Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

A convolutional neural network-based system to detect malignant findings in FDG PET/CT examinations

Keisuke Kawauchi
Hokkaido University

Sho Furuya
Hokkaido University

Kenji Hirata (khirata@med.hokudai.ac.jp)
Hokkaido University

Chietsugu Katoh
Hokkaido University

Osamu Manabe
Hokkaido University

Kentaro Kobayashi
Hokkaido University

Shiro Watanabe
Hokkaido University

Tohru Shiga
Hokkaido University

Research article

Keywords: FDG, PET, convolutional neural network, deep learning

Posted Date: August 31st, 2019

DOI: https://doi.org/10.21203/rs.2.13839/v1

License: 🅭 🅯 This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at BMC Cancer on March 17th, 2020. See the published version at https://doi.org/10.1186/s12885-020-6694-x.

Abstract

Background As the number of PET/CT scanners increases and FDG PET/CT becomes a common imaging modality for oncology, the demands for automated detection systems on artificial intelligence (AI) to prevent human oversight and misdiagnosis are rapidly growing. We aimed to develop a convolutional neural network (CNN)-based system that can classify whole-body FDG PET as 1) benign, 2) malignant, or 3) equivocal.

Methods This retrospective study investigated 3,485 sequential patients with malignant or suspected malignant disease, who underwent whole-body FDG PET/CT at our institute. All the cases were classified into the 3 categories by a nuclear medicine physician. A residual network (ResNet)-based CNN architecture was built for classifying patients into the 3 categories. This network was trained with PET images. Five-fold cross-validations were carried out to estimate the classification performance. In addition, we examined whether the CNN could determine the location of the malignant uptake, be it in the head-and-neck region, chest, abdomen, or pelvic region.

Results There were 1,280 (37%), 1,450 (42%) and 755 (22%) patients classified as benign, malignant and equivocal, respectively. In patient-based analysis, the CNN predicted benign and malignant images with 99.4% and 99.4% accuracy, respectively. Furthermore, in region-based analysis, the prediction was correct with the probability of 97.3% (head-and-neck), 96.6% (chest), 92.8% (abdomen) and 99.6% (pelvic region), respectively.

Conclusion The CNN-based system reliably classified FDG PET images into 3 categories, indicating that it would be helpful for physicians as a double-checking system to prevent oversight and misdiagnosis.

Background

FDG PET/CT is widely used to detect metabolically active lesions, especially in oncology.[1, 2] PET/CT scanners are becoming widespread because of their usefulness, whereas the number of FDG PET/CT examinations has also increased. In Japan, the number of institutes that have installed a PET/CT scanner has increased by 177 (212 to 389) from 2007 to 2017, with examinations increasing 72% from 414,300 to 711,800.[3] In the current clinical practice, FDG PET/CT images require interpretation by specialists in nuclear medicine. As the physicians’ burden of interpreting images increases, the risk of oversight or misdiagnosis also increases. Therefore, there is a demand for an automated system that can prevent such incidents.

Image analysis using a convolutional neural network (CNN), a machine learning method, has attracted a great deal of attention as a method of artificial intelligence (AI) in the medical field.[4–7] CNN is a branch of deep neural network (so-called deep learning) techniques and is known to be feasible for image analysis because of its high performance at image recognition.[8] In a previous study using a CNN, tuberculosis was automatically detected on chest radiographs.[9] The use of a CNN also enabled brain tumor segmentation and prediction of genotype from magnetic resonance images.[10] Another study showed high diagnostic performance in the differentiation of liver masses by dynamic contrast agent-enhanced computed tomography.[11] CNN methods have also been applied to PET/CT, with successful results.[12–14]

We hypothesized that introducing an automated system to detect malignant findings would prevent human oversight/misdiagnosis. In addition, the system would be useful to select patients who need urgent interpretation by radiologists. Physicians who are inexperienced in nuclear medicine would particularly benefit from such a system.

In this research, we aimed to develop a CNN-based diagnosis system that classifies whole-body FDG PET images into 3 categories: 1) benign, 2) malignant, and 3) equivocal; such a system would allow physicians performing radiology-based diagnosis to double-check their opinions. In addition, we examined whether the CNN could determine the location of the malignant uptake, whether in the head-and-neck region, chest, abdomen or pelvic region.

Methods

Subjects

This retrospective study included 3,485 sequential patients (mean age ± SD, 63.9 ± 13.6 y; range, 24–95 y) who underwent whole-body FDG PET/CT and were classified by a nuclear medicine physician into 3 categories: 1,280 benign patients, 755 malignant patients and 1,450 equivocal patients. All patients were scanned on either Scanner 1 (N = 2,864, a Biograph 64 PET/CT scanner, Asahi-Siemens Medical Technologies Ltd., Tokyo) or Scanner 2 (N = 621, a GEMINI TF64 PET/CT scanner, Philips Japan, Ltd., Tokyo) at our institute between January 2016 and December 2017.

The institutional review board of Hokkaido University Hospital approved the study (#017–0365) and waived the need of written informed consent from each patient because the study was conducted retrospectively.

Labeling

An experienced nuclear medicine physician classified all cases into 3 categories: 1) benign, 2) malignant and 3) equivocal, based on the FDG PET maximum intensity projection (MIP) images and diagnostic reports. The criteria of classification were as follows.

1) The patient was labeled as malignant when any malignant uptakes were observed by the labeling physician and the corresponding description was found in the radiology report.

2) The patient was labeled as benign when no malignant uptakes were observed by the labeling physician, also confirmed by the radiology report. Inflammatory and physiological accumulations were considered benign.

3) The patient was labeled as equivocal when some abnormal accumulation was observed but it difficult to differentiate malignant from benign, or when the labeling physician did not agree with the conclusion of the radiology report.

