HistoDyn::BioGater, Biologically Interpretable Dynamic Gating Architecture for Computational Pathology

Diogen Babuc*

Faculty of Informatics, West University of Timişoara, Bd. V. Pârvan no. 4, Timişoara 300223, Timiş, Romania

Abstract

Large-scale analysis of histopathological images for disease identification and characterization is made possible by computational pathology, which is revolutionizing clinical diagnostics. Nevertheless, current deep learning models frequently have poor interpretability and large computational costs. Therefore, this work presents HistoDyn::BioGater, a dynamically sparse vision architecture for high-performance histopathology that is physiologically interpretable. BioGater substitutes a lightweight combination of physiologically grounded operations, such as edge polarity, contrast amplification, and stain normalization, regulated by top-k learning gating, for dense convolutional blocks. BioGater, which was trained on colorectal histology, performs better than RandomOp Mixer and ConvMixer in terms of accuracy, calibration, and energy efficiency. Sparsity and specialization are measured by a new interpretability metric called *Gated Activation Diversity Score*. BioGater offers a single framework for effective, comprehensible, and scalable computational pathology. Our findings confirm that dynamic conditional execution and physiologically relevant transformations are useful tools for dependable clinical AI implementation.

Keywords

Computational pathology, dynamic gating, lightweight vision models, interpretability.

1. Introduction

Advances in computational pathology are revolutionizing the analysis of histological slides in cancer and precision medicine [1]. Histopathological examination of tissue samples is often used to identify colorectal cancer, one of the most prevalent malignancies worldwide. Recent digital pathology pipelines record these samples as gigapixel-scale whole-slide images (WSIs), producing petabyte-level data in hospitals and research facilities [2]. Effective processing of such data requires scalable, interpretable calculations that can run on distributed GPU and HPC clusters, as well as accurate image recognition models.

Although deep convolutional and transformer-based vision architectures have demonstrated exceptional accuracy in biological imaging, their decision-making procedures are still computationally intensive and frequently opaque [3]. Deployment in large-scale pathology processes is hampered by their huge memory footprints and communication overheads due to their rich parameterization and global attention activities. Furthermore, the majority of current models function as "black boxes", learning abstract representations that have nothing to do with biological processes like gland shape, tissue contrast enhancement, or stain normalization. When it comes to clinical adoption, where interpretability and repeatability are just as important as raw performance, the lack of transparency presents significant obstacles.

To address these challenges, we propose *HistoDyn::BioGater*, a lightweight, dynamically sparse, and biologically interpretable vision architecture for high-performance computational pathology. HistoDyn is built upon the BioGater module, a multi-head dynamic gating mechanism that replaces heavy convolutional or attention blocks with collections of inexpensive tensor operations, referred to as biological transforms. Each transform performs a simple yet biologically meaningful computation, such

EXPLIMED 2026 - Third Workshop on Explainable Artificial Intelligence for the medical domain - 15-17 August 2026, Bremen, Germany

*Corresponding author.

diogen.babuc00@e-uvt.ro (D. Babuc)

https://sites.google.com/e-uvt.ro/babucdiogen (D. Babuc)

0009-0000-5126-6480 (D. Babuc)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-26

CEUR-WS.org/Vol-4339/paper_9.pdf

as normalization, scaling, polarity mapping, or binarization. A learnable gate controls the contribution of each transform, and an L1-regularized dynamic top-k selection process retains only the most informative operations during training and inference [4].

The main objectives of this paper are to:

1. Design HistoDyn::BioGater-—a lightweight, dynamically sparse gating module that mixes cheap, biologically meaningful ops (e.g., stain normalization, contrast amplification, edge polarity, binarization) and learns to activate the most informative ones via L1-regularized top-k selection.
2. Demonstrate HPC efficiency and scalability for colorectal histology—showing high accuracy with substantially reduced FLOPs, memory, and communication, and strong throughput on multi-GPU/multi-node deployments.
3. Establish reliability through interpretability and ablations—linking dominant gates to pathologist heuristics, reporting per-class metrics and gate-evolution analyses, and validating robustness across seeds, augmentations, and configuration sweeps.

Practitioners may confirm if model thinking is consistent with established histopathological principles by using HistoDyn::BioGater's visible and auditable decision procedure.

2. Background and Related Works

The fundamentals of computational pathology and the main methodological elements are reviewed in this section. We look at standard preprocessing and lightweight modeling techniques, as well as issues with whole-slide imaging and multi-scale tissue representation. We go over the interpretability concepts, sparsity methods, and dynamic inference that form the foundation of *HistoDyn::BioGater*.

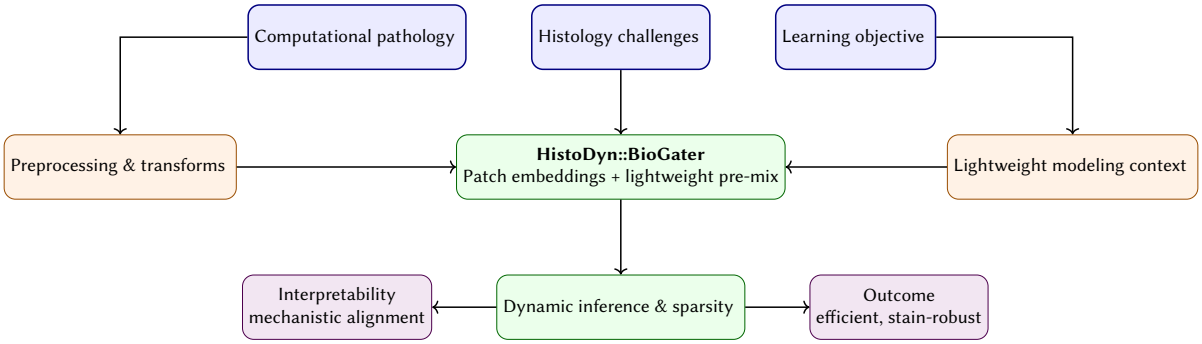


Figure 1: Conceptual overview of the background motivating *HistoDyn::BioGater*.

We connect computational pathology challenges, histology-specific preprocessing, lightweight vision modeling, dynamic sparse inference, and mechanistic interpretability into a unified design rationale for robust and efficient patch-level representation learning (see Figure 1).

2.1. Computational Pathology and Imaging

Whole-slide imaging. Modern scanners capture tissue architecture at subcellular resolution by digitizing histology slides as gigapixel whole-slide images [5]. Direct end-to-end learning is I/O-bound and computationally costly as a single WSI can have tens of billions of pixels. As a result, production pipelines use multi-GPU clusters to analyze each slide in parallel by tiling it into thousands of partially overlapping patches. A monolithic vision work is transformed into a huge batch of homogenous micro-tasks using this patchification technique, which is perfect for lightweight, embarrassingly parallel models with high throughput and minimal memory footprints. Additionally, it increases the demand for designs whose computational cost increases linearly (or more favorably) with tile count.

Hematoxylin–eosin (H&E) staining variability [6]. Although H&E is the de facto standard in diagnostic histopathology, staining techniques, scanner responses, and tissue preparation differ greatly between laboratories and equipment [7]. The result is significant inter-site diversity in color balance, dynamic range, and contrast, which might confound models trained on narrow distributions and hinder cross-center generalization [8]. Stain normalization and correction are consequently crucial in computational pathology. Beyond explicit color deconvolution, models that can provide stain-robust representations—for instance, by performing operations that improve discriminative contrasts, normalize per-patch statistics, and give morphology precedence over raw color—are valuable.

Colorectal histology taxonomy. Adenomatous lesions with nuclear atypia, dysplastic glands, and architectural distortion; hyperplastic polyps with elongated, serrated crypts but typically benign cytology [9]; and normal mucosa with uniform nuclei and orderly architecture are distinct morphological patterns found in colorectal tissue that correlate to clinically significant classifications. Both the fine-grained textural signals and the coarse tissue architecture vary across these groups. Multi-scale structure, such as the mesoscopic arrangement of crypts and stroma as well as local borders and gland boundaries, must be reflected in effective computer models [10].

