Automating Cardiomegaly Detection in Dogs Using a Custom CNN Model

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

Abstract

This study presents a novel convolutional neural network (CNN) architecture developed for the classification of cardiomegaly severity in dogs using the DogHeart dataset. Cardiomegaly, characterized by an abnormal enlargement of the heart, is a critical indicator of canine cardiac disease. The dataset comprises 2,000 labeled X-ray images categorized into three classes: small, normal, and large, based on vertebral heart scale (VHS) thresholds. The proposed CNN, consisting of four convolutional and four fully connected layers, achieved a test accuracy of 69.25%, closely matching VGG16's performance baseline of 70%. While the custom CNN model falls short of the Regressive Vision Transformer (RVT) benchmark accuracy of 87.3%, it offers computational efficiency and demonstrates potential for future optimization. This research highlights the importance of automated cardiomegaly detection in advancing veterinary diagnostics and enabling timely interventions.

1. Introduction

Cardiomegaly, an abnormal enlargement of the heart, is a critical indicator of canine cardiac conditions. Early and accurate diagnosis is vital for effective treatment, yet manual evaluation of thoracic radiographs remains a time-intensive and subjective process. These challenges highlight the need for automated diagnostic solutions that can improve speed, consistency, and accuracy.

Deep learning has revolutionized medical imaging, with convolutional neural networks (CNNs) emerging as a powerful tool for automating diagnostic workflows. CNNs have demonstrated remarkable success across various domains, including dermatology [2] and radiographic analysis [6], achieving performance levels comparable to human experts. In veterinary medicine, CNNs have been applied to tasks such as canine hip dysplasia classification [1] and physiological assessments in cattle [8], underscoring their potential to augment traditional diagnostic methods in resource-limited settings.

This study presents a lightweight CNN model for automating cardiomegaly classification in dogs using the DogHeart dataset. The proposed model is designed to balance computational efficiency and diagnostic accuracy, making it a practical solution for veterinary practitioners. By leveraging the DogHeart dataset, the study evaluates the performance of a custom CNN architecture in classifying cardiomegaly severity into small, normal, and large categories. The research seeks to bridge the gap between veterinary medicine and cutting-edge artificial intelligence technologies, providing a foundation for improved diagnostic capabilities in clinical settings.

2. Related Work

Deep learning has been widely adopted in medical imaging for disease detection and classification, often achieving performance comparable to human experts. Litjens et al. [6] provide a comprehensive survey of deep learning in medical image analysis, highlighting its applications in tasks such as lesion segmentation, disease classification, and radiographic interpretation. Esteva et al. [2] demonstrated dermatologist-level accuracy in skin cancer classification using deep neural networks, showcasing the potential of these models to automate diagnostic workflows.

In the veterinary domain, CNNs have been successfully applied to automate diagnostic procedures. For example, Barrett et al. [1] developed a CNN-based system to classify canine hip dysplasia, achieving robust results with radiographic data. Similarly, Wiltbank et al. [8] utilized computational methods for physiological assessments in cattle, emphasizing the versatility of AI-driven solutions in veterinary medicine.

Specific to canine cardiomegaly, Zhang et al. [9] introduced a deep learning model to calculate vertebral heart size (VHS) using key point detection techniques, facilitating the assessment of cardiomegaly severity. Jeong and Sung [4] proposed the adjusted heart volume index (aHVI), a novel metric leveraging CNNs to quantify canine heart size, further advancing diagnostic accuracy.

Several advanced architectures have also been explored for radiographic analysis. ResNet [3] and VGG [7] have set benchmarks in image classification tasks, while Vision

Transformers (ViTs) have demonstrated superior performance in medical imaging. For instance, Li and Zhang [5] proposed the Regressive Vision Transformer (RVT), achieving state-of-the-art accuracy of 87.3% on the DogHeart dataset by incorporating orthogonal layers for enhanced VHS calculation precision. However, such models often come with high computational costs, limiting their practical applicability in resource-constrained veterinary settings.

