



Open-Access Fully Automated Intravenous Contrast Detection and Body Part Classification for Computed Tomography Scans: The FALCON Model

Julian A. Westphal^{1,2} · Philipp Kaess^{2,3} · Lea Mantz^{2,4} · Rashid A. Barnawi² · Sami Saidi^{2,3} · Matthias P. Fabritius¹ · Alexander Ziller³ · Georg Kaissis³ · Daniel Rueckert³ · Florian J. Fintelmann^{2,5}

Received: 27 July 2025 / Revised: 13 January 2026 / Accepted: 27 January 2026
© The Author(s) 2026

Abstract

Presence of intravenous contrast on computed tomography (CT) scans is often unreliably documented, especially in large research datasets. FALCON is an open-access fully automated deep learning model enabling large-scale intravenous contrast detection and body part classification for CT scans of the head and neck (HN), chest, abdomen, and pelvis (AP). This study used six independent datasets consisting of 3138 CT scans of the HN, chest, and AP of 3126 patients from five institutions between 1996 and 2023 to train and validate four CNN models for intravenous contrast detection and body part classification. The ground truth of intravenous contrast presence was verified by a radiologist. We used ResNet9 network architecture and integrated the four models into a graphical user interface. We assessed FALCON's performance with *F1* scores and compared FALCON's annotation time to manual annotation by human experts. In the external test set containing 1348 scans, the *F1* score for intravenous contrast detection was 99.4% (95%CI: 98.8, 99.9) for HN CT, 98.3% (95%CI: 96.9, 99.5) for chest CT, and 98.1% (95%CI: 96.9, 99.1) for AP CT. The *F1* score for body part classification alone on unseen data was 100% for HN, chest, and AP CT. Compared to human experts, annotation of a single scan with FALCON required 1.3 s vs. 21 s for HN CT, 1.8 s vs. 33 s for chest CT, and 3.7 s vs. 1.6 s for AP CT. The open-access FALCON model (<https://github.com/FintelmannLabDevelopmentTeam/Falcon>) quickly and reliably detects intravenous contrast and classifies body part on CT scans.

Keywords Computed tomography · Intravenous contrast · Body part classification · Deep learning · Open access

Julian A. Westphal and Philipp Kaess shared first authorship.

✉ Julian A. Westphal
julian.westphal@campus.lmu.de

Philipp Kaess
ph.cheese@web.de

Lea Mantz
lmantz@mgh.harvard.edu

Rashid A. Barnawi
rbarnawi@mgh.harvard.edu

Sami Saidi
sami.saidi2000@outlook.com

Matthias P. Fabritius
matthias.fabritius@med.uni-muenchen.de

Alexander Ziller
alex.ziller@tum.de

Georg Kaissis
g.kaissis@tum.de

Daniel Rueckert
daniel.rueckert@tum.de

Florian J. Fintelmann
fintelmann@mgh.harvard.edu

¹ Department of Radiology, LMU Munich, Munich, Germany

² Department of Radiology, Division of Thoracic Imaging and Intervention, Massachusetts General Hospital, Boston, MA, USA

³ Chair for AI in Healthcare and Medicine, Technical University of Munich (TUM), and TUM University Hospital, Munich, Germany

⁴ Department of Diagnostic and Interventional Radiology, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany

⁵ Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA

Introduction

Large-scale computed tomography (CT) datasets are increasingly available for quantitative image analysis including radiomics, body composition analysis, and development of machine learning algorithms. Manually input Digital Imaging and Communications in Medicine (DICOM) metadata is often unreliable and inconsistent, with Gueld et al. reporting that 15.3% of all studies are incorrectly labeled as a result of human error [1–4].

Annotation of CT scans regarding intravenous contrast presence and imaged body part is time consuming. In the absence of a clinical report, a human expert may need to review the DICOM images, an expensive proposition.

Deep learning algorithms can detect intravenous contrast. However, some models require region localization and have not been externally validated [5, 6], while others are not open-access [7, 8]. Similarly, published deep learning algorithms providing automated body part classification are not open-access [9–13]. An open-access tool for automated intravenous contrast detection and body part classification would be valuable for annotation of large-scale datasets of unknown quality [14–19], especially in the research setting.

The purpose of this manuscript is to describe an open-access fully automated tool for intravenous contrast detection and body part classification, named FALCON (Fully Automated Labeling of CT anatomy and intravenous CONTRast), utilizing full volumetric CT scans of the HN, chest, and AP, that can be deployed on standard portable computers. We hypothesize that FALCON will result in significant time savings compared with annotation by human experts.

Materials and Methods

This HIPAA-compliant retrospective multicenter cohort study was approved by the institutional review board at Massachusetts General Hospital, and the need for written informed consent was waived.

Datasets

Six independent, non-overlapping datasets consisting of full volumetric CT scans of the HN, chest, and AP from five institutions and 3126 patients were used. Images were acquired between 1996 and 2023 on 68 different scanner models from six different manufacturers (Table S1). Prior studies have reported on 958 patients included in the current study [20, 21].

Datasets I ($n = 558$) and IV ($n = 598$) consisted of different, independent HN CT scans from 1151 patients made

available through The Cancer Imaging Archive [22, 23] with ground truth contrast labels curated by Ye et al. [24]. Datasets II ($n = 601$) and V ($n = 400$) consisted of chest CT scans [21] from 994 patients, while datasets III ($n = 631$) and VI ($n = 350$) consisted of CT scans of the AP from 981 patients. Two trained research assistants (LM, JAW) and a radiologist with four years of experience (RAB) established ground truth of body part classification and intravenous contrast presence in datasets II, III, V, and VI. Ground truth of datasets I and IV was provided with the datasets and was manually reviewed by residency-trained radiologists. A multi-body part dataset ($n = 1020$), dataset VII, was created by randomly sampling 170 series with and 170 series without intravenous contrast from datasets I, II, and III.

Datasets I, II, III, and VII were used for training, internal validation, and internal testing (Fig. S1). Each dataset was split 80:20 to produce a development and an internal testing subset. Datasets IV, VI, and VI were used for external testing of the final models (Fig. S2). If a patient contributed a scan to the training dataset, their other scans were excluded from testing.

