

# Prediction of Response to Stereotactic Radiosurgery for Brain Metastases Using Convolutional Neural Networks

YU JIN CHA<sup>1,2</sup>, WON IL JANG<sup>1</sup>, MI-SOOK KIM<sup>1</sup>, HYUNG JUN YOO<sup>1</sup>,  
EUN KYUNG PAIK<sup>1</sup>, HEE KYUNG JEONG<sup>1</sup> and SANG-MIN YOUN<sup>3</sup>

*Departments of <sup>1</sup>Radiation Oncology, and <sup>3</sup>Neurosurgery,  
Korea Institute of Radiological & Medical Sciences, Seoul, Republic of Korea;  
<sup>2</sup>Department of Bio and Brain Engineering,  
Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea*

**Abstract.** *Background: It is unclear whether radiomic phenotypes of brain metastases (BM) are related to radiation therapy prognosis. This study assessed whether a convolutional neural network (CNN)-based radiomics model which learned computer tomography (CT) image features with minimal preprocessing, could predict early response of BM to radiosurgery. Materials and Methods: Tumor images of 110 BM post stereotactic-radiosurgery (SRS) (within 3 months) were assessed (Response Evaluation Criteria in Solid Tumor, version 1.1) as responders (complete or partial response) or non-responders (stable or progressive disease). Datasets were axial planning CT images containing the tumor center, and the tumor response. Datasets were randomly assigned to training, validation, or evaluation groups repeatedly, to create 50 dataset combinations that were classified into five groups of 10 different dataset combinations with the same evaluation datasets. The CNN learned using training-group images and labels. Validation datasets were used to choose the model that best classified evaluation images as responders or non-responders. Results: Of 110 tumors, 57 were classified as responders, and 53 as non-responders. The area under the receiver operating characteristic curve (AUC) of each CNN model for 50 dataset combinations ranged from 0.602 [95% confidence interval (CI)=36.5-83.9%] to 0.826 [95% CI, 64.3-100%]. The AUC of ensemble models, which averaged prediction results of 10 individual models within the same group, ranged from 0.761 (95% CI=55.2-97.1%) to 0.856 (95%*

*CI=68.2-100%). Conclusion: A CNN-based ensemble radiomics model accurately predicted SRS responses of unlearned BM images. Thus, CNN models are able to predict SRS prognoses from small datasets.*

The classical use and role of tumor images in radiation therapy planning has been restricted to defining treatment-target boundaries. However, as the field of medical-image analysis matures, it is expected that tumor imaging will include tumor biological and physiological information, which could inform treatment strategies. Radiomics is an emerging field of imaging analysis that quantifies information hidden in tumor images (1, 2). This concept is based on the conversion of tumor phenotypes, as represented in digital tumor images, into mineable high-dimensional feature data (3). Several studies have applied radiomics to the prediction of cancer prognosis, and demonstrated the potential of radiomic features to improve decision-making (4-6). To predict clinical prognosis using radiomic feature data, machine learning algorithms are applied, which are trained on feature data to recognize patterns and construct classification models after feature extraction (7). Traditional radiomics approaches rely on manually pre-defined features and hand-crafted feature extractions (Figure 1); recent studies have indicated that such approaches can be effective (4-6). However, the radiomics workflow is relatively complicated and cumbersome, and it is impossible to uncover features beyond those that were pre-defined. These reasons may explain the low acceptability of radiomics in clinical investigations.

*Correspondence to:* Won Il Jang, MD, Department of Radiation Oncology, Korea Institute of Radiological & Medical Sciences, 75 Nowon-ro, Nowon-gu, Seoul 01812, Republic of Korea. Tel: +82 29702484, Fax: +82 29702412, e-mail: zzang11@kirams.re.kr

**Key Words:** Brain metastases, radiomics, machine learning, radiosurgery, convolutional neural networks.

*The convolutional neural network (CNN) classifier.* An artificial neural network is a machine-learning technique that emulates an animal's neural structure for the classification of high-dimensional non-linear data or pattern recognition. Neural networks comprise several vertical layers, and each layer comprises numerous neural units. Each neural unit is connected to those of the next layer with constant weight and

bias. The initial weight and bias have random values. In a general neural-network structure, the first layer is called the input layer, while the final layer is called the output layer, as it expresses the results of the task. Between the input and output layers, there are hidden layers that perform constant operations or data processing. The number and arrangement of the hidden layers may be varied, or a layer for special operations can be inserted depending on the features of the task. When training data are input to the neural network, the neural units of the input layer multiply the input data by the connection weights and relay the product to the neural units of the next layer. The neural units of the next layer then generate an activation value by adding the sums of the values generated from the previous layer to the bias and inputting the value to an activation function. The activation value in turn becomes the inputted value for the next layer. Through this method, the output layer of a neural network presents the final value for the input data of a neural network. If the final output value differs from the targeted value, the connection weight is adjusted in a process called learning. If novel arbitrary data are inputted to an adequately trained neural network, the neural network can generate a value close to the target value according to the learned pattern.

