

<https://doi.org/10.15388/vu.thesis.947>

<https://orcid.org/0000-0003-4818-8316>

VILNIUS UNIVERSITY

Julius Drachneris

# Modeling of Urothelial Bladder Carcinoma Microenvironment by Digital Pathology and Artificial Intelligence Methods

**DOCTORAL DISSERTATION**

Medicine and Health Sciences,  
Medicine (M 001)

VILNIUS 2026

The dissertation was prepared between 2021 and 2025. The research was supported by Research Council of Lithuania.

**Academic Supervisor** – Prof. Dr. Arvydas Laurinavičius (Vilnius University, Medicine and Health Sciences, Medicine – M 001).

**Academic Consultant** – Prof. Dr. Feliksas Jankevičius (Vilnius University, Medicine and Health Sciences, Medicine – M 001).

This Doctoral Dissertation will be defended in a Public/Closed Meeting of the Dissertation Defence Panel:

**Chairman** – Prof. Dr. Dainius Characiejus (Vilnius University, Medicine and Health Sciences, Medicine – M 001),

**Members:**

Assoc. Prof. Dr. Vaida Baltrūnienė (Vilnius University, Medicine and Health Sciences, Medicine – M 001),

Prof. Dr. Ilze Štrumfa (Riga Stradiņš University, Medicine and Health Sciences, Medicine – M 001),

Dr. Marius Kinčius (National Cancer Center, Medicine and Health Sciences, Medicine – M 001),

Prof. Dr. Dalia Pangonytė (Lithuanian University of Health Sciences, Medicine and Health Sciences, Medicine – M 001),

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 11:00 on 26<sup>th</sup> of June 2026 in Medical Science Centre of the Vilnius University Faculty of Medicine. Address: Žaliųjų Ežerų g. 2, 08406, Vilnius, Lithuania

<https://doi.org/10.15388/vu.thesis.947>

<https://orcid.org/0000-0003-4818-8316>

VILNIAUS UNIVERSITETAS

Julius Drachneris

Šlapimo pūslės urotelio vėžio  
mikroaplinkos modeliavimas  
skaitmeninės patologijos ir dirbtinio  
intelektu metodais

**DAKTARO DISERTACIJA**

Medicinos ir sveikatos mokslai,  
Medicina (M 001)

VILNIUS 2026

Disertacija rengta 2021–2025.

Mokslinius tyrimus rėmė Lietuvos mokslo taryba.

**Mokslinis vadovas** – prof. dr. Arvydas Laurinavičius (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina – M 001).

**Mokslinis konsultantas** – prof. dr. Feliksas Jankevičius (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina – M 001).

Gynimo taryba:

**Pirmininkas** – prof. dr. Dainius Characiejus (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina – M 001).

**Nariai:**

doc. dr. Vaida Baltrūnienė (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina – M 001),

prof. dr. Ilze Štrumfa (Rygos Stradinio universitetas, medicinos ir sveikatos mokslai, medicina – M 001),

dr. Marius Kinčius (Nacionalinis vėžio centras, medicinos ir sveikatos mokslai, medicina – M 001),

prof. dr. Dalia Pangonytė (Lietuvos sveikatos mokslų universitetas, medicinos ir sveikatos mokslai, medicina – M 001).

Disertacija ginama viešame Gynimo tarybos posėdyje 2026 m. Birželio mėn. 26 d. 11 val. Vilniaus universiteto Medicinos fakulteto Medicinos mokslo centre. Adresas: Žaliųjų Ežerų g. 2, 08406, Vilnius, Lietuva

## LIST OF ABBREVIATIONS

BC	Bladder cancer
UC	Urothelial carcinoma
PUC	Papillary urothelial carcinoma
CIS	<i>Carcinoma in situ</i>
MIBC	Muscle-invasive bladder cancer
NMIBC	Non-muscle-invasive bladder cancer
TURBT	Transurethral resection of bladder tumor
BCG	Bacille Calmette–Guérin
NMIPUC	Non-muscle-invasive papillary urothelial carcinoma
RFS	Relapse-free survival
PFS	Progression-free survival
EAU	European Association of Urology
H&E	Hematoxylin and eosin
TIL	Tumor-infiltrating lymphocyte
TME	Tumor microenvironment
TAM	Tumor-associated macrophage
WSI	Whole slide image
DIA	Digital image analysis
DL	Deep learning
IZ	Interface zone
CM	Center of mass
ID	Immunodrop
IDR	Interface density ratio
CI	Confidence interval
CNN	Convolutional neural network
MIL	Multiple instance learning
RFS	Relapse-free survival
mIS	Modified immunoscore
AI	Artificial Intelligence
VUHSK	Vilnius University Hospital Santaros Clinics
NCP	National Center of Pathology
FFPE	Formalin-fixed paraffin-embedded
IHC	Immunohistochemical
C-index	Harrell’s concordance index
AIC	Akaike information criterion
Adam	Adaptive Moment Estimation
HR	Hazard ratio

Re-TUR repeated transurethral resection  
TLS Tertiary lymphoid structures

## TABLE OF CONTENTS

List of abbreviations.....	5
Table of contents .....	7
Introduction .....	10
Epidemiology and classification of bladder cancer.....	10
Adjuvant treatment of NMIBC .....	10
NMIBC risk assessment.....	11
Deep learning-based risk stratification.....	12
Tumor microenvironment .....	13
Tumor-infiltrating lymphocytes (TILs).....	13
Tumor-associated macrophages .....	13
Spatial analysis.....	14
Study hypothesis .....	15
The goal of the study .....	16
The objectives of the study.....	16
Statements to be defended.....	17
Materials and methods.....	18
Bioethics permit and funding .....	18
Patient cohort and sample collection.....	18
Immunohistochemical slide preparation .....	20
Slide digitization and tissue classification .....	20
Survival analysis .....	21
Objective 1: outcome prediction using abstract image features .....	22
Methods.....	22
Image data sets .....	22
Image feature extraction.....	24
Deep multiple instance learning (MIL) to predict patient relapse.....	25
Deep multiple instance survival learning (MISL) to predict the risk of relapse .....	25

Survival analysis .....	26
Results.....	27
Optimization of feature extraction network .....	27
Prediction of NMIPUC relapse by deep MIL .....	28
Prediction of risk of NMIPUC relapse by deep MISL.....	29
Clinicopathological variables and survival analysis .....	30
Objective 2: adaptation and optimization of immune response indicators...	33
Methods.....	33
Results.....	34
Univariable Cox regression analysis.....	34
Multivariable Cox regression analysis .....	35
Kaplan-Meier survival analysis.....	37
Objective 3: evaluation of prognostic value of immune cell subclasses and their distribution in the tumor microenvironment .....	38
Methods.....	38
Assessment of the spatial distribution of the immune cells .....	38
Statistical analysis .....	39
Results.....	40
Optimization of IZ width .....	40
Univariate Cox regression for prediction of RFS.....	41
Multiple Cox regression.....	43
Kaplan-Meier RFS analysis of the selected features.....	44
Objective 4: development of a comprehensive model .....	46
Methods.....	46
Results.....	47
Univariate Cox regression.....	47
Multivariate Cox regression.....	47
Kaplan-Meier analysis .....	49
Discussion .....	52

Clinical and pathologic variables .....	52
DL-based risk assessment .....	53
Adaptation of tumor-host interaction indicators .....	54
Prognostic value of immune cell subclasses .....	55
Development of a comprehensive model .....	57
Limitations .....	58
Conclusions .....	59
Further research directions .....	60
Recommendations .....	60
References .....	61
Supplementary material.....	73
Santrauka.....	74
Įvadas .....	74
Tyrimo hipotezė .....	76
Tyrimo tikslas.....	76
Tyrimo uždaviniai .....	76
Ginamieji teiginiai.....	76
Tiriamoji medžiaga ir tyrimo metodai.....	77
1 uždavinys. Išgyvenamumo predikcija pagal abstrakčius vaizdo požymius .....	78
2 uždavinys. Imuninio atsako indikatorių adaptavimas ir optimizavimas....	79
3 uždavinys. Imuninių ląstelių subklasių ir jų erdvinio pasiskirstymo prognostinės vertės tyrimas .....	81
4 uždavinys. Išsamaus modelio kūrimas .....	82
Išvados.....	83
Apibendrinimas .....	84
CURRICULUM VITAE .....	85
List of doctoral publications and presentations .....	86

# INTRODUCTION

## Epidemiology and classification of bladder cancer

Bladder cancer (BC) is the 9<sup>th</sup> most common type of cancer globally, with approximately 614,000 new cases and 220,000 deaths occurring in 2022. Among males, it is the 6<sup>th</sup> most frequently diagnosed cancer and the 9<sup>th</sup> leading cause of cancer-related death [1]. Although more recent national data is limited, in 2017, according to the cancer registry of the National Cancer Institute, there were 405 newly diagnosed cases in Lithuania and 263 BC-related deaths [2].

The most significant modifiable risk factor for BC development is tobacco smoking, which increases the risk up to fourfold due to carcinogens like beta-naphthylamine and polycyclic aromatic hydrocarbons [3]. Occupational exposure to aromatic amines and other industrial chemicals in sectors such as dye, rubber, and paint manufacturing also contributes substantially to bladder carcinogenesis [4]. In endemic regions, Schistosomiasis haematobium infection is a major cause of squamous cell carcinoma of the bladder, driven by chronic inflammation and nitrosamine production [5]. Additionally, obesity, particularly abdominal adiposity, has emerged as an independent risk factor, likely due to chronic inflammation and altered metabolic signaling [3].

The vast majority (>90%) of BC constitutes urothelial carcinoma (UC) – a malignant epithelial tumor derived from specific epithelium lining the urinary tract, referred to as transitional epithelium or urothelium [6]. Based on the differences in molecular pathways, UC can be classified into papillary tumors and flat tumors. Papillary tumors, including papillary urothelial carcinoma (PUC), are characterized by the presence of finger-like fibrovascular projections lined by neoplastic urothelium. These structures preserve cell-matrix interaction and provide a larger interface for metabolic interplay for continuous proliferation [7]. Papillary tumors are more likely to harbor FGFR3 mutations and have a lower risk of progression, while non-papillary tumors, including urothelial carcinoma in situ (CIS), have a higher incidence of TP53 and RB1 mutations and more aggressive behavior [8,9]. This classification aligns with the molecular taxonomy proposed by the Cancer Genome Atlas Research Network [10].

## Adjuvant treatment of NMIBC

Based on involvement of the detrusor muscle, BC can be categorized into non-muscle-invasive bladder cancer (NMIBC), which accounts for ~75% of

diagnoses, and muscle-invasive bladder cancer (MIBC) [11]. Treatment strategies for NMIBC differ significantly from those for MIBC due to differences in tumor invasiveness and prognosis. MIBC is typically managed with radical cystectomy, whereas NMIBC is treated with transurethral resection (TURBT) [11].

After TURBT, NMIBC carries a risk of recurrence and progression to muscle-invasive disease, which may eventually lead to metastasis if untreated. To reduce this risk, based on established risk stratification models, adjuvant therapy following TURBT is tailored to individual patient risk. For intermediate and high-risk patients, this treatment includes intravesical chemotherapy (mitomycin C, gemcitabine, anthracyclines) or intravesical bacillus Calmette-Guérin (BCG) immunotherapy [12]. While BCG immunotherapy is the most effective type of adjuvant treatment [13] it is associated with a higher incidence of side effects [14], which range from common side effects (e.g., cystitis, fever, hematuria) and rare but serious ones (e.g., BCG sepsis) [14]. However, selected very high-risk NMIBC can be treated with radical cystectomy [11].

#### NMIBC risk assessment

The risk of NMIBC recurrence and progression after TURBT is highly variable across patient groups. Current systems of tumor risk assessment include patient demographic and clinical history variables (patient age, recurrence rate), cystoscopic findings (tumor size, number of tumors), and pathologic findings (grade, stage, presence of concomitant CIS) [12,15–17]. However, while these models are widely accepted, they lack accuracy in predicting relapse-free survival (RFS) and progression-free survival (PFS) in NMIBC patients treated with BCG [18].

Histological grading is a crucial component of risk stratification in NMIBC. Two main grading systems are currently in use: the World Health Organisation (WHO) 1973 and the WHO 2004/2016 classifications. The WHO 1973 system uses a three-tiered scale (Grade 1, 2, and 3), while the WHO 2004/2016 system employs a two-tiered approach, categorizing tumors as low-grade or high-grade, with an additional category for papillary urothelial neoplasm of low malignant potential (PUNLMP). Although the newer system was introduced to improve clarity and clinical relevance, a major limitation of both grading systems is their poor interobserver reproducibility. Studies have shown substantial variability among pathologists when assigning grades, particularly for intermediate lesions, due to the subjective nature of nuclear

atypia and architectural disorganization. This inconsistency can lead to misclassification, which in turn affects treatment decisions and patient outcomes. A large multicenter study by the European Association of Urology (EAU) found that the WHO 1973 system had slightly better prognostic value for progression, while the WHO 2004/2016 system was more reproducible but less predictive in some settings. As a result, current EAU guidelines recommend using both systems in parallel to enhance prognostic accuracy [19].

To improve patient risk stratification, different approaches have been studied, including the assessment of biomarkers in the urine, assessing the tumor microenvironment, and improving histological tumor assessment, including digital image analysis methods [20,21].

### Deep learning-based risk stratification

Recent studies have demonstrated encouraging outcomes in the application of deep learning (DL) techniques for automated tumor grading [22–28]. Zhang et al. developed a deep learning-based system that outperformed pathologists in tumor grading. Additionally, the system generated descriptive annotations of histologic features in NMIPUC tissue [24]. However, this approach depends on histologic features defined by pathologists; therefore, it rather reproduces current medical knowledge. To evade this limitation, Lucas et al. performed a study predicting 1-year and 5-year relapse of NMIPUC using features extracted by a pre-trained VGG16 neural network, reaching 0.61 and 0.67 accuracy, respectively, thus demonstrating that the prediction of NMIPUC patient outcomes directly from hematoxylin and eosin (H&E) is possible [29].

The WSI- or patient-level prediction is complicated by the gigapixel size of WSI, which requires image partitioning of WSIs into smaller tiles or patches to enable efficient processing and feature extraction [30]. Moreover, intratumoral heterogeneity poses a challenge in assessing tumors from whole-slide images (WSI) of full-face histology sections. Previous studies have demonstrated the importance of identifying focal areas of higher grade for assessing NMIPUC risk [31–34]. The attention-guided multiple instance learning (MIL) framework addresses this problem by incorporating an attention layer, which helps focus on the most important image areas while also addressing the variability in the number of image patches [35]. Furthermore, the deep attention-guided multiple instance survival learning approach in lung and colorectal tumors developed by Yao et al. uses complete survival data (survival time and censoring data), thus better representing

patients' outcomes in comparison to the assessment of relapse-free survival (RFS) in the specific timeframe (e.g., one year) [36].

### Tumor microenvironment

The tumor microenvironment (TME) is crucial for cancer progression, patient survival, and treatment response. It comprises a complex network of immune cells, stromal elements, extracellular matrix, and signaling molecules that interact dynamically with tumor cells [37]. Importantly, the composition and spatial organization of immune cells within the TME significantly affect the efficacy of immunotherapies [38], including intravesical BCG in NMIBC and immune checkpoint inhibitors in advanced disease [21]. Understanding and quantifying these microenvironmental features are therefore essential for developing predictive biomarkers and optimizing personalized treatment strategies.

### Tumor-infiltrating lymphocytes (TILs)

Tumor-infiltrating lymphocytes (TILs) play a crucial role in the host immune response against cancer. Their presence often reflects the immune system's attempt to recognize and eliminate tumor cells, making them valuable biomarkers for prognosis and therapeutic response. Thus, TILs and their distribution patterns within the tumor microenvironment (TME) are extensively studied as important predictors of patient outcomes across various tumors, alongside efforts to standardize their assessment [39,40]. In NMIBC studies focusing on TIL subclasses, Th2 cell density was predictive of response to BCG immunotherapy [41,42], while a high density of regulatory T cells was associated with shorter RFS [43]. CD8-positive TILs were associated with tumor stage [44] and were co-localized with PD-L1-expressing cells [45]; however, their density in the tissue of NMIBC did not correlate with survival probability [41,45–47].

The prognostic significance of inducible co-stimulator (ICOS) positive lymphocytes in the context of immunotherapy has been demonstrated in several tumor types [28], including BC [29]. However, this biomarker has not been investigated in NMIBC patients treated with intravesical BCG to date.

### Tumor-associated macrophages

Tumor-associated macrophages (TAM) are an important constituent of the TME that affects tumor growth, immune suppression, chemoresistance, tumor

angiogenesis, and metastasis [48]. Macrophages are highly adaptable cells that, in response to signals from their microenvironment, such as chemokines and cytokines, differentiate/polarize into distinct phenotypes with specific functions [49]. The unpolarized macrophages, traditionally called M0 (naïve), can differentiate into M1 and M2 phenotypes, which can be further subdivided into subclasses. M1 macrophages are essential for pro-inflammatory responses, host defense against pathogens, and anti-tumor activities. In contrast, M2 macrophages primarily function in tissue repair, wound healing, and immunoregulation [49,50].

Hanada et al. first explored the significance of TAM and microvasculature density in BC. They found that a higher density of TAM is linked to muscle-invasive disease, vascular invasion, increased metastasis rates, and poor survival outcomes [51]. Further studies reported higher densities of TAM to be associated with stromal invasion [52], failure of BCG immunotherapy [47], a higher rate of recurrence [43,47,53], and progression to MIBC [54,55]. However, the meta-analysis performed by Wu et al. did not find a significant correlation between general (CD68-positive) TAM population density and BC patient outcomes.

Subsequent studies revealed the importance of macrophage polarization towards M1 macrophages and M2 macrophages: higher M1 macrophage densities were associated with better RFS [56] while high M2 macrophage densities were associated with worse RFS [47,56] and BCG immunotherapy failure [57].

### Spatial analysis

Recent advancements in digital image analysis (DIA) have enabled high-throughput quantification of TILs, including spatial distribution and subtype associations [58,59]. Furthermore, it enabled various methodologies for the assessment of the spatial distribution of TILs, including metrics of spatial heterogeneity, cluster analysis enabling hot-spot detection, metrics of cellular colocalization, and assessment of TIL distribution in specific tumor compartments [60]. The latter approach enables data-driven replication of grouping tumors into groups proposed by Chen [61] as “Immune-excluded”, “immune-desert” and “immune-inflamed”, paralleling findings of Herbs et al. [62].

The most notable method, the Immunoscore system, was proposed by Galon et al. [63,64] for colorectal cancer, based on cell density in the “core tumor” and “invasive margin”. This method was later validated in a large multicentric study [65] and adapted to other tumor types [66]. The primary

drawback of this method is the manual annotation of tumor compartments, which can lead to impaired interobserver variability.

In our institution, Rasmusson [67] proposed an automated tumor-stroma interface zone (IZ) sampling method. It enabled assessment of spatial immune cell analysis, evading manual tissue annotation with subsequent computation of Immunogradient indicators (Center of mass (COM) and Immunodrop (ID)). This method enables selective and extensive sampling of the tumor-host interaction area with quantification of the TIL density gradient across it. These indicators were tested as independent prognostic computational biomarkers in colorectal and breast cancer patients [67–71]. However, the performance of this methodology has not been investigated in the context of immunotherapy.

In BC, the direct application of “Immunoscore” was published by Rejeb et al. In their limited study (including only 30 patients), while Immunoscore showed associations with patients' age, lympho-vascular invasion status, and the absence of specific histological subtype, it lacked prognostic significance in the survival analysis [72]. Assessing immune cell densities in the different tumor compartments, Ariafar et al. showed associations with tumor stage; however, they showed no significant association with patient survival [73]. As an adaptation of “Immunoscore,” Bieri et al. developed a modified Immunoscore (mIS) using digital image analysis, which demonstrated its independent prognostic value in patients with muscle-invasive bladder cancer (MIBC) [74]. However, in a subsequent study in NMIBC patients, this indicator was of prognostic significance only in the high-risk patient subgroup [75]. Furthermore, the studies by Bieri et al. were performed on tissue microarrays, which lack a wider spatial context.

The research on the spatial distribution of macrophages is more limited. In the study evaluating prognostic significance of TAM in NMIBC in the context of BCG immunotherapy, Ajili et al. reported that a higher ratio of CD68-positive macrophage density in the epithelial versus stromal compartments was associated with improved recurrence-free survival (RFS) [53]. However, the assessment was performed using conventional microscopy, thus limiting its replicability.

## STUDY HYPOTHESIS

Digital pathology and AI-based models evaluating tumor tissue and tumor-host interaction can improve prognostic stratification of NMIPUC patients in the context of BCG immunotherapy.

## THE GOAL OF THE STUDY

To develop an integrated prognostic model for assessing the microenvironment and immune response of NMIPUC and predicting response to BCG immunotherapy, based on digital pathology and AI methods.

## THE OBJECTIVES OF THE STUDY

1. Study of NMIPUC tissues using abstract image features learned in artificial neural networks.
2. Study and optimize tumor-host interaction indicators in tissue samples of NMIPUC patients.
3. Identify morphological and microenvironmental features of tumor tissue that are significant for predicting the effect of BCG immunotherapy.
4. Develop comprehensive predictive models to help decide on the appointment of BCG immunotherapy to NMIPUC patients.

## STATEMENTS TO BE DEFENDED

1. Multiple instance survival learning (MISL) applied to histological images offers a distinct and independent contribution to enhancing relapse risk prediction in NMIPUC patients undergoing BCG immunotherapy. However, these findings require external validation.
2. Immune cell density gradient indicators adapted to the specific histologic architecture of NMIPUC can be used for the assessment of immune cell distribution in the epithelial-stromal interface to stratify NMIPUC patients into prognostic groups in the context of BCG immunotherapy. However, these findings require external validation.
3. Positive epithelial-directed gradient of proinflammatory CD11c macrophages, quantified by the interface density ratio (IDR), provides superior and independent prognostic value for NMIPUC relapse risk—surpassing simple cell density metrics and underscoring the importance of spatial immune analytics and anti-tumor cell populations (CD11c and CD8) in enhancing risk stratification alongside conventional clinicopathologic features.

## MATERIALS AND METHODS

### Bioethics permit and funding

This research was funded by a grant (“Artificial intelligence-driven prediction of BCG immunotherapy response in patients with non-muscle invasive papillary urothelial carcinoma”, No. S-MIP-21-31) from the Research Council of Lithuania.

This study was conducted in accordance with the Declaration of Helsinki and approved by the Vilnius Regional Biomedical Research Ethics Committee (research permit No. 2021/11-1394-867 issued on 5 November 2021). Patient consent was waived by the Lithuanian Bioethics Committee in accordance with the International Ethical Guidelines for Health-Related Research Involving Humans [76].

### Patient cohort and sample collection

To achieve a more homogeneous cohort for the purpose of this study, we have decided to include only NMIPUC patients, excluding flat tumors (CIS, invasive non-papillary carcinoma), which, as discussed previously, demonstrate more aggressive behavior.

We have collected pathologic and clinical data of all 230 BC patients treated with BCG immunotherapy in Vilnius University Hospital Santaros Clinics (VUHSC). For the study image analysis, we selected NMIPUC patients who met the following inclusion criteria:

- Diagnosis of papillary urothelial carcinoma
- pTa or pT1 stage
- Full (6-week) BCG induction course
- Available clinical and follow-up data
- Available archival tumor material at the National Center of Pathology (NCP).

A total of 166 patients met the inclusion criteria and had available H&E slides. 165 of them had available formalin-fixed paraffin-embedded (FFPE) tissue blocks for further immunohistochemical (IHC) analysis.

