BJS Open, 2024, zrae127
https://doi.org/10.1093/bjsopen/zrae127
Original Article

Classifying histopathological growth patterns for resected colorectal liver metastasis with a deep learning analysis

Diederik J. Höppener[1], Witali Aswolinskiy[2], Zhen Qian[1], David Tellez[2], Pieter M.H. Nierop[1], Martijn Starmans[3,4], Iris D. Nagtegaal[2], Michail Doukas[4], Johannes H.W. de Wilt[5], Dirk J. Grünhagen[1], Jeroen A.W.M. van der Laak[2], Peter Vermeulen[6], Francesco Ciompi[2] and Cornelis Verhoef[1,*]

[1]Department of Surgical Oncology and Gastrointestinal Surgery, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, The Netherlands
[2]Departments of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands
[3]Department of Radiology and Nuclear Medicine, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, The Netherlands
[4]Department of Pathology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, The Netherlands
[5]Departments of Surgery, Radboud University Medical Center, Nijmegen, The Netherlands
[6]Translational Cancer Research Unit6, GZA Hospital Sint-Augustinus & University of Antwerp, Antwerp, Belgium

*Correspondence to: Cornelis Verhoef, Department of Surgical Oncology and Gastrointestinal Surgery, Erasmus MC Cancer Institute, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA, Rotterdam, The Netherlands (e-mail: c.verhoef@erasmusmc.nl)

Abstract

Background: Histopathological growth patterns are one of the strongest prognostic factors in patients with resected colorectal liver metastases. Development of an efficient, objective and ideally automated histopathological growth pattern scoring method can substantially help the implementation of histopathological growth pattern assessment in daily practice and research. This study aimed to develop and validate a deep-learning algorithm, namely neural image compression, to distinguish desmoplastic from non-desmoplastic histopathological growth patterns of colorectal liver metastases based on digital haematoxylin and eosin-stained slides.

Methods: The algorithm was developed using digitalized whole-slide images obtained in a single-centre (Erasmus MC Cancer Institute, the Netherlands) cohort of patients who underwent first curative intent resection for colorectal liver metastases between January 2000 and February 2019. External validation was performed on whole-slide images of patients resected between October 2004 and December 2017 in another institution (Radboud University Medical Center, the Netherlands). The outcomes of interest were the automated classification of dichotomous hepatic growth patterns, distinguishing between desmoplastic hepatic growth pattern and non-desmoplatic growth pattern by a deep-learning model; secondary outcome was the correlation of these classifications with overall survival in the histopathology manual–assessed histopathological growth pattern and those assessed using neural image compression.

Results: Nine hundred and thirty-two patients, corresponding to 3.641 whole-slide images, were reviewed to develop the algorithm and 870 whole-slide images were used for external validation. Median follow-up for the development and the validation cohorts was 43 and 29 months respectively. The neural image compression approach achieved significant discriminatory power to classify 100% desmoplastic histopathological growth pattern with an area under the curve of 0.93 in the development cohort and 0.95 upon external validation. Both the histopathology manual–scored histopathological growth pattern and neural image compression-classified histopathological growth pattern achieved a similar multivariable hazard ratio for desmoplastic *versus* non-desmoplastic growth pattern in the development cohort (histopathology manual score: 0.63 *versus* neural image compression: 0.64) and in the validation cohort (histopathology manual score: 0.40 *versus* neural image compression: 0.48).

Conclusions: The neural image compression approach is suitable for pathology-based classification tasks of colorectal liver metastases.

Introduction

Colorectal cancer (CRC) is the third most common cancer and second cause of cancer mortality worldwide[1,2]. Approximately one-third of these patients are afflicted with metastatic disease, with the liver representing the most predominant metastatic site[3,4]. The presence of CRC distant metastases itself does not preclude potentially curative treatment[5–12]. Although half of all patients with colorectal liver metastases (CRLM) may now be eligible for local treatment[13], the results are still unsatisfactory, with only a quarter of patients achieving a long-term cure[14,15]. This has garnered a longstanding interest in the prediction of prognosis and treatment effect, with the ultimate goal of guiding patient selection and improving outcome[16].

In the search for new biomarkers, histological evaluation of liver metastases has emerged as a promising candidate. Light-microscopic evaluation of resected metastases allows for the determination of distinct histopathological growth patterns (HGPs)[17]. The most clinically relevant distinction between HGPs is desmoplastic *versus* non-desmoplastic HGP, according to the

Received: July 21, 2024. **Revised:** September 09, 2024. **Accepted:** September 12, 2024

Rotterdam 50% cut-off. A desmoplastic HGP is recognized with an approximate two-fold reduction in mortality and cancer recurrence[18,19]. Beside prognosis, several studies suggest that HGP is also predictive for treatment effect[2,20,21]. Although HGPs have been shown to describe the biological properties of the tumour relating to therapy response and prognosis, they are not routinely scored yet. Expertise is required because there are several caveats in scoring[22]. Moreover, as HGP scoring requires a pathologist to score the full interface between the liver and the tumour cell by cell, the task is time-consuming. The lack of an efficient, objective and ideally automated HGP classification method substantially limits the implementation of HGPs in daily practice and research.

Developments in the application of artificial intelligence, and specifically deep learning, to high-resolution digitalized whole-slide images (WSI) has led to a rapidly growing research field at the interface of medical and computer sciences[23,24]. Several deep-learning models are already approaching or even surpassing dedicated pathologists in histology-based marker determination tasks[25–32]. Moreover, deep-learning models can predict prognosis by learning directly from the histology slides, effectively creating novel AI-based computational biomarkers[32].

This study aims to assess whether a novel state-of-the-art deep-learning approach can be employed for the automated classification of the desmoplastic HGP in resected CRLM.