Patient sex, age, and image acquisition parameters were retrieved from DICOM attributes or provided with the dataset (Table 1). Race and ethnicity were self-reported. To compare FALCON with results from Ye et al. [24] and DICOM attributes, we extracted intravenous contrast presence from DICOM solely for comparison.

Pipeline Overview, Preprocessing, and Model Architecture

Three-dimensional volumetric CT data were fed into the model and preprocessed to select a single central axial slice, defined as the middle slice in the series (Fig. 1). For CT scans of the HN, axial 1.25 mm or 3 mm cuts were used; for chest, axial 3 mm cuts; and for AP, 5 mm axial cuts. The selected slice entered a body part classifier sub-algorithm, which selects one of three distinct contrast classification models, one for each body part (HN, chest, AP). Subsequently, the selected slice entered its corresponding contrast detection sub-algorithm to predict intravenous contrast presence with a confidence level. Preprocessing, based on related studies [24–27] and streamlined for computational efficiency (Fig. 2), was utilized in model development and is part of the final model (Fig. 3). The ResNet9 architecture was used for all four models: body part, HN contrast, chest contrast, and AP contrast (Fig. S3).

Model Training

To train the four models, we used a large, specified range (Table 2) of axial slices from the preprocessed CT scans

Table 1 Participant characteristics stratified by dataset

	Dataset I	Dataset II	Dataset III	Dataset IV	Dataset V	Dataset VI	Dataset VII
Body part	Head and neck	Chest	Abdomen and pelvis	Head and neck	Chest	Abdomen and pelvis	Multi-body part
Patient sex							
Female	111	335	278	84	214	139	400
Male	447	266	353	509	179	211	620
Total	558	601	631	593	393	350	1020
Age (years)							
Median	59.8	68.3	68.0	57.0	70.0	61.0	68.0
Mean	60.5	67.0	66.5	57.8	67.6	58.4	64.8
Standard deviation	10.0	10.1	10.9	9.4	14.1	14.5	10.6
Range	33.0–89.2	27.3–91.8	26.0–94.0	24.0–91.0	18.0–96.0	19.0–86.0	27.0–93.0
NA	0	0	0	1	0	8	0
Height (cm)							
Median	–	167.6	169.0	174.0	–	165.0	168.0
Range	–	139.7–200.7	109.0–196.0	146.0–197.0	–	137.0–195.0	139.7–200.7
NA	558	0	270	406	393	8	340
Race*							
Asian	–	27	1	–	73	–	15
Black	–	54	2	–	49	–	32
Caucasian	–	516	6	–	214	–	289
Other	–	2	–	–	54	–	2
Declined or NA	558	2	622	593	3	350	682
Ethnicity*							
Hispanic	–	2	51	–	13	–	23
Not Hispanic	–	362	124	–	354	–	277
Declined or NA	558	237	456	593	26	350	720

Dataset VII was created by randomly sampling 170 series with and 170 series without intravenous contrast from datasets I, II, and III. Dash (–) indicates data was not provided with the dataset. *NA* not available

*Self-reported by the patient

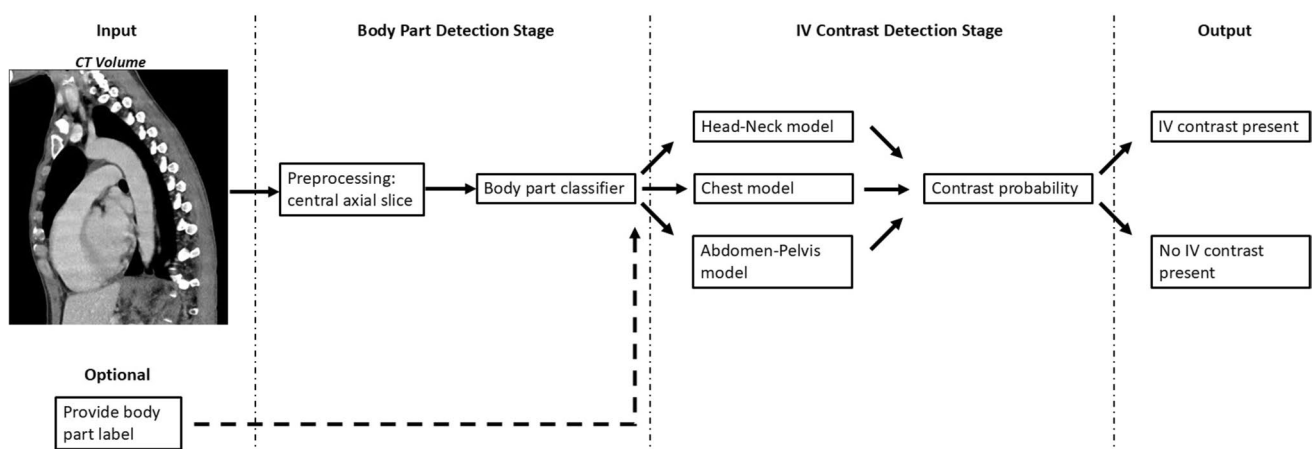


Fig. 1 Overview of the FALCON model. First, a three-dimensional computed tomography volume is fed into the body part detection stage. Second, the CT is preprocessed to yield a single central axial slice. Third, this slice enters the body part classifier, which identifies the image belonging to the head-neck, chest, or abdomen-pelvis

region. Alternatively, body part labels can be provided, and thus, the body part detection stage can be skipped. Fourth, the probability of intravenous (IV) contrast presence is calculated according to each body part model during the IV contrast detection stage. This results in a binary output: intravenous contrast presence or absence

