

# Spurious Correlation Mitigation in CXR Images via Reinforcement learning and Self-Supervision

Weichen Huang<sup>1</sup>, Kathleen M. Curran<sup>2</sup>

<sup>1</sup>*St Andrew's College Dublin, Dublin, Ireland*

<sup>2</sup>*School of Computer Science, University College Dublin, Dublin, Ireland*

## Abstract

In the medical domain, accurate interpretation of chest X-ray (CXR) images is critical for diagnosis and treatment decisions. However, deep learning models trained on large datasets can be susceptible to spurious correlations, leading to erroneous interpretations and potentially harmful decisions. This study aims to address this issue in the CXR domain by proposing the use of reinforcement learning techniques and semi-supervised training. These methods actively select relevant CXR data samples while mitigating the influence of spurious correlations. The results demonstrate the effectiveness of these approaches in improving prediction accuracy and decision-making performance compared to traditional data selection methods. This research contributes to the advancement of both technical state-of-the-art and clinical applications of deep learning in healthcare.

**Keywords:** Medical Imaging, Machine Vision, Reinforcement Learning, Self-Supervised Learning, Spurious Correlations

## 1 Introduction

The article tackles the challenge of spurious correlations in deep learning, specifically in medical domains such as CXR image classification [Nguyen et al., 2021]. Spurious correlations can lead to inaccurate diagnoses and treatment decisions. To address this issue, the article proposes a reinforcement learning-based data selection framework that integrates self-supervised training [Calude and Longo, 2017]. By actively selecting relevant data samples, the framework enhances the accuracy of the classification model and improves diagnostic reliability. Empirical evaluations demonstrate the effectiveness of this approach in mitigating spurious correlations and enhancing prediction accuracy. In the medical domain, this framework shows promise in improving the trustworthiness and effectiveness of diagnostic models, resulting in accurate predictions and improved patient outcomes [Nguyen et al., 2021]. This work offers several contributions in tackling spurious correlations in the medical domain. Firstly, it emphasizes the importance of developing solutions to mitigate these correlations, particularly in diagnostic tasks. Secondly, the work proposes a unified reinforcement learning (RL) based framework that incorporates an adaptable data relevance assessment system. This framework considers the inter-dependence between task-specific data relevance assessment and the target task, aiming to reduce the impact of spurious correlations. Lastly, the framework is evaluated on a public dataset of chest X-ray images, with a specific focus on the diagnostic task of pneumonia detection.

## 2 Related Work

Addressing spurious correlations in data is a critical challenge in AI and machine learning. Previous methods, such as feature selection and engineering [Deng et al., 2023], regularization techniques [Kirichenko et al., 2023], and causal inference [Cui and Athey, 2022], have limitations in complex scenarios. Our proposed approach

combines reinforcement learning (RL) and self-supervised learning (SSL) techniques to overcome these limitations. Feature selection and engineering, while effective, are manual processes that may not capture all relevant features in complex scenarios [Deng et al., 2023]. In contrast, our RL-based approach automatically learns the most relevant features and discards spurious correlations, reducing the need for manual intervention. Causal inference techniques rely on observational data and human intervention, limiting their applicability in complex scenarios [Cui and Athey, 2022]. Our RL-based method actively selects informative data points, reducing bias and spurious correlations without solely relying on observational data. By combining RL and SSL, our approach effectively addresses spurious correlations [Cui and Athey, 2022]. RL enables dynamic data selection based on rewards and penalties, while SSL reduces the influence of spurious correlations and improves generalization capabilities. Our approach overcomes the shortcomings of previous methods, enhancing the model’s feature selection, interpretability, and robustness [Deng et al., 2023, Kirichenko et al., 2023, Cui and Athey, 2022]. It provides an effective solution for mitigating spurious correlations in data, improving the reliability of AI and machine learning models.

### 3 Methodology

#### 3.1 Problem formulation

Spurious correlations refer to statistical relationships that appear to exist between variables but are not causally related. These correlations are often coincidental or arise due to confounding factors.

Let’s assume we have two variables,  $X$  and  $Y$ , and we denote their sample sets as  $S_X$  and  $S_Y$  respectively. A correlation between  $X$  and  $Y$  can be quantified using the Pearson correlation coefficient, denoted by  $r(X, Y)$ . The Pearson correlation coefficient measures the linear relationship between two variables and ranges from -1 to 1.

