

Article

Improvement in the Convolutional Neural Network for Computed Tomography Images

Keisuke Manabe ¹, Yusuke Asami ¹, Tomonari Yamada ¹ and Hiroyuki Sugimori ^{2,*} 

¹ Graduate School of Health Sciences, Hokkaido University, Sapporo 060-0812, Japan; ksk0843@eis.hokudai.ac.jp (K.M.); yusuke12@eis.hokudai.ac.jp (Y.A.); tomonarihandball@eis.hokudai.ac.jp (T.Y.)

² Faculty of Health Sciences, Hokkaido University, Sapporo 060-0812, Japan

* Correspondence: sugimori@hs.hokudai.ac.jp; Tel.: +81-11-706-3410

Abstract: Background and purpose. This study evaluated a modified specialized convolutional neural network (CNN) to improve the accuracy of medical images. Materials and Methods. We defined computed tomography (CT) images as belonging to one of the following 10 classes: head, neck, chest, abdomen, and pelvis with and without contrast media, with 10,000 images per class. We modified the CNN based on the AlexNet with an input size of 512×512 . We resized the filter sizes of the convolution layer and max pooling. Using these modified CNNs, various models were created and evaluated. The improved CNN was evaluated to classify the presence or absence of the pancreas in the CT images. We compared the overall accuracy, which was calculated from images not used for training, to that of the ResNet. Results. The overall accuracies of the most improved CNN and ResNet in the 10 classes were 94.8% and 89.3%, respectively. The filter sizes of the improved CNN for the convolution layer were (13, 13), (7, 7), (5, 5), (5, 5), and (5, 5) in order from the first layer, and that of max-pooling was (7, 7). The calculation times of the most improved CNN and ResNet were 56 and 120 min, respectively. Regarding the classification of the pancreas, the overall accuracies of the most improved CNN and ResNet were 75.75% and 58.25%, respectively. The calculation times of the most improved CNN and ResNet were 36 and 55 min, respectively. Conclusion. By optimizing the filter size of the convolution layer and max-pooling of 512×512 images, we quickly obtained a highly accurate medical image classification model. This improved CNN can be useful for classifying lesions and anatomies for related diagnostic aid applications.

Keywords: deep learning; convolutional neural network; image classification



Citation: Manabe, K.; Asami, Y.; Yamada, T.; Sugimori, H. Improvement in the Convolutional Neural Network for Computed Tomography Images. *Appl. Sci.* **2021**, *11*, 1505. <https://doi.org/10.3390/app11041505>

Academic Editor: Chuan-Ming Liu

Received: 12 January 2021

Accepted: 5 February 2021

Published: 7 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Image classification is a typical technology of image analysis that uses artificial intelligence. In computed tomography (CT) images, it has been used to classify pulmonary nodules [1,2], slice positions [3,4], and calcaneal fractures [5,6]. Both AlexNet [7] and ResNet [8] are examples of image classification models. Because AlexNet has fewer layers than ResNet, its accuracy is low but the calculation time is short. Learning images with a large pixel size results in graphics processing unit (GPU) memory shortage, as the calculation cost is large. Therefore, images resized to 224×224 , which is the default pixel size of most image classification models, are often used for training [9,10]. In the study by Santin et al. [9], augmentation of data and transfer learning were used to improve the accuracy and robustness of the model, but the pixel size of the input images was $224 \times 224 \times 3$; thus, the pixel size was not examined. However, in medical images, the general pixel size is 512×512 . Reducing the image size may reduce the number of features as a result of the compression of image information. Because several tens of thousands of images are required for medical image classification, learning takes a lot of time [11]. The existing image classification model, AlexNet, has low accuracy because it is an early model, but the calculation time is short because it has few layers. If AlexNet can be specialized to

512×512 , two problems may be solved: the reduction of features due to resizing and the long calculation time, which is a weakness of models with a large number of parameters. Being able to train with the original size images is useful because, in actual diagnosis, micro lesions need to be detected. In this study, the parameters of AlexNet were customized for medical images, and the accuracy and calculation time of the convolutional neural network (CNN) were evaluated. The generalization capability of the improved CNN was evaluated by classifying the presence or absence of the pancreas, which is considered difficult [12].

2. Materials and Methods

2.1. Subjects and Datasets

In this study, we targeted 118,000 512×512 axial CT images. These images were approved by the Hokkaido University Hospital Ethics Committee. In the 10 classes for adjusting AlexNet parameters, 100,000 images were used for training, and 10,000 images were used for accuracy verification. In the classification of the presence or absence of the pancreas to evaluate the generalization capability of the improved CNN, 6000 images were used for training, and 2000 images were used for accuracy verification.

2.2. Ten Classes for Adjusting AlexNet Parameters

We defined training and accuracy verification images as the following 10 classes: head, neck, chest, abdomen, and pelvis with and without contrast media. Training images comprised 10,000 images per class, and accuracy verification images included 1000 per class. The original AlexNet was trained with 224×224 and 512×512 images, and the overall accuracy and calculation time of these two models were compared. The respective models were named original (input image size: 224×224) and original (input image size: 512×512). The AlexNet parameters adjusted for 512×512 images were the filter sizes of the convolution layer and max pooling. Because the filter sizes of the convolution layer had many change patterns, we divided them into five groups—A, B, C, D, and E—and named each model as A1 and A2 and so on. Figure 1 presents the change patterns of groups A to E. Figure 2 shows the original and changed values of the filter size of the convolution layer. On the other hand, the original value of the filter size of max-pooling was (3, 3), which we changed to an odd number of 5 to 15. We calculated the overall accuracies of these models using accuracy verification images. Various models were created with all combinations of parameters exceeding the overall accuracy of the original (input image size: 512×512), and the overall accuracy was calculated. The names of these models were “group name of convolution layer”–“filter size of max pooling,” such as A1-5. Next, the original ResNet was trained with 224×224 images, and the overall accuracy and calculation time were calculated. We compared the overall accuracy, confusion matrix, and calculation time of the model with the highest overall accuracy among the created models to those of the ResNet.

