

Evaluation of the accuracy of heart dose prediction by machine learning for selecting patients not requiring deep inspiration breath-hold radiotherapy after breast cancer surgery

RYO KAMIZAKI^{1,2*}, MASAHIRO KURODA^{1*}, WLLA E. AL-HAMMAD^{3,4*},
 NOUHA TEKIKI³, HINATA ISHIZAKA¹, KAZUHIRO KURODA^{1,5}, KOHEI SUGIMOTO^{1,6},
 MASATAKA OITA⁶, YOSHINORI TANABE¹, MAJD BARHAM⁷, IRFAN SUGIANTO⁸, YUKI NAKAMITSU¹,
 MASAKI HIRANO^{1,9}, YUKI MUTO^{1,10}, HIROKI IHARA¹¹ and SOICHI SUGIYAMA¹²

¹Department of Radiological Technology, Graduate School of Health Sciences, Okayama University, Okayama 700-8558; ²Department of Radiology, Matsuyama Red Cross Hospital, Matsuyama, Ehime 790-8524; ³Department of Oral and Maxillofacial Radiology, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama 700-8558, Japan; ⁴Department of Oral Medicine and Oral Surgery, Faculty of Dentistry, Jordan University of Science and Technology, Irbid 22110, Jordan; ⁵Department of Health and Welfare Science, Graduate School of Health and Welfare Science, Okayama Prefectural University, Okayama 719-1197; ⁶Graduate School of Interdisciplinary Sciences and Engineering in Health Systems, Okayama University, Okayama 770-8558, Japan; ⁷Department of Dentistry and Dental Surgery, College of Medicine and Health Sciences, An-Najah National University, Nablus 44839, Palestine; ⁸Department of Oral Radiology, Faculty of Dentistry, Hasanuddin University, Sulawesi 90245, Indonesia; ⁹Department of Radiology, Osaka Red Cross Hospital, Osaka 543-8555; ¹⁰Department of Radiology, Oomoto Hospital, Okayama 700-0924; Departments of ¹¹Radiology and ¹²Proton Beam Therapy, Faculty of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama 700-8558, Japan

Received June 8, 2023; Accepted September 1, 2023

DOI: 10.3892/etm.2023.12235

Abstract. Increased heart dose during postoperative radiotherapy (RT) for left-sided breast cancer (BC) can cause cardiac injury, which can decrease patient survival. The deep

inspiration breath-hold technique (DIBH) is becoming increasingly common for reducing the mean heart dose (MHD) in patients with left-sided BC. However, treatment planning and DIBH for RT are laborious, time-consuming and costly for patients and RT staff. In addition, the proportion of patients with left BC with low MHD is considerably higher among Asian women, mainly due to their smaller breast volume compared with that in Western countries. The present study aimed to determine the optimal machine learning (ML) model for predicting the MHD after RT to pre-select patients with low MHD who will not require DIBH prior to RT planning. In total, 562 patients with BC who received postoperative RT were randomly divided into the training (n=449) and external (n=113) test datasets for ML using Python (version 3.8). Imbalanced data were corrected using synthetic minority oversampling with Gaussian noise. Specifically, right-left, tumor site, chest wall thickness, irradiation method, body mass index and separation were the six explanatory variables used for ML, with four supervised ML algorithms used. Using the optimal value of hyperparameter tuning with root mean squared error (RMSE) as an indicator for the internal test data, the model yielding the best F2 score evaluation was selected for final validation using the external test data. The predictive ability of MHD for true MHD after RT was the highest among all algorithms for the deep neural network, with a RMSE of 77.4, F2 score of 0.80 and area under the curve-receiver operating characteristic of 0.88, for a cut-off value of 300 cGy.

Correspondence to: Professor Masahiro Kuroda, Department of Radiological Technology, Graduate School of Health Sciences, Okayama University, 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8558, Japan
 E-mail: kurodamd@cc.okayama-u.ac.jp

*Contributed equally

Abbreviations: AI, artificial intelligence; BC, breast cancer; BMI, body mass index; CV, cross validation; CWT, chest wall thickness; DIBH, deep inspiration breath-hold technique; DNN, deep neural network; η , Eta analysis correlation coefficient; FIF, field-in-field method; FIF-IRP, one-reference point FIF method; FIF-2RP, FIF with two RP; FN, false negative; MHD, mean heart dose; ML, machine learning; RF, random forest; RMSE, root mean squared error; r_s , Spearman's correlation coefficient; RT, radiotherapy; SEP, separation; SMOGN, synthetic minority oversampling with Gaussian noise; W, wedge method; XGB, extreme gradient boosting

Key words: BC, RT, heart dose, ML, DNN, DIBH

The present study suggested that ML can be used to pre-select female Asian patients with low MHD who do not require DIBH for the postoperative RT of BC.

