Research in Veterinary Science 175 (2024) 105317

Contents lists available at ScienceDirect

Research in Veterinary Science

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

Artificial intelligence in veterinary diagnostic imaging: Perspectives and limitations

Silvia Burti [a,*], Tommaso Banzato [a], Simon Coghlan [b], Marek Wodzinski [c,d], Margherita Bendazzoli [a], Alessandro Zotti [a]

[a] Department of Animal Medicine, Production and Health, University of Padua, Viale dell'Università 16, Legnaro, 35020 Padua, Italy
[b] School of Computing and Information Systems, Centre for AI and Digital Ethics, Australian Research Council Centre of Excellence for Automated Decision-Making and Society, University of Melbourne, 3052 Melbourne, Australia
[c] Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Krakow, 30059 Kraków, Poland
[d] Information Systems Institute, University of Applied Sciences - Western Switzerland (HES-SO Valais), 3960 Sierre, Switzerland

ARTICLE INFO

Keywords:
Artificial intelligence
Machine learning
Deep learning
Veterinary diagnostic imaging
Convolutional neural network
Ethics

ABSTRACT

The field of veterinary diagnostic imaging is undergoing significant transformation with the integration of artificial intelligence (AI) tools. This manuscript provides an overview of the current state and future prospects of AI in veterinary diagnostic imaging.

The manuscript delves into various applications of AI across different imaging modalities, such as radiology, ultrasound, computed tomography, and magnetic resonance imaging. Examples of AI applications in each modality are provided, ranging from orthopaedics to internal medicine, cardiology, and more. Notable studies are discussed, demonstrating AI's potential for improved accuracy in detecting and classifying various abnormalities.

The ethical considerations of using AI in veterinary diagnostics are also explored, highlighting the need for transparent AI development, accurate training data, awareness of the limitations of AI models, and the importance of maintaining human expertise in the decision-making process. The manuscript underscores the significance of AI as a decision support tool rather than a replacement for human judgement.

In conclusion, this comprehensive manuscript offers an assessment of the current landscape and future potential of AI in veterinary diagnostic imaging. It provides insights into the benefits and challenges of integrating AI into clinical practice while emphasizing the critical role of ethics and human expertise in ensuring the wellbeing of veterinary patients.

1. Introduction

Nowadays, diagnostic imaging is a fundamental step in the clinical evaluation of veterinary patients, and this is driving a growing demand for veterinarians specialized in this field. Diagnostic imaging is also complementary to several other disciplines, such as internal medicine, surgery, neurology, oncology, and obstetrics. Given this context the quick and reliable interpretation of veterinary imaging data is evidently paramount for every busy veterinary clinical practice. Although no official data is available on this occurrence, those people currently working in small animal clinics commonly find that the number of diagnostic imaging studies produced far exceeds specialists' ability to assess them. Hence, the need for tools and solutions to assist general practitioners during their in daily routine is steadily increasing, and artificial intelligence (AI) tools are gaining popularity, due to the promise of their ease of use and potentially unlimited scope. In fact, the AI tools market is swiftly expanding as more and more potential applications are tested out and the number of market players is growing.

The core idea of AI is to create machines capable of performing actions that typically require human abilities (Lecun et al., 2015). AI is a broad term including a variety of different algorithm types, such as machine learning (ML), which in turn encompasses the most complex deep learning (DL) algorithms. ML algorithms often require manual feature extraction and validation, while DL leverages deep neural networks to automatically extract complex features from raw data, thus overcoming the need for manual feature engineering. The recent flourishing of both research and commercial applications using deep neural networks for the automatic interpretation of diagnostic images has been

* Corresponding author.
E-mail address: silvia.burti@unipd.it (S. Burti).

https://doi.org/10.1016/j.rvsc.2024.105317
Received 14 March 2024; Received in revised form 22 May 2024; Accepted 29 May 2024
Available online 31 May 2024

driven in part by the lack of available certified veterinary radiologists, and the need to support overworked specialists and, therefore, prone to interpretation errors. Last but not least, the opportunity to establish an economically viable enterprise represents another important driving force.

In human medicine, the "effect of" of diagnostic imaging errors has been thoroughly investigated, mainly in terms the potential impact of such errors on the patients and, ultimately, on hospital administrations (Siewert et al., 2008; Yun et al., 2017a). Oversight involving a lesion in an overlooked anatomical area (often referred to as a blind spot), failure to detect lesions due to improper use of the window settings, superimposition of lesions, or atypical presentation of a lesion are all examples of factors that can lead to errors in interpretation in diagnostic imaging (Yun et al., 2017a). Despite advancements in experience, knowledge and technology over recent decades, the error rate in diagnostic imaging has remained constant (e.g. <15% for thoracic radiographic studies, and 16% for abdominal tomographic studies) (Yun et al., 2017b). Several strategies to reduce human error - such as using dynamic (e.g. tomographic) rather than static (e.g. radiographic) images to mitigate the effects of fatigue, or having a second reader double-read the images - have been proposed (Degnan et al., 2019; Krupinski et al., 2012). The deployment of AI-based products to assist radiologists during image acquisition and interpretation is recommended by several authors in human medicine literature (Degnan et al., 2019; Hardy and Harvey, 2020). The prevalence of interpretation errors in veterinary medicine has not been studied as extensively as in human medicine (Cohen et al., 2023; Lamb et al., 2007).

The literature on DL algorithms applied in the field of veterinary image diagnostics is of a relatively limited quantity compared to what is available in human medical literature. The development of these technologies has indeed reached outstanding levels across all application domains in human medicine, facilitated by extensive databases and driven by the need for advanced tools to support radiologists. Conversely, the available veterinary literature is still quite patchy, and only a few studies span different application fields.

The purposes of this review are: to assess the current state of the art of AI in veterinary diagnostic imaging; to provide the reader with an overview of the possible future directions of AI in both research and the market; and to review the AI ethics that all developers and retailers should respect.

2. Applications of AI in veterinary diagnostic imaging

2.1. Conventional radiology

Twenty-four papers on conventional radiology AI applications are presented below and synthetically reported in Table 1.

2.1.1. Orthopaedics

There is a lack of studies exploring AI use on orthopaedic imaging; the few available papers focus on the evaluation of hip radiographs. A handful of articles have been published on stifle joints and long- bone fractures.

A recent online search of the PubMed, Scopus, and Web Of Science databases from January 2000 to November 2023, using the keywords ("radiology" OR "radiographic" AND "elbow" AND "veterinary" OR "veterinary medicine") failed to reveal AI use for the automatic evaluation of elbow dysplasia. McEvoy and Amigo (Mcevoy and Amigo, 2013) developed a deep neural network (Artificial Neural Network- ANN) and a linear regression model to determine whether the hip was present in an image. The ML model performed better, showing a higher sensitivity (89% vs. 86%) and a lower classification error (6.7% vs. 8.9%) than the ANN did, despite the database being small (256 hip X-ray images from 60 dogs) for a neural network. McEvoy et al. (McEvoy et al., 2021) fine-tuned a pre-trained convolutional neural network (CNN), YOLO v3-Tiny, using a database of almost 16,300 hip radiographs classified according to Fédération Cynologique Internationale (FCI) scores, both for hip joint detection and for binary classification of dysplasia (grades A-B vs. C-E). The model showed a poor sensitivity (53%), but high specificity, positive and negative predictive values (92%, 91% and 81%, respectively). Ergun and Guney (Ergun and Guney, 2021) used 3437 images of long bones (femur, tibia, humerus, radius-ulna) to train three different CNNs to solve three tasks: determining dog age; dating fractures; detecting fractures. The CNN based on the ResNet-50 architecture showed the best performance (F1 = 0.80, 0.81 and 0.89 respectively for the three tasks), with high accuracy (0.80, 0.83 and 0.89, respectively). A CNN to classify common canine stifle joint diseases (patellar deviation, drawer sign, osteophyte formation, and presence of joint effusion) was developed by Shim et al. (Shim et al., 2023). The study was designed with two steps. During the first step. 200 cropped X-rays were used to train an Inception-v2-based model to detect the main components of the joint (stifle joint region itself, patellar, and infrapatellar fat pad). The results were good, with an average precision (AP) between 0.99 and 1, meaning that the model was able to correctly classify the joint components in almost all cases. During the second step 2218 radiographs were used to train a ResNet-based model to classify stifle abnormalities: accuracy values ranged between 81.25% (for drawer sign) and 93.18% (for patellar deviation), and sensitivity between 79.41% (for drawer sign) and 89.7% (for patellar deviation). These results were then compared with the classifications performed by expert radiologists, with congruency values >80% for all the stifle joint diseases considered.