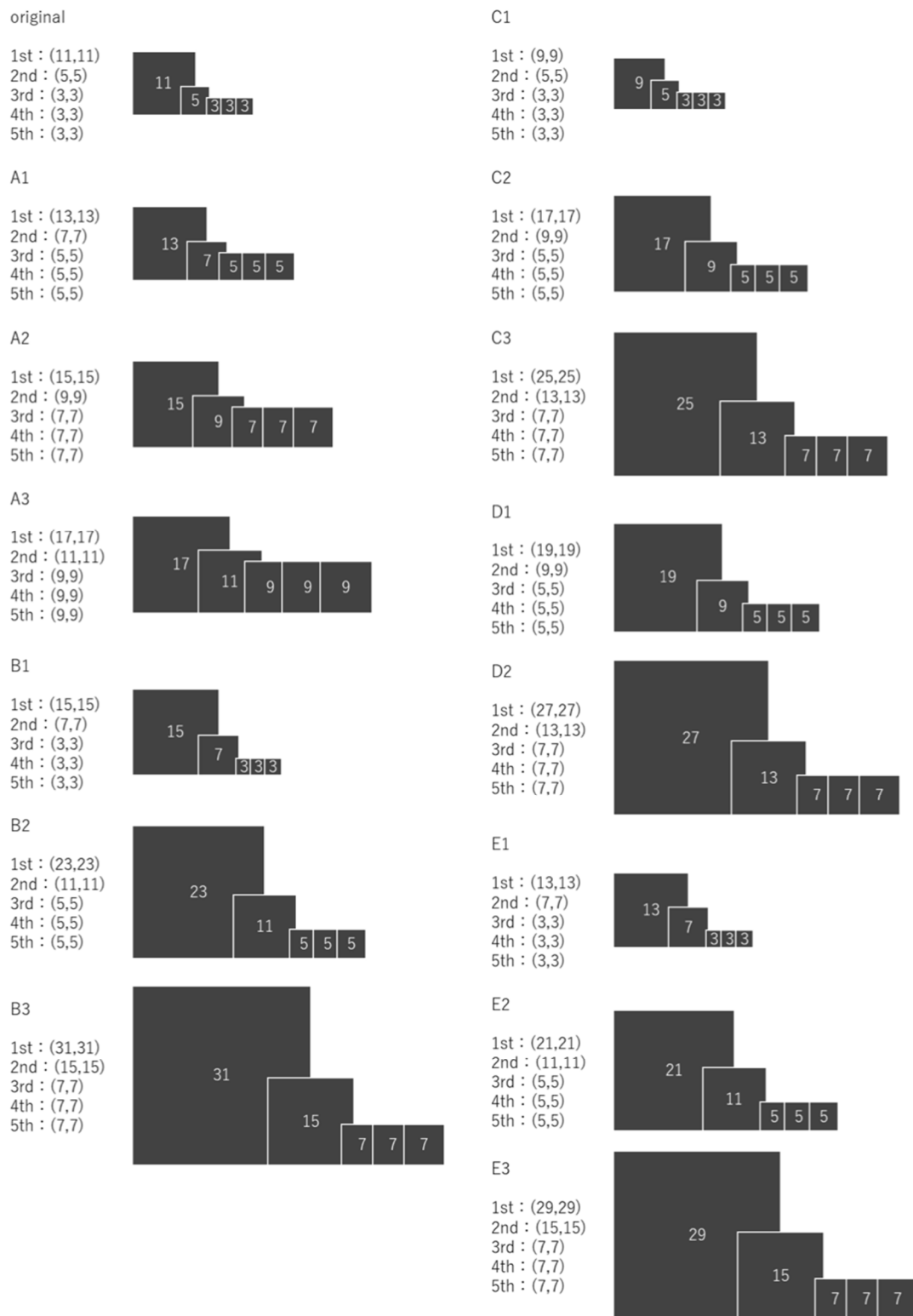


Figure 2. Original and changed values of the filter size of the convolution layer. [(A–E): the change patterns of group name (groups A to E)].

2.3. Classification of the Presence or Absence of the Pancreas to Evaluate the Generalization Capability of the Improved CNN

We defined training and accuracy verification images as the following four classes: the presence or absence of the pancreas with and without contrast media. Training images

were 1500 images per class, and accuracy verification images were 500 per class. ResNet and the model with the highest overall accuracy among the created models were trained, and we compared the overall accuracy, confusion matrix, and calculation time. Figure 3 shows the entire learning and evaluation process. For the training, we used a PC with NVIDIA GeForce GTX TITAN X 12GB (NVIDIA Corporation, Santa Clara, CA, USA).

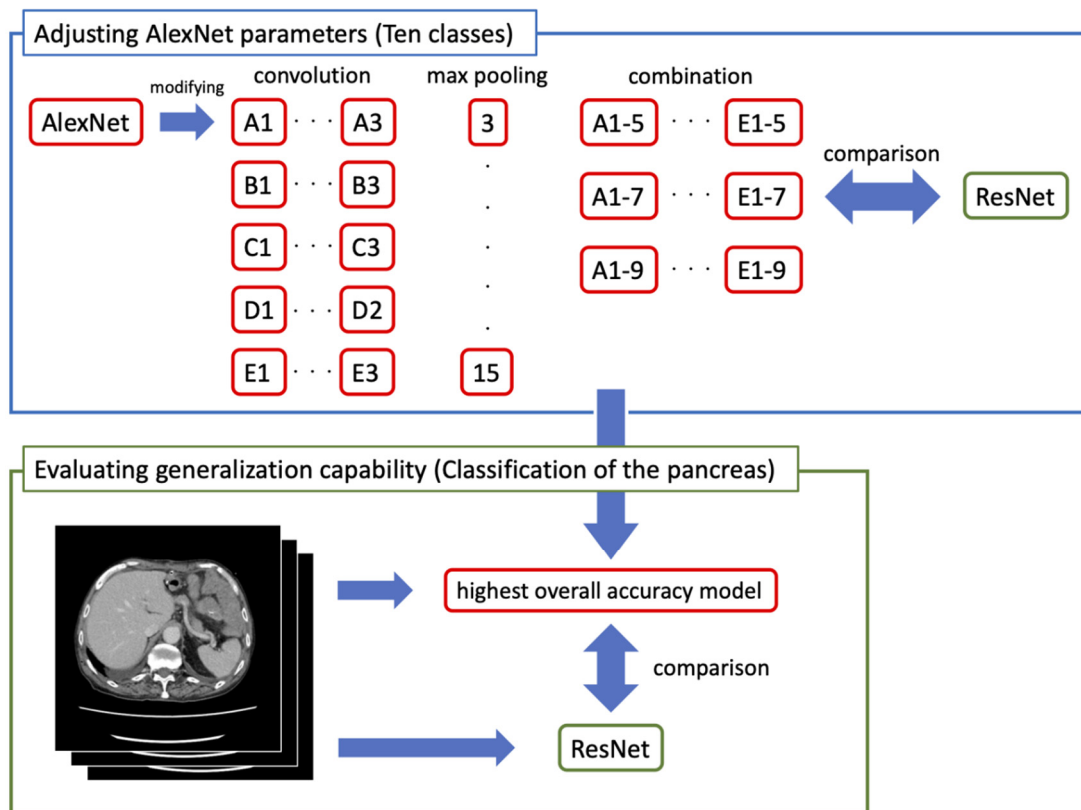


Figure 3. The entire learning and evaluation process.