Methods

The current study adheres to the REporting recommendations for tumour MARKer prognostic studies (REMARK)[33]. Institutional ethical review was obtained from both the medical ethics committee of the Erasmus Medical Centre (MEC-2018-1743), which granted a waiver for (renewed) informed consent, and the Ethical Committee of the Radboud University Medical Centre (MEC 2015–1637).

Patient cohorts and sample preparation

The patient cohort used for development consisted of patients undergoing surgical treatment of CRLM at the Erasmus MC Cancer Institute, Rotterdam, the Netherlands, between January 2000 and February 2019. For external validation purposes patients treated in a similar time frame (October 2004 to December 2017) at a different centre, the Radboud University Medical Centre, Nijmegen, the Netherlands, were selected. All available haematoxylin and eosin–stained slides of all resection specimens were requested from the respective pathology departments and subsequently digitalized. Patients were included only if they underwent first curative intent CRLM resection (that is, resection specimens for recurrent disease were excluded, and patients had to have had curative intent local treatment of all known cancerous disease at time of first liver surgery). Follow-up was obtained through the electronic patient record as patients are scheduled for regular follow-up after resection.

Histopathological growth patterns determination

All slides were scanned at the pathology department of the Radboud UMC using a 3DHistech P1000 scanner at a spatial resolution of 0.25 µm/pixel. Digital assessment of all WSI was performed by a trained observer (DJH) to confirm slide content and assess WSI quality.

The HGP was previously determined in accordance with international consensus guidelines within the context of retrospective cohort studies[18,19,34]. The HGPs represent distinct histomorphological tumour–liver interface phenotypes of resected liver metastasis (*Fig. S1*), and can be grossly divided into two classes. The desmoplastic HGP is characterized by a broad band of desmoplastic stroma barring tumour–liver cell contact, and often displays a dense lymphocytic infiltrate peripherally to this desmoplastic stroma. The non-desmoplastic types most often exhibit cell-to-cell contact between tumour and liver cells, with the replacement of hepatocytes by tumour cells retaining the liver-cell plate architecture, that is the 'replacement' HGP. Although HGPs can appear in conjunction, we performed classification of the dichotomous presence of any non-desmoplastic HGP (*Fig. S1*) rather than relative abundance for the development and validation of the model, as this best distinguishes prognosis and is therefore clinically most relevant[17–19].

Neural image compression algorithm with multitask learning and attention pooling

For the classification of WSI we developed a neural image compression (NIC) algorithm with a supervised multitask-learning encoder framework (*Fig. 1*), building upon previous work[35]. The multitask NIC pipeline consists of two steps.

First, subregions of the entire gigapixel WSI are compressed into low-dimensional embedding vectors using a convolutional neural network (CNN), the encoder. These vectors are subsequently organized to form a compressed representation of the WSI, maintaining the spatial arrangement of the original WSI. The encoder model is responsible for gleaning high-level discriminatory information contained in the WSI for a variety of downstream tasks, while simultaneously suppressing image noise and spurious correlations[35,36]. To improve the extraction of high-level discriminatory factors that are transferable between a variety of tasks, we initially developed a supervised multitask learning architecture and trained an encoder on four histopathological tasks[35]. This approach demonstrated increased performance when compared to an unsupervised single-task framework. Independently, another author developed a similar multitask encoder, trained on 22 classification tasks and with validated performance increase compared to non-histopathological pretrained encoders[37]. In this work, we therefore use the new multitask encoder[37], which compresses a tile of size $256 \times 256 \times 3$ into a vector of size 2048. As input, we use here tiles at resolution $5\times$ (2 µm/px).

Second, a second CNN is trained on the entire compressed WSI as input to predict an outcome of interest, for example the HGP. For the CNN classifier, we adapted the attention-based architecture introduced in previous works (*Fig. 1*)[38,39]. In the context of neural networks, the term 'attention' refers to the capability of a network to learn to focus, that is to attend to specific regions of the input image. Using attention allows neural networks to make efficient use of training data as well as provide visually interpretable outputs *via* so-called 'attention maps'. In one of the authors' previous works, they demonstrated the performance advantage of attention on a task for lung cancer subtyping compared to a convolutional architecture without attention.[40] After a single layer, an attention block is applied, resulting in a score for each compressed tile. It follows a matrix multiplication of the attention map with the output of the first layer ('attention pooling'), resulting in a single vector which is then fed to the final classification layer. In the attention block, a dropout rate of 0.25 was used. The attention maps were used to visualize what is relevant for the network's

