COPYRIGHT © BY THE JOURNAL OF BONE AND JOINT SURGERY, INCORPORATED

http://dx.doi.org/10.2106/JBJS.21.01229

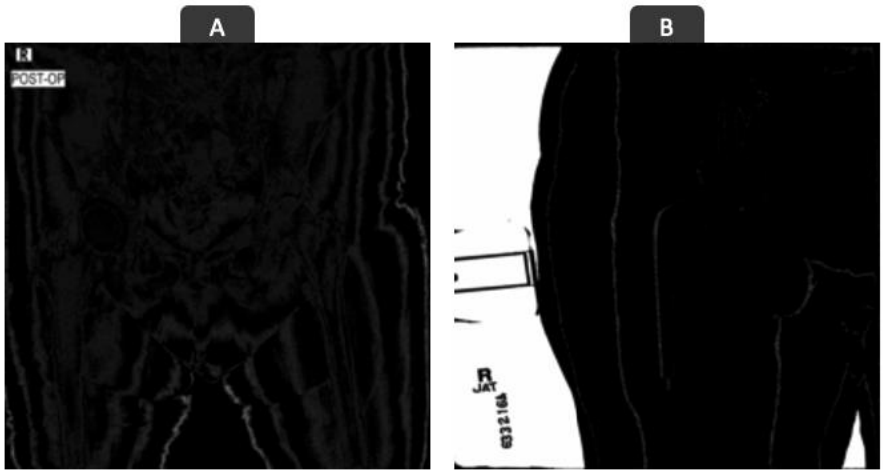
The following content was supplied by the authors as supporting material and has not been copy-edited or verified by JBJS.

Supplement 1

In this section, we explain additional tips and implementation details for each of the three steps of building our registry. Please note that all our processes were run on Python 3.7, and multiprocessing using 18 CPU cores on a computer machine with 80 Gigabytes (GB) of Random Access Memory (RAM) was done whenever appropriate, to reduce the duration of operations.

- **Step 1: DICOM Screening and Image Retrieval:**
    - To save the imaging data from DICOMs, we extracted that data as NumPy arrays, clipped the pixel values of each image between their 2.5 - 97.5 percentiles, padded the image with zero pixels to make it square-shaped, resized it to 224*224-pixel size, and rescaled it to 8-bit format. The final image was then saved to disk using the Pillow package running in Python. For those DICOMs that had *Monochrome1* value in their *PhotometricInterpretation* tag, the imaging array was first inverted before doing the above operations.
    - Most x-ray images that are taken with digital machines have a DICOM tag named "VOI LUT" (0028, 3010). This tag denotes some linear or nonlinear function that should be applied to the raw pixel values of a DICOM image in order to make it visible in the form of an image that is both printable and viewable by a computer screen. The linear form of this operation performs windowing using window center value (*c*) and window width value (*w*) in a way that pixel values that are either greater than $c + w/2$ or smaller than $c - w/2$ are set to zero (black) and

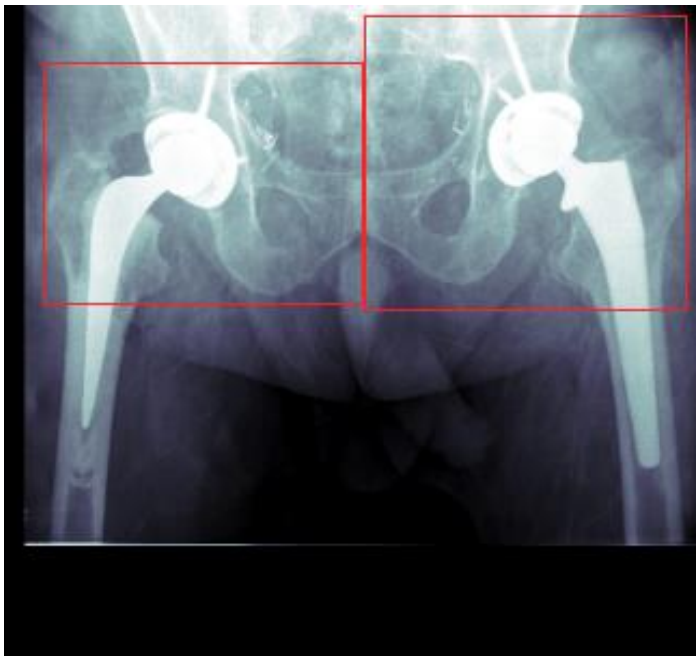
pixels in between are mapped linearly between (0, 255), so that can be viewed in a monitor or be printed. The non-linear transformations use a sigmoid function to create values in the range of (0, 255). Applying these functions before feeding an image to a deep learning model would make the image more sensible. However, there are some images that have corrupted VOI LUT values which distort the image. By doing an overview of 10,000 images, we found that more than 90% of these corrupt images could be filtered from our initial pool by counting the number of unique pixel values after applying the VOI LUT function. If the image had less than 100 unique values, it would be filtered out as a non-standard image. For the remaining DICOMs with corrupted VOI LUT tag that could not be excluded in this step, we relied on our EfficientNetB3 Classifier model to detect and further exclude them in step 2. The figure below shows two examples of such images that have corrupted VOI LUT tags with 85 (A) and 92 (B) unique values, respectively.



