The Veterinary Journal 262 (2020) 105505

Contents lists available at ScienceDirect

The Veterinary Journal

journal homepage: www.elsevier.com/locate/tvjl

# Use of deep learning to detect cardiomegaly on thoracic radiographs in dogs

S. Burti, V. Longhin Osti, A. Zotti, T. Banzato*

*Department of Animal Medicine, Productions and Health, University of Padua, Viale Dell'Università 16, 35020 Legnaro, Padua, Italy*

ARTICLE INFO

*Article history:*
Accepted 6 July 2020

*Keywords:*
Computer aided diagnosis
Convolutional neural network
Vertebral heart scale

ABSTRACT

The purpose of this study was to develop a computer-aided detection (CAD) device based on convolutional neural networks (CNNs) to detect cardiomegaly from plain radiographs in dogs. Right lateral chest radiographs ($n$ = 1465) were retrospectively selected from archives. The radiographs were classified as having a normal cardiac silhouette (No-vertebral heart scale [VHS]-Cardiomegaly) or an enlarged cardiac silhouette (VHS-Cardiomegaly) based on the breed-specific VHS. The database was divided into a training set (1153 images) and a test set (315 images). The diagnostic accuracy of four different CNN models in the detection of cardiomegaly was calculated using the test set.

All tested models had an area under the curve >0.9, demonstrating high diagnostic accuracy. There was a statistically significant difference between Model C and the remainder models (Model A vs. Model C, $P$ = 0.0298; Model B vs. Model C, $P$ = 0.003; Model C vs. Model D, $P$ = 0.0018), but there were no significant differences between other combinations of models (Model A vs. Model B, $P$ = 0.395; Model A vs. Model D, $P$ = 0.128; Model B vs. Model D, $P$ = 0.373). Convolutional neural networks could therefore assist veterinarians in detecting cardiomegaly in dogs from plain radiographs.



## Introduction

Error in the interpretation of plain radiographs is common both in human (Kelly et al., 2016; Waite et al., 2016) and veterinary medicine (Alexander, 2010). The average error rate reported in human medicine is about 26% (Herman and Hessel, 1975) and, despite increasingly higher levels of training radiologists, has remained virtually unchanged during the last five decades (Berlin, 2007). The main causes of interpretation errors are perceptual errors, omission errors, cognitive biases, fatigue, and distraction (Waite et al., 2016). Level of expertise plays a major role in the error rate (Gatt et al., 2003); nonetheless, expertise alone is not sufficient to significantly reduce the error rate in a clinical setting (Bruno et al., 2015). Similar causes of interpretation error are recognized in veterinary medicine (Alexander, 2010); however, studies on the rate and impact of radiographic interpretation error are, to the best of the authors' knowledge, not currently available in veterinary literature.

Strategies for error reduction in radiology include: structured reports and checklists, standardization of the interpretation process, strategies for cognitive debiasing, and metacognitive interventions (Croskerry, 2009). In recent decades, research into this topic has been largely focused on the development of computer-aided detection (CAD) methods, especially those based on artificial intelligence (Lin et al., 1993; Monnier-Cholley et al., 1998; Kakeda et al., 2004; Suzuki et al., 2005; Cicero et al., 2017; Wang et al., 2017; Dunnmon et al., 2019; Seah et al., 2019). The increasing availability of large-scale databases and servers with high computing power constitute the game-changing approach for the future development of these technologies (Yasaka and Abe, 2018).

In veterinary medicine, the availability of highly-trained radiologists is limited and most clinical practises do not have such professionals amongst their staff. Further, interpretation of radiographs in the emergency setting is often performed by non-specialist veterinarians who may be having to care for multiple patients concurrently (i.e. are distracted) and under time constraints. In such a scenario, the availability of CAD methods to assist the clinician in interpreting of basic radiographic changes could be beneficial in reducing interpretation error. In veterinary medicine, the use of CAD methods has been explored in ultrasonography (Drost et al., 2000; Ivancic and Mai, 2008; Banzato et al., 2015; Zotti et al., 2015; Banzato et al., 2017a,c; Banzato et al., 2018a; Da Silva et al., 2018; Rodrigues Simões et al., 2018; Mattei et al., 2019), radiology (Vaccaro et al., 2012), magnetic resonance imaging (Banzato et al., 2017b, 2018b,c) and computed tomography (Marschner et al., 2017; Longo et al.,

* Corresponding author.
*E-mail address:* tommaso.banzato@unipd.it (T. Banzato).

http://dx.doi.org/10.1016/j.tvjl.2020.105505

2018). Nevertheless, technical limitations of the CAD methods described in these studies have hindered the routine use of CAD in clinical practice.

Cardiac disease is a leading cause of death in pet dogs (O'Neill et al., 2013; Inoue et al., 2015; Boland et al., 2018). Cardiomegaly is a frequent radiographic finding associated with cardiac disease (Lamb et al., 2001). A method to index heart size to body size, and more accurately determine cardiomegaly, is known as the vertebral heart scale (VHS), where dimensions of the cardiac silhouette are measured in terms of thoracic vertebral body lengths on plain radiographs (Buchanan and Bücheler, 1995). Although such a method is available, VHS may not be routinely used to quantitatively evaluate heart size, and radiographic diagnoses of cardiomegaly (especially if mild to moderate) might be missed by veterinarians with limited training in veterinary radiology or cardiology. Thus, the aim of this study was: (1) to develop a CAD method to detect cardiomegaly from plain radiographs in dogs; and (2) to test the accuracy of the developed CAD method.

