



An Automated Deep Learning Method for Tile AO/OTA Pelvic Fracture Severity Grading from Trauma whole-Body CT

David Dreizin¹ · Florian Goldmann² · Christina LeBedis³ · Alexis Boscak¹ · Matthew Dattwyler¹ · Uttam Bodanapally¹ · Guang Li⁴ · Stephan Anderson³ · Andreas Maier² · Mathias Unberath⁵

Received: 29 April 2020 / Revised: 14 October 2020 / Accepted: 10 November 2020 / Published online: 21 January 2021
© Society for Imaging Informatics in Medicine 2021

Abstract

Admission trauma whole-body CT is routinely employed as a first-line diagnostic tool for characterizing pelvic fracture severity. Tile AO/OTA grade based on the presence or absence of rotational and translational instability corresponds with need for interventions including massive transfusion and angioembolization. An automated method could be highly beneficial for point of care triage in this critical time-sensitive setting. A dataset of 373 trauma whole-body CTs collected from two busy level 1 trauma centers with consensus Tile AO/OTA grading by three trauma radiologists was used to train and test a triplanar parallel concatenated network incorporating orthogonal full-thickness multiplanar reformat (MPR) views as input with a ResNeXt-50 backbone. Input pelvic images were first derived using an automated registration and cropping technique. Performance of the network for classification of rotational and translational instability was compared with that of (1) an analogous triplanar architecture incorporating an LSTM RNN network, (2) a previously described 3D autoencoder-based method, and (3) grading by a fourth independent blinded radiologist with trauma expertise. Confusion matrix results were derived, anchored to peak Matthews correlation coefficient (MCC). Associations with clinical outcomes were determined using Fisher's exact test. The triplanar parallel concatenated method had the highest accuracies for discriminating translational and rotational instability (85% and 74%, respectively), with specificity, recall, and F1 score of 93.4%, 56.5%, and 0.63 for translational instability and 71.7%, 75.7%, and 0.77 for rotational instability. Accuracy of this method was equivalent to the single radiologist read for rotational instability (74.0% versus 76.7%, $p = 0.40$), but significantly higher for translational instability (85.0% versus 75.1, $p = 0.0007$). Mean inference time was < 0.1 s per test image. Translational instability determined with this method was associated with need for angioembolization and massive transfusion ($p = 0.002$ – 0.008). Saliency maps demonstrated that the network focused on the sacroiliac complex and pubic symphysis, in keeping with the AO/OTA grading paradigm. A multiview concatenated deep network leveraging 3D information from orthogonal thick-MPR images predicted rotationally and translationally unstable pelvic fractures with accuracy comparable to an independent reader with trauma radiology expertise. Model output demonstrated significant association with key clinical outcomes.

Keywords Pelvic fracture · Pelvic ring disruption · Pelvic instability · Tile classification · Deep learning · Convolutional neural network

✉ David Dreizin
daviddreizin@gmail.com

Florian Goldmann
Florian.goldmann@fau.de

Guang Li
guangli@umm.edu

Andreas Maier
Andreas.maier@fau.de

² Friedrich-Alexander University, Schloßplatz, Erlangen, Germany

³ Department of Radiology, Boston Medical Center, Boston University School of Medicine, Baltimore, MD, USA

⁴ Department of Diagnostic Radiology and Nuclear Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

⁵ Department of Computer Science, Malone Center for Engineering in Healthcare, Johns Hopkins University, Baltimore, MD, USA

¹ Emergency and Trauma Imaging, Department of Diagnostic Radiology and Nuclear Medicine, R Adams Cowley Shock Trauma Center, University of Maryland School of Medicine, Baltimore, MD, USA

Abbreviations

LSTM RNN	Long short-term memory recurrent neural network
TP	True positive
TN	True negative
FP	False positive
FN	False negative
TPR	True positive rate
TNR	True negative rate
PPV	Positive predictive value
NPV	Negative predictive value
FNR	False negative rate
FPR	False positive rate
FDR	False discovery rate
FOR	False omission rate
DOR	Diagnostic odds ratio

Introduction

Fractures of the pelvic ring are encountered in approximately 10% of patients admitted to level 1 trauma centers with blunt injury mechanisms [1] and 20% of patients with high-velocity polytrauma [2]. Pelvic fractures are potentially lethal due to the risk of exsanguination [3, 4]. The leading cause of death in the first 6 h after pelvic fracture is abdominopelvic hemorrhage [5]. In-hospital mortality is as high as 20–50% in patients with mechanically unstable pelvic fractures and hemodynamic instability [6–10]. Widespread deployment of trauma bay-adjacent CT scanners and improvements

in resuscitation techniques have led to the routine use of admission trauma whole-body CT as the first-line diagnostic tool for characterizing pelvic fracture severity in all but a small subset of patients that require immediate surgery for severe refractory hemodynamic collapse [2, 6, 11–13]. A recent study across 11 major trauma centers reported use of admission trauma CT in 85% of patients with pelvic fractures admitted in shock [6].

Basic Anatomy and Biomechanics of Pelvic Fractures

To understand the rationale for algorithm development in this work, a basic understanding of pelvic anatomy, biomechanics, and instability grading is necessary. The pelvis is a complex ring structure comprised of the sacrum and two innominate bones that articulate at the sacroiliac joints posteriorly and pubic symphysis anteriorly [2, 12]. Because of its multipart nature, the bony ring has little inherent stability [9]. Stability is derived almost entirely from the strong soft-tissue envelope of primary and secondary stabilizing ligaments about the posterior sacroiliac complex that bind the sacrum and innominate bones together [2]. Pelvic stability requires an intact posterior sacroiliac complex [9]. Disruption of the posterior ligaments will result in a variety of abnormal bony relationships and unilateral or bilateral rotational or translational anatomic distortions of the pelvis [9]. An accounting of pelvic fracture lines is much less important for pelvic fracture severity grading than global assessment of pelvic distortions. Some consider “pelvic ring disruption” a more apt term than “pelvic fracture” for these injuries [2].

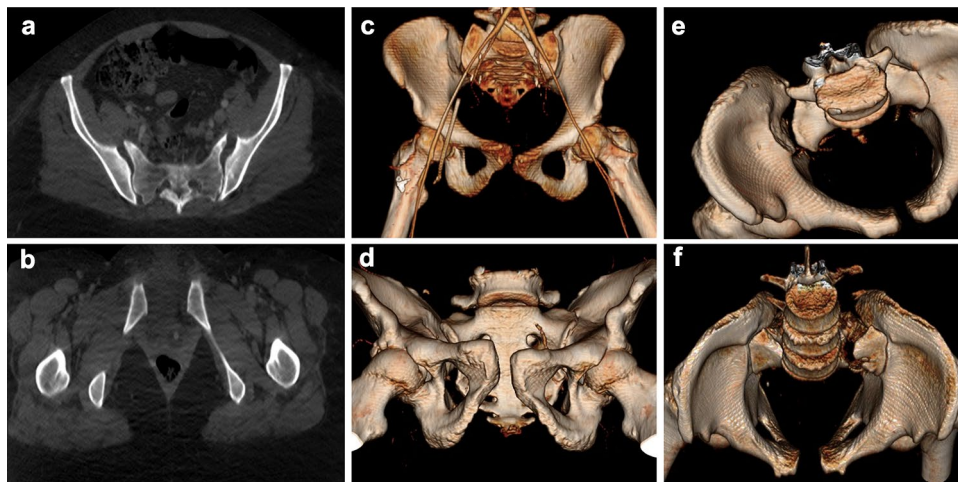


Fig. 1 **a** Illustrates anteriorly divergent sacroiliac diastasis hinged about an intact posterior SI ligament. Rotational but not translational instability is the hallmark of a Type B Tile AO/OTA injury. This is associated with pubic symphysis widening in **b** (same patient). Pubic symphysis widening can be seen in both Type B and C injuries and is not as discriminative as the bony relationships about the SI joint. **c–f** are Type

