HHS Public Access

Author manuscript

Emerg Radiol. Author manuscript; available in PMC 2023 July 01.

Published in final edited form as:

Emerg Radiol. 2023 February ; 30(1): 41–50. doi:10.1007/s10140-022-02099-1.

Toward automated interpretable AAST grading for blunt splenic injury

Haomin Chen[1], **Mathias Unberath**[1], **David Dreizin**[2]

[1]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

[2]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

Abstract

Background—The American Association for the Surgery of Trauma (AAST) splenic organ injury scale (OIS) is the most frequently used CT-based grading system for blunt splenic trauma. However, reported inter-rater agreement is modest, and an algorithm that objectively automates grading based on transparent and verifiable criteria could serve as a high-trust diagnostic aid.

Purpose—To pilot the development of an automated interpretable multi-stage deep learning-based system to predict AAST grade from admission trauma CT.

Methods—Our pipeline includes 4 parts: (1) automated splenic localization, (2) Faster R-CNN-based detection of pseudoaneurysms (PSA) and active bleeds (AB), (3) nnU-Net segmentation and quantification of splenic parenchymal disruption (SPD), and (4) a directed graph that infers AAST grades from detection and segmentation results. Training and validation is performed on a dataset of adult patients (age ≥ 18) with voxelwise labeling, consensus AAST grading, and hemorrhage-related outcome data (n = 174).

Results—AAST classification agreement (weighted κ) between automated and consensus AAST grades was substantial (0.79). High-grade (IV and V) injuries were predicted with accuracy, positive predictive value, and negative predictive value of 92%, 95%, and 89%. The area under the curve for predicting hemorrhage control intervention was comparable between expert consensus and automated AAST grading (0.83 vs 0.88). The mean combined inference time for the pipeline was 96.9 s.

Conclusions—The results of our method were rapid and verifiable, with high agreement between automated and expert consensus grades. Diagnosis of high-grade lesions and prediction of hemorrhage control intervention produced accurate results in adult patients.

Keywords

Deep learning; Artificial intelligence; Machine learning; Interpretable AI; Explainable AI; Splenic trauma; Spleen; Blunt splenic trauma; Abdominal trauma; Computed tomography

✉David Dreizin, daviddreizin@gmail.com.
Mathias Unberath is the co-last author.

Conflict of interest The authors declare no competing interests.

Introduction

Splenic injury is the most common solid organ injury in adult blunt abdominal trauma [1, 2]. It is routinely evaluated on admission CT when abdominal trauma is suspected, both in stable patients and in those with hemodynamic compromise that demonstrate transient response to fluid resuscitation [3]. In 2018, the American Association for the Surgery of Trauma (AAST) Patient Assessment Committee (PAC) introduced an updated AAST splenic organ injury scale (OIS) for treatment decision-making based on admission two-phase abdominopelvic CT examination. While AAST grading has existed as a research tool for decades, the recent update reflects an attempt to operationalize this grading system for point-of-care use and standardize management practices. Treatment options vary by injury severity, including routine observation for low-grade injuries, and urgent angioembolization or splenectomy to control hemorrhage in high-grade injuries. Patient selection is critical as hemorrhage control interventions require mustering limited resources and staff and are not without the potential for both short- and long-term morbidity, including from catheter-related complications and an increased lifetime risk of overwhelming post-splenectomy infection (OPSI) [4, 5]. Clinical decision-making should be rapid and based on objective criteria, as the spleen is a highly vascular organ, and severe splenic injury can potentially lead to exsanguination, multi-organ system failure, and death [6].

CT-based AAST grading enjoys widespread adoption among trauma surgeons; however, in a survey of AAST member practices, only 45% of respondents reported routine use of AAST grading for blunt splenic trauma by radiologists at their institutions [7]. Even in an ideal circumstance of ubiquitous radiologist adoption and reporting, classification systems are prone to variability in the perceived grades among readers with varying experience and specialization, and reported agreement for the splenic AAST OIS has been modest under research conditions [8, 9]. In practice, radiologists are subject to shifting circumstances in their clinical environment with respect to study volume, reading room distractions, and fatigue-related performance degradation, such as from circadian rhythm disruptions after multiple consecutive night shifts [10–14]. Furthermore, admission trauma CT interpretation is time consuming. Among expert trauma radiologists, interpretation turnaround times for severely injured patients commonly exceed 20–30 min [10].

Automated AAST grading could potentially provide a rapid, objective, and accurate second-reader capability, but there has been limited automation research involving the individual CT features of splenic injury, and, to our knowledge, no work has described automated AAST grading to date [15–19].

Black box methods are prone to spurious causal inference [20] and are unaccountable to end users as the reasoning used to arrive at a decision cannot be verified. For an intelligent system to be considered responsible, ethical, and trustworthy—a requirement in the high-stakes setting of trauma care—it must at a minimum include a layer of explainability to ensure that the decisions made are justifiable [21, 22]. Since AAST grading is a multi-stage process, the intermediate steps of an automated method should be interpretable to end users, giving them agency to verify individual model assumptions should they choose to

do so. Interpretability involves the provision of packets of symbolic information that are semantically similar to the common-sense causal reasoning that would be used by an expert to arrive at a decision—in this case, at a given splenic injury grade [23, 24].

To this end, we leverage transparent deep learning approaches that are based on clinical grading standards to develop a novel automated multi-stage deep learning (DL) method that predicts the AAST splenic OIS using the most salient features of the grading system, namely, active bleeding, pseudoaneurysm, and splenic parenchymal disruption [25].

Materials and methods

Datasets

The work was conducted as part of an IRB-approved study and utilized two deidentified single institution datasets. The primary clinical dataset is previously described [3] and consists of 174 dual-phase trauma CT scans from consecutively selected adult patients (age ≥ 18) collected between 2017 and 2019 and archived at 1.5–3-mm section thickness, with voxelwise labeling of pseudoaneurysm (PSA), active bleed (AB), and splenic parenchymal disruption (SPD). PSAs in this context are vascular injuries contained by splenic parenchyma with densities similar to or slightly higher than the blood pool [1]. Foci of AB extend beyond the splenic parenchyma and typically increase in size on the portal venous phase. The portal venous phase is optimal for the delineation of SPD and AB, whereas PSA is best detected on the arterial phase [26, 27].