2.1.2. Thorax

The automatic evaluation of thoracic radiographs is the most widely investigated application of CNN-based models in veterinary literature. Interestingly, most of the studies focus on dogs, while just two are on cats. Yoon et al. (Yoon et al., 2018) compared a ML algorithm (bag of feature - BOF) with a CNN to evaluate - using a binary classification of normal vs. abnormal - the cardiac silhouette, lung, presence of mediastinal shift, and pleural space. In particular, the algorithms were trained to distinguish between the normal or abnormal appearance of each area separately. Unsurprisingly, the CNN model showed higher accuracy (92.9% - 96.9% vs. 79.6% - 96.9% for BOF) and sensitivity (ranging from 92.1% - 100% vs. 74.1–94.8% for BOF).

In Kim et al. (Kim et al., 2022) the authors used a commercially a commercially available AI-based software (Vetology Innovations, San Diego, CA, USA) for possible cardiogenic pulmonary edema (CPE) on 500 canine thoracic radiographs. The same radiographs were also evaluated by a certified radiologist, whose reports were used as reference standards for final diagnoses. The AI Vetology software discarded 19 images for technical or quality issues, and obtained accuracy, specificity and sensitivity of 92.3%, 92.4% and 91.3% respectively. On the other hand, a negative predictive value (NPV) of 99% was reached, with a positive predictive value (PPV) of 56%, suggesting that if a diagnosis of presence of CPE is given by the software, a more careful evaluation of the patient should be done.

The same AI-based software was used in two other studies. The first one by Muller et al. (Müller et al., 2022) took into consideration the presence of pleural effusion on a relatively small dataset of radiographs (62 dogs, 21 as control group and 41 as confirmed pleural effusion cases). In this case, the software showed accuracy, sensitivity and specificity of 88.7%, 90.2% and 85.7% respectively, with NNP and PPV of 81.8% and 92.5%. The second one by Pomerantz et al. (Pomerantz et al., 2023) used Vetology AI-based software to evaluate the CT-confirmed presence of pulmonary nodules/masses in 56 dogs. A control group of 32 dogs with normal thorax was also included in the study. The AI model showed accuracy, sensitivity and specificity of 69.3%, 55.4% and 93.7% respectively. This study reiterated the capacity of these algorithms to assist in veterinary daily routines while underscoring the ongoing importance of patient assessment by an expert radiologist.

In Ott et al. (Ott et al., 2021), the authors trained and compared five different DL models for detecting pulmonary coccidioidomycosis

Table 1
Peer-reviewed veterinary AI publications concerning conventional radiography.

Topic	Reference	Task	Species	Model	Results
Orthopaedics					
	Mcevoy and Amigo, 2013	hip identification	dog	ML	sensitivity 89%
					classification error 6.7%
				DL	sensitivity 86%
					classification error 8.9%
	McEvoy et al., 2021	hip dysplasia classification	dog	DL	sensitivity 53% specificity 92% PPV 91% NPV 81%
	Ergun and Guney, 2021	determining age from long bones images	dog	DL	accuracy 80%; F1 0.80
		dating long bones fractures			accuracy 83%; F1 0.81
		detecting long bones fractures			accuracy 89%; F1 0.89
	Shim et al., 2023	stiffle joint components identification	dog		avarage precision >0.99
		stiffle joint abnormalities classification		DL	accuracy 81.25% - 93.18% sensitivity 79.41% - 89.70%
Thorax					
	Yoon et al., 2018	normal vs abnormal cardiac silhouette and thoracic portions	dog	ML	accuracy 79.6% - 96.9%
					sensitivity 92.9% - 96.9%
				DL	accuracy 74.1–94.8%
					sensitivity 92.1% - 100%
	Kim et al., 2022	presence/absence of cardiogenic pulmonary edema	dog	Vetology®	accuracy 92.3% sensitivity 91.3% specificity 92.4%
	Müller et al., 2022	presence of pleural effusion	dog	Vetology®	accuracy 88.7% sensitivity 90.2% specificity 85.7%
	Pomerantz et al., 2023	presence of pulmonary nodules/masses	dog	Vetology®	accuracy 69.3% sensitivity 55.4% specificity 93.7%
	Ott et al., 2021	detecting pulmonary coccidioidomycosis lesions	dog	DL	AUC of 0.99
	Banzato et al., 2021a	detecting common radiographic findings	dog	DL	accuracy >90%
	Fitzke et al., 2021	detecting thoracic and extra-thoracic radiographic abnormalities	dog+cat	DL	AUROC 0.687–0.994
					FPR 0–0.057 Sensitivity 0–0.962
	Banzato et al., 2021b	detecting common radiographic findings	dog	DL	AUC >0.5 - >0.8
	Dumortier et al., 2022	normal vs abnormal	cat	DL	accuracy 82% sensitivity 88% specificity 75%
	Boissady et al., 2020	primary thoracic lesions classfications	dog	DL	overall error rate 10.7%
				Radiologist	overall error rate 17.2%
				DL + Radiologist	overall error rate 16.8%
	Hespel et al., 2022	primary thoracic lesions classfications	dog	DL	overall error rate 15.8%
				Radiologist	overall error rate 13%
	Celniak et al., 2023	primary thoracic lesions classfications	dog+human	DL	ROC AUC 0.77 (LL radiographs) ROC AUC 0.66 (DV radiographs)
Cardiac silhouette					
	Li et al., 2020	detecting left atrial enlargement	dog	DL	accuracy 82.71% sensitivity 68.42% specificity 87.09% concordance with radiologist 85.19%
	Burti et al., 2020	classification of cardiomegaly based on VHS value	dog	DL	AUC of 0.97
	Boissady et al., 2021	automatically measuring VHS	dog + cat	DL and Radiologist	concordance >0.9
	Zhang et al., 2021	identification of landmarks for calculating VHS	dog	DL	average performance 90.9%
	Jeong and Sung, 2022	determining adjusted heart volume index (aHVI)	dog	DL	ROC AUC 0.77–0.83
	Valente et al., 2023	classification of canine MMVD stages	dog	DL	AUC 0.77–0.88 for different stages
Image quality analysis					
	Tahghighi et al., 2023	assessing proper collimation	dog + cat	DL	accuracy 83.17%
	Banzato et al., 2023	determining the most common technical errors	dog	DL	accuracy 81.5% (LL radiographs) accuracy 75.7% (sagittal radiographs)

ML: machine learning; DL: deep learning; PPV: positive predictive value; NPV: negative predictive value; AUC: area under the curve; AUROC: area under the receiver operating characteristic; FPR: false positive rate; ROC: receiver operating characteristic; VHS: vertebral heart score; aHVI: adjusted heart volume index; MMVD: myxomatous mitral valve disease; LL: latero-lateral; DV: dorso-ventral.

lesions, a parasitic infection and potential zoonosis that is increasingly prevalent in the southern United States. Radiographic signs of coccidioidomycosis are an interstitial nodular pattern with hilar lymphadenopathy. Once again, ResNet outperformed the remaining networks, with an (Area Under the Curve) AUC of 0.99. Banzato et al. (Banzato et al., 2021a) trained two different CNNs (ResNet-50 and DenseNet-121) on canine thoracic radiographs to detect some of the most common radiographic findings. The best-performing CNN was based on ResNet-50 with an overall AUC >0.8. Accuracy was particularly high (>0.9) for alveolar pattern, interstitial pattern, megaesophagus, and pneumothorax.

