UNIVERSITÄT
ZU KÖLN

Deep Learning for Prediction of Recurrence of Diffuse Large B-Cell Lymphoma

Inaugural-Dissertation zur Erlangung des Doktorgrades der Mathematisch-Naturwissenschaftlichen Fakultät der Universität zu Köln

vorgelegt von Hussein Naji
aus Bagdad, Irak

Köln, 02.05.2024

Berichterstattende:	Prof. Dr. Katarzyna Bożek
	Prof. Dr. Johannes Berg
Prüfungsvorsitzender:	Prof. Dr. Achim Tresch

Tag der mündlichen Prüfung: 12.07.2024

Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Köln, 02.05.2024

Unterschrift

Abstract

Diffuse large B-cell lymphoma (DLBCL) is an aggressive and highly heterogeneous tumor which is fatal without treatment. It is the most common non-Hodgkin lymphoma in adults. Up to today, its treatment presents a serious challenge as up to half of the treated patients suffer from recurrence of this cancer. Therefore, an early detection of those patients who do not benefit from standard first-line treatments would be of high significance as treatment procedures could be more personalized to prevent a recurrence. Previous research suggested that clinical data, gene expression markers, and genetic abnormalities are strong prognostic factors for recurrence of various cancer types. Recently, deep learning (DL) has shown increasing success in predicting recurrence of cancer from histological images. DL models, despite being highly accurate predictors, are often criticized for their black box characteristic. However, state-of-the-art research, such as models that use multiple instance learning, attention-based mechanisms, and/or class activation mapping methods, indicates that it is possible to make models more explainable. From a biological perspective, the morphology of tumor cells has also been used for several medical prediction and classification tasks because it is crucial to grasp a full picture of the tumor microenvironment. All experiments conducted within this project are based on histological images of DLBCL. Firstly, it is shown that morphological features extracted from cell nuclei are strong indicators to distinguish between different cell types in DLBCL tumors. Secondly, cell nuclei are automatically segmented and classified to close the gap of non-existing segmentation approaches for lymphoma images. The last part deals with the prediction of recurrence of DLBCL and combines the methods and findings of the previous approaches with attention-based mechanisms of DL models to biologically interpret the model's prediction decision. To this end, it was not only possible to distinguish non-recurred from recurred patients but also to explain which morphological features of cell nuclei were indicative for the prediction task. This allows for both, an upfront detection of recurred patients and to find distinct morphological patterns between both groups of interest.

Zusammenfassung

Das diffuse großzellige B-Zell-Lymphom (DLBCL) ist ein aggressiver und sehr heterogener Tumor, der ohne Behandlung tödlich verläuft. Es ist das häufigste Non-Hodgkin-Lymphom bei Erwachsenen. Bis heute stellt seine Behandlung eine ernste Herausforderung dar, da bei bis zur Hälfte der behandelten Patienten ein Rezidiv dieser Krebsart auftritt. Daher wäre eine frühzeitige Erkennung derjenigen Patienten, die von den Standard-Erstlinientherapien nicht profitieren, von großer Bedeutung, da die Behandlungsverfahren stärker personalisiert werden könnten, um ein Rezidiv zu verhindern. Frühere Forschungen haben gezeigt, dass klinische Daten, Genexpressionsmarker und genetische Anomalien starke prognostische Faktoren für das Rezidiv verschiedener Krebsarten sind. In jüngster Zeit hat Deep Learning (DL) zunehmenden Erfolg bei der Vorhersage von Rezidiven aus histologischen Bildern gezeigt. Obwohl DL-Modelle hochpräzise Vorhersagen ermöglichen, werden sie oft wegen ihrer Blackbox-Eigenschaften kritisiert. Der aktuelle Stand der Forschung, wie z.B. Modelle, die Mehrfachinstanz-Lernen (Multiple Instance Learning), aufmerksamkeitsbasiertes Lernen (Attention-based Learning) und/oder Klassenaktivierungsmapping-Methoden (z.B. Grad-CAM) verwenden, zeigt jedoch, dass es möglich ist, die Modelle besser zu erklären. Aus biologischer Sicht wurde ebenfalls die Morphologie von Tumorzellen für verschiedene medizinische Vorhersage- und Klassifizierungsaufgaben verwendet, da sie entscheidend ist, um ein vollständiges Bild der Mikroumgebung des Tumors zu erhalten. Alle im Rahmen dieses Projekts durchgeführten Experimente beruhen auf histologischen Bildern von DLBCL. Zunächst wird gezeigt, dass morphologische Merkmale, die aus Zellkernen extrahiert wurden, starke Indikatoren zur Unterscheidung zwischen verschiedenen Zelltypen in DLBCL-Tumoren sind. Zweitens werden Zellkerne automatisch segmentiert und klassifiziert, um die Lücke nicht vorhandener Segmentierungsansätze für Lymphombilder zu schließen. Der letzte Teil befasst sich mit der Vorhersage eines Rezidivs von DLBCL und kombiniert die Methoden und Erkenntnisse der vorherigen Ansätze mit aufmerksamkeitsbasierten Mechanismen von DL-Modellen, um die Vorhersageentscheidung des Modells biologisch zu interpretieren. Dadurch war es möglich, nicht nur rezidivierte von nicht-rezidivierten Patienten zu unterscheiden, sondern auch zu erklären, welche morphologischen Merkmale der Zellkerne für die Vorhersage ausschlaggebend waren. Dies ermöglicht sowohl eine frühzeitige Erkennung von rezidivierten Patienten als auch das Auffinden distinkter morphologischer Muster zwischen beiden Interessengruppen.

Acknowledgements

Firstly, I would like to thank my supervisor Prof. Dr. Katarzyna Bożek, who has been a great PI throughout my whole PhD. Your scientific guidance helped me enormously to conduct my research. Additionally, on a more personal level, you always made sure that all members of our group feel comfortable and welcome which has contributed significantly to our great lab atmosphere.

Thank you also to my colleagues, who I do not only consider colleagues but also friends. You always supported and helped me and each other with all sorts of challenges, from organizational tasks to scientific questions. I could not have wished for better group and I will hold my time as a PhD student in this group in great memory, thanks to you all.

To my friends outside of work: thank you for always listening to my worries and complaints (never about work of course) and significantly contributing to reducing my stress level, especially when I had deadlines.

I want to thank my brothers and parents for always supporting me in any kind of matter and providing me with everything I need (especially delicious arabic food).

Lastly, it is also very important for me to mention, that as somebody who wasn't born in Germany, I don't take it for granted that I am allowed to live and work in a country without war. Especially when watching the news nowadays, I feel extremely privileged and grateful to be able to live here in peace while so many other people sadly can't. This is where a huge chunk of my motivation to pursue this PhD comes from. Not only my PhD but basically everything I pursue has the intention that I want to give something back even if the impact is minor. Being a citizen here is a debt I will never be able to repay and I will forever be grateful for this.

”If everything you try works, you aren’t trying hard enough.”

– Gordon Moore

Abbreviations

AJI aggregated Jaccard index

AUC area under the curve

CLAM clustering constrained attention multiple instance learning

CNN convolutional neural network

DL deep learning

DLBCL diffuse large B-cell lymphoma

H&E hematoxylin and eosin

IHC immunohistochemistry

MIL multiple instance learning

ML machine learning

NHL non-Hodgkin lymphoma

PQ panoptic quality

SOTA state-of-the-art

TCGA The Cancer Genome Atlas

TMA tissue microarray

WSI whole slide image

Publication Preface

The contributions presented in this thesis are based on previous publications and manuscripts.

Publications:

[31] Hussein Naji, Lunas Hahn, Katarzyna Bożek. 2023. **Lymphoma Cell Nuclei Classification using Color and Morphology Features**. *Current Directions in Biomedical Engineering*. 9(1):210-212. doi: 10.1515/cdbme-2023-1053.

[32] Hussein Naji, Lucas Sancere, Adrian Simon, Reinhard Büttner, Marie-Lisa Eich, Philipp Lohneis, Katarzyna Bożek. 2024. **HoLy-Net: Segmentation of histological images of diffuse large B-cell lymphoma.** *Computers in Biology and Medicine*. 170:107978. doi: 10.1016/j.compbiomed.2024.107978.

Manuscript under review:

[33] Hussein Naji, Juan I. Pisula, Stefano Ugliano, Adrian Simon, Reinhard Büttner, Katarzyna Bożek. 2024. **Deep learning-based interpretable prediction of recurrence of diffuse large B-cell lymphoma**.

Figures and tables displayed in this thesis (including all publications and manuscripts) that have no reference were designed by the author. Relevant code that was implemented for conducting experiments for the included publications and manuscripts can be found at: https://github.com/najihu94/dlbcl_project.

Contribution Statement

Hussein Naji is the main author of all of the publications and manuscripts included in this thesis. As the lead author, he took the main responsibility for design, implementation, data collection and analysis, and publishing the results in peer-reviewed venues. His and the co-authors' contributions to the included publications/manuscripts is described next using the Contributor Roles Taxonomy (CRediT):

[31] **Hussein Naji**: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Lunas Hahn**: Formal Analysis, Investigation, Methodology, Software. **Katarzyna Bożek**: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

[32] **Hussein Naji**: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Lucas Sancere**: Methodology, Software. **Adrian Simon**: Data curation. **Reinhard Büttner**: Data curation, Funding acquisition, Project administration. **Marie-Lisa Eich**: Data curation. **Philipp Lohneis**: Data curation. **Katarzyna Bożek**: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

[33] **Hussein Naji**: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Juan I. Pisula**: Methodology, Writing – review & editing. **Stefano Ugliano**: Visualization. **Adrian Simon**: Data curation. **Reinhard Büttner**: Data curation, Funding acquisition, Project administration, Writing – review & editing. **Katarzyna Bożek**: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Contents

Abbreviations **ix**

Publication Preface **xi**

Contribution Statement **xiii**

1 Introduction **1**
1.1 Diffuse Large B-Cell Lymphoma 1
1.2 Histopathology 1
1.3 Deep Learning in Histopathology 6
1.4 Outline 7

2 Lymphoma Cell Nuclei Classification using Color and Morphology Features **9**

3 Segmentation of Histological Images of Diffuse Large B-Cell Lymphoma **15**

4 Deep Learning-based Interpretable Prediction of Recurrence of Diffuse Large B-Cell Lymphoma **23**

5 Discussion **35**
5.1 Lymphoma Cell Nuclei Classification using Color and Morphology Features 35
5.2 Segmentation of Histological Images of Diffuse Large B-Cell Lymphoma . . 36
5.3 Deep Learning-based Interpretable Prediction of Recurrence of Diffuse Large B-Cell Lymphoma 36
5.4 Conclusion and Outlook 37

Bibliography **41**

A Supplementary Material for Chapter 3 **A.1**

List of Tables

A.1 Configuration of Mask R-CNN model. A.1
A.2 Configuration of HoLy-Net model. A.1

List of Figures

1.1 Sample preparation process 3
1.2 Example images of H&E- and IHC-stained tissue 4
1.3 Example of a WSI 5
1.4 Example of a TMA 5

Chapter 1

Introduction

This thesis addresses the hypothesis that human tissue samples have characteristics that are predictive of recurrence of diffuse large B-cell lymphoma (DLBCL). This chapter provides an overview of the medical background of DLBCL (section 1.1). Then, the role of histopathology in the context of medical research, including data acquisition, is described (section 1.2). Subsequently, current trends and challenges in the application of deep learning in histopathology are presented (section 1.3). Lastly, an outline of the main contributions of this thesis is given (section 1.4).

1.1 Diffuse Large B-Cell Lymphoma

Until today, cancer presents one of the most serious medical challenges and is the second leading cause of death globally, accounting for approximately 10 million deaths every year [1]. DLBCL is a type of lymphoid neoplasy and its treatment presents a significant challenge due to its highly aggressive and heterogeneous nature. It is the most common non-Hodgkin lymphoma (NHL) in adults, encompassing approximately 30-40% of all NHLs. The aggressive nature of DLBCL causes a fast progression which necessitates immediate and intensive treatment protocols. Despite advancements in high-dose chemotherapy with autologous stem cell support, long-term disease control remains challenging as up to 50% of patients suffer from recurrence of disease which significantly decreases the survival probability [2]. What makes recurrence so challenging is that recurred tumors tend to be more aggressive and often develop a resistance against therapies [3]. This highlights the necessity of more effective diagnostic and treatment strategies. Predicting recurrence, which is the main goal of this project, is not just a matter of extending survival but enhancing the quality of life for patients that do not benefit from standard frontline regimen.

1.2 Histopathology

Histopathology plays a crucial role in the diagnosis and study of diseases of tissues, including cancers like DLBCL. It typically involves the examination of patient biopsies by a pathologist after it has been processed, which allows for the detailed visualization of the tissue's composition on cell-level. The process from harvesting of the tissue to the digitization of high resolution images consists of the following steps (Fig.1.1):

1. **tissue harvesting**: A biopsy is carried out in which a small piece of tissue within the tumor region is cut out from the patient.
2. **fixation**: The tissue sample is fixed in a solution (usually formalin) to preserve the tissue's cellular structure and prevent degradation.

3. **dehydration**: The fixed tissue sample is dehydrated through a series of alcohol baths.

4. **embedding**: The processed tissue sample is embedded in a block of paraffin wax.

5. **sectioning**: A microtome is used to cut the paraffin block into thin slides (typically 4-5 µm). The slides are then placed onto a hotplate to ensure a uniform spreading of the sample. After the heating process, the residual liquid is removed. The slides are then dried at room temperature.

6. **deparaffinization**: Paraffin penetrated into the tissue is removed. Then, the slides are hydrated in a series of alcohol baths were the alcohol concentration is gradually reduced until only water is used.

7. **staining**: Staining increases contrasts in order to identify and differentiate the different components of the biological material. The most common staining technique is hematoxylin and eosin (H&E). Hematoxylin stains cell nuclei blue/purple, while eosin stains the cytoplasm and the surrounding tissue pink/red. This staining is critical for distinguishing between different cell types and structures. Another popular staining technique is immunohistochemistry (IHC) which is used to identify specific biomarkers.

8. **mounting**: After staining, the tissue sections are covered with a glass coverslip using a mounting medium, which protects the tissue and ensures adhesion.

9. **digitization**: Finally, the prepared slides are digitized using a slide scanner, creating high resolution images called whole slide images (WSIs).

