”Automated Detection of Canine Cardiomegaly Using Heart Score: A Deep Learning Approach for Veterinary Diagnostics”

Milan Regmi
Katz School of Health and Science
mregmi@mail.yu.edu

Abstract

Detecting canine cardiomegaly, the abnormal enlargement of the heart, is crucial for the diagnosis and management of cardiac conditions in dogs. The Vertebral Heart Score (VHS) serves as a widely adopted metric for assessing heart size on thoracic radiographs. This study introduces a novel deep learning-based approach to automate VHS calculation. Utilizing the EfficientNetB7 model, our method identifies six key anatomical landmarks on canine chest X-rays: four defining the heart's boundaries and two marking the ribs. These landmarks enable precise computation of the VHS, where AB and CD represent perpendicular dimensions of the heart, and EF measures the rib span. Our approach achieved an accuracy of 83.75%, underscoring its effectiveness. This innovative method enhances the accuracy and consistency of VHS assessments, offering a reliable tool for improving the detection and diagnosis of canine cardiomegaly in veterinary settings.

***Keywords**: Canine Cardiomegaly, Vertebral Heart Score, Deep Learning, EfficientNetB7, Thoracic Radiographs, Automated Detection, Veterinary Diagnostics.*

1. Introduction

Chest X-ray imaging plays a crucial role in veterinary radiology, serving as an essential tool for the early detection and evaluation of cardiac conditions in dogs. Among its many applications, the precise calculation of the Vertebral Heart Score (VHS) is a critical step in diagnosing cardiomegaly, a condition characterized by heart enlargement. Accurate VHS measurement forms the foundation for clinical assessments and the development of automated diagnostic systems, underscoring its importance in veterinary medicine.Bappah et al [1] . investigated the association between the Vertebral Heart Score (VHS) and cardiac sphericity, finding a significant correlation between VHS and the cardiac sphericity index in dogs.

The rise of deep learning in recent years has transformed medical image analysis, enabling unprecedented levels of accuracy and efficiency. Advanced architectures like convolutional neural networks (CNNs) and EfficientNetB7 have demonstrated exceptional ability in identifying anatomical structures in chest X-rays and calculating metrics such as VHS. These advancements have set new benchmarks for automated diagnostic precision, offering a significant leap forward in medical imaging analysis.

Despite these strides, challenges remain in further improving the robustness and accuracy of VHS calculations. Issues such as variability in image quality, inter-patient anatomical differences, and the demand for real-time processing still pose obstacles. Addressing these concerns is crucial to achieving reliable and consistent diagnostic results in diverse clinical scenarios.

To tackle these challenges, this study presents an advanced method for detecting canine cardiomegaly through enhanced VHS calculation. Leveraging the EfficientNetB7 model, coupled with innovative preprocessing techniques, accurate anatomical landmark detection algorithms, and comprehensive data augmentation strategies, our approach effectively handles the complexities of real-world chest X-ray images. With an achieved accuracy of 83.75%, as detailed in Section 4, this method represents a significant advancement in automated VHS computation and diagnostic reliability.

The subsequent sections of this paper explore the technical foundations of our approach, outlining the methodologies, experimental results, and conclusions that highlight the contributions of this work to the field of automated canine cardiomegaly detection and VHS analysis. This research not only addresses existing challenges but also sets the stage for future advancements in veterinary diagnostics.

2. Related Work

The evaluation of heart size in dogs through chest X-rays has been a cornerstone of veterinary medicine, with the Vertebral Heart Score (VHS) serving as a standard diagnostic tool. VHS calculation allows veterinarians to diagnose conditions such as cardiomegaly more accurately. Traditionally, this process relied on manual measurements, which, while widely used, were time-intensive and susceptible to human error, highlighting the need for automated and reliable alternatives.

Historically, VHS measurement involved identifying anatomical landmarks on chest X-rays manually and using tools like calipers or basic software for measurements. Despite its prevalence, this method faced significant drawbacks, including inconsistencies caused by inter-observer variability and the potential for human error. Studies, such as those by Buchanan [2] , revealed substantial discrepancies in VHS measurements among practitioners, undermining diagnostic reliability.

