IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 20, 2018, accepted August 22, 2018, date of publication September 20, 2018, date of current version October 19, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2871502

Learning Radiologist's Step-by-Step Skill for Cervical Spinal Injury Examination: Line Drawing, Prevertebral Soft Tissue Thickness Measurement, and Swelling Detection

YOUNG HAN LEE[1], SEWON KIM[2], JIN-SUCK SUH[1], AND DOSIK HWANG[2], (Member, IEEE)

[1]Department of Radiology, Research Institute of Radiological Science, YUHS-KRIBB Medical Convergence Research Institute and Center for Clinical Imaging Data Science, College of Medicine, Yonsei University, Seoul 03722, South Korea

[2]School of Electrical and Electronic Engineering, College of Engineering, Yonsei University, Seoul 03722, South Korea

Corresponding author: Dosik Hwang (dosik.hwang@yonsei.ac.kr)

This work was supported in part by the National Research Foundation (NRF) Grant through the Korean Government, Ministry of Science, ICT & Future Planning (MSIP), under Grant 2015R1A2A1A05001887 and Grant 2018R1A2B6009076, and in part by the NRF Grant through the Korean Government (MSIP) under Grant 2016R1A2B4015016.

ABSTRACT Radiologists examine lateral view radiographs of the cervical spine to determine the presence of cervical spinal injury. In this paper, we demonstrate that an artificial intelligence neural network can learn the steps employed by a radiologist when examining these radiographs for possible injury. We deconstructed the decision-making strategy into three steps: line drawing, prevertebral soft tissue thickness (PSTT) measurement, and swelling detection. After training neural networks to be guided by the radiologist's intermediate labels, the networks successfully performed comparable line drawings to those of the radiologists, and subsequent PSTT measurement and swelling detection were successful. Quantitative comparison of PSTT measurements between our proposed method and radiologists showed a high correlation ($r = 0.8663$, $p < 0.05$, and intraclass correlation coefficient $= 0.9283$ at the C2 level; $r = 0.7720$, $p < 0.05$, and intraclass correlation coefficient $= 0.8667$ at the C6 level). Using the radiologist's diagnosis as the reference point, the sensitivity, specificity, and accuracy of swelling detection by our proposed method were 100%, 98.37%, and 98.48, respectively. We conclude that our neural networks successfully learned the sequence of skills used by radiologists when interpreting radiographs for injury of the cervical spine.

INDEX TERMS Artificial intelligence, machine learning, computer assisted radiographic image interpretation, vertebrae, cervical, radiography.

I. INTRODUCTION

Cervical spine injury is a common problem, and its severity ranges from minor ligamentous injury to severe spinal cord injury (SCI) [1]. Cervical lateral view radiographs are lateral projection X-ray images of the cervical spine that are routinely taken as the first-line imaging for patients with suspected cervical SCI in most clinical settings including an emergency environment [2]–[4]. Radiologists examine these radiographs to determine whether or not injury is present in the cervical spinal before recommending further scanning by computed tomography (CT). However, when CT is unavailable, such as in rural areas, these radiographs will be the only available imaging.

Correct and early diagnosis of cervical spine injuries is imperative because delayed or missed diagnoses increase both morbidity and mortality [5]. Measuring the prevertebral soft tissue thickness (PSTT) on cervical spine lateral view images is a simple and quick method for examining potential cervical injury [6], [7], with abnormal thickening being strongly associated with acute injury [7]. Although an initial trauma series is routinely obtained in the emergency setting for patients with trauma, the rate of misdiagnosis or delayed diagnosis based on cervical spine radiographs is as high as 44%–47% [8], [9]. In addition, time pressures in emergency departments is becoming an ever greater problem for emergency physicians and radiologists as case volumes increase. Therefore, an automated method for measuring critical parameters, such as the PSTT, and subsequent detection of the swelling in radiograph exams would be invaluable in the workup of patient with trauma.

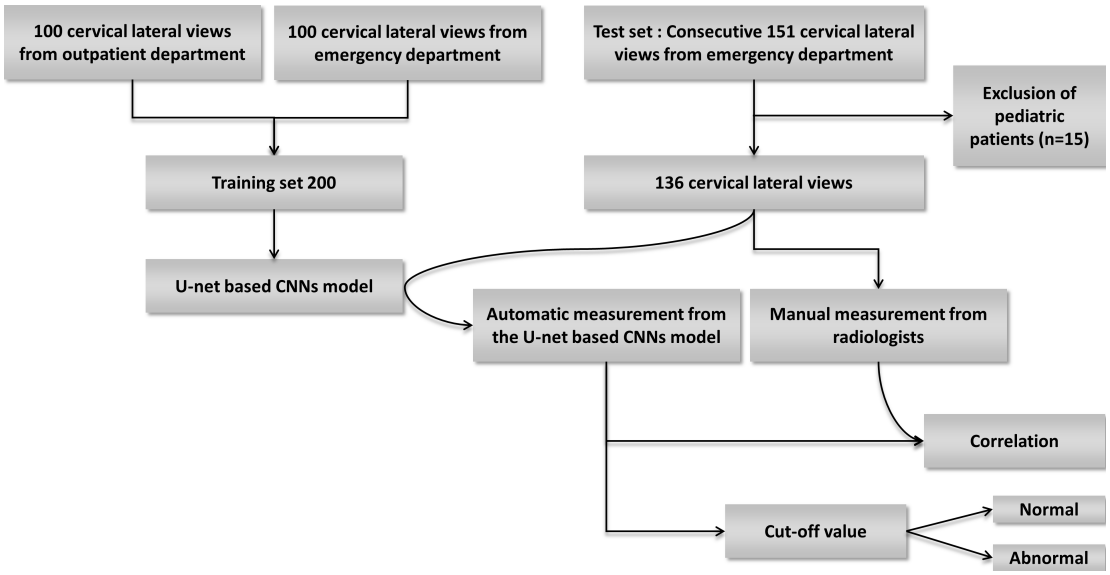

FIGURE 1. Flow chart of patient selection and grouping.

