



# The application of machine learning for predicting recurrence in patients with early-stage endometrial cancer: a pilot study

Munetoshi Akazawa, MD<sup>1</sup>, Kazunori Hashimoto, PhD<sup>1</sup>, Katsuhiko Noda, DE<sup>2</sup>, Kaname Yoshida, DE<sup>2</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, Tokyo Women's Medical University Medical Center East, <sup>2</sup>SIOS Technology Inc., Tokyo, Japan

## Objective

Most women with early stage endometrial cancer have a favorable prognosis. However, there is a subset of patients who develop recurrence. In addition to the pathological stage, clinical and therapeutic factors affect the probability of recurrence. Machine learning is a subtype of artificial intelligence that is considered effective for predictive tasks. We tried to predict recurrence in early stage endometrial cancer using machine learning methods based on clinical data.

## Methods

We enrolled 75 patients with early stage endometrial cancer (International Federation of Gynecology and Obstetrics stage I or II) who had received surgical treatment at our institute. A total of 5 machine learning classifiers were used, including support vector machine (SVM), random forest (RF), decision tree (DT), logistic regression (LR), and boosted tree, to predict the recurrence based on 16 parameters (age, body mass index, gravity/parity, hypertension/diabetic, stage, histological type, grade, surgical content and adjuvant chemotherapy). We analyzed the classification accuracy and the area under the curve (AUC).

## Results

The highest accuracy was 0.82 for SVM, followed by 0.77 for RF, 0.74 for LR, 0.66 for DT, and 0.66 for boosted trees. The highest AUC was 0.53 for LR, followed by 0.52 for boosted trees, 0.48 for DT, and 0.47 for RF. Therefore, the best predictive model for this analysis was LR.

## Conclusion

The performance of the machine learning classifiers was not optimal owing to the small size of the dataset. The use of a machine learning model made it possible to predict recurrence in early stage endometrial cancer.

**Keywords:** Machine learning; Recurrence; Endometrial cancer; Probability learning

## Introduction

Endometrial cancers are the most common gynecologic malignancies in developed countries and the second most common in developing countries [1]. The vast majority of women with endometrial cancer are diagnosed with early stage tumors that are associated with a good prognosis. However, a subgroup of women with early stage endometrial cancer face recurrence and are at an increased risk of death.

Prediction of prognosis is one of the biggest problems in cancer therapy. Accurate prediction of recurrence or survival after treatment could help guide additional treatment strate-

Received: 2020.08.18. Revised: 2020.10.22. Accepted: 2020.11.03.

Corresponding author: Munetoshi Akazawa, MD

Department of Obstetrics and Gynecology, Tokyo Women's Medical University Medical Center East, 2 Chome-1-10 Nishiogu, Arakawa-ku, Tokyo 116-8567, Japan

E-mail: navirez@yahoo.co.jp

<https://orcid.org/0000-0003-3378-8546>

Articles published in Obstet Gynecol Sci are open-access, distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2020 Korean Society of Obstetrics and Gynecology

gies and follow-up. Traditionally, gynecologists have relied mainly on the final International Federation of Gynecology and Obstetrics (FIGO) stage to estimate oncologic outcomes. However, it is well known that other factors such as age, histology, final grade, and additional treatment, such as chemotherapy or radiotherapy could play equally important roles in the prognosis [1]. For predictions consisting of multiple variables, many previous studies have used nomograms [2,3]. However because the relationship among the variables is linear in the nomogram, the algorithm is considered insufficient for the expression of real relationships in the predictive task.

Artificial intelligence (AI) is considered to be a new and possible predictive technique. In particular, deep learning, a subtype of AI, has been developed recently and has shown good performance in many realistic predictive tasks. In medicine, the application of AI has progressed mainly to image-recognition tasks, such as the diagnosis of imaging data. Previously, several reports have demonstrated the excellent accuracy of AI in diagnosis, such as for head computed tomography (CT) scans [4], skin cancer [5], and retinopathy in diabetic patients [6]. The application of AI would spread more widely in medicine and be effective not only for the diagnostic prediction, but also for therapeutic prediction, such as the prediction of a prognosis.

