CS482/682 Final Project Report Group 7
Domain Generalization for Chest X-Ray Imaging

Rena Bi (rbi4), Andrea Cheng (acheng41), Emily Guan (eguan2), Aprina Wang (awang100)

December 10, 2024

1 Introduction

Background In real-world healthcare settings, variations in imaging practices create domain shifts that can reduce the model performance when applied to new clinical sites. Challenges in medical data privacy make collecting comprehensive datasets from all domains infeasible, and labeling datasets can be extremely labor intensive. Our goal is to design a domain-generalizable classification model to identify 3 pathological states on chest X-rays: pneumonia, non-pneumonia disease, and normal.

Related Work Overall, previous work focuses on two aspects for domain generalization of medical imaging data: architecture and training scheme. A skin lesion classification model introduced latent feature learning, and another Chest X-Ray model created image and feature-level perturbations [1, 2]. Other work modifies the training scheme to mock domain generalization, like using semi-supervised training and different domains in episodic training [3, 4]. Thus, we can make modifications to the architecture and training style to enhance generalizability.

2 Methods

Dataset We used multiple datasets, each reflecting unique domains. The datasets contain labeled chest X-Rays of varying image sizes and resolutions from varying hospitals, X-Ray technology, and collection years. The datasets include Nigeria Chest X-Ray (Nigeria, 2024) [5], COVID-19 Radiography Database (Various Hospitals, 2020) [6], Chest X-Ray Image (Bangladesh, 2023) [7], RSUA Chest X-Ray Dataset (Indonesia, 2023) [8]. The datasets will be referred to as Domains A, B, C, and D respectively. We randomly sampled a subset of Domain B (due to training capacity limits), which consists of various datasets that do not overlap with other domains. Classes were approximately balanced in each dataset.

Architecture We used AlexNet as the backbone CNN architecture, and our models were initialized with pre-trained ImageNet weights [9].

Style Randomization Modules We implemented style randomization modules (SRMs) consisting of image-level (IL) and feature-level (FL) perturbations in AlexNet [1]. SRM-IL transforms the visual image style using a random sample of image pixel values. SRM-FL changes the style of intermediate feature representations using transformation parameters learned via a StyleNet composed of two small ConvNets. This StyleNet is trained separately with its own loss function at the beginning of each epoch on the same data as the overall classifier.

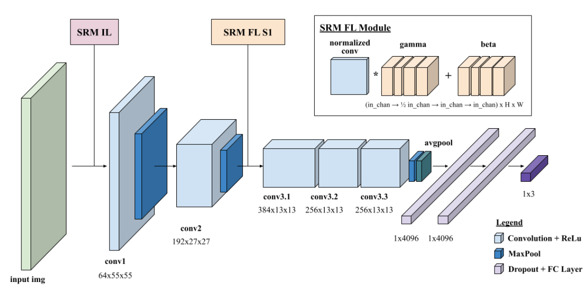

Figure 1: AlexNet Architecture with SRMs.

Training and Evaluation We trained models with cross entropy loss (softmax applied to the model's output layer) on Domains A and B, validated on Domain C, and tested on Domain D to evaluate the efficacy of our methods to enhance generalizability to new domains.

Episodic Learning We employed an episodic training scheme for our model with SRMs [4]. In episodic learning, the model trains and validates on meta tasks in different domains. We adapted and incorporated this paradigm into our SRM network by training StyleNet on Domain A and the final classifier on Domain B. Thus, the model learns FL transformation parameters encoding the style of A and applies them to Domain B, as illustrated in Figure 2.

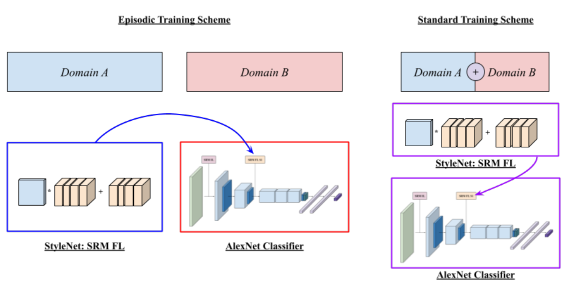

Figure 2: Episodic and Standard Training Scheme.

Evaluation A "negative control" model was trained without generalization schemas. A "positive control" model was trained, validated, and tested on a combination of all domains.

