Veterinary and Animal Science 33 (2026) 100709

Contents lists available at ScienceDirect

Veterinary and Animal Science

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

Evaluation of radiographic heart size measurement repeatability and reproducibility among veterinary students with varying training levels

Radu Andrei Baisan [*], Vasile Vulpe

Faculty of Veterinary Medicine, "Ion Ionescu de la Brad" Iasi University of Life Sciences, 700490 Iasi, Romania

A R T I C L E I N F O

Keywords:
Canine
Cardiomegaly
Heart
Myxomatous mitral valve disease
Radiography
Vertebral heart scale
Vertebral left atrial size

A B S T R A C T

Vertebral heart scale (VHS) and vertebral left atrial size (VLAS) are commonly used to assess the heart size. It remains unknown if observers with different levels of training within the same protocol would improve the repeatability and reproducibility. This study aims to describe the measurement errors and to assess if different levels of experience can influence the outcome.

Radiographs of healthy dogs and dogs with myxomatous mitral valve disease were included. Nine veterinary students were divided into three groups and followed a learning phase by measuring a set of 10, 30 and 50 images respectively, followed by an assessment of measurement error. Next, each group measured 20 radiographs twice, and the measurements of an experienced observer were used to calculate the reproducibility and repeatability.

For VHS and VLAS measurements, 67.8% and 54.1% respectively, had at least one error. Interobserver agreement for VHS was comprised between 0.948 and 0.993 while for VLAS between 0.819 and 0.978. Intraobserver agreement for VHS was comprised between 0.959 and 0.995, while for VLAS between 0.71 and 0.994.

This study shows that the measurement errors are common, however all students obtained a substantial to almost perfect agreement, suggesting that measuring experience does not influence the result.

1. Introduction

Cardiac disease is highly prevalent in dogs, reaching up to 10% of total cases presented to primary veterinary care, of which 75% being represented by myxomatous mitral valve disease (MMVD) (Keene et al., 2019). Thoracic radiography remains one of the most common methods for assessment of the heart size and diagnosis of left sided congestive heart failure in dogs, among general veterinary practitioners (Zhang et al., 2021). The diagnosis and monitoring of animals at increased risk of heart failure or death related to cardiac disease can be pursued using both echocardiography and radiography (Dukes-McEwan et al., 2003; Keene et al., 2019). Nevertheless, echocardiography is the gold standard for diagnosing and staging heart disease in dogs, however a complete ultrasonographic evaluation performed by a cardiologist has limited availability in primary practice, requires specific training for the clinician and the costs are higher compared to radiological examination. Radiographic heart and left atrial size assessment in dogs can be quantified using vertebral heart scale (VHS) and vertebral left atrial size (VLAS) measurements routinely used by general practitioners for triage and follow up.

The American College of Veterinary Internal Medicine (ACVIM) recommendations for diagnosis and treatment of MMVD suggests annual or bi-annual radiologic evaluation for asymptomatic dogs which present echocardiographic signs of mitral regurgitation (Keene et al., 2019).

Both VHS and VLAS are dedicated radiologic methods for assessment of cardiac and left atrium dimensions, with available breed-specific reference ranges (Baisan & Vulpe, 2024, 2025; Lamb et al., 2001). Specifically, VHS is a measurement used mainly for assessment of the general cardiac size without taking into account which exact heart chamber is enlarged, while VLAS is a measurement focused on the left atrial size on radiographic image. In addition, there are more and more studies demonstrating that VHS is highly dependent of breed and thoracic conformation (Jepsen-Grant et al., 2013). Multiple studies have assessed the generalizability and repeatability of these measurements with good results, suggesting that both methods are easy to perform (Levicar et al., 2022) and the observer's experience has a low impact on the outcome, even when assessed among students, veterinarians with or without experience or specialists in cardiology or diagnostic imaging

* Corresponding author.
E-mail address: baisan.andrei_mv@yahoo.com (R.A. Baisan).

https://doi.org/10.1016/j.vas.2026.100709

Available online 23 May 2026

(Baisan & Vulpe, 2022; Levicar et al., 2022; Malcolm et al., 2018; Poad et al., 2020; Vezzosi et al., 2020). However, Hansson et al., 2005 observed that the VHS method for heart size is independent of observer experience but dependent on individual observers' selection of reference points (Hansson et al., 2005). Moreover, VLAS was shown to have a lower performance when assessed by multiple observers, one study results showing only a moderate agreement (Wiegel et al., 2022). Also, Bagardi et al., 2021 found differences in VLAS values among observers with different experience (Bagardi et al., 2021).

To date, it remains unknown whether observers with different levels of training within the same protocol would improve the repeatability and reproducibility of measuring VHS and VLAS.

We hypothesized that (a) the agreement between observers will increase with the level of experience according to the proposed learning process for both measurements and (b) the bias will decrease when comparing examiners with higher levels of expertise between each other and in comparison to a reference observer. Therefore, this observer agreement study aims (a) to describe the main errors that can influence the results and (b) to observe if different levels of experience regarding radiologic examination can influence the outcome.

2. Materials and methods

2.1. Radiographs

This study included 70 right lateral radiographs of healthy dogs and dogs with MMVD, retrospectively selected from the radiology database, divided into two main sets: the learning, composed of 50 images and the test set composed of 20 images. All dogs meeting the inclusion criteria underwent a clinical evaluation, a complete cardiological examination including echocardiography and at least one right lateral chest radiography as a preestablished cardiologic evaluation protocol. The exclusion criteria was the inability to follow the cardiac silhouette margin (eg. dogs with pulmonary edema, lung neoplasia, pleural effusion), inappropriate positioning or radiographs from dogs with vertebral malformations.

2.2. Observers

Nine 4th year veterinary medicine students were randomly selected to participate in this study after completing the radiology module and all had the same experience in radiographs interpretation provided during the radiology module. Students were randomly distributed in 3 groups: G1 (n= 3), G2 (n= 3) and G3 (n= 3) by using the excel random function. All students were voluntarily enrolled in this study. Furthermore, one veterinary doctor (R.A.B.) with >10 years of experience in veterinary cardiology participated as the reference observer (RO).

