



Published in final edited form as:

*J Bone Joint Surg Am.* 2022 September 21; 104(18): 1649–1658. doi:10.2106/JBJS.21.01229.

## Applying Deep Learning to Establish a Total Hip Arthroplasty Radiography Registry:

### A Stepwise Approach

Pouria Rouzrokh, MD, MPH, MHPE<sup>1,2</sup>, Bardia Khosravi, MD, MPH, MHPE<sup>1,2</sup>, Quinn J. Johnson, BA<sup>2,3</sup>, Shahriar Faghani, MD<sup>1</sup>, Diana V. Vera Garcia, MD<sup>1,2</sup>, Bradley J. Erickson, MD, PhD<sup>1</sup>, Hilal Maradit Kremers, MD<sup>2,4,5</sup>, Michael J. Taunton, MD<sup>2,5</sup>, Cody C. Wyles, MD<sup>2,5,6</sup>

<sup>1</sup>Radiology Informatics Laboratory, Department of Radiology, Mayo Clinic, Rochester, Minnesota

<sup>2</sup>Orthopedic Surgery Artificial Intelligence Laboratory (OSAIL), Mayo Clinic, Rochester, Minnesota

<sup>3</sup>Mayo Clinic Alix School of Medicine, Mayo Clinic, Rochester, Minnesota

<sup>4</sup>Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota

<sup>5</sup>Department of Orthopedic Surgery, Mayo Clinic, Rochester, Minnesota

<sup>6</sup>Department of Clinical Anatomy, Mayo Clinic, Rochester, Minnesota

### Abstract

**Background:** Establishing imaging registries for large patient cohorts is challenging because manual labeling is tedious and relying solely on DICOM (digital imaging and communications in medicine) metadata can result in errors. We endeavored to establish an automated hip and pelvic radiography registry of total hip arthroplasty (THA) patients by utilizing deep-learning pipelines. The aims of the study were (1) to utilize these automated pipelines to identify all pelvic and hip radiographs with appropriate annotation of laterality and presence or absence of implants, and (2) to automatically measure acetabular component inclination and version for THA images.

**Methods:** We retrospectively retrieved 846,988 hip and pelvic radiography DICOM files from 20,378 patients who underwent primary or revision THA performed at our institution from 2000 to 2020. Metadata for the files were screened followed by extraction of imaging data. Two deep-learning algorithms (an EfficientNetB3 classifier and a YOLOv5 object detector) were developed to automatically determine the radiographic appearance of all files. Additional deep-learning algorithms were utilized to automatically measure the acetabular angles on anteroposterior pelvic

---

corresponding author: wyles.cody@mayo.edu.

Investigation performed at the Orthopedic Surgery Artificial Intelligence Laboratory (OSAIL), Department of Orthopedic Surgery, Mayo Clinic, Rochester, Minnesota

**Disclosure:** The **Disclosure of Potential Conflicts of Interest** forms are provided with the online version of the article (<http://links.lww.com/XXXXXXX>).

Appendix

Supporting material provided by the authors is posted with the online version of this article as a data supplement at [jbs.org](http://jbs.org) (<http://links.lww.com/XXXXXXX>).

and lateral hip radiographs. Algorithm performance was compared with that of human annotators on a random test sample of 5,000 radiographs.

**Results:** Deep-learning algorithms enabled appropriate exclusion of 209,332 DICOM files (24.7%) as misclassified non-hip/pelvic radiographs or having corrupted pixel data. The final registry was automatically curated and annotated in <8 hours and included 168,551 anteroposterior pelvic, 176,890 anteroposterior hip, 174,637 lateral hip, and 117,578 oblique hip radiographs. The algorithms achieved 99.9% accuracy, 99.6% precision, 99.5% recall, and a 99.6% F1 score in determining the radiograph appearance.

**Conclusions:** We developed a highly accurate series of deep-learning algorithms to rapidly curate and annotate THA patient radiographs. This efficient pipeline can be utilized by other institutions or registries to construct radiography databases for patient care, longitudinal surveillance, and large-scale research. The stepwise approach for establishing a radiography registry can further be utilized as a workflow guide for other anatomic areas.

**Level of Evidence:** Diagnostic Level IV. See Instructions for Authors for a complete description of levels of evidence.

---

Patient registries are organized systems that allow institutions to collect uniform data that can be utilized to evaluate specific outcomes for a population defined by a disease or exposure, and facilitate the achievement of predetermined scientific, clinical, or policy purposes<sup>1</sup>. These registries serve as data repositories for studies with a variety of purposes, including but not limited to understanding the natural history of a disease, evaluating diagnostic or therapeutic interventions, monitoring benefits and harms, measuring the quality of care, and/or identifying disparities among patient populations<sup>2</sup>. In contrast to a database that can be regarded as an unstructured pool of data, a registry is a structured and organized collection that can be further utilized for research purposes<sup>3,4</sup>.

