Journal of Radiation Research, Vol. 66, No. 2, 2025, pp. 144–156
https://doi.org/10.1093/jrr/rraf004
Advance Access Publication: 7 March 2025

Development of a deep learning-based model to evaluate changes during radiotherapy using cervical cancer digital pathology

Masaaki Goto[1,2,*], Yasunori Futamura[3,4], Hirokazu Makishima[1,5], Takashi Saito[1], Noriaki Sakamoto[6], Tatsuo Iijima[7], Yoshio Tamaki[8], Toshiyuki Okumura[9], Tetsuya Sakurai[3,4] and Hideyuki Sakurai[1]

[1]Department of Radiation Oncology & Proton Medical Research Center, Institute of Medicine, University of Tsukuba, 2-1-1 Amakubo, Tsubuka, Ibaraki 305-8576, Japan
[2]Department of Radiation Oncology, Japan Red Cross Medical Center, 4-1-22 Hiroo, Shibuya, Tokyo 150-8935, Japan
[3]Center for Artificial Intelligence Research, University of Tsukuba, 1-1-1 Tennodai, Tsubuka, Ibaraki 305-8577, Japan
[4]Department of Computer Science, University of Tsukuba, 1-1-1 Tennodai, Tsubuka, Ibaraki 305-8577, Japan
[5]QST Hospital, National Institute for Quantum Science and Technology, 4-9-1 Anagawa, Inage, Chiba 263-8555, Japan
[6]Department of Diagnostic Pathology, Institute of Medicine, University of Tsukuba, 1-1-1 Tennodai, Tsubuka, Ibaraki 305-8575, Japan
[7]Department of Diagnostic Pathology, Ibaraki Prefectural Central Hospital, 6528 Koibuchi, Kasama, Ibaraki 309-1793, Japan
[8]Department of Radiation Oncology, Fukushima Rosai Hospital, 3 Numaziri, Uchigotsuzuramachi, Iwaki, Fukushima 973-8403, Japan
[9]Department of Radiation Oncology, Ibaraki Prefectural Central Hospital, 6528 Koibuchi, Kasama, Ibaraki 309-1793, Japan
*Corresponding author. Department of Radiation Oncology & Proton Medical Research Center, Institute of Medicine, University of Tsukuba, Department of Radiation Oncology, Japan Red Cross Medical Center, 2-1-1, Amakubo, Tsukuba, Ibaraki 305-8576, Japan. Tel: +81 298537100; Fax: +81 298537102; Email: goto@pmrc.tsukuba.ac.jp
(Received 22 October 2024; revised 20 December 2024; editorial decision 22 January 2025)

ABSTRACT

This study aims to create a deep learning-based classification model for cervical cancer biopsy before and during radiotherapy, visualize the results on whole slide images (WSIs), and explore the clinical significance of obtained features. This study included 95 patients with cervical cancer who received radiotherapy between April 2013 and December 2020. Hematoxylin–eosin stained biopsies were digitized to WSIs and divided into small tiles. Our model adopted the feature extractor of DenseNet121 and the classifier of the support vector machine. About 12 400 tiles were used for training the model and 6000 tiles for testing. The model performance was assessed on a per-tile and per-WSI basis. The resultant probability was defined as radiotherapy status probability (RSP) and its color map was visualized on WSIs. Survival analysis was performed to examine the clinical significance of the RSP. In the test set, the trained model had an area under the receiver operating characteristic curve of 0.76 per-tile and 0.95 per-WSI. In visualization, the model focused on viable tumor components and stroma in tumor biopsies. While survival analysis failed to show the prognostic impact of RSP during treatment, cases with low RSP at diagnosis had prolonged overall survival compared to those with high RSP ($P = 0.045$). In conclusion, we successfully developed a model to classify biopsies before and during radiotherapy and visualized the result on slide images. Low RSP cases before treatment had a better prognosis, suggesting that tumor morphologic features obtained using the model may be useful for predicting prognosis.

Keywords: radiation therapy; deep learning; digital pathology; cervical cancer

INTRODUCTION

Radiotherapy plays a pivotal role in cancer therapy and is the first-line treatment for over 30% of types of cancer [1]. Radiotherapy affects tumor cells and the tumor microenvironment, ultimately inducing tumor cell death. Tumors that have not been treated typically contain viable tumor cells, whereas irradiated tumors generally have residual tumor cells within necrotic or fibrotic tissue. These radiation-induced histopathological changes in tumor cells and surrounding tissue are used for evaluation and grading schemes for assessment of the effects of radiotherapy on different types of cancer [2, 3]. However, describing irradiated tumor cells and their microenvironment quantitatively remains challenging and the prognostic implications are largely unknown.

Deep learning is a field of artificial intelligence and a form of machine learning that has recently been applied to medical imaging. While conventional machine learning requires features to be defined manually, deep learning can automatically extract features from training data, which is advantageous for tasks with complex or unknown features [4, 5]. Deep learning models using radiological imaging have been shown to be useful [6–9] and studies on automated segmentation are well-known, especially among radiation oncologists [10, 11]. Deep learning is also being applied to digital pathology and may be able to extract pathological features beyond human visual perception [12, 13]. Such clinical applications of deep learning are likely to improve diagnosis, prognosis, and predictive clinical decision-making in cancer treatment.

To our knowledge, there have been no studies of deep learning adapted to histopathological analysis of irradiated tissues. Thus, it is unclear if radiotherapy-induced changes in tumors can be captured by deep learning. If this is possible and deep learning can distinguish naïve from irradiated tumors, this may lead to quantitative evaluation of treatment responses, extraction of novel prognostic characteristics, and provision of a better understanding of the mechanism of the tumor response to radiotherapy. Thus, the purpose of this study is to create a classification model for cervical tumor biopsy treated with radiotherapy, visualize the results on whole slide images (WSIs), and explore the clinical significance of morphology changes during radiotherapy.

MATERIAL AND METHODS

Study design

This retrospective exploratory study aims to investigate whether deep learning can detect pathological features affected by radiotherapy. We utilized a deep learning-based model to classify cervical cancer biopsy images taken before and during radiotherapy. Tumor biopsy images were collected retrospectively from patients who underwent radiotherapy between April 2013 and December 2020 at a single center. After model development, visualization was performed to interpret the decisions of the model, and the prognostic impact of quantitative values generated by the model was analysed statistically. The overall study pipeline is shown in Fig. 1.

This study was approved from the institutional review board (IRB) of the participating hospital (IRB no. 937, approved on 10 March 2021), and was carried out in accordance with the Declaration of Helsinki. The IRB waived the requirement of written informed consent, and consent was obtained in the form of an opt-out on a website. This manuscript was constructed according to the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) guidelines [14].

Data

The study included cervical cancer patients who underwent definitive radiotherapy with two biopsies, one before and one during treatment, from 1 April 2013 to 1 December 2020. These patients received a combination of external beam radiotherapy (EBRT) of 50 Gy in 25 fractions with 5 fractions of central shielding, and 2–5 times high-dose-rate intracavitary brachytherapy (HDR-ICBT) of 6 Gy with or without cisplatin [15, 16]. Clinical data, including age, performance status (Eastern Cooperative Oncology Group: ECOG), histological type, staging, and treatment method, were collected retrospectively from medical records. Since there were some cases lacking high-risk clinical target volume (HR-CTV) dose data, we estimated the total dose to the cervix as a substitute for the HR-CTV dose. The total dose to the cervix was calculated by adding the EBRT dose, excluding the central shield, to the HDR-ICBT dose prescribed to point A. Overall survival (OS) was defined as time from the start of treatment to death from any cause and progression-free survival (PFS) was defined as time from the start of treatment to disease progression at any site or death from any cause. Cervical tumor biopsies taken before treatment were used for histological diagnosis. Biopsies taken during treatment were obtained about 3 weeks after the start of external beam radiotherapy to assess the treatment response. While tumors that responded well to treatment could potentially lead to false negative results, imaging and clinical findings during treatment were used to determine the biopsy site, and multiple samples were obtained if necessary to minimize sampling errors. All slides used in the study were hematoxylin and eosin (H&E) stained and reviewed by a pathologist.

