AI & MEDICAL IMAGE PROCESSING

Risk-Aware Adaptive Image Transmission for Reliable Skin Lesion Screening in Low-Bandwidth Settings

Pranjal Jha, Vishesh Panghal

Department of Artificial Intelligence and Data Science, Poornima Institute of Engineering and Technology (PIET), Jaipur, Rajasthan, India

Published November 2026

Abstract

Transmitting high-resolution images may cause long delays in bandwidth-limited tele-dermatology, whereas uniform lossy compression may cause small malignant lesions to be missed. This work explores a risk-aware two-stage adaptive transmission heuristic for skin lesion screening. An aggressively compressed image (Q = 30) is first sent from the edge client; if a composite of referral risk and predictive entropy exceeds an empirically derived threshold, a high-quality image (Q = 90) is requested from the edge buffer. On the leakage-free DermaMNIST-C benchmark (28 × 28), the adaptive strategy achieves a referral sensitivity of 96.0% (143/149), within 0.7 percentage points of continuous Q = 90 transmission (144/149), while reducing entropy-coded scan data by 30.8%. A separate model at native resolution (600 × 450) estimates a 57.9% reduction in average payload (24.6 KB vs. 58.4 KB); diagnostic performance on native images was not evaluated. The adaptive strategy achieved a specificity of 99.0% for routine cases, a precision of 92.9% for referred cases and a melanoma recall of 98.6% (69/70), and only one test case bypassed retransmission (0.7%; 95% CI: 0.02%–3.68%). These results show that risk- and entropy-based triage heuristics, combined with failure auditing, can be beneficial in bandwidth-limited tele-screening.

Keywords: Tele-dermatology; adaptive transmission; image compression; rate-distortion; uncertainty estimation; low-bandwidth networks; DermaMNIST-C

1 Introduction

Dermatology related cancers such as skin melanoma, Basal Cell Carcinoma (BCC) and Actinic Keratoses (AKIEC) have been increasingly becoming a problem globally. Early diagnosis and removal plays an important role in the prognosis of these diseases [1],[2]. If melanomas are diagnosed at an early stage then there is over a 99% chance of patients surviving for over 5 years. However, if they aren't diagnosed until later stages then less than 30% of patients will survive. In rural areas there is often a lack of board certified dermatologists creating large backlogs in diagnosis. Patients often do not get referred to specialist dermatologists for many months.

Recent advances in Deep Convolution Neural Networks and Vision Transformers have led to very high diagnostic performance on complex dermoscopic classification tasks [3–5]. This has resulted in store and forward tele-dermatology becoming a critical paradigm whereby community-based health workers can obtain digital dermoscopy images using mobile attachments at local clinics. These are then transmitted to remote diagnostic servers via wireless connections [6–8].

However, telemedicine suffers due to poor communication. Modern dermoscopy devices provide high resolution images but tele dermoscopy occurs in remote locations where data can only be sent via low bandwidth asymmetric connections such as old cellular connections (2G/3G), wireless or IoT satellite connections with very strict bandwidth limits [8–10]. Transmitting high resolution data will cause long delays, congestion and expensive data costs.

To reduce the impact of high network usage, fixed lossy compression methods are used including JPEG [11] or modern representations [12, 13].

Traditional lossy techniques aim to preserve the image quality at a pixel level [14]. However this is not good for dermoscopy. Lossy compression can cause significant problems, especially when using high levels of spatial compression. High frequency transform coefficients are removed which can reduce the quality of the image by removing fine details [15, 16]. This can cause issues with diagnostic features. The aggressive compression can cause malignant features to become smoother and appear benign. This can increase the risk of missing malignant lesions as they can be mistaken for benign macules.

This issue is further complicated by concerns about benchmark validity. Landmark studies have shown that common benchmarks such as HAM10000 [1] and MedMNIST v2 [17] exhibit severe lesion-level data leakage when random split strategies are used [18, 19]. Duplicate captures are distributed between partitions, causing classifiers to memorise acquisition artefacts [16, 19]. To measure rate-distortion performance reliably, partitions must be completely lesion-disjoint, as implemented in the DermaMNIST-C benchmark [18].

We propose a feedback-driven two-stage adaptive framework incorporating rate-distortion theory [20] and task informed Information Bottleneck principles [21]. Although importance-aware re-transmission techniques exist for general wireless edge learning scenarios [22], conventional protocols do not consider clinical prioritization or distortion incurred by lossy compression. Our approach considers that not all bits have equal value in a clinical scenario. Bandwidth is therefore requested dynamically depending on how complex/urgent the case is. This enables cases with clear benign features to be triaged using low bit-rates whilst ambiguous or high risk lesions can be transmitted in high fidelity. An ultra compressed image (Q = 30) is sent from the client-side. The server then uses a deep network to calculate the total referral probability and normalised predictive information entropy of the input ($P_{\text{referral}}(x)$ and $H_{\text{norm}}(\mathbf{p}(x))$). A total score is calculated based on these values, $s(x) = \alpha P_{\text{referral}}(x) + (1 - \alpha)H_{\text{norm}}(\mathbf{p}(x))$. If this score reaches a threshold ($s(x) \geq \tau$), the high-quality candidate (Q = 90) stored on the client is requested so that a definitive triage can be performed. We use selective classification [23] and predictive calibration [24] to derive and analyse the silent failure rate (SFR): the critical failure scenario in which overconfident false negatives occur because of compression distortion.

