European Radiology (2019) 29:3705–3713
https://doi.org/10.1007/s00330-019-6013-6

CHEST

Potential of a machine-learning model for dose optimization in CT quality assurance

Axel Meineke[1] · Christian Rubbert[2] · Lino M. Sawicki[2] · Christoph Thomas[2] · Yan Klosterkemper[2] · Elisabeth Appel[2] · Julian Caspers[2,3] · Oliver T. Bethge[2] · Patric Kröpil[2] · Gerald Antoch[2] · Johannes Boos[2]

Received: 30 November 2018 / Accepted: 17 January 2019 / Published online: 19 February 2019

Abstract

Objectives To evaluate machine learning (ML) to detect chest CT examinations with dose optimization potential for quality assurance in a retrospective, cross-sectional study.

Methods Three thousand one hundred ninety-nine CT chest examinations were used for training and testing of the feed-forward, single hidden layer neural network (January 2016–December 2017, 60% male, 62 ± 15 years, 80/20 split). The model was optimized and trained to predict the volumetric computed tomography dose index ($CTDI_{vol}$) based on scan patient metrics (scanner, study description, protocol, patient age, sex, and water-equivalent diameter (D_W)). The root mean-squared error (RMSE) was calculated as performance measurement. One hundred separate, consecutive chest CTs were used for validation (January 2018, 60% male, 63 ± 16 years), independently reviewed by two blinded radiologists with regard to dose optimization, and used to define an optimal cutoff for the model.

Results RMSE was 1.71, 1.45, and 1.52 for the training, test, and validation dataset, respectively. The scanner and D_W were the most important features. The radiologists found dose optimization potential in 7/100 of the validation cases. A percentage deviation of 18.3% between predicted and actual $CTDI_{vol}$ was found to be the optimal cutoff: 8/100 cases were flagged as suboptimal by the model (range 18.3–53.2%). All of the cases found by the radiologists were identified. One examination was flagged only by the model.

Conclusions ML can comprehensively detect CT examinations with dose optimization potential. It may be a helpful tool to simplify CT quality assurance. CT scanner and D_W were most important. Final human review remains necessary. A threshold of 18.3% between the predicted and actual $CTDI_{vol}$ seems adequate for CT quality assurance.

Key Points

- *Machine learning can be integrated into CT quality assurance to improve retrospective analysis of CT dose data.*
- *Machine learning may help to comprehensively detect dose optimization potential in chest CT, but an individual review of the results by an experienced radiologist or radiation physicist is required to exclude false-positive findings.*

Keywords Radiation dosage · Multidetector computed tomography · Thorax · Quality assurance, health care · Machine learning

Abbreviations

$CTDI_{vol}$	Volumetric computed tomography dose index
DLP	Dose length product
DRLs	Diagnostic reference levels
D_W	Water-equivalent diameter
ML	Machine learning
QA	Quality assurance
RMSE	Root mean-squared error

Axel Meineke and Christian Rubbert contributed equally to this work.

✉ Christian Rubbert
christian.rubbert@med.uni-duesseldorf.de

[1] Cerner HS Deutschland GmbH, 13629 Berlin, Germany

[2] Department of Diagnostic and Interventional Radiology, Medical Faculty, University Dusseldorf, 40225 Dusseldorf, Germany

[3] Research Centre Jülich, Institute of Neuroscience and Medicine (INM-1), 52425 Jülich, Germany

Introduction

The number of computed tomography (CT) examinations has increased during the last decades, and CT accounts for a large amount of medical radiation dose exposition. Thus, CT dose management is an important part of CT quality assurance.

CT dose is typically reported as volumetric computed tomography dose index ($CTDI_{vol}$) and dose length product (DLP). Various countries and institutions, e.g., the American College of Radiologists (ACR), provide diagnostic reference levels (DRLs) for specific anatomic regions or CT protocols [1–3]. While national DRLs can be used for overall benchmarking, additional institutional CT dose analysis is an important part of CT quality assurance [4–7]. Apart from the patient size, institutional specifics, such as different CT scanner hardware and post-processing software, especially scanner-adjusted CT protocols and the local clinical setup may influence the dose of CT examinations [8]. Detection of CT dose optimization potential and dose outliers may be cumbersome and difficult.

To perform a comprehensive CT quality assurance, typically a large amount of data has to be screened, which is time exhaustive and staff intensive [3, 7, 9]. Modern CT dose management systems and automatically acquired patient-specific parameters such as the water-equivalent patient diameter may help to improve CT dose analysis [1, 6, 10]; however, analysis of large numbers of CT examinations remains challenging.

Machine learning is used as an umbrella term for statistical models, which are trained based on existing data in order to generalize on previously unseen data [11]. Machine learning models are able to detect patterns in large datasets, which, amongst other potential applications, can be used to detect outliers [12, 13]. Machine learning has been evaluated for different aspects of radiology including lung nodule detection, outcome prediction, and CT dose optimization [11, 14, 15].

We hypothesized that machine learning may be beneficial for retrospective dose analysis as part of CT quality assurance to detect examinations with potential for dose optimization. To our knowledge, the benefit of machine learning to assess large CT dose data sets as part of CT quality assurance has so far not been investigated.

