Received: 17 February 2023 | Revised: 13 July 2023 | Accepted: 17 July 2023

DOI: 10.1111/ejh.14066

ORIGINAL ARTICLE

European Journal of Haematology

WILEY

Deep-learning based classification of a tumor marker for prognosis on Hodgkin's disease

Ila Motmaen[1] | Sergej Sereda[1] | Alexander Brobeil[2] | Ananth Shankar[3] | Andreas Braeuninger[1] | Dirk Hasenclever[4] | Stefan Gattenlöhner[1]

[1]Department of Pathology, Justus-Liebig-University, University Hospital Giessen and Marburg GmbH, Giessen, Germany

[2]Department of Pathology, University Hospital Heidelberg, Heidelberg University, Heidelberg, Germany

[3]Children and Young People's Cancer Services, University College London Hospitals NHS Foundation Trust, London, UK

[4]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany

Correspondence
Ila Motmaen, Department of Pathology, University Hospital Giessen and Marburg (UKGM), Giessen, Germany.
Email: imotmaen@t-online.de

Funding information
Elternverein fuer leukaemie- und krebskranke Kinder, Giessen e.V. and Tour der Hoffnung e.V.; Deutsche Krebshilfe, Grant/Award Number: DKH-70114098

Abstract

Purpose: Hodgkin's disease is a common malignant disorder in adolescent patients. Although most patients are cured, approximately 10%–15% of patients experience a relapse or have resistant disease. Furthermore, there are no definitive molecular predictors for early identification of patients at high risk of treatment failure to first line therapy. The aim of this study was to evaluate the deep learning-based classifier model of medical image classification to predict clinical outcome that may help in appropriate therapeutic decisions.

Methods: Eighty-three FFPE biopsy specimens from patients with Hodgkin's disease were stratified according to the patient's qPET scores, stained with picrosirius red dye and digitalized by whole slide image scanning. The resulting whole slide images were cut into tiles and annotated by two classes based on the collagen fibers' degree of coloring with picrosirius red. The neural network (YOLOv4) was then trained with the annotated data. Training was performed with 30 cases. Prognostic power of the weakly stained picrosirius red fibers was evaluated with 53 cases. The same neural network was trained with MMP9 stained tissue slides from the same cases and the quantification results were compared with the variant from the picrosirius red cases.

Results: There was a weak monotonically increasing relationship by parametric ANOVA between the qPET groups and the percentages of weakly stained fibers ($p = .0185$). The qPET-positive cases showed an average of 18% of weakly stained fibers, and the qPET-negative cases 10%–14%. Detection performance showed an AUC of 0.79.

Conclusions: Picrosirius red shows distinct associations as a prognostic metric candidate of disease progression in Hodgkin's disease cases using whole slide images but not sufficiently as a prognostic device.

KEYWORDS
digital pathology, Hodgkin lymphoma, MMP9, pediatrics, picrosirius red

1 | INTRODUCTION

Hodgkin's disease is the most common lymphoma disease in adolescents, with approximately 80 000–100 000 cases per year worldwide.[1] Typical signs of this malignancy are B-symptoms, lymphopenias as well as increased inflammation levels.[1,2] Treatment of Hodgkin's disease usually consists of multi-agent combination chemotherapy with or without radiotherapy.[3] The choice of therapy depends on the stage of disease and prognostic parameters, that is, the patient's risk of relapse.[4,5] Currently, only clinical criteria including affected lymph node areas, B symptoms, treatment response measured by qualitative PET analyses, volume, occurrence of bulky disease or mediastinal disease exist to predict the risk of progression or recurrence.[1,2,6] A potential histological parameter—the quantification of MMP9-positive cells—was currently discussed in several publications.[7–9] MMP9 is a matrixin belonging to the zinc-metalloproteinases family involved in collagenous degradation of the extracellular matrix.[10] The discovery of a robust and reliable prognostic marker would help especially in the selection of therapy for younger patients. Here we face a trade-off between efficacy and side effects caused by therapy often leading to severe side effects in later stages of life,[11] a more conservative approach for patients with better prognosis would help to diminish the risk of the later complications. A prerequisite for the success and acceptance of the use of a histologic prognostic marker is an accurate quantification capability.[12] The usual methods of quantifying, for example, MMP9 or other histological parameters with the microscope by the practitioner himself turns out to be highly subjective and extremely time-consuming. As data volumes continue to grow, artificial intelligence algorithms are becoming increasingly important where human performance is reaching its limits in terms of time and computational capacity.[13–15] For the quantification of histological parameters, convolutional deep learning networks (CNN) have become increasingly important serving as a suitable deep learning architecture for analyzing tissue structures on whole slide images.[16–18] Due to MMP9 quantification in classical Hodgkin lymphoma had shown significant prognostic power[7] we hypothesized that picrosirius red collagen staining of Hodgkin's lymphoma tissue might serve as a prognostic signal in young patients. As it was almost impossible to accurately quantify the absolute amount of faintly stained fibers through quantification of the microscopic field of view, we decided to use a deep learning approach with the CNN model YOLOv4 which has been successfully used in several medical computer vision tasks regarding object detection on whole slides.[19–21] We chose patients from the EuroNet-PHL-C2 study that were representative for different degrees of response to perform automated analysis and another fraction to test our hypothesis statistically.

