Inferring CAF-1/p60 Expression from Hematoxylin and Eosin Stained Images in Oral Squamous Cell Carcinoma

1st Cristian Tommasino
Department of Electrical Engineering and Information Technology
University of Naples Federico II
Naples, Italy
cristian.tommasino@unina.it

2nd Cristiano Russo
Department of Electrical Engineering and Information Technology
University of Naples Federico II
Naples, Italy
cristiano.russo@unina.it

3rd Angela Crispino
Department of Advanced Biomedical Sciences, Pathology Unit
University of Naples Federico II
Naples, Italy
angela.crispino@unina.it

4th Stefania Staibano
Department of Advanced Biomedical Sciences, Pathology Unit
University of Naples Federico II
Naples, Italy
stefania.staibano@unina.it

5th Antonio M. Rinaldi*
Department of Electrical Engineering and Information Technology
University of Naples Federico II
Naples, Italy
antoniomaria.rinaldi@unina.it

6th Francesco Merolla*
Department of Medicine and Health Sciences "V. Tiberio"
University of Molise
Campobasso, Italy
francesco.merolla@unimol.it

***Abstract*—Hematoxylin and Eosin (H&E) staining remains a standard of histopathological diagnostics due to its efficiency, affordability, and ubiquity. However, its lack of molecular specificity often necessitates complementary immunohistochemistry (IHC) to detect protein-level biomarkers critical for diagnosis and treatment planning. In scenarios where IHC is impractical, due to cost, limited tissue, or logistical constraints, computational alternatives may provide viable solutions. In this study, we investigate the feasibility of predicting CAF-1/p60 protein expression directly from H&E-stained slides using deep learning. Leveraging a curated dataset of oral squamous cell carcinoma samples with nucleus-level annotations, we classify nuclei into three categories: (i) Tumor-CAF-1/p60-positive, (ii) Tumor-CAF-1/p60-negative, and (iii) stromal. Our results demonstrate that Convolutional Neural Networks (CNNs) can learn morphological cues indicative of protein expression, achieving promising classification performance at the single-cell level. This approach may enable more accessible molecular profiling from routine H&E slides, reducing dependency on specialized staining procedures in resource-constrained or time-sensitive settings.**

***Index Terms*—Classification, Caf-1/p6, Hematoxylin and Eosin, Computational pathology, Molecular marker inference, Nuclei classification**

I. Introduction

Hematoxylin and Eosin (H&E) staining remains the histopathological gold standard for visualizing tissue architecture and cellular morphology [1]. Previous studies have demonstrated the effectiveness of deep learning and feature representation techniques for histopathology image analysis, particularly in the context of content-based image retrieval (CBIR) and stain normalization strategies [2], [3]. Moreover, graph-based models have been successfully employed to enhance semantic retrieval and representation across biomedical datasets [4], motivating the integration of both local morphology and global context in single-cell analysis. However, it lacks molecular specificity, making immunohistochemistry (IHC) indispensable for biomarker detection that is yet expensive, time-consuming, and often impractical in low-resource clinical settings [5], [6].

Recently, deep learning (DL) methods have shown that routine H&E whole-slide images (WSIs) can predict molecular biomarkers, such as protein expression, gene mutations, and molecular subtypes [7]–[9]. For instance, pan-cancer weakly supervised models trained on over 12,000 WSIs achieved AUC up to 0.83 for PD-L1 expression prediction [10]. Moreover, hist2RNA demonstrated strong correlation for gene-expression prediction in breast cancer from H&E [9]. Contrastive and cross-modality learning frameworks (e.g., HECLIP [11] and Cross-Modality Learning [12]) further improved imputation of transcriptomic profiles. Despite these successes, single-cell resolution inference from H&E remains largely unexplored.

A promising target is CAF-1/p60, a chromatin assembly factor subunit associated with aggressive oral squamous cell carcinoma (OSCC) and poor patient prognosis [13]. Prior studies using registered H&E–IHC images and segmentation (e.g., StarDist) reported effective nucleus-level detection of CAF-1/p60 [10], [14], but these approaches still rely on IHC at inference.