## Materials and methods

### *Database construction*

All the methods were carried out in accordance with the relevant guidelines and regulations. This study was conducted respecting Italian Legislative Decree 26/2014 (transposing the EU directive 2010/63/EU regarding animal welfare and protection). As the data used in this study were part of routine clinical activity, no ethical committee approval was needed. Informed consent regarding personal data processing was obtained from the owners.

All thoracic radiographs of dogs referred for specialist examination at the Veterinary Teaching Hospital of the University of Padua between January 2014 and October 2019 were reviewed. Signalment was obtained from digital imaging and communications in medicine (DICOM) files and recorded for each patient. The radiography system was replaced during the observation period: with a computed radiography system (Kodak Point of Care CR-360 System, Carestream Health Inc.) used between January 2014 and June 2018; and a direct radiography system (FDR D-EVO 1200 G43, Fujifilm Corporation) used between July 2018 and October 2019. The radiographs from the computed radiography system were used to train the network (training set) and the radiographs from the direct radiography system were used to test the accuracy of the CAD method (test set). Initially, radiographic quality was checked, and poorly positioned or poorly exposed radiographs were excluded from further analysis. Cases in which the cardiac silhouette was not clearly visible (e.g. when pleural effusions, overlying masses or focal/disseminated alveolar pattern were superimposed over the heart) or was displaced (e.g. when pneumothorax was present) were excluded from further analysis. In order to reduce the intrinsic variability of the data (Greco et al., 2008), only right lateral projections were included in the database, with all other views excluded from further analysis. No patients had manual inflation of their lungs.

The VHS was calculated by two experienced operators (TB and AZ; 10 years' and 20 years' experience in veterinary diagnostic imaging, respectively) jointly evaluating each radiograph. The images were classified as having a normal cardiac silhouette (No-VHS-Cardiomegaly) or having an enlarged cardiac silhouette (VHS-Cardiomegaly), based on the calculated VHS alone, regardless the presence of other pathologies (e.g. masses, alveolar patterns, etc.). Breed-specific VHS reference intervals, as reported in the available literature, were used for the classification of pedigree dogs (Lamb et al., 2001), whereas a VHS reference interval of 9.2–10.2 was used for the classification of mixed-breed dogs or pedigree dogs where no breed-specific reference interval were available (Buchanan and Bücheler, 1995). The images were stored in separate folders according to their VHS classification (No-VHS-Cardiomegaly; VHS-Cardiomegaly), and whether they were to be used for training or for testing.

### *Image processing*

All the radiographic images were exported in an 8-bit Joint Photographic Experts Group (JPEG) format using a freely available image visualization and analysis software (Horos, Nimble). Images were then manually cropped to a square format in order to avoid geometric distortions. Lastly, the images were resized using commercially available software (MATLAB and Statistics Toolbox, release 2019b, The MathWorks) to meet the specific requirements of the different convolutional neural network (CNN) architectures tested.

### *Deep learning*

The deep-learning analysis was performed on a dedicated workstation (Linux operating system; Ubuntu 18.04, Canonical) equipped with four graphic processing units (Tesla V100; NVIDIA), a 2.2 GHz processor (Intel Xeon E5-2698 v4; Intel) and 256 GB random access memory. A deep-learning toolbox (MATLAB and Statistics Toolbox, release 2019b) was used for image analysis. The accuracy of four different models, based on different CNN architectures was tested. Specifically, Model A was based on Inception V3 (Szegedy et al., 2015; Szegedy et al., 2017)., Model B was based on Inception-ResNet V2 (Szegedy et al., 2015), Model C was based on VGG-19 (Simonyan and Zisserman, 2015), and Model D was based on ResNet-101 (He et al., 2016). All the CNN architectures were pre-trained on a large-scale database[1] of everyday colour images were used for image classification. The learn rate was automatically adjusted using an adaptive learn rate function. Image normalization was performed prior to the analysis. During training, the images in the training set were randomly split into a training set comprising 80% of the images and a validation set comprising 20% of the images. Training was performed for 10 epochs, where an epoch is a complete iteration of the CNN on the images. The training and validation sets were randomly shuffled after every epoch. The following network hyperparameters were always used: optimizer stochastic gradient descent; batch size of 32; and initial learn rate of 3e-4. An adaptive learn rate option was selected. No data augmentation technique was used. The predicted class (i.e. No-VHS-Cardiomegaly vs. VHS-Cardiomegaly) of each case in the test set was recorded.

### *Statistics and data analysis*

Normal distribution of the descriptive data was tested with the Shapiro–Wilk test. Normally distributed data were reported as mean and standard deviation, whereas non-normally distributed data were reported as median with the limits of the overall range. A contingency table with the results of the test set was created and the area under the curve (AUC), sensitivity, specificity, accuracy, positive and negative likelihood ratios (PLR and NLR, respectively) and positive and negative predictive values (PPV and NPV, respectively) were calculated using commercially available software (MedCalc v. 15.05, MedCalc Software; and MATLAB and Statistics Toolbox, release 2018a). Confidence intervals (95% CI) were calculated as exact Clopper-Pearson confidence intervals. The AUC value as a criterion of discrimination accuracy was classified as low (0.5–0.7), moderate (0.7–0.9) or high (>0.9; McCrum-Gardner, 2008). The AUC of the individual architectures was compared with the DeLong method (DeLong et al., 1988).