Therefore, the aim of our study was to evaluate the feasibility of a neural network for detection of dose optimization potential and dose outliers in quality assurance of chest CT.

Materials and methods

Study design

This retrospective, cross-sectional study was approved by the local ethics committee. The requirement for written informed consent was waived.

A feed-forward neural network regression model with a single hidden layer was trained to predict an expected $CTDI_{vol}$ value considering a variety of CT examination parameters and patient parameters in order to detect CT chest examinations with potential for dose optimization and dose application errors. A total of 3199 CT chest examinations acquired between January 2016 and December 2017 (60.1% male, mean age 62.4 ± 15.2 years) were used for model training and testing. A stratified random split, based on the $CTDI_{vol}$, into a training (80%, $n = 2561$) and a test dataset (20%, $n = 638$) was performed.

The trained model was further validated using 100 independent, consecutive CT chest examinations acquired at our institute in January 2018 (60% male, aged 62.8 ± 16.1). Two radiologists independently reviewed the examinations for dose optimization potential and dose application errors.

CT-guided interventions, pediatric patients, and dual-energy CT examinations were excluded (Fig. 1).

The characteristics of the patient population (shown in Table 1) were plotted using histograms (not shown) and compared between the training, test, and validation dataset using pairwise, unpaired Wilcoxon tests for non-normally distributed characteristics (i.e., skewed, $CTDI_{vol}$, and DLP) as well as t tests for normally distributed characteristics (age and D_W), each with pooled standard deviation and Bonferroni correction for multiple testing, and the χ^2 test for count statistics (sex, kVp, and scanner). Statistical significance is set to $p < 0.05$.

CT examinations

CT examinations were performed on four different CT scanners (scanner 1 = Definition Edge, scanner 2 = Definition Flash, scanner 3 = Definition AS+, and scanner 4 = Definition AS with sliding gantry; Siemens Healthineers). Automated tube current modulation and automated tube voltage selection (CareDose 4D and CareKV, Siemens Healthineers) were used in all examinations. All protocols that were assigned to the scan region "chest" in the Digital Imaging and Communications Radiation Dose Structured Report were included in the study. Images were stored in the local Picture Archiving and Communication System (PACS, IDS7, Sectra) in clinical routine. Dose data from the DICOM-radiation dose structured report is automatically stored in a local dose management system (DoseIntelligence, Pulmokard).

Patient diameter measurements

Water-equivalent patient diameter (D_W) was automatically calculated for all patients according to the recommendations of the American Association of Physicists in Medicine Task Group 220, based on the mean D_W from all slices of the CT

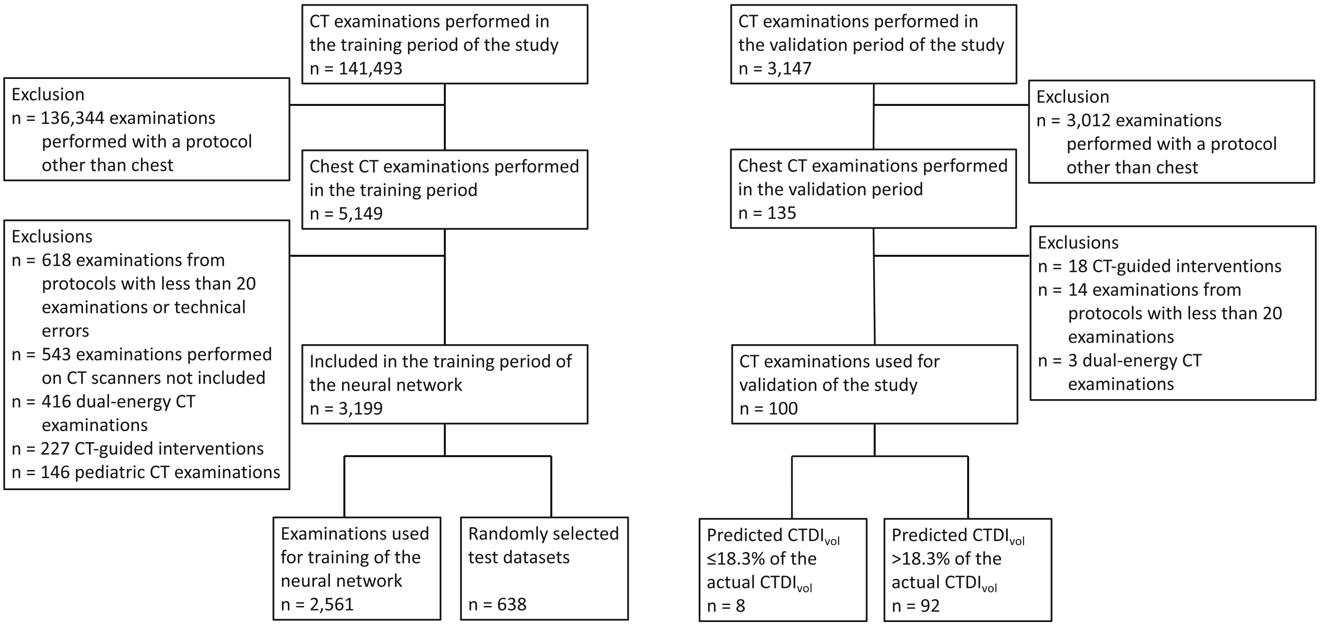

Fig. 1 Standards for Reporting of Diagnostic Accuracy Studies (STARD) flowchart of the patient collective

