR China

ARTICLE INFO

[Keywords]:
Myofascial pelvic pain syndrome
Random forest
Logistic regression
Modified Oxford muscle strength grading
Pelvic floor pressure assessment

[ABSTRACT]

Objective: The objective is to construct a random forest model for predicting the occurrence of Myofascial pelvic pain syndrome (MPPS) and compare its performance with a logistic regression model to demonstrate the superiority of the random forest model.

Methods: We retrospectively analyze the clinical data of female patients who underwent pelvic floor screening due to chronic pelvic pain at the Pelvic Floor Rehabilitation Center of the Third Affiliated Hospital of Zhengzhou University from January 2021 to December 2023. A total of 543 female patients meeting the study's inclusion and exclusion criteria are randomly selected from this dataset and allocated to the MPPS group. Furthermore, 702 healthy female patients who underwent pelvic floor screening during routine physical examinations within the same time-frame are randomly selected and assigned to the non-MPPS group. Chi-square test and rank-sum test are used to select demographic variables, pelvic floor pressure assessment data variables, and modified Oxford muscle strength grading data for logistic univariate analysis. The selected variables are further subjected to multivariate logistic regression analysis, and a random forest model is also established. The predictive performance of the two models is evaluated by comparing their accuracy, sensitivity, specificity, precision, receiver operating characteristic (ROC) curve, and area under the curve (AUC) area.

Results: Based on a dataset of 1245 cases, we implement the random forest algorithm for the first time in the screening of MPPS. In this investigation, the Logistic regression model forecasts the accuracy, sensitivity, specificity, and precision of MPPS at 69.96 %, 57.46 %, 79.63 %, and 68.57 % respectively, with an AUC of the ROC curve at 0.755. Conversely, the random forest prediction model exhibits accuracy, sensitivity, specificity, and precision rates of 87.11 %, 90.66 %, 90.91 %, and 83.51 % respectively, with an AUC of the ROC curve at 0.942. The random forest model showcases exceptional predictive performance during the initial screening of MPPS.

Conclusion: The random forest model has exhibited exceptional predictive performance in the initial screening evaluation of MPPS disease. The development of this predictive framework holds significant importance in refining the precision of MPPS prediction within clinical environments and elevating treatment outcomes. This research carries profound global implications, given the potentially elevated misdiagnosis rates and delayed diagnosis proportions of MPPS on a worldwide scale, coupled with a potential scarcity of seasoned healthcare providers. Moving forward,

* Corresponding author.
** Corresponding author.
E-mail addresses: beileili2004@126.com (L. Li), xinbin_li@163.com (X. Li).

https://doi.org/10.1016/j.heliyon.2024.e31928
Received 31 January 2024; Received in revised form 23 May 2024; Accepted 24 May 2024
Available online 27 May 2024

continual refinement and validation of the model will be imperative to further augment the precision of MPPS risk assessment, thereby furnishing clinicians with more dependable decision-making support in clinical practice.

Myofascial pelvic pain syndrome (MPPS) is characterized by pain, muscle stiffness, and spasms in the fascia of the lower back, sacrum, buttocks, and legs, accompanied by highly sensitive myofascial trigger points (MTrPs) [1]. The prevalence of MPPS in females ranges from 5.7 % to 26.6 % [2], while in males, it ranges from 2.9 % to 9.7 % [3]. Owing to limited clinical awareness, MPPS is frequently underdiagnosed or misdiagnosed, resulting in ineffective treatment and impairments in women's quality of life and psychological well-being [4]. Presently, scholars both domestically and internationally lack consensus on the etiology and pathogenesis of MPPS [5]. Since there are no specific laboratory or imaging findings, MPPS lacks unified diagnostic criteria. Therefore, the diagnosis of MPPS is exclusive, requiring women to present with clinical complaints of pelvic pain, palpable MTrPs, and the exclusion of other factors causing chronic pelvic pain through medical history, physical examination, and relevant auxiliary examinations [6]. Therefore, an effective predictive model is essential to predict the occurrence of MPPS. We propose a random forest model that utilizes women's clinical baseline parameters, pelvic floor pressure assessment, and modified Oxford muscle grading assessment data to predict the risk of developing MPPS, providing a theoretical basis for the clinical diagnosis of MPPS.

1. Objectives and methods

1.1. Objectives

This study is a retrospective analysis. The clinical data of female patients who underwent pelvic floor screening due to chronic pelvic pain at the Pelvic Floor Rehabilitation Center of the Third Affiliated Hospital of Zhengzhou University from January 2021 to December 2023 is retrospectively analyzed. A total of 543 female patients meeting the study's inclusion and exclusion criteria are randomly selected from this dataset and allocated to the MPPS group. Furthermore, 702 healthy female patients who underwent pelvic floor screening during routine physical examinations within the same timeframe are randomly selected and assigned to the non-MPPS group. Ethical approval for this study was granted by the Ethics Committee of the Third Affiliated Hospital of Zhengzhou University (Ethics Approval Number: 2023-206-01, Approval Date: August 23, 2023).

