Diagnostic and Interventional Imaging (2019) 100, 177—183

ORIGINAL ARTICLE / *Research and new developments*

Convolutional neural network evaluation of over-scanning in lung computed tomography

M. Colevray[a], VM. Tatard-Leitman[b,c], S. Gouttard[a], P. Douek[a,c], L. Boussel[a,b,*]

[a] *Department of radiology, hôpital de la Croix-Rousse, 103, Grande rue de la Croix-Rousse, 69004 Lyon, France*
[b] *Unité CNRS UMR 5220, CREATIS, Inserm U1206, Insa Lyon, université Lyon 1, université Jean-Monnet Saint-Étienne, 7, avenue Jean-Capelle, 69100 Villeurbanne, France*
[c] *Department of radiology, Louis-Pradel hospital, 59, boulevard Pinel, 69500 Bron, France*

KEYWORDS
Artificial intelligence;
Convolutional neural network;
Computed tomography (CT);
Over-scanning;
Ionizing radiation

Abstract
Introduction: The purpose of this study was to develop a convolutional neural network (CNN) to determine the extent of over-scanning in the Z-direction associated with lung computed tomography (CT) examinations.
Materials and methods: The CT examinations of 250 patients were used to train the machine learning software and 100 were used to validate the results. Each lung CT examination was divided into cervical, lung, and abdominal areas by the CNN and 2 independent radiologists, and the length of each area was measured. Every part above or below the lung marks was labeled as over-scanning. The accuracy of the CNN was calculated after the training phase and agreement between CNN and radiologists was assessed using kappa statistics during the validation phase. After validation the software was used to estimate the length of each of the three areas and the total over-scanning in further 1000 patients.
Results: An accuracy of 0.99 was found for the testing dataset and a very good agreement (kappa = 0.98) between the CNN and the radiologists' evaluation was found for the validation dataset. Over-scanning was 22.8% with the CNN and 22.2% with the radiologists. The degree of over-scanning was 22.6% in 1000 lung CT examinations.
Conclusion: Our study shows a substantial over estimation of the length of the area to be scanned during lung CT and thus an unnecessary patient's over-exposure to ionizing radiation. This over-scanning can be assessed easily, reliably and quickly using CNN.

* Corresponding author. Department of radiology, hôpital de la Croix-Rousse, 103, Grande rue de la Croix-Rousse, 69004 Lyon, France.
E-mail address: loic.boussel@chu-lyon.fr (L. Boussel).

https://doi.org/10.1016/j.diii.2018.11.001
2211-5684/

Computed tomography (CT) is one of the most prevalent imaging methods with over a hundred million CT examinations performed every year in the world. However, CT is a radiating imaging technique [1,2]. For this reason, the dose of ionizing radiation is measured for each CT examination. This information is particularly important for chest CT examination, as the organs exposed during this procedure are highly sensitive to radiations.

The automated recording of the radiation dose during CT scans has led to the publications of several studies showing that for similar procedures between different or identical scanner types the dose delivered to the patients can be very variable [3—6]. This inconsistency can be explained at least in part by the many factors that influence the level of radiation, such as tube rotation speed, helical pitch, collimation, filtration, image reconstruction, and patient weight and diameter [7,8]. Over the last ten years many of these parameters have been optimized and the introduction of iterative reconstruction [9] or automatic exposure control (AEC) [10] for example has led to a significant decrease in radiation exposure. Nevertheless, there is still one major cause for variation of the dose delivered to the patients that has not much improved over the years, which is the selected length of the region to be scanned.

