

Helper AI V1 - Scope and Implementation Plan

Date: 2/17/2026

Summary

Helper AI V1 focuses on building a lightweight solution for classifying CT series using image content rather than relying solely on inconsistent DICOM metadata. The system generates a representative mosaic for each series and applies a multi-head ResNet-18 classifier to predict modality, view, and body region with associated confidence scores. The outputs are structured in a machine-readable format to support downstream metadata standardization, including narrowing to relevant Radlex Series Names.

The initial phase emphasizes building a stable end-to-end prototype and validating performance against a manually annotated test set. By early April, the goal is to have a working Helper AI V1 pipeline that operates on anonymized CT head and CT chest data and produces structured outputs aligned with Radlex naming conventions.

1. Background and Motivation

In practice, DICOM metadata is often incomplete, inconsistent, or unreliable for accurately describing imaging studies. Fields such as BodyPartExamined, SeriesDescription, and ViewPosition can vary significantly across institutions, be inconsistently populated, or contain ambiguous terminology. Because of this, relying solely on DICOM tags for study classification and standardized metadata mapping does not scale in a reliable or reproducible way. Manual annotation can resolve these inconsistencies, but it is time-intensive and not scalable.

Helper AI V1 is designed to address this gap by shifting the primary source of truth from DICOM metadata to image content. Instead of assuming the tags are correct, the system analyzes representative image content from each series and generates structured, confidence-scored labels. These outputs can then be used to support standardized workflows and improve metadata consistency.

The long-term objective is to enable consistent, automated study characterization that integrates cleanly with the RSNA Anonymizer and supports downstream metadata standardization efforts.

2. Problem Statement

We need a scalable and reproducible way to classify imaging studies by:

- Modality
- View / Plane
- Body Region

without relying exclusively on unreliable DICOM metadata.

The tool must:

- Run locally
- Be free to use
- Integrate cleanly into the anonymization workflow
- Produce structured outputs suitable for downstream metadata standardization

3. Scope of V1 (CT Head and CT Chest Only)

V1 will focus exclusively on CT head and CT chest series. For each CT series, the system will classify:

Contrast Usage (single label)

- With contrast
- Without contrast

Body Region (single label)

- Head
- Chest

From these predictions, the tool will output standardized Radlex series names such as:

- CT Head Without Contrast
- CT Head With Contrast
- CT Chest Without Contrast

The system will generate confidence scores for each prediction and flag low-confidence cases for review.

4. Technical Approach

4.1 Series-Level Mosaic Representation

For each anonymized CT series:

1. Load all slices within the series
2. Sample slices evenly across the volume
3. Resize and tile them into a single mosaic image

This mosaic acts as a compact visual summary of the entire series. The goal is to provide the model with sufficient anatomical coverage and contrast characteristics without requiring 3D modeling.

4.2 Classification Model

A lightweight CNN (ResNet-18) will be used as the backbone architecture.

The model will include multiple prediction heads:

- Contrast Usage
- Body Region

ResNet-18 is selected because it is computationally efficient and small enough to run locally. Predictions will be generated at the series level using the mosaic images.

4.3 Training Strategy

Training will initially use labels derived from DICOM metadata (weak supervision). Only cases with sufficient metadata will be included in the training set.

A smaller manually reviewed gold test set (~ 200–300 series) will be created to evaluate performance.

The model will be iteratively improved by:

- Reviewing misclassifications

- Improving the logic used to derive reliable training labels from DICOM metadata
- Expanding the ground truth set as needed

4.4 Output Format

For each series, the system will produce a structured JSON output including:

- Predicted contrast usage
- Predicted body region
- Confidence scores
- Radlex Series Name
- 'Needs review' flag if thresholds are not met

This structured output makes it easier to plug into downstream workflows, including Radlex Series Name mapping.

5. Implementation Timeline

Phase 1 Prototype Development: February 11 - February 25

The first phase is focused on getting the full pipeline working end to end.

This includes:

- Finalizing the labels for body region and contrast 
- Building the mosaic generation pipeline
- Generating mosaics for CT head and chest dataset
- Extracting high-confidence labels from metadata to use for training 
 - Developed automated extraction script that parses DICOM series metadata (SeriesDescription, BodyPartExamined, ProtocolName, ContrastBolusAgent)
 - Uses keyword-based inference with high-confidence filtering
 - Example: "CT HEAD WITHOUT CONTRAST" → body_region=head, contrast=without_contrast
 - Labels ready for training
- Implementing the ResNet-18 classification architecture 
 - Train a multi-head ResNet-18 classifier to predict:
 - Body region (head vs chest)
 - Contrast usage (with vs without)
 - Input: mosaic images

Backbone: Pretrained ResNet-18 (~11M parameters)

- Multi-head outputs:
 - Body region classification
 - Contrast classification
- Trained simultaneously using a combined loss function
- Training config
 - 80/20 train-validation split
 - Data augmentation: Flips, rotations
 - Optimizer: Adam + ReduceLROnPlateau scheduler
 - Early stopping
 - Batch size 32
- Training a baseline model
- Producing initial structured JSON outputs

Deliverable: A working CT prototype that produces structured classification outputs:

JSON example:

```
{  
  "series_name": "CT_HEAD_0042",  
  "body_region": "head",  
  "contrast": "without_contrast",  
  "radlex_series_name": "CT Head Without Contrast",  
  "body_region_confidence": 0.95,  
  "contrast_confidence": 0.89  
}
```

Phase 2 Validation and Refinement: February 26 - March 12

This phase will focus on:

- Creating a manually annotated gold test set of approximately 200 - 300 series
- Evaluating model performance against clean, human-reviewed labels

- Updating training data and metadata-derived labeling rules
- Adding confidence scoring and quality control logic

Deliverable: A validated V1 CT classifier with measurable performance.

Phase 3 Integration with Metadata Standardization: March 13 to April 1

Once the CT Head and Chest classifier is stable, the next step is to use its predictions to support standardized coding workflows in a structured way.

This phase will focus on:

- Loading Radlex Series Names locally as a structured reference table
- Using predicted contrast and body region to filter and narrow the set of relevant Radlex Series Names
- Extending the JSON output to include a shortlist of candidate Radlex Series Names based on this structured filtering

Deliverable: A classification pipeline that produces structured CT series labels along with Radlex series names.

6. Expected Outcome by Early April

By early April, the goal is to have:

- A stable CT series classification pipeline
- Confidence-based quality control logic
- Structured JSON outputs aligned with Radlex series names
- A demonstrable Helper AI V1 workflow