We tried to predict recurrence in early stage endometrial cancer, based on clinical data, using machine learning methods.

## Materials and methods

### 1. The dataset

A total of 75 patients with early stage endometrial cancers were enrolled, including 60 in stage 1 and 15 in stage 2 of endometrial cancer. All patients underwent surgeries in Tokyo Women's Medical University Medical Center East between December 2013 and January 2019 and received a pathological diagnosis. The inclusion criterion for the study was a case of endometrial cancer that had undergone surgical staging in our institute and confirmed the FIGO stage. After the primary treatment, we checked the patient every 3–6 months using gynecologic examinations, tumor markers in blood examinations, and CT examinations. In cases of recurrence, all cases of recurrence within 5 years of primary treatment were included in this analysis, regardless of the length of follow-up

periods. In the case of non-recurrence, the case was included when the 5-year disease-free survival was confirmed after a follow-up over 5 years. Exclusion criteria was for the case of a patient who could not receive 5 years of follow-up in our institute or had insufficient preoperative clinical data.

Each patient in this dataset had 16 features. The features included the following: 1) age (years), 2) gravidity, 3) parity, 4) body mass index ( $\text{kg}/\text{m}^2$ ), 5) FIGO stage, 6) grade, 7) hypertension, 8) diabetes, 9) carbohydrate antigen 125 (CA125; U/mL), 10) (carbohydrate antigen 19-9 (CA19-9; U/mL), 11) carcinoembryonic antigen (CEA) (ng/mL), 12) approach to hysterectomy, 13) pelvic lymphadenectomy, 14) para-aortic lymphadenectomy, 15) omentectomy, and 16) the number of postoperative chemotherapy. FIGO stage was defined by FIGO 1988 and the grade was the pathological grade. Tumor markers (CA125/CA19-9/CEA) were examined preoperatively in all cases. Hysterectomies were categorized into 2 approaches: radical hysterectomies or simple hysterectomies. The operating team decided whether or not they performed a pelvic lymphadenectomy, para-aortic lymphadenectomy, or omentectomy, after considering the intraoperative pathological results. The chemotherapy regimen was TC therapy (paclitaxel plus carboplatin). The missing values were noted in the categories of tumor markers and replaced with the median values.

### 2. Model of machine learning classifiers

We developed 5 machine learning classifiers, including support vector machine (SVM), random forest (RF), decision tree (DT), logistic regression (LR), and boosted tree, predicting recurrence from 16 features as mentioned above. The 75 cases were randomly assigned to the training data (80%) and test data (20%) through a generator of random numbers. The robustness of these analyses was examined using classification accuracy and the area under the curve (AUC) with a 5-fold cross-validation method. The implementation of machine learning was performed using Python as a programming language by using the Turi Create machine learning package.

### 3. Evaluation technique

An accuracy score was used to assess the test performance. The accuracy was calculated as follows:  $\text{accuracy} = \frac{\text{correctly predicted as non-recurrence in non-recurrence case} + \text{correctly predicted as recurrence case in recurrence}}{\text{Total case}} (n=75)$ . We also used the AUC of the receiver-operating characteristic.

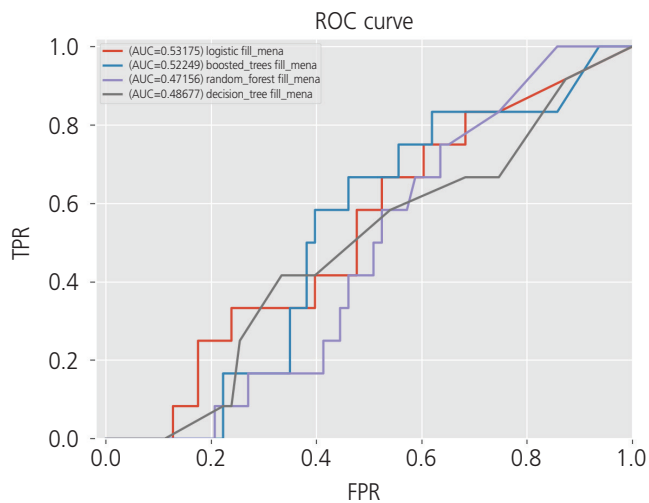
Statistical analyses were performed using the Turi Create machine learning package of Python and R statistical software (R Foundation, Vienna, Austria). For continuous variables, the *t*-test was used, and the data were reported as

medians and ranges. For categorical variables, Pearson's  $\chi^2$  test was used, and the data were reported as percentages. Two-sided *P*-values <0.05 were considered significant.

