What patient positioning in chest X-ray is still acceptable? An empirical study on thresholding quality metrics

Berg, J. von; Hergaarden, K.F.M.; Englmaier, M.; Pfeiffer, D.; Wieberneit, N.; Krönke-Hille, S.; ... ; Lamb, H.J.

AKADÉMIAI KIADÓ

What patient positioning in chest X-ray is still acceptable? – An empirical study on thresholding quality metrics

IMAGING

ORIGINAL ARTICLE

JENS VON BERG[1†*], KENNETH F. M. HERGAARDEN[2†], MAX ENGLMAIER[3], DANIELA PFEIFFER[3,4], NATALY WIEBERNEIT[5], SVEN KRÖNKE-HILLE[1], TIM HARDER[1], ANDRÉ GOOßEN[1], DANIEL BYSTROV[1], MATTHIAS BRUECK[1], STEWART YOUNG[1] and HILDO J. LAMB[2]

[1] Philips Innovative Technologies, Röntgenstr 24–26, 22335 Hamburg, Germany

[2] Department of Radiology, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands

[3] Department of Diagnostic and Interventional Radiology, School of Medicine & Klinikum Rechts der Isar, Technical University of Munich, Ismaningerstr 22, 81675 Munich, Germany

[4] Institute for Advanced Study, Technical University of Munich, Lichtenbergstrase 2a, 85748 Garching bei München, Germany

[5] Philips Medical Systems DMC, Röntgenstr 24–26, 22335 Hamburg, Germany

Received: November 17, 2023 • Revised manuscript received: February 27, 2024 • Accepted: March 29, 2024

ABSTRACT

Background and Aim: Issues in patient positioning during chest X-ray (CXR) acquisition impair diagnostic quality and potentially increase radiation dose. Automated quality assessment was proposed to address this. Our objective is to determine thresholds on some quality control metrics following international guidelines, that represent expert knowledge and can be applied in a comprehensible and explainable AI approach for such an automatic quality assessment.

Materials and Methods: An AI-method estimating collimation distance to the ribcage, balancing between both clavicle heads, and number of ribs above the diaphragm as metrics for collimation, rotation, and inhalation quality was applied on 64,315 posteroanterior CXR images from a public dataset (ChestX-ray8). From this set 920 CXR images were sampled and manually annotated to gain additional trusted reference metrics. Seven readers from different institutions then classified the acquisition quality of these images independently into okay, inadequate, or unacceptable following the criteria of international guidelines. Optimal thresholds on the metrics were determined to reproduce these classes using the metrics only.

Results: A fair to moderate agreement between the experts was found. When disregarding all inadequate rates a classification on the metrics was able to separate okay rated cases from unacceptable cases for collimation (AUC > 0.97), rotation (AUC = 0.93) and inhalation (AUC = 0.97).

Conclusion: Suitable thresholds were determined to reproduce expert opinions in the assessment of the most important quality criteria in CXR acquisition. These thresholds were finally applied on the AI-method's estimates to automatically classify image acquisition quality comprehensibly and according to the guidelines.

KEYWORDS

chest radiography, image quality, patient positioning, explainable artificial intelligence, quality management

IMAGING 16 (2024) 1, 41–49
DOI: 10.1556/1647.2024.00187

†J.v.B. and K.H. contributed equally to this work as first authors.

*Corresponding author.
Tel.: +49 40 34971 1321.
E-mail: Jens.von.Berg@philips.com

AKJournals

Introduction

Proper patient positioning is crucial for diagnostic quality of radiographs [1] and relevant quality criteria and recommendations have been formulated in official standards [2]. The European guidelines on quality criteria for diagnostic radiographic images list three image criteria concerning patient positioning [3]: (1) Performed at full inspiration (as assessed by the position of the ribs above the diaphragm — either 6 anteriorly or 10 posteriorly) (2) Symmetrical reproduction of the thorax as shown by central position of the spinous process between the medial ends of the clavicles (3) Reproduction of the whole rib cage above the diaphragm. Furthermore, the ALARA ("as low as reasonably achievable") principle states that unnecessary X-ray exposure should be avoided, which requires collimation always to be balanced between completeness and dose considerations.

