

Aus der Klinik für Hämatologie, Onkologie und
Immunologie
Geschäftsführender Direktor: Herr Univ.-Prof. Dr. med.
Andreas Burchert

des Fachbereichs Medizin der Philipps-Universität
Marburg

Titel der Dissertation:

**Identifizierung prognostischer Zellpopulationen bei
Patienten mit Chronischer lymphatischer Leukämie
durch Anwendung von erklärbarer künstlicher
Intelligenz auf durchflusszytometrische Daten**

Inaugural-Dissertation zur Erlangung des Doktorgrades
der gesamten Humanmedizin
dem Fachbereich Medizin der Philipps-Universität
Marburg
vorgelegt von

Semil Eminovic
aus Darmstadt
Marburg, 2024

Angenommen vom Fachbereich Medizin der Philipps-
Universität Marburg am:
30.09.2024

Gedruckt mit Genehmigung des Fachbereichs Medizin

Dekanin:	Frau Prof. Dr. Denise Hilfiker- Kleiner
Referent:	Herr Priv.-Doz. Dr. Jörg Hoffmann
1. Korreferent:	Herr Prof. Dr. Martin Hirsch

Gewidmet
meinen Eltern Ismeta und Ujkan sowie meinen
Geschwistern Edis und Dzeneta.

Inhaltsverzeichnis

1. Einleitung	16
1.1 Chronische lymphatische Leukämie	16
1.1.1 Epidemiologie	16
1.1.2 Diagnosestellung der CLL	16
1.1.3 Prognose der CLL	26
1.1.4 Therapie der CLL	33
1.1.5 Relevanz des Themas	38
1.2 Der durchflusszytometrische Datensatz	39
1.3 Machine-Learning-Verfahren in der Hämatologie	46
1.4 Explainable Artificial Intelligence (XAI) ...	53
2. Patienten, Materialien und Methoden.....	61
2.1 Patienten, Datenerhebung	61
2.2 Antigen-Panel, durchflusszytometrische Färbung und Analyse	64
2.2.1 Datenkompilierung ALPODS	67
2.2.2 Statistik	70

2.3 Algorithmic Population Descriptions (ALPODS)	72
3. Ergebnisse.....	81
3.1 Patientencharakteristika	81
3.2 Datenverarbeitung und Zellpopulationen von ALPODS	84
3.3 Identifizierung der XAI-Populationen	95
3.4 Charakterisierung der prädiktiven Untergruppen innerhalb der CLL-Zellen	99
3.5 Klinische Anwendbarkeit.....	102
4. Diskussion.....	106
4.1 Diskussion klinischer Ergebnisse	107
4.2 Diskussion des Patientenkollektivs.....	111
4.3 Diskussion von ALPODS und weiterer Clusteralgorithmen.....	113
5. Zusammenfassung	117
6. Summary	122

Literaturverzeichnis	126
Anhang.....	157
I. Lebenslauf.....	157
II. Verzeichnis der akademischen Lehrer/- innen	158
III. Danksagung.....	160
IV. Ehrenwörtliche Erklärung	161

Abbildungsverzeichnis

Abbildung 1 A-F: Beispielhaftes CLL-Gating	42
Abbildung 2 (angelehnt an Chaddad et al. ¹⁸): Kategorisierung verschiedener XAI-Methoden....	55
Abbildung 3A und B (angelehnt an Das und Rad ²⁴): Illustration eines intrinsischen sowie eines Post-hoc XAI-Modells.....	56
Abbildung 4 (von Ultsch et al. ¹³²): VisPanel in Form von MPFC-Punktdiagrammen (Streudiagramme) von Ereignissen aus dem Trainingssatz für die Regel „Regel1: Relevante Ereignisse in Population 1 zur Identifizierung des Knochenmarks sind SS++ und CD33+ und CD13- “	76
Abbildung 5: Kaplan-Meier-Kurve der CLL- Patientenkohorte	83
Abbildung 6: Arbeitsablauf und Datenverarbeitung mit ALPODS	85
Abbildung 7A-F: Box-Plots zur Verteilung der Zellpopulationen T1C0011, T1C0016, T1C0020,	

T1C0023, T2C0004 sowie T2C0018 zu IPI-Scores ≤ 1 und ≥ 4	90
Abbildung 8A und 8B: Multiple logistische Regression mit ROC-Kurvenanalyse.....	94
Abbildung 9A und 9B: Identifizierung und Beschreibung der XAI-Populationen in zweidimensionalen durchflusszytometrischen Punktwolkendiagrammen	96
Abbildung 10A, 10B und 10C: Vergleich der medianen Antigenexpression und Streuungseigenschaften der für TTF prädiktiven XAI-Populationen mit den CLL-Zellen eines durchschnittlichen Patienten	100

Tabellenverzeichnis

Tabelle 1: Initiale Diagnostik der CLL (nach Hallek et al. ³³).....	22
Tabelle 2: Stadieneinteilung nach Binet.....	27
Tabelle 3: Rai-Klassifikation.....	28
Tabelle 4: Merkmale von Prognosemodellen bei CLL mit mindestens drei Validierungsstudien und Gesamtüberleben als Endpunkt (nach Molica ⁶⁷) ..	33
Tabelle 5: Schema der CLL-Erstlinientherapie (nach Hallek und Al-Sawaf ³²)	36
Tabelle 6: Schema der CLL-Zweitlinientherapie (nach Hallek und Al-Sawaf ³²)	37
Tabelle 7: Exemplarische Darstellung durchflusszytometrischer Rohdaten	45
Tabelle 8: Ausgewählte Machine-Learning-Algorithmen, die zur durchflusszytometrischen Analyse genutzt werden (nach Hu, Bhattacharya und Butte ⁴¹).....	47
Tabelle 9: Verwendete Antikörperkombinationen („Panels“) „T1“ und „T2“ mit Antikörpern „Duraclones“ der Firma Beckman Coulter.....	67

Tabelle 10: Patientencharakteristika.....	82
Tabelle 11: Vorhersagewerte der XAI-Populationen bezüglich des IPI-Scores	88
Tabelle 12: Vorhersagewerte der XAI-Populationen bezüglich TTF 1 und TTF 0	92
Tabelle 13: Zuordnung wichtiger XAI- Zellpopulationen.....	98
Tabelle 14: Vorhersagefähigkeit für das Ergebnis (TTF 0 vs. TTF 1) von drei Modellen	104

Abkürzungsverzeichnis

ALPODS	Algorithmic Population Description Approach
AUC	Area Under Curve
BCL2	B-Cell Lymphoma 2
BIRC3	Baculoviral IAP Repeat Containing protein3
BTK	Brutontyrosinkinase
CAR-T-Zellen	Chimäre-Antigen-Rezeptor- T-Zellen
CD	Cluster of Differentiation
CellCNN	Cell Convolutional Neural Network
CITRUS	Cluster Identification, Characterization, and Regression
CLL	Chronische lymphatische Leukämie
CLL-IPI	International Prognostic Index For Patients With

	Chronic Lymphocytic Leukaemia
CNN	Convolutional Neural Network
CTLA-4	Cytotoxic T-Lymphocyte- Associated Protein 4
Deep CNN	Deep Convolutional Neural Network
DGCyTOF	Deep learning with graphic cluster visualization
ECOG	Eastern Cooperative Oncology Group
FISH	Fluoreszenz-in-situ- Hybridisierung
FLOCK	FLOw Clustering without K
FloReMi	Flow Density Survival Regression Using Minimal Feature Redundancy
FlowSOM	Flow Self Organizing Map
FS	Vorwärtsstreuung
Hb	Hämoglobin

IGHV	Immunglobulin-variable-heavy-chain
iwCLL	International Workshop on Chronic Lymphocytic Leukemia
KI	Künstliche Intelligenz
LASSO	Least Absolute Shrinkage and Selection Operator
LDA	Linear Discriminant Analysis
LDT	Lymphozytenverdopplungszeit
MBL	Monoklonale B-Lymphozytose
MPFC	Multiparameter-Durchflusszytometrie
MRD	Minimale Resterkrankung
MRT	Magnetresonanztomographie
NK-Zellen	Natürliche Killerzellen
NOTCH1	Neurogenic Locus Notch Homolog Protein 1

PCA	Principle Component Analysis
PD-1	Programmed Cell Death Protein 1
PET	Positronen-Emissions- Tomografie
PI3K	Phosphatidylinositol 3- Kinase
ROC	Receiver Operation Characteristics
SDH	Smoothed Data Histograms
SF3B1	Splicing Factor 3b Subunit 1
SI	Simpson-Index
SLL	Kleinzellige, lymphozytäre Lymphome
SOM	Self Organizing Map
SS	Seitwärtsstreuung
TIGIT	T cell immunoreceptor with Ig and ITIM domains
tSNE	t-Distributed Stochastic Neighbor Embedding
TP53	Tumorprotein 53

TTF	Time-To-First-Line- Treatment-Failure
WHO	World Health Organization
XAI	Explainable Artificial Intelligence
ZAP-70	Zeta-Chain-Associated Protein Kinase 70

1. Einleitung

1.1 Chronische lymphatische Leukämie

1.1.1 Epidemiologie

Die Chronische lymphatische Leukämie (CLL) ist die häufigste leukämische Erkrankung in den westlichen Ländern¹¹⁶, wobei sie in Deutschland mit einer Inzidenz von etwa 3-5 pro 100.000 Einwohner auftritt.⁷² In Deutschland versterben jährlich etwa 1000 Männer und 850 Frauen an den Folgen einer CLL.¹²⁷ Die World Health Organization (WHO)-Klassifikation von 2022 ordnet die CLL den prä-neoplastischen und neoplastischen klein-lymphozytären Neoplasien zu, einer Untergruppe der reifen B-Zell-Neoplasien.⁵

1.1.2 Diagnosestellung der CLL

Der International Workshop on Chronic Lymphocytic Leukemia (iwCLL) hat in ihren letzten

Konsensleitlinien von 2018 Empfehlungen für die diagnostischen Untersuchungen bei Patienten mit CLL erarbeitet.³⁹ Die Empfehlungen zur Diagnostik in der klinischen Routine und im Rahmen von klinischen Studien sind in *Tabelle 1* aufgeführt. Dabei empfiehlt die iwCLL zur Diagnosestellung der CLL die Durchführung eines Differentialblutbildes zur morphologischen Begutachtung sowie die Bestimmung eines Immunphänotyps zur Abgrenzung gegenüber anderen Lymphomen mit kleinzelliger Morphologie. Die Ermittlung des Immunphänotyps erfolgt mittels Multiparameterdurchflusszytometrie (MPFC).^{70,46} In seltenen Fällen können genetische Untersuchungen bei der Differentialdiagnostik eine Rolle spielen.³⁹

Morphologie

Für die Diagnose einer CLL müssen mindestens drei Monate lang $\geq 5 \times 10^9/L$ B-Lymphozyten im peripheren Blut nachgewiesen werden können. Im Blutausschlag zeigen sich die CLL-Zellen

typischerweise als kleine, reife Lymphozyten mit einem schmalen Zytoplasmasaum, einem dichten Zellkern ohne erkennbare Nukleoli und mit teilweise aggregiertem Chromatin. Charakteristische Begleiterscheinungen bei CLL sind Gumprecht-Kernschatten oder Smudge-Zellen, die als Zelltrümmer identifiziert werden können.³⁹ Es kann eine geringe Anzahl von größeren, atypischen Zellen oder auch Prolymphozyten mit den typischen CLL-Zellen vermischt sein. Übersteigt der Anteil der Prolymphozyten jedoch $\geq 15\%$ spricht man von einer prolymphozytischen Variante der CLL (WHO 2022⁵). Bei einem Anteil von $\geq 10\%$ Prolymphozyten an allen zirkulierenden Zellen deutet dies darauf hin, dass die CLL mutmaßlich aggressiver ist (mit NOTCH1- oder TP53-Mutationen).⁸⁶

Differentialdiagnostisch ist es wichtig die CLL von einer monoklonalen B-Lymphozytose (MBL) abzugrenzen: diese liegt vor, wenn ein Patient keinerlei CLL-typische Symptome zeigt (keine

Lymphadenopathie, keine Organomegalie, keine weiteren assoziierten Zytopenien oder Symptome), ein absoluter Anstieg der klonalen B-Lymphozyten besteht, jedoch $<5 \times 10^9/L$ B-Lymphozyten im peripheren Blut nachweisbar sind.⁶⁸ Sobald eine Zytopenie besteht, die durch ein typisches Knochenmarksinfiltrat verursacht wird, spricht man von einer CLL, unabhängig von der Anzahl der B-Lymphozyten im peripheren Blut oder Befall der Lymphknoten. In 1-2% der Patienten pro Jahr entwickelt sich eine MBL zu einer CLL.^{95,115} MBL- und CLL-Patienten scheinen beide ein erhöhtes Risiko für sekundäre Krebserkrankungen zu haben, insbesondere der Haut¹¹⁵ (Melanome und Nicht-Melanome wie Basalzellkarzinome oder Plattenepithelkarzinome^{67,111}).

