Classification of Asbestosis in CT Imaging Data Using Convolutional LSTM

Doohyung Kim

Korea University

Jun-Pyo Myong

Catholic University of Korea

Sung Won Han (✉ swhan@korea.ac.kr)

Korea University

Research Article

Keywords: Long-term Recurrent Convolution Networks (LRCN), CT, asbestos-related lung diseases (ARDs)

Posted Date: August 24th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-816493/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Classification of Asbestosis in CT Imaging Data Using Convolutional LSTM

DooHyung Kim^{1,+}, Jun-Pyo Myong^{2,+}, and Sung Won Han^{1,*}

¹School of Industrial and Management Engineering, Korea University, Seoul 02841, South Korea

²Department of Occupational and Environmental Medicine, Seoul St. Marys Hospital, College of Medicine, The Catholic University of Korea, Seocho-gu, Seoul 06591, Republic of Korea

*corresponding. swhan@korea.ac.kr

⁺these authors contributed equally to this work

ABSTRACT

Asbestosis is a disease with a high rate of mismatch between the readers. In this study, we use a deep learning framework to develop a model that can check the presence and lesion of asbestosis using medical CT data. The data was collected from 469 patients who had been tested for asbestosis. Out of the 469 patients, 284 were tested negative for asbestosis, while 185 were asbestosis positive. A CT image of a supine positioned lung setting was used. This study sought to solve the problem of image classification of CT data through Convolutional LSTM. Long-term Recurrent Convolution Networks (LRCN), a model that uses video format data as input, has been applied in this study. The model showed 83.3% Accuracy, 81.578% true positive and 86% true negative. In addition, a model that can verify validity by assisting a specialist with a Grad-CAM that can visualize the judgment was also developed. We hope that the model of this study will be able to work with specialists by acting as a judgment assistant.

1 Introduction

Asbestos has been widely used for the purpose of insulation, anti-abrasion, and anticorrosive because of its mechanical, chemical, and fire-proofing characteristics as well as its cheap cost. However, there are also significant public health problems. Inhaled asbestos fiber causes chronic irritation on the respiratory bronchiole and its structures. These chronic irritations were likely to introduce health outcomes such as asbestos, lung cancer, and any other pleural diseases (pleural plaque, mesothelioma, and etc.). For this reason, the government is banning the usage of materials as building materials. However, the long latent periods for developing asbestos-related lung diseases prevent physicians and researchers from estimating the causation of specific diseases with previous exposure to asbestos. Previous literature had shown that several radiologic findings were key features for asbestosis vs. idiopathic pulmonary lung diseases.^{1,2} The persistent debates on key features of asbestosis are ongoing. The Korean government compensates those victims with asbestos-related lung diseases (ARDs) due to previous asbestos exposure from occupational and environmental sources. ARDs were not prevalent. Because of few experts with clinical or radiological diagnosis experience, there are several problems in the judgment of ARDs.

Recently, however, a model that can assist a specialist by in the use of deep learning has been developed to show the possibility of using it as a specialist judgment³. Although CNN is used a lot for image classification problems, medical CT data are not appropriate for applying normal CNN to CT data because there are multiple images per patient. To apply data to deep learning models, the number of images must be the same, but CT images differ from patient to patient. The typical reason is that each patient has a different lung height, length, and width. Therefore, it is necessary to match the number of CT images.

In this study, CT data were collected in video format and had the same frame to match the number of images. For the classification of asbestosis, the deep learning structure (DenseNet)⁴, which showed excellent performance in image classification, was used in this study. The structure of DenseNet was changed according to this study, and Convolutional LSTM was used to combine LSTM with CNN. Through this, a model was created that could help determine asbestosis, which even specialists find difficult to find. Using Conv LSTM, experiments were conducted to verify various hypotheses to find the optimal conditions. In the medical field, judgment errors are directly linked to the patients health. Therefore, it is difficult for a specialist to trust a deep learning model just because the accuracy is high without knowing the basis of the judgment. This causes the utilization of deep learning in real clinical trials to be reduced. For the deep learning model to be used as an aid for determining sclerosis, the following must be satisfied, as noted in the preceding study: first, a specialist who uses the asbestos lung classification model should have a basis for judgment of the model. Second, it is necessary to inform what characteristics

of the data entered into the model have affected the results⁵.

