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

Min Wook Joo, Yong-Suk Lee, Zhegao Piao, Younghyun Koo, Ilkyu Han, Jewoo Lee, Minpyo Lee, Yang-Guk Chung

INTRODUCTION: Pathologic fractures are devastating events in patients with advanced cancer. They often result in impaired mobility, prolonged hospitalization, and a serious decline in quality of life, with some patients ultimately dying. Early surgical intervention can reduce avoidable morbidity and mortality if high-risk lesions are accurately identified in advance. However, conventional clinical scoring tools such as Mirels' scoring system exhibit limited reproducibility and insufficient predictive power. Meanwhile, abdominopelvic CT scans are frequently incorporated into routine surveillance or systemic evaluation in cancer patients, but valuable skeletal and other imaging-derived features within them are typically overlooked and underutilized. Recognizing this opportunity, we aimed to develop and validate a three-dimensional deep learning model using these opportunistic scans to predict impending proximal femoral fractures within a timeframe relevant for patient management.

METHODS: A total of 2,933 abdominopelvic CT scans from 540 patients with proximal femur metastases identified on current or prior imaging were collected from four tertiary academic hospitals between January 2011 and December 2023. Since patients with bilateral proximal femoral metastatic lesions yielded two different cases from a single CT scan, 3,621 cases were finally included. Among them, 1,956 and 1,665 were from male and female patients, respectively. The mean age of the cases was 63 ± 11 years. The CT scans were assigned to fracture (F) or non-fracture (N) groups based on whether a pathologic fracture occurred within three months of the scan: 3570 N- and 51 F-labelled. All CT images were pre-processed by clipping using a bone window setting (level = 500, width = 2000) and normalizing to a 0–255 voxel intensity range, and processed into three-dimensional volumes. To standardize anatomical localization, a U-Net segmentation model identified the anterior inferior iliac spine. From this landmark, 32 consecutive axial slices were extracted to create three-dimensional input volumes. 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-labelled scans. Four three-dimensional deep learning models were trained and evaluated: ResNet-18, ResNet-34, DenseNet-121, and Swin Transformer. Model performance was assessed using accuracy, sensitivity, specificity, precision, and F1 score.

RESULTS:

DenseNet-121 showed the highest external test performance (accuracy 0.9315, sensitivity 0.8483, specificity 0.9703, precision 0.9315, and F1 score 0.9315). ResNet-34 followed closely (accuracy 0.9242, sensitivity 0.7743, specificity 0.9942, precision 0.9242, and F1 score 0.9242), showing better performance than ResNet-18 (accuracy 0.8782, sensitivity 0.8201, specificity 0.9053, precision 0.8782, and F1 score 0.8782). Swin Transformer showed lower performance (accuracy 0.5244, sensitivity 0.3686, specificity 0.5972, precision 0.5244, and F1 score 0.5244).

DISCUSSION AND CONCLUSION: This preliminary report demonstrates the potential feasibility and accuracy of a three-dimensional deep learning model using abdominopelvic CT scans routinely acquired for patients with advanced cancer, for evaluating imminent risk of proximal femoral pathologic fractures. Among the tested models, DenseNet-121 seemed to show superior performance and consistency. This opportunistic imaging-based approach could eliminate the need for dedicated skeletal imaging, offering a practical, non-invasive, and cost-effective method for early pathologic fracture risk stratification. To improve the model's robustness and generalizability, an extensive dataset will be derived and secured from the Clinical Data Warehouse.