Das kleinzellige lymphozytäre Lymphom (SLL) entspricht biologisch und immunphänotypisch CLL-Zellen, die jedoch ohne wesentliche Ausschwemmung auf Lymphknoten begrenzt sind: die Diagnosestellung erfordert eine

Lymphadenopathie, das Fehlen von Zytopenien, die durch klonale Knochenmarksinfiltrate verursacht sind sowie die Anzahl von klonalen B-Lymphozyten im peripheren Blut $<5 \times 10^9/L$.³⁹

Immunphäotypisierung

CLL-Zellen koexprimieren das Oberflächenantigen CD5 zusammen mit den B-Zell-Antigenen CD19, CD20 und CD23. Im Vergleich zu normalen B-Zellen ist die Expression von den Oberflächen-Immunglobulinen CD20 und CD79b charakteristisch vermindert.^{70,80,34} Wie bei anderen reifen Lymphomen ist der Klon der CLL-Zelle im Regelfall auf die Expression einer Leichtkette beschränkt (κ oder λ).³⁴ Die Expression von CD5 kann jedoch auch bei anderen lymphatischen Malignomen beobachtet werden, wie z. B. dem Mantelzell-Lymphom.⁸¹ In der Regel reicht ein Panel von CD5, CD19, CD20, CD23, Ig κ oder Ig λ zur Diagnosestellung aus.⁹⁶ In Grenzfällen können zusätzliche Marker wie CD43, CD79b, CD81,

CD200, CD10 oder ROR1 die Diagnosestellung präzisieren.⁹⁶

Molekulargenetik

Molekulargenetisch lässt sich durch Anwendung einer Fluoreszenz-in-situ-Hybridisierung (FISH) auf periphere Lymphozyten in > 80% der CLL-Fälle eine zytogenetische Auffälligkeit finden.²⁵ Die häufigste chromosomale Veränderung bei der CLL ist die Deletion des langen Arms von Chromosom 13 (del(13q)) wobei die Trisomie von Chromosom 12 sowie Deletionen im langen Arm von Chromosom 11 (del(11q)) und im kurzen Arm des Chromosoms 17 (del(17p)) weitere häufige Aberrationen sind.²⁵ Die FISH und die konventionelle Karyotypisierung können bei der Differenzialdiagnostik verschiedener lymphoproliferativen Erkrankungen eine Rolle spielen.³⁹

Tabelle 1: Initiale Diagnostik der CLL (nach Hallek et al.³⁹)

Diagnostik	Allgemeine Praxis	Klinische Studien
Tests zur Diagnosestellung		
Großes Blutbild	Immer	Immer
Immunphänotypisierung der peripheren Blutlymphozyten	Immer	Immer
Untersuchung vor Therapieeinleitung		
Anamnese, körperliche Untersuchung, Leistungsstatus	Immer	Immer
Großes Blutbild und Differentialblutbild	Immer	Immer
Knochenmarkspunktion und -biopsie	Wenn klinisch indiziert (unklare Zytopenie)	Wünschenswert
Serumchemie, Serumimmunoglobulin und direkter Antiglobulintest	Immer	Immer
Röntgen-Thorax	Immer	Immer
Infektionskrankheitsstatus	Immer	Immer
Zusätzliche Tests vor der Behandlung		
Molekulargenetik (FISH) für del(13q), del(11q), del(17p), add(12) in Lymphozyten des peripheren Bluts	Immer	Immer
Konventionelle Karyotypisierung in Lymphozyten des peripheren Bluts (mit spezifischer Stimulation)	Nicht allgemein indiziert	Wünschenswert
TP53-Mutation	Immer	Immer
IGHV-Mutationsstatus	Immer	Immer
Serum-β2-Mikroglobulin	Wünschenswert	Wünschenswert
CT-Thorax, -Abdomen, - Becken	Nicht allgemein indiziert	Wünschenswert
MRT, PET	Nicht allgemein indiziert	Nicht allgemein indiziert
Ultraschall des Abdomens	möglich	Nicht allgemein indiziert

Es gibt weitere diagnostische Tests (Molekulargenetik, IGHV, weitere Immunphänotypisierung, Serummarker, Knochenmarksuntersuchungen, klinische Merkmale), die zur Diagnosestellung nicht

essentiell sind, jedoch helfen die Prognose vorherzusagen (siehe Kapitel 1.1.3).

Kriterien einer symptomatischen oder aktiven CLL

Gemäß iwCLL-Leitlinien³⁹ gibt es mehrere Kriterien zur Definition einer symptomatischen oder aktiven CLL-Erkrankung, welche eine Indikation zur Therapieeinleitung sind:

1. Anzeichen eines fortschreitenden Knochenmarkversagens, das sich in der Entwicklung oder Verschlimmerung einer Anämie und/oder Thrombozytopenie äußert. Grenzwerte von Hb <10 g/dL oder Thrombozytenzahlen von $<100\ 000/\mu\text{L}$ werden im Allgemeinen als Indikation für eine Behandlung angesehen. Es sollte jedoch darauf hingewiesen werden, dass bei einigen Patienten Thrombozytenzahlen von $<100\ 000/\mu\text{L}$ über

einen langen Zeitraum stabil bleiben können; diese Situation erfordert nicht automatisch eine Therapie

2. Massive (d. h. ≥ 6 cm unterhalb des linken Rippenbogens) oder progressive oder symptomatische Splenomegalie
3. Massive Knoten (d. h. ≥ 10 cm im längsten Durchmesser) oder progressive oder symptomatische Lymphadenopathie
4. Progressive Lymphozytose mit einem Anstieg von $\geq 50\%$ über einen Zeitraum von zwei Monaten oder eine Lymphozytenverdopplungszeit (LDT) von weniger als sechs Monaten. Die LDT kann durch lineare Regressionsextrapolation der absoluten Lymphozytenzahlen ermittelt werden, die in Abständen von zwei Wochen über einen Beobachtungszeitraum von zwei bis drei Monaten ermittelt werden; bei Patienten mit anfänglichen Lymphozytenzahlen von weniger als $30\,000/\mu\text{L}$ im Blut ist möglicherweise ein längerer Beobachtungszeitraum erforderlich, um

die LDT zu bestimmen. Andere Faktoren als die CLL, die zur Lymphozytose beitragen (z. B. Infektionen, Steroidverabreichung), sollten ausgeschlossen werden

5. Autoimmunkomplikationen einschließlich Anämie oder einer Thrombozytopenie, die schlecht auf Kortikosteroide ansprechen
6. Symptomatische oder funktionelle extranodale Beteiligung (z. B. Haut, Niere, Lunge, Wirbelsäule)
7. Krankheitsbedingte Symptome im Sinne einer der folgenden Definitionen:
 - a. Unbeabsichtigter Gewichtsverlust $\geq 10\%$ innerhalb der letzten sechs Monate
 - b. Erhebliche Müdigkeit (d.h. Eastern Cooperative Oncology Group Performance Status (ECOG PS)⁸⁵ 2 oder schlechter; kann nicht arbeiten oder ist nicht in der Lage, die üblichen Aktivitäten durchzuführen)
 - c. Fieber $\geq 100,5$ °F oder $38,0$ °C für zwei oder mehr Wochen ohne Nachweis einer Infektion

- d. Nachtschweiß für ≥ 1 Monat ohne Nachweis einer Infektion

1.1.3 Prognose der CLL

Die Prognose einer CLL ist sehr heterogen. Ein Teil der Patienten benötigt keine Therapie, bei einem anderen Teil kann die CLL sehr schnell und aggressiv fortschreiten. Dabei kann es unter anderem zur sogenannten Richter-Transformation kommen, der Umwandlung der CLL in ein aggressives und hochgradiges Lymphom mit einem medianen Überleben von nur 1-2 Jahren.^{88,101}

Klassifikationssysteme Binet und Rai

Traditionell gibt es zwei CLL-Klassifikationssysteme auf deren Basis die Tumorlast bestimmt wird, die zumindest historisch in Patientenkohorten unterteilt und eine erste Einschätzung der Prognose erlaubt (Klassifikation nach Binet¹¹ in *Tabelle 2* sowie Rai-Klassifikation⁹³

in *Tabelle 3*). Berücksichtigt werden hierbei die Anzahl der Lymphozyten sowie Thrombozyten, eine vorliegende Lymphadenopathie mit der Anzahl betroffener Lymphknotenregionen, eine mögliche Hepato- und/oder Splenomegalie und der Hämoglobin-Wert.^{11,93} Vorteile dieser beiden Staging-Systeme sind die einfache, kostengünstige sowie weltweit anwendbare und damit einhergehend gut vergleichbare Anwendungsmöglichkeit.³⁹

Tabelle 2: Stadieneinteilung nach Binet

Stadieneinteilung nach Binet ¹¹			
Stadium	Vergrößerte Lymphknoten-/Organregionen*	Hämoglobin	Thrombozyten
A	< 3 Regionen befallen	≥ 10 g/dl	≥ 100 000/μl
B	≥ 3 Regionen befallen	≥ 10 g/dl	≥ 100 000/μl
C	Unabhängig	< 10 g/dl	und/oder < 100 000/μl

* Als Region gelten jeweils die Lymphknoten im Hals-, Achsel- oder Leistenbereich (ein- oder beidseitig) sowie die Leber und die Milz. Lymphknoten gelten als vergrößert, wenn sie einen Durchmesser von mehr als 1cm haben. Entscheidend ist der Befund der körperlichen Untersuchung, nicht die Befunde der Bildgebung, also z.B. Ultraschall- oder CT-Untersuchungen.