Thus, in this study, the Gradient-Weighted Class Activation Mapping (Grad-CAM)⁶, developed for the visual interpretation of CNN Model, was used to visualize the basis of the model judgment. In addition, this study did not use the last feature on CNN, which is commonly used by Grad-CAM. The model was built in collaboration with specialists to show higher performance and better visualization of CT data. This allows a more effective visualization of the results of the learned model with which parts of the entered asbestos lung data were responded to. The purpose of the model using deep learning was to enable the interpretation of the basis for judging asbestosis. In addition, comparisons with Grad-CAM visualizations produced in general structures showed that the models created in this study were better models for visualization of asbestosis. Through this, we wanted to demonstrate that the model proposed in this study could be more helpful than the existing method in clinical practice. The present study used about 60,000 CT data taken from about 578 patients collected through Catholic University St. Marys Hospital. All the data were obtained from individual patients, and the experts verification ensured the reliability of the correct labels in the data. In practice, it was determined that filtering out abnormalities were the most important part of the clinical trial. Therefore, the focus of this study was on increasing true negative rather than true positive in confusion matrix.

The composition of this paper is as follows. Chapter 2 introduces a prior study using CT to detect lung abnormalities. Chapter 3 introduces the models and techniques used in this study. Chapter 4 introduces the data set and the pretreatment process. Chapter 5 describes the learning process and evaluation methods. Chapter 6 explains the results of the experiment, and Chapter 7 describes the conclusion of the paper and the direction of development.

2 Related Works

Figure 1. (a) represents the layer of the Convolutional Neural Network. (b) represents a module in the Long-Short Term Memory (LSTM). (c) represents the appearance of the Convolutional LSTM.

There is no dictionary paper analyzing this disease using deep learning, and this is the first research. Therefore, in the related chapter in the paper, I would like to mention the research that deep learning methodology used in computer vision and detects lung abnormalities.

Convolutional Neural Network (CNN)⁷ is one of the most popular deep learning techniques in the field of imagery. Vanilla

CNN consists of Convolutional Layer, Pooling Layer, and Fully Connected Layer. Convolutional Layer finds the characteristics of the image through the filter, as shown in Fig.1 (a). Proceed with convolution operation to find the image. The symbol $\otimes$ means convolution operation. Pooling Layer plays a role in reducing large amounts of computation because it can capture features that are strongly expressed in the image. Fully Connected Layer means performing Deep Neural Network operations after CNN to flatten features.

Long-short Term Memory (LSTM)⁸ is one of the methods of the Recurrent Neural Network (RNN) and is a model studied to solve the problem of long-Term Dependencies of RNN. The LSTM is divided into several gates. Detailed illustrations can be found in Fig.1 (b). The Input Gate is a gate for remembering current information. Delete Gate is a gate for deleting previously entered information. Cell State is called a long-term state in the LSTM. The C_{t-1} at the previous time, $t-1$, and the purge gate are multiplied by element. This indicates how much the previous Cell State will be preserved. Finally, there is the output gate and hidden state. The hidden state is sometimes referred to as the short-term state, where the value of the long-term state passes through the tanh function and becomes a value between -1 and 1, which is computed with the value of the output gate, resulting in a filter effect. Values in the short-term state are also directed to the output layer. Finally, convolutional LSTM⁹ refers to the insertion of a three-dimensional tensor rather than a one-dimensional tensor input to the existing LSTM. In this study, a three-dimensional Tensor refers to an image. As you can see from Fig.1 (c), you can see that the image performs a convolution operation and passes information to the next time.