2. Inclusion and exclusion criteria

2.1. Inclusion criteria

To be eligible for participation, individuals have to meet specific criteria: (1) Exhibit a history of chronic pelvic pain lasting a minimum of 3–6 months, diagnosed as Myofascial Pelvic Pain Syndrome (MPPS) [7]. (2) Not have undergone any other treatment pertaining to this condition in the preceding month. (3) Be aged 18 years or above. (4) Patients with a history of sexual activity who can undergo vaginal examinations.

2.2. Exclusion criteria

Patients meeting any of the following criteria are excluded from the study: (1) Patients with chronic pelvic pain (CPP) attributed to known factors like infection, adenomyosis, uterine fibroids, etc. (2) Patients who have undergone gynecological examination, gynecological and urological ultrasound, examination of vaginal secretions and urine pathogens, and have been excluded due to acute infectious diseases or other organic diseases. (3) Pregnant patients or those within 3 months postpartum are excluded. (4) Patients with implanted cardiac pacemakers, metal intrauterine devices, or allergies to electrical stimulation are excluded. (5) Patients with active vaginal bleeding, local pelvic skin/mucosal damage, or infection are excluded. (6) Patients with other pelvic floor dysfunction diseases like urinary incontinence or pelvic organ prolapse are excluded. (7) Patients with neurological or psychiatric disorders or other severe illnesses that could impede cooperation are excluded. (8) Patients with a history of spinal disc disease, sciatica, or other neurological disorders are excluded. (9) Patients with incomplete clinical baseline data are excluded.

3. Data collection

In this study, we gather case data from the study participants, encompassing age, height, weight, body mass index (BMI), pregnancy history, and childbirth history [8].

Participants included in the study underwent pelvic floor pressure assessment administered by trained therapists specializing in pelvic floor care. The assessment utilized the Myotrac biofeedback device manufactured by Nanjing Vishee Medical Technology Co., Ltd. Various parameters are automatically recorded, including the average value and the coefficient of variation during the pre-resting phase, the maximum value and relaxation time during the fast contraction phase, the average value, coefficient of variation, and relaxation time during the tension contraction phase, and the average value and coefficient of variation are recorded during both the endurance contraction phase and the post-resting phase.

The modified Oxford grading scale is used for pelvic floor muscle assessment. The evaluation criteria for pelvic floor muscle strength assessment [9] are as follows: Grade I indicates a slight contraction of the vaginal muscles during testing. Grade II signifies that the vaginal muscles can sustain a contraction for 2 s and repeat it twice. Grade III denotes that the vaginal muscles can sustain a contraction for 3 s and repeat it three times. Grade IV indicates that the vaginal muscles can sustain a contraction for 4 s and repeat it four times. Grade V reflects significant contraction of the vaginal muscles, with the ability to sustain a contraction for 5 s or longer and repeat it five times or more. These assessments are conducted to gather comprehensive data on pelvic floor muscle pressure and strength in the study participants.

4. Statistical analysis methods

Statistical analysis is conducted using *SPSS* 26.0 and *MATLAB* (version 2020b) software. The process modeling and analysis method is illustrated in Fig. 1. Different methods are employed for various types of data: For continuous variables, data are presented as median (P_{25}, P_{75}), and between-group comparisons utilize the Mann-Whitney *U* test. For categorical variables, data are presented as counts and percentages (%), and between-group comparisons are conducted using the chi-square test. Logistic regression and random forest algorithms are utilized to develop a prediction model for Myofascial Pelvic Pain Syndrome (MPPS). The predictive accuracy, ROC curve, and AUC (Area Under the Curve) are compared for these models. A significance level of $P < 0.05$ is deemed as statistically significant.

