Learning to Offload with Low Regret under Asymmetric Misclassification Costs

Anonymous Authors[1]

Abstract

We consider edge AI systems designed for binary classification via *hierarchical inference* (HI): a compact local model on a resource-constrained device classifies each incoming sample, and when uncertain, forwards it to a more powerful model on a remote server over a costly wireless link, a decision called *offloading*. The device receives a stream of periodically generated samples and must classify each one in real time, deciding online whether to act locally or offload, trading offloading cost against misclassification risk. The challenge is that neither the local model's true accuracy at each confidence level nor the offloading cost is known in advance and must be learned from experience. We focus on the setting where the two types of misclassifications carry different costs and propose an offloading algorithm for which we prove the first $O(\log T)$ regret guarantee for asymmetric-cost hierarchical inference. We supplement our analytical results via experiments on four real-world datasets.

1. Introduction

Modern AI systems designed for classification tasks increasingly face a deployment tension: powerful models are accurate but computationally heavy, while compact models fit on resource-constrained devices but sacrifice accuracy. Hierarchical inference (HI) resolves this tension by combining both. A lightweight local model deployed on an end device, such as a sensor, smartphone, or medical instrument, handles most classification tasks directly. When the local model is uncertain about a sample, it forwards the sample to a more accurate model running on a remote edge server over a wireless link, a decision called *offloading*. Offloading improves accuracy but incurs an offloading cost that varies with wireless channel conditions. The system must therefore decide, for each incoming sample, whether the risk of a local misclassification outweighs the cost of offloading to the server.

This offloading decision is complicated by two sources of uncertainty. First, the local classification model's confidence score is not a calibrated probability: the same score can correspond to very different true accuracies depending on how the model was trained. The true mapping from confidence score to accuracy is unknown at deployment and must be estimated online. Second, the offloading cost fluctuates randomly with channel conditions and its average is likewise unknown. Crucially, both quantities can only be estimated from offloaded samples: when the system decides to act locally, it receives no feedback. This creates a classic exploration-exploitation tension: offloading too often wastes communication resources, while offloading too rarely leaves both unknowns poorly estimated.

A key but overlooked aspect of real deployments is that the two types of classification error carry very different costs. In dynamic spectrum access, a missed primary user causes harmful interference while a false alarm merely delays one slot. In medical diagnosis, a missed malignancy is far more serious than an unnecessary follow-up. Moothedath et al. (2026) addressed this asymmetric-cost setting and showed that the optimal offloading policy requires *two* confidence thresholds rather than one, but their algorithm, which makes no structural assumptions on the relationship between confidence score and local accuracy, achieves only polynomial regret. Prior work on the symmetric-cost case (Al-Atat et al., 2024; Chattopadhyay et al., 2025) achieves logarithmic regret by exploiting the monotonicity of this relationship. This raises a natural question: does the same assumption suffice to achieve logarithmic regret in the asymmetric-cost case? *This paper answers that question affirmatively.*

1.1. Contributions

We characterize the optimal offline policy under asymmetric misclassification costs, proving it has a two-threshold form (Section 3). We propose HI-LCB-2D (Section 4), which learns both thresholds online without prior knowledge of local accuracy or channel cost, under the mild monotonicity assumption. We prove the first logarithmic regret guarantee for asymmetric-cost hierarchical inference (Section 5;

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

Theorem 5.1), showing that monotonicity is sufficient to move from polynomial to logarithmic regret. We also show that this regret guarantee is order-optimal. Experiments on four real-world datasets confirm that HI-LCB-2D achieves performance competitive with or superior to the prior state of the art across varying channel conditions and cost parameters (Section 6).

Remark 1.1 (Technical Novelty). In the symmetric case of Chattopadhyay et al. (2025), a single confidence interval on local accuracy suffices to certify the unique offloading threshold. With asymmetric costs, the two boundaries require separate conditions that combine estimates of local accuracy and channel cost in different ways, yet both share the same offload observations and running channel cost estimate, making naive decoupling impossible. HI-LCB-2D resolves this by maintaining independent per-class confidence bounds while sharing a single channel cost estimate across both conditions.

1.2. Related Work

Hierarchical and early-exit inference. Early-exit networks (Teerapittayanon et al., 2016) and device-server split computing (Kang et al., 2017) optimize the partition point to minimize latency and energy. Our setting differs in that the split is fixed and the offloading cost is a stochastic wireless quantity learned online.

Online learning for hierarchical inference. Al-Atat et al. (2024) achieved polynomial regret for symmetric-cost HI; Chattopadhyay et al. (2025) improved this to logarithmic regret by exploiting monotonicity of the calibration function, with a matching lower bound. Moothedath et al. (2026) extended this to asymmetric costs, identifying the two-threshold optimal policy structure and achieving polynomial regret. The logarithmic barrier for asymmetric-cost HI remained open until this work, which achieves logarithmic regret under the same monotonicity assumption.

Partial monitoring and censored feedback. The offload-only feedback model is structurally a partial monitoring problem (Cesa-Bianchi et al., 2006). The closest analogue is the threshold bandit with censored feedback (Abernethy et al., 2016), where rewards are observed only when exceeding a threshold; our contribution achieves logarithmic regret in a two-threshold variant of this setting.

Cost-sensitive learning. Elkan (2001) showed that optimal classifiers under asymmetric penalties threshold the posterior at the cost ratio. Our two-threshold structure is the HI analogue, with offloading introducing a third action and two distinct cost-ratio boundaries.

Model calibration. Post-hoc calibration methods such as temperature scaling (Guo et al., 2017) estimate the mapping from confidence scores to true accuracy before deployment. Our algorithm instead estimates this mapping online from offload-only feedback, under the standard monotonicity assumption that is empirically well-supported across modern classifiers.

2. System Model and Problem Formulation

We consider a two-tier system for binary classification in which an end device and a remote edge server collaborate to classify a stream of periodically generated samples. The end device runs a pre-trained lightweight local model and must decide, for each incoming sample, whether to classify it locally or forward it to the more accurate server. While the local parameters remain fixed, the central goal is to learn an offloading policy that minimizes the cumulative gap in expected cost against an oracle that knows the local model's true accuracy and the average wireless offloading cost in advance. We now describe the system components, cost structure, and the formal regret objective in turn.

Per-sample decision. The ED receives a stream of samples $\{x_t\}_{t\geq 1}$ one at a time. For each sample, the system must choose one of three actions:

- *Accept class 0*: use the local prediction $h_l(x_t) = 0$.
- *Accept class 1*: use the local prediction $h_l(x_t) = 1$.
- *Offload*: send x_t to the ES over the communication link and use the server's prediction $h_r(x_t)$ as the final answer.

Accepting locally avoids offloading cost but risks a misclassification. Offloading obtains a more reliable label but incurs an offloading cost.

