Title

Performance of the Winning Algorithms of the RSNA 2022 Cervical Spine Fracture Detection Challenge.

Journal

Radiology: Artificial Intelligence, 6(1)

Authors

Lee, Ghee
Flanders, Adam
Richards, Tyler
et al.

Publication Date

2024

DOI

10.1148/ryai.230256

Peer reviewed

Radiology: Artificial Intelligence

TECHNICAL DEVELOPMENT

Performance of the Winning Algorithms of the RSNA 2022 Cervical Spine Fracture Detection Challenge

Ghee Rye Lee, MMSc, MM • *Adam E. Flanders, MD* • *Tyler Richards, MD* • *Felipe Kitamura, MD, PhD* • *Errol Colak, MD* • *Hui Ming Lin, HBSc* • *Robyn L. Ball, PhD* • *Jason Talbott, MD, PhD* • *Luciano M. Prevedello, MD, MPH*

From the Department of Radiology, Ohio State University Wexner Medical Center, 395 W 12th Ave, Columbus, OH 43210 (G.R.L., L.M.P.); Department of Radiology, Thomas Jefferson University, Philadelphia, Pa (A.E.F.); Department of Radiology, University of Utah School of Medicine, Salt Lake City, Utah (T.R.); Dasalnova, Diagnósticos da América, São Paulo, Brazil (F.K.); Department of Diagnostic Imaging, Universidade Federal de São Paulo, São Paulo, Brazil (F.K.); Department of Medical Imaging, Unity Health Toronto, University of Toronto, Toronto, Canada (E.C., H.M.L.); The Jackson Laboratory, Bar Harbor, Maine (R.L.B.); and Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, Calif (J.T.). Received July 8, 2023; revision requested August 17; final revision received November 24; accepted December 12. **Address correspondence to** L.M.P. (email: *Luciano.prevedello@osumc.edu*).

Authors declared no funding for this work.

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2024; 6(1):e230256 • https://doi.org/10.1148/ryai.230256 • Content codes: MK CT AI

Purpose: To evaluate and report the performance of the winning algorithms of the Radiological Society of North America Cervical Spine Fracture AI Challenge.

Materials and Methods: The competition was open to the public on Kaggle from July 28 to October 27, 2022. A sample of 3112 CT scans with and without cervical spine fractures (CSFx) were assembled from multiple sites (12 institutions across six continents) and prepared for the competition. The test set had 1093 scans (private test set: $n = 789$; mean age, 53.40 years ± 22.86 [SD]; 509 males; public test set: $n = 304$; mean age, 52.51 years ± 20.73; 189 males) and 847 fractures. The eight top-performing artificial intelligence (AI) algorithms were retrospectively evaluated, and the area under the receiver operating characteristic curve (AUC) value, F1 score, sensitivity, and specificity were calculated.

Results: A total of 1108 contestants composing 883 teams worldwide participated in the competition. The top eight AI models showed high performance, with a mean AUC value of 0.96 (95% CI: 0.95, 0.96), mean F1 score of 90% (95% CI: 90%, 91%), mean sensitivity of 88% (95% CI: 86%, 90%), and mean specificity of 94% (95% CI: 93%, 96%). The highest values reported for previous models were an AUC of 0.85, F1 score of 81%, sensitivity of 76%, and specificity of 97%.

Conclusion: The competition successfully facilitated the development of AI models that could detect and localize CSFx on CT scans with high performance outcomes, which appear to exceed known values of previously reported models. Further study is needed to evaluate the generalizability of these models in a clinical environment.

Supplemental material is available for this article.