## Results

### *Database*

A diagram of the workflow used in constructing the database is provided in Fig. 1. Of a total of 6282 canine thoracic radiographs were retrieved, 1153 images were included in the final training set and 315 images were included in the final test set. The VHS of both final training and test sets were normally distributed. Of the included images, 281 of the final training set (24.4%) were classified as VHS-Cardiomegaly (mean VHS 12.2, ±0.9) and 872 as No-VHS-Cardiomegaly (mean VHS 9.6, ±0.4); whilst 70 of the final test set (22.2%) were classified as VHS-Cardiomegaly (mean VHS 11.7, ±1.0) and 245 as No-VHS-Cardiomegaly (mean VHS 10.1, ±0.4). A bar-plot showing the breed distribution of dogs classified as No-VHS-Cardiomegaly or VHS-Cardiomegaly in both the final training and test sets is reported in Supplementary document 1. The distribution of the different radiographic findings in both the final training and test sets are graphically depicted in Supplementary document 2.

### *Classification accuracy*

Model A showed an AUC of 0.965 (95% CI: 0.943–0.988), Model B showed an AUC of 0.953 (95% CI: 0.923–0.983), Model C showed an AUC of 0.904 (95% CI: 0.858–0.950), and Model D showed an AUC of 0.973 (95% CI: 0.958–0.988). There were significant differences in the performances between Model C and the other models (Model A vs. Model C, $P = 0.0298$; Model B vs. Model C, $P = 0.003$; Model C vs. Model D, $P = 0.0018$); whereas, no significant differences were evident for the other comparisons (Model A vs. Model B, $P = 0.395$; Model A vs. Model D, $P = 0.128$; and Model B vs. Model D, $P = 0.373$). The complete results for the classification

[1] See: ImageNet. www.image-net.org. (Accessed 3 July 2020).

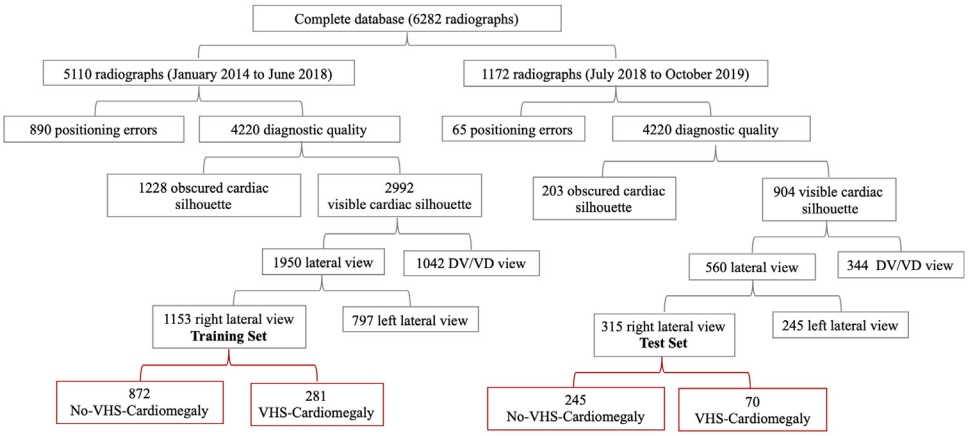


**Fig. 1.** Diagram of the workflow used in building the database to train the convolutional neural network. L, lateral view; RL, right lateral view; LL, left lateral view; DV, dorsoventral view; VD, ventrodorsal view; VHS, vertebral heart scale.

accuracy of the architectures tested in the test set are reported as Supplementary material. The comparison of the ROC curves is reported in Fig. 2.

Examples of radiographs correctly and incorrectly classified by the Model A model are reported in Figs. 3 and 4, respectively. The Model A misclassified 31 radiographs in total (three VHS-Cardiomegaly and 28 No-VHS-Cardiomegaly). The VHS was 11.2 or higher in all three false negative cases, furthermore, in two there were concomitant radiographic findings consistent with cardiogenic pulmonary oedema. In the 28 misclassified No-VHS-Cardiomegaly cases no common features or concomitant radiographic finding were identified. Model B misclassified 27 cases (seven VHS-Cardiomegaly and 20 No-VHS-Cardiomegaly). In three cases the VHS was 11.1 or higher whereas in four cases the VHS was between 10.5 and 11. Also for Model B there were no common features or radiographic findings in the false positive cases. Model C showed the overall worse performance with 39 misclassified cases (10 VHS-Cardiomegaly and 29 No-VHS-Cardiomegaly) with no clear pattern in both false positive and false negative. Finally, Model D misclassified 26 cases (four VHS-Cardiomegaly and 22 No-VHS-Cardiomegaly). The VHS of the false negative cases was between 10.6 and 11.1. As for the other models, there was no clear pattern in the false positive cases.

