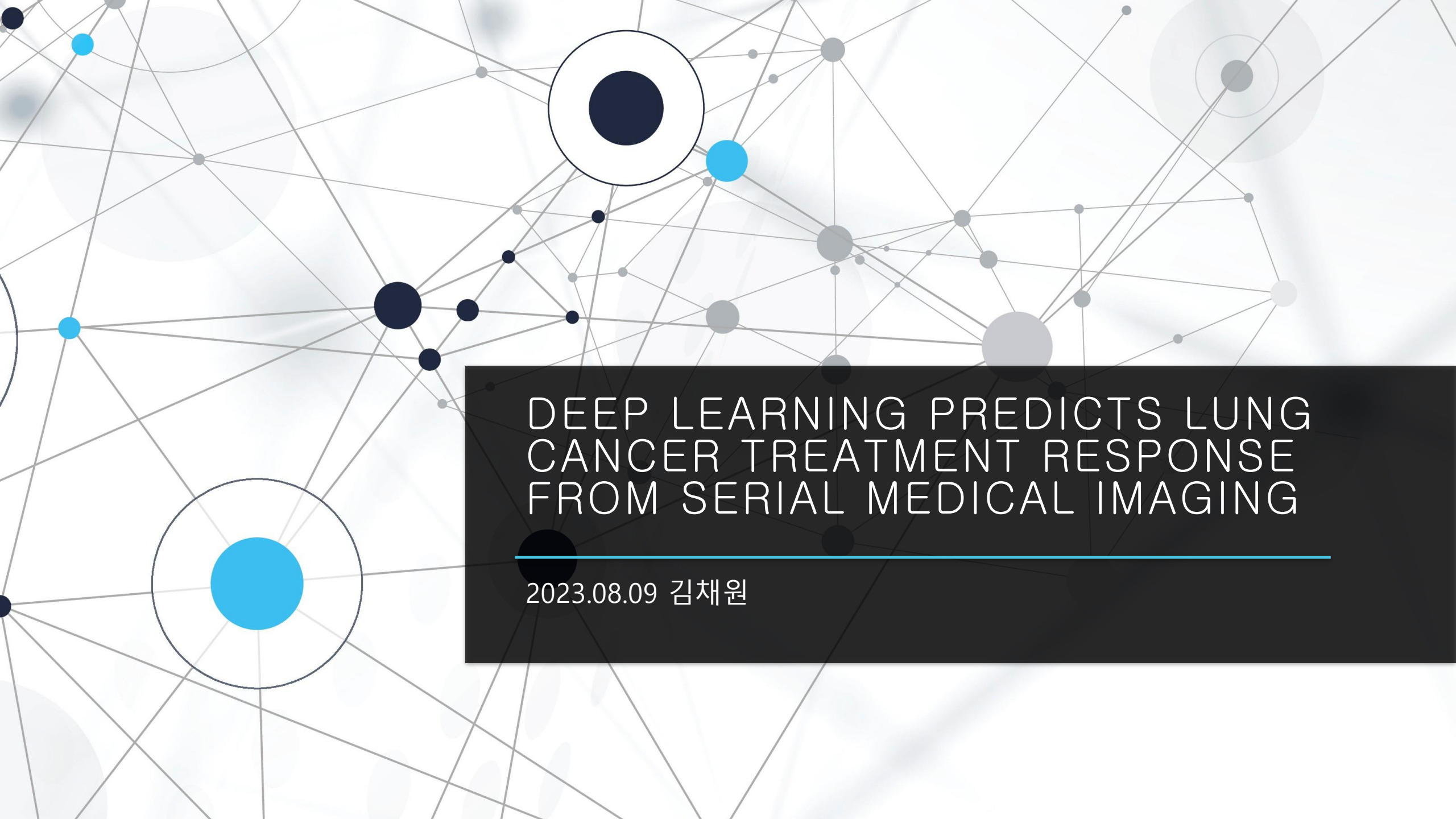
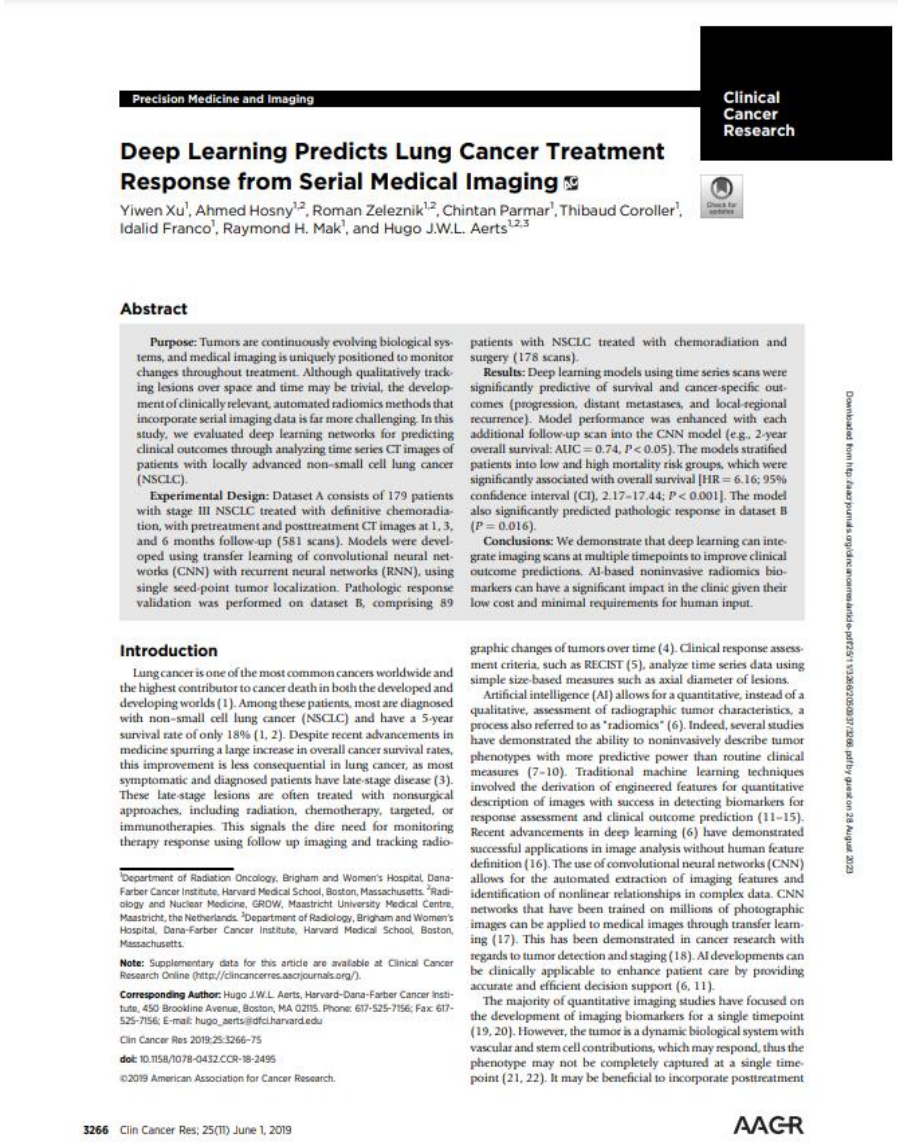


DEEP LEARNING PREDICTS LUNG CANCER TREATMENT RESPONSE FROM SERIAL MEDICAL IMAGING

2023.08.09 김채원



Precision Medicine and Imaging

Clinical Cancer Research

Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging

Yiwen Xu[1], Ahmed Hosny[1,2], Roman Zeleznik[1,2], Chintan Parmar[1], Thibaud Coroller[1], Idalid Franco[1], Raymond H. Mak[1], and Hugo J.W.L. Aerts[1,2,3]

Abstract

Purpose: Tumors are continuously evolving biological systems, and medical imaging is uniquely positioned to monitor changes throughout treatment. Although qualitatively tracking lesions over space and time may be trivial, the development of clinically relevant, automated radiomics methods that incorporate serial imaging data is far more challenging. In this study, we evaluated deep learning networks for predicting clinical outcomes through analyzing time series CT images of patients with locally advanced non–small cell lung cancer (NSCLC).

Experimental Design: Dataset A consists of 179 patients with stage III NSCLC treated with definitive chemoradiation, with pretreatment and posttreatment CT images at 1, 3, and 6 months follow-up (581 scans). Models were developed using transfer learning of convolutional neural networks (CNN) with recurrent neural networks (RNN), using single seed-point tumor localization. Pathologic response validation was performed on dataset B, comprising 89 patients with NSCLC treated with chemoradiation and surgery (178 scans).

Results: Deep learning models using time series scans were significantly predictive of survival and cancer-specific outcomes (progression, distant metastases, and local-regional recurrence). Model performance was enhanced with each additional follow-up scan into the CNN model (e.g., 2-year overall survival: AUC = 0.74, $P < 0.05$). The models stratified patients into low and high mortality risk groups, which were significantly associated with overall survival [HR = 6.16; 95% confidence interval (CI), 2.17–17.44; $P < 0.001$]. The model also significantly predicted pathologic response in dataset B ($P = 0.016$).

Conclusions: We demonstrate that deep learning can integrate imaging scans at multiple timepoints to improve clinical outcome predictions. AI-based noninvasive radiomics biomarkers can have a significant impact in the clinic given their low cost and minimal requirements for human input.