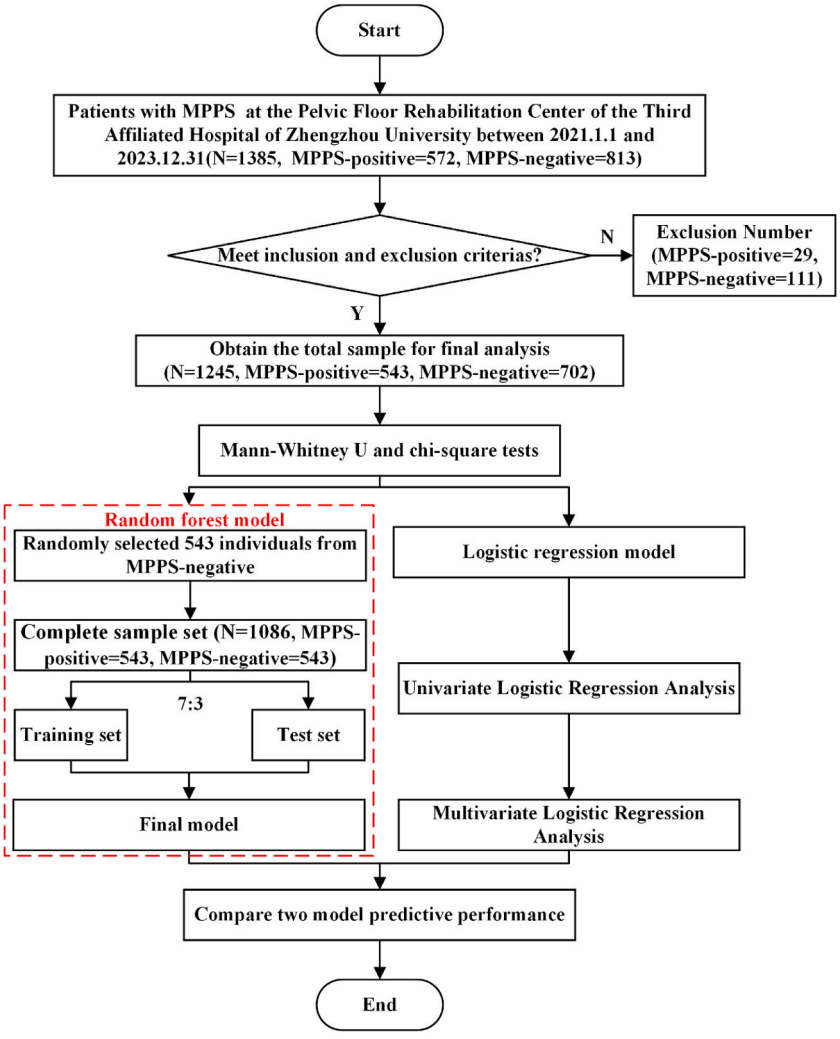

Fig. 1. Process of modeling and analysis method.

5. Results

5.1. Univariate and multivariate logistic analysis of factors influencing MPPS

The results of the chi-square test and Mann-Whitney *U* test revealed statistically significant differences ($P < 0.05$) between the two groups in the following variables: parity, gravidity, number of vaginal deliveries, perineal laceration, average value and the coefficient of variation during the pre-resting phase, the maximum value and relaxation time during the fast contraction phase, the average value and coefficient of variation during the tension contraction phase, the average value and coefficient of variation during the endurance contraction phase, and the coefficient of variation during the post-resting phase. Please refer to Table 1 and Table 2 for detailed analysis results.

5.1.1. Univariate logistic regression analysis results

The results of the univariate logistic regression analysis indicates significant differences ($P < 0.05$) between the two groups in various variables, including parity, gravidity, number of vaginal deliveries, perineal laceration, maximum value and relaxation time in the fast contraction phase, the average value and coefficient of variation in the tension contraction phase, the average value and coefficient of variation in the endurance contraction phase, and the coefficient of variation in the post-resting phase. These results suggest a significant association between these factors and MPPS. For detailed analysis results, please refer to Table 3.

5.1.2. Multivariate logistic regression analysis results

In the multivariate logistic regression analysis, the factors that demonstrated a significant association ($P < 0.05$) with MPPS in the chi-square test, Mann-Whitney *U* test, and univariate logistic regression analysis are included. The results indicates that the following factors independently influence MPPS: parity (OR = 1.475, $P < 0.05$), average value in the pre-resting phase (OR = 1.347, $P < 0.05$), coefficient of variation in the tension contraction phase (OR = 7.043, $P < 0.05$), average value in the endurance contraction phase (OR = 0.83, $P < 0.05$), coefficient of variation in the endurance contraction phase (OR = 4.919, $P < 0.05$), and coefficient of variation in the post-resting phase (OR = 14.569, $P < 0.05$). These results indicate that these factors play an independent role in the development of MPPS. For detailed analysis results, please refer to Table 4.

5.2. Random forest model

In this study, the prediction model is developed using *MATLAB* (version 2020b). A total of 1245 cases are collected, comprising 543 MPPS-positive and 702 MPPS-negative cases. To enhance result accuracy when employing the RF method for binary classification,

Table 1
General demographic characteristics of MPPS and non-MPPS groups (Example (%)).