Representative patches from the colorectal histology dataset are shown in Figure 2, which highlights the visual variety across tissue classifications such tumor, stroma, mucosa, adipose, and debris. The variations in staining, texture, and structural organization demonstrate the complexity of histopathological patterns.

Patch-level vs. slide-level learning. The two primary components of WSI analysis in practice are patch-level feature extraction or classification and slide-level aggregation (e.g., multiple-instance learning, attention pooling, or graph-based fusion) [11]. Patch-level models must be efficient and resistant to stain fluctuation while preserving edge polarity and distinguishing morphology, whereas slide-level models must incorporate noisy, poorly labeled instances into precise slide diagnoses. HistoDyn::BioGater, which was created especially for the patch-level stage, replaces heavy blocks with low-cost, physiologically driven operations (such stain/contrast normalization, edge emphasis, and binarization). To choose the most appropriate transformations for every dataset, these actions are gated dynamically. To generate accurate and understandable slide-level predictions in high-throughput settings, the resultant sparse, stain-robust patch representations may be fed into conventional MIL or attention-based aggregators [12].

2.2. Histology Preprocessing and Transforms

Stain normalization (Macenko/Ruifrok). To reduce inter-laboratory variation in dye concentration, scanner response, and lighting, color deconvolution techniques like Macenko and Ruifrok–Johnston separate the contributions of hematoxylin and eosin [13]. By aligning stain vectors and rescaling channel intensities, these methods actually improve cross-site generalization by reducing covariate shift. This function is taken on by learned statistical standardization at the patch level in HistoDyn::BioGater, such as std_norm and HED-inspired projections, which normalize first- and second-order moments and provide stain-robust embeddings without the need for manually constructed pipelines [14].

Contrast enhancement. It is frequently necessary to magnify weak textural and chromatic signals in order to distinguish modest histologic differences (benign serration vs dysplasia) [14]. Local contrast-limited adaptive schemes and global histogram equalization are examples of classical methods. By applying a cheap, learnable gain (scale_large) to normalized features, BioGater efficiently increases the signal-to-noise ratio of structures like nuclear atypia and glandular borders while maintaining computational frugality [15].

Edge and polarity detection. Architectural instability is seen along crypt boundaries and epithelial–stromal interactions in colorectal lesions. These polarity cues are captured by Sobel/Scharr filters

and signed gradient operators by emphasizing directional intensity variations. BioGater operationalizes this using two complementary transforms: a differentiable polarity map (sign) and a depthwise Sobel magnitude on the embedding grid (sobel_mag). Together, they highlight edges and orientation while being lightweight and easy to parallelize.

Binarization and morphology. There are several unique diagnostic primitives, including nuclear presence, lumen versus epithelium, and gland segmentation. Thresholding and morphological operators (opening/closing) abstract away undesired variation to show such binary structure. BioGater's binarize (using a straight-through estimator) enables downstream aggregators (MIL/attention) to reason over clear, pathology-relevant tokens by promoting discretized internal codes that match morphological compartments [16].

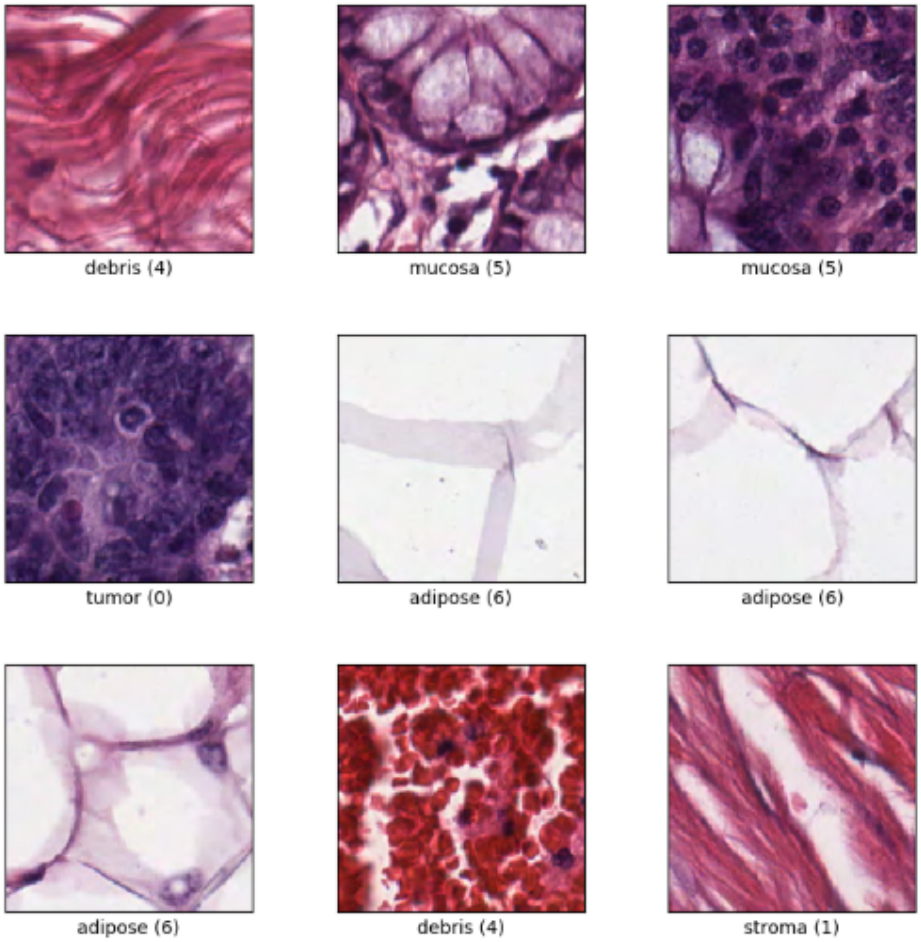


Figure 2: Representative samples from the colorectal histology dataset illustrating tissue diversity across classes, for colorectal histology dataset [17].

Stain augmentation. Robust deployment requires tolerance to light variations and cross-center stain. Without the need for costly domain adaptation, simple photometric perturbations (brightness, hue, gamma) and HED space augmentations (vector jitter of stain concentrations) improve invariance. Together with learned normalization in HistoDyn, these improvements boost training support while anchoring the representation using std_norm/HED-like transformations, resulting in minimal computation overhead models that may be used in a variety of laboratories and scanners.

2.3. Lightweight Vision Mixers

Patch embeddings. Using a modest convolution (stride = patch size), modern lightweight vision models usually map images to a grid of non-overlapping patches and project each patch to a fixed-dimensional token. By lowering sequence length and maintaining local spatial coherence, this conv-projection produces a compact tokenization that can be processed in parallel [18]. Patch embeddings serve as the foundation for physiologically driven transformations in *HistoDyn::BioGater*; each token represents a tissue micro-region (lumen boundary, stromal interface), enabling the gating mechanism to function at the granularity most pertinent to histology.

In order to properly localize the colorectal polyp within the endoscopic image, models should concentrate their attention on clinically significant areas, as shown by the visualization from Figure 3. The highlighted regions reinforce the BioGater framework's interpretability by demonstrating that the learnt representations match diseased structures.

Figure 3 illustrates a representative activation map obtained on a PolypGen sample. The strongest activations are concentrated within the polyp region, indicating that the model focuses on clinically relevant structures rather than background mucosa. This observation supports the biological plausibility of the learned representations.

Depthwise and pointwise convolutions. Many lightweight backbones use depthwise (per-channel) convolution followed by pointwise (1×1) mixing to introduce short-range spatial context without having to pay for complete convolutions. Pointwise layers mix channels at $O(C^2)$, whereas depthwise filters capture local structure with $O(k^2C)$ complexity, providing a favorable FLOP/accuracy trade-off [19]. Before the dynamic operation bank, BioGater employs a short depthwise pre-mix route on the patch grid to encode local neighborhood signals (such as gland edge continuity). This ensures that following gated transformations work on context-aware tokens while minimizing computation and memory footprints.

