Received: 11 July 2021 | Revised: 22 January 2022 | Accepted: 23 January 2022

DOI: 10.1111/vru.13089

ORIGINAL INVESTIGATION

WILEY

Accuracy of artificial intelligence software for the detection of confirmed pleural effusion in thoracic radiographs in dogs

Thiago Rinaldi Müller[1] | Mauricio Solano[1] | Mirian Harumi Tsunemi[2]

[1]Department Clinical Sciences, Tufts University Cummings School of Veterinary Medicine, North Grafton, Massachusetts, USA

[2]Department of Biostatistics, São Paulo State University. R. Prof. Dr. Antônio Celso Wagner Zanin, São Paulo, Brazil

Correspondence
Thiago Muller, Clinical Sciences Department, Cummings School of Veterinary Medicine, 200 Westboro Rd, North Grafton, MA 01536, USA.
Email: Thiago.Muller@tufts.edu

Previous presentation or publication disclosure: The results of this study were presented at the 2021 ACVR Annual Scientific Meeting, Virtual meeting, October 20–23, 2021.

Reporting guideline disclosure: The following was used for guidance in the preparation of this manuscript: Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell 2020; 2(2):e200029.
https://doi.org/10.1148/ryai.2020200029

Abstract

The use of artificial intelligence (AI) algorithms in diagnostic radiology is a developing area in veterinary medicine and may provide substantial benefit in many clinical settings. These range from timely image interpretation in the emergency setting when no boarded radiologist is available to allowing boarded radiologists to focus on more challenging cases that require complex medical decision making. Testing the performance of artificial intelligence (AI) software in veterinary medicine is at its early stages, and only a scant number of reports of validation of AI software have been published. The purpose of this study was to investigate the performance of an AI algorithm (Vetology AI®) in the detection of pleural effusion in thoracic radiographs of dogs. In this retrospective, diagnostic case–controlled study, 62 canine patients were recruited. A control group of 21 dogs with normal thoracic radiographs and a sample group of 41 dogs with confirmed pleural effusion were selected from the electronic medical records at the Cummings School of Veterinary Medicine. The images were cropped to include only the area of interest (i.e., thorax). The software then classified images into those with pleural effusion and those without. The AI algorithm was able to determine the presence of pleural effusion with 88.7% accuracy ($P < 0.05$). The sensitivity and specificity were 90.2% and 81.8%, respectively (positive predictive value, 92.5%; negative predictive value, 81.8%). The application of this technology in the diagnostic interpretation of thoracic radiographs in veterinary medicine appears to be of value and warrants further investigation and testing.

KEYWORDS

imaging algorithms, machine learning, convolutional neural network, pleural effusion, thorax, radiograph, dog, canine, X-ray

1 | INTRODUCTION

Recently, technological innovations in artificial intelligence (AI) and machine learning (ML) software have been shown to be useful to medical and veterinary professionals.[1,2]. Artificial intelligence can analyze images to recognize objects of interest and distinguish certain features in a variety of medical conditions in different imaging modalities.[3–9]

Abbreviations: AI, artificial intelligence; CNN, convolutional neural network; DL, deep learning; ML, machine learning.

AI can be defined as a set of computer algorithms that attempt to simulate the problem-solving capacity and cognitive function of the human brain. AI software tries to recreate the cognitive function of the human brain.[10] ML is a form of AI that applies a specific algorithm to observational data points without the need for additional programming by software developers. Another form of AI is representation learning (RL). In RL, the algorithm detects certain features that aim to classify a certain condition. For example, in the context of this paper, the features could be the abnormal radiographic findings that a radiologist uses to conclude that there is pleural effusion. The latter is viewed as the condition. RL tends to become more accurate as more data become available. Deep learning (DP), another subset of AI, uses multiple algorithms to analyze data. To develop an efficient DP system, multiple images need to be available to the software to improve accuracy when compared to the "correct answer." The correct answer can be viewed as the gold standard, which in the field of AI, is known as the ground truth.[10,11] A complete explanation of how AI works is beyond the scope of this article. However, two premier articles on how AI works can be found in the list of references at the end of this paper.[10,11] There is considerable potential for AI to provide an initial interpretation of medical images in thoracic radiographs in human medicine. For instance, in human patients, AI algorithms have been created to evaluate thoracic radiographs for pneumothorax and tuberculosis or to detect other key extrathoracic findings, such as bony fractures[12–14]. Furthermore, AI has been demonstrated to aid in decision-making, predicting the mortality risk in patients with COVID-19 with high accuracy using the patients' physiological conditions, symptoms, preexisting conditions, and demographic information.[15] AI has also demonstrated the ability to recognize ischemic strokes on MRI.[16] Previous studies have shown that, for some specific tasks, AI systems are already outperforming humans in the detection of breast cancer using digital mammography.[17] The quantity of peer-reviewed articles published annually involving deep learning (DL) or convolutional neural networks (CNNs), another commonly used term that groups different AI methodologies, has increased exponentially in the past 5 years.[18,19] Advances in AI have yet to be assessed comprehensively in veterinary diagnostic imaging. To date, few studies testing the application of deep learning in veterinary medicine have been published.