C injuries. In **c** and **d**, there is vertical translational displacement of the right innominate bone with respect to the sacrum. **e** Illustrates parallel widening of the SI joint characteristic of posterior SI ligament disruption. **f** Illustrates a variant where the juxta-articular sacrum is widely diastatic. In this instance the posterior SI ligament is intact, but functionally incompetent, with posterior translation of the left hemipelvis

Grading of Instability

There are several competing pelvic fracture classification systems for grading the spectrum of instability [9, 14, 15]. The Orthopedic Trauma Association (OTA) and AO Foundation have adopted the imaging-based Tile classification system for pelvic instability grading [12, 16]. The Tile AO/OTA comprehensive classification system differentiates fractures into three major first-order classes based primarily on the integrity of the posterior sacroiliac (SI) complex as it appears on CT [17, 18]. In all patients and permutations of injury, the integrity of the posterior sacroiliac complex (the SI joints and the juxta-articular sacrum and ilium) remains the most critical element of instability grading [9]. Features of instability and Tile AO/OTA grade are illustrated in Fig. 1. Type A fractures are stable as the SI joint is effectively intact. Type B fractures are rotationally unstable but translationally stable—the weaker anterior SI ligament is compromised and the affected hemipelvis can be either internally rotated (from lateral compression mechanism), or externally rotated (from anteroposterior compression) in the axial plane and anteriorly divergent widening of the SI joint giving a classic “open book” appearance. The intact fibrous posterior sacroiliac ligament acts as a hinge that prevents any translation in the vertical or AP direction. Type C injuries do not occur without massive force transmission that can overcome the structural capacity of the posterior sacroiliac ligament. Type C injuries are characterized by complete translational instability and result in vertical and AP translational distortions and parallel widening about the SI joint [2, 9, 12, 17, 19]. As a ring structure, an unstable pelvis is typically disrupted anteriorly as well [9, 12, 20], manifesting as pubic symphysis diastasis in externally rotated injuries, and pubic body override in severely internally rotated injuries [9, 21]. Pelvic ring disruptions may involve either hemipelvis and are often bilateral [17]. Pelvic binders can reduce the pelvis and have been shown to hinder discrimination of rotational and translational instability [8, 22]. Various fracture permutations of the four pubic rami are quite common, even in stable fracture patterns and have limited utility for severity grading [9]. In published series from major trauma centers, Type A and B fractures make up 70–80% of all pelvic ring disruptions and Type C, 20–30% [9, 23, 24].

The Tile AO/OTA fracture grade and outcomes—Tile C fractures in particular—have been shown to be independently predictive of mortality, transfusion requirement, and arterial hemorrhage requiring hemostatic intervention [14, 23, 25–28]. However, grading pelvic fractures and distortions is a challenging non-trivial 3D problem [22]. Mental integration of findings on orthogonal multiplanar reformats into a single 3D image is nontrivial, and manipulation of

volumetric images using post-processing software can be difficult in a pressured time-sensitive setting [8, 19]. The concepts of instability become intuitive with experience and trauma subspecialization but are difficult for less experienced readers or those who do not regularly encounter high energy pelvic ring disruptions in their practice. An explainable automated method that accurately and objectively rules in rotational or translational instability learned from consensus patient-level annotation by experts could improve outcome by reducing treatment delays and would be highly beneficial for point of care triage in this critical time sensitive setting.

Tile AO/OTA Classification Algorithm Pre-requisites

Numerous works have described relatively high accuracies of convolutional neural networks for detecting orthopedic fractures in different body regions (typically in the range of 80–90%), but automatic grading of fracture severity has remained an elusive task as very high-level abstract outputs must be distilled from a large number of parameters and 3D contextual data. A successful algorithm for Tile AO/OTA grading must have reasonable performance and association with outcomes that is similar to grading by human readers. Traditional 2D slice-based methods are suboptimal given the degree of global contextual information needed for classification. On the other hand, while 3D convolutional neural networks are highly advantageous, they are computationally expensive. Given the high co-occurrence between pelvic fracture grade and prevalence of multiple injuries in other body regions [11, 13, 23] as well as the increased parameter count compared with 2D convolutional neural networks, more data is needed for comparable performance to reduce overfitting behavior. Traumatic brain injury, abdominal organ injuries with hemoperitoneum, spine fractures, and extremity fractures are commonly encountered together with unstable pelvic fractures [17, 29–33]. Pelvic fractures should certainly be considered in the context of polytrauma management [9], but it is easy to imagine how overfitting to injuries that correlate with Tile classification, but are unrelated to the classification task itself, can occur if the learning task is not modeled and supervised accordingly with this risk in mind.

The purpose of our study was to determine the feasibility of and compare performance between automated multistage machine learning methods that first partition the pelvic region from a whole-body CT using template matching and then use one of three techniques to efficiently glean global 3D information for Tile AO/OTA classification: (i) a parallel concatenated multi-view deep learning method, (ii) a recurrent neural network (RNN)-based multiview deep learning method, and (iii) a 3D deep convolutional neural network classifier.

Materials and Methods

Dataset and Study Population

The work was IRB-approved with waiver of informed consent and included a dataset of abdominopelvic CT scans, routinely performed as part of a contrast-enhanced whole-body trauma CT, in 373 adult patients (age ≥ 18 years) with bleeding pelvic fractures from two major level I trauma centers. Studies were performed on 40-, 64-, and dual source 128-section CT scanners and archived at 1.25–3-mm section thickness. Those patients who had already undergone damage control laparotomy or angioembolization prior to CT were excluded. No studies represented follow-up imaging in the same patient. Patient-level annotation was performed using first-order Tile AO/OTA grading (A—stable, B—rotationally unstable, or C—translationally unstable) by three trauma-subspecialized radiologists at one of the level I trauma centers with a best-two-out-of-three consensus approach with arbitration of disagreement by the senior-most reader. Blinded interpretation for comparison with algorithm performance was undertaken by a fourth radiologist with trauma expertise at the second level I trauma center.

Deep Learning Method and Neural Network Architecture: Prior Art and Rationale

Evolution of Multiview Approaches for Orthopedic Fracture Detection: Plain Radiographs

There is a small but growing body of literature reporting results of deep learning methods for fracture detection using plain radiographs of different anatomic regions including the proximal femur, elbow, wrist, and ankle with deep networks [34–41]. There has been growing emphasis on explainability, and this is currently realized through object detection methods and saliency maps. Early fracture detection methods performed binomial classification of a single view as positive or negative despite routine acquisition of two or more views in clinical practice [34–37]. At the time of writing, we are aware of three groups that have implemented approaches that consolidate output from multiple views [38, 40, 41]. Kitamura et al. used an ensemble method with majority voting based on three plain radiographic views of the ankle [40]. Thian et al. [38] determined per-patient accuracy of region-proposal networks applied to AP and lateral views of the wrist, and Rayan et al. [41] combined a convolutional neural network with a recurrent a neural network (a long short-term memory (LSTM) network) for detection of fractures

on multiple views of pediatric elbows. The use of bounding boxes in the former and saliency maps in the latter provided location information, making the networks more transparent and believable.

Convolutional Neural Networks for Fracture Severity Grading

There are few examples of CNNs being implemented for classification of clinically relevant injury grade. Chung et al. implemented a ResNet architecture using one shoulder radiograph image per patient to classify proximal humerus fractures according to Neer's classification with 65–86% accuracy [42]. We have previously reported preliminary results of Tile grading from CT scans using a 3D VGG-based autoencoder to reduce dimensionality and improve efficiency and utilization of global contextual information [43], however overfitting to pelvic soft-tissue findings including pelvic hematoma [44], another independent predictor of bleeding-related outcomes but one that does not contribute to Tile grading likely adversely affected our results.

