ARTICLE   OPEN

Deep learning-based interpretable prediction of recurrence of diffuse large B-cell lymphoma

Hussein Naji[1,2,5], Paul Hahn[1,2,5], Juan I. Pisula[1,2], Stefano Ugliano[1,2], Adrian Simon[3], Reinhard Büttner[3] and Katarzyna Bozek[1,2,4 ✉]

BACKGROUND: The heterogeneous and aggressive nature of diffuse large B-cell lymphoma (DLBCL) presents significant treatment challenges as up to 50% of patients experience recurrence of disease after chemotherapy. Upfront detection of recurring patients could offer alternative treatments. Deep learning has shown potential in predicting recurrence of various cancer types but suffers from lack of interpretability. Particularly in prediction of recurrence, an understanding of the model's decision could eventually result in novel treatments.
METHODS: We developed a deep learning-based pipeline to predict recurrence of DLBCL based on histological images of a publicly available cohort. We utilized attention-based classification to highlight areas within the images that were of high relevance for the model's classification. Subsequently, we segmented the nuclei within these areas, calculated morphological features, and statistically analyzed them to find differences between recurred and non-recurred patients.
RESULTS: We achieved an f1 score of 0.88 indicating that our model can distinguish non-recurred from recurred patients. Additionally, we found that features that are the most predictive of recurrence include large and irregularly shaped tumor cell nuclei.
DISCUSSION: Our work underlines the value of histological images in predicting treatment outcomes and enhances our understanding of complex biological processes in aggressive, heterogeneous cancers like DLBCL.

BJC Reports; https://doi.org/10.1038/s44276-025-00147-0

INTRODUCTION

Diffuse large B-cell lymphoma (DLBCL) represents the most prevalent form of non-Hodgkin's lymphoma (NHL) in adults, accounting for 30–40% of all NHL cases. It originates from B-lymphocytes and is characterized by its remarkable diversity in terms of morphology, biology, and clinical outcomes. Despite treatment strategies that include high-dose chemotherapy coupled with autologous stem cell transplantation, achieving long-term remission remains challenging. Approximately half of all patients experience a recurrence, following initial therapy. Due to its aggressiveness, recurred patients have a critically low survival probability [1]. Thus, there is a pressing need for a reliable upfront detection of those patients that will not benefit from standard frontline regimens.

In the recent past, deep learning has made remarkable progress in predicting recurrence of various cancer types (including bladder [2], breast [3–6], colorectal [7, 8], hepatocellular carcinoma [9, 10], lung [11, 12], and prostate cancer [13, 14]) based on histological images.

Previous studies have used deep learning models (mostly variations of ResNet architectures [15]) to predict recurrence based on images as well as clinical features such as tumor grade, stage, age, etc. [2–5, 9, 11]. While offering good performance, these models did not provide insights into the determinants of the patient outcomes.

Explainability of recurrence prediction models is of tremendous medical importance. Identifying which image regions are predictive of specific clinical outcomes could potentially lead to discovery of novel visual biomarkers as well as to a better understanding of the underlying resistance mechanisms. Several recent studies have included interpretability analyses using multiple instance learning (MIL) [12, 13] or gradCAM heatmaps [6] to highlight image patches that were important for the model's decision. Although these methods allowed for qualitative insights into the relevant image regions, they did not perform an in-depth quantitative analysis of those regions.

There is relatively little research on the image-based prediction of DLBCL patient outcomes. Up to 2019, the primary prognostic features were clinical features, gene expression markers, and genetic abnormalities [16]. Wang et al. [17], Fan et al. [18], and Xing et al. [19] used traditional machine learning algorithms (such as random forest support vector machine, regression, gaussian mixture model clustering) to predict recurrence in patients with DLBCL using clinical features. Shankar et al. [20] introduced LymphoML, an interpretable deep learning method that identifies morphological features that are predictive of lymphoma subtypes.

