

## ORIGINAL RESEARCH

# Fertility-sparing treatments decision in patients with endometrial cancer based on machine learning

Yue Sun<sup>1</sup>, Zhi Li<sup>1</sup>, Li Gao<sup>2</sup>, Wenhan Yuan<sup>3</sup>, Fan Yang<sup>3,\*</sup>

<sup>1</sup>College of Electronics and Information Engineering, University of Sichuan, 610065 Chengdu, Sichuan, China

<sup>2</sup>Department of Neurology, Chengdu Third People's Hospital, 610031 Chengdu, Sichuan, China

<sup>3</sup>Department of Gynecology and Obstetrics, Key Laboratory of Obstetric and Gynecologic and Pediatric Diseases and Birth Defects of Ministry of Education, West China Second Hospital, University of Sichuan, 610041 Chengdu, Sichuan, China

**\*Correspondence**

sharry48@163.com

(Fan Yang)

**Abstract**

Although many studies have been recently performed on fertility-sparing treatments in patients with endometrial cancer (EC) and endometrial atypical hyperplasia (EAH), most of the corresponding studies were retrospective and small sample research. However, it is essential to more thoroughly assess the necessity of hysterectomy in EC patients using current accumulated experience. With the development of machine learning (ML), it has been gradually integrated into oncologic research but seldom applied to predict the efficacy of hysterectomy due to an insufficient number of patients who did not undergo hysterectomy, leading to a learning imbalance. Thus, the commonly used machine learning models cannot provide satisfying performance. In this study, we aimed to develop ML models to predict whether hysterectomy is necessary for EC patients and help gynecologists determine the possibility of fertility-preserving treatment in EC patients. A clinical dataset of 1534 women with EC was analyzed. The Borderline-SMOTE algorithm was employed to solve imbalanced learning issues. Then, the Adaptive Boosting (AdaBoost) algorithm, which is less susceptible to overfitting than other machine learning algorithms, was used to build a high-performance ensemble classification model. The findings indicated that the method outperformed conventional machine learning models and provided a realistic strategy to make fertility-preserving treatment decisions. The proposed model provides a platform for physicians to precisely predict the efficacy of fertility-sparing therapy in EC patients, allows gynecologists to select the optimal treatment for a patient, and reduces resource waste and risks of overtreatment.

**Keywords**

Endometrial cancer; Machine learning; Borderline-SMOTE; AdaBoost

## 1. Introduction

Endometrial cancer (EC) is among the most common malignant tumor of the female reproductive system, ranking second in incidence among gynecological malignant tumors in China and first in developed countries [1, 2]. The yearly incidence rate is increasing and is trending towards the younger population. Although EC occurs predominantly in postmenopausal women, it affects around 7% of patients aged between 20 and 44 years old. At the same time, the delay in child-bearing has led to an increase in EC in nulliparous women at child-bearing age. The common risk factors of EC are associated with fertility, hormones, metabolism, and physiological behaviors [3]. Despite hysterectomy being the standard treatment, many young women seek fertility-sparing therapy to prevent permanent loss of fertility. To meet patients' needs, it is therefore essential to precisely evaluate the necessity of hysterectomy to provide superior postoperative conservation treatment. Endometrial atypical hyperplasia (EAH) is a precancerous lesion of the endometrium and may progress to cancer. Therefore,

patients with EAH were included in our study.

Artificial intelligence mainly falls into two major categories. The first category includes machine learning (ML) techniques that analyze structured data such as imaging, genetic and electrophysiological (EP) data. In medical applications, the ML procedures attempt to cluster patients' traits or infer the probability of disease outcomes. The second category includes natural language processing (NLP) methods that extract information from unstructured data, such as clinical notes/medical journals, to supplement and enrich structured medical data. The NLP procedures turn texts into machine-readable structured data that can be analyzed by ML techniques [4]. Deep learning is a subtype of artificial intelligence whereby sufficient training data can provide greater contributions and more accurate results to the medical field. For instance, Serj *et al.* [5] proposed a deep convolutional neural network for lung cancer diagnosis. Lin *et al.* [6] proposed using multiphoton microscopy and deep learning to automatically classify the differentiation of hepatocellular carcinoma. Yan *et al.* [7] performed deep combination learning of multi-level features

for Magnetic Resonance (MR) prostate segmentation. Hermes *et al.* [8] proposed deep feature learning of soft tissue sarcoma classification using MR images through transfer learning. Ghassemi *et al.* [9] reported a deep neural network with pre-training to generate a confrontation network for brain tumor classification based on MR images. Artificial intelligence has been applied in EC. Pergialiotis *et al.* [10] conducted a retrospective case-control study on the pathological reports of 106 EC and 72 healthy postmenopausal women to predict the risk of EC using parameters including age, menopause, diabetes, hypertension, smoking and obesity. They found that the correct diagnostic rates of artificial neural networks (ANNs), regression, and cart algorithms were 85.4%, 72.5%, and 77.5%, respectively, indicating that the ANNs algorithm had the best accuracy, while regression analysis showed the poorest prediction accuracy. Vezzo *et al.* [11] used the demographics data, tumor markers (serum human epididymis secretory protein 4 (HE4) and cancer antigen 125 (CA125)) and histological grades of 293 EC patients to predict extrauterine disease (The International Federation of Gynecology and Obstetrics (FIGO) stage > I) by implementing a new statistical method of representative regression tree (RERT) combined with a random forest algorithm to predict the risk of late FIGO stage in EC patients. The model was shown to be superior to logistic regression, with a 90% correct diagnosis rate and 76% specificity.