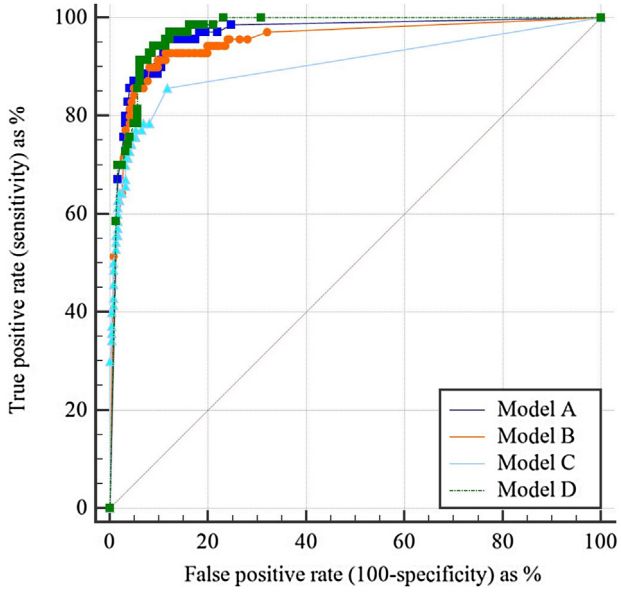


**Fig. 2.** Comparison of the receiver operating characteristic curves of Models A, B, C, and D.

## Discussion

The results of this study suggest that the application of CNN-based CAD methods could, prospectively, assist veterinarians in the interpretation of thoracic radiographs. In this paper, due to the limited size of the available database, only a single radiographic finding (cardiomegaly) was included. Interestingly, a similar classification accuracy was obtained by Lakhani and Sundaram (2017) in the detection of tuberculosis on thoracic radiographs from a small-sized (1007 images) database. Other authors (Cicero et al., 2017; Rajpurkar et al., 2017) have developed and tested CNNs capable of identifying several radiographic abnormalities. In the paper by Cicero et al. (2017), six categories were included (Normal, Cardiomegaly, Effusion, Consolidation, Oedema, and Pneumothorax), whilst in the paper by Rajpurkar et al. (2017) 14 categories of pathology were included (Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural Thickening, and Hernia). In both studies massive databases were available (65,000 radiographs in the first paper and 112,120 images in the latter). Creating databases of such size is only feasible using machine-learning techniques capable of extracting relevant information from reports and automatically assigning tags to the associated images. Interestingly, the accuracy in detecting cardiomegaly as reported by Rajpurkar et al. (2017 was similar to the best classification results obtained in this study.

Another interesting aspect of this study was that the radiographs used to train the CNNs were classified either as No-VHS-Cardiomegaly or VHS-Cardiomegaly based on the available breed-specific VHS reference intervals as reported in the literature. Indeed, the relative dimensions of the normal cardiac silhouette, and therefore the VHS reference intervals, have a significant breed-related variability (Lamb et al., 2001). For example, boxers, Labrador retrievers, and cavalier King Charles spaniels have a relatively larger cardiac silhouette, as compared to other breeds (Lamb et al., 2001). The high classification accuracy obtained in the test set reveals that the CNN was able to account for the high variability in the morphology of the thorax and the cardiac silhouette, typical of the canine species, during classification of the images. These algorithms are often described as black boxes because, even if the CNN was capable to identify high-level features to correctly classify the radiographs regardless of the