Local model confidence and the calibration function. The local model outputs a softmax score vector $\bar{\phi}_t = (\phi_0(x_t), \phi_1(x_t))$ with $\phi_0 + \phi_1 = 1$, where $\phi_c(x_t)$ reflects the model's raw confidence in class c. We write $\phi_t := \phi_1(x_t) \in [0, 1]$ as the score for class 1; by complementarity, the score for class 0 is $1 - \phi_t$. A higher value of ϕ_t indicates greater local confidence that the true label is 1, and a lower value indicates greater confidence in class 0. However, these scores are not necessarily calibrated probabilities: the same score can correspond to very different true accuracies depending on the model architecture, training procedure, and data distribution (Guo et al., 2017). For instance, a model that consistently outputs scores near 0.9 may be correct only 70% of the time. Ignoring this miscalibration and treating the raw score as a probability would lead to systematically wrong offloading decisions. We therefore model the true posterior via a calibration function $f : [0, 1] \to [0, 1]$, assumed non-decreasing (Chattopadhyay et al., 2025), such that

$$\mathrm{P}\{y_t = 1 \mid \bar{\phi}_t\} = f(\phi_t). \tag{1}$$

By complementarity, $\mathrm{P}\{y_t = 0 \mid \bar{\phi}_t\} = 1 - f(\phi_t)$. Refer to

the appendix for empirical results backing the assumption that $f(\cdot)$ is non-decreasing for the binary classifiation. In a perfectly calibrated model f equals the identity. In practice f can differ substantially, and its exact form is unknown at deployment. Estimating f online from offloaded samples is one of the two central challenges the algorithm must solve.

Misclassification costs. The two types of classification error carry different costs, set by domain knowledge (Elkan, 2001). Since the true label y_t is unobserved at inference time, we follow Moothedath et al. (2026) in treating the edge server prediction $h_r(x_t)$ as a perfect label proxy, i.e., $h_r(x_t) = y_t$. This assumption is reasonable in settings where the server has access to substantially more compute and data than the end device, enabling the construction of highly accurate classifiers that serve as reliable ground-truth proxies. Relaxing this assumption to account for server errors is an open problem. Accepting the local prediction incurs a misclassification cost

$$\eta_t = \begin{cases} \delta_1 & \text{if } h_l(x_t) = 1,\ h_r(x_t) = 0 \text{ (false +ve)}, \\ \delta_{-1} & \text{if } h_l(x_t) = 0,\ h_r(x_t) = 1 \text{ (false −ve)}, \\ 0 & \text{otherwise}, \end{cases} \quad (2)$$

where $\delta_1 > 0$ and $\delta_{-1} > 0$ are the respective penalty constants, assumed known to the operator.

Offloading cost and total loss. Offloading incurs a offloading cost $\Gamma_t \in [0, 1]$, drawn i.i.d. each round with unknown mean γ, reflecting stochastic wireless channel conditions (Al-Atat et al., 2024). The mean γ is not known in advance and must be estimated online alongside $f(\cdot)$; estimating it is the second central challenge the algorithm must solve. The total per-slot cost under policy π combines both cost sources:

$$l_t(\pi) = \Gamma_t \, \mathbf{1}\{D_t^\pi = \text{off}\} + \eta_t \, \mathbf{1}\{D_t^\pi \in \{0, 1\}\}. \quad (3)$$

Note that exactly one of the two terms is nonzero at each round: offloading avoids misclassification risk but incurs offloading cost, while acting locally avoids offloading cost but risks misclassification.

What is observed and what must be learned. At each round t, the algorithm observes the softmax score ϕ_t and knows the penalty parameters δ_1, δ_{-1} (operator-specified). It does *not* know $f(\cdot)$ or γ. When the algorithm offloads x_t, it observes the ES output $h_r(x_t)$ (used as a label proxy to update the estimate of f at level ϕ_t) and the realized channel cost Γ_t (used to update the estimate of γ). When it accepts locally, no feedback is received. The online learning challenge is thus: learn f and γ fast enough, from offload-only feedback, that cumulative cost is close to what an oracle knowing both would achieve.

Threshold policies and regret. We restrict attention to *threshold-based policies* parameterized by a pair $\bar{\theta} = (\theta_l, \theta_u)$ with $0 \le \theta_l \le \theta_u \le 1$: offload when $\theta_l < f(\phi_t) < \theta_u$, accept class 0 when $f(\phi_t) \le \theta_l$, and accept class 1 when $f(\phi_t) \ge \theta_u$. This is without loss of generality: Section 3 proves that the optimal offline policy belongs to this family (Lemma 3.3). Let $L_T(\bar{\theta}) = \sum_{t=1}^{T} l_t(\bar{\theta})$ be the cumulative cost and $\bar{\theta}^* = \arg\min_{\bar{\theta}} \mathbb{E}[L_T(\bar{\theta})]$ the best fixed threshold pair. The *regret* of an online policy π is its cumulative cost gap against this offline optimum:

$$R_T(\pi) = \mathbb{E}[L_T(\pi)] - \mathbb{E}[L_T(\bar{\theta}^*)]. \quad (4)$$

Discrete confidence set. In practice, softmax outputs are quantized (e.g., to 8-bit precision), so ϕ_t takes values in a finite set $\Phi = \{\phi_1 < \cdots < \phi_N\} \subset [0, 1]$ (Al-Atat et al., 2024). This is not a restriction: it matches how models are deployed and makes per-level accuracy estimates $\hat{f}(\phi_i)$ well-defined without binning. The regret bound in Theorem 5.1 depends on $|\Phi| = N$ through the gap-dependent sum; for 8-bit precision $N \le 256$.

3. Optimal Offline Policy

Before designing an online algorithm, we characterize the optimal policy available to an oracle that knows both $f(\cdot)$ and γ in advance. This offline policy serves as the benchmark against which the regret of any online algorithm is measured; a small cumulative gap against this oracle is the central goal of HI-LCB-2D.

When $f(\cdot)$ and γ are both known, the optimal per-sample decision follows from comparing the three expected costs.

Lemma 3.1 (Optimal Offline Decision)**.** *For a sample x_t with confidence score ϕ_t, the expected costs of the three decisions are*

$$L_t(0) = \delta_{-1} f(\phi_t), L_t(1) = \delta_1\big(1 - f(\phi_t)\big), L_t(\text{off}) = \gamma. \quad (5)$$

The offline optimal policy selects $D_t^{\pi^*} = \arg\min_d L_t(d)$.

The three costs have natural interpretations. $L_t(0) = \delta_{-1} f(\phi_t)$ is the expected false-negative cost: accepting class 0 is wrong whenever the true label is 1, which happens with probability $f(\phi_t)$, incurring penalty δ_{-1}. Symmetrically, $L_t(1) = \delta_1(1 - f(\phi_t))$ is the expected false-positive cost. Offloading incurs the expected offloading cost γ regardless of the sample.

Lemma 3.2 (No-Offload Condition)**.** *The offline optimal policy never offloads when $\gamma > \delta_1\delta_{-1}/(\delta_1 + \delta_{-1})$, i.e., when the wireless cost exceeds the harmonic half-mean of the two class penalties.*

The harmonic half-mean $\delta_1\delta_{-1}/(\delta_1 + \delta_{-1})$ is the maximum value that $\min\{\delta_{-1} f(\phi_t), \delta_1(1 - f(\phi_t))\}$ can take over all ϕ_t. When γ exceeds this value, the offloading cost is higher than the best possible local decision cost for every sample, regardless of its confidence score. Offloading is therefore

never optimal. For the symmetric case $\delta_1 = \delta_{-1} = \delta$, this reduces to $\gamma > \delta/2$, consistent with Al-Atat et al. (2024); Chattopadhyay et al. (2025).

