

AI-based image quality assessment in CT

Lars Edenbrandt^{1,2,3}, Elin Trägårdh^{4,5}, and Johannes Ulén⁶

¹Region Västra Götaland, Sahlgrenska University Hospital, Department of Clinical Physiology, Gothenburg, Sweden

²Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

³SliceVault AB, Malmö, Sweden

⁴Department of Clinical Physiology and Nuclear Medicine, Skåne University Hospital and Lund University, Malmö, Sweden

⁵Wallenberg Center for Molecular Medicine, Lund University, Malmö, Sweden

⁶Eigenvision AB, Malmö, Sweden

ABSTRACT

Medical imaging, especially computed tomography (CT), is becoming increasingly important in research studies and clinical trials and adequate image quality is essential for reliable results. The aim of this study was to develop an artificial intelligence (AI)-based method for quality assessment of CT studies, both regarding the parts of the body included (i.e. head, chest, abdomen, pelvis), and other image features (i.e. presence of hip prosthesis, intravenous contrast and oral contrast).

Approach: 1,000 CT studies from eight different publicly available CT databases were retrospectively included. The full dataset was randomly divided into a training ($n = 500$), a validation/tuning ($n = 250$), and a testing set ($n = 250$). All studies were manually classified by an imaging specialist. A deep neural network was then trained to directly classify the 7 different properties of the image.

Results: The classification results on the 250 test CT studies showed accuracy for the anatomical regions and presence of hip prosthesis in the interval 98.4% to 100.0%. The accuracy for intravenous contrast was 89.6% and for oral contrast 82.4%.

Conclusions: We have shown that it is feasible to develop an AI-based method to automatically perform a quality assessment regarding if correct body parts are included in CT scans, with a very high accuracy.

Keywords: Artificial Intelligence, machine learning, quality assessment, diagnostic imaging

1. INTRODUCTION

Medical imaging, especially computed tomography (CT), is becoming increasingly important in research studies and clinical trials. Large projects and trials could include hundreds or thousands of CT studies and adequate image quality is essential for reliable results. This is a particular concern in multi-center trials, which often provide detailed imaging guides that must be followed in order to correctly include patients. Problems related to imaging could lead to either exclusion of patients or false image data incorporated in study or trial results. A quality check of images selected for a study is therefore an important process. Today, this is performed manually. Often, the quality check must be performed promptly by the clinical research organization before the patient can be enrolled. In both clinical trials and in large retrospective studies, this could be tedious work. The quality check of images could ensure for example that:

- the correct part of the body is visible in the CT, e.g., the chest in a lung cancer study,
- CT artefacts are not present, e.g., hip prostheses causing artefacts which usually prevent a proper analysis of the prostate,

Corresponding author Lars Edenbrandt (E-mail: lars.edenbrandt@gu.se).

- the CT is acquired according to the study protocol regarding the use of intravenous or oral contrast.

Information about a CT study should ideally be described in the DICOM tags. However, experience shows that it is not possible to rely only on this information. This is true especially in research projects and clinical trials when important information in the DICOM tags could be deleted in the pseudonymization process.

The use of artificial intelligence (AI) to solve clinical problems has been intensely studied in recent times.¹ Deep learning, in particular, has gained attention as a method of obtaining complex information from medical images. AI could potentially be trained to help with image quality assessment and could be an important tool in this important, otherwise manual, task.

The aim of this study was to develop an AI-based method for quality assessment of CT studies, both regarding the parts of the body included (i.e. head, chest, abdomen, pelvis), and other image features (i.e. presence of hip prosthesis, intravenous contrast and oral contrast).

2. METHODS

2.1 Patients

We retrospectively included 1,000 CT studies from eight publicly available CT databases, see Table 1.

Database	Number of images	References
C4KC-KiTS	204	2–4
ACRIN 6668	203	2, 5, 6
CT Lymph Nodes	176	2, 7–9
CT-ORG	117	2, 10–12
NSCLC-Radiomics	115	2, 13, 14
Task 07 Pancreas	100	15
Task 03 Liver	52	15
Anti-PD-1 Immunotherapy Lung	33	2, 16
Total	1,000	

Table 1. The number of images selected from each publicly available database.