We suppose that DL models can classify CAF-1/p60 positivity at the nuclear level using only H&E, bypassing IHC during

This study was partially supported by the PNRR MUR project PE0000013-FAIR. This work also was partially supported by the PNRR MUR project PE00000001-RESTART.

*Authors jointly supervised this work.

inference. To test this, we curated a nucleus-level dataset of paired H&E–IHC OSCC slides annotated for CAF-1/p60 positivity. Leveraging advanced segmentation tools (CellViT, StarDist, InstantSeg, NuLite, HoVer-UNet) [15]–[20], our pipeline focuses solely on classification based on morphological features.

Our work makes three key contributions. First, we assembled a nucleus-level dataset consisting of paired H&E and IHC images, meticulously annotated to identify CAF-1/p60–positive nuclei. Second, we developed a deep-learning classification pipeline that predicts protein expression solely from morphological features captured in H&E-stained images. Finally, we provide a quantitative evaluation of our approach, demonstrating that deep-learning models can accurately detect CAF-1/p60 positivity at the single-cell level.

Our framework achieves three significant advancements in computational pathology. First, it operates at nucleus-level granularity, enabling inference with a precision unattainable by patch- or slide-level models, an important step forward in single-cell histological analysis. Second, it benefits from strong supervision, relying on IHC-derived ground truth to train and validate predictions of molecular expression, thereby enhancing both accuracy and trustworthiness. Finally, the approach ensures clinical feasibility, as it is entirely compatible with routine H&E-only workflows and requires no additional staining during deployment, making it suitable for integration into existing diagnostic pipelines.

The remainder of this paper is organized as follows. Section II describes dataset creation and model architecture. Section III presents experimental results. Section IV discusses clinical relevance and limitations. Section V concludes and explores future work.

II. Related Work

Deep learning (DL) has made significant strides in predicting molecular biomarkers from routine H&E whole-slide images (WSIs), typically operating at the patch or slide level. A notable example is Cheng et al.'s pan-cancer approach, which used weakly supervised, teacher–student multiple-instance learning on a dataset of 12,299 WSIs. It reached an excellentr results for PD-L1 expression prediction across 20 tumor types, validating the broad potential of H&E-based biomarker modeling [10], [21]. In a different domain, the hist2RNA model demonstrated strong correlation between predicted gene-expression values and ground-truth RNA-seq data from breast cancer WSIs, revealing the feasibility of inferring transcriptomic information from H&E morphology [9], [14]. Moreover, recent advances in contrastive and cross-modality learning, such as HECLIP and related frameworks, have improved the accuracy of transcriptome imputation from H&E images, outperforming earlier methods [11], [12].

Despite these advances, these approaches focus on low-resolution input and output (i.e., patches or entire slides), which limits their granularity and hinders the ability to derive insights at the level of individual cells. In parallel, the field of nuclei segmentation has rapidly evolved, introducing several high-performance models, such as CellViT, StarDist, InstantSeg, and NuLite, that enable precise, cell-level delineation in H&E images [15]–[19]. However, these tools primarily facilitate morphological analyses rather than biomarker inference.

A small but growing number of studies has begun to bridge this gap, especially by leveraging paired H&E–IHC slides. For example, Varricchio et al. and Crispino et al. employed StarDist-based pipelines on oral squamous cell carcinoma (OSCC) samples to detect CAF-1/p60–positive nuclei, highlighting the potential of morphological cues in H&E when supervised by IHC labels [10], [14]. Although promising, these approaches still require IHC to generate labels for inference, which restricts their applicability in real-world clinical settings.

In contrast, our work targets single-cell resolution inference of CAF-1/p60 expression from H&E images alone at inference time, without relying on IHC. By curating a high-quality, nucleus-level dataset annotated via IHC and combining it with state-of-the-art segmentation and classification pipelines, we propose a fully H&E-based method that achieves single-cell molecular inference in OSCC, a novel and clinically feasible contribution to computational pathology.