All studies in this dataset had accompanying AAST consensus grading by three expert trauma radiologists and outcome data including whether patients underwent angioembolization or splenectomy. Foci of active bleed and pseudoaneurysm occupy a small fraction of an abdominal CT volume when present. To address AB and PSA class imbalance, the dataset was augmented with labeled dual-phase CTs from a second existing blunt splenic injury dataset with 68 consecutively selected patients who underwent splenic hemorrhage control intervention and had AB, PSA, or both on CT between 2007 and 2016 [28] (ABPSA dataset). A third dataset with the subset of 41 labeled splenic normals from the medical segmentation decathlon challenge [29] (SMSD dataset) was employed to develop an initial automated localization step of injured spleens.

Summary of the AAST splenic organ injury scale and clinical relevance

All patients with intraperitoneal AB receive a grade of V in this system, while any patient with PSA but no AB receives a grade of IV [25]. This assignment is irrespective of the size and number of splenic vascular lesions. High-grade (IV and V) injuries are considered to necessitate angioembolization (AE) for hemorrhage control at a minimum, and surgeons may opt instead for early splenectomy. Rates of failure for attempted splenic salvage for high-grade injuries are historically high, ranging from 20% to over 60% [5, 30–33], but are improved with the liberal use of angioembolization [31]. The surgeon's judgment and institution-specific guidelines play an important role in the choice of hemorrhage control intervention [32]. In patients without splenic vascular injury, the AAST splenic OIS grade is primarily determined by the extent of visually estimated SPD [25, 34] using

diameter measurement cut-offs established using rules of thumb in the original 1994 AAST classification [34]. Management of grade 3 injuries (with greater than approximately 3 cm estimated SPD depth but less than 25% parenchymal involvement) is highly variable. At many institutions, these injuries are considered low grade and are routinely managed conservatively [32]; however, some investigators report improved salvage rates with routine angiographic screening followed by AE if a vascular injury is seen on the image intensifier [5, 35], and still others recommend the routine use of AE as a precautionary measure [36]. Variability in practice patterns lies in the potential for missed small or subtle vascular injury on CT due to variable scan timing and transient vessel thrombosis or spasm [36]. Low-grade (grade I and II) injuries have less than approximately 3 cm SPD depth. Conservative management is considered the standard of care for low-grade injuries across institutions [32, 33].

Automated splenic injury grading pipeline: overview

The complete pipeline for our proposed automated AAST OIS grade prediction method is shown in Fig. 1. The pipeline begins with an automated 3D cropping step aimed at (a) reducing irrelevant background which could otherwise contribute to false positive results and (b) increasing the proportion of positive voxels in the data given small target volumes of AB, PSA, and SPD. 2D Faster R-CNN [37] is then applied to the detection of AB on portal venous phase axial slices and PSA on arterial axial slices, leveraging the optimal phase for the detection of each feature [26, 27]. SPD is segmented using nnU-Net [38] and quantified using voxel counting. Vascular injury detections and SPD volumes are then fed into a hierarchical rules-based system to derive the predicted AAST grade.

Step 1: automated splenic localization

A semi-supervised method using the noisy student algorithm [39] was employed to derive whole-organ label masks for injured spleens. The method initially utilizes a 3D U-Net trained on the external medical segmentation decathlon challenge (SMSD) “teacher” splenic normal labels to generate pseudo labels in the in-house clinical and ABPSA datasets. The labels and pseudo labels are used to create improved spleen segmentations in an iterative process. The 3D U-Net predicted segmentations from a given iteration are used as ground truth training cases in the next iteration, resulting in gradual segmentation refinement. We used the MONAI platform (40) to construct the 3D U-Net [41], with a learning rate of 1e-4. All training was performed using an RTX 3090 NVIDIA GPU with 24 GB of RAM. Training involved six iterations with 600 epochs per iteration. Subsequently, the splenic volume is dilated by 30 voxels to include relevant perisplenic soft tissue structures, and axial images above and below the dilated volume are excluded. Following this pre-processing step, visual inspection of each CT study indicated that all foci of vascular injury and SPD were contained within the cropped range and there were no failures.

Step 2: vascular lesion detection

Following splenic localization, 2D Faster R-CNN [37], a two-stage object detection network, was used to detect PSA on axial arterial phase images and AB on axial portal venous phase images. The network first extracts image features and generates region proposals. Second, it fine-tunes the box proposal size and location and classifies each proposal. We

used ResNeXt-101 with Feature Pyramid Network (FPN) as the backbone given prior best performance in the COCO object detection task [42, 43]. Training was augmented using the non-overlapping ABPSA dataset, within which each CT study includes at least one focus of pseudoaneurysm or active bleed. Faster R-CNN was trained in five-fold cross-validation, splitting the combined dataset evenly into 5 independent subsets to avoid data leak, and using each fold for validation and the remaining 4 folds for training. Thresholds were selected to achieve the highest possible sensitivity. Faster R-CNN was implemented in the detectron2 platform [44], with the following parameters: learning rate of 0.02, a 10 × decay at 15,000 iterations, and a total of 30,000 training iterations.

Step 3: splenic parenchymal disruption segmentation and quantification

We applied nnU-Net [38] to segment splenic parenchymal disruption (SPD) due to its state-of-the-art performance across a large variety of segmentation tasks. nnU-Net trains four models in five-fold cross-validation (2D U-Net, low-resolution 3D U-Net, high-resolution 3D U-Net, and a low and high-resolution cascaded 3D U-Net) and determines the best-performing model or ensemble of models for inference. Design choices including hyperparameter selection and pre-processing steps are made automatically from specific dataset properties known as the dataset and pipeline fingerprint [38]. Voxel counting is then applied to automated label masks to determine laceration volumes.

Step 4: AAST grade determination

The automated detections and segmentation-derived volumes are directly applied to a hierarchical AAST OIS-based system of rules (Fig. 1). The directed graph starts from the highest-grade decision, proceeding toward the lowest grade in a manner similar to how AAST grading is employed in clinical practice (Table 1). First, if AB is detected, the patient receives a grade of V. If no AB is detected, but PSA is detected, the patient receives a grade of IV. If no vascular injury is present, the grade is determined by the extent of splenic parenchymal disruption. In our method, grades are stratified by the automated SPD volume in place of visual estimation of injury depth using a logistic regression-derived cut-off. Since low-grade lesions are managed conservatively, we combined grades I and II into a single “low grade” class. A laceration volume of 14 mL optimally discriminated between low-grade (I and II) and grade III lesions.

