ARTICLES
Archives of Veterinary Science
UFPR
https://www.ufpr.br/

Effectiveness of two neural networks (Resnet50v2 and Inception V3) in identifying left atrial enlargement in adult dogs

Lorena Tavares de Brito Nery Jaworski[1], Guilherme de Moraes Restani[2], Lucas Ferrari de Oliveira[3], Tilde Rodrigues Froes[4]*

Submitted: *10/04/2024*
Accepted: *23/06/2024*

[1]*Department of Veterinary Medicine, Federal University of Paraná. R: dos Funcionários, 1540. CEP 80.035-050. Curitiba, PR, Brazil. ORCID: https://orcid.org/0000-0002-6582-5797*
[2] *Department of Informatics, Federal University of Paraná. R: Evaristo F. Ferreira da Costa, 383-391. CEP 81.530-015.Curitiba, PR, Brazil. ORCID: https://orcid.org/0009-0008-8723-717X*
[3]*Department of Informatics, Federal University of Paraná. R: Evaristo F. Ferreira da Costa, 383-391. CEP 81.530-015.Curitiba, PR, Brazil. ORCID: https://orcid.org/0000-0002-8198-0877*
[4]*Department of Veterinary Medicine, Federal University of Paraná. R: dos Funcionários, 1540. CEP 80.035-050. Curitiba, PR, Brazil. ORCID: https://orcid.org/0000-0002-7802-2598*

*Correspondence email: lorenajaworski@gmail.com.

Abstract: Mitral valve endocardiosis (MVE) accounts for approximately 75% of cases of canine heart disease. Currently, the gold standard for diagnosis of MVE is echocardiography. However, radiography provides valuable information for rapid clinical staging of the disease, such as detection of left atrial enlargement, the presence or absence of pulmonary congestion, and cardiogenic pulmonary edema. This study evaluated the effectiveness of two neural networks, Resnet50v2 and Inception V3, in detecting left atrial enlargement. A database was built from 902 lateral thoracic canine radiographic images, separated into three groups: 198 images with left atrial enlargement (LAA) alone, 124 images with left atrial enlargement and concurrent pulmonary edema (LAA_PE), totaling 322 images with left atrial enlargement, and 580 images with a normal left atrium (N). After creation of the dataset, the images were divided into training, validation, and test sets. The algorithms were then trained, and their performance evaluated using appropriate metrics (accuracy, precision, sensitivity, specificity, F-score, ROC, area under the curve - AUC). Respective accuracies were: Resnet50V2, 94.77%, and the Inception V3, 92.84%. In conclusion, artificial neural networks can be used to classify left atrial enlargement from radiographs in dogs to assist veterinary radiologists in screening patients. The study limitations were the lack of labeled data availability and extensive anatomical and size variations in animals (i.e., breed and age-related). The next steps include segmenting the regions of interest in the radiographs so that classification can be improved and also to conduct studies to assess how the models could perform in real-world situations.
Keywords: Artificial intelligence – AI, Convolutional Neural Network – CNN, heart, radiology, enlarged left atrium.

1. Introduction

Myxomatous mitral valve disease (MMVD) is the most common acquired heart disease and the most frequent cause of congestive heart failure in dogs (Bogarelli et al., 2012; Lam et al., 2021), accounting for approximately 75% of cardiac conditions in veterinary practice (Kenne et al., 2019). It is a progressive disease that initially causes degenerative lesions in the mitral valve (Li et al., 2020), preventing it from functioning correctly and resulting in mitral regurgitation (Boswood et al., 2016). Mitral valve regurgitation imposes a chronic volume overload that results in a detectable enlargement of the left atrium, which may progress to signs of congestive heart failure, i.e., pulmonary venous congestion and cardiogenic pulmonary edema (Boswood et al., 2016). Confirmation of MMVD is usually by echocardiographic examination. However, although allowing early identification of disease pathology eg mitral valve thickening and left atrial enlargement, echocardiography requires specialist equipment, trained personnel for accurate assessment, and the use of specific software for analysis of parameters (Boswood et al., 2016; Li et al., 2020).

Thoracic radiography remains an important screening test in veterinary practice. It is rapid, non-invasive, and widely available and can be used to identify left atrial enlargement and cardiogenic pulmonary edema. This method is especially effective when combined with the patient's medical history and clinical presentation (Kim et al., 2022). Accurate and effective diagnosis is essential as delays in diagnosis and management increase morbidity and mortality (LI et al., 2020; Kim et al., 2022). Veterinary radiology is an ideal target for the application of artificial intelligence (AI) which has the potential to provide a quick and accurate diagnosis. This rapid diagnosis is important in screening the large number of exams carried out daily which is complicated by the high degree of complexity of some diseases (Garcia and Maciel, 2020).