In 2018, a study comparing two strategies to separate normal versus abnormal thoracic radiographs in dogs showed good performance of the AI software in detecting abnormalities and assisting general practitioners.[20] In 2020, a pilot study used AI techniques to screen thoracic radiographs for the detection of canine left atrial enlargement and compared the results with those of veterinary radiologist interpretations. The overall accuracy of the CNN algorithm and veterinary radiologists in that study was identical.[21] Studies have demonstrated classification accuracy in the detection of thoracic abnormalities such as generalized cardiomegaly[22], tracheal collapse, left atrial enlargement, alveolar pulmonary patterns, pneumothorax, and pulmonary masses in dogs.[20] Another report also showed that CNN was able to identify multiple thoracic lesions in canine as well as feline radiographs.[2] One of the main challenges of AI software is in obtaining a large quality data set for training. Another challenge faced by AI is training for rare diseases, as there are few examples from which AI software can learn, and validation can be difficult given the requirement for large sample sizes.[13] While recent publications evaluating thoracic radiographs in dogs using DL approaches[2,20,23] provided strong evidence for the utility and accuracy of AI, these studies compared radiology reports without confirmation of the various disease processes investigated, including pleural effusion. Whereas this is an accepted methodology, pitfalls may occur with this method, as abnormal findings that are the focus of any given study might not be systematically mentioned in the radiologist's reports. Generally, the radiology report can be subjective and may not be supported by later evidence, such as histopathology or surgery findings. Although the performance of AI algorithms in veterinary medicine is beginning to be tested, they are still lagging when compared to human medicine, especially in validating the accuracy of the technology.[18,19] The authors believe this new technology should be validated before its application in day-to-day veterinary medicine. However, commercially available products are already being offered to private veterinary practitioners and practices. Hence, validation of the technology has a sense of urgency.

A variety of diseases can lead to the abnormal accumulation of fluid in the pleural space.[24,25,26,27] Thoracic radiographs are arguably the most efficient method to detect and subjectively quantify pleural effusion (PE), and the commonly seen radiographic signs of PE have been well described. [28] Positioning, adipose tissue, accumulation, and diseases such as pleural nodules or masses can sometimes be misinterpreted as PE.[29]

The AI software selected for this study, Vetology AI Guardian (Vetology Innovations, San Diego, CA, USA), was created to produce diagnostic reports for the evaluation of canine thoracic radiographs. This software uses multiple CNN algorithms and has been developed using deep learning best practices. The testing and training involve comparison to the ground truth of an ACVR-certified veterinary radiologist's report. The AI-based software is directed to identify a variety of routinely assessed features in radiographic images of the canine thorax. The purpose of this study was to investigate the performance of an AI algorithm for the detection of confirmed PE in thoracic radiographs of dogs. The authors hypothesized that AI may have satisfactory accuracy as a screening method in the detection of PE.

2 | MATERIALS AND METHODS

2.1 | Experimental design and selection of subjects

The present retrospective, diagnostic, case–control study was conducted under the approval of the Foster Hospital for Small Animals Hospital Director's office. This includes approval for the use of the data. All the images were individually evaluated by two of the authors (T.M., diagnostic imaging resident, and M.S., ACVR-certified veterinary radiologist) with over 10 and 30 years of experience in small animal diagnostic imaging, respectively. The sample size in each group was defined by power analysis. A retrospective analysis of the electronic medical records at the Cummings School of Veterinary Medicine at Tufts University between January 2009 and October 2020 was reviewed. Inclusion criteria were availability of diagnostic quality orthogonal radiographic projections, radiographic reports that contained a diagnosis of PE, and confirmation of pleural fluid by thoracocentesis (17), thoracic ultrasound (13), surgery (9), CT (1) or MRI (1). None of the patients had manual inflation of their lungs or were under general anesthesia for acquisition of the radiographs.