Every year in the United States, more than 1.5 million spine fractures occur, resulting in more than 17730 spinal cord injuries (1). Of these fractures, cervical spine fractures (CSFx) are common, with an overall incidence of 5.0%, and, if not treated properly, can lead to neurologic deterioration and death (2,3). CT is the primary imaging modality for CSFx detection due to its high diagnostic accuracy (4). However, inaccurate diagnoses can occur because of challenges with interpretation such as superimposed degenerative disease in elderly patients and congenital developmental anomalies, among others (5). Delayed diagnosis can result from high demand for imaging from emergency departments (6). The high demand also places a huge burden on radiologists (7,8). Accurate diagnosis and early intervention are key to reducing morbidity and mortality in patients with CSFx, and methods that might improve accuracy and efficiency in establishing the diagnosis can potentially enhance the overall workflow and quality of patient care.

Artificial intelligence (AI) is a promising tool for assisting radiologists in rendering a prompt, accurate diagnosis. Prior reports have investigated the use of AI to detect fractures in the hip, femur, humerus, wrist, and ankle (9). A recent systematic review of AI fracture detection in multiple body parts and using different modalities showed that AI has high diagnostic accuracy that in some cases has been shown to be comparable to nonradiologist interpretation and can assist clinicians and improve clinician performance (9). However, to our knowledge, very few studies have evaluated the performance of AI in detecting CSFx, and studies that have done so showed variable performance (10,11). One obstacle to conducting research in this area is the lack of publicly available, high-quality, diverse, annotated cervical spine scans gathered from multiple institutions.

For the past 7 years, the Radiological Society of North America (RSNA) has been sponsoring a series of AI challenges to address specific diagnostic problems in medical imaging. The RSNA Cervical Spine Fracture AI Challenge invited participants to develop AI models that can accurately detect, identify, and localize fractures in the cervical spine. A multi-institutional and multinational dataset consisting of 3112 annotated CT scans was prepared for the

Abbreviations

AI = artificial intelligence, AUC = area under the receiver operating characteristic curve, CSFx = cervical spine fractures, RSNA = Radiological Society of North America

Summary

The RSNA Cervical Spine Fracture AI Challenge promoted the development of artificial intelligence models that showed high performance in detecting cervical spine fractures on CT scans.

Key Points

- The mean area under the receiver operating characteristic curve value of the top eight algorithms of the RSNA Cervical Spine Fracture AI Challenge was 0.96 (95% CI: 0.95, 0.96), compared with the highest previously reported value of 0.85.
- The mean F1 score of the top eight algorithms of the competition was 90% (95% CI: 90%, 91%), compared with the highest reported value of 81% from previous literature for a machine learning algorithm.

Keywords

Cervical Spine, Fracture Detection, Machine Learning, Artificial Intelligence Algorithms, CT, Head/Neck

competition: A training set of 2019 scans were made available to the participants and the public, while 1093 were reserved for model-testing purposes. The goal of this competition was to stimulate innovation in the field of medical imaging AI as well as to foster collaboration among radiologists and data scientists to enhance diagnostic care. Herein, we evaluated the performance of the competition's top AI models.

Materials and Methods

Data Assembly and Curation

RSNA partnered with the American Society of Neuroradiology and the American Society of Spine Radiology to host the RSNA Cervical Spine Fracture AI Challenge on Kaggle *(https://www.kaggle.com/competitions/rsna-2022-cervical-spine-fracture-detection)*. The competition was open to the public internationally and was run from July 28 to October 27, 2022.

A ground truth dataset was created by collecting 3112 cervical spine CT studies from 12 institutions on six continents (five institutions in North America, one in South America, three in Europe, one in Africa, one in Asia, and one in Australia). These studies included a set of labels indicating which levels of the cervical spine (C1–C7) contained a fracture based on the radiology report. A quality review was performed by radiologists within the task force for each of the examinations. A group of 40 volunteer board-certified spine radiology specialists from the American Society of Neuroradiology and the American Society of Spine Radiology also annotated a subset of the CT studies (235 examinations) by drawing bounding boxes that encompassed fractures on axial images using a web-based annotation platform (MD.ai). A total of 1445 CT scans were positive for CSFx and 1667 were negative. The training set was composed of 2019 scans, with 961 scans (47.6%) containing at least one fracture and a total of 1444 fractures in the cervical spine. The positive cases had an average of 1.50 ± 0.82 (SD) fractures. The training data, which were freely downloadable by contestants, included the axial CT images and the vertebral levels positive for fracture, as well as a subset containing bounding boxes and segmentations that provided spatial information in the z-axis for locating fracture levels. The private test set was composed of 789 scans, with 362 scans (45.9%) containing at least one fracture and 672 fractures in total in the cervical spine. The positive cases had an average of 2.86 ± 1.04 fractures. The public test set was composed of 304 scans, with 122 scans (40.1%) containing at least one fracture and 175 fractures in total in the cervical spine.

