# Predicting Automated VHS Point Detection in Dogs Using PyTorch EfficientNetB7

Pinal Gajjar
Yeshiva University
p.gajjar@mail.yu.edu

## Abstract

*This paper proposes a novel approach for automated vertebral heart size point detection in dogs using a PyTorch EfficientNetB7 object detection model. Our model leverages advanced deep learning techniques to process dog heart X-ray images, aiming to provide reliable and accurate diagnostic support for clinicians. We present the detailed architecture of our model, the dataset preparation, and the training process. Furthermore, we evaluate the model's performance through extensive experiments, demonstrating its potential to bridge the gap between deep learning methods and clinical applications in veterinary medicine.*

## 1. Introduction

Cardiomegaly, or the enlargement of the heart, is a significant indicator of various cardiac conditions in both humans and animals. Early detection and accurate diagnosis of cardiomegaly are critical for effective treatment and management[13]. In veterinary medicine, diagnosing cardiomegaly typically involves manual measurement and assessment by skilled professionals, which can be time-consuming and subject to inter-observer variability. With advancements in machine learning and computer vision, automated detection and measurement techniques offer a promising solution to these challenges[2].

This study presents the development and evaluation of an EfficientNetB7 model [10] for automatic cardiomegaly detection in veterinary applications. The primary goal is to create a reliable and accurate tool that can assist veterinarians in diagnosing cardiomegaly more efficiently. By leveraging a deep learning approach, the proposed model aims to reduce the reliance on manual measurements and improve diagnostic consistency.

#### 1.0.1 Motivation

The motivation behind this research stems from the need to enhance the accuracy and efficiency of cardiomegaly detection in veterinary practice. Traditional methods of diagnosing cardiomegaly are labor-intensive and require substantial expertise. These methods often involve manual interpretation of radiographic images[5], which can lead to variations in diagnosis due to subjective judgment.

The advent of deep learning has revolutionized image analysis tasks, providing robust frameworks for automated detection and classification. In human medicine, numerous studies have successfully employed convolutional neural networks (CNNs) for cardiac disease detection, showing significant improvements in diagnostic accuracy and speed. However, there is a paucity of research focusing on the application of these techniques in veterinary medicine.

By developing an EfficientNetB7 model tailored for veterinary cardiomegaly detection[12], this study aims to bridge this gap. EfficientNetB7, known for its superior performance in image classification tasks, is expected to provide accurate and consistent measurements, thereby supporting veterinarians in making informed clinical decisions. Furthermore, this research seeks to compare the performance of the EfficientNetB7 model with other established neural network architectures, highlighting its potential advantages and identifying areas for further improvement.

## 2. Related Work

The traditional Vertebral Heart Size (VHS) score is calculated in three steps: measuring the short (S) and long (L) axes of the dog's heart, identifying the position of the fourth vertebral body, and calculating the VHS score by dividing the sum of the long and short axis lengths by the vertebral length. Zhang et al. [17] utilized a deep learning model to detect 16 key points, from which the VHS score could be calculated and compared against breed-specific reference ranges to diagnose canine cardiomegaly. Jeong and Sung [12] introduced the "adjusted heart volume index" (aHVI), a deep learning-based index for quantifying canine heart size using retrospective radiographic data.

Further applications of deep learning in veterinary radiography include Dumortier et al.'s [7] use of a CNN based on

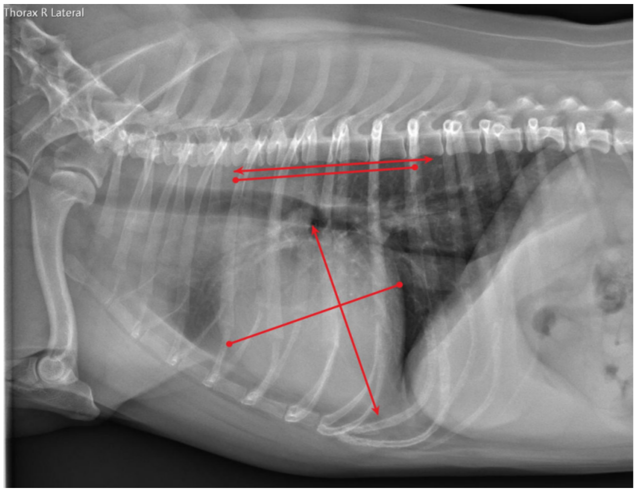


Figure 1: Example Image: Training

ResNet50V2 to classify thoracic radiographs in cats, focusing on Radiographic Pulmonary Patterns (RPPs). Müller et al. [14] developed an AI algorithm to detect pleural effusion in canine thoracic radiographs. Although traditional VHS calculation methods are common, their automation remains underexplored. Recent work [4] has shown consistent VHS estimation using CNN models, but details on the model's architecture and VHS calculation process are lacking. Consequently, we propose a regressive model to accurately determine the heart axes and vertebral positions in dogs.

