# Automatic AI CT Quality Assurance Artifact Detection using Residual-Enhanced Teacher–Student Learning

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

Shanli Ding[1,2]; Tinsu Pan, PhD[2]; Frank Dong, PhD[2]; Osama Mawlawi, PhD[2]
[1] Graduate School of Biomedical Sciences, The University of MD Anderson Cancer Center, Houston, TX 77030, USA
[2] Department of Imaging Physics, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

## ABSTRACT

**Purpose**: Routine computed tomography (CT) quality assurance (QA) relies on visual inspection, where subtle non-uniformities can be difficult to identify consistently across scanners and readers. We developed an efficient, interpretable artificial intelligence-assisted tool that detects and localizes artifacts in CT QA images to reduce manual review burden.

**Methods**: We propose a CT QA artifact detection model (AutoCTAD), a teacher–student anomaly detector with residual-based multi-scale enhancement to amplify subtle non-uniformities while preserving smooth shading trends. We randomly selected 16,734 normal studies (49,996 images) spanning 2023-01 to 2025-12 and curated 300 artifact images including ring, shading, strike, and low-contrast smudge patterns. Normal images were defined by visual inspection and using a CT-number-based QA criterion. Artifact images were identified by physicist visual inspection and included cases with either passing or failing CT-number criteria, to capture clinically relevant artifacts that may not be reflected by CT number alone. Normal images were split into train/validation/test sets of 39,837/9,959/200 images. An image-level decision threshold was set using a percentile criterion derived from train-normal scores. Pixel-level heatmaps were generated for localization and review. The model was trained and evaluated on a Linux system with a NVIDIA A100 GPU.

**Results**: On the test set (300 artifact, 200 normal), AutoCTAD achieved 97.8% accuracy, 98.0% precision, and 98.2% F1-score (295/300 artifacts detected; 194/200 normals correctly rejected). Mean end-to-end inference latency was 143 ms/image (p95: 146 ms) for 512×512 inputs on the NVIDIA A100, including residual-enhancement preprocessing. False negatives were subtle central smudges with acceptable CT number, whereas false positives were normal floods with mild non-uniformity or smudge-like texture. Heatmaps localized suspicious regions consistent with visual findings and supported rapid adjudication.

**Conclusion**: AutoCTAD provides high-performance, interpretable artifact detection with localized heatmaps across diverse artifact types, while identifying failure modes for contrast-limited smudges that motivate future work on scanner-aware thresholds and sensitivity tuning.

## CONTACT

Shanli Ding
Department of Imaging Physics
University of Texas MD Anderson Cancer Center
1155 Pressler St, Houston TX 77030
sding2@mdanderson.org
713-563-8928

## INTRODUCTION

Routine CT quality assurance relies on visual review and quantitative image quality checks, including CT number accuracy, uniformity, low-contrast performance, and artifact evaluation. However, subtle non-uniformities, faint smudge-like artifacts, and early scanner-related image degradation may be difficult to identify consistently by visual inspection alone. These findings can also be missed when the CT-number criterion remains within the acceptable range.

To develop AutoCTAD, an efficient and interpretable AI-assisted CT QA artifact detection model that detects image-level artifacts and localizes suspicious regions using pixel-level heatmaps.

## METHODS AND MATERIALS

- Dataset: The dataset was extracted from our institution across 73 scanners, including CT, PET/CT, and SPECT/CT systems from both GE and Siemens, although this study focused specifically on CT scanning quality control images. A total of 16,734 normal CT QA studies containing 49,996 images were randomly selected from January 2023 to December 2025. Normal images were defined by physicist visual inspection and CT-number-based QA criteria. A separate artifact set included 300 curated artifact images identified by physicist visual inspection. Artifact patterns included ring artifacts, shading artifacts, streak artifacts, and low-contrast smudge-like artifacts. Artifact images included both CT-number-passing and CT-number-failing cases to capture visually relevant artifacts that may not be reflected by CT-number measurements alone.
  - Training: **39,837** normal images
  - Validation: **9,959** normal images
  - Testing: **200** normal + **300** artifact

- Model Design: AutoCTAD was designed as a teacher–student anomaly detection framework. The model was trained using normal CT QA images so that the student network learned to reproduce teacher-derived representations of normal image appearance. During inference, images that deviated from the learned normal appearance produced higher anomaly responses. Residual-based multi-scale enhancement was incorporated before model inference to amplify subtle non-uniformities while preserving broad smooth shading trends. This step was designed to improve sensitivity to faint artifacts such as central smudges, mild rings, and spatially diffuse non-uniformity.

- Decision Threshold and Localization: An image-level anomaly score was calculated for each input image. The decision threshold was determined using a percentile criterion derived from train-normal anomaly scores. Pixel-level heatmaps were generated to localize suspicious regions and support physicist review.

- Implementation: All experiments were performed on a Linux workstation using an NVIDIA A100 GPU. End-to-end inference latency was measured on 512 × 512 input images and included residual-enhancement preprocessing.

## RESULTS

AutoCTAD achieved strong image-level artifact detection performance on the independent test set of 300 artifact images and 200 normal images.

- ✓ Accuracy: 97.8%
- ✓ Precision: 98.0%
- ✓ F1-score: 98.2%
- ✓ Artifact detection: 295/300
- ✓ Normal rejection: 194/200

Based on the confusion matrix, AutoCTAD achieved **98.3% sensitivity** and **97.0% specificity**. The model missed 5 artifact images and falsely flagged 6 normal images.

Inference was efficient, with a mean end-to-end latency of **143 ms/image** and a p95 latency of 146 ms/image for **512 × 512** inputs on the NVIDIA A100 GPU.

