Proceedings of Machine Learning Research 299:1–24, 2025 Conference on Applied Machine Learning for Information Security

Causal Reinforcement Learning for Labelling Optimization in Cyber Anomaly Detection*

Susan Babirye[†1] and **Yu Gong**[‡1] and **Hideyasu Shimadzu**[§2] and **Konstantinos G. Kyriakopoulos**[¶ 1]

[1] *Wolfson School of Mechanical, Electrical and Manufacturing, Loughborough University, LE11 3TU Loughborough, United Kingdom.*

[2] *Department of Data Science, Kitasato University, Kanagawa, Japan.*

Editor: Edward Raff and Ethan M. Rudd

Abstract

The application of machine learning (ML) for cyber anomaly detection has attracted significant research attention. However, existing detection systems often face major challenges, including rigid feature discretisation, black-box classification, biased learning from confounded data, and lack of robustness, which collectively compromise interpretability, fairness, and predictive accuracy. Causal inference offers a robust approach to estimating intervention effects by isolating spurious correlations from true cause-effect relationships, crucial for reliable decision making under uncertainty. In contrast, reinforcement learning (RL) enables agents to learn optimal adaptive policies through interaction with dynamic environments. To address the aforementioned challenges, this work proposes a paradigm that leverages a RL framework to drive causal inference into the anomaly detection pipeline. Specifically, an RL agent is trained to optimize binning thresholds for confounded numerical features, guided by a reward function that incorporates both causal effect estimation and predictive accuracy. This approach enables the agent to learn feature discretisation strategies that avoid spurious associations induced by confounders, resulting in thresholds that are both causally aware and statistically effective. The optimized binning policy is then applied to transform the dataset, and a decision tree classifier is trained on the resulting unbiased features. This produces a model that is interpretable, robust to confounding, and sensitive to causal structures. Experimental results show that the proposed approach improves robustness and interpretability in unseen environments. This work highlights the potential of combining causal reasoning with adaptive learning to produce high-performance, transparent, optimal feature discretisation, and bias-aware cyber defence models.

Keywords: Confounding, Feature Labelling, Causal Inference, Reinforcement Learning, Decision Tree.

1. Introduction

Cyber defence systems rely on the detection of network traffic anomalies, which indicate the presence of attacks, such as botnets and data exfiltration (Choppadandi et al., 2021).

* Corresponding author
† S.Babirye@lboro.ac.uk, 0009-0002-0920-9946
‡ Y.Gong@lboro.ac.uk, 0000-0002-3985-8691
§ shimadzu.hideyasu@kitasato-u.ac.jp, 0000-0003-0919-8829
¶ K.Kyriakopoulos@lboro.ac.uk, 0000-0002-7498-4589

Intrusion Detection Systems (IDS) use labelled datasets for effective training, identifying deviations from expected network behaviour(Aparicio-Navarro et al., 2014).

Machine learning techniques such as supervised classifiers and unsupervised clustering have replaced traditional statistical methods in the detection of malicious traffic patterns (Jadidi et al., 2013). However, these methods still face three persistent problems: opaque decision boundaries and the black-box nature of neural detectors, hidden confounding bias between features, and significant degradation due to dataset shifts (Javaid et al., 2016), (Koh et al., 2021). Security analysts need to audit and explain why flows are flagged, while models can unfairly penalize valid traffic, resulting in false alerts (Casper et al., 2024; Malik, 2024; Ananth and Schisterman, 2017).

Decision trees and ensemble methods are popular due to their interpretability and high performance (Costa and Pedreira, 2023). They rely heavily on feature engineering, particularly for numeric feature data. Their efficacy depends on discretising continuous variables, which can be a heuristic or quantile-based approach (Liu et al., 2002). Poor binning can introduce bias, especially in class-imbalanced datasets (Huang et al., 2020). Defining optimal threshold values for bins is challenging and current methods may unintentionally reinforce biases (Mariooryad and Busso, 2015).

In this work, we propose a causal reinforcement learning (RL) framework that learns optimal threshold values for feature discretisation, while simultaneously guarding against confounding bias, with the goal of reducing false classifications and enhancing model explainability in intrusion detection tasks. Feedback on the effectiveness of the learned binning policy, specifically how well it distinguishes malicious traffic from benign traffic, is provided to the RL agent as a reward signal. The agent's action involves assigning a given feature value to a particular bin.

To incorporate causal reasoning into the training process, we apply the backdoor adjustment method using a known confounder variable (Fang et al., 2024). This allows us to estimate the causal effect of assigning a feature to a specific bin on the likelihood that a traffic flow is labelled malicious. The estimated causal effect is integrated into the reward function, embedding causal inference within the RL optimization loop.

The agent is rewarded positively when its binning strategy produces a stable and significant causal effect and high predictive accuracy. This encourages policies that align with true causal relationships while discouraging strategies that exploit spurious correlations. The causally-aware bins produced by the agent are used to train a decision tree classifier, resulting in a final model characterised by interpretable decision rules that can be inspected and audited by analysts.

The contributions in this paper are as follows.

- A causal reinforcement learning approach that learns optimal discretisation thresholds for numerical features by embedding do-calculus causal estimation into the reward function, leading to interpretable and causally robust representations for cyber anomaly detection.

- An interpretable detection framework that integrates causal fairness principles into a decision tree classifier, resulting in high predictive performance, transparency, and explainable decision making.

- A detailed empirical evaluation using standard metrics including recall and false positive rate to assess the effectiveness of the proposed approach. The causal RL framework is benchmarked against traditional quantile-based binning strategies, demonstrating improved robustness to confounding, lower false positive rates, and enhanced interpretability in both training and unseen test environments.
- Comprehensive empirical evaluations show that the proposed causal RL-driven binning strategy improves malicious traffic detection by up to 10% and benign traffic by 15% compared to the numerically based classifier model, when evaluated on unseen data. Additionally, it achieves higher balance of recall and precision than the fixed-bins based model.

The rest of the sections of this paper are as follows: Section 2 reviews related work on causal learning and cyber anomaly detection. The preliminaries are written and the problem and details of our causal RL framework are formalized in Sections 3 and 4, respectively. The experimental setup, result analysis, and discussions are presented in Section 5. The paper concludes in Section 6.

2. Related work

2.1. Causal Inference

The authors in (Lu et al., 2023) introduced causal inference using Bayesian networks and do-calculus to differentiate between causal and statistical reasoning. Causal inference is a deductive process, whereas statistical inference is an iterative process that involves deduction and induction (Pearl, 2009).

In (Zeng et al., 2022), the authors address the challenge of diagnosing the causal relationship between attack and traffic features using existing ML detection methods. They propose a distributed denial-of-service framework based on causal reasoning, using counterfactual diagnosis for attack detection.

The study in (Dasgupta et al., 2019), presents a model-free meta-learning algorithm for causal reasoning, utilizing RL to generate and use various data types, including counterfactual predictions, active intervention, and inferring causal inferences from passive observations.

2.2. Causal Reinforcement Learning

The integration of causal inference into RL has gained popularity for enhancing robustness, decision-making, and sampling effectiveness in confounding variable environments. Pearl's do-calculus framework (Pearl, 2009), provides a theoretical foundation for RL agents to learn cause-effect relationships, rather than simple correlations, in interventions.

The deconfounded optimistic value iteration (DOVI) algorithm was proposed in (Wang et al., 2021), to enhance online sampling efficiency by identifying causal effects and estimating action value functions, adjusting for confounding bias when observed.

Causal RL goes beyond associational learning by using interventions with 'do-operator' ($do(T = t)$) to separate correlation from causation, useful in biased action selection mechanisms or latent confounders. Causal RL addresses challenges such as sample inefficiency,

biased reward estimation, and policy generalization. To address hidden confounders, the authors in (Zhu et al., 2024), corrected biased action distributions by learning structural causal models.

The work in (Zhang et al., 2019), further demonstrated that causal structure improves a policy's robustness to changes in distribution. To enable the agent to represent the cause and effect of its own actions and learn from experience, the authors of (Purves et al., 2024) integrated a causal reward model into the RL environment.