Additional normal thoracic radiographs were obtained from patients without clinical evidence of thoracic disease and with no previous history of PE. A study was considered normal when no abnormalities were present in the lungs, cardiovascular structures, pleural space, or mediastinum. These patients were presented to preoperatory examination or preanesthesia examination with unrelated thoracic abnormalities such as cervical spinal pain, spinal cord lesions and dental disease.

Before all cases were submitted for AI analysis, the authors (TM and MS) jointly evaluated each of the radiographs and, by consensus, graded the degree of severity of the pleural effusion as mild, moderate, or severe. Patients without pleural effusion and a normal thorax were assigned to group 1. The confirmed cases of pleural effusion were assigned to group 2.

2.2 | Data recording and analysis

Digital radiographs were received by the Vetology AI software in standard DICOM format. Image processing included the application of intensity normalization, denoising, and gamma correction to the extracted images. The images were cropped to include only the area of interest (i.e., thorax). The software then classified images into those with pleural effusion and those without. The CNN was trained on approximately 2000 images of pleural effusion and approximately 2000 images of normal patients on the TensorFlow platform. The normal set did not contain other diseases. TensorFlow is an open-source platform that is used for machine learning. The platform provides a user interface for executing AI algorithms.[30] A broad range of digital radiographs from clinical cases with varied canine breeds, ages, geographies, and digital X-ray systems were used to source training data representative of diverse real-world cases. The training set of images was selected and labeled for disease status by technicians under the supervision of board-certified radiologists. A k-fold ($k = 3$) cross-validation algorithm assessed model performance. The CNN architecture is VGG16.[31] The CNN is a binary classifier that provides a probability of pleural effusion. The specific threshold was selected prior to this study based on internal testing.

2.3 | Statistical analysis

The results generated by the AI software for groups 1 and 2 were statistically evaluated. Statistical significance was considered for a P-value < 0.05. Efficiency in discriminating effusion was assessed by statistics on sensitivity, specificity, predictive values, and accuracy.[32,33] Confidence intervals (CIs) of 95% of the agreement and efficiency parameters were constructed. All statistical analyses were performed by an independent statistician (MT) using a statistical software package (R Core Team, 2020. Vienna, Austria). Factors such as age, weight, and breed were not included in the statistical analysis given that the software was trained to recognize and interpret varied canine breeds of different ages.

3 | RESULTS

All studies were produced using commercially available DR equipment (Canon Digital Radiography Systems, CXDI 17 × 17 flat panel detector and image postprocessing software Sound Smart DR, Carlsbad, CA) with exposure techniques that varied according to the thickness of the animal, with kVp ranging from 80–90 and mAs from 3.5–8.0. The DR equipment uses grid suppression software. The inclusion criteria were met by 62 dogs of different breeds. The median age of dogs in the normal group was 10 years (range 3 to 14 years), and in the pleural effusion group, it was 12 years (range of 4 to 24 years). The median weight of dogs in the normal group was 16 kg (range 6.0–47.7 kg). The median weight of dogs in the pleural effusion group was 32 kg (range 7.5–55.0 kg). Dog breeds included American Pit Bull Terrier, American Staffordshire Terrier, Beagle, Border Collie, Boston Terrier, Chihuahua, Cocker Spaniel, French Bulldog, German Shepherd, Goldendoodle, Grayhound, Labrador, Maltese, Miniature Poodle, mixed breed, Portuguese Water Dog, Pug, Schnauzer, Scottish Deer Hound, Siberian Husky and Tibetan Terrier.

Forty-one included dogs had confirmed pleural effusion, and 21 dogs had normal thoracic radiographs. A total of 173 images were evaluated from the 62 patients. This included 62 right lateral images, 49 left lateral images, and 62 ventrodorsal images. Evaluators were aware of the animal's presenting complaint for group 1 (n: 21). This group included dogs undergoing a preanesthesia workup with unrelated thoracic abnormalities. Group 2 (n: 41) included animals with a history of traumatic metastatic disease, cardiomyopathy, thoracic neoplasia, and lung lobe torsion.

The final diagnosis leading to pleural effusion included diaphragmatic hernia (3), chylothorax (4), lung neoplasia (6), heart base mass (1), lung lobe torsion (4), rib neoplasia (4), mediastinal mass (6), metastatic disease (6), fungal disease with pleuritis (1), and heart failure (6). The authors classified the confirmed pleural effusion cases as having mild (9), moderate (22), and severe (10) volumes of pleural effusion.