Table 1 Patient and examination characteristics and $CTDI_{vol}$ of the training, test, and validation patient collective. In case of a skewed distribution (plotted histograms not shown), median and interquartile ranges are reported. Otherwise, mean and standard deviation are shown. Values were not significantly different between the collectives (every $p > 0.05$), except for the scanners ($p = 0.0048$), since the Definition AS+ CT scanner was not used during the validation period of the study, and for kVp ($p = 0.0456$)

Population characteristics						
	Training		Test		Validation	
	n = 2561		*n* = 638		*n* = 100	
Sex (m/f)	1531 (59.8%)	1030 (40.2%)	390 (61.1%)	248 (38.9%)	60 (60%)	40 (40%)
	mean ± SD	range	mean ± SD	range	mean ± SD	range
Age (years)	62.3 ± 15.1	18.2–99	62.5 ± 15.3	18.4–97.2	62.8 ± 16.1	19.5–89.7
D_W (mm)	278.5 ± 34.0	78.1–427.4	276.2 ± 33.1	189.5–413.8	279.2 ± 36.5	186.9–386.4
	Median (25th–75th quartile)	Range	Median (25th–75th quartile)	Range	Median (25th–75th quartile)	Range
$CTDI_{vol}$ (mGy)	4.4 (3.1–6.6)	0.7–25.7	4.6 (3.1–6.7)	1.3–22.6	3.9 (2.9–6.4)	1.4–20.9
DLP (mGy*cm)	141.3 (97.1–215.6)	4.5–1378.2	146.0 (99.7–217.1)	14.7–983.0	126.2 (97.3–201.5)	20.3–673.9
	n (%)		*n* (%)		*n* (%)	
kVp						
80	40 (1.6%)		9 (1.4%)		0 (0%)	
100	1778 (69.4%)		458 (71.8%)		59 (59%)	
120	743 (29.0%)		1971(26.8%)		41 (41%)	
Scanner (SOMATOM Definition ...)						
Flash	969 (37.8%)		262 (41.1%)		44 (44%)	
AS	638 (24.9%)		148 (23.2%)		25 (25%)	
Edge	620 (24.2%)		143 (22.4%)		31 (31%)	
AS+	334 (13%)		85 (13.3%)		–	

$CTDI_{vol}$ volumetric computed tomography dose index, *D_W* patient water-equivalent diameter, *AS* AS with sliding gantry

scan volume, using an in-house developed MATLAB (version R2014a, MathWorks Inc.) application [16, 17].

Training of the neural network

A feed-forward neural network regression model with a single hidden layer was trained on the training dataset to predict $CTDI_{vol}$. Hyperparameter optimization and training was done using R v3.5.1, caret v6.0-80 and nnet v7.3-12. Input parameters were patient age, sex, D_W, acquisition protocol, study description, and CT scanner (Table 1). Patient age and D_W were centered and scaled. There was no missing data.

A random search with 100 runs was used for hyperparameter optimization [18]. Resampling was done using bootstrapping with 25 repetitions. Weight decay was applied. The optimal model was selected based on the root mean-squared error (RMSE). The maximum number of iterations was set to 500. The RMSE for the training dataset was 1.71 using six units in the hidden layer and a weight decay of 5.228329. To assess the model's performance, the $CTDI_{vol}$ of the previously unseen test dataset was predicted and compared to the actual $CTDI_{vol}$.

The importance of each input feature was determined based on Garson weights [19].

Validation of the model

Two radiologists (JB and LM with 6 and 5 years of experience in CT), blinded to the results of the machine learning model, independently reviewed the 100 examinations in the validation dataset with regard to potential for CT dose optimization and possible dose application errors.

To reflect the institutional quality assurance (QA) procedure, the radiologists reviewed $CTDI_{vol}$; relative $CTDI_{vol}$ compared to the national diagnostic reference levels (10 mGy for CT chest [20]); usage of automated tube current modulation and tube voltage adaption; and patient information such as age, sex, D_W, the CT scanner, reconstruction parameters, patient position in the gantry, and the selected CT protocol. As part of the QA review, the images were reviewed using the PACS. Overall image quality was rated on a 5-point Likert scale (5 = excellent, optimal image quality, very low image noise, no artifacts, optimal delineation of pulmonary and soft tissue structures; 4 = good image quality, low image noise, minimal artifacts, good delineation of pulmonal and soft tissue structures; 3 = moderate image quality, moderate image noise, image artifacts without impairment of image interpretation, slightly reduced delineation of pulmonary and mediastinal structures; 2 = poor image quality, high noise, severe artifacts, partial impairment of image interpretation; 1 = non-diagnostic image quality, image interpretation cannot be performed).