In a noteworthy study by Fitzke et al.(Fitzke et al., 2021), a multicentric dataset comprising over 2.5 million canine and feline thoracic and extra-thoracic radiographs was employed to train a DL model for the recognition of various abnormalities. A natural language processing (NLP) algorithm was used to extract labels from the corresponding reports. The study reported promising results in terms of AUROC (ranging from 0.687 to 0.994), false positive rate (ranging from 0 to 0.057), and sensitivity (ranging from 0 to 0.962). However, the most noteworthy achievement was a significant improvement in the generalizability of the overall results.

The two studies on the analysis of feline thoracic radiographs are by Banzato et al. (Banzato et al., 2021b) and Dumortier et al. (Dumortier et al., 2022). In Banzato et al. (Banzato et al., 2021b), two neural networks (ResNet 50 and Inception V3) were trained on the detection of some of the most common radiographic findings (bronchial pattern, pleural effusion, mass, alveolar pattern, pneumothorax, cardiomegaly, no findings)(Banzato et al., 2021b). Both networks achieved good performances (AUC >0.8) for almost all findings, except for cardiomegaly (AUC > 0.7) and mass (AUC > 0.5). Dumortier et al. (Dumortier et al., 2022) used the ResNet50V2 neural network to classify feline thoracic images as normal or abnormal by using the contrast-enhanced method and manually segmented images(Dumortier et al., 2022). Despite its promising results, the small database (500 radiographs) and the need for manual image segmentation severely limit application of this method in clinical practice.

In Boissady et al. (Boissady et al., 2020) the authors aimed to establish the error rate for three different CNNs in classifying 15 primary thoracic lesions. Once the best-performing network was selected, the error rates for the neural network alone, the veterinarians alone, and the CNN-assisted veterinarians were compared and analysed. The CNN alone showed a significantly lower global error rate for some lesions (e.g.: cardiac enlargement and bronchial pattern) than the unassisted veterinarians or the CNN-assisted veterinarians did. The authors hypothesized this may have been due to the veterinarians' poor confidence in the relatively new CNN tool and its proposed classification. Hespel et al. (Hespel et al., 2022) instead compared the error rates for four pre-trained CNNs with those for 13 veterinary radiologists in classifying 15 primary thoracic lesions. In this research, the radiologists' error rate was equal to (and sometimes lower than) that of at least one of the four pre-trained networks, except in the case of oesophageal dilation, where two of the four networks had significantly lower error rates than their human counterpart.

Celniak et al. (Celniak et al., 2023) recently pre-trained a DL algorithm on a large-scale database containing both human and canine data: using a self-supervised learning approach, this achieved higher classification accuracy for some lesions (pleural effusion, pneumothorax) than customary DL approaches did (Banzato et al., 2021a).

2.1.3. Cardiac silhouette

Several studies specifically on developing AI systems for the automatic evaluation of cardiac silhouette have been published.

Li et al. (Li et al., 2020) developed a CNN (Visual Geometry Group 16) to detect left atrial enlargement on lateral thoracic radiographs. A database of 792 patients' echocardiograms classified as "positive" or "negative" for left atrial enlargement was used, and the classifications by the CNN and by the board-certified radiologists were performed. The developed CNN-based model classified the images with an 82.71% accuracy, a 68.42% sensitivity and an 87.09% specificity, and with an 85.19% concordance between CNN and veterinary radiologists. Burti et al. (Burti et al., 2020) compared four different CNNs' accuracy for the binary (positive or negative) classification of cardiomegaly based on Vertebral Heart Score (VHS): the best performing CNN was based on ResNet-101 and had an AUC of 0.97. Boissady et al. (Boissady et al., 2021) trained a 121-layer DenseNet CNN to automatically measure VHS on lateral radiographs, albeit with a relatively small test set (30 canine and 30 feline thoracic radiographs). Agreement between the VHS measurements by the two board-certified radiologists involved and the AI algorithm-measured VHS was high (>0.9) for both the canine and the feline radiographs. Zhang et al. (Zhang et al., 2021) published a technical study on training neural networks to identify the necessary landmarks for calculating VHS in lateral radiographs. More specifically, HRNets were trained to locate 12 thoracic vertebrae and the four points on the cardiac silhouette required for tracing the long and short axes, resulting in an average performance of 90.9%. In everyday clinical practice, this could assist veterinarians in calculating VHS more objectively. In Jeong and Sung (Jeong and Sung, 2022), a new automated method for evaluating cardiac dimensions based on the estimated entire cardiac area, namely adjusted heart volume index (aHVI), was developed. This method showed a similar performance to VHS in predicting left atrial (LA) enlargement, left ventricular (LV) enlargement and combined LA + LV enlargement. In Valente et al. (Valente et al., 2023), a ResNet18-based CNN was trained to classify different stages (B1, B2, C + D) of canine myxomatous mitral valve disease based on the American College of Veterinary Internal Medicine guidelines. The algorithm was trained on two distinct sets of radiographs, consisting of 728 sagittal and 514 lateral images. Notably, the best classification performance was achieved for the lateral images, with AUC values of 0.87, 0.77 and 0.88 for the B1, B2 and C + D stages, respectively.

2.1.4. Radiograph image quality analysis

Obtain high-quality images is paramount to achieve a correct diagnosis, regardless of which diagnostic imaging technique is used. Two studies focused on the application of DL algorithms for assessing radiographic image quality exist in the veterinary literature. In the first study, by Taghighi et al. (Tahghighi et al., 2023), a multilayer perceptron neural network was trained to assess proper collimation on a group of 900 sagittal chest radiographs of dogs and cats. The algorithm was trained to recognize the inclusion of lung fields by evaluating the cranial and caudal boundaries within the previously segmented thoracic area. The model achieved accuracy values of 83.17% and an F1 score of 87%. The second study, by Banzato et al. (Banzato et al., 2023), focused more specifically on determining the most common technical errors in canine thoracic radiographs. To do this, the authors fine-tuned the pre-trained ResNet-50 network on a dataset of 6028 lateral and 4053 sagittal radiographs, aiming to classify each image as correct or as having at least one of the following errors: rotation, underexposure, overexposure, limb mispositioning, neck mispositioning, blurriness, cut-off, and presence of foreign objects/medical devices. The algorithm showed an overall accuracy of 81.5% for lateral radiographs, and 75.7% for sagittal images, with the most common technical errors both in lateral and sagittal radiographs being limb mispositioning and underexposure.

2.2. Ultrasound

Only two studies to date have highlighted of AI-based algorithms' potential in the field of veterinary ultrasonography, focusing particularly on ML with texture analysis, and with both studies primarily on liver-related applications. The first study (Banzato et al., 2015) combined ultrasound and cytology to assess diagnostic efficacy in cases of diffuse liver disease in dogs and cats. The second study (Banzato et al., 2016) used texture analysis in combination to estimate the triacylglycerol content of the livers of dairy cows.