The AI software detected pleural effusion in 37/41 of the confirmed pleural effusion cases and correctly classified 18/21 of the normal cases.

The sensitivity of the AI model to recognize pleural effusion was 90.2% (95% CI 0.768–0.972), and it showed a specificity of 85.7% (95% CI 0.636–0.969). The positive and negative predictive values of the AI software for predicting pleural effusion were 92.5% (95% CI 0.796–0.984) and 81.8% (95% CI 0.597–0.948), respectively. The diagnostic accuracy of the AI software was 88.7% (95% CI 0.781–0.953).

Examples of radiographs correctly and incorrectly classified by the AI model are reported in Figures 1 and 2, respectively.

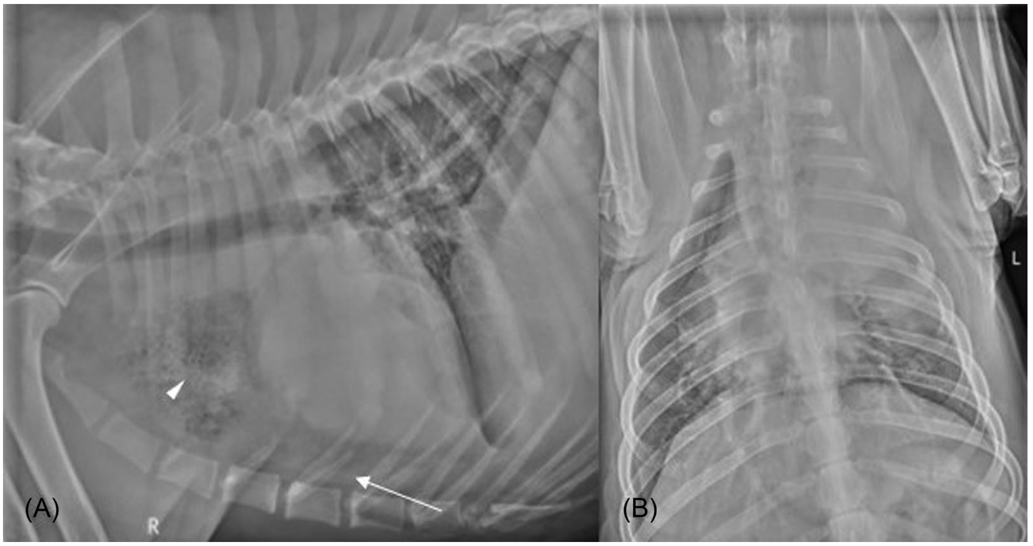

FIGURE 1 Examples of radiographs correctly classified by the AI model as having pleural effusion. A, Right lateral (kVp 80, mAs 6.5) and B, ventrodorsal (kVp 90, mAs 6.5) radiographic projections of a dog with mild and unilateral signs of pleural effusion. There is a vesicular pattern in the cranioventral aspect of the lung fields on the lateral projection (arrowhead). The free fluid accumulates ventral to the heart, increasing the radiographic opacity of the mediastinal fat (arrow). This dog had confirmation of left cranial lung lobe torsion and pleural effusion on surgery