2 | MATERIALS AND METHODS

2.1 | Data collective

All of the 83 cases used for the investigations were selected from the EuroNet-PHL-C2 study, an international cross-group study of classical Hodgkin's lymphoma in children and adolescents. The treatment regimen based on a risk- and response-adapted treatment approach.[22,23] Patients were all administered two cycles of induction chemotherapy by vincristine, etoposide, prednisolone, and doxorubicin (OEPA), followed by early response assessment (ERA), including whole-body CT or MRI and FDG-PET scan.[24] Patients who achieved complete metabolic remission received one to four cycles of further chemotherapy according to diagnostic stage and stratification of the treatment risk group.[25] All other patients received radiation to the affected site after completion of their planned chemotherapy regimen (ISRT). We used qPET, a quantitative extension of the Deauville scale used in evaluating treatment response of FDG-PET scans in lymphoma patients, after initial primary chemotherapy as an indicator of abnormal tumor metabolism and stratified cases using the qPET thresholds proposed by Hasenclever et al. (2014).[26] Accordingly, the cases were selected to represent four qPET groups ("D 1–4") (Table 1). For some analyses, the D-groups were additionally unified into a qPET-negative group (D1 + D2 as qPET-Neg., qPET < 1.3, median age = 14 years, range = 3–17 years, 23 males, 11 females) and a qPET-positive group (D3 + D4 as qPET-Pos., qPET > 1.3, median age = 15 years, range = 6–17 years, 13 males, 6 females) according to the qPET threshold value of 1.3 (Table 1). The biopsy formalin fixed and paraffin embedded (FFPE) material was cut into sections of about 5 µm thickness using a Leica SM 2000R sliding microtome (Leica Microsystems, Wetzlar, Germany). Selected slide-mounted sections were deparaffinized, incubated with primary antibody against MMP9 (mouse anti-human monoclonal Ab, clone 5G3, dilution 1:5000, antigen retrieval with EDTA, Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA), and visualized with 3,3′-Diaminobenzidine (DAB) using Bond Polymer Refine Detection Kit (DS9800, Leica Biosystems, Newcastle, UK) on a Leica BOND-MAX automated staining platform (Leica Microsystems, Wetzlar, Germany) according to manufacturer's protocol. Another set of deparaffinized slide-mounted sections from the same cases underwent manufacturer's picrosirius red staining protocol (ab150681—Picro Sirius Red Stain Kit (Connective Tissue Stain), Abcam, Cambridge, UK). Afterwards, the stained slides were scanned using a Panoramic SCAN II slide scanner (3DHISTECH, Budapest, Hungary) equipped with a 20× plan apochromat objective and a 5-megapixel CMOS camera mounted on a 0.63× C-mount adapter. The obtained whole slide images (WSI) in MRSX format had a resolution of 0.2749 × 0.2749 µm/pixel.

2.2 | CNN training

From the selected 83 cases, 30 WSIs were used to provide training data for a neural network and 53 WSIs to evaluate the prognostic significance of the biomarker candidate. The whole slide images were first divided into 320 × 320 pixel tiles in original full resolution with the open source software QuPath 0.3.2[27] to provide suitable input format for the convolutional neural network YOLOv4.[28] Training data annotation was performed using LabelImg[29] by drawing rectangles around the objects of interest including them in so-called bounding boxes. The annotated objects were weakly and strongly stained fibers in the case of picrosirius red staining, MMP9-positive and -negative cells for MMP9

TABLE 1 Picrosirius red detection results grouped by qPET values.