The primary contributions of this paper are summarised as follows:

- **Risk-Entropy Triage Heuristic:** A heuristic is proposed for use in practice in cloud-edge systems which balances the risk of referral alongside predictive Shannon entropy for deciding when to selectively re-transmit data in order to save bandwidth.
- **Leakage-Controlled Benchmark Evaluation:** We provide experimental rate-distortion results using the leakage-free DermaMNIST-C benchmark [18]. This therefore demonstrates that the trade-off between diagnostic performance is due to true generalization in the absence of any leakage.

- **Latency and Metric Modelling:** An analytical expression is derived for the uplink latency which depends on the network bandwidth B and round-trip time (RTT). The application of adaptive triage to full resolution medical imaging data gives an estimated reduction of 57.9% of the payload, whilst preserving 96.0% of the benchmark sensitivity for 28 × 28 sized images. This is only 0.7 percentage points less than Q = 90 (143/149 versus 144/149).
- **Silent Failure Auditing:** A new metric, the Silent Failure Rate (SFR), has been defined which separates the false negatives from the classifier due to the compression being bypassed. An exact 95% confidence interval has been calculated on the SFR using Clopper Pearson for both images and lesions for auditing purposes.

The rest of this paper is structured as follows. The next section describes the system model and the triage policy. The experimental setup and telecommunication profiles are then described, followed by the results for rate-distortion, latency and failures. The final section concludes the paper.

2 System model and problem formulation

Operational label grouping and screening priority. The diagnostic problem is defined in terms of $K = 7$ dermoscopy classes, where the label set is $\mathcal{Y} = \{1, \ldots, K\}$ = {melanoma (MEL), basal cell carcinoma (BCC), actinic keratoses and intraepithelial carcinoma (AKIEC), benign keratosis-like lesions (BKL), melanocytic nevi (NV), dermatofibroma (DF), vascular lesions (VASC)}. We define operational class groups: those requiring referral for further diagnostic consideration ($\mathcal{Y}_{\text{ref}}$ = {MEL, BCC, AKIEC}) and those observed routinely ($\mathcal{Y}_{\text{routine}} = \mathcal{Y}\ \mathcal{Y}_{\text{ref}}$ = {BKL, NV, DF, VASC}). For an input image x, a convolutional classifier outputs a normalised posterior probability distribution across the classes, $\mathbf{p}(x) = [p_1(x), \ldots, p_K(x)]$. The cumulative posterior probability that a sample belongs to one of the referral classes is

$$P_{referral}(x) = \sum_{c \in Y_{ref}} p_c(x) \tag{1}$$

Lossy quantisation and rate-distortion modelling. Let x_{raw} be an RGB image of size $H \times W \times 3$, representing the raw dermoscopic image taken at the point of care. Using the standard JPEG codec, x_{raw} is transformed into the appropriate colour space, and each 8 × 8 block of pixels is transformed using the discrete cosine transform (DCT), giving coefficients $C_{\text{u,v}}$ that represent the frequency content of x_{raw}. Spatial compression is controlled by a quality factor $q \in [1, 100]$, which scales the standard quantisation matrices $Q_{\text{u,v}}(q)$ for luminance and chrominance:

$$\widehat{C_{u,v}} = \text{round}\left(\frac{C_{u,v}}{Q_{u,v}(q)}\right) \tag{2}$$

where $C_{\text{u,v}}$ is the DCT coefficient at frequency (u, v), $Q_{\text{u,v}}(q)$ is the corresponding quality-scaled quantisation step and $\hat{C}_{\text{u,v}}$ is the quantised coefficient.

Aggressive compression ($q \ll 50$) quantises high-frequency transform coefficients to zero, thereby removing fine morphological features such as fragile atypical pigment networks, streaks and arborising telangiectasia.

From a rate-distortion perspective, let $R(x_q)$ = Bytes(x_q) be the transmission rate. The diagnostic distortion $D(\pi) = 1 - \text{Sens}(\mathcal{Y}_{\text{ref}})$ is the error rate of the referral classification. The aim of tele-screening becomes

$$\min_{\pi} \mathrm{E}_{x\sim \mathrm{D}}\left[\text{Payload}\,(x)\right] \text{s.t.} \text{Sens}\left(\mathrm{Y}_{ref}\right) \geq 1 - \epsilon \tag{3}$$

where $\epsilon \in (0, 1)$ is the permissible bound on clinical sensitivity degradation.

Predictive uncertainty and composite decision metric. Compression artefacts frequently induce inter-class predictive ambiguity. To quantify the model's distributional uncertainty independently of class identity, we compute the normalised Shannon entropy:

$$H_{norm}(p(x)) = -\frac{1}{\log K}\sum_{k=1}^{K} p_k(x)\log p_k(x) \tag{4}$$

The overall architecture of the proposed framework is shown in Figure 1. The edge dermatoscope captures a high-resolution image, transmits an aggressively compressed version (Q = 30) in Stage 1 and awaits feedback. The diagnostic server evaluates the composite score $s(x)$; if $s(x) \geq \tau$, a 1-byte feedback packet requests high-quality retransmission (Q = 90), otherwise the fast screening triage is confirmed. The payloads shown in the figure are modelled native-resolution values.

