Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer

Sebastian Foersch[1] , Christina Glasner[1], Ann-Christin Woerl[1,2],Markus Eckstein[3], Daniel-Christoph Wagner[1], Stefan Schulz[1], Franziska Kellers[1,4], Aurélie Fernandez[1], Konstantina Tserea[1], Michael Kloth[1], Arndt Hartmann[3], Achim Heintz[5], Wilko Weichert[6], Wilfried Roth[1], Carol Geppert[3], Jakob Nikolas Kather[7,8,9] & Moritz Jesinghaus[6,10]

1 Institute of Pathology, University Medical Center Mainz, Mainz, Germany.
2 Institute of Computer Science, Johannes Gutenberg University Mainz, Mainz, Germany.
3 Institute of Pathology and Comprehensive Cancer Center EMN, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany.
4 Department of Pathology, University Hospital Schleswig-Holstein, Kiel, Germany.
5 Department of General Visceral and Vascular Surgery, Marien Hospital Mainz, Mainz, Germany.
6 Institute of Pathology, Technical University Munich, Munich, Germany.
7 Department of Medicine III, University Hospital RWTH Aachen, Aachen, Germany.
8 Pathology and Data Analytics, Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, UK.
9 Else Kroener Fresenius Center for Digital Health, Medical Faculty Carl Gustav Carus, Technical University Dresden, Dresden, Germany.
10 Institute of Pathology, University Hospital Marburg, Marburg, Germany.

INTRODUCTION AND PURPOSE

Although it has been long known that the immune cell composition has a strong prognostic and predictive value in colorectal cancer (CRC), the relevant scoring systems are not well established. Thus, we established and evaluated a multi stain deep learning model (MSDLM) utilizing artificial intelligence (AI) to determine the AImmunoscore (AIS) in more than 1,000 patients with CRC. AIS was found to be a strong predictor of patient outcomes, and their response to pre-surgery treatment. It outperformed clinical, molecular, and immune cell-based parameters. Hence, AIS provides clinicians with a valuable decision-making tool based on the tumor immune microenvironment.

METHODS

Study design:

- 4 cohorts of patients.
  - Inclusion criteria: patients with the diagnosis of a colorectal adenocarcinoma of any UICC stage.
  - For each patient, the initial, pre-therapeutic, diagnostic biopsy specimen, as well as information about tumor regression from the final surgical specimen were collected.
  - All experiments were in accordance with the Declaration of Helsinki.
- Immunohistochemistry, digitalization and preprocessing: TMAs were created from FFPE tissue blocks, with cores from tumor centers and invasive margins.
- Immunoscore: 3 methods which calculated the density of CD3 and CD8 T cells in tumor tissue to define immune scores.
- BRAF and MMR status
- Model architecture: Multistain Deep Learning Model (MSDLM) to analyze multiple stained tissue slides. This model utilizes individual neural networks (ResNet18) for each stain.
- Training and evaluation: The models were trained on 3-year relapse-free survival (RFSS) data, with relapse or no relapse as the target.

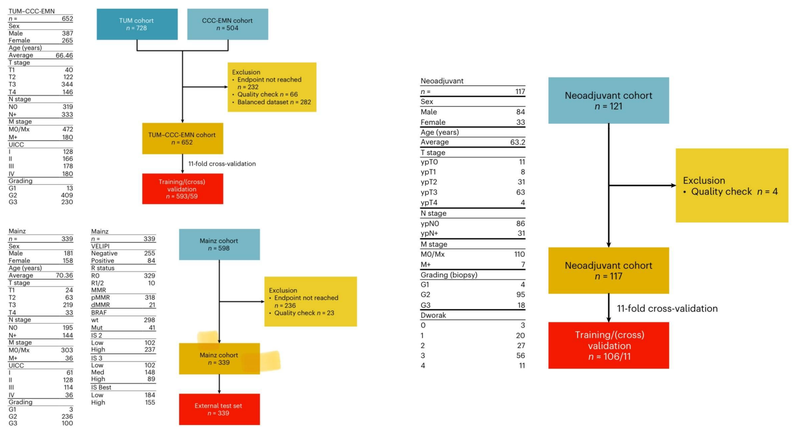

***Figure 1**:* Clinical characteristics and CONSORT diagrams for the prognostic cohorts (left) and for the neoadjuvant cohort (right)