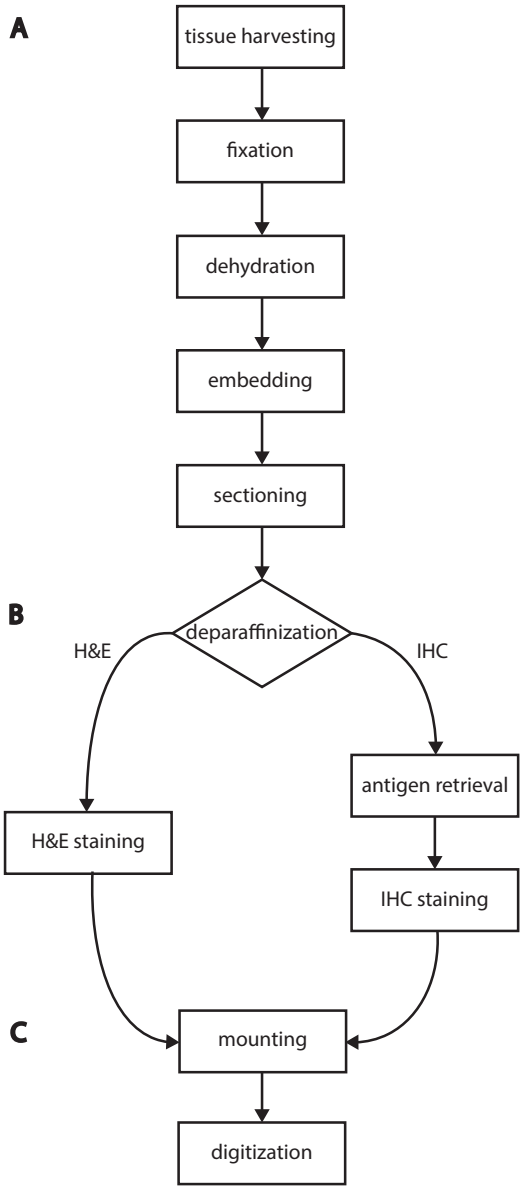
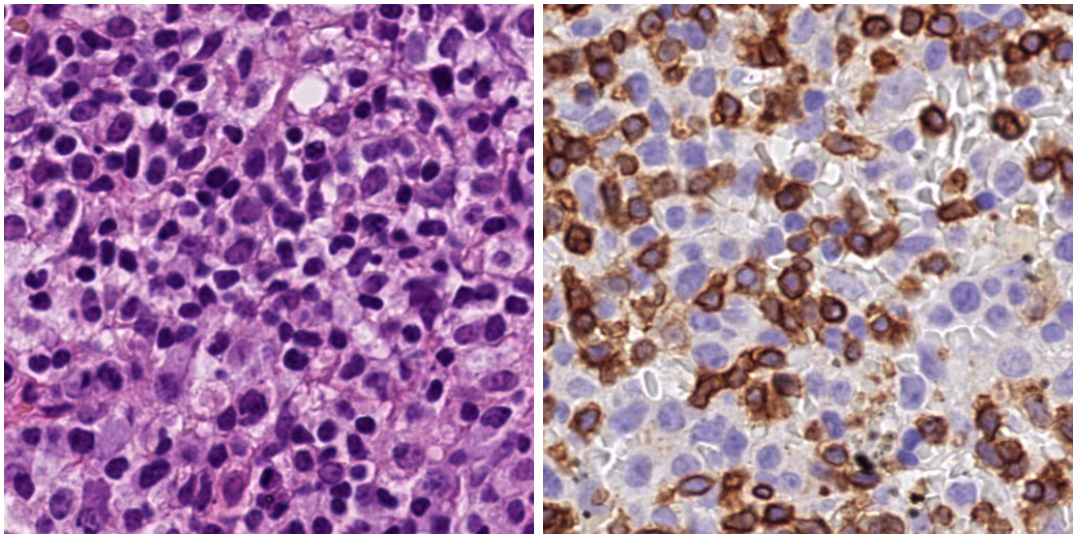


Fig. 1.1: Flowchart of the sample preparation process from harvesting of the tissue to the digitization to obtain WSIs.

Both, H&E and IHC staining, are widely used techniques and have unique purposes [4, 5]. H&E staining (Fig.1.2A) offers insights into the morphology and histology of a tissue sample. It is valuable for detecting abnormalities, such as tumors, inflammation, and necrosis [4, 6].
IHC staining (Fig.1.2B) highlights specific proteins of interest, allowing to determine the presence of tumor markers, diagnose tumors as benign or malignant, assess the tumor's stage, and identify specific cell types. It involves the usage of antibodies that bind to specific antigens in the tissue [7, 8]. When preparing IHC slides, the deparaffinization step is followed by an antigen retrieval, where the specific antigens are exposed to the tissue (Fig. 1.1B) [5].



A B

Fig. 1.2: Example images of H&E- and IHC- stained DLBCL tissue. The image sizes are 512 x 512 pixels (or 118 x 118 µm). (a) H&E-stained image. (b) IHC-stained image with CD3 marker. Brown cells indicate marker-positive cells.

WSIs (Fig.1.3) allow pathologists and researchers to examine slides in a virtual environment, offering certain advantages over traditional microscopy, such as easy access, the possibility to analyze and process the images, and remote collaborations. In the context of cancers like DLBCL, WSIs enable detailed analysis of the tumor microenvironment, providing insights into the spatial relationships and heterogeneity within the tumor [9].
A popular method to analyze multiple tissue samples simultaneously are tissue microarrays (TMAs) (Fig. 1.4). For acquiring TMAs, a puncher is used to extract samples from regions of interest of the original paraffin block, resulting in multiple tissue cores (with diameters of approximately 0.6 mm). These tissue cores are then inserted in a recipient paraffin block, assembled in an array. This enables simultaneous and high-throughput analysis, allowing to study multiple cases at once under identical experimental conditions. In cancer research, this procedure is of high value as it is well-suited for the evaluation of biomarkers, therapeutic targets, and understanding of disease mechanisms [10, 11].

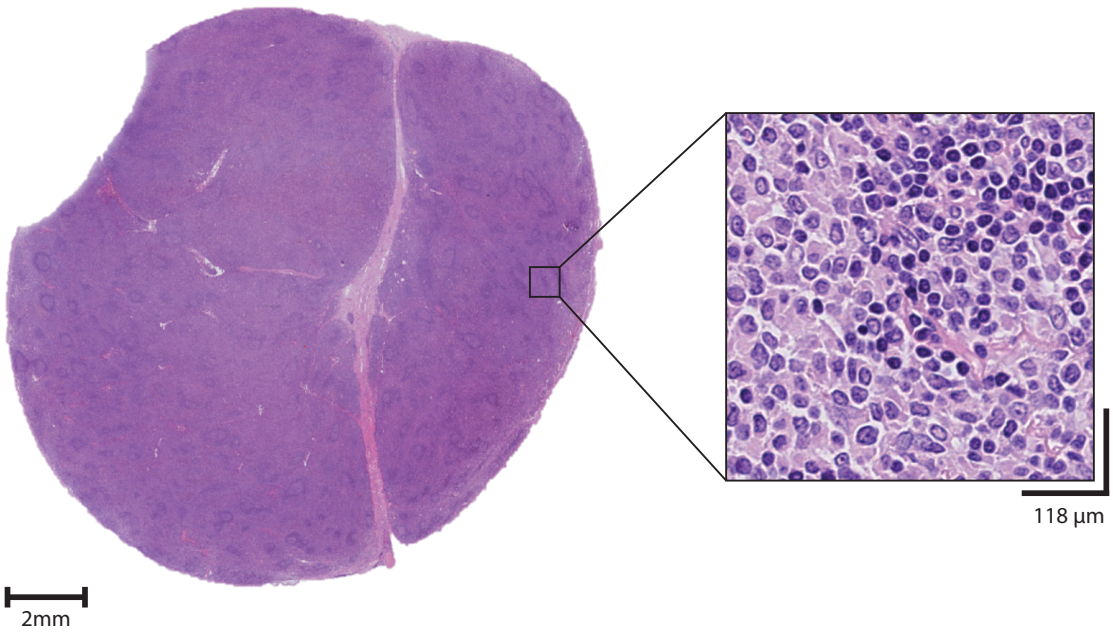


Fig. 1.3: Example of an H&E-stained WSI of DLBCL tissue.

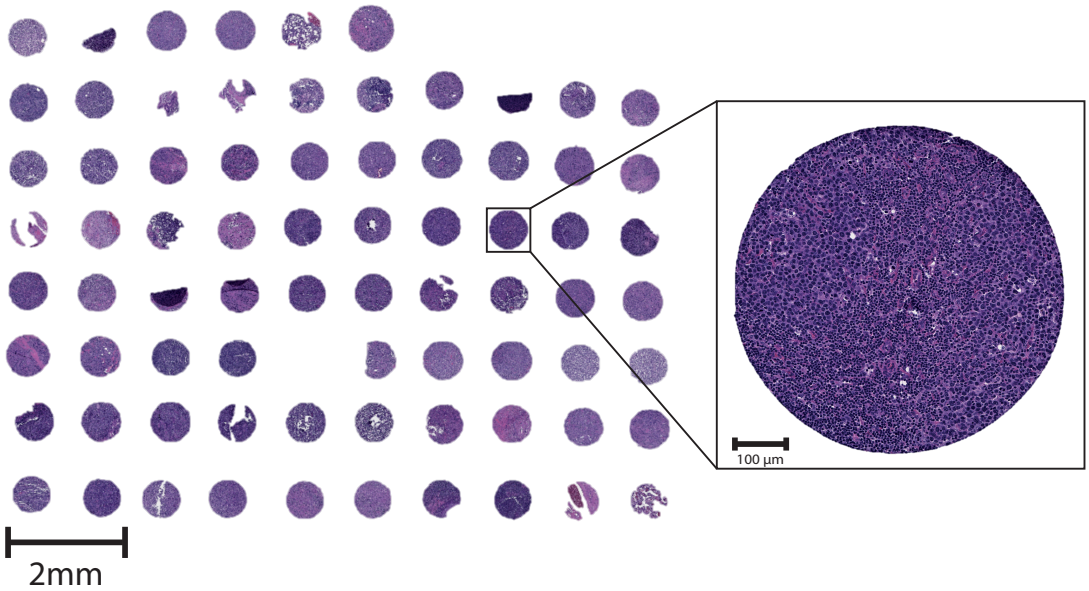


Fig. 1.4: Example of an H&E-stained TMA of DLBCL tissue.

1.3 Deep Learning in Histopathology

In the recent decades, deep learning (DL) has revolutionized various fields by leveraging convolutional neural network (CNN) architectures to automatically analyze and interpret large datasets as well as make meaningful predictions from them [12, 13, 14, 15, 16]. In histopathology, DL algorithms process and learn from vast amounts of image data, especially WSIs. This has shown significant successes in improving diagnostic accuracy, prognostication, and understanding of disease mechanisms. DL models handle the complexity and variability of histopathological images well and can be utilized to automatically detect distinct patterns and features within the WSIs that may not be easily detectable by human eyes, even for trained experts.

One widely-used application of DL in histopathology is the classification of cancers or subtypes of cancer, supporting pathologists in making more precise diagnoses [15, 17, 18, 19]. For instance, Korbar et al. [17] used a residual neural network [20] to classify colorectal polyps on WSIs, achieving an overall accuracy of 0.95. Coudray et al. utilized a CNN architecture (inception v3 [21]) to distinguish between adenocarcinoma, squamous cell carcinoma, and normal lung tissue, based on WSIs obtained from The Cancer Genome Atlas (TCGA)[1]. They report an average area under the curve (AUC) of 0.97.

Another prevalent use case is the segmentation of specific tissue components (e.g. tumor cell nuclei), which can be leveraged to analyze the tumor microenvironment [22, 23, 24, 25, 26]. A significant milestone in biomedical image segmentation was set in 2015 with the development of U-Net, a U-shaped CNN architecture with symmetric encoding and decoding paths connected by skip connections, enabling the precise localization and the gathering of contextual information within images [22]. It performed well for larger objects but struggled to correctly segment small dense objects as the model tends to merge instances that are in close proximity to each other. With the development of HoVer-Net, a network that predicts the horizontal and vertical distances of pixels within each instance to their respective center of mass (in addition to a binary mask and instance class), the borders of small objects such as nuclei could be captured much more precisely [23].

DL also has been used to predict treatment outcomes, including recurrence and survival rates [13, 27, 28, 29]. For instance, Shi et al. predicted early breast cancer recurrence using TMAs from the Carolina Breast Cancer Study[2]. A CNN was used to extract features from the images and then these features were used as an input for a support vector machine to predict recurrence. Their model yields a mean classification accuracy of 0.62 [28]. Xiao et al. developed a weakly supervised DL model to predict five-year recurrence-free survival in colorectal cancer. The model is based on a multiple instance learning (MIL) framework, that instead of receiving a set of individually labeled instances the network receives a set of labeled bags, each containing many instances. This approach is especially useful for WSIs, since the model is able to capture the heterogeneous tissue structures within WSIs and it can help to locate areas within WSIs that are indicative for the prediction. They report AUCs of 0.83 and 0.72 on the validation cohort and external test cohort, respectively [29]. Using histopathological images as input allows DL models to identify certain morphological patterns associated with higher risks of recurrence, which may not be apparent in traditional histopathological analysis. This includes recognizing variations in cell shape, size, and organization (cell morphology). As more data becomes available, DL models can be further trained and refined to improve their performance. This ensures

[1] https://www.cancer.gov/ccg/research/genome-sequencing/tcga, last access: 28.04.2024
[2] https://unclineberger.org/cbcs, last access: 28.04.2024

that the models remain relevant and accurate over time, adapting to new discoveries and treatment outcomes.
All of the before-mentioned applications demonstrate DLs tremendous potential for clinical practices. By providing more accurate and personalized risk assessments, DL can guide treatment decisions, identify patients who may benefit from more aggressive or novel therapies, and ultimately improve treatment outcomes. The ongoing development and validation of DL models in this area represent a promising avenue for research and application in personalized medicine. This work aims to make use of DL's potential in histopathology by using different architectures to analyze and process histopathological images and to predict recurrence of DLBCL.

1.4 Outline

The scientific contribution of this thesis encompasses three milestones. Firstly, features characterizing the morphology of cell nuclei of H&E- stained images of DLBCL are determined and used as an input for machine learning (ML) models to identify different cell types. A cluster analysis is applied and the feature importances are determined to show that the morphology of cells is a significant indicator for medical classification tasks (chapter 2).
Previous research predominantly focused on solid tumors, leaving a research gap for the analysis of cancers of the lymphatic system, such as DLBCL. Up to this day, there is no study, that emphasizes the segmentation of cell nuclei of lymphoma tissue. Furthermore, there is a lack of segmentation approaches for IHC-stained WSIs. Thus, in the second step, DL models are utilized to segment and classify cell nuclei in H&E- and IHC-stained WSIs of DLBCL. The models are trained to distinguish tumor and non-tumor cells in H&E- and marker-positive and marker-negative cells in IHC- stained WSIs. This enables a slide-level analysis in which tumor areas are highlighted and patterns of tumor immune activation are shown (chapter 3).
So far, clinical features, gene expression markers, and genetic abnormalities were used to predict recurrence of cancers [30]. Methods for image-based predictions are sparse and there is no existing approach that utilizes histological images to predict recurrence of lymphomas. Terefore, as a last step, a pipeline to predict recurrence of DLBCL is developed based on images from H&E-stained TMAs. Through MIL and attention-based prediction, the identification of image regions that were significant for the prediction is enabled. Integrating the methods and knowledge gained from chapter 2 and 3 into the pipeline, the nuclei of cells in image regions with the highest attention scores for recurred and non-recurred patients are segmented, respectively. Relating thereto, morphological features from all segmented cell nuclei are determined and analyzed statistically to find patterns in both groups of interest, aiming to not only being able to predict recurrence but also to interpret what characteristics the DL model focuses on based on the cell morphologies (chapter 4).
Lastly, the main findings are summarized and put into context, including a discussion of the study limitations and future directions (chapter 5).

Chapter 2

Lymphoma Cell Nuclei Classification using Color and Morphology Features

The following study [31] underlines the high information content of cell morphology with regard to the classification of different cell types, which enables the identification of distinct morphological patterns for recurred and non-recurred patients, respectively.

Hussein Naji*, Lunas Hahn*, and Katarzyna Bozek

Lymphoma Cell Nuclei Classification using Color and Morphology Features

https://doi.org/10.1515/

Abstract: Diffuse large B-cell lymphoma (DLBCL) is the most common subtype of Non-Hodgkin Lymphoma, presenting a great challenge for treatment due to its highly heterogeneous nature. DLBCL is diagnosed based on microscopy images of patient tissue samples. To help gain a better understanding of DLBCL, we developed an automated computer vision method to analyze morphological and color-based information within patient biopsies. We analyzed a dataset of whole slide images of DLBCL by segmenting individual cells and representing cell morphologies through a set of engineered features. The features were evaluated using a variety of visualization and machine learning (ML) classification techniques. Current state-of-the-art deep learning methods use images as the input in classification tasks achieving high performance but lacking in interpretability. A big challenge lies in finding out what features the pixel-based deep learning methods utilize in prediction. Here, we present a technique that not only yields high prediction accuracy but also provides insights into which of the features are key for prediction. We show that the color-based features have the highest importance for cell classification, allowing for the accurate identification of various cell types with an accuracy of 84% in a multi-class and 91% in a binary classification. Our results provide valuable insights for exploring cell image datasets to gain an in-depth view of the tumor microenvironment.