Dose application errors were defined as a $CTDI_{vol}$ of more than twice the national DRL (i.e., 20 mGy) in cases without reasons for increased dose (e.g., very obese patients). Suboptimal dose application was defined as examinations with suboptimal scan parameters or mismatch between the different evaluated parameters. In cases of potential for dose optimization or dose application errors, the reason was noted.

Both reviewers used the composite outcome to screen for dose application errors and CT examinations with potential for dose optimization and independently formed a decision for each examination.

The percentage deviation between predicted and actual $CTDI_{vol}$ was calculated for all examinations:

$$\frac{(predicted - actual)}{|actual|} \times 100$$

Based on the classifications of the radiologists, an optimal cutoff of the percentage deviation between predicted and actual $CTDI_{vol}$ was determined according to Youden's Index in order for the model to identify examinations with dose optimization potential [21]. Examinations detected by both the radiologists and the model, by the radiologists only, and by the model only were analyzed in detail. RMSE was calculated for the validation dataset, as well.

Results

Patients

Characteristics of the training, test, and validation dataset are shown in Table 1. There was no significant difference between patient and examination parameters between the different datasets ($p > 0.05$ for all), except for the scanners ($p = 0.0013$), since scanner 3 was not used during the validation period of the study, and kVp ($p = 0.0456$).

Comparison of predicted and actual $CTDI_{vol}$

The predicted $CTDI_{vol}$ was 5.3 ± 3.0 mGy (range 1.4–21.7 mGy) compared to 5.3 ± 3.2 mGy (1.3–22.6 mGy) for the actual $CTDI_{vol}$ in the test dataset and 5.2 ± 3.1 mGy (range 1.0–16.7 mGy) predicted $CTDI_{vol}$ compared to 4.9 ± 3.0 mGy (range 1.4–20.9 mGy) for the actual $CTDI_{vol}$ in the validation dataset. The RMSE in the test dataset was 1.45 and 1.52 in the validation dataset (Fig. 2).

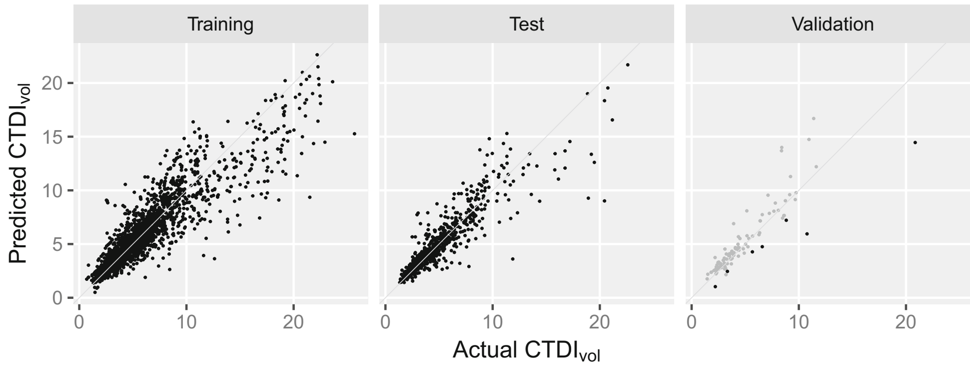

Fig. 2 Predicted versus actual CTDI$_{vol}$ in the training (RMSE = 1.71), test (RMSE = 1.45), and validation (RMSE = 1.52) dataset. In the graph for the validation dataset, examinations flagged for potential for dose optimization are shown in black, the others in gray

Analysis by the radiologists

According to our definition, none of the examinations entailed dose application errors. 7/100 (7%) of the CT examinations were reported to have potential for dose optimization by the radiologists. Both radiologists reported the same seven cases.

On average, image quality was 4.3 ± 0.72 (range 3–5). The seven examinations found to have potential for dose optimization by the radiologists had an image quality of 4.4 ± 0.7 (range 3–5).

Determination of the optimal cutoff and analysis of the cases

The actual CTDI$_{vol}$ exceeded the predicted CTDI$_{vol}$ in 30/100 cases (30%). The optimal threshold was found to be − 18.3% (area under the receiver operator characteristic curve = 0.992), i.e., a predicted CTDI$_{vol}$ which is at least 18.3% lower than the actual CTDI$_{vol}$ is optimal to detect CT examinations with dose optimization potential when using the model.

When applying the 18.3% threshold, 8/100 (8%) of the cases were found to have potential for dose optimization by the model (62.5% male, aged 59.1 ± 21.6 years, Fig. 3). These cases included a wide range of D_W (218.6–330.8 mm, mean 267.2 ± 35.5 mm) and actual CTDI$_{vol}$ (2.2–20.9 mGy, mean 7.7 ± 6.1 mGy). The predicted CTDI$_{vol}$ in cases exceeding the 18.3% threshold ranged from 1 to 14.4 mGy (5.3 ± 4.2 mGy) with a tube voltage of 100 and 120 kVp in 4/8 (50%) of the cases, respectively. The examinations were acquired on three scanners (scanner 1 = 1/8 (12.5%), scanner 2 = 3/8 (37.5%), scanner 4 = 4/8 (50%)). Scanner 3 was not used during the period of the validation scans. CT protocols with potential for dose optimization were CT chest without contrast (5/8, 62.5%), CT chest with and without contrast (1/8, 12.5%), a venous phase CT chest (1/8, 12.5%), and a venous phase CT chest additionally performed as part of a CT stroke imaging protocol (1/8, 12.5%).

