

Task-Oriented Semantic Communication for Digital Pathology over Wireless Channels

Domenico Benfenati*, Giovanni Maria De Filippis*, Antonio Maria Rinaldi*,
Cristiano Russo†, and Cristian Tommasino*

*Department of Electrical Engineering and Information Technology (DIETI),
University of Naples Federico II, Via Claudio 21, 80125 Naples, Italy

†Department of Humanities, University of Naples Federico II, Naples, Italy

Email: {domenico.benfenati, giovannimaria.defilippis, antoniomaria.rinaldi, cristiano.russo, cristian.tommasino}@unina.it

Abstract—Streaming histopathology tiles from edge scanners to cloud-based classifiers faces stringent bandwidth and latency constraints. Since diagnostic inference requires only the predicted clinical outcome, faithful image reconstruction is unnecessary. This work investigates task-oriented semantic communication for digital pathology, focusing on where semantic compression should be performed and how reliable inference can be maintained over noisy channels. We evaluate the proposed approach across five frozen backbones and three public pathology datasets, BRACS, NCT-CRC, and LC25000, covering both general-purpose and pathology-specific representations. Compressing learned feature representations rather than raw image tiles reaches the linear-probe accuracy ceiling using only 32 bytes per tile. This yields a bitrate reduction of approximately three orders of magnitude compared with JPEG and JPEG2000, and nearly $300\times$ compared with modern learned image codecs and HEVC-Intra at matched downstream accuracy, with only a negligible gap from the upper bound. To enhance robustness, we formulate the pipeline as a deep joint source-channel coding (JSCC) system, jointly optimizing feature encoding, channel adaptation, and classification end-to-end. The proposed method retains at least 95% of ceiling accuracy down to -5 dB SNR, while conventional separated digital transmission degrades sharply below its operating threshold. Ablation studies show that rate savings stem from feature-space compression, whereas channel robustness results from end-to-end joint optimization. These findings highlight semantic communication as a promising framework for efficient and reliable digital pathology streaming.

Index Terms—semantic communication, joint source-channel coding, digital pathology, histopathology classification, edge-cloud, deep learning, mobile health, explainable AI.

I. INTRODUCTION

Digital pathology is steadily moving away from the centralized laboratory. Portable and benchtop whole-slide scanners, telepathology [1], and AI-assisted triage increasingly operate at the *edge*, offloading heavy analysis to the cloud over mobile and wireless links (5G, Wi-Fi 6) [2]. This shift is driven as much by necessity as by technology: a global shortage of pathologists and the uneven distribution of expertise make timely diagnosis increasingly dependent on moving images, rather than patients or specialists, across a network. The

wireless link therefore acts as a clinical enabler rather than an implementation detail, and its bandwidth, energy, and latency budgets directly shape what mobile health can deliver at the point of care.

This raises a communication bottleneck. A whole-slide image (WSI) is gigapixel-scale and yields thousands of high-resolution tiles, so transmitting raw or generically compressed imagery wastes bandwidth and energy and inflates end-to-end latency. Yet for a given diagnostic task, the receiver does not need a pixel-faithful reconstruction, only the *semantics* sufficient to determine the label. This is the premise of task-oriented (goal-oriented) semantic communication [2], [3].

The proposed pipeline is generic by design: it can be adapted to any tile-level classification task by modifying only the tile collection and label taxonomy. We evaluate its generality across three public datasets spanning diverse tissues and classification granularities: *BRACS* [4] for breast-lesion subtyping (7 classes; Fig. 1), a challenging clinical setting characterized by highly similar atypical lesions; *NCT-CRC* [5] for colorectal tissue classification (9 classes); and *LC25000* [6] for lung and colon pathology classification (5 classes). Each dataset is evaluated using five frozen feature backbones.

Instead of transmitting raw image tiles, the edge device extracts task-relevant representations and maps them into channel symbols through a deep joint source-channel coding (JSCC) framework [7]. Deep feature representations of histopathology tiles are established task descriptors, having proven effective for downstream problems such as content-based image retrieval in digital pathology archives [8], [9], which motivates their use as the semantic payload transmitted over the channel. By incorporating the channel model within the computational graph, gradients can propagate through the communication module during training, enabling the cloud-side classifier to operate directly on noisy received representations.

