Received: 20 May 2021 | Revised: 8 December 2021 | Accepted: 8 December 2021

DOI: 10.1111/vru.13062

ORIGINAL INVESTIGATION

WILEY

Comparison of artificial intelligence to the veterinary radiologist's diagnosis of canine cardiogenic pulmonary edema

Eunbee Kim[1] | Anthony J. Fischetti[1] | Pratheev Sreetharan[2] | Joel G. Weltman[3] | Philip R. Fox[4]

[1] Department of Diagnostic Imaging, The Animal Medical Center, New York, New York, USA

[2] Vetology Innovations, LLC., San Diego, California, USA

[3] Department of Emergency and Critical Care, The Animal Medical Center, New York, New York, USA

[4] Department of Cardiology, The Animal Medical Center, New York, New York, USA

Correspondence
Anthony Fischetti, DVM, MS, DACVR, The Animal Medical Center – Diagnostic Imaging, 510 E 62nd Street, New York, NY 10065, USA.
Email: anthony.fischetti@amcny.org

Abstract

Application of artificial intelligence (AI) to improve clinical diagnosis is a burgeoning field in human and veterinary medicine. The objective of this prospective, diagnostic accuracy study was to determine the accuracy, sensitivity, and specificity of an AI-based software for diagnosing canine cardiogenic pulmonary edema from thoracic radiographs, using an American College of Veterinary Radiology-certified veterinary radiologist's interpretation as the reference standard. Five hundred consecutive canine thoracic radiographs made after-hours by a veterinary Emergency Department were retrieved. A total of 481 of 500 cases were technically analyzable. Based on the radiologist's assessment, 46 (10.4%) of these 481 dogs were diagnosed with cardiogenic pulmonary edema (CPE+). Of these cases, the AI software designated 42 of 46 as CPE+ and four of 46 as cardiogenic pulmonary edema negative (CPE−). Accuracy, sensitivity, and specificity of the AI-based software compared to radiologist diagnosis were 92.3%, 91.3%, and 92.4%, respectively (positive predictive value, 56%; negative predictive value, 99%). Findings supported using AI software screening for thoracic radiographs of dogs with suspected cardiogenic pulmonary edema to assist with short-term decision-making when a radiologist is unavailable.

KEYWORDS
Artificial intelligence, congestive heart failure, convolutional neural network, myxomatous mitral valve disease, thoracic radiograph

1 | INTRODUCTION

Heart disease occurs in approximately 10% of dogs visiting primary care veterinary clinics.[1] The most common acquired cardiac disorder, myxomatous mitral valve degeneration (MMVD), affects approximately up to 75% of these cases.[1,2] MMVD frequently results in congestive heart failure (CHF) characterized by acute cardiogenic pulmonary edema (left-sided CHF; CPE) and respiratory distress—a rapidly progressive medical emergency that is fatal if misdiagnosed or if treatment is delayed. Thoracic radiography has long been considered the gold standard for diagnosing CPE.[3,4] Thoracic radiography remains a widely available, non-invasive, rapid test to determine the presence of CPE, especially when combined with the medical history and clinical presentation. Nevertheless, diagnosing CPE from thoracic radiographs can be challenging when patients display atypical radiographic features or subtle findings. While Doppler echocardiography and point of care lung ultrasonography can aid in the diagnosis of canine and feline CPE,

Abbreviations: ACVR, American College of Veterinary Radiology; AI, artificial intelligence; CHF, congestive heart failure; CI, confidence interval; CNN, convolutional neural network; CPE, cardiogenic pulmonary edema negative; CPE+, cardiogenic pulmonary edema positive; ECVR, European College of Veterinary Radiology; ML, machine learning; MMVD, myxomatous mitral valve disease; NPV, negative predictive value; OR, odds ratio; PPV, positive predictive value; RGT, report generation time.

