Strategy to develop convolutional neural network-based classifier for diagnosis of whole-body FDG PET images

Keisuke Kawauchi[1], Kenji Hirata[2], Seiya Ichikawa[1], Osamu Manabe[2], Kentaro Kobayashi[2], Shiro Watanabe[2], Miki Haseyama[3], Takahiro Ogawa[3], Ren Togo[3], Tohru Shiga[2], Chietsugu Katoh[1]

1. Graduate School of Biomedical Science and Engineering, School of Medicine,
2. Department of Nuclear Medicine, Graduate School of Medicine,
3. Graduate School of Information Science and Technology,
Hokkaido University, Sapporo, Japan

SNMMI June 25, 2018 Philadelphia, PA

Backgrounds

- As the number of PET-CT scanners increases and FDG PET-CT becomes a common study in oncology, the physicians' burden of interpreting images is increased.
- Recently, image analysis using the convolutional neural network (CNN) is rapidly becoming popular in medical imaging.

Backgrounds

- CNN is one of the deep learning techniques and is known to be feasible to image classification by recognizing complex visual patterns in a similar way to human perceptions.
- The demand on AI is rapidly growing in order to the relative decrease of radiologists/nuclear medicine physicians and to prevent human oversight/misdiagnosis.

Objectives

- We aimed to develop a CNN-based diagnosis system that classifies whole-body FDG PET-CT images into 1) normal, 2) abnormal, 3) equivocal.

Patient population

- We investigated sequential 1,053 studies of FDG PET-CT at our institute.
- A nuclear medicine physician classified all the cases into the following 3 categories:

(1) No malignant uptake

N=302

(2) Malignant uptake

N=146

(3) Equivocal

N=605

Methods; Image acquisition and reconstruction

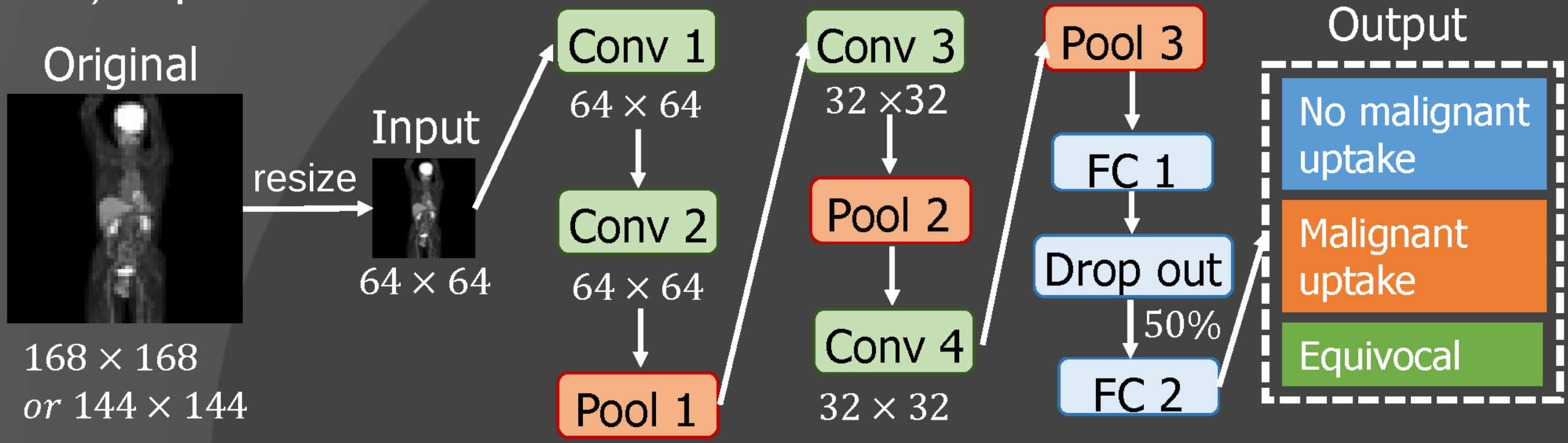
Scanner	Biograph 64 (Siemens, N = 4,794)	GEMINI-TF 64 (Philips, N = 1,641)
Reconstruction method	TrueX (OSEM)	3D-RAMLA (OSEM)
Time of flight	No	Yes
Point spread function	Yes	No
Matrix size	168×168	144×144

Methods; Image acquisition and reconstruction

- Maximum intensity projection (MIP) images (matrix size, 168 × 168 or 144 × 144) were generated on each scanner's workstation at various angles.
- In the current study, CT images were used only for attenuation correction but not for classification.