A CNN is a type of feed-forward artificial neural network that is generally applied to image recognition. Categorizing medical images using conventional machine-learning algorithms requires a procedure involving pre-processing of the images, defining the inherent features, and extracting the output. In contrast, CNNs can automatically extract high-dimensional features from the original images, learn the patterns, and classify them (8). CNN was inspired by the neural structures of the animal visual cortex. The fundamental principle behind efficient learning of images by a CNN is the presence of a hidden layer that performs convolution and pooling operations, which is similar to the actual visual cortex (9).

*CNN-based radiomics for brain metastases (BMs).* CNNs have shown high performance in the classification of both natural and medical images (10-13). CNNs have the potential to improve the acceptability and applicability of radiomics because the approach requires minimal image preprocessing and no feature detector (Figure 1). To our knowledge, CNN-based radiomics has been used for tumor diagnosis (12) and predicting chemotherapy outcome for lung cancer (13), but not to predict radiation-therapy response. In this study, for the first time, we investigated the performance of CNN-based radiomics for predicting response after radiation therapy. We aimed to predict the response to stereotactic radiosurgery (SRS) for BMs, using CNN-based radiomics.

BMs, which may originate from any primary site, are a major cause of reduced quality of life, cognitive function, and life expectancy (14). As only a minority of patients with BMs are eligible for surgical resection (15), and SRS is a

relatively short, convenient, and noninvasive treatment course, it has become a common treatment modality for BMs. Moreover, BMs generally have a round shape and clear tumor boundary; thus, they have an appropriate geometry for radiomics research. Although several studies have assessed factors that predict prognosis of SRS for BM, including enhancement of patterns in CT images (16,17), as far as we are aware, no attempts have been made to use radiomics to predict the radio-response of BMs.

The objectives of this study were, firstly, to predict the SRS response of BMs using tumor images extracted from SRS planning CT, and secondly, to demonstrate the applicability of CNN-based radiomics combined with ensemble learning in the clinical radiation therapy field, using a small dataset.

## Materials and Methods

*Patients, target tumors, and treatment.* Medical records of all patients who received SRS for BMs at the Korea Institute of Radiological and Medical Sciences between 2007 and 2015 were analyzed. Patients with no imaging follow-up within 3 months of receiving SRS were excluded. Additional exclusions were patients with primary brain cancer, receipt of SRS combined with whole-brain radiation therapy, multiple BMs (more than five), skull-infiltrative BMs, and tumor with a longest diameter larger than 3.5 cm or smaller than 1 cm. Finally, 89 patients and 110 tumors were included in this study. All tumors were diagnosed by magnetic resonance (MR) imaging. SRS was performed using CyberKnife (Accuracy Inc., Sunnyvale, CA, USA) for 100 tumors, and RapidArc (Varian Medical System, Palo Alto, CA, USA) for 10 tumors. The gross tumor volume (GTV) was defined as the visible tumor extent on planning CT. The planning target volume (PTV) was defined as the GTV plus a 0-2 mm margin. Radiation doses were prescribed to the 80-95% isodose line of the maximum dose to cover the PTV. Tumor and treatment characteristics are summarized in Table I.

*Images and dataset acquisition.* For this study, two kinds of data were acquired from the treatment record for the 110 tumors. Firstly, one axial slice of each tumor was extracted. This contained the central point on the SRS planning CT images, as saved in the picture archiving communication system and extracted with the same CT window levels and widths for all tumors. Using the same scale (151×151 pixels; c. 40 mm × 40 mm), a square section of the image with tumor with a small portion of surrounding tissue was cropped. The cropped image contained the axial tumor centrally and the surrounding tissue marginally. The greatest tumor diameter enrolled in our study was less than 35 mm, so that each cropped image contained the entire axial tumor. No other image processing was carried out. Compared to MR, CT images are acquired under consistent, controlled conditions, such that if CT images are extracted for model learning, bias can be minimized in comparisons of images. MR data may have radiomic variations in images acquired due to differences in contrast time, equipment upgrades over time, and changes contrast agent use.

Secondly, the treatment response of each tumor in the 3 months following SRS was obtained. As per the Response Evaluation Criteria in Solid Tumor version 1.1 (RECIST 1.1) (18), the treatment results for the 110 tumor images were labeled as either

Table I. *Tumor characteristics.*

Characteristic	Responder (N=57)	Non-responder (N=53)	<i>p</i> -Value
Size, median (range), cm	2.0 (1.0-3.3)	2.2 (1.2-3.3)	0.205*
Primary site, n (%)			
Lung	20 (35%)	16 (30%)	0.051 <sup>†</sup>
Breast	23 (40%)	13 (25%)	
Rectum	4 (7%)	6 (11%)	
Stomach	3 (5%)	0 (0%)	
Kidney	2 (4%)	5 (9%)	
Other	5 (9%)	13 (25%)	
Location of BM, n (%)			
Cerebrum	43 (75%)	37 (70%)	0.520 <sup>†</sup>
Cerebellum	13 (23%)	13 (25%)	
Brain stem	1 (2%)	3 (5%)	
Patient age, years			
Median (range)	59 (14-92)	57 (43-76)	0.598*
Patient gender, n (%)			
Male	22 (39%)	22 (42%)	0.430 <sup>†</sup>
Female	35 (61%)	31 (58%)	
KPS Scale			
Median (range) <sup>‡</sup>	90 (40-100)	90 (50-100)	0.758*
Primary disease control, n (%) <sup>‡</sup>			
Yes	32 (56%)	37 (73%)	0.076 <sup>†</sup>
No	25 (44%)	14 (27%)	
Presence of extra-cranial metastases, n (%) <sup>‡</sup>			
Yes	29 (51%)	37 (73%)	0.021 <sup>†</sup>
No	28 (49%)	14 (27%)	
Systemic therapy, n (%) <sup>‡</sup>			
Yes	27 (47%)	14 (28%)	0.033 <sup>†</sup>
No	30 (53%)	37 (72%)	