Group	Group size	Age in years [median(range)]	Sex (male/female)	qPET range	qPET mean (SD)	PicLow percent mean (SD)
D1	19	14 (3–17)	12/7	<0.95	0.52 (0.41)	9.86 (4.91)
D2	15	14 (4–17)	11/4	0.95–1.3	1.12 (0.09)	12.41 (5.71)
qPET-Neg.	34	14 (3–17)	23/11	<1.3	0.78 (0.43)	10.98 (5.35)
D3	12	15 (6–17)	9/3	1.3–3.0	1.64 (0.41)	14.09 (3.86)
D4	7	15 (10–17)	4/3	>3	5.43 (3.43)	18.12 (6.15)
qPET-Pos.	19	15 (6–17)	13/6	>1.3	3.04 (2.74)	15.58 (5.06)

Note: Groups were formed by qPET ranges as D1–D4. D groups were additionally paired by qPET threshold value 1.3 into a qPET-negative group and a qPET-positive group.

staining. Transfer-learning based training of the model YOLOv4 was carried out with pre-trained weights (MS-COCO).[30] The complete training set for picrosirius red included 7134 tiles. For MMP9, the complete training set consisted of 5788 tiles. The training sets were split into a training subset and a validation subset fixed at a ratio of 80:20 in both settings (picrosirius red and MMP9). The achieved mean average precision was 73.4% for picrosirius red and 87.6% for MMP9 WSIs. The trained neural network was used to verify the biomarker candidate as a suitable prognostic entity using the preselected 53 WSIs and their clinical data with 953.068 tiles from picrosirius red and 409.406 tiles from the MMP9 WSIs.

2.3 | Final analysis and development of fiber quantification technique

Trained YOLOv4 models were used to analyze tiles from 53 WSIs dedicated to the final analysis. In the case of MMP9 the result of the analysis consisted of coordinates of the bounding boxes—network generated frames enclosing detected cells. For the further analysis, we quantified (counted) the number of bounding boxes corresponding to the detected (MMP9 positive) cells per WSI. In the case of picrosirius red staining, the enclosing detected objects were weakly or strongly stained fibers. As the length of the detected fibers differed among each WSI and between WSIs, we decided to quantify the length of the fibers enclosed by bounding boxes rather than the number of boxes as such. At the time of manuscript preparation, there was no established method to reliably quantify fiber lengths on WSIs. We decided to measure the diagonals of the bounding boxes of the weakly colored fibers cumulatively (in pixels). In a preliminary investigation, the actual fiber lengths (ground truth) in pixels, measured by ruler tool in QuPath, within the bounding boxes of each tile were compared to the length of the diagonals of the bounding boxes. Each pair of fiber length inside a bounding box and its diagonal length (Figure 1) were compared using *T*-test ($p > .05$). Thus, the diagonal proved to be a suitable proxy for the fiber length. For each case, the total length of diagonals of the bounding box of all detected weakly stained fibers was summed and calculated as a percentage of the total area of the WSI with a custom Python script ("PicLow Percent," Table 1).

2.4 | Statistics

Parametric ANOVA was performed using R programming language[31] to compare means of weakly stained fiber percentage in cases across the groups stratified by qPET thresholds (D 1–4). Logistic regression with binarized dependent variable (qPET ≥ 1.3 = 1, qPET < 1.3 = 0) and the percentage of weakly stained fibers as independent variable was performed. The qPET thresholds were represented in the results as a PET-positive group (qPET ≥ 1.3) and a PET-negative group (qPET < 1.3 = 0). In order to measure detection performance for discriminating clinically favorable and unfavorable cases based on the percentage of weakly stained fibers receiver operating characteristic curve (ROC) and the area under the curve (AUC) of the regression was conducted. The model's goodness of fit was evaluated by McFadden pseudo R^2 and Chi-Square score (X^2). For ROC, the confidence interval based on bootstrapping with 2000 replications was constructed at a confidence level of 95%. An AUC of 69% or above was considered statistically significant. The Seaborn package,[32] Matplotlib[33] and pROC package[34] was used for graphical plotting. A p value <.05 was considered statistically significant.

