

Multi-classification of Osteoporosis Grading Stages Using Abdominal Computed Tomography with Clinical Variables : Application of Deep Learning with a Convolutional Neural Network

Tae Jun Ha¹, Hee Sang Kim², Seong Uk Kang³, DooHee Lee⁴, Woo Jin Kim^{5,6,10},
Ki Won Moon^{5,6,10}, Hyun-Soo Choi^{6,7}, Jeong Hyun Kim⁸, Yoon Kim^{4,9}, So Hyeon Bak^{13,*}, Sang Won Park^{10,11,12,*}

¹Medical Business Unit, DOUZONE BIZON Co., Ltd., Chuncheon, Republic of Korea

²Institute of Technology, POSOD Co., Ltd., Seoul, Republic of Korea

³Department of Biomedical Research Institute, Kangwon National University Hospital, Chuncheon, Re-public of Korea

⁴Department of Research and Development, ZIOVISION Co. Ltd., Chuncheon, Republic of Korea

⁵Department of Internal Medicine, Kangwon National University Hospital, Chuncheon, Republic of Korea

⁶Department of Internal Medicine, School of Medicine, Kangwon National University, Chuncheon, Republic of Korea

⁷Department of Computer Science and Engineering, Seoul National University of Science and Technology, Seoul, Republic of Korea

⁸Department of Urology, Kangwon National University School of Medicine, Chuncheon, Republic of Korea

⁹Department of Computer Science and Engineering, Kangwon National University, Republic of Korea

¹⁰Department of Medical Informatics, School of Medicine, Kangwon National University, Chuncheon, Republic of Korea

¹¹Institute of Medical Science, School of Medicine, Kangwon National University, Chuncheon, Republic of Korea

¹²Department of Data science, Weknew Co. Ltd., Chuncheon, Republic of Korea

¹³Department of Radiology, Asan Medical Center, Seoul, Republic of Korea

Received: April 26, 2024. Revised: June 05, 2024. Accepted: June 30, 2024.

ABSTRACT

Osteoporosis is a major health issue globally, often remaining undetected until a fracture occurs. To facilitate early detection, deep learning (DL) models were developed to classify osteoporosis using abdominal computed tomography (CT) scans. This study was conducted using retrospectively collected data from 3,012 contrast-enhanced abdominal CT scans. The DL models developed in this study were constructed for using image data, demographic/clinical information, and multi-modality data, respectively. Patients were categorized into the normal, osteopenia, and osteoporosis groups based on their T-scores, obtained from dual-energy X-ray absorptiometry, into normal, osteopenia, and osteoporosis groups. The models showed high accuracy and effectiveness, with the combined data model performing the best, achieving an area under the receiver operating characteristic curve of 0.94 and an accuracy of 0.80. The image-based model also performed well, while the demographic data model had lower accuracy and effectiveness. In addition, the DL model was interpreted by gradient-weighted class activation mapping (Grad-CAM) to highlight clinically relevant features in the images, revealing the femoral neck as a common site for fractures. The study shows that DL can accurately identify osteoporosis stages from

* Corresponding Author: So Hyeon Bak E-mail: arsgnm17@gmail.com Tel: +82-33-258-9109
Address: Dept. Radiology, Asan Medical Center, 88, Olympic-ro 43, Songpa, Seoul, Republic of Korea

* Co-Corresponding Author: Sang Won Park E-mail: chicwon229@kangwon.ac.kr Tel: +82-033-258-9109
Address: Dept. Medical informatics, 156, Baengnyeong-ro, Chuncheon, Gangwon, Republic of Korea

clinical data, indicating the potential of abdominal CT scans in early osteoporosis detection and reducing fracture risks with prompt treatment.

Keywords: Artificial intelligence, Bone marrow density, Osteoporosis, Clinical Decision Support Systems, Deep learning

I. INTRODUCTION

Osteoporosis has emerged as a growing global health concern exacerbated by an aging population and longer life spans^[1]. It is a systemic skeletal disease that decreases bone density and weakens the bone micro-architecture, increasing the risk of fractures^[2]. Although it is seen in all age groups, gender, and races, it is more common in Caucasians (white race), older people, and women. Currently, it has been estimated that more than 200 million people are suffering from osteoporosis. According to recent statistics from the International Osteoporosis Foundation, worldwide, 1 in 3 women over the age of 50 years and 1 in 5 men will experience osteoporotic fractures in their lifetime^[1,3]. Furthermore, it often remains asymptomatic until a fracture occurs, creating significant challenges for early diagnosis. This delay in detection can result in heightened risks of fractures, reduced quality of life, and elevated mortality rates^[4,5]. Among the older people aged > 70 years, the incidence of the disease is 18% in males and 68.5% in females, and the prevalence of osteoporosis appears to increase rapidly after menopause in females aged > 50 years^[4,6,7].

Dual-energy X-ray absorptiometry (DXA) is a commonly employed and established method for diagnosing osteoporosis, which delivers highly accurate measurements of bone mineral density (BMD) while minimizing radiation exposure^[8]. Its ability to focus on key areas, such as the hip and spine, makes it invaluable for assessing the risk of fracture in clinical settings. However, owing to the characteristics of the DXA, it is more time-consuming than other imaging scans, and requires specific conditions for accuracy, such as the patient's ability to supine position correctly. Consequently, for patients

with hip joint abnormalities or scoliosis, the scan can accompany painful, with the degree of discomfort varying from one individual to another. In other words, it may not provide a comprehensive view of bone health^[9,10]. This limitation, combined with the low screening rates, highlights the impracticality of relying solely on DXA for early diagnosis^[10-13]. Therefore, there is a need to develop additional methods for detecting osteoporosis. One possible way to improve osteoporosis identification rates is to use bone data obtained from abdominal computed tomography (CT) performed for other indications^[13-17]. Patients who undergo abdominal CT have a potential opportunity for BMD screening of the femur without the need for any additional imaging, radiation exposure, or patient time^[14]. Previous studies have demonstrated the feasibility of predicting osteoporosis by examining the femoral region through abdominal CT scans^[13,18,19]. Therefore, abdominal CT may be considered a valuable tool for assessing the risk of osteoporosis and osteopenia, as well as for distinguishing the different stages of the disease. With an exponential increase in computing power in the era of big data, deep-learning (DL) approaches have been rapidly adopted for the diagnosis of bone diseases, including osteoporosis^[20,21]. This DL-based artificial intelligence (AI) analysis can be used to elucidate the complex relationships between diverse features in medical images for osteoporosis and to make computer-aided diagnosis (CAD) by providing rapid results^[2,22-25].