Today, no automated system allows a precise and reproducible selection of regions to be scanned. As a consequence, the radiologic technologist hand selects the area to be examined on the 2-dimensional scout view. Most of the times, especially for technologists with little experience, the length of the area to be scanned is over-estimated in order to not miss any possible valuable information but this leads to an unnecessary over-exposure to X-Rays. Accordingly, some studies have reported that up to 80% of chest and abdomen scans had excessive coverage, resulting in higher effective doses for patients [11,12]. In these studies the radiologists assessed the over-scanning manually by going through a lot of chest CT performed in their institutions. This method makes it impossible to analyze an extensive number of patients so it is difficult to extend the results to the general population. A much more efficient process would be to take advantage of artificial intelligence, such as convolutional neural networks (CNN) that are increasingly being studied for radiology applications [13].

Therefore, the purpose of this study was to develop a CNN to determine the extent of over-scanning in the Z-direction associated with lung CT examinations.

Materials and methods

Population

All CT studies included in this study were extracted from the picture archiving and communication system (PACS) of the Hospices Civils de Lyon. Data usage policy of the Hospices Civils de Lyon in terms of confidentiality, anonymization and security was applied for each study and approval was obtained from our local committee.

The database consisted of 1350 randomly selected pulmonary CT examinations containing the keywords ''thorax'', ''pulmonary embolism'' or ''mediastinum'' (one CT examination per-patient), performed at our institution between January 2017 and May 2018. All patients were over 18 years of age.

The chest CT examinations of 350 patients were used to train and validate the machine learning software. Amongst those, 250 CT examinations making the ''learning dataset'' were used to train the system and 100 CT examinations making the ''validation dataset'' were used for its validation. In the training dataset, 48 (19.2%) examinations were obtained after intravenous administration of iodinated contrast material. The 100 CT examinations used for the validation dataset were obtained in 100 patients. There were 66 men for 34 women, with a mean age of 64.2 ± 15.4 (SD) years (range: 25—99 years). Thirty-one of the 100 examinations (31/100; 31%) were obtained after intravenous administration of iodinated contrast material. Finally, after the validation process, the trained machine learning system was used on a cohort of 1000 patients (595 men, 405 women; mean age, 62.0 ± 16.3 [SD] years; range, 18—95 years). One hundred and eighty seven of the 1000 examinations (187/1000; 18.7%) were obtained after intravenous administration of iodinated contrast material.

CT protocol

Several helical CT units were used, including Revolution GSI® (General-Electric healthcare), Brilliance 40® (Philips), Brilliance 64® (Philips), iCT 256® (Philips), Ingenuity CT® (Philips), IQon — Spectral CT® (Philips), Somatom Definition AS® (Siemens Healthineers), Somatom Definition AS+® (Siemens Healthineers). The scanning parameters were as follows: tube voltage (mean, 121 ± 9 [SD] kVp; range: 100—140 kVp) and slice thickness (mean, 1.8 ± 0.8 [SD] mm; range: 0.9—3 mm).

Training and validation processes

Each CT examination of the learning and validation datasets was labeled by a radiologist with 20 years of experience in chest imaging (LB) using custom made software that allows the rater to mark the first and last slices of the examination including at least one part of a lung (i.e. from lung apices to costodiaphragmatic sinus). Each area was then labeled as ''lung'' between the first and last slices, ''cervical'' above the first slice, and ''abdomen'' below the last slice. A second radiologist also labeled the validation dataset independently. In case of discordance between the two radiologists on the position of the marks, a consensual reading was performed to obtain the final marks. The two datasets of labels were used to calculate the inter-rater reproducibility for the labeling.

After being labeled, the learning dataset (15,1006 images) was split into a training (132,136/151,006 images; 87.5%) and a testing (18,870/151,006/images; 12.5%) dataset. The training data set contained 8866 images in the cervical area, 102,966 images in the lung area and 20,304 images in the abdominal area. The testing dataset contained 1267 images in the cervical area, 14,708 images in the lung area and 2895 images in the abdominal area.

The validation dataset consisted in 37,915 images (from the series of 100 examinations) distributed as follows: cervical area = 2609, lung area = 29,678 and abdominal area = 5628.