Results

Descriptive statistics for the clinical dataset are provided in Table 2. The weighted Cohen’s κ between automated and consensus grades in the clinical dataset of 174 patients was 0.79.

Using radiologist expert consensus grading as the reference standard, diagnosis of high-grade (IV and V) splenic injuries—those that require urgent hemorrhage control intervention for splenic salvage [33, 45]—was achieved with an accuracy, sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) of 92%, 93%, 92%, 95%, and 89%. The proposed method correctly identified almost all high-grade injuries. Only 1 of 69 patients with high-grade (IV and V) injuries was underestimated as a grade III injury. 9 grade III injuries were overestimated as high grade. All patients classified as grades I and

II by radiologists (n = 91) were correctly predicted as low-grade injuries using our method. This indicates that among patients who would normally be managed conservatively, there were no false positive severe injuries. In subanalysis of high-grade injuries, 19 out of 51 consensus grade IV patients are overestimated by our method as grade V, but only 2 out of 18 grade V patients are underestimated as grade IV. The area under the curve (AUC) for predicting a composite outcome of intervention with angioembolization or splenectomy was 0.83 for automated grades, comparable to an AUC of 0.88 for consensus grading.

Performance of detection and segmentation tasks

For AB detection in the clinical and ABPSA datasets, Faster R-CNN achieved an AUC, accuracy, sensitivity, specificity, NPV, and PPV of 0.84, 88%, 82%, 91%, 91%, and 83% (Table 3). For PSA, Faster R-CNN achieved an AUC, accuracy, sensitivity, specificity, NPV, and PPV of 0.79, 83%, 91%, 78%, 94%, and 71%. Examples of AB and PSA box detections are shown in Fig. 2.

A review of AB false positive detections revealed that in 13 of 19 patients, the detection network misclassified pseudoaneurysm as active bleeding on the portal venous phase, owing to lingering pseudoaneurysm blush which otherwise characteristically washes out and is inconspicuous on this phase [1, 26]. This could be attributable to imperfect CT timing using a descending thoracic aorta ROI trigger threshold or variability in cardiac output between patients [3]. Nevertheless, box proposals of these lesions provide transparent results which would allow radiologists, interventionalists, or surgeons to reject detections they disagree with.

Patients in the clinical dataset had labeled laceration with a range of volumes between 0.1 and 255.1 mL (median volume: 1.8 mL). All patients with < 1 mL had low-grade injuries. Automated segmentations in patients with ≥ 1 mL SPD and no vascular injury that would take absolute priority in injury grading had a volume similarity (VS) index of 0.68, and dice similarity coefficient of 0.54 with respect to manual labels, corresponding with high-saliency visual results that conformed to the margins of laceration (Fig. 3). Pearson's r between manual and automated volumes was 0.89 ("excellent" range).

Mean inference times for our method included 1.5 s for automated splenic localization, 4.2 s for Faster R-CNN, and 91.2 s for ensembled nnU-Net, for a total of 96.9 s.

Discussion

The AAST splenic organ injury scale is often used to guide surgical management decisions. High-grade (AAST IV and V) lesions typically require angioembolization or splenectomy for hemorrhage control [31–33, 45, 46]. Low-grade (AAST I and II) lesions are routinely managed conservatively [32, 33]. Management of grade III lesions remains variable and institution dependent [5, 35, 36]. AAST grading is limited by modest interobserver agreement, inconsistent reporting, and the long interpretation and reporting times of admission trauma CT [7–10]. An interpretable automated system could augment objective decision-making as a second-reader diagnostic aid, producing verifiable visual results that could be accepted or rejected by the end user.

To date, we are not aware of previous attempts to automate AAST splenic OIS grading, either with black box or interpretable methods. Few studies report automated methods for the detection or segmentation of individual features relevant to the grading of splenic injury. Several works describe whole-spleen segmentation in trauma patients using semi-automated [15] and automated methods, such as with 3D active shape contours and probabilistic atlases [16, 17]; however, whole-spleen volumes are of unknown clinical import in trauma. Other work describes black-box detection but not quantification of splenic parenchymal disruption using a random forest method and a convolutional neural network with a long short-term memory (LSTM) model [18]. One group examined detection of active bleed but not pseudoaneurysm using a hand-crafted feature engineering-based method. Of 30 splenic injury subjects, 4 had active bleeding. The method had a detection accuracy and PPV of only 73% and 33%, respectively [19]. In more recent work using deep learning segmentation methods, automated liver parenchymal disruption volumes predicted angiopositivity on subsequent conventional angiography [47]. A variety of robust methods have emerged using DL for non-trivial hemorrhage-related tasks, including hemoperitoneum, extraperitoneal pelvic hematoma, and hemothorax quantitative visualization [48–51]. Additional DL-related work in the spleen has demonstrated the feasibility of splenic vascular injury segmentation, however, without an initial detection step or the ability to differentiate between pseudoaneurysm and active bleeding [28]. Quantification of vascular injury burden is presently not included in the AAST OIS framework, and detection of vascular injury is made in a binary fashion on the patient level.

In the present work, we pilot the development of an interpretable automated AAST splenic grade prediction pipeline using the dual-phase imaging protocol currently recommended by the AAST Patient Assessment Committee [25]. Dual-phase imaging is optimized for the delineation of splenic disruption and detection of active bleeding on the portal venous phase and detection of pseudoaneurysm on the arterial phase [26, 27].

Our pipeline begins with robust splenic localization, followed by detection of pseudoaneurysm and active bleed using Faster R-CNN using a RexNeXt-101 backbone with FPN and nnU-Net segmentation of splenic parenchymal disruption. The detections and splenic volumes are then fed into an intuitive rules-based system guided by major concepts of the AAST OIS combined with consideration of clinical evidence and expert knowledge. Using the Landis and Koch scheme [52], we achieved substantial agreement with expert consensus ground truth AAST grading (weighted κ of 0.79) and a 92% accuracy for predicting high-grade (IV and V) lesions. The AUC of automated grades for predicting a composite outcome of angioembolization or splenectomy was comparable to the prediction of the same outcome using expert consensus AAST grading. The method is much more rapid in inference compared with reported admission trauma CT interpretation times.