Introduction

Lung cancer is one of the most common cancers worldwide and the highest contributor to cancer death in both the developed and developing worlds (1). Among these patients, most are diagnosed with non–small cell lung cancer (NSCLC) and have a 5-year survival rate of only 18% (1, 2). Despite recent advancements in medicine spurring a large increase in overall cancer survival rates, this improvement is less consequential in lung cancer, as most symptomatic and diagnosed patients have late-stage disease (3). These late-stage lesions are often treated with nonsurgical approaches, including radiation, chemotherapy, targeted, or immunotherapies. This signals the dire need for monitoring therapy response using follow up imaging and tracking radiographic changes of tumors over time (4). Clinical response assessment criteria, such as RECIST (5), analyze time series data using simple size-based measures such as axial diameter of lesions.

Artificial intelligence (AI) allows for a quantitative, instead of a qualitative, assessment of radiographic tumor characteristics, a process also referred to as "radiomics" (6). Indeed, several studies have demonstrated the ability to noninvasively describe tumor phenotypes with more predictive power than routine clinical measures (7–10). Traditional machine learning techniques involved the derivation of engineered features for quantitative description of images with success in detecting biomarkers for response assessment and clinical outcome prediction (11–15). Recent advancements in deep learning (6) have demonstrated successful applications in image analysis without human feature definition (16). The use of convolutional neural networks (CNN) allows for the automated extraction of imaging features and identification of nonlinear relationships in complex data. CNN networks that have been trained on millions of photographic images can be applied to medical images through transfer learning (17). This has been demonstrated in cancer research with regards to tumor detection and staging (18). AI developments can be clinically applicable to enhance patient care by providing accurate and efficient decision support (6, 11).

The majority of quantitative imaging studies have focused on the development of imaging biomarkers for a single timepoint (19, 20). However, the tumor is a dynamic biological system with vascular and stem cell contributions, which may respond, thus the phenotype may not be completely captured at a single timepoint (21, 22). It may be beneficial to incorporate posttreatment

[1]Department of Radiation Oncology, Brigham and Women's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts. [2]Radiology and Nuclear Medicine, GROW, Maastricht University Medical Centre, Maastricht, the Netherlands. [3]Department of Radiology, Brigham and Women's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Corresponding Author: Hugo J.W.L. Aerts, Harvard–Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02115. Phone: 617-525-7156; Fax: 617-525-7156; E-mail: hugo_aerts@dfci.harvard.edu

Clin Cancer Res 2019;25:3266–75

doi: 10.1158/1078-0432.CCR-18-2495

AACR

Journal : AACR (American Association for Cacner Reasearch)

Published date: 2018.08.15

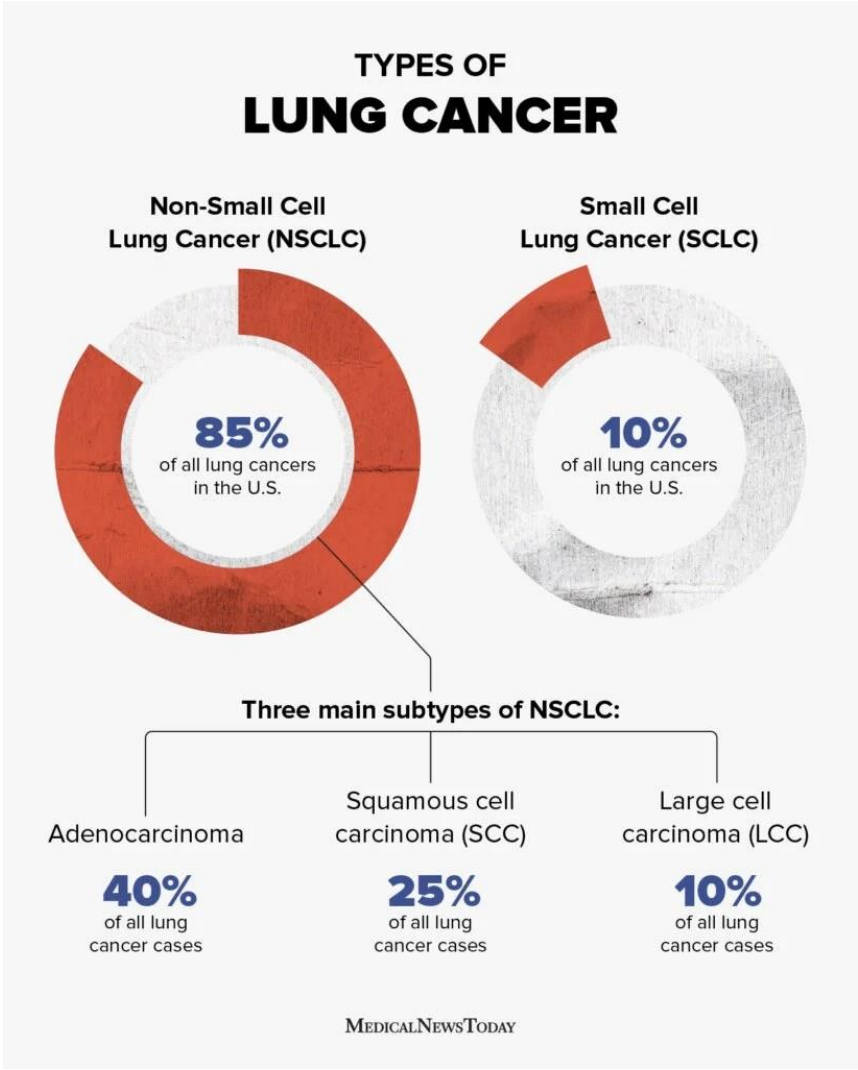
Author : Yiwen Xu, Ahmed Hosny, Chintan Parmar et al.

(Dana-Farber Cacner Institute, Harvard Medical School)

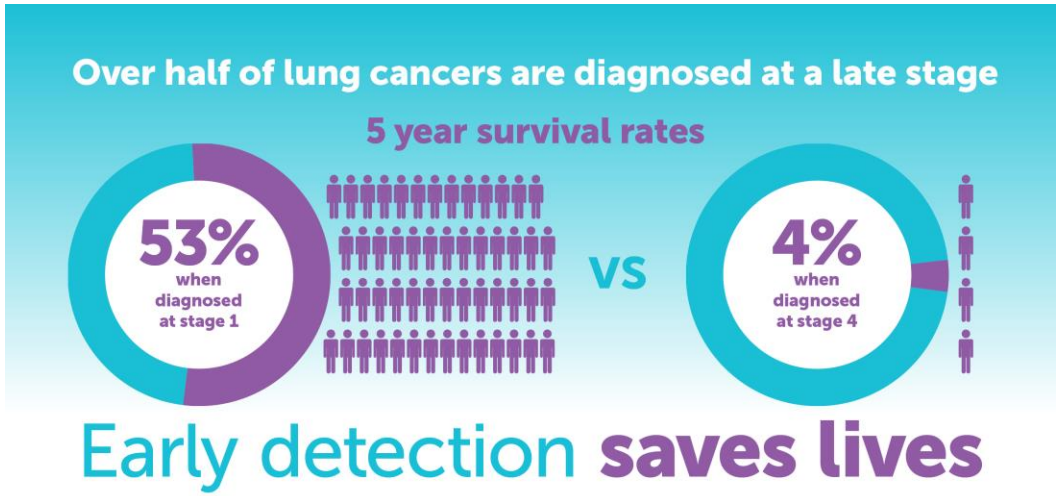
1.Introduction

- Lung cancer is one of the most common cancers worldwide and the highest contributor to cancer death in both the developed and developing worlds

- Among these patients, most are diagnosed with non–small cell lung cancer (NSCLC) and have a 5-year survival rate of only 18%.



1.Introduction



- Despite recent advancements in medicine spurring a large increase in overall, cancer survival rates, this improvement is less consequential in lung cancer, as most symptomatic and diagnosed patients have late-stage disease.

1.Introduction

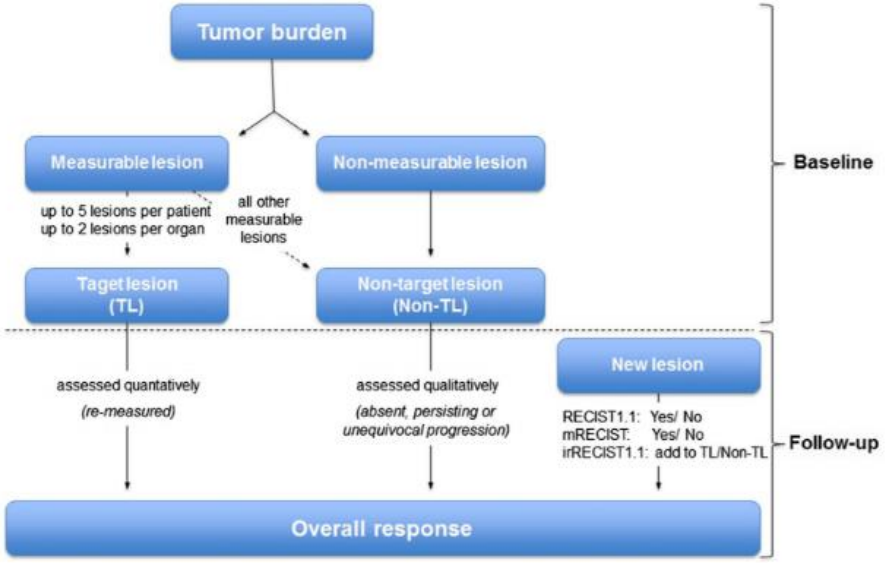
- late-stage lesions are often treated with nonsurgical approaches

1. radiation
2. chemotherapy
3. targeted (표적치료)
4. immunotherapies

1.Introduction

- Clinical response assessment criteria, such as RECIST, analyze time series data using simple size-based measures such as axial diameter of lesions.