Deep learning is a form of machine learning wherein neural networks with multiple hidden layers are trained to perform specific tasks [10]. More recently, it is being actively investigated for use in a range of medical fields, including radiology [10], [11]. However, to date, deep learning has not been investigated as a potential aid when examining suspected cervical injury on cervical lateral view radiographs, which are important when diagnosing cervical SCI.

In this study, we aimed to develop an effective method of examining and interpreting radiographs for possible cervical injury, based on the approach of radiologists, but using deep learning. Due to the limited number of training data, we deconstructed the interpretation process used by a radiologist into three steps and trained our neural network to learn the Please submit all of the following in the list below and note that all files intended for publication need to be uploaded during this step, even if some files are unchanged from your previous submission. If all files are not submitted with final files, it will delay the publication of your article. intermediate step, which can be effectively achieved with small amount of datasets. In the subsequent evaluations, we demonstrated that PSTT values estimated by the trained neural networks are highly correlated with those manually measured by a radiologist. This produced high accuracy for the detection of swelling by the neural networks when using the radiologist's interpretation as the point of reference.

II. MATERIALS AND METHODS

A. STUDY POPULATION

For the training dataset, 200 lateral cervical radiographs were collected from the image database at our institution for the period between January and February 2016. Among these, 100 radiographs each were obtained from outpatient and emergency department settings. The training dataset included 186 digital radiograph (DR) images (Discovery XR656, GE Healthcare, Milwaukee, WI, n = 20; Definium 8000, GE Healthcare, n = 17; Innovision-SH, Shimadzu, Kyoto, Japan, n = 20; UD150B-30, Shimadzu, n = 127; GC85A, Samsung Electronics, Suwon, Korea, n = 1; DRS, Listem, Seoul, Korea, n = 1) and 14 computed radiography (CR) images (FCR5000, Fujifilm, Tokyo, Japan, n = 13; DirectView CR Systems, Carestream/Kodak, Rochester, NY, n = 1).

For the test dataset, we included 136 consecutive patients aged ≥16 years who underwent cervical lateral radiography in the emergency department between December 2016 and February 2017. Younger patients were excluded because lateral radiographs of pediatric patients differ from those of adult patients. All images were obtained using a DR imager (UD150B-30, Shimadzu, Kyoto, Japan, n = 103; MobileDart Evolution, Shimadzu, n = 32; Optima XR220, GE Healthcare, n = 1). The patient selection flow chart for the training and test datasets is shown in Fig. 1. The study protocol was reviewed and approved by the relevant institutional review board.

B. RADIOLOGIST DECISION-MAKING PROTOCOL

To determine whether patients have a cervical spinal injury on radiographs, the three steps illustrated in Fig. 2 are usually performed by radiologists: 1) they carefully draw both the prevertebral line and the anterior vertebral line on cervical lateral radiographs; 2) they measure the distances between these two lines at the C2 and C6 levels; and 3) they determine an abnormal PSTT (i.e., swelling) when this distance is higher than a predefined cutoff value of 7 mm at the C2 level.

C. INTERPRETATION SCHEME BASE ON DEEP LEARNING

Due to the limited number of training data in this study (only 200 images), we could not use end-to-end classification deep learning techniques, which would require tens of thousands of images for proper training, cannot be applied for this type of examination. Therefore, instead of using end-to-end classification techniques, we deconstructed the interpretation sequence used by radiologists into the three steps outlined in section II.B. The neural networks were then trained to learn the radiologist's intermediate interpretation step (i.e., line drawing), which could be achieved using the small dataset. Next, a PSTT-measuring algorithm was developed to calculate the distances between the prevertebral and the anterior vertebral lines. Finally, a simple thresholding technique with the established cutoff value of 7 mm was applied to determine if the PSTT was abnormal as radiologists did. In this way, we overcame the problem of a shortage of training datasets, and our deep learning-based interpretation networks successfully learned the strategy used by radiologists in decision-making.

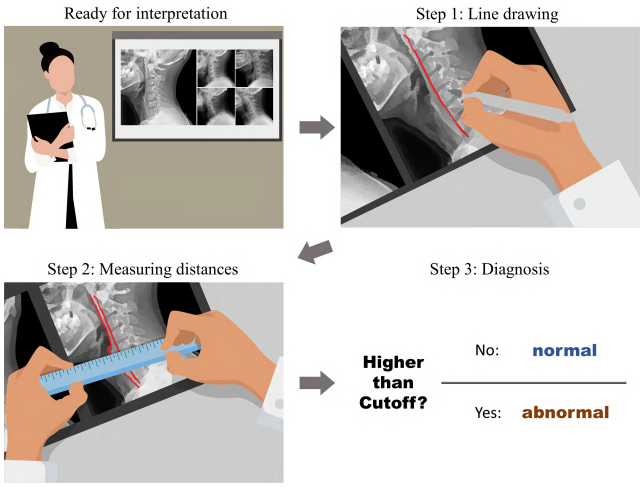

FIGURE 2. The three steps used by radiologists when interpreting lateral cervical spine images. Step 1) Draw prevertebral and anterior vertebral lines. Step 2) Measure the distances between the two lines at the C2 and C6 levels. Step 3) Diagnose any abnormality.