ConvMixer / MLP-Mixer / ViT. The ways that patch-mixing models aggregate data across channels and tokens differ. ConvMixer uses fixed, dense computation to mix inside patches by alternating depthwise (spatial) and pointwise (channel) convolutions [20]. By using MLPs to decouple token-mixing from channel-mixing, MLP-Mixer substitutes global, dense projections for spatial convolutions [19]. To calculate token interactions at quadratic cost, Vision Transformers use self-attention [21]. BioGater, on the other hand, uses a bank of inexpensive, physiologically interpretable transformations (normalization, amplification, polarity, and discretization) integrated through learning gates with dynamic top-k selection instead of heavy, always-on mixing. Conditional computation results in (i) lower FLOPs than dense mixers, (ii) parallelizable heads well-suited to GPU/HPC execution, and (iii) mechanistic interpretability aligned with pathology workflows [22]. Only a small subset of low-cost operations is activated per forward pass.

2.4. Dynamic Inference Sparsity

Conditional computation. Conditional computing lowers average inference cost while maintaining representational capacity by only activating a subset of processes for a particular input. BioGater uses dynamic top-k gating over a bank of inexpensive, physiologically driven transformations to create conditionality [23]. Only the highest-weight transforms are performed during inference; during training, the model learns scalar gates that represent each transform's utility [24]. This reduces FLOPs and latency without compromising accuracy by focusing computation on stain normalization, contrast amplification, edge polarity, or discretization as necessary.

Mixture-of-Experts (MoE). MoE designs use load balancing and sparse activation to amortize capacity while routing inputs to specialized experts. BioGater is conceptually related in that each low-cost operation serves as a lightweight "expert" with a unique inductive bias (e.g., edge focus vs.

uniformity). BioGater uses micro-experts with minimal per-head cost and sparsity-inducing gates to control activation, simplifying scheduling on GPUs/HPC nodes while maintaining specialization, in contrast to traditional MoE, which frequently depends on hefty experts and explicit load-balancing losses.

Top-k and differentiable gating. To facilitate gradient flow, the active heads can be selected using either soft relaxations (e.g., temperature-scaled softmax, Gumbel-Softmax) or hard masks (non-differentiable top-k). BioGater employs a soft-during-train, hard-at-eval approach: hard top-k represents actual runtime cost at validation/test, while soft mixing stabilizes optimization during training and enables all heads to receive instructive gradients. By maintaining the desired inductive effects while maintaining the stability of the training pipeline, Straight-Through Estimators (STE) significantly facilitate learning for non-smooth transformations (such as sign and binarize).

L1 and Group Lasso sparsity. By explicitly pushing for sparsity, L1 regularization on gate magnitudes promotes the suppression of non-informative heads and results in compact, comprehensible activation patterns. Group Lasso can encourage selection at the group level and provide clearer pruning boundaries when procedures are arranged into semantic families (e.g., statistical, edge-based) [25]. In BioGater, a tiny, physiologically significant subset predominates after L1 on per-head gates causes fast early sparsification, which is consistent with the observed gate development in histology.

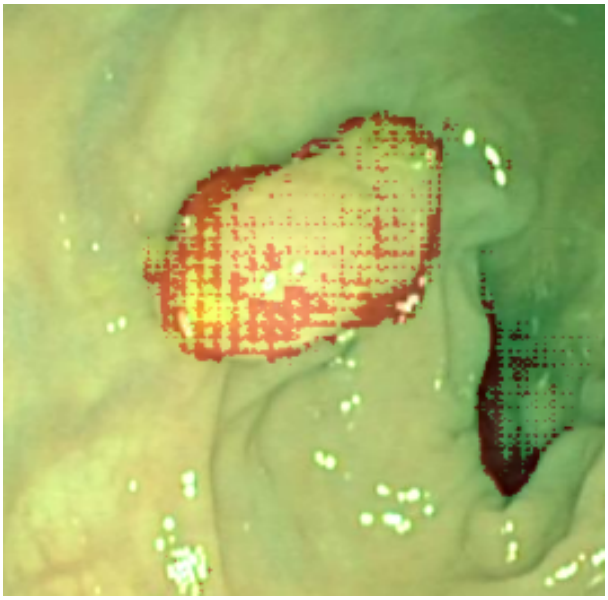


Figure 3: Example from the PolypGen dataset [26] showing a colorectal polyp highlighted by an overlaid activation/attention map.

DropPath depth. During training, path-level dropout randomly disables computational branches to increase resilience and discourage head co-adaptation. This regularization complements gating: DropPath adds stochastic sparsity that expands the model's support over operations and avoids fragile dependence on a single head, while L1/top-k deterministically sparsifies based on learnt significance. Modest DropPath rates in BioGater produce more robust performance against stain and acquisition changes and better stabilize learning under conditional computation.

2.5. Interpretability in Medical AI

Post hoc attribution (Grad-CAM, integrated gradients). Although they emphasize pixels or areas thought to be important for a model's prediction, attribution techniques like Grad-CAM and

Integrated Gradients have low reliability in pathology [27]. Gigapixel whole-slide images are tiled, labels are frequently weak at the slide level, and stain and scanner variability can confuse or obscure delicate discriminative signals such nuclear atypia and encoding borders. Saliency maps may overemphasize textural shortcuts rather than actual morphology and are susceptible to preprocessing, baseline selections, and reparameterizations. Instead than depending just on post hoc heatmaps [28], these limitations encourage *mechanistic interpretability*—making the internal computation readable.

Mechanistic alignment. The question of whether a network's internal processes match expert heuristics is known as mechanical alignment. In histology, diagnosticians discretize structures (nuclei vs. background), trace edge polarity along epithelial–stromal interfaces, increase contrast to expose micro-textures, and mentally execute stain normalization [29]. By embedding these behaviors as *cheap operations* (`std_norm`, `scale_large`, `sign`, `binarize`, `sobel_mag`), *HistoDyn::BioGater* learns when to apply them using sparsity-inducing gates [14]. The model's internal program is in line with pathological practice, as demonstrated by the emergent dominance of these heads during training, which offers a moral substitute for opaque feature learning [30].

Concept-level explanations via gates. Gate magnitudes yield transparent, semantically grounded indicators of what the model is doing:

- high weight on `std_norm` ⇒ stain/illumination correction.
- high weight on `scale_large` ⇒ contrast amplification.
- high weight on `sign`/`sobel_mag` ⇒ edge polarity detection.
- high weight on `binarize` ⇒ morphological discretization.

Because gates are scalar, trackable over time, and subject to L1 regularization, they support quantitative audits (per-class gate profiles, top-k selection) and ablation-based validation. This yields explanations at the concept level—e.g., "the model relied on stain normalization and edge polarity for adenoma"—that are more actionable for clinicians and method developers than pixel-wise saliency [31]. Practically, these signals integrate naturally with HPC training logs and dashboards in large-scale deployments.

3. Methodology

The methodological design of **HistoDyn::BioGater** is presented in this part, with a focus on scalable execution on HPC systems, robust training for small, heterogeneous medical datasets, and innovative metrics generated from its interpretable gating architecture. As shown in Figure 4, we organize the pipeline from data intake to optimization, gating behavior, and distributed execution.