Introduction

Postoperative radiotherapy (RT) forms a significant component in the standard treatment regimen of breast cancer (BC). Postoperative RT for early-stage BC can not only reduce postoperative recurrence, but it can also improve the survival rate (1,2). However, increased cardiac irradiation dose during postoperative RT in patients with left-sided BC has been reported to cause cardiac injury, which is classified as a late adverse event and can lead to a decreased survival rate (3). Therefore, the deep inspiration breath-hold technique (DIBH) is becoming an increasingly preferred method for reducing the mean heart dose (MHD) in patients with left-sided BC (4-6). In DIBH, the patients are requested to hold their breath with deep inspiration during irradiation, so that the lungs between the anterior chest wall and heart are filled with air (7). This causes the heart to move away from the anterior chest wall, which then reduces the irradiation dose to the heart (4-6). However, treatment planning for DIBH is time-consuming, laborious and costly for both the patient and the radiation therapist in terms of treatment planning and implementation (8). In particular, Asian countries tend to have fewer personnel and facilities for RT compared with the United States (9,10), limiting the application of DIBH. Furthermore, although MHD is greatly influenced by breast volume (11,12), patients of Asian ethnicities such as Japanese (12), Korean (13) and Chinese (14) tend to have less breast volume (12-14) compared with patients in the United States and Europe (11,15,16). Therefore, the MHD also tends to be lower in patients of Asian ethnicities (12) compared with that in the United States and Europe, limiting the number of patients with left BC with high MHD who require DIBH. Only ~26% of Asian patients with left-sided BC receive high MHD (>300 cGy) after left-sided BC RT, with an average MHD of 304 cGy following calculation by the wedge method (W) and 251 cGy by the field-in-field method (FIF) (12). Therefore, it is desirable to select Asian patients who may not require DIBH prior to RT planning for left-sided BC. DIBH can be used more efficiently if MHD can be predicted in advance, where patients with MHD who do not require DIBH can be accurately selected.

Over the past decade, artificial intelligence (AI) and machine learning (ML) techniques are increasing being applied in the field of RT (17-20). However, few studies have applied ML for predicting the MHD during RT using patient information (21-23). Therefore, the purpose of the present study is to compare the various ML models for predicting MHD to select patients who may not require DIBH and to present the optimal model for predicting MHD.

Materials and methods

Patients and RT. The present study included 577 female (a mean age of 55 years and standard deviation of 11 years) patients with BC who received RT at Okayama University Hospital (Okayama, Japan) between April 2009 and March

2016. Fifteen patients were excluded based on the exclusion criteria of some missing data. All patients underwent whole-breast irradiation after partial breast resection, where 167 patients underwent the wedge method and 395 patients underwent the FIF method. In FIF, two types of methods were used. The one-reference point FIF method (FIF-1RP) was used for 142 patients, where one reference point (RP) was set at the mid-level between the upper and lower edges of the irradiation field or 2 cm apart from the deepest point and upper edge of the irradiation field. The other method was the FIF with two RP (FIF-2RP) method (24), which was applied on 253 patients. The FIF-2RP method involves 2RPs set for each patient, specifically one RP for the main beam at a point 2 cm apart from the deepest point and upper edge of the irradiation field, the other RP for the FIF at the mid-level between the upper and lower edges of the irradiation field (24). All patients were irradiated with 2 Gy per fraction, 25 fractions, for a total of 50 Gy. Some patients were irradiated with an additional 10 Gy boost on the tumor bed. The heart dose during the 50-Gy whole-breast irradiation (12) was the subject of the present study.

Data collection. As explanatory variables, data including right-left, tumor site (upper-inner quadrant, lower-inner quadrant, upper-outer quadrant, lower-outer quadrant and central portions) (25), chest wall thickness (CWT), irradiation method (W, FIF-1RP and FIF-2RP), body mass index (BMI), separation (SEP), age, height and weight were collected retrospectively, whilst as an objective variable, MHD (12) was collected retrospectively. CWT and SEP were measured using a nipple-level one-slice simulated CT image for treatment planning (Fig. 1). SEP was defined as the distance along the posterior edge of the tangent fields at the nipple level. CWT was defined as the distance from the nipple surface to the lung, on a perpendicular line of breast separation. Data on the right and left sides, tumor site, irradiation method and BMI were collected from clinical records, whilst MHD was collected from the RT planning system. In the present study, MHD \geq 300 cGy was defined as high MHD, whereas MHD <300 cGy would be defined as low MHD, following the QUANTEC cardiac guidelines (3). There were 76 patients (14%) of high MHD and 486 patients (86%) of low MHD.