D. PREPARATION OF IMAGE DATA AND MANUAL LINE DRAWING

A local picture archiving and communication system (PACS) was queried for cervical lateral radiographs, which were then exported to a local computer, where they were anonymized and standardized. The obtained images were cropped, centered on the cervical spinal cords, and resized to a standard image matrix of 256 × 256 with an average resolution of 0.800 ± 0.062 (range, 0.639–1.020) mm. This allowed efficient training and testing. Based on pixel information in the DICOM (Digital Imaging and Communications in Medicine) header and resize ratio, pixel sizes were recorded proportionally.

For the training dataset, prevertebral and anterior vertebral lines were manually drawn on each image by a musculoskeletal radiologist with 10 years' experience in clinical diagnosis and scientific research. The manually drawn lines were converted to a binary mask as a label image for neural network training. A total of 200 image pairs were prepared (a radiograph as the input and a line mask as the label) for training.

E. THE NEURAL NETWORK ARCHITECTURE FOR LINE DRWAING

In a clinical setting, PSTT measurement involves manual drawing of prevertebral and anterior vertebral lines on cervical lateral radiographs (as shown in Fig. 2), and then measuring the distances between these two lines at the C2 and C6 levels. Although we trained our neural networks to learn this protocol, the direct learning of line drawings on radiographs is inefficient, especially for a small number of training image sets. Indeed, we failed to train our networks to learn the line drawing skill with 200 training image sets, so instead, we converted the line drawing problem into a segmentation problem.

The space between the two lines on the binary mask image was filled with 1's, forming a region-of-interest (ROI) that corresponded to the prevertebral soft tissues. Now, the input line label for the training became an ROI mask, which is commonly used in segmentation neural networks. U-net architecture [12] is known to perform well for such image segmentation tasks, so we constructed our neural networks to learn the segmentation task with the converted ROI labels based on this architecture. Because the U-net architecture can effectively learn both local and global features by a pooling process, our network also effectively learned both the local contrast features and the global shape features of the prevertebral soft tissue ROIs. Later, the boundary lines of the segmented ROIs were extracted to form the prevertebral and anterior vertebral lines. These extracted lines were given to the radiologist, who was asked to verify the efficacy of the proposed method.

The proposed training network for extracting prevertebral soft tissue ROIs is shown in Fig. 3. This network structure comprises the following: 33 convolutional (C) layers, a 3 × 3 convolutional kernel, three 2 × 2 max pooling (MP) layers, a stride of two for down-sampling, three 3 × 3 up-convolutional (UC) layers for up-sampling, and three concatenation (CC) layers. The learning process was divided into two parts: one that extracted the features while reducing the resolution of the feature maps, and another that segmented the objects while increasing the resolution again. In the first part, the resolution of the input image and its subsequent feature maps were reduced by half using three MP layers, each with four C layers, forward and backward. Each C layer was followed by a rectified linear unit (ReLU) as the activation function. The feature maps entering the second part returned to the original resolution through three UC layers. All UC layers had one CC layer and four prior C layers.

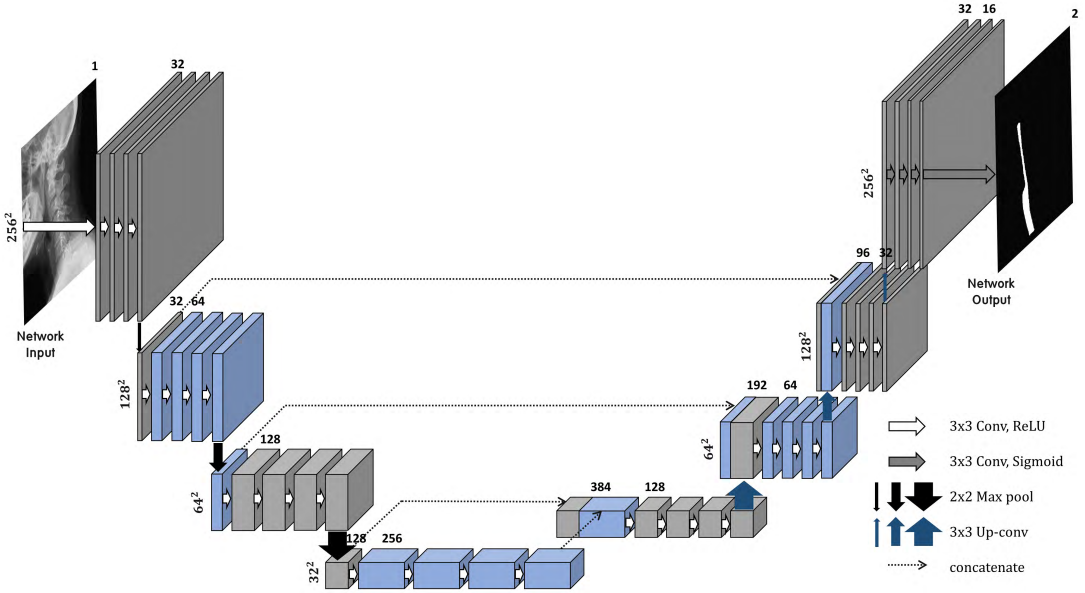

FIGURE 3. Neural network architecture for prevertebral soft tissue extraction (U-net based). The structure comprises two parts: the first involves a learning process during reduction of the feature map resolution; the second involves increasing the resolution back to the original size. Each box represents a group of data containing multichannel feature maps. The number of channels is noted at the top of each box, and the resolution of each feature map is denoted at the left edge of the box. The width of the box corresponds to the number of channels within the box.