The next lemma is the central structural result of this section. It characterizes the full partition of the confidence space under the optimal offline policy (see Figure 1).

Lemma 3.3 (Two-Threshold Structure)**.** *Assume $\gamma < \delta_1\delta_{-1}/(\delta_1 + \delta_{-1})$. Under the offline optimal policy, the confidence space partitions as $\Phi_0 \prec \Phi_{\text{amb}} \prec \Phi_1$, where*

$$\Phi_0 = \{\phi : f(\phi) \le \gamma/\delta_{-1}\}, \tag{6}$$

$$\Phi_1 = \{\phi : f(\phi) \ge 1 - \gamma/\delta_1\}, \tag{7}$$

$$\Phi_{\text{amb}} = \Phi \setminus (\Phi_0 \cup \Phi_1). \tag{8}$$

The two optimal thresholds are $\theta_l^* = f^{-1}(\gamma/\delta_{-1})$ and $\theta_u^* = f^{-1}(1 - \gamma/\delta_1)$.

Each threshold has a clear interpretation. The lower threshold θ_l^* is the confidence level at which the expected false-negative cost $\delta_{-1}f(\phi)$ equals the offloading cost γ. Below this level, accepting class 0 is cheaper than offloading. The upper threshold θ_u^* is where the expected false-positive cost $\delta_1(1 - f(\phi))$ equals γ. Above this level, accepting class 1 is cheaper. In the ambiguous region Φ_{amb} between the two thresholds, neither local decision is cheap enough, so offloading is optimal.

Remark 3.4 (Connection to the Symmetric Case). The case $\theta_l^* = \theta_u^*$ implies $\gamma(\delta_1 + \delta_{-1}) = \delta_1\delta_{-1}$, which defines the boundary of the offloading regime (Lemma 3.2). For γ strictly below this boundary, the two thresholds are distinct unless $\delta_1 = \delta_{-1}$. In this symmetric subcase, $\theta_l^* = \theta_u^* = f^{-1}(1/2)$, and the two-threshold structure degenerates to the single-threshold case of Al-Atat et al. (2024) and Chattopadhyay et al. (2025).

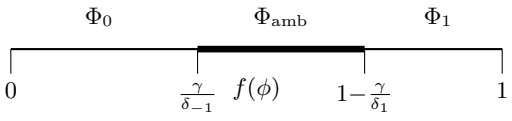


Figure 1. Two-threshold partition of the confidence space. The lower boundary γ/δ_{-1} is the point at which the expected false-negative cost equals the offloading cost. The upper boundary $1 - \gamma/\delta_1$ does the same for the false-positive cost. The two boundaries coincide only when $\delta_1 = \delta_{-1}$.

4. Our Algorithm: HI-LCB-2D

In practice, neither $f(\cdot)$ nor γ is known. The algorithm must therefore learn both online, from the feedback generated by its own decisions. This is complicated by the fact that feedback is only available when the algorithm offloads: local decisions reveal nothing about f or γ.

Confidence intervals. The algorithm maintains an empirical estimate $\hat{f}(\phi_i)$ and offload count O_{ϕ_i} for each $\phi_i \in \Phi$, along with upper and lower confidence bounds

$$\text{LCB}_{\phi_i} = \hat{f}(\phi_i) - \sqrt{\frac{\alpha \log t}{O_{\phi_i}}}, \text{UCB}_{\phi_i} = \hat{f}(\phi_i) + \sqrt{\frac{\alpha \log t}{O_{\phi_i}}}. \tag{9}$$

These are standard Hoeffding-based confidence intervals. The exploration parameter $\alpha > 0.5$ controls the width of these intervals; larger α gives wider intervals and more cautious certification, at the cost of more exploration. The constraint $\alpha > 0.5$ is required to ensure that error probabilities are summable over time. For the offloading cost, with $O_\gamma = \sum_{\phi_i} O_{\phi_i}$ total offloads and running estimate $\hat{\gamma}$, the lower confidence bound is

$$\text{LCB}_\gamma = \hat{\gamma} - \sqrt{\frac{\alpha \log t}{O_\gamma}}. \tag{10}$$

Note that $O_\gamma \ge O_{\phi_i}$ for all ϕ_i, since $\hat{\gamma}$ accumulates over all offloads while O_{ϕ_i} counts only offloads at a single confidence level. This means the channel cost estimate $\hat{\gamma}$ is always at least as well-estimated as any individual $\hat{f}(\phi_i)$.

Design Principle. The key design principle of HI-LCB-2D is to replace the unknown quantities in the offline optimality conditions of Lemma 3.1 with pessimistic confidence bound estimates. If the condition holds under these pessimistic substitutions, it is safe to accept class 0 without offloading. The same reasoning applies to class 1, replacing $f(\phi_t)$ with $\text{LCB}_{\phi(t)}$ (the smallest plausible value of f). This yields the two conditions:

$$\mathcal{C}_0(t) : \delta_{-1}\,\text{UCB}_{\phi(t)} < \text{LCB}_\gamma \;\&\; \text{UCB}_{\phi(t)} < \frac{\delta_1}{\delta_1 + \delta_{-1}}, \tag{11}$$

$$\mathcal{C}_1(t) : \delta_1\big(1 - \text{LCB}_{\phi(t)}\big) < \text{LCB}_\gamma \;\&\; \text{LCB}_{\phi(t)} \ge \frac{\delta_1}{\delta_1 + \delta_{-1}}. \tag{12}$$

Each condition has two sub-conditions serving distinct roles. In $\mathcal{C}_0$: the first sub-condition certifies that the worst-case false-negative cost is still below the best-case offloading cost, so accepting class 0 is safe; the second sub-condition ensures the sample genuinely belongs to Φ_0 rather than Φ_1, preventing the algorithm from accepting the wrong class. Similar logic applies to $\mathcal{C}_1$.

Shared cost estimate. A critical design decision is that both $\mathcal{C}_0$ and $\mathcal{C}_1$ use the *same* lower confidence bound LCB_γ, built from all offload observations regardless of confidence level. This shared estimate is both necessary and sufficient: necessary because running two decoupled copies of the symmetric algorithm would require two independent cost estimates, which is impossible since all offloads contribute to the same $\hat{\gamma}$; sufficient because the coupling through LCB_γ does not prevent independent certification of the two boundaries, as the per-class confidence bounds $\text{UCB}_{\phi(t)}$ and $\text{LCB}_{\phi(t)}$ are

maintained separately for each class. This is the key algorithmic insight that allows the regret bound to decompose into two independent $O(\log T)$ terms, one per boundary.

Decision rule. The full decision rule is:

$$D^\pi(t) = \begin{cases} 0 & \text{if } \mathcal{C}_0(t), \\ 1 & \text{if } \mathcal{C}_1(t), \\ \text{offload} & \text{otherwise.} \end{cases} \tag{13}$$

The formal algorithm is given in Algorithm 1.

