PAFL: Personalized and Adaptive Federated Learning for Canine Cardiomegaly Keypoint Prediction

Jialu Li, M.S. in Artificial Intelligence

FACULTY MENTOR: Honggang Wang, Ph.D.

Katz
Katz School
of Science and Health

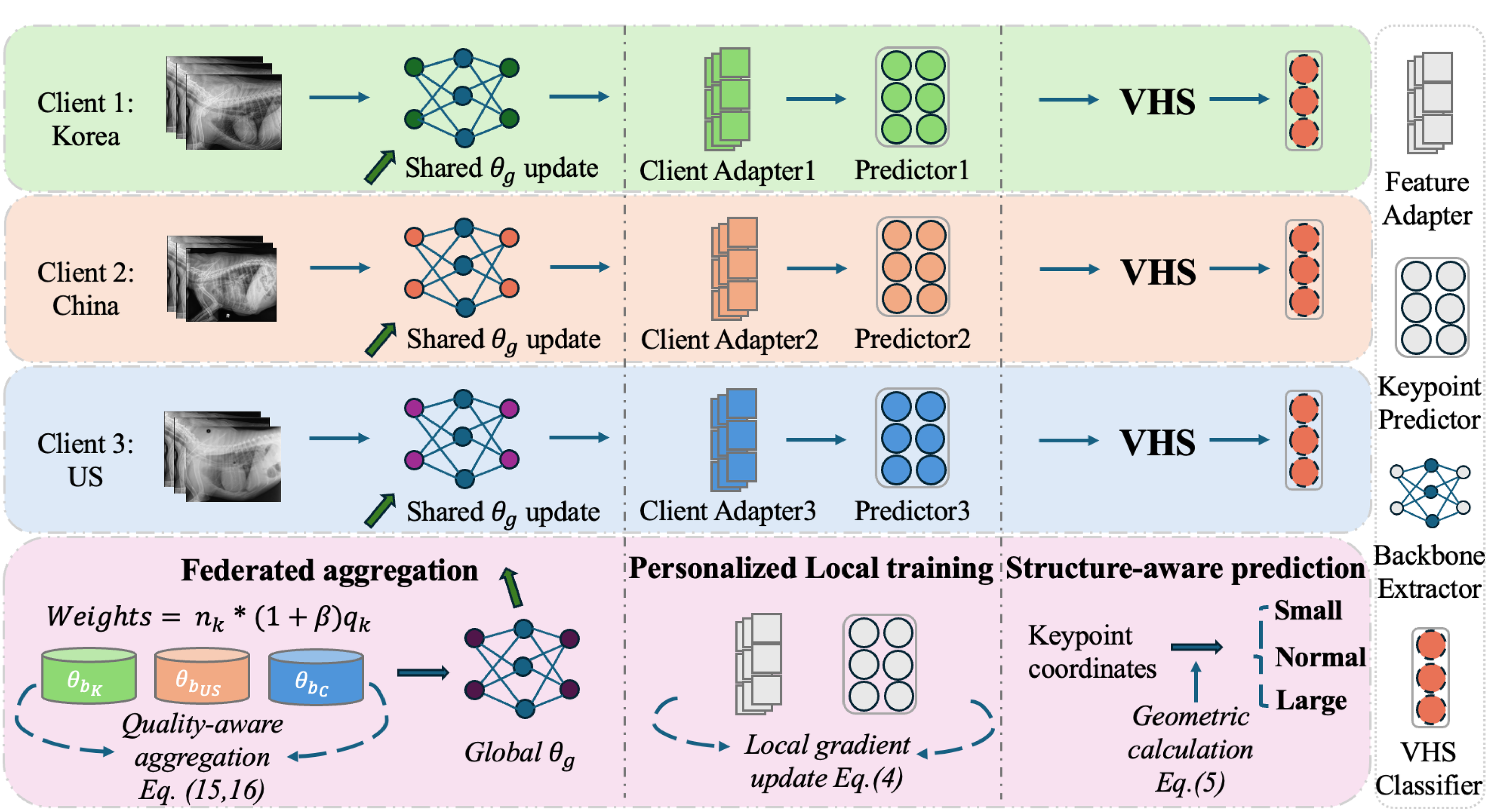
Introduction

Training high-performing deep learning models for medical image analysis typically requires *large-scale, diverse, and annotated datasets*. In clinical environments, such datasets are often difficult to obtain due to privacy restrictions and limited annotations.

Federated learning (FL) has recently emerged as a promising framework for collaborative model training without requiring the cross-sharing of raw data. While effective, directly applying personalized FL may lead to **structural inconsistency** and **domain inconsistency** in medical imaging tasks, where heterogeneities always occur.

