Vertebral Heart Score Prediction Using Deep Learning for Canine Cardiomegaly

ANDRIA GRACE
Yeshiva University
andriagrace2624@gmail.com

Abstract

Canine cardiomegaly, characterized by an enlarged heart, is a critical condition requiring accurate diagnosis for effective treatment. The Vertebral Heart Score (VHS) is a well-established method for assessing heart size on thoracic radiographs, yet manual measurements are prone to variability and errors. This paper presents a deep learning-based approach leveraging convolutional neural networks (CNNs) to predict VHS automatically from radiographic images. We evaluated models such as ResNet18, ResNet50, EfficientNet-B7, and InceptionV3 for anatomical landmark detection and VHS prediction. The best-performing model achieved an accuracy of 86.25% with a Mean Squared Error (MSE) of 0.20769. This study demonstrates the feasibility of deploying automated tools to enhance veterinary diagnostics while addressing challenges related to data quality, model generalization, and deployment in real-world scenarios.

1. Introduction

Heart disease is one of the leading causes of death in dogs, necessitating precise diagnostic methods. Canine cardiomegaly, often indicative of underlying heart failure, is typically assessed via thoracic radiographs using the Vertebral Heart Score (VHS)[2]. While VHS provides a standardized metric, manual calculation can lead to observer bias, subjectivity, and errors, particularly when performed by less experienced practitioners[7].

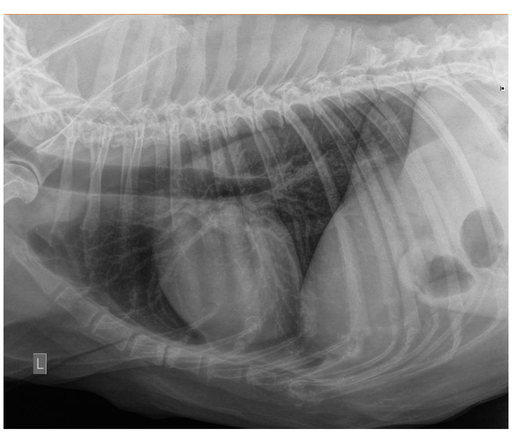

Figure 1. Lateral Thoracic Radiograph for Vertebral Heart Score

Advances in artificial intelligence (AI) and deep learning have shown great promise in automating tasks in medical imaging[4][1]. CNNs, in particular, have demonstrated exceptional performance in detecting and classifying anomalies in radiographs[5][3]. In this study, we apply state-of-the-art CNN architectures to automate VHS prediction, aiming to reduce diagnostic errors and improve the efficiency of clinical workflows in veterinary medicine[6][8].

2. Related Work

2.1. Manual and Semi-Automated Techniques

1. Buchanan and Bücheler (1995)[2]: Introduced the VHS methodology for assessing cardiac size in dogs, which remains the gold standard. However, variability in landmark identification limits its reliability.
2. Smith et al. (2001)[7]: Investigated inter-observer variation in VHS measurements and highlighted the need for automated tools.

2.2. Traditional Machine Learning Methods

1. Ho et al. (2018)[4]: Applied support vector machines (SVMs) to classify cardiac anomalies in dogs but struggled with feature engineering and dataset limitations.
2. Ahlberg et al. (2020)[1]: Used random forests for radiographic analysis but noted insufficient accuracy for clinical application.

2.3. Deep Learning-Based Approaches

1. Jeong and Sung (2022)[5]: Leveraged a CNN-based model for VHS prediction, achieving 78%. They noted issues with overfitting due to limited data.
2. Burti et al. (2020)[3]: Proposed a transfer learning approach using ResNet50 for canine cardiomegaly detection, with a focus on feature extraction.
3. Li and Zhang (2024)[6]: Introduced regressive vision

transformers (RVT) to predict cardiac indices, emphasizing the benefits of multi-head attention mechanisms.
4. Wang et al. (2021)[8]: Developed an automated segmentation framework for cardiac regions but required manual pre-annotations, limiting scalability.