Algorithm 1 HI-LCB-2D

Require: Samples $\{x_t\}$, confidence set Φ, penalties δ_1, δ_{-1}, exploration parameter $\alpha{>}0.5$

Initialize: $\hat{\gamma} \leftarrow 0$, $O_\gamma \leftarrow 0$; $\hat{f}(\phi_i) \leftarrow 0$, $O_{\phi_i} \leftarrow 0$ $\forall\, \phi_i \in \Phi$
for $t = 1, 2, \ldots$ **do**
  Receive local prediction $h_l(x_t)$ and score $\phi(t) \in \Phi$
  Compute $\mathrm{LCB}_{\phi(t)}$, $\mathrm{UCB}_{\phi(t)}$ via (9); LCB_γ via (10)
  Set $D^\pi(t)$ via (13)
  if $D^\pi(t) = \text{off}$ **then**
    Offload x_t; observe label proxy y_t and channel cost Γ_t
    $\hat{f}(\phi(t)) \leftarrow \frac{O_{\phi(t)}\hat{f}(\phi(t)) + \mathbb{1}\{h_l(x_t) = y_t\}}{O_{\phi(t)}+1}$; $O_{\phi(t)} \leftarrow O_{\phi(t)}+1$
    $\hat{\gamma} \leftarrow \frac{O_\gamma \hat{\gamma} + \Gamma_t}{O_\gamma + 1}$; $O_\gamma \leftarrow O_\gamma + 1$
  end if
end for

5. Regret Analysis

We now analyze the cumulative regret of HI-LCB-2D. For each $\phi_i \in \Phi$, define the sub-optimality gaps $\Delta_{\phi_i}^{A0\text{-}O} = |\delta_{-1} f(\phi_i) - \gamma|$, $\Delta_{\phi_i}^{O\text{-}A1} = |\gamma - \delta_1(1 - f(\phi_i))|$, $\Delta_{\phi_i} = |f(\phi_i) - \delta_1/(\delta_1 + \delta_{-1})|$, $\beta_{\phi_i} = \gamma - \delta_{-1} f(\phi_i)$ for $\phi_i \in \Phi_0$, and $\kappa_{\phi_i} = \gamma - \delta_1(1 - f(\phi_i))$ for $\phi_i \in \Phi_1$. These measure how far each confidence level is from the relevant decision boundaries.

Theorem 5.1. *For $\alpha > 0.5$, the regret of* HI-LCB-2D *satisfies*

$$R_T(\pi) \leq O(1) + 4\alpha \log T \times \left(\sum_{\phi_i \in \Phi_0} \Delta_{\phi_i}^{A0\text{-}O} \cdot \max\left\{ \Delta_{\phi_i}^{-2},\, (\delta_{-1}+1)^2 \beta_{\phi_i}^{-2} \right\} + \sum_{\phi_i \in \Phi_1} \Delta_{\phi_i}^{O\text{-}A1} \cdot \max\left\{ \Delta_{\phi_i}^{-2},\, (\delta_1+1)^2 \kappa_{\phi_i}^{-2} \right\} \right).$$

It follows that, $R_T(\pi) = O(\log T)$.

Proof sketch. We decompose regret across the three regions. In Φ_{amb}, offloading is optimal; incorrect local acceptances occur only when a Hoeffding confidence bound fails, contributing $O(1)$ in expectation. In Φ_0 and Φ_1, the algorithm offloads until $\mathcal{C}_0$ or $\mathcal{C}_1$ is certified. The expected offload count at each ϕ_i before certification is $O(\log T)$, governed by whichever of the two sub-conditions is harder to satisfy: resolving the cost comparison (governed by β_{ϕ_i} or κ_{ϕ_i}) or identifying the correct class (governed by Δ_{ϕ_i}). Multiplying the offload count by the per-step regret and summing over Φ_0 and Φ_1 gives the two logarithmic terms above. Since $|\Phi| < \infty$ and all gaps are strictly positive, the total is $O(\log T)$. Full details are in Appendix E. □

Remark 5.2. When $\delta_1 = \delta_{-1}$, the two terms $R_{\Phi_0}(T)$ and $R_{\Phi_1}(T)$ collapse into the single logarithmic term of Chattopadhyay et al. (2025), confirming that HI-LCB-2D is a strict generalization of the symmetric-cost algorithm. The $\Omega(\log T)$ lower bound of Chattopadhyay et al. (2025) suggests the rate is tight.

6. Experiments

To evaluate the performance of our approach, we conduct experiments on the following dataset-model pairs, which are identified hereafter by the name of the dataset.

- The Breast Cancer Histopathological Image Classification *(BreakHis)* (Spanhol et al., 2015) contains 7909 images of benign (class 1) and malignant cells. The LDL model is a lightweight MobileNet-based classifier with additional dense and dropout layers.
- *Chest* (Mohamed, 2025) contains 844 chest CT scan images re-categorised into healthy and cancerous (class 1) classes for our use. The LDL is based on MobileNet, with a size of 13.3 MB including trainable parameters.
- *Phishing* (Tiwari, 2025; Tan, 2018) includes data collected from 5000 phishing and as many legitimate websites, with additional samples for testing. A logistic regression model (56 bytes) is used as LDL.
- *BreaCh* is created by classifying all 7909 BreakHis samples using the model trained on Chest CT scan. It mimics an OOD data scenario with a domain shift for the model trained on Chest CT scan. As a result, unbeknownst to the user, the accuracy drops below 50%, resulting in the misdiagnosis of 38% cancerous tumors.

We implement the following algorithms for benchmarking. *No Offload*: the LDL inference is accepted as is, *Full Offload*: all samples are offloaded, $\vec{\theta}^*$: the offline optimal two-threshold policy. *H2T2* (Moothedath et al., 2026): the current state-of-the-art two threshold HI policy. *HI-LCB-2D*: our proposed policy.

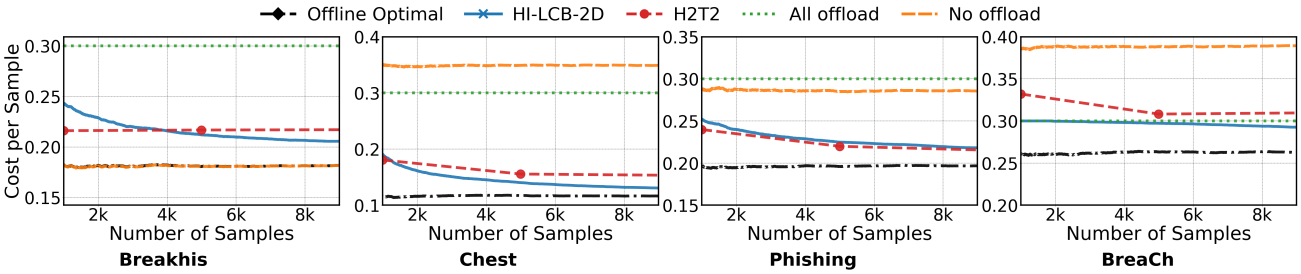


Figure 2. Per-sample cost vs. number of samples across four datasets. HI-LCB-2D converges to the Offline Optimal after a short exploration phase; H2T2 converges more slowly due to its $O(T^{2/3})$ rate ($\delta_1 = 0.7, \delta_{-1} = 1, \alpha = 0.52, \gamma = 0.3$).

