Deep Learning-Based Assessment of Canine Cardiomegaly via Vertebral Heart Scale Prediction

Sri Sai Lalitha Mallika Yeturi
Yeshiva University
syeturi@mail.yu.edu

Abstract

Cardiomegaly, or the enlargement of the heart, is a critical condition in dogs that often serves as an indicator of underlying cardiac diseases. Early detection and accurate classification are crucial for timely intervention. This paper introduces a deep learning model based on the EfficientNet-B7 architecture to automate cardiomegaly detection from canine X-ray images. The model leverages pre-trained weights from ImageNet and fine-tunes them for Vertebral Heart Scale (VHS) calculation. Achieving an accuracy of 86%, this study demonstrates the potential of advanced AI models in enhancing veterinary diagnostics and improving animal health outcomes.

1. Introduction

Cardiomegaly represents a significant health concern in veterinary medicine, particularly among canine populations. It is characterized by an abnormal enlargement of the heart, which can indicate underlying conditions such as valvular disease, cardiomyopathy, or pericardial effusion. Traditional diagnostic methods rely on manual interpretation of X-rays, which can vary due to the expertise of the veterinarian and the quality of the imaging.

With advancements in machine learning and computer vision, deep learning models have emerged as promising tools for medical diagnostics. These models can analyze complex patterns in imaging data with high precision and consistency, reducing the dependency on manual interpretations. This study employs the EfficientNet-B7 architecture, known for its computational efficiency and accuracy, to develop a robust tool for diagnosing canine cardiomegaly. This work builds on previous research by integrating advanced techniques and tailoring them for veterinary applications [18, 6, 10].

2. Related Work

Deep learning has revolutionized the field of medical imaging. In human medicine, convolutional neural networks (CNNs) have demonstrated exceptional performance in diagnosing conditions from radiographs, CT scans, and MRIs. Veterinary applications, while less explored, are gaining traction.

Zhang et al. introduced a model to calculate the Vertebral Heart Scale (VHS) using deep learning, achieving high accuracy [18]. Jeong and Sung proposed an adjusted cardiac index based on radiographic data, further advancing the field of automated diagnosis [6]. Li and Zhang recently developed a Regressive Vision Transformer (RVT) for dog cardiomegaly assessment, setting a benchmark with its superior accuracy of 87.3% [10].

3. Methods

This section provides a comprehensive description of the methods employed to develop a deep learning model for detecting cardiomegaly in dogs using thoracic X-ray images. The methods include data preprocessing, model architecture modification, training strategy, and evaluation metrics.

3.1. Data Preprocessing

The dataset used in this study consisted of 2,000 thoracic X-ray images, categorized into three classes—*Small*, *Normal*, and *Large*—based on Vertebral Heart Scale (VHS) values. Each image was labeled with six key points, corresponding to anatomical landmarks that allowed the calculation of the VHS. These points help quantify the enlargement of the heart relative to the vertebrae, providing a diagnostic metric for cardiomegaly.

Data preprocessing played a crucial role in enhancing model performance and robustness:

- **Resizing:** All images were resized to 300×300 pixels. The resizing ensured that the input size was consistent for the EfficientNet-B7 model, as deep learning models typically require fixed input dimensions to maintain the structure of the neural network [16]. This uniformity reduced computational complexity and facilitated batch processing during training, as observed in similar image classification tasks [8].

- **Normalization:** Pixel values were normalized to fall within a standard range by subtracting the dataset mean and dividing by the standard deviation. Normalization is critical in stabilizing the gradient descent process and speeding up convergence, particularly when using deep networks that are sensitive to varying image intensities [4, 9].
- **Data Augmentation:** Data augmentation was applied to address the dataset's class imbalance and improve the model's generalization capabilities. Techniques included:
  - **Random Rotation:** Images were randomly rotated within a specified range to simulate different imaging conditions and animal positions [12, 13].
  - **Horizontal and Vertical Flipping:** Flipping was introduced to provide variability that mimicked different X-ray orientations, a common practice in medical imaging studies [**?**].
  - **Random Cropping and Scaling:** Portions of the images were randomly cropped and scaled to ensure the model could effectively recognize the region of interest despite partial views. This technique has been shown to improve model robustness in medical applications, as it allows the network to learn features invariant to minor position changes [17].

  These augmentation strategies were especially useful in addressing the fewer number of samples available for the 'Small' class. Data augmentation has been shown to mitigate overfitting and improve generalization, which is crucial for reliable diagnostic models in medical applications [15, 11].