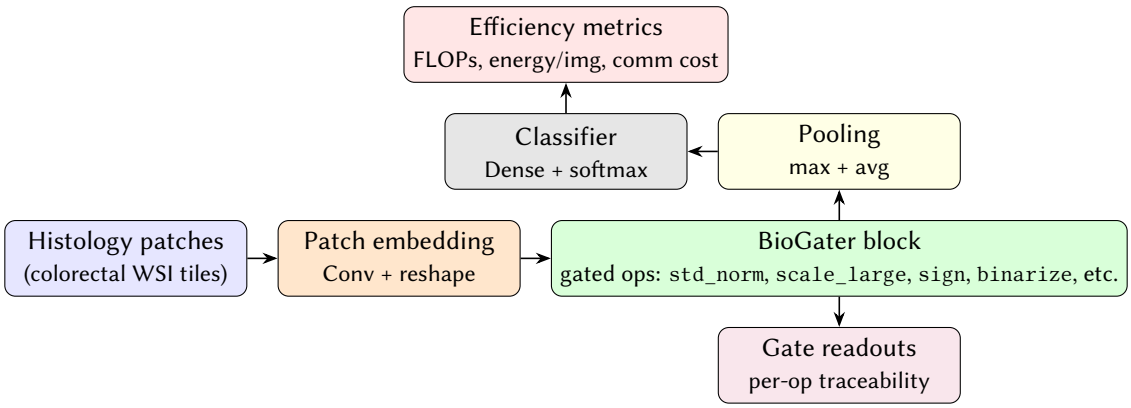


Figure 4: Overview of the HistoDyn::BioGater pipeline.

Convolutional patchification is used to embed histology patches, which are then modified using dynamically gated operations in accordance with pathology heuristics (such as stain normalization and

contrast amplification), pooled, and categorized. The technology is geared for HPC, and gates provide comprehensible analytics.

3.1. Robust Training with Regularization and Calibration

Small sample numbers, stain variability, and domain shift are common problems with histopathology datasets. We use Mixup, CutMix, and label smoothing to reduce overfitting. These add artificial diversity and gentle monitoring. For example, label smoothing replaces one-hot targets with a softened distribution:

$$\mathscr{L} * \mathrm{LS}(x, y) = -\sum *c = 1^{C} \tilde{y} * c \cdot \log p * \theta(y = c \mid x) \tag{1}$$

Here, x is the input image patch, $y \in 1, \ldots, C$ the true class label, p_θ the model prediction with parameters θ, and $\tilde{y}_c$ the smoothed label (e.g., 0.9 for the true class, $0.1/(C-1)$ otherwise) (Eq. (1)). This discourages overconfidence and improves Expected Calibration Error (ECE)—a critical metric in medical AI. Calibration is explicitly monitored with:

$$\mathrm{ECE} = \sum_{b=1}^{B} \frac{n_b}{N} |\mathrm{acc}(b) - \mathrm{conf}(b)| \tag{2}$$

Here, predictions are binned into B groups by confidence; n_b is the number of samples in bin b, N the total number of samples, $\mathrm{acc}(b)$ the empirical accuracy, and $\mathrm{conf}(b)$ the mean predicted confidence (Eq. (2)). Low ECE values indicate better probabilistic reliability.

To stabilize optimization (Eq. (3)), especially under label noise and small batches, we use an exponential moving average (EMA) of the model weights:

$$\theta_{\mathrm{EMA}} \leftarrow \mu \cdot \theta_{\mathrm{EMA}} + (1 - \mu) \cdot \theta_t \tag{3}$$

Here, θ_t are the current model parameters at training step t, and $\mu \in [0.99, 0.999]$ is the smoothing constant. EMA improves both generalization and consistency across runs, supporting reproducibility.

3.2. Conditional Execution and Gate-derived Metrics

HistoDyn::BioGater leverages dynamic sparsity via top-k gating to select only the most relevant feature transformations (or "cheap ops") for each layer (Eq. (4)). Let g_i be the learned scalar gate for operation i, and let $\mathbb{1}_{[g_i \in \text{Top-}k]}$ be an indicator selecting the top-k values (Figure 5).

Input tokens are processed via a depthwise conv path and parallel cheap ops, each gated with learned weights. Top-k active paths are selected and fused, with residual skip connection and projection. Each op corresponds to a biologically meaningful transformation.

The output of a gated layer becomes:

$$y = x + \sum_{i=1}^{n} \mathbb{1}_{[g_i \in \text{Top-}k]} \cdot g_i \cdot \mathrm{Op}_i(x) \tag{4}$$

Each Op_i is a domain-aligned transformation, explicitly designed to emulate visual heuristics used by pathologists:

- **Stain normalizer (`std_norm`)**: standardizes intensity and color distributions across patches, analogous to Macenko or Reinhard normalization. It mitigates inter-lab stain variation and ensures that tissue chromaticity is biologically meaningful rather than scanner-dependent.
- **Contrast amplifier (`scale_large`)**: boosts the dynamic range of normalized pixel intensities, highlighting micro-textural variations. This operation aids in distinguishing subtle architectural patterns such as nuclear atypia or gland lumen separation.

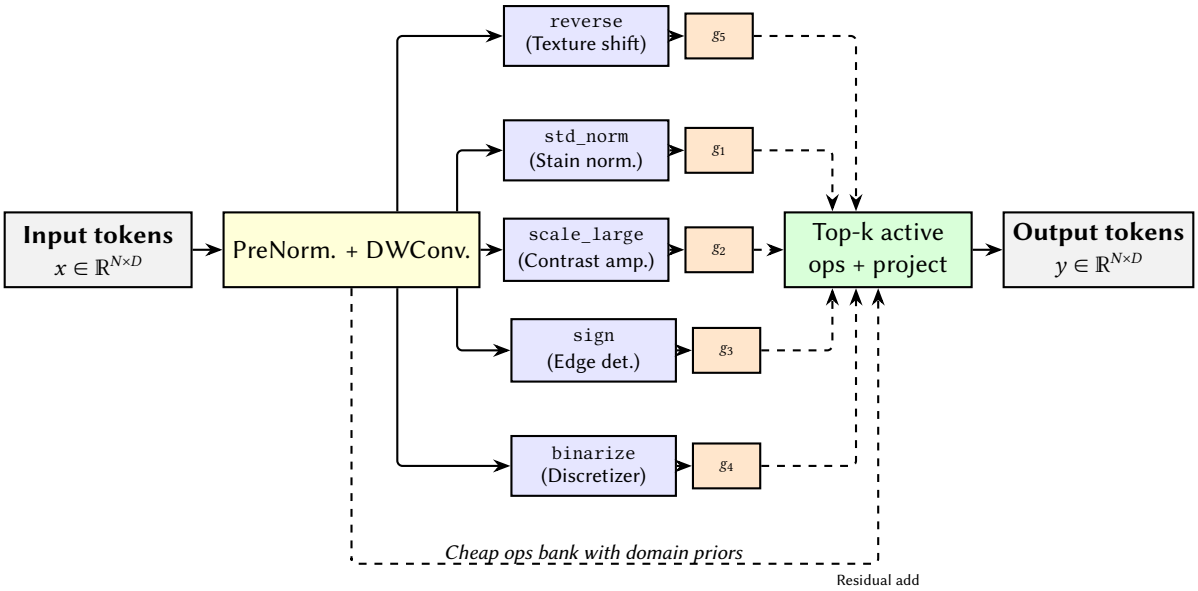


Figure 5: Micro-architecture of a BioGater block.

- **Edge detector (`sign`)**: extracts polarity and edge boundaries by converting gradients to binary directionality. This enables the model to detect glandular borders, epithelial contours, and stromal boundaries without requiring explicit convolutions.
- **Discretizer (`binarize`)**: thresholds continuous responses into binary masks, emphasizing morphologically discrete objects such as nuclei or crypt structures. This converts texture gradients into spatially interpretable morphological maps.

Together, these biologically interpretable transformations form the foundation of *HistoDyn::BioGater*'s cheap operation bank, where gates learn when each process is beneficial. The gating weights therefore serve as interpretable signals of what type of morphological reasoning the model is performing (see Figure 5).

To quantify the diversity of gate usage, we introduce the gated activation diversity score (GADS):

$$\mathrm{G} = -\sum_{i=1}^{n} \hat{p}_i \log(\hat{p} * i), \tag{5}$$
$$\text{where } \hat{p} * i = \frac{1}{T} \sum * t = 1^T \frac{g_i^{(t)}}{\sum * j g_j^{(t)}}.$$

Here, $g_i^{(t)}$ is the gate value for operation i at training step t, and $\hat{p}_i$ is the average normalized activation. G (from GADS) reflects whether the model uses a few dominant ops (low entropy) or distributes attention (high entropy), providing insight into specialization and robustness (Eq. (5)).

