Scientific Article

Medicina VETERINÁRIA
e-ISSN 2675-6617
v. 19 n. 1 (2025)

Detection of left atrial enlargement in dogs using deep learning in thoracic radiographs

Detecção de aumento atrial esquerdo em cães usando deep learning em radiografias torácicas

Lucas Auto **Lopes**[1*], Anaemilia das Neves **Diniz**[1], Helio **Silva Neto**[2]

[1]*Department of Animal Science, Federal University of Alagoas (UFAL), Rio Largo-AL, Brazil.*
[2]*Department of Informatics Engineering, University of Coimbra, Coimbra, Centro, Portugal.*
**Corresponding author: lucas_auto@hotmail.com*

Article info

Keywords
Computer-aided diagnosis
Heart disease
Image classifier

DOI
10.26605/medvet-v19n1-6639

Citation
Lopes, L. A., Diniz, A. N., & Silva Neto, H. (2025) Detection of left atrial enlargement in dogs using deep learning in thoracic radiographs. *Medicina Veterinária*, *19*(1), 31-37. https://doi.org/10.26605/medvet-v19n1-6639

Received: January 23, 2024
Accepted: October 20, 2024

Abstract

The study aimed to develop a tool to assist veterinarians in diagnosing left atrial enlargement on chest X-rays in dogs. The model utilized a total of 652 images all used in training and testing divided into two categories "positive" and "negative". Three algorithms were used, obtaining the following results: the accuracy of the neural network was 89.7%, sensitivity of 90%, specificity of 89.5%, and Area Under the Curve (AUC) 95.8%. The accuracy of logistic regression was 88.2%, sensitivity 88.7%, specificity 87.8%, and AUC 94.1%. The decision tree accuracy was 69.6%, sensitivity 68.0%, specificity 71.0%, and AUC 69.6%. The classifier model with different algorithms can help radiologists improve the analysis of medical images by reducing errors, starting a selective double reading.

Resumo

Objetivou-se desenvolver uma ferramenta para auxiliar o médico veterinário no diagnóstico do aumento do átrio esquerdo em radiografias de tórax em cães. O modelo contém um total de 652 imagens, todas usadas no treino e no teste, divididas em duas categorias: "positiva" e "negativa". Foram usados três algoritmos obtendo os seguintes resultados: a acurácia da rede neural foi de 89,7%, sensibilidade de 90%, especificidade de 89,5% e área sob a curva (AUC), 95,8%. A acurácia da regressão logística foi de 88,2%, sensibilidade de 88,7%, especificidade de 87,3% e AUC 94,4%. A acurácia da árvore foi de 69,6%, sensibilidade de 68,0%, especificidade de 71,0% e AUC 69,6%. O modelo classificador com diferentes algoritmos poderá ajudar os radiologistas a melhorar na análise de imagens médicas com redução de erros, iniciando a leitura dupla seletiva.
Pavaras-chave: diagnóstico auxiliado por computador; cardiopatia; classificador de imagens.

1 | Introduction

Chest radiography is a routine diagnostic imaging test in small animals (Lamb and Nelson, 2015) because of its wide availability, affordability, and easy performance. The clinic conducts clinical investigations supported by interpretations from experienced professionals, and may direct, if necessary, to laboratory tests or other imaging tests (Liang et al., 2019).

Left atrial enlargement is a common cardiac alteration found in the routine of dog radiology and is usually associated with left atrial dilation (Bahr, 2015). However, pulmonary congestion and edema are common in dogs with left heart failure (Van Vleet and Ferrans, 1998). In chest radiography, this edema can be identified by the symmetric increase in radiopacity in the perihilar region (Kealy and McAllister, 2005; Abbott, 2006). However, in acute cases, the distribution of the liquid may be irregular (Abbott, 2006). If there is an increase in a single cardiac chamber, this condition will be for a short time because of the action of compensation mechanisms that promote the growth of other cardiac chambers. Therefore, veterinarians rarely observe an isolated increase in one cardiac chamber (Kealy and McAllister, 2005).

