Applying Artificial Neural Networks to Develop a Decision Support Tool for Tis–4N0M0 Non–Small-Cell Lung Cancer Treated With Stereotactic Body Radiotherapy

Takafumi Nemoto, MD[1,2]; Atsuya Takeda, MD, PhD[1]; Yukinori Matsuo, MD, PhD[3]; Noriko Kishi, MD[3]; Takahisa Eriguchi, MD, PhD[1]; Etsuo Kunieda, MD, PhD[4]; Ryusei Kimura, MS[5]; Naoko Sanuki, MD, PhD[1]; Yuichiro Tsurugai, MD, PhD[1]; Masamichi Yagi, MS[6]; Yousuke Aoki, BS[1]; Yohei Oku, PhD[1]; Yuto Kimura, MS[1]; Changhee Han, PhD[7]; and Naoyuki Shigematsu, MD, PhD[2]

PURPOSE Clear evidence indicating whether surgery or stereotactic body radiation therapy (SBRT) is best for non–small-cell lung cancer (NSCLC) is lacking. SBRT has many advantages. We used artificial neural networks (NNs) to predict treatment outcomes for patients with NSCLC receiving SBRT, aiming to aid in decision making.

PATIENTS AND METHODS Among consecutive patients receiving SBRT between 2005 and 2019 in our institution, we retrospectively identified those with Tis–T4N0M0 NSCLC. We constructed two NNs for prediction of overall survival (OS) and cancer progression in the first 5 years after SBRT, which were tested using an internal and an external test data set. We performed risk group stratification, wherein 5-year OS and cancer progression were stratified into three groups.

RESULTS In total, 692 patients in our institution and 100 patients randomly chosen in the external institution were enrolled. The NNs resulted in concordance indexes for OS of 0.76 (95% CI, 0.73 to 0.79), 0.68 (95% CI, 0.60 to 0.75), and 0.69 (95% CI, 0.61 to 0.76) and area under the curve for cancer progression of 0.80 (95% CI, 0.75 to 0.84), 0.72 (95% CI, 0.60 to 0.83), and 0.70 (95% CI, 0.57 to 0.81) in the training, internal test, and external test data sets, respectively. The survival and cumulative incidence curves were significantly stratified. NNs selected low-risk cancer progression groups of 5.6%, 6.9%, and 7.0% in the training, internal test, and external test data sets, respectively, suggesting that 48% of patients with peripheral Tis–4N0M0 NSCLC can be at low-risk for cancer progression.

CONCLUSION Predictions of SBRT outcomes using NNs were useful for Tis–4N0M0 NSCLC. Our results are anticipated to open new avenues for NN predictions and provide decision-making guidance for patients and physicians.

INTRODUCTION

Lung cancer is the most common malignancy and the most frequent cause of cancer-related deaths worldwide.[1] Non–small-cell lung cancer (NSCLC) represents 85% of all lung cancers, out of which 25%-28% are diagnosed as stage I.[2,3] Recently, numbers of patients with early-stage NSCLC have been rising.[4] For early-stage NSCLC, surgical resection is the standard treatment, whereas stereotactic body radiotherapy (SBRT) is a standard treatment for medically inoperable patients.[5] However, clear evidence indicating whether surgery or SBRT is best for NSCLC is lacking.[6] SBRT enables a high percentage of patients with early-stage NSCLC to achieve a complete cure.[7] Furthermore, SBRT has many advantages; it is less painful, has few side effects, and preserves a high quality of life after treatment, as well as being cost-effective and that it can be performed on an outpatient basis during a short treatment period, features favored by governments and insurance companies.[8-10]

For these reasons, patients with NSCLC have increasingly been treated with SBRT in the United States.[11] In the Netherlands, the frequency of SBRT is very similar to that of surgery (41% v 47%).[12] This suggests that there may be an increase in the number of operable patients who choose SBRT, along with rising use in the elderly and those who are ineligible for surgery. In clinical practice, we often make decisions that are not on the basis of solid evidence, instead relying on experience. However, outcomes following SBRT are hard to predict in all patients and there are limits in making decisions on the basis of experience. Were it possible to inform patients and physicians of recurrence risks on SBRT on the basis of advanced tools specific to a patient's own data, such information would be helpful to both of them while making the decisions related to undergoing treatment or not, facilitating a shared decision-making process.

Risk factors for poor overall survival (OS) reportedly include high T-stage, high comorbidity score, squamous

CONTEXT

Key Objective

To develop and validate artificial neural network (NN) models that predict systemic treatment outcomes for patients with Tis–4N0M0 non–small-cell lung cancer receiving stereotactic body radiotherapy and stratify the patients into three risk groups, aiming to aid in decision making.

Knowledge Generated

The devised NNs predicted outcomes more accurately than other statistical and machine learning models and significantly stratified the survival and cumulative incidence curves into high-, intermediate-, and low-risk cohorts for overall survival and cancer progression.

Relevance

The NNs suggested that approximately half of the patients with peripheral Tis–4N0M0 non–small-cell lung cancer expected to receive our current stereotactic body radiotherapy delivery with a high maximum dose and steep gradient should be informed beforehand of the likelihood of belonging to the low-risk cancer progression group. Our results are anticipated to open new avenues for NN predictions and provide decision-making guidance for patients and physicians.

cell histology, and low biologically effective dose (BED).[13-15] We previously analyzed prognostic factors in patients with NSCLC treated with SBRT, and demonstrated pleural contact, as well as long solid component diameter, to be associated with worse cancer-specific mortality and OS.[16] In addition, artificial intelligence (AI) such as artificial neural networks (ANNs, often simply referred to as NNs) and support vector machine (SVM) approaches, which include information obtained from computed tomography (CT)-based radiomics, providing prognostic information for patients with NSCLC.[17-19] These findings motivated us to investigate whether combining multiple clinical and treatment factors can predict patients' outcomes more precisely, and whether we can identify low-risk patients among those with NSCLC undergoing SBRT. Furthermore, we have experience and reported about SBRT for clinical T3-4N0M0 NSCLC at our institution.[20] In light of the above considerations, we reanalyzed prognostic factors in our data set that comprised patients with Tis–4N0M0 NSCLC treated with SBRT, developed a tool that estimates treatment outcomes from clinical information using NNs for not only physicians but also patients, and tested this tool using both an internal test and an external test data set, endeavoring to verify whether NNs were superior to conventional methods.

PATIENTS AND METHODS

Patients

This was a retrospective multicohort study of consecutive patients with clinical Tis–4N0M0 NSCLC on the basis of the eighth TNM staging system, treated with SBRT in Ofuna Chuo Hospital (OCH) or Kyoto University Hospital (KUH) between January 2005 and August 2019. The inclusion criteria were as follows: clinically staged as Tis–4N0M0; total peripheral dose or minimum dose that covered 95% of the planning target volume of 40-60 Gy; 4-10 fractions; and no history of SBRT (if SBRT had been performed twice or more for metachronous lung cancers, the first SBRT was described). Patients lacking follow-up were excluded. Although principally biopsy was proposed to patients, some refused, others could not undergo a biopsy because of technical or clinical difficulties, and yet others were examined but without pathologic confirmation. For patients lacking pathologic confirmation, clinical diagnoses of NSCLC were made by the lung cancer board on the basis of clinical information. This detail and dose prescriptions are described in the Data Supplement. All patients provided written informed consent for their treatments and the retrospective use of their data for future investigations. The review boards of both institutions approved this study (OCH, No. 2019-013; KUH, No. R2292).

