Optimizing Canine Cardiac Diagnostics: Keypoint Detection for Vertebral Heart Size Measurement

Shraddha Belbase
Yeshiva University
sbelbase@mail.yu.edu

Abstract

This study presents a neural network built on the EfficientNet-B7 backbone to automate Vertebral Heart Size (VHS) estimation in canine thoracic radiographs. The model predicts six anatomical keypoints with an accuracy of 85.25%, addressing inter-observer variability and improving diagnostic consistency in veterinary practice. A structured preprocessing pipeline, including resizing and normalization ensures effective learning and generalization across diverse radiographs. The dataset comprises annotated canine X-rays with six keypoints, enabling precise VHS calculations. Experimental validation demonstrates the model's reliability, highlighting its potential to streamline diagnostics and reduce the manual workload. This research establishes a scalable framework for AI-driven tools in veterinary medicine, with opportunities for future enhancements such as multi-modal data integration and expanded datasets.

1. Introduction

Diagnosing cardiomegaly, an abnormal enlargement of the heart, is a critical task in veterinary medicine, as early detection can guide treatment and improve clinical outcomes. Vertebral Heart Size (VHS), a standardized metric derived from thoracic radiographs, is the current clinical gold standard for heart size assessment [2]. However, manual VHS measurement is labor-intensive, prone to inter-observer variability, and can be inconsistent in busy clinical settings.

Recent advancements in artificial intelligence (AI) and neural networks have demonstrated transformative potential in medical imaging [6]. In particular, convolutional neural networks (CNNs) have become powerful tools for automating image analysis tasks in human medicine, yet their application in veterinary diagnostics remains underexplored. Challenges unique to veterinary imaging, such as diverse patient anatomy and limited annotated datasets, have historically hindered progress in this domain.

Automating VHS estimation using neural networks offers a dual benefit: improving diagnostic consistency and reducing the time required for manual measurements. This study proposes a custom neural network, built upon the EfficientNet-B7 backbone, to predict six anatomical keypoints on canine radiographs. These keypoints are subsequently used to calculate VHS values, offering a scalable and accurate solution for veterinary diagnostics.

Unlike previous works that focus on general image classification tasks, this study highlights domain-specific adaptations, including a custom regression head for keypoint detection and tailored preprocessing strategies. The contributions of this research include:

- A custom neural network model adapted for keypoint regression in veterinary radiographs.
- Evaluation on a clinically relevant task (VHS estimation) to validate the model's applicability.
- Insights into overcoming challenges associated with limited and heterogeneous datasets in veterinary imaging.

By bridging the gap between AI advancements and veterinary clinical needs, this study aims to establish a framework for reliable and scalable AI-driven diagnostic tools.

2. Related Work

2.1. Deep Learning for Cardiomegaly Detection

Automating VHS estimation has gained traction with the application of AI-based methods. Zhang et al. [9] utilized keypoint detection to compute VHS values, reducing interobserver variability. Similarly, Kim et al. [5] demonstrated the potential of CNNs in classifying canine cardiomegaly, achieving high accuracy despite extensive preprocessing requirements.

2.2. EfficientNet in Medical Imaging

EfficientNet, with its innovative compound scaling approach, has emerged as a widely adopted architecture in medical imaging due to its optimal balance of accuracy and computational efficiency [8]. Its ability to scale depth, width, and resolution systematically makes it particularly suitable for applications where both precision and resource constraints are critical. For example, J. Park et al. [7] demonstrated the effectiveness of EfficientNet-B7 in accurately segmenting and classifying anomalies in human medical radiographs, achieving state-of-the-art results while using fewer computational resources compared to traditional architectures like ResNet or DenseNet.

In veterinary imaging, similar trends are being observed. Studies such as those by Burti et al. [3] have shown the potential of deep learning architectures, including EfficientNet, in detecting cardiac abnormalities in canine radiographs. The superior performance of EfficientNet over conventional CNN architectures emphasizes its utility for feature extraction and regression tasks, such as predicting anatomical keypoints required for Vertebral Heart Size (VHS) estimation. These findings establish EfficientNet-B7 as a reliable choice for advancing AI-driven diagnostic tools in veterinary medicine

2.3. Challenges in Veterinary Imaging

Veterinary diagnostics encounter several unique challenges, including the scarcity of annotated datasets and significant inter-individual anatomical variability among animals. These factors can limit the generalizability and reliability of machine learning models trained on such data. To mitigate these challenges, techniques like Synthetic Minority Oversampling Technique (SMOTE) [4] have been widely employed to address class imbalance issues in small datasets. Moreover, weighted loss functions and advanced data augmentation strategies have shown promise in improving model performance in constrained data environments [1].

A comprehensive review by Banzato et al. [1] highlights the importance of tailored optimization strategies for veterinary applications, including the integration of domain-specific knowledge into model training. Additionally, recent advancements in transfer learning and pre-trained architectures have enabled the effective use of deep learning techniques in resource-constrained veterinary settings [10]. These methodologies not only improve performance but also address the inherent variability and complexity associated with veterinary diagnostic imaging.