The distribution of the CSFx in each dataset is shown in Table 1. The mean patient age was 53.65 years ± 21.57 in the training dataset, 52.51 years ± 20.73 in the public test set, and 53.40 years ± 22.86 in the private test set. The ratio of males to females was 1.72:1 in the training dataset, 1.64:1 in the public test, and 1.82:1 in the private test. To mitigate potential confounding, stratified random sampling based on site, age group, sex, and fracture level was used to partition the data into the training, public test, and private test sets (12). Details of the eligibility, exclusion, and inclusion criteria used in creating the dataset can be found in work by Lin et al (12).

AI Challenge Submission Evaluation

All the submitted algorithms were evaluated on two hidden datasets (ie, the test sets) that were available to the participants only through the challenge's platform; the performance results of the algorithms on each dataset were posted on a public and private leaderboard. The inference results of the algorithms on the public test set were posted publicly for the competitors to view their standing on this dataset. The inference results on the private dataset were available only to the competition hosts and were used to determine the final winner at the end of the contest. A weighted log loss scoring system was developed to score the algorithms based on their ability to detect any fractures in the cervical spine of a patient ("patient level") as well as at each vertebral level of the cervical spine. The binary weighted log loss function for label j on examination i was specified as follows:

$$L_{ij} = -w_j * \left[y_{ij} * log\left(p_{ij}\right) + \left(1 - y_{ij}\right) * log\left(1 - p_{ij}\right) \right].$$

Descriptions of the variables in the equation can be found in Appendix S1.

The binary weighted log loss for each of the labels (eight per patient) was then averaged across all labels and examinations in the test dataset to determine the final score. Penalties were given for a missed fracture at each vertebral level (C1–C7), and a heavier penalty (seven times higher) was given if the algorithm classified a patient with one or more fractures as negative for fracture. At the end of the competition, the eight participating teams with the highest scores were identified as the winners of the competition. RSNA traditionally awards eight to 10

Table 1: Distribution of Fractures by Cervical Spine Level in the Training Set, Public Test Set, and Private Test Set of the RSNA Cervical Spine Fracture AI Challenge

	Distribution of Fractures		
Cervical Spine Level	Training Set	Public Test Set	Private Test Set
C1	10.1 (146/1444)	14.9 (26/175)	10.2 (69/672)
C2	19.7 (285/1444)	18.3 (32/175)	15.5 (104/672)
C3	5.1 (73/1444)	4.6 (8/175)	8.5 (57/672)
C4	7.5 (108/1444)	4.0 (7/175)	12.2 (82/672)
C5	11.2 (162/1444)	9.7 (17/175)	15.8 (106/672)
C6	19.2 (277/1444)	17.1 (30/175)	19.0 (128/672)
C7	27.2 (393/1444)	31.4 (55/175)	18.8 (126/672)

Note.—Values shown as percentage, with proportion in parentheses.

