European Radiology (2019) 29:6741–6749
https://doi.org/10.1007/s00330-019-06265-x

MOLECULAR IMAGING

Prediction of local relapse and distant metastasis in patients with definitive chemoradiotherapy-treated cervical cancer by deep learning from [^{18}F]-fluorodeoxyglucose positron emission tomography/computed tomography

Wei-Chih Shen[1] · Shang-Wen Chen[2,3,4] · Kuo-Chen Wu[1] · Te-Chun Hsieh[5,6] · Ji-An Liang[2,7] · Yao-Ching Hung[3,8] · Lian-Shung Yeh[3,8] · Wei-Chun Chang[3,8] · Wu-Chou Lin[3,8] · Kuo-Yang Yen[5,6] · Chia-Hung Kao[5,7,9]

Received: 4 January 2019 / Revised: 18 April 2019 / Accepted: 3 May 2019 / Published online: 27 May 2019

Abstract

Background We designed a deep learning model for assessing ^{18}F-FDG PET/CT for early prediction of local and distant failures for patients with locally advanced cervical cancer.

Methods All 142 patients with cervical cancer underwent ^{18}F-FDG PET/CT for pretreatment staging and received allocated treatment. To augment the amount of image data, each tumor was represented as 11 slice sets each of which contains 3 2D orthogonal slices to acquire a total of 1562 slice sets. In each round of k-fold cross-validation, a well-trained proposed model and a slice-based optimal threshold were derived from a training set and used to classify each slice set in the test set into the categories of with or without local or distant failure. The classification results of each tumor were aggregated to summarize a tumor-based prediction result.

Results In total, 21 and 26 patients experienced local and distant failures, respectively. Regarding local recurrence, the tumor-based prediction result summarized from all test sets demonstrated that the sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 71%, 93%, 63%, 95%, and 89%, respectively. The corresponding values for distant metastasis were 77%, 90%, 63%, 95%, and 87%, respectively.

Conclusion This is the first study to use deep learning model for assessing ^{18}F-FDG PET/CT images which is capable of predicting treatment outcomes in cervical cancer patients.

Wei-Chih Shen and Shang-Wen Chen contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-019-06265-x) contains supplementary material, which is available to authorized users.

✉ Chia-Hung Kao
d10040@mail.cmuh.org.tw; dr.kaochiahung@gmail.com

[1] Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan

[2] Department of Radiation Oncology, China Medical University Hospital, Taichung, Taiwan

[3] School of Medicine, College of Medicine, China Medical University, Taichung, Taiwan

[4] Department of Radiology, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[5] Department of Nuclear Medicine and PET Center, China Medical University Hospital, Taichung, Taiwan

[6] Department of Biomedical Imaging and Radiological Science, China Medical University, Taichung, Taiwan

[7] Graduate Institute of Biomedical Sciences, School of Medicine, College of Medicine, China Medical University, No. 2, Yuh-Der Road, Taichung 404, Taiwan

[8] Department of Obstetrics and Gynecology, China Medical University Hospital, Taichung, Taiwan

[9] Department of Bioinformatics and Medical Engineering, Asia University, Taichung, Taiwan

Key Points

- *This is the first study to use deep learning model for assessing ^{18}F-FDG PET/CT images which is capable of predicting treatment outcomes in cervical cancer patients.*
- *All 142 patients with cervical cancer underwent ^{18}F-FDG PET/CT for pretreatment staging and received allocated treatment. To augment the amount of image data, each tumor was represented as 11 slice sets each of which contains 3 2D orthogonal slices to acquire a total of 1562 slice sets.*
- *For local recurrence, all test sets demonstrated that the sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 71%, 93%, 63%, 95%, and 89%, respectively. The corresponding values for distant metastasis were 77%, 90%, 63%, 95%, and 87%, respectively.*

Keywords AI (artificial intelligence) · Neural network models · Fluorodeoxyglucose F18 · PET-CT scan · Cervical cancer

Abbreviations

^{18}F-FDG PET/^{18}F-fluorodeoxyglucose positron emission tomography-computed tomography	
CRT	Chemoradiotherapy
CTV	Clinical target volume
HGRE	High gray-level run emphasis
MTV	Metabolic tumor volume
PLNs	Pelvic lymph nodes
SUV	Standardized uptake value
VOI	Volume of interest

Introduction

Although chemoradiotherapy (CRT) is a mainstay of advanced cervical cancer treatment, patient outcomes have not been satisfactory in terms of local control or distant metastasis [1, 2]. Currently, ^{18}F-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography (PET)–computed tomography (CT) has been commonly used for staging and monitoring the treatment outcomes of various cancers, including cervical cancer [3]. The measurement of intratumoral metabolic heterogeneity using radiomics on ^{18}F-FDG PET/CT has recently been suggested to supplement standard indices for predicting the response of tumors to various cancer therapies [4–12] and might provide certain hints of the underlying tumor phenotype. However, the performance of radiomics highly depends on the methods used in delineating the metabolic tumor volume (MTV) of a tumor and in discretizing the uptake values within the delineated volume [13, 14]. In several PET-related studies for cervical cancer [8–12], the methodologies lack a standard technical approach. As a result, the predictive performance or representative features for outcome prediction were not consistent.