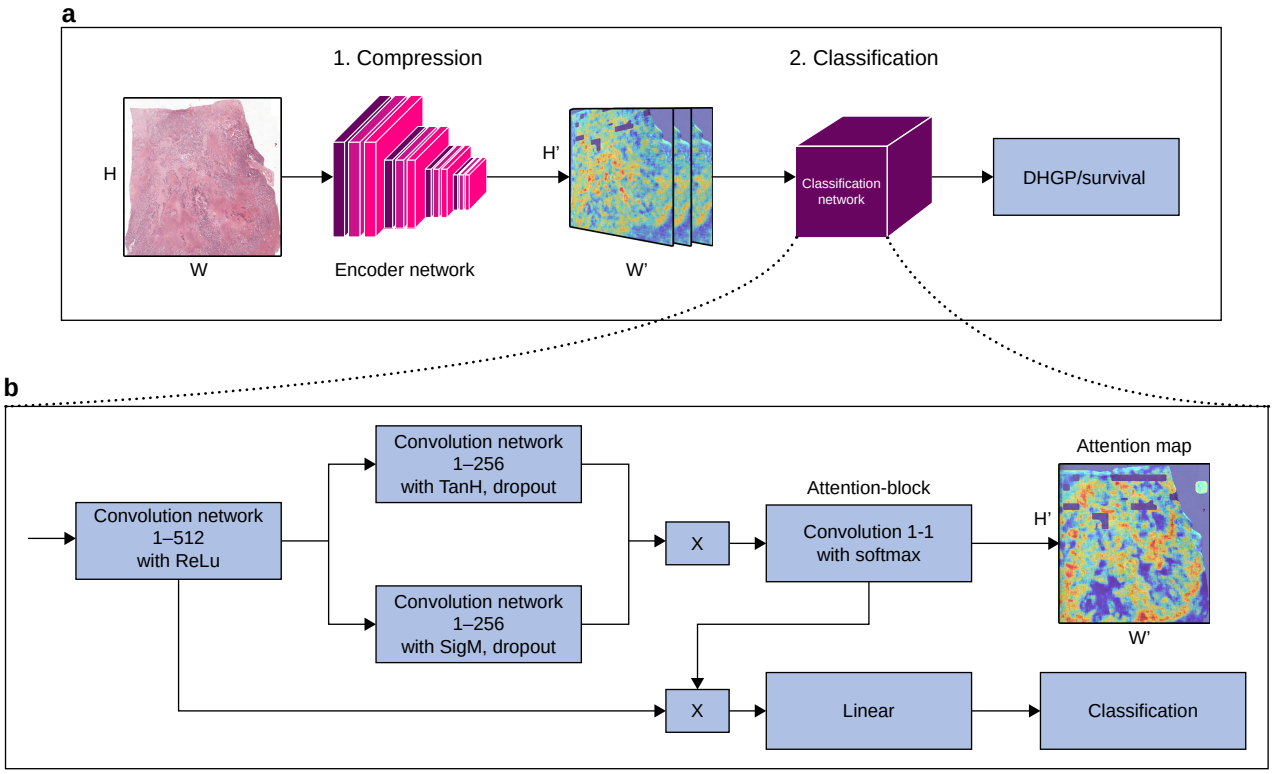

Fig. 1 Neural image compression pipeline (A) with a supervised multitask learning encoder framework and convolutional neural networks classifier (B) Neural image compression with attention pipeline. First the slide is compressed, then classified. The classification architecture consists of four 1 × 1 convolutional layers and a final linear layer starting with a 1 × 1 convolution reducing the input channels from 2048 to 512 (conv1-512). H and W stand for height and width of the image respectively. H' and W' are the height and width of the compressed images respectively with H' << than H and W' << than W.

prediction and thus contributes to the interpretability of the model.

Experimental setup

Following the compression of the slides using the encoder model, we trained the CNN with cross-entropy loss minimization to predict the image label of interest (that is the HGP). Development was performed using a five-fold cross-validation (three folds for training, one for validation, one for testing). The training was done with balanced sampling, batch size of one, and early stopping with 25 epoch patience using the validation ROC-AUC (receiver operating characteristic area under the curve) as stopping criteria. External validation was performed on previously unseen slides of the Nijmegen cohort by averaging the predictions of the five models. A patient-level score was subsequently obtained by averaging the scores of all slides belonging to a single patient.

Outcomes of interest

The primary outcome of interest was the classification of dichotomous hepatic growth patterns, distinguishing between desmoplastic hepatic growth pattern and non-desmoplastic growth pattern by a deep-learning model. The secondary outcome was the correlation of these classifications with overall survival, irrespective of the underlying cause of death.

Statistical analysis

All statistical analyses were performed using the R project for statistical computing (https://www.r-project.org/). A complete case analysis was performed because of a low percentage of missing data (<5%) and large sample size. Categorical variables are reported as absolute numbers with corresponding percentages and non-parametric ordinal or numerical variables as medians with corresponding interquartile ranges, and were compared using the chi^2 or Kruskall Wallis tests respectively. Assessment of HGP classifier performance was done through ROC curve analysis with the slide-level ensemble score and observer-based HGP as the predictor and label respectively and the AUC with corresponding 95% c.i. as the performance metric. Given the class imbalance (roughly 80% of patients have a non-desmoplastic HGP), the optimality criteria were modified according to the prevalence of desmoplastic samples in the development cohort as proposed by others[41]. This threshold was subsequently applied in the external validation cohort to the patient-level ensemble scores, using the balanced accuracy as a performance metric[42]. Kaplan–Meier and Cox proportional regression survival analyses were performed to assess the prognostic value of the histopathology observer–based HGP and NIC-classified HGP. Multivariable models were corrected for age, sex, pT-stage, pN-stage, right-sided colorectal cancer, disease-free interval, number of liver metastasis, diameter of largest liver metastasis, preoperative carcinoembryonic antigen level and extrahepatic disease.

Results

Of 1254 patients treated at the development institution, 965 met the inclusion criteria and 932 were eligible for analysis. On the other hand, of 305 patients treated at the validation centre, 294 were eligible for analysis. A timeline of patients' enrolment over the years at the development and the validation centres is presented respectively in *Fig. S2* and *Fig. S3*.

Patient and treatment characteristics of the original and validation cohort are provided in *Table S1*. The development

cohort comprised a total of 3.641 WSI from 932 patients (median follow-up time 43 months) undergoing first curative intent surgical treatment for CRLM. For external validation, a total of 870 WSI from 294 patients were available (median follow-up time 29 months). Fifty-five per cent of the patients in the development cohort received neo-adjuvant chemotherapy and 72.1% in the validation cohort. pT-stage did not differ significantly between the two cohorts ($P = 0.94$); however, a higher proportion of pN0-stage primary tumour was observed in the development cohort ($P = 0.02$). No statistically significant difference in HGP proportions was observed between the two cohorts.