Work on pelvic fractures has primarily been relegated to fracture detection on plain radiographs, 2D CT, and 3D volumetric images using handcrafted and rule-based approaches including active shape models, discrete wavelet transform, and graph cut methods in small cohorts [45–49]. Unlike these methods, convolutional neural networks deal well with variations in patient size, anatomy, and complexity of the fracture including substantial pelvic distortions. Wang et al. developed a weakly supervised cascaded coarse-to-fine method for pelvic fracture detection on plain radiographs using a deep CNN backbone to produce a fracture probability map, followed by a second network that performs localized analyses in the high probability zones to detect fractures [39]. Unlike fracture detection, instability grade in the pelvis is more of a global than a local problem. Fracture lines are typically seen in all three Tile AO/OTA grades, and detection of fracture lines on plain radiographs is of much less relevance for severity grading than abnormal spatial relationships between the sacrum and innominate bones on CT. The high dimensionality of CT data, the non-localized abstract nature of pelvic distortions in three dimensions, and large volumes of whole-body CT datasets pose unique challenges to the classification of pelvic fractures in polytrauma patients. We believed that a multiview approach using full thickness multiplanar reformats of the cropped pelvis offered a promising and computationally efficient solution for Tile grading since each view offers valuable and complementary information regarding rotation or translation of the pelvic ring.

A Concatenated Triplanar Thick-MPR Network for Tile AO/OTA Grading

We are not aware of methods that have concatenated multiple 2D volumetric views from 3D CT datasets for fracture classification. Assessment of pelvic fracture stability is inherently a 3D problem that requires a global approach. However, recent work in natural image processing has shown that 3D representations can be simulated using a compilation of 2D views with similar accuracy under current hardware constraints to native 3D shape classifiers such as point clouds or polygonal meshes that are prone to overfitting from high dimensionality [50]. Su et al. used parallel concatenated networks that take a limited number of 2D views from different angles as inputs [50]. While recognition rates increase with the number of views provided, relatively few are needed to draw effective inference about 3D shapes [50].

Orthogonal full-thickness MPRs serve to maximize representative 3D spatial information in a single image through in-plane aggregation of information over multiple slices [51]. The use of bone windowing is meant to prevent the network from focusing on irrelevant soft tissue information. SI joint translation or widening and pubic symphysis diastasis are well characterized on volumetric CT images reconstructed in three orthogonal planes [8]. We reasoned that three orthogonal bone-window thick-MPR 2D representations trained jointly with a view-pooling fully connected layer could be used to compile the 3D information necessary to discriminate pelvic fractures by Tile AO/OTA grade. The full-thickness bone-window multiplanar reformats in three canonical orientations (axial, coronal, and sagittal planes) are passed independently down each deep network while the fully connected layer ensures that model updates for each view are based on synthesized information from all three views.

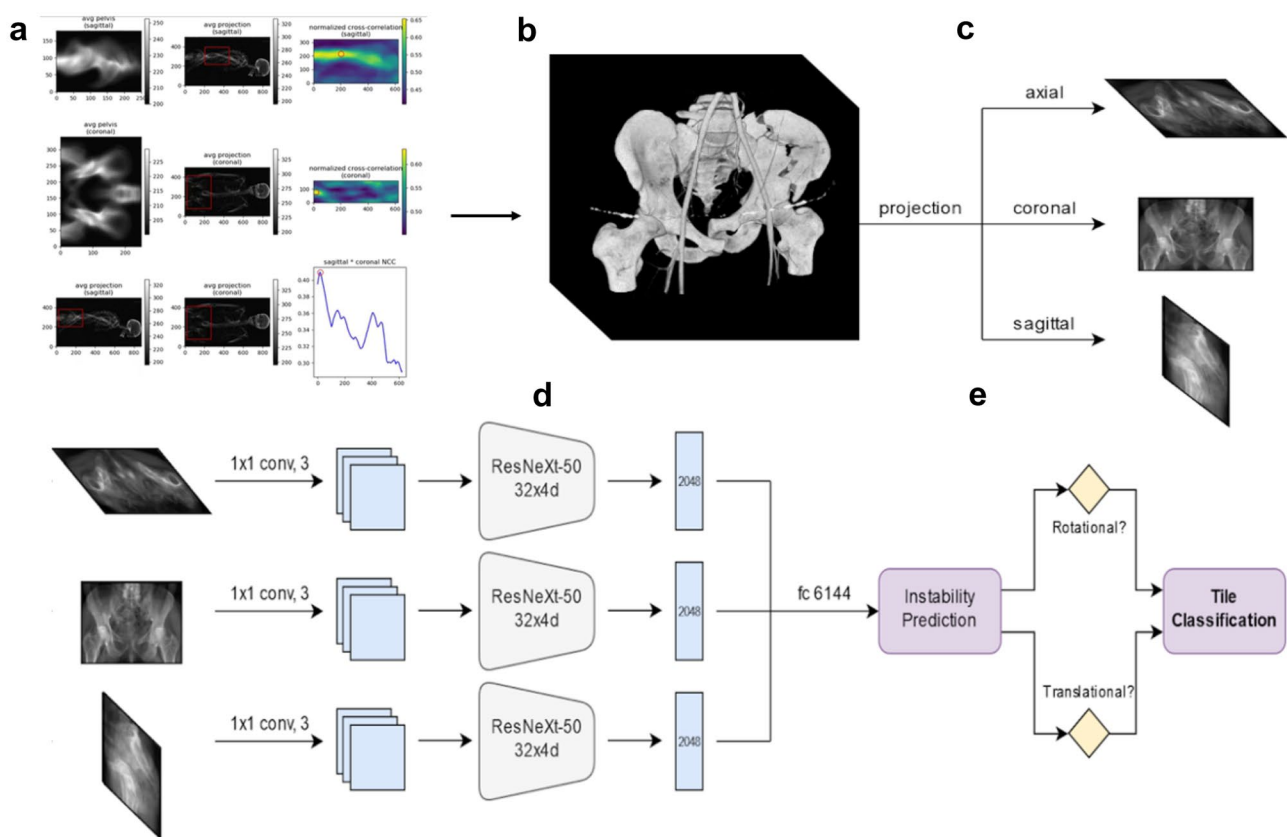


Fig. 2 Flow diagram of automated deep learning method for instability prediction and Tile AO/OTA grading. **a** Automated partitioning of the pelvis from the whole-body CT. A bounding cube is created around the pelvis following NCC-based registration, and **b** the pelvic region is cropped. **c** Full thickness bone-window MPRs are generated from the cropped pelvic CT in the three orthogonal planes. **d** These are passed down three parallel ResNeXt-50 networks which are joined by

a fully connected “view-pooling” layer that synthesizes information from all three views during discriminative learning and inference. **e** The network predicts whether rotational or translational instability are present or absent. The Tile AO/OTA grade is then determined. If there is no instability of either form, the fracture is Type A. If rotational instability is present but not vertical instability, the fracture is Type B. If translational instability is present, the fracture is type C

Template Matching for Registration and Cropping of the Pelvis

We used a template-matching procedure with normalized cross-correlation (NCC) as an efficient and robust registration method for matching correspondence between subject whole-body or torso CTs and pelvic template image patches within a sliding window of subregions from coronal and sagittal thick MPR images of entire CT volumes [52]. In natural image processing, NCC is a standard and commonly used metric for evaluating the degree of similarity between two image patches [52, 53]. With this technique, the greatest weight is placed on image patches that are matched with the greatest certainty [52]. The calculated NCC values can be plotted to create an intuitive, map-like display to visualize the areas of peak similarity used to partition the pelvis from the remainder of the study which is cropped out. A flow diagram illustrating our parallel concatenated method beginning with template matching using NCC is shown in Fig. 2.