The rapid development of artificial intelligence and its implementation into routine clinical imaging will cause a major transformation to the practice of radiology [15]. Currently, artificial intelligence is already widely employed in various medical roles. One such technique that has exhibited great success is deep learning [16], which can be described as the application of multilayered convolutional neural network (CNN) to solving a wide range of problems [17–22]. Compared with traditional human-engineered radiomics, which strongly relies on feature quantification, CNN functions by learning relevant features directly from image databases. Additionally, the optimal features of conventional radiomics are neither well-defined nor externally validated in other study methods. Although the potential advantages of deep learning in the field of medical imaging are well known, its major challenges include the difficulty clinicians encounter in trying to understand the reasoning of a CNN, which may limit their confidence in its advice [23].

Determining whether patients are predisposed to tumor recurrence is crucial for being able to effectively counsel them on their treatment. In patients with cervical cancer, pertinent studies on finding the relationship between ^{18}F-FDG PET/CT radiomics and CRT-based outcome by using machine learning algorithms are still limited. Additionally, most applications of deep learning to medical images had the objective of identifying the nature of radiological lesions [18, 20–22] not providing predictive information. On the basis of comparing the predictive performance using human-engineered radiomics for cervical cancer, we introduced a novel deep learning model associated with a strategy of data processing for outcome prediction by analyzing baseline ^{18}F-FDG PET/CT. This approach aims to circumvent the major limitation of deep learning: its need for a large quantity of medical images.

Materials and methods

Study population

This retrospective cohort study included 142 patients newly diagnosed with cervical cancer between July 2009 and December 2015. All patients had undergone ^{18}F-FDG PET/CT for pretreatment staging and had received allocated external beam radiotherapy and intracavitary brachytherapy. The eligibility criterion regarding the minimum size for primary tumors was a greater diameter of ≥ 2 cm on a CT scan because

small tumor volume would be a confounding factor in the radiomic analyses [24]. This study was approved by a local institutional review board (DMR99-IRB-010(CR6)). Tumors were staged in accordance with the International Federation of Gynecology and Obstetrics (FIGO) staging system and observed that 37, 78, and 27 patients had stages I, II, and III cervical cancer, respectively. The histological type was squamous cell carcinoma in 114 patients and non-squamous cell carcinoma in 28 patients. The diagnoses of lymph node metastases were made using PET/CT. In 37 patients with stage I disease, 6 were diagnosed to have stage IB1 tumors, and the reasons of receiving CRT were old age at the diagnosis ($n = 4$) or presence of pelvic lymph node metastasis ($n = 2$). We excluded patients who presented with para-aortic lymph node metastasis on the ^{18}F-FDG PET/CT scan because they were categorized as distant nodal metastasis. The patients' characteristics are listed in Table 1.

Study endpoints

This study focused on a novel deep learning model (DLM) for outcome prediction. We categorized treatment failures into two types of events that could jeopardize survival, namely local relapse and distant metastasis. Local relapse was defined as evidence of central residual or recurrent tumors discovered after CRT; distant metastasis was defined as evidence of recurrent tumors at non-regional lymph node or distant organs [2]. A diagnosis of residual or recurrent local disease was based on radiographic examinations or pathological confirmation, whereas distant metastases were mostly confirmed by PET/CT.

Table 1 Patient characteristics ($N = 142$)

Variables	Value
Age (year)	Median 55 (range, 28~81)
FIGO stage	
IB1–IB2	37 (26%)
IIA–IIB	78 (55%)
IIIA–IIIB	27 (19%)
Histology	
Squamous cell carcinoma	114 (80%)
Adenocarcinoma	28 (20%)
Pelvic LN metastasis	
Negative	81 (57%)
Positive	61 (43%)
SUV_{max}	Mean 11.3 ± 6.0 (range, 2.9~37.0)
MTV (mL)	Mean 28.1 ± 42.7 (range, 2.5~450.0)

FIGO, International Federation of Gynecology and Obstetrics; *LN*, lymph node

PET/CT imaging

A baseline PET/CT scanner (PET/CT-16 slice; Discovery STE; GE Medical System) was applied for all patients. The scans were performed approximately 60 min after the administration of 370 MBq of ^{18}F-FDG before more than 4 h of fasting [25]. The uptake of ^{18}F-FDG was determined by calculating the standardized uptake value (SUV). The matrix size was 128 voxels × 128 voxels × length of interest, and the voxel size was 5.47 mm × 5.47 mm × 3.27 mm.

Overview of workflow

The workflow of this study was arranged as training and test processes based on k-fold cross-validation, as illustrated in Fig. 1. A data augmentation approach that slices each tumor into 11 slice sets was employed to reduce the limitation of sample size. The patients were randomly partitioned into k sets to perform cross-validation. In each round of cross-validation, k-1 sets were aggregated to train a DLM (Fig. 1a). After training, each slice set in the training set was input to the DLM to determine the probability of event occurrence. A slice-based optimal threshold was derived from all probabilities to determine whether a slice set had the event occurrence. Subsequently, the remaining set was regarded as a test set to verify the trained DLM (Fig. 1b). Each slice set in the test set was input to the DLM to acquire a probability of event occurrence and classified as a slice-based prediction according to the slice-based optimal threshold. For each tumor, the slice-based predictions were summarized as a tumor-based probability of event occurrence. Finally, all tumor-based probabilities were used to derive a tumor-based optimal threshold, which was used to determine whether a patient had event occurrence.