Perception is an important component of radiographic interpretation (Gunderman and Patel, 2019; Trall, 2015). Veterinarians perform this assessment empirically, but may be inaccurate as it depends on the veterinarian’s eyesight. An optical illusion may lead to an incorrect perception of inexperienced radiologists (Trall, 2015).

In 2018, Malcolm developed a tool, Vertebral Left Atrial Size (VLAS), to measure the left atrium size on radiographs. It is a quantitative method that uses pixels with a unit of measurement and compares the atrium with the vertebral body of the fourth thoracic vertebra, so that a moving structure can be measured anchored by a rigid structure.

Machine learning is a subfield of artificial intelligence that aims to enable computers to extract information, identify patterns, and generalize. This approach can solve classification problems, when the patterns that define classes may be difficult for humans to realize, or regression problems, by defining mathematical functions that allow for statistical analysis of the data (Alpaydin, 2014).

Digital radiography consists of a matrix of pixels with different shades of grey (White, 2004). Deep learning, a subfield of machine learning, uses algorithms to extract information directly from image pixels. The patterns identified in the extracted features are then used to create a prediction tool that can classify new images into specific categories (Webb, 2018).

This study proposes a classifier model that recognizes patterns in X-ray images using machine learning techniques to help veterinarians diagnose atrial enlargement in dogs.

2 | Materials and Method

2.1 | Dataset

For our retrospective study located in five cities in Brazil, we collected images from multiple institutions with different models of machines, operated by different technicians, and dogs of different breeds and ages. Diversity reduces bias in the model.

This retrospective study was performed with images provided by the Imavet, Focus, JR Diagnosticos, Vet-x and Uniradio imaging centres in Maceió (AL), Recife (PE), São Paulo (SP), Rio de Janeiro (RJ) and Lajes (SC). In the period between 2019 and 2021, a total of 326 right and left lateral thorax images were obtained. All images were oriented in the cranial (rostral) direction to the left. The dogs were of different breeds and ranged in age from 0.9 to 20 years. The radiographs were taken with different radiographic equipment in DICOM (Digital Imaging and Communications in Medicine) format. The annotations indicating the projections of the images were maintained.

2.2 | Formatting of the Dataset

We converted the images to 8-bit grayscale JPG format. On the 8-bit grayscale, the size of the image file ranged from 14.2 KB to 1.03 MB, and the size of the image matrix ranged from 242 184 to 3750 2572 pixels. The file size and image matrix changed according to the detector density of the radiographic detector plate used during image acquisition.

2.3 | Dataset Classification

We classified images as 'positive' (Figure 1) or 'negative' (Figure 2) for atrial enlargement using radiological criteria. This classification was based on left atrial enlargement and supported by the conclusions in the relevant radiological reports. We also used the VLAS method to split images. Thus, the total dataset for training the model was calculated using 150 positive and 176 negative images.

2.4 | Pre-processing the Dataset

For the training images of the classifier model, we used the following techniques for data preprocessing: data augmentation is a technique for generating new examples of training data (Shorten and Khoshgoftaar, 2019). We duplicated the images, resulting in 300 positive and 352 negative images.

We used manual resizing to define anatomical boundaries. This involves setting anatomical boundaries: the diaphragm, as the caudal boundary, the vertebrae, as the dorsal boundary, the external bones, as the ventral boundary, and the scapula as the cranial boundary.