Before training the model, the full image set of 1,000 CT studies was randomly divided into a training ($n = 500$), a validation/tuning ($n = 250$), and a testing set ($n = 250$). The test set was reserved for model evaluation.

2.2 Manual classification/Ground truth definition

All CT studies were classified by a nuclear medicine specialist experienced in hybrid imaging. Each case was classified based on the presence of the following seven features:

- *Head* The cranium is visible at least partly. Head is not present if only part of the mandible is visible.
- *Chest* The lungs are visible. Only very minor parts may be missing.
- *Abdomen* Main parts of liver, spleen, and the kidneys are visible.
- *Pelvis* The hip bones are visible.
- *Hip prosthesis* Uni- or bilateral hip prosthesis including implants for fixation of hip fractures.
- *Intravenous (IV) contrast* Signs of intravenous contrast including different phases.
- *Oral contrast*: Signs of oral contrast including different phases.

An overview of the distribution of the different classes in the dataset is given in Table 2.

Class	Positive count	Negative count
Head	256	744
Chest	603	397
Abdomen	903	97
Pelvis	702	298
Hip prosthesis	25	975
Intravenous contrast	256	744
Oral contrast	422	578

Table 2. Positive and negative examples for each class in the dataset.

2.3 AI tool

The AI tool consists of a 3D-ResNet,¹⁷ a deep neural network designed for classification of 3D images. The network have an input shape of $110 \times 110 \times 110 \times 1$ pixels with 7 output channels each with its own sigmoid activation. Each output channel represents one of the classes defined in Section 2.2.

Many CT images contain quite a lot of air, which is not helpful for classification. In order to remove air, the images are smoothed using a Gaussian kernel with standard deviation 5mm^3 . An axis-aligned bounding box is then fitted to all pixels with Hounsfield unit (HU) above -800 in the smoothed image. The original image is then cropped to this bounding box.

The cropped CT images are pre-processed by clamping the HU values to the range $[-1000, 3000]$ and then normalized to $[-1, 1]$. Furthermore, the CT volumes are re-sized to resolution $5 \times 6 \times 12$ mm (or the smallest possible pixel shape with the same aspect ratio making the full image fit) and placed in the middle of the input volume.

2.3.1 Sampling

The classes are quite imbalanced as seen in Table 2. In order to sample uncommon examples more often each image i is sampled proportional to a weight w_i defined as:

$$w_i = \sum_{\ell \in L} \begin{cases} 1/p_\ell & \text{if } \ell \text{ is true for image } i, \\ 1/n_\ell & \text{if } \ell \text{ is false for image } i, \end{cases} \quad (1)$$

where L is the set of labels, p_ℓ the total number of positive examples, and n_ℓ the total number of negative examples for label ℓ . p_ℓ and n_ℓ are calculated individually for the training and validation sets.

2.3.2 Training

Binary cross-entropy is used as loss function and the network is optimized using the ADAM optimizer¹⁸ with Nesterov momentum and an initial learning rate of 1×10^{-5} . Each training and validation epoch consists of 2,000 samples and 400 samples respectively. If the validation loss has not improved for 10 epochs the learning rate is halved until it reaches a minimum value of 1×10^{-8} . The training stops when validation loss has not improved for 20 epochs.

During training the images are augmented using rotations (-0.1 to 0.1 radians), scaling (-10 to 10%) and an intensity shift of (-100 to $+100$ HU).

3. RESULTS

The classification results on the test set of is presented in Table 3. The accuracy for the anatomical regions and presence of hip prosthesis were 98.4% to 100.0%. The accuracy for intravenous contrast was 89.6% and for oral contrast 82.4%. Figure 1 and 2 show patient examples with correct and non-correct classifications.

The execution time on a high-end desktop computer with a Nvidia RTX 3090 graphics card is about 2 seconds per image.

Classification task	TP	TN	FP	FN	Accuracy
Head	70	176	0	4	98.4%
Chest	146	104	0	0	100.0%
Abdomen	222	24	0	4	98.4%
Pelvis	167	80	0	3	98.8%
Hip prosthesis	6	244	0	0	100.0%
Intravenous contrast	149	75	13	13	89.6%
Oral contrast	71	135	15	29	82.4%

Table 3. Result for the 250 test CT studies. True positive (TP), True negative (TN), False positive (FP), False negative (FN), and accuracy.



