Research article

Machine-learning diagnostic models for ovarian tumors

Yuwei Sun [1], Bin Wen [*]

Department of Gynecology, Guangdong Women and Children Hospital, Guangzhou City, Guangdong Province, China

ABSTRACT

Purpose: To create a diagnostic framework for clinical behavior and pathological tissue prognosis in ovarian cancer by using machine-learning (ML) methods based on multiple biomarkers.
Experimental design: Overall, 713 patients with ovarian tumors at Sun Yat Sen Memorial Hospital were randomized into training and test cohorts. Four supervised ML classifiers, namely Support Vector Machine, Random Forest, k-nearest neighbor, and logistic regression were used to derive diagnostic and prognostic information from 10 parameters commonly available from pretreatment peripheral blood tests and age. The best prediction model was selected and validated by comparing the accuracy and the area under the ROC curve of each prediction model and by applying the external data of Guangdong Maternal and Child Health Center.
Results: ML techniques were superior to conventional regression-based analyses in predicting multiple clinical parameters pertaining to ovarian tumor. Ensemble methods combining weak decision trees and RF showed the best reference in diagnosis, especially for malignant ovarian cancer. The values for the highest accuracy and area under the ROC curve for malignant ovarian cancer from benign or borderline ovarian tumors with RF were 99.82 % and 0.86 (micro-average ROC curve), respectively. The greatest accuracy and AUC for the diagnosis of pathological tissue with logistic regression curve were 78.0 % and 0.95 (micro-average ROC curve), respectively. In external validation, the random forest prediction model had an accuracy of 0.789 for applying data from external centers to verify tumor benignity and malignancy, and the logistic regression model had an accuracy of 0.719 for predicting the nature of the tumor.
Conclusions: An ovarian tumor can be diagnosed and characterized before initial treatment via ML systems to provide critical diagnostic and prognostic information. The use of predictive algorithms can facilitate customized treatment options with patient preprocessing stratification.

1. Introduction

Ovarian neoplasms are one of the most common malignancies found in women and are responsible for maximum deaths related to reproductive system tumors. This is likely because ovarian tumors lack specific manifestations at the early stage, and most patients already have advanced-stage cancer at the time of diagnosis and often present with remote metastases [1]. The 2024 National Comprehensive Cancer Network (NCCN) ovarian cancer guideline considers that the main treatment option for ovarian cancer is surgery [2]. The type of ovarian tumor can only be confirmed after histopathological samples have been collected intraoperatively. This means that patients cannot receive treatment before undergoing surgery, which causes significant trauma to patients. And for the

---

* Corresponding author. Guangdong Women and Children Hospital, China.
E-mail address: sums_wenbin@126.com (B. Wen).
[1] Most significant contributions to the study.

patients who are diagnosed with advanced tumor will have to pay more for the treatment. Therefore, it is important for the early detection of ovarian cancer.

Diagnosing ovarian cancer and developing a suitable treatment plan remains a clinical challenge as pathological samples can only be obtained after carrying out invasive surgery. An increasing number of studies are now dedicated to predicting the diagnosis prior to surgery. The advances in machine-learning methodologies and artificial intelligence may likely hold the answers to this clinical predicament. In recent years, machine learning (ML) has found widespread application, the disease prognosis and prediction to identifying potential adverse drug events [3] It is not surprising that decision-support based on ML models is becoming increasingly common in clinical settings. Because of the complexity of ovarian tumor diagnosis, an increasing number of scientists are now trying to apply ML methods to build predictive models for ovarian tumor diagnosis.

As early as 1990, Jacobs et al. used ultrasound scoring, menopausal status, and serum cancer antigen 125 (CA125) concentration to establish a risk of malignancy index (RMI) [4]. The model estimated the probability that pelvic masses would be benign or malignant. It provides the basis for the development of a predictive model for the diagnosis of ovarian tumors. In 2005, Zhangs et al. found that serum concentrations of CA125, transferrin, transthyretin, apolipoprotein α1, and β2-microglobulin were closely associated with ovarian tumors; apolipoprotein α1 and transthyretin showed a down-regulated trend, while a cleaved fragment of the inter-trypsin inhibitor heavy chain H4 showed an up-regulated trend in patients with cancer[5]. Based on this finding in 2009, an ovarian malignancy risk index system (OVA1) was developed with CA125, β2-microglobulin, transferrin, apolipoprotein α1, and transthyretin [5]. The importance of tumor markers for the prediction of ovarian tumors was demonstrated in 2015 when OVA1 was licensed by the FDA for the screening of ovarian tumors [6]. In 2011, Moore et al. proposed the risk of ovarian malignancy algorithm (ROMA) [7] by assessing patients' menopausal status, serological CA125 and human epididymis protein 4 (HE4) levels. ROMA had a sensitivity of 93.8 % and a specificity of 74.9 % for the diagnosis of ovarian cancer in the whole population [8,9]. This study attracted widespread attention, and subsequent studies showed that serological tumor markers HE4 and CA125 [10,11] had a high sensitivity in predicting ovarian tumors, while value of tumor markers in predicting ovarian tumors is gradually recognized. In 2019, Kawakami applied Gradient Boosting Machine, Support Vector Machine, Random Forest, Conditional random forest, Bayesian machine-learning methods, including neural network and elastic network regression, to investigate the superiority of clinical factors in predicting the diagnosis of ovarian tumors, determine the likelihood of ovarian tumor resection, and assess prognosis [12].

The human brain is limited in its ability to integrate and process large amounts of sample data, and artificial intelligence has shown to be more efficient, faster, and more accurate to integrate, analyze, and synthesize complex clinical data. As ML methods are widely used to predict the diagnosis, recurrence, and prognosis of malignant tumors, early diagnosis of ovarian tumors has become a challenge for gynecological oncologists owing to their early symptoms and lack of specific presentation [13–15]. ML is an emerging research area that offers a variety of useful methodologies that can handle large dimensional datasets, and it excels in providing methods that can efficiently and effectively evaluate a large number of variables to construct an accurate model for prediction [16,17]. In 2021, Farinell
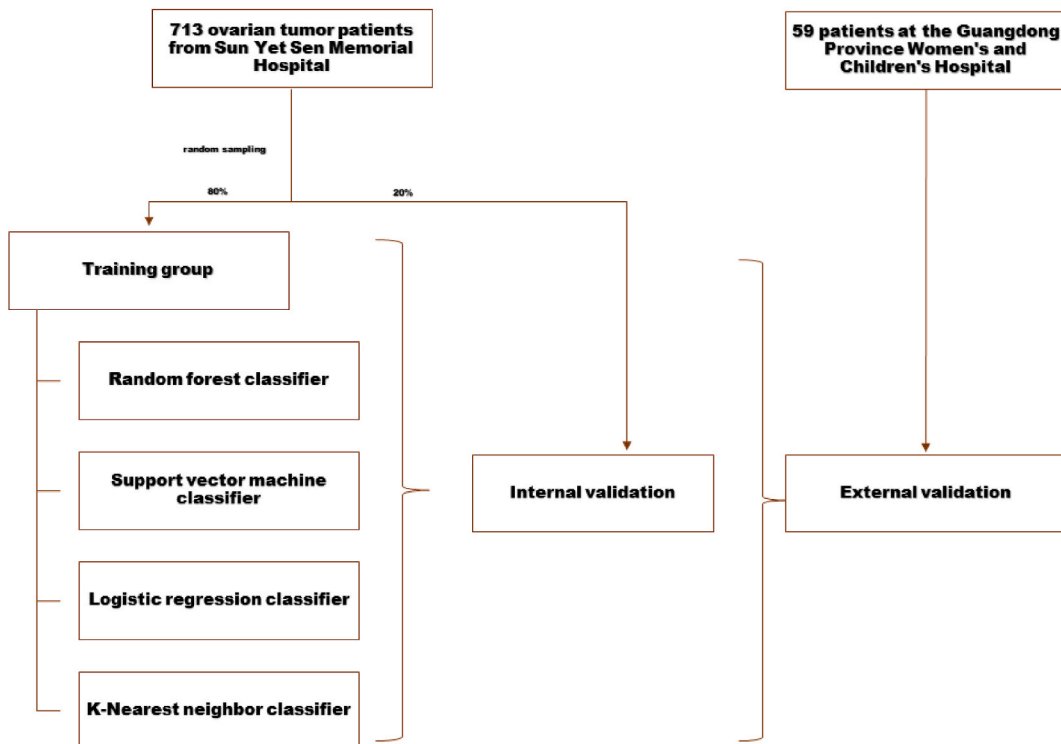

