

Automatic Assessment of Radiological Parameters of the Distal Radius Using a Hybrid Approach Combining Deep Learning and a Computer-Aided Diagnostic Algorithm

Sang-Jeong Lee, PhD, Minji Kang, MS, Jae-Sung Lee, PhD*, Kyu-Tae Kang, MD*, Hyoung-Seok Jung, PhD^{†,‡}

Multimodal AX Business Team, LG CNS, Seoul,

**Department of Orthopedic Surgery, Chung-Ang University Hospital, Seoul,*

[†]Department of Orthopedic Surgery, Chung-Ang University Gwangmyeong Hospital, Gwangmyeong,

[‡]Biomedical Research Institute, Chung-Ang University Hospital, Seoul, Korea

Background: The use of deep learning algorithms in medical imaging has increased rapidly. This study aimed to develop an automated, hybrid approach combining a deep learning architecture and a conventional computer-aided diagnostic method to detect anatomical landmarks and measure radiological parameters in wrist radiography.

Methods: Overall, 487 wrist radiographs were randomly sampled for training and validation, and 100 radiographs collected from 2 institutions were used as the test set. Anatomical landmarks for 4 commonly used parameters, namely, radial inclination (RI), radial length (RL), volar tilt (VT), and ulnar variance (UV), were identified and labeled. A 2-step hybrid method combining a deep learning model with conventional computer-aided diagnosis was developed to measure radiological parameters using these anatomical landmarks. Measurements were obtained from the test set. The mean value of each parameter determined by 2 hand surgeons served as the reference standard. The performance of the deep learning algorithm was evaluated using the successful detection rate (SDR), mean absolute error (MAE), intraclass correlation coefficient (ICC), and Pearson correlation coefficient (r).

Results: The SDR of the model ranged from 97% to 98% at the 1 mm threshold and 99% at the 2 mm threshold. The RI, RL, VT, and UV generated by the model were $26.41^\circ \pm 4.01^\circ$, 12.63 ± 2.72 mm, $14.01^\circ \pm 8.02^\circ$, and 2.09 ± 2.76 mm, respectively. The overall MAEs for RI, RL, VT, and UV between the manually and automatically measured parameter values were $1.62^\circ \pm 1.26^\circ$, 1.56 ± 1.23 mm, $1.88^\circ \pm 1.68^\circ$, and 0.43 ± 0.41 mm, respectively. In the correlation analysis, good or high reliability was observed for RI, VT, and UV (ICC: 0.86, 0.95, and 0.98; $r = 0.86$, $r = 0.95$, and $r = 0.98$, respectively), and moderate reliability was observed for RL (ICC: 0.75, $r = 0.78$).

Conclusions: This novel automated hybrid method can accurately identify landmarks on wrist radiographs and automatically generate the radiological parameters of the distal radius. This method saves time and reduces human labor in creating datasets for training segmentation models and developing image processing algorithms.

Keywords: Radius, Wrist, Radiography

Received October 11, 2024; Revised May 16, 2025;

Accepted May 16, 2025

Correspondence to: Hyoung-Seok Jung, PhD

Department of Orthopedic Surgery, Chung-Ang University Gwangmyeong Hospital, 110 Deokan-ro, Gwangmyeong 14353, Korea

Tel: +82-2-2610-6641

E-mail: ksjs5739@naver.com

The wrist is one of the most complex joints in the human body. The wrist coordinates the synchronous movements of independent bones, allowing rotational and gliding motions in the coronal and sagittal planes. Although adjunctive modalities such as computed tomography (CT) and magnetic resonance imaging are helpful in providing detailed information about the morphology of the distal

radius, plain radiographic studies are essential for the evaluation of wrist injuries, especially at the initial presentation, during operative treatment, and during routine assessments necessary for follow-up care.¹⁾

Radiographic parameters that characterize the morphology of the distal radius can be used to assess fracture displacement and have predictive value for fracture prognosis. Therefore, surgeons use these parameters in conjunction with patient factors as tools to determine treatment.²⁻⁴⁾ Various radiological parameters have been proposed for assessing the morphology of the distal radius based on wrist radiographs, including radial inclination (RI), radial length (RL), volar tilt (VT), and ulnar variance (UV). The measurement of these radiological parameters mainly involves manual labeling, which is time-consuming for orthopedic surgeons. Manual measurements, relying on clinician experience, may be imprecise and vary significantly among clinicians.⁵⁾

The use of deep learning algorithms in medical imaging has rapidly increased. These algorithms can provide diagnostic information by detecting landmarks using cephalograms and CT scans of the craniomaxillofacial area.^{6,7)} With respect to the distal radius, deep learning has been applied to estimate bone age and detect and localize distal radius fractures.^{8,9)} However, few studies have examined the automatic measurement of radiological parameters of the distal radius using deep learning algorithms.^{10,11)} In addition, since a deep learning algorithm alone is an incomplete method for measuring radiological parameters for diagnosis, a hybrid method has been proposed that applies an image processing algorithm to the outputs of the deep learning algorithm.¹²⁾ Based on experience from clinical practice, the key to precisely measuring radiological parameters is the correct identification and localization of the landmarks. This study aimed to develop an automated

hybrid approach using a deep learning architecture and a conventional computer-aided diagnostic (CAD) approach to detect anatomical landmarks and measure radiological parameters on wrist radiography. We also validated the fully automated measurement of radiological parameters by comparing them with the measurements made by 2 hand surgeons (HSJ and JSL).

METHODS

This retrospective study was approved by the Institutional Review Board of Chung-Ang University Hospital (IRB No. 2201-005-19400). The requirement for informed consent was waived because of the retrospective nature of the study.

