Machine Learning Algorithms in Histology and Radiology for Cancer Drug Discovery and Development

Partha Paul, Yuki Shimahara, Diasuk Takamiya

Disclosures

I am an employee of LPIXEL and own stock options of LPIXEL

Lung Cancer: significant unmet medical need

- About 14% of all new cancers are lung cancers, but 24% of all cancer deaths.
- About 234,030 new cases of lung cancer (121,680 in men and 112,350 in women) About 154,050 deaths from lung cancer (83,550 in men and 70,500 in women)
- One in 16 people in the US will be diagnosed with lung cancer in their lifetime.
- 10% to 15% of new lung cancer cases are among never-smokers.
- Only 19% of all people diagnosed with lung cancer survive 5 years or more, BUT if it's caught before it spreads, the chance for 5-year survival improves dramatically.
- About 80% to 85% of lung cancers are NSCLC. The main subtypes of NSCLC are adenocarcinoma, squamous cell carcinoma, and large cell carcinoma.

There is a significant need for improved diagnosis and drug therapies

Radiological detection of lung nodules

- X-ray
  - Chest x-ray usually is the first test that is done to look at abnormalities in the lung.
- CT
  - The American Cancer Society (ACS) has a lung cancer screening guideline for people with a higher risk of getting lung cancer
  - The ACS recommends yearly lung cancer screening with LDCT scans for people who are 55 to 74 years old in high risk category but, are in fairly good health.
- Lung-Rads score
  - Lung-RADS® is a tool designed to standardize lung cancer screening CT reporting and management recommendations, reduce confusion in lung cancer screening CT interpretations, and facilitate outcome monitoring.
- PET/CT
  - PET is used to distinguish between benign (noncancerous) and malignant (cancerous) masses. PET can also be used for staging.

Challenges with radiology detections

- Interpretation of CR has problems with overlooking.
- The reported error rates for missed lung cancer on CR are 20%-50% (Diagn Interv Radiol 2017;23:118-126; Am J Roentgenol2007; 188:1173-1178)
- In CT studies, reported sensitivities range from 30–97% with false positive counts of 0.6 – 2.1 per patient (G. Rubin, J Thorac Imaging. 2015 March ; 30(2): 130–138.)
- LDCT and subsequent Lung-Rads classification has high FP rates
  - Classified as Grade 4A : Solid nodules ≥ 8 to < 15 mm at baseline OR new 6 to < 8 mm Additional diagnostic testing and/or tissue sampling is recommended. Likelihood of finding cancer: 5-15%
  - Classified as Grade 4 B : solid nodule ≥ 15 mm OR new ≥ 8 mm. Additional diagnostic and tissue sampling is ordered. Likelihood of finding cancer: >15%

Deep learning techniques can potentially overcome performances of conventional reads. (Jpn J Radiol 2019; 37:15-33)

A deep convolutional neural network based software in a multicenter study involving 12 radiologists improved sensitivity by an average of 5.2%, 65.1 to 70.3 (Y. Sim, Radiol Nov12)