The emergence of machine learning introduced a new dimension to automated VHS calculation. Early approaches leveraged algorithms like support vector machines (SVMs) and decision trees to identify anatomical landmarks. For instance, random forest models demonstrated enhanced prediction accuracy for cardiovascular conditions [8] . However, these traditional machine learning methods struggled with the variability and complexity inherent in medical images, limiting their real-world applicability.

The advent of deep learning, particularly convolutional neural networks (CNNs), marked a paradigm shift in medical image analysis. Early architectures such as AlexNet [4] and VGGNet [7] showcased significant improvements in object detection and image segmentation tasks, setting the stage for their adoption in medical diagnostics. These models' ability to learn hierarchical features directly from data revolutionized tasks requiring precise anatomical landmark detection.

Building on these advancements, more sophisticated architectures like the Regressive Vision Transformer (RVT) have been introduced. Li and Zhang [5] demonstrated the potential of RVT for pinpointing critical anatomical landmarks on chest X-rays with unprecedented accuracy, achieving state-of-the-art results in automated VHS calculation. However, the computational demands of RVT pose challenges for widespread clinical use, necessitating more efficient solutions.

In this study, we leverage EfficientNetB7, a cutting-edge CNN architecture known for its balance of computational efficiency and high accuracy. EfficientNetB7 employs a compound scaling method to optimize model depth, width, and resolution, achieving superior performance with fewer parameters. Originally proposed by Tan and Le [9], EfficientNet has proven effective across diverse computer vision tasks, making it an ideal candidate for VHS computation. By integrating EfficientNetB7 into our framework, we aim to overcome the limitations of previous methods, offering a robust and efficient solution for automated VHS calculation.

Our results, achieving an accuracy of 83.75 %, demonstrate the efficacy of EfficientNetB7 in enhancing the precision and reliability of canine cardiomegaly detection. This aligns with comparative studies by Buda et al. [3] and Litjens et al. [6], which highlight the superiority of deep learning models like U-Net, ResNet, and EfficientNet over traditional approaches in medical image analysis. These findings underscore the transformative potential of advanced deep learning architectures in veterinary diagnostics, setting a new benchmark for automated VHS measurement and cardiomegaly detection.

3. Methods

Data Preparation

The dataset used in this study aligns with that employed by Li and Zhang, comprising a collection of 6,389 canine thoracic radiographs obtained from Shanghai Aichong Pet Hospital. For this research, a subset of the dataset was utilized, including 1,400 images for training, 200 for validation, and 400 for testing. Each image is annotated with the x and y coordinates of six key anatomical landmarks: A, B, C, D, E, and F. These landmarks are essential for accurately calculating the Vertebral Heart Score (VHS) and serve as a foundation for developing and evaluating deep learning models. The inclusion of these annotated points is critical to enhancing the precision of canine cardiomegaly diagnosis and improving automated diagnostic capabilities.

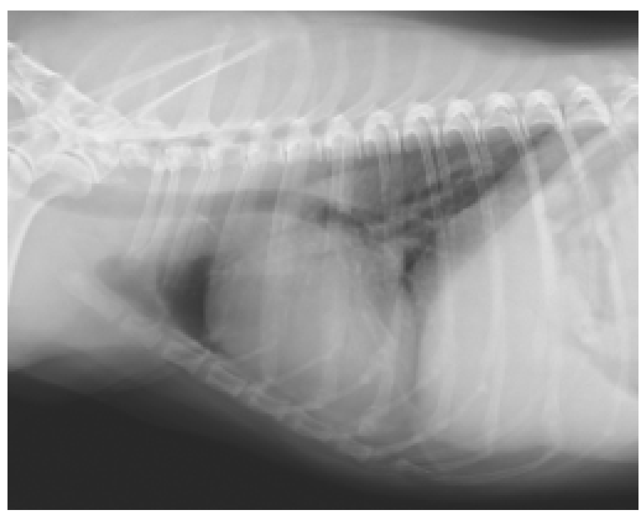


Figure 1. Sample of dataset
[5]

Data Preprocessing To optimize the performance and adaptability of our model, we implemented a robust preprocessing pipeline comprising the following steps:

