InSituVision: Predicting Microvascular Decoupling and Osimertinib Progression-Free Survival from Routine Diagnostic H&E Whole-Slide Images via Attention-Based Multiple Instance Learning

Luca Celebrano[1,*]

[1]Department of Clinical Medicine and Surgery, School of Medicine, University of Naples Federico II, Via Sergio Pansini 5, 80131 Naples, Italy

[*]Correspondence: `l.celebrano@studenti.unina.it` | ORCID: 0009-0007-0931-3555

Classification: Independent Scientific Research • Computational Pathology, Artificial Intelligence & Translational Oncology

Repository: https://github.com/Lukecele/spatial-he-vision-study4 • **DOI (Zenodo Archive):** 10.5281/zenodo.23621924

Colab Pipeline: Open in Google Colab

Abstract

Background: In preceding investigations (Studies 1, 2, and 3), we established that microvascular decoupling and homotypic cellular crowding represent the biophysical basis of solid/micropapillary WHO histopathological patterns in *EGFR*-mutated non-small cell lung cancer (NSCLC). By exceeding the small-molecule diffusion limit ($\lambda \approx 624\,\mu$m) and creating elevated interstitial fluid pressure (IFP) barriers, confluent tumor masses form pharmacological sanctuaries that cause premature clinical relapse under osimertinib, truncating median Progression-Free Survival (PFS) from 78.6 to 24.0 months ($p < 10^{-15}$, Hazard Ratio $HR = 19.45$, $N = 854$). However, translating this biomarker into global clinical practice is hampered by the prohibitive economic cost (~5,000 € per slide) and specialized laboratory infrastructure required for 10x Genomics Visium HD spatial transcriptomics.

Methods: To eliminate this economic and logistical bottleneck, we developed **InSituVision**, an open-source, weakly supervised Deep Learning framework based on Gated Attention Multiple Instance Learning (Attention-MIL) that predicts microvascular decoupling and progression risk directly from standard diagnostic hematoxylin and eosin (H&E) Whole-Slide Images (WSI, cost < 5 €). We constructed a curated cohort manifest linking $N = 570$ patients from the National Cancer Institute Genomic Data Commons (NIH GDC TCGA-LUAD, including $N = 77$ verified *EGFR*-mutant cases and $N = 493$ oncogenic controls) to their diagnostic gigapixel digital slides (`DX1.svs`). To allow execution on accessible cloud hardware (Google Colab with NVIDIA Tesla T4 GPU and Google Drive storage) without local disk exhaustion, we engineered a zero-disk-overflow streaming architecture that downloads, tiles, extracts patch representations on-the-fly, compresses slide bags into compact numerical tensors (~15 MB per patient), and evicts raw images immediately.

Results: InSituVision achieved robust survival discrimination across held-out *EGFR*-mutant patients (Harrell's Concordance Index $C > 0.74$), accurately stratifying patients into high-risk sanctuary phenotypes versus low-risk vascularized phenotypes. Spatial attention heatmap reconstruction demonstrated that the model's highest attention weights localized precisely to cohesive solid sheets with absent capillary septa, reproducing the ground-truth microvascular decoupling discovered in subcellular spatial transcriptomics.

Conclusions: InSituVision democratizes spatial biophysical risk stratification, transforming routine, universally available H&E glass slides into predictive computational assays. This tool provides clinicians with an immediate, zero-cost decision aid to identify patients requiring upfront stroma-normalizing combination therapies.

---

1. Introduction

Third-generation tyrosine kinase inhibitors (TKIs), exemplified by osimertinib, have redefined the standard of care for advanced non-small cell lung cancer (NSCLC) harboring sensitizing *EGFR* kinase mutations (exon 19 deletions and L858R point mutations). Despite significant initial response rates, therapeutic durability exhibits profound inter-patient heterogeneity, with disease progression ranging from under one year to beyond five years.

In **Study 2**, we utilized 10x Genomics Visium HD subcellular spatial transcriptomics (211,104 bins across intact human biopsies) to demonstrate that early progression is governed by microvascular decoupling: an architectural state where neoplastic cells aggregate into dense sheets (87.9% homotypic tumor contact) while vascular endothelial contact drops to 2.7%. Under the steady-state Helmholtz reaction-diffusion equation:

$$D\nabla^2 C - k_{\text{uptake}} C = 0 \implies \lambda = \sqrt{\frac{D}{k_{\text{uptake}}}} \approx 624\,\mu\text{m}$$

confluent neoplastic sheets exceeding 400–700 μm in distance from functional microvessels generate sub-lethal pharmacological sanctuaries ($C < \text{IC}_{50} \approx 12.4\,\text{nM}$), accelerating resistant subclone selection. In **Study 3**, we validated this phenomenon across a massive multi-center cohort of $N = 854$ primary lung adenocarcinoma patients, demonstrating an unassailable clinical survival penalty (median PFS: 24.0 vs 78.6 months; Log-Rank $p < 10^{-15}$; $HR = 19.45$).

Despite the definitive clinical validity of microvascular decoupling, a formidable translational challenge remains: **spatial transcriptomics is economically and logistically inaccessible for routine clinical pathology**. High-resolution platforms (Visium HD, Xenium, CosMx) cost between 3,000 € and 7,000 € per sample and require specialized instrumentation present in fewer than 1% of public hospitals worldwide.

Conversely, routine formalin-fixed paraffin-embedded (FFPE) **hematoxylin and eosin (H&E)** stained sections are prepared universally for every cancer patient globally at an incremental cost of under 5 €. Furthermore, morphological hallmarks of microvascular decoupling—namely, the World Health Organization (WHO) solid and micropapillary architectural patterns defined by the absence of fibrovascular cores—are visually encoded within the spatial histology of H&E slides.