Drug therapy challenges in NSCLC

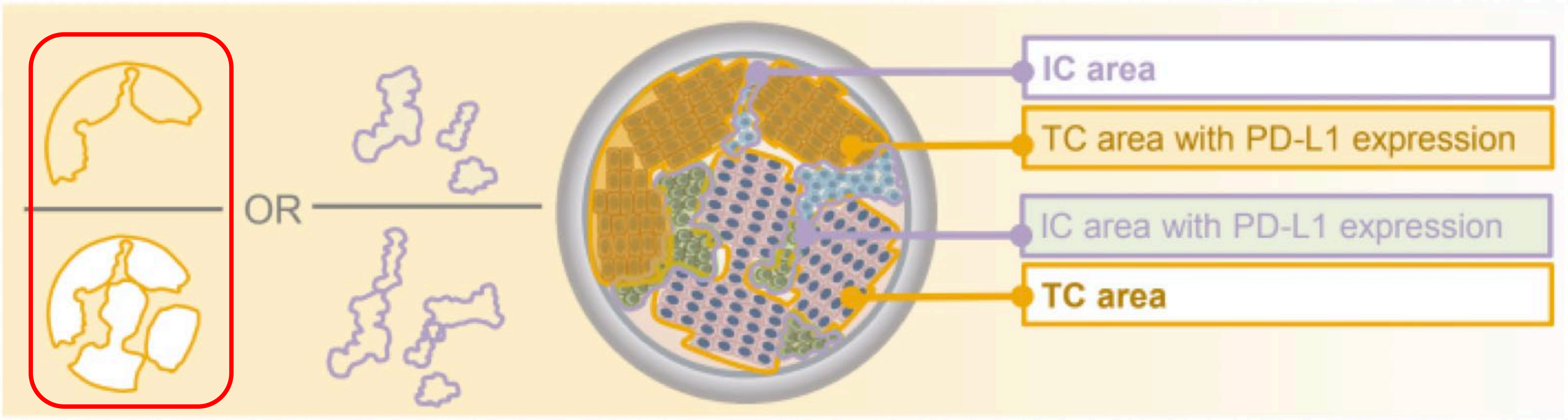
- Significant progress has been made in targeted and immunotherapies to treat NSCLC
- Therapies targeting angiogenesis, Epidermal growth factor receptor (EGFR), Anaplastic lymphoma kinase (ALK), ROS1 gene and NTRK gene fusion are available
- Immune checkpoint inhibitors blocking PD-1/PD-L1 pathways such as nivolumab (Opdivo), pemrolizumab (Keytruda), durvalumab (Imfinzi) and atezolimumab (Tecentriq) are standard treatment option for advanced NSCLC.
- A response rate of 14-23% is observed in unselected population with immunotherapies
- Response rate of 16-48% in patients in high PD-1 expression

PD-1/PD-L1 is an innovative therapy; however, it fails to work in many patients

Deep learning tool for accurate PD-L1 expression

- PD-L1 expression in the tumor microenvironment is one of the efficacy predictors of cancer treatment using immune checkpoint inhibitors.
- Tumor proportion score (the ratio of PD-L1 expressed area within the total tumor area) is often used as a measure to distinguish low and high PD-L1 expression
- However, due to difficulty in quantitative measurement of microscopic images, the proportion score is prone to be evaluated subjectively

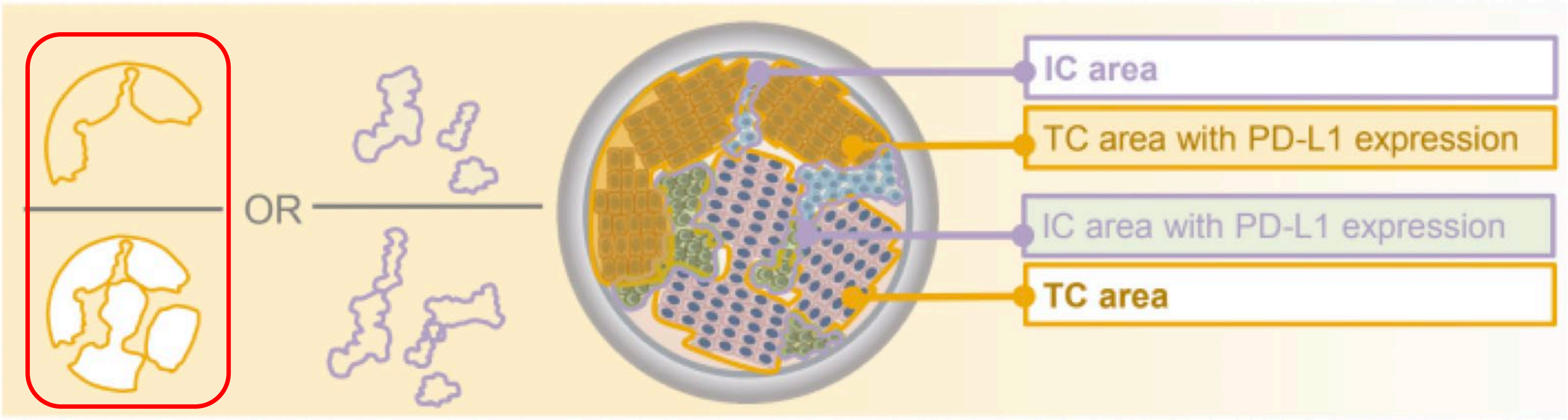
Ilustration of ”tumor proportion score”: Tumor area with PD-L1 expression / Total tumor area



