Confident Pseudo-labeled Diffusion Augmentation for Canine Cardiomegaly Detection

Shiman Zhang, M.S. in Artificial Intelligence

FACULTY MENTOR: Youshan Zhang, Ph.D.

Katz School of Science and Health

Introduction

- Canine cardiomegaly presents serious health risks, and traditional vertebral heart score (VHS) methods are limited by subjectivity and inconsistency, emphasizing the need for automated and accurate diagnostic approaches.
- Recent advances in deep learning (DL) and convolutional neural networks (CNNs) have shown promise in medical imaging (Li & Zhang, 2024), but existing veterinary applications often rely on small, poorly annotated datasets, leading to poor generalization.
- To address data scarcity, we propose the Confident Pseudo-labeled Diffusion Augmentation (CDA) model, which uses Latent Diffusion Models (Rombach et al., 2022) to generate 3,000 synthetic canine chest X-rays annotated with VHS key points, significantly expanding and diversifying the training data.
- By integrating pseudo-labeling (Lee, 2013) and uncertainty estimation via Monte Carlo Dropout (Gal & Ghahramani, 2016), CDA selectively incorporates high-confidence labels to iteratively refine the training set, improving accuracy and robustness over traditional approaches.

Method

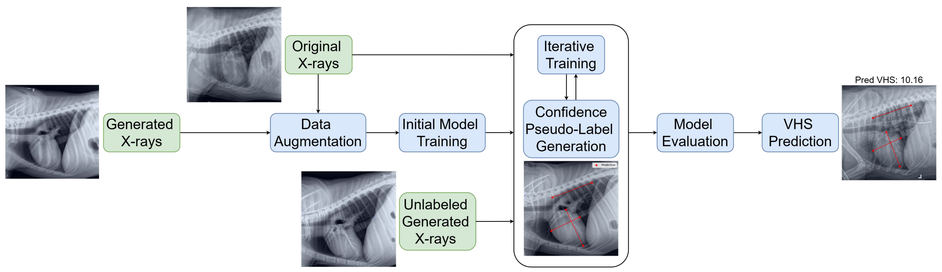

Figure 1. Overall framework of the Confident Pseudo-labeled Diffusion Augmentation (CDA) model.

We propose the Confident Pseudo-labeled Diffusion Augmentation (CDA) model, an AI-based method for automated Vertebral Heart Score (VHS) prediction from canine chest X-rays. CDA combines synthetic data generation via diffusion models, pseudo-labeling with Monte Carlo Dropout to exploit unlabeled data, and a task-optimized neural network architecture. The complete training pipeline is detailed in Algorithms 1 and 2.

Algorithm 1 Training Initial CDA Model

Require: Canine X-ray images $X = \{x_i\}_{i=1}^n$ and corresponding VHS-labeled key points $P = \{y_i\}_{i=1}^n$, where n is the total number of images.
Ensure: Predicted VHS scores and classifications for heart size.
for epoch = 1 to num_epochs **do**
  for each mini-batch $B(X)$ and $B(P)$ sampled from X and P **do**
    Compute true VHS scores and classification labels using Eq. (1).
    Apply model f to predict VHS scores and heart size classes.
    Calculate total loss $\mathcal{L}$ using Eq. (2).
    Update model parameters via backpropagation and optimizer step.
  end for
end for
return Final trained model for VHS prediction and cardiomegaly classification.

Algorithm 2 Pseudo-Labeling with Monte Carlo Dropout

Require: Pre-trained model f, labeled dataset $\mathcal{D}_{\text{Labeled}}$, unlabeled dataset $\mathcal{D}_{\text{Unlabeled}}$, confidence threshold τ, number of MC Dropout passes K.
Ensure: Trained model with improved performance.
for epoch = 1 to num_epochs **do**
  Train f on $\mathcal{D}_{\text{Labeled}}$ using the current loss function.
  Enable MC Dropout during inference for $\mathcal{D}_{\text{Unlabeled}}$.
  for each image $x \in \mathcal{D}_{\text{Unlabeled}}$ **do**
    Perform K stochastic forward passes to obtain $\{f_k(x)\}_{k=1}^K$.
    Compute mean prediction $\mu(x)$ and uncertainty $\sigma(x)$ using Eqs. (3) and (4).
  end for
  Select high-confidence samples $\mathcal{C}$ based on $\max(\sigma(x)) < \tau$.
  Combine $\mathcal{C}$ with $\mathcal{D}_{\text{Labeled}}$ to create an expanded training dataset.
end for
return Trained model f.