3. Results

3.1. Ten Classes for Adjusting AlexNet Parameters

Table 1 shows the overall accuracies and calculation times of the original AlexNet trained with 224×224 and 512×512 images. In the comparison between original (input image size: 224×224) and original (input image size: 512×512), original (input image size: 512×512) had higher overall accuracy and a shorter calculation time. Figure 4 presents the overall accuracies and calculation times of the models in which the filter sizes of the convolution layer were changed, and Figure 5 shows the overall accuracies and calculation times of the models in which the filter sizes of max-pooling were changed. The models exceeding the overall accuracy of original (input image size: 512×512) were A1~3, B1~3, C1~2, D1~2, and E1, and the models with filter sizes of max-pooling were 5, 7, and 9. Figures 6–8 show the overall accuracies and calculation times of the models with all combinations of these parameters. Among these models, the highest overall accuracy was in model A1-7. The filter sizes of the convolution layer for A1-7 were (13, 13), (7, 7), (5, 5), (5, 5), and (5, 5) in order from the first layer, and that of max-pooling was (7, 7). The overall accuracy of A1-7 was 94.40%, and the calculation time was 56 min. Figure 9 displays the confusion matrix. In contrast, the overall accuracy of ResNet was 88.80%, and calculation time was 120 min. The confusion matrix is shown in Figure 10. In the comparison of A1-7 and ResNet, A1-7 was superior in both overall accuracy and calculation time.

Table 1. Overall accuracies and calculation times of the original AlexNet.

	Overall Accuracy [%]	Calculation Time [min]
Original (input image size: 224 × 224)	82.5	41
Original (input image size: 512 × 512)	86.7	29

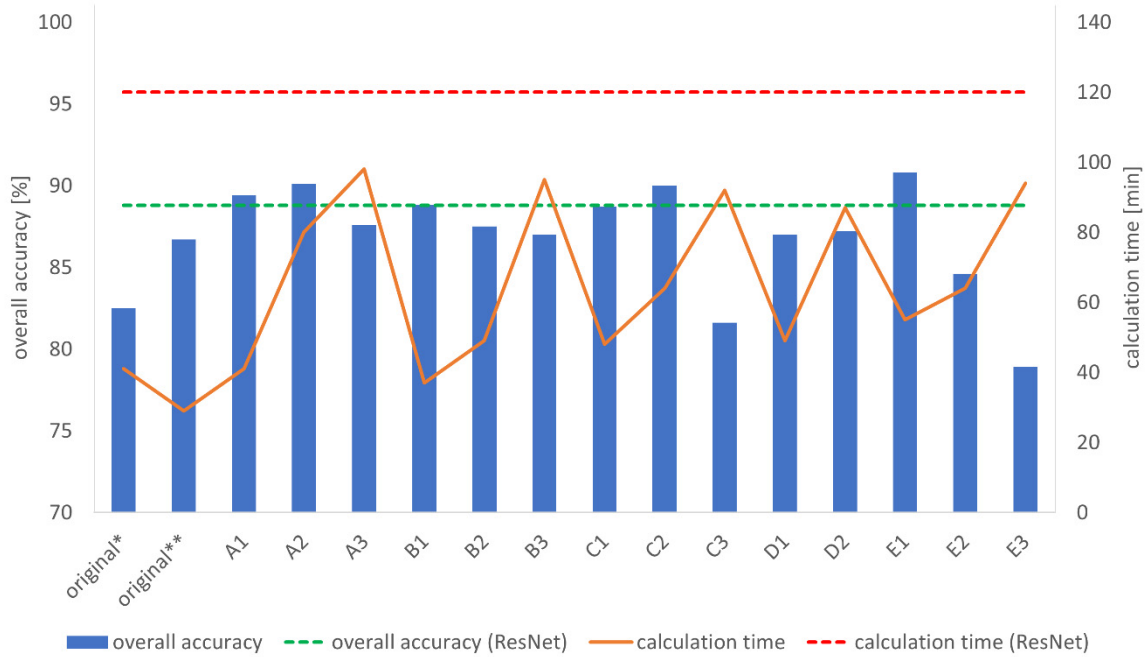


Figure 4. Overall accuracies and calculation times of the models in which the filter sizes of the convolution layer were changed (original *: original (input image size: 224 × 224), original **: original (input image size: 512 × 512)).