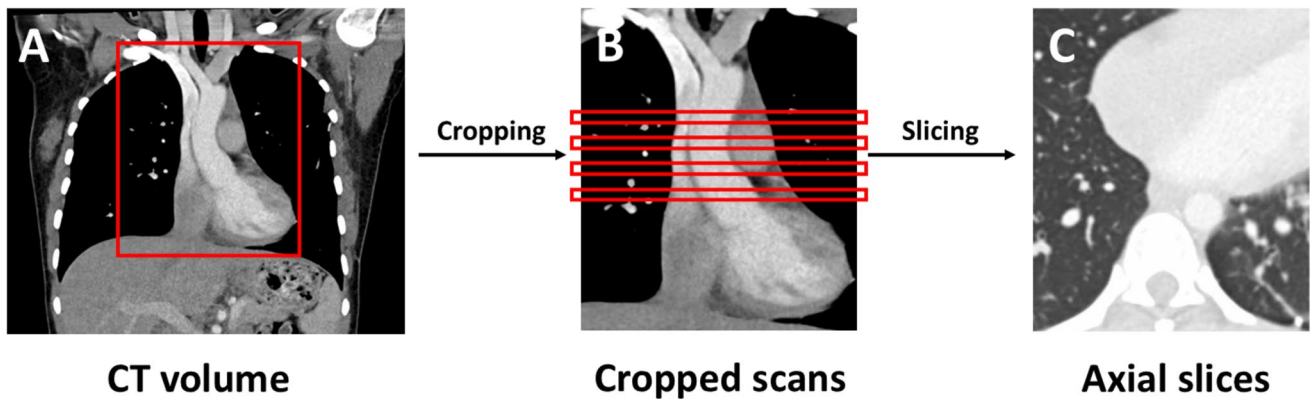


Fig. 2 Preprocessing image selection workflow. Initially, DICOM scans were converted to Hounsfield units, values were clipped, scans were respaced, and the scans were cropped around the center of mass (A) to yield the region of interest (B). This cropped volume is subse-

quently decomposed into its two-dimensional axial slices (C), which are used for training and prediction. The exact slice range used for training and prediction was optimized based on the internal validation set

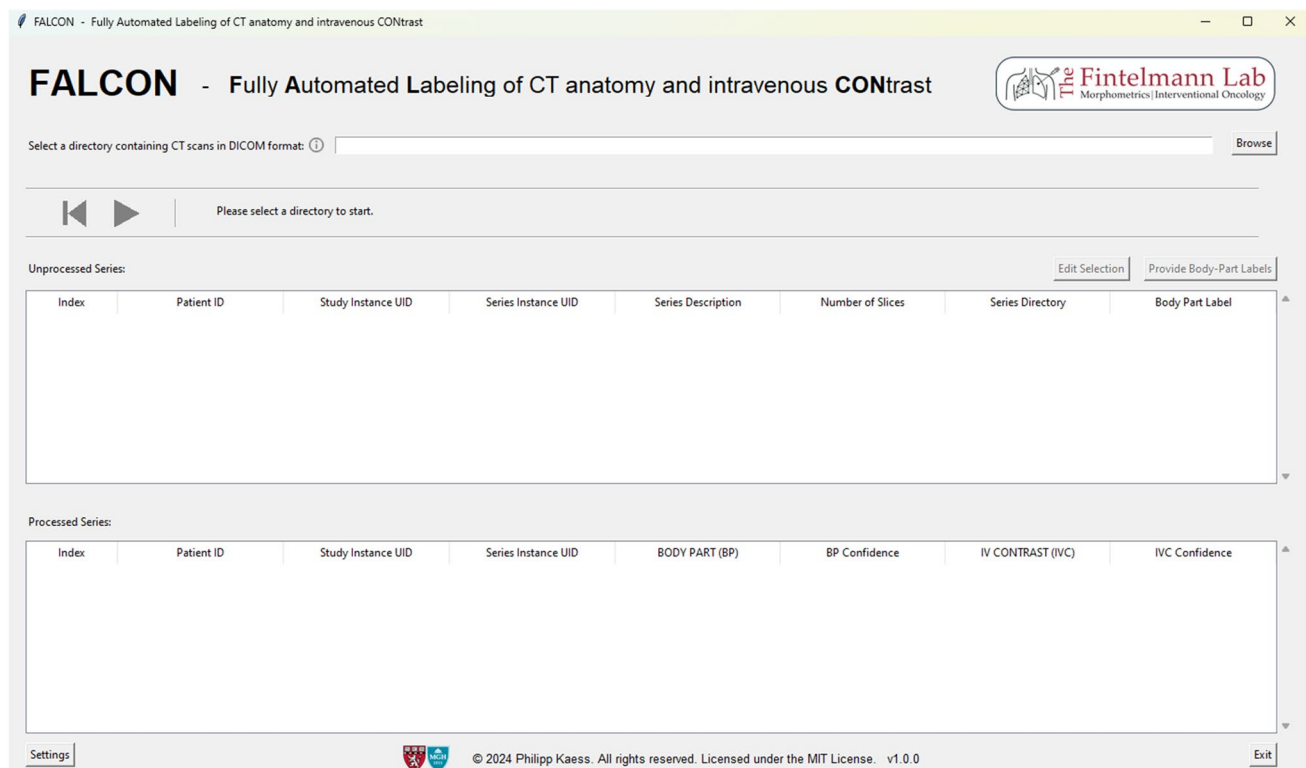


Fig. 3 FALCON (Fully Automated Labeling of CT anatomy and intravenous CONTRast) graphical user interface. Screenshot shows the graphical user interface at startup. A directory containing computed tomography scans can be selected and loaded into the tool. FALCON will provide body part labeling and the probability of intravenous

contrast presence for each CT scan. If the user provides the body part labels of all input scans, the first sub-algorithm can be skipped, and FALCON only provides the probability of intravenous contrast presence

(Fig. S4). Model training utilized high-performance computing clusters to analyze the full volumetric data of each CT scan; details are provided in the Supplementary Information.