Acquisition of Radiographs and Dataset Preparation

Between January 2022 and December 2023, 1,864 patients who underwent wrist radiography in anteroposterior (AP) and lateral (LAT) views were included in our study. All patients were aged 18 years or older to ensure inclusion of only skeletally mature individuals. The premise of manual labeling of the wrist is that orthopedic surgeons can accurately identify landmarks. To ensure confidence in manual labeling, cases with unrecognizable wrist landmarks were excluded. The following exclusion criteria were used to ensure that only radiographs with clear landmarks of the wrists were considered: distal radius or ulnar fractures ($n = 388$), history of distal radius fractures ($n = 51$), wrist osteoarthritis (Kellgren-Lawrence grade > 2 , $n = 156$), and radiographs with other detected abnormalities ($n = 782$). Finally, 487 sets of wrist radiographs were included in the internal dataset. Fifty sets of AP and LAT radiographs were randomly selected as the test set, and the remaining 437 sets were used for training and validation (Fig. 1). To assess the generaliz-

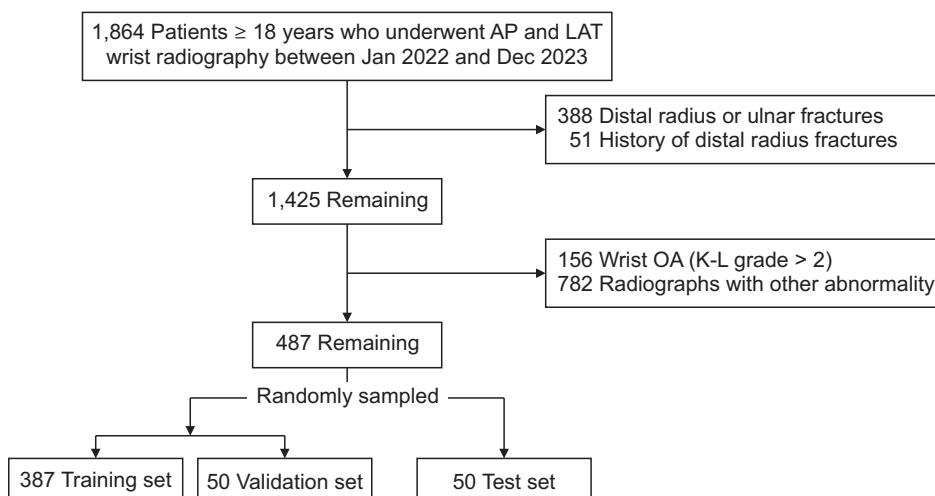


Fig. 1. Flowchart of the patient selection process for this study. AP: anteroposterior, LAT: lateral, OA: osteoarthritis, K-L grade: Kellgren-Lawrence grade.

ability of the proposed method, we performed external validation using an independent dataset collected from a different institution and device. A total of 100 sets of wrist radiographs were obtained from a second institution and served as the external dataset. Among these, 50 sets were randomly selected and added to the training and validation sets of the original dataset, and the remaining 50 sets were added to the original test set. A total of 587 sets of wrist radiographs were collected across the 2 institutions after applying the inclusion and exclusion criteria. From this combined dataset, 487 sets were randomly sampled for training and validation, and the remaining 100 radiographs, consisting of 50 internal and 50 external test cases, were used for testing.

Annotation of Anatomical Landmarks and Measurement of Radiological Parameters

Four commonly used and widely accepted radiological parameters based on AP and LAT wrist radiography were selected to evaluate the model: RI, RL, UV, and VT.^{3-5,13-16} A free, open-source software, 3-dimensional (3D) Slicer,¹⁷ was used to annotate the landmarks. The radial styloid process (RSP), medial corner of the articular surface of the distal radius (MCDR), and most distal point of the articular surface of the ulnar head (DPUH) were identified on AP radiographs. The distal radial shaft center (DRSC-AP) and proximal radial shaft center (PRSC-AP) on AP radiographs were defined as the position at 3 and 5 cm below the proximal lunate articular surface, respectively. On LAT radiographs, the most distal point of the volar radius (DPVR), most distal point of the dorsal radius (DPDR), distal radial shaft center on the LAT view (DRSC-LAT), and proximal radial shaft center on the LAT view (PRSC-LAT) were identified using the same method. The longitudinal axis of the radius was determined by the line between the DRSC-AP and PRSC-AP on AP radiographs and between the DRSC-LAT and PRSC-LAT on LAT radiographs.

The RI was measured as the angle between 2 lines. The first line was perpendicular to the longitudinal axis of the radius. The other line was between the RSP and MCDR. The RL was defined as the distance between the line perpendicular to the longitudinal axis of the radius passing through the RSP and the line perpendicular to the longitudinal axis of the radius passing through the MCDR. The UV was defined as the distance between the line perpendicular to the longitudinal axis of the radius passing through the MCDR and the line perpendicular to the longitudinal axis of the radius passing through the DPUH. The VT was measured on LAT radiographs and defined as the angle between the line perpendicular to the longitudinal axis of the radius

and the line between the DPVR and DPDR (Fig. 2).

A wrist radiograph labeled by a hand surgeon (S1; HSJ), with 10 years of clinical experience, was used to train and validate the deep learning algorithm. Following training, the model predicted the location of a set of landmarks without annotation. To evaluate performance, S1 and another hand surgeon (S2; JSL), with 20 years of clinical experience, labeled the remaining images as the test set. The mean of the measurement parameters obtained by the 2 hand surgeons served as the reference standard.¹⁸ All manual annotations were performed using 3D Slicer.

Automatic Measurement of Parameters with the Deep Learning Hybrid Method

We developed a 2-step hybrid method to measure radiological parameters using anatomical landmarks by combining a deep learning model with a conventional CAD approach (Fig. 3). To detect anatomical landmarks on AP and LAT radiographs using the deep learning model, we used ResNet-50 with a convolutional block attention module as a feature extractor and adopted an attentive feature pyramid fusion (AFPF) module and a regression voting method to predict landmark locations. The model predicted landmark coordinates by voting for the highest probabilities in the heat map, which was regressed using the AFPF module. Unlike U-Net,¹⁹ optimized for