Mean CTDI$_{vol}$ of the examinations as detected by the model was 76.8% ± 60.8% of the national DRLs (range 22–209%) with only 2/8 (25%) of these CT examinations exceeding the national DRLs (Fig. 4).

In the seven cases detected by both the model and the radiologists, the actual CTDI$_{vol}$ was significantly higher than the predicted CTDI$_{vol}$ (8.3 ± 6.3 mGy vs. 5.7 ± 4.4 mGy, p = 0.01834). Patients were 22.2–31.1 years of age and D_W ranged from 218 to 330 mm. In 4/7 cases flagged by the model and the radiologists, a

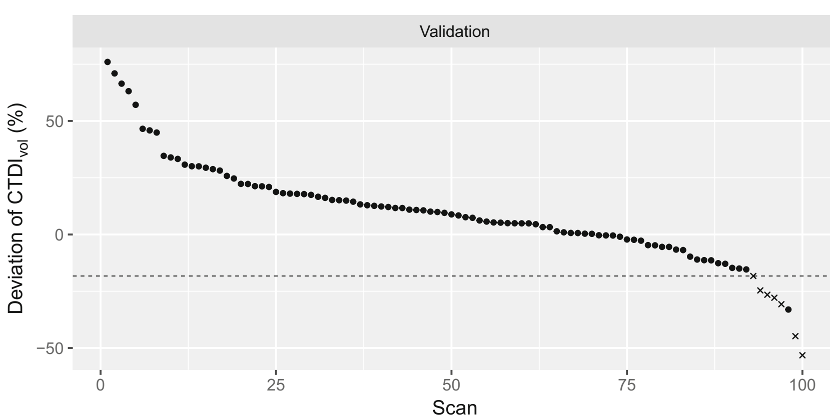

Fig. 3 Percentage deviation between actual and predicted CTDI$_{vol}$. Scans are sorted by the percentage deviation. Examinations with dose optimization potential as determined by the two radiologists are shown as crosses. All but one of the CT examinations determined negative by the machine learning model (deviation of more than 18.3%, indicated by the gray dashed line) were found to be inadequate by the radiologists. This demonstrates the additional value of ML for CT quality assurance

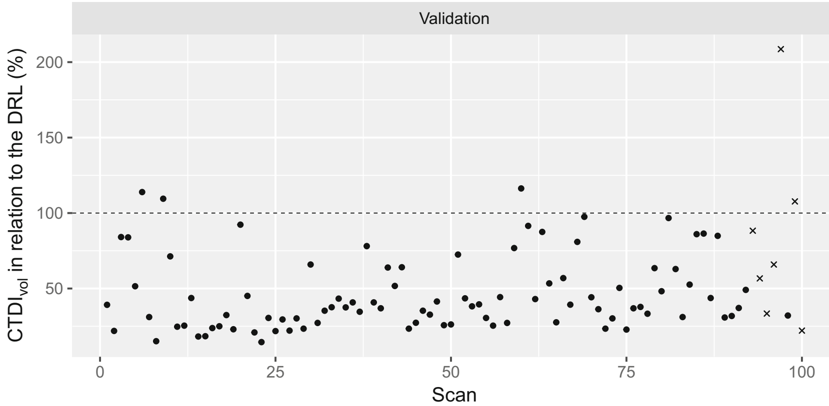

Fig. 4 Comparison of actual $CTDI_{vol}$ to the nation DRLs (10 mGy, marked by the gray dashed line). The same order of the examinations as in Fig. 3 was used. Examinations with dose optimization potential as determined by the two radiologists are shown as crosses. Of note, only two CT scans that were found to be suboptimal by the model actually exceeded national DRLs while five CT scans deemed insufficient by the model were within the DRLs

suboptimal tube voltage of 100 kVp (2/7) or 120 kVp (2/7) was used in very slim patients (D_w of 21.8–27.7 cm). In 2/7 patients, the suboptimal $CTDI_{vol}$ was caused by an insufficient patient position with arms next to the chest. In neither patient a reason for arm positioning had been documented. In 1/7 patient, a chest CT was performed as part of a stroke imaging and the technicians did not select the adequate chest CT protocol.

A single case was flagged by the ML-model only. In this case, a 63-year-old female patient underwent non-contrast chest CT for evaluation of pneumothoraces on scanner 2. The predicted $CTDI_{vol}$ was 33.1% lower than the actual $CTDI_{vol}$ (2.15 vs. 3.21 mGy). A tube voltage of 120 kVp had been selected. The patient suffered from severe soft tissue emphysema (Fig. 5), and scan settings and dose were regarded as adequate by both radiologists.