Causal RL has not been widely adopted in cyber security, where confounding features of the network and protocol are common, despite its relevance in regulated fields like robotics and healthcare.

2.3. Dataset Labelling

Supervised cyber anomaly detection relies on optimized discretised bins for continuous features, but obtaining them is challenging. Early datasets such as DARPA / Lincoln Labs used expert manual notation, but were criticized for synthetic nature (Lippmann et al., 2000). The CIC-IDS and CIC-DDoS datasets (Sharafaldin et al., 2018), improved realism by replaying malware, but still rely on knowledge of the time window under attack.

The Snorkel framework (Ratner et al., 2020), uses data programming to encode domain heuristics as labelling functions, improving recall in imbalanced botnet datasets. Currently RL models have been applied to cyber defence problems like moving target defence, dynamic firewalls, intrusion response, and network hardening (Abdel-Basset et al., 2023; Kheddar et al., 2024). However, research on exploring RL for learning feature discretisation thresholds with causal inference remains limited.

This work integrates causal inference into the reinforcement learning process for feature discretisation, thus guiding the binning of numerical features, which is a critical preprocessing step in tree-based anomaly detection systems. By incorporating causal fairness constraints during the optimization of binning thresholds, the proposed approach enhances both the interpretability and robustness of the resulting models. To our knowledge, this is the first study to combine causal reinforcement learning-based binning with interpretable decision tree classifiers in the context of network traffic analysis for cybersecurity applications.

3. Preliminaries

3.1. Causal Inference

Causal inference improves probabilistic modelling by enabling interventions and cause-effect relationship estimation. It supports counterfactual reasoning and robust conclusions about system variables, crucial in anomaly detection where spurious correlations can lead to misleading decisions (Lu et al., 2023).

A foundational principle in causal reasoning is Reichenbach's Common Cause Principle (Hitchcock and Rédei, 2020), which states that if two variables T and Y are statistically not independent ($T \not\perp\!\!\!\perp Y$) in other words, they share statistical information, there exists a third variable C that causally influences both (C may coincide with T or Y), as shown in Figure 1.

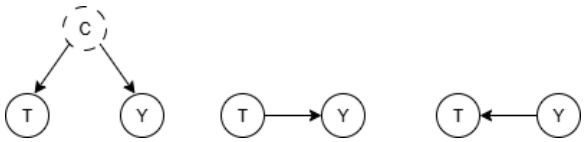


Figure 1: Reichenbach's common cause principle established a link between statistical properties and causal structures. T and Y are statistically independent, conditioned on C.

Conditioning on C can render T and Y statistically independent ($T \perp\!\!\!\perp Y \mid C$), highlighting the need to distinguish between direct associations and those induced by confounding.

Causal structures can be represented using directed acyclic graphs (DAGs), where edges represent assumed causal relations between variables (Figure 2(a)). The joint probability of the triplet can be factorized as

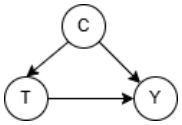
$$P(Y,T,C) = P(C)P(T|C)P(Y|T,C) \tag{1}$$

Within this framework, the confounder (C) simultaneously affects both the treatment T and the outcome Y, making it difficult to estimate causal effects directly from observed data. To overcome this, interventional / change distributions (Peters et al., 2017), are defined using Pearl's do-operator, denoted by $do(T = t)$, which simulates an intervention by setting the treatment variable T to a constant value t, independent of its natural causes and creates a mutilated model as depicted in Figure 2(b). The intervention involves giving the entire population the same treatment t, while conditioning on $T = t$ focuses on the subset of the population for those who receive treatment t (Neal, 2020). When we intervene in the treatment, the truncated factorization (1) gives the following (Lu et al., 2023), (Oliveira et al., 2021):

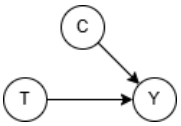
$$P(Y,C|do(T = t)) \triangleq P(Y,C|do(t)) \tag{2}$$

$$P(Y,C|do(t)) = \sum_{C_i} P(Y|t,C_i)P(C_i) \tag{3}$$

where i are the different scenarios of C.



(a) A causal structure where condition C is a common cause of treatment T and target variable Y.



(b) A mutilated model where $do(t)$ replaces all functions f_i determining T with the constant function $T = t$.

Figure 2: Causal structural relationships with a confounder influencing the treatment feature and the outcome.

3.2. Decision Tree

Decision trees are supervised learning models valued for their accuracy, interpretability, and robustness (Costa and Pedreira, 2023). They offer additional benefits, including low computational cost, tolerance to missing data, and native support for mixed data types. Structurally, decision trees are binary trees that comprise internal nodes and leaves. Internal nodes perform logical tests, typically of the form $'attribute \leq value'$ for numerical features and $'attribute = value'$ for categorical ones, resulting in binary branching. The leaves contain either class labels or constant values and represent the predictions of the model (Ying et al., 2015).

Inference involves routing an observation from the root to a leaf via a sequence of deterministic splits, making the decision process fully transparent (Rudin, 2019). With the increasing adoption of machine learning in automated systems, there is growing interest in explainable models, with decision trees being a primary example due to their inherent transparency (Roscher et al., 2020).

3.3. Reinforcement Learning

RL (Sutton et al., 1998), is a paradigm of machine learning for solving control tasks or decision problems by building agents that interact and learn from the environment through trial and error and receiving rewards as feedback. The RL process shown in Figure 3 consists of state, action, reward, and next state iteratively.

The states or observations are the information that the agent gets from the environment. The action space is the set of all possible actions in an environment, and the reward is the feedback from the agent. The agent's role is to learn taking actions that maximize the expected cumulative reward in an environment, often modelled as a Markov decision process (MDP). MDP is a mathematical framework that is used to describe the decision-making process in an environment. The agent needs only the current state to decide what action to take, not the history of all previous states and actions.

The MDP decision process is defined as $M = (S, A, P, R, \gamma)$, where S and A are sets of states and actions. P denotes the transition probability of the state $P(s^{'}|s, a)$ which defines the probability of transitioning from the state s to the state $s^{'}$ by taking action a. R is the reward function, $R(s, a)$, which gives the immediate reward after performing the action a in state s. γ is the discount factor, $\gamma \in [0, 1]$, that weighs the importance of future rewards.

Proximal policy optimization (PPO) is a deep learning technique used for quantitative security evaluation of large-scale enterprise networks (Schulman et al., 2017) and (Engstrom

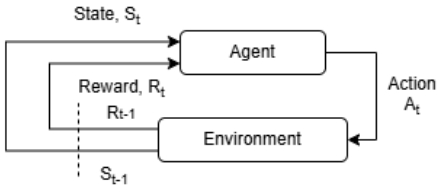


Figure 3: RL process

et al., 2020). This policy gradient algorithm improves the stability of the training by avoiding large policy updates and conservatively updating the policy (Jia et al., 2024).

The PPO clipped surrogate objective ensures stable learning by preventing excessively large updates:

$$L^{\text{CLIPP}}(\theta) = \hat{\mathbb{E}}_t \left[\min(r_t(\theta)\hat{A}_t), \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t) \right] \tag{4}$$

PPO optimizes a policy $\pi_\theta(a|s)$ using clipped objective functions to ensure stable updates, where $r_t(\theta) = \frac{\pi_\theta(a|s)}{\pi_{\theta_{old}}(a|s)}$ is the probability ratio, A_t is the advantage estimate, ϵ is the clipping parameter.

4. Methodology

4.1. Problem Formulation

Let $M = \{(\mathrm{c_i}, \mathrm{t_i}, \mathrm{y_i})\}_{i=1}^{n}$ be a dataset of network traffic flows, where $c_i \in \mathbb{R}^d$ represents observed features, $t_i \in \mathbb{R}$ is a specific numerical value feature that we aim to discretise and $y_i \in \{0, 1\}$ is the binary class label indicating whether the flow is benign or malicious attacks.