The location of any malignant uptake was determined as A) head and neck, B) chest, C) abdomen, D) pelvic region. For the classification, the physician was blinded to the CT images and parameters such as maximum standardized uptake value (SUVmax). Diagnostic reports were made based on several factors including SUVmax, diameter of tumors, visual contrast between the tumors, location of tumors, and changes over time by 2+ physicians each with more than 8 years' experience in nuclear medicine.

Image acquisition and reconstruction

All clinical PET/CT studies were performed with either Scanner 1 or Scanner 2. All patients fasted for ≥6 hr before the injection of FDG (approx. 4 MBq/kg), and the emission scanning was initiated 60 min post-injection. For Scanner 1, the transaxial and axial fields of view were 68.4 cm and 21.6 cm, respectively. For Scanner 2, the transaxial and axial fields of view were 57.6 cm and 18.0 cm. Three-min emission scanning in 3D mode was performed for each bed position. Attenuation was corrected with X-CT images acquired without contrast media. Images were reconstructed with an iterative method integrated with (Scanner 1) or without (Scanner 2) a point spread function.

Each reconstructed image had a matrix size of 168 × 168 with the voxel size of 4.1 × 4.1 × 2.0 mm for Scanner 1, and a matrix size of 144 × 144 with the voxel size of 4.0 × 4.0 × 4.0 mm for Scanner 2. MIP images (matrix size 168 × 168) were generated by linear interpolation. MIP images were created at increments of 10-degree rotation for up to 180 or 360 degrees. Therefore, 18 or 36 angles of MIP images were generated per patient. In this study, CT images were used only for attenuation correction, not for classification.

Convolutional neural network (CNN)

A neural network is a computational system that simulates neurons of the brain. Every neural network has input, hidden, and output layers. Each layer has a structure in which multiple nodes are connected by edges. A "deep neural network" is defined as the use of multiple layers for the hidden layer. Machine learning using a deep neural network is called "deep learning." A convolutional neural network (CNN) is a type of deep neural network that has been proven to be highly efficient in image recognition. A CNN does not require predefined image features. We propose the use of a CNN to classify the images of FDG PET examination.

Architectures

In this study, we used a network model with the same configuration as ResNet. In the original ResNet, the output layer was classified into 1000 classes. We modified the number of classes to 3. We used this network model to classify whole-body FDG PET images into 1) benign, 2) malignant and 3) equivocal categories. Here we provide details on CNN architectures with the techniques used in this study. The detailed architecture is shown in Figure 1 and Table 1. Each neuron in a layer is connected to the corresponding neurons in the previous layer. The architecture of the CNN used in the present study contained five convolutional layers. This network also applied a rectified linear unit (ReLU) function, local response normalization, and softmax layers. The softmax function is defined as follows:

$$\mathrm{F}(x_i) = \frac{exp(x_i)}{\sum_j exp(x_j)}$$

where xi is the output of the neuron i (i = 1, 2, …, n, with n being the number of neurons belonging to the layer).

Model training and testing

Experiment 1 (Overall): First, input images were enlarged to (224, 224) to match the input size of the network. After that, we trained the CNN using data from the FDG PET images. The CNN was trained and validated. A 5-fold cross-validation scheme was used to validate the model. Subsequently, we tested the model. In the model-training phase, we used “early stopping” and “dropout” to prevent overfitting. Early stopping is a function used to monitor the loss function of training and validation and to stop the learning before falling into excessive learning.[15, 16] Early stopping and dropout have been adopted in various machine-learning methods.[17–19]

Experiment 2 (Region-based analysis): In this experiment, the neural network was given a subset of image data for training. Specifically, the CNN was trained for each of the four regions: A) head and neck, B) chest, C) abdomen, D) pelvic region. As a result, a model capable of predicting 3 categories was constructed for each of the 4 body parts. The configuration of the network was the same as in Experiment 1.

Experiment 3 (Grad-CAM): We carried out additional experiments using the Grad-CAM technique, which visualizes the part activating the neural network, or in other words, the part of the image that the neural network responds to. The same image as the original image used in Experiment 1 was used as the input image.

Hardware and software environments

This experiment was performed under the following environment:

OS, Windows 10 pro 64 bit; CPU, intel Core i7–6700K; GPU, NVIDIA GeForce GTX 1070 8GB; Framework, Keras 2.2.4 and TensorFlow 1.11.0; Language, Python 3.6.7; CNN, the same configuration as ResNet; Optimizer, Adam[20].

Results

We retrospectively collected a total of 3,485 patients, of whom 1,280 (37%) were labeled “benign”, 1,450 (42%) “malignant” and 755 (22%) “equivocal”. Figure 2 shows typical images of each category. A total of 76,785 maximum intensity projection (MIP) images were investigated. The number of images of benign patients, malignant patients, and equivocal patients were 28,688, 31,751 and 16,346, respectively.

Experiment 1 (Overall analysis)

In the image-based prediction, the model was trained for 30 epochs using an early stopping algorithm. The CNN process spent 3.5 hours for training and <0.1 second / image for prediction. When images of benign patients were given to the learned model, the accuracy was 96.6%. Similarly, the accuracies for images of malignant and equivocal patients were 97.3% and 77.8%, respectively. The results are shown in Figure 3 (a) and Table 1 (a).

The patient-based prediction was performed based on the following algorithm.

1. If more than 1/3 of the total images of the patient were judged as malignant, the patient was judged as being malignant.
2. If less than 1/3 of the total images were judged as malignant and more than 1/3 of the total images were judged as equivocal, the patient was judged as being equivocal.
3. If none of the above were satisfied, the patient was judged as being benign.