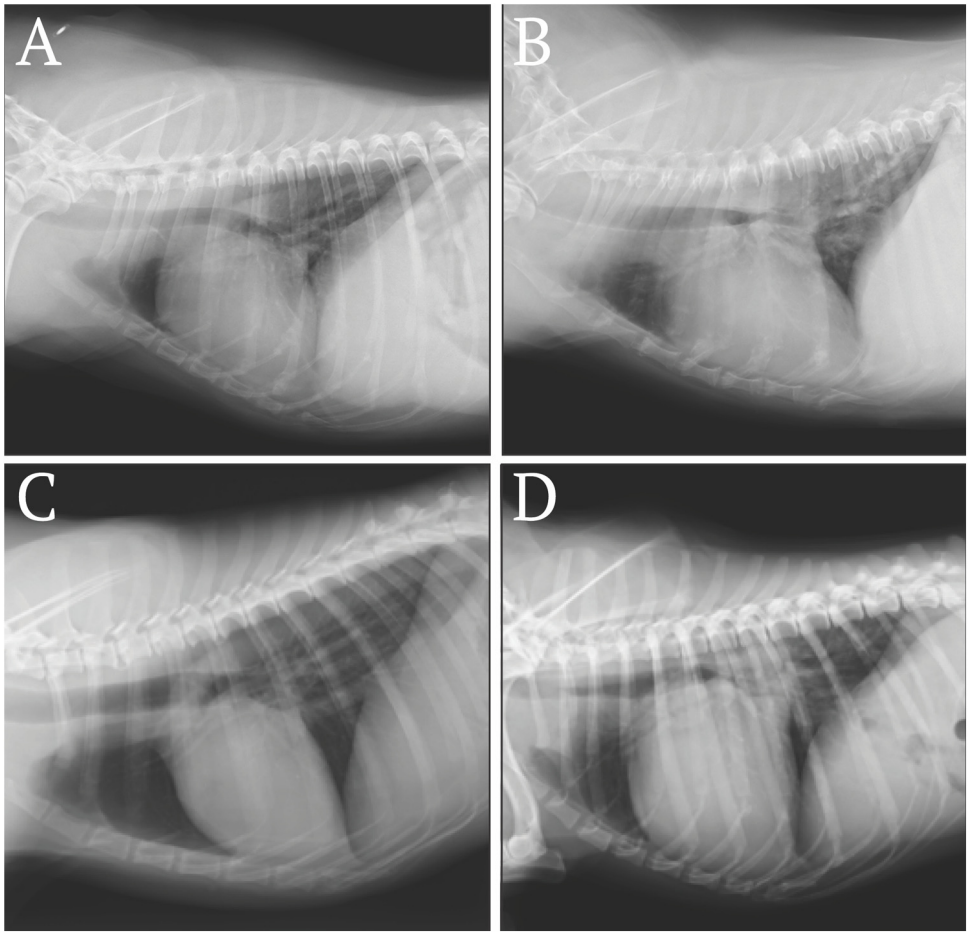


**Fig. 3.** Examples of radiographs accurately classified by Model A as having either a normal cardiac silhouette (A, D) or cardiomegaly (B, D), using measured vertebral heart scale (VHS) as the reference standard. A, lateral radiograph of an English bulldog with a normal cardiac silhouette (VHS 9.9; breed-specific reference interval 9.2–10.2). B, lateral radiograph of an English bulldog with an enlarged cardiac silhouette (VHS 11.3). C, lateral radiograph of a boxer dog with a normal cardiac silhouette (VHS 10.8; breed-specific reference interval 10.8–12.4). D, lateral radiograph of a boxer dog with an enlarged cardiac silhouette (VHS 12.8).

breed, the parameters used for classification are not directly accessible to the operator (Prieto et al., 2016).

The diagnostic accuracy of several deep-learning models was tested in this work and only Model C displayed a significantly lower accuracy compared to remainder tested models. The actual reasons underlying the performance differences of the deep-learning models on a specific database are difficult to fully assess and, therefore, we cannot offer an in-depth explanation for such differences. For instance, the performance of one CNN (not included in this study) is affected by noise, blur, missing pixels and brightness (Grm et al., 2018). By comparison, Model A has been reported to outperform other CNNs in the classification of diagnostic images in other studies (Banzato et al., 2018a; Noguchi et al., 2018; Banzato et al., 2019). Compared to the 'everyday life' photographs used to pre-train the CNNs, diagnostic images are usually acquired in a highly standardized fashion by trained personnel, and, therefore, differences in image quality are minimal among the images acquired with the same scanner.

The main strengths of this work are firstly, that the training set and test set were unrelated. In fact, these two sets of radiographs were acquired through different scanners with different resolutions and an overall different image quality. The high classification accuracy obtained on the test set suggests that the developed CAD methods could become a reliable, ready-to-use tool in clinical practice. However, as recently demonstrated by Zech et al. (2018), CNNs may have significantly varied performances on different sites, primarily related to differences in disease prevalence. Therefore, the on-site accuracy should be tested prior to the introduction of these CAD methods in clinical practice. Secondly, the CAD methods developed in this study could be directly applied to new images to detect cardiomegaly. Indeed, the limited pre-processing required and the use of CNNs enable straightforward use of this CAD method.

A limitation of this study is that the radiographs were classified based only on the VHS from a single radiographic view. It is possible that dogs with a normal VHS might have a significant cardiomyopathy and dogs with a VHS above the reference interval might have normal cardiac function (Lamb et al., 2000, 2001). Further, the VHS reference intervals used were based upon a group mean plus one standard deviation, which would identify 16% of normal dogs as having an increased VHS (Buchanan and Bücheler, 1995), whilst not all pedigree dogs have breed-specific reference intervals available, which might result in the over- or under-estimation of cardiomegaly in these dogs. The aim of the present study was not to overcome the well-known limitations of plain radiographs in the detection of cardiac disease, but simply to develop a CAD method to automatically classify canine thoracic radiographs as either No-VHS-Cardiomegaly or alternatively VHS-Cardiomegaly suggesting to the veterinarian that a more accurate examination of