Although most orthopaedic registries utilize clinical data, medical-imaging registries in other fields have also been described<sup>3</sup>. Although image analysis is a pillar of orthopaedic evaluation and longitudinal surveillance, current registries are limited by inadequate linkage of clinical and radiographic data. Radiographs serve as the cornerstone of patient evaluation and surveillance, and merging clinical and radiographic information is critical to maximize the potential of registries. For example, our own institution has a well-characterized total joint arthroplasty registry that has followed patients since 1969; however, most studies generated from this database have included limited radiographic outcomes or have relied on tedious manual review of a subset of patients of interest. Furthermore, the mere process of identifying relevant radiographs is cumbersome and not standardized, thus impeding the ability to answer clinical questions expeditiously. Ideally, the creation of radiography registries would be linked with clinical registries to leverage a powerful combination of information.

Most medical-imaging registries store and transfer information as DICOM (digital imaging and communications in medicine) files<sup>5</sup>. In addition to imaging information, most DICOM files include text metadata about the patient (e.g., name, sex, and age) and/or the imaging process (e.g., site, date, time, and magnification). Such metadata are stored as “tags” or “attributes” in those files to ensure that patient images will never be separate from their

metadata<sup>6</sup>. Furthermore, DICOM is both a data-storage format and a universal transfer protocol; thus, imaging and metadata from most modern imaging devices can be easily integrated with a picture archiving and communication systems (PACS), regardless of its manufacturer<sup>7</sup>.

Computer algorithms can assess DICOM tags and organize them according to their metadata. Although this approach seems like an intuitive solution to automatically build and organize medical-imaging registries, there are 2 critical shortcomings. First, several DICOM tags are entered manually and, because of different data-entry protocols or human error, can be inaccurate or incomplete. Second, imaging registries typically need to be organized according to some imaging features that are not documented or reliably identifiable in DICOM tags. For example, DICOM files are not annotated with important features, such as the presence or absence of devices or pathology. Therefore, building medical-imaging registries necessitates more comprehensive strategies to screen DICOM files and enhance imaging data with corresponding clinical and text data.

Artificial intelligence and deep-learning algorithms are emerging technologies that are widely utilized in both medical and nonmedical computer-based image processing<sup>8</sup>; their application in medical imaging has included enhancement of diagnosis, prognosis, lesion detection, and generating synthetic medical imaging<sup>9–11</sup>. Deep-learning algorithms are also powerful tools to screen DICOM pools and classify images. These tools usually work in the background to facilitate clinical workflows, such as routing specific radiographic studies toward certain specialists, but their research application for establishing large registries from retrospective studies is less discussed. The present study endeavored to utilize deep-learning algorithms to automatically construct a registry of hip and pelvic radiographs for an existing clinical registry of total hip arthroplasty (THA) patients. We hypothesized that a pipeline based on DICOM file processing and deep-learning algorithms could accurately identify and organize hip and pelvic radiographs according to selected features such as the radiographic view, laterality, presence or absence of a prosthesis, and annotation of salient prosthesis features. The aims of the study were (1) to utilize these automated pipelines to identify all pelvic and hip radiographs with appropriate annotation of laterality and presence or absence of implants and (2) to automatically measure acetabular component inclination and version for THA images. Herein, we report our stepwise approach, hoping that it will facilitate similar efforts at other interested institutions.

## Materials and Methods

Following institutional review board approval, we retrospectively utilized our total joint registry to identify 20,378 THA patients who underwent 1 or more primary or revision THA procedures at our institution between 2000 and 2020. We subsequently retrieved 846,988 radiography DICOM files of these patients from our institutional PACS, referring to this initial pool of file as the “institutional radiography database.” This collection of files included all radiographic studies (hip or otherwise) for these patients over that timeframe. Advanced imaging such as computed tomography and magnetic resonance imaging was excluded. We then developed a 3-step deep-learning-based pipeline to identify hip and pelvic DICOM files from the above database, process them, and organize their data

into a formal THA radiography registry (Fig. 1). Each step is briefly introduced below. Implementation details for all steps are described in Appendix Supplement 1.

### Step 1: File Screening and Image Retrieval

To better understand the data at hand, we programmatically processed the metadata of all DICOM files in the institutional radiography database. Specific metadata values relevant for constructing a radiography registry (Table I) were recorded and described statistically. All files that had missing, inaccessible, or corrupted imaging data were excluded. We retrieved the imaging data for all remaining DICOM files from the previous step and saved them as grayscale images in PNG format. All of the DICOM files were anonymized and saved after running the pipeline, to facilitate potential future research with outside institutions.