Objective

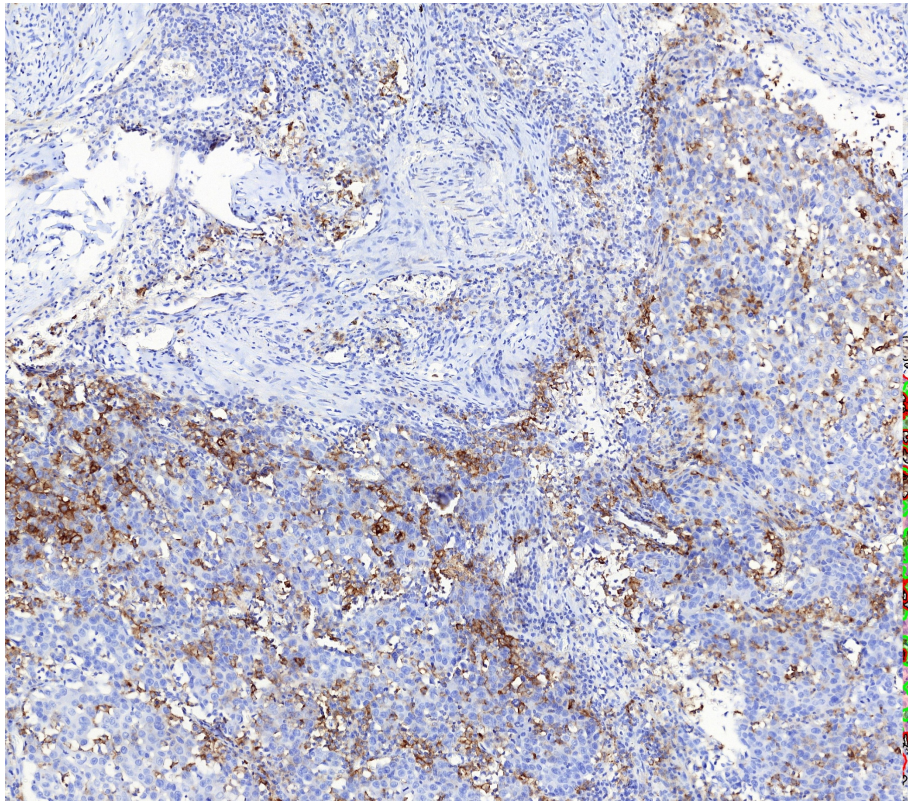
- Our objective is to create an image analysis system to automatically recognize the “total tumor area” and “tumor area with PD-L1 expression” in order to derive a quantitatively measured “tumor proportion score”

Ilustration of ”tumor proportion score”: Tumor area with PD-L1 expression / Total tumor area