Fig. 1. Flowchart representing the modeling procedure in this study.

et al. applied the public Clinical Proteomic Tumor Analysis Consortium (CPTAC) databases to explore new serological markers for ovarian tumors with the expectation of improving diagnostic sensitivity and specificity for ovarian tumor diagnosis [18]. Despite the study finding that new tumor markers have better diagnostic value for ovarian cancer, clinical validation is still lacking. At the same time, a number of researchers have pursued DNA methylation to identify patients with early-stage ovarian tumors. Li et al. analyzed the DNA methylation profile of malignant ovarian tissues and non-malignant tissues. However, the differences showed poor diagnostic value in early ovarian tumors [19]. The diagnostic accuracy rate is significantly improved when DNA methylation is combined with CA125 levels, indicating the significance of tumor markers.

Early diagnosis of ovarian tumors is a still long way off. Previous studies have focused on a single type of epithelial ovarian tumor to the exclusion of other histological types of ovarian tumors, and the large number of predictors has made it difficult to apply the model in clinical practice.

This study builds on previous research by using tumor markers from 713 patients with ovarian tumor prior to initial surgical treatment to develop multiple ML models and validates the models with data from 69 patients with ovarian tumor from another study center to assess the optimal predictive model for ovarian tumor diagnosis for clinical application. The model was validated using data from 59 patients with ovarian tumor from another study center to assess the optimal predictive model for clinical application.

2. Materials and methods

2.1. Patients and tumor characteristics

The data source of this study came from 713 patients with ovarian tumor from Sun Yet Sen Memorial Hospital between September 25, 2009 and December 22, 2020 and 59 patients with ovarian tumor from Guangdong Province Women and Children's Hospital between September 8, 2020 and May 8, 2022. This study was approved by the ethics committee of Guangdong Women and Children Hospital. Baseline characteristics were retrospectively collected from electronic medical. The data from Sun Yet Sen Memorial Hospital was used as the training data for model development and validated the model to select the best model. The 59 patients from Guangdong Hospital were used as analytical data to validate the model's efficacy (Fig. 1). All patients were diagnosed based on the pathological findings following the surgical procedure. None of the patients with ovarian cancer received preoperative chemotherapy or radiotherapy. The histological type was classified based on the criteria of the World Health Organization (WHO). Patient demographic and histology data are listed in Table 1.

The study concentrated on establishing a diagnostic model for ovarian neoplasms using tumor markers in the clinic. In all, 10 clinical predictors were collected, which included age (when diagnosed with ovarian neoplasm),height, weight, BMI (admission index), and common tumor markers (within 3 months before surgery for primary ovarian neoplasm). The post-operative pathological tumor type was included as an outcome variable.

2.2. Inclusion and exclusion criteria

The inclusion criteria were:

(1) The patient was admitted for examination of a pelvic mass;
(2) The patient received their first surgical treatment at our research center;
(3) Appropriate tumor markers were analyzed/quantified in our hospital prior to surgery;

Table 1
Clinical characteristics of the 713 patients with ovarian tumor.

	Benign ovarian tumor			Borderline ovarian cancer			Malignant ovarian cancer			P-value
N	304			98			311			
Histologic stage										
Epithelial	94			98			263			
Germ cell	197			0			36			
Sex cord stromal	9			0			10			
Rare type	4			0			2			
Clincical characteristics										
	mean	median	IQR	mean	median	IQR	mean	median	IQR	P-value
Age (year)	34.25	30.00	25.00–42.00	39.33	33.5	24.00–54.00	48.08	49.00	40.0–59.0	$<2.2e^{-16}$
BMI (kg/m^2)	21.75	21.41	19.56–23.33	22.28	21.55	19.39–24.19	22.32	22.21	20.07–24.09	0.032
Tumor marker										
AFP (ng/mL)	2.84	1.95	1.34–2.9	3.139	2.05	1.462–2.788	1980.16	2.52	1.76–3.56	0.019
CEA (ng/mL)	1.579	1.40	0.90–2.00	24.56	1.4	0.825–2.227	42.85	1.5	0.8–2.75	0.300
CA125 (U/mL)	28.17	15.10	11.57–22.82	237.47	30.50	19.75–85.53	116.1	303	66.8–12.02	$<2.2e^{-16}$
CA199 (U/mL)	27.96	14.7	7.80–27.70	179.86	15.15	7.825–41.805	524.9	12.6	5.9–37.4	0.31
CA72-4 (U/mL)	4.606	1.6	1.1–3.4	6.44	2.5	1.30–7.38	51	7.9	2.25–37.25	$2.7xe^{-6}$
HE4 (pmol/mL)	47.04	44.83	38.88–50.09	104.6	55.77	47.23–77.75	456.51	1921.4	73.41–547.15	$<2.2e^{-16}$

(4) The post-surgery pathology results were suggestive of a tumor of ovarian origin.

   The exclusion criteria were:

(1) Patients with missing post-operative pathological outcomes or pathological reports that could not clearly identify ovarian tumors or presence of metastatic tumors of any other origin;
(2) Patients with rare pathologic ovarian tumor types (total sample size: <5 cases);
(3) Inclusion of patients with three or more missing key preoperative examination data (including height, weight, and tumor markers)
(4) Patients who refused to apply clinical information to clinical trials.

3. ML analysis

3.1. Missing values [20]

To ensure that the data can be completely used, we retained more complete original data as much as possible during data screening. The miceforest project (miceforest · PyPI)was applied for multiple imputation for missing data. Multiple imputation creates m > 1 complete data sets. Each of these datasets is analyzed by standard analysis software. The m results are pooled into a final point estimate plus standard error by pooling rules ("Rubin's rules"). Fig. 1 illustrates the three main steps in multiple imputation: imputation, analysis, and pooling.

3.2. Feature scaling

Tumor markers should be standardized for smooth data analysis. This technique is to re-scale feature's value with the distribution value between 0 and 1 is useful for the optimization algorithms, such as gradient descent, that are used within ML algorithms that weight inputs.