Model Performance

We measured the average time FALCON required to assess intravenous contrast presence on a randomly

Table 2 Technical characteristics stratified by dataset

	Dataset I	Dataset II	Dataset III	Dataset IV	Dataset V	Dataset VI	Dataset VII
Body part	Head and neck	Chest	Abdomen and pelvis	Head and neck	Chest	Abdomen and pelvis	Multi-body part
Number of CT scans	558	601	631	598	400	350	1020
Intravenous contrast							
No	172	213	256	194	201	54	590
Yes	386	388	375	404	199	296	430
Number of slices per series							
Median	175.0	138.0	97.0	247.0	135.0	317.5	155.0
Range	130–227	53–430	27–1521	43–549	74–197	69–681	30–979
Slice thickness (mm)							
Median	2.5	3.0	5.0	1.25	2.5	3.0	3.0
Range	2.0–2.5	1.0–5.0	0.5–5.0	1.0–5.0	2.5–3.0	1.0–5.0	0.625–5.0
NA	0	0	0	16	0	0	0
Slice range used for training*	35–75	50–80	20–80	–	–	–	35–65
Slice used for prediction*	55	65	63	–	–	–	50
Pixel spacing							
Median	0.98	0.76	0.86	0.49	0.81	0.92	0.89
Range	0.76–1.17	0.54–1.19	0.58–1.37	0.34–1.07	0.58–1.27	0.60–1.37	0.54–1.37
Tube voltage (kVp)							
Median	120.0	120.0	120.0	120.0	120.0	120.0	120.0
Range	120.0–120.0	80.0–140.0	80.0–140.0	120.0–140.0	80.0–150.0	120.0–140.0	80.0–140.0
NA	0	0	0	17	0	0	0
Manufacturer							
GE Medical Systems	399	169	245	399	212	29	493
Philips	3	141	36	180	35	316	98
Toshiba	156	17	27	10	0	5	109
Siemens	0	98	321	8	153	0	224
Neusoft	0	0	2	0	0	0	0
Hitachi	0	0	0	1	0	0	0
NA	0	176	0	0	0	0	96

Dataset VII was created by randomly sampling 170 series with and 170 series without intravenous contrast from datasets I, II, and III. Dash (–) indicates data was not used for training. NA not available

*Both the slice ranges for training and the index of the slice used for prediction have been optimized on the internal validation set independently for each of the four models

selected scan from the external test set for each body part. In the first part, FALCON calculated the contrast probability with the body part label provided. In the second part, we measured the time for both body part classification and contrast detection on randomly selected scans. All assessments were performed with and without a dedicated GPU (graphics processing unit). Discrepant cases between the algorithmic predictions and the reference contrast labels were analyzed to characterize failure modes.

Graphical User Interface

A graphical user interface (Fig. 3) was developed using Python 3.9.18 to facilitate clinical implementation.

Outcomes

The primary outcome was the accuracy and F1 score of FALCON compared to ground truth for intravenous contrast detection and body part classification. The secondary

outcome was the average time required to annotate a CT scan of the HN, chest, and AP using FALCON compared to data published by Ye et al. [24] and to ground truth data obtained from a radiologist with four years of experience. The reference annotation time per CT scan was obtained from a previously published study [24], which reported the time to manually annotate 1315 HN scans and 664 chest scans by two radiation oncologists with four and seven years of clinical experience, which we refer to as expert clinicians. The reference annotation time per CT scan for AP was obtained by a radiologist with four years of experience through manually annotating 981 scans from Datasets III and VI, preloaded into RadiAnt DICOM viewer version 2025.2 to minimize loading times.

Statistical Analysis

We calculated sensitivity, specificity, *F1* score, overall accuracy, positive predictive value, and negative predictive value for internal and external test sets. Sensitivity and specificity were calculated with a 50% contrast probability threshold, defined during training. For the external test set, mean *F1* score, accuracy, and Matthews Correlation Coefficient with 95% CIs were estimated via bootstrapping (1000 iterations).

Model accuracy was evaluated using confusion matrices, accuracy, and *F1* scores. A χ^2 test compared accuracy to ground truth, and agreement was assessed with Yule's *Q*. Receiver-operating-characteristic (ROC) curves and area-under-the-curve (AUC) analyses were performed for the development dataset to discriminate contrast presence. Model speed was evaluated using relative differences.

Cohen's kappa coefficient was computed for Datasets II, III, V, and VI to quantify interreader reliability between a research assistant and ground truth provided by a radiologist with four years of experience.

Statistical analyses were generated in Python 3.9.18 and R 4.4.1. *P*-values of <0.05 were considered statistically significant. Missing data was excluded from statistical

analyses. All source code and the model can be found at <https://github.com/FintelmannLabDevelopmentTeam/Falcon>.

Results

Patient and CT Characteristics

Datasets I and IV consisted of 1151 patients (mean age 59.1 years, 195 female) with 1156 CT scans of the HN, including 790 scans with intravenous contrast (68.3%). Datasets II and V consisted of 994 patients (mean age 67.2 years, 549 female) with 1001 CT scans of the chest, including 587 scans with intravenous contrast (58.6%). Datasets III and VI consisted of 981 patients (mean age 63.7, 417 female) with 981 scans of the AP, including 671 scans with intravenous contrast (68.4%) (Table 1 and Table 2). Each dataset contributed 523 scans on average (range, 350–631).

FALCON Accuracy and *F1* Score

Model performance distinguishes how well the algorithm detects the presence of intravenous contrast on CT scans. For the internal validation dataset, FALCON yielded *F1* scores for contrast detection of the HN, chest, and AP datasets of 98.7%, 99.4%, and 95.5%, respectively. Sensitivity and specificity for contrast detection of the HN, chest, and AP datasets are 98.7% and 97.1%, 100% and 97.7%, and 98.7% and 88.5%, respectively. FALCON yielded *F1* scores for body part classification of the HN, chest, and AP datasets of 100% for all datasets. Sensitivity and specificity of body part classification of the HN, chest, and AP datasets are 100% for all datasets (Table S2). Outputs are shown in a confusion matrix (Fig. 4).