Previous presentation or publication disclosure: The content of this paper has not been previously published, in full or in abstract form, nor has it been presented at a scientific meeting or congress.

these techniques require specific equipment, specialized training, and are user dependent.[4,5]

Artificial intelligence (AI) is a broad term used to describe computer algorithms that perform tasks to mimic cognitive function, such as learning or problem solving. Within artificial intelligence are more specific fields including machine learning (ML), representation learning, and deep learning. Machine learning refers to AI that uses observational data without specific programming by narrowing the algorithm's parameters to optimize the relationship between the input and output.[6–8] In representation learning, the computer algorithm learns features to facilitate classification with performance, generally improving as more data are added.[6–8] Deep learning is a subfield utilizing multiple layers of algorithms to analyze data. To train the deep learning systems, hundreds of thousands of images are presented to the computer for the software to then guess, compare, and re-calibrate to improve its accuracy compared to the ground truth. Ground truth is a term that defines what the computer should consider as the correct answer.[8]

Growing interest in AI-based imaging software has been stimulated by the expansion of telemedicine, greater volumes of diagnostic images, and demand for more efficient report generation. In human medicine, AI systems have been used effectively to detect various medical conditions by multiple diagnostic imaging modalities.[9–11] Recently, deep learning has been applied to veterinary medicine for the detection of lesions from feline and canine thoracic radiographs and the presence of left atrial enlargement in dogs.[12–14] Furthermore, the American College of Veterinary Radiology (ACVR) and European College of Veterinary Radiology (ECVDI) created the AI Education and Development Committee to evaluate the current and future impact of AI within the speciality.[15] The role of AI in veterinary radiology is rapidly growing and evolving but the accuracy of the various AI systems and clinical applications remains undetermined.

A commercially available AI-based software (Vetology Innovations, San Diego, CA, USA) has been developed to review canine thoracic radiographs and generate diagnostic reports. The software relies on convolutional neural networks (CNNs), a type of deep learning network that uses layers of algorithms. Convolutional neural networks contain connected nodes designed to mimic neurons in the brain and are commonly used for deep learning imaging analysis.[6] This software was developed following deep learning best practices with proprietary training, testing, and comparison to the ground truth of a board-certified veterinary radiologist's report. Multiple CNN models incorporated into this AI-based software are trained to detect various features of radiographic images of the canine thorax. The software automatically produces an output report containing a collection of findings. In particular, these output reports determine the presence or absence of CPE associated with left-sided CHF from canine radiographic images.

The objective of this study was to determine the accuracy, sensitivity, and specificity of an AI-based software for diagnosing canine cardiogenic pulmonary edema from thoracic radiographs, using an ACVR-certified veterinary radiologist's interpretation as the reference standard.

2 | MATERIALS AND METHODS

2.1 | Experimental design and subject selection

The study was a prospective, diagnostic accuracy design. Canine thoracic radiographs made after-hours by the Emergency Department at the Animal Medical Center between January 15, 2020 and June 14, 2020 were considered for inclusion. Approval from the Animal Medical Center's Institutional Animal Care and Use Committee was obtained. All thoracic radiographs made between the hours of 18:00 and 8:00 were considered eligible, during which time the Diagnostic Imaging Department was unavailable. The inclusion criteria for this study were determined by an ACVR-certified veterinary radiologist (A.F.): the radiographic examination was required to have at least one view of the thorax. Age, breed, sex, body weight, and number of radiographic images were recorded for each enrolled patient as directed by the veterinary radiologist (A.F.).

2.2 | Data recording

All eligible thoracic radiograph DICOM files were electronically transferred at 8:00 each morning to the AI software server (Vetology Innovations, San Diego, CA, USA) independent of patient history and radiologist interpretation. The specific CNN architecture, training methods, and training datasets were key performance drivers of CNN-based software algorithms.[16] The selected AI software evaluated radiographs for the presence of a range of diseases and produced a plain language report as output. We specifically examined a single feature of this output: the determination of either a CPE+ or CPE− state. The algorithm to determine CPE disease was previously trained on radiographs of varied canine breeds, ages, geographies, and digital X-ray systems from diverse real-world cases with a broad range of CPE severities. The software performed whole image analysis and crafted plain language radiology reports based on internal CNN model results. If the CPE+ likelihood exceeded internally defined thresholds, the software indicated CPE+ status within the output report, thus providing the end user with a classification between CPE+ and CPE−. The selected AI software ran on individual servers containing an 8-core computer (Intel Xeon CPU E5-2660 v3 @ 2.60 GHz, 48GB of DDR4 2666 MHz RAM) and solid-state drives. A load balancer distributed AI processing work between four servers for this trial. The AI server recorded a timestamp upon initiating image processing from a fully received DICOM and a second timestamp when the software generated a report. Lastly, the times when the study request was received by the AI software's server and when the AI-generated report was electronically returned were recorded to calculate the report generation time (RGT). The RGT value was calculated as the difference between these two timestamps and thus did not include data transit times to and from the servers.