3.3. Data splitting

The dataset from Sun Yet Sen Memorial Hospital was split into training and validation cohorts with repeated random sampling. Briefly, 80 % data was established as the training group, others are used to build validation. There was no significant difference (P ≥ 0.20). Data from The Guangdong Province Women's and Children's Hospital were used as analytical data for external validation of the model's efficacy.

3.4. Feature selection [21]

There are 11 variables in our research, including histology, age, BMI, and various tumor markers as shown in Table 1. To effectively prepare this high-dimensional data for ML and facilitate clinical work, we employed a data dimensionality reduction strategy, namely feature selection, to help derive a clean and most relevant subset of data to predict the outcome. Feature selection is a major problem in ML, where the purpose is to find the optimal subset of features. Feature selection eliminates irrelevant or redundant features, reduces the number of features, enhances model accuracy, and reduces runtime. This study used sequential backward selection (SBS) to reduce the dimensionality of the initial feature space under the constraint of minimized performance degradation of the classifier to improve the computational efficiency of the model.

3.5. Supervised ML classifiers

In this study, four types of supervised machine learning classifiers were implemented via Python: Support Vector Machine (SVM), Random Forest (RF), k-nearest neighbor (KNN), and Logistic Regression (LR). Classifiers were trained using a training dataset, and their predictive performance was assessed in the test dataset. The correctness of the classification, and the receiver operating characteristic curve (ROC) for validation determined the suitability of the grader.

3.5.1. RF classifier [22]
Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. It is based on two ML techniques: bagging and random feature selection. In bagging, each tree is trained using a bootstrap sample of training data. During the process of training, each tree is grown using a particular bootstrap sample. By virtue of these techniques, the RF classifier avoids overfitting and stratifies samples by considering complex interactions between variables.

3.5.2. SVM classifier
A SVM is a binary classification model whose basic design is a linear classifier defined by maximizing the interval on the feature space. This distinguishes it from a perceptron; the SVM also includes a kernel trick, which makes it essentially a non-linear classifier.

The learning algorithm for SVM is an optimization algorithm for solving convex quadratic programming. SVM has been extensively used for classification, regression, novelty detection tasks, and feature reduction.

3.5.3. LR classifier

LR is a generalized linear model. It assumes that the dependent variable follows a Bernoulli distribution, whereas linear regression assumes that the dependent variable follows a Gauss distribution. It can be argued that LR is theoretically supported by linear regression, but LR introduces a non-linear element through the Sigmoid function, so it can easily handle classification problems. Furthermore, LR algorithms are ML algorithms that are more widely used in different domains. This study focuses on the multi-classification problem by mirroring the soft max function and the output of multiple neurons, mapped into the (0,1) interval, thus applied to the multi-classification problem.

3.5.4. KNN classifier

The KNN rule is a classical, yet powerful technique in non-parametric classification. It is commonly used in a variety of applications such as grouping, feature selection, object detection, and model recognition. Because of its simplicity, effectiveness, and intuitiveness, the KNN classifier tries to identify as many KNNs of a test sample in the functional space of the training data set. It assigns a class label to the test sample via a majority vote among its KNNs. Despite its simplicity, the KNN classifier offers a number of interesting advantages. First, the performance of the KNN classification depends only on a parameter k, which is the number of closest neighbors found. Second, as a non-parametric grader, it does not require prior knowledge of sample likelihood distributions in the classification problem. Finally, when the constraint k/N is close to zero (where N is the total number of training samples), the KNN classification accuracy can converge to twice the accuracy of the optimal Bayesian classifier.

4. Results

4.1. Statistics

Differences in predictors between benign, malignant, and intersectional ovarian tumors were examined by Students' *t*-test. Significant differences have been observed between benign and intersectional ovarian cancer for age, CA125, CEA, CA199, and HE4 (P <
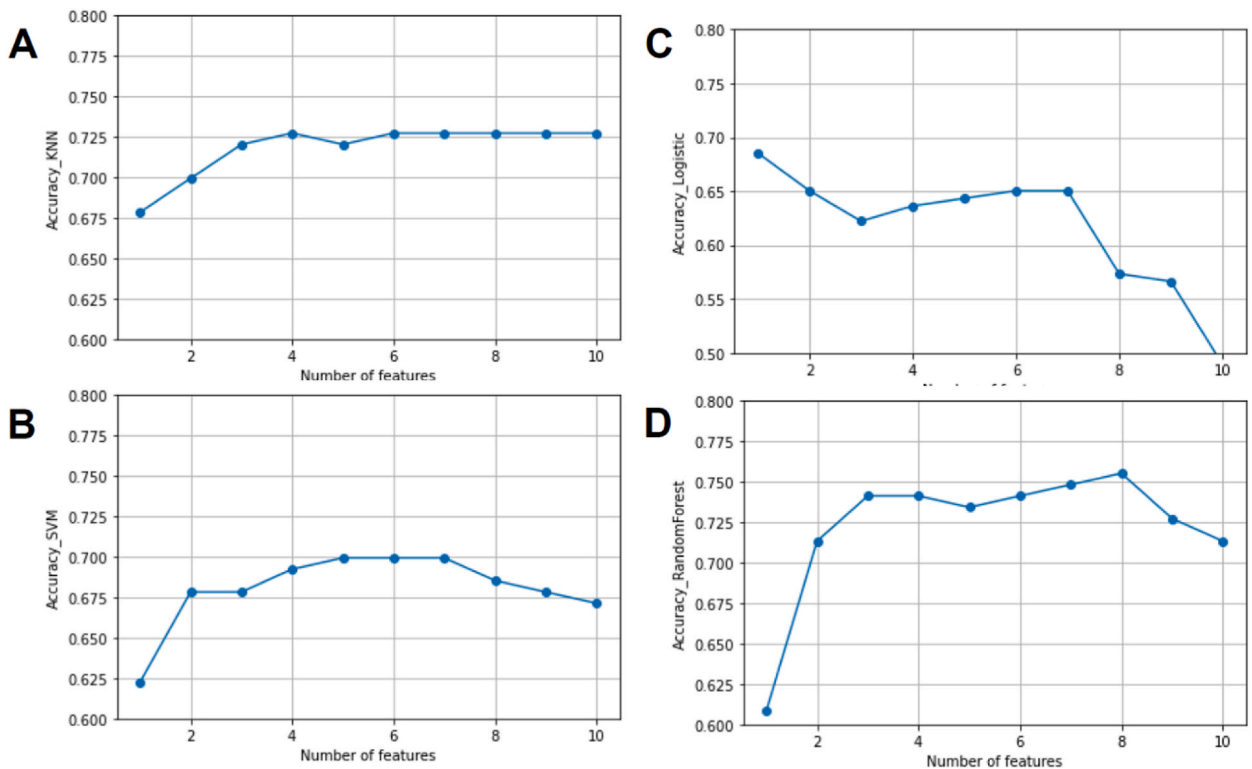

Fig. 2. Different prediction models differ between benign, borderline, and malignant ovarian tumors based on sequence backward selection (SBS). A. The KNN model differ between benign, borderline, and malignant ovarian tumors based on SBS. B. The SVM diagnostic model differ between benign, borderline, and malignant ovarian tumors based on SBS. C. The LR model differ between benign, borderline, and malignant ovarian tumors based on SBS. D. The RF model differ between benign, borderline, and malignant ovarian tumors based on SBS.