BM: Brain metastasis; KPS: Karnofsky Performance Status. \*Mann-Whitney *U*-test, <sup>†</sup>Chi-squared test. <sup>‡</sup>N=51.

‘responder’ for those showing complete (CR) or partial (PR) response, or as ‘non-responder’ for those showing stable (SD) or progressive (PD) disease. The cropped image of the tumor and the treatment results together defined a data point, resulting in a dataset of 110 such points.

**Implementation and computation.** In this study, CNNs featured an architecture comprising a hidden layer and two convolution/max-pooling layers (Figure 2). Image data were converted to a matrix of pixels to be introduced in the input layer. Training used the drop-out technique, which probabilistically selects and trains a subset of the neurons in the hidden layer. This technique reduces overfitting in the process of creating a multilayer-deep neural-network learning model, thereby boosting overall performance (19). Between the final pooling layer and output layer were two fully connected neural layers. These performed additional feature extractions, and did not incorporate the drop-out technique. The final output layer consisted of two neural units. Outputs from the two neural units were compared to classify the input data as ‘responder’ or ‘non-responder.’ A stochastic gradient-descent algorithm, which is generally used in deep artificial neural networks for processing large-scale data, was used to update the connection weights of the neurons during learning (20). This neural network was based on Google’s machine-learning framework called TensorFlow (version 0.9) and configured in the Python

language. Amazon’s cloud GPU computing was used to enhance performance during learning operations.

**Training, validation, and evaluation.** One hundred and ten datasets were randomly classified into training, validation, and evaluation groups. This random assignment was repeated 50 times to create 50 independent combinations of datasets. Note that the following conditions were met for the random assignment: Firstly, the 50 combinations were clustered into five groups (A to E) of 10 combinations. Each group was assigned the same evaluation datasets but different training and validation datasets. Secondly, the ratio of responder to non-responder datasets remained consistent across the five groups.

Next, the experiment for each dataset combination proceeded as follows. Training datasets were used to train the neural networks. The neural networks learned the radiomic features of responders and non-responders by learning the labels that matched the training images. However, learning performance can vary according to the initial connection weights, which were randomly determined. Hence, the validation datasets were used to validate performance during learning. Validation proceeded by periodically presenting a validation image at a fixed trial-interval to the neural network undergoing the learning process. The network then predicted the matched label. When the accuracy of prediction was low, learning was reinitialized and restarted. When the network demonstrated good performance or did

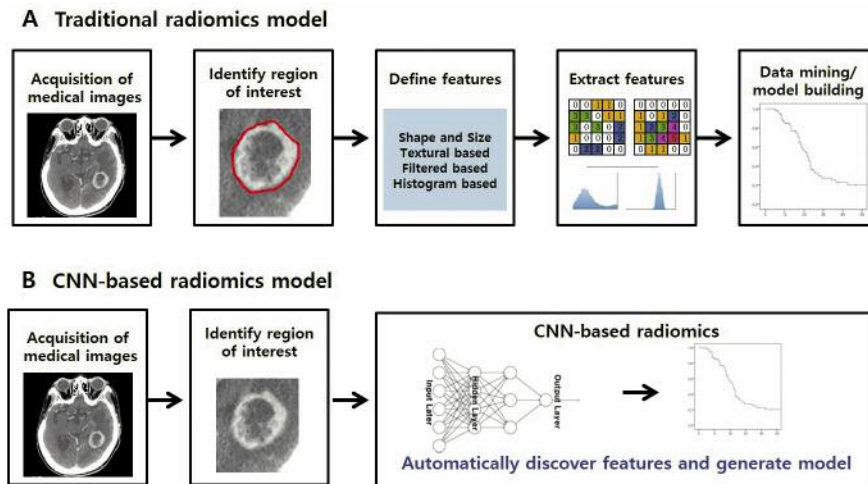


Figure 1. Overview of methodological process of traditional (A) and convolutional neural networks (CNN)-based (B) radiomics. The process of defining and extracting features to generate output is performed directly by the investigator in traditional radiomics but performed automatically by an algorithm in a CNN-based model.

