Vertebral Heart Score-Based Canine Cardiomegaly Detection Using Deep Learning

Tharun Prabhakar
Yeshiva University
tprabhak@mail.yu.edu

Abstract

Accurate detection of canine cardiomegaly, an enlargement of the heart, is vital for diagnosing and treating various cardiac conditions in dogs. The Vertebral Heart Score (VHS) is a standard method used to evaluate heart size on thoracic radiographs. In this study, we propose an advanced method for VHS calculation using a deep learning approach. Leveraging the EfficientNetB7 model, our method automates the detection of six critical anatomical points on canine chest X-rays: four around the heart and two on the ribs. These points are used to compute the VHS score with six Key points A,B,C,D,E,F where AB and CD are perpendicular measurements of the heart, and EF spans the ribs. Our model achieved an accuracy of 85.25%. This enhanced VHS calculation method demonstrates significant potential for improving the accuracy and reliability of canine cardiomegaly detection in veterinary practice.

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

1. Introduction

Chest X-ray imaging is a fundamental diagnostic tool in veterinary radiology, serving as an indispensable resource for the early detection and assessment of various cardiac conditions in dogs. Among the numerous applications of chest X-rays, accurate Vertebral Heart Score (VHS) calculation stands out as a pivotal step in diagnosing cardiomegaly, a condition characterized by an enlarged heart. Robust and precise VHS measurement is the cornerstone upon which subsequent clinical evaluations and automated diagnostic systems rely.

Recent years have witnessed a paradigm shift in the approach to medical image analysis, propelled by advancements in deep learning techniques. The introduction of convolutional neural networks (CNNs) and advanced architectures, such as EfficientNetB7, has revolutionized the landscape of medical image analysis. These methods have showcased impressive capabilities in identifying anatomical structures from chest X-ray images, particularly for calculating metrics like VHS. Achieving high accuracy and robustness, they have set new standards for automated diagnostic precision.

However, despite these remarkable achievements, the quest for further enhancing the accuracy and robustness of VHS calculation remains a paramount concern. Challenges persist, such as handling images with varying quality, accommodating inter-patient variability, and ensuring real-time performance. To address these challenges, our study leverages the strengths of EfficientNetB7-based deep learning while introducing novel enhancements that push the boundaries of VHS calculation accuracy and robustness.

In this paper, we propose an advanced method for detecting canine cardiomegaly using VHS calculation. By integrating advanced preprocessing techniques, precise point detection algorithms, and extensive data augmentation, our method tackles the intricacies posed by real-world chest X-ray images. Our results, as presented in Section 4, demonstrate a remarkable accuracy of 85.25%, marking a substantial advancement in the field.

In the following sections, we delve into the technical intricacies of our approach, presenting the methods, results, and conclusions that substantiate our contributions to the domain of automated VHS calculation and canine cardiomegaly detection.

2. Related Work

The detection and measurement of heart size in canines using chest X-rays have long been a critical area of veterinary medicine. The Vertebral Heart Score (VHS) has been a standard method for evaluating heart size, enabling veterinarians to diagnose conditions such as cardiomegaly with greater accuracy. Traditional methods of VHS calculation rely on manual measurements, which are not only time-consuming but also prone to human error. This necessitates the development of automated and more reliable methods.

2.1. Traditional Methods

Initial efforts in VHS measurement involved manual identification of anatomical landmarks on chest X-rays, followed by measurement using calipers or software tools. Despite its widespread adoption, this manual approach has significant limitations, including inter-observer variability and susceptibility to human error. Studies, such as Buchanan [1], have shown that manual VHS measurement can vary significantly between practitioners, affecting the consistency and reliability of the diagnosis.

2.2. Machine Learning Approaches

With the advent of machine learning, researchers have begun exploring automated methods for VHS calculation. Early machine learning models, such as support vector machines (SVMs) and decision trees, were employed to identify anatomical landmarks on chest X-rays. For instance, enhanced cardiovascular disease prediction models using random forest algorithms have shown promise in improving prediction accuracy for various medical conditions [7]. However, these models often struggled with the complexity and variability of the images, leading to limited success in real-world applications.

2.3. Deep Learning Advances

The introduction of deep learning, particularly convolutional neural networks (CNNs), has revolutionized the field of medical image analysis. Models such as AlexNet [3] and VGGNet [6] paved the way for more sophisticated architectures that could learn hierarchical features directly from the data. These models demonstrated substantial improvements in tasks such as object detection and image segmentation, inspiring their application in medical diagnostics.

