J Appl Biomed
DOI: 10.32725/jab.2020.013
journal homepage: http://jab.zsf.jcu.cz

Journal of Applied Biomedicine

Original research article

Automatic multi-class intertrochanteric femur fracture detection from CT images based on AO/OTA classification using faster R-CNN-BO method

Sun-Jung Yoon [1#], Tae Hyong Kim [2#], Su-Bin Joo [3], Seung Eel Oh [4*]

[1] *Jeonbuk National University, Research Institute of Clinical Medicine, Department of Orthopedic Surgery, Jeonju, South Korea*
[2] *Sungkyunkwan University, College of Biotechnology and Bioengineering, Suwon, South Korea*
[3] *Korea Institute of Machinery & Materials, Daegu Research Center for Medical Devices and Green Energy, Daegu, South Korea*
[4] *Korea Food Research Institute, Research Group of Consumer Safety, Wanju, South Korea*

Abstract
Intertrochanteric (IT) femur fractures are the most common fractures in elderly people, and they lead to significant morbidity, mortality, and reduced quality of life. The different types of fractures require a careful definition to ensure accurate surgical planning and reduce the operation time, healing time, and number of surgical failures. In this study, a deep learning-based automatic multi-class IT fracture detection model was developed using computed tomography (CT) images and based on the AO/OTA classification method. The original CT image was resized and rearranged according to the fracture location and an unsharp masking filter was applied. A multi-class classification of nine different types of IT fractures and no fracture was performed using the faster regional-convolutional neural network (R-CNN). Bayesian optimization was also implemented to determine the optimal hyperparameter values for the faster R-CNN algorithm. In our proposed model, IT fractures classified into two classes showed an average accuracy of 0.97 ± 0.02, which was 0.90 ± 0.02 when classified into ten classes. Additionally, the detected region of interest from our proposed model showed minimum root mean square error and intersection over union values of 16.34 ± 47.01 pixels and 0.87 ± 0.12, respectively. In the future, our proposed automatic multi-class IT femur fracture detection model could allow clinicians to identify the fracture region and diagnose different types of femur fractures faster and more accurately. This will increase the probability of correct surgical treatment and minimize postoperative complications.

Keywords: AO/OTA classification method; Computer-aided Diagnostic Detection; Deep learning; Intertrochanteric femur fracture; Optimization

Highlights:
- A system to automatically detect femur fracture types and their regions of interest was developed.
- Our proposed intertrochanteric femur fracture detection model using deep faster R-CNN-BO allows detection of ten different types of femur fractures.
- Similar systems can be applied to detect possible fracture cases to assist clinicians' interpretations.

Introduction

Intertrochanteric (IT) femur fractures are some of the most common fractures in the elderly and are usually due to falls. The incidence rate of IT fractures grows continuously, and it is expected to double by 2040 as the elderly population increases (Braun et al., 2018; Faisal and Nistane, 2016). As IT fractures often lead to significant morbidity, mortality, or reduced quality of life (Boone et al., 2014), it is important to provide an operative treatment that allows the restoration of efficient mobility while minimizing the risk of complications (Segal et al., 2018). IT femur fractures can be determined by medical imaging techniques such as X-rays or computed tomography (CT), and the type of operation or fixation device must be selected based on the characteristics of the fracture. Three primary factors should be considered when classifying the type of fractures: stability evaluation, reduction, and lateral or posteromedial wall integrity (Cho et al., 2018). In particular, stable fractures require treatment with sliding hip screw surgery, unlike unstable fractures, which show reverse obliquity fracture line characteristics and no lateral buttress intactness, making this treatment unsuitable for them. Instead, this type of fracture can be treated using an intramedullary nail to provide the buttress (Mears and Kates, 2015). Therefore, good IT fracture classification systems are required for accurate surgical planning.

*** Corresponding author:** Seung Eel Oh, Korea Food Research Institute 245, Research Group of Consumer Safety, Research Division of Strategic Food Technology, Nongsaengmyeong-ro, Iseo-myeon, Wanju-gun, Jeollabuk-do 55365, South Korea; e-mail: dr51@kfri.re.kr
[#] These authors made equal contributions to this work.
http://doi.org/10.32725/jab.2020.013
Submitted: 2019-08-19 • Accepted: 2020-08-26 • Prepublished online: 2020-09-22
J Appl Biomed 18/4: 97–105 • EISSN 1214-0287 • ISSN 1214-021X

In general, there are two main clinical classification methods that determine the types of IT fractures based on the physiological differences of the causes of fractures. Among several IT fracture classification systems, such as AO/OTA or Evans modified by Jensen (EVJE), the AO classification system is the most reproducible and reliable (Fung et al., 2007). In this system, A1 refers to a stable two-part fracture, A2 refers to an unstable fracture with three or more fragments, and A3 refers to the most unstable fracture with transverse or reverse obliquity fractures. However, even the AO classification system shows high inter- and intra-expert variability, especially when the fracture is examined by less experienced physicians (Fung et al., 2007; Jin et al., 2005). Additionally, human classification of these fractures is extremely time-consuming because of the multiple image readings (Wu et al., 2012). These problems can lead to surgical process failures, longer operation times, and a more difficult healing process. An automatic fracture classification system is important given the need for variability reduction, high classification accuracy, and quick detection. With the help of such a system, physicians can achieve better interpretations and reduce the interpretation times.