- **Step 2: Appearance Classifications:**
    - Training details for the EfficientNetB3 Algorithm: One author (PR) manually labeled the appearance of 25,000 random images retrieved in step 1. To develop the EfficientNetB3 model, we considered all AP radiographs as a single category, yielding a dataset of 10 different classes (described in the main text). Any radiograph without a standard AP pelvis, AP hip, lateral hip, or oblique hip appearance was considered non-standard. Following the pre-processing of labeled images, we used stratified random sampling to split the data into 80% of training and 20% of validation subsets, based on the PatientIDs. The initial weights for the EfficientNetB3 model were pre-trained on the ImageNet database. We retrained all the weights using an Adam Optimizer, a learning rate of 0.001, a categorical cross-entropy loss function, and a batch size of 64. Augmentations included a random rotation in a range of 15 degrees, random zooming of 0 - 10%, and the addition of gaussian noise. The training was done for 20 epochs and the version of the model with the least validation loss was saved as the final model.
    - Training details for the YOLOv5 Algorithm: One author (QJ Medical Student) went through the labeled AP radiographs from the previous step and further annotated them by manually drawing a bounding box around the joint area(s). Boxes were plotted in a way that their medial boundary crossed the pubic symphysis, their lateral boundary was tangent to the lateral border of the ipsilateral femur, their inferior boundary was tangent to the ipsilateral ischial tuberosity, and their superior border was either tangent to the most superior part

of the ipsilateral prosthesis (if a prosthesis was present in the joint) or the most superior part of the ipsilateral acetabulum (if no prosthesis was present in the joint). The figure below shows a sample annotation as described. Each bounding box was assigned one of the following labels: right-side-no-prosthesis, right-side-with-prosthesis, left-side-no-prosthesis, left-side-with-prosthesis. Following the pre-processing of labeled images and their annotated bounding boxes, we used them to train a YOLOv5 deep learning object detection algorithm to automatically localize and label hip(s) on any given AP pelvis or hip radiograph from our database (i.e., to automatically draw bounding boxes around the joint area and give it one of the four mentioned labels). The training was done using the Ultralytics implementation of YOLOv5, for 50 epochs, on a batch size of 64, with a startling learning rate of 0.01, which was reduced gradually and we used Adam as the optimizer. All default augmentations were applied except for random horizontal flipping which could change the left or right labels. The model that achieved the highest mean Average Precision (mAP0.5) was selected as the final model. The output of the YOLOv5 algorithm could further be used to classify a given AP radiograph into one of eight appearances: right hip without prosthesis, right hip with prosthesis, left hip without prosthesis, left hip with prosthesis, pelvis with no prosthesis, pelvis with left side prosthesis, pelvis with right side prosthesis, and pelvis with prosthesis on both sides. For example, suppose the algorithm detects one right-sided joint with prosthesis and one left-sided joint without prosthesis on a given AP radiograph. In that case, the image will be