Tabelle 3: Rai-Klassifikation

Rai-Klassifikation ⁹³				
Stadium	Lymphadenopathie	Hepato- oder Splenomegalie	Hämoglobin	Thrombozyten
0	Nein	Nein	≥ 11 g/dl	≥ 100 000/μl
I	Ja	Nein	≥ 11 g/dl	≥ 100 000/μl
II	Irrelevant	Ja	≥ 11 g/dl	≥ 100 000/μl
III	Irrelevant	Irrelevant	< 11 g/dl	≥ 100 000/μl
IV	Irrelevant	Irrelevant	Irrelevant	< 100 000/μl

Wichtige Prognosemarker

Neben dem Tumorstadium bestimmen verschiedene weitere Faktoren die Prognose: genetische Faktoren wie eine TP53-Mutation, die 17p- oder 11q-Deletion gehen mit einer schlechteren Prognose einher während eine 13q14-Deletion und Trisomie 12 mit einer besseren Prognose einhergehen.^{25,20,144} Mutationen in NOTCH1-, SF3B1- und BIRC3-Genen sind assoziiert mit einer geringeren Lebenserwartung.^{136,100,92,9} Eine Expression von CD38 und ZAP-70 auf CLL-Zellen ist mit einer unmutierten Immunglobulin-variable-heavy-chain (IGHV) und höheren Werten von beta2-

Mikroglobulin assoziiert, was eine ungünstigere Prognose zur Folge hat.^{22,42,41,57,94,107,44}

Patienten mit 17p-Deletion und/oder TP53-Mutation sprechen schlecht auf Chemo- und Immuntherapien an, profitieren jedoch deutlich von einer Therapie mit nicht-chemotherapeutischen Wirkstoffen, wie beispielsweise Brutontyrosinkinase (BTK)-, Phosphatidylinositol-3 (PI3K)-Kinase- oder B-Cell-Lymphoma-2 (BCL2)-Inhibitoren.³⁹ Es konnte gezeigt werden, dass das progressionsfreie Überleben von CLL-Patienten mit einer 17p-Deletion und Patienten mit einer TP53-Mutation ohne 17p-Deletion ähnlich sind.¹⁴⁵

Zeigen sich neben mutierten IGHV-Genen zusätzliche positive Prognosefaktoren wie beispielsweise eine günstige Zytogenetik (u.a. keine 11q- oder 17p-Deletion¹⁰²), dann sprechen die Patienten gut und lange auf eine Chemoimmuntherapie mit Fludarabin,

Cyclophosphamid und Rituximab an.^{102,32,118} Ungefähr ein Drittel der CLL-Patienten kann basierend auf gemeinsamen Sequenzen innerhalb der IGHV-Region 3 in Subgruppen eingeteilt werden.¹¹³ Diese Subgruppen zeigen teilweise ähnliche Prognosen, beispielsweise kann die Expression des IGHV3-21 Gens (des stereotypen Subsets 2) unabhängig vom IGHV-Mutationsstatus mit einer schlechteren Prognose einhergehen.^{119,8}

Im Rahmen einer großen, multizentrischen Studie zeigte sich CD49d (α-Kette des α4β1-Integrin-Heterodimer) mit einer ungünstigen Prognose assoziiert und als stärkster durchflusszytometrischer Prädiktor für das Gesamtüberleben und das behandlungsfreie Überleben.^{14,21}

Zusätzlich zum β2-Mikroglobulin, das sich als unabhängiger prognostischer Marker bereits in verschiedenen Prognosescores^{117,138,24,112} etabliert hat (siehe *Tabelle 4*), sind auch die Serummarker

Thymidinkinase sowie das lösliche CD23 mit einem niedrigeren Gesamt- und progressionsfreiem Überleben assoziiert.^{40,97,106,64,69}

Prognosemodelle bei CLL

2016 wurde der sogenannte International Prognostic Index For Patients With Chronic Lymphocytic Leukaemia (CLL-IPI) eingeführt mit dem eine Unterteilung in vier prognostische Subgruppen gelingt. Die Einteilung basiert auf fünf unabhängigen Faktoren: TP53-Status, IGHV-Mutationsstatus, beta2-Mikroglobulin im Serum, klinisches Stadium und das Alter. Mittels CLL-IPI erfolgt eine Einteilung von Patienten in vier Risikogruppen, bei denen das 5-Jahres-Gesamtüberleben von 93,2% (geringes Risiko) bis 23,3% (sehr hohes Risiko) liegen kann. Dabei entspricht ein Score von 0-1 einem niedrigen Risiko mit 5-Jahres-Gesamtüberleben von 93,2%, Score 2-3 einem mittleren Risiko mit 5-Jahres-Gesamtüberleben von 79,3%, Score 4-6 einem hohen Risiko mit einem 5-Jahres-Gesamtüberleben

von 63,3% und Score 7-10 einem sehr hohen Risiko mit 5-Jahres-Gesamtüberleben von 23,3%.¹¹⁷ Ziel des CLL-IPI ist anhand weniger Parameter Patienten einer Prognosegruppe zuzuordnen.

Weitere prognostische Modelle unterscheiden sich in der Anzahl der berücksichtigten Parameter und sind beispielsweise der 2007 MDACC oder B-B Score (siehe *Tabelle 4*).⁷⁹

Tabelle 4: Merkmale von Prognosemodellen bei CLL mit mindestens drei Validierungsstudien und Gesamtüberleben als Endpunkt (nach Molica⁷⁹)

Prognose-Modell	Variablen	Analysierter Endpunkt	Patientenanzahl in Entwicklungs- sowie Validierungsstudien
2007 MDACC (MD Anderson Cancer Center) ¹³⁸	<ul style="list-style-type: none"> • Alter • Absolute Lymphozytenzahl • $\beta 2$-Mikroglobulin • Lymphknoten-Gruppen • Rai-Stadium • Geschlecht 	Gesamtüberleben	7246
CLL-IPI ¹¹⁷	<ul style="list-style-type: none"> • Alter • $\beta 2$-Mikroglobulin • IGHV-Mutationsstatus • Rai- oder Binet-Stadium • TP53-Status 	Gesamtüberleben	6003
B-B (Barcelona-Brno) ²⁴	<ul style="list-style-type: none"> • IGHV-Mutationsstatus • FISH-Zytogenetik 11q-Deletion und/oder 17p-Deletion 	Gesamtüberleben	2696
BALL ¹¹²	<ul style="list-style-type: none"> • $\beta 2$-Mikroglobulin • LDH • Hämoglobin • Beginn der letzten Therapie < 24 Monate 	Gesamtüberleben	2475

1.1.4 Therapie der CLL

Zur Therapie der CLL kann auf vielfältige Medikamentengruppen zurückgegriffen werden, darunter fallen Zytostatika (Chlorambucil, die

Purinanaloga Fludarabin, Pentostatin und Cladribin, Bendamustin), monoklonale Antikörper (Anti-CD20-Antikörper Rituximab, Ofatumumab, Obinutuzumab oder der Anti-CD52-Antikörper Alemtuzumab), Medikamente die auf den Signalweg der CLL-Zellen und deren Mikroumgebung abzielen (PI3K-Inhibitoren wie Idelalisib, Duvelisib oder Umbralisib sowie BTK-Inhibitoren wie Ibrutinib, Acalabrutinib, Pirtobrutinib, Zanubrutinib), Lenalidomid als Thalidomidanalogon, BCL2-Inhibitoren wie Venetoclax, Checkpoint-Inhibitoren wie Pembrolizumab als PD-1-blockierender Antikörper sowie CAR-T-Zell-Therapien. Ein wichtiges Grundprinzip der immer effizienter werdenden CLL-Therapie besteht in der Kombinationstherapie dieser Medikamente, um synergistische Effekte zu nutzen, möglichst ohne überlappende toxische Nebenwirkungen.³⁸

Bevor ein Therapieansatz vorgeschlagen wird, müssen folgende Parameter berücksichtigt werden:

Stadium der Erkrankung, die Symptome, Fitnesszustand und Begleiterkrankungen des Patienten („go go“ steht für einen guten, „slow go“ für einen eingeschränkten körperlichen Zustand) insbesondere im Hinblick auf die potenzielle Organtoxizität der neueren und zielgerichteteren Wirkstoffe, das genetische Risiko der Leukämie, die aktuelle Behandlungssituation (Erst- oder Zweitlinientherapie, Ansprechen oder Nichtansprechen auf die letzte Behandlung).³⁸ *Tabelle 5* und *6* illustrieren die aktuellen Empfehlungen zur Erst- und Zweitlinientherapie bei CLL.

Tabelle 5: Schema der CLL-Erstlinientherapie (nach Hallek und Al-Sawaf³⁸)

Stadium	Del(17p) oder TP53mut	Fitness	IGHV	Therapie
Inaktive Erkrankung, Binet A-B, Rai 0-II	Irrelevant	Irrelevant	Irrelevant	Keine
Aktive Erkrankung oder Binet C oder Rai III-IV	Ja	Irrelevant	Irrelevant	Ibrutinib / Acalabrutinib ¹ oder Venetoclax + Obinutuzumab oder Idelalisib-Rituximab (falls Kontraindikationen für andere Optionen)
	Nein	Go go	M	FCR (BR > 65 Jahre) oder Ibrutinib / Acalabrutinib ¹ oder Venetoclax + Obinutuzumab ²
			U	Ibrutinib / Acalabrutinib ¹ oder Venetoclax + Obinutuzumab oder FCR (BR > 65 Jahre)
		Slow Go	M	Venetoclax + Obinutuzumab oder Ibrutinib / Acalabrutinib ^{1,2} oder Chlorambucil-Obinutuzumab
			U	Venetoclax + Obinutuzumab oder Ibrutinib / Acalabrutinib ^{1,2} oder Chlorambucil-Obinutuzumab

M: Mutiert; U: Unmutiert; FCR: Fludarabin, Cyclophosphamid, Rituximab; BR: Bendamustin + Rituximab

¹ Hinzufügen von Obinutuzumab zu Acalabrutinib erwägen

² Erwägen und mit dem Patienten diskutieren: Kontinuierliche vs. Therapie mit fester Dauer, spezifische Nebenwirkungen von Medikamentengruppen

(Myelosuppression, Infektionen, maligne Zweitneoplasien bei Chemoimmuntherapien, kardiale Toxizität und Blutungen für BTK-Inhibitoren (Acalabrutinib < Ibrutinib), Tumorlysesyndrom und Infektionen für Venetoclax + Obinutuzumab; Autoimmunerkrankungen und opportunistische Infektionen für Idelalisib

Tabelle 6: Schema der CLL-Zweitlinientherapie (nach Hallek und Al-Sawaf⁶³⁸)

Ansprechen auf Erstlinientherapie	Fitness	Therapie
Rückfall oder Progress innerhalb von drei Jahren	Go go	Wechsel auf: Venetoclax + Rituximab oder Ibrutinib oder Acalabrutinib Weitere Optionen: Idealisib + R, FA, FCR (nach BR), Venetoclax, A-Dex, Lenalidomid (+ R), BR (nach FCR). Diskussion einer Konsolidierung mit allogener Stammzelltherapie.
	Slow Go	Wechsel auf: Venetoclax + Rituximab oder Ibrutinib oder Acalabrutinib Weitere Optionen: Idealisib + R, A, FCR-lite, BR, Lenalidomid (+ R), Ofatumumab, HD-R
Progress nach drei Jahren	Alle	Wiederholung der Erstlinientherapie kann erwogen werden, falls zuvor Chemotherapie genutzt wurde, jedoch Wechsel auf zielgerichtete Therapie

A-Dex: Alemtuzumab-Dexamethason; B: Bendamustin;
FA: Fludarabin-Alemtuzumab; FCR: Fludarabin-Cyclophosphamid-Rituximab; HD: hochdosiert; R: Rituximab

1.1.5 Relevanz des Themas

Die CLL hat insbesondere bei älteren Patienten eine hohe Prävalenz. Aufgrund der sehr heterogenen Prognose ist es wichtig, dass es aufgrund einer zu restriktiven Einschätzung der Prognose anhand der wenigen Parameter in den klassischen Prognosescores nicht zu einer Übertherapie kommt. Auch innerhalb der klassischen Prognosegruppen wäre es wünschenswert, wenn die komplexe heterogene Biologie der CLL durch eine Vielzahl an Parametern individueller erfasst wird.

