**Three-Dimensional Deep Learning Model for Predicting Proximal Femoral Pathologic Fracture Using Opportunistic Abdominopelvic CT in Patients With Advanced Cancer: A Preliminary Report**

Yong-Suk Lee¹, Min Wook Joo², Zhegao Piao 3, Yeonghyeon Gu3, Ilkyu Han4, Seung Han Shin5, Jewoo Lee², Minpyo Lee², Yang-Guk Chung5

*¹Department of Orthopaedic Surgery, Incheon St. Mary’s Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea*

*²Department of Orthopaedic Surgery, St. Vincent’s Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea*

*³Department of Artificial Intelligence and Data Science, Sejong University, Seoul, Korea*

*⁴Department of Orthopaedic Surgery, Seoul National University Hospital, Seoul, Korea.*

*5Department of Orthopaedic Surgery, Seoul St. Mary’s Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea*

**Introduction**

Pathologic fractures of the proximal femur are devastating complications in patients with advanced cancer. These fractures often result in, impaired mobility, prolonged hospitalization, and a marked decrease in quality of life, with some patients ultimately leading to death. Early surgical intervention can meaningfully reduce avoidable morbidity and mortality if high-risk lesions are accurately identified in advance. Traditional clinical scoring systems, such as Mirels’ criteria and the Harrington classification, have been widely used to guide clinical decision-making. However, these approaches are limited by modest sensitivity and specificity, subjective interpretation, and poor interobserver reproducibility, which restrict their clinical utility.

Recent advances in artificial intelligence, particularly deep learning, have demonstrated strong potential in analyzing medical images that are often overlooked by conventional methods. In particular, convolutional neural networks (CNNs) have shown promising results in musculoskeletal oncology for tasks involving lesion detection and risk stratification. Extending these methods to volumetric CT data may allow more comprehensive assessment of bone architecture and lesion morphology, thereby improving predictive accuracy and clinical applicability. While prior studies have shown that deep learning models applied to CT-derived images can predict fracture risk with reasonable accuracy, most approaches relied on two-dimensional projections that do not fully exploit the volumetric information contained in CT scans. A three-dimensional deep learning approach has the potential to provide more comprehensive assessment of bone architecture and lesion morphology, thereby improving predictive accuracy and clinical applicability.

Abdominopelvic CT scans are already routinely performed for staging and surveillance in patients with cancer. Despite their prevalence, however, valuable skeletal information within these scans remains overlooked and underutilized. Utilizing these existing scans for opportunistic risk prediction could provide a non-invasive, cost-effective, and widely accessible method to identify patients at impending risk of proximal femoral fracture. Building upon these considerations, the present study aimed to develop and validate a three-dimensional deep learning model, including a segmentation-based approach, to predict impending proximal femoral fractures in patients with advanced cancer.

**Methods**

A total of 2,933 abdominopelvic CT scans of 540 patients were collected from four tertiary academic hospitals. Since patients with bilateral proximal femoral metastatic lesions contributed two different cases from a single CT scan, 3,621 cases were finally included. (Fig.1 STROBE 이용석 작성) Among them, 1,956 and 1,665 were from male and female patients, respectively. The mean age of the cases was 63 ± 11 years. Computed tomography scans were assigned into fracture (F) or non-fracture (N) groups based on whether a pathologic fracture occurred within three months of the scan. All CT images were pre-processed by clipping using a bone window setting (level = 500, width = 2000) and normalizing voxel intensities to a 0–255 range, and processed into three-dimensional volumes. To standardize anatomical localization, a U-Net segmentation model was used to identify the anterior inferior iliac spine. From this landmark, 48(32) consecutive axial slices were extracted to create three-dimensional input volumes. (Fig.2 process 또는 framework관련 imaging - 박철호박사님) The training dataset, from a single institution, included 1,884 N- and 12 F-labelled scans, and the latter was augmented 20-fold through affine transformations. Internal validation used 472 N- and 12 augmented F-labelled scans. An external test set comprised 1,241 scans from the other institutions: 1,214 N- and 27 F-labeled scans. Four three-dimensional deep learning models were trained and evaluated: ResNet-18, ResNet-34, DenseNet-121, and Swin Transformer.