Automated HGP classification

Using a five-fold cross-validation the NIC classifier achieved an AUC of 0.93 (95% c.i. 0.93 to 0.94) in the original cohort to classify the slide-level HGP (*Fig. 2*). Applying the optimal threshold for the ensemble score (0.69) based on the ROC curve/Youden's J statistic (*Fig. 2*) resulted in a patient-level sensitivity of 82%, a specificity of 93% and a balanced accuracy of 88% (*Fig. 2*, *Table 1*). Upon external validation in the 870 previously unseen WSI of the validation cohort the NIC classifier achieved a similar AUC of 0.95 (95% c.i. 0.93 to 0.96) to classify the slide-level HGP (*Fig. 2*). Application of the optimal threshold from the development cohort achieved a patient-level sensitivity of 87%, a specificity of 91% and a balanced accuracy of 89% when compared to the observer-based HGP (*Fig. 2*).

Survivals

Table 2 reports the survival estimates and regression results for the observer-based and the NIC-classified HGP in both the development and external validation cohort, and *Fig. 3* and *Fig. 4* display the respective overall survival (OS) curves with stratification for chemo-naïve and pretreated. Overall, the NIC-classified HGP exhibited similar prognostic impact on OS as the histopathology observer–based HGP, also upon external validation. For example, the adjusted hazards ratio (95% c.i.) for desmoplastic *versus* non-desmoplastic patients based on the NIC-classified HGP was 0.64 (0.51 to 0.79) in the original cohort and 0.48 (0.28 to 0.83) upon external validation, compared to 0.63 (0.50 to 0.79) and 0.40 (0.22 to 0.75) respectively for the observer-based HGP (*Table 1*). *Figure S4* shows examples of attention maps of four different histological slides of liver tissue samples paired with their corresponding attention maps. The attention maps are generated using predictive models to visualize the areas of importance for classifying the hepatic growth pattern, thus providing insights into the model's decision-making process. An initial analysis of the attention maps shows that the model is indeed mainly focusing on the tumour–stroma border to determine the HGP.

Discussion

In this study the authors developed and validated a deep-learning–based pipeline with compression and attention to classify HGP on a large data set of digitalized WSI of resected CRLM without manual input from a clinician. The developed NIC classifier performed similarly across the development and previously unseen external validation cohort, achieving high levels of classifier performance and demonstrating generalizability with a balanced accuracy of ≥ 88%. In addition, the NIC-classified HGP demonstrated similar prognostic impact in terms of OS when compared to observer-based pathologist determination with the added benefit of faster output.

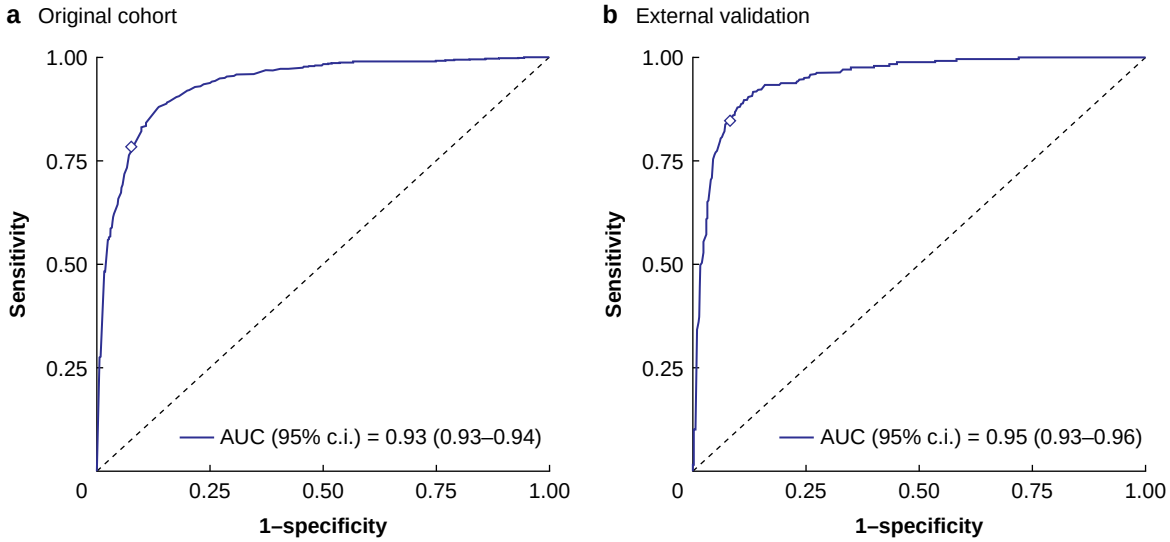

Fig. 2 ROC curves of the automated histopathological growth pattern (HGP) classification in the original (a) and in the external validation cohort (b)

Table 1 NIC HGP classification performance in the development and validation cohorts

	TP	TN	FP	FN	Sens.	Spec.	PPV	NPV	Bal. Acc.
Development—patient level ($n = 932$)	180	662	51	39	82%	93%	78%	94%	88%
Validation—patient level ($n = 294$)*	52	213	21	8	87%	91%	71%	96%	89%

*According to the predefined classification cut-off determined in the development cohort. Bal. Acc., balanced accuracy; FN, false negative; FP, false positive; NIC, neural image compression; NPV, negative predictive value; PPV, positive predictive value; Sens., sensitivity; Spec., specificity; TN, true negative; TP, true positive.

Table 2 Survival analyses on the ground-truth and NIC-classified HGP

Desmoplastic *versus* non-desmoplastic	Non-desmoplastic 5-year OS (95% c.i.)	Desmoplastic 5-year OS (95% c.i.)	Desmoplastic *versus* non-desmoplastic	
			Univariable HR (95% c.i.)	Multivariable HR (95% c.i.)*
Development cohort ($n = 932$)				
Ground-truth HGP	40% (36,44)	63% (57,70)	0.57 (0.47,0.70)	0.63 (0.50,0.79)
NIC-classified HGP	40% (37,44)	60% (54,67)	0.61 (0.50,0.75)	0.64 (0.51,0.79)
Validation cohort ($n = 294$)				
Ground-truth HGP	64% (58,71)	80% (70,91)	0.51 (0.30,0.86)	0.40 (0.22,0.75)
NIC-classified HGP	66% (60,72)	73% (63,84)	0.64 (0.41,1.02)	0.48 (0.28,0.83)

*Corrected for age, sex, primary tumour location, pT-stage, pN-stage, disease-free interval, number of CRLM, diameter of largest CRLM, preoperative CEA, and extrahepatic disease. CEA, carcinoembryonic antigen; CRLM, colorectal liver metastasis; HGP, histopathological growth pattern; NIC, neural image compression; OS, overall survival.