The AI software inspected the species field of the DICOM labels to reject any indicated as noncanine. All other included metadata in the

DICOM labels were discarded. The anonymized radiograph image data were extracted and forwarded for automated image cropping. The AI model was applied to assess these auto-cropped images. If an image could not successfully be auto-cropped, the AI software company's human technicians intervened to troubleshoot and provide an image that could be evaluated by the AI software. After the images were evaluated, the AI software then translated the results into a single plain language radiograph report. Finally, the report was automatically sent back to the requesting clinician. By reading these reports, we extracted a binary classification of each case between CPE+ and CPE− disease states attributed to the AI software. Images were excluded from statistical analysis if a report was not successfully generated by the AI software. Radiographs participating in this study were not distinguished from the software's regular workload.

All images assessed by AI software were independently evaluated by one ACVR board-certified veterinary radiologist (A.F.) with 15 years of post-graduate experience. The radiologist was blinded to the patient's history, signalment, and the results of the AI-generated report. All results were separated into two possible categories: cardiogenic pulmonary edema positive (CPE+) or cardiogenic pulmonary edema negative (CPE−).

2.3 | Data analysis

Descriptive analyses were performed by a data scientist (P.S.) with 2 years of formal training in statistics using commercial statistical software (R version 4.0.3, Vienna, Austria; Microsoft Excel 2021, Redmond, WA, USA). Estimated prevalence of CPE in the study population was calculated using a retrospective review of the hospital's electronic medical records database for all dogs who obtained thoracic radiographs and furosemide for treatment of presumptive left-sided CHF. In 2019, 1693 canine patients were evaluated using thoracic radiographs through the after-hours Emergency Department between the hours of 18:00 and 8:00 and 183 of these patients were treated with furosemide during the same visit. Based on these data, the prevalence of CPE was calculated to be 10.8%. Sample size calculations were then made to target high sensitivity, given the importance of minimizing the risk of false negative in this disease population. Utilizing Buderer's formula we calculated an estimated sample size of 375 thoracic radiographic studies.[17] Standard calculations of accuracy, sensitivity, and specificity were performed, using the radiologist's assessment as the reference standard. For cases diagnosed by a radiologist as CPE+, positive AI diagnoses were classified as True Positives (TP) while negative AI diagnoses were classified as False Negatives (FN). For cases diagnosed by a veterinary radiologist as CPE−, negative AI diagnoses were classified as True Negatives (TN), while positive AI diagnoses were classified as False Positives (FP) (Table 1). The sensitivity was calculated as TP/(TP+FN). The specificity was calculated as TN/(TN+FP). Overall accuracy was calculated as (TP+TN)/(TP+TN+FP+FN). Confidence interval bounds were calculated using the Wilson score interval. The positive predictive value (PPV) and negative predictive value (NPV) were calculated using the following equations respectively: PPV = TP/(TP+FP) and NPV = TN/(TN+FN) (Table 1). Lastly, the Youden's index was calculated as sensitivity + specificity − 1.

TABLE 1 Basis for sensitivity and specificity calculations displaying which studies were considered true positives, true negatives, false positives, and false negatives

	Radiologist CPE+	Radiologist CPE−
AI CPE+	True Positive (TP)	False Positive (FP)
AI CPE-	False Negative (FN)	True Negative (TN)

Abbreviations: AI, artificial intelligence; CPE− cardiogenic pulmonary edema negative.; CPE+, cardiogenic pulmonary edema positive.

Regression analyses were performed by an American College of Veterinary Emergency Critical Care-certified veterinary specialist (J.W.) with a doctoral degree in biosciences using commercial statistical software (Stata SE v15.1, College Station, TX, USA). A stepwise logistic regression analysis was applied to assess for independent association between patient characteristics and disagreement of AI and radiologist assessment of the images. The model and goodness of fit were evaluated by likelihood ratio and comparison to Chi square assessment and all values evaluated as statistically significant demonstrated appropriate models. Bonferroni correction was applied for multiple comparisons. Statistical significance was set at a P value of <0.05.