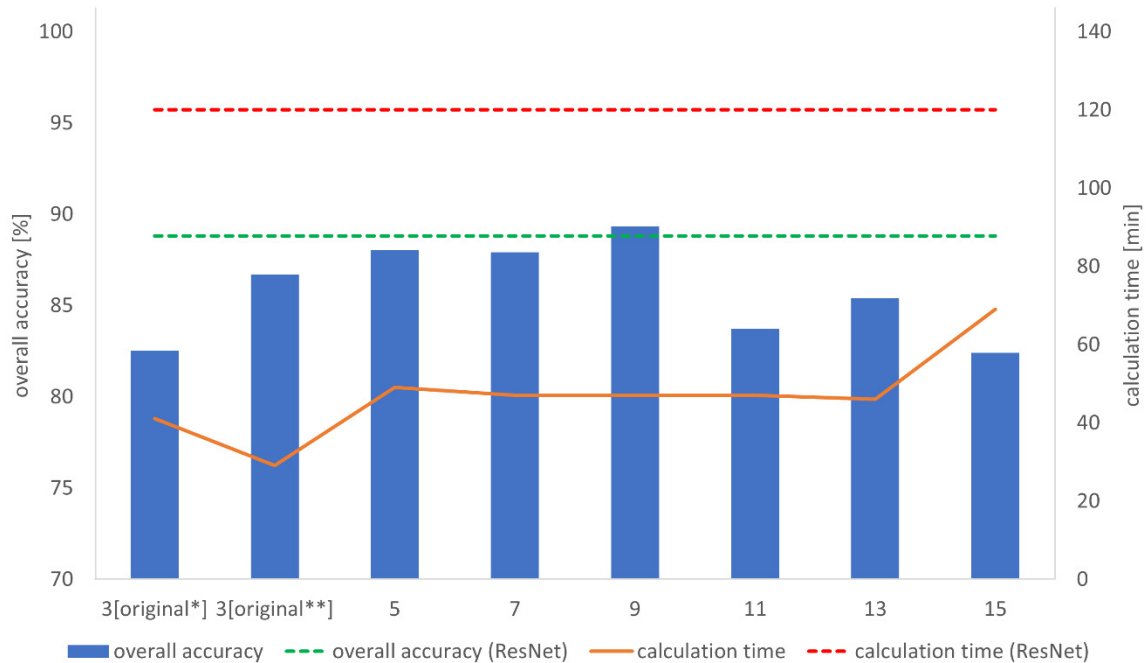


Figure 5. Overall accuracies and calculation times of the models with the filter sizes of max-pooling (original *: original (input image size: 224 × 224), original **: original (input image size: 512 × 512)).

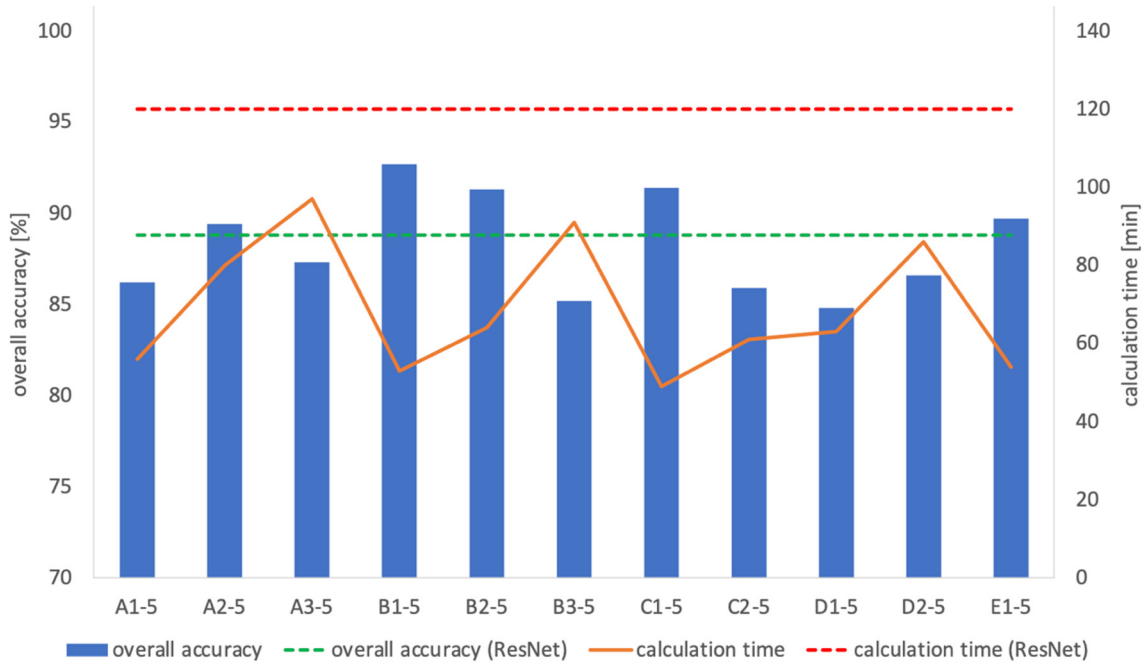


Figure 6. Overall accuracies and calculation times of the models with combination (the filter size of max-pooling: (5, 5)).

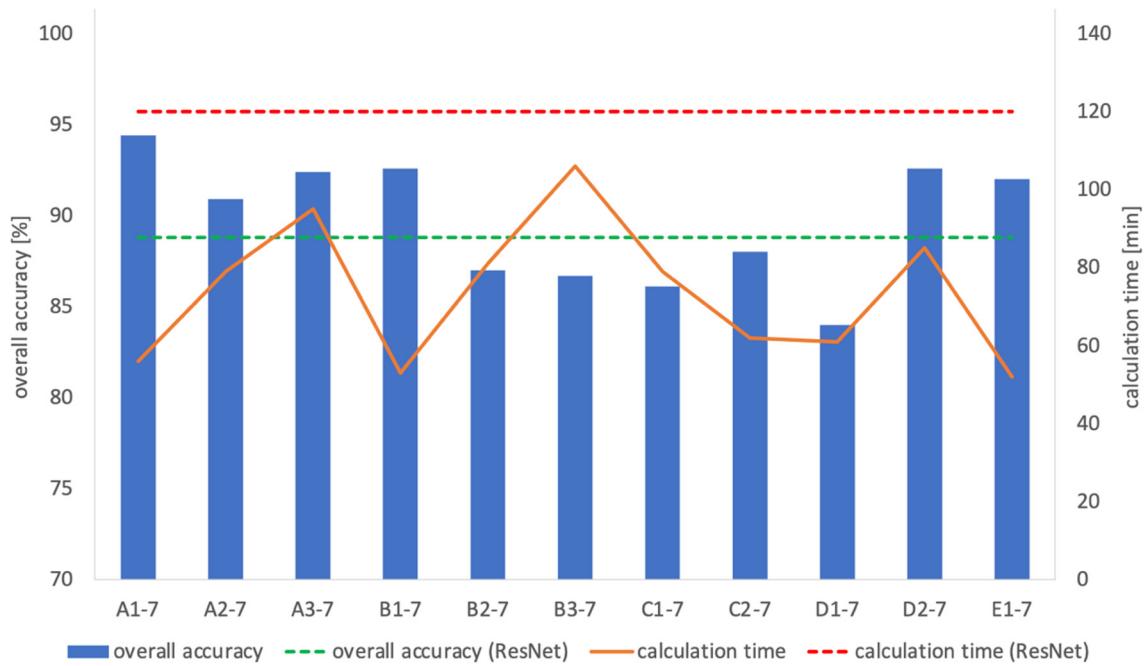


Figure 7. Overall accuracies and calculation times of the models with combination (the filter size of max-pooling: (7, 7)).