In this study, we developed a machine learning method trained using basic information, high-risk factors and hospital laboratory examination data of EC and EAH patients to construct a decision-making model for predicting the efficacy of fertility-sparing treatments in EC patients based on BorderlineSMOTE-AdaBoost algorithms. Data augmentation was used to synthesize minority samples to reduce imbalances in the sample.

## 2. Materials and methods

This section describes the methodological approach, including data pre-processing and the methods used for data extraction and data augmentation.

### 2.1 Ensemble learning

Ensemble learning produces collaborative decisions on assignments by constructing numerous learners, also referred to as a multi-classifier system, committee-based learning, among others. The primary objective is to cultivate a large number of weak learners using particular integration tactics. The primary idea is first to generate numerous weak learners *via* specialized integration processes, then to obtain model results *via* combination methods of joint judgment [12–14].

In ensemble learning, multiple weak learners (or base models) are often homogeneous, each constructed using the same algorithm. Then, the idea of ensemble methods is to try reducing the bias and/or variance of such weak learners by combining several of them to create a strong learner (or ensemble model), thereby achieving better performances. Bagging and boosting are the two principal ensemble learning methods used in ensemble learning algorithms, and the Adaptive Boosting

(AdaBoost) algorithm is selected to construct the integrated classifier based on the classification accuracy requirement.

#### 2.1.1 Boosting

Boosting [15] is an ensemble modeling technique that attempts to build a strong classifier from the number of weak classifiers. It is performed by building a model using a serial number of weak models. Fig. 1 shows a model built from the training data. Then, the second model is built to correct the errors present in the first model. This procedure is continued, and models are added until the complete training data set is predicted correctly or the maximum number of models has been added. AdaBoost, a statistical classification meta-algorithm developed by Yoav Freund and Robert Schapire, is the most typical algorithm in the Boosting family [16]. AdaBoost is adaptive as it can adjust sample weights based on their performance following the training of each weak learner. Examples that the previous weak learners erroneously classified are more heavily weighted. In contrast, successful models are allocated a reduced weight, and the revised datasets are used to train the succeeding weak learners. Being an iterative process, it continues to add learner(s) until a limit is reached in the number of models or accuracy. AdaBoost's final strong learner is determined, as shown in Eqn. 1, which represents the weighted linear combination of the weak learners:

$$H(x) = \sum_{t=1}^T a_t h_t(x) \quad (1)$$

Here,  $T$  represents the number of weak learners,  $h_t(x)$  represents each weak learner, and  $a_t$  represents the weight of each weak learner.

Eqn. 2 was implemented to minimize the exponential loss function:

$$\ell_{\text{exp}}(H | D) = E_{X-D} \left[ e^{-f(x)H(x)} \right] \quad (2)$$

Here,  $D$  and  $f(x)$  represent the original data set and the objective function.

AdaBoost has several advantages, including its easy use and fewer tweaking, compared with algorithms such as support vector machine (SVM). In addition, AdaBoost can be used with SVM. Theoretically, AdaBoost is not prone to overfitting though no concrete proof exists. It could be because the parameters are not jointly optimized, causing stage-wise estimation to slow down the learning process. AdaBoost can improve weak classifiers' accuracy, hence making them flexible. It has now been extended beyond binary classification and has found use cases in text and image classification. Some of the disadvantages of AdaBoost include: (1) As boosting technique learns progressively, it is essential to ensure quality data; (2) it is also extremely sensitive to noisy data and outliers, and; (3) AdaBoost has also been proven to be slower than bagging.

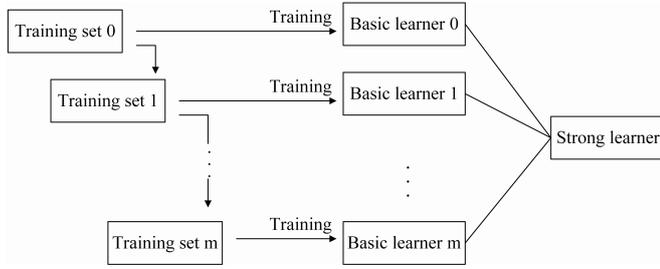


FIGURE 1. Illustration of Boosting ensemble classifiers.

### 2.1.2 The support vector machine

SVM is an extensively used machine learning algorithm for binary classification tasks. The fundamental concept of the algorithm is to locate a hyperplane in the data set that will classify the data into precisely two classes. The Support Vectors are comprised of the sample points of each data class closest to the hyperplane. In many instances, multiple hyperplanes may be able to classify a data set into two groups. A support vector machine is tasked with selecting the ideal hyperplane from these hyperplanes. The criterion for selecting the hyperplane and the support vector is that the margin between the hyperplane and the support vector should be as large as possible so that the spacing between the sample points on the boundary of the different classes is maximized, thereby minimizing classification error and allowing the classification model to achieve optimal generalization performance [12].

In the sample space, the equation of hyperplane can be represented *via* Eqn. 3:

$$w^T + b = 0 \quad (3)$$

Here,  $w$  represents the normal vector, and  $b$  represents the displacement. It can be seen that these two parameters jointly determine a hyperplane.