### Step 2: Appearance Classification

We developed 2 deep-learning algorithms to automatically determine the anatomic area captured in the radiograph (i.e., hip, pelvis, or neither), the view of the radiograph (i.e., anteroposterior hip, anteroposterior pelvis, lateral hip, or oblique hip), the visible joint (i.e., left, right, or both), and the presence of a prosthesis (i.e., yes or no). First, an EfficientNet-B3<sup>12</sup> deep-learning algorithm was developed to look at an input radiograph and automatically categorize it into 1 of the following 10 categories: anteroposterior hip or pelvis, right lateral hip without prosthesis, right lateral hip with prosthesis, left lateral hip without prosthesis, left lateral hip with prosthesis, right oblique hip without prosthesis, right oblique hip with prosthesis, left oblique hip without prosthesis, left oblique hip with prosthesis, or not a standard hip or pelvic radiograph. Next, a YOLOv5<sup>13</sup> deep-learning object detection algorithm was developed to further organize the curated anteroposterior hip or pelvis radiographs into 8 categories: right anteroposterior hip without prosthesis, right anteroposterior hip with prosthesis, left anteroposterior hip without prosthesis, left anteroposterior hip with prosthesis, anteroposterior pelvis without prosthesis, anteroposterior pelvis with prosthesis on the right, anteroposterior pelvis with prosthesis on the left, and anteroposterior pelvis with prostheses on both sides. Together, the 2 deep-learning models could classify any input radiographs into 1 of the 17 categories (16 of which shown in Fig. 2).

### Step 3: Annotation of Acetabular Angles

To further enrich our registry, we utilized 2 previously developed deep-learning algorithms to measure acetabular component inclination and anteversion angles on postoperative anteroposterior pelvic and lateral hip radiographs, respectively<sup>14</sup>. All ground truth measurements of the acetabular components were performed by 2 orthopaedic surgeons. The development process of these algorithms, along with the definition of measured acetabular angles, are available in the original source and briefly described in Appendix Supplement 1. We applied these algorithms on all anteroposterior pelvic and lateral radiographs obtained from the previous step, except for those considered nonstandard according to the implementation protocol of the mentioned algorithms.

## Implementation and Validation

To build the final registry, the outputs of the above steps were collated, and the obtained data set was saved as a single file in CSV format. To validate the registry and following appropriate sample-size estimation (see Appendix Supplement 1), we collected a random test sample of 5,000 DICOM files from the radiography database (none of which were previously utilized for training or validating the deep-learning algorithms), and 2 authors (P.R. and B.K.) abstracted the same data points as a gold-standard comparison. The total and class-wise accuracy, precision, recall, and F1 score, along with the confusion matrix for all comparisons, were reported for the performance of the deep-learning algorithms utilized in step 2. The deep-learning algorithms used in step 3 were previously validated by Rouzrokh et al.<sup>14</sup> and Hevesi et al.<sup>15</sup>.

## Source of Funding

This work was supported by National Institutes of Health (NIH) grants R01AR73147 and P30AR76312.

## Results

Of the 846,988 DICOM files from 20,378 THA patients in the institutional radiography database, 110,672 (13.1%) were excluded in step 1 for missing, inaccessible, or corrupted imaging data. The remaining 736,316 files were saved in PNG format to be further analyzed by deep-learning algorithms. Appendix Supplement 2 shows the availability of metadata tags described in Table I for all DICOM files from the institutional radiography database. Although some metadata such as “PatientID,” “PatientSex,” and “StudyDate” are mandatory and therefore were consistently present in all files, several other metadata fields were missing for a substantial number of files.

The performance of the trained EfficientNetB3 and YOLOv5 deep-learning algorithms is described in Appendix Supplement 2. Following the application of these algorithms in step 2, an additional 98,660 (11.6%) DICOM files were excluded because they were identified by the deep-learning algorithms to be nonstandard hip or pelvic radiographs, yielding a total of 209,332 (24.7%) excluded DICOM files in total. Among those, some were from other anatomic areas (e.g., knee or ankle) that could not be detected with use of DICOM metadata on retrieval. In addition, 45 DICOM files had metadata indicating non-hip/pelvic anatomic areas (and would have been excluded if they had been screened solely on the basis of their metadata), yet the classifier accurately identified them as hip or pelvic radiographs. Other excluded files were hip or pelvic radiographs that had nonstandard views (e.g., those captured in the operating room) or were of poor quality. The remaining files included 168,551 anteroposterior pelvic, 176,890 anteroposterior hip, 174,637 lateral hip, and 117,578 oblique hip radiographs from 20,266 patients. In step 3, a total of 92,445 acetabular component inclination angles and 92,367 acetabular component anteversion angles were annotated on postoperative anteroposterior pelvic and lateral hip radiographs, respectively, in 16,955 patients.

The final THA radiography registry was assembled in <8 hours by processing every DICOM file from the institutional radiography database according to steps 1 through 3. Table II demonstrates 4 sample rows from the CSV file generated for the final radiography registry. This file includes all desired patient-level, study-level, and imaging-level data for every DICOM file from the radiography database.

For the test sample of 5,000 DICOM files, no files with imaging data were mistakenly excluded during the file screening and image retrieval step (step 1). Together, the 2 algorithms (EfficientNetB3 and YOLOv5) achieved an accuracy, precision, recall, and F1 score of 99.9%, 99.6%, 99.5%, and 99.6%, respectively (Table III). Only 27 of the 5,000 files were mistakenly classified compared with gold-standard manual annotations. Of these, 26 errors were the result of classifying a nonstandard radiograph as standard or vice versa, and only 1 error occurred as a result of a misclassification of an anteroposterior hip radiograph as an oblique hip radiograph. Examples of errors are demonstrated in Figure 3. Appendix Supplement 3 plots the confusion matrix for all 5,000 comparisons.