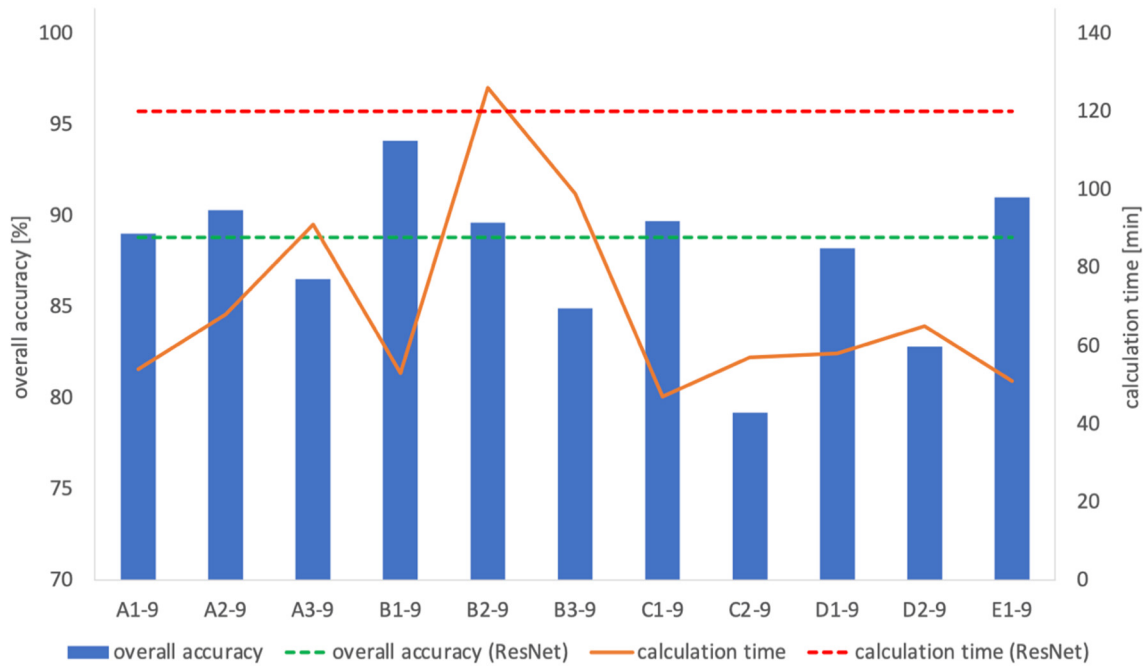


Figure 8. Overall accuracies and calculation times of the models with combination (the filter size of max-pooling: (9, 9)).

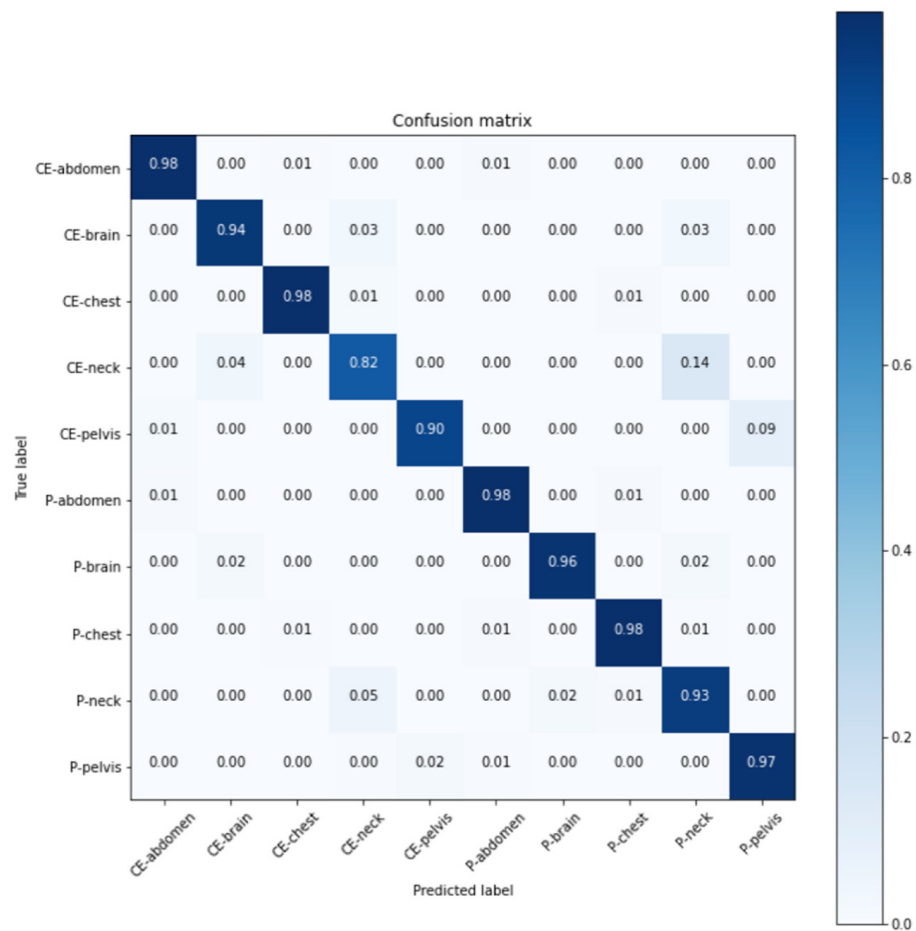


Figure 9. Confusion matrix of A1-7. (CE: contrast-enhanced, P: plain).