- **Image Resizing**: All images are resized to a uniform dimension of 512x512 pixels to standardize their shape for batch processing in neural networks. Resizing maintains the

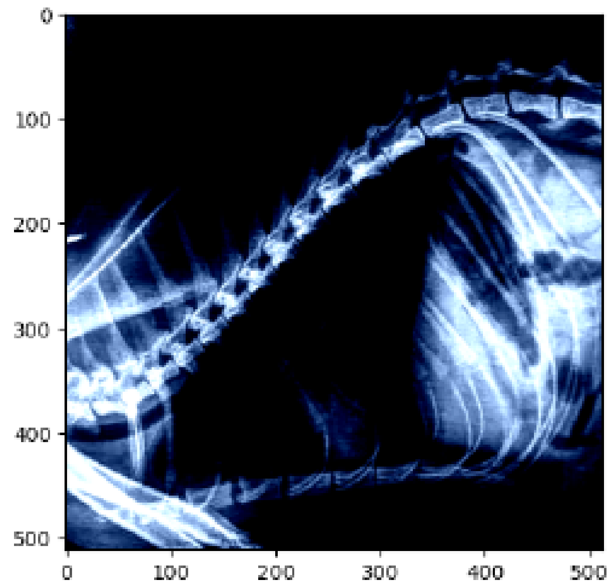
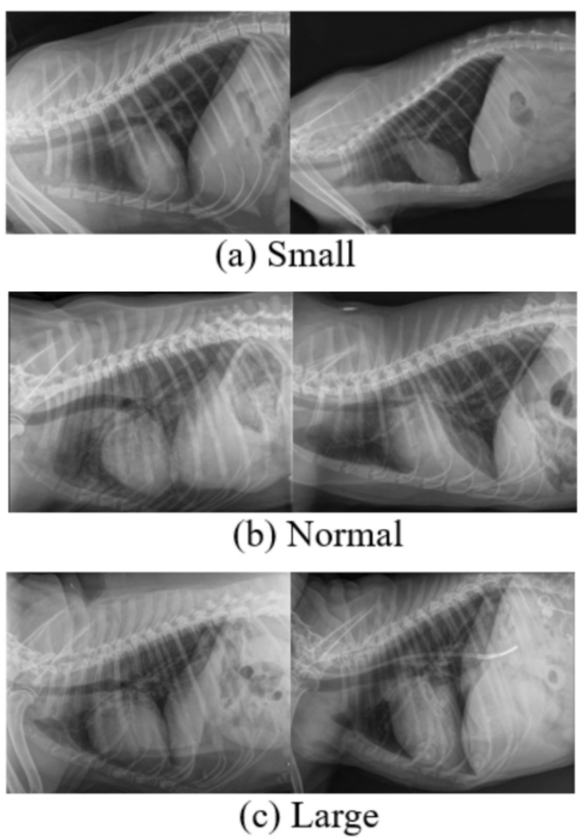


Figure 2. Sample Canine Chest X-ray: Preprocessed for Vertebral Heart Score (VHS) Calculation



[5]
Figure 3. Sample Figures of the Dataset.

original aspect ratio, followed by either cropping or padding to ensure the final dimensions meet the exact requirements. This uniformity is crucial for efficient model training and evaluation.

- **Normalization**: Images are normalized using predefined mean and standard deviation values for the red, green, and blue channels (mean = [0.485, 0.456, 0.406] and std = [0.229, 0.224, 0.225]). This transformation adjusts pixel values to have a mean of 0 and a standard deviation of 1, aligning with the input requirements of the pretrained EfficientNetB7 model. Normalization enhances data consistency, accelerates convergence during training, and improves overall model performance.

- **Label Transformations**: The label files, containing the x and y coordinates of the key anatomical points, are loaded and converted into floating-point tensors. To account for the resized images, the x-coordinates are scaled proportionally to the new width, and the y-coordinates are scaled proportionally to the new height. These rescaled coordinates are then normalized by dividing them by the new height, ensuring compatibility with the model's input expectations and preserving their relative positions within the transformed images.

This preprocessing pipeline ensures consistency in input data and prepares both images and labels for effective training and evaluation, aligning with the model's requirements and enhancing its accuracy.