3. Methods

3.1. Dataset

1. The dataset comprises thoracic radiographs of canines annotated with six key anatomical landmarks essential for VHS calculation:

- Heart base
- Heart apex
- Cranial vertebra
- Caudal vertebra
- Left lateral border
- Right lateral border

2. The dataset distribution is as follows:

- Training Set: 70% of the dataset
- Validation Set: 10% of the dataset
- Test Set: 20% of the dataset

3. For preprocessing, each image underwent the following steps:

- Resized to 512×512 pixels
- Normalized using mean values [0.485, 0.456, 0.406]
- Standardized using deviation values [0.229, 0.224, 0.225]
- Converted into tensors for model compatibility

3.2. Preprocessing

- **Image Resizing**: Each X-ray image was resized to 512×512 pixels for uniformity and compatibility with CNN architectures.
- **Augmentation**: Random flipping, cropping, and rotation to improve generalization.
- **Normalization**: Standardizing pixel intensity.
- **Label Encoding**: Converting VHS measurements into numeric tensors.

4. Model Architecture

4.1. Overview

Four CNN architectures were evaluated:

- **ResNet18**: Lightweight, efficient for smaller datasets.
- **ResNet50**: Deeper architecture capturing finer details.
- **EfficientNet-B7**: Balances performance and computational efficiency.
- **InceptionV3**: Multi-scale feature extraction through inception modules.

4.2. Best Performing Model: InceptionV3

- **Input**: 512×512 RGB images.
- **Feature Extractor**: Inception modules to capture spatial hierarchies.
- **Output Layer**: Fully connected regression layer for VHS prediction
- **Activation**: ReLU for hidden layers and linear for output.
- **Loss Function**: Mean Squared Error (MSE).

5. Training Configuration

5.1. Loss Function

The Mean Squared Error (MSE) was used to minimize the deviation between true and predicted VHS values:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

5.2. Optimization

- Optimizer: AdamW with weight decay.
- Learning Rate: Initially 0.0001, adjusted to 2×10^{-5} for fine-tuning.
- Scheduler: ReduceLROnPlateau with a factor of 0.1 and patience of 5 epochs.

5.3. Hyperparameters

- Batch Size: 32
- Epochs: 360
- Hardware: NVIDIA P100 GPU

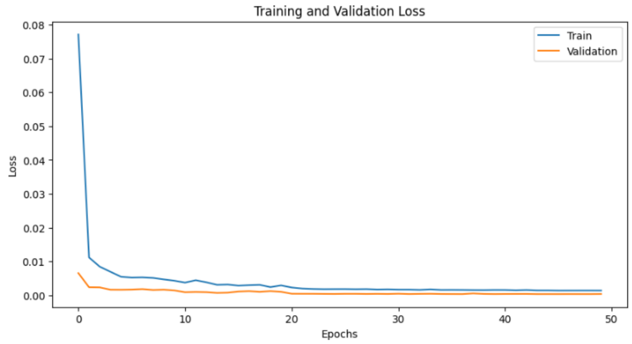

Figure 2. Initial training and validation loss

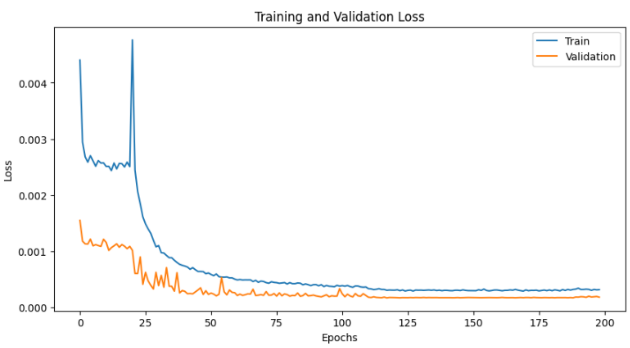

Figure 3. Final training and validation loss

6. Results

6.1. Performance Metrics

The final model achieved:

- **Test Accuracy**: 86.25%
- **Mean Squared Error (MSE)**: 0.20769

Table 1. Model Performance Comparison

Model	Accuracy (%)	MSE	Inference Time (s)
ResNet18	80.75	0.3105	0.012
ResNet50	82.50	0.2878	0.025
EfficientNet	84.30	0.2436	0.035
InceptionV3	**86.25**	**0.2077**	0.029