Data augmentation approach

The volume of interest (VOI) centered by SUV_{max} and with the dimensions of 7 cm × 7 cm × 7 cm was extracted from a PET image. A voxel was defined as a local maximum if its SUV was not less than those of neighboring voxels. For an input PET image, the area of maximum uptake was automatically detected. The highest SUV within the tumor was manually identified and defined as SUV_{max}, which served as a label in later processes. Interpolation was employed to improve the resolution of PET images. The original PET images were interpolated to a resolution of 2.5 mm × 2.5 mm × 2.5 mm by using a splice method. Data processing and augmentation are presented in Fig. 2.

As the data augmentation technique for deep learning has been employed for CT scan [26], a novel augmentation approach for PET/CT was proposed to expand the limited sample size (Fig. 2b, c). The principle of augmentation was to balance the solution for a lack of sufficient training data and

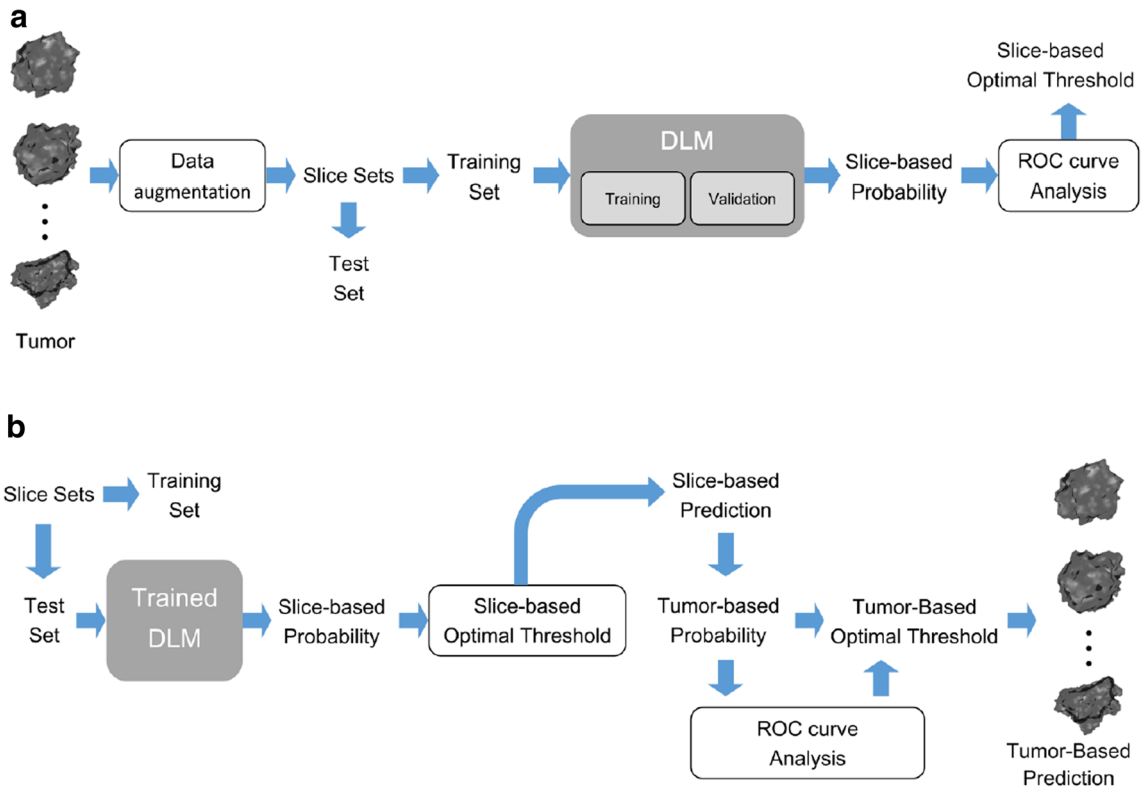

Fig. 1 Workflow of the proposed training process (**a**) and test process (**b**)

the preservation of FDG uptake heterogeneity surrounding the SUV_{max}. Each tumor was re-sampled into one base slice set and 10 rotated slice sets for augmenting the amount of imaging data. The orthogonal XY, XZ, and YZ planes passing through SUV_{max} were used to sample SUVs from the VOI and were stacked to form a base slice set. The aim of stacking three orthogonal slices was to partially preserve the 3D heterogeneity information. Then, the XY plane of the base slice set was rotated by 15, 30, 45, 60, 75, 105, 120, 135, 150, and 165° counterclockwise around the *y*-axis to acquire ten slice sets. A total of 1562 slice sets from 142 tumors (11 sets for each) were acquired. Besides, each slice set was labeled with the status of two events, local relapse or distant metastasis, by the corresponding patient's outcome.