Data preprocessing

All slides were digitized using a NanoZoomer (Hamamatsu Photonics), which produces a final WSI with a pixel-level resolution of 227 nm/pixel (111 894 DPI) at 40× magnification. Each WSI was saved in NDPI image format and consists of as many as tens of billions of pixels, which is hard to analyse without preprocessing. To use large-size WSIs for deep learning while not losing information at the cellular level, the entire WSI is divided into tiles of a specific number of pixels (17). We used the Python module Histprep to divide WSIs into tiles of 224 × 224 pixels according to the size of the input data. [17]. In this module, preprocessing metrics such as hue, saturation, and brightness are calculated and saved as metadata. Tiles include both tumor and non-tumor areas but exclude background and artifact tiles using the metadata hue (hue $0.75 > 130$). In the process of creating tiles, each one received a unique identifier that included a patient-specific ID, which guaranteed traceability and prevented any data leakage while dividing the data. In addition to this identifier, each tile was also given coordinate data within the WSI, which served as a locator for the tiles in later visualization stages.

Ground truth labeling

WSIs were labeled as "pre-treatment" or "mid-treatment" according to the timing of acquisition. Biopsies taken before radiotherapy were defined as "pretreatment," and those taken during radiotherapy at

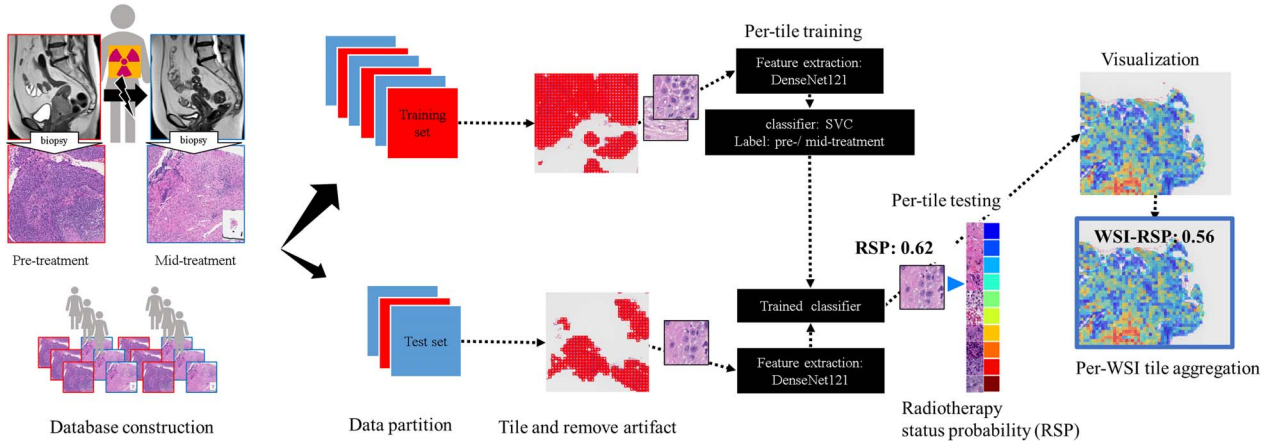

Fig. 1. Pipeline of deep learning-based model using whole slide images (WSIs) for classifying radiotherapy status. A total of 95 patients with cervical cancer treated with radiotherapy were enrolled in the study. Data were divided on a patient basis at a 2:1 ratio by time series. Digitized WSIs were preprocessed to small tiles (224 × 224 pixels). These tiles were used as input data with a label of 'pre-treatment' or 'mid-treatment' depending on the sampling time. In the model, a pre-trained convolutional neural network (DenseNet 121) was used as a feature extractor, and a support vector machine was used as a classifier (SVC). The training set included 12 400 tiles and the test set used 6000 tiles. The outcome of the model was defined as the Radiotherapy Status Probability (RSP), which was used in visualization and survival analysis.

about 3 weeks to evaluate the treatment response were defined as "mid-treatment". The timing of mid-treatment biopsies varied, but brachytherapy had not yet begun. Since this information was not accompanied by subjective judgments, no blinding procedures were performed.

Data partitioning

All labeled data were divided into a training set and a test set on a patient basis at a 2:1 ratio. The purpose of splitting the dataset per patient is to prevent data leakage during model training. Traceable patient ID information in tiles ensured that there was no overlap of tiles from the same patient across the training and test sets. These data were divided by time series and temporal validation was performed using the test set. While random data splitting is a common method for internal validation, it may lead to significant biases and result in less robust models. The most unbiased form of data splitting is external validation using datasets from other institutions. The temporal validation method is considered to be intermediate between internal and external validation [18, 19].

Model

The model consisted of two components: a convolutional neural network (CNN) for feature extraction using deep learning and a support vector machine (SVM) for classification through machine learning. Thus, our model is a deep learning-based hybrid, and this combination allows us to utilize the strong ability of deep learning to learn features of unknown tasks and the robust classification skills of machine learning for predicting results.

Convolutional neural network for feature extraction

CNNs utilize convolutional operations to extract features from input images for a diverse range of tasks, including image classification, domain segmentation, and regression [20, 21]. One of the advantages of CNNs is the direct utilization of image inputs. Within the convolutional layers, basic features like edges and colors are preserved in the forward layer, while highly abstracted features are stored in the backward layer [22]. Eventually, these features are combined in the fully connected layer and used for classification. CNN models, such as VGG, ResNet, and DenseNet [23–25], are trained on large image datasets (ImageNet) and have demonstrated high performance. These pre-trained CNN models can also function as image feature extractors by employing feature vectors up to the fully connected layer [26–28]. In this study, DenseNet121, which had the best performance in preliminary experiments, was used as the feature extractor. The results of preliminary experiments comparing CNN models are summarized in Table A in the supplementary material. Input data were image data of 224 × 224 pixels, and the 1024-D feature vector was extracted from the last layer just before the fully connected layer. By describing the pathology image as a feature vector and using dimension reduction for the vector, we were also able to capture an overview of the whole dataset. The obtained feature vectors were used as input for the subsequent classification model.

Support vector machine for classification

SVM is a technique used in classifying unknown data and has relatively good performance [29]. There are several reports on methods of training SVMs using feature vectors extracted from CNNs as input in medical imaging analysis [30, 31]. In this study, we trained the SVM as a binary classification. The C-Support Vector Classification (SVC) model within the Python machine learning library, scikit-learn, was utilized for the SVM [32].

Training

In the training process, 100 tiles per WSI were randomly selected as the training set, and a total of 12 400 tiles were used for training the model.

The purpose of the present study was to evaluate the pre-treatment and mid-treatment status of the entire tissue, and tiles were selected without distinguishing between tumor and non-tumor areas. We did not use all the tiles because increasing the number of training tiles did not significantly improve performance. The training was done using the SVM, as the classification part of the model; the parameters of DenseNet121, which was used for feature extraction, were not trained. To ensure that the trained model was not overfitting, we also performed a 10-fold cross-validation within the training dataset and compared the respective accuracy.

Model test

For the test data set, 100 tiles per WSI were randomly selected, for a total of 6000 tiles. The classification performance of the trained model was assessed for the training and test sets.

Evaluation

To evaluate the model, the sensitivity, specificity accuracy, positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating characteristic (ROC) curve (AUC) were calculated on a per-tile basis, using all tiles from both the training and test sets separately. The same evaluation was performed on a per-WSI basis based on the average of the predicted probabilities of the tiles. In addition, per-tile and per-WSI confusion matrices of the trained model were generated for the training and test sets.