[1]Institute for Biomedical Informatics, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [2]Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [3]Institute of Pathology, University Hospital Cologne, Cologne, Germany. [4]Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany. [5]These authors contributed equally: Hussein Naji, Paul Hahn. ✉email: k.bozek@uni-koeln.de
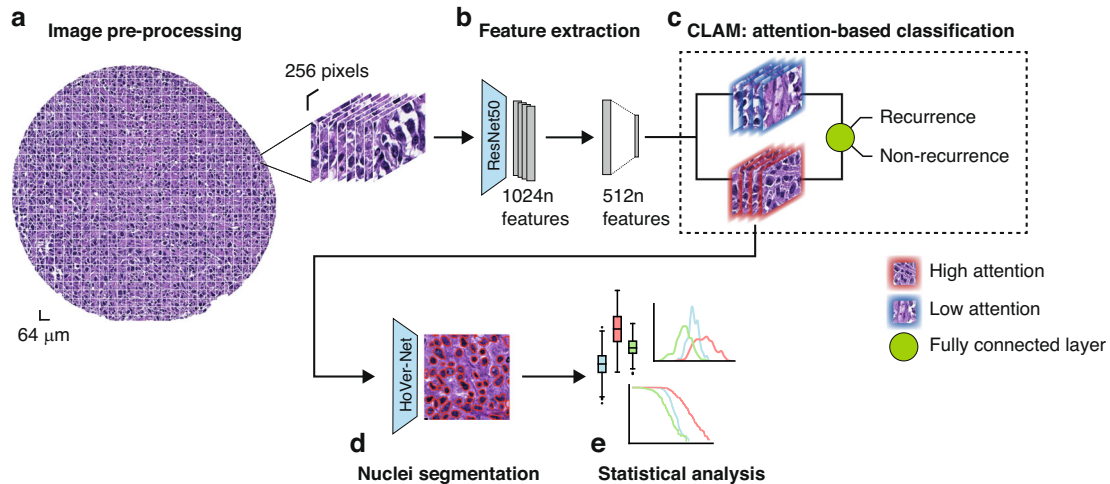
Fig. 1 Deep learning pipeline. a H&E-stained core images were segmented and cut into non-overlapping patches of size 256 ×256 pixels. **b** Features were extracted from all patches of a core image using a ResNet50 encoder. The final feature vectors had a size of 1024. **c** CLAM model to classify the cores into 'recurrence' or 'non-recurrence'. The model assigns attention scores to each patch of a core image, with blue frames indicating low attention patches and red frames indicating high attention patches. **d** Segmentation of the nuclei of the top 20 patches (based on the attention score) of each class using a pre-trained modified HoVer-Net architecture [24]. **e** We calculate morphological features from the segmented cell nuclei and statistically analyze them to find significant differences between recurred and non-recurred patients.

Their model points to concrete single-cell-based morphological features that are distinctive for different lymphoma subtypes, such as minor axis length and nuclei area. However, how these features could be leveraged to predict recurrence in DLBCL patients is an open question.

In this work, we address the challenge of prediction of recurrence after first-line treatment in DLBCL and interpretation of the prediction model. We developed a deep learning-based prediction pipeline based on tissue microarrays (TMAs) stained with hematoxylin and eosin (H&E). We extracted the model's high attention areas of the core images, applied nuclei segmentation on these areas, and determined single-cell morphological features which are distinct for recurred and non-recurred patients. While the causes of DLBCL relapse are still unknown, here, we point to cell morphology features which might be determinant for patient response to treatment.

METHODS
Patient cohort and dataset
Our dataset is based on a publicly available dataset of the Stanford Cancer Institute [21]. Of the 42 TMAs in the original dataset, prepared with six different stains, we are using the seven TMAs stained with hematoxylin and eosin (H&E). After discarding cores with low quality and incomplete patient information, our dataset contained 167 patients with 306 cores (1–2 cores per patient) from the seven H&E-stained TMAs. The TMA slides were scanned at 40× magnification (0.25 μm per pixel) on an Aperio AT2 scanner (Leica Biosystems, Nussloch, Germany). Each TMA has a thickness of 4 μm and includes a formalin-fixed, paraffin-embedded (FFPE) section of tumors arranged in a grid. Each core within the TMAs has a diameter of 0.6 mm [21]. The clinical data shows that 47 patients of the cohort did suffer from a recurrence, whereas 120 patients stayed recurrence-free.