Instruments used for ML. Python (version 3.8, Python Software Foundation) and the Python open source ML library; scikit-learn (version 0.24.1, <https://scikit-learn.org/stable/index.html>), TensorFlow (version 1.15.3, <https://www.tensorflow.org/install?hl=ja>) and extreme gradient boosting (XGB, version 1.4.2) (26), were used.

Data partitioning and model building. Fig. 2 shows the data preparation process. All data were split into datasets used for model building using Python by ML, hereafter defined as the 'trainval' dataset and data used for evaluating the prediction model of final MHD, hereafter defined as the 'external test' dataset. The trainval dataset and the external test dataset were achieved by random splitting, to a ratio of 80:20 (27). A ratio of 80:20 is generally recommended in machine learning. In the pre-study, when the trainval dataset was reduced to other ratios such as 70:30 and the external test dataset was

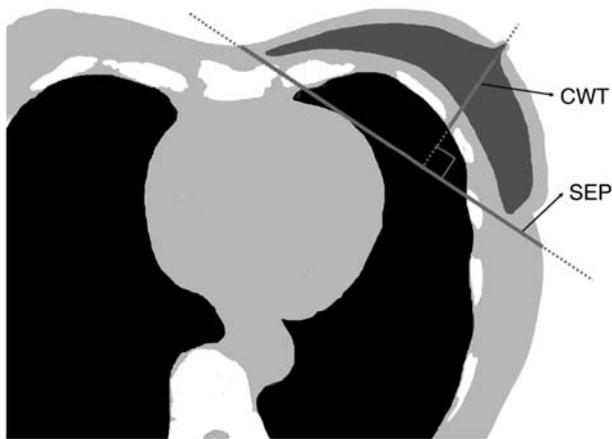


Figure 1. SEP and CWT in the CT image at the nipple level. SEP was defined as the distance along the posterior edge of the tangent fields at the nipple level. CWT was defined as the distance from the nipple surface to the lung, on a perpendicular line of breast separation. SEP, breast separation; CWT, chest wall thickness.

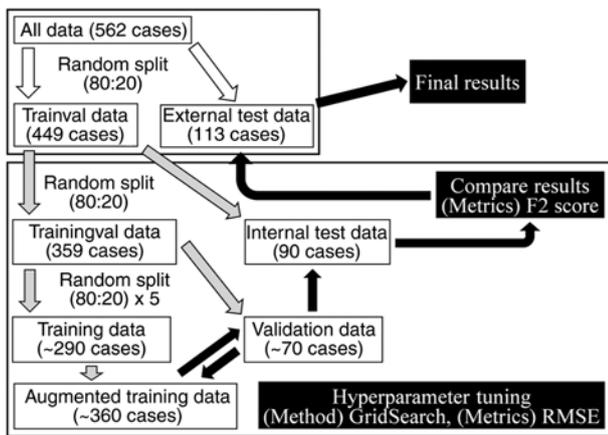


Figure 2. Process of model building. White arrows, data preparation; gray arrows, machine learning data preparation; black arrows, machine learning process. Hyperparameters were adjusted using GridSearch for decision tree, random forest and extreme gradient boosting, whereas those for deep neural network were adjusted manually. RMSE, root mean squared error.

increased, the prediction performance deteriorated (data not shown) due to the occurrence of learning loss, so a ratio of 80:20 was selected.

Correlation analysis of the explanatory and objective variables. For the analysis of correlation between the explanatory and objective variables, the statistical software SPSS (v27.0, IBM Corp.) was used to calculate the Spearman's correlation coefficient (r_s) for interval scaled variables and the Eta analysis correlation coefficient (η) for nominal scaled variables. Correlation coefficients of 1.0-0.8, 0.8-0.6, 0.6-0.4, 0.4-0.2 and 0.2-0.0 would be adjudged to be 'very strong', 'strong', 'normal', 'weak' and 'no', respectively, for the correlation strength. Since right-left, tumor site and irradiation method are categorical variables, they were converted to 0 or 1 and used as dummy variables. The six explanatory variables used for ML were right-left, tumor site, CWT, irradiation method, BMI and SEP.

Process of building models in ML.

ML algorithms. In total, four supervised ML algorithms were used: i) decision tree; ii) random forest (RF); iii) XGB; and iv) deep neural network (DNN).

Dealing with unbalanced data sets. Since the MHD data, the collected objective variable, are unbalanced data, the Python library synthetic minority oversampling with Gaussian noise (SMOBN) (28) was used to augment the number of patients of high MHD in the training dataset (Fig. 2). SMOBN is a common machine learning method for increasing the number of small number of high MHD cases (28).