Each C layer was also followed by an ReLU function. All CC layers combined the incoming data from the preceding lower layers and data with the same resolution from the first part (indicated by dotted arrows in Fig. 3). When all feature maps return to their original resolution, they pass through the remaining four C layers, followed by the activation function to generate the final output.

After prevertebral soft tissue ROIs were segmented using the trained neural networks, the prevertebral and anterior vertebral lines were derived using a Canny edge detection algorithm [13], and the resulting lines were later compared with those manually drawn by radiologists. In this way, we were able to evaluate the performance of the networks.

F. PSTT MEASUREMENTS AT THE C2 AND C6 LEVELS

Radiologists routinely measure PSTTs at the C2 and C6 levels when examining cervical spinal injury. This is because the C2 level represents the cranial portion of prevertebral soft tissue, characterized by a thickness of <7 mm, and the C6 level represents the caudal portion of prevertebral soft tissue, characterized by a more substantial thickness. In our proposed interpretation method for measuring PSTTs of the cranial and caudal portions of prevertebral soft tissues, all PSTTs from the top of the cranial portion to the bottom of the caudal portion were first calculated as illustrated in Fig. 4, before two representative PSTTs of the thin cranial and thick caudal portions were determined statistically. This mathematical process is explained by the following.

In the Cartesian coordinate, as shown in Fig. 4(a), let $y = al(x)$ and $y = pl(x)$ denote the anterior vertebral line (al) and prevertebral line (pl), respectively, and let $dl(i)$ denote the distance between two lines at the location $(x_i, al(x_i))$. Then, $dl(i)$ can be obtained as follows:

$$dl\,(\mathrm{i}) = |\mathbf{a} - \mathbf{p}|\,, \quad \mathbf{a} = \begin{bmatrix} \mathrm{x_i} \\ al\,(\mathrm{x_i}) \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} \mathrm{x_j} \\ pl\,(\mathrm{x_j}) \end{bmatrix}, \quad 1 \le \mathrm{i, j} \le n \tag{1}$$

where,

$$x_j = argmin\sqrt{(x - x_i)^2 + (pl(x) - dif\,(x, x_i)^2}, \quad dif\,(x, x_i) = -\left.\frac{dx}{dal\,(x)}\right|_{x=x_i}(x - x_i) + al\,(x_i) \tag{2}$$

Fig. 4(b) shows an example of the distribution of $dl(i)$ for the prevertebral soft tissue region from the top of the cranial portion to the bottom of the caudal portion. The midpoint, p, which divides the cranial and caudal portions of prevertebral soft tissues, can be determined by examining the variation of dl, as follows:

$$p = \underset{i}{argmax}\,(std\,(dl\,(i)))\,, \{i|\,3 \le i \le n-2\} \tag{3}$$

where, $std\,(f(\mathrm{i})) = \sqrt{\frac{1}{5}\sum_{\mathrm{t=i2}}^{\mathrm{i+2}}(f(\mathrm{t}) - \mu_{f(\mathrm{i})})}$ and $\mu_{f(\mathrm{i})} = \frac{1}{5}\,(f\,(\mathrm{i}-2) + f\,(\mathrm{i}-1) + f\,(\mathrm{i}) + f\,(\mathrm{i}+1) + f\,(\mathrm{i}+2))\,.$

PSTTs of the cranial and caudal portions were estimated using a statistical measure of central tendency for each cranial and caudal portion of dl distribution. For this, we used the median value to avoid any undesirable effects from outliers or skewed data. PSTTs of the cranial and caudal portions were estimated as $median\,(dl\,(\mathrm{i}))\,, \{\mathrm{i}|\,1 \le i < \mathrm{p}\}$ and

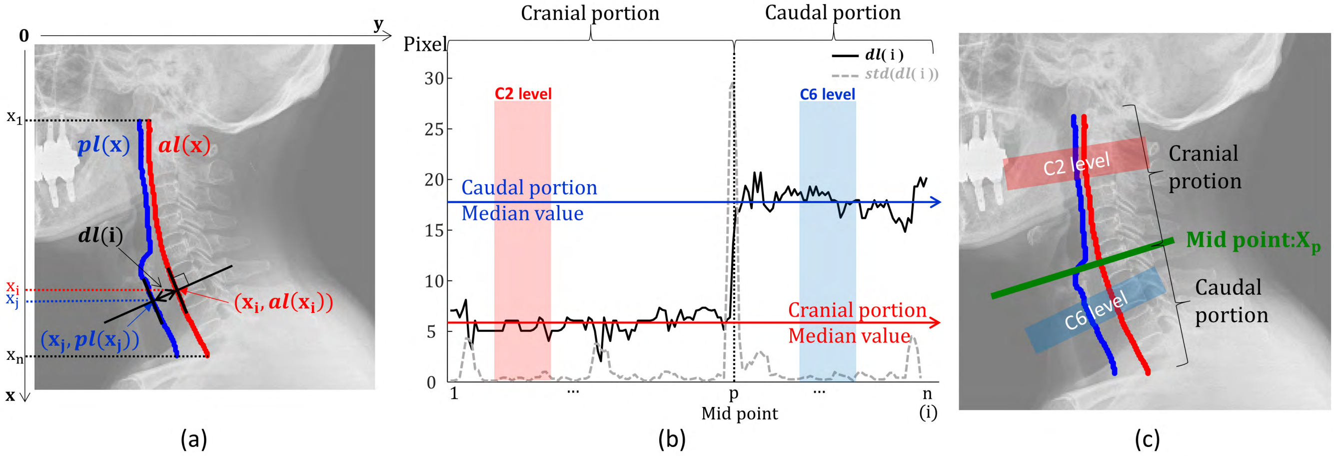

FIGURE 4. Definition of dl and the process of calculating PSTTs for the cranial and caudal portions of the prevertebral soft tissues. (a) The red line denotes the anterior vertebral line and the blue line denotes the prevertebral line. dl is the distance between al and pl. (b) An example of the distribution of dl for prevertebral soft tissues from the cranial to the caudal portion in pixel units (black solid line). The gray dotted line denotes std(dl(i)). The graphs can be separated into cranial and caudal portions by the midpoint (vertical dotted line). The horizontal arrows denote the median PSTT values calculated from the separated portions. The red and blue areas indicate the locations of C2 and C6. (c) The midpoint location, the cranial and caudal portions, and the C2 and C6 levels are illustrated on a representative radiograph.