Deep learning model

The proposed deep learning model was based on the architecture Network in Network [27]. The model consisted of three mlpconv layers and one global-average pooling layer (Fig. 3). Batch normalization (BN) was employed to reduce internal covariate shifts in this model to prevent overfitting [28]. Because the model was trained by slice sets, the output of the model was a slice-based probability of event occurrence.

The proposed model was implemented using Matlab 2015b, Python 3.6.4, and Keras 2.1.3 on a workstation installed with an NVIDIA Quadro M4000.

Statistical method

Receiver operating characteristic (ROC) curve analysis was performed to calculate predictive performance. Additionally, the ROC curve was used to ascertain the optimal threshold for maximizing prediction accuracy. The predictive indices were sensitivity (SE), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), and accuracy. The statistical analysis was performed using MedCalc 18.2.1.

Experimental method

K-fold cross-validation was employed to verify the robustness of the proposed model. The patients were also randomly grouped into seven sets, each containing a similar number of patients with or without event occurrence. Each set was regarded as a test set only once, and the remaining sets were merged together to form the training set for training the model (Fig. 1a). During the training process, all slice sets contained in the training set were partitioned into training and validation sets with a ratio of 8:2. The model with the highest performance was preserved by checking val_loss. The number of epochs was set to 500. After training, an optimal threshold was derived from the probabilities of event prediction in all slice sets of the training set by using ROC curve analysis (Fig. 1a). Then, each slice set in the test set was input to the trained model to acquire the probability of event occurrence (Fig. 1b). The slice sets were classified into two

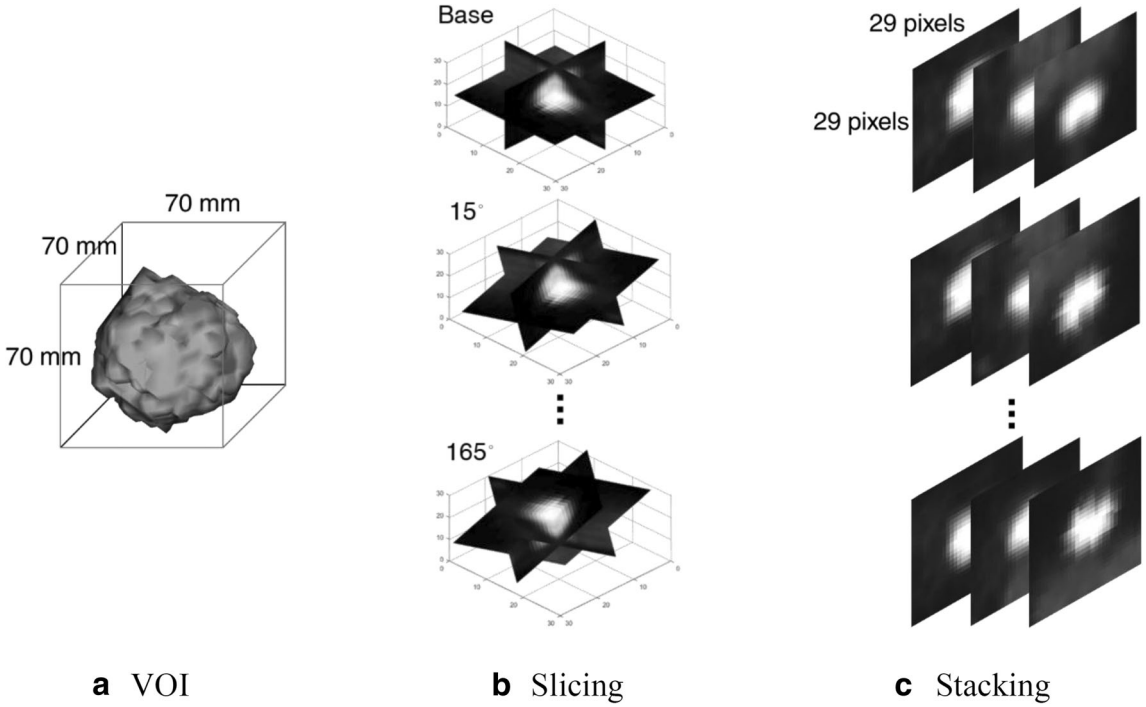

Fig. 2 Data processing and augmentation: For each tumor, the coordinate of labeled SUV_{max} in the original PET image was converted into a new coordinate in the interpolated image. Then, a volume of interest (VOI), centered by SUV_{max} and with dimensions of 70 mm × 70 mm × 70 mm, was fetched from the interpolated image (**a**). The orthogonal XY, XZ, and YZ planes passing through SUV_{max} were used to sample SUVs from the VOI and were stacked to form a base slice set (**b**, **c**)

groups with or without occurrence by their correlation with the optimal threshold value. Accordingly, the slice-based performance of the proposed model was acquired. For each tumor in the test set, the frequency belonging to the group with events was aggregated and then divided by 11 to acquire a tumor-based probability. An optimal threshold for identifying events was derived from all tumor-based probabilities in the test set by using ROC analysis, which led to an assessment of the tumor-based performance of the model. Finally, the slice-based performance and tumor-based performance in all test sets were respectively merged to establish the overall performance of the DLM.

Calculation of human-engineered radiomic features

The definition of MTV, discretization of SUVs within the MTV, and requirements for calculating textural indices have already been detailed in Appendix 1. ROC curves were plotted to determine the highest predictive performance among the various textural indices.

