SIIM CAIMI26 Conference on Artificial Intelligence in Medical Imaging
OCT 26-27 | University of Pennsylvania, Philadelphia, PA

Zero-Shot Series-Level MRI Artifact Localization Using Pretrained Foundation Model Embeddings and Weak Study-Level Supervision

Dvij Sharma, MS, University of Wisconsin – Madison; Xue Li, PhD; Carl Kashuk, MS; Tracy He; Hugh Pemberton, PhD; Erhan Bas, PhD; Krisztian Koos, PhD; Dattesh Shanbhag, PhD; Noel Dsouza, MS; Pranay Doshi, MS; Iman Estakhraji, PhD; Marc Label, PhD; Orhan Unal, PhD; Richard Bruce, MD; Alan McMillan, PhD; John Garrett, PhD

Introduction/Background

MRI artifacts can reduce diagnostic confidence and hospital throughput. While study-level artifact labels can be obtained from radiologist reports, series-level annotations are costly and often unavailable in routine practice. We investigated whether representations learned by a pretrained MRI foundation model could support both study-level artifact detection and zero-shot transfer to series-level localization without requiring series-level annotations.

Methods/Intervention

A pre-trained 3D MRI FM (GE HealthCare) generated embeddings from 21,780 MRI studies (11,424 with artifacts and 10,356 without) spanning multiple anatomies and pulse sequences. A lightweight multilayer perceptron (MLP) adapter was trained on study-level labels, with a 50-setting Optuna search. Data were partitioned at the study level into training, validation, and test (60:20:20). The resulting classifier was evaluated for study-level artifact detection and zero-shot series-level localization. For localization, 17 randomly selected studies comprising ~220 series (at least one of each artifact, see Figure 2A) were reviewed by an imaging physicist to identify artifact-bearing series; only primary sequences were retained, with localizers, reformats, and derived parametric maps (e.g. ADC) omitted. The trained classifier was then applied directly to individual series, which were ranked by logit. For each study, the number of ranked series needed to find the physicist-flagged series was recorded.

Results/Outcome

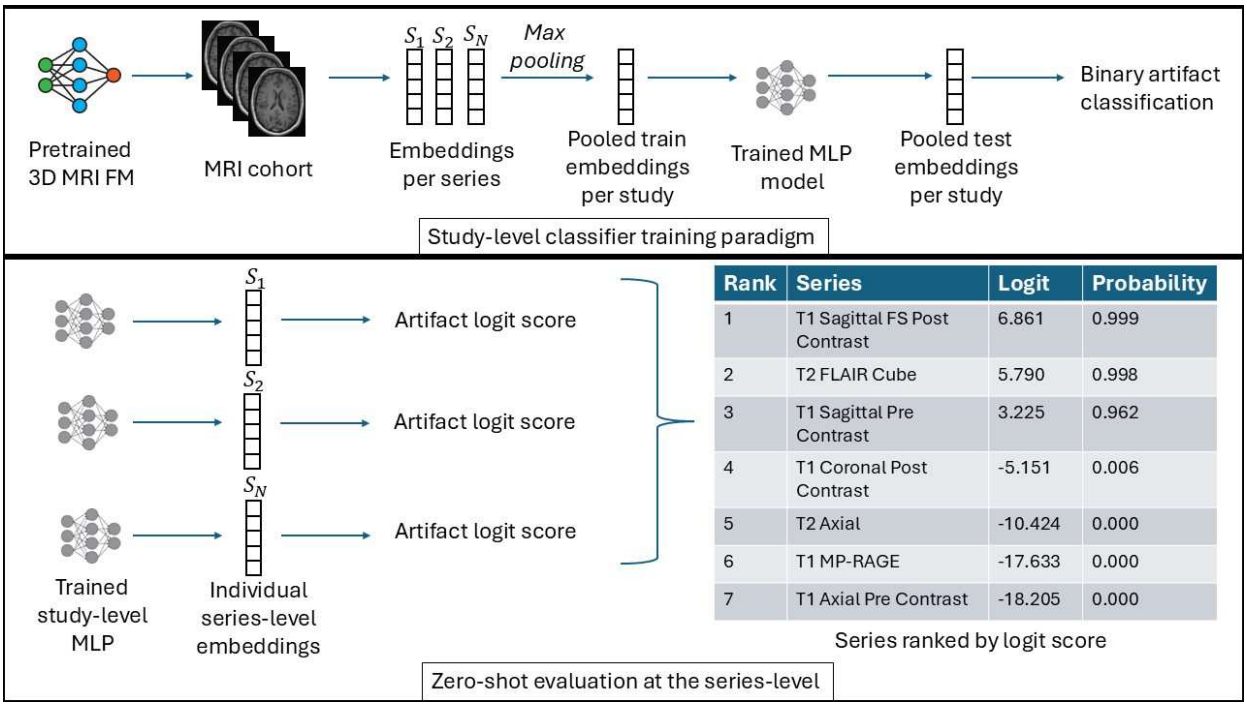
The study-level classifier achieved 97.2% accuracy, 99.4% AUROC, and 99.4% PR-AUC on the held-out test set (Figure 2A), and t-SNE showed good embedding separability between artifact and non-artifact studies (Figure 2C). When applied at the series level without additional training, the classifier identified physicist-labeled artifact-bearing series within the top 3 rankings for 94.1% of studies (Figure 2B), typically within the first two ranked cases (median=1.00, mean=1.88). An example annotation is shown in Figure 3.

Conclusion

An adapter-based FM achieved strong study-level artifact detection and demonstrated promising zero-shot transfer to series-level artifact localization. Future work will evaluate performance in a more extensively annotated cohort.