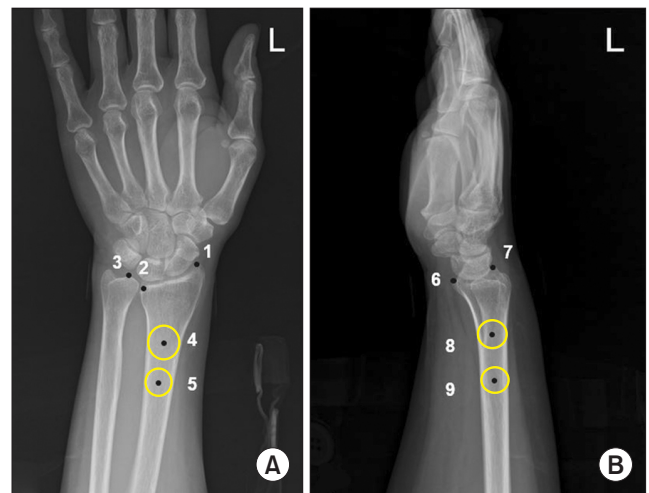


Fig. 2. Definitions of landmarks and parameters. Each landmark is assigned to a specific name. (A) 1: radial styloid process, 2: medial corner of the articular surface of the distal radius, 3: most distal point of the articular surface of the ulnar head, 4: distal radial shaft center on the anteroposterior view, and 5: proximal radial shaft center on the anteroposterior view. (B) 6: most distal point of the volar radius, 7: most distal point of the dorsal radius, 8: distal radial shaft center on the lateral view, and 9: proximal radial shaft center on the lateral view.

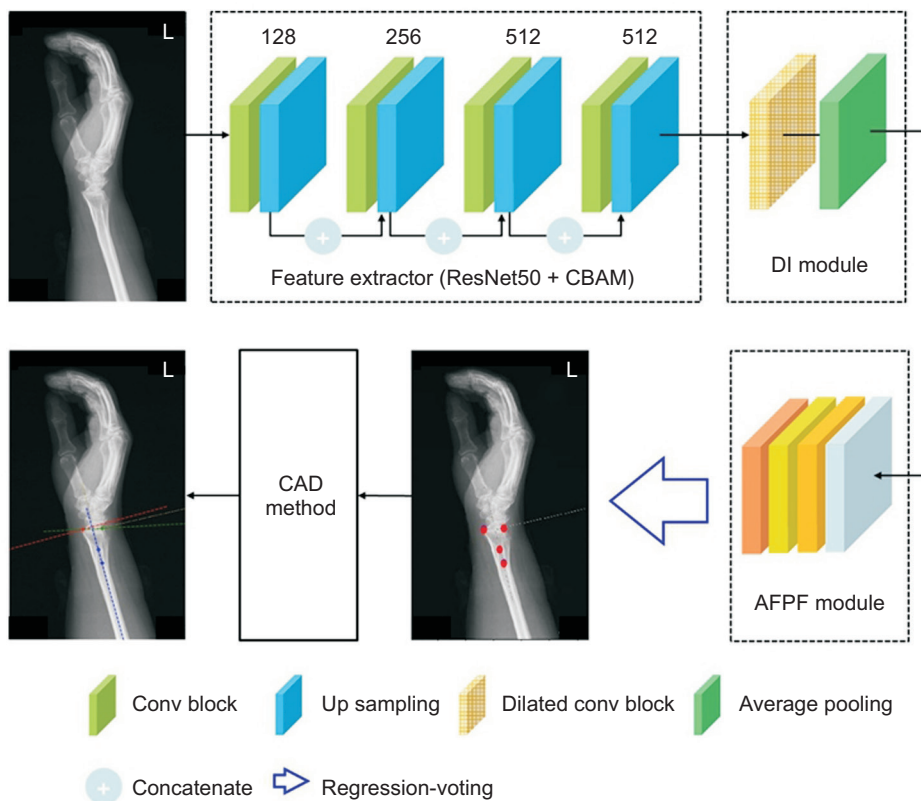


Fig. 3. Workflow of the model implementation. CBAM: convolutional block attention module, DI: dilated convolutional block, CAD: computer-aided diagnostic, AFPF: attentive feature pyramid fusion.

dense segmentation, our method focuses on sparse key point regression, benefiting from the multi-scale semantic attention and reduced spatial redundancy of ResNet-based backbones. To mitigate the risk of overfitting, we implemented a 5-fold cross-validation approach in which the dataset was randomly split into 5 subsets. Each fold was trained independently for 100 epochs to ensure a reliable performance assessment across the entire dataset. Hyperparameters were optimized using a validation split within each fold. In addition, early stopping was employed based on the validation loss with a patience of 5 epochs to prevent overfitting. The model was trained using a workstation equipped with an NVIDIA RTX A6000 GPU (48 GB VRAM), an AMD EPYC 7413 24-Core Processor, and 755 GB of RAM. In the second step, a CAD algorithm was developed to measure radiological parameters using the predicted landmarks according to the method described in the previous section. We then connected the output of the deep learning model to the input of the CAD algorithm to automatically measure radiological parameters. The model and CAD algorithm were implemented in Python using the PyTorch deep learning framework (version 2.0.0).

Reliability Evaluation of Manual Annotations

To evaluate the interobserver reliability of the measurement of radiological parameters, 2 hand surgeons (HSJ and JSL) independently evaluated the wrist radiographs. The percentages within the 1-, 2-, 3-, and 4-mm discrepancies between landmark thresholds and the mean radial error (MRE) were evaluated to determine the reliability of the annotations.

Evaluation of Landmark Detection Performance

We evaluated the developed model using a test set of 50 AP and LAT radiographs. The performance in detecting the 5 AP and 4 LAT anatomical landmarks was evaluated using the MRE and successful detection rate (SDR).²⁰⁾ The MRE was calculated as the mean of the absolute differences between the centers of the predicted landmarks and the reference standard. The SDR indicates the percentage of successfully detected landmarks in the ranges of 1-, 2-, 3-, and 4-mm differences with respect to the reference standard. S1 manually measured the reference standard coordinates.

Evaluation of Automatic Measurements of Radiological Parameters

To evaluate the radiological parameters, the automatically measured parameter values were compared with those

measured by the 2 surgeons. We used the mean values of each radiological parameter, as manually measured by the 2 hand surgeons, as reference standards. The paired *t*-test was used to compare the algorithm-measured parameters with those obtained manually. We compared the mean absolute errors (MAEs) between the parameter values obtained using the automatic method and the reference standards. We also calculated the intraclass correlation coefficient (ICC) and Pearson correlation coefficient (*r*) between parameter values. The ICC was used to evaluate consistency; ICC values less than 0.5 indicated poor reliability,

values between 0.5 and 0.75 indicated moderate reliability, and values greater than 0.75 indicated good reliability.²¹⁾ For the Pearson correlation coefficient, $0.6 \leq |r| \leq 0.79$ indicated a moderate correlation and $|r| \geq 0.8$ indicated a high correlation.²²⁾ Additionally, Bland-Altman plots were created to show the mean difference, standard deviation, and 95% limit of agreement between the values.