For the external test dataset, FALCON yielded *F1* scores for contrast detection after bootstrapping for the

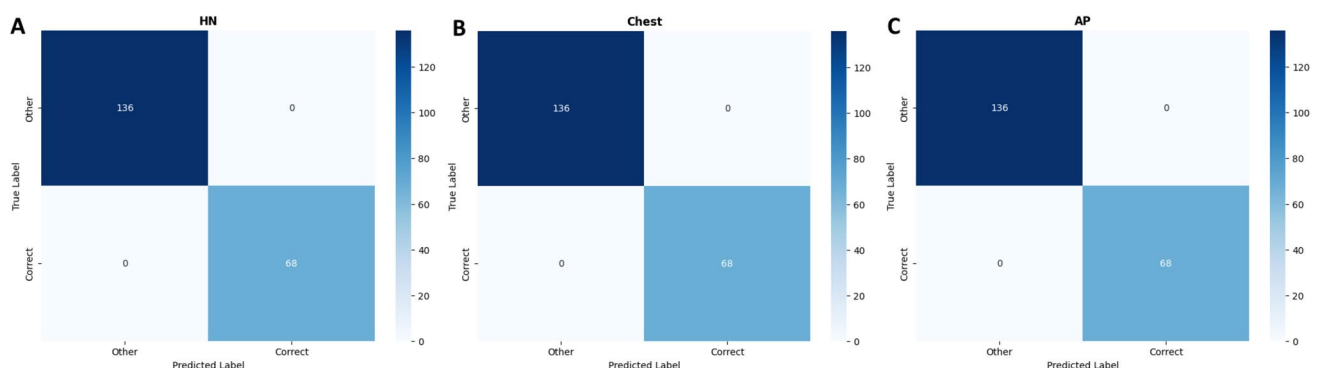


Fig. 4 Confusion matrices of development datasets for body part classification. Confusion matrices of patient-level model performance for body part classification in dataset I (A), dataset II (B), and dataset III (C). HN = head and neck. AP = abdomen and pelvis

HN, chest, and AP datasets of 99.4% (95%CI: 98.8, 99.9), 98.3% (95%CI: 96.9, 99.5), and 98.1% (95%CI: 96.9, 99.1), respectively (Table S3). Accuracy of the HN, chest, and AP datasets yielded 99.2% (95%CI: 98.5, 99.8), 98.1% (95%CI: 97.0, 99.5), and 96.9% (95%CI: 94.9, 98.6), respectively. The DICOM header *F1* scores yielded 95.7% and 93.8% for HN and chest, respectively (Table 3). Metrics were calculated with respect to only the final contrast prediction; body part predictions were ignored. A χ^2 test comparing predictions to ground truth yielded a statistically significant difference ($P < 0.001$) for each external test dataset. The robustness of the model is further demonstrated with a

strong correlation ($Q > 0.99$) for all datasets with Yule's *Q* statistic (Table S4). Outputs for internal validation and external testing are shown in a confusion matrix (Fig. 5, Fig. S5).

ROC curves at the patient level for the internal testing dataset of HN, chest, and AP show AUCs of 99%, 100%, and 99%, respectively (Fig. 6).

Interreader variability between the research assistant and a radiologist with four years of experience yielded Cohen's kappa coefficients of 0.97 ($P < 0.001$) for Dataset II, 0.85 ($P < 0.001$) for Dataset III, 0.99 ($P < 0.001$) for Dataset V, and 0.97 ($P < 0.001$) for Dataset VI.

Table 3 FALCON external test set performance for intravenous contrast detection and body part classification

	FALCON sensitivity	FALCON specificity	FALCON <i>F1</i> score	FALCON accuracy	FALCON PPV	FALCON NPV	Ye et al. external test <i>F1</i> score	DICOM header <i>F1</i> score	DICOM header accuracy
Head-neck Dataset IV body part	99.3%	100%	99.7%	99.7%	100%	99.5%	—	—	—
Head-neck Dataset IV contrast	99.5%	98.5%	99.4%	99.2%	99.3%	99.0%	99.9%	95.7%	94.5%
Chest Dataset V body part	99.3%	97.5%	96.7%	98.0%	94.3%	99.7%	—	—	—
Chest Dataset V contrast	99.0%	97.5%	98.3%	98.3%	97.5%	99.0%	92.3%	93.8%	94.0%
Abdomen-pelvis Dataset VI body part	93.7%	99.5%	96.0%	98.0%	98.5%	97.8%	—	—	—
Abdomen-pelvis Dataset VI contrast	97.0%	96.3%	98.1%	96.9%	99.3%	85.2%	—	—	—

Calculations were performed with respect to either body part prediction or intravenous contrast detection. We compared FALCON results to Ye et al. and DICOM headers. *DICOM* Digital Imaging and Communications in Medicine, *NPV* negative predictive value, *PPV* positive predictive value

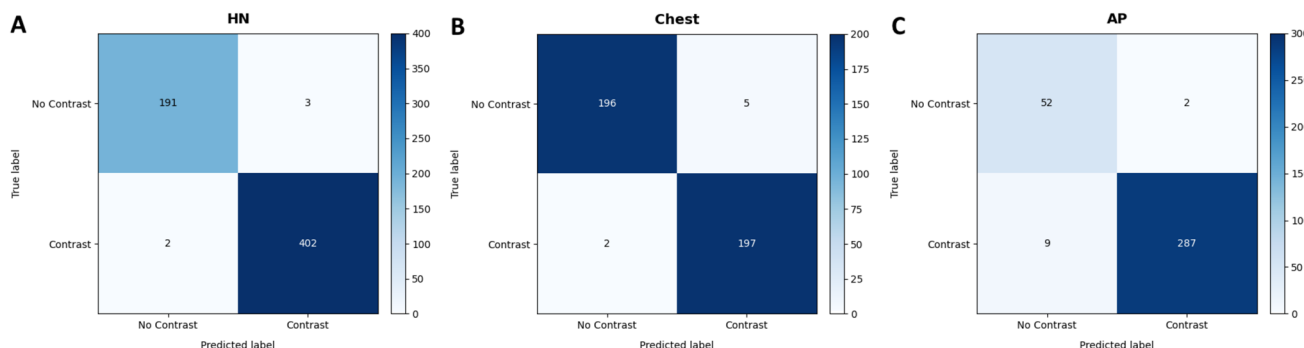


Fig. 5 Confusion matrices of external test datasets for intravenous contrast detection. Confusion matrices of patient-level model performances for contrast detection in dataset IV (A), dataset V (B), and

dataset VI (C). For external test results, the entire pipeline was used end-to-end. HN= head and neck. AP= abdomen and pelvis

