



Contents lists available at ScienceDirect

## Journal of Orthopaedic Translation

journal homepage: [www.journals.elsevier.com/journal-of-orthopaedic-translation](http://www.journals.elsevier.com/journal-of-orthopaedic-translation)

## Original Article

## Machine learning–based identification of hip arthroplasty designs

Yang-Jae Kang<sup>a,b</sup>, Jun-Il Yoo<sup>c,\*</sup>, Yong-Han Cha<sup>d</sup>, Chan H. Park<sup>e</sup>, Jung-Taek Kim<sup>f</sup><sup>a</sup> Division of Applied Life Science Department at Gyeongsang National University, PMBBRC, Jinju, Republic of Korea<sup>b</sup> Division of Life Science Department at Gyeongsang National University, Jinju, Republic of Korea<sup>c</sup> Department of Orthopaedic Surgery, Gyeongsang National University Hospital, Jinju, South Korea<sup>d</sup> Department of General Surgery, Eulji University Hospital, Daejeon, South Korea<sup>e</sup> Department of Orthopedic Surgery, Yeungnam University Medical Center, Daegu, South Korea<sup>f</sup> Department of Orthopedic Surgery, Ajou University School of Medicine, Ajou Medical Center, Suwon, South Korea

## ARTICLE INFO

## Keywords:

Hip arthroplasty  
Identification  
Machine learning

## ABSTRACT

**Background:** The purposes of this study were to develop a machine learning–based implant recognition program and to verify its accuracy.**Methods:** Postoperative anteroposterior (AP) X-rays ( $\geq 300$  dpi) were collected of patients who underwent total hip arthroplasty. X-rays with a wire or plate added and those without a true anteroposterior view were excluded. A total of 170 X-ray images of hip implants from 29 brands were collected from five hospitals and a Google image search. These collected images were manually reorganised to ensure appropriate labelling. Collected images were preprocessed to have grey-scaled pixels with histogram equalisation for efficient training. Images varied by  $+10^\circ$  and  $-10^\circ$ , and 3606 unique images derived from the original 170 images were created for training. Discussion of the validation set being derived 25% of training set. The recognition model structure consisted of two steps: object detection and clustering. Model training was performed with Keras deep learning platform.**Results:** The 170 X-ray images of hip implants were used to build a stem detection model using YOLOv3. Manually labelled images were successfully trained into the stem detection model. Evaluation of 58 newly labelled X-ray images showed highly accurate stem detection (mean average precision  $> 0.99$ ). Fully connected layers generated 29 class outputs. After training, a receiver operating characteristic curve was generated with a test set containing 25% of all stem-cropped images, yielding an area under the curve of 0.99.**Conclusion:** Femoral stem identification in patients with total hip arthroplasty was very accurate. This technology could be used to collect large-scale implant information.**The translational potential of this article:** This program has the following clinical relevance. First, we can prepare the implants needed for revision surgery by identifying the old types of implants. Second, it can be used to diagnose peripheral osteolysis or periprosthetic fracture by further developing the ability to sensitise implant detection. Third, an automated implant detection system will help organise imaging data systematically and easily for arthroplasty registry construction.

## Introduction

The rates of hip disease and fragility fracture are increasing as the population ages. One of the best methods for treating elderly patients is total hip arthroplasty (THA) [1,2]. Many kinds of implants are available, and the rate of revision is increasing [3,4]. Revision surgery success depends on the preparation of the device including collecting informa-

tion on previous implants such as the size and type. Because the number of revision patients has increased in the past 30 years, patients may not know information about their previous surgery [5]. For such cases, it is difficult to prepare stem revision preoperatively.

Stem type, taper size, and design differences among appliance companies are very important for selecting a surgical option. It is therefore very important to know such information beforehand based on preop-

\* Corresponding author. Department of Orthopaedic Surgery, Gyeongsang National University Hospital, 90 Chilamdong, Jinju, Gyeongnamdo, 660-702, South Korea.