▶ RECIST



1.measurable

X-ray, CT, MRI 영상에서 가장 긴 직경의 길이 ≥ 20mm

or Spiral CT 영상에서 가장 긴 직경의 길이 ≥ 10mm

2.nonmesurable

그 외의 다른 모든 병변

1.Introduction

- Artificial intelligence (AI) allows for a quantitative, instead of a qualitative, assessment of radiographic tumor characteristics, a process also referred to as “Radiomics”

- Indeed, several studies have demonstrated the ability to noninvasively describe tumor phenotypes with more predictive power than routine clinical measures

1.Introduction

- Traditional machine learning techniques required human feature definition to extract biomarkers, whereas recent advancements don’t require human feature definition.

- The use of CNN allows for the automated extraction of imaging features and identification of nonlinear relationships in complex data.

1.Introduction

- CNN networks that have been trained on millions of photographic images can be applied to medical images through transfer learning.

- This has been demonstrated in cancer research with regards to tumor detection and staging(classification).

(Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 2016;35:1285–98.)

- AI developments can be clinically applicable to enhance patient care by providing accurate and efficient decision support

1.Introduction

- The majority of quantitative imaging studies have focused on the development of imaging biomarkers for a single timepoint

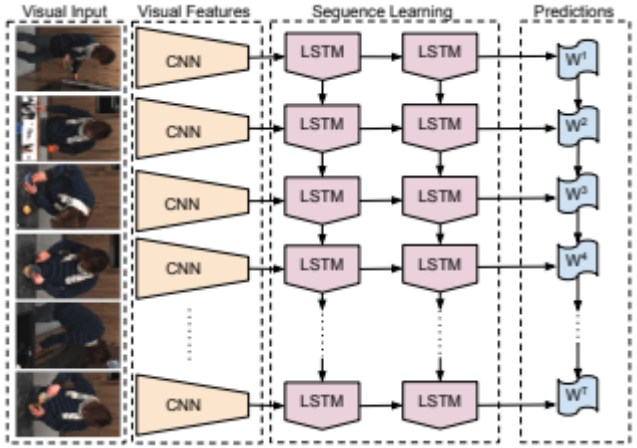
- However, the tumor is a dynamic biological system with vascular and stem cell contributions, thus the phenotype may not be completely captured at a single timepoint.

- It may be beneficial to incorporate post-treatment CT scans from routine clinical follow-up as a means to tracking changes in phenotypic characteristics after radiation therapy.

1.Introduction

- SOTA model in video classification and natural language processing have utilized recurrent neural networks (RNN) to incorporate longitudinal data

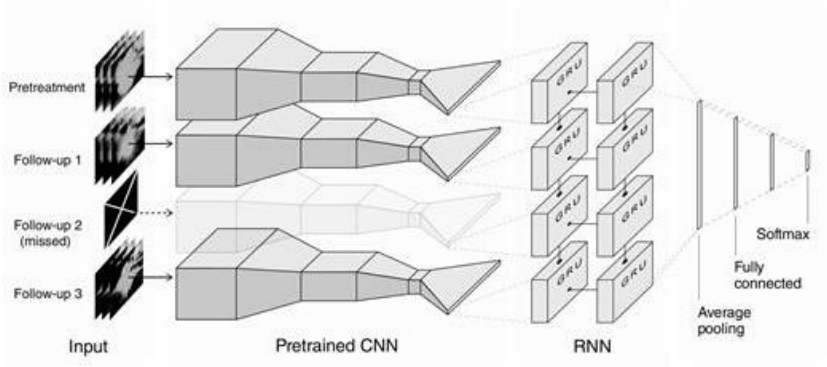
Long-term Recurrent Convolutional Networks (2015)



- However, only a few studies have applied these advanced computational approaches in radiology

1.Introduction

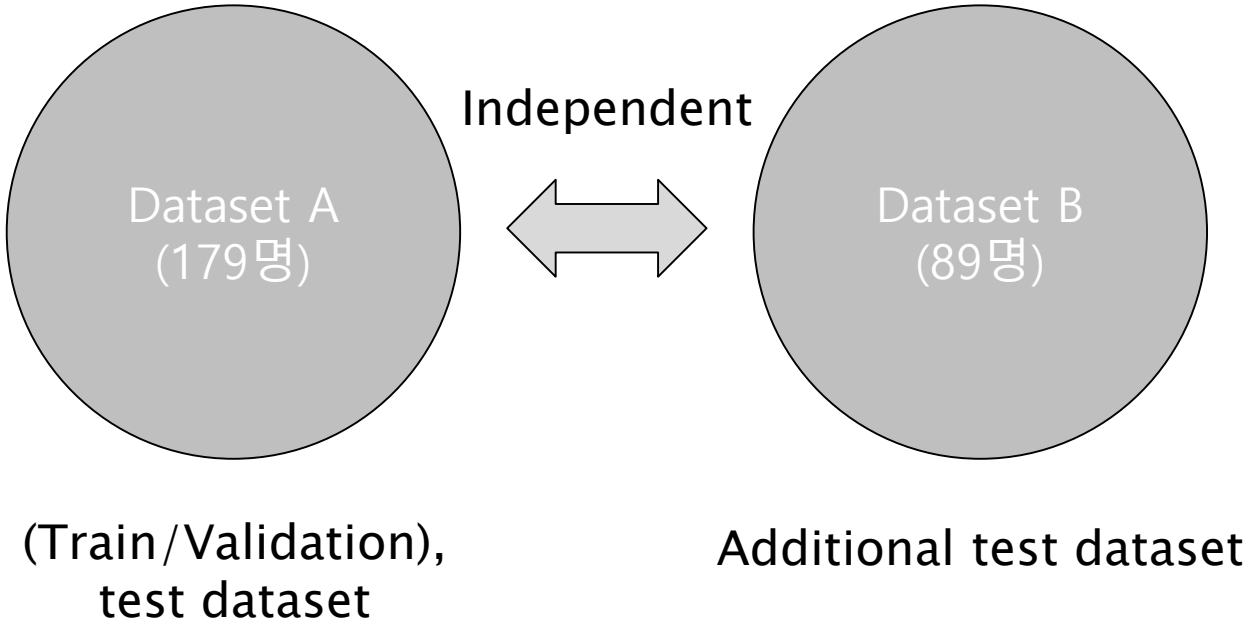
- In this study, we use AI in the form of deep learning, specifically CNNs and RNNs, to predict survival and other clinical endpoints of patients with NSCLC by incorporating pretreatment and follow up CT images.



- This work has implications for the use of AI-based imaging biomarkers in the clinic, as they can be applied noninvasively, repeatedly, at low cost, and requiring minimal human input.

2.Method

- Patient cohorts



2.Method

▶ Patient cohorts – Dataset A

Dataset A	
환자군	2003년~2014년 Brigham and Women's/Dana-Farber Cancer Center에서 치료 받은 179명의 NSCLC 환자
치료방법	Radiation, chemotherapy
영상 갯수	581장(CT-PET : 125장, diagnostic CT : 456장)
Follow up 기간	1, 3, 6개월
특이사항	1. 최소 1회 이상의 follow-up CT를 촬영 2. CT-PET 영상은 요오드 조영제 없이 촬영되었으며 흉부 CT는 환자 별로 가이드라인에 따라 투약됨 3. 치료 전이나 치료 후에 수술을 받은 환자는 dataset에서 제거됨

- Dataset A was randomly split 2:1 into training/tuning (n = 107) and test (n = 72).
- Overall survival was assessed along with three other clinical endpoints for the definitive radiation therapy cohort: distant metastases, locoregional recurrence, and progression

2.Method

▶ Patient cohorts – Dataset B

Dataset B	
환자군	2001년~2003년 Brigham and Women's/Dana-Farber Cancer Center에서 치료받은 89명의 NSCLC 환자
치료방법	Radiation, chemotherapy, surgical (trimodality)
영상 갯수	178장(diagnostic CT : 178장)
Follow up	방사선 치료 전, 방사선 치료 후
특이사항	1. 전이가 발생했거나 chemotherapy와 surgical 사이에 120일 텀이 있는 환자 제거 2. 생존 데이터가 없는 환자 제거

- To use model generalizability and additional test dataset

2.Method

- Patient cohorts – Both

- No histologic exclusions were applied

- For localization of the tumors, only single-click seed points were needed without volumetric segmentations, demonstrating the ease of incorporating a large number of scans at several timepoints into deep learning analyses.