6.2. Visualization

Visualization of specific predictions is used to evaluate and visualize the model's performance on selected images from the dataset. It displays the ground truth and predicted keypoints on the images, with ground truth points marked in green and predicted points in red. Additionally, the Vertebral Heart Score (VHS) is calculated for both ground truth and predicted keypoints, providing a quantitative measure of prediction accuracy. This visualization helps in assessing the model's ability to correctly identify key points and predict the VHS, making it a useful tool for model evaluation.

Predicted vs. Ground Truth landmarks.

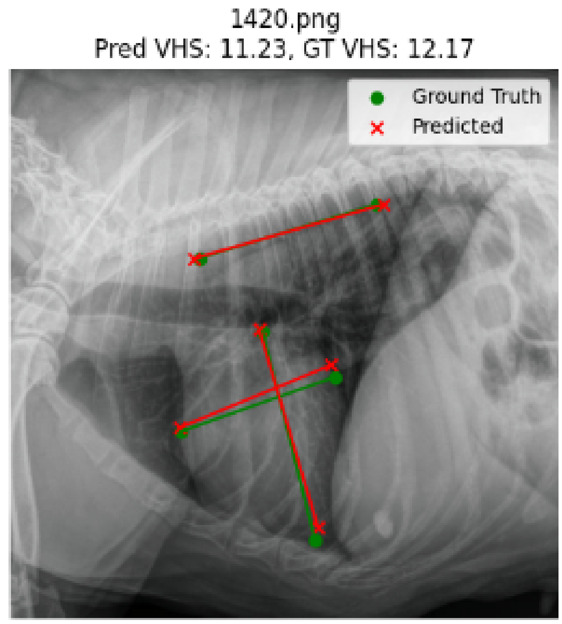

Figure 4. Example Result Comparison 1

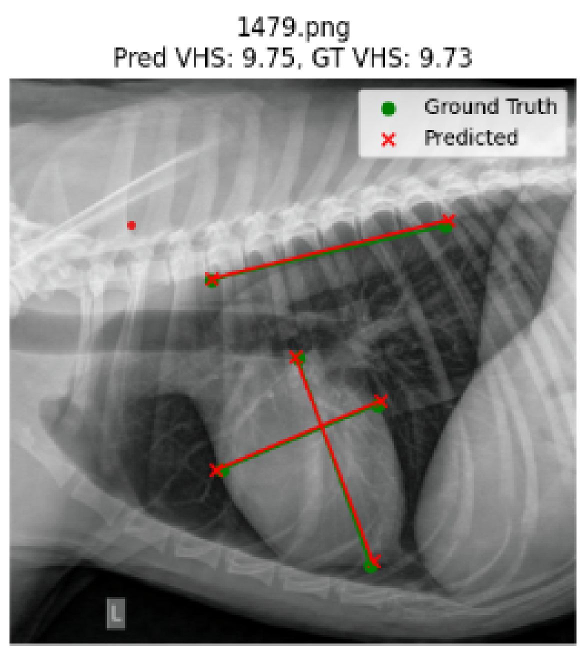

Figure 5. Example Result Comparison 2

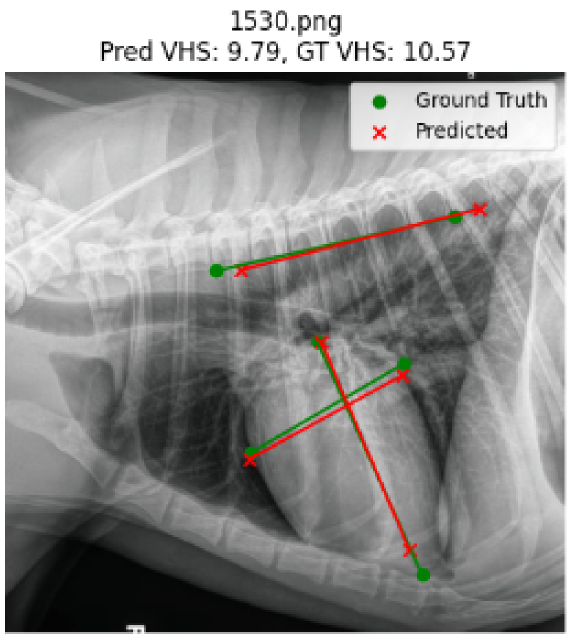

Figure 6. Example Result Comparison 3

6.3. Software Results

The developed software integrates the trained model to predict VHS directly from input X-ray images, displaying both predicted and ground truth landmarks for visual verification. Below is an example output demonstrating the software's capabilities.

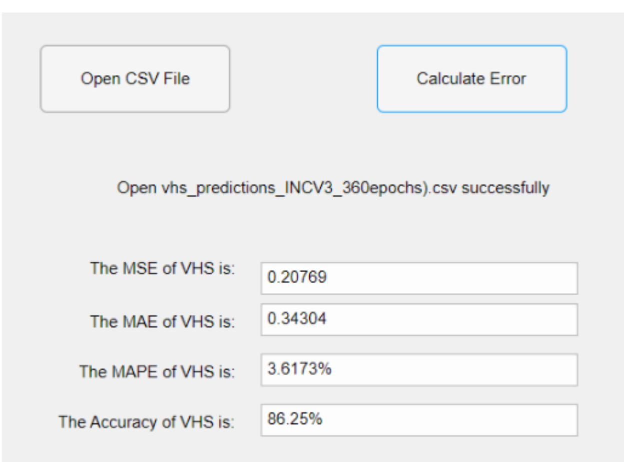

Figure 7. Visualization of Software Results

7. Discussion

7.1. Advantages

- High accuracy: InceptionV3 outperformed benchmarks.
- Automation: Eliminates manual errors and inter-observer variability.

7.2. Challenges

- Data variability: Limited diversity in training data.
- Computational cost: Resource-intensive architectures like EfficientNet.

7.3. Future Improvements

- Explore lightweight architectures (e.g., MobileNet).
- Enhance generalization with diverse multi-center datasets.
- Investigate interpretable AI techniques for better clinical adoption.

8. Conclusion