Pathologist Julius Drachneris (JD) reviewed original slides, and the most informative slides (demonstrating the highest tumor grade and areas of stromal invasion in invasive tumors) with corresponding FFPE tissue blocks were selected for further analysis. After slide preparation and initial image analysis, subsets of the original dataset with sufficient material without artifacts were selected for the specific tasks (tasks 1 and 2 – 157 patients, tasks 3 and 4 –

155 patients). Data from specific patient datasets are summarized in the corresponding publications [77–79].

*Table 1. Demographic, clinical, and pathologic data of patients included in our studies*

Characteristic	Value (%)
Patients	155 (100%)
Age, years	
Median (range)	69.8 (33–89)
Gender	
Male	127 (81.9%)
Female	28 (18.1%)
RFS time, months	
Median (range)	16.3 (1.6–60)
Recurrences (BCG failures)	46 (30%)
Tumor grade WHO 1973	
G1	5 (3.2%)
G2	67 (43.2%)
G3	83 (53.6%)
Tumor grade WHO 2004	
Low	12 (7.7%)
High	143 (92.2%)
pT stage	
Ta	94 (60.6%)
T1	61 (39.4%)
Carcinoma in situ association	8 (5.2%)
Positive re-TUR	55 (45.5%)
Recurrent tumor	45 (29%)
Multiple tumors	76 (49%)
Tumor size > 30 mm	43 (31.9%)
EORTC risk group	
Intermediate	75 (48.4%)
High	71 (45.8%)
Very High	5 (3.2%)

For the training of the feature extraction network, an additional cohort of primary NMIPUC patients was selected without clinical or demographic data collection. This cohort included 981 patients diagnosed with NMIPUC in the same period as the cohort of patients treated with BCG immunotherapy. A single most informative (covering invasive area if present and areas with the highest grade) H&E slide per patient was selected for image analysis.

## Immunohistochemical slide preparation

For immunohistochemical analysis from the selected most representative blocks (one block per patient), 3  $\mu\text{m}$ -thick sections were cut and stained using a standard procedure. After deparaffinization, sections were subjected to heat-induced epitope retrieval and incubation with primary monoclonal antibodies. Data on antibodies used in the study are summarized in the Table 2. Detection was performed using the ultraView Universal DAB Detection kit (Ventana Medical Systems, Oro Valley, AZ, USA), followed by counterstaining with hematoxylin. CD163, CD11c, and CD8 stainings were performed on a Roche Ventana Benchmark Ultra autostainer (Roche Diagnostics, Mannheim, Germany), while CD20 and ICOS stainings were performed on a Dako Autostainer Link 48+ Dako, Glostrup, Denmark).

*Table 2. Antibodies used in the IHC preparation in our studies*

<b>Antibody</b>	<b>Clone</b>	<b>Dilution</b>	<b>Manufacturer</b>
CD8	C8/144B	1:100	Dako
CD20	L26	1:500	Dako
ICOS	D1K2T	1:500	Cell Signaling Technology
CD163	MRQ-26	1:50	Cell Marque
CD11c	5D11	1:100	Leica

## Slide digitization and tissue classification

All IHC and original H&E slides were digitized at 20x magnification (0.5  $\mu\text{m}$  per pixel) using an Aperio® AT2 DX scanner (Leica Aperio Technologies, USA).

We utilized HALO® AI (Indica Labs, USA) Densenet v2 classifiers to classify tissues based on their characteristics. We excluded any artifacts, primarily coagulation resulting from the TURBT procedure, and classified good-quality tissues into epithelium and stroma classes. Additionally, for CD20 staining, we trained a classifier segmenting tertiary lymphoid structures. Tissue classifiers were trained iteratively by adding manual annotations and training the classifier network. The quality of tissue classification masks produced by HALO® AI was visually assessed by a pathologist (JD). We performed tissue classification in a sequential manner, first applying the glass-tissue classifier, then applying the artifact exclusion classifier on the tissue areas, followed by the stroma-epithelium classifier or TLS classifier (see

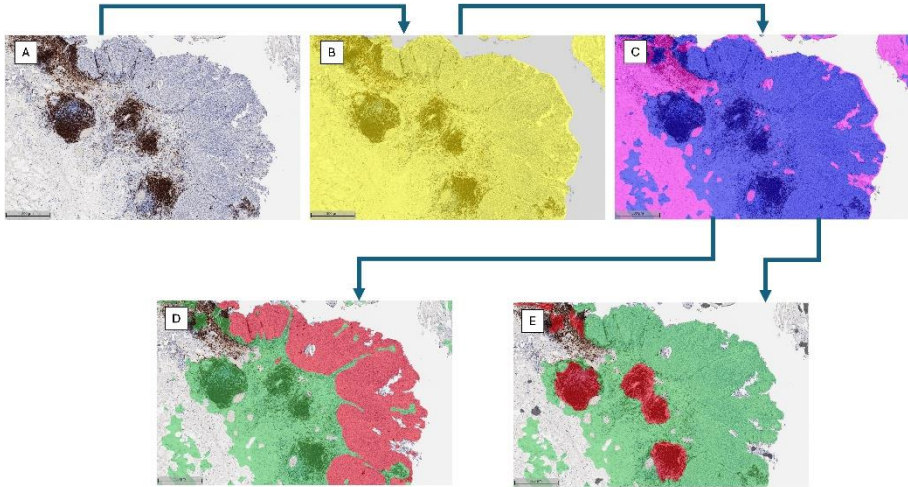


Figure 1).

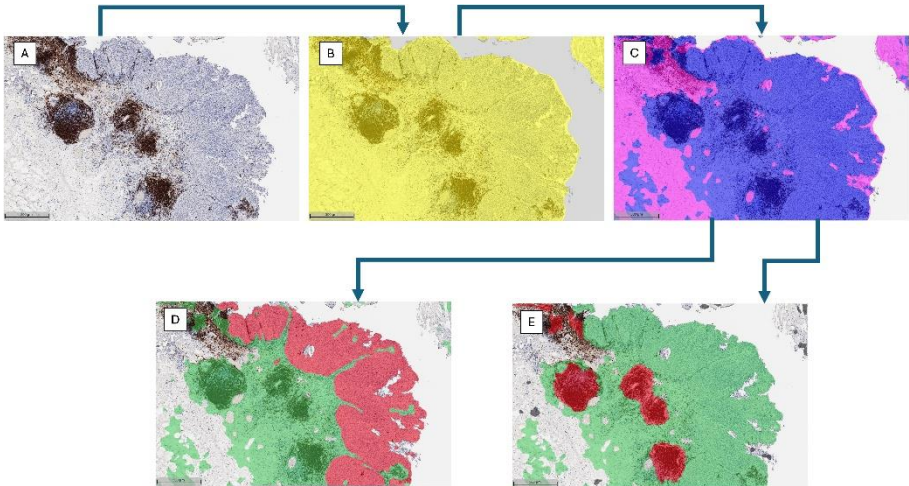


Figure 1. Tissue classification workflow. Image (A) undergoes tissue detection classification (B) followed by artifact exclusion (C). Remaining good quality tissue is further subjected to epithelium-stroma classification (D) or, in the case of CD20 IHC staining, to TLS detection (E).

### Survival analysis

The univariable Cox model was used to evaluate the performance of individual features and to select features for multivariable Cox regression. For multivariable analysis, we selected factors from univariable analysis with a p-value  $<0.1$ , constructing all possible combinations of these variables for modeling. Multivariable Cox regression models using these constellations of variables were then fitted to the training set to select models consisting of

independent variables (with HR p-values  $< 0.05$  for all covariates). On these selected models, we performed cross-validation, using the mean Harrell's concordance index (C-index) on the validation set as a performance indicator, along with the Akaike information criterion (AIC). The Kaplan–Meier survival estimator was used to estimate survival for patients stratified by factors with statistically significant associations with recurrence hazard in the univariate Cox analysis, and the log-rank test was used to compare patient survival across groups. Statistical data analysis and visualisation was performed using Python libraries (Pandas, Scikit-learn, SciPy, Lifelines, Matplotlib).

## OBJECTIVE 1: OUTCOME PREDICTION USING ABSTRACT IMAGE FEATURES

Image features extracted using neural networks can be used for various classification and regression tasks. In this study, we investigated various approaches to tumor tissue image sampling and patient outcome prediction, as well as variations in the training of the feature extraction neural network. We compared the performance of this approach in classification and survival regression tasks.

### Methods

#### Image data sets

We compared several different methods of image sampling. To assess different levels of histological image detail, we examined 1024-, 512-, and 256-pixel-sized patches separately. To combine image features from different patch sizes, we used a multiresolution patch sampling approach. In computer vision, image patches are usually used with dimensions that are powers of two. These sizes align with the hierarchical subdivision framework, enabling a comprehensive analysis of the image's structural hierarchy. The choice to start with a patch size of 256 pixels was considered suitable because it effectively balances the extraction of fine details with the preservation of the overall context of the image. Smaller patches (256 pixels) are expected to emphasize tumor cytology details, whereas larger patches (1024 pixels) represent tissue microarchitecture. These differences in image characteristics can be visually appreciated in Figure 2. To prepare image data for multiresolution analysis, a hierarchical subdivision technique was utilized, enabling the smooth integration of patches at different resolutions. Initially, 1024-pixel-sized image patches were extracted from the WSIs. To ensure the analysis focused on relevant tissue regions, we filtered the selected patches to retain only those with at least 50% tissue content as determined by the ratio of pixels belonging to the 'stroma' or 'epithelium' classes and the total number of pixels in a corresponding patch region in the predicted HALO® AI classifier mask (as described previously). This tissue content control mechanism is a crucial safeguard against including artifacts or non-tissue regions, enhancing the reliability and relevance of our findings. Subsequently, the 1024-pixel patches meeting the tissue content criterion were subdivided into 512-pixel patches, followed by a further division into 256-pixel patches. This procedure resulted

in a hierarchical series of image patches at different resolutions (see Figure 2), providing multiscale representations for subsequent analyses.

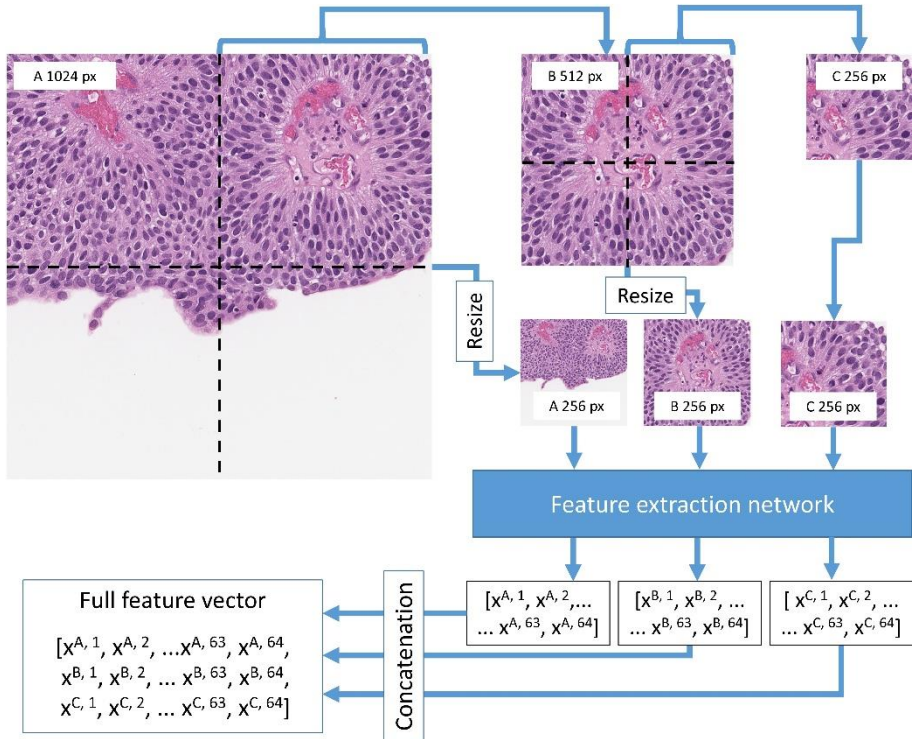


Figure 2. Hierarchical patch subdivision for multiresolution analysis. 1024-pixel patch subdivided into four 512-pixel patches, which were further subdivided into 256-pixel patches. 1024 and 512-pixel patches were resized to 256-pixel patches. Patches were passed to the feature extraction network, and for each 256-pixel patch, the corresponding larger-sized patch feature vectors were concatenated.

Patches were grouped into 3 clusters according to stroma and epithelium content, defined by HALO® AI tissue classifier (C1 cluster being predominantly (>50%) composed of stroma, C2 cluster – having <50% of stroma and <50% of epithelium, and C3 being predominantly (>50%) composed of epithelium). Similarly, multiresolution patches were assigned into clusters according to 1024-pixel patch epithelium-stroma content. We included stromal areas of tissue to include not only features of epithelial tissue, which are more important in describing tumor features (usually conceptualized as tumor grade), but also cover the features of host response, which can be seen as changes in tumor stroma (fibroblastic response, fibrosis, stromal infiltration of immune cells).

## Image feature extraction

To predict patient outcomes from readily available, routinely H&E-stained histology tissue WSIs, we converted the sampled image patches into feature vectors. This feature extraction was performed according to the method published by Rawat et al. [80]. Similarly, we designed our feature extraction model based on the InceptionResNetV2 architecture and tasked it to assign the same identity index correctly to all patches sampled from the same patient WSI. To train the feature extraction network, we composed a dataset of neighboring patch pairs sampled from the training cohort WSIs (see Figure 3). One patch from each pair was reserved for training the feature extraction network, while the second was only used to validate the training (resulting in a 0.5 training-validation data ratio).

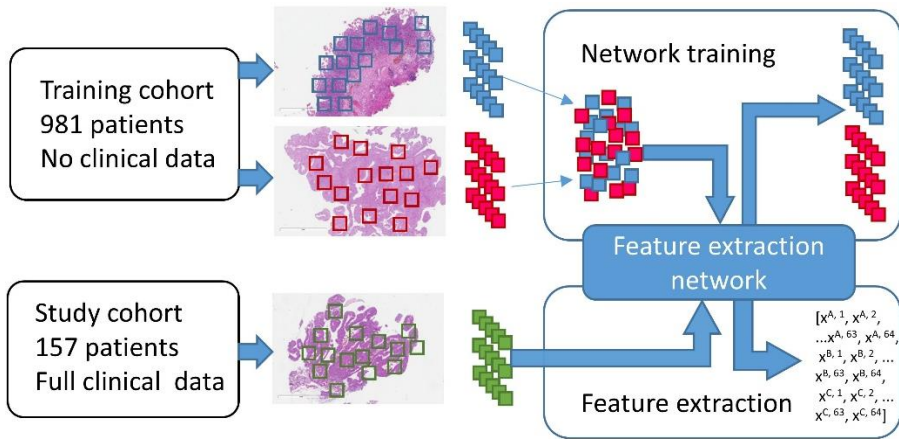


Figure 3. Schematic representation of feature extraction network training and feature extraction.

To optimize the feature extraction network, we have run experiments with 256 pixel-sized patches employing variations of the dataset and the mode of feature extraction (Figure 4). We tested the patch pair matching accuracy for different numbers of patch pairs per WSI used to train feature extraction models. We prepared datasets by extracting three patch pairs, ten patch pairs, and 100 patch pairs per single WSI. When 100 patch pairs per WSI were unavailable, we employed all available patch pairs. To reduce the length of extracted feature vectors, we employed an additional compression layer - the last layer before the decision layer, with two variations utilizing the 2D convolutional or the dense layers and conditioning these layers with a different number of output features - 1536, 1024, 512, 256, 128, 64, 32, 16, 8, 4, and 2.

## Deep multiple instance learning (MIL) to predict patient relapse

We adopted the MIL implementation as proposed by Ilse [35] to train a simple convolutional neural network (CNN) model. This classical MIL assumption involves feeding the network with batches of data extracted from WSIs, providing the ability for binary classification based on clinical data categories. We modified the original method to accept image feature vectors instead of image patches. Therefore, each MIL batch comprises a set of feature vectors from image patches originating from the same WSI, and the batch label was derived from the associated clinical data value for the corresponding WSI. The model is capable of handling bags of varying lengths. Training involves a small CNN with the Adam optimization algorithm, terminating when validation loss remains unchanged for 200 epochs. As per the original method, an attention-based MIL pooling layer is incorporated before the model's final layer. The objective function is the negative log-likelihood of the Bernoulli distribution.

To evaluate the performance of our MIL model, we implemented a 5-fold cross-validation scheme, randomly dividing the patient cohort into five equal folds. The model was trained on four folds and evaluated on the remaining fold, repeating this process for each fold. This procedure provided a robust assessment of the model's generalizability to unseen data.

To balance the MIL training dataset due to a low number of patients with relapse (positive class), we employed minority class oversampling. Oversampling was achieved by repeatedly drawing random positive cases from the minority pool to balance the class distribution.

## Deep multiple instance survival learning (MISL) to predict the risk of relapse

An image feature vector-based standard MIL assumption for binary classification can be adapted to model patient survival by utilizing a loss function based on survival probability. For MISL, we have adopted a method published by Yao et al. [36]. The MISL method adopts the negative partial log-likelihood as a loss function and an average concordance index as a training metric.

To train the MISL model effectively for patient survival prediction, we designed an oversampling technique that specifically addresses the imbalance in the patient survival status distribution. By augmenting the minority class, which represents patients with poorer survival outcomes, our oversampling

technique ensures that the model is adequately exposed to the diverse patterns associated with shorter survival times.

The technique involves identifying the longest follow-up time ( $t_{\max} = 1826$  days) among all patients in the cohort. Subsequently, we artificially extend the follow-up period for each patient to match  $t_{\max}$ , ensuring consistent evaluation of survival status across patients. Next, we define a fixed time interval ( $t_{\text{step}} = 30$  days) to check the survival status of each patient at regular intervals. A cohort-wide survival matrix ( $T$ ) with dimensions  $m \times n$  is constructed, where  $m$  represents the number of patients and  $n$  represents the number of  $t$ -steps. The matrix is filled row-wise by assigning the appropriate survival status (left-censored, event, or right-censored) for each patient at each  $t$ -step. The resulting over-sampled MISL training set, represented by the completed survival matrix  $T$ , provides a balanced representation of patient survival status, facilitating effective training of the MISL model. We trained the MISL model in a 3-fold cross-validation setting.

### Survival analysis

Both MIL and MISL predictions from each fold of a  $k$ -fold cross-validation were aggregated to reconstruct the entire cohort's survival statistics and were assessed by Kaplan-Meier survival analysis. A one-sided Log-rank test was used to assess the difference between patient groups in MIL and MISL experiments, since both models, MIL and MISL, predicted high-risk groups and low-risk groups, and the correct allocation of patients was important, not only the survival difference between the groups. To compare different outcome prediction methods, we have stratified patients into two groups according to the cut-off value with the lowest MISL-predicted log-rank test  $p$ -value. We performed multiple Cox regressions to analyze the MISL model prediction performance in the context of other clinical features. We have used the partial Akaike information criterion (AIC) for assessment of the models' prediction error and the concordance index (C-index) to assess predictive performance of the models. We used Kaplan-Meier survival analysis in patient groups defined by other prognostic features. Additionally, differences in the distribution of prediction values between patient subgroups were assessed using Kruskal-Wallis and Mann-Whitney U tests where applicable.

## Results

### Optimization of feature extraction network

The original methodology proposed for feature extraction by Rawat et. al. [80] used a 512 feature vector for the characterization of the image patch. To reduce the dimensionality of the image feature space and eliminate redundancy, we conducted an extraction network optimization experiment. We aimed to retain a low number of highly informative features. The accuracy of patch pairing was the only metric used to measure the performance of feature extraction models. Overall, the highest accuracies were obtained by models trained on a maximum number of patch pairs per WSI and models utilizing convolutional feature compression (see Figure 4). Models using the dense feature compression layer in all scenarios resulted in a lower patch pairing accuracy. The accuracy of models trained on 100 patch pairs per WSI remained stable as the convolutional feature extraction layer was compressed from its original width of 1536 features down to 64 features. However, further compression to 32 features and below resulted in a rapid decline in accuracy. In this setting, the dense feature extraction resulted in an even earlier decline in patch pairing accuracy. Even though the accuracy of models trained on lower numbers of patch pairs per WSI was significantly lower in the whole range of extracted features, the effect of compression of the feature extraction layer was quite the opposite - in this setting, the patch pairing accuracy increased in the range from 1536 features down to 64 features retained.

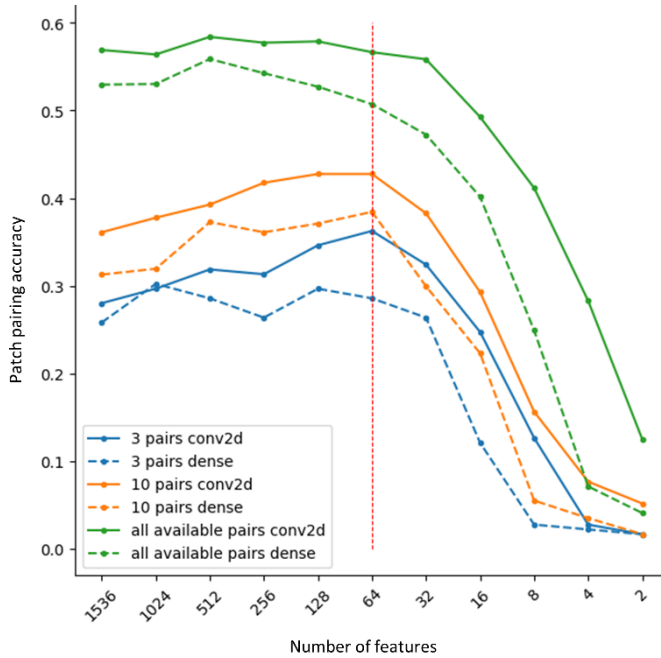


Figure 4. Feature extraction network optimization. The best results were achieved using all available patch pairs in comparison with three and ten pairs; also, the convolutional 2d architecture outperformed the dense layer architecture constantly. Decreasing the number of features to 64 did not hinder the performance of the network, while further decreases led to a fast deterioration in performance.

We ran the optimization experiments on a dataset prepared from 256 pixel-sized image patches. Based on these observations, for our further experiments, we utilized image feature vectors produced by models trained on 100 patch pairs per WSI using convolutional feature compression and a 64-feature-long image feature vector.

### Prediction of NMIPUC relapse by deep MIL

Table 3 summarizes the results of the cross-validation metrics for different image resolution deep MIL models in patient relapse prediction, as well as the log-rank statistics for survival differences between the two predicted groups.

Table 3. Performance of multiple instance learning models predicting relapse in 2-year and 5-year periods by different input image (patch) sizes

Patch size (pixels)	2-year relapse prediction			5-year relapse prediction		
	f-1 score	Accur acy	Log-rank p-value	f-1 score	Accur acy	Log-rank p-value
Multiresolution	0.618	0.622	0.208	0.422	0.446	0.916
1024	0.590	0.610	0.423	0.412	0.438	0.427
1024 resized to 256	0.654	0.672	0.257	0.476	0.492	0.486
512	0.572	0.566	0.134	0.472	0.492	0.579
512 resized to 256	0.592	0.592	0.323	0.494	0.502	0.613
256	0.626	0.620	0.441	0.480	0.490	0.268

The best results of 2-year relapse prediction were obtained with image features extracted from down-sampled 1024-pixel patches (resized to 256 pixels). Although this experimental setting allowed for the highest F1 score (0.654) and accuracy (0.672), the survival differences between the predicted groups were not statistically significant (log-rank p-value 0.257). The features extracted from the multiscale and the 256-pixel-sized patches also allowed reasonable prediction results, achieving both F1 scores and accuracies above 0.6. However, the relatively high log-rank p-values of 0.208 and 0.441, respectively, suggest lower predictive reliability.