In addition, we constructed a ~~novel~~ UNet encoder classification model for segmental-based approach. The novel method consists of two main steps: (1) localizing the proximal femur through segmentation, and (2) classifying fracture status based on the features extracted from the UNet encoder. Proximal femur masks were generated from CT scans of 72 patients (43 fracture, 29 non-fracture) and used to train a three-dimensional UNet model (Depth=32, 300 epochs). Encoder weights from the trained UNet were transferred to a novel classification model (UNet-Encoder) with added fully connected layers, where only the final two encoder layers and classifier were fine-tuned. The classification model was trained on CT scans from Seoul St. Mary’s Hospital. (Fig.3 segmentation관련 imaging - 박철호박사님)

위 method에서segmentation 방법 기술이 맞는지요? (박철호 박사님)

**Results**

DenseNet-121 achieved the highest external test performance (accuracy 0.9315, sensitivity 0.8483, specificity 0.9703, precision 0.9315, F1 score 0.9315). ResNet-34 followed closely (accuracy 0.9242, sensitivity 0.7743, specificity 0.9942, precision 0.9242, F1 score 0.9242), showing better performance than ResNet-18 (accuracy 0.8782, sensitivity 0.8201, specificity 0.9053, precision 0.8782, F1 score 0.8782). Swin Transformer showed worst performance (accuracy 0.5244, sensitivity 0.3686, specificity 0.5972, precision 0.5244, F1 score 0.5244). In the segmentation-based model, the UNet-Encoder classifier achieved predictive performance comparable to the baseline three-dimensional deep learning models. While segmentation did not yield superior accuracy over DenseNet-121, it provided anatomically focused classification and confirmed that femur-targeted modeling enhances robustness and interpretability.

(Table 1 결과표 이용석 작성)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | AUROC | Accuracy | Sensitivity | Specificity | Precision | F1 score |
| Swin Transformer |  | 0.5244 | 0.3686 | 0.5972 | 0.5244 | 0.5244 |
| ResNet-18 |  | 0.8782 | 0.8201 | 0.9053 | 0.8782 | 0.8782 |
| ResNet-34 |  | 0.9242 | 0.7743 | 0.9942 | 0.9242 | 0.9242 |
| DenseNet-121 |  | 0.9315 | 0.8483 | 0.9703 | 0.9315 | 0.9315 |
| UNet-Encoder (segmentation) |  |  |  |  |  |  |

(Table 1 결과표 박철호 수정 – 이용석 교수님 검토 부탁드립니다.)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | AUROC (95% CI) | Accuracy (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) | Precision (95% CI) | F1 score (95% CI) |
| Swin Transformer | 0.4359± 0.0536  (0.3369 to 0.5407) | 0.5903± 0.0139 (0.5641 to 0.6181) | 0.2629± 0.0853 (0.1153 to 0.4348) | 0.5976± 0.014 (0.5705 to 0.6242) | 0.0143± 0.0052 (0.0058 to 0.0248) | 0.0271± 0.0097 (0.011 to 0.0466) |
| ResNet-18 | 0.692± 0.062  (0.5672 to 0.8076) | 0.8944± 0.009 (0.8759 to 0.9106) | 0.4045± 0.0966 (0.2174 to 0.6) | 0.9054± 0.0086 (0.8875 to 0.9216) | 0.0873± 0.0251 (0.0431 to 0.1395) | 0.1429± 0.0385 (0.073 to 0.2182) |
| ResNet-34 | 0.7584± 0.0477  (0.6632 to 0.8461) | 0.9766± 0.0043 (0.9678 to 0.9847) | 0.1833± 0.0736 (0.0556 to 0.3333) | 0.9943± 0.0021 (0.9901 to 0.9983) | 0.4203± 0.1445 (0.1667 to 0.7143) | 0.2502± 0.09 (0.0888 to 0.4255) |
| DenseNet-121 | 0.6487± 0.0712  (0.5049 to 0.7778) | 0.9572± 0.0059 (0.946 to 0.9678) | 0.3676± 0.0954 (0.1905 to 0.5518) | 0.9704± 0.0049 (0.9597 to 0.9795) | 0.2175± 0.0615 (0.1064 to 0.3421) | 0.2706± 0.069 (0.1408 to 0.4) |
| UNet-Encoder (segmentation) | 0.8585 ± 0.0453  (0.7656 to 0.9411) | 0.9485 ± 0.0061 (0.9363 to 0.9597) | 0. 6662± 0.0901 (0.4838 to 0.84) | 0.9547± 0.0059 (0.9426 to 0.9662) | 0.245± 0.0504 (0.1492 to 0.3388) | 0.3559± 0.0613 (0.2353 to 0.4673) |