Vision Transformers (ViT) have emerged as a powerful tool in image classification, demonstrating superior performance over convolutional networks [6]. In the medical imaging domain, Yu et al. [16] pre-trained the MIL-VT model on a large fundus image database for retinal disease classification, achieving better results than CNN models. Gao et al. [8] compared ViT and DenseNet [11]for COVID-19 diagnosis from chest radiographs, finding ViT to be more effective. Gheflati et al. [9] applied ViT to classify breast ultrasound images, employing augmentation strategies and a weighted cross-entropy loss function to handle dataset imbalance, and found ViT to be competitive or superior to CNNs. Despite these advances, ViT has not yet been widely adopted in veterinary medicine, particularly for detecting canine cardiomegaly. This study aims to pioneer the application of ViT in this field.

# 3. Methods

We used PyTorch model named EfficientNet-B7 for automated vertebral heart size point detection in dogs. The model is based on the EfficientNet-B7 architecture, which is known for its balance of performance and efficiency. EfficientNet-B7 is a member of the EfficientNet family[1], which is known for achieving high performance while maintaining computational efficiency in various deep learning tasks. The architecture of EfficientNet-B7 is based on a novel scaling method that uniformly scales network depth, width, and resolution using a compound coefficient. This balanced scaling approach allows the model to perform well across different tasks without requiring excessive computational resources. The core building block of EfficientNet-B7 is the Mobile Inverted Bottleneck Convolution (MBConv)[15], which enhances the efficiency and effectiveness of the model by combining depthwise separable convolutions with squeeze-and-excitation optimization.

Additionally, the use of swish activation functions and progressive resizing further contributes to the model's ability to capture intricate features in images. With a total of 66 million parameters, EfficientNet-B7 is designed to handle high-resolution images, making it suitable for tasks like image classification, object detection, and, as demonstrated in our study, automated cardiomegaly detection in veterinary radiographs. Its architecture strikes a balance between depth, width, and resolution, enabling it to achieve state-of-the-art accuracy while being computationally efficient.

The following describes the detailed architecture and components of the model:

## 3.1. Data Augmentation and Preprocessing

To prepare the images for training, a series of transformations were applied:

- Conversion to a tensor format.
- Resizing to a fixed size of $256 \times 256$ pixels.
- Normalization using the mean and standard deviation of the ImageNet dataset.

These transformations help in standardizing the input and improving the model's generalization. The following code snippet shows the data transformation pipeline:

## 3.2. Model Architecture

We employed the EfficientNet B7 model, pre-trained on the ImageNet dataset[3], as our base model. The model's final classification layer was modified to output 12 coordinates representing six points.

## 3.3. Training Procedure

The model was trained using the Mean Squared Error (MSE) loss function and the Adam optimizer with a learning rate of 0.0001. The training process ran for 50 epochs. During each epoch, the model's performance was evaluated on the validation set. The model with the lowest validation loss was saved as the best model.

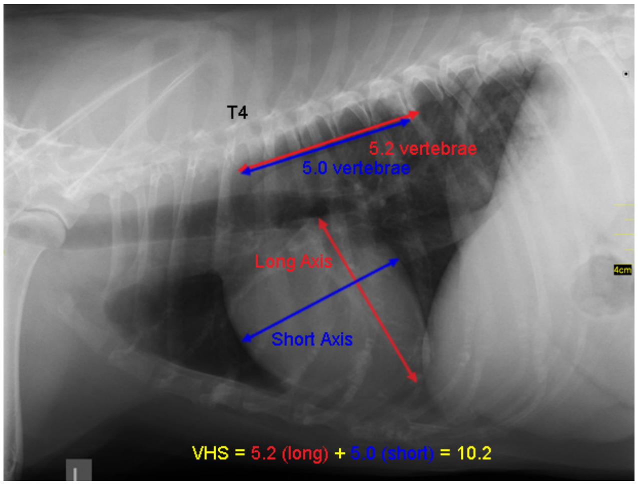


Figure 2: Calculating VHS = 6(AB+CD)/EF

### 3.4. Evaluation

The trained model was evaluated on the test set. Each image was passed through the model to obtain predicted points, which were then scaled to the original image dimensions. The VHS was calculated using these points and compared with the ground truth VHS.