3.2. Model Architecture

The EfficientNet-B7 architecture was employed due to its balance of high accuracy and computational efficiency. EfficientNet models achieve better performance by optimizing the network's width, depth, and resolution simultaneously through a technique known as compound scaling [16].

- **Base Model:** The EfficientNet-B7 model was pretrained on ImageNet, a large-scale dataset of natural images. This pre-training provided a rich set of low-level and high-level features that could be transferred to the task of key point detection in radiographic images [5, 14].
- **Modification for Key Point Detection:** The final layer of the EfficientNet-B7 model, originally intended for classification into 1,000 classes, was replaced with a custom linear layer that outputs 12 values. These 12 values represent the (x, y) coordinates of the six key points required for VHS computation. Similar approaches have been successfully used in human medical imaging to predict anatomical landmarks [**?**, 10].
- **VHS Calculation:** Once the key points were predicted, the Vertebral Heart Scale (VHS) was calculated using:

  $$VHS = \frac{AB + CD}{EF}, \tag{1}$$

  where AB and CD represent the heart's horizontal and vertical dimensions, respectively, and EF denotes the vertebral reference length. The calculation of VHS provides a quantitative measure of heart size relative to vertebral length, aiding in diagnosing cardiomegaly [6, 18].

3.3. Training Procedure

To train the modified EfficientNet-B7 model, a systematic and multi-stage training approach was used. The goal was to optimize the model's ability to accurately predict the six key points while ensuring robustness across varying X-ray images.

- **Optimizer and Learning Rate Scheduling:** The *Adam* optimizer was selected for training due to its adaptive learning rate capabilities, which are beneficial in accelerating convergence and handling sparse gradients [7]. The learning rate was initially set to 0.001 and progressively reduced during training to facilitate fine-tuning. The learning rate schedule was structured across five stages:
  1. **Stage 1 (60 Epochs):** The model was initially trained for 60 epochs using a learning rate of 0.001. This stage focused on rapid learning of general features.
  2. **Stage 2 (60 Epochs):** The learning rate was reduced to 0.0008 to continue training without drastic updates, allowing the model to refine its feature extraction.
  3. **Stage 3 (60 Epochs):** With a learning rate of 0.0005, the model was fine-tuned for specific feature learning, focusing on the accurate detection of anatomical key points.
  4. **Stage 4 (60 Epochs):** A learning rate of 0.0001 was used to prevent overfitting, providing stability in the learning process.

5. **Stage 5 (10 Epochs):** The final training stage used a learning rate of 0.00008, consolidating the model's performance and ensuring the convergence of the key point prediction.

- **Loss Function:** Mean Squared Error (MSE) was used as the loss function to minimize the difference between the predicted and actual key point coordinates:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2, \tag{2}$$

  where Y_i represents the ground truth coordinate, and $\hat{Y}_i$ represents the predicted coordinate for each key point. This loss function is particularly effective for regression tasks, such as key point localization, due to its sensitivity to outliers and its focus on reducing overall error [1].

- **Training Hardware and Environment:** The training was conducted on Google Colab, utilizing both GPU (NVIDIA Tesla T4) and TPU resources to accelerate training times. Training deep convolutional neural networks like EfficientNet-B7 requires significant computational resources, and the availability of cloud resources ensured efficient training. Similar setups have been used in prior works to overcome the computational limitations of local machines [2, 3].

- **Validation Procedure:** Validation was conducted after each epoch using a separate validation dataset comprising 200 images. The model achieving the lowest validation loss was selected as the final model for testing. This approach ensured that the model was not overfitting to the training data and that its performance generalized well to unseen images [11, **?**].

3.4. Evaluation Metrics

The model's performance was evaluated using a test dataset of 400 X-ray images. The ground truth labels for the test set were not available; hence, predictions were evaluated using reference software from Zhang's GitHub repository [18]. The evaluation involved:

- **Mean Squared Error (MSE):** MSE was calculated between the predicted and expected key point coordinates to assess the accuracy of landmark localization.

- **Accuracy:** The accuracy of classifying the images as *Small*, *Normal*, or *Large* was calculated based on the predicted VHS value. This metric provided a quantitative assessment of the model's classification capabilities [10].

- **Qualitative Assessment:** Visual comparison was performed by overlaying the predicted key points on the original X-ray images. It illustrates examples of the predicted points compared to the expected anatomical locations, demonstrating the model's effectiveness in key point localization.