2.Method

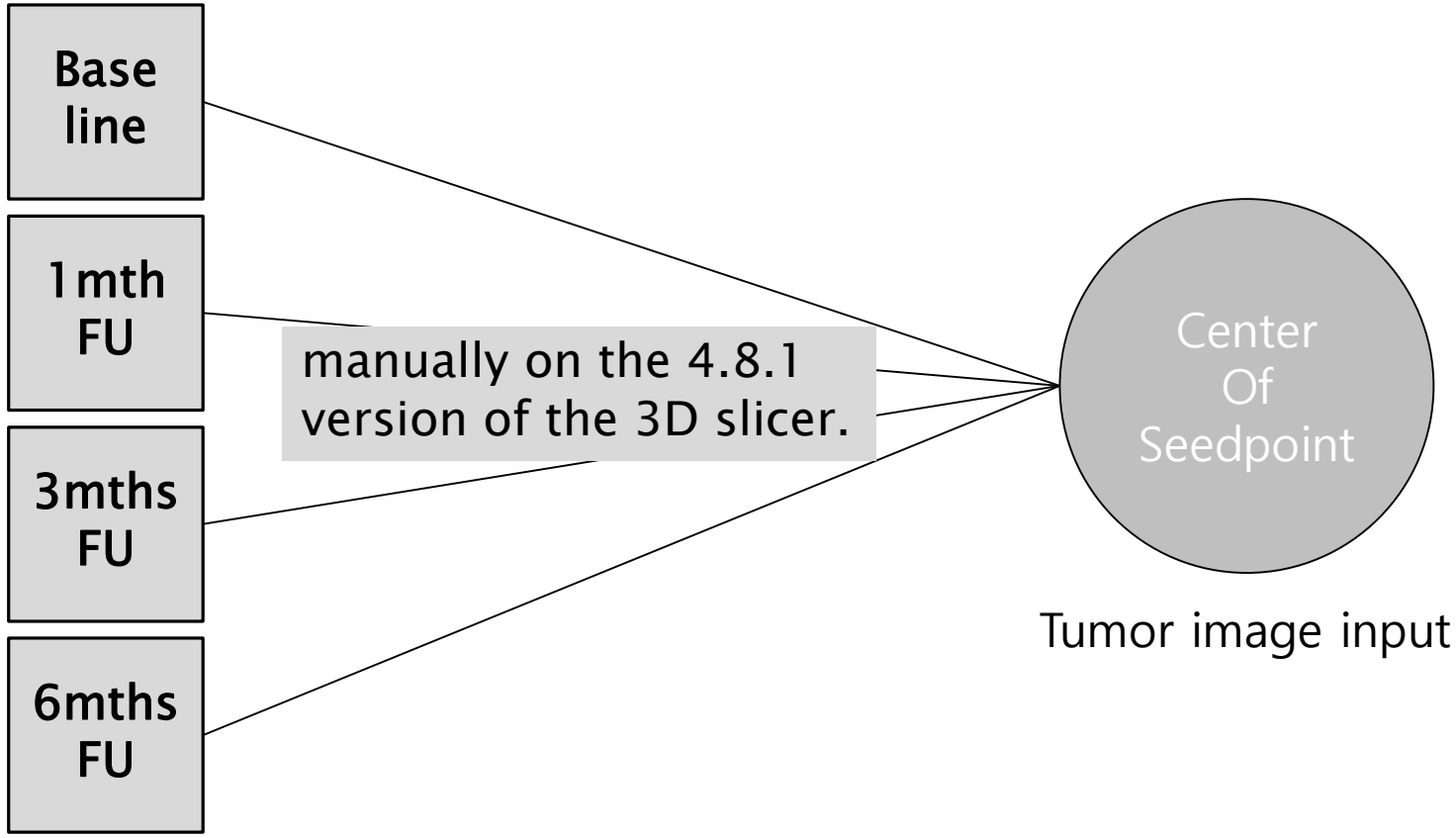
▶ CT acquisition and image preprocessing

- CTs were acquired according to standardized scanning protocols at our institution, using a GE “Lightspeed” CT scanner

- The follow-up scans consisted of different axial spacing and a portion of the images are from PET–CT acquisitions

2.Method

- CT acquisition and image preprocessing

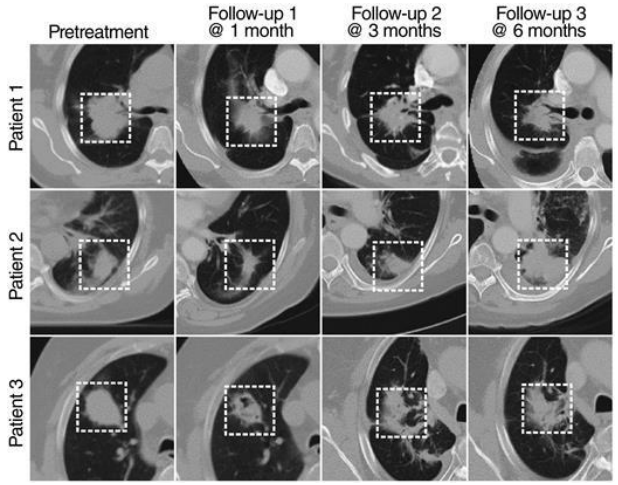


2.Method

- CT acquisition and image preprocessing

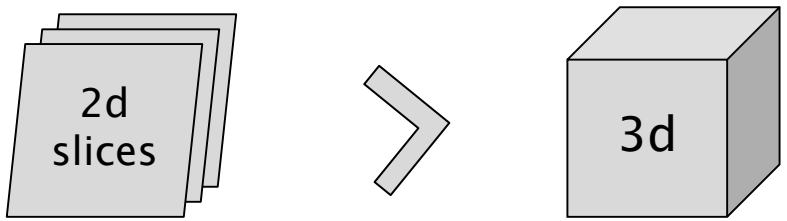
- To have a stable input for the proposed architecture, it was necessary to interpolate the imaging data to homogeneous resolution.

- Three axial slices of 50 × 50 mm^2 centered on the selected seed point were used as inputs to the model



2.Method

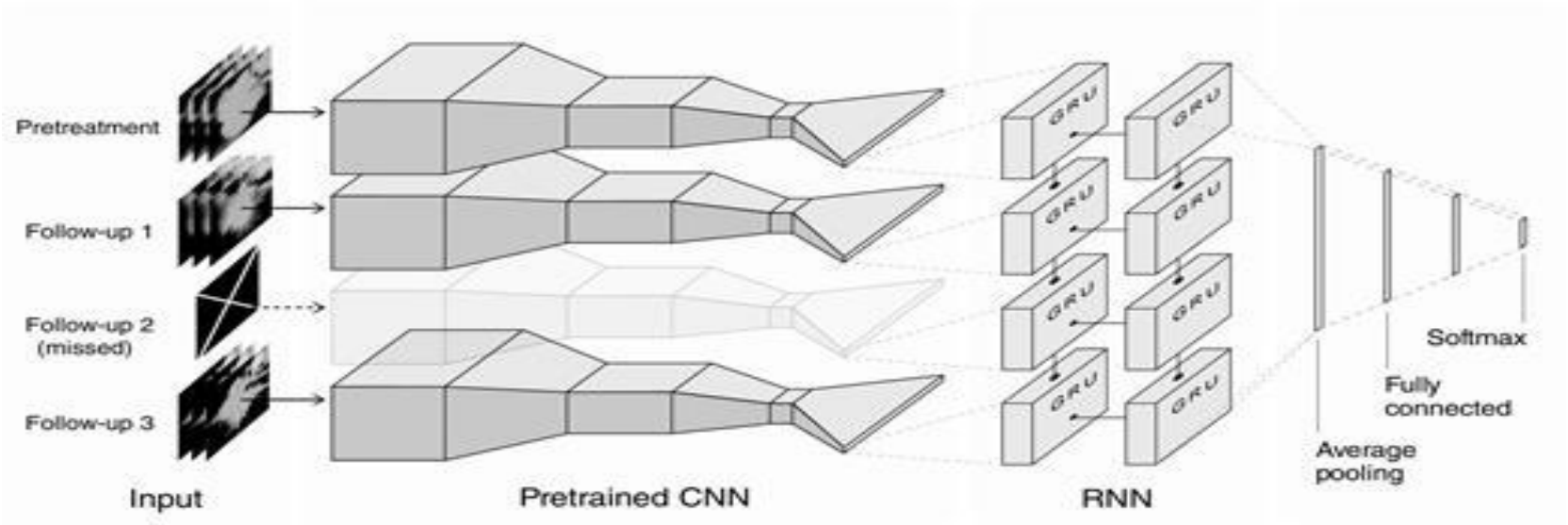
- CT acquisition and image preprocessing
  - Using three 2D slices keeps the number of features lower than a full 3D approach and reduces GPU memory usage and training time, as well as limits the overfitting.