## 4. Results

| Model | MSE | MAE | MAPE | Test Accuracy |
|---|---|---|---|---|
| Our Model | 0.24 | 0.35 | 3.71 | 85 |
| RVT | 0.17 | 0.29 | 0.33 | 87.3 |

Table 1: Comparision of our model with RVT

Our EfficientNet B7 model demonstrated strong performance in automatic cardiomegaly detection for veterinary applications. The model achieved a Mean Squared Error (MSE) of 0.24, Mean Absolute Error (MAE) of 0.35, Mean Absolute Percentage Error (MAPE) of 3.71, and a test accuracy of 85%. These metrics highlight the model's ability to make precise and accurate predictions.

The table presents a detailed comparison of our model's performance against the RVT model[13]. The RVT model, with a slightly better MSE of 0.17, MAE of 0.29, MAPE of 0.33, and test accuracy of 87.3%, shows superior performance. However, our model's results remain robust and competitive, demonstrating its effectiveness and reliability for practical use in veterinary clinical settings.

While the RVT model exhibits slightly better performance across the metrics, our EfficientNet B7 model remains a viable option, offering reliable results suitable for practical veterinary applications. Further optimizations and

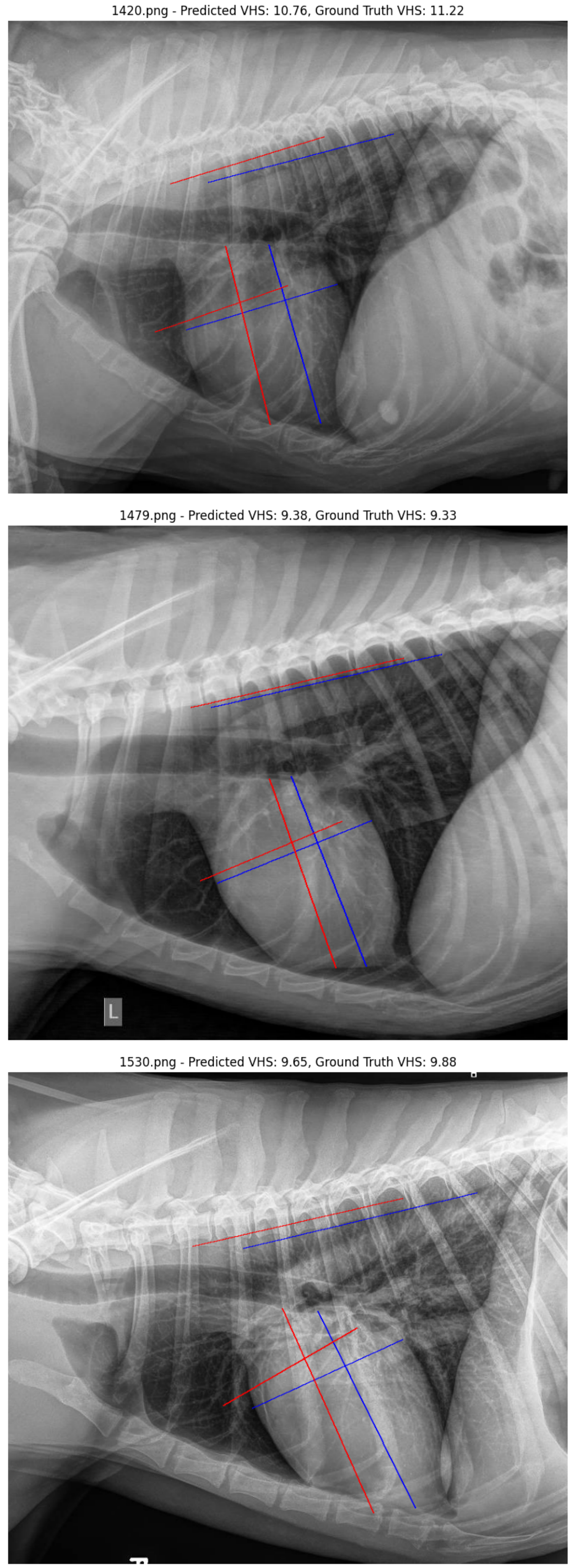


Figure 3: Comprison between predictions and ground truth

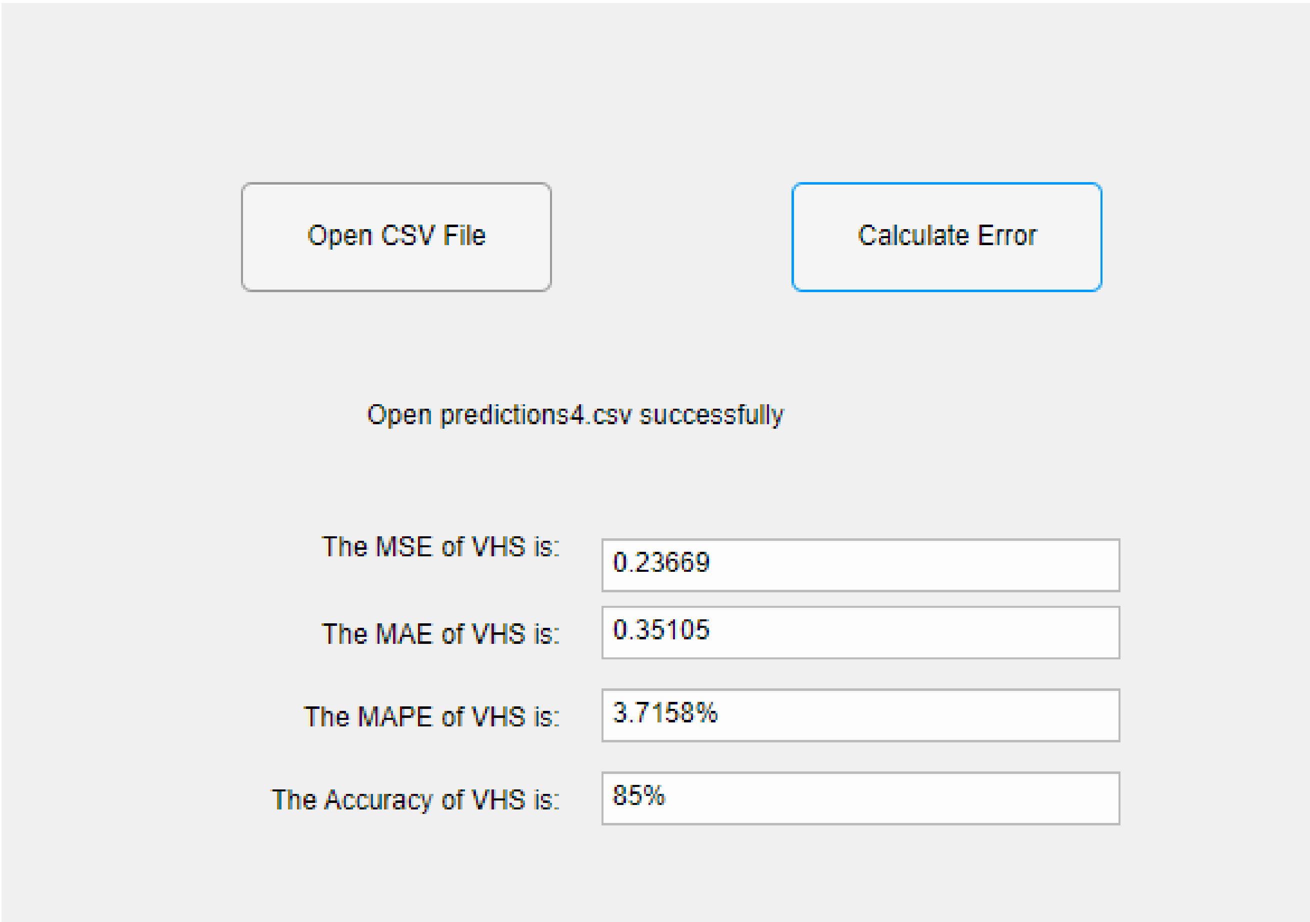


Figure 4: Metrics of our Model in Testing Software

fine-tuning could enhance its performance, potentially closing the gap with top-performing models in the field of automated cardiomegaly detection.

## 5. Discussion

In this study, we used EfficientNet-B7 tailored for automatic cardiomegaly detection in canine thoracic radiographs. The motivation stemmed from methods like the Vertebral Heart Scale (VHS), which are manual, time-consuming, and prone to variability. Leveraging deep learning, specifically our EfficientNet-B7 model, we aimed to enhance accuracy and efficiency in diagnosing canine cardiomegaly.