**Table 1.** The details of the patients' information and pathological examination for 75 patients with endometrial cancers

Characteristic	Non-recurrence (n=63)	Recurrence (n=12)	P-value
Patients' demographic			
Age (yr)	57.8	69.1	0.005
Gravidity	1.9	1.2	0.240
Parity	1.36	0.92	0.270
BMI (kg/m <sup>2</sup> )	24.3	23.1	0.450
Medical history			
HT	0.22 (14/63)	0.33 (4/12)	0.460
DM	0.12 (7/63)	0.12 (2/12)	1.000
Tumor markers			
CA125	17.4 (5.7–178.0)	25.7 (8.2–141.0)	0.430
CA19-9	15.4 (0.5–99.0)	18.6 (0.9–29.5)	0.520
CEA	1.6 (0.4–7.8)	2.9 (1.0–5.5)	0.120
Pathologic factors			
FIGO			
Stage 1	0.82 (52/63)	0.66 (8/12)	0.240
Stage 2	0.18 (11/63)	0.33 (4/13)	
Rate of endometrioid carcinoma	0.79 (58/63)	0.66 (8/12)	0.031
Detail of histological diagnosis			
Endometrioid (n=66)			
Grade 1	0.44 (26/58)	0.50 (4/8)	1.000
Grade 2	0.44 (26/58)	0.50 (4/8)	
Grade 3	0.11 (6/58)	0	
Other subtype (n=9)			
Serous	1	1	
Small cell	2	0	
Mixed	1	0	
Carcinosarcoma	1	3	
Therapeutic factors			
Surgical approach			
TAH	0.98 (62/63)	1.00 (12/12)	1.000
PLA	0.55 (35/63)	0.75 (9/12)	0.380
PALA	0.17 (11/63)	0.16 (2/12)	1.000
OMT	0.06 (4/63)	0.25 (3/12)	0.076
No. of course of chemotherapy	0.85	1.83	0.130

BMI, body mass index; HT, hypertension; DM, diabetes mellitus; CA125, carbohydrate antigen 125; CA19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen; FIGO, International Federation of Gynecology and Obstetrics; TAH, total abdominal hysterectomy; PLA, pelvic lymphadenectomy; PALA, para-aortic lymphadenectomy; OMT, omentectomy.