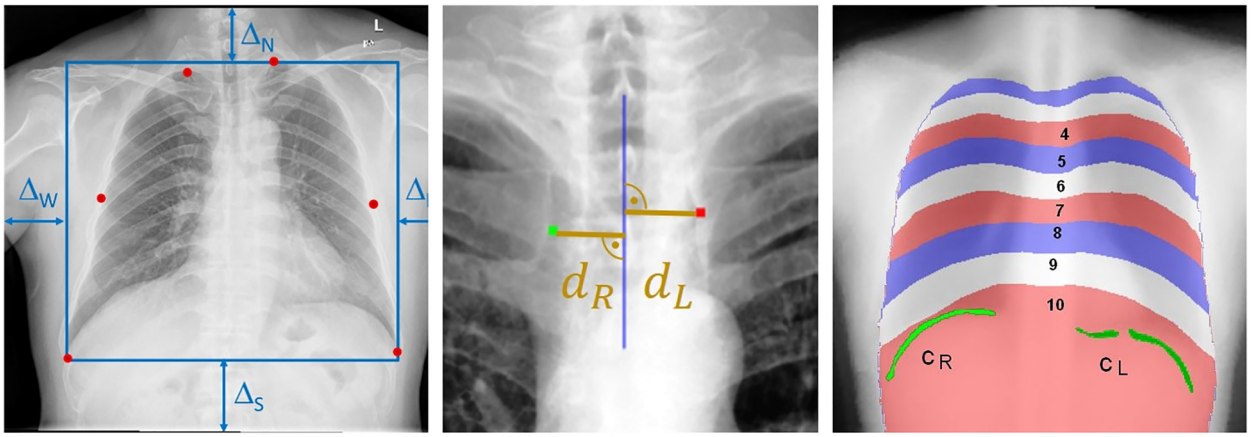
In daily routine, there are many factors causing deviations from these standards, such as high workload, lack of training, missing education, missing feed-back, and lack of rewards. A study found approx. 15%, approx. 5%, and approx. 50% of 116 images were violating the above three criteria when applied strictly [4]. Such images may cause extra workload for the reading radiologist, mean potential harm to the patient through delayed or missed diagnosis and avoidable exposure by too wide collimation or even retakes. Typically, a radiology department board occasionally discusses the quality of only a limited sample of images in order to maintain or improve certain quality levels [5, 6]. Quality monitoring based on well reproducible measurements [7] applied to every acquired image would be preferable, but it requires an automated robust method to evaluate adherence to the most relevant quality criteria [8]. Any automation on quality control that tries to reproduce expert performance is challenged by the fact that different experts may apply different criteria and use different thresholds on them to judge a given image. Some studies report a moderate agreement (Cohen's Kappa $0.4 < \kappa \leq 0.6$) between experts [9], while others report a fair agreement ($0.2 < \kappa \leq 0.4$) only [4, 10]. Also, the differences between radiologists from different countries have been addressed [11].

There are two approaches reported recently on automatic evaluation of chest radiography positioning quality. In the first approach, a number of training images were classified by an expert to be either insufficient (too narrow), appropriate, or excessive (too wide) in each single edge of the image [12, 13]. Further on, inspiration and rotation were both rated appropriate or not among some further criteria. A classification neural network was trained end-to-end to reproduce these ratings directly from the image. In contrast, the second approach used machine learning to derive quality metrics based on comprehensible measurements like estimating the number of ribs above the diaphragm or measuring specific distances in the image like those between a clavicle and the spinous processes or the abdominal space below the lung base being exposed to radiation [14, 15]. Some thresholds may be applied on these quality metrics to finally come to a classification (e.g. ok or unacceptable). That has been done in a prospective study giving immediate feedback to the radiographers whether one of the three quality aspects leads to imprecise positioning [16]. The thresholds used for the AI-method in that study were listed but their choice was not motivated nor have the consequences of this choice been discussed. This second approach is followed here because it appears more comprehensible and configurable than end-to-end learning. For quality classification consistent with experts however it is based on the assumption that for each of the three quality aspects thresholds can be established that separate images considered ok from images considered unacceptable based on these metrics only. Here, objective thresholds are to be established that can be applied to the three quality metrics in an automated context. To that end, a multi-site multi-reader study was performed that aims at testing the hypothesis that (i) the automatically computable quality metrics correspond to expert's quality impressions of the image, (ii) empirical quality classes are separable using these quality metrics, and (iii) variability among experts' assessments does not counteract using such an automatic method.

Materials and methods

All adult patients' CXRs imaged in an upright and posterioanterior orientation were selected from the publicly available chest X-ray8 data base [17] from the National Institute of Health giving 64,315 such images. These images were processed by an AI-method capable of automated quantification of six quality metrics aimed at assessing the quality of collimation, rotation and inhalation through anatomical feature detection [14]. The collimation metrics ($\Delta N, \Delta S, \Delta E, \Delta W$) measure the distance for each image border to the bounding box surrounding landmarks on the lung borders (Fig. 1a). A wide collimation is indicated by a positive value and an intersection between the bounding box and the collimation is indicated by a negative value. For the rotation metric based on the distances between both clavicle heads to the spinous process line (Fig. 1b) a rotation asymmetry metric α is calculated from $\alpha = \frac{d_R - d_L}{d_R + d_L}$. This value is bounded to $[-1, 1]$ and quantifies deviations from the desired balanced position at $\alpha = 0$. For the metrics on inhalation status the AI-method uses an anatomical atlas matching the rib cage and approximating the amount of posterior ribs visible above both diaphragms as c_L and c_R (Fig. 1c). These numbers may also contain fractions of a rib. In the original approach c is used as the sum of both, here c_{max} the highest number of both sides is used, as pathology or anatomical variance has turned out to complicate the assessment of inhalation status on one side.