Figure 1. Example of a correctly classified image from ref. 5. Head, chest, abdomen, pelvis, hip prosthesis, and oral contrast were present. Intravenous contrast was not present.

4. DISCUSSION

In this study we have shown that it is feasible to develop an AI-based tool to automatically check that the correct body parts are visible in the CT studies, with a very high accuracy. The AI-based method was also able to accurately detect hip prosthesis even though the number of positive cases in the training and validation sets were limited ($n = 19$).

A limitation of this study was that the AI-based tool only made a classification regarding presence of contrast or not. Many different phases of contrast enhancement exist,¹⁹ typically early arterial phase (15–25 s post injection), late arterial phase (30–40 s post injection), hepatic or late portal venous phase (70–90 s post injection), nephrogenic phase (85–120 s post injection) and excretory or delayed phase (5–10 min post injection). No clearly defined times post injection of the contrast agent exist, but with a large number of images with different contrast phases in the training group, it would probably be possible to train an AI-method to categorize the contrast phase in more detail than we did in this study. Other potential problems related to intravenous contrast is different amounts of contrast agent administered, for example reduced doses in patients with kidney disease. Problems with oral contrast for this type of task include different timings of contrast as well as different contrast agents (for example barium or iodine-based agents). A more comprehensive classification of contrast would most likely require a much larger training set. Some of the false negative cases of our test set represented very late intravenous phases with low contrast in the aorta but contrast in the kidneys or urinary bladder (Figure 2). This

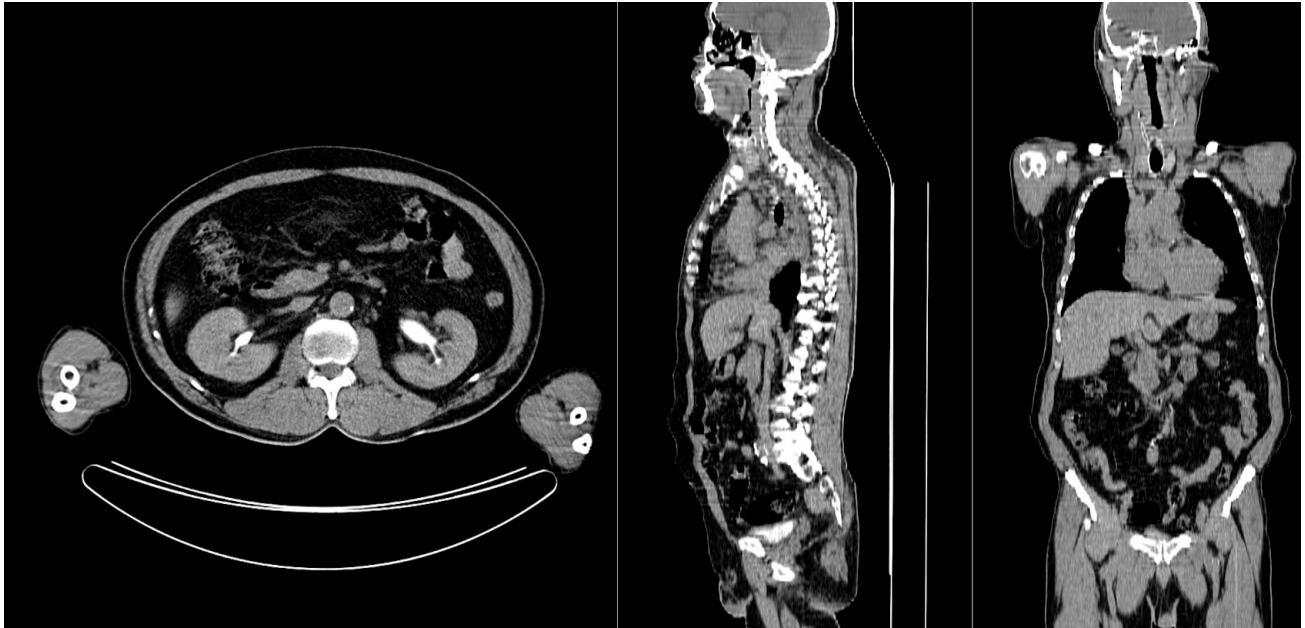


Figure 2. Example of image with a miss-classification from ref. 5. Excretory intravenous contrast phase is present, but not detected by AI. All other classifications were correct.