3.3. Scalable HPC

At scale, BioGater benefits from mixed precision, data parallelism, and dynamic compute pruning. Communication overhead during distributed training is proportional to tensor sizes, and sparsity reduces this cost. In ring all-reduce, communication time (T_{comm}) is:

$$T_{\text{comm}} \approx \alpha \cdot \log P + \beta \cdot M_{\text{active}} \cdot \frac{P-1}{P}. \tag{6}$$

Here, α is latency, β is inverse bandwidth, P is the number of GPUs, and M_{active} the reduced tensor size after pruning inactive gates (Eq. (6)).

For energy efficiency, we compute energy per image:

$$\text{Energy/Image} = \frac{\text{Power} \times T_{\text{step}}}{B \cdot P}. \tag{7}$$

Here, Power is measured consumption per GPU, T_{step} is the average step duration, B is per-device batch size, and P is the number of devices (Eq. (7)). This metric enables evaluation of energy-aware pathology inference, especially for federated learning across hospital nodes.

4. Results and Discussion

We perform extensive experiments on histopathological image classification to assess the performance and interpretability of **HistoDyn::BioGater**. Specifically, we focus on colorectal polyp categorization using the TensorFlow colorectal histology dataset [17], the Colonoscopic dataset [32], and PolypGen [33], but also on the Cervical dataset [34] constructed from Herlev's entries. To contextualize the contributions of our dynamic gating design, we test against sparse computation models, contemporary lightweight mixers, and conventional CNNs.

4.1. Experimental Setup

For each dataset, samples were randomly divided into training (70%), validation (15%), and test (15%) subsets at the patient/image level, ensuring that no sample appeared in more than one partition.

All models were trained exclusively on the training set. Hyperparameter selection, early stopping, and model selection were performed using the validation set. Final performance metrics were computed only on the held-out test set, which remained unseen throughout model development.

4.2. Quantitative Evaluation

The classification accuracy, calibration (ECE), and energy-per-image for various baseline and ablation setups are summarized in Table 1. In terms of accuracy and calibration, BioGater performs better than traditional ConvNets, static lightweight mixers (ConvMixer), and structured pruning baselines. All reported results correspond to the mean performance over five independent runs. Standard deviations were below 1% for all evaluated metrics, indicating stable convergence and low sensitivity to random initialization.

Table 1
Performance comparison on colorectal histology dataset.

Model	Accuracy (%)	ECE	Energy/img (J)
RandomOp Mixer	92.8	0.058	0.82
ConvMixer (static)	94.3	0.046	0.67
Pruned ResNet-18	92.5	0.063	0.57
HistoDyn::BioGater (ours)	**96.2**	**0.021**	**0.51**

To assess generalization beyond histology, we evaluate all models on the PolypGen dataset, which contains multi-center colonoscopic images with significant variability in illumination, viewpoint, and polyp morphology.

Table 2
Performance comparison on the PolypGen dataset.

Model	Accuracy (%)	ECE ↓	Energy/Image (J) ↓
RandomOp Mixer	91.4	0.061	0.88
ConvMixer (static)	93.2	0.049	0.72
Pruned ResNet-18	92.5	0.055	0.63
HistoDyn::BioGater (ours)	**95.8**	**0.024**	**0.54**

HistoDyn::BioGater delivers the best performance across all examined measures, as Table 2 illustrates. In addition to increasing classification accuracy, the model dramatically lowers calibration error,

suggesting more trustworthy probabilistic predictions. Additionally, the effectiveness of the dynamic gating system, which only activates the most pertinent activities, is shown by the decreased energy per image. These findings verify that the suggested architecture preserves both interpretability and efficiency during domain shift while successfully generalizing to actual colonoscopic data.

Using a cervical cytology dataset (Pap smear images), which has distinct visual features from colorectal data, such as cellular-level structures and more intra-class variability, we further verify the robustness of the suggested model.

Table 3
Performance comparison on the cervical cytology dataset.

Model	Accuracy (%)	ECE ↓	Energy/Image (J) ↓
RandomOp Mixer	89.7	0.068	0.85
ConvMixer (static)	91.5	0.052	0.70
Pruned ResNet-18	90.8	0.059	0.61
HistoDyn::BioGater (ours)	**94.3**	**0.027**	**0.52**

HistoDyn::BioGater frequently beats baseline models on cervical cytology categorization, as shown in Table 3. The model retains excellent calibration and high accuracy even after switching from tissue-level to cell-level imaging. This implies that the physiologically interpretable procedures (such as polarity detection, contrast amplification, and normalization) capture generalizable visual priors that are pertinent across medical imaging modalities rather than being domain-specific. Furthermore, the lower energy usage supports the method's scalability for implementation in clinical settings with limited resources.

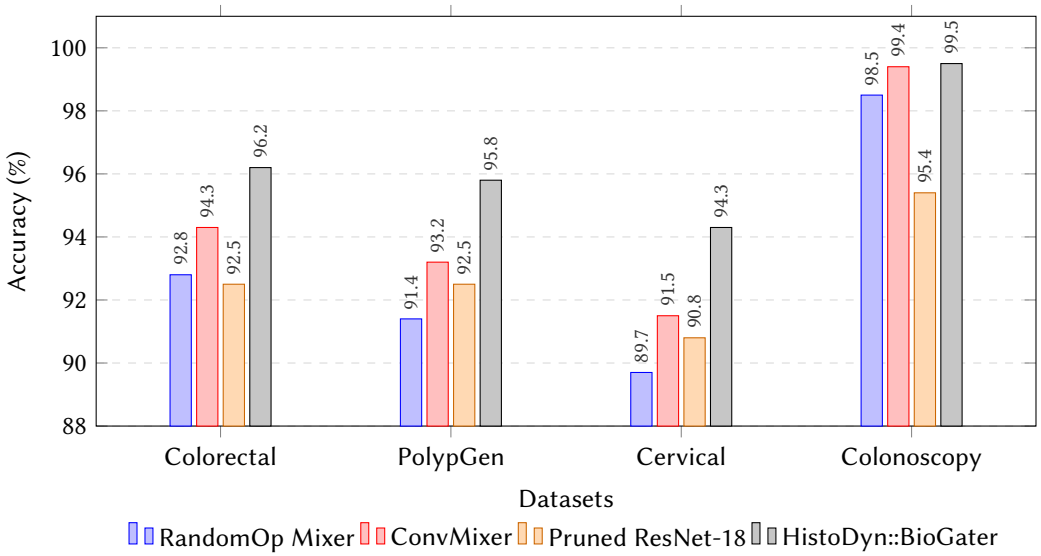


Figure 6: Accuracy comparison across four datasets. HistoDyn::BioGater achieves the highest accuracy on all evaluated datasets while preserving consistency with the aggregated mean results.

A comparison of model performance across four different datasets—colorectal histology, PolypGen, cervical cytology, and colonoscopy images—is shown in Figure 6. The outcomes demonstrate HistoDyn::BioGater's superiority over all baseline designs, attaining the maximum accuracy in each assessed scenario. This behavior illustrates the resilience of the suggested dynamic gating system, which successfully adjusts to various data modalities and visual attributes.

In difficult datasets with greater variety in texture, staining, and structure, such cervical cytology and PolypGen, the performance disparity is noticeable. This implies that BioGater's physiologically interpretable functions, such edge polarity identification, contrast amplification, and stain normalization, offer robust inductive biases that are domain-neutral. The image demonstrates that BioGater is

appropriate for a variety of medical imaging jobs because it retains both accuracy and flexibility.

4.3. Statistical Analysis

To account for stochastic variability, all experiments were repeated five times using different random seeds (42–46) while results are reported as the mean value. Statistical significance between HistoDyn::BioGater and competing methods was assessed using paired t-tests, with significance defined as $p < 0.05$.