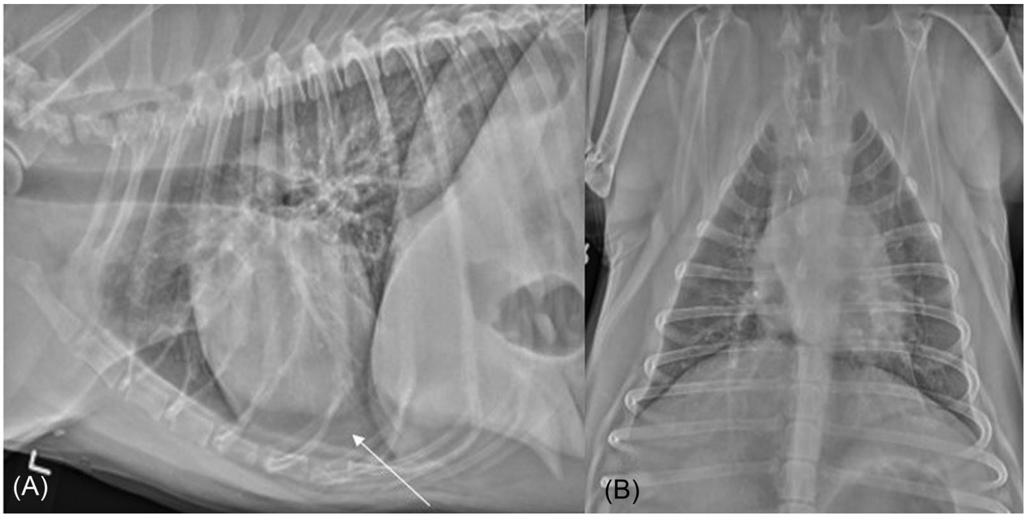

FIGURE 2 Examples of radiographs incorrectly classified by the AI model as having pleural effusion. A, Left lateral (kVp 80, mAs 6.5) and B, ventrodorsal (kVp 90, mAs 6.5) radiographic projections of a dog without pleural effusion. Note the fat opacity ventral to the heart (arrow). The excessive fat accumulation in the mediastinum is mistakenly assigned a yes-effusion value, likely as the radiographic density of fat allowing visualization of the apex of the heart was not considered by the algorithm

The AI model failed to identify four cases of pleural effusion that were previously classified by the two authors as mild (3) (Figure 3) and moderate (1) (Figure 4).

4 | DISCUSSION

The authors use the term validation of the software specifically to determine the sensitivity and specificity of the AI algorithm to detect pleural effusion in test groups A and B. This is limited to a population of 41 abnormal and 21 normal subjects. It is worth highlighting in this discussion that there is a separate and independent data set used by the developers of any AI software to train their software. The latter is often a larger data set in the thousands that contain a myriad of pathological processes, including pleural effusion. However, just because AI software has been trained with a larger data set, using a myriad of pathological processes does not mean that the software has been validated against an external independent data set. The results of

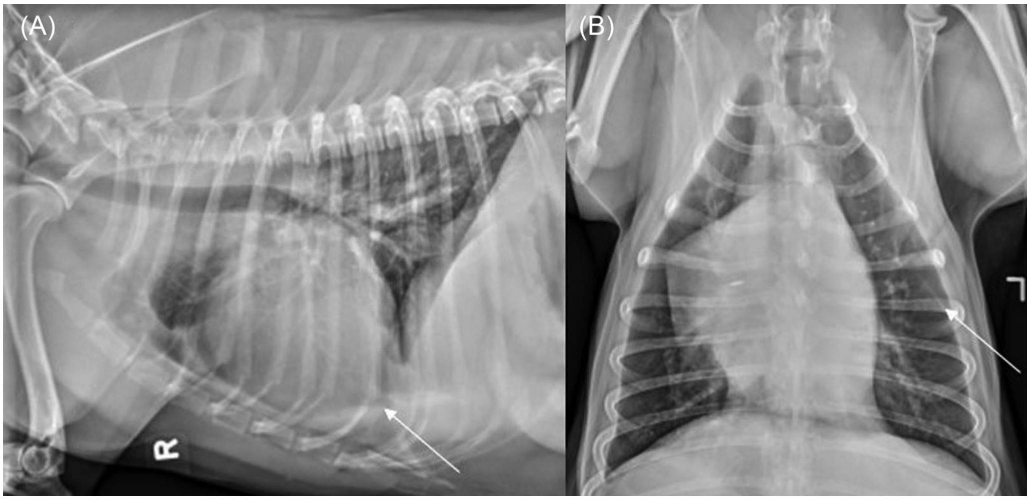

FIGURE 3 Examples of radiographs incorrectly classified by the AI model as not having pleural effusion. A, Right lateral (kVp 80, mAs 4.5) and B, ventrodorsal (kVp 90, mAs 4.5) radiographic projections of a dog with mild and unilateral pleural effusion. Arrows indicate fluid lines indicating a minimal amount of fluid within the pleural space. The amount of fluid in the lateral projection obscures the apex of the heart. This dog had confirmation of a heart base mass by echocardiography. Pleural effusion was confirmed by thoracic ultrasound