Therefore, in this study, we implemented a DL framework for the multi-classification of osteoporosis using abdominal CT. We also compared the performance of abdominal CT with demographic and clinical information acquired from BMD and further investigated the effect of combining the two-modality

information for osteoporosis diagnosis performance.

II. MATERIAL AND METHODS

1. Study design and participant

A total of 3,012 image data were collected from 2,126 patients who underwent contrast-enhanced abdominal CT and DXA between January 2015 and October 2021. All CT images were acquired within \pm 3 months of DXA, and the data were labeled using the T-score from DXA. We excluded patients with foreign bodies resulting from femoral surgery, those with implanted artificial joints, or those lacking coronal CT phases from our study. Furthermore, as the majority of patients who underwent both DXA and CT within the study period exhibited abnormal bone density, we decided to include additional data from patients in their 20s who had exclusively undergone contrast-enhanced abdominal CT under identical conditions. A qualified radiologist assessed these patients, excluded those who did not have a fracture or chronic disease, and included those with normal bone density. Consequently, we developed models to evaluate the patients' bone density risk utilizing their CT images or demographic/clinical variables, such as sex, age, height, weight, and body mass index (BMI).

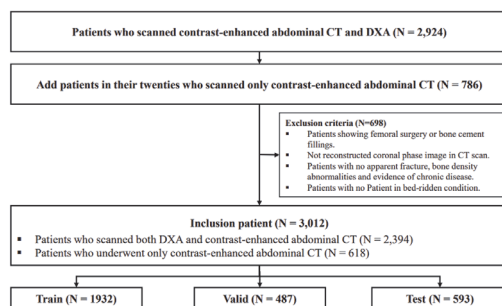


Fig. 1. Patient classification flowchart for modeling. Patients who scanned abdominal in CT in each group (normal, osteopenia, and osteoporosis) were classified. Within the training data, 20% was used as test data. Abbreviation: CT, computed tomography; DXA, dual-energy x-ray absorptiometry.

For a more comprehensive analysis, we also constructed a multi-modal model that incorporated

both types of data. The patient recruitment process is illustrated in Fig. 1.

All data used in this study was approved by the Institutional Review Board (IRB) (IRB number: A-2021-03-020) and the requirement for informed consent was waived because of the non-interventional observational nature of the study.

2. Image acquisition and measurements

Fig. 2 shows images for each step. All enhanced CT images were acquired using dual-source CT scanners with 64 and 128 detectors (SOMATOM Definition and SOMATOM Definition Flash; Siemens Healthineers, Forchheim, Germany), with the following parameters: detector collimation, 1.2 mm or 0.6 mm; field of view, 50 cm; tube voltage and tube current for x-ray exposure, 80–140 kVp; 125 mA; beam width, 38.4 mm or 28.8 mm; beam pitch, 0.6; and slice thickness 2 - 4 mm^[26]. All scanned images were obtained 80 s after the administration of the contrast agent at a rate of 2.6 ml/s using an auto-injector. Subsequently, 10 ml/s saline was injected at a rate of 2.5 ml/s. DXA was performed using Lunar Prodigy and Lunar Prodigy Advance (GE Healthcare Systems, Wauwatosa, WI, USA), and the T-score was calculated as the difference between the measured BMD and the mean BMD of females aged 20–40 years. According to the World Health Organization definitions, patients are classified based on their DXA-derived T-Score as follows: normal (T-Score \geq -1.0), osteopenia ($-2.5 < \text{T-Score} < -1.0$), and osteoporosis (T-Score ≤ -2.5)^[26].



(A) Normal (B) Osteopenia (C) Osteoporosis

Fig. 2. Femur images from abdominal coronal CT. The images were cropped to 250*250 at the bottom left and right depending on the inspection area.

3. Image preprocessing

Acquisition of the phase, including the femoral neck, head, and body, during contrast-enhanced abdominal CT was carried out following a thorough review by a radiologist. The selected digital imaging and communications in the medicine (DICOM) files were converted into grayscale images that could be used for training. The brightness and contrast of the images were adjusted using the specified window value parameters. By setting the window width and center parameters, the maximum and minimum values of the window can be defined as follows:

$$w_{\max} = w_{\text{center}} + (w_{\text{width}}/2) \quad (1)$$

$$w_{\min} = w_{\text{center}} - (w_{\text{width}}/2) \quad (2)$$

It can also be converted to the original unit of measure in the device generated using

$$\text{rescale}_{\text{pixel}} = \text{pixel} \times \text{rescale}_{\text{slope}} - \text{rescale}_{\text{intercept}} \quad (3)$$

In this study, by setting to 500 and to 2000, Equation (4) was used to convert the bones into an image such that they stood out.

$$\text{pixel} = \frac{\text{rescale}_{\text{pixel}} - \text{window}_{\min}}{\text{window}_{\max} - \text{window}_{\min}} \times 255 \quad (4)$$

Because the images obtained all had different sizes depending on the characteristics of the patient, and many other parts besides the femur were included, it was necessary to obtain the regions of interest. All the images were cropped to 250×250 pixels to include the femoral neck and head. During the training, augmentation was performed on the preprocessed images to prevent overfitting. The training image, rotation, zoom in/out, and translation

(up, down, left, and right) methods were randomly changed from -10 to 10 , and a vertical/horizontal flip method was applied. The probability of applying each augmentation method was set to 0.5.

4. Implementation models

We divided the training dataset into two subsets: training (80%) and validation (20%) datasets. All hyper-parameters were tuned during the validation phase. The validation set consisted of 20% of the training set. The proposed model was divided into segments designed to learn from both images and demographic information. Specifically, we constructed a model dedicated to processing CT images and demographic data while also conducting a comprehensive analysis by combining both datasets. In addition, we conducted a multi-modal data analysis by combining both sets of data. As depicted in Fig. 3, the DL model architecture for extracting image features consists of six convolutional layers (16,32,64,128,256,512); kernel_size = 3, same padding, and Maxpool2D were applied to each layer to use the activation function of rectified linear unit (ReLU). Dropout (0.2) was applied to the last hidden layer and three fully connected layers (4096, 1024, 128) were constructed. The adaptive moment estimation (Adam) a first-order gradient-based probability optimization algorithm optimizer was used with the mini-batch size was set to 32 and the epoch was set to 500. The learning rate was set to 0.001 with decay rate of 0.96 and adjusted to 70% if the performance did not improve after 30 epochs. Whereas, the model for only using demographic information consists of two fully connected layers. In addition, the DL model for multi-modality was constructed by concatenating extracted features from images with extracted features from clinical information. It is used same convolutional layers used for DL using CT images and constructed two fully connected layers (64,16). Each modality is learned and transformed into a feature map, which is the key to the classification.