For active bleed detection, Faster R-CNN achieved an accuracy, NPV, and PPV of 88%, 91%, and 83%, and for pseudoaneurysm detection, an accuracy, NPV, and PPV of 83%, 94%, and 71%. nnU-Net results demonstrated reasonably high DSC and volume similarity for a task involving small target volumes, with excellent Pearson correlation between manual and automated volumes (0.89) and highquality visual results. A threshold of 14 mL optimally distinguished between low-grade (I and II) and grade III lesions.

Our study had some limitations. In our clinical dataset, we were not able to determine a threshold distinguishing between grade III and the subset of grade IV patients without detected pseudoaneurysms as there were only two such patients, both with volumes that overlapped with the volume distribution of grade III injuries. A larger multi-institutional dataset will contain more such patients and likely allow determination of additional cut-offs for grade IV and V injuries. The 2018 AAST OIS is not without controversy. Despite the recommendation of dual-phase scanning, this protocol is not widely adopted at present. Some high-volume level I trauma centers perform scanning of the abdomen in the portal venous phase only or use a single-phase split bolus protocol [53–55]. Additional features included in the AAST OIS were selected using heuristics without strong evidence. For example, the estimated size of subcapsular hematoma is included, although our review of the literature yielded few studies supporting this as a univariate predictor of outcome [56], and we are not aware of studies showing that this feature is independently predictive when accounting for SPD and vascular injury. Laceration and intraparenchymal hematoma (blood pooled within the interstices of a laceration) are typically indistinguishable and grouped as splenic parenchymal injury or disruption in scientific works and clinical practice [26, 57]. The 2018 AAST discriminates between intra- and extraperitoneal active bleeding even though the extensive literature on vascular lesions does not differentiate between active bleeding confined to or extending beyond the capsule into the peritoneal cavity [26, 27, 58, 59]. Additionally, the AAST OIS currently includes capsular tear for discriminating between grade I and II lesions [25]. The capsule is microscopic and not directly visible on CT. Capsular laceration is only implied by the presence of SPD, and this feature may be redundant. By including only those features in our simplified system supported by strong evidence [26, 32, 45, 58, 60], we managed to achieve substantial agreement with human expert grading. The internal and clinical validity of each feature of the 2018 AAST splenic OIS is an active area of investigation by the American Society of Emergency Radiology (ASER) Splenic Trauma Expert Panel [61], and we will consider inclusion of additional features as dictated by its ongoing conclusions and further emergence of scientific evidence. Finally, our method is trained and validated using only adult patients, and future work will ultimately need to include pediatric trauma victims.

Other future avenues of investigation may include collaboration with participants of the ASER panel, which has curated a large, as yet unlabeled, multicenter CT dataset with studies performed using a variety of protocols on a wide range of scanner makes and models. The method could be refined, retrained, and tested on a hold-out sample. A simulated deployment study comparing multileader agreement and diagnostic performance with and without our interpretable method as a diagnostic support system is also planned.

In conclusion, in this single-center pilot study, we developed a rapid interpretable automated method for grading splenic injury using the most salient features of the AAST splenic OIS. The method achieved high agreement with and accuracy compared to consensus expert AAST grading in cross-validation. Prediction of hemorrhage control intervention was comparable between automated and consensus grading. Future avenues include scaling to a larger dataset, conducting a simulated deployment study, and assessing user acceptance.

Funding

NIH K08 EB027141-01A1 (PI: David Dreizin, MD).

References

1. Dreizin D, Munera F (2012) Blunt polytrauma: evaluation with 64-section whole-body CT angiography. Radiographics 32(3):609–631 [PubMed: 22582350]
2. Chahine AH, Gilyard S, Hanna TN, Fan S, Risk B, Johnson JO, Duszak R Jr, Newsome J, Xing M, Kokabi N (2021) Management of splenic trauma in contemporary clinical practice: a national trauma data bank study. Acad Radiol 28:S138–S147 [PubMed: 33288400]
3. Dreizin D, Champ K, Dattwyler M, Bodanapally U, Smith EB, Li G, Singh R, Wang Z, Liang Y (2022) Blunt splenic injury in adults: association between volumetric quantitative CT parameters and intervention. J Trauma Acute Care Surg. 10.1097/ta.0000000000003684
4. Zarzaur BL, Kozar R, Myers JG, Claridge JA, Scalea TM, Neideen TA, Maung AA, Alarcon L, Corcos A, Kerwin A (2015) The splenic injury outcomes trial: an American Association for the Surgery of Trauma multi-institutional study. J Trauma Acute Care Surg 79(3):335–342 [PubMed: 26307863]
5. Haan JM, Biffl W, Knudson MM, Davis KA, Oka T, Majercik S, Dicker R, Marder S, Scalea TM (2004) Committee WTAM-IT Splenic embolization revisited: a multicenter review. J Trauma Acute Care Surg 56(3):542–547
6. Krausz MM, Hirsh M (2003) Bolus versus continuous fluid resuscitation and splenectomy for treatment of uncontrolled hemorrhagic shock after massive splenic injury. J Trauma Acute Care Surg 55(1):62–68
7. Zarzaur BL, Kozar RA, Fabian TC, Coimbra R (2011) A survey of American Association for the Surgery of Trauma member practices in the management of blunt splenic injury. J Trauma Acute Care Surg 70(5):1026–1031
8. Barquist ES, Pizano LR, Feuer W, Pappas PA, McKenney KA, LeBlang SD, Henry RP, Rivas LA, Cohn SM (2004) Inter-and intrarater reliability in computed axial tomographic grading of splenic injury: why so many grading scales? J Trauma Acute Care Surg 56(2):334–338
9. Clark R, Hird K, Misur P, Ramsay D, Mendelson R (2011) CT grading scales for splenic injury: Why can't we agree? J Med Imaging Radiat Oncol 55(2):163–169 [PubMed: 21501405]
10. Banaste N, Caurier B, Bratan F, Bergerot J-F, Thomson V, Millet I (2018) Whole-body CT in patients with multiple traumas: factors leading to missed injury. Radiology 289(2):374–383 [PubMed: 30084754]
11. Watchorn J, Miles R, Moore N (2013) The role of CT angiography in military trauma. Clin Radiol 68(1):39–46 [PubMed: 22824572]
12. Glover M IV, Almeida RR, Schaefer PW, Lev MH, Mehan WA Jr (2017) Quantifying the impact of noninterpretive tasks on radiology report turn-around times. J Am Coll Radiol 14(11):1498–1503 [PubMed: 28916177]
13. Hunter TB, Taljanovic MS, Krupinski E, Ovitt T, Stubbs AY (2007) Academic radiologists' on-call and late-evening duties. J Am Coll Radiol 4(10):716–719 [PubMed: 17903757]
14. Hanna TN, Loehfelm T, Khosa F, Rohatgi S, Johnson J-O (2016) Overnight shift work: factors contributing to diagnostic discrepancies. Emerg Radiol 23(1):41–47 [PubMed: 26475281]
15. Cruz-Romero C, Agarwal S, Abujudeh HH, Thrall J, Hahn PF (2016) Spleen volume on CT and the effect of abdominal trauma. Emerg Radiol 23(4):315–323 [PubMed: 27166964]
16. Wood A, Soroushmehr SR, Farzaneh N, Fessell D, Ward KR, Gryak J, Kahrobaei D, Na K. Fully automated spleen localization and segmentation using machine learning and 3D active contours. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC): IEEE, 2018; p. 53–56.
17. Dandin O, Teomete U, Osman O, Tulum G, Ergin T, Sabuncuoglu MZ (2016) Automated segmentation of the injured spleen. Int J Comput Assist Radiol Surg 11(3):351–368 [PubMed: 26337443]