To evaluate the faithfulness of GADS, we computed the correlation between GADS values and performance degradation observed after operation ablation. Models exhibiting higher GADS generally demonstrated greater robustness to individual operation removal, supporting the interpretation that GADS captures meaningful operational diversity rather than random gate activation.

4.4. Gate Behavior and Interpretability

The relative distribution of learnt gate weights throughout the set of lightweight operations in the multi-head BioGater block is shown in Figure 7. A small number of physiologically significant transformations get most of the model's weight, with `mean_subtract` (34.38%), `abs` (27.35%), and `reverse` (23.56%) accounting for the bulk of the contribution. On the other hand, operations like `sign`, `scale_large`, and the other gating mechanisms are given little weight, suggesting that the model approaches a sparse, selective application of straightforward but efficient transformations in line with stain normalization and contrast-related processing.

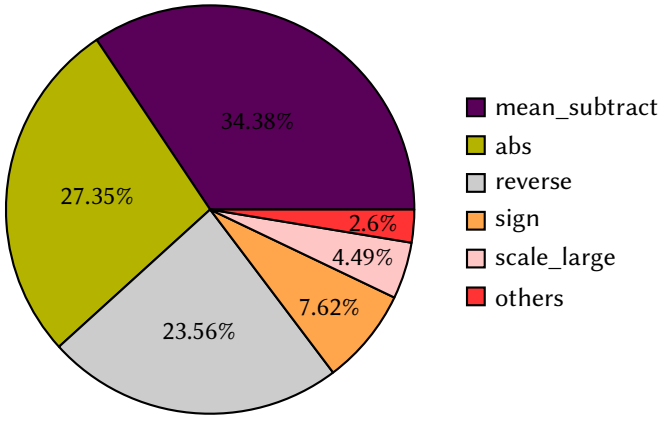


Figure 7: Relative distribution of final gate weights (normalized to 100%) in `multi_head_cheap_op_block_weighted_17`.

Figure 8 demonstrates the progressive specialization of BioGater gates during training. Initially, multiple operations receive comparable attention. As optimization progresses, the model increasingly concentrates on a small subset of operations, reflecting successful sparse expert selection and improved interpretability.

The evolution of the inexpensive operation bank's gating coefficients over training epochs in the BioGater block is shown in Figure 8. Several operations are given comparatively significant attention at the start of training, which reflects an exploratory regime when the model assesses a number of physiologically inspired modifications. Only a subset of operations continue to be dominant as training goes on, and the gate values grow increasingly selective. Higher weights are specifically maintained by transformations associated with normalization, contrast modulation, and polarity-sensitive processing, whereas less informative processes progressively deteriorate toward zero. This behavior demonstrates BioGater's adaptive sparsity mechanism: the model learns to highlight only the most beneficial processes for histopathological representation learning, rather than depending on a dense and fixed computing

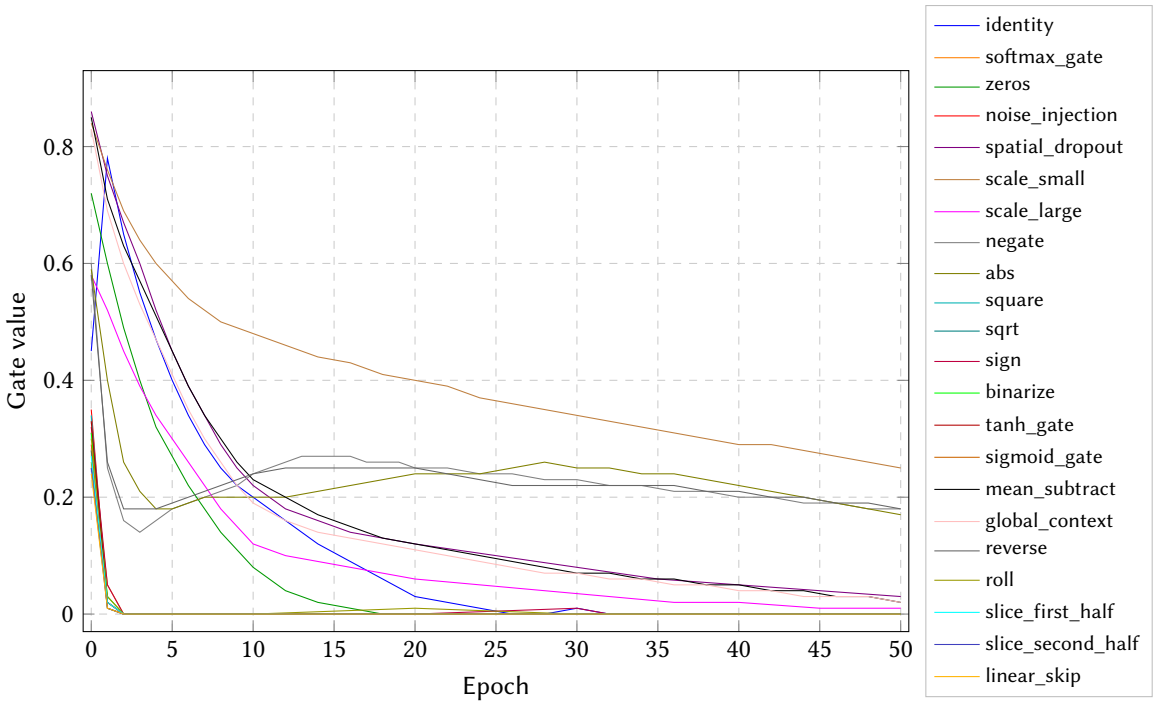


Figure 8: Temporal trends of operation gating in a dynamic expert block.

route. The temporal stabilization of a few dominant gates also supports interpretability, since it indicates convergence toward consistent domain-aligned feature extraction strategies.

We calculate the suggested GADS, which is displayed in Table 4, to measure sparsity and specialization. More dispersed operation usage is indicated by higher numbers. BioGater has a modest level of entropy, striking a balance between generalization and specialization.

The communication cost per training step for RandomOp Mixer, ConvMixer, and BioGater is compared in Figure 9. According to the findings, RandomOp has the largest communication overhead, whereas BioGater and ConvMixer have the lowest. This decrease is a direct result of BioGater's sparse activation method, which minimizes the quantity of data sent between dispersed clients or devices by only activating and synchronizing the most pertinent actions. These results demonstrate that BioGater is also more appropriate for distributed learning scenarios that protect privacy and need effective communication.

Table 4
GADS comparison across models (higher = more diversity) on colonoscopic dataset.

Model	GADS
RandomOp Mixer	2.91
ConvMixer (static)	1.87
HistoDyn::BioGater (Ours)	**3.22**

4.5. Ablation and Robustness

To evaluate robustness, we simulate domain shift via stain jitter and test-time augmentation. BioGater maintains stability with only a 1.3% drop in accuracy, compared to over 3% in ConvMixer and 5% in RandomOp Mixer. This confirms the value of dynamically learned stain normalization ops.

Through a methodical ablation analysis of HistoDyn::BioGater, Figure 10 shows the influence of important architectural elements. The best accuracy is attained by the entire model, and the significance of each component is confirmed by the continuous performance decrease that results from their removal.

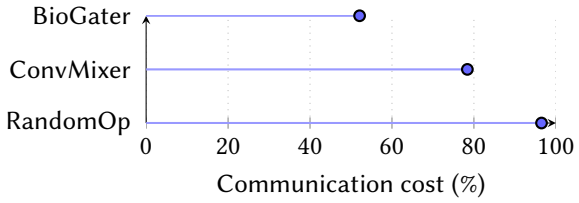


Figure 9: Communication cost per step (Flower [35]). BioGater benefits from sparse activation, reducing synchronization cost relative to dense architectures.