3 | RESULTS

Thoracic radiographs had been acquired using one of two available digital radiographic systems (Quantum Medical Imaging, Quantum HF Radiographic Imaging System, Ronkonkoma, NY). Both units had the same flat panel digital radiographic image detector (Canon U.S.A., Inc., Canon CXDI-50G, Melville, NY) with the same postprocessing algorithms optimized for thoracic studies. The technique varied depending on the thickness of the animal with the kVp ranging from 85–100 and the mAs from 4–5. All 500/500 cases received a CPE+ or CPE- diagnosis from the veterinary radiologist. The AI software produced a CPE diagnosis for 481/500 cases, with an analyzability rate of 96.2%. The 19/500 cases that did not receive a generated report from the AI software were excluded from further analyses. In 12 of the 19 non-analyzable cases, the AI software did not make a fully automated CPE diagnosis due to a specific failure to automatically crop images. In these 12 cases, a human technician manually cropped images, after which AI software produced and delivered a CPE diagnosis. The remaining 7/19 cases did not have an AI generated report due to internal server error. Of the remaining 481 image sets, patient age ranged from 1 month to 18 years (median, 9.2 years). Seventy one different breeds were represented with 233 females (40 intact, 193 spayed), and 248 males (57 intact, 191 neutered). Body weight ranged from 1 kg to 82 kg (median, 9.3 kg). The number of images per study ranged from 1 to 6 with a median, mean, and mode of 3.0. A total of 1441 radiographic images of the 481 image sets were included in this study.

TABLE 2 Comparison of CPE diagnosis of the ACVR-certified radiologist and AI software (n = 500) where radiologist provided a diagnosis for all 500 cases and the AI software provided a diagnosis for 481 cases

	Radiologist	AI
CPE+	49	75
CPE–	451	406
Total	500	481

Abbreviations: AI, artificial intelligence; CPE–, cardiogenic pulmonary edema negative.; CPE+, cardiogenic pulmonary edema positive.

TABLE 3 Comparison of the CPE diagnosis of the ACVR-certified radiologist and AI software in the analyzable cases (n = 481) where the radiologist and AI software agreed on CPE+ for 42 cases and CPE– for 402 cases

	Radiologist CPE+	Radiologist CPE–
AI CPE+	42	33
AI CPE–	4	402
Total	46	435

Abbreviations: AI, artificial intelligence; CPE–, cardiogenic pulmonary edema negative; CPE+, cardiogenic pulmonary edema positive.

Radiographic evaluation by the veterinary radiologist reported 49/500 CPE+ and 451/500 CPE- cases (Table 2). Of these, 3 CPE+ and 16 CPE- image sets were excluded; therefore, the radiologist reported 46/481 CPE+ and 435/481 CPE-. The AI software reported 75/481 CPE+ and 406/481 CPE- (Table 2). Of the 46 CPE+ diagnoses by the radiologist, the AI software agreed with the diagnosis on 42 (92.3%) of the cases (Table 3). Comparing diagnosis by the radiologist with that from AI, 42/481 were diagnosed CPE+ and 4/481 were diagnosed CPE- by AI software (91.3% sensitivity, 95% confidence interval, 79.7% to 96.6%). Of the 435 cases diagnosed by the radiologist as CPE-, AI diagnosed 402 as CPE- and 33 as CPE+ (reflecting 92.4% specificity, 95% confidence interval, 89.5% to 94.5%). The positive predictive value was 56% and negative predictive value was 99% based upon AI diagnosis of cardiogenic pulmonary edema from canine thoracic radiographs. The Youden's index was 0.84.

The AI software reported true diagnosis in all 444/481 cases, false diagnosis in 37/481 (accuracy, 92.3%, 95% confidence interval, 89.6% to 94.4%). Of the 37 cases with a false diagnosis, 33 were FP and 4 were FN. The FP cases included radiographic findings suggesting a normal thorax (n = 17), pleural effusion (n = 4), pneumonia (n = 4), cardiomegaly (n = 3), pulmonary neoplasia/metastasis (n = 2), chronic lower airway disease (2), and non-cardiogenic pulmonary edema (n = 4); a few of the studies displayed more than one of the listed abnormalities.