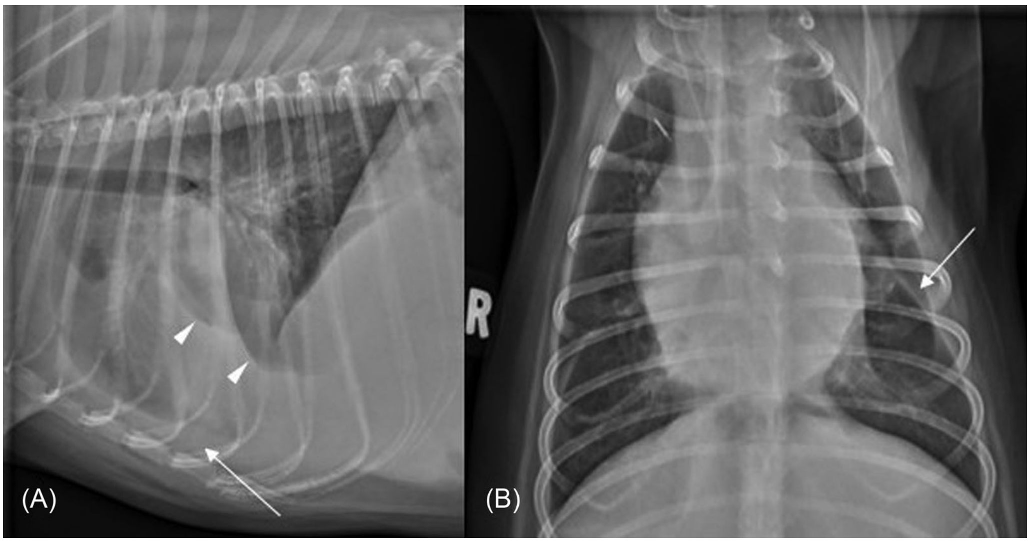

FIGURE 4 Examples of radiographs incorrectly classified by the AI model as not having pleural effusion. A, Right lateral (kVp 80, mAs 6.5) and B, ventrodorsal (kVp 90, mAs 6.5) radiographic projections of a dog with moderate and bilateral pleural effusion (arrows). This dog had confirmation of pleuritis by thoracentesis due to coccidioidomycosis. The software did not recognize the soft tissue opacity obscuring portions of the heart as free pleural fluid. The lung borders are highlighted by the fluid (scalloping), as noted by the arrowheads in the lateral view

this study suggest that the application of the AI model could assist veterinarians in the detection of pleural effusion in thoracic radiographs and possibly serve as a screening tool for triaging a patient. Remarkably, a similar sensitivity (91%) and specificity (91%) was obtained by other authors[9] in detecting pleural effusion in human patients with the use of CNN. A similar performance was also found by CNN in veterinary patients.[20,23] However, these previous studies relied only on the radiologist report, and confirmation of the disease was not obtained. Using exclusively the radiologist's report as the ground truth for this analysis can create some pitfalls[9], as the accuracy of the radiologist in detecting pleural effusion is estimated to vary from 67% to 92%.[34,35] There are no reports evaluating the accuracy of a board-certified radiologist in detecting pleural effusion in the veterinary literature. Pleural effusion can be misinterpreted in cases of pleural masses, adipose tissue accumulation within the mediastinum, and positioning of the patient. In other words, the authors speculate that the radiographic report is not the ideal ground truth, as it is subject to the accuracy of a radiologist. The test group in this work (confirmed cases of pleural effusion) was selected using various methods, such as thoracentesis or surgery. While the authors reviewed the abnormal cases and by consensus agreed that all subjects in the test group had radiographic evidence of pleural effusion, the software performance was tested against the proven presence of pleural effusion.

The AI model used, tested on a small database of radiographic images, showed a high classification accuracy in the detection of pleural effusion in dogs and, as importantly, in the detection of normal thoracic radiographs.

One false-negative case of pleural effusion exhibited subtle changes. This case was classified as mild by the authors. This highlights the fact that the CNN may also encounter difficulties in correctly identifying subtle sets of abnormalities and assigning the incorrect output, similar to what a radiologist may encounter in practice. On the other hand, another false negative case had a severe volume of pleural effusion. The cause of this misdiagnosis remains unclear, as it was deemed a case of severe pleural effusion by the authors.

The aim of the present study was to test the ability of the AI model to detect PE as an isolated abnormality and not to determine the cause or classify the severity of PE. The software was programmed to automatically classify canine thoracic radiographs as either no-pleural effusion or Yes-pleural effusion. In other words, the ability of the software to discern pleural effusion as the result of right heart failure or lung lobe torsion was not tested. This requires collecting a larger set of a specific disease process to achieve statistical significance. Moreover, even with a training data set in the thousands, the developers of the software do not have enough data points to classify effusion as mild, moderate, and severe. Further grading the severity of the pleural effusion will improve the analysis and understanding of the software limitations, but it remains beyond the capabilities of the software.