Implementation Details

All CT datasets were de-identified and converted to NifTI format. Each scan was rescaled to a voxel size of $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$

using linear interpolation and re-oriented into a default RAS+ orientation. The intensities were clamped by applying a bone window (Window width: 1300 HU level: 650 HU), followed by z-score normalization to achieve zero mean and a standard deviation of 1.

As our deep network backbone, we used ResNeXt-50. ResNeXt is a state-of-the-art deep 2D CNN which has shown improvement over ResNet101 and 152. The ResNet architecture allows much greater depth by using residual shortcut connections to bypass activation function layers, thereby alleviating the problem of vanishing and exploding gradients. However, increasing capacity through depth (number of layers) and width (number of nodes) has diminishing returns. ResNeXt improves performance without increasing complexity by branching and then aggregating data into and from multiple parallel low-dimensional building blocks of identical depth [54]. A comparison between ResNet and ResNeXt is shown in Fig. 3. The final convolutional outputs of the three ResNeXt networks (one for each orthogonal thick-MPR image) were concatenated, and class probabilities were determined by using a fully connected layer preceding a softmax layer. The algorithm determines whether there is rotational or translational instability by application of an individual threshold anchored to the Matthews correlation coefficient [55, 56]. If there is neither, the injury is Tile A.

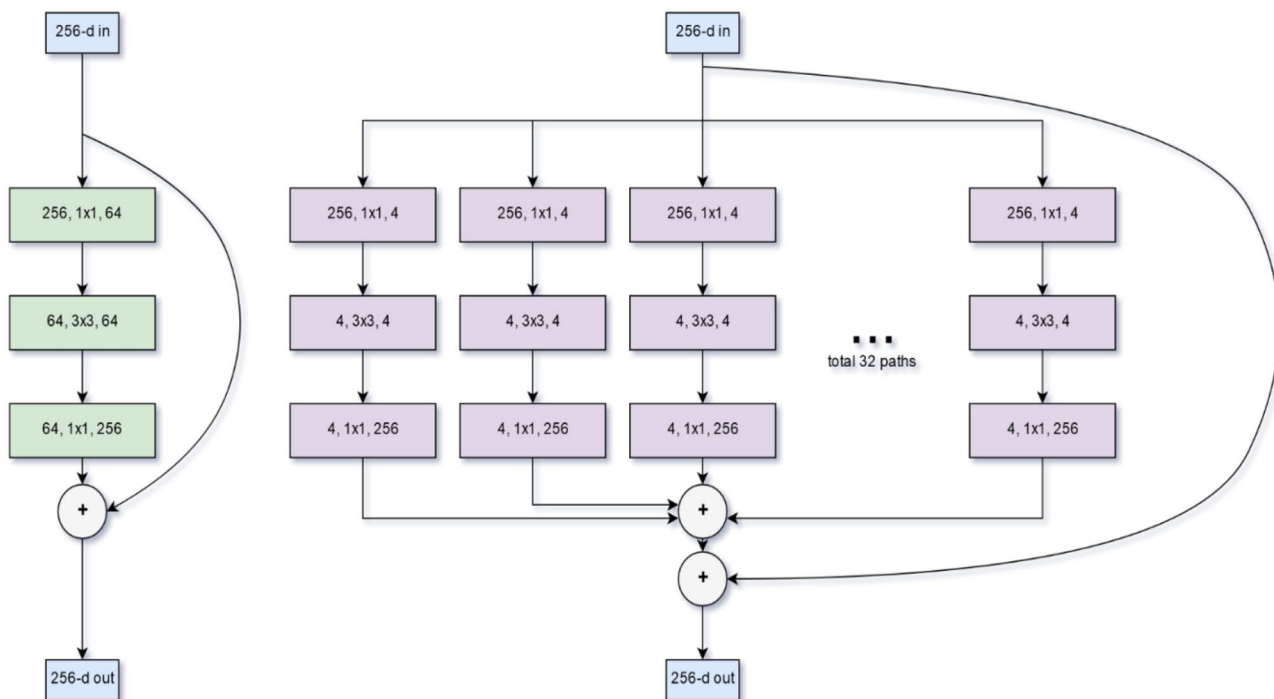


Fig. 3 Comparison between ResNet and ResNeXt backbone building blocks. Figure modified from Xie et al. (54). ResNet is shown left, ResNeXt shown right. Both use shortcut residual connections that allow very deep networks by alleviating vanishing and exploding gradients. ResNeXt (right) uses multiple reduced dimensionality blocks

in a split-transform-merge strategy. The parallelization of building blocks is referred to as cardinality. The method improved performance over ResNet at ILSVRC 2016 and achieved second place for classification tasks

If there is only rotational instability but no translational instability, the injury is Tile B, and if there is translational instability, the injury is Tile C.

We performed five-fold cross validation to assess the algorithm's performance and generalization capabilities. In each fold, 80% of the data was used on de novo fracture classification models pretrained using the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [57], and 20% of the data was used for validation in five non-overlapping combinations. Model convergence was monitored using a cross-entropy loss function. AdamW optimizer was used in combination with a cyclic learning rate for gradient descent optimization. Models were trained on a Quadro RTX 8000 GPU machine with a minibatch size of 42, and a cyclic learning rate between $1E-5$ and $5E-5$ for 100 epochs with a weight decay of $1E-2$. A total of 44 epochs were required to achieve convergence. Random flip, additive Gaussian noise, random crop, and random rotations were used for data augmentation. The network was implemented using the Pytorch deep learning platform (version 1.3.1).

We generated class activation (saliency) maps that display a heat map with higher intensities corresponding with the parts of an image most discriminative for the network [58]. Saliency maps for patients with rotational and translational instability

showed maximum attention at injured sacroiliac complexes and pubic symphyses.

Comparison Experiments

Triplanar ResNeXt Combined with an LSTM Network

We combined ResNeXt with a convolutional LSTM recurrent neural network (RNN) in a fashion analogous to the work of Rayan et al. [41] for multiview detection of pediatric elbow fractures. A block-flow diagram is shown in Fig. 4. LSTM uses temporal memory to store salient information learned from each sequentially analyzed view as it learns from the next. This serves as another method to aggregate information from all three processed orthogonal thick-MPR views in a unified way during discriminative learning. A batch size of 42 was again employed, and all other hyperparameters were also kept constant.

3D Autoencoder-Based Method

3D convolutional neural networks are potentially highly advantageous but very computationally expensive.

