**PATHOLOGICA 2024;**116:390-403;
DOI: 10.32074/1591-951X-1069

Original article

# A Digital Workflow for Automated Assessment of Tumor-Infiltrating Lymphocytes in Oral Squamous Cell Carcinoma Using QuPath and a StarDist-Based Model

Angela Crispino[1*], Silvia Varricchio[1*], Gennaro Ilardi[1], Daniela Russo[1], Rosa Maria Di Crescenzo[1], Stefania Staibano[1**], Francesco Merolla[2**]

[1] *Department of Advanced Biomedical Sciences, Pathology Unit, University of Naples "Federico II", Naples, Italy;*
[2] *Department of Medicine and Health Sciences "V. Tiberio", University of Molise, Campobasso, Italy*
**These authors equally contributed to this work*
***co-senior authors*

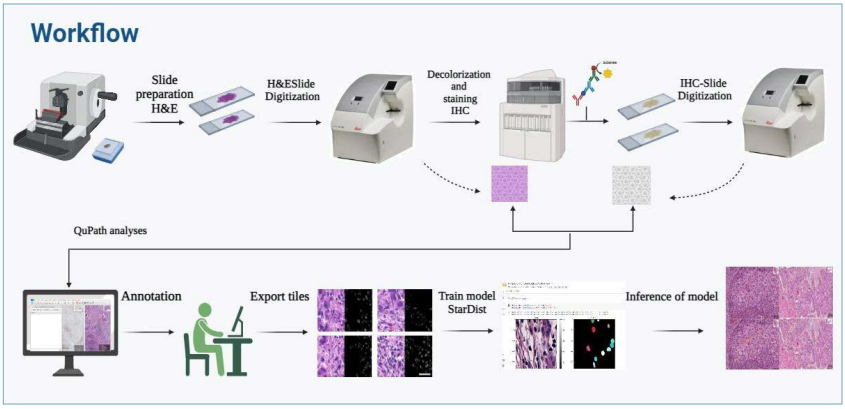
## Summary

The search for reliable prognostic markers in oral squamous cell carcinoma (OSCC) remains a critical need. Tumor-infiltrating lymphocytes (TILs), particularly T lymphocytes, play a pivotal role in the immune response against tumors and are strongly correlated with favorable prognoses. Computational pathology has proven highly effective for histopathological image analysis, automating tasks such as cell detection, classification, and segmentation.

In the present study, we developed a StarDist-based model to automatically detect T lymphocytes in hematoxylin and eosin (H&E)-stained whole-slide images (WSIs) of OSCC, bypassing the need for traditional immunohistochemistry (IHC). Using QuPath, we generated training datasets from annotated slides, employing IHC as the ground truth. Our model was validated on Cancer Genome Atlas-derived OSCC images, and survival analyses demonstrated that higher TIL densities correlated with improved patient outcomes.

This work introduces an efficient, AI-powered workflow for automated immune profiling in OSCC, offering a reproducible and scalable approach for diagnostic and prognostic applications.

**Key words:** OSCC**,** T lymphocyte cells, IA, Digital Pathology, Computational Pathology, deep learning, QuPath, StarDist




**Graphical Abstract.** The proposed workflow: from tissue slide preparation to model inference. (Created with BioRender.com)

Received: August 12, 2024
Accepted: December 10, 2024

**Correspondence**
Francesco Merolla
E-mail: francesco.merolla@unimol.it

**How to cite this article:** Crispino A, Varricchio S, Ilardi G, et al. A Digital Workflow for Automated Assessment of Tumor-Infiltrating Lymphocytes in Oral Squamous Cell Carcinoma Using QuPath and a StarDist-Based ModelPathologica 2024;116:390-403. https://doi.org/10.32074/1591-951X-1085

## Introduction

Oral squamous cell carcinoma (OSCC), while a relatively rare neoplasm, poses significant clinical challenges due to its aggressive nature, low response rates to standard therapies, and high levels of drug resistance [1-3]. Several studies have emphasized the tumor microenvironment's (TME) pivotal role in regulating tumor progression and therapeutic responses. Interactions between tumor cells and components of the TME, particularly immune cells, are critical in shaping these processes [4,5]. In this context, innate immunity acts as the first line of defense, followed by adaptive immunity, where lymphocytes play an active role. In OSCC, T lymphocytes influence tumor behavior, therapeutic outcomes, and patient survival [6,7].

Given their central role in antitumor responses, the accurate detection and quantification of tumor-infiltrating lymphocytes (TILs) is crucial. The presence of TILs is strongly associated with improved prognosis and enhanced responses to immunotherapies [8-11]. Traditionally, T-cell identification relies on immunohistochemistry (IHC) using specific antibodies to detect relevant lymphocyte subsets. However, this approach is prone to significant inter- and intra-observer variability, and manual quantification can be labor-intensive and time-consuming [12].

Integrating artificial intelligence (AI) offers a promising solution to these challenges. Advances in digital pathology, supported by whole-slide imaging (WSI), have enabled cell detection and quantification automation. Deep learning techniques, particularly convolutional neural networks (CNNs), have shown exceptional accuracy in analyzing hematoxylin and eosin (H&E)-stained slides. These methods provide a viable alternative to IHC, delivering reproducible and standardized metrics while reducing manual effort and observer bias [13,14].

Our study aims to develop an AI-powered tool for the automated segmentation and detection of lymphocytic T cells in WSI of OSCC stained with H&E. By streamlining the diagnostic workflow, this approach seeks to improve diagnostic efficiency and precision. Leveraging AI and WSI technology, we aim to enhance the clinical decision-making process, enabling pathologists to perform immune profiling more accurately and efficiently. Furthermore, our work builds upon a previously established strategy for predicting CAF-1/p60-positive cells in OSCC WSI-H&E tumor tissue [15]

Our approach to computational pathology emphasizes creating user-friendly diagnostic tools that aid in risk stratification and support more personalized treatment planning for OSCC patients.

## Materials and methods

### Slides preparation

We selected 10 OSCC formalin-fixed paraffin embedded (FFPE) cases from the archive of the Pathology Unit of the Department of Advanced Biomedical Sciences of the University of Naples "Federico II." To ensure precise detection of T cells, we performed immunohistoochemical (IHC) staining for the T cell marker CD3 to build a robust ground truth. We restained previously H&E-stained tissue sections with CD3-IHC to guarantee more accurate segmentation masks, as described in [15].

Briefly, H&E stained and digitized tissue sections were destained and immunostained with anti-CD3 antibody. We digitized H&E and IHC slides using a Leica Aperio AT2 scanner at 40x magnification. We performed immunostaining with the Ventana Benchmark Ultra automatic stainer (Ventana Medical Systems Inc., 1910 Innovation patk drive Tucson, AZ, USA) using the rabbit monoclonal anti-CD3 antibody (2GV6), according to the manufacturer's instructions.

Before digitizing, we cleaned the slides with sterile gauze and 90% ethanol to eliminate any impurities that might affect our analysis.

### Quality check: HistoQC

We utilized HistoQC, an open-source quality control (QC) tool for digital pathology slides, to automate the quality assessment of whole-slide images (WSIs) [16,17]. HistoQC efficiently detects and classifies artifacts while calculating visual metrics associated with the images. Its modular pipeline allows for the analysis of image properties such as color, brightness, and contrast, generating visual masks and metadata to identify areas of interest and anomalies within the datasets. The tool produces masks that can be overlaid on the original images, with specific colors denoting regions of interest:

- **Pink masks**: Indicate artifact-free, computable areas suitable for AI model training and inference.
- **Green masks**: Highlight regions compromised by significant artifacts excluded from further analysis.
- For our study, we exclusively used the pink-masked regions of interest (ROIs) in all subsequent analyses.

### Dataset generation

After quality control, we generated the training and validation sets with QuPath (v. 0.5.1 version) [18] We created a project and inserted our H&E-stained WSIs and the paired CD3-stained ones. We registered the paired images with an affine registration by using a

QuPath extension (https://github.com/qupath/qupath-extension-align. Last accessed 06/08/2024) (Fig. 1).
Once we aligned H&E with the paired IHC, we performed a positive cell detection by the mean of DAB OD features. Then, we exported the entire set of detected objects from IHC in geojson format and re-imported them onto the H&E images. We obtained CD3-positive segmentation masks that perfectly overlapped the H&E-stained cells (Fig. 2).
Two pathologists visually examined the samples to validate the annotated images (SS and FM). After this check, we used a Tile Exporter groovy script to export 256x256px tiles and the corresponding label images. We modified the Groovy script on the QuPath documentation web page ("https://qupath.readthedocs.io/en/stable/," last accessed: 10/04/2024). The full script edit is available in **Supplementary I.**

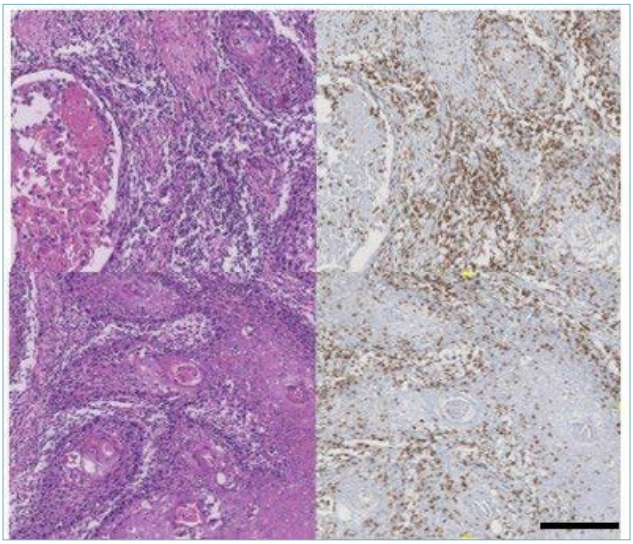


**Figure 1.** Representative example of the aligned image pairs showing both the H&E images and the corresponding IHC images (Scale bar: 50 pixels).