The 5-year relapse prediction analysis did not reveal any significant results, with an accuracy above 0.5 achieved only using the features extracted from down-sampled 512-pixel patches (resized to 256 pixels). However, none of the experiments reached an F1 score of at least 0.5.

### Prediction of risk of NMIPUC relapse by deep MISL

Table 4. Performance of multiple instance survival learning models by different input image (patch) sizes

Patch size (pixels)	Concordance index	Log-rank p-value
Multiresolution	0.574	0.010
1024	0.564	0.007
1024 resized to 256	0.562	0.687

512	0.569	0.053
512 resized to 256	0.579	0.126
256	0.532	0.095

MISL results are summarized in Table 4. Using a multiresolution approach and features extracted from 1024-pixel-sized patches, we were able to stratify patients into risk groups with similar statistical significance in survival difference (log-rank p-value < 0.05), with the multiresolution approach showing slightly higher C-index. In general, features obtained from smaller patches (256-pixel size or larger patches resized to 256-pixel size) performed worse in survival prediction. In contrast, the MISL models trained with features extracted from downsized 512-pixel patches (resized to 256-pixel) achieved the highest C-index (0.579).

#### Clinicopathological variables and survival analysis

Significant relapse hazard differences were observed only by a re-TUR, tumor grade, and stage, with hazard ratios 5.018, 1.9902, and 1.8545, respectively (Table 5), with significant patient stratification on Kaplan Meier analysis (see Figure 5). These features were selected for multiple Cox regression together with stratified MISL prediction.

*Table 5. Univariate Cox regression results of clinicopathological data and stratified multiple instance survival learning prediction results*

Feature	Hazard ratio	p-value
Positive re-TUR	5.018	0.0001
pT1 stage	1.9902	0.0187
MISL prediction >-0.0082	2.1849	0.0237
G3 (1973 grading system)	1.8545	0.0451
High grade (2004 grading system)	2.6389	0.1807
Association with carcinoma in situ	1.8076	0.2586
EORTC high or very high-risk group	1.3548	0.3012
Multiple tumors	1.3395	0.3489
Recurrent tumor	1.255	0.4549
Muscle presence in TUR	1.2098	0.6668
Sex	1.1822	0.683
Tumor size >30 mm	1.1281	0.7376

Notably, tumor grade demonstrated a significant difference in relapse hazard stratification (p-value 0.0451) only when assigned using the WHO 1973 classification system, but not when assessed using the WHO 2004 system (p-value 0.1807).

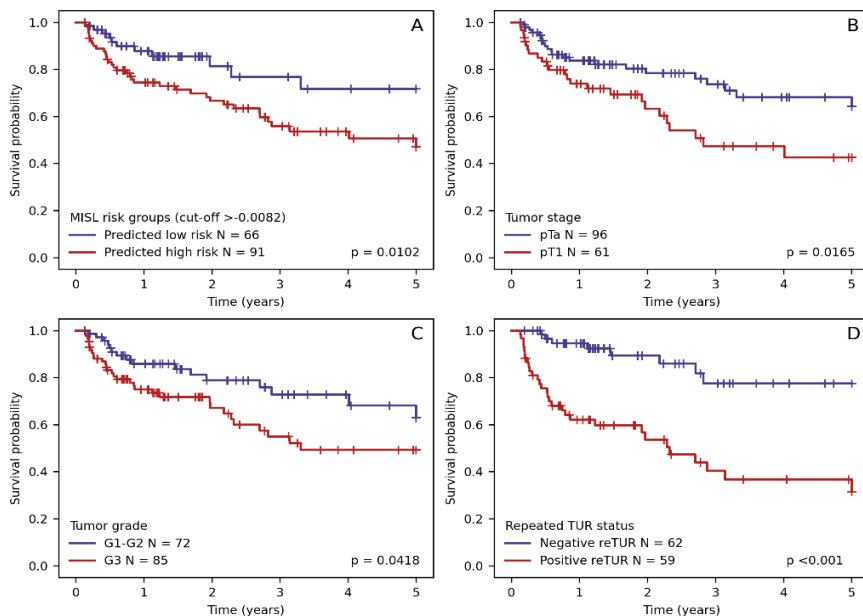


Figure 5. Kaplan-Meier RFS plots stratified according to: A. Stratified multiple instance survival learning (MISL) risk prediction results, B. Tumor stage (pTa vs. pT1), C. Tumor grade (1973 grading system), D. Presence of tumor in re-TUR.

We generated all possible combinations of selected features and evaluated them using multiple Cox regression. Three Cox regression models yielded significant individual features (p-values < 0.05). All three models (see Table 6) consisted of two independent predictive features combining MISL-based risk stratification with one of the selected clinico-pathologic features (stage, grade, and re-TUR status).

Table 6. Multiple Cox regression models with p-values of individual features < 0.05.

Features	Hazard ratio	95% CI	p-value
Model 1: AIC = 302.40; C-index = 0.73			
Positive re-TUR	4.907	2.245-10.726	<0.001
MISL prediction	2.181	1.058-4.499	0.035
Model 2: AIC = 418.59; C-index = 0.64			
G3 grade (WHO 1973)	2.026	1.105-3.716	0.023

MISL prediction	2.374	1.202-4.688	0.013
Model 3: AIC = 418.76; C-index = 0.63			
pT1 stage	1.969	1.109-3.495	0.021
MISL prediction	2.164	1.099-4.263	0.026

This finding prompted further investigation of the relationship between MISL prediction and histology-derived features, as well as an evaluation of MISL-based risk stratification within tumor grade and stage subgroups. We found MISL prediction able to successfully stratify patients in the pTa, pT1 stage, and G1-G2 grade groups. Additionally, we found a similar distribution of MISL predictions scalar value between tumor stage and grade groups ( $p = 0.876$  and  $p = 0.365$ , respectively), supporting independence of this feature (see Figure 6)

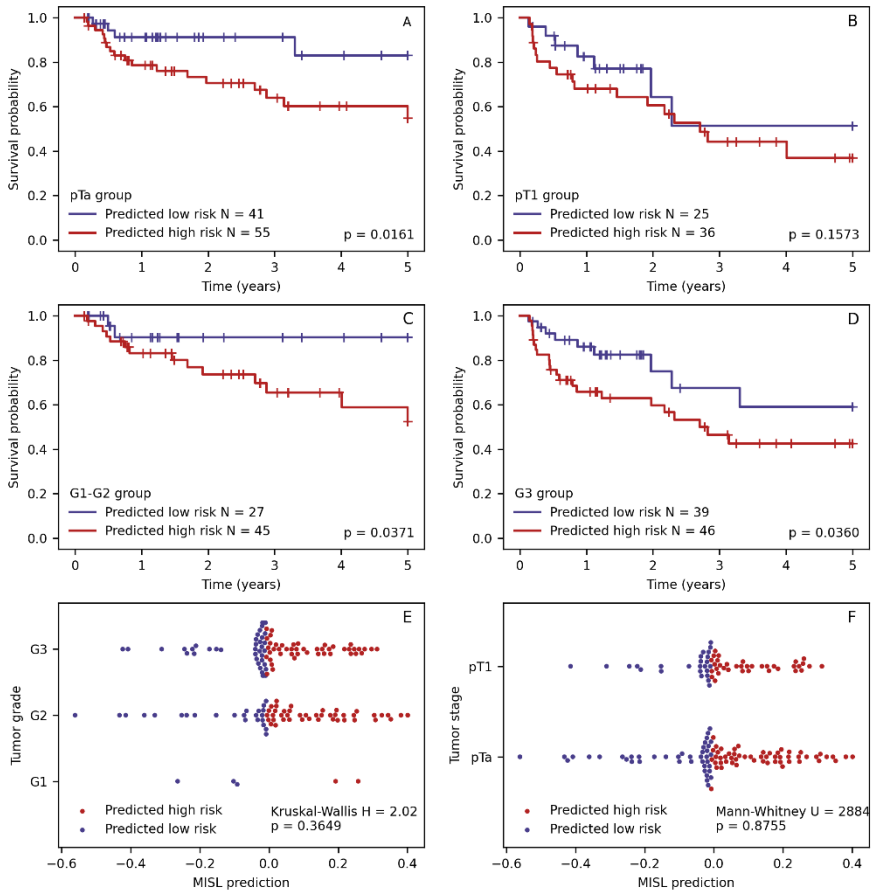


Figure 6. Kaplan-Meier RFS plots of patient survival stratified by multiple instance survival learning (MISL) in patient stage (A, B) and grade (C, D) subgroups. MISL prediction value distribution between tumor grade (E) and stage (F) groups.



## OBJECTIVE 2: ADAPTATION AND OPTIMIZATION OF IMMUNE RESPONSE INDICATORS

In this study, we have adapted the concepts of immunogradient indicators introduced by Rasmusson et al. to the specific histological architecture of NMIPUC, comparing the prognostic performance of these factors to immune cell densities in the overall NMIPUC and its compartments. We assessed the spatial distribution of CD8-positive T cells – the most widely used class of immune cells in the tumor microenvironment for cancer outcome prediction, which was also used by Rasmusson et al [67].

### Methods

Tissue classification was performed as discussed previously. The images and tissue classification masks were reviewed by a pathologist (JD), and epithelial region boundaries shorter than 1000  $\mu\text{m}$  were removed to reduce noise caused by small, incorrectly classified foci. The tissue classification was followed by CD8-positive cell segmentation and assessment of CD8-positive cell distribution in the stroma–epithelium interface zone using HALO® Multiplex IHC and Spatial Analysis modules (Indica Labs, Albuquerque, NM, USA), respectively. The spatial analysis was performed on a 150  $\mu\text{m}$  zone divided into 10  $\mu\text{m}$  width bands in both the epithelial and stromal sides, assigning ranks according to the distance from the epithelium–stroma interface (e.g., rank 1—covering area ranging from the interface (0 distance) to 10  $\mu\text{m}$ ; rank 2—from 10  $\mu\text{m}$  to 20  $\mu\text{m}$ ; rank 3—from 20  $\mu\text{m}$  to 30  $\mu\text{m}$ , etc.) with a negative or positive rank value assigned to the stromal or to epithelial aspect, respectively.

The immunodrop (ID) and center of mass (CM) density gradient indicators [30] were adapted to the results of the infiltration analysis. ID was calculated as the ratio of CD8-positive cell density in a corresponding pair of stromal and epithelial bands (e.g., ID (5) is the ratio between CD8 cell density in the stromal band with rank  $-5$  and the epithelial band with rank  $r$  where  $\rho$  represents CD8-positive cell density in the band, and  $r$  represents the rank of the band number.

$$ID = \frac{\rho_{-r}}{\rho_r}$$

CM was calculated as the ratio between the sum of the products of the indices and densities of bands in the IZ and the total sum of CD8-positive cell densities in the IZ as reported originally by Rasmusson et al [67].

$$CM = \frac{\sum r_i r_i \times \rho_i}{\sum r_i \rho_i}$$

Additionally, absolute CD8-positive cell densities were calculated in the stromal and epithelial compartments, as well as the overall IZ area. To determine the optimal width of the IZ, the absolute densities of CD8-positive cells and CM gradient indicators were assessed at various IZ widths and ranks, ranging from 20 to 300  $\mu\text{m}$ .

The dataset was randomly split into a training set (117 patients) and a hold-out test set (40 patients) with a similar proportion of patients with tumor recurrence in both sets. At this point, we had multiple variants of ID, CM, and absolute densities in tumor compartments. The variant with the lowest p-value was selected. For multivariable analysis, we selected factors from univariable analysis with a p-value  $<0.05$ . Multivariable Cox regression models using these constellations of variables were then fitted on the training set to select models consisting of independent variables (with p-values of HR for all covariates being  $<0.05$ ). On these selected models, we performed 5-fold cross-validation as described previously. We then selected the best-performing model and tested it on a hold-out test set. The Kaplan–Meier survival estimator was used to investigate the survival of patients stratified by factors with statistically significant associations with the recurrence hazard in the univariate Cox analysis. The log-rank test was used for pairwise comparison of patient survival between groups. For prognosticators with continuous data (ID and CM), we have arbitrarily stratified patients into three equal groups (low, medium, and high).

## Results

### Univariable Cox regression analysis

Tumor stage and G3 tumor grade were significantly associated with increased recurrence risk, as summarized in Table 7. Other traditional stratification prognosticators like tumor size, concurrent CIS, tumor multifocality, and demographic data (age and gender) did not show a statistically significant association.

Both CD8-positive cell density gradient indicators (CM and ID) were significantly associated with patient outcomes, whereas absolute CD8-positive cell densities in the stromal, epithelial, and overall IZ compartments did not show a significant association with RFS. The best-performing variation of ID included a ratio of ranks 10-20  $\mu\text{m}$  on stromal and epithelial

sides, suggesting that the changes closest to the epithelial-stromal interface are most indicative of patient outcomes. Worse performance of ID variant using band ranks next to the epithelial-stromal interface (0-10  $\mu\text{m}$ ) might be associated with minor inconsistencies in tissue classifier performance. Similarly, the best-performing CM measure was obtained from an IZ covering the 0-20  $\mu\text{m}$  interval from the interface on both epithelial and stromal aspects (ranks -2 to 2).

Table 7. Univariable Cox regression results

<b>Feature</b>	<b>p-value</b>	<b>HR</b>
Male gender	0.6229	0.7845
Age	0.4051	1.0000
Immunodrop	0.0031	12.2830
Center of mass	0.0082	0.0660
CD8 density epithelial	0.1140	0.9971
CD8 density stromal	0.2718	0.9993
CD8 density overall	0.1659	0.9979
pT1 stage	0.0126	2.6092
G1	0.9959	0.0000
G2	0.0757	0.4773
G3	0.0159	2.7387
Concurrent CIS	0.4793	1.5417
Tumor size > 30 mm	0.5781	0.7686
Multiple tumors	0.4050	1.3858
Positive re-TUR*	0.0009	3.6726
Recurrent tumor**	0.3955	1.3945
Positive re-TUR* or recurrent tumor	0.0016	5.4702
EORTC Intermediate risk	0.0655	0.4765
EORTC High risk	0.0766	1.9712
EORTC Very high risk	0.2071	2.5514

\*NMIPUC identified on repeated transurethral resection, \*\* not the primary NMIPUC identified on the first transurethral resection

Of medical history data, the positive re-TUR but not the recurrent tumor was associated with significantly increased recurrence hazard. Interestingly, the combination of these two factors formed an even stronger predictor of BCG failure (HR of patients having positive re-TUR and/or recurrent tumor was 5.4702 in comparison with only having positive re-TUR patients HR of 3.6726).

## Multivariable Cox regression analysis

A total of 11 multivariable Cox regression models could be obtained from the univariate prognostic features (Table 8). The best performing model from the 5-fold cross-validation included ID, G3 tumor grade, and positive re-TUR or recurrent tumor (

Table 9). Other models showed slightly higher AIC and lower mean C-index in the validation splits. Some of these models included CM, but without the cooccurrence of ID in the same model. The strongest models included the covariates of any medical history parameter, positive anamnesis of tumor (positive re-TUR and/or recurrent tumor), thus showing the high predictive value of this feature.

Table 8. Performance of multivariable Cox regression models (CM – center of mass, ID – immunodrop)

Model covariates	Mean validation set C-index	AIC
Positive re-TUR* or recurrent tumor** + G3 + ID	0.7837	173.3428
Positive re-TUR* or recurrent tumor** + G3	0.7397	174.6718
Positive re-TUR* or recurrent tumor** + ID	0.7370	174.7917
Positive re-TUR* or recurrent tumor** + pT1 + ID	0.7388	172.5348
Positive re-TUR* or recurrent tumor** + pT1	0.7355	174.4835
Positive re-TUR* or recurrent tumor** + CM	0.7308	174.9942
G3 + ID	0.7028	179.8105
G3 + CM	0.7028	179.8105
Positive re-TUR + ID	0.6613	178.3879
pT1 + CM	0.6551	180.2151
pT1 + ID	0.6438	178.4539

\*NMIPUC identified on repeated transurethral resection, \*\* not the primary NMIPUC identified on the first transurethral resection

The best performing model from the 5-fold cross-validation included ID, G3 tumor grade, and positive re-TUR or recurrent tumor (

Table 9) (log likelihood ratio = 22.76,  $p < 0.005$ ). Of note, this model included one anamnestic factor, one histological factor, and one tumor microenvironment factor. The C-index for the test set was 0.7429, slightly lower than the training set C-index of 0.7579, which excludes the possibility of overfitting this model.

Table 9. Best performing multivariable Cox regression model.

Covariate	p-value	HR
Positive re-TUR* or recurrent tumor**	0.0063	4.4492
G3	0.0457	2.3672
ID	0.0455	5.5072

\*NMIPUC identified on repeated transurethral resection, \*\* not the primary NMIPUC identified on the first transurethral resection

### Kaplan-Meier survival analysis

Non-parametric survival analyses once again supported our finding from univariable Cox regression analysis, with all six features separating patient groups with statistically significant differences in RFS (see Figure 7. Kaplan-Meier RFS plots stratified according to: A. Immunodrop (ID), B. Center of mass (CM), C. Presence of tumor in re-TUR, D. Presence of tumor in repeated transurethral resection (re-TUR) or recurrent tumor, E. Tumor stage (pTa vs pT1), F. Tumor grade (G3 vs G1-G2). The Low ID group of patients showed significantly longer RFS, similar to the high CM group, which was expected due to the strong inverse correlation of these indicators (data not shown). The traditional pathological factors, tumor stage, and grade, also separated groups with significantly different RFS. The presence of a tumor in re-TUR was associated with significantly shorter RS; however, the combined factor of recurrent tumor and/or presence of tumor in re-TUR extracted a larger group of patients with shorter RFS, resulting in a somewhat better balanced risk stratification.

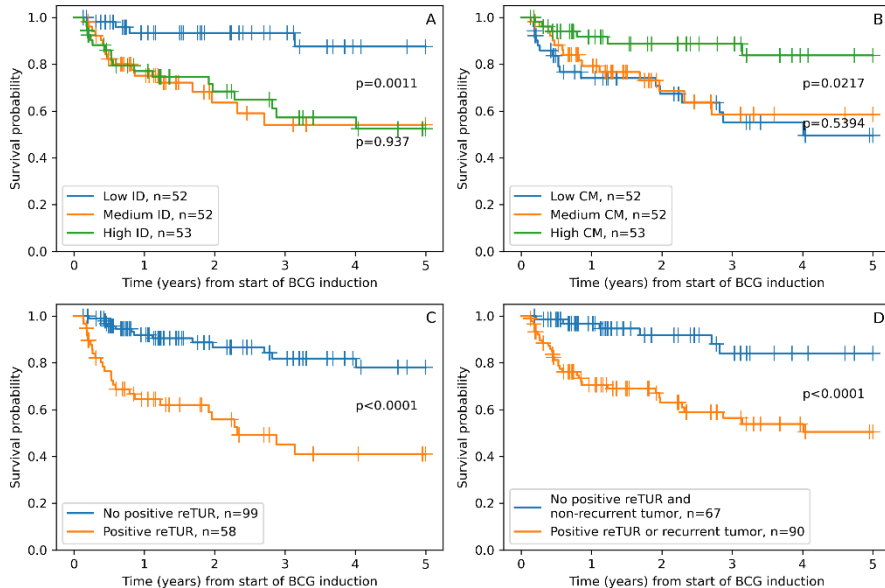


Figure 7. Kaplan-Meier RFS plots stratified according to: A. Immunodrop (ID), B. Center of mass (CM), C. Presence of tumor in re-TUR, D. Presence of tumor in repeated transurethral resection (re-TUR) or recurrent tumor, E. Tumor stage (pTa vs pT1), F. Tumor grade (G3 vs G1-G2).

### OBJECTIVE 3: EVALUATION OF PROGNOSTIC VALUE OF IMMUNE CELL SUBCLASSES AND THEIR DISTRIBUTION IN THE TUMOR MICROENVIRONMENT

In this study, we have assessed the prognostic significance of immune cell densities and their distribution in the epithelial-stromal interface, as well as the formation of tertiary lymphoid structures in the tumor tissue of NMIPUC. We analyzed CD8 and ICOS-positive populations of T cells, CD20-positive B cells, as well as CD11c and CD163-positive TAM populations representing the M1 and M2 classes, respectively.

#### Methods

##### Assessment of the spatial distribution of the immune cells

Immunohistochemical slide preparation, digitization, and tissue classification were performed as described previously.

To detect CD8, CD20, and ICOS-positive lymphocytes, we utilized the HALO® AI Multiplex IHC module. Because of the irregularities of the cytoplasm of macrophages, the CD11c and CD163 macrophages were segmented using the HALO® Nuclei seg classifier, followed by the HALO® Nuclei Phenotyper classifier.

We performed spatial analysis of the immune cell infiltrates using the HALO® AI Spatial Analysis module within the epithelial-stromal interface zone extending up to 150  $\mu\text{m}$  into epithelial and stromal compartments. Further, the width of the interface zone was decreased in 10  $\mu\text{m}$  increments to search for the most informative width for the predictive modeling. We generated a set of indicators to characterize the cell infiltration (density in the entire interface zone, stromal, and epithelial compartments) as well as cell density gradients distribution across the epithelial-stromal interface.

To assess the immune cell gradients (two illustrative cases are shown in **Error! Reference source not found.**) we have decided to do a simpler assessment, which is measured as the ratio of the cell density in the epithelial compartment and that in the stromal compartment of the interface zone. We have called this feature the infiltrate density ratio (IDR) across the epithelial-stromal interface.

$$IDR = \frac{\rho_{\text{epithelial}}}{\rho_{\text{stromal}}}$$

We have considered this approach easier to replicate and a more parsimonious way of assessing immune cell distribution in the interface zone.

Similar to absolute density assessment, to optimise the interface zone width, we calculated IDR in interface zone widths ranging from 10  $\mu\text{m}$  to 150  $\mu\text{m}$  (in 10  $\mu\text{m}$  increments) in the epithelial and stromal sides.

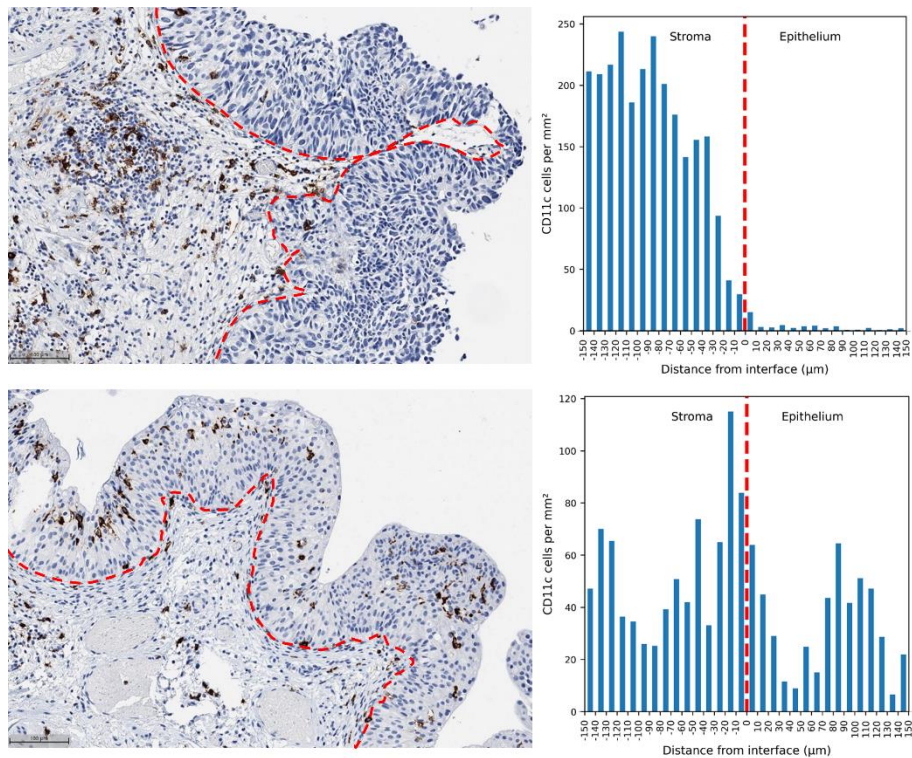


Figure 8. CD11c-positive cell infiltration density profiles in the NMIPUC of two selected patients are shown to illustrate different infiltration patterns. Upper panel, patient A with predominant infiltration in the tumor stroma (IDR = 0.074): a representative IHC image area (left) and a bar plot (right) representing the CD11c cell density distribution across the interface. Lower panel, patient B with relatively higher CD11c cell density in the epithelium (IDR = 0.424). The red dotted line represents the epithelial-stromal interface. 100  $\mu\text{m}$  measure added for the reference. Negative distances in the plot represent the stromal aspect of the interface zone.

