scientific reports

OPEN

The construction of HMME-PDT efficacy prediction model for port-wine stain based on machine learning algorithms

Hongxia Yan[1,2], Yixin Tan[1,2], Fan Qiao[1,2,3], Zhuotong Zeng[1,2], Yaqian Shi[1,2], Xueqin Zhang[1,2], Lu Li[1,2], Ting Zeng[1,2], Yi Zhan[1,2], Ruixuan You[1,2], Xinglan He[1,2], Rong Xiao[1,2✉] & Xiangning Qiu[1,2✉]

Hematoporphyrin monomethyl ether-photodynamic therapy (HMME-PDT) is a safe and effective treatment for port-wine stain (PWS). Comprehensive methods for predicting HMME-PDT efficacy based on clinical factors are lacking. This study aims to develop and validate two machine learning models to predict the therapeutic effect of HMME-PDT for PWS. We conducted a retrospective study of 131 facial PWS patients treated with single HMME-PDT at the Second Xiangya Hospital from May 2022 to January 2025. The patients were divided into the training cohort and the validation cohort based on the order of their enrollment. Key clinical features were selected using recursive feature elimination (RFE). We developed and validated prediction models with Extreme Gradient Boosting (XGBoost) and Random Forest (RF) algorithms. Model performance was assessed using confusion matrix and evaluation metrics. RFE identified the top predictive factors: dermoscopy vascular pattern, immediate fluorescence intensity (IFI) after HMME-PDT, the facial port-wine stain area and severity index score, and age. In the training cohort, both models demonstrated strong predictive performance, with accuracies, F1 scores, and AUC values exceeding 0.8. The XGBoost model outperformed with an accuracy of 0.8750, F1 score of 0.8750, and AUC of 0.8636. In the validation cohort, XGBoost model achieved an accuracy and F1 score both greater than 0.73, with an AUC value of 0.7672. It had the better comprehensive performance. Our findings suggest these models are promising for predicting HMME-PDT efficacy in PWS. This is the first study to explore IFI after HMME-PDT in efficacy assessment.

Keywords Port-wine stain, HMME-PDT, Machine learning algorithms, Efficacy prediction model, Immediate fluorescence intensity at the lesion site after HMME-PDT

Port-wine stain (PWS), also known as nevus flammeus, is a benign congenital vascular malformation characterized by the dilation and malformation of superficial dermal capillaries and post-capillary venules. The incidence of PWS in newborns is approximately 0.3–0.5%[1]. Further, PWS frequently occurs in exposed areas, such as the face and neck, and generally does not regress spontaneously. If not treated promptly or appropriately, the lesions may enlarge, thicken, and darken as the patient ages, potentially leading to spontaneous bleeding[2]. PWS is a disfiguring condition that can lead to psychosocial issues and reduce the overall quality of life of patients. Furthermore, the pathogenesis of PWS remains unclear[3] making it difficult to prevent the condition at its root cause. Current research focuses primarily on optimizing treatment strategies.

In recent years, photodynamic therapy (PDT) has been rapidly developed and shown to be an effective treatment for PWS, which can induce photochemical reactions leading to cell death by apoptosis, autophagy, or necrosis[4]. Hematoporphyrin monomethyl ether (HMME), also called Hemoporfin, is a novel second-generation porphyrin photosensitizer. It offers advantages, such as safety, minimal side effects, non-invasiveness, and high selectivity for abnormal blood vessels. Hematoporphyrin monomethyl ether-PDT (HMME-PDT) has been approved for the treatment of PWS in China since 2017[5]. The efficacy of HMME-PDT has been confirmed in

[1]Department of Dermatology, The Second Xiangya Hospital, Central South University, Changsha 410011, Hunan, China. [2]Institution of Dermatology, Central South University, Changsha 410011, Hunan, China. [3]Department of Dermatology, The Second Affiliated Hospital, Henan University of Science and Technology, Luoyang 471000, Henan, China. ✉email: xiaorong65@csu.edu.cn; xiangningqiu@csu.edu.cn

both children and adults by years of clinical observation[6]. It is worth noting that the application of HMME-PDT is not limited to the treatment of PWS. Studies have shown that it also demonstrates significant efficacy in combating fungal infections and drug-resistant bacterial infections[7,8]. Due to the heterogeneity of vascular malformations and individual differences among patients, there are significant variations in the efficacy of HMME-PDT among different patients[9].

Many non-invasive skin examination techniques are gradually being applied to predict, assesss and intraoperatively monitor treatment outcomes in PWS. These include approaches such as dermoscopy, reflectance confocal microscopy, and high-frequency ultrasound. Research has shown that different vascular patterns identified using dermatoscopy can predict the efficacy of HMME-PDT[10]. Real-time fluorescence spectroscopy can monitor the skin photosensitizer concentration during PDT for PWS and predict treatment responses[11]. However, there is still a lack of comprehensive methods to predict PDT efficacy based on various clinical factors in PWS patients, which is far from meeting clinical needs. Therefore, there is an urgent need for precise and comprehensive new methods to predict patient outcomes, which can assist clinicians in treatment decisions and achieve accurate and personalized therapy.

Machine learning has played a crucial role in medical research and clinical practice, given the advancement and development of big data technology. Machine learning techniques can autonomously learn the identification of hidden relationships in existing data and use this knowledge to predict outcomes for unknown data. Further, machine learning has garnered global attention due to its efficient data processing, deep analytical capabilities, and self-optimization attributes. Currently, machine learning has a unique value in several fields, such as basic medical research, clinical practice, and epidemiological studies[12–14]. For example, Mungeret al. used the random forest machine learning algorithm in an observational cohort study to identify the main predictors of non-calcified coronary plaque burden in psoriasis patients[15]. In another report, Maintz et al. employed a machine learning-gradient boosting approach to analyze the factors associated with the severity of atopic dermatitis[16]. Zheng et al. were the first to construct risk prediction model for postherpetic itch using machine learning methods, including logistic regression, random forest, k-nearest neighbor, gradient boosting decision tree and neural network[17]. Furthermore, machine learning algorithms use non-parametric models that do not rely on assumptions about data distribution in contrast to traditional statistical methods. These algorithms can flexibly adapt to the nonlinearity and complexity of the data and automatically adjust without the need for human intervention. They can efficiently and objectively reveal the complex and hidden relationships between clinical characteristics and patient outcomes, as well as the relationships among various features[18].

In summary, this study collected data on clinical characteristics of HMME-PDT of PWS patients, including the immediate fluorescence intensity (IFI) at the lesion site after HMME-PDT, dermoscopy vascular pattern, and the facial port-wine stain area and severity index (FSASI) score[19]. Feature selection was performed using the Recursive Feature Elimination (RFE) method, and efficacy prediction models for HMME-PDT were constructed by applying Extreme Gradient Boosting (XGBoost) and Random Forest (RF) algorithms. These models were comprehensively evaluated to determine the superior model. The relationships between various clinical features and the efficacy of HMME-PDT were explored by leveraging these models. The work aimed to establish a foundation for timely and accurate assessment and adjustment of HMME-PDT treatment plans.

Materials and methods
Research design and objectives

This study retrospectively analyzed clinically diagnosed facial PWS patients ($n = 131$) who received a single HMME-PDT session at the Second Xiangya Hospital of Central South University from May 2022 to January 2025. Patients enrolled between May 2022 and April 2024 were assigned to the training cohort ($n = 78$), while patients enrolled between May 2024 and January 2025 were assigned to the validation cohort ($n = 53$). Exclusion criteria included: (1) presence of other vascular malformations; (2) without clear clinical images for efficacy evaluation. This study was approved by the clinical research ethics committee of Second Xiangya Hospital of Central South University (LYEC2024-K0152). The research was performed in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from all patients before treatment. The patients mentioned in this manuscript have given written informed consent to the release of images that may lead to identification. Specially, minor patients was obtained consent from their parents and legal guardians.