Low feature
Reduce gpu usage & train time
Limit overfitting

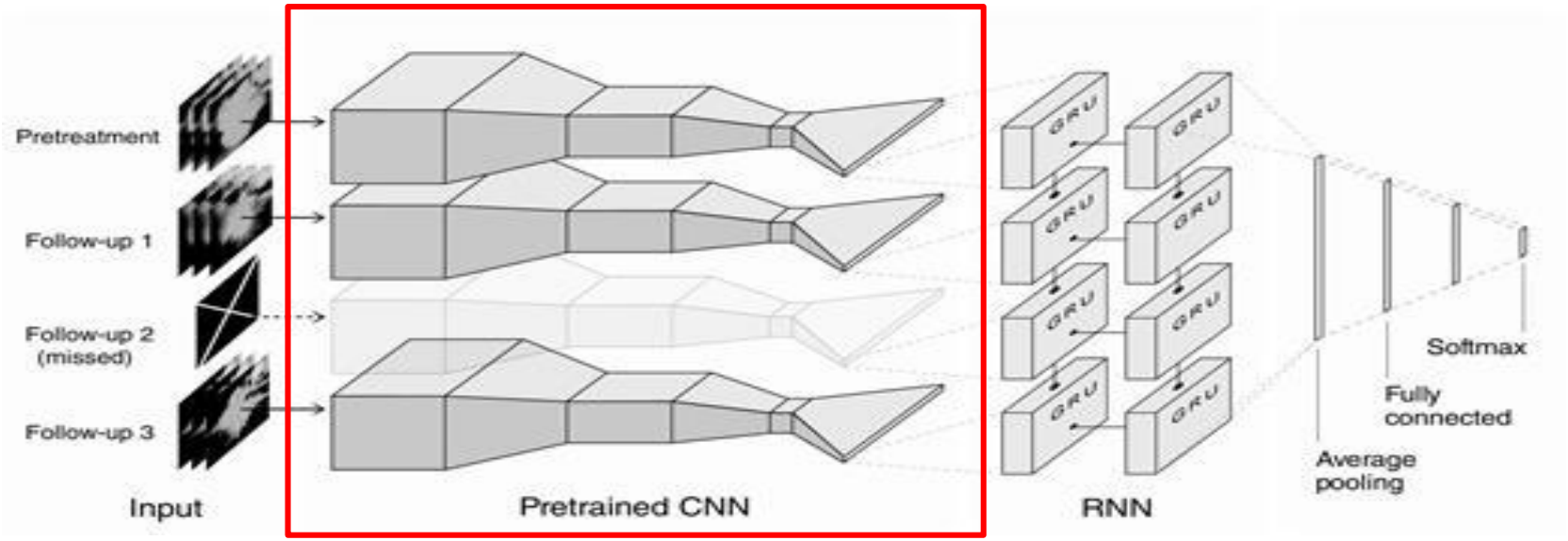
2.Method

- Neural network structure



2.Method

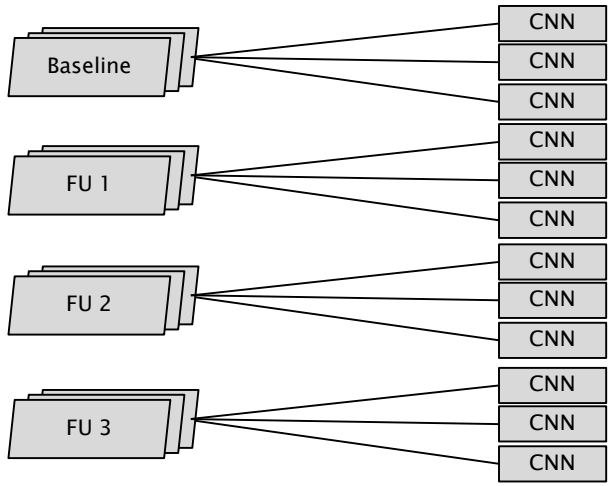
- Neural network structure

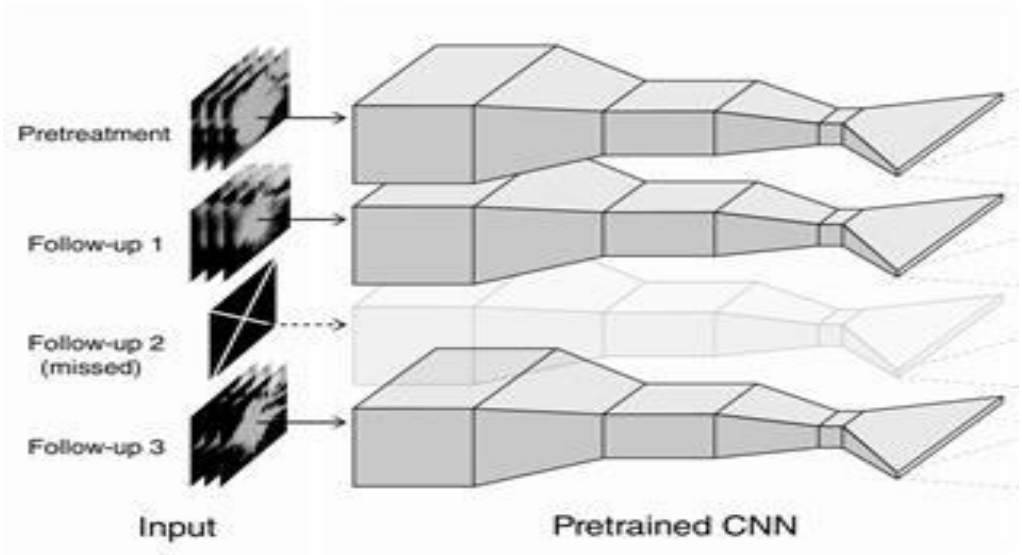


- The proposed network structure has a base ResNet CNN trained on the ImageNet database containing over 14 million natural images.

2.Method

- Neural network structure

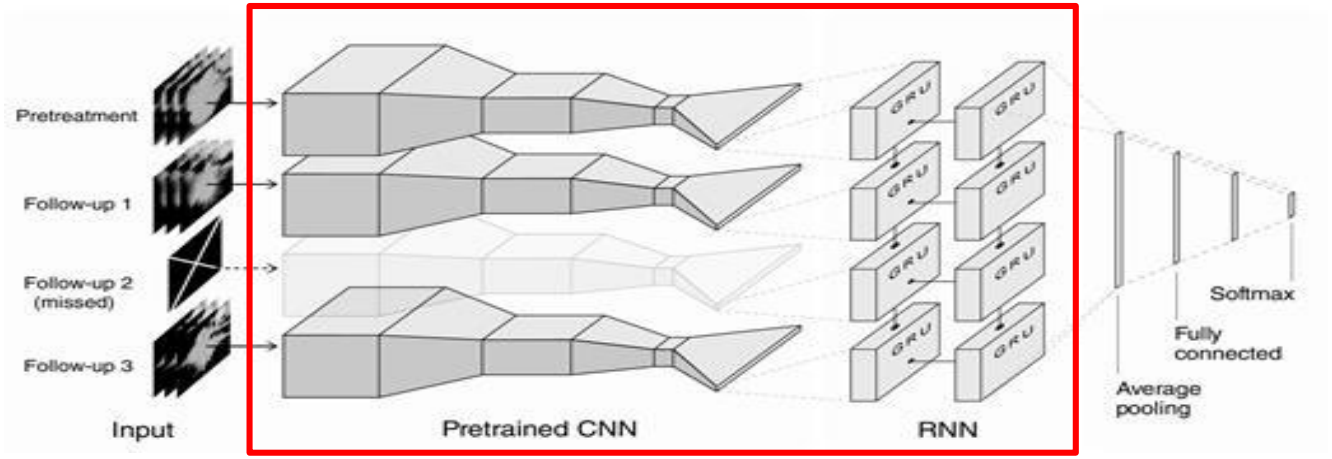




- One CNN was defined for each timepoint input, such that an input with scans at three timepoints would involve input into three CNNs.

2.Method

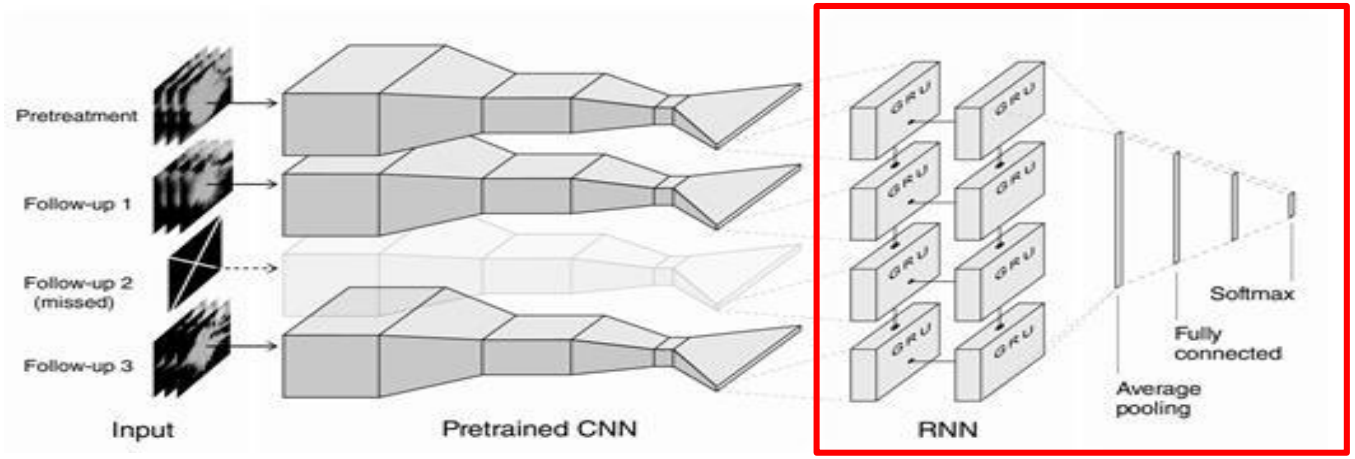
- Neural network structure



- The output of the pretrained network model was then input into recurrent layers with gated recurrent units (GRU), which takes the time domain into account.
- RNN algorithm has the ability to capture multiple time steps and learn from missing samples at specific time steps.

2.Method

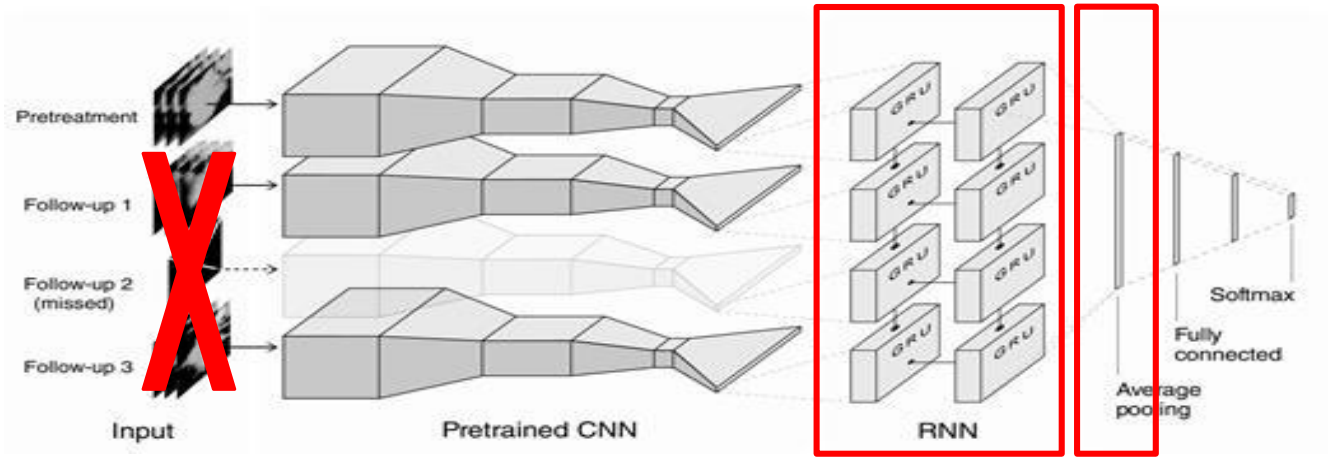
- Neural network structure



- After passing through a GRU with batch normalization, the output goes through average pooling and a fully connected layer.
- To prevent overfitting, dropout is applied after each fully connected layer, and ultimately, the output passes through a softmax layer to produce binary output.