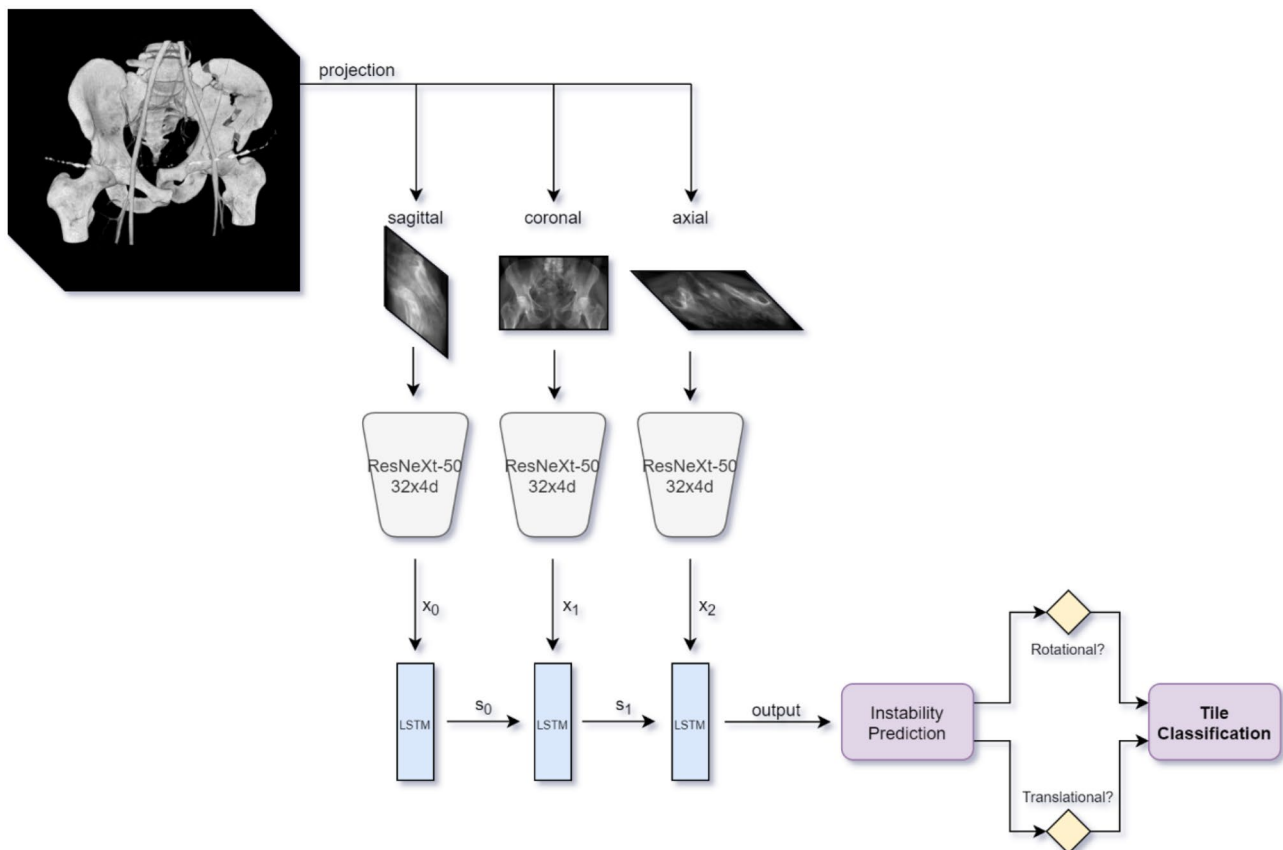


Fig. 4 Block flow diagram of recurrent neural network using ResNext-50 backbone. We implemented LSTM cells with three layers and a hidden size of 2048

Auto-encoders reduce dimensionality, regularize the latent space, and increase descriptive capacity in an efficient manner. Prior work on vertebral body fractures used autoencoders and point cloud reconstructions for efficient detection of vertebral body fractures on CT [59]. The network consisted of a volumetric CNN with residual connections, in this case based on ResNet-50. Volumetric ResNets have previously been implemented for segmentation tasks [60]. To our knowledge, 3D autoencoders have not been used to classify fracture severity. For classifying pelvic fractures using CT volumes, we employed a 3D ResNet-50-based architecture. Visual confirmation that the dimensionally reduced latent space retained a meaningful representation of the bony pelvis and was not overfitting to unrelated objects in the CT was achieved using a decoder and root mean squared error method during training. The classifier predicts rotational and translational instability from the generated latent space representation.

Statistical and Data Analysis

We assessed sensitivity (recall), specificity, positive predictive value (precision), negative predictive value, false omission and discovery rates, F1 score, and accuracy, anchored to the Matthews correlation coefficient for each network, and results were compared. Performance metrics were compared for statistical significance using McNemar's test for paired proportions. We then assessed for significant differences between automated Tile grades in terms of mortality, massive transfusion (≥ 10 U of packed red cells in 24 h or ≥ 4 U packed red cells in 4 h) and decision to perform angioembolization for arterial hemorrhage using the Fisher exact test.

Results

Exactly 159 patients in the dataset were graded by radiologist consensus as Tile A, 129 were Tile B, and 85 were Tile C. Ninety patients had pelvic binders in place. The distribution is representative of previous reports including Tile's original series [17]. Median patient age was 47, and median injury severity score (ISS) was 24.

Template Matching Method

The normalized cross-correlation template matching method was very robust, resulting in successful partitioning of the pelvic region in all 373 cases.

Deep Learning Method Performance

Confusion matrix results are presented for the three networks from highest to lowest model accuracy in Table 1. Accuracy of the additional blinded reader is also included.

Table 1 Confusion matrix results are provided for discrimination of rotational and translational instability by the three algorithms (triplanar concatenated, LSTM-based recurrent RNN, and 3D encoder method) and the independent radiologist reader

Model	Class	Confusion matrix results														
		TP	TN	FP	FN	Sensitivity (recall/TPR)	Specificity (1NR)	Precision (PPV)	NPV	Miss rate (FNR)	Fall-out rate (FPR)	FDR	FOR	Accuracy	F1	DOR
Triplanar parallel concatenated ResNet50	Rotational instability	162	114	45	52	75.7%	71.7%	78.3%	68.7%	24.3%	28.3%	21.7%	31.3%	74.0%	0.77	7.9
	Translational instability	48	269	19	37	56.5%	93.4%	71.6%	87.9%	43.5%	6.6%	28.4%	12.1%	85.0%	0.63	18.4
ResNet50+LSTM-based RNN	Rotational instability	137	120	39	77	64.0%	75.5%	77.8%	60.9%	36.0%	24.5%	22.2%	39.1%	68.9%	0.70	5.5
	Translational instability	38	278	10	47	44.7%	96.5%	79.2%	85.5%	55.3%	3.5%	20.8%	14.5%	84.7%	0.57	22.5
3D-ResNet50 encoder method	Rotational instability	61	147	12	153	28.5%	92.5%	83.6%	49.0%	71.5%	7.5%	16.4%	51.0%	55.8%	0.43	4.9
	Translational instability	32	273	15	53	37.6%	94.8%	68.1%	83.7%	62.4%	5.2%	31.9%	16.3%	81.8%	0.48	11.0
Independent radiologist reading	Rotational instability	174	112	47	40	81.3%	70.4%	78.7%	73.7%	18.7%	29.6%	21.3%	26.3%	76.7%	0.80	10.4
	Translational instability	68	212	76	17	80.0%	73.6%	47.2%	92.6%	20.0%	26.4%	52.8%	7.4%	75.1%	0.59	11.2

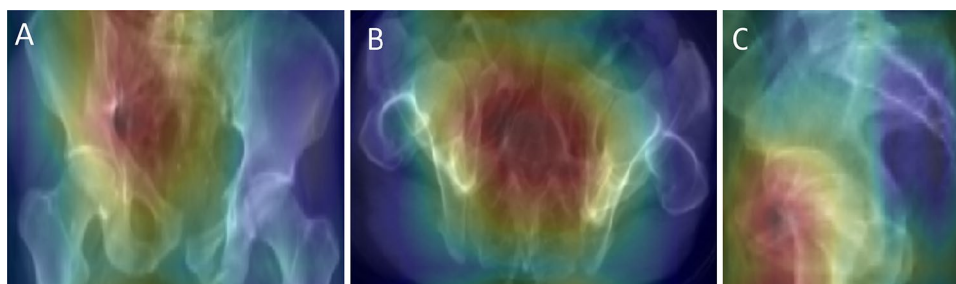


Fig. 5 Activation maps are shown in three views: **a** anteroposterior, **b** axial, and **c** sagittal for a correctly classified Tile C patient using the concatenated triplanar ResNeXt architecture. In figure part **a**, attention is primarily focused on the SI joint and posterior pelvic ring of the

superiorly translated right hemipelvis. In figure part **b**, attention is maximally focused centrally about the SI joints as well as the pubic symphysis which demonstrates AP translation. Maximal attention is focused on the abnormal pubic symphysis in figure part **c**