### Model training

We built the StarDist-based model as described [19-21] ("https://github.com/stardist/stardist"; last access: 04.03.2024).
StarDist is a Python package implementing star-convex object detection for 2D and 3D images. It is a deep learning-based method for nuclei/cell detection and segmentation in microscopy images. The general approach for 2D images involves training a model to densely predict the distances to the object boundary along a fixed set of rays and object probabilities. This produces an overcomplete set of candidate polygons for a given input image. The final result is obtained via non-maximum suppression (NMS) of these candidates. In particular, it recognizes round objects such as nuclei or cells through the "star-convex polygons" system to represent their edges. Each pixel is associated with the star-convex polygon defined by the distances from the pixel to the object's edge along the various predefined radial directions.
The training was conducted on a Google Colab virtual machine with a V100 GPU with 16 GB of VRAM and 51 GB of RAM. All coding and statistical analyses were performed using Python 3.10.
Our model was trained using 43908 objects (i.e., CD3 positive cells) in 7750 tiles of 256x256 pixels. We split the dataset into train/validation/test using a cross-vali-

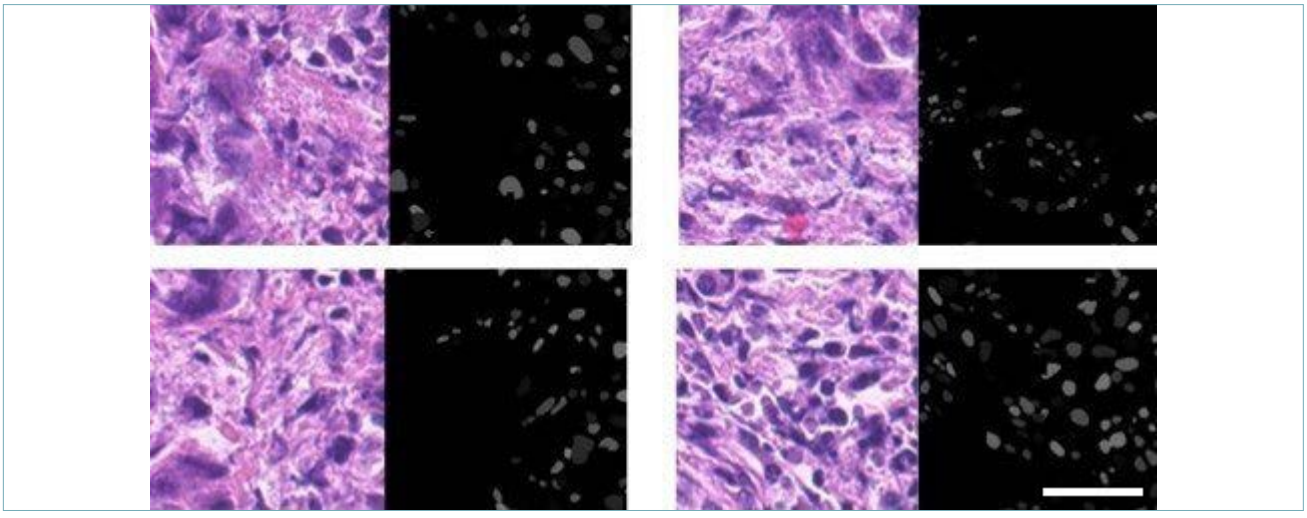


**Figure 2.** An example of H&E tiles and masks exported from QuPath after annotation alignment (Scale bar: 50 pixels).

**Table I.** Versions of specific packages used for model training.

| Package | Version |
|---|---|
| StarDist | (0.9.1) |
| csbdeep | (0.8.0) |
| scikit-image | (0.19.3) |
| numpy | (1.25.2) |
| scipy | (1.11.4) |
| matplotlib | (3.7.1) |
| tifffile | (2024.6.18) |
| tqdm | (4.66.4) |
| gputools | 0. 2. 14 |

dation method. All the details of the training are shown in **Supplementary VI**. The full code used to train the model is available upon request at: https://github.com/Frank19782023/paperCD3OSCC

### Model Testing

We tested our model to detect and segment $CD3^+$ T cells in H&E-stained WSI on test OSCC samples stained with H&E and CD3. We then validated our model visually in QuPath, comparing the prediction results with the IHC.

Figure 3 (B and D) shows the model-based prediction on H&E, with the recognized lymphocytes highlighted in yellow and validated by IHC (A and C).

### Model Inference with QuPath

Once the model was built, it was converted into portable formats such as Protocol Buffers (.pb) format and ImageJ-compatible zip files. This conversion was performed in an Anaconda virtual environment using TensorFlow 1.15 and tfonnx to ensure compatibility and ease of use across various platforms.

### Dataset for model inference

We used a dataset derived from the Cancer Genome Atlas (TCGA) to infer our model, precisely 50 cases of OSCC stained with hematoxylin and eosin (Tab. II.).

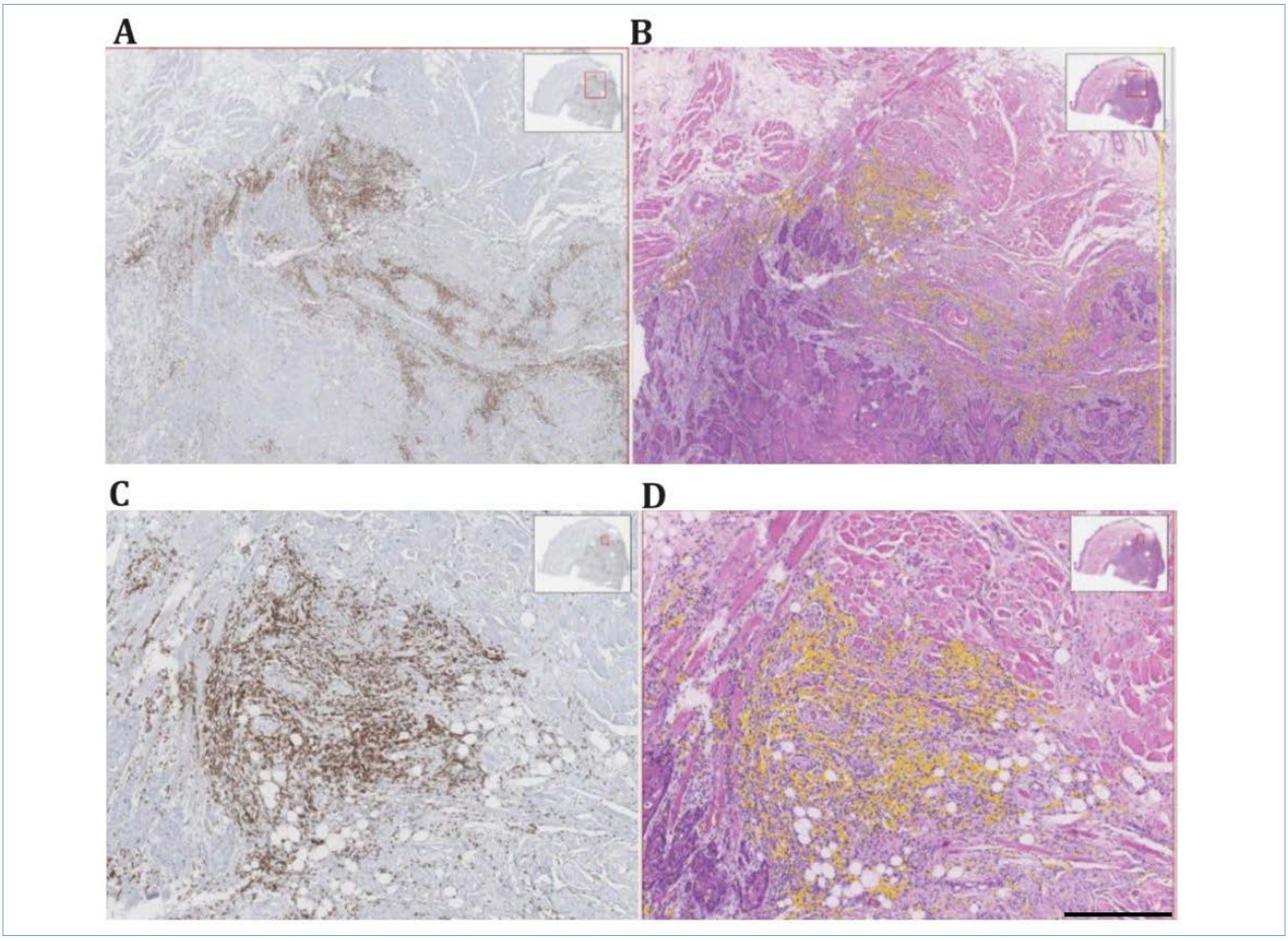



**Figure 3.** (A and C) IHC staining identifying CD3-positive lymphocytes. (B and D) Hematoxylin and eosin (H&E) staining was applied from the model, with recognized lymphocytes highlighted in yellow. Images C and D are magnified at 6.70x.(Scale bar:100 pixels).