When images of benign patients were given to the learned model, the accuracy was 99.4%. The accuracy for images of malignant patients was also 99.4%. The accuracy was lower (87.5%) when images of equivocal patients were given. The results are shown in Figure 3 (b) and Table 1 (b). The prediction showed a tendency to fail especially when strong physiological accumulation (e.g., in the larynx) or mild malignant accumulation was present. Typical cases where the neural network failed to predict the proper category are shown in Figure 4.

Experiment 2 (Region-based analysis)

The same population was used in this experiment as was used in Experiment 1. The model was trained for 33–45 epochs for each dataset using an early stopping algorithm. The CNN process spent 4–5 hours for training and <0.1 second / image for prediction.

In the experiment for the head-and-neck region, a new labeling system was introduced to classify the images into 3 categories: 1) benign in the head-and-neck region, 2) malignant in the head-and-neck region, and 3) equivocal in the head-and-neck region. When images from “malignant in the head-and-neck region” patients were given to the learned model, the accuracy was 97.3%. The accuracy was 97.8% and 96.2% for “benign in the head-and-neck region” patients and “equivocal in the head-and-neck region” patients, respectively.

Similar experiments were performed for the chest, abdominal, and pelvic regions. The details of the results are shown in Figure 3 (c)-(f) and Table 2 (c)-(f). The accuracy was higher for the pelvic region (95.3—99.7%) than for the abdominal region (91.0—94.9%).

Experiment 3 (Grad-CAM[21])

We employed Grad-CAM to identify the part of the image from which the neural network extracted the largest amount of information. Typical examples are shown in Figure 5. Grad-CAM reasonably highlighted the area of malignant uptake by which physicians may have made a diagnosis.

Discussion

In patient-based classification, the neural network predicted correctly both the malignant and benign categories with 99.4% accuracy, although the accuracy for equivocal patients was 87.5%. Therefore, an average probability of 95.4% suggests that a CNN may be useful to predict 3-category classification from MIP images of FDG PET. Furthermore, in the prediction of the malignant uptake region, it was classified correctly with probabilities of 97.3% (head-and-neck), 96.6% (chest), 92.8% (abdomen) and 99.6% (pelvic region), respectively. These results suggested that the system may have the potential to help radiologists avoid oversight and misdiagnosis.

To clarify the reasons for the classification failure, we investigated some cases that were incorrectly predicted in Experiment 1. As expected, the most frequent patterns we encountered were strong physiological uptake and weak pathological uptake. In the case shown in Fig. 3a, the physiological accumulation in the oral region was relatively high, which might have caused erroneous prediction. In contrast, another case (Fig. 3b) showed many small lesions with low-to-moderate intensity accumulation, which was erroneously predicted as benign despite the true label being malignant. The equivocal category was more difficult for the neural network to predict; the accuracy was lower than for the other categories. The results may be due to the definition; though common in clinical settings, “equivocal” is a kind of catch-all or “garbage” category for all images not clearly belonging to “malignant” or “benign”; thus, a greater variety of images was included in the equivocal category. We speculate that such a wide range may have made it difficult for the neural network to extract consistent features.

We also conducted patient-based predictions in this study. In patient-based prediction, the accuracy was higher than in image-based prediction by an ensemble effect. This approach takes advantages of MIP images generated from various angles.

The CNN focuses on some features of the images. Grad-CAM is a technology that visualizes the region of interest. The results of Experiment 3 suggested that CNN responded to the part of the malignant uptake if presented. Grad-CAM results would provide physicians information on the mechanisms of the CNN; such information would help physicians decide whether to accept or reject the CNN's diagnosis.

The computational complexity becomes enormous when a CNN directly learns with 3D images.[22–26] Although we employed MIP images in the current study, an alternative approach may be to provide each slice to the CNN. However, even in the case of 'malignant' or 'equivocal', the tumor is usually localized in some small area and thus most of the slices do not contain abnormal findings. Consequently, a positive vs. negative imbalance problem would disturb efficient learning processes. In this context, MIP seems to be advantageous for a CNN as most MIP images of malignant patients contain accumulation in the image somewhere unless a stronger physiological accumulation (e.g., brain or bladder) hides the malignant uptake.

We believe that this system will be useful in various clinical situations. First, it can reduce oversight and misdiagnosis by physicians as an automated double-check system. Second, the system can assist less experienced physicians, especially residents, complete radiology reports. Third, it can be used as a triage system to determine priority cases for a radiologist's review. The radiologist would read those images the CNN classifies as malignant before reading the images of benign-classified patients. This could be highly useful in case that urgent care is needed.

In this study, we used only 2 scanners, but further studies are needed to reveal what will happen when more scanners are investigated. For instance, what if the numbers of examinations from various scanners is imbalanced? What if a particular disease is imaged by some scanners but not by the other scanners? There is a possibility that AI system cannot make a correct evaluation in such cases. The AI system should be tested using "real-world data" before using in clinical settings.

Some approaches could further improve the accuracy. In this research, in order to reduce the learning cost, we used a network that is equivalent to ResNet−34[27], which is a relatively simple version of the "ResNet" family. In fact, ResNet systems with deeper layers can be built technically. More recently, various networks based on ResNet have been developed and demonstrated to have high performance.[28, 29] From the viewpoint of big-data science, it is also important to increase the number of images for further improvement in diagnostic accuracy.

This study has some limitations. First, this model can only deal with FDG PET MIP images in the imaging range from the head to the knees; correct prediction is much more difficult when spot images or whole-body images from the head to the toes are given. Future studies will use regional CNN (RCNN) to solve the problem. Second, low-accumulation lesions such as pancreatic cancer cannot be classified only with MIP images, and there is a possibility that it cannot be labeled correctly. Third, the cases were classified by a nuclear medicine physician but were not based on a pathological diagnosis.