Komplexe durchflusszytometrische Daten werden anders als z.B. genetische Daten fast bei jedem Patienten mit CLL erhoben, da sie entscheidend bei der Diagnosestellung sind. Durchflusszytometrische Datensätze sind hochindividuell und beinhalten neben Informationen zum Immunphänotyp der Leukämiezellen auch Informationen über die reaktiven, nicht malignen Immunzellen.

In dieser Arbeit wird daher die Frage gestellt, ob es in den durchflusszytometrischen Daten bisher nicht berücksichtigte Informationen gibt, welche die genetische, pathophysiologische und prognostische Heterogenität einer CLL besser abbilden können als bisherige Prognosewerkzeuge und damit eine genauere Prognose ermöglichen.

1.2 Der durchflusszytometrische Datensatz

Die Immunphänotypsierung mittels Durchflusszytometrie wird zur Diagnosestellung, Verlaufsbeobachtung und auch zur Bestimmung der messbaren oder minimalen Resterkrankung (MRD) eingesetzt.

Der aktuelle Standard zur Analyse von durchflusszytometrischen Daten ist das manuelle „Gating“, bei dem man eine Reihe zweidimensionaler Diagramme zur Visualisierung der Daten sowie hierarchische Gates zur

Identifizierung von Zellpopulationen nutzt. Dabei wird in dieser Dissertation der Begriff Gating aufgrund eines fehlenden deutschen prägnanten Begriffs zur Beschreibung dieses Prozesses verwendet.

Nachdem die Blutprobe eines Patienten mit Hilfe des Durchflusszytometrie-Geräts prozessiert wurde, analysiert der hämatologische Experte die Zellen eines Patienten zumeist in Form von sog. Dot-Plots (Synonym: Punktwolkendarstellungen). Ein Dot-Plot beschreibt die zweidimensionale Darstellung von Messwerten, bei der jede Zelle entsprechend ihren x- und y-Signalintensitäten als einzelner Punkt eingezeichnet wird. Diese Dot-Plots können nun Schritt für Schritt in „Gates“ unterteilt werden, d.h. in Regionen oder eine Kombination von Regionen, die zur Auswahl von Zellen in einer erneuten Darstellung dienen.¹⁰³

In *Abbildung 1A-F* sieht man den charakteristischen Prozess des manuellen Gatings bei der CLL-

Diagnostik. Dabei grenzt ein Experte die einzelnen Zellpopulationen händisch mit Hilfe von Trennungslinien bzw. geometrischen Formen voneinander ab. In diesem Beispiel werden 95533 Zellen („Zelluläre Events“) in mehreren Schritten so voneinander abgegrenzt, dass sich bei Vorliegen einer CLL am Ende des Prozesses typischerweise monoklonale, CD19+, CD20+, CD5+ und CD23+ B-Lymphozyten zeigen.

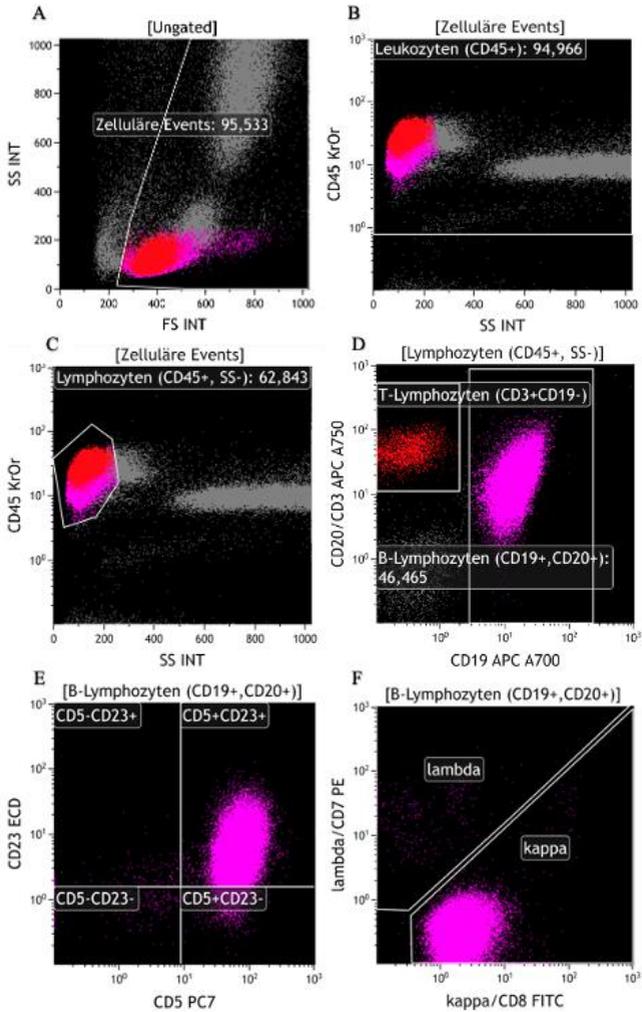


Abbildung 1 A-F: Beispielhaftes CLL-Gating

IA Alle Zellen der Blutprobe eines CLL-Patienten (95533 „zelluläre Events“).

IB Abgrenzung von Leukozyten (diese zeichnen sich durch eine geringere Größe und eine geringe Granularität aus).

IC Gating nach CD45+ Leukozyten mit einer geringen Granularität (CD45+, SS-).

ID Da die CLL unter die B-Zell-Lymphome fällt, erfolgt nun die Unterscheidung zwischen T- und B-Lymphozyten: leukämische B-Lymphozyten exprimieren CD19+ und CD20+, wohingegen T-Lymphozyten CD19- und CD3+ exprimieren.

IE CLL-Zellen koexprimieren das Oberflächenantigen CD5 zusammen mit den B-Zell-Antigenen CD19, CD20 und CD23, also filtert man die CD19+ und CD20+ leukämischen B-Lymphozyten auf CD5 und CD23, um diese innerhalb der B-Zell-Leukämien zu differenzieren.

IF Nachweis der CLL-charakteristischen Monoklonalität der Lymphozyten durch Nachweis einer Leichtkettenrestriktion (Ig κ oder Ig λ), was vorzugsweise durch Doppelmarkierung erfolgt (hier mit Ig κ /CD8 und Ig λ /CD7). Zeigen sich die bisher gefilterten B-Lymphozyten (CD19+, CD20+,

CD5+, CD23+) nun auch noch monoklonal, so ist das typisch für eine CLL.

Der individuelle Gating-Prozess macht deutlich wie abhängig die Diagnostik von der Erfahrung der Experten ist. Jede Zelle weist eine individuelle Kombination der analysierten Parameter auf (Größe, Granularität, Oberflächenantigene, Leichtketten, etc.). Es gibt also multiple, potenziell relevante Dimensionen, in denen man eine Zelle betrachten kann, d.h. letztendlich ist ein zelluläres Event eine Koordinate mit X-Dimensionen. Jedoch erlaubt das Gating lediglich eine zweidimensionale Darstellung der Zellen in den Dot-Plots, sodass es im Gegensatz zu einer multidimensionalen und ganzheitlichen Analyse zu einem Informationsverlust kommt.

In *Tabelle 7* wird ein Überblick über die aufgenommenen Rohdaten bei der Durchflusszytometrie gegeben. Jede Zeile in *Tabelle 7* beschreibt ein zelluläres Event mit Expression der gleichzeitig gemessenen Parameter.

Die einzelnen Kästchen der Tabelle beschreiben die Ausprägung eines Parameters in absoluten Werten. Neben der Anzahl der gemessenen Parameter, spielt auch die Anzahl der gemessenen zellulären Events eine Rolle, diese liegt häufig bei mehr als 10^5 . Hierdurch können auch kleinere Zellpopulationen in durchflusszytometrischen Daten bewertet werden.

Tabelle 7: Exemplarische Darstellung durchflusszytometrischer Rohdaten

FS INT	SS INT	kappa/CD8 FITC	lambda/CD7 PE	CD23 ECD	CD79b/CD4 PCS.5	CD5 PC7	CD38 APC	CD19 APC A700	CD20/C D3 APC A750	FMC7/CD2 PB	CD45 KrOr	TIME
389694	83440	2168,954	424,5437	7099,759	24090,57	45811,11	14274,42	4931,434	20287,52	306,0995	11294,42	1
403454	86800	1686,052	580,1818	15205	16594,6	58623,84	4608,263	18990,48	21125,04	521,9395	14973,94	12
378393	86852	3313,233	293,406	3598,635	31914,35	85191,91	4328,751	24016,36	25747,51	675,4004	22235,39	13
376951	84573	1441,043	89,17203	2809,73	16520,21	27912,52	19064,74	4834,637	4570,28	235,9285	8737,648	16
399839	66767	1871,18	167,4395	3790,986	16993,53	41415,46	1785,966	4834,637	7241,313	850,8527	9106,435	19
326559	86049	1354,931	113,6332	2543,828	11322,01	40767,44	5281,381	9392,916	9292,505	140,6217	9639,401	24
367025	71762	1152,98	327,7216	6396,49	8491,189	49690,53	3930,963	13276,68	22000,38	359,214	11944,06	26
455073	126895	5552,391	500,2131	11197,78	52212,73	140984	16594,59	43686,57	30308,58	1255,659	22487,33	27
398621	78339	1820,036	98,18158	3546,336	17153,16	48502,43	10492,4	8915,839	8481,861	363,3344	16015,87	29
327811	109723	1242,625	341,5334	7205,478	13389,83	77218,38	9574,314	12499,57	13245,54	316,0679	14331,7	34

In dieser Arbeit erfolgt die Analyse der durchflusszytometrischen Daten von diagnostischen B-Zell-Panels mit Hilfe von Machine-Learning-Verfahren und Algorithmen der Erklärbaren Künstlichen Intelligenz (XAI), um prognostische

Zellpopulationen für die CLL in den durchflusszytometrischen Daten zu finden.

1.3 Machine-Learning-Verfahren in der Hämatologie

In den letzten 20 Jahren kam es kontinuierlich zu einer steigenden Anwendung von Machine-Learning-Algorithmen auf durchflusszytometrische Datensätze.⁵³ In der Literatur werden verschiedene Algorithmen zur Analyse der Durchflusszytometrie-Daten beschrieben, deren Ziele ganz unterschiedlich sind: Insbesondere in der onkologischen und hämatologischen Forschung erfolgt die Anwendung häufig zur Reduktion der Dimensionalität, zum Datenclustering sowie zur Vorhersagemodellierung.⁵³ In *Tabelle 8* finden sich ausgewählte Machine-Learning-Algorithmen, die zur durchflusszytometrischen Analyse genutzt werden.