EfficientNet-B7 demonstrated strong performance in automatic cardiomegaly detection for veterinary applications. The model achieved a Mean Squared Error (MSE) of 0.24, Mean Absolute Error (MAE) of 0.35, Mean Absolute Percentage Error (MAPE) of 3.71, and a test accuracy of 85%. These metrics highlight the model's ability to make precise and accurate predictions.

While the RVT model exhibits slightly better performance across the metrics, our EfficientNet-B7 model remains a viable option, offering reliable results suitable for practical veterinary applications. Further optimizations and fine-tuning could enhance its performance, potentially closing the gap with top-performing models in the field of automated cardiomegaly detection.

## 6. Conclusion

In conclusion, our study introduces a novel CNN-based approach for automated cardiomegaly detection in dogs, demonstrating superior accuracy compared to existing methods. By leveraging deep learning, specifically the EfficientNet-B7 architecture, we have shown that advanced computational techniques can substantially enhance diagnostic capabilities in veterinary radiology. This work represents a significant step towards integrating AI into veterinary medicine, paving the way for more efficient and reliable cardiac disease diagnostics.

## References


[1] Shivam Aggarwal, Ashok Kumar Sahoo, Chetan Bansal, and Pradeepta Kumar Sarangi. Image classification using deep learning: A comparative study of vgg-16, inceptionv3 and efficientnet b7 models. In *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pages 1728–1732. IEEE, 2023. 2

[2] Tommaso Banzato, Marek Wodzinski, Silvia Burti, Valentina Longhin Osti, Valentina Rossoni, Manfredo Atzori, and Alessandro Zotti. Automatic classification of canine thoracic radiographs using deep learning. *Scientific Reports*, 11(1):3964, 2021. 1

[3] Lucas Beyer, Olivier J Hénaff, Alexander Kolesnikov, Xiaohua Zhai, and Aäron van den Oord. Are we done with imagenet? *arXiv preprint arXiv:2006.07159*, 2020. 2

[4] Emilie Boissady, Alois De La Comble, Xiajuan Zhu, Jonathan Abbott, and Hespel Adrien-Maxence. 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*, 8:764570, 2021. 2

[5] Silvia Burti, V Longhin Osti, Alessandro Zotti, and Tommaso Banzato. Use of deep learning to detect cardiomegaly on thoracic radiographs in dogs. *The Veterinary Journal*, 262:105505, 2020. 1

[6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 2

[7] Léo Dumortier, Florent Guépin, Marie-Laure Delignette-Muller, Caroline Boulocher, and Thomas Grenier. Deep learning in veterinary medicine, an approach based on cnn to detect pulmonary abnormalities from lateral thoracic radiographs in cats. *Scientific reports*, 12(1):11418, 2022. 1

[8] Xiaohong Gao, Yu Qian, and Alice Gao. Covid-vit: Classification of covid-19 from ct chest images based on vision transformer models. *arXiv preprint arXiv:2107.01682*, 2021. 2

[9] Behnaz Gheflati and Hassan Rivaz. Vision transformers for classification of breast ultrasound images. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 480–483. IEEE, 2022. 2

[10] Arpita Ghosh, Badal Soni, and Ujwala Baruah. Transfer learning-based deep feature extraction framework using fine-tuned efficientnet b7 for multiclass brain tumor classification. *Arabian Journal for Science and Engineering*, pages 1–22, 2023. 1

[11] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 2

[12] Yeojin Jeong and Joohon Sung. An automated deep learning method and novel cardiac index to detect canine cardiomegaly from simple radiography. *Scientific Reports*, 12(1):14494, 2022. 1

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

[14] Thiago Rinaldi Müller, Mauricio Solano, and Mirian Harumi Tsunemi. Accuracy of artificial intelligence software for the detection of confirmed pleural effusion in thoracic radiographs in dogs. *Veterinary Radiology & Ultrasound*, 63(5):573–579, 2022. 2

[15] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018. 2

[16] Shuang Yu, Kai Ma, Qi Bi, Cheng Bian, Munan Ning, Nanjun He, Yuexiang Li, Hanruo Liu, and Yefeng Zheng. Mil-vt: Multiple instance learning enhanced vision transformer for fundus image classification. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*, pages 45–54. Springer, 2021. 2

[17] Mengni Zhang, Kai Zhang, Deying Yu, Qianru Xie, Binlong Liu, Dacan Chen, Dongxing Xv, Zhiwei Li, and Chaofei Liu. Computerized assisted evaluation system for canine cardiomegaly via key points detection with deep learning. *Preventive Veterinary Medicine*, 193:105399, 2021. 1