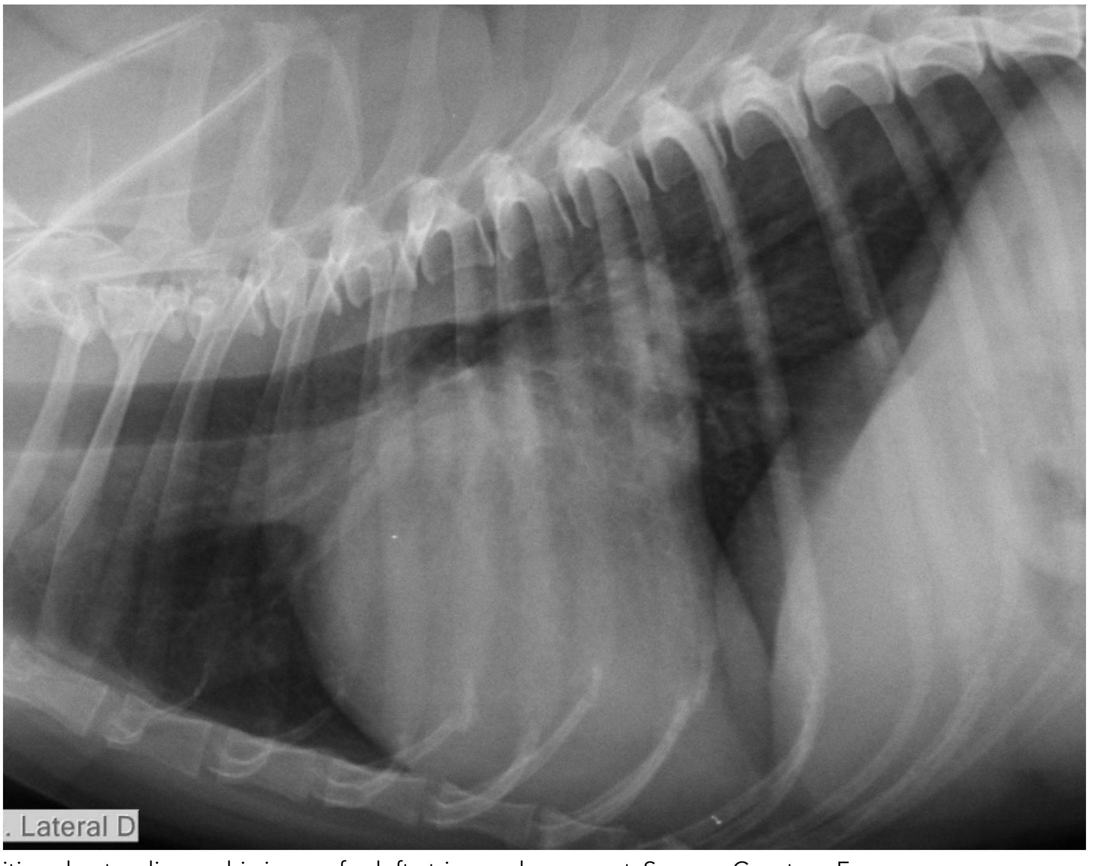

Figure 1. Positive chest radiographic image for left atrium enlargement. Source: Courtesy Focus.