2.3. Protocol

First, folders with each student's name and phases of the study were created and each set of images were added in the designated folder. Images were exported as DICOM and all measurements were performed in a DICOM viewer (RadiAnt DICOM Viewer, version 2020.2.3, Medixant, Poznań, Poland). After each measurement, the student was asked to save a JPG image with VHS and VLAS measurement within the designated folder. All students had to participate in a general meeting to explain the study protocol (Fig. 1) and tasks to be performed, which lasted for 2 h. Furthermore, measuring methods for VHS and VLAS was explained using a PDF presentation by one of the authors (R.A.B.). In particular, details were offered with regard to the reference points for drawing the lines for both VHS and VLAS. Also, the method for calculating was explained according to literature, using the modified VHS protocol (Hansson et al., 2005) and the VLAS protocol (Malcolm et al., 2018). After the briefing session, each student was asked to measure VHS and VLAS on two images, for practice, apart from the images

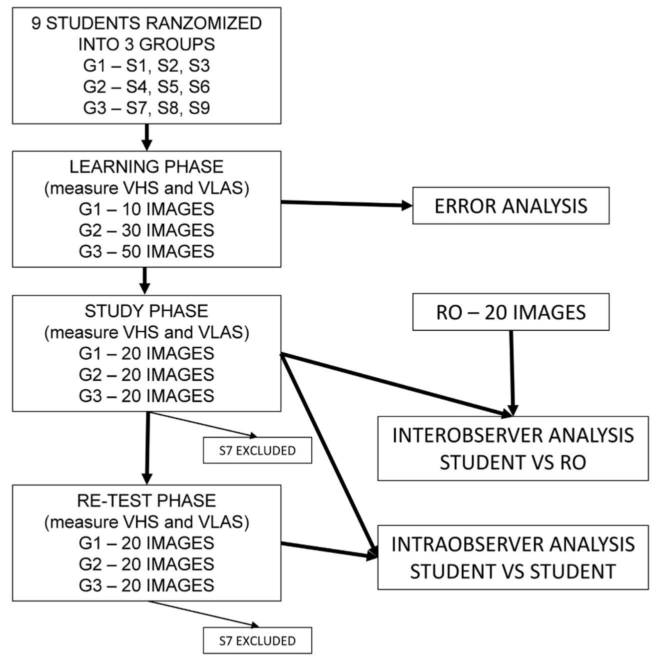

Fig. 1. – Flowchart representing the observers grouping and the following study protocol; G1 – group 1 of observers, G2 – group 2 of observers, G3 – group 3 of observers, S1-S9 - assigned student's number, RO – reference observer.

included in this study. The measurement practice was performed under the supervision of R.A.B. and in the presence of the whole observers group and corrections and discussions based on their measurements were offered. Afterwards, during each student's measurement for learning, test and re-test phases, students were asked to be alone. All observers and RO were blinded to the clinical, echocardiographic data and diagnosis of the dogs.

2.4. Learning phase

The learning phase (LP) included a different number of images in each group for measuring VHS and VLAS. Students in G1 had a set of 10 images, G2 had a set of 30 images (10 images of G1 and 20 additional images) and G3 had a set of 50 images (30 from G2 and 20 additional images). Students from different groups had to practice VHS and VLAS measurements on a different number of images. Based on the images saved as JPG, a measurement error analysis was performed.

The test phase (TP) consisted of a set of 20 images, different from those in the learning phase, but the same images were used for all 3 groups. In this phase the RO also measured VHS and VLAS. An interclass correlation coefficient was calculated for each student from each group with regard to the RO.

The re-test phase (RTP) consisted of the same set of 20 images used in TP. However, all students were asked for a time frame of at least one week between the TP and the RTP. These data were used to calculate the interclass correlation coefficient within each student.

2.5. Errors

Errors were assessed for each measured image. For VHS, the following errors were taken into account: T4-VHS error represented the inability to overlap the long or short axis, or both from the cranial edge of T4; carina error represented the inability to draw the long axis of the heart from the most ventral point of the tracheo-bronchial bifurcation; cardiac apex error represented the inability to stop the long axis at the cardiac apex, or confusing the cardiac apex with other structures such as fat or sternum; short axis position error represented the positioning of the short axis lower than the upper third of the heart silhouette in the middle height of the caudal vena cava (CVC); short axis perpendicularity error represented the inability to draw the short axis perpendicular to the long axis of the heart; short axis cranial point error represented the inability to place the cranial point of the short axis over the cranial cardiac silhouette; short axis caudal point error represented the inability to place the caudal point of the short axis over the caudal cardiac silhouette (Fig. 2A-D). For VLAS, the following errors were taken into account: T4-VLAS error represented the inability to overlap the measured line to the cranial edge of T4; carina error represented the inability to draw the VLAS line from the most ventral point of the traheobronchial bifurcation; CVC error represented the inability to stop the VLAS line at the intersection of the dorsal aspect of the caudal vena cava to the caudal aspect of the cardiac silhouette (Fig. 2E, F). Furthermore, a sub-analysis of measurement error was conducted, comparing the frequency of errors associated with VHS and VLAS assessments on radiographs obtained from clinically normal and preclinical dogs, with those from clinically affected dogs.

2.6. Statistical analysis

Statistical analysis was performed using IBM SPSS v. 21. Demographic and clinical data was tested for normality using Shapiro-Wilk test. Normally distributed data are presented as mean and standard deviation while non-normally distributed data are presented as median and quartiles (Q1-Q3). Errors in establishing the reference points are presented as values and proportions. Differences in MMVD classes among sets of radiographies as well as the error proportion between normal and preclinical dogs group and clinical dogs was assessed using Pearson Chi-Square test. An intraclass correlation (ICC) was used to assess intra- and interobserver measurement agreements for VHS and VLAS by a two-way mixed effect model for absolute agreement. The agreement of the investigators was considered "poor," "fair," "moderate," "substantial," and "almost perfect" when values were in range 0–0.2, 0.21–0.4, 0.41–0.6, 0.61–0.8, and 0.81–1. Finally, the degree of bias for VHS and VLAS between each student and reference observer was assessed by generating Bland-Altman plos, using the online free online software (*BA-plotteR – A web tool for generating Bland-Altman plots & constructing limits of agreement*). A p-wave < 0.05 was considered statistically significant.