In the research by Banzato et al. (Banzato et al., 2018a), a transfer-learning approach was applied to modify and enhance a pre-trained neural network (AlexNet) to detect degenerative hepatic diseases from ultrasound images of canine livers. This study involved 48 dogs suspected of having liver disease, confirmed by standard histopathology results. The new model's output demonstrated greater diagnostic accuracy than did the results from serum biochemical markers (alanine aminotransferase and aspartate aminotransferase), hepatic cytology, and histopathology: the model correctly classified 82% of the test-set images, with an AUC of 0.91, a sensitivity of 100%, and a specificity of 83%.

2.3. Computed tomography

AI-based methods for the automatic evaluation of veterinary computed tomography (CT) images have only seldom been explored in veterinary literature. The main AI applications in CT are for automatic classification of images and in for automatic segmentation of lesions. Eleven papers are presented below and synthetically reported in Table 2.

2.3.1. Automatic classification of abdominal and pulmonary lesions

To date all the available papers using AI for classification purposes have been based on applying ML algorithms to manually crafted descriptors (lesion Hounsfield Unit, lesion dimensions, margin characteristics, etc..) or to lesion texture analysis for case classification. Burti et al. (Burti et al., 2021) proposed a decision tree to classify focal liver lesions based on several imaging descriptors such as well-defined margins, irregular surface, or cyst-like appearance. The developed decision tree had a moderate accuracy (62%) for classifying lesions. In particular, this decision tree had a high specificity in classifying a category named "other benign lesions" (containing all the benign lesion other than nodular hyperplasia), and hepatocarcinoma. Pey et al. (Pey et al., 2022) established a 7-point scale for predicting the vascular invasion of adrenal tumors. Here a decision tree was developed, based on imaging features, to distinguish between pheochromocytoma and adrenocortical tumour. Unfortunately, several histopathological categories were not included in the decision tree, significantly restricting its clinical applications. Burti et al. (Burti et al., 2022) also developed a decision tree for classifying splenic lesions from their CT features. In this case too, some histopathological categories were not included in the decision tree, largely limiting its clinical applications. Shaker et al. (Shaker et al., 2021) used ML on texture features extracted from CT images to distinguish between benign and malignant hepatic lesions, deploying a quadratic discriminant analysis model for lesion classification. This model had a moderate accuracy (0.73) in distinguishing between benign and malignant focal liver lesions. Marschner et al. (Marschner et al., 2017) used texture analysis of CT lung images to detect thromboembolism-related changes in dogs. Here the authors used least square discriminant analysis and support vector machines. In this study, the authors have used Least Square Discriminant Analysis and Support Vector Machines to distinguish between normal lungs, diseased lungs affected by pulmonary thromboembolism, and diseased lungs without pulmonary thromboembolism. The developed method had high accuracy in a binary classification task (normal vs diseased) but failed to distinguish between diseased lungs affected by pulmonary thromboembolism and diseased lungs without pulmonary thromboembolism. Lastly, Choi et al. (Choi et al., 2023a) developed a ML method based on texture analysis to distinguish between splenic nodular hyperplasia and splenic hemangiosarcoma. However, the small number of cases included in the study (23) and the limited number of histopathological classes included (nodular hyperplasia and splenic hemangiosarcoma) both make the usefulness of this method questionable.

2.3.2. Automatic segmentation of organs and lesions

The other important application of AI on CT images is lesion detection and segmentation. This task is important mainly for radiation oncology but can be also implemented in other clinical settings. A special type of DL architecture called U-Net (Ronneberger et al., 2015). The U-Net model and its variants are examples of hierarchical encoder-

Table 2
Peer-reviewed veterinary AI publications concerning computed tomography.

Topic	Reference	Task	Species	Model	Results
Automatic classification of abdominal and pulmonary lesions					
	Burti et al., 2021	classification of focal liver lesions	dog	ML	accuracy 62%
	Pey et al., 2022	distinction between pheochromocytoma and adrenocortical tumour	dog	ML	CVC invasion: accuracy 81% - 91% RV invasion: accuracy 60% - 88% PAV invasion: accuracy 58% - 60%
	Burti et al., 2022	classification of focal splenic lesions	dog	ML	overall accuracy 0.67
	Shaker et al., 2021	distinguish between benign and malignant hepatic lesions	dog	ML	accuracy 73%
	Marschner et al., 2017	detection of pulmonary thromboembolism	dog	ML (PLS-DA)	sensitivity 94% specificity 96%
				ML (SVM)	sensitivity 99% specificity 100%
	Choi et al., 2023a	distinction between splenic nodular hyperplasia and hemangiosarcoma	dog	ML	accuracy 95.7%
Automatic segmentation of organs and lesions					
	Ji et al., 2022	evaluation of kidney volume	dog	DL	Lin's CCC 0.95
	Ji et al., 2023	detecting kidney stones on pre-contrast CT	dog	DL	DSC 0.74
	Park et al., 2021	segmentation of head and neck organs	dog	DL	DSC 0.83
	Groendahl et al., 2023	segmentation of canine head and neck cancer	dog	DL	overall DSC 0.52
	Schmid et al., 2022	segmentation of the medial retropharyngeal lymph nodes	dog	DL	AIOU of 36% ± 20%

ML: machine learning; CVC: caudal vena cava; RV: renal vein; PAV: phrenicoabdominal vein; PLS-DA: partial least square discriminant analysis; SVM: support vector machine; DL: deep learning; Lin's CCC: Lin's concordance correlation coefficient; CT: computed tomography; DSC: Dice similarity coefficient; AIOU: average intersection over union.

decoder architecture where the encoder is responsible for automatic feature extraction, and the decoder learns how to aggregate the features to calculate the segmentation maps. When paired with the residual connections, this hierarchical structure enables the U-Net architecture to learn features at different scales and complexities, thus resulting in robust and reliable feature extraction and gradient propagation. The literature offers several contributions related to U-Net architecture use in veterinary imaging. Ji et al. (Ji et al., 2022) developed a DL algorithm for the automated evaluation of kidney volume. Here the authors have implemented several different variations of U-Net, and the best-performing architecture had a Lin's concordance correlation coefficient of 0.95 thus enabling a very accurate estimation of kidney volume. Shortly afterwards, the same authors developed a U-Net architecture capable of detecting kidney stones on pre-contrast CT scans in dogs (Ji et al., 2023). Other authors have instead focused on the automatic segmentation of head and neck structures. In particular, Park et al. (Park et al., 2021) developed and tested a DL algorithm for the automatic segmentation of some of the main head and neck organs for radiotherapy (e.g. the brain, the mandibular salivary glands, the pharynx), with an average Dice Similarity Score (DSC) of 0.83. Groendahl et al. (Groendahl et al., 2023) proposed a very interesting inter-species approach to developing a DL algorithm capable of automatically segmenting canine head and neck cancer. Here the authors referred to both human and canine patients affected by head and neck cancer to create a mixed CT database for the automatic segmentation of head and neck cancer. The overall DSC was not very high (0.52) but was higher for nasal tumors (0.69). Despite the modest overall results, the mixed human-canine training could be a game-changing approach for the development of AI based tools both in human and veterinary medicine. Lastly, Schmid et al. (Schmid et al., 2022) developed an algorithm for the automatic segmentation of the medial retropharyngeal lymph nodes in dogs: its accuracy was fair, despite the limited number of cases (40) included in the study.

2.4. Magnetic resonance imaging

Eight papers on magnetic resonance imaging AI applications are presented below and synthetically reported in Table 3.

2.4.1. Distinction between different types of brain diseases

Different approaches have been proposed in the literature for the distinction between brain diseases on magnetic resonance imaging (MRI) scans. Banzato et al. (Banzato et al., 2018b) used a DL-based approach to distinguish between meningiomas and gliomas. The best classifier was developed by combining post-contrast T1 images and GoogleNet, and showed a 91% accuracy in the test set. Wanamarker et al. (Wanamaker et al., 2021) used a texture-based approach to distinguish between neoplastic and inflammatory brain disease on MRI images. This approach showed a high accuracy (85%) in the distinction between inflammatory and neoplastic disease, but the authors reported a low accuracy in distinguishing between different inflammatory subtypes.