The triplanar parallel concatenated method had the highest overall accuracies for discriminating translational and rotational instability (translational, 85.0%, rotational 74.0%) with associated specificity, recall, and F1 score of 93.4%, 56.5%, and 0.63 for translational instability and 71.7%, 75.7%, and 0.77 for rotational instability. On the whole, across the three models translational instability was diagnosed with higher accuracy than rotational instability. Comparing methods, differences in overall accuracies for discriminating translational instability did not reach significance (85.0%–parallel concatenated, 84.7%–RNN and 81.8% 3D-encoder, $p = 0.23$ – 0.92), while the parallel concatenated method had significantly higher accuracy for rotational instability than the 3D encoder method—74.0% versus 55.8% ($p < 0.0001$). At peak MCC, specificity for diagnosing translational instability was very high for all three algorithms (93.4–96.5%); however, sensitivity was limited at the same threshold, ranging from 37.6 to

56.5%. A representative activation map for a translationally unstable (Tile C) patient using the parallel concatenated method is shown in Fig. 5. For the independent blinded reader-accuracy, sensitivity, and specificity were 76.7, 81.3, and 70.4% for rotational instability and 75.1, 80.0, 73.6% for vertical instability. Respective p -values showed equivalence for prediction of rotational instability ($p = 0.40$, 0.13, 0.80), but improved performance for the triplanar concatenated model over radiologist for translational instability ($p = 0.0007$, 0.001, < 0.0001). Automated instability grade using the parallel concatenated method was significantly associated with both the decision to perform angioembolization ($p = 0.002$) and need for massive transfusion ($p = 0.008$). This mirrored associations between the same outcomes and consensus grade (Table 2). Conversely, both RNN and 3D decoder-graded Tile C injuries failed to reach a significant association with need for angioembolization ($p = 0.063$ – 0.079). Tile grade was

Table 2 Association of rotational and translational instability with outcomes (by method)

Outcomes/groups compared	Method			
	Ground truth (consensus read)	Parallel concatenated	Recurrent neural network (LSTM)	Volumetric ResNet
Rotationally unstable (Tile B and C versus Tile A)				
<i>Angioembolization</i>	0.015*	0.002*	0.003*	0.074
<i>Massive transfusion</i>	0.011*	0.002*	<0.0001*	0.004*
<i>Mortality</i>	0.139	0.154	0.140	0.182
Translationally unstable (Tile C versus Tile A and B)				
<i>Angioembolization</i>	0.001*	0.008*	0.063	0.079
<i>Massive transfusion</i>	0.011*	0.002*	0.0004*	0.028*
<i>Mortality</i>	0.149	0.201	0.172	0.233

p -values for associations between degree of instability (determined by consensus read and three automated methods) and outcomes (i.e., mortality, need for angioembolization, need for massive transfusion) are presented *Indicates statistically significant p -values ($p < 0.05$) corresponding with significant associations between degree of instability and outcomes

not associated with an increased mortality rate using either ground truth or automated grading.

Assessment of saliency maps derived from the parallel concatenated method showed that the network focused attention on the sacroiliac complex and the pubic symphysis, which fits the paradigm of the Tile AO/OTA classification system. The mean runtime per epoch was 15 s with total training time of 3.5 h. Mean inference time was < 0.1 s per test image.

Discussion

Pelvic instability grading has not been previously explored using machine learning approaches. To our knowledge this is the first study to present a method for automated Tile AO/OTA classification from CT scans. Overall, our method employing a triplanar concatenated ResNeXt architecture demonstrated performance similar to expert radiologists for predicting the degree of instability and demonstrated the same associations with clinical outcomes (need for angioembolization and massive transfusion) as did our patient-level ground truth consensus labels.

This represents an advance over prior work limited to pelvic fracture detection using either handcrafted features or deep learning methods. Automated methods of orthopedic fracture detection using plain radiographs have achieved accuracies typically ranging between 80–90% [35–39]. Wang et al. recently achieved an accuracy of 91% for pelvic fracture detection on plain films [39].

However, identification of a pelvic fracture line is not sufficient to assess severity of injury [49]. Pelvic stability is largely a function of ligamentous integrity which is assessed by observing spatial relationships between the two innominate bones and sacrum about the SI joints and pubic symphysis in three orthogonal planes. Automated classification of pelvic fracture severity is therefore a very challenging task that involves integration of complex and global spatial information. The task is made all the more challenging by presence of pelvic binders in a substantial proportion of patients, which are known to reduce and mask pelvic ring disruptions.

In prior work, experienced trauma and musculoskeletal radiologist interpretations result in sensitivity (53 and 86%) and specificity (82 and 100%) for rotational and translational instability respectively [8]. The results of our parallel concatenated algorithm (specificity 93.4%, sensitivity 56.5%, accuracy 85.0%—translational instability; specificity 71.7%, sensitivity 75.7%, accuracy 74.0%—rotational instability) are comparable to these results. Accuracy for the independent blinded reader was equivalent for prediction of rotational instability (74.0%—algorithm versus 76.1—reader; $p = 0.40$), but lower than the triplanar concatenated model for

translational instability (85.0%—algorithm, versus 75.1%—reader; $p = 0.0007$). In many cases, trauma whole-body CT will initially be interpreted by trainees or radiologists without trauma subspecialization during after-hours care when limited staffing and high workloads result in fatigue and distraction. A rapid automated algorithm that is able to predict instability with equivalent or improved accuracy to radiologists as well as the need for major interventions—specifically angioembolization and massive transfusion—may assist with rapid objective decision support in these settings and could improve patient outcomes by facilitating earlier catheter-directed intervention and a more aggressive transfusion strategy, especially in the face of the most severe (Tile C) injury type.

The use of orthogonal thick MPR volumetric images appears to be a viable and computationally efficient alternative to native 3D CT volumes for Tile AO/OTA grading, especially for discrimination of translationally unstable (Tile C) injuries. Bone segmentation has been considered an important initializing step for DL approaches for fracture detection using CT, but not for plain radiographs [38, 41]. Unlike prior hand-crafted methods [45, 47, 48], our method is effective without requiring an explicit and potentially error-prone bone segmentation step. Our method relies on multiple thick-MPR views rather than native 3D data. Efficient 3D methods currently require techniques to reduce dimensionality and may still result in overfitting to soft tissues. This likely explains the improved performance over our 3D autoencoder method.

Recent work in orthopedic fracture identification has emphasized the use of class activation maps as a transparent means of illustrating where the algorithm is focusing its attention during decision making [41]. We found that with the use of a fully connected “view-pooling” layer, the algorithm learns to triangulate its attention primarily to the posterior sacroiliac complex and pubic symphysis, on which Tile AO/OTA grading is based.