3. Results

3.1. Demographics

Seventy radiographs belonging to dogs referred to the cardiology and radiology unit were retrospectively selected. Healthy dogs and dogs with different classes of MMVD according to ACVIM classification were included. The final inclusion decision was approved by R.A.B. and V.V. There were mix breed dogs (*n*= 30), Maltese (*n*= 29), Poodle (*n*= 4),

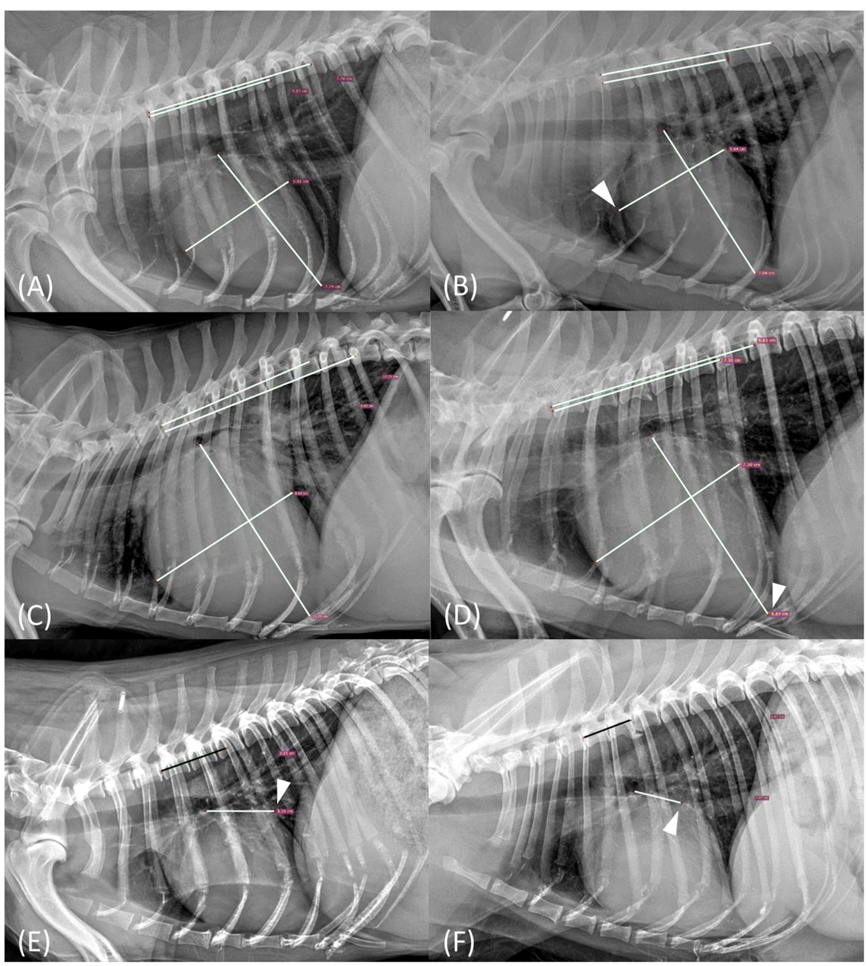

Fig. 2. – Measurement errors for VHS (A-D) and VLAS (E, F). (A) – error in tracing the short axis perpendicular to the long axis; (B) – Short axis cranial point positioned too cranial with regard to the cardiac silhouette (arrowhead); (C) – Short axis positioned lower than the upper third of the heart silhouette and ventral to CVC; (D) – error in positioning the apex reference point on the cardiac silhouette; (E) – Caudal point exceeds the point of intersection between the caudal aspect of the heart and the dorsal aspect of the caudal vena cava (arrowhead); (F) – Caudal point positioned within the silhouette of the left atrium (arrowhead).

Yorkshire terrier (n= 2) and one of each Miniature schnauzer, Shih-tzu, Beagle, Cavalier King Charles Spaniel and Chihuahua. Dogs' demographics are presented in Table 1.

In LP, G1 comprised 30% normal dogs, 30% dogs in class B1, 20% dogs in class B2 and 20% dogs in class C. Group 2 comprised 23.3% normal dogs, 33.3% dogs in class B1, 13.3 dogs in class B2 and 30% dogs in class C. Group 3 comprised 18% normal dogs, 30% dogs in class B1, 8% dogs in class B2 and 44% dogs in class C. In the test group there were 25% healthy and B1 dogs, 20% dogs in class B2 and 30% dogs in class C. There was no statistical difference between MMVD classes among sets of images in the learning and test phases.

3.2. Errors

Of the total 540 radiographic measurements in the study phase, two VHS and two VLAS measurements were not saved and therefore removed from the analysis. Of the remaining 536 radiographic measurements, there were 581 measurement errors in total, with a mean 1.34 (range 0–6) errors per image for VHS and a mean of 0.8 (range 0–3) errors per image for VLAS. Of the 268 VHS measurements, 183 (67.8%) had at least one measurement error (Table 2), while of the 268 VLAS measurements, 146 (54.1%) had at least one measurement error (Table 3). For VHS, G1 had 60% of the measurements with at least one error, G2 had 57.8% and G3 had 75%, while for VLAS, G1 had 43.3%, G2 had 57.8% and G3 had 58%. There was a difference for errors in measuring VHS between groups of students, where G3 had a significantly higher percentage of errors compared to G1 and G2 (P= 0.007). However, there was no difference in percentage of errors between groups when VLAS was measured (P= 0.221). When error occuurence was compared between images from clinical dogs (Class C) and all other dogs, the VHS measurement reveald a statistically significant higher proportion of errors in clinical dogs for short axis perpendicularity ($P<$ 0.01), for short axis cranial point ($P<$ 0.01), and for short axis caudal point ($P<$ 0.01). Furthermore, the same analysis for VLAS revealed a statistically significant higher proportion of error in clinical dogs for CVC point ($P<$ 0.01).