type of cases was not common in the training set. Also the appearance of oral contrast on CT showed substantial variation. In most cases contrast was clearly visible in the small bowel. In other cases only the stomach or colon showed contrast.

Medical imaging is often a key asset in clinical trials, as it can provide efficacy evaluation and safety monitoring.²⁰ It is also often used as screening for eligible patients to include. Medical imaging can also improve clinical trial efficacy and reduce the time to complete a specific trial, by offering imaging biomarkers that can act as a surrogate endpoint. In order to do so, good image quality is crucial and therefore it is necessary to monitor image quality throughout different stages of a study. A step in the quality assessment could be to determine if correct body parts are included and if the images contain contrast agent or not. Further development of automated image quality assessment could also include image properties such as noise level and patient motion. Evaluation by a human observer is both time consuming and subjective. AI-based tools could help minimize both issues.

5. CONCLUSIONS

We have shown that it is feasible to develop an AI-based method to automatically perform a quality assessment regarding if correct body parts are included in CT scans, with a very high accuracy.

6. ACKNOWLEDGMENT

We would like to thank Måns Larsson and Olof Enqvist for fruitful discussions regarding this study and closely related topics.

REFERENCES

- [1] D. Jarrett, E. Stride, K. Vallis, and M. J. Gooding, “Applications and limitations of machine learning in radiation oncology,” *The British Journal of Radiology* **92**, p. 20190001, Aug. 2019.
- [2] K. W. Clark, B. A. Vendt, K. E. Smith, J. B. Freymann, J. S. Kirby, P. Koppel, S. M. Moore, S. R. Phillips, D. R. Maffitt, M. Pringle, L. Tarbox, and F. W. Prior, “The cancer imaging archive (tcia): Maintaining and operating a public information repository,” *J. Digital Imaging* **26**(6), pp. 1045–1057, 2013.