0.05). Between benign and malignant ovarian tumors, there were differences in age, CA125, CA72-4, and HE4 (P < 0.05). There were differences in age, height, BMI, CA125, AFP, CA72-4, and HE4 between benign and malignant tumors (P < 0.05). For each pathological histological type, differences in age, BMI, AFP, CA125, CA72-4, and HE4 were observed between epithelial and germ-line ovarian tumors (P < 0.05), but only HE4 was different between epithelial and intersex stromal tumors or types of rare tumors (P < 0.05). In contrast, there were differences in age, BMI, and CA125 between germ cell tumors and gonadal interstitial cell tumor of the mesenchymal cells of the sex cords. For the rare tumor types, differences existed in age, BMI, CA125, and HE4 for germ cell tumors and in age, BMI, and CA199 for ovarian sex cord stromal tumor (P < 0.05).

4.2. Features selection

Sequential backward selection (SBS) was performed on the training dataset with 570 patients' data. The X-axis represents the number of features used in each experiment, while the Y-axis is the accuracy of the classification prediction models. Fig. 2 shows the feature selection of ML to diagnose benign, borderline, and malignant ovarian tumors. During the progress of behavior prediction, the KNN model had the highest accuracy when the model incorporated up to four predictors, namely age, height, CA125, and CA72-4. The accuracy did not improve after adding more predictors. The SVM predictive model incorporated five predictors per SBS, including age, weight, AFP, CA72-4, and HE4. The LR model had the highest prediction accuracy with only one predictor, i.e., HE4. The accuracy decreased when more predictors were added. The RF incorporated eight predictors after SBS, namely height, weight, AFP, CEA, CA125, CA199, CA72-4, and HE4. Fig. 3 presents the outcome of feature selection of models that predict the type of tumor pathology. A KNN prediction model of ovarian tumor pathological tissue type was developed, and the model showed best prediction when six predictors were included in the SBS namely age, weight, BMI, CA125, CEA, and HE4 (Fig. 4). The SVM predictive model for pathological tissue type incorporated six predictors per SBS including height, weight, AFP, CA199, CA72-4, and HE4.

4.3. Difference between benign, borderline, and malignant ovarian tumors

Different prediction models differ in their ability to differentiate between benign, borderline, and malignant ovarian tumors through SBS. As shown in Fig. 2A, the KNN model had the highest accuracy when the model incorporated up to four predictors, namely age, height, CA125, and CA72-4. The accuracy did not improve after adding further predictors. The KNN predictive model had an accuracy of 0.748 for internal validation and 0.748 for external validation in benign, borderline, and malignant ovarian tumors. Its
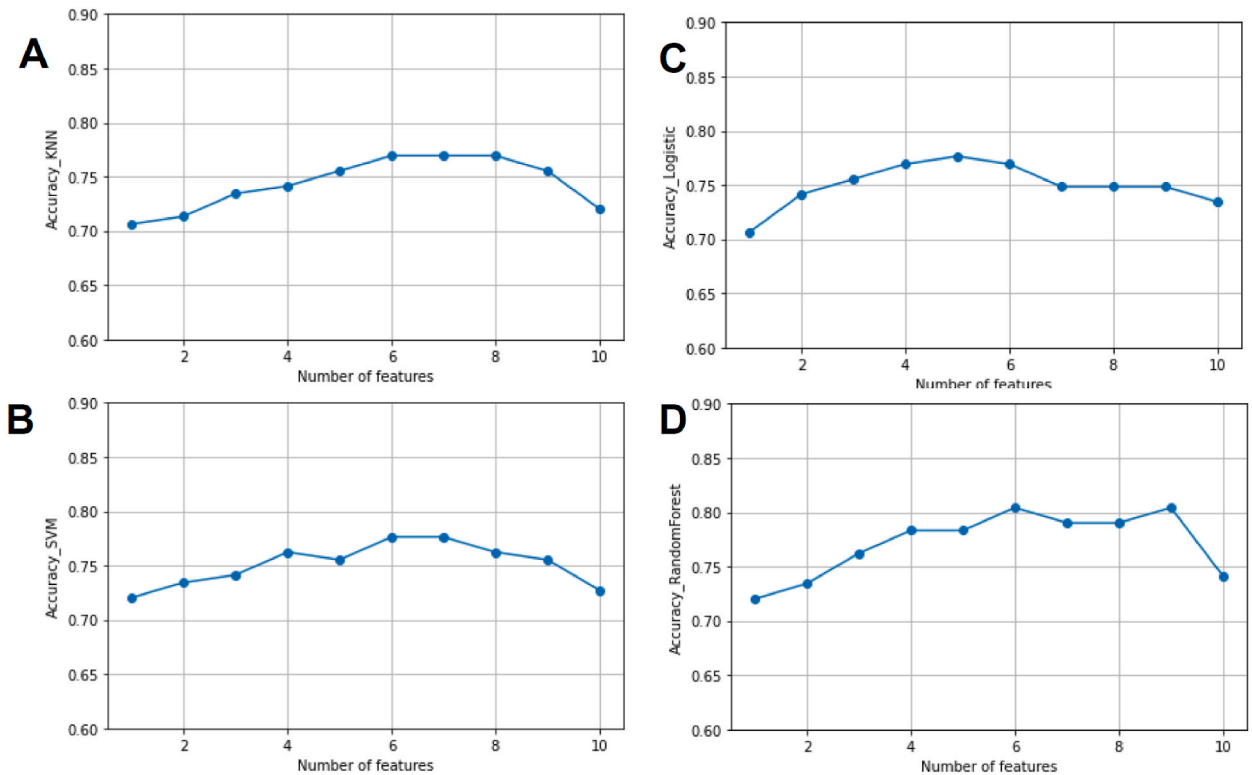