Treatment

All the patients received the exact same treatment composed of external beam radiotherapy and brachytherapy regardless of the tumor stage or histological type. All patients were treated with intensity-modulated radiotherapy technique. The total dose applied to the pelvis was 45 Gy, administered in 25 fractions over a 5-week period. Following pelvic irradiation, the bilateral parametrium was boosted from 50.4 to 54 Gy [29]. In patients presented with metastatic pelvic lymph nodes, the involved nodes were further irradiated with IMRT to a cumulative dose of 60 to 64 Gy.

When cervical tumor regressed, high-dose-rate intracavitary brachytherapy was performed once or twice a week by using an

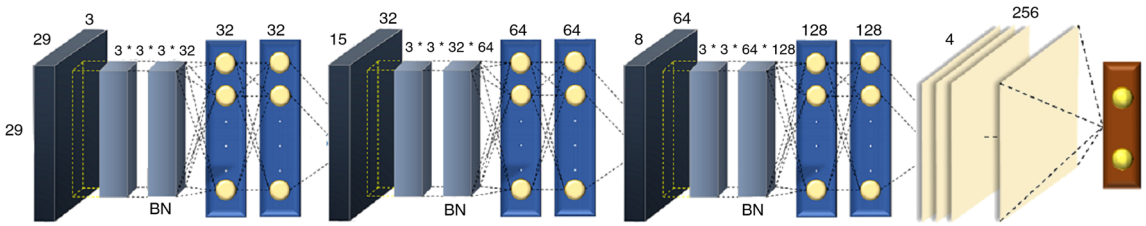

Fig. 3 Overall structure of the proposed deep learning model. "BN" indicates the batch normalization layer

Ir-192 remote afterloading technique concurrently with pelvic irradiation or parametrial boosting. Before January 2013, the standard prescribed dose of each session of brachytherapy was 6.0 Gy to point A, with 4 to 5 sessions. After January 2013, 77 patients were treated with three-dimensional (3D) image-based brachytherapy according to the recommendations of the Groupe Européen de Curiethérapie and the guidelines specified by the European Society for Radiotherapy and Oncology [30]. The mean high-risk CTV (HR-CTV) was 32.8 ± 17.5 mL, whereas the mean cumulative EQD2 of D90 of HR-CTV (Gy_{10}) was 87.1 ± 9.2 Gy.

Chemotherapy consisted of weekly 40 mg/m^2 doses of cisplatin, administered intravenously and accumulating to a total dose of 60 mg.

Follow-up

After completion of radiotherapy, the patients were regularly followed up every 2 months for the first year, and every 3 to 4 months thereafter. Besides a routine pelvic examination, the levels in serum of tumor markers, namely squamous cell carcinoma and carcinoembryonic antigens, were examined during each follow-up. Additionally, a radiographic examination was performed every 6 months.

Results

Treatment outcomes

After a median follow-up duration of 40 months, 101 patients displayed no evidence of disease progression, whereas 41 patients did have disease progression (infield recurrence, distant metastasis, and both in 21, 26, and 9 patients, respectively). Overall, 21 and 26 patients experienced local and distant failures, respectively. The tumors were categorized as the events.

Partitioning of patients

The 142 patients included in this study were randomly partitioned into seven sets on the basis of the number of events that occurred, each set containing three patients with local relapse and 17 or 18 patients without event occurrence. Likewise, the cohort was randomly partitioned into seven sets with the same ratio of patients presenting with and without distant metastasis.

Heatmap

A heatmap was employed to visually observe the discriminative regions used by proposed DLM to identify a specific event in a slice set. The uterus is adjacent to anatomic structures, such as the bladder, colon, and lymph nodes; these organs, if they also exhibit a high uptake of ^{18}F-FDG, could disrupt the focus of the deep learning model and lead to inaccurate analyses. Thus, the heatmap represented the activations of slice sets in the last layer of our proposed model (Fig. 4). The heatmaps demonstrate a remarkable ability to differentiate the uterus from adjacent anatomic structures. Moreover, the characteristics that signified event occurrences were derived from the areas of the cervical tumors.

Slice-based prediction

Regarding local recurrence, the classification results of all slice sets in each test set are summarized in Appendices 2 and 3. All SPs were greater than 88%, and SE in five of the sets ranged from 36 to 58%. When the classification results of all slice sets were summarized, the SE, SP, PPV, NPV, and accuracy were 35%, 93%, 45%, 89%, and 84%, respectively.

Regarding distant metastasis, the classification results of all slice sets in each test set are presented in Appendices 4 and 5. All SPs were greater than 76%, and six of the SEs were greater than 33%. After summarizing the results of all slice sets, the SE, SP, PPV, NPV, and accuracy were 51%, 89%, 50%, 89%, and 82%, respectively.

Tumor-based prediction

The classification results regarding local recurrence for all tumors in the seven test sets are summarized in Table 2. The SE, SP, PPV, NPV, and accuracy were 71%, 93%, 63%, 95%, and 89%, respectively. The high SP and NPV exhibited by the model demonstrated its strong ability to reject the possibility of local recurrence after treatment.