Several studies present classifications of femur fractures or detection of fracture lines using various methods, such as segmentation based on statistical shape models or traditional machine learning algorithms such as random forest (Erickson et al., 2017). The traditional automatic detection methods of previous studies are limited by the requirements of manual labeling of data, determination of the optimal threshold for segmentation, feature extraction process, etc. However, the application of deep learning methods such as convolutional neural networks (CNNs), which is now possible owing to the enhancement of hardware and algorithms, has led to improvements in terms of reduction of pre-processing steps with no manual feature extraction (Erickson et al., 2017; Ren et al., 2015; Shen et al., 2017). Bayram and Çakıroğlu (2016) developed automatic diaphyseal femur fracture classification methods for the first time; to improve their performance, a pre-processing method, called the support vector machine (SVM)-based sensitive noise remover, was applied along with feature extraction methods known as bone completeness indicators and fractured region mappings. Fractures were classified into nine classes based on the AO/OTA classification system using a multi-class SVM classifier with an accuracy of 89.87%. However, the disadvantage of this method is the use of X-ray imaging, which requires image enhancement and noise reduction, thus increasing the classification time cost.

To overcome the abovementioned limitation, deep learning has been actively applied to medical image analysis because of its ability to discover high correlation relationships within large medical data sets by automatically learning significant low- to high-level features (Kim and MacKinnon, 2018). As discussed above, in previous studies, machine learning allowed the determination of object locations in an image by drawing a bounding box or detecting single or multiple objects of different specific classes (Ker et al., 2017). From a clinical perspective, different tasks using deep learning algorithms are not crucial; it would be preferable to allow the detection and localization of objects in a single workflow (Voulodimos et al., 2018). Kazi et al. (2017) developed an automatic femur fracture detection model using an unsupervised spatial transformer network (USTN) with a CNN algorithm to classify six femur fracture types along with the fracture regions. However, their results showed that when the fracture types increased for classification, the object detection accuracy decreased from 84% to 82% with a smaller specificity value. Additionally, the localization performance was relatively poor, with a mean average precision (MAP) of 0.47. This may be owing to the application of an unsupervised ROI (region of interest) detection algorithm; the optimization of such a model is highly challenging.

Therefore, based on our knowledge, we have pioneered the development of an automatic femur fracture detection model that can classify a maximum of nine different fracture types and detect the femur location using CT imaging with a deep faster R-CNN model. The accuracy of the classification results and detection was evaluated. In the future, the developed model is expected to be applied by physicians and residents to determine the type of fracture and location in a single step. This will benefit the reproducibility of the AO/OTA classification method and increase the chances of accurate surgical treatment choices to allow early patient mobilization.

Materials and methods

Subjects

A total of 85 patients, 55 males and 30 females, were recruited from the Medical School of Chonbuk National University from 2016 to 2018. They were admitted to the hospital with IT fractures. The patients were excluded from the study if they had a history of surgery near the femur region. The present study received approval from the institutional review board (IRB No.CUH2018-01-228) of Chonbuk Hospital, and informed written consent was obtained from all patients.

Materials and fracture assessment

In this research, a total of 3343 CT images of size 512 × 512 × 3 pixels were collected from patients who had IT fractures. All CT images were obtained in digital imaging and communications in medicine (DICOM) format for the purpose of classification.

The IT-fracture CT images were reviewed independently by orthopedic surgeons, who were asked to classify and locate the regions of interest for the femur fracture for each image independently according to the AO/OTA classification (Sinno et al., 2010). The AO/OTA classification is primarily divided into three groups of fracture types, and then into nine classes of subgroups based on increasing fracture severity, as shown in Fig. 1.

Data pre-processing

The CT images collected from the hospital consisted only of femur fracture images; none of them were obtained directly from patients, as in Fig. 2A. Therefore, the dataset was increased by duplicating the original images with no fracture and fracture images. As shown in Fig. 2B, the original IT femur fracture CT image is resized to 224 × 224 pixels to reduce the computation time by considering general hardware specification and to fit the CNN architecture (Pranata et al., 2019; Urakawa et al., 2019; Toderici et al., 2017). To make the femur fracture line clear, an unsharp masking filter was applied, which allowed sharpening of the image; it functions by subtracting a blurred version of the original image from the original image to detect the edges. The contrast was increased along the edges using this mask. A radius value of two and contrast level of 20% was applied (Fig. 2B, left bottom corner).

Next, the resized CT image was marked based on the fracture location by dividing it into three regions (left side, middle, and right side). For the fracture within the left side region, the image was cropped from pixel 1 to 112 on the x-axis, and the rest of the image was left in black, as shown in the right upper corner of Fig. 2B. The image was cropped from pixel 56 to 168

Fig. 1. AO/OTA classification method with the types of femur fractures that were used for classification

on the x-axis for the middle region fractures and from pixel 112 to 224 on the x-axis for images showing fractures on the right-hand region. This cropped image was used to train the dataset in our proposed model.

For each image that was manually labeled with the class of the femur fracture along with the ROI of the fracture, we relabeled the IT femur fracture CT image dataset by dividing it into five groups based on the manual classification results: (i) two classes: no fracture and fracture; (ii) three classes: no fracture, A1.1 to A2.1, and A2.2 to A3.3; (iii) four classes: no fracture, A1, A2, and A3; (iv) seven classes: no fracture, each type of A1.1 to A1.3, each type of A2.1 to A2.3, and A3; (v) ten classes: no fracture, each type of A1.1 to A1.3, each type of A2.1 to A2.3, and each type of A3.1 to A3.3 (Crijns et al., 2018).

Our proposed femur fracture classification and ROI detection model

As mentioned above, the deep faster R-CNN deep learning approach was applied to determine the nine different fracture types and the femur location using a CT image and in a single process, unlike the previous general computer-aided diagnosis (CAD) techniques, as shown in Fig. 3A and B (Ren et al., 2015). It calculates the feature map of the input image and automatically obtains the features of each ROI using a region proposal network (RPN). The RPN uses the feature map of the last CNN convolution layer to determine the output ROI. Detecting the ROI requires an anchor, which is rectangular with various scale and aspect ratios, to generate different types of ROIs in the feature map. For each ROI, the object score and its bounding box are calculated and sent to the classification and bounding box regression layer. From these two layers, the class of the detected object is determined. Ren et al. (2015) describe the faster R-CNN in more detail.