E-mail address: [furim@daum.net](mailto:furim@daum.net) (J.-I. Yoo).

<https://doi.org/10.1016/j.jot.2019.11.004>

Received 17 May 2019; Received in revised form 11 September 2019; Accepted 21 November 2019

Available online 20 December 2019

2214-031X/© 2019 The Authors. Published by Elsevier (Singapore) Pte Ltd on behalf of Chinese Speaking Orthopaedic Society. This is an open access article under the

CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

erative X-rays. This knowledge will enable the management of implants that will be developed in the future. Image-based machine learning programs have recently been applied to diagnose many diseases [6–9]. In particular, patient care quality is increasing because of the use of image-based machine learning programs in cancer diagnosis and diabetic retinopathy (DR) [9].

The purposes of this study were to develop a machine learning-based implant recognition program and to verify its accuracy.

## Materials and methods

### Materials

Postoperative hip anteroposterior (AP) X-rays were collected after THA. True AP hip radiographs are taken in a supine position, one of the most common mistakes is image distortion as the hip is externally rotated. Thus, either both patellae should be facing forward or lower extremities should be internally rotated by 15° and –20° to accommodate femoral anteversion in AP hip radiographs [10]. Only those with ≥300 dpi were included in this study. X-rays with a wire or plate added and pictures without a true AP were excluded. Twenty-nine products from Stryker (Mahwah, NJ 07430, United States), DePuy (Warsaw, Indiana, United States), Zimmer Biomet (Warsaw, Indiana, United States), Smith & Nephew (London, United Kingdom), Lima (Villanova, San Daniele del Friuli, Italy), Corentec (South Korea), and other manufacturers were collected from five hospitals.

### Implant classification model construction

A total of 170 X-ray images of hip implants from 29 brands were collected from five hospitals (Table 1). These collected images were manually reorganised to ensure correct labelling. Collected images were preprocessed to have grey-scaled pixels using a histogram equalisation enhancing the outlines of images. This method allows the images to be properly compared for efficient training. Collected images were augmented by flipping horizontally, and image orientation was changed from –10 to 10°.

The recognition model structure consisted of two steps: (1) object detection by YOLOv3 (<https://pjreddie.com/darknet/yolo/>).

For YOLOv3 object detection, stem portions of hip implants in X-ray images were square-labelled using a piece of software (<https://github.com/tzutalin/labelImg>). Model training was performed with the Keras deep learning platform. Beta-evaluation was performed using Flask (<https://github.com/pallets/flask>).

## Results

We first collected 170 X-ray images of hip implants and manually square-labelled the stems to build a stem detection model using YOLOv3.

**Table 1**  
Number of collected images for each brand.

Brand	Count	Brand	Count
Accolade II	18	DePuy-AML-CDH	4
Aesculap Bicontact	2	DePuy-Coraill	4
AML	10	DePuy-ICS	3
Bencox II	17	DePuy-SROM	4
BENCOX Long stem revised	8	Implantcast-EcoFit	1
Biomet-mallory	3	LimaCorporate-C2	7
Biomet-microplasty	2	M-stem	12
Biomet-Stanmore	5	MLtaper	11
Biomet-Taperloc	5	SNN-Echelon	4
CLS	12	SNN-Spectron	2
Corentec-Bencox II	4	SNN-Synergy	3
DePuy Tri-lock BPS	2	STRYKER-Howmedica PCA	4
DePuy C-STEM	4	Zimmer-Fitmore	4
DePuy Proxima	10	Zimmer-FMT	5
Total			170

Because of the small number of images, we tried to overfit the object detection model on all images; however, the validation of 58 newly labelled X-ray images showed high accurate stem detection (mean average precision, > 0.99).