Tabelle 8: Ausgewählte Machine-Learning-Algorithmen, die zur durchflusszytometrischen Analyse genutzt werden (nach Hu, Bhattacharya und Butte⁴⁹)

Machine-Learning-Typ	Name	Beschreibung
Dimensionsreduktion	Principle Component Analysis (PCA) ²	PCA projiziert die hochdimensionalen Daten in niedrigere Dimensionen, wobei ein möglichst großer Teil der Variation der Daten erhalten bleibt.
	t-Distributed Stochastic Neighbor Embedding (tSNE) ⁵²	t-SNE ist ein nichtlineares Verfahren zur Dimensionalitätsreduktion. t-SNE transformiert die paarweisen Abstände in Wahrscheinlichkeiten, die auf der t-Verteilung basieren, und betont so die Erhaltung der lokalen Struktur der Daten.
	Uniform Manifold Approximation and Projection (UMAP) ⁷⁴	UMAP ist eine Methode zur Dimensionsreduzierung unter Verwendung von Techniken des Mannigfaltigkeitslernens. Ähnlich wie bei tSNE liegt bei UMAP der Schwerpunkt auf der Erhaltung der lokalen Struktur der Daten.
Unüberwachte Methoden zur Zellpopulations-Identifizierung	Flow Clustering without K (FLOCK) ²⁶	FLOCK identifiziert Zellpopulationen durch dichte-basiertes Clustering.
	Flow Self Organizing Map (FlowSOM) ¹³³	FlowSOM bildet Zellen auf selbstorganisierenden Karten ab und verwendet konsensbasiertes hierarchisches Clustering, um die Zellpopulationen zu identifizieren.
	PhenoGraph ³⁹	PhenoGraph konstruiert zunächst einen Graphen der nächsten Nachbarn der einzelnen Zellen auf der Grundlage ihrer phänotypischen Ähnlichkeit und teilt den Graph anschließend mit Hilfe eines Algorithmus in verschiedene Cluster auf.
Überwachte Methoden zur Zellpopulations-Identifizierung	Linear Discriminant Analysis (LDA) für Zytometriedaten ^{1,33,7}	Die Methode trainiert einen Klassifikator der LDA zur Identifizierung von Zellpopulationen.
	Deep learning with graphic cluster (DGCytoF) visualization ¹⁸	DGCytoF trainiert ein Deep-Learning-Modell, um Zellpopulationen zu identifizieren. Eine Rückkopplungsschleife sorgt für die Anpassung zwischen neuen und unbekanntem Zellpopulationen.
	Cluster Identification, Characterization, and Regression (CITRUS) ¹³	CITRUS verwendet hierarchisches Clustering zur Identifizierung einer großen Anzahl kleiner Zelluntergruppen von Zytometriedaten und verwendet ein „Least Absolute Shrinkage and Selection Operator“ (LASSO) ¹²⁵ -Modell um klinische Ergebnisse vorherzusagen.
Stichproben-Klassifizierung unter Verwendung von Zelluntergruppen-Informationen	Flow Density Survival Regression Using Minimal Feature Redundancy (FloReMi) ¹³⁴	FloReMi ist eine Pipeline für die Vorverarbeitung von Daten, die Identifizierung von Zelluntergruppen, die Auswahl von Merkmalen und die prädiktive Modellierung von Zytometriedaten unter Verwendung eines Random-Forest-Modells zur Vorhersage klinischer Ergebnisse.
Klassifizierung von Proben anhand von Einzelzelldaten	Cell Convolutional Neural Network (CellCNN) ⁶	CellCNN verwendet eine Faltungsstruktur für neuronale Netze zur Vorhersage klinischer oder biologischer Ergebnisse direkt anhand von Einzelzelldaten.
	Deep Convolutional Neural Network (Deep CNN) ⁵⁰	Deep CNN verwendet eine Faltungsnetzwerkstruktur, um klinische oder biologische Ergebnisse direkt anhand von Einzelzelldaten vorherzusagen. Das Modell enthält eine höhere Anzahl interner Schichten, so dass es die komplexen Wechselwirkungen zwischen Zellmarkierungen in den Zytometriedaten besser erfassen kann.

Bruggner et al.¹³ haben beispielsweise die Methode CITRUS entwickelt zur Identifizierung von Zellpopulationen in der multidimensionalen Durchflusszytometrie, deren Verhalten mit bestimmten Probenendpunkten korreliert. Dabei konnte CITRUS mehrere Zelluntergruppen identifizieren, die mit dem AIDS-freien Überleben bei HIV-infizierten Patienten in Verbindung gebracht werden.

Um die hochdimensionalen Daten fassbarer und interpretierbar zu machen, werden beispielsweise dimensionsreduzierende Techniken wie Self-Organizing Maps (SOM) genutzt.^{133,105} Eine SOM ist eine Art unüberwachtes künstliches neuronales Netz mit einem Netz von Knoten, die Punkte im mehrdimensionalen Raum darstellen.⁵⁶

Es gibt zwei verschiedene SOM-Anwendungen.⁵⁵ Man unterscheidet SOM mit wenigen Neuronen (k-means SOM) und SOM mit vielen Neuronen (Emergent SOM, ESOM¹²⁸). Bei dem Clusteralgorithmus K-means SOM werden die Neuronen mit Clustern im Datenraum identifiziert.

ESOM hingegen werden primär zur Visualisierung eines ansonsten unzugänglichen hochdimensionalen Datenraums genutzt. Charakteristisch für ESOM ist die große Anzahl der verwendeten Neuronen (mehrere Tausend). Sie ermöglichen die Herausbildung von intrinsischen Strukturmerkmalen eines Datenraumes und sind letztendlich selbstorganisierte Projektionen von hochdimensionalen Daten auf eine zweidimensionale Karte.¹²⁹

Zhao et al. haben SOMs um Convolutional Neural Networks (CNN) erweitert, um verschiedene Lymphom-Entitäten zu differenzieren.¹⁴⁷ Ein CNN beschreibt eine Deep-Learning-Netzarchitektur, die häufig für die Mustererkennung in medizinischen Bilddaten genutzt wird und eine Standardtechnik insbesondere für die Klassifikation von Bilddaten ist.⁶¹

Die bisherigen Anwendungsfälle von Machine-Learning-Algorithmen auf durchflusszytometrische

Datensätze bei einer CLL beschränken sich insbesondere auf die Diagnose.

Ng und Zuromski⁸² konnten mit Hilfe einer UMAP⁷⁴ zur Dimensionsreduktion der durchflusszytometrischen Daten und anschließender Random-Forest-Klassifikation¹² eine Diagnosegenauigkeit von mehr als 95% bei allen B-Zell-Malignomen erreichen, (Fragestellung: B-Zell-Malignom oder nicht; 2026 Trainingsfälle, 516 Testfälle) bei der Diagnose einer CLL sogar mit einer Genauigkeit von 96.90% (Fragestellung: CLL oder nicht; 491 Trainingsfälle, 99 Testfälle). Ein Random Forest ist ein Algorithmus bestehend aus einer Sammlung von mehreren Klassifikationsbäumen, welche unabhängig voneinander eine Art Stimme für die beliebteste Klasse abgeben können.¹²

Der aktuelle Stand der wissenschaftlichen Forschung zeigt, dass es insbesondere an Machine-Learning-Algorithmen fehlt, die die Menge an

durchflusszytometrischen Informationen nutzen, um die Genauigkeit der CLL-Prognose zu verbessern.

In dieser Arbeit wurde der XAI-Algorithmus Algorithmic Population Description Approach (ALPODS) genutzt, um prognostische Zellpopulationen in den durchflusszytometrischen Daten der CLL-Patienten zu finden. Die Entwicklung von ALPODS ist aus einer Kooperation mit Prof. Dr. Ultsch und PD Dr. Thrun vom Fachbereich Mathematik der Philipps-Universität Marburg entstanden.¹³⁰

Mit ALPODS wurden bereits Immunzellpopulationen visualisiert, welche zur Unterscheidung von Knochenmarksproben und peripheren Blutproben genutzt werden konnten.¹³⁰ Ultsch et al.¹³⁰ konnten in ihrer Arbeit zeigen, dass ALPODS für große ($n \geq 100.000$) und multivariate Daten ($d \geq 10$) im Vergleich zu anderen XAI-Algorithmen für dieselbe Aufgabe (Differenzierung von Blutproben nach den Materialquellen

Knochenmark vs. peripheres Blut) für den Menschen zugängliche und verständliche Ergebnisse liefern konnte. Sehr wichtig für den klinischen Alltag ist die Fähigkeit von ALPODS die komplexen hochdimensionalen Daten auf die essentiellen disjunkten Zellpopulationen zu reduzieren, denn nur diese Informationen können von Menschen genutzt werden und somit im klinischen Alltag unterstützen.

Darüber hinaus nutzen Hoffmann et al.⁴⁷ ALPODS um Zellpopulationen zu identifizieren, die es erlauben eine Vorhersage zur Hämodilution von Knochenmarksproben mit peripherem Blut zu treffen, denn dies kann je nach Aspirationsvolumen im Zuge der Diagnostik einer MRD bei Patienten mit akuter myeloischer Leukämie zu falsch-negativen Ergebnissen führen.

1.4 Explainable Artificial Intelligence (XAI)

Für Entscheidungsträger deren Entscheidungen sich auf eine Künstliche Intelligenz (KI) stützen ist die Erklärbarkeit einer KI essentiell, denn die Entscheidung muss gerechtfertigt und begründbar sein. Basiert ein Algorithmus auf einem einfachen statistischen Modell so können Entscheidungsträger dies verstehen und ihre Entscheidung rechtfertigen. Bei Systemen des Machine und Deep Learnings ist dies heutzutage jedoch meist nicht möglich.⁴⁵ Diese Differenz zwischen KI-Modellen und dem menschlichen Verständnis wird „Black Box“ genannt.¹³⁵

Viele KI-Systeme sind subsymbolisch, insbesondere wenn sie auf neuronalen Netzen beruhen.^{15,120,124} Diese subsymbolischen Systeme basieren auf der korrelativen Analyse großer Datenmengen und nicht auf der Externalisierung von Schlussfolgerungswissen, was sie zu „Black Boxes“⁴⁴ macht, d.h. zu Systemen deren interne Repräsentation von Wissen nicht ohne Weiteres

externalisiert und den Nutzern zugänglich gemacht werden kann.¹⁵ Sie liefern also weder Erklärungen für ihre Entscheidungen in Form von Wissen und Argumenten, noch begründen sie ihre Entscheidungen.^{121,35}

In der Medizin ist das Verständnis der Gründe für eine Entscheidung essentiell und Voraussetzung für eine vertrauenswürdige Interaktion zwischen KI-Systemen und Ärzten. Aufgrunddessen verlangen die Gesetze der Europäischen Union eine Begründung für die von KI-Systemen getroffenen Entscheidungen.^{114,48} Die Wichtigkeit dieser Erklärbarkeit zeigt sich auch in der deutlichen Zunahme von Forschung zu dem Thema XAI. Zusammengefasst besteht das Hauptziel der XAI-Modelle darin, dass Menschen dem Modell vertrauen können.¹⁷ Zur Förderung des Vertrauens hat die Europäische Union sieben essentielle Anforderungen an KI-Modelle vorgeschlagen: (1) menschliches Handeln und Aufsicht, (2) technische Robustheit und Sicherheit, (3) Datenschutz, (4) Transparenz, (5) Vielfalt, Nicht-Diskriminierung

und Fairness; (6) soziales und ökologisches Wohlergehen sowie (7) Rechenschaftspflicht.¹¹⁰

Abbildung 2 (angelehnt an Chaddad et al.¹⁷) stellt eine Möglichkeit dar, um verschiedene Klassifizierungskriterien von XAI-Methoden zusammenzufassen.^{60,27,141}