Our method still leaves room for improvement in accuracy. Our algorithm was trained on a total of 1119 thick-MPR views from 373 CTs provided by two level I trauma centers. This is a relatively small number of patients and images. We will need to expand the size of the dataset through further collaboration with investigators at other institutions. A larger number of rotating thick MPR views reconstructed at small (e.g. 15°) intervals would be expected to improve the 3D representation generated by multiple parallel concatenated networks and should improve Tile/OTA grading; however, with current hardware, the memory footprint is substantial and would require compromises such as decreased batch size. Multi-parcellated labeling of characteristic features of pelvic fractures (e.g. pubic symphysis widening, SI joint widening, and a variety of fracture sites) was impracticable and risky without an initial study demonstrating the feasibility of automated Tile AO/OTA classification. However, this is

an area we are now exploring. Object detection networks exploiting these features could result in a single unified method for both anomaly detection and grading. Tile classification by expert radiologists offers a relatively objective but still imperfect reference standard. Learned network features can only be as good as the annotation used for training [38]. Label noise is unavoidable and is undoubtedly responsible for some of our misclassifications. Binders contribute significantly to label uncertainty due to their potential masking effect and were used in a substantial proportion of patients. Another limitation of this study is our exclusive focus on Tile AO/OTA classification for bleeding risk prediction. The grading system is also important for orthopedic fixation of pelvic fractures. These decisions are usually made by seasoned orthopedic traumatologists once life-threatening vascular injuries and hemorrhage are already controlled and may depend on additional information garnered from stress testing and manipulation under anesthesia [2, 12, 13]. Utility of Tile AO/OTA grading for guiding the reduction and fixation strategy is beyond the scope of the current work.

Prior methods have not addressed the reality that life-threatening pelvic fractures are routinely assessed as part of a whole-body CT. We offer a computationally efficient template-matching method for the purpose of automated pelvic region extraction. Automated cropping and classification supported by explainable activation maps are necessary elements of a fully automated end-to-end workflow. However, in the future, a pelvic CT fracture detection step is also required, and this is an avenue we are further pursuing. Automated segmentation algorithms have already been developed for pelvic contrast extravasation (CE) and extraperitoneal hematoma. Pelvic fracture grade, pelvic hematoma volume, and CE are the most important determinants of outcome [23, 28, 61] and could ultimately be combined into a single automated method for precise objective forecasting and decision support in patients with potentially life-threatening bleeding pelvic fractures.

Conclusion

A multiview concatenated deep network leveraging 3D information from orthogonal thick-MPR images was effective in capturing global 3D CT information necessary to predict both rotational and translational instability with accuracy on par with radiologist prediction. The approach demonstrated significantly improved performance over a 3D classifier. Further improvements may be possible by increasing the number of views. A fully end-to-end clinical workflow will require addition of a pelvic fracture detection step. Pixel-level annotations and object detection networks could be used to both detect pelvic fractures and classify pelvic ring disruptions in the future.

Funding 1. NIH K08 EB027141-01A1 (PI: David Dreizin, MD). 2. RSNA Research Scholar Grant (#RSCH1605) (PI: David Dreizin, MD). 3. Accelerated Translational Incubator Pilot (ATIP) award, University of Maryland (PI: David Dreizin, MD).

References

- Coccolini F, Stahel PF, Montori G, Biffi W, Horer TM, Catena F, Kluger Y, Moore EE, Peitzman AB, Ivatury R. Pelvic trauma: WSES classification and guidelines. *World J of Emerg Surg* 12(1):15,2017
- Garlapati AK, Ashwood N. An overview of pelvic ring disruption. *Trauma* 14(2):169-178,2012
- Dreizin D. Commentary on “Multidetector CT in Vascular Injuries Resulting from Pelvic Fractures”. *RadioGraphics* 39(7):2130-2133,2019
- Raniga SB, Mittal AK, Bernstein M, Skalski MR, Al-Hadidi AM. Multidetector CT in Vascular Injuries Resulting from Pelvic Fractures: A Primer for Diagnostic Radiologists. *RadioGraphics* 39(7):2111-2129,2019
- Vaidya R, Scott AN, Tonnos F, Hudson I, Martin AJ, Sethi A. Patients with pelvic fractures from blunt trauma. What is the cause of mortality and when? *Am J Surg* 211(3):495–500,2016
- Costantini TW, Coimbra R, Holcomb JB, Podbielski JM, Catalano R, Blackburn A, Scalea TM, Stein DM, Williams L, Conflitti J. Current management of hemorrhage from severe pelvic fractures: results of an American Association for the Surgery of Trauma multi-institutional trial. *J Trauma Acute Care* 80(5):717-725,2016
- Yoshihara H, Yoneoka D. Demographic epidemiology of unstable pelvic fracture in the United States from 2000 to 2009: trends and in-hospital mortality. *J Trauma Acute Care Surg* 76(2):380-385,2014
- Dreizin D, Nascone J, Davis DL, Mascarenhas D, Tirada N, Chen H, Bodanapally UK. Can MDCT unmask instability in binder-stabilized pelvic ring disruptions? *Am J Roentgenol* 2016;207(6):1244-1251,2016
- Tile M. Acute pelvic fractures: I. Causation and classification. *JAAOS-J Am Acad Orthop Surg* 4(3):143–151,1996
- Hanson PB, Milne JC, Chapman MW. Open fractures of the pelvis. Review of 43 cases. *J Bone Joint Surg Brit* 73(2):325–329,1991
- Slater S, Barron D. Pelvic fractures—A guide to classification and management. *Eur J Radiol* 74(1):16-23,2010
- Cooper J. Pelvic ring injuries. *Trauma* 8(2):95-110,2006
- Van Vugt A, Van Kampen A. An unstable pelvic ring: the killing fracture. *J Bone Joint Surg Brit* 88(4):427-433,2006
- Osterhoff G, Scheyerer MJ, Fritz Y, Bouaicha S, Wanner GA, Simmen H-P, Werner CM. Comparing the predictive value of the pelvic ring injury classification systems by Tile and by Young and Burgess. *Injury* 45(4):742-747,2014
- Berger-Groch J, Thiesen DM, Grossterlinden LG, Schaevel J, Fensky F, Hartel MJ. The intra- and interobserver reliability of the Tile AO, the Young and Burgess, and FFP classifications in pelvic trauma. *Arch Orthop Trauma Surg* 139(5):645-650,2019
- Alton TB, Gee AO. Classifications in brief: Young and Burgess classification of pelvic ring injuries. Springer, 2014
- Tile M. Pelvic ring fractures: should they be fixed? *J Bone Joint Surgery Brit* 70(1):1-12,1988
- Zingg T, Uldry E, Omoumi P, Clerc D, Monier A, Pache B, Moshebah M, Butti F, Becce F. Interobserver reliability of the Tile classification system for pelvic fractures among radiologists and surgeons. *Eur Radiol* 2020, pp 1–9
- Koo H, Leveridge M, Thompson C, Zdero R, Bhandari M, Kreder HJ, Stephen D, McKee MD, Schemitsch EH. Interobserver