Table 2: Weighted Log Loss Scores of the Eight Top-performing Algorithms

Rank	Log Loss Score	Architecture
1	0.2047	Segmentation + 2D CNN + RNN
2	0.2389	Segmentation + 2D CNN + RNN
3	0.2412	Segmentation + 2D CNN + RNN
4	0.2456	Channel-separated CNN
5	0.2580	2.5D CNN + 3D CNN
6	0.2631	3D CNN + 2D CNN with transformer
7	0.2634	3D U-Net + 3D CNN
8	0.2657	2.5D CNN + RNN

Note.—CNN = convolutional neural network, RNN = recurrent neural network, 3D = three-dimensional, 2D = two-dimensional, 2.5D = 2.5-dimensional.

top-performing algorithms, with the intent to recognize a large number of participants for their contributions to the field and give visibility to a large number of creative solutions. The winners were recognized at the 2022 RSNA Annual Meeting, and their algorithms have been made publicly available at *https://www.kaggle.com/competitions/rsna-2022-cervical-spine-fracture-detection/leaderboard.*

Evaluation of Top-scoring Algorithms and Statistical Analysis

Members of the challenge committee with access to both the submissions of the Kaggle participants and the private test dataset conducted a retrospective analysis to further evaluate the eight top-scoring models. Note that the private test dataset, curated for the purpose of the challenge, remains undisclosed to the public. Performance of the algorithms was evaluated using the area under the receiver operating characteristic curve (AUC) for each vertebral level and the patient level. F1 score, sensitivity, and specificity were measured by identifying individualized thresholds for each of the winning algorithms using the Youden *J* statistic to balance the true-positive rate and the false-positive rate. A total of 64 thresholds were identified to account for the seven vertebral levels and overall study performance for each of the top eight algorithms. JMP (SAS Institute) was used to analyze model performance, and GraphPad Prism (GraphPad Software) was used to run statistical analyses and generate figures. A Kruskal-Wallis test was used to compare the performance of the eight top-scoring models among the different vertebral levels. Two authors (G.R.L. and L.M.P.) performed the statistical analysis.

Results

AI Challenge Participants

A total of 1108 competitors composing 883 teams worldwide participated, with 12 871 entries submitted. The eight top-performing algorithms were selected based upon their weighted log loss performance in the private test set (Table 2). The architecture of each algorithm is shown in Table 2. Further details, including how the algorithms were tuned for hyperparameter optimization, can be found at *https://www.kaggle.com/competitions/rsna-2022-cervical-spine-fracture-detection/discussion*.

Performance of Top-scoring Algorithms

On the private dataset, the mean patient-level AUC across the top eight algorithms was 0.96 (95% CI: 0.95, 0.96). Figure 1 shows the median AUC values and IQRs at each vertebral level and at the patient level. Performance at C4 was significantly lower than that at C1 ($P < .05$), C2 ($P < .0001$), and C7 ($P < .01$).

The mean F1 score at the patient level was 90% (95% CI: 90%, 91%). Median F1 scores and IQRs are shown in Figure 2. Variable performance across the seven vertebrae was observed, with significantly lower performance at C4 compared with C2 ($P < .0001$) and C7 ($P < .001$).

The mean sensitivity and specificity at the patient level were 88% (95% CI: 86%, 90%) and 94% (95% CI: 93%, 96%), respectively. Median sensitivity and specificity values with corresponding IQRs are shown in Figure 3. Performance was variable across the cervical spine vertebrae, and sensitivity and specificity values at C4 were significantly lower than those at C2 ($P < .05$ and $P < .01$, respectively).

Discussion

The RSNA Cervical Spine Fracture AI Challenge was RSNA's seventh AI competition. In addition to fostering collaboration, these competitions accelerate discoveries in the field and generate innovative high-performing algorithms that are open-source for the advancement of AI research in radiology (13).

To our knowledge, the performance of the top eight algorithms in the RSNA Cervical Spine Fracture AI Challenge appears to exceed the previously reported study-level algorithm performance of individually trained models in the literature, with a