The distance  $r$  from a sample point  $x$  in the sample space to the hyperplane can be calculated using Eqn. 4:

$$r = \frac{|w^T + b|}{\|w\|} \quad (4)$$

The sum of the distances from two different categories of support vectors to the hyperplane is:

$$\gamma = \frac{2}{\|w\|} \quad (5)$$

Here,  $\gamma$  represents the margin. From the SVM principle, the best hyperplane can be determined to maximize the interval between it and the support vector, *i.e.*, to get the optimal  $w$ ,  $b$ , so that  $\gamma$  has the greatest value.

The aforementioned ideas have been explored in the context of linearly divided datasets. In practice, however, datasets are frequently linearly indivisible, indicating that a better hyperplane cannot be identified to segregate all samples under the original sample space. An appropriate hyperplane cannot be

discovered in the initial two-dimensional space for a dataset in which the sample points of two categories are scattered in a circular pattern with the same center. Nonetheless, if the actual distance can be mapped to a three-dimensional feature space, it can be straightforward to discern the hyperplane separating the samples of these two groups. To address this issue, kernel functions were added as they can translate the models from the original sample space to a higher-dimensional feature space, eventually locating an ideal hyperplane inside this higher-dimensional feature space. Numerous studies have shown that the SVM algorithm outperforms other classification algorithms in numerous academic disciplines. Further, as the SVM method was shown to outperform other classification algorithms in several studies, this has led to its adoption in various research fields. Therefore, our present study used this algorithm to build each weak learner in the AdaBoost integrated learner.

## 2.2 Data

This study collected the relevant data of EC and EAH patients in West China Second Hospital between 01 January 2000, and 31 December 2020, for ML. As one of the centers of Chinese gynecological hospitals, our institution treats patients from all over China, including different provinces and nationalities, fully reflecting the diversity of the investigated samples and might represent the general population in China, to a certain extent. As illustrated in Fig. 2, each patient was diagnosed with EC or endometrial atypical hyperplasia (EAH). Patients without inpatient and outpatient records and those missing >25 important data points were excluded. After patient selection, a total of 1534 patients were eligible for this study, and their data were analyzed.

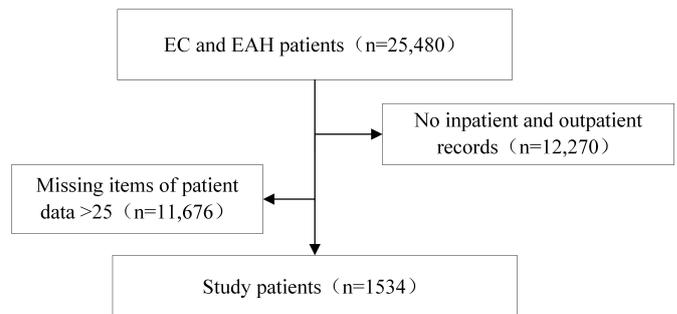


FIGURE 2. Flow chart of patient selection. EC: endometrial cancer; EAH: endometrial atypical hyperplasia.

### 2.2.1 Data extraction of clinical diagnosis, treatment records and laboratory examination items

The electronic medical record (EMR) of inpatients and outpatients contains a large amount of data. The data of interest extracted from the electronic medical record included: current medical history, menstruation, marriage and childbirth history, history, discharge diagnosis, inpatient diagnosis, outpatient diagnosis, preliminary diagnosis, personal history, and laboratory information system (LIS) test data. However, the

extracted data could not be directly used for machine learning and the task of predicting the effectivity of hysterectomy.

In recent years, some researchers have applied natural language processing and word embedding models to the representation of medical concepts, transforming unstructured clinical records into structural data. Choi *et al.* [17] used a Skip-gram algorithm to learn low-dimensional vector representation of medical insurance data of health insurance companies. The Medical Language Extraction and Encoding System (MEDLEE system), designed by Carol Friedman of Columbia University and others, is a very successful medical information extraction system. As an independent module of the clinical information system (CIS), it is used in New York Presbyterian Hospital. It can convert medical records into encoded data to promote breast cancer research, which is conducive to improving patient care quality [18]. In this study, we used ML to extract data from electronic medical records and used pandas to clean, prepare and regularize the data [19]. Through exploratory analysis of patient data by Exploratory Data Analysis (EDA), the age density distribution of patients was obtained, as shown in Fig. 3.

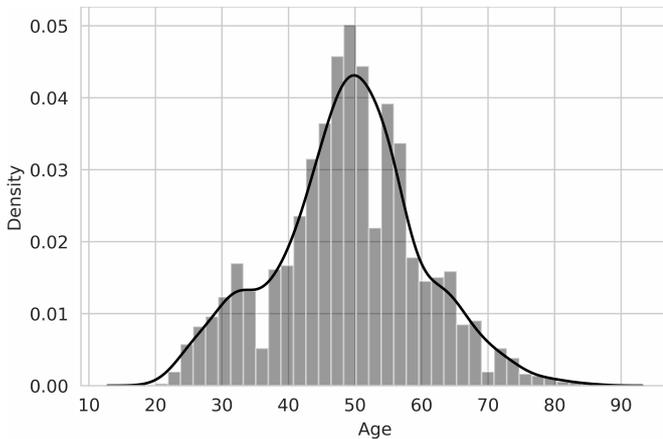


FIGURE 3. Patient age density distribution.

### 2.2.2 Missing value processing

Many variables in the data collection table derived from outpatient, inpatient, and laboratory information system (LIS) data have NaN values due to incomplete or missing patient records. Similar to machine learning models, AdaBoost requires complete data. Thus, first consider the following two methods: the deletion method and the data completion method (K Nearest Neighbors (KNN) interpolation, Linear interpolation, Kalman interpolation [20], and Multiple Imputation by Chained Equations (MICE) [21]).