The presence of TLS was assessed using a tissue classifier on CD20 IHC images. The cases with any amount of tissue classified were considered as having present TLS. We assessed it as a binary feature. And for the assessment of CD20-positive lymphocyte distributions, we excluded areas of TLS.

### Statistical analysis

To determine the optimal width of the interface zone for each indicator, we conducted univariate Cox regression using 50 iterations of 10-fold cross-

validation. This approach helped us obtain stable results, and we used the mean concordance index (C-index) from the validation splits as our performance metric.

Univariate Cox regression performed on the full dataset was used to select the most relevant features for multiple Cox regression. Along with the clinical and pathology parameters, we tested tumor microenvironment aspects, including the presence of TLS. We selected the features with a p-value of less than 0.2 and employed them in multiple Cox regression models. From all possible feature combinations, models with independent prognostic significance of features ( $p < 0.05$  of all features in the model) were selected for further analysis. To assess the performance of these models, we used 10-fold cross-validation.

We performed a Kaplan-Meier survival analysis with the features included in selected multivariate models to assess their impact on patient risk stratification regarding NMIPUC recurrence. The statistical significance of the results was assessed using the log-rank test, with a p-value of 0.05 used as the threshold. For continuous variables, we used the median value as a cut-off to stratify the data.

## Results

### Optimization of IZ width

Immunohistochemical (IHC) assessment of tumor tissue by digital image analysis revealed variable patterns of the immune cell distributions at the epithelial/stromal interface. They ranged from cases with predominant infiltration in the stroma represented by low IDR (Figure 8. Patient A) to cases with similar density in the epithelial and stromal aspects of the interface, represented by high IDR (Figure 8. Patient B).

The results of the optimization experiments are presented in Table 10. CD11c epithelial-stromal interface density ratio (IDR) achieved the highest C-index on validation splits in the cross-validation. CD8 IDR was the second-best feature; IZ of 50 $\mu$ m thickness was found optimal for both cell subtypes. In general, IDR performed better than cell densities in the entire interface zone (IZ) or in any other (epithelial or stromal) compartment of the tumor microenvironment.

Table 10. The optimal interface zone width for the cell distribution features according to the mean concordance index (C-index) in the cross-validation. Immune cell infiltrate density ratio (IDR) across the epithelial-stromal interface

<b>IHC marker</b>	<b>Feature</b>	<b>Optimal IZ width (<math>\mu\text{m}</math>)</b>	<b>C-index mean (SD)</b>
CD8	Total density	10	0.591 (0.157)
	Stromal density	10	0.574 (0.156)
	Epithelial density	130	0.62 (0.146)
	IDR	50	0.637 (0.146)
CD20	Total density	150	0.488 (0.162)
	Stromal density	150	0.503 (0.152)
	Epithelial density	140	0.51 (0.149)
	IDR	150	0.566 (0.151)
ICOS	Total density	10	0.503 (0.1417)
	Stromal density	10	0.504 (0.141)
	Epithelial density	30	0.513 (0.142)
	IDR	40	0.583 (0.162)
CD11c	Total density	10	0.518 (0.148)
	Stromal density	10	0.493 (0.147)
	Epithelial density	140	0.572 (0.157)
	IDR	50	0.64 (0.134)
CD163	Total density	40	0.51 (0.149)
	Stromal density	10	0.427 (0.132)
	Epithelial density	70	0.567 (0.142)
	IDR	140	0.603 (0.149)

Of note, for CD8, ICOS, and CD11c, the optimal IZ was narrower (50  $\mu\text{m}$ , 40  $\mu\text{m}$ , and 50  $\mu\text{m}$ , respectively) than in CD20 and CD163 (150  $\mu\text{m}$  and 140  $\mu\text{m}$ , respectively). However, the latter two features showed a lower mean C-index overall.

#### Univariate Cox regression for prediction of RFS

Results of univariate Cox regression of features predictive of RFS with a p-value below 0.2, selected for multiple Cox regression, are summarized in Table 11.

Table 11. Results of univariate Cox regression with p-value lower than 0.2. IDR - immune cell infiltrate density ratio across the epithelial-stromal interface. "Density stromal", "density epithelial", and "density total" correspond to the density in the specific compartment of the interface zone or the total area of the interface zone.

<b>Feature</b>	<b>HR</b>	<b>p-value</b>
Positive re-TUR	4.9321	0.0001
CD11c IDR	0.0012	0.0002
CD8 IDR	0.0379	0.005
pT1	2.0445	0.0159
ICOS IDR	0.0768	0.0388
G3 grade (WHO 1973)	1.8254	0.0522
CD163 IDR	0.0712	0.0549
CD8 density total	0.9984	0.0648
CD8 density epithelial	0.996	0.0857
CD8 density stroma	0.9988	0.0988
Tertiary lymphoid structures	1.6915	0.1033
ICOS density epithelial	0.9963	0.1375
High grade (WHO 2004)	2.5873	0.1899

Patients with a history of positive re-TUR (repeated transurethral resection) had a significantly higher hazard ratio (HR) of 4.93 (p-value = 0.0001). Also, stage and grade emerged as potential prognostic factors. Patients with pT1 stage tumors (HR: 2.04, p-value = 0.0159) and high-grade tumors, according to the 1973 WHO classification (HR: 1.83, p-value = 0.0522), had a higher risk of recurrence, although the latter association was of borderline significance.

Several immune cell indicators revealed potential prognostic value. Low CD11c IDR (HR: 0.0012, p-value = 0.0002) and low CD8 IDR (HR: 0.0379, p-value = 0.005) were associated with an increased risk of recurrence. Conversely, no significant association was observed for the density of these cells in either the epithelial or the stromal compartments. Stage and grade emerged as potential prognostic factors. Patients with pT1 stage tumors (HR: 2.04, p-value = 0.0159) and G3 grade according to the 1973 WHO classification (HR: 1.83, p-value = 0.0522) had a higher risk of recurrence, although the latter association was of borderline significance.

We have identified TLS in the TUR specimens of 100 patients (65.5%). The presence of TLS did not show a statistically significant association with recurrence (HR: 1.69, p-value = 0.1033). Similarly, high grade, according to the 2004 WHO classification and ICOS density in the epithelium, did not reach statistical significance (p-values > 0.1); nevertheless, we tested them in the multivariate analysis (see below).

## Multiple Cox regression

Multiple Cox regression analyses generated several models with moderate to good discriminative ability, with a mean C-index ranging between 0.59 and 0.74. The best-performing model (see Table 12) consisted of two factors that were also the best-performing in the univariate Cox regression analysis, namely, a history of positive re-TUR and CD11c IDR, with a mean C-index of 0.7427. Another strong model (C-index > 0.7) included CD11c IDR and pT1 tumor stage. Interestingly, CD11c was included only in these two models, while re-TUR status was included only in the best-performing model.

*Table 12. Two Cox regression models with a concordance index >0.7. IDR - infiltrate density ratio across the epithelial-stromal interface.*

<b>Features</b>	<b>Hazard ratio</b>	<b>95% confidence interval</b>	<b>p-value</b>
<b>Model: positive re-TUR + CD11c IDR</b>			
Positive re-TUR	4.3411	1.9616-9.6072	<0.001
CD11c IDR	0.0282	0.00097-0.824	0.038
<b>Model: pT1 stage + CD11c IDR</b>			
pT1 stage	2.2524	1.2449-4.075	0.007
CD11c IDR	0.00067	0.000017-0.268	<0.001

The weaker models included all features with p-values < 0.05 from the univariate analysis, as well as the presence of TLS, G3 tumor grade, CD163 IDR, and CD8 cell density in the epithelial compartment (data are summarized in Table 13).

Table 13. Results of multivariate Cox regression with a p-value of individual features lower than 0.05. IDR - immune cell infiltrate density ratio across the epithelial-stromal interface. TLS – presence of tertiary lymphoid structures

<b>Model</b>	<b>AIC</b>	<b>Mean C-index</b>
CD11c IDR + Positive re-TUR	257.8785	0.7427
CD11c IDR + pT1	338.225	0.703
CD8 IDR + TLS	352.2153	0.6449
ICOS IDR + G3	361.1213	0.6364
CD8 IDR + pT1	354.4931	0.6308
CD163 IDR + TLS	358.9342	0.6204
ICOS IDR + TLS + G3	356.6093	0.6143
CD163 IDR + G3	359.7073	0.6102
CD8 density total + TLS	360.0685	0.6084
ICOS IDR + pT1 + TLS	357.7706	0.6078
ICOS IDR + pT1	359.8559	0.6019
ICOS density epithelial + G3	360.1979	0.5974
CD163 IDR + pT1	358.9573	0.5972

#### Kaplan-Meier RFS analysis of the selected features

The greatest difference between the risk groups of patients, similar to the results of univariate Cox regression, was observed in cases with positive restaging transurethral resection (re-TUR) and CD11c IDR ( $p < 0.001$ ). CD8 IDR, ICOS IDR, and tumor stage also showed statistically significant stratification of patient relapse risk ( $p < 0.05$ ), which was also significantly associated with patients' RFS in univariate Cox regression. Additionally, G3 tumor grade, stratified CD163 IDR, and absolute CD8 cell density reached statistically significant ( $p < 0.05$ ) stratification in the Kaplan-Meier analysis of RFS (see Figure 9). This can be attributed to their non-linear effect on the risk of NMIPUC relapse. The presence of TLS showed a tendency for prognostic stratification, but it did not reach statistical significance ( $p = 0.0996$ ). On the other hand, other features, such as the WHO 2004 grade, ICOS density in the epithelial compartment, and CD8 density in the stromal compartment, showed worse performance, paralleling the findings of the multivariate Cox regression analysis.

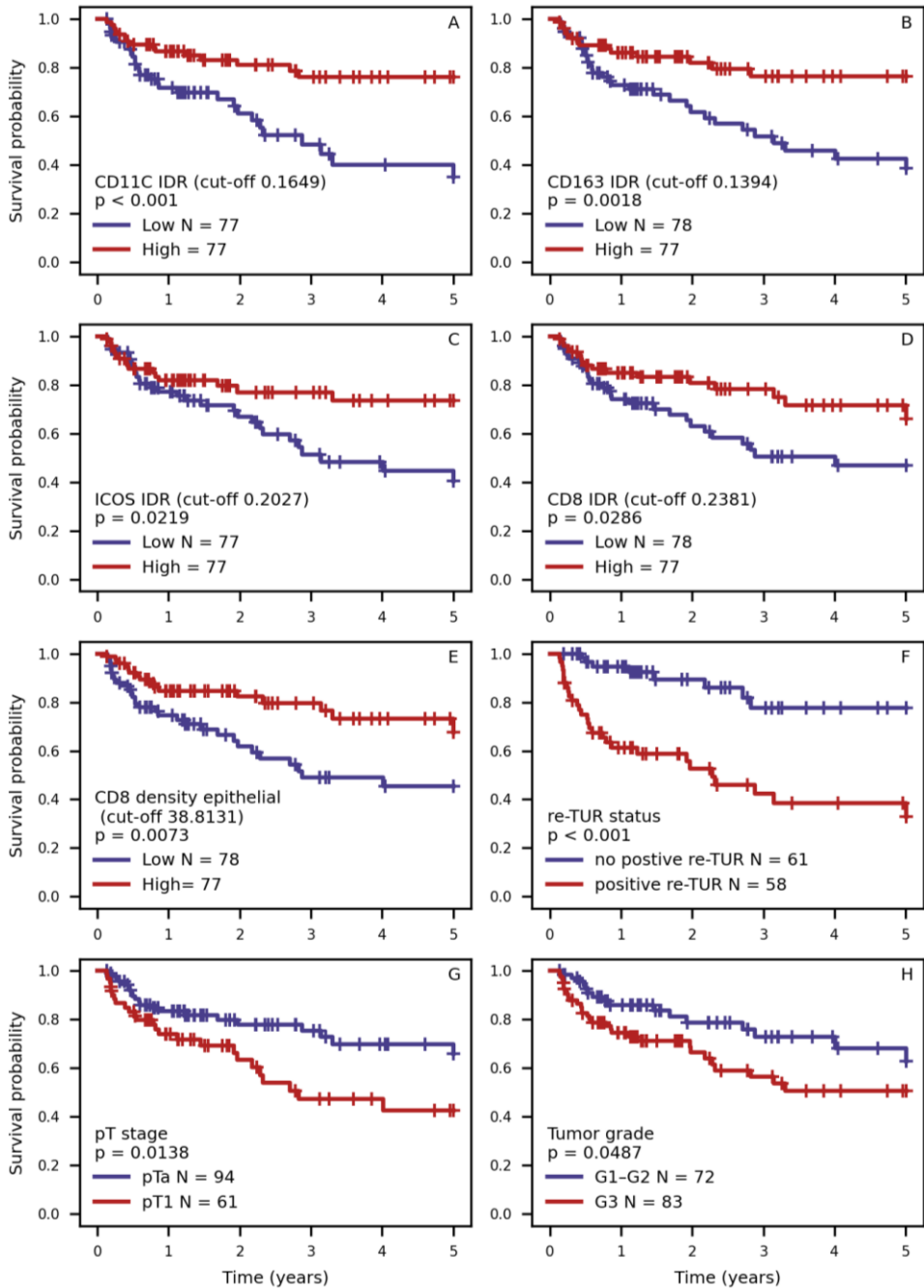


Figure 9. Kaplan-Meier RFS plots stratified on: A. CD11c infiltrate density ratio (IDR) across the epithelial-stromal interface, B. CD163 IDR, C. ICOS IDR, D. CD8 IDR, E. Density of CD8 cell in the epithelial compartment of interface zone, F. Presence of tumor in repeated TUR (re-TUR), G. Presence of tumor in repeated transurethral resection (re-TUR), H. Tumor grade. Continuous variables (plots A through E) are stratified according to the median value of the indicator.

## OBJECTIVE 4: DEVELOPMENT OF A COMPREHENSIVE MODEL

In this study, we explored the possibility of integrating well-established clinicopathologic variables with our previously developed prognostic features to determine whether DL-based image analysis features and tumor microenvironment features can be combined to improve patient outcome prediction.

### Methods

We performed univariate Cox regression analysis in our previous study (task 3) to evaluate the prognostic performance of immune microenvironment variables, as well as pathologic variables (tumor stage and grade). We used positive re-TUR status as a feature for all patients (as in task 2), even if re-TUR was not performed (coded as 1 if the patient had a tumor on re-TUR, and 0 if re-TUR was tumor-free or the procedure was not performed), since the re-TUR procedure is not indicated for all patients; thus, including only patients who underwent re-TUR would exclude a subgroup of patients with a relatively lower risk. We used an aggregated patient history feature coded as 1 if the patient had either a tumor at re-TUR or a documented tumor recurrence, and 0 otherwise (no positive re-TUR as described above, and no tumor recurrence). We also combined tumor stage and grade into a single feature (pT1 or G3), as they were strongly dependent ( $\chi^2 = 23.912$ ,  $p < 0.001$ ) and did not appear as independent variables in previous multiple Cox regression models. Univariate Cox regression models were fitted to the dataset with 155 patients (see Materials and Methods), which was also used in task 3. Features with p-values less than 0.1 were selected for further multivariate analysis. Multivariate models were constructed and fitted as described previously, and the models with independent covariates (with HR p-values  $< 0.05$ ) were then subjected to 10-fold cross-validation (see Materials and Methods). We selected the best-performing model based on the highest mean C-index across cross-validation test splits. To assess the performance of patient risk stratification of this model, we predicted the relative risk for all patients, organized them into risk groups, and evaluated the survival difference between groups using the log-rank test.

## Results

### Univariate Cox regression

*Results of univariate analysis are summarized in*

Table 14. While positive re-TUR retained the lowest p-value, when combined with tumor recurrence status prior to treatment, it showed a higher tumor recurrence hazard (HR = 5.2528). Interestingly, when combined, tumor grade and stage showed a higher HR (2.5934) and a lower p-value (0.0078) in comparison to stage and grade individually. MISL prediction demonstrated slightly lower HR (2.1554 vs. 2.1849) and slightly higher p-value (0.0267 vs. 0.0237) compared to task 1 results, which was likely due to a slight difference between the datasets.

*Table 14. Results of univariate Cox regression with p-value lower than 0.1. IDR - immune cell infiltrate density ratio across the epithelial-stromal interface. "density stromal", "density epithelial", and "density total" correspond to the density in the specific compartment of the interface zone or the total area of the interface zone.*

<b>Feature</b>	<b>HR</b>	<b>p-value</b>
Positive re-TUR	3.7953	<0.0001
Positive re-TUR or recurrent tumor	5.2528	0.0001
CD11c IDR	0.0012	0.0002
CD8 IDR	0.0379	0.005
Combined pT1/G3	2.5934	0.0078
pT1	2.0445	0.0159
MISL prediction	2.1554	0.0267
ICOS IDR	0.0768	0.0388
G3 (WHO 1973)	1.8254	0.0522
CD163 IDR	0.0712	0.0549
CD8 density total	0.9984	0.0648
CD8 density epithelial	0.996	0.0857
CD8 density stromal	0.9988	0.0988

### Multivariate Cox regression

Features selected for multivariate analysis included all features selected in Task 3, as well as DL-based (MISL) prediction, and features that combined tumor stage with grade and re-TUR status with prior recurrence status. A total of 8178 feature combinations were tested, resulting in 41 models with

independent covariates. Results for the 10 best-performing models, based on mean C-index across cross-validation test splits, are summarized in the Table 15.

*Table 15. Results of multivariate Cox regression with a p-value of individual features lower than 0.05. IDR - immune cell infiltrate density ratio across the epithelial-stromal interface.*

<b>Model</b>	<b>AIC</b>	<b>Mean C-index</b>
CD11c IDR, MISL prediction, Positive re-TUR, Combined pT1/G3	325.2265	0.8096
CD11c IDR, Positive re-TUR, Combined pT1/G3	328.2969	0.7769
MISL Positive re-TUR or recurrent tumor	267.0194	0.7714
CD11c IDR, MISL prediction, Positive re-TUR	329.2075	0.7640
CD11c IDR, Positive re-TUR or recurrent tumor	265.8861	0.7596
pT1', CD11c IDR, Positive re-TUR	330.76	0.7573
CD8 IDR, MISL prediction, Positive re-TUR, Combined pT1/G3	339.3849	0.7569
MISL prediction, Positive re-TUR, Combined pT1/G3	340.473	0.7529
CD11c IDR, Positive re-TUR	332.134	0.7295
MISL prediction, Positive re-TUR	342.7532	0.7278

The best-performing models most commonly included CD11c IDR from microenvironment features, as in task 3. A combination of re-TUR and recurrence status appeared only in two-variable models, which, while showing lower mean AIC scores, were outperformed by larger models that included only re-TUR status as an individual variable. A combination of tumor grade and stage was included in the best-performing models, whereas the individual features appeared only in weaker models.

The best-performing model (mean AIC = 325.2265; mean C-index = 0.8096; see Table 16), together with CD11c, re-TUR status, and combined tumor grade and stage, included MISL prediction, supporting the independent impact of tumor microenvironment, classical pathology, clinical, and tumor histology image features on patient outcomes.

Table 16. Summary statistics of the best-performing model. IDR - immune cell infiltrate density ratio across the epithelial-stromal interface. MISL – multiple instance survival learning.

Feature	Hazard ratio	95% confidence interval	p-value
CD11c IDR	0.0057	0.0002 - 0.1680	0.0028
MISL prediction	2.2613	1.1370 - 4.4974	0.02
Positive re-TUR	2.9169	1.5534 - 5.4769	0.0009
Combined pT1/G3	2.3481	1.1538 - 4.7785	0.0185

### Kaplan-Meier analysis

Initially, we organized the patients into quartiles according to predicted relative recurrence risk (see Figure 10) using the best-performing model. This resulted in a statistically significant separation of the low-risk group (pair-wise log-rank p-value = 0.045 between Q1 and Q2 groups) and high-risk group (pair-wise log-rank p-value = 0.0026 between Q3 and Q4 groups), while Q2 and Q3 quartiles did not show a statistically significant difference.

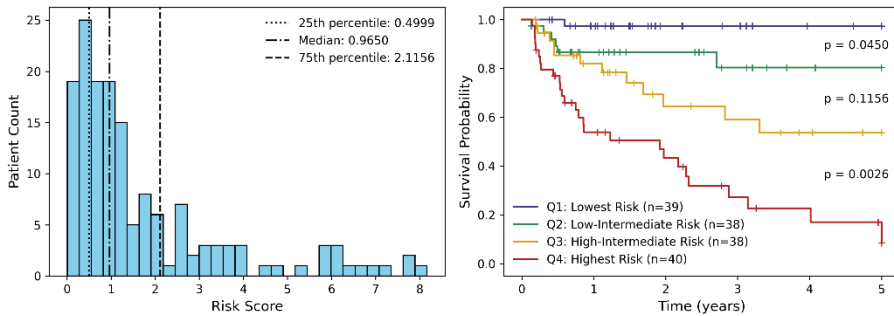


Figure 10. Left - the histogram of the distribution of patients according to the predicted relative recurrence risk with cut-off values separating the cohort into quartiles. Right – The Kaplan-Meier RFS plot with patient risk groups (quartiles) Q1 through Q4 from lowest risk to highest.

However, after initial analysis, we observed a bimodal distribution of risk scores, which was confirmed by statistical testing (bimodality coefficient = 0.7495). Thus, we stratified patients using an optimal cut-off value determined by maximizing the silhouette score (cut-off value = 4.5401; maximum silhouette score = 0.7578; Ashman's D = 3.8462). This stratification (see fig.) separated a smaller group of patients with an extremely bad prognosis (median recurrence-free survival time 216 days).

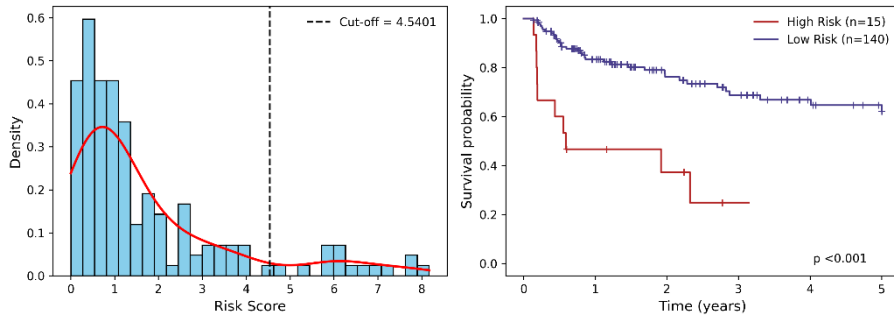


Figure 11. Left - the histogram of the distribution of patients according to the predicted relative recurrence risk, with the KDE curve highlighting the bimodal distribution and optimal cut-off value. Right - the Kaplan-Meier RFS plot with patient risk groups (high vs low) based on the optimal cut-off value (4.5401).