이전 연구 table 표 예시

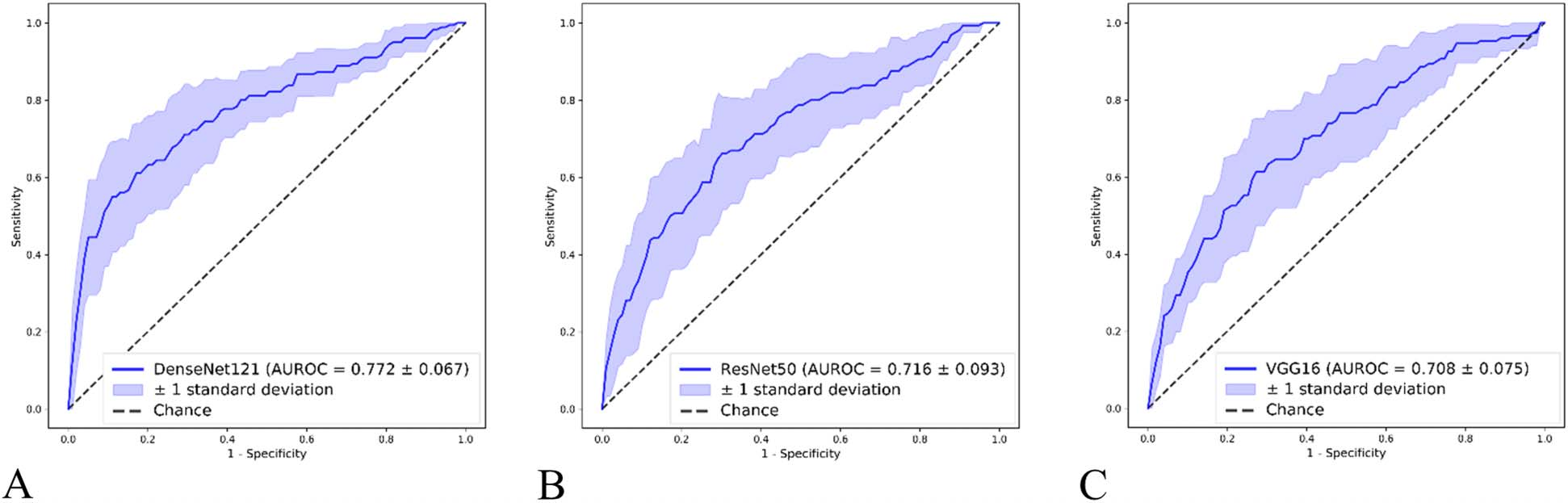
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | AUROC (95% CI) | Accuracy (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) | Precision (95% CI) | F1 score (95% CI) |
| VGG16 | 0.71 ± 0.08 (0.69 to 0.73) | 0.85 ± 0.02 (0.85 to 0.86) | 0.09 ± 0.07 (0.07 to 0.11) | 0.99 ± 0.02 (0.98 to 0.99) | 0.59 ± 0.37 (0.49 to 0.70) | 0.16 ± 0.09 (0.14 to 0.18) |
| ResNet50 | 0.72 ± 0.09 (0.69 to 0.74) | 0.84 ± 0.02 (0.83 to 0.84) | 0.14 ± 0.08 (0.12 to 0.16) | 0.98 ± 0.02 (0.97 to 0.98) | 0.61 ± 0.09 (0.58 to 0.63) | 0.23 ± 0.11 (0.20 to 0.26) |
| DenseNet121 | 0.77 ± 0.07 (0.75 to 0.79) | 0.85 ± 0.02 (0.84 to 0.85) | 0.22 ± 0.07 (0.20 to 0.24) | 0.98 ± 0.01 (0.98 to 0.99) | 0.72 ± 0.19 (0.67 to 0.77) | 0.34 ± 0.10 (0.31 to 0.37) |