CT data are now available for convolution of the data set itself through learning spatial features with 3D convolution networks¹⁰. The study wanted to create a model that uses CT data to classify lung abnormalities using 3D convolution¹¹. However, the results of fine tuning using medical imaging data were not good. To solve this problem, only lung abnormalities were isolated from CT data through watershed operation, and the nodule data of the separated lungs were learned through U-Net. The study conducted 3D convolution of nodular data of the learned lungs¹². Recently, a deep learning model was studied in Korea that detecting medical lesion using a modified GoogLeNet by combining CT data¹³.

However, in the case of deep learning models, there is a disadvantage that it is difficult for people to grasp the criteria for judging models. Therefore, research was actively conducted to find out the criteria for judgment of deep learning. For CNN, Grad-CAM⁶ is one of the most widely known methods and has eliminated the structural limitations of CAM. Many medical imaging analyses now use Grad-CAM to identify areas of response that affect classification. A study was also conducted to interpret the basis for judging abnormalities classification models. The study applied Grad-CAM to CT data using features learned through the model and provided evidence that the image was considered abnormalities¹⁴. To our knowledge only, no case has been found in which Grad-CAM is applied to visualize all characteristic maps to detect and classify asbestosis. In most studies, CT data were used as a single image, and for studies using CT data sets, such as 3D Convolution, papers that provided evidence for judgment, such as Grad-CAM, could not be verified.

Therefore, in this study, we wanted to develop a supervised binary classification model that can use CT data, which do not need to set up bounding box in advance and can more accurately classify asbestosis with the only label. Furthermore, Grad-CAM was used to provide evidence that was determined to be asbestosis.

3 Model Architecture

In this study, the Conv LSTM structure was used to use CT data sets. To implement the structure of these models, the structure of the preceding study Long-term Recurrent Convolution Networks for Visual Recreation and Description¹⁰ was used. LRCN, as shown in Fig.1, will result in each image being classified via CNN and LSTM. In this study, the backbone of LRCN was created using one of CNNs models, Densely Connected Converged Network (DenseNet)⁴. Because DenseNet has a skip-connection structure, information on the input part of the model is not lost and can be delivered well to the end, and even when back-propagation operations are performed, the computation of the last part of the model is delivered well to the front part, alleviating the problem of gradient-disappearance, which is one of the problems of deep learning, showing high performance. The DenseNet model expected CT data for asbestosis, where the location of the outbreak is at the bottom of the lung, to be able to detect movement of the data changing as it moves downward.

Because CT images vary from patient to patient, it was necessary to size the input data equally to use the deep learning model. After equalizing the amount of data, the images were studied through the Conv LSTM. As Fig.2 each image passes through CNN and updates the feature. Unlike conventional DenseNet, the process after the last fully-connected layer was removed to connect with LSTM. In this case, if you pass the DenseNet, you will see the feature at the end, which was flattened and inserted into LSTM. After this process for all images, the average of the values in the last layer was measured to determine whether or not asbestosis was present. It also sought to obtain an optimal model for classifying CT data. After creating and learning from low-complexity models to high-complex models in turn, the scores for validation data sets and training data sets were compared. Through this, we found the best layer depth and hyperparameter at the balance point, where the issue of overfitting and underfitting is minimal. This produced an optimal Conv LSTM model for asbestosis data. Through this, an optimal Conv LSTM model for asbestosis data was created. The hyperparameter of DenseNet used in this study is

Visual-LRCN for Asbestosis

Figure 2. Model Architecture(Visual-LRCN for Asbestosis). Conduct a Convolutional LSTM from each image to predict the class. Before flattening the feature map for convolutional LSTM, extract the feature map and proceed with Grad-CAM. This enables visualization

the hyperparameter that showed the best performance in the validation data set: Learning Rate = 0.0005, Batch Size = 32, Backbone = DenseNet161, Epoch = 600, Latend Dimension = 512, LSTM Layer = 2, Hidden Size = 256 was used with NVIDIA V100 16GB $\times$ 8.