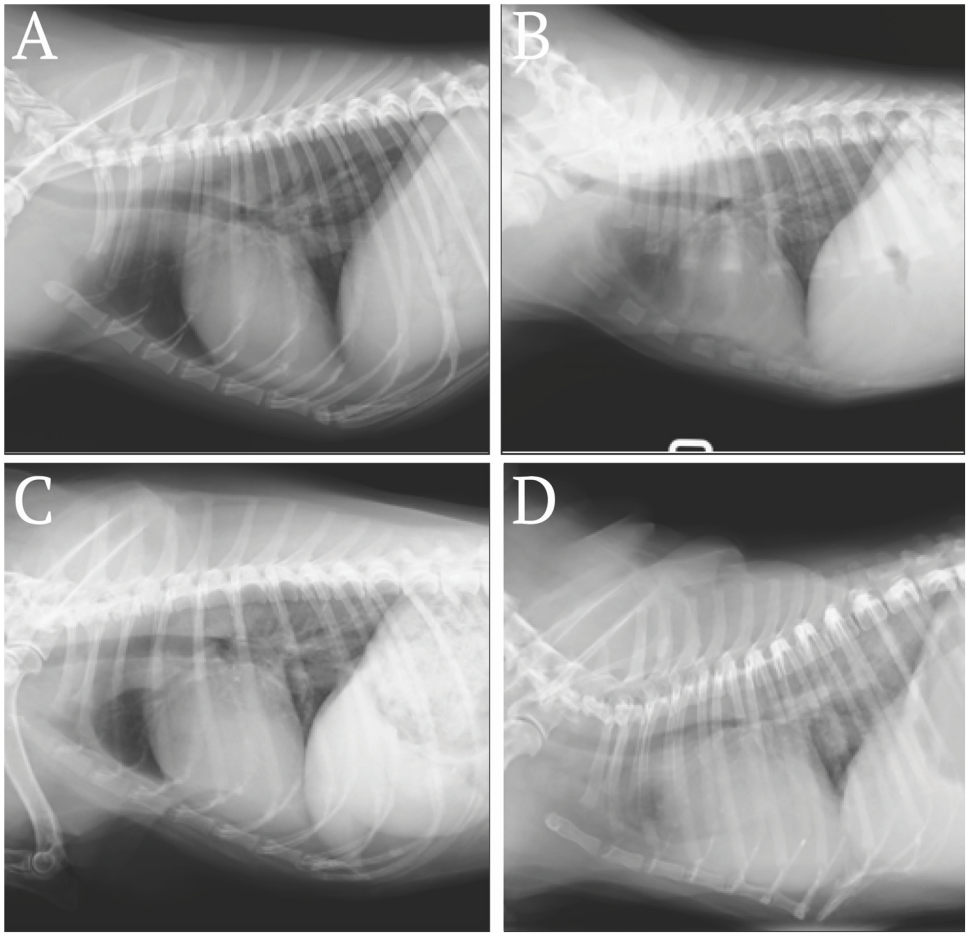


**Fig. 4.** Examples of radiographs incorrectly classified by Model A as having cardiomegaly when they had a normal cardiac silhouette (A, B), or as being normal when they had cardiomegaly (C, D), using measured vertebral heart scale (VHS) as the reference standard. A, lateral radiograph of a mixed breed dog with a normal cardiac silhouette (VHS 9.9; reference interval 9.2–10.2). B, lateral radiograph of a mixed breed dog with a normal cardiac silhouette (VHS 9.9). C, lateral radiograph of a mixed breed dog with an enlarged cardiac silhouette (VHS 10.7). D, lateral radiograph of a pug dog with an enlarged cardiac silhouette (VHS 11.7).

the cardiac function could be necessary. Further studies, classifying the thoracic radiographs based on the results of a full cardiological examination might help in the development of a CNN aimed to detect cardiac disease.

Another limitation of this study is that cases in which the cardiac silhouette was obscured were excluded from analysis. Further studies, possibly including a larger number of radiographs (ideally from several different veterinary clinics) would enable development of a CAD method capable of detecting several different radiographic abnormalities (e.g. nodules, effusions, pneumothorax, etc.). A major challenge in developing this kind of CAD method is that increasing the number of categories requires an increasingly higher number of images to accurately train the CNN; this problem is called 'the curse of dimensionality'. i.e. when a large number of features, which is the case of CNNs, are extracted from an insufficient number of images, the observations for each category become sparse, and the algorithm is unable to converge to an optimal solution, thus resulting in poor detection accuracy (Liang et al., 2015).

## Conclusions

The CNNs used, trained on a relatively small database of radiographic images, showed a high classification accuracy in the detection of cardiomegaly in dogs. These CAD methods developed stand as the first step in structuring a tool to assist the routine clinical work of veterinarians in detecting cardiomegaly in dogs.

## Conflict of interest

None of the authors has any financial or personal relationships that could inappropriately influence or bias the content of the paper.

## Acknowledgements

This paper is part of a project funded by two research grants from the Department of Animal Medicine, Production and Health — MAPS, University of Padua, Italy. Details of the grants are as follows: (1) SID-Zotti 2018, entitled: 'Application of deep-learning algorithms in pet animal diagnostic imaging'; and (2) SID-Banzato 2019, entitled 'Development of an algorithm for the automatic classification and identification of the lesions on the radiographs of the thorax in dogs.' The authors also hold one grant from the University of Padua 'Talents in Research@University of Padua', entitled 'Prediction of the histological grading of human meningiomas using MR images texture and deep learning: a translational application of a model developed on spontaneously occurring meningiomas in dogs'. In addition, the authors would like to thank the NVIDIA Corporation for donation of the GPU card used in this study.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.tvjl.2020.105505.

## References


Alexander, K., 2010. Reducing error in radiographic interpretation. Canadian Veterinary Journal 51, 533–536.

Banzato, T., Gelain, M.E., Aresu, L., Centelleghe, C., Benali, S.L., Zotti, A., 2015. Quantitative analysis of ultrasonographic images and cytology in relation to histopathology of canine and feline liver: an ex-vivo study. Research in Veterinary Science 103, 164–169.

Banzato, T., Zovi, G., Milani, C., 2017a. Estimation of fetal lung development using quantitative analysis of ultrasonographic images in normal canine pregnancy. Theriogenology 96, 158–163.

Banzato, T., Bernardini, M., Cherubini, G.B., Zotti, A., 2017b. Texture analysis of magnetic resonance images to predict histologic grade of meningiomas in dogs. American Journal of Veterinary Research 78, 1156–1162.

Banzato, T., Bonsembiante, F., Aresu, L., Zotti, A., 2017c. Relationship of diagnostic accuracy of renal cortical echogenicity with renal histopathology in dogs and cats, a quantitative study. BMC Veterinary Research 13, 24.