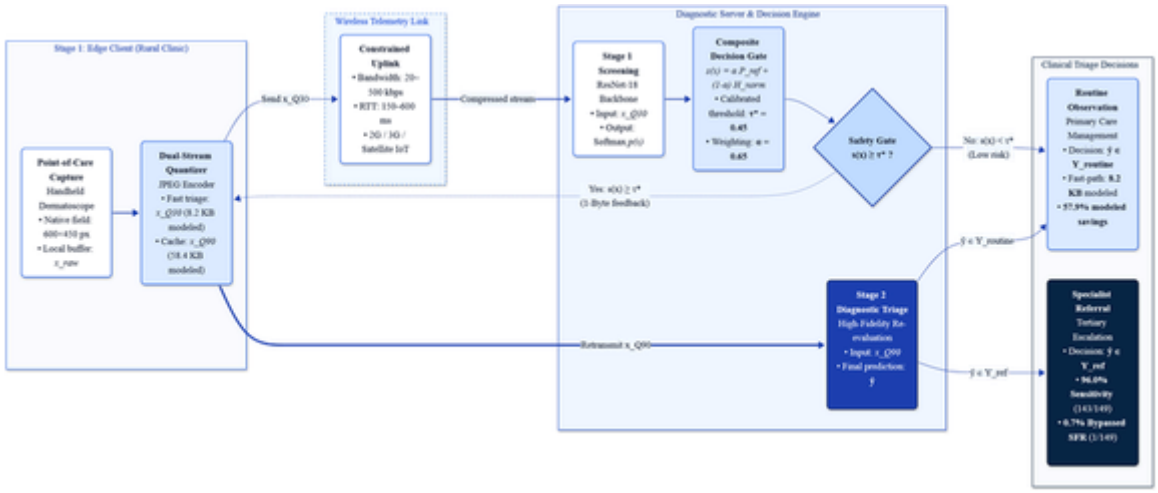


Figure 1. System architecture of the proposed risk-aware closed-loop adaptive transmission framework.

We define the composite transmission score $s(x)$ as a convex combination of the clinical referral probability and the predictive entropy:

$$s(x) = \alpha \cdot P_{referral}(x) + (1-\alpha)\cdot H_{norm}(p(x)) \tag{5}$$

where $\alpha \in [0, 1]$ is a tunable weighting hyperparameter balancing high stakes clinical prioritization against epistemic and aleatoric predictive ambivalence.

Two-stage closed-loop adaptive transmission policy. Let x_{Q30} denote the aggressively compressed initial screening transmission (q = 30) and x_{Q90} the high-fidelity representation (q = 90) cached locally in the edge client buffer. The diagnostic server computes the composite score $s(x_{Q30})$ and executes the acquisition decision policy:

$$\text{Request}(x) = \begin{cases} 0, & s(x_{Q30}) < \tau \text{(fast path)} \\ 1, & s(x_{Q30}) \geq \tau \text{(retransmit } x_{Q90}\text{)} \end{cases} \tag{6}$$

When $s(x_{Q30}) \geq \tau$, the server transmits a 1-byte control packet requesting x_{Q90}. Upon receipt, the edge client offloads the cached candidate x_{Q90}, and the server executes Stage 2 high-fidelity inference $\hat{y}(x_{Q90})$. The resulting triage action is determined by whether the definitive prediction $\hat{y}$ maps to the referral-priority subset $\mathcal{Y}_{\text{ref}}$ (specialist referral) or the routine subset $\mathcal{Y}_{\text{routine}}$ (outpatient observation):

$$\text{Triage}(x) = \begin{cases} \text{Specialist referral}, & \hat{y}(x^*) \in \text{Y}_{ref} \\ \text{Routine observation}, & \hat{y}(x^*) \in \text{Y}_{routine} \end{cases} \tag{7}$$

where $x^* = x_{Q90}$ if $s(x_{Q30}) \geq \tau$, and x_{Q30} otherwise, explicitly decoupling adaptive acquisition from specialist referral. The hyperparameter pair (α^*, τ^*) is calibrated on the disjoint validation set $\mathcal{D}_{\text{val}}$ (N = 573; 68 referral and 505 routine cases) by constrained empirical optimisation relative to full-fidelity Q = 90:

$$(\alpha^*, \tau^*) = \arg\min_{(\alpha,\tau)} \sum_{i \in \text{D}_{val}} \text{Payload}(x_i) \text{s.t.} \text{Sens}(\text{Y}_{ref}) \geq \eta_{target} \tag{8}$$

where full-fidelity Q = 90 attains $\text{Sens}_{Q90}(\mathcal{D}_{\text{val}})$ = 97.1% (66/68 cases, 97.06%), and we apply an empirical validation constraint η_{target} = 96.0% (an operational margin of ≈ 1.1 percentage points from Q = 90 on Dval), serving as a tuning constraint rather than established clinical non-inferiority. Coarse 2D grid search (α, τ step 0.05) and 1D fine refinement (Δ = 0.01) along this boundary selected (α^* = 0.65, τ^* = 0.45), minimizing validation payload (23.9 KB) at $\text{Sens}(\mathcal{D}_{\text{val}}) \geq 96.0\%$. Single-metric ablations selected τ_{ent} = 0.48 (α = 0.0) and τ_{risk} = 0.35 (α = 1.0).