18. Wang J, Wood A, Gao C, Najarian K, Gryak J (2021) Automated Spleen Injury Detection Using 3D Active Contours and Machine Learning. Entropy 23(4):382 [PubMed: 33804831]
19. Teomete U, Tulum G, Ergin T, Cuce F, Koksal M, Dandin O, Osman O (2018) Automated computer-aided diagnosis of splenic lesions due to abdominal trauma. Hippokratia 22(2):80 [PubMed: 31217680]
20. DeGrave AJ, Janizek JD, Lee S-I (2021) AI for radiographic COVID-19 detection selects shortcuts over signal. Nature Machine Intelligence 3(7):610–619
21. Zapaishchykova A, Dreizin D, Li Z, Wu JY, Roohi SF, Unberath M. An interpretable approach to automated severity scoring in pelvic trauma. arXiv preprint arXiv:210510238 2021.
22. Chen H, Gomez C, Huang C-M, Unberath M. INTRPRT: A systematic review of and guidelines for designing and validating transparent AI in medical image analysis. arXiv preprint arXiv:211212596 2021.
23. Vlontzos A, Rueckert D, Kainz B. A review of causality for learning algorithms in medical image analysis. arXiv preprint arXiv:220605498 2022.
24. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, García S, Gil-López S, Molina D, Benjamins R (2020) Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion 58:82–115
25. Kozar RA, Crandall M, Shanmuganathan K, Zarzaur BL, Coburn M, Cribari C, Kaups K, Schuster K, Tominaga GT, Committee APA (2018) Organ injury scaling 2018 update: spleen, liver, and kidney. J Trauma Acute Care Surg 85(6):1119–1122 [PubMed: 30462622]
26. Boscak AR, Shanmuganathan K, Mirvis SE, Fleiter TR, Miller LA, Sliker CW, Steenburg SD, Alexander M (2013) Optimizing trauma multidetector CT protocol for blunt splenic injury: need for arterial and portal venous phase scans. Radiology 268(1):79–88 [PubMed: 23449955]
27. Uyeda JW, LeBedis CA, Penn DR, Soto JA, Anderson SW (2014) Active hemorrhage and vascular injuries in splenic trauma: utility of the arterial phase in multidetector CT. Radiology 270(1):99–106 [PubMed: 24056401]
28. Zhou Y, Dreizin D, Wang Y, Liu F, Shen W, Yuille AL. External attention assisted multi-phase splenic vascular injury segmentation with limited data. IEEE Transactions on Medical Imaging 2021.
29. Antonelli M, Reinke A, Bakas S, Farahani K, Landman BA, Litjens G, Menze B, Ronneberger O, Summers RM, van Ginneken B. The medical segmentation decathlon. arXiv preprint arXiv:210605735 2021.
30. Cooney R, Ku J, Cherry R, Maish GO III, Carney D, Scorza LB, Smith JS (2005) Limitations of splenic angioembolization in treating blunt splenic injury. J Trauma Acute Care Surg 59(4):926–932
31. Bhullar IS, Frykberg ER, Siragusa D, Chesire D, Paul J, Tepas JJ III, Kerwin AJ (2012) Selective angiographic embolization of blunt splenic traumatic injuries in adults decreases failure rate of nonoperative management. J Trauma Acute Care Surg 72(5):1127–1134 [PubMed: 22673236]
32. Crichton JCI, Naidoo K, Yet B, Brundage SI, Perkins Z (2017) The role of splenic angioembolization as an adjunct to nonoperative management of blunt splenic injuries: a systematic review and meta-analysis. J Trauma Acute Care Surg 83(5):934–943 [PubMed: 29068875]
33. Requarth JA, D'Agostino RB Jr, Miller PR (2011) Nonoperative management of adult blunt splenic injury with and without splenic artery embolotherapy: a meta-analysis. J Trauma Acute Care Surg 71(4):898–903
34. Moore EE, Cogbill TH, Jurkovich GJ, Shackford SR, Malangoni MA, Champion HR (1995) Organ injury scaling: spleen and liver (1994 revision). J Trauma Acute Care Surg 38(3):323–324
35. Haan JM, Bochicchio GV, Kramer N, Scalea TM (2005) Nonoperative management of blunt splenic injury: a 5-year experience. J Trauma Acute Care Surg 58(3):492–498
36. Miller PR, Chang MC, Hoth JJ, Mowery NT, Hildreth AN, Martin RS, Holmes JH, Meredith JW, Requarth JA (2014) Prospective trial of angiography and embolization for all grade III to V blunt splenic injuries: nonoperative management success rate is significantly improved. J Am Coll Surg 218(4):644–648 [PubMed: 24655852]

37. Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems 2015;28.
38. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18(2):203–211 [PubMed: 33288961]
39. Xie Q, Luong M-T, Hovy E, Le QV. Self-training with noisy student improves imagenet classification. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition2020; p. 10687–10698.
40. Consortium M. Project MONAI. Zenodo. Available online: https://zenodoorg/record/4323059#YXaMajgzaUk (accessed on 25 May 2020) 2020.
41. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention: Springer, 2015; p. 234–241.
42. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL (2014) Microsoft coco: Common objects in context. Springer, European conference on computer vision, pp 740–755
43. Zhang R, Tian Z, Shen C, You M, Yan Y. Mask encoding for single shot instance segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition2020; p. 10226–10235.
44. Wu Y, Kirillov A, Massa F, Lo W-Y, Girshick R. Detectron2 (2019). 2019.
45. Bhangu A, Nepogodiev D, Lal N, Bowley DM (2012) Meta-analysis of predictive factors and outcomes for failure of non-operative management of blunt splenic trauma. Injury 43(9):1337–1346 [PubMed: 21999935]
46. Dreizin D, Yu T, Motley K, Li G, Morrison JJ, Liang Y. Blunt splenic injury: assessment of follow-up CT utility using quantitative volumetry. Frontiers in radiology 2022:23.
47. Dreizin D, Chen T, Liang Y, Zhou Y, Paes F, Wang Y, Yuille AL, Roth P, Champ K, Li G. Added value of deep learning-based liver parenchymal CT volumetry for predicting major arterial injury after blunt hepatic trauma: a decision tree analysis. Abdominal Radiology 2021:1–11. [PubMed: 33104826]
48. Dreizin D, Zhou Y, Fu S, Wang Y, Li G, Champ K, Siegel E, Wang Z, Chen T, Yuille AL (2020) A multiscale deep learning method for quantitative visualization of traumatic hemoperitoneum at CT: assessment of feasibility and comparison with subjective categorical estimation. Radiology Artificial Intelligence 2(6):e190220 [PubMed: 33330848]
49. Dreizin D, Zhou Y, Chen T, Li G, Yuille AL, McLenithan A, Morrison JJ (2020) 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 88(3):425–433 [PubMed: 32107356]
50. Dreizin D, Zhou Y, Zhang Y, Tirada N, Yuille AL (2020) Performance of a deep learning algorithm for automated segmentation and quantification of traumatic pelvic hematomas on CT. J Digit Imaging 33(1):243–251 [PubMed: 31172331]
51. Dreizin D, Nixon B, Hu J, Albert B, Yan C, Yang G, Chen H, Liang Y, Kim N, Jeudy J. A pilot study of deep learning-based CT volumetry for traumatic hemothorax. Emergency Radiology 2022:1–8.
52. Landis JR, Koch GG. The measurement of observer agreement for categorical data. biometrics 1977:159–174. [PubMed: 843571]
53. Morell-Hofert D, Primavesi F, Fodor M, Gassner E, Kranebitter V, Braunwarth E, Haselbacher M, Nitsche UP, Schmid S, Blauth M (2020) Validation of the revised 2018 AAST-OIS classification and the CT severity index for prediction of operative management and survival in patients with blunt spleen and liver injuries. Eur Radiol 30(12):6570–6581 [PubMed: 32696255]
54. Jeavons C, Hacking C, Beenen LF, Gunn ML (2018) A review of split-bolus single-pass CT in the assessment of trauma patients. Emerg Radiol 25(4):367–374 [PubMed: 29478119]
55. Beenen LF, Sierink JC, Kolkman S, Nio CY, Saltzherr TP, Dijkgraaf MG, Goslings JC (2015) Split bolus technique in polytrauma: a prospective study on scan protocols for trauma analysis. Acta Radiol 56(7):873–880 [PubMed: 25033993]

56. Lopez JM Jr, McGonagill PW, Gross JL, Hoth JJ, Chang MC, Parker K, Requarth JA, Miller PR (2015) Subcapsular hematoma in blunt splenic injury: a significant predictor of failure of nonoperative management. J Trauma Acute Care Surg 79(6):957–960 [PubMed: 26488320]

57. Scatamacchia SA, Raptopoulos V, Fink MP, Silva WE (1989) Splenic trauma in adults: impact of CT grading on management. Radiology 171(3):725–729 [PubMed: 2717742]

58. Rowell SE, Biffl WL, Brasel K, Moore EE, Albrecht RA, DeMoya M, Namias N, Schreiber MA, Cohen MJ, Shatz DV (2017) Western Trauma Association Critical Decisions in Trauma: Management of adult blunt splenic trauma—2016 updates. J Trauma Acute Care Surg 82(4):787–793 [PubMed: 27893644]

59. Coccolini F, Montori G, Catena F, Kluger Y, Biffl W, Moore EE, Reva V, Bing C, Bala M, Fugazzola P (2017) Splenic trauma: WSES classification and guidelines for adult and pediatric patients. World J Emerg Surg 12(1):1–26 [PubMed: 28070213]

60. Boscak A, Shanmuganathan K (2012) Splenic trauma: what is new? Radiologic Clinics 50(1):105–122 [PubMed: 22099490]

61. Lee JT, Slade E, Uyeda J, Steenburg SD, Chong ST, Tsai R, Raptis D, Linnau KF, Chinapuvvula NR, Dattwyler MP (2021) American Society of Emergency Radiology multicenter blunt splenic trauma study: CT and clinical findings. Radiology 299(1):122 [PubMed: 33529133]