Efforts have been made to develop neural networks that mimic the information processing of a human brain (Chassagnon et al., 2020) and can identify cardiomegaly (Burti et al., 2020; Boissady et al., 2021; Solomon et al., 2023), left atrial enlargement (Li et al., 2020) and even grade mitral myxomatous disease based on the American College of Veterinary Internal Medicine, AVCVIM guidelines (Valente et al., 2023). However, further studies of the data processing are required to validate the technique. One of the most popular types of deep artificial neural networks (ANN) is convolutional neural networks (CNN). CNN is able to automatically extract features from input data using 2D convolutional layers, eliminating manual activity (LeCun et al., 2015). ANN are adaptive machines that can be implemented as electronic components or software modeled on brain behavior. Like the brain, knowledge is acquired from the environment, a process called learning, and is stored by weighting the connections between neurons (Haykin et al., 2008). In this context, Resnet50v2 and InceptionV3 have demonstrated satisfactory results for detecting cardiomegaly (Banzato

et al., 2021) in cats. However, no studies have evaluated the effectiveness of these two protocols in detecting the left atrium in dogs. This study aimed to evaluate the effectiveness of Resnet50v2 and Inception V3 in detecting left atrial enlargement in skeletally mature dogs.

2. Materials e Methods

2.1. Creating the database

A retrospective study was carried out using thoracic radiographic images from the image bank of the Veterinary Hospital of the Federal University of Paraná (HV-UFPR) combined with a selection of thoracic radiographic images of dogs selected by Brazilian radiologists from another five veterinary hospitals. The inclusion criteria were well-positioned images of good diagnostic quality in which it was possible to delineate the entire cardiac silhouette.

Images (n=2,974) were selected from 1,487 patients. These images were then screened by two experienced radiologists, one with five years of practice in radiographic reading and the other with over 20 years of training and certified by the Brazilian College of Veterinary Radiology. From the total, 2,072 images were excluded because they did not meet the inclusion criteria for the study. A final 902 images were included: 882 thoracic images from the HV-UFPR and 20 from other radiology services. The images were separated into three groups: 198 images with left atrial enlargement (LAA), 124 images with left atrial enlargement and concurrent pulmonary edema (LAA_PE), totaling 322 images with left atrial enlargement, and 580 images considered normal (N) (Figure 1).

Figure 1 – Right-lateral thoracic radiographic images. (A): normal left atrium, (B): enlarged left atrium and (C): enlarged left atrium and pulmonary edema. The black arrows show the enlargement of the left atrium, while the white arrow shows the caudal lung lobe indicating radiographic signs compatible with pulmonary edema.

Only the right or left lateral projections were assessed. Left atrial enlargement was defined using the Vertebral Left Atrial Size (VLAS) method where > 2.3v was considered enlarged and, the Vertebral Heart Size (VHS) of the patients was measured, with VHS >10.5 v being considered abnormal, depending on the breed (Buchanan and Bucheler, 1995). Atrial enlargement was confirmed by echocardiographic examination in all included cases.

2.2. Experimental data

2.2.1. Characteristics of the images and equipment

The images had resolutions of between 4248 × 3480 pixels and 2328 × 1728 pixels, with 72 x 72 ppi and 8-bit grayscale depth. Originally in DICOM format, these images were converted to PNG and resized to a width of 640 pixels while maintaining the original aspect ratio. The study was conducted on the Google Colab platform, on which a computer with a 2.30GHz Intel(R) Xeon(R) CPU, 13GB RAM, and an NVIDIA Tesla T4 GPU was allocated. The operating system was the Linux distribution Ubuntu 22.04.2 LTS, with the CUDA library version 11.8.

The CNN used, Resnet50v2 and InceptionV3, were implemented using the Python programming language, version 3.10.12, and the tensor flow and Keras libraries, version 2.14.0. The metrics were generated using the scikit-learn libraries, version 1.2.2. To perform the binary classification, the ResNet50V2 and InceptionV3 models had their classification modules removed, and the following layers were added in their places: GlobalAveragePooling2D, a Dense layer of 1024 neurons with a 'relu' activation function, and a Dense layer of one neuron with a 'sigmoid' activation function (Figure 2). The SGD (Stochastic gradient descent) algorithm was used as the optimizer, and the binary cross entropy was chosen as the objective function. The learning rate was 0.01. The input dimensions of the networks were 416 x 416, with a batch size of 4. Each training session was conducted for 50 epochs, each consisting of a complete iteration through all the training data.