III. Methods

We assembled our dataset using nine whole-slide images (WSIs) and a tissue microarray (TMA) comprising 55 oral squamous cell carcinoma (OSCC) cores archived by the Pathology Unit at the University of Naples "Federico II," following procedures published in [10], [14]. All WSI and TMA samples underwent standard Hematoxylin and Eosin (H&E) staining and were digitized at 40× magnification ($0.25\mu m/pixel$) with a Leica Aperio AT2 scanner.

To label CAF-1/p60 expression, H&E-stained slides were destained, rehydrated, and re-stained with an anti-CAF-1/p60 antibody using a Ventana Benchmark Ultra system. These IHC slides were scanned under identical imaging conditions. Positivity was determined based on DAB chromogen intensity.

We then registered H&E and IHC images to align corresponding tissue regions. Within each aligned tile, nuclei were detected using optical density thresholding on the DAB channel. Two pixel-level masks were produced per tile: one delineating all nuclei and another classifying nuclei into three categories, CAF-1/p60–positive, CAF-1/p60–negative tumor cells, and stromal cells. Expert pathologists identified regions of interest (ROIs) from which we extracted individual nuclei, resulting in a dataset of 46,251 samples: 3,186 CAF-1/p60–positive, 12,312 CAF-1/p60–negative, and 30,753 stromal nuclei (see Figure 1 for a representative sample).

To perform classification, we fine-tuned two CNN backbones, ResNet and EfficientNet, pre-trained on ImageNet. These architectures were selected based on their demonstrated effectiveness in capturing local morphological patterns relevant to histopathology [22], [23]. Model input consisted of

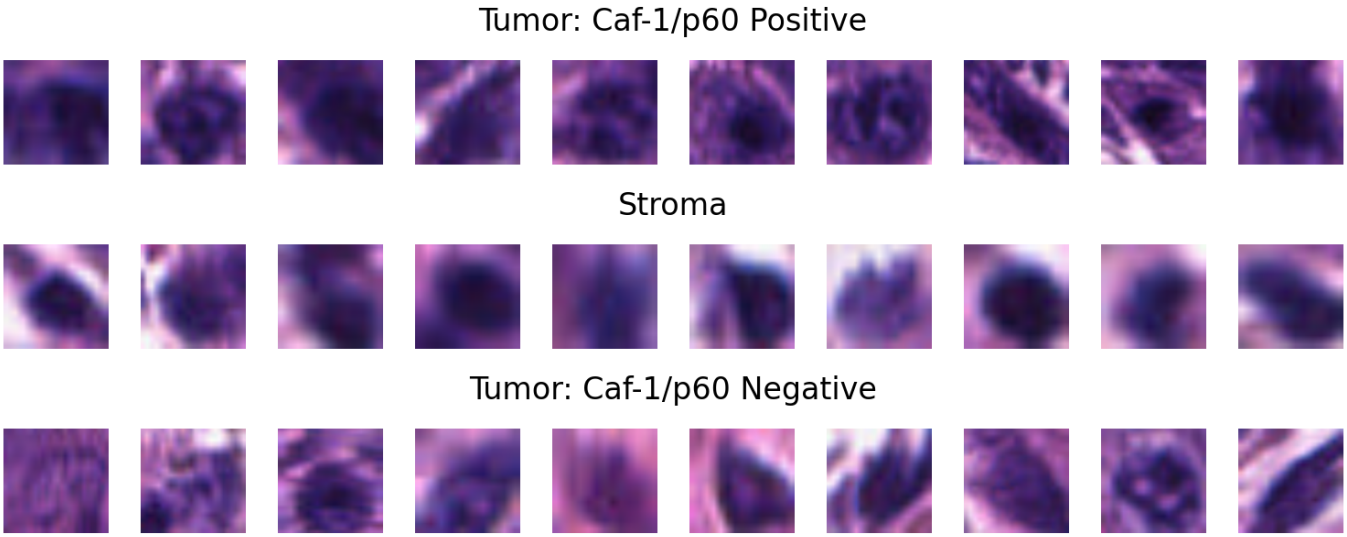

Fig. 1. Example nuclei patches extracted from H&E–IHC–paired tiles in OSCC.