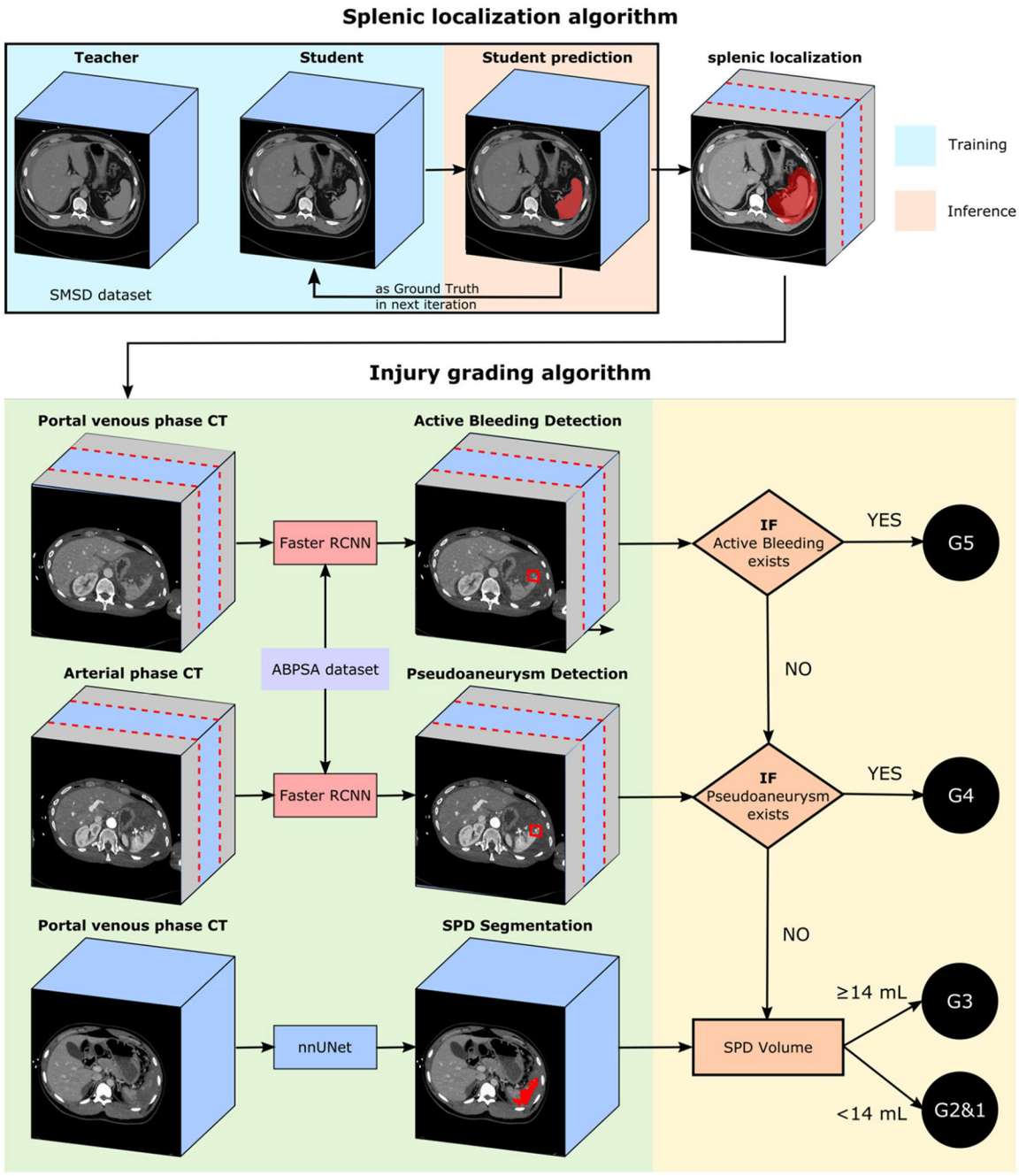

Fig. 1.

Overall pipeline of the proposed automatic splenic AAST grading algorithm. Splenic localization is first performed, which crops irrelevant slices on abdominopelvic CT cranial and caudal to the spleen and neighboring soft tissue that may harbor foci of active bleeding. Active bleeding is detected on portal venous CT scans, and pseudoaneurysm is detected on arterial CT scans by Faster RCNN on localized axial sections. SPD is segmented by nn-UNet, and volume is calculated. Active bleeding and pseudoaneurysm detection and SPD volume is then fed into a directed graph for AAST grading prediction

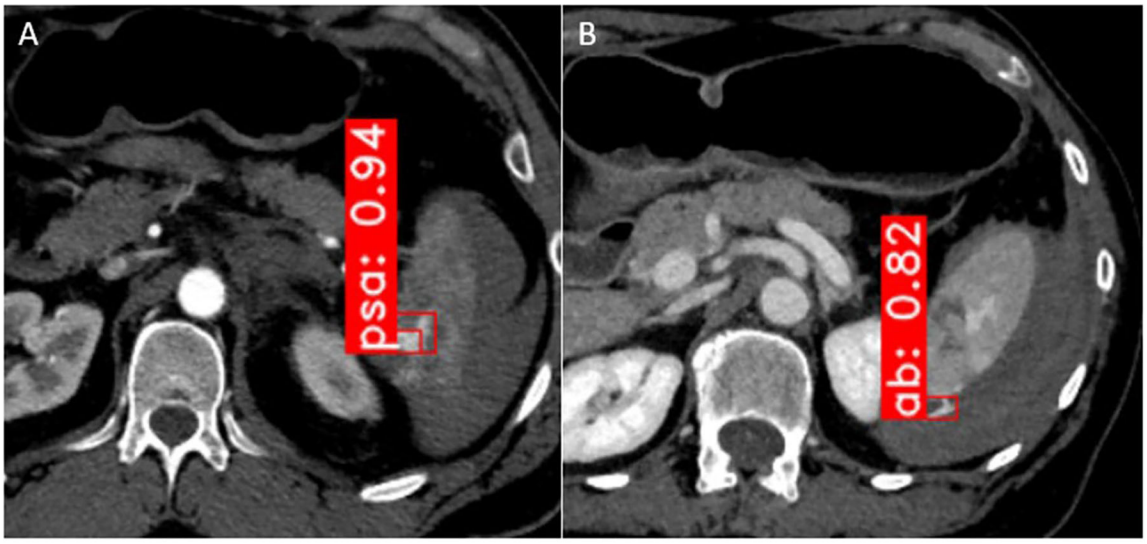

Fig. 2.