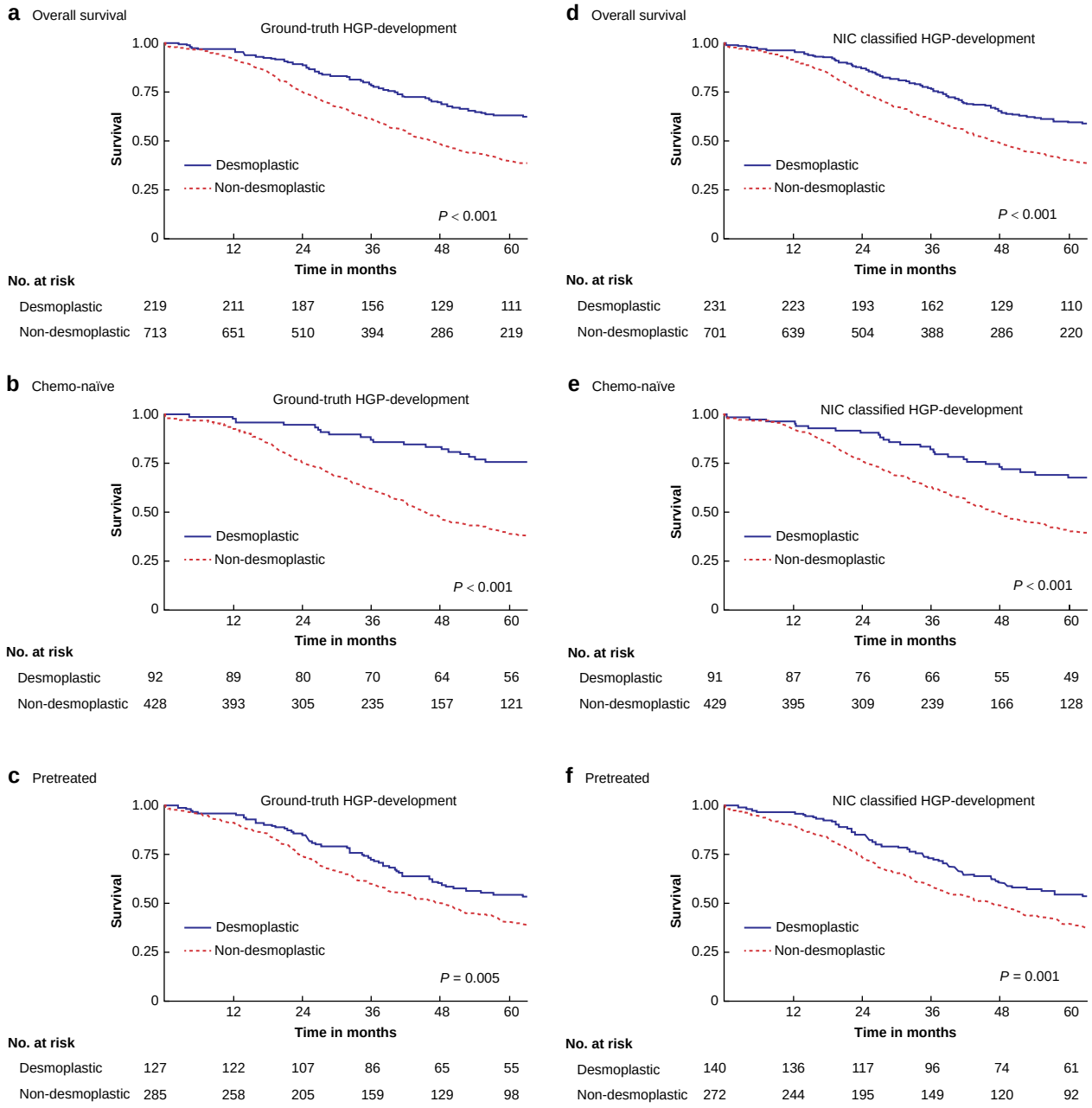

Fig. 3 Overall survival (OS) curves for the observer-based (a–c) and neural image compression (NIC) (d–f) classified histopathological growth pattern (HGP) in the development cohort and stratified for pretreatment (c,f) and chemo-naïve (b,e) patients