The mean difference (and standard deviation) between human-level and machine-level measurements of the fully automated (i.e., no human annotation) acetabular component inclination and anteversion annotation algorithms was 1.3 (1.0°) and 1.4 (1.3°) for inclination and anteversion angles, respectively. Differences of 5° or more between human-level and machine-level measurements were observed in <2.5% of cases (n = 14).

## Discussion

This study utilized a deep-learning-based pipeline to build a radiography registry of hip and pelvic radiographs from a pool of 846,988 DICOM files derived from a large cohort of 20,378 THA patients from an institutional arthroplasty registry. Our approach was based on interpreting DICOM metadata and relying on multiple deep-learning algorithms to appropriately curate and annotate radiographs in a fully automated fashion. Once developed and validated, the collection of algorithms was able to construct the final radiography registry of 637,656 included radiographs in <8 hours with an accuracy of 99.9%. Furthermore, our deep-learning pipeline was able to accurately identify DICOM files by imaging appearance even in the presence of inaccurate metadata values, highlighting the utility of the automated method in resource-constrained environments and the ability to approximate human annotator performance for labeling radiographs.

The process for creating our THA radiography registry using a deep-learning pipeline is a stepwise approach that can be replicated and refined by other institutions. In fact, we are presently creating radiography registries for other anatomic areas with use of similar methodologies. In addition to supporting patient care and research within individual institutions, establishing a facile workflow for radiography registry development may potentiate multicenter collaboration or support the enhancement of national registries. Through techniques such as natural language processing-enhanced data abstraction and deep-learning image characterization<sup>16–19</sup>, registries such as the American Joint Replacement Registry can be poised to answer questions on a scale not previously possible.

The efforts to curate this database are analogous to an investment in critical infrastructure. Although the registry itself does not answer clinical questions of interest, it is a critical foundational effort to enable a myriad of studies and more integrated longitudinal surveillance. In the preoperative realm, a database of this type can now be utilized to understand the radiographic natural history of the hip prior to arthroplasty. By comparisons of preoperative and postoperative radiographs, questions such as those regarding the restoration of leg length and offset can be explored. In the postoperative phase, the artificial intelligence pipelines from this registry can be built on to evaluate the change in implant position or change in bone quality on serial radiographs. Indeed, we have since performed a study with this registry showing that deep-learning algorithms can detect femoral component subsidence of as little as 0.1 mm<sup>20</sup>.

The present article should be interpreted with consideration of the following limitations. First, although the deep-learning algorithms were highly accurate, a small number of errors did occur; however, these errors were far more infrequent than human errors discovered in DICOM metadata. Furthermore, most of the algorithm errors were the result of radiographs that were nonstandard or of poor quality. Importantly, performance was superior to previous error rates of deep-learning algorithms on other medical imaging data sets<sup>21,22</sup>. Second, our deep-learning algorithms were trained on imaging data from 1 institution and during a specific time frame (2000 to 2020). A natural limitation to all deep-learning algorithms is that they may not generalize with the same level of accuracy to imaging data from other institutions or even images from another time frame at the same institution with substantially different qualities. However, the fine-tuning of an already-trained algorithm on images from different institutions is much more straightforward than de novo algorithm development and can improve the overall performance and generalizability of the algorithm with each external fine-tuning event.

In conclusion, we constructed an accurate and organized hip and pelvic radiography registry for THA patients with use of a pipeline of deep-learning algorithms for performing this task in a fully automated fashion. The stepwise approach detailed in this report may enable the establishment of similar radiography registries by other institutions, or even national registries. Similarly, the methodology can be applied to other anatomic areas in orthopaedic surgery with a similar workflow. As imaging data represent a noted limitation to most current large databases, this process has implications for enabling integrated clinical and radiographic research and improving patient characterization and surveillance.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## References

1. Gliklich R, Leavy M. Patient registries and rare diseases. *Appl Clin Trials*. 2011;20(3):1–5.
2. Gliklich RE, Dreyer NA, Leavy MB, editors. *Registries for Evaluating Patient Outcomes: A User's Guide*. 3rd ed. Rockville: Agency for Healthcare Research and Quality (US); 2014.
3. Hachamovitch R, Peña JM, Xie J, Shaw LJ, Min JK. Imaging Registries and Single-Center Series. *JACC Cardiovasc Imaging*. 2017 Mar;10(3):276–85. [PubMed: 28279375]