The deletion method can provide a complete data subset by deleting the data containing missing values. Data completion is usually based on the distribution of other data values in the information table to fill a blank value.

MICE is a robust informative method dealing with missing data in datasets. The procedure ‘fills in’ (imputes) missing data in a dataset through an iterative series of predictive models. In each iteration, each specified variable in the dataset is added using the other variables. These iterations are run until convergence has been met.

### 2.2.3 Borderline-SMOTE

An insufficient number of patients without hysterectomy often leads to learning imbalance. Machine learning usually requires a large number of balanced datasets. Therefore, when applied to complex datasets with unbalanced class distribution, these networks often provide inaccurate prediction performance and cannot correctly represent data distribution characteristics [22].

Synthetic Minority Over-Sampling Technique (SMOTE) algorithm was first proposed by Chawla *et al.* [23]. It allows oversampling by linear interpolation between minority samples and their nearest neighbors to synthesize new minority samples, effectively preventing overfitting problems.

However, one of the drawbacks of the SMOTE algorithm is that it can lead to sample aliasing. When synthesizing pieces, the SMOTE method does not account for the location and distribution of minority class samples. For a minority class sample at the edge of the minority class sample set, the new selection generated through it will likely be at the edge of the minority class sample set, increasing the difficulty of classifying boundary samples [22, 24]. Han *et al.* [25] presented the Borderline-SMOTE algorithm to overcome this issue. As illustrated in Fig. 4, during the operation of this algorithm, unlike SMOTE which only considers minority class samples when selecting the nearest neighbors, the Borderline-SMOTE sampling process divides a small number of samples into three categories: safe (more than half of the closest neighbor samples are minority samples, as shown in sample point A in the figure), danger (more than half of the nearest neighbor samples are majority samples, as shown in sample point B in the figure) and noise (all the nearest neighbor samples are majority samples, as shown in sample point C in the figure). Only the danger category was used to generate synthetic instances. Thus, the method only over-sample the borderline examples of the minority class to improve the class distribution of the samples.

This study’s nearest neighbor number of minority samples was set to 5. The data distribution before and after data enhancement is shown in Fig. 5, including 1826 samples, 1496 samples with hysterectomy (76.6%), and 330 samples without hysterectomy (23.4%). The enhanced data were only added to the training set, while the other not enhanced data were randomly divided into a training dataset and a test set. Therefore, only accurate clinical data were present in the test set, resulting in 1365 (74.75%) patients in the training set and 461 (25.25%) in the test set.

### 2.3 Feature extraction

Before processing the training data set, feature analysis was performed to identify and output highly associated features to increase accuracy and reduce errors. Here, we first used the Pearson correlation coefficient (Pearson) to perform correlation analysis on all features. Next, features strongly correlated with different categories were selected at a significance level of  $\leq 0.01$  ( $p \leq 0.01$ ) and were used as input parameters.

The Pearson correlation coefficient, which precisely measures the degree of linear correlation between two variables, was produced by dividing the covariance with the standard

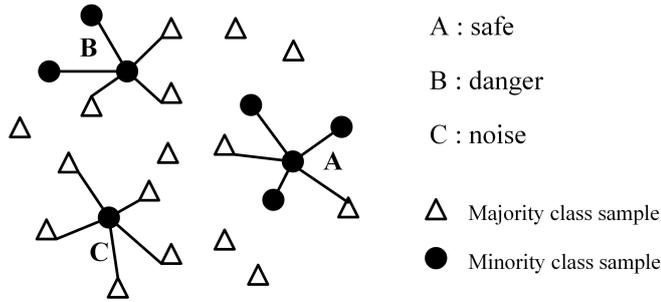


FIGURE 4. Illustration of Borderline-SMOTE algorithm.

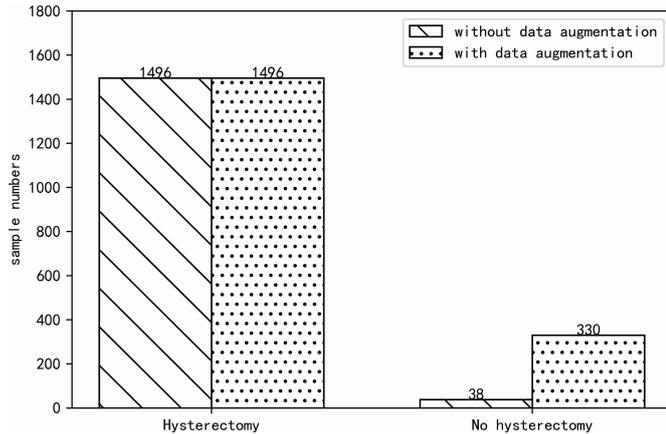


FIGURE 5. Distribution comparison of data with and without Borderline-SMOTE-based data augmentation.

deviation of the two variables. Its value ranged from  $-1$  to  $1$ . When the correlation between two attributes was highly significant, its value was near  $1$  or  $-1$ . A coefficient of  $0$  indicated no linear correlation [26]. The equation for determining the Pearson coefficient is presented in Eqn. 6:

$$r_p = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y} \quad (6)$$

Here,  $\sigma_x$  and  $\sigma_y$  represents the standard deviation of characteristic  $x$  and  $y$ , respectively;  $\sigma_{xy}^2$  represents the covariance of distinct  $x$  and  $y$ . The calculation formulas of standard deviation and covariance were as follows, respectively:

$$\sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (7)$$

$$\sigma_{xy}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (8)$$

Here,  $x_i$  represents each input feature,  $\bar{x}$  represents the average value of this input feature,  $y_i$  represents the actual situation of each patient sample (irrespective of whether the uterus was removed or not),  $\bar{y}$  represents the average value of

all models, and  $n$  represents the number of pieces.