Regarding distant metastasis, the classification results for all tumors are summarized in Table 3. The SE, SP, PPV, NPV, and accuracy were 77%, 90%, 63%, 95%, and 87%, respectively. The high NPV indicated that the model accurately excluded the development of distant metastasis.

Comparison with the performance of human-engineered radiomic features

Four groups of radiomic features were calculated, as presented in Appendix 1. ROC analysis revealed that the high gray-level run emphasis (HGRE) derived from the gray-level run-length matrix most accurately and consistently predicted local relapse. HGRE values demonstrated an inverse correlation with local recurrence, and the area under the ROC curve was 0.31 ± 0.06 ($p = .006$).

None of the radiomic features demonstrated the ability to predict distant metastasis. One potential reason for this is that, because of technical difficulties, the textures of metastatic lymph nodes were not included in the analysis. In addition, clinical stage was not a predictor for local relapse or distant

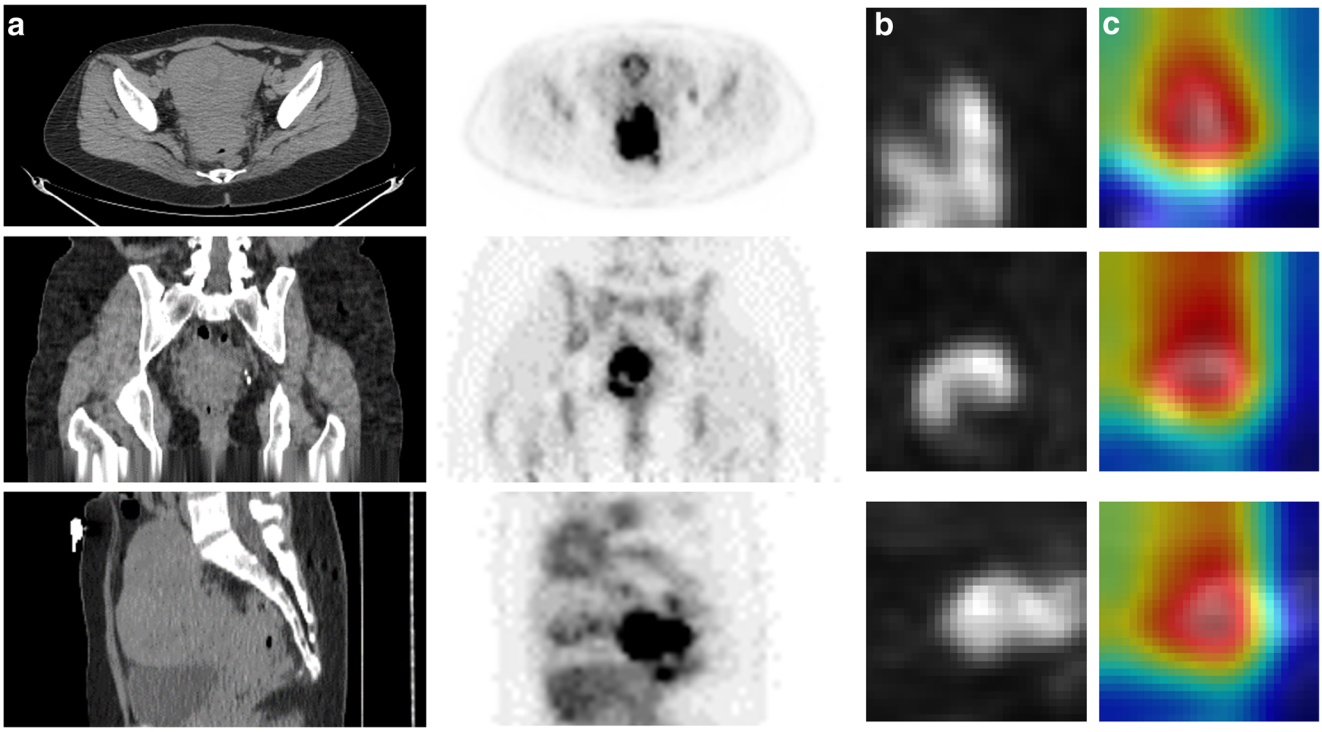

Fig. 4 Representative heatmap. PET/CT images (**a**). Base slice set (**b**). Heatmap (**c**)

metastasis. The information about the outcomes in relation to the baseline stage is presented in Appendix 6.

Discussion

Simple machine learning methods are easily applied to problems with well-structured data or well-defined optimal features [18]. To date, neither the optimal radiomic features for specific biological endpoints nor the most powerful discretization approach for radiomic analysis has been validated. Therefore, the introduction of deep learning techniques in cancer treatment can assist oncologists in creating a new approach to the classification of imaging phenotypes. However, the most prominent limitation of such an approach is that deep learning is an intensely data-hungry technology; learning from scratch the weights in a large network requires numerous labeled samples to achieve accurate classification [18]. Its augmentation of data by rotating the plane of labeled images, as well as the merging of slice-based and tumor-based models, distinguishes this study as the first to conclude that deep learning applied to ^{18}F-FDG PET/CT images can play a crucial role in predicting treatment outcomes, even when the model is trained using a small sample size. The deep learning model can provide previsional information regarding the risk patients have for treatment failures.