Training process and database

Images were first converted into 16-bits portable network graphics (PNG) format and down sampled from 512×512 to 128×128 pixels. An image augmentation method consisting in a random horizontal flip to account for patients with only one lung or unilateral pleural effusion was applied on the training images. A shuffle method was then applied to obtain a random sorting of the images.

The CNN was implemented using TensorFlow 1.5 (Google, USA) and Python 3.6.3 (Python Software Foundation, USA). It consisted in the succession of 2 consecutive convolution kernels (size 3×3), a max pool (reduction factor = 2), another 2 consecutive convolution kernels (size 5×5), a max pool (reduction factor = 2), two fully connected layers of 1024 elements and a final class prediction layer of 3 classes (cervical, lung and abdomen). A rectified linear unit activation function was used for the convolution kernels and the fully connected layer. Loss function consisted in a softmax classifier with cross entropy. The images were handled in batches of 32. A drop out of 0.75 was applied on the fully connected layers and 20 epochs were performed. All this process was performed using a GPU NVIDIA GeForce GTX 1080 Ti (NVIDIA Corporation). Finally, the testing dataset was used to assess the accuracy of the CNN.

Validation process

The validation was performed following the training phase. The data obtained from the CNN and the radiologists' evaluations were used to calculate the lengths of the cervical, lung and abdominal areas by multiplying the number of slices of each category by the space between slices. The total over-scanning was calculated using the following equation:

Total over-scanning = (number of cervical slices + number of abdominal slices)/total number of slices.

The mean absolute difference between CNN and radiologists' evaluations were calculated for each area as well as for the total over-scanning. To account for any possible oscillations of the CNN response at the edges, we took a window of 10 slices (5 before the transition and 5 after the transition) at the cervical-to-lung and lung-to-abdominal transitions and calculated the inter-observers as well as the CNN versus human agreements, using kappa statistics.

Retrospective evaluation of 1000 patients

For the cohort of 1000 patients the length of cervical, lung and abdominal areas and the total over-scanning were calculated using only the validated machine learning software. During this step, the computational time was recorded.

Furthermore, in order to be comparable with previous studies [10,11], margins of 2 cm were subtracted to the cervical and the abdominal areas, and the length of the resulting over-margin scans as well as the proportion of scanners above these margins were reported. These 2 cm margins were added to account for 1 cm for differences in patients' respiration between the acquisition and the topogram, and 1 cm for ''operator-dependent'' estimation of the imaging range on the topogram.

Statistical analysis

Statistical analyses were performed using Intercooled Stata 11 (StataCorp LP). Quantitative variables were expressed as mean ± standard deviation (SD) and ranges. Qualitative variables were expressed as raw numbers, proportions or percentages. For kappa values, the following rating was used for significance: Poor = 0—0.20; Fair = 0.21—0.40; Moderate = 0.41—0.60; Good = 0.61—0.80; Very good = 0.81—1.00.

During testing and validation phases, accuracy was calculated as the sum of true positive and true negative divided by the total number of items. A value of one would mean a perfect prediction. The comparison between the results of the CNN and the consensual labeling of the radiologists was assessed using accuracy and kappa statistics. Kappa statistics was also used to calculate the inter-observer reproducibility between the two radiologists.

A Bland-Altman representation and a linear regression analysis were also performed for over-range as well as cervical, lung and abdominal length assessed during the validation phase.

Results

Testing and validation

The total time of the training phase of 20 epochs was 8 hours. The accuracy of the testing dataset (accuracy = 99%) was considered very good so the validation process was performed.

During the validation phase, the CNN and the two radiologists assessed all the 37,915 images. The inter-rater reproducibility was very good ($kappa_{INTER\text{-}RATER} = 0.981$). Agreement between CNN and radiologists' evaluation was very good ($kappa_{VALIDATION} = 0.957$) (Figs. 1 and 2). Regarding agreement on the cervical-to-lung and lung-to-abdomen transition zones, kappa of 0.81 and 0.75 for inter-observers and 0.87 and 0.48 for CNN versus radiologists, were found. The mean abdominal length difference between CNN and radiologists was 2.8 ± 5.3 mm. Finally, CNN had an accuracy of 98.4%.