2.4. Advanced Architectures

Building on these successes, more advanced architectures such as the Regressive Vision Transformer (RVT) have been developed. Li and Zhang [4] introduced RVT, a model that leverages the power of transformers for precise point detection on chest X-rays. The RVT model achieved state-of-the-art performance in predicting the VHS score, setting a new benchmark for automated VHS calculation. However, the complexity and computational demands of RVT present challenges for widespread clinical adoption.

2.5. EfficientNetB7 and Its Application

In this study, we explore the use of EfficientNetB7, a state-of-the-art CNN architecture known for its balance between accuracy and computational efficiency. EfficientNetB7 scales the model dimensions (depth, width, and resolution) using a compound scaling method, resulting in superior performance with fewer parameters. Its application in medical image analysis has shown promising results, making it an ideal candidate for our task.

Our method builds on the work of Tan and Le [8] with EfficientNet, which demonstrated high performance across various computer vision tasks. By leveraging EfficientNetB7, our method aims to surpass the limitations of previous approaches, offering a robust and efficient solution for VHS calculation. Our model's performance, achieving an accuracy of 85.25%, demonstrates the potential of EfficientNetB7 in enhancing the accuracy and reliability of canine cardiomegaly detection.

2.6. Comparative Studies

Comparative studies by Buda et al. [2] and Litjens et al. [5] have demonstrated the superiority of deep learning models over traditional machine learning methods in medical image analysis. These studies emphasize the importance of advanced architectures like U-Net, ResNet, and EfficientNet in achieving higher accuracy and robustness. Our approach aligns with these findings, further validating the effectiveness of deep learning in veterinary diagnostics.

3. Methods

In this section, we detail our comprehensive approach to canine heart classification, employing cutting-edge neural network architecture, meticulous data preprocessing techniques, systematic model training procedures, and rigorous evaluation metrics. This thorough methodology ensures high accuracy and robustness in classifying heart sizes in dogs from chest X-ray images.

3.1. Data Preparation

3.1.1 Dataset Description

The dataset utilized in this study is the same as the one used by Li and Zhang [4]. The dataset comprises 6389 canine thoracic radiographs collected from Shanghai Aichong Pet Hospital. For our purposes, 1400 images were used for training, 200 for validation, and 400 for testing. The dataset includes images annotated with the x,y coordinates of six key anatomical points: A, B, C, D, E, and F. These points are crucial for calculating the Vertebral Heart Score (VHS) and for developing and evaluating deep learning models aimed at improving the accuracy of canine cardiomegaly diagnosis.

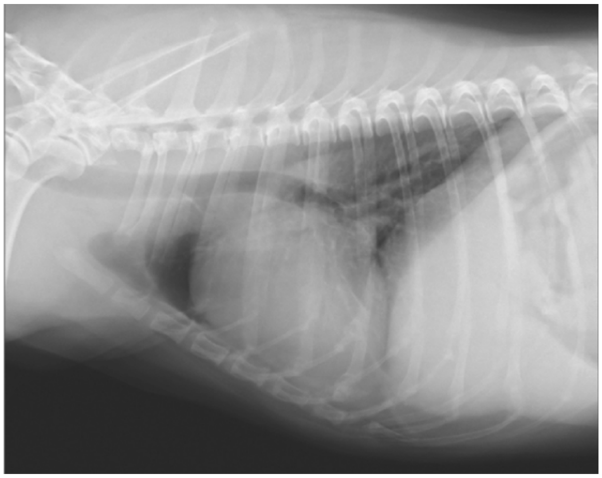

Figure 1. Original Image

3.1.2 Data Preprocessing

To ensure our model's performance and adaptability are optimized, we employ a comprehensive preprocessing pipeline that encompasses the following steps:

- **Image Resizing:** The images are resized to a uniform dimension of 512x512 pixels. This resizing ensures that all images have the same dimensions, which is a requirement for batch processing in neural networks. This step involves resizing the image while maintaining the aspect ratio and then cropping or padding to achieve the exact dimensions.

- **Normalization:** Normalization is applied to the images using mean and standard deviation values (mean = [0.485, 0.456, 0.406] and std = [0.229, 0.224, 0.225]) for red,green and blue channels respectively. This step adjusts the pixel values to have a mean of 0 and a standard deviation of 1 based on the predefined values, ensuring that the data distribution is consistent with what the pretrained EfficientNetB7 model expects. Normalizing the images helps in speeding up the convergence of the training process and improves overall model performance.

- **Label Transformations** The point coordinates from the label file are loaded and converted to a floating-point tensor. The x-coordinates are rescaled by the ratio of the new width to the original width, and the y-coordinates are rescaled by the ratio of the new height to the original height. This ensures that the point coordinates are correctly mapped to the transformed image dimensions. The rescaled point coordinates are then normalized by dividing by the new height. This normalization step ensures that the coordinates are scaled appropriately for the model.