Pipeline
We developed a pipeline (Fig. 1) that starts with a preprocessing step (Fig. 1a) in which the cores are cut into non-overlapping patches of size 256 × 256 pixels. For each patch of a core, a low-dimensional feature representation is determined using a ResNet50 model pretrained on ImageNet. This results in a 1024-dimensional feature vector which is then further compressed to a 512-dimensional vector (Fig. 1b) through linear transformation of a fully connected layer. To predict recurrence from the images, we use CLAM (clustering-constrained attention multiple instance

learning) [22]. CLAM leverages attention-based learning to localize image regions that are of significant diagnostic importance. That is, the model uses an attention-based pooling function to aggregate patch-level features into sample-level representations to classify a sample into 'recurrence' or 'non-recurrence'. The classification decision depends on the attention score that is assigned to each patch and indicates its importance to the sample-level representation for the respective class (Fig. 1c) [22].

For training the classification model, we used 60% of patients for training, 20% for validation and 20% for final testing. We further applied a cross validation with 4 iterations to further validate our model's performance. Hence, we express the performance as the mean of all iterations. Out of the correctly classified cores (true positives), we utilized the top 20 patches with the highest attention scores for further analysis (for both groups, respectively). We segmented the nuclei within each patch using a modified version of HoVer-Net [23], that has specifically been trained on H&E-stained DLBCL whole slide images (Fig. 1d) [24]. We excluded cut-off nuclei at the patch margins, quantified morphological features of the remaining segmented nuclei, and applied statistical analysis to identify differences between the features of recurred and non-recurred patients (Fig. 1e). We quantified the following shape- and size-based nuclei features:

1. area: area of the nucleus.
2. major axis length: length of the major axis of the ellipse that has the same second central moment as the nucleus.
3. minor axis length: length of the minor axis of the ellipse that has the same second central moment as the nucleus.
4. roughness: variance in the length of a vector that is centered at the centroid of a nucleus as it rotates along with each boundary point. The higher the value, the higher the degree of nucleus shape irregularity [25].

RESULTS
Relapse prediction
Our model achieves a classification accuracy of 0.79, a precision of 0.79, a recall of 0.98 and an f1-score of 0.88. A confusion matrix is shown in Fig. 2. Despite the class imbalance of the dataset with approximately 28% of patients that experience recurrence, the errors are equally distributed among the two patient groups.

Feature encoder and model hyperparameters were chosen based on performance. We have tested several models for comparison, shown in Table 1. Each model was trained for a maximum of 200 epochs with early stopping enabled after 50

epochs. Instance clustering was performed using eight patches with the CLAM method. Using a higher number of patches was not possible due to the patch extraction of CLAM, which extracted eight patches for some cores available. Class weighting was enabled to balance the dataset. As a patient can have two cores in the dataset, DLBCL is predicted to recur for a patient if DLBCL is predicted to recur for at least one core.

Statistical analysis

We identified nuclei area, major and minor axis length, and roughness as nuclei features that differ between the patients that have recurred and those that have not (Fig. 3). In the high-attention image areas for recurred patients, the nuclei area, major and minor axis length, and roughness are higher, with a difference of 10.8 μm^2 ($p < 0.0001$, t-test), 230 nm ($p = 0.0036$, t-test), 321 nm ($p < 0.0001$, t-test), and 706 nm ($p = 0.0002$, t-test), respectively.

Figure 4 shows examples of tissue areas from recurrence and non-recurrence patients in which we visualize differences in

Fig. 2 Confusion matrix of the recurrence prediction on the hold-out test set ($n = 32$). The matrix was normalized and shows the average performance across all cross validation folds.

these morphological features. The elevated values of the nuclei area and roughness parameters suggest that patients who relapsed have bigger nuclei and more nuclei with irregular-shaped cells.

Survival prediction

We next inspected whether the visual features identified by our predictive model are predictive of patient survival. To this end, we used the overall survival and the follow-up status from the clinical data of the patients to visualize the survival curves with a Kaplan–Meier plot.