The means and ranges of length and over-scanning for the cervical, lung and abdomen estimated by the CNN and the radiologists are presented in Table 1. The mean absolute and range differences between CNN and radiologists for the cervical, lung and abdominal areas are presented in Table 2. The total over-scanning involved predominantly the abdominal (68.7% for CNN and 66.9% for radiologists) than the cervical area (31.3% for CNN and 33.1% for radiologists).

Cohort of 1000 patients

The mean lengths and ranges for the cervical, lung and abdominal areas calculated by the machine learning software are presented in Table 3. Including margins described above, the total mean extra-margins were 38.9 ± 31.0 (SD) mm, with 7.8 ± 13.1 (SD) mm for the cervical area and 31.1 ± 26.2 (SD) mm for the abdominal area. Altogether, the total percentage of over-scanning was 22.6 ± 7.9 (SD) %. Additionally, the proportion of over-scanning in cervical, abdominal and both directions were 66.1%, 87.5% and

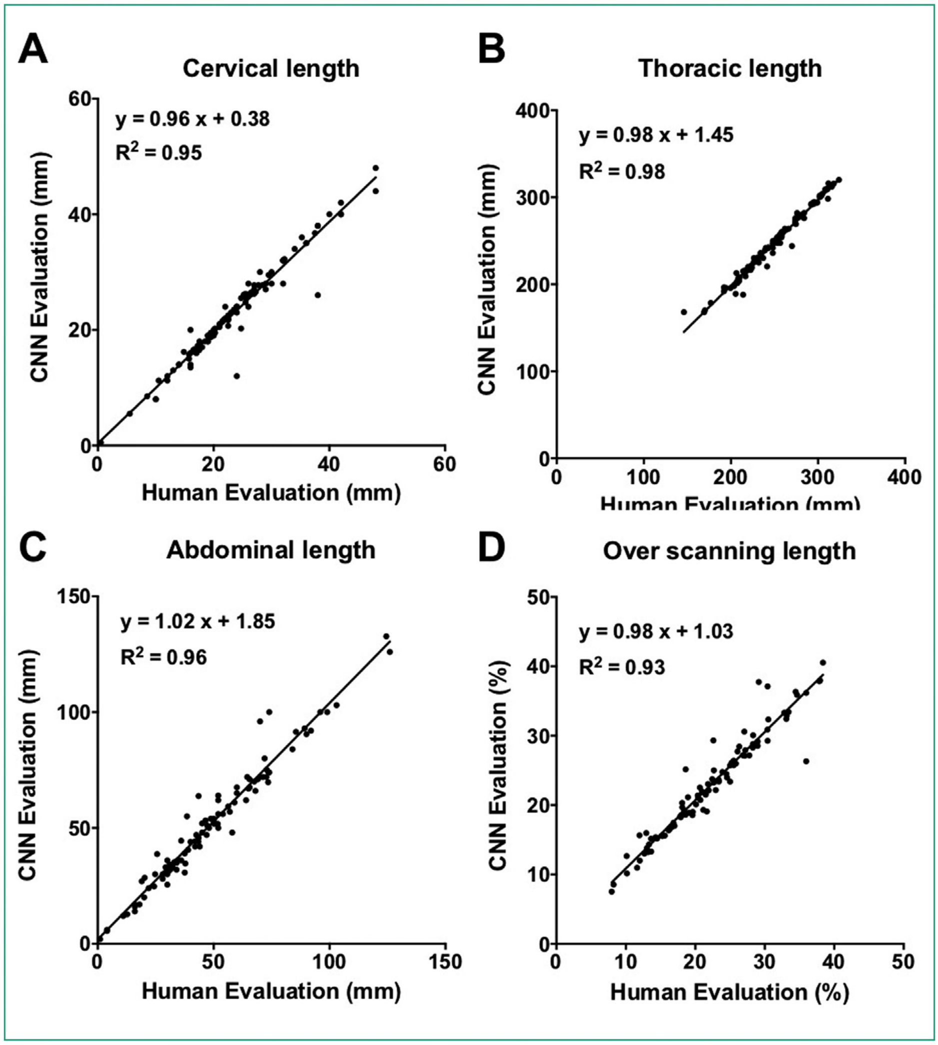

Figure 1. Graphs show correlations between the radiologists (Human Evaluation) and convolutional neural network (CNN) evaluations of cervical (A), thoracic (B), abdominal (C) and over-scanning (D) lengths (mm).

57.8%, respectively. Finally, the mean computation time was 0.05 seconds per slice, resulting in a mean computation time of 23.04 s per examination.

Discussion

Our study demonstrates that CNN can accurately recognize anatomic levels on chest CT and calculate the over-scanning associated with inappropriate scan length. Over the last decades, a lot of progress has been made by the manufacturers to decrease the radiation dose in chest CT. Automatic