Additionally, we have performed an overall survival (OS) analysis (see Figure 12). The initial stratification of patients into quartiles on the multivariate log-rank test failed to show statistically significant stratification ( $p$ -value = 0.0739, see Figure 12 panel A). However, when comparing the isolated Q1 group with the combined Q2-Q4 groups, the survival difference was statistically significant (see Figure 12 panel B). The survival difference between Q4 and the combined groups Q1-Q3 was not statistically significant, while the high-risk group defined by the predicted relative risk score of more than 4.5401 showed significantly worse OS (median survival time 5.56 years) in comparison to the low-risk group (median survival time 10.5 years).

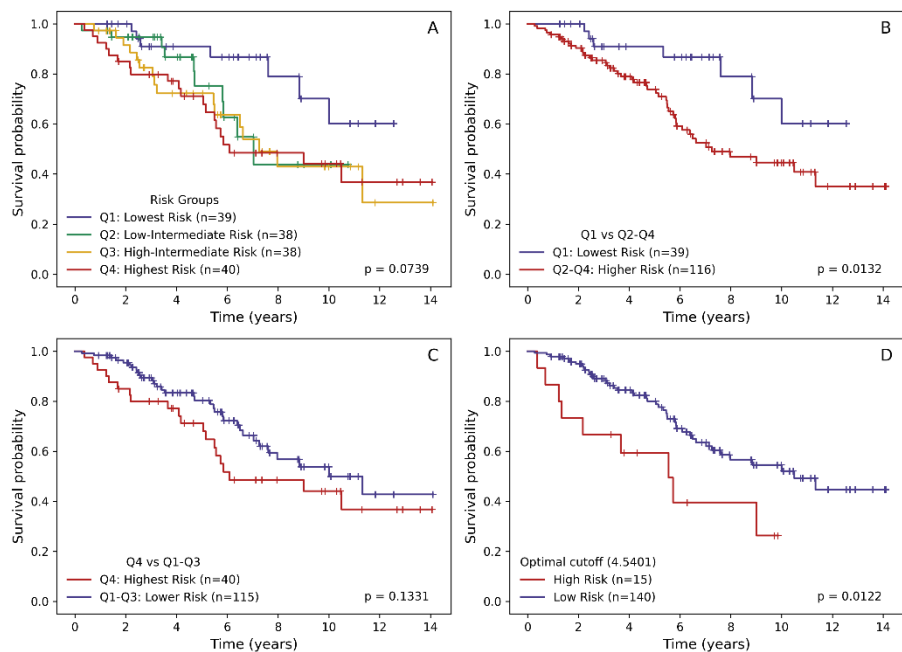


Figure 12. Kaplan-Meier plots for overall survival (OS). A. Kaplan-meier plot for OS according to risk groups (Q1 through Q4). B. Kaplan-Meier plot of OS by risk groups (Q1 vs Q2-Q4). C. Kaplan-Meier plot of OS by risk groups (Q4 vs Q1-Q3). D. Kaplan-Meier plot of OS by risk groups defined by the optimal cut-off value (low risk vs high risk).

## DISCUSSION

The primary aim of this study was to explore whether digital pathology combined with AI can improve prognostic stratification for patients with NMIPUC treated with BCG immunotherapy. Current risk assessment models rely heavily on clinicopathological features, which often fail to capture the complexity of tumor biology and microenvironment. By combining DL-based image analysis with spatial immune profiling, this research aimed to develop more precise, clinically relevant predictive models.

Deep learning models outperformed traditional methods in predicting recurrence risk, with multiresolution analysis and MISL architecture showing the best results. Standard clinical factors, such as stage and grade, remained important, while spatial immune indicators, especially CD8 and CD11c gradients, added independent prognostic value. Combining DL-derived features with clinical, pathologic, and immune data achieved the highest risk-stratification accuracy.

### Clinical and pathologic variables

While it was not the focus of our studies, from clinical and pathological variables, tumor stage, G3 tumor grade defined by the 1973 WHO system, and the finding of residual tumor at re-TUR, showed prognostic significance in both univariate and multivariable models. However, patient age, sex, association with CIS, tumor size, and multifocality did not show a significant association with patients' outcomes. In part, this might be explained by the limited size of our cohort and the exclusion of non-papillary tumors from our studies, resulting in a lower proportion of patients with CIS.

Positive re-TUR in combination with a history of recurrence or as a standalone feature remained the strongest predictor of patient outcomes in all our studies. Its prognostic significance was already established by other studies [81–83]; however, it is not utilized as a prognostic indicator, despite strong evidence supporting its use.

The difference in accuracy of prognostic performance between the 1973 WHO and the 2004/2016 WHO grading systems may be explained by the fact that the threshold between G2 and G3 grades defined by the 1973 WHO system might be more relevant in our cohort of patients than the groups defined by the 2004/2016 WHO. Since our study primarily includes patients from intermediate or higher risk groups, the latter system separates only a small subgroup of low-grade NMIPUC tumors. This observation supports the

current NMIPUC grading approach by reporting both grading systems in the diagnostic workup, particularly in instances where further subgrouping of high-grade tumors is of importance.

Interestingly, the combination of tumor grade and stage performed better in task 4, the development of a multiparametric model. This aligns with current risk assessment methods, where the WHO 1973 G3 grade or pT1 stage is typically used in combination. Thus, including these features as a combined rather than concurrent variable increases their applicability in patient risk stratification.

### DL-based risk assessment

In study 1, we developed a DL-based NMIPUC risk stratification model to enhance the prediction of tumor relapse during BCG immunotherapy. Comparing two DL approaches for patient risk prediction, we found that models utilizing time-dependent survival probability data outperformed those based on a simple dichotomous prediction of relapse within a specific time period. In other words, survival data, including relapse status and the exact timing of the event, offer a more precise definition of patient outcomes, thereby improving the prediction of the tumor's more aggressive behavior.

By comparing MIL models that predict tumor relapse in 2- and 5-year periods, we found that models based on predicting relapse in a shorter period performed better, despite having fewer events. This observation contradicts the results reported by Lucas et al. [29], who used deep learning to predict relapse in 1-year and 5-year periods. Our findings may be attributed to the fact that more aggressive tumors tend to relapse earlier, and these tumors might have a higher degree of architectural and cytological atypia at the histological level, which makes these tumors more straightforward to identify with image analysis.

Comparison of the models based on different patch sizes revealed that experiments with larger patch sizes and multiresolution approaches generally performed better. The performance advantage of the multiresolution method may be attributed to a more comprehensive representation of tumor tissue properties, which extracts both cytological details and architectural features at lower and higher resolutions, respectively. On the other hand, the difference in performance of single-resolution inputs might indicate that architectural features (large patches) might be more representative of tumor biological behavior.

The values of the MISL multiresolution approach did not show a significant difference in distribution between the stage and grade groups of patients; therefore, they are not associated with these known indicators and are likely to represent a novel computational biomarker for predicting relapse. Furthermore, the stratification of patients by the MISL indicator in most subgroups (see Figure 6) retained significant differences in relapse probabilities. Overall, this highlights the potential to improve risk stratification based on histology image analysis data.

Despite the promising results of our work in DL-based outcome prediction due to rapid advancement in DL-based DIA, their applicability is increasingly limited. Our approaches rely on patch-level feature extraction and simple attention-based aggregation, which fail to capture the full complexity of histological context and spatial relationships within WSI. Recent advances have introduced pathology-specific foundation models, such as UNI [84], Virchow [85], and GPFM [86], which leverage large-scale pretraining on histopathology datasets to learn domain-specific representations that outperform generic CNN-based models. Furthermore, state-of-the-art aggregation techniques now incorporate spatially aware mechanisms and graph-based learning, enabling models to represent tissue architecture and cellular interactions more effectively [87]. These innovations have demonstrated superior performance, highlighting the need to transition from traditional DL pipelines toward more modern approaches.

#### Adaptation of tumor-host interaction indicators

Our study reveals that CD8-positive cell density gradient indicators, ID and CM, were significantly associated with RFS of patients treated with BCG immunotherapy for NMIPUC, highlighting the importance of the spatial distribution of CD8-positive cells across the tumor interface. Similar work was published recently by Bieri et al. [74,75] explored the prognostic significance of TILs in bladder cancer by introducing the mIS concept for the assessment of TILs in the tumor tissue. In both of their studies, mIS enabled significant risk stratification only in subsets of patients (progression-free survival and cancer-specific survival stratification in MIBC after cystectomy - the AJCC stage IIIa group, RFS stratification in NMIPUC treated with BCG – EORTC high-risk group). In contrast, the CD8-positive density gradient indicators provided significant stratification in the entire patient cohort.

An important advantage of our method is that it generates the CD8-positive cell density data from the epithelium-stroma interface with high selectivity

and capacity, also maintaining the spatial context of the tumor-host interaction area. In univariate analyses, the best-performing variants of both CM and ID were generated from the IZ within a range of 20  $\mu\text{m}$ , extending into both the stromal and epithelial aspects. DIA performed a precise selection of areas of interest, paired with the high-throughput nature of the method in WSI, which enabled an optimized solution to assess the tissue immune response in this tumor with its peculiar papillary microarchitecture. In contrast to other tumor types, where immunogradient indicators and Immunoscore were found to be prognostic, NMIPUC, in most cases, does not have a wide invasive border; instead, a tumor-host interaction takes place in very thin, elongated papillary tumor structures, requiring a more delicate approach. Our study demonstrates that this can be achieved through AI-based pixel-level tissue classification, followed by computational immunohistochemistry assessment.

Another important observation emerging from our study is that CD8-positive cell density gradient indicators were significant prognosticators of RFS, while none of the absolute CD8-positive cell densities in any tumor tissue compartment showed a significant impact. This further supports the importance of spatial analytics in studying the tumor microenvironment, rather than relying solely on the quantification of cell densities in tumor tissue compartments. Whereas all patients in our cohort received BCG immunotherapy, our computational models enable assessment of RFS probability and, with appropriate study design, can be tested as predictive biomarkers for immunotherapy modalities.

#### Prognostic value of immune cell subclasses

In this study, we report computational models to predict NMIPUC relapse after BCG immunotherapy based on standard clinicopathologic factors and tumor-infiltrating immune cell densities and their density ratios (gradients) across the epithelial-stromal interface. Remarkably, we found that, in general, the ratio of immune cell density gradients was more predictive of the disease recurrence than absolute immune cell densities in any tumor tissue compartment or the entire interface zone. Nevertheless, the extended investigation has shown that the CD11c gradient IDR was the best performing in the multivariate prognostic models, in combination with re-TUR and tumor stage. Meanwhile, CD20 and CD163 cell densities, or their gradient ratio, did not reveal significant prognostic associations in our dataset.

CD11c gradient density ratio was the strongest prognostic feature from the immune tumor microenvironment features, followed by the CD8 gradient

IDR. CD11c is expressed in M1-polarized macrophages and has been previously used in renal cell carcinoma [28], hepatocellular carcinoma [29], and breast carcinoma [30] to assess M1 macrophage infiltration. These studies revealed a higher density of CD11c-positive macrophages, which is associated with better patient outcomes. Dendritic cells also express CD11c, and multiplex immunotyping of the CD11c-positive cell population could enhance the specificity of TAM assessment. However, according to the study by Ayari et al. [52], macrophages and dendritic cells have different distributions in NMIBC, and the narrow (50  $\mu$ m) interface zone used in our study predominantly covers the tissue area infiltrated by TAM. Nevertheless, our finding may be influenced by a minor component of CD11c-positive dendritic cells, as these cells play a crucial role in orchestrating antitumor immunity. Both CD11c-positive M1 macrophages and CD8-positive cytotoxic T lymphocytes are directly involved in anti-tumor immune response. Thus, their positive gradients towards (or higher relative densities at) the epithelial aspect of the interface can be interpreted as representing the higher intensity of the immune response against the tumor. Conversely, other cells (M2 macrophages and B cells in our study) did not provide the prognostic value.

We found that the gradient density ratio of ICOS cells was associated with the RFS in univariate analyses; ICOS expression can be present in both CD4+ helper T cells and CD8-positive cytotoxic T cells. It is, therefore, difficult to interpret this finding; one could speculate that a higher ICOS IDR (relatively higher intraepithelial ICOS-positive cell density) might reflect an increased proportion of ICOS-positive cytotoxic T cells rather than helper T cells, considering the distribution of CD8 cells in our investigation. Studies subtyping ICOS cell populations, preferably based on multiplex spatial immunoprofiling, are needed to further investigate the role of ICOS expression. Nevertheless, the prognostic significance of ICOS IDR in our study highlights the importance of these cells in the context of BCG immunotherapy outcomes. Also, this may reveal the predictive potential of ICOS IDR in the framework of upcoming ICOS/ICOSL immunotherapy options (REF).

Detection of TLS was not significantly associated with the RFS in our univariate analysis. Exploring the phenomenon and variability of TLS was outside the scope of the present study, as it was treated as a binary feature and therefore lacked prognostic value. Tissue sampling and intratumoral heterogeneity may also affect the detection and quantification of TLS. Nevertheless, in our multivariate analysis, TLS appeared in some weaker models, suggesting its relevance in the broader context of assessing the tumor

microenvironment. Studies investigating the composition of TLS, similar to the study of Dijk et al. [88], might reveal more details on the importance of more specific subsets of TLS.

### Development of a comprehensive model

Multivariate Cox regression analysis revealed that clinical, pathologic, TME, and abstract image features provide independent prognostic information, helping to better stratify patients into risk groups. The best model included already known clinical and pathologic findings, together with our newly developed features – CD11c IDR and MISL prediction. This model helped identify a high-risk group of patients who could benefit from a more aggressive management approach. A bimodal distribution of patient risk suggests that this high-risk group may represent a biologically distinct subset of tumors; however, deeper analysis at the molecular and tumor microenvironment levels is needed to support this finding. Moreover, we identified a patient group (the lowest quartile of the predicted relative risk) with a very good prognosis, which could be further investigated to assess the need for adjuvant therapy in these patients.

To compare our model to the current routine risk assessment strategy, we evaluated the performance of the EORTC risk stratification algorithm [15] in our patient cohort (see Figure 13). In the RFS analysis, the EORTC risk groups did not show a statistically significant difference (multivariate log-rank test,  $p$ -value = 0.1592). This underperformance of the EORTC scheme might be explained by some shift towards more aggressive tumors in our patients, who are eligible for BCG immunotherapy. However, in the OS analysis, risk groups showed significant survival differences in the multivariate log-rank test ( $p$ -value < 0.001) and in pairwise log-rank tests (intermediate vs high risk –  $p$ -value = 0.0402; high vs very high risk –  $p$ -value < 0.001), showing the relevance of this widely accepted risk assessment methodology.

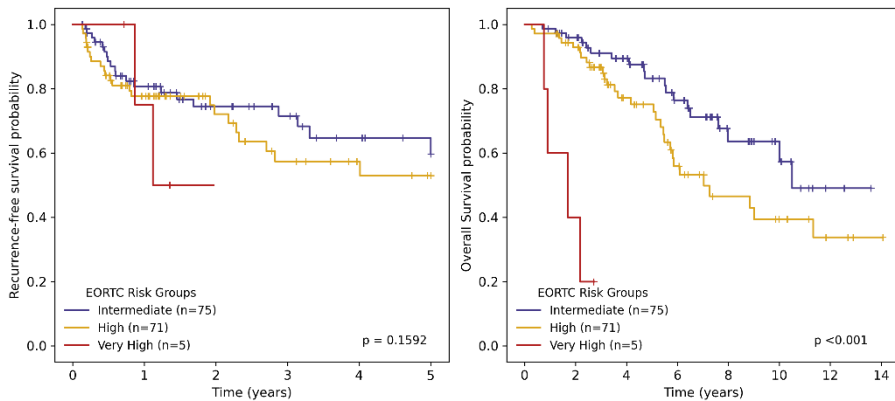


Figure 13. Kaplan-Meier plots for RFS (left) and OS (right) stratified by EORTC risk groups.

The evidence on the use of early radical cystectomy for high-risk non-muscle-invasive bladder cancer that can be performed upfront or in a delayed setting after BCG failure remains controversial [89]. Many recent reports have shown that in patients with BCG-unresponsive HGT1 disease, cancer-specific and overall survival were lower after delayed (>2 years) versus early radical cystectomy [90–92]. However, the retrospective series suggested that in patients with T1G3 tumors, there was only a small difference in recurrence rate between the BCG-treated and the non-BCG-treated group (70 vs. 75%) [93]. Nearly 40% of patients in our study harbored HGT1 disease and, therefore, were less likely to respond to BCG therapy than high-grade Ta tumors. In this context, any improvement in pretreatment prognostic stratification may have significant clinical importance, as it can improve oncological outcomes in a substantial number of patients and/or spare them from unnecessary radical cystectomy.

## LIMITATIONS

The most significant limitation of our study is that our relatively small retrospective cohort consisted of patients from a single center, which may lead to relative homogeneity in patient management, H&E staining quality, and pathologist evaluation, thereby limiting the generalizability of our findings. We look forward to collaborating with other laboratories in the field as we plan to validate these findings in our prospective patient cohort and refine the implementation of MISL to facilitate more explainable predictions.

Additional studies, preferably based on multiplex immunofluorescence and spatial analytics, could enable more precise classification of tumor-infiltrating immune cell populations (T cell subclasses, dendritic cells),

improve definitions of prognostic features, and provide new insights into the biological mechanisms underlying them.

During the study of adapting density gradient indicators, small inconsistencies in our tissue classifier performance led to misclassification of minor epithelial areas in the stroma, necessitating manual data curation. It was performed in a standardized manner; nevertheless, this made the DIA not entirely automated. Another issue in urothelial tumors is cytological similarity of malignant and nonmalignant urothelium, which is why the interface zone in our study includes any urothelium. Therefore, all tissue has been classified as 'stroma' or 'epithelium'. However, to minimize the impact of data derived from normal mucosa, we have selected tissue samples for the study that have a predominance of tumor epithelium over normal urothelium.

## CONCLUSIONS

Abstract image features extracted from H&E images using a deep neural network provide a strong backbone for further prognostic work in NMIPUC, and MISL is a promising approach for patient outcome prediction, providing additional information beyond the well-known clinical and pathologic variables. The use of full survival data proved to be a more effective approach for predicting patient outcomes. However, the rapid pace of progress in the application of AI technology in digital pathology leaves room for further improvement in the methodology of this task.

We demonstrated that density gradient indicators tailored to the histological architecture of NMIPUC can be used to assess immune cell distribution at the epithelial-stromal interface and serve as a strong prognostic factor. Interestingly, this tumor type required a much more delicate approach, with a very narrow interface zone, compared to other tumor types reported previously.

Our further investigation of the immune tumor microenvironment reveals the independent and superior prognostic value of proinflammatory CD11c macrophages with a positive gradient (relatively dense infiltration) toward the epithelial aspect of the tumor. We have demonstrated the high predictive value of immune cell distribution at the tumor epithelium-stroma interface, as assessed by the IDR indicator, which outweighs the simple measurement of cell densities within the tumor. The strongest predictive value was found in the cell populations directly involved in anti-tumor immunity (CD11c and CD8), highlighting the IDR as a measure of immune system mobilization. These findings further support the value of spatial analytics in the tumor

microenvironment while also providing useful means for NMIPUC patient risk stratification along with conventional clinicopathologic features.

## FURTHER RESEARCH DIRECTIONS

Our work yielded strong results in both DL-based outcome prediction and TME-derived feature assessments. A further validation study is currently underway, involving the collection of a new, independent patient cohort. Additionally, we are exploring new methods for feature extraction from tumor images, employing pathology-specific visual transformer-based foundation models, more effective fine-tuning methods, advanced aggregation techniques, and federated learning. Preliminary results are showing promise to improve DL-based prognostic capabilities from H&E images.

TME assessment with our current methodology requires IHC for immune cell identification, thus limiting its application in everyday practice. To address this limitation, we plan to apply the methodology of assessing spatial distribution to the immune cells identified in readily available H&E images at the epithelial-stromal interface of NMIPUC.

Another research possibility could be a deeper multimodal analysis incorporating genomic and transcriptomic data, and analyzing more immune cell populations could yield even stronger prognostic models, as well as giving deeper insights into NMIPUC tumor biology and its clinical relevance.

Of course, the ultimate goal is real-time clinical decision support; however, before achieving this, multicenter validation, improvements in automation, reductions in manual intervention, and enhanced scalability across pathology labs are required. This would lead to improved clinical decision-making and patient management.

## RECOMMENDATIONS

1. Spatial immune cell distribution (CD8-positive and CD11c-positive) in NMIPUC tissue, assessed by digital image analysis, can be used to predict NMIPUC patient outcomes after intravesical BCG therapy.
2. DL-based relapse prediction can be used to improve NMIPUC patient stratification and clinical decision-making; however, adaptation of newer, more robust methods of feature extraction and aggregation would likely increase the predictive ability of this modality.

## REFERENCES

1. Bray, F.; Laversanne, M.; Sung, H.; Ferlay, J.; Siegel, R.L.; Soerjomataram, I.; Jemal, A. Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* **2024**, *74*, 229–263, doi:10.3322/caac.21834.
2. Vincerževskienė, I.; Stankevič, A.; Zinkevičiūtė, E.; Steponavičienė, L.; Obžigailov, J.; Kalvaitis, R. Vėžys Lietuvoje 2017 metais 2023.
3. Larsson, S.C.; Chen, J.; Ruan, X.; Li, X.; Yuan, S. Genome-Wide Association Study and Mendelian Randomization Analyses Reveal Insights into Bladder Cancer Etiology. *JNCI Cancer Spectr* **2025**, *9*, pkaf014, doi:10.1093/jncics/pkaf014.
4. Letašiová, S.; Medved'ová, A.; Šovčíková, A.; Dušinská, M.; Volkovová, K.; Mosoiu, C.; Bartonová, A. Bladder Cancer, a Review of the Environmental Risk Factors. *Environmental Health* **2012**, *11*, S11, doi:10.1186/1476-069X-11-S1-S11.
5. Zaghoul, M.S. Bladder Cancer and Schistosomiasis. *Journal of the Egyptian National Cancer Institute* **2012**, *24*, 151–159, doi:10.1016/j.jnci.2012.08.002.
6. Board, W.C. of T.E. *Urinary and Male Genital Tumours*; ISBN 978-92-832-4512-4.
7. Schultz, L. The Papilla as a Biomarker in the Molecular Era of Bladder Oncology. *Surgical and Experimental Pathology* **2018**, *1*, 10, doi:10.1186/s42047-018-0023-6.
8. Guo, C.C.; Czerniak, B. Bladder Cancer in the Genomic Era. *Arch Pathol Lab Med* **2019**, *143*, 695–704, doi:10.5858/arpa.2018-0329-RA.
9. Neuzillet, Y.; Paoletti, X.; Ouerhani, S.; Mongiat-Artus, P.; Soliman, H.; The, H. de; Sibony, M.; Denoux, Y.; Molinie, V.; Herault, A.; et al. A Meta-Analysis of the Relationship between FGFR3 and TP53 Mutations in Bladder Cancer. *PLOS ONE* **2012**, *7*, e48993, doi:10.1371/journal.pone.0048993.
10. Weinstein, J.N.; Akbani, R.; Broom, B.M.; Wang, W.; Verhaak, R.G.W.; McConkey, D.; Lerner, S.; Morgan, M.; Creighton, C.J.; Smith, C.; et al. Comprehensive Molecular Characterization of Urothelial Bladder Carcinoma. *Nature* **2014**, *507*, 315–322, doi:10.1038/nature12965.
11. Kamat, A.M.; Hahn, N.M.; Efstathiou, J.A.; Lerner, S.P.; Malmström, P.-U.; Choi, W.; Guo, C.C.; Lotan, Y.; Kassouf, W. Bladder Cancer. *The Lancet* **2016**, *388*, 2796–2810, doi:10.1016/S0140-6736(16)30512-8.