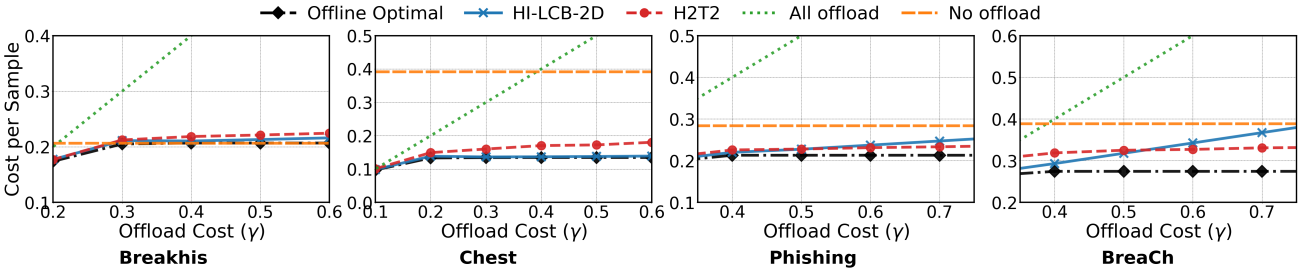


Figure 3. Per-sample cost vs. mean offloading cost γ across four datasets. HI-LCB-2D tracks the Offline Optimal; H2T2 deviates, and All Offload grows linearly with γ ($\delta_1 = 0.7, \delta_{-1} = 1, \alpha = 0.52$).

All experimental results are reported as averages over 10 independent rounds, with each round consisting of 100,000 independent copies of images obtained from individual datasets.

Convergence (Figure 2). After a brief exploration phase, HI-LCB-2D rapidly tracks the offline optimum across all datasets; H2T2 closes the gap more slowly, consistent with its $O(T^{2/3})$ rate.

Sensitivity to γ (Figure 3). HI-LCB-2D closely follows the offline optimum across the full feasible range of γ, including near the no-offload transition (Lemma 3.2); H2T2 adapts less cleanly near this boundary.

Effect of α (Figure 4). Cost increases with α as predicted by the 4α factor in Theorem 5.1; moderate values just above the theoretical minimum of 0.5 give near-optimal performance without over-exploring.

Sensitivity to δ_1 *and* δ_{-1} (Figures 7a–7b, Appendix C). As each penalty grows, the corresponding acceptance region shrinks and offloading increases; HI-LCB-2D matches or outperforms H2T2 across most datasets and cost configurations, with the OOD dataset under high δ_{-1} being the notable exception; its centrally-peaked confidence distribution causes most samples to fall in Φ_{amb}, driving elevated offloading regardless of penalty (Appendix B).

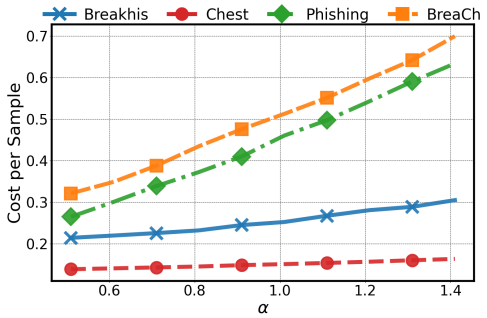


Figure 4. Average per-sample cost vs. α across four datasets. The linear increase confirms the 4α scaling in Theorem 5.1 ($\delta_1 = 0.7, \delta_{-1} = 1, \gamma = 0.35$).

7. Conclusions

We studied online learning for hierarchical inference under asymmetric misclassification costs, a setting arising in spectrum sensing, medical diagnosis, and intrusion detection. The central question, left open by Moothedath et al. (2026), was whether the mild monotonicity assumption on local accuracy is sufficient to move from polynomial to logarithmic regret. We answered this affirmatively with HI-LCB-2D. which learns two class-specific offloading thresholds online via independent confidence bounds without prior knowledge of local accuracy or channel cost, and proved the first $O(\log T)$ regret guarantee for asymmetric-cost hierarchical inference. Experiments on real-world datasets demonstrate that HI-LCB-2D matches or surpasses prior state-of-the-art performance across diverse cost configurations.

References

Abernethy, J. D., Amin, K., and Zhu, R. Threshold bandits, with and without censored feedback. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 4889–4897, 2016.

Al-Atat, G., Datta, P., Moharir, S., and Champati, J. P. Regret bounds for online learning for hierarchical inference. In *ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pp. 281–290, 2024.

Cesa-Bianchi, N., Lugosi, G., and Stoltz, G. Regret minimization under partial monitoring. *Mathematics of Operations Research*, 31(3):562–580, 2006.

Chattopadhyay, S., Sutar, V., Champati, J. P., and Moharir, S. Low-regret and low-complexity learning for hierarchical inference. *arXiv preprint arXiv:2508.08985*, 2025.

Elkan, C. The foundations of cost-sensitive learning. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 973–978, 2001.

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pp. 1321–1330, 2017.

Kang, Y., Hauswald, J., Gao, C., Rovinski, A., Mudge, T., Mars, J., and Tang, L. Neurosurgeon: Collaborative intelligence between the cloud and mobile edge. In *22nd International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, pp. 615–629, 2017.

Mohamed, H. Chest ct-scan images dataset, 2025. URL https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images.

Moothedath, V. N., Agarwal, U., N, U., Gross, J. R., Champati, J. P., and Moharir, S. Inference offloading for cost-sensitive binary classification at the edge. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-26)*, pp. 24449–24457, Mar. 2026.

Spanhol, F. A., Oliveira, L. S., Petitjean, C., and Heutte, L. A dataset for breast cancer histopathological image classification. *Ieee transactions on biomedical engineering*, 63(7):1455–1462, 2015.

Tan, C. L. Phishing dataset for machine learning: Feature evaluation. *Mendeley Data*, 1(8), 2018.

Teerapittayanon, S., McDanel, B., and Kung, H. BranchyNet: Fast inference via early exiting from deep neural networks. In *23rd International Conference on Pattern Recognition (ICPR)*, pp. 2464–2469. IEEE, 2016.

Tiwari, S. Phishing dataset for machine learning, 2025. URL https://www.kaggle.com/datasets/shashwatwork/phishing-dataset-for-machine-learning.

A. Empirical Validation of the Monotonicity Assumption

We verify the non-decreasing monotonicity of $f(\phi) = \mathrm{P}\{y_t = 1 \mid \bar{\phi}_t\}$ (Section 2) on the **Phishing** dataset (Tiwari, 2025; Tan, 2018) (5,000 phishing + 5,000 legitimate samples, logistic regression local model). Samples are grouped into equal-width bins of ϕ_1; within each bin $f(\phi_t)$ is the fraction labelled class 1 by the edge server.

Figure 5 confirms the assumption: $f(\phi_t)$ rises monotonically from 0 to 1 as ϕ_1 increases (left panel), and falls symmetrically as $\phi_0 = 1 - \phi_1$ increases (right panel). This ordinal reliability is what allows HI-LCB-2D to certify local decisions from confidence intervals on $f(\phi_i)$ and achieve $O(\log T)$ regret.

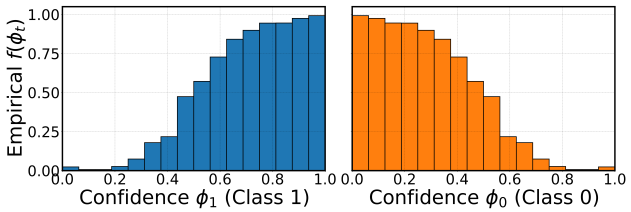


Figure 5. Empirical $f(\phi_t)$ vs. confidence score (Phishing, logistic regression). $f(\phi_t)$ is monotone increasing in ϕ_1 (left) and monotone decreasing in $\phi_0 = 1 - \phi_1$ (right), validating the non-decreasing assumption on $f(\cdot)$.