It is very hard to collect high-quality X-ray image of every brands of stem. To overcome this problem, we planned to build the classification model to extract the features from images as feature extractor. Then, we can extract the feature vectors from any X-ray images of stem (object extracted by YOLOv3) that can even include the new stem from the outside of training set. Using the output feature vectors, we implement the clustering among the stems in our database and the query stem to find the nearest stem brands.

To carry out the plan, first, image augmentations including histogram equalisation, flipping, and rotation were applied to increase the amount of training data, resulting in 3606 augmented X-ray images (Fig. 1). Then, we split the images into the train and test set with ratio of 3:1, respectively. Manually labelled images were successfully trained into the stem classification model; a feature extractor based on the CNN architecture. Using the output feature vectors from the model, we implement clustering that could properly group the query image into the known brands (Fig. 2).

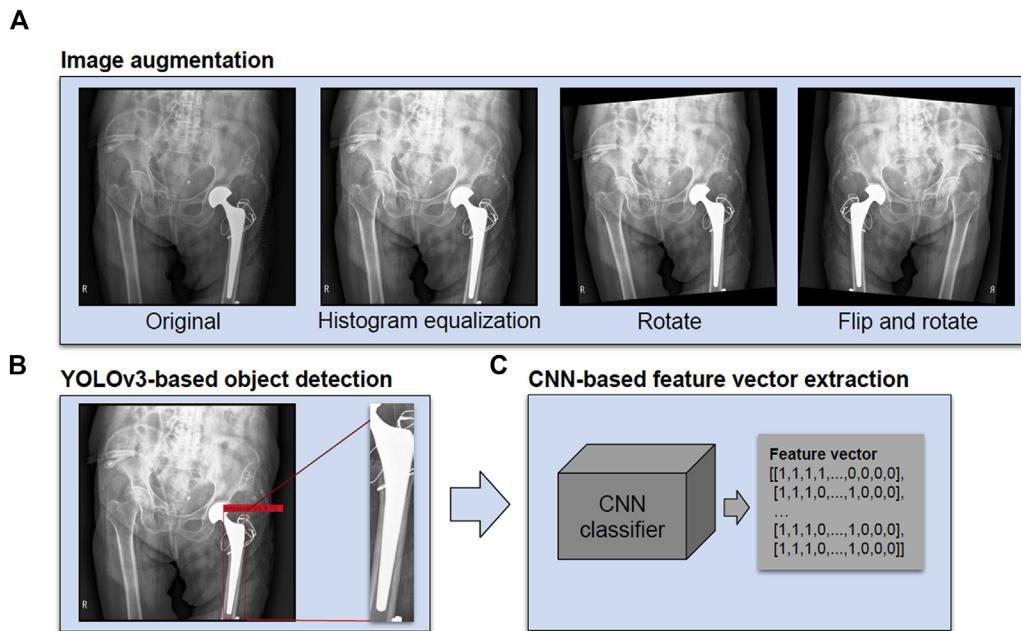
Our simple CNN architecture consisted of six layers. Inputs (224 × 224 grey-scale images) were passed through two layers of convolution and one max pooling layer to generate a feature map and feed into two fully connected layers that generated 29 class outputs. After training, the receiver operating characteristic curve was drawn with test set containing 25% of all stem-cropped images, showing an area under the curve (AUC) of 0.99 (Fig. 3).