Banzato, T., Bonsembiante, F., Aresu, L., Gelain, M.E., Burti, S., Zotti, A., 2018a. Use of transfer learning to detect diffuse degenerative hepatic diseases from ultrasound images in dogs: a methodological study. The Veterinary Journal 233, 35–40.

Banzato, T., Bernardini, M., Cherubini, G.B., Zotti, A., 2018b. A methodological approach for deep learning to distinguish between meningiomas and gliomas on canine MR-images. BMC Veterinary Research 14, 317.

Banzato, T., Cherubini, G.B., Atzori, M., Zotti, A., 2018c. Development of a deep convolutional neural network to predict grading of canine meningiomas from magnetic resonance images. The Veterinary Journal 235, 90–92.

Banzato, T., Causin, F., Puppa, Alessandro Della, Cester, G., Mazzai, L., Zotti, A., Della Puppa, A., Cester, G., Mazzai, L., Zotti, A., 2019. Accuracy of deep learning to differentiate the histopathological grading of meningiomas on MR images: a preliminary study. Journal of Magnetic Resonance Imaging 50, 1152–1159.

Berlin, L., 2007. Accuracy of diagnostic procedures: has it improved over the past five decades? American Journal of Roentgenology 188, 1173–1178.

Boland, M.R., Kraus, M.S., Dziuk, E., Gelzer, A.R., 2018. Cardiovascular disease risk varies by birth month in canines. Scientific Reports 8, 1–11.

Bruno, M.A., Walker, E.A., Abujudeh, H.H., 2015. Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. RadioGraphics 35, 1668–1676.

Buchanan, J.W., Bücheler, J., 1995. Vertebral scale system to measure canine heart size in radiographs. Journal of the American Veterinary Medical Association 206, 194–199.

Cicero, M., Bilbily, A., Colak, E., Dowdell, T., Gray, B., Perampaladas, K., Barfett, J., 2017. Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. Investigative Radiology 52, 281–287.

Croskerry, P., 2009. Clinical cognition and diagnostic error: applications of a dual process model of reasoning. Advices in Health Science Education 14, 27–35.

Da Silva, P.D.A., Uscategui, R.A.R., Santos, V.J.C., Taira, A.R., Mariano, R.S.G., Rodrigues, M.G.K., Simões, A.P.R., Maronezi, M.C., Avante, M.L., Vicente, W.R.R., et al., 2018. Qualitative and quantitative ultrasound attributes of maternal-foetal structures in pregnant ewes. Reproduction in Domestic Animals 53, 725–732.

DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44, 837–845.

Drost, W.T., Henry, G.A., Meinkoth, J.H., Woods, J.P., Lehenbauer, T.W., 2000. Quantification of hepatic and renal cortical echogenicity in clinically normal cats. American Journal of Veterinary Research 61, 1016–1020.

Dunnmon, J.A., Yi, D., Langlotz, C.P., Ré, C., Rubin, D.L., Lungren, M.P., 2019. Assessment of convolutional neural networks for automated classification of chest radiographs. Radiology 290, 537–544.

Gatt, M.E., Spectre, G., Paltiel, O., Hiller, N., Stalnikowicz, R., 2003. Chest radiographs in the emergency department: is the radiologist really necessary? Postgraduate Medical Journal 79, 214–217.

Greco, A., Meomartino, L., Raiano, V., Fatone, G., Brunetti, A., 2008. Effect of left vs. right recumbency on the vertebral heart score in normal dogs. Veterinary Radiology and Ultrasound 49, 454–455.

Grm, K., Struc, V., Artiges, A., Caron, M., Ekenel, H.K., 2018. Strengths and weaknesses of deep learning models for face recognition against image degradations. IET Biometrics 7, 81–89.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27th–30th June 2016, pp. 770–778.

Herman, P.G., Hessel, S.J., 1975. Accuracy and its relationship to experience in the interpretation of chest radiographs. Investigative Radiolology 10, 62–67.

Inoue, M., Hasegawa, A., Hosoi, Y., Sugiura, K., 2015. A current life table and causes of death for insured dogs in Japan. Preventive Veterinary Medicine 120, 210–218.

Ivancic, M., Mai, W., 2008. Qualitative and quantitative comparison of renal vs. hepatic ultrasonographic intensity in healthy dogs. Veterinary Radiolology and Ultrasound 49, 368–373.

Kakeda, S., Moriya, J., Sato, H., Aoki, T., Watanabe, H., Nakata, H., Oda, N., Katsuragawa, S., Yamamoto, K., Doi, K., 2004. Improved detection of lung nodules on chest radiographs using a commercial computer-aided diagnosis system. American Journal of Roentgenology 182, 505–510.

Kelly, B.S., Rainford, L.A., Darcy, S.P., Kavanagh, E.C., Toomey, R.J., 2016. The development of expertise in radiology: in chest radiograph interpretation, 'expert' search pattern may predate 'expert' levels of diagnostic accuracy for pneumothorax identification. Radiology 280, 252–260.

Lakhani, P., Sundaram, B., 2017. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 284, 574–582.

Lamb, C.R., Tyler, M., Boswood, A., Skelly, B.J., Cain, M., 2000. Assesment of the value of the vertebral heart scale in the the radiographic diagnosis of cardiac disease in dogs. Veterinary Record 146, 687–690.

Lamb, C.R., Wikeley, H., Boswood, A., Pfeiffer, D.U., 2001. Use of breed-specific ranges for the vertebral heart scale as an aid to the radiographic diagnosis of cardiac disease in dogs. Veterinary Record 148, 707–711.