Treatment protocol
Pre-treatment preparation

All patients completed blood count, liver and kidney function test, and ECG examination to evaluate treatment risks. An experienced operator captured lesion images using VISIA-CR™ system (Canfield Scientific, Inc., United States) under consistent angle and light intensity. Dermoscope images of lesion were detected by the hand-held dermoscope with 50X polarized light (Guangzhou Chuanghong Medical Technology Co. Ltd., China) in the same area. Each patient was assessed for the subtype of lesion and the FSASI score.

HMME-PDT treatment

Hemoporfin solution (Shanghai Fudan Zhangjiang Biopharmaceutical Co. Ltd., China) was prepared at a dosage of 5 mg/kg. The patient was made to lie down, the hair at the treatment site was shaved, and the lesions were fully exposed and cleaned. Normal skin within the irradiation area was covered with a light-blocking cloth while the operator wore protective goggles. The total administration time was 20 min for adults (speed of the infusion pump speed was 150 ml/h) and 5 min for children weighing ≤ 40 kg (speed of the infusion pump speed was 240 ml/h). Illumination was started at the 8th minute of drug infusion for adults and the 3rd minute for children. The lesion was irradiated using 532 nm LED green light (Wuhan YaGe Photoelectric Technology Co. Ltd., China) for 15–20 min for each spot. An auxiliary treatment head was used simultaneously for larger or

more dispersed lesions. After treatment, a routine cooling spray or cold compress was applied for 30 min, and patients were given postoperative care instructions.

Post-treatment and follow-up

Immediately after HMME-PDT, the lesion was irradiated to induce fluorescence response using UVA (385 ± 10 nm) as the excitation light source. Fluorescence images were recorded with an Apple iPhone XS mobile phone. After the single HMME-PDT, the patient underwent follow-up at the outpatient department. The same VISIA-CR™ system was utilized to capture lesion images under constant conditions.

Data collection

The clinical baseline data of each patient was collected, including age; gender; the subtype of the lesion; history of previous pulsed dye laser (PDL) treatment; pre-treatment FSASI[19] score; dermatoscopic images before treatment; immediate fluorescence images after HMME-PDT; clinical images before and after HMME-PDT; and the interval between two treatment. The efficacy assessment followed the traditional classification method. Depending on the degree of regression in treated lesions, the clinical efficacy was categorized into two classes: effective groups ($\geq 25\%$ improvement) and ineffective groups ($< 25\%$ improvement). The post-PDT fluorescence was qualitatively classified into three levels. It was defined as weak when little fluorescence was observed, and intense when homogeneous, bright fluorescence was observed (Fig. 1). The intermediate fluorescence fell in between these two levels. We focused on three main dermoscopic patterns, as follows the superficial type included point globular and short clubbed blood vessels; the deep type included mainly reticular and linear blood vessels (Fig. 1). If the images included both these vascular patterns, it was defined as a mixed type. All these mentioned evaluation techniques were independently assessed and recorded by two dermatologists with extensive clinical experience.
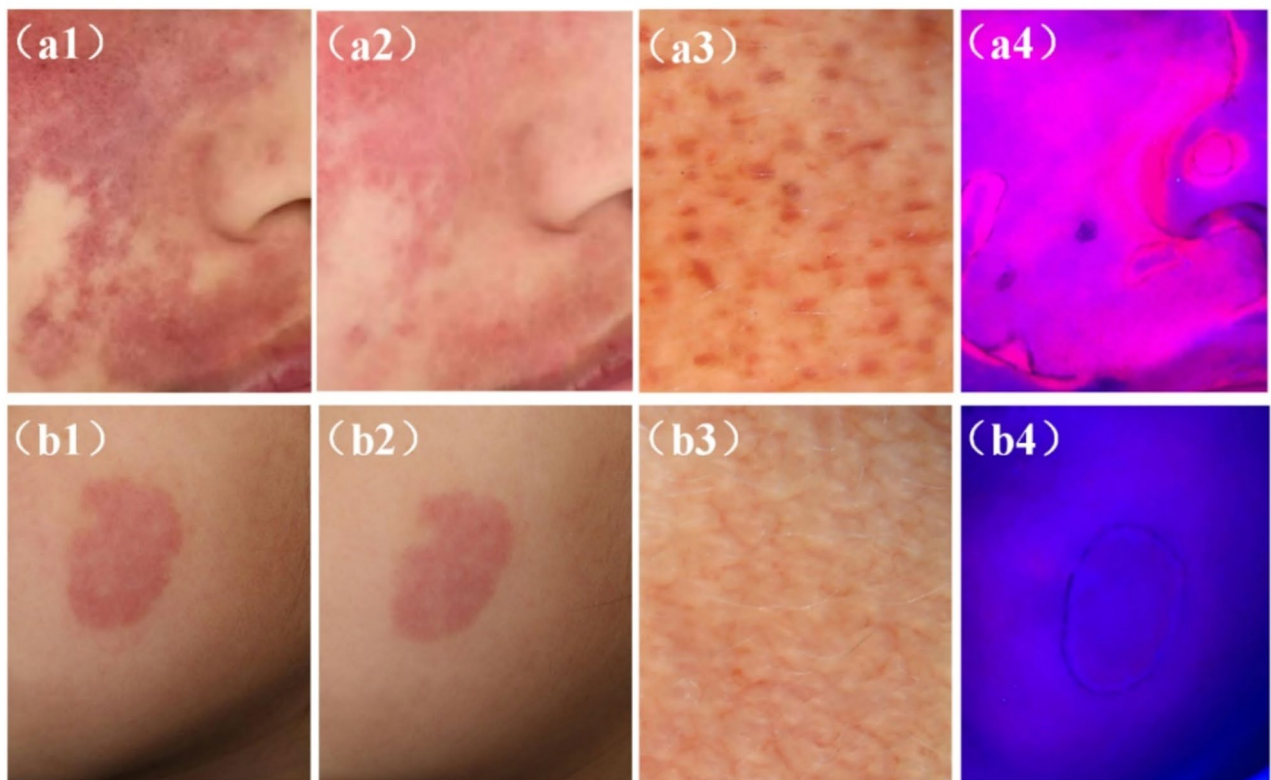
Fig. 1. Clinical pictures, dermoscopy and fluorescence imaging of port-wine stain (PWS)lesions before and after one session of hematoporphyrin monomethyl ether-photodynamic therapy (HMME-PDT). (**a**) A PWS patient that showed $\geq 25\%$ improvement (effective). (**a1**) Before treatment and (**a2**) after HMME-PDT. (**a3**) Superficial type: dotted and globular, and short clubbed vessels were seen under dermoscopy before treatment. (**a4**) Immediate fluorescence intensity (IFI) at the lesion site after HMME-PDT where bright and homogeneous fluorescence was detected in the lesion. (**b**) A PWS patient that showed $< 25\%$ improvement (ineffective). (**b1**) Before treatment and (**b2**) after HMME-PDT. (**b3**) Deep type: reticular and linear vessels were seen under dermoscopy before treatment. (**b4**) Immediate fluorescence intensity at the lesion site after HMME-PDT where little fluorescence was detected in the lesion.

Machine learning development process

Data preprocessing

This study enrolled a dataset of 78 patients with PWS, including 9 variables (6 categorical and 3 continuous). The categorical variables were gender, subtype of lesion, PDL treatment history, IFI after HMME-PDT, dermoscopy vascular pattern, and the therapeutic effect. The continuous variables were age (integer), the interval between two treatment sessions (integer), and the FSASI score (decimal number). Depending on the presence of sequential relationships and the number of categories, one-hot encoding was used for unordered categorical variables, while ordinal encoding was applied to ordered categorical variables. During the data processing phase, missing values are detrimental to the generalization of predictive models and can reduce their practical utility to some extent. Therefore, it is essential to conduct appropriate data cleaning to address this issue[20,21]. In this study, data cleaning was necessary due to a small number of missing values in the dermatoscopy data. The data were categorical variables with values missing completely at random. The primary method used for data cleaning was mode imputation.