Network payload, latency and silent failure modelling. The effective transmission payload per screened subject is given by

$$\text{Payload}(x) = \text{Bytes}(x_{Q30}) + \text{I}(s(x_{Q30}) \geq \tau) \cdot \text{Bytes}(x_{Q90}) \tag{9}$$

where $\mathbb{I}(\cdot)$ is the indicator function. For an uplink channel with bandwidth B (bps) and round-trip time RTT (s), the analytical transmission latency model $T(x)$ captures over-the-air serialisation and signalling delay:

$$\begin{aligned} T(x) = & \left(\frac{8 \cdot \text{Bytes}(x_{Q30})}{B} + RTT \right) \\ & + \text{I}(s(x_{Q30}) \geq \tau) \left(\frac{8 \cdot \text{Bytes}(x_{Q90})}{B} + RTT \right) \end{aligned} \tag{10}$$

Over constrained uplinks ($B \leq 100$ kbps), serialisation delay dominates response time; benchmark profiling (batch size 1, 1000 runs) shows that on-device CPU JPEG encoding of 28 × 28 captures (≈ 2.4 ms on an ARM Cortex-A53, steady-state CPU timers) and server GPU inference on resized 224 × 224 inputs (≈ 12 ms on an NVIDIA T4, CUDA event timing) contribute negligible overhead.

A safety-critical failure mode in adaptive medical transmission is the occurrence of overconfident false negatives that bypass retransmission, measured by the silent failure rate (SFR). Under the two-stage policy, the total referral false-negative rate decomposes into two mutually exclusive pathways, $\text{FNR}_{\text{total}} = 1 - \text{Sens}(\mathcal{Y}_{\text{ref}}) = \text{SFR} + \text{FNR}_{\text{post}}$, where SFR accounts for malignant lesions ($y \in \mathcal{Y}_{\text{ref}}$) misclassified as routine ($\hat{y} \in \mathcal{Y}_{\text{routine}}$) with a low risk–uncertainty score ($s(x_{Q30}) < \tau$) that bypass retransmission, while FNR_{post} captures residual classification errors after triggered high-fidelity retransmission ($s(x_{Q30}) \geq \tau$):

$$SFR = \text{P}(\hat{y}(x_{Q30}) \in \text{Y}_{routine} \wedge s(x_{Q30}) < \tau \mid y \in \text{Y}_{ref}) \tag{11}$$

$$FNR_{post} = \text{P}(\hat{y}(x_{Q90}) \in \text{Y}_{routine} \wedge s(x_{Q30}) \geq \tau \mid y \in \text{Y}_{ref}) \tag{12}$$

To address statistical uncertainty, we evaluate two cohorts: (i) image level (N_{ref} = 149 referral test images) and (ii) lesion level ($N_{\text{ref,lesion}}$ = 131 unique lesions). Each physical lesion maps to one of three

mutually exclusive outcomes: (1) successfully triaged (at least one capture triggers referral, $\hat{y} \in \mathcal{Y}_{\text{ref}}$); (2) silent failure (every capture misclassified as routine on the fast path, bypassing escalation); or (3) post-retransmission error (retransmission triggered, but the final prediction remained routine). For k silent failures, exact Clopper–Pearson 95% binomial confidence intervals are computed from beta quantiles: [Beta^{-1}(0.025; k, $N - k + 1$), Beta^{-1}(0.975; $k + 1$, $N - k$)].

3 Experimental setup

3.1 Benchmark integrity: DermaMNIST-C dataset

To eliminate data leakage present in legacy benchmarks [18, 17], we evaluate on the leak-free DermaMNIST-C benchmark v1.1 [18] (10,015 images, 7,470 unique lesions across 7 classes: MEL, NV, BKL, BCC, AKIEC, VASC, DF). DermaMNIST-C enforces strict lesion deduplication: 8,215 train, 573 validation, and 1,227 test images (N_{ref} = 149 cases across 131 lesions: MEL 70/63, BCC 44/39, AKIEC 35/29; N_{routine} = 1,078 cases across 1,018 lesions: NV 948, BKL 105, VASC 17, DF 8), ensuring duplicate captures never cross partition boundaries.

3.2 Model architecture and reproducibility protocol

We fine-tune the backbone network (ResNet-18 [25]) for 30 epochs. The Adam optimizer is used with learning rate (η = 10−4). Weight decay is set to 10−4 while a cosine learning schedule is used. We train on 64 samples per batch with inverse class-frequency weighting, image augmentation (including flipping, rotation (±15 ◦) and colour jitter (0.1)). Checkpoints were selected using the validation macro F1 score (N = 573). The temperature scaling is tuned using the validation set (five seeds, 42–46). The temperature (T = 1.18 ± 0.04) is tuned to minimize NLL on the validation set. Similarly thresholding parameters are tuned (α = 0.65, τ = 0.45±0.02). This leads to a decrease in ECE scores from 4.8% to 2.1% (M = 15 evenly spaced bins) at test time on Dtest [26]. Table 1 displays results for seed 42 (same average over 5 seeds). Model weights, split indices and decision thresholds are included to ensure full reproducible performance. The diagnostic interface and telemetry signals use different resolution. Therefore, all classification results are reported as benchmark results while modeled communications results are presented. There is no assumption made regarding dermatology transfer. Diagnostic inference results are obtained by evaluating the DermaMNIST-C benchmark images (28 × 28 × 3). The format is: libjpeg (Q: 30/90) decoding, resizing to 224 × 224, then passing into ResNet-18. The relevant entropy encoded data can be extracted by parsing through the byte stream and extracting data between the Start of Scan (SOS, 0xFFDA) header segment and the End of Image (EOI, 0xFFD9) marker. This provides purely scan data with a total size of 100.8±17.0 B at Q = 30 and 244.8±44.1 B at Q = 90.