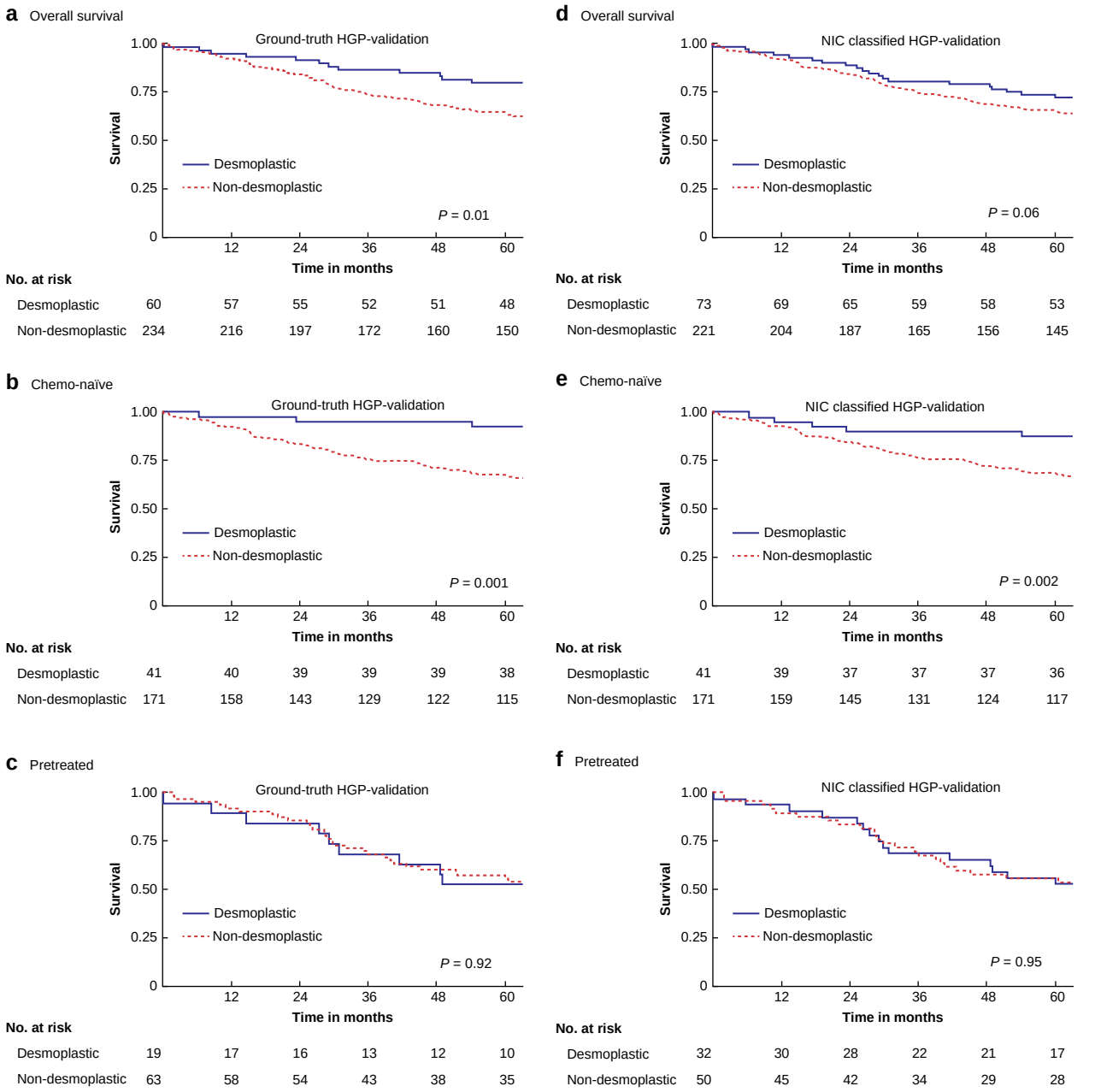

Fig. 4 Overall survival (OS) curves for the observer-based (a–c) and neural image compression (NIC) (d–f) classified histopathological growth pattern (HGP) in the validation cohort and stratified for pretreatment (c,f) and chemo-naïve (b,e) patients

Literature shows that HGP is an independent prognostic factor for survival and there are studies suggesting HGP as a predictive factor for therapeutic effectiveness, making it a clinically highly relevant biomarker[2,15,21,43]. It is of the utmost importance that such a biomarker is objective and reproducible, independent of the scoring physician. It is known that scoring of HGPs has several caveats, so expertise is necessary. The results of this study demonstrate high levels of HGP classification performance in both the development and validation cohorts (AUC $\geq$ 0.93), suggesting the development of an objective and reproducible clinically relevant scoring method that can automatically be determined. This will substantially help the implementation of HGPs in daily practice and research.

The attention maps illustrate that the NIC model concentrates on various regions of the slide. By analysing these maps, we can understand which features or regions of the slide are most influential in the model's assessment of the HGP. This can be particularly useful for identifying potential areas for improvement in the model or for validating that the model is focusing on clinically relevant regions or even potentially to discover new histopathological biomarkers.

Although promising, these results also suggest the limits of the NIC classification pipeline with incorporation of even larger data sets and different immunohistochemical staining. This study includes only two tertiary university hospitals with very high performance both in the development and validation cohorts, which could be a sign of model overfitting to these specific data sets. Additional development and validation cohorts of different centres in multiple countries could improve this model even further and alleviate this problem. A recent study has suggested

that a more granular, non-dichotomous approach could potentially offer enhanced prognostic value and stratify patient survival even further than the dichotomous classification. Lastly, this study did not explore how the deep-learning model could be seamlessly integrated into existing clinical workflows. Addressing the practical challenges of implementation in daily clinical practice is essential for ensuring the model's effective use in real-world settings. Further research is necessary to validate these results[44].

In conclusion, these experimental results show that automated NIC-based models are promising to objectively classify HGP following surgical treatment of CRLM.

Funding

This study was partly funded by a grant from Stichting Coolsingel, Rotterdam, the Netherlands, and by the European Union through the Horizon 2020 framework under grant agreement No. 825292 (ExaMode, htttp://www.examode.eu/).

Acknowledgements

D.J.H. and W.A. are contributed equally. F.C. and C.V. are jointly supervised this work.

Disclosure

David Tellez is now affiliated with Aiosyn BV, the Netherlands. Jeroen van der Laak was a member of the advisory boards of Philips, the Netherlands and ContextVision, Sweden, and received research funding from Philips, the Netherlands, ContextVision, Sweden, and Sectra, Sweden in the last 5 years. He is chief scientific officer and a shareholder of Aiosyn BV, the Netherlands. Francesco Ciompi was Chair of the Scientific and Medical Advisory Board of TRIBVN Healthcare, France, and received advisory board fees from TRIBVN Healthcare, France in the last 5 years. He is a shareholder of Aiosyn BV, the Netherlands.