3.3. Interobserver and intraobserver analysis

Interobserver and intraobserver analysis revealed substantial to almost perfect agreement. One student (S7) from G3 was excluded from inter- and intraobserver analysis due to the large number of errors in measurement. In this case, the student erroneously confused the cranial edge of T4 with T3 or T5 in 8 cases for VHS measurement and 9 cases of VLAS measurement during the TP and 11 cases for VHS and 9 cases in VLAS for RTP. For VHS the interobserver agreement between students and RO was comprised between 0.948 (S1G1) (95%CI: 0.759–0.983; $P<$ 0.01) and 0.993 (S8G3) (95%CI: 0.981–0.997, $P<$ 0.01), while for VLAS was comprised between 0.819 (S3G1) (95%CI: 0.552–0.928; $P<$ 0.01) and 0.978 (S8G3) (95%CI: 0.944–0.991; $P<$ 0.01) during TP. In the next stage of our study (RTP), the interobserver agreement between students and RO for VHS was comprised between 0.968 (S4G2) (95%CI: 0.920–0.987; $P<$ 0.01) and 0.99 (S8G3) (95%CI: 0.976–0.996; $P<$ 0.01), while for VLAS was comprised between 0.845 (S3G1) (95%CI: 0.544–0.942; $P<$ 0.01) and 0.979 (S8G3) (95%CI: 0.946–0.992; $P<$ 0.01). The intra-observer variability within student, between TP and RTP, for VHS was comprised between 0.959 (95%CI: 0.899–0.984; $P<$ 0.01) (S1G1) and 0.995 (95%CI: 0.988–0.998; $P<$ 0.01) (S8G3) while for VLAS was comprised between 0.71 (S3G1) (95%CI: 0.273–0.885; $P<$ 0.01) and 0.994 (S8G3) (95%CI: 0.984–0.997; $P<$ 0.01). Interobserver and intraobserver agreement was similar among the 3 groups for VHS, however an increasing trend among groups was observed for interobserver and intraobserver agreement for VLAS measurement (Fig. 3). The interobserver, intraobserver coefficient, 95% CI and statistical significance for each student measuring VHS and VLAS are presented in Supplementary file 1.

Table 1
– Demographic, echocardiographic and diagnosis data of study population.

Measurement	Dogs (n= 70)
Age	11 (8.5–13)
BW	6.6 (5.15–8.7)
Sex (M/F)	48/22
LA/Ao	1.62 (1.37–2.2)
LVIDD_n	1.75 ± 0.44
Normal/MMVD	14/56
MMVD class (B1/B2/C)	20/8/28

BW – body weight, LA/Ao – left atrial to aorta ratio, LVIDD_n – left ventricular internal diameter in diastole normalized to body-weight, MMVD – myxomatous mitral valve disease;.

Table 2
– The value and percentage of errors of the total number of radiographic measurements for VHS.

VHS measurement	Value (% of errors)
T4-VHS	35 (13%)
Carina	60 (22.2%)
Cardiac apex	80 (29.6%)
Short axis position	25 (9.3%)
Short axis perpendicularity	73 (27%)
Short axis cranial point	40 (14.8%)
Short axis caudal point	50 (18.5%)

T4-VHS – the reference point of the cranial aspect of the 4th thoracic vertebra used to measure vertebral heart scale.

Table 3
– The value and percentage of errors of the total number of radiographic measurements for VLAS.

VLAS measurement	Value (% of errors)
T4-VLAS	41 (15.2%)
Carina	65 (24.1%)
CVC	112 (41.5%)

T4-VLAS – the reference point of the cranial aspect of the 4th thoracic vertebra used to measure vertebral left atrial size, CVC – the reference point of the interection between dorsal aspect of the caudal vena cava and cardiac silhouette.

The mean value of agreement per group for VHS and VLAS measurement is depicted in Fig. 3.

3.4. Bias analysis

Bland-Altman graphics were created for assessing the VHS and VLAS measurement bias between RO and each student, excluding S7. For the VHS measurement, the bias ranged between −0.07 v. (RO vs S3 from G1) and −0.44 v. (RO vs S3 from G1). For VLAS measurement, the bias ranged between ±0.03 v. (RO vs S1 from G1 and S9 from G3) and 0.13 v. (RO vs S3 from G1). The bias and limits of agreement for VHS and VLAS are depicted in Fig. 4 and 5.

4. Discussions

This study evaluated the influence of learning experience on the outcome of measuring VHS and VLAS on thoracic radiographs in healthy dogs and dogs with different classes of MMVD by comparing the results of the trained observers with an experienced observer. Moreover, an analysis of the potential errors in choosing the reference points based on the learning process was assessed.

The learning phase for each student was used to assess the errors for choosing the reference points for both VHS and VLAS. Errors with the highest percentage for measuring VHS were represented by defining the