We structure our analysis around two fundamental design choices in semantic communication: *where* compression should be applied (pixel space versus feature space) and *how* robustness to channel impairments should be achieved (separate source-channel coding versus joint optimization). Through carefully designed control experiments, we isolate the contribution of each component and quantify its impact

This study was supported by PON CREST - Computing Research infrastructure for Engineering, Science and Technology, CUP: E63C26000360007.

Corresponding author: C. Tommasino (email: cristian.tommasino@unina.it).

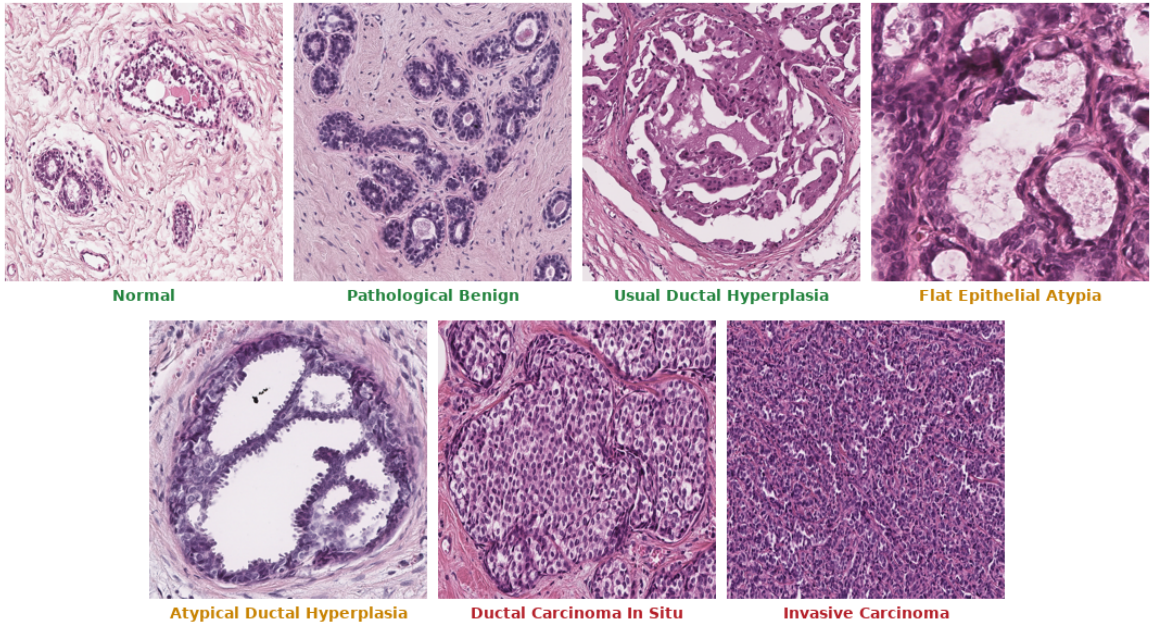


Fig. 1. The seven BRACS subtypes (one example RoI each), color-coded by group: benign (green: Normal, Pathological Benign, Usual Ductal Hyperplasia), atypical (orange: Flat Epithelial Atypia, Atypical Ductal Hyperplasia), and malignant (red: Ductal Carcinoma In Situ, Invasive Carcinoma). Subtypes differ by subtle nuclear/architectural cues, making fine-grained subtyping difficult and sensitive to lossy transmission.

on rate efficiency and reliability.

We formulate digital pathology streaming as a task-oriented semantic communication problem, where diagnostic inference is performed directly from compressed and channel-distorted representations rather than reconstructed images. We develop an end-to-end JSCC framework that jointly optimizes semantic representation extraction, communication, and classification, enabling reliable inference under bandwidth and channel constraints. Finally, we establish a comprehensive experimental framework for evaluating semantic compression in histopathology, systematically comparing pixel-space and feature-space transmission strategies across multiple datasets, backbone architectures, and communication baselines.

II. SYSTEM MODEL

We cast the diagnostic offload as a three-stage edge–channel–cloud pipeline (Fig. 2). The edge converts a tile into a compact set of channel symbols through a frozen backbone g_θ and a semantic encoder E_ϕ ; a differentiable simulated channel corrupts these symbols; and the cloud recovers a task representation with a decoder D_ψ and emits the label with a classifier C_ω . The defining choice is that the channel remains inside the computational graph, allowing the encoder and decoder to be optimized jointly for the channel encountered at deployment rather than for pixel fidelity. The three subsections that follow detail the edge, the channel, and the cloud in turn, and finally the two mechanisms that set the operating point.