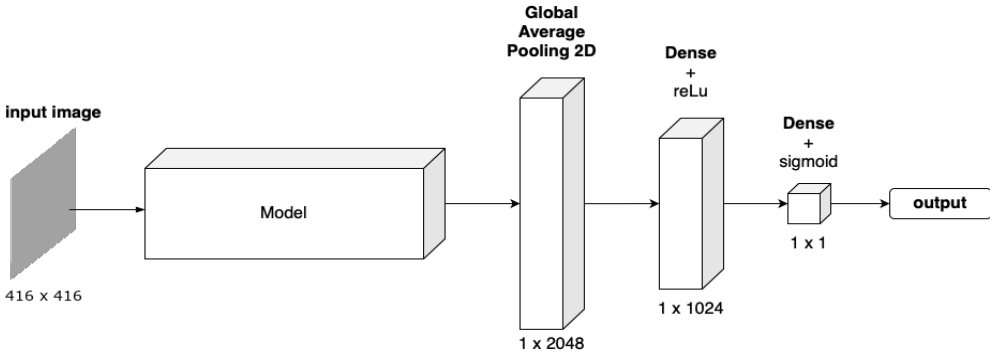

Figure 2 – Diagram of the models implemented. The input image is resized to 416 x 416 pixels, then fed into the chosen model (ResNet50V2 or InceptionV3) for extraction of features. The result of this stage is used as input by the new structure added. The new layers are presented in the image with their respective names and activation functions on the top and their shapes on the bottom.

CNN models often require thousands of image samples to maximise accuracy, so we used the data augmentation technique to increase the number of samples. The data augmentation functions were carried out following a methodology previously established by Kam et al. (2018), using the following parameters: (i) rotation of ± 20 degrees; (ii) zoom in (+) and out (-) by 5%; (iii) vertical displacement of ± 20%; and (iv) horizontal displacement of ± 20%; and adding Gaussian noise.

Using the cross-validation technique (k-fold), the images were manually, but randomly, divided into five-fold partitions (Table 1). Five training sessions were carried out for each network without data augmentation and five training sessions with data augmentation. At the end of each training session, the evaluation metrics and their respective validation sets were generated. Thus, four folds were used for each training session, and one was used as a test set. Finally, the average of the metrics obtained in each training session was used for the model's final metrics.

Fold[a]	AAE[b]	AAE_PE[c]	AAE + AAE_PE[d]	N[e]
1	38	26	64	116
2	40	24	64	116
3	40	24	64	116
4	40	24	64	116
5	40	26	66	116
Total	198	124	322	580

Table 1 – Allocation of the number of images used to train the networks in five different groups. Fold[a]: groups of images used to train the networks. AAE[b]: group of images with enlarged left atrium. AAE_PE[c]: group of images that had an enlarged left atrium at the same time as pulmonary edema. AAE + AAE_PE[d]: total number of images used in the AAE groups together with AAE_PE. N[e]: patients with images considered normal.

The learning of the networks was assessed using a set of metrics capable of summarizing the performance of a model (Hicks et al., 2022). Initially, a (i) confusion matrix was constructed, in which the values obtained by the network were compared with those obtained by the evaluators, obtaining values referring to data classified as false positive, true positive, false negative, and true negative. From these values, it was possible to assess (ii) accuracy, (iii) precision, (iv) sensitivity, (v) specificity, (vi) F-score, (vii) receiver operating characteristic (ROC) curve, and the area under the curve (AUC). These are all critical performance metrics used to evaluate the effectiveness of classification models in machine learning.