128×128 pixel nuclei patches, normalized using standard H&E color deconvolution techniques.

Given the pronounced class imbalance, particularly for the CAF-1/p60–positive class, we implemented a weighted random sampler during training. The sampling weight assigned to each nucleus i follows:

$$w_{nuclei}(i,\lambda) = \frac{N_{\text{train}}}{\lambda\left(\sum_{j=1}^{N_{\text{train}}} \mathbf{1}(c_j = c_i)\right) + (1-\lambda)N_{\text{train}}} \quad (1)$$

where N_{train} is the total number of training samples, $\sum_{j \in C_{c_i}} 1$ counts samples in class c_i, and $\lambda \in [0,1]$ controls oversampling intensity. Setting $\lambda = 1$ results in full class-balance sampling, while $\lambda = 0$ produces uniform sampling.

During training, we used a batch size of 64 and Adam optimizer with an initial learning rate of 1e-4, decayed by a factor of 0.1 every 10 epochs. Early stopping on a balanced validation loss prevented overfitting. We evaluated performance using per-class accuracy, precision, recall, and macro-averaged F1 and AUC scores.

In addition to sampling, we also experimented with focal loss [24], and data augmentation techniques including random rotations, flips, and color jitter. We chose the configuration that achieved the highest macro-F1 across five-fold cross-validation for final comparison.

This methodology ensures robust nucleus-level classification, handling data imbalance carefully to preserve minority signal detection while avoiding bias and overfitting.

IV. Experimental Results

We evaluated our classification models by computing macro-averaged and per-class Precision, Recall, and F1-score, and we additionally assessed the Area Under the ROC Curve (AUC) specifically for the CAF-1/p60–positive class to better characterize its detection capabilities. Training was performed with the Adam optimizer (initial learning rate 1×10^{-3}, weight decay 1×10^{-4}), using an exponential learning rate scheduler ($\gamma = 0.9$), and employing a weighted random sampler with $\lambda = 0.8$ (see Equation 1).

Table I summarizes the performance achieved by five architectures: ResNet18, ResNet34, ResNet50, EfficientNet-V2-S, and EfficientNet-V2-M. All architectures reached similar macro F1-scores, around 0.70–0.71, demonstrating overall robustness across the dataset. However, substantial variability emerged when examining the CAF-1/p60–positive class, which remains the rarest and most challenging to detect.

EfficientNet-V2-S and ResNet50 achieved the highest precision (0.62) for CAF-1/p60 positivity, indicating strong specificity. However, their recall scores (0.53 and 0.51, respectively) remained modest, resulting in mid-0.50s F1-scores, a pattern suggesting a conservative threshold favoring precision over sensitivity. In contrast, EfficientNet-V2-M prioritized recall (0.61) over precision (0.53), offering a different balance and achieving a similarly moderate F1 (0.57). These results underscore the difficulty of detecting rare biomarkers based on subtle morphological patterns and highlight how architectural choices can influence the precision–recall trade-off.

In comparison, the stromal class demonstrated outstanding performance across all tested models, consistently achieving F1-scores of 0.87–0.88. This likely reflects the clear and abundant morphological cues available in stromal nuclei. The CAF-1/p60–negative class exhibited stable yet moderate performance (F1 0.65–0.67), likely due to its heterogeneity and partial overlap with both positive and stromal nuclei.