Statiscal analysis:

- Unpaired t-test and One-way ANOVA with the Dunnett's test to compare between two or more groups
- Uni- and multivariable Cox regression was used for prognosis analyses. Wald test to calculate statistical significance.
- Image analysis and deep learning experiments were conducted using QuPath and Python/PyTorch, respectively.

RESULTS

The diagnostic accuracy of the MSDLM was greater than that of SSDLMs. (see Figure 2 below) The MSDLM also exceeded the prognostic capabilities of classical machine learning methods, which relied solely on immune cell counts.

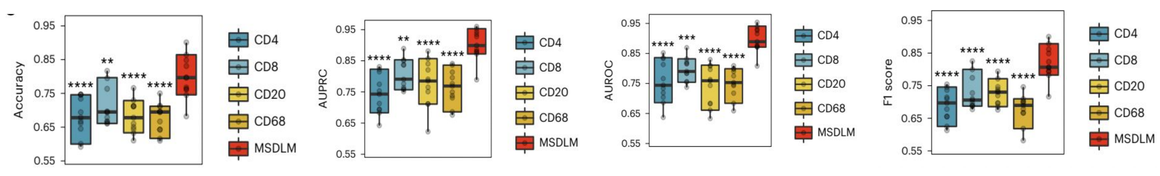

Figure 2: Training and cross-validation of the MSDLM

The AIS, derived from the MSDLM, is proven to provide better patient stratification than traditional methods based on kaplan-meier analysis and was validated by an independent external cohort, where it produced robust prognostic accuracy ($P \leq 0.0001$). Patients with a low AIS (AISlow) had significantly poorer relapse-free survival (hazard ratio: 6.24, $P \leq 0.0001$) than those with a high AIS (AIShigh).

Furthermore, the MSDLM showed promise in predicting responses to neoadjuvant radiochemotherapy in rectal cancer patients achieving accuracy, AUPRC, AUROC and F1 scores higher that SSDLMs. (see Figure 3 below) These findings highlight the potential of deep learning models, particularly the MSDLM, as powerful tools for prognostic and predictive applications in colorectal cancer and beyond.

	MSDLM	CD4 SSDLM (P-value)	CD8 SSDLM (P-value)	CD68 SSDLM (P-value)	CD20 SSDLM (P-value)
Accuracy (%)	79.6 ± 6.13	68.2 ± 5.63 (P ≤ 0.0001)	70.9 ± 5.65 (P ≤ 0.01)	67.6 ± 4.37 (P ≤ 0.0001)	68.3 ± 4.74 (P ≤ 0.0001)
AUPRC	0.902 ± 0.047	0.742 ± 0.063 (P ≤ 0.0001)	0.806 ± 0.048 (P ≤ 0.01)	0.765 ± 0.055 (P ≤ 0.0001)	0.776 ± 0.075 (P ≤ 0.0001)
AUROC	0.896 ± 0.040	0.757 ± 0.070 (P ≤ 0.0001)	0.796 ± 0.039 (P ≤ 0.001)	0.752 ± 0.049 (P ≤ 0.0001)	0.742 ± 0.068 (P ≤ 0.0001)
F1 Score	0.819 ± 0.051	0.690 ± 0.048 (P ≤ 0.0001)	0.756 ± 0.050 (P ≤ 0.0001)	0.678 ± 0.045 (P ≤ 0.0001)	0.726 ± 0.036 (P ≤ 0.0001)

Figure 3. Table of accuracy, AUPRC, AUROC and F1 scores of MSDLM and SSDLM

DISCUSSION

The reasons underlying differences in predictive power between MSDLM and SSDLM can be summarised in Figure 3 below.