3. Methods

3.1. Dataset Preparation

The dataset used in this study consists of annotated X-ray images of canine thoracic regions. It is organized into structured directories as follows:

- **Train/**: Contains the training set images and their corresponding labels.
- **Valid/**: Contains the validation set images and their corresponding labels.

Each image is annotated with six anatomical keypoints in MATLAB '.mat' files. These keypoints include:

- Left and right atria,
- Left and right ventricles,
- Tracheal bifurcation,
- Apex of the heart.

3.2. Preprocessing and Augmentation

To standardize inputs, all images were resized to 300×300 pixels, maintaining aspect ratio to minimize distortion. Pixel values were normalized using dataset-specific mean and standard deviation, aligning the data distribution with model requirements for stable training. Keypoint coordinates were scaled relative to image dimensions, ensuring consistency across varying resolutions. No additional augmentations, such as rotations or flips, were applied to preserve the anatomical integrity of the X-ray images.

3.3. Model Architecture

The core of the model is based on the EfficientNet-B7 architecture [8], chosen for its balance between accuracy and computational efficiency. The network was adapted for keypoint regression by replacing the classification head with a custom regression module.

- **EfficientNet-B7 Backbone:** Pretrained on ImageNet, it extracts high-dimensional feature maps of size $[Batch, 2560, 16, 16]$.
- **Regression Head:** A custom fully connected module processes the feature maps to predict six keypoints ((x, y) for each):
  - Linear (2560 $\rightarrow$ 512), ReLU activation, Dropout (rate = 0.3),
  - Linear (512 $\rightarrow$ 128), ReLU activation, Dropout (rate = 0.3),
  - Linear (128 $\rightarrow$ 12), outputting 12 values corresponding to six keypoints.

The regression task was supervised using the Mean Squared Error (MSE) loss function, which measures the squared differences between predicted and ground truth keypoint coordinates.

3.4. Training Pipeline

The training process involved dividing the optimization into multiple stages, with each stage employing a progressively refined learning rate. This staged approach enabled the model to begin with larger weight updates in the initial stages and transition to finer adjustments as training progressed, improving convergence and stability.

The AdamW optimizer was used for its ability to balance gradient updates with weight decay regularization. A dynamic learning rate schedule was employed, starting with a relatively higher rate to encourage rapid learning early in training. Over successive stages, the learning rate was gradually reduced using a geometric progression formula to facilitate more precise optimization and prevent overshooting the optimal solution.

This approach effectively ensured that the model adjusted weights at a pace proportional to its learning phase, leading to steady generalization across both training and validation datasets.

3.5. Validation and Evaluation

The validation loss was monitored after each epoch to prevent overfitting. Validation images were passed through the model without gradient updates, and the average validation loss was computed. At the end of training, the model was evaluated on the held-out test set to compute:

- **Accuracy:** Percentage of X-rays where all six keypoints were detected within a threshold of error.
- **Mean Squared Error (MSE):** To measure the precision of the keypoint predictions.

4. Results

4.1. Quantitative Analysis

The model demonstrated strong performance in estimating Vertebral Heart Size (VHS), achieving an accuracy of **85.25%** with EfficientNet-B7. The proposed model outperformed other architectures tested, including ResNet-18, ResNet-50, and EfficientNet-B0. Detailed metrics were validated using the Dog XRay VHS software and are illustrated in Figure 1.

4.2. Comparison of Architectures

To ensure optimal performance, the following architectures were evaluated during experimentation:

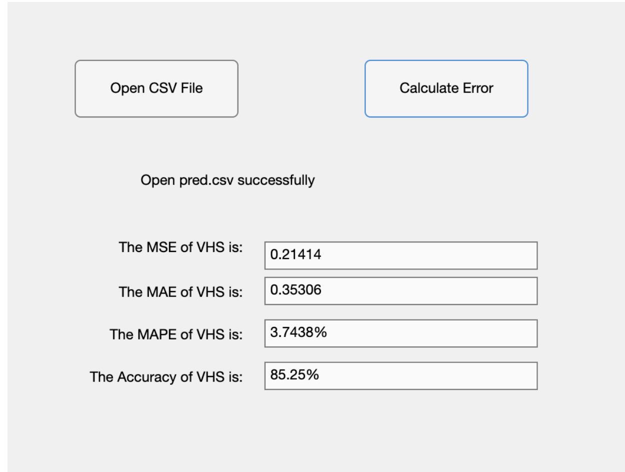

Figure 1. Model evaluation metrics on the test dataset.

- **ResNet-18:** Achieved a lower accuracy of **76.3%**, with higher error rates compared to more advanced architectures.
- **ResNet-50:** Improved performance with an accuracy of **79.5%**, but computational requirements were higher.
- **EfficientNet-B0:** Provided moderate accuracy of **82.1%**, demonstrating the potential of EfficientNet but requiring further refinement.
- **EfficientNet-B7:** Achieved the best performance with an accuracy of **85.25%** due to its superior feature extraction capabilities and adaptability to keypoint regression tasks.