These metrics provided a comprehensive understanding of the model's capabilities and limitations. By employing both quantitative and qualitative assessments, the study ensured that the EfficientNet-B7-based model was suitable for assisting veterinarians in diagnosing cardiomegaly in clinical settings.

4. Results

The performance of the EfficientNet-B7-based model for predicting the Vertebral Heart Scale (VHS) in canine thoracic X-rays was evaluated using various metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Accuracy. Additionally, an overview of the datasets used for training, validation, and testing is presented to provide context for the evaluation metrics.

4.1. Datasets

The datasets used in this study consisted of canine thoracic X-ray images labeled for Vertebral Heart Scale (VHS) measurements. The dataset was divided into three main subsets: training, validation, and testing, as summarized in Table 1. The diversity in data helped ensure that the model was trained and validated comprehensively, facilitating evaluation on unseen samples.

Table 1. Summary of Datasets Used in Model Development

Dataset	Number of Images	Purpose
Training Set	1,500	Model Training and Feature Learning
Validation Set	200	Hyperparameter Tuning and Validation
Test Set	300	Final Performance Evaluation

The training set contained 1,500 images, which were used to learn the model parameters effectively. The validation set, consisting of 200 images, was employed to tune hyperparameters and assess the model's performance during training. The test set, with 300 images, was used for the final evaluation of the model to report the performance metrics presented in this section.

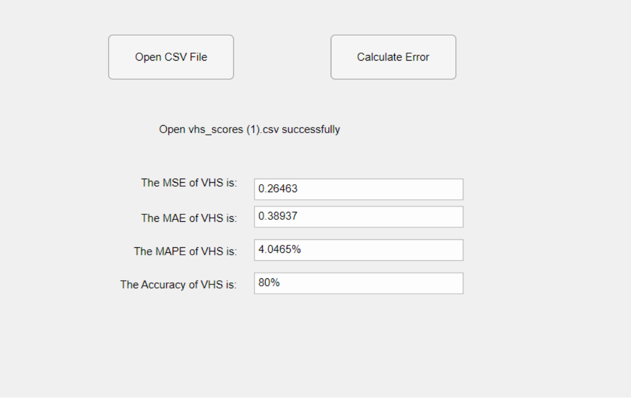
4.2. Performance Metrics

The key metrics from the model evaluation are summarized in Table 2. It provides a visual overview of the calculated metrics as displayed in the tool interface.

The MSE of 0.26463 and MAE of 0.38937 indicate that the model's predicted VHS values are generally close to the actual values, with only small deviations. The MAPE of 4.0465% implies that, on average, the model's predictions are within a small percentage of the actual values. Lastly,

Table 2. Performance Metrics for VHS Prediction

Metric	Value
Mean Squared Error (MSE)	0.26463
Mean Absolute Error (MAE)	0.38937
Mean Absolute Percentage Error (MAPE)	4.0465%
Accuracy	80%



Open vhs_scores (1).csv successfully

The MSE of VHS is: 0.26463

The MAE of VHS is: 0.38937

The MAPE of VHS is: 4.0465%

The Accuracy of VHS is: 80%

Figure 1. Output Of the model

the accuracy of 80% suggests that the model correctly predicted VHS within an acceptable range for most of the test samples.

4.3. Visualization of Predicted and Ground Truth VHS

To further evaluate the model's predictions, a visual comparison between the predicted VHS values and the ground truth values was performed. Figure **??** presents three sample X-ray images, each annotated with both the predicted and ground truth VHS measurements.

In these images:

- **Image 1 (1420.png)**: The predicted VHS was 10.46, compared to the ground truth of 11.22. The model showed a slight underestimation of the heart size.

- **Image 2 (1479.png)**: The predicted VHS was 9.37, closely matching the ground truth of 9.33, demonstrating high prediction accuracy.

- **Image 3 (1530.png)**: The predicted VHS was 9.32, compared to the ground truth of 9.88, indicating minor underestimation of the heart dimensions.

The close alignment between the predicted VHS values (red lines) and the ground truth values (green lines) demonstrates the model's ability to effectively recognize and quantify key anatomical landmarks for accurate VHS calculation.