RESULTS

Reliability of the Manual Measurement of Landmarks by 2 Surgeons

The percentages of interobserver landmark discrepancies within 4 mm for the AP and LAT radiographs were 81.6% and 86.5 %, respectively (Table 1).

Model Generalization via 5-Fold Cross-Validation

To evaluate the generalization performance of the model, we performed 5-fold cross-validation, training each fold for 100 epochs (a total of 500 epochs). The learning curves for the AP and LAT models were plotted using training and validation losses, as illustrated in Fig. 4. For the AP model, the validation loss decreased sharply in the early

Table 1. Interobserver Reliabilities and MREs of Landmark Annotations on AP and LAT Radiographs

Image	MRE (mm), mean $\pm$ SD	Interobserver reliability (%)			
		1 mm	2 mm	3 mm	4 mm
AP	4.97 $\pm$ 19.41	22	57	73	82
LAT	2.8 $\pm$ 2.02	11	36	66	87

MRE: mean radial error, AP: anteroposterior, LAT: lateral, SD: standard deviation.

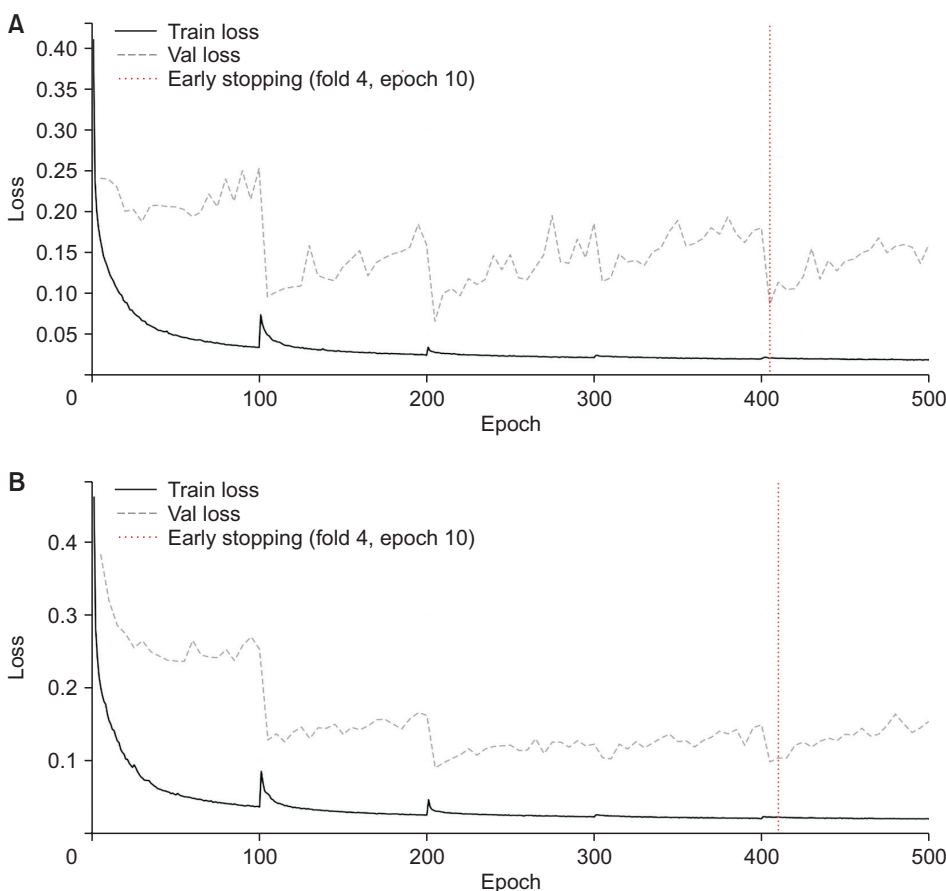


Fig. 4. Training and validation loss curves for the anteroposterior (A) and lateral (B) models over 500 epochs (5-fold cross-validation). The red dotted vertical lines indicate the early stopping points.

epochs and reached its lowest point at the fifth fold (epoch 5). Subsequent epochs often led to marginal or negative gains in the validation performance, highlighting early convergence and potential overfitting risks (Fig. 4A). For the LAT model, the training loss steadily decreased across all folds. However, the validation loss reached its minimum during the fifth fold at epoch 10, after which it exhibited notable fluctuations indicative of overfitting. To mitigate overfitting, we applied early stopping based on the validation MRE. These early stopping points were also marked on the learning curves (Fig. 4B). Notably, the AP model converged significantly faster across all folds, whereas the LAT model demonstrated more gradual im-

provements over the extended training, particularly in the first and second folds.

Performance of the Deep Learning Model for Landmark Detection

We evaluated the landmark detection performance of the developed model by comparing the predicted positions with the reference standard positions manually annotated by the surgeon. Table 2 shows the performance of the developed deep learning model in detecting anatomical landmarks in the test set. We achieved 98%, 99%, 100%, and 100% SDRs for measurements of 1, 2, 3, and 4 mm, respectively, on AP radiographs, and 97%, 99%, 100%, and 100% SDRs for measurements of 1, 2, 3, and 4 mm, respectively, on LAT radiographs.