Traditional classification approaches often discretise t_i using fixed values (e.g., quantiles or equal-width binning), then pass the transformed features to a classifier f_θ. However, when the value in c_i potentially influences both the binning of t and the label, this process can introduce confounding bias, obscure meaningful thresholds, reduce accuracy and interpretability, and lead to suboptimal classification performance due to rigid or misaligned bins.

The objective of this work is to learn optimal binning thresholds $b = \{b_1, b_2, ..., b_k\}$ for feature t such that the resulting classifier $f_\theta(c, Bin_b(t))$ achieves high predictive accuracy, causal fairness by removing spurious associations between treatment $T = Bin_b(t)$ and outcome Y.

4.2. Causal Discovery of features

Causal discovery is a model that illustrates the relationship between features and a target output. There are several tools that can be used for causal discovery, for example, NOTEARS (Wang et al., 2023), PC-algorithm (Le et al., 2016), causal-learn or correlation-based (Zheng et al., 2024), causal discovery can be used. Some direct causal factors or features (parents of the attack node) that are direct causal influences on the attack outcome are enumerated using the PC technique and correlation causal findings. From the Canadian Institute of Cyber Security dataset, CSE-CIC-IDS, we obtained the causal relationship between the mean forward inter-arrival (Fwd IAT Mean) which is the mean tile between forward packets, flow duration and the label (benign or malicious).

- FWd IAT Mean: This indicates abnormal burst patterns. The Inter-Arrival Time between packets can reveal timing-based attacks.
- Flow Duration (FD): The total time a network flow lasted from the arrival of the first packet to the last packet in that flow. The unusually long or short connection times can indicate various attacks.

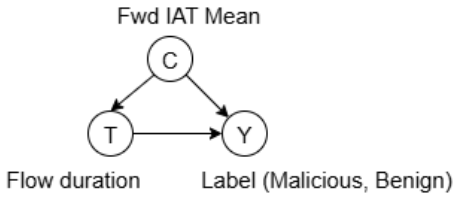


Figure 4: Typical causal structure of features in a dataset having a confounder (Fwd IAT Mean), treatment feature (flow duration) and the outcome (label).

- Attack: Used to indicate the characteristic of the label of traffic noted as benign (0) or malicious (1).

The causal paths leading to attack success are obtained.

In Figure 4, the correlation between the variables and the label in the path of direct causal effect Fwd IAT Mean → Label, focusing on the Fwd IAT Mean and the traffic label. The distribution of the Fwd IAT mean in the dataset influences the flow duration and the success of the attack. Timing anomalies can lead to unusual connection durations, making it necessary to control for confounding variables.

Long flow durations may correlate with attacks due to persistent connections, but timing patterns may confound this. Controlling for confounding variables is, therefore, necessary to determine the true causal effect of flow duration on attacks.

The common cause Fwd IAT mean creates a confounding bias, making flow duration appear less or more strongly associated with attacks than it is. The study uses backdoor adjustment with C as a common cause to estimate causal effects from T to Y.

4.3. Binning strategy

A decision tree classifier splits numerical feature values such as c or T into a tree structure for predicting the outcome (Y) label of the traffic. Numerical feature decision trees are very accurate but suffer from a lack of robustness when used on unseen data. These numerical values can be transformed into discrete feature labels with a process called discretisation to improve the robustness of the classifier (Kotsiantis and Kanellopoulos, 2006). The discretisation process involves classification of the numerical feature values into labels L_C and L_f, essential for categorizing numerical variables into bins. For this work, quantile-based binning is used to divide numerical data into bins with a similar number of observations, ensuring balance in skewed datasets and reducing variance in estimation, using percentiles to determine the boundaries of the bins (Labovich, 2025). In this work, we used 3 quantile bins of which 33.3% of the data points are contained from (5). This ensures balance in skewed datasets to reduce the variance in estimation. The decision tree classifier is trained with fixed bins and the performance observed.

$$Q\_i = \text{Percentile}(\text{data}, i * (100/n\_bins) \tag{5}$$

for $i = 1, 2, ..., n\_bins - 1)$. The threshold assignment function is:

$$\text{assign\_bin(x, thresholds)} = min\{i : x \leq \text{threshold}_i\} \tag{6}$$

4.4. Backdoor adjustment of dataset features

Fixed quantile binning improves robustness but sacrifices granularity and the causal structure of the features. In this work, static binning is replaced by causally aware binning optimized with reinforcement learning.

Causal inference improves decision making under uncertainty by understanding cause-effect relationships between actions and outcomes, distinguishing true effects from spurious correlations using the backdoor adjustment formula (7), which estimates the causal effect of treatment T on outcome Y, by adjusting for confounder L_C (Pearl, 2009).

$$P(Y = 1|do(T = t)) = \sum_{L_C}(P(Y = 1|T = t, L_C = c) * P(L_C = c)) \tag{7}$$

where Y is the label (1 for Bot, 0 for Benign), T is the treatment (flow duration bin), L_C is the confounder bins (Forward IAT bin) and $do(T = t)$ is the intervention to set a value t to the treatment random variable T.

The $do-operator$ ($do(T = t)$) gives the true effect of action T (discretising the numerical values of the flow duration into bins) on the outcome (Y) by controlling for the confounding effect (C).

With the backdoor adjustment formula, the causal quantities $P(Y = 1|do(T = 1), L_C = c) * P(L_C = c)$ and $P(Y = 1|do(T = 0), L_C = c) * P(L_C = c)$ are estimated to obtain the average treatment effect (ATE) (Neal, 2020):

$$\begin{aligned}\text{ATE}_{L_C} = \sum_{L_C}(&P(Y = 1|do(T = 1), L_C = c) * P(L_C = c) \\ &- (P(Y = 1|do(T = 0), L_C = c) * P(L_C = c)\end{aligned} \tag{8}$$

Expressed in expectation form:

$$\text{ATE}_{L_C} = \sum_{L_C}(\mathbb{E}\left[Y|do(T = 1)\right] - \mathbb{E}\left[Y|do(T = 0)\right]) \tag{9}$$

4.5. Causal Reinforcement Learning framework

A Markov Decision Process (MDP) is defined in the binning space: The state space, s_t, is the numerical values of the confounder c_i (Fwd IAT Mean) and the treatment t_i (flow duration) features. In each observation, the agent observes c_i and t_i. The PPO agent assigns treatment values t_i, to a fixed number of k bins with the action space ($a_t \in \{0, 1, ..., k - 1\}$) and receives a reward based on how well that decision helps identify malicious traffic.

The goal is to maximize the reward if the agent's bin assignment results in a meaningful difference in the outcome. To align this with causal reasoning, the reward is informed by the average treatment effect (ATE) for the current instance and the true label ($Y = 1$ for malicious, $Y = 0$ for benign). The reward ($\text{ATE}_{backdoor}$) is computed using the backdoor formula (10). The next state space is the next treatment-confounder pair (t_{i+1}, c_{i+1}) and the episode ends after completing all samples in the dataset.

Under the backdoor criterion: The RL agent learns the bin boundaries for a numerical treatment variable such that placing values into these bins leads to the maximized treatment effect (ATE) under:

$$\text{ATE}_{backdoor} = \sum_{L_C} (\mathbb{E}\left[Y|T=1, L_C=c\right] - \mathbb{E}\left[Y|T=0, L_C=c\right]) * P(L_C=c) \tag{10}$$

This can be interpreted with the question: what is the expected difference in the outcome if we do treat (assign to bin =1) versus if we do not treat (assign to bin = 0), keeping the confounder constant?

ATE measures how important treatment is for this particular individual (flow duration). If the ATE is large, the agent's binning decision has a great impact. The ATE tells the RL agent how useful its binning decision is in terms of changing the outcome. A high ATE_{L_C} indicates a strong causal link and a low ATE_{L_C} shows a weak or harmful bin decision. This makes the RL policy causally aware. The reward r_i given to the agent for its action is:

$$r_i = \begin{cases} + \left|\text{ATE}_{L_{C_i}}\right| & \text{if assigned bin corresponds to T = 1} \\ - \left|\text{ATE}_{L_{C_i}}\right| & \text{if assigned bin corresponds to T = 0} \end{cases}$$

where L_{C_i} = the features of the i-th sample of the confounder, $\text{ATE}_{L_{C_i}}$ = the estimated individual treatment effect using the backdoor model.