Model Development and Evaluation

We constructed two NN types (Fig 1 and Data Supplement). First, we implemented DeepSurv[21] for OS, designated NN1. DeepSurv is a deep feed-forward NN that follows the Cox proportional hazards (CPH) model and predicts the hazard function. Next, we constructed another NN for a binary classification of cancer progression in the first 5 years after SBRT, depicted as NN2. The cancer progression category included local recurrence, regional lymph node recurrence, pleural cavity recurrence, and distant metastasis. Five sixths of the eligible patients in OCH served as a training data set and one sixth as an internal test data set. The eligible patients in KUH served as an external test data set.

All of the factors shown in Table 1 except the maximum standardized uptake value (SUVmax) were input into the NNs. In histology, nonspecific NSCLC represented

FIG 1. Flowchart of patient recruitment and analysis. We created NN1 and CPH models for OS, and NN2 and LR models for cancer progression using OCH training data set. The models were tested using the internal (OCH) and external (KUH) data sets. CPH, Cox proportional hazards; KUH, Kyoto University Hospital; LR, logistic regression; NN, neural network; NSCLC, non–small-cell lung cancer; OCH, Ofuna Chuo Hospital; OS, overall survival; SBRT, stereotactic body radiotherapy.
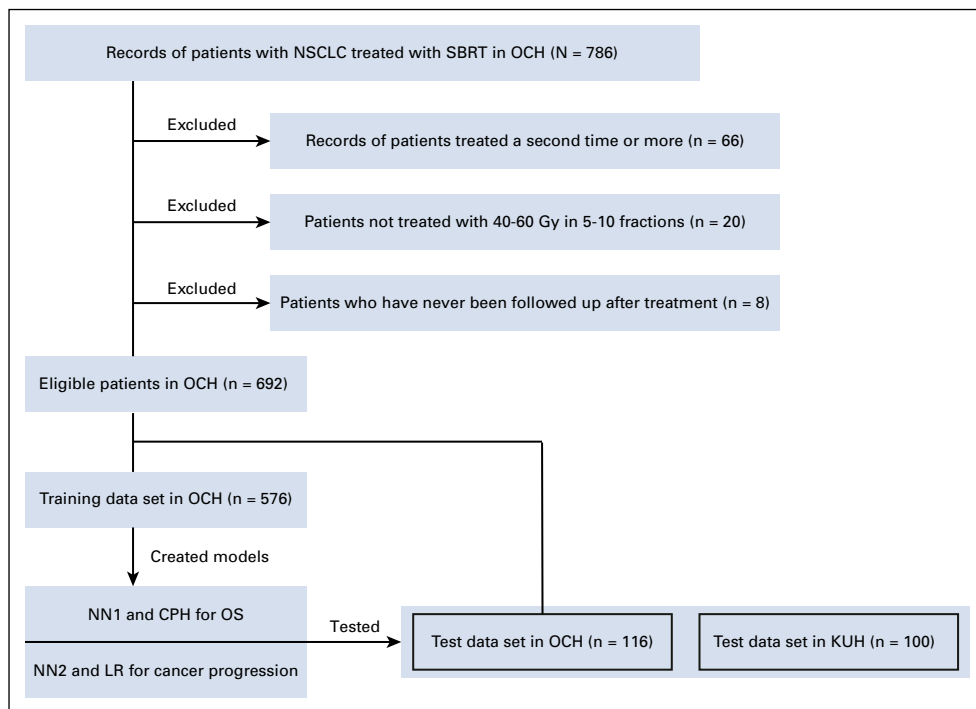
NSCLC with no specific diagnosis such as adenocarcinoma and squamous cell carcinoma. We excluded SUVmax because there are concerns about the consistency of values obtained using different [^{18}F] fluorodeoxyglucose positron emission tomography-CT machines. Tumor characteristics[22] were classified as pure ground-glass opacity (GGO), part-solid (GGO + solid), and pure-solid. Radiation oncologists (T.E. and A.T. in OCH; N.K. and Y.M. in KUH) retrospectively reviewed pretreatment CT images.

For comparison, we constructed CPH, elastic net penalized CPH,[23] random survival forest (RSF),[24] logistic regression (LR), SVM, random forest (RF), and extreme gradient boosting (XGBoost).[25] Elastic net is a mixture of lasso and ridge regression penalties. RSF is a survival model that applies RF. XGBoost is an algorithm of gradient boosted decision trees designed for speed and performance. Regarding CPH, LR, and elastic net, the variables were selected using a stepwise method on the basis of the Akaike information criterion. Since our goal was to create a predictive model, we strictly adopted the output factors of the stepwise selection without making any adjustments on the basis of clinical judgment. Concordance index (C-index) was evaluated for OS and the area under the receiver operating characteristic curve (AUC) for cancer progression. Finally, risk group stratification and a feature analysis were performed. As subgroup analyses, groups of Tis–2N0M0 NSCLC and histologically proven NSCLC were evaluated. Further details are given in the Data Supplement.

Statistical Analysis

For comparison of baseline characteristics between the data sets, the chi-squared test was used for categorical variables, and the Mann-Whitney U test for continuous variables. Median follow-up was calculated using the reverse Kaplan-Meier method. The log-rank test was used for OS. Gray's test was used for cancer progression, with nonspecific death being regarded as a competing risk. For all tests, a two-tailed P-value of < .05 was considered to indicate a statistically significant difference. Clopper-Pearson 95% CIs were calculated for the proportions of each factor (Data Supplement). All statistical analyses were performed with Python 3.7.4, SAS 9.4, and R 3.6.3.

RESULTS

In total, 692 patients in OCH and 100 patients randomly chosen in KUH satisfying the criteria were enrolled (Fig 1). The 692 OCH patients were randomly divided into 576 as a training data set and 116 as an internal test data set. All 100 KUH patients served as an external test data set. The characteristics of the eligible patients are shown in Table 1. Median follow-up durations in OCH and KUH were 68.6 and 65.3 months, respectively. Although some characteristics differed significantly between the two institutions, for example, the OCH patients had larger tumors, more severe T-classification, and a higher BED at the planning target volume center, OS and the cumulative cancer progression incidence of the institutions did not differ significantly (Fig 2). There was