In veterinary medicine, the generalization of deep learning models remains limited. **To enable robust training of deep learning models, we propose a structure-aware, personalized federated learning framework, PAFL, for automatic vertebral heart score (VHS) estimation and canine cardiomegaly classification using X-rays.**

Methods



- **Structure-aware personalized federated learning**. The proposed framework combines anatomical keypoint modeling with shared and client-specific parameter decomposition. This drives robust learning under cross-institutional data heterogeneity assisted by a domain adapter module.
- **Quality-aware aggregation strategy.** The aggregation accounts for both dataset size and model quality based on relative model performance, improving training stability and model generalization across heterogeneous settings.

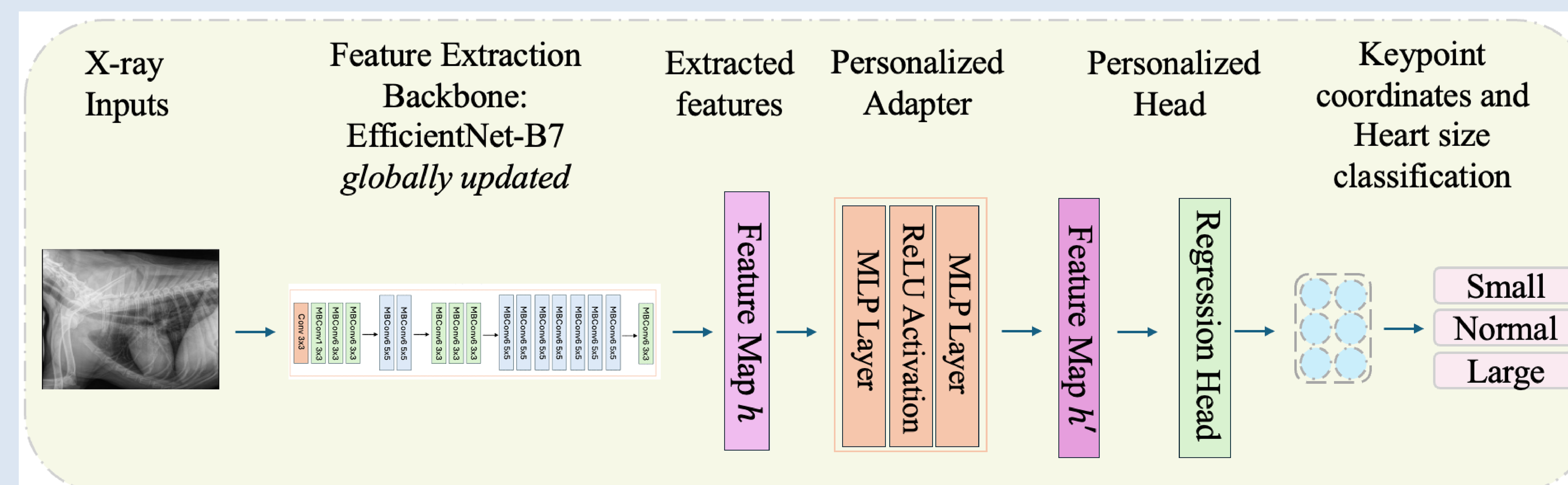
$$q_k = \frac{\max_j \mathcal{L}_j^{\mathrm{val}} - \mathcal{L}_k^{\mathrm{val}}}{\max_j \mathcal{L}_j^{\mathrm{val}} - \min_j \mathcal{L}_j^{\mathrm{val}} + \epsilon}, \qquad (15)$$

$$w_k = \frac{n_k\,(1+\beta q_k)}{\sum_{j=1}^{N} n_j\,(1+\beta q_j)}, \qquad (16)$$

- **Three canine X-ray datasets**. Our experiments work with three X-ray image datasets collected from three countries, China, Korea, and U.S., ensuring the robustness of model training towards heterogeneities of medical imaging.

Methods (continued)

Figure 2. Proposed client architecture for the VHS keypoint prediction model.



- **EfficientNet-B7 as model backbone for image feature extraction.**
- **A residual adapter handling data heterogeneity.** A multi-perception-based adapter was designed in addition to the original model to dynamically adapt client models to their local feature distributions when applying the shared global backbone weights.

Results

Figure 3. Model results of predicted keypoints after applying federated learning for three datasets (Blue – ground truths; Cyan – predictions).