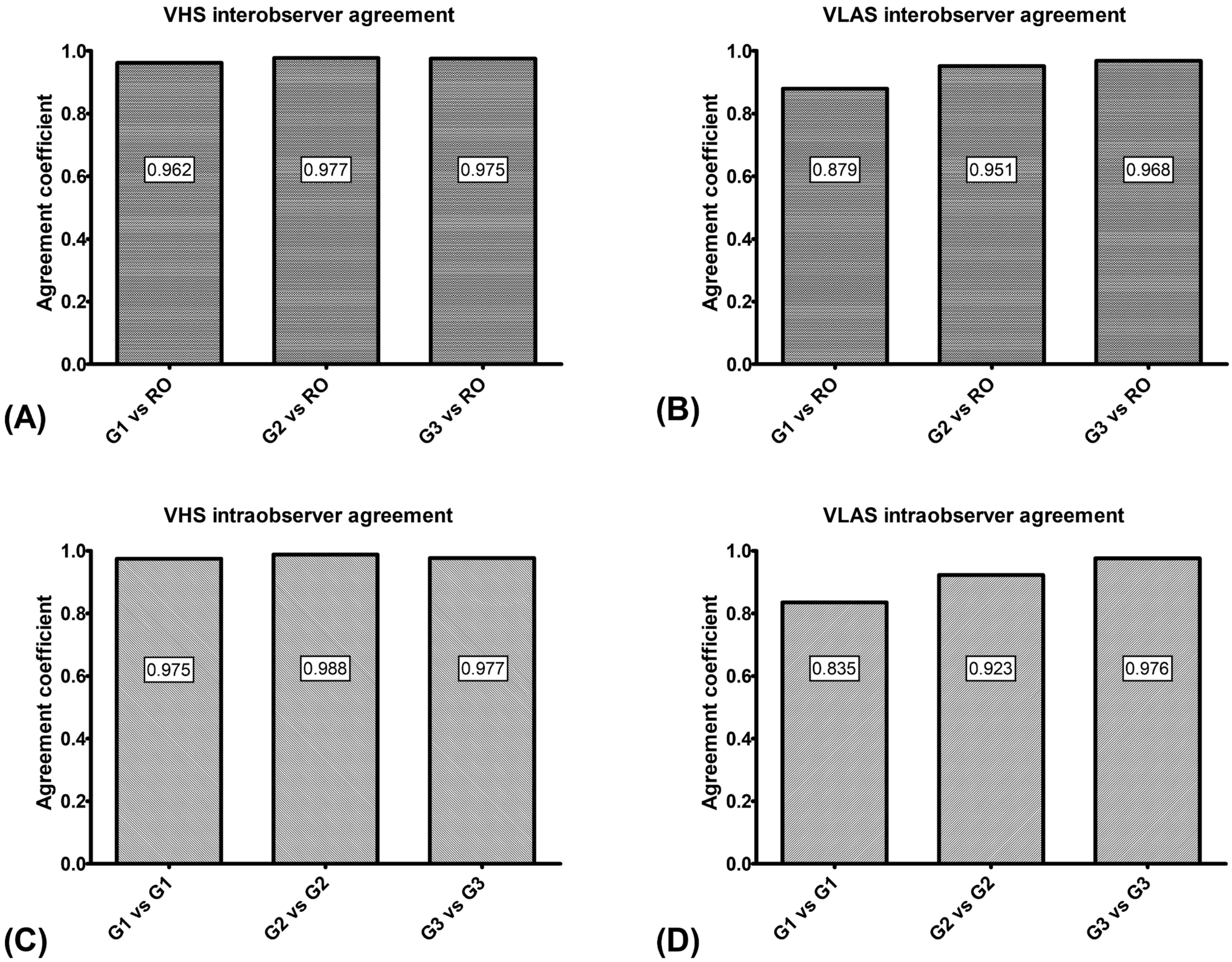

Fig. 3. - Mean value of agreement coefficients per group for VHS and VLAS: A – mean group interobserver agreement between each group and reference observer for VHS; B - mean group interobserver agreement between each group and reference observer for VLAS; C – mean group intraobserver agreement for VHS; D – mean group interobserver agreement for VLAS.

cardiac apex (29.6%) and tracing short axis perpendicular to the long axis (27%), while the lowest error percentage was found in positioning of the short axis over the cardiac hight (9.3%) and defining the cranial edge of T4 (13%). Cranial aspect of the cardiac silhouette of the short axis had a lower error percentage (14.8%) as compared to defining the caudal aspect (18.5%), as the cranial contour of the heart is the easiest point to identify due to the high contrast between the lung parenchyma and the myocardium (Levicar et al., 2022). The inability to identify the caudal point of the short axis (18.5%) halfway between the dorsal and the ventral border of the caudal vena cava can be a consequence of the inability to identify the CVC. The wrong positioning of this landmark can further lead to the inability of drawing the short and long axis perpendicular to each other (as detected in 27% of the VHS measurements). To our understanding, the errors occurring from the wrong drawing of the short axis can be a consequence of either over- or underestimation of the maximal horizontal length of the cardiac silhouette. Another reason can be the difficulty of freehand tracing the two axes perpendicular. For this reason, we suggest that this error could be eliminated by using designated radiological softwares which include this feature. The high error percentage obtained for the cardiac apex (29.6%) might derive from the decreased contrast between cardiac silhouette, ventral thoracic limit and cranial limit of the abdominal cavity, as well as the overlap of the costal cartilages and the sternum.

For VLAS measurement, the highest percentage of errors was defining the CVC point (41.5%) and the lowest was represented by defining the cranial aspect of T4 (15.2%). In our perspective, defining the dorsal point of the CVC can be challenging in some cases especially when perihilar infiltrate is present such as in dogs with pulmonary edema as well as in those where exposure is performed during expiration. Nevertheless, we assume that the lower number of possible errors for VLAS measurement can ease the learning process. Interestingly, there were slight differences in the error percentage for defining cranial aspect of T4 (13%vs 15.25) and carina (22.2%vs 24.1%) between VHS and VLAS although both measurements were performed on the same image and one following the other. In our opinion, this could only be explained by the observer's personal decision of choosing the reference points. The analysis of error proportions comparing radiographic images from clinically affected dogs with those from normal and other MMVD groups demonstrated a higher frequency of measurement errors in the clinical cohort. Notably, identification of the cranial and caudal landmarks defining the short axis in VHS assessment proved more challenging in these cases, likely due to partial effacement of the cardiac silhouette secondary to pulmonary infiltrates. Similarly, accurate localization of the CVC reference point for VLAS measurement was more difficult. This limitation may be attributed to the respiratory status of affected dogs—particularly those with congestive heart failure, who