From the processed dataset a sampling of 920 images was done. 600 images were specifically sampled for the collimation test in total, 150 for each of the four image borders.



a) ΔN, ΔS, ΔW, ΔE measure the distances of the lung bounding box to the image borders.

b) α balances the clavicle distances d_L and d_R to spinous process line.

c) A rib cage atlas is registered to the image and rib counts c_L and c_R are sampled along the half-diaphragms (green).

Fig. 1. Given the localization of the anatomical landmarks, the quality metrics ΔN, ΔS, ΔW and ΔE, α are geometrically derived and rib counts c_L and c_R are estimated

200 additional images were sampled for the rotation test and 120 for the inhalation status test. The sampling was random apart from the one parameter relevant for each test, the metric with the aim to oversample images with interesting such values. This means the share of cases with values close to an anticipated threshold should be higher than in the entire dataset but the whole parameter range should still be covered. Figure 2 shows the result of this sampling strategy, the distribution of all quality metrics calculated atomatically on the entire data set, and the distribution of the trusted reference quality metrics on each of the corresponding samples.

To exclude potential errors in the anatomical feature detection by the AI-method, landmark placement for the inhalation metric (lung borders), rotation metric (medial clavicle heads and spinous processes) or rib counts above the diaphragms were done manually for all 920 cases by some of the authors. In this case unlike in the AI-method rib counts were set to be integer values for practical reasons. To gain further insights on the quality of the automated assessment on this dataset this manually annotated reference standard was compared to the AI results. The absolute differences between trusted and automatic metrics are on average (median) deviations of ΔN: 3.2 mm (2.7 mm), ΔS: 7.3 mm (6.1 mm), $\Delta W/\Delta E$: 3.9 mm (3.0 mm), α: 0.11 (0.07), rounded c_{max}: 0.51 ribs (0 ribs). Figure 3 shows a comparison of these reference quality metrics with those provided by the AI-method automatically.

The seven readers were three radiologists from a university medical center in the Netherlands, two residents in radiology from a university hospital in Germany and two experienced radiographers certified to read CXR images (reading radiographers) from a regional supply hospital in Denmark. The raters were instructed on the acquisition quality criteria following the European guidelines [3] and had access to guidance documents at any time. The images were presented to the readers in randomized batches, in which the collimation, rotation and inhalation quality were separately scored as being 'OK', 'inadequate' or 'unacceptable'. For the collimation per image border additional options 'too narrow' and 'too wide' were given. A ROC analysis was done for each of the metrics on how it separates the images for these different scores. Optimal thresholds were determined to maximize the accuracy of a classifier separating 'OK' from 'unacceptable' cases on votes of subsets of the raters and finally on the votes of all seven raters to represent a kind of 'super reader'.

As an extension, these thresholds were applied to classify both the six samples separately as well as the entire data set into 'OK' and 'unacceptable' fully automatically based on the metrics of the AI-method for all the quality criteria.

Results

Ratings related to reference quality metrics

Figure 4 shows the distribution of expert quality classes on the corresponding reference quality metrics. The classes appear generally ordered by the reference quality metrics, some of them overlap a little. The responses of the different raters appear generally similar to each other, sometimes there is a little shift between their responses that means some tend to apply stricter criteria than others. Most strikingly, two raters (R4 and R6) did not choose 'too wide, unacceptable' at any lateral Field-of-View case at all.