B. Confidence Distribution and OOD Behaviour

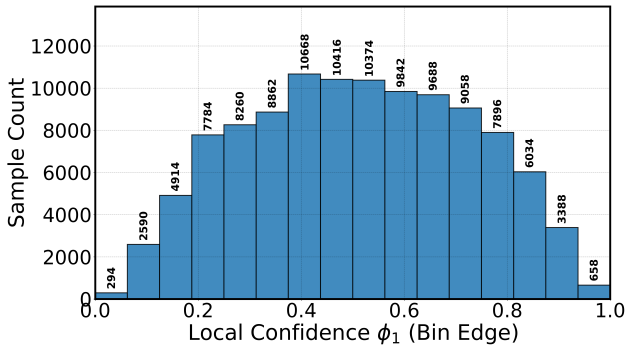


Figure 6. Sample count per confidence bin, BreaCh (OOD) dataset. The centrally-peaked distribution reflects model uncertainty under domain shift, pushing most samples into Φ_{amb} and driving elevated offloading.

The partition of samples across Φ_0, Φ_{amb}, and Φ_1 depends directly on the shape of the confidence distribution. For the **BreaCh** dataset, constructed by applying the Chest CT-scan model to all 7,909 BreakHis images, creating a deliberate OOD scenario, Figure 6 shows the sample count per equal-width bin of ϕ_1.

The distribution is centrally peaked ($\approx$10,000 samples per bin near $\phi_1 \in [0.3, 0.6]$; only 294 and 658 at the extremes), reflecting model uncertainty under domain shift. Consequently, most samples fall in Φ_{amb} and are offloaded, producing the near-linear cost growth with γ in the experiments.

C. Sensitivity to Asymmetric Cost Parameters

Figure 7 shows per-sample cost as δ_1 and δ_{-1} are varied independently. As each penalty grows, the corresponding threshold shifts and the optimal acceptance region shrinks. HI-LCB-2D adapts each boundary independently and matches or outperforms H2T2 across most datasets and penalty values; the advantage is most pronounced at higher penalties where accurate boundary estimation matters. The OOD dataset under high δ_{-1} is an exception where H2T2 converges to the better steady-state policy faster, likely due to its confidence distribution concentrating near the lower threshold.

D. Detailed Proofs for Regret Analysis

This section provides the full statements and proofs of the three lemmas that underpin Theorem 5.1. Together they establish the confidence bound validity guarantee, bound the number of incorrect local decisions, and bound the exploration count at each confidence level.

Sub-optimality Gaps

For each $\phi_i \in \Phi$, the five sub-optimality gaps are

$$\Delta_{\phi_i}^{A0\text{-}O} = |\delta_{-1} f(\phi_i) - \gamma|, \tag{14}$$

$$\Delta_{\phi_i}^{O\text{-}A1} = |\gamma - \delta_1(1 - f(\phi_i))|, \tag{15}$$

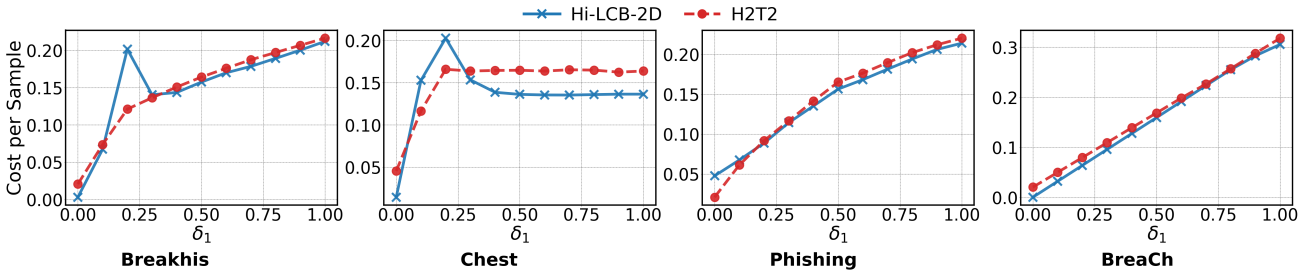
$$\Delta_{\phi_i} = \left| f(\phi_i) - \frac{\delta_1}{\delta_1 + \delta_{-1}} \right|, \tag{16}$$

$$\beta_{\phi_i} = \gamma - \delta_{-1} f(\phi_i) \quad (\phi_i \in \Phi_0), \tag{17}$$

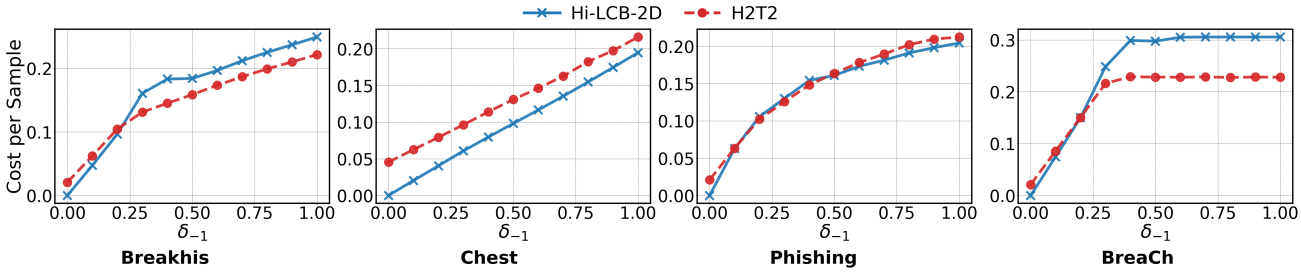
$$\kappa_{\phi_i} = \gamma - \delta_1(1 - f(\phi_i)) \quad (\phi_i \in \Phi_1). \tag{18}$$

$\Delta_{\phi_i}^{A0\text{-}O}$ is the per-step regret of an unnecessary offload at $\phi_i \in \Phi_0$; $\Delta_{\phi_i}^{O\text{-}A1}$ plays the same role in Φ_1. Δ_{ϕ_i} measures how far $f(\phi_i)$ is from the class boundary $\delta_1/(\delta_1 + \delta_{-1})$; a small value means the second sub-condition of $\mathcal{C}_0$ or $\mathcal{C}_1$ is hard to certify, requiring many offloads before the algorithm can distinguish which class the sample belongs to. β_{ϕ_i} and κ_{ϕ_i} measure how far the expected local cost is from the offloading cost at levels in Φ_0 and Φ_1 respectively; a small value means the first sub-condition of the certification condition is hard to satisfy, requiring more exploration. Together, these five quantities fully characterize the difficulty of learning the two thresholds.

Lemma D.1 (Confidence Bound Validity)**.** *For any t and $\phi_i \in \Phi$, each one-sided deviation, for example* $\{\mathrm{LCB}_{\phi_i} >$



(a) Cost vs. false-positive penalty δ_1 ($\delta_{-1} = 0.7, \alpha = 0.52, \gamma = 0.35$).