	Fold 01	Fold 02	Fold 03	Fold 04	Fold 05	Mean	STD Dev
Accuracy	0.8490	0.8050	0.8246	0.8981	0.8443	0.8442	0.0312
Precision	0.8394	0.8025	0.8011	0.8994	0.8526	0.8390	0.0363
Sensitivity	0.9426	0.9403	0.9658	0.9597	0.9301	0.9477	0.0131
Specificity	0.6857	0.5294	0.5732	0.8209	0.6667	0.6552	0.1011
F-score	0.8880	0.8660	0.8758	0.9286	0.8896	0.8896	0.0213
AUC	0.8981	0.7823	0.8572	0.9025	0.9172	0.8715	0.0488

Table 2 – Metrics obtained from training the ResNet50V2 network with no data augmentation for detecting left atrial enlargement in dogs. Description of the accuracy, precision, sensitivity, specificity, F-score and AUC values. The columns labeled "Fold" from 01 to 05 refer to the folds removed for validation in that training. Finally, the means of the results obtained in each training session and their respective standard deviations are shown. The best results are highlighted in bold. * STD Dev: Standard deviation.

http://dx.doi.org/10.5380/avs.v29i3.95142

3. Results

Four hundred and fifty-one adult dogs met the pre-established criteria to be included in the study, totaling 902 lateral thoracic radiographs. Two image acquisition systems were used: computed radiography (CR) and digital radiography (DR). A total of 322 images showed left atrial enlargement, with 198 images showing left atrial enlargement, 124 images showing atrial enlargement with concurrent signs of pulmonary edema, and 580 radiographic images were considered normal by the evaluators.

The ResNet50V2 neural network was initially tested in two ways, first without data augmentation and then using data augmentation. Tables 2 and 3 show the metrics and their respective values. According to the data shown, the best performance was obtained in "Fold 4", where the area under the curve (AUC) showed a good performance with an AUC of 0.90, accuracy of 0.8981, precision of 0.8994, sensitivity of 0.9597 and specificity of 0.8209.

	Fold 01	Fold 02	Fold 03	Fold 04	Fold 05	Mean	STD Dev
Accuracy	0.8385	0.7900	0.8640	0.8843	0.8821	0.8518	0.0350
Precision	0.8421	0.7744	0.8808	0.8563	0.9103	0.8528	0.0455
Sensitivity	0.9180	0.9621	0.9110	1.0000	0.9167	0.9416	0.0345
Specificity	0.7000	0.4500	0.7805	0.6269	0.8116	0.6738	0.1291
F-score	0.8784	0.8581	0.8956	0.9226	0.9135	0.8937	0.0234
AUC	0.8799	0.7980	0.9051	0.8700	0.9294	0.8765	0.0443

Table 3 – Metrics obtained from training the ResNet50V2 network using data augmentation for detecting left atrial enlargement in dogs. Description of the accuracy, precision, sensitivity, specificity, F-score and AUC values. The columns labeled "Fold" from 01 to 05 refer to the folds removed for validation in that training. Finally, the means of the results obtained in each training session and their respective standard deviations are shown. The best results are highlighted in bold. *STD Dev: Standard deviation.

The AUC is essential for evaluating machine learning models applied to classification tasks, especially in medical contexts. AUC values closer to 1 indicate the model's greater ability to correctly distinguish between classes (in this case, images without left atrial enlargement vs. images with left atrial enlargement). Using the data augmentation technique, similar data was observed, with the best performance also being obtained in "Fold 4", where an AUC of 0.87, accuracy of 0.8843, precision of 0.8563, sensitivity of 1.0, and specificity of 0.9226 were obtained. In addition, ROC curves were also generated for each training session, which can be seen in Figures 3a and 3b, respectively.