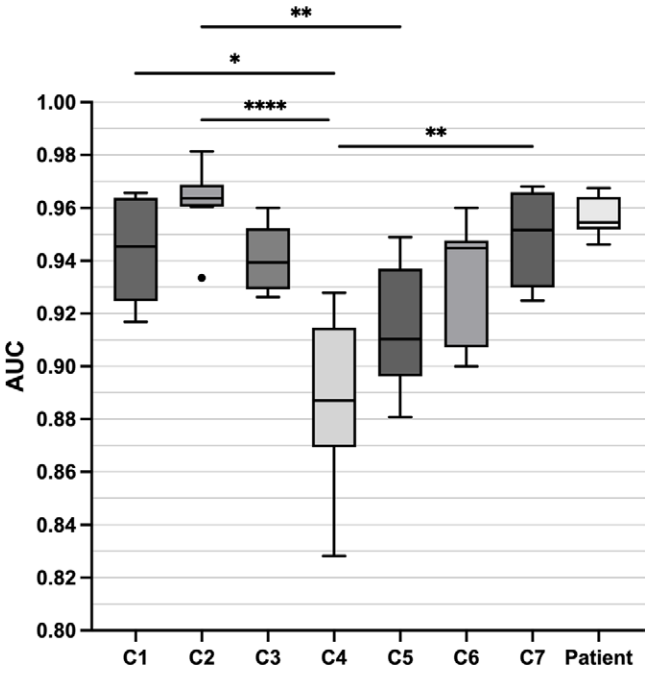

Figure 1: Box and whisker plot of the area under the receiver operating characteristic curve (AUC) values of the top eight algorithms at each vertebral level (C1–C7) and the patient level. Box borders indicate the 25th and 75th percentiles, midlines indicate the median, and whiskers indicate the minimum and maximum values. An outlier for C2 is shown as a dot. Statistical analysis comparing algorithm performance at the different vertebral levels was performed using the Kruskal-Wallis test followed by the Dunn multiple comparisons test. * = $P < .05$, ** = $P < .01$, *** = $P < .0001$.

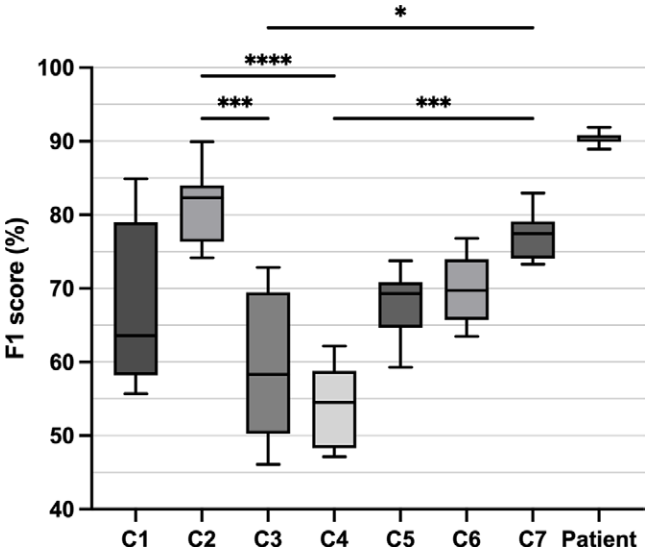

Figure 2: Box and whisker plot of F1 scores of the top eight algorithms at each vertebral level (C1–C7) and the patient level. Box borders indicate the 25th and 75th percentiles, midlines indicate the median, and whiskers indicate the minimum and maximum values. Statistical analysis comparing algorithm performance at the different vertebral levels was performed using the Kruskal-Wallis test followed by the Dunn multiple comparisons test. * = $P < .05$, ** = $P < .001$, *** = $P < .0001$.

mean AUC of 0.96 for the top eight algorithms versus 0.85 for the algorithm of Zhang et al (14), followed by 0.72 for the algorithm of Salehinejad et al (15). Like the RSNA competition's