Grad-CAM used in this study is a further development of Class Activation Mapping (CAM), a typical methodology for visualizing Convolutional Neural Networks. Grad-CAM obtains the partial differential value between the feature map and the last classification layer that you want to visualize and then averages it by channel to obtain the value. It is a method of interpreting important parts of the feature map along with the original data by multiplying this value by the characteristic map you want to visualize and then averaging it in the channel direction and extending the obtained value to match the original data size. CAM can only be used by Global Average Pools at the end of the classification model, but Grad-CAM can visualize the character map of any desired area. This can help you visualize each layer of a characteristic map to understand how the input value of the synthetic product neural network affects the final classification and to help you identify model errors. Using the Grad-CAM model in this study, using the last Feature on CNN did not show very good performance. To show higher performance and better visualization of lung asbestos data, Grad-CAM was applied to all layers to determine which layer best represents asbestosis symptoms.

In this study, we would like to change the structure so that Grad-CAM can be used in Conv LSTM and call the structure that finds optimal asbestos parameters for visual-LRCN. In this study, the training visual-LRCN for asbestosis model removed the full-connected layer included in the basic DenseNet structure and changed the structure to allow Grad-CAM to be performed before the last feature is flattened and delivered to LSTM.

4 Data sets

The data used in this study are CT data of about 578 patients collected at St. Marys Hospital of the Catholic University of Korea. In this study, reliable data were used after verification by skilled specialists in St. Marys Hospital. Our study was approved by the IRB of the Seoul St. Marys Hospital, College of Medicine, the Catholic University of Korea (KC17ENSI0379). Those CT data were extracted without any information on the CT scan. CT scans were labeled as asbestosis with reading results or not. A supine lung setting was utilized in this model development. The data labels were classified as nonrecognition, suspicious, initial, and progressive, and three categories were used except suspected. There is no asbestosis in the CT scan, which is

classified as non-asbestosis, and the number of data is 284. The suspicious cases that are not clinically defined and providing confusion in the decision are excluded ($n = 99$). The initial and progressive types are classified as asbestosis, and the number of data is 185. In this study, we wanted to focus on the classification of two categories.

The results of CT imaging vary from patient to patient. First of all, parts that do not include the lungs were removed with a specialist. In addition, it was necessary to match the same length for each patient to use as an input for deep learning. To this end, the same number of inputs was extracted using the methods shown in equation (1) later.

$$L_k = \text{index}[F(N)/K * (k)] \quad (1)$$

L is the image to be inserted at the k th, and N is the total number of CT data per patient. K is the number of images to be extracted from CT data. $F(N)$ is a function that extracts the lung image below first in extracting the image. The same number of the image was extracted with index from this equation. In this study, the data collected were transformed into a model suitable for learning. Because the base model, the input format of the LRCN, was video, the CT data, which is the image format, was converted to a video format of 1 Frame per Second (FPS) to convert it into the video. FPS refers to the number of frames moving per second, and a typical video has frames of more than 30. However, since each CT data is meaningful, it is set to 1FPS so that one picture can be seen per second. To fit the same length, 15 frames were extracted from each video. Extractions should be made at equal intervals and were extracted using the equation (1) to give weight to the lower part of the lung. Because video is a continuous flow of image graphs, if a constant frame is extracted, different patients can be matched to the same length.

5 Training

For model train and test, 469 total data were divided into 8:2 and separated into the train set and test set. Totally 80% classified as train set was divided again and 8:2 and 80% were used as a train set and 20% as a validation set. Therefore, 64% of the total data were used as a train set and 16% as a validation set, and the remaining 20% was extended as test data set. The number of train sets is 369, the number of validation sets is 93, and the number of test sets is 116. Each data was selected randomly, and the data splitting was performed by fixing the random seed for fair model comparison. Adam optimizer was used to study the model, and the initial value of the reading rate, which is the parameter of the optimizer, was specified as 0.005. Since the data are images with 15 frames each, they are converted to 15 images before entering the model. These transformed data are learned in the model with 32 patient sets in batch size. The loss function used relative cross entropy. Category cross entropy is most widely used method is the loss function of the classification model, by comparing the predicted values of the model with the correct label. In this study, the model was studied under the above conditions in order to find the optimum parameters.