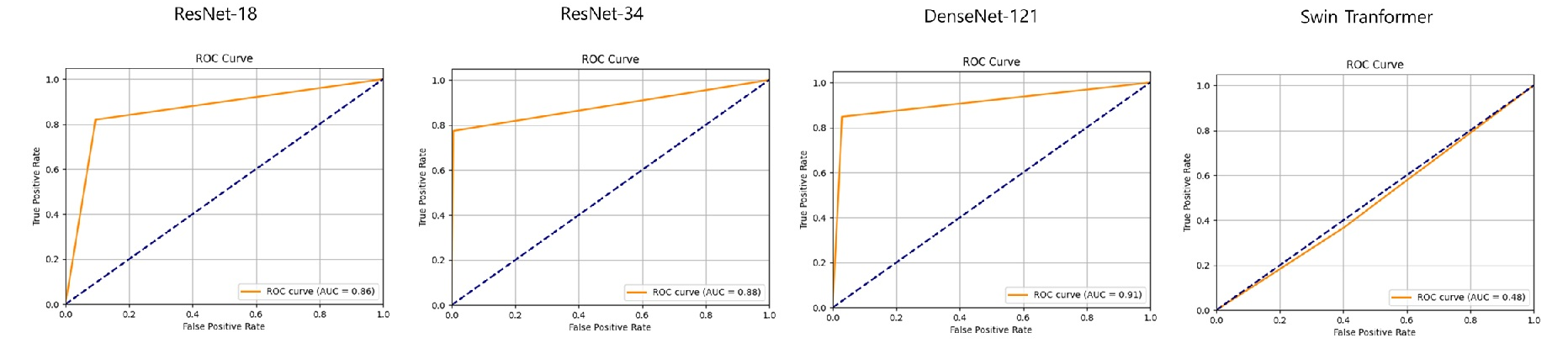
UNet Encoder모델에서 N과 F를 합친 결과 필요합니다 (박철호박사님)

Result 부분을 아래 처럼 작성하는 것이 필요합니다.

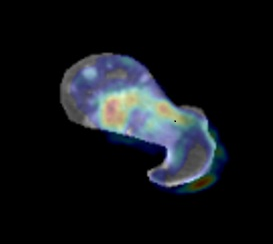
DenseNet121 showed better performance in identifying pathologic fractures than VGG16 and ResNet50 (Table 1); the area under the receiver operating characteristic curve for DenseNet121 was larger than those for VGG16 (0.77 6 0.07 [95% CI 0.75 to 0.79] versus 0.71 6 0.08 [95% CI 0.69 to 0.73]); p = 0.001) and ResNet50 (0.77 6 0.07 [95% CI 0.75 to 0.79] versus 0.72 6 0.09 [95% CI 0.69 to 0.74]; p = 0.001) (Fig. 4). Specifically, DenseNet121 scored the highest in sensitivity (0.22 6 0.07 [95% CI 0.20 to 0.24]), precision (0.72 6 0.19 [95% CI 0.67 to 0.77]), and F1 score (0.34 60.10 [95% CI 0.31 to 0.37]), and it focused accurately on the region with the expected pathologic fracture (Fig. 5)