Liang, M., Li, Z., Chen, T., Zeng, J., 2015. Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach. IEEE/ACM Transactions in Computer Biology Bioinformatics 12, 928–937.

Lin, F.J., Ligomenides, P.A., Freedmant, M.T., Munt, S.K., 1993. Application of artificial neural networks for reduction of false-positive detections in digital chest radiographs. Proceedings of the Annual Symposium on Computer Applications in Medical Care, Washington, DC, USA, 30th October–3rd November 1993, pp. 434–438.

Longo, F., Nicetto, T., Banzato, T., Savio, G., Drigo, M., Meneghello, R., Concheri, G., Isola, M., 2018. Automated computation of femoral angles in dogs from three-dimensional computed tomography reconstructions: comparison with manual techniques. The Veterinary Journal 232, 6–12.

Marschner, C.B., Kokla, M., Amigo, J.M., Rozanski, E.A., Wiinberg, B., McEvoy, F.J., 2017. Texture analysis of pulmonary parenchymateous changes related to pulmonary thromboembolism in dogs — a novel approach using quantitative methods. BMC Veterinary Research 13, 219.

Mattei, C., Pelander, L., Hansson, K., Uhlhorn, M., Olsson, U., Häggström, J., Ljungvall, I., Ley, C.J., 2019. Renal ultrasonographic abnormalities are associated with low glomerular filtration rate calculated by scintigraphy in dogs. Veterinary Radiology and Ultrasound 60, 432–446.

McCrum-Gardner, E., 2008. Which is the correct statistical test to use? British Journal of Oral and Maxillofacial Surgery 46, 38–41.

Monnier-Cholley, L., MacMahon, H., Katsuragawa, S., Morishita, J., Ishida, T., Doi, K., 1998. Computer-aided diagnosis for detection of interstitial opacities on chest radiographs. American Journal of Roentgenology 171, 1651–1656.

Noguchi, T., Higa, D., Asada, T., Kawata, Y., Machitori, A., Shida, Y., Okafuji, T., Yokoyama, K., Uchiyama, F., Tajima, T., 2018. Artificial intelligence using neural network architecture for radiology (AINNAR): classification of MR imaging sequences. Japanese Journal of Radiology 5–11.

O'Neill, D.G., Church, D.B., McGreevy, P.D., Thomson, P.C., Brodbelt, D.C., 2013. Longevity and mortality of owned dogs in England. The Veterinary Journal 198, 638–643.

Prieto, A., Prieto, B., Martinez Ortigosa, E., Ros, E., Pelayo, F., Ortega, J., Rojas, I., 2016. Neural networks: an overview of early research, current frameworks and new challenges. Neurocomputing 214, 242–268.

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., et al., 2017. CheXNet: Radiologist-level Pneumonia Detection on Chest X-rays with Deep Learning. arXiv:1711.05225. .

Rodrigues Simões, A.P., Rossi Feliciano, M.A., Maronezi, M.C., Uscategui, R.A.R., Bartlewski, P.M., de Almeida, V.T., Oh, D., do Espírito Santo Silva, P., da Silva, L.C. G., Russiano Vicente, W.R., 2018. Elastographic and echotextural characteristics of foetal lungs and liver during the final 5 days of intrauterine development in dogs. Animal Reproduction Science 197, 170–176.

Seah, J.C.Y., Tang, J.S.N., Kitchen, A., Gaillard, F., Dixon, A.F., 2019. Chest radiographs in congestive heart failure: visualizing neural network learning. Radiology 290, 514–522.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7th–9th May 2015, pp. 1–14.

Suzuki, K., Shiraishi, J., Abe, H., MacMahon, H., Doi, K., 2005. False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network. Academic Radiology 12, 191–201.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2015. Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27th–30th June 2016, pp. 2818–2826.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4, inception-ResNet and the impact of residual connections on learning. Proceedings of the 31st AAAI Conference on Artificial Intelligence, San Fancisco, CA, USA, 4th–9th February 2017, pp. 4278–4284.

Vaccaro, C., Busetto, R., Bernardini, D., Anselmi, C., Zotti, A., 2012. Accuracy and precision of computer-assisted analysis of bone density via conventional and

digital radiography in relation to dual-energy X-ray absorptiometry. American Journal of Veterinary Research 73, 381–384.
Waite, S., Kolla, S., Reede, D., Gale, B., Fuchs, T., Scott, J., 2016. Interpretive error in radiology. American Journal of Roentgenology 208, 739–749.
Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M., 2017. ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. Proceedings of the IEEE Conference on Computer Vision Pattern Recognition, Honolulu, HI, USA, 21st–26th July 2017,, pp. 3462–3471.
Yasaka, K., Abe, O., 2018. Deep learning and artificial intelligence in radiology: current applications and future directions. PLoS Medicine 15, 2–5.
Zech, J.R., Badgeley, M.A., Liu, M., Costa, A.B., Titano, J.J., Oermann, E.K., 2018. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLoS Medicine 15, 1–17.
Zotti, A., Banzato, T., Gelain, M.E., Centelleghe, C., Vaccaro, C., Aresu, L., 2015. Correlation of renal histopathology with renal echogenicity in dogs and cats: an ex-vivo quantitative study. BMC Veterinary Research 11, 99.