$$y_t = \begin{cases} 0 & \text{if VHS} < 8.2, \\ 1 & \text{if } 8.2 \leq \text{VHS} \leq 10, \\ 2 & \text{if VHS} > 10. \end{cases} \quad (1)$$

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}\Big(10 \cdot \mathcal{L}_{\text{L1}}(f(x_i), y_i^{\text{points}}) + 0.1 \cdot \mathcal{L}_{\text{L1}}(\text{calc\_vhs}(f(x_i)), y_i^{\text{VHS}}) + \mathbb{1}_{\{\text{epoch}>10\}} \cdot \mathcal{L}_{\text{L1}}(f(x_i), y_i^{\text{soft}})\Big), \quad (2)$$

where $\mathcal{L}_{\text{L1}}$ represents the mean absolute error (L1 loss).

$$\mu(x) = \frac{1}{K}\sum_{k=1}^{K} f_k(x), \quad (3)$$

$$\sigma(x) = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(f_k(x) - \mu(x))^2}, \quad (4)$$

where $\mu(x)$ represents the pseudo-label, and $\sigma(x)$ represents the model's uncertainty.

Results

Model	Valid Acc	Test Acc
GoogleNet	77.5	74.8
VGG16	78.5	75.0
ResNet50	80.0	78.2
DenseNet201	77.0	80.8
Inceptionv3	79.0	80.0
Xception	78.5	75.2
InceptionResnetV2	77.5	78.8
NasnetLarge	80.0	82.5
EfficientNetB7	82.0	84.5
Vision transformer	80.0	77.5
CONVT	82.0	85.3
Beit_large	71.0	74.3
RVT	84.9	87.3
ConvNeXt	89.5	89.8
Vim	73.5	71.5
MambaVision	87.5	86.8
CDA w/o CPL	88.5	91.0
CDA	89.5	92.8

Table 1. Results comparisons of different methods (Acc: accuracy, CPL: Confident Pseudo labels).

Table 1 compares the performance of state-of-the-art models, showing that our CDA model achieves the highest test accuracy of 92.8%, surpassing ConvNeXt (89.8%), MambaVision (86.8%), and ViM (71.5%). Without pseudo-labeling, CDA reaches 91.0%, highlighting the significant performance gain from leveraging high-confidence pseudo-labeled samples.

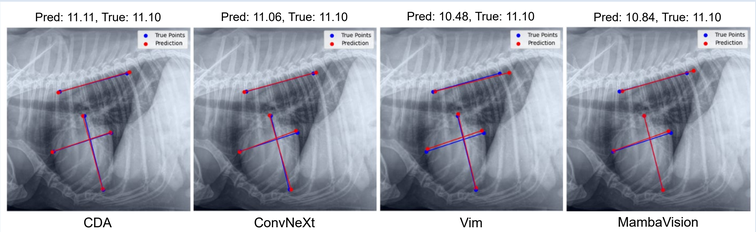

Figure 2. Prediction comparisons of different models.

Figure 2 presents a visual comparison of predicted key points and VHS values across different models. The CDA model shows superior alignment between the predicted key points (red) and the true anatomical key points (blue), resulting in a predicted VHS value that most closely matches the ground truth.

Conclusions

- Proposed a novel deep learning approach for automated Vertebral Heart Score (VHS) prediction and canine cardiomegaly classification.
- Integrated pseudo-labeling with Monte Carlo Dropout and synthetic data augmentation to address the scarcity of labeled veterinary imaging data.
- Achieved a test accuracy of 92.8%, outperforming state-of-the-art methods and demonstrating potential for improving early detection and clinical outcomes in canine cardiology.

Limitations:

- Synthetic data quality affects model performance, and low-quality samples may introduce bias.
- Fixed pseudo-labeling thresholds could exclude useful samples.
- High computational costs may limit deployment in low-resource settings.

Future Work:

- Explore adaptive pseudo-labeling and more advanced synthetic data generation.
- Investigate cross-modality learning and external validation to enhance robustness and clinical adoption.

Acknowledgements

I would like to express my sincere gratitude to Professor Youshan Zhang for his invaluable guidance, encouragement and support throughout this research. I also extend my appreciation to the Katz School of Science and Health for providing the resources and academic environment that made this work possible.

References

Li, J., & Zhang, Y. (2024). Regressive vision transformer for dog cardiomegaly assessment. *Scientific Reports, 14*(1), 1539.

Lee, D. H. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML* (Vol. 3, No. 2, p. 896).

Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning* (pp. 1050-1059). PMLR.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10684-10695).