Keywords: nuclei classification, machine learning, feature engineering

1 Introduction

Up to this day, cancer remains a challenging disease causing millions of deaths yearly. Lymphomas are cancers that develop in the lymphatic system and are classified into Hodgkin lymphomas (HL) and non-Hodgkin lymphomas (NHL). HLs are rare and highly curable, while NHLs are more unpredictable and heterogeneous, with diffuse large B-cell lymphoma (DLBCL) being the most common subtype. The tumor microenvironment - the tissue around and within the tumor consisting of various types of immune cells, molecules, fibroblasts, and blood vessels - plays an important role in disease progression and treatment response [1]. A better understanding of the tumor environment could allow to develop new, more efficient treatment options. Currently, state-of-the-art deep learning methods show high performance in image-based prediction tasks using pixels as input. However, such models act as black boxes and provide limited insights into their decision-making. This black box problem is one of the most prevalent challenges in ML and necessitates tools to make ML more explainable. Especially in the field of medicine, interpretation of the prediction process is important for increasing the trustworthiness of the automated methods and improving our understanding of the disease.

Here, we propose an approach offering increased explainability via interpretable morphological features of cells. We analyze diagnostic slides of DLBCL patient biopsies. These biopsies were stained with hematoxylin and eosin (H&E). Hematoxylin is a chemical compound that is used to highlight cell nuclei, whereas eosin highlights the surrounding tissue. The stained samples were digitized into whole slide images (WSIs). Our dataset consists of 6,273 cell nuclei whose contours and cell types were labeled. Cell types include lymphoma (tumor), lymphocyte, leukocyte, stroma, and epithelial cells. We quantified features of these cell nuclei and utilized them to predict the cell type for each nucleus.

Quantifying cell morphology has been emphasized by previous studies [2–5]. Features such as cell nuclei area, perimeter, eccentricity, aspect ratio, etc. were used to differentiate various cell types, e.g. centroblasts and non-centroblasts [2], to classify five different types of white blood cells [3], cells in different stages of cell cycle [4], or to explore the cell morphology via clustering [5].

We demonstrate a classification accuracy of 91% in a binary classification of tumor and non-tumor cells and 84% in classification of all five cell types. We additionally inspect our features to find the key ones of our model's decision-making. We found that especially the color-based features proved to be of high importance.

***Corresponding author: Hussein Naji, Lunas Hahn,** Institute for Biomedical Informatics, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany, e-mail: hnaji@uni-koeln.de

Katarzyna Bozek, Institute for Biomedical Informatics, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany

Tab. 1: Number of annotated cell nuclei per cell type.

cell type	number of annotated nuclei
epithelial	109
leukocyte	826
lymphocyte	2,310
stroma	111
lymphoma (tumor)	3,764

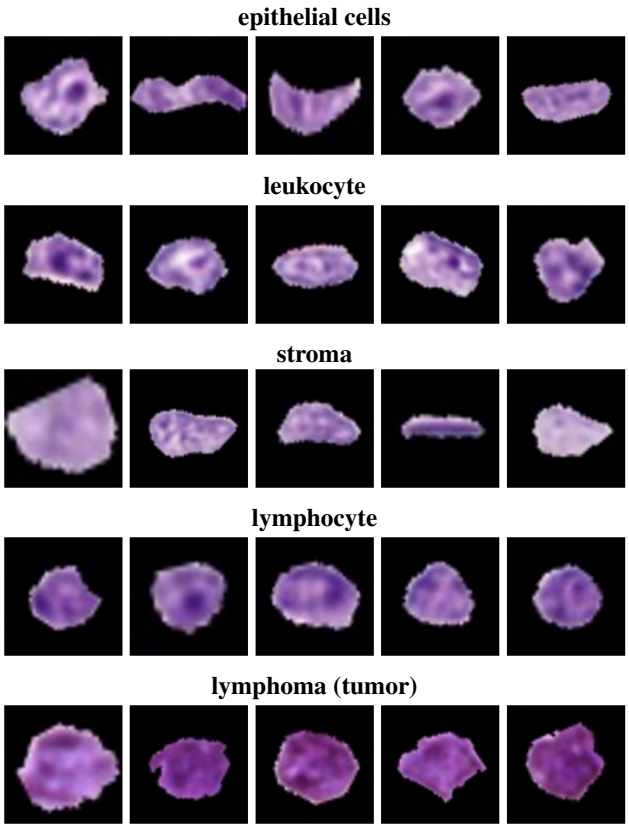


Fig. 1: Sample images of cell nuclei for each cell type.

2 Materials and Methods

2.1 Dataset

Due to the lack of publicly available annotated datasets of DLBCL, we generated a labeled dataset consisting of 40 image tiles of size 512 x 512 pixels at 40x magnification. These tiles were cropped from H&E-stained WSIs from eight different DLBCL cases. For each tile, two trained pathologists labeled the data, i.e. they drew contours around the cell nuclei and assigned class labels to each contour. Using the contours, we cut out each nucleus and aligned them so that their maximum diameter is aligned horizontally. Since the number of contour points varied among cells, we used interpolation to obtain 50 contour points for every nucleus. Furthermore, we masked the background - i.e. the surrounding tissue - leaving the nuclei as foreground in the images. Lastly, we rescaled the images to a uniform size of 32 x 32 pixels. Table 1 shows the composition of our dataset and Fig.1 shows some examples of cell nuclei for each cell type.

2.2 Feature Engineering

We calculated morphological features of each cell nuclei based on their contours and color-based features based on the cut-out nuclei images. To calculate the color-based features, we used the original RGB channels of the images and additionally applied a color deconvolution technique to obtain the color channels of the H&E staining [6]. In total, we use 26 features. In the following, we list all of the features that we calculated [7–9]:

- **area**: area of the nucleus.
- **perimeter**: perimeter of the nucleus.
- **radius**: radius of the nucleus.
- **aspect ratio**: ratio of width and height of the bounding box around the nucleus.
- **extent**: ratio of the area to the area in the bounding box around the nucleus.
- **solidity**: ratio of pixels in the nucleus to pixels of the convex hull image (smallest convex polygon that encloses the nucleus).
- **equivalent diameter**: diameter of the circle whose area is the same as the nucleus area.
- **eccentricity**: ratio of the focal distance (distance between focal points) to the major axis length of the nucleus. The value is in the interval $[0, 1)$ and the closer it is to 0 the more the ellipse approximates a circle.
- **curvature**: degree by which a curve deviates from being a straight line. The curvature is calculated as the reciprocal of the radius of a circle fitted at each contour point. We calculated the mean and standard deviation of the curvature.
- **roughness**: variance in the length of a vector that is centered at the geometric centroid of a nucleus as it rotates along with each boundary point.
- **pixel value**: color pixel value. We calculated the mean, minimum (min), and maximum (max) value for the r-, g-, and b-channel of the RGB images, and for the h- and e-channel after color deconvolution.

2.3 Cell Nuclei Classification

We utilized the random forest algorithm [10] to predict cell type based on the shape and color features and denote the model CoMorphLy (color and morphology feature-based cell

type prediction for lymphoma images). We used 80% of images for training the model and 20% for final evaluation. Furthermore, we used 20% of the training set for validation to find the best set of hyperparameters. For the test set, we randomly selected 20% of images of each cell type respectively to guarantee an equal proportional distribution of classes. We trained two separate models. The first one was trained to classify cells into all five cell types (Table 1). The second one was trained to perform binary classification to distinguish between the classes tumor and non-tumor, where all cell types except lymphoma are labeled as the class non-tumor.

We further analyzed our model feature's impact on the classification by determining the permutation feature importance which describes the difference between the classification accuracy for one feature and the accuracy when the entries of that feature are randomly shuffled [10].

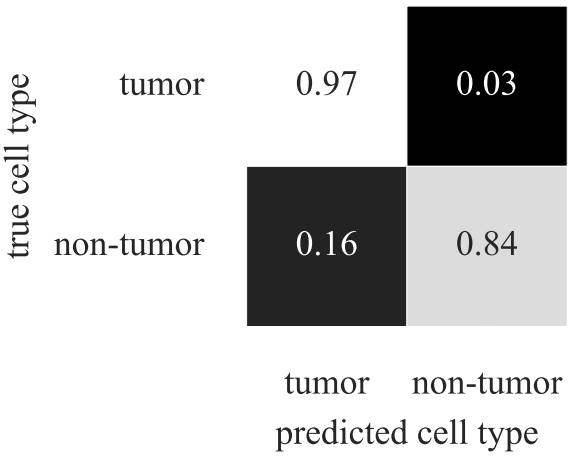


Fig. 3: Confusion matrix of binary prediction.

3 Results

In the multi-class prediction, CoMorphLy yields an accuracy of 84%, whereas in the binary prediction 91%. Fig.2 and Fig.3 show the confusion matrices of both models.

The feature importances are shown in Fig.4. The feature *max pixel value hematoxylin* has the highest importance of 0.1 and eight features have an importance of 0.

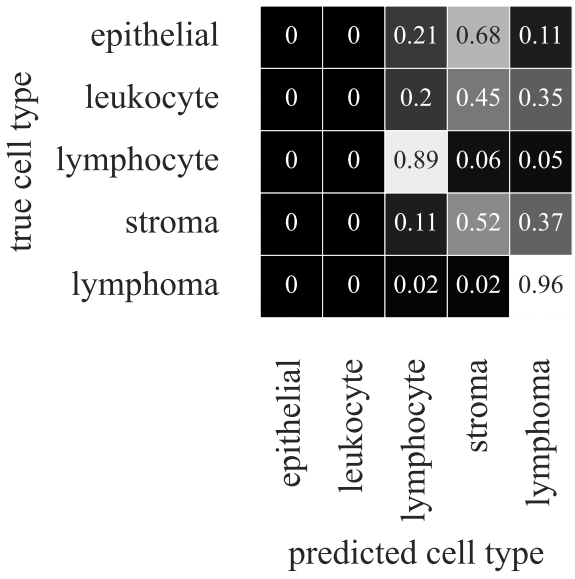


Fig. 2: Confusion matrix of multi-class prediction.

4 Discussion and Conclusion

Fig.2 shows that 96% of lymphoma cells and 89% of lymphocytes are classified correctly. On the other hand, epithelial cells and leukocytes are incorrectly classified. In the case of epithelial cells, most of them (68%) are classified as stroma whereas leukocytes are more evenly distributed within the lymphocyte, stroma, and lymphoma classes. Indeed, epithelial and stroma cells resemble each other in shape (Fig.1) since both cell types show higher degrees of non-circularity. Another important source of errors is the class imbalance. CoMorphLy achieves high performance in predicting the classes lymphocytes and lymphomas which make up almost 97% of all samples in the dataset. Conversely, the model results in low performance for the underrepresented cell types. This indicates that more samples of the underrepresented classes might improve the accuracy of classification of cells of these classes.

Among all the features, color-based features show the highest importance for prediction. The maximum value of the hematoxylin color channel is the most predictive of the cell class. The primary purpose of hematoxylin is to visually highlight the cell nuclei and their information content which might be the key difference between tumor cells and other cell types.

Visualizing the feature space using a principle component analysis (PCA) (Fig.5) supports the conclusion that our features are highly descriptive of the cell nuclei. Different cell types form clusters, especially the overrepresented ones such as lymphoma (tumor), lymphocyte, and leukocyte. Here, we also hypothesize that more samples of the underrepresented classes might allow a clearer distinction between these types of cells.

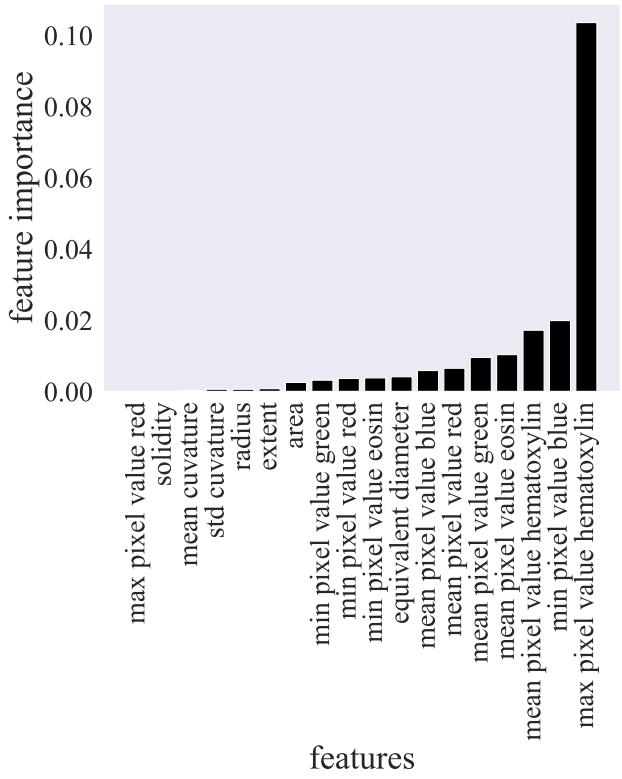


Fig. 4: Permutation feature importances of the CoMorphLy model. Features with an importance of 0 are excluded from this figure.

We showed that our model is able to correctly classify two types of cells that are very frequent in DLBCL images (lymphoma and lymphocytes) which allows to quantitatively describe tumor samples. Further studies should focus on annotation of additional cells of the underrepresented as well as other cell types to increase the inter-class generalizability of CoMorphLy. Combining CoMorphLy with accurate nuclei segmentation models allows for a quantitative description of the tumor microenvironment at a single cell level and potentially for linking such a description with disease characteristics or treatment outcomes.

Author Statement

Research funding: The author state no funding involved. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

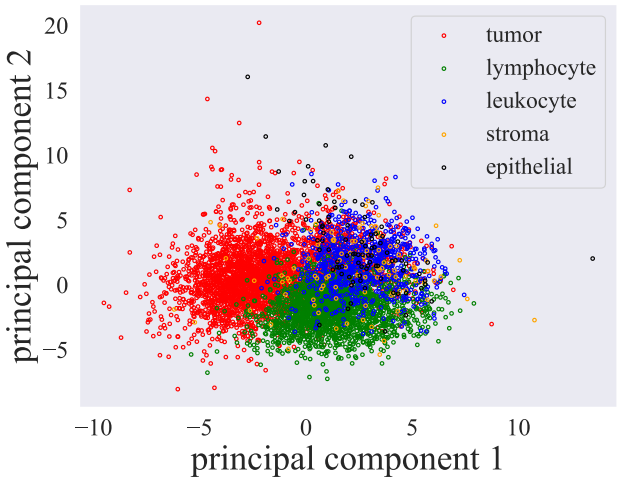


Fig. 5: PCA of our features.