tube-current modulation, new scanning protocols with kVp adjusted to the size and weight of the patient, prospective ECG-triggering as well as iterative reconstruction have allowed substantial dose saving [14—23]. Nevertheless, they do not consider the inappropriate anatomic scan coverage, which is an important source of dose increase and variation. Indeed, Litmanovich et al. have estimated that appropriate adjustment of the z-axis can result in a radiation dose savings of more than 40% [23].

For comparability issues with two previous studies that have reported measurements of over-scanning, we used the same extra-scanning margin of 2 cm in the cranial and caudal directions [11,12]. Zanca et al. have reported a mean extra-imaging length of 1.8 cm and 2.9 cm at the top and the bottom of chest CT, respectively [11]. Moreover, in their large study involving 600 patients over 6 hospitals, Schwartz et al. confirmed these results and also observed an important variation of over-scanning between the reporting centers [12]. In these two studies, the calculation of the over-scanning was performed manually, using the two-dimensional scout view or a coronal reconstruction of the chest CT.

In our study, the use of CNN provided result similar to those obtained in the aforementioned studies but in a fully automated and much faster mode. Indeed, we recorded mean rostral and caudal over-scanning lengths of 0.8 cm and 3.1 cm, respectively, by comparison with 1.8 cm and 2.9 cm reported by Zanca et al. [11]. However, we observed a more systematic over-scanning of the abdomen with an incidence of 87.5% in our study versus 53% in Zanca et al. [11], and 4—60% in Schwartz et al. [12].