To identify criteria highly correlated with hysterectomy, Pearson correlation coefficients were used to explore the most significant characteristics supporting or rejecting the need for hysterectomy. The SPSS (v22.0, International Business Machines Corporation (IBM), Armonk, NY, USA) software was used for statistical analysis. The strength of the link between the various parameters and whether or not the patient should undergo hysterectomy is displayed in Table 1, which was based on the Pearson correlation coefficient.

## 2.4 The decision-making model based on BorderlineSMOTE-AdaBoost

This study employed AdaBoost to determine the need for hysterectomy in EC patients. The Borderline-SMOTE algorithm was employed to solve imbalanced learning issues. Then, an AdaBoost model was developed using the augmented data to predict whether hysterectomy is necessary for EC patients.

Fig. 6 illustrates the framework of the decision-making model for fertility-sparing therapy in patients with EC based on BorderlineSMOTE-AdaBoost. Based on data synthesis utilizing the Borderline-SMOTE algorithm, this paper introduced the idea of ensemble learning in constructing a classifier. As shown in Table 1, age (Age), parity (Para), a history of drugs (Pha), diabetes (Dia), hypertension (Hyp), a history of surgery (OH), menstruation (MS), neutrophil count (NEUT), Lymphocyte count (LYMPH), lymphocyte percentage (LYMPH%), albumin (ALB), serum urea (ure) and other parameters were significantly correlated with hysterectomy ( $p < 0.01$ ). Therefore, Age, Para, Pha, Dia, Hyp, OH, MS, NEUT, LYMPH, LYMPH%, ALB and ure were selected as the input parameters. Our method was based on AdaBoost, the weak classifier adopted the SVM algorithm based on the probabilistic mode, and in each round of weak classifier training, parameter optimization was performed. The maximum number of weak classifiers in the experiment was set to 500, and the integrated classifier was constructed using the AdaBoost method. In the construction of the classification model, the sample numbers for the patients who have not undergone hysterectomy were insufficient, resulting in an imbalanced learning problem. Thus, the Borderline-SMOTE was employed to solve this issue. All these augmented data were only put into the training set. Then, we randomly divided the other data without data augmentation into the training and testing sets.

## 3. Results

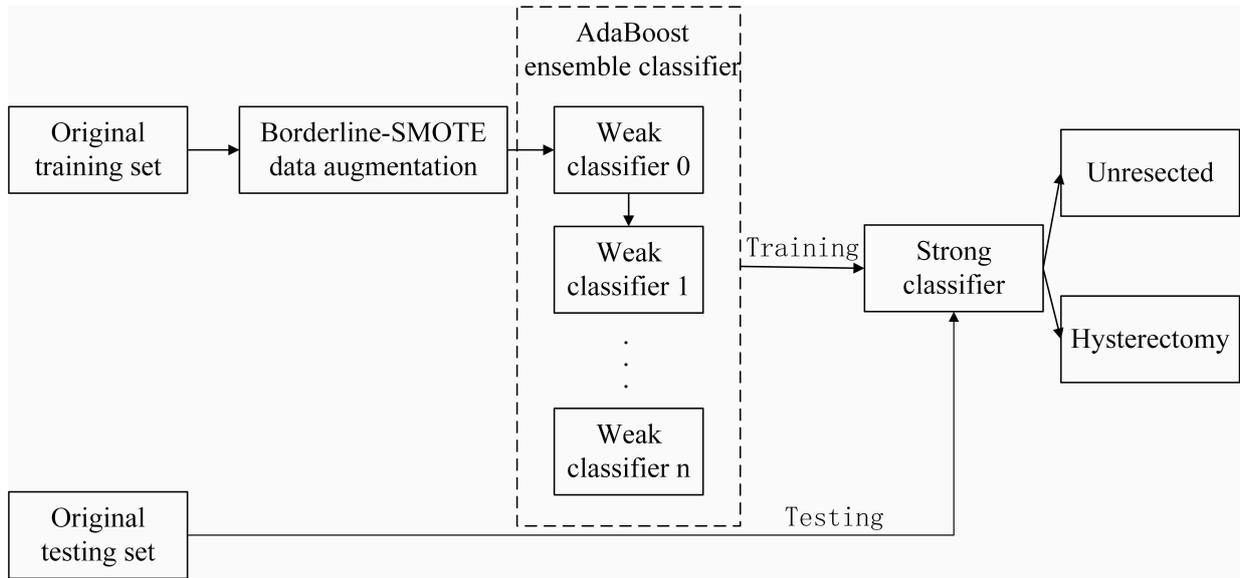
### 3.1 Performance comparison of imbalanced and balanced data sets

We constructed an AdaBoost-based prediction model to examine the impact of the data synthesis technique on the classification effect of the two categories. The original data without data synthesis and the data produced by data synthesis were imported into the training and testing model. To ensure the model's viability, K-fold cross-validation was used to assess the performance of our model by dividing the data into K divisions of the same size. Fig. 7 shows the outcomes of the data sets. Due to the imbalance issue, to improve the prediction