A. Edge–Channel–Cloud Pipeline

The architecture is shown in Fig. 2. Let t be a tile and $f = g_\theta(t) \in \mathbb{R}^D$ its embedding extracted by a frozen backbone

g_θ (ResNet-50 [10] or a pathology foundation model [11]). The edge applies a semantic encoder $E_\phi : \mathbb{R}^D \rightarrow \mathbb{C}^k$ to produce k complex channel symbols $x = E_\phi(f)$ under the average-power constraint $\frac{1}{k} \mathbb{E}|x|^2 \leq 1$. The symbol count k sets the number of channel uses per tile, and hence the bandwidth (or, at a fixed symbol rate, the transmission time); the power constraint ensures that link quality depends only on the signal-to-noise ratio (SNR). In Section IV, we express k as an equivalent byte budget for comparison with digital baselines.

B. Differentiable Channel

The channel is simulated and differentiable, so any of the following models can sit inside the training graph and let the diagnostic loss back-propagate through it (Section III). For AWGN, $y = x + n$, with $n \sim \mathcal{CN}(0, \sigma^2 I)$ and $\sigma^2 = 10^{-\text{SNR}/10}$; we also model Rayleigh block fading $y = h \odot x + n$ with perfect-CSI equalization, and packet erasure that zeroes symbol blocks with probability p . Together, these span the impairments a hospital edge link is likely to face, from a clean, well-provisioned link (AWGN), to indoor multipath (Rayleigh), to the bursty losses of a congested or mobile link (erasure). We sample the SNR per batch over a configurable range during training, so a single model learns to operate across the whole range it will meet in deployment rather than being tuned to one nominal condition.

C. Cloud Decoder and Classifier

A decoder $D_\psi : \mathbb{C}^k \rightarrow \mathbb{R}^D$ estimates $\hat{f} = D_\psi(y)$ and a classifier C_ω outputs the 7-class posterior $\hat{p} = \text{softmax}(C_\omega(\hat{f}))$. With SNR-adaptive coding, the (estimated) SNR is concatenated to the encoder and decoder inputs.

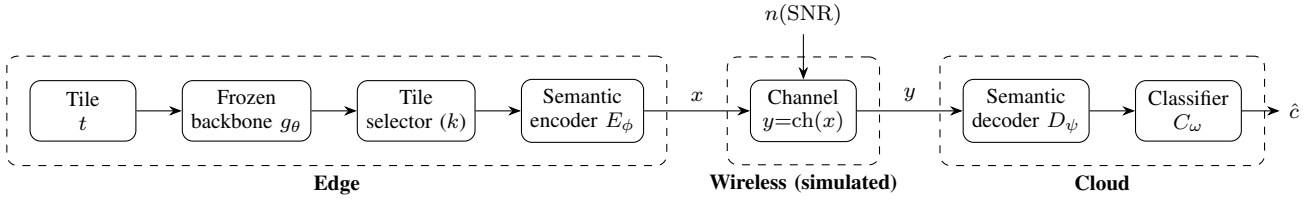


Fig. 2. Task-oriented semantic communication pipeline. The edge extracts tile features with a frozen backbone, optionally keeps the top- k tiles, and encodes them into k complex channel symbols x . A differentiable simulated channel (AWGN/Rayleigh/erasure) yields y ; the cloud decoder and classifier output the class label $\hat{c}$. The whole chain is trained end-to-end with the loss in (1), so gradients flow through the channel.

D. Compression Mechanisms

(i) The *symbol dimension* k sets the bandwidth ratio (channel uses per tile); (ii) *SNR-adaptive* coding shares one model across the SNR range [12].

III. METHOD

Our method comprises two components. We first define a single end-to-end training objective that optimizes the entire pipeline by evaluating the classification loss *after* the channel, ensuring that the encoder, channel model, decoder, and classifier are jointly optimized for the diagnostic task rather than image reconstruction. We then introduce a set of controlled baselines, each removing one key component of the full system (learned coding, joint optimization, or feature-space transmission), which lets us isolate the contribution of each design choice. We describe the training objective first and the baselines second.