The higher NPVs compared with the PPVs poses a challenge for establishing clinical utility. One plausible cause for this discrepancy is that only a small percentage of the studied patients experienced local or distant failures. In general, a positive correlation exists between PPV and disease prevalence, in the sense that PPV is higher in people with higher disease prevalence, whereas NPV is negatively associated with disease prevalence. To maximize the predictive ability of the model, our future goal is to reassess the PPV by increasing the sample size or event number.

A well-known pitfall of machine learning is overfitting, where a model learns individual statistical variations of the training set rather than generalized patterns for a particular

Table 2 Classification results of tumors in all seven test sets, with comparison of patients with and without local recurrence

Prediction	Local recurrence		Indices
	P	N	
P	15	9	63%
N	6	112	95%
Indices	71%	93%	89%

Table 3 Classification results of tumors in all seven test sets, with comparison of patients with and without distant metastasis

Prediction	Distant Metastasis		Indices
	P	N	
P	20	12	63%
N	6	104	95%
Indices	77%	90%	87%

feature. To circumvent this potential bias, k-fold cross-validation was employed to verify the robustness of our model. Because of the limited number of patients presenting with events, the value of k was set to seven. The prediction results derived from merging the seven test sets demonstrated the robustness of our model's ability to identify both local recurrence and distant metastasis.

Delineation of the spatial extent of a tumor was defined as the identification of pixels or voxels composing a particular target structure. In machine learning models, delineation is regarded a pixel-level classification task, the success of which depends on the criteria for determining whether a given pixel belongs to the background or a target lesion. In human-engineered radiomics, MTVs being connected to the bladder pose a major problem for most segmentation methods [24]. Deep learning models overcome these confounding factors. By assessing the heatmap generated from the input layer, we discovered that the model selectively focused on FDG hotspots in PET images and ignored the adjacent organs. Thus, it was beneficial to explore the biological characteristics on the heatmap that were accepted by the deep learning model.

Because this study used a limited data set from a single institution, our findings should be interpreted cautiously. External validation studies that produce similar training findings by using independent imaging data, and a range of different scanners, resolution settings, and reconstruction algorithms are necessary to confirm our results. Additionally, disparities between studies might be a result of variance in training- or test-model structures. Therefore, more studies are required to verify the prediction capability of the deep learning model. Furthermore, a general approach to avoid overfitting on deep learning model is to artificially enlarge the dataset [16]. A useful technique to duplicate the imaging data is to rotated, flip, shift, crop, zoom in or out, or distort the figures. In this study, a novel augmentation approach was proposed to represent the VOI by stacking three orthogonal slices acquired from different angles. Because each slice is sampled from the VOI, the concern about the validity of label assignments could be minimized. Moreover, the 3D heterogeneity was partially preserved on the slice set. As a result, the information loss from transforming VOI to 2D slices could be reduced. Nonetheless, validation studies are certainly required to ensure that the diverse augmentations did not change the validity of label assignments. Finally, to establish the clinical utility, it is imperative to compare this model with the standard clinical and imaging assessment such as tumor size, tumor volume, FIGO stage, and histological subtype. Nevertheless, our findings constitute a pivotal step toward being able to tailor CRT to specific imaging phenotypes for patients with cervical cancer. Regardless, CNN-based features must be correlated with any specific underlying biological mechanism. Particularly, it would be interesting to know why some patients experienced both local and distant recurrences. Future validation studies should integrate imaging phenotypes based on deep learning with other documented clinical parameters to give oncologists the ability to rapidly assess the feasibility of salvage treatments for high-risk patients.

Conclusion

In patients with definitive-CRT-treated uterine cervical cancer, the proposed DLM for assessing ^{18}F-FDG PET/CT images obtained pretreatment can predict treatment outcome. By using the slice-based model for deep learning to augment collected data, high accuracy and NPV can be achieved even with a limited sample size. Future external validation studies with independent data sets are required for establishing the model's clinical utility.

Funding This work was supported by grants from the Ministry of Health and Welfare, Taiwan (MOHW107-TDU-B-212-123004); China Medical University Hospital (DMR-107-192, CRS-106-036, CRS106-039, CRS106-040, CRS106-041); Asia University (DMR-106-150); Academia Sinica Stroke Biosignature Project (BM10701010021); MOST Clinical Trial Consortium for Stroke (MOST 107-2321-B-039-004-); Tseng-Lien Lin Foundation, Taichung, Taiwan; and Katsuzo and Kiyo Aoshima Memorial Funds, Japan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding was received for this study.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Chia-Hung Kao, MD, Graduate Institute of Biomedical Sciences and School of Medicine, College of Medicine, China Medical University, No. 2, Yuh-Der Road, Taichung 404, Taiwan. E-mail: d10040@mail.cmuh.org.tw; dr.kaochiahung@gmail.com

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors (Prof. Shang-Wen Chen) has significant statistical expertise.

Informed consent This is a retrospective study for images' analyses. The IRB also specifically waived the consent requirement.

Ethical approval This study was approved by the local institutional review board (certificate numbers CMUH102-REC2-74 and DMR99-IRB-010-1).