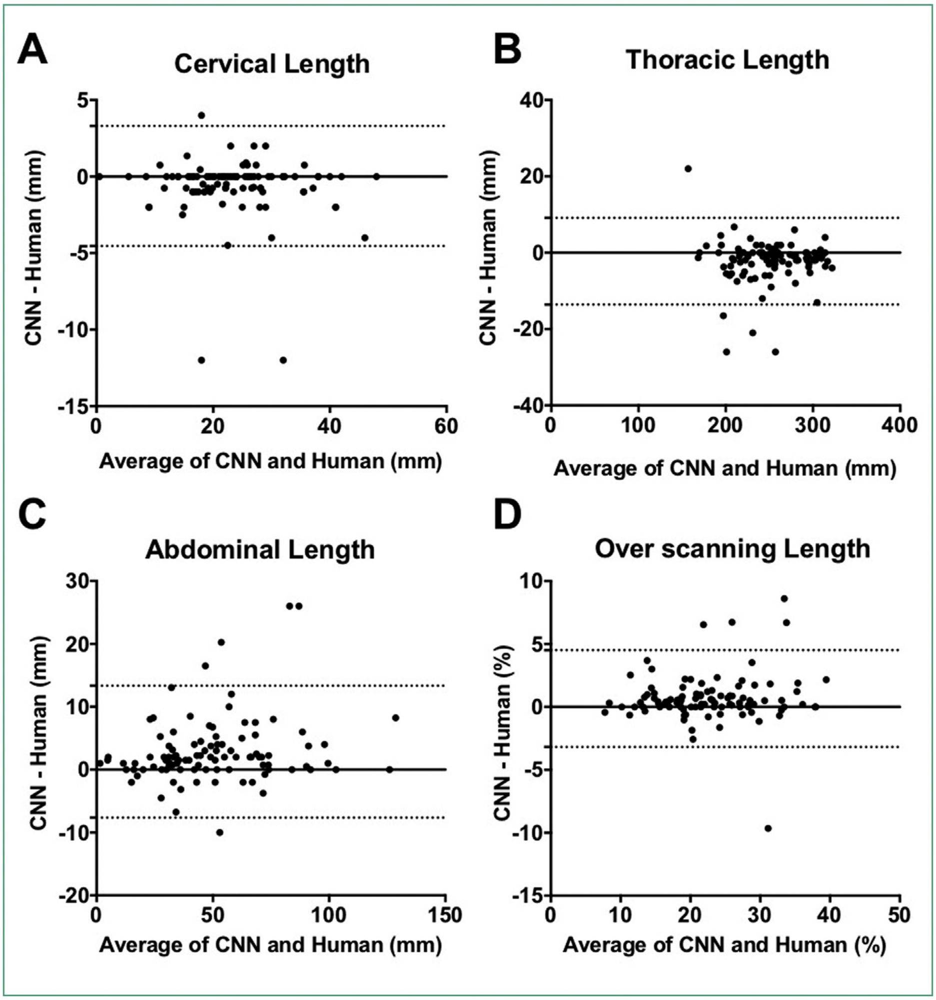

Figure 2. Diagrams show Bland-Altman representation of the comparison between radiologist and convolutional neural network evaluations of cervical (A), thoracic (B), abdominal (C) and over-scanning (D) lengths.

Table 1 Length and over-scanning measured by the convolutional neural network and the radiologists.

	CNN	Radiologists
Cervical (mm)	23.3 ± 8.7 [0.5–48]	23.9 ± 8.8 [0.5–48]
Lung (mm)	250 ± 38.5[167.8–320]	252.2 ± 38.6[146–324]
Abdominal (mm)	51.2 ± 25.9 [2–132.7]	48.3 ± 24.9 [1–126]
Over-scanning (%)	22.8 ± 7.3 [7.5–40.5]	22.1 ± 7.1 [8–38.4]

Results are expressed as mean ± standard deviation. Numbers in brackets are ranges. CNN indicates convolutional neural network

The short computational time (23.04 s per examination) and the automated calculation allowed by the CNN may have several important advantages. First, the CNN method could be implemented to prospectively stop the acquisition when the end of the lung is reached. This would require an integrated complex solution in addition to a fast hardware setting but seems reachable with the current GPU processors. Second, we could consider in the near future after further improvement of the computational time to provide the results of the over-scanning immediately after the examination, which could help the technologists improve their practice. Indeed, they mostly rely on the landmarks seen on the topogram to determine the start and end of CT data acquisition and naturally tend to increase the total coverage to be sure to include the whole lung. This is particularly true for the lower part of the acquisition since the costodiaphragmatic sinus is often difficult to localize precisely on the topogram. Accordingly, in our study, the proportion of over-scanning was found to be approximately two third at the abdominal level.

Table 2 Absolute mean and range differences as measured by the convolutional neural network and the radiologists.

Areas	Differences
Cervical (mm)	0.6 ± 2.2[0–12]
Lung (mm)	2.2 ± 5.8[0–26]
Abdominal (mm)	2.8 ± 5.3[0–26]

Results are expressed as mean ± standard deviation. Numbers in brackets are ranges.

Table 3 Mean length and ranges measured by the machine learning software on the 1000 patients cohort.