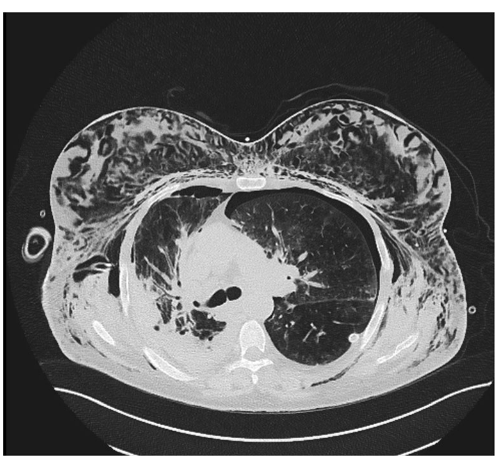

Fig. 5 63-year-old female patient that underwent a non-contrast chest CT examination for evaluation of pneumothoraces. The CT examination was performed with 120 kVp and a $CTDI_{vol}$ of 3.21 mGy. The machine-learning model predicted a $CTDI_{vol}$ of 2.15 mGy (deviation of − 33.1% from the established threshold). The examination was not flagged by the radiologists who found the $CTDI_{vol}$ to be adequate

Feature importance

The calculation of feature importance showed that the CT scanner (mean score of 6.1 ± 1.0) and D_W (4.6) were most important. The acquisition protocol (2.5 ± 1.6), sex (1.8), the study description (0.9 ± 0.5), and age (0.4) were found to be less important.

Discussion

In this study, we demonstrated the feasibility of using a neural network regression model for large-scale CT dose data analysis to detect inadequate and suboptimal CT dose application. The analysis demonstrates the value of a machine learning-based process in the analysis of a large number of CT examinations. Multiple variables and scan parameters can be taken into account to assess whether or not a CT scan was performed suboptimally. The determined feature importance demonstrates the importance of the characteristics of the CT scanner and of the patient water-equivalent diameter (D_W) in CT dose data analysis.

Neural networks are inspired by theories on how the brain works, computationally representing the intake, processing, and output of data on multiple layers. Each neuron has a forward connection to the subsequent layer. The input layer connects to units in the hidden layer, which are then linearly combined to the output neuron in this single-layer approach, which gives the results of the prediction. The units in the hidden layer are using nonlinear, i.e., sigmoidal, activation functions in a classification setting, and linear activation functions in a regression model. It has been shown that network-based regression models may capture complex relationships between variables better than more classical regression-based methods, which are still widely used in research and clinical routine [22]. One of the main concerns in machine learning is the problem of over-fitting the model to the training data. However, by using weight decay, the comparable RMSE over the training, test, and validation dataset indicate that no relevant overfitting is present.

ML models are currently of great interest to improve different aspects of modern radiology, such as cancer prediction in patients with incidental lung nodules [23], radionomics [24], identification of novel biomarkers [25] or cardiac-CT-derived fractional flow reserve assessment [26]. Furthermore, ML can be used beyond image interpretation, e.g., to predict outcome [11] or as a tool for quality assurance [27, 28].

We demonstrated feasibility of ML for the detection of dose optimization potential in a large dataset as part of CT quality assurance. CT dose optimization has been in the focus of interest for some years and various dose management systems have been introduced [29]. Although typically a large number of CT examinations, each with multiple scan- and dose parameters, have to be analyzed, to our knowledge, the value of machine learning for CT dose optimization potential has not been investigated. Previous works, for example by Smith-Bindman et al, have used intricate statistical modeling to evaluate the influence of different patient and vendor parameters on dose (most important: patient size), but have not applied predictive modeling to generalize on previously unseen data [8].

Our model was able to predict $CTDI_{vol}$ for a single examination by taking into account the CT scanner, the CT protocol, and patient-specific parameters such as age, sex and D_W. The performance measurement RMSE, 1.4–1.7 for the different datasets, has to be interpreted in the context of the range of the dependent variable (overall 0.7–25.7) and was therefore considered “good”. A threshold of 18.3% deviation between predicted (expected) and actual $CTDI_{vol}$ was found to be optimal for detecting examinations with suboptimal dose application. However, carefully reviewing the threshold led to one additional case which was flagged incorrectly by the ML model. Thus, a machine-learning model may help to screen large CT dose datasets by reducing the workload on the staff for CT quality assurance. A future study manually reviewing a larger number of examinations, thereby increasing the number of dose outliers as detected by the radiologists, should be performed to confirm or refine this threshold. Even though the neural network comprehensively detected the examinations with potential for dose optimization, a final human review remains mandatory. This is in accordance with previous studies assessing machine learning models in radiology, such as in the detection of stroke [30].