2.Method

- Neural network structure - test without input

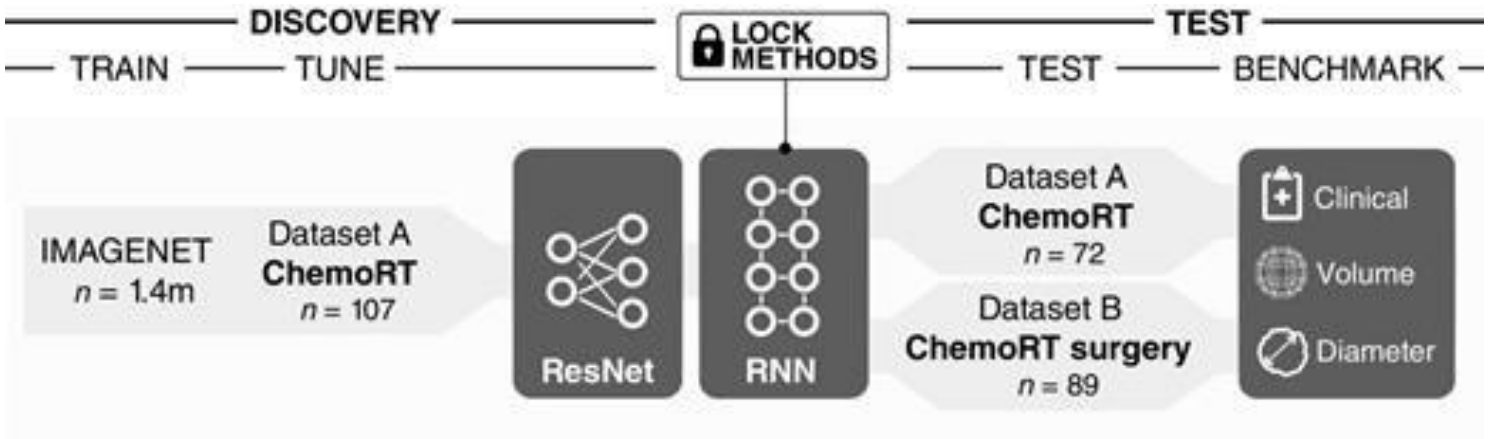


Fully connected layer로 대체

- To test the model without follow-up data input, when only baseline images are provided as input, the RNN and average pooling layers are replaced with fully connected layers.

2.Method

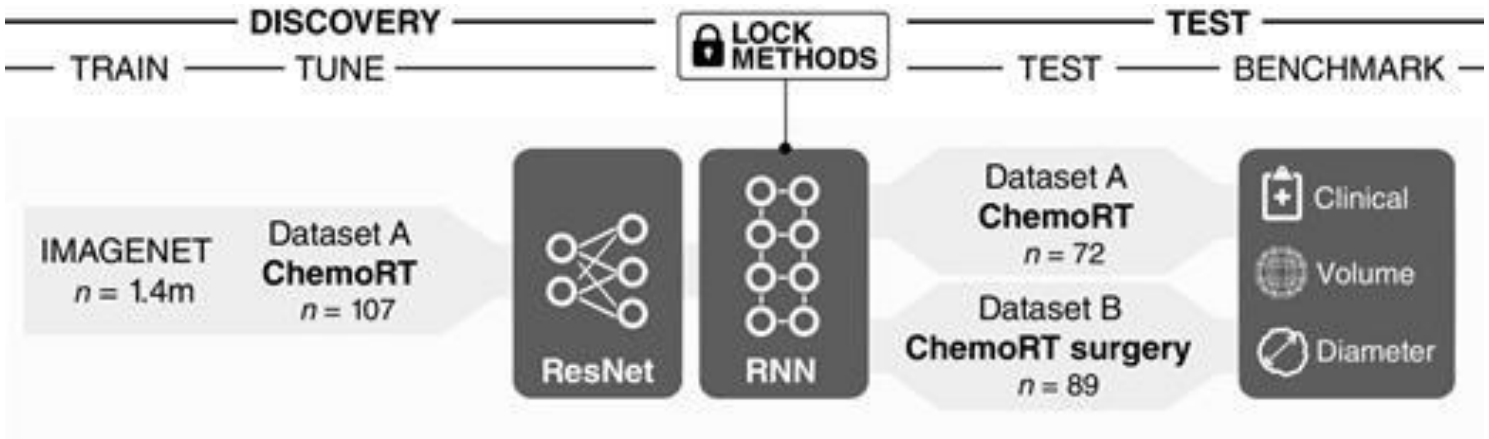
▶ Transfer Learning(ResNet CNN)



- The ResNet CNN was initialized with weights pre-trained on a dataset of 14 million 2D color images (ImageNet). Subsequent weights added after the CNN were randomly initialized for transfer learning purposes.

2.Method

- Transfer Learning(ResNet CNN)

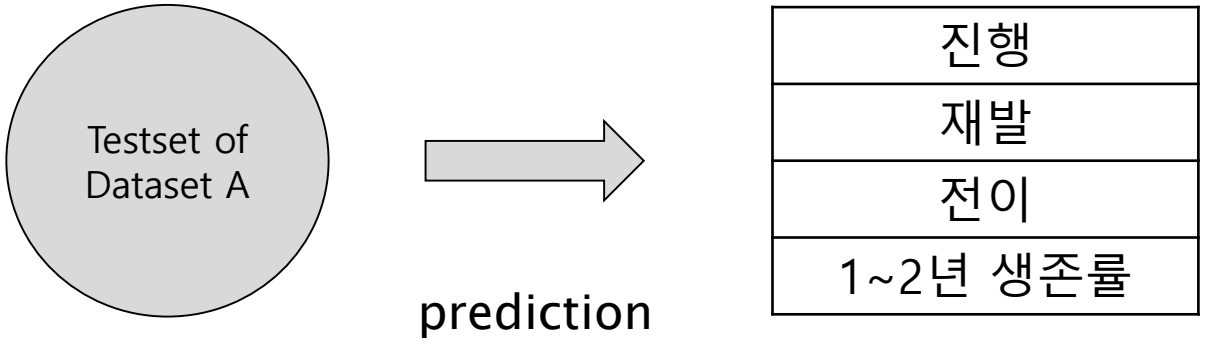


- During training, Monte Carlo cross-validation was employed, and the dataset of 107 patients was split into 10 subsets. Training was conducted for 300 epochs.

2.Method

▶ Statistical analysis

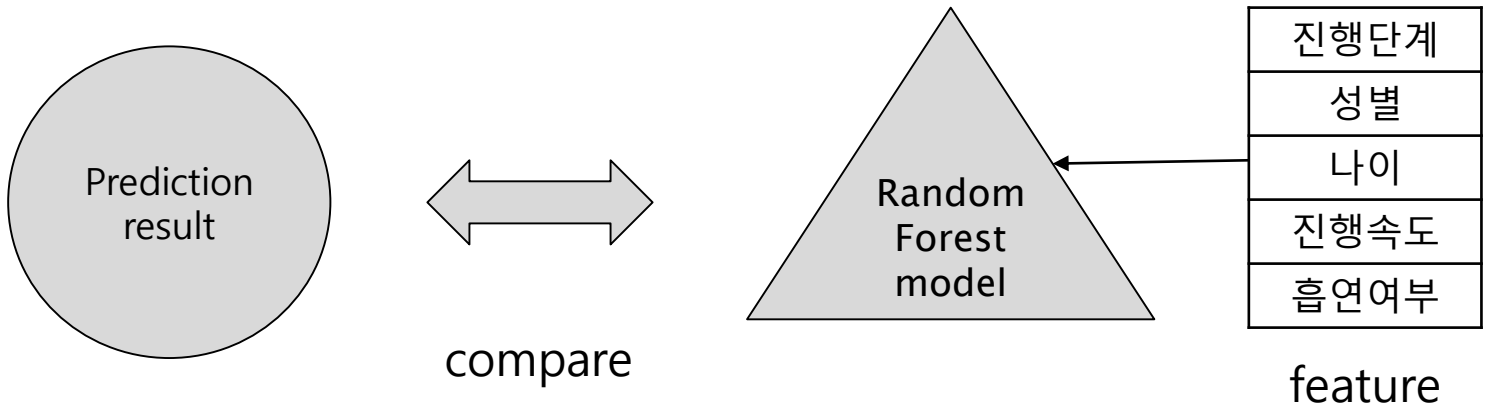
- All predictions were evaluated on the independent test set of dataset A for survival and for prognostic factors after definitive radiation therapy.



2.Method

▶ Statistical analysis

- The analyses were compared with a random forest clinical model with features of stage, gender, age, tumor grade, performance, smoking status, and clinical tumor size



2.Method

▶ Statistical analysis

- Statistical differences between the positive and negative survival groups in dataset A were evaluated using the area under the receiver operator characteristic curve (AUC) and the Wilcoxon rank sum test.