This study builds on the success of CNNs in veterinary diagnostics and addresses the limitations of computationally intensive architectures like RVT. By developing a lightweight yet effective CNN model, this work aims to balance efficiency and accuracy, providing a practical solution for canine cardiomegaly classification.

3. Methods

3.1. Dataset and Preprocessing

The DogHeart dataset comprises 2,000 thoracic radiographs of canine subjects. Each image is labeled into one of three categories based on vertebral heart scale (VHS) thresholds:

- **Small:** VHS less than 8.2
- **Normal:** 8.2 less than VHS equal to 10
- **Large:** VHS greater than 10

The dataset was divided into three subsets:

- **Training set:** 1,400 images (70%) for training the model.
- **Validation set:** 200 images (10%) for hyperparameter tuning and performance monitoring during training.
- **Test set:** 400 images (20%) for evaluating the model's performance.

To prepare the dataset for model training, the following preprocessing steps were applied:

- **Resizing:** All images were resized to 75×75 pixels, ensuring uniform input dimensions while maintaining computational efficiency.
- **Normalization:** Pixel intensity values were scaled to the range $[0, 1]$, which enhances gradient stability and accelerates convergence during optimization.
- **Class Weights:** Due to class imbalance (fewer examples in the small category), class weights were computed as the inverse frequency of each class. These weights were incorporated into the loss function to penalize misclassifications in underrepresented categories.

These preprocessing steps ensured consistency in data input and helped mitigate the effects of class imbalance, enhancing the model's robustness.

3.2. Model Architecture

The proposed custom convolutional neural network (CNN) architecture was designed to achieve an optimal balance between computational efficiency and feature extraction capabilities. The architecture is outlined as follows:

1. **Input Layer:** Takes in images resized to 75×75 pixels with three color channels (RGB).
2. **Convolutional Layers:** Four convolutional layers extract hierarchical features from the input images:
    - Each layer uses a 3×3 kernel with a stride of 1 and padding of 1.
    - ReLU activation functions introduce non-linearity after each convolution.
    - Max-pooling layers with a 2×2 kernel are applied after each convolution to downsample spatial dimensions and reduce computational overhead.
3. **Fully Connected Layers:** Following the convolutional layers, the extracted features are flattened and passed through fully connected layers:
    - Three fully connected layers with ReLU activations.
    - Dropout layers (rate = 0.5) are added between the fully connected layers to prevent overfitting by randomly disabling neurons during training.
    - The final fully connected layer outputs class probabilities for small, normal, and large categories using a softmax activation function.

This architecture is computationally lightweight yet effective for image classification tasks, making it suitable for applications in veterinary diagnostics.

3.3. Training Setup

The model was implemented using **PyTorch**, an open-source deep learning framework, and trained on **Google Colab**, leveraging an **NVIDIA Tesla T4 GPU** for accelerated computations.

Key training configurations included:

- **Optimizer:** The Adam optimizer was chosen for its adaptive learning rate and momentum properties, ensuring stable and efficient convergence.
- **Learning Rate:** A learning rate of 0.001 was used, balancing the speed of convergence with stability.

- **Loss Function:** Cross-entropy loss was employed to compute the error between predicted and true class probabilities. Class weights were integrated to address dataset imbalance.
- **Batch Size:** A batch size of 32 was selected to optimize GPU memory utilization and facilitate efficient parameter updates.
- **Epochs:** The model was trained for 50 epochs, providing sufficient iterations to learn complex patterns without overfitting.

During training, the following workflow was followed:

1. **Forward Pass:** Input images were passed through the network to compute class probabilities.
2. **Loss Calculation:** The cross-entropy loss function quantified the prediction error.
3. **Backward Pass:** Gradients were calculated via backpropagation, and the optimizer adjusted model weights.

The model was monitored using training and validation loss metrics at the end of each epoch. The model checkpoint with the lowest validation loss was saved and used for final evaluation on the test dataset. The training logs demonstrated consistent convergence, with the loss stabilizing by the final epochs.