3 | RESULTS

The parametric ANOVA revealed that there was a weak monotonically increasing relationship between groups D1–D4 and the percentages of weakly stained collagen fibers ($p = .0185$) (Figure 2). Among unified qPET groups, the qPET-positive cases (unfavorable prognosis) had an average of 18% of weakly stained fibers, and the qPET-negative cases (favorable prognosis) had 10%–14%. Based on the logistic regression performed on unified qPET groups, the coefficients for "PicLow Percent" means were found to have a value of 0.256 (standard error: 1.39, $p = .006$) and an intercept of −4.14 (standard error: 0.09, $p = .003$). The McFadden pseudo R^2 was 0.255 and Chi-Square score (X^2) was 11.698 with $p = .0006$. Receiver operating characteristic (ROC) analysis showed an area under the curve (AUC) of 79.2% with a confidence interval of 67.3%–91.2% constructed with 2000 bootstraps (Figure 3). Comparing qPET-negative cases (D1, D2)

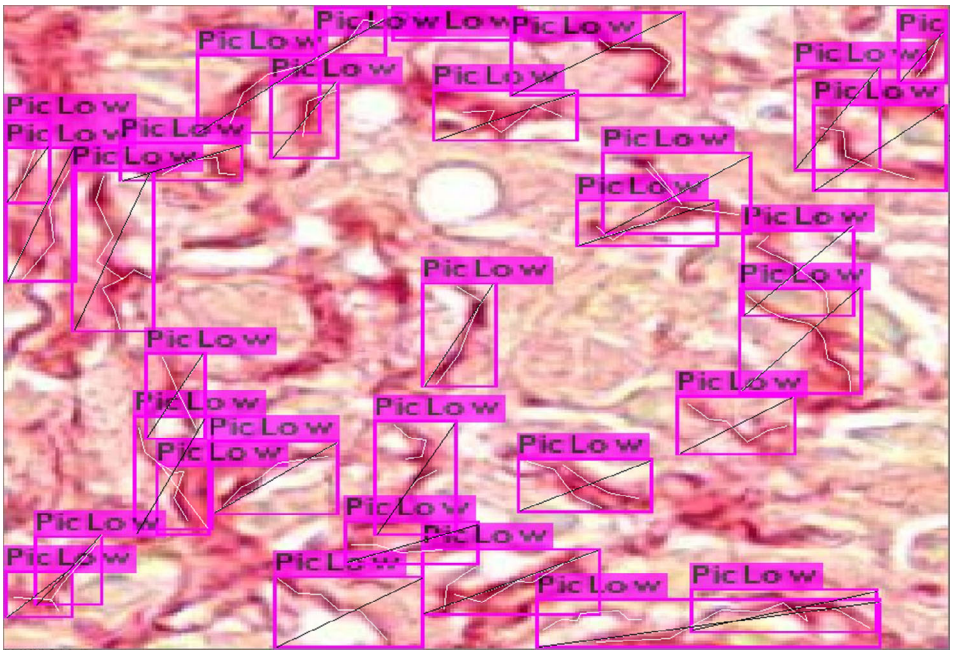

FIGURE 1 Bounding box diagonal and ground truth. Actual weakly stained fiber lengths on picrosirius red sections and the diagonals of the corresponding bounding boxes (purple rectangles) were measured. On each bounding box, its diagonal length was compared to the true fiber length inside the box and length difference between each of them was tested for statistical significance (*T*-test) to verify bounding box diagonals as a suitable proxy for actual fiber length inside each bounding box. In the course of the investigations weakly stained picrosirius fibers were labeled as "PicLow." The thin yellow lines represent the measured fiber lengths and the thin black lines the diagonal of each bounding box.

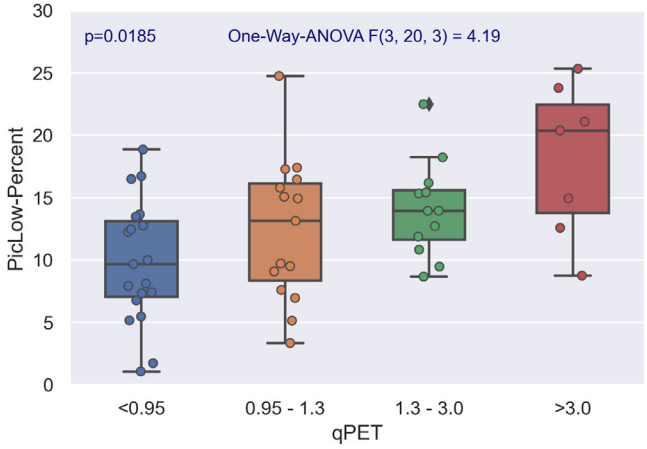

FIGURE 2 Box plot and result of analysis of variance of the 53 cases grouped by qPET ranges. Colored dots = individual measurements, black bars in boxes = median of groups, boxes = upper and lower quartiles, whiskers = data extremes within 1.5 times the quartiles. There were significant differences between the means of qPET ranges.