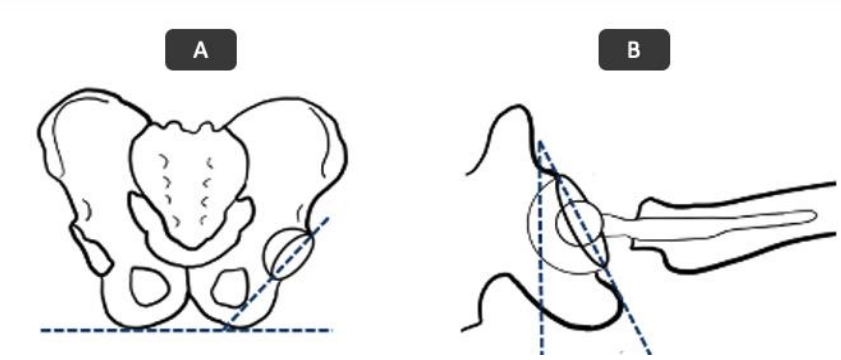
classified as a pelvis radiograph with a prosthesis on the right side (pelvis with right side prosthesis).



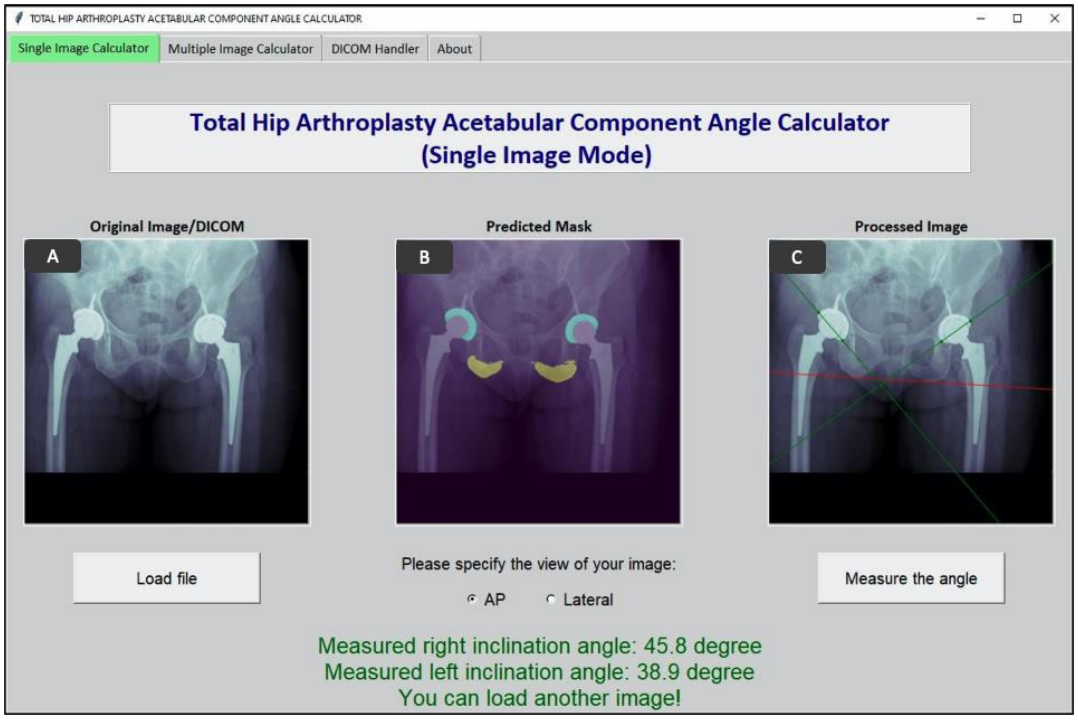
- **Step 3: Annotation of Acetabular Angles:**
    - Details of the deep learning algorithms used in this step are described in (14). Briefly, two cohorts of 600 AP pelvis and 600 cross-table lateral hip postoperative radiographs were used to train deep learning models to segment the acetabular component (on both AP pelvis and cross table lateral hip radiographs) and the ischial tuberosities (only on AP pelvis radiographs). Cohorts were manually annotated and randomly split to train-validation-test datasets on an 8:1:1 basis. Following appropriate augmentation, two U-Net convolutional neural networks (CNN) models (one for AP and one for cross-table lateral radiographs) were trained for 50 epochs. Acetabular component angles were then measured by

applying image processing on the predicted masks for the anatomical landmarks in radiographs.