A. Task-Oriented Training

The whole chain is trained end-to-end by minimizing a classification loss evaluated *after* the channel:

$$\begin{aligned} \mathcal{L} &= \mathbb{E}_{(t,c) \sim \mathcal{D}} \mathbb{E}_{\text{SNR} \sim \mathcal{U}} \ell(\hat{p}, c), \\ \hat{p} &= C_\omega(D_\psi(\text{ch}(E_\phi(g_\theta(t))))), \end{aligned} \quad (1)$$

where each factor plays a distinct role. The pair (t, c) is a labeled training tile drawn from the dataset $\mathcal{D}$, and the outer expectation $\mathbb{E}_{(t,c) \sim \mathcal{D}}$ averages the loss over the training set. The inner expectation $\mathbb{E}_{\text{SNR} \sim \mathcal{U}}$ averages over channel conditions: the SNR is resampled per batch from a uniform range $\mathcal{U}$ (Section II), so a single set of weights is optimized for the entire operating range rather than for one nominal SNR. The argument of ℓ is the full transmit–receive chain, read from the inside out: the frozen backbone g_θ maps the tile to the feature $f = g_\theta(t)$; the semantic encoder E_ϕ turns f into the k channel symbols $x = E_\phi(f)$; the differentiable channel $\text{ch}(\cdot)$ (AWGN, Rayleigh, or erasure) corrupts them into $y = \text{ch}(x)$; the decoder D_ψ reconstructs a task feature $\hat{f} = D_\psi(y)$; and the classifier C_ω produces the posterior $\hat{p} = \text{softmax}(C_\omega(\hat{f}))$ over the classes. The per-sample term $\ell(\hat{p}, c)$ is the focal loss [13] with inverse-frequency class weights, chosen over plain cross-entropy to counter class imbalance by down-weighting easy, well-classified tiles. Because the channel lies inside the computational graph, gradients of $\mathcal{L}$ propagate back through C_ω , D_ψ , $\text{ch}(\cdot)$, and E_ϕ , so the encoder and decoder are jointly

optimized for the channel encountered at deployment; only θ is held fixed. Keeping the backbone frozen also means training operates on cached features, making a sweep over the symbol budget k inexpensive.

B. Baselines as Controls

Each baseline is constructed to remove exactly one ingredient of the full system, so that the resulting accuracy gap isolates the contribution of that ingredient. We group them by the question they answer.

Reference upper bound. (a) The frozen-feature *ceiling* is a linear/MLP probe on the uncompressed backbone features. It is the accuracy attainable with no rate constraint and no channel, and it serves as the reference against which every compressed or noisy system is measured. We emphasize that it is a frozen-feature reference, not a true information-theoretic upper bound.

Where to compress: pixels vs. features. (b) *Image codecs*, namely JPEG [14], JPEG2000 [15], BPG [16], and a learned CompressAI model [17], [18] evaluated at multiple quality levels, compress the transmitted-resolution *tile* and thereby represent the conventional pixel-domain approach. Including modern codecs verifies whether the rate advantage persists against state-of-the-art image compression. (c) *Non-neural feature compression* applies PCA and product vector quantization to the same frozen features at matched byte budgets. Compared with image codecs, it isolates the benefit of feature-space compression; compared with our learned encoder, it quantifies the value of learned semantic coding over fixed feature transformations.

How to survive the channel: separate vs. joint. (d) A *separate neural autoencoder* is trained for reconstruction without channel noise, and its latent representation is transmitted over the analog channel at test time. Compared with our method, it isolates the effect of joint end-to-end optimization. (e) A *digital baseline with capacity-achieving channel coding* provides error-free transmission above the Shannon threshold and fails below it, representing the strongest possible separate source–channel coding system. This baseline highlights the characteristic digital cliff that our joint analog approach is designed to avoid.

IV. EXPERIMENTAL SETUP

Having defined the pipeline (Section II) and its training objective (Section III), we now describe the protocol used to eval-

uate them: the datasets, the frozen backbones and classification head, and the metrics that jointly quantify communication rate and diagnostic accuracy. The evaluation follows a uniform protocol across all configurations. Each of the 15 backbone–dataset combinations is assessed under identical conditions, namely the same data splits, preprocessing, training schedule, rate budgets, and channel models, so that any observed difference is attributable to the choice of backbone or dataset rather than to the evaluation procedure. Only the label taxonomy and the input tiles differ across configurations.