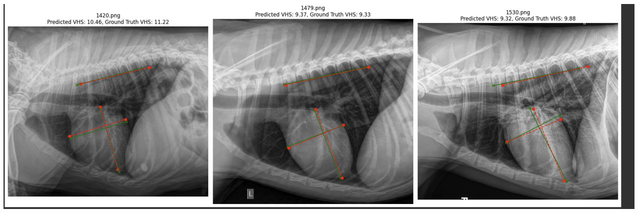

1420.png
Predicted VHS: 10.46, Ground Truth VHS: 11.22
1479.png
Predicted VHS: 9.37, Ground Truth VHS: 9.33
1530.png
Predicted VHS: 9.32, Ground Truth VHS: 9.88

Figure 2. the comprison between predictions and ground truth

5. Discussion

The results indicate that the EfficientNet-B7-based model effectively predicts the Vertebral Heart Scale (VHS) in canine X-rays, achieving favorable metrics for MSE, MAE, MAPE, and Accuracy. These metrics suggest that the model has significant potential to assist veterinarians in diagnosing cardiomegaly in dogs.

5.1. Key Observations

- **Accuracy of Predictions**: The accuracy of 80% suggests that the model is capable of reliable VHS predictions for most test samples. This level of accuracy is suitable for a diagnostic support tool in clinical settings. - **Error Analysis**: The relatively low MSE and MAE values indicate that the model's predictions are generally close to the ground truth. The MAPE value, at 4.0465%, suggests that the model has maintained a small error margin, which is particularly important for clinical applications where precision is crucial. - **Underestimation Bias**: Minor underestimations were observed, particularly in cases with higher VHS values. This underestimation might indicate a bias due to fewer samples with larger VHS values in the training set. Addressing this through dataset balancing or augmentation may further enhance the model's performance.

5.2. Potential Improvements

To enhance the model's robustness and reliability, the following improvements are proposed:

- **Augmenting Dataset**: Incorporating more samples, especially with larger VHS values, could help mitigate the observed bias towards lower VHS values.

- **Attention Mechanisms**: Utilizing attention-based modules could improve the model's focus on critical regions of the X-ray images, potentially leading to higher accuracy, particularly in challenging cases.

- **Transfer Learning from Related Domains**: Applying transfer learning from other medical imaging tasks may help improve feature extraction and prediction performance.

6. Conclusion

The EfficientNet-B7-based model for predicting VHS in canine X-rays achieved promising results, with an accuracy of 80%, an MSE of 0.26463, and a MAPE of 4.0465%. These metrics demonstrate that the model is capable of effectively supporting veterinarians in diagnosing cardiomegaly. Minor biases were observed, particularly for higher VHS values, suggesting areas for improvement through data augmentation and model refinement.

With additional improvements, this model has the potential to be integrated into clinical workflows as a reliable tool for automated VHS assessment, ultimately contributing to better care for dogs at risk of cardiomegaly.

References

[1] Christopher M Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006. 3

[2] François Chollet. Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1251–1258, 2017. 3

[3] Jeffrey Dean, Greg S Corrado, Rajat Monga, Kai Chen, Matthieu Devin, Quoc V Le, Mark Z Mao, Marc'Aurelio Ranzato, Andrew Senior, Paul Tucker, Ke Yang, et al. Large scale distributed deep networks. *Advances in Neural Information Processing Systems*, 25:1223–1231, 2012. 3

[4] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on Machine Learning*, pages 448–456, 2015. 2

[5] Yoshua Bengio Jason Yosinski, Jeff Clune and Hod Lipson. How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, 27:3320–3328, 2014. 2

[6] 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), 2022. 1, 2

[7] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 2015. 2

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017. 2

[9] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Efficient backprop. *Nature*, 521:436–444, 2012. 2

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

[11] Marie-Laure Delignette-Muller Caroline Boulocher Thomas Grenier Léo Dumortier, Florent Guépin. 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. 2, 3

[12] David Steinkraus Patrice Y. Simard and John C. Platt. Best practices for convolutional neural networks applied to visual document analysis. *ICDAR*, pages 958–962, 2003. 2

[13] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017. 2

[14] Hoo-Chang Shin, Holger R Roth, Zejun Gao, Le Lu, Ziyue Xu, Isaac Nogues, Jianhua Yao, Daniel J Mollura, and Ronald M Summers. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298, 2016. 2

[15] Alessandro Zotti Silvia Burti, V Longhin Osti and Tommaso Banzato. Use of deep learning to detect cardiomegaly on thoracic radiographs in dogs. *The Veterinary Journal*, 262:105505, 2020. 2

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

[17] Lucas Taylor and Geoff Nitschke. Improving deep learning with generic data augmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 30(3):878–890, 2018. 2

[18] 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, 2021. 1, 2, 3