4. Bhatt DL, Drozda JP Jr, Shahian DM, Chan PS, Fonarow GC, Heidenreich PA, Jacobs JP, Masoudi FA, Peterson ED, Welke KF. ACC/AHA/STS Statement on the Future of Registries and the Performance Measurement Enterprise: A Report of the American College of Cardiology/American Heart Association Task Force on Performance Measures and The Society of Thoracic Surgeons. *J Am Coll Cardiol*. 2015 Nov 17;66(20):2230–45. [PubMed: 26440214]
5. Graham RNJ, Perriss RW, Scarsbrook AF. DICOM demystified: a review of digital file formats and their use in radiological practice. *Clin Radiol*. 2005 Nov;60(11):1133–40. [PubMed: 16223609]
6. Varma DR. Managing DICOM images: Tips and tricks for the radiologist. *Indian J Radiol Imaging*. 2012 Jan;22(1):4–13. [PubMed: 22623808]
7. Bidgood WD Jr, Horii SC, Prior FW, Van Syckle DE. Understanding and using DICOM, the data interchange standard for biomedical imaging. *J Am Med Inform Assoc*. 1997 May-Jun;4(3):199–212. [PubMed: 9147339]
8. Chai J, Zeng H, Li A, Ngai EWT. Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*. 2021 Dec 15;6:100134.
9. Erickson BJ, Korfiatis P, Kline TL, Akkus Z, Philbrick K, Weston AD. Deep Learning in Radiology: Does One Size Fit All? *J Am Coll Radiol*. 2018 Mar;15(3 Pt B):521–6. [PubMed: 29396120]
10. Cai L, Gao J, Zhao D. A review of the application of deep learning in medical image classification and segmentation. *Ann Transl Med*. 2020 Jun;8(11):713. [PubMed: 32617333]
11. Zhou SK, Greenspan H, Davatzikos C, Duncan JS, Van Ginneken B, Madabhushi A, et al. A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises. *Proc IEEE*. 2021 May;109(5):820–38.
12. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2020.
13. Jocher G, Stoken A, Borovec J. NanoCode012, Chaurasia A, TaoXie, et al. ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations. Zenodo; 2021. <https://zenodo.org/record/4679653#.Yp9-pnbMI2x>
14. Rouzrokh P, Wyles CC, Philbrick KA, Ramazanian T, Weston AD, Cai JC, et al. Deep Learning Tool for Automated Radiographic Measurement of Acetabular Component Inclination and Version Following Total Hip Arthroplasty. *J Arthroplasty*. 2021;36(7):2510–2517. [PubMed: 33678445]
15. Hevesi M, Wyles CC, Rouzrokh P, Erickson BJ, Lewallen DG, Taunton MJ, et al. Redefining the 3D Topography of the Acetabular Safe Zone. A Multivariable Study Evaluating Prosthetic Hip Stability. *J Bone Joint Surg Am*. 2022 Feb 2;104(3):239–45. [PubMed: 34958643]
16. Fu S, Wyles CC, Osmon DR, Carvour ML, Sagheb E, Ramazanian T, Kremers WK, Lewallen DG, Berry DJ, Sohn S, Kremers HM. Automated Detection of Periprosthetic Joint Infections and Data Elements Using Natural Language Processing. *J Arthroplasty*. 2021 Feb;36(2):688–92. [PubMed: 32854996]
17. Sagheb E, Ramazanian T, Tafti AP, Fu S, Kremers WK, Berry DJ, Lewallen DG, Sohn S, Maradit Kremers H. Use of Natural Language Processing Algorithms to Identify Common Data Elements in Operative Notes for Knee Arthroplasty. *J Arthroplasty*. 2021 Mar;36(3):922–6. [PubMed: 33051119]
18. Tibbo ME, Wyles CC, Fu S, Sohn S, Lewallen DG, Berry DJ, Maradit Kremers H. Use of Natural Language Processing Tools to Identify and Classify Periprosthetic Femur Fractures. *J Arthroplasty*. 2019 Oct;34(10):2216–9. [PubMed: 31416741]
19. Wyles CC, Tibbo ME, Fu S, Wang Y, Sohn S, Kremers WK, Berry DJ, Lewallen DG, Maradit-Kremers H. Use of Natural Language Processing Algorithms to Identify Common Data Elements in Operative Notes for Total Hip Arthroplasty. *J Bone Joint Surg Am*. 2019 Nov 6;101(21):1931–8. [PubMed: 31567670]
20. Rouzrokh P, Wyles CC, Kurian SJ, Ramazanian T, Cai JC, Huang Q, Zhang K, Taunton MJ, Maradit Kremers H, Erickson BJ. Deep Learning for Radiographic Measurement of Femoral Component Subsidence Following Total Hip Arthroplasty. *Radiol Artif Intell*. 2022 May 4;4(3):e210206. [PubMed: 35652119]
21. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases [abstract]. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017 July 21–26. Honolulu: IEEE; 2017.