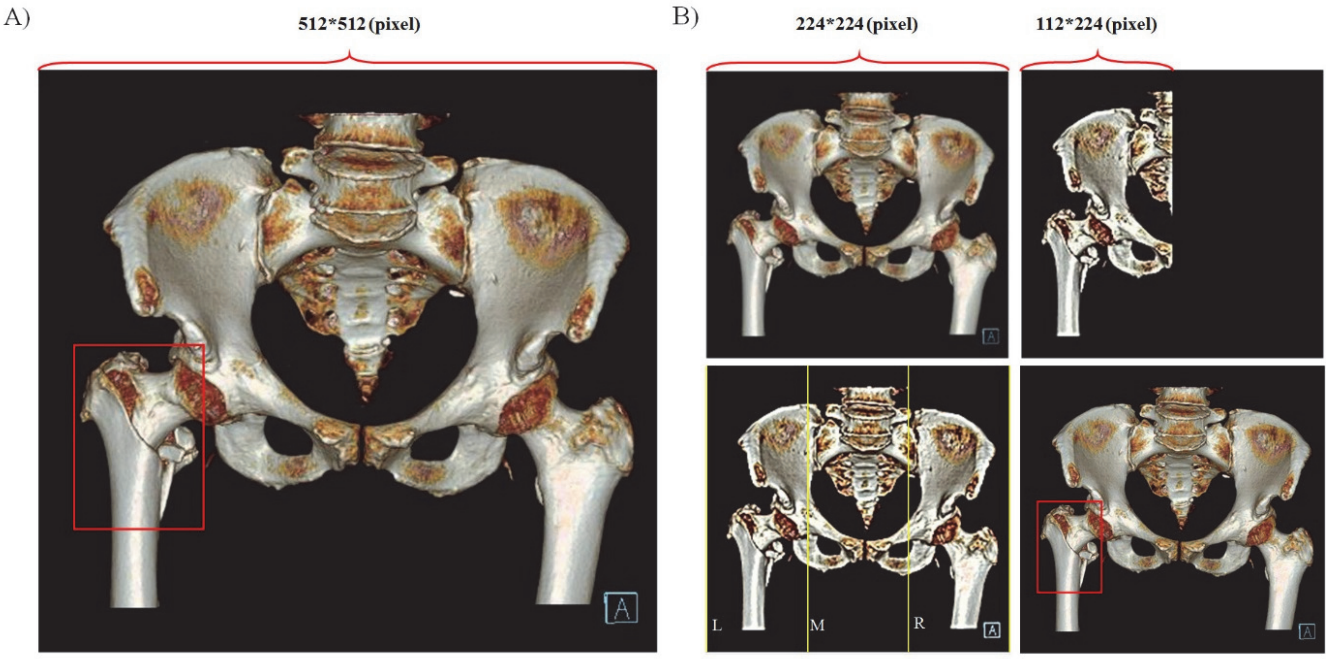

Fig. 2. Original CT image of a femur fracture. (**A**) The fracture line can be observed within the ROI (red box), which was manually annotated by experts. (**B**) The CT images were resized and cropped into left- and right-side images to collect normal IT images. The resized image was divided into three regions depending on the fracture site location, and the final reconstructed image is shown in the right bottom corner.

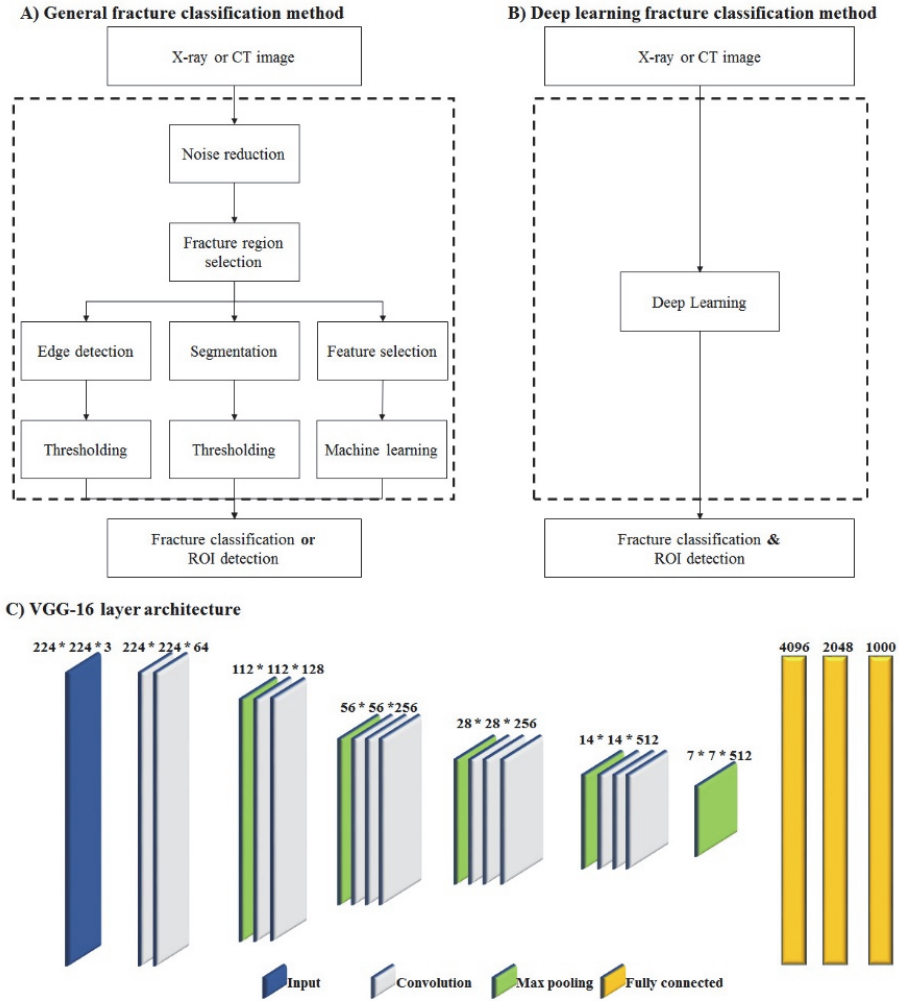

Fig. 3. Schematic diagrams of the previous general and the proposed femur fracture classification methods. (**A**) General CAD classification methods consisting of pre-processing, region selection, and annotation. (**B**) Deep learning algorithm showing simple CAD classification process. (**C**) VGG-16 architecture or layer for our proposed faster R-CNN-BO model.