Feature	Total number of people(n = 1245)	The non-MPPS group(n = 702)	The MPPS group (n = 543)	χ^2	*P*
Age (years)				1.134	0.567
<30	175(14.06 %)	88(12.54 %)	87(16.02 %)		
30~40	629(50.52 %)	338(48.15 %)	291(53.59 %)		
>40	441(35.42 %)	276(39.32 %)	165(30.39 %)		
BMI(Kg/m^2)				1.015	0.602
<18.5	73(5.86 %)	38(5.41 %)	35(6.45 %)		
18.5~<24	770(61.85 %)	431(61.40 %)	339(62.43 %)		
≥24	402(32.29 %)	233(33.19 %)	169(31.12 %)		
Number of pregnancies (times)				78.994	<0.001
0	33(2.65 %)	30(4.27 %)	3(0.55 %)		
1~2	448(35.98 %)	299(42.59 %)	149(27.44 %)		
>2	764(61.37 %)	373(53.13 %)	391(72.01 %)		
Number of production (times)				37.939	<0.001
0	123(9.88 %)	51(7.26 %)	72(13.26 %)		
1~2	965(77.51 %)	562(80.06 %)	403(74.22 %)		
>2	157(12.61 %)	89(12.67 %)	68(12.52 %)		
Number of vaginal births (times)				18.846	<0.001
0	395(31.73 %)	207(29.49 %)	188(34.62 %)		
1~2	733(58.88 %)	427(60.83 %)	306(56.46 %)		
>2	117(9.40 %)	68(9.69 %)	49(9.02 %)		
Number of cesarean sections (times)				1.017	0.601
0	871(69.96 %)	495(70.51 %)	376(69.24 %)		
1~2	363(29.16 %)	200(28.49 %)	163(30.02 %)		
>2	11(0.88 %)	7(1.00 %)	4(0.74 %)		
History of perineal laceration				7.027	0.008
Yes	54(4.34 %)	21(2.99 %)	33(6.08 %)		
No	1191(95.66 %)	681(97.01 %)	510(93.92 %)		
History of hysterectomy				0.203	0.653
Yes	47(3.78 %)	25(3.56 %)	22(4.05 %)		
No	1198(96.22 %)	677(96.44 %)	521(95.95 %)		

Table 2
Univariate analysis of height, weight, pelvic floor pressure assessment, and modified Oxford muscle strength assessment between MPPS and non-MPPS groups.

Feature	The non-MPPS group(n = 702)	The MPPS group (n = 543)	*Z*	*P*
Height(cm)	161(158, 165)	161(158,165)	−0.491	0.623
Weight (Kg)	60(54, 65)	59(53, 64)	−1.900	0.057
The pre-resting stage				
Average value (mmHg)	3.55(2.97, 4.53)	3.87(3.34,4.53)	−4.095	<0.001
Variable coefficient	0.15(0.11, 0.21)	0.17(0.13,0.22)	−3.302	0.001
Rapid contraction phase				
Maximum value (mmHg)	13.57(9.62, 19.11)	12.24(7.89, 18.21)	−3.713	0.001
Relaxation time(s)	0.92(0.14, 1.61)	0.54(0.42, 1.90)	−3.011	0.003
the tension contraction phase				
Average value (mmHg)	11.77(9.39,14.14)	9.84(7.44, 13.03)	−7.757	<0.001
Coefficient of variation	0.17(0.11, 0.25)	0.18(0.12, 0.26)	−2.668	0.008
Relaxation time(s)	0.82(0.09, 1.35)	0.22(0.07, 2.03)	−0.955	0.339
the endurance contraction phase				
Average value (mmHg)	9.35(7.59, 11.67)	7.41(5.41, 9.93)	−9.670	<0.001
Coefficient of variation	0.18(0.14, 0.23)	0.21(0.16, 0.26)	−5.987	<0.001
the post-resting phase				
Average value (mmHg)	3.99(2.57, 5.50)	4.23(2.82, 5.73)	−1.731	0.083
Coefficient of variation	0.19(0.17, 0.20)	0.21(0.14, 0.27)	−1.999	0.046
Deep class I muscle strength	2(1, 3)	2(1, 3)	−1.079	0.28
Superficial class I muscle strength	2(1, 3)	2(1, 3)	−1.359	0.174

Table 3
Univariate Logistic regression analysis of MPPS.

Feature index	Feature	*b*	SE	Wald $\chi 2$	*P*	OR
1	Age	−0.008	0.006	2.055	0.152	0.992
2	Height	−0.007	0.009	0.54	0.462	0.993
3	Weight	−0.01	0.007	2.455	0.117	0.99
4	BMI	−0.002	0.013	0.02	0.889	0.998
5	Number of pregnancies	0.381	0.038	101.611	<0.001	1.463
6	Number of production	0.332	0.062	28.899	<0.001	1.394
7	Number of vaginal births	0.217	0.053	16.794	<0.001	1.242
8	Number of cesarean sections	0.075	0.078	0.916	0.339	1.078
9	History of perineal laceration	0.741	0.285	6.752	0.009	2.098
10	History of hysterectomy	0.134	0.298	0.202	0.653	1.143
11	Average value during the pre-resting phase	0.089	0.046	3.788	0.052	1.093
12	Coefficient of variation during the pre-resting phase	0.493	0.329	2.245	0.134	1.638
13	Maximum value during the fast contraction phase	−0.021	0.007	7.933	0.005	0.979
14	Relaxation time during the fast contraction phase	0.09	0.035	6.719	0.01	1.094
15	Average value during the tension contraction phase	−0.081	0.015	30.956	<0.001	0.922
16	Coefficient of variation during the tension contraction phase	1.243	0.561	4.914	0.027	3.465
17	Relaxation time during the tension contraction phase	0.034	0.041	0.685	0.408	1.035
18	Average value during the endurance contraction phase	−0.137	0.018	57.535	<0.001	0.872
19	Coefficient of variation during the endurance contraction phase	2.256	0.684	10.878	0.001	9.542
20	Average value during the post-resting phase	0.043	0.027	2.575	0.109	1.044
21	Coefficient of variation during the post-resting phase	3.495	0.806	18.784	<0.001	32.958
22	Deep class I muscle strength	−0.089	0.059	2.231	0.135	0.915
23	Deep class II muscle strength	−0.024	0.054	0.205	0.651	0.976
24	Superficial class I muscle strength	−0.081	0.062	1.713	0.191	0.922
25	Superficial class II muscle strength	−0.051	0.054	0.897	0.344	0.95