**Table II.** Study population from TCGA. Synoptic report. All cases were OSCC.

| | | |
|---|---|---|
| **Age** | **Mean**<br>**Median**<br>**Range** | **64.82**<br>**65.5**<br>**24-87** |
| Sex | F<br>M | 15<br>35 |
| Stage | I<br>II<br>III<br>IV | 1<br>13<br>5<br>25 |
| Grade | G1<br>G2<br>G3 | 12<br>29<br>8 |
| Days to death | Mean<br>Median<br>Range | 3248.62<br>4211<br>11-8760 |
| Vital Status | Alive<br>Dead | 25<br>25 |

Using a modified Groovy script, we used QuPath with the StarDist extension (https://github.com/qupath/qupath-extension-stardist. Last accessed 06/08/2024) to infer the model on these 50 WSI (details in **Supplementary II**).

Our goal was to identify T lymphocytes in both the tumor and stroma. Before applying the script, we normalized all WSIs using QuPath's auto-parameters for color normalization and performed manual annotations. We annotated the invasive tumor margin and the central tumor region, while for the stroma we focused only on regions near the invasive tumor margin and the central tumor region. Using these annotations, we applied the script to extract the quantity of T lymphocytes.

### Definitions of "Cold" and "Hot"

We first determined the percentage of positive cells. Using the Cell Detection function of QuPath, we calculated the total cell count of both tumor and stroma in the specified ROIs. We further predicted the T lymphocytes by applying our model. From the total number of T lymphocytes identified, we derived a percentage using the following formula:

$$Percentage\ of\ T\ lymphocytes\ = \frac{Number\ of\ T\ lymphocytes}{Total\ number\ of\ cells} \times 100$$

Using the percentages, cutoffs were defined using the Cutoff Finder software [22], an online accessible platform for data optimization and visualization (available at the link: https://molpathoheidelberg.shinyapps.io/CutoffFinder_v1/). The cutoffs were calculated on the basis of the median values to distinguish between high and low T lymphocyte expression. This methodology allowed us to determine the cutoff for T cells infiltrating the intratumoral area and for T cells in the stroma.

### Statistical analyses

Statistical analysis was performed using the latest version (v29) of IBM SPSS Statistics for Windows (IBM Corp., Armonk, NY, USA). Survival data present in the TCGA were used for this purpose. First, we created Kaplan-Meier survival curves to estimate patient survival in the different groups (high and low T-cell expression). The curves were compared using the Log-rank (Mantel-Cox) test to determine if the differences were statistically significant ($p < 0.05$).

A Cox regression model was used to examine the effect of the selected variables on patient survival. An Omnibus test was performed to assess the model's overall significance. It was also checked whether adding new variables to the model causes a change in the Chi-square test that significantly improves the model. Once the optimal model was identified, the regression coefficients of the included variables were interpreted. A p-value less than 0.05 was considered statistically significant.

## Results

### Quality check

Analysis by HistoQC allowed us to distinguish computable regions from regions compromised by artifacts effectively. Figures 4 and 5 show that the pink-masked tissue portions represent the areas free of artifacts; conversely, the green-masked regions contain artifacts such as folds or altered staining. WSIs from the TCGA dataset showed a higher frequency of "unusable" regions than the in-house prepared images.

### Model building

We created our training dataset once the WSIs were prepared and annotated, as described in the Materials and Methods and **Supplementary VI**. The WSI images were tiled in 256x256px tiles, generating 7750 pairs (H&E and the correspondent mask). For a total of 15,500 tiles, each pair included one tile from the source image and the matching segmentation mask. We obtained 43,908 objects (i.e. CD3 positive cells) that were finally used to train the model. Figure 6 shows an example of an image tile with its segmentation mask.

After threshold optimization, the model's optimized values were as follows: {'prob': 0.4097943603992462, 'nms': 0.5}. The intersection over union (IoU) parameter was used to determine which dataset values

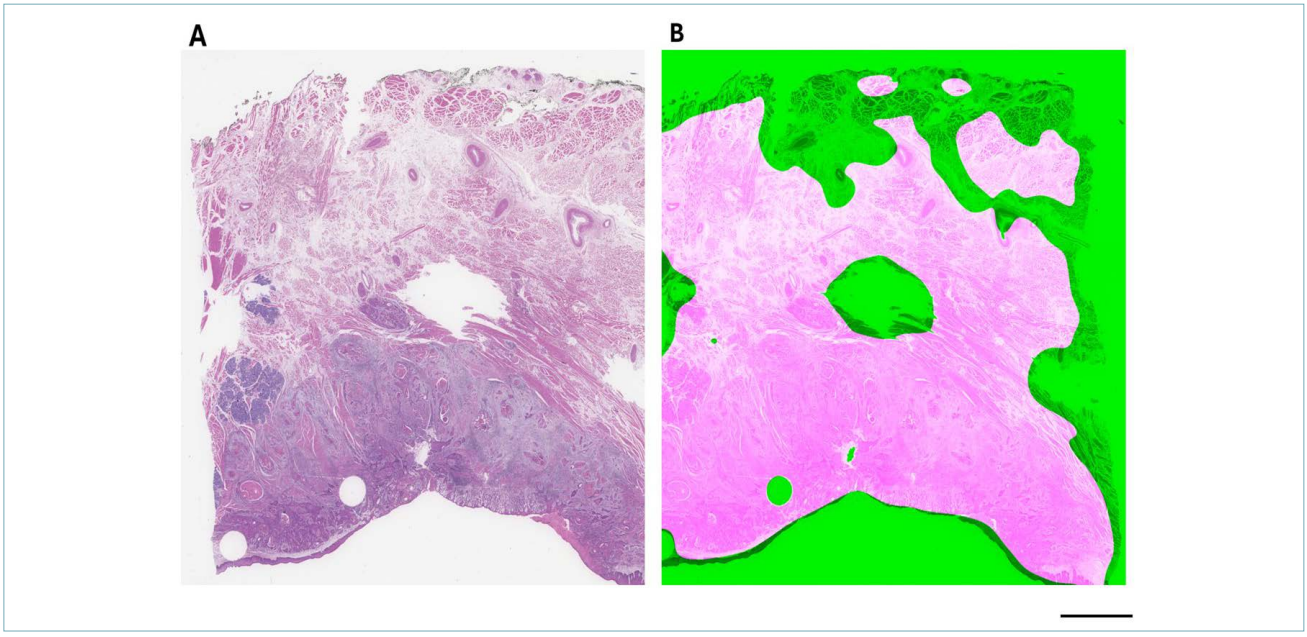



**Figure 4.** An example of a WSI from the training dataset analyzed by HistoQC. (A) shows the original WSI; (B) shows the final mask generated by HistoQC. Areas colored in pink indicate sections considered suitable for analysis, while those in green were classified as unsuitable (Scale bar: 100 pixels).

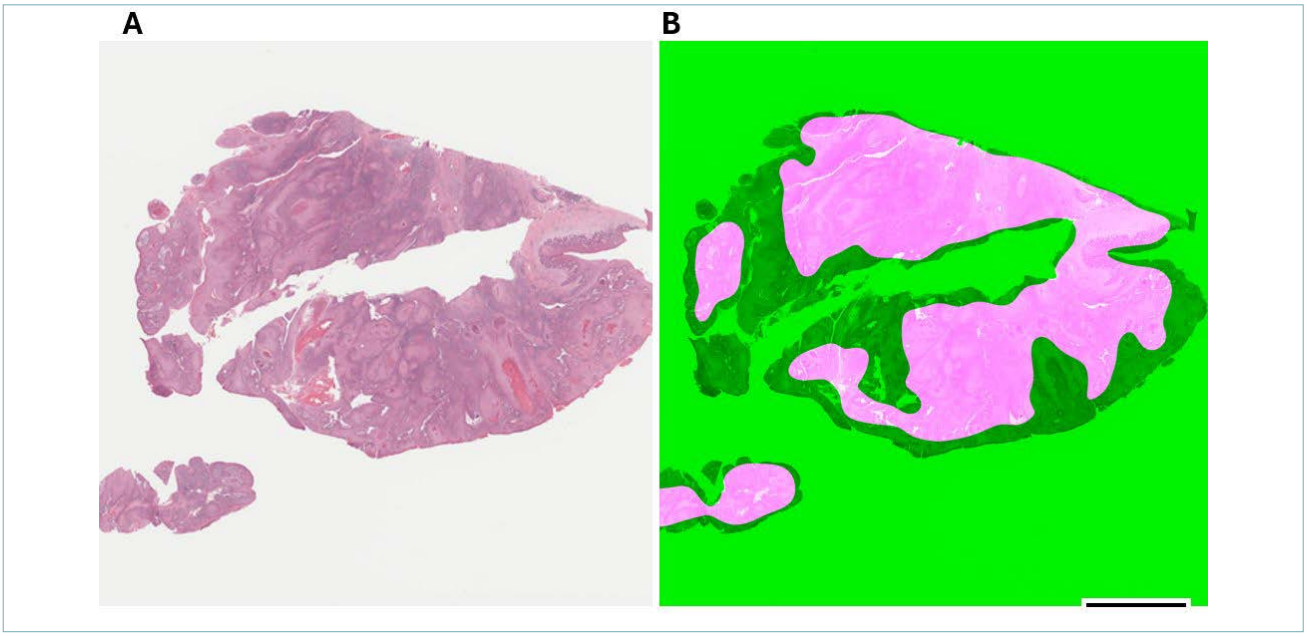



**Figure 5.** An example of a WSI from the inference dataset analyzed by HistoQC. (A) Original WSI; (B) final mask generated by HistoQC. Areas colored in pink indicate sections considered suitable for analysis, while those in green were classified as unsuitable (Scale bar: 100 pixels).