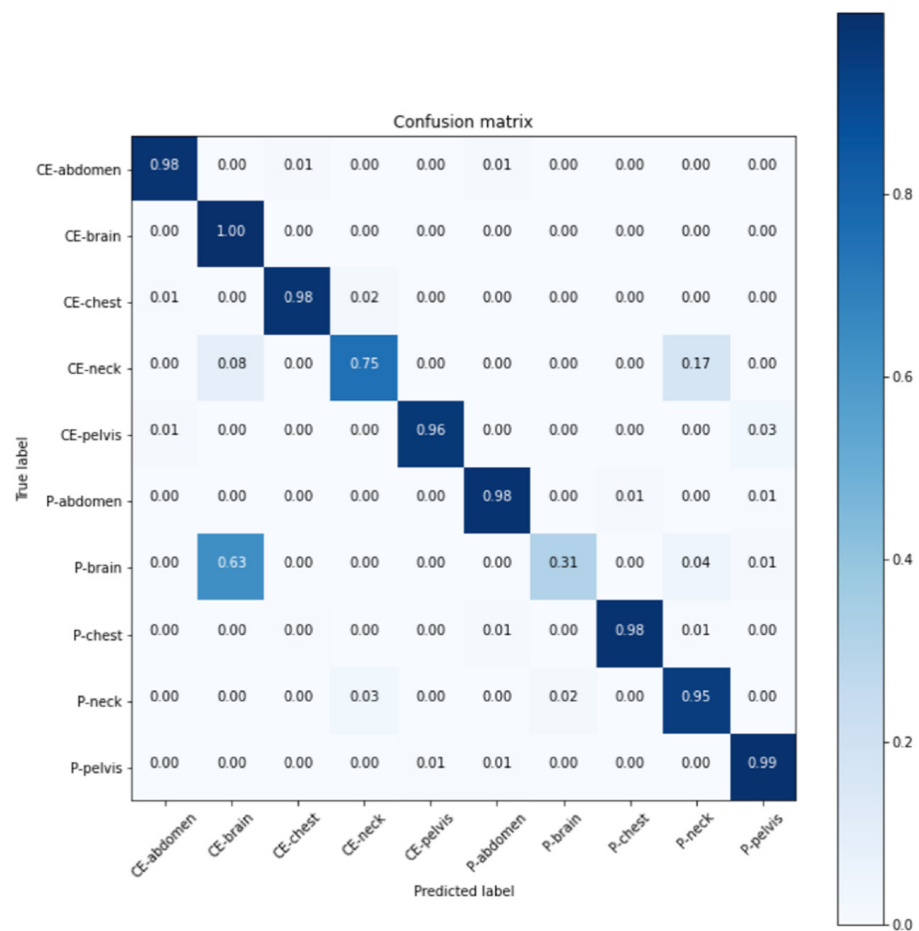


Figure 10. Confusion matrix of ResNet (CE: contrast-enhanced, P: plain).

3.2. Classification of the Presence or Absence of the Pancreas for the Evaluation of the Generalization Capability of the Improved CNN

The overall accuracy of A1-7 was 75.75%, and the calculation time was 36 min. Figure 11 shows the confusion matrix. The overall accuracy of ResNet was 58.25%, and the calculation time was 55 min. Figure 12 displays the confusion matrix. In the comparison of A1-7 and ResNet, A1-7 was superior to ResNet in both overall accuracy and calculation time, as in the 10 classes.

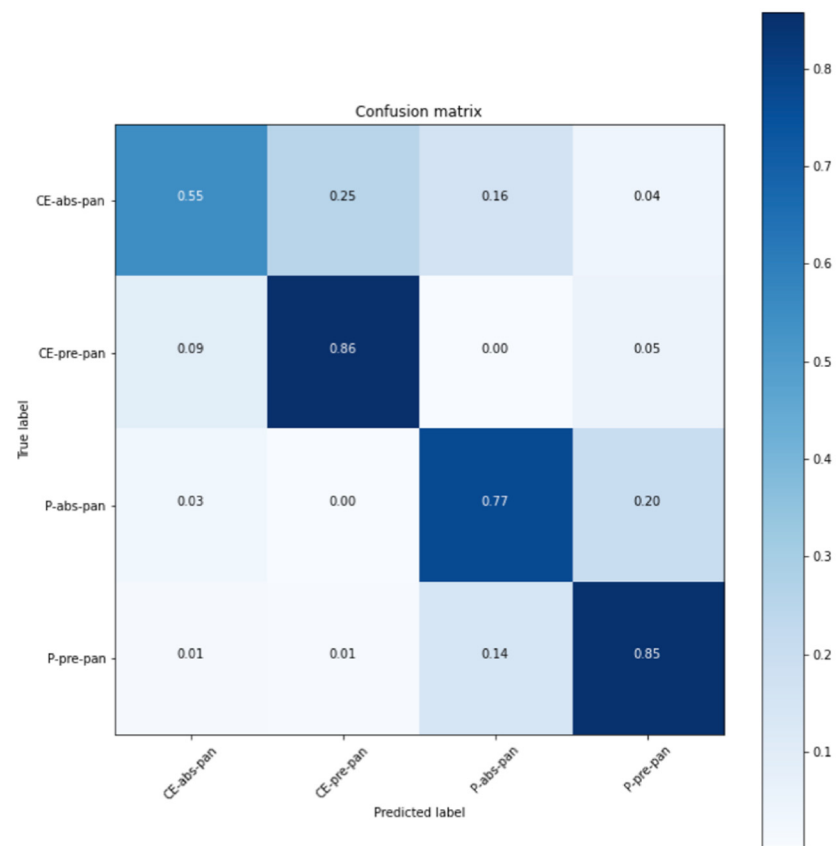


Figure 11. Confusion matrix of A1-7. (CE: contrast-enhanced, P: plain).

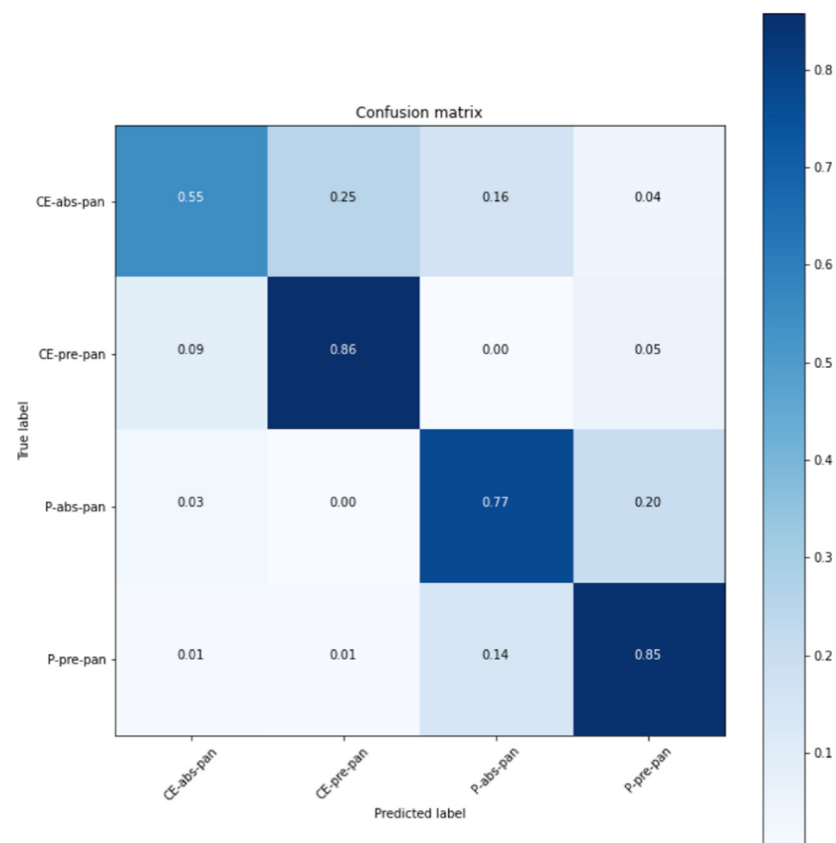


Figure 12. Confusion matrix of ResNet (CE: contrast-enhanced, P: plain).