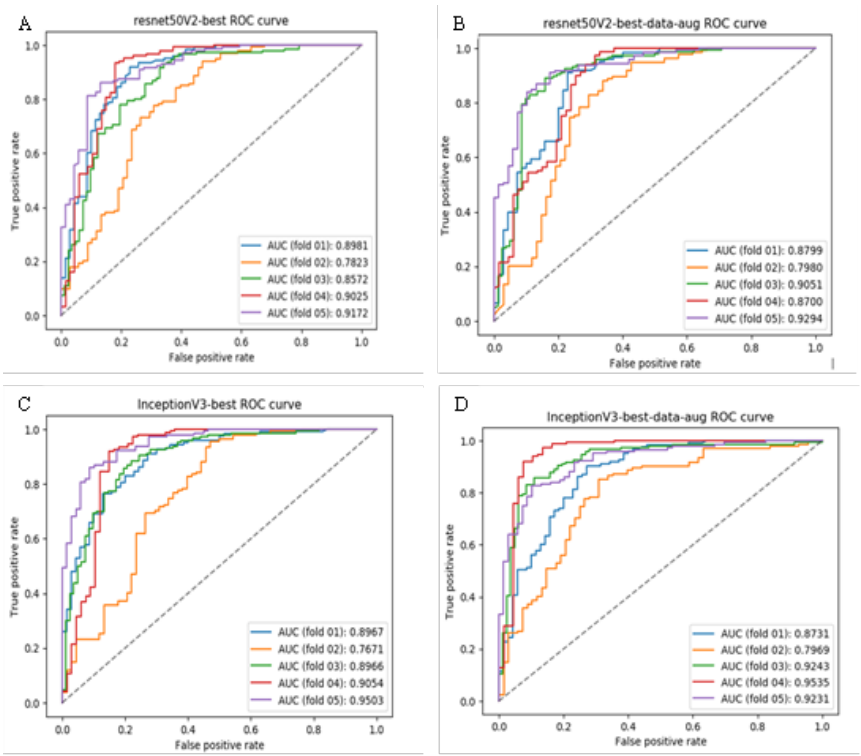

Figure 3 – A) ROC curves and AUC obtained for each training session of the ResNet50V2 network. In the image, each color represents a training session, while the dashed line represents a random classifier. B) ROC curves and AUC obtained for each training of the ResNet50V2 network with data augmentation. C) ROC curves and AUC obtained for each training session of the InceptionV3 network. D) ROC curves and AUC obtained for each training of the Inception V3 network, using data augmentation.

Next, the InceptionV3 neural network, was tested in two ways, similar to testing of ResNet50V2; first without data augmentation and then using data augmentation. Tables 4 and 5 show the evaluation of better performance in "Fold 4". The data obtained without the data augmentation technique had an AUC of 0.9054, an accuracy of 0.9074, a precision of 0.9057, a sensitivity of 0.9664, and a specificity of 0.7761. Superior results were observed using the data augmentation technique. However, the best performance was also obtained in "Fold 4", with an AUC of 0.9535 and specificity of 0.8507. Furthermore, ROC curves were also generated for each training session, which can be seen in Figures 3c and 3d.

Table 6 was generated to facilitate comparison of the results obtained in the previous steps. It shows the means of the metrics obtained for the ResNet50V2 and InceptionV3 networks, with and without data augmentation (DA). The results showed a slight improvement in most metrics obtained when the data augmentation technique was used (Table 6). In this case, the mean AUC obtained was 0.8942 for InceptionV3 and 0.8765 for ResNetV2.

	Fold 01	Fold 02	Fold 03	Fold 04	Fold 05	Mean	STD Dev
Accuracy	0.8333	0.8100	0.8421	0.9074	0.8821	0.8550	0.0350
Precision	0.8583	0.7975	0.8767	0.9057	0.8782	0.8633	0.0362
Sensitivity	0.8862	0.9545	0.8767	0.9664	0.9580	0.9284	0.0386
Specificity	0.7429	0.5294	0.7805	0.7761	0.7246	0.7107	0.0930
F-score	0.8720	0.8690	0.8767	0.9351	0.9164	0.8938	0.0268
AUC	0.8967	0.7671	0.8966	0.9054	0.9503	0.8832	0.0614

Table 4 – Metrics obtained from training the InceptionV3 network without data augmentation. Description of the accuracy, precision, sensitivity, specificity, F-score and AUC values. The columns labeled "Fold" from 01 to 05 refer to the folds removed for validation in that training. Finally, the means of the results obtained in each training session and their respective standard deviations are shown. The best results are highlighted in bold. * STD Dev: Standard deviation.

	Fold 01	Fold 02	Fold 03	Fold 04	Fold 05	Mean	STD Dev
Accuracy	0.8333	0.7950	0.8684	0.9352	0.8679	0.8600	0.0463
Precision	0.8462	0.8321	0.8973	0.9355	0.8859	0.8794	0.0370
Sensitivity	0.9016	0.8636	0.8973	0.9732	0.9231	0.9118	0.0361
Specificity	0.7143	0.6618	0.8171	0.8507	0.7536	0.7595	0.0682
F-score	0.8730	0.8476	0.8973	0.9539	0.9041	0.8952	0.0355
AUC	0.8731	0.7969	0.9243	0.9535	0.9231	0.8942	0.0551

Table 5 – Metrics obtained from training the InceptionV3 network with increased data. Description of the accuracy, precision, sensitivity, specificity, F-score and AUC values. The columns labeled "Fold" from 01 to 05 refer to the folds removed for validation in that training. Finally, the means of the results obtained in each training session and their respective standard deviations are shown. The best results are highlighted in bold. * STD Dev: Standard deviation.