matched the computed threshold. The optimal IoU threshold chosen is thresh = 0.4. This allows us to balance precision, accuracy, and recall best, thus maintaining a low number of false negatives and false positives. The model's performance was evaluated with the creation of a graph where the performance

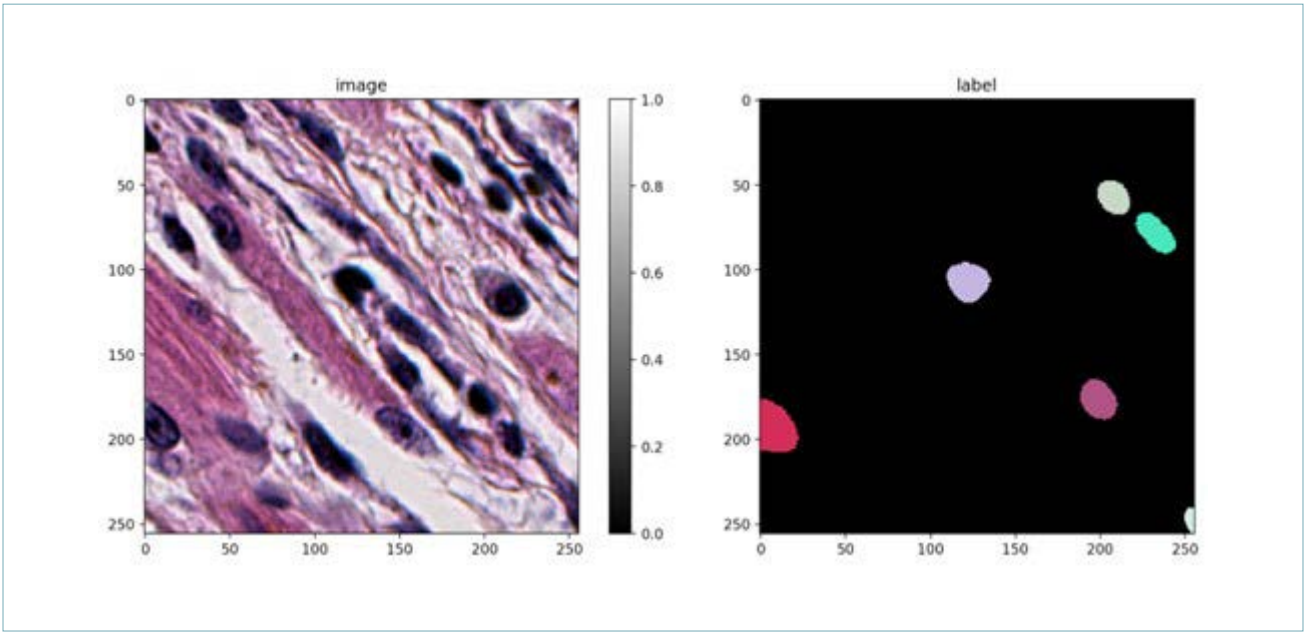



**Figure 6.** Example of a WSI image tile with the corresponding segmentation mask.

of our model is evaluated against a series of threshold values, as shown in **Supplementary III**.

### Inference of the trained model

After building the model, we inferred WSIs downloaded from TCGA (TCGA-HNSC project, link:https://portal.gdc.cancer.gov/).

We analyzed 50 cases of OSCC stained with hematoxylin and eosin. The objective was to evaluate the effectiveness of our model in recognizing T lymphocytes by comparing the data extracted from the model's application with patient outcomes provided by TCGA. As described in the Materials and Methods the analysis was conducted using QuPath (Fig. 7).

As described in the Materials and Methods section (2.8 Definitions of "Cold" and "Hot"), the following categories were defined (Fig. 8):

In the intra-tumor analysis, tumors were classified based on TIL percentages as follows:

- Intratumoral TIL Cold: Tumors with a TIL percentage below the median (4%).
- Intratumoral TIL Hot: Tumors with a TIL percentage above the median (4%).
- In the stroma analysis, regions were classified based on lymphocyte T percentages as follows:
- Stroma Cold: Regions with a lymphocyte T percentage below the median (11%).
- Stroma Hot: Regions with a lymphocyte T percentage above the median (11%).
- We then combined these data points to classify all tumors slides as follows:
- Cold Tumor: Intratumoral TIL percentage ≤ 4% and stromal T lymphocyte percentage ≤ 11%
- Hot Tumor: Either intratumoral TIL percentage > 4% or stromal T lymphocyte percentage > 11%

We generated Kaplan-Meier survival curves using the specified cutoffs, patient outcome data (alive or dead), and the time from the study's start to death.

First, we analyzed the expression of TILs (Fig. 9). Kaplan-Meier survival analysis revealed that the "TIL-cold" group experienced a rapid decline in survival (blue curve, Fig. 10A), reflecting a significantly lower survival rate compared to the "TIL-hot" group (red curve), which exhibited notably higher survival rates. This pattern persisted on the logarithmic scale (Fig. 10B).

The Log-Rank (Mantel-Cox) test further confirmed a significant difference in survival between the two cohorts (Fig. 10C), yielding a chi-square ($X^2$) value of 34.324 and a p-value < 0.001, indicating robust statistical significance.

We evaluated the data extracted from the stroma only (Fig. 11). Kaplan-Meier survival analysis (Fig. 12A) demonstrated that the "stroma-cold" group had a rapid initial decline in survival, stabilizing at low levels, whereas the "stroma-hot" group showed a more gradual decline and higher survival rates. Therefore, observations show that patients with low expression of T lymphocytes in stroma have a worse outcome than those with high expression. It is again observed

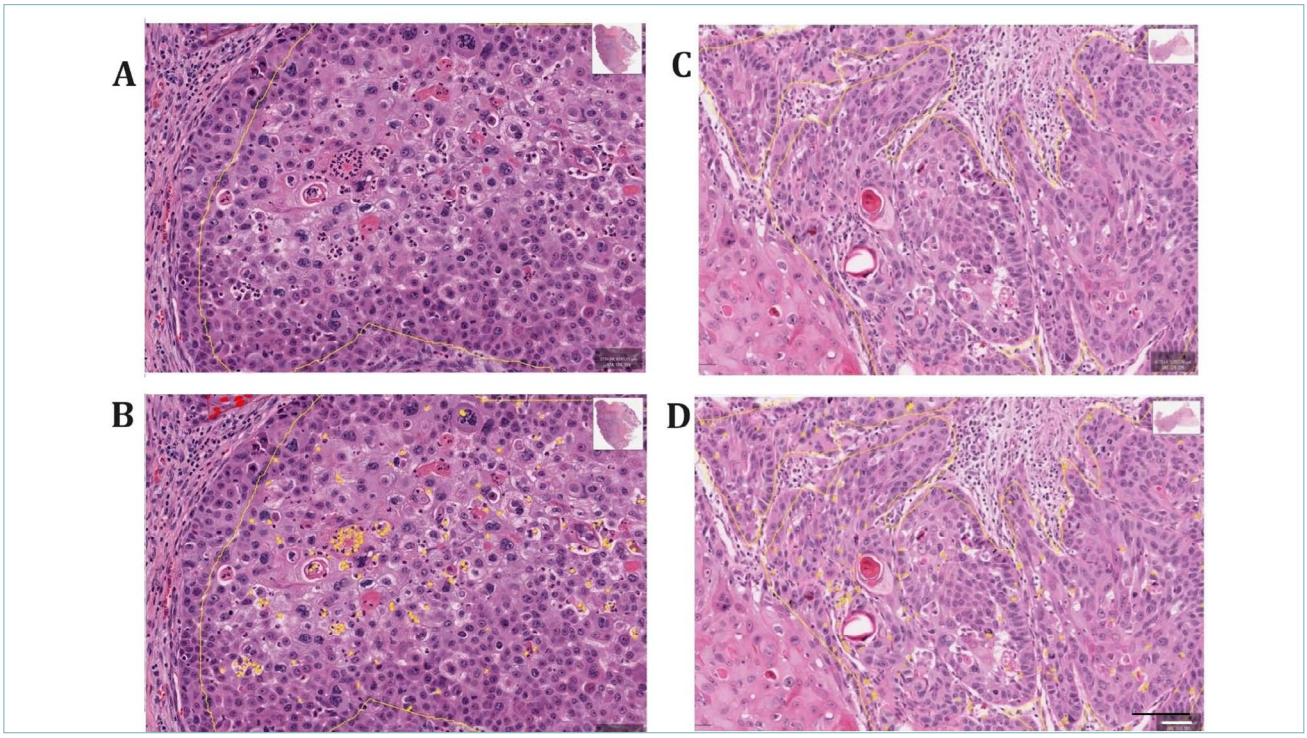



**Figure 7.** An example of model inference on WSI. (A) and (C) represent the original image with annotation ROIs. (B) and (D) are the same as in A and C WSI with overlaid model predictions (yellow overlays). (Scale bar: 100 pixels)