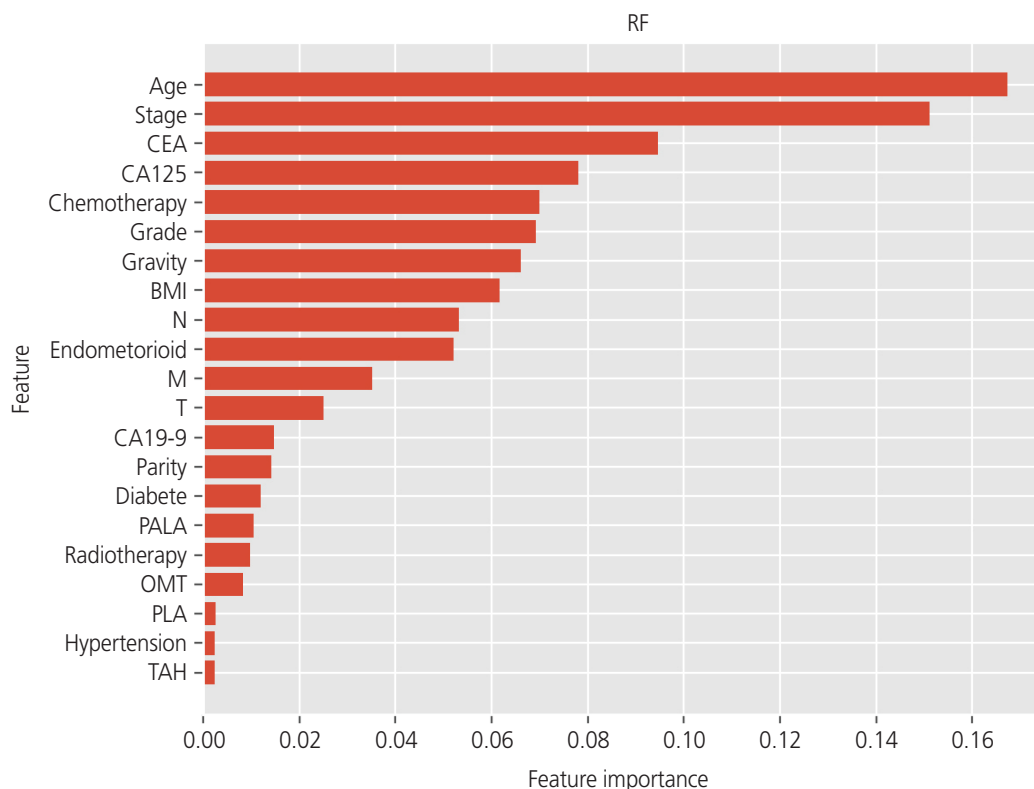
**Fig. 1.** The receiver operating characteristic (ROC) curve of the 5 algorithms. The highest area under the curve (AUC) was 0.53 for logistic regression, followed by 0.52 for boosted trees, 0.48 for decision tree, and 0.47 for random forest. FPR, false positive rate; TPR, true positive rate.

## Results

### 1. Patient and tumor characteristics

A total of 75 patients with endometrial cancers were enrolled, of which 63 had no recurrence 5 years after the primary surgery and 12 had recurrence. The values of the patients' information and pathological examination divided by the state of prognosis are summarized in Table 1.

Among the patients' demographic, a significant difference was noticed in the "age" of the patients. Compared with patients in the non-recurrence group (median age: 57 years), the median age in the recurrence groups was 10 years greater (median age: 69 years). Although a significance was not observed, the median values for all tumor markers were higher in the recurrence groups. Among the pathologic factors, a significant difference was noticed in the "rate of endometrioid carcinoma." In the non-recurrence group, the endometrioid carcinoma was more frequently included (79%) than in the non-recurrence group (66%).



**Fig. 2.** In the analysis of the importance of factors on the prediction of the recurrence, the random forest (RF) classifier showed that "age", "stage", and "carcinoembryonic antigen (CEA)" were the most valuable factors. CA125, carbohydrate antigen 125; BMI, body mass index; N, nodes; M, metastasis; T, tumor; CA19-9, carbohydrate antigen 19-9; PALA, para-aortic lymphadenectomy; OMT, omentectomy; PLA, pelvic lymphadenectomy; TAH, total abdominal hysterectomy.

Although no difference was noted, stage 2 occurred more in the recurrence groups (33%) than in the non-recurrence groups (18%). In the analysis of the other subtypes of pathology, carcinosarcoma was more common in the recurrence groups. There was no significant difference observed in any of the therapeutic factors (surgical content and adjuvant chemotherapy).