Areas	Mean length
Cervical (mm)	26.1 ± 14.5 [0–162]
Lung (mm)	255.8 ± 35.6[166–377]
Abdominal (mm)	50.0 ± 27.9 [0–326]

Results are expressed as mean ± standard deviation. Numbers in brackets are ranges.

In our study, the training cases were randomly selected to provide a representative population. While this can make the task of the CNN more complicated during inference, it is also an important reason explaining the performance of our CNN that despite being relatively simple still allowed reaching a high accuracy of 98.4%. It also helped make the results independent of the CT scanner vendor and the presence or absence of contrast injection. Additionally, a very good agreement was observed between the CNN and the observers. This could be partially related to the fact that the mean length of the lung represents about two third of the total acquisition length and the CNN response is expected to be relatively stable in the central part of each anatomical levels. Therefore, we also calculated the agreement on the transition zones (i.e. cervical-to-lung or lung-to-abdomen) and report that the CNN shows a lower agreement with the observers at the lung-to-abdomen level. This difference is due to oscillations of the response of the CNN in this transition area caused by the high heterogeneity of the images at the lower part of the lung. The impact of this problem is globally small and probably does not affect too much our measurement on the global population with regard to the 2 cm margins that was used in the calculation of the over-scanning. However, this issue would have to be considered for per-patient individual analysis.

Our study has some other limitations, such as the fact that we did not take the medical indications into account when selecting the patients. Therefore, it is possible that in some patients the over-scanning was actually needed for diagnostic purpose, such as adrenal gland assessment in patient with pulmonary cancer. Nevertheless, it is unlikely that it was the case for every examination over the full cohort of 1000 patients. Nevertheless, adding the clinical indication of the CT examination will certainly help interpret individual over-scanning results. Another limitation of our study is the absence of radiation dose calculation. Indeed, our goal was to develop and validate a tool enabling a systematic and automatic calculation of the extra-Z coverage in chest CT, regardless of any scan parameters used during acquisition. Furthermore, the estimation of the excess of dose related to over-scanning has already been published [11,12].

In conclusion, our study demonstrates that it is possible to automatically measure the over-scanning related to the over estimation of the area to be scanned during lung CT scans. In addition to the other common dose parameters, it could allow a decrease in radiation dose exposure.

Human and animal rights

The authors declare that the work described has been carried out in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments involving humans as well as in accordance with the EU Directive 2010/63/EU for animal experiments.

Funding

This work did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Contribution of authors

M. Colevray contributed to the conception and design of the study, the acquisition, analysis and interpretation of data, and the critical revision of the article for important intellectual content.

VM. Tatard-Leitman contributed to the analysis and interpretation of data, and the drafting and critical revision of the article for important intellectual content.

S. Gouttard contributed to the conception and design of the study, the analysis and interpretation of data, and the critical revision of the article for important intellectual content.

L. Boussel contributed to the conception and design of the study, the acquisition, analysis and interpretation of data, and the drafting and critical revision of the article for important intellectual content.

P. Douek contributed to the conception and design of the study and the critical revision of the article for important intellectual content. All authors gave their approval of the final version to be submitted.

Acknowledgements

We would like to thank Christopher for its technical help and Adeline Mansuy for her assistance in obtaining all authorizations necessary for conducting the study.

Disclosure of interest

The authors declare that they have no competing interest.

References

[1] Doss M. Linear no-threshold model may not be appropriate for estimating cancer risk from CT. Radiology 2014;270:307—8.

[2] Larbi A, Orliac C, Frandon J, Pereira F, Ruyer A, Goupil J, et al. Detection and characterization of focal liver lesions with ultra-low dose computed tomography in neoplastic patients. Diagn Interv Imaging 2018;99:311—20.

[3] Shrimpton PC, Wall BF. Reference doses for paediatric computed tomography. Radiat Prot Dosim 2000;90:249—52.

[4] Aldrich JE, Bilawich AM, Mayo JR. Radiation doses to patients receiving computed tomography examinations in British Columbia. Can Assoc Radiol J 2006;57:79—85.