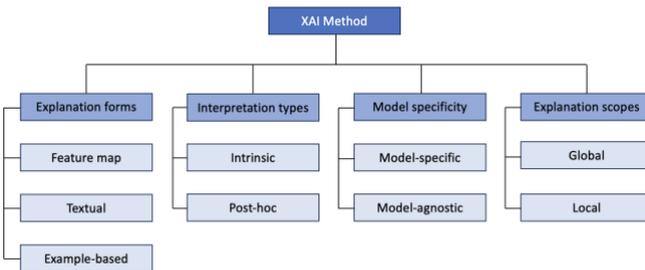


Abbildung 2 (angelehnt an Chaddad et al.¹⁷): Kategorisierung verschiedener XAI-Methoden

Interpretierbarkeit

Die Hauptunterscheidung von XAI-Modellen erfolgt nach Interpretierbarkeit in intrinsische und Post-hoc Modelle.^{3,17,23} Intrinsische Modelle werden so strukturiert, dass sie selbst interpretierbar sind ohne zusätzliche Informationen, im Gegensatz dazu

bieten post-hoc Modelle nachträgliche Erklärungen für bereits trainierte Modelle.^{17,146} Bei intrinsischen Modellen ist die Erklärung in den KI-Algorithmus integriert und in der Regel nicht auf andere Architekturen transferierbar. Bei Post-hoc Modellen ist der XAI Algorithmus nicht abhängig von der Modellarchitektur, sodass er auf andere, bereits trainierte Architekturen wie beispielsweise neuronale Netzwerke angewandt werden kann.²³ Der Unterschied wird anhand der *Abbildungen 3A und B* (nach Das und Rad²³) verdeutlicht.

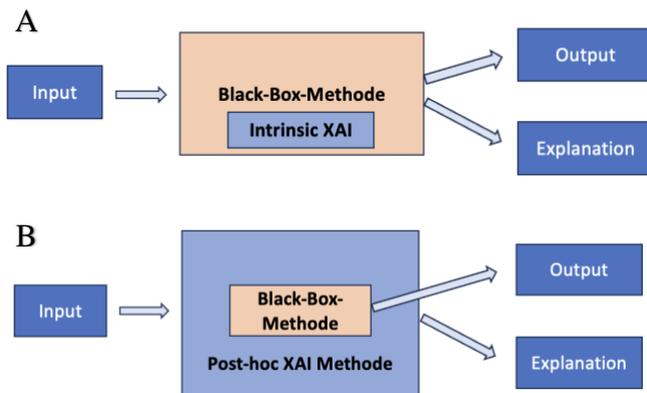


Abbildung 3A und B (angelehnt an Das und Rad²³): Illustration eines intrinsischen sowie eines Post-hoc XAI-Modells

3A Die Erklärbarkeit ist unmittelbar in die Black-Box-Methode eingebettet.

3B Der XAI-Algorithmus wird auf die Black-Box-Methode angewandt und erklärt diese extern.

Modellspezifität

XAI-Modelle können in modellspezifische und modellagnostische Modelle eingeteilt werden. Modellspezifische XAI-Modelle werden auf einen bestimmten Anwendungsbereich angewandt und benötigen eine bestimmte Struktur. Ein modellagnostischer Ansatz hingegen stellt keine besonderen Anforderungen an das Modell, vielmehr wird dieser in den meisten XAI-Modellen verwendet.¹⁷

Erklärungsumfang

Eine XAI kann den Entscheidungsprozess für das gesamte Modell oder nur für bestimmte Inputs und

Outputs erklären, entsprechend unterscheidet man zwischen globalen und lokalen Erklärungen.¹⁷ Eine lokaler Erklärungsumfang fokussiert sich auf die Erklärung einzelner Dateninstanzen mit Erstellung einer Erklärungskarte pro Dateninstanz, ein globaler Erklärungsansatz hingegen versucht das gesamte Modell zu verstehen und generiert eine oder mehrere Erklärungskarten für eine Gruppe von Dateninstanzen.²³

Erklärungsformen

Eine XAI kann verschiedene Erklärungsformen generieren, darunter fallen merkmalsbasierte, textbasierte und beispielesbasierte Erklärungen. Merkmalsbasierte Erklärungen können zeigen welche Merkmale des Modells den größten Einfluss auf das Ergebnis haben.¹⁷ Bei Algorithmen in der medizinischen Bildgebung geschieht dies beispielweise durch Überlagerung des Originalbildes mit einer sogenannten Salienzkarte, die anzeigt welche Merkmale des Bildes besonders

wichtig sind für ein bestimmtes Ergebnis bzw. Entscheidung.^{109,143} Textbasierte Erklärungen beschreiben für den Menschen verständliche Erklärungen in Textform. Beispielbasierte Erklärungen nutzen Beispiele zur Modellerklärung.¹⁷

Beispielalgorithmen

Typische intrinsische XAI-Methoden sind beispielweise die logistische Regression⁵⁴ oder Entscheidungsbäume.¹⁴⁶ Die meisten aktuell verwendeten XAI-Algorithmen fallen jedoch unter die post-hoc Modelle aufgrund der leichteren Integration und größeren Reichweite.^{17,23} Typische Post-hoc XAI-Methoden sind beispielweise Local interpretable model-agnostic explanations (LIME)⁹⁸, Class Activation Mapping (CAM)¹⁴⁸, Shapley Additive exPlanations (SHAP)⁶³ oder die PCA (siehe *Tabelle 8*)^{89,2}. LIME⁹⁸ ist eine lokale, modellagnostische Methode, welche die Vorhersagen bzw. Ergebnisse des ursprünglichen

Black-Box-Modells erklärt, indem sie das komplexe Modell durch ein lokal interpretierbares Ersatzmodell ersetzt. Dabei versucht LIME die lokale Nachbarschaft jeder Vorhersage zu modellieren. Die Eingangsdaten werden gering verändert (z.B. Überschreiben von Bilddaten), anschließend wird ein interpretierbares Modell auf diesem Datensatz trainiert und daraus eine Erklärung gewonnen. Nun wird gemessen wie nah die Erklärung der Vorhersage des ursprünglichen Black-Box-Modells kommt und versucht damit zu erklären wieso dieses eine bestimmte Vorhersage getroffen hat.^{17,146} CAM¹⁴⁸ verwendet ein globales Durchschnittspooling und einen sogenannten Fully-Connected-Layer um neuronale Netze zu visualisieren und die wichtigsten, zur Entscheidung beitragenden Merkmale, abzubilden. SHAP⁶³ ist eine von der Spieltheorie inspirierte XAI-Methode, welche die Relevanz jedes einzelnen Merkmals für eine bestimmte Vorhersage berechnet unter Nutzung sogenannter Shapley-Werte, die berechnet werden müssen.

2. Patienten, Materialien und Methoden

2.1 Patienten, Datenerhebung

Es wurden die MPFC-Daten aus dem peripheren Blut von 157 CLL-Patienten analysiert und mit klinischen Daten abgeglichen. Die MPFC-Daten wurden erhoben für die routinemäßige diagnostische Analyse am Universitätsklinikum Marburg von 2014 bis 2020. Die klinischen Daten umfassen folgendes: CLL-IPI (d.h. TP53-Mutationsstatus, IGHV-Mutationsstatus, Alter, Binet, beta2-Mikroglobulin), Geschlecht, ECOG-Stadium, Richter-Transformation ja/nein, Behandlung ja/nein, Todesdatum, Behandlungsversagen und Datum der letzten Nachuntersuchung. Im Falle unvollständiger oder unbekannter CLL-IPI-Parameter wurde die Hälfte der Punkte für die fehlenden Parameter vergeben. Der modifizierte CLL-IPI-Score wird als „Assumed CLL-IPI-Score“ bezeichnet. Die Studie wurde von der lokalen

Ethikkommission Marburg genehmigt. Die Patienten wurden in eine Gruppe mit schlechterem und eine Gruppe mit besserem klinischen Verlauf eingeteilt. Patienten, die während der Nachbeobachtung verstarben sowie diejenigen, die einen Progress auf die erste Linie der Systemtherapie hatten, wurden als TTF 1 (Time-to-First-Line-Treatment-Failure bzw. Zeit bis zum Versagen der Erstlinientherapie) kategorisiert, d. h. der Gruppe mit schlechterem klinischen Verlauf zugewiesen. Alle anderen Patienten wurden als TTF 0 eingestuft, die einen besseren klinischen Verlauf beschreibt.

Der ALPODS-XAI-Algorithmus¹³⁰ wurde verwendet, um Zellpopulationen in den Daten der Durchflusszytometrie zu identifizieren, die bei Patienten mit schlechterem (TTF 1) oder besserem klinischen Verlauf (TTF 0) über- oder unterrepräsentiert waren (siehe *Tabelle 12*). Zur Kontrolle etablierter Marker wurde der prädiktive Wert der XAI-Populationen mit der Häufigkeit von

CD38-positiven CLL-Zellen und CLL-IPI auf den Receiver Operating Characteristic (ROC)-Kurven verglichen. Zudem wurden die Vorhersagewerte der XAI-Populationen bezüglich des IPI-Scores bestimmt (siehe *Tabelle 11*). Eine multiple logistische Regressionsanalyse wurde für die Kombination von mehr als einer unabhängigen Variable durchgeführt, um dichotome Gruppen vorherzusagen. ALPODS liefert nur disjunkte Zellpopulationen, die in den üblichen zweidimensionalen Punktdiagrammen der Durchflusszytometrie mit FCS-Datenfiles visualisiert werden können. Dies ermöglicht es den klinischen Experten und Betreibern von Durchflusszytometern die identifizierten Zellpopulationen ihren biologischen Gegenstücken zuzuordnen.

2.2 Antigen-Panel, durchflusszytometrische Färbung und Analyse

Bei einer Durchflusszytometrie werden Zellen einer Suspension (z. B. Blut oder Liquor) einzeln an einem Laserstrahl vorbei geführt. Während sie den Lichtstrahl passieren, entsteht Streulicht, das sich wie folgt unterteilen lässt: Vorwärtsstreulicht („forward scatter“), das Informationen über die Zellgröße liefert, und Seitwärtsstreulicht („side scatter“), das Aufschluss über die Granularität einer Zelle gibt. Dies ermöglicht bereits in einer Vollblutprobe die Unterscheidung verschiedener Zelltypen. Moderne Durchflusszytometer verfügen neben den Streulichtdetektoren über mehrere Laser und Fluoreszenzdetektoren, die Licht bei unterschiedlichen Wellenlängen aussenden oder emittiertes Fluoreszenzlicht erfassen. Durch die Verwendung von Antikörpern, die mit verschiedenen Fluoreszenzfarbstoffen markiert sind, kann innerhalb einer Probe die Anwesenheit

mehrerer unterschiedlicher Oberflächenantigene auf den Zellen nachgewiesen werden. Der Einsatz von Antikörpern, die zellspezifische Antigene erkennen, ermöglicht die eindeutige Identifizierung und Abgrenzung von Zellen in einem Zellgemisch.⁸⁴

Zur Diagnose der CLL und anderen B-Zell-Lymphomen wurde in dieser Arbeit eine B-Zelle verwendet, die aus zwei getrennten Röhrchen mit unterschiedlicher Fluoreszenz-Antikörperzusammensetzung bestand. Das erste Röhrchen (T1) enthielt Fluoreszenz-Antikörper gegen B-Zell-Antigene (CD19, CD20, FMC7, CD79b, CD23, leichte Ketten von kappa und lambda), T-Zell-Antigene (CD3, CD5, CD2, CD7, CD4, CD8) und den Aktivierungsmarker CD38.²² Die Kombination von Antikörpern gegen B-Zellen und T-Zellen sollte verhindern, dass in der Routinediagnostik ein T-Zell-Lymphom übersehen wird. Das zweite Röhrchen (T2) enthielt B-Zell-Antigene (CD19, CD20, IgM), Marker für eine Haarzelleukämie (CD103, CD11c, CD25), ein

follikuläres Lymphom und ein hochgradiges Lymphom (CD10) sowie zusätzliche Marker, um die Diagnose CLL sicherzustellen (CD43, CD200). Das vollständige Antikörperpanel, die Klone und die Fluoreszenzfarbstoffe sind in *Tabelle 9* abgebildet. Die zumeist mehr als 10^5 zellulären Events pro Patient werden in dieser Arbeit also auf mehr als 20 quantifizierbare Eigenschaften (zelluläre Oberflächenantigene) untersucht.