maintaining a 1:1 ratio between negative and positive samples is recommended. Consequently, 543 individuals from the negative cases were randomly chosen to create a balanced dataset of 1086 individuals, combined with the positive cases. Subsequently, the dataset was split into a 7:3 ratio, with 760 cases assigned to the training set and 326 cases to the cross-validation set. The modeling process of the random forest model is depicted in Fig. 1. To assess the classification model's reliability, various metrics including the confusion matrix, accuracy, sensitivity, specificity, precision, receiver operating characteristic (ROC) curve, and area under the curve (AUC) value are analyzed.

The out-of-bag (OOB) classification error of the classification model is illustrated in Fig. 2. It is observed that the OOB error stabilizes at a value below 0.15 when the number of trees in the RF algorithm exceeds 100. This suggests that the classification model in this study exhibits strong performance.

To examine the correlation between different clinical features and the incidence rate of MPPS, the significance of each clinical feature is assessed, as depicted in Fig. 3. The feature indices and their definitions are provided in Table 3. The top five significant

Table 4
Multivariate Logistic regression analysis of MPPS.

Feature	*b*	SE	Wald $\chi 2$	*P*	OR
Number of pregnancies	0.389	0.05	61.66	<0.001	1.475
Number of production	−0.008	0.112	0.005	0.943	0.992
Number of vaginal births	−0.084	0.089	0.895	0.344	0.919
History of perineal laceration	0.42	0.328	1.645	0.2	1.523
average value during the pre-resting phase	0.298	0.062	23.219	<0.001	1.347
Coefficient of variation during the pre-resting phase	0.711	0.414	2.955	0.086	2.037
Maximum value during the fast contraction phase	0.007	0.017	0.156	0.693	1.007
Relaxation time during the fast contraction phase	0.084	0.044	3.615	0.057	1.088
Average value during the tension contraction phase	0.007	0.04	0.027	0.871	1.007
Coefficient of variation during the tension contraction phase	1.952	0.812	5.778	0.016	7.043
Average value during the endurance contraction phase	−0.186	0.035	27.709	<0.001	0.83
Coefficient of variation during the endurance contraction phase	1.593	0.811	3.86	0.049	4.919
Coefficient of variation during the post-resting phase	2.679	0.884	9.179	0.002	14.569
constant	−2.614	0.417	39.345	<0.001	0.073

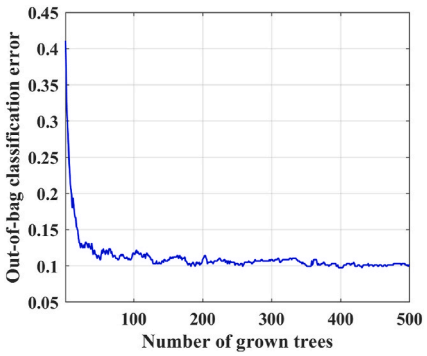

Fig. 2. The OOB classification error of the classification model.

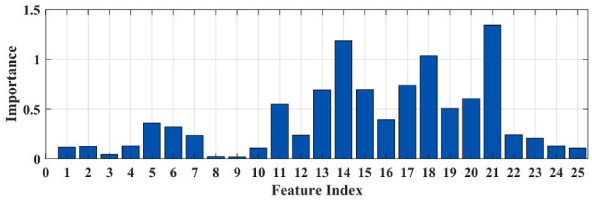

Fig. 3. Importance of the analyzed clinical features.

features, listed in descending order of importance, include the coefficient of variation during the resting phase, relaxation time during rapid contraction, average value during endurance contraction, average value during sustained contraction, and maximum value during rapid contraction.

5.3. Comparative analysis of logistic regression and random forest model

The accuracy, sensitivity, specificity and precision of the classification model are usually used as the evaluation metrics of the prediction model. The definitions of the above indicator are [10,11].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

where, TP, TN, FP, and FN are the numbers of true positive, true negative, false positive, and false negative, respectively.

The comparisons of accuracy, sensitivity, specificity, and precision between the random forest and logistic regression models are presented in Fig. 4. The accuracy, sensitivity, specificity, and precision of the random forest model are much higher than those of logistic regression model. The differences are 0.1715, 0.332, 0.1048, and 0.1494, respectively. These results indicate that the random forest model exhibits significantly superior predictive performance compared to the logistic regression model.

The ROC curves for the logistic regression and random forest models are provided in Fig. 5 (a) and (b), respectively, with their corresponding evaluation metric comparisons presented in Table 5. The logistic regression analysis yielded a sensitivity of 57.46 %, specificity of 79.63 %, and an AUC of 0.7550 for the MPPS model. In contrast, the random forest model achieved a sensitivity of 90.66 %, specificity of 90.91 %, and an AUC of 0.9426 for MPPS. Evidently, the random forest model demonstrates superior predictive performance in comparison to the logistic regression model.