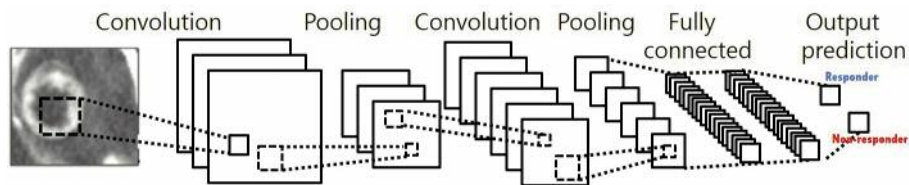


Figure 2. Convolutional neural networks architecture for computed tomography images of brain metastases in our study. This neural network comprised an input layer, two convolution/pooling layers, and two fully connected layers. The output layer generated a classification output as either R (responder) or N (non-responder) for the input image.

not show improved performance, learning was stopped and the model was accepted and proceeded to the evaluation stage. That is, validation was used to discover the best model among the candidate models during training. Finally, an evaluation image was presented to the neural network that completed learning to generate a classification output. Each neural network was designed such that it generated an output of 1 for an image determined to be a non-responder and 0 for an image determined to be a responder.

Independently of the classification outputs from the 50 individual neural network models, the classification outputs generated by the neural networks of the same group (which shared the same evaluation datasets) were averaged to determine the final result. This model is defined as an ensemble model (Figure 3). Performance of the ensemble model was evaluated using the area under the receiver operating characteristic curve (AUC).

## Results

**Tumor and treatment characteristics.** The neural networks classified 57 images as responders and 53 images as non-responders. The maximum tumor diameter was 1.0-3.3 cm in the responder group (median=2.0 cm) and 1.2-3.3 cm in the

non-responder group (median=2.2 cm), with no significant differences between the two groups. With regard to the primary tumor site, the greatest number of tumors in the responder group were primarily of breast (40%), followed by lung (35%), rectum (7%), and other sites. In the non-responder group, the most frequent site was lung (30%), followed by breast (25%), rectum (11%), and other sites. The difference in the distribution of primary tumors between the two groups approached statistical significance ( $p=0.051$ ). There was no statistically significant difference between the responder and non-responder groups in terms of the tumor's intracranial position. Patient age ranged from 14-92 years (median=59 years) in the responder group and 43-76 years (median=57 years) in the non-responder group; there was no significant difference in age between the two groups. The two groups also did not significantly differ in terms of patient sex and Karnofsky performance status scale. Furthermore, there was no difference with respect to whether the primary disease site was completely controlled by surgery or radiation therapy compared to cases without such control. However, there was a significant difference in response of

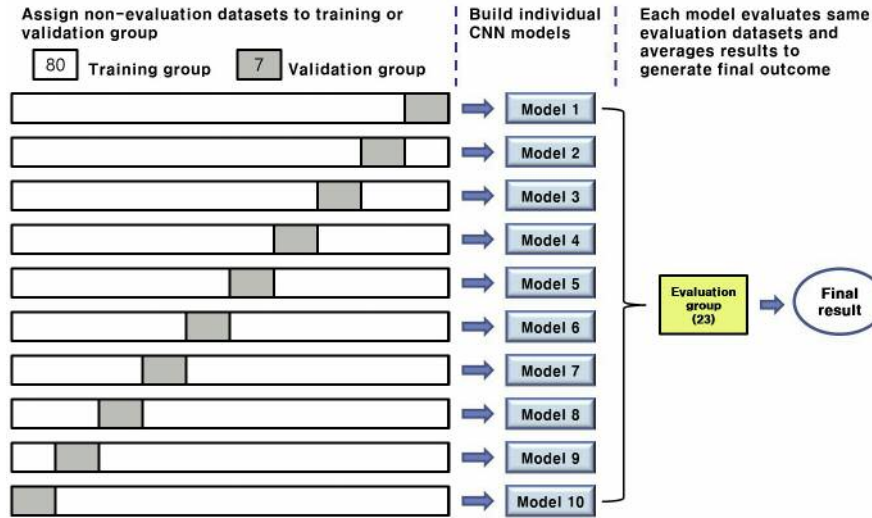


Figure 3. Outline of the ensemble prediction model. Eighty-seven datasets that were not assigned to the evaluation group were assigned to the training and validation groups in 10 varying combinations to create 10 individual convolutional neural networks models. The final result was generated by taking the average of the classification results for the same evaluation image from 10 individual models.