- The figure below demonstrates how the acetabular inclination (A) and anteversion (B) angles were defined for the above measurements:



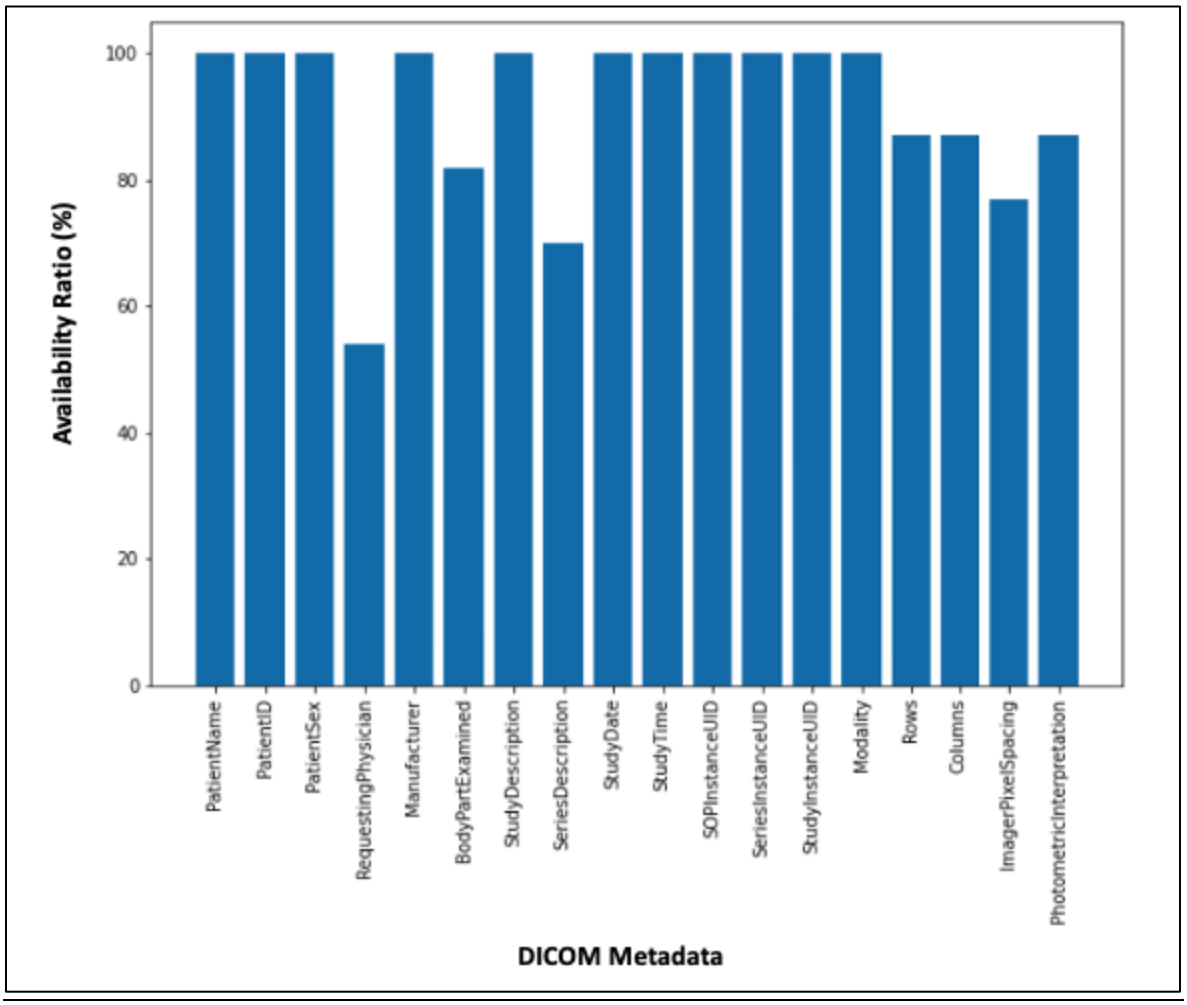
- The figure below demonstrates the graphic user interface (GUI) put together for automatic measurement of acetabular angles on postoperative radiographs. (A) is an input AP pelvis radiograph, (B) is the segmentation mask generated for A by a deep learning model, and (C) shows how the algorithm measured the acetabular inclination angles on B (while overlapping that measurement on A for clarity).