## 2. Performance of machine learning classifiers

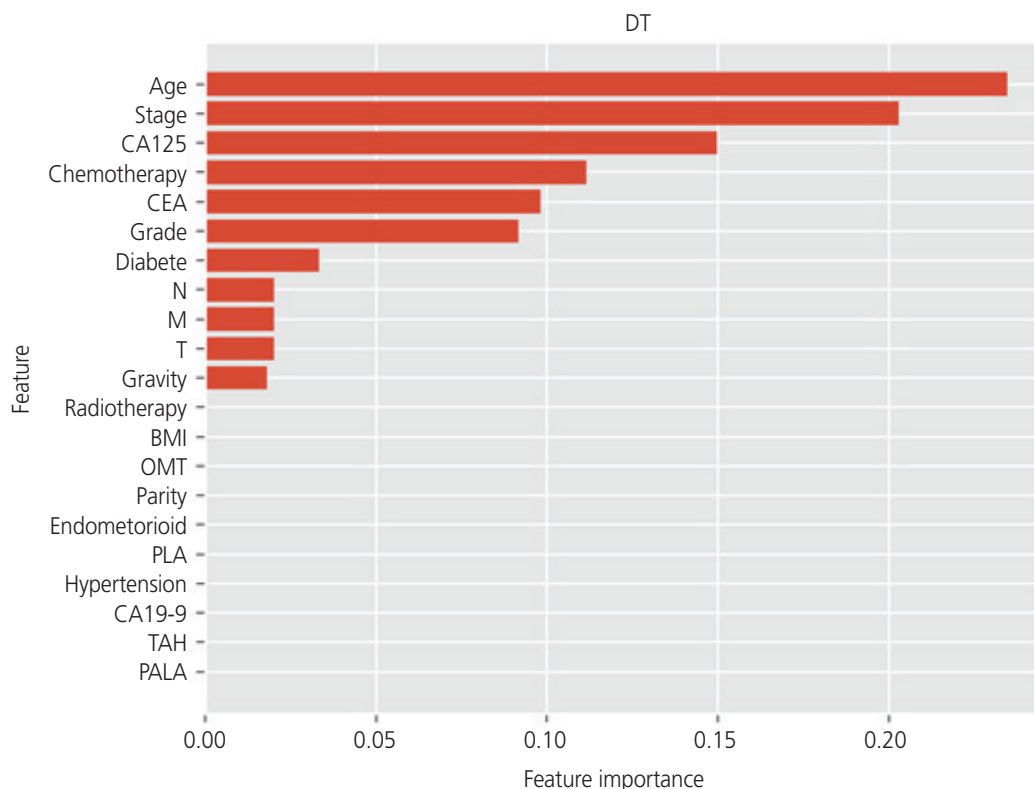
The highest accuracy was 0.80 for RF, followed by 0.77 for SVM, 0.73 for LR, 0.70 for boosted trees, and 0.66 for DT. The highest AUC was 0.53 for LR, followed by 0.52 for boosted trees, 0.48 for DT, and 0.47 for RF. The receiver operating characteristic curve of the algorithms is shown in Fig. 1.

Using tree-based classifiers (RF, DT, and boosted tree), we analyzed the importance of each clinical factor (16 features) in predicting recurrence. In the RF classifier, "age", "stage", and "CEA" were the most valuable factors for the prediction (Fig. 2). Similarly, in the DT classifier, "age", "stage", and

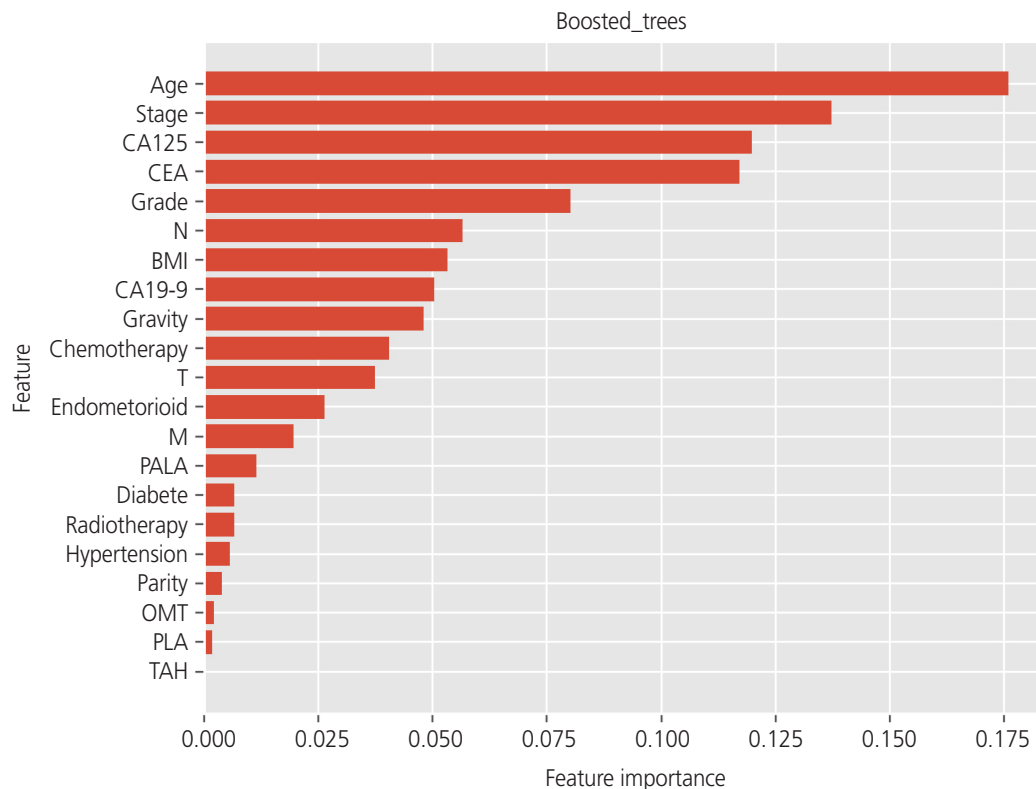
"CA125" were the most valuable factors on the prediction (Fig. 3). In the boosted trees classifier, "age", "stage", and "CEA" were the most valuable factors on the prediction (Fig. 4). In addition, the number of courses of chemotherapy was also considered a valuable factor for all 3 classifiers.