Hyper-parameter tuning. The trainval dataset was randomly divided into the 'trainingval' dataset and 'internal test' dataset at a ratio of 80:20. The training dataset was randomly divided into the training dataset and validation dataset using 5-fold cross validation (5-fold CV) at a ratio of 80:20 to avoid overfitting. The training dataset was used to increase the number of patients of high MHD using SMOBN, hereafter defined as the 'augmented' training dataset. Using the augmented training dataset and validation dataset, hyperparameter tuning, which is a process of selecting the optimal parameters for each algorithm, was performed. GridSearchCV in scikit-learn (version 0.24.1, <https://scikit-learn.org/stable/index.html>) was used for all algorithms except for DNN. In DNN, hyperparameter tuning was performed manually. The root mean squared error (RMSE) was used as the evaluation metric for prediction. Hyperparameters with the best RMSE were determined for each algorithm.

Creating a model using the F2 score as the evaluation index.

In the present study, a predictive model and RMSE as the evaluation metric were used. Since the aim of the present study was to select patients with low MHD to whom DIBH are not applicable, emphasis was placed on learning to minimize false negatives (FN), preventing the false reporting of high MHD as low MHD. Therefore, the F2 score obtained using the confusion matrix was used in conjunction with RMSE as the evaluation metric. The model was created to have the best F2 score using the optimal value of hyperparameter tuning with RMSE and the internal test data.

Final model validation using the external test data. The final model was validated using external test data from 113 patients who were not used to train and build the model. In the final model validation, RMSE, MSE, MAE, r_s , accuracy, precision, recall, specificity, AUC-ROC, F1 score, and F2 score were evaluated for each algorithm using the confusion matrix.

Results

Correlations between explanatory and objective variables.

Table I shows the r_s - and η -values between the explanatory variables and MHD as the objective variable. Among the explanatory variables, a strong correlation was found between right-left and MHD ($P < 0.001$). The correlation coefficients for the other variables were low, but a significant correlation was found between CWT and MHD ($P = 0.005$).

Table I. Spearman's correlation coefficient (r_s) and Eta correlation ratio (η) between explanatory variables and MHD.

Variables	r_s or η	P-value
Right and left	0.780 ^a	<0.001
Tumor site	0.126 ^a	0.063
Chest wall thickness	0.118 ^b	0.005
Irradiation method	0.103 ^a	0.051
Age	-0.039 ^b	0.357
Body mass index	0.038 ^b	0.370
Separation	-0.027 ^b	0.522
Weight	0.025 ^b	0.560
Height	-0.024 ^b	0.572

Tumor site was defined according to the International Classification of Diseases for Oncology (third edition). Chest wall thickness (cm) was defined as the distance from the nipple surface to the lung, on a perpendicular line of breast separation. Irradiation methods are wedge or one-reference point field-in-field, field-in-field with 2 reference points. Body mass index is calculated as weight (kg)/height (m)². Separation (cm) was defined as the distance along the posterior edge of the tangent fields at the nipple level. ^aEta correlation ratio (η); ^bSpearman's correlation coefficient (r_s). MHD, mean heart dose (cGy).

Parameter optimization through hyperparameter tuning. Table II shows the optimal values of the hyperparameters for each algorithm. Each model with these optimal hyperparameters was then evaluated with an internal test dataset.

Evaluation of models in each algorithm using internal test data. Table III summarizes the optimal results out of 5-fold CV for each algorithm using the internal test dataset, with F2 score as the evaluation metric. In addition to the F2 score, the results of other evaluation metrics, such as RMSE, were also indicated. The results of RMSE were the lowest for XGB at 67.4, followed by DNN at 69.5 and RF at 81.2. For the results of F2 scoring, DNN was the highest at 0.64, followed by RF at 0.60 and decision tree at 0.48.

Final evaluation of the model of each algorithm using external test data. Table IV shows the results of the final evaluation of the model for each algorithm using the external test dataset. For RMSE, DNN had the lowest score of 77.5, followed by XGB with 85.6. For the F2 score, DNN had the highest score of 0.80, whilst RF had 0.64.

Fig. 3A shows the correlation between the true and predicted MHDs in the external test dataset using the DNN, where a strong correlation was observed with a r_s of 0.77. Fig. 3B shows the correlation between the true and predicted MHDs for the FIF-2RP patients among the external test data using DNN, where a potent correlation was observed, with a r_s of 0.83.

Fig. 4A shows the confusion matrix for the predicted MHD in the external test dataset using DNN, with 16 true positives and only 2 FN in the DNN, with a F2 score of 0.80. Fig. 4B shows the confusion matrix for predicted MHD in the

FIF-2RP patients of the external test dataset using the DNN. There were 10 true positives and only one FN, with a F2 score of 0.89.