Conclusion

The CNN-based system successfully classified whole-body FDG PET images into 3 categories. In the region-based analysis, the CNN successfully determined the location of the malignant findings. The system would be useful for preventing physicians' oversight and misdiagnosis, for helping physicians inexperienced in nuclear medicine to make an accurate diagnosis, and for selecting cases requiring urgent radiologist interpretation.

Declarations

Ethics approval and consent to participate

The institutional review board of Hokkaido University Hospital approved the study (#017–0365) and waived the need of written informed consent from each patient because the study was conducted retrospectively registered.

Consent for publication

Not applicable

Availability of data and materials

Images used in this study cannot be released from the viewpoint of personal information protection. In addition, the amount of data is too large to publish.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was partly supported by the Center of Innovation Program from Japan Science and Technology Agency Grant Number H30W16 to purchase a computer for data analysis, and hard disks for data storage, and to compose the manuscript.

Authors' contributions

KKa, KH, and TS conceived the study concept. KKa conducted the experiments. KKa, KH, and SF interpreted the results. KKa, KH, and SF wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This research is partially supported by the Center of Innovation Program from Japan Science and Technology Agency, JST (Grant Number H30W16). We thank Eriko Suzuki for her support.

References

1. Mandelkern M, Raines J. Positron emission tomography in cancer research and treatment. Technol Cancer Res Treat. 2002;1:423–39. doi:10.1177/153303460200100603.
2. Nabi HA, Zubeldia JM. Clinical applications of (18)F-FDG in oncology. J Nucl Med Technol. 2002;30:1–3. https://www.ncbi.nlm.nih.gov/pubmed/11948260.
3. Kinuya K, Nishiyama Y, Kato R, Kayano D, Sato S, Tashiro M, et al. The 8th survey report on national nuclear medicine clinical practice in Japan. Radioisotopes. 2018;67:339–87.
4. Komeda Y, Handa H, Watanabe T, Nomura T, Kitahashi M, Sakurai T, et al. Computer-Aided Diagnosis Based on Convolutional Neural Network System for Colorectal Polyp Classification: Preliminary Experience. Oncology. 2017;93 Suppl 1:30–4. doi:10.1159/000481227.
5. Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng. 2017;19:221–48. doi:10.1146/annurev-bioeng–071516–044442.
6. Kahn Jr. CE. From Images to Actions: Opportunities for Artificial Intelligence in Radiology. Radiology. 2017;285:719–20. doi:10.1148/radiol.2017171734.
7. Dreyer KJ, Geis JR. When Machines Think: Radiology's Next Frontier. Radiology. 2017;285:713–8. doi:10.1148/radiol.2017171183.
8. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436. doi:10.1038/nature14539.
9. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology. 2017;284:574–82. doi:10.1148/radiol.2017162326.
10. Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. Sci Rep. 2017;7:5467. doi:10.1038/s41598–017–05848–2.
11. Yasaka K, Akai H, Abe O, Kiryu S. Deep Learning with Convolutional Neural Network for Differentiation of Liver Masses at Dynamic Contrast-enhanced CT: A Preliminary Study. Radiology. 2018;286:887–96. doi:10.1148/radiol.2017170706.
12. Xu L, Tetteh G, Lipkova J, Zhao Y, Li H, Christ P, et al. Automated Whole-Body Bone Lesion Detection for Multiple Myeloma on (68)Ga-Pentixafor PET/CT Imaging Using Deep Learning Methods. Contrast Media Mol Imaging. 2018;2018:2391925. doi:10.1155/2018/2391925.
13. Blanc-Durand P, Van Der Gucht A, Schaefer N, Itti E, Prior JO. Automatic lesion detection and segmentation of 18F-FET PET in gliomas: A full 3D U-Net convolutional neural network study. PLoS One. 2018;13:e0195798. doi:10.1371/journal.pone.0195798.
14. Ypsilantis PP, Siddique M, Sohn HM, Davies A, Cook G, Goh V, et al. Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks. PLoS One. 2015;10:e0137036. doi:10.1371/journal.pone.0137036.
15. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15:1929–58.

16. Haykin S. Neural Networks: A Comprehensive Foundation. Prentice Hall; 1994.
17. Yen C-W, Young C-N, Nagurka M. A vector quantization method for nearest neighbor classifier design. 2004. doi:10.1016/j.patrec.2004.01.012.
18. Karimpouli S, Fathianpour N, Roohi J. A new approach to improve neural networks' algorithm in permeability prediction of petroleum reservoirs using supervised committee machine neural network (SCMNN). J Pet Sci Eng. 2010;73:227–32. doi:https://doi.org/10.1016/j.petrol.2010.07.003.
19. Kahou SE, Michalski V, Konda K, Memisevic R, Pal C. Recurrent Neural Networks for Emotion Recognition in Video. Proc 2015 ACM. 2015;:467–74. doi:10.1145/2818346.2830596.
20. Diederik PK, Jimmy B. Adam: A Method for Stochastic Optimization. In arXiv:14126980. 2014.
21. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In arXiv:161002391v3. 2017.
22. Nie D, Cao X, Gao Y, Wang L, Shen D. Estimating CT Image from MRI Data Using 3D Fully Convolutional Networks. Deep Learn Data Label Med Appl. 2016;2016:170–8. doi:10.1007/978–3–319–46976–8_18.
23. Choi H, Lee DS, Alzheimer's Disease Neuroimaging I. Generation of Structural MR Images from Amyloid PET: Application to MR-Less Quantification. J Nucl Med. 2018;59:1111–7. doi:10.2967/jnumed.117.199414.
24. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. Med Phys. 2017;44:1408–19. doi:10.1002/mp.12155.
25. Martinez-Murcia FJ, Górriz JM, Ramírez J, Ortiz A. Convolutional Neural Networks for Neuroimaging in Parkinson's Disease: Is Preprocessing Needed? Int J Neural Syst. 2018;28:1850035. doi:10.1142/S0129065718500351.
26. Zhou Z, Chen L, Sher D, Zhang Q, Shah J, Pham N-L, et al. Predicting Lymph Node Metastasis in Head and Neck Cancer by Combining Many-objective Radiomics and 3-dimensioal Convolutional Neural Network through Evidential Reasoning. Conf Proc. Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf. 2018;2018:1–4. doi:10.1109/EMBC.2018.8513070.
27. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. p. 770–8. doi:10.1109/CVPR.2016.90.
28. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and. 2016. http://arxiv.org/abs/1602.07360. Accessed 7 Mar 2019.
29. Zagoruyko S, Komodakis N. Wide Residual Networks. 2016. http://arxiv.org/abs/1605.07146. Accessed 7 Mar 2019.