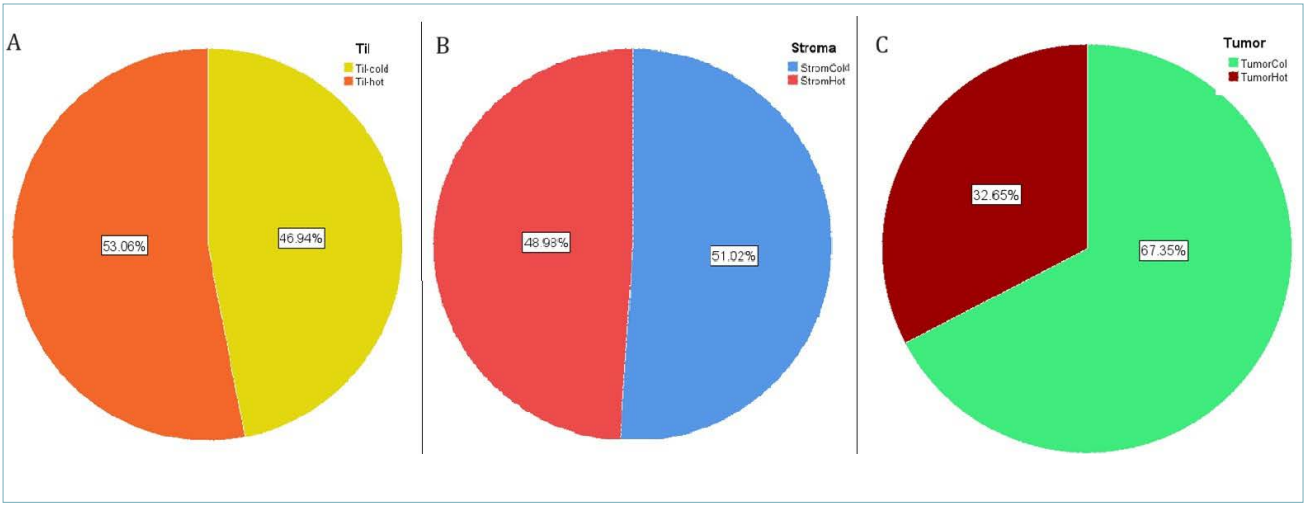



**Figure 8.** Percentage distribution of the three different categories analyzed. (A) Intratumor TIL is the subdivision between TIL-cold (53.06%) and TIL-hot (46.94%). (B) Stroma is the subdivision between StromaCold (48.98%) and StromaHot (51.02%). (C) Tumor is the subdivision between TumorCold (32.65%) and TumorHot (67.35%).

in "Logarithmic Survival Function" (Fig. 12B) that the survival of the "stroma-cold" group is lower than that of the "stroma-hot".

The Log Rank (Mantel-Cox) test has an $X^2$ value of 2.904, a degree of freedom (df) of 1, and a p-value of 0.088. Although the p-value of 0.088 suggests that we cannot say with certainty that this difference is statistically significant, given that it is close to the significance level. This suggests that with a larger sample, the difference could become statistically significant.

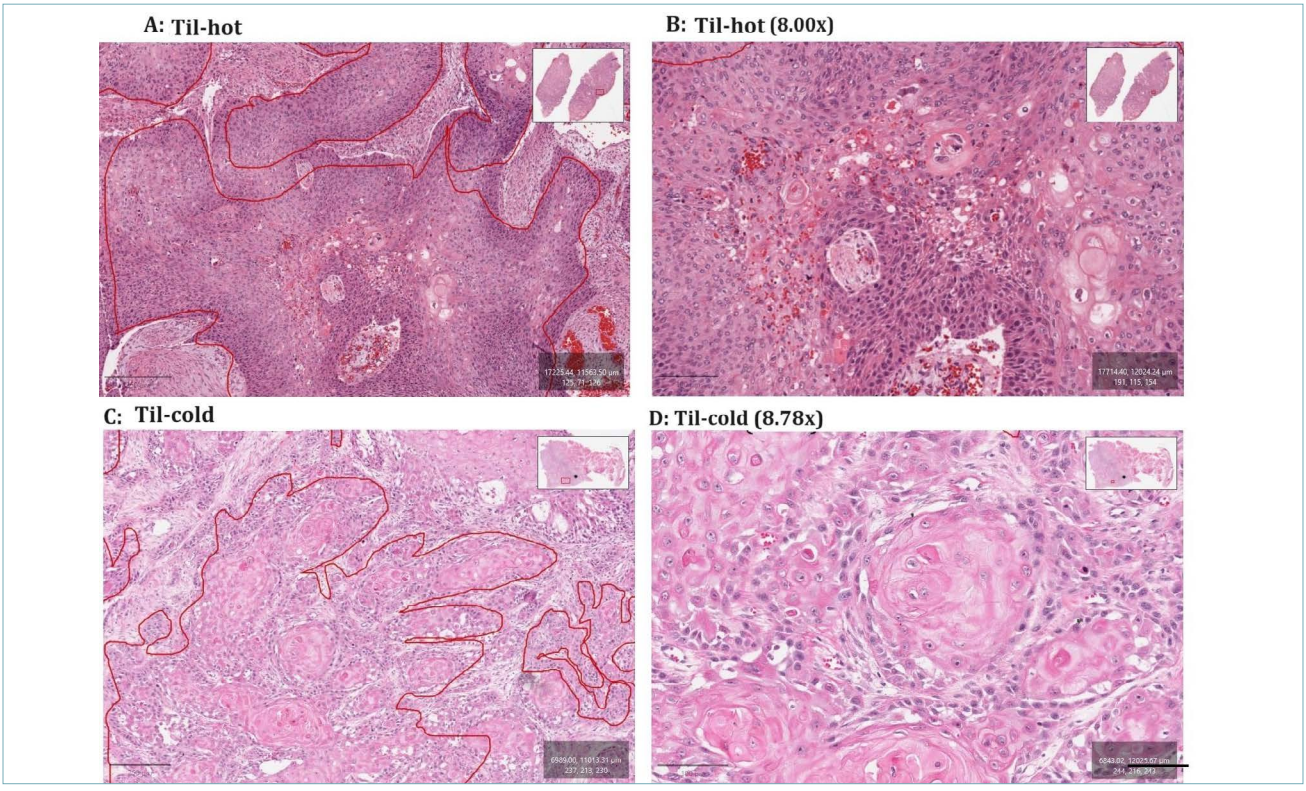
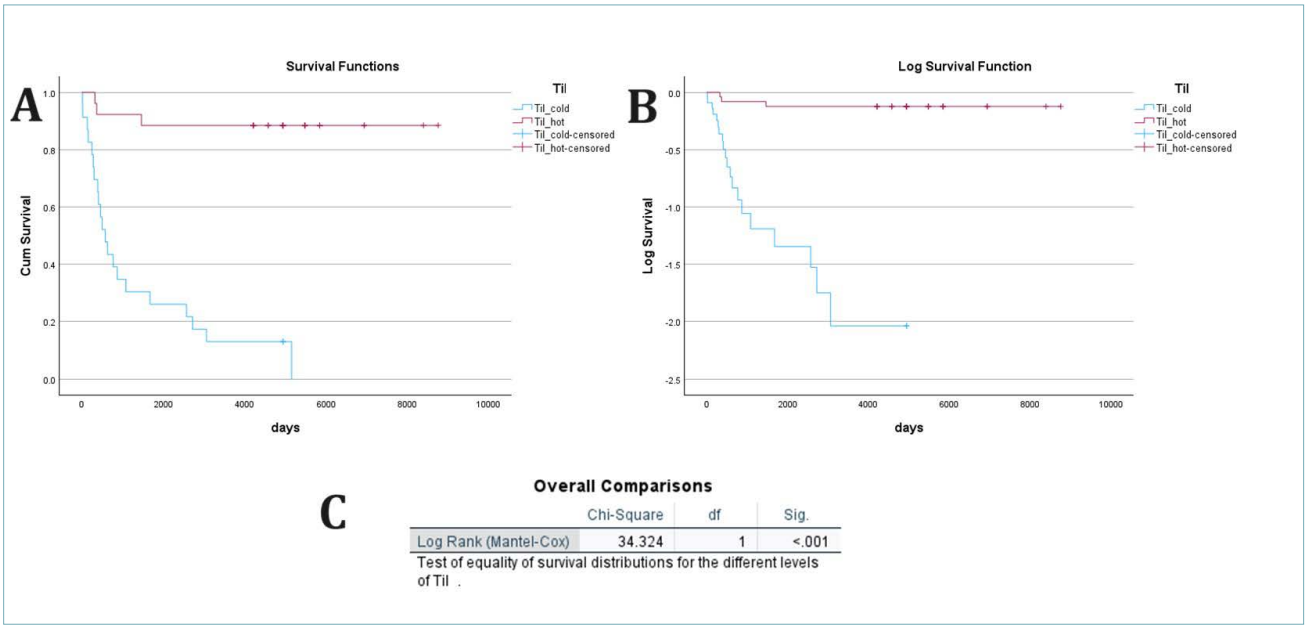



**Figure 10.** Representative images of TIL-hot and TIL-cold with red annotations indicating the tumor ROI. (A) TIL-hot at low magnification, (B) TIL-hot at 8x magnification, (C) TIL-cold at low magnification, (D) TIL-cold at 8.78x magnification. (Scale bar:100 pixels).