(b) Cost vs. false-negative penalty δ_{-1} ($\delta_1 = 0.7, \alpha = 0.52, \gamma = 0.35$).

Figure 7. Sensitivity to asymmetric cost parameters across four datasets (Breakhis, Chest, Phishing, OOD). HI-LCB-2D adapts each boundary independently and outperforms H2T2 across both sweeps.

$f(\phi_i)\}$, *has probability at most* $t^{-2\alpha}$. *The same bound holds for the deviations of* $\hat{\gamma}$ *and* LCB_γ.

Proof. Fix $\phi_i \in \Phi$. The correctness indicators $\mathbb{1}\{h_l(x_s) = y_s\}$ for the $O_{\phi_i}(t)$ offloads at confidence level ϕ_i up to time t are i.i.d. Bernoulli with mean $f(\phi_i)$. By Hoeffding's inequality, for any $\varepsilon > 0$,

$$\mathbb{P}\Big\{\hat{f}(\phi_i) - f(\phi_i) \geq \varepsilon\Big\} \leq \exp\big(-2\, O_{\phi_i}(t)\, \varepsilon^2\big).$$

Setting $\varepsilon = \sqrt{\alpha \log t / O_{\phi_i}(t)}$ gives probability $\leq \exp(-2\alpha \log t) = t^{-2\alpha}$. The one-sided bound $\{\mathrm{UCB}_{\phi_i} < f(\phi_i)\}$ follows by the same argument in the other direction. Identical reasoning applies to $\hat{\gamma}$ and LCB_γ since the Γ_t are i.i.d. in $[0, 1]$. □

Lemma D.2 (Opposite-Class and Ambiguous-Region Errors)**.** *For* $\alpha > 0.5$*, the expected number of rounds in which the algorithm accepts the wrong class for any fixed* ϕ_i *is at most* $\zeta(2\alpha)$*, where* ζ *is the Riemann zeta function. For* $\phi_i \in \Phi_{\mathrm{amb}}$*, the total expected incorrect local acceptances are at most* $3\zeta(2\alpha) = O(1)$*. The constraint* $\alpha > 0.5$ *ensures* $\zeta(2\alpha) < \infty$*, so both error counts are finite and contribute only a constant to total regret.*

Proof. **Opposite-class errors.** Consider $\phi_i \in \Phi_0$. The algorithm accepts class 1 (condition $\mathcal{C}_1$) only if $\mathrm{LCB}_{\phi_i} \geq \delta_1/(\delta_1 + \delta_{-1})$. Since $f(\phi_i) \leq \gamma/\delta_{-1} < \delta_1/(\delta_1 + \delta_{-1})$ (by Lemma 3.3), this requires $\mathrm{LCB}_{\phi_i} > f(\phi_i)$, which by Lemma D.1 occurs with probability at most $t^{-2\alpha}$. The expected number of such events over all $t \geq 1$ is $\sum_{t=1}^{\infty} t^{-2\alpha} = \zeta(2\alpha) < \infty$ for $\alpha > 1/2$. The same bound holds for Φ_1 by symmetry.

Ambiguous-region errors. For $\phi_i \in \Phi_{\mathrm{amb}}$, offloading is optimal. The algorithm accepts class 0 (condition $\mathcal{C}_0$) only if $\delta_{-1}\,\mathrm{UCB}_{\phi_i} < \mathrm{LCB}_\gamma$. Since $\delta_{-1} f(\phi_i) > \gamma$ (as $\phi_i \notin \Phi_0$), this requires either $\mathrm{UCB}_{\phi_i} < f(\phi_i)$ or $\mathrm{LCB}_\gamma > \gamma$, each with probability at most $t^{-2\alpha}$ by Lemma D.1. By a union bound, the probability that $\mathcal{C}_0$ fires incorrectly is at most $2t^{-2\alpha}$. Similarly for $\mathcal{C}_1$. The total expected incorrect local acceptances in Φ_{amb} is at most $3\sum_{t=1}^{\infty} t^{-2\alpha} = 3\zeta(2\alpha) = O(1)$. □

Lemma D.3 (Offload Count Bound)**.** *The algorithm offloads at* ϕ_i *until both sub-conditions of the certification condition are satisfied. The first sub-condition, certifying that the worst-case local cost is below the best-case offloading cost, requires the confidence intervals on* $f(\phi_i)$ *and* γ *to narrow enough that their overlap is resolved; this is governed by* β_{ϕ_i} *(or* κ_{ϕ_i}*). The second sub-condition, certifying which class the sample belongs to, requires the interval on* $f(\phi_i)$ *to lie clearly on one side of the class boundary; this is governed by* Δ_{ϕ_i}*. Whichever sub-condition is harder to satisfy determines the total offload count. Formally, for* $\phi_i \in \Phi_0$*:*

$$\mathbb{E}[O_{\phi_i}(T)] \leq 4\alpha \log T \cdot \max\Big\{\Delta_{\phi_i}^{-2},\, (\delta_{-1}+1)^2 \beta_{\phi_i}^{-2}\Big\} \tag{19}$$

$$+\, 3\zeta(2\alpha). \tag{20}$$

An analogous bound holds for $\phi_i \in \Phi_1$ *with* κ_{ϕ_i} *replacing*

β_{ϕ_i} *and* δ_1 *replacing* δ_{-1}.

Proof. Fix $\phi_i \in \Phi_0$. The algorithm offloads at ϕ_i in round t when $\mathcal{C}_0$ does not hold (ignoring the negligible opposite-class events bounded in Lemma D.2). Condition $\mathcal{C}_0$ fails when at least one of the following holds:

$$\mathcal{E}_1(t) : \delta_{-1}\,\mathrm{UCB}_{\phi_i}(t) \geq \mathrm{LCB}_{\gamma}(t),$$
$$\mathcal{E}_2(t) : \mathrm{UCB}_{\phi_i}(t) \geq \delta_1/(\delta_1+\delta_{-1}).$$

Event $\mathcal{E}_1$. When both estimates are within their confidence intervals (which fails with probability at most $2t^{-2\alpha}$ by Lemma D.1), $\mathrm{UCB}_{\phi_i} \leq f(\phi_i) + \sqrt{\alpha \log t / O_{\phi_i}}$ and $\mathrm{LCB}_{\gamma} \geq \gamma - \sqrt{\alpha \log t / O_{\gamma}}$. Event $\mathcal{E}_1$ then requires

$$\delta_{-1}\Big(f(\phi_i) + \sqrt{\tfrac{\alpha \log t}{O_{\phi_i}}}\Big) \geq \gamma - \sqrt{\tfrac{\alpha \log t}{O_{\gamma}}}.$$

Since $\beta_{\phi_i} = \gamma - \delta_{-1} f(\phi_i) > 0$, this rearranges to $\delta_{-1}\sqrt{\alpha \log t / O_{\phi_i}} + \sqrt{\alpha \log t / O_{\gamma}} \geq \beta_{\phi_i}$. Using $O_{\gamma} \geq O_{\phi_i}$, the condition can only hold while $O_{\phi_i} \leq 4\alpha(\delta_{-1}+1)^2 \beta_{\phi_i}^{-2} \log T$.