Spiteri et al. (Spiteri et al., 2019) developed developed an ML-based approach to identify candidate biomarkers of the morphological changes associated with Chiari-like malformations in Cavalier King Charles Spaniels. Such an approach is particularly meaningful as it not only provides a potentially helpful tool in the diagnosis of Chiari-like malformations but also takes a further step towards understanding the pathogenesis of this disease.

2.4.2. Prediction of the grading of certain intracranial diseases

Predicting intracranial neoplasias grade has been used both for meningiomas and gliomas. Banzato et al. (Banzato et al., 2017) developed a texture analysis-based tool to predict meningiomas grading in dogs. This model was developed starting from 58 histologically proven meningiomas and had a very high accuracy (96.8%) in distinguishing between benign and atypical-anaplastic meningiomas. Nevertheless, the limited database variability and the complex analysis required both restrict the clinical usefulness of this model. The following year the same authors developed a DL-based model for the same task (Banzato et al., 2018c). Despite its lower overall accuracy (80%) its ease of implementation makes this approach much more promising. Barge et al. (Barge et al., 2023) used texture analysis to predict the histopathological grading of canine gliomas. Despite the limited database (38 dogs with 40 lesions in total), the overall case classification accuracy was 77% when using a leave-one-out classification scheme. Interestingly, the authors reported that peritumoral oedema on T1 weighted images and non-enhancing parts of the tumour were the most discriminative features, indicating that texture analysis could be directly used in a clinical context.

Table 3
Peer-reviewed veterinary AI publications concerning magnetic resonance imaging.

Topic	Reference	Task	Species	Model	Results
Distinction between different types of brain diseases					
	Banzato et al., 2018a	distinction between meningiomas and gliomas	dog	DL	accuracy 91%
	Wanamaker et al., 2021	distinguish between neoplastic and inflammatory brain disease	dog	DL	accuracy 85%
	Spiteri et al., 2019	identification of morphological changes associated with Chiari-like malformations	dog (CKCS)	DL	AUC 0.78–0.82 sensitivity 82% - 93% specificity 67% - 69%
Prediction of the grading of certain intracranial diseases					
	Banzato et al., 2017	predicting meningiomas grading	dog	ML	accuracy 96.8%
	Banzato et al., 2018b	predicting meningiomas grading	dog	ML	accuracy 80%
	Barge et al., 2023	prediction of gliomas histopathological grading	dog	ML	accuracy 77%
Detection of spinal cord diseases					
	Biercher et al., 2021	detection of spinal cord diseases	dog	DL	sensitivity 90.8% - 100% specificity 95.1% - 98.98%
Improvement of MR image quality					
	Choi et al., 2023b	reduce scan time and improve image quality	dog	DL	scan time reduction up to 75%

DL: deep learning; CKCS: Cavalier King Charles Spaniel; AUC: area under the curve; ML: machine learning; MR: machine resonance.

2.4.3. Detection of spinal cord diseases

To date, only one study exploring DL for use in the detection of spinal cord disease on MRI is available (Biercher et al., 2021). Biercher et al. (Biercher et al., 2021) developed a CNN capable of detecting intervertebral disc extrusion (IVDE), intervertebral disc protrusion (IVDP), fibrocartilaginous embolism (FCE)/acute non-compressive nucleus pulposus extrusion (ANNPE), syringomyelia and neoplasia. The developed CNN had a very high accuracy in detecting IVDP and IVDE with a 100% sensitivity and 95.1% specificity, and a 90.8% sensitivity and 98.98% specificity, respectively. The CNN's accuracy for the other diseases was lower but still acceptable.

2.4.4. Improvement of MR image quality

One of the most investigated applications of AI regarding MRI is improved image quality with a shorter scanning time (Yutong Chen et al., 2022); since it is a very technical aspect it is usually investigated in specialty journals. Nevertheless, Choi et al. (Choi et al., 2023b) developed a DL-based algorithm to improve specific as well as overall image quality; due to the lower number of excitations used it significantly reduced scanning time.

3. Ethical considerations in AI-driven veterinary diagnostics

Veterinary professionals have a basic duty to promote their animal patients' wellbeing while avoiding doing harm (Coghlan, 2018; Hernandez et al., 2018; Rollin, 2006). Promoting patient interests can also enhance human-animal bonds (Walsh, 2009) and improve professional satisfaction. Little has been written about veterinary AI ethics, but see Coghlan and Quinn (Coghlan and Quinn, 2023a); Cohen and Gordon (Cohen and Gordon, 2022).

AI can have problematic effects in healthcare (Morley et al., 2020; Topol, 2019). For instance, one ML model designed to predict human pneumonia risk incorrectly gave lower risk scores for asthma sufferers due to features of the medical record data used in training the model, thereby potentially causing serious patient harm (Caruana et al., 2015). Veterinary radiological AI too carries risks alongside its potential benefits (Coghlan and Quinn, 2023b; Cohen and Gordon, 2022).

Radiological AI that approaches or exceeds the performance of experts may sometimes allow more accurate and speedier diagnosis (Basran and Appleby, 2022). In general practice, AI might serve as a useful 'second opinion' when specialist radiologists are unavailable. AI may theoretically reduce misdiagnoses and missed diagnoses involving rarer conditions or unusual presentations. Improved workflows from speedier diagnosis could benefit patients and clients, while fostering job satisfaction.

However, AI's risks are equally important (Raymond Geis et al., 2019). ML models can have inaccuracies due to issues with their training data (Coghlan and Quinn, 2023b). For example, insufficient and unrepresentative data in model training can lead to false positive and negative classifications (Goisauf and Cano Abadía, 2022). So too can the application of ML models to patient groups that differ significantly from training datasets. Factors such as patient age, breed, and conformation, plus differences in scanning and labelling techniques (Raymond Geis et al., 2019), may be associated with model bias. If widely used, biased and inaccurate models could negatively affect very many veterinary patients. An essential consideration warrants attention at this point: AI algorithms are typically trained and validated using extensive databases in human medicine. On the contrary, datasets available in veterinary medicine are typically more limited. Consequently, it is the authors' opinion that further studies are imperative to attain a robust level of result generalizability.

Promoting patient wellbeing will only occur if practitioners avoid overtrusting or undertrusting AI models (Jacovi et al., 2021). Overtrust in AI may derive from 'automation bias' (Parasuraman and Riley, 1997) and underappreciating machine learning's limitations. Too little trust is problematic when AI is relatively trustworthy. Trust in AI can drop off, sometimes legitimately, with DL systems (such as CNNs) that are uninterpretable despite being generally accurate (Goisauf and Cano Abadía, 2022). Such blackbox AI (Quinn et al., 2022) stops practitioners knowing the reasons for, say, the classification of a radiological anomaly (e.g., in a thoracic X-ray). Other ethical issues include privacy of client data in radiological AI models (Raymond Geis et al., 2019) and carbon emissions from energy intensive AI (Crawford, 2021).

Given these issues, we offer the following ethical recommendations. Crucially, AI developers must appreciate veterinary medicine's goals of promoting patients' interests and human-animal relationships. AI models should therefore be trained and rigorously tested—not just in the lab but in real-life circumstances—to ensure reliability (Cohen and Gordon, 2022). Furthermore, radiological AI systems, in our view, should come with information about their training data and appropriate warnings about their proper use and risks. This would facilitate informed decision-making by veterinarians.