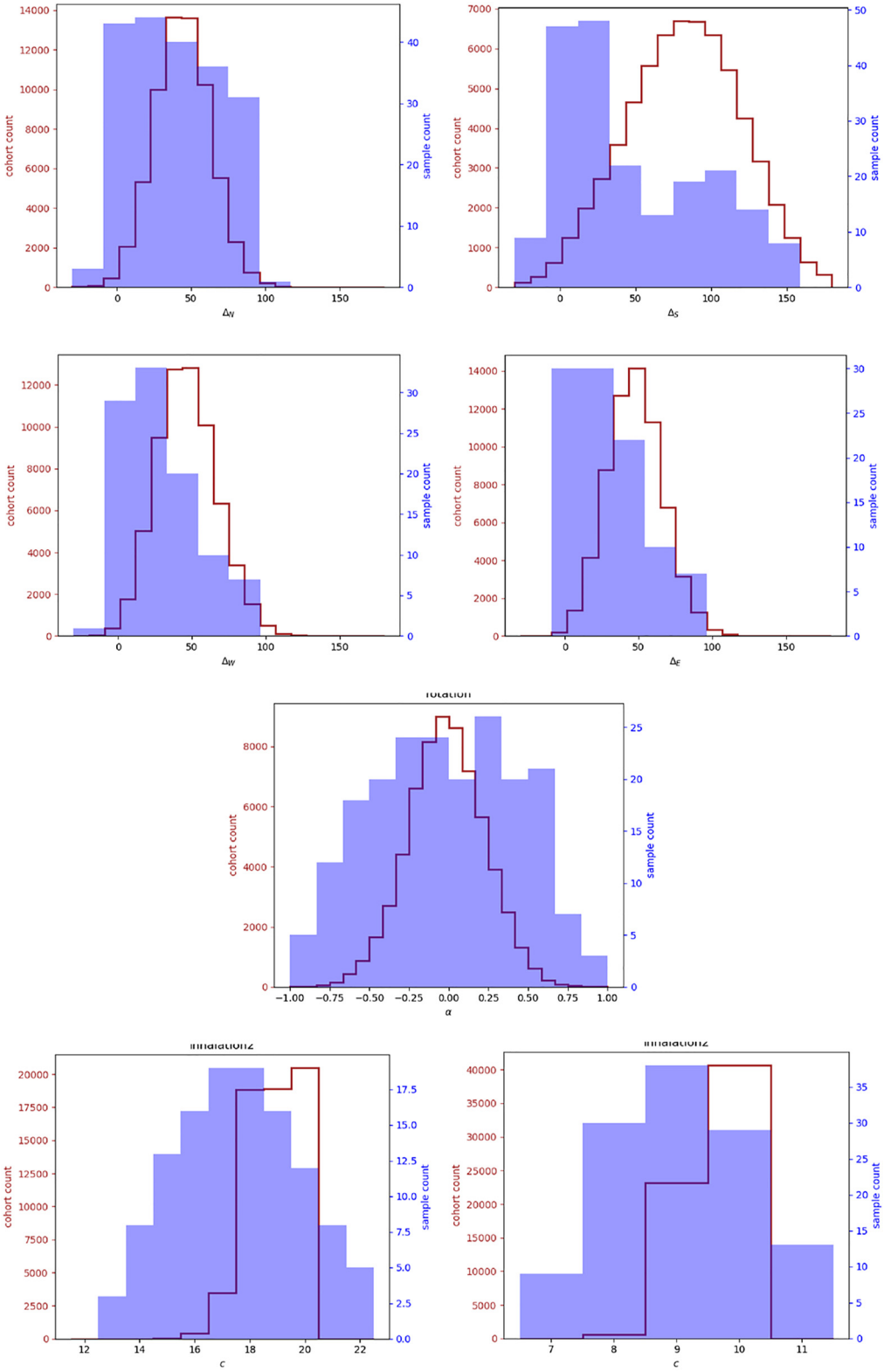


Fig. 2. Distribution of quality metrics (ΔN, ΔS, ΔW, ΔE, α, c, c_{max}) in the entire data set (64,315 cases, brown line) and in the six different samples of 150, 150, 150, 150, 200, and 120 cases to be used in the corresponding study tasks (blue). For the inhalation sample the two different metrics c and c_{max} are both analyzed separately