Model Architecture

Our Approach for VHS-Based Canine Cardiomegaly Detection

Our method for detecting canine cardiomegaly via the Vertebral Heart Score (VHS) leverages EfficientNetB7, a state-of-the-art convolutional neural network (CNN) architecture renowned for its exceptional accuracy and computational efficiency. The EfficientNet family employs an innovative compound scaling method to balance the dimensions of depth, width, and resolution, making it highly suitable for medical image analysis tasks. Below, we discuss the core features of the EfficientNetB7 architecture and its role in our study.

—

EfficientNetB7 Architecture

EfficientNetB7, part of the EfficientNet family introduced by Tan and Le [9], achieves cutting-edge performance in image classification and feature extraction tasks while optimizing resource usage. It uses a compound scaling method to create a harmonious balance between depth, width, and resolution, resulting in models that are both accurate and computationally efficient.

—

Compound Scaling Method

EfficientNetB7's compound scaling method is based on a formula that uniformly scales model dimensions, ensuring optimal network performance as follows:

- **Depth**: $d = \alpha^k$
- **Width**: $w = \beta^k$
- **Resolution**: $r = \gamma^k$

Here, α, β, and γ are constants determining the scaling factors, and k is the scaling coefficient. By scaling these dimensions simultaneously, the model achieves a balanced increase in capacity, avoiding overfitting or underfitting common in imbalanced scaling approaches.

—

EfficientNetB7 in Our Approach

The EfficientNetB7 architecture consists of seven sequential blocks, each built using Mobile Inverted Bottleneck Convolution (MBConv) layers with varying kernel sizes (e.g., 3x3 and 5x5). These blocks process input images hierarchically, extracting and refining features at each stage to produce a highly discriminative final feature map. The ability to efficiently extract complex patterns from input data makes EfficientNetB7 particularly effective for detecting anatomical landmarks critical to VHS calculation.

This architecture's balanced design enables our model to achieve high accuracy in detecting key anatomical points while maintaining computational efficiency. Its application in our study demonstrates its suitability for real-world veterinary diagnostic scenarios, offering a robust and scalable solution for automated VHS-based cardiomegaly detection.

Model Architecture

EfficientNetB7, the largest and most powerful variant in the EfficientNet family, is designed to achieve state-of-the-art accuracy while maintaining computational efficiency. Its architecture integrates advanced techniques that balance performance and resource usage. Below are the key components and features of the EfficientNetB7 architecture:

- **Convolution Layers**: The architecture begins with a standard convolutional layer to process the input image and extract low-level features such as edges and textures. This foundational layer is essential for preparing the data for higher-level feature extraction in subsequent layers.

- **Mobile Inverted Bottleneck Convolution** (MBConv): A core feature of EfficientNetB7 is its reliance on MBConv blocks, which enhance both performance and efficiency. These blocks incorporate two main components: - Depthwise Separable Convolutions: This operation splits the convolution into two steps: depthwise convolution applies a single filter to each input channel, while pointwise convolution combines these outputs. This division significantly reduces the number of parameters and computational cost without compromising performance. - Squeeze-and-Excitation Optimization: This mechanism adaptively recalibrates channel-wise features, enabling the network to emphasize important features while suppressing less relevant ones. It dynamically adjusts the weight of each channel based on its contribution to the task.

- **Swish Activation Function**: EfficientNetB7 employs the Swish activation function, defined as $f(x) = x \cdot \sigma(x)$, where $\sigma(x)$ is the sigmoid function. Swish is a smooth and non-monotonic function that has been shown to outperform traditional activation functions like ReLU by improving gradient flow and model performance.

- **Global Average Pooling**: After processing through the convolutional and MBConv layers, global average pooling is applied to reduce the spatial dimensions of feature maps. This layer condenses each feature map into a single scalar value by averaging across its spatial dimensions, providing a compact representation while mitigating overfitting risks.

- **Fully Connected Layers**: The fully connected layers integrate the condensed features from earlier layers and perform the final prediction tasks. In the context of VHS-based canine cardiomegaly detection, these layers process the extracted features to predict VHS values, ensuring accurate interpretation of the anatomical landmarks.