The two generated feature maps were expressed by concatenation and classified into normal, osteopenia, and osteoporosis groups through two fully connected layers. Each layer uses a ReLU activation function and is configured to learn quickly using batch normalization. If learning did not improve after 50 epochs, it was ended early to shorten the experimental time. A deep learning model was developed using one RTX A5000 GPU and programmed by using Python 3.8.10 version and Tensorflow 2.4.0.

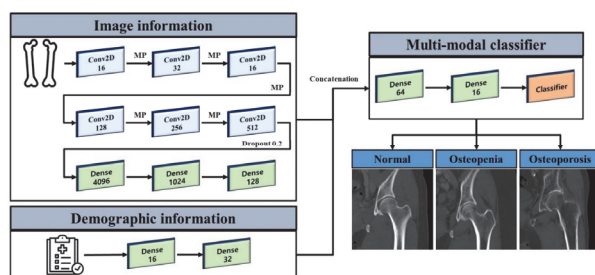


Fig. 3. The process for layer stacking of deep learning model

5. Informative feature identification for multi-classification

To identify informative features extracted through convolutional neural network (CNN) models, gradient-weighted class activation mapping (Grad-CAM) was used to provide a clinical interpretation of the results. The GRAD-CAM can be used to identify informative features extracted through DL models^[27]. To identify informative features extracted through CNN models, Grad-CAM was used. The feature map could be visualized with the average pixel value up to the final layers. We identified regions in the femur by applying the ReLU activation function to visualize important parts of the model during the analysis process. In addition, a radiologist's judgment was used to identify whether it matched the greater trochanter and femoral neck regions, which are areas where osteoporosis occurs.

6. Statistical analysis

The Chi-squared (χ^2) test and analysis of variance (ANOVA) were used to confirm the differences in ratio and mean among the three groups (normal, osteopenia, osteoporosis). After conducting Levene's test to assess the equality of variances, an analysis of variance (ANOVA) was performed to test the difference in the means among the three groups. In addition, if the assumption of equal variance was not satisfied, Welch's ANOVA was performed to test for mean differences among groups, and the Games-Howell test was conducted to confirm post-hoc analysis.

III. RESULTS

1. Patient characteristics

The characteristics of all patients included in this study are presented in Table 1. The mean age of all patients was 58.1 years, with the average age of each group as follows: normal individuals = 36.6 years, individuals with osteopenia = 62.5 years, and individuals with osteoporosis = 76.5 years. Among all patients, 2,367 were females (78.6%), of whom 777 (72.8%) were normal, 791 (85.5%) had osteopenia, and 799 (78.4%) had osteoporosis. Overall, the weight of the patients was decreased as the disease severity was increased; the mean weight of normal patients was 64.2 kg. For patients with osteopenia, it was 59.7 kg, and for patients with osteoporosis, it was 53.7 kg. The differences among the three groups for all variables were statistically significant ($P < 0.001$), as determined by ANOVA. Post-hoc analysis of individual groups revealed statistically significant differences between all groups, except for BMI between the normal and osteopenia groups and sex between the normal and osteoporosis groups.

Table 1. Patients characteristics

	Normal (N=1,068)	Osteopenia (N=925)	Osteoporosis (N=1,019)	Total (N=3,012)
Age*	36.6 ± 15.6	62.5 ± 11.9	76.5 ± 9.6	58.1 ± 21.1
Sex*				
Male	291 (27.2%)	134 (14.5%)	220 (21.6%)	645 (21.4%)
Female	777 (72.8%)	791 (85.5%)	799 (78.4%)	2367 (78.6%)
Height*	1.6 ± 0.1	1.6 ± 0.1	1.5 ± 0.1	1.6 ± 0.1
Weight*	64.2 ± 13.6	59.7 ± 9.7	53.7 ± 9.1	59.3 ± 11.9
BMI*	24.1 ± 4.4	24.5 ± 3.7	23.0 ± 3.7	23.9 ± 4.0

The superscript * indicates that the variable has statistical significance ($P < 0.001$), and the superscript ** indicates that the variable has significance ($P < 0.05$) among the three groups according to Welch's ANOVA and Chi-squared test. Post-hoc analysis of individual groups revealed statistically significant differences between all groups. However, there were no statistically significant differences in BMI between the normal and osteopenia groups and in sex between the normal and osteoporosis groups. Continuous variables are presented as the mean ± SD, and categorical variables are presented as the count (%) unless otherwise stated. Abbreviation: BMI, Body mass index.

2. Model performance

Table 2 suggests the model performance results. All metrics in the multi-modal data showed the highest performance, which was more significant compared with that of the model using only demographic information for grading stage classification. The area under the receiver operating characteristic curve (AUC) of the multi-modal model was the highest (0.94) with an accuracy (ACC) of 0.80, indicating a 10% better performance than that of the model using

only demographic information. The model using demographic information only showed the worst performance with an AUC of 0.85 and an ACC of 0.68. In addition, there was a larger difference in sensitivity than in specificity between the two models. The model using a multi-modal dataset had a sensitivity and specificity of 0.80 and 0.90, respectively, whereas the model using only demographic data had a sensitivity and specificity of 0.69 and 0.84, respectively. Similar results were obtained for precision and recall.

Table 2. Model performance

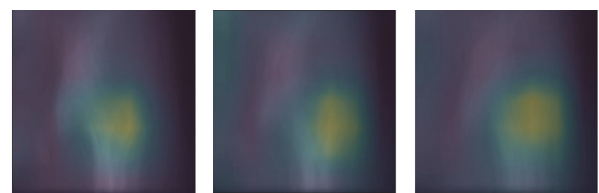
Model	Data modality	ACC	Sensitivity	Specificity	Precision	F1-Score	AUC
6-layer CNN	Image	0.794	0.795	0.897	0.801	0.798	0.933
2-layer FNN	Demographic / clinical variables	0.689	0.690	0.844	0.692	0.691	0.859
6-layer CNN with 2-layer FNN	Multi-modality	0.802	0.802	0.901	0.804	0.803	0.939

The total dataset consisted of 3,012 CT scans. The training dataset comprised 2,419 scans, and the test data comprised 593 scans. Abbreviation: ACC, Accuracy; AUC, Area under the receiver operating characteristic curve; CNN, Convolutional neural network; FNN, Fully connected neural network.