The box plot illustrating the random forest and logistic regression models is displayed in Fig. 6. The median, 25th percentile, 75th percentile, maximum value, and minimum value of the random forest model are 0.482, 0.423, 0.5468, 0.726, and 0.252, respectively. Corresponding values for the logistic regression model are 0.3982, 0.164, 0.6862, 1, and 0.00889, respectively. The random forest model demonstrates a distribution closer to normal, suggesting its superiority over the logistic regression model.

6. Discussion

This study is grounded in the clinical variables of the patients enrolled in the research, in conjunction with assessments of pelvic floor pressure and data from the modified Oxford muscle strength evaluation. The study analyzes the pertinent risk factors and constructs a random forest model. This model introduces a preliminary risk assessment approach for MPPS in patients. A comparison between the performance of the random forest model and the logistic regression prediction model illustrates the superior predictive performance of the random forest model. In descending order of importance, the top five factors are identified as the coefficient of variation in the late resting phase, relaxation time in the fast contraction phase, mean value of the endurance contraction phase, mean value of the tension contraction phase, and maximum value in the fast contraction phase. Consequently, these indicators merit significant attention during the clinical screening process.

MPPS is a non-inflammatory disease that occurs in the pelvic floor muscles and fascia. It presents with local adhesions and spasms resulting from pelvic floor muscles and fascia damage, leading to persistent chronic pain [12]. The current diagnosis of MPPS continues to depend on clinical assessment, guided by existing research and clinical guidelines [13,14]. Within the clinical diagnostic process, the diagnosis of MPPS lacks specificity, and a single clinical feature is insufficient for a definitive diagnosis. A comprehensive analysis of multiple clinical factors is necessary, posing a significant challenge to the clinical experience of healthcare providers. These factors can influence treatment outcomes, patient quality of life, and even mental health implications [15,16]. While depending on experienced healthcare providers can enhance diagnostic accuracy for MPPS, the complexity of the condition and the scarcity of experienced providers render this approach impractical. Therefore, the development and implementation of reliable predictive models are crucial for timely identification of these patients and providing effective interventions to improve their prognosis [17]. Currently, machine learning models are widely used in the diagnosis and treatment of various clinical conditions [18–22], and research by Wenhui Jiang and colleagues has shown that the random forest model demonstrates good predictive performance in clinical practice [18].

Based on a dataset of 1245 cases, we implement the random forest algorithm for the first time in the screening of MPPS. In this investigation, the Logistic regression model forecasts the accuracy, sensitivity, specificity, and precision of MPPS at 69.96 %, 57.46 %, 79.63 %, and 68.57 % respectively, with an AUC of the ROC curve at 0.755. Conversely, the random forest prediction model exhibits accuracy, sensitivity, specificity, and precision rates of 87.11 %, 90.66 %, 90.91 %, and 83.51 % respectively, with an AUC of the ROC curve at 0.942. The random forest model showcases exceptional predictive performance during the initial screening of MPPS. The development of this predictive model holds significant importance in enhancing the predictive accuracy and treatment outcomes in clinical MPPS. Research findings highlight that the random forest model displays superior clinical predictive performance in identifying risk factors among MPPS patients, offering vital support and guidance for bolstering early identification and personalized treatment for MPPS. In the early stages of MPPS, treatment costs are minimal, necessitating solely rehabilitation therapy without clinical intervention. Patients in this phase typically exhibit subtle symptoms, presenting a notable diagnostic challenge for healthcare professionals. Through the predictive model proposed in this study, timely diagnosis in the early stages of MPPS becomes achievable, arresting the progression of the condition and eliminating the need for clinical intervention. This model aids in reducing economic burdens, enhancing patients' quality of life, and is pivotal in boosting the efficiency and quality of healthcare services. By enabling accurate prediction and timely intervention, it becomes feasible to significantly enhance patients' treatment outcomes and quality of life.

The factors contributing to the superior predictive performance of the random forest model over the logistic regression model are multifaceted. Firstly, the model utilize in this study encompasses a multitude of features, posing challenges for the logistic regression model in managing a problem of such complexity. In contrast, the Random Forest model remains unaffected by the abundance of features, rendering it well-suited for addressing multi-feature complexities. Secondly, the logistic regression model operates under the assumption of relatively independent features, adhering to a linear framework. However, in complex scenarios such as predicting MPPS, where features demonstrate intricate interdependencies and significant non-linearity, the predictive performance of the logistic regression model is compromised. The Random Forest model comprises numerous independent decision trees. During its construction

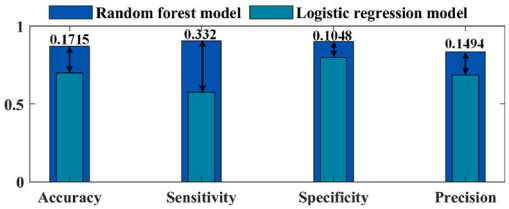

Fig. 4. Four evaluation metrics of the random forest and logistic regression models.