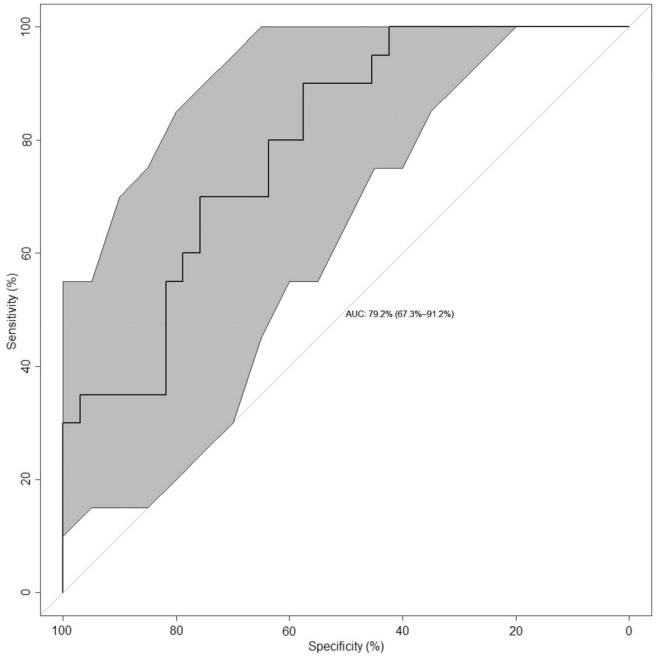

FIGURE 3 Receiver operating characteristic curve (ROC; thick stepped line) of the logistic regression of PicLow on qPET-groups (negative = D1, 2; positive = D3, D4). The thin diagonal line as the boundary represents an algorithm that would make only random-based decisions (AUC 0.5). The area under the curve (AUC) value is 79.2% with a CI of 67.3%–91.2% (grey area), constructed by bootstrapping, at a confidence level of 95% which indicates a satisfying evaluation performance with a qPET threshold value of 1.3.

and qPET-positive cases (D3, D4) showed differences in the studied areas not only in terms of staining intensity between cases containing a lot of weakly stained fibers (qPET-positive) and such cases with only few weakly stained but with a higher content of strongly stained fibers (qPET-negative). In addition, strongly stained fibers showed a more homogeneous, smoother texture with uniform bright red coloration, whereas weakly stained fibers showed a more frayed texture with inhomogeneous color distribution in the pink to purple range. Furthermore, a visibly higher number as well as density of MMP9-positive cells could be observed in the immediate vicinity of the weakly stained fibers on the picrosirius red slides analogously to the MMP9 WSIs (Figure 4).

4 | DISCUSSION

Our preliminary observations on a limited number of 9 cases showed that cases with complicated treatment response and/or relapse had a

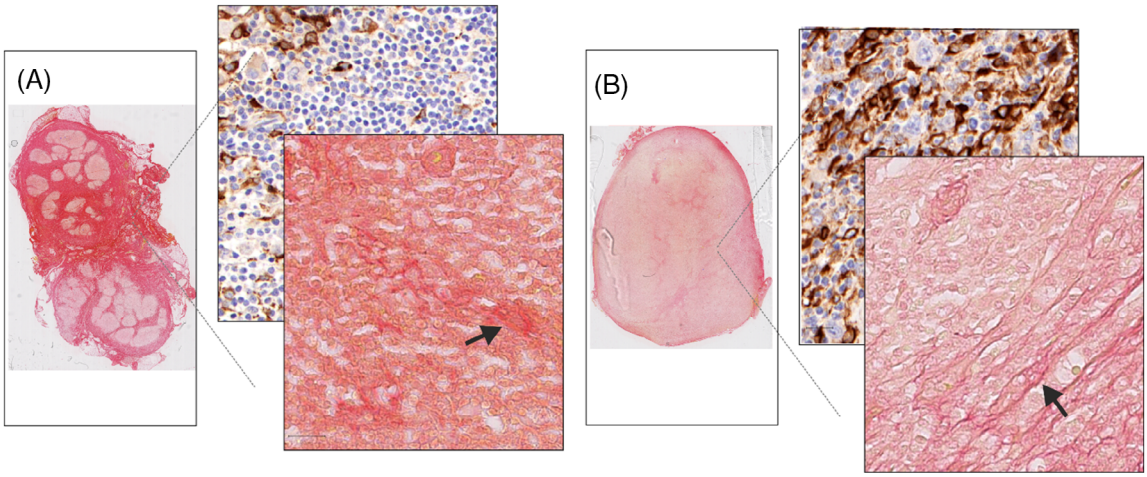

FIGURE 4 Relationship between the amount of collagen fibers weakly stained with picrosirius red staining and qPET groups. (A) Case K620-16 representing a patient with complete remission (qPET < 1.3). Detailed image center top: histological MMP9 staining. Detailed image center bottom: picrosirius red staining. Black arrow shows a strongly (bright red) stained collagen fiber. (B) Case K600-16 representing a relapsed patient (qPET > 1.3). Detailed image center top: histological MMP9 staining. Detail image center bottom: picrosirius red staining. Black arrow indicates a weakly (pinkish-purple) stained collagen fiber.