Equally, practitioners, radiologists, and clinic managers must take steps to appreciate the benefits and risks of AI tools used or purchased (Coghlan and Quinn, 2023b). This includes risks of overtrusting and undertrusting AI, turning to AI instead of radiology experts, employing 'blackbox' models (London, 2019), exposing or uploading private client information, and contributing to pollution. To mitigate risk, veterinarians and clinic managers should ask questions about particular models and insist that AI purveyors, not least profit-driven companies, are transparent about them (Larsson and Heintz, 2020). This is especially important given that veterinary technologies, unlike medical technologies, are relatively unregulated (Cohen and Gordon, 2022). The authors recommend the reader of this paper to refer to the official publication of the ACVR and ECVDI, where an AI-published supplement focused on AI is available.

Finally, we recommend that radiological AI never entirely replace human judgement. Removing a 'human-in-the-loop' (Goisauf and Cano Abadía, 2022) imperils patient wellbeing and abrogates veterinary responsibility. Practitioners and radiologists should remain accountable for patient care even when using sophisticated AI tools, since those tools can still get it wrong. We therefore advise that veterinarians treat radiological AI as a decision support tool, not a decision-making substitute. Veterinary professionals should employ only those AI tools which they know will support good medical practice (Reddy et al., 2019).

4. Conclusion

AI-based models, in particular deep learning models, could act as effective supports in the evaluation of medical images for both specialized radiologists and general practitioners. Nevertheless, these technologies should not replace veterinary experience and knowledge. On the contrary, AI products have the potential to empower radiologists to deliver increased value in a more efficient way.

Funding

The present paper is part of a project funded by a research grant from the Department of Animal Medicine, Production and Health – MAPS, University of Padua, Italy: SID- Banzato 2023 (€ 20000; Development of an algorithm for the automatic classification and identification of the lesions on the radiographs of the thorax in dogs).

Author contributions

TB conceived the study and drafted the manuscript; SB, SC, MW, MB and AZ drafted the manuscript.

CRediT authorship contribution statement

Silvia Burti: Writing – original draft. **Tommaso Banzato:** Writing –

original draft, Supervision, Conceptualization. **Simon Coghlan:** Writing – original draft. **Marek Wodzinski:** Writing – original draft. **Margherita Bendazzoli:** Writing – original draft. **Alessandro Zotti:** Writing – original draft.

Declaration of competing interest

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

References

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. Res. Vet. Sci. 103, 164–169. https://doi.org/10.1016/j.rvsc.2015.10.002.

Banzato, T., Fiore, E., Morgante, M., Manuali, E., Zotti, A., Manu-, E., Zotti, A., Manuali, E., Zotti, A., 2016. Texture analysis of B-mode ultrasound images to stage hepatic lipidosis in the dairy cow: a methodological study. Res. Vet. Sci. 108, 71–75. https://doi.org/10.1016/j.rvsc.2016.08.007.

Banzato, T., Bernardini, M., Cherubini, G.B., Zotti, A., 2017. Texture analysis of magnetic resonance images to predict histologic grade of meningiomas in dogs. Am. J. Vet. Res. 78, 1156–1162. https://doi.org/10.2460/ajvr.78.10.1156.

Banzato, Tommaso, Bernardini, M., Cherubini, G.B., Zotti, A., 2018a. A methodological approach for deep learning to distinguish between meningiomas and gliomas on canine MR-images. BMC Vet. Res. 14, 317. https://doi.org/10.1186/s12917-018-1638-2.

Banzato, T., Bonsembiante, F., Aresu, L., Gelain, M.E., Burti, S., Zotti, A., 2018b. Use of transfer learning to detect diffuse degenerative hepatic diseases from ultrasound images in dogs: a methodological study. Vet. J. 233 https://doi.org/10.1016/j.tvjl.2017.12.026.

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. Vet. J. 235, 90–92. https://doi.org/10.1016/j.tvjl.2018.04.001.

Banzato, T., Wodzinski, M., Burti, S., Osti, V.L., Rossoni, V., Atzori, M., Zotti, A., 2021a. Automatic classification of canine thoracic radiographs using deep learning. Sci. Rep. 11, 1–8. https://doi.org/10.1038/s41598-021-83515-3.

Banzato, T., Wodzinski, M., Tauceri, F., Donà, C., Scavazza, F., Müller, H., Zotti, A., 2021b. An AI-based algorithm for the automatic classification of thoracic radiographs in cats. Front. Vet. Sci. 8, 1–7. https://doi.org/10.3389/fvets.2021.731936.

Banzato, T., Wodzinski, M., Burti, S., Vettore, E., Muller, H., Zotti, A., 2023. An AI-based algorithm for the automatic evaluation of image quality in canine thoracic radiographs. Sci. Rep. 13 https://doi.org/10.1038/s41598-023-44089-4.

Barge, P., Oevermann, A., Maiolini, A., Durand, A., 2023. Machine learning predicts histologic type and grade of canine gliomas based on MRI texture analysis. Veterin. Radiol. Ultrasound 64, 724–732. https://doi.org/10.1111/vru.13242.

Basran, P.S., Appleby, R.B., 2022. The unmet potential of artificial intelligencin veterinary medicine. Am. J. Vet. Res. 83, 385–392. https://doi.org/10.2460/ajvr.22.03.0038.

Biercher, A., Meller, S., Wendt, J., Caspari, N., Schmidt-Mosig, J., De Decker, S., Volk, H. A., 2021. Using deep learning to detect spinal cord diseases on thoracolumbar magnetic resonance images of dogs. Front. Vet. Sci. 8, 1–9. https://doi.org/10.3389/fvets.2021.721167.

Boissady, E., de La Comble, A., Zhu, X., Hespel, A.M., 2020. Artificial intelligence evaluating primary thoracic lesions has an overall lower error rate compared to veterinarians or veterinarians in conjunction with the artificial intelligence. Veterin. Radiol. Ultrasound 61, 619–627. https://doi.org/10.1111/vru.12912.

Boissady, E., De La Comble, A., Zhu, X., Abbott, J., Adrien-Maxence, H., 2021. Comparison of a deep learning algorithm vs. humans for vertebral heart scale measurements in cats and dogs shows a high degree of agreement among readers. Front. Vet. Sci. 8, 1–6. https://doi.org/10.3389/fvets.2021.764570.

Burti, S., Longhin Osti, V., Zotti, A., Banzato, T., 2020. Use of deep learning to detect cardiomegaly on thoracic radiographs in dogs. Vet. J. 262, 105505 https://doi.org/10.1016/j.tvjl.2020.105505.

Burti, S., Zotti, A., Bonsembiante, F., Contiero, B., Banzato, T., 2021. Diagnostic accuracy of delayed phase post contrast computed tomographic images in the diagnosis of focal liver lesions in dogs: 69 cases. Front. Vet. Sci. 8, 1–10. https://doi.org/10.3389/fvets.2021.611556.

Burti, S., Zotti, A., Bonsembiante, F., Contiero, B., Banzato, T., 2022. A Machine Learning-Based Approach for Classification of Focal Splenic Lesions Based on their CT Features, 9, pp. 1–12. https://doi.org/10.3389/fvets.2022.872618.

Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., Elhadad, N., 2015. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, pp. 1721–1730. https://doi.org/10.1145/2783258.2788613.

Celniak, W., Wodziński, M., Jurgas, A., Burti, S., Zotti, A., Atzori, M., Müller, H., Banzato, T., 2023. Improving the classification of veterinary thoracic radiographs through inter-species and inter-pathology self-supervised pre-training of deep learning models. Sci. Rep. 13, 19518. https://doi.org/10.1038/s41598-023-46345-z.

Choi, H., Lee, S., Choi, Hojung, Lee, Y., Lee, K., 2023a. Deep learning-based reconstruction for canine brain magnetic resonance imaging could improve image quality while reducing scan time. Vet. Radiol. Ultrasound 64, 873–880. https://doi.org/10.1111/vru.13279.