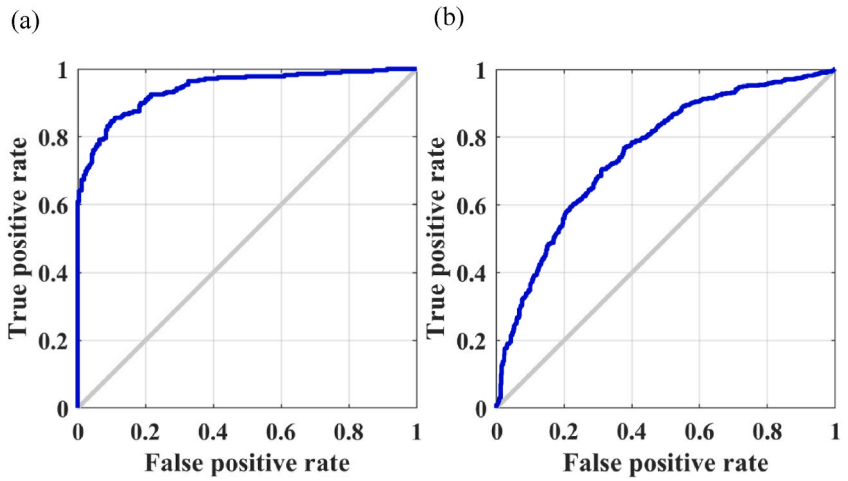

Fig. 5. ROC curve of (a) RF model (AUC = 0.9426) and (b) logistic model (AUC = 0.755).

Table 5
Evaluation metrics of the prediction models.

	Accuracy	Sensitivity	Specificity	Precision
RF model	0.8711	0.9066	0.9091	0.8351
Logistic regression model	0.6996	0.5746	0.7963	0.6857

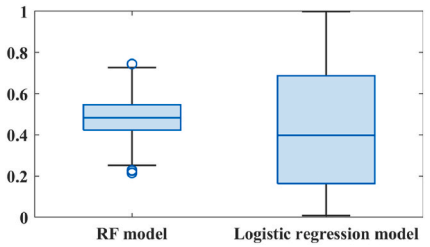

Fig. 6. Box plot of the random forest and logistic regression models.

phase, the model randomly selects subsets of samples and features. This capability enables the model to identify interactions among features, thereby improving classification accuracy and generalization abilities [23]. Finally, it is imperative to acknowledge that errors may arise during the screening process conducted by healthcare professionals, potentially introducing biases into the dataset. Such biases can significantly impact the performance of the logistic regression model. Nonetheless, as previously noted, the Random Forest model comprises numerous independent decision trees, endowing it with a degree of resilience against external influences [24]. These factors contribute to the superior predictive performance exhibited by the Random Forest model in MPPS prediction.

Although the model in this study demonstrated excellent performance, it is imperative to acknowledge its inherent limitations. The restricted sample size originating solely from a singular research center may introduce selection bias, thereby potentially compromising the model's generalizability and applicability. Future endeavors will necessitate a more extensive dataset for external validation to substantiate the model's effectiveness. Our forthcoming objective is to augment data collection efforts and diversify the array

of predictive variables to enhance the model's comprehensiveness and predictive prowess. The continuous integration of updates and additional sample data is poised to bolster the model's stability and accuracy. Furthermore, external validation of this predictive framework will be systematically conducted to fortify the reliability of our research findings. Through the ongoing processes of refinement and validation, we aim to refine the precision of MPPS risk assessment, furnishing clinicians with dependable decision-making support in clinical settings.

7. Conclusion

The random forest model has exhibited exceptional predictive performance in the initial screening evaluation of MPPS disease. The development of this predictive framework holds significant importance in refining the precision of MPPS prediction within clinical environments and elevating treatment outcomes. This research carries profound global implications, given the potentially elevated misdiagnosis rates and delayed diagnosis proportions of MPPS on a worldwide scale, coupled with a potential scarcity of seasoned healthcare providers. Moving forward, continual refinement and validation of the model will be imperative to further augment the precision of MPPS risk assessment, thereby furnishing clinicians with more dependable decision-making support in clinical practice.

Funding

Foundation of He'nan Educational Committee (24A320036).

Ethical approval statement

This study is approved by the Ethics Committee of The Third Affiliated Hospital of Zhengzhou University (Ethics Approval Number: 2023-206-01, Approval Data: August 23, 2023). The patients whose data are included in this manuscription have consented for all clinical data and other data included in the manuscript to be published.

Ethics Approval Number

2023-206-01.

Data availability statement

The authors do not have permission to share data.