4. Discussion

In the comparison between original (input image size: 224×224) and original (input image size: 512×512), original (input image size: 512×512) was found to have higher overall accuracy and a shorter calculation time. This means that training with large pixel-sized images may improve the overall accuracy and increasing the pixel size does not always increase the calculation time.

As a result of changing the filter size of the convolution layer, the overall accuracy was improved when the filter size was slightly increased, and it was decreased when the filter size was further increased in groups A to E. This result suggests that when training with images whose pixel size is larger than the original CNN, the overall accuracy is improved by appropriately increasing the filter size of the convolution layer. In the 512×512 image, the original filter size has a narrow range for the extraction of features, which makes it difficult to extract the overall features. Therefore, increasing the filter size improved the overall accuracy. However, if the filter sizes are made too large, detailed features cannot be extracted, and the overall accuracy is decreased. On the other hand, the calculation time became longer as the filter size increased. This is because the filter size was increased without changing the stride. The number of feature extractions was the same, but the range increased; thus, the calculation time became longer.

When the filter sizes of max-pooling were changed to an odd number from 5 to 15, the models with filter sizes 5, 7, and 9 exceeded original (input image size: 512×512), and the models after 11 were less than original (input image size: 512×512). This is because the overall features can be extracted by increasing the filter sizes of max-pooling to fit the 512×512 images. However, if the filter sizes of max-pooling are made too large, extraction of detailed features is not possible, and the overall accuracy is reduced. On the other hand, the calculation time was almost constant. Max-pooling is an operation used for the extraction of the maximum value, and the amount of the calculation is not large. Therefore, the calculation time was constant, even if the filter size was increased. When combining the filter sizes of the convolution layer and max-pooling, which had high overall accuracy, the overall accuracy of A1-7 was the highest. A1 and the model in which the filter size of max-pooling was 7 were not the most accurate models when changed separately. Combining the parameters of the model with the highest overall accuracy does not mean that a model with higher overall accuracy can be created. Thus, the overall accuracy varies depending on the affinity of the parameter combinations.

In the comparison of A1-7 and ResNet, A1-7 was superior in both overall accuracy and calculation time. According to the confusion matrix, ResNet often misclassified heads without contrast media as heads with contrast media. In contrast, the accuracy of A1-7 was 96%, which is about three times higher. Older CNNs, such as AlexNet, can exceed the overall accuracy of the relatively new ResNet by specializing the parameters to 512×512 images. Therefore, by specializing the parameters to the uncompressed pixel size, the overall accuracy of CNNs trained with compressed images might be improved.

Regarding the pancreas classification, A1-7 was superior to ResNet in both overall accuracy and calculation time, as in the 10 classes. According to the confusion matrix, the classification of images with the pancreas had about the same accuracy, but the accuracy of A1-7 in the classification of images without the pancreas was about twice that of ResNet. For this reason, A1-7 is not a CNN specialized for only 10 classes but is rather a generalized CNN.

In the study by Lakhani et al. [13], the authors used AlexNet and GoogLeNet [14] and created four types of image classification models with and without transfer learning, and the accuracy of the classification of tuberculosis was compared. Although the creation of multiple models and the accuracy comparison were similar to those in this study, Lakhani et al. did not change the CNN parameters, such as the filter size of the convolution layer. AlexNet, GoogLeNet, and transfer learning are technologies developed for general images and are not specialized for medical imaging. Therefore, in this study, we improved the

accuracy by specializing the filter sizes of the convolution layer and max-pooling for medical images.

This study has five limitations. First, the comparison target was only ResNet. The created CNN was compared with ResNet, which is a typical example of high-performance CNN; however, the latest CNN is more accurate [15–20], so the created CNN should have been compared with the latest CNN. Second, we did not evaluate some parameters. However, considering that there is a limit to the number of parameters that can be evaluated individually and that we obtained a result of 94.40%, we believe that the number of parameters was sufficient. Third, we used the holdout validation. Ideally, k-fold cross-validation should be used, but we chose the holdout validation study because there were too many models to validate. In general, the ratio of data sets for holdout validation is 80:20 [5,21], and the mean performance is obtained from multiple data sets. In this study, the ratio of data sets was not 80:20, and the number of data sets was one. However, since some papers use a 90:10 ratio of data sets [22,23] and others use various ratios [1,6,24], this is not considered a problem. As for the number of data sets, it is not a problem since there are papers that validate with only one [3,4,25]. Fourth, only one generalization capability test was used in this study. By classifying small lesions, it is possible to verify the ability to respond to minute changes and clinical practicality. Since the accuracy in actual diagnosis is not verified, the classification of lesions and malignancy needs to be verified before it is used in clinical practice. However, being able to train with the original pixel size without resizing is useful, because it may capture more minute features. Finally, we tested the generalization capability only on the model with the highest overall accuracy. Testing with other models could reveal the relationship between accuracy and parameters.

The CNN we created in this study can quickly create a model, even if it is trained with a large number of medical images. This feature has the potential to create an image classification model that can be updated daily by training with images taken on the same day. As a result, the created CNN can be optimized for the imaging method, rules, radiologist habits, and patient tendency for each facility and can contribute to the creation of a diagnostic support application specialized for each facility.

5. Conclusions

By optimizing the filter size of the convolution layer and max-pooling of 512×512 images, we were able to quickly obtain a highly accurate medical image classification model. This improved CNN can be useful for the classification of lesions and anatomies for related diagnostic aid applications.

Author Contributions: K.M. contributed to data analysis, algorithm construction, and the writing and editing of the article; Y.A. and T.Y. contributed to reviewing and editing the paper; H.S. proposed the idea and contributed to data acquisition, performed supervision, project administration, and reviewed and edited the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Hokkaido University Hospital Ethics Committee.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The created models in this study are available on request from the corresponding author. However, the image datasets presented in this study are not publicly available due to ethical reasons, e.g., containing information that could compromise the privacy of research participants.