In this study, the performance of asbestosis classification was compared using the accuracy and confusion matrix for model evaluation. The ratio of non-asbestosis to asbestosis data used in this study is not a perfect 1:1 ratio. Thus, comparisons of models with accuracy, and evaluation indicator that has a significant impact on categories that account for a large percentage, are not correct. Therefore, the performance was compared in this study using a confusion matrix that could reflect the ratio of the data. False-negative rate, the rate at which asbestosis data are actually asbestosis but classified as non-asbestosis, is a very important indicator in the medical field, and results were compared for each model. The information for each indicator is shown in Table 1.

Confusion matrix		Predicted	
		Positive	Negative
Actual	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Table 1. Confusion matrix

The study was repeated to find the best parameters in each model structure, and the test set was evaluated using the parameters that showed the best last name in the validation set. For comparison of each model, visualization of the model was carried out so that people could compare the results with their own eyes.

To compare the basis of judgment of the model, the visualization of the basis of judgment was carried out using Grad-CAM for non-asbestosis and asbestosis data. Grad-CAM is well known. Grad-CAM, as is well known, can be used by inserting it into FCLayer. However, it is not possible to apply Grad-CAM because LRCN deploys all elements through convolutional LSTM after CNN. To solve this problem, we referred to the paper that studied it earlier and modified it to fit this research¹⁵. In this study, a verification was carried out with a specialist to verify that the visualization using Grad-CAM was well done.

In using Grad-CAM, the layer is set up, which usually uses the last layer to use the weight of the last layer. However, in this study, the performance of Grad-CAM was not satisfactory when using the last layer, and to solve this problem, a layer was set up with a specialist.

Model	Visual-LRCN for Asbestosis			C3D
Backbone	ResNet152	DenseNet161	DenseNet201	-
Loss	0.5198	0.4957	0.5502	0.4812
Accuracy	80.159%	80.952%	75.397%	73.834%
True Positive Rate	84.21%	78.94%	68.42%	35.664%
Treu Negative Rate	74%	84%	86%	89%

Table 2. Performance for each model

6 Results and Experiments

Table 2 is a table comparing the performance by the classification model for the test data set. The actual classification results for test data set for each model can be found in true positive and true negative. The model shown in Table 2 is a type of CNN model used as a backbone in LRCN used in this study, using ResNet152, DenseNet161, and DenseNet201^{4,16}. C3D conducted a study that applied a solution to three-dimensional data¹⁷. This model is used in most prior studies using CT data to find abnormalities¹⁸. The model used in the previous study, C3D, had an Acc of 73%, but the true positive showed a very low performance of 35%. Using the LRCN models conducted in this study, Acc could be found to be more than 75% in all cases, regardless of CNN models, but decided to choose between ResNet152 and DenseNet161 over the target of 80%. When choosing between these two, a specialist and a specialist decided to choose DenseNet161, which has a high true negative.

Figure 3. Grad-CAM Application Results. Images in the first row are images of Grad-CAM applied to CT data sets of patients diagnosed with asbestosis, and images in the second row are images of Grad-CAM applied to CT data sets of patients who have not been diagnosed with asbestosis

	15 pieces	20 pieces	25 pieces
Loss	0.4957	0.4727	0.5721
Accuracy	80.952%	83.333%	73.810%
True Positive	78.94%	81.578%	84.21%
True Negative	84%	86%	58%
This experiment used the parameters shown in Table 4			

Table 3. Results of Experiment the Number of Input Images with Visual-LRCN for Asbestosis

In this study, additional experiments were conducted after establishing hypotheses. The hypothesis is the result of the number of input images. We assumed that increasing the number of Input Images by more than 15 would result in better

results. In general, deep learning is known to perform better with more information. The results for the hypothesis can be seen in Table 3. The results showed that using 20 pictures showed better results than using 15 pictures(base). However, if you used 25 photos, you might see worse results. We concluded that the results were not good, as many parts were not related to the manifestation location of asbestosis.