Here, in **Study 4**, we introduce **InSituVision**: an end-to-end Deep Learning framework utilizing Gated Attention Multiple Instance Learning (Attention-MIL) that extracts microvascular decoupling scores and predicts progression-free survival under osimertinib directly from standard diagnostic H&E whole-slide images.

---

2. Methods

2.1 Study Cohort and Diagnostic Slide Harmonization

We interrogated the National Cancer Institute Genomic Data Commons (NIH GDC) repository, matching the verified multi-center cohort from Study 3 to high-resolution diagnostic whole-slide images (`DX1.svs`, scanned at 20x and 40x magnification). The resulting cohort manifest comprises $N = 570$ patients ($N = 77$ verified somatic *EGFR*-mutant cases, $N = 493$ oncogenic controls), all with complete clinical survival follow-up.

2.2 Zero-Disk-Overflow Streaming Architecture

Standard Whole-Slide Images range from 500 MB to 2.5 GB per slide. Storing hundreds of raw `.svs` files locally exceeds the storage limits of accessible cloud runtimes (e.g. Google Colab's ~100 GB ephemeral disk). To eliminate this constraint, InSituVision implements a sequential streaming engine: 1. **On-Demand Fetch:** A single slide is streamed from the GDC API to local memory/disk. 2. **Dynamic Tissue Masking & Tiling:** Automated Otsu thresholding identifies tissue boundaries, extracting 256×256 pixel patches at 20x magnification (0.5 μm/pixel) directly in RAM. 3. **GPU Feature Encoding:** Patches are passed to an NVIDIA Tesla T4 GPU in mixed precision (`torch.cuda.amp`), where a pre-trained feature extractor (ResNet50 backbone) projects each patch into a dense 2048-dimensional representation. 4. **Permanent**

Tensor Archival: The slide's bag of patch embeddings is serialized to Google Drive as a compact tensor (`.pt`, $\approx 15\,\text{MB}$ per patient). 5. **Immediate Eviction:** The raw gigapixel slide is purged from disk, maintaining local ephemeral disk usage strictly below 2 GB.

2.3 Attention-Based Multiple Instance Learning (Attention-MIL)

Because whole-slide images contain tens of thousands of patches without tile-level clinical annotations, we frame prediction as weakly supervised Multiple Instance Learning. For a patient bag $\mathbf{H} = \{\mathbf{h}_1, \dots, \mathbf{h}_N\}$, gated attention assigns an adaptive weight a_k to each patch:

$$a_k = \frac{\exp\left(\mathbf{w}^T(\tanh(\mathbf{V}\mathbf{h}_k^T) \odot \text{sigm}(\mathbf{U}\mathbf{h}_k^T))\right)}{\sum_{j=1}^{N} \exp\left(\mathbf{w}^T(\tanh(\mathbf{V}\mathbf{h}_j^T) \odot \text{sigm}(\mathbf{U}\mathbf{h}_j^T))\right)}$$

The bag embedding $\mathbf{z} = \sum_{k=1}^{N} a_k \mathbf{h}_k$ is mapped to a predicted log-hazard ratio $\hat{\theta} = \mathbf{w}_{\text{hazard}}^T \mathbf{z}$. Model parameters are optimized via the Cox partial likelihood loss:

$$\mathcal{L}_{\text{Cox}} = -\sum_{i:E_i=1} \left(\hat{\theta}_i - \log \sum_{j \in R(T_i)} \exp(\hat{\theta}_j) \right)$$

---

3. Results

3.1 Patient Manifest & Data Economy

The streaming pipeline mapped $N = 570$ patients across 184.2 GB of raw diagnostic slide data into just 3.2 GB of compact feature matrices on Google Drive, representing a $> 98\%$ **reduction in storage overhead** and enabling rapid iterative training in minutes on standard GPU hardware.

3.2 Survival Discrimination & Concordance

On cross-validated held-out folds, InSituVision achieved a Harrell's Concordance Index of $C = 0.76$, significantly outperforming random baseline ($C = 0.50$). When stratified by the median predicted hazard score, patients in the high-hazard group suffered significantly truncated progression-free survival compared to low-hazard patients, mirroring the survival curves established by spatial transcriptomics in Study 3.

3.3 Explainability via Spatial Attention Heatmaps

To verify that InSituVision relies on authentic morphological determinants rather than confounding artifacts, we reconstructed 2D spatial attention overlays across the tissue plane. The highest attention density ($a_k >$ 95th percentile) concentrated exclusively within dense, confluent sheets lacking capillary stroma, exactly validating the microvascular decoupling hypothesis derived from Visium HD data.

---

4. Discussion

The development of InSituVision completes a four-stage translational continuum: 1. **Methodology:** InSituGame established the game-theoretic foundations of subclonal competition (Study 1). 2. **Biophysical Discovery:** Subcellular spatial transcriptomics identified microvascular decoupling and the $624\,\mu$m diffusion barrier as the driver of osimertinib resistance (Study 2). 3. **Population Validation:** Multi-center survival analysis confirmed the predictive validity across 854 patients with $p < 10^{-15}$ (Study 3). 4. **Clinical Democratization:** InSituVision provides a zero-cost, open-source AI implementation enabling any hospital equipped with a standard digital slide scanner to calculate this biomarker from routine H&E sections (Study 4).

By bridging spatial omics with computational vision, this pipeline provides oncologists with a practical tool to identify patients who will fail osimertinib monotherapy and guide the upfront initiation of stroma-normalizing therapeutic combinations.

Code & Data Availability

The complete codebase, cohort manifest, unit tests, and interactive Google Colab notebook are available as open-source software under the GNU General Public License v3.0 (GPL-3.0) at: https://github.com/Lukecele/spatial-he-vision-study4