## Discussion

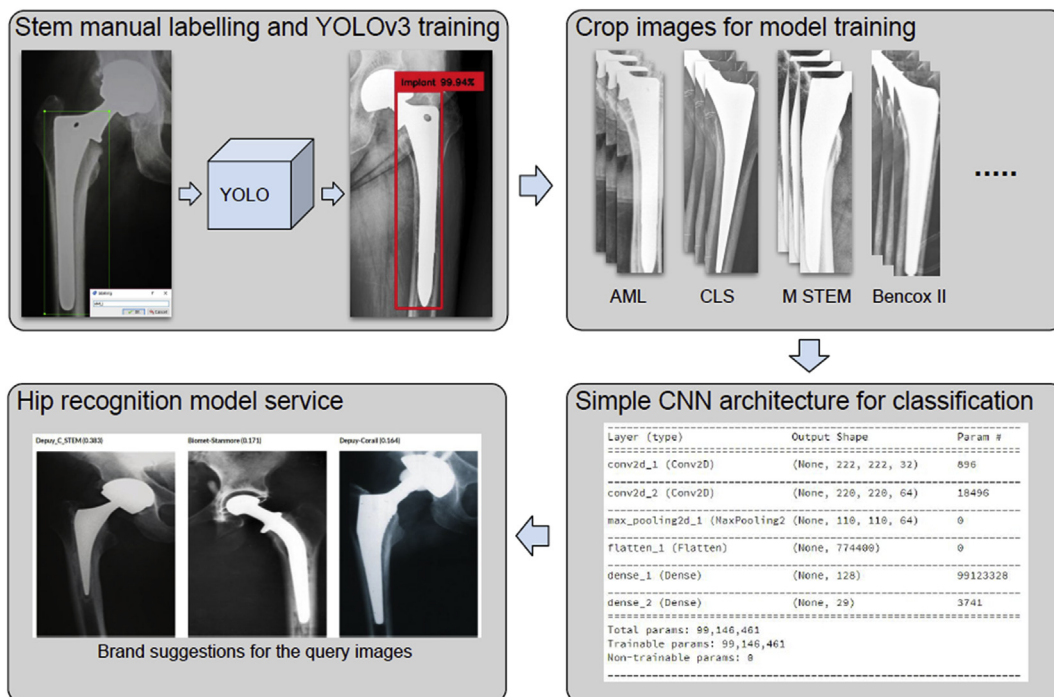
The main result of this study was that instrument identification was very accurate, with an AUC of 0.99 despite the small amount of data. DR is the most commonly used disease in deep learning-based diagnosis. Verbraek et al [9] performed a study to determine the real-world accuracy of diagnosing DR in 1616 patients with type 2 diabetes. They reported that the hybrid deep learning-enhanced device's sensitivity/specificity against the reference standard was 100%/97.8% for vision-threatening DR and 79.4%/93.8% for more than mild DR.

Conventionally, practices in orthopaedic surgery performed based on the subjective opinion of surgeons are changing thanks to the use of machine learning and deep learning. Milimonfared et al [11] performed a study to implement a computer-based visual scoring method for objectively rating corrosion damage to metallic stem tapers of retrieved hip implants. In addition, they have tailored existing methodologies to develop an intelligent algorithm, using digital image processing and machine learning so that corrosion scores can be automatically and objectively assigned for retrieval studies. In the present study, we were able to quantify taper damage more objectively, which could affect the postoperative results. In the present study, the AUC for X-rays in the test set was very accurate. In addition, object detection was very accurate when image quality was good. However, if similar product and version were changed, the accuracy of the delineation was reduced. In the future, it will be necessary to construct a database using X-rays published in articles and provided by the company. Nonetheless, this study will help surgeons plan revision surgeries.

The selection of a surgical instrument for revision THA is very difficult if there is no hospital record from the original procedure. When revision THA is performed with stem retention, it is important to identify the stem type preoperatively to select the bearing. In addition, it would be possible to presort revision options and prepare bearing options for each company in advance. Furthermore, if the application information is provided in the implant database, surgery preparation will be much easier. Through these techniques, it is possible to construct large-scale implant datasets by systematically collecting various information and identifying problems or improvements.



**Figure 1.** Schematic prediction procedures (A) image augmentations; rotation, flipping, and histogram equalisation; (B) object detection for stem part extraction using YOLOv3; and (C) feature extraction from stem image by trained CNN classifier.