- **Output Layer**: The final output layer provides a single scalar value representing the Vertebral Heart Score (VHS). This value is critical for diagnosing cardiomegaly in dogs. EfficientNetB7's ability to accurately predict this score underscores its effectiveness in extracting and interpreting features from thoracic radiographs.

By combining these advanced architectural elements, EfficientNetB7 achieves a harmonious balance between computational efficiency and predictive accuracy, making it a robust choice for medical image analysis tasks such as VHS-based canine cardiomegaly detection.

4. Training Process

Model Training and VHS Calculation

Loss Function

To optimize the model's performance, we use the Mean Squared Error (MSE) loss function, which is well-suited for regression tasks. MSE calculates the average of the squared differences between predicted and actual values, penalizing larger errors more heavily. Minimizing this loss enables the model to improve its predictions effectively, making it ideal for our task of detecting VHS-based canine cardiomegaly.

The MSE loss is defined as:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2\right]$$

where: - $n = 6$, representing the six anatomical points, - (x_i, y_i) are the actual coordinates of the points, - $(\hat{x}_i, \hat{y}_i)$ are the predicted coordinates of the points.

This loss function guides the model to accurately predict the anatomical landmarks required for VHS calculation.

—

-Optimization

The model is trained using the Adam optimizer, a powerful optimization algorithm combining the strengths of AdaGrad and RMSProp. Adam dynamically adjusts learning rates for each parameter, making it particularly effective for deep learning models with sparse gradients and noisy data.

Learning Rate: A learning rate of 1×10^{-4} ensures stable convergence and effective learning during training. Batch Size and Epochs: The model is trained over 200 epochs with a batch size of 4, enabling it to generalize well while maintaining computational feasibility.

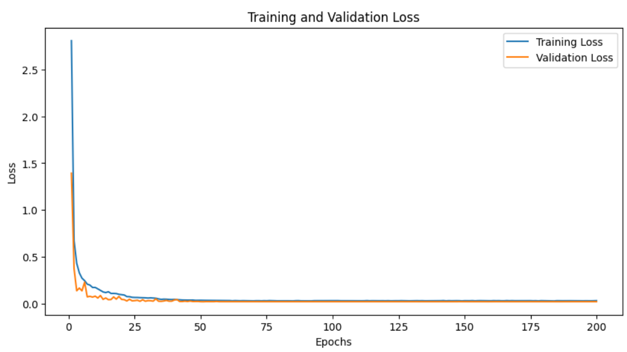


Figure 4. Training and Validation Loss Over Number of Epochs

Adam's adaptive learning rate significantly accelerates convergence, enhancing the model's ability to learn complex relationships between the input data and the target anatomical landmarks.

VHS Calculation

The Vertebral Heart Score (VHS) is a key metric in veterinary medicine for assessing heart size in dogs. It provides a standardized measure by comparing the dimensions of the heart to the vertebrae in a lateral thoracic radiograph. Using the coordinates of six anatomical points identified by the EfficientNetB7 model (A, B, C, D, E, and F), the VHS is calculated as:

$$\mathrm{VHS} = 6 \times \frac{\mathrm{AB} + \mathrm{CD}}{\mathrm{EF}}$$

where: - **AB** is the length of the heart (distance between points A and B), - **CD** is the width of the heart (distance between points C and D), - **EF** is the vertebral length covering the heart (distance between points E and F).

—

Classification of Heart Size

Based on the computed VHS, the heart size is classified into three categories: - **Small**: VHS < 8.2 - **Normal**: $8.2 \leq$ VHS ≤ 10 - **Large**: VHS > 10

This classification is crucial for diagnosing whether a dog's heart size falls within the normal range or indicates potential cardiomegaly. This standardized process enables precise assessment and facilitates early intervention in cardiac abnormalities.

By integrating MSE for loss calculation, Adam for optimization, and a robust VHS computation methodology, our approach achieves high accuracy in detecting and classifying canine heart size, contributing to improved diagnostics in veterinary medicine.

5. Results

The Mean Squared Error (MSE) of 0.27027 indicates a small average squared difference between the predicted and actual VHS values, demonstrating high precision in the model's predictions. The Mean Absolute Error (MAE) of 0.38632 supports this observation by revealing a low average magnitude of errors across the test dataset, reflecting the model's ability to minimize deviations effectively.