different microscopic appearance of collagen fibers stained with picrosirius red than cases with regular treatment response: Tissue slides from patients with worse therapy outcomes appeared to have higher amounts of collagen fibers with weaker dye and irregular consistency than their counterparts with better therapy outcome. The results of the performed analysis might be mitigated by an inconsistent or unprecise fiber detection algorithm. As a crucial step towards prevention, the training material must be annotated as well as possible.[35,36] Annotating the training images depends on the perception of the person performing the training and on his knowledge of how to distinguish regular collagen fibers from abnormal fibers which was the responsibility of a well-trained physician.[37] These results were statistically significant observations in the differences in the proportion of weakly stained picrosirius red collagen fibers found by ANOVA when cases were grouped by qPET values. Furthermore, they indicated a positive correlation between qPET values and the proportion of weakly stained collagen fibers. However, the statistical connection between these two parameters is not large enough to provide a confident prognosis for a less favorable clinical course in Hodgkin lymphoma. Although, the detected relationship may represent a disease course relevant connection between fiber staining patterns and prognostic outcome among cases. This relationship could be based on the differences in collagen metabolism in the qPET-positive cases associated with poorer clinical outcome: an additional investigation revealed that the collagen modulating enzyme MMP9 was more expressed in the cases with an increased proportion of weakly stained fibers than in the cases with fewer such fibers. Increased MMP9 expression correlated with reduced overall survival in 148 young adult cases in one study.[7] Various other works indicated MMP9 as a correlate to collagen alteration in lymphomas.[38–40] MMP9 contributes to a family of enzymes that are involved in the degradation of the extracellular matrix and subsequently in the processes of invasion as well as metastasis of many human tumors.[41] These results confirm our assumption that the detected faintly stained fibers are not artifacts but a result of biological processes modulating collagen fibers. There are other published studies in literature that relate collagen modifications in lymphomas to clinical outcome.[42,43]

In further addition, it should be mentioned that the YOLOv4 neural network in particular has already been successfully used in other fields, such as the textile industry, for the detection of fibers making it also suitable for the detection of collagen fibers.[44,45] Our results appear promising and probably indicates a likely correlation between the percentage of weakly dyed collagen fibers and picrosirius red staining. Validation of these findings in a larger cohort will be necessary to confirm our results.

AUTHOR CONTRIBUTIONS

Ila Motmaen and Sergej Sereda developed the workflow. Ila Motmaen and Sergej Sereda wrote the original manuscript. Ila Motmaen and Alexander Brobeil prepared and performed annotations. Stefan Gattenlöhner conceptualized the research and obtained funding. Andreas Braeuninger helped with constructive criticism on the initial draft. Stefan Gattenlöhner and Ananth Shankar helped with interpretation of the results. Dirk Hasenclever advised in statistics.

ACKNOWLEDGMENTS

We kindly thank Elternverein fuer leukaemie- und krebskranke Kinder, Giessen e.V. and Tour der Hoffnung e.V. for their generous financial support. SG was supported by the Deutsche Krebshilfe through grant DKH-70114098. We also would like to thank Alexander Brobeil for constructive criticism of the initial draft. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

All authors have declared no conflict of interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Ila Motmaen https://orcid.org/0000-0002-9094-8756
Sergej Sereda https://orcid.org/0000-0003-3212-8722
Alexander Brobeil https://orcid.org/0000-0002-1003-8577
Ananth Shankar https://orcid.org/0000-0001-9885-5776
Andreas Braeuninger https://orcid.org/0000-0003-1666-8163
Dirk Hasenclever https://orcid.org/0000-0003-3224-1610
Stefan Gattenlöhner https://orcid.org/0000-0002-8673-6777