Algorithm training and validation

After data preprocessing, we used RFE with an XGBoost model as the estimator to iteratively remove the least important features[22]identifying the most significant and clinically relevant predictive factors. Based on the RFE feature selection results, the dataset of 78 PWS patients was randomly divided into training and testing sets with an 8:2 ratio using a random seed. This study applied a Bayesian optimization algorithm to select the optimal hyperparameters. We utilized the Gaussian Process as the surrogate model and the Expected Improvement (EI) as the acquisition function, with negative accuracy as the objective function return value. The hyperparameter search space for the XGBoost and RF algorithms was defined separately (Table 3). The iteration count was set to 100, and the termination criterion was reaching the maximum number of iterations (Supplementary Fig. 2–4). The synthetic minority over-sampling technique (SMOTE) method was employed (setting strategy = auto、k_ neighbors = 3) for oversampling to balance the sample size disparity between groups. The prediction models were developed using XGBoost and RF methods, respectively. We comprehensively evaluated model performance and selected the optimal model by plotting the confusion matrix, receiver operating characteristic curve (ROC), and evaluation measures, such as Accuracy, Precision, Recall, F1-Score, and the area under the curve (AUC). The validation cohort was employed to assess the performance of both models using the same ways. And we interpreted the impact of each feature in the model using the Feature Importance and SHapley Additive exPlanations (SHAP) method by applying Tree Explainer[23].

Statistical analysis

The baseline characteristics of the patients were compared between the different groups (Table 1, Supplementary Table 1). Categorical data were compared using Pearson's chi-square or likelihood ratio chi-square tests, with results presented as frequency and percentage. Continuous variables were compared using the Mann-Whitney U test, and the results were expressed as M (Q1, Q3). The data analysis was conducted using SPSS 26.0 software. Feature selection, model training, validation, evaluation, and interpretation analyses were performed using Python 3.9.13 software. The primary data analysis packages included xgboost 1.7.6, scikit-learn 1.1.3, and shap 0.44.0. The model interpretation also utilized the built-in Feature Importance provided by XGBoost algorithms. A p-value<0.05 was deemed statistically significant.

Results

Patient characteristics

This study enrolled 131 facial PWS patients, with 78 assigned to the training cohort and 53 to the validation cohort. The baseline characteristics in the training cohort are presented in Table 1. There were 36 male and 42 female patients, with the age ranging between 2 and 47 years. The treatment of training cohort was ineffective

Classifier models	Hyperparameters	Optimal value	Search space
Extreme gradient boosting	learning rate	0.351	0.01,0.5
	n estimators	401	10,500
	gamma	0.315	0.01,1
	reg alpha	0.18	0.01,1
	reg lambda	0.184	0.01,1
Random forest	n estimators	677	1, 1000
	min samples split	2	2,10
	min samples leaf	5	1,5
	min weight fraction leaf	0.0511	0,0.5
	min impurity decrease	0.1428	0,1
	max samples	0.5451	0.1,1
	max depth	97	1,100

Table 3. Bayesian optimization search space and optimal values for hyperparameters in extreme gradient boosting (XGBoost) and random forest (RF) efficacy prediction models.

Characteristics	Total ($n = 78$)	Ineffective ($n = 21$)	Effective ($n = 57$)	p value
Gender, n (%)				0.503
Male	36(46.154)	11(52.381)	25(43.860)	
Female	42(53.846)	10(47.619)	32(56.140)	
Age (years)	10(5, 21)	11(5.5, 19)	10(5, 21)	0.731
Interval (months)	4(3, 6)	5(3.5, 7)	4(3, 6)	0.13
Treatment history, n (%)				0.952
Yes	45(57.692)	12(57.143)	33(57.895)	
No	33(42.308)	9(42.857)	24(42.105)	
Subtype of lesion, n (%)				0.25
Pink	27(34.616)	9(42.857)	18(31.579)	
Purple	45(57.692)	12(57.143)	33(57.895)	
Thick	6(7.692)	0(0)	6(10.526)	
Dermoscopy vascular pattern, n (%)				<0.001
Superficial	47(60.257)	6(28.571)	41(71.930)	
Mixed	6(7.692)	1(4.762)	5(8.772)	
Deep	25(32.051)	14(66.667)	11(19.298)	
Fluorescence intensity, n (%)				0.003
Weak	15(19.231)	9(42.857)	6(10.527)	
Intermediate	26(33.333)	7(33.333)	19(33.333)	
Intense	37(47.436)	5(23.810)	32(56.140)	
FSASI (scores)	2.95(1.58, 5.85)	1.8(1.2, 2.95)	3.6(2.3, 7.2)	0.001

Table 1. Patients' baseline clinical characteristics analysis in the training cohort. *P* value < 0.05 indicates statistical significance.

in 21 patients and effective in 57 patients. There were no statistically significant differences between the two groups regarding gender, lesion subtype, PDL treatment history, age, and the interval between treatment sessions ($p > 0.05$). In contrast, significant differences were observed between the groups in terms of the dermoscopy vascular pattern, immediate fluorescence intensity following HMME-PDT, and the FSASI score ($p < 0.05$). This suggests that these factors may have an impact on the treatment efficacy.

Feature selection
We performed feature selection using XGBoost-RFE to obtain meaningful features. Further optimization of the feature combination was achieved using 3-fold cross-validation. The accuracy of the RFE model using XGBoost as the estimator peaked when retaining 4 or 5 features (Fig. 2, Table 2). Ultimately, the algorithm opted for a concise subset comprising 4 features. At this stage, the important features were dermoscopy vascular pattern, IFI after HMME-PDT, the FSASI score, and age.

Comparing machine learning performance
The Gaussian process was chosen as the surrogate model using the four important features previously selected. The acquisition function for Bayesian optimization within the search space described above was EI, and the optimal hyperparameter values were obtained (Table 3, Supplementary Fig. 2–4). The HMME-PDT efficacy prediction models were developed on the training set using XGBoost and RF algorithms under the optimal hyperparameter combination. After fitting the models on the test set, their performance was evaluated (Table 4). The overall accuracy of the XGBoost prediction model was 0.875. The weighted average precision, recall, and F1 score were more than 0.85, demonstrating its excellent fitting performance. The overall accuracy of the RF prediction model was 0.8125, with weighted average precision, recall, and the F1 score exceeding 0.80. These results were also considered excellent. We evaluated the accuracy of the prediction models for treatment efficacy using ROC curve analysis. The ROC curves of both models were significantly above the baseline of random classification, represented by the 45-degree diagonal line. The AUC values were 0.8638 for the XGBoost model and 0.8818 for the RF model (Fig. 3). In summary, both models exhibited an excellent predictive performance. However, the overall performance of the XGBoost model was better. Given its higher F1 score and accuracy, the XGBoost model had a better and more accurate performance predicting both effective and ineffective outcomes.