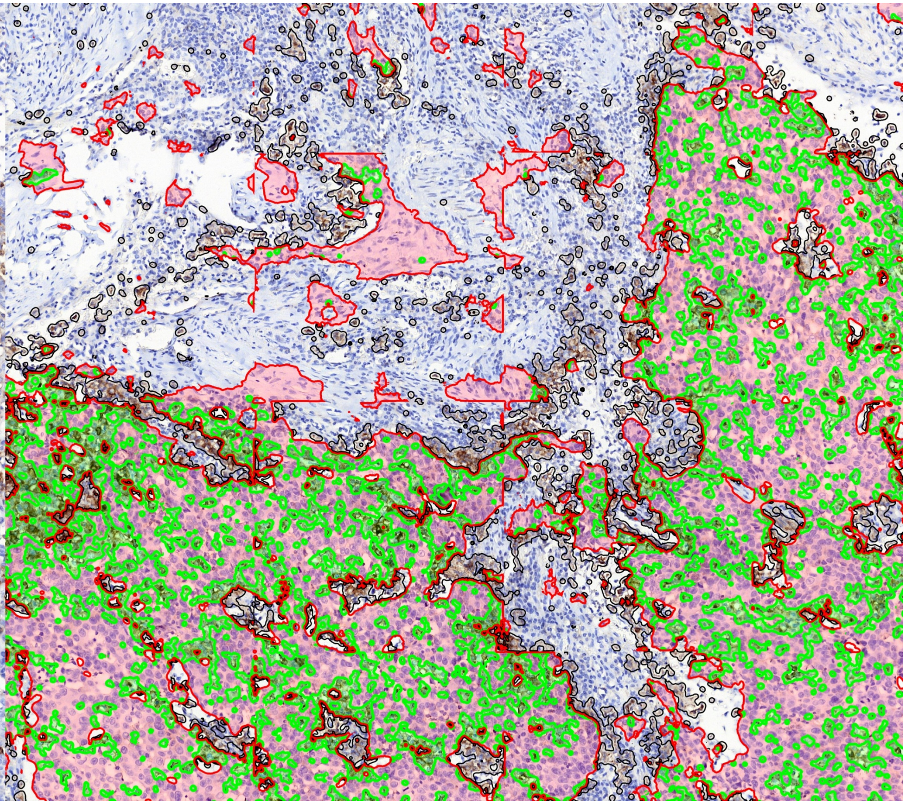
Segmentation result

Original image



~3600px

Segmentation result



Calculated ratio

Total-cancer to total-tissue ratio: 53.41%

PD-L1 cancer(green) to total-cancer(red+green) ratio: 39.21%

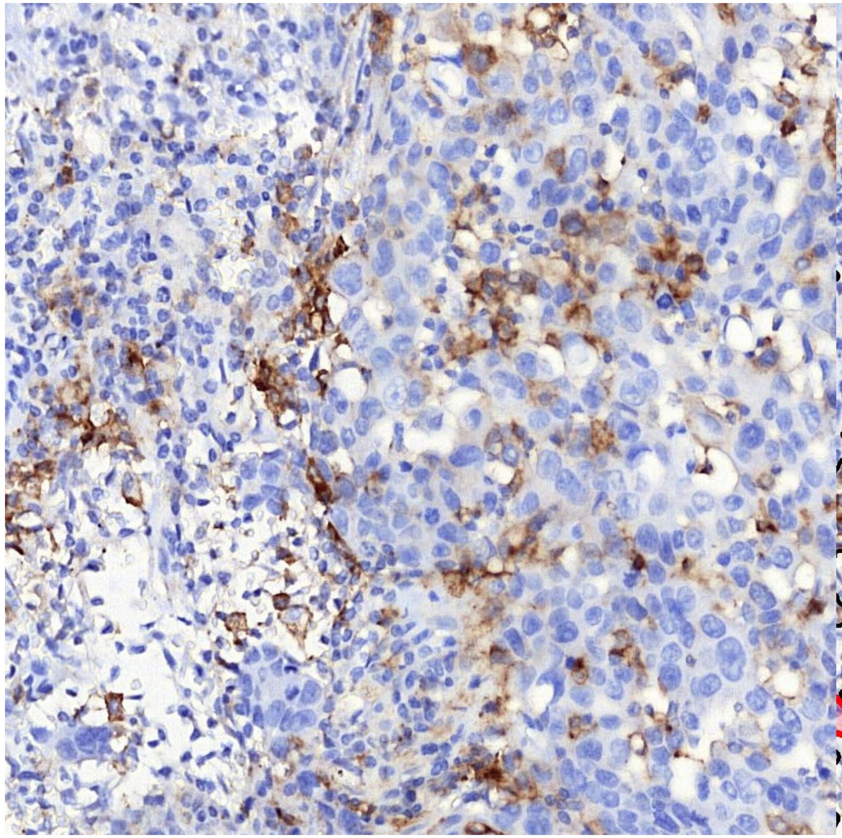
Red: TC area

Green: TC area with PD-L1 expression

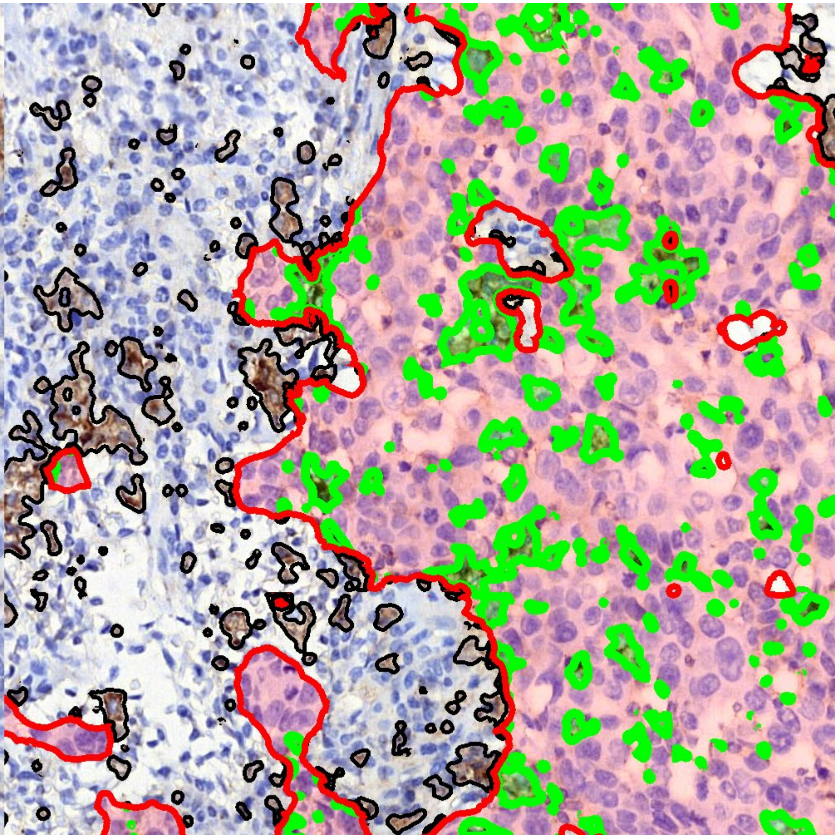
Black outline: non-TC area with PD-L1 expression

Segmentation results (magnified)

Original image



Segmentation result



~1000px

PD- L 1 specifically expressed in cancer cells can be quantified

Summary of PD – L1 analysis

We have demonstrated that application of deep learning to microscopic pathology images has potential in the following tasks

- Distinguishing cancerous regions from other tissue regions
- Segmentation of regions with PD-L1 expression
- Quantitative evaluation of tumor proportion score

Addressing the following limitations in data could further improve performance

- Quantity of training data
  - training has been done with only 34 patches of 1000 x 1000 pixel images
- Quality of training data
  - Annotations by a staff pathologist is needed
- Variation of training data
  - Training patches are extracted from slides from multiple institutions would help generalization ability of the model