Datasets. Three public histopathology tile-classification sets, used identically: BRACS RoI [4] (7 breast subtypes, official patient-disjoint split 3657/312/570), NCT-CRC [5] (9-class colorectal, official train/test), and LC25000 [6] (5-class lung/colon). Only the label taxonomy and tiles change; all are resized to 224×224 (antialiased) with a patient/slide-disjoint split where available.

Backbones and head. We evaluate five frozen feature extractors: ResNet-50 pretrained on ImageNet and four pathology foundation models, UNI [11], Virchow2 [19], H-optimus-0 [20], and Prov-GigaPath [21]. A linear/MLP classification head is trained on top of each frozen backbone, yielding the 15 backbone–dataset configurations evaluated in this study.

Metrics and rate. Our primary metric is macro-F1 over the 7 classes (mean $\pm$ std over five seeds); we also report 3-group accuracy, balanced accuracy, per-class F1, and confusion matrices. The transmission rate of the analog JSCC link is measured by the number of channel uses (k complex symbols). For comparison with digital codecs, we also report a byte-equivalent assuming $b = 16$ bits per real symbol, corresponding to $2kb/8$ bytes, while noting that this is only an equivalent communication budget rather than a file size. Finally, we report the per-tile latency, decomposed into encoding, transmission, and inference. As the rate decreases, the transmission time shrinks while encoding and inference remain constant, making the system compute-bound. Accordingly, JSCC trades additional edge computation for reduced bandwidth, unlike conventional image codecs.

V. RESULTS

All 15 backbone–dataset configurations, comprising five frozen backbones (ResNet-50 and the pathology foundation models UNI [11], Virchow2, H-optimus-0, and Prov-GigaPath) evaluated on three datasets (BRACS, NCT-CRC [5], LC25000 [6]), follow the identical evaluation protocol of Section IV, with only the encoder and the input tiles varying between them. An ImageNet-pretrained ViT-B baseline is excluded, as its frozen features proved near-random on all three datasets. Unless stated otherwise, the reported values are averaged over the 15 configurations, and the BRACS $\times$ UNI configuration is additionally repeated over five random seeds to provide error bars. The results are organized around three findings: the magnitude of the rate reduction (Finding 1), its underlying source (Finding 2), and the distinct contribution of end-to-end joint training (Finding 3).

A. Finding 1: Feature-space coding attains the accuracy ceiling at a $\sim 10^3 \times$ lower rate

At only 32 bytes/tile, JSCC attains the frozen-feature ceiling to within a fraction of a macro-F1 point in every configuration (Table I, Fig. 3). Repeating each configuration over five seeds confirms that this residual gap, although statistically significant (paired t -test $p=0.007$), is negligible in magnitude: it remains well below one macro-F1 point and closes almost entirely when the budget is raised only slightly (for instance, from 32 to 128 bytes on BRACS $\times$ UNI). The result is therefore a *near-match at three orders of magnitude lower rate*, rather than an improvement in accuracy.

Table II quantifies this gap on the rate axis: to reach the same accuracy, JPEG and JPEG2000 require tens of kilobytes on every dataset. A *modern* pixel codec (BPG or a learned CompressAI model) narrows but does not eliminate the gap, matching the accuracy of JSCC only in the low tens of kilobytes, still roughly $300\times$ the JSCC budget, and, on BRACS, never reaching the ceiling at any tested rate. Fig. 3 corroborates this behavior on the log-rate axis: the JSCC curve is already flat against the ceiling at its leftmost point, whereas the JPEG and JPEG2000 curves reach that ceiling only after three additional decades of rate. This horizontal separation *at matched accuracy*, rather than any difference in peak accuracy, constitutes the principal result, and it holds consistently across all three datasets.

A second, consistent effect is that *pathology foundation models systematically outperform the ImageNet CNN*, most notably on BRACS and NCT-CRC (Table I). LC25000, by contrast, is essentially saturated for every backbone and serves primarily as a sanity check; harder slide-level benchmarks such as CAMELYON [22] are left to future work.

B. Finding 2: The rate reduction is primarily attributable to the feature space, not to the learned network

We next examine whether the gain stems from the learned code itself or from the feature space in which it operates. For the most part, it stems from the latter: a non-neural PCA/product-VQ of the *same* frozen features, at the same 32-byte budget, already surpasses JPEG and JPEG2000 by the full three orders of magnitude, and modern codecs by the full $\sim 300\times$ (Table II), confirming that the saving originates from compressing features rather than pixels. The learned code is nonetheless not redundant: at the tightest budget it retains a small but consistent margin over PCA/VQ, which narrows as the budget increases until the methods converge within a few hundred bytes. In the *noiseless* regime, the efficiency gain is thus a property of the feature space itself, and the frozen-feature “ceiling” is best interpreted as a reference rather than a true upper bound. The *decisive* advantage of the learned code emerges instead once the channel becomes non-ideal, precisely the regime in which joint training is consequential.