Methods; Architectures

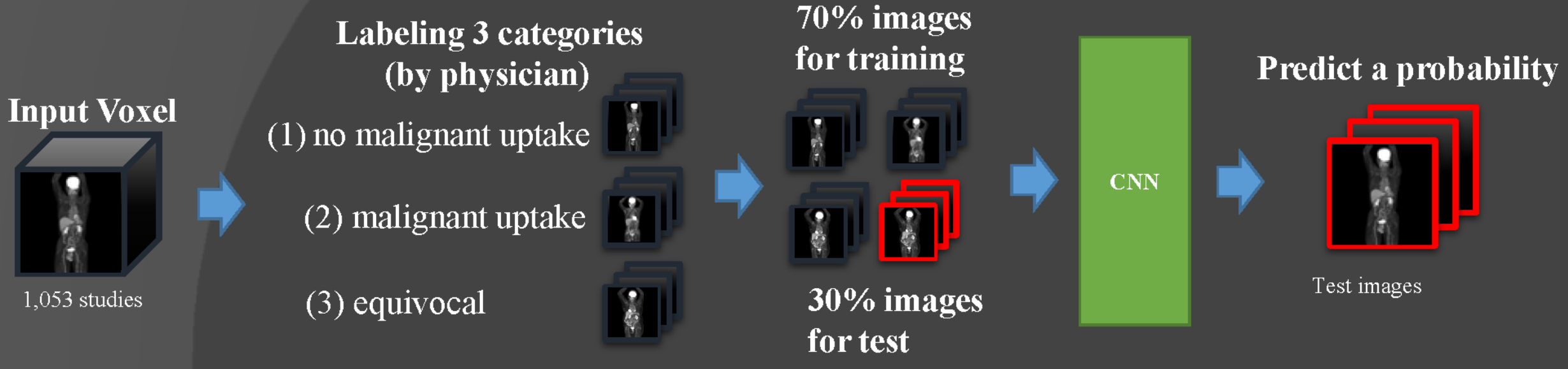
- We constructed a CNN to classify whole-body FDG PET-CT images into 1) No malignant uptake, 2) Malignant uptake, 3) Equivocal.



· Conv = Convolutional Layer · Pool = Max Pooling Layer · FC = Fully Connected Layer

Methods; Model training and test

- We trained the CNN using data from the FDG PET images.
- CNN was trained and validated using the randomly selected 70% images while the data of the remaining 30% images were used for test purpose. The process was repeated 5 times to calculate the accuracy.