$median\,(dl\,(i))\,, \{i|\,p \leq i \leq n\}$, respectively. To increase the accuracy when estimating PSTTs using median values, three maximum and three minimum values were excluded. Fig. 4(b) and 4(c) show examples of the midpoint, C2, and C6 in the distance graph and cervical lateral radiographs, respectively. In Fig. 4(b), the midpoint is the highest point of the $std\,(dl\,(i))$ graph (gray dash line), which is the point at which the $dl(i)$ graph (black line) rises steeply as shown by the green line in Fig. 4(c). Fig. 4(b) also shows that the median PSTT values of the caudal (blue horizontal line) and cranial (red horizontal line) portions represent the C2 (red band) and C6 (blue band) levels, respectively. Finally, abnormal swelling was defined when the PSTT exceeded the predefined cutoff value of 7 mm at the C2 level.

The design and training of the neural networks were implemented using Python 2.7 and the Google TensorFlow library (Google, Mountain View, CA, available at https://www.tensorflow.org) on a computer with the following specifications: an NVIDIA GeForce GTX 1080 GPU (NVIDIA Corp., Santa Clara, CA), a 3.60 GHz octa core CPU (Xeon, Intel, Santa Clara, CA), and 32 GB memory. The post-processing of the network output was performed in MATLAB (MathWorks, Natick, MA).

G. EVALUATION OF PSTT MEASUREMENT AND DETECTION ACCURACY

To validate the trained model, PSTT values that were measured using the trained neural networks were compared with those measured by the radiologist (reference values). Pearson's correlation coefficient (r) and intraclass correlation coefficient (ICC) tests were performed, including their 95% confidence intervals (CIs), to assess the correlation between values obtained by these two methods. To evaluate diagnostic performance, the ability of the trained networks to detect increased PSTT was evaluated in a test dataset of 132 cervical lateral radiographs of trauma patients from emergency department records. The cutoff value for normal and abnormal PSTTs was set to 7 mm at the C2 level, as recommended [7]. The sensitivity, specificity, accuracy, false-positive predictive, and false-negative predictive values of the neural network-based measurements to detect increased PSTTs were calculated with respect to the radiologist's decisions.

All statistical analyses were performed using the statistical software R package 3.1.2 (the R foundation for statistical computing, Vienna, Austria). A P value of <0.05 was considered to indicate statistically significant differences.

III. RESULTS

Training comprised 500 iterations, which required 7.25 h for 200 images, with training loss converging after 400 iterations. In the qualitative review by radiologist, the trained networks were confirmed to have successfully drawn the prevertebral and anterior vertebral lines at the C2 (n = 132/136; 97.06%) and C6 (n = 126/136; 92.65%) levels.

Fig. 5 shows examples of the automated drawing of the prevertebral and anterior vertebral lines in cervical lateral radiographs. As can be seen, both lines were successfully drawn from the top of the cranial portion to the bottom of the caudal portion of the prevertebral soft tissues, in various lordotic angles, reversed lordotic curves, or straightening of the cervical radiographs. However, there were four failures at the C2 level and six failures at the C6 level (Fig. 6). Close observation of these revealed that they were of poor quality, making it difficult and too confusing to draw clear lines (even for radiologists) due to marked deviations from standard imaging requirements.