4. Discussion

To implement the deep learning algorithms, two ANN architectures were chosen, InceptionV3 and ResNet50V2, which have both been used in both human medicine (Banzato et al., 2019) and veterinary medicine (Banzato et al., 2018a; Banzato et al., 2018b; Banzato et al., 2021). Based on the results, the InceptionV3 and ResNet50V2 performed better when trained using data augmentation. This suggests that this technique, by artificially expanding the sample size in the training set, allowed the model to generalize better, reducing the false positive and false negative rates. In addition, ROC curves were generated for each training set, which allowed us to visualize the model's ability to distinguish between classes graphically. The results show that, as with ResNet50V2, InceptionV3 also showed diagnostic improvement when trained with more data, as indicated by the higher metrics and ROC curves. This reinforces the benefits of using this technique for neural networks in medical imaging. During the training sessions, it was observed that, in most cases, the metric curves saturated in the first half of the training sessions, even with data augmentation.

For this reason, we chose the parameters from the session for each model trained that achieved the best validation accuracy. Also, the data augmentation technique improved every metric except for sensitivity, which slightly worsened. Given that sensitivity measures how well the model can detect positive cases, this metric is affected by the number of false negatives. The models became somewhat more prone to false negative results with data augmentation. However, the results recorded for sensitivity were still high, more than 0.9 in all cases, with the best result (0.9477) achieved by the ResNet50V2 model. Data augmentation is a technique used to artificially expand the size of the training dataset for a neural network. This is achieved by applying various transformations to the original data, such as rotations, flips, scaling, and other modifications (Shorten et al., 2022). Data augmentation is responsible for improving the performance of ANN network models because it enables the models to better generalize the problem by creating a new set of images derived from the original ones, which are then incorporated into the original database (Khan et al., 2018).

	ResNet50V2	ResNet50V2 (DA)	InceptionV3	InceptionV3 (DA)
Accuracy	0.8442	0.8518	0.8550	0.8600
Precision	0.8390	0.8528	0.8633	0.8794
Sensitivity	0.9477	0.9416	0.9284	0.9118
Specificity	0.6552	0.6738	0.7107	0.7595
F-score	0.8896	0.8937	0.8938	0.8952
AUC	0.8715	0.8765	0.8832	0.8942

Table 6 – Comparison of the means for each metric obtained for the ResNet50V2 and InceptionV3 networks, with and without data augmentation (DA). DA* With data increase. STD Dev: Standard deviation.

Although ResNet50V2 performed slightly less well than InceptionV3, both had an AUC > 0.8, similar to studies previously described in the literature (Banzato et al., 2021). Furthermore, this data confirms that ANNs can effectively detect left atrial enlargement, serving as a screening tool for cases requiring immediate report release and facilitating faster and more effective clinical management (Li et al., 2020). The rapid identification of left atrial enlargement by radiographic examination in dogs is essential for the diagnosis, staging, and consequent early treatment of heart diseases such as mitral valve insufficiency (Malcom et al., 2018; Li et al., 2020; Lam et al., 2021). Rapid radiographic screening allows for early implementation of complementary tests and mitral valve disease staging in clinics where echocardiography is not available (Ross et al., 2023). However, experience is required for a radiologist's to make a visual assessment of this parameter without diagnostic errors (Ross et al., 2023). Machine learning models have emerged as an auxiliary tool in this process, helping to automatically detect alterations in atrial size (Li et al., 2020). However, the responsible integration of AI into veterinary care requires veterinarians to expand their knowledge and skills beyond the traditional clinical domain (Joslyn and Alexander, 2022).

Based on the results of this study and a review of the literature, we confirmed the viability of using ANN to automatically classify left atrial enlargement in dog radiographs (Li et al., 2020). However, the high accuracy in this study may be related to the standardization of radiographic images. Studies with poorly positioned radiographic images with a high degree of rotation, inadequate techniques and the presence of thoracic diseases that made it impossible to analyze the cardiac silhouette, specifically the analysis of the dorso-caudal cardiac border and the topography of the left atrium like pleural or pericardial effusion, intra-thoracic or extra-thoracic masses, and congenital heart disease were excluded. This methodology was reported in previous studies (Banzato et al., 2021; Jeong and Sung, 2022), and does not necessarily represent the use of these networks in routine clinics.. Studies with poorly positioned radiographic images and inadequate techniques are therefore necessary to assess the degree of accuracy of ANN in a routine clinical situation.

Although the effectiveness of the ResNet50V2 and InceptionV3 networks in detecting radiographic images with and without left atrial enlargement has been validated, this study has some limitations. The sample size was relatively smal, although 1052 images were analyzed these were only from 451 patients. In addition, the images were collected from various hospitals without standardized acquisition protocols, and these variations in radiographic settings could impact the model's performance.