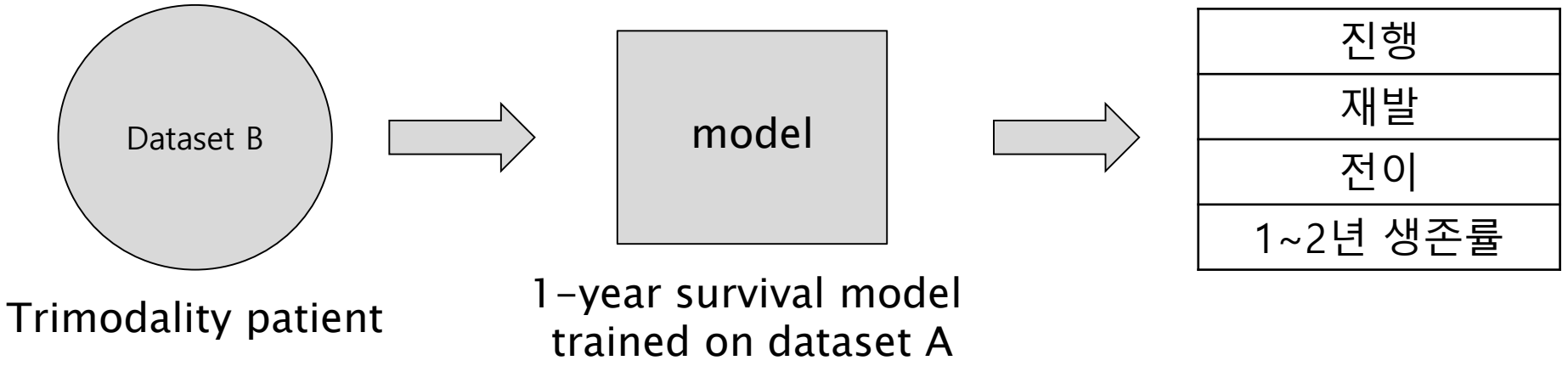
- Prognostic and survival estimates were calculated using the Kaplan–Meier method between low and high mortality risk groups, stratified at the median prediction probability of the training set and controlled using a log-rank test.

- Hazard ratios were calculated through the Cox proportional-hazards model.

2.Method

▶ Statistical analysis

- An additional test was conducted using dataset B, and survival rate predictions were made using the 1-year survival model trained on dataset A.



3.Results

- Clinical characteristics

Patient characteristics		Train	Test
Age		63 (35 – 85)	67 (32 – 93)
Gender	Female	48 (0.45)	37 (0.51)
	Male	58 (0.55)	36 (0.49)
Ethnicity	African-American	7 (0.07)	4 (0.05)
	Asian	3 (0.03)	---
	Caucasian	94 (0.89)	67 (0.92)
	Other	2 (0.02)	2 (0.02)
Smoking	Never	9 (0.08)	9 (0.13)
	Current	41 (0.39)	23 (0.32)
	Former	56 (0.53)	40 (0.56)
Pack-years		40 (1.2 – 160)	40 (0.05 – 165)
Performance status	0	39 (0.37)	25 (0.34)
	1	56 (0.53)	43 (0.59)
	2	10 (0.09)	5 (0.07)
	3	1 (0.01)	---
Tumor characteristics			
Overall stage	IIIA	46 (0.43)	35 (0.49)
	IIIB	60 (0.57)	37 (0.51)
Histology	Adenocarcinoma	58	42
	Squamous cell carcinoma	29	20
	Undifferentiated NSCLC	16	8
	Other (Large Cell)	3	3
	Time to event (months)	8.4 (1.9 – 90.5)	9.4 (2.2 – 153.4)

Treatment characteristics		Train	Test
Prescribed radiation dose (Gy)		66 (45 – 70)	66 (46 – 70)
Radiation dose delivered (Gy)		66 (33 – 70)	66 (6 – 70)
Number of radiation fract ions		33 (15 – 37)	33 (27 – 38)
Delivered biologically effe ctive dose (Gy)		79.2 (48 – 84)	79.2 (64.8 – 84)
Clinical outcomes			
Distant metastasis (DM)	No/ Yes	37/69	16/57
	Time to event (months)	8.7 (1.9 – 90.5)	11.1 (2.2 – 154.5)
Locoregional recurrence (LRR)	No/ Yes	54/52	40/33
	Time to event (months)	15.4 (1.9 – 121.5)	12.3 (2.5 – 153.4)
Progression	No/ Yes	26/80	16/57
	Time to event (months)	8.4 (1.9 – 90.5)	9.4 (2.2 – 153.4)
Survival	No/ Yes	30/76	19/54
	Time to event (months)	15.1 (2.8 – 98.9)	15.0 (4.1 – 154.8)

- There was no significant difference between the patient parameters in the training and test sets of dataset A

3.Results

▶ Deep learning-based prognostic biomarker development and evaluation

- The model that utilized only baseline as input for predicting the overall 2-year survival rate exhibited lower performance compared to the model that incorporated follow-up data

- 2년 생존율 예측 시

  baseline만 사용 (AUC = 0.58; *P* = 0.3; Wilcoxon test)

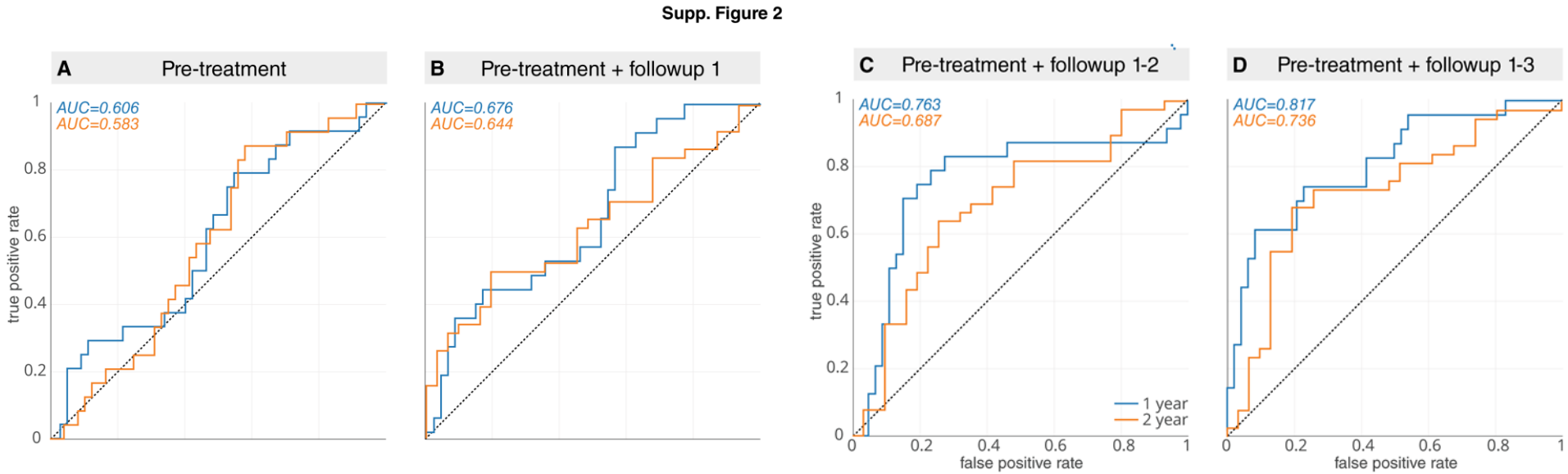
  1month fu 사용 (AUC = 0.64, *P* = 0.04)

  1,3 month fu 사용 (AUC = 0.69, *P* = 0.007)

  1,3,6 month fu 사용 (AUC = 0.74, *P* = 0.001)