**Fugue 10.** (A) Survival Function: The vertical (Y) axis represents cumulative survival (probability of survival). The horizontal (X) axis represents time (in days). "Til-cold" curve (blue curve) percentage of T-lymphocytes 4%, and "Til-hot" (red curve) percentage of T-lymphocytes > 4%. Censored symbols indicate patients for whom no more data are available (censoring). (B) Logarithmic Survival Function: The vertical (Y) axis represents survival on a logarithmic scale. The horizontal (X) axis represents time (in days). The "Til-cold" curve (blue curve) percentage of T-lymphocytes 4%, and "Til-hot" (red curve) percentage of T-lymphocytes > 4%. (C) Log Rank (Mantel-Cox) test used to compare survival curves between two cohorts. The values presented include the Chi-square ($X^2$), degrees of freedom (df), and the p-value, indicating the statistical significance of the differences observed between the curves.

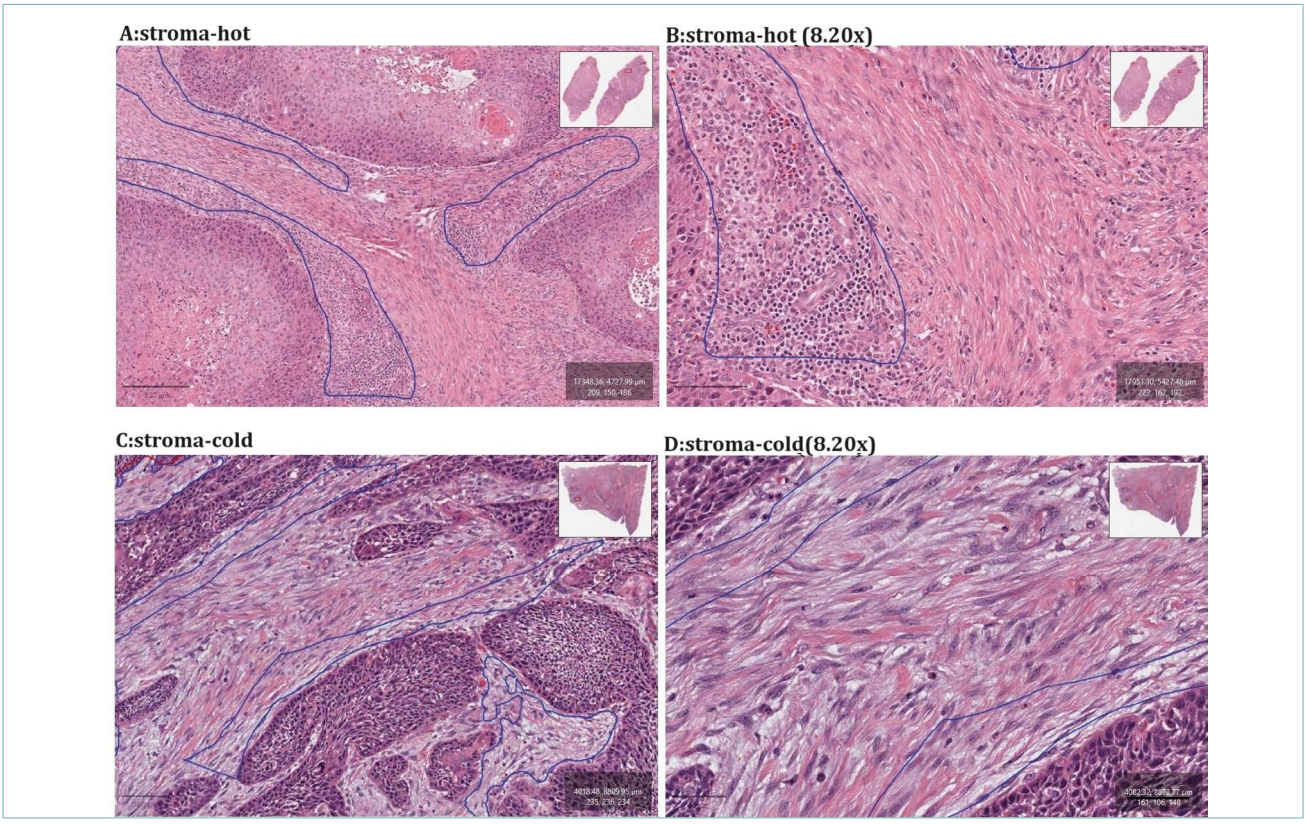



**Figure 11.** Representative images of stroma-hot and stroma-cold with blu annotations indicating the stroma ROI. (A) stroma-hot at low magnification, (B) stroma-hot at 8.20x magnification, (C) stroma-cold at low magnification, (D) stroma-cold at 8.20x magnification. (Scale bar:100 pixels).

Finally, we combined the results to evaluate whether high expression in both behaviors has positive outcomes as in the two separate regions (Fig. 13). The Kaplan-Meier survival curves (Fig. 14A) indicated that the "Tumor-cold" group had worse survival outcomes, with a faster decline and lower survival values compared to the "Tumor-hot" group, which maintained higher survival probabilities. The logarithmic curves also show this identical trend (Fig. 14B).

The Log Rank (Mantel-Cox) test (Fig. 14C) demonstrated high statistical significance ($p < 0.001$), and the high Chi-square value (11.642) indicates that the observed difference between the curves is significant. This suggests that low and high expression of T cells, both intratumorally and in the stroma, significantly impacts patient survival.

The Cox regression model confirmed that the percentage of TILs significantly predicted patient survival. The regression coefficient (B) for TILs was -2.574, with a standard error of 0.630. The Wald test yielded a value of 16.664, with a p-value < 0.001, indicating strong statistical significance. The hazard ratio (Exp(B)) was 0.076, suggesting that an increase in the percentage of TILs reduces the risk of death by 92.4%.

The Omnibus test confirmed the overall significance of the model (p-value < 0.001). Adding the TIL variable resulted in a $X^2$ change of 26.420, with a p-value < 0.001, demonstrating that including TILs significantly improved the model.

Other variables, such as tumor stage and grade, did not show significant p-values and did not contribute meaningfully to the model.

These findings underscore that the intratumoral percentage of TILs is a critical factor in patient survival, exerting a strong protective effect against the risk of death. (Detailed results are provided in Supplementary IV.)

## Discussion

The immune system is a critical player in the defense against tumor cells, engaging innate immunity initially

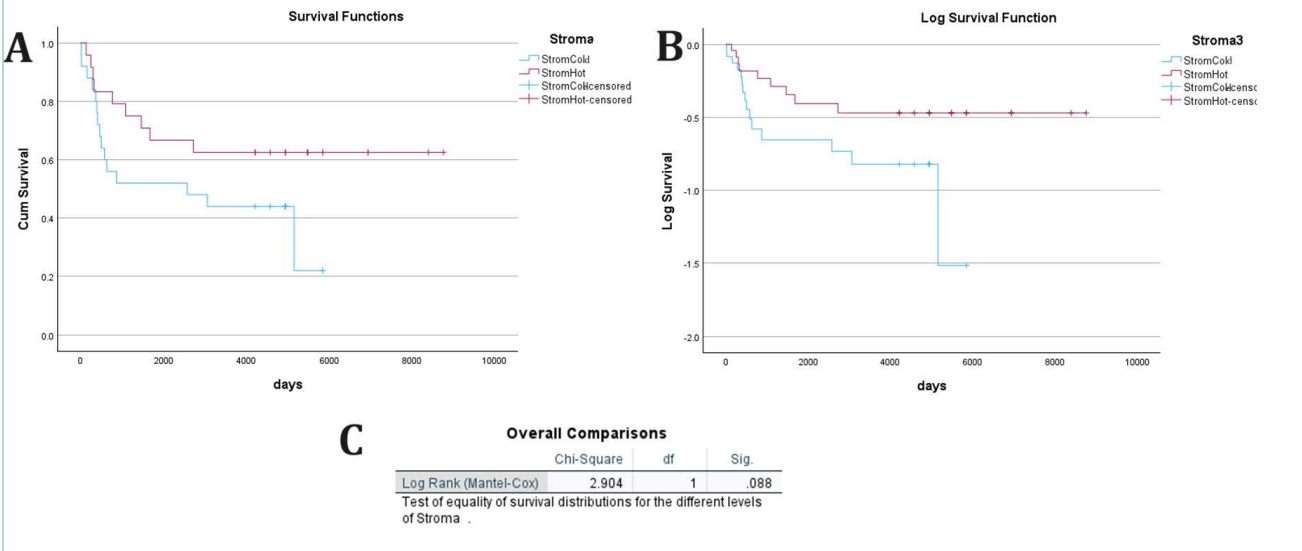



**Figura 12.** (A) Survival Function: The vertical (Y) axis represents cumulative survival (probability of survival). The horizontal (X) axis represents time (in days). "stroma-cold" curve (blue curve) percentage of T-lymphocytes 11%, and the "stroma-hot" (red curve) percentage of T-lymphocytes > 11%. Censored symbols indicate patients for whom no more data are available (censoring). (B) Logarithmic Survival Function: The vertical (Y) axis represents survival on a logarithmic scale. The horizontal (X) axis represents time (in days)."stroma-cold" curve (blue curve) percentage of T-lymphocytes 11%, and the "stroma-hot" (red curve) percentage of T-lymphocytes > 11%. (C) Log Rank (Mantel-Cox) test used to compare survival curves between two cohorts. The values presented include the Chi-square ($X^2$), degrees of freedom (df), and the p-value, indicating the statistical significance of the differences observed between the curves.