Fig. 3. Different prediction models have different pathologic tissue ovarian tumors based on sequence backward selection (SBS). A. The KNN model different pathologic tissue ovarian tumors based on SBS. B. The SVM diagnostic model different pathologic tissue ovarian tumors based on SBS. C. The LR model different pathologic tissue ovarian tumors based on SBS. D. The RF model different pathologic tissue ovarian tumors based on SBS.
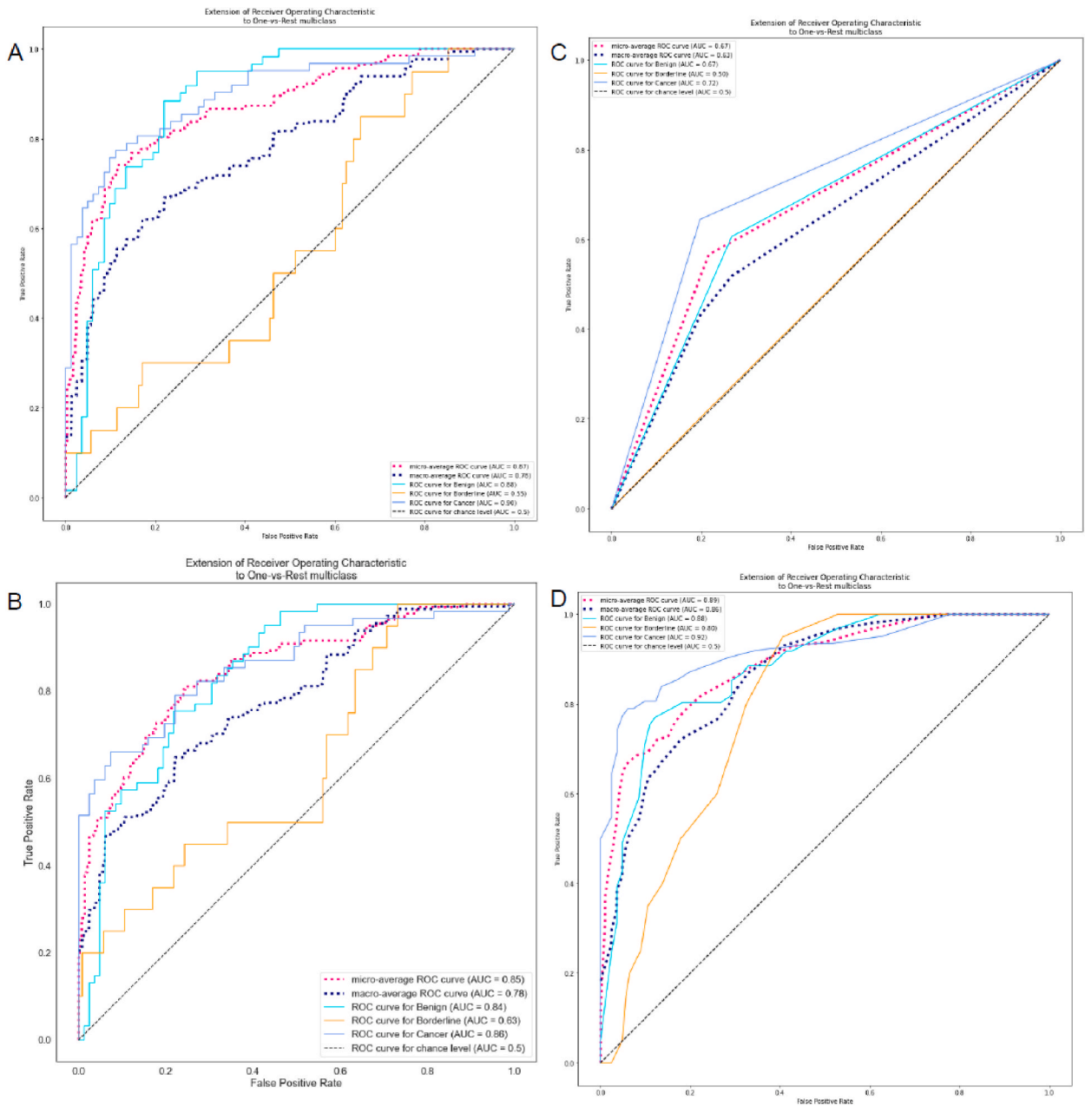
Fig. 4. Difference between benign, borderline, and malignant ovarian tumors. A. The KNN diagnostic model was well predicted for both benign and malignant tumors with AUCs of 0.88, 0.55, and 0.90. B. The SVM diagnostic model was well predicted for both benign and malignant tumors with AUCs of 0.84 and 0.85. C. The LR model was predicted benign, borderline, and malignant tumors with AUCs of 0.67, 0.50, and 0.72. D. The RF model was predicted benign, borderline, and malignant ovarian tumors with AUCs of 0.86, 0.88, and 0.80.

ROC to predict tumor behavior is shown in Fig. 4A, with AUCs of 0.88, 0.55, and 0.90 for benign, borderline, and malignant tumors, respectively. The SVM diagnostic model incorporated five predictors per SBS including age, weight, AFP, CA72-4, and HE4 (Fig. 2B). The SVM diagnostic model had an internally validated prediction accuracy of 0.696 for benign, borderline, and malignant ovarian tumors and an externally validated prediction accuracy of 0.692. The SVM diagnostic model was well predicted for both benign and malignant tumors with AUCs of 0.84 and 0.85, respectively (Fig. 4B), but the diagnostic ability for borderline ovarian tumors was poor at 0.63. Within the LR model, the SBS process is as described in Fig. 2C. The highest prediction accuracy was seen with only one predictor, namely HE4. The accuracy reduced when other predictors were added. The AUCs of the LR model were 0.67, 0.50, and 0.72 for benign, borderline, and malignant tumors, respectively (Fig. 4C). RF incorporated eight predictors after SBS (Fig. 2D), namely height, weight, AFP, CEA, CA125, CA199, CA72-4, and HE4. The RF model has an internal validation accuracy of 0.99 and an external

validation accuracy of 0.727. Its ROC is shown in Fig. 4D, and AUCs for benign, borderline, and malignant ovarian tumors were 0.86, 0.88, and 0.80, respectively. The LR model applied age, weight, CA19, CA125, CA72-4, and the RF model was built based on age, BMI, AFP, CEA, CA125, and CA72-4.

4.4. Diagnosis of histologic types of ovarian tumors with classifiers

A KNN prediction model of ovarian tumor pathological tissue type was developed and the model showed best prediction when the six predictors included in the SBS were age, weight, BMI, CA125, CEA, and HE4(Fig. 3A), with an accuracy of 0.79 for internal validation and 0.76 for external validation. The AUCs of epithelial, germ cell, interstitial sex cord, and rare types of ovarian tumors
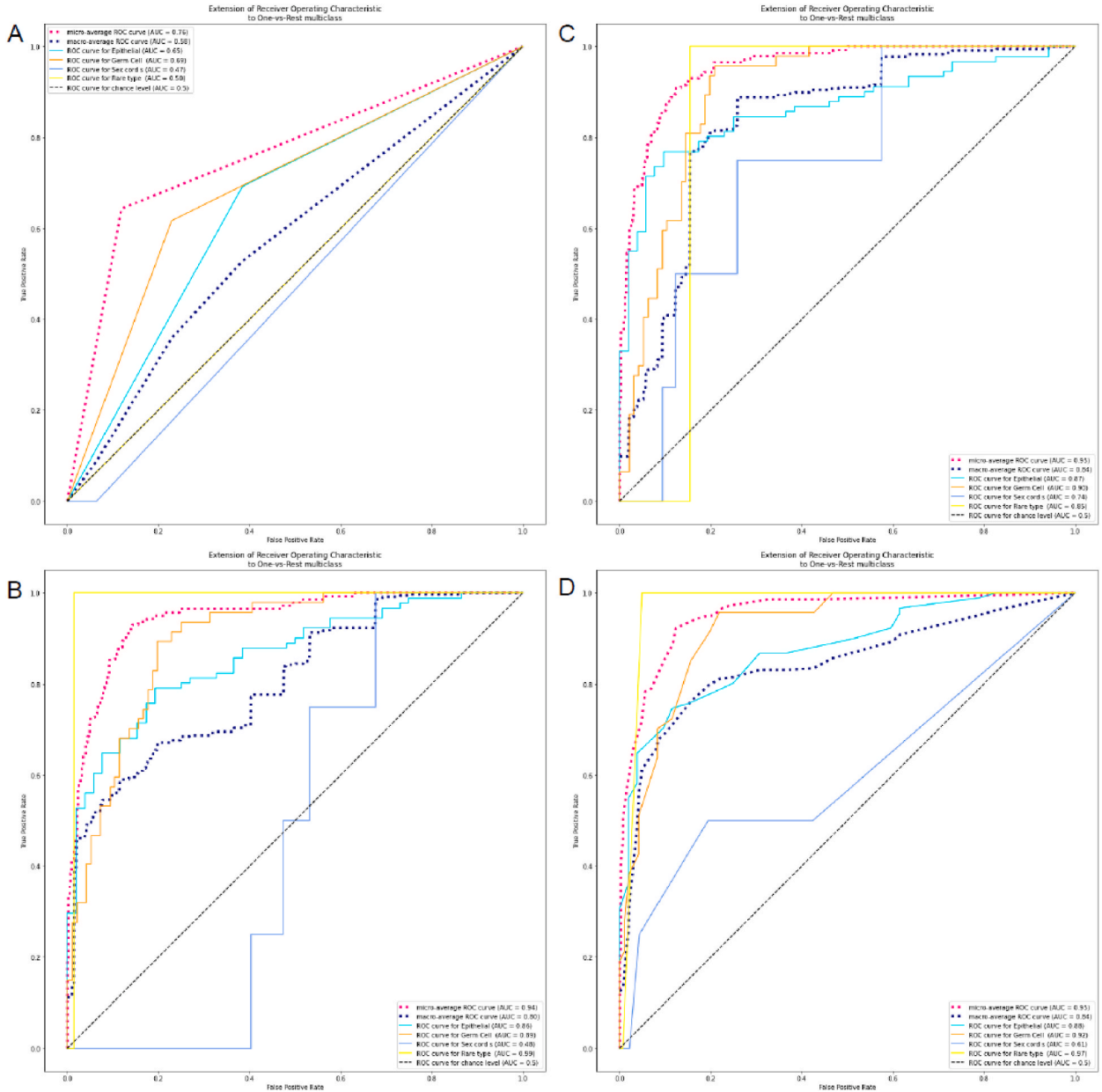