Figure 5 indicates that patients predicted to recur have lower survival probabilities. Additionally, the survival curve of patients predicted as recurring closely matches the curve of patients that actually experienced recurrence based on the clinical follow up data.

DISCUSSION

Deep learning presents tremendous opportunities for medically relevant prediction tasks, such as recurrence or survival. Additionally, attention-based mechanisms provide insights into the prediction process of the model by enabling quantitative analyses of the tumor morphology and microenvironment features that were crucial to the model's decision.

In this work, we perform image-based prediction of recurrence of DLBCL. We show, that, despite limited dataset size and class imbalance, our model is able to distinguish recurred from non-recurred patients, indicating that H&E-stained images contain meaningful information for such prediction tasks. Our statistical analysis further suggests that the model's prediction decision can be explained by morphological characteristics of the cell nuclei within the high-attention areas of the images. Namely, the model focuses on areas where nuclei of recurred patients tend to be bigger and more irregular in shape as the nuclei of non-recurred patients (Figs. 3 and 4). These features indicate a higher degree of nuclear pleomorphism, which has previously been shown as
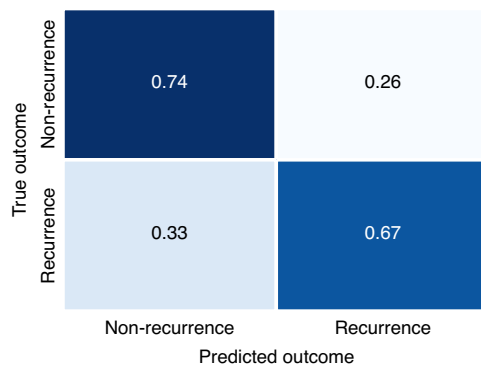
Table 1. Prediction accuracy depending on hyperparameter values and feature encoders.

Learning rate	Weight decay	Metrics	ResNet	UNI	CONCH	CTrans-Path	Prov-GigaPath
0.0001	0.00001	Accuracy	0.7879	0.7500	0.7500	0.7273	0.7424
		B. Acc.	0.5531	0.6626	0.5776	0.6147	0.5883
		AUC	0.5524	0.6146	0.6546	0.5775	0.6149
		F1-score	0.8768	0.8370	0.8481	0.8252	0.8378
0.0001	0.0001	Accuracy	**0.7879**	0.7348	0.7348	0.7197	0.7197
		B. Acc.	**0.5979**	0.6130	0.5732	0.5998	0.6246
		AUC	**0.5994**	0.6304	0.6681	0.5627	0.682
		F1-score	**0.8728**	0.8328	0.8367	0.8189	0.8199
0.00001	0.00001	Accuracy	0.7576	0.7348	0.6818	0.6894	0.7121
		B. Acc.	0.5106	0.6343	0.5458	0.6243	0.5986
		AUC	0.5024	0.7172	0.559	0.5956	0.6549
		F1-score	0.8602	0.8291	0.7962	0.7844	0.8153
0.00001	0.0001	Accuracy	0.7576	0.7045	0.6970	0.6894	0.7273
		B. Acc.	0.5000	0.5819	0.5829	0.6243	0.6165
		AUC	0.5074	0.6650	0.5889	0.6193	0.6619
		F1-score	0.8614	0.8112	0.8039	0.7844	0.8251