- reliability of the young-burgess and tile classification systems for fractures of the pelvic ring. *J Orthop Trauma* 22(6):379-384,2008
20. Buchholz R. The pathological anatomy of Malgaigne fracture-dislocations of the pelvis. *J Bone Joint Surg Am* 63(3):400-404,1981
 21. Dreizin D, Bodanapally U, Mascarenhas D, O'Toole RV, Tirada N, Issa G, Nascone J. Quantitative MDCT assessment of binder effects after pelvic ring disruptions using segmented pelvic haematoma volumes and multiplanar caliper measurements. *Eur Radiol* 28(9):3953-3962,2018
 22. Gabbe BJ, Esser M, Bucknill A, Russ MK, Hofstee D-J, Cameron P, Handley C, deSteiger RN. The imaging and classification of severe pelvic ring fractures: experiences from two level 1 trauma centres. *Bone Joint J* 95(10):1396-1401,2013
 23. Dreizin D, Bodanapally U, Boscak A, Tirada N, Issa G, Nascone JW, Bivona L, Mascarenhas D, O'Toole RV, Nixon E. CT prediction model for major arterial injury after blunt pelvic ring disruption. *Radiology* 287(3):1061-1069,2018
 24. Gänsslen A, Pohlemann T, Paul C, Lobenhoffer P, Tscherne H. Epidemiology of pelvic ring injuries. *Injury* 27:13-20,1996
 25. Hussami M, Grabherr S, Meuli RA, Schmidt S. Severe pelvic injury: vascular lesions detected by ante-and post-mortem contrast medium-enhanced CT and associations with pelvic fractures. *Int J Legal Med* 131(3):731-738,2017
 26. Ruatti S, Guillot S, Brun J, Thony F, Bouzat P, Payen J, Tonetti J. Which pelvic ring fractures are potentially lethal? *Injury* 46(6):1059-1063,2015
 27. Agri F, Bourgeat M, Becce F, Moerenhout K, Pasquier M, Borens O, Yersin B, Demartines N, Zingg T. Association of pelvic fracture patterns, pelvic binder use and arterial angio-embolization with transfusion requirements and mortality rates; a 7-year retrospective cohort study. *BMC Surg* 17(1):104,2017
 28. Dreizin D, Zhou Y, Chen T, Li G, Yuille AL, McLenithan A, Morrison JJ. Deep learning-based quantitative visualization and measurement of extraperitoneal hematoma volumes in patients with pelvic fractures: potential role in personalized forecasting and decision support. *J Trauma Acute Care Surg*, 2019
 29. Dreizin D, Munera F. Blunt polytrauma: evaluation with 64-section whole-body CT angiography. *Radiographics* 32(3):609-631,2012
 30. Battey TW, Dreizin D, Bodanapally UK, Wnorowski A, Issa G, Iacco A, Chiu W. A comparison of segmented abdominopelvic fluid volumes with conventional CT signs of abdominal compartment syndrome in a trauma population. *Abdominal Radiol*, 2019, pp 1–8
 31. Dreizin D, Bodanapally UK, Munera F. MDCT of complications and common postoperative findings following penetrating torso trauma. *Emerg Radiol* 22(5):553-563,2015
 32. Dreizin D, Menaker J, Scalea TM. Extracorporeal membranous oxygenation (ECMO) in polytrauma: what the radiologist needs to know. *Emerg Radiol* 22(5):565-576,2015
 33. Dreizin D, LeBedis CA, Nascone JW. Imaging Acetabular Fractures. *Radiologic Clinics* 57(4):823-841,2019
 34. Gan K, Xu D, Lin Y, Shen Y, Zhang T, Hu K, Zhou K, Bi M, Pan L, Wu W. Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments. *Acta Orthop*, 2019, pp 1–12
 35. Olczak J, Fahlberg N, Maki A, Razavian AS, Jilert A, Stark A, Sköldenberg O, Gordon M. Artificial intelligence for analyzing orthopedic trauma radiographs: deep learning algorithms—are they on par with humans for diagnosing fractures? *Acta Orthop* 88(6):581-586,2017
 36. Kim D, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clin Radiol* 73(5):439-445,2018
 37. Urakawa T, Tanaka Y, Goto S, Matsuzawa H, Watanabe K, Endo N. Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. *Skeletal Radiol* 48(2):239-244,2019
 38. Thian YL, Li Y, Jagmohan P, Sia D, Chan VEY, Tan RT. Convolutional neural networks for automated fracture detection and localization on wrist radiographs. *Radiol Artif Intell* 1(1):e180001,2019
 39. Wang Y, Lu L, Cheng C-T, Jin D, Harrison AP, Xiao J, Liao C-H, Miao S. Weakly Supervised Universal Fracture Detection in Pelvic X-Rays. *International Conference on Medical Image Computing and Computer-Assisted Intervention: Springer*, 2019, pp 459–467
 40. Kitamura G, Chung CY, Moore BE. Ankle fracture detection utilizing a convolutional neural network ensemble implemented with a small sample, de novo training, and multiview incorporation. *J Digit Imaging*, 2019, pp 1–6
 41. Rayan JC, Reddy N, Kan JH, Zhang W, Annapragada A. Binomial classification of pediatric elbow fractures using a deep learning multiview approach emulating radiologist decision making. *Radiol Artif Intell* 1(1):e180015,2019
 42. Chung SW, Han SS, Lee JW, Oh K-S, Kim NR, Yoon JP, Kim JY, Moon SH, Kwon J, Lee H-J. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop* 89(4):468-473,2018
 43. Dreizin D, Goldmann F, Chen T, M U. Automated CT pelvic fracture severity grading with deep learning: association with clinical outcomes. *Conference on Machine Intelligence in Medical Imaging (C-MIMI)*. Austin, TX: SIIM, 2019
 44. Dreizin D, Bodanapally UK, Neerchal N, Tirada N, Patlas M, Herskovits E. Volumetric analysis of pelvic hematomas after blunt trauma using semi-automated seeded region growing segmentation: a method validation study. *Abdominal Radiol* 41(11):2203-2208,2016
 45. Ligisha P BS. A survey on pelvic bone fracture detection. *Int J Adv Res Sci Eng Technol* 5(22):19-22,2016
 46. Najarian K, Wu J, Davuluri P, Ward K, Hargraves RH. Automated computer-aided decision support for traumatic pelvic and abdominal injuries
 47. Smith R, Najarian K, Ward K. A hierarchical method based on active shape models and directed Hough transform for segmentation of noisy biomedical images; application in segmentation of pelvic X-ray images. *BMC Med Inform Decis Mak* 9(1):S2,2009
 48. Chowdhury AS, Burns JE, Mukherjee A, Sen B, Yao J, Summers RM. Automated detection of pelvic fractures from volumetric CT images. 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI): IEEE, 2012, pp 1687–1690
 49. Wu J, Davuluri P, Ward KR, Cockrell C, Hobson R, Najarian K. Fracture detection in traumatic pelvic CT images. *J Med Imaging* 2012;1,2012
 50. Su H, Maji S, Kalogerakis E, Learned-Miller E. Multi-view convolutional neural networks for 3d shape recognition. *Proceedings of the IEEE international conference on computer vision*, 2015, pp 945–953
 51. Ritter D, Orman J, Schmidgunst C, Graumann R. 3D soft tissue imaging with a mobile C-arm. *Comput Med Imaging Graph* 31(2):91-102,2007
 52. Rao YR, Prathapani N, Nagabhooshanam E. Application of normalized cross correlation to image registration. *Int J Adv Res Sci Eng Technol* 3(5):12-16,2014
 53. Sullivan J, Blake A, Isard M, MacCormick J. Bayesian object localisation in images. *Int J Comp Vis* 44(2):111–135,2001 <https://doi.org/10.1023/A:1011818912717>
 54. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp 1492–1500
 55. Boughorbel S, Jarray F, El-Anbari M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PloS one* 12(6),2017

56. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure* 405(2):442–451,1975
57. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. *Computer Vision and Pattern Recognition*, 2009, pp 248-255
58. Springenberg JT, Dosovitskiy A, Brox T, Riedmiller MA. Striving for Simplicity: The All Convolutional Net. *International Conference on Learning Representations*, 2015
59. Sekuboyina A, Rempfler M, Valentinitich A, Loeffler M, Kirschke JS, Menze BH. Probabilistic point cloud reconstructions for vertebral shape analysis. *Medical Image Computing and Computer-Assisted Intervention*, 2019, pp 375–383
60. Yu L, Yang X, Chen H, Qin J, Heng PA. Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images. *Thirty-first AAAI conference on artificial intelligence*, 2017
61. Anderson SW, Soto JA, Lucey BC, Burke PA, Hirsch EF, Rhea JT. Blunt trauma: feasibility and clinical utility of pelvic CT angiography performed with 64–detector row CT. *Radiol* 246(2):410-419,2008

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.