Automatic Parameter Measurement Performance of Computer-Aided Diagnosis

The measured values from the 2 hand surgeons were compared with the model estimates for the 4 parameters (Table 3). In the correlation analysis, RI, VT, and UV demonstrated good or high reliability (ICC: 0.86, 0.95, and 0.98; $r = 0.86$, $r = 0.95$, and $r = 0.98$, respectively), while RL exhibited moderate reliability (ICC: 0.75, $r = 0.78$, respectively). In the subgroup analysis, the internal group

Table 2. MREs and SDRs of Landmark Annotations on AP and LAT Radiographs

Image	MRE (mm), mean $\pm$ SD	SDR (%)			
		1 mm	2 mm	3 mm	4 mm
AP	0.19 $\pm$ 0.21	98	99	100	100
LAT	0.27 $\pm$ 0.30	97	99	100	100

MRE: mean radial error, SDR: successful detection rate, SD: standard deviation, AP: anteroposterior, LAT: lateral.

Table 3. Comparison of Radiological Parameters Measured Manually and Automatically Using Deep Learning

Institution	Variable	Manual method	Automatic method	<i>p</i> -value	ICC	<i>r</i>	Mean difference	95% CI	MAE
Whole	RI (°)	26.68 $\pm$ 3.58	26.41 $\pm$ 4.01	0.188	0.86	0.86	0.27	−0.13 to 0.68	1.62 $\pm$ 1.26
	RL (mm)	13.64 $\pm$ 2.09	12.63 $\pm$ 2.72	0.000	0.75**	0.78**	1.01	0.67 to 1.35	1.56 $\pm$ 1.23
	VT (°)	14.56 $\pm$ 7.72	14.01 $\pm$ 8.02	0.029	0.95*	0.95*	0.55	0.06 to 1.04	1.88 $\pm$ 1.68
	UV (mm)	2.18 $\pm$ 2.88	2.09 $\pm$ 2.76	0.144	0.98	0.98	0.09	−0.03 to 0.20	0.43 $\pm$ 0.41
Internal	RI (°)	25.67 $\pm$ 4.30	25.35 $\pm$ 4.73	0.321	0.88	0.88	0.32	−0.32 to 0.96	1.76 $\pm$ 1.42
	RL (mm)	12.91 $\pm$ 2.29	11.39 $\pm$ 2.55	0.000	0.78**	0.79**	1.52	1.07 to 1.97	1.67 $\pm$ 1.43
	VT (°)	11.32 $\pm$ 9.13	10.94 $\pm$ 9.30	0.285	0.96	0.96	0.38	−0.33 to 1.09	2.02 $\pm$ 1.47
	UV (mm)	3.43 $\pm$ 3.18	3.27 $\pm$ 2.98	0.117	0.97	0.98	0.16	−0.04 to 0.35	0.55 $\pm$ 0.44
External	RI (°)	27.69 $\pm$ 2.32	27.46 $\pm$ 2.78	0.395	0.74	0.76	0.23	−0.30 to 0.74	1.48 $\pm$ 1.08
	RL (mm)	14.37 $\pm$ 1.60	13.86 $\pm$ 2.32	0.042	0.64*	0.68*	0.51	0.02 to 0.99	1.45 $\pm$ 1.00
	VT (°)	17.79 $\pm$ 3.95	17.08 $\pm$ 4.91	0.045	0.85*	0.87*	0.71	0.02 to 1.41	1.73 $\pm$ 1.86
	UV (mm)	0.92 $\pm$ 1.85	0.91 $\pm$ 1.91	0.790	0.97	0.97	0.02	−0.11 to 0.15	0.31 $\pm$ 0.34

Values are presented as mean $\pm$ standard deviation.

ICC: intraclass correlation coefficient, *r*: Pearson correlation coefficient, MAE: mean absolute error, RI: radial inclination, RL: radial length, VT: volar tilt, UV: ulnar variance.

* $p < 0.05$, ** $p < 0.001$.

demonstrated excellent agreement for UV (ICC: 0.97, $r = 0.98$) and VT (ICC: 0.96, $r = 0.96$). The RL showed consistent underestimation, with a statistically significant mean difference of 1.52 (95% CI, 1.07–1.97, $p < 0.001$). The VT, although highly correlated, exhibited greater variability across measurements, as reflected in the Bland-Altman plots (Fig. 5). In the external group, the automatic method maintained high performance for UV (ICC = 0.97, $r = 0.97$) and showed significant agreement for VT (ICC: 0.85, $r = 0.87$). Statistically significant but smaller biases were

observed for RL (mean difference: 0.51, $p = 0.042$) and VT (mean difference: 0.71, $p = 0.045$). Agreement for RI was high, and the differences were not statistically significant in either group ($p < 0.3$). To evaluate the measurement of radiological parameters, we calculated the MAE between the parameter values measured manually and automatically. In the subgroup analysis, the MAE remained the lowest for UV in both the internal (0.55 ± 0.44 mm) and external (0.31 ± 0.34 mm) groups, indicating consistent assessment accuracy for this parameter.