In order to assess for patient characteristics influencing the likelihood of disagreement between AI and radiologist assessment of images, a regression model was built to include disagreement (yes vs. no) as a dependent variable. Independent variables included age and patient body weight. A significant model emerged demonstrating that

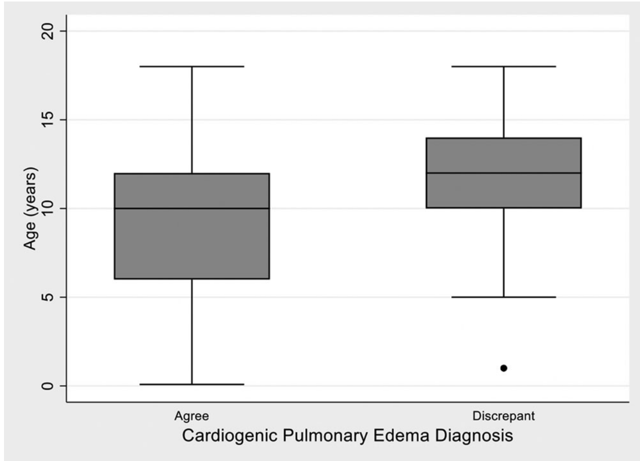

FIGURE 1 Box plot displaying the significantly higher median patient age in the cases with discrepant CPE diagnosis compared to the between the board-certified radiologist and AI software compared to in cases where they agreed. The central line is median, upper and lower limits of the box represent 25% and 75% quantiles, the whiskers represent the range minus the outliers, and the outliers are designated by dot

patient age influences the likelihood of disagreement (agreed: median 10 y, range 0.08-18 vs disagreed: median 12 y, range 1–18; OR: 1.13; CI: 1.03-1.25; p = 0.005). This indicated that for each increase in age by 1 year, the chances of disagreement increased by about 13%. (Figure 1) No difference was found between agreement groups in regard to patient body weight (agreed: 8.5 kg, range 0.8-82 vs disagreed: 9.3 kg, range 1.6-51).

The AI report generation time (RGT) was available for 460/481 cases. If images were manually cropped and resubmitted by human technicians, the AI time stamp would reset so the RGT was not recorded. For the 96% of cases for which timing data existed, the average RGT was 2.45 minutes with a maximum of 8.3 minutes and a minimum of 0.65 minutes.

4 | DISCUSSION

The present study demonstrated that AI software correctly identified CPE from canine thoracic radiographs in approximately every 9 out of 10 cases diagnosed with CPE by a board-certified veterinary radiologist. We identified a high NPV reflecting that the CPE- cases identified by AI had a high probability of being negative. This indicated that under present study parameters, an AI report indicating CPE- represented a very high likelihood (99%) to be in agreement with the radiologist's interpretation. On the other hand, the low PPV for CPE+ indicated that compared to a radiologist's diagnosis, the AI assessment of true positive CPE+ was realized 56% of the time. Thus, if an AI-generated CPE– report is received by a clinician, left-sided congestive heart failure is unlikely, and consideration should be given for further clinical investigation or diagnostic testing. In general, our data from 481 radiographic studies suggests that in our study population, the AI software could be considered a good screening tool in the low CHF+ prevalence canine population. These findings also suggest that the acuity of the present AI software versions may be enhanced when used in conjunction with a clinician or radiologist's assessment.

Older patient age was linked to a higher likelihood of discrepancy between the board-certified radiologist and AI in CPE diagnosis. Radiographic changes seen in the canine lungs of aging dogs without clinical evidence of disease have long been reported and include pleural thickening, increased nonvascular linear markings, and nodular lesions.[18] A more recent study has also confirmed the increased likelihood of the presence of osseous metaplasia and lung collapse in the aging canine population using CT.[19] It is possible that radiographic changes that are considered to be normal age-related changes may have led to misinterpretation by the AI software. Future, prospective evaluations designed to specifically investigate age are indicated based on these findings. As there was no difference found between agreement groups in relation to the patient body weight, breed differences were not investigated as they were unlikely to yield statistically significant data.

Veterinary literature reporting a clinical application of AI-based software is limited. In one retrospective study, the reported accuracy, sensitivity, and specificity were similar between the ACVR-certified veterinary radiologist and AI when investigating left atrial enlargement from 81 canine thoracic radiographs.[13] A recent prospective study investigated the use of AI software in detecting thoracic lesions from 120 canine and feline thoracic radiographs. Investigators demonstrated a significantly lower error rate in AI image interpretation as compared to that of veterinarians with varying levels of experience and veterinarians aided by the AI software that were all held against a reference standard of an ECVDI or ACVR board-certified radiologist's interpretation.[14] Similar to the present study, these studies support that AI can provide adjunctive data that may be useful in guiding a clinical diagnosis.