- [3] N. Heller, N. Sathianathen, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, J. Dean, M. Tradewell, A. Shah, R. Tejpaul, Z. Edgerton, M. Peterson, S. Raza, S. Regmi, N. Papanikolopoulos, and C. Weight, “C4kc kits challenge kidney tumor segmentation dataset,” 2019.
- [4] N. Heller, F. Isensee, K. H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han, G. Yao, Y. Gao, Y. Zhang, Y. Wang, F. Hou, J. Yang, G. Xiong, J. Tian, C. Zhong, J. Ma, J. Rickman, J. Dean, B. Stai, R. Tejpaul, M. Oestreich, P. Blake, H. Kaluzniak, S. Raza, J. Rosenberg, K. Moore, E. Walczak, Z. Rengel, Z. Edgerton, R. Vasdev, M. Peterson, S. McSweeney, S. Peterson, A. Kalapara, N. Sathianathen, N. Papanikolopoulos, and C. Weight, “The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge,” *Medical Image Analysis* **67**, p. 101821, 2021.
- [5] P. Kinahan, M. Muzi, B. Bialecki, B. Herman, and L. Coombs, “Data from the acrin 6668 trial nsclc-fdg-pet,” 2019.
- [6] M. Machtay, F. Duan, B. A. Siegel, B. S. Snyder, J. J. Gorelick, J. S. Reddin, R. Munden, D. W. Johnson, L. H. Wilf, A. DeNittis, *et al.*, “Prediction of survival by [18f] fluorodeoxyglucose positron emission tomography in patients with locally advanced non-small-cell lung cancer undergoing definitive chemoradiation therapy: results of the acrin 6668/rtog 0235 trial,” *Journal of clinical oncology* **31**(30), p. 3823, 2013.
- [7] H. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers, “A new 2.5 d representation for lymph node detection in ct,” 2015.
- [8] H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers, “A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations,” in *International conference on medical image computing and computer-assisted intervention*, pp. 520–527, Springer, 2014.
- [9] A. Seff, L. Lu, K. M. Cherry, H. R. Roth, J. Liu, S. Wang, J. Hoffman, E. B. Turkbey, and R. M. Summers, “2d view aggregation for lymph node detection using a shallow hierarchy of linear classifiers,” in *International conference on medical image computing and computer-assisted intervention*, pp. 544–552, Springer, 2014.
- [10] B. Rister, K. Shivakumar, T. Nobashi, and D. L. Rubin, “Ct-org: A dataset of ct volumes with multiple organ segmentations,” 2019.
- [11] B. Rister, D. Yi, K. Shivakumar, T. Nobashi, and D. L. Rubin, “Ct organ segmentation using gpu data augmentation, unsupervised labels and iou loss,” 2018.
- [12] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C. Fu, X. Han, P. Heng, J. Hesser, S. Kadoury, T. K. Konopczynski, M. Le, C. Li, X. Li, J. Lipková, J. S. Lowengrub, H. Meine, J. H. Moltz, C. Pal, M. Piraud, X. Qi, J. Qi, M. Rempfler, K. Roth, A. Schenk, A. Sekuboyina, P. Zhou, C. Hülsemeyer, M. Beetz, F. Ettliger, F. Grün, G. Kaissis, F. Lohöfer, R. Braren, J. Holch, F. Hofmann, W. H. Sommer, V. Heinemann, C. Jacobs, G. E. H. Mamani, B. van Ginneken, G. Chartrand, A. Tang, M. Drozdal, A. Ben-Cohen, E. Klang, M. M. Amitai, E. Konen, H. Greenspan, J. Moreau, A. Hostettler, L. Soler, R. Vivanti, A. Szeskin, N. Lev-Cohain, J. Sosna, L. Joscowicz, and B. H. Menze, “The liver tumor segmentation benchmark (lits),” *CoRR abs/1901.04056*, 2019.
- [13] B. Rister, K. Shivakumar, T. Nobashi, and D. L. Rubin, “Ct-org: A dataset of ct volumes with multiple organ segmentations,” 2019.
- [14] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebbers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin, “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach,” *Nature Communications* **5**, June 2014.
- [15] M. Antonelli, A. Reinke, S. Bakas, K. Farahani, AnnetteKopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers, B. van Ginneken, M. Bilello, P. Bilic, P. F. Christ, R. K. G. Do, M. J. Gollub, S. H. Heckers, H. Huisman, W. R. Jarnagin, M. K. McHugo, S. Napel, J. S. G. Pernicka, K. Rhode, C. Tobon-Gomez, E. Vorontsov, H. Huisman, J. A. Meakin, S. Ourselin, M. Wiesenfarth, P. Arbelaez, B. Bae, S. Chen, L. Daza, J. Feng, B. He, F. Isensee, Y. Ji, F. Jia, N. Kim, I. Kim, D. Merhof, A. Pai, B. Park, M. Perslev, R. Rezaiifar, O. Rippel, I. Sarasua, W. Shen, J. Son, C. Wachinger, L. Wang, Y. Wang, Y. Xia, D. Xu, Z. Xu, Y. Zheng, A. L. Simpson, L. Maier-Hein, and M. J. Cardoso, “The medical segmentation decathlon,” 2021.
- [16] M. Patnana, S. Patel, and A. S. Tsao, “Data from anti-pd-1 immunotherapy lung,” 2019.

- [17] C. Yang, A. Rangarajan, and S. Ranka, “Visual explanations from deep 3d convolutional neural networks for alzheimer’s disease classification,” *CoRR* **abs/1803.02544**, 2018.
- [18] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014.
- [19] D. Fleischmann and A. Kamaya, “Optimal vascular and parenchymal contrast enhancement: The current state of the art,” *Radiologic Clinics of North America* **47**, pp. 13–26, Jan. 2009.
- [20] P. Murphy, “Imaging in clinical trials,” *Cancer Imaging* **10**(1A), pp. S74–S82, 2010.