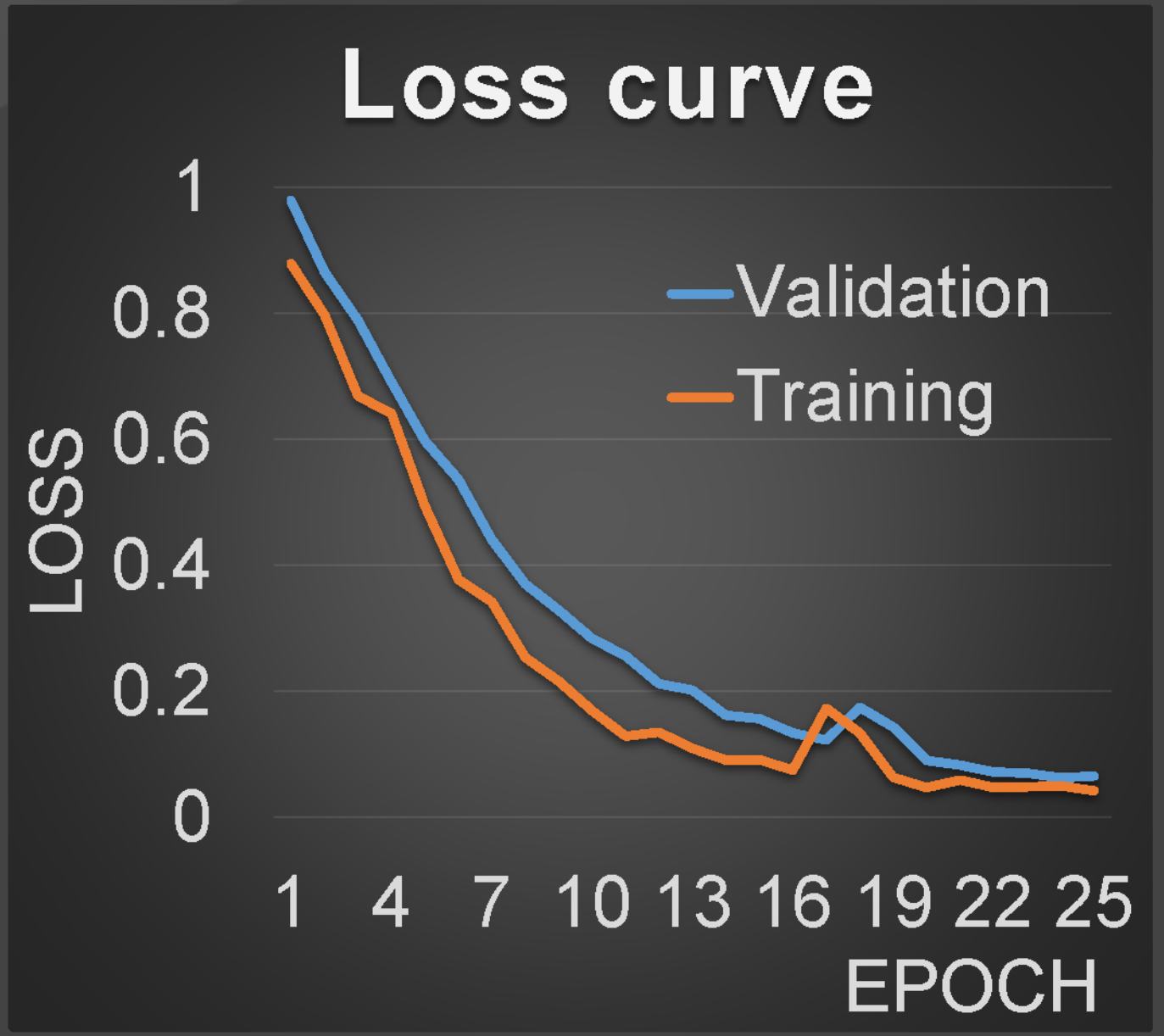
Methods; Hardware and software environments

- These experiments were performed the following environment:

OS	Windows 10 pro 64 bit
CPU	intel Core i7-6700K
GPU	2 × NVIDIA GeForce GTX 1060 6GB
Framework	Keras 2.0.2 and TensorFlow 1.3.0
Language	Python 3.5.2