## Discussion

Endometrial cancer is the most common gynecologic malignancy in developed countries, and its incidence is increasing [1]. Most endometrial cancers (75%) are diagnosed at an early stage (FIGO stages I or II) with a better prognosis and a 5-year overall survival rate ranging from 74% to 91% [1]. However, even among patients with early stage endometrial cancer, there is a subset of patients who develop recurrence after the initial treatment of the primary tumor. For those in whom endometrial cancer recurs or progresses to distant sites, complete recovery is difficult and palliative care is considered.



**Fig. 3.** In the analysis of the importance of factors on the prediction of the recurrence, the decision tree (DT) classifier showed that "age", "stage", and "carbohydrate antigen 125 (CA125)" were the most valuable factors. CEA, carcinoembryonic antigen; BMI, body mass index; N, nodes; M, metastasis; T, tumor; CA19-9, carbohydrate antigen 19-9; PALA, para-aortic lymphadenectomy; OMT, omentectomy; PLA, pelvic lymphadenectomy; TAH, total abdominal hysterectomy.



**Fig. 4.** In the analysis of the importance of factors on the prediction of the recurrence, the boosted tree classifier showed that “age”, “stage”, and “carbohydrate antigen 125 (CA125)” were the most valuable factors. CEA, carcinoembryonic antigen; BMI, body mass index; N, nodes; M, metastasis; T, tumor; CA19-9, carbohydrate antigen 19-9; PALA, para-aortic lymphadenectomy; OMT, omentectomy; PLA, pelvic lymphadenectomy; TAH, total abdominal hysterectomy.

Prediction of the prognosis is one of the biggest problems in cancer therapy. Accurate prediction of the recurrence or survival after treatment could help guide additional treatment and follow-up strategies. Traditionally, the prognosis of endometrial carcinoma is determined primarily by the final FIGO stage. In addition, previous studies analyzed many clinical factors that affected the postoperative course. The other predictive factors were reported as tumor grade, age, comorbidities, tumor diameter, American Society of Anesthesiologists score, lymphovascular space involvement, and postoperative complications at 30 days [1]. Naturally, postoperative treatment content, such as chemotherapy or radiotherapy, could play equally important roles in prognosis.

To predict prognosis, nomograms have been introduced as predictive models in the region of endometrial cancer. A nomogram is a predictive tool that creates a simple graphical representation of a statistical model that generates the numerical probability of a clinical event. It has also been described as a chart representing numerical relationships or

a graphic calculation tool [2]. Several studies have shown that nomograms have better individual discrimination than current staging systems. Abu-Rustum et al. [2] developed a nomogram based on 5 easily available clinical characteristics, such as FIGO stage, grade, histologic subtype, age, lymph node metastasis, and 3- and 5-year OS with a high concordance probability ( $0.746 \pm 0.011$ ). The authors stated that incorporating other clinical variables is important for a more accurate prediction of patients' individualized outcomes. Zhu et al. [3] used age, race, year of diagnosis, histologic grade, clinical stage, tumor size, and developing a nomogram for the prediction of 3- and 5-year OS. As the model of Abu-Rustum et al. [2], the nomogram showed a high concordance probability (0.782). Although these nomograms are easily used and calculated, the linear relationship of limited variables in a nomogram is insufficient in the predictive task, considering that real-life factors are numerous and have a complex nonlinear relationship.