As shown in Figure 2, all models achieve comparable macro-level performance (F1 $\approx$ 0.70), but exhibit a marked drop when specifically evaluated on CAF-1/p60-positive nuclei, underscoring the difficulty of this rare-class prediction. Figure 3 further elucidates this trade-off: architectures like EfficientNet-V2-S cluster towards higher precision, while

EfficientNet-V2-M shifts towards higher recall—highlighting the importance of selecting the appropriate backbone depending on whether minimizing false positives or false negatives is prioritized, a commonly observed phenomenon in other histopathology studies.

We also evaluated discrimination power using AUC: EfficientNet-V2-S and V2-M achieved AUCs of 0.74 and 0.75, respectively. These results indicate acceptable classification performance considering the low prevalence and challenging visual similarity of the target class. The use of EfficientNet architectures aligns with recent findings that they combine high discriminative ability with computational efficiency in histopathology tasks.

Overall, our results provide critical insights into nucleus-level classification performance. First, CNN architectures, particularly EfficientNet-V2 variants—demonstrate adaptable trade-offs between precision and recall depending on the specific task requirements. Second, accurately identifying rare biomarkers such as CAF-1/p60 at the single-cell level remains challenging, primarily due to class imbalance and overlapping morphological features. Third, the consistently strong performance observed across CAF-1/p60-negative and stromal nuclei attests to the robustness of our segmentation and patch-extraction pipeline.

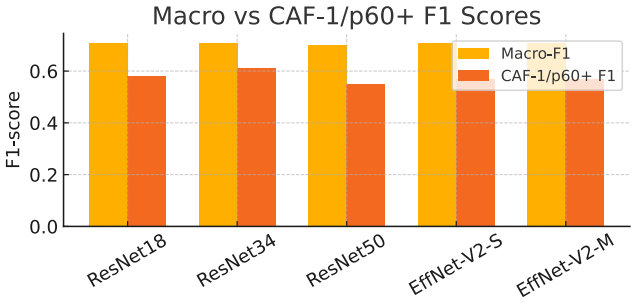

Fig. 2. Comparison of Macro-F1 and CAF-1/p60+ F1 scores across different CNN architectures.

V. Conclusion

This study demonstrates the feasibility of classifying CAF-1/p60–positive nuclei using only H&E-stained images, thereby eliminating reliance on immunohistochemistry (IHC) in specific diagnostic scenarios. Leveraging our custom-built nucleus-level dataset and a suite of convolutional neural networks (CNNs), we show that molecular markers can be inferred from purely morphological cues with encouraging performance metrics.

Our experimental results suggest that computational pathology workflows can be simplified, reducing dependency on costly and time-consuming IHC protocols, without compromising diagnostic accuracy for key molecular markers. This has important implications for high-throughput clinical settings and low-resource environments.

Looking ahead, we plan to pursue several enhancements aimed at increasing both model performance and clinical

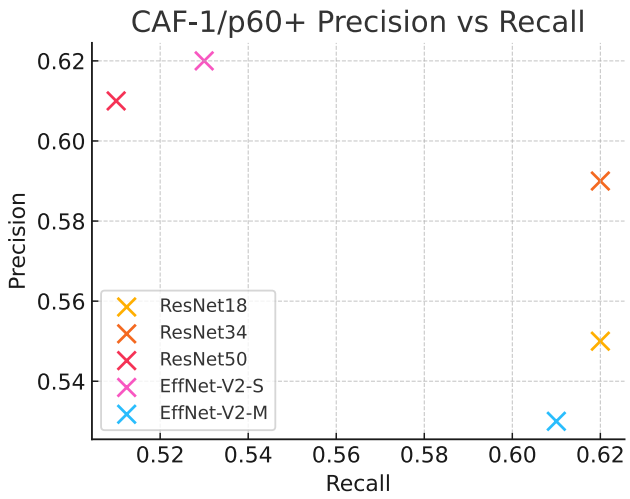

Fig. 3. Precision vs. Recall for CAF-1/p60+ nuclei classification, highlighting the trade-off for each model.

applicability. First, we will investigate graph neural networks (GNNs) to capture spatial and contextual relationships among nuclei, which may improve classification of ambiguous morphology. Second, we intend to develop an end-to-end deep learning framework that integrates nuclei instance segmentation with joint CAF-1/p60 classification to streamline the inference pipeline and reduce cumulative error. Finally, expanding the dataset with additional WSIs and diverse OSCC cases will strengthen generalizability and robustness. Together, these efforts aim to advance towards scalable, H&E-only diagnostic tools capable of reliable molecular inference, supporting more accessible, cost-effective, and automated pathology workflows.