12. Babjuk, M.; Burger, M.; Čapoun, O.; Cohen, D.; Compérat, E.; Escrig, J.L.D.; Gontero, P.; Liedberg, F.; Masson-Lecomte, A.; Mostafid, H.; et al. European Association of Urology Guidelines on Non-Muscle-Invasive Bladder Cancer (Ta, T1, and Carcinoma in Situ). *European Urology* **2021**, doi:10.1016/j.eururo.2021.08.010.
13. Sylvester, R.J.; Brausi, M.A.; Kirkels, W.J.; Hoeltl, W.; Calais Da Silva, F.; Powell, P.H.; Prescott, S.; Kirkali, Z.; van de Beek, C.; Gorlia, T.; et al. Long-Term Efficacy Results of EORTC Genito-Urinary Group Randomized Phase 3 Study 30911 Comparing Intravesical Instillations of Epirubicin, Bacillus Calmette-Guérin, and Bacillus Calmette-Guérin plus Isoniazid in Patients with Intermediate- and High-Risk Stage Ta T1 Urothelial Carcinoma of the Bladder. *European Urology* **2010**, *57*, 766–773, doi:10.1016/j.eururo.2009.12.024.
14. Koch, G.E.; Smelser, W.W.; Chang, S.S. Side Effects of Intravesical BCG and Chemotherapy for Bladder Cancer: What They Are and How to Manage Them. *Urology* **2021**, *149*, 11–20, doi:10.1016/j.urology.2020.10.039.
15. Sylvester, R.J.; van der Meijden, A.P.M.; Oosterlinck, W.; Witjes, J.A.; Bouffieux, C.; Denis, L.; Newling, D.W.W.; Kurth, K. Predicting Recurrence and Progression in Individual Patients with Stage Ta T1 Bladder Cancer Using EORTC Risk Tables: A Combined Analysis of 2596 Patients from Seven EORTC Trials. *European Urology* **2006**, *49*, 466–477, doi:10.1016/j.eururo.2005.12.031.
16. Fernandez-Gomez, J.; Madero, R.; Solsona, E.; Unda, M.; Martinez-Piñeiro, L.; Gonzalez, M.; Portillo, J.; Ojea, A.; Pertusa, C.; Rodriguez-Molina, J.; et al. Predicting Nonmuscle Invasive Bladder Cancer Recurrence and Progression in Patients Treated With Bacillus Calmette-Guerin: The CUETO Scoring Model. *The Journal of Urology* **2009**, *182*, 2195–2203, doi:10.1016/j.juro.2009.07.016.
17. Chang, S.S.; Boorjian, S.A.; Chou, R.; Clark, P.E.; Daneshmand, S.; Konety, B.R.; Pruthi, R.; Quale, D.Z.; Ritch, C.R.; Seigne, J.D.; et al. Diagnosis and Treatment of Non-Muscle Invasive Bladder Cancer: AUA/SUO Guideline. *The Journal of Urology* **2016**, doi:10.1016/j.juro.2016.06.049.
18. Ślusarczyk, A.; Garbas, K.; Pustuła, P.; Zapła, Ł.; Radziszewski, P. Assessing the Predictive Accuracy of EORTC, CUETO and EAU Risk Stratification Models for High-Grade Recurrence and Progression after Bacillus Calmette–Guérin Therapy in Non-Muscle-Invasive Bladder Cancer. *Cancers* **2024**, *16*, 1684, doi:10.3390/cancers16091684.

19. Soukup, V.; Čapoun, O.; Cohen, D.; Hernández, V.; Babjuk, M.; Burger, M.; Compérat, E.; Gontero, P.; Lam, T.; MacLennan, S.; et al. Prognostic Performance and Reproducibility of the 1973 and 2004/2016 World Health Organization Grading Classification Systems in Non-Muscle-Invasive Bladder Cancer: A European Association of Urology Non-Muscle Invasive Bladder Cancer Guidelines Panel Systematic Review. *European Urology* **2017**, *72*, 801–813, doi:10.1016/j.eururo.2017.04.015.
20. Shalata, A.T.; Shehata, M.; Van Bogaert, E.; Ali, K.M.; Alksas, A.; Mahmoud, A.; El-Gendy, E.M.; Mohamed, M.A.; Giridharan, G.A.; Contractor, S.; et al. Predicting Recurrence of Non-Muscle-Invasive Bladder Cancer: Current Techniques and Future Trends. *Cancers* **2022**, *14*, 5019, doi:10.3390/cancers14205019.
21. Kamat, A.M.; Roger Li; Li, R.; O'Donnell, M.A.; Choyke, P.L.; Black, P.C.; Rouprêt, M.; Catto, J.W.F.; Comperat, E.; Ingersoll, M.A.; et al. Predicting Response to Intravesical Bacillus Calmette-Guérin Immunotherapy: Are We There Yet? A Systematic Review. *European Urology* **2018**, *73*, 738–748, doi:10.1016/j.eururo.2017.10.003.
22. Jansen, I.; Lucas, M.; Bosschieter, J.; de Boer, O.J.; Meijer, S.L.; van Leeuwen, T.G.; Marquering, H.A.; Nieuwenhuijzen, J.A.; de Bruin, D.M.; Savci-Heijink, C.D. Automated Detection and Grading of Non-Muscle-Invasive Urothelial Cell Carcinoma of the Bladder. *The American Journal of Pathology* **2020**, *190*, 1483–1490, doi:10.1016/j.ajpath.2020.03.013.
23. Wetteland, R.; Kvikstad, V.; Eftestol, T.; Tossebro, E.; Lillesand, M.; Janssen, E.A.M.; Engan, K. Automatic Diagnostic Tool for Predicting Cancer Grade in Bladder Cancer Patients Using Deep Learning. *IEEE Access* **2021**, *9*, 115813–115825, doi:10.1109/ACCESS.2021.3104724.
24. Zhang, Z.; Chen, P.; McGough, M.; Xing, F.; Wang, C.; Bui, M.; Xie, Y.; Sapkota, M.; Cui, L.; Dhillon, J.; et al. Pathologist-Level Interpretable Whole-Slide Cancer Diagnosis with Deep Learning. *Nat Mach Intell* **2019**, *1*, 236–245, doi:10.1038/s42256-019-0052-1.
25. Spyridonos, P.; Cavouras, D.; Ravazoula, P.; Nikiforidis, G. Neural Network-Based Segmentation and Classification System for Automated Grading of Histologic Sections of Bladder Carcinoma. *Anal Quant Cytol Histol* **2002**, *24*, 317–324.
26. Spyridonos, P.R., D. Cavouras, K. Berberidis, G. Nikiforidis, P. Computer-Based Grading of Haematoxylin-Eosin Stained Tissue Sections of Urinary Bladder Carcinomas. *Medical Informatics and the Internet in Medicine* **2001**, *26*, 179–190, doi:10.1080/14639230110065757.

27. Choi, H.; Jarkrans, T.; Bengtsson, E.; Vasko, J.; Wester, K.; Malmström, P.-U.; Busch, C. Image Analysis Based Grading of Bladder Carcinoma. Comparison of Object, Texture and Graph Based Methods and Their Reproducibility. *Anal Cell Pathol* **1997**, *15*, 1–18, doi:10.1155/1997/147187.
28. Shalata, A.T.; Alksas, A.; Shehata, M.; Khater, S.; Ezzat, O.; Ali, K.M.; Gondim, D.; Mahmoud, A.; El-Gendy, E.M.; Mohamed, M.A.; et al. Precise Grading of Non-Muscle Invasive Bladder Cancer with Multi-Scale Pyramidal CNN. *Sci Rep* **2024**, *14*, 25131, doi:10.1038/s41598-024-77101-6.
29. Lucas, M.; Jansen, I.; van Leeuwen, T.G.; Oddens, J.R.; de Bruin, D.M.; Marquering, H.A. Deep Learning–Based Recurrence Prediction in Patients with Non–Muscle-Invasive Bladder Cancer. *European Urology Focus* **2022**, *8*, 165–172, doi:10.1016/j.euf.2020.12.008.
30. Chen, S.; Campanella, G.; Elmas, A.; Stock, A.; Zeng, J.; Polydorides, A.D.; Schoenfeld, A.J.; Huang, K.; Houldsworth, J.; Vanderbilt, C.; et al. Benchmarking Embedding Aggregation Methods in Computational Pathology: A Clinical Data Perspective. In Proceedings of the Proceedings of the MICCAI Workshop on Computational Pathology; PMLR, November 17 2024; pp. 38–50.
31. Cheng, L.; Neumann, R.M.; Nehra, A.; Spotts, B.E.; Weaver, A.L.; Bostwick, D.G. Cancer Heterogeneity and Its Biologic Implications in the Grading of Urothelial Carcinoma. *Cancer* **2000**, *88*, 1663–1670, doi:10.1002/(SICI)1097-0142(20000401)88:7<1663::AID-CNCR21>3.0.CO;2-8.
32. Gofrit, O.N.; Pizov, G.; Shapiro, A.; Duvdevani, M.; Yutkin, V.; Landau, E.H.; Zorn, K.C.; Hidas, G.; Pode, D. Mixed High and Low Grade Bladder Tumors—Are They Clinically High or Low Grade? *Journal of Urology* **2014**, *191*, 1693–1696, doi:10.1016/j.juro.2013.11.056.
33. Reis, L.O.; Taheri, D.; Chaux, A.; Guner, G.; Mendoza Rodriguez, M.A.; Bivalacqua, T.J.; Schoenberg, M.P.; Epstein, J.I.; Netto, G.J. Significance of a Minor High-Grade Component in a Low-Grade Noninvasive Papillary Urothelial Carcinoma of Bladder. *Human Pathology* **2016**, *47*, 20–25, doi:10.1016/j.humpath.2015.09.007.
34. Schubert, T.; Danzig, M.R.; Kotamarti, S.; Ghandour, R.A.; Lascano, D.; Dubow, B.P.; Decastro, G.J.; Benson, M.C.; McKiernan, J.M. Mixed Low- and High-Grade Non-Muscle-Invasive Bladder Cancer: A Histological Subtype with Favorable Outcome. *World J Urol* **2015**, *33*, 847–852, doi:10.1007/s00345-014-1383-5.

35. Ilse, M.; Tomczak, J.; Welling, M. Attention-Based Deep Multiple Instance Learning. In Proceedings of the Proceedings of the 35th International Conference on Machine Learning; PMLR, July 3 2018; pp. 2127–2136.
36. Yao, J.; Zhu, X.; Jonnagaddala, J.; Hawkins, N.; Huang, J. Whole Slide Images Based Cancer Survival Prediction Using Attention Guided Deep Multiple Instance Learning Networks. *Medical Image Analysis* **2020**, *65*, 101789, doi:10.1016/j.media.2020.101789.
37. Arneth, B. Tumor Microenvironment. *Medicina* **2020**, *56*, 15, doi:10.3390/medicina56010015.
38. Monkman, J.; Moradi, A.; Yunis, J.; Ivison, G.; Mayer, A.; Ladwa, R.; O’Byrne, K.; Kulasinghe, A. Spatial Insights into Immunotherapy Response in Non-Small Cell Lung Cancer (NSCLC) by Multiplexed Tissue Imaging. *J Transl Med* **2024**, *22*, 239, doi:10.1186/s12967-024-05035-8.
39. Hendry, S.; Salgado, R.; Gevaert, T.; Russell, P.A.; John, T.; Thapa, B.; Christie, M.; van de Vijver, K.; Estrada, M.V.; Gonzalez-Ericsson, P.I.; et al. Assessing Tumor Infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method from the International Immuno-Oncology Biomarkers Working Group. *Adv Anat Pathol* **2017**, *24*, 235–251, doi:10.1097/PAP.000000000000162.
40. Hendry, S.; Salgado, R.; Gevaert, T.; Russell, P.A.; John, T.; Thapa, B.; Christie, M.; van de Vijver, K.; Estrada, M.V.; Gonzalez-Ericsson, P.I.; et al. Assessing Tumor Infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method from the International Immuno-Oncology Biomarkers Working Group. *Adv Anat Pathol* **2017**, *24*, 311–335, doi:10.1097/PAP.000000000000161.
41. Pichler, R.; Gruenbacher, G.; Culig, Z.; Brunner, A.; Fuchs, D.; Fritz, J.; Gander, H.; Rahm, A.; Thurnher, M. Intratumoral Th2 Predisposition Combines with an Increased Th1 Functional Phenotype in Clinical Response to Intravesical BCG in Bladder Cancer. *Cancer Immunology, Immunotherapy* **2017**, *66*, 427–440, doi:10.1007/s00262-016-1945-z.
42. VILLOLDO, G.M.; POMBO, M.T.; ARIS, M.; CHEMI, J.; MANDÓ, P.; NAGARAJU, S.; CAMEAN, J.; BURIONI, A.; EGEE, D.; AMAT, M.; et al. A Th2-Score in the Tumor Microenvironment as a Predictive Biomarker of Response to Bacillus Calmette Guérin in Patients with Non-Muscle Invasive Bladder Carcinoma: A Retrospective Study. *Oncol Res* **31**, 207–220, doi:10.32604/or.2023.028163.
43. Miyake, M.; Tatsumi, Y.; Gotoh, D.; Ohnishi, S.; Owari, T.; Iida, K.; Ohnishi, K.; Hori, S.; Morizawa, Y.; Itami, Y.; et al. Regulatory T Cells and Tumor-Associated Macrophages in the Tumor Microenvironment in Non-

Muscle Invasive Bladder Cancer Treated with Intravesical Bacille Calmette-Guérin: A Long-Term Follow-Up Study of a Japanese Cohort. *IJMS* **2017**, *18*, 2186, doi:10.3390/ijms18102186.

44. Masson-Lecomte, A.; Maillé, P.; Pineda, S.; Soyeux, P.; Sagraera, A.; Rava, M.; Lopez de Maturana, E.; Márquez, M.; Tardón, A.; Carrato, A.; et al. CD8+ Cytotoxic Immune Infiltrate in Non-Muscle Invasive Bladder Cancer: A Standardized Methodology to Study Association with Clinico-Pathological Features and Prognosis. *Bladder Cancer* **2019**, *5*, 159–169, doi:10.3233/BLC-180206.

45. Kates, M.; Matoso, A.; Choi, W.; Baras, A.S.; Daniels, M.J.; Lombardo, K.; Brant, A.; Mikkilineni, N.; McConkey, D.J.; Kamat, A.M.; et al. Adaptive Immune Resistance to Intravesical BCG in Non-Muscle Invasive Bladder Cancer: Implications for Prospective BCG-Unresponsive Trials. *Clin Cancer Res* **2020**, *26*, 882–891, doi:10.1158/1078-0432.CCR-19-1920.

46. Lillesand, M.; Kvikstad, V.; Mangrud, O.M.; Gudlaugsson, E.; van Diermen-Hidle, B.; Skaland, I.; Baak, J.P.A.; Janssen, E.A.M. Mitotic Activity Index and CD25+ Lymphocytes Predict Risk of Stage Progression in Non-Muscle Invasive Bladder Cancer. *PLoS One* **2020**, *15*, e0233676, doi:10.1371/journal.pone.0233676.

47. Pichler, R.; Fritz, J.; Zavadil, C.; Claudia Zavadil; Schäfer, G.; Culig, Z.; Brunner, A. Tumor-Infiltrating Immune Cell Subpopulations Influence the Oncologic Outcome after Intravesical Bacillus Calmette-Guérin Therapy in Bladder Cancer. *Oncotarget* **2016**, *7*, 39916–39930, doi:10.18632/oncotarget.9537.

48. Jinushi, M.; Komohara, Y. Tumor-Associated Macrophages as an Emerging Target against Tumors: Creating a New Path from Bench to Bedside. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **2015**, *1855*, 123–130, doi:10.1016/j.bbcan.2015.01.002.

49. Hourani, T.; Holden, J.A.; Li, W.; Lenzo, J.C.; Hadjigol, S.; O'Brien-Simpson, N.M. Tumor Associated Macrophages: Origin, Recruitment, Phenotypic Diversity, and Targeting. *Front. Oncol.* **2021**, *11*, doi:10.3389/fonc.2021.788365.

50. Bai, B.; Xie, S.; Wang, Y.; Wu, F.; Chen, Y.; Bian, J.; Gao, X. Development of Anti-Cancer Drugs for Tumor-Associated Macrophages: A Comprehensive Review and Mechanistic Insights. *Front. Mol. Biosci.* **2024**, *11*, doi:10.3389/fmolb.2024.1463061.

51. Hanada, T.; Nakagawa, M.; Emoto, A.; Nomura, T.; Nasu, N.; Nomura, Y. Prognostic Value of Tumor-Associated Macrophage Count in

Human Bladder Cancer. *International Journal of Urology* **2000**, *7*, 263–269, doi:10.1046/j.1442-2042.2000.00190.x.

52. Krpina, K.; Babarović, E.; Španjol, J.; Đorđević, G.; Maurer, T.; Jonjić, N. Correlation of Tumor-Associated Macrophages and NK Cells with Bladder Cancer Size and T Stage in Patients with Solitary Low-Grade Urothelial Carcinoma. *Wien Klin Wochenschr* **2016**, *128*, 248–252, doi:10.1007/s00508-015-0907-3.

53. Ajili, F.; Kourda, N.; Darouiche, A.; Chebil, M.; Boubaker, S. Prognostic Value of Tumor-Associated Macrophages Count in Human Non-Muscle-Invasive Bladder Cancer Treated by BCG Immunotherapy. *Ultrastructural Pathology* **2013**, *37*, 56–61, doi:10.3109/01913123.2012.728688.

54. Boström, M.M.; Irjala, H.; Mirtti, T.; Taimen, P.; Kauko, T.; Ålgars, A.; Jalkanen, S.; Boström, P.J. Tumor-Associated Macrophages Provide Significant Prognostic Information in Urothelial Bladder Cancer. *PLoS ONE* **2015**, *10*, e0133552, doi:10.1371/journal.pone.0133552.

55. Ayari, C.; LaRue, H.; Hovington, H.; Decobert, M.; Harel, F.; Bergeron, A.; Têtu, B.; Lacombe, L.; Fradet, Y. Bladder Tumor Infiltrating Mature Dendritic Cells and Macrophages as Predictors of Response to Bacillus Calmette-Guérin Immunotherapy. *European Urology* **2009**, *55*, 1386–1396, doi:10.1016/j.eururo.2009.01.040.

56. Suriano, F.; Santini, D.; Perrone, G.; Amato, M.; Vincenzi, B.; Tonini, G.; Muda, A.O.; Boggia, S.; Buscarini, M.; Pantano, F. Tumor Associated Macrophages Polarization Dictates the Efficacy of BCG Instillation in Non-Muscle Invasive Urothelial Bladder Cancer. *J Exp Clin Cancer Res* **2013**, *32*, 87, doi:10.1186/1756-9966-32-87.

57. Lima, L.; Oliveira, D.; Tavares, A.; Amaro, T.; Cruz, R.; Oliveira, M.J.; Ferreira, J.A.; Santos, L. The Predominance of M2-Polarized Macrophages in the Stroma of Low-Hypoxic Bladder Tumors Is Associated with BCG Immunotherapy Failure. *Urol Oncol* **2014**, *32*, 449–457, doi:10.1016/j.urolonc.2013.10.012.

58. Steele, K.E.; Tan, T.H.; Korn, R.; Dacosta, K.; Brown, C.; Kuziora, M.; Zimmermann, J.; Laffin, B.; Widmaier, M.; Rognoni, L.; et al. Measuring Multiple Parameters of CD8+ Tumor-Infiltrating Lymphocytes in Human Cancers by Image Analysis. *J Immunother Cancer* **2018**, *6*, 20, doi:10.1186/s40425-018-0326-x.

59. Lim, Y.; Choi, S.; Oh, H.J.; Kim, C.; Song, S.; Kim, S.; Song, H.; Park, S.; Kim, J.-W.; Kim, J.W.; et al. Artificial Intelligence-Powered Spatial Analysis of Tumor-Infiltrating Lymphocytes for Prediction of Prognosis in

Resected Colon Cancer. *npj Precis. Onc.* **2023**, 7, 124, doi:10.1038/s41698-023-00470-0.

60. Page, D.B.; Broeckx, G.; Jahangir, C.A.; Verbandt, S.; Gupta, R.R.; Thagaard, J.; Khuroy, R.; Kos, Z.; Abduljabbar, K.; Acosta Haab, G.; et al. Spatial Analyses of Immune Cell Infiltration in Cancer: Current Methods and Future Directions: A Report of the International Immuno-Oncology Biomarker Working Group on Breast Cancer. *The Journal of Pathology* **2023**, 260, 514–532, doi:10.1002/path.6165.

61. Chen, D.S.; Mellman, I. Elements of Cancer Immunity and the Cancer–Immune Set Point. *Nature* **2017**, 541, 321–330, doi:10.1038/nature21349.

62. Herbst, R.S.; Soria, J.-C.; Kowanetz, M.; Fine, G.D.; Hamid, O.; Gordon, M.S.; Sosman, J.A.; McDermott, D.F.; Powderly, J.D.; Gettinger, S.N.; et al. Predictive Correlates of Response to the Anti-PD-L1 Antibody MPDL3280A in Cancer Patients. *Nature* **2014**, 515, 563–567, doi:10.1038/nature14011.

63. Galon, J.; Pagès, F.; Marincola, F.M.; Angell, H.K.; Angell, H.K.; Thurin, M.; Lugli, A.; Zlobec, I.; Berger, A.; Bifulco, C.; et al. Cancer Classification Using the Immunoscore: A Worldwide Task Force. *Journal of Translational Medicine* **2012**, 10, 205–205, doi:10.1186/1479-5876-10-205.

64. Galon, J.; Costes, A.; Sanchez-Cabo, F.; Kirilovsky, A.; Mlecnik, B.; Lagorce-Pagès, C.; Tosolini, M.; Camus, M.; Berger, A.; Wind, P.; et al. Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome. *Science* **2006**, 313, 1960–1964, doi:10.1126/science.1129139.

65. Pagès, F.; Mlecnik, B.; Marliot, F.; Bindea, G.; Ou, F.S.; Ou, F.-S.; Bifulco, C.; Bifulco, C.; Lugli, A.; Zlobec, I.; et al. International Validation of the Consensus Immunoscore for the Classification of Colon Cancer: A Prognostic and Accuracy Study. *The Lancet* **2018**, 391, 2128–2139, doi:10.1016/s0140-6736(18)30789-x.

66. Angell, H.K.; Bruni, D.; Barrett, J.C.; Herbst, R.; Galon, J. The Immunoscore: Colon Cancer and Beyond. *Clinical Cancer Research* **2020**, 26, 332–339, doi:10.1158/1078-0432.CCR-18-1851.

67. Rasmusson, A.; Zilenaite, D.; Nestarenkaite, A.; Augulis, R.; Laurinaviciene, A.; Ostapenko, V.; Poskus, T.; Laurinavicius, A. Immunogradient Indicators for Antitumor Response Assessment by Automated Tumor-Stroma Interface Zone Detection. *The American Journal of Pathology* **2020**, 190, 1309–1322, doi:10.1016/j.ajpath.2020.01.018.