Furthermore, the data set of the present study only contained images of normal vs. enlarged left atrium, with or without cardiogenic edema. The absence of other common concomitant heart or lung diseases in the training data may compromise the model's performance when incorporated into clinical practice. This is because these rare lesions are likely to be poorly represented in the training data, and consequently, the algorithm's performance will be limited in their detection and classification (Joslyn and Alexander., 2022).

Another significant limitation is the lack of standardization in the image data regarding the breed, sex, age, and weight of the dogs. This heterogeneity arose from the fact that the images were acquired from diverse sources, which often made it impossible to collect these specific details. This variability in the input data can pose a substantial challenge for training an effective AI model. If the model is trained on a dataset that does not represent the full diversity seen in the real world, ie if it is skewed towards certain breeds or age groups, the AI system may develop biases and perform poorly on underrepresented subgroups (Norori et al., 2021).

5. Conclusion

We conclude that the Inception V3 and ResNet50V2 models successfully detected left atrial enlargement in a curated set of thoracic radiographs. The models can serve as an auxiliary tool, assisting in the automated identification of atrial abnormalities in dogs. However, despite the promising preliminary results, further prospective studies with larger datasets are still required. These future investigations should incorporate the patient's clinical history and account for other thoracic conditions.

Acknowledgments: We thank the Coordination for the Improvement of Higher Education Personnel (CAPES), Brazil, for granting the doctoral scholarship through the Graduate Program in Veterinary Sciences at UFPR to the first author. We also thank all the veterinarians and radiologists who contributed by sending images to compose the database.

6. References

Banzato T, Bonsembiante F, Aresu L, Gelain ME, Burti S & Zotti A. Use of transfer learning to detect diffuse degenerative hepatic diseases from ultrasound images in dogs: a methodological study. The Veterinary Journal. 233:35–40, 2018. doi: 10.1016/j.tvjl.2017.12.026(a)

Banzato T, Cherubini GB, Atzori M. & Zotti A. Development of a deep convolutional neural network to predict grading of canine meningiomas from magnetic resonance images. The Veterinary Journal. 235:90–2, 2018. doi: 10.1016/j.tvjl.2018.04.001 (b)

Banzato T, Causin F, Puppa A.D, Cester G, Mazzai L & Zotti A. Accuracy of deep learning to differentiate the histopathological grading of meningiomas on MR images: a preliminary study. Journal of Magnetic Resonance Imaging, 2019. 50:1152–9. doi: 10.1002/jmri.26723

Banzato T, Wodzinski M, Tauceri F, Donà C, Scavazza F, Müller H. & Zotti A. An AI-Based Algorithm for the Automatic Classification of Thoracic Radiographs in Cats. Frontiers in Veterinary Science, 2021. Oct 15(8):731936. doi: 10.3389/fvets.2021.731936.

Boissady E, De La Comble A, Zhu X, Abbott J & Adrien-Maxence H. 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. Frontiers in Veterinary Science, 2021. Dec 9(8):764570. doi: 10.3389/fvets.2021.764570.

Borgarelli M & Buchanan JW. Historical review, epidemiology and natural history of degenerative mitral valve disease. Journal of Veterinary Cardiology. 14:93-101, 2012. Doi.org/10.1016/j.jvc.2012.01.011.

Boswood A, Häggström J, Gordon SG, Wess G, et al., Effect of Pimobendan in Dogs with Preclinical Myxomatous Mitral Valve Disease and Cardiomegaly: The EPIC Study—A Randomized Clinical Trial. Journal of Veterinary Internal Medicine. 30: 1765-1779 2016. https://doi.org/10.1111/jvim.14586

Buchanan JW. & Bucheler J. Vertebral heart size standards in dogs: attempting to differentiate normal versus abnormal cardiomegaly. Journal of Veterinary Internal Medicine. 9(4):224–230, 1995.

Burti S. Osti LV, Zotti A & BANZATO T. Use of deep learning to detect cardiomegaly on thoracic radiographs in dogs. The Veterinary Journal. 262:1-7, 2020. https://doi.org/10.1016/j.tvjl.2020.105505

Chassagnona G, Vakalopoulou M, Paragios N & Revel MP. Artificial intelligence applications for thoracic imaging. European Journal of Radiology. 123:18774, 2020. doi: 10.1016/j.ejrad.2019.108774

Garcia ML. & Maciel NF. Inteligência artificial no acesso a saúde: Reflexões sobre a utilização da telemedicina em tempos de pandemia. Revista Eletrônica Direito e Política. 15:2, n.2, 2023. https://siaiap32.univali.br/seer/index.php/rdp/article/view/16866/9581.