Explanatory analysis of the prediction model
For the global interpretation, the feature importance scores of each feature ranked from highest to lowest in the XGBoost model were as follows: IFI after HMME-PDT, dermoscopy vascular pattern, FSASI score, and age (Fig. 4). We used SHAP to quantify the contribution of each feature to the prediction results. Based on their SHAP absolute values, the features ranked in descending order include dermoscopy vascular pattern, IFI after HMME-PDT, FSASI score, and age (Figs. 5 and 6). Additionally, we utilized SHAP for local interpretation of the data. The SHAP distributions between groups exhibited significant imbalance for categorical variables. Superficial type under dermoscopy and intense fluorescence intensity both had SHAP values > 0, indicting
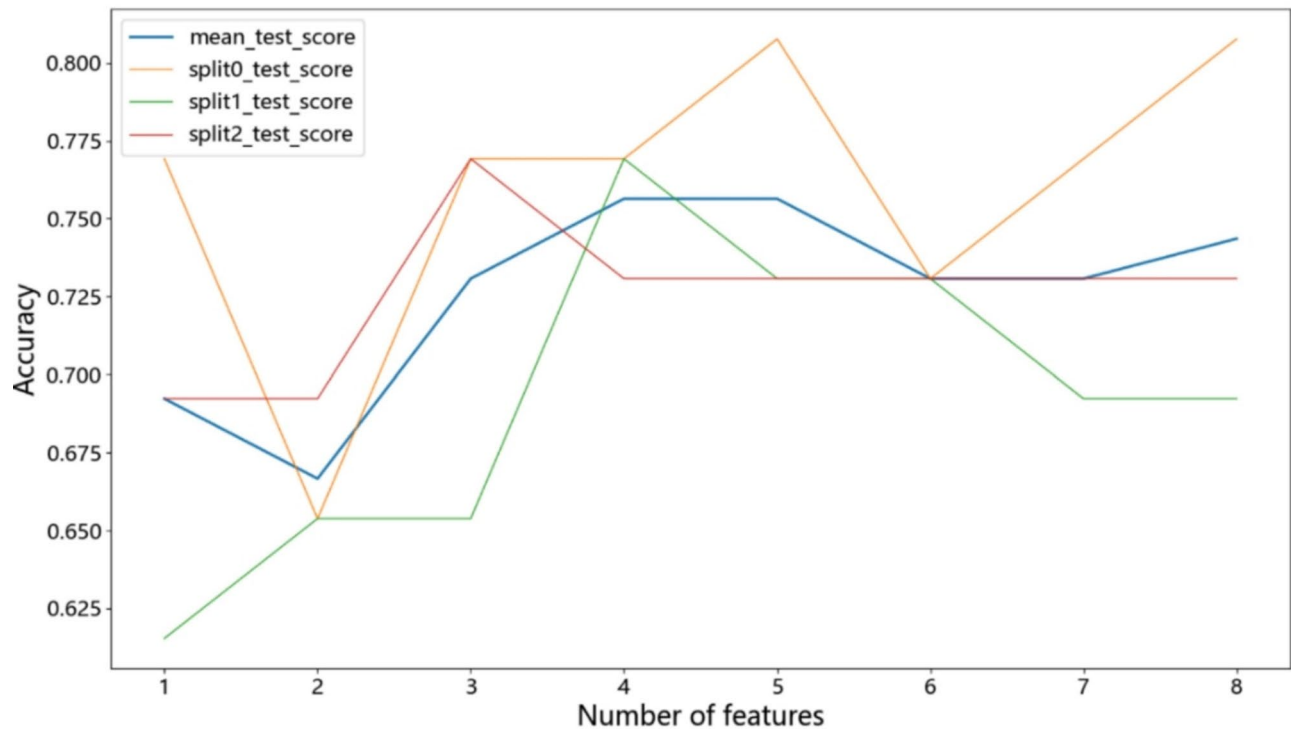
Fig. 2. Flowchart of the recursive feature elimination (RFE) with an Extreme Gradient Boosting (XGBoost) model as the estimator (XGBoost-RFE). The four curves in the figure represent the performance of the XGBoost-RFE model in 3-fold cross-validation and its mean performance. Based on the blue curve, the accuracy peaked when 4 or 5 features were retained.

Number of features	Mean test score	Std test score	Split0 test score	Split1 test score	Split2 test score
1	0.692	0.063	0.769	0.615	0.692
2	0.667	0.018	0.654	0.654	0.692
3	0.731	0.054	0.769	0.654	0.769
4	0.756	0.018	0.769	0.769	0.731
5	0.756	0.036	0.808	0.731	0.731
6	0.731	0.000	0.731	0.731	0.731
7	0.731	0.031	0.769	0.692	0.731
8	0.744	0.048	0.808	0.692	0.731

Table 2. Performance evaluation results of the XGBoost-RFE model with varying numbers of retained features.

Models	Testing dataset		Accuracy	Precision	Recall	F1-score	AUC
	Predict ineffective	Predict effective					
Extreme Gradient Boosting (XGBoost)							
ineffective	4	1	0.8750	0.8750	0.8750	0.8750	0.8636
effective	1	10					
Random Forest (RF)							
ineffective	4	1	0.8125	0.8271	0.8125	0.8166	0.8818
effective	2	9					

Table 4. Comparison of the performance of XGBoost and RF models by confusion matrix and evaluation measures. AUC the area under the curve.
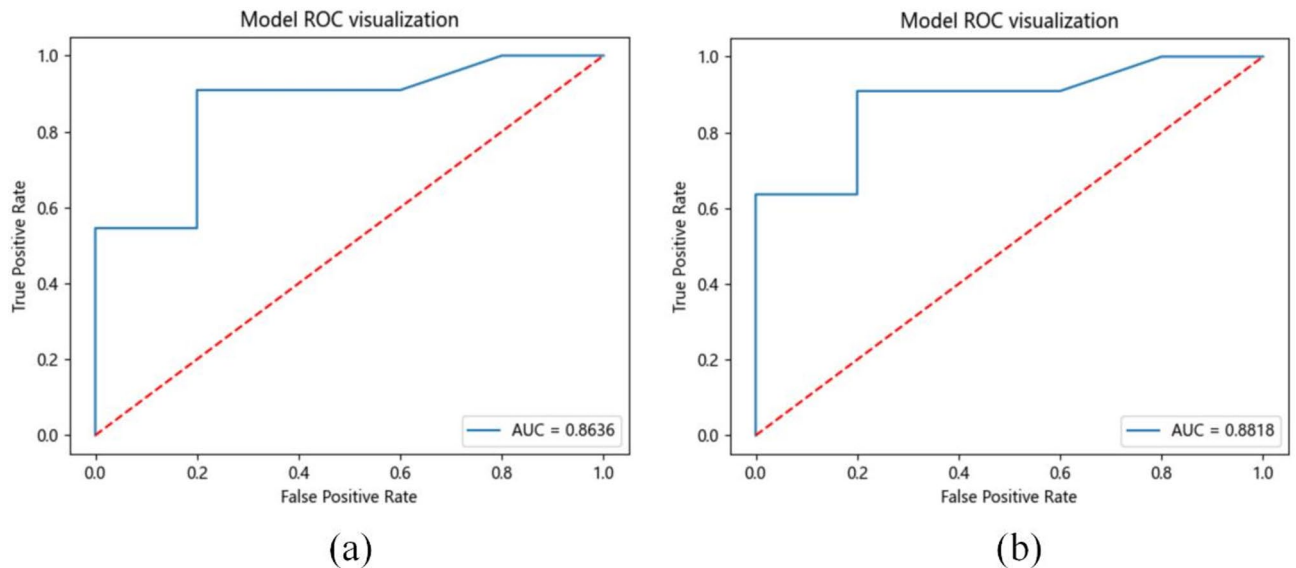
Fig. 3. Receiver operating characteristic (ROC) curve of (**a**) XGBoost and (**b**) Random Forest (RF) on the testing dataset.

correlation with better efficacy (Fig. 7a-b). There was a clear trend in the distribution of the FSASI score and age data points (Fig. 7c-d). The SHAP values for the FSASI score increased with feature statistics and reached a peak, demonstrating a positive correlation with efficacy. And age was negatively correlated with efficacy.

Model validation

A comparison of baseline characteristics between the training cohort ($n = 78$) and validation cohort ($n = 53$) is presented in Supplementary Table 1. No statistically significant differences were observed between the two groups in terms of gender, age, history treatment of PDL, lesion type, IFI, and FSASI scores ($p > 0.05$). In the validation cohort, the AUC values were 0.7672 for the XGBoost model and 0.7557 for the RF model (Supplementary Fig. 1). The results show the comparison of F1 score, accuracy, precision, and recall among the two models (Supplementary Table 2). The XGBoost model achieved values greater than 0.73 for all of the above metrics and it had the better comprehensive performance.