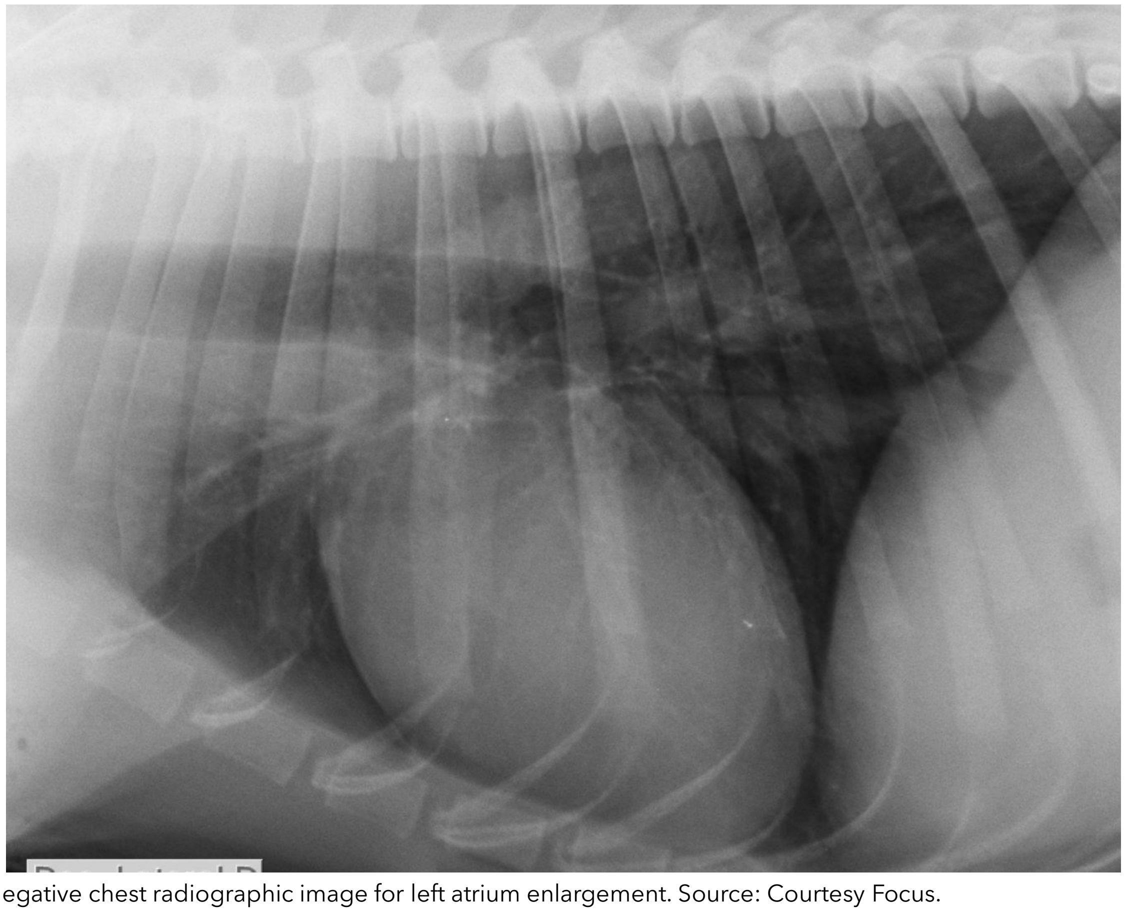

Figure 2. Negative chest radiographic image for left atrium enlargement. Source: Courtesy Focus.

2.5 | Embedding from the Dataset

To transform the unstructured images into structured (tabular) data, we used the feature extraction technique with the convolutional neural network Squeezenet, a technique that aims to filter the most relevant features for the classification step. The images are presented as input and we use the output of the penultimate layer of the convolutional neural network as a reference vector, generating a vector of 1000 elements (Alpaydin, 2014).

2.6 | Variable Selection and Processing

We used ANOVA to rank the variables from the embedding process. The top 900 variables were selected for further analysis.

As part of the processing of the selected variables, we normalized the data by transforming all variables to the same order of magnitude, where the pattern of variable values resulted in a mean of zero (0) and a standard deviation of one (1).

2.7 | Classifier Model Training

For the model training, we selected three machine learning algorithms: neural network, logistic regression, and decision tree.

We configured the neural network with 64 neurons, a rule for hyperbolic tangent activation function and Adam optimizer. For regularization, it was subjected to a penalty of $\alpha = 0.0001$ and performed a maximum of 200 interactions. The logistic regression algorithm had a weak lasso regularization configuration c=1000, and the decision tree with the following parameters: does not split subsets smaller than seven, and limits the maximum tree depth to 100 for sorting to stop when the majority reaches 90%.

2.8 | Model Evaluation Technique

To evaluate the performance of machine learning models, we used a stratified cross-validation technique. This technique consists of randomly dividing the data into ten subsets of similar size: nine were used for training, and one subset was used in the test. This process was repeated ten times, alternating the test subset at each cycle, and the model accuracy in each cycle was calculated. After ten cycles, we calculated the model's final accuracy based on errors. Stratification ensures an equal distribution of data, even if there is an imbalance between categories, both in training and testing.

2.9 | Tool Used

We used Orange Data Mining to allow image analysis using visual programming (Godec et al., 2019). We developed the classification model and performed statistical analysis using Orange Data Mining version 3.26, which comprises a group of open-source software developed in Python.

3 | Results

The final analytical dataset consisted of 652 images, 300 positives, and 352 negatives taken from dogs of different breeds and ages.