Fig. 5. Diagnosis of histologic types ovarian tumors using the ML model. A. The KNN diagnostic model was predicted germ cell, interstitial sex cord, and rare types of ovarian tumors with AUCs of 0.65, 0.69, 0.47, and 0.50. B. The SVM diagnostic model was predicted germ cell, interstitial sex cord, and rare types of ovarian tumors with AUCs of 0.86, 0.89, 0.99, and 0.48. C. The LR model was predicted germ cell, interstitial sex cord, and rare types of ovarian tumors with AUCs of 0.87, 0.90, 0.74, and 0.85. D. The RF model was predicted germ cell, interstitial sex cord, and rare types of ovarian tumors with AUCs of 0.8, 0.,2, 0,61, and 0.97.

were 0.65, 0.69, 0.47, and 0.50 (Fig. 5A), respectively. The SVM predictive model for pathological tissue type incorporated six predictors per SBS (Fig. 3B), including height, weight, AFP, CA199, CA72-4, and HE4. The SVM model was excellent in identification of epithelial, germ cell, and rare type tumors, but underperformed in interstitial sex cord ovarian tumor, with an accuracy of 0.65 for internal validation and 0.68 for external validation. The ROC of the SVM model is as shown in Fig. 5B. The AUCs of epithelial, germ cell, interstitial sex cord, and rare types of ovarian tumors were 0.86, 0.89, 0.99, and 0.48, respectively. The LR model showed the best outcome in the diagnosis of pathological tissue type. The model applied age, weight, CA199, CA125, and CA72-4 in the SBS (Fig. 3C). The accuracy of internal validation was 0.747 and that of external validation was 0.783. The ROC of the LR model is as shown in Fig. 5C. The AUCs of epithelial, germ cell, interstitial sex cord, and rare types of ovarian tumors were 0.87, 0.90, 0.74, and 0.85, respectively. Surprisingly, the RF model did not perform well in diagnosing pathological tissue types, especially for interstitial sex cord ovarian tumor. A RF model built by SBS incorporating age, BMI, AFP, CEA, CA125, and CA72-4 (Fig. 3D) had an internally validated accuracy of 0.99 and an externally validated accuracy of 0.72 to diagnose ovarian tumor pathological tissues. The ROC of the RF model is as shown in Fig. 5D. The AUCs of epithelial, germ cell, interstitial sex cord, and rare types of ovarian tumors were 0.8, 0.,2, 0,61, and 0.97, respectively.

4.5. External data validation model

To further explore the potential for use of ML models in clinical work, this study innovatively validated the diagnostic predictive ability of ML models by introducing multicentric data. This study applied clinical data of ovarian tumor patients from 2020 to 2022 at Guangdong Women's and Children's Hospital to validate the predictive effect of the ML model. After applying the inclusion and exclusion criteria, 58 patients were finally included to validate the model predicted efficiency. These included eight cases of benign tumors, seven of borderline tumors, and 43 malignant tumors. There were 42 epithelial tumors, five germ cell tumors, nine sex-cord mesenchymal cell tumors, and two rare types of tumors. Due to the lack of preoperative CA72-4 data for patients in Guangdong Maternal and Child Hospital, the mean CA72-4 value for the training group was used instead. Finally, as shown in Table 2, the accuracy of diagnosis of benign, borderline, and malignant ovarian tumors from external data by the RF prediction model was 0.789. The accuracy of applying LR models to predict diagnosis based on pathologic tissues of epithelial cells, germ cells, sex-cord mesenchymal cell, and rare types of ovarian tumors from external data was 0.719 (Table 3).

5. Discussion

During the past decade, with the greatest advancements seen in clinical medicine, accurate prediction of the diagnosis of a disease has become the focus of significant research studies to alleviate challenges for doctors. There is now a considerably great deal of new technology available and significant cancer data has been collected and is available for the medical research community. With the development of artificial intelligence, ML has become a popular tool for medical researchers. These techniques allow us to discover and identify models and their inter-relationships, based on complex data sets, while being able to effectively predict the future outcomes of a type of cancer. Clearly, this innovative approach is an important tool in the area of precision medicine that can facilitate the choice of optimal treatment strategies [23]. By manipulating many factors in the data at once, it may lead to a better understanding of the complex mechanisms underlying cancer genesis and progression.

Ovarian tumors are among the most complex pathological types known, and treatment options vary according to the pathological type of tumor [24]. In practice, however, the inability to predict the nature of the tumor often leads to inadequate preparation for surgery or misdiagnosis of the pathological type, which can cause significant harm to the patient. Early diagnosis of ovarian tumors is therefore paramount. This study launched the comparison of multiple supervised learning algorithms to identify the approach with the most favorable performance. We believe we have innovatively introduced multi-center data to validate the prediction effect of ML models. The importance of HE4 and CA125 in determining the nature of tumors has been highlighted in previous predictive models for ovarian tumors [8,10,11,25]. Another study included 32 clinical datas to determine the nature of ovarian tumors [12]. However, past studies have neglected practical clinical scenarios where too many predictors could over-adapt a predictive model and render it unnecessary for general application. These studies were also limited by their single-center design, which has limited relevance for the treatment of ovarian tumors.