Discussion

PWS is a congenital disfiguring condition with an incidence rate of approximately 0.3–0.5%. Although PDL is the gold standard clinical treatment for PWS, only 10-20% of patients achieve near-complete lesion resolution after multiple treatments, with a high risk of recurrence[24]. Therefore, efficient treatment strategies are needed to reduce or eliminate skin lesions in PWS patients. HMME-PDT was introduced into clinical practice in 2017 as a novel technology. Due to its superior efficacy and safety, HMME-PDT is increasingly used in the treatment of PWS in China. It is noteworthy that the efficacy of HMME-PDT is influenced by various factors, such as the type of PWS, dermoscopic pattern, lesion area, site and thickness, patient's age at treatment, and previous history of PDL treatment. Numerous studies have evaluated and predicted the efficacy of HMME-PDT in treating PWS. However, a single, precise indicator for accurately predicting HMME-PDT efficacy in real-world settings still needs to be identified. This limitation makes it difficult for clinicians to achieve precise and personalized treatment based on patient characteristics.

Machine learning has demonstrated extensive potential applications in clinical research and medical practice by leveraging its advantages in accuracy, large-scale data processing, and self-learning optimization. It has improved diagnostic accuracy and extended the comprehensive management of complex diseases, meeting various medical needs. In this study, we developed machine learning-based prediction models for predicting the efficacy of HMME-PDT in PWS patients. This lays the foundation for developing an artificial intelligence system that can be widely applied in clinical practice, offering a new approach for early efficacy assessment and timely intervention.

XGBoost is an efficient supervised learning algorithm suitable for regression, classification, and ranking problems. It is based on the principle of Gradient Boosting Trees, continuously adding decision trees to correct prediction errors and approximate the true values. The algorithm uses parallel computing and optimization techniques to reduce training costs and automatically handle missing values, enhancing the model's predictive capability[25,26]. The other algorithm RF is an ensemble learning method that constructs multiple decision trees and combines their prediction results to augment the model's performance[27,28]. Although both XGBoost and RF are decision tree-based machine learning algorithms, they differ significantly in their computational processes. The RF algorithm uses the Bootstrap Aggregating method to perform random sampling and independently train decision trees in parallel. In contrast, XGBoost builds each decision tree based on the performance of the previous tree, creating a serial relationship between the trees. Furthermore, the XGBoost algorithm has
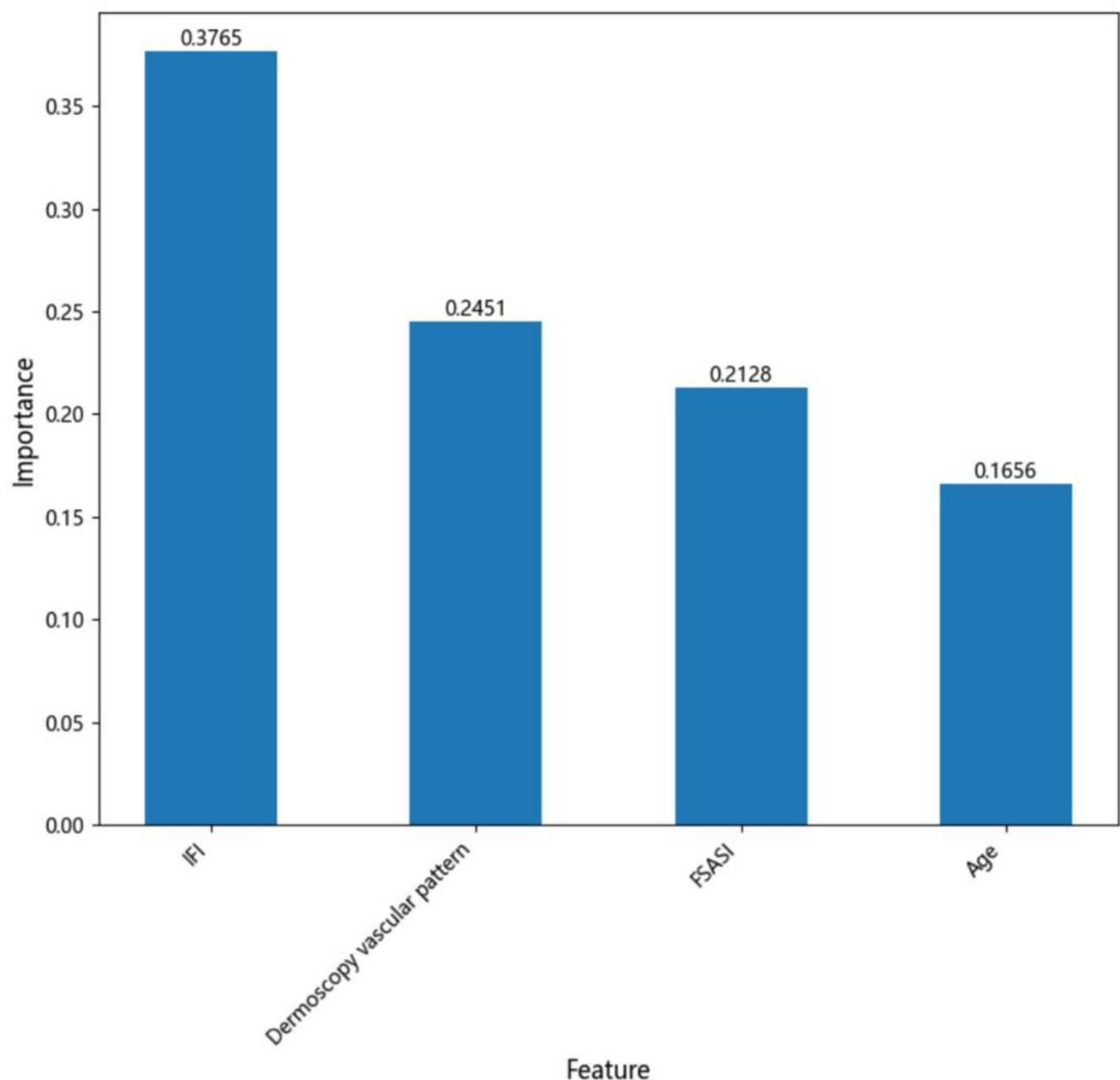
**Fig. 4**. Importance of each feature during the training process of the XGBoost model.

low training costs, and fast computation speed and also offers a rich set of hyperparameters for optimizing performance. Although RF is a powerful ensemble learning method, its performance can be affected by certain factors. If there are fewer samples of certain classes in the dataset, RF may have difficulty adequately learning the features of the few classes. XGBoost can better handle the sample imbalance problem by weighting the loss function. In addition, XGBoost is able to automatically capture the non-linear interactions between features, while RF is weaker in this aspect, which may lead to a decrease in prediction performance[29,30]. These were also demonstrated in the present study.