TABLE 1. Patient, Tumor, and Treatment Characteristics in the Eligible Data Sets

Characteristic	OCH (n = 692)	OCH Training Data Set (n = 576)	OCH Test Data Set (n = 116)	KUH (n = 100)	OCH v KUH P	OCH Training v OCH Test P
Follow-up, months (range)	68.6 (0.1-162.7)	68.6 (0.1-162.7)	66.8 (2.4-136.9)	65.3 (6.3-152.5)	.33	.39
Age, years (IQR)	78 (74-83)	78 (74-83)	79 (73-82)	78 (73-83)	.70	.78
Sex, female/male	221/471 (32/68)	186/390 (32/68)	35/81 (30/70)	30/70 (30/70)	.78	.74
Height, cm (IQR)	159.5 (151.9-165.6)	159.7 (151.4-166.6)	158.8 (154.2-164.5)	158.5 (151.8-164.2)	.19	.71
Weight, kg (IQR)	54.9 (47.0-62.0)	54.9 (47.0-62.0)	54.2 (46.8-60.0)	53.6 (44.8-61.5)	.29	.59
Treated tumor size, cm (IQR)	2.5 (1.9-3.3)	2.5 (1.9-3.3)	2.5 (1.8-3.2)	2.0 (1.5-2.9)	< .001	.50
T-category					< .001	.32
Tis/1mi/1a/1b/1c	24/7/35/179/145 (3/1/5/26/21)	17/5/30/154/121 (3/1/5/27/21)	7/2/5/25/24 (6/2/4/22/21)	8/2/6/39/27 (8/2/6/39/27)		
T2a/2b/3/4	167/31/77/27 (24/4/11/4)	133/24/67/25 (23/4/12/4)	34/7/10/2 (29/6/9/2)	16/2/0/0 (16/2/0/0)		
Histology					.46	.52
Ad/SCC/LCC/nonspecific NSCLC/unproven	192/104/3/41/352 (28/15/0.4/6/51)	166/83/3/33/291 (29/14/1/6/51)	26/21/0/8/61 (22/18/0/7/53)	29/18/1/2/50 (29/18/1/2/50)		
Tumor texture					.008	.06
Pure GGO/GGO + solid/pure solid	29/105/558 (4/15/81)	20/92/464 (3/16/81)	9/13/94 (8/11/81)	9/6/85 (9/6/85)		
Pleural contact, yes/no	274/418 (40/60)	234/342 (41/59)	40/76 (34/66)	34/66 (34/66)	.34	.26
PET/CT SUVmax (range)	3.0 (0-30.1)	2.9 (0-30.1)	3.1 (0-16.4)	5.1 (0.6-30.0)	.01	.68
Charlson comorbidity index					.99	.72
0/1/2/3	95/161/185/126 (14/23/27/18)	79/141/153/100 (14/25/27/17)	16/20/32/26 (14/17/28/22)	17/19/31/18 (17/19/31/18)		
4/5/6-12/unknown	56/37/31/1 (8/5/4/0.1)	45/30/27/1 (8/5/5/0.2)	11/7/4/0 (9/6/3/0)	8/4/3/0 (8/4/3/0)		
GOLD classification					< .001	.53
0/I/II/III/IV/unknown	366/53/149/102/19/3 (53/8/22/15/3/0.4)	306/41/123/85/18/3 (53/7/21/15/3/1)	60/12/26/17/1/0 (52/10/22/15/1/0)	47/16/17/6/1/13 (47/16/17/6/1/13)		
Performance status					.10	.65
0/1/2/3/4/unknown	407/193/71/18/2/1 (59/28/10/3/0.3/0.1)	339/161/57/17/2/0 (59/28/10/3/0.3/0)	68/32/14/1/1 (59/28/12/1/1)	50/42/7/1/0/0 (50/42/7/1/0/0)		
History of lung cancer	166 (24)	141 (24)	25 (22)	27 (27)	.60	.58
History of nonlung cancer	151 (22)	128 (22)	23 (20)	33 (33)	.02	.66
Operability, yes/no	151/541 (22/78)	120/456 (21/79)	31/85 (27/73)	30/70 (30/70)	.09	.20
Dose prescription at PTV periphery					NA	.77
40 Gy/5 fr	98 (14)	78 (14)	20 (17)	0 (0)		
50 Gy/5 fr	489 (71)	410 (71)	79 (68)	0 (0)		
60 Gy/5 fr	80 (12)	66 (11)	14 (12)	0 (0)		
60 Gy/10 fr	20 (3)	17 (3)	3 (3)	0 (0)		
42 Gy/4 fr	0 (0)	0 (0)	0 (0)	42 (42)		

(Continued on following page)

TABLE 1. Patient, Tumor, and Treatment Characteristics in the Eligible Data Sets (Continued)

Characteristic	OCH (n = 692)	OCH Training Data Set (n = 576)	OCH Test Data Set (n = 116)	KUH (n = 100)	OCH v KUH P	OCH Training v OCH Test P
49 Gy/4 fr	0 (0)	0 (0)	0 (0)	7 (7)		
50 Gy/4 fr	0 (0)	0 (0)	0 (0)	38 (38)		
52.5 Gy/8 fr	0 (0)	0 (0)	0 (0)	13 (13)		
Others	5 (0.7)	5 (1)	0 (0)	0 (0)		
Dose prescription at PTV center, Gy (range)	83 (50-100)	83 (50-100)	83 (50-100)	60 (48-76)	< .001	.65

NOTE. Data are presented as the median (IQR) or n (%) with the exception of follow-up, PET/CT SUVmax, and dose prescription at the PTV center, which are presented as median (range) values. All of the factors except SUVmax were input into the NNs.

Abbreviations: Ad, adenocarcinoma; fr, fractions; GGO, ground-glass opacity; GOLD, global initiative for chronic obstructive lung disease; IQR, interquartile range; KUH, Kyoto University Hospital; LCC, large cell carcinoma; NA, not applicable; NSCLC, non–small-cell lung cancer; OCH, Ofuna Chuo Hospital; PET/CT, [18F] fluorodeoxyglucose positron emission tomography–computed tomography; PTV, planning target volume; SCC, squamous cell carcinoma; SUVmax, the maximum standardized uptake value.

no significant difference between the OCH training and OCH test data sets.

The prognostic performances are shown in Table 2. The multivariate CPH regression model consisted of the following 13 factors: squamous cell carcinoma, operability, T-category, tumor texture, maximum dose, fraction number, total tumor size, solid-part tumor size, age, Charlson comorbidity index, performance status, height, and weight, and the LR model consisted of the following seven factors: large cell carcinoma, nonspecific NSCLC, operability, T-category, tumor texture, age, and history of nonlung cancer. The performance of the NNs was consistently superior to that of the traditional statistical models and the other machine learning models in both the OCH and KUH test data sets. The C-index of OS for the training data set in OCH using univariate (T-category) CPH regression, multivariate CPH regression, and NN were 0.64 (95% CI, 0.61 to 0.67), 0.71 (95% CI, 0.68 to 0.74), and 0.76 (95% CI, 0.73 to 0.79), respectively. The AUC of cancer progression for the training data set using univariate (T-category) and multivariate LR and NN were 0.65 (95% CI, 0.60 to 0.71), 0.68 (95% CI, 0.63 to 0.73), and 0.80 (95% CI, 0.75 to 0.84), respectively. Figure 3 shows stratification of the outcomes into high-risk, intermediate-risk, and low-risk groups on the basis of the tertile values of the training data set derived using the NN models. In the OCH training data set, the 5-year OS rates for the high-risk, intermediate-risk, and low-risk groups were 17.8% (95% CI, 12.2 to 24.3), 55.1% (95% CI, 46.6 to 62.7), and 77.2% (95% CI, 68.8 to 83.6), respectively. The corresponding 5-year cancer progression rates were 56.6% (95% CI, 48.1 to 63.2), 22.3% (95% CI, 16.3 to 28.9), and 5.6% (95% CI, 2.5 to 10.3). The survival and cumulative incidence curves also showed significant stratification for both of OS and cancer progression, in both the internal and external test data sets. Subgroup analyses of groups of Tis–2N0M0 NSCLC and histologically proven NSCLC revealed reasonable performance and risk stratifications of the NNs (Table 2, Data Supplement). Calibration plots of the NN models generally followed the 45-degree diagonal line, which is widely considered to be ideal (Data Supplement).