The presence of a spurious correlation means that  $X$  and  $Y$  exhibit a non-causal, coincidental association. In other words, their correlation arises due to the influence of an unaccounted third variable,  $Z$ , which affects both  $X$  and  $Y$  independently.

Mathematically, we can describe spurious correlations as follows:

Given  $X$ ,  $Y$ , and  $Z$ , the observed correlation between  $X$  and  $Y$ , denoted as  $r_{XY}$ , can be expressed as:

$$r_{XY} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

Where:  $X_i$  and  $Y_i$  represent individual data points from the sample sets  $S_X$  and  $S_Y$  respectively.  $\bar{X}$  and  $\bar{Y}$  are the mean values of the sample sets  $S_X$  and  $S_Y$  respectively.

However, if we consider the correlation between  $X$  and  $Y$  while controlling for the influence of  $Z$ , denoted as  $r_{XY.Z}$ , we would perform partial correlation. The partial correlation coefficient is used to measure the relationship between  $X$  and  $Y$  after accounting for the effect of  $Z$ .

The spurious correlation between  $X$  and  $Y$  is present if  $r_{XY.Z}$  is substantially different from  $r_{XY}$  (the observed correlation without considering  $Z$ ). In this case, the correlation between  $X$  and  $Y$  in the presence of  $Z$  disappears or becomes significantly weaker, indicating that the original correlation was spurious.

Our approach combines self-supervised learning and spurious feature correction to simultaneously train a task predictor  $f(x; w)$  and a data selection controller  $h(x; \theta)$ . Reinforcement learning is used to modify a parameter associated with the target task through the controller, aiming to maximize task performance. We utilize a recurrent neural network (RNN) for adaptability. The integrated approach incorporates self-supervised learning, emphasizes spurious feature correction, and operates within a meta-learning framework. In this section, we address spurious correlations by assessing data relevance using the task predictor and controller functions. The controller assigns data relevance scores to improve task performance over time, guided by task performance feedback.

### 3.2 Justification of Reinforcement Learning Approach

The proposed data relevance framework can be formulated as the following bi-level minimization problem:

$$\begin{aligned} & \min_{\theta} \mathbb{E}_{(x,y) \sim \mathcal{P}_{xy}} [L_h(f(x; w^*), y) h(x; \theta)], \\ \text{s.t. } & w^* = \arg \min_w \mathbb{E}_{(x,y) \sim \mathcal{P}_{xy}} [L_f(f(x; w), y) h(x; \theta)], \\ & \mathbb{E}_{x \sim \mathcal{P}_x} [h(x; \theta)] \geq c > 0. \end{aligned}$$

This problem can be restructured to allow sampling or selection based on controller outputs by considering the data  $x$  and  $(x, y)$  to be sampled from the controller-selected or -sampled distributions  $P^h(X)$  and  $P^h(XY)$ , with probability density functions  $p_h(x) \propto p(x)h(x; \theta)$  and  $p_h(x, y) \propto p(x, y)h(x; \theta)$ , respectively. Thus, reformulating to facilitate sampling or selection, we can rewrite the bi-level minimization problem as follows:

$$\begin{aligned} & \min_{\theta} \mathbb{E}_{(x,y) \sim p_{xy}^+} [L_h(f(x; w^*), y)], \\ \text{s.t. } & w^* = \arg \min_w \mathbb{E}_{(x,y) \sim p_{xy}'} [L_f(f(x; w), y)], \\ & \mathbb{E}_{x \rightarrow p_x^+} [1] \geq c > 0. \end{aligned}$$

The formulated data relevance assessment problem can be learned in a RL-based meta-learning framework. In this work, we outline a general RL-based meta-learning framework to learn adaptable data relevance assessment.

The proposed formulation can be modeled as a finite-horizon Markov decision process (MDP) [Puterman, 1990], with the controller interacting with, and influencing, an 'environment,' which contains the task predictor and the data used to train such a function. The MDP environment for this data relevance problem consists of the data from  $\mathcal{P}_X$  and the target task predictor  $f(-; w)$ . At time-step  $t$ , the observed state of the environment  $s_t = (f(-; w_t), \mathcal{B}_t)$  is composed of the target task predictor  $f(-; w)$  and a batch of samples  $\mathcal{B}_t = \{x_i\}_{i=1}^B$  from a train set  $\mathcal{D}_{\text{train}} = \{x_i\}_{i=1}^N$  from the distribution  $\mathcal{P}_X$ . If each MDP environment is defined as  $M_k$ , the distribution and task predictor within the environment can be defined as  $\mathcal{P}_{X,k}$  and  $f_k(\cdot; w_k)$ , respectively. However, in further analysis, we omit  $k$  from these expressions for notational convenience.