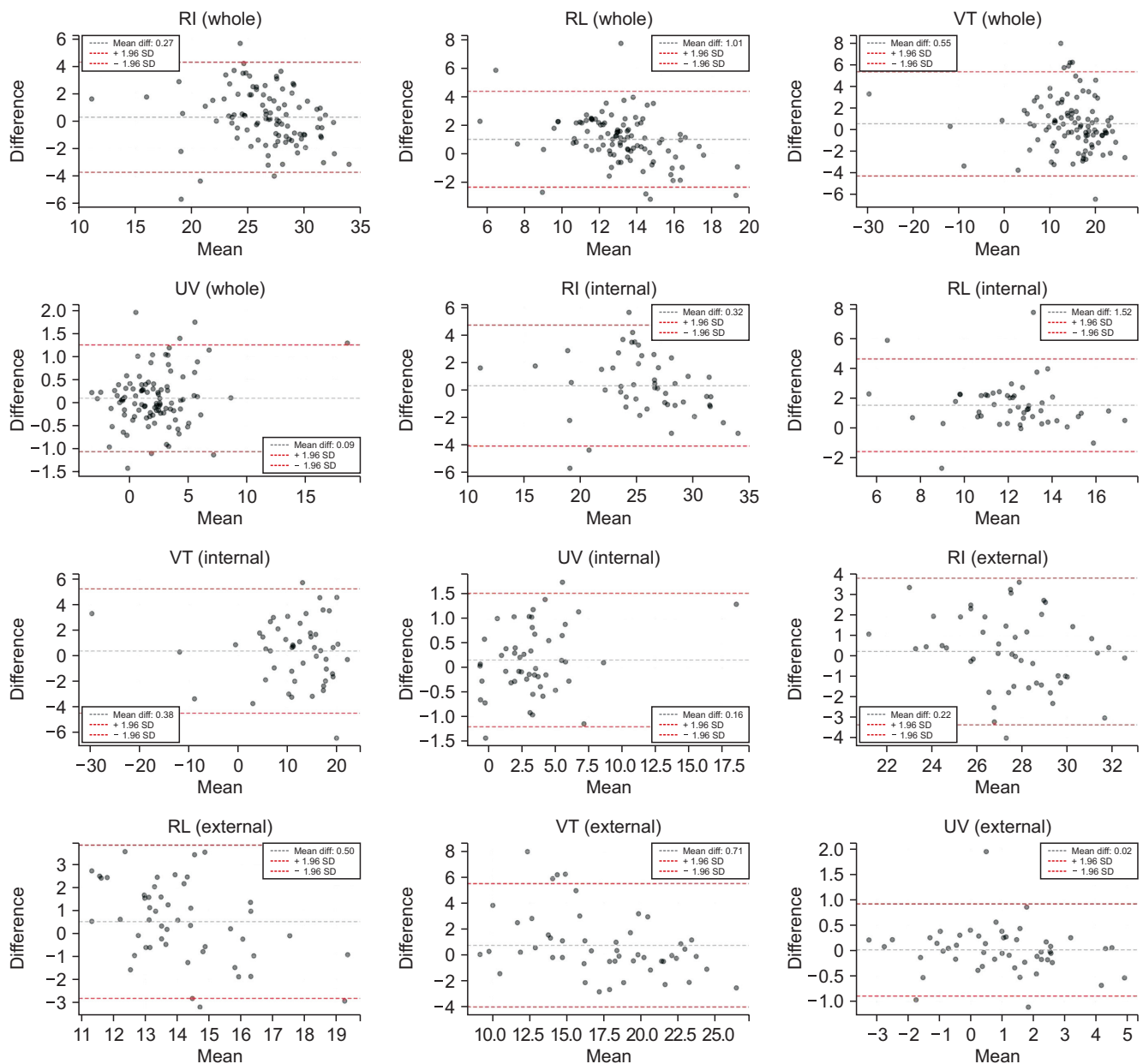


Fig. 5. Bland-Altman plots showing the mean differences between the model and reference standards. X-axis: the means of the model and reference standard, Y-axis: the difference between the model and reference standard. SD: standard deviation, Mean: mean of the difference (diff), RI: radial inclination, RL: radial length, VT: volar tilt, UV: ulnar variance.

Computational Efficiency

Training and inference were performed using an NVIDIA RTX A6000 GPU. The system required 17 seconds to process 100 wrist radiographs with an average inference time of approximately 0.17 seconds per image. This speed demonstrates the feasibility of near real-time integration into clinical workflows.

DISCUSSION

Wrist radiographs and a deep learning algorithm were used in this study to develop an automatic measurement model. Compared with the reference standard, the results highlighted that (1) the developed method achieved 94% and 96% SDRs within the clinically acceptable range of 2.0 mm and (2) the model showed good reliability for RI, VT, and UV, and moderate reliability for RL. Radiographic parameters play an important role in evaluating wrist pathologies. However, measuring these radiological parameters obtained by manually annotating AP and LAT radiographs is time-consuming and subject to observer variability. A previous study reported interobserver ICCs of 0.74 for VT, 0.38 for RI, and 0.44 for RL. Based on manual measurements of healed distal radius fractures, intraobserver ICCs were 0.71 for VT, 0.39 for RI, and 0.49 for RL.²³⁾ Another recent study reported that interobserver ICC values for radiographic measurements of the normal distal radius indicated good reliability for UV (> 0.8) and poor reliability for RI and VT (< 0.5).²⁴⁾ The ICC for RI measured by the algorithm and manual method was 0.86, with 0.95 for VT, 0.75 for RL, and 0.98 for UV. These results indicate that our deep learning model performed comparably to manual measurements performed by an experienced orthopedic surgeon.

Parameters dependent on distance, such as RL, showed ICCs of 0.75, which were lower than those of the angular parameters. The main reason for this finding was the difficulty in detecting landmarks located in the central area of the shaft, such as the DRSC and PRSC. For the DRSC and PRSC, unlike other landmarks, the central shaft area lacks anatomical features to accurately identify their locations, resulting in insufficient detection accuracy for correctly determining the longitudinal axis in terms of the angle for some patient radiographs. Approximately 95% of the landmarks in the test sets showed errors of less than 2 mm, indicating that only small angular errors occurred, which did not cause a significant problem in determining the lines that were nearly parallel to the longitudinal axis of the radius. Therefore, for the angular parameters, no significant differences were observed, and good reliability

was noted between the algorithm and manual method. In contrast, when calculating RL, a small angular error could lead to larger differences in the coordinates of the feet perpendicular to the longitudinal axis of the radius; consequently, the distance error between the perpendicular lines also increased. For patient images with high landmark detection errors of the DRSC and PRSC, the RL differences between the algorithm and manual methods were calculated as 7.76 mm and 5.88 mm for 2 cases, respectively. The occurrence of this outlier explains why the ICC of this parameter was lower than that of the angular parameters, although no significant differences were observed in the ICCs of the angular parameter between the algorithm and manual method. Although these 2 outlier cases occurred, they were not removed from the calculations as they were within the clinically acceptable RL range.

The clinical acceptability of measurement errors was also considered. Previous studies suggested that thresholds of < 2 mm for RL and $< 5^\circ$ for VT are clinically acceptable. Our MAE values of 1.56 mm (RL) and 1.88° (VT) were within this range, supporting their relevance in clinical practice. For example, Watson et al. reported that interobserver differences for RL and VT measurements of distal radius fractures commonly exceeded 2 mm and 5° , respectively, highlighting that errors within these ranges are acceptable for clinical purposes.⁵⁾ Similarly, Suojärvi et al.²⁴⁾ demonstrated considerable variability among physicians interpreting VT and UV on 2-dimensional (2D) radiographs, reinforcing the practical adequacy of our model's precision.