Choi, B.K., Park, S., Lee, G., Chang, D., Jeon, S., Choi, J., 2023b. Can CT texture analysis parameters be used as imaging biomarkers for prediction of malignancy in canine splenic tumors? Veterin. Radiol. Ultrasound 64, 224–232. https://doi.org/10.1111/vru.13175.

Coghlan, S., 2018. Strong patient advocacy and the fundamental ethical role of veterinarians. J. Agric. Environ. Ethics 31, 349–367. https://doi.org/10.1007/s10806-018-9729-4.

Coghlan, S., Quinn, T., 2023a. Ethics of using artificial intelligence (AI) in veterinary medicine. AI & Soc. https://doi.org/10.1007/s00146-023-01686-1.

Coghlan, S., Quinn, T., 2023b. Ethics of using artificial intelligence (AI) in veterinary medicine. AI & Soc. https://doi.org/10.1007/s00146-023-01686-1.

Cohen, E.B., Gordon, I.K., 2022. First, do no harm. Ethical and legal issues of artificial intelligence and machine learning in veterinary radiology and radiation oncology. Veterin. Radiol. Ultrasound. https://doi.org/10.1111/vru.13171.

Cohen, J., Fischetti, A.J., Daverio, H., 2023. Veterinary radiologic error rate as determined by necropsy. Veterin. Radiol. Ultrasound 64, 573–584. https://doi.org/10.1111/vru.13259.

Crawford, K., 2021. The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press.

Degnan, A.J., Ghobadi, E.H., Hardy, P., Krupinski, E., Scali, E.P., Stratchko, L., Ulano, A., Walker, E., Wasnik, A.P., Auffermann, W.F., 2019. Perceptual and interpretive error in diagnostic radiology—causes and potential solutions. Acad. Radiol. 26, 833–845. https://doi.org/10.1016/j.acra.2018.11.006.

Dumortier, L., Guépin, F., Delignette-Muller, M.L., Boulocher, C., Grenier, T., 2022. Deep learning in veterinary medicine, an approach based on CNN to detect pulmonary abnormalities from lateral thoracic radiographs in cats. Sci. Rep. 12, 1–12. https://doi.org/10.1038/s41598-022-14993-2.

Ergun, G.B., Guney, S., 2021. Classification of canine maturity and bone fracture time based on X-ray images of long bones. IEEE Access 9, 109004–109011. https://doi.org/10.1109/ACCESS.2021.3101040.

Fitzke, M., Stack, C., Dourson, A., Santana, R.M.B., Wilson, D., Ziemer, L., Soin, A., Lungren, M.P., Fisher, P., Parkinson, M., 2021. RapidRead: Global Deployment of State-of-the-Art Radiology AI for a Large Veterinary Teleradiology Practice.

Goisauf, M., Cano Abadía, M., 2022. Ethics of AI in radiology: a review of ethical and societal implications. Front. Big Data. https://doi.org/10.3389/fdata.2022.850383.

Groendahl, A.R., Huynh, B.N., Tomic, O., Søvik, Å., Dale, E., Malinen, E., Skogmo, H.K., Futsaether, C.M., 2023. Automatic gross tumor segmentation of canine head and neck cancer using deep learning and cross-species transfer learning. Front. Vet. Sci. 10 https://doi.org/10.3389/fvets.2023.1143986.

Hardy, M., Harvey, H., 2020. Artificial Intelligence in Diagnostic Imaging: Impact on the Radiography Profession, p. 1.

Hernandez, E., Fawcett, A., Brouwer, E., Rau, J., Turner, P.V., 2018. Speaking up: veterinary ethical responsibilities and animalwelfare issues in everyday practice. Animals 8. https://doi.org/10.3390/ani8010015.

Hespel, A.-M., Boissady, E., De La Comble, A., Acierno, M., Alexander, K., Auger, M., Biller, D., de Swarte, M., Fuerst, J., Green, E., Hoey, S., Koernig, K., Lee, A., MacLellan, M., McAllister, H., Rechy Jr., J., Xiaojuan, Z., Zarelli, M., Morandi, F., 2022. Comparison of error rates between four pretrained DenseNet convolutional neural network models and 13 board-certified veterinary radiologists when evaluating 15 labels of canine thoracic radiographs. Veterin. Radiol. Ultrasound 63, 456–468. https://doi.org/10.1111/vru.13069.

Jacovi, A., Marasović, A., Miller, T., Goldberg, Y., 2021. Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. In: FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery, Inc, pp. 624–635. https://doi.org/10.1145/3442188.3445923.

Jeong, Y., Sung, J., 2022. An automated deep learning method and novel cardiac index to detect canine cardiomegaly from simple radiography. Sci. Rep. 12, 1–10. https://doi.org/10.1038/s41598-022-18822-4.

Ji, Y., Cho, H., Seon, S., Lee, K., Yoon, H., 2022. A deep learning model for CT-based kidney volume determination in dogs and normal reference definition. Front. Vet. Sci. https://doi.org/10.3389/fvets.2022.1011804.

Ji, Y., Hwang, G., Lee, S.J., Lee, K., Yoon, H., 2023. A deep learning model for automated kidney calculi detection on non-contrast computed tomography scans in dogs. Front. Vet. Sci. 10 https://doi.org/10.3389/fvets.2023.1236579.

Kim, E., Fischetti, A.J., Sreetharan, P., Weltman, J.G., Fox, P.R., 2022. Comparison of artificial intelligence to the veterinary radiologist's diagnosis of canine cardiogenic pulmonary edema. Veterin. Radiol. Ultrasound 63, 292–297. https://doi.org/10.1111/vru.13062.

Krupinski, E.A., Berbaum, K.S., Caldwell, R.T., Schartz, K.M., Madsen, M.T., Kramer, D. J., 2012. Do long radiology workdays affect nodule detection in dynamic CT interpretation? J. Am. Coll. Radiol. 9, 191–198. https://doi.org/10.1016/j.jacr.2011.11.013.