3. Identification of informative features for classification

To identify informative features extracted through the CNN, Grad-CAM was used (Fig. 4), allowing the visualization of the results during the analysis process. Fig. 4 shows the images and gradients extracted from the data, which are presented at an opacity of 0.7 and 0.3, respectively, for the normal, osteopenia, and osteoporosis groups.

Based on the Grad-Cam results, distinctive regions extracted from CT scans were associated with osteoporosis and fracture induced by osteoporosis, such as the greater trochanter and femoral neck regions.



(A) Normal (B) Osteopenia (C) Osteoporosis

Fig. 4. The results presented by Grad-CAM for each grading stage. The images used and the extracted gradients were averaged to obtain the overall results. The averaged image and gradient were mixed and presented at opacity of 0.7 and 0.3, respectively.

The regions that contributed significantly to the model results are shown in green; notably, the femoral neck was observed in all stages.

IV. DISCUSSION

In this study, the multi-classification of osteoporosis into different stages (normal, osteopenia, and osteoporosis) was conducted through DL with abdominal CT, which is already widely performed in clinical practice. Using the developed multi-modal DL model, CT images captured for diverse medical purposes can be used to screen latent patients at risk of osteoporosis without additional costs and radiation exposure. Our model, implemented for osteoporosis classification, demonstrated superior performance by incorporating both imaging and demographic/clinical information from individual patients. The model could provide an interpretable DL method to enhance clinician decision support systems (CDSS). This classification approach can aid clinical decision-making by estimating the probability of osteoporosis from abdominal CT images that were initially acquired for other medical reasons. Here, modeling was performed separately using CT images, demographic data, and a combination of CT images and demographic data. The model using multi-modal data performed well in multi-classification with an AUC and ACC of 0.94 and 0.80, respectively, by combining the data from two different modalities.

In general, osteoporosis has a high prevalence and is known to present with difficulty in detecting fractures before disease progression^[25]. Therefore, most older patients miss the opportunity to combat the risk of fracture owing to decreased bone density. Women > 65 and men > 70 years of age are exposed to many factors that can cause osteoporosis, such as low body weight and a history of previous fractures^[26]. Although DXA is a representative method

used to diagnose osteoporosis, it has some disadvantages, such as a low utility rate and radiation exposure. Moreover, practical challenges, such as the requirement for specialized equipment knowledge and limited accessibility due to low penetration rates exist. Consequently, an alternative approach to osteoporosis detection involves utilizing bone data derived from abdominal CT scans, a method widely recognized as a valuable tool for osteoporosis screening in general^[17,23,28,29]. Abdominal CT is a medical imaging modality that can examine the spine and femur and is known to be useful for accurately measuring the risk of osteoporosis in the area of interest. Proximal femoral fractures are fatal fractures with high morbidity and mortality rates despite representing only a small proportion of osteoporotic fractures.

Importantly, our study offers several novelties for the multi-classification of osteoporosis into three stages based on the femur region on abdominal CT images. First, our study results provide an opportunity to overcome the shortcomings of DXA and quickly respond to the potential risk of the disease using widely used and easily obtained CT images and demographic data. In particular, we utilized real-world clinical data to address the challenge of concurrently classifying individuals into the normal, osteopenia, and osteoporotic stages, leveraging observations from the femur region in abdominal CT. Previous studies have primarily focused on classifying individuals as either normal or having osteoporosis, with limited attention given to classifying osteopenia—a stage that falls between the two disease categories—as either normal or osteoporotic. Moreover, few studies have classified osteopenia, which lies between the two disease stages, as normal or osteoporosis (Table 3).

Table 3. Comparison results of this study with those of other studies

Models	Modality	Dataset (Train/Test)	ROI	Model	Normal vs Osteopenia		Normal vs Osteoporosis		Osteopenia vs Osteoporosis	
					AUC	ACC	AUC	ACC	AUC	ACC
Liu et al. ^[31]	X-ray	89 (NR)	NR	U-net CNN	0.88	0.81	0.87	0.82	0.75	0.71
Yamamoto et al. ^[24]	X-ray	1133 (598/535)	femur	ResNet-(18,34), GoogleNet, EfficientNet	-	-	0.93	0.88	-	-
Yasaka et al. ^[32]	CT	1760 (1665/95)	lumbar spine	4-layers CNN	0.95	0.86	0.97	0.96	-	-
This study	CT	3046 (2446/600)	femur	6-layers CNN	0.91	0.82	0.99	0.97	0.93	0.86

We obtained an AUC of 0.94, ACC of 0.81, precision and sensitivity of 0.8, and specificity and recall of 0.1 upon utilizing a multi-modal dataset. These outcomes surpass those reported in several recent studies^[20,30]. Notably, our model could exceed the performance achieved by machine learning-based X-ray image analysis in various existing studies^[24,30,31]. Zhang et al. performed multi-classification using the same CNN model used by our team but showed a performance of AUC 0.81 and ACC 0.6, lower than our results^[30]. Liu et al. performed binary classification for each of the three groups (normal, osteopenia, osteoporosis) and presented results of AUC 0.88 for classification between normal and osteopenia, AUC 0.87 for classification between normal and osteoporosis, and AUC 0.75 for classification between osteopenia and osteoporosis^[31]. Yamamoto et al. presented relatively high model performance results of AUC 0.93 and ACC 0.88 as a result of the classification between normal and osteoporosis^[24]. However, compared with our study, these studies used X-ray images and were conducted on a smaller number of patients. In addition, a study by Yasaka et al. showed superior performance in classifying between normal and osteopenia; however, generalization would be difficult because the dataset for validation was very small in that study^[32] in contrast to our study. Their study may be similar to our study in that it used abdominal CT images;

however, it differs from our study in that it examined BMD based on the vertebrae. It is generally difficult to perform CT and DXA for patients with compression fractures or comminuted fractures of the vertebrae. In this study, we present a model with high accessibility in clinical practice with enhanced clinical applicability through multi-classification. As we used CT images, information on the femur region can be obtained from various angles. Furthermore, through abdominal CT, the disease can be identified much more easily in the DL model according to X-ray absorbance differences, which can be distinguished according to a decrease in bone density.