Subgroup analysis showed that the automated method performed reliably across institutions, but agreement varied by parameter and institution. The consistent underestimation of RL in the internal group raises concerns regarding the calibration or true variability across institutions. This discrepancy, although statistically significant, may reflect differences in image acquisition or labeling methods, highlighting the importance of local validation. In contrast, the external group had lower bias for RL and VT, suggesting greater generalizability or model compatibility with external data. Despite statistical significance, the small size of these differences may limit clinical impact. Overall, UV was calculated as the most robust parameter with excellent agreement across all subgroups and minimal bias. The MAE analysis further supports the robust performance of the model for UV, consistently showing the smallest absolute error across institutions. These results reinforce the potential utility of the automatic method for quantifying UV, while minimizing deviation from expert-derived measurements. The relatively high

MAE in VT, especially in the internal dataset, may be due to the inherent variability of the parameters or sensitivity to image artifacts such as motion. Similarly, the high MAE in RL shows that although correlations may remain high, absolute agreement is more variable, which is clinically significant when precision is important. The slightly improved MAE values in the external group indicated better model generalization or more homogeneous image features, highlighting the importance of multicenter validation in assessing model generalizability.

The key challenge in measuring the distal radius is determining the location of landmarks. In this study, most of the interobserver distances fell within 4 mm (82%–87%), demonstrating that the 2 hand surgeons were relatively reliable. However, the interobserver distances were within 1 mm, indicating poor performance (8%–28%). As mentioned above, manual labeling is subjective even for experienced surgeons, and the accuracy and consistency of labeling between surgeons can differ significantly, especially for high-accuracy requirements. Although intraobserver reliability was not quantified in this study, we acknowledge this as a limitation and plan to incorporate test-retest analysis in future work. In terms of landmark prediction, our deep learning model showed good performance, and the SDR at the 2–4 mm threshold ranged from 99% to 100%. However, some errors remain in the results. In particular, as previously mentioned, the center of the shaft could not be easily detected because of the normal bowing of the radius on AP radiographs and overlap with the ulnar shaft on LAT radiographs.²⁵⁾

Previous studies have adopted a method of calculating radiological parameters by initially extracting bone images, which are the subject of fracture diagnosis, using segmentation models, and then sequentially applying image processing and mathematical calculations to the extracted bone images.^{10,11)} Although this method demonstrates high accuracy for standardized data, it can show deviations in the image processing results for images with between-subject variations, such as patient data (e.g., bone shape and length). Typically, image processing and rule-based algorithms output results based on the rules set by the user. However, not all patient images satisfy these rules, which can potentially lead to significant measurement errors in radiological parameters. Moreover, the possibility of error accumulation exists during the implementation of these 3 steps, which means that segmentation models required a large number of samples to enhance reliability. Typically, a landmark detection model can directly identify anatomical landmarks from images without any segmentation or image processing steps.²⁶⁾ In this study,

we propose a novel automated method for the assessment of radiological parameters of the distal radius using a hybrid approach combining deep learning-based landmark detection and a CAD algorithm with mathematical calculations. We developed a deep learning model to identify anatomical landmarks to derive radiological parameters that can be used to assess fracture displacement by applying a CAD algorithm. Our method directly identified anatomical bone landmarks from patient radiographs using a landmark detection model and immediately applied the extracted landmark coordinates to derive radiological parameters, thereby eliminating potential errors that may occur in image processing. In addition, our deep learning model exhibited good performance in landmark detection. Even with a relatively small number of samples, this performance by our deep learning model led to satisfactory reliability of the values of radiologic parameters by simplifying the problem of calculating radiologic parameters compared to that in previous studies.

However, this study had several limitations. First, our automated assessment method for the distal radius requires further development to enhance its accuracy by increasing the amount of patient data required to train the model and reducing landmark detection errors. However, the advantage of our assessment is that it exhibits good performance with a small number of samples. Second, to correct the landmark detection, the training set involved a large group of normal wrist radiographs consisting mainly of clear landmarks. However, patients with distal radius measurements frequently present with displaced fractures or significant arthritis. The premise of manual labeling of the wrist is that hand surgeons can accurately identify the landmarks. To ensure confidence in manual labeling, cases with unrecognizable wrist landmarks, such as fractures, surgical history, and arthritic changes, were excluded. Despite adopting standardized manual measurement techniques, intraobserver and interobserver consistency in distal radius fractures was varied in previous studies.^{4,5)} Thus, to apply our algorithm to pathologic cases, a larger sample size is required; however, this is an extremely difficult task. We intend to include pathological cases in the future and believe that the effective method established in this study will be gradually applied in the clinic after the inclusion of cases with more lesion types. When developing an enhanced model for future studies, this study will be a useful guide to help reduce the sample size in the training set to meet clinically acceptable accuracy by simplifying the process for calculating radiologic parameters. Finally, our study did not consider measuring radiological parameters using 3D CT analysis. Although simple radiography

remains the primary imaging modality for measuring radiological parameters in clinical practice, a recent study reported that CT analysis is reliable for determining radiographic parameters of the normal distal radius, whereas physicians demonstrated considerable variability in interpreting radiographs.²⁴⁾ Previous studies have demonstrated that 3D CT provides higher precision in anatomical landmark detection than traditional 2D imaging modalities.^{27,28)} In particular, a recent review showed that deep learning methods can reliably automate 3D cephalometric landmarking using CT data, achieving accuracy superior to manual reference standards.²⁹⁾ Thus, future research using CT images is necessary to enable a more comprehensive analysis of the distal radius morphology.

In conclusion, a novel automated hybrid method can accurately identify landmarks on wrist radiographs and automatically generate radiological parameters of the distal radius. This saves time and human labor when creating datasets for training segmentation models and developing image processing algorithms, even with a relatively small number of samples.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGEMENTS

This research was supported by research grant from Biomedical Research Institute, Chung-Ang University Hospital (2021).