To train the faster R-CNN we implemented the pretrained VGG-16 architecture from deep learning toolbox of MathWorks (R2018a, MathWorks, USA). In this research, the number of layers is same as the pretrained VGG-16 model. The changes of pretrained VGG16 model for our research is the layer of 39 to 41 such as the weight and bias value is replaced by optimized weight and bias values in fully connected layer or the number of classes for classification layer. The last max pooling layer is replaced by ROI max pooling layer with output size of 7 by 7. The more details such as size of stride or padding values of the VGG16 net architecture that was implemented are depicted in Fig. 3C (Zhang et al., 2016).

Previous studies have examined the influence of the classifier hyperparameters on performance; the particular hyperparameter values should be modified or determined to prevent a decrease in the performance of the model. As presented in Table 1, a total of ten hyperparameters are modified, and seven of them (which are shaded in gray in the table) are selected as optimized hyperparameters. The range of the selected hyperparameters of the faster R-CNN is also described.

Table 1. Hyperparameters and their details used for faster R-CNN

	Hyperparameter	Details
Fixed	Solver optimizer	Stochastic gradient descent with momentum
	Mini batch size	32
	Maximum epoch	40
Variable	Initial learning rate	0.0001–0.1
	Momentum	0.8–0.95
	L2 Regularization	0.0001–0.01
	Negative overlap range	0.1–0.4
	Positive overlap range	0.4–0.9
	Weights	0.0001–0.01
	Box pyramid scale	1.0–1.9

In this study, the sequential model-based optimization, also known as the Bayesian optimization (BO) technique, was applied to fine-tune the hyperparameters of the faster R-CNN classifier. The pseudocode for the BO is presented in Table 2. The maximum number of objective function evaluations was set to 30 (Shahriari et al., 2015).

Table 2. BO pseudocode for optimizing hyperparameters of faster R-CNN

Algorithm 1 Bayesian optimization	
1:	**for** t = 1,2, ... **do**
2:	select new X_{t+1} by optimizing acquisition function α $X_{t+1} = \arg\max_X \alpha(X; D_t)$
3:	query objective function to obtain y_{t+1}
4:	augment data $D_{t+1} = \{D_t, (X_{t+1}, y_{t+1})\}$
5:	update statistical model
6:	**end for**

Previous research implementations

To compare the performance of our automatic femur fracture model with that of other techniques, we implemented the method reported in the previous research by Kazi et al. (2017). We used our femur fracture CT image rather than the X-ray images of size 2500 × 2048 pixels that Kim and MacKinnon (2018) applied to classify femur fractures into a maximum of six subgroups (A1 to B3). Histogram normalization was applied during image pre-processing. In this research, we used three different approaches, the lower bound model (LBM), upper bound model (UPM), and USTN. MatConvNet was used to train the network with a learning rate from 10^{-5} to 10^{-4} for a maximum of 80 epoch, momentum with 0.9, and batch size of 10; backpropagation and stochastic gradient descent were used (Vedaldi and Lenc, 2015).

Statistical methods

To evaluate the performance of our developed classifiers, 5-fold cross validation was applied. The total number of used datasets was 3343. When using the 5-fold cross validation technique, the datasets were randomly divided into 2675 (80% of a total of 3343) for training and 668 (the remaining 20% of a total of 3343) for testing. From the result, we sorted out the validation group, which showed median classification accuracy, and carried out the process mentioned above five times to adjust for possible deviations in the results. The average classification accuracy, sensitivity (the number of true positives (TPs) over the number of TPs plus the number of false negatives (FNs)), and positive predictive value (PPV) (the number of TPs over the number of TPs plus the false positives (FPs)) were calculated from a confusion matrix. The intersection over union (IoU), which represents how well the area of the ground truth bounding box (A_{gt}) and that of the predicted bound box (A_p) overlap, was subsequently calculated.

$$\text{IoU between Ap and Agt} = \frac{Ap \cap Agt}{Ap \cup Agt}$$

Additionally, the root mean square error (RMSE) for the detected and original ROIs of diagonal length was calculated to determine the difference, where y_i is the true value and p_i is the predicted value.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2}$$

The obtained results were compared and analyzed through ANOVA using Tukey's post hoc analysis to verify the differences. The significance level was set to $p < 0.05$, and all statistical analyses were conducted using PASW Statistics 18 (v. 18, SPSS Inc, Chicago, IL, USA).

All the data processing was performed using MATLAB programs (R2018a, MathWorks, USA). The networks were trained and tested on a computer system running Windows with an Intel(R) Core(TM) i7-5930K @ 3.50 GHz processor and an 8 GB NVIDIA GeForce GTX 1080 graphics card.

Results

Classification accuracy of femur fracture from CT images using faster R-CNN-BO vs previous research method

In this research, we classified nine different types of femur fractures based on the AO/OTA classification method along with no femur fracture. Table 3 shows that, as compared to the accuracy when the number of femur fracture classification types increases, binary classification (no fracture vs fracture) shows the highest accuracy in our proposed model and in the other three methods.

As shown in Fig. 4, our proposed model had a significantly higher average classification accuracy of 0.97 ± 0.02 and 0.90 ± 0.02 when dividing the femur fractures into two and ten classes, respectively, as compared to 0.89 ± 0.02 and 0.78 ± 0.01 in the USTN model ($p < 0.01$), respectively.

