Development and Validation of a Deep Learning Model for
Preoperative Virtual Reduction of Proximal Humerus Fractures

Young Dae Jeon[1], Ki-Bong Park[1], Kwang-Hwan Jung[1], Moo-Sub Kim[2], Hyo-Hun Kim[2], Do-Kun Yoon[2]
[1]Department of Orthopaedic Surgery, University of Ulsan College of Medicine, Ulsan University Hospital, Ulsan, Republic of Korea
[2]Industrail R&D Center, KAVILAB Co. Ltd., Seoul, Republic of Korea

INTRODUCTION: Deep learning has a strong ability to provide useful information in field of orthopaedic surgery over a diagnosis. And now it can automatically establish preoperative surgical plan for orthopaedic trauma including segmentation of fracture fragments, the restored bone shape after automatic reduction. If reduction images of proximal humeral fractures can be provided in advance using a deep learning model, these would contribute to optimize preoperative surgical planning. In this study, our objective was to develop a deep learning model that can automatically establish a virtual reduction in proximal humerus fractures using 3D computed tomography (CT) images, and to evaluate the performance of this deep learning model and its clinical validity.

METHODS: This study was approved by an Ethics Committee. We developed an automatic segmentation method using deep learning and a simulation-based automatic virtual reduction system: deep learning performed the segmentation of fracture fragments and a Monte Carlo simulation completed the virtual reduction to determine the best model. Twenty CT scans of Neer classification 3- or 4-part proximal humerus fractures were used to establish a virtual reduction with manual virtual reduction (MVR) or deep learning virtual reduction (DLVR). The performance outcomes included final accuracies of the segmentation and reduction model, the Dice similarity coefficient (DSC) score for virtual reduction image and postoperative 3D CT image, and working time. The lengths of the inferomedial screws were precisely measured during the DLVR images, and compared with the used screws during surgery. The clinical validation outcomes included the clinical validation agreement (CVA) and reduction quality scores (RQS), which were evaluated by 20 orthopaedic surgeons.

RESULTS SECTION: The DSC of the MVR and the DLVR were 0.78 ± 0.06 and 0.68 ± 0.07, respectively ($p < 0.0001$). The DLVR (49 s) had a shorter working time than the MVR (3,477 s). For the prediction of length for inferomedial support screws, the Pearson's correlation coefficient between the measured screws and the screws used during surgery was 0.9719 ($p < 0.001$). The mean CVA of all cases was $80.50\% \pm 14.47\%$ (DLVR images, $81.13\% \pm 15.04\%$; MVR images, $78.00\% \pm 12.29\%$; $p = 0.547$). The mean RQS was significantly higher in the DVLR images than in the MVR images (91.47 ± 1.12 vs. 89.30 ± 1.62; $p = 0.040$).

DISCUSSION: The model demonstrated high-quality performance for automatic virtual reduction of proximal humeral fractures. To date, the use of deep learning in orthopaedic surgery has been limited to diagnosis, risk analysis, clinical decision support, and outcome prediction; however, we believe that the field of deep learning can be extended to preoperative planning and can replace time-consuming manual work and 3D printing involved in preoperative planning. Nevertheless, this study had some limitations. First, this study did not conduct a real clinical trial evaluating efficiency, cost-effectiveness, and patient-reported outcome measurements but rather analyzed only virtual reduction images. Second, if virtual reduction of fractures is automatically performed before surgery using a deep learning model, surgeons may not benefit from developing 'mental memory' training before surgery. In conclusion, this deep learning model showed high virtual reduction performance and quality with faster working times. This technology may change the paradigm of preoperative planning in orthopaedic surgery.

SIGNIFICANCE/CLINICAL RELEVANCE: (1-2 sentences): This deep learning-based automatic virtual reduction will change the paradigm of preoperative surgical planning in the orthopaedic trauma surgery.
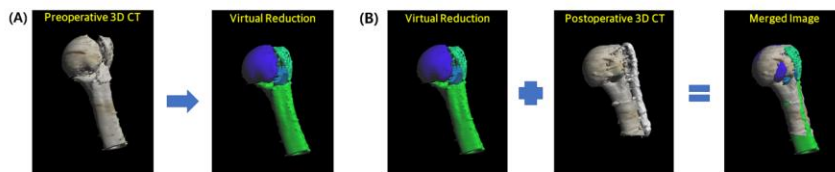
IMAGES AND TABLES:

Figure 1. Preoperative 3D CT images and virtual reduction images. (A) Virtual reduction was carried out by manually or using a deep learning-based method. (B) After manual virtual reduction and deep learning virtual reduction using the preoperative CT image, the virtual reduction images were merged with the postoperative CT image, except for implants and screws, using iterative closest-point algorithm. 3D, three dimensional; CT, computed tomography.
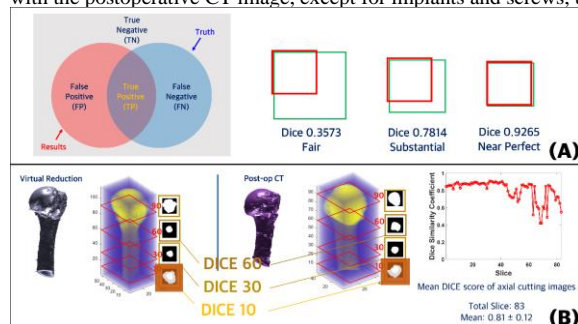

Figure 2. Evaluation of the Dice similarity score (DSC) in the proximal humerus fracture. (A) As the overlapping area increases, the DSC score approaches 1, and becomes near perfect. (B) The DSC scores in each axial cutting image of the virtual reduction and postoperative three-dimensional CT are calculated. After calculating the DSC scores for each merged axial image, the mean DSC score of the whole merged axial image is calculated.
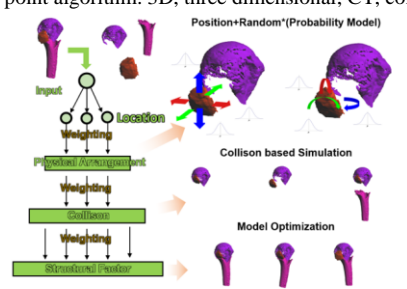

Figure 3. Outline of the Monte Carlo simulation and decision tree for automatic virtual reduction. Monte Carlo simulation provides a random number for the location of the fragments, the physical arrangement (rotation), and the detection threshold for the collision as weighting factors. The decision tree finds the best simulation model from the numerus simulation scenarios using the Dice similarity coefficient score.