Lamb, C.R., Pfeiffer, D.U., Mantis, P., 2007. Errors in radiographic interpretation made by veterinary students. J. Vet. Med. Educ. 34, 157–159. https://doi.org/10.3138/jvme.34.2.157.
Larsson, S., Heintz, F., 2020. Transparency in artificial intelligence. Internet Pol. Rev. 9, 1–16. https://doi.org/10.14763/2020.2.1469.
Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.
Li, S., Wang, Z., Visser, L.C., Wisner, E.R., Cheng, H., 2020. Pilot study: application of artificial intelligence for detecting left atrial enlargement on canine thoracic radiographs. Veterin. Radiol. Ultrasound 61, 611–618. https://doi.org/10.1111/vru.12901.
London, A.J., 2019. Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hast. Cent. Rep. 49, 15–21. https://doi.org/10.1002/hast.973.
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 Vet. Res. 13 https://doi.org/10.1186/s12917-017-1117-1.
Mcevoy, F.J., Amigo, J.M., 2013. Using machine learning to classify image features from canine pelvic radiographs: evaluation of partial least squares discriminant analysis and artificial neural network models. Veterin. Radiol. Ultrasound 54, 122–126. https://doi.org/10.1111/vru.12003.
McEvoy, F.J., Proschowsky, H.F., Müller, A.V., Moorman, L., Bender-Koch, J., Svalastoga, E.L., Frellsen, J., Nielsen, D.H., 2021. Deep transfer learning can be used for the detection of hip joints in pelvis radiographs and the classification of their hip dysplasia status. Veterin. Radiol. Ultrasound 62, 387–393. https://doi.org/10.1111/vru.12968.
Morley, J., Machado, C.C.V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., Floridi, L., 2020. The ethics of AI in health care: a mapping review. Soc. Sci. Med. https://doi.org/10.1016/j.socscimed.2020.113172.
Müller, T.R., Solano, M., Tsunemi, M.H., 2022. Accuracy of artificial intelligence software for the detection of confirmed pleural effusion in thoracic radiographs in dogs. Veterin. Radiol. Ultrasound 63, 573–579. https://doi.org/10.1111/vru.13089.
Ott, J., Bruyette, D., Arbuckle, C., Balsz, D., Hecht, S., Shubitz, L., Baldi, P., 2021. Detecting pulmonary Coccidioidomycosis with deep convolutional neural networks. Machine Learn. Appl. 5, 100040 https://doi.org/10.1016/j.mlwa.2021.100040.
Parasuraman, R., Riley, V., 1997. Humans and automation: use, misuse, disuse, abuse. Human Fact. J. Human Fact. Ergonom. Soc. 39, 230–253. https://doi.org/10.1518/001872097778543886.
Park, J., Choi, B., Ko, J., Chun, J., Park, I., Lee, J., Kim, Jayon, Kim, Jaehwan, Eom, K., Kim, J.S., 2021. Deep-learning-based automatic segmentation of head and neck organs for radiation therapy in dogs. Front. Vet. Sci. 8 https://doi.org/10.3389/fvets.2021.721612.
Pey, P., Specchi, S., Rossi, F., Diana, A., Drudi, I., Zwingenberger, A.L., Mayhew, P.D., Pisoni, L., Mari, D., Massari, F., Dalpozzo, B., Fracassi, F., Nicoli, S., 2022. Prediction of vascular invasion using a 7-point scale computed tomography grading system in adrenal tumors in dogs. J. Vet. Intern. Med. 36, 713–725. https://doi.org/10.1111/jvim.16371.
Pomerantz, L.K., Solano, M., Kalosa-Kenyon, E., 2023. Performance of a commercially available artificial intelligence software for the detection of confirmed pulmonary nodules and masses in canine thoracic radiography. Veterin. Radiol. Ultrasound 64, 881–889. https://doi.org/10.1111/vru.13287.
Quinn, T.P., Jacobs, S., Senadeera, M., Le, V., Coghlan, S., 2022. The three ghosts of medical AI: can the black-box present deliver? Artif. Intell. Med. 124 https://doi.org/10.1016/j.artmed.2021.102158.
Raymond Geis, J., Brady, A.P., Wu, C.C., Spencer, J., Ranschaert, E., Jaremko, J.L., Langer, S.G., Kitts, A.B., Birch, J., Shields, W.F., van den Hoven Genderen, R., Kotter, E., Gichoya, J.W., Cook, T.S., Morgan, M.B., Tang, A., Safdar, N.M., Kohli, M., 2019. Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. Radiology 293, 436–440. https://doi.org/10.1148/radiol.2019191586.
Reddy, S., Fox, J., Purohit, M.P., 2019. Artificial intelligence-enabled healthcare delivery. J. R. Soc. Med. https://doi.org/10.1177/0141076818815510.
Rollin, B., 2006. An Introduction to Veterinary Medical Ethics — Theory and Cases, 2nd ed. Blackwell Publishing, Ames, Iowa, USA.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation.
Schmid, D., Scholz, V.B., Kircher, P.R., Lautenschlaeger, I.E., 2022. Employing deep convolutional neural networks for segmenting the medial retropharyngeal lymph nodes in CT studies of dogs. Veterin. Radiol. Ultrasound 63, 763–770. https://doi.org/10.1111/vru.13132.
Shaker, R., Wilke, C., Ober, C., Lawrence, J., 2021. Machine learning model development for quantitative analysis of CT heterogeneity in canine hepatic masses may predict histologic malignancy. Veterin. Radiol. Ultrasound 62, 711–719. https://doi.org/10.1111/vru.13012.
Shim, H., Lee, J., Choi, S., Kim, Jayon, Jeong, J., Cho, C., Kim, H., Kim, Jee In, Kim, Jaehwan, Eom, K., 2023. Deep learning-based diagnosis of stifle joint diseases in dogs. Veterin. Radiol. Ultrasound 64, 113–122. https://doi.org/10.1111/vru.13181.
Siewert, B., Sosna, J., McNamara, A., Raptopoulos, V., Kruskal, J.B., 2008. Quality initiatives: missed lesions at abdominal oncologic CT: lessons learned from quality assurance. Radiographics 28, 623–638. https://doi.org/10.1148/rg.283075188.
Spiteri, M., Knowler, S.P., Rusbridge, C., Wells, K., 2019. Using machine learning to understand neuromorphological change and image-based biomarker identification in cavalier king Charles spaniels with Chiari-like malformation-associated pain and syringomyelia. J. Vet. Intern. Med. 33, 2665–2674. https://doi.org/10.1111/jvim.15621.
Tahghighi, P., Norena, N., Ukwatta, E., Appleby, R.B., Komeili, A., 2023. Automatic classification of symmetry of hemithoraces in canine and feline radiographs. J. Med. Imag. (Bellingham) 10, 044004. https://doi.org/10.1117/1.JMI.10.4.044004.
Topol, E.J., 2019. High-performance medicine: the convergence of human and artificial intelligence. Nat. Med. https://doi.org/10.1038/s41591-018-0300-7.
Valente, C., Wodzinski, M., Guglielmini, C., Poser, H., Chiavegato, D., Zotti, A., Venturini, R., Banzato, T., 2023. Development of an artificial intelligence-based method for the diagnosis of the severity of myxomatous mitral valve disease from canine chest radiographs. Front. Vet. Sci. 10 https://doi.org/10.3389/fvets.2023.1227009.
Walsh, F., 2009. Human-animal bonds I: the relational significance of companion animals. Fam. Process 48, 462–480. https://doi.org/10.1111/j.1545-5300.2009.01296.x.
Wanamaker, M.W., Vernau, K.M., Taylor, S.L., Cissell, D.D., Abdelhafez, Y.G., Zwingenberger, A.L., 2021. Classification of neoplastic and inflammatory brain disease using MRI texture analysis in 119 dogs. Veterin. Radiol. Ultrasound 62, 445–454. https://doi.org/10.1111/vru.12962.
Yoon, Y., Hwang, T., Lee, H., 2018. Prediction of radiographic abnormalities by the use of bag-of-features and convolutional neural networks. Vet. J. 237, 43–48. https://doi.org/10.1016/j.tvjl.2018.05.009.
Yun, S.J., Kim, H.C., Yang, D.M., Kim, S.W., Rhee, S.J., Ahn, S.E., 2017a. Diagnostic errors when interpreting abdominopelvic computed tomography: a pictorial review. Br. J. Radiol. 90, 2–7. https://doi.org/10.1259/bjr.20160928.
Yun, S.J., Kim, H.C., Yang, D.M., Kim, S.W., Rhee, S.J., Ahn, S.E., 2017b. Diagnostic errors when interpreting abdominopelvic computed tomography: a pictorial review. Br. J. Radiol. 90, 2–7. https://doi.org/10.1259/bjr.20160928.
Yutong Chen, B., Schönlieb, C., Liò, P., Leiner, T., Luigi Dragotti, P., Ieee, F., Wang, G., Rueckert, D., Firmin, D., Yang, G., Member IEEE, S, 2022. AI-Based Reconstruction for Fast MRI-A Systematic Review and Meta-Analysis and also with the. Proceedings of the IEEE, 110. https://doi.org/10.1109/JPROC.2022.3141367.
Zhang, M., Zhang, K., Yu, D., Xie, Q., Liu, B., Chen, D., Xv, D., Li, Z., Liu, C., 2021. Computerized assisted evaluation system for canine cardiomegaly via key points detection with deep learning. Prev. Vet. Med. 193, 105399 https://doi.org/10.1016/j.prevetmed.2021.105399.