This study primarily involved data preprocessing, feature selection, algorithm training and testing, and performance evaluation steps. Further validation was conducted using patient data from different time periods. After data preprocessing, the XGBoost-RFE method identified the dermoscopy vascular pattern classification, IFI after HMME-PDT, FSASI score, and age as predictive factors for HMME-PDT treatment efficacy in PWS. Two prediction models for assessing PWS therapeutic effect were developed using XGBoost and RF machine learning methods. Model performance was compared, with the XGBoost model showing a slightly lower AUC value but higher accuracy and F1 score compared to the RF model. Additionally, the XGBoost model demonstrated a more balanced performance between effective and ineffective groups. In the validation cohort, the XGBoost model outperformed the RF model, achieving higher values for all of the above metrics.
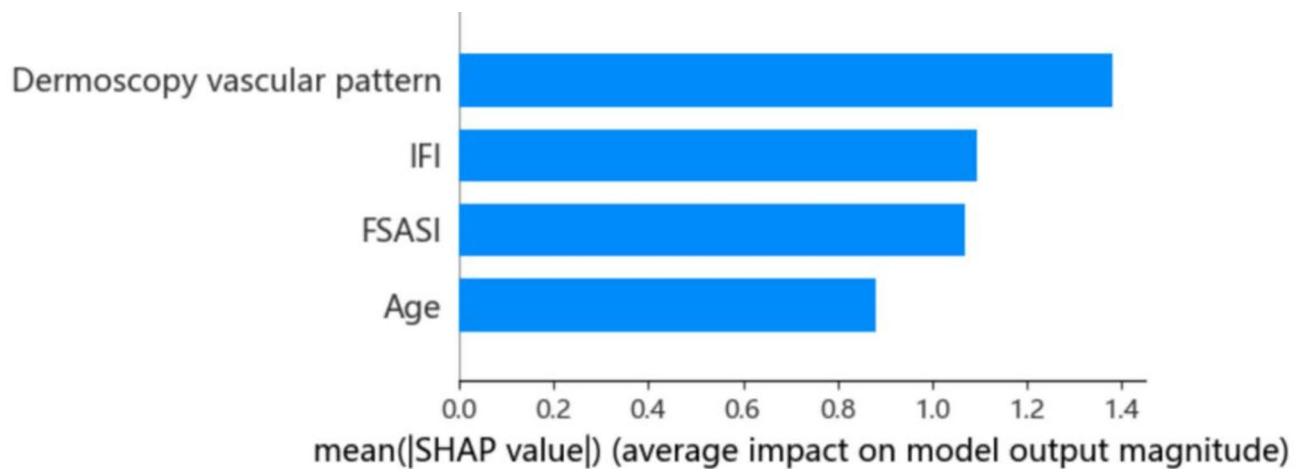
Fig. 5. Ranking of the average absolute SHapley Additive exPlanations (SHAP) values for each feature. The x-axis represents the absolute value of the average SHAP values for each feature, indicating the degree of its influence on the prediction results. Higher SHAP values indicate a greater impact of the feature on predicting a specific outcome.
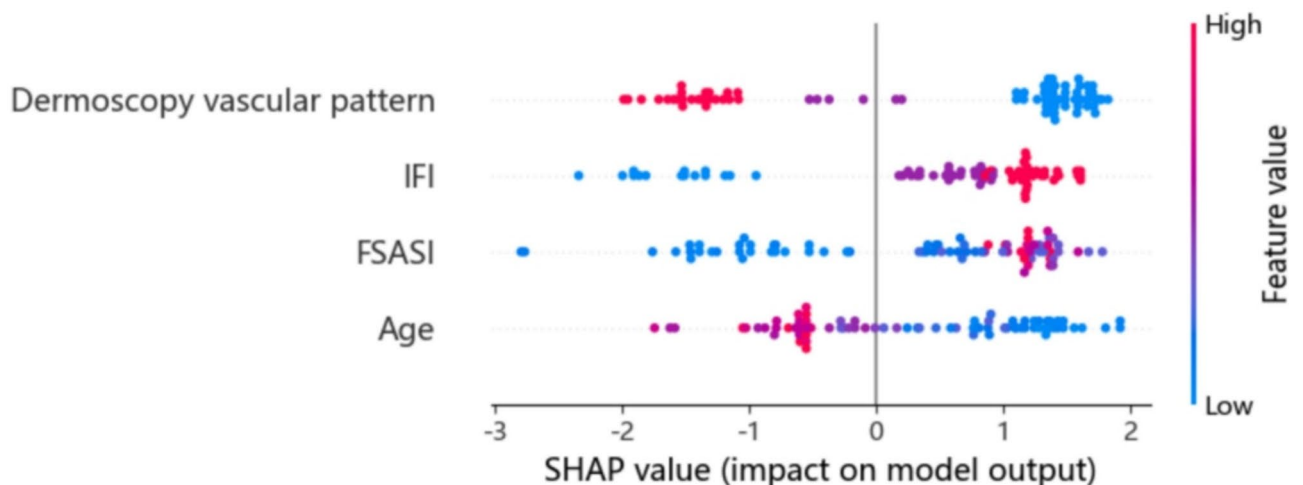
Fig. 6. Overview of SHAP distribution for each feature in the XGBoost model for PWS. The x-axis represents the SHAP values attributed to each feature, and the y-axis lists the features in descending order of importance. Each point in the plot represents a patient, with the color indicating the feature value in the dataset: red denotes higher values, while blue represents lower values. Positive SHAP values suggest that the feature prompts the model to predict the therapy as effective, while negative SHAP values suggest the feature causes the model to predict the therapy as ineffective.

Machine learning algorithms have significant advantages in analyzing complex datasets and fitting potential relationships between features. However, they also inevitably need help with the drawback of poor interpretability[31]. This study utilized feature importance analysis and SHAP to explain the decision-making process and outcomes of the model to understand the model[32]. We conducted global and local analyses to assess the impact and positive or negative correlations of each clinical feature on the prediction model. The feature importance analysis and SHAP results were inconsistent in this study. The rankings of dermoscopy vascular pattern and IFI after HMME-PDT varied, possibly due to differences in the theoretical foundations of both assessment methods employed. Feature importance analysis evaluates the importance of features during node splitting in the model. It is often computed from the gain from node splitting, which measures the improvement in model performance after splitting on a particular feature. This evaluation reflects the degree to which a feature influences model performance during the decision tree construction[33]. On the other hand, SHAP takes the interactions and nonlinear relationships between features into account. It calculates the weighted sum of contributions from each feature to the overall model based on their permutations and combinations. This approach emphasizes the importance of the impact of features on the actual predictive outcomes of the model,
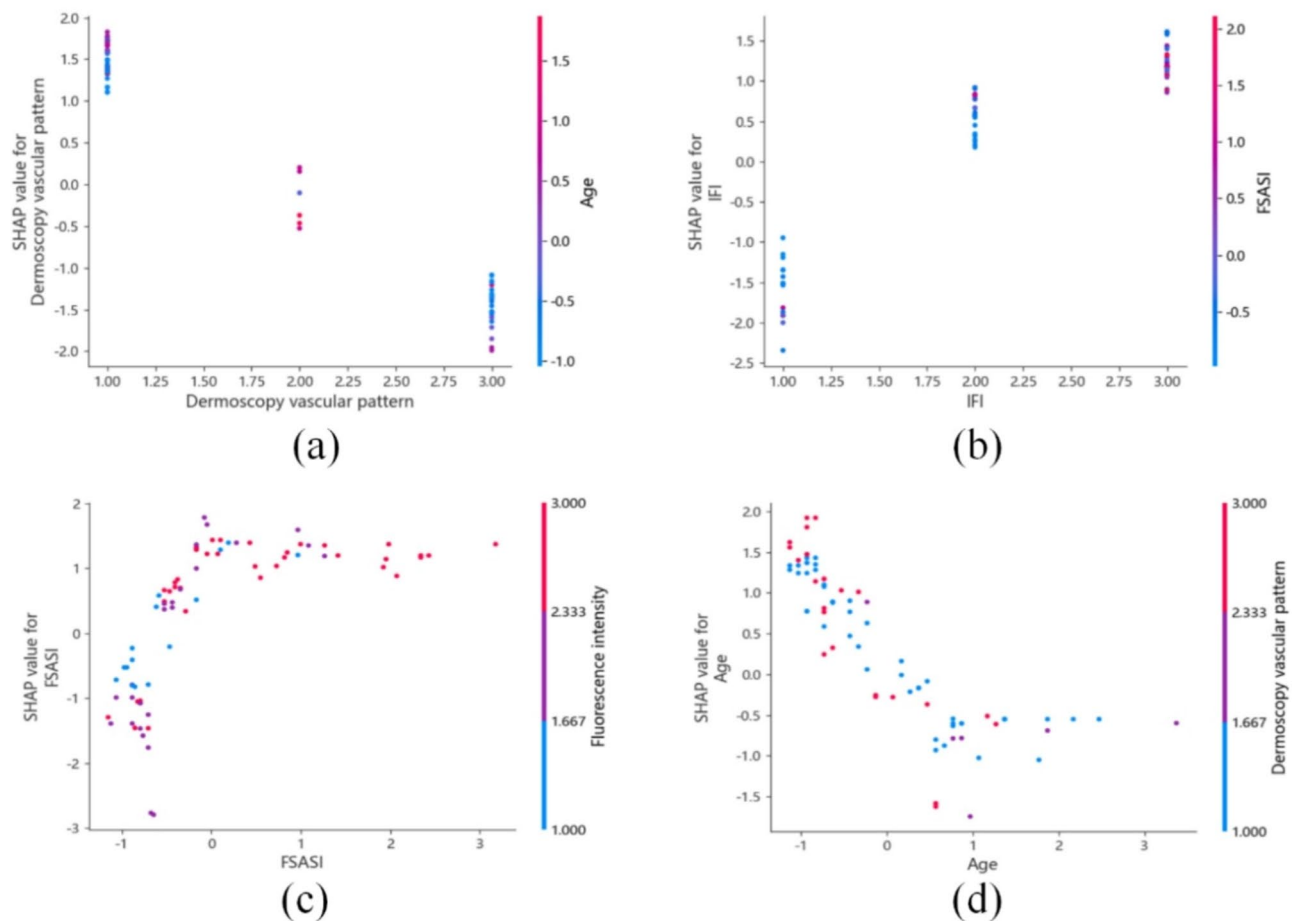
Fig. 7. SHAP distribution of key clinical features in the XGBoost efficacy prediction model for PWS. (**a**) Dermoscopy vascular pattern; (**b**) IFI at the lesion site after HMME-PDT; (**c**) The facial port-wine stain area and severity index (FSASI) score; (**d**) Age. The x-axis represents the statistical values of each feature (with the x-axis in (**a**) and (**b**) showing the encoded values of the features), and the y-axis shows the SHAP values attributed to each feature. Each point on the plot indicates the marginal contribution of a sample to the model's prediction. The color gradient, ranging from blue to red, illustrates the value of another feature that may interact with the primary feature displayed on the x-axis. The color scale is consistent across all subplots, with blue indicating lower values and red indicating higher values of the secondary feature.