Discussion

In the present study, four different ML algorithms were used based on the factors obtained from a single CT image slice and clinical factors to create models for predicting the MHD. Specific focus was placed on low MHD, which is not an indication for DIBH. The prediction performance of each model was then evaluated and compared. Among the algorithms tested, DNN was found to show the highest performance, with an F2 score of 0.80 and an area under the curve-receiver operating characteristic score of 0.88. The present study revealed that DNN is the optimal model for predicting MHD to select patients who are less likely to require DIBH. Previously, FIF-2RP was reported as a novel method of FIF that can significantly reduce the incidence of adverse skin events whilst slightly reducing MHD, compared with conventional FIF-1RP (24). In DNN, which was the optimal predictor of MHD, the prediction accuracy for FIF-2RP was higher compared with the analysis accuracy for all irradiation methods, where DNN appeared to be useful for selecting patients for whom DIBH was not applicable even for FIF-2RP.

In the postoperative treatment of BC, RT contributes to the reduction of postoperative local recurrence and improves survival (1,2). However, RT for BC can also reduce the survival rate due to late cardiac adverse events in some patients with left BC (3). DIBH, which reduces MHD, is becoming used more frequently in clinical practice for reducing cardiac adverse events in left BC treatment (4-6). DIBH involves asking the patient to hold their breath with deep inspiration during irradiation, causing the lung to expand with air and to enter between the heart and the chest wall (7). This dislodges the heart from the irradiation field, with the resultant reduction of heart dose (4-6). However, DIBH imposes several burdens on both patients and the RT staff, such as the additional breath-held CT imaging with deep inspiration and irradiation with respiratory synchronization, complex RT planning, extension treatment time and increasing costs (8). Furthermore, in Asian women, for FIF, the MHD of left BC is 257±90 cGy for FIF-1RP and 248±76 cGy for FIF-2RP (24), such that only ~14% of the patients have high MHD requiring DIBH (24). Therefore, a simple and accurate MHD prediction method prior to RT planning is needed for selecting patients for DIBH in Asian women. The present study revealed the highly effective utility of DNN among the ML models tested.

In recent years, AI use is becoming increasingly common in radiological practice (17,18), including AI-assisted imaging, RT planning and contouring, radiation exposure reduction and quality assurance. A number of studies have attempted the prediction of MHD in RT for BC (21-23). Koide *et al* (21) used a convolutional neural network to predict the difference between MHD with and without DIBH and MHD without DIBH, using preoperative frontal and lateral chest radiographs of 103 patients with BC. The advantage of using chest radiographs is that they are simpler compared with CT. However,

Table II. Hyperparameters-tuning results.

Machine learning algorithm	Hyperparameter name	Best value
Decision tree	max_depth	15
	min_samples_leaf	3
	max_leaf_nodes	15
Random forest	max_depth	5
	max_features	auto
	min_samples_split	2
	n_estimators	100
Extreme gradient boosting	colsample_bytree	0.700
	eta	0.05
	eval_metric	RMSE
	max_depth	11
	min_child_weight	6
	subsample	0.500
Deep neural network	Number of hidden layers	4
	Number of neurons in each hidden layer	256
	Activation function in the hidden layers	ReLU
	Activation function in the output layer	linear
	Loss function	MSE
	Optimizer	Adam

RMSE, root mean squared error; ReLU, rectified linear unit; MSE, mean squared error.

Table III. Best evaluation results of each algorithm in 5 fold-cross validation using internal test data.

Metrics	Decision tree	Random forest	Extreme gradient boosting	Deep neural network
RMSE	85.5	81.2	67.4	69.5
MSE	7311	6600	4545	4834
MAE	56.2	57.1	47.2	49.9
r_s	0.722	0.770	0.749	0.804
Accuracy	0.75	0.72	0.79	0.76
Precision	0.35	0.34	0.30	0.40
Recall	0.53	0.73	0.20	0.75
Specificity	0.80	0.72	0.91	0.76
AUC-ROC	0.67	0.73	0.55	0.75
F1 score	0.42	0.47	0.24	0.52
F2 score	0.48	0.60	0.21	0.64

RMSE, root mean squared error; MSE, mean squared error; MAE, mean absolute error; r_s , Spearman's correlation coefficient; AUC-ROC, area under receiver operating characteristic curve.

in this report, the correlation coefficient between true and predicted MHD without DIBH was 0.46 and the specificity was 0.77 (21), suggesting that results from the present study using the DNN-based method were superior. Another report (22) of the use of ML for the RT of BC showed the dose distribution of volumetric modulated arc therapy was well predicted by deep learning, with resultant improvement of the radiation treatment process by reducing the time required for planning, while maintaining plan quality. For using ML, CT

at DIBH was synthesized without imaging, where the effect of MHD reduction by DIBH was examined using MHD at DIBH calculated based on the synthesized CT (23).