Radiotherapy status probability and visualization

The computed probabilities of possible outcomes for input tiles by the trained model were defined as the 'radiotherapy status probability (RSP). RSP is a continuous variable ranging from 0 to 1, with values closer to 0 representing a more pre-treatment' state and values closer to 1 representing a more 'mid-treatment' state. In the visualization process, RSP was assigned to each tile in the WSI and overlaid on a color map using coordinate data in the tile. In this color map, red indicates lower RSP; pre-treatment status, and blue indicates higher RSP; mid-treatment status. Finally, whole slide image RSP (WSI-RSP), the average value of RSP in the WSI was calculated for subsequent statistical analysis.

Survival analysis

We performed a survival analysis using the RSP of both mid-treatment and pre-treatment tissues, based on the hypothesis that mid-treatment RSP could serve as an indicator of the effectiveness of radiotherapy and impact prognosis. All cases regardless of data partitioning were divided into two groups based on the median RSP at pre-treatment and mid-treatment and these groups were compared by log-rank test. Statistical significance was defined as $P < 0.05$. These analyses were performed using EZR [33], a free software for R commander.

RESULTS

Datasets

From April 1st, 2013, to December 1st, 2020, a total of 166 patients underwent radiotherapy for cervical cancer. Of these cases, 95 in which both pre-treatment and mid-treatment biopsies were taken were included in the study. The training set included 64 patients, with a total of 12 400 randomly selected tiles (100 tiles per WSI) from 1 091 283 pre-treatment and 550 568 mid-treatment tiles). The test set included 31 patients, with a total of 6000 randomly selected tiles (100 tiles per WSI) from 463 611 pre-treatment and 182 804 mid-treatment tiles). A flowchart of the dataset criteria is shown in Fig. 2 and a summary of the patient characteristics for both training and test datasets is given in Table 1.

Cross-validation in the training set

In the training set, a 10-fold cross-validation was performed and evaluated the performance of each subset to ensure the model was not overfitting. The class distribution for each fold and the accuracy results from each fold are summarized in Table B in the supplementary material. The overall cross-validation accuracy was 69.98% with a standard deviation of 3.86%, indicating consistent model performance.

Model evaluation

The model trained on all data in the training set was evaluated. In this evaluation, RSPs exceeding 0.5 were judged as mid-treatment status, while values at or below 0.5 as pre-treatment status. The AUC, sensitivity, specificity, accuracy, PPV, and NPV for the trained model on a per-tile basis were 0.76, 63%, 75%, 68%, 72%, and 65%, respectively, for the test set; and 0.86, 80%, 75%, 78%, 76%, and 80%, respectively, for the training set. The model was applied to all tiles of tissue, excluding the background, and WSIs were evaluated by the average of those tiles. On a per-WSI basis, the model had an AUC, sensitivity, specificity, accuracy, PPV, and NPV of 0.95, 71%, 97%, 83%, 96%, and 76%, respectively, for the test set; and 0.99, 95%, 95%, 95%, 95%, and 95%, respectively, for the training set. Per-tile and per-WSI basis ROC curves and performance evaluation are shown in Fig. 3 and Table C in the supplementary material.

Radiotherapy status probability

The trained model outputs continuous values from 0 (pre-treatment) to 1 (mid-treatment) for each tile image input. This value was defined as the RSP. WSI-RSP was calculated as the average of all RSPs of tiles in the WSI. The median pre- and mid-treatment WSI-RSPs were 0.38 (range 0.13 to 0.61) and 0.63 (0.47 to 0.82), respectively, in the training set; and 0.36 (0.13 to 0.56) and 0.56 (0.34 to 0.75), respectively, in the test set. A scattered boxplot of the pre-treatment and mid-treatment WSI-RSPs for each set is shown in Fig. 4.

Visualization

The model was applied to all tiles and visualization was performed on all WSIs. An example of the visualization results is shown in Fig. 5. In the image of representative tiles arranged according to their RSP, viable tumor cells with a high nuclear-to-cytoplasmic ratio displayed a low RSP, represented by a strong red color, while areas of stroma and necrosis showed a high RSP, represented by a strong blue color. In the images of WSI-RSP and a WSI example, it was observed that tiles with extremely high or low RSP predominantly contained either tumor or stromal tissues. Meanwhile, WSIs with intermediate RSP exhibited a heterogeneous distribution, showing various RSPs in different regions.

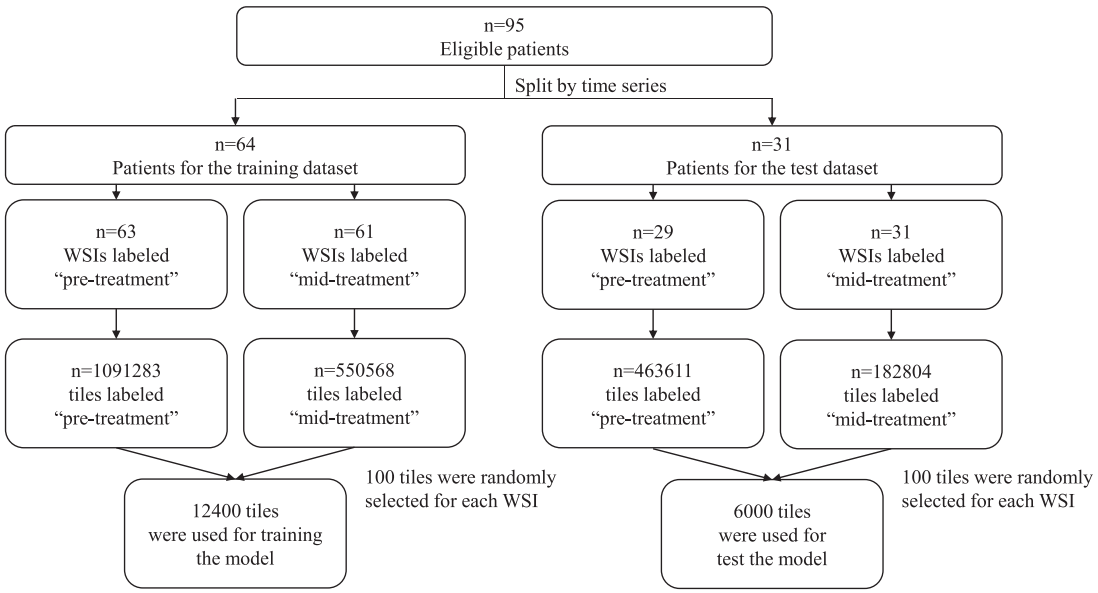

Fig. 2. Flowchart of the datasets. The flowchart shows the number of patients, whole slide images (WSIs), and tiles in each dataset in the study.

Table 1. Summary of the patient characteristics for training and test set

Item	Training set		Test set		
Total patients	64		31		
Age	62	(31–89)	54	(29–80)	$P = 0.015$
PS (ECOG)					$P = 0.43$
0	23	(36%)	17	(55%)	
1	34	(53%)	13	(42%)	
≥2	4	(6%)	1	(3%)	
Histological type					$P = 0.42$
SCC*	58	(91%)	30	(97%)	
not SCC	6	(9%)	1	(3%)	
Staging (FIGO2008)					$P = 0.006$
I	4	(6%)	7	(23%)	
II	29	(45%)	19	(61%)	
III	27	(42%)	4	(13%)	
IV	4	(6%)	1	(3%)	
N staging (UICC 8th)					$P = 0.75$
N0	45	(70%)	20	(65%)	
N1	18	(28%)	11	(35%)	
M staging (UICC 8th)					$P = 0.85$
M0	58	(91%)	30	(97%)	
M1	5	(8%)	1	(3%)	
Treatment					$P = 0.61$
Chemoradiotherapy	48	(75%)	25	(81%)	
Radiotherapy alone	16	(25%)	6	(19%)	
Total dose to the cervix (EQD2(10)[a])	63 Gy	(55–78 Gy)	62 Gy	(55–72 Gy)	$P = 0.46$
Dose at the time of mid-treatment biopsy	32 Gy	(26–44 Gy)	30 Gy	(24–43 Gy)	$P = 0.04$

SCC * = Squamous cell carcinoma, EQD2 [10]
[a]Equivalent dose in 2 Gy/fr for α/β ratio of 10 Gy.