- **Validation:**
  - To estimate the sample size needed for validation of our pipeline, we manually annotated the radiograph appearance of 5,000 randomly selected DICOMs from the radiograph database. The sample size needed to validate the pipeline's performance was then estimated using the following formula: $n = N*X / (X + N - 1)$, where, $X = Z_{\alpha/2}^2 -*p*(1-p) / MOE^2$, and $Z_{\alpha/2}$ is the critical value of the Normal distribution at $\alpha/2$ (we picked a confidence level of 95%; hence α is 0.05 and the critical value is 1.96), *MOE* is the margin of error which we picked at 1%, *p* is the maximum proportion which was observed to be 15.4% in our sample of 1000 radiographs, and *N* is the population size (=846,988), needing us to hold a sample of 4977 DICOMs.

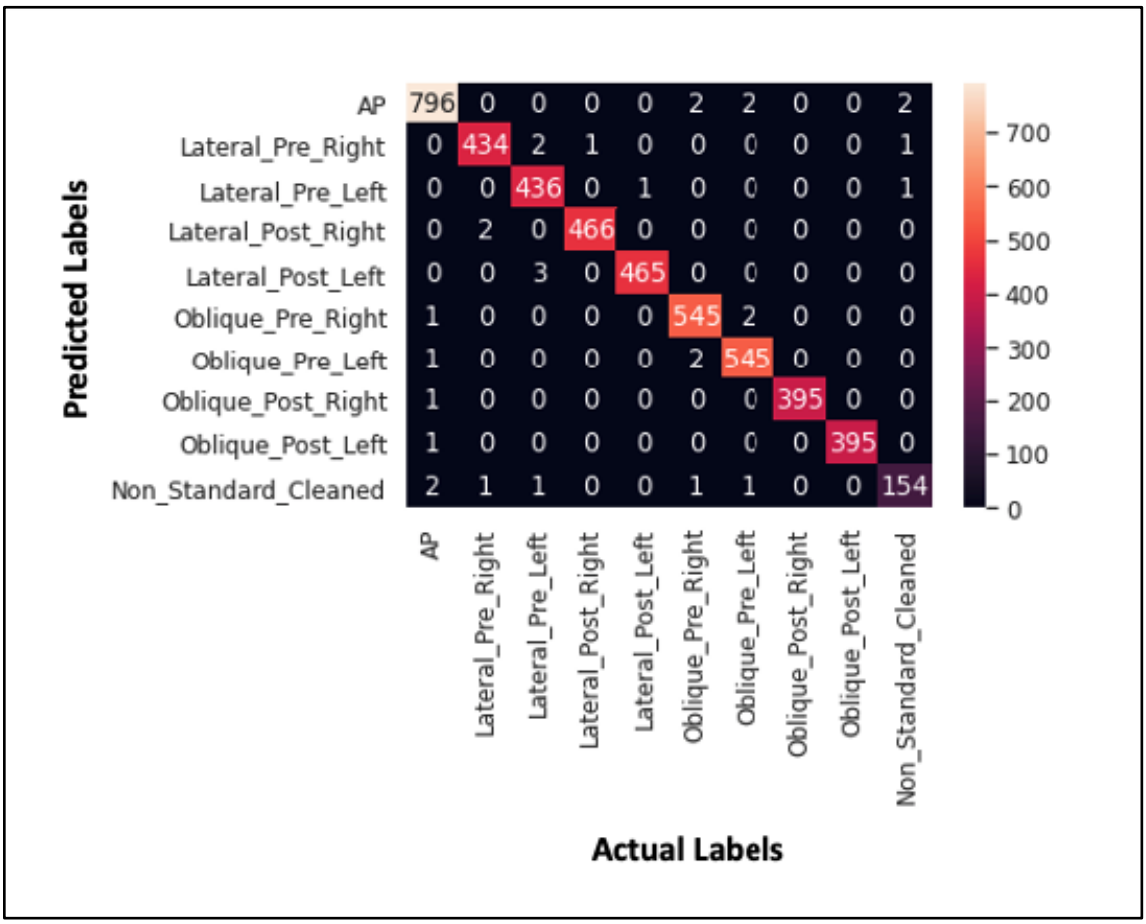
Supplement 2

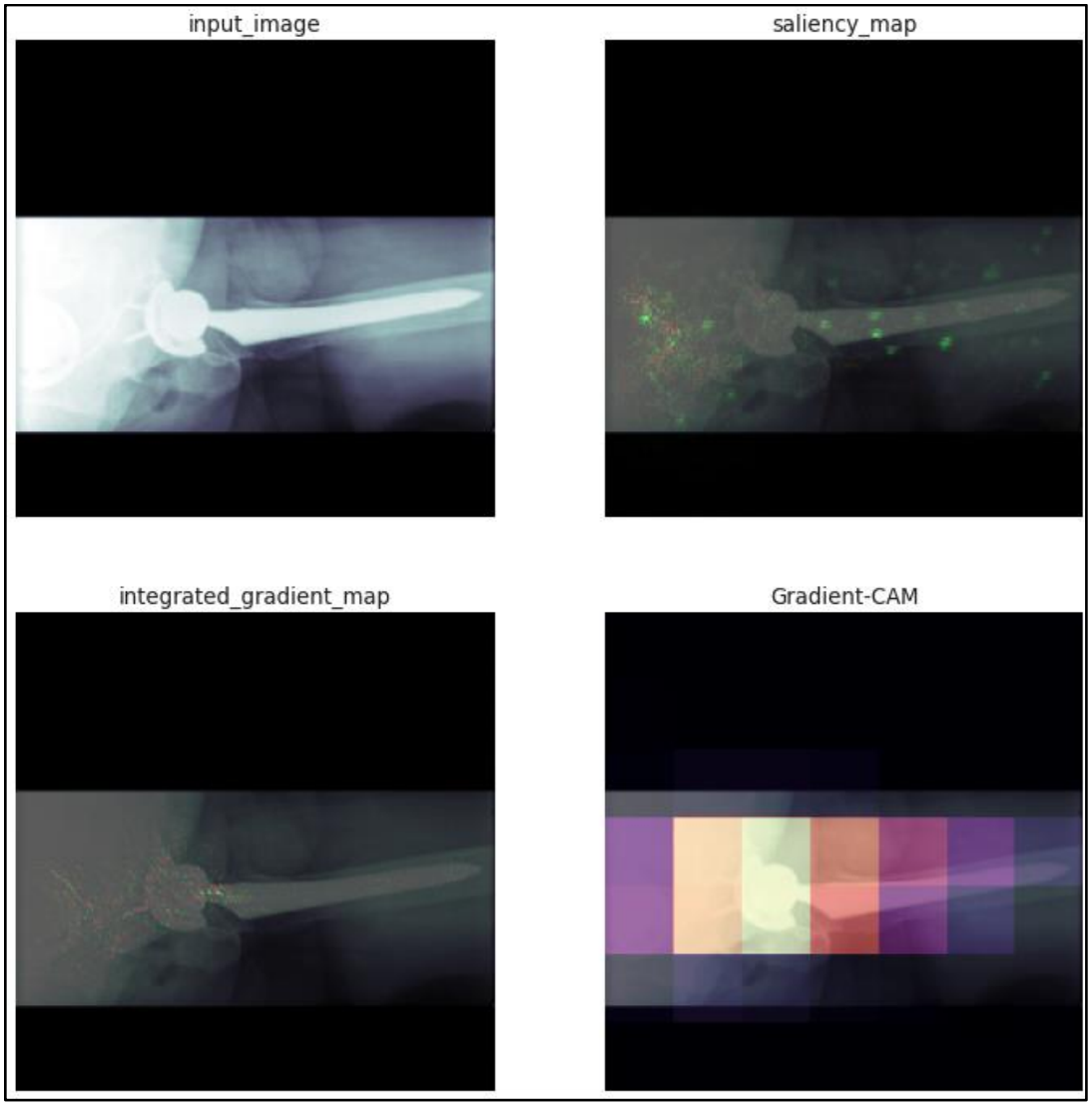
- Availability of metadata tags described in Table 1 for all DICOMs from the institutional radiography database.