A unique feature of the present study was the use of different ML models for prediction, which was able to predict the absolute value of MHD for each individual patient. In addition, the evaluation of the confusion matrix used in the classification model was incorporated into the model creation process during learning. To reduce the number of FNs that

Table IV. The final evaluation results of each algorithm using external test data.

Metrics	Decision tree	Random forest	Extreme gradient boosting	Deep neural network
RMSE	97.6	88.8	85.6	77.5
MSE	9527	7878	7325	6000
MAE	57.9	58.6	50.8	48.7
r_s	0.645	0.724	0.701	0.770
Accuracy	0.77	0.75	0.81	0.88
Precision	0.38	0.37	0.36	0.57
Recall	0.67	0.78	0.22	0.89
Specificity	0.79	0.75	0.93	0.87
AUC-ROC	0.73	0.76	0.57	0.88
F1 score	0.48	0.50	0.28	0.70
F2 score	0.58	0.64	0.24	0.80

RMSE, root mean squared error; MSE, mean squared error; MAE, mean absolute error; r_s , Spearman's correlation coefficient; AUC-ROC, area under receiver operating characteristic curve.

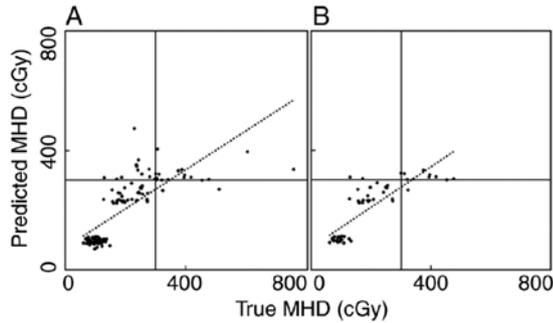


Figure 3. Spearman's correlation analysis between true and predicted MHDs in the external test data. (A) DNN (all external test data). (B) DNN (external test data of field-in-field-2 reference point). The dotted line indicates the regression line between true and predicted MHDs. MHD, mean heart dose; DNN, deep neural network.

Predicted MHD	True MHD		True MHD	
	Low	High	Low	High
High	FP 12	TP 16	FP 2	TP 10
Low	TN 83	FN 2	TN 44	FN 1

Figure 4. Confusion matrix between true and predicted MHDs in the external test data. (A) DNN (all external test data). (B) DNN (external test data of field-in-field-2 reference point). MHD ≥ 300 cGy was defined as high MHD, whereas MHD < 300 cGy would be defined as low MHD. MHD, mean heart dose; DNN, deep neural network; FP, false positive; TP, true positive; TN, true negative; FN, false negative.

incorrectly predicted patients with high MHD to be low MHD and to minimize the number of missed patients with high MHD for the selection of patients with low MHD who are not

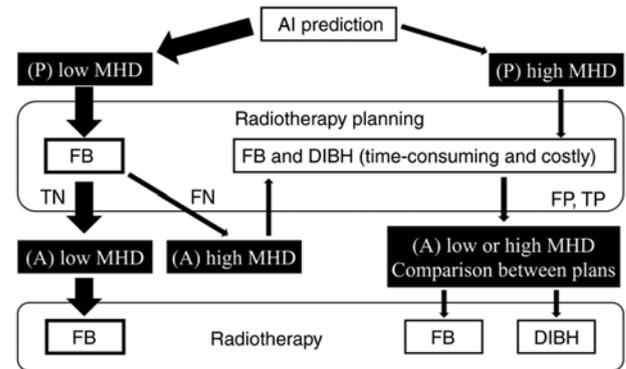


Figure 5. Radiotherapy flow using AI models in clinical practice. MHD ≥ 300 cGy was defined as high MHD, whereas MHD < 300 cGy would be defined as low MHD. AI, artificial intelligence; P, predictive; MHD, mean heart dose; FB, free-breathing; DIBH, deep inspiration breath-hold; TN, true negative; FN, false negative; FP, false positive; TP, true positive; A, actual.

candidates for DIBH, the F2 score was used as the metric of the confusion matrix in the learning process.

Using the model created in the present study, the proposed RT flow for Asian women with left BC is shown in Fig. 5, where only a portion of patients have high MHD requiring DIBH. First, a simulated CT is taken during free breathing, before the MHD is predicted using the explanatory variables from a nipple-level, one-slice simulated CT image and the present model. If the patient is predicted to have low MHD, which is expected to be the case in $>50\%$ of all patients, then treatment planning should be done by free-breathing CT. If the treatment planning result is low MHD, RT without DIBH would be administered according to the plan. If the treatment plan results in a rare FN with high MHD, then a simulated CT imaging for DIBH would be added and DIBH would be performed according to the DIBH treatment planning. For patients with high MHD, who are predicted to represent $<50\%$ of all patients, a simulated CT for DIBH would be taken and DIBH would be performed according to the DIBH treatment

planning. This RT flow should assist in reducing the number of patients who will undergo both DIBH and free-breathing treatment planning, increasing the cost and time effectiveness.