68. Zilenaite, D.; Rasmusson, A.; Augulis, R.; Besusparis, J.; Laurinaviciene, A.; Plancoulaine, B.; Ostapenko, V.; Laurinavicius, A. Independent Prognostic Value of Intratumoral Heterogeneity and Immune Response Features by Automated Digital Immunohistochemistry Analysis in Early Hormone Receptor-Positive Breast Carcinoma. *Frontiers in Oncology* **2020**, *10*, 950, doi:10.3389/fonc.2020.00950.
69. Radziuviene, G.; Rasmusson, A.; Augulis, R.; Grineviciute, R.B.; Zilenaite, D.; Laurinaviciene, A.; Ostapenko, V.; Laurinavicius, A. Intratumoral Heterogeneity and Immune Response Indicators to Predict Overall Survival in a Retrospective Study of HER2-Borderline (IHC 2+) Breast Cancer Patients. *Frontiers in Oncology* **2021**, *11*, doi:10.3389/fonc.2021.774088.
70. Nestarenkaite, A.; Fadhil, W.; Rasmusson, A.; Susanti, S.; Hadjimichael, E.; Laurinaviciene, A.; Ilyas, M.; Laurinavicius, A. Immuno-Interface Score to Predict Outcome in Colorectal Cancer Independent of Microsatellite Instability Status. *Cancers* **2020**, *12*, 2902, doi:10.3390/cancers12102902.
71. Stulpinas, R.; Zilenaite-Petrulaitiene, D.; Rasmusson, A.; Gulla, A.; Grigonyte, A.; Strupas, K.; Laurinavicius, A. Prognostic Value of CD8+ Lymphocytes in Hepatocellular Carcinoma and Perineoplastic Parenchyma Assessed by Interface Density Profiles in Liver Resection Samples. *Cancers* **2023**, *15*, 366, doi:10.3390/cancers15020366.
72. Rejeb, S.B.; Elfekih, S.; Kouki, N.; Boulma, R.; Khouni, H. Immunochemistry-Based Quantification of Tumor-Infiltrating Lymphocytes and Immunoscore as Prognostic Biomarkers in Bladder Cancer. *Journal of the Egyptian National Cancer Institute* **2024**, *36*, 9, doi:10.1186/s43046-024-00212-8.
73. Ariafar, A.; Sanati, A.; Ahmadvand, S.; Shekarkhar, G.; Safaei, A.; Shayan, Z.; Faghieh, Z. Prognostic Significance of Immunoscore Related Markers in Bladder Cancer. *BMC Urology* **2022**, *22*, 133, doi:10.1186/s12894-022-01085-6.
74. Bieri, U.; Buser, L.; Wettstein, M.S.; Eberli, D.; Saba, K.; Moch, H.; Hermanns, T.; Poyet, C. Modified Immunoscore Improves Prediction of Survival Outcomes in Patients Undergoing Radical Cystectomy for Bladder Cancer—A Retrospective Digital Pathology Study. *Diagnostics* **2022**, *12*, 1360, doi:10.3390/diagnostics12061360.
75. Bieri, U.; Enderlin, D.; Buser, L.; Wettstein, M.S.; Eberli, D.; Moch, H.; Hermanns, T.; Poyet, C. Modified Immunoscore Improves the Prediction

of Progression-Free Survival in Patients with Non-Muscle-Invasive Bladder Cancer: A Digital Pathology Study. *Frontiers in Oncology* **2022**, *12*.

76. Council for International Organizations of Medical Sciences (CIOMS) *International Ethical Guidelines for Health-Related Research Involving Humans*; Council for International Organizations of Medical Sciences (CIOMS), 2016;

77. Drachneris, J.; Rasmusson, A.; Morkunas, M.; Fabijonavicius, M.; Cekauskas, A.; Jankevicius, F.; Laurinavicius, A. CD8+ Cell Density Gradient across the Tumor Epithelium–Stromal Interface of Non-Muscle Invasive Papillary Urothelial Carcinoma Predicts Recurrence-Free Survival after BCG Immunotherapy. *Cancers* **2023**, *15*, 1205, doi:10.3390/cancers15041205.

78. Drachneris, J.; Morkunas, M.; Fabijonavicius, M.; Cekauskas, A.; Jankevicius, F.; Laurinavicius, A. Prediction of Non-Muscle Invasive Papillary Urothelial Carcinoma Relapse from Hematoxylin–Eosin Images Using Deep Multiple Instance Learning in Patients Treated with Bacille Calmette–Guérin Immunotherapy. *Biomedicines* **2024**, *12*, 360, doi:10.3390/biomedicines12020360.

79. Julius Drachneris; M. Morkunas; M. Fabijonavicius; Albertas Čekauskas; F. Jankevicius; A. Laurinavicius Spatial Distribution of Macrophage and Lymphocyte Subtypes within Tumor Microenvironment to Predict Recurrence of Non-Muscle-Invasive Papillary Urothelial Carcinoma after BCG Immunotherapy. *International Journal of Molecular Sciences* **2024**, doi:10.3390/ijms25094776.

80. Rawat, R.R.; Ortega, I.; Roy, P.; Sha, F.; Shibata, D.; Ruderman, D.; Agus, D.B. Deep Learned Tissue “Fingerprints” Classify Breast Cancers by ER/PR/Her2 Status from H&E Images. *Sci Rep* **2020**, *10*, 7275, doi:10.1038/s41598-020-64156-4.

81. Ferro, M.; Vartolomei, M.D.; Cantiello, F.; Lucarelli, G.; Di Stasi, S.M.; Hurle, R.; Guazzoni, G.; Busetto, G.M.; De Berardinis, E.; Damiano, R.; et al. High-Grade T1 on Re-Transurethral Resection after Initial High-Grade T1 Confers Worse Oncological Outcomes: Results of a Multi-Institutional Study. *Urologia Internationalis* **2018**, *101*, 7–15, doi:10.1159/000490765.

82. Tae, B.S.; Chang Wook Jeong; Jeong, C.W.; Cheol Kwak; Kwak, C.; Kim, H.H.; Kyung Chul Moon; Moon, K.C.; Ku, J.H. Pathology in Repeated Transurethral Resection of a Bladder Tumor as a Risk Factor for Prognosis of High-Risk Non-Muscle-Invasive Bladder Cancer. *PLOS ONE* **2017**, *12*, doi:10.1371/journal.pone.0189354.

83. Han, K.S.; Joung, J.Y.; Cho, K.S.; Seo, H.K.; Chung, J.; Park, W.S.; Lee, K.H. Results of Repeated Transurethral Resection for a Second Opinion in Patients Referred for Nonmuscle Invasive Bladder Cancer: The Referral Cancer Center Experience and Review of the Literature. *J Endourol* **2008**, *22*, 2699–2704, doi:10.1089/end.2008.0281.
84. Chen, R.J.; Ding, T.; Lu, M.Y.; Williamson, D.F.K.; Jaume, G.; Song, A.H.; Chen, B.; Zhang, A.; Shao, D.; Shaban, M.; et al. Towards a General-Purpose Foundation Model for Computational Pathology. *Nat Med* **2024**, *30*, 850–862, doi:10.1038/s41591-024-02857-3.
85. Vorontsov, E.; Bozkurt, A.; Casson, A.; Shaikovski, G.; Zelechowski, M.; Severson, K.; Zimmermann, E.; Hall, J.; Tenenholtz, N.; Fusi, N.; et al. A Foundation Model for Clinical-Grade Computational Pathology and Rare Cancers Detection. *Nat Med* **2024**, *30*, 2924–2935, doi:10.1038/s41591-024-03141-0.
86. Ma, J.; Guo, Z.; Zhou, F.; Wang, Y.; Xu, Y.; Li, J.; Yan, F.; Cai, Y.; Zhu, Z.; Jin, C.; et al. A Generalizable Pathology Foundation Model Using a Unified Knowledge Distillation Pretraining Framework. *Nat. Biomed. Eng* **2025**, 1–20, doi:10.1038/s41551-025-01488-4.
87. Chen, S.; Campanella, G.; Elmas, A.; Stock, A.; Zeng, J.; Polydorides, A.D.; Schoenfeld, A.J.; Huang, K.; Houldsworth, J.; Vanderbilt, C.; et al. Benchmarking Embedding Aggregation Methods in Computational Pathology: A Clinical Data Perspective. In Proceedings of the Proceedings of the MICCAI Workshop on Computational Pathology; PMLR, November 17 2024; pp. 38–50.
88. Dijk, N. van; Gil-Jimenez, A.; Silina, K.; Montfoort, M.L. van; Einerhand, S.; Jonkman, L.; Voskuilen, C.S.; Peters, D.; Sanders, J.; Lubeck, Y.; et al. The Tumor Immune Landscape and Architecture of Tertiary Lymphoid Structures in Urothelial Cancer. *Frontiers in Immunology* **2021**, *12*.
89. Diamant, E.; Roumiguié, M.; Ingels, A.; Parra, J.; Vordos, D.; Bajeot, A.-S.; Chartier-Kastler, E.; Soulié, M.; de la Taille, A.; Rouprêt, M.; et al. Effectiveness of Early Radical Cystectomy for High-Risk Non-Muscle Invasive Bladder Cancer. *Cancers* **2022**, *14*, 3797, doi:10.3390/cancers14153797.
90. Jäger, W.; Thomas, C.; Haag, S.; Hampel, C.; Salzer, A.; Thüroff, J.W.; Wiesner, C. Early vs Delayed Radical Cystectomy for ‘High-Risk’ Carcinoma Not Invading Bladder Muscle: Delay of Cystectomy Reduces Cancer-Specific Survival. *BJU International* **2011**, *108*, E284–E288, doi:10.1111/j.1464-410X.2010.09980.x.

91. Canter, D.; Egleston, B.; Wong, Y.-N.; Smaldone, M.C.; Simhan, J.; Greenberg, R.E.; Uzzo, R.G.; Kutikov, A. Use of Radical Cystectomy as Initial Therapy for the Treatment of High-Grade T1 Urothelial Carcinoma of the Bladder: A SEER Database Analysis. *Urologic Oncology: Seminars and Original Investigations* **2013**, *31*, 866–870, doi:10.1016/j.urolonc.2011.07.009.

92. Tully, K.H.; Roghmann, F.; Noldus, J.; Chen, X.; Häuser, L.; Kibel, A.S.; Sonpavde, G.P.; Mossanen, M.; Trinh, Q.-D. Quantifying the Overall Survival Benefit With Early Radical Cystectomy for Patients With Histologically Confirmed T1 Non–Muscle-Invasive Bladder Cancer. *Clinical Genitourinary Cancer* **2020**, *18*, e651–e659, doi:10.1016/j.clgc.2020.03.013.

93. Shahin, O.; Thalmann, G.N.; Rentsch, C.; Mazzucchelli, L.; Studer, U.E. A Retrospective Analysis of 153 Patients Treated With or Without Intravesical Bacillus Calmette-Guerin for Primary Stage T1 Grade 3 Bladder Cancer: Recurrence, Progression and Survival. *Journal of Urology* **2003**, *169*, 96–100, doi:10.1016/S0022-5347(05)64044-X.

## SUPPLEMENTARY MATERIAL

*Supplementary Table 1. Summary of datasets used in the study*

Study cohort	Patients count	Clinicopathological information	Image data
Objective 1 cohort A	981	Only pathological information (tumor type, stage, grade) for sample selection	Original HE slides
Objective 1 cohort A	157	Full demographic, clinical, and pathologic data	Original HE slides
Objective 2	157	Full demographic, clinical, and pathologic data	CD8 IHC slides
Objective 3	155	Full demographic, clinical, and pathologic data	CD8, ICOS, CD20, CD11c, CD163 IHC slides
Objective 4	155	Full demographic, clinical, and pathologic data	Outputs of study objectives 1 and 3

# SANTRAUKA

## Įvadas

Šlapimo pūslės vėžys (ŠPV) yra devinta dažniausia piktybinių navikų lokalizacijų. 2022 m. pasaulyje užregistruota daugiau nei 600 000 naujų atvejų ir daugiau nei 220 000 mirčių, tarp kurių vyravo vyrai [1]. Lietuvoje šlapimo pūslės vėžys taip pat yra didelė onkologinė problema – 2017 m. diagnozuota daugiau nei 400 ŠPV [2]. Nepaisant diagnostikos ir gydymo pažangos, klinikiniai rezultatai nevienodi, ypač pacientams, kurių liga dar neišplitusi į raumenį sluoksnį.

Svarbiausias modifikuojamas šlapimo pūslės vėžio rizikos veiksnys dėl sąlyčio su kancerogeniniais junginiais yra rūkymas [3]. Papildomi etiologiniai veiksniai yra profesinis aromatinių aminų poveikis, lėtinis parazitinių infekcijų, pavyzdžiui, *Schistosoma haematobium*, sukeltas uždegimas endeminiuose regionuose ir metaboliniai veiksniai, tarp jų ir nutukimas [4,5]. Daugiau nei 90 proc. šlapimo pūslės vėžio atvejų priskiriami urotelinei karcinomai, kuri išsivysto iš šlapimo takų gleivinę klojančio urotelio [6].

Urotelinė karcinoma biologiškai yra labai heterogeniška. Pagal morfologines ir molekulinės savybes urotelio karcinoma skirstoma į papiliarinius ir nepapiliarinius (plokščiuosius) navikus. Papiliarinės urotelinės karcinomos dažnai siejamos su FGFR3 valdomais molekuliniais keliais ir paprastai lėčiau progresuoja. Priešingai, plokšti pažeidimai, įskaitant karcinomą *in situ*, susiję su TP53 ir RB1 pažaidomis ir kliniškai agresyvesni [7–9]. Dvigubo kelio modelis atitinka šiuolaikinės molekulinės klasifikacijos, įskaitant tas, kurias siūlo „The Cancer Genome Atlas“ [10].

Maždaug trys ketvirtadaliai šlapimo pūslės vėžio atvejų diagnozuojami kaip raumens neinfiltruojantis šlapimo pūslės vėžys (NMIBC), kuris neinfiltrouoja šlapimo pūslės raumeninio sluoksnio. Dažniausias jo variantas – raumenis neinfiltruojanti papiliarinė urotelio karcinoma (NMIPUC) [11]. Standartinis pradinis NMIBC gydymas yra transuretrinė šlapimo pūslės naviko rezekcija (TURBT) [11]. Tačiau, nepaisant visiškos rezekcijos, NMIBC būdingas didelis recidyvų dažnis ir progresavimo į raumenis invazinės ligos rizika. Mažinant šią riziką, adjuvantinė intravezikinė terapija skiriama pagal paciento rizikos stratifikaciją. *Bacillus Calmette-Guérin* (BCG) imunoterapija yra veiksmingiausias adjuvantinis vidutinės ir didelės rizikos NMIBC gydymas, tačiau jis gali lemti pašalinius poveikius ir būti neefektyvus, dėl to labai rizikingais atvejais svarstyta ir ankstyva radikali

cistektomija [12–14]. Todėl optimizuojant NMIBC gydymą būtinas tikslus rizikos stratifikavimas.

Dabartiniai prognostiniai modeliai apima klinikinius duomenis, cistoskopinius radinius ir patologines naviko savybes: naviko stadiją, diferenciacijos laipsnį, dydį ir navikų skaičių. Šiame procese itin svarbus histologinis įvertinimas [12,15–17]. Tačiau šiuo metu naudojamos naviko diferenciacijos laipsnio nustatymo sistemos PSO 1973 ir PSO 2004/2016 pasižymi ribotu atkartojamumu ir neoptimaliu prognostiniu tikslumu. Dėl to modeliai neatspindi NMIBC biologinio kompleksiško ir nesugeba patikimai numatyti gydymo atsako [18].

Pastaraisiais metais skaitmeninė patologija ir dirbtinis intelektas (DI) tapo perspektyviomis priemonėmis, galinčiomis geriau identifikuoti naviko savybes ir prognozuoti pacientų išėitis [22–28]. Gilaus mokymosi metodai automatiškai analizuoja viso mikropreparato (angl. *whole slide image, WSI*) histologinius vaizdus ir išskiria abstrakčius vaizdo požymius, kurių aiškiai neapibrėžia stebintys žmonės. Dauginių atvejų mokymosi (angl. *multiple instance learning, MIL*) metodika, sujungdama informaciją iš daugybės naviko vaizdo sričių, padeda įveikti WSI dydžio ir naviko heterogeniškumo keliamus iššūkius [35]. Dar efektyviau prognozuoja išgyvenamumu pagrįsti MIL metodai, pavyzdžiui, dauginių atvejų išgyvenamumo mokymas (angl. *multiple instance survival learning, MISL*), į mokymą įtraukdamas išgyvenamumo, o ne binarizuotus duomenis [36].

Kartu su skaitmeninės patologijos pažanga nuspėjant vėžio klinikinę eigą ir gydymo atsaką reikšminga naviko mikroaplinka (TME) [37]. Imuninių ląstelių sudėtis ir erdvinė organizacija TME veikia naviko elgseną ir imunoterapijos, įskaitant intravezikinę BCG imunoterapiją, veiksmingumą. NMIBC atveju naviką infiltruojantys limfocitai ir su naviku susiję makrofagai buvo siejami su klinikiniais rezultatais; tačiau vien absoliutaus imuninių ląstelių tankio prognostinė vertė – ribota [21,38]. Naujesni metodai, paremti imuninių ląstelių pasiskirstymo vertinimu, pavyzdžiui, *Immunoscore* sistema, pasiūlyta Galono [63,64] ar *Immunogradiiento* indikatorių metodika, sukurta mūsų įstaigoje Rasmussono [67], parodė nepriklausomą prognostinę reikšmę įvairių piktybinių navikų pacientų imtyse [67–71].

## Tyrimo hipotezė

Skaitmeninė patologija ir dirbtiniu intelektu pagrįsti modeliai, vertinantys naviko audinį, naviko ir šeimininko sąveiką, gali pagerinti NMIPUC pacientų prognostinę stratifikaciją BCG imunoterapijos kontekste.

## Tyrimo tikslas

Sukurti integruotą prognostinį kepenų navikų mikroaplinkos ir imuninio atsako vertinimo modelį, remiantis skaitmeninės patologijos ir dirbtinio intelekto (DI) metodais.

## Tyrimo uždaviniai

1. NMIPUC audinių tyrimas abstrakčiais vaizdų požymiais, išmoktais dirbtiniuose neuroniniuose tinkluose.
2. Tirti ir optimizuoti naviko ir paciento organizmo sąveikos indikatorius NMIPUC pacientų audinių mėginiuose.
3. Identifikuoti BCG imunoterapijai predikciškai reikšmingų naviko audinio morfologinius ir mikroaplinkos faktorius.
4. Sukurti išsamius predikcinius modelius, padedančius nuspręsti, ar skirti BCG imunoterapiją NMIPUC pacientams.

## Ginamieji teiginiai

1. Dauginių atvejų išgyvenamumo mokymo (MISL) metodika pagrįstas histologinių vaizdų vertinimas suteikia nepriklausomą indėlį gerinant recidyvo rizikos prognozavimą NMIPUC pacientams, kuriems taikoma BCG imunoterapija.
2. Imuninių ląstelių tankio gradiento rodikliai, pritaikyti specifinei NMIPUC histologinei architektūrai, gali būti naudojami imuninių ląstelių pasiskirstymui epitelio ir stromos sąveikos zonoje (angl. *interface zone*, IZ) įvertinti, skirstant BCG imunoterapija gydytus NMIPUC pacientus į prognostines grupes.
3. CD11c teigiamų makrofagų gradientas, kiekybiškai įvertinamas pagal sąveikos zonos tankio santykį (IDR), yra nepriklausomas prognostinis NMIPUC recidyvo rizikos rodiklis, pranokstantis paprastus ląstelių tankio rodiklius. Taip pat pabrėžtina, kad erdvinė priešnavikinių ląstelių populiacijų (CD11c ir CD8) analizė kartu su įprastiniais klinikiniais ir patologiniais požymiais gerina rizikos stratifikaciją

## Tiriamoji medžiaga ir tyrimo metodai

Tyrimas iš dalies finansuotas pagal Lietuvos mokslo tarybos projektą („Raumens neinfiltruojančio šlapimo pūslės urotelio vėžio atsako į BCG imunoterapiją numatymas panaudojant dirbtinį intelektą“, S-MIP-21-31).

Vilniaus regioninis biomedicininų tyrimų etikos komitetas (leidimas Nr. 2021/11-1394-867, išduotas 2021 m. birželio 29 d.) patvirtino atlikto tyrimo protokolą, kuriuo tyrėjai buvo atleisti nuo pareigos gauti kiekvieno paciento sutikimą.

Atrinkdami pacientus apsiribojome NMIPUC atvejais, kad rezultatai būtų homogeniškesni. Šie navikai pasižymi specifiška histologine architektūra ir kiek vangesne klinicine eiga. Peržiūrėjome visų 230 šlapimo pūslės vėžio pacientų, gydytų BCG imunoterapija Vilniaus universiteto ligoninės Santaros klinikose. Įtraukimo kriterijus (papiliarinės urotelinės karcinomos diagnozė, pTa arba pT1 stadija, bent 6 savaičių BCG indukcijos kursas, klinikiniai duomenys ir naviko medžiaga) atitiko 166 pacientai. Iš jų 165 atvejais turėjome formaline fiksuotą į parafiną įlietą (FFPE) medžiagą. Atliekant vaizdo analizę ir pašalinus atvejus su nepakankamu naviko kiekiu, toliau tirti 157 pacientai (1, 2 uždaviniai) ir 155 pacientai (3, 4 uždaviniai).

Vaizdo požymių išgavimo (angl. *feature extraction*) neuroninio tinklo mokymui atskirai surinkome 981 pacientą, iš kurių atrinkome po vieną informatyviausią hematoksilino ir eozino (H&E) metodu nudažytą preparatą. Šioje imtyje klinikiniai duomenys rinkti nebuvo.

Imunohistocheminei analizei panaudoti CD163, CD11c, CD8, ICOS ir CD20 antikūnai. Visi imunohistocheminiai ir originalūs H&E preparatai skaitmenizuoti 20x didinant (0,5 μm/pikseliui) „Aperio® AT2 DX“ skeneriu.

Audiniai klasifikuoti etapais. Pirmiausia identifikuoti audiniai, tuomet pašalinti artefaktai, galiausiai suklasifikuoti į epitelio ir stromos klases ar identifikuotos tretinės limfoidinės struktūros. Audinių klasifikacija atlikta „HALO® AI Densenet v2“ klasifikatoriumi. CD20, CD8 ir ICOS teigiamiems limfocitams identifikuoti panaudotas „HALO® AI Multiplex IHC“ modulis, o CD11c ir CD163 teigiamiems makrofagams – „HALO® AI Nuclei Phenotyper“ klasifikatorius.

## Statistinė analizė

Kintamųjų įtaka išgyvenamumui be atkryčio (angl. *recurrence free survival*, RFS) pirmiausia vertinta vienmate Kokso regresija, taip pat atrenkant kintamuosius su p reikšme, mažesne 0,1. Iš atrinktų kintamųjų

sukurtos visos galimos kombinacijos, iš kurių atrinkti modeliai su visais kintamaisiais, nepaisant susijusių su RFS ( $p$  reikšmė  $< 0,05$ ). Šie modeliai palyginti atliekant daugialypės Kokso regresijos kryžminę validaciją, vertinant vidutinį Harrellio konkordancijos indeksą (C indeksą) ir vidutinį Akaike informacijos kriterijų (AIC) validavimo rinkiniuose. Išgyvenamumo funkcijoms palyginti tarp grupių atlikti Kaplano ir Mejerio kreivių analizė ir Log Rank testas. Statstinė analizė atlikta „Python“ programavimo kalba, naudojant „Pandas“, „Scikit-learn“, „SciPy“, „Lifelines“, „Matplotlib“ bibliotekas.

## 1 UŽDAVINYS. IŠGYVENAMUMO PREDIKCIJA PAGAL ABSTRAKČIUS VAIZDO POŽYMIUS

### Metodai

Pirmiausia suklasifikuoti vaizdai, kaip aprašyta aukščiau, ir atrinktos tinkamos naviko epitelinio komponento, stromos ir mišrios sudėties vaizdo iškarpos. Neuroniniam tinklui optimizuoti atlikome eksperimentus su skirtingu vaizdo požymių skaičiumi (2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 1536) ir neuronio tinklo paskutinio sluoksnio variacijomis (2d kovoliucinis, visiškai sujungtas). Neuroninis tinklas apmokytas sugrupuoti sumaišytas skirtingų atvejų 256 pikselių kraštinės ilgio kvadratinės vaizdo iškarpas. Eksperimentuojant vertinimo rodiklis buvo iškarpų grupavimo atkartojimo tikslumas.