Tables

Table 1. Details of Experiment 1, 2.

	Experiment 1			
Prediction	**(a) Image-based**	Correct Label		
		Benign	Malignant	Equivocal
	Benign	96.6%	2.4%	10.1%
	Malignant	0.3%	97.3%	12.1%
	Equivocal	3.2%	0.2%	77.8%
	(b) Patient-based	Correct Label		
		Benign	Malignant	Equivocal
	Benign	99.4%	0.6%	3.8%
	Malignant	0.0%	99.4%	8.8%
	Equivocal	0.6%	0.0%	87.5%
	Experiment 2			
	(c) Head and Neck	Correct Label		
		Benign	Malignant	Equivocal
	Benign	97.8%	1.7%	3.0%
	Malignant	1.5%	97.3%	0.8%
	Equivocal	0.7%	1.1%	96.2%
	(d) Chest	Correct Label		
		Benign	Malignant	Equivocal
	Benign	98.4%	1.8%	5.9%
	Malignant	0.6%	96.6%	1.6%
	Equivocal	1.0%	1.6%	92.5%
	(e) Abdomen	Correct Label		
		Benign	Malignant	Equivocal
	Benign	94.9%	5.7%	7.0%
	Malignant	1.1%	92.8%	2.0%
	Equivocal	4.1%	1.5%	91.0%
	(f) Pelvic region	Correct Label		
		Benign	Malignant	Equivocal
	Benign	99.7%	0.4%	2.8%
	Malignant	0.1%	99.6%	1.9%
	Equivocal	0.3%	0.0%	95.3%

Table 2. Details of architecture.

Layer	Filter Size	Stride	Repeat count	Output Size
Input				(224, 224, 3)
Convolutional	(7, 7)	(2, 2)	1	(112, 112, 64)
Max pooling	(3, 3)	(2, 2)	1	(56, 56, 64)
Residual 1	(3 x 3, 64) (3 x 3, 64)	(1, 1)	3	(56, 56, 64)
Residual 2	(3 x 3, 128) (3 x 3, 128)	(2, 2)	4	(28, 28, 128)
Residual 3	(3 x 3, 256) (3 x 3, 256)	(2, 2)	6	(14, 14, 256)
Residual 4	(3 x 3, 512) (3 x 3, 512)	(2, 2)	3	(7, 7, 512)
Average pooling	(7, 7)	(1, 1)	1	(1, 1, 1024)
Fully connected				(3)

"Residual" contains the following structure. “1. Convolutional layer1, 2. Batch normalization1, 3. Activation layer1 (ReLU), 4. Convolutional layer2, 5. Batch normalization2, 6. Merge layer (Add), 7. Activation layer2 (ReLU).”