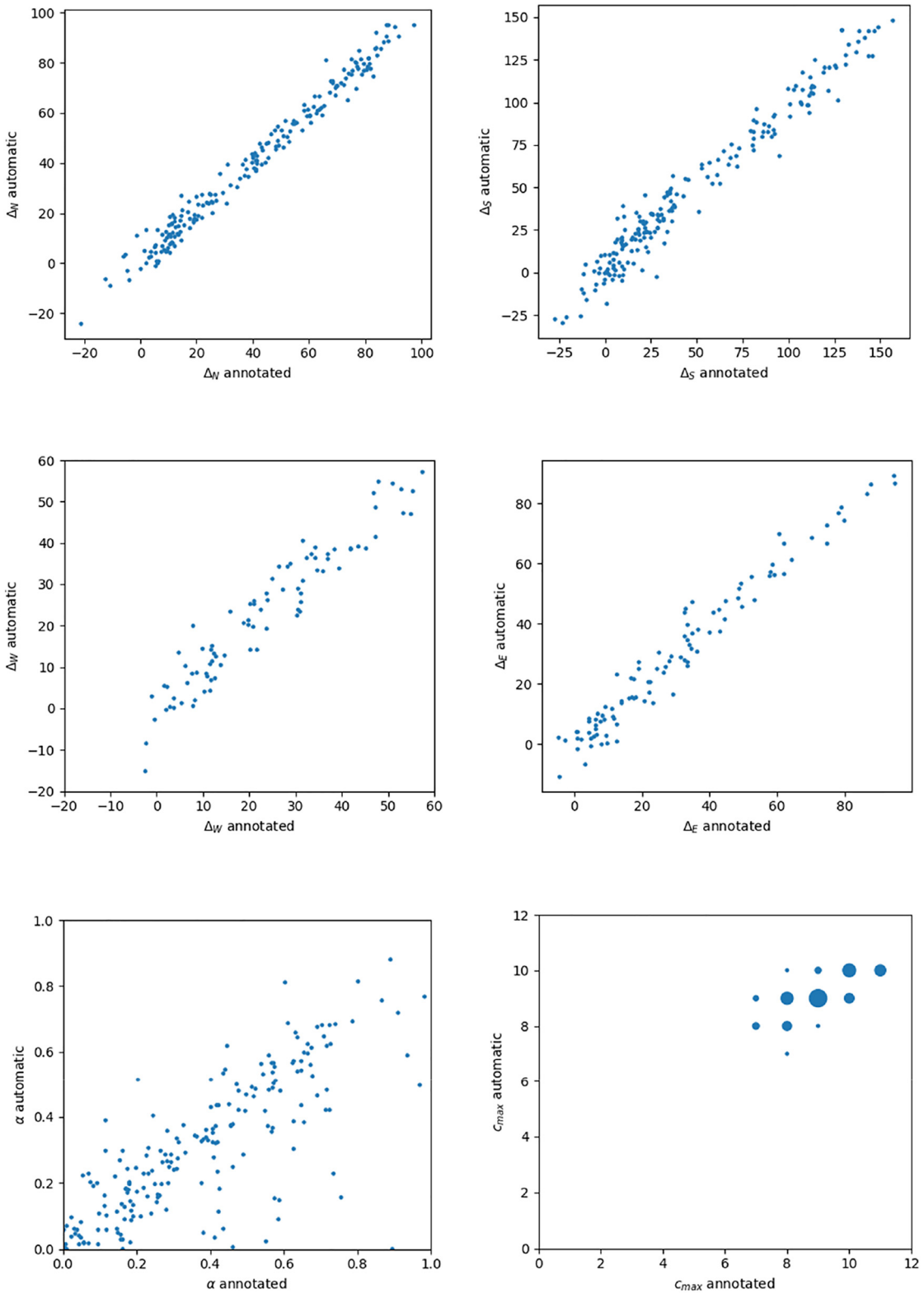


Fig. 3. Comparison of reference quality metrics calculated from manually localized anatomical landmarks (x-axis) and those estimated by the AI-method (y-axis) on the corresponding data set