Naudojant optimizuotą požymių išgavimo tinklą skirtingo dydžio vaizdo iškarpomis (256, 512 ir 1024 pikselių ilgio kraštinėmis) išgauti vaizdo požymių vektoriai. Taip pat panaudota multirezoliucinė metodika, sujungusi hierarchiškai susijusių iškarpų vaizdo savybių vektorius.

5 ir 2 metų išgyvenamumo be atkryčio klasifikacijai mokytai pasitelkta MIL metodika, pasiūlyta Ilse [35]. Metodika modifikuota taip, kad kaip įvestis naudotos ne vaizdo iškarpos, o vaizdo savybių vektoriai.

Išgyvenamumo regresijos užduočiai panaudota MISL metodika, pasiūlyta Yao [36]. Dėl nepakankamos pacientų su trumpu išgyvenamumu reprezentacijos buvo taikyta klasių balansavimo metodika ir taip pakartotinai atrinkti tyrimui pacientai.

Pacientų grupių išgyvenamumo analizei pasitelktas vienašalis Log Rank kriterijus, nes vertinant svarbus ne tik išgyvenamumo skirtumas tarp grupių, bet ir grupei teisingai priskirta rizika. Kadangi MISL rezultatas yra tolydus dydis, pacientai suskirstyti į grupes pagal ribinę vertę, pasižyminčią žemiausia

Log Rank testo p reikšme. Predikcijos reikšmės tarp pacientų grupių papildomai palygintos pagal Kruskalo ir Valio, Mano ir Vitnio U kriterijus.

## Rezultatai

Optimizavimo eksperimente tiksliausi buvo neuroniniai tinklai su 2d konvoliuciniu sluoksniu, o rezultatai buvo stabilūs mažinant požymių skaičių iki 64, toliau mažinant požymių skaičių tikslumas sparčiai prastėjo.

Geriausi rezultatai pasiekti naudojant multirezoliucinę vaizdo atranką ir MISL metodiką (C indeksas – 0,574,  $p = 0,01$ ). Stratifikavus duomenis ribine verte –0,0082 vienmatėje Kokso regresijoje, gauta HR – 2,1849,  $p = 0,0237$ .

Asociacijos su atkryčio rizika vienmatės Kokso regresijos analizės duomenimis, nustatyta teigiamos pakartotinės transuretinės rezekcijos (re-TUR) (HR – 5,018,  $p = 0,0001$ ), naviko pT1 stadijos (HR – 1,9902,  $p = 0,0187$ ) ir G3 naviko diferenciacijos laipsnis, nustatytas pagal PSO 1973 sistemą (HR – 1,8545,  $p = 0,0451$ ).

Daugialypė Kokso regresijos analizė identifikavo tris modelius su nepriklausomais kintamaisiais. Jie sudaryti iš MISL rizikos grupės kombinacijose su teigiama re-TUR (C indeksas – 0,73), naviko diferenciacijos laipsniu G3 (C indeksas – 0,64) ir pT1 naviko stadija (C indeksas – 0,63). Kadangi tiek naviko stadija, tiek diferenciacijos laipsnis nustatyti histologinio vaizdo analize, atlikus papildomą analizę nenustatytas statistiškai reikšmingas MISL rizikos pasiskirstymas tarp diferenciacijos laipsnio grupių (Kruskalo ir Valio H kriterijus = 2,02,  $p = 0,3649$ ) ir naviko stadijos (Mano ir Vitnio U kriterijus = 2884,  $p = 0,8755$ ).

## 2 UŽDAVINYS. IMUNINIO ATSAKO INDIKATORIŲ ADAPTAVIMAS IR OPTIMIZAVIMAS

### Metodai

Imunohistocheminiai dažymai, skaitmeninimas, vaizdo klasifikacija, CD8 teigiamų limfocitų identifikavimas atlikti, kaip aprašyta aukščiau. Erdvinė analizė atlikta pagal „HALO® Spatial Analysis“ modulio „Infiltration analysis“ metodą. Buvo tiriama IZ epitelio ir stromos jungtyje, apimanti iki 150  $\mu\text{m}$  epitelio ir stromos. Ši zona padalinta į 10  $\mu\text{m}$  pločio juostas, kurios perkoduotos į rangus pagal Rasmussono pasiūlytą metodiką [67], taikomą šešiakampėms gardelėms (epitelio pusėje juosta nuo 0  $\mu\text{m}$  iki 10  $\mu\text{m}$  – 1 rangas, 10  $\mu\text{m}$  iki 20  $\mu\text{m}$  – 2 rangas, ir t. t., stromos pusėje priskirti atitinkamai neigiami rangai). Pagal šių juostų ląstelių tankius apskaičiuoti imuninio atsako

indikatorių variantai, adaptuoti NMIPUC, apimantys skirtingo pločio sąveikos zonos variantus. Imunonuokryčio indikatorius (angl. *Immunodrop*, ID) apskaičiuotas kaip neigiamo (stomos) ir atitinkamo teigiamo (epitelio) rangų tankių santykis. Masės centras (angl. *ceter of mass*, CM) apskaičiuotas kaip aritmetinis rangų ir jų tankių sandaugų vidurkis. Taip pat įvertinti ir absoliutūs ląstelių tankiai bendrai sąveikos zonose bei epitelinėje ir stomos dalyse.

Duomenų rinkinys padalintas į mokymo ( $n = 117$ ) ir testavimo ( $n = 40$ ) imtis. Mokymo imties statistinė išgyvenamumo analizė atlikta, kaip aprašyta aukščiau. Geriausi indikatorių variantai atrinkti pagal žemiausią univariacinės analizės  $p$  reikšmę. Taip pat analizuotas kombinuotas teigiamos re-TUR ir naviko atkryčio anamnezėje požymis. Modelis su aukščiausia C indekso reikšme buvo įvertintas testuojant. Kaplano ir Mejerio analizei ID ir CM arbitraliai sugrupuoti į tris lygias grupes (žema, vidutinė, aukšta).

## Rezultatai

Geriausi imuninio atsako indikatorių rezultatai buvo siauroje IZ (ID variantas, naudojantis 10–20  $\mu\text{m}$  arba 2 ir –2 rangą, HR – 12,283,  $p = 0,0031$ ; CM variantas zonoje, apimančioje 20  $\mu\text{m}$  epitelio ir stomos pusėje, arba rangus nuo 2 iki –2, HR – 0,066,  $p = 0,0082$ ). Iš klinikinių ir patologinių faktorių stipriausiai su naviko atkryčio rizika buvo susijęs kombinuotas teigiamos re-TUR ir atkryčio anamnezėje požymis (HR – 5,4702,  $p = 0,0016$ ), taip pat statistiškai reikšmingi buvo atskirai teigiama re-TUR (HR – 3,6726,  $p = 0,0009$ ), naviko pT1 stadija (HR – 2,6092,  $p = 0,0126$ ), G3 diferenciacijos laipsnis pagal PSO 1973 (HR – 2,7387,  $p = 0,0159$ ).

Daugialypės Kokso regresijos analizė iš viso identifikavo 11 modelių su nepriklausomais kintamaisiais. Kryžminė validacija nustatė geriausią modelį (vidutinis C indeksas validavimo imtyse – 0,7837, AIC – 173,3428), sudarytą iš trijų kintamųjų: kombinuoto teigiamo re-TUR ir atkryčio anamnezėje požymio, G3 naviko diferenciacijos laipsnio ir ID imuninio atsako indikatoriaus. Testuojamas šis modelis taip pat taip pat patvirtino stabilius duomenis (C indeksas – 0,7429).

### 3 UŽDAVINYS. IMUNINIŲ LĄSTELIŲ SUBKLASIŲ IR JŲ ERDVINIO PASISKIRSTYMO PROGNOSTINĖS VERTĖS TYRIMAS

#### Metodai

Imunohistocheminiai dažymas, skaitmeninimas, vaizdo klasifikacija, imuninių ląstelių identifikavimas ir erdvinė analizė atlikti, kaip aprašyta aukščiau. Imuninių ląstelių pasiskirstymui IZ panaudotas supaprastintas metodas, kurį pavadino sąveikos tankių santykiu (angl. *interface density rasion*, IDR), apskaičiuojamu kaip vienodo pločio epitelinės IZ ir stromos IZ pusių ląstelių tankių santykis. Taip pat įvertinome, ar yra tretinių limfoidinių struktūrų (identifikuota pagal audinių klasifikatorių, kaip aprašyta aukščiau).

Optimalų IZ plotį nustatėme pagal ląstelių tankio rodiklius (absoliutūs tankiai ir IDR), apskaičiuotus sąveikos zonose nuo 20 µm iki 300 µm pločio (iki 150 µm į epitelinę ir stromos puses). Su šiais rodikliais atlikome vienmatės Kokso regresijos kryžminę validaciją kaip atkartojamumo rodiklį pagal C indeksą. Tolesnė statistinė analizė atlikta, kaip aprašyta aukščiau. Kaplano Mejerio analizei tolydieji kintamieji buvo suskirstyti į grupes pagal medianą kaip ribinę vertę.

#### Rezultatai

Imuninių ląstelių tankio rodiklių vienmatės Kokso regresijos analizei statistiškai reikšmingi su atkryčio rizika asocijuoti tik CD11c teigiamų makrofagų IDR (CD11c IDR, HR = 0,0012, p = 0,0002), CD8 teigiamų limfocitų IDR (CD8 IDR, HR = 0,0379, p = 0,005) ir ICOS teigiamų limfocitų IDR (ICOS IDR, HR = 0,0768, p = 0,0388). Tretinės limfoidinės struktūros identifikuotos 100 atveju, tačiau jos nebuvo statistiškai reikšmingai asocijuotos su RFS (HR – 1,69, p = 0,0522). Iš klinikinių ir pataloginių kintamųjų statistiškai reikšminga asociacija su RFS nustatyta teigiamos re-TUR (HR – 4,9321, p = 0,0001) ir pT1 naviko stadijos (HR – 2,0445, p = 0,0159).

Daugialypės Kokso regresijos analizė nustatė 13 modelių su nepriklausomais kintamaisiais. Iš jų dviejuose kryžminės validacijos metu vidutinis C indeksas buvo daugiau nei 0,7. Šiuos modelius sudarė CD11c IDR ir teigiamos re-TUR (C indeksas 0,7427) arba pT1 naviko stadijos (C indeksas 0,703) kombinacijos.

## 4 UŽDAVINYS. IŠSAMIAUS MODELIO KŪRIMAS

### Metodai

Išsamiam modeliui sukurti mes atlikome daugialypės Kokso regresijos analizę pagal aukščiau aprašytą metodiką. Į modeliavimą įtraukėme visus statistiškai reikšmingai su RFS susijusius požymius ir 3 uždavinio stratifikuotą MISL metodu nustatytą rizikos grupę; sujungtą teigiamos re-TUR ir atkryčio anamnezės požymį; ir sujungtą naviko G3 diferenciacijos laipsnio arba naviko pT1 stadijos požymį. Pastarąjį nusprendėme sujungti, kadangi naviko stadija ir diferenciacijos laipsnis statistiškai stipriai susiję ( $\chi^2 = 23,912$ ,  $p < 0,001$ ), dėl to ankstesniuose uždaviniuose nepatekdavo kartu į daugialypės Kokso regresijos modelius kaip nepriklausomi kintamieji. Univariacinė ir multivariacinė analizės atliktos, kaip aprašyta aukščiau. Geriausią modelį (aukščiausias C indeksas kryžminėje validacijoje) panaudojome atkryčio rizikos predikcijai. Tada pagal šiuos duomenis suskirstėme pacientus į rizikos grupes.

Papildomai atlikome pacientų bendro išgyvenamumo (OS) analizę pagal Kaplano Mejerio kreives ir Log Rank testą.

### Rezultatai

Vienmatėje Kokso regresijoje, be 3 uždavinyje aprašytų kintamųjų, statistiškai reikšmingai su išgyvenamumu buvo susiję ir MISL rizikos grupė (HR = 2,1554,  $p = 0,0267$ ), kombinuotas teigiamos re-TUR ir atkryčio anamnezėje požymis (HR = 5,2528,  $p = 0,0001$ ), ir kombinuotas G3 diferenciacijos laipsnio arba naviko pT1 stadijos požymis (HR = 2,5934,  $p = 0,0078$ ).

Daugialypėje Kokso regresijoje stipriausią modelį sudarė keturi kintamieji: CD11c IDR, teigiama re-TUR, kombinuotas G3 diferenciacijos laipsnis arba pT1 stadijos požymis ir MISL rizikos grupė. Pagal šį modelį įvertinus riziką ir sugrupavus pacientus į kvartilius (Q1-Q4) nustatyta, kad pacientų išgyvenamumas statiškai reikšmingai skyrėsi Q1 ir Q2 kvartiliuose ( $p = 0,045$ ) bei Q3 ir Q4 kvartiliuose ( $p = 0,0026$ ). Taip pat aptikta mažesnė pacientų, turinčių itin aukštą atkryčio riziką ( $p < 0,001$ ), grupė.

Bendro išgyvenamumo analizė nustatė statistiškai reikšmingai ilgesnį RFS Q1 pacientų grupėje, palyginti su sujungta Q2-Q4 grupe ( $p = 0,0132$ ), o itin aukštos rizikos pacientų grupėje – statiškai reikšmingai trumpesnį RFS ( $p = 0,0122$ ).

## IŠVADOS

Šios disertacijos tikslas – sukurti išsamius modelius, padedančius nuspėti NMIPUC pacientų prognozę BCG imunoterapijos kontekste panaudojant skaitmeninės patologijos ir dirbtinio intelekto metodus. Tyrimas atliktas Vilniaus universiteto ligoninės Santaros klinikose, tirti BCG imunoterapija gydyti RNIPUK pacientai (155–157 pacientai) pagal dirbtinių neuroninių tinklų metodiką ir adaptuotą Valstybinio patologijos centro erdvinio imuninių ląstelių pasiskirstymo metodiką, grįstą skaitmenine vaizdo analize.

Atliekant abstrakčių vaizdo požymių išgavimu pagrįstą tyrimą nustatyta kad optimaliausia yra multirezoliucinės vaizdo iškarpų metodika, kuri sujungia tiek smulkias ląstelių lygio vaizdo, tiek stambesnes histologinės architektūros ir audinio organizacijos detales. Palyginti su MIL metodika, MISL metodu geriau paskirstyti pacientai į rizikos grupes, tai galėtų būti susiję su detalesniais mokymo duomenimis. Taip pat šiuo metodu nustatytas rizikos faktorius parodė nepriklausomą prognostinę reikšmę multivariacinėje analizėje.

Adaptuojant imuninių ląstelių vertinimo metodiką specifinei NMIPUC, audinio architektūra nustatyta neįprastai siaura, palyginti su kitais navikais IZ [67–71]. Nepaisant metodikos skirtumo pavyko adaptuoti abiejų (ID ir CM) ląstelių erdvinio pasiskirstymo indikatorių koncepcijas NMIPUC imuninei aplinkai įvertinti. Abu indikatoriai buvo statiškai reikšmingai, asocijuoti su NMIPUC pacientų RFS BCG imunoterapijos kontekste.

Imuninių ląstelių subpopuliacijų tyrimu nustatėme, kad IDR metodika įvertintas CD11c teigiamų makrofagų, CD8 teigiamų ir ICOS teigiamų limfocitų pasiskirstymas IZ suteikė reikšmingos papildomos informacijos, padedančios nuspėti NMIPUC pacientų išgyvenamumą be atkryčio. Kadangi CD11c makrofagai ir CD8 T limfocitai pasižymi priešnavikiniu veikimu, galima daryti išvadą, kad IDR metodika įvertintas ląstelių pasiskirstymas atspindi priešnavikinę imuninės sistemos mobilizaciją. Taip atkreiptinas dėmesys į tai, kad 2 ir 3 uždaviniais nustatytos erdvinio ląstelių pasiskirstymo sąsajos su pacientų išgyvenamumu buvo daug stipresnės, palyginti su navikinio audinio absoliučiais ląstelių tankiais. Tai pabrėžia, kad svarbu vertinti šio naviko erdvinio ląstelių pasiskirstymą.

Taip pat reikšmingai su pacientų išgyvenamumu be atkryčio buvo susiję ir anksčiau žinomi klinikiniai ir patologiniai kriterijai. Naviko stadija ir diferenciacijos laipsnis yra plačiai taikomi NMIPUC rizikos vertinimo kriterijai. Tačiau naviko diferenciacijos laipsnis, nustatytas pagal PSO 1973 sistemą, buvo stipresnis prognostinis faktorius galimai dėl specifinės BCG

gydytų pacientų imties, į kurią patenka aukštesnės rizikos pacientai, todėl PSO 2004/2016 sistema nustatytų aukšto ir žemo laipsnio navikų grupių pasiskirstymas buvo nesubalansuotas, o tai lėmė menkesnę prognostinę vertę. Stipriausiai su NMIPUC pacientų išgyvenamumu buvo asocijuota teigiama re-TUR. Šis faktorius taip pat yra jau anksčiau identifikuotas, tačiau galimai dėl skirtingos re-TUR taktikos nėra visuotinai priimtas rizikos vertinimo kriterijus [81–83].

Paskutinio uždavinio sukurtas išsamus modelis apėmė tiek imuninės mikroaplinkos faktorių (CD11c IDR), tiek abstrakčiais vaizdo požymiais paremtą požymį (MISL rizikos grupę) kartu su teigiama re-TUR ir sujungto naviko stadijos ir diferenciacijos laipsnio požymį. Tokia modelio sudėtis parodo visų kriterijų (klinikinių, patologinių, naviko mikroaplinkos ir abstrakčių vaizdo požymių) nepriklausomą prognostinę vertę.

Pagrindinis šio tyrimo trūkumas – nebuvo išorinės validacijos. Dėl to šie radiniai negali būti tiesiogiai perkelti į klinikinę praktiką, pirmiausia reikėtų validuoti kitose įstaigose. Be to, būtų naudingi papildomi tyrimai, giliau analizuojantys biologinius procesus, lemiančius imuninės sistemos atsaką į NMIPUS BCG imunoterapijos kontekste. Taip pat vienas iš pritaikomumą ribojančių faktorių galėtų būti tai, kad procesas automatizuotas ne iki galo, tai galėtų lemti prastesnę atkartojamumą.

### Apibendrinimas

Šiame tyrime taikyti dirbtinio intelekto ir naviko mikroaplinkos vertinimo metodai suteikia papildomos informacijos apie NMIPUC pacientų prognozę, kuri kartu su plačiai naudojamais rizikos faktoriais galėtų būti integruojama į pacientų rizikos vertinimą. Tačiau tam, kad plačiau būtų taikomi šie metodai, būtina išorinė jų validacija. Šis tyrimas atskleidė mikroaplinkos ląstelių pasiskirstymo vertinimo svarbą ir dirbtinio intelekto metodų galimybes iš histologinio vaizdo išgauti reikšmingą prognostinę informaciją, kuri yra nepriklausoma nuo žmogaus vertinimu paremtų klasikinių rizikos faktorių.

# CURRICULUM VITAE

## Education

- 2021 – 2025 Doctorate studies, Vilnius University
- 2010 – 2014 Residency of Pathology, Vilnius University
- 2009 – 2010 Internship at Klaipėda Seaman's hospital
- 2003 – 2009 Master's degree in medicine, Vilnius University
- 2001 – 2003 Secondary education, Saugų Jurgio Mikšo Secondary School
- 1991– 2001 Primary education, Vilkyčių primary School

## Professional Experience

- 2014 – 2026 Pathologist specializing in hematopathology, urologic pathology, soft tissue pathology and pulmonary pathology. Supervisor of molecular pathology at National Center of Pathology, Affiliate of Vilnius University Hospital Santaros Klinikos
- 2016 – 2026 Lecturer and junior assistant manager at Faculty of Medicine, Institute of Biomedical Sciences, Department of Pathology and Forensic Medicine, Vilnius University
- 2010 – 2014 Resident of pathology at National Center of Pathology, Affiliate of Vilnius University Hospital Santaros Klinikos
- 2010 – 2014 Pathologist's assistant at National Center of Pathology, Affiliate of Vilnius University Hospital Santaros Klinikos
- 2009 – 2010 Emergency room doctors assistant at Klaipėda Seaman's hospital

## LIST OF DOCTORAL PUBLICATIONS AND PRESENTATIONS

### Publications

1. Drachneris, J.; Rasmusson, A.; Morkunas, M.; Fabijonavicius, M.; Cekauskas, A.; Jankevicius, F.; Laurinavicius, A. CD8+ Cell Density Gradient across the Tumor Epithelium–Stromal Interface of Non-Muscle Invasive Papillary Urothelial Carcinoma Predicts Recurrence-Free Survival after BCG Immunotherapy. *Cancers* 2023, 15, 1205, doi:10.3390/cancers15041205.
2. Drachneris, J.; Morkunas, M.; Fabijonavicius, M.; Cekauskas, A.; Jankevicius, F.; Laurinavicius, A. Prediction of Non-Muscle Invasive Papillary Urothelial Carcinoma Relapse from Hematoxylin–Eosin Images Using Deep Multiple Instance Learning in Patients Treated with Bacille Calmette–Guérin Immunotherapy. *Biomedicines* 2024, 12, 360, doi:10.3390/biomedicines12020360.
3. Drachneris, J.; Morkunas, M.; Fabijonavicius, M.; Cekauskas, A.; Jankevicius, F.; Laurinavicius, A. Spatial Distribution of Macrophage and Lymphocyte Subtypes within Tumor Microenvironment to Predict Recurrence of Non-Muscle-Invasive Papillary Urothelial Carcinoma after BCG Immunotherapy. *International Journal of Molecular Sciences* 2024, 25, 4776, doi:10.3390/ijms25094776.

### Presentations

1. Digital Pathology & AI Congress (2022.12.07-08 London, UK). Poster presentation. “Indicators of CD8+ Cell Gradients across the Tumour Epithelium Stromal Interface of Non Muscle Invasive Papillary Urothelial Carcinoma predicts Relapse Free Survival after BCG Immunotherapy”.
2. European Congress on Digital Pathology ECDP2023 (2023.06.14-17 Budapest, Hungary). Poster presentation. “Prediction of Non Muscle Invasive Papillary Urothelial Carcinoma Relapse from Hematoxylin Eosin Images using Deep Multiple Instance Learning in patients treated with BCG immunotherapy”.
3. Dirbtinio intelekto technologijos medicinoje: tyrimai ir diagnostika (2023.10.20, Vilnius Lithuania). Poster presentations. “Indicators of CD8+ Cell Gradients across the Tumour Epithelium Stromal Interface of Non Muscle Invasive Papillary Urothelial Carcinoma predicts Relapse Free Survival after BCG

Immunotherapy” and “Prediction of Non Muscle Invasive Papillary Urothelial Carcinoma Relapse from Hematoxylin Eosin Images using Deep Multiple Instance Learning in patients treated with BCG immunotherapy”.

4. European Congress on Digital Pathology ECDP2024 (2024.06.5-8 Vilnius, Lithuania). Oral presentation. “Prognostic Significance of the Spatial Distribution of TAMs and TILs in Non-muscle-invasive Papillary Urothelial Carcinoma”.
5. Evolutionary Medicine: How evolutionary thinking can contribute to the medical and health sciences (2024.06.18-21, Vilnius, Lithuania). Oral presentation. “Digital Image Analysis of the Bladder Cancer: Deeper and More Precise Insights into the Tumor”

Vilniaus universiteto leidykla  
Saulėtekio al. 9, III rūmai, LT-10222 Vilnius  
El. p. [info@leidykla.vu.lt](mailto:info@leidykla.vu.lt), [www.leidykla.vu.lt](http://www.leidykla.vu.lt)  
[bookshop.vu.lt](http://bookshop.vu.lt), [journals.vu.lt](http://journals.vu.lt)  
Tiražas 35 egz.