The first limitation of the present study is the relatively small number of patients (562 patients). In existing reports on MHD prediction, even fewer patients were included compared with the present study, such as 103 (21) and 94 (23). In addition, the data of the present study were unbalanced, with a high number of patients with low MHD. In response to this, data augmentation using SMOGN was performed, but it may be necessary to study with additional patients with high MHD to restore the balance.

To conclude, the present study enables the accurate prediction of MHD prior to RT planning by DNN using factors obtained from a single CT image slice and factors based on patient information. The present method is expected to be beneficial for selecting Asian patients with low MHD who do not require DIBH.

Acknowledgements

The authors would like to thank Professor Ken'ichi Morooka (Faculty of Environmental, Life, Natural Science and Technology, Okayama University, Okayama, Japan) and assistant Professor Ryohei Fukui (Faculty of Health Sciences, Okayama University, Okayama, Japan) for their useful advice.

Funding

The present study was partially supported by the Grant-in-Aid for Scientific Research (grant no. 23K07063) from the Ministry of Health, Labour and Welfare of Japan.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

RK, MK and WEHA participated in research design. RK, MK, WEA, NT, HinI, KK, KS, MO, YT, YN, MH, YM, HirI and SS performed the experiments and collected data. RK, MK and WEA analyzed the data and were major contributors in writing the manuscript. RK, MK, WEA, MB and IS analyzed data and confirm the authenticity of all the raw data. All the authors read and approved the final version of the manuscript.

Ethics approval and consent to participate

The present study was approved by the Ethics Committee of Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, and Okayama University Hospital (approval no. 2103-024). Patients provided written informed consent for undergoing RT and for the anonymous use of their data for scientific studies. The institutional informed consent forms for treatment included consent for the use of patient data and materials for research purposes. The present study was conducted in accordance with The Declaration of Helsinki, as revised in 2013.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

- Clarke M, Collins R, Darby S, Davies C, Elphinstone P, Evans V, Godwin J, Gray R, Hicks C, James S, *et al*: Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: An overview of the randomised trials. *Lancet* 366: 2087-2106, 2005.
- Darby S, McGale P, Correa C, Taylor C, Arriagada R, Clarke M, Cutter D, Davies C, Ewertz M, Godwin J, *et al*: Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: Meta-analysis of individual patient data for 10,801 women in 17 randomised trials. *Lancet* 378: 1707-1716, 2011.
- Beaton L, Bergman A, Nichol A, Aparicio M, Wong G, Gondara L, Speers C, Weir L, Davis M and Tyldesley S: Cardiac death after breast radiotherapy and the QUANTEC cardiac guidelines. *Clin Transl Radiat Oncol* 19: 39-45, 2019.
- Lu Y, Yang D, Zhang X, Teng Y, Yuan W, Zhang Y, He R, Tang F, Pang J, Han B, *et al*: Comparison of deep inspiration breath hold versus free breathing in radiotherapy for left sided breast cancer. *Front Oncol* 12: 845037, 2022.
- Falco M, Masojć B, Macała A, Łukowiak M, Woźniak P and Malicki J: Deep inspiration breath hold reduces the mean heart dose in left breast cancer radiotherapy. *Radiol Oncol* 55: 212-220, 2021.
- Yamauchi R, Mizuno N, Itazawa T, Saitoh H and Kawamori J: Dosimetric evaluation of deep inspiration breath hold for left-sided breast cancer: Analysis of patient-specific parameters related to heart dose reduction. *J Radiat Res* 61: 447-456, 2020.
- Stowe HB, Andruska ND, Reynoso F, Thomas M and Bergom C: Heart sparing radiotherapy techniques in breast cancer: A focus on deep inspiration breath hold. *Breast Cancer-Targets Ther* 14: 175-186, 2022.
- Darapu A, Balakrishnan R, Sebastian P, Hussain MR, Ravindran P and John S: Is the deep inspiration breath-hold technique superior to the free breathing technique in cardiac and lung sparing while treating both left-sided post-mastectomy chest wall and supraclavicular regions? *Case Rep Oncol* 10: 37-51, 2017.
- Teshima T, Owen JB, Hanks GE, Sato S, Tsunemoto H and Inoue T: A comparison of the structure of radiation oncology in the United States and Japan. *Int J Radiat Oncol Biol Phys* 34: 235-242, 1996.
- Nakamura K, Konishi K, Komatsu T, Sasaki T and Shikama N: Patterns of radiotherapy infrastructure in Japan and in other countries with well-developed radiotherapy infrastructures. *Jpn J Clin Oncol* 48: 476-479, 2018.
- Morganti AG, Cilla S, de Gaetano A, Panunzi S, Digesù C, Macchia G, Massaccesi M, Deodato F, Ferrandina G, Cellini N, *et al*: Forward planned intensity modulated radiotherapy (IMRT) for whole breast postoperative radiotherapy. Is it useful? When? *J Appl Clin Med Phys* 12: 3451, 2011.
- Ishizaka H, Kuroda M, Tekiki N, Khasawneh A, Barham M, Hamada K, Konishi K, Sugimoto K, Katsui K, Sugiyama S, *et al*: Investigation into the effect of breast volume on irradiation dose distribution in Asian women with breast cancer. *Acta Med Okayama* 75: 307-314, 2021.
- Kim S, Kim M and Kim M: The affecting factors of breast anthropometry in Korean women. *Breastfeed Med* 9: 73-78, 2014.
- Li X, Zhou C, Wu Y and Chen X: Relationship between formulaic breast volume and risk of breast cancer based on linear measurements. *BMC Cancer* 20: 989, 2020.
- Tortorelli G, Di Murro L, Barbarino R, Cicchetti S, di Cristino D, Falco MD, Fedele D, Ingrosso G, Janniello D, Morelli P, *et al*: Standard or hypofractionated radiotherapy in the postoperative treatment of breast cancer: A retrospective analysis of acute skin toxicity and dose inhomogeneities. *BMC Cancer* 13: 230, 2013.