The higher the causal effect, the more this treatment matters, thus a bigger reward. The reward is positive if the bin represents a treated ($T = 1$) action, encouraging high-effect interventions. The reward is negative for the control ($T = 0$), so the agent learns to prefer bins that maximize causal impact. The absolute value $|\text{ATE}_{C_i}|$ ensures that the magnitude of the treatment effect drives learning, regardless of the direction.

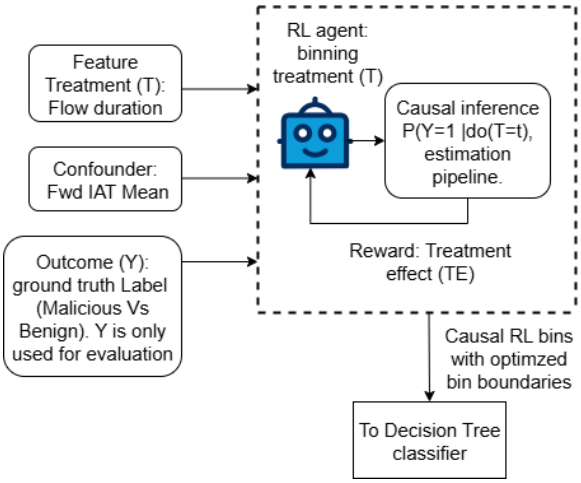


Figure 5: Step-by-step RL causally-guided feature discretisation process

In this work, a PPO agent learns treatment binning policies $\pi_\phi(s)$ that separate flows in a causally meaningful and classification-effective way tailored to real effects from a confounder

shown in Figure 5. Static binning is replaced with an RL agent that learns treatment assignments optimized for the true relevance of the outcome, guided by causal relationships. The learned discrete causalRL bins maintain the causal structure of the numerical feature to improve the accuracy and robustness of the decision tree classifier.

5. Results

The results in this section show the evaluation performance of the decision tree classifier across three strategies: raw numerical features (Numerical), fixed quantile binning (Fixed bins), and RL-based causal binning (CausalRL bins) in the prediction of malicious traffic with the CSE-CIC-IDS2018 dataset with 1,034,236 instances (748,045 benign and 286,191 malicious) (for Cybersecurity, b), and the CSE-CIC-IDS-2017 dataset with 684,483 instances (434,844 benign and 249,639 malicious) (for Cybersecurity, a) (Sharafaldin et al., 2018) of the Canadian Institute of Cyber security with compromised benign and bot instances. The CSE-CIC-IDS2018 dataset was split into the training set 60% (448826 benign, 171715 malicious) and the validation set 20% (149609 benign, 57238 malicious). The test set 20% (86968 benign, 49928 malicious) was obtained from the CSE-CIC-IDS2017 evaluation dataset.

The performance of the decision tree classifier is evaluated using six metrics in Table 1, to provide a complete view of accuracy and robustness in three scenarios (Beechey et al., 2021).

Table 1: Key Performance Metrics

Metric	**Decision Tree scenarios**
Accuracy	Correct predictions of a ML model, or intrusion detection system.
Recall	Number of positive instances that were correctly classified as positive amongst all predicted instances.
Precision	Fraction of all actual positive classifications amongst all predicted positive instances.
F1-score	Harmonic mean of both precision and recall metrics to the measure of a test's accuracy.
Specificity	Fraction of all actual negative classifications amongst all predicted instances.
ROC AUC	Measures the classifier's ability to distinguish between malicious and benign classes. Higher AUC represents better performance.

5.1. Result Analysis

5.1.1. Decision Tree classifier with one input feature

Table 2, Figure 6 and Figure 7 describe the performance of the decision tree classifier scenarios in the validation set and are provided with only the flow duration feature as input.

Table 2: Key Results Comparison of the decision tree classifier scenarios with one feature (only flow duration) on Validation set.

Metric	Validation set: one feature		
	Numerical	***Fixed bins***	***CausalRL bins***
Accuracy	0.9496	0.6732	0.8832
Recall (malicious)	0.97	0.82	0.99
Precision (malicious)	0.86	0.45	0.70
F1-score (malicious)	0.91	0.58	0.82
Specificity	0.94	0.62	0.8408
ROC AUC	0.9865	0.7175	0.9182

The classifier using raw numerical features yielded the highest performance, achieving an accuracy of 94.96%, an F1 score of 0.91, and a receiver operating characteristic (ROC) area under curve (AUC) of 0.9865, correctly identifying 55,594 out of 57,238 malicious flows and 140,825 out of 149,609 benign flows.

In contrast, the fixed binning approach demonstrated substantially lower effectiveness, with an accuracy of 67.32%, an F1 score of 0.58, and a ROC AUC of 0.7175, correctly classifying 46,739 malicious and 92,515 benign instances.

The proposed causalRL binning method significantly improved the fixed binning baseline, achieving an accuracy of 88.32%, an F1 score of 0.82, and an ROC AUC of 0.9182, with 56,906 correctly predicted malicious flows and 125,788 correctly identified benign flows. These results highlight the potential of causally informed RL to improve both predictive accuracy and robustness in security-sensitive classification tasks.

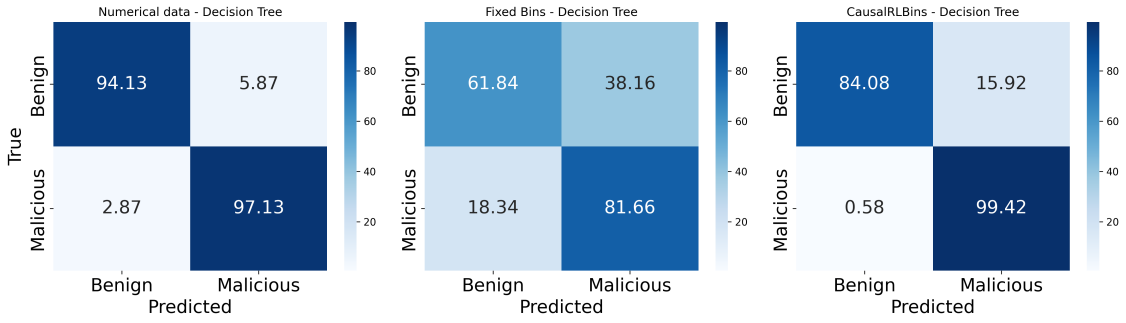
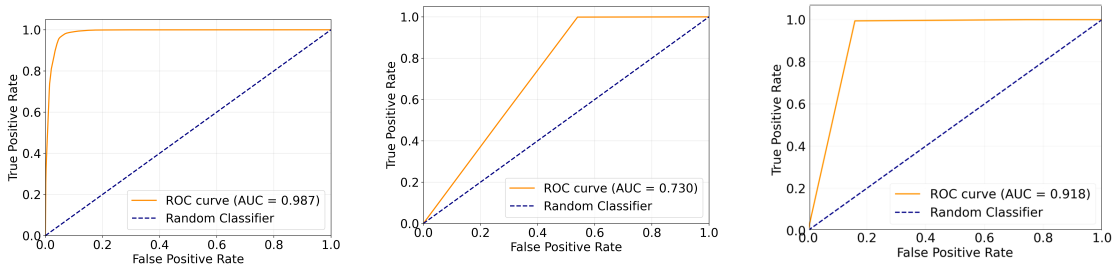


Figure 6: Confusion matrices for three models with one input feature (flow duration) evaluated on the validation dataset

The generalization performance of the decision tree classifier was assessed in a domain shift setting, where models trained on one dataset were evaluated on an unseen test set using only the flow duration feature as input. Table 3, Figure 8 and Figure 9 present the results of this evaluation.



(*a*) Decision Tree.

(*b*) Binned features Decision Tree.

(*c*) Causal bins decision Tree.

Figure 7: Decision tree classifier ROC for one input feature (flow duration) evaluated on the Validation dataset.