Reinforcement learning allows training of a controller to maximize a reward obtained based on controller-environment interactions, considered as an MDP. In RL, the MDP is represented as a 5-tuple  $(S, \mathcal{A}, p, r, \pi)$ .  $S$  is the state space and  $\mathcal{A}$  is the continuous action space.  $p: S \times S \times \mathcal{A} \rightarrow [0, 1]$  is the state transition distribution conditioned on state-actions, where  $p(s_{t+1} | s_t, a_t)$  represents the probability of the next state  $s_{t+1} \in S$  given the current state  $s_t \in S$  and action  $a_t \in \mathcal{A}$ .

The reward function is denoted by  $r: S \times \mathcal{A} \rightarrow \mathbb{R}$ , and  $R_t = r(s_t, a_t)$  denotes the reward given the current state  $s_t$  and action  $a_t$ . The policy,  $\pi(a_t | s_t): S \times \mathcal{A} \in [0, 1]$ , represents the probability of performing action  $a_t$  given the state  $s_t$ . The controller interacting with an environment creates a trajectory of states, actions, and rewards,  $(s_1, a_1, R_1, s_2, a_2, R_2, \dots, s_T, a_T, R_T)$ , where the subscript indicates the time-step.

The goal of the agent is to maximize the cumulative reward over a trajectory. The cumulative reward is the discounted sum of accumulated rewards starting from time-step  $t$ :  $Q^\pi(s_t, a_t) = \sum_{k=0}^T \gamma^k R_{t+k}$ , where the discount factor  $\gamma \in [0, 1]$  is used to discount future rewards. The objective of the controller is to learn a parameterized policy  $\pi_\theta$  that maximizes the expected return  $J(\theta) = \mathbb{E}_{\pi_\theta} [Q^\pi(s_t, a_t)]$ . The central optimization problem in RL can be expressed as:

$$\theta^* = \arg \max_{\theta} J(\theta),$$

where  $\theta^*$  denotes optimal policy parameters.

We propose to train the controller using RL, where the controller outputs sampling probabilities  $\{h(x_{i,t}, \theta)\}_{i=1}^B$  based on the input images. The action  $a_t = \{a_{i,t}\}_{i=1}^B \in \{0, 1\}^B$  leads to a sample selection decision for target task

predictor training if  $a_{i,t} = 1$ . The selection is done based on  $a_{i,t} \sim \text{Bernoulli}(h(x_{i,t}; \theta))$ . The policy  $\pi_\theta(a_t | s_t)$  is defined as:

$$\log \pi_\theta(a_t | s_t) = \sum_{i=1}^B h(x_{i,t}; \theta) a_{i,t} + (1 - h(x_{i,t}; \theta)) (1 - a_{i,t}).$$

In this formulation, the reward  $R_t$  is formulated based on the metric function that measures the performance of the target task,  $L_h$ .

The proposed reinforcement learning (RL) approach addresses the issue of data relevance by training a controller to select or sample data points based on their relevance to the target task. The RL agent learns an optimal policy by maximizing the expected return, effectively identifying and emphasizing the most relevant data during training. This iterative process improves the performance of the target task predictor by efficiently utilizing informative examples while minimizing the negative impact of noisy or irrelevant data. Ultimately, the RL-based data relevance assessment optimizes the data selection strategy, resulting in improved training efficiency and better task performance.

### 3.3 Optimizing the task predictor

We propose SimCLR [Chen et al., 2020], a self-supervised learning technique, to optimize the task predictor  $f(x; w)$ . SimCLR maximizes similarity between augmented views of the same image while minimizing similarity between views of different images, enabling robust and transferable image representations. After pre-training, we perform supervised fine-tuning on a small labeled dataset selected based on the controller’s scores. In fine-tuning, we utilize cross-entropy loss to adapt the pre-trained CNN to the labeled dataset. Additionally, we experiment with a purely supervised learning method, leveraging all labels from the training set.