ORCID

Sang-Jeong Lee <https://orcid.org/0000-0002-1350-9434>
 Minji Kang <https://orcid.org/0000-0001-8800-2585>
 Jae-Sung Lee <https://orcid.org/0000-0002-3745-9048>
 Kyu-Tae Kang <https://orcid.org/0000-0002-8458-0596>
 Hyoung-Seok Jung <https://orcid.org/0000-0003-4756-6258>

REFERENCES

1. Mauck BM, Swigler CW. Evidence-based review of distal radius fractures. *Orthop Clin North Am*. 2018;49(2):211-22.
2. Egol KA, Walsh M, Romo-Cardoso S, Dorsky S, Paksima N. Distal radial fractures in the elderly: operative compared with nonoperative treatment. *J Bone Joint Surg Am*. 2010;92(9):1851-7.
3. Jung HS, Chun KJ, Kim JY, Lee JS. Necessity of acceptable radiologic alignment by preoperative closed reduction for unstable distal radius fractures treated with volar locking plates. *Eur J Trauma Emerg Surg*. 2021;47(6):1881-7.
4. Park MJ, Kim JP, Lee HI, Lim TK, Jung HS, Lee JS. Is a short arm cast appropriate for stable distal radius fractures in patients older than 55 years? A randomized prospective multicentre study. *J Hand Surg Eur Vol*. 2017;42(5):487-92.
5. Watson NJ, Asadollahi S, Parrish F, Ridgway J, Tran P, Keating JL. Reliability of radiographic measurements for acute distal radius fractures. *BMC Med Imaging*. 2016;16(1):44.
6. Yang S, Song ES, Lee ES, Kang SR, Yi WJ, Lee SP. Ceph-net: automatic detection of cephalometric landmarks on scanned lateral cephalograms from children and adolescents using an attention-based stacked regression network. *BMC Oral Health*. 2023;23(1):803.
7. Yang S, Lee SJ, Yoo JY, et al. V2-net: an attention-guided volumetric regression network for tooth landmark localization on CT images with metal artifacts. *Annu Int Conf IEEE Eng Med Biol Soc*. 2023;2023:1-5.
8. Lee JH, Kim YJ, Kim KG. Bone age estimation using deep learning and hand X-ray images. *Biomed Eng Lett*. 2020;10(3):323-31.
9. Oka K, Shiode R, Yoshii Y, Tanaka H, Iwahashi T, Murase T. Artificial intelligence to diagnosis distal radius fracture using biplane plain X-rays. *J Orthop Surg Res*. 2021;16(1):694.
10. Suna A, Davidson A, Weil Y, Joskowicz L. Automated computation of radiographic parameters of distal radial metaphyseal fractures in forearm x-rays. *Int J Comput Assist Radiol Surg*. 2023;18(12):2179-89.
11. Lee S, Kim KG, Kim YJ, et al. Automatic segmentation and radiologic measurement of distal radius fractures using deep learning. *Clin Orthop Surg*. 2024;16(1):113-24.
12. Chang HJ, Lee SJ, Yong TH, et al. Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis. *Sci Rep*. 2020;10(1):7531.
13. Jung HW, Hong H, Jung HJ, et al. Redisplacement of distal radius fracture after initial closed reduction: analysis of prognostic factors. *Clin Orthop Surg*. 2015;7(3):377-82.
14. Walenkamp MM, Aydin S, Mulders MA, Goslings JC, Schep NW. Predictors of unstable distal radius fractures: a systematic review and meta-analysis. *J Hand Surg Eur Vol*. 2016;

- 41(5):501-15.
15. LaMartina J, Jawa A, Stucken C, Merlin G, Tornetta P. Predicting alignment after closed reduction and casting of distal radius fractures. *J Hand Surg Am.* 2015;40(5):934-9.
 16. Jung HS, Jung HS, Baek SH, Lee JS. How many screws are needed for reliable stability of extra-articular nonosteoporotic distal radius fractures fixed with volar locking plates? *Clin Orthop Surg.* 2020;12(1):22-8.
 17. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging.* 2012;30(9):1323-41.
 18. Yang W, Ye Q, Ming S, et al. Feasibility of automatic measurements of hip joints based on pelvic radiography and a deep learning algorithm. *Eur J Radiol.* 2020;132:109303.
 19. Zunair H, Ben Hamza A. Sharp u-net: depthwise convolutional network for biomedical image segmentation. *Comput Biol Med.* 2021;136:104699.
 20. Wang CW, Huang CT, Hsieh MC, et al. Evaluation and comparison of anatomical landmark detection methods for cephalometric x-ray images: a grand challenge. *IEEE Trans Med Imaging.* 2015;34(9):1890-900.
 21. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155-63.
 22. Rusakov DA. A misadventure of the correlation coefficient. *Trends Neurosci.* 2023;46(2):94-6.
 23. Kreder HJ, Hanel DP, McKee M, Jupiter J, McGillivray G, Swiontkowski MF. X-ray film measurements for healed distal radius fractures. *J Hand Surg Am.* 1996;21(1):31-9.
 24. Suojärvi N, Lindfors N, Höglund T, Sippo R, Waris E. Radiographic measurements of the normal distal radius: reliability of computer-aided ct versus physicians' radiograph interpretation. *J Hand Surg Eur Vol.* 2021;46(2):176-83.
 25. Firl M, Wünsch L. Measurement of bowing of the radius. *J Bone Joint Surg Br.* 2004;86(7):1047-9.
 26. Song Y, Qiao X, Iwamoto Y, Chen Y-w. Automatic cephalometric landmark detection on x-ray images using a deep-learning method. *Appl Sci.* 2020;10(7):2547.
 27. Bajaj K, Rathee P, Jain P, Panwar VR. Comparison of the reliability of anatomic landmarks based on pa cephalometric radiographs and 3d ct scans in patients with facial asymmetry. *Int J Clin Pediatr Dent.* 2011;4(3):213-23.
 28. Serafin M, Baldini B, Cabitza F, et al. Accuracy of automated 3D cephalometric landmarks by deep learning algorithms: systematic review and meta-analysis. *Radiol Med.* 2023;128(5):544-55.
 29. Yun HS, Hyun CM, Baek SH, Lee SH, Seo JK. A semi-supervised learning approach for automated 3D cephalometric landmark identification using computed tomography. *PLoS One.* 2022;17(9):e0275114.