Proportions of features in the low-, intermediate-, and high-risk groups in the training data set are shown in Figure 4 and the Data Supplement. In the high-risk group, T-category, proportion of positive pleural contacts, and proportion of solid tumor textures were higher, and the BED at alpha/beta of 10 (BED_{10}) of the maximum dose was lower for both OS and cancer progression, whereas age, performance status, and Charlson comorbidity index were higher only for OS. As to cancer progression, NN2-derived low risk accounted for 48% of the group with BED_{10} of the
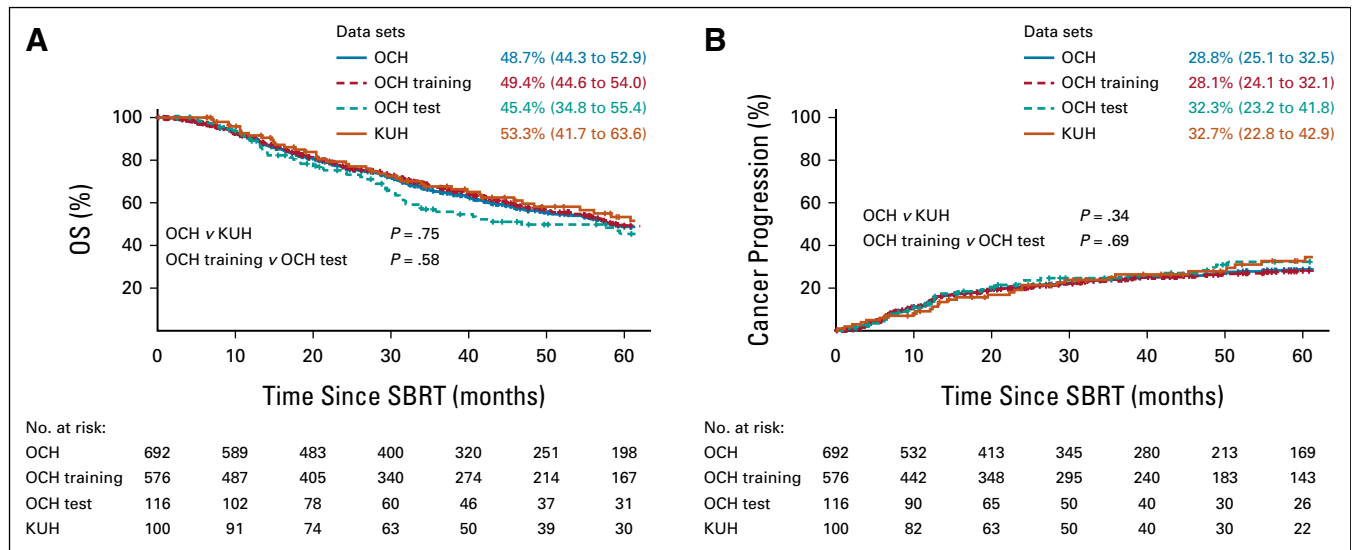
FIG 2. Survival and cumulative incidence curves. (A) OS and (B) cumulative cancer progression incidence are indicated by solid lines for OCH (sum of OCH training data set and OCH test data set) and KUH and dashed lines for OCH training data set and OCH test data set. Values at 5 years are shown with 95% CIs. KUH, Kyoto University Hospital; OCH, Ofuna Chuo Hospital; OS, overall survival; SBRT, stereotactic body radiotherapy.

maximum dose ≥ 200, which is equivalent to the dose currently prescribed for peripheral tumors in OCH (Data Supplement).

DISCUSSION

To our knowledge, this is the first study attempting to predict systemic treatment outcomes of patients with NSCLC undergoing SBRT, using NNs with input of pretreatment information. This is among the largest studies of SBRT for lung cancer, given the patient number: the internal data set was composed of 692 patients, and the independent external test data set was composed of 100 patients (Fig 1). The median follow-up duration was longer than 5 years and 73% of the patients had more than 5 years of follow-up or died before the 5th year, enhancing the reliability of the results (Table 1). In this research, we focused on patients with a wide range of staging, Tis–T4N0M0. We were interested to predict the prognosis of a wide range of patients without being limited to T-category using factors such as pleural contact and tumor size. Although SBRT for T3-4N0M0 is unusual, we have reported good results with SBRT alone for patients who were inoperable or refused surgery at our hospital.[20] The 5-year cancer progression rates in the NN2-derived low-risk group, which accounted for one third of the training data set, were 5.6%, 6.9%, and 7.0% of the OCH training, OCH internal test, and KUH external test data sets, respectively (Fig 3). Since 2011, OCH has been treating peripheral NSCLC patients with SBRT using a high dose of 50-60 Gy/5 fractions with a steep dose gradient, which is over 200 Gy of the BED_{10} of the maximum dose. Nearly half of these patients, 48%, were categorized into the NN2-derived low-risk group for cancer progression in

the training data set (Data Supplement). Therefore, the results of this study raise the possibility that, in the future, approximately half of patients with peripheral Tis–4N0M0 NSCLC recommended to undergo SBRT with an equivalent dose prescription could be informed beforehand of the prediction of a low cancer progression rate on the basis of the group into which they were categorized.

NNs are highly scalable, and DeepSurv is a NN optimizing the hazard function of CPH.[21] DeepSurv was reported to have a higher C-index for survival prediction than RSF and CPH in studies of patients with cancer.[26,27] Similar to these previous studies, we used DeepSurv to predict OS, and found the C-index of DeepSurv (NN1) to be greater than those of CPH, elastic net, and RSF in the test data sets (Table 2). TNM staging is the most widely used prognostic scale and facilitates selecting appropriate treatment on the basis of stages. However, because of its simplicity, each stage contains various cancer and patient states. Traditional statistical approaches such as CPH and LR are powerful for analyzing and explaining data, but they are not necessarily as good at prediction as NNs. It also difficult for the traditional approaches to handle a very large number of dimensions because multicollinearity needs to be avoided, and they cannot process imaging or language data unless there is feature extraction.[28] NN is a method that can overcome these shortcomings.