The authors argue that any AI software that is being developed reflects the accuracy of the data it was trained against, which was created by specialists. This includes the ability of the imaging experts to accurately provide the key features that are present in any given pathological process. This is one of the key issues that should be at the forefront of veterinary practitioners in successfully understanding the role of AI in day-to-day practice. While validation of the software to detect pleural effusion has been shown, correlating the abnormal radiographic finding of pleural effusion with other abnormal radiographic findings to determine the etiology of the effusion, e.g., a pulmonary mass, diaphragmatic hernia, or heart disease, was not possible due to the small number of cases.

It is the opinion of the authors that AI software as applied to the veterinary imaging field is here to stay. However, validating such technology is essential to ensure the correct use of the technology in day-to-day practice. Although there is no evidence that AI can make complex problem-solving decisions typical of a radiologist as reflected in radiology reporting, AI shows promise to at least be used as a screening tool for general practitioners. The user of such technology must, however, have proper expectations and a clear understanding of the current pitfalls and advantages of such technology.

ACKNOWLEDGMENTS

The authors would like to acknowledge Vetology for providing the AI software for testing.

LIST OF AUTHOR CONTRIBUTIONS

Category 1

(a) Conception and Design: Solano, Müller, Tsunemi.
(b) Acquisition of Data: Müller
(c) Analysis and Interpretation of Data: Solano, Müller, Tsunemi

Category 2

(a) Drafting the Article: Müller
(b) Revising Article for Intellectual Content: Solano, Müller, Tsunemi

Category 3

(a) Final Approval of the Completed Article: Solano, Müller, Tsunemi

Category 4

(a) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: Solano, Müller, Tsunemi

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

ORCID

Thiago Rinaldi Müller https://orcid.org/0000-0002-7494-8588

REFERENCES

1. Ci_han P, Gokçe E, Kalipsiz O. A review of machine learning applications in veterinary field. *Kafkas Univ Vet Fak Der* 2017;23:673–80.
2. Boissady E, de La Comble A, Zhu X, Hespel A-M. Artificial intelligence evaluating primary thoracic lesions has an overall lower error rate compared to veterinarians or veterinarians in conjunction with the artificial intelligence. *Vet Radiol Ultrasound*. 2020;1–9
3. Vinicki K, Ferrari P, Belic M, Turk R. Using convolutional neural networks for determining reticulocyte percentage in cats. http://arxiv.org/abs/1803.04873.*arXiv*: 1803.04873; 2018.
4. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60–88.
5. Vetology launches pet X-ray-reading ai software [accessed 20.01.21], https://www.veterinarypracticenews.com/vetology-launches-pet-X-ray-reading-aisoftware/; 2018.
6. Artificial intelligence powers new X-ray software [accessed 20.01.21], https://todaysveterinarybusiness.com/artificial-intelligence-powers-new-X-raysoftware/; 2018
7. Chassagnon G, Vakalopoulou M, Paragios N, Revel MP. Artificial intelligence applications for thoracic imaging. *Eur J Radiol* 2020;123:1-6.
8. Winkel DJ, Heye T, Weikert TJ, Boll DT, Stieltjes B. Evaluation of an AI-based detection software for acute findings in abdominal computed tomography scans toward an automated work list prioritization of routine CT examinations. *Invest Radiol* 2019;54:55-9
9. Cicero M, Bilbily A, Colak E, Dowdell T, Gray B, Perampaladas K, Barflett J. Training and validating a deep convolutional neural network