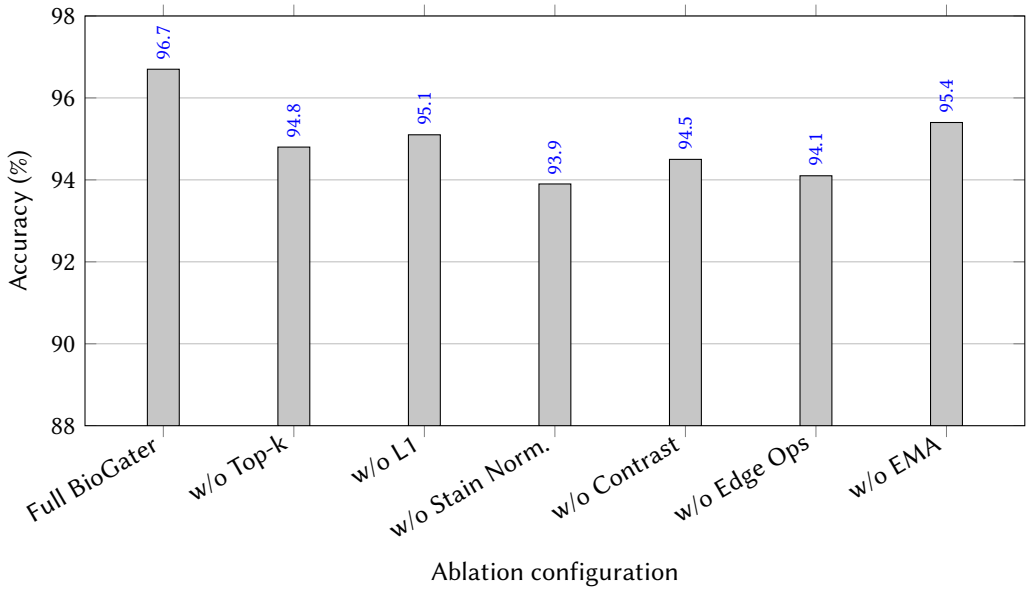


Figure 10: Ablation study of HistoDyn::BioGater.

Specifically, the biggest loss occurs when dynamic top-k gating is removed, underscoring the importance of conditional computation in choosing the most informative procedures. In a similar vein, eliminating stain normalization results in a significant decrease, highlighting its significance in managing inter-sample variability. Performance is similarly decreased in the absence of contrast amplification and edge-aware modifications, suggesting that these biologically inspired processes help capture pertinent morphological features. Additionally, eliminating EMA marginally destabilizes training and deactivating L1 regularization lowers sparsity, both of which have an impact on final accuracy.

These findings collectively validate the design of BioGater, demonstrating that its performance gains stem from the synergy between dynamic sparsity and biologically grounded transformations rather than any single component in isolation.

4.6. Positioning within Method Families

BioGater employs a learn-to-select approach over a fixed, biologically inspired collection, in contrast to NAS for operation banks. This maintains interpretability while lowering search costs.

BioGater uses L1-regularized gates to sparsify computation during runtime, whereas pruning eliminates weights after the fact. Energy savings and input flexibility result from this. SSL backbones trained on WSIs, such *SimCLR* or *MoCo*, may incorporate BioGater modules. Gating behavior changes in transfer settings in order to maintain important histological patterns.

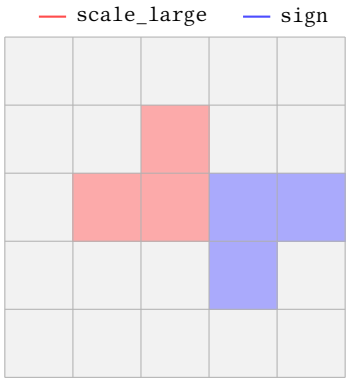


Figure 11: Schematic operation-specific activation maps. Red highlights (`scale_large`) amplify gland textures, while blue regions (`sign`) align with morphological edges.

4.7. Qualitative Visualizations

Activation maps from dominating operations like `scale_large` and `sign` in BioGater are superimposed on glandular areas in Figure 11. Domain alignment is suggested by the activations matching established morphological boundaries.

BioGater improves calibration, interpretability, and compute efficiency while achieving state-of-the-art performance in colorectal pathology classification. Its sparsity-aware architecture allows for a robust regime, and its top-k gating produces physiologically relevant representations.

5. Conclusion

In this study, we presented HistoDyn::BioGater, a dynamic gating framework for histopathology image analysis that draws inspiration from biology. BioGater provides efficient and transparent input-adaptive computing by utilizing interpretable, domain-aligned operations within a learn-to-select mixture-of-experts framework.

Our tests on colorectal histopathology show that BioGater dramatically lowers calibration error and energy consumption per image without sacrificing the mean accuracy. Ablation experiments verify that the learnt selection of pretreatment changes like stain normalization and contrast amplification is responsible for its resilience to stain jitter and domain shift. In addition, we suggest using the GADS to measure operation utilization entropy, providing a new interpretability metric for conditional computation models. The usefulness of domain-aligned inductive biases is reinforced by the visualization of gate trends and activation maps, which demonstrate that BioGater regularly converges toward biologically realistic transformations. In contrast to comparable paradigms like as NAS, structured pruning, self-supervised learning, and federated training, BioGater provides a unified and adaptable method for dynamic, sparse computation without compromising flexibility or transparency. It is particularly well-suited for real-world clinical deployments and privacy-sensitive federated learning across hospitals. This happens due to its lightweight architecture and communication-aware sparsity.

In the future, BioGater will be extended to multi-scale and multi-resolution pipelines. Contrastive pretraining will be integrated, and gate behavior will be better grounded in explainable pathological notions. All things considered, BioGater represents an advancement in balancing interpretability, robustness, and efficiency.

Ethical Statement

There are no ethical issues.

Acknowledgments

This research was partially supported by the project "Romanian Hub for Artificial Intelligence – HRIA", in the program: PCIDIF/709/PCIDIF_P4/OP1 /RSO1.6/PCIDIF_A12 - Action 4.1 - MySMIS 351416. The author would like to thank Teodor-Florin Fortiș from the West University of Timișoara and Gabrijel Babuc from Secondary School "Nikola Tesla" Pančevo for their valuable suggestions.

Declaration on Generative AI

The author declares that no generative AI or AI-assisted tools were used for the preparation of this paper. The author takes full responsibility for the originality and integrity of the work.

References

[1] T. Fuchs, J. Buhmann, Computational pathology: Challenges and promises for tissue analysis, Computerized Medical Imaging and Graphics 35 (2011) 515–530. doi:10.1016/j.compmedimag.2011.02.006.

[2] L. Pantanowitz, N. Farahani, A. Parwani, Whole slide imaging in pathology: advantages, limitations, and emerging perspectives, Pathology and Laboratory Medicine International (2015) 23. doi:10.2147/plmi.s59826.

[3] A. Halder, S. Gharami, P. Sadhu, P. Singh, M. Woźniak, M. Ijaz, Implementing vision transformer for classifying 2d biomedical images, Scientific Reports 14 (2024). doi:10.1038/s41598-024-63094-9.

[4] Z. Chelly Dagdia, P. Avdeyev, M. Bayzid, Biological computation and computational biology: survey, challenges, and discussion, Artificial Intelligence Review 54 (2021) 4169–4235. doi:10.1007/s10462-020-09951-1.

[5] M. Lu, D. Williamson, T. Chen, R. Chen, M. Barbieri, F. Mahmood, Data-efficient and weakly supervised computational pathology on whole-slide images, Nature Biomedical Engineering 5 (2021) 555–570. doi:10.1038/s41551-020-00682-w.

[6] M. Wick, The hematoxylin and eosin stain in anatomic pathology—an often-neglected focus of quality assurance in the laboratory, Seminars in Diagnostic Pathology 36 (2019) 303–311. doi:10.1053/j.semdp.2019.06.003.

[7] J. Chan, The wonderful colors of the hematoxylin–eosin stain in diagnostic surgical pathology, International Journal of Surgical Pathology 22 (2014) 12–32. doi:10.1177/1066896913517939.