AI is considered a novel diagnostic technique for medical

diagnosis and is different from traditional computer programming [2]. A previous general programming algorithm produces outputs using the input data and given rules. In contrast, AI can produce rules using input and output data. Given the input and output data of the existing dataset, the AI algorithm can derive rules and patterns hidden in data [3]. Furthermore, using the newly found rules and patterns, AI can also predict the output prospectively from other input data. AI prediction has been applied and studied in various scientific areas. Machine learning or deep learning is a sub-type of AI. In particular, for the imaging tasks, deep learning showed excellent predictive performance. In medicine, several reports have shown deep learning to have high accuracy in diagnostics of imaging examinations, such as in head CT scans [4], skin cancer [5], and retinopathy in diabetic patients [6].

In the gynecologic region, Matsuo et al. [7] analyzed the prediction of survival length in patients with cervical cancer using a deep learning model. Compared with the traditional Cox regression model, the author showed better predictive performance for deep learning. The author commented that the strengths of the deep-learning model existed in the following 3 points. First, the model exhibits an improved fit for variables with a nonlinear relationship, which is applicable when examining real-life factors. Deep-learning approaches can model nonlinear risk functions that are present in survival data. Second, deep-learning models can not only automatically learn feature representations from raw clinical data without explicit feature engineering but can also fit censored survival data with the use of nonlinear risk functions. In other words, deep-learning models are powerful for learning nonlinear relationships that are present in the data, and they can easily handle censoring in survival data. Thus, selection bias due to the process of demographic grouping can be eliminated in the deep-learning model. Third, the performance of the deep-learning model is superior when large feature sets are used. The strength of the deep-learning model in handling large feature sets because of its ability to learn feature representation, may be beneficial particularly in biomedical research because the inclusion of many variables in conventional linear regression models may result in overfitting.

In this study, we showed the possibility of AI prediction in patients with endometrial cancer. Because of limited data, the AUC was not high in order to incorporate the prediction to clinical situations. However, with the large dataset, the performance of AI will improve. We used the data from

our institute, so the size of the dataset had several hundred patients. Considering that AI works well with big data, data over ten thousand patients should be prepared. Research in a multi-institute or using a database of one area and one country is necessary for accuracy.

## Conflict of interest

No potential conflict of interest relevant to this article was reported.

## Ethical approval

The dataset of patients with endometrial cancer from Tokyo Women's Medical University Medical Center East was used after obtaining the approval of the Institutional Review Board (IRB).

## Patient consent

Informed consent was waived by the Institutional Review Board since this study was retrospective, and the personal information in the data was blinded.

## Funding information

None

## References

1. Morice P, Leary A, Creutzberg C, Abu-Rustum N, Darai E. Endometrial cancer. *Lancet* 2016;387:1094-108.
2. Abu-Rustum NR, Zhou Q, Gomez JD, Alektiar KM, Hensley ML, Soslow RA, et al. A nomogram for predicting overall survival of women with endometrial cancer following primary therapy: toward improving individualized cancer care. *Gynecol Oncol* 2010;116:399-403.
3. Zhu L, Sun X, Bai W. Nomograms for predicting cancer-specific and overall survival among patients with endometrial carcinoma: a SEER based study. *Front Oncol*

2020;10:269.

4. Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet* 2018;392:2388-96.
5. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115-8.
6. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;316:2402-10.
7. Matsuo K, Purushotham S, Jiang B, Mandelbaum RS, Takiuchi T, Liu Y, et al. Survival outcome prediction in cervical cancer: Cox models vs deep-learning model. *Am J Obstet Gynecol* 2019;220:381.e1-14.