Predicting Biochemical Recurrence from Prostatectomy Slides - the LEOPARD Challenge

Khrystyna Faryna[1]
Clément Grisi[1]
Nefise Uysal[1]
[1] *Department of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands*
Jeroen van der Laak[1,2]
Geert Litjens[1]
[1] *Department of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands*
[2] *Center for Medical Image Science and Visualization, Linköping University, Linköping, Sweden*
The LEOPARD Challenge Consortium[3]
[3] *The complete list of consortium authors is provided in Appendix A.*

Editors: Under Review for MIDL 2026

Abstract

Prostate cancer affects over 1.4 million men yearly, with approximately 30% experiencing biochemical recurrence (BCR) after prostatectomy. Current clinical tools predict risk at the population level but are less accurate for individuals. The LEarning biOchemical Prostate cAncer Recurrence from histopathology sliDes (LEOPARD) challenge benchmarked AI models to predict time to BCR directly from H&E-stained prostatectomy slides using 2,181 cases from four countries. Sixteen teams participated; nine models were selected for final evaluation. Top AI models achieved a C-index of 0.740, comparable to histopathological grading (0.739) and slightly below Cancer of the Prostate Risk Assessment Post-surgical (CAPRA-S) - 0.785. Combining AI with histopathological grading improved performance to 0.766, and with CAPRA-S to 0.799, significantly enhancing BCR risk prediction.

Keywords: Prostate Cancer, Biochemical Recurrence, Deep Learning, AI, Histopathology

1. Introduction

Prostate cancer affects 1.4 million men annually (Sung et al., 2021). After prostatectomy, PSA levels are used to monitor recurrence, typically dropping below 0.1 ng/mL within weeks (Goonewardene et al., 2014). However, 30% of patients develop biochemical recurrence (BCR), linked to worse outcomes (Freedland et al., 2005; Han et al., 2001). While PSA screening is debated (Force, 2018; Heijnsdijk et al., 2018), it remains essential post-surgery. Risk stratification relies on models like Cancer of the Prostate Risk Assessment (CAPRA), pre-operative (Cooperberg et al., 2005) and Cancer of the Prostate Risk Assessment Surgical (CAPRA-S), post-operative (Cooperberg et al., 2011), which combine clinical and pathological factors to group patients by risk (Cornford et al., 2024). Despite clinical value, these models are limited at the individual level due to interreader variability and potential incomplete use of available data (Epstein, 2010; Pierorazio et al., 2013).

Recently, it was shown that AI can predict BCR directly from histopathology slides, potentially capturing subtle morphological features not incorporated into conventional grading (Pinckaers et al., 2022; Eminaga et al., 2024; Dietrich et al., 2021).

We introduce the LEarning biOchemical Prostate cAncer Recurrence from histopathology sliDes (LEOPARD) challenge to benchmark AI models for predicting time to BCR from H&E-stained prostatectomy slides.