Table 3: Key Results Comparison of the decision tree classifier scenarios with one feature (only flow duration) on the test set

Metric	**Test set: one feature**		
	Numerical	***Fixed bins***	***CausalRL bins***
Accuracy	0.6259	0.2821	0.5451
Recall (malicious)	0.03	0.35	0.11
Precision (malicious)	0.36	0.21	0.24
F1-score (malicious)	0.06	0.26	0.15
Specificity	0.97	0.24	0.80
ROC AUC	0.3546	0.2842	0.5803

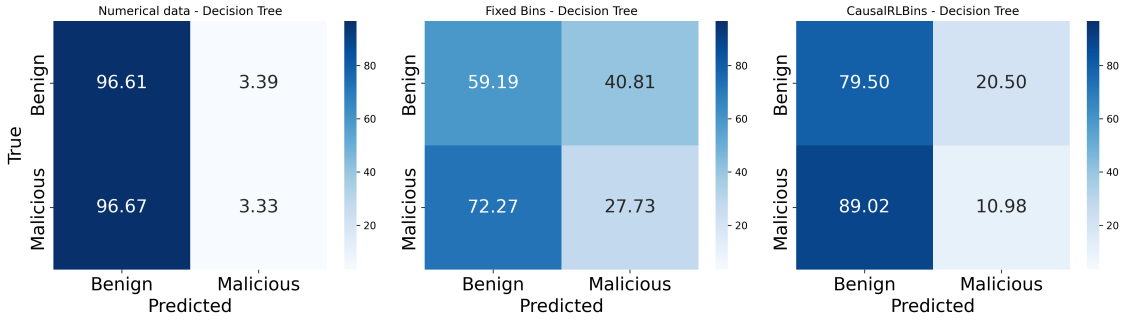
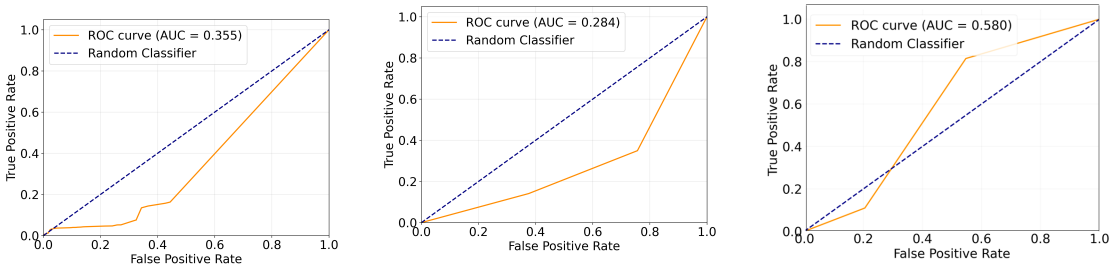


Figure 8: Confusion matrices for three models with one input feature (flow duration) evaluated on the test dataset



(*a*) Decision Tree.

(*b*) Binned features Decision Tree.

(*c*) Causal bins decision Tree.

Figure 9: Decision tree classifier ROC for one input feature (flow duration) evaluated on the test dataset.

The classifier using raw numerical values yielded poor generalization, with an accuracy of 62.59%, a notably low F1 score of 0.06, and a ROC AUC of 0.3546, correctly identifying 1,661 of 49,928 malicious flows and 84,023 of 86,968 benign flows.

The fixed binning strategy further deteriorated performance, achieving only 28.21% accuracy, an F1 score of 0.26, and a ROC AUC of 0.2842, with 17,452 true positives and 21,169 true negatives.

In contrast, the causally guided RL (causalRL) binning method demonstrated improved generalization over the fixed binning approach, reaching 54.51% accuracy, an F1 score of 0.15, and an ROC AUC of 0.5803, while correctly predicting 5,483 malicious and 69,143 benign flows.

These findings suggest that causalRL binning enhances out-of-distribution robustness relative to conventional binning methods, although performance remains suboptimal under significant distributional shifts.

5.1.2. Decision Tree classifier with two input features

Table 4, Figure 10 and Figure 11 describe the performance of the decision tree classifier scenarios in the validation set and are provided with two input features (flow duration and Fwd IAT mean).

When the decision tree classifier is trained using two input features, flow duration and Fwd IAT Mean, its performance improves substantially across all evaluation metrics. The numerical feature representation yields near-optimal results, with an accuracy of 99.25%, an F1 score of 0.99, a ROC AUC of 0.9978, and high classification correctness for malicious (56,963/57,238) and benign (148,336/149,609) flows.

Although the fixed binning approach performs moderately well, it results in a reduced accuracy of 82.07%, an F1 score of 0.76, and an ROC AUC of 0.9438, with 57,170 true positives and 112,592 true negatives. The causalRL bins achieve slightly higher performance than the fixed binning, with an accuracy of 85.45%, an F1 score of 0.79, and a ROC AUC of 0.9255, correctly identifying 57,172 malicious and 119,570 benign flows.

Table 4: Key Results Comparison of the decision tree classifier scenarios with two features (flow duration and Fwd IAT Mean) on the Validation set

Metric	Validation set: two features		
	Numerical	***Fixed bins***	***CausalRL bins***
Accuracy	0.9925	0.8207	0.8545
Recall (malicious)	1.00	1.00	1.00
Precision (malicious)	0.98	0.61	0.66
F1-score (malicious)	0.99	0.76	0.79
Specificity	0.9915	0.7551	0.7992
ROC AUC	0.9978	0.9438	0.9255

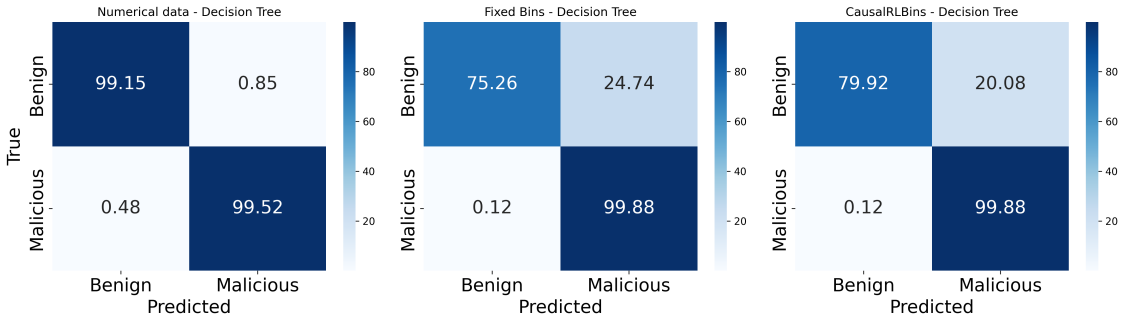
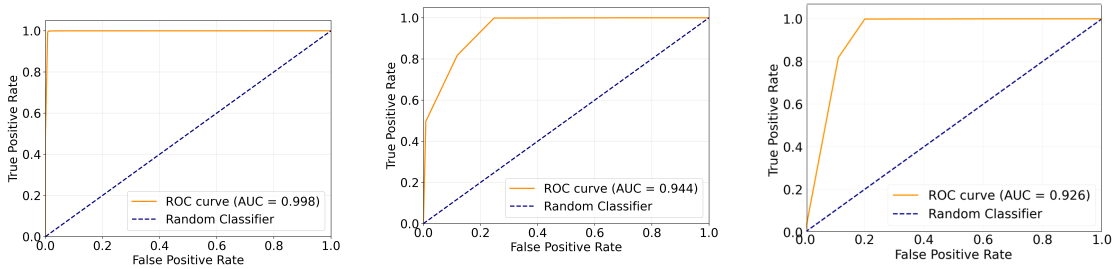


Figure 10: Confusion matrices for three models with two input feature (flow duration and Fwd IAT Mean) evaluated on the validation dataset



(*a*) Decision Tree.

(*b*) Binned features Decision Tree.

(*c*) Causal bins decision Tree.

Figure 11: Decision tree classifier ROC for two input features (flow duration and Fwd IAT Mean) evaluated on the Validation dataset.