Acknowledgment

References

[1] Andrew H Fischer, Kenneth A Jacobson, Jack Rose, and Rolf Zeller, "Hematoxylin and eosin staining of tissue and cell sections," *Cold spring harbor protocols*, vol. 2008, no. 5, pp. pdb–prot4986, 2008.

[2] Cristian Tommasino, Francesco Merolla, Cristiano Russo, Stefania Staibano, and Antonio Maria Rinaldi, "Histopathological image deep feature representation for cbir in smart pacs," *Journal of Digital Imaging*, vol. 36, no. 5, pp. 2194–2209, 2023.

[3] Antonio M Rinaldi, Cristiano Russo, and Cristian Tommasino, "Effects of color stain normalization in histopathology image retrieval using deep learning," in *2022 IEEE International Symposium on Multimedia (ISM)*. IEEE, 2022, pp. 26–33.

[4] Giovanni M De Filippis, Antonio M Rinaldi, Cristiano Russo, and Cristian Tommasino, "Enhanced semantic understanding with graph-based information retrieval," in *International Workshop on Graph-Based Approaches in Information Retrieval*. Springer, 2024, pp. 11–24.

[5] Minh-Khang Le, Naoki Oishi, Kunio Mochizuki, and Tetsuo Kondo, "Immunohistochemical detection of cancer genetic abnormalities," *Pathology-Research and Practice*, p. 155109, 2024.

[6] Dinku Yigzaw and Gashaw Getaneh Dagnaw, "Review of immunohistochemistry techniques: Applications, current status, and future perspectives," in *Seminars in Diagnostic Pathology*. Elsevier, 2024.

[7] Heather D Couture, "Deep learning-based prediction of molecular tumor biomarkers from h&e: A practical review," *Journal of Personalized Medicine*, vol. 12, no. 12, pp. 2022, 2022.

TABLE I

PERFORMANCE COMPARISON OF CNN ARCHITECTURES FOR CAF-1/P60 POSITIVE NUCLEUS CLASSIFICATION. THIS TABLE PRESENTS THE MACRO-AVERAGE AND CLASS-SPECIFIC PRECISION, RECALL, AND F1 SCORES FOR FIVE CNN ARCHITECTURES, RESNET18, RESNET34, RESNET50, EFFICIENTNET-V2-S, AND EFFICIENTNET-V2-M.

Model	Macro			p60 - Pos			Stroma			p60 - Neg		
	P	R	$F1$	P	R	$F1$	P	R	$F1$	P	R	$F1$
ResNet18	0.70	0.71	0.71	0.55	0.62	0.58	0.89	0.87	0.88	0.67	0.66	0.66
ResNet34	0.71	0.71	0.71	0.59	0.62	0.61	0.89	0.87	0.88	0.65	0.63	0.64
ResNet50	0.71	0.68	0.70	0.61	0.51	0.55	0.89	0.85	0.87	0.62	0.69	0.66
EfficientNet-V2-S	0.71	0.71	0.71	0.62	0.53	0.57	0.9	0.84	0.87	0.62	0.73	0.67
EfficientNet-V2-M	0.70	0.71	0.71	0.53	0.61	0.57	0.89	0.88	0.88	0.69	0.65	0.67

[8] Jakob Nikolas Kather, Alexander T Pearson, Niels Halama, Dirk Jäger, Jeremias Krause, Sven H Loosen, Alexander Marx, Peter Boor, Frank Tacke, Ulf Peter Neumann, et al., "Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer," *Nature medicine*, vol. 25, no. 7, pp. 1054–1056, 2019.