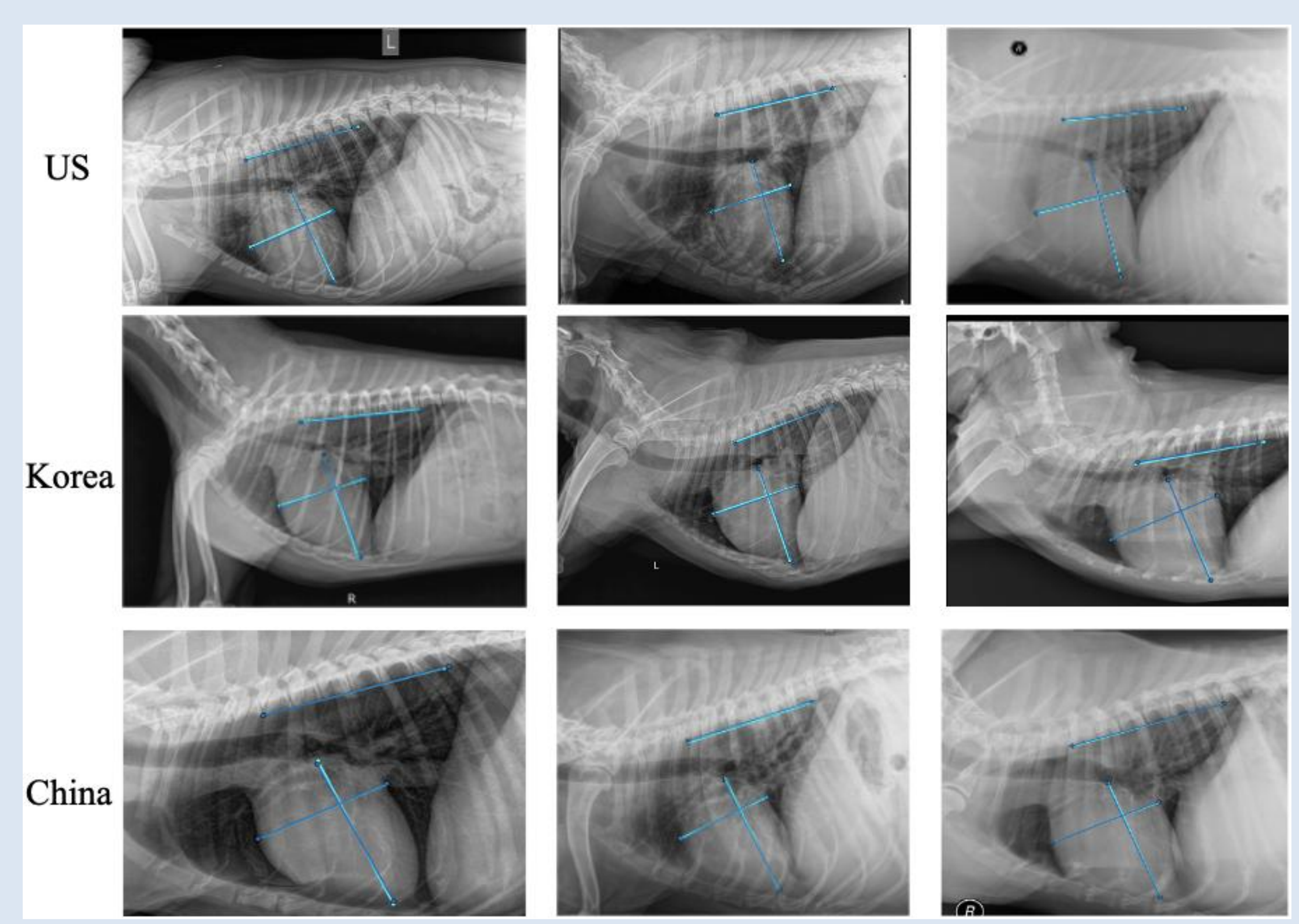


Table 1: Federated learning results on local datasets with 30 rounds of global training.

Federated Learning	Server 1 (China)		Server 2 (US)		Server 3 (Korea)	
	Validation	Test	Validation	Test	Validation	Test
Round 1 (initial)	77.5%	76.0%	72.0%	71.0%	74.3%	76.0%
Round 5	78.5%	78.0%	79.7%	80.6%	81.0%	82.0%
Round 10	82.5%	82.5%	78.6%	82.2%	81.0%	87.0%
Round 15	82.0%	85.5%	82.7%	83.5%	82.5%	85.5%
Round 20	81.3%	86.5%	80.6%	84.5%	82.0%	88.0%
Round 25	83.6%	**88.5%**	83.0%	**86.0%**	82.5%	**89.5%**
Round 30	81.0%	88.0%	81.0%	85.2%	84.0%	89.0%

Results (continued)

- The performance of individual client models peaked around round 25, reaching an accuracy of +86%.
- Figure 3 presents the alignment of the predicted key points with the original ground truths. The figure shows demonstrable overlaps between predictions and ground truths, indicating solid performance of the FL framework.

Ablation Studies

- The experiments focus on comparisons of raw and relative quality computation, personalized and non-personalized model aggregation, and traditional FedAvg and quality-aware FedAvg.
- Relative quality metrics across clients yields the highest accuracy score.
- Personalized models, aggregating only backbones, performed better than fully aggregated models that share weights across all model components.
- Quality-aware aggregation helps prevent premature convergence, enabling the model to reach better performance.

Conclusions

We present a structure-aware, personalized, and adaptive federated learning framework (PAFL) for automatic Vertebral Heart Score estimation from canine thoracic X-rays.

Extensive experiments on multi-institutional veterinary datasets demonstrate that our proposed PAFL framework achieves improved prediction performance by around 10% across all clients as compared to isolated training and outperformed conventional federated learning approaches.

It demonstrated the promise of applying a personalized and quality-aware FL framework to medical imaging tasks to address training data scarcity and heterogeneity.

References

Scan Me