References

[1] Sandra Susanibar-Adaniya, Stefan K. Barta. Update on Diffuse large B cell lymphoma: A review of current data and potential applications on risk stratification and management. American Journal of Hematology 2019, 58

[2] Michail, E., Dimitropoulos, K., Koletsa, T., Kostopoulos, I., Grammalidis, N. Morphological and textural analysis of centroblasts in low-thickness sliced tissue biopsies of follicular lymphoma. Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2014, 3374–3377

[3] A. Gautam and H. Bhadauria. Classification of white blood cells based on morphological features. International Conference on Advances in Computing, Communications and Informatics (ICACCI) 2014, 2363-2368

[4] Li, Y., Nowak, C.M., Pham, U., Nguyen, K., Bleris, L. Cell morphology-based machine learning models for human cell state classification. npj Syst Biol Appl 7:23, 2021

[5] Ilmari Ahonen, Ville Härmä, Hannu-Pekka Schukov, Matthias Nees, Jaakko Nevalainen. Morphological Clustering of Cell Cultures Based on Size, Shape, and Texture Features. Statistics in Biopharmaceutical Research 2016, 8:2, 217-228

[6] Ruifrok, A. C., Johnston, D. A. Quantification of histochemical staining by color deconvolution. Analytical and quantitative cytology and histology 2001, 23:4, 291–299

[7] Phillip, J.M., Han, KS., Chen, WC. et al. A robust unsupervised machine-learning method to quantify the morphological heterogeneity of cells and nuclei. Nat Protoc 2021, 16, 754–774

[8] Lenoir, A. Fast estimation of mean curvature on the surface of a 3D discrete object. International Conference on Discrete Geometry for Computer Imagery. Lecture Notes in Computer Science 1997, 1347.

[9] Takacs, G. Sucker-Rod Pumping Handbook: Production Engineering Fundamentals and Long-Stroke Rod Pumping, Gulf Professional Publishing 2015, 13-56

[10] Breiman, L. Random Forests. Machine Learning 2001, 455–32

Chapter 3

Segmentation of Histological Images of Diffuse Large B-Cell Lymphoma

In the following study [32], DL models are used to segment and classify cell nuclei in WSIs which is a crucial step to analyze the tumor microenvironment. These models can then be utilized to find differences in the tumor microenvironment of recurred and non-recurred patients.

Computers in Biology and Medicine 170 (2024) 107978

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

HoLy-Net: Segmentation of histological images of diffuse large B-cell lymphoma

Hussein Naji [a,b,*], Lucas Sancere [a,b], Adrian Simon [d], Reinhard Büttner [d], Marie-Lisa Eich [d], Philipp Lohneis [d,e], Katarzyna Bożek [a,c]

[a] Institute for Biomedical Informatics, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany
[b] Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany
[c] Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Germany
[d] Institute of Pathology, University Hospital Cologne, Germany
[e] Hämatopathologie Lübeck, Germany

ARTICLE INFO

Keywords:
Diffuse large B-Cell lymphoma
Nuclei segmentation
Immunohistochemistry
Hematoxylin and eosin
Whole slide image

ABSTRACT

Over the last years, there has been large progress in automated segmentation and classification methods in histological whole slide images (WSIs) stained with hematoxylin and eosin (H&E). Current state-of-the-art (SOTA) techniques are based on diverse datasets of H&E-stained WSIs of different types of predominantly solid cancer. However, there is a scarcity of methods and datasets enabling segmentation of tumors of the lymphatic system (lymphomas). Here, we propose a solution for segmentation of diffuse large B-cell lymphoma (DLBCL), the most common non-Hodgkin's lymphoma. Our method applies to both H&E-stained slides and to a broad range of markers stained with immunohistochemistry (IHC). While IHC staining is an important tool in cancer diagnosis and treatment decisions, there are few automated segmentation and classification methods for IHC-stained WSIs. To address the challenges of nuclei segmentation in H&E– and IHC-stained DLBCL images, we propose HoLy-Net - a HoVer-Net-based deep learning model for lymphoma segmentation. We train two different models, one for segmenting H&E– and one for IHC-stained images and compare the test results with the SOTA methods as well as with the original version of HoVer-Net. Subsequently, we segment patient WSIs and perform single cell-level analysis of different cell types to identify patient-specific tumor characteristics such as high level of immune infiltration. Our method outperforms general-purpose segmentation methods for H&E staining in lymphoma WSIs (with an F1 score of 0.899) and is also a unique automated method for IHC slide segmentation (with an F1 score of 0.913). With our solution, we provide a new dataset we denote LyNSeC (lymphoma nuclear segmentation and classification) containing 73,931 annotated cell nuclei from H&E and 87,316 from IHC slides. Our method and dataset open up new avenues for quantitative, large-scale studies of morphology and micro-environment of lymphomas overlooked by the current automated segmentation methods.

1. Introduction

Diffuse large B-cell lymphoma (DLBCL) is a tumor of B lymphocytes and is the most common non-Hodgkin's lymphoma (NHL) in adults, with a relative frequency of 30%–40% among all NHLs. DLBCL is a morphologically, biologically, and clinically an extraordinarily heterogeneous disease. Although there are treatment approaches based on high-dose chemotherapy with autologous stem cell support, long-term disease control is not always possible as up to 50% of all patients suffer from relapses after first-line treatment [1]. To better understand the biology of this highly heterogeneous disease and to improve its treatment, there is a pressing need for quantitative tools to describe the morphology, composition, and variability of DLBCLs.

The main source of information for DLBCL diagnosis is paraffin-embedded patient biopsies stained with hematoxylin and eosin (H&E). Furthermore, immunohistochemical (IHC) staining is used, which highlights specific proteins of interest in addition to cell nuclei, allowing to determine the presence of tumor markers, diagnose tumors as benign or malignant, assess the tumor's stage, and identify specific cell types [2]. There can be millions of cells per section which makes the diagnosis

* Corresponding author. Institute for Biomedical Informatics, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany.
E-mail address: hnaji@uni-koeln.de (H. Naji).

https://doi.org/10.1016/j.compbiomed.2024.107978
Received 13 July 2023; Received in revised form 30 October 2023; Accepted 8 January 2024
Available online 11 January 2024
0010-4825/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

time-consuming. Describing such a tumor sample comprehensively and quantitatively necessitates dedicated computational tools.

Automated segmentation of histology images [3,4], especially H&E slides [5–7], has improved dramatically over the recent years. Existing methods allow, not only to accurately determine nuclei contours but additionally, to correctly recognize several cell types in the tumor microenvironment. State-of-the-art (SOTA) H&E whole slide image (WSI) segmentation methods focus on a large variety of cancer types. Until now, however, few approaches exist for segmenting diagnostic slides of lymphomas and leukemias [8–10]. Several lymphoma-dedicated segmentation methods involve algorithmic solutions with various types of adaptive thresholding, watershed transform, and/or K-means clustering [9,10]. Although these methods correctly resolve cell areas, they fail to discern overlapping cells. Masoudi et al. [8] proposed a method to detect Hodgkin's lymphoma by identifying a cancer cell subtype named Reed-Sternberg. This algorithm aims to improve segmentation at the margins of overlapping cells based on stain intensity and size of the segmented region. However, neither of the multi-step algorithmic approaches fully resolves segmentation of overlapping cells in lymphoma.

The major focus of methods for segmentation of tumor microscopy images is H&E-stained images. Only a few methods allow to segment IHC-stained images, including algorithmic approaches [11,12] and those that combine segmentation with classification [11]. Oscanoa et al. [11] reported 93.1% accuracy in the classification of detected nuclei into marker-positive (cell where the target antigen is present in the cell and the antibody/marker used in the staining process binds to the antigen) and marker-negative (cell that does not show an expression of the target antigen). However, this metric includes correctly detected nuclei without taking into account the segmentation performance, which is an incomplete quality assessment, especially in cases of overlapping nuclei. Carreras et al. [13] quantified the immunohistochemical expression for prognosis assessment of DLBCL. The nuclei were segmented using Waikato Environment for Knowledge Analysis (Weka), a library that utilizes various machine learning algorithms on a set of selected image features to produce pixel-based segmentation [14]. The authors quantified the presence of PD-L1 only and did not perform evaluation of the segmentation performance.

Deep learning has opened up possibilities to improve IHC image segmentation. Swiderska-Chadaj et al. [15] compared four methods to detect lymphocytes in IHC images with two markers: CD3 and CD8. The best results were achieved by a U-Net-based architecture [16] with an F1-score of 0.78.

Mahanta et al. [17] developed IHC-Net which is based on a fully convolutional neural network for nuclei segmentation of different types of hormone receptor-positive cells in breast cancer. The method yields a dice score of 0.825 in segmentation. Both approaches exceeded previous SOTA methods regarding segmentation quality but did not address the question of overlapping nuclei. Furthermore, they focus on either one or two markers that have similar visual patterns, their performance on other markers has not been tested.

Here, we address the lack of methods for segmentation of lymph node tumors as well as for the IHC-stained samples by introducing HoLy-Net - a HoVer-Net-based [6] automated solution for DLBCL patient sample description. We adapt and optimize the network to automatically segment and classify lymphoma cell nuclei in H&E– and IHC-stained WSIs. To this end, we generated a comprehensively labeled dataset we denote LyNSeC (lymphoma nuclei segmentation and classification) including 152,795 annotated nuclei in both types of staining.

2. Methods

2.1. LyNSeC dataset

Due to the lack of publicly available annotated datasets of lymphoma, we generated a labeled dataset and denoted it LyNSeC (lymphoma nuclear segmentation and classification). LyNSeC comprises three subsets: LyNSeC 1 consisting of 379 IHC images of size 512 x 512 pixels at 40x magnification. In the images, we annotated the contours of each cell nuclei and the cell class: marker-positive or marker-negative.

In total, LyNSeC 1 contains 87,316 annotated cell nuclei of four different patients, with 48,171 of them assigned the class negative and 39,145 positive. We included three markers in this dataset showing visually different staining patterns: cluster of differentiation 3 (CD3), Ki67 as a marker of proliferation, and erythroblast transformation-specific (EST)-related gene (ERG) (Fig. 1). CD3 is commonly expressed on the surface of all T cells [18]. Ki67 marks cells in all phases of mitosis and, therefore, reflects the proliferation rate of tumors Schonk, [19]. ERG is a protein widely expressed in endothelial cells and various cancers, and allows to distinguish between several hematopoietic cells and lymphoma cells, like T-ALL (acute lymphoblastic leukemia) cells [20,21].

LyNSeC 2 and 3 contain H&E-stained images of 70 different patients (64 patients for LyNSeC 2 and 6 for LyNSeC 3). LyNSeC 2 consists of 281 images (the sample size for 63 patients is in the range of 1–7 images and one patient has a sample size of 16 images) and LyNSeC 3 of 40 images (the sample size for 39 patients is in the range of 2–6 images and one patient has a sample size of 20 images) of size 512 x 512 pixels at 40x magnification. 65,479 and 8,452 nuclei were annotated in LyNSeC 2 and 3, respectively. In LyNSeC 3, the nuclei were also assigned a class label (tumor and non-tumor). 3,747 nuclei were identified as tumors and 4,705 as non-tumors. In the annotation procedure, the contours of the H&E images (LyNSeC 2 and LyNSeC 3) were annotated by two pathologists and by two students (trained by the pathologists). Annotation of the cell classes in LyNSeC 3 was done by the pathologists only. LyNSeC 1 was annotated by the two students who were additionally trained to annotate the contours and to distinguish marker-positive and marker-negative cells. The pathologists inspected and (if necessary) adjusted the LyNSeC 3 annotations. In the following, we refer to LyNSeC 1 as $\text{LyNSeC}_{\text{IHC}}$ and to LyNSeC 2 and 3 as $\text{LyNSeC}_{\text{H\&E}}$. Our dataset is available at https://zenodo.org/record/8065174.

2.2. Preprocessing

We first down sampled and binarized the WSIs to determine the contour of the entire tissue section through step-wise morphology operations (thresholding, opening, and closing). Next, we split the foreground region into a grid of square-shaped image tiles of 512 x 512 pixels with an overlap of 100 pixels. The overlap between grid parts allows to resolve contours of cells on the image borders. Only tiles that contain at least 10% of foreground were kept for segmentation.

2.3. Segmentation models

Mask R-CNN. As in previous approaches [22–24], we utilized Mask R-CNN [25] – an instance segmentation approach originally designed for natural images – and adapted it to the problem of nuclei segmentation (we used and modified the detectron2 implementation [25]). The method predicts a binary mask together with a bounding box, delineating each of the segmented objects, and a class label. Our parametrization of Mask R-CNN can be found in the supplementary material. We found that changing the anchor size was key to enabling the model to detect small and dense objects such as cell nuclei.

HoLy-Net. Additionally, we introduce HoLy-Net: HoVer-Net-based network [6] for lymphoma segmentation. HoVer-Net is an instance segmentation network predicting horizontal and vertical distances of pixels within each nucleus to their respective center of mass. Subsequently, these distances are used to separate individual nuclei. A separate part of the network predicts the class of each cell. HoVer-Net's performance was evaluated on three labeled nuclei segmentation datasets: CoNSeP [6] which consists of image tiles that were extracted from 16 colorectal adenocarcinoma WSIs, Kumar [26] that includes cases of

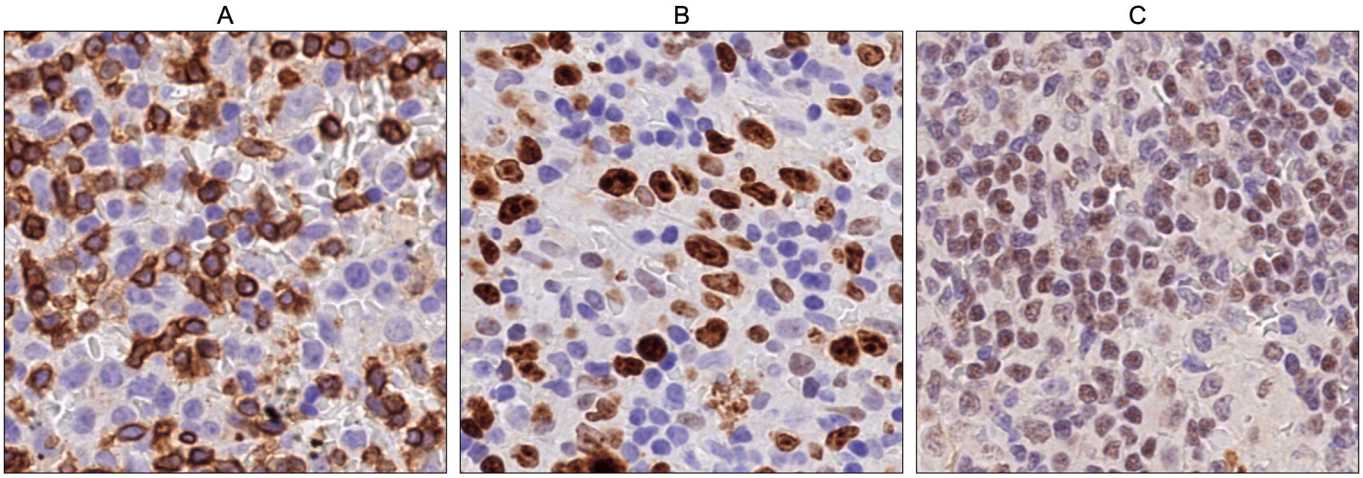


Fig. 1. Examples of IHC-stained DLBCL images with three different markers. (A) CD3, (B) Ki67, and (C) ERG.

breast, liver, kidney, prostate, bladder, colon and gastric cancer, and CPM [27], consisting of lung cancer, head and neck squamous cell carcinoma, glioblastoma multiforme, and lower grade glioma. None of these datasets contain any type of lymphoma. We adapted this architecture to segment lymphoma cell nuclei with several modifications. We added dropout after each residual unit of the down sampling branch. Additionally, we changed the default learning rate scheduler to a 'reduce on plateau' scheduler to reduce the learning rate only when the model stops improving for a certain number of epochs. The full configuration of the model is listed in the supplementary material. We varied the initial learning rate between values of 10^{-2} and 10^{-7}, with 10^{-4} showing to be the best fit. Lower learning rates lead to too minor improvements in the loss curve causing the model to stagnate or to get stuck in a local minimum, whereas high learning rates caused the model to diverge as the parameter adjustments overshot the optimal values. We used batch sizes from 2 to 64, with 4 showing the best results. Batch sizes that were too large resulted in poor generalization due to overfitting of the model. We trained the model for 100 epochs to achieve optimal convergence. Fewer epochs were insufficient and resulted in underfitting, whereas values over 100 didn't have any effect since the model had already converged. The model has 37,639,431 parameters.