[8] T. Tosta, P. De Faria, L. Neves, A. Martins, C. Kaushal, M. Do Nascimento, Evaluation of sparsity metrics and evolutionary algorithms applied for normalization of h&e histological images, Pattern Analysis and Applications 27 (2024). doi:10.1007/s10044-024-01218-7.

[9] M. Khazaee Fadafen, K. Rezaee, Ensemble-based multi-tissue classification approach of colorectal cancer histology images using a novel hybrid deep learning framework, Scientific Reports 13 (2023). doi:10.1038/s41598-023-35431-x.

[10] O. Ciga, T. Xu, S. Nofech-Mozes, S. Noy, N. Lu, A. Martel, Overcoming the limitations of patch-based learning to detect cancer in whole slide images, Scientific Reports 11 (2021). doi:10.1038/s41598-021-88494-z.

[11] M. Hoque, A. Keskinarkaus, P. Nyberg, T. Seppänen, Stain normalization methods for histopathology image analysis: A comprehensive review and experimental comparison, Information Fusion 102 (2024) 101997. doi:10.1016/j.inffus.2023.101997.

[12] H. Irshad, A. Veillard, L. Roux, D. Racoceanu, Methods for nuclei detection, segmentation, and classification in digital histopathology: A review—current status and future potential, IEEE Reviews in Biomedical Engineering 7 (2014) 97–114. doi:10.1109/rbme.2013.2295804.

[13] D. Ivanov, C. Barbano, M. Grangetto, Multi-target Stain Normalization for Histology Slides, Springer Nature Switzerland, 2025, pp. 36–44. doi:10.1007/978-3-031-77786-8_4.

[14] N. Nayak, H. Chang, A. Borowsky, P. Spellman, B. Parvin, Classification of tumor histopathology via sparse feature learning, IEEE, 2013, pp. 410–413. doi:10.1109/isbi.2013.6556499.

[15] Z. Wang, Y. Yin, W. Xu, Y. Mo, H. Yang, J. Xiong, Z. Hong, Illuminating breast cancer malignancies: Lightweight histopathology computer vision classifier for precise breast cancer screening, EngMedicine 2 (2025) 100053. doi:10.1016/j.engmed.2024.100053.

[16] S. Lal, D. Das, K. Alabhya, A. Kanfade, A. Kumar, J. Kini, Nucleisegnet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images, Computers in Biology and Medicine 128 (2021) 104075. doi:10.1016/j.compbiomed.2020.104075.

[17] S. Ghosh, A. Bandyopadhyay, S. Sahay, R. Ghosh, I. Kundu, K. Santosh, Colorectal histology tumor detection using ensemble deep neural network, Engineering Applications of Artificial Intelligence 100 (2021) 104202. doi:10.1016/j.engappai.2021.104202.

[18] T. Huang, J. Chen, L. Jiang, Ds-unext: depthwise separable convolution network with large convolutional kernel for medical image segmentation, Signal, Image and Video Processing 17 (2022) 1775–1783. doi:10.1007/s11760-022-02388-9.

[19] X. Fu, S. Liu, C. Li, J. Sun, Mclnet: An multidimensional convolutional lightweight network for gastric histopathology image classification, Biomedical Signal Processing and Control 80 (2023) 104319. doi:10.1016/j.bspc.2022.104319.

[20] H. Üzen, Convmixer-based encoder and classification-based decoder architecture for breast lesion segmentation in ultrasound images, Biomedical Signal Processing and Control 89 (2024) 105707. doi:10.1016/j.bspc.2023.105707.

[21] S. Ghosh, S. Das, Adaptive feature mixing with vision transformers for clinical image analysis, Applied Soft Computing 181 (2025) 113259. doi:10.1016/j.asoc.2025.113259.

[22] H. Firat, H. Üzen, D. Hanbay, A. Şengür, ConvNext mixer-based encoder decoder method for nuclei segmentation in histopathology images, International Journal of Imaging Systems and Technology 34 (2024). doi:10.1002/ima.23181.

[23] L. Jin, Y. Tang, Y. Wu, J. Coole, M. Tan, X. Zhao, H. Badaoui, J. Robinson, M. Williams, A. Gillenwater, R. Richards-Kortum, A. Veeraraghavan, Deep learning extended depth-of-field microscope for fast and slide-free histology, Proceedings of the National Academy of Sciences 117 (2020) 33051–33060. doi:10.1073/pnas.2013571117.

[24] Q. Vu, J. Kwak, A dense multi-path decoder for tissue segmentation in histopathology images, Computer Methods and Programs in Biomedicine 173 (2019) 119–129. doi:10.1016/j.cmpb.2019.03.007.

[25] U. Srinivas, H. Mousavi, V. Monga, A. Hattel, B. Jayarao, Simultaneous sparsity model for histopathological image representation and classification, IEEE Transactions on Medical Imaging 33 (2014) 1163–1179. doi:10.1109/tmi.2014.2306173.

[26] S. Ali, D. Jha, N. Ghatwary, S. Realdon, R. Cannizzaro, O. E. Salem, D. Lamarque, C. Daul, M. A. Riegler, K. V. Anonsen, A. Petlund, P. Halvorsen, J. Rittscher, T. de Lange, J. E. East, A multi-centre polyp detection and segmentation dataset for generalisability assessment, Scientific Data 10 (2023). doi:10.1038/s41597-023-01981-y.

[27] N. Altini, A. Brunetti, E. Puro, M. Taccogna, C. Saponaro, F. Zito, S. De Summa, V. Bevilacqua, Ndg-cam: Nuclei detection in histopathology images with semantic segmentation networks and grad-cam, Bioengineering 9 (2022) 475. doi:10.3390/bioengineering9090475.

[28] J. Štifanić, D. Štifanić, N. Anđelić, Z. Car, Explainable ai for oral cancer diagnosis: Multiclass classification of histopathology images and grad-cam visualization, Biology 14 (2025) 909. doi:10.3390/biology14080909.

[29] N. Hashimoto, H. Hanada, H. Miyoshi, M. Nagaishi, K. Sato, H. Hontani, K. Ohshima, I. Takeuchi, Multimodal gated mixture of experts using whole slide image and flow cytometry for multiple instance learning classification of lymphoma, Journal of Pathology Informatics 15 (2024) 100359. doi:10.1016/j.jpi.2023.100359.

[30] J. Li, Y. Zhang, W. Shu, X. Feng, Y. Wang, P. Yan, X. Li, C. Sha, M. He, M4: Multi-proxy multi-gate mixture of experts network for multiple instance learning in histopathology image analysis, Medical Image Analysis 103 (2025) 103561. doi:10.1016/j.media.2025.103561.

[31] W. Alyami, A. Kyme, R. Bourne, Histological validation of mri: A review of challenges in registration of imaging and whole-mount histopathology, Journal of Magnetic Resonance Imaging 55 (2020) 11–22. doi:10.1002/jmri.27409.

[32] P. Mesejo, D. Pizarro, A. Abergel, O. Rouquette, S. Beorchia, L. Poincloux, A. Bartoli, Computer-aided classification of gastrointestinal lesions in regular colonoscopy, IEEE Transactions on Medical Imaging 35 (2016) 2051–2063. doi:10.1109/tmi.2016.2547947.

[33] S. Liu, Z. Chen, Q. Yang, W. Yu, D. Dong, J. Hu, Y. Yuan, Polyp-gen: Realistic and diverse polyp image generation for endoscopic dataset expansion, in: 2025 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2025, pp. 15776–15782. doi:10.1109/icra55743.2025.11128145.

[34] R. Austin, R. Parvathi, Cnn based method for classifying cervical cancer cells in pap smear images, Scientific Reports 15 (2025). doi:10.1038/s41598-025-10009-x.

[35] M. Naseri, J. Fernandez-Marques, Y. Gao, H. Pan, Privacy-preserving federated learning using flower framework, in: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '24, ACM, 2024, pp. 6422–6423. doi:10.1145/3637528.3671447.