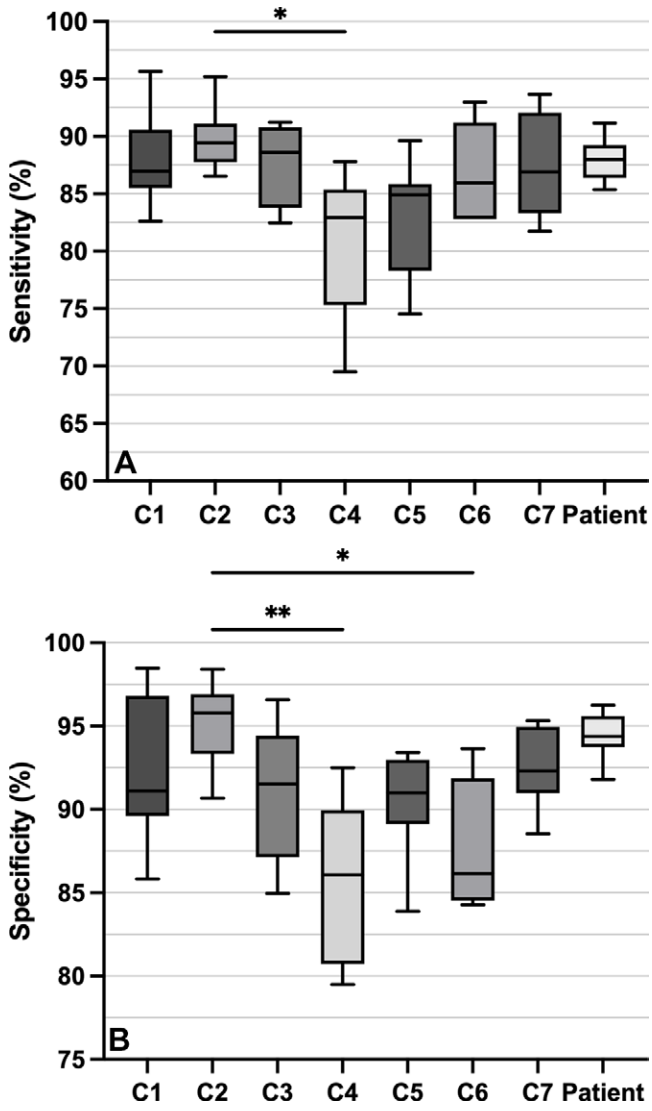

Figure 3: Box and whisker plots of **(A)** sensitivity and **(B)** specificity of the top eight algorithms at each vertebral level (C1–C7) and the patient level. Box borders indicate the 25th and 75th percentiles, midlines indicate the median, and whiskers indicate the minimum and maximum values. Statistical analysis comparing algorithm performance at the different vertebral levels was performed using the Kruskal-Wallis test followed by the Dunn multiple comparisons test. * = $P < .05$, ** = $P < .01$.

winning algorithms, the previous models in the literature were trained to identify CSFx on CT scans. Additionally, the majority of the winning algorithms used a 2.5-dimensional or three-dimensional convolutional neural network (CNN); similarly, Zhang et al (14) used a three-dimensional CNN with the feature pyramid network architecture, and Salehinejad et al (15) used a deep CNN with a bidirectional long-short term memory layer. However, there are some notable differences. For example, the RSNA competition's dataset had similar ratios of positive and negative scans in the training and test sets. However, the dataset in the study by Zhang et al (14), composed of 259 positive and 1088 negative scans, had a ratio of 1:4.5 in the training set and 1:2.6 in the validation set. Furthermore, their dataset contained scans collected from a single institution (14), while the RSNA competition's dataset was curated from multiple institutions. Thus, while the algorithms and datasets share some similarities,

they have their own idiosyncrasies, and algorithm performance should be compared with caution.

Another category of studies has focused on external testing of commercial models that detect CSFx. These studies used sensitivity and specificity instead of AUC to measure the performance of models. The highest sensitivity and specificity values found in the literature were 76% (95% CI: 68%, 83%) and 97% (95% CI: 95%, 98%), respectively (10). The top eight algorithms had a mean sensitivity of 88% (95% CI: 86%, 90%) and mean specificity of 94% (95% CI: 93%, 96%). To better compare the results of the top eight algorithms with that of previous algorithms in the literature in a way that balances the characteristics of sensitivity and specificity in a single metric, we calculated the F1 score: The mean F1 score of the top eight algorithms was 90%, compared with an F1 score of 81% for a previous algorithm by Small et al (10). It is worth noting that other studies used different methods to create their ground truth, which can affect model performance. For example, Zhang et al, similar to our approach, used radiologists' annotations as the reference standard, while Small et al used additional data such as from MRI scans, to detect missed fractures by looking at marrow edema. This could lead to more false negatives and decreased performance by models, relative to methods that used radiologists' interpretation as ground truth, because some fractures detectable on MRI scans may not be visible on CT scans (10,14).