MSDLM	SSLDM
Utilises multiple immune cell markers from the TIME	Uses a limited number of TIME variables
Through the incorporation of convolutional neural networks (CNN), it is sensitive to how each parameter relate to one another non-linearly	Does not use CNN in its workings

Figure 4: Table showing the differences in how MSDLM and SSDLM operate

As shown below, Figure 5 depicts the differences in accuracy between SSLDM and MSLDM in predicting CRC prognosis.

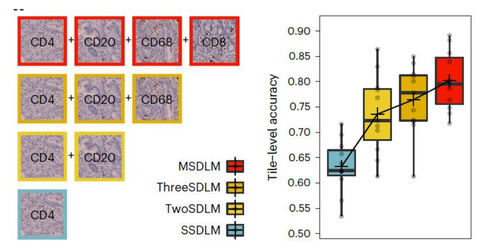

Figure 5: The boxplots compare the predictive accuracies of different models

Apart from showing the differences in predictive power between MSLDM and SSDLM, this study raises two interesting points as well.

- Firstly, MSLDM outperforms traditional methods like staging and resection status in predicting neoadjuvant therapy outcomes. This shows that it surpasses established CRC prognostic tools, in addition to models like SSLDM.

- Secondly, MSLDM predictions are aligned with established immuno-oncological principles. For example, the positive prognostic impact of CD8+ T cells and CD20+ B cells are reflected in the model's workings[2,4]. Likewise, the negative association of monocytes and macrophages with CRC progression are considered by the MSDLM[5] (see Figure 6 below).

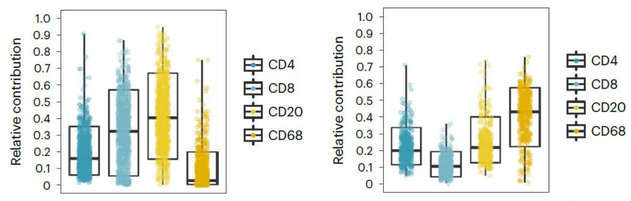

Figure 6: These two boxplots compare the relative contributions of each immune cell type to the overall prediction of 'no relapse' (left) and 'relapse' (right) made by the MSLDM.

Having explained the benefits of MSLDM, there are two limitations to using such a model.

- Firstly, the predictions of the model has to be confirmed by additional prospective multicenter studies[6]. Coupled with the lack of external validation of the results, there is a possibility that the MSLDM is overfitting the data.

- Secondly, the use of additional image processing techniques or multiplexed images could further refine predictions made by the MSLDM.

CONCLUSION

To conclude, MSLDM can be a valuable addition to current methods in assessing the TIME for the prognostication of CRC. The model incorporates IHC stains that are cheap and ubiquitously found. Given that we are in the digital age, MSLDM can be seamlessly added to current workflows that are used to predict the prognosis of CRC.

Specifically, a digital application hosting MSLDM could be envisaged for doctors to submit histological slides and obtain an AIS for their patients. In this way, MSLDM represents a significant step forward in CRC prognosis and therapy planning, aligning state-of-the-art AI with clinical needs.

REFERENCES

1. Galon, J. et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science 313, 1960–1964 (2006).
2. Bindea, G. et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity 39, 782–795 (2013).
3. Pages, F. et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet 391, 2128–2139 (2018).
4. Fridman, W. H. et al. B cells and cancer: to B or not to B? J. Exp. Med. 218, e20200851 (2021).
5. Aras, S. & Zaidi, M. R. TAMeless traitors: macrophages in cancer progression and metastasis. Br. J. Cancer 117, 1583–1591 (2017).
6. Kleppe, A. et al. Designing deep learning studies in cancer diagnostics. Nat. Rev. Cancer 21, 199–211 (2021).