Pseudoaneurysm ("psa," part **A**) and active bleed ("ab," part **B**) box detection on arterial and portal venous phase images, respectively. Detection is achieved using Faster R-CNN with a very deep (ResNeXt-101 with FPN) backbone. Numbers shown refer to probability of correct detection as a fraction of 1

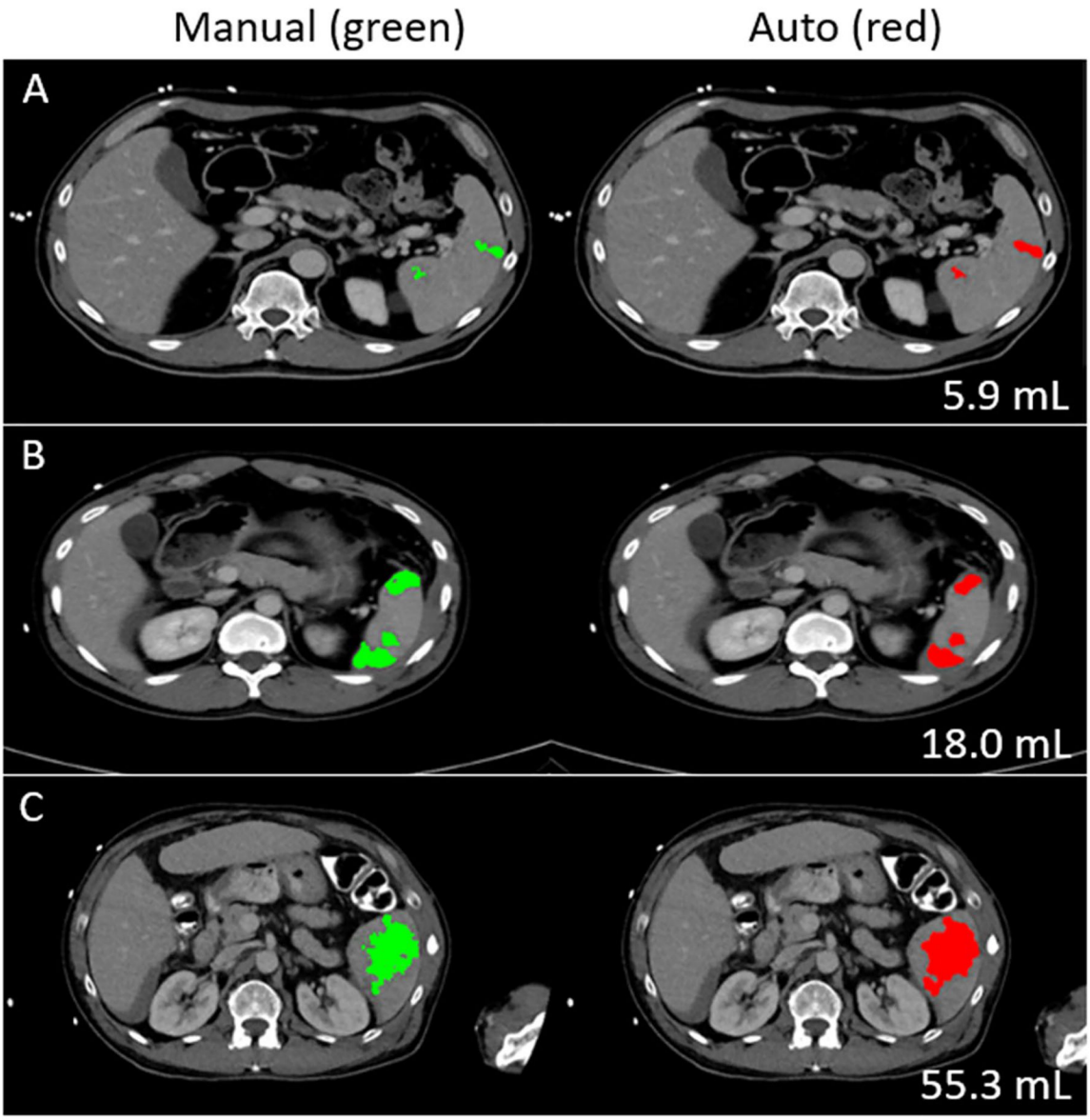

Fig. 3.

Splenic parenchymal disruption (SPD) segmentation/quantitative visualization for a range of volumes. On regression, an optimal cut-off of 14 mL distinguished between low-grade (I and II) and grade III lesions

Table 1

Comparison of the 2018 spleen AAST splenic organ injury scale (OIS) grading criteria and lesions in our method selected using available evidence and expert knowledge

	AAST grading criteria	Lesions in our system
Grade I	Subcapsular hematoma < 10% surface area Parenchymal laceration < 1 cm depth Capsular tear[1]	SPD < 14 mL
Grade II	Subcapsular hematoma[2] 10–50% surface area Intraparenchymal hematoma < 5 cm Parenchymal laceration 1–3 cm	
Grade III	Subcapsular hematoma > 50% surface area Ruptured subcapsular/intraparenchymal hematoma[3] ≥ 5 cm Parenchymal laceration[4] > 3 cm depth	SPD ≥ 14 mLwith no PSA or AB
Grade IV	Any injury in the presence of a contained splenic vascular injury[5] Parenchymal laceration producing > 25% devascularization[6] Active bleeding confined within splenic capsule[7]	PSA
Grade V	Any injury in the presence of a splenic vascular injury with active bleeding extended beyond the spleen into the peritoneum[8]	AB

Capsular tear[1] and subcapsular hematoma[2] are ignored in our simplified system (see “Discussion”). Laceration and intraparenchymal hematoma[3,4] are grouped as splenic parenchymal disruption. Contained splenic vascular injury[5] is synonymous with pseudoaneurysm on CT. In our dataset, only two patients with ground truth consensus grade of IV had laceration with no detected pseudoaneurysm, precluding derivation of a data-driven cut-off differentiating grade III and IV lesions by SPD volume[6]. Current literature does not differentiate between active bleeding confined to or extending beyond the capsule[7,8]

Abbreviations: SPD- splenic parenchymal disruption; PSA- pseudoaneurysm; AB- active bleeding

Table 2

Ground truth descriptive statistics of the clinical splenic injury dataset

	n **(%)**
Total	174 (100)
Active bleeding	28 (16)
Pseudoaneurysm	54 (31)
PSD ≥ 1 cm^3	93 (53)
Expert consensus grade	
Grade V	18 (10)
Grade IV	50 (29)
Grade III	15 (9)
Grades I & II	91 (52)

All CT studies in the ABPSA dataset (n = 68ven) had pseudoaneurysm (n = 31, 46%), active bleed (n = 54, 80%), or both (n = 17, 25%).

Table 3

Faster RCNN detection results for active bleeding and pseudoaneurysm

	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
Active bleeding	0.84	88%	82%	91%	91%	83%
Pseudoaneurysm	0.79	83%	91%	78%	94%	71%