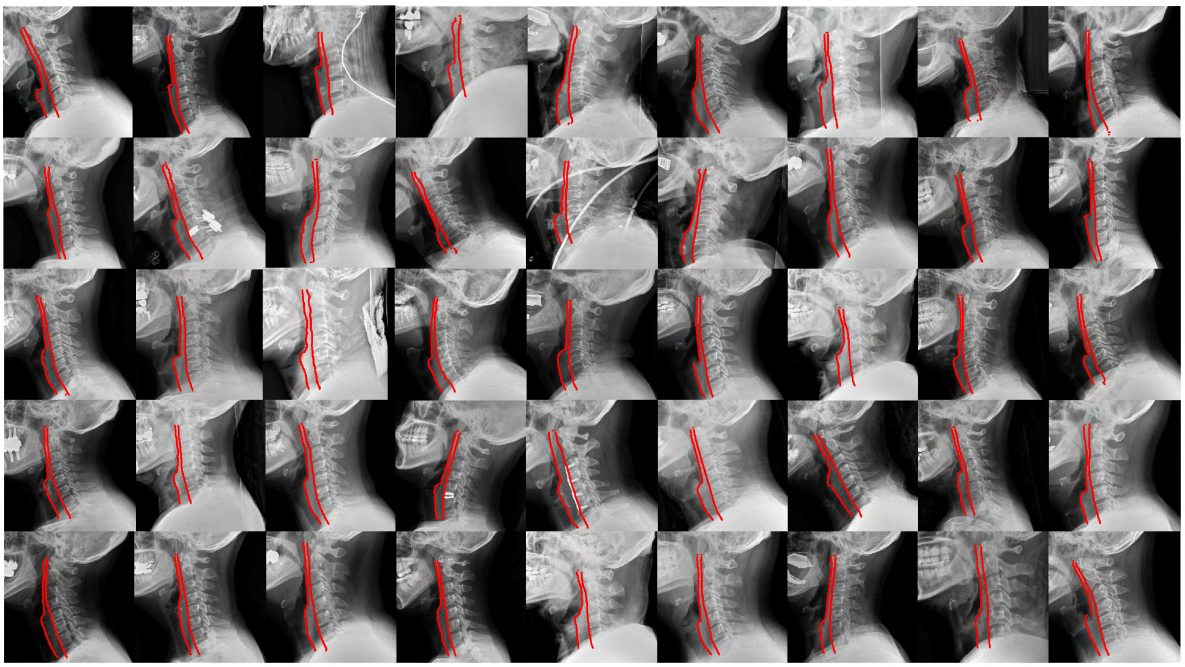

FIGURE 5. Various examples of the automated drawing of the prevertebral and anterior vertebral lines in cervical lateral radiographs. Both lines were successfully drawn from the top of the cranial portion to the bottom of the caudal portion of the prevertebral soft tissues at different lordotic angles, with reversed lordotic curves, or straightening of the cervical radiograph.

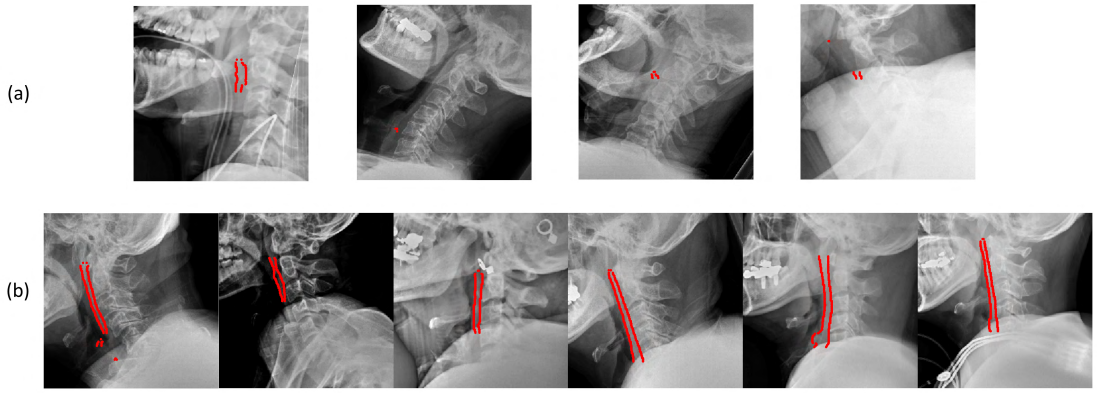

FIGURE 6. Failed cases of automatic line drawing at the C2 and C6 levels. (a) Four failed cases at the C2 level due to poor quality imaging, which prevented even the radiologists from drawing lines. Of note, the second radiograph was severely rotated beyond the coverage of the network structure. (b) Six failed cases at the C6 level were caused by excessive shoulder shadowing or poor image quality.

Therefore, 132 radiographs were evaluated to determine the correlation and diagnostic performance of our proposed method with respect to the decisions of the radiologist. The measurements obtained by the trained networks showed a high correlation with those manually obtained by the radiologist at the C2 level ($r = 0.8663$, 95% CI 0.8162–0.9034, $p < 0.05$; ICC = 0.9283, 95% CI, 0.8989–0.9492) and a fair correlation at the C6 level ($r = 0.7720$, 95% CI 0.6904–0.8343, $p < 0.05$; ICC = 0.8667, 95% CI, 0.8104–0.9062) for both normal and abnormal PSTTs (Fig. 7).

In terms of detecting the abnormal increased PSTTs (i.e., swelling) at the C2 level, among 132 radiographs, our deep learning-based interpretation method gave only two false positives (positive on automatic measurement and negative on manual measurement; n = 2/132; Table 1). Using the radiologist's decisions as the reference point, the sensitivity, specificity, positive predictive, negative predictive, and accuracy values of increased PSTT detection of the neural network-based method were 100% (95% CI, 66.37–100), 98.37% (95% CI, 94.25–99.80), 81.81% (95% CI, 53.23–94.68), 100% (95% CI, –100), and 98.48% (95% CI, 94.63–99.82), respectively.

IV. DISCUSSION

In this study, we demonstrated that deep learning could be used for radiographic measurement and interpretation of PSTT, with a relatively small training dataset (i.e., 200).

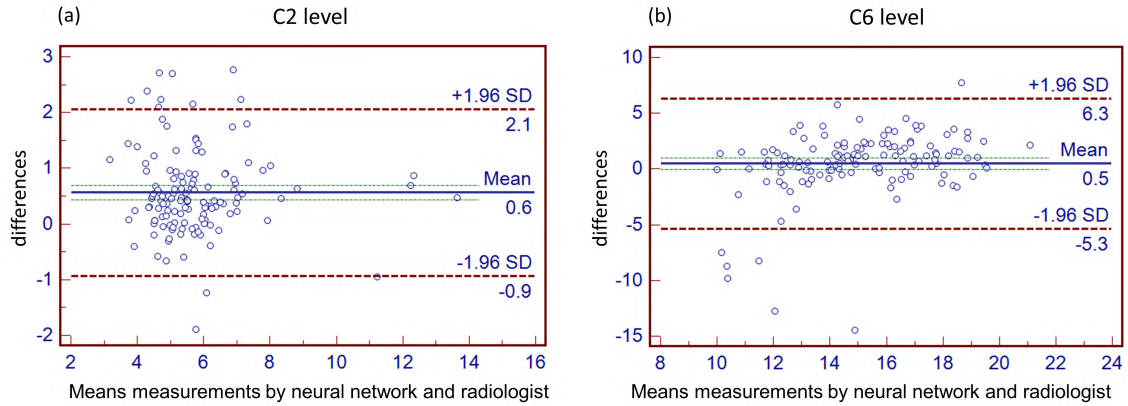

FIGURE 7. Bland–Altman plot comparing the means measurements by the neural network and by the radiologist. (a) C2 level: Pearson's correlation coefficient = 0.8663 and ICC = 0.9283. (b) C6 level: Pearson's correlation coefficient = 0.7720 and ICC = 0.8667.