We trained both, Mask R-CNN and HoLy-Net for the two separate tasks of H&E-stained and IHC-stained image segmentation. Both networks were trained twice, once on the $\text{LyNSeC}_{\text{H\&E}}$ dataset (Mask $\text{R-CNN}_{\text{H\&E}}$ and $\text{HoLy-Net}_{\text{H\&E}}$) and once on the $\text{LyNSeC}_{\text{IHC}}$ (Mask $\text{R-CNN}_{\text{IHC}}$ and $\text{HoLy-Net}_{\text{IHC}}$). We applied a 5-fold cross validation and for each fold, we used 80% of each dataset for training and 20% for testing. We used the models for both, nuclei segmentation and classification. Mask $\text{R-CNN}_{\text{H\&E}}$ and $\text{HoLy-Net}_{\text{H\&E}}$ training was performed in two steps. In the first step, a binary segmentation model was trained on LyNSeC 2. In the second step, the pre-trained network was fine-tuned using LyNSeC 3 to segment and classify the cells as tumor and non-tumor.

2.4. Evaluation metrics

To evaluate our models' segmentation performance, we used four metrics: F1 score, dice score, aggregated Jaccard index (AJI), and panoptic quality (PQ). The first three metrics measure the segmentation accuracy and tend to over-penalize overlapping regions (in our case overlap of different cell nuclei) which is why we included PQ to obtain a more complete assessment of the overall segmentation quality. PQ combines the segmentation quality, which evaluates how closely matched instances are with their ground truths (instance segmentation), with the detection quality that assess how good the model detects objects (semantic segmentation) [28].

We further used the F_c^t score (*c* stands for classification task and *t* refers to a specific class type) [6] to assess the classification performance of our models. It uses a unique matching that splits all available instances into matched pairs (true positives), unmatched ground truth instances (false negatives) and unmatched prediction instances (false positives) which allows for a joint quality assessment of segmentation and cell type classification performance.

3. Results

3.1. Segmentation performance

The accuracy of our segmentation methods on the LyNSeC dataset is shown in Table 1 and examples of our results in Fig. 2. The segmentation results of HoLy-Net surpass those of Mask R-CNN in almost all metrics. The largest differences between the methods are in the AJI, PQ, and F_c^t metric. With a PQ of 0.765 and classification accuracy of 0.924 the results of HoLy-Net on H&E images match current SOTA results in other tumor types [6].

3.2. Comparison to SOTA segmentation methods

We further compared HoLy-Net to three baseline HoVer-Net models. These baseline models were trained on the CPM, Kumar, and CoNSeP datasets, respectively. Since our dataset contains different class labels than the baseline datasets, we only compare segmentation performance (Fig. 3). Furthermore, we compared HoLy-Net trained on our IHC dataset to the SOTA method for IHC nuclei segmentation based on U-Net [15].

For $\text{LyNSeC}_{\text{H\&E}}$, Holy-Net outperforms the best baseline model ($\text{HoVer-Net}_{\text{CPM}}$) by 0.05 in dice coefficient, 0.086 in AJI, and 0.074 in the PQ metric (Fig. 3A). These results underline the lack of generalizability of the current SOTA methods to images of lymphoma. For $\text{LyNSeC}_{\text{IHC}}$, HoLy-Net outperforms the baseline model by 0.069 in the dice coefficient, 0.257 in AJI, and 0.326 in the PQ metric (Fig. 3B). The high value of the dice metric and low values of AJI and PQ of the U-Net-based IHC segmentation model [16] indicate that the model indeed correctly segments nuclei areas but lacks the ability to separate overlapping instances. The trained models $\text{HoLy-Net}_{\text{H\&E}}$ and $\text{HoLy-Net}_{\text{IHC}}$ are available at https://zenodo.org/record/8065174.

3.3. Single-cell level analysis of lymphoma WSIs

Lastly, we applied HoLy-Net across several WSIs of DLBCL with different functional markers in the IHC stainings. These markers were not part of the training set but, nevertheless, appear as correctly segmented (Fig. 4).

Table 1
Segmentation and classification results on our datasets. accc is the classification accuracy, Fctu, Fcnt, Fc p, and Fcn are the respective F_c^t scores for the classes *tumor*, *non-tumor*, *positive*, and *negative*. In the column dataset, on which the models were trained and evaluated, H&E refers to $LyNSeC_{H\&E}$ and IHC to $LyNSeC_{IHC}$. The values indicate the results of a 5-fold cross validation (mean ± standard deviation).

Dataset	Model	Dice	F1	AJI	PQ	Accc	$^{tu}_{Fc}$	$^{nt}_{Fc}$	$^{p}_{Fc}$	$^{n}_{Fc}$
H&E	HoLy-Net	**0.892 ± 0.008**	**0.876 ± 0.016**	**0.749 ± 0.035**	**0.755 ± 0.023**	**0.923 ± 0.041**	**0.820 ± 0.164**	**0.645 ± 0.182**	–	–
	Mask R-CNN	0.747 ± 0.012	0.720 ± 0.025	0.525 ± 0.241	0.608 ± 0.127	0.851 ± 0.036	0.659 ± 0.283	0.448 ± 0.344	–	–
IHC	HoLy-Net	**0.864 ± 0.019**	**0.895 ± 0.019**	**0.732 ± 0.026**	**0.720 ± 0.024**	0.964 ± 0.012	–	–	**0.846 ± 0.035**	**0.825 ± 0.029**
	Mask R-CNN	0.769 ± 0.052	0.822 ± 0.016	0.614 ± 0.094	0.646 ± 0.087	**0.971 ± 0.011**	–	–	0.835 ± 0.061	0.714 ± 0.043

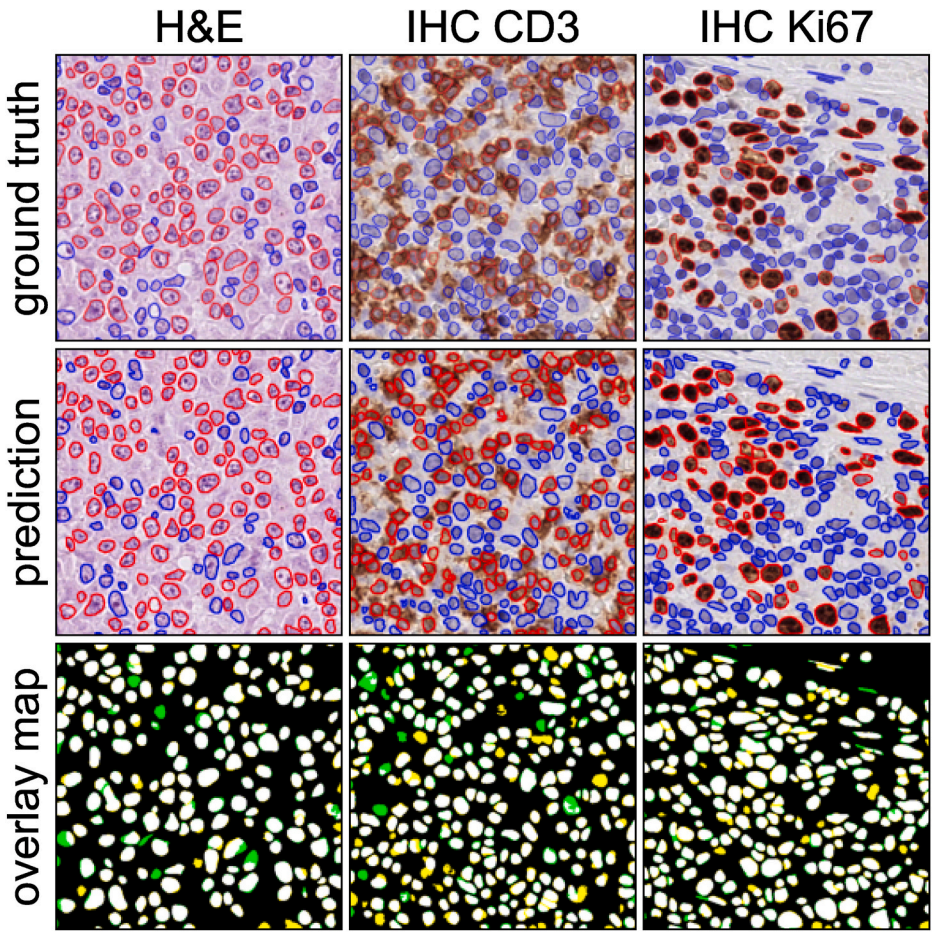


Fig. 2. Segmentation and classification results on the LyNSeC datasets. The red contours surround tumor cells in the H&E images and marker-positive cell in the IHC images, whereas blue indicates the non-tumor and marker-negative cells in the H&E and IHC images, respectively. The images in the third row show the overlay of segmentation maps of the respective ground truth labeling and predictions. White color indicates correctly classified foreground nuclei pixels (true positive), yellow color background pixels that were classified as foreground (false positives) and green color foreground pixels that were classified as background (false negatives). For these examples, the mean dice, F1, and AJI scores are 0.881, 0.882, and 0.789.

Fig. 5 depicts two lymphoma samples whose consecutive sections were stained with H&E and a panel of immune cell and immune activation markers. The heatmaps illustrate the numbers of tumor cells (H&E) or marker-positive cells (IHC) in each image region. The two cases show several prominent differences. Case 1 is highly infiltrated by T cells, whereas case 2 appears T cell-free, indicative of low immune activation in this tumor and its low visibility to the immune system of a patient. Inversely, case 2 shows high proliferation rates compared to the much lower density of Ki67-positive cells in case 1. Importantly, these results are based on single cell counts, providing quantitative and broad insights into the two lymphoma cases.

4. Discussion

Automated segmentation of diagnostic microscopy images of cancer presents tremendous opportunities for quantitative studies of the tumor microenvironment, cancer morphology, immune infiltration, and activation. In a tumorous lymph node tissue there is no clear structure, the tissue is composed of very densely packed immune cells of various types. This cell density as well as different cell type compositions requires development of dedicated methods for their correct segmentation.

In this work, we expand the array of tumor types, that can be automatically segmented, with lymphoma – a highly prevalent cancer type. Lymphoma visually differs from solid tumors in that it is more dense and

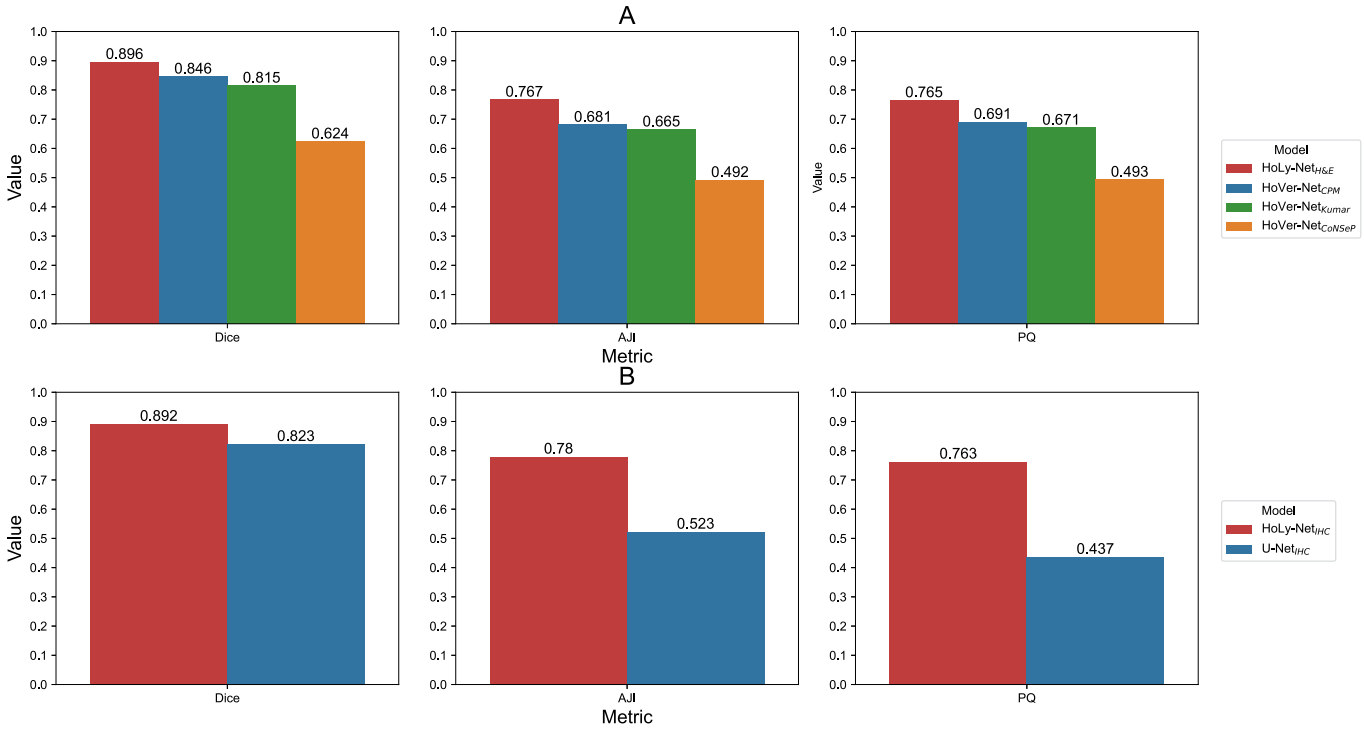


Fig. 3. Comparison of segmentation results to baseline and SOTA methods. (A) H&E images. (B) IHC images. U-Net$_{IHC}$ indicates a U-Net model trained on LyNSeC$_{IHC}$.

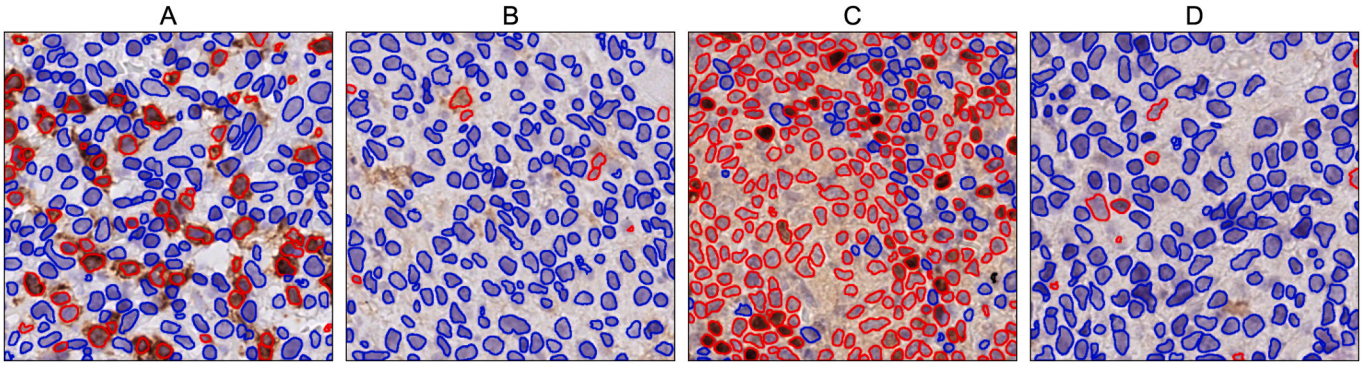


Fig. 4. Segmentation results on additional markers that were not part of the training set. (A) CD8, (B) PD-L1, (C) FOXP3, and (D) LAG3. The nuclei with red contours indicate marker-positive cells, whereas blue indicates marker-negative cells.

shows different cell composition and morphology. It hence represents itself particular challenges for automated segmentation that are not yet addressed by the SOTA segmentation methods. Our method shows comparable results in H&E– and IHC-stained lymphoma image segmentation to the quality of segmentation that is currently possible in other types of tumors (more visual depictions of our results can be found in Fig. 1 of the supplementary material).