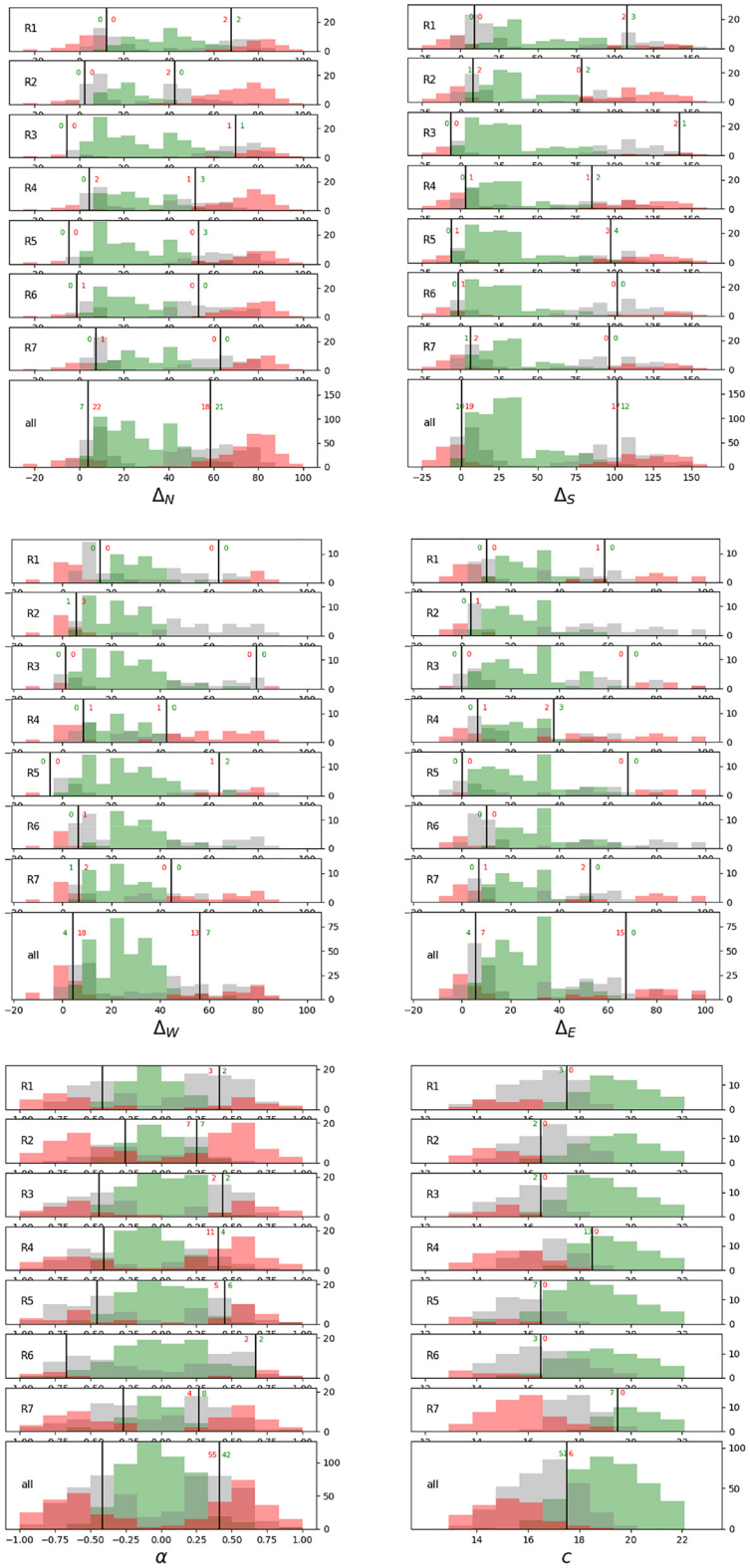


Fig. 4. Distribution of raters' quality classes plotted on the axis of the corresponding reference quality metric that was geometrically derived from the manual landmark localizations. Collimation (ΔN, ΔS, ΔW and ΔE) – from left to right: too wide – unacceptable (red), too wide – inadequate (grey), OK (green), too narrow – inadequate (grey), too narrow – unacceptable (red). Rotation (α) and Inhalation (c): unacceptable (red), inadequate (grey), OK (green). Thresholds are indicated that minimize the number of misclassified rates (numbers also shown close to the threshold line). All seven raters are analyzed separately and in the bottom line their pooled rates are used to derive common thresholds

Separability of quality classes

The agreement among the seven raters is moderate for most of the quality criteria (Cohen's kappa between 0.36 and 0.50). When applied to the institution's subgroups this does not substantially increase the agreement. It was observed quite often that one rater chose 'inadequate' but another chose 'Ok' or 'unacceptable' for the same image. This is reflected in Fig. 4 by overlapping distributions (red with grey or grey with green). A separation between 'inadequate' and 'unacceptable' on all seven experts' pooled votes could be done by the metric with an area under the ROC curve (AUC) of values ranging from 0.73 to 0.89, and between 'Ok' and 'inadequate' ranging from 0.84 to 0.94. When disregarding all 'inadequate' (grey) votes a separation of all seven experts' pooled votes between 'Ok' and 'unacceptable' is possible completely or almost completely with an AUC of 0.97 and above, but with 0.93 for Rotation (see Table 1).

Establishing thresholds

The thresholds between quality classes have been determined by minimizing the absolute number of rates misclassified by it. If both distributions do not overlap (which corresponds to AUC = 1.0) this minimum number becomes zero. Table 1 shows all thresholds for ratings of a single reader and for groups of readers pooled by institution. Finally, it provides the thresholds based on all readers' pooled rates representing a 'super reader'. For two of the readers (R2 and R6) a threshold for 'too wide' could not be determined on left and right patient side because they decided not to use any 'Unacceptable' rating there. The reference quality metric values c_{max} from manual annotation are integer by study design. To separate a set of our study cases by the integer reference quality metrics any value between two integers would serve equally well as threshold. The AI-method however estimates interpolated floating point numbers for c_{max}, so here the exact choice of the threshold would make a difference. Thresholds on c_{max} were chosen at half integer values unless there were multiple minima. The thresholds are also indicated in Fig. 4. There, the number of cases that are mis-classified to either side of the threshold are shown as well. The percentages of correctly classified cases (accuracy) by the 'super reader' over all individual rates are 93.5%, 93.9%, 94.3%, and 91.1% for Field-of-View (North, South, East, and West), 88.8% for Rotation, and 89.0% for Inhalation on the respective samples.

Making the AI-method an additional reader

Once the thresholds are established, the AI-method can be used to classify the study cases as well. Table 1 shows AUC values indicating how well the classes formed by the AI-method can be separated using the reference quality metrics as a criterion. Both, the expert group's and the AI-method's classes can be separated well, and similarly well to each other by the reference quality metrics. In detail, the AI-method appears more consistent (higher AUC values) in the Field-of-View task than the expert group, but less consistent in the Inhalation task.

Applying the AI-method to the entire data set

Given these thresholds, the AI-method can be used to classify the entire data set. Table 1 shows the percentages of cases that would fall beyond these thresholds. Violating any of the four 'too narrow' would result in 6.6%. Violating any of the four 'too wide' would result in 61.1%, only cranially or abdominally too wide would result in 56.3%, and violating any of the Field-of-View criteria would result in 65.0%. Excluding any lateral 'too wide' that was not considered relevant by all raters would reduce this to 47.3%. Finally, 76.0% of all cases are violating any criterion in Table 1.