The different cores per patient were first transformed into feature vectors using a ResNet-50 trained on ImageNet and different feature encoders trained on H&E-stained images: UNI by Chen et al. [29], CONCH by Lu et al. [30], CTransPath by Wang et al. [28] and Prov-GigaPath by Xu et al. [31]. The length of the feature vectors varies across the different feature encoders. The pretrained Resnet-50 and UNI generate 1024-dimensional features, CONCH produces features of size 512, CTransPath of size 768 and Prov-GigaPath outputs features of length 1536. B. Acc. stands for balanced accuracy. The results of our model are marked in bold.
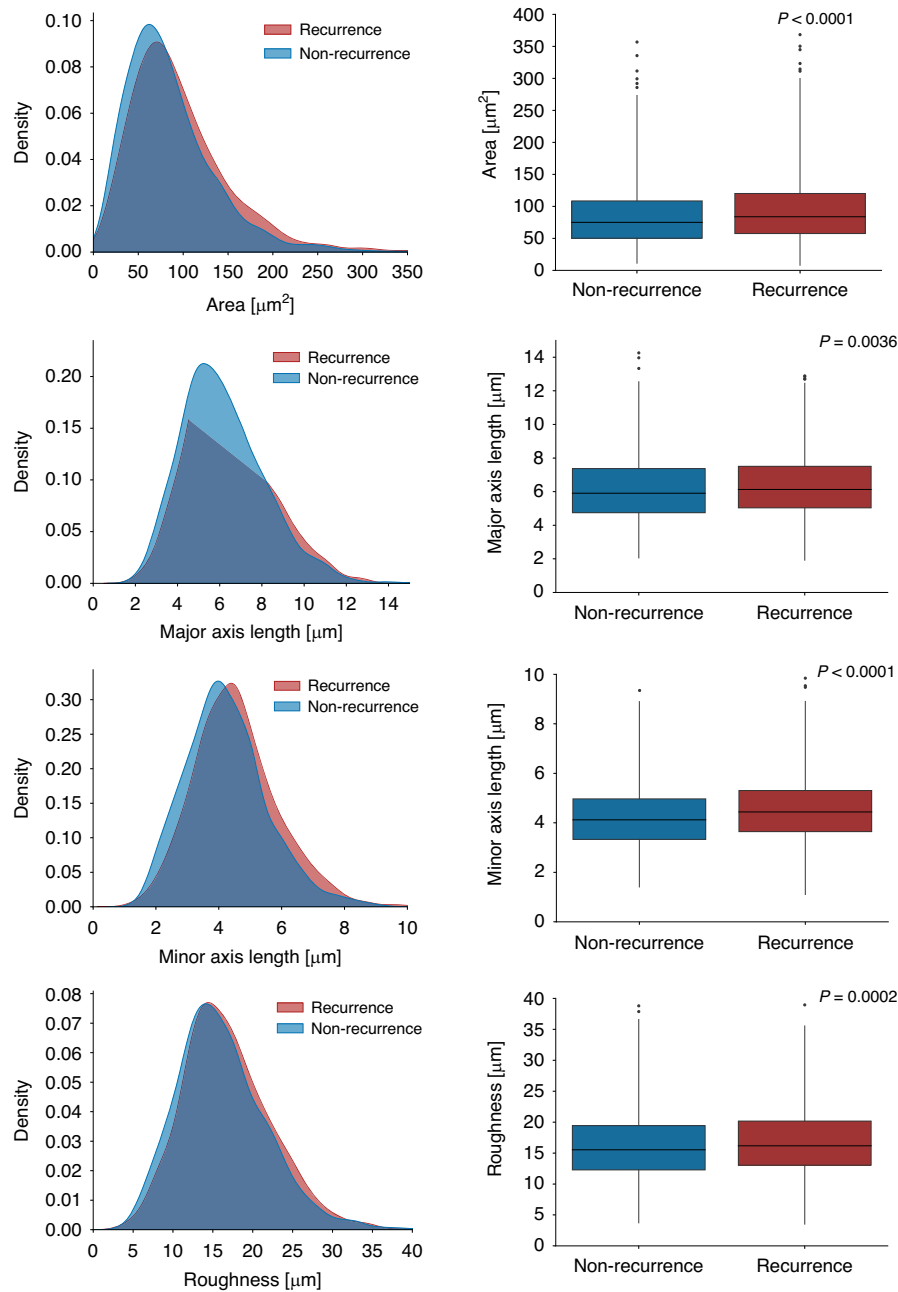
Fig. 3 Distribution and boxplots of the determined morphological features for recurred ($n = 18$, with an average of 63 nuclei per patient) and non-recurred ($n = 20$, with an average of 65 nuclei per patient) patients. We calculated the features 'area', 'major axis length', 'minor axis length', and 'roughness'. The p values indicate the t-test results to highlight statistically significant differences between the features of recurred and non-recurred patients.

related to tumor aggressiveness [26, 27]. Lastly, our model is indirectly predicting patient survival as well (Fig. 5).