16. Kim T, Reardon K, Trifiletti DM, Geesey C, Sukovich K, Crandley E, Read PW and Wijesooriya K: How dose sparing of cardiac structures correlates with in-field heart volume and sternal displacement. *J Appl Clin Med Phys* 17: 60-68, 2016.
17. Siddique S and Chow JCL: Artificial intelligence in radiotherapy. *Rep Pract Oncol Radiother* 25: 656-666, 2020.
18. Kang J, Schwartz R, Flickinger J and Beriwal S: Machine learning approaches for predicting radiotherapy outcomes: A clinician's perspective. *Int J Radiat Oncol Biol Phys* 93: 1127-1135, 2015.
19. Luo Y, Chen S and Valdes G. Machine learning for radiation outcome modeling and prediction. *Med Phys* 47: e178-e184, 2020.
20. Brodin NP, Schulte L, Velten C, Martin W, Shen S, Shen J, Basavatia A, Ohri N, Garg MK, Carpenter C and Tomé WA: Organ-at-risk dose prediction using a machine learning algorithm: Clinical validation and treatment planning benefit for lung SBRT. *J Appl Clin Med Phys* 23: e13609, 2022.
21. Koide Y, Aoyama T, Shimizu H, Kitagawa T, Miyauchi R, Tachibana H and Kodaira T. Development of deep learning chest X-ray model for cardiac dose prediction in left-sided breast cancer radiotherapy. *Sci Rep* 12: 13706, 2022.
22. Ahn SH, Kim E, Kim C, Cheon W, Kim M, Lee SB, Lim YK, Kim H, Shin D, Kim DY, *et al*: Deep learning method for prediction of patient-specific dose distribution in breast cancer. *Radiat Oncol* 16: 154, 2021.
23. Koide Y, Shimizu H, Wakabayashi K, Kitagawa T, Aoyama T, Miyauchi R, Tachibana H and Kodaira T: Synthetic breath-hold CT generation from free-breathing CT: A novel deep learning approach to predict cardiac dose reduction in deep-inspiration breath-hold radiotherapy. *J Radiat Res* 62: 1065-1075, 2021.
24. Tekiki N, Kuroda M, Ishizaka H, Khasawneh A, Barham M, Hamada K, Konishi K, Sugimoto K, Katsui K, Sugiyama S, *et al*: New field-in-field with two reference points method for whole breast radiotherapy: Dosimetric analysis and radiation-induced skin toxicities assessment. *Mol Clin Oncol* 15: 193, 2021.
25. International Classification of Diseases for Oncology. Available from: https://apps.who.int/iris/bitstream/handle/10665/96612/9789241548496_eng.pdf.
26. Chen T and Guestrin C: XGBoost: A Scalable Tree Boosting System. In proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. pp785-794, 2016.
27. Joseph, VR: Optimal ratio for data splitting. *Stat Anal Data Min* 15: 531-538, 2022.
28. Branco P, Torgo L and Ribeiro RP: SMOGN: A pre-processing approach for imbalanced regression. *Proc Mach Learn Res* 74: 36-50, 2017.