Because the thoracic radiographs in the present study were submitted manually and consecutively for AI analysis during morning hours, this may have created a bottleneck effect for AI response time. Accordingly, it may be useful to consider the software's capability for automatic downloading of thoracic radiographs to the AI software at the completion of each study in order to try to reduce software wait time. The authors are not aware of published data regarding the length of time for board-certified veterinary radiologists to interpret a set of thoracic radiographs and to generate a report. One study in human medicine tested the capability of radiology residents to read thoracic radiographs and estimated that the study could be appropriately evaluated in 1.5 min.[20] This is below the average of 2.45 min that it took the AI software to perform in our study. However, this study only measured the time required for the radiology resident to identify the lesions and not to generate a report. It would be straightforward for the AI software to produce a CPE determination directly from CNN outputs, but, as currently designed, only the plain language report is provided as output. Direct reporting of the binary classification for CPE instead of producing a plain language report could be considered in an effort to shorten RGT achieved by AI software. Further analysis of RGT in AI software as compared to radiologist are warranted.

Numerous studies investigating the application of AI in humans have reported a range of utility and outcomes when specific AI algorithms are evaluated to detect a specific disease or condition. As of October 2020, 64 AI/ML-based algorithms have been approved by the Food and Drug Administration (FDA).[21] Application of AI systems has been suggested for optimizing image acquisition and processing, enhancing patient care, and optimizing workflow.[22] Nevertheless, the search for a role of AI in human medicine remains a work in progress,[22–24] and a seamless integration and clear-cut role of AI in human and veterinary diagnostic imaging has yet to be defined.[25,26]

The present study has limitations that warrant discussion. Some cases had to be excluded from analyses due to failure of the AI software to generate a fully automated diagnosis. Possible causes could have included animal rotation or malpositioning, poor image quality, or failure to automatically crop for the specific area being evaluated. While human intervention helped to resolve this problem in our sample, these cases were excluded from analyses to minimize potential bias. We chose to have a single radiologist assess for CPE, therefore generalizability for our findings remains unknown. This decision was made to minimize possible confounding factors between investigators that could have introduced type 2 errors. In future studies, it may be beneficial to include multiple ACVR- or ECVDI-certified veterinary radiologists. Another limitation was that the study was designed to compare the ability of the AI to a veterinary radiologist who was similarly blinded to all patient information, clinical findings, and history. Final diagnostic imaging reports can list multiple differentials for lung pathology. To avoid ambiguity, the board-certified radiologist was given a binary option of CPE+ or CPE−. For these reasons, the AI software and board-certified radiologist were not compared to the final diagnostic imaging report. Lastly, this study was not designed to assess the impact of AI software on clinical outcomes in dogs with CPE. Future investigations of the effect of AI on clinical decision-making and patient outcomes may help to better understand the integration of these systems in the management of dogs with CPE.

In conclusion, findings from the present study supported the use of AI software for an initial screening assessment of canine thoracic radiographs for CPE when a veterinary radiologist is not available. The AI software's interpretations had a high NPV in this study, however, the comparatively low PPV suggested that the AI software had limitations and that confirmation of AI assessment by a veterinary radiologist remains important. Future studies are needed to assess the generalizability of these study findings for varying hospitals and veterinary radiologist observers.

LIST OF AUTHOR CONTRIBUTIONS

Category 1

(a) Conception and Design: Kim, Fischetti, Weltman
(b) Acquisition of Data: Kim, Fischetti
(c) Analysis and Interpretation of Data: Sreetharan, Kim, Fischetti, Fox

Category 2

(a) Drafting the Article: Kim, Fischetti, Sreetharan, Fox, Weltman
(b) Revising the Article for Intellectual Content: Kim, Fischetti, Sreetharan, Fox, Weltman

Category 3

(a) Final Approval of the Completed Article: Kim, Fischetti, Sreetharan, Fox, Weltman

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: Kim, Fischetti, Sreetharan, Fox, Weltman

CLAIM DISCLOSURE

Along with the advice of the ACVR and ECVDI's AI Education and Development Committee, the following was used for guidance in preparation of this manuscript: Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell*. 2020 Mar 25;2(2):e200029.

CONFLICT OF INTEREST

Pratheev Sreetharan is a member of the development team of the artificial intelligence software used in this study.