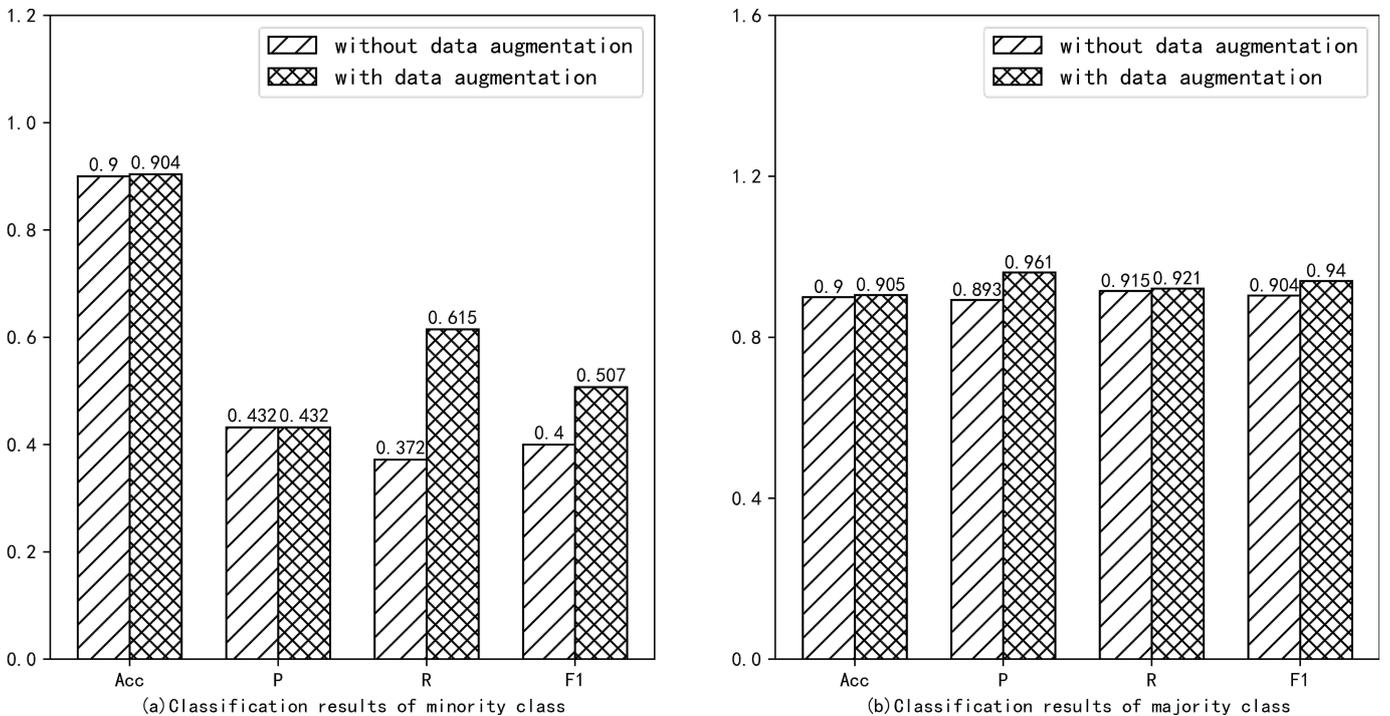
**TABLE 1. Pearson correlation coefficient between hysterectomy and patient characteristics.**

	Pearson correlation coefficient					
	Age	Para	Pha	Dia	Hyp	OH
Whether or not to have a hysterectomy	0.323**	0.186**	0.085**	0.077**	0.117**	0.315**
	MS	NEUT	LYMPH	LYMPH%	ALB	ure
	0.124**	0.153**	0.154**	0.168**	0.141**	0.078**

“\*\*” indicates a strong correlation between this feature and whether the patient needed a hysterectomy. Para: parity; Pha: a history of drugs; Dia: diabetes; Hyp: hypertension; OH: a history of surgery; MS: menstruation; NEUT: neutrophil count; LYMPH: Lymphocyte count; ALB: albumin; ure: serum urea.



**FIGURE 6. The decision-making model of fertility-sparing treatments in patients with EC based on BorderlineSMOTE-AdaBoost.**

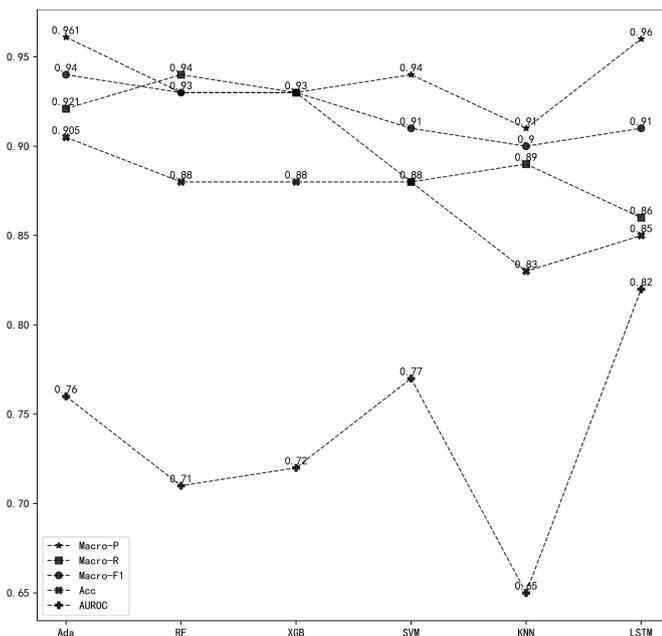


**FIGURE 7. Evaluating the effects of the proposed model before and after data augmentation. P: accuracy; R: recall; F1: F1 value; ACC: accuracy.**

performance of the minority class, we primarily compared the results of the minority class sample, and the results showed that the proposed model's data augmentation had improved the predictive performance of the minority group of patients without hysterectomy. Thus, using the balanced dataset to make fertility-sparing treatment decisions for EC patients was found to be more beneficial than using the original unbalanced dataset.