This excludes the SOS header segment and container overhead (APP0, DQT, SOF0 and DHT tables) of approximately 600 B (totals: 700.8 ± 17.0 B and 844.8 ± 44.1 B, respectively). Using the representative mean model ($L_{90|\text{req}} \approx L_{90}$), adaptive screening transmits 100.8 + (344/1227)(244.8) = 169.4 B of scan data (30.8% less scan data overall). This 30.8% reduction refers only to entropy-coded scan data and does not represent an overall reduction in the transferred 28 × 28 file size. If non-adaptive container files are sent, 700.8 + (344/1227)(844.8) ≈ 937.6 B must be transmitted, an 11.0% increase over fixed Q = 90 (844.8 B) because of duplicated headers (≈ 600 B). An out-of-band shared-table approach (e.g. RFC 2035 RTP transport) could reduce this. For native captures (600 × 450 × 3, HAM10000 source, 810 KB raw), fixed overhead is minimal (< 1–7%) relative to image entropy. Applying the representative mean model to complete native payloads (8.2, 16.5, 34.1 and

58.4 KB at Q = 30, 50, 70 and 90) gives 8.2 + (344/1227)(58.4) = 24.6 KB of adaptive cohort data (57.9% saved). The gating score $s(x)$ uses only the low-quality softmax outputs; ground truth is not used.

3.3 Wireless uplink channel profiles

We model four telecommunication channels: (1) legacy 2G/EDGE (B = 50 kbps, RTT = 300 ms); (2) constrained 3G (B = 384 kbps, RTT = 150 ms); (3) rural satellite IoT (B = 128 kbps, RTT = 600 ms); and (4) standard 4G/LTE (B = 2 Mbps, RTT = 40 ms).

4 Results and discussion

4.1 Rate-distortion and clinical screening evaluation

Table 1 and Figure 2 summarise the evaluation across static compression, heuristic transmission baselines and the proposed framework, reporting empirical diagnostic classification on the DermaMNIST-C benchmark (28 × 28, N = 1,227, N_{ref} = 149) alongside modelled native-resolution (600 × 450) communication payloads.

Table 1. Comparison of policies: 28 × 28 benchmark versus modelled native telemetry (N = 1,227, Nref = 149, seed 42).

Policy	Modelled payload (KB)*	Sens ($\mathcal{Y}_{ref}$)	Spec ($\mathcal{Y}_{rout}$)	PPV (Prec)	Bypassed SFR	MEL re-call	Macro F1
Fixed Q = 30	8.2	81.2%	98.0%	84.6%	18.8%	82.9%	0.817
Fixed Q = 50	16.5	87.9%	98.3%	87.9%	12.1%	91.4%	0.866
Fixed Q = 70	34.1	93.3%	98.7%	90.8%	6.7%	97.1%	0.907
Fixed Q = 90	58.4	96.6%	99.0%	92.9%	0.0%	98.6%	0.941
Random (p = 0.35)	29.8	83.9%	98.3%	87.4%	12.8%	84.3%	0.852
Entropy-only (α = 0.0)	25.2	94.0%	98.1%	87.5%	3.4%	95.7%	0.861
Risk-only (α = 1.0)	22.4	94.0%	98.1%	87.5%	3.4%	95.7%	0.860
Proposed (α = 0.65)	24.6	96.0%	99.0%	92.9%	0.7%	98.6%	0.874

Note: *Modelled native-resolution payload. Macro F1 evaluated before rounding.