The CT scanner and D_W were found to be the parameter with the highest importance in the ML model. Currently, national or institutional DRLs are frequently used for benchmarking of CT dose parameters [1, 5–7, 31]. Such DRLs may help to detect severe dose outliers and systematic inadequate CT dose applications. Because typical DRLs are provided for a normally sized patient, the value to detect individual inadequate dose applications is limited, which especially applies to underweight patients where inadequately high doses may not lead to transgression of DRLs. Recently, size-specific DRLs have been introduced, which allow for a more detailed analysis of CT dose application [1, 5, 6, 10]. The value of size-specific DRLs is in accordance with our results indicating that D_W is of high importance to detect CT dose optimization potential when using an ML model. Besides the advantage of considering patient size, many other factors that may reflect inadequate dose application cannot be detected by simply using DRLs for CT quality assurance. Of note, we found the CT scanner to be of high importance in the ML model. $CTDI_{vol}$ across different CT scanners may vary, e.g., in CTs with or without statistical or model-based iterative reconstruction which is automatically incorporated into our ML model analysis. Our results demonstrate that ML models can be used to improve CT dose analysis as various parameters can be integrated into a single model.

Our study has limitations. First, the study was performed retrospectively and a review of only 100 CT examinations was performed subjectively. However, this reflects clinical routine of CT quality assurance, since manual review is staff intensive. To screen for dose optimization potential, thresholds such as national or institutional DRLs are typically used, which may be inadequate in individual cases. Second, given the promising results by the simple feed-forward neural network with a single hidden layer, we chose not to evaluate other, more intricate machine learning models. Neural networks are known to perform well in regression tasks, and accuracy benefits from even larger training data [22, 32]. Nevertheless, given larger and probably more heterogeneous datasets with site-specific protocols and more variance in scanner hardware, it might be beneficial to explore other machine learning models. Third, since the study was performed in a single center, the variance in scanner hardware could not be fully addressed. Factors such as scanner age, maintenance intervals/status and number of detectors were not considered in the model, which could limit generalizability of the current model and should be addressed in a larger, multi-center study. Fourth, the training dataset has not been manually reviewed for dose optimization potential or dose application errors, and the results of the manual review of the validation dataset suggest that such outliers are present in the training data. Still, given the large number of studies included in the training dataset, the effect of learning on a somewhat biased training dataset should not have had a great effect on the final model, since a systematic bias is unlikely due to ongoing local quality assurance and review of the protocols. The results of model training, testing and validation suggest that there was no substantial effect on the model by these potential outliers. Potentially, manual review of a full training dataset could increase accuracy in very large, heterogeneous datasets and should be evaluated before applying such a model in clinical routine. Fifth, the machine learning based neural network may take into account various parameters to detect dose optimization potential and the optimal set of parameters might not have been determined, yet. Sixth, the machine learning model can potentially be used to also detect inadequate low CT dose application, which was not part of our study. Last, although automated image quality analysis by machine learning approaches is feasible, image quality of the CT exams was not assessed by our model.

In conclusion, a machine-learning-based neural network can comprehensively detect chest CT examinations with potential for dose optimization. It may be a helpful tool to simplify CT quality assurance, making CT dose data analysis more comprehensive and less staff intensive. A final human review of cases suspicious for suboptimal dose applications remains necessary.

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Johannes Boos.

Conflict of interest The authors of this manuscript declare relationships with the following companies: AM—Cerner HS Deutschland GmbH (employee) and Pulmokard GmbH (consultant).

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
- retrospective
- cross-sectional study
- performed at one institution