Event $\mathcal{E}_2$. When the confidence bound is valid, $\mathcal{E}_2$ requires $f(\phi_i) + \sqrt{\alpha \log t / O_{\phi_i}} \geq \delta_1/(\delta_1+\delta_{-1})$. Since $f(\phi_i) < \delta_1/(\delta_1+\delta_{-1}) - \Delta_{\phi_i}$, this can only hold while $O_{\phi_i} \leq 4\alpha \Delta_{\phi_i}^{-2} \log T$.

Combining, the expected number of offloads at ϕ_i before $\mathcal{C}_0$ is permanently certified is

$$\mathbb{E}[O_{\phi_i}(T)] \leq 4\alpha \log T \cdot \max\{\Delta_{\phi_i}^{-2},\, (\delta_{-1}+1)^2 \beta_{\phi_i}^{-2}\} + 3\zeta(2\alpha),$$

where the last term accounts for rounds where confidence-bound failures or opposite-class errors occur. The bound for Φ_1 follows by replacing δ_{-1} with δ_1 and β_{ϕ_i} with κ_{ϕ_i}. □

E. Proofs

E.1. Proof of Lemma 3.1 (Optimal Offline Decision)

The expected cost of accepting class 0 for a sample with confidence score ϕ_t is

$$\mathcal{L}_t(0) = \delta_{-1}\,\mathbb{P}\{y_t = 1 \mid \bar{\phi}_t\} = \delta_{-1} f(\phi_t),$$

since a false negative (true class 1, predicted class 0) costs δ_{-1}. Symmetrically, the expected cost of accepting class 1 is

$$\mathcal{L}_t(1) = \delta_1\,\mathbb{P}\{y_t = 0 \mid \bar{\phi}_t\} = \delta_1\big(1 - f(\phi_t)\big).$$

Offloading incurs cost Γ_t with $\mathbb{E}[\Gamma_t] = \gamma$, so $\mathcal{L}_t(\mathrm{off}) = \gamma$. Since the three costs are additive and independent, the myopically optimal decision is $D_t^{\pi^*} = \arg\min_d \mathcal{L}_t(d)$. Under i.i.d. sample arrivals, this per-sample optimum also minimizes the cumulative expected cost. □

E.2. Proof of Lemma 3.2 (No-Offload Condition)

Offloading is optimal for sample x_t only if $\gamma < \min\{\delta_{-1} f(\phi_t),\, \delta_1(1 - f(\phi_t))\}$. The maximum of $\min\{\delta_{-1} p,\, \delta_1(1-p)\}$ over $p \in [0,1]$ is achieved at the crossing point $\delta_{-1} p^* = \delta_1(1-p^*)$, giving $p^* = \delta_1/(\delta_1+\delta_{-1})$ and peak value $\delta_1\delta_{-1}/(\delta_1+\delta_{-1})$.

If $\gamma > \delta_1\delta_{-1}/(\delta_1+\delta_{-1})$, then for every $p \in [0,1]$ we have $\gamma > \min\{\delta_{-1} p,\, \delta_1(1-p)\}$, so a local decision (0 or 1) is always at least as cheap as offloading. Offloading is therefore never optimal. □

E.3. Proof of Lemma 3.3 (Two-Threshold Structure)

We identify when each decision is optimal using Lemma 3.1.

Region Φ_0. $D^* = 0$ requires $\delta_{-1} f(\phi) \leq \delta_1(1 - f(\phi))$ and $\delta_{-1} f(\phi) \leq \gamma$:

(i) $\delta_{-1} f \leq \delta_1(1-f)$ iff $f \leq \delta_1/(\delta_1+\delta_{-1})$.
(ii) $\delta_{-1} f \leq \gamma$ iff $f \leq \gamma/\delta_{-1}$.

Under the assumption $\gamma < \delta_1\delta_{-1}/(\delta_1+\delta_{-1})$, dividing by δ_{-1} gives $\gamma/\delta_{-1} < \delta_1/(\delta_1+\delta_{-1})$, so condition (ii) is binding. Thus $\Phi_0 = \{\phi : f(\phi) \leq \gamma/\delta_{-1}\}$ with $\theta_l^* = f^{-1}(\gamma/\delta_{-1})$.

Region Φ_1. $D^* = 1$ requires $\delta_1(1-f) \leq \delta_{-1} f$ and $\delta_1(1-f) \leq \gamma$:

(i) $\delta_1(1-f) \leq \delta_{-1} f$ iff $f \geq \delta_1/(\delta_1+\delta_{-1})$.
(ii) $\delta_1(1-f) \leq \gamma$ iff $f \geq 1 - \gamma/\delta_1$.

Under the stated assumption, $1 - \gamma/\delta_1 > \delta_1/(\delta_1+\delta_{-1})$, so condition (ii) binds: $\Phi_1 = \{\phi : f(\phi) \geq 1 - \gamma/\delta_1\}$ with $\theta_u^* = f^{-1}(1 - \gamma/\delta_1)$.

Coincidence condition. $\theta_l^* = \theta_u^*$ iff $\gamma/\delta_{-1} = 1 - \gamma/\delta_1$, i.e., $\gamma(\delta_1+\delta_{-1}) = \delta_1\delta_{-1}$. This is the boundary of the offloading regime (Lemma 3.2). For γ strictly below this boundary, the two thresholds are distinct unless $\delta_1 = \delta_{-1}$. □

E.4. Proof of Theorem 5.1 ($O(\log T)$ Regret)

Decompose the cumulative regret by region:

$$R_T(\pi) = \underbrace{\sum_{\phi_i \in \Phi_0} \Delta_{\phi_i}^{A_0\text{-}O}\, \mathbb{E}[O^{\phi_i}(T)]}_{R_{\Phi_0}(T)} + R_{\Phi_{\mathrm{amb}}} + \underbrace{\sum_{\phi_i \in \Phi_1} \Delta_{\phi_i}^{O\text{-}A_1}\, \mathbb{E}[O^{\phi_i}(T)]}_{R_{\Phi_1}(T)}.$$

Φ_0 contribution. Each unnecessary offload at $\phi_i \in \Phi_0$ costs $\Delta_{\phi_i}^{A_0\text{-}O} = \gamma - \delta_{-1} f(\phi_i) = \beta_{\phi_i}$ more than the optimal local action. Applying Lemma D.3 and multiplying by $\Delta_{\phi_i}^{A_0\text{-}O}$ gives $R_{\Phi_0}(T)$ term.

Φ_1 **contribution.** The same argument with $\kappa_{\phi_i} = \Delta_{\phi_i}^{O\text{-}A_1}$ gives $R_{\Phi_1}(T)$ term.

Φ_{amb} **contribution.** In Φ_{amb} offloading is optimal, so correct offloads incur no regret. Incorrect local acceptances are bounded by $3\zeta(2\alpha) \cdot \max_d \mathcal{L}_t(d) = O(1)$ by Lemma D.2.

Assembling the bound. Since $|\Phi| < \infty$ and all gaps Δ_{ϕ_i}, β_{ϕ_i}, κ_{ϕ_i} are strictly positive (the regions Φ_0, Φ_1, Φ_{amb} are well-separated by Lemma 3.3), both sums are finite. The dominant term scales as $4\alpha \log T$, giving $R_T(\pi) = O(\log T)$. □