We further analyzed cancer progression, because progression of the malignancy was the second most serious event next to mortality for patients. For cancer progression, nonspecific death is a competing risk, and
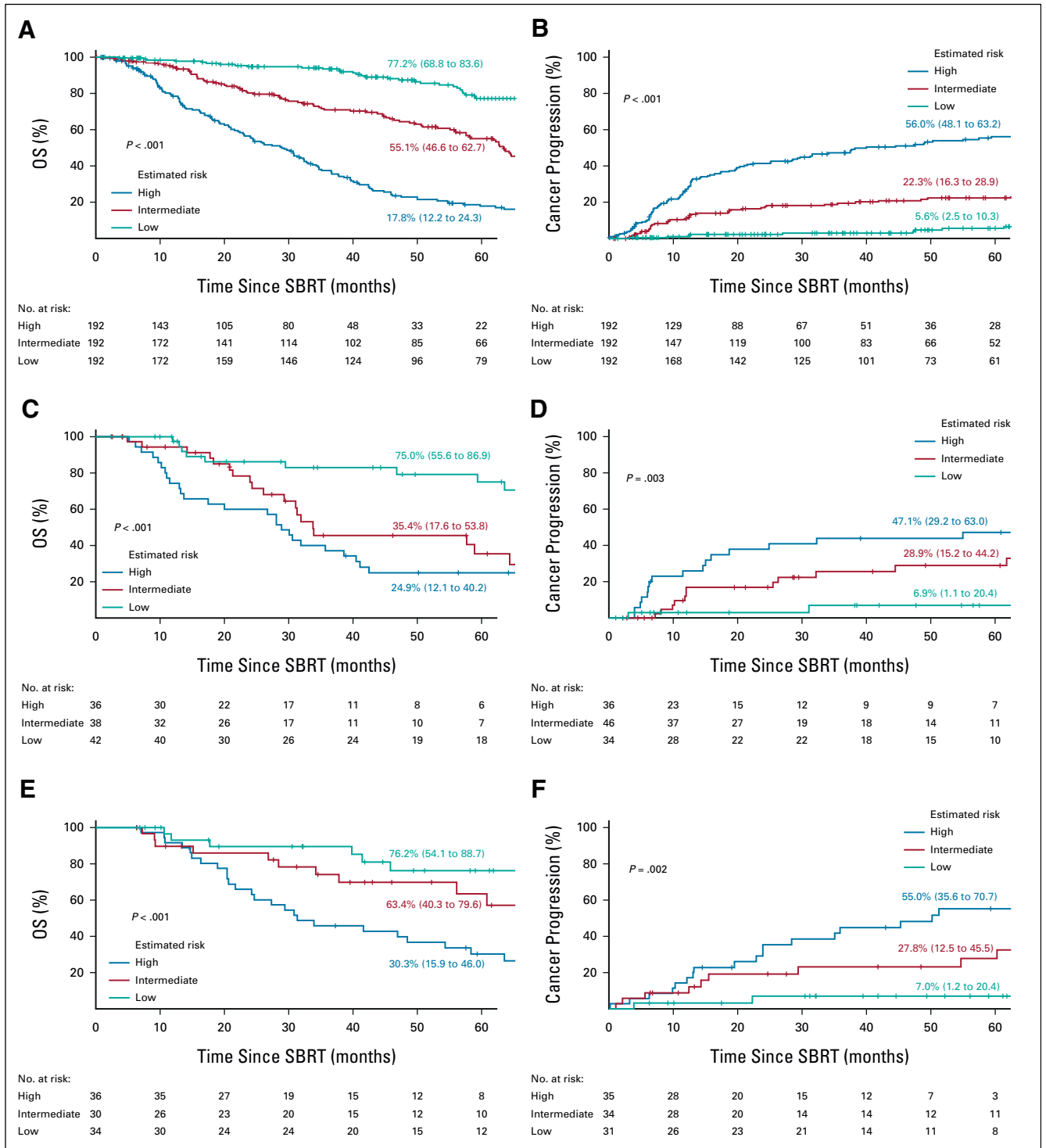
FIG 3. OS and cumulative incidence curves apportioned by the estimated risk. (A, C, E) OS and (B, D, F) cancer-progression in the first 5 years after SBRT, stratified into high-risk, intermediate-risk, and low-risk groups on the basis of the tertile values obtained in the training data set by the NN models. The top row (A, B) shows OCH training data set, the middle row (C, D) OCH test data set, and the bottom row (E, F) KUH test data set. Predicted values at 5 years are shown with 95% CIs. KUH, Kyoto University Hospital; NN, neural network; NSCLC, non–small-cell lung cancer; OCH, Ofuna Chuo Hospital; OS, overall survival; SBRT, stereotactic body radiotherapy.

must be considered for correctly predicting cancer progression because almost half of patients with NSCLC treated with SBRT die of other diseases. However, no standard NN has as yet been established representing the cumulative incidence function (CIF) considering time-to-event and competing risks. In this study, we

TABLE 2. Prognostic Performances of NNs and Classic Regression Methods

Outcome	Model	OCH Training Data Set	OCH Test Data Set	KUH Test Data Set
OS (C-index)	Univariate (T-category) CPH	0.64 (0.61-0.67)	0.62 (0.54-0.69)	0.57 (0.49-0.65)
	Multivariate CPH	0.71 (0.68-0.74)	0.64 (0.57-0.71)	0.67 (0.59-0.75)
	Elastic net	0.71 (0.68-0.74)	0.64 (0.57-0.72)	0.68 (0.60-0.76)
	RSF	0.79 (0.77-0.82)	0.66 (0.59-0.73)	0.65 (0.57-0.73)
	NN1 (Tis–2N0M0)	0.76 (0.72-0.79)	0.66 (0.58-0.73)	0.69 (0.61-0.76)
	NN1 (histologically proven NSCLC)	0.76 (0.73-0.80)	0.68 (0.56-0.79)	0.68 (0.58-0.78)
	NN1	0.76 (0.73-0.79)	0.68 (0.60-0.75)	0.69 (0.61-0.76)
Cancer progression within 5 years (AUC)	Univariate (T-category) LR	0.65 (0.60-0.71)	0.61 (0.49-0.73)	0.63 (0.50-0.74)
	Multivariate LR	0.68 (0.63-0.73)	0.64 (0.52-0.76)	0.69 (0.55-0.81)
	SVM	0.64 (0.59-0.69)	0.59 (0.49-0.70)	0.62 (0.50-0.74)
	RF	0.65 (0.60-0.70)	0.61 (0.50-0.72)	0.60 (0.48-0.72)
	XGBoost	0.87 (0.83-0.90)	0.64 (0.52-0.77)	0.62 (0.47-0.74)
	NN2 (Tis–2N0M0)	0.81 (0.75-0.85)	0.74 (0.62-0.85)	0.70 (0.57-0.81)
	NN2 (histologically proven NSCLC)	0.78 (0.71-0.84)	0.66 (0.50-0.80)	0.61 (0.42-0.78)
	NN2	0.80 (0.75-0.84)	0.72 (0.60-0.83)	0.70 (0.57-0.81)

NOTE. The performances of NN, CPH, and LR were evaluated using the C-index for OS and the AUC for cancer progression. The values presented are means and 95% CIs derived by a bootstrap approach of 2,000 iterations.