3 Results and Discussion

Table 1 shows the SRM model trained with episodic learning exhibited the best test accuracy, followed by the model with only SRMs. Both showed slight improvements over the negative control model. Validation accuracy in Figure 3 suggests that SRMs successfully enhance variability in training images (IL) and feature representations of the images (FL). Episodic learning further enhances generalizability by introducing generalization tasks via meta-tasks during training. Figure 4 shows the models most frequently classify images as "non-pneumonia disease," which may be harder to distinguish because this class includes multiple diseases, including specific diseases that were not included in the training dataset.

No.	Model	Accuracy (%)
I	SRMs (Perturbations)	72.0
II	SRMs + Episodic Learning	73.0
III	Negative Control	71.4
IV	Positive Control	72.9

Table 1: Accuracy Results for Different Models.

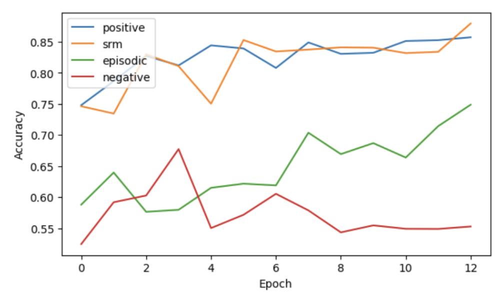

Figure 3: Validation Accuracy for Different Models.

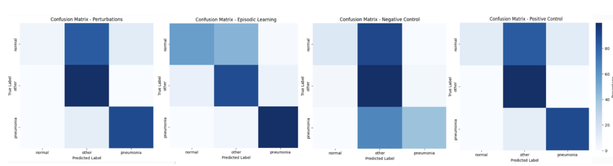

Figure 4: Confusion Matrix for Different Models: I, II, III, IV (left to right)

Our results show higher classification accuracy with the application of generalization schemas. Using a smaller subset of images from Domain A may have limited the models' potential learning ability. Additionally, more datasets from other domains would allow greater variation in episodic learning meta-tasks. Currently, StyleNet is only trained on one domain; however, previous literature has created meta-tasks from multiple domains [4]. We can implement this by using multiple domains to train the style network.

Source Code

https://github.com/acheng41/MLDLFA24-CXR/tree/main

References

[1] M. Zunaed, M. A. Haque, and T. Hasan, "Learning to Generalize towards Unseen Domains via a Content-Aware Style Invariant Model for Disease Detection from Chest X-rays," IEEE Journal of Biomedical and Health Informatics, 2024. doi:10.1109/JBHI.2024.3372999.

[2] H. Li, Y. Wang, R. Wan, S. Wang, T.Q.Li, and A. C. Kot, "Domain Generalization for Medical Imaging Classification with Linear-Dependency Regularization," IEEE Conference Proceedings, Rapid-Rich Object Search Lab, Nanyang Technological University, Singapore, 2022.

[3] R. Zhang, Q. Xu, C. Huang, Y. Zhang, and Y. Wang, "Semi-Supervised Domain Generalization for Medical Image Analysis,"in Proceedings of the IEEE 19th International Conference on Medical Image Analysis, 2022.

[4] C. Li, X. Lin, Y. Mao, W. Lin, Q. Qi, X. Ding, Y. Huang, D. Liang, and Y. Yu, "Domain generalization on medical imaging classification using episodic training with task augmentation," Computers in Biology and Medicine, vol. 137, no. 105144, 2021. doi: 10.1016/j.compbiomed.2021.105144.

[5] A. Musa, M. I. Adamu, H. A. Kakudi, and Y. Lawal, "Nigeria Chest X-ray Dataset," Kaggle, 2024. doi: 10.34740/KAGGLE/DSV/9370352.

[6] T. Rahman, M. Chowdhury, A. Khandakar (2020), "COVID-19 Radiography Database," Kaggle. Available: https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database

[7] M. A. Talukder, "Chest X-Ray Image," Mendeley Data, vol. 1, 2023. doi: 10.17632/m4s2jn3csb.1

[8] Radiologi RSUA, "RSUA Chest X-Ray Dataset," Mendeley Data, vol. 1, 2023. doi: 10.17632/2jg8vfdmpm.1

[9] PyTorch Documentation, "AlexNet Weights," [Online]. Available: https://pytorch.org/vision/main/models/generated/torchvision.models.alexnet.html#torchvision.models.AlexNet_Weights.