In the neural network model, with 652 images in the test dataset, the neural network correctly predicted 270 (90.0%) positive images; 30 (10.0%) positive images were incorrectly predicted as negative; 37 (10.5%) negative images were incorrectly predicted as positive and 315 (89.5%) negative images were correctly predicted as negative (Table 1). The area under the curve (AUC) was 95.8% (Figure 3), and the overall accuracy of the neural network model was 89.7%, with sensitivity of 90.0% and specificity of 89.8% (Table 2).

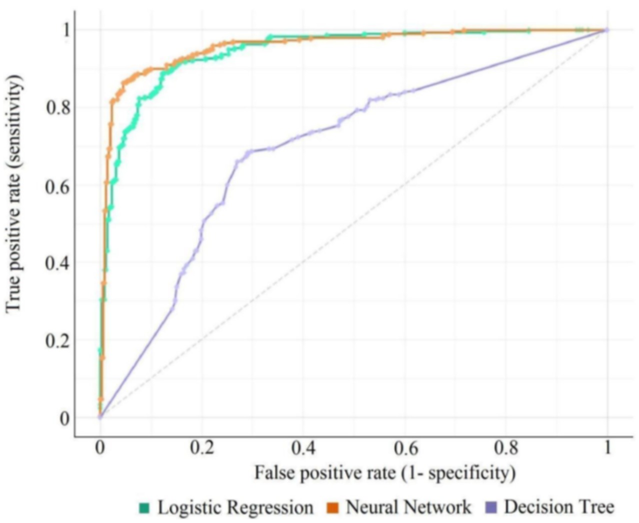

Figure 3. Comparison of the receiver operating characteristic curves of models Neural Network, Logistic Regression and Decision Tree.

Table 1. Prediction result of the accuracy Neural Network model

	Predict Positive	**Predict Negative**	**Total**
True Positive	270	30	300
True Negative	37	315	352
Total	307	345	652

Table 2. Comparison of models performance

Model	**AUC**	**Accuracy**	**Sensitivity**	**Specificity**
Neural Network	95.8%	89.7%	90.0%	89.8%
Logistic Regression	94.4%	88.2%	88.7%	87.8%
Decision Tree	69.6%	69.6%	68.0%	71.0%

In the logistic regression model, with 652 images in the test dataset, 266 (88.7%) positive images were correctly predicted as positive; 34 (11.3%) positive images were incorrectly predicted as negative; 43 (12.2%) negative images were incorrectly predicted as positive, and 309 (87.8%) negative images were correctly predicted as negative (Table 3). The AUC was 94.4% (Figure 3), and the overall accuracy of the logistic regression model was 88.2%, with a sensitivity of 88.7% and a specificity of 87.8% (Table 2).

In the decision tree model, with 652 images in the test dataset, 204 (68.0%) positive images were correctly predicted as positive; 96 (32.0%) positive images were incorrectly predicted as negative; 102 (29.0%) negative images were incorrectly predicted as positive, and 250 (71.0%) negative images were correctly predicted as negative (Table 4). The AUC was 69.6% (Figure 3), and the overall accuracy of the tree model was 69.6%, with a sensitivity of 68.0% and a specificity of 71.0% (Table 2).

4 | Discussion

Li et al. (2020), in a pilot study, developed a model to identify left atrial enlargement using artificial intelligence in a pilot study, with an accuracy of 79.01%; sensitivity of 73.68%, and specificity of 80.64%. The radiologists' performance for this task was 82.71%, sensitivity of 68.42%, and specificity of 87.09%. The results of our study surpassed those of the pilot study.

Diversity reduces bias in the model. Our retrospective study included images from five Brazilian cities, using various machines operated by distinct technicians, and featured dogs of diverse breeds and ages.

Li et al. (2020) removed the image annotations, but the present study chose to keep them. Therefore, it has approached the real working scenario of a veterinarian who will use the proposed solution. In this way, the images were maintained without editing, preserving the original characteristics used by the radiologists, which gave a high reliability to the evaluations.