2. Materials and Methods

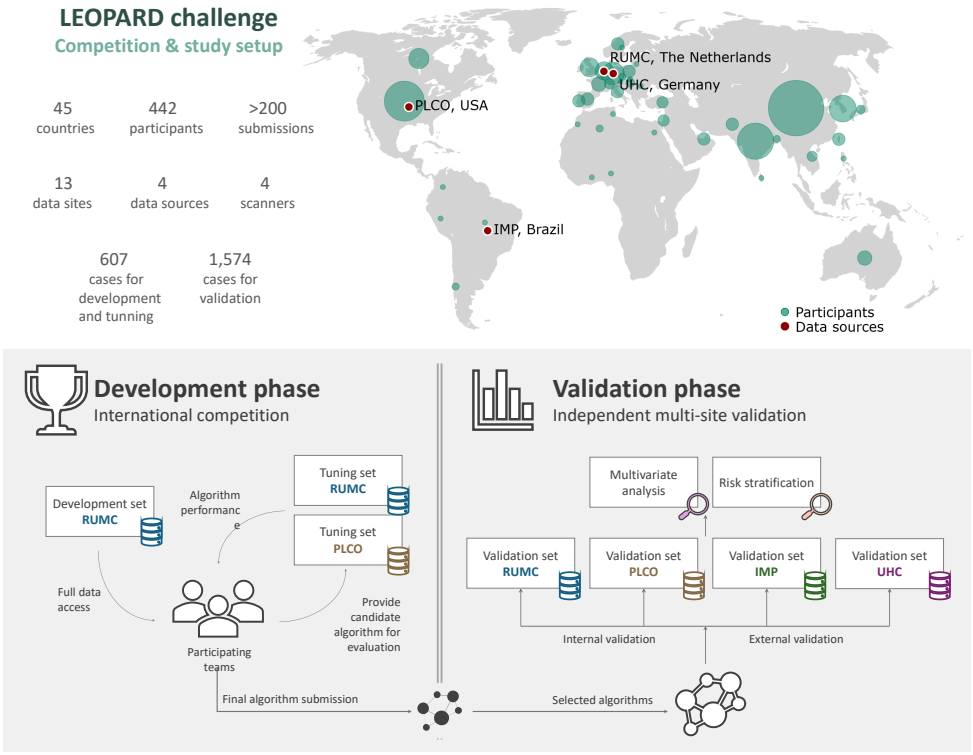

Figure 1: Study setup and implementation.

We retrospectively collected whole-slide images of prostate biopsies, along with clinical and pathological data, from Radboud University Medical Center (RUMC), the Netherlands, Institute Mario Penna (IMP), Brazil, University hospital Cologne (UHC), Germany; in addition, the data from the PLCO Cancer Screening Trial was used in this study(Gohagan et al., 2000; Andriole et al., 2009). The dataset comprised of prostatectomy slides from 2,181 patients, with relevant clinical parameters such as Gleason grade, TNM stage, PSA levels, BCR status, and time to BCR or last follow-up. BCR was defined as a postoperative PSA rise, with center-specific thresholds: RUMC, PLCO, and IMP used $\geq$ 0.2 ng/mL confirmed by a second measurement, while UHC used a more sensitive threshold of $>$ 0.1 ng/mL based on EAU-guided follow-up. The dataset was split into development, tuning, and internal and external validation sets to develop and assess predictive models.

The RUMC cohort consisted of 657 patients who underwent prostatectomy between 1992 and 2012. The slides were scanned with 3DHistech P1000 scanner at $0.25\mu m/pixel$ resolution. The radical prostatectomy PLCO cohort data was collected between November 1993 and July 2008 from ten medical centers across the USA. The final test cohort consisted of 773 patient cases. The slides were scanned using a Leica Biosystems scanner at $0.25\mu m/pixel$. The IMP cohort included data from 421 patients who underwent radical

prostatectomy at IMP, São Paulo, Brazil, in 2016. The cases were followed for 6 years, with 2022 being the last year of follow-up. The slides were scanned using Motic Easy Scan Infinity 60N scanner at a resolution of $0.26\mu m/pixel$. The UHC dataset cohort comprises radical prostatectomy specimens collected between 2015 and 2020 from 330 patients treated at UHC, Cologne, Germany. Follow-up extended through 2023. The WSIs of prostate tissue specimens were scanned using the Hamamatsu Nanozoomer S360 digital slide scanner at a resolution of $0.23\mu m/pixel$.

Participating teams were granted access to the training dataset from RUMC, and AI models were developed to predict time to BCR. During the model development phase, teams trained their algorithms using the training set and submitted models to an evaluation platform for performance assessment. Intermediate performance estimates were provided using a tuning dataset comprised of 49 patients' prostatectomies from RUMC and 50 patients' prostatectomies from PLCO.

The final model evaluation was conducted on a blinded validation set of 823 patients' prostatectomies from the RUMC and PLCO. The challenge was handled via https://grand-challenge.org/ platform. After initial evaluation, selected teams participated in an external validation phase, during which their models were tested on two independent datasets: one comprising prostatectomies of 330 patients from the UHC and another containing prostatectomies of 421 patients from IMP.

To assess potential added value to existing tools, models' predictions were combined with histological grading and CAPRA-S (clinical variable based on blood PSA levels, histological grading, lymph node status, seminal vesicle invasion, and extracapsular extension) using Cox Proportional Hazard models.

Model performance was evaluated using concordance indices (C-indices) for recurrence prediction. 95% Confidence intervals were computed using bootstrapping. Finally, we conducted a two-sided permutation test to determine whether the CAPRA-S + AI ensemble significantly outperformed CAPRA-S alone.

3. Results and Discussion

The LEOPARD challenge, Figure 1, (April–September 2024) attracted 442 participants from 40+ countries, with over 206 submissions. Sixteen teams submitted final models, of which eight were selected for external evaluation. Most approaches utilized Foundational models and Multiple Instance Learning.