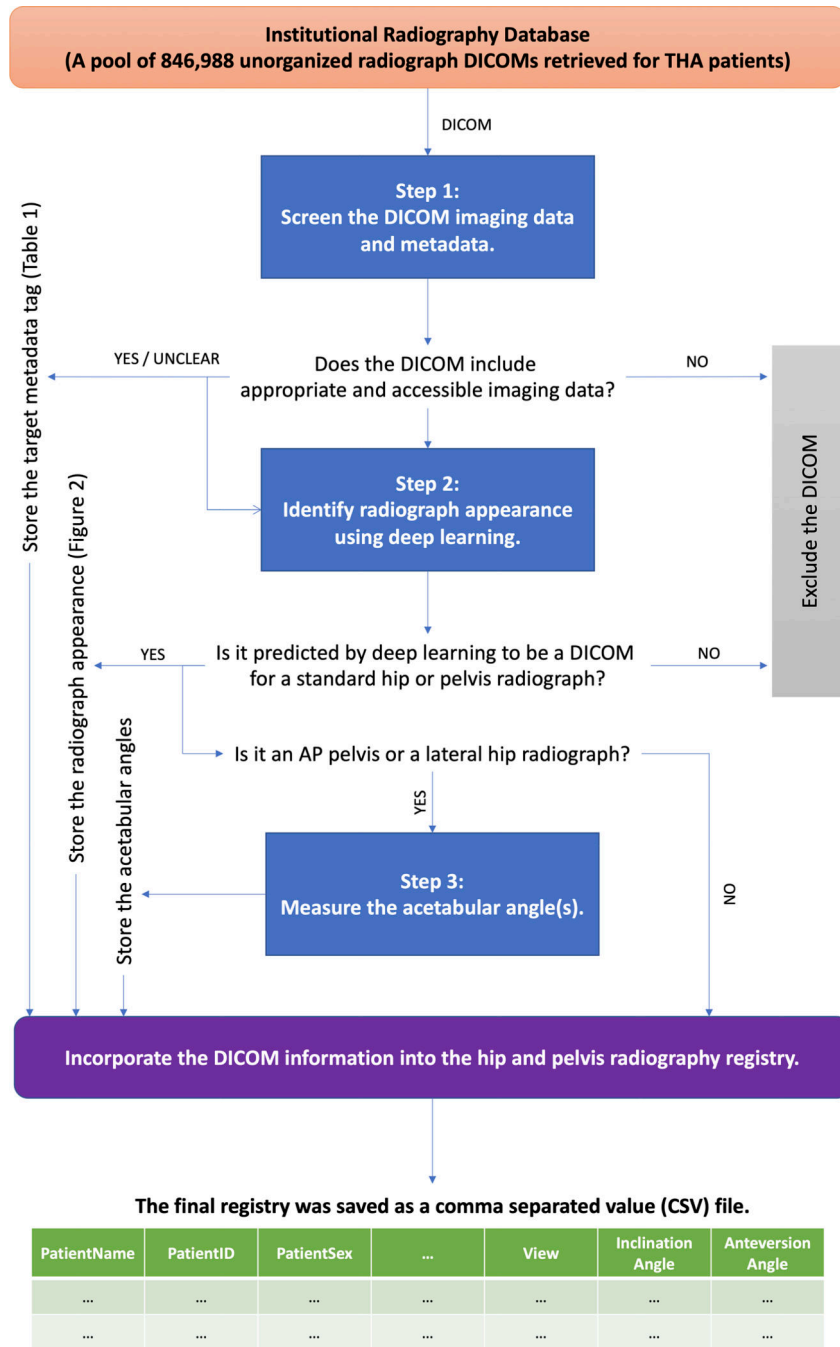
22. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, et al. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. 2019.