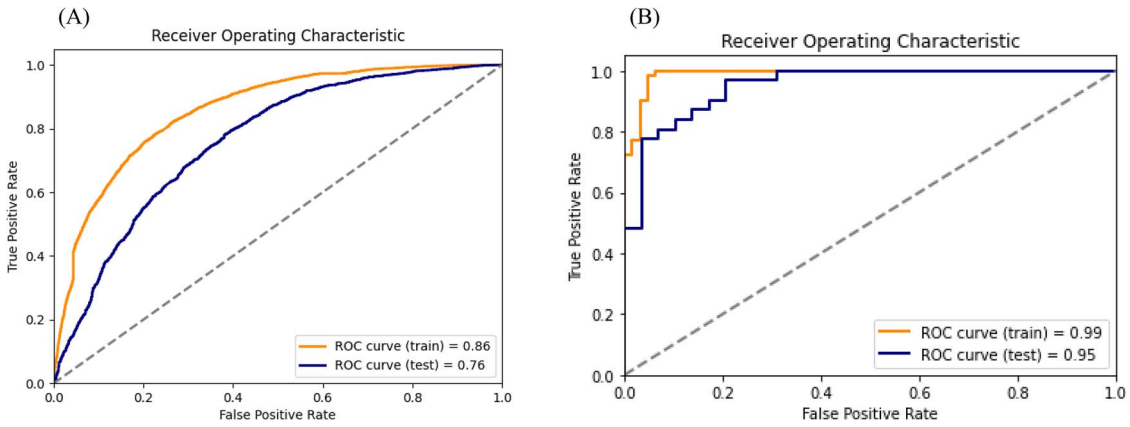

Fig. 3. Receiver operating characteristic (ROC) curves of the trained model. ROC curves show model performance on (A) per-tile and (B) per-whole slide image (WSI) basis. In both figures, the ROC curve for the training and test sets are distinguished as indicated in the legend at the bottom right.

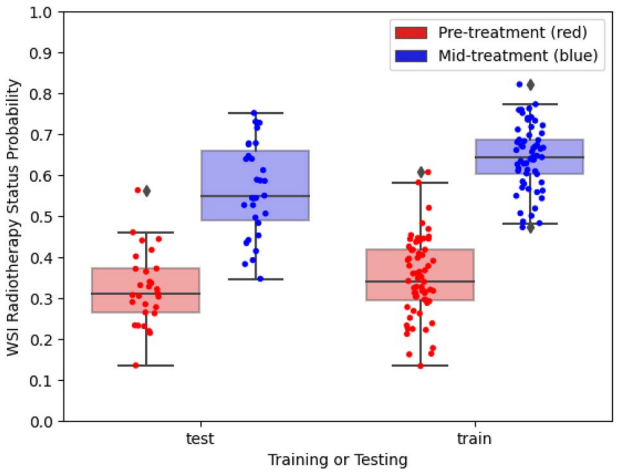

Fig. 4. Scattered boxplot of WSI-RSP of the pre-treatment and mid-treatment group in training and test set. This figure shows the distribution of Whole Slide Image Radiotherapy Status Probability (WSI-RSP), the average RSP value of the WSIs. The pre-treatment and mid-treatment values are distinguished as indicated in the legend at the top right.

In the representative visualization result that correctly judged the pre-treatment and mid-treatment of the same patient, the pre-treatment WSI predominantly showed squamous cell carcinoma, whereas the mid-treatment WSI mostly displayed stromal and necrotic areas, with tumor cells at the margins. Conversely, in the other example where the model incorrectly judged its status, the majority of the pre-treatment tissue was hematoma, displaying a red color indicating a high RSP, which led to a misjudgment of the WSI as mid-treatment. The mid-treatment WSI was characterized by inflammation and degenerated tumor cells with minimal stroma. The small size of the WSI and the sole presence of this condition may have caused the model to incorrectly classify it as pre-treatment.

Survival analysis based on radiotherapy status probability value

We hypothesized that mid-treatment RSP would reflect the effectiveness of radiotherapy and be associated with prognosis. The patients were divided into groups with high and low RSP, using the median value of 0.62 for the mid-treatment WSI-RSP as a threshold. 92 cases without missing data were included in the analysis. Patient characteristics of low and high RSP in mid-treatment biopsies are shown in Table 2. For the low vs. high RSP groups, the median OS periods were not reached (NR) (95% CI, NR-NR) vs. 91 months (95% CI, 64-NR) ($P = 0.76$), and the median PFS periods were NR (95% CI, 43-NR) vs. 64 months (95% CI, 50-NR) ($P = 0.72$). However, these results indicate that mid-treatment RSP was not associated with prognosis. These findings suggest that mid-treatment RSP may not be a reliable indicator for predicting prognosis during radiotherapy. Kaplan–Meier curves for OS and PFS in each group based on mid-treatment biopsies are shown in Fig. 6A and B.

The same analyses were also conducted in the pre-treatment biopsies, with the threshold for the pre-treatment WCI-RSP of 0.36. Patient characteristics of low and high RSP in pre-treatment biopsies are shown in Table 3. For the low vs. high RSP groups, the median OS periods were 90 months (95% CI, NR-NR) vs. 83 months (95% CI, 83-NR) ($P = 0.045$) and the 2-year OS rates were 95% (95% CI, 83–99) vs. 87% (95% CI, 72–95); the median PFS periods were 91 months (95% CI, 64-NR) vs. 50 months (95% CI, 34-NR) ($P = 0.045$) and the 2-year PFS rates were 84% (95% CI, 70–92) vs. 67% (95% CI, 51–79). In contrast to mid-treatment RSP, pre-treatment RSP was associated with prognosis, as the low RSP group, which had stronger pre-treatment characteristics at diagnosis, had significantly longer OS and PFS compared to the high RSP group. Kaplan–Meier curves for OS and PFS in each group based on pre-treatment biopsies are shown in Fig. 6C and D.

Changes in RSP (ΔRSP) between pre-treatment and mid-treatment were analysed for their correlation with OS and PFS, using the Pearson correlation coefficient. A positive correlation was observed between ΔRSP and PFS ($r = 0.21$, $P = 0.04$), but not between ΔRSP and OS ($r = 0.17$, $P = 0.09$). Survival analysis comparing the