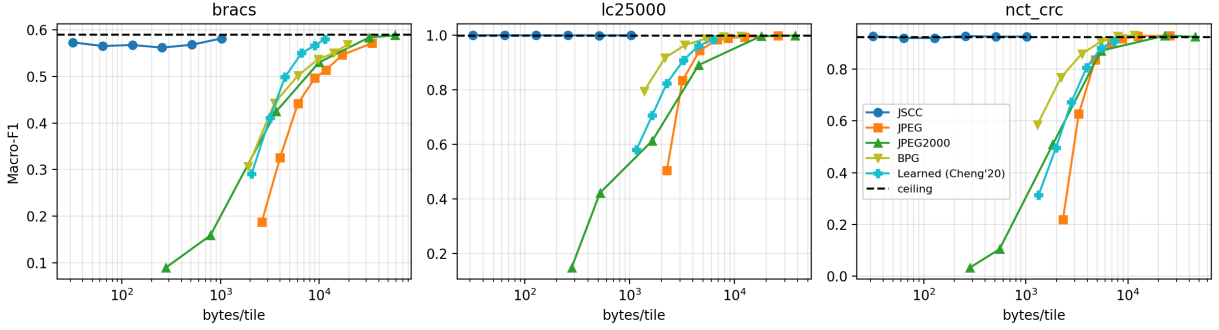


Fig. 3. Rate-accuracy per dataset (log-rate), averaged over the five backbones. JSCC reaches the (per-dataset) ceiling at 32 bytes/tile, while JPEG/JPEG2000 need tens of kilobytes (the same $\sim 10^3\times$ gap on all three datasets) and even the strongest pixel codec (BPG or the learned Cheng’20 model) still needs ≈ 10 kB, two to three decades to the right.

TABLE I

ACROSS BACKBONES (COLUMNS) AND DATASETS (ROWS): FROZEN-FEATURE CEILING MACRO-F1, WITH **JSCC AT 32 BYTES/TILE** IN PARENTHESES. PATHOLOGY FOUNDATION MODELS DOMINATE THE IMAGENET CNN, AND JSCC CLOSELY TRACKS THE CEILING IN EVERY CELL AT THREE ORDERS OF MAGNITUDE LOWER RATE. SINGLE SEED PER CELL; ERROR BARS OVER 5 SEEDS (PER CELL, AND FOR BRACS $\times$ UNI) ARE IN THE TEXT.

	ResNet-50	UNI	Virchow2	H-optimus	GigaPath
BRACS (7)	.51 (.51)	.60 (.61)	.61 (.58)	.61 (.57)	.62 (.59)
NCT-CRC (9)	.88 (.87)	.93 (.94)	.94 (.95)	.94 (.94)	.94 (.94)
LC25000 (5)	.99 (1.0)	1.0 (1.0)	1.0 (1.0)	1.0 (1.0)	1.0 (1.0)

C. Finding 3: End-to-end joint training eliminates the channel cliff

This subsection addresses the question raised by Finding 2: what does joint training contribute once the channel is non-ideal? Fig. 4 reports the fraction of each configuration’s ceiling F1 retained as the channel degrades, averaged over all configurations and normalized by the per-configuration ceiling so that datasets of markedly different difficulty remain comparable. As the channel quality decreases (from left to right in the figure), the informative feature is the *shape* of each curve, a gradual roll-off as opposed to an abrupt drop, rather than its absolute value at any single SNR.

JSCC degrades gracefully across essentially the entire tested SNR range and exhibits comparable resilience under packet loss. By contrast, the digital baseline, a *capacity-achieving* code (genie FEC) at spectral efficiency $\eta=2$ bits/symbol that represents the strongest attainable digital system rather than an uncoded link, exhibits a sharp *cliff* in every configuration: it matches the ceiling above the Shannon waterfall (SNR $\geq 10 \log_{10}(2^\eta - 1) \approx 4.8$ dB), but its accuracy collapses to chance once the link enters outage below that threshold. The channel-agnostic separate autoencoder degrades earlier still, well before JSCC incurs any comparable loss. Only the jointly trained code avoids the cliff, and this robustness, rather than the noiseless rate, constitutes the distinctive benefit of JSCC.