Sensitivity and positive predictive values of proposed model when classifying different types of femur fractures

As presented in Table 4, the sensitivity and PPV of our proposed model were described using a confusion matrix. The overall average sensitivity was higher than the PPV for all five different classification results, indicating that the number of FPs was higher. The sensitivity and PPV for the two-class classification results ranged between 0.99 and 0.94. For the ten-class classification, the lowest sensitivity was 0.75, for the B2.2 fracture type, and the lowest PPV was 0.65, for the A1.1 fracture type.

The result of ROI localization for fracture detection

Our proposed femur fraction ROI detection model was evaluated by measuring IoU, as presented in Table 5. The average IoU of the classification of the femur fractures into two and ten classes showed minimum values of 0.88 ± 0.13 and 0.87 ± 0.12, respectively. Additionally, the diagonal length RMSE of the detected and ground truth of the ROI was calculated, and the three- and ten-class classification resulted in the lowest average RMSE value of 16.34 ± 47.01 and the highest average RMSE value of 31.48 ± 54.83, respectively. As shown in Fig. 5A, the higher IoU was achieved when there was sufficient overlap between the detected and ground truth of the fracture region.

Table 3. Comparison of classification accuracy for different types of femur fracture detection between previous studies and our proposed model

	LBM		UBM		USTN		Our proposed model
	(Ref 19)	(rep)	(Ref 19)	(rep)	(Ref 19)	(rep)	
2 class	0.81	0.90 ± 0.01	0.88	0.92 ± 0.01	0.84	0.89 ± 0.02	0.97 ± 0.02 (0.01)
3 class	0.83	0.89 ± 0.01	0.86	0.90 ± 0.01	0.83	0.85 ± 0.01	0.95 ± 0.02 (0.01)
4 class	–	0.85 ± 0.01	–	0.88 ± 0.00	–	0.83 ± 0.01	0.94 ± 0.01 (0.01)
7 class	0.86	0.83 ± 0.02	0.89	0.85 ± 0.01	0.82	0.81 ± 0.02	0.92 ± 0.01 (0.01)
10 class	–	0.80 ± 0.01	–	0.82 ± 0.00	–	0.78 ± 0.01	0.90 ± 0.02 (0.01)

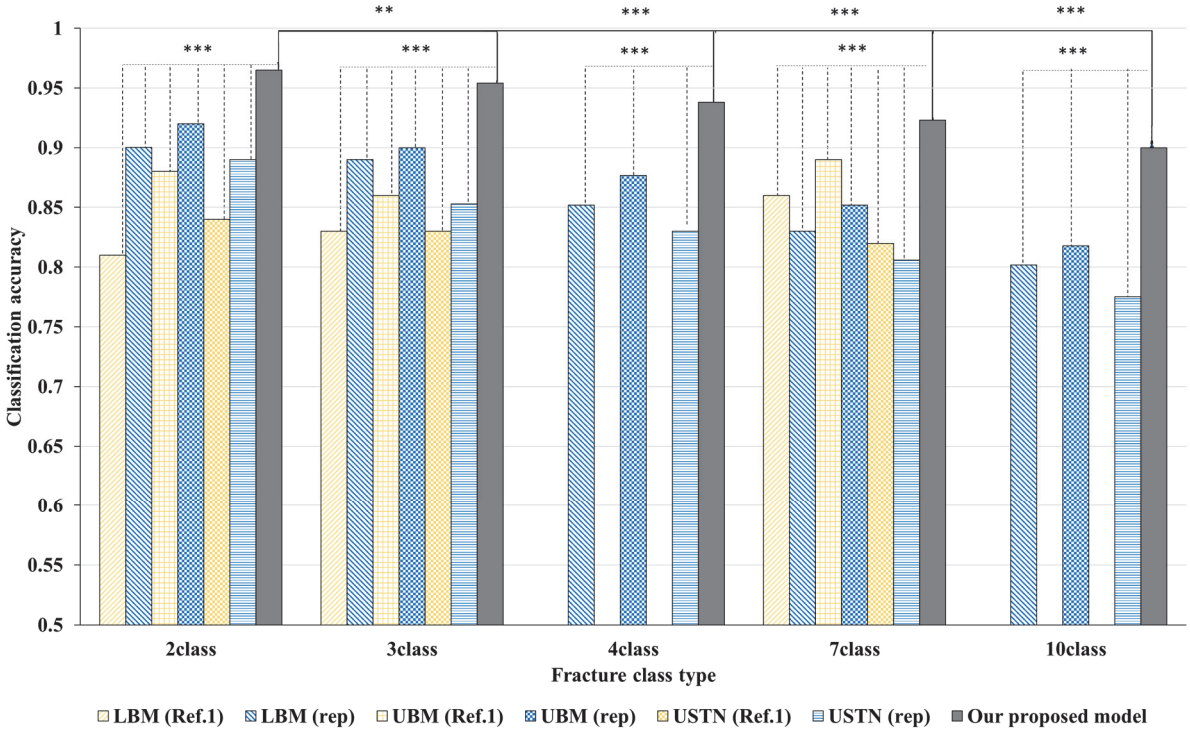

Fig. 4. Comparison of multi-class classification accuracy of femur fracture between previous research and our proposed model. LBM, UBM, and USTN represent the lower boundary model, upper boundary model, and unsupervised spatial transformer network, respectively.

Discussion

This study aimed to develop an automatic multi-class IT femur fracture detection model using deep learning to classify the type and location of a fracture in a single flow. Based on the annotated expert data, we modified the CT fracture images into three regions to train data for the deep faster R-CNN model. It could detect nine different types of femur fracture classes and no-fracture regions with relatively high accuracy and low RMSE and IoU.