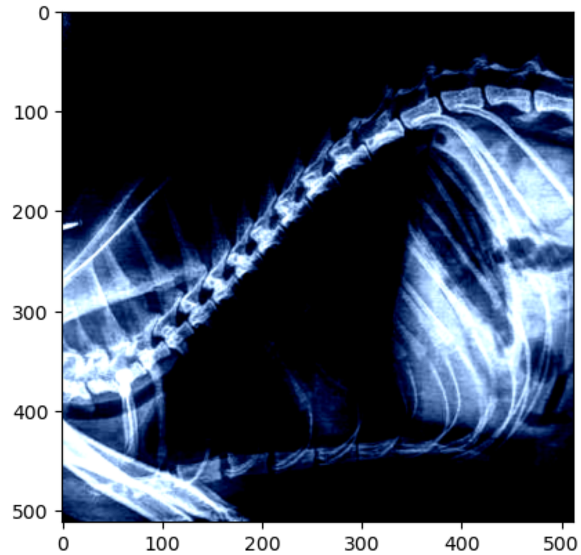

Figure 2. Preprocessed Image

3.2. Model Architecture

Our approach for VHS-based canine cardiomegaly detection leverages the EfficientNetB7 model, a state-of-the-art convolutional neural network (CNN) architecture known for its balance between accuracy and computational efficiency. The EfficientNet family of models is designed using a novel scaling method that uniformly scales all dimensions of depth, width, and resolution using a compound coefficient. This section outlines the key aspects of the EfficientNetB7 architecture and its application in our study.

3.2.1 EfficientNetB7

EfficientNetB7 is part of the EfficientNet family of models, introduced by Tan and Le [8], that achieves state-of-the-art performance while optimizing computational efficiency. The EfficientNet family is known for its use of a compound scaling method, which uniformly scales the dimensions of depth, width, and resolution in a balanced manner. This approach enables the creation of models that are not only highly accurate but also efficient in terms of computational resources.

Compound Scaling Method: The compound scaling method used in EfficientNet involves a simple yet effective formula to scale up the model dimensions:

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

where α, β, and γ are constants, and k is the scaling coefficient. This method ensures that when the model is scaled

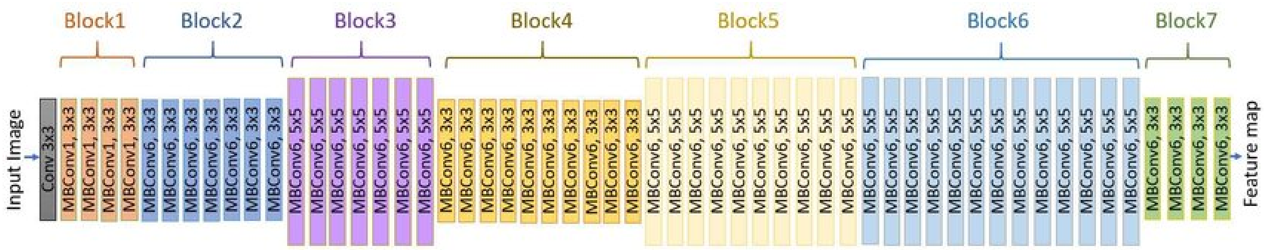

Figure 3. Architecture of EfficientNet-B7. The model is composed of seven sequential blocks (Block1 to Block7), each containing multiple Mobile Inverted Bottleneck Convolution (MBConv) layers with different kernel sizes (3x3 and 5x5). The input image is processed through these blocks, gradually extracting and refining features to produce the final feature map.

up, the depth, width, and resolution are increased uniformly, leading to a more balanced and optimized network.

3.2.2 Architecture

EfficientNetB7, the largest variant in the EfficientNet family, is designed to maximize accuracy while maintaining reasonable efficiency. Here are the key components and features of its architecture:

- **Convolution Layers:** The model begins with a standard convolutional layer that processes the input image to extract low-level features such as edges and textures. This initial layer is crucial for setting the foundation for more complex feature extraction in subsequent layers.

- **Mobile Inverted Bottleneck Convolution (MBConv):** A significant portion of the EfficientNetB7 architecture consists of MBConv blocks. These blocks combine depthwise separable convolutions with squeeze-and-excitation optimization:

  Depthwise Separable Convolutions: This technique separates the convolution operation into two parts: depthwise convolution, which applies a single filter per input channel, and pointwise convolution, which combines these filtered channels. This approach significantly reduces the number of parameters and computational cost.