Table 5: Key Results Comparison of the decision tree classifier scenarios with two features (flow duration and Fwd IAT Mean) on the test set

Metric	Test set: two features		
	Numerical	***Fixed bins***	***CausalRL bins***
Accuracy	0.6395	0.4491	0.5040
Recall (malicious)	0.02	0.14	0.16
Precision (malicious)	0.66	0.18	0.24
F1-score (malicious)	0.05	0.16	0.19
Specificity	0.9931	0.6255	0.7013
ROC AUC	0.5046	0.5053	0.5766

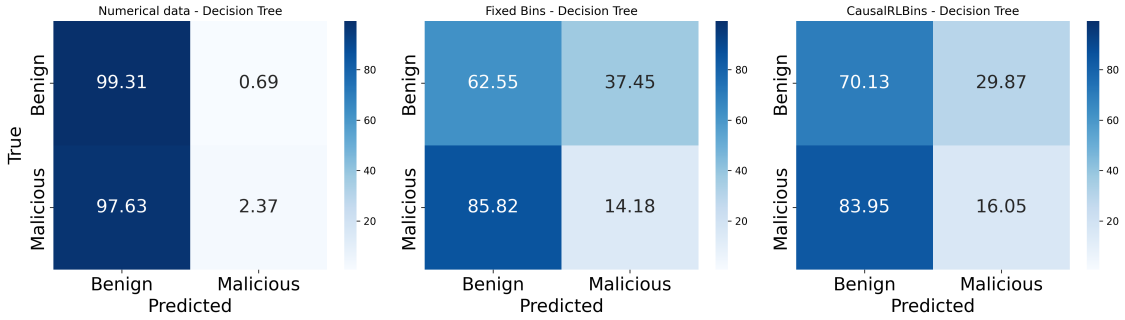


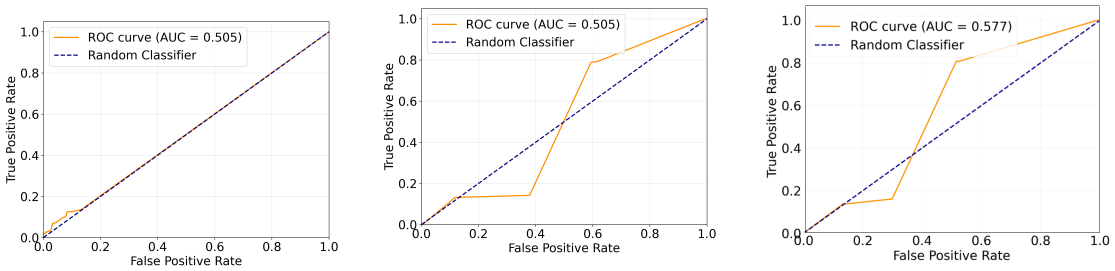
Figure 12: Confusion matrices for three models with two input feature (flow duration and Fwd IAT Mean) evaluated on the test dataset

These findings highlight that while raw numerical features offer the best performance in this context, causalRL-based discretisation offers a viable alternative with improved performance over standard fixed binning.

Table 5, Figure 12 and Figure 13 describe the performance of the decision tree classifier scenarios trained using a different dataset, evaluated on the test set (unseen data) and are provided with two input features (flow duration and Fwd IAT mean).

In the test dataset, the performance of the decision tree classifier varies considerably depending on the input representation. When using the raw numerical treatment values, the classifier achieves an accuracy of 63.95%, an F1 score of 0.05, and a ROC AUC of 0.5046, correctly identifying only 1,182 out of 49,928 malicious flows and 86,365 out of 86,968 benign flows.

The fixed binning approach shows diminished performance with an accuracy of 44.91%, an F1 score of 0.16, and an ROC AUC of 0.5053, yielding 7,078 true positives and 54,402 true negatives. In contrast, the causalRL binning strategy outperforms both alternatives in terms of discrimination capability, achieving a higher ROC AUC of 0.5766 along with an accuracy of 50.40% and an F1 score of 0.19. It correctly classifies 8,013 malicious and 60,989 benign flows.



(*a*) Decision Tree.

(*b*) Binned features Decision Tree.

(*c*) Causal bins decision Tree.

Figure 13: Decision tree classifier ROC for two input features (flow duration and Fwd IAT Mean) evaluated on the test dataset

These results suggest that, while overall classification remains challenging in unseen data, the causalRL binning improves generalization and maintains a better balance between sensitivity and specificity compared to raw numerical and fixed binning methods.

5.1.3. Decision Tree classifier with four input features

Table 6, Figure 14 and Figure 15 describe the performance of the decision tree classifier scenarios in the validation set and are provided with four input features (flow duration, Fwd IAT mean, Flow IAT Mean, backward (Bwd) IAT Mean).

When trained using four input features the decision tree classifier demonstrates varying levels of performance depending on the representation of the input features. The model using raw numerical values achieves near-optimal performance, with an accuracy of 99.53%, F1 score of 0.99, and ROC AUC of 0.9983, correctly identifying 57,127 malicious and 146,681 benign flows.

The fixed binning approach results in reduced effectiveness, achieving an accuracy of 84.24%, F1 score of 0.78, and ROC AUC of 0.9614, with 57,140 true positive and 115,362 true negative classifications. In contrast, the causalRL binning method provides a strong balance between binning abstraction and predictive performance. It achieves an accuracy of 97.77%, F1 score of 0.96, and an ROC AUC of 0.9884, while correctly classifying 56,950 malicious and 143,246 benign instances.

These results suggest that causalRL binning preserves the advantages of feature discretisation while maintaining high classification fidelity, offering a robust alternative to both raw numerical and fixed binning strategies in multi-feature decision-making.

Table 7, Figure 16 and Figure 17 describe the performance of the decision tree classifier scenarios trained using a different dataset, evaluated on the test set (unseen data) and provided with two input features (flow duration and Fwd IAT mean).

When evaluated in the test dataset using four input features flow duration, Fwd IAT Mean, flow IAT Mean, and backward (Bwd) IAT Mean, the decision tree classifier demonstrates varying degrees of performance depending on the feature encoding strategy. The model utilizing raw numerical features yields an accuracy of 64.14%, an F1 score of 0.06,

Table 6: Key Results Comparison of the decision tree classifier scenarios with two features (flow duration and Fwd IAT Mean) on the test set

Metric	Validation set: four features		
	Numerical	***Fixed bins***	***CausalRL bins***
Accuracy	0.9953	0.8424	0.9777
Recall (malicious)	1.00	1.00	0.99
Precision (malicious)	0.99	0.64	0.93
F1-score (malicious)	0.99	0.78	0.96
Specificity	0.9942	0.7820	0.9710
ROC AUC	0.9983	0.9614	0.9884

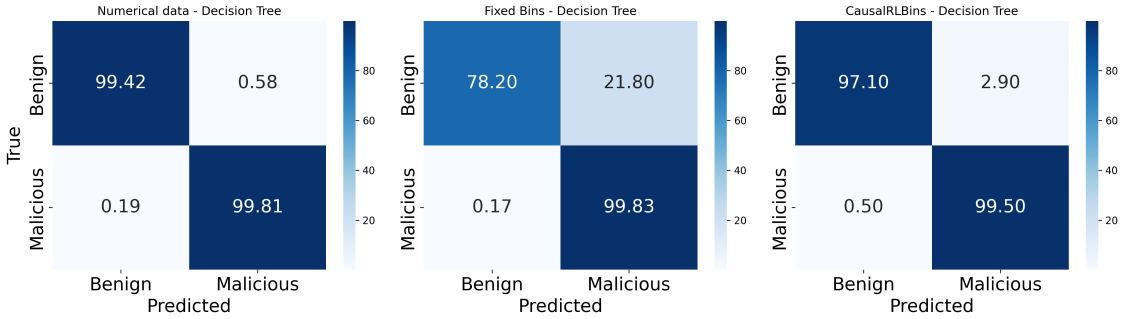
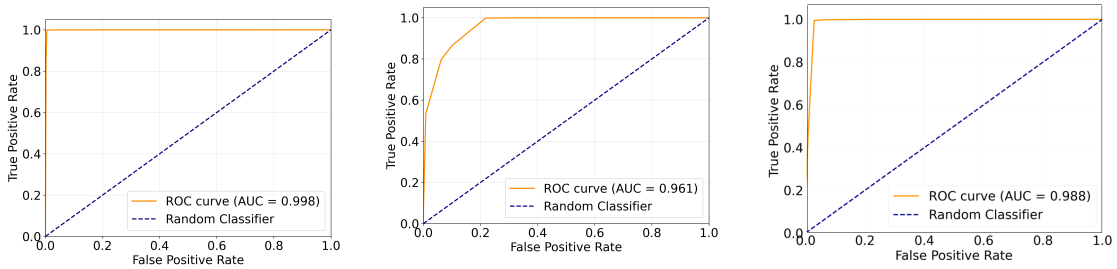


Figure 14: Confusion matrices for three models with four input feature (flow duration and Fwd IAT Mean, flow IAT Mean, Bwd IAT Mean) evaluated on the validation dataset



(*a*) Decision Tree.