A major limitation of this study is the small size of the dataset and lack of an external validation cohort. To the best of our knowledge the used dataset is the only available one which contains the patient recurrence information. Our cross-validation test however suggests that the information on the disease progression might indeed be encoded in the patient diagnostic images. This finding should be further verified as more DLBCL patient data becomes publicly available in the future.

In addition to the feature extractor pre-trained on ImageNet, we also used the Transformer-based feature extractor CTransPath [28] – a model that has been pretrained on histological images. We

found, however, that the model pre-trained on ImageNet yielded better results.

One limitation of our work is the lack of validation on an independent cohort from a different clinical center. To our knowledge, there is however no other publicly available DLBCL histology image data with reported patient clinical outcomes. We aimed to mitigate the model by training and testing the model in several cross-validation iterations.

Our study highlights the importance and high information content of histological images for prediction of treatment outcome. In combination with the ability to make our model biologically interpretable, it allows for descriptive, quantitative, and thorough distinction between recurred and non-recurred patients offering the
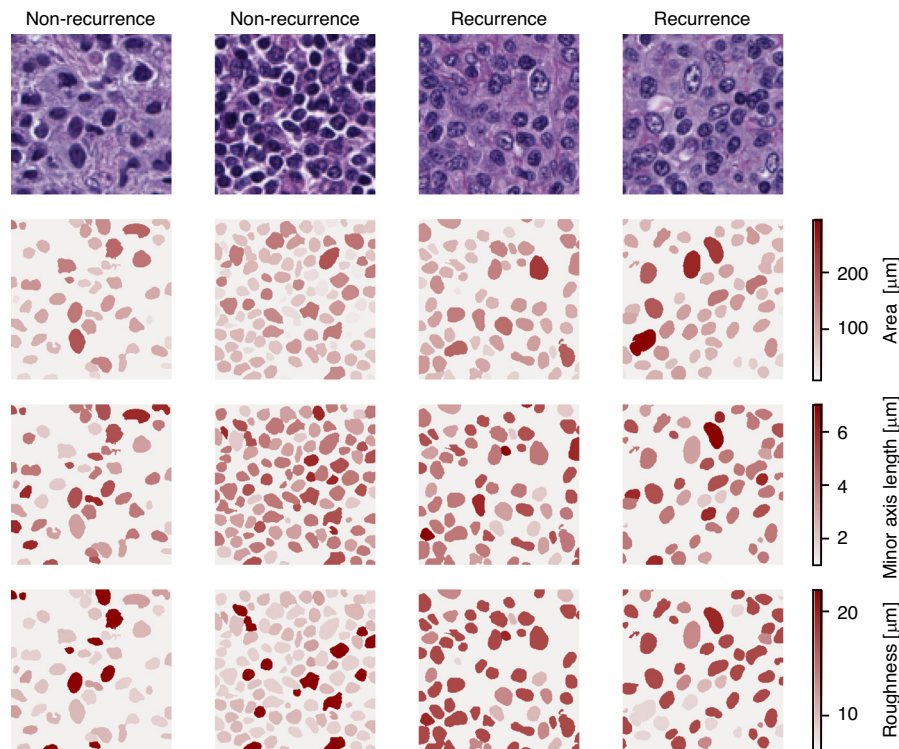
Fig. 4 Heatmap visualization of feature values. For the top two patches of recurred and non-recurred patients, we visualize the strengths of the features 'area', 'minor axis length', and 'roughness'. The darker the color of a cell the higher its respective feature value.
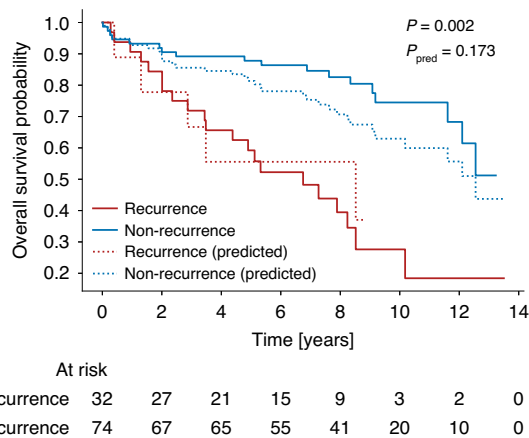