Our proposed model showed slight decrease in accuracy, from 97% during binary classification to 90% during classification into ten subgroups. However, the previous research performed by Kazi et al. (2017) showed increased accuracy when the number of classes increased in the LBM model. This result may be due to the fact that the number of datasets was normalized to be identical for each class, indicating a higher number of training datasets, such as in the A2 class. Additionally, they applied the USTN model, which is a CNN containing one or several spatial transformer modules. This allowed the network to be spatially invariant for input images, unlike in the max pooling layer used in CNN. This spatial transformer network allowed the determination of an object's class and location in an unsupervised manner (Jaderberg et al., 2015). The significant PPV decrease was also determined based on the number of classes to classify the average PPV values from 96% to 79%. Additionally, the results showed that the PPV values for each of the five different models of our proposed model were lower than the sensitivity values. This could indicate that more FPs exist than for every TP than FNs.

Some previous studies have utilized CAD with medical imaging to detect or classify diseases such as fractures (Kim and MacKinnon, 2018; Yates et al., 2018). Most of them used X-ray images instead of CT images as an input data. X-ray images are single plane, mainly anterior/posterior or medial/lateral views; however, the complexity of bone morphology, especially in the posterolateral area and fracture lines, is obstructed

Table 4. Confusion matrix for the results of classification of the IT femur fracture into two to ten classes. The green and orange in the table represent maximum and minimum values for positive predictive values and sensitivity, respectively

		Positive predictive values										
Sensitivity	2 class	369.4	4.9	**0.99**								
		16.3	278.2	**0.94**								
		0.96	**0.98**									
	3 class	372.6	2.4	1.2	**0.99**							
		11.6	159.4	2.6	0.92							
		5.2	8.1	105.7	**0.89**							
		0.96	**0.94**	**0.97**								
	4 class	371.2	2.1	0.9	0.4	**0.99**						
		12.0	127.5	2.6	0.2	0.90						
		7.8	7.8	84.6	0.6	**0.84**						
		4.1	1.0	0.5	45.6	0.89						
		0.94	**0.92**	0.95	**0.97**							
	7 class	422.4	0.3	2.0	0.0	0.2	0.4	0.6	**0.99**			
		5.7	22.8	1.8	0.0	0.0	0.6	0.0	0.74			
		9.1	0.0	79.2	0.0	0.6	0.0	0.0	0.89			
		4.2	0.0	0.4	16.8	0.0	0.6	0.0	0.76			
		5.1	0.2	1.4	0.2	26.6	0.0	0.0	0.79			
		5.7	0.0	0.2	0.0	0.6	35.6	0.4	0.84			
		4.9	0.0	0.0	1.0	0.6	1.2	17.8	**0.70**			
		0.92	**0.98**	0.93	0.93	0.93	0.93	0.95				
	10 class	362.4	1.3	4.0	1.0	0.5	0.4	0.5	0.7	0.5	0.2	**0.98**
		5.2	20.3	4.2	0.0	0.6	0.4	0.0	0.0	0.4	0.0	**0.65**
		13.6	1.2	70.4	0.6	1.2	0.0	0.2	0.0	0.0	0.4	0.80
		0.2	0.2	0.4	18.6	1.2	1.8	0.4	0.2	0.1	0.2	0.80
		1.3	1.2	2.8	0.6	27.5	0.4	0.2	0.0	0.0	0.2	0.80
		0.7	0.2	3.2	1.0	2.6	33.3	0.6	0.1	0.4	0.2	0.79
		1.6	0.5	0.6	1.0	1.8	2.0	19.8	0.0	0.1	0.0	0.72
		1.7	0.0	0.2	0.0	0.2	0.0	0.0	6.8	0.4	0.4	0.70
		0.6	0.2	1.2	0.2	1.0	0.2	0.4	0.0	25.4	0.2	0.86
		0.7	0.2	1.0	0.0	0.0	0.0	0.0	0.2	0.8	10.3	0.78
		0.93	0.80	0.80	0.81	**0.75**	0.86	0.90	0.85	0.90	0.85	

Table 5. IoU and RMSE of ROI diagonal length for femur fracture between detected ROI area and expected ROI area

	Our proposed detection model (deep faster R-CNN-BO)	
	RMSE of diagonal length of ROI (avg ± std, pixel)	IoU (avg ± std)
2 class	23.02 ± 39.17	0.88 ± 0.13
3 class	16.34 ± 47.01	0.91 ± 0.10
4 class	16.84 ± 46.34	0.89 ± 0.08
7 class	23.18 ± 66.58	0.88 ± 0.06
10 class	31.48 ± 54.83	0.87 ± 0.12

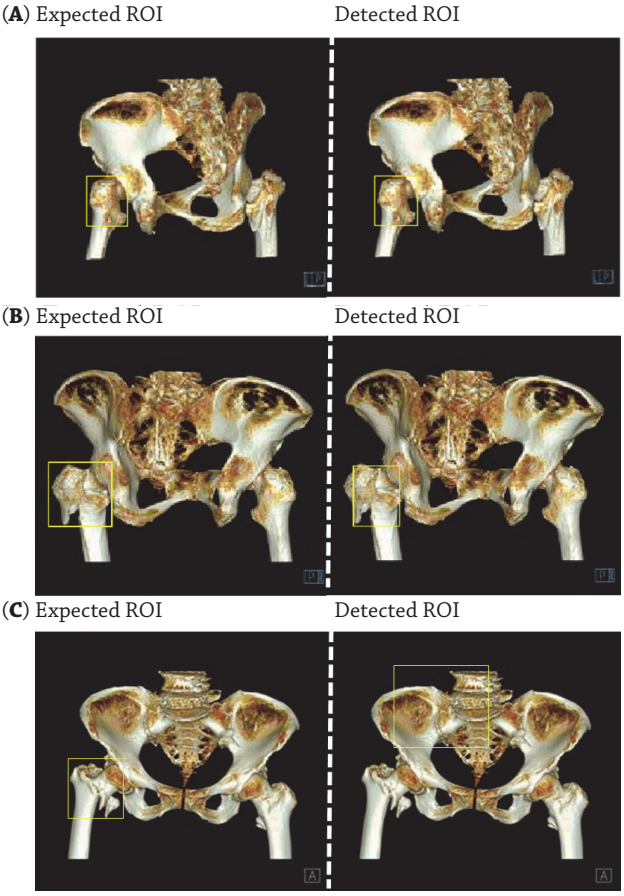

Fig. 5. Schematic diagrams of the previous and proposed femur fracture classification methods. (**A**) represents a well overlapped detected and expected ROI that shows relatively high IoU, (**B**) represents a less overlapped detected and expected ROI that shows relatively low IoU and (**C**) represents low IoU along with high RMSE value.