making it more relevant to clinical research and practical applications. Given these advantages, our feature ranking was primarily based on SHAP.

In local interpretations based on SHAP, when superficial dermoscopic vascular patterns and strong fluorescence intensity both have SHAP values > 0, the treatment efficacy is generally higher. Previous studies have shown that superficial vascular patterns in PWS correlate with better treatment responses[34,35]. Furthermore, a prospective cohort study of 163 PWS patients found a significant correlation between PDT efficacy and dermoscopic vascular patterns. Dot-like and short rod-like vascular patterns were highly associated with beneficial outcomes and cures, assisting in predicting treatment efficacy[10]. Consistent with these reports, our study found that superficial patterns correspond to better treatment outcomes. Hence, the dermoscopic assessment of vascular features in PWS helps to predict the response to PDT and manage patient expectations.

Fluorescence intensity related to PDT has been previously proposed as an indicator for observing treatment response in clinical settings. For example, Wang et al. established a fluorescence method by measuring real-time fluorescence spectra of the skin during PDT to determine the local concentration of the photosensitizer. A therapeutic effect correlation index (TECI) was proposed as the area under the photosensitizer concentration-time curve during PDT. The correlation between TECI and PDT outcomes in 31 PWS patients revealedthat fluorescence spectroscopy can monitor the concentration of the photosensitizer in the skin during PDT and predict treatment response[11]. However, it has not yet been widely adopted in clinical practice due to the complexity of this technique. Therefore, to the best of our knowledge, our work is the first to propose IFI at the lesion site after HMME-PDT as a new indicator while evaluating its role in predicting therapeutic efficacy.

Our results also indicated that the IFI after HMME-PDT was the second most important factor for HMME-PDT. When the fluorescence intensity was categorized as intermediate or intense, all the SHAP values were greater than 0, indicating a significant impact on model prediction and association with better therapeutic

outcomes. To address the varying fluorescence patterns of skin lesions among different patients, we hypothesize the following possible mechanisms. There is variability in the accumulation of HMME due to factors such as heterogeneity in vascular abnormalities of the lesions and individual patient differences[36]. After PDT, HMME and its photobleaching products can absorb UVA energy and emit fluorescence. Hence, variations in HMME accumulation result in differences in fluorescence intensity. To summarize, our study indicates that higher IFI after HMME-PDT is associated with better therapeutic outcomes. However, the specific mechanisms underlying the fluorescence response and associated influencing factors require further investigation.

The FSASI score is an area and severity index scoring system for facial PWS that divides the face into four regions: forehead, right malar, left malar, and perioral. It provides a comprehensive assessment of PWS lesions based on the percentage of the area affected, lesion color, and thickness. The total scores range from 0 to 42, with higher scores indicating greater severity of the condition. As the facial PWS lesions are lightened, the FSASI scores gradually decrease[19]. SHAP analysis indicated that the FSASI score significantly impacts the PWS efficacy prediction model, and the SHAP values for individual samples show a positive correlation with the outcome for PWS. Previous studies have shown that larger areas and greater thickness of PWS lesions are related to poorer curative effects[37,38]. However, our results are contrary to these reports. This discrepancy might be attributed to smaller, lighter, and thinner lesions showing little changes in the FSASI scores after a single HMME-PDT session. The difference may also reflect a limitation of the FSASI scoring method. Future research should focus on developing more objective and practical efficacy assessment methods for clinical use.

The age of PWS patients undergoing PDT also influences the efficacy. Studies have shown that treatment outcomes for PWS are better in patients under the age of 18[39]. Our study also reveals that younger PWS patients may benefit more from HMME-PDT, which aligns with existing research. This result suggests that early interventions can improve treatment outcomes.

This study presents the IFI after HMME-PDT as a novel metric and applies it to predict the efficacy of HMME-PDT in PWS. As a direct indicator of the distribution and activation status of the photosensitizer in the lesion tissue, fluorescence intensity may be closely associated with treatment outcomes. Our study innovatively employs two machine learning algorithms to construct predictive models for the efficacy of HMME-PDT for PWS. The prediction models can help physicians to better predict patients' response to treatment, thereby optimizing the treatment plan and improving treatment outcomes. In addition, our study provides new ideas for building predictive models for treating PWS, especially by applying machine learning algorithms.

Our study had some limitations. As a retrospective study, the potential biases in the data collection process may lead to a decrease in model stability and an increased risk of overfitting. Furthermore, this was the first study to explore the IFI after HMME-PDT to assess and predict the efficacy of PWS. Thus, our study was limited by a relatively small sample size and the lack of external validation data. This may limit the model's ability to fully understand and learn from the data, leading to increased instability and a potential risk of overfitting. These limitations underscore the necessity for future research, particularly the conduct of large-scale, multi-center, prospective cohort studies, to further validate our findings and provide more robust and reliable evidence to support the treatment of PWS.

Conclusions

Our study utilized the XGBoost-RFE feature selection method to identify four key clinical features and built two predictive models for the therapeutic effect of HMME-PDT in PWS patients using two different machine learning algorithms. The XGBoost prediction model demonstrated better performance. Furthermore, for the first time, we proposed and explored the relationship between IFI at the lesion site after HMME-PDT and the treatment efficacy. Future studies could increase sample size and include multicenter data to validate model generalizability and robustness. This would allow physicians to quickly assess patients' treatment responses in clinical practice.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References
1. Cordoro, K. M. et al. Physiologic changes in vascular birthmarks during early infancy: mechanisms and clinical implications. *J. Am. Acad. Dermatol.* **60**(4), 669–675 (2009).
2. Minkis, K., Geronemus, R. G. & Hale, E. K. Port wine stain progression: a potential consequence of delayed and inadequate treatment? *Lasers Surg. Med.* **41**(6), 423–426 (2009).
3. Huikeshoven, M. et al. Redarkening of port-wine stains 10 years after pulsed-dye-laser treatment. *N Engl. J. Med.* **356**(12), 1235–1240 (2007).
4. Diao, P. et al. Hematoporphyrin monomethyl ether photodynamic therapy of Port wine stain: narrative review. *Clin. Cosmet. Investig Dermatol.* **16**, 1135–1144 (2023).
5. Zhang, X. et al. Hemoporfin-mediated photodynamic therapy for the treatment of port-wine stain: A multicenter, retrospective study. *Photodiagnosis Photodyn Ther.* **42**, 103545 (2023).
6. Zhao, Y. et al. Hemoporfin photodynamic therapy for Port-Wine stain: A randomized controlled trial. *PLoS One.* **11**(5), e0156219 (2016).
7. Cui, Z. et al. Polymyxin B-targeted liposomal photosensitizer cures MDR A. baumannii burn infections and accelerates wound healing via M1/M2 macrophage polarization. *J. Control Release.* **366**, 297–311 (2024).
8. Cui, Z. et al. Antifungal effect of antimicrobial photodynamic therapy mediated by haematoporphyrin monomethyl ether and Aloe Emodin on malassezia furfur. *Front. Microbiol.* **12**, 749106 (2021).