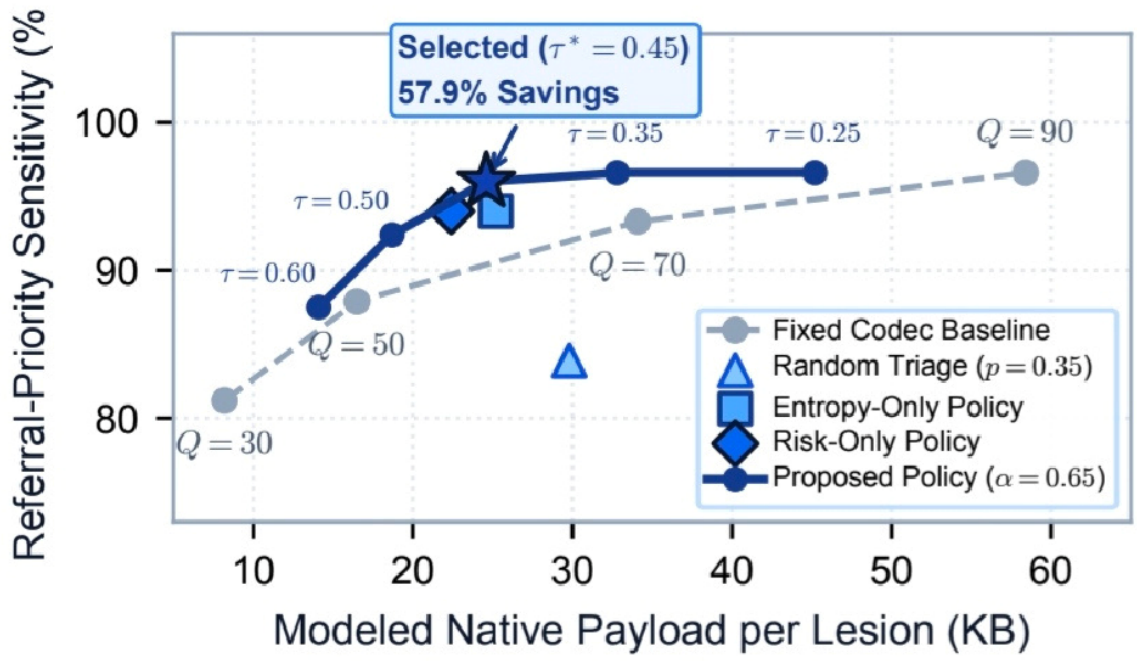


Figure 2. Referral-priority sensitivity (28 × 28 benchmark) versus modelled average transmitted payload (KB, native 600 × 450 resolution). The proposed policy (α = 0.65, τ = 0.45) attains 96.0% benchmark sensitivity at an estimated 24.6 KB per lesion.

Random triage (p = 0.35) yields 83.9% benchmark sensitivity. Entropy-only (τ_{ent} = 0.48) and risk-only (τ_{risk} = 0.35) policies reach 94.0% sensitivity at 25.2 KB and 22.4 KB modelled payload, respectively.

Our proposed composite policy (α = 0.65, τ* = 0.45) delivers 96.0% referral sensitivity (143/149) and a 0.874 Macro F1 score on the benchmark, while reducing 28 × 28 entropy-coded scan data by 30.8% (169.4 B vs. 244.8 B, excluding container overhead). In the separate native telemetry model, the policy yields an average payload of just 24.6 KB (an estimated 57.9% reduction relative to continuous Q = 90 transmission of 58.4 KB). The sensitivity difference is only 0.7 percentage points (144/149 vs. 143/149). Across 5 seeds (42–46), sensitivity is 96.0% ± 0.5 percentage points (TP counts: 143, 143, 144, 142, 143 out of 149; sample s = 0.47 pp, pop. σ = 0.42 pp) and payload 24.6 ± 0.3 KB vs. 96.6% ± 0.5 pp at Q = 90 (TP: 144, 144, 145, 143, 144; s = 0.47 pp, σ = 0.42 pp). On Seed 42, paired outcomes show 143 concordant referrals, 1 discordant case (b = 1, c = 0), and 5 concordant missed referrals: uncorrected McNemar χ^2 = 1.0 (p = 0.317, indicating lack of significant degradation rather than formal equivalence), continuity-corrected χ^2 = 0.0 (p = 1.0), exact binomial p = 1.0, and sensitivity difference −0.67 pp (paired Wald 95% CI: [−1.98, +0.64] pp; empirical tuning target ηtarget = 96.0%).

Adaptive retransmission ensures benign ambiguities are correctly classified so that the specificity of routine classification is 99.0%. Referral precision is improved to 92.9% (same as full fidelity Q = 90). This reduces unnecessary referrals by half to 11 cases. Continuous Q = 90 achieves a high F1 score for all classes (MEL: 0.939, recall: 98.6%; NV: 0.965; BKL: 0.769; BCC: 0.932, recall: 93.2%; AKIEC: 0.986, recall: 97.1%; VASC: 1.000; DF: 1.000), achieving an average of 0.941 (Macro F1) (mean before rounding 0.9413). Adaptive triage maintains the same high recall of MEL cases as in full fidelity (recall 98.6%, 69/70 cases; F1 score 0.939) and BCC cases (recall 93.2%) at the same level as Q = 90. AKIEC recall drops to 94.3% when using adaptive triage (33/35 cases, F1 score 0.971 versus 0.986) giving an average of 0.874 (Macro F1) (actual value: 0.8737). Therefore giving lower overall F1 scores

(MEL: 0.939, NV: 0.882, BKL: 0.393, BCC: 0.932, AKIEC: 0.971, VASC: 1.000, DF: 1.000). This decrease of 0.067 happens due to the amount of ambiguity between classes within routine classification. 72.0% fast route cases classified at Q = 30 have been incorrectly classified (NV as BKL). This occurs within Yroutine and does not affect referral.

4.2 Ablation analysis of the risk–entropy trade-off

The composite metric incorporates the benefits of both P_{referral} and H_{norm}. On Dval, Pure Entropy (α = 0.0, τ_{ent} = 0.48) is inefficient at identifying abnormal benign lesions, but fails to detect subtle malignancies (3.4% SFR). Pure Risk (α = 1.0, τ_{risk} = 0.35) focusses on malignancies but fails to detect compressed boundaries (3.4% SFR). Using α = 0.65 high risk and ambiguous cases are detected, reducing failed silent cases to 0.7%.