### 3.2 Accuracy comparison of different prediction models

We compared the proposed model with some basic models, including Random Forest (RF), eXtreme Gradient Boosting (XGB), SVM, KNN, and Long Short Term Memory (LSTM). Their accuracy (ACC), accuracy (P), recall (R), F1 value (F1) and area under ROC curve (AUROC) values were computed and compared using the same test set (Fig. 8). The results showed that the proposed model was superior to the basic machine learning methods. From the prediction results of LSTM, it can be seen that LSTM can be used as a new method to develop prediction models and provide a machine learning method for subsequent processing of fertility-sparing treatments in EC patients.



**FIGURE 8. Comparison of results of other models.** AUROC: area under ROC curve.

## 4. Discussion

Hysterectomy is the standard treatment for EC. Although adverse events associated with hysterectomy are becoming less severe due to advancements in minimally invasive techniques, complete removal of the uterus *via* hysterectomy renders young nulliparous EC patients permanently infertile, with irreversible effects on their quality of life and mental health. Young EC patients usually face the dilemma of standard cancer treatment and the desire to keep fertility.

In recent years, fertility-sparing treatments for EC patients have attracted considerable interest. Therefore, obstetricians and gynecologists must accurately evaluate the need for hysterectomy in patients with EC to provide the optimal treatment that would not only fulfill the patient's desire but also maximize treatment outcomes. Most currently reported studies on fertility-sparing treatments in patients with EC are retrospective and small sample research. Eleven national and international organizations have released guidelines and expert consensus on fertility management in EC in recent years, some of which differ due to the lack of evidence-based medical proof and controversies in actual clinical care. Substantial discoveries were also made from studies investigating the molecular type of EC, distinct gene expression, and prognosis.

In this study, we validated the validity of Borderline-SMOTE (Fig. 7). For clinical settings where an original imbalance dataset (hysterectomy cases > non-hysterectomy cases) is used as the training set in a prediction model, the results would be more likely skewed towards the larger class group and demonstrate unreliable results for the minority class of patients. In this present study, Borderline-SMOTE was used for data augmentation to reconstitute a balanced training dataset, resulting in a more accurate performance prediction of the minority group of patients who did not undergo hysterectomy. The results (Fig. 7) also indicated that increasing the sample number of patients who did not undergo hysterectomy would also improve the estimation performance.

Further analysis also showed that the proposed method was superior to the other fundamental models (Fig. 8). Compared with cervical and ovarian malignancies, fewer studies on EC have implemented artificial intelligence, and even fewer have focused on fertility-sparing treatments. Our research highlights the promising application of AI in fertility-sparing therapy in patients with EC.

There were some limitations associated with this study that should be recognized. Many machine learning models require complete data, but the original dataset has many missing values, so the handling of missing values is also very important. In future studies, we will explore specific responses associated with the missing values. Second, although many metrics were retrieved from the patients' records, including hemoglobin (HGB), platelets (PLT), prothrombin time (PT), albumin (ALB), CA125, and CA199, based on the expert's clinical experience and Pearson correlation analysis, Age, Para, Pha, Dia, Hyp, OH, MS, NEUT, LYMPH, LYMPH%, ALB and ure were used as the input parameters for the proposed model. However, it was discovered that different parameters had varying degrees of correlation with the need for hysterectomy and that different combinations of parameters might affect the prediction results. Thus, additional strategies should be implemented to improve the model's accuracy by adding more valid parameters, such as combining the patient's CT, pathology and ultrasound reports. The correct combination of patient parameters is essential for developing more accurate estimating models. Further, the dataset created by the Borderline-SMOTE data improvement was well-balanced, and the proposed model improved the estimation accuracy at the extreme weight ranges (hysterectomy versus no hysterectomy). Nonetheless, there were also several

issues with the algorithm. Different strategies to define the DANGER examples and automated adaptive determination of the number of examples in DANGER might be valuable. The combination of this method with under-sampling methods and integrating data mining algorithms are subsequent research directions. Besides, these augmented samples cannot replace real EC patients. Thus, multicenter data collection and analysis would be needed to minimize such issues in future studies.

This study represents the initial phase of fertility-sparing treatments for EC patients. We plan to conduct multivariate time-series data studies of EC patients at different stages of treatment (Patients participating in fertility preservation treatment are treated in a three-month cycle) to examine the progression of patients' conditions and treatments to more accurately assess the risk of disease progression, provide early warning, and guide physicians in proposing the optimal fertility-sparing therapy with the best benefit-risk ratio.

## 5. Conclusions

In this study, machine learning was employed to develop a learning model that could assist doctors in making fertility-sparing treatment decisions (*i.e.*, weighing the need for hysterectomy versus fertility-sparing therapy). Borderline-SMOTE was used to handle imbalance data issues, AdaBoost to classify imbalanced datasets, and MICE to manage missing values. Data analysis showed that the model outperformed standard prediction models such as RF, XGB, SVM and KNN. Thus, the proposed model may help physicians rationalize conserving fertility treatment while minimizing the waste of resources and adverse effects of overtreatment in EC and EAH patients, leading to an optimal benefit-risk ratio. In future work, we aim to collect outpatient, inpatient and screening test data from patients at different time points and perform time series analysis to improve the accuracy of the prediction model and improve decision-making for fertility preservation treatments.