It is important to say that we divided the image group into two categories for training and testing. These categories were extracted from the conclusive results of the radiographic reports and the VLAS. To define the classes, we adopted both subjective and objective criteria, as well as Thian et al. (2019), Burti et al. (2020), and Li et al. (2020).

The study showed the algorithms of machine learning neural network, decision tree and logistic regression can potentially learn and detect left atrial dilation pathology. The algorithms were able to extract and classify information from radiological images, as evidenced by AUC and accuracy results.

The three algorithms processed the image in different ways, and a comparison of the analyses minimized both false positives and false negatives. False positive diagnoses can harm animals by misleading clinical decisions (Lamb, 2016); therefore, error reduction is critical. This strategy has been successfully employed in other studies, as reported by Liang et al. (2019).

Among the three studied algorithms, the best performance concerning predictive metrics was the neural network algorithm (Table 2 and Figure 3) once the results are satisfactory for a classifier algorithm that presents a low risk of false positives and false negatives (Table 1). Logistic regression (Table 3) showed performance like a neural network. However, the decision tree presented a high number of false positives and false negatives (Table 4), making its prediction unreliable.

Table 3. Prediction result of the accuracy Logistic Regression model

	Predict Positive	**Predict Negative**	**Total**
True Positive	266	34	300
True Negative	43	309	352
Total	309	343	652

Table 4. Prediction result of the accuracy Decision Tree model

	Predict Positive	**Predict Negative**	**Total**
True Positive	204	96	300
True Negative	102	250	352
Total	306	346	652

The neural network and logistic regression models identified atrial dilation both in the right lateral projection and in the left lateral view. Burti et al. (2020) and Li et al. (2020) performed studies using only the right lateral projection.

The model was effective in identifying both positive and negative cases, and a correct diagnosis allows the initiation of more appropriate and efficient measures for the treatment of the patient. This model differs from the model proposed by Li et al. (2020), which showed greater efficiency in identifying negative cases.

The need for computational and infrastructure resources is proportional to the complexity of the algorithm (Du and Ko, 2011). Thus, the neural network requires greater processing power compared to the decision tree and logistic regression algorithm; however, it presents a lower error rate in predictions.

SqueezeNet (Ucar and Korkmaz, 2020), a lightweight pre-processing algorithm, enabled efficient image processing on low-resource systems. Unlike the studies by Derkatch et al. (2019) and Thian et al. (2019), who developed radiograph classifiers using processors with greater computational power.

Despite its efficiency, the model does not replace the assessment made by radiologists, as it is limited to identifying only the left atrium enlargement. Other findings present in the image are not identified. Therefore, this is complementary information that enriches and contributes to a more accurate diagnosis.

This study presents limitations. We did not train algorithms to analyze ventrodorsal or dorsoventral projections. There is the possibility of applying such an improvement solution for future studies. Future studies could incorporate additional projections to improve accuracy.

It is important to emphasize that radiographic diagnosis is not the primary method for detecting atrial enlargement, and we did not perform the gold-standard examination, the echocardiogram, to detect left atrial enlargement.

We did not evaluate the images external to the machine learning model. In future studies, we will observe the performance of the model using images of new patients.

Considering the data sampling, we need to increase the number of images compared to other studies, such as Rajpurkar et al. (2018), who used 112,120 images to train their classifier model.

Finally, the model did not consider the patient's history (age, sex, race, and clinical signs). We performed the analysis using only the information contained in the image. In future studies, it would be useful to enrich the model with information outside the images.

5 | Conclusion

This article demonstrates the implementation of a classifier model with different algorithms that can help radiologists to improve the analysis of medical images by reducing the error of selective double reading.

The model thus complements, but does not replace, the radiologist's assessment. As it is limited to identify a single change, an enlargement of the left atrium, it does not detect other findings that may be present in the image. This is additional information that enriches the diagnosis and makes it more accurate. However, researchers should carry out other studies, particularly in comparison with echocardiography to further evaluate the diagnostic capabilities of the model.

6 | Conflict of Interest Statement

The authors declare no conflict of interest.

7 | Acknowledgements

The authors would like to thank Imavet, Focus, JR Diagnósticos, Vet-x e Uniradio for the images used in this study.