Qualitative review showed that heatmaps localized suspicious regions consistent with visual findings, including ring-like patterns, shading regions, streak artifacts, and smudge-like abnormalities. These heatmaps provided interpretable evidence for rapid adjudication rather than only producing a binary pass/fail output.

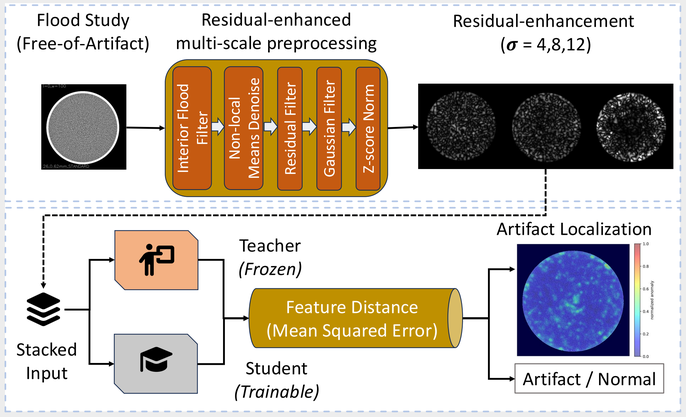


**Figure 1.** AutoCTAD architecture (teacher-student feature matching with residual-enhanced multi-scale inputs).

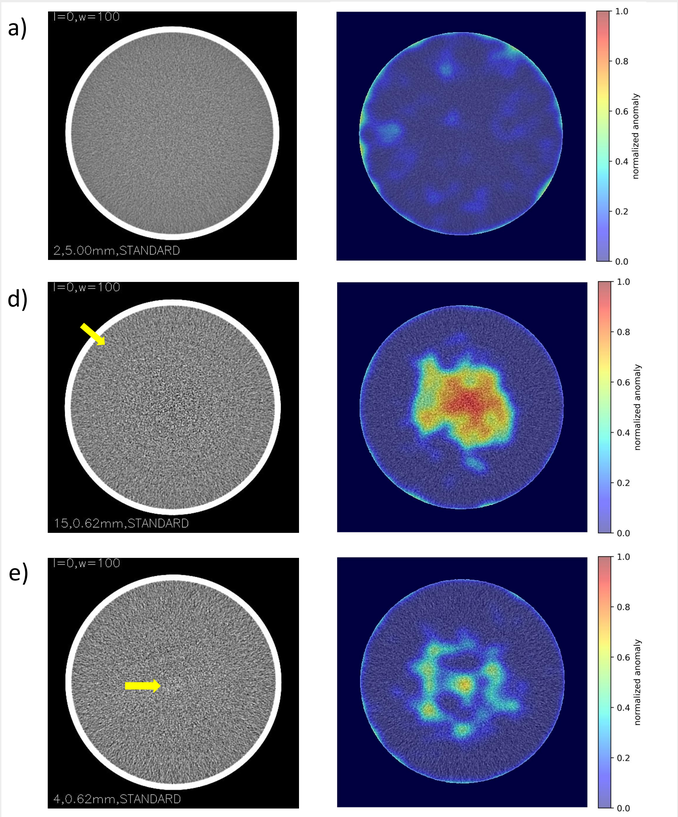


**Figure 2.** Representative examples of normal and artifact CT flood images with model detections. (a) Normal flood image meeting CT-number QA criteria. (b) Ring artifact. (c) low-contrast smudge patterns that passes CT-number criteria.

## DISCUSSION

AutoCTAD demonstrated high artifact detection performance across multiple artifact types while maintaining low false-positive rates on normal QA images. The results suggest that teacher–student anomaly detection is well suited for CT QA because the model can be trained primarily on normal images, reducing the need for extensive artifact annotation during training.

The residual-enhancement strategy was important for this task because many CT QA artifacts are subtle and low contrast. By emphasizing local deviations from expected smooth image appearance, the model improved visibility of faint non-uniformities that may be difficult to identify during routine manual review.

The inclusion of CT-number-passing artifact cases is clinically important. These cases show that quantitative CT-number criteria alone may not capture all visually meaningful artifacts. AutoCTAD may therefore serve as a complementary review tool rather than a replacement for standard QA metrics.

Failure analysis showed two main limitations. False-negative cases were primarily subtle central smudges with acceptable CT-number measurements, suggesting that contrast-limited artifacts remain challenging. False-positive cases were normal images with mild non-uniformity or smudge-like texture, indicating that scanner-specific appearance and acquisition variability may affect threshold selection.

Future work will focus on scanner-aware thresholding, sensitivity tuning for low-contrast smudge artifacts, and prospective validation across additional scanners and QA acquisition settings.

## CONCLUSIONS

AutoCTAD provides high-performance, efficient, and interpretable AI-assisted CT QA artifact detection. The model detects diverse artifact patterns, localizes suspicious regions with heatmaps, and identifies subtle failure modes that may not be captured by CT-number criteria alone. By combining quantitative anomaly scoring with visual localization, AutoCTAD may help standardize routine QA review, reduce manual inspection burden, and support earlier recognition of scanner-related image quality issues. Because AutoCTAD was trained and evaluated using multi-vendor, multi-scanner institutional data, its strong performance supports its potential as a vendor-agnostic CT QA artifact detection tool. Future prospective evaluation across additional scanners and acquisition protocols will further assess its robustness, generalizability, and clinical workflow value. These results support further development of AutoCTAD as a complementary tool for routine CT QA review.

## REFERENCES

1. American College of Radiology. CT Quality Control Manual. ACR; 2017, revised 2025.
2. Bergmann P, Fauser M, Sattlegger D, Steger C. Uninformed Students: Student–Teacher Anomaly Detection With Discriminative Latent Embeddings. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020:4183–4192.

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com