Abbreviations: AUC, area under the curve; C-index, concordance index; CPH, Cox proportional hazards; KUH, Kyoto University Hospital; LR, logistic regression; NN, neural network; NSCLC, non–small-cell lung cancer; OCH, Ofuna Chuo Hospital; OS, overall survival; RF, random forest; RSF, random survival forest; SVM, support vector machine; XGBoost, extreme gradient boosting.

constructed a binary classification NN (NN2) for cancer progression and stratified risks by the output of NN2, plotting CIF to evaluate the probability. NN2 showed better predictive performance than LR, SVM, RF, and XGBoost in the test data sets and significantly stratified risk assignment, that is, NN2 worked satisfactorily (Table 2 and Fig 3). In the near future, it is expected that more efficient NN methods dealing with CIF would be developed.

As there is a black box within the NNs, we demonstrated the compositions of each risk group stratified by NNs for each factor considered to be clinically important (Fig 4). As mentioned in the results section, there were factors that seemed to be related to both OS and the cancer progression, and factors apparently related only to OS. These results are clinically plausible. Histology correlated with both OS and cancer progression, but more strongly with cancer progression; the low-risk group for cancer progression had a higher proportion of cases with unproven histology. We can speculate as to several possible mechanisms. Approximately half of the patients included in this study had histologically unproven tumors. These were usually small and often GGO, and therefore more likely to be classified into the low-risk group for cancer progression. Moreover, biopsies are frequently skipped in patients who are elderly and/or frail, and these patients often die of other diseases.

Indeed, the possibility that the group with histologically unproven cancers included tumors that were not truly malignant cannot be ignored. However, we can reasonably speculate that inflammatory diseases including tuberculosis and benign tumors would not have been particularly high in the group with histologically unproven tumors, because the diagnosis of lung cancer was carefully made by a multidisciplinary lung cancer review board after thorough consideration of these diseases. Overall, the results produced by the NNs were within the range of what is currently understood by experienced clinicians.

Various types of data have been widely input using AI, not only to predict outcomes but also to identify links relevant to the treatment decision-making process.[19] A number of studies input CT imaging features extracted by radiomics, and recent studies have directly input images using a convolutional neural network , a novel NN technology that can automatically extract features of an image and optimize them.[19,29,30] Yet other studies have used pathologic images or genes as input.[30,31] Also, for lung cancer, there has been AI prediction research. She et al developed a DeepSurv model to predict cancer-specific survival for patients with NSCLC receiving surgery and reported that the C-index was 0.739 for the internal test and 0.742 for the external test.[17] Zhou et al[18] predicted distant metastasis for patients with stage I NSCLC receiving SBRT using a
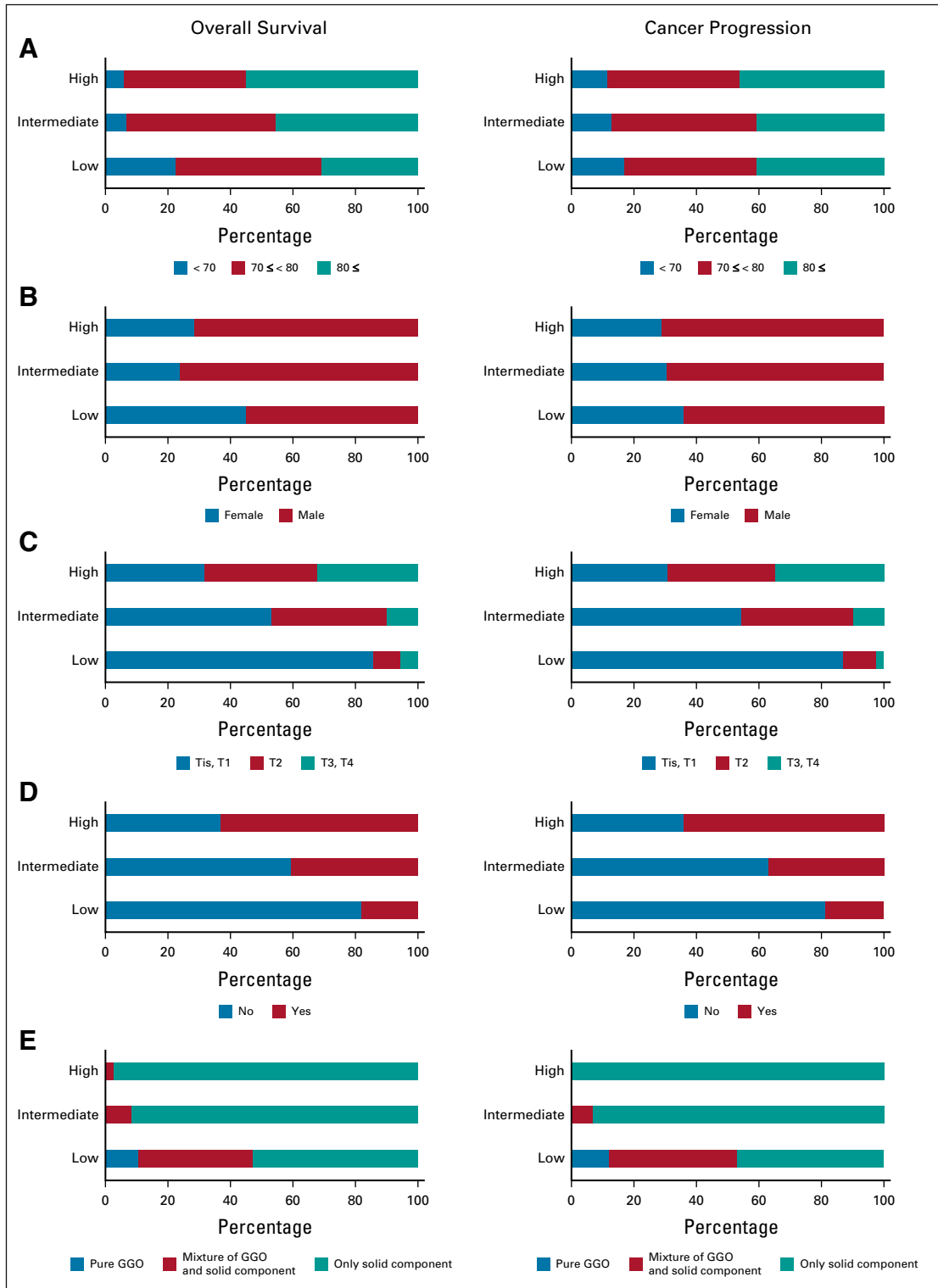
FIG 4. Feature analysis in OCH training data set. For clinically important features, the proportion of features included in each risk group was determined. The left column shows OS and the right column shows cancer progression. What attributes each risk group was composed of was shown: (A) age, (B) sex, (C) T-category, (D) pleural contact, (E) tumor texture, (F) histology, (G) PS, (H) CCI, and (I) BED_{10} of maximum dose. Ad, adenocarcinoma; BED_{10}, biological effective dose at alpha/beta of 10; CCI, Charlson comorbidity index; GGO, ground-glass opacity; LCC, large cell carcinoma; NSCLC, non–small-cell lung cancer; OCH, Ofuna Chuo Hospital; OS, overall survival; PS, performance status; SCC, squamous cell carcinoma.
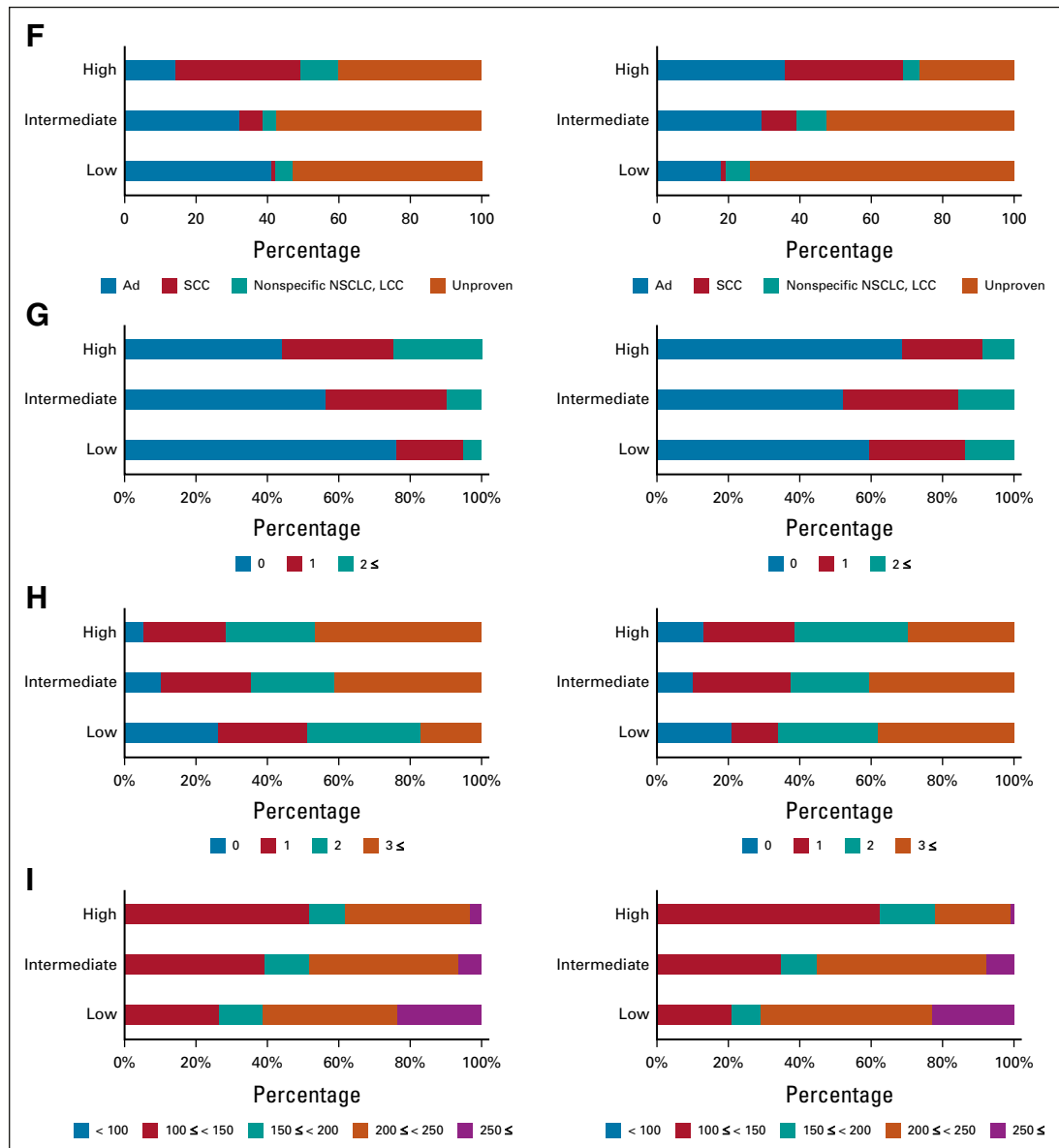
FIG 4. (Continued)