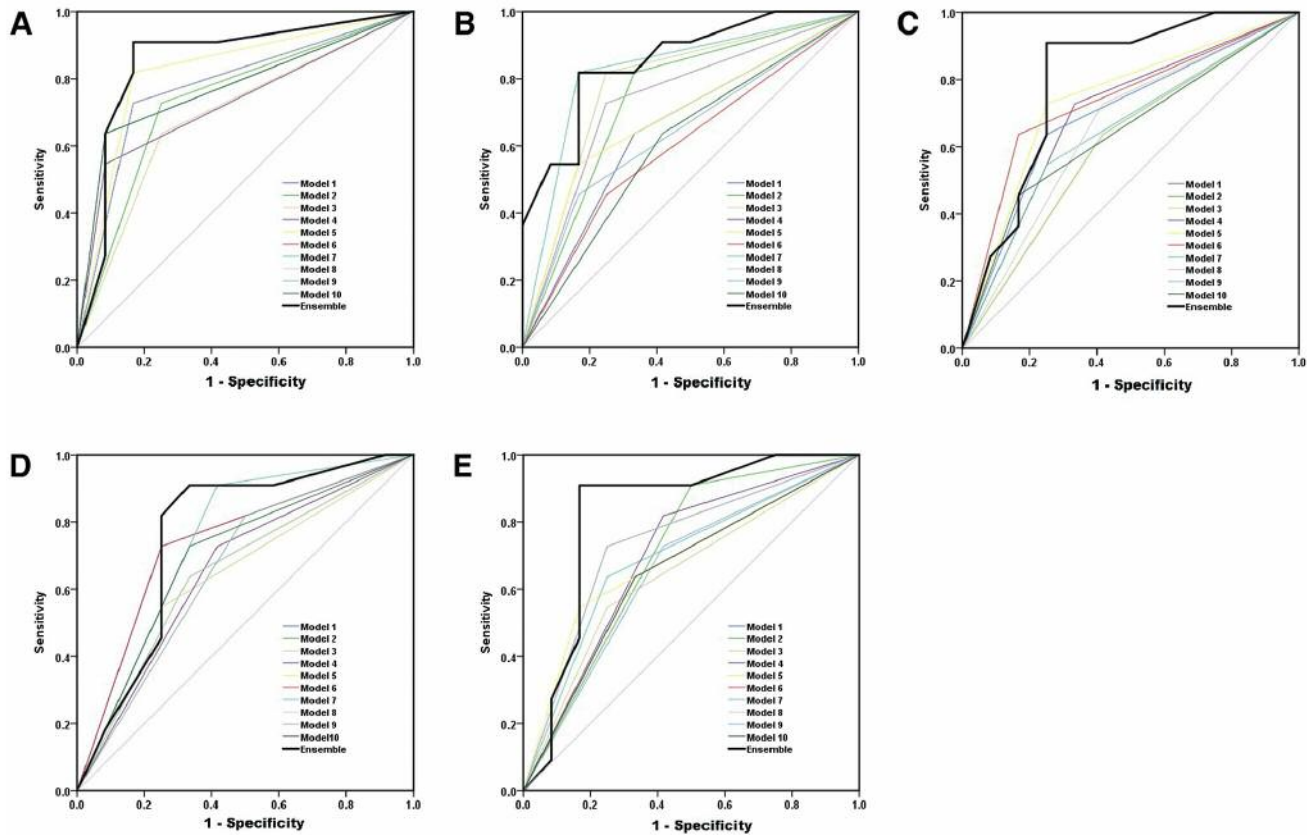


Figure 4. Ensemble model for predicting early responses to stereotactic radiosurgery and the receiver operating characteristic curves for individual neural-network model groups A to E, respectively, that comprised the ensemble model. The area under the curve data are given in Table III.

Table II. *Treatment characteristics.*

Characteristic	Responder (N=57)	Non-responder (N=53)	p-Value
Total dose, median (range), Gy	24 (19-48)	23 (16-36)	0.395*
No. of fractions, median (range)	3 (1-5)	3 (1-4)	
Biological effective dose, median (range) Gy10	60 (33-94)	60 (30-100)	0.094*
Treatment platform, n (%)			0.587†
Cyberknife	51 (89%)	49 (92%)	
Linac (RapidArc)	6 (11%)	4 (8%)	

\*Mann-Whitney *U*-test, †Chi-squared test.

extra-cranial metastases, except for the primary site, and for those who did not receive systemic therapy for BMs (Table I). Systemic therapy refers to chemotherapy, hormone therapy, and target therapy, depending on the type of primary tumor used alone or in combination. Specific analysis of systemic therapy regimen was not performed.

The total dose of radiation ranged from 19-48 Gy (median 24 Gy) in 1-5 fractions in the responder group and from 16-36 Gy in 1-4 fractions in the non-responder group, with no significant differences between the two groups. The two groups also did not significantly differ in the biologically effective dose, or treatment platform (Table II).

**Prediction results.** The independent treatment prediction results for groups A to E were as follows. The AUC of the predictions generated by each of the 10 neural networks in group A ranged from 0.693 [95% confidence interval (CI)=47.0-91.6%] to 0.826 (95% CI=70.3-100%) (median 0.778). The AUC of the ensemble model, which determined the result based on the average of the outputs generated by each of the 10 neural networks, was 0.856 (95% CI=68.2-100%). When the cut-off point was set to 0.5 for the ensemble model, the sensitivity and specificity for prediction of non-responders were 82% and 83%, respectively. The results for groups B to E are detailed in Table III.

In summary, the AUC generated by the ensemble model for the evaluation data of groups A through E were 0.856 (95% CI=68.2-100%,  $p=0.004$ ), 0.856 (95% CI=70.2-100%,  $p=0.004$ ), 0.799 (95% CI=60.8-99.0%,  $p=0.015$ ), 0.761 (95% CI=55.2-97.1%,  $p=0.034$ ), and 0.826 (95% CI=63.7-100%,  $p=0.008$ ), respectively, which were either comparable to or superior to the results generated from individual neural networks (Figure 4).