4.3 Analytical latency across telecommunication channels

Figure 3 plots analytical latency from Equation (10) across bandwidths under a 300 ms RTT, matching the 2G channel profile. Specifically, on a 2G/EDGE link (50 kbps, 300 ms RTT), continuous Q = 90 transmission requires 9.64 s per lesion, whereas our adaptive policy completes screening in 4.32 s (55.2% reduction). Over rural Satellite IoT (128 kbps, 600 ms RTT), latency drops from 4.25 s to 2.31 s (45.7% reduction), enabling responsive point-of-care tele-screening.

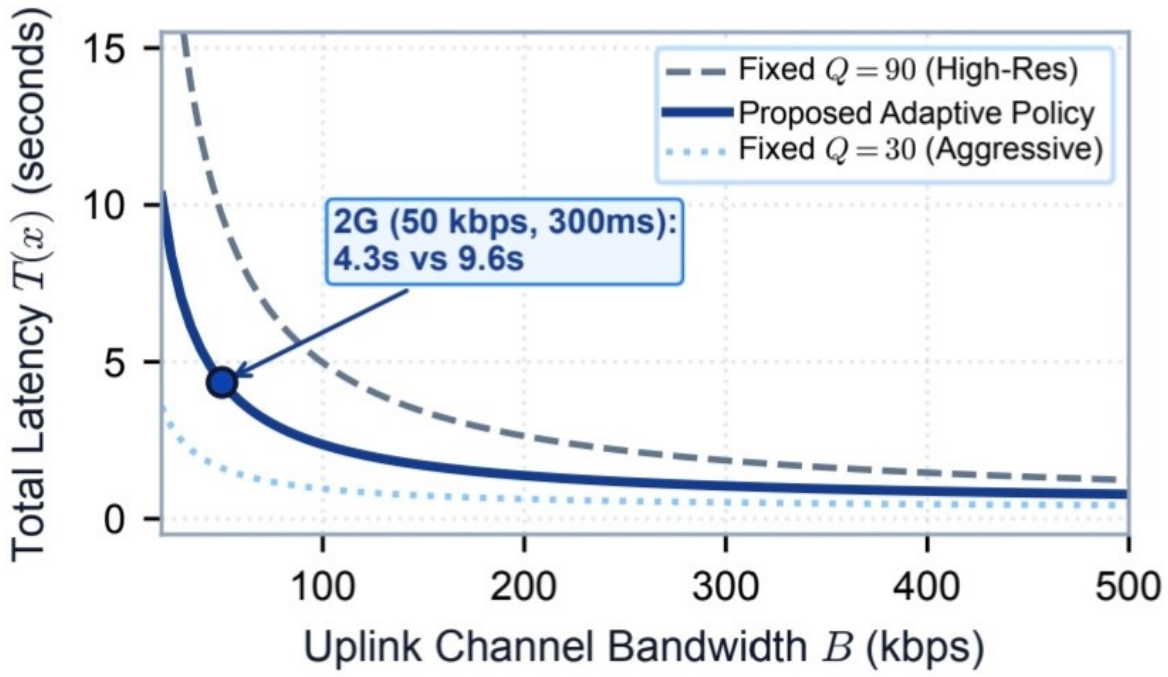


Figure 3. Analytical end-to-end transmission latency T(x) versus uplink bandwidth B at 300 ms RTT (2G/EDGE profile). Over low-bandwidth uplinks (20–100 kbps), adaptive transmission reduces latency by more than 55% compared with continuous high-resolution transmission.

4.4 Silent failure audit and uncertainty communication

Audit of Silent Failure Rate (SFR) is integral to safe tele-screening; the false-negative rate from referral can be separated into by-passed silent failures (SFR) and post-retransmission failures (FNRpost) using Equations (11) and (12). The risk-uncertainty gate provides only one observed bypassed failure (1/149 = 0.67% to one decimal place, 95% exact Clopper-Pearson Confidence

Interval = [0.02%, 3.68%]) as shown in Table 1. It results in only one silent failure (1/131 = 0.76%, 95% CI = [0.02%, 4.18%]) across $N_{\text{ref,lesion}}$ = 131 unique lesions. A total of 5 post-retransmission failures are also observed (Ppost = 5/131 = 3.82%) and 125 lesions were correctly triaged (T = 125/131 = 95.42%). The large Confidence Interval indicates the low number of lesions tested, so the residual risk cannot yet be considered clinically negligible.

Only one lesion, HAM_0001786 (ISIC_0027047), was bypassed, with a subtle AKIEC atypia which was resolved under Q = 90. This means there was only a small number of referred false negatives (3.4%, N=5/149). These failures were due to the classifier failing at baseline Q = 90 rather than the adaptive failure within transmission. The routing method should therefore be regarded as an aid to clinicians rather than a fully automated system.