8 | References

Abbott, J.A. Cardiomiopatia dilatada canina. In:_____. **Segredos em cardiologia de pequenos animais**. Porto Alegre: Artmed, 2006. p.304-314.

Alpaydin, E. Multilayer Perceptrons. In: **Introduction to machine learning**. 3rd ed. London: MIT Press, 2014. p.267-316.

Bahr, R.O Coração e os Vasos Pulmonares. In: Thrall, D.E. (Ed.). **Diagnóstico de radiologia veterinária**, 6ª ed. Rio de Janeiro: Elsevier, 2015. p1220-1261.

Burti, S. et al. Use of deep learning to detect cardiomegaly on thoracic radiographs in dogs. **The Veterinary Journal**, 262: 1-7, 2020.

Derkatch, S. et al. Identification of vertebral fractures by convolutional neural networks to predict nonvertebral and hip fractures: a registry-based cohort study of dual x-ray absorptiometry. **Radiology**, 293(2): 1-7, 2019.

Du, D.Z.; Ko, K.I. Models of computation and complexity classes. In:_____. **Theory of computational complexity**. 2nd ed. Hoboken: John Wiley & Sons, 2011. p.3-44.

Godec, P. et al. Democratized image analytics by visual programming through integration of deep models and small-scale machine learning. **Nature Communications,** 10 (1): 1-7, 2019.

Gunderman, R.B.; Patel, P. Perception's crucial role in radiology education. **Academic Radiology**, 26(1): 141–143. 2019.

Kealy, J.K.; McAllister, H. O tórax. In:_____. **Radiologia e ultrassonografia do cão e do gato**. 3ª ed. Barueri: Manole, 2005. p.149-252.

Lamb, C.R. Veterinary diagnostic imaging: Probability, accuracy and impact. **The Veterinary Journal**, 215: 55-63, 2016.

Lamb, C.R.; Nelson, J.R. Diagnostic accuracy of tests based on radiologic measurements of dogs and cats: a systematic review. **Veterinary Radiology & Ultrasound**, 56(3): 231-244, 2015.

Li, S. et al. Pilot study: Application of artificial intelligence for detecting left atrial enlargement on canine thoracic radiographs. **Veterinary Radiology & Ultrasound**, 61(6): 611-618, 2020.

Liang, C.H. et al. Identifying pulmonary nodules or masses on chest radiography using deep learning: external validation and strategies to improve clinical practice. **Clinical Radiology**, 75(1): 38-45, 2019.

Malcolm, E.L. et al. Diagnostic value of vertebral left atrial size as determined from thoracic radiographs for assessment of left atrial size in dogs with myxomatous mitral valve disease. **Journal of the American Veterinary Medical Association**, 253(8): 1038-1045, 2018.

Rajpurkar, P. et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. **PLoS Med**, 15(11): 1-17, 2018.

Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. **Journal of Big Data**, 60(1): 1-48, 2019.

Thian, Y.L. et al. Convolutional Neural Networks for Automated Fracture Detection and Localization on Wrist Radiographs. **Radiology: Artificial Intelligence**, 1(1): 1-8, 2019.

Thrall, D.E. Introdução à Interpretação Radiográfica. In:_____. **Diagnóstico de radiologia veterinária**. 6ª ed. Rio de Janeiro: Elsevier, 2015. p.175-209.

Ucar, F.; Korkmaz, D. COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. **Medical Hypotheses**, 140: 1-12, 2020.

Van Vleet, J.F.; Ferrans, V.J. Patologia do sistema cardiovascular. In: Carlton, W.W., McGavian, M.D., (Ed.). **Patologia veterinária especial**. 2ª ed. Rio Grande do Sul: Artmed, 1998. p.194-227.

Webb, S. Technology feature deep learning for biology. **Nature**, 554(7693): 555-557, 2018.

White, S.C. Cysts of the jaws. In: White, S.C.; Pharoah, M.J. **Oral radiology: principles and interpretation**. 5th ed. St. Louis: Mosby, 2004. p.355-377.