9. Wang, X. et al. Non-invasive detection technology in port-wine stain treatment. *Chin. Med. J. (Engl)*. **135**(21), 2535–2537 (2022).
10. Wang, X. et al. Correlation between the hemoporfin-mediated photodynamic treatment response and the dermoscopy vascular pattern in patients with a port-wine stain: a prospective study. *J. Eur. Acad. Dermatol. Venereol*. **34**(12), 2795–2801 (2020).
11. Wang, Y. et al. Fluorescence monitoring of a photosensitizer and prediction of the therapeutic effect of photodynamic therapy for Port wine stains. *Exp. Biol. Med. (Maywood)*. **235**(2), 175–180 (2010).
12. Peiffer-Smadja, N. et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clin. Microbiol. Infect*. **26**(5), 584–595 (2020).
13. Tuppad, A. & Patil, S. D. Machine learning for diabetes clinical decision support: a review. *Adv. Comput. Intell*. **2**(2), 22 (2022).
14. Cearns, M. et al. Machine learning probability calibration for high-risk clinical decision-making. *Aust N Z. J. Psychiatry*. **54**(2), 123–126 (2020).
15. Munger, E. et al. Application of machine learning to determine top predictors of noncalcified coronary burden in psoriasis: an observational cohort study. *J. Am. Acad. Dermatol*. **83**(6), 1647–1653 (2020).
16. Maintz, L. et al. Machine Learning-Based deep phenotyping of atopic dermatitis: Severity-Associated factors in adolescent and adult patients. *JAMA Dermatol*. **157**(12), 1414–1424 (2021).
17. Lin, Z. et al. Construction of a disease risk prediction model for postherpetic pruritus by machine learning. *Front. Med*. **11**, 1454057 (2024).
18. Greener, J. G. et al. A guide to machine learning for biologists. *Nat. Rev. Mol. Cell. Biol*. **23**(1), 40–55 (2022).
19. Ren, J. et al. A proposed scoring system for facial port-wine stain evaluation: facial port-wine stain area and severity index. *J. Cosmet. Dermatol*. **21**(7), 2931–2938 (2022).
20. Zhou, P. et al. Balanced spectral feature selection. *IEEE Trans. Cybern*. **53**(7), 4232–4244 (2023).
21. Gould, A. L., Baumgartner, R. & Zhao, A. Bayesian screening for feature selection. *J. Biopharm. Stat*. **32**(6), 832–857 (2022).
22. Li, L., Ching, W-K. & Liu, Z-P. Robust biomarker screening from gene expression data by stable machine learning-recursive feature elimination methods. *Comput. Biol. Chem*. **100**, 107747 (2022).
23. Lundberg, S. M. et al. From local explanations to global Understanding with explainable AI for trees. *Nat. Mach. Intell*. **2**(1), 56–67 (2020).
24. Savas, J. A. et al. Pulsed dye laser-resistant port-wine stains: mechanisms of resistance and implications for treatment. *Br. J. Dermatol*. **168**(5), 941–953 (2013).
25. Zheng, J. et al. Clinical data based XGBoost algorithm for infection risk prediction of patients with decompensated cirrhosis: a 10-year (2012–2021) multicenter retrospective Case-control study. *BMC Gastroenterol*. **23**(1), 310 (2023).
26. Lu, D. et al. Dielectric property measurements for the rapid differentiation of thoracic lymph nodes using XGBoost in patients with non-small cell lung cancer: a self-control clinical trial. *Transl Lung Cancer Res*. **11**(3), 342–356 (2022).
27. Zhu, X-W., Xin, Y-J. & Ge, H-L. Recursive random forests enable better predictive performance and model interpretation than variable selection by LASSO. *J. Chem. Inf. Model*. **55**(4), 736–746 (2015).
28. Liu, Y. & Zhao, H. Variable importance-weighted random forests. *Quant. Biol*. **5**(4), 338–351 (2017).
29. Gu, Q. et al. A novel random forest integrated model for imbalanced data classification problem. *Knowl. Based Syst*. **250** (2022).
30. Li, X. W. et al. Exploring interactive and nonlinear effects of key factors on intercity travel mode choice using XGBoost. *Appl. Geogr*. **166** (2024).
31. Amarasingham, R. et al. Implementing electronic health care predictive analytics: considerations and challenges. *Health Aff (Millwood)*. **33**(7), 1148–1154 (2014).
32. Zhang, G. et al. A machine learning model based on ultrasound image features to assess the risk of Sentinel lymph node metastasis in breast cancer patients: applications of scikit-learn and SHAP. *Front. Oncol*. **12**, 944569 (2022).
33. Mendoza, T. et al. Random forest for automatic feature importance Estimation and selection for explainable postural stability of a Multi-Factor clinical test. *Sens. (Basel)* ;**21**(17) (2021).
34. Kwiek, B. et al. Predictive value of dermoscopy for the treatment of port-wine stains with large spot 532 Nm laser. *Lasers Surg. Med*. **51**(7), 569–583 (2019).
35. Vázquez-López, F. et al. Dermoscopy of port-wine stains. *Arch. Dermatol*. **143**(7), 962 (2007).
36. Xu, J. et al. Study of pharmacokinetics and cutaneous photosensitization of Hemoporfin in healthy volunteers. *Photodiagnosis Photodyn Ther*. **26**, 116–123 (2019).
37. Liu, J. et al. Retrospective analysis of Hemoporfin-mediated photodynamic therapy in the treatment of Naïve port-wine stains. *Photodiagnosis Photodyn Ther*. **39**, 103003 (2022).
38. Liao, Z., Nie, J. & Sun, P. The impact of particulate matter (PM2.5) on skin barrier revealed by transcriptome analysis: focusing on cholesterol metabolism. *Toxicol. Rep*. **7**, 1–9 (2020).
39. Khalaf, A. T. et al. Photodynamic therapy using HMME for Port-Wine stains: clinical effectiveness and sonographic appearance. *Biomed. Res. Int*. **2020**, 6030581 (2020).

Acknowledgements

Author contributions

H.Y: Data curation, Writing-Original draft preparation, Formal analysis, Visualization. Y.T: Writing–review & editing, Methodology. F.Q: Software, Formal analysis, Methodology. Z.Z: Methodology, Formal analysis. Y.S: Methodology. X.Z: Project administration, Investigation. L.L: Project administration, Investigation. T.Z: Project administration, Investigation. Y.Z: Project administration, X.H: Data curation. R.Y: Data curation. R.X: Conceptualization, Methodology, Supervision, Validation, Funding acquisition. X.Q: Conceptualization, Methodology, Resources, Validation, Investigation.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

This study was approved by the clinical research ethics committee of Second Xiangya Hospital of Central South University (LYEC2024-K0152). The research was performed in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from all patients before treatment. The patients mentioned in this manuscript have given written informed consent to the release of images that may lead to identification. Specially, minor patients was obtained consent from their parents and legal guardians.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-06589-3.

Correspondence and requests for materials should be addressed to R.X. or X.Q.

Reprints and permissions information is available at www.nature.com/reprints.

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