All other authors have no conflict of interests to declare.

Supplementary material

Supplementary material is available at *BJS Open* online.

Data availability

All clinical data and corresponding digital WSI used in this study are not publicly available but can be requested from and may be provided at the discretion of the corresponding authors of each respective centre and under the provision of appropriate data and material transfer agreements.

All developed deep-learning models are published online and are freely accessible at https://grand-challenge.org/algorithms/colorectal-liver-metastases-survival-prediction upon request. We also provide the code to create the correct input (slide as.tif or.svs file with correct background mask as.tif file) for the algorithm: https://grand-challenge.org/algorithms/tissue-segmentation-and-packing. The source code for NIC is available at https://github.com/DIAGNijmegen/pathology-whole-slide-learning.

Author contributions

Diederik Höppener (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Validation, Visualization, Writing—original draft, Writing—review & editing), Witali Aswolinskiy (Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Visualization, Writing—original draft, Writing—review & editing), Zhen Qian (Visualization, Writing—review & editing), David Tellez (Writing—review & editing), Pieter Nierop (Writing—review & editing), Martijn Starmans (Writing—review & editing), Iris Nagtegaal (Writing—review & editing), Michail Doukas (Writing—review & editing), Johannes De Wilt (Writing—review & editing), Dirk Grünhagen (Writing—review & editing), Jeroen van der Laak (Writing—review & editing), Peter Vermeulen (Writing—review & editing), Francesco Ciompi (Conceptualization, Funding acquisition, Project administration, Supervision, Writing—review & editing), and Cornelis Verhoef (Conceptualization, Funding acquisition, Project administration, Supervision, Writing—review & editing)

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;**68**:394–424
2. Buisman FE, van der Stok EP, Galjart B, Vermeulen PB, Balachandran VP, Coebergh van den Braak RRJ *et al.* Histopathological growth patterns as biomarker for adjuvant systemic chemotherapy in patients with resected colorectal liver metastases. *Clin Exp Metastasis* 2020;**37**:593–605
3. Manfredi S, Lepage C, Hatem C, Coatmeur O, Faivre J, Bouvier AM. Epidemiology and management of liver metastases from colorectal cancer. *Ann Surg* 2006;**244**:254–259
4. Engstrand J, Nilsson H, Strömberg C, Jonas E, Freedman J. Colorectal cancer liver metastases—a population-based study on incidence, management and survival. *BMC Cancer* 2018;**18**:78
5. Gootjes EC, Buffart TE, Tol MP, Burger J, Grunhagen DJ, van der Stok EP *et al.* The ORCHESTRA trial: a phase III trial of adding tumor debulking to systemic therapy *versus* systemic therapy alone in multi-organ metastatic colorectal cancer (mCRC). *J Clin Oncol* 2016;**34**:TPS788–TPS788
6. Moris D, Ronnekleiv-Kelly S, Rahnemai-Azar AA, Felekouras E, Dillhoff M, Schmidt C *et al.* Parenchymal-sparing *versus* anatomic liver resection for colorectal liver metastases: a systematic review. *J Gastrointest Surg* 2017;**21**:1076–1085
7. Capussotti L, Muratore A, Baracchi F, Lelong B, Ferrero A, Regge D *et al.* Portal vein ligation as an efficient method of increasing the future liver remnant volume in the surgical treatment of colorectal metastases. *Arch Surg* 2008;**143**:978–982; discussion 982
8. Sandström P, Røsok BI, Sparrelid E, Larsen PN, Larsson AL, Lindell G *et al.* ALPPS improves resectability compared with conventional two-stage hepatectomy in patients with advanced colorectal liver metastasis: results from a Scandinavian multicenter randomized controlled trial (LIGRO trial). *Ann Surg* 2018;**267**:833–840
9. Bismuth H, Adam R, Lévi F, Farabos C, Waechter F, Castaing D *et al.* Resection of nonresectable liver metastases from colorectal cancer after neoadjuvant chemotherapy. *Ann Surg* 1996;**224**:509–520; discussion 520–502
10. Huiskens J, van Gulik TM, van Lienden KP, Engelbrecht MR, Meijer GA, van Grieken NC *et al.* Treatment strategies in colorectal cancer patients with initially unresectable liver-only