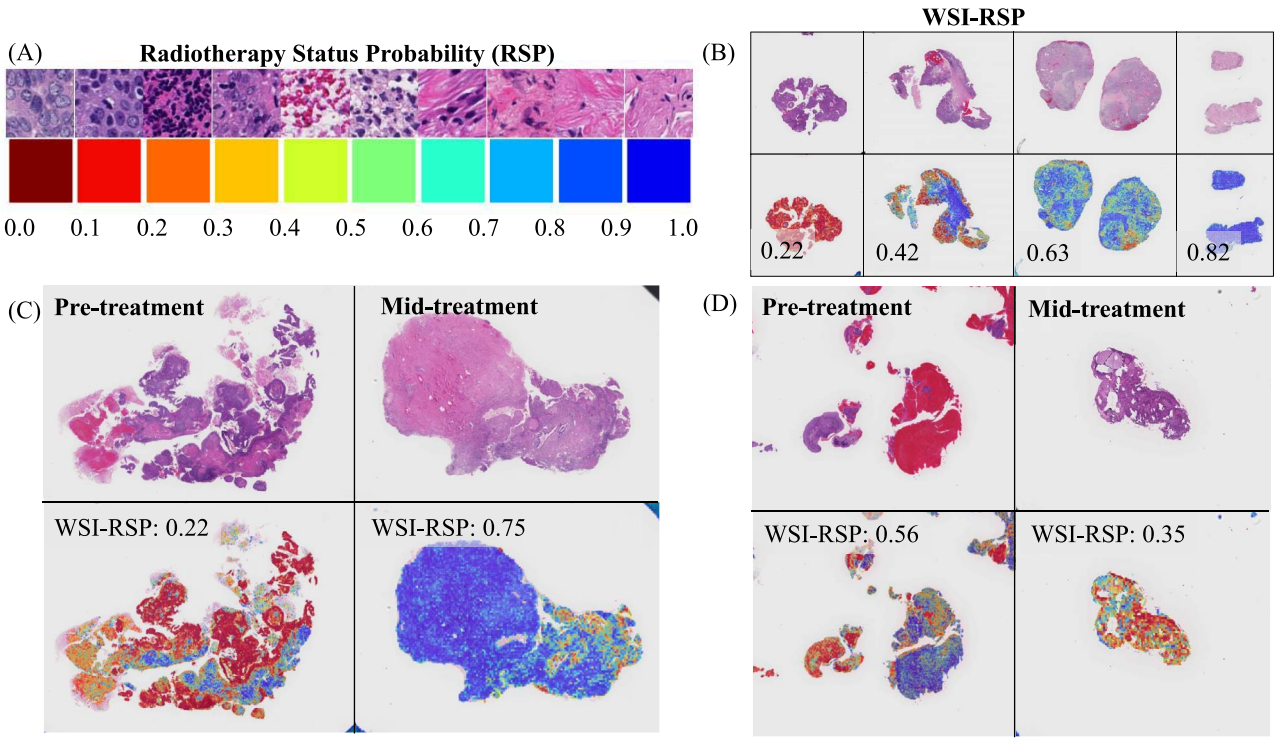

Fig. 5. Examples of visualized tiles and whole slide images. (A) is a scale showing RSP values and representative tiles. Lower RSP values indicate that the model's judgment is closer to pre-treatment, and higher RSP values indicate that the model's judgment is closer to mid-treatment. In the visualization, the RSP values are color-coded as shown in the figure. (B) These visualized biopsies show examples of WSI-RSP and WSIs. The numbers shown on the visualized slides are WSI-RSP. (C) This example shows cases where the model's judgment was correct. The pre-treatment WSI predominantly showed squamous cell carcinoma, whereas the mid-treatment WSI mostly displayed stromal and necrotic areas, with tumor cells at the margins. (D) This example shows cases where the model's judgment was incorrect. The presence of hematoma, which indicated a high RSP, in pre-treatment biopsy and the small size of mid-treatment biopsy may have caused the failed judgment.

high-ΔRSP and low-ΔRSP groups showed no significant difference in OS and PFS (see Supplementary Fig. A for details).

DISCUSSION

We successfully developed a deep learning-based model that can differentiate pre- and mid-treatment specimens. The model achieved an AUC of 0.76, sensitivity of 63%, and specificity of 75% for tile-based classification, and an AUC of 0.95, sensitivity of 71%, and specificity of 97% for WSI-based classification. The model's performance on a per-tile basis is not high but improves significantly when evaluated on a per-WSI basis. This improvement suggests that individual tiles fail to capture the broader context of the tumor and surrounding environment, which are necessary for accurate pre-treatment and mid-treatment status prediction. The result of a scattered boxplot of WSI-RSP distribution showed that the difference in RSP is greater in the training set compared to the test set. This may reflect the fact that the model performed slightly worse in the test group. The distribution of WSI-RSPs with variance within each group is undesirable as a result of the machine learning task but is thought to reflect the heterogeneous nature of the pathological tissue.

Visualization results showed that our trained model focused on tumor cells and stroma in pathological tissues. Through this visualization, it was made understandable to us that the model consistently extracted relevant features from input tiles. While the model correctly classified tumor cell-rich tissues before treatment and stroma-rich tissues during treatment, in certain cases, the model's predictions were affected by poor biopsy specimens, leading to misjudgments. Survival analysis showed that high RSP in pre-treatment WSI was associated with a poor prognosis. These results indicate that our model which quantified histopathology on the pre-treatment and mid-treatment axis can lead to surrogate assessment of tumor cells and stromal tissues, and suggest that the extracted characteristics could be used as prognostic factors.

Prognostic prediction using tumor biopsy is a fundamental approach to managing malignant tumors. Unlike surgical treatment with tumor resection, radiation therapy is not able to obtain an entire tumor sample. As a result, there is an increasing number of studies focusing on image evaluation and radiomics assessments in the field of radiation oncology [34, 35]. However, tumor pathology provides unique insights into tumor characteristics at the cellular level that imaging cannot capture. This information remains valuable