Study subjects or cohorts overlap This work was partially presented at NVIDIA GTC Taiwan 2018 Poster Contest.

Methodology
- Retrospective
- Diagnostic or prognostic study
- Performed at one institution

References

1. Rose PG, Java J, Whitney CW et al (2015) Nomograms predicting progression-free survival, overall survival, and pelvic recurrence in locally advanced cervical cancer developed from an analysis of identifiable prognostic pactors in patients from NRG Oncology/Gynecologic Oncology Group randomized trials of chemoradiotherapy. J Clin Oncol 33:2136–2142
2. Kang S, Nam BH, Park JY et al (2012) Risk assessment tool for distant recurrence after platinum-based concurrent chemoradiation in patients with locally advanced cervical cancer: a Korean Gynecologic Oncology Group Study. J Clin Oncol 30:2369–2374
3. Kidd EA, Siegel BA, Dehdashti F, Grigsby PW (2007) The standardized uptake value for F18 fluoro-deoxyglucose is a sensitive predictive biomarker for cervical cancer treatment response and survival. Cancer 110:1738–1744
4. Tixier F, Le Rest CC, Hatt M et al (2011) Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med 52:369–378
5. Cook GJ, Yip C, Siddique M et al (2013) Are pretreatment 18F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? J Nucl Med 54:19–26
6. Hatt M, Majdoub M, Vallières M et al (2015) ^{18}F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. J Nucl Med 56:38–44
7. Ohri N, Duan F, Snyder BS et al (2016) Pretreatment 18FDG-PET textural features in locally advanced non-small cell lung cancer: secondary analysis of ACRIN 6668/RTOG 0235. J Nucl Med 57:842–848
8. Kidd EA, Grigsby PW (2008) Intratumoral metabolic heterogeneity of cervical cancer. Clin Cancer Res 14:5236–5524
9. Yang F, Thomas MA, Dehdashti F, Grigsby PW (2013) Temporal analysis of intratumoral metabolic heterogeneity characterized by textural features in cervical cancer. Eur J Nucl Med Mol Imaging 40:716–727
10. Ho KC, Fang YH, Chung HW et al (2016) A preliminary investigation into textural features of intratumoral metabolic heterogeneity in FDG PET for overall survival prognosis in patients with bulky cervical cancer treated with definitive concurrent chemoradiotherapy. Am J Nucl Med Mol Imaging 6:166–175
11. Lucia F, Visvikis D, Desseroit MC et al (2018) Prediction of outcome using pretreatment ^{18}F-FDG PET/CT and MRI radiomics in locally advanced cervical cancer treated with chemoradiotherapy. Eur J Nucl Med Mol Imaging 45:768
12. Chen SW, Shen WC, Hsieh TC et al (2018) Textural features of cervical cancers on FDG-PET/CT associate with survival and local relapse in patients treated with definitive chemoradiotherapy. Sci Rep 8:11859
13. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D (2017) Characterization of PET/CT images using texture analysis: the past, the present any future? Eur J Nucl Med Mol Imaging 44:151–165
14. Sollini M, Cozzi L, Antunovic L, Chiti A, Kirienko M (2017) PET Radiomics in NSCLC: state of the art and a proposal for harmonization of methodology. Sci Rep 7:358
15. Liew C (2018) The future of radiology augmented with artificial intelligence: a strategy for success. Eur J Radiol 102:152–156
16. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1097–1105
17. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444
18. Chartrand G, Cheng PM, Vorontsov E et al (2017) Deep learning: a primer for radiologists. Radiographics 37:2113–2131
19. Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542:115–118
20. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S (2016) Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans Med Imaging 35:1207–1216
21. Kallenberg M, Petersen K, Nielsen M et al (2016) Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. IEEE Trans Med Imaging 35:1322–1331
22. Setio AA, Ciompi F, Litjens G et al (2016) Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans Med Imaging 35:1160–1169
23. Fazal MI, Patel ME, Tye J, Gupta Y (2018) The past, present and future role of artificial intelligence in imaging. Eur J Radiol 105:246–250
24. Brooks FJ, Grigsby PW (2014) The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. J Nucl Med 55:37–42
25. Delbeke D, Coleman RE, Guiberteau MJ et al (2006) Procedure guideline for tumor imaging with 18F-FDG PET/CT 1.0. J Nucl Med 47:885–895
26. Yasaka K, Akai H, Abe O, Kiryu S (2018) Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. Radiology 286:887–896
27. Lin M, Chen Q, Yan S (2014) Network In Network. arXiv: 1312.4400 [cs.NE] Available via https://arxiv.org/abs/1312.4400
28. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv: 1502.03167 [cs.LG] Avaliable via https://arxiv.org/abs/1502.03167
29. Chen SW, Liang JA, Hung YC et al (2013) Does initial 45Gy of pelvic intensity-modulated radiotherapy reduce late complications in patients with locally advanced cervical cancer? A cohort control study using definitive chemoradiotherapy with high-dose rate brachytherapy. Radiol Oncol 47:176–184
30. Pötter R, Haie-Meder C, Van Limbergen E et al (2006) Recommendations from gynaecological (GYN) GEC ESTRO working group (II): concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy-3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology. Radiother Oncol 78:67–77