Figures

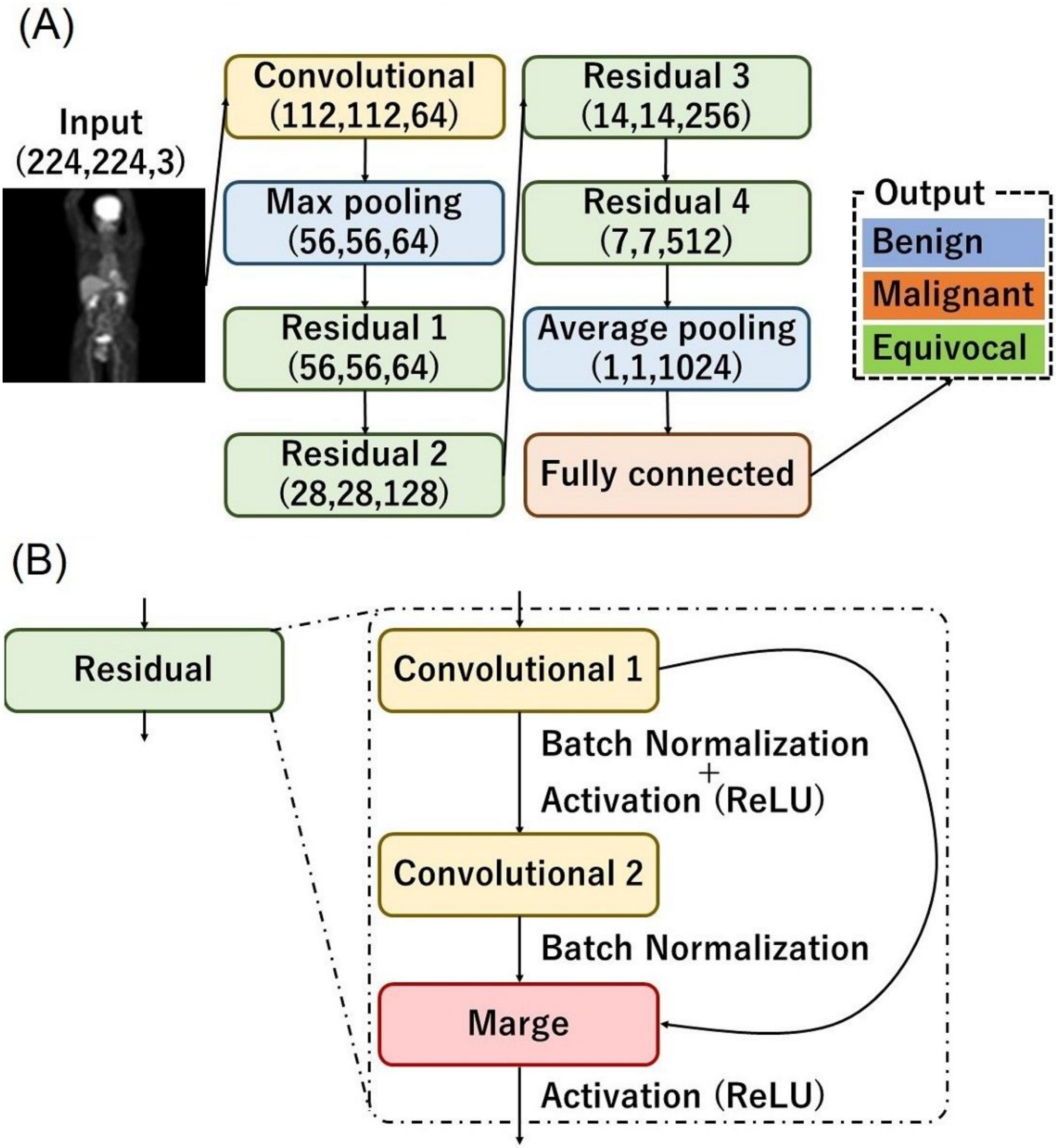


Figure 1

The functional architecture of the CNN. (A) The detailed structure of the CNN used in this study. (B) An internal structure of the residual layer.

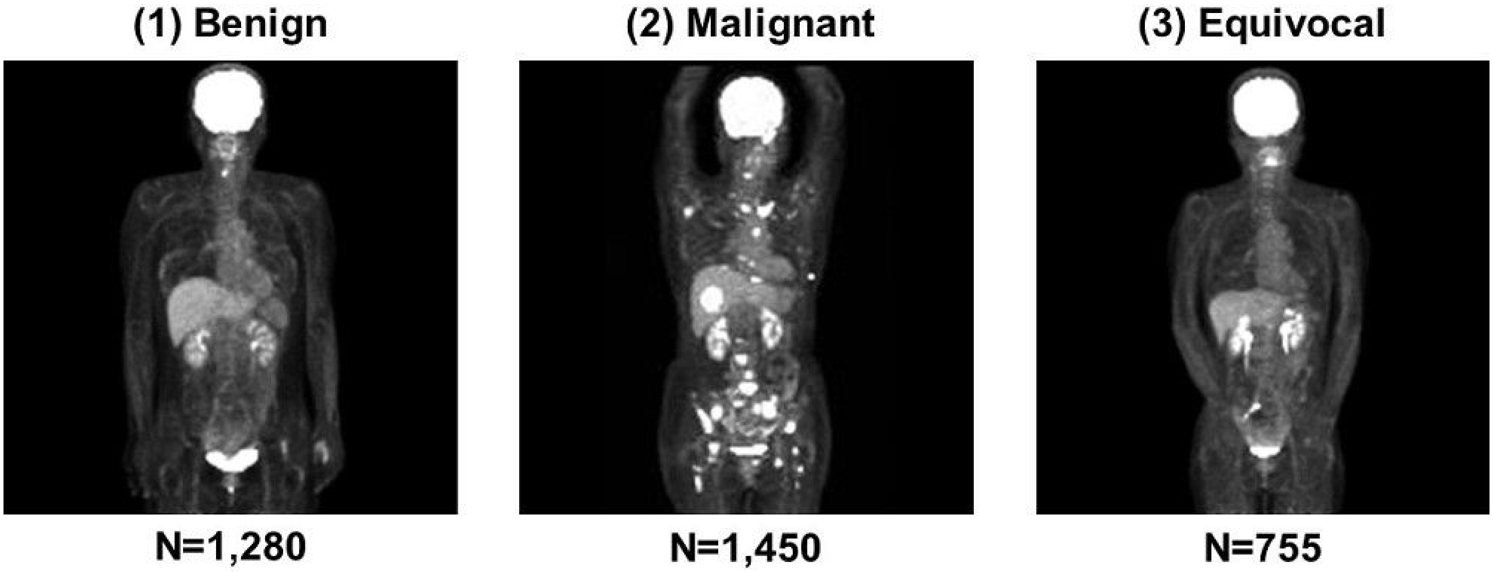


Figure 2

Typical cases in this study. (1) benign patient with physiological uptake in the larynx, (2) malignant uptake patient with multiple metastasis to bones and other organs, and (3) equivocal patient with abdominal uptake that was indeterminant between malignant or inflammatory foci.