3.Results

- Deep learning-based prognostic biomarker development and evaluation

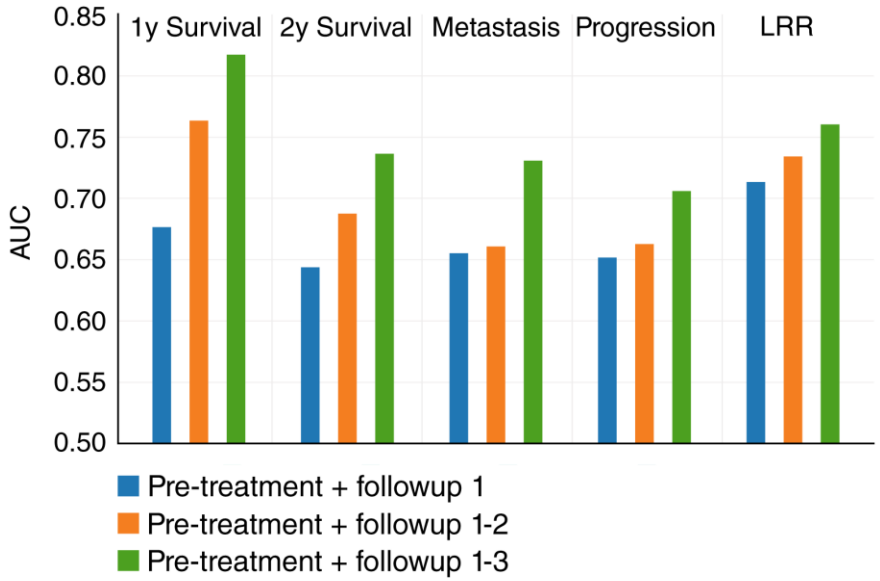


Blue : 1년 생존율
Orange : 2년 생존율

3.Results

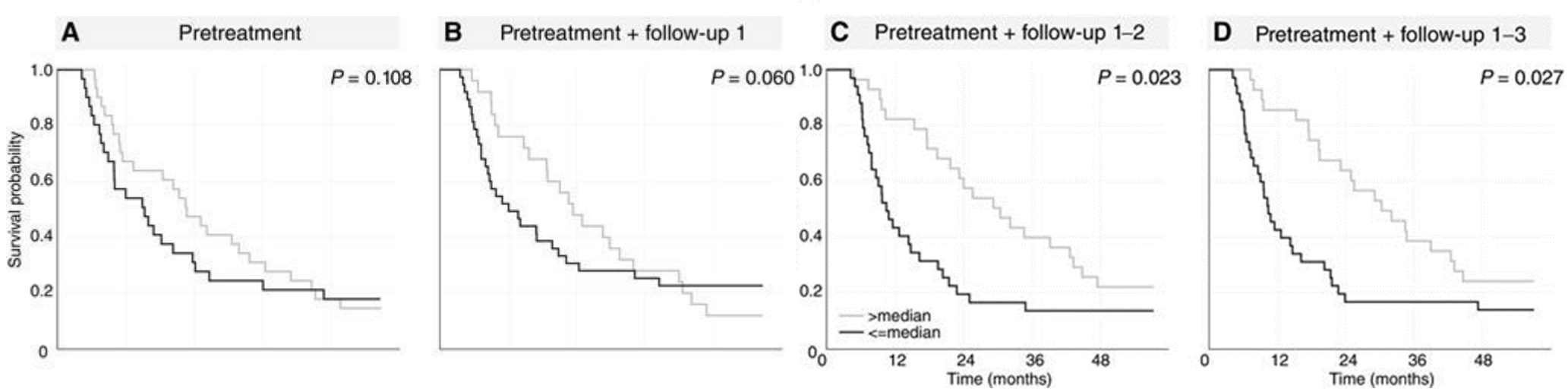
- Deep learning-based prognostic biomarker development and evaluation
  - Similar trends were observed for other clinical variables such as 1-year survival, survival, metastasis, progression, and recurrence.

Supp. Figure 3

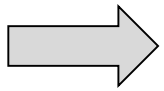


3.Results

- Deep learning-based prognostic biomarker development and evaluation
  - Further survival analyses were performed with Kaplan–Meier estimates for low and high mortality risk groups based on median stratification of patient prediction scores



Follow up 1-2 (P = 0.023, log-rank test)
Follow up 1-3 (P = 0.027, log-rank test)



significant differences

3.Results

- Deep learning-based prognostic biomarker development and evaluation

- Comparable results were found for the following predictions with their respective hazard ratios

  1. 1-year overall survival (6.16; 95% CI, 2.17–17.44]; *P* = 0.0004),
  2. distant metastasis free (3.99; 95% CI, 1.31–12.13; *P* = 0.01),
  3. progression free (3.20; 95% CI, 1.16–8.87; *P* = 0.02),
  4. no locoregional recurrence (2.74; 95% CI, 1.18–6.34; *P* = 0.02),

- Each exhibited significant differences at all three follow-up timepoints scans.

3.Results

▶ Predicting pathologic response

- purpose : To evaluate the relationship between delta imaging analysis and pathologic response

  Delta image: the difference or change in an image compared to a reference image or baseline.

- Method : 1. For survival prediction evaluation, the model was tested on dataset B

  2. To match the number of input timepoints, use 1 year survival model with the baseline and follow up 1month was used

- Result : This model significantly predicted metastasis, recurrence, and progression.

  But overall survival there were a low number of events (30 of 89), the model was trending towards making a prediction for 3-year overall survival in dataset B.

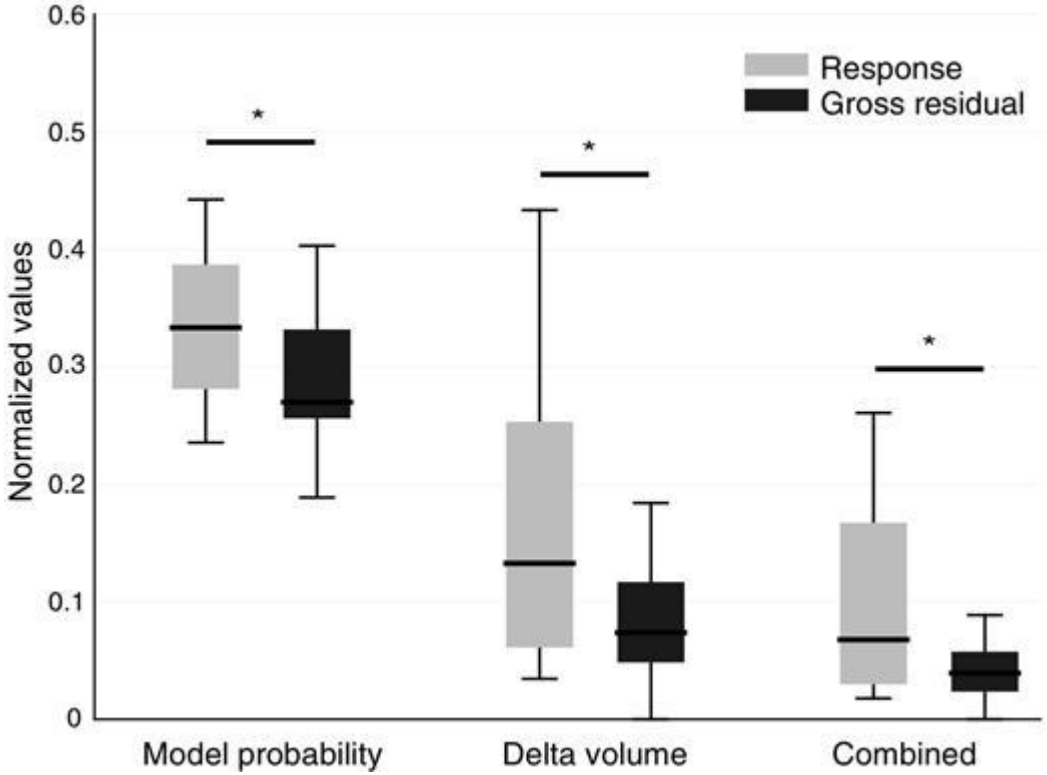
3.Results

- Predicting pathologic response

		Dataset B Test (n = 89)	
		AUC	Rank-sums
Overall survival	1 year	0.601	0.182
	2 year	0.625	0.066
Distant Metastasis Free	1 year	0.635	0.031
	2 year	0.657	0.014
Progression Free	1 year	0.642	0.022
	2 year	0.672	0.008
Locoregional Recurrence Free	1 year	0.656	0.015
	2 year	0.643	0.023

3.Results

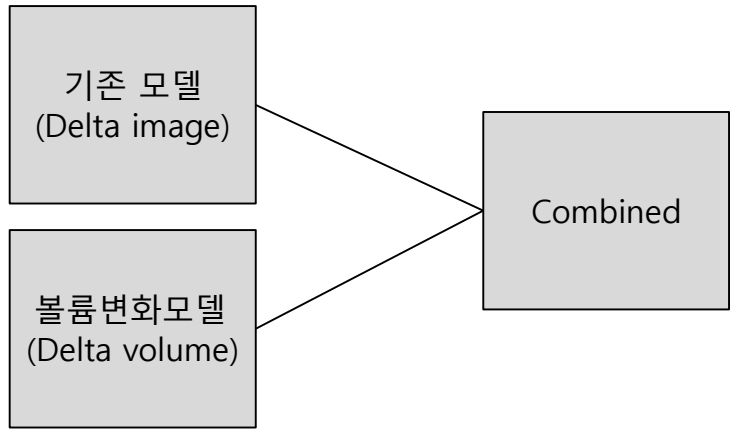
- Predicting pathologic response
  - Responders vs Gross residual disease



Model probabilities : AUC of 0.65; n = 89; P = 0.016; Wilcoxon test

Change in volume : AUC of 0.65; n = 89; P = 0.017; Wilcoxon test

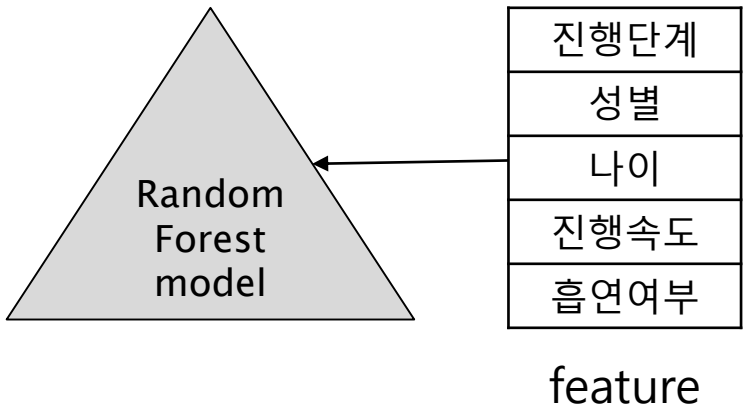
Combined : AUC of 0.67; n = 89; P = 0.006; Wilcoxon test



3.Results

▶ Predicting pathologic response

- A clinical model, involving parameters of stage, gender, age, tumor grade, performance, smoking status, and clinical tumor size, did not yield a statistically significant prediction for pathologic response (P = 0.42; Wilcoxon test).



4.Conclusion

1. Baseline and follow-up data demonstrated the impact of deep learning on post-definitive radiation therapy tumor tracking. By utilizing CNN and RNN networks and incorporating timepoints, the performance of survival and prognosis prediction was enhanced.

2. When compared to models utilizing clinical factors, this enhancement wasn't statistically significant.

3. The survival prediction model could predict pathological response in patients undergoing trimodality treatment after radiation therapy.

4.Conclusion

4. This model employed a single seed point at the tumor's center as input, obviating the need for volume segmentation.

   Our results showed similar performance to using manually masked volumes.

5. Through non-invasive tracking of tumor phenotypes, prediction of survival, prognosis, and pathological response is achievable, holding potential clinical implications for adaptive and personalized treatments.