References

1. Butler PF, Kanal KM (2018) Diagnostic reference levels for adult patients in the United States. J Am Coll Radiol 15:932–933. https://doi.org/10.1016/j.jacr.2017.12.012
2. Alhailiy AB, Ekpo EU, Ryan EA, Kench PL, Brennan PC, McEntee MF (2018) Diagnostic reference levels for cardiac CT angiography in Australia. Radiat Prot Dosimetry. https://doi.org/10.1093/rpd/ncy112
3. Appel E, Kröpil P, Bethge OT et al (2018) Quality assurance in CT: implementation of the updated national diagnostic reference levels using an automated CT dose monitoring system. Clin Radiol 73: 677.e13–677.e20. https://doi.org/10.1016/j.crad.2018.02.012
4. Roch P, Célier D, Dessaud C, Etard C (2018) Using diagnostic reference levels to evaluate the improvement of patient dose optimisation and the influence of recent technologies in radiography and computed tomography. Eur J Radiol 98:68–74. https://doi.org/10.1016/j.ejrad.2017.11.002
5. Klosterkemper Y, Appel E, Thomas C et al (2018) Tailoring CT dose to patient size: implementation of the updated 2017 ACR size-specific diagnostic reference levels. Acad Radiol 25:1624–1631. https://doi.org/10.1016/j.acra.2018.03.005
6. Boere H, Eijsvoogel NG, Sailer AM et al (2018) Implementation of size-dependent local diagnostic reference levels for CT angiography. AJR Am J Roentgenol 210:W226–W233. https://doi.org/10.2214/AJR.17.18566
7. MacGregor K, Li I, Dowdell T, Gray BG (2015) Identifying institutional diagnostic reference levels for CT with radiation dose index monitoring software. Radiology 276:507–517. https://doi.org/10.1148/radiol.2015141520
8. Smith-Bindman R, Wang Y, Yellen-Nelson TR et al (2016) Predictors of CT radiation dose and their effect on patient care: a comprehensive analysis using automated data. Radiology 282:182–193. https://doi.org/10.1148/radiol.2016151391
9. Demb J, Chu P, Nelson T et al (2017) Optimizing radiation doses for computed tomography across institutions: dose auditing and best practices. JAMA Intern Med 177:810–817. https://doi.org/10.1001/jamainternmed.2017.0445
10. Boos J, Thomas C, Appel E et al (2018) Institutional computed tomography diagnostic reference levels based on water-equivalent diameter and size-specific dose estimates. J Radiol Prot 38:536–548. https://doi.org/10.1088/1361-6498/aaa32c
11. Rubbert C, Patil KR, Beseoglu K et al (2018) Prediction of outcome after aneurysmal subarachnoid haemorrhage using data from patient admission. Eur Radiol 28:4949–4958. https://doi.org/10.1007/s00330-018-5505-0
12. Kim GB, Jung KH, Lee Y et al (2018) Comparison of shallow and deep learning methods on classifying the regional pattern of diffuse lung disease. J Digit Imaging 31:415–424. https://doi.org/10.1007/s10278-017-0028-9
13. Dreyer KJ, Geis JR (2017) When machines think: radiology's next frontier. Radiology 285:713–718. https://doi.org/10.1148/radiol.2017171183
14. Nishio M, Sugiyama O, Yakami M et al (2018) Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. PLoS One 13:e0200721. https://doi.org/10.1371/journal.pone.0200721
15. Chen B, Xiang K, Gong Z, Wang J, Tan S (2018) Statistical iterative CBCT reconstruction based on neural network. IEEE Trans Med Imaging 37:1511–1521. https://doi.org/10.1109/TMI.2018.2829896
16. Anam C, Haryanto F, Widita R, Arif I, Dougherty G (2016) Automated calculation of water-equivalent diameter (DW) based on AAPM task group 220. J Appl Clin Med Phys 17:6171
17. Boos J, Kröpil P, Bethge OT et al (2018) Accuracy of size-specific dose estimate calculation from center slice in computed tomography. Radiat Prot Dosimetry 178(1):8–19
18. Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. J Mach Learn Res 13:281–305
19. Gevrey M, Dimopoulos I, Lek S (2003) Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol Modell 160:249–264. https://doi.org/10.1016/S0304-3800(02)00257-0
20. BfS - diagnostic reference levels. http://www.bfs.de/EN/topics/ion/medicine/diagnostics/reference-levels/reference-levels_node.html;jsessionid=F14BA4ABEE11D08B75B07F8720579B3C.1_cid339. Accessed 16 Aug 2018

21. Feinstein S (1975) The accuracy of diver sound localization by pointing. Undersea Biomed Res 2:173–184
22. Zhang X, Yuan Z, Ji J, Li H, Xue F (2016) Network or regression-based methods for disease discrimination: a comparison study. BMC Med Res Methodol 16:100. https://doi.org/10.1186/s12874-016-0207-2
23. Kadir T, Gleeson F (2018) Lung cancer prediction using machine learning and advanced imaging techniques. Transl Lung Cancer Res 7:304–312. https://doi.org/10.21037/tlcr.2018.05.15
24. Saha A, Harowicz MR, Grimm LJ et al (2018) A machine learning approach to radiogenomics of breast cancer: a study of 922 subjects and 529 DCE-MRI features. Br J Cancer. https://doi.org/10.1038/s41416-018-0185-8
25. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J (2015) Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. Neuroimage 104:398–412. https://doi.org/10.1016/j.neuroimage.2014.10.002
26. Coenen A, Kim YH, Kruk M et al (2018) Diagnostic accuracy of a machine-learning approach to coronary computed tomographic angiography-based fractional flow reserve: result from the MACHINE Consortium. Circ Cardiovasc Imaging 11:e007217. https://doi.org/10.1161/CIRCIMAGING.117.007217
27. Lakhani P, Prater AB, Hutson RK et al (2018) Machine learning in radiology: applications beyond image interpretation. J Am Coll Radiol 15:350–359. https://doi.org/10.1016/j.jacr.2017.09.044
28. Küstner T, Gatidis S, Liebgott A et al (2018) A machine-learning framework for automatic reference-free quality assessment in MRI. Magn Reson Imaging. https://doi.org/10.1016/j.mri.2018.07.003
29. Parakh A, Kortesniemi M, Schindera ST (2016) CT radiation dose management: a comprehensive optimization process for improving patient safety. Radiology 280:663–673. https://doi.org/10.1148/radiol.2016151173
30. Guberina N, Dietrich U, Radbruch A et al (2018) Detection of early infarction signs with machine learning-based diagnosis by means of the Alberta Stroke Program Early CT Score (ASPECTS) in the clinical routine. Neuroradiology. https://doi.org/10.1007/s00234-018-2066-5
31. Brink JA, Miller DL (2015) U.S. national diagnostic reference levels: closing the gap. Radiology 277:3–6. https://doi.org/10.1148/radiol.2015150971
32. Bataineh M, Marler T (2017) Neural network for regression problems with reduced training sets. Neural Netw 95:1–9. https://doi.org/10.1016/j.neunet.2017.07.018