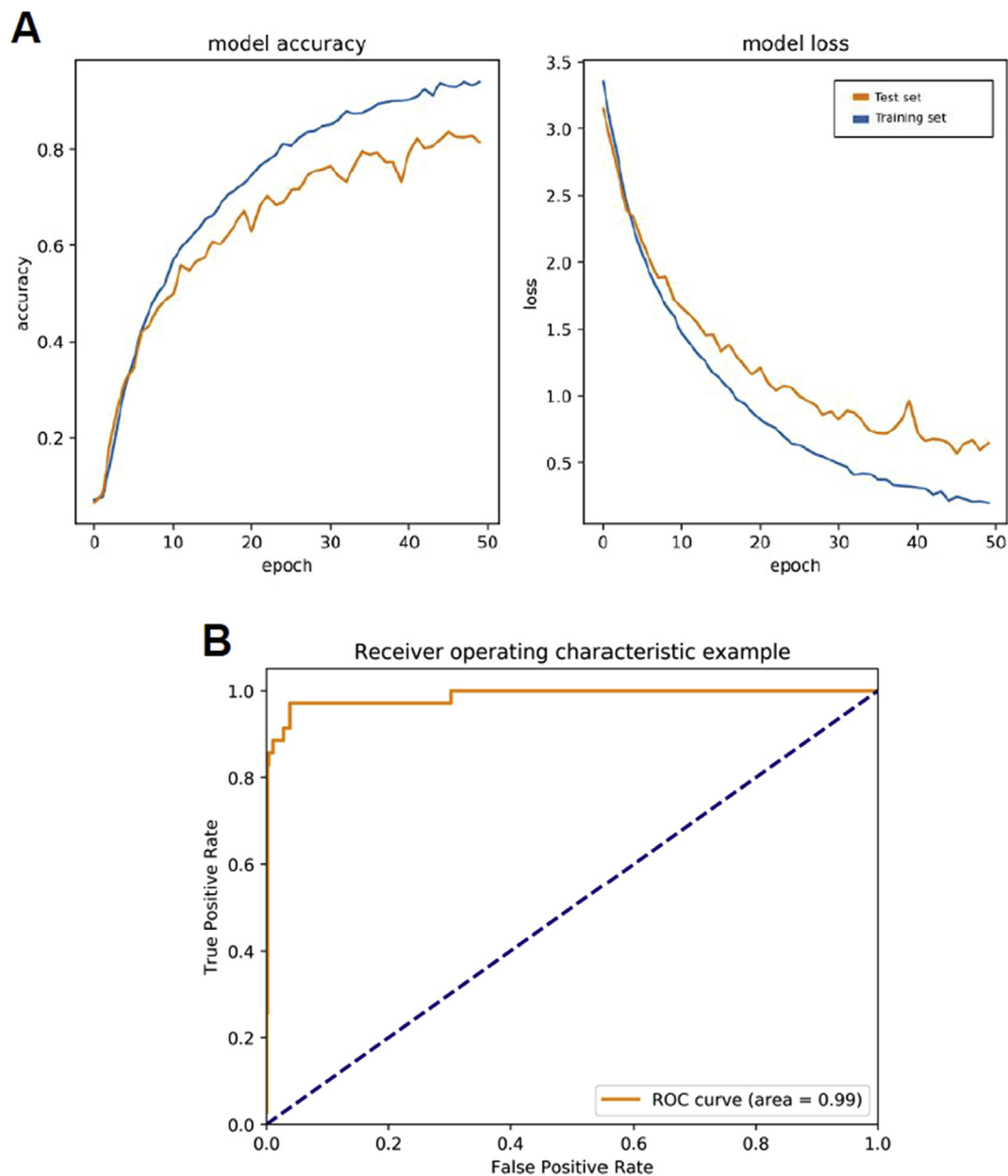


**Figure 2.** Strategy for CNN classifier training by stem-extracted images and hip implant brand suggestion service scheme.