## Discussion

The emergence of radiomics has provided an impetus to research assessing whether tumor images can predict radiation therapy prognosis. Mattonen *et al.* reported that radiomics analysis significantly predicts the probability of

recurrence of lung cancer after radiation therapy (21). However, there has been no study that applied CNN-based radiomics to predict the prognosis of radiation therapy. Hence, this study analyzed radiation therapy planning CT images with CNNs to determine whether CNNs can predict early response to SRS.

We made two assumptions prior to beginning this study. The first was that microenvironments within tumors are determined by the pathophysiology of the tumor, and thereby are related to the tumor's response to treatment (22). The key concept of radiomics is that if the microenvironment of a tumor is exhibited as a characteristic phenotype in images, analyzing the images would enable the prediction of treatment response. George *et al.* (16) and Goodman *et al.* (17) reported that local control post-SRS is associated with the contrast patterns in BM and MRI, but these studies are limited in that they only qualitatively classified the contrast patterns. Evidence for the presence of a correlation between BM pathophysiology and their phenotypes in images is lacking in the literature. Therefore, we assumed that phenotypes of BMs shown in CT images reflect the tumor microenvironment, which determines the prognosis of radiation therapy.

The second assumption was that BMs are generally round with clear margins (23). Thus, we assumed that BMs display similar radiomic phenotypic patterns radially from a central point of the tumor. Based on this assumption, we would be able to identify the overall radiomic features of the tumor using a cross-sectional CT image containing the center of the tumor.

In this analysis, the CNN-based machine-learning model was taught pairs of tumor images and responses to SRS and then predicted SRS responses for unlearned images. Fifty independent neural networks used for the analysis showed a prediction AUC ranging from 0.602 to 0.826 (median=0.695). However, the AUC values of the ensemble models, each of which comprised 10 individual neural networks that assessed the same evaluation datasets, ranged from 0.761 (95% CI=55.2-97.1%) to 0.856 (95% CI=68.2-100%), thus showing better predictive performance than the individual neural networks. We cannot compare the performance of this model to that of another model, as no studies have used radiomics to predict BM responses to

Table III. Performance (prediction of treatment responses) of individual convolutional neural networks (CNN) and ensemble models.

Group	Individual CNN model										Ensemble model
	1	2	3	4	5	6	7	8	9	10	
A AUC	0.780	0.739	0.693	0.731	0.826	0.780	0.777	0.780	0.780	0.777	0.856
p-Value	0.023	0.053	0.117	0.060	0.008	0.023	0.025	0.023	0.023	0.025	0.004
95% CI	0.580-0.981	0.527-0.951	0.470-0.916	0.515-0.947	0.643-1.000	0.580-0.981	0.574-0.979	0.580-0.981	0.580-0.981	0.574-0.979	0.682-1.000
Sensitivity (%)	73	73	64	55	82	73	64	73	73	64	82
Specificity (%)	83	75	75	92	83	83	92	83	83	92	83
B AUC	0.652	0.742	0.784	0.739	0.689	0.602	0.826	0.739	0.644	0.610	0.856
p-Value	0.218	0.049	0.021	0.053	0.124	0.406	0.008	0.053	0.242	0.372	0.004
95% CI	0.422-0.882	0.532-0.953	0.586-0.982	0.527-0.951	0.465-0.914	0.365-0.839	0.643-1.000	0.527-0.951	0.412-0.876	0.375-0.845	0.702-1.000
Sensitivity (%)	64	82	82	73	55	45	82	73	45	64	82
Specificity (%)	67	67	75	75	83	75	83	75	83	58	83
C AUC	0.693	0.610	0.610	0.697	0.739	0.735	0.648	0.655	0.693	0.644	0.799
p-Value	0.117	0.372	0.372	0.110	0.053	0.056	0.230	0.207	0.117	0.242	0.015
95% CI	0.470-0.916	0.375-0.845	0.375-0.845	0.475-0.918	0.527-0.951	0.521-0.949	0.417-0.879	0.426-0.884	0.470-0.916	0.412-0.876	0.608-0.990
Sensitivity (%)	64	64	64	73	64	66	55	73	64	45	91
Specificity (%)	75	58	58	67	75	83	75	58	75	83	75
D AUC	0.739	0.652	0.648	0.655	0.659	0.739	0.746	0.652	0.659	0.697	0.761
p-Value	0.053	0.218	0.230	0.207	0.196	0.053	0.045	0.218	0.196	0.110	0.034
95% CI	0.527-0.951	0.422-0.882	0.417-0.879	0.426-0.884	0.431-0.887	0.527-0.951	0.538-0.954	0.422-0.882	0.431-0.87	0.475-0.918	0.552-0.971
Sensitivity (%)	73	64	55	73	82	73	91	64	82	73	82
Specificity (%)	75	67	75	58	50	75	58	67	50	67	75
E AUC	0.739	0.705	0.648	0.701	0.689	0.652	0.693	0.739	0.655	0.652	0.826
p-Value	0.053	0.097	0.230	0.103	0.124	0.218	0.117	0.053	0.207	0.218	0.008
95% CI	0.527-0.951	0.486-0.923	0.417-0.879	0.481-0.921	0.465-0.914	0.422-0.882	0.470-0.916	0.527-0.951	0.426-0.884	0.422-0.882	0.637-1.000
Sensitivity (%)	73	91	55	82	55	64	64	73	73	64	82
Specificity (%)	75	50	75	58	73	67	75	75	58	67	83