### 3.4 Optimizing the controller model

The controller is optimized by minimizing the weighted metric function  $L_h : Y \times Y \rightarrow R_{\geq 0}$  on the validation set. The controller predicts lower data relevance scores for samples with higher metric function values, indicating lower task performance. A constraint prevents the trivial solution. The data relevance assessment problem is learned using a RL-based meta-learning framework modeled as a finite-horizon Markov decision process (MDP) [Puterman, 1990]. Reinforcement learning trains the controller to maximize the reward obtained through controller-environment interactions within the MDP framework. The MDP is defined by its state space  $S$ , continuous action space  $\mathcal{A}$ , state transition distribution  $p$ , reward function  $r$ , and policy  $\pi$ . The goal is to learn a parameterized policy  $\pi_\theta$  that maximizes the expected return  $J(\theta)$ .

### 3.5 Data Environment

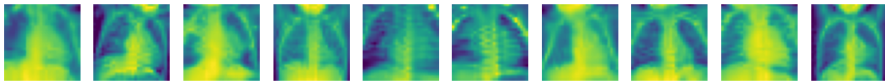


Figure 1: These images are examples of normal chest X-ray images from the PneumoniaMNIST dataset.

The proposed framework’s ability to detect spurious features was evaluated using the publicly available PneumoniaMNIST dataset [Yang et al., 2023], consisting of a total of 5856 chest X-ray images with binary labels for pneumonia diagnosis, split into train, validation and holdout sets. This dataset allows for evaluating the framework’s performance in pneumonia detection while demonstrating its applicability beyond a single modality, dataset, or task. Gaussian noise with random intensities and random obstructions were added to simulate real-world variations and occlusions in imaging data. These corruptions allowed assessing the framework’s ability to detect and mitigate spurious features in pneumonia detection using chest X-ray images, ensuring relevance and comparability to existing studies. The evaluation focused on the framework’s capability to detect

and exclude irrelevant data. A separate holdout set was used for evaluation, where samples were sorted based on controller predictions. The holdout set rejection ratio varied from 0% to 100% in 10% increments to assess the controller’s ability to detect spurious features. The remaining training data was used to train the controller.

### 3.6 Training

The controller is trained using the DDPG algorithm [Li et al., 2019] with empirically configured hyperparameters. An Alex-Net-style architecture [Yan et al., 2015] serves as the target task predictor, trained with cross-entropy loss and classification accuracy-based rewards. The controller underwent 100 episodes of training with a batch size of 64, while the task predictor underwent 100 epochs of training with the same batch size.

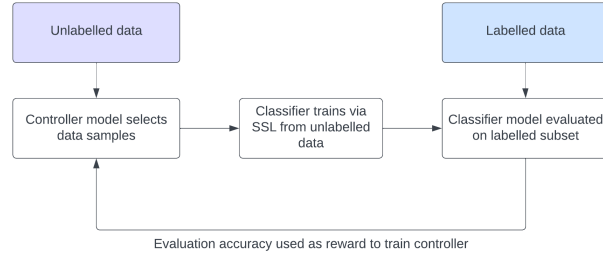


Figure 2: Diagram of the proposed framework. The controller takes in the data and outputs a ranking of data relevance. The task predictor takes in the data and outputs a prediction of the target task. The controller is trained to maximise the reward based on the task predictor output.

## 4 Evaluation

### 4.1 Evaluation Procedure

The goal is to evaluate the task-predictor accuracy based on reinforcement learning data selection and random data selection. We also evaluate the efficacy of self-supervised learning vs supervised learning of the task-predictor.

These methods are evaluated on the holdout dataset. The procedure for evaluation is the following:

- The trained controller model takes in the holdout set and generates the data-relevance rankings.
- A proportion (with value  $k$ ,  $0 < k < 1$ ) of the data samples are removed based on the ranking ( $nk$  data samples are removed for a holdout set of size  $n$ )
- The remaining data is used to evaluate the task predictor model which outputs evaluation metrics.

### 4.2 Results

Reject ratio	0.0	0.2	0.4
RL (Supervised)	<b><math>0.835 \pm 0.0251</math></b>	<b><math>0.836 \pm 0.0325</math></b>	<b><math>0.821 \pm 0.031</math></b>
RL (Unsupervised)	$0.8109 \pm 0.0317$	<b><math>0.824 \pm 0.0354</math></b>	$0.8027 \pm 0.0346$
No selection (Supervised)	<b><math>0.815 \pm 0.032</math></b>	$0.815 \pm 0.032$	<b><math>0.815 \pm 0.032</math></b>
Random selection (Supervised)	$0.810 \pm 0.0272$	$0.800 \pm 0.03$	$0.792 \pm 0.034$

Table 1: This table shows the accuracy of each technique on the holdout set. The standard deviation of each value is also shown.