binary classification NN, SVM, and LR. The AUCs for the NN, SVM, and LR were 0.75, 0.80, and 0.73, respectively. However, the results of Zhou et al were not evaluated using an internal and/or an external test data set. To our knowledge, this is the first study to apply NNs including DeepSurv to systemic treatment outcomes after SBRT for NSCLC, in contrast to the DeepSurv model of surgical data studied by She et al. We used both internal and external tests, thereby evaluating the generalization performance of the devised models. Our results showed the accuracies for the test data set to be lower than those of the training data set; however, the difference was modest because not only cross-validation but also generalizing techniques such as regularization and dropout worked efficiently. We obtained approximately 0.7-8 for both

the C-index of OS and the AUC of the cancer progression rate, and these values were similar to those obtained in previously reported studies (Table 2).

The reason why NNs are important, despite the effectiveness of regression analysis, is that NNs are scalable to multidimensional image and language input without the need for manual feature extraction. As our next project, we are considering integrating a convolutional neural network for images into this research. We believe that imaging information such as CT scans, contoured structures, and dose distribution would be particularly rich and prognostically important in radiation therapy. Health care is shifting toward a more participative, patient-centered approach.[32] We hope that this research will lead to the development of

AI-based tools to aid in making treatment decisions, thereby contributing to both shared decision making and precision medicine.

This study has limitations. First, although we previously reported that SUVmax was a significant prognostic factor for OS,[33] we excluded it from the input features because there are concerns about the consistency of values obtained using different positron emission tomography-CT machines. Next, the data included records lacking pathologic confirmation, which may bias the result. Then, although the number of cases was larger than in previous studies and independent external tests were performed, the models were trained by the data from a single institution and the data had variation in clinical practice over a long period. Finally,

as all patients here underwent SBRT, this study did not provide a way to discriminate about those who should go to surgery. There are concerns about applying this prognostic approach uniformly to patients with operable cancers since many of those undergoing SBRT had inoperable tumors, and even if operable, many of the patients were elderly and/or frail.

In conclusion, the devised NNs predicted outcomes more accurately than other statistical and machine learning models methods and were also able to stratify patients with Tis–4N0M0 NSCLC treated with SBRT into high-, intermediate-, and low-risk cohorts. Our results are anticipated to open new avenues for NN predictions and provide decision-making guidance for patients and physicians.

AFFILIATIONS
[1]Radiation Oncology Center, Ofuna Chuo Hospital, Kamakura, Kanagawa, Japan
[2]Department of Radiology, Keio University School of Medicine, Tokyo, Japan
[3]Department of Radiation Oncology and Image-applied Therapy, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[4]Radiation Oncology Center, Tokyo General Hospital, Tokyo, Japan
[5]Biostatistics Unit, Clinical and Translational Research Center, Keio University Hospital, Tokyo, Japan
[6]Fujitsu Limited, Tokyo, Japan
[7]Department of Health Sciences, Saitama Prefectural University, Saitama, Japan

CORRESPONDING AUTHOR
Atsuya Takeda, MD, PhD, Department of Radiation Oncology, Ofuna Chuo Hospital, 6-2-24 Ofuna, Kamakura-shi, Kanagawa 247-0056, Japan; e-mail: takeda@1994.jukuin.keio.ac.jp.