Another finding of the study is the variable performance of the top eight algorithms at different vertebral levels. Performance was consistently significantly lower at C4 compared with other levels. The variability may be due to several factors. One factor may be the limited number of CT scans with a C4 fracture in the training set (108 of 1444 CSFx; 7.5%), as CSFx occur predominantly at the C2 and C7 levels (2). Zhang et al (14) reported a similar observation—that errors came from underpredicting fractures in less common locations, as well as overpredicting C2 fractures. Other previous algorithms experienced variable performance at different vertebral levels as well. Small et al (10) reported higher incidence of false-negative fractures along the lower cervical spine, and Voter et al (11) observed higher rates of errors for C5 fractures than for C2 fractures. Considering the notably small number of C4 vertebral fractures present in the dataset, it is plausible that participants received insufficient feedback regarding this specific vertebral segment from the public test set. This would have limited their capacity to make necessary adjustments to their models.

Some of the roadblocks in AI-driven CSFx detection research have been the lack of large, geographically diverse datasets and the limited collaboration at a multi-institutional level. Such roadblocks can limit the generalizability of AI models and their clinical applicability. In fact, in a meta-analysis by Kuo et al (9) of 42 studies that developed an Al algorithm for fracture detection, half of them were found to have high concern for generalizability due to the lack of external testing studies and the small sample sizes used for internal testing. It is critical for AI models to be generalizable in a clinical setting. For the RSNA Cervical Spine Fracture AI Challenge, RSNA compiled 3112 CT scans from 12 institutions on six continents—the most multi-institutional and multicontinental dataset made available to the public to our knowledge. The leading algorithms demonstrated strong performance on an extensive and varied dataset. However, it is essential to explore their generalizability in other medical centers or clinical situations, such as postoperative spine imaging, where they were not specifically designed to function. Future studies should further investigate ther generalizability.

While the outcome of the RSNA competition is promising, it is important to note that the research still remains in a very early stage, and more rigorous studies are needed to assess the potential utility of such algorithms in a clinical environment. The winning algorithms' performance on a new external dataset needs to be investigated. Also, their potential for clinical utility is limited due to the criteria used for dataset assembly, such as exclusion of postsurgical scans and strict inclusion of only noncontrast 1-mm-thick axial section images (12). Further limitations of the dataset are discussed by Lin et al (12). There are other areas for improvement, including creating datasets with a more balanced representation of fractures at all vertebral levels and developing models that consistently demonstrate high performance across the entire cervical spine. With these limitations in mind, practicing radiologists and data scientists can use this study's findings to develop datasets and algorithms that could potentially enhance the efficiency and workflow of patient care.

In conclusion, the eight top-performing algorithms of the RSNA Cervical Spine Fracture AI Challenge generated open-source algorithms with extremely high performance, appearing to surpass many of the previously reported AI algorithms in the literature. This outcome showcases the value of open competition, multi-institutional and multinational collaboration, and large, heterogeneous datasets. The successful outcome of the competition highlights the potential of such endeavors to spur innovation and move the field of AI research and patient care forward.