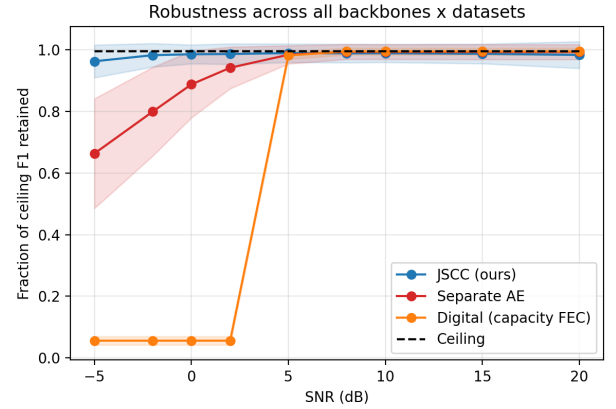


Fig. 4. Robustness across all 15 backbone $\times$ dataset configurations: fraction of the per-configuration ceiling F1 retained vs. SNR (mean; band = std across configurations). JSCC degrades gracefully; the capacity-achieving digital baseline (genie FEC) and the analog separate AE do not.

D. Bandwidth, Latency, and Ablation

The rate gap translates directly into bandwidth and energy savings, though not necessarily into wall-clock time: transmission collapses from seconds to milliseconds per slide, yet the overall per-tile latency becomes *compute-bound*, dominated by encoding and inference rather than by transmission. JSCC additionally relocates the backbone to the edge, trading edge computation for bandwidth; lightweight and fast pathology models suitable for such on-device inference are increasingly available [23], [24]. Finally, SNR-adaptive coding yields no consistent gain over a single fixed model across the SNR range, which indicates that one model already suffices to cover the operating range required by this task.

VI. DISCUSSION AND CONCLUSION

We asked *where* to compress and *how* to survive the channel for task-oriented histopathology transmission, and answered both through controlled experiments spanning all 15 backbone $\times$ dataset configurations. *Where*: coding in feature space reaches the frozen-feature ceiling at ~ 32 bytes/tile, three

TABLE II

PER-DATASET METHOD COMPARISON (MACRO-F1, AVERAGED OVER THE FIVE BACKBONES). JSCC AND FEATURE COMPRESSION AT 32 BYTES/TILE; IMAGE CODECS AT THEIR BEST QUALITY WITH THE BYTES REQUIRED.

Dataset	Ceiling	JSCC@32B	PCA@32B	VQ@32B	JPEG (F1@kB)	JPEG2k (F1@kB)	Modern (F1@kB)
BRACS (7)	0.590	0.573	0.535	0.560	0.572@31k	0.588@59k	0.580@11k
NCT-CRC (9)	0.925	0.928	0.907	0.911	0.931@20k	0.930@23k	0.933@10k
LC25000 (5)	0.998	0.999	0.989	0.996	0.997@26k	0.998@26k	0.997@11k

orders of magnitude below image codecs, although a non-neural PCA/VQ control shows this to be a property of feature space itself rather than of the learned network. *How*: only end-to-end joint training removes the digital cliff, retaining most of the ceiling accuracy down to -5 dB, where a capacity-achieving digital link and a channel-agnostic autoencoder both collapse; this robustness, which holds in every configuration, is JSCC’s distinctive contribution. Domain-pretrained pathology backbones raise the achievable accuracy further still. The “ceiling” is a frozen-feature reference rather than a true upper bound, the gigapixel-WSI motivation is addressed at the tile/RoI level, and the channel is simulated. Avenues for future work include whole-slide MIL bags, an information-bottleneck rate penalty, learned channel-aware separate coding, a quantitative study of saliency- and calibration-preservation under channel noise, and over-the-air validation.