This study has several limitations. First, implant classification accuracy was poor for low-quality photographs. A hip implant X-ray image database needs to be developed using a well-designed quantitative image collection protocol for the internet. Second, the number of images for each brand was highly varied because of data collection limitations. We tried to overcome this by performing image augmentation, but this can result in overfitting problems. Hospital-wide image collection efforts would be necessary for continuous update of our feature extractor. Third, there was a personal information problem associated with creating applications. Each country is subjected to artificial intelligence-based privacy protection regulations. Consensus agreements will be required to

overcome this barrier. Moreover, it may not be cost-effective to simply use the application for implant identification. In the future, it will be used to construct a large-scale database of artificial joints and expanded to various other implants. In addition, we are developing software that can help users plan surgeries by constructing databases of information and precautions for different implants. In addition, high-quality image databases such as computed tomography scans can be used to help develop next-generation implants.

This program has the following clinical relevance. First, we can prepare the implants needed for revision surgery by identifying the old types of implants. Therefore, it is possible to prevent stem removal caused by



**Figure 3.** Result of CNN classifier training. (A) Accuracy increasing and loss decreasing for training and test set; (B) averaged ROC curve for the predicted classes showing 0.99 mean AUC. AUC = area under the curve; ROC = receiver operating characteristic.

improper preparation of the bearing option. Second, it can be used to diagnose peripheral osteolysis or periprosthetic fracture by further developing the ability to sensitise implant detection. Third, an automated implant detection system will help organise imaging data systematically and easily for arthroplasty registry construction. Finally, in the near future, implant information such as taper size, stem size, and angle included in the implant detection system will help make our surgical planning easier and more accurate.

In conclusion, the identification of selected femoral stems in patients with THA was very accurate with an AUC of 0.99. In the future, this technology can be used to collect large-scale implant information.

#### Ethics statement

The study was approved by the Ethics Committee of GNUH, and written informed consents were obtained from the participant.

#### Conflict of interest

The authors have no conflicts of interest to disclose in relation to this article.

#### Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. NRF-2019R1F1A1059208).

#### Funding/support statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors, and no material support of any kind was received.

## References

- [1] Grau L, Summers S, Massel DH, Rosas S, Ong A, Hernandez VH. Operative trends in the treatment of hip fractures and the role of arthroplasty. *Geriatr Orthop Surg Rehabil* 2018;9. 215145931876063.
- [2] Gasbarra E, Perrone FL, Celi M, Rao C, Feola M, Cuozzo N, et al. Total hip arthroplasty revision in elderly patients. *Aging Clin Exp Res* 2013;25:61–3.
- [3] Lee J-M. The current concepts of total hip arthroplasty. *Hip Pelvis* 2016;28:191.
- [4] Schwartz BE, Paponov HI, Helder CW, Mayers WF, Gonzalez MH. Revision total hip arthroplasty in the United States: national trends and in-hospital outcomes. *Int Orthop* 2016;40:1793–802.
- [5] Wolf BR, Lu X, Li Y, Callaghan JJ, Cram P. Adverse outcomes in hip arthroplasty: long-term trends. *J Bone Joint Surg Am* 2012;94:e103–8.
- [6] Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: a literature review. *Front Bioeng Biotech* 2018;6.
- [7] Bini SA. Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? *J Arthroplast* 2018;33:2358–61.
- [8] Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, et al. Deep neural network improves fracture detection by clinicians. *Proc Natl Acad Sci* 2018;115:11591–6.
- [9] Verbraak FD, Abramoff MD, Bausch GCF, Klaver C, Nijpels G, Schlingemann RO, et al. Diagnostic accuracy of a device for the automated detection of diabetic retinopathy in a primary care setting. *Diabetes Care* 2019;42:651–6.
- [10] Lim S-J, Park Y-S. Plain radiography of the hip: a review of radiographic techniques and image features. *Hip Pelvis* 2015;27:125–34.
- [11] Milimonfared R, Oskouei RH, Taylor M, Solomon LB. An intelligent system for image-based rating of corrosion severity at stem taper of retrieved hip replacement implants. *Med Eng Phys* 2018;61:13–24.