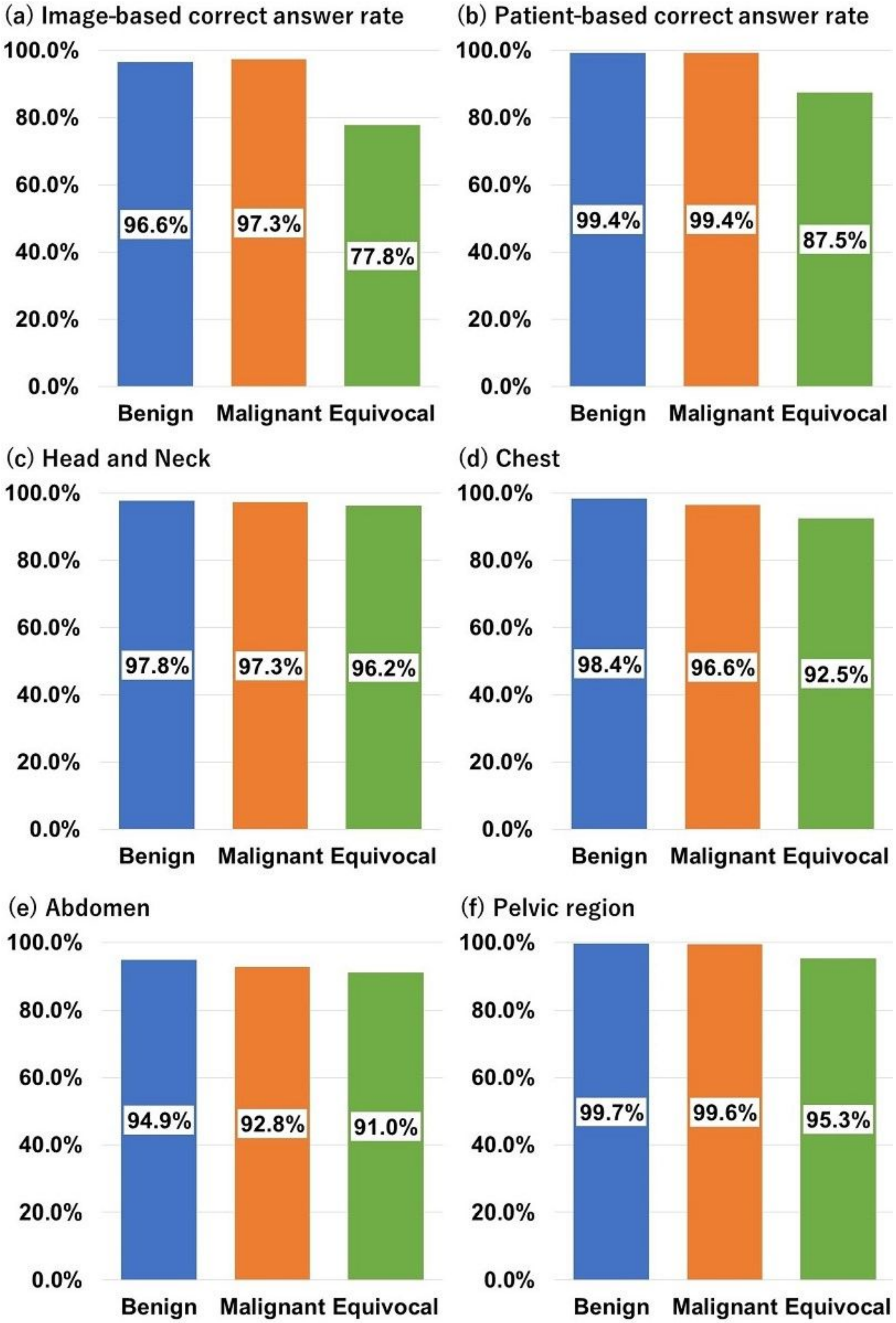
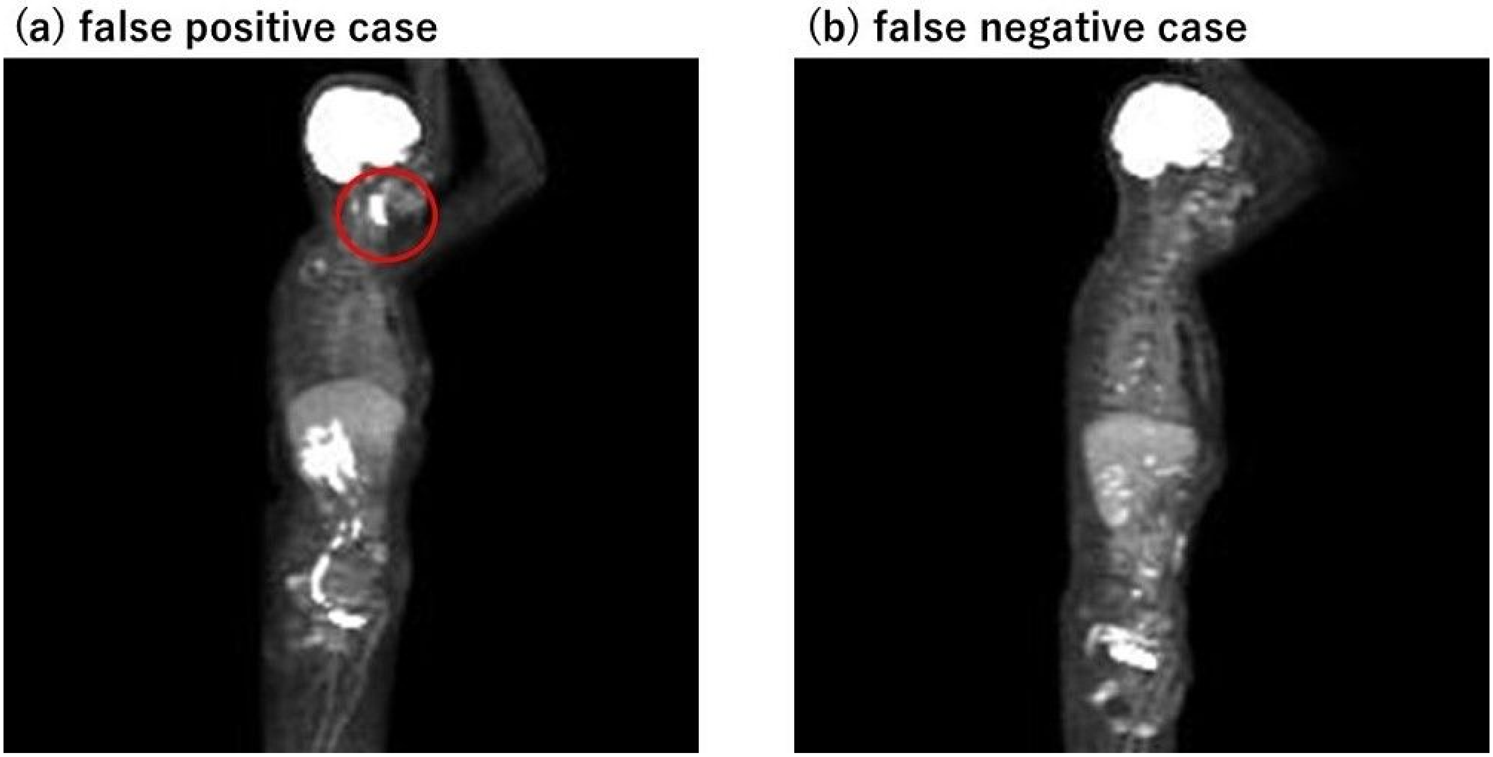