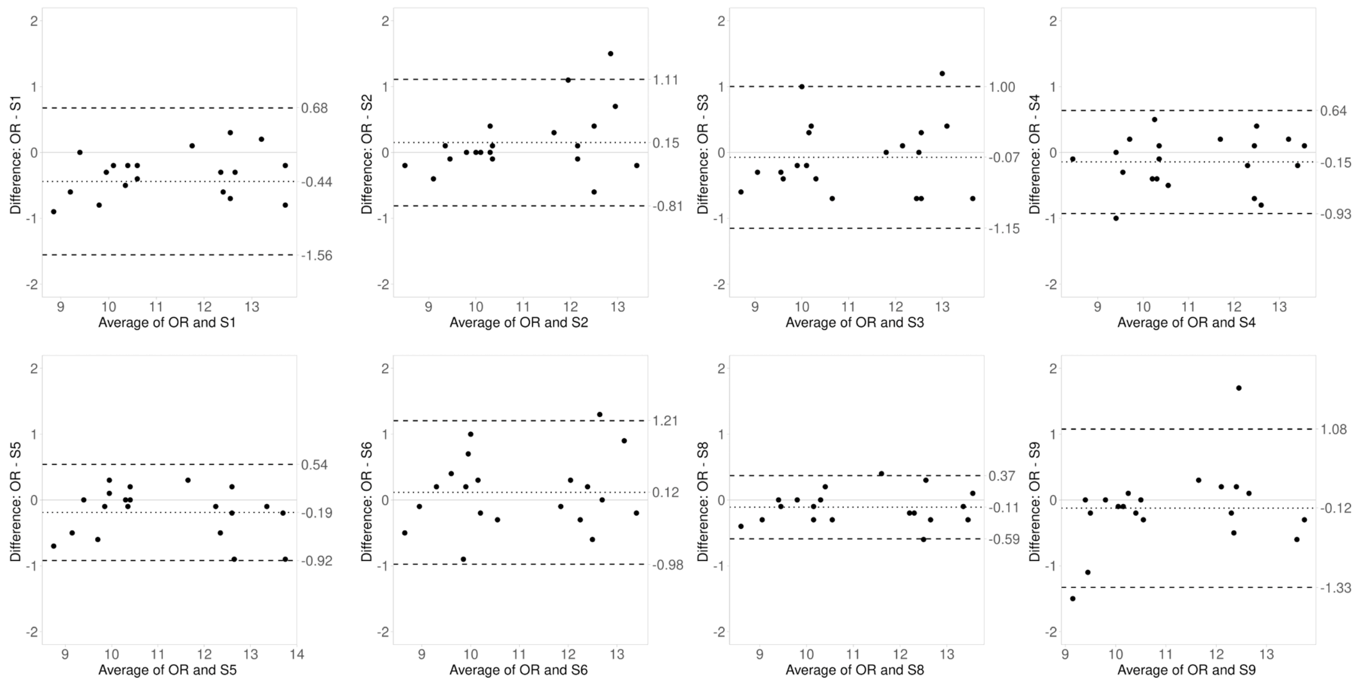

Fig. 4. - Bland–Altman plot evaluating agreement between reference observer and each student for mesuring VHS on thoracic radiography; dotted line represent the mean difference viewed relative to zero (solid line) and dashed line reprezent the 95% limits of agreement; the Y axis range is set to −2/+2 for consistency.

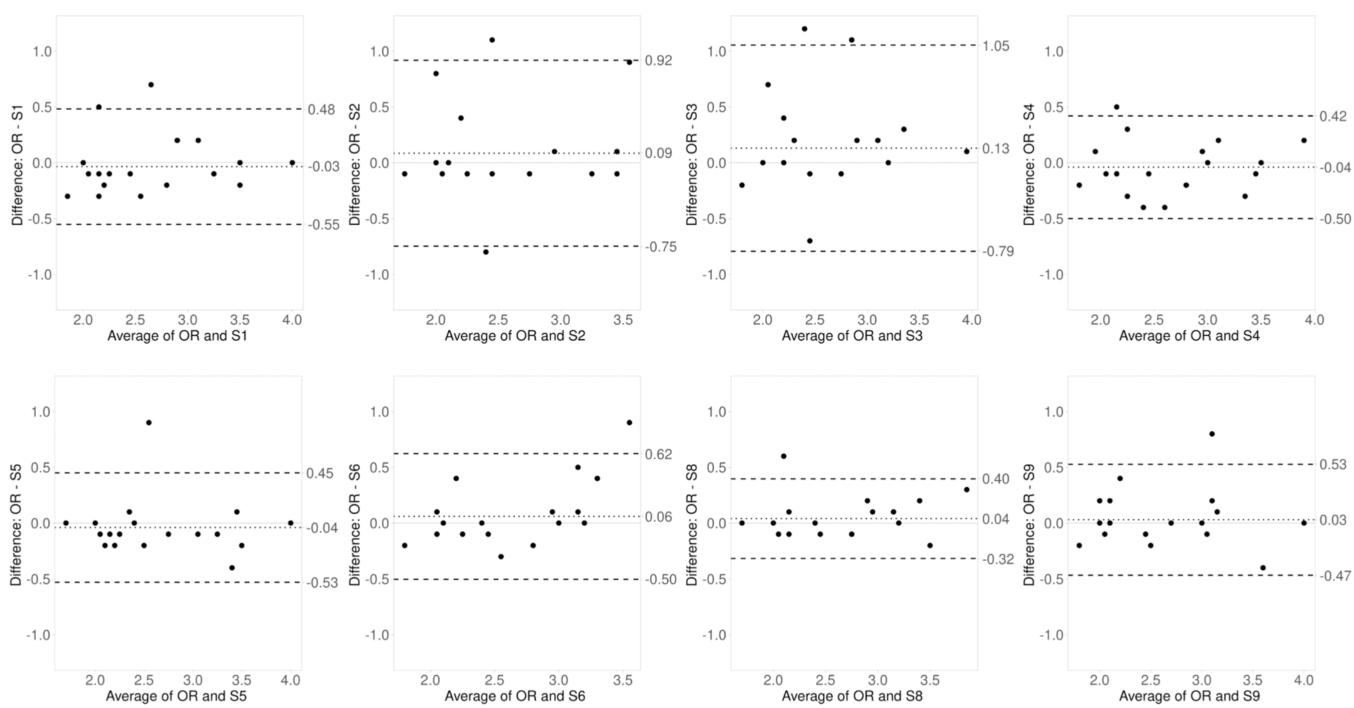

Fig. 5. - Bland–Altman plot evaluating agreement between reference observer and each student for measuring VLAS on thoracic radiography; dotted line represent the mean difference viewed relative to zero (solid line) and dashed line reprezent the 95% limits of agreement; the Y axis is set to −1.2/+1.2 for consistency.

often exhibit tachypnea and reduced lung inflation compared to eupneic individuals. Additionally, the presence of interstitial or alveolar pulmonary patterns may further obscure the CVC landmark, contributing to increased measurement variability.

The inter- and intraobserver agreement for VHS and VLAS have been previously evaluated between observers with the same level of experience for dogs with MMVD (Lam et al., 2021) and dogs with different body condition score (Bruining-Staal et al., 2024), or observers with different levels of experience regarding radiography interpretation for normal dogs (Vezzosi et al., 2020) and dogs with MMVD (Bagardi et al., 2021; Chhoey et al., 2020; Levicar et al., 2022), for assessment of radiological left atrial enlargement (LAE) in comparison with echocardiography (Duler et al., 2018), for specific dog breeds, such as Cavalier King Charles Spaniels (Lord et al., 2011), Maltese (Baisan & Vulpe, 2022), Chihuahuas (Puccinelli et al., 2021), Pugs (Wiegel et al., 2022) and Dachshund (Birks et al., 2017) and in comparison to artificial intelligence (Boissady et al., 2021).