Author Manuscript

Author Manuscript

Author Manuscript

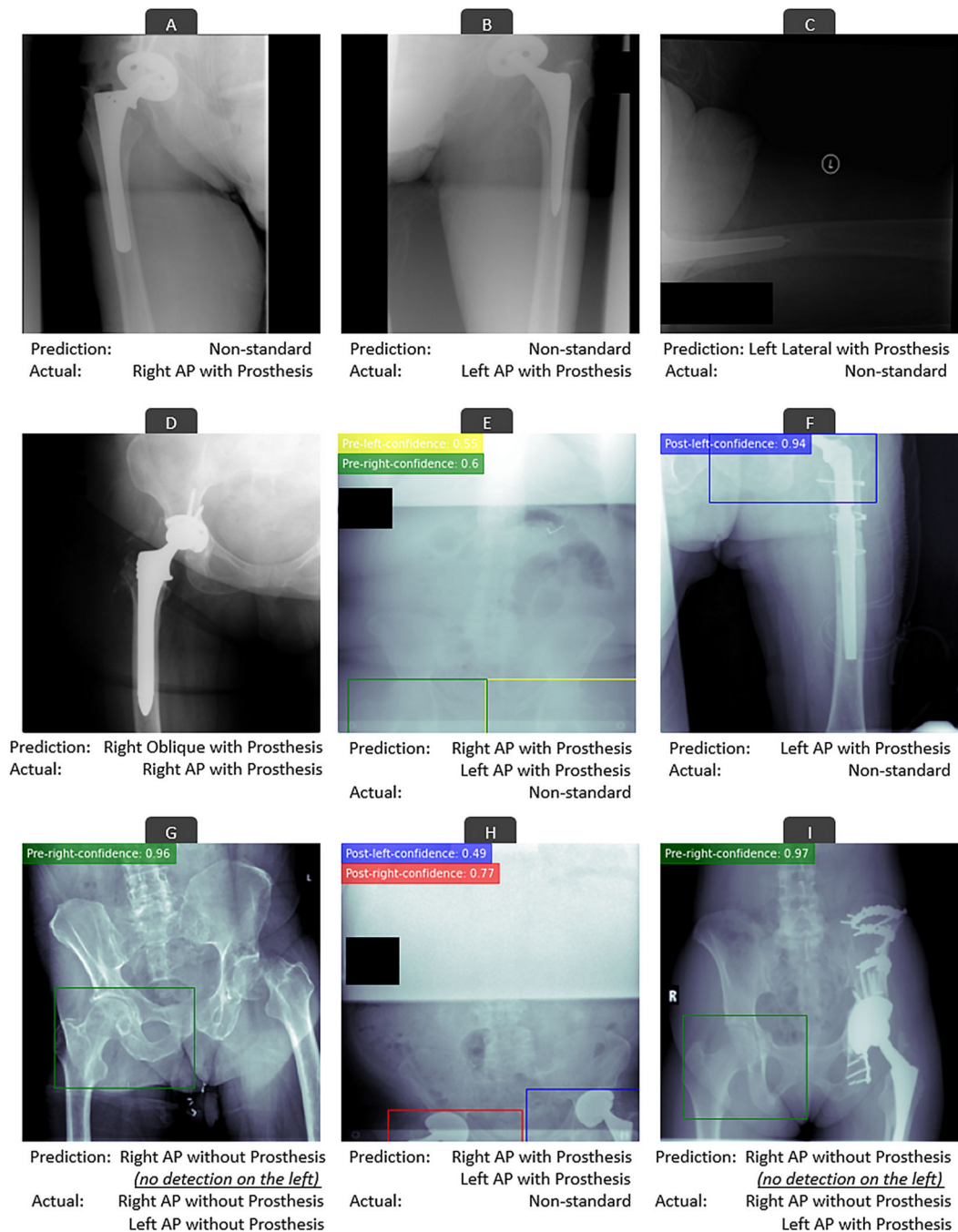
Author Manuscript



**Fig. 1.** The 3-step deep-learning-based pipeline designed to build a hip and pelvic radiography registry for THA patients. Every DICOM file from the institutional radiography database goes through the 3 steps in this pipeline to be processed and merged into the THA radiography registry. AP = anteroposterior.



**Fig. 2.** Examples of radiographic appearance classes (RACs) assigned to every DICOM file in the radiography database. Lateral, oblique, and anteroposterior hip radiographs will each have 4 categories based on the presence of the prosthesis and the side of the visible joint (Figs. 2-A **through** 2-L). Anteroposterior pelvic radiographs have 4 possible RACs based on the presence of the prosthesis on neither side, right side, left side (not shown on the figure for brevity), or both sides of the pelvis (Figs. 2-M, 2-N, **and** 2-O). Poorly taken or non-hip/pelvic radiographs were considered as nonstandard (Fig. 2-P).

**Fig. 3.**

Examples of errors of the deep-learning-based pipeline in classifying the radiographic appearance of DICOM files. Colored rectangles on anteroposterior (AP) radiographs show how the YOLOv5 deep-learning algorithm detected hips with or without prostheses. These detections were utilized for further classifying the anteroposterior hip or pelvis radiographs. The numbers on the top left side of AP radiographs show how confident the YOLOv5 algorithm was in its predictions. Most errors occurred as a result of nonstandard radiographs (Figs. 3-C, 3-E, 3-F, and 3-H), radiographs containing less prevalent hardware variations

(Fig. 3-I), or radiographs with lower quality or malpositioning (Figs. 3-A, 3-B, **and** 3-G). Nevertheless, the underlying reason for a few errors are not clear (Fig. 3D).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**TABLE I**

Important DICOM Metadata for Building the Radiography Registry of Hip and Pelvic Radiographs \*