In addition, the prevention of fractures is of significant importance. Our model, which can be applied using CT scans commonly obtained in clinical settings for other reasons, allows the early diagnosis of osteoporosis in patients. Specifically, the ability to identify potential osteoporosis during an abdominal CT conducted for various medical reasons can encourage patients to undergo more precise diagnostic examinations such as BMD tests for osteoporosis and seek appropriate treatment. Consequently, these individuals may undergo treatment aimed at preventing fatal fractures, which includes receiving guidance on proper dietary intake, adopting fall prevention strategies, and benefiting from pharmacological interventions^[3]

In particular, we used the results of the DXA scan,

which is used as a diagnostic standard for osteoporosis in clinical practice, are used as labels for each disease group based on review by a radiologist, suggesting a high possibility of clinical application.

Second, our study presents the explanatory potential of DL models. One prevalent issue with DL-based approaches is the 'black box' nature of these models, which hinders a clear understanding of their internal processes^[33]. Misinterpretations by AI-based models can lead to incorrect diagnoses, emphasizing the need for model validation. In our study, we employed Grad-CAM to visualize and elucidate the model's inferred rationale, confirming the alignment between the regions identified within the model and clinically relevant areas (Fig.4)^[34-36]. Based on the results of Grad-CAM analysis, we focused on feature extraction, with a specific emphasis on the femoral neck. The consistency with clinical findings highlights the femoral neck as the most vulnerable area to osteoporosis. Furthermore, our study carries implications for expanding the clinical applications of the model. By collecting data spanning all age groups in the range of 20–70 years, the multi-classification of osteoporosis for various age brackets may be possible. Moreover, our findings remain applicable even for individuals with scoliosis or those who have undergone femur-related surgery, as we leverage the entire thigh region observable in abdominal CT scans. Additionally, by harnessing all femoral bone images within abdominal CT scans, categorized into the neck, head, and torso, we obtain input data with minimal need for preprocessing in most cases. This presents significant potential for the rapid proliferation of computer-aided diagnosis. Lastly, we conducted a comprehensive analysis to identify errors that may have arisen from differences in data distribution while classifying osteoporosis (Fig. 5). It is reasonable to encounter classification challenges at the boundary points that demarcate these stages, given that disease risks can vary within the same stage based on T-score values. To visualize these discrepancies, we used the

T-score. Our analysis reaffirmed the accuracy of most classifications between normal and osteoporotic tissues while highlighting that errors primarily concentrate on the transitional boundaries between disease stages.

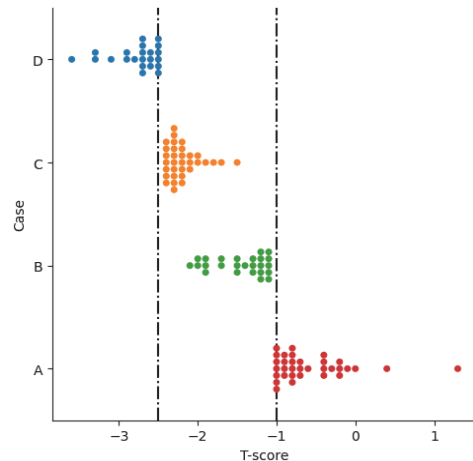


Fig. 5. T-score distribution showing errors between the predicted and the correct results.

However, this study has several limitations. First, the performance was guaranteed only for contrast-enhanced abdominal CT data. Although CT is a common imaging modality, it seldom provides BMD information in the clinic owing to technical difficulties. Therefore, DXA is required to measure BMD at the expense of additional radiation exposure. However, DXA may not be readily accessible in all healthcare facilities. Additionally, patients might not undergo this examination if specialists do not suspect them of having osteoporosis. Nevertheless, abdominal CT is generally performed for patients who do not have kidney function abnormalities or contrast agent side effects. Accordingly, we conducted this study using contrast-enhanced abdominal CT scans, which demonstrated high accuracy in osteoporosis classification. Second, there was a time gap in the data collected in this study. All patient data were obtained using a concomitant CT scan within 3 months before or after the DXA scan to collect as much data as possible. Therefore, based on a single DXA examination, two or more CT images may be matched for the same patient. Third, data were

collected from a single medical institution. It is difficult to prove this effect using data obtained from other institutions or CT equipment. The quality of CT images can be influenced by factors such as the type of imaging equipment, the imaging protocol used, and the manufacturer of the equipment. In general, using only CT images for diagnosing or screening osteoporosis has definite clinical limitations. Specifically, for the purpose of diagnosis for osteoporosis, DXA testing, which has the potential to become an international diagnostic standard, is necessary. However, given that contrast-enhanced abdominal CT is predominantly used unless there is impaired kidney function, our osteoporosis assessment was conducted using contrast-enhanced abdominal CT images. We anticipate that future studies involving multiple centers will be essential to validate and extend the findings of our research, which was based on CT images obtained from a single institution. In the future, we plan to collect more data from several machines and hospitals to reduce bias and increase robustness. Furthermore, to reinforce the reliability of the internal validation results obtained in this study, it is crucial to carry out external validation of the model. Fourth, we confirmed that using clinical structured data and unstructured CT image data simultaneously improved the performance compared to using individual data independently. However, the results obtained using only images were not significantly different from those obtained using multi-modal data. This finding suggests that the demographic data used in this study had only a minor effect. Although demographic information shows a small effect in improving the overall performance of the model, this can indicate the possibility of generalization of the model and its versatility in clinical practice environments by using only the initial screening information of subjects. Improved performance can be expected by additionally using clinical variables directly related to bone density, such as drugs and disease history; however, this may result

in a trade-off for widely used in clinical practice. Finally, this study has also a shallow structural part of the model. Although we used various models such as ResNet and Mini ResNet to conduct for this study, there was an overfitting problem. Therefore, although it is a six-layer CNN based model, we implemented and suggested a model optimized for the data used in this study. Fifth, our study based on a specific population and imaging may not be broadly applicable to diverse global populations due to variations in demographic, genetic, and environmental factors. This indicates that to enhance the robustness of the model, it will be necessary to gather more data in the future and to take into accounts the characteristics of various population groups. Sixth, given that most osteoporosis occurs in the elderly population, the model was built by including images from patients in their 20s who did not have a DXA scan to ensure the correct classification of osteoporosis and osteopenia through classification from a normal group with high bone density. We added a group of patients who were judged to be normal to minimize bias in the results and to ensure the correct classification of osteopenia and osteoporosis. In general, in clinical practice, DXA is not performed in the absence of obvious osteoporosis findings. Although we included a group of patients with normal bone density based on radiologist findings and diagnosis, the lack of DXA may be a limitation of this study. Further research may be needed to improve the model to be more robust by obtaining DXA results from a population with normal bone density.