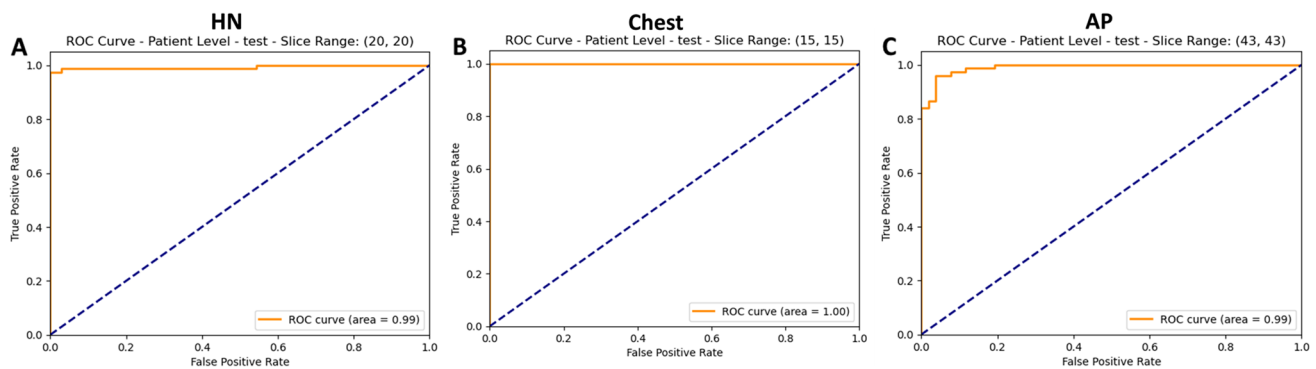


Fig. 6 Receiver operating characteristic curves for the FALCON development dataset. Receiver operating characteristic (ROC) curves for the ResNet9 model on datasets used to develop FALCON. All

ROC curves show high areas under the curve: 0.99 for the head and neck (HN) CT dataset (A), 1.00 for the chest CT dataset (B), and 0.99 for the abdomen and pelvis (AP) dataset (C)

Failure Case Analysis

The failure case analysis demonstrated that intravenous contrast presence was incorrectly labeled in 1.4% (19/1348 scans) between FALCON outputs and the ground truth contrast labels in the external test set. Discrepant cases in the external test dataset included five (0.84%) HN, four (1.0%) chest, and ten (2.9%) AP scans. No failures occurred in body part classification.

Comparison of FALCON for Intravenous Contrast Detection to Historical Data

On a computer without a dedicated GPU, the average time required for the FALCON model was 1.3, 1.8, and 3.7 s for a CT scan of the HN, chest, and AP, respectively. The average number of slices per series varies among datasets with 143.1, 143.0, and 300.6, respectively, exhibiting a clear correlation between the amount of information processed (number of slices) and the necessary computation time. Results slightly decreased on a computer with a dedicated GPU, with a mean relative difference of 16.5%. Compared to manual contrast annotation of an expert clinician, the algorithm performs significantly faster than an expert clinician (Table 4, Table S5).

Graphical User Interface

The open-access FALCON model (Fig. 3) was designed to run in the background of any radiology workstation to detect intravenous contrast presence and classify the body part of CT scans. A file directory can be selected in the FALCON graphical user interface to add multiple CT scans in the queue. FALCON will provide results of each scan in real-time.

Table 4 Average time in seconds required to annotate each computed tomography scan for the presence of intravenous contrast

	Expert Clinician	FALCON Computer (GPU)	FALCON Computer (CPU only)
Head-neck	21	2.1	1.3
Chest	33	2.6	1.8
Abdomen-pelvis	1.6	3.1	3.7

Time (seconds) was measured for FALCON to process one CT scan for the presence of intravenous contrast. We compared results obtained on a computer with and without a dedicated GPU, as well as with an expert clinician, as reported by Ye et al. *CPU* central processing unit, *GPU* graphics processing unit

Discussion

We present the development and validation of the open-access fully automated end-to-end FALCON pipeline that can reliably annotate intravenous contrast and body part of CT scans. FALCON is faster than human experts, does not require region localization, and can run without supervision.

FALCON stands out due to the large, diverse, and validated ground truth training data and near-perfect *F1* scores on external validation sets. External validation is critical for assessing an algorithm's transferability to unseen datasets. While some studies report perfect *F1* scores for contrast classification in chest CT scans, the use of small external validation sets with fewer than 100 samples limits statistical and clinical significance [9]. Other studies were limited to one particular body part to identify contrast enhancement [7, 8, 28]. By leveraging FALCON's open-source design, the algorithm could be adapted to differentiate between various contrast phases. Other models that identify intravenous contrast presence were limited in their application due to low accuracy performance metrics [29] or small cohort size [5].

Patient and CT Characteristics

To reduce selection bias, external datasets from multiple institutions were added to the training pipeline, to include differing patient groups and ethnicities. A large variety of CT scans from diverse manufacturers and clinical settings were included to ensure generalizability. The external test set included 1348 independent CT scans and the development dataset includes 1790 independent CT scans. Model training data included diagnostic and attenuation correction CT scans; thus, the algorithm was validated on a broad spectrum of technical differences and acquisition parameters.

FALCON Accuracy and F1 Score

FALCON uses lightweight model architecture and a single slice for prediction rather than all slices in a CT series. These design choices allow FALCON to run on portable computers with limited computing resources, while still achieving near-perfect performance on the training datasets. We chose a barrier-free approach to maintain the already near-perfect predictive performance. Differences in processing time per scan between computers with and without a dedicated GPU may be attributed to hardware acceleration, GPU overhead, or access time. Access time may vary between machines depending on the storage device such as a hard disk drive or solid-state drive.