Top models achieved strong internal performance (C-index up to 0.723 RUMC, 0.732 PLCO) and generalized well externally (0.773 IMP, 0.702 UHC).

The top-five algorithms were combined into an ensemble. The AI ensemble reached an overall C-index of 0.740 (95% CI, 0.716–0.765), comparable to 0.739 ISUP grading and 0.785 CAPRA-S. Combining AI with clinical tools results in improved risk stratification performance. AI + ISUP: C-index 0.766 (95% CI, 0.743–0.788) and AI + CAPRA-S: C-index 0.799 (95% CI, 0.778–0.819), significantly outperforming CAPRA-S alone.

AI models ensemble matched traditional grading performance using only histopathology and showed consistent predictive performance across international cohorts. Combining AI with ISUP or CAPRA-S improved prognostic accuracy, demonstrating complementary value beyond existing clinical tools.

Acknowledgments

We thank the Grand-Challenge team for their support during the course of the challenge.

References

Gerald L. Andriole, E. David Crawford, Robert L. Grubb, Saundra S. Buys, David Chia, Timothy R. Church, Mona N. Fouad, Edward P. Gelmann, Paul A. Kvale, Douglas J. Reding, Joel L. Weissfeld, Lance A. Yokochi, Barbara O'Brien, Jonathan D. Clapp, Joshua M. Rathmell, Thomas L. Riley, Richard B. Hayes, Barnett S. Kramer, Grant Izmirlian, Anthony B. Miller, Paul F. Pinsky, Philip C. Prorok, John K. Gohagan, and Christine D. Berg. Mortality results from a randomized prostate-cancer screening trial. *New England Journal of Medicine*, 360(13):1310–1319, 2009. doi: 10.1056/NEJMoa0810696. URL https://www.nejm.org/doi/full/10.1056/NEJMoa0810696.

Matthew R. Cooperberg, David J. Pasta, Eric P. Elkin, Mark S. Litwin, David M. Latini, Janeen Du Chane, and Peter R. Carroll. The University of California, San Francisco Cancer of the Prostate Risk Assessment score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy. *J. Urol.*, 173(6):1938–1942, June 2005. doi: 10.1097/01.ju.0000158155.33890.e7. URL https://pubmed.ncbi.nlm.nih.gov/15879786/.

Matthew R. Cooperberg, Joan F. Hilton, and Peter R. Carroll. The CAPRA-S score: A straightforward tool for improved prediction of outcomes after radical prostatectomy. *Cancer*, 117(22):5039–5046, November 2011. doi: 10.1002/cncr.26169. URL https://pubmed.ncbi.nlm.nih.gov/21647869/.

Philip Cornford, Roderick C.N. van den Bergh, Erik Briers, Thomas Van den Broeck, Oliver Brunckhorst, Julie Darraugh, Daniel Eberli, Gert De Meerleer, Maria De Santis, Andrea Farolfi, Giorgio Gandaglia, Silke Gillessen, Nikolaos Grivas, Ann M. Henry, Michael Lardas, Geert J.L.H. van Leenders, Matthew Liew, Estefania Linares Espinos, Jan Oldenburg, Inge M. van Oort, Daniela E. Oprea-Lager, Guillaume Ploussard, Matthew J. Roberts, Olivier Rouvière, Ivo G. Schoots, Natasha Schouten, Emma J. Smith, Johan Stranne, Thomas Wiegel, Peter-Paul M. Willemse, and Derya Tilki. Eau-eanm-estro-esur-isup-siog guidelines on prostate cancer—2024 update. part i: Screening, diagnosis, and local treatment with curative intent. *European Urology*, 86(2):148–163, 2024. ISSN 0302-2838. doi: https://doi.org/10.1016/j.eururo.2024.03.027. URL https://www.sciencedirect.com/science/article/pii/S0302283824022541.

Esther Dietrich, Patrick Fuhlert, Anne Ernst, Guido Sauter, Maximilian Lennartz, H. Siegfried Stiehl, Marina Zimmermann, and Stefan Bonn. Towards explainable end-to-end prostate cancer relapse prediction from hamp;e images combining self-attention multiple instance learning with a recurrent neural network. In Subhrajit Roy, Stephen Pfohl, Emma Rocheteau, Girmaw Abebe Tadesse, Luis Oala, Fabian Falck, Yuyin Zhou, Liyue Shen, Ghada Zamzmi, Purity Mugambi, Ayah Zirikly, Matthew B. A. McDermott, and Emily Alsentzer, editors, *Proceedings of Machine Learning for Health*, volume 158

of *Proceedings of Machine Learning Research*, pages 38–53. PMLR, 04 Dec 2021. URL https://proceedings.mlr.press/v158/dietrich21a.html.