[5] Hatziioannou K, Papanastassiou E, Delichas M, Bousbouras P. A contribution to the establishment of diagnostic reference levels in CT. Br J Radiol 2003;76:541—5.

[6] Koller CJ, Eatough JP, Bettridge A. Variations in radiation dose between the same model of multislice CT scanner at different hospitals. Br J Radiol 2003;76:798—802.

[7] McNitt-Gray MF. AAPM/RSNA physics tutorial for residents: topics in CT. Radiation dose in CT. Radiographics 2002;22:1541—53.

[8] Kubo T, Lin PJ, Stiller W, et al. Radiation dose reduction in chest CT: a review. AJR Am J Roentgenol 2008;190:335—43.

[9] Fillon M, Si-Mohamed S, Coulon P, Vuillod A, Klahr P, Boussel L. Reduction of patient radiation dose with a new organ based dose modulation technique for thoraco-abdominopelvic computed tomography (CT) (Liver dose right index). Diagn Interv Imaging 2018;99:483—92.

[10] Singh S, Kalra MK, Thrall JH, Mahesh M. Automatic exposure control in CT: applications and limitations. J Am Coll Radiol 2011;8:446—9.

[11] Zanca F, Demeter M, Oyen R, Bosmans H. Excess radiation and organ dose in chest and abdominal CT due to CT acquisition beyond expected anatomical boundaries. Eur Radiol 2012;22:779—88.

[12] Schwartz F, Stieltjes B, Szucs-Farkas Z, Euler A. Over-scanning in chest CT: comparison of practice among six hospitals and its impact on radiation dose. Eur J Radiol 2018;102:49—54.

[13] Yamashita R, Nishio M, Do K, Togashi K. Convolutional neural networks: an overview and application in radiology. Eur Radiol 2017;27:611—29.

[14] McCollough CH, Bruesewitz MR, Kofler Jr JM. CT dose reduction and dose management tools: overview of available options. Radiographics 2006;26:503—12.

[15] McCollough CH, Primak AN, Braun N, Kofler J, Yu L, Christner J. Strategies for reducing radiation dose in CT. Radiol Clin North Am 2009;47:27—40.

[16] Lee CH, Goo JM, Ye HJ, Park CM, Chun EJ, Im JG. Radiation dose modulation techniques in the multidetector CT era: from basics to practice. Radiographics 2008;28:1451—9.

[17] Yu L, Li H, Fletcher JG, McCollough CH. Automatic selection of tube potential for radiation dose reduction in CT: a general strategy. Med Phys 2010;37:234—43.

[18] Deak PD, Langner O, Lell M, Kalender WA. Effects of adaptive section collimation on patient radiation dose in multisection spiral CT. Radiology 2009;252:140—7.

[19] Silva AC, Lawder HJ, Hara A, Kujak J, Pavlicek W. Innovations in CT dose reduction strategy: application of the adaptive statistical iterative reconstruction algorithm. AJR Am J Roentgenol 2010;194:191—9.

[20] Leipsic J, Labounty TM, Heilbron B, Min JK, Mancini GBJ, Lin FY, et al. Estimated radiation dose reduction using adaptive statistical iterative reconstruction in coronary CT angiography: the ERASIR study. AJR Am J Roentgenol 2010;195:655—60.

[21] Leipsic J, Nguyen G, Brown J, Sin D, Mayo JR. A prospective evaluation of dose reduction and image quality in chest CT using adaptive statistical iterative reconstruction. AJR Am J Roentgenol 2010;195:1095—9.

[22] Leipsic J, Heilbron BG, Hague C. Iterative reconstruction for coronary CT angiography: finding its way. Int J Cardiovasc Imaging 2012;28:613—20.

[23] Litmanovich DE, Tack DM, Shahrzad M, Bankier AA. Dose reduction in cardiothoracic CT: review of currently available methods. Radiographics 2014;34:1469—89.