Conflicts of Interest: The authors declare that no conflicts of interest exist.

References

1. Chae, K.J.; Jin, G.Y.; Ko, S.B.; Wang, Y.; Zhang, H.; Choi, E.J.; Choi, H. Deep Learning for the Classification of Small (≤ 2 cm) Pulmonary Nodules on CT Imaging: A Preliminary Study. *Acad. Radiol.* **2020**, *27*, e55–e63. [[CrossRef](#)] [[PubMed](#)]
2. Nishio, M.; Sugiyama, O.; Yakami, M.; Ueno, S.; Kubo, T.; Kuroda, T.; Togashi, K. Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. *PLoS ONE* **2018**, *13*, e0200721. [[CrossRef](#)] [[PubMed](#)]
3. Sugimori, H. Evaluating the overall accuracy of additional learning and automatic classification system for CT images. *Appl. Sci.* **2019**, *9*, 682. [[CrossRef](#)]
4. Sugimori, H. Classification of Computed Tomography Images in Different Slice Positions Using Deep Learning. *J. Healthc. Eng.* **2018**, *2018*, 1753480. [[CrossRef](#)] [[PubMed](#)]
5. Pranata, Y.D.; Wang, K.C.; Wang, J.C.; Idram, I.; Lai, J.Y.; Liu, J.W.; Hsieh, I.H. Deep learning and SURF for automated classification and detection of calcaneus fractures in CT images. *Comput. Methods Programs Biomed.* **2019**, *171*, 27–37. [[CrossRef](#)] [[PubMed](#)]
6. Aghnia Farda, N.; Lai, J.Y.; Wang, J.C.; Lee, P.Y.; Liu, J.W.; Hsieh, I.H. Sanders classification of calcaneal fractures in CT images with deep learning and differential data augmentation techniques. *Injury* **2020**. [[CrossRef](#)] [[PubMed](#)]
7. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
8. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
9. Santin, M.; Brama, C.; Théro, H.; Ketheeswaran, E.; El-Karoui, I.; Bidault, F.; Gillet, R.; Gondim Teixeira, P.; Blum, A. Detecting abnormal thyroid cartilages on CT using deep learning. *Diagn. Interv. Imaging* **2019**, *100*, 251–257. [[CrossRef](#)] [[PubMed](#)]
10. Kamiya, K.; Ayatsuka, Y.; Kato, Y.; Fujimura, F.; Takahashi, M.; Shoji, N.; Mori, Y.; Miyata, K. Keratoconus detection using deep learning of colour-coded maps with anterior segment optical coherence tomography: A diagnostic accuracy study. *BMJ Open* **2019**, *9*, 1–7. [[CrossRef](#)] [[PubMed](#)]
11. Nasrullah, N.; Sang, J.; Alam, M.S.; Mateen, M.; Cai, B.; Hu, H. Automated lung nodule detection and classification using deep learning combined with multiple strategies. *Sensors* **2019**, *19*, 3722. [[CrossRef](#)] [[PubMed](#)]
12. Roth, H.R.; Lu, L.; Farag, A.; Shin, H.C.; Liu, J.; Turkbey, E.B.; Summers, R.M. Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. *Lect. Notes Comput. Sci.* **2015**, *9349*, 556–564. [[CrossRef](#)]
13. Lakhani, P.; Sundaram, B. Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* **2017**, *284*, 574–582. [[CrossRef](#)] [[PubMed](#)]
14. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
15. Xie, Q.; Luong, M.T.; Hovy, E.; Le, Q.V. Self-training with noisy student improves imagenet classification. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10684–10695.
16. Kolesnikov, A.; Beyer, L.; Zhai, X.; Puigcerver, J.; Yung, J.; Gelly, S.; Houlsby, N. Big Transfer (BiT): General Visual Representation Learning. *arXiv* **2019**, arXiv:1912.11370.
17. Xie, C.; Tan, M.; Gong, B.; Wang, J.; Yuille, A.L.; Le, Q.V. Adversarial examples improve image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 816–825. [[CrossRef](#)]
18. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning Transferable Architectures for Scalable Image Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8697–8710.
19. Touvron, H.; Vedaldi, A.; Douze, M.; Jégou, H. Fixing the train-test resolution discrepancy. *arXiv* **2019**, arXiv:1906.06423.
20. Liu, C.; Zoph, B.; Neumann, M.; Shlens, J.; Hua, W.; Li, L.J.; Fei-Fei, L.; Yuille, A.; Huang, J.; Murphy, K. Progressive Neural Architecture Search. *Lect. Notes Comput. Sci.* **2018**, *11205 LNCS*, 19–35. [[CrossRef](#)]
21. Wang, G.; Sun, Y.; Wang, J. Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning. *Comput. Intell. Neurosci.* **2017**, *2017*. [[CrossRef](#)] [[PubMed](#)]
22. Ding, Y.; Sohn, J.H.; Kawczynski, M.G.; Trivedi, H.; Harnish, R.; Jenkins, N.W.; Lituiev, D.; Copeland, T.P.; Aboian, M.S.; Aparici, C.M.; et al. A deep learning model to predict a diagnosis of Alzheimer disease by using 18 F-FDG PET of the brain. *Radiology* **2019**, *290*, 456–464. [[CrossRef](#)] [[PubMed](#)]
23. Zhou, Q.Q.; Wang, J.; Tang, W.; Hu, Z.C.; Xia, Z.Y.; Li, X.S.; Zhang, R.; Yin, X.; Zhang, B.; Zhang, H. Automatic detection and classification of rib fractures on thoracic ct using convolutional neural network: Accuracy and feasibility. *Korean J. Radiol.* **2020**, *21*, 869–879. [[CrossRef](#)] [[PubMed](#)]
24. Chilamkurthy, S.; Ghosh, R.; Tanamala, S.; Biviji, M.; Campeau, N.G.; Venugopal, V.K.; Mahajan, V.; Rao, P.; Warier, P. Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study. *Lancet* **2018**, *392*, 2388–2396. [[CrossRef](#)]
25. Tomita, N.; Cheung, Y.Y.; Hassanpour, S. Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. *Comput. Biol. Med.* **2018**, *98*, 8–15. [[CrossRef](#)] [[PubMed](#)]