Okyaz Eminaga, Fred Saad, Zhe Tian, Ulrich Wolffgang, Pierre I. Karakiewicz, Véronique Ouellet, Feryel Azzi, Tilmann Spieker, Burkhard M. Helmke, Markus Graefen, Xiaoyi Jiang, Lei Xing, Jorn H. Witt, Dominique Trudel, and Sami-Ramzi Leyh-Bannurah. Artificial intelligence unravels interpretable malignancy grades of prostate cancer on histology images. *npj Imaging*, 2(1):6, Mar 2024. ISSN 2948-197X. doi: 10.1038/s44303-023-00005-z. URL https://doi.org/10.1038/s44303-023-00005-z.

Jonathan I Epstein. An update of the gleason grading system. *J. Urol.*, 183(2):433–440, February 2010.

US Preventive Services Task Force. Screening for prostate cancer: Us preventive services task force recommendation statement. *JAMA*, 319(18):1901–1913, 05 2018. ISSN 0098-7484. doi: 10.1001/jama.2018.3710. URL https://doi.org/10.1001/jama.2018.3710.

Stephen J. Freedland, Elizabeth B. Humphreys, Leslie A. Mangold, Mario Eisenberger, Frederick J. Dorey, Patrick C. Walsh, and Alan W. Partin. Risk of prostate cancer–specific mortality following biochemical recurrence after radical prostatectomy. *JAMA*, 294(4):433–439, 07 2005. ISSN 0098-7484. doi: 10.1001/jama.294.4.433. URL https://doi.org/10.1001/jama.294.4.433.

John K. Gohagan, Philip C. Prorok, Richard B. Hayes, and Barnett-S. Kramer. The prostate, lung, colorectal and ovarian (plco) cancer screening trial of the national cancer institute: History, organization, and status. *Controlled Clinical Trials*, 21(6, Supplement 1):251S–272S, 2000. ISSN 0197-2456. doi: https://doi.org/10.1016/S0197-2456(00)00097-0. URL https://www.sciencedirect.com/science/article/pii/S0197245600000970.

Sanchia S. Goonewardene, Jaspal S. Phull, Amit Bahl, and Raj A. Persad. Interpretation of psa levels after radical therapy for prostate cancer. *Trends in Urology & Men's Health*, 5(4):30–34, 2014. doi: https://doi.org/10.1002/tre.407. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/tre.407.

Misop Han, Alan W. Partin, Charles R. Pound, Jonathan I. Epstein, and Patrick C. Walsh. Long-term biochemical disease-free and cancer-specific survival following anatomic radical retropubic prostatectomy. the 15-year johns hopkins experience. *Urologic Clinics of North America*, 28(3):555–565, 2001. ISSN 0094-0143. doi: https://doi.org/10.1016/S0094-0143(05)70163-4. URL https://www.sciencedirect.com/science/article/pii/S0094014305701634.

Eveline A.M. Heijnsdijk, Chris H. Bangma, Josep M. Borràs, Tiago M. de Carvalho, Xavier Castells, Martin Eklund, Josep A. Espinàs, Markus Graefen, Henrik Grönberg, Iris Lansdorp-Vogelaar, Pim J. van Leeuwen, Vera Nelen, Franz Recker, Monique J. Roobol, Pieter Vandenbulcke, and Harry J. de Koning. Summary statement on screening for prostate cancer in europe. *International Journal of Cancer*, 142(4):741–746, 2018. doi: https://doi.org/10.1002/ijc.31102. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.31102.

Phillip M. Pierorazio, Patrick C. Walsh, Alan W. Partin, and Jonathan I. Epstein. Prognostic gleason grade grouping: data based on the modified gleason scoring system. *BJU International*, 111(5):753–760, 2013. doi: https://doi.org/10.1111/j.1464-410X.2012.11611.x. URL https://bjui-journals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1464-410X.2012.11611.x.

Hans Pinckaers, Jolique van Ipenburg, Jonathan Melamed, Angelo De Marzo, Elizabeth A. Platz, Bram van Ginneken, Jeroen van der Laak, and Geert Litjens. Predicting biochemical recurrence of prostate cancer with artificial intelligence. *Communications Medicine*, 2(1):64, Jun 2022. ISSN 2730-664X. doi: 10.1038/s43856-022-00126-3. URL https://doi.org/10.1038/s43856-022-00126-3.

Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3):209–249, 2021. doi: https://doi.org/10.3322/caac.21660. URL https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21660.

Appendix A. The Leopard Challenge Consortium

Khrystyna Faryna[1], **Clément Grisi**[1], **Nefise Uysal**[1], **Hans Pinckaers**[1], **Joep Bogaerts**[1], **Solène-Florence Kammerer-Jacquet**[2], **Pierre Allaume**[2], **Vittorio Agosti**[1,3], **Enrico Munari**[3,4], **Paulo Guilherme de Oliveira Salles**[5], **Yuri Tolkach**[6], **Reinhard Büttner**[6], **Sofiya Semko**[7], **Maksym Pikul**[7], **Axel Heidenreich**[7], **Adam Kowalewski**[1,8,9], **Leslie Tessier**[1,10,11], **Frédérique Meeuwsen**[1], **Jolique van Ipenburg**[1], **Till Nicke**[12], **Johannes Lotz**[12], **Jan Raphael Schäfer**[12], **Matthew McNeil**[13], **Anne L. Martel**[14], **Witali Aswolinskiy**[15], **Christian Aichmüller**[15], **Zijie Fang**[16], **Songhan Jiang**[17], **Xiangming Yan**[17], **Qi Ouyang**[18], **Yongbing Zhang**[18], **Surya Achanta**[19], **Nilanjan Chattopadhyay**[19], **Raja Fida Mohammad**[19], **Aditya Vartak**[19], **Nitin Singhal**[19], **Li Zhang**[20], **Aaron Kutzer**[20], **Narmin Ghaffari Laleh**[20,21], **Jakob Nikolas Kather**[20,21,22], **Doanh Cao Bui**[23], **Anh Tien Nguyen**[23], **Sunhong Park**[23], **Jin Tae Kwak**[23], **Suhang You**[24,25], **Sanyukta Adap**[24], **Siddhesh Thakur**[24], **Bhakti Baheti**[24,25], **Spyridon Bakas**[24,25,26], **Jeroen van der Laak**[1,27], **Geert Litjens**[1].

[1] Department of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands
[2] Department of Pathology, CHU de Rennes, Rennes, France
[3] Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy
[4] Surgical Pathology Unit, Verona University Hospital Trust, Verona, Italy
[5] Anatomical Pathology Service, Instituto Mário Penna, Belo Horizonte, Brazil
[6] Institute of Pathology, University Hospital Cologne, Cologne, Germany
[7] Clinic of Urology, University Hospital Cologne, Cologne, Germany
[8] Faculty of Medicine, Bydgoszcz University of Science and Technology, Bydgoszcz, Poland
[9] Department of Tumor Pathology, Oncology Centre Prof. Franciszek Łukaszczyk Memorial

Hospital, Bydgoszcz, Poland
[10] Pathology Department, Institut du Cancer de l'Ouest, Angers, France
[11] Pathology Department, Centre hospitalier universitaire d'Angers, Angers, France
[12] Fraunhofer Institute for Digital Medicine MEVIS, Bremen, Germany
[13] Department of Medical Biophysics, University of Toronto, Toronto, Canada
[14] Sunnybrook Research Institute, Toronto, Canada
[15] PAICON GmbH, Heidelberg, Germany
[16] Tsinghua Shenzhen International Graduate School, Tsinghua University, China
[17] School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China
[18] School of Science, Harbin Institute of Technology, Shenzhen, China
[19] AIRA Matrix, Mumbai, India
[20] Else Kroener Fresenius Center for Digital Health, TUD Dresden University of Technology, Dresden, Germany
[21] Medical Oncology, National Center for Tumor Diseases, University Hospital Heidelberg, Heidelberg, Germany
[22] Department of Medicine, University Hospital Dresden, Dresden, Germany
[23] School of Electrical Engineering, Korea University, Seoul, Republic of Korea
[24] Division of Computational Pathology, Department of Pathology & Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA
[25] Indiana University Melvin and Bren Simon Comprehensive Cancer Center, Indianapolis, IN, USA
[26] Department of Computer Science, Luddy School of Informatics, Computing, and Engineering, Indiana University, Indianapolis, IN, USA
[27] Center for Medical Image Science and Visualization, Linköping University, Linköping, Sweden