The parameters and conditions finally determined are as shown in Table 4. Epoch is 600, Model uses LRCN, and LRCN is DenseNet161 + ConvLSTM. The number of images used as input is 20, and the learning rate is 0.0005. The last Dim parameter of ConvLSTM, 512 LSTM Layer, used two hidden size 256. Batch size is 32. The following parameters enabled us to achieve the target of 80% with 0.4727, Acc 83.333%, true positive 81.578%, and true negative 86% at test data set. The loss and accuracy of each epoch of the model are shown in Fig.4. The following are the results for Grad-CAM. If you look at Fig.3 images in the first row are images of Grad-CAM applied to CT data sets of patients diagnosed with asbestosis, and images in the second row are images of Grad-CAM applied to CT data sets of patients who have not been diagnosed with asbestosis. Looking at the images of the two patients, we can see that the colors inside the lungs are significantly different. This was done with a specialist, and the same result could be seen in almost every patient. Therefore, we could confirm that the results of the Grad-CAM application in this study were significant.

Figure 4. Accuracy and Loss of Result

Parameter	Value
Epoch	600
Model	DenseNet161
Number of Pictures	20 pieces
Learning Rate	0.0005
Latend Dim.	512
LSTM Layer	2
Hidden Size	256
Batch Size	32
Loss	0.4727
Accuracy	83.333%
True Positive	81.578%
True Negative	86%

Table 4. Final Parameters and Results.

7 Conclusion

In this study, CT data of patients were used to create a model that showed high performance in determining whether or not asbestosis was present. In the medical field, mistakes are very dangerous, and the cost of the redetermination is high. The study,

which proposed a highly sensitive model, could help specialists determine asbestos lung disease more efficiently. In addition, to overcome the limitations of this model that derive only non-asbestosis and asbestosis results, the model's shortcomings were supplemented by visualizing asbestosis lesions using Grad-CAM.

It is expected that this will serve as a basis for specialists to use models as decision-making tools in the treatment of asbestos. In this study, the research was carried out by using classification without using segmentation or object detection. This model can be used as long as it is classified without the need to label diseases such as asbestos lung disease that even specialists find difficult to detect lesions. This will be a great advantage when using this model at the actual medical site. Because, there is no need to label new data. This study used only data on asbestosis, but if data from a variety of diseases can be collected, a decision-making aid tool is expected to be available at various actual medical sites.

Acknowledgments

The authors of this manuscript have reviewed and approved its content and have all made substantive intellectual contributions to the work. Further, they all affirm that the article is original, is not under consideration by another journal, has not been previously reviewed or published. Informed consent was obtained from all participants or, if participants are under 16, from a parent and/or legal guardian A statement to confirm that all methods were carried out in accordance with relevant guidelines and regulations. The authors confirm that all methods were carried out in accordance with relevant guidelines and regulations.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2017R1C1B1004010). This research was also supported by National Research Foundation of Korea (NRF-2019R1F1A1060250) and Brain Korea 21 FOUR. This research was supported by Korea University Grant (K2107521, K2011531).

Author contributions

Kim, D.H. : Data Curation, Formal Analysis, Methodology, Writing-Original Draft, Visualization. Myong, J.P. : Data Curation, Formal Analysis, Methodology, Writing-Original Draft. Han, S.W. : Conceptualization, Methodology, Writing Review & Editing, Supervision.