metastases, a study protocol of the randomised phase 3 CAIRO5 study of the Dutch Colorectal Cancer Group (DCCG). *BMC Cancer* 2015;**15**:365
11. Stang A, Fischbach R, Teichmann W, Bokemeyer C, Braumann D. A systematic review on the clinical benefit and role of radiofrequency ablation as treatment of colorectal liver metastases. *Eur J Cancer* 2009;**45**:1748–1756
12. Mahadevan A, Blanck O, Lanciano R, Peddada A, Sundararaman S, D'Ambrosio D *et al.* Stereotactic body radiotherapy (SBRT) for liver metastasis—clinical outcomes from the international multi-institutional RSSearch® patient registry. *Radiat Oncol* 2018;**13**:26
13. Meyer Y, Olthof PB, Grünhagen DJ, de Hingh I, de Wilt JHW, Verhoef C *et al.* Treatment of metachronous colorectal cancer metastases in the Netherlands: a population-based study. *Eur J Surg Oncol* 2022;**48**:1104–1109
14. Tomlinson JS, Jarnagin WR, DeMatteo RP, Fong Y, Kornprat P, Gonen M *et al.* Actual 10-year survival after resection of colorectal liver metastases defines cure. *J Clin Oncol* 2007;**25**:4575–4580
15. Buisman FE, Giardiello D, Kemeny NE, Steyerberg EW, Höppener DJ, Galjart B *et al.* Predicting 10-year survival after resection of colorectal liver metastases; an international study including biomarkers and perioperative treatment. *Eur J Cancer* 2022;**168**:25–33
16. Kanas GP, Taylor A, Primrose JN, Langeberg WJ, Kelsh MA, Mowat FS *et al.* Survival after liver resection in metastatic colorectal cancer: review and meta-analysis of prognostic factors. *Clin Epidemiol* 2012;**4**:283–301
17. Latacz E, Höppener D, Bohlok A, Leduc S, Tabariès S, Fernández Moro C *et al.* Histopathological growth patterns of liver metastasis: updated consensus guidelines for pattern scoring, perspectives and recent mechanistic insights. *Br J Cancer* 2022;**127**:988–1013
18. Galjart B, Nierop PMH, van der Stok EP, van den Braak RRJC, Höppener DJ, Daelemans S *et al.* Angiogenic desmoplastic histopathological growth pattern as a prognostic marker of good outcome in patients with colorectal liver metastases. *Angiogenesis* 2019;**22**:355–368
19. Höppener DJ, Galjart B, Nierop PMH, Buisman FE, van der Stok EP, Coebergh van den Braak RRJ *et al.* Histopathological growth patterns and survival after resection of colorectal liver metastasis: an external validation study. *JNCI Cancer Spectr* 2021;**5**:pkab026
20. Nierop PMH, Galjart B, Höppener DJ, van der Stok EP, Coebergh van den Braak RRJ, Vermeulen PB *et al.* Salvage treatment for recurrences after first resection of colorectal liver metastases: the impact of histopathological growth patterns. *Clin Exp Metastasis* 2019;**36**:109–118
21. Frentzas S, Simoneau E, Bridgeman VL, Vermeulen PB, Foo S, Kostaras E *et al.* Vessel co-option mediates resistance to anti-angiogenic therapy in liver metastases. *Nat Med* 2016;**22**:1294–1302
22. van Dam PJ, van der Stok EP, Teuwen LA, Van den Eynden GG, Illemann M, Frentzas S *et al.* International consensus guidelines for scoring the histopathological growth patterns of liver metastasis. *Br J Cancer* 2017;**117**:1427–1441
23. van der Laak J, Litjens G, Ciompi F. Deep learning in histopathology: the path to the clinic. *Nat Med* 2021;**27**:775–784
24. Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br J Cancer* 2021;**124**:686–696
25. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B *et al.* Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol* 2020;**21**:233–241
26. Nagpal K, Foote D, Tan F, Liu Y, Chen PC, Steiner DF *et al.* Development and validation of a deep learning algorithm for Gleason grading of prostate cancer from biopsy specimens. *JAMA Oncol* 2020;**6**:1372–1380
27. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D *et al.* Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018;**24**:1559–1567
28. Ehteshami Bejnordi B, Mullooly M, Pfeiffer RM, Fan S, Vacek PM, Weaver DL *et al.* Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. *Mod Pathol* 2018;**31**:1502–1512
29. Mercan E, Mehta S, Bartlett J, Shapiro LG, Weaver DL, Elmore JG. Assessment of machine learning of breast pathology structures for automated differentiation of breast cancer and high-risk proliferative lesions. *JAMA Netw Open* 2019;**2**:e198777
30. Wu M, Yan C, Liu H, Liu Q. Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks. *Biosci Rep* 2018;**38**:BSR20180289
31. Hekler A, Utikal JS, Enk AH, Solass W, Schmitt M, Klode J *et al.* Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *Eur J Cancer* 2019;**118**:91–96
32. Skrede O-J, De Raedt S, Kleppe A, Hveem TS, Liestøl K, Maddison J *et al.* Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet* 2020;**395**:350–360
33. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 2005;**93**:387–391
34. Höppener DJ, Nierop PMH, Herpel E, Rahbari NN, Doukas M, Vermeulen PB *et al.* Histopathological growth patterns of colorectal liver metastasis exhibit little heterogeneity and can be determined with a high diagnostic accuracy. *Clin Exp Metastasis* 2019;**36**:311–319
35. Tellez D, Höppener D, Verhoef C, Grünhagen D, Nierop P, Drozdzal M *et al.* Extending unsupervised neural image compression with supervised multitask learning. *Proc Machine Learning Res* 2020;**121**:770–783
36. Tellez D, Litjens G, Bándi P, Bulten W, Bokhorst JM, Ciompi F *et al.*Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Med Image Anal* 2019;**58**:101544
37. Mormont R, Geurts P, Marée R. Multi-task pre-training of deep neural networks for digital pathology. *IEEE J Biomed Health Inform* 2020;**25**:412–421
38. Maximilian I, Jakub T, Max W. Attention-based Deep Multiple Instance Learning. In: *Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden*. PMLR, 2018
39. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng* 2021;**5**:555–570
40. Witali A, David T, Gabriel R, Lieke van der W, Monika L-S, Jeroen van der L *et al.* Neural image compression for non-small cell lung cancer subtype classification in H&E stained whole-slide images. In: *Proc SPIE, International Society for Optics and Photonics, 2021, San Diego Convention Center/San Diego, California, United States*. 2021

41. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol* 2006; **163**:670–675
42. Kleppe A, Skrede OJ, De Raedt S, Liestøl K, Kerr DJ, Danielsen HE. Designing deep learning studies in cancer diagnostics. *Nat Rev Cancer* 2021;**21**:199–211
43. Zaharia C, Veen T, Lea D, Kanani A, Alexeeva M, Søreide K. Histopathological growth pattern in colorectal liver metastasis and the tumor immune microenvironment. *Cancers (Basel)* 2022;**15**:181
44. Fernández Moro C, Geyer N, Harrizi S, Hamidi Y, Söderqvist S, Kuznyecov D *et al.* An idiosyncratic zonated stroma encapsulates desmoplastic liver metastases and originates from injured liver. *Nat Commun* 2023;**14**:5024