Author contributions: Guarantors of integrity of entire study, **G.R.L., F.K., L.M.P.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agree to ensure any questions related to the work are appropriately resolved, all authors; literature research, **G.R.L., E.C., L.M.P.**; clinical studies, **G.R.L., H.M.L.**; experimental studies, **J.T., L.M.P.**; statistical analysis, **G.R.L., F.K., E.C., R.L.B.**; and manuscript editing, all authors

Disclosures of conflicts of interest: **G.R.L.** No relevant relationships. **A.E.F.** On RSNA Board of Directors and reviewer for *Radiology: Artificial Intelligence.* **T.R.** No relevant relationships. **F.K.** Consultant for MD.ai; speaker for Sharing Progress in Cancer Care and Mayo Clinic; early career consultant to the editor for *Radiology*; associate editor for *Radiology: Artificial Intelligence*; and vice-chair of the Machine Learning Committee for the Society of Imaging Informatics in Medicine. **E.C.** No relevant relationships. **H.M.L.** No relevant relationships. **R.L.B.** Statistical consulting fee for assistance with RSNA Kaggle challenges; subcontract to The Jackson Laboratory from RSNA; grants or contracts from Medical Imaging and Data Resource Center; member of the RSNA Machine Learning Steering Committee; and chair elect of the American Statistical Association Statistical Consulting Section. **J.T.** No relevant relationships. **L.M.P.** Associate editor for *Radiology: Artificial Intelligence.*

References

1. Jain NB, Ayers GD, Peterson EN, et al. Traumatic spinal cord injury in the United States, 1993–2012. JAMA 2015;313(22):2236–2243.
2. Passias PG, Poorman GW, Segreto FA, et al. Traumatic fractures of the cervical spine: analysis of changes in incidence, cause, concurrent injuries, and complications among 488,262 patients from 2005 to 2013. World Neurosurg 2018;110:e427–e437.

3. Malik SA, Murphy M, Connolly P, O'Byrne J. Evaluation of morbidity, mortality and outcome following cervical spine injuries in elderly patients. Eur Spine J 2008;17(4):585–591.
4. Holmes JF, Akkinepalli R. Computed tomography versus plain radiography to screen for cervical spine injury: a meta-analysis. J Trauma 2005;58(5): 902–905.
5. Parizel PM, van der Zijden T, Gaudino S, et al. Trauma of the spine and spinal cord: imaging strategies. Eur Spine J 2010;19(Suppl 1):8–17.
6. Perotte R, Lewin GO, Tambe U, et al. improving emergency department flow: reducing turnaround time for emergent CT scans. AMIA Annu Symp Proc 2018;2018:897–906.
7. Mendoza D, Bertino FJ. Why radiology residents experience burnout and how to fix it. Acad Radiol 2019;26(4):555–558.
8. Jalal S, Parker W, Ferguson D, Nicolaou S. Exploring the role of artificial intelligence in an emergency and trauma radiology department. Can Assoc Radiol J 2021;72(1):167–174.
9. Kuo RYL, Harrison C, Curran TA, et al. Artificial intelligence in fracture detection: a systematic review and meta-analysis. Radiology 2022;304(1):50–62.
10. Small JE, Osler P, Paul AB, Kunst M. CT cervical spine fracture detection using a convolutional neural network. AJNR Am J Neuroradiol 2021;42(7):1341–1347.
11. Voter AF, Larson ME, Garrett JW, Yu JJ. Diagnostic accuracy and failure mode analysis of a deep learning algorithm for the detection of cervical spine fractures. AJNR Am J Neuroradiol 2021;42(8):1550–1556.
12. Lin HM, Colak E, Richards T, et al. The RSNA cervical spine fracture CT dataset. Radiol Artif Intell 2023;5(5):e230034.
13. Prevedello LM, Halabi SS, Shih G, et al. Challenges related to artificial intelligence research in medical imaging and the importance of image analysis competitions. Radiol Artif Intell 2019;1(1):e180031.
14. Zhang M, Kim L, Cheong R, et al. 218. Deep-learning artificial intelligence model for automated detection of cervical spine fracture on computed tomography (CT) imaging. In: 2019 AANS Annual Scientific Meeting. J Neurosurg 2019;131(1):2–116.
15. Salehinejad H, Ho E, Lin HM, et al. Deep sequential learning for cervical spine fracture detection on computed tomography imaging. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). Nice, France: IEEE, 2021; 1911–1914.