One limitation of our study is the low inter-patient variability in LyNSeC 1, with only four different cases. Retraining our model on larger datasets of IHC data will allow to obtain a higher degree of method generalizability. Similarly, the method might need to be adapted to varying staining and imaging procedures.

Our approach allows to quantitatively describe patterns of tumor immune activation and its composition at single cell level from routine diagnostic images of lymphoma samples (Fig. 5). In combination with the H&E-derived information on tumor cells and their location, the segmentation results of our methods allow for comprehensive, quantitative analyses of large patient datasets collected in clinical centers. Such datasets are typically assessed qualitatively for diagnostic purposes. Segmentation maps generated by HoLy-Net allow to extract morphological and spatial features from the WSIs at a single-cell level, such as the relative ratios and densities of cell types, distances between cells, their sizes, shapes, etc., opening up new opportunities for systematic study of lymphoma patient cohorts. Features quantified using HoLy-Net can be utilized to identify groups of tumors characterized by similar properties, and to search for those properties that show clinical relevance. In contrast to image-based predictive or representation-learning models, models based on interpretable features provide a descriptive overview of the data and allow to generate hypotheses about the underlying biological mechanisms in cancer development and progression.

Funding

This work was supported by the Federal Ministry of Education and Research (BMBF) [grant number 01ZX1917B]; Ministry for Culture and

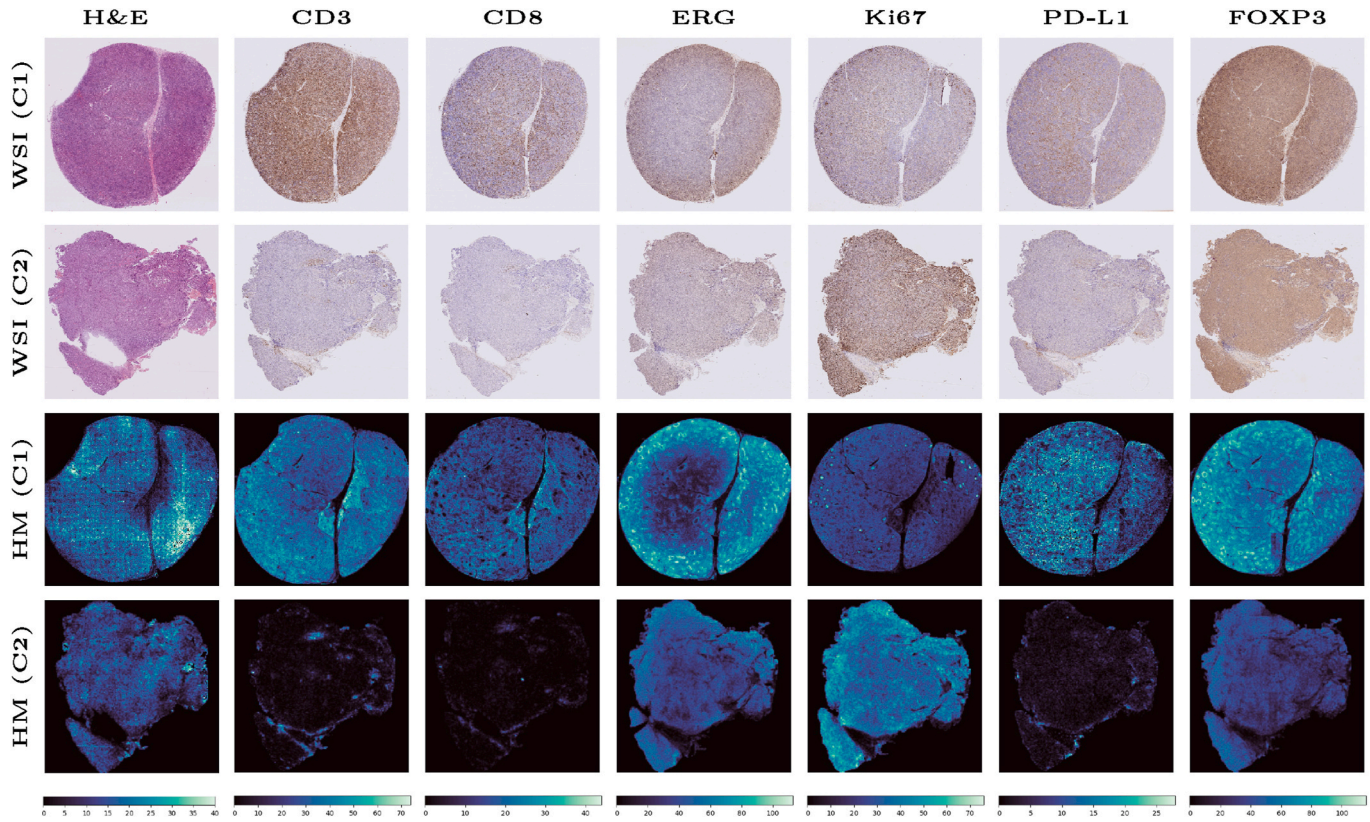


Fig. 5. Heatmaps (HM) of the number of tumor cells in the H&E-stained WSIs and numbers of marker-positive cells in the IHC-stained WSIs. Shown are two patient cases (C1 and C2) and the stainings represent consecutive slides of their tissue samples. The color bars apply to the heatmaps of both cases. The values in the H&E heatmaps were capped at 40, due to regions with high numbers of artifacts.

Science (MKW) of the State of North Rhine-Westphalia [grant number 311-8.03.03.02-147635]; and the German Research Foundation (DFG) [grant number 325931972].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2024.107978.

References

[1] S. Susanibar-Adaniya, S.K. Barta, 2021 Update on Diffuse large B cell lymphoma: a review of current data and potential applications on risk stratification and management, Am. J. Hematol. 96 (5) (2021) 617–629, https://doi.org/10.1002/ajh.26151.
[2] I.B. Buchwalow, W. Böcker, Immunohistochemistry: Basics and Methods, Springer, Berlin, Heidelberg, 1999.
[3] E. Upschulte, et al., Contour proposal networks for biomedical instance segmentation, Med. Image Anal. 77 (2022) 102371, https://doi.org/10.1016/j.media.2022.102371.
[4] C. Wang, et al., Weakly supervised bilayer convolutional network in segmentation of HER2 related cells to guide HER2 targeted therapies, Comput. Med. Imag. Graph. 108 (2023) 102270, https://doi.org/10.1016/j.compmedimag.2023.102270.
[5] H. Chen, et al., DCAN: deep contour- aware networks for accurate Gland segmentation, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2487–2496, https://doi.org/10.1109/CVPR.2016.273.
[6] S. Graham, et al., HoVer-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images, Med. Image Anal. 58 (2019) 101563, https://doi.org/10.1016/j.media.2019.101563.
[7] P. Naylor, et al., Segmentation of nuclei in Histopathology images by deep Regression of the distance map, IEEE Trans. Med. Imag. 38 (2) (2019) 448–459, https://doi.org/10.1109/TMI.2018.2865709.
[8] M.H. Masoudi, M. Mikaeili, Diagnosis of Hodgkin's Disease by Identifying Reed-Sternberg Cell Nuclei in Histopathological Images of Lymph Nodes Stained with Hematoxylin and Eosin, 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME), 2020, pp. 298–304, https://doi.org/10.1109/ICBME51989.2020.9319460.
[9] T. Tosta, et al., Segmentation methods of H&E-stained histological images of lymphoma: a review, Inform. Med. Unlocked 9 (2017) 35–43, https://doi.org/10.1016/j.imu.2017.05.009.
[10] T. Tosta, et al., Computational method for unsupervised segmentation of lymphoma histological images based on fuzzy 3-partition entropy and genetic algorithm, Expert Syst. Appl. 81 (2017) 223–243, https://doi.org/10.1016/j.eswa.2017.03.051.
[11] J. Oscanoa, et al., Automated segmentation and classification of cell nuclei in immunohistochemical breast cancer images with estrogen receptor marker, 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2016) 2399–2402, https://doi.org/10.1109/EMBC.2016.7591213.
[12] H. Zhou, K.Z. Mao, Adaptive successive erosion-based cell image segmentation for p53 immunohistochemistry in bladder inverted papilloma, IEEE Engineering in Medicine and Biology 27th Annual Conference (2005) 6484–6487, https://doi.org/10.1109/IEMBS.2005.1615984.
[13] J. Carreras, et al., A Combination of Multilayer Perceptron, Radial Basis Function Artificial Neural Networks and Machine Learning Image Segmentation for the Dimension Reduction and the Prognosis Assessment of Diffuse Large B-Cell Lymphoma. AI, vol. 2, 2021, pp. 106–134, https://doi.org/10.3390/ai2010008.
[14] M. Hall, et al., The WEKA data mining software: an update, SIGKDD Explor. Newsl. 11 (2008) 10–18, https://doi.org/10.1145/1656274.1656278.
[15] Z. Swiderska-Chadaj, et al., Learning to detect lymphocytes in immunohistochemistry with deep learning, Med. Image Anal. 58 (2019) 101547, https://doi.org/10.1016/j.media.2019.101547.
[16] O. Ronneberger, et al., U-net: convolutional networks for biomedical image segmentation, Medical Image Computing and Computer-Assisted Interventions (MICCAI) 9351 (2015) 2980–2988, https://doi.org/10.1007/978-3-319-24574-4_28.
[17] L. Mahanta, et al., IHC-Net: a fully convolutional neural network for automated nuclear segmentation and ensemble classification for Allred scoring in breast

pathology, Appl. Soft Comput. 103 (2021) 107136, https://doi.org/10.1016/j.asoc.2021.107136.
[18] H. Clevers, et al., The T cell receptor/CD3 complex: a dynamic protein ensemble, Annu. Rev. Immunol. 6 (1) (1988) 629–662, https://doi.org/10.1146/annurev.iy.06.040188.003213.
[19] D.M. Schonk, et al., Assignment of the gene(s) involved in the expression of the proliferation-related Ki-67 antigen to human chromosome 10, Hum. Genet. 83 (3) (1989) 297–299, https://doi.org/10.1007/BF00285178.
[20] J. Thoms, et al., ERG promotes T-acute lymphoblastic leukemia and is transcriptionally regulated in leukemic cells by a stem cell enhancer, Blood 117 (26) (2011) 7079–7089, https://doi.org/10.1182/blood-2010-12-317990.
[21] M. Miettinen, et al., ERG transcription factor as an immunohistochemical marker for vascular endothelial tumors and prostatic carcinoma, Am. J. Surg. Pathol. 35 (3) (2011) 432–441, https://doi.org/10.1097/PAS.0b013e318206b67b.
[22] S. Fujita, X.H. Han, Cell detection and segmentation in microscopy images with improved mask R-CNN, Asian Conference on Computer Vision 12628 (2020) 58–70, https://doi.org/10.1007/978-3-030-69756-3_5.
[23] Lv Guofeng, et al., Nuclei R-CNN: improve mask R-CNN for nuclei segmentation, IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP) (2019) 357–362, https://doi.org/10.1109/ICICSP48821.2019.8958541.
[24] X. Xie, et al., Robust segmentation of nucleus in histopathology images via mask R-CNN. Brainlesion: glioma, multiple sclerosis, Stroke and Traumatic Brain Injuries 11383 (2018) 428–436, https://doi.org/10.1007/978-3-030-11723-8_43.
[25] K. He, et al., Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980–2988, https://doi.org/10.1109/ICCV.2017.322.
[26] N. Kumar, et al., A dataset and a technique for generalized nuclear segmentation for computational pathology, IEEE Trans. Med. Imag. 36 (7) (2017) 550–1560, https://doi.org/10.1109/TMI.2017.2677499.
[27] Q. Dang Vu, et al., Methods for segmentation and classification of digital microscopy tissue images, Front. Bioeng. Biotechnol. 7 (2019) 2296–4185, https://doi.org/10.3389/fbioe.2019.00053.
[28] A. Kirillov, et al., Panoptic Segmentation. 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9396–9405, https://doi.org/10.1109/CVPR.2019.00963.

Chapter 4

Deep Learning-based Interpretable Prediction of Recurrence of Diffuse Large B-Cell Lymphoma

---

The last study [33] focuses on the DL-based prediction of recurrence of DLBCL and utilizes the methods and findings of the previous studies [31, 32] to interpret the prediction results by finding morphological patterns that are characteristic for recurred and non-recurred patients.

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

Hussein Naji[1,2], Juan I. Pisula[1,2], Stefano Ugliano[1,2], Adrian Simon[4], Reinhard Büttner[4], and Katarzyna Bożek[1,2,3]

[1] Institute for Biomedical Informatics, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany
[2] Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany
[3] Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Germany
[4] Institute of Pathology, University Hospital Cologne, Germany

Abstract. Diffuse large B-cell lymphoma (DLBCL) is the most common non-Hodgkin lymphoma in adults. Its morphologically, genetically, and clinically heterogeneous nature presents significant challenges for treatment as up to 50% of patients do not benefit from frontline chemoimmune therapy and experience recurrence of disease, characterized by alarmingly low survival rates. Upfront detection of recurring patients could offer them alternative treatments. However, there are no known biomarkers distinguishing these patients from those who respond to treatment. Deep learning has shown potential in predicting recurrence of various cancer types, however, deep black box classifiers suffer from lack of interpretability. In particular in prediction of patient outcomes, an understanding of the model's decision could result in novel biomarkers and eventually treatments. To address these challenges, we developed a deep learning-based pipeline to predict recurrence of DLBCL based on hematoxylin and eosin (H&E)-stained tissue slides 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 decision. Subsequently, we segmented the nuclei within these high attention areas, calculated morphological features, and statistically analyzed them to find differences between recurred and non-recurred patients. We achieved an f1 score of 0.83 indicating that our model is indeed able to 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. 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.

Keywords: diffuse large B-cell lymphoma · recurrence prediction · attention based multiple instance learning · hematoxylin and eosin · morphological features.

1 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. Their model indeed shows high degree of interpretability by utilizing single-cell-based morphological features. 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 its interpretation. We developed a deep learning-based prediction pipeline based on tissue microarrays (TMAs) stained with hema-

toxylin 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-based morphological features which are distinct for recurred and non-recurred patients. While the causes of DLBCL recurrence are still unknown, here, we point to cell morphology features which might be determinant for patient response to treatment.

2 Methods

2.1 Patient Cohort and Dataset

We use a publicly available dataset of the Stanford Cancer Institute [21]. The data is based on a study cohort with 167 patients, containing 306 cores (1-2 cores per patient) from 7 H&E-stained TMAs. The TMA slides were scanned at 40x 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.

2.2 Pipeline

We developed a pipeline (Figure 1) that starts with a preprocessing step (Figure 1A) in which the cores are cut into non-overlapping patches of size 256 x 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 (Figure 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. Additionally, it employs clustering at the level of individual instances within these key regions to restrict and refine the feature space (Figure 1C) [22].
For training the classification model, we used 60% of patients for training, 20% for validation and 20% for final testing. We applied a patient-specific split, that is, samples of patients can only be present in one subset (training, validation, or test). 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 (Figure 1D) [24]. Subsequently, 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 (Figure 1E). We quantified the following shape- and size-based nuclei features:

- area: area of the nucleus.
- major axis length: length of the major axis of the ellipse that has the same second central moment as the nucleus.
- minor axis length: length of the minor axis of the ellipse that has the same second central moment as the nucleus.
- 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].