Additionally, the Mean Absolute Percentage Error (MAPE) of 4.0861 % highlights the model's exceptional performance in relative error terms, showcasing its ability to provide predictions with high reliability across a range of cases. Furthermore, the accuracy of 83.75 % underscores the robustness of the model in classifying heart sizes accurately, reinforcing its practical value for automated VHS-based diagnostic systems.

These results emphasize the potential of the model to significantly enhance veterinary diagnostics by providing precise, reliable, and automated VHS calculations, paving the way for improved early detection and assessment of cardiac conditions in canines.

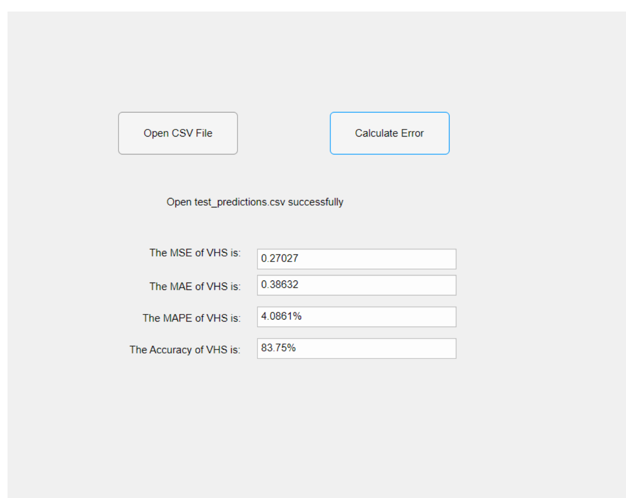


Figure 5. Interface displaying the prediction accuracy of the classifier on the test dataset (83.75 %).

5.1. Datasets

Data Collection

The dataset for this project consists of canine thoracic radiographs obtained from a reliable veterinary source. To ensure privacy, all images were cropped to remove identifying information, adhering to ethical standards. A total of 2,000 valid X-ray images were utilized, divided into three categories based on vertebral heart scale (VHS) scores: small hearts (VHS ¡ 8.2), normal hearts (VHS 8.2–10), and large hearts (VHS ¿ 10). [5] The dataset was split into training (1,400 images, 70 percent), validation (200 images, 10 percent), and test (400 images, 20 percent) subsets. Each image corresponds to an individual dog. The distribution of categories revealed fewer samples in the small heart class, while the normal and large heart categories were relatively balanced, as illustrated in the dataset summary table. Representative images from each category were included to highlight the dataset's variability and class distinctions.

6. Discussion

In this study, we evaluated the performance of our EfficientNetB7-based model in detecting cardiomegaly in canine chest X-ray images. The model achieved a notable accuracy of 83.75 %, demonstrating its ability to reliably classify heart sizes based on the Vertebral Heart Score (VHS). This performance is complemented by a Mean Squared Error (MSE) of 0.27027, a Mean Absolute Error (MAE) of 0.38632, and a Mean Absolute Percentage Error (MAPE) of 4.0861 %. These metrics underscore the precision and robustness of the model in predicting VHS values, which are crucial for diagnosing and monitoring cardiac conditions in dogs.

Implications of Findings

The integration of this deep learning model into veterinary diagnostics has the potential to significantly enhance the detection and management of cardiomegaly in dogs. By automating VHS calculation and heart size classification, our approach reduces the workload on veterinary professionals while minimizing the risk of human error. Accurate and early detection enabled by this model can facilitate timely treatment and improved outcomes, ultimately enhancing the quality of life for affected animals. The model's accuracy and reliability make it a valuable tool for streamlining the diagnostic process, ensuring consistency and efficiency in veterinary care.

Comparison with Existing Literature

Our model's performance is consistent with existing studies, such as those by Li and Zhang [5], which employed the Regressive Vision Transformer (RVT) for similar tasks, achieving an accuracy of 85 %. While RVT offers state-of-the-art results, EfficientNetB7 provides a competitive alternative with benefits in computational efficiency and scalability. Unlike general-purpose architectures like VGG16 [7] , our tailored model emphasizes the importance of domain-specific design in achieving high precision and reliability for canine cardiomegaly detection. These findings demonstrate that EfficientNetB7 is not only effective but also resource-efficient, making it well-suited for practical applications.