Failure Case Analysis

Failure to correctly identify intravenous contrast occurred in 1.4% of cases while no failures occurred in body part classification. In HN CT scans, failures were most likely due to streak artifacts from dental implants or heavy carotid artery calcification, which introduce scatter and alter surrounding tissue attenuation. In chest CT scans, failures were largely attributed to streak artifacts related to the presence of coronary artery stents, central venous catheters, pacemaker leads, and surgical clips related to coronary artery bypass grafting. However, the reasons behind discrepancies in AP CT scans remain unclear, though they are unlikely to be related to calcified aortic plaque.

Use Cases

FALCON was developed with the intention to streamline annotation of CT scans used for research. While intravenous contrast labeling in the clinical setting was not the motivation to develop FALCON, the tool might be used to inform hanging protocols on PACS workstations. Accurate reporting of intravenous contrast use has implications for medical

billing, and failure to identify intravenous contrast may result in medical malpractice claims [30]. Therefore, there may be additional clinical use cases for FALCON.

Limitations

The results of our study must be interpreted in the context of its limitations. First, FALCON is not designed to process whole-body scans. For detection of body parts not defined within our scope, the model would require retraining. Second, while contrast phase is not captured, this could be achieved with retraining. Third, the cascading architecture of body part classification followed by intravenous contrast detection may impact performance. However, body part classification demonstrated consistent high performance across all datasets, suggesting that this design choice is unlikely to represent a critical vulnerability. Lastly, the study incorporates ground truth and time assessment from both the literature and our team. Longer assessment times for expert clinicians in the literature may be attributable to the DICOM viewer used and whether the scan needed to be downloaded from PACS to a local radiology workstation. In our study, we utilized a local DICOM viewer in which downloaded DICOM files were preloaded prior to assessing contrast presence.

Conclusion

The open-access FALCON pipeline reliably classifies the body part and presence of intravenous contrast on CT scans without user input. This has the potential to enhance reliability, speed, and efficiency in a clinical and research workflow.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10278-026-01865-8>.

Acknowledgements Julian Westphal was supported by Rolf W. Günther Stiftung für Radiologische Wissenschaften. We thank Dr. Zezhong Ye for kindly providing us with the ground truth labels they had created for their publication Ye Z, Qian JM, Hosny A, Zeleznik R, Plana D, Likitlersuang J, et al. Deep Learning-based Detection of Intravenous Contrast Enhancement on CT Scans. *Radiol Artif Intell*. 2022 May 4 PMID: 35652117.

Author Contribution All authors contributed to the study conception and design. Data collection and analysis were performed by Philipp Kaess, Julian Westphal, Lea Mantz, Rashid Barnawi, and Florian Fintelmann. Coding was performed by Philipp Kaess. The first draft of the manuscript was written by Julian Westphal, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the National Institute of Health (NIH) grant R01-CA298002 (F.J.F.).

Data Availability The public datasets used in this study are openly available through The Cancer Imaging Archive. The institutional dataset is not publicly available.

Declarations

Ethics Approval Approval was obtained from the ethics committee of Mass General Brigham. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Consent to Participate This study has obtained IRB approval from the Mass General Brigham Institutional Review Board, and the need for informed consent was waived.

Consent for Publication For this type of study, consent for publication is not required.

Competing Interests Dr. Fintelmann receives research support from the William M. Wood Foundation and serves as a paid consultant for Boston Scientific Corporation, Varian Medical Systems, Tubulis GmbH, and Becton Dickinson for unrelated work. The other authors declare no relevant conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Guelde MO, Kohnen M, Keysers D, Schubert H, Wein BB, Bredno J, et al. Quality of DICOM header information for image categorization. In: Siegel EL, Huang HK, editors. *Medical Imaging 2002: PACS and Integrated Medical Information Systems: Design and Evaluation* [Internet]. San Diego, CA; 2002 [cited 2025 Apr 19]. p. 280–7. <https://doi.org/10.1117/12.467017>
- Harvey H, Glocker B. A Standardised Approach for Preparing Imaging Data for Machine Learning Tasks in Radiology. In: Ranschaert ER, Morozov S, Algra PR, editors. *Artificial Intelligence in Medical Imaging* [Internet]. Cham: Springer International Publishing; 2019 [cited 2025 Apr 19]. p. 61–72. https://doi.org/10.1007/978-3-319-94878-2_6
- Neu SC, Crawford KL, Toga AW. Practical management of heterogeneous neuroimaging metadata by global neuroimaging data repositories. *Front Neuroinform*. 2012;6:8. <https://doi.org/10.3389/fninf.2012.00008>
- Mackenzie A, Lewis E, Loveland J. Successes and challenges in extracting information from DICOM image databases for audit and research. *Br J Radiol*. 2023;96:20230104. <https://doi.org/10.1259/bjr.20230104>
- Criminisi A, Juluru K, Pathak S. A Discriminative-Generative Model for Detecting Intravenous Contrast in CT Images. In: Fichtinger G, Martel A, Peters T, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*. Berlin, Heidelberg: Springer; 2011. p. 49–57. https://doi.org/10.1007/978-3-642-23626-6_7
- Sofka M, Wu D, Sühling M, Liu D, Tietjen C, Soza G, et al. Automatic contrast phase estimation in CT volumes. *Med Image Comput Comput Assist Interv*. 2011;14:166–74. https://doi.org/10.1007/978-3-642-23626-6_21
- Philbrick KA, Yoshida K, Inoue D, Akkus Z, Kline TL, Weston AD, et al. What Does Deep Learning See? Insights From a Classifier Trained to Predict Contrast Enhancement Phase From CT Images. *AJR Am J Roentgenol*. 2018;211:1184–93. <https://doi.org/10.2214/AJR.18.20331>
- Tang Y, Lee HH, Xu Y, Tang O, Chen Y, Gao D, et al. Contrast Phase Classification with a Generative Adversarial Network. *Proc SPIE Int Soc Opt Eng*. 2020;11313:1131310. <https://doi.org/10.1117/12.2549438>
- Na S, Sung YS, Ko Y, Shin Y, Lee J, Ha J, et al. Development and validation of an ensemble artificial intelligence model for comprehensive imaging quality check to classify body parts and contrast enhancement. *BMC Med Imaging*. 2022;22:87. <https://doi.org/10.1186/s12880-022-00815-4>
- Raffy P, Pambrun J-F, Kumar A, Dubois D, Patti JW, Cairns RA, et al. Deep Learning Body Region Classification of MRI and CT Examinations. *J Digit Imaging*. 2023;36:1291–301. <https://doi.org/10.1007/s10278-022-00767-9>
- Roth HR, Lee CT, Shin H-C, Seff A, Kim L, Yao J, et al. Anatomy-specific classification of medical images using deep convolutional nets. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI) [Internet]. 2015 [cited 2025 Mar 18]. p. 101–4. <https://doi.org/10.1109/ISBI.2015.7163826>
- Yan K, Lu L, Summers RM. Unsupervised body part regression via spatially self-ordering convolutional neural networks. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) [Internet]. 2018 [cited 2025 Mar 18]. p. 1022–5. <https://doi.org/10.1109/ISBI.2018.8363745>
- Ouyang Z, Zhang P, Pan W, Li Q. Deep learning-based body part recognition algorithm for three-dimensional medical images. *Med Phys*. 2022;49:3067–79. <https://doi.org/10.1002/mp.15536>
- Perez AA, Pickhardt PJ, Elton DC, Sandfort V, Summers RM. Fully automated CT imaging biomarkers of bone, muscle, and fat: correcting for the effect of intravenous contrast. *Abdom Radiol (NY)*. 2021;46:1229–35. <https://doi.org/10.1007/s00261-020-02755-5>
- Boutin RD, Kaptuch JM, Batani CP, Chalfant JS, Yao L. Influence of IV Contrast Administration on CT Measures of Muscle and Bone Attenuation: Implications for Sarcopenia and Osteoporosis Evaluation. *AJR Am J Roentgenol*. 2016;207:1046–54. <https://doi.org/10.2214/AJR.16.16387>
- Fuchs G, Chretien YR, Mario J, Do S, Eikermann M, Liu B, et al. Quantifying the effect of slice thickness, intravenous contrast and tube current on muscle segmentation: Implications for body composition analysis. *Eur Radiol*. 2018;28:2455–63. <https://doi.org/10.1007/s00330-017-5191-3>
- He L, Huang Y, Ma Z, Liang C, Liu Z. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci Rep*. 2016;6:34921. <https://doi.org/10.1038/srep34921>
- Troschel AS, Troschel FM, Fuchs G, Marquardt JP, Ackman JB, Yang K, et al. Significance of Acquisition Parameters for Adipose Tissue Segmentation on CT Images. *AJR Am J Roentgenol*. 2021;217:177–85. <https://doi.org/10.2214/AJR.20.23280>