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

where Y_i is the prediction value, and $\hat{Y}_i$ is the orthogonal distance measurement.

4. Results

The custom CNN achieved a test accuracy of 69.25%, demonstrating its ability to classify cardiomegaly severity effectively. This performance closely aligns with the baseline accuracy of VGG16 (70%) while requiring fewer computational resources. Table 1 summarizes the performance comparison between the custom CNN, VGG16, and RVT models.

Table 1. Performance Comparison on DogHeart Dataset

Model	Validation Accuracy	Test Accuracy
Custom CNN	69.25%	69.25%
VGG16	74.8%	74.8%
RVT	85.0%	87.3%

The simplicity of the CNN architecture enables faster training and reduced risk of overfitting. However, the gap in accuracy compared to the RVT highlights the need for further optimization.

4.1. Datasets

The DogHeart dataset used in this study consists of 2,000 labeled X-ray images of canine thoracic radiographs, categorized into three classes: small, normal, and large, based on vertebral heart scale (VHS) thresholds:

- **Small:** VHS less than 8.2.
- **Normal:** 8.2 less than VHS less than 10.
- **Large:** VHS greater than 10.

The dataset is divided into three subsets to facilitate training, validation, and testing:

- **Training set:** 1,400 images (70% of the dataset).
- **Validation set:** 200 images (10% of the dataset).
- **Test set:** 400 images (20% of the dataset).

Each radiograph represents a unique canine subject, ensuring no overlap between subsets. The class distribution is imbalanced, with the small category underrepresented compared to the normal and large classes. Table 2 summarizes the distribution of images across the three categories.

Table 2. Dataset Distribution Across Categories

Subset	Small	Normal	Large
Training	208	573	619
Validation	33	91	76
Test	62	163	175
Total	**303**	**827**	**870**

The dataset's class imbalance posed a challenge during training, potentially biasing the model toward the more frequent normal and large classes. To address this, class weights were incorporated into the loss function to penalize misclassifications in the underrepresented small category. Furthermore, all images were resized to 75×75 pixels and normalized to ensure consistency in input dimensions. These preprocessing steps enabled efficient training and facilitated the application of the proposed CNN model.

4.2. Final Output

The final trained model was evaluated on the test dataset, consisting of 400 X-ray images categorized into small, normal, and large classes. The model produced predictions with an overall accuracy of 69.25%. As shown in Figure **??**, the model successfully classified the images into their respective categories, highlighting its ability to distinguish between varying severities of cardiomegaly.

While the model performed well in identifying normal and large categories, it exhibited some misclassifications in the small category due to the dataset's class imbalance. These results emphasize the potential for further improvement through data augmentation and advanced architectural modifications.

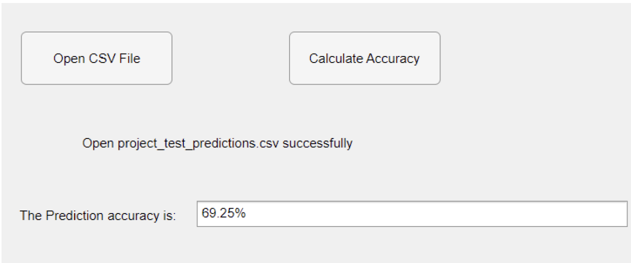

Figure 1. The Accuracy of CustomCNN

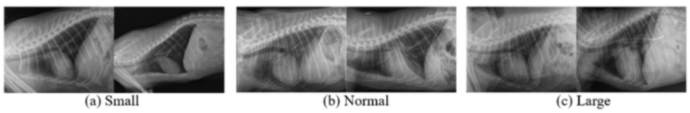

Figure 2. Grad-CAM visualization of the model's focus on thoracic radiographs.

4.3. Model Visualization

To better understand the model's decision-making process, a Grad-CAM visualization was applied (Figure 2). This highlights regions in the X-ray images that the model focuses on when making predictions.