5 Conclusion and future work

This paper investigated a risk-aware, two-stage adaptive transmission heuristic for low-bandwidth tele-dermatology. Evaluated on the 28 × 28 DermaMNIST-C benchmark (N = 1,227, 149 referral cases across 131 lesions), the policy achieves 96.0% referral sensitivity (within 0.7 percentage points of Q = 90, p = 0.317), 99.0% specificity, 98.6% melanoma recall, a 30.8% scan data reduction, and limits bypassed silent failures to one test case (0.7%, 95% CI: [0.02%, 3.68%]). In a separate telemetry model on native resolution captures, it yields an estimated 57.9% payload reduction (24.6 KB vs. 58.4 KB) and over 55% lower 2G latency. Because diagnostic performance on native captures remains untested, these modeled communication estimates demonstrate triage feasibility and failure auditing rather than native clinical efficacy. Future work will examine end-to-end models on clinical video feeds.

References

[1] Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*. 2018;5(1):1–.

[2] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–.

[3] dos Santos E, Cunha K, Mota R, Papais F, Bezerra T, Lopes N, et al. Bridging research and practice: A systematic evaluation of generalist and dermatology specific models in clinical skin lesion classification. *arXiv* preprint arXiv:2610.03193. 2026.

[4] Fan FX, Tan C, Ong YS, Wattenhofer R. Authority preserving evaluation of medical vision-language assistants. *arXiv* preprint arXiv:2609.22302. 2026.

[5] Shahriar MM, Mridha MF. LENS-GRF: Permutation-invariant lesion evidence network with gated residual fusion for acne severity grading and multi-rater clinical oracle analysis. *arXiv* preprint arXiv:2610.00294. 2026.

[6] Pal A, Patidar G, Santra B. FedDermaSeg: Federated learning for dermatological image segmentation. *arXiv* preprint arXiv:2610.08574. 2026.

[7] Barua A, Khan MMA, Iktidar A, Islam MS. Masked privileged-information distillation for multimodal skin lesion classification under missing clinical metadata. *arXiv* preprint arXiv:2610.03991. 2026.

[8] Alvarez D, Thorne S, Gupta R. Progressive image streaming for bandwidth-constrained edge healthcare AI. *IEEE Internet of Things Journal*. 2024;11(4):5890–5901.

[9] Bause O, Werner J, Bringmann O. CapsuleMotion: A lightweight real-time visual motion predictor for capsule endoscopy. *arXiv* preprint arXiv:2609.15367. 2026.

[10] Sirjani MS. Hardware-aware functional Kolmogorov–Arnold networks for efficient medical image enhancement and segmentation. *arXiv* preprint arXiv:2609.36134. 2026.

[11] Wallace GK. *The JPEG still picture compression standard*. IEEE Transactions on Consumer Electronics. 1992;38(1):xviii–xxxiv.

[12] Bai Y, Zhao Y, Wang K, Du Y, Cheng J, Fang T, et al. Practical lossless volumetric medical image compression via triplane context tree learning. *arXiv* preprint arXiv:2608.13897. 2026.

[13] Chou HH, Maly J, Rauhut H. Compressive sensing – introduction and relations to deep learning. *arXiv* preprint arXiv:2608.24211. 2026.

[14] Gorski N, Liang X, Guo H, Wang B. FeatureZ: A general framework for feature-preserving compression via pointwise bounds and star classification. *arXiv* preprint arXiv:2610.12371. 2026.

[15] Li Y, Bedros S, Bigliardi P, Qi MB, Morellas V, Papanikolopoulos N. MDSkin-Net: Multi-task skin lesion analysis driven by pattern analysis priors and spatial alignment regularization. *arXiv* preprint arXiv:2609.30855. 2026.

[16] Attia Y, Mukherjee D. CAMEO: A class-activation-mapped equitable overlay framework for fair and robust deep learning-based skin condition diagnosis. *arXiv* preprint arXiv:2609.36400. 2026.

[17] Yang J, Shi R, Wei D, Liu Z, Zhao L, Ke B, et al. MedMNIST v2: A large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Scientific Data*. 2023;10(1):41.

[18] Abhishek K, Jain A, Hamarneh G. Investigating the quality of DermaMNIST and Fitzpatrick17k dermatological image datasets. *Scientific Data*. 2025;12:1–14.

[19] Vu TV, Hai NC, Nguyen NP, Tu NH, Bao VNQ, Anh NT. Less is more: A leakage-controlled study of dermoscopic preprocessing for joint skin lesion classification and segmentation with YOLO26. *arXiv* preprint arXiv:2610.08570. 2026.

[20] Shannon CE. A mathematical theory of communication. *The Bell System Technical Journal*. 1948;27(3):379–423.

[21] Zhang J, Li Y, Han L, Anaissi A, Tran NH. Rethinking the information bottleneck: Structured decomposition under label-induced partitions. *arXiv* preprint arXiv:2610.01175. 2026.

[22] Liu D, Zhu G, Zhang J, Huang K. Wireless data acquisition for edge learning: Data-importance aware retransmission. *IEEE Transactions on Wireless Communications*. 2020;19(1):406–420.

[23] Geifman Y, El-Yaniv R. Selective classification for deep neural networks. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2017. p. 4878–4887.

[24] Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. In: *International Conference on Machine Learning (ICML)*. PMLR. 2017. p. 1321–1330.

[25] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. p. 770–778.

[26] Rangnekar A, Tapias Gomez J, Deasy JO, Veeraraghavan H. Parameter-efficient pretrained-CT-to-MRI transfer for rectal cancer segmentation: Performance-calibration trade-offs. *arXiv* preprint arXiv:2608.27178. 2026.