Zwei 5 ml Polystyrol-FACS-Röhrchen mit Fluoreszenz-Antikörpern in einer getrockneten Schicht (DuraClone-Technology, Beckman Coulter, Krefeld, Deutschland) wurden für 15 Minuten bei Raumtemperatur in 100 µl vorgewaschenem peripheren Blut inkubiert. Nach der Antikörperfärbung wurden die Erythrozyten 10 Minuten in 2 ml Versalyse (Beckman Coulter) lysiert, mit 3 ml gepufferter Phosphatsalzlösung (PBS Biochrom, Berlin, Deutschland) gewaschen und 5 Minuten bei 300 g zentrifugiert. Das Zellpellet wurde in 500 µl PBS resuspendiert und auf einem

Navios Flow Cytometer (Beckman Coulter) gemessen. Insgesamt wurden bis zu 1×10^5 Zellen erfasst.

Table 9: Verwendete Antikörperkombinationen („Panels“) „T1“ und „T2“ mit Antikörpern „Duraclones“ der Firma Beckman Coulter

405 Excitation			488 Excitation						633 Excitation			
PPP	Pacific Blue	Krome Orange	FITC	PE	ECD	PC5	PC5.5	PC7	APC	APC-AF700	APC-AF750	
T1	CD2 + FMC7	CD45	CD8 + Kappa	CD7 + lambda	CD23		CD4 + CD79b	CD5	CD38	CD19	CD3 + CD20	
Klone	39C1.5, FMC7	J33	B9. 11, RAHK	8 H8. 1, RAHL	9 P25		13B8.2, CB3.1	BL 1A	LS 198-4-3	JR. 119	B9E9 (HRC20), UCHT1	
T2	IgM	CD19	CD103	CD43	CD25		CD10	CD200		CD11c	CD20	
Klone	OX-104	BU15	SA-DA4	2G5.1	J3.119		DFT1	B1.49.9		ALB1	B9E9 (HRC20)	

2.2.1 Datenkompilierung ALPODS

Die rohen MPFC-Daten wurden kompensiert und logarithmiert. Ereignisse mit sehr hoher seitlicher Streuung (d. h. hauptsächlich Granulozyten) wurden gekappt, um die Menge an Daten zu reduzieren, die wenig informativen Wert haben. Anschließend wurden die Daten auf der Grundlage der adaptierten Milligan-Cooper-Standardisierung^{76,122} auf Werte zwischen null und sechs standardisiert. Diese

standardisierten MPFC-Daten und die Ergebnisgruppe (TTF 1 oder TTF 0) wurden als Eingabeinformationen für den XAI-Algorithmus ALPODS verwendet (siehe schematischer Arbeitsablauf in *Abbildung 6*). Aus der Gesamtzahl der aufgezeichneten Zellereignisse wurde eine zufällige Datenstichprobe aus 1% der Ereignisse jeder Sonde gezogen. Unter Verwendung dieser 1%-Stichprobe für das Training von ALPODS wurde eine 1000-fache Kreuzvalidierung durchgeführt, um eine Überanpassung auf den Trainingsdatensatz zu verhindern und die Generalisierbarkeit¹⁰ von ALPODS für weitere MPFC-Daten zu erhalten. Die Kreuzvalidierung ist eine der am weitesten verbreiteten Methoden zur Schätzung des Vorhersagefehlers von Modellen sowie zur Anpassung von Modellparametern: bei einer k-fachen Kreuzvalidierung (hier: $k = 1000$) wird der Lerndatensatz in k (hier 1000) Teilmengen ähnlicher Größe unterteilt. Anschließend wird das Modell mit $k - 1$ (hier 999) Teilmengen trainiert, die gemeinsam die Trainingsmenge bilden. Danach wird das Modell

mit der verbleibenden Teilmenge validiert (hier 1000-999 = 1 Teilmenge, sog. „Validierungsmenge“) und damit die Performance des Modells beurteilt. Dieses Prozedere wiederholt sich nun solange bis jede der k Teilmengen als Validierungsmenge gedient hat und letztendlich entspricht der Durchschnitt der k Leistungsmessungen der kreuzvalidierten Leistung.¹⁰

Die Populationen, die für die Unterscheidung zwischen TTF 1 und TTF 0 relevant waren, wurden von ALPODS ausgewählt, und die wichtigsten Populationen anhand des Cohen's D-Effektgrößenmaßes gefiltert.^{19, 139} Cohen's D ist eine statistische Effektgröße, die der standardisierten mittleren Differenz zwischen zwei Gruppen entspricht.^{19, 58}

Eine ABC-Analyse wählte optimale Grenzen für die Aufteilung der Untergruppen, indem sie die mathematischen Eigenschaften im Zusammenhang mit der Verteilung der analysierten Items ausnutzte. Die ABC-Analyse unterteilt die Daten in drei

disjunkte Teilmengen A, B und C. Dabei umfasst Teilmenge A die profitabelsten Werte bzw. die wenigen wichtigen ("the important few"), Teilmenge B umfasst Werte, bei denen der Ertrag dem Aufwand entspricht, der nötig ist, um ihn zu erhalten, Teilmenge C umfasst nicht profitable Werte.¹³¹

2.2.2 Statistik

Die Diagramme und Statistiken wurden erstellt mit Excel 2016 (Microsoft Corporation, Redmond Washington, USA) sowie im R-Paket ggplot2¹³⁷ und DataVisualizations¹²³, GraphPad Prism® Version 9.4.1 (GraphPad Software, San Diego California, USA), R (Programmiersprache), www.R-project.org (Zugriff am 06/10/2022).

Zwei nachfolgend erklärte und zur Beurteilung der Ergebnisse dieser Dissertation notwendige Parameter sind die Receiver-Operating-Characteristics (ROC) sowie die Area-Under-the-

ROC-Curve (AUC). Mittels ROC und AUC kann die Unterscheidungsgenauigkeit eines diagnostischen Tests anschaulich visualisiert werden.^{31,30} Es kann schwierig sein einen Trade-Off zu finden, bei dem ein bestimmter Test eine gute Sensitivität (Richtig-positiv-Rate) aber auch eine gute Spezifität (Richtig-negativ-Rate) hat. Die ROC-Kurve beschreibt dabei graphisch die Änderung der Sensitivität und Spezifität bei Änderung des Grenzwerts (x-Achse: Spezifität 1-0 (100% bis 0%); y-Achse: Sensitivität von 0-1 (0 bis 100%)), wobei die Wertepaare jedes angenommenen Grenzwertes in dieses Diagramm eingetragen werden und durch eine Linie (die ROC-Kurve) verbunden werden. Die Area-Under-the-ROC-Curve (AUC) beschreibt die Fläche unterhalb der ROC-Kurve, deren maximaler Wert bei 1,0 liegt, einem (theoretisch) perfektem Test (mit 100% Sensitivität und 100% Spezifität). Ein AUC-Wert von 0,5 sagt aus, dass ein Test keinerlei Unterscheidungsfähigkeit hat. Es gibt zwar mehrere AUC-Interpretations-Skalen, i.d.R. sind jedoch

Tests mit AUC-Werten $\leq 0,75$ klinisch nicht relevant.³⁰ Ziel ist es also diagnostische und prognostische Tests mit möglichst hohen AUC-Werten zu erstellen.

2.3 Algorithmic Population Descriptions (ALPODS)

Algorithmic Population Description (ALPODS) ist ein Algorithmus, der basierend auf dem Bilden von Clustern die Klassifizierung von hochdimensionalen Daten ermöglicht, womit beispielsweise relevante Subpopulationen in MPFC-Daten gefunden werden können. Für eine klinische Anwendung ist es besonders wichtig, dass Ärztinnen und Ärzte die Entscheidungsfindung eines Algorithmus verstehen können. ALPODS kann Entscheidungen anhand von Zellpopulationen in durchflusszytometrischen Diagrammen („dot plots“) darstellen und lässt sich damit in die Systeme der XAI einordnen.^{130,4}

Zunächst werden die für die Klassifizierung relevanten Teilpopulationen innerhalb eines Trainingsdatensatzes identifiziert. Diese werden rekursiv durch eine Folge von Entscheidungen für jede Variable in den Daten erzeugt.

Das Bayes-Theorem⁷³ wird verwendet, um sämtliche potenziellen Entscheidungen für alle Marker in einer gegebenen Population abzuleiten. Dies stellt sicher, dass die optimale Entscheidung hinsichtlich der Kosten (bzw. des Risikos) getroffen wird.¹⁰⁴

Die Bewertung der Bedingungen (Split-Kriterium) erfolgt mit Hilfe des Simpson-Index (SI), der die erwartete gemeinsame Wahrscheinlichkeit darstellt, dass zwei aus der Population entnommene Entitäten die gleichen oder unterschiedliche Typen repräsentieren. Inspiriert ist der SI biologisch vom Konzept der erwarteten Heterozygotie in der Populationsgenetik und der Wahrscheinlichkeit eines interspezifischen Zusammentreffens.⁵¹

Die Rekursion beginnt mit dem vollständigen Datensatz als Population. Der Rekursionsprozess endet, wenn die Klassenzuordnungen für alle Mitglieder einer Teilpopulation gleich sind oder wenn die Größe der Teilpopulation einen vordefinierten Anteil der Daten unterschreitet (i.d.R. $< 1\%$).

In Anlehnung an das Konzept der Fast-and-Frugal-Trees⁶² wird die Abfolge von Entscheidungen für eine Population zu einer algorithmischen Populationsbeschreibung vereinfacht, indem alle Bedingungen, die denselben Marker nutzen, zu einer einzigen Bedingung aggregiert werden. Diese aggregierte Bedingung stellt üblicherweise ein Intervall innerhalb des Markerbereichs dar.

Die Vereinfachung soll Populationen so beschreiben, dass sie für menschliche Experten verständlich sind. Darüber hinaus führt die Vereinfachung zu Teilpopulationen, die zum Einen relevant für die Klassifizierung sind, außerdem durch die Anwendung der Beschreibungsregel auf

den Datensatz algorithmisch entscheidbar sind und nicht zuletzt auch in einer Form dargestellt werden, die für Domänenexperten nachvollziehbar ist.