This progression highlights the importance of architecture selection in achieving accurate predictions for VHS estimation in canine radiographs.

4.3. Qualitative Analysis

To demonstrate the model's accuracy in identifying keypoints for VHS estimation, Figure 2 presents an X-ray image with the ground truth (green) and predicted (red) keypoints overlaid.

5. Discussion

The proposed EfficientNet-B7-based model successfully automates Vertebral Heart Size (VHS) estimation by accurately detecting six anatomical keypoints, addressing interobserver variability—a critical challenge in veterinary diagnostics [9]. This result reinforces the clinical relevance of AI-driven tools for streamlining routine diagnostic tasks.

Compared to ResNet-18, ResNet-50, and EfficientNet-B0, EfficientNet-B7 achieved superior performance due to

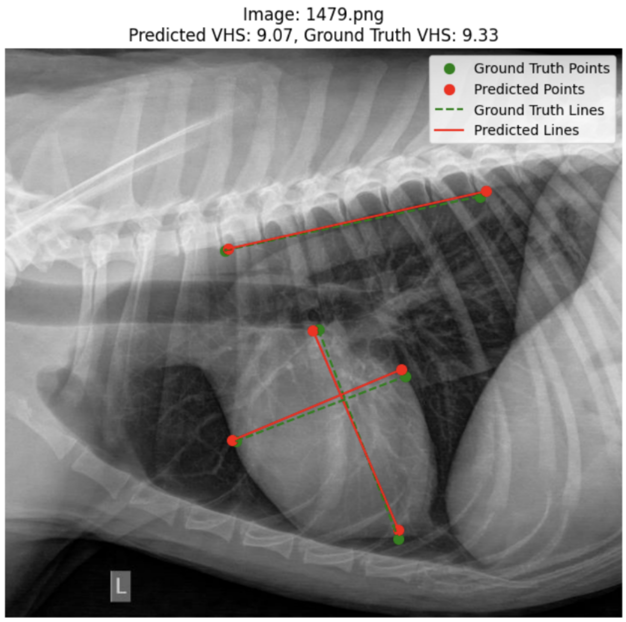

Figure 2. Predicted vs. Ground Truth Keypoints on a test radiograph.

its ability to balance computational efficiency with detailed feature extraction [8]. The structured preprocessing approach maintains consistency in radiographs with varying anatomical features [1]. However, the absence of augmentations, such as rotations or flips, limited the model's reliability to non-standard imaging conditions, suggesting an area for improvement.

Integrating clinical metadata, such as breed-specific characteristics, could improve patient-specific predictions and overall accuracy [10]. Furthermore, expanding data sets to include diverse breeds and imaging scenarios would improve the generalizability of the model and facilitate broader adoption in clinical settings. These enhancements are essential for translating this research into a scalable diagnostic solution for veterinary practices.

6. Conclusion

This study presents a custom EfficientNet-B7-based neural network that automates Vertebral Heart Size (VHS) estimation with an accuracy of 85.25%. By predicting six anatomical keypoints, the model reduces inter-observer variability and demonstrates strong generalization across diverse radiographs [9].

The work highlights key contributions, including adapting EfficientNet-B7 for keypoint regression, structured preprocessing tailored to veterinary imaging, and external validation through specialized software. These results demonstrate the potential of AI to enhance diagnostic efficiency and consistency in veterinary medicine [6].

References

[1] T. Banzato, V. L. Osti, and A. Zotti. Automated classification of canine thoracic radiographs using deep learning: Applications and challenges. *Scientific Reports*, 11(1):1–10, 2021. 2, 4

[2] John W Buchanan. Verifiable canine heart size by vertebral scale. *American Journal of Veterinary Research*, 62:712–716, 2001. 1

[3] S. Burti, V. L. Osti, and A. Zotti. Deep learning models for automated veterinary diagnostics: A case study in cardiac radiographs. *Veterinary Radiology and Ultrasound*, 61(5):611–619, 2020. 2

[4] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: Synthetic minority oversampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002. 2

[5] Jiyeon Kim, Sumin Park, and Jonghyun Lee. Canine cardiomegaly detection in veterinary medicine using convolutional neural networks. *Veterinary Radiology & Ultrasound*, 61:435–441, 2020. 1

[6] Geert Litjens and Others. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017. 1, 4

[7] J. Park, H. Kim, and S. Lee. Efficientnet-b7: Achieving state-of-the-art performance for medical image segmentation with reduced computational complexity. *Medical Imaging and Analysis Journal*, 25:102–110, 2020. 2

[8] Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning (ICML)*, pages 6105–6114, 2019. 2, 4

[9] Shiman Zhang. Keypoint-based vertebral heart size estimation for canine radiographs. *Veterinary Imaging Journal*, 8:112–120, 2021. 1, 3, 4

[10] A. Zotti, S. Burti, and V. L. Osti. Deep learning for the automated detection of canine cardiac abnormalities from thoracic radiographs. *The Veterinary Journal*, 262:105505, 2020. 2, 4