AUROC curve 넣기 (이용석 작성) 예시 – 박철호박사님





Grad-CAM 넣기 (이용석 작성) – 박철호박사님



**Discussion**

In this study, we developed a three-dimensional deep learning model using abdominopelvic CT scans to predict the risk of pathologic fractures in the proximal femur and compared the performance of multiple architectures. DenseNet-121 demonstrated the most favorable and consistent performance, suggesting that convolutional neural network–based models are particularly effective in capturing the spatial continuity of bone structures and lesion characteristics. By contrast, the Swin Transformer exhibited poor performance, which may reflect the data-intensive nature of transformer-based models and their limited generalizability when trained on relatively small datasets. This finding underscores the importance of aligning model complexity with dataset size and clinical feasibility.

We also introduced a segmentation-based approach using the UNet-Encoder classifier, which provided comparable predictive performance to the baseline models while enhancing anatomical focus and interpretability. Although segmentation did not result in higher accuracy than DenseNet-121, it demonstrated practical advantages by confining model attention to the proximal femur and confirming the utility of both expert- and laboratory-generated segmentation masks. This approach may reduce the cost and effort required for large-scale dataset construction and facilitate broader clinical application.

DenseNet-121과 UNet-Encoder classifier 의 차이점 기술이 필요할까요? (구영현 교수님, 박철호 박사님)

From a clinical standpoint, our findings highlight the potential of opportunistic imaging. Abdominopelvic CT scans are routinely obtained for staging and surveillance in patients with cancer, and our results suggest that these existing scans can be repurposed to provide actionable insights into fracture risk without additional imaging, cost, or radiation exposure. Unlike conventional clinical scoring systems, which suffer from observer variability and limited reproducibility, deep learning–based models offer objective and reproducible risk stratification. Such tools may assist clinicians in identifying high-risk patients and guiding timely prophylactic interventions to prevent morbidity and mortality associated with completed fractures.

Several limitations should be acknowledged. The dataset was imbalanced, with relatively few fracture cases, which likely contributed to reduced sensitivity. The study population was limited to a single ethnic group, restricting generalizability, and clinical variables such as tumor histology and treatment history were not incorporated, which could further enhance predictive accuracy. Moreover, the segmentation-based approach requires additional preprocessing steps, which may limit immediate clinical adoption.

Future work should focus on prospective validation and the inclusion of larger, more diverse multicenter datasets. Integrating clinical and imaging data through ensemble learning may further improve predictive performance. In addition, the systematic use of explainable AI techniques could enhance transparency and foster trust among clinicians. With these advancements, three-dimensional deep learning models have the potential to become practical clinical tools for fracture risk prediction and decision-making in patients with advanced cancer.

**Conclusion**

This preliminary report demonstrated the potential feasibility and accuracy of a three-dimensional deep learning model using abdominopelvic CT scans routinely acquired for patients with advanced cancer for identifying imminent risk of proximal femoral pathologic fractures. Among tested models, DenseNet-121 seemed to show superior performance and consistency, while the UNet-Encoder segmentation-based approach ensured anatomical precision. This opportunistic imaging-based approach could eliminate the need for dedicated skeletal imaging, offering a practical, non-invasive, and cost-effective method for pathologic fracture risk stratification. To improve the model’s robustness and generalizability, an extensive dataset will be derived from the Clinical Data Warehouse. The code for our model is publicly available online at ??? 🡪 저희 모델도 코드를 online으로 올릴수 있는지요? (박철호박사님)