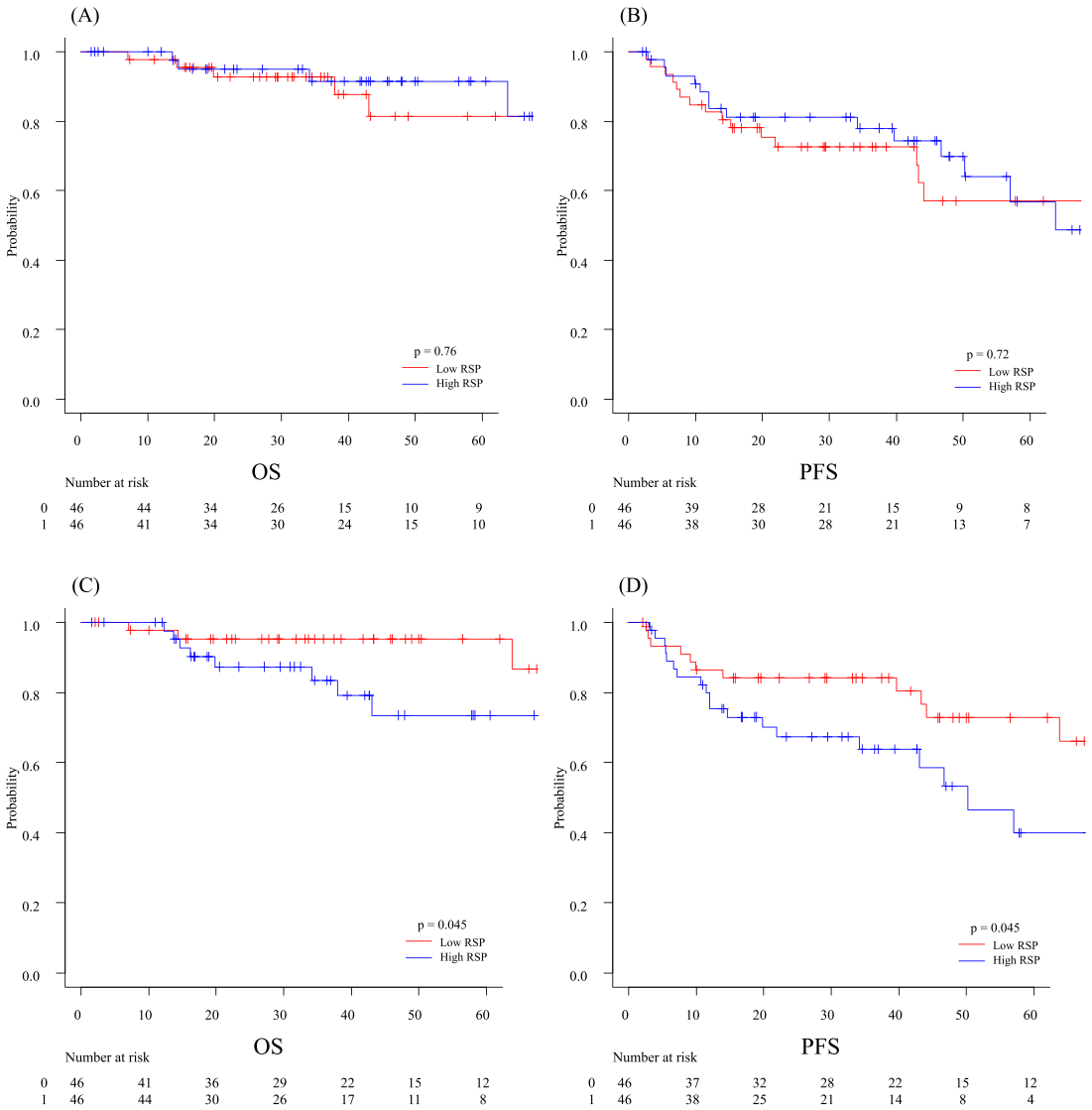

Fig. 6. (A, B) Kaplan–Meier curves for overall survival (OS) and progression-free survival (PFS) based on Radiotherapy Status Probability (RSP) on mid-treatment biopsies. The Kaplan–Meier curves for (A) overall survival and (B) progression-free survival for two patient groups stratified by RSP on mid-treatment biopsies. The red line represents the low-RSP group while the blue line represents those with high RSP during treatment. (C, D) Kaplan–Meier curves for overall survival (OS) and progression-free survival (PFS) based on Radiotherapy Status Probability (RSP) on pre-treatment biopsies. The Kaplan–Meier curves for (C) overall survival and (D) progression-free survival for two patient groups stratified by RSP on pre-treatment biopsies. The red line represents the low-RSP group while the blue line represents those with high RSP at diagnosis.

and should also be studied in the field of radiation therapy. In this study, pre-treatment biopsies are collected for diagnosis. If further model improvement allows identification of robust prognostic features, they could serve as non-invasive tissue biomarkers. Mid-treatment biopsy is considered invasive because it is not routinely recommended. Although the current model did not show the prognostic value of mid-treatment biopsy, evaluating treatment response pathologically remains important. It is expected to detect residual tumors and early recurrence that imaging studies fail to identify. It also serves as a ground truth label in studies using deep learning-based classification to evaluate treatment response. While biopsy results during or after treatment may vary depending on the site of sampling, targeted biopsies, combining imaging studies or radiomics, could achieve more accurate evaluation [36]. It is important to comprehensively analyse tumor information from both pathological and imaging findings to contribute to oncology research.

To the best of our knowledge, this is the first study to create and evaluate a classification model of histopathological changes induced by

Table 2. Summary of patient characteristics low and high radiotherapy status probability (RSP) cases on mid-treatment biopsies

Item	Low RSP**		High RSP		
Total patients	46		46		
Age	60	(29–88)	61	(35–89)	$P = 0.296$
PS (ECOG)					$P = 0.87$
0	21	(45%)	19	(41%)	
1	21	(45%)	24	(52%)	
≥2	3	(6%)	2	(4%)	
Histological type					$P = 1$
SCC*	43	(93%)	42	(91%)	
not SCC	3	(6%)	4	(9%)	
Staging (FIGO2008)					$P = 0.21$
I	8	(17%)	3	(6%)	
II	21	(46%)	26	(57%)	
III	16	(35%)	13	(28%)	
IV	1	(2%)	4	(9%)	
N staging (UICC 8th)					$P = 0.36$
N0	30	(65%)	34	(74%)	
N1	16	(35%)	11	(24%)	
M staging (UICC 8th)					$P = 1$
M0	42	(91%)	43	(93%)	
M1	4	(9%)	2	(4%)	
Treatment					$P = 0.62$
Chemoradiotherapy	37	(80%)	34	(74%)	
Radiotherapy alone	9	(20%)	12	(26%)	
Total dose to the cervix (EQD2(10)[a])	62 Gy	(55–74 Gy)	62 Gy	(58–78 Gy)	$P = 0.55$
Dose at the time of biopsy	30 Gy	(26–44 Gy)	30 Gy	(24–44 Gy)	$P = 0.97$

RSP** = Radiotherapy Status Probability, SCC* = Squamous cell carcinoma, EQD2 [10]
[a]Equivalent dose in 2 Gy/fr for α/β ratio of 10 Gy.

radiotherapy. Several studies have evaluated gene or protein expression during radiotherapy for cervical cancer and showed a correlation with clinical outcomes [37–39]. While histopathological changes during treatment were successfully quantified as RSP, our initial hypothesis that mid-treatment RSP would be associated with prognosis was not supported by our result. The small size of the mid-treatment biopsies, as evidenced by the lower number of tiles, may have led to an inaccurate evaluation. Nonetheless, it is important to note that our findings align with previous studies, which have similarly reported that initial treatment response does not necessarily correlate with survival [40]. Interestingly, when our model was applied to pre-treatment biopsies, we found that the low-RSP group had better survival outcomes. In this group, biopsy tissue was characterized by dense tumor cells and less stroma and fibrosis. This unexpected result suggests that pre-treatment tissue characteristics, as quantified by RSP, may hold prognostic value, even though this was not the original focus of our analysis. Although there was no significant difference in prognosis between the high and low ΔRSP groups, ΔRSP was found to correlate with PFS. This correlation with PFS may be influenced by pre-treatment RSP.

Previous studies have shown that non-keratinizing type SCC cervical cancer has a better prognosis, with a good response to radiotherapy compared to keratinizing type SCC [41]. The presence of CD8+ lymphocytes in the tumor microenvironment before treatment is a favorable factor for successful CRT and a good prognosis [42, 43]. We did not analyse keratinization or lymphocyte infiltration directly, but these are the main changes seen after radiotherapy and match the difference between non-keratinizing and keratinizing SCC. It is of note that such features were extracted without the labels of histological classification, but only based on pre- and mid-treatment labels. Although multivariate analysis was not performed due to the limited number of patients, age and concurrent chemotherapy as possible confounders differed between the two groups in pre-treatment [44, 45]. Further adjustment for confounding factors and inclusion of histopathological features would allow for a more detailed discussion of histopathologic morphology and prognosis. In this study, we employed a hybrid model approach, utilizing deep learning CNNs for feature extraction and machine learning SVMs for classification tasks. In the application of deep learning, purely deep learning methods such as fine-tuning CNN parameters and using fully connected layers for classification have been explored [46, 47]. Compared to these purely deep learning methods, our hybrid model maximizes the benefits of SVM as a stable classifier with few parameters and a convex loss function, especially for binary classification tasks [48]. This approach would overcome issues like overfitting to the training data and local minima problems, which are common in purely end-to-end trained deep learning models.

Table 3. Summary of patient characteristics for low and high radiotherapy status probability (RSP) cases on pre-treatment biopsies

Item	Low RSP**		High RSP		
Total patients	46		46		
Age	59	(31–81)	66	(29–89)	$P = 0.004$
PS (ECOG)					$P = 0.59$
0	20	(43%)	17	(37%)	
1	24	(52%)	23	(50%)	
≥2	1	(2%)	4	(9%)	
Histological type					$P = 0.43$
SCC	44	(96%)	41	(89%)	
not SCC	2	(4%)	5	(11%)	
Staging (FIGO2008)					$P = 0.28$
I	7	(15%)	4	(9%)	
II	25	(54%)	20	(43%)	
III	13	(28%)	16	(34%)	
IV	1	(2%)	4	(9%)	
N staging (UICC 8th)					$P = 0.82$
N0	31	(67%)	32	(70%)	
N1	15	(32%)	13	(28%)	
M staging (UICC 8th)					$P = 1$
M0	43	(93%)	43	(93%)	
M1	3	(7%)	2	(4%)	
Treatment					$P = 0.045$
Chemoradiotherapy	40	(87%)	31	(67%)	
Radiotherapy alone	6	(13%)	15	(33%)	
Total dose to the cervix (EQD2 [10][a])	62 Gy	(55–76 Gy)	62 Gy	(60–78 Gy)	$P = 0.59$

RSP** = Radiotherapy Status Probability, SCC = Squamous cell carcinoma, EQD2 [10]
[a]Equivalent dose in 2 Gy/fr for α/β ratio of 10 Gy.