Die Bewertung der Relevanz einer Subpopulation für die Diagnose erfolgt mithilfe eines Effektstärkemaßes¹³⁹, wobei standardmäßig der absolute Wert von Cohen's d verwendet wird¹⁹. Die berechnete ABC-Analyse¹³¹ der Effektgrößen wird rekursiv angewendet, um die m relevantesten Subpopulationen auszuwählen. Dieser Prozess wird fortgesetzt, bis die Anzahl der ausgewählten Populationen im Bereich des Miller-Optimums m des menschlichen Verständnis liegt, d.h. $m = 7 \pm 2$.⁷⁵

Die Beschreibungen der Teilpopulationen werden dem Fachexperten vorgelegt, der jeder relevanten Population einen aussagekräftigen Namen zuordnet. Während dieser Interaktion wurde deutlich, dass Regeln mit Bedingungen, die sich aus einem XAI-System ergeben, alleine nicht ausreichen, um die von der Regel ausgewählte Teilpopulation zu

verstehen. Um die „Bedeutung“ einer spezifischen Zellpopulation besser zu erfassen, wurde den klinischen Experten ein sog. Visualisierungspanel (VisPanel) präsentiert, bestehend aus einer Matrix von farbigen Streudiagrammen der Klassen. Ein beispielhaftes VisPanel für die Regel "Regel1: Relevante Ereignisse in Population 1 zur Identifizierung des Knochenmarks sind SS++ und CD33+ und CD13-" sieht man in *Abbildung 4*.

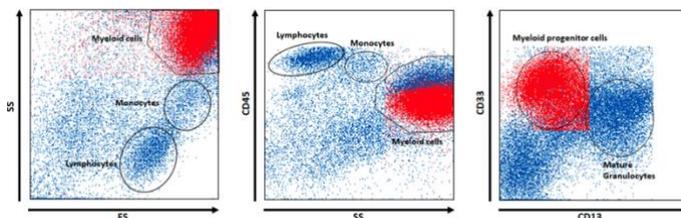


Abbildung 4 (von Ultsch et al.¹³⁰): VisPanel in Form von MPFC-Punktdiagrammen (Streudiagramme) von Ereignissen aus dem Trainingssatz für die Regel „Regel1: Relevante Ereignisse in Population 1 zur Identifizierung des Knochenmarks sind SS++ und

Die roten Punkte bezeichnen die Subpopulation 1, die von ALPODS als relevant für die Identifizierung des Knochenmarks erkannt wurde. Die Ereignisse der "Population 1" sind in der Abbildung rot markiert, wobei diese Zellen in Knochenmarksproben durchschnittlich mit 43 % und im peripheren Blut mit 5 % auftreten. Das linke Feld von *Abbildung 4* zeigt, dass die Zellpopulation im Bereich der myeloischen Zellen mit hoher Seiten- (SS) und Vorwärtsstreuung (FS) zu finden ist, daneben befinden sich Lymphozyten und Monozyten. Das mittlere Feld zeigt, dass die Zellpopulation eine schwache CD45-Expression innerhalb der myeloischen Zellen aufweist. Schließlich konnte Population 1 aufgrund der schwachen CD13- und starken CD33-Expression als myeloische Vorläuferzellen (z. B. Myelozyten) identifiziert werden (rechtes Feld).

Die Anzahl und Anordnung des VisPanels können entweder nach den Gewohnheiten der Experten im klinischen Alltag oder alternativ anhand der größten

absoluten Wahrscheinlichkeitsunterschiede (ProbDiff) festgelegt werden. ProbDiff wird für alle Paare (X,Y) der d Variablen in den Daten berechnet. Für jedes Paar (X,Y) von Variablen wird die Wahrscheinlichkeitsdichte einer Klasse C $p(X,Y|C)$ unter Verwendung geglätteter Datenhistogramme (SDHs)²⁸ geschätzt. Die absolute Wahrscheinlichkeitsdifferenz wird dann wie folgt berechnet: $\text{ProbDiff}(X,Y,C) = |\text{SDH}(X,Y|C) - \text{SDH}(X,Y|\text{not}(C))|$.

Das VisPanel¹³¹ wird durch die A-Menge gebildet, welche die größte Anzahl einer berechneten ABC-Analyse über alle ProbDiffs beinhaltet. Die Klasse mit der höchsten Klassenwahrscheinlichkeit (z. B. Knochenmark in Regel 1) wird durch rote Punkte über den Daten außerhalb dieser Klasse dargestellt (im Beispiel der *Abbildung 4* peripheres Blut als andere Klasse). Mittels dieser Methode der differenziellen Populationsvisualisierung (DIPOLVIS) können die klinischen Experten alle dargestellten Populationen verstehen und jedem

XAI-Ergebnis können sinnvolle Beschreibungen zugeordnet werden. Im obigen Beispiel konnte mit der Regel 1 die Population 1 als "myeloische Vorläuferzellen" identifiziert werden. Der klinische Experte konnte also durch die Regel 1 nachvollziehen, wieso ALPODS diese Patientenprobe als Knochenmarksprobe identifiziert hat.

Fuzzy Reasoning⁶⁵ bildet die Grundlage für die Kombination der relevanten Populationen für ein XAI-Diagnosesystem: dabei werden die relativen Anteile der Zugehörigkeit der relevanten Subpopulationen zu einer ausgewählten Patientendaten-Teilmenge (erweiterte Lernmenge) berechnet. Die Fuzzy-Zugehörigkeitsfunktionen werden als Posterior-Wahrscheinlichkeiten berechnet, woraus sich ein Satz von XAI-Experten für jede der Populationen ergibt. Das Mehrheitsvotum der XAI-Experten entscheidet über die Klassifizierung (Diagnose). Der große Vorteil dieses Ansatzes besteht darin, dass sowohl

unterstützende als auch gegensätzliche Argumente für eine Diagnose dem Domänenexperten in verständlichen Begriffen erklärt werden können. Eine Erklärung, warum eine Probe mehr peripheres Blut als Knochenmark enthält, sieht zum Beispiel so aus: `many(Thrombozytenaggregate)` und `few(Vorläufer-B-Zellen)`, wobei die Funktionen `many()` und `few()` als Fuzzy-Set-Zugehörigkeitsfunktionen definiert sind.⁶⁵

3. Ergebnisse

3.1 Patientencharakteristika

Von den 157 CLL-Patienten hatten N = 42 ein inferiores Ergebnis (TTF 1: Tod und/oder Therapieversagen in der Erstlinie), N = 115 Patienten erreichten die definierten Endpunkte nicht (TTF 0). Das mittlere Alter der Gesamtkohorte betrug 68 Jahre (26-91 Jahre), 62 (39,5%) der Patienten waren weiblich und 95 (60,5%) männlich. 83 (52,9%) der Patienten waren mit Binet A, 24 (15,3%) mit Binet B und 12 (7,6%) mit Binet C diagnostiziert. Die Nachbeobachtungszeit betrug im Median 31,5 Monate (Interquartilsbereich 9-65 Monate). Bei vier Patienten (2,5%) kam es zu einer malignen Richter-Transformation. Zusätzliche Patientencharakteristika für die Gesamtkohorte sowie getrennt für TTF 1 oder TTF 0 sind in *Tabelle 10* angegeben.

Tabelle 10: Patientencharakteristika

	Total (N = 157)	TTF 1 (N = 42)	TTF 0 (N = 115)
Alter (Jahre)			
Median (range)	68 (26-91)	70 (50-88)	67 (26-91)
Geschlecht			
Weiblich	62 (39,5%)	17 (40,5%)	45 (39,1%)
Männlich	95 (60,5%)	25 (59,5%)	70 (60,9%)
Binet			
A	83 (52,9%)	15 (35,7%)	68 (59,1%)
B	24 (15,3%)	7 (16,7%)	17 (14,8%)
C	12 (7,6%)	8 (19,0%)	4 (3,5%)
Unbekannt	31 (19,7%)	9 (21,4%)	22 (19,1%)
CLL-IPI			
Niedrig	84 (53,5%)	9 (21,4%)	75 (65,2%)
Mittel	53 (33,8%)	20 (47,6%)	33 (28,7%)
Hoch	16 (10,2%)	11 (26,2%)	5 (4,3%)
Sehr hoch	4 (2,5%)	2 (4,8%)	2 (1,7%)
Follow up			
Median (IQR) – Monate	31 (9-65)	71 (38-114)	24 (7-46)
Tod (N)	16 (10,2%)	16 (38,1%)	0 (0%)
Erstlinien-Therapieversagen (N)	34 (21,7%)	34 (80,1%)	0 (0%)
Richter-Syndrom (N)	4 (2,5%)	3 (7,1%)	1 (0,9%)
Therapie (Erstlinie)			
R-Bendamustin	25 (15,9%)	17 (40,5%)	8 (7,0%)
Ibrutinib	12 (7,6%)	6 (14,3%)	6 (5,2%)
Andere Therapie	24 (15,3%)	15 (35,7%)	9 (7,8%)
Keine Therapie	80 (51,0%)	0 (0%)	80 (70,0%)
Unbekannt	16 (10,2%)	4 (9,5%)	12 (10,4%)

Abkürzungen: N = Nummer (Anzahl); CLL = Chronisch lymphatische Leukämie; IPI = international prognostic index; TTF = time to first line treatment failure (Zeit bis zum Versagen der Erstlinientherapie); IQR = Interquartil-Range; R = Rituximab.

Die Kaplan-Meier-Kurven für TTF 1 und TTF 0 sind in *Abbildung 5* dargestellt. Die Unterschiede zwischen den Gruppen sind im Log-Rank-Test

hochsignifikant ($p < 0,0001$). Damit sind die Gruppen geeignet, Zellpopulationen zu definieren, die Unterschiede im Ergebnis erklären könnten.

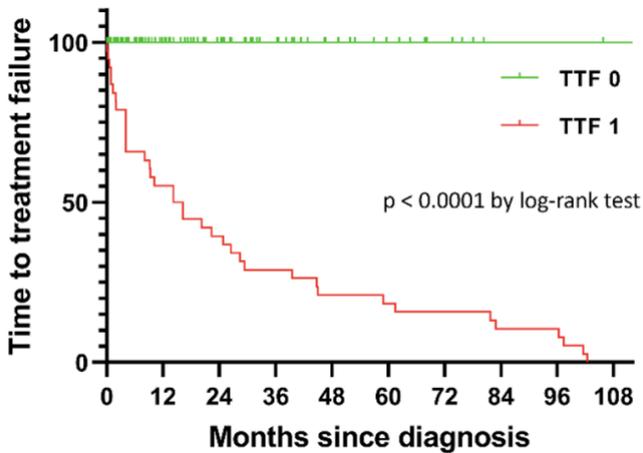


Abbildung 5: Kaplan-Meier-Kurve der CLL-Patientenkohorte

TTF 1 umfasst Patienten mit dem Endpunkt Tod oder Therapieversagen und TTF 0 umfasst alle anderen Patienten.

3.2 Datenverarbeitung und Zellpopulationen von ALPODS

ALPODS identifizierte $N = 17$ charakteristische Zellpopulationen in den MPFC-Daten, die in der TTF 1-Patientenkohorte mit schlechterem Ergebnis überrepräsentiert ($N = 14/17$) oder unterrepräsentiert ($N = 3/17$) waren. 7 von 17 Populationen wurden im ersten Röhrchen (T1) des diagnostischen B-Zell-Panels der MPFC identifiziert. 10 von 17 Populationen wurden im zweiten Röhrchen (T2) des B-Zell-Panels identifiziert. Arbeitsablauf und Datenverarbeitung mit ALPODS ist in *Abbildung 6* nachvollziehbar.