(*b*) Binned features Decision Tree.

(*c*) Causal bins decision Tree.

Figure 15: Decision tree classifier ROC for four input features (flow duration and Fwd IAT Mean, flow IAT Mean, Bwd IAT Mean) evaluated on the validation dataset.

Table 7: Key Results Comparison of the decision tree classifier scenarios with two features (flow duration and Fwd IAT Mean) on the test set

Metric	Test set: four features		
	Numerical	***Fixed bins***	***CausalRL bins***
Accuracy	0.6414	0.4699	0.5868
Recall (malicious)	0.03	0.12	0.12
Precision (malicious)	0.66	0.17	0.32
F1-score (malicious)	0.06	0.14	0.17
Specificity	0.9911	0.6693	0.8547
ROC AUC	0.4170	0.5410	0.6113

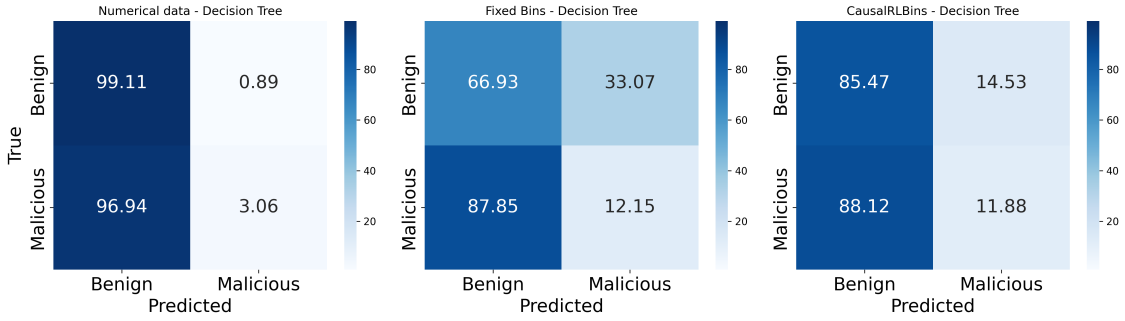
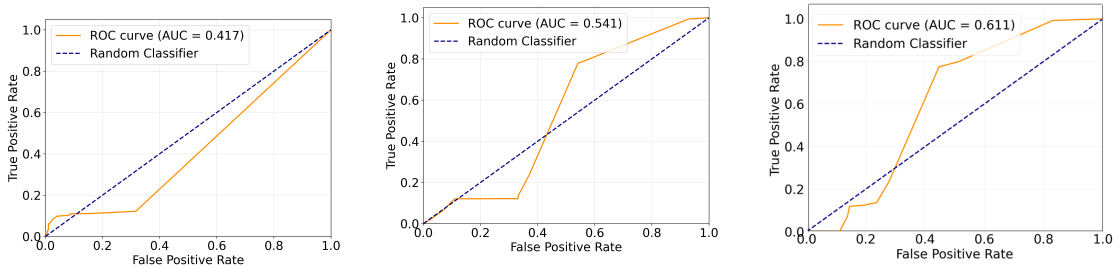


Figure 16: Confusion matrices for three models with four input feature (flow duration and Fwd IAT Mean, flow IAT Mean, Bwd IAT Mean) evaluated on the test dataset



(*a*) Decision Tree.

(*b*) Binned features Decision Tree.

(*c*) Causal bins decision Tree.

Figure 17: Decision tree classifier ROC for four input features (flow duration and Fwd IAT Mean, flow IAT Mean, Bwd IAT Mean) evaluated on the test dataset.

and a ROC AUC of 0.4170, with correct classification of 1,511 malicious and 85,417 benign instances.

The fixed binning approach results in reduced performance, achieving an accuracy of 46.99%, F1 score of 0.14, and ROC AUC of 0.5410, correctly predicting 5,992 malicious and 57,685 benign samples. In contrast, the causalRL binning approach improves generalization, attaining an accuracy of 58.68%, F1 score of 0.17, and ROC AUC of 0.6113, with 5,863 malicious and 73,664 benign flows accurately identified.

5.2. Discussions

The results demonstrate the advantages of using RL with causal inference in discretising numerical features into bins. CausalRL bins improve performance across multiple evaluation metrics, including AUC, precision, recall, and specificity, especially in false positive cases.

It can be observed that the performance of the decision tree classifier in predicting traffic labels improves when more than one feature is input into the model. In the validation set, the numerical model outperforms fixed bins in classifier training, with causal RL achieving the best accuracy and detection rate. It also minimizes false negatives and reduces false positives. The causal RL model offers a high area under the curve, indicating strong class separation. The fixed bins perform poorly due to information loss.

In the test set, the results reveal that the numerical model struggles with malicious recall, whereas the fixed bins model improves recall but struggles with other areas (specificity). The causalRL bins model shows better robustness, accuracy, and ROC AUC. The causal RL bins model is more robust than other scenarios, as it considers causal relationships rather than spurious ones.

The causalRL-based binning strategy demonstrates an improved balance between minimizing false positives and preserving high true positive rates. By learning bin boundaries that are aligned with causally informative thresholds, this approach improves the model's ability to generalize to unseen data. Unlike fixed binning, which can obscure meaningful patterns by imposing arbitrary discretisation, the causal RL method adapts to the underlying data distribution in a manner that preserves treatment-outcome relationships critical for reliable decision making.

6. Conclusion

This study explores a novel causal RL framework for adaptive binning of numerical features in network intrusion detection. By formulating the binning task as a decision-making problem, we employ an RL agent guided by backdoor-adjusted treatment effects to learn optimal discretisation of a numerical variable. These learned bins were used as input features for a decision tree classifier to detect malicious attacks. Our proposed model integrates causal inference into an RL model that assigns rewards based on causal reasoning. This allows the agent to learn statistically informed causal bin assignments enabling agents to learn more effectively, unlike traditional quantile-based strategies. Empirical results demonstrate that although the numerical features are strongest with a known dataset, causalRL binning significantly improves over fixed binning and offers higher accuracy, robustness, and generalization in unseen data. This work focused on a single treatment, confounder, outcome, and

future research will extend the framework to multiple treatments and higher-dimensional confounding structures.

Acknowledgments

This work was supported by the Schlumberger Foundation Faculty for the Future Fellowship.

References

Mohamed Abdel-Basset, Reda Mohamed, Karam M Sallam, and Ibrahim M Hezam. Multi-objective task scheduling method for cyber–physical–social systems in fog computing. *Knowledge-Based Systems*, 280:111009, 2023.

Cande V Ananth and Enrique F Schisterman. Hidden biases in observational epidemiology: the case of unmeasured confounding. *BJOG: an international journal of obstetrics and gynaecology*, 125(6):644, 2017.

Francisco J Aparicio-Navarro, Konstantinos G Kyriakopoulos, and David J Parish. Automatic dataset labelling and feature selection for intrusion detection systems. In *2014 IEEE military communications conference*, pages 46–51. IEEE, 2014.

Matthew Beechey, Konstantinos G Kyriakopoulos, and Sangarapillai Lambotharan. Evidential classification and feature selection for cyber-threat hunting. *Knowledge-Based Systems*, 226:107120, 2021.