Tag	Keyword	Importance
(0010, 0010)	PatientName	Identifying the patient
(0010, 0020)	PatientID	Identifying the patient
(0010, 0040)	PatientSex	Identifying the patient
(0032, 1032)	RequestingPhysician	Identifying the corresponding physician for the patient
(0008, 0082)	InstitutionNumber	Identifying the location of radiographs (radiographs from different institutes may have unique properties important for further research)
(0008, 0070)	Manufacturer	Identifying the manufacturer of the radiography device (radiographs from different manufacturers may have unique properties important for further research)
(0008, 0015)	BodyPartExamined	Identifying the body parts targeted for radiography
(0008, 1030)	StudyDescription	Identifying the body parts targeted for radiography
(0008, 103E)	SeriesDescription	Identifying the body parts targeted for radiography
(0008, 0020)	StudyDate	Identifying the date of radiography and sorting of radiographs
(0008, 0030)	StudyTime	Identifying the time of radiography and sorting of radiographs
(0008, 0018)	SOPInstanceUID	Identifying and sorting of radiographs; each DICOM has a unique SOPInstanceUID (e.g., anteroposterior view of the hip in a patient)
(0020, 000E)	SeriesInstanceUID	Identifying and sorting of radiographs; several radiographs (e.g., lateral and anteroposterior view of the hip in a patient) could be captured in a series and thus have similar SeriesInstanceUIDs
(0020, 000D)	StudyInstanceUID	Identifying and sorting of radiographs; multiple series (e.g., anteroposterior and lateral views of the hip and anteroposterior and lateral views of the knee) can be captured for a single patient in a single study and thus, all DICOM files from that study will have similar StudyInstanceUIDs
(0008, 0060)	Modality	Identifying the type of radiograph (e.g., conventional vs. digital radiography); digital and conventional radiographs have different qualities that may be important for future research
(0028, 0010)	Rows	Identifying the resolution of the radiograph
(0028, 0011)	Columns	Identifying the resolution of the radiograph
(0018, 1164)	ImagerPixelSpacing	Identifying the real (physical) distance in the patient between the centers of adjacent pixels and estimating the real measurements on radiographs
(0028, 0004)	PhotometricInterpretation	Pixel data from DICOM files with a "Monochrome 1" value for this metadata element are inverted (i.e., actual black pixels are visualized as white in the radiograph and vice versa); such pixel data need to be re-inverted to the actual value before saving to disc

\* Keyword, and the reason for importance, are mentioned for each metadata tag. Please note that not all mentioned tags are present in every DICOM file and that files contain more tags than shown in the above table. Only those deemed important for building a radiography registry are listed.

**TABLE II**  
 Sample Entries from a CSV-Format File Created for the Final Hip and Pelvic Radiography Registry\*

Patient Name	Patient Sex	Patient ID	Series Description	Surgery Date	Study Date	Photometric Interpretation	DICOM Path	Pixel Array Shape: (Height, Width)	Radiograph Appearance	Inclination Angle (°)	Anteversion Angle (°)
J.S.	Male	12342	Ant. hip x-ray	20100402	20100409	Monochrome2	././dicom1.dcm	(3000, 2980)	AP pelvis with prosthesis on right	20.1	
N.A.	Female	56781	Anterior – left side	20150101	20141104	Monochrome2	././dicom2.dcm	(2048, 2092)	AP left hip without prosthesis		
W.W.	Male	19823	Lateral hip x-ray	20090222	20090229	Monochrome1	././dicom3.dcm	(3100, 2200)	Lateral right hip with prosthesis		14.3
E.E.	Female	45232		20040810	20010110	Monochrome2	././dicom4.dcm	(3024, 2048)	Oblique left hip without prosthesis		

\* The real version of this file contains rows in which the entries are separated by commas, and could easily be opened with use of any software for reading tabular data or data sets. Please note that (1) not all columns are shown for brevity, yet this dataset stores values for all metadata introduced in Table I; (2) values in the final 3 columns are generated with use of deep-learning algorithms; (3) values in the “Series Description” column are directly extracted from the DICOM metadata and as explained in the text, do not necessarily follow a consistent formatting; and (4) all values are synthetic to keep real patient data confidential. AP = anteroposterior.

**TABLE III**

Performance of the Deep-Learning-Based Pipeline for Classification of the Radiographic Appearance of 5,000 Radiographs from the Test Sample \*

<b>Radiographic Appearance Class</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1 Score</b>
Anteroposterior left hip without prosthesis	99.9%	98.0%	97.0%	97.5%
Anteroposterior right hip without prosthesis	99.9%	95.7%	100.0%	97.8%
Anteroposterior left hip with prosthesis	99.9%	97.9%	98.5%	98.2%
Anteroposterior right hip with prosthesis	99.9%	100.0%	97.5%	98.7%
Anteroposterior pelvis without prosthesis	99.9%	98.7%	98.0%	98.0%
Anteroposterior pelvis with prosthesis on right	99.9%	99.8%	99.8%	99.8%
Anteroposterior pelvis with prosthesis on left	99.9%	100.0%	98.9%	99.4%
Anteroposterior pelvis with prosthesis on both sides	99.9%	99.5%	99.7%	99.6%
Lateral left hip with prosthesis	100.0%	100.0%	100.0%	100.0%
Lateral right hip with prosthesis	100.0%	100.0%	100.0%	100.0%
Lateral left hip without prosthesis	100.0%	100.0%	100.0%	100.0%
Lateral right hip without prosthesis	100.0%	100.0%	100.0%	100.0%
Oblique left hip with prosthesis	100.0%	100.0%	100.0%	100.0%
Oblique right hip with prosthesis	100.0%	99.7%	100.0%	99.8%
Oblique left hip without prosthesis	100.0%	98.3%	100.0%	99.1%
Oblique right hip without prosthesis	100.0%	100.0%	100.0%	100.0%
Nonstandard	99.5%	98.9%	98.2%	99.0%
Total	99.9%	99.6%	99.5%	99.6%

\* The reported metrics are neither for the EfficientNetB3 nor the YOLOv5 algorithms, but rather for the outcomes of a pipeline that is based on both. The independent performance of each algorithm is separately discussed in Appendix Supplement 2.