V. CONCLUSION

In conclusion, we developed a DL model for the multi-classification of osteoporosis using real-world clinical data combining CT scanned images with variables. Additionally, we discussed important features selected based on Grad-CAM technology. This implies that DL can be fully applied to medical

data for the classification of osteoporosis. In addition, our results suggest that abdominal CT could be used as important data in osteoporosis screening and lead to appropriate treatment for the reduction of osteoporotic fractures.

Acknowledgement

Author Contributions

Conceptualization, S.H.B.; methodology, T.J.H. and H.S.K.; software, validation, formal analysis, S.H.B., T.J.H., S.W.P., and H.S.K.; investigation, T.J.H., D.H.L., N.Y.Y., and S.U.K.; data curation, S.H.B., T.J.H. and H.S.K.; writing – original draft preparation, S.H.B., T.J.H. and S.W.P.; writing – review and editing, W.J.K., H.S.C., J.H.K., Y.K and K.W.M.; visualization, S.W.P.; supervision, S.W.P.; project administration, funding acquisition, S.W.P and S.H.B. All authors have read and agreed to the published version of the manuscript.

Funding

This research was supported by "Institute for Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT)(No.RS-2022-II221196, Regional strategic Industry convergence security core talent training business) and supported by the Promotion of Innovative Businesses for Regulation-Free Special Zones funded by the Ministry of Small and Medium-sized Enterprises (SMEs) and Startups (MSS, South Korea) (P0020626).

Institutional Review Board Statement

This study was approved by the Institutional Review Board (IRB) (IRB number: A-2021-03-020) and the requirement for informed consent was waived because of the non-interventional observational nature of the study.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets analyzed in the current study are available from the corresponding

author on reasonable request because they are not publicly available data.

Conflicts of Interest

The authors declare no conflicts of interest.

There is a preprinted version of the above paper (<https://doi.org/10.21203/rs.3.rs-3440051/v1>)

Reference

- [1] T. Sözen, L. Özışık, N. Ç. Başaran, "An overview and management of osteoporosis", *European Journal of Rheumatology*, Vol. 4, No. 1, pp. 46-56, 2017. <http://dx.doi.org/10.5152/EURJRHEUM.2016.048>
- [2] H. Abbouchie, N. Raju, A. Lamanna, C. Chiang, N. Kutaiba, "Screening for osteoporosis using L1 vertebral density on abdominal CT in an Australian population", *Clinical Radiology*, Vol. 77, No. 7, pp. e540-e548, 2022. <http://dx.doi.org/10.1016/j.crad.2022.04.002>
- [3] F. Cosman, S. J. de Beur, M. S. LeBoff, E. M. Lewiecki, B. Tanner, S. Randall, R. Lindsay, "Clinician's Guide to Prevention and Treatment of Osteoporosis", *Osteoporosis International*, Vol. 25, No. 10, pp. 2359-2381, 2014. <http://dx.doi.org/10.1007/s00198-014-2794-2>
- [4] NIH Consensus Development Panel on Osteoporosis Prevention, Diagnosis, and Therapy, "Osteoporosis prevention, diagnosis, and therapy", *The Journal of the American Medical Association*, Vol. 285, No. 6, pp. 785-795, 2001. <http://dx.doi.org/10.1001/jama.285.6.785>
- [5] M. T. Löffler, A. Jacob, A. Scharr, N. Sollmann, E. Burian, M. El Husseini, et al., "Automatic opportunistic osteoporosis screening in routine CT: improved prediction of patients with prevalent vertebral fractures compared to DXA", *European Radiology*, Vol. 31, No. 8, pp. 6069-6077, 2021. <http://dx.doi.org/10.1007/s00330-020-07655-2>
- [6] J. S. Yu, N. G. Krishna, M. G. Fox, D. G. Blankenbaker, M. A. Frick, S. T. Jawetz, et al., "ACR Appropriateness Criteria® Osteoporosis and Bone Mineral Density: 2022 Update", *Journal of the American College of Radiology*, Vol. 19, No. 11, pp. 417-S432, 2022. <http://dx.doi.org/10.1016/j.jacr.2022.09.007>
- [7] S. H. Ahn, S. M. Park, S. Y. Park, J. I. Yoo, H. S.