ORCID

Eunbee Kim https://orcid.org/0000-0002-6756-4324

REFERENCES

1. Keene BW, Atkins CE, Bonagura JD, et al. ACVIM consensus guidelines for the diagnosis and treatment of myxomatous mitral valve disease in dogs. *J Vet Intern Med*. 2019; 33:1127-1140.
2. Buchanan JW. Prevalence of cardiovascular disorders. In: Sisson D, Sydney Moïse N, Fox PR, eds. *Textbook of Canine and Feline Cardiology: Principles and Clinical Practice*. WB Saunders. 1999; 457-470.
3. Thrall DE, Widmer WR. *Textbook of Veterinary Diagnostic Radiology*. Elsevier; 2018;684-709.
4. Schober KEE, Hart TM, Stern JA, et al. Detection of congestive heart failure in dogs by Doppler echocardiography. *J Vet Intern Med*. 2010; 24:1358-1368.
5. Ward JL, Lisciandro GR, Keene BW, Tou SP, DeFrancesco TC. Accuracy of point-of-care lung ultrasonography for the diagnosis of cardiogenic pulmonary edema in dogs and cats with acute dyspnea. *J Am Vet Med Assoc*. 2017; 250:666-675.
6. Chartrand G, Cheng PM, Voronstov E, et al. Deep learning: a primer for radiologists. *Radiographics*. 2017; 37:2113-2131.
7. Choy G, Khalizadeh O, Michalski M, et al. Current applications and future impact of machine learning in radiology. *Radiology*. 2018; 288:318-328.
8. Willemink MJ, Koszek WA, Hardell C, et al. Preparing medical imaging data for machine learning. *Radiology*. 2020; 295:4-15.
9. Chassagnon G, Vakalopoulou M, Paragios N, Revel MP. Artificial intelligence applications for thoracic imaging. *Eur J Radiol*. 2020; 123:1-6.
10. Cicero M, Bilbily A, Colak E, et al. Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. *Invest Radiol*. 2017; 52:281-287.
11. 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; 54:55-59.
12. Yoon Y, Hwang T, Lee T. Prediction of radiographic abnormalities by the use of bag-of-features and convolutional neural networks. *Vet J*. 2018; 237:43-48.
13. 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; 61:1-8.
14. Boissady E, de La Comble A, Zhu X, Hespel A. Artificial intelligence evaluating primary thoracic lesions has an overall lower error rate compared to veterinarians in conjunction with the artificial intelligence. *Vet Radiol Ultrasound*. 2020; 61:1-9.
15. Artificial intelligence. American College of Veterinary Radiology website. May 7, 2021. Accessed May 12, 2021. https://acvr.org/artificial-intelligence-in-veterinary-diagnostic-imaging-and-radiation-oncology/
16. Simonyan K, Zisserman A, Very deep convolutional networks for large-scale image recognition. Paper presented at: International Conference on Learning Representations 2015; May 7–9, 2015; San Diego, USA.
17. Haijan-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform*. 2014; 48:193-204.
18. Reif JS, Rhodes WH. The lungs of aged dogs. A radiographic-morphologic correlation. *J Am Vet Radiol Soc*. 1966; 7:5-11.
19. Hornby NL, Lamb CR. Does the computed tomographic appearance of the lung differ between young and old dogs?. *Vet Radiol Ultrasound*. 2017; 1-6.
20. Fabre C, Proisy M, Chapuis C, et al. Radiology residents' skill level in chest x-ray reading. *Diagn Interv Imaging*. 2018; 99:361-370.
21. Benjamens S, Dhunnoo P, Mesko B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med*. 2020; 118:1-8.
22. Hardy M, Harvey H. Artificial intelligence in diagnostic imaging: impact on the radiography profession. *Br J Radiol*. 2020; 93:1108.
23. Oren O, Gersh BJF, Bhatt DL. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. *Lancet Digit Health*. 2020; 2:486-488.
24. Gampala S, Vankeshwaram V, Gadula SSP. Is artificial intelligence the new friend for radiologists? A review article. *Cureus*. 2020; 12:1-7.
25. O'Neill TJ, Xi Y, Browning T, et al. Active reprioritization of the reading worklist using artificial intelligence has a beneficial effect on the turnaround time for interpretation of head CT with intracranial hemorrhage. *Radiology*. 2020; 3:2.
26. Retson TA, Masutani EM, Golden D. Clinical performance and role of expert supervision of deep learning for cardiac ventricular volumetry: a validation study. *Radiology*. 2020; 2:4.

How to cite this article: Kim E, Fischetti AJ, Sreetharan P, Weltman JG. Fox PR Comparison of artificial intelligence to the veterinary radiologist's diagnosis of canine cardiogenic pulmonary edema. *Vet Radiol Ultrasound*. 2022;1-7. https://doi.org/10.1111/vru.13062