Additionally, the extracted features by the CNN can be applied not only to classification but also to unsupervised analyses such as clustering [49]. The potential limitations of our hybrid model approach include the increased complexity of the model due to the combination of two different models, and it is necessary to consider the optimal model design according to the specific task in future research.

This study has several limitations. In the model training part, first, a tumor biopsy does not represent the entire tumor and is only a partial assessment of the tumor. Second, the validation cohort was internal. Even though we made a temporal validation dataset to reduce bias, external validation would make the model more robust. Third, the study design is retrospective. While most patients were treated based on Japanese radiation oncology guidelines and basically standardized, radiation oncologists multilaterally evaluated patients and optimized treatment case-by-case, especially for the brachytherapy dose. In the survival time analysis, the result was somewhat affected by data leakage because the trained model was applied to all cases, including the training set. The small sample size, retrospective nature, and lack of multivariate analysis including confounding factors for prognostic factors are the limitations of survival time analysis of this study.

To ensure the robustness of our model, it is necessary to externally validate the model using cervical cancer cases treated in other institutes. To improve the performance of our model and to understand the biological effects of radiation on tumor cells, it is necessary to design an integrated model that not only labels status before and during treatment but also classifies tissues (tumor, stroma, etc.) and allows evaluation of changes in each tissue. Finally, it would be necessary to adapt an established model to a prospective cohort and include confounding and pathological factors, to examine the prognostic value in patients treated with definitive radiotherapy.

CONCLUSIONS

In this study, we developed a deep learning-based model to classify cervical cancer pathology tissues at diagnosis and during radiotherapy. The model showed high classification performance, especially on per-WSI basis, and visualization revealed that the model focused on tumor cells and stroma. While survival analysis failed to show the prognostic impact of RSP during treatment, cases with low RSP at diagnosis were suggested to be potential prognostic factors. To further validate the model, external validation is essential, and we intend to focus on this as a future research task.

ABBREVIATIONS

AUC, Area under the receiver operating characteristic curve; CLAIM, Checklist for Artificial Intelligence in Medical Imaging; CNN, Convolutional neural network; ECOG, Eastern Cooperative Oncology Group; H&E, Hematoxylin and eosin; HDR-ICBT, High-dose-rate intracavitary brachytherapy; HR-CTV, High-risk clinical target volume; IRB, Institutional review board; NPV, Negative predictive

value; OS, Overall survival; PFS, Progression-free survival; PPV, Positive predictive value; ROC, Receiver operating characteristic; RSP, Radiotherapy Status Probability; SVC, Support Vector Classification; SVM, Support vector machine; WSI, Whole slide image; WSI-RSP, Whole Slide Image Radiotherapy Status Probability.

ACKNOWLEDGEMENTS

The author is grateful to Toshiyuki Terunuma, Proton Medical Research Center, Institute of Medicine, University of Tsukuba for providing expert opinions about deep learning on the study. This work was partially supported by Grants-in-Aid for Scientific Research (B) (19 H03596) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

SUPPLEMENTARY DATA

Supplementary data is available at *Journal of Radiation Research* online.

CONFLICT OF INTEREST

None.

FUNDING

This work was partially supported by Grants-in-Aid for Scientific Research (B) (19 H03596) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

PRESENTATION AT A CONFERENCE

Poster presentation at ASTRO2023 and Oral presentation at the 3rd International Congress of the Asian Oncology Society.

REFERENCES

1. Royce TJ, Qureshi MM, Truong MT. Radiotherapy utilization and fractionation patterns during the first course of cancer treatment in the United States from 2004 to 2014. *J Am Coll Radiol* 2018;15:1558–64. https://doi.org/10.1016/j.jacr.2018.04.032.
2. Evans DB, Rich TA, Byrd DR *et al.* Preoperative chemoradiation and pancreaticoduodenectomy for adenocarcinoma of the pancreas. *Arch Surg* 1992;127:1335–9. https://doi.org/10.1001/archsurg.1992.01420110083017.
3. Chatterjee D, Katz MH, Rashid A *et al.* Histologic grading of the extent of residual carcinoma following neoadjuvant chemoradiation in pancreatic ductal adenocarcinoma: a predictor for patient outcome. *Cancer* 2012;118:3182–90. https://doi.org/10.1002/cncr.26651.
4. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44. https://doi.org/10.1038/nature14539.
5. Hinton G. Deep learning-a technology with the potential to transform health care. *JAMA* 2018;320:1101–2. https://doi.org/10.1001/jama.2018.11100.
6. Wasserthal J, Breit HC, Meyer MT *et al.* TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. *Radiol Artif Intell* 2023;5:e230024. https://doi.org/10.1148/ryai.230024.
7. Wang M, Fu F, Zheng B *et al.* Development of an AI system for accurately diagnose hepatocellular carcinoma from computed tomography imaging data. *Br J Cancer* 2021;125:1111–21. https://doi.org/10.1038/s41416-021-01511-w.
8. Calabrese E, Rudie JD, Rauschecker AM *et al.* Feasibility of simulated Postcontrast MRI of glioblastomas and lower-grade gliomas by using three-dimensional fully convolutional neural networks. *Radiol Artif Intell* 2021;3:e200276. https://doi.org/10.1148/ryai.2021200276.
9. Urushibara A, Saida T, Mori K *et al.* Diagnosing uterine cervical cancer on a single T2-weighted image: comparison between deep learning versus radiologists. *Eur J Radiol* 2021;135:109471. https://doi.org/10.1016/j.ejrad.2020.109471.
10. van Rooij W, Dahele M, Ribeiro Brandao H *et al.* Deep learning-based delineation of head and neck organs at risk: geometric and Dosimetric evaluation. *Int J Radiat Oncol Biol Phys* 2019;104:677–84. https://doi.org/10.1016/j.ijrobp.2019.02.040.
11. Rigaud B, Anderson BM, Yu ZH *et al.* Automatic segmentation using deep learning to enable online dose optimization during adaptive radiation therapy of cervical cancer. *Int J Radiat Oncol Biol Phys* 2021;109:1096–110. https://doi.org/10.1016/j.ijrobp.2020.10.038.
12. Baxi V, Edwards R, Montalto M, Saha S. Digital pathology and artificial intelligence in translational medicine and clinical practice. *Mod Pathol* 2022;35:23–32. https://doi.org/10.1038/s41379-021-00919-2.
13. Xing F, Yang L. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. *IEEE Rev Biomed Eng* 2016;9:234–63. https://doi.org/10.1109/RBME.2016.2515127.
14. Mongan J, Moy L, Kahn CE Jr. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell* 2020;2:e200029. https://doi.org/10.1148/ryai.2020200029.
15. Nakano T, Kato S, Ohno T *et al.* Long-term results of high-dose rate intracavitary brachytherapy for squamous cell carcinoma of the uterine cervix. *Cancer* 2005;103:92–101. https://doi.org/10.1002/cncr.20734.
16. Murakami N, Kasamatsu T, Wakita A *et al.* CT based three dimensional dose-volume evaluations for high-dose rate intracavitary brachytherapy for cervical cancer. *BMC Cancer* 2014;14:447Published 2014 Jun 17. https://doi.org/10.1186/1471-2407-14-447.
17. Pohjonen J, Ariotta V *et al.* HistoPrep: Preprocessing large medical images for machine learning made easy!GitHub repository. 2022; https://github.com/jopo666/HistoPrep.
18. Ramspek CL, Jager KJ, Dekker FW *et al.* External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* 2020;14:49–58. https://doi.org/10.1093/ckj/sfaa188.
19. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594. https://doi.org/10.1136/bmj.g7594.

20. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun ACM* 2017;60:84–90. https://doi.org/10.1145/3065386.
21. François C. Xception: Deep Learning with Depthwise Separable Convolutions. F. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA*. IEEE, New York, NY, USA, 2017, pp. 1800–1807. https://doi.org/10.1109/CVPR.2017.195
22. Zeiler MD, Fergus R. Visualizing and Understanding Convolutional Networks. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds). *Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science*, vol. 8689. Springer, Cham. https://doi.org/10.1007/978-3-319-10590-1_53
23. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: *3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA*. International Conference on Learning Representations, pp. 1–14. https://doi.org/10.48550/arXiv.1409.1556
24. Kaiming H, Zhang X, Ren S *et al.* Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA*. IEEE, New York, NY, USA, 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90
25. Huang G, Liu Z, Weinberger KQ *et al.* Densely Connected Convolutional Networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA*. IEEE, New York, NY, USA, 2017, pp. 2261–2269. https://doi.org/10.1109/CVPR.2017.243
26. Muhammad AA, Assef J, Nada G. Pre-trained CNNs as feature-extraction modules for image captioning: an experimental study. *Electron Lett Comput Vis Image Anal* 2022;21:1–16. https://doi.org/10.5565/rev/elcvia.1436.
27. Rajaraman S, Antani SK, Poostchi M *et al.* Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ* 2018;6:e4568. https://doi.org/10.7717/peerj.4568.
28. Awais M, Long X, Yin B *et al.* Can pre-trained convolutional neural networks be directly used as a feature extractor for video-based neonatal sleep and wake classification? *BMC Res Notes* 2020;13:507. https://doi.org/10.1186/s13104-020-05343-4.
29. Evgeniou T, Pontil M. Support Vector Machines: Theory and Applications. In: Paliouras G, Karkaletsis V, Spyropoulos CD (eds). *Machine Learning and Its Applications. ACAI 1999. Lecture Notes in Computer Science*, vol. 2049. pp. 249–257. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-44673-7_12
30. Rathore S, Chaddad A, Iftikhar MA *et al.* Combining MRI and histologic imaging features for predicting overall survival in patients with glioma. *Radiol Imaging Cancer* 2021;3:e200108. https://doi.org/10.1148/rycan.2021200108.
31. Gupta A, Das S, Khurana T *et al.* Prediction of lung cancer from low-resolution nodules in CT-scan images by using deep features. In: *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India*. IEEE, New York, NY, USA, 2018, pp. 531–537. https://doi.org/10.1109/ICACCI.2018.8554951.
32. Pedregosa F, Varoquaux G, Gramfort A *et al.* Scikit-learn: machine learning in python. *J Mach Learn Res* 2011;12:2825–30.
33. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplant* 2013;48:452–8. https://doi.org/10.1038/bmt.2012.244.
34. Hoebers FJP, Wee L, Likitlersuang J *et al.* Artificial intelligence research in radiation oncology: a practical guide for the clinician on concepts and methods. *BJR Open* 2024;6:tzae039. https://doi.org/10.1093/bjro/tzae039.
35. Bibault JE, Giraud P, Burgun A. Big data and machine learning in radiation oncology: state of the art and future prospects. *Cancer Lett* 2016;382:110–7. https://doi.org/10.1016/j.canlet.2016.05.033.
36. Beer L, Martin-Gonzalez P, Delgado-Ortet M *et al.* Ultrasound-guided targeted biopsies of CT-based radiomic tumour habitats: technical development and initial experience in metastatic ovarian cancer. *Eur Radiol* 2021;31:3765–72. https://doi.org/10.1007/s00330-020-07560-8.
37. Cosper PF, McNair C, González I *et al.* Decreased local immune response and retained HPV gene expression during Chemoradiotherapy are associated with treatment resistance and death from cervical cancer. *Int J Cancer* 2020;146:2047–58. https://doi.org/10.1002/IJC.32793.
38. Klopp AH, Jhingran A, Ramdas L *et al.* Gene expression changes in cervical squamous cell carcinoma after initiation of chemoradiation and correlation with clinical outcome. *Int J Radiat Oncol Biol Phys* 2008;71:226–36. https://doi.org/10.1016/j.ijrobp.2007.10.068.
39. Choi CH, Chung JY, Kang JH *et al.* Chemoradiotherapy response prediction model by proteomic expressional profiling in patients with locally advanced cervical cancer. *Gynecol Oncol* 2020;157:437–43. https://doi.org/10.1016/j.ygyno.2020.02.017.
40. Katanyoo K, Sanguanrungsirikul S, Manusirivithaya S. Comparison of treatment outcomes between squamous cell carcinoma and adenocarcinoma in locally advanced cervical cancer. *Gynecol Oncol* 2012;125:292–6. https://doi.org/10.1016/j.ygyno.2012.01.034.
41. Kumar S, Shah JP, Bryant CS *et al.* Prognostic significance of keratinization in squamous cell cancer of uterine cervix: a population based study. *Arch Gynecol Obstet* 2009;280:25–32. https://doi.org/10.1007/s00404-008-0851-9.
42. Someya M, Tsuchiya T, Fukushima Y *et al.* Prediction of treatment response from the microenvironment of tumor immunity in cervical cancer patients treated with chemoradiotherapy. *Med Mol Morphol* 2021;54:245–52. https://doi.org/10.1007/s00795-021-00290-w.
43. Martins PR, Machado CMT, Coxir SA *et al.* Cervical cancer patients that respond to chemoradiation therapy display an intense tumor infiltrating immune profile before treatment. *Exp Mol Pathol* 2019;111:104314. https://doi.org/10.1016/j.yexmp.2019.104314.
44. Datta NR, Stutz E, Liu M *et al.* Concurrent chemoradiotherapy vs. radiotherapy alone in locally advanced cervix cancer: a systematic review and meta-analysis. *Gynecol Oncol* 2017;145:374–85. https://doi.org/10.1016/j.ygyno.2017.01.033.

45. Quinn BA, Deng X, Colton A *et al.* Increasing age predicts poor cervical cancer prognosis with subsequent effect on treatment and overall survival. *Brachytherapy* 2019;18:29–37. https://doi.org/10.1016/j.brachy.2018.08.016.
46. Rajpurkar P, Irvin J, Zhu K *et al.* CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning*arXiv preprint*, arXiv:1711.05225. 2017. https://doi.org/10.48550/arXiv.1711.05225.
47. Afshar P, Mohammadi A, Plataniotis KN. Brain Tumor Type Classification via Capsule Networks. In: *2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece*. IEEE, New York, NY, USA, 2018, pp. 3129–3133. https://doi.org/10.1109/ICIP.2018.8451379
48. Zhao L, Mammadov M, Yearwood J. From Convex to Nonconvex: A Loss Function Analysis for Binary Classification. In: *2010 IEEE International Conference on Data Mining Workshops, Sydney, NSW, Australia*. IEEE, New York, NY, USA, 2010, pp. 1281–1288. https://doi.org/10.1109/ICDMW.2010.57
49. Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. *Comput Struct Biotechnol J* 2018;16:34–42. https://doi.org/10.1016/j.csbj.2018.01.001.