References

1. Arakawa, H. *et al.* Asbestosis and other pulmonary fibrosis in asbestos-exposed workers: High-resolution ct features with pathological correlations. *Eur. Radiol.* **26**, 1485–1492 (2016).
2. Akira, M. & Morinaga, K. The comparison of high-resolution computed tomography findings in asbestosis and idiopathic pulmonary fibrosis. *Am. J. Ind. Medicine* **59**, 301–306 (2016).
3. Rajpurkar, P., Hannun, A. Y., Haghpahani, M., Bourn, C. & Ng, A. Y. Cardiologist-level arrhythmia detection with convolutional neural networks. *Clin. Orthop. Relat. Res.* **Abs/1707.01836** (2017). [1707.01836](#).
4. Huang, G., Liu, Z. & Weinberger, K. Q. Densely connected convolutional networks. *Clin. Orthop. Relat. Res.* **Abs/1608.06993** (2016). [1608.06993](#).
5. Holzinger, A., Biemann, C., Pattichis, C. S. & Kell, D. B. What do we need to build explainable AI systems for the medical domain? *Clin. Orthop. Relat. Res.* **Abs/1712.09923** (2017). [1712.09923](#).
6. Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization. In *IEEE International Conference on Computer Vision, ICCV, Venice, Italy, October 22-29, 2017*, 618–626, DOI: [10.1109/ICCV.2017.74](#) (IEEE Computer Society, 2017).
7. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
8. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural computation* **9**, 1735–1780 (1997).
9. Xingjian, S. *et al.* Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, 802–810 (2015).
10. Donahue, J. *et al.* Long-term recurrent convolutional networks for visual recognition and description. *IEEE Transactions on Pattern Analysis Mach. Intell.* **39**, 677–691, DOI: [10.1109/TPAMI.2016.2599174](#) (2017).
11. Alakwaa, W., Nassef, M. & Badr, A. Lung cancer detection and classification with 3d convolutional neural network (3d-cnn). *Int. J. Adv. Comput. Sci. Appl.* **8**, 409 (2017).

12. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th Proceedings, Part III*, vol. 9351 Navab N, Hornegger J, III, WMW, Frangi AF (eds) International Conference Munich, Germany, October 5-9, 2015 of *Lecture Notes in Computer Science*, 234–241, DOI: [10.1007/978-3-319-24574-T28](https://doi.org/10.1007/978-3-319-24574-T28) (Springer, 2015).
13. Han, S., Hwang, S. I. & Lee, H. J. The classification of renal cancer in 3-phase CT images using a deep learning method. *J. Digit. Imaging* **32**, 638–643, DOI: [10.1007/s10278-019-00230-2](https://doi.org/10.1007/s10278-019-00230-2) (2019).
14. Qin, R. *et al.* Fine-grained lung cancer classification from PET and CT images based on multidimensional attention mechanism. *Complex*. **2020**, 6153657:1–6153657:12, DOI: [10.1155/2020/6153657](https://doi.org/10.1155/2020/6153657) (2020).
15. Yamashita, N., Iwahashi, N., Nakano, S., Sakai, T. & Hamano, M. Visually explaining videos using recurrent neural networks with gradient-based localization. *80* **(6)**, 03 (2018).
16. 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, Las Vegas, NV, USA, June 27-30, 2016*, 770–778, DOI: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90) (IEEE Computer Society, 2016).
17. Tran, D., Bourdev, L. D., Fergus, R., Torresani, L. & Paluri, M. Learning spatiotemporal features with 3d convolutional networks. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, December, Santiago, Chile 7-13*, 4489–4497, DOI: [10.1109/ICCV.2015.510](https://doi.org/10.1109/ICCV.2015.510) (IEEE Computer Society, 2015).
18. Jung, H., Kim, B., Lee, I., Lee, J. & Kang, J. Classification of lung nodules in CT scans using three-dimensional deep convolutional neural networks with a checkpoint ensemble method. *BMC Medical Imaging* **18**, 48, DOI: [10.1186/s12880-018-0286-0](https://doi.org/10.1186/s12880-018-0286-0) (2018).