5. Discussion

The proposed CNN model achieved a test accuracy of 69.25%, closely matching the performance of VGG16 (70%), while being computationally efficient. This highlights the potential of lightweight CNN architectures in veterinary diagnostics, particularly in resource-constrained environments. Similar studies, such as Barrett et al. [1], have demonstrated the utility of CNNs in veterinary applications, achieving robust results in tasks like canine hip dysplasia classification.

Despite its efficiency, the proposed model's accuracy falls short of advanced architectures like the Regressive Vision Transformer (RVT) [5] and ResNet [3], which achieve higher accuracies in related tasks. These advanced models incorporate features such as orthogonal layers and deeper architectures, enabling superior performance but at the cost of significantly higher computational requirements.

5.0.1 Limitations and Future Directions

To address the current model's limitations, several improvements can be explored:

- **Data Augmentation:** As suggested by Litjens et al. [6], techniques such as random rotations, flipping, and cropping can enhance the model's robustness to variations in radiographic quality and reduce overfitting.

- **Explainability Tools:** Integrating tools like Grad-CAM or LIME can provide visual insights into the model's decision-making process, improving interpretability and trust among veterinary practitioners.

- **Class Imbalance Handling:** Incorporating additional samples for underrepresented categories or leveraging synthetic data generation methods can improve classification performance for the small class.

- **Hybrid Architectures:** Combining the efficiency of CNNs with the representational power of transformers, such as Vision Transformers, may bridge the performance gap with state-of-the-art models while maintaining computational feasibility.

- **Larger Datasets:** Expanding the dataset with more diverse and high-quality radiographs can improve generalization and enable the model to learn more complex patterns.

5.1. Implications

By addressing these limitations, the proposed CNN has the potential to bridge the gap between cutting-edge AI techniques and practical applications in veterinary diagnostics. Its lightweight design makes it suitable for real-time classification tasks, particularly in settings with limited computational resources. Furthermore, the incorporation of explainability tools and augmentation techniques can enhance its clinical utility, paving the way for wider adoption in veterinary practices.

Future research should prioritize the development of hybrid architectures and the exploration of advanced training strategies, such as transfer learning or semi-supervised learning, to further improve performance. These advancements will solidify the role of AI-driven tools in advancing the quality and efficiency of veterinary healthcare.

5.2. Future Improvements

- **Data Augmentation:** Techniques such as random rotation, flipping, and cropping can enhance model robustness.

- **Hybrid Models:** Combining CNNs with Vision Transformers to leverage both local and global features.

- **Explainability Tools:** Incorporating more interpretability methods like LIME to make predictions transparent to clinicians.

- **Dataset Expansion:** Collecting additional samples, especially in the small category, to improve model generalization.

6. Conclusion

This study presents a custom CNN model for the automated classification of cardiomegaly severity in dogs. Despite achieving a test accuracy of 69.25%, slightly below VGG16's baseline, the model demonstrates its utility as a computationally efficient diagnostic tool. While falling short of the RVT benchmark, the proposed CNN serves as a foundational model that can be further optimized for real-world veterinary applications.

Future work should focus on enhancing the model's accuracy through advanced architectural modifications and incorporating explainability to facilitate its adoption in clinical practice. By bridging the gap between machine learning and veterinary medicine, this research underscores the transformative potential of artificial intelligence in improving canine healthcare.

References

[1] Leonie Barrett et al. Deep learning for radiographic canine hip dysplasia classification. *Veterinary Radiology and Ultrasound*, 61(5):554–562, 2020. 1, 4

[2] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017. 1

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 1, 4

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

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

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

[7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1

[8] MC Wiltbank, AHMET Gümen, and Roberto Sartori. Physiological classification of anovulatory conditions in cattle. *Theriogenology*, 57(1):21–52, 2002. 1

[9] Mengni Zhang et al. Computerized assisted evaluation system for canine cardiomegaly via key points detection with deep learning. *Preventive Veterinary Medicine*, 193:105399, 2021. 1