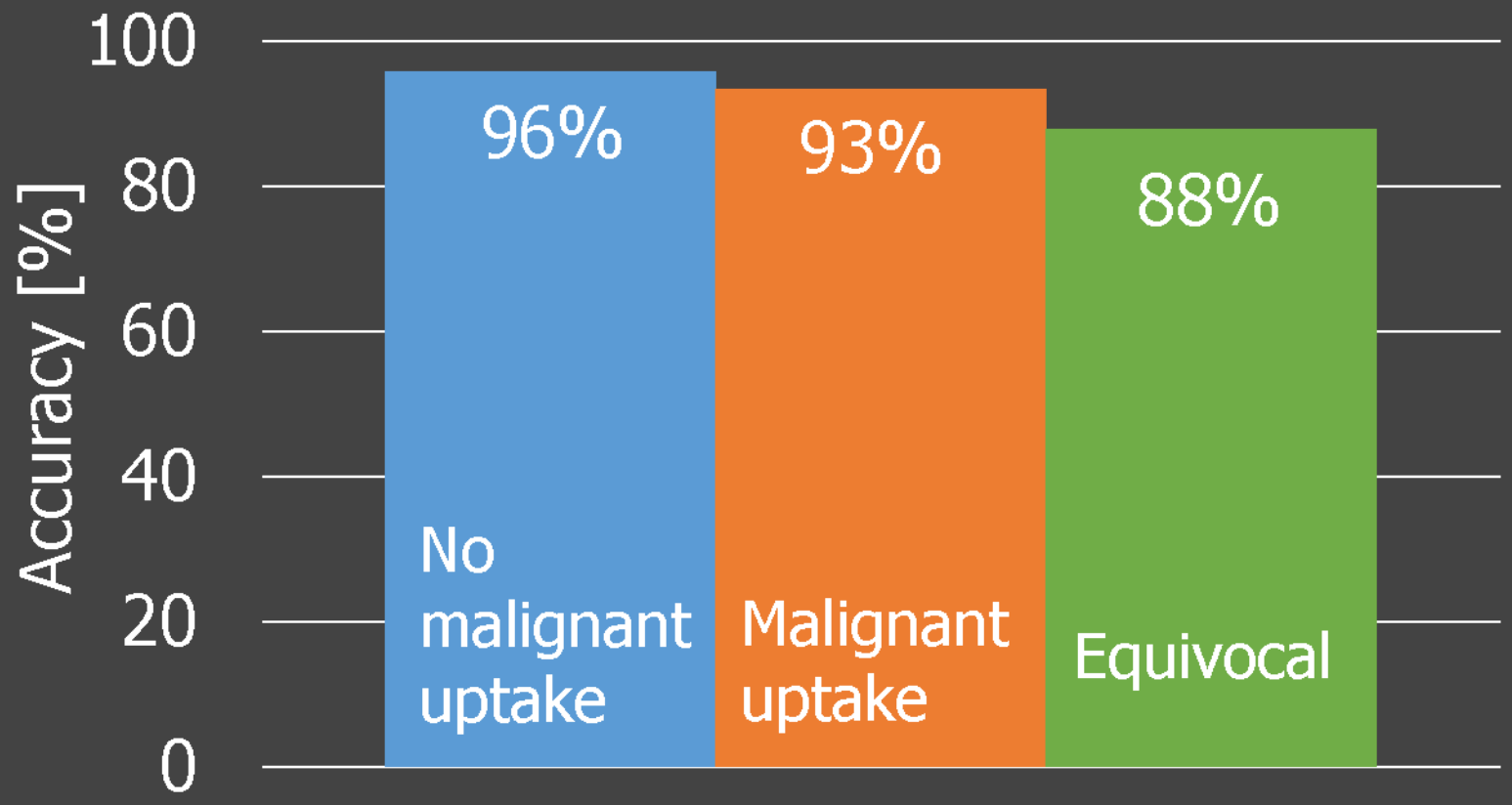
Results

- The model was trained for 25 epochs.
- The CNN process spent 15 minutes for training each fold dataset and <0.1 second / patient for prediction.



Results

- When images of no malignant uptake were given to the learned model, the accuracy was 96%.
- Similarly, the accuracy for images of malignant uptake and images of equivocal were 93% and 88% respectively.



Results - Detailed results -

- For no malignant uptake case, 96% cases were correctly predicted as no malignant uptake.

Prediction	Correct Label		
	(1) no malignant uptake	(2) malignant uptake	(3) equivocal
(1)	96%	4%	8%
(2)	1%	93%	4%
(3)	3%	3%	88%

Results - AI failed to predict -

- This case was "No malignant uptake " by physician’s diagnosis.
- However, the prediction of AI was "Malignant uptake".
- In this case, physiological accumulation of head and neck was high, possibly mimicking malignancy.

Summary

- In this research using CNN, for classification, FDG PET-CT images were classified into 3 categories.
- The overall accuracy was around 90%.

Discussions

- The overall accuracy of 90% was not enough to reduce physicians' burden.
- Especially, the accuracy was lower for "equivocal" than for other categories, possibly because "equivocal" group contained both malignant and non-malignant images and thus was difficult for AI to define; in other words "heterogenous group".

Discussion

- There is still room for improvement. As shown in the table, the percentage of false negatives is high.

Prediction	Correct Label: (1) no malignant uptake	(2) malignant uptake	(3) equivocal
(1)	96%	4%	8%
(2)	1%	93%	4%
(3)	3%	3%	88%

Discussion

- Future studies will make false negatives as small as possible.
- Such an automated diagnosis system will

  - select cases where radiologist further inspection is not necessary.
  - be effective for medical checkup.

Conclusion

- We developed and tested a convolutional neural network-based system to classify FDG-PET images into 1) normal, 2) abnormal, 3) equivocal.
- Although further improvement is needed, such a system will contribute to reducing physicians' burden and preventing oversight.