[9] Raktim K. Mondol, Ewan K. A. Millar, Peter H. Graham, Lois Browne, Arcot Sowmya, and Erik Meijering, "hist2rna: An efficient deep learning architecture to predict gene expression from breast cancer histopathology images," *Cancers*, vol. 15, no. 9, pp. 2569, 2023.

[10] Silvia Varricchio, Gennaro Ilardi, Daniela Russo, Rosa Maria Di Crescenzo, Angela Crispino, Stefania Staibano, and Francesco Merolla, "Leveraging deep learning for identification and segmentation of "caf-1/p60-positive" nuclei in oral squamous cell carcinoma tissue samples," *Journal of Pathology Informatics*, vol. 15, pp. 100407, 2024.

[11] Qing Wang, Wen-jie Chen, Jing Su, Guangyu Wang, and Qianqian Song, "Heclip: histology-enhanced contrastive learning for imputation of transcriptomics profiles," *Bioinformatics*, p. btaf363, 2025.

[12] A. Das, N. Tomita, K. J. Syme, W. Ma, P. O'Connor, K. N. Corbett, B. Ren, X. Liu, and S. Hassanpour, "Cross-modality learning for predicting ihc biomarkers from h&e-stained whole-slide images," *arXiv*, vol. arXiv:2506.15853, 2025.

[13] Andrew Volk and John D Crispino, "The role of the chromatin assembly complex (caf-1) and its p60 subunit (chaf1b) in homeostasis and disease," *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1849, no. 8, pp. 979–986, 2015.

[14] Angela Crispino, Silvia Varricchio, Daniela Russo, Rosa Maria Di Crescenzo, Stefania Staibano, and Francesco Merolla, "A digital workflow for automated assessment of tumor-infiltrating lymphocytes in oral squamous cell carcinoma using qupath and a stardist-based model," *Pathologica-Journal of the Italian Society of Anatomic Pathology and Diagnostic Cytopathology*, vol. 116, 2024.

[15] Fabian Hörst, Moritz Rempe, Lukas Heine, Constantin Seibold, Julius Keyl, Giulia Baldini, Selma Ugurel, Jens Siveke, Barbara Grünwald, Jan Egger, et al., "Cellvit: Vision transformers for precise cell segmentation and classification," *Medical Image Analysis*, vol. 94, pp. 103143, 2024.

[16] Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers, "Cell detection with star-convex polygons," in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II*, 2018, pp. 265–273.

[17] Martin Weigert, Uwe Schmidt, Robert Haase, Ko Sugawara, and Gene Myers, "Star-convex polyhedra for 3d object detection and segmentation in microscopy," in *The IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2020.

[18] Thibaut Goldsborough, Ben Philps, Alan O'Callaghan, Fiona Inglis, Leo Leplat, Andrew Filby, Hakan Bilen, and Peter Bankhead, "Instanseg: an embedding-based instance segmentation algorithm optimized for accurate, efficient and portable cell segmentation," *arXiv preprint arXiv:2408.15954*, 2024.

[19] Cristian Tommasino, Cristiano Russo, and Antonio Maria Rinaldi, "Nulite–lightweight and fast model for nuclei instance segmentation and classification," *arXiv preprint arXiv:2408.01797*, 2024.

[20] Cristian Tommasino, Cristiano Russo, Antonio Maria Rinaldi, and Francesco Ciompi, "" hover-unet": Accelerating hovernet with unet-based multi-class nuclei segmentation via knowledge distillation," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2024, pp. 1–4.

[21] Y. Cheng, X. Zhan, L. Hou, X. Luo, Z. Feng, Z. Jiang, K. Yu, Y. Gu, T. Fan, J. Yao, et al., "A weakly supervised deep learning model for pan-cancer prediction of pd-l1 expression from histopathology slides," *Nature Communications*, vol. 15, no. 1, pp. 5097, 2024.

[22] Mingxing Tan and Quoc Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[24] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.