REFERENCES

1. Zhou L, Deng Y, Li N, et al. Global, regional, and national burden of Hodgkin lymphoma from 1990 to 2017: estimates from the 2017 Global Burden of Disease study. *J Hematol Oncol*. 2019; 12(1):107.
2. 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(6):394-424.
3. Specht L, Yahalom J, Illidge T, et al. Modern radiation therapy for Hodgkin lymphoma: field and dose guidelines from the international lymphoma radiation oncology group (ILROG). *Int J Radiat Oncol Biol Phys*. 2014;89(4):854-862.
4. Morschhauser F, Brice P, Fermé C, et al. Risk-adapted salvage treatment with single or tandem autologous stem-cell transplantation for first relapse/refractory Hodgkin's lymphoma: results of the prospective multicenter H96 trial by the GELA/SFGM Study Group. *J Clin Oncol*. 2008;26(36):5980-5987.
5. Moskowitz CH, Nademanee A, Masszi T, et al. Brentuximab vedotin as consolidation therapy after autologous stem-cell transplantation in patients with Hodgkin's lymphoma at risk of relapse or progression (AETHERA): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*. 2015;385(9980):1853-1862.
6. Shanbhag S, Ambinder RF. Hodgkin lymphoma: a review and update on recent progress. *CA Cancer J Clin*. 2018;68(2):116-132.
7. Campos AH, Vassallo J, Soares FA. Matrix metalloproteinase-9 expression by Hodgkin-Reed-Sternberg cells is associated with reduced overall survival in young adult patients with classical Hodgkin lymphoma. *PLoS One*. 2013;8(9):e74793.
8. Hazar B, Polat G, Seyrek E, Bagdatoglglu O, Kanik A, Tiftik N. Prognostic value of matrix metalloproteinases (MMP-2 and MMP-9) in Hodgkin's and non-Hodgkin's lymphoma. *Int J Clin Pract*. 2004;58(2): 139-143.
9. Kuittinen O, Soini Y, Turpeenniemi-Hujanen T. Diverse role of MMP-2 and MMP-9 in the clinicopathological behavior of Hodgkin's lymphoma. *Eur J Haematol*. 2002;69(4):205-212.
10. Nagase H, Woessner JF Jr. Matrix metalloproteinases. *J Biol Chem*. 1999;274(31):21491-21494.
11. Aleman BM, van den Belt-Dusebout AW, Klokman WJ, Van't Veer MB, Bartelink H, van Leeuwen FE. Long-term cause-specific mortality of patients treated for Hodgkin's disease. *J Clin Oncol*. 2003; 21(18):3431-3439.
12. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: a review. *IEEE Rev Biomed Eng*. 2009;2:147-171.
13. Atasoy H, Greenwood BN, McCullough JS. The digitization of patient care: a review of the effects of electronic health records on health care quality and utilization. *Annu Rev Public Health*. 2019;40(1): 487-500.
14. Baumgärtel M, Riessen R, John S. Digitalisierung in der Intensivmedizin. *DMW – Deutsche Medizinische Wochenschrift*. 2019;144(7): 436-441.
15. Li KC, Marcovici P, Phelps A, et al. Digitization of medicine. *Acad Radiol*. 2013;20(12):1479-1494.
16. Ghazvinian Zanjani F, Zinger S, Piepers B, Mahmoudpour S, Schelkens P, de With PHN. Impact of JPEG 2000 compression on deep convolutional neural networks for metastatic cancer detection in histopathological images. *J Med Imaging (Bellingham)*. 2019;6(2): 027501.
17. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015; 521(7553):436-444.
18. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88.
19. Zheng Z, Zhang X, Ding J, et al. Deep learning-based artificial intelligence system for automatic assessment of glomerular pathological findings in lupus nephritis. *Diagnostics (Basel)*. 2021;11(11):1983.
20. Nair LS, Prabhu R, Sugathan G, Gireesh KV, Nair AS, eds. Mitotic nuclei detection in breast histopathology images using YOLOv4. *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE; 2021.
21. Huang Y, Xue Y, Lan J, et al., eds. Deep learning framework for detecting positive lymph nodes of gastric cancer on histopathological images. *2021 6th International Conference on Biomedical Imaging, Signal Processing*. Association for Computing Machinery (ACM); 2021.
22. Mauz-Korholz C, Landman-Parker J, Fernandez-Teijeiro A, et al. Response-adapted omission of radiotherapy in children and adolescents with early-stage classical Hodgkin lymphoma and an adequate response to vincristine, etoposide, prednisone, and doxorubicin (EuroNet-PHL-C1): a titration study. *Lancet Oncol*. 2023;24(3): 252-261.
23. Mauz-Korholz C, Landman-Parker J, Balwierz W, et al. Response-adapted omission of radiotherapy and comparison of consolidation chemotherapy in children and adolescents with intermediate-stage and advanced-stage classical Hodgkin lymphoma (EuroNet-PHL-C1): a titration study with an open-label, embedded, multinational, non-inferiority, randomised controlled trial. *Lancet Oncol*. 2022;23(1): 125-137.
24. Schulte M, Brecht-Krauss D, Heymer B, et al. Grading of tumors and tumorlike lesions of bone: evaluation by FDG PET. *J Nucl Med*. 2000; 41(10):1695-1701.
25. Lo AC, Dieckmann K, Pelz T, et al. Pediatric classical Hodgkin lymphoma. *Pediatr Blood Cancer*. 2021;68(Suppl 2):e28562.
26. Hasenclever D, Kurch L, Mauz-Korholz C, et al. qPET – a quantitative extension of the Deauville scale to assess response in interim FDG-PET scans in lymphoma. *Eur J Nucl Med Mol Imaging*. 2014;41(7): 1301-1308.
27. Bankhead P, Loughrey MB, Fernandez JA, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep*. 2017;7(1): 16878.
28. Bochkovskiy A, Wang C-Y, Liao H-YM. YOLOv4: optimal speed and accuracy of object detection. arXiv Preprint arXiv:2004.10934. 2020.
29. Tzutalin D. LabelImg is a graphical image annotation tool and label object bounding boxes in images. 2015. https://github.com/tzutalin/labelImg
30. Lin TY, Maire M, Belongie S, et al. Microsoft COCO: common objects in context. *Lect Notes Comput Sc*. 2014;8693:740-755.
31. Team RC. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2012:2021.
32. Waskom M. seaborn: statistical data visualization. *J Open Source Softw*. 2021;6(60):3021.

33. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 2007;9(3):90-95.
34. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
35. Zhang WL, Li RJ, Zeng T, et al. Deep model based transfer and multi-task learning for biological image analysis. *Kdd'15: Proceedings of the 21st Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery (ACM); 2015: 1475-1484.
36. Hawkins DM. The problem of overfitting. *J Chem Inf Comput Sci*. 2004;44(1):1-12.
37. Chollet F. *Deep Learning with Python*. Manning Publications Company; 2017.
38. Aoudjit F, Masure S, Opdenakker G, Potworowski EF, St-Pierre Y. Gelatinase B (MMP-9), but not its inhibitor (TIMP-1), dictates the growth rate of experimental thymic lymphoma. *Int J Cancer*. 1999; 82(5):743-747.
39. Bouchard F, Belanger SD, Biron-Pain K, St-Pierre Y. EGR-1 activation by EGF inhibits MMP-9 expression and lymphoma growth. *Blood*. 2010;116(5):759-766.
40. Citak EC, Oguz A, Karadeniz C, Akyurek N. Role of gelatinases (MMP-2 and MMP-9), TIMP-1, vascular endothelial growth factor (VEGF), and microvessel density on the clinicopathological behavior of childhood non-Hodgkin lymphoma. *Pediatr Hematol Oncol*. 2008;25(1):55-66.
41. Vihinen P, Kahari VM. Matrix metalloproteinases in cancer: prognostic markers and therapeutic targets. *Int J Cancer*. 2002;99(2):157-166.
42. Alfano M, Locatelli I, D'Arrigo C, et al. Lysyl-oxidase dependent extracellular matrix stiffness in Hodgkin lymphomas: mechanical and topographical evidence. *Cancers (Basel)*. 2022;14(1):259.
43. Esteve-Puig R, Climent F, Pineyro D, et al. Epigenetic loss of m1A RNA demethylase ALKBH3 in Hodgkin lymphoma targets collagen, conferring poor clinical outcome. *Blood*. 2021;137(7):994-999.
44. Dlamini S, Kao C-Y, Su S-L, Jeffrey Kuo C-F. Development of a real-time machine vision system for functional textile fabric defect detection using a deep YOLOv4 model. *Text Res J*. 2021;92(5–6): 675-690.
45. Zhang C, Li T, Zhang W. The detection of impurity content in machine-picked seed cotton based on image processing and improved YOLO V4. *Agronomy*. 2021;12(1):66.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Motmaen I, Sereda S, Brobeil A, et al. Deep-learning based classification of a tumor marker for prognosis on Hodgkin's disease. *Eur J Haematol*. 2023;111(5): 722-728. doi:10.1111/ejh.14066