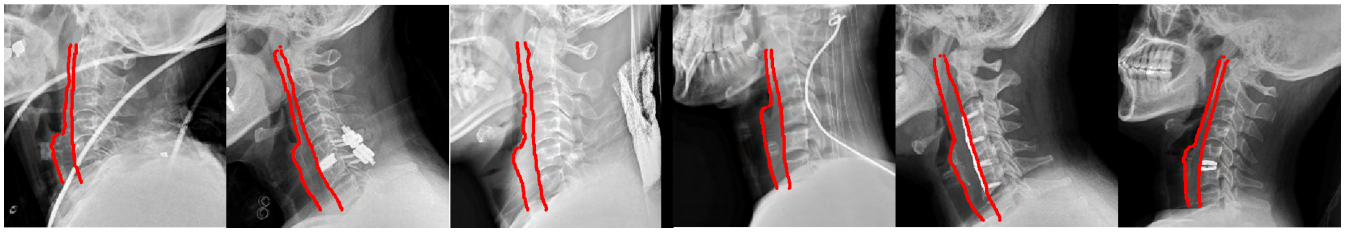

FIGURE 8. Example of successful line drawings, even when the radiographs contained metallic fixation devices, intravenous lines, or oxygen lines.

TABLE 1. Confusion matrix for the performance of the trained network compared with that of a radiologist.

		Radiologist		
		Normal	Abnormal	Total
Trained networks	Normal	121	0	121
	Abnormal	2	9	11
	Total	123	9	132

If a greater number of datasets are used for training, we anticipate that accuracy will be increased further.

Radiographs from eight different radiographic devices, including CR and DR, were used in our training and testing to ensure the relevance of our results to clinical settings. Although CR and DR provided different image textures and qualities, there was no significant difference in performance of our proposed method regarding the detection success rate. If we had restricted our study to a single image type or scanner with a comparable number of images, the training and testing performance may have been higher than reported in this study. This is because the dataset would have been more consistent, and therefore, the training would have been more efficient.

The training group also included patients from both outpatient and emergency settings. In a crowded emergency department with limited availability of radiologists, automated PSTT measurement and swelling detection can be invaluable in the workup of patients with cervical trauma. A warning system or an alarm function using this automated system could be implemented in a radiographic imager or in the image viewing system of PACS to ensure the fast and immediate detection of possible cervical spinal injury.

We used only minor data preparation steps, such as cropping and resizing of cervical radiograph images to make them uniform for training. No rotation, magnification, or panning was performed during this step. The acquired cervical radiographs were cropped such that the images ran from the top of the maxillary sinus to the clavicle. To reduce the training time and to make the size of input images uniform, we also resized all images to a 256 × 256 matrix. These steps could be automated by deep learning technology, provided the availability of enough data.

In clinical settings, it is often inevitable that artifacts such as wire lines or devices, e.g., will be overlaid onto radiographs. In our study, we encountered cases with intravenous or oxygen line shadows (n = 7), surgical fixation devices (n = 2), and endotracheal intubation (n = 2). Even with these artifacts, the neural networks performed well in the line drawings, except for one case that contained wires and endotracheal tubes that were aligned in the same direction as the prevertebral line (first image in Fig. 6(a)). This radiograph was even too confusing for the radiologist to draw lines. In all other cases with artifacts, automated line drawing was successful (Fig. 8). Another apparent issue was with the shoulder shadow affecting measurement at the C6 level.

In this study, we had six measurement failures for this reason, which caused invisible lower cervical vertebrae (Fig. 6(b)). In these cases, even the radiologist could not draw the lines properly.

PSTTs were successfully measured on machine-drawn lines at the C2 level, and these correlated almost perfectly with the manual measurements. When using an established cutoff value (7 mm), the detection rate of increased PSTTs was 100% (sensitivity of 100% and specificity of 98.37%). Correlation strengths between the automated and manual measurements were less robust at the C6 level, probably because of the patient's shoulder shadow. The size and extent of the shoulder shadow varied among patients because of differences in body sizes.

This study had several limitations. First, our deep learning-based interpretation method was trained to follow the radiologist's decisions on radiographic readings only. We did not intend to estimate further diagnostic information or confirmation that could be found in more advanced imaging exams, such as CT or magnetic resonance imaging. This is because radiologists cannot make such estimates without examining tomographic images.

Second, we used a relatively small dataset for training (200 radiographs). To ensure our proposed interpretation method is applicable to a wider selection of institutions, datasets from multiple sites should be used for further training and testing. Nevertheless, we are confident that the performance of our neural network system was excellent in the current datasets and that its efficacy would increase if the number of training datasets increased.