- Jung, J. H. Nho, et al., "Osteoporosis and Osteoporotic Fracture Fact Sheet in Korea", *Journal of Bone Metabolism*, Vol. 27, No. 4, pp. 281-290, 2020. <http://dx.doi.org/10.11005/jbm.2020.27.4.281>
- [8] H. P. Dimai, "Use of dual-energy X-ray absorptiometry (DXA) for diagnosis and fracture risk assessment; WHO-criteria, T- and Z-score, and reference databases", *Bone*, Vol. 104, pp. 39-43, 2017. <http://dx.doi.org/10.1016/j.bone.2016.12.016>
- [9] A. B. King, D. M. Fiorentino, "Medicare payment cuts for osteoporosis testing reduced use despite tests' benefit in reducing fractures", *Health Affairs (Millwood)*, Vol. 30, No. 12, pp. 2362-2370, 2011. <http://dx.doi.org/10.1377/hlthaff.2011.0233>
- [10] M. Pazianas, B. Abrahamsen, Y. Wang, R. G. Russell, "Incidence of fractures of the femur, including subtrochanteric, up to 8 years since initiation of oral bisphosphonate therapy: a register-based cohort study using the US MarketScan claims databases", *Osteoporosis International*, Vol. 23, No. 12, pp. 2873-2884, 2012. <http://dx.doi.org/10.1007/s00198-012-1952-7>
- [11] A. L. Amarnath, P. Franks, J. A. Robbins, G. Xing, J. J. Fenton, "Underuse and Overuse of Osteoporosis Screening in a Regional Health System: a Retrospective Cohort Study", *Journal of General Internal Medicine*, Vol. 30, No. 12, pp. 1733-1740, 2015. <http://dx.doi.org/10.1007/s11606-015-3349-8>
- [12] J. R. Curtis, L. Carbone, H. Cheng, B. Hayes, A. Laster, R. Matthews, et al., "Longitudinal trends in use of bone mass measurement among older Americans, 1999-2005", *Journal of Bone and Mineral Research*, Vol. 23, No. 7, pp. 1061-1067, 2008. <http://dx.doi.org/10.1359/jbmr.080232>
- [13] S. J. Lee, P. A. Anderson, P. J. Pickhardt, "Predicting Future Hip Fractures on Routine Abdominal CT Using Opportunistic Osteoporosis Screening Measures: A Matched Case-Control Study", *American Journal of Roentgenology*, Vol. 209, No. 2, pp. 395-402, 2017. <http://dx.doi.org/10.2214/AJR.17.17820>
- [14] R. M. Summers, N. Baecher, J. Yao, J. Liu, P. J. Pickhardt, J. R. Choi, et al., "Feasibility of Simultaneous Computed Tomographic Colonography and Fully Automated Bone Mineral Densitometry in a Single Examination", *Journal of Computed Assisted Tomography*, Vol. 35, No. 2, pp. 212-216, 2011. <http://dx.doi.org/10.1097/RCT.0b013e3182032537>
- [15] P. J. Pickhardt, L. J. Lee, A. M. del Rio, T. Lauder, R. J. Bruce, R. M. Summers, et al., "Simultaneous screening for osteoporosis at CT colonography: bone mineral density assessment using MDCT attenuation techniques compared with the DXA reference standard", *Journal of Bone and Mineral Research*, Vol. 26, No. 9, pp. 2194-2203, 2011. <http://dx.doi.org/10.1002/jbmr.428>
- [16] P. J. Pickhardt, B. D. Pooler, T. Lauder, A. M. del Rio, R. J. Bruce, N. Binkley, "Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications", *Annals of Internal Medicine*, Vol. 158, No. 8, pp. 588-595, 2013. <http://dx.doi.org/10.7326/0003-4819-158-8-201304160-00003>
- [17] S. J. Lee, N. Binkley, M. G. Lubner, R. J. Bruce, T. J. Ziemlewicz, P. J. Pickhardt, "Opportunistic screening for osteoporosis using the sagittal reconstruction from routine abdominal CT for combined assessment of vertebral fractures and density", *Osteoporosis International*, Vol. 27, No. 3, pp. 1131-1136, 2016. <http://dx.doi.org/10.1007/s00198-015-3318-4>
- [18] M. P. Rosen, B. Siewert, D. Z. Sands, R. Bromberg, J. Edlow, V. Raptopoulos, "Value of abdominal CT in the emergency department for patients with abdominal pain", *European Radiology*, Vol. 13, No. 2, pp. 418-424, 2003. <http://dx.doi.org/10.1007/s00330-002-1715-5>
- [19] P. J. Pickhardt, T. Lauder, B. D. Pooler, A. Muñoz Del Rio, H. Rosas, R. J. Bruce, N. Binkley, "Effect of IV contrast on lumbar trabecular attenuation at routine abdominal CT: correlation with DXA and implications for opportunistic osteoporosis screening", *Osteoporosis International*, Vol. 27, No. 1, pp. 147-152, 2016. <https://doi.org/10.1007/s00198-015-3224-9>
- [20] B. Suh, H. Yu, H. Kim, S. Lee, S. Kong, J. W. Kim, et al., "Interpretable Deep-Learning Approaches for Osteoporosis Risk Screening and Individualized Feature Analysis Using Large Population-Based Data: Model Development and Performance Evaluation", *Journal of Medical Internet Research*, Vol. 25, e40179, 2023. <http://dx.doi.org/10.2196/40179>
- [21] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning", *Nature*, Vol. 521, No. 7553, pp. 436-444, 2015. <https://doi.org/10.1038/nature14539>

- [22] A. Tarekegn, F. Ricceri, G. Costa, E. Ferracin, M. Giacobini, "Predictive Modeling for Frailty Conditions in Elderly People: Machine Learning Approaches", *JMIR Medical Informatics*, Vol. 8, No. 6, e16678, 2020. <https://doi.org/10.2196/16678>
- [23] J. Smets, E. Shevroja, T. Hügle, W. D. Leslie, D. Hans, "Machine Learning Solutions for Osteoporosis-A Review", *Journal of Bone and Mineral Research*, Vol. 36, No. 5, pp. 833-851, 2021. <http://dx.doi.org/10.1002/jbmr.4292>
- [24] N. Yamamoto, S. Sukegawa, A. Kitamura, R. Goto, T. Noda, K. Nakano, et al., "Deep Learning for Osteoporosis Classification Using Hip Radiographs and Patient Clinical Covariates", *Biomolecules*, Vol. 10, No. 11, 1534, 2020. <http://dx.doi.org/10.3390/biom10111534>
- [25] M. Zeytinoglu, R. K. Jain, T. J. Vokes, "Vertebral fracture assessment: Enhancing the diagnosis, prevention, and treatment of osteoporosis", *Bone*, Vol. 104, pp. 54-65, 2017. <http://dx.doi.org/10.1016/j.bone.2017.03.004>
- [26] J. A. Kanis, "Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a WHO report", *Osteoporosis International*, Vol. 4, No. 6, pp. 368-381, 1994. <http://dx.doi.org/10.1007/BF01622200>
- [27] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization", In *Proceedings of the IEEE international conference on computer vision*, pp. 618-626, 2017.
- [28] G. Yang, H. Wang, Z. Wu, Y. Shi, Y. Zhao, "Prediction of osteoporosis and osteopenia by routine computed tomography of the lumbar spine in different regions of interest", *Journal of Orthopaedic Surgery and Research*, Vol. 17, No. 1, pp. 454, 2022. <http://dx.doi.org/10.1186/s13018-022-03348-2>
- [29] Y. W. Kim, J. H. Kim, S. H. Yoon, J. H. Lee, C. H. Lee, C. S. Shin, et al., "Vertebral bone attenuation on low-dose chest CT: quantitative volumetric analysis for bone fragility assessment", *Osteoporosis International*, Vol. 28, No. 1, pp. 329-338, 2017. <http://dx.doi.org/10.1007/s00198-016-3724-2>
- [30] B. Zhang, K. Yu, Z. Ning, K. Wang, Y. Dong, X. Liu, et al., "Deep learning of lumbar spine X-ray for osteopenia and osteoporosis screening: A multicenter retrospective cohort study", *Bone*, Vol. 140, 115561, 2020. <http://dx.doi.org/10.1016/j.bone.2020.115561>
- [31] J. Liu, J. Wang, W. Ruan, C. Lin, D. Chen, "Diagnostic and Gradation Model of Osteoporosis Based on Improved Deep U-Net Network", *Journal of Medical Systems*, Vol. 44, No. 1, 15, 2019.
- [32] K. Yasaka, H. Akai, A. Kunimatsu, S. Kiryu, O. Abe, "Prediction of bone mineral density from computed tomography: application of deep learning with a convolutional neural network", *European Radiology*, Vol. 30, No. 6, pp. 3549-3557, 2020. <http://dx.doi.org/10.1007/s00330-020-06677-0>
- [33] Z. C. Lipton, "The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery", *Queue*, Vol. 16, No. 3, pp. 31-57, 2018. <https://doi.org/10.1145/3236386.3241340>
- [34] G. Jones, T. Nguyen, P. Sambrook, P. J. Kelly, J. A. Eisman, "Progressive loss of bone in the femoral neck in elderly people: longitudinal findings from the Dubbo osteoporosis epidemiology study", *BMJ*, Vol. 309, No. 6956, pp. 691-695, 1994. <http://dx.doi.org/10.1136/bmj.309.6956.691>
- [35] N. C. Wright, A. C. Looker, K. G. Saag, J. R. Curtis, E. S. Delzell, S. Randall, et al., "The recent prevalence of osteoporosis and low bone mass in the United States based on bone mineral density at the femoral neck or lumbar spine", *Journal of Bone and Mineral Research*, Vol. 29, No. 11, pp. 2520-2526, 2014. <http://dx.doi.org/10.1002/jbmr.2269>
- [36] J. A. Kanis, "Diagnosis of osteoporosis and assessment of fracture risk", *Lancet*, Vol. 359, No. 9321, pp. 1929-1936, 2002. [http://dx.doi.org/10.1016/S0140-6736\(02\)08761-5](http://dx.doi.org/10.1016/S0140-6736(02)08761-5)