In this study, we constructed a diagnostic model based on previous studies of ovarian tumor prediction models, using the most diagnostically useful tumor markers in ovarian tumors as predictors. In the analysis of benign, borderline, and malignant data, we found that the SVM model performed poorly, possibly because of the poor performance of characteristic tumor markers for different tumor types. This may be because the SVM model is mainly used for linearly separable classification; whereas, different types of tumor cells secrete different characteristic tumor markers. On the other hand, KNN and LR are unique in their best performance in the diagnosis of benign and malignant tumors, but they do not give good results in the prediction of borderline tumors. Unfortunately, even the best-performing RF model failed to identify borderline tumors in the multicenter validation, which may be because of the

Table 2

The accuracy of predicting benign, borderline, and malignant ovarian tumors from external data by the RF model.

	Benign tumors	Borderline tumors	Malignant tumors	N
True	8	7	43	58
Pred_randomforest	6	0	52	58

Table 3
The accuracy of applying LR models to predict diagnosis based on pathologic tissues.

	Epithelial tumors	Germ cell tumors	Sex-cord mesenchymal cell	Rare types
True	42	5	9	2
Pred_Logistic regression	47	11	0	0

small number of borderline tumors in the training group and the fact that all borderline tumors are epithelial in nature.

To establish diagnostic models for pathological types, the non-parametric predictive models represented by KNN showed decadent performance in the determination of pathological tumor tissue types. However, other parametric example model types had better diagnostic predictive ability for both epithelial ovarian tumors and germ cell tumors, with AUCs of 0.86, 0.88, and 0.87 for epithelial tumors and 0.89, 0.92, and 0.90 for germ cell tumors for SVM, LR, and RF models, respectively. Interestingly, however, even the RF model based on the stimulation method lacked the precision to distinguish between interstitial tumor and sex cords. Likewise, even the LR model that was most successful in diagnosing the type of pathology in the single center data lacked judgment of gonadal mesenchymal cells when validated against multi-center data. The reason for the poor diagnostic prediction of borderline ovarian tumors and sex cord mesenchymal stromal cell tumors is the lack of sex cord mesenchymal cell data in the training group. It can be seen that the ML model requires very strict training group data and needs to be further improved with large samples and multi-center data to make up for this shortcoming.

Because the approach used in this study did not incorporate information from imaging studies or pre-treatment biopsies, the ability to accurately predict clinical behavior and treatment results prior to the procedure were limited. Therefore, further validation efforts are required by increasing the number of input variables based on the robust ML approach to spill over into a larger independent cohort. As tumors develop over time, the signaling between the tumor and its microenvironment, composed of fibroblasts, infiltrating immune cells, and endothelial cells also evolves. It is believed that further future studies will fill this gap.

Although the excellent performance of ML models in ovarian tumor diagnosis relies on a large amount of training data, this study only collected data on ovarian tumors from two centers, and there is less data on junctional tumors and gonadal interstitial tumors, which leads to poor performance of the prediction model for the diagnosis of this type of tumor. We are also concerned that ovarian tumors are often detected by imaging, such as ultrasound, computed tomography, and magnetic resonance imaging; thus, imaging is also crucial for the diagnosis of ovarian tumors. A recent study in 2022 by Gao et al. used the ultrasound imaging data of 3755 patients with ovarian tumors to automatically evaluate ultrasound images by developing a deep convolutional neural network (DCNN) model [26] and facilitated the diagnosis of ovarian cancer more accurately than existing methods. Based on this study, larger multicenter studies on ovarian cancer are underway and are expected to incorporate additional predictors of diagnostic significance such as imaging reports. Machine learning, which is a data-driven science, has opened up a promising path towards an evolving healthcare system that is filled with exciting opportunities for precision oncology. Despite this, the technique still has some limitations. The training data used to train ML models is what makes them as good as they are. The model cannot function properly without the standardization and realism of the train groups. The principal architecture of current models is designed with performance evaluated by accuracy, AUC, sensitivity, specificity, the expense of interpretability, and transparency. Although artificial intelligence frequently exceeds traditional interpretable models, the perceived lack of transparency and lack of quantified uncertainty significantly undermines the belief in artificial intelligence [27].

In summary, despite some shortcomings, AI-based algorithms can be effective tools to diagnose and evaluate patients with ovarian tumor before initiating intervention, as demonstrated in this study.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

Ethics declarations

The study was reviewed and approved by the ethics committee of Guangdong Women and Children Hospital, with the approval number: 202301384. All participants/patients (or their proxies/legal guardians) provided informed consent for the publication of their anonymized case details and images.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author [Bin Wen].

CRediT authorship contribution statement

Yuwei Sun: Writing – original draft, Software, Resources, Methodology, Funding acquisition. **Bin Wen:** Writing – review & editing, Funding acquisition, Data curation, 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.