## AUTHOR CONTRIBUTIONS

YS and FY—designed the research study. YS—performed the research. ZL, FY and YS—analyzed the data. FY, LG and WHY—offer clinical advice. FY and YS—wrote the manuscript. All authors read and approved the final manuscript.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

## ACKNOWLEDGMENT

Not applicable.

## FUNDING

This research was funded by the Sichuan Provincial Science and technology program “Risk warning and treatment guidance of retaining reproductive function in patients with endometrial cancer and endometrial atypical hyperplasia by machine learning”, grant number 2021YJ0137, and the project fund of Chengdu Science and Technology Bureau, grant number 2019-YF09-00086-SN.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

- [1] Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, *et al.* Cancer statistics in China, 2015. CA: A Cancer Journal for Clinicians. 2016; 66: 115–132.
- [2] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA: A Cancer Journal for Clinicians. 2017; 67: 7–30.
- [3] Wortman BG, Creutzberg CL, Putter H, Jürgenliemk-Schulz IM, Jobsen JJ, Lutgens LCHW, *et al.* Ten-year results of the PORTEC-2 trial for high-intermediate risk endometrial carcinoma: improving patient selection for adjuvant therapy. British Journal of Cancer. 2018; 119: 1067–1074.
- [4] Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, *et al.* Artificial intelligence in healthcare: past, present and future. Stroke and Vascular Neurology. 2017; 2: 230–243.
- [5] Serj MF, Lavi B, Hoff G, Valls DP. A deep convolutional neural network for lung cancer diagnostic. Available at: <https://arxiv.org/abs/1804.08170> (Accessed: 22 April 2018).
- [6] Lin H, Wei C, Wang G, Chen H, Lin L, Ni M, *et al.* Automated classification of hepatocellular carcinoma differentiation using multiphoton microscopy and deep learning. Journal of Biophotonics. 2019; 12: e201800435.
- [7] Yan K, Wang X, Kim J, Khadra M, Fulham M, Feng D. A propagation-DNN: deep combination learning of multi-level features for MR prostate segmentation. Computer Methods and Programs in Biomedicine. 2019; 170: 11–21.
- [8] Hermessi H, Mourali O, Zagrouba E. Deep feature learning for soft tissue sarcoma classification in MR images *via* transfer learning. Expert Systems with Applications. 2019; 120: 116–127.
- [9] Ghassemi N, Shoeibi A, Rouhani M. Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images. Biomedical Signal Processing and Control. 2020; 57: 101678.
- [10] Pergialiotis V, Pouliakis A, Parthenis C, Damaskou V, Chrelias C, Papantoniou N, *et al.* The utility of artificial neural networks and classification and regression trees for the prediction of endometrial cancer in postmenopausal women. Public Health. 2018; 164: 1–6.
- [11] Vezzoli M, Ravaggi A, Zanotti L, Miscioscia RA, Bignotti E, Ragnoli M, *et al.* RERT: a novel regression tree approach to predict extrauterine disease in endometrial carcinoma patients. Scientific Reports. 2017; 7: 10528.
- [12] Zhou ZH. Machine Learning. 1st ed. Tsinghua University Press: Beijing. 2016.
- [13] Oza NC, Russell SJ. Online bagging and boosting. International Workshop on Artificial Intelligence and Statistics. 2001: R3: 229–236.
- [14] Li X, Liu X, Yu L. Aggregative model-based classifier ensemble for improving land-use/cover classification of Landsat TM images. International Journal of Remote Sensing. 2014; 35: 1481–1495.
- [15] Nordhausen K. Ensemble methods: foundations and algorithms by Zhi-Hua Zhou. International Statistical Review. 2013; 81: 470.
- [16] Freund Y, Schapire RE. ‘Experiments with a new boosting algorithm,’ Machine Learning: Proceedings of the Thirteenth International Conference. Morgan Kaufmann Publishers: San Francisco. 1996.
- [17] Choi Y, Chiu CY, Sontag D. Learning low-dimensional representations of

- medical concepts. AMIA Summits on Translational Science Proceedings. 2016; 2016: 41–50.
- [18] Ying L. Research on text medical record information extraction method [master's thesis]. Hangzhou: Zhejiang University. 2009.
- [19] McKinney W. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. 1st ed. O'Reilly Media, Inc.: Sebastopol. 2012.
- [20] Kalman RE. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*. 1960; 82: 35–45.
- [21] Van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *Journal of Statistical Software*. 2011; 45: 1–67.
- [22] Haibo He, Garcia EA. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*. 2009; 21: 1263–1284.
- [23] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*. 2002, 16: 321–357.
- [24] Fernández A, Garcia S, Herrera F, Chawla NV. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research*. 2018; 61: 863–905.
- [25] Han H, Wang W, Mao B. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. *Lecture Notes in Computer Science*. 2005; 3644: 878–887.
- [26] Jia JP, He XQ, Jin YJ. *Statistics*. 2nd Ed. Chinese Renmin University Press: Beijing. 2004.

**How to cite this article:** Yue Sun, Zhi Li, Li Gao, Wenhan Yuan, Fan Yang. Fertility-sparing treatments decision in patients with endometrial cancer based on machine learning. *European Journal of Gynaecological Oncology*. 2022; 43(5): 91-99. doi: 10.22514/ejgo.2022.046.