for computer-aided detection and classification of abnormalities on frontal chest radiographs. *Invest Radiol* 2017;52:281-7.
10. Chartrand G, Cheng PM, Voronstov E, Drozdzal M, Turcotte S, Pal CJ, Kadoury S, Tang A. Deep learning: a primer for radiologists. *Radiographics* 2017;37:2113-2131.
11. Choy G, Khalizadeh O, Michalski M, Do S, Samir AEE, Pianykh OS, Geis JR, Pandharipande PV, Brink JA, Dreyer KJ. Current applications and future impact of machine learning in radiology. *Radiology* 2018;288:318-328.
12. Burns JE, Yao J, Muñoz H, Summers RM. Automated detection, localization, and classification of traumatic vertebral body fractures in the thoracic and lumbar spine at CT. *Radiology*. 2016;278(1):64-73.
13. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*. 2017;284(2):574-582.
14. Taylor AG, Mielke C, Mongan J. Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: a retrospective study. *PLoS Med*. 2018;15(11):e1002697
15. Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health*, 2021; 20. https://doi.org/10.1016/j.smhl.2020.100178.
16. Maier O, Schröder C, Forkert ND, Martinetz T, Handels H. Classifiers for ischemic stroke lesion segmentation: a comparison study.*PLoS One*. 2015;10(12):e0145118
17. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al. Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. *J Natl Cancer Inst*. 2019;111(9):916-922. https://doi.org/10.1093/jnci/djy222
18. Sahiner B, Pezeshk A, Hadjiiski LM, et al. Deep learning in medical imaging and radiation therapy. *Med Phys*. 2019;46:e1-e36
19. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88
20. Yoon Y, Hwang T, Lee H. Prediction of radiographic abnormalities by the use of bag-of-features and convolutional neural networks. *Vet J*. 2018;237:43-48. https://doi.org/10.1016/j.tvjl.2018.05.009
21. Li S, Wang Z, Visser LC, Wisner ER, Cheng H. Pilot study: Application of artificial intelligence for detecting left atrial enlargement on canine thoracic radiographs. *Vet Radiol Ultrasound*. 2020;61:611–618. https://doi.org/10.1111/vru.12901
22. Burti S., Longhin Osti V., Zotti A., Banzato T.. Use of deep learning to detect cardiomegaly on thoracic radiographs in dogs. *The Veterinary Journal*. 2020:262
23. Banzato T, Wodzinski M, Burti S, et al. Automatic classification of canine thoracic radiographs using deep learning. *Nature Scientific Reports* (2021) 11:3964 | https://doi.org/10.1038/s41598-021-83515-3
24. O'Brien PJ and Lumsden JH. Cytologic examination of body cavity fluids. *Semin Vet Med Surg (Small Anim)* 1988; 3: 140–156.
25. Dempsey SM and Ewing PJ. A review of the pathophysiology, classification, and analysis of canine and feline cavitary effusions. *J Am Anim Hosp Assoc* 2011; 47: 1–11.
26. Noone KE. Pleural effusions and diseases of the pleura. *Vet Clin North Am Small Anim Pract* 1985; 15: 1069–1084.
27. Forrester SD, Troy GC and Fossum TW. Pleural effusions – pathophysiology and diagnostic considerations. *Comp Cont Ed Pract Vet* 1988; 10: 121–136.
28. Lynch KC, Oliveira CR, Matheson JS, et al: Detection of pneumothorax and pleural effusion with horizontal beam radiography, *Vet Radiol Ultrasound* 53(1):38, 2012.
29. Thrall DE. Canine and Feline Pleural Space. In: Thrall DE (7th ed): *Textbook of veterinary diagnostic radiology*. St. Louis, MO: Saunders Elsevier, 2018; 670–678.
30. Abadi M, Barham P, Chen J, et al. "TensorFlow: A System for Large-Scale Machine Learning TensorFlow: A system for large-scale machine learning," in 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16), 2016, pp. 265–284
31. Simonyan K, Zisserman A. "Very deep convolutional networks for large-scale image recognition." Presented as a conference paper international: conference on learning representation ICLR, 2015, San Diego., CA: arXiv preprint arXiv:1409.1556 (2014).
32. Altman DG, Machin D, Bryant TN, and Gardner MJ (2000). *Statistics with Confidence*, second edition. British Medical Journal, London, pp. 28-29.
33. Collett D (1999). *Modeling Binary Data*. Chapman & Hall/CRC, Boca Raton Florida, pp. 24.
34. Ruskin J, Gurney JW, Thorsen MK, Goodman LR. Detection of pleural effusions on supine chest radiographs. *American Journal of Radiology*. 1987; 148: 681–683.
35. Kitazono MT, Lau CT, Parada AN, Renjen P, Miller WT Jr. Differentiation of pleural effusions from parenchymal opacities: Accuracy of bedside chest radiography. *American Journal of Radiology*. 2010; 194: 407–412.

How to cite this article: Müller TR, Solano M, Tsunemi MH. Accuracy of an artificial intelligence software for detection of confirmed pleural effusion in thoracic radiographs in dogs. *Vet Radiol Ultrasound*. 2022;1–7. https://doi.org/10.1111/vru.13089