Table 1. Optimal thresholds between 'Ok' and 'Unacceptable' on the corresponding metric for collimation (Δ_N, Δ_S, Δ_W, Δ_E for the four image borders, w:too wide, n:too narrow), rotation (α) and inhalation (c_{max}) per rater R1 to R7, pooled by institution, and on the rates of all seven raters pooled. AUCs for separating all study cases 'Ok' from 'unacceptable' by thresholds and AUCs formed by applying the AI-method as an 'AI rater'. Percentage of images of the entire data set violating each single threshold-based quality criterion. The c_{max} rejection rate may appear too high for reasons discussed in the text

	$\Delta_{N,w}$	$\Delta_{N,n}$	$\Delta_{S,w}$	$\Delta_{S,n}$	$\Delta_{E,w}$	$\Delta_{E,n}$	$\Delta_{W,w}$	$\Delta_{W,n}$	α	c_{max}
R1	68	12	108	9	59	10	64	16	±0.41	9.5
R2	43	2	79	8		4		6	±0.26	9.5
R3	70	−5	142	−6	68	−0	79	1	±0.44	9.5
R4	52	5	85	3	38	6	43	9	±0.40	9.5
R5	53	−5	97	−6	68	0	64	−5	±0.45	8.5
R6	53	−1	102	−2		10		6	±0.67	9.0
R7	63	7	97	6	53	7	44	7	±0.27	10.5
R1 to R3	62	4	102	1	68	6	76	3	±0.39	9.5
R4, R5	53	−1	88	−1	62	2	51	4	±0.42	9.5
R6, R7	59	5	101	1	70	7	47	6	±0.44	9.5
all	**59**	**4**	**102**	**1**	**67**	**6**	**56**	**4**	**±0.41**	**9.5**
AUC (all raters)	0.99	0.97	0.99	0.98		0.99		0.98	0.93	0.97
AUC (AI as rater)	1.00	0.99	0.99	1.00		0.99		1.00	0.93	0.91
rejection rate	18.9%	2.0%	29.5%	2.3%	13.9%	1.0%	27.7%	1.5%	4.1%	36.0%

Discussion

In order to establish an automated quality assessment on patient positioning for posteroanterior CXR images a method is pursued that applies metrics for the three most relevant quality issues concerning collimation, patient rotation, and inhalation. International guidelines already introduce some formal criteria to do such quality assessment, but typically not providing exact numbers for decision making. An AI-method was adapted that establishes and implements some formal metrics for all these quality criteria. It calculates these metrics geometrically from anatomical landmarks that it localizes in the image. This way the method estimates the distances of the rib cage to the image borders, balances the distance of the clavicles to the spinous processes, it estimates and interpolates the number of ribs above the diaphragm. An adherence to guidelines is given to this method by design. Unlike other known approaches the method is comprehensible because it does not only provide a binary decision but also all decision criteria. It may also be configured and parameterized to specific requirements by adapting the thresholds. What was missing so far is to prove whether such an automated method also corresponds to experts' assessment of positioning quality, especially for interesting borderline cases. Also, thresholds on these metrics were missing that optimally reproduce the experts' ability to classify images into okay and unacceptable.

A careful approach was followed that tries to separate the metrics themselves from the implemented AI-method by introducing additional trusted reference quality metrics. They are also calculated geometrically from the landmarks, but based on landmark positions that are manually determined by some of the authors in the images, so basically in a pen and paper act without any machine learning involved. The inhalation reference quality metrics was determined by counting ribs. The study cases were carefully sampled with interesting borderline cases being over-represented, almost a thousand cases in total.

The main result of the study is that when sorting the study cases based upon their reference quality metrics this separates the quality classes as given by the experts, even when all their ratings are pooled. Especially for the strong binary decision on 'Ok' versus 'Unacceptable' there was actually not much overlap of these two classes. Thresholds could be established to do such an optimal separation by each of the quality metrics based on the empirical data of all raters. Two of the seven raters did not choose 'too wide, unacceptable' at any lateral Field-of-View case at all and explained that later by the fact that they do not see a strong negative effect in extending the radiated area into the empty space to the patient side. Considering that, both quality metrics with their corresponding thresholds $\Delta_{W,w}$ and $\Delta_{E,w}$ appear not as relevant as the other ones.

A limitation of the study design lies in the annotation of rib counts by integer values. It restricts possible reasonable thresholds on c_{max} for Inhalation practically to be set either between eight and nine or between nine and ten ribs. The latter resulted from the raters' votes, very much in line with the guideline. A threshold of 9.5 ribs resulted in a rejection rate of 36% on the entire data set. The study results would justify any threshold between nine and ten ribs and using 9.0 ribs reduces the rejection rate to 24.8%.

When using the AI-method as a reader it showed high AUC values for classifying all study cases rated ok or unacceptable. Its performance is similar to what could be achieved when applying the determined common thresholds on the trusted reference metrics to reproduce all pooled expert rates ok or unacceptable. When comparing these values it should be considered that the particular impacting factor for the expert group is disagreement among each other while for the AI-method it is the inaccuracies in landmark localizations with respect to those of the trusted reference landmarks.

When applying the empiric thresholds based on guidelines and expert opinions to the entire data set of 64,315 cases, this gives percentages of cases that fail the corresponding quality criteria. For the too narrow Field-of-View this is 6.6%, for the too wide Field-of-View it is 61.1%. For Rotation it is 4.1%, and 24.8% for Inhalation. These numbers fit well into the variety found in earlier studies. Also, in our large data set the majority of diagnosed cases fail at least one quality criterion. This shows a high discrepancy between expectations and real practice in positioning quality with implications on dose, potentially missed diagnosis, and additional workload. That turned into a positive statement, we see a big potential for improvements in radiology departments that may be addressed by automated quality feedback and control on an every acquisition basis. Further studies are planned to prove practical appropriateness of the AI-method with the given thresholds in a real radiology setting on the everyday image data at different institutions.