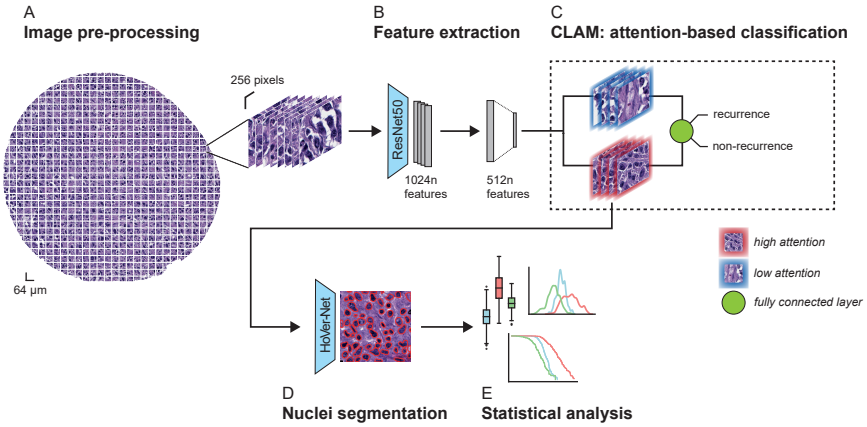


Fig. 1. Deep learning pipeline. **A** H&E-stained core images were segmented and cut into non-overlapping patches of size 256 x 256 pixels. **B** Features were extracted from all patches of a core image using a ResNet50 encoder. The final feature vectors had a size of 1024. **C** CLAM model to classify 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.

3 Results

3.1 Relapse Prediction

Our model achieves a classification accuracy of 0.73, a precision of 0.74, a recall of 0.96, and an f1 score of 0.83. A confusion matrix is shown in Figure 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.

Fig. 2. Confusion matrix of the recurrence prediction on the hold-out test set (n=63). The matrix was normalized and shows the average performance across all cross validation iterations.

3.2 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 (Figure 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 with marked differences in these morphological features. Elevated values of nuclei area, minor axis length, and roughness suggest that patients who recurred have bigger nuclei and more nuclei with irregular-shaped cells.

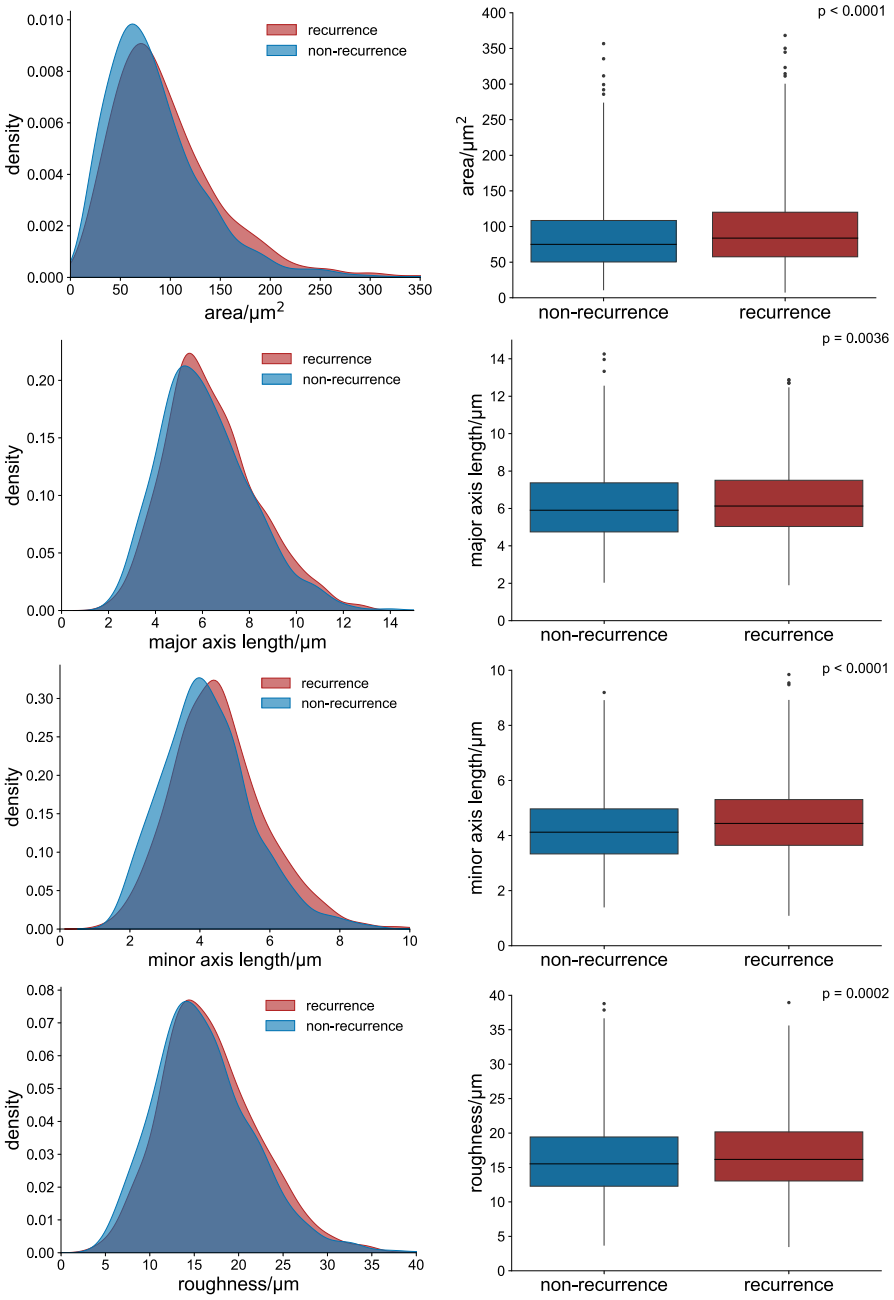


Fig. 3. Distribution and boxplots of the determined nuclei-based morphological features for recurred (n=1139) and non-recurred (n=1317) patients.We calculated the features 'area', 'major axis length', 'minor axis length', and 'roughness'. The p values indicate the t-test results to highlight statistically significant differences between the features of recurred and non-recurred patients.

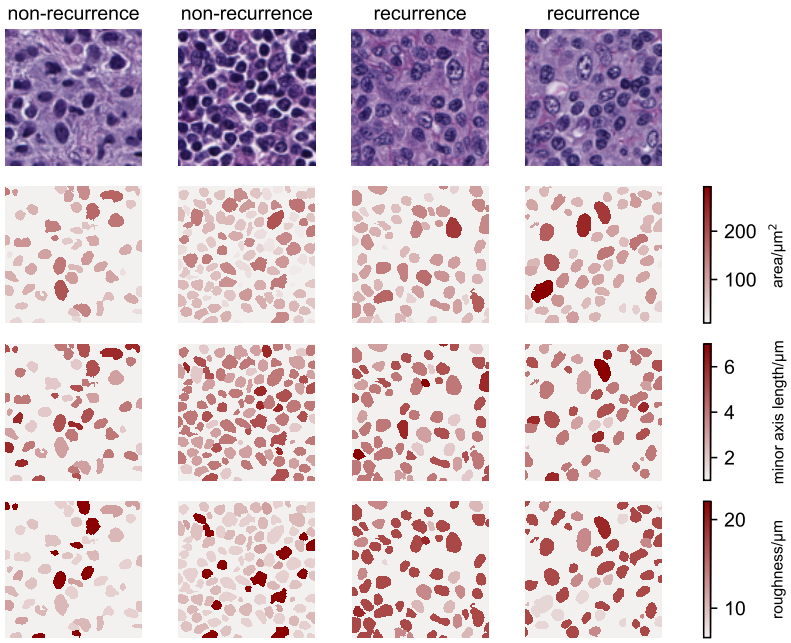


Fig. 4. Heatmap visualization of feature values. For the top two patches of recurred and non-recurred patients, we visualize the strengths of the features 'area', 'minor axis length', and 'roughness'. The darker the color of a cell the higher its respective feature value.

3.3 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 experience recurrence based on the clinical follow up data.

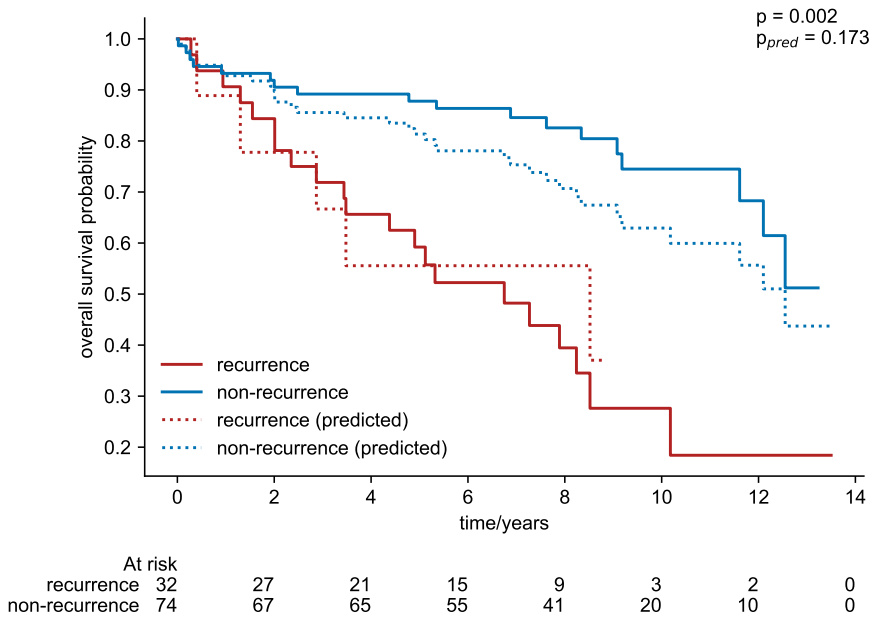


Fig. 5. Kaplan–Meier plot for overall survival. The plot shows the overall survival probability of recurred and non-recurred patients (true and predicted) over time. The variable p indicates the t-test-based p-value for true recurrence and non-recurrence and p_{pred} the p value for predicted recurrence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

4 Discussion

Deep learning presents tremendous opportunities for medically relevant prediction tasks, such as recurrence or survival. Additionally, attention-based mechanisms enable quantitative analyses of the tumor morphology and microenvironment of image areas that were crucial to the model's prediction decision, allowing enhanced explainability of such models.

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. To our knowledge, there is no other existing DLBCL dataset with available clinical data (regarding recurrence). In the future, additional datasets from different sources could serve as an external test cohort and could help to improve the generalizability of the prediction model. Our statistical analysis further suggests that the model's prediction decision can be explained by morphological characteris-

tics 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 than the nuclei of non-recurred patients (Figures 3-4), indicating a higher degree of nuclear pleomorphism, which correlates with tumor aggressiveness [26,27]. Lastly, our model is indirectly predicting patient survival as well (Figure 5).

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

References

1. Susanibar-Adaniya, S., Barta, S.K.. 2021 Update on Diffuse large B cell lymphoma: A review of current data and potential applications on risk stratification and management. American Journal of Hematology96(5), 617-629 (2021).
2. Lucas, M., Jansen, I., van Leeuwen, T.G., Oddens, J.R., de Bruin, D.M., Marquering, H.A.,. Deep Learning–based Recurrence Prediction in Patients with Non–muscle-invasive Bladder Cancer. European Urology Focus 8(1), 165-172 (2022).
3. Yang. Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. Computational and Structural Biotechnology Journal 20, 333-342 (2022).
4. Howard, F.M., 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 9, 25 (2023).
5. Shi, Y., Olsson, L.T., Hoadley, K.A., Calhoun, B.C., Marron, J.S., Geradts, J. et al. Predicting early breast cancer recurrence from histopathological images in the Carolina Breast Cancer Study. npj Breast Cancer 9, 92 (2023).
6. Su, Z., Niazi, M.K.K., Tavolara, T.E., Niu, S., Tozbikian G.H., Wesolowski, R. et al. BCR-Net: A deep learning framework to predict breast cancer recurrence from histopathology images. PLoS ONE 18(4), e0283562 (2023).
7. Foersch, S., Glasner, C., Woerl, A.C., Eckstein, M., Wagner, D.C., Schulz, S. et al. Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. Nat Med 29, 430–439 (2023).
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 130, 951–960 (2024).

9. Yamashita, R., Long, J., Saleem, A., Rubin, D.L., Shen, J. Deep learning predicts postsurgical recurrence of hepatocellular carcinoma from digital histopathologic images. Sci Rep 11, 2047 (2021).
10. Qu, W.F., Tian, M.X., Qiu, J.T., Guo, Y.C., Tao, C.Y., Liu, W.R. et al. Exploring pathological signatures for predicting the recurrence of early-stage hepatocellular carcinoma based on deep learning. Front. Oncol 12, 968202 (2022).
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. 11, 768 (2020).
12. Akram, F., Wolf, J.L., Trandafir, T.E., Dingemans, A.M.C., Stubbs, A.P. von der Thüsen, J.H. et al. Artificial intelligence-based recurrence prediction outperforms classical histopathological methods in pulmonary adenocarcinoma biopsies. Lung Cancer 186, 107413 (2023).
13. Dietrich, E. Fuhlert, P., Ernst, A., Sauter, G., Lennartz, M., Stiehl, H.S. 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. Proceedings of Machine Learning Research 158, 38–53 (2021).
14. Eminaga, O, Saad, F., Tian, Z., Wolffgang, U., Karakiewicz, P.I., Ouellet, V. et al. Artificial Intelligence Helps to Predict Recurrence and Mortality for Prostate Cancer Using Histology Images. bioRxiv, Preprint posted online 20 August 2023, doi: https://doi.org/10.1101/2023.07.27.550781 (2023).
15. He, K., Zhang, X., Ren, S., Sun, J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778 (2016).
16. Harkins, R.A., Chang, A., Patel, S.P., Lee, M.J., Goldstein, J.S., Merdan, S. et al. Remaining challenges in predicting patient outcomes for diffuse large B-cell lymphoma. Expert Review of Hematology, 12(11), 959-973 (2019).
17. Wang, L., Zhao, Q., 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. Computer Methods and Programs in Biomedicine 196, 105567 (2020).
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 Mining 14, 38 (2021).
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. Computer Methods and Programs in Biomedicine 226, 107103 (2022).
20. Shankar, V., Yang, X., Krishna, V., Tan, B., Silva, O., Rojansky, R. et al. LymphoML: An interpretable artificial intelligence-based method dentifies morphologic features that correlate with lymphoma subtype. Proceedings of Machine Learning Research 225, 528–558 (2023).

21. Vrabac, D., Smit, A., Rojansky, R., Natkunam, Y., Advani, R.H., Ng, A.Y. et al. DLBCL-Morph: Morphological features computed using deep learning for an annotated digital DLBCL image set. Sci Data 8, 135 (2021).
22. Lu, M.Y., Williamson, D.F.K., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F. et al. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng 5, 555–570 (2021).
23. Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W. Kwak, J.T. et al. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Medical Image Analysis 58, 101563 (2019).
24. Naji, H., Sancere, L., Simon, A., Büttner, R., Eich, M.L., Lohneis, P. et al. HoLy-Net: Segmentation of histological images of diffuse large B-cell lymphoma. Computers in Biology and Medicine 170, 107978 (2024).
25. Phillip, J.M., Han, KS., Chen, WC., Wirtz, D., Wu, P.H. A robust unsupervised machine-learning method to quantify the morphological heterogeneity of cells and nuclei. Nat Protoc 16, 754–774 (2021).
26. Rashid, F., Haque, A.U. Frequencies of different nuclear morphological features in prostate adenocarcinoma. Annals of Diagnostic Pathology 15(6), 414-421 (2011).
27. Fischer, E.G. Nuclear Morphology and the Biology of Cancer Cells. Acta Cytologica 2 64 (6), 511–519 (2020).

Chapter 5

Discussion

In this chapter, the main contributions of this work are summarized and put into context of the research hypothesis, including an elaboration on study limitations (sections 5.1-5.3). Lastly, a conclusion and future perspectives (section 5.4) are given.