CRediT authorship contribution statement

Hang Yu: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Hongguo Zhao:** Investigation, Data curation, Conceptualization. **Dongxia Liu:** Data curation, Conceptualization. **Yanhua Dong:** Formal analysis. **Manman Nai:** Validation. **Yikun Song:** Validation. **Jiaxi Liu:** Validation. **Luwen Wang:** Writing – review & editing. **Lei Li:** Funding acquisition. **Xinbin Li:** Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] V. Arora, S. Mukhopadhyay, E. Morris, Painful sex (dyspareunia): a difficult symptom in gynecological practice, Obstet. Gynaecol. Reprod. Med. 30 (9) (2020) 269–275.
[2] A. Ahangari, Prevalence of chronic pelvic pain among women: an updated review, Pain Physician 17 (2) (2014) E141–E147.
[3] J.N. Krieger, S.W.H. Lee, J. Jeon, et al., Epidemiology of prostatitis, Int. J. Antimicrob. Agents 31 (2008) 85–90.
[4] S.G.R. Klotz, G. Ketels, B. Löwe, et al., Myofascial Findings and psychopathological factors in patients with chronic pelvic pain syndrome, Pain Med. 21 (2) (2020) e34–e44.
[5] Q.W. Cao, B.G. Peng, L. Wang, et al., Expert consensus on the diagnosis and treatment of myofascial pain syndrome, World journal of clinical cases 9 (9) (2021) 2077.
[6] J. Patel, S. Javed, Myofascial pain syndrome and SARS-CoV-2: a case series, Pain Manag. 12 (3) (2021) 255–260.
[7] L.T. Akhmedzhanova, A.N. Barinov, M.S. Leontyeva, et al., Diagnosis and treatment of chronic pelvic pain syndrome, Neurology, Neuropsychiatry, Psychosomatics 14 (4) (2022) 54–61.
[8] A. Shrikhande, C. Ullger, K. Seko, et al., A physiatrist's understanding and application of the current literature on chronic pelvic pain: a narrative review, Pain Reports 6 (3) (2021) e949.
[9] C.H.J. Ferreira, P.B. Barbosa, F. de Oliveira Souza, et al., Inter-rater reliability study of the modified Oxford grading scale and the peritron manometer, Physiotherapy 97 (2) (2011) 132–138.
[10] L. Yang, H. Wu, X. Jin, et al., Study of cardiovascular disease prediction model based on random forest in eastern China, Sci. Rep. 10 (1) (2020) 5245.
[11] C. Iwendi, A.K. Bashir, A. Peshkar, et al., COVID-19 patient health prediction using boosted random forest algorithm, Front. Public Health 8 (2020) 357.

[12] D.S. Engeler, A.P. Baranowski, P. Dinis-Oliveira, et al., The 2013 EAU guidelines on chronic pelvic pain: is management of chronic pelvic pain a habit, a philosophy, or a science? 10 years of development, Eur. Urol. 64 (3) (2013) 431–439.
[13] Q.W. Cao, B.G. Peng, L. Wang, et al., Expert consensus on the diagnosis and treatment of myofascial pain syndrome, World journal of clinical cases 9 (9) (2021) 2077.
[14] J. Patel, S. Javed, Myofascial pain syndrome and SARS-CoV-2: a case series, Pain Manag. 12 (3) (2021) 255–260.
[15] F. Itza, D. Zarza, L. Serra, et al., Myofascial pain syndrome in the pelvic floor: a common urological condition, Actas Urol. Esp. 34 (4) (2010) 318–326.
[16] K. Grinberg, I. Weissman-Fogel, L. Lowenstein, et al., How does myofascial physical therapy attenuate pain in chronic pelvic pain syndrome? Pain Res. Manag. (2019) 2019.
[17] S.C. Kapurubandara, B. Lowes, U.M. Sansom-Daly, et al., A systematic review of diagnostic tests to detect pelvic floor myofascial pain, International urogynecology journal 33 (9) (2022) 2379–2389.
[18] Y. Wu, B. Xin, Q. Wan, et al., Risk factors and prediction models for cardiovascular complications of hypertension in older adults with machine learning: a cross-sectional study, Heliyon 10 (6) (2024) e27941.
[19] Q. Wang, J. Sun, X. Liu, et al., Comparison of risk prediction models for the progression of pelvic inflammatory disease patients to sepsis: cox regression model and machine learning model, Heliyon 10 (1) (2024) e23148.
[20] S.A. Suha, M.N. Islam, Exploring the dominant features and data-driven detection of polycystic ovary syndrome through modified stacking ensemble machine learning technique, Heliyon 9 (3) (2023) e14518.
[21] S. Bendifallah, A. Puchar, S. Suisse, et al., Machine learning algorithms as new screening approach for patients with endometriosis, Sci. Rep. 12 (1) (2022) 639.
[22] N. Al Mudawi, A. Alazeb, A model for predicting cervical cancer using machine learning algorithms, Sensors 22 (11) (2022) 4132.
[23] A. Parmar, R. Katariya, V. Patel, A Review on Random Forest: an Ensemble classifier//International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018, Springer International Publishing, 2019, pp. 758–763.
[24] M. Schonlau, R.Y. Zou, The random forest algorithm for statistical learning, STATA J. 20 (1) (2020) 3–29.