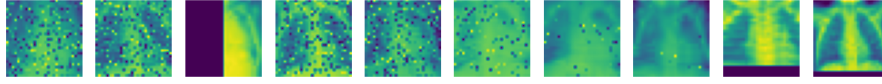


Figure 3: This is a spectrum of images from lowest score to highest score based on the controller model’s predictions. This shows that the model can highlight images that have little to no distortions (right side), and can classify when there are spurious correlations in the images (left side).

Comparing results from Table 1 in the classification task, both proposed RL-based algorithms show significant improvements over non-selective baselines. However, the unsupervised method has lower accuracy due to the lack of labeled data. While it still provides insights, its accuracy is generally lower than supervised methods. The peak accuracy for unsupervised learning was 82.4% at 0.2 rejection ratio, while supervised learning achieved 83.6% at the same ratio. Figure 3 provides a visual interpretation of the controller model’s predictions. Images on the left have low selection scores, while images on the right have high scores. The model effectively highlights images with minimal distortions (right) and detects spurious correlations (left). This empirical evidence supports the model’s ability to identify images with and without spurious correlations.

## 5 Conclusions and Future Work

The study concludes that RL techniques and self-supervised training effectively select relevant data samples from large unlabeled datasets given a small labeled dataset. The proposed RL-based framework addresses spurious correlations in deep learning by considering task-specific data relevance assessment and the target task. Unsupervised methods provide valuable insights but have slightly lower accuracy compared to supervised methods due to the lack of labeled data. The observation of peak performance before decreasing raises questions about the reasons behind this behavior, such as prediction variance, RL algorithm overfitting, and dataset limitations. Future work can explore different RL algorithms or self-supervised learning techniques for data selection, evaluate on larger datasets and complex models for scalability, and incorporate explanatory supervision for enhanced efficacy and interpretability. In the medical domain, the RL-based framework has potential to improve diagnostic models, reducing misdiagnosis risks and improving patient outcomes. Future work in medicine could focus on fine-tuning the framework for medical imaging tasks, incorporating domain-specific knowledge and evaluating on larger and diverse medical datasets.

## 6 Acknowledgements

We would like to express our gratitude to the MedMNIST team for providing their data, which was instrumental in conducting this research. Additionally, we acknowledge the contributions of previous works in this field that have paved the way for our study. Furthermore, we extend our thanks to our colleagues and collaborators for their valuable feedback and support throughout this project, as their insights have greatly contributed to the success of this research.

## References

- [Calude and Longo, 2017] Calude, C. S. and Longo, G. (2017). The deluge of spurious correlations in big data. *Foundations of science*, 22:595–612.
- [Chen et al., 2020] Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR.

- [Cui and Athey, 2022] Cui, P. and Athey, S. (2022). Stable learning establishes some common ground between causal inference and machine learning. *Nature Machine Intelligence*, 4(2):110–115.
- [Deng et al., 2023] Deng, Y., Yang, Y., Mirzasoleiman, B., and Gu, Q. (2023). Robust learning with progressive data expansion against spurious correlation.
- [Kirichenko et al., 2023] Kirichenko, P., Izmailov, P., and Wilson, A. G. (2023). Last layer re-training is sufficient for robustness to spurious correlations. In *The Eleventh International Conference on Learning Representations*.
- [Li et al., 2019] Li, S., Wu, Y., Cui, X., Dong, H., Fang, F., and Russell, S. (2019). Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 4213–4220.
- [Nguyen et al., 2021] Nguyen, T., Nagarajan, V., Sedghi, H., and Neyshabur, B. (2021). Avoiding spurious correlations: Bridging theory and practice. In *NeurIPS 2021 Workshop on Distribution Shifts: Connecting Methods and Applications*.
- [Puterman, 1990] Puterman, M. L. (1990). Markov decision processes. *Handbooks in operations research and management science*, 2:331–434.
- [Yan et al., 2015] Yan, L. C., Yoshua, B., and Geoffrey, H. (2015). Deep learning. *nature*, 521(7553):436–444.
- [Yang et al., 2023] Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., and Ni, B. (2023). Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data*, 10(1):41.