5.1 Lymphoma Cell Nuclei Classification using Color and Morphology Features

A dataset of 6,273 annotated cell nuclei was used for ML models to distinguish between five different types of cells: tumor (lymphoma), lymphocyte, leukocyte, stroma, and epithelial cells. The annotated nuclei stem from H&E-stained WSIs of DLBCL. The goal was to accurately predict the cell type based on the morphology of the cells. Hence, morphological features were calculated based one the contours of the nuclei, as well as color-based features determined from the cell nuclei images, as an input for the ML models. In a binary classification task, where tumor and non-tumor cells were distinguished, the model yielded a classification accuracy of 0.91, showing its strong ability to identify tumor cells. In the multi-classification task, where all five before-mentioned cell types were classified, the accuracy is 0.84, indicating a strong performance as well. However, the confusion matrix showed, that the model struggled to correctly classify underrepresented cell-types. The performance of the algorithm and feature analysis indicated that the cell morphology is a strong prognosticator for ML tasks, allowing the hypothesis that the morphology could also be a significant characteristic for predictive tasks, such as the prediction of recurrence of DLBCL.
A limitation here was the high imbalance of cell types. Tumor cells and lymphocytes constitute almost 97% of all samples, leaving only 3% for the remaining cell types (epithelial, leukocyte, and stroma). More samples of the underrepresented classes could improve the class-wise classification accuracy. Furthermore, the color features should be used with caution. Although improving the classification, these features can vary highly in different datasets as the colors do not only depend on the staining procedure (which has a non-neglectable degree of variability between different sources) but also on the scanner to digitize the images.
However, this study clearly shows the significance of the cell morphology for classification tasks and allows for the assumption that the morphology could also be a helpful feature to detect patterns for recurred and non-recurred patients.

5.2 Segmentation of Histological Images of Diffuse Large B-Cell Lymphoma

Nuclei segmentation is crucial for analyzing the tumor microenvironment and there is a scarcity of datasets and methods that allow to segment lymphoma cells. The proposed method outperformed other existing segmentation models for H&E-stained WSIs of lymphoma (with a dice score of 0.89), and is also a unique technique for the segmentation of IHC-stained WSIs (with a dice score of 0.86). For H&E slides, the segmented nuclei were classified into tumor and non-tumor cells and for the IHC slides, a classification of marker-positive and marker-negative was applied. With an accuracy of 0.92 (H&E) and 0.96 (IHC), the models were reliable in correctly classifying the segmented cell nuclei. The methods were applied on a data set that was published in [32], containing 73,931 annotated cell nuclei from H&E and 87,316 from IHC slides. In a further analysis, WSIs were segmented and visualized as a heatmap, highlighting areas of high tumor/marker-positive cell densities, allowing to quantitatively analyze the tumor microenvironment at single-cell level.
The nuclei segmentation of IHC-stained images is based on a labeled dataset containing four cases. Although showing high performance, the generalizability of the model could be improved by adding samples from more patients to the dataset. For the H&E dataset, nuclei were labeled as either tumor or non-tumor. A more precise labeling for the non-tumor class (e.g. as in the first study) could yield an even better overview of the tumor microenvironment.
Nevertheless, this approach closes the gap of non-existing nuclei segmentation techniques for lymphoma datasets and contributes significantly to the segmentation of IHC-stained WSIs, that has not been emphasized by previous studies.

5.3 Deep Learning-based Interpretable Prediction of Recurrence of Diffuse Large B-Cell Lymphoma

Recurrence of DLBCL was predicted using a MIL model, yielding an f1 score of 0.83, indicating that the model was able to distinguish non-recurred from recurred patients. The integrated attention mechanisms of the DL model made it possible to extract areas of high-resolution images that were of high significance for the model's decision. Utilizing the knowledge and methods from the previous experiments, the nuclei of the high-attention areas were segmented and morphological features were calculated based on the resulting contours. Subsequently, a statistical analysis was applied in which differences between features from recurred and non-recurred patients were highlighted. This analysis concluded that nuclei of recurred patients tend to be bigger and more irregular in shape than the nuclei of non-recurred patients. Big and irregular cells indicate a higher degree of nuclear pleomorphism which correlates with tumor aggressiveness. Consequently, a higher tumor aggressiveness can be assessed with a higher risk of recurrence. [34, 35]. Thus, the method not only was able to predict recurrence but also offered the possibility to interpret the prediction in a biologically relevant matter.
These experiments were based on a previously published dataset [36] with a high imbalance between the targets of interest (recurrence and non-recurrence). Additionally, the overall number of samples is relatively low. Removing the bias and increasing the total

sample size could help to gain more significant outcomes. What is more, the data stems from TMA cores wich encompass certain areas of the whole tissue sample. A dataset containing the whole sample would increase the number of patches per patient and may also contribute to a better performing prediction model.
Combining the recurrence prediction with the methods and findings discussed in section 5.1 and 5.2, it can be concluded that tissue samples of humans indeed have characteristics, that are predictive of recurrence/non-recurrence. Here, the characteristics were shown to be distinct patterns within the cell morphology of recurred and non-recurred patients, respectively.

5.4 Conclusion and Outlook

A common challenge of all conducted experiments was the class imbalance and/or low amount of total samples. Additionally, all methods were applied on one dataset from the same source, respectively. Since generalizability is a significant aspect for DL models more data from different sources would help the models to be agnostic to the data source. However, well annotated histopathological data is rare (especially for lymphoma) as the labeling procedure is time-consuming and requires the expertise of trained pathologists. Therefore, well-annotated datasets together with DL methods can help to improve outcomes and diminish the burden on pathologists. In summary, the studies encompassed in this thesis showed that histological images contain meaningful information and can be utilized for various classification, segmentation, and prediction tasks. In this regard, image-based analyses, including traditional ML as well as DL approaches are shown to be powerful tools to inspect the tumor microenvironment, allowing to generate hypotheses about the underlying biological mechanisms of cancer. Existing methods predominantly focus on solid tumors, leaving a research gap for the analysis of cancers of the lymphatic system. This project contributes to closing this gap by focusing on a highly aggressive and heterogeneous type of lymphoma, namely DLBCL, by developing novel tools for the identification of tumor cells, segmentation of nuclei in WSIs, predicting recurrence of DLBCL, and finding distinct patterns in the cell morphology of recurred and non-recurred patients, respectively.
This work and state-of-the-art research has shown the reliability and great potential of DL models in histopathology. In the medical field, patient insights are constructed from multiple data modalities such as:

- **clinical data**: encompasses medical history, diagnosis, medications, treatment assessments, allergies, age, gender, etc.
- **genomic data**: provides information about the genetic material of an organism (DNA and RNA), often in the form of sequencing data.
- **biometric data**: includes information derived from sensors, such as ECGs, EEGs, blood pressure, etc.

All these modalities are useful for DL models. In the future, the development of multimodal models (which is a growing field of research) would allow to combine different

modalities to achieve better performances. Relating to the prediction of recurrence, especially biometric and sequencing data of a patient could contain meaningful information that is not visible in imaging data and thus, a DL architecture that is able to process all of these modalities together would likely perform better than using one of these modalities alone. The challenge here lies in how to merge modalities that are completely different from each other and how to merge them in a way that does not lead to high information loss.

The aim of all digital pathology studies is the same: beating cancer. With new annotated datasets that are published regularly and continuously improving methods, it can eventually be possible to develop DL models that are not only agnostic to data from different sources but also to different cancer types.

Bibliography

[1] Sung H, Ferlay J, Siegel RL, et al. 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *American Cancer Society.* 71(3):209-249. doi: 10.3322/caac.21660.

[2] Susanibar-Adaniya S, Barta SK. 2021. 2021 Update on Diffuse large B cell lymphoma: A review of current data and potential applications on risk stratification and management. *American Journal of Hematology.* 96(5):617-629. doi: 10.1002/ajh.26151.

[3] Esmatabadi MJ, Bakhshinejad B, Motlagh FM, et al. 2016. Therapeutic resistance and cancer recurrence mechanisms: Unfolding the story of tumour coming back. *Journal of Biosciences.* 41(3):497-506. doi: 10.1007/s12038-016-9624-y.

[4] Bancroft JD, Gamble M. 2007. Theory and Practice of Histological Techniques (Sixth Edition). *Churchill Livingstone.* ISBN: 9780443102790.

[5] Aziz, SJ, Zeman-Pocrnich CE. 2022. Tissue Processing. *Immunohistochemistry and Immunocytochemistry. Methods in Molecular Biology.* 2422:47–63 doi: 10.1007/978-1-0716-1948-3_4.

[6] Titford M. 2005. The long history of hematoxylin. *Biotechnic & Histochemistry*, 80(2):73–78. doi: 10.1080/10520290500138372.

[7] Duraiyan J, Govindarajan R, Kaliyappan K, et al. 2012. Applications of immunohistochemistry. *Journal of Pharmacy and Bioallied Sciences.* 4(2):307-309, August 2012. doi: 10.4103/0975-7406.100281.

[8] Binch A, Snuggs J, Le Maitre CL. 2020 Immunohistochemical analysis of protein expression in formalin fixed paraffin embedded human intervertebral disc tissues. *JOR Spine.* 3(3):e1098. doi: 10.1002/jsp2.1098.

[9] Kumar N, Gupta R, Gupta S. 2020. Whole Slide Imaging (WSI) in Pathology: Current Perspectives and Future Directions. *J Digit Imaging.* 33(4):1034-1040. doi: 10.1007/s10278-020-00351-z.

[10] Camp RL, Neumeister V, Rimm DL. 2008. A Decade of Tissue Microarrays: Progress in the Discovery and Validation of Cancer Biomarkers. *Journal of Clinical Oncology.* 26(34):5630-5637. doi:10.1200/JCO.2008.17.3567.

[11] Gremel G, Bergman J, Djureinovic D. 2014. A systematic analysis of commonly used antibodies in cancer diagnostics. *Histopathology.* 64(2):293-305. doi: 10.1111/his.12255.

[12] Litjens G, Sánchez C, Timofeeva N, et al. 2016. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep.* 6:26286. doi: 10.1038/srep26286.

[13] Bera K, Schalper KA, Rimm DL, et al. 2019. Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol.* 16:703–715. doi: 10.1038/s41571-019-0252-y.

[14] Mobadersany P, Yousefi S, Amgad M, et al. 2018 Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences*. 115(13):E2970-E2979. doi: 10.1073/pnas.1717139115.

[15] Hermsen M, de Bel T, den Boer M, et al. 2019. Deep Learning-Based Histopathologic Assessment of Kidney Tissue. *JASN*. 30(10):1968-1979. doi: 10.1681/ASN.2019020144.

[16] Janowczyk A, Madabhushi A. 2016. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics*. 7(1):29. doi: 10.4103/2153-3539.186902.

[17] Korbar B, Olofson AM, Miraflor AP, et al. 2017. Deep Learning for Classification of Colorectal Polyps on Whole-slide Images. *Journal of Pathology Informatics*. 8(1):30. doi: 10.4103/jpi.jpi_34_17.

[18] Coudray N, Ocampo PS, Sakellaropoulos T, et al. 2018. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nat Med*. 24:1559–1567. doi: 10.1038/s41591-018-0177-5.

[19] Fourkioti O, De Vries M, Bakal C. 2023. CAMIL: Context-Aware Multiple Instance Learning for Cancer Detection and Subtyping in Whole Slide Images. *arXiv*. Preprint, lastly accessed on 18.04.2024. doi: 10.48550/arXiv.2305.05314.

[20] He K, Zhang X, Ren S, et al. 2016. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp.770-778. doi: 10.1109/CVPR.2016.90.

[21] Szegedy C, Liu W, Jia Y, et al. 2015. Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp.1-9. doi: 10.1109/CVPR.2015.7298594.

[22] Ronneberger O, Fischer P, Brox T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science*. 9351:234–241. doi: 10.1007/978-3-319-24574-4_28.

[23] Graham S, Vu QD, Raza SEA, et al. 2019. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*. 58:101563. doi:10.1016/j.media.2019.101563.

[24] Kim H, Yoon H, Thakur N, et al. 2021. Deep learning-based histopathological segmentation for whole slide images of colorectal cancer in a compressed domain. *Sci Rep*. 11:22520. doi: 10.1038/s41598-021-01905-z.

[25] Zhao T, Fu C, Tian Y, et al. 2023. GSN-HVNET: A Lightweight, Multi-Task Deep Learning Framework for Nuclei Segmentation and Classification. Bioengineering. 10(3):393. doi: 10.3390/bioengineering10030393.

[26] Naylor P, Lae M, Reyal F, et al. 2019. Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map. *IEEE Transactions on Medical Imaging*. 38(2):448-459. doi: 10.1109/TMI.2018.2865709.

[27] Zhang L, Xia W, Yan ZP, et al. 2020. Deep Learning Predicts Overall Survival of Patients With Unresectable Hepatocellular Carcinoma Treated by Transarterial Chemoembolization Plus Sorafenib. *Front Oncol.* 10:593292. doi: 10.3389/fonc.2020.593292.

[28] Shi Y, Olsson LT, Hoadley KA, et al. 2023. Predicting early breast cancer recurrence from histopathological images in the Carolina Breast Cancer Study. *npj Breast Cancer.* 9:92. doi: 10.1038/s41523-023-00597-0.

[29] Xiao H, Weng Z, Sun K, et al. 2024. Predicting 5-year recurrence risk in colorectal cancer: development and validation of a histology-based deep learning approach. *British Journal of Cancer.* 130:951–960. doi: 10.1038/s41416-024-02573-2.

[30] Harkins RA, Chang A, Patel SP, et al. 2019. Remaining challenges in predicting patient outcomes for diffuse large B-cell lymphoma. *Expert Review of Hematology*, 12(11):959-973. doi: 10.1080/17474086.2019.1660159.

[31] Naji H, Hahn L, Bożek K. 2023. Lymphoma Cell Nuclei Classification using Color and Morphology Features. *Current Directions in Biomedical Engineering.* 9(1):210-212. doi: 10.1515/cdbme-2023-1053.

[32] Naji H, Sancere L, Simon A, et al. 2024. HoLy-Net: Segmentation of histological images of diffuse large B-cell lymphoma. *Computers in Biology and Medicine.* 170:107978. doi: 10.1016/j.compbiomed.2024.107978.

[33] Naji H, Pisula JI, Ugliano S, et al. 2024. Deep learning-based interpretable prediction of recurrence of diffuse large B-cell lymphoma. Under review for publication in the *British Journal of Cancer.*

[34] Rashid F, Haque AU. 2011. Frequencies of different nuclear morphological features in prostate adenocarcinoma. *Annals of Diagnostic Pathology.* 15(6):414-421. doi: 10.1016/j.anndiagpath.2011.06.002.

[35] Fischer EG. 2020. Nuclear Morphology and the Biology of Cancer Cells. *Acta Cytologica 2.* 64(6):511–519. doi: 10.1159/000508780.

[36] Vrabac D, Smit A, Rojansky R, et al. 2021. DLBCL-Morph: Morphological features computed using deep learning for an annotated digital DLBCL image set. *Sci Data.* 8:135. doi: 10.1038/s41597-021-00915-w.

Appendix A

Supplementary Material for Chapter 3

parameter	value
pre-trained weights	ResNet101
learning rate	10^{-5}
batch size	2
weight decay	10^{-6}
nms threshold	0.8
test detections per image	400
anchor size	8, 16, 32, 64
iterations	40000

Table A.1: Configuration of Mask R-CNN model.

parameter	value
pre-trained weights	Preact ResNet50
learning rate	10^{-4}
batch size	4
weight decay	10^{-6}
reduce on plateau factor	0.8
reduce on plateau patience	25
dropout factor	0.2
epochs	100

Table A.2: Configuration of HoLy-Net model.