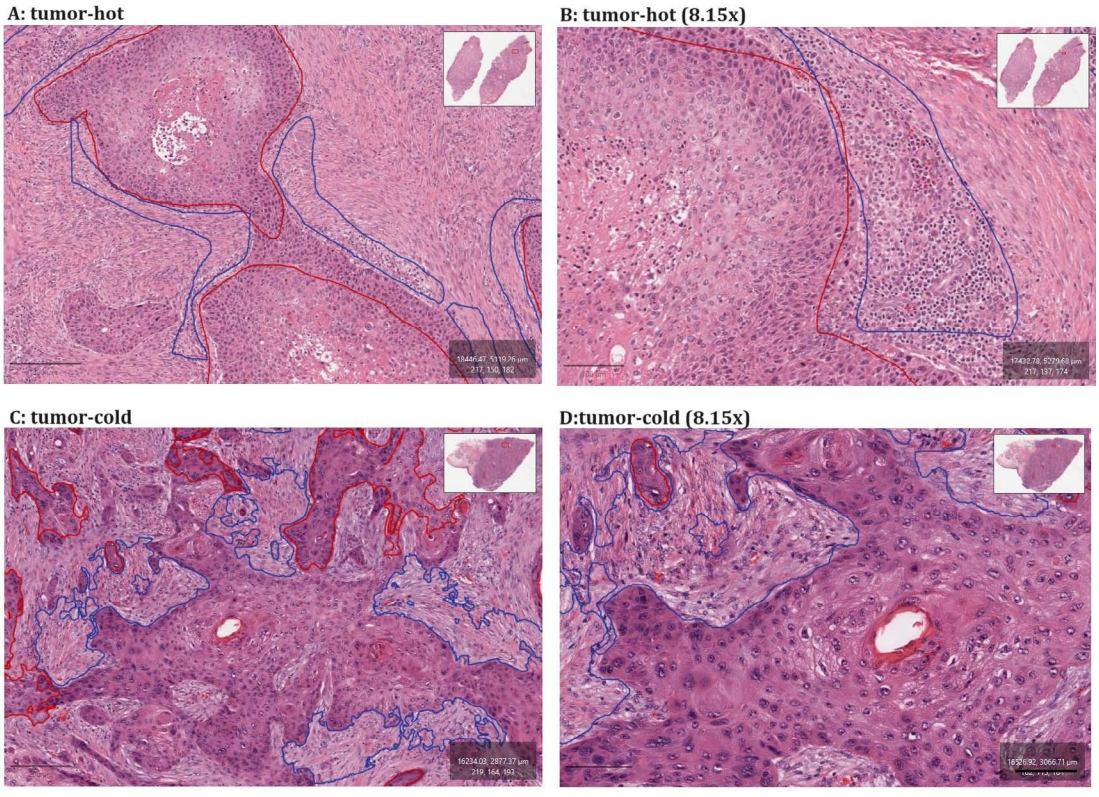



**Figure 13.** Representative images of tumor-hot and tumor-cold with blu annotations indicating the stroma ROI and red annotations indicating the tumor ROI. (A) Tumor-hot at low magnification, (B) tumor-hot at 8.15x magnification, (C) tumor-cold at low magnification, (D) tumor-cold at 8.15x magnification. (Scale bar:100 pixels).

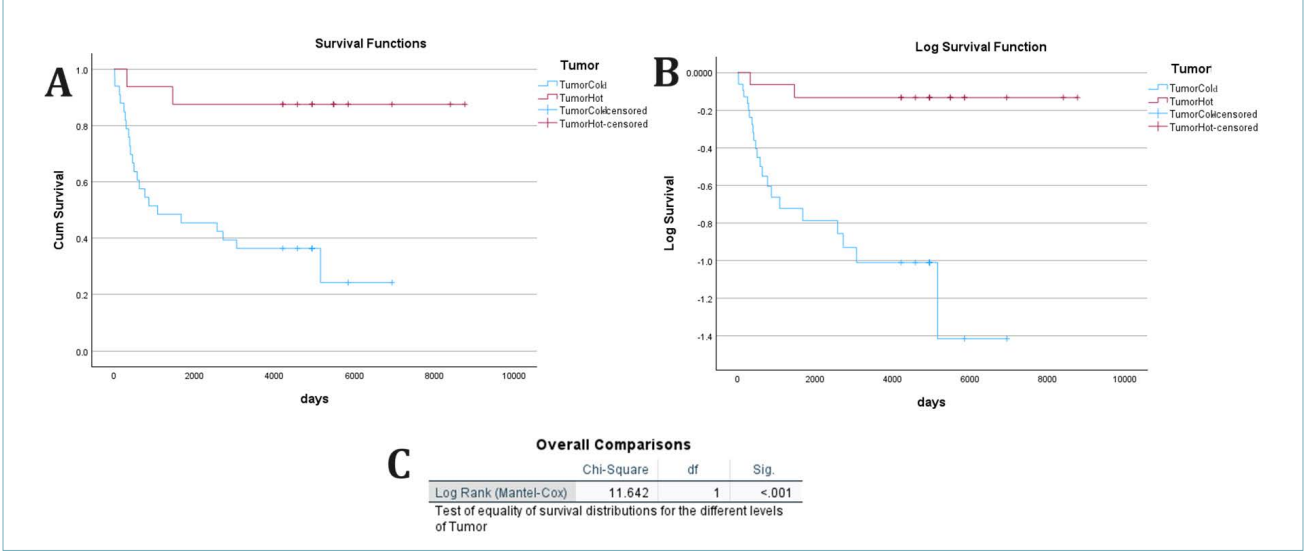



**Figure 14.** (A) Survival Function: The vertical (Y) axis represents cumulative survival (probability of survival). The horizontal (X) axis represents time (in days). The blue curve represents "Tumor-cold" (Til 4% and stroma 11%); the red curve represents "Tumor-hot" (Til > 4% and stroma > 11%). Censored symbols indicate patients for whom no more data are available (censoring). (B) Logarithmic Survival Function: The vertical (Y) axis represents survival on a logarithmic scale. The horizontal (X) axis represents time (in days). The blue curve represents "Tumor-cold" (Til 4% and stroma 11%); the red curve represents "Tumor-hot" (Til > 4% and stroma > 11%). (C) Log Rank (Mantel-Cox) test used to compare survival curves between two cohorts. The values presented include the Chi-square ($X^2$), degrees of freedom (df), and the p-value, indicating the statistical significance of the differences observed between the curves.

and recruiting adaptive immune responses. T lymphocytes are central to this process since they able to recognize and induce apoptosis in tumor cells. Numerous studies have demonstrated that a high density of T lymphocytes in tumors correlates with a better prognosis in OSCC. Given OSCC's malignancy and resistance to conventional treatments, identifying reliable prognostic biomarkers and developing targeted immunotherapies are imperative. Traditional methods, such as IHC, while accurate, are labor-intensive and prone to inter- and intra-observer variability, underscoring the need for automated solutions.

Artificial intelligence (AI) models have emerged as transformative tools in histopathology, enabling automated identification and quantification of TILs on hematoxylin and eosin (H&E)-stained WSIs. This study demonstrates the effectiveness of an AI-based workflow using QuPath and the StarDist model to segment and detect T lymphocytes. By leveraging IHC-derived annotations as ground truth, we trained a StarDist model capable of performing "virtual immunohistochemistry" on H&E-stained WSIs, bypassing the need for traditional IHC while maintaining accuracy.

Previous studies have validated the potential of AI in histopathology. For instance, Martino et al. (2020) [23] employed machine learning to predict the positivity of proliferation markers directly from H&E-stained tumor sections, and later, their deep-learning approach enhanced this capability [24]. Similarly, Saltz et al. (2018) used deep learning to quantify spatial patterns of lymphocytes in multiple cancer types [25], while Shaban et al. (2019) introduced a digital score for TILs in OSCC [26]. Budginaite et al. efficiently used Micro-Net architecture to segment lymphocytic cells in H&E-stained colorectal and breast cancer tissue images [27]. These efforts underscore the growing application of AI in immune profiling and its prognostic significance. Building on these advances, our study integrates AI-driven cell detection with a user-friendly digital pathology platform, bridging computational power and clinical utility.

Our analysis confirms that intratumoral T lymphocytes are significantly associated with improved survival in OSCC. Kaplan-Meier curves demonstrated that patients with "TIL-hot" tumors have better survival rates than those with "TIL-cold" tumors. Although the prognostic significance of stromal T lymphocytes approached statistical significance, their role in isolation may require further exploration with larger datasets. These findings align with established literature high-

lighting the prognostic value of T lymphocytes, particularly CD3-positive subsets, in OSCC and other malignancies [28,29].

One of the key strengths of our workflow is its simplicity and adaptability. By combining QuPath with StarDist, we developed a pipeline that can be seamlessly integrated into clinical laboratories [30]. The model was easily applied to TCGA WSIs, enabling high-throughput immune profiling. The flexibility to export models in portable formats enhances the potential for widespread adoption. However, our study has limitations. The current model is limited to single-class detection, focusing solely on CD3-positive cells. Expanding the model to identify multiple immune cell subsets would provide a more comprehensive analysis of the tumor immune microenvironment.

Future directions involve refining the model into a multi-class framework capable of detecting other key immune cell populations linked to patient outcomes, such as CD8+ and CD4+ cells. Additionally, incorporating advanced statistical techniques, larger and diverse datasets, and external validation steps could improve the model's robustness and generalizability.

## Conclusions

In conclusion, our study highlights the potential of AI-driven workflows for TIL detection in OSCC. The proposed methodology offers a scalable, efficient, and reproducible alternative to traditional IHC, enhancing diagnostic precision and prognostic stratification. This approach sets the stage for more personalized treatment strategies in OSCC and other cancers by reducing variability and manual labor.