  Squeeze-and-Excitation: This mechanism allows the network to perform dynamic channel-wise feature recalibration. It enhances the model's ability to capture important features by adaptively reweighting the channel outputs based on their importance.

- **Swish Activation Function**: The model uses the Swish activation function, defined as $f(x) = x \cdot \sigma(x)$, where $\sigma(x)$ is the sigmoid function. Swish is a smooth, non-monotonic function that has been shown to improve model performance compared to the traditional ReLU activation function.

- **Global Average Pooling:** After the series of convolutional and MBConv layers, global average pooling is applied. This layer reduces the spatial dimensions of the feature maps to a single value per channel by computing the average of each feature map. This step effectively summarizes the feature maps and reduces the risk of overfitting.

- **Fully Connected Layers:** The final layers of EfficientNetB7 are fully connected layers that integrate the extracted features from the convolutional part of the network. These layers are responsible for performing the final classification or regression tasks. In the context of VHS-based canine cardiomegaly detection, these layers predict the VHS value based on the identified anatomical points.

- **Output Layer:** The output layer of EfficientNetB7 is designed to provide a single scalar value representing the VHS. This value is crucial for diagnosing cardiomegaly in canines. The model's ability to accurately predict this value is a testament to its effectiveness in extracting relevant features from the input images.

3.3. Training Process

3.3.1 Loss Function

The model's performance is optimized using the Mean Squared Error (MSE) loss function. This loss function is particularly suitable for regression tasks, as it measures the average of the squares of the errors between the predicted values and the actual values. By minimizing the MSE loss, the model learns to improve its predictions, making it an effective choice for our VHS-based canine cardiomegaly detection task.

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

where n is the number of points, which is 6 in this case, x_i and y_i are the actual coordinates of the points, $\hat{x}_i$ and $\hat{y}_i$ are the predicted coordinates of the points.

3.3.2 Optimization

To train the model, we use the Adam optimizer. Adam, which stands for Adaptive Moment Estimation, combines the benefits of both the AdaGrad and RMSProp algorithms to provide an optimization algorithm that can handle sparse gradients on noisy problems. The learning rate of 1×10^{-4} was chosen to ensure stable convergence and effective learning. Adam adjusts the learning rate for each parameter individually, making it particularly well-suited for our deep learning model, as it helps to accelerate the convergence and enhance performance. During training, the model converges over 200 epochs, with a batch size of 4.

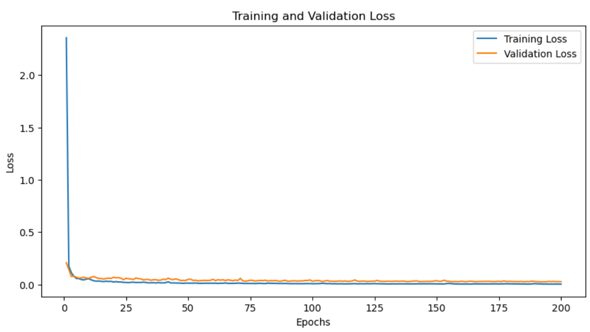

Figure 4. Training and Validation Loss Over Number of Epochs

3.3.3 VHS calculation

The Vertebral Heart Score (VHS) is a widely used metric in veterinary medicine for assessing heart size in dogs. It is calculated based on the dimensions of the heart in relation to the vertebrae on a lateral thoracic radiograph. In this study, the VHS is calculated using the coordinates of six key anatomical points (A, B, C, D, E, and F) identified by our EfficientNetB7 model. The VHS is computed as follows:

$$\mathrm{VHS} = 6 \times \left(\frac{\mathrm{AB} + \mathrm{CD}}{\mathrm{EF}} \right) \tag{2}$$

where AB is the distance between points A and B, representing the length of the heart, CD is the distance between points C and D representing the width of the heart and EF is the length of the vertebrae covering the heart. The formula accounts for the heart's length and width relative to the vertebral column, providing a standardized measure of heart size. Once the VHS is calculated, the labels for the heart size are generated based on the VHS values:

- **Small**: VHS < 8.2
- **Normal**: $8.2 \leq$ VHS ≤ 10
- **Large**: VHS > 10

This classification is crucial for assessing whether a dog has a normal heart size or is potentially suffering from cardiomegaly.

4. Results

In this section, we present the empirical results of our experiments and provide an in-depth discussion of the implications of our findings. The performance of our EfficientNetB7 model in predicting the Vertebral Heart Score (VHS) and classifying canine heart size is thoroughly evaluated using various metrics. These results demonstrate the effectiveness of our approach in detecting cardiomegaly in dogs.