Figure 3

(a), (b) The overall correct answer rate of Experiments 1; (c) to (f) the correct answer rate for each malignant uptake region.



(a) false positive case

(b) false negative case

Figure 4

Typical cases whose category were incorrectly classified (a, false positive case; b, false negative case).

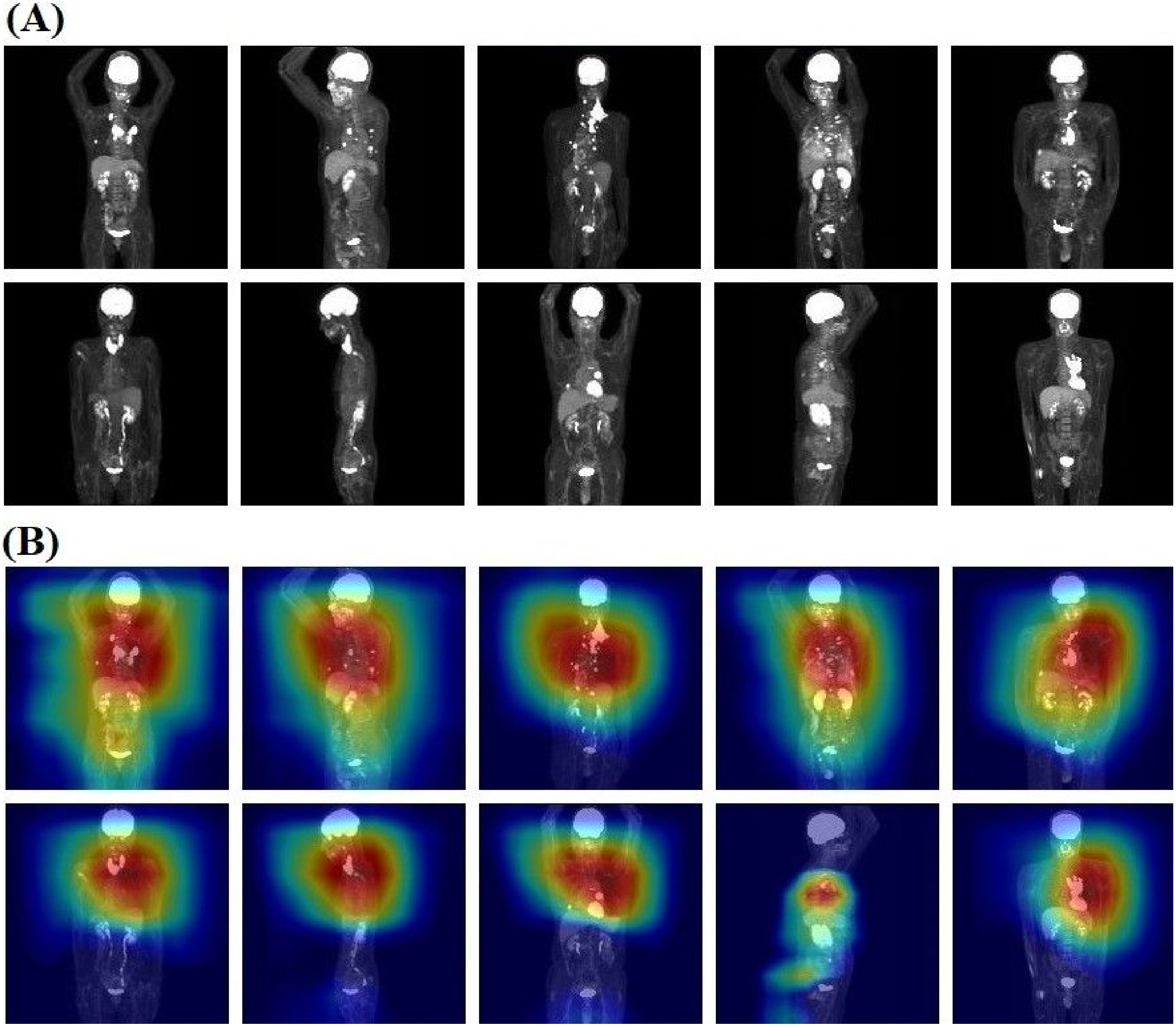


Figure 5

Visualization of classification standard of CNN. (A) Examples of original images input to CNN. (B) Examples of images output by Grad-CAM, highlighting the area of malignant uptake.