- Training Results for the EfficientNetB3 Model: The least validation loss for the EfficientNetB3 classifier algorithm was achieved after 11 epochs of training. The final algorithm reached an accuracy of 98.3% on the validation set. The following figures show the confusion matrix plotted for the validation samples (first image) and a combination of model prediction interpretation maps (saliency map, integrated gradient

map, and class activation map) for a sample image from the validation set (second image).

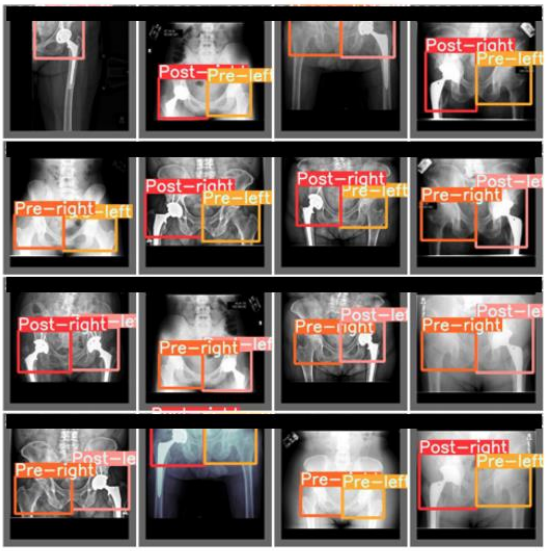




- Training results for the YOLOv5 model: The highest mAP for the YOLOv5 object detection algorithm was achieved on epoch 32. The final algorithm reached a mean average precision (mAP) of 99.4%, a precision of 99.8%, and a recall of 99.3%. The image below compares the ground truths and predictions of the YOLOv5 model on one sample batch of the validation data:

COPYRIGHT © BY THE JOURNAL OF BONE AND JOINT SURGERY, INCORPORATED

http://dx.doi.org/10.2106/JBJS.21.01229



Ground Truth

Predictions

Supplement 3

The confusion matrix for all comparisons made between the manual and automatic appearance labeling of the 5,000 test sample radiographs. The classes in the matrix are coded as follows: (1) AP left hip without prosthesis, (2) AP right hip without prosthesis, (3) AP left hip with prosthesis, (4) AP right hip with prosthesis, (5) AP pelvis without prosthesis, (6) AP pelvis with prosthesis on the right, (7) AP pelvis with prosthesis on the left, (8) AP pelvis with prosthesis on both sides, (9) Lateral left hip with prosthesis, (10) Lateral right hip with a prosthesis, (11) Lateral left hip without prosthesis, (12) Lateral right hip without prosthesis, (13) Oblique left hip with prosthesis, (14) Oblique right hip with prosthesis, (15) Oblique left hip without prosthesis, (16) Oblique right hip without prosthesis, (17) Non-standard.

	Predicted																	
Actual	**Class**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**Total**
	1	98	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	101
	2	0	90	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90

	3	0	0	192	0	0	0	0	0	0	0	0	0	0	0	0	3	195
	4	0	0	0	238	0	0	0	1	0	0	0	0	0	1	0	0	4	244
	5	0	1	0	0	153	0	0	0	0	0	0	0	0	0	0	0	2	156
	6	0	0	0	0	0	465	0	0	0	0	0	0	0	0	0	0	1	466
	7	0	3	0	0	0	0	349	0	0	0	0	0	0	0	0	0	1	353
	8	0	0	0	0	0	0	0	399	0	0	0	0	0	0	0	0	1	400
	9	0	0	0	0	0	0	0	0	398	0	0	0	0	0	0	0	0	398
	10	0	0	0	0	0	0	0	0	0	450	0	0	0	0	0	0	0	450
	11	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	100
	12	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	99
	13	0	0	0	0	0	0	0	0	0	0	0	0	299	0	0	0	0	299
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	313	0	0	0	313

	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	57	0	0	57
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45	0	45
	17	2	0	4	0	2	0	0	1	0	0	0	0	0	0	1	0	1224	1234
	Total	100	94	196	238	155	466	349	401	398	450	100	99	299	314	58	45	1238	5000