on X-ray images, resulting in poor readability on the sagittal plane when using them to evaluate the class of a fracture (Isida et al., 2015). For acute trauma in an IT femur fracture, CT scanning image is required to detect intra-articular fragments and surface fractures to detect the fracture pattern for surgical planning. The clinical and radiological results showed that the use of CT images is recommended to evaluate an unstable IT fracture. Detection of the correct class, especially for unstable type 2 IT fractures, will decrease the operation time and the risk of surgical failure (Han et al., 2010).

In this research, the deep faster R-CNN was applied for femur fracture detection. This is a two-stage detection model that uses an RPN instead of selection search methods. This method was faster than previous models such as R-CNN or fast R-CNN (Zhang et al., 2016); however, two-stage R-CNN models have identical disadvantages such as a complex pipeline, lack of real time feasibility, and difficulty in optimizing each hyperparameter. In this search, we applied the Bayesian optimization model to overcome the abovementioned limitation. In the future, a one-stage R-CNN model such as YOLO might be applicable to increase the object detection speed if the accuracy or MAP shows no difference compared to that of our model (Redmon et al., 2016). Additionally, in this study, the CNN architecture of VGG-16 was applied as the base feature extractor, which is the main detection framework applied in object detection models. While it has the advantage of accurate classification performance, it is somewhat complex because the convolution layers of VGG-16 calculate 30.69 billion floating point operations for a single image of size 224 × 224 pixels (Wu et al., 2017). One of the main aims of object detection models is improving the speed of object detection. Yu et al. (2016) reported that the VGG-16 model showed higher performance as compared to Alexnet; however, it has a twofold larger model size. Recent studies have applied other architectures to detect or classify fractures from medical images, such as the Squeezenet and Resnet, which may increase the performance. However, for instance, with Googlenet in multi box, the number of parameters used for the output layer is twofold as compared to that in VGG-16; this might not be suitable when the number of images used for training is small, as it might increase the overfitting. Therefore, the shared CNN architecture for deep faster R-CNN must be considered in the future to secure performance while considering various environments, such as computational load or number of images. Lastly, the total number of CT of IT femur fractures collected from patients was 3343, which may seem to be quite small. Several previous studies have increased the number of fracture images for training purposes by applying data augmentation techniques. For instance, Guan et al. (2020) applied data augmentation techniques such as horizontal flipping and random rotation of original X-ray images. Therefore, in the future, data augmentation should be considered to increase the number of images for training.

Conclusions

In our study, the type of IT femur fracture and its ROI were automatically determined from a CT image in a single step based on the AO/OTA classification method using the deep faster R-CNN-BO algorithm. It showed higher performance than that obtained in previous studies when classifying and detecting the ROI. In conclusion, our developed automatic femur fracture detection model can reduce the diagnostic differences owing to surgeons' experience levels and provide deeper insight when selecting or planning the treatment of a femur fracture.

Conflict of interests

The authors have no conflict of interests to declare.

Acknowledgement

This research was supported with funding from the Biomedical Research Institute, Chonbuk National University Hospital and Main Research Program (E0162500) of the Korea Food

Research Institute (KFRI), funded by the Ministry of Science and ICT.

References

Bayram F, Çakıroğlu M (2016). Diffract: Diaphyseal femur fracture classifier system. Biocybern Biomed Eng 36(1): 157–171. DOI: 10.1016/j.bbe.2015.10.003.

Boone C, Carlberg KN, Koueiter DM, Baker KC, Sadowski J, Wiater PJ, et al. (2014). Short versus long intramedullary nails for treatment of intertrochanteric femur fractures (OTA 31-A1 and A2). J Orthop Trauma 28(5): 96–100. DOI: 10.1097/BOT.0b013e3182a7131c.

Braun BJ, Holstein JH, Pohlemann T (2018). Intertrochanteric Hip Fracture: Intramedullary Nails. In: Egol KA, Leucht P (Eds), Proximal Femur Fractures. Springer, Cham, pp. 85–100.

Cho YC, Lee PY, Lee CH, Chen CH, Lin YM (2018). Three-dimensional CT Improves the Reproducibility of Stability Evaluation for Intertrochanteric Fractures. Orthop Surg 10(3): 212–217. DOI: 10.1111/os.12396.

Crijns TJ, Janssen SJ, Davis JT, Ring D, Sanchez HB, Althausen P, et al. (2018). Reliability of the classification of proximal femur fractures: Does clinical experience matter? Injury 49(4): 819–823. DOI: 10.1016/j.injury.2018.02.023.

Erickson BJ, Korfiatis P, Akkus Z, Kline TL (2017). Machine learning for medical imaging. Radiographics 37(2): 505–515. DOI: 10.1148/rg.2017160130.

Faisal M, Nistane P (2016). Proximal Femoral Nailing vs. Dynamic Hip Screw in unstable Intertrochanteric Fracture of Femur – A comparative analysis. Int J Biomed Adv Res 7(10): 489–492. DOI: 10.7439/ijbar.

Fung W, Jönsson A, Bühren V, Bhandari M (2007). Classifying intertrochanteric fractures of the proximal femur: does experience matter? Med Princ Practi 16(3): 198–202. DOI: 10.1159/000100390.