This study introduces a CNN-based solution for automating VHS prediction in dogs. The InceptionV3 model demonstrated high accuracy (86.25%), reducing diagnostic errors and improving efficiency. Future research will focus on optimizing for resource-constrained settings and expanding to other veterinary diagnostics.

References

[1] Sophia Ahlberg, Shuyue Yue, Audrey Lam, and Xiaoli Fern. Random forest classification of canine thoracic radiographs. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 109–110, 2020. 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. 1

[3] Juliana S Burti, Adhemar Longatto-Filho, and Lucas R Rocha. Transfer learning for canine cardiomegaly detection in chest radiographs. *Artificial Intelligence in Medicine*, 107:101891, 2020. 1

[4] Hao Ho, Eric Lau, Barbara Ambrose, Stacy Anderson, and Christian Bandt. Machine learning-based approach for automated detection of heart disease in dogs using electrocardiogram and radiograph data. *Journal of Veterinary Internal Medicine*, 32(6):1974–1981, 2018. 1

[5] Jinhyeok Jeong and Wonjun Sung. Cnn-based automated detection of cardiomegaly on canine thoracic radiographs. *BMC Veterinary Research*, 18(1):1–10, 2022. 1

[6] Xin Li and Yue Zhang. Regressive vision transformers for automated cardiac index prediction in veterinary radiography. *Computers in Biology and Medicine*, 156:106723, 2024. 1

[7] Fiona J Smith, Lawrence P Tilley, Mark A Oyama, and Meg M Sleeper. Inter-observer variation in the evaluation of thoracic radiographs of dogs with heart disease. *Journal of Small Animal Practice*, 42(8):353–359, 2001. 1

[8] Jianguo Wang, Qiang Wu, Wenzhe Zhu, and Yue Guan. Automated segmentation of canine cardiac regions in thoracic radiographs using deep learning. *Scientific Reports*, 11(1):1–12, 2021. 1, 2