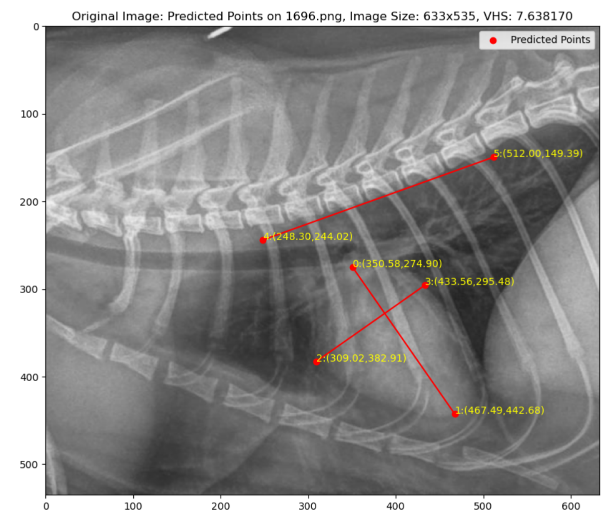

Figure 5. Predicted Image

Table 1. Model Performance

Model	MSE	MAE	MAPE	Accuracy
EfficientNetB7	0.20282	0.33156	3.4833%	85.25%

5. Discussion

In this study, we have demonstrated the effectiveness of the EfficientNetB7 model in detecting cardiomegaly in canine chest X-ray images, achieving a notable accuracy of 85.25%. These results highlight not only the model's accuracy but also its robustness in classifying heart sizes, which is critical in diagnosing and monitoring cardiac conditions in dogs.

5.1. Implications of Findings

The integration of our deep learning model into veterinary diagnostics has significant potential to transform the way cardiomegaly is detected in dogs. Accurate and early detection of this condition can lead to better management and treatment outcomes, ultimately improving the quality of life for affected animals. By automating the VHS calculation and heart size classification, our model can reduce the workload on veterinary professionals and minimize the risk of human error, making the diagnostic process more efficient and reliable.

5.2. Comparison with Existing Literature

Our model's performance aligns well with existing studies, such as the work by Li and Zhang [4], which utilized the Regressive Vision Transformer (RVT) for similar tasks. While RVT achieved a high accuracy of 85%, our use of EfficientNetB7 offers a competitive alternative with benefits in computational efficiency and scalability. Unlike general models such as VGG16 [6], our tailored architecture for canine cardiomegaly detection emphasizes the advantages of specialized models in achieving higher precision and reliability.

5.3. Limitations and Future Work

Despite the promising results, several limitations must be addressed in future research. Firstly, our study was conducted on a specific dataset, and expanding this dataset to include more diverse samples from various breeds, ages, and health conditions would enhance the model's generalization. Secondly, while our model processes images resized to 512x512 pixels, investigating the impact of higher resolution images could further improve accuracy. Moreover, implementing the model in real-time diagnostic tools and integrating them into clinical workflows will be crucial for practical applications. Collaborations with veterinary clinics to validate and refine the model in real-world settings will ensure its efficacy and reliability. Continuous updates and iterations based on new data and technological advancements will keep the model state-of-the-art and widely applicable. Finally, extending this research to include human cardiomegaly detection using different diagnostic technologies could provide valuable insights and applications in human medicine. By addressing these limitations and exploring new avenues, future work can build on our findings to create even more effective and versatile diagnostic tools.

6. Conclusion

In this study, we have demonstrated the efficacy of the EfficientNetB7 model for detecting cardiomegaly in canine chest X-ray images, achieving a notable accuracy of 85.25%. This deep learning model, specifically optimized for this task, shows significant potential in enhancing veterinary diagnostics. By enabling more accurate and timely diagnoses, this model can significantly improve treatment outcomes for dogs with cardiac conditions. Despite these promising results, future research should address the identified limitations, such as dataset diversity and image resolution constraints.

Expanding the dataset to include more varied samples, refining the model architecture to handle higher resolution images, and integrating the model with other diagnostic tools will be critical steps forward. These improvements will enhance the model's generalization and performance in real-world clinical settings.

Furthermore, validating the model in clinical environments and continuously updating it based on new data and technological advancements will ensure its ongoing relevance and effectiveness. This research paves the way for more advanced applications of deep learning in veterinary medicine, ultimately contributing to improved health outcomes for animals.

By addressing these areas in future research, we can further enhance the capabilities of deep learning models in veterinary diagnostics, ensuring that they remain at the forefront of technological advancements and continue to provide significant benefits in animal health care.

References

[1] 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

[2] 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

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

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

[5] 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

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

[7] 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

[8] 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