Statement of Impact

Our findings suggest that pretrained MRI FMs may transfer artifact knowledge learned from study-level labels to series-level localization using only lightweight adapters, reducing dependence on costly series-level annotations. This could help technologists rapidly identify problematic studies and series, mitigating costs due to reduced throughput, facilitating quality assurance, and improving workflow efficiency.



Rank	Series	Logit	Probability
1	T1 Sagittal FS Post Contrast	6.861	0.999
2	T2 FLAIR Cube	5.790	0.998
3	T1 Sagittal Pre Contrast	3.225	0.962
4	T1 Coronal Post Contrast	-5.151	0.006
5	T2 Axial	-10.424	0.000
6	T1 MP-RAGE	-17.633	0.000
7	T1 Axial Pre Contrast	-18.205	0.000

Figure 1: (top) Schematic demonstrating method used to train study-level MLP adapter head using study-level radiologist annotations. (bottom) workflow for using trained MLP adapter for zero-shot evaluation at the series level, with sample output from our test set.

A

Study-level summary

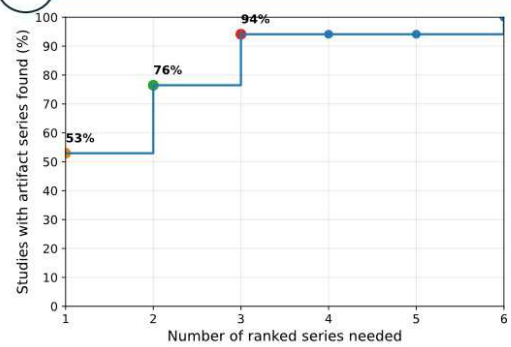
Unweighted mean (macro)						
Adapter	Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC
MLP	97.245%	97.241%	97.236%	97.238%	99.412%	99.404%

Support-weighted mean (weighted)						
Adapter	Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC
MLP	97.245%	97.245%	97.245%	97.245%	99.412%	99.404%

Artifact	Count
Aliasing	215
Gibbs Ringing	105
Inhomogeneity	209
Motion	13,440
Susceptibility	1,333
Flow	269
Other	1,145

B

Series-level summary



Metric	Value
Median series needed	1.00
Mean series needed	1.88
Top-1 localization	52.9%
Top-2 localization	76.5%
Top-3 localization	94.1%

C

Embeddings summary

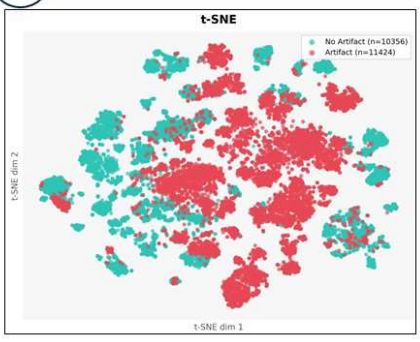
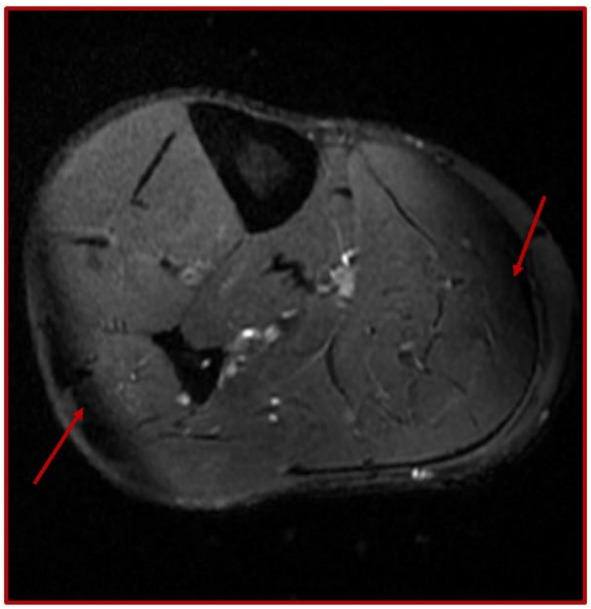


Figure 2: (A) Tables of (left) standard performance metrics (macro and weighted averages) on study-level artifact detection, and (right) breakdown of artifacts in dataset and associated counts. (B) (left) Cumulative plot showing number of studies with physicist-identified artifact series found at each ranking, and (right) table of mean / median series needed, and top-k localization. (C) t-SNE plot showing good separation between artifact and non-artifact embeddings.



Study-level report label: Inhomogeneity

Physicist annotation: Inhomogeneity shading
Notes: "Shading in other series but worst in axial T2"

Model study-level prediction: artifact present

Rank	Series	Logit	Probability
1	T2 Axial FS	6.553	0.999
2	T1 Coronal FS Post Contrast	5.665	0.997
3	Coronal STIR	5.133	0.994
4	Sagittal STIR	4.311	0.987
5	T1 Axial Pre Contrast	0.171	0.543
6	T1 Coronal Pre Contrast	-0.021	0.495
7	T1 Sagittal Pre Contrast	-3.614	0.026
8	T1 Sagittal FS Post Contrast	-4.563	0.010

Figure 3: Example model ranking of a test case with an 'inhomogeneity' study-level label from radiologist report. PACS-derived image of top model-ranked series is shown, displaying clear inhomogeneity shading, as identified by the medical physicist. In this case, only 1 ranked series was needed to match physicist annotation.

Keywords

Foundation Models; MRI Artifact Detection; Weakly Supervised Learning; Zero-Shot Learning; Imaging Workflow Optimization