To overcome the shortage of training datasets, we decomposed the radiologist's decision-making strategy into three steps and trained our networks to learn each step, making the strategy possible even with the small number of datasets. If there were enough datasets for training, direct learning between the input images and the radiologists' decisions might have been possible; however, that would require far more datasets than could be obtained from even multiple hospitals. Furthermore, because our networks learned the intermediate steps in diagnosis, it could provide intermediate results for radiologists to verify, such as the prevertebral and anterior vertebral lines, the PSTT distribution along the cervical spine, and the final PSTT values (swelling or not) at the C2 level. This capability offers radiologists with opportunities to verify whether the conclusions are accurate based on intuitive steps. Many deep learning applications suffer from the critics of the so-called "black box" issue, which suggests that radiologists should not blindly accept the results of deep learning systems because they cannot check how the system reached its conclusions. However, by providing detail of the results at intermediate steps that are familiar to radiologists, which are therefore easily understood and verified, our system can provide greater confidence in the final decision.

V. CONCLUSION

In conclusion, we demonstrated that our neural network-based interpretation method successfully learned the radiologist's step-by-step skills when interpreting radiographs for potential injury of the cervical spinal. The correlation and accuracy of our method were high with respect to the radiologist's decisions, implying that they will be clinically relevant for determining PSTTs and swelling. These results support the use of deep learning techniques to assist radiologists in their work, showing that they can provide timely and highly accurate warnings that require only rapid human confirmation.

REFERENCES

[1] L. H. Sekhon and M. G. Fehlings, "Epidemiology, demographics, and pathophysiology of acute spinal cord injury," *Spine*, vol. 26, no. 24S, pp. S2–S12, Dec. 1964.

[2] B. D. Gerrelts, E. U. Petersen, J. Mabry, and S. R. Petersen, "Delayed diagnosis of cervical spine injuries," *J. Trauma*, vol. 31, no. 12, pp. 1622–1626, Dec. 1991.

[3] H. H. Bohlman, "Acute fractures and dislocations of the cervical spine. An analysis of three hundred hospitalized patients and review of the literature," *J. Bone Joint Surg.*, vol. 61, no. 8, pp. 1119–1142, Dec. 1979.

[4] J. A. Torretti *et al.*, "Cervical spine trauma," *Indian J. Orthopaedics*, vol. 41, no. 4, pp. 255–267, Dec. 2007, doi: 10.4103/0019-5413.36985.

[5] J. W. Davis, D. L. Phreaner, D. B. Hoyt, and R. C. Mackersie, "The etiology of missed cervical spine injuries," *J. Trauma*, vol. 34, no. 3, pp. 342–346, Mar. 1993.

[6] L. Penning, "Prevertebral hematoma in cervical spine injury: Incidence and etiologic significance," *Amer. J. Roentgenol.*, vol. 136, no. 3, pp. 553–561, Mar. 1981, doi: 10.2214/ajr.136.3.553.

[7] L. D. Matar and A. J. Doyle, "Prevertebral soft-tissue measurements in cervical spine injury," *Australas. Radiol.*, vol. 41, no. 3, pp. 229–237, Mar. 2008, doi: 10.1111/j.1440-1673.1997.tb00665.x.

[8] D. C. Reid, R. Henderson, L. Saboe, and J. D. Miller, "Etiology and clinical course of missed spine fractures," *J. Trauma*, vol. 27, no. 9, pp. 980–986, Sep. 1987.

[9] P. Platzer *et al.*, "Delayed or missed diagnosis of cervical spine injuries," *J. Trauma Acute Care Surg.*, vol. 61, no. 1, pp. 150–155, Jul. 2006, doi: 10.1097/01.ta.0000196673.58429.2a.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[11] D. Forsberg, E. Sjöblom, and J. L. Sunshine, "Detection and labeling of vertebrae in MR images using deep learning with clinical annotations as training data," *J. Digit. Imag.*, vol. 30, no. 4, pp. 406–412, Jan. 2017, doi: 10.1007/s10278-017-9945-x.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[13] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.

YOUNG HAN LEE received the M.D. and Ph.D. degrees from Yonsei University in 2001 and 2007, respectively. His Ph.D. thesis was on metal-induced CT perfusion images. He is currently with the Radiology and Musculoskeletal Fellowship with the Department of Radiology, Severance Hospital, Yonsei University. His research interests include musculoskeletal radiology and computer application in medical imaging.

SEWON KIM received the B.S. degree in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2016, where he is currently pursuing the Ph.D. degree in electrical and electronic engineering. His research interests include image processing, computer vision, machine learning, and medical imaging.

JIN-SUCK SUH received the M.D. degree from Yonsei University in 1979 and the Ph.D. degree from Ajou University with his thesis in the area of MR artifacts reduction in 1999. He has been an Avison Distinguished Professor with Yonsei University in 2011. He is currently with the Department of Radiology, Severance Hospital, Yonsei University. He is also the Director of the Imaging Development Projects of Medical Convergence Research Institute, Yonsei University. He is a Principal Investigator and a Korea's Frontier Research Scientist, named by the Korean Academy of Science and Technology, in 2011. His research interests include musculoskeletal radiology and molecular imaging. For his Ph.D. degree, he received the Korea's Highest Scientific Technical Award in 2007.

DOSIK HWANG received the B.S. and M.S. degrees in electrical engineering from Yonsei University, Seoul, South Korea, in 1997 and 1999, respectively, and the Ph.D. degree in bioengineering from The University of Utah, Salt Lake City, UT, USA, in 2006.

From 2006 to 2008, he was a Researcher with the University of Colorado Health Science Center, Denver, CO, USA. He is currently an Associate Professor with Yonsei University. His research interests include biomedical signal processing, MRI, single-photon emission computed tomography, ultrasound imaging, computed-tomography reconstruction, biometrics, and computer-aided diagnosis.