Agreement analysis for our study did not reveal differences among groups. Interobserver and intraobserver agreement for VHS was almost perfect for all groups. However, VLAS revalead a trend of increase with experience for both inter- and intraobserver agreement, with substantial agreement in G1 and almost perfect agreement in G2 and G3. Our findings are in agreement with previous research, in which VHS has been described as the most reproducible radiographic measurement for cardiomegaly, while for VLAS an observer-related influence was detected (Levicar et al., 2022). Similarly, in a previous study performed on healthy Maltese dogs, our research team found a higher interobserver agreement for VHS as compared to VLAS (Baisan & Vulpe, 2022). The agreement between an Italian Recognized Specialist in Veterinary Radiology and a first-year PhD student in veterinary radiology (0.95; 95%CI: 0.83–0.99) was higher for VHS, emphasizing the idea that the protocol for determining VHS has been well clarified in literature and many veterinary practitioners are familiar with the methodology (Vezzosi et al., 2020). Moreover, a study published by Levicar et al., obtained an almost perfect level of agreement (ICC of 0.962; 95%CI: 0.948–0.973) among three different levels of experience observers (a recent graduate, a veterinarian with > 10 years of experience in small animal medicine with a focus on small animal cardiology and a specialist in veterinary radiography (Levicar et al., 2022). In contrast to our findings, Bagardi et al. found differences in radiographic measurements between clinicians with >15 years of experience and both veterinary students or PhD students. Moreover, the same study found differences in VLAS measurements between the veterinary cardiology specialist with higher level of expertise and all the other observers (Bagardi et al., 2021). In our study, all students were trained by the same veterinarian and followed the same training lecture, however their experience did not differ as much as the observers from the previous study.

Importantly, regardless of the high percentage of errors, all participants in this study obtained substantial or almost perfect values of agreement. Previously, it had been suggested that VHS is independent of observer experience but dependent on individual observers' selection of reference points and the transformation of long and short axis dimensions into VHS units (Hansson et al., 2005). The results of our study suggest that observers' personal decision of choosing the reference points does not have a major impact on the measurement outcome.

Bias analysis between each student and reference observer revealed no association with learning experience. For VHS, both students with lowest and highest bias belonged to G1, the group with the lowest experience, while for VLAS, the lowest bias was obtained by two students, one from G1 and one from G3, while the highest bias value was found in a student from G1. Interestingly, the mean VHS bias obtained from all 8 students (- 0,1 v.) was very similar to the bias value (- 0.09 v.) calculated between veterinary cardiology specialists and a machine learning algorithm (Solomon et al., 2023).

Based on the error percentages observed, this study can help improve the feasibility of radiological softwares for VHS and VLAS measurements, implementing a user-friendly methodology with increased accuracy of the results. Moreover, the error analysis can be applied to perfectioning AI in performing cardiac radiological measurements.

This study has several limitations. First, all students participated in one training session which may influence the results with regard to the generalizability of the learning process among different veterinarians from different regions or countries. Secondly, the learning phase consisted of a low number of measured images (a minimum of 10 images in G1 and a maximum of 50 images in G3) and there was no assistance or validation for these measurements within this stage. Therefore, students might have acquired errors during the learning phase which might have persisted in the test and re-test phase. However, when the results were compared to the RO, the correlation coefficient was substantial to almost perfect in all students for both VHS and VLAS. Lastly, the quantifications of the errors detected in the learning phase was subjectively assessed by the RO, and the RO reference points preference could have also been a source of bias.

5. Conclusions

The results of this study reveal that occurrence of errors when performing radiological measurements of the heart is not uncommon, however, these errors do not have a major impact on the outcome. Furthermore, all students regardless of experience obtained a substantial to almost perfect agreement with the reference observer, suggesting that measuring experience does not influence the result. These findings suggest that both VHS and VLAS methods are reliable measurements with good reproducibility and repeatability.

Funding

The authors received no specific funding for this work.

Data integrity statement

All authors have full access to all data in the study and take responsibility for the integrity of the data and accuracy of data analysis.

Ethical statement

This study was based exclusively on the retrospective evaluation of archived thoracic radiographs obtained during routine veterinary clinical practice. No live animals were used, and no procedures were performed specifically for research purposes. Consequently, the study did not require ethical approval or informed consent under applicable institutional and regulatory guidelines.

CRediT authorship contribution statement

Radu Andrei Baisan: Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Vasile Vulpe:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the students who voluntarily participated as observers in this study, to the PhD students Ștefan Acomi, Laura Marina Bilboc and Mălina Maftei and DVM Cătălina Andreea Turcu for assistance and help with the protocol and logistics.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.vas.2026.100709.

Data availability

The data generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

References

BA-plotteR – A web tool for generating Bland-Altman plots and constructing limits of agreement. In.

Bagardi, M., Manfredi, M., Zani, D. D., Brambilla, P. G., & Locatelli, C. (2021). Interobserver variability of radiographic methods for the evaluation of left atrial size in dogs. *Veterinary Radiology & Ultrasound : The Official Journal Of The American College Of Veterinary Radiology And The International Veterinary Radiology Association, 62*(2), 161–174. https://doi.org/10.1111/vru.12930

Baisan, R. A., & Vulpe, V. (2022). Vertebral heart size and vertebral left atrial size reference ranges in healthy Maltese dogs. *Veterinary Radiology & Ultrasound : The Official Journal Of The American College Of Veterinary Radiology And The International Veterinary Radiology Association, 63*(1), 18–22. https://doi.org/10.1111/vru.13027

Baisan, R. A., & Vulpe, V. (2024). Determination of radiographic vertebral heart score and vertebral left atrial size cutoffs based on echocardiographic left atrial size in dogs with myxomatous mitral valve disease. *Veterinary Radiology & Ultrasound : The Official Journal Of The American College Of Veterinary Radiology And The International Veterinary Radiology Association, 65*(1), 54–61. https://doi.org/10.1111/vru.13324

Baisan, R. A., & Vulpe, V. (2025). Vertebral heart score and Vertebral left atrial size as radiographic measurements for cardiac size in dogs-A literature review. *Animals : An Open Access Journal From MDPI, 15*(5), 683. https://doi.org/10.3390/ani15050683

Birks, R., Fine, D. M., Leach, S. B., Clay, S. E., Eason, B. D., Britt, L. G., & Lamb, K. E. (2017). Breed-specific vertebral heart scale for the Dachshund. *Journal of the American Animal Hospital Association, 53*(2), 73–79. https://doi.org/10.5326/JAAHA-MS-6474