AUTHOR CONTRIBUTIONS
Conception and design: Takafumi Nemoto, Atsuya Takeda, Takahisa Eriguchi, Etsuo Kunieda, Naoko Sanuki, Masamichi Yagi, Naoyuki Shigematsu
Financial support: Atsuya Takeda, Naoyuki Shigematsu

Administrative support: Atsuya Takeda, Naoyuki Shigematsu
Provision of study materials or patients: Atsuya Takeda, Naoyuki Shigematsu
Collection and assembly of data: Takafumi Nemoto, Atsuya Takeda, Yukinori Matsuo, Noriko Kishi, Takahisa Eriguchi, Yuichiro Tsurugai, Masamichi Yagi, Yousuke Aoki, Yohei Oku, Yuto Kimura, Naoyuki Shigematsu
Data analysis and interpretation: Takafumi Nemoto, Atsuya Takeda, Takahisa Eriguchi, Ryusei Kimura, Naoko Sanuki, Masamichi Yagi, Changhee Han, Naoyuki Shigematsu
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Atsuya Takeda
Consulting or Advisory Role: Accuray Japan K.K.
Research Funding: Varian Medical Systems

Yukinori Matsuo
Research Funding: Varian Medical Systems

No other potential conflicts of interest were reported.

REFERENCES
1. Leksell L: Stereotactic radiosurgery. J Neurol Neurosurg Psychiatry 46:797-803, 1983
2. Duma N, Santana-Davila R, Molina JR: Non-small cell lung cancer: Epidemiology, screening, diagnosis, and treatment. Mayo Clin Proc 94:1623-1640, 2019

3. Morgensztern D, Ng SH, Gao F, et al: Trends in stage distribution for patients with non-small cell lung cancer: A national cancer database survey. J Thorac Oncol 5:29-33, 2010

4. Brunelli A: It is not just about surgery versus stereotactic ablative radiotherapy, it is about curing as many patients with lung cancer as possible. J Thorac Cardiovasc Surg 156:1247-1248, 2018

5. Guerrero M, Li XA: Extending the linear-quadratic model for large fraction doses pertinent to stereotactic radiotherapy. Phys Med Biol 49:4825-4835, 2004

6. Subramanian MP, Meyers BF: Surgical resection versus stereotactic body radiation therapy for stage I NSCLC: Can randomized trials provide the solution? Cancers (Basel) 10:310, 2018

7. Shinde A, Li R, Kim J, et al: Stereotactic body radiation therapy (SBRT) for early-stage lung cancer in the elderly. Semin Oncol 45:210-219, 2018

8. Wolff HB, Alberts L, van der Linden N, et al: Cost-effectiveness of stereotactic body radiation therapy versus video assisted thoracic surgery in medically operable stage I non-small cell lung cancer: A modeling study. Lung Cancer 141:89-96, 2020

9. Stokes WA, Bronsert MR, Meguid RA, et al: Post-treatment mortality after surgery and stereotactic body radiotherapy for early-stage non-small-cell lung cancer. J Clin Oncol 36:642-651, 2018

10. Chen H, Louie AV, Boldt RG, et al: Quality of life after stereotactic ablative radiotherapy for early-stage lung cancer: A systematic review. Clin Lung Cancer 17:e141-e149, 2016

11. Dalwadi SM, Szeja SS, Bernicker EH, et al: Practice patterns and outcomes in elderly stage I non-small-cell lung cancer: A 2004 to 2012 SEER analysis. Clin Lung Cancer 19:e269-e276, 2018

12. Damhuis R, Senan S, Khakwani A, et al: Utilization rates of stereotactic body radiation therapy for the treatment of stage I NSCLC in three European countries. Ann Oncol 30:ii27-ii28, 2019

13. Abel S, Hasan S, White R, et al: Stereotactic ablative radiotherapy (SABR) in early stage non-small cell lung cancer: Comparing survival outcomes in adenocarcinoma and squamous cell carcinoma. Lung Cancer 128:127-133, 2019

14. Moreno AC, Fellman B, Hobbs BP, et al: Biologically effective dose in stereotactic body radiotherapy and survival for patients with early-stage NSCLC. J Thorac Oncol 15:101-109, 2020

15. Amin SA, Alam M, Baine MJ, et al: The impact of stereotactic body radiation therapy on the overall survival of patients diagnosed with early-stage non-small cell lung cancer. Radiother Oncol 155:254-260, 2020

16. Eriguchi T, Takeda A, Tsurugai Y, et al: Pleural contact decreases survival in clinical T1N0M0 lung cancer patients undergoing SBRT. Radiother Oncol 134:191-198, 2019

17. She Y, Jin Z, Wu J, et al: Development and validation of a deep learning model for non-small cell lung cancer survival. JAMA Netw Open 3:e205842, 2020

18. Zhou Z, Folkert M, Cannon N, et al: Predicting distant failure in early stage NSCLC treated with SBRT using clinical parameters. Radiother Oncol 119:501-504, 2016

19. Lou B, Doken S, Zhuang T, et al: An image-based deep learning framework for individualizing radiotherapy dose. Lancet Digit Health 1:e136-e147, 2019

20. Narita A, Takeda A, Eriguchi T, et al: Stereotactic body radiotherapy for primary non-small cell lung cancer patients with clinical T3-4N0M0 (UICC 8th edition): Outcomes and patterns of failure. J Radiat Res 60:639-649, 2019

21. Katzman JL, Shaham U, Cloninger A, et al: DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol 18:24, 2018

22. Wang H, Schabath MB, Liu Y, et al: Semiquantitative computed tomography characteristics for lung adenocarcinoma and their association with lung cancer survival. Clin Lung Cancer 16:e141-e163, 2015

23. Simon N, Friedman J, Hastie T, et al: Regularization paths for Cox's proportional hazards model via coordinate descent. J Stat Softw 39:1-13, 2011

24. Ishwaran H, Kogalur UB, Blackstone EH, et al: Random survival forests. Ann Appl Stat 2:841-860, 2008

25. Chen T, Guestrin C: XGBoost: A Scalable Tree Boosting System, 2016, arXiv:1603.02754

26. Kim DW, Lee S, Kwon S, et al: Deep learning-based survival prediction of oral cancer patients. Sci Rep 9:6994, 2019

27. Bice N, Kirby N, Bahr T, et al: Deep learning-based survival analysis for brain metastasis patients with the national cancer database. J Appl Clin Med Phys 21:187-192, 2020

28. Bzdok D, Altman N, Krzywinski M: Statistics versus machine learning. Nat Methods 15:233-234, 2018

29. Thawani R, McLane M, Beig N, et al: Radiomics and radiogenomics in lung cancer: A review for the clinician. Lung Cancer 115:34-41, 2018

30. Saillard C, Schmauch B, Laifa O, et al: Predicting survival after hepatocellular carcinoma resection using deep-learning on histological slides. Hepatology 72:2000-2013, 2020

31. Yousefi S, Amrollahi F, Amgad M, et al: Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. Sci Rep 7:11707, 2017

32. Walsh S, de Jong EEC, van Timmeren JE, et al: Decision support systems in oncology. JCO Clin Cancer Inform 3:1-9, 2019

33. Takeda A, Yokosuka N, Ohashi T, et al: The maximum standardized uptake value (SUVmax) on FDG-PET is a strong predictor of local recurrence for localized non-small-cell lung cancer after stereotactic body radiotherapy (SBRT). Radiother Oncol 101:291-297, 2011