Stephen Casper, Carson Ezell, Charlotte Siegmann, Noam Kolt, Taylor Lynn Curtis, Benjamin Bucknall, Andreas Haupt, Kevin Wei, Jérémy Scheurer, Marius Hobbhahn, et al. Black-box access is insufficient for rigorous ai audits. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 2254–2272, 2024.

Ashok Choppadandi, Jagbir Kaur, Pradeep Kumar Chenchala, Akshay Agarwal, Varun Nakra, and Pandi Kirupa Gopalakrishna Pandian. Anomaly detection in cybersecurity: Leveraging machine learning algorithms. *ESP Journal of Engineering & Technology Advancements*, 1(2):34–41, 2021.

Vinícius G Costa and Carlos E Pedreira. Recent advances in decision trees: an updated survey. *Artificial Intelligence Review*, 56(5):4765–4800, 2023.

Ishita Dasgupta, Jane Wang, Silvia Chiappa, Jovana Mitrovic, Pedro Ortega, David Raposo, Edward Hughes, Peter Battaglia, Matthew Botvinick, and Zeb Kurth-Nelson. Causal reasoning from meta-reinforcement learning. *arXiv preprint arXiv:1901.08162*, 2019.

Logan Engstrom, Andrew Ilyas, Shibani Santurkar, Dimitris Tsipras, Firdaus Janoos, Larry Rudolph, and Aleksander Madry. Implementation matters in deep policy gradients: A case study on ppo and trpo. *arXiv preprint arXiv:2005.12729*, 2020.

Junpeng Fang, Gongduo Zhang, Qing Cui, Caizhi Tang, Lihong Gu, Longfei Li, Jinjie Gu, and Jun Zhou. Backdoor adjustment via group adaptation for debiased coupon recommendations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 11944–11952, 2024.

Canadian Institute for Cybersecurity. Intrusion detection evaluation dataset (cic-ids2017). *URL https://www.unb.ca/cic/datasets/ids-2017.html*, a.

Canadian Institute for Cybersecurity. Cse-cic-ids2018 on aws. *URL https://www.unb.ca/cic/datasets/ids-2018.html*, b.

Christopher Hitchcock and Miklós Rédei. Reichenbach's common cause principle. 2020.

Lanlan Huang, Junkai Zhao, Bing Zhu, Hao Chen, and Seppe Vanden Broucke. An experimental investigation of calibration techniques for imbalanced data. *Ieee Access*, 8: 127343–127352, 2020.

Zahra Jadidi, Vallipuram Muthukkumarasamy, Elankayer Sithirasenan, and Mansour Sheikhan. Flow-based anomaly detection using neural network optimized with gsa algorithm. In *2013 IEEE 33rd international conference on distributed computing systems workshops*, pages 76–81. IEEE, 2013.

Ahmad Javaid, Quamar Niyaz, Weiqing Sun, and Mansoor Alam. A deep learning approach for network intrusion detection system. In *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, pages 21–26, 2016.

Lu Jia, Binglin Su, Du Xu, and Yewei Wang. Proximal policy optimization with an activated policy clipping. In *2024 International Conference on Energy and Electrical Engineering (EEE)*, pages 1–5. IEEE, 2024.

Hamza Kheddar, Diana W Dawoud, Ali Ismail Awad, Yassine Himeur, and Muhammad Khurram Khan. Reinforcement-learning-based intrusion detection in communication networks: A review. *IEEE Communications Surveys & Tutorials*, 2024.

Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, et al. Wilds: A benchmark of in-the-wild distribution shifts. In *International conference on machine learning*, pages 5637–5664. PMLR, 2021.

Sotiris Kotsiantis and Dimitris Kanellopoulos. Discretization techniques: A recent survey. *GESTS International Transactions on Computer Science and Engineering*, 32(1):47–58, 2006.

Asher Labovich. A case for library-level k-means binning in histogram gradient-boosted trees. *arXiv preprint arXiv:2505.12460*, 2025.

Thuc Duy Le, Tao Hoang, Jiuyong Li, Lin Liu, Huawen Liu, and Shu Hu. A fast pc algorithm for high dimensional causal discovery with multi-core pcs. *IEEE/ACM transactions on computational biology and bioinformatics*, 16(5):1483–1495, 2016.

Richard P Lippmann, David J Fried, Isaac Graf, Joshua W Haines, Kristopher R Kendall, David McClung, Dan Weber, Seth E Webster, Dan Wyschogrod, Robert K Cunningham, et al. Evaluating intrusion detection systems: The 1998 darpa off-line intrusion detection

evaluation. In *Proceedings DARPA Information survivability conference and exposition. DISCEX'00*, volume 2, pages 12–26. IEEE, 2000.

Huan Liu, Farhad Hussain, Chew Lim Tan, and Manoranjan Dash. Discretization: An enabling technique. *Data mining and knowledge discovery*, 6(4):393–423, 2002.

Yonggang Lu, Qiujie Zheng, and Daniel Quinn. Introducing causal inference using bayesian networks and do-calculus. *Journal of Statistics and Data Science Education*, 31(1):3–17, 2023.

Shoaib Malik. Explainable ai for cybersecurity: Improving transparency in automated threat detection systems. 2024.

Soroosh Mariooryad and Carlos Busso. The cost of dichotomizing continuous labels for binary classification problems: Deriving a bayesian-optimal classifier. *IEEE Transactions on Affective Computing*, 8(1):119–130, 2015.

Brady Neal. Introduction to causal inference. *Course lecture notes (draft)*, 132, 2020.

Eduardo E Oliveira, Vera L Miguéis, and José L Borges. Understanding overlap in automatic root cause analysis in manufacturing using causal inference. *IEEE Access*, 10:191–201, 2021.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT press, 2017.

Tom Purves, Konstantinos G Kyriakopoulos, Sian Jenkins, Iain Phillips, and Tim Dudman. Causally aware reinforcement learning agents for autonomous cyber defence. *Knowledge-Based Systems*, 304:112521, 2024.

Alexander Ratner, Stephen H Bach, Henry Ehrenberg, Jason Fries, Sen Wu, and Christopher Ré. Snorkel: rapid training data creation with weak supervision. *The VLDB Journal*, 29(2):709–730, 2020.

Ribana Roscher, Bastian Bohn, Marco F Duarte, and Jochen Garcke. Explainable machine learning for scientific insights and discoveries. *Ieee Access*, 8:42200–42216, 2020.

Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215, 2019.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Iman Sharafaldin, Arash Habibi Lashkari, Ali A Ghorbani, et al. Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSp*, 1(2018): 108–116, 2018.

Richard S Sutton, Andrew G Barto, et al. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

Hairui Wang, Junming Li, and Guifu Zhu. A data feature extraction method based on the notears causal inference algorithm. *Applied Sciences*, 13(14):8438, 2023.

Lingxiao Wang, Zhuoran Yang, and Zhaoran Wang. Provably efficient causal reinforcement learning with confounded observational data. *Advances in Neural Information Processing Systems*, 34:21164–21175, 2021.

LU Ying et al. Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27(2):130, 2015.

ZengRi Zeng, Wei Peng, Detian Zeng, Chong Zeng, and YiFan Chen. Intrusion detection framework based on causal reasoning for ddos. *Journal of Information Security and Applications*, 65:103124, 2022.

Amy Zhang, Zachary C Lipton, Luis Pineda, Kamyar Azizzadenesheli, Anima Anandkumar, Laurent Itti, Joelle Pineau, and Tommaso Furlanello. Learning causal state representations of partially observable environments. *arXiv preprint arXiv:1906.10437*, 2019.

Yujia Zheng, Biwei Huang, Wei Chen, Joseph Ramsey, Mingming Gong, Ruichu Cai, Shohei Shimizu, Peter Spirtes, and Kun Zhang. Causal-learn: Causal discovery in python. *Journal of Machine Learning Research*, 25(60):1–8, 2024.

Xinyuan Zhu, Yang Zhang, Fuli Feng, Xun Yang, Dingxian Wang, and Xiangnan He. Mitigating hidden confounding effects for causal recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 36(9):4794–4805, 2024.