Haykin S. Neural Networks and Learning Machines. [S.l.: s.n.], 2008. v. 3, p. 906, 2020.

Hicks SA, Strumke I, Thambawita V, Hammou M, Riegler M.A, Halvorsen P & Parasa S. On evaluation metrics for medical applications of artificial intelligence. Scientific Reports. 12:5979, 2022.

Joslyn S, Alexander K. Evaluating artificial intelligence algorithms for use in veterinary radiology. Veterinary Radiology & Ultrasound. 63 (1):871-879, 2022. doi: 10.1111/vru.13159.

Keene BW, Atkins CE, Bonagura JD, Fox PR, Häggström J, Fuentes VL, Oyama MA, Rush JE, Stepien R & Uechi M. ACVIM consensus guidelines for the diagnosis and treatment of myxomatous mitral valve disease in dogs. Journal of Veterinary Internal Medicine. 33(3):1127-1140, 2019. doi: 10.1111/jvim.15488.

Kim E, Fischetti AJ, Sreetharan P, Weltman JG. & Fox PR. Comparison of artificial intelligence to the veterinary radiologist's diagnosis of canine cardiogenic pulmonary edema. Veterinary Radiology & Ultrasound. 63(3): 292-297, 2022. doi: 10.1111/vru.13062.

Lam C, Gavaghan BJ. & Meyers FE. Radiographic quantification of left atrial size in dogs with myxomatous mitral valve disease. Journal of Veterinary Internal Medicine. 35(2): 747-754, 2021. doi: 10.1111/jvim.16073.

Lecun Y, Bengio Y & Hinton G. Deep learning. Nature. 521:436–444, 2015. https://doi.org/10.1038/nature14539

Li S, Wang Z, Visser LC, Wisner ER. & Cheng H. Pilot study: Application of artificial intelligence for detecting left atrial enlargement on canine thoracic radiographs. Veterinary Radiology & Ultrasound. 61(6): 611-618, 2020. doi: 10.1111/vru.12901.

Jeong, Y., Sung, J. An automated deep learning method and novel cardiac index to detect canine cardiomegaly from simple radiography. *Scientific Reports* 12(14494), 2022. doi.org/10.1038/s41598-022-18822-4

Malcolm EL., Visser LC., Phillips KL. & Johnson LR. 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 American Veterinary Medical Association. 253: 1038–1045, 2018. doi: 10.2460/javma.253.8.1038.

Mongan J, Moy L & Kahn CE. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers. Radiology: Artificial Intelligence. 2(2), 2020. doi.org/10.1148/ryai.2020200029.

Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. Patterns 8;2(10):100347,2021.doi:10.1016/j.patter.2021.100347.

Rahmani H, Khan S, Shah SAA. & Bennamoun M. A Guide to Convolutional Neural Networks for Computer Vision. 1. ed. [S.l.]: Morgan & Claypool. P. 207, 2018.

Ross ES, Visser LC, Sbardellati N, Potter BM, Ohlendorf A, Scansen BA. Utility of vertebral left atrial size and

vertebral heart size to aid detection of congestive heart failure in dogs with respiratory signs. Journal of Veterinary Internal Medicine. 37(6):2021-2029, 2023. doi: 10.1111/jvim.16918.

Shorten C, Khoshgoftaar TM, Furht B. Text Data Augmentation for Deep Learning. J Big Data. 8(1):101,2021. doi: 10.1186/s40537-021-00492-0.

Solomon J, Bender S, Durgempudi P, Robar C, Cocchiaro M, Turner S, Watson C, Healy J, Spake J & Szlosek D. Diagnostic validation of vertebral heart score machine learning algorithm for canine lateral chest radiographs. Journal of Small Animal Practice. 64(12): 769-775, 2023. doi: 10.1111/jsap.13666.

Valente C, Wodzinski M, Guglielmini C, Poser H, Chiavegato D, Zotti A, Venturini R, & Banzato T. Development of an artificial intelligence-based method for the diagnosis of the severity of myxomatous mitral valve disease from canine chest radiographs. Frontiers in Veterinary Science. v.10, 2023. doi: 10.3389/fvets.2023.1227009