REFERENCES

- [1] S. Zia, I. Z. Yildiz-Aktas, F. Zia, and A. V. Parwani, “An update on applications of digital pathology: Primary diagnosis, telepathology, education and research,” *Diagnostic Pathology*, vol. 20, no. 1, p. 17, 2025.
- [2] D. Gündüz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, “Beyond transmitting bits: Context, semantics, and task-oriented communications,” *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 5–41, 2023.
- [3] E. C. Strinati and S. Barbarossa, “6g networks: Beyond shannon towards semantic and goal-oriented communications,” *Computer Networks*, vol. 190, p. 107930, 2021.
- [4] N. Brancati, A. M. Anniciello, P. Pati, D. Riccio, G. Scognamiglio, G. Jaume, G. De Pietro, M. Di Bonito, A. Foncubierta, G. Botti, M. Gabrani, F. Feroce, and M. Frucci, “BRACS: A dataset for BReAst carcinoma subtyping in H&E histology images,” *Database*, vol. 2022, p. baac093, 2022.
- [5] J. N. Kather, J. Krisam, P. Charoentong *et al.*, “Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study,” *PLOS Medicine*, vol. 16, no. 1, p. e1002730, 2019.
- [6] A. A. Borkowski, M. M. Bui, L. B. Thomas, C. P. Wilson, L. A. DeLand, and S. M. Mastorides, “Lung and colon cancer histopathological image dataset (LC25000),” *arXiv preprint arXiv:1912.12142*, 2019.
- [7] E. Boursoulatz, D. B. Kurka, and D. Gündüz, “Deep joint source-channel coding for wireless image transmission,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567–579, 2019.
- [8] C. Tommasino, F. Merolla, C. Russo, S. Staibano, and A. M. Rinaldi, “Histopathological image deep feature representation for CBIR in smart PACS,” *Journal of Digital Imaging*, vol. 36, no. 5, pp. 2194–2209, 2023.
- [9] C. Tommasino, C. Russo, and A. M. Rinaldi, “Effects of color stain normalization in histopathology image retrieval using deep learning,” in *IEEE International Symposium on Multimedia (ISM)*, 2022.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [11] R. J. Chen, T. Ding, M. Y. Lu *et al.*, “Towards a general-purpose foundation model for computational pathology,” *Nature Medicine*, vol. 30, pp. 850–862, 2024.
- [12] M. Ding, J. Li, M. Ma, and X. Fan, “SNR-adaptive deep joint source-channel coding for wireless image transmission,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 1555–1559.
- [13] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [14] G. K. Wallace, “The JPEG still picture compression standard,” *Communications of the ACM*, vol. 34, no. 4, pp. 30–44, 1991.
- [15] A. Skodras, C. Christopoulos, and T. Ebrahimi, “The JPEG2000 still image compression standard,” *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 36–58, 2001.
- [16] F. Bellard, “BPG image format,” 2018, libbpg 0.9.8, HEVC-intra still-image codec.
- [17] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, “Learned image compression with discretized Gaussian mixture likelihoods and attention modules,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 7939–7948.
- [18] J. Bégaïnt, F. Racapé, S. Feltman, and A. Pushparaja, “CompressAI: A PyTorch library and evaluation platform for end-to-end compression research,” *arXiv preprint arXiv:2011.03029*, 2020.
- [19] E. Zimmermann, E. Vorontsov, J. Viret, A. Casson, M. Zelechowski, G. Shaikovski, N. Tenenholtz, J. Hall, D. Klimstra, R. Yousfi, T. Fuchs, N. Fusi, S. Liu, and K. Severson, “Virchow2: Scaling self-supervised mixed magnification models in pathology,” *arXiv preprint arXiv:2408.00738*, 2024.
- [20] C. Saillard, R. Jenatton, F. Llinas-López, Z. Mariet, D. Cahané, E. Durand, and J.-P. Vert, “H-optimus-0,” 2024, open-source histopathology foundation model, Biopitimus.
- [21] H. Xu, N. Usuyama, J. Bagga, S. Zhang, R. Rao, T. Naumann, C. Wong *et al.*, “A whole-slide foundation model for digital pathology from real-world data,” *Nature*, vol. 630, no. 8015, pp. 181–188, 2024.
- [22] B. E. Bejnordi, M. Veta, P. J. van Diest *et al.*, “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,” vol. 318, no. 22, 2017, pp. 2199–2210.
- [23] C. Tommasino, C. Russo, and A. M. Rinaldi, “NuLite: Lightweight and fast model for nuclei instance segmentation and classification,” *Biomedical Signal Processing and Control*, 2026, arXiv:2408.01797.
- [24] C. Tommasino, C. Russo, A. M. Rinaldi, and F. Ciompi, “HoVer-UNET: Accelerating HoVerNet with UNet-based multi-class nuclei segmentation via knowledge distillation,” in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2024.