Deep Learning Algorithm for Detecting Lung Cancers in Chest Radiographs

LPIXEL

- A training dataset was obtained between January 2006 and June 2017 in our institution.
- A test dataset was obtained between July 2017 and June 2018 in our institution.
- Patients who had surgery and pathologically proven lung cancers.
- CR had one or more malignant nodules or/and masses.
- CR was performed with an upright posterior-anterior view using a digital radiography.
- Chest CT was performed within a month of taking the CR.
- CRs with consolidation, atelectasis, or pleural effusion were excluded

Image segmentation and Algorithm development

- All malignant nodules and masses in CRs were manually segmentated by two radiologists with reference to CT images.
- Manual segmentations were made with ITK-SNAP version 3.6.0.

- The algorithm was developed by segmentation method with encoder-decoder architecture.
- The encoder used Inception-ResNet-v2 model as the backbone algorithm to extract features.
- The decoder was designed to compute upconvolution to detect lung nodule or mass with segmentation.

Statistical Analysis

LPIXEL

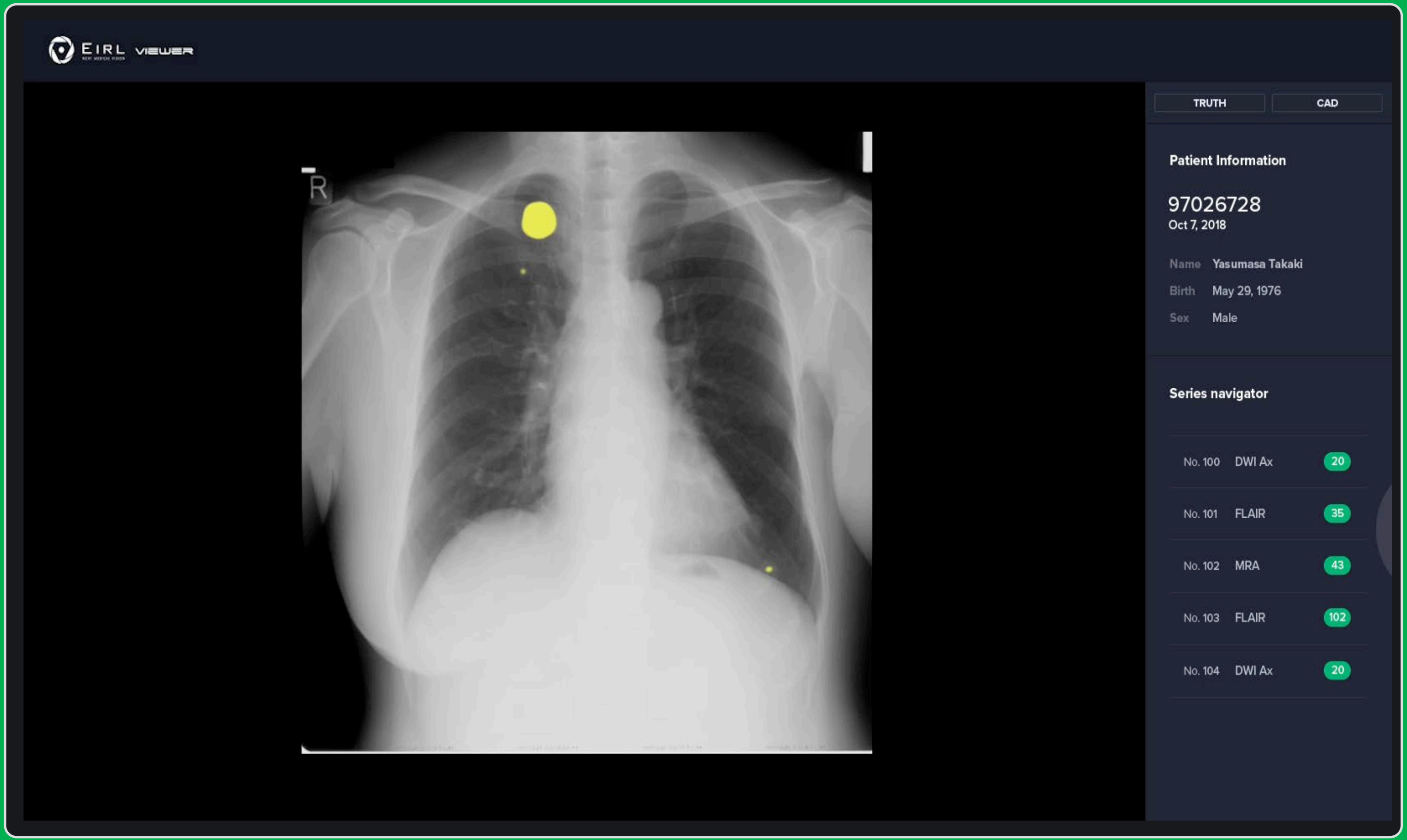
- The two radiologists determined true positive (TP) or false positive (FP) of the lesions which were segmentated by the algorithm.
- The algorithm was evaluated by its sensitivity, positive predictive value (PPV), false positive per image (FPI), and free-response receiver operating characteristic curve (FROC).
- The size, location, and shape of false negative(FN) candidates ware evaluated.