### Abbreviations

AI: artificial intelligence
CNN: convolutional neural networks
DAB :Diaminobenzidine
df: degrees of freedom
FFPE: formalin-fixed, paraffin-embedded
fn: false negatives
fp: false positives
GT: ground truth
H&E: hematoxylin and eosin
HNSCC: head and neck squamous cell carcinomas
IHC: immunohistochemistry
IoU: Intersection over Union
OSCC: oral squamous cell carcinoma
ROI: Region of Interest
SE: standard error
TCGA: Cancer Genome Atlas
TIL: tumor-infiltrating lymphocytes
TME: tumor microenvironment
tn: true negative
tp: true positives
WSI: Whole Slide Imaging

### Glossary

See Supplementary material (Glossary).

### Conflict of interest statement

The authors declare no conflicts of interest.

### Funding

Rare cancers of the head and neck: a comprehensive approach combining genomic, immunophenotypic and computational aspects to improve patient prognosis and establish innovative preclinical models – RENASCENCE " (project code PNRR-TR1-2023-12377661)".

### Authors contribution

AC: Formal Analysis, Investigation, Writing original draft, Software, Visualization. SV: Investigation, Writing original draft, Analysis. GI: Project administration. DR: Validation, Writing review, and editing. RMDC: Validation, Writing review, and editing. SS: Funding acquisition, Validation. FM: Conceptualization, Methodology, Data Curation, Supervision, Writing review and editing.

### Ethical consideration

The study was performed according to the Declaration of Helsinki and following Italian law for studies based only on retrospective analyses of routine archival FFPE tissue; written informed consent from the living patient, following the indication of Italian DLgs No. 196/03 (Codex on Privacy), as modified by UE 2016/679 law of the European Parliament and Commission, was obtained at the time of surgery.

## References

1 de Melo Filho MR, Rocha BA, de Oliveira Pires MB, et al. Quality of life of patients with head and neck cancer. Brazilian Journal of otorhinolaryngology. 2013;79(1):82-88. https://doi.org/doi: 10.5935/1808-8694.20130014

2 Skvortsov S, Dudás J, Eichberger P, et al. Rac1 as a potential therapeutic target for chemo-radioresistant head and neck squamous cell carcinomas (HNSCC). Br J Cancer. 2014;110(11):2677-2687. https://doi.org/10.1038/bjc.2014.221

3 Vermorken JB, Herbst RS, Leon X, et al. Overview of the efficacy of cetuximab in recurrent and/or metastatic squamous cell carcinoma of the head and neck in patients who previously failed platinum-based therapies. Cancer. 2008;112(12):2710-2719. https://doi.org/10.1002/cncr.23442

4 Wei R, Liu S, Zhang S, et al. Cellular and Extracellular Components in Tumor Microenvironment and Their Application in Early Diagnosis of Cancers. Analytical Cellular Pathology. 2020;2020(1):6283796. https://doi.org/10.1155/2020/6283796

[5] Rodrigo JP, Sánchez-Canteli M, López F, et al. Tumor-Infiltrating Lymphocytes in the Tumor Microenvironment of Laryngeal Squamous Cell Carcinoma: Systematic Review and Meta-Analysis. Biomedicines. 2021;9(5):486. https://doi.org/10.3390/biomedicines9050486

[6] Wongpattaraworakul W, Choi A, Buchakjian MR, et al. Prognostic Role of Tumor-Infiltrating Lymphocytes in Oral Squamous Cell Carcinoma. BMC Cancer. 2024;24(1):766. https://doi.org/10.1186/s12885-024-12539-5

[7] Hadler-Olsen E, Wirsing AM. Tissue-infiltrating immune cells as prognostic markers in oral squamous cell carcinoma: a systematic review and meta-analysis. Br J Cancer. 2019 2019;120(7):714-727. https://doi.org/10.1038/s41416-019-0409-6

[8] Fang J, Li X, Ma D, et al. Prognostic significance of tumor infiltrating immune cells in oral squamous cell carcinoma. BMC Cancer. 2017;17(1):375. https://doi.org/10.1186/s12885-017-3317-2

[9] Nguyen N, Bellile E, Thomas D, et al. Tumor infiltrating lymphocytes and survival in patients with head and neck squamous cell carcinoma. Head Neck. 2016;38(7):1074-1084. https://doi.org/10.1002/hed.24406

[10] Almangush A, Leivo I, Mäkitie AA. Biomarkers for Immunotherapy of Oral Squamous Cell Carcinoma: Current Status and Challenges. Front Oncol. 2021;11. https://doi.org/10.3389/fonc.2021.616629

[11] Wang CW, Biswas PK, Islam A, et al. The Use of Immune Regulation in Treating Head and Neck Squamous Cell Carcinoma (HNSCC). Cells. 2024;13(5):413. https://doi.org/10.3390/cells13050413

[12] Ramos-Vara JA. Technical aspects of immunohistochemistry. Veterinary pathology. 2005;42(4):405-426. https://doi.org/10.1354/vp.42-4-405

[13] Swiderska-Chadaj Z, Pinckaers H, van Rijthoven M, et al. Learning to detect lymphocytes in immunohistochemistry with deep learning. Medical Image Analysis. 2019;58:101547. https://doi.org/10.1016/j.media.2019.101547

[14] Keren Evangeline I, Glory Precious J, Pazhanivel N, et al. Automatic Detection and Counting of Lymphocytes from Immunohistochemistry Cancer Images Using Deep Learning. J Med Biol Eng. 2020;40(5):735-747. https://doi.org/10.1007/s40846-020-00545-4

[15] Varricchio S, Ilardi G, Russo D, et al. Leveraging deep learning for identification and segmentation of "CAF-1/p60-positive" nuclei in oral squamous cell carcinoma tissue samples. Journal of Pathology Informatics. 2024:100407. https://doi.org/10.1016/j.jpi.2024.100407

[16] Chen Y, Zee J, Smith A, et al. Assessment of a computerized quantitative quality control tool for whole slide images of kidney biopsies. J Pathol. 2021;253(3):268-278. https://doi.org/10.1002/path.5590

[17] Janowczyk A, Zuo R, Gilmore H, et al. HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides. JCO Clin Cancer Inform. 2019;(3):1-7. https://doi.org/10.1200/CCI.18.00157

[18] Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: Open source software for digital pathology image analysis. Sci Rep. 2017;7(1):16878. https://doi.org/10.1038/s41598-017-17204-5

[19] Weigert M, Schmidt U. Nuclei instance segmentation and classification in histopathology images with StarDist. In: 2022 IEEE International Symposium on Biomedical Imaging Challenges (ISBIC). IEEE, 2022. p. 1-4.

[20] Weigert M, Schmidt U, Haase R, et al. Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2020. p. 3666-3673.

[21] Schmidt U, Weigert M, Broaddus C, et al. Cell detection with star-convex polygons. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Proceedings, Part II 1. Granada, Spain: Springer International Publishing; 2018. p. 265-273.

[22] Budczies J, Klauschen F, Sinn BV, et al. Cutoff Finder: A Comprehensive and Straightforward Web Application Enabling Rapid Biomarker Cutoff Optimization. Van Diest P, ed. PLoS ONE. 2012;7(12):e51862. https://doi.org/10.1371/journal.pone.0051862

[23] Martino F, Varricchio S, Russo D, et al. A Machine-learning Approach for the Assessment of the Proliferative Compartment of Solid Tumors on Hematoxylin-Eosin-Stained Sections. Cancers (Basel). 2020;12(5):1344. https://doi.org/10.3390/cancers12051344

[24] Martino F, Ilardi G, Varricchio S, et al. A deep learning model to predict Ki-67 positivity in oral squamous cell carcinoma. J Pathol Inform. 2023;15:100354. https://doi.org/10.1016/j.jpi.2023.100354

[25] Saltz J, Gupta R, Hou L, et al. Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. Cell Reports. 2018;23(1):181-193.e7. https://doi.org/10.1016/j.celrep.2018.03.086

[26] Shaban M, Khurram SA, Fraz MM, et al. A Novel Digital Score for Abundance of Tumour Infiltrating Lymphocytes Predicts Disease Free Survival in Oral Squamous Cell Carcinoma. Sci Rep. 2019;9(1):13341. https://doi.org/10.1038/s41598-019-49710-z

[27] Budginaitė E, Morkūnas M, Laurinavičius A, et al. Deep Learning Model for Cell Nuclei Segmentation and Lymphocyte Identification in Whole Slide Histology Images. Informatica. 2021;32(1):23-40. https://doi.org/10.15388/20-INFOR442

[28] Shimizu S, Hiratsuka H, Koike K, et al. Tumor-infiltrating CD8+ T-cell density is an independent prognostic marker for oral squamous cell carcinoma. Cancer Medicine. 2019;8(1):80-93. https://doi.org/doi:10.1002/cam4.1889

[29] Huang Z, Xie N, Liu H, et al. The prognostic role of tumour-infiltrating lymphocytes in oral squamous cell carcinoma: A meta-analysis. Journal of Oral Pathology & Medicine. 2019;48(9):788-798. https://doi.org/doi:10.1111/jop.12927

[30] Perez-Lopez R, Ghaffari Laleh N, Mahmood F, et al. A guide to artificial intelligence for cancer researchers. Nat Rev Cancer. 2024;24(6):427-441. https://doi.org/10.1038/s41568-024-00694-7