Guan Q, Huang Y, Zhong Z, Zheng Z, Zheng L, Yang Y (2020). Thorax disease classification with attention guided convolutional neural network. Pattern Recognition Letters 131: 38-45. DOI: 10.1016/j.patrec.2019.11.040.

Han SK, Lee BY, Kim YS, Choi NY (2010). Usefulness of multi-detector CT in Boyd-Griffin type 2 intertrochanteric fractures with clinical correlation. Skelet Radiol 39(6): 543–549. DOI: 10.1007/s00256-009-0795-6.

Isida R, Bariatinsky V, Kern G, Dereudre G, Demondion X, Chantelot C (2015). Prospective study of the reproducibility of X-rays and CT scans for assessing trochanteric fracture comminution in the elderly: a series of 110 cases. Eur J Orthop Surg Traumatol 25: 1165–1170. DOI: 10.1007/s00590-015-1666-6.

Jaderberg M, Simonyan K, Zisserman A (2015). Spatial transformer networks. In: Cortes C, Lee DD, Garnett R, Lawrence ND, Sugiyama M (Eds), Advances in neural information processing systems 28. Montreal, pp. 2017–2025.

Jin WJ, Dai LY, Cui YM, Zhou Q, Jiang LS, Lu H (2005). Reliability of classification systems for intertrochanteric fractures of the proximal femur in experienced orthopaedic surgeons. Injury 36(7): 858–861. DOI: 10.1016/j.injury.2005.02.005.

Kazi A, Albarqouni S, Sanchez AJ, Kirchhoff S, Biberthaler P, Navab N, et al. (2017). Automatic classification of proximal femur fractures based on attention models. In: Wang Q, Shi Y, Suk HI, Suzuki K (Eds), Machine Learning in Medical Imaging. MLMI 2017. Lecture Notes in Computer Science, vol. 10541. Springer, Cham. 70–78. DOI: 10.1007/978-3-319-67389-9_9.

Ker J, Wang L, Rao J, Lim T (2017). Deep learning applications in medical image analysis. IEEE Access 6: 9375–9389. DOI: 10.1109/ACCESS.2017.2788044.

Kim DH, MacKinnon T (2018). Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clin Radiol 73(5): 439–445. DOI: 10.1016/j.crad.2017.11.015.

Mears SC, Kates SL (2015). A guide to improving the care of patients with fragility fractures, edition 2. Geriatr Orthop Surg Rehab 6(2): 58–120. DOI: 10.1177/2151458515572697.

Pranata YD, Wang WC, Wang JC, Idram I, Lai JY, Liu JW, et al. (2019). Deep learning and SURF for automated classification and detection of calcaneus fractures in CT images. Computer Methods and Programs in Biomedicines 171: 27–37. DOI: 10.1016/j.cmpb.2019.02.006.

Redmon J, Divvala S, Girshick R, Farhadi A (2016). You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788.

Ren S, He K, Girshick R, Sun J (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems, pp. 91–99.

Segal D, Palmonovich E, Faour A, Marom E, Feldman V, Yaacobi E, et al. (2018). Routine early post-operative X-ray following internal fixation of intertrochanteric femoral fractures is unjustified: a quality improvement study. J Orthop Surg Res 13(1):189. DOI: 10.1186/s13018-018-0896-9.

Shahriari B, Swersky K, Wang Z, Adams RP, de Freitas N (2015). Taking the human out of the loop: A review of Bayesian optimization. Proceedings of the IEEE 104(1): 148–175. DOI: 10.1109/JPROC.2015.2494218.

Shen D, Wu G, Suk HI (2017). Deep learning in medical image analysis. Annu Rev Biomed Eng 19: 221–248. DOI: 10.1146/annurev-bioeng-071516-044442.

Sinno K, Sakr M, Girard J, Khatib H (2010). The effectiveness of primary bipolar arthroplasty in treatment of unstable intertrochanteric fractures in elderly patients. N Am J Med Sci 2(12): 561. DOI: 10.4297/najms.2010.2561.

Toderici G, Vincent D, Johnston N, Jin H, Minnen D, Shor J, et al. (2017). Full resolution image compression with recurrent neural network. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 5306–5314.

Urakawa T, Tanaka Y, Goto, S, Matsuzawa H, Watanabe K, Endo N (2019). Detecting intertrochanteric hip fracture with orthopedist-level accuracy using a deep convolutional neural network. Skeletal Radiol 48(2): 239–244. DOI: 10.1007/s00256-018-3016-3.

Vedaldi A, Lenc K (2015). Matconvnet: Convolutional neural networks for matlab. In: Proceedings of the 23rd ACM international conference on Multimedia, pp. 689–692.

Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E (2018). Deep learning for computer vision: A brief review. Computational Intelligence and Neuroscience. DOI: 10.1155/2018/7068349.

Wu B, Iandola F, Jin PH, Keutzer K (2017). Squeezedet: Unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 129–137.

Wu J, Davuluri P, Ward KR, Cockrell C, Hobson R, Najarian K (2012). Fracture detection in traumatic pelvic CT images. Int J Biomed Imaging. DOI: 10.1155/2012/327198.

Yates EJ, Yates LC, Harvey H (2018). Machine learning "red dot": open-source, cloud, deep convolutional neural networks in chest radiograph binary normality classification. Clin Radiol 73(9): 827–831. DOI: 10.1016/j.crad.2018.05.015.

Yu W, Yang K, Bai Y, Xiao T, Yao H, Rui Y (2016). Visualizing and comparing AlexNet and VGG using deconvolutional layers. In: Proceedings of the 33rd International Conference on Machine Learning.

Zhang L, Lin L, Liang X, He K (2016). Is faster R-CNN doing well for pedestrian detection? In: European conference on computer vision October, pp. 443–457. DOI: 10.1007/978-3-319-46475-6_28.