Characteristics Training dataset Test dataset

Characteristics	Training Dataset	Test Dataset
No. of chest radiographs	629	151
Men	408 (65%)	94 (62%)
Women	221 (35%)	57 (38%)
Mean age ±SD (year)	69±8.87	69±8.92
No. of malignant nodules and masses	652	160
Size (cm)		
<=1	5 (0.77%)	6 (3.8%)
1.1 – 1.5	45 (6.9%)	21 (13%)
1.6 – 2.0	68 (10%)	27 (17%)
2.1 – 2.5	87 (13%)	23 (14%)
2.6-3.0	111 (17%)	19 (12%)
3.1- 4.0	133 (20%)	25 (16%)
4.1 – 5.0	73 (11%)	13 (8.1%)
>5.0	130 (20%)	26 (16%)

Lung Nodule Detection in X-ray

Detection of lung nodules from chest X-rays using segmentation-based and object detection algorithms



Results

The sensitivity, PPV, and FPI were 73.0%, 85.3%, and 0.13, respectively.

Size	Number of Lung Cancer	No. of TPs	No. of FNs	Sensitivity (%)
Size (cm)				
<=1	6	0	6	0
1.1 – 1.5	21	8	13	38.1
1.6 – 2.0	27	14	13	51.9
2.1 – 2.5	23	19	4	82.6
2.6-3.0	19	15	4	78.9
3.1- 4.0	25	25	0	100
4.1 – 5.0	13	13	0	100
>5.0	26	22	4	84.6

Conclusion, limitations and future direction

LPIXEL

- In previous studies using deep learning techniques

  Sensitivity and FPI were 69.6-82.0% and 0.02-0.34, respectively (Radiology, 2019; 290:218-228)

  Sensitivity and FPI were 70.3% and 0.18, respectively (Radiology 2019, Nov 12, online)
- We could achieve high performance with few FPs using relatively small training data.
  - Sensitivity is poor for nodules <1.0 cm
  - Lung cancers larger than 5cm widely overlapped with the normal structures.
- Small dataset
- Further testing in other institutions is needed

Acknowledgements

Osaka City Hospital, Department of Diagnostic and Interventional Radiology

Akitoshi Shimazaki

Daiju Ueda

Takashi Honjo

Akira, Yamamoto

Yukio Miki

Department of Thoracic Surgery, Osaka City University Hospital

Noritoshi Nishiyama

Nobuhiro Izumi

Takuma Tsukioka

LPIXEL, Inc.

Takayuki Okamoto,

Yuki Shimahara,

Kazumi Hakamada

Hiroki Kawai

Daisuk Takamiya