Conclusion

Thresholds on quality metrics were determined in a reader study that allow reproducing expert classifications between ok and unacceptable for three relevant patient positioning criteria in chest radiography: collimation, patient rotation, and inhalation. An AI-method was introduced and validated that is able to estimate these metrics from the image according to the guidelines. Applying the threshold to these metrics achieves automatic but comprehensible quality classification for quality management that is consistent with expert judgements.

Authors' contribution: J.v.B., K.H., N.W., T.H., S.Y., H.L. contributed to the conception and design of the work, J.v.B., K.H., M.E., D.P., N.W., S. K.-H., T.H., S.Y., H.L. contributed to data collection, J.v.B., K.H., N.W., S.K.-H., T.H., A.G., D.B., M.B., S.Y., H.L. contributed to data analysis, while all authors contributed to manuscript writing. All authors

reviewed the final version of the manuscript and agreed to submit it to IMAGING for publication.

Conflict of interest: J.v.B, N.W, S.K.-H, T.H, A.G, D.B, M.B, and S.Y are all employees of Philips.

Funding sources: D.P. was funded by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder, the German Research Foundation (GRK2274), as well as by the Technical University of Munich–Institute for Advanced Study.

Ethical statement: The study has been conducted in accordance with the Declaration of Helsinki and according to requirements of all applicable local and international standards.

REFERENCES

[1] Foos DH, Sehnert WJ, Reiner B, Siegel EL, Segal A, Waldman DL: Digital radiography reject analysis: data collection methodology, results, and recommendations from an in-depth investigation at two hospitals. J Digit Imaging 2009; 22(1): 89–98.

[2] American College of Radiology (ACR): ACR-SPR practice parameter for the performance of chest radiography Res. 56-2011, amended 2014 (Res. 39).

[3] Carmichael JHE: European guidelines on quality criteria for diagnostic radiographic images. Office for Official Publications of the European Communities, 1996.

[4] Grewal RK, Young N, Collins L, Karunaratne N, Sabharwal R: Digital chest radiography image quality assessment with dose reduction. Australas Phys Eng Sci Med 2012; 35(1): 71–80.

[5] Tesselaar E, Dahlström N, Sandborg M: Clinical audit of image quality in radiology using visual grading characteristics analysis. Radiat Prot Dosim 2016; 169(1–4): 340–6.

[6] Little KJ, Reiser I, Liu L, Kinsey T, Sánchez AA, Haas K, et al.: Unified database for rejected image analysis across multiple vendors in radiography. J Am Coll Radiol 2017; 14(2): 208–16.

[7] Alpert HR, Hillman BJ: Quality and variability in diagnostic radiology. J Am Coll Radiol 2004; 1(2): 127–32.

[8] Reiner BI: Automating quality assurance for digital radiography. J Am Coll Radiol 2009; 6(7): 486–90.

[9] Whaley JS, Pressman BD, Wilson JR, Bravo L, Sehnert WJ, Foos DH: Investigation of the variability in the assessment of digital chest X-ray image quality. J Digit Imaging 2013; 26(2): 217–26.

[10] Kjelle E, Schanche AK, Hafskjold L: To keep or reject, that is the question-a survey on radiologists and radiographers' assessments of plain radiography images. Radiol 2021; 27(1): 115–9.

[11] Kjelle E, Chilanga C: The assessment of image quality and diagnostic value in x-ray images: a survey on radiographers' reasons for rejecting images. Insights imaging 2022; 13(1): 1–6.

[12] Nousiainen K, Mäkelä T, Piilonen A, Peltonen JI: Automating chest radiograph imaging quality control. Physica Medica 2021; 83: 138–45.

[13] Dasegowda G, Bizzo BC, Gupta RV, Kaviani P, Ebrahimian S, Ricciardelli D, et al.: Radiologist-trained AI model for identifying suboptimal chest-radiographs. Acad Radiol 2023.

[14] Berg JV, Krönke S, Gooßen A, Bystrov D, Brück M, Harder T, et al.: Robust chest X-ray quality assessment using convolutional neural networks and atlas regularization. In: *Medical imaging 2020: image processing*; 2020. Vol. 11313:113131L. International Society for Optics; Photonics.

[15] Meng Y, Ruan J, Yang B, Gao Y, Jin J, Dong F, et al.: Automated quality assessment of chest radiographs based on deep learning and linear regression cascade algorithms. Eur Radiol 2022: 1–11.

[16] Poggenborg J, Yaroshenko A, Wieberneit N, Harder T, Gossmann A: Impact of AI-based real time image quality feedback for chest radiographs in the clinical routine. medRxiv 2021; 2021–06.

[17] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM: Chestx-Ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2097–106.