Acknowledgments

References

[1] D.K. Armstrong, R.D. Alvarez, J.N. Bakkum-Gamez, L. Barroilhet, K. Behbakht, A. Berchuck, L.M. Chen, M. Cristea, M. DeRosa, E.L. Eisenhauer, D. M. Gershenson, H.J. Gray, R. Grisham, A. Hakam, A. Jain, A. Karam, G.E. Konecny, C.A. Leath, J. Liu, H. Mahdi, L. Martin, D. Matei, M. McHale, K. McLean, D. S. Miller, D.M. O'Malley, S. Percac-Lima, E. Ratner, S.W. Remmenga, R. Vargas, T.L. Werner, E. Zsiros, J.L. Burns, A.M. Engh, Ovarian cancer version 2.2020, NCCN clinical practice guidelines in oncology, J. Natl. Compr. Canc. Netw. 19 (2021 Feb 2) 191–226, https://doi.org/10.6004/jnccn.2021.0007.
[2] E.M. Wilson, R.N. Eskander, P.S. Binder, Recent therapeutic advances in gynecologic oncology: a review, Cancers 16 (4) (2024 Feb 13) 770, https://doi.org/ 10.3390/cancers16040770. PMID: 38398161; PMCID: PMC10887183.
[3] Y. Jiang, C. Wang, S. Zhou, Artificial intelligence-based risk stratification, accurate diagnosis and treatment prediction in gynecologic oncology, Semin. Cancer Biol. 96 (2023) 82–99, https://doi.org/10.1016/j.semcancer.2023.09.005.
[4] I. Jacobs, D. Oram, J. Fairbanks, J. Turner, C. Frost, J.G. Grudzinskas, A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer, Br. J. Obstet. Gynaecol. 97 (1990) 922–929, https://doi.org/10.1111/J.1471-0528.1990.TB02448.X.
[5] R.G. Moore, D.S. McMeekin, A.K. Brown, P. DiSilvestro, M.C. Miller, W.J. Allard, W. Gajewski, R. Kurman, R.C. Bast, S.J. Skates, A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass, Gynecol. Oncol. 112 (2009) 40–46, https://doi.org/ 10.1016/j.ygyno.2008.08.031.
[6] D.G. Grenache, K.A. Heichman, T.L. Werner, Z. Vucetic, Clinical performance of two multi-marker blood tests for predicting malignancy in women with an adnexal mass, Clin. Chim. Acta 438 (2015) 358–363, https://doi.org/10.1016/J.CCA.2014.09.028.
[7] R.G. Moore, M.C. Miller, P. Disilvestro, L.M. Landrum, W. Gajewski, J.J. Ball, S.J. Skates, Evaluation of the diagnostic accuracy of the risk of ovarian malignancy algorithm in women with a pelvic mass, Obstet. Gynecol. 118 (2011) 280–288, https://doi.org/10.1097/AOG.0b013e318224fce2.
[8] L. Zhang, Y. Chen, K. Wang, Comparison of CA125, HE4, and ROMA index for ovarian cancer diagnosis, Curr. Probl. Cancer 43 (2019) 135–144, https://doi.org/ 10.1016/j.currproblcancer.2018.06.001.
[9] M.T. Sandri, F. Bottari, D. Franchi, S. Boveri, M. Candiani, S. Ronzoni, M. Peiretti, D. Radice, R. Passerini, M. Sideri, Comparison of HE4, CA125 and ROMA algorithm in women with a pelvic mass: correlation with pathological outcome, Gynecol. Oncol. 128 (2013) 233–238, https://doi.org/10.1016/j. ygyno.2012.11.026.
[10] P. Leandersson, A. Åkesson, I. Hedenfalk, S. Malander, C. Borgfeldt, A multiplex biomarker assay improves the diagnostic performance of HE4 and CA125 in ovarian tumor patients, PLoS One 15 (2020 October 19) e0240418, https://doi.org/10.1371/journal.pone.0240418.
[11] M. Lycke, B. Kristjansdottir, K. Sundfeldt, A multicenter clinical trial validating the performance of HE4, CA125, risk of ovarian malignancy algorithm and risk of malignancy index, Gynecol. Oncol. 151 (2018) 159–165, https://doi.org/10.1016/j.ygyno.2018.08.025.
[12] E. Kawakami, J. Tabata, N. Yanaihara, T. Ishikawa, K. Koseki, Y. Iida, M. Saito, H. Komazaki, J.S. Shapiro, C. Goto, Y. Akiyama, R. Saito, M. Saito, H. Takano, K. Yamada, A. Okamoto, Application of artificial intelligence for preoperative diagnostic and prognostic prediction in epithelial ovarian cancer based on blood biomarkers, Clin. Cancer Res. 25 (2019) 3006–3015, https://doi.org/10.1158/1078-0432.CCR-18-3378.
[13] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction, Comput. Struct. Biotechnol. J. 13 (2015) 8–17, https://doi.org/10.1016/j.csbj.2014.11.005.
[14] R.D. Riley, Correction to: minimum sample size for developing a multivariable prediction model: part II-binary and time-to-event outcomes, Stat. Med. 38 (2019) 5672, https://doi.org/10.1002/sim.8409.
[15] M. Lu, Z. Fan, B. Xu, L. Chen, X. Zheng, J. Li, T. Znati, Q. Mi, J. Jiang, Using machine learning to predict ovarian cancer, Int. J. Med. Inf. 141 (2020) 104195, https://doi.org/10.1016/j.ijmedinf.2020.104195.
[16] F. Farinella, M. Merone, L. Bacco, A. Capirchio, M. Ciccozzi, D. Caligiore, Machine Learning analysis of high-grade serous ovarian cancer proteomic dataset reveals novel candidate biomarkers, Sci. Rep. 12 (2022) 3041, https://doi.org/10.1038/s41598-022-06788-2.
[17] J.A. Cruz, D.S. Wishart, Applications of machine learning in cancer prediction and prognosis, Cancer Inf. 2 (2007) 59–77, https://doi.org/10.1177/ 117693510600200030.
[18] G.A. Atallah, N.H. Abd Aziz, C.K. Teik, M.N. Shafiee, N.C. Kampan, New predictive biomarkers for ovarian cancer, Diagnostics 11 (2021), https://doi.org/ 10.3390/DIAGNOSTICS11030465.
[19] N. Li, X. Zhu, W. Nian, Y. Li, Y. Sun, G. Yuan, Z. Zhang, W. Yang, J. Xu, A. Lizaso, B. Li, Z. Zhang, L. Wu, Y. Zhang, Blood-based DNA methylation profiling for the detection of ovarian cancer, Gynecol. Oncol. 167 (2022) 295–305, https://doi.org/10.1016/j.ygyno.2022.07.008.
[20] R.W. Krause, M. Huisman, C. Steglich, T. Snijders, Missing data in cross-sectional networks – an extensive comparison of missing data treatment methods, Soc. Netw. 62 (2020) 99–112, https://doi.org/10.1016/J.SOCNET.2020.02.004.
[21] C. Tang, T. Gao, Y. Li, B. Chen, EEG channel selection based on sequential backward floating search for motor imagery classification, Front. Neurosci. 16 (2022) 1045851, https://doi.org/10.3389/FNINS.2022.1045851.
[22] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[23] R.G. Moore, A. Blackman, M.C. Miller, K. Robison, P.A. DiSilvestro, E.E. Eklund, R. Strongin, G. Messerlian, Multiple biomarker algorithms to predict epithelial ovarian cancer in women with a pelvic mass: can additional makers improve performance? Gynecol. Oncol. 154 (2019 July) 150–155, https://doi.org/10.1016/ j.ygyno.2019.04.006.
[24] A.A. Ahmed, A.M. Abdou, Diagnostic accuracy of CA125 and HE4 in ovarian carcinoma patients and the effect of confounders on their serum levels, Curr. Probl. Cancer 43 (2019) 450–460, https://doi.org/10.1016/J.CURRPROBLCANCER.2018.12.004.
[25] Y. Gao, S. Zeng, X. Xu, H. Li, S. Yao, K. Song, X. Li, L. Chen, J. Tang, H. Xing, Z. Yu, Q. Zhang, S. Zeng, C. Yi, H. Xie, X. Xiong, G. Cai, Z. Wang, Y. Wu, J. Chi, X. Jiao, Y. Qin, X. Mao, Y. Chen, X. Jin, Q. Mo, P. Chen, Y. Huang, Y. Shi, J. Wang, Y. Zhou, S. Ding, S. Zhu, X. Liu, X. Dong, L. Cheng, L. Zhu, H. Cheng, L. Cha, Y. Hao, C. Jin, L. Zhang, P. Zhou, M. Sun, Q. Xu, K. Chen, Z. Gao, X. Zhang, Y. Ma, Y. Liu, L. Xiao, L. Xu, L. Peng, Z. Hao, M. Yang, Y. Wang, H. Ou, Y. Jia, L. Tian, W. Zhang, P. Jin, X. Tian, L. Huang, Z. Wang, J. Liu, T. Fang, D. Yan, H. Cao, J. Ma, X. Li, X. Zheng, H. Lou, C. Song, R. Li, S. Wang, W. Li, X. Zheng, J. Chen, G. Li, R. Chen, C. Xu, R. Yu, J. Wang, S. Xu, B. Kong, X. Xie, D. Ma, Q. Gao, Deep learning-enabled pelvic ultrasound images for accurate diagnosis of ovarian cancer in China: a retrospective, multicentre, diagnostic study, Lancet Digit. Health 4 (2022 March) e179–e187, https://doi.org/10.1016/S2589-7500(21) 00278-8.
[26] B.H. Kann, A. Hosny, H.J.W.L. Aerts, Artificial intelligence for clinical oncology, Cancer Cell 39 (2021) 916–927, https://doi.org/10.1016/j.ccell.2021.04.002.