AUC, Area under the receiver operating characteristics curve; CI=confidence interval.

radiation therapy. However, the sensitivity and specificity of the ensemble models for recognizing non-responders were superior when the cut-off value was set to 0.5. As the ensemble technique increases the number of datasets used for learning (24,25), we speculate that this approach could compensate for the problem of small datasets in radiomics research.

Furthermore, considering the promising predictive performance found in this study, the assumption that radiological phenotypes are correlated with microenvironments that determine the tumor's treatment response may be supported. There were no statistically significant differences in other major parameters between the responder and non-responder groups, except for the presence of extra-cranial metastases and systemic therapy (Tables I and II). Thus, tumor radiological phenotypes may be more likely predictive of treatment response than other clinical factors. However, it seems that the presence of extra-cranial metastases and receipt of systemic therapy may affect intra-cranial tumor burden, and may partially relate to local control of BMs.

We only used one axial image, which contained the center of the tumor, for analysis instead of using the entire image data. When limited to BMs, it is reasonable to assume that one cross-sectional image containing the center of the tumor represents the overall radiomic features of the tumor. Furthermore, the use of one cross-sectional image extracted from the planning CT images for analysis, as in this study, may improve clinical applicability.

However, the design of this study has some limitations. Firstly, evidence for the effectiveness of a single fraction SRS for BMs has been established, but there is a lack of evidence for effectiveness of multi-fraction SRS for BMs. Nevertheless, a linear-quadratic model was introduced to simplify the problem in order to compensate for differences in biological treatment effects between the two groups. However, multi-fraction SRS has a high fraction size, which makes it difficult to compare strict doses between single and multi-fraction SRS. Second, systemic therapy needs to be considered as an important factor affecting local control of tumors. However, analysis of local control taking into consideration the types of systemic therapies and the details of regimens was not performed in this study. Thirdly, we used CT images that can be obtained under consistent conditions and that most appropriately reflect the timing of treatment initiation in model analysis. However, if MR images were available with consistent conditions and imaging parameters, it might be possible to improve the predictive accuracy by developing MR-based machine learning models that incorporate more detailed radiomic information.

We also need to review data that failed to accurately predict treatment outcome. Although we did not perform dataset-specific analysis of prediction failures, the following reasons may be involved. Firstly, CNN is known to be capable of learning high-dimensional features of images, but a great deal

of learning data is required to achieve good performance (26). However, our data only comprised 87 sets, which is a relatively small number, even considering the ensemble learning approach. The small data size hindered the ability of the predictive model to learn generalized boundaries of radiomic phenotypes that distinguished responders from non-responders. Secondly, tumor responses PR and SD were labeled as responder and non-responder, respectively, as per RECIST version 1.1. However, the demarcations for PR and SD are very thin, and it is unclear whether tumors that show PR and those that show SD can be viewed as having different radiomic properties. Thus, it is possible that the predictive models generated conflicting results for tumors that were between PR and SD. Thirdly, other predictors of tumor treatment may exist in addition to the clinical factors analyzed in this study. As we did not control for potential predictors, our results should be carefully interpreted. Fourthly, the findings of progression at the 3-month follow-up imaging of BMs did not exclude the possibility of pseudoprogression due to edema or hemorrhage. Indeed, making a clinical distinction between progression and pseudoprogression is not easy. Despite these limitations, this study demonstrated the possibility of predicting BM response to SRS through CNN-based radiomic analysis, even in a small dataset. Future studies should address these limitations and utilize multicenter data to improve the predictive performance of the model so as to produce data useful for clinical treatment decisions.

## Conclusion

A CNN-based ensemble radiomics model that learned SRS from planning CT images for BMs and known early responses predicted the SRS responses of unlearned images of BMs with high accuracy. Hence, the radiomic phenotypes of BMs shown in CT images might be correlated with tumor response to SRS. This study is the first to suggest the use of CNN models to predict radiation therapy prognoses with small-scale data. However, additional studies are required to improve the performance of this model so that it may be utilized to help make clinical treatment decisions.

## Funding

None.

## Conflicts of Interest

The Authors have no conflicts of interest that are relevant to this report.

## Acknowledgements

This study was supported by a grant from the Korea Institute of Radiological and Medical Science (KIRAMS), funded by the Ministry of Science and ICT, Republic of Korea (50543-2017).