Fig. 5 Kaplan–Meier plot for overall survival. The plot shows the overall survival probability of recurred and non-recurred patients (true and predicted) over time. The variable p indicates the t-test-based p-value for true recurrence and non-recurrence and p_{pred} the p value for predicted recurrence and non-recurrence.

potential for better diagnostics and better understanding of the complexity of fundamental biological processes in aggressive, heterogeneous cancer types, such as DLBCL.

DATA AVAILABILITY
All data supporting the results reported in this article can be found here: https://github.com/stanfordmlgroup/DLBCL-Morph.

REFERENCES
1. Susanibar-Adaniya S, Barta SK. 2021 Update on Diffuse large B cell lymphoma: a review of current data and potential applications on risk stratification and management. Am J Hematol. 2021;96:617–29.

2. Lucas M, Jansen I, van Leeuwen TG, Oddens JR, de Bruin DM, Marquering HA. Deep learning–based recurrence prediction in patients with non–muscle-invasive bladder cancer. Eur Urol Focus. 2022;8:165–72.

3. Yang. Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. Computational Struct Biotechnol J. 2022;20:333–42.

4. Howard FM, Dolezal J, Kochanny S, Khramtsova G, Vickery J, Srisuwananukorn A, et al. Integration of clinical features and deep learning on pathology for the prediction of breast cancer recurrence assays and risk of recurrence. npj Breast Cancer. 2023;9:25.

5. Shi Y, Olsson LT, Hoadley KA, Calhoun BC, Marron JS, Geradts J, et al. Predicting early breast cancer recurrence from histopathological images in the Carolina Breast Cancer Study. npj Breast Cancer. 2023;9:92.

6. Su Z, Niazi MKK, Tavolara TE, Niu S, Tozbikian GH, Wesolowski R, et al. BCR-Net: a deep learning framework to predict breast cancer recurrence from histopathology images. PLoS ONE. 2023;18:e0283562.

7. Foersch S, Glasner C, Woerl AC, Eckstein M, Wagner DC, Schulz S, et al. Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. Nat Med. 2023;29:430–9.

8. Xiao H, Weng Z, Sun K, Shen J, Lin J, Chen S, et al. Predicting 5-year recurrence risk in colorectal cancer: development and validation of a histology-based deep learning approach. Br J Cancer. 2024;130:951–60.

9. Yamashita R, Long J, Saleem A, Rubin DL, Shen J. Deep learning predicts post-surgical recurrence of hepatocellular carcinoma from digital histopathologic images. Sci Rep. 2021;11:2047.

10. Qu WF, Tian MX, Qiu JT, Guo YC, Tao CY, Liu WR, et al. Exploring pathological signatures for predicting the recurrence of early-stage hepatocellular carcinoma based on deep learning. Front Oncol. 2022;12:968202.

11. Wu Z, Wang L, Li C, Cai Y, Liang Y, Mo X, et al. DeepLRHE: a deep convolutional neural network framework to evaluate the risk of lung cancer recurrence and metastasis from histopathology images. Front Genet. 2020;11:768.

12. Akram F, Wolf JL, Trandafir TE, Dingemans AC, Stubbs AP, von der Thüsen JH. Artificial intelligence-based recurrence prediction outperforms classical histopathological methods in pulmonary adenocarcinoma biopsies. Lung Cancer. 2023;186:107413.

13. Dietrich E, Fuhlert P, Ernst A, Sauter G, Lennartz M, Stiehl HS, et al. Towards explainable end-to-end prostate cancer relapse prediction from H&E images combining self-attention multiple instance learning with a recurrent neural network. Proc Mach Learn Res. 2021;158:38–53.