19. Marquardt JP, Tonnesen PE, Mercaldo ND, Graur A, Allaire B, Bouxsein ML, et al. Subcutaneous and Visceral Adipose Tissue Reference Values From the Framingham Heart Study Thoracic and Abdominal CT. *Invest Radiol*. 2025;60:95–104. <https://doi.org/10.1097/RLI.0000000000001104>
20. Mantz L, Mercaldo ND, Simon J, Kaess P, Yang K, Dietrich A-SW, et al. Preoperative Chest CT Myosteosis Indicates Worse Postoperative Survival in Stage 0-IIIB Non-Small Cell Lung Cancer. *Radiology*. 2025;314:e240282. <https://doi.org/10.1148/radiol.240282>
21. Best TD, Mercaldo SF, Bryan DS, Marquardt JP, Wrobel MM, Bridge CP, et al. Multilevel Body Composition Analysis on Chest Computed Tomography Predicts Hospital Length of Stay and Complications After Lobectomy for Lung Cancer: A Multicenter Study. *Ann Surg*. 2022;275:e708–15. <https://doi.org/10.1097/SLA.0000000000004040>
22. Grossberg A, Elhalawani H, Mohamed A, Mulder S, Williams B, White AL, et al. HNSCC [Internet]. The Cancer Imaging Archive; 2020 [cited 2024 Nov 10]. <https://doi.org/10.7937/K9/TCIA.2020.A8SH-7363>
23. Kwan JYY, Su J, Huang SH, Ghoraie LS, Xu W, Chan B, et al. Data from Radiomic Biomarkers to Refine Risk Models for Distant Metastasis in Oropharyngeal Carcinoma [Internet]. The Cancer Imaging Archive; 2019 [cited 2024 Nov 10]. <https://doi.org/10.7937/TCIA.2019.8DHO2GLS>
24. Ye Z, Qian JM, Hosny A, Zeleznik R, Plana D, Likitlersuang J, et al. Deep Learning-based Detection of Intravenous Contrast Enhancement on CT Scans. *Radiol Artif Intell*. 2022;4:e210285. <https://doi.org/10.1148/ryai.210285>
25. Kaess P, Rückert D, Ziller A. Fairness Impact of Private Workflows in Large Real-World Medical AI Models. Technische Universität München; 2024.
26. Dozat T. Incorporating Nesterov Momentum into Adam. *ICLR*. 2016;
27. Klaus H, Ziller A, Rueckert D, Hammernik K, Kaissis G. Differentially private training of residual networks with scale normalisation [Internet]. arXiv; 2022 [cited 2025 Apr 19]. <https://doi.org/10.48550/ARXIV.2203.00324>
28. Muhamedrahimov R, Bar A, Laserson J, Akselrod-Ballin A, Elnekave E. Using Machine Learning to Identify Intravenous Contrast Phases on Computed Tomography. *Comput Methods Programs Biomed*. 2022;215:106603. <https://doi.org/10.1016/j.cmpb.2021.106603>
29. Sugimori H. Classification of Computed Tomography Images in Different Slice Positions Using Deep Learning. *J Healthc Eng*. 2018;2018:1753480. <https://doi.org/10.1155/2018/1753480>
30. Khan A, Bajaj S, Khunte M, Payabvash S, Wintermark M, Gandhi D, et al. Contrast Agent Administration as a Source of Liability: A Legal Database Analysis. *Radiology*. 2023;308:e230802. <https://doi.org/10.1148/radiol.230802>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.