# References

- 1 Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebbers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ and Lambin P: Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 5: 4006, 2014.
- 2 Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJWL, Dekker A, Fenstermacher D, Goldof DB, Hall LO, Lambin P, Balagurunathan Y, Gatenby RA and Gillies RJ: Radiomics: the process and the challenges. *Magn Reson Imaging* 30: 1234-1248, 2012.
- 3 Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A and Aerts HJWL: Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 48(4): 441-446, 2012.
- 4 Gevaert O, Xu J, Hoang C, Leung AN, Xu Y, Quon A, Rubin DL, Napel S and Plevritis SK: Nonsmall cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data. *Radiology* 2(4): 387-396, 2012.
- 5 Khalvati F, Wong A and Haider MA: Automated prostate cancer detection *via* comprehensive multiparametric magnetic resonance imaging texture feature models. *BMC Med Imaging* 15(1): 27, 2015.
- 6 Fried DV, Tucker SL, Zhou S, Liao Z, Mawlawi O, Ibbott G and Court LE: Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer. *Int J Radiat Oncol Biol Phys* 90: 834-842, 2014.
- 7 Gillies RJ, Kinahan PE and Hricak H: Radiomics: Images are more than pictures, they are data. *Radiology* 278(2): 563-577, 2016.
- 8 Cireşan DC, Meier U, Masci J, Gambardella LM and Schmidhuber J: Flexible, high-performance convolutional neural networks for image classification. In: *IJCAI'11 Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence-Volume Two*. Barcelona, Catalonia, Spain: AAAI Press pp. 1237-1242, 2011.
- 9 Cadieu CF, Ha H, Yamins DLK, Pinto N, Ardila D, Solomon EA, Majaj NJ and DiCarlo JJ: Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLOS Comput Biol* 10: e1003963, 2014.
- 10 Krizhevsky A, Sutskever I and Hinton G: ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems* 25. Pereira F, Burges C, Bottou L and Weinberger K (eds.) Nevada: Curran Associates, Inc., pp. 1097-1105, 2012.
- 11 Ciresan DC, Meier U and Schmidhuber J: Multi-column deep neural networks for image classification. In: *Proceedings of the 25th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3642-3649, 2012.
- 12 Cireşan DC, Giusti A, Gambardella LM and Schmidhuber J: Mitosis detection in breast cancer histology images with deep neural networks. *Med Image Comput Comput Assist Interv* 16: 411-418, 2013.
- 13 Ypsilantis P-P, Siddique M, Sohn H-M, Davies A, Cook G, Goh V and Montana G: Predicting response to neoadjuvant chemotherapy with PET imaging using convolutional neural networks. *PLoS ONE* 10(9): e0137036, 2015.
- 14 Tsao MN, Lioyd NS, Wong RK, Rakovitch E, Chow E and Laperriere N: Radiotherapeutic management of brain metastases: A systematic review and meta-analysis. *Cancer Treat Rev* 31: 256-273, 2005.
- 15 DeAngelis LM: Management of brain metastases. *Cancer Invest* 12: 156-165, 1994.
- 16 George R, Jaap Z, Andrew W and Lagerwaard F: Recursive partitioning analysis for the prediction of stereotactic radiosurgery brain metastases lesion control. *Oncologist* 18: 330-335, 2013.
- 17 Goodman KA, Sneed PL, McDermott MW, Shiau CY, Lamborn KR, Chang S, Park E, Wara WM and Larson DA: Relationship between pattern of enhancement and local control of brain metastases after radiosurgery. *Int J Radiat Oncol Biol Phys* 50: 139-146, 2001.
- 18 Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D and Verweij J: New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 45(2): 228-247, 2009.
- 19 Srivastava N, Hinton GE, Krizhevsky A, Sutskever. I and Salakhutdinov R: Dropout: a simple way to prevent neural networks from overfitting. *J Machine Learn Res* 15: 1929-1958, 2014.
- 20 Bordes. A, Bottou L and Gallinari P: SGD-QN: careful quasi-newton stochastic gradient descent. *J Machine Learn Res* 10: 1737-1754, 2009.
- 21 Mattonen SA, Palma DA, Johnson C, Louie AV, Landis M, Rodrigues G, Chan I, Etemad-Rezai R, Yeung TP, Senan S and Ward AD: Detection of local cancer recurrence after stereotactic ablative radiation therapy for lung cancer: physician performance versus radiomic assessment. *Int J Radiat Oncol Biol Phys* 94(5): 1121-1128, 2016.
- 22 Son B, Lee S, Youn H, Kim E, Kim W and Youn B: The role of tumor microenvironment in therapeutic resistance. *Oncotarget* 8(3): 3933-3945, 2017.
- 23 Adler JR, Cox RS, Kaplan I and Martin DP: Stereotactic radiosurgical treatment of brain metastases. *J Neurosurg* 76: 444-449, 1992.
- 24 Rokach L: Ensemble-based classifiers. *Artif Intell Rev* 33: 1, 2010.
- 25 Ortiz A, Munilla J, Gorris JM and Ramírez J: Ensembles of deep learning architectures for the early diagnosis of Alzheimer's disease. *Int J Neural Syst* 26(7): 1650025, 2016.
- 26 Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB and Liang J: Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Trans Med Imag* 35(5): 1299-1312, 2016.

Received July 13, 2018

Revised July 27, 2018

Accepted July 30, 2018