14. Eminaga, O, Saad, F, Tian, Z, Wolffgang, U, Karakiewicz, PI, Ouellet, V et al. Artificial intelligence helps to predict recurrence and mortality for prostate

cancer using histology images. Preprint. 2023, https://doi.org/10.1101/2023.07.27.550781.

15. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: IEEE conference on computer vision and pattern recognition. Las Vegas 2016:770–8. https://doi.org/10.1109/CVPR.2016.90.

16. Harkins RA, Chang A, Patel SP, Lee MJ, Goldstein JS, Merdan S, et al. Remaining challenges in predicting patient outcomes for diffuse large B-cell lymphoma. Expert Rev Hematol. 2019;12:959–73.

17. Wang L, Zhao Z, Luo Y, Yu H, Wu S, Ren X, et al. Classifying 2-year recurrence in patients with dlbcl using clinical variables with imbalanced data and machine learning methods. Comput Methods Prog Biomed. 2020;196:105567.

18. Fan S, Zhao Z, Zhang Y, Yu H, Zheng C, Huang X, et al. Probability calibration-based prediction of recurrence rate in patients with diffuse large B-cell lymphoma. BioData Min. 2021;14:38.

19. Xing M, Zhang Y, Yu H, Yang Z, Li X, Li Q, et al. Predict DLBCL patients' recurrence within two years with Gaussian mixture model cluster oversampling and multi-kernel learning. Comput Methods Prog Biomed. 2022;226:107103.

20. Shankar V, Yang X, Krishna V, Tan BT, Silva O, Rojansky R, et al. LymphoML: an interpretable artificial intelligence-based method dentifies morphologic features that correlate with lymphoma subtype. Proc Mach Learn Res. 2023;225:528–58.

21. Vrabac D, Smit A, Rojansky R, Natkunam Y, Advani RH, Ng AY, et al. DLBCL-Morph: Morphological features computed using deep learning for an annotated digital DLBCL image set. Sci Data. 2021;8:135.

22. 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;5:555–70.

23. Graham S, Vu QD, Raza SEA, Azam A, Tsang YW, Kwak JT, et al. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med Image Anal. 2019;58:101563.

24. Naji H, Sancere L, Simon A, Büttner R, Eich ML, Lohneis P, et al. HoLy-Net: Segmentation of histological images of diffuse large B-cell lymphoma. Comput Biol Med. 2024;170:107978.

25. Phillip JM, Han KS, Chen WC, Wirtz D, Wu PH. A robust unsupervised machine-learning method to quantify the morphological heterogeneity of cells and nuclei. Nat Protoc. 2021;16:754–74.

26. Rashid F, Haque AU. Frequencies of different nuclear morphological features in prostate adenocarcinoma. Ann Diagnostic Pathol. 2011;15:414–21.

27. Fischer EG. Nuclear morphology and the biology of cancer cells. Acta Cytologica. 2020;64:511–9.

28. Wang X, Yang S, Zhang J, Wang M, Zhang J, Yang W, et al. Transformer-based unsupervised contrastive learning for histopathological image classification. Med Image Anal. 2022;81:102559.

29. Chen RJ, Ding T, Lu MY, Williamson DFK, Jaume G, Song AH, et al. Towards a general-purpose foundation model for computational pathology. Nat Med. 2024;30:850–62.

30. Lu MY, Chen B, Williamson DF, Chen RJ, Liang I, Ding T, et al. A visual-language foundation model for computational pathology. Nat Med. 2024;30:863–74.

31. Xu H, Usuyama N, Bagga J, Zhang S, Rao R, Naumann T, et al. A whole-slide foundation model for digital pathology from real-world data. Nature. 2024;630:181–8.

AUTHOR CONTRIBUTIONS

HN and PH performed data analysis, paper writing and software development. JIP and SU contributed to software development and data analysis. AS contributed to data analysis. RB and KB designed the study. KB also wrote the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All methods were performed in accordance with the relevant guidelines and regulations.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Katarzyna Bozek.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.