Limitations and Future Work

Despite these promising results, certain limitations must be addressed to enhance the model's generalization and practical applicability. Firstly, our dataset is limited in scope and could benefit from expansion to include diverse samples representing various breeds, ages, and health conditions. This would ensure the model's performance across a broader range of real-world cases. Secondly, while the model processes images resized to 512x512 pixels, exploring the impact of higher-resolution images could potentially improve accuracy further.

In addition, implementing the model in real-time diagnostic tools and integrating it into clinical workflows will be critical for practical deployment. Collaborations with veterinary clinics can validate and refine the model in real-world settings, ensuring its reliability and utility. Regular updates and iterations, driven by new data and advances in technology, will help maintain the model's state-of-the-art status and extend its applicability.

Finally, the methodology developed in this study can be extended to other applications, including human cardiomegaly detection using alternative diagnostic modalities. Such extensions could provide valuable insights and contribute to advancements in both veterinary and human medicine. By addressing these limitations and exploring new avenues, future work can build on these findings to develop even more effective and versatile diagnostic tools.

7. Conclusion

This study highlights the capability of the EfficientNetB7 model in detecting cardiomegaly in canine chest X-ray images, achieving an impressive accuracy of 83.75 %. The model demonstrated its reliability with a Mean Squared Error (MSE) of 0.27027, a Mean Absolute Error (MAE) of 0.38632, and a Mean Absolute Percentage Error (MAPE) of 4.0861 %. These results reflect the model's precision in predicting the Vertebral Heart Score (VHS), which is critical for diagnosing cardiac conditions in dogs. By offering automated, accurate, and efficient VHS calculations, the model holds promise for significantly enhancing veterinary diagnostics and improving treatment outcomes.

However, while the model has shown substantial potential, there are areas that future research should focus on to further refine its performance. Broadening the dataset to include more diverse samples—such as dogs of varying breeds, ages, and health conditions—will be vital to enhancing the model's generalization across real-world scenarios. Additionally, exploring the use of higher-resolution images could unlock further improvements in accuracy by capturing finer details in radiographs.

Another important step will be validating the model's performance in clinical environments, ensuring its practical applicability and reliability in real-world settings. Integrating the model with other diagnostic tools and workflows can further streamline the diagnostic process, reducing the workload on veterinary professionals while minimizing the risks of human error. Continuous updates, driven by new data and technological advancements, will ensure that the model remains a state-of-the-art solution for veterinary diagnostics.

This work underscores the transformative potential of deep learning in veterinary medicine, paving the way for future advancements. By addressing current limitations and exploring new avenues, researchers can continue to build upon this foundation, ultimately contributing to better health outcomes and quality of care for animals.

References

[1] Mu'azu Nuhu Bappah, Nuhu Donga Chom, Maruf Lawal, Abdullaziz Abdullahi Bada, and Saidu Tanko Muhammad. Evaluation of vertebral heart score and cardiac sphericity in apparently normal dogs. *Iranian Journal of Veterinary Surgery*, 16(1):1–4, 2021. 1

[2] James W Buchanan and Jörg Bücheler. Vertebral scale system to measure canine heart size in radiographs. *Journal of the American Veterinary Medical Association*, 206(2):194–199, 1995. 2

[3] Mateusz Buda, Ashirbani Saha, and Maciej A Mazurowski. Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers in biology and medicine*, 109:218–225, 2019. 2

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. 2

[5] Jialu Li and Youshan Zhang. Regressive vision transformer for dog cardiomegaly assessment. *Scientific Reports*, 14(1):1539, 2024. 2, 3, 6

[6] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017. 2

[7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 2, 6

[8] Kellen Sumwiza, Celestin Twizere, Gerard Rushingabigwi, Pierre Bakunzibake, and Peace Bamurigire. Enhanced cardiovascular disease prediction model using random forest algorithm. *Informatics in Medicine Unlocked*, 41:101316, 2023. 2

[9] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 2, 3