멀티 모달리티 데이터 활용을 통한 골다공증 단계 다중 분류 시스템 개발: 합성곱 신경망 기반의 딥러닝 적용

하태준¹, 김희상², 강성욱³, 이두희⁴, 김우진^{5,6,10}, 문기원^{5,6,10}, 최현수^{4,7},
김정현⁸, 김 윤^{4,9}, 박소현^{13,*}, 박상원^{10,11,12,*}

¹더존비즈온 플랫폼사업부

²㈜포소드 기술연구소

³강원대학교병원 차세대정보산업실

⁴㈜지오비전

⁵강원대학교병원 내과

⁶강원대학교 의과대학 내과학교실

⁷서울과학기술대학교 컴퓨터공학과

⁸강원대학교 의과대학 비뇨의학과

⁹강원대학교 컴퓨터공학과

¹⁰강원대학교 의과대학 의료정보학과

¹¹강원대학교 의과학연구소

¹²㈜위뉴 데이터본부

¹³서울아산병원 영상의학과

요 약

골다공증은 전 세계적으로 주요한 건강 문제임에도 불구하고, 골절 발생 전까지 쉽게 발견되지 않는 단점을 가지고 있습니다. 본 연구에서는 골다공증 조기 발견 능력 향상을 위해, 복부 컴퓨터 단층 촬영(Computed Tomography, CT) 영상을 활용하여 정상-골감소증-골다공증으로 구분되는 골다공증 단계를 체계적으로 분류할 수 있는 딥러닝(Deep learning, DL) 시스템을 개발하였습니다. 총 3,012개의 조영제 향상 복부 CT 영상과 개별 환자의 이중 에너지 X선 흡수 계측법(Dual-Energy X-ray Absorptiometry, DXA)으로 얻은 T-점수를 활용하여 딥러닝 모델 개발을 수행하였습니다. 모든 딥러닝 모델은 비정형 이미지 데이터, 정형 인구 통계 정보 및 비정형 영상 데이터와 정형 데이터를 동시에 활용하는 다중 모달 방법에 각각 모델 구현을 실현하였으며, 모든 환자들은 T-점수를 통해 정상, 골감소증 및 골다공증 그룹으로 분류되었습니다.

가장 높은 정확도를 갖는 모델 우수성은 비정형-정형 결합 데이터 모델이 가장 우수하였으며, 수신자 조작 특성 곡선 아래 면적이 0.94와 정확도가 0.80를 제시하였습니다. 구현된 딥러닝 모델은 그라디언트 가중치 클래스 활성화 매핑(Gradient-weighted Class Activation Mapping, Grad-CAM)을 통해 해석되어 이미지 내에서 임상적으로 관련된 특징을 강조했고, 대퇴 경부가 골다공증을 통해 골절 발생이 높은 위험 부위임을 밝혔습니다. 이 연구는 DL이 임상 데이터에서 골다공증 단계를 정확하게 식별할 수 있음을 보여주며, 조기에 골다공증을 탐지하고 적절한 치료로 골절 위험을 줄일 수 있는 복부 컴퓨터 단층 촬영 영상의 잠재력을 제시할 수 있습니다.

중심단어: 인공지능, 골밀도, 골다공증, 임상 의사결정 지원 시스템, 딥러닝

연구자 정보 이력

	성명	소속	직위
(제1저자)	하태준	(주) 더존비즈온 플랫폼사업부	연구원
(공동저자)	김희상	(주) 포소드 기술연구소	연구원
	강성욱	강원대학교병원 차세대정보산업실	연구원
	이두희	(주) 지오비전	이사
	김우진	강원대학교 의과대학 내과학교실	교수
	문기원	강원대학교 의과대학 내과학교실	교수
	최현수	서울과학기술대학교 컴퓨터공학과	조교수
	김정현	강원대학교 의과대학 비뇨의학과	교수
	김 윤	강원대학교 컴퓨터공학과	교수
	(교신저자)	박소현	서울아산병원 영상의학과
(교신저자)	박상원	강원대학교 의과대학 의료정보학과	연구조교수