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. *Frontiers in Veterinary Science, 8*, Article 764570. https://doi.org/10.3389/fvets.2021.764570

Bruining-Staal, K., Tiemessen, I., Vernooij, J. C. M., & Beijerink, N. (2024). Effect of body condition score on the interobserver variability of vertebral heart scale assessment in dogs. *Veterinary Radiology & Ultrasound : The Official Journal Of The American College Of Veterinary Radiology And The International Veterinary Radiology Association, 65*(6), 694–701. https://doi.org/10.1111/vru.13414

Chhoey, S., Lee, S. K., Je, H., Jung, J. W., Jang, Y., & Choi, J. (2020). Effect of expiratory phase for radiographic detection of left heart enlargement in dogs with mitral regurgitation. *Veterinary Radiology & Ultrasound : The Official Journal Of The American College Of Veterinary Radiology And The International Veterinary Radiology Association, 61*(3), 291–301. https://doi.org/10.1111/vru.12850

Dukes-McEwan, J., Borgarelli, M., Tidholm, A., Vollmar, A. C., & Haggstrom, J. (2003). Proposed guidelines for the diagnosis of canine idiopathic dilated cardiomyopathy. *Journal Of Veterinary Cardiology : The Official Journal Of The European Society Of Veterinary Cardiology, 5*(2), 7–19. https://doi.org/10.1016/S1760-2734(06)70047-9

Duler, L., LeBlanc, N. L., Cooley, S., Nemanic, S., & Scollan, K. F. (2018). Interreader agreement of radiographic left atrial enlargement in dogs and comparison to echocardiographic left atrial assessment. *Journal Of Veterinary Cardiology : The Official Journal Of The European Society Of Veterinary Cardiology, 20*(5), 319–329. https://doi.org/10.1016/j.jvc.2018.07.004

Hansson, K., Haggstrom, J., Kvart, C., & Lord, P. (2005). Interobserver variability of vertebral heart size measurements in dogs with normal and enlarged hearts. *Veterinary Radiology & Ultrasound : The Official Journal Of The American College Of Veterinary Radiology And The International Veterinary Radiology Association, 46*(2), 122–130. https://doi.org/10.1111/j.1740-8261.2005.00024.x

Jepsen-Grant, K., Pollard, R. E., & Johnson, L. R. (2013). Vertebral heart scores in eight dog breeds. *Veterinary Radiology & Ultrasound : The Official Journal Of The American College Of Veterinary Radiology And The International Veterinary Radiology Association, 54*(1), 3–8. https://doi.org/10.1111/j.1740-8261.2012.01976.x

Keene, B. W., Atkins, C. E., Bonagura, J. D., Fox, P. R., Haggstrom, J., Fuentes, V. L., Oyama, M. A., Rush, J. E., Stepien, R., & Uechi, M. (2019). ACVIM consensus guidelines for the diagnosis and treatment of myxomatous mitral valve disease in dogs. *Journal of Veterinary Internal Medicine, 33*(3), 1127–1140. https://doi.org/10.1111/jvim.15488

Lam, C., Gavaghan, B. J., & Meyers, F. E. (2021). Radiographic quantification of left atrial size in dogs with myxomatous mitral valve disease. *Journal of Veterinary Internal Medicine, 35*(2), 747–754. https://doi.org/10.1111/jvim.16073

Lamb, C. R., Wikeley, H., Boswood, A., & Pfeiffer, D. U. (2001). Use of breed-specific ranges for the vertebral heart scale as an aid to the radiographic diagnosis of cardiac disease in dogs. *Veterinary Record, 148*(23), 707–711. <Go to ISI>://000169602000008.

Levicar, C., Nolte, I., Granados-Soler, J. L., Freise, F., Raue, J. F., & Bach, J. P. (2022). Methods of radiographic measurements of heart and left atrial size in dogs with and without myxomatous mitral valve disease: Intra- and interobserver agreement and practicability of different Methods. *Animals : An Open Access Journal From MDPI, 12* (19), 2531. https://doi.org/10.3390/ani12192531

Lord, P. F., Hansson, K., Carnabuci, C., Kvart, C., & Haggstrom, J. (2011). Radiographic heart size and its rate of increase as tests for onset of congestive heart failure in Cavalier King Charles Spaniels with mitral valve regurgitation. *Journal Of Veterinary Internal Medicine, 25*(6), 1312–1319. https://doi.org/10.1111/j.1939-1676.2011.00792.x

Malcolm, E. L., Visser, L. C., Phillips, K. L., & Johnson, L. R. (2018). 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. https://doi.org/10.2460/javma.253.8.1038

Poad, M. H., Manzi, T. J., Oyama, M. A., & Gelzer, A. R. (2020). Utility of radiographic measurements to predict echocardiographic left heart enlargement in dogs with preclinical myxomatous mitral valve disease. *Journal Of Veterinary Internal Medicine, 34*(5), 1728–1733. https://doi.org/10.1111/jvim.15854

Puccinelli, C., Citi, S., Vezzosi, T., Garibaldi, S., & Tognetti, R. (2021). A radiographic study of breed-specific vertebral heart score and vertebral left atrial size in Chihuahuas. *Veterinary Radiology & Ultrasound : The Official Journal Of The American College Of Veterinary Radiology And The International Veterinary Radiology Association, 62*(1), 20–26. https://doi.org/10.1111/vru.12919

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

Vezzosi, T., Puccinelli, C., Tognetti, R., Pelligra, T., & Citi, S. (2020). Radiographic vertebral left atrial size: A reference interval study in healthy adult dogs. *Veterinary Radiology & Ultrasound : The Official Journal Of The American College Of Veterinary Radiology And The International Veterinary Radiology Association, 61*(5), 507–511. https://doi.org/10.1111/vru.12896

Wiegel, P. S., Mach, R., Nolte, I., Freise, F., Levicar, C., Merhof, K., & Bach, J. P. (2022). Breed-specific values for vertebral heart score (VHS), vertebral left atrial size (VLAS), and radiographic left atrial dimension (RLAD) in pugs without cardiac disease, and their relationship to brachycephalic obstructive airway syndrome (BOAS). *PLoS One, 17*(9), Article e0274085. https://doi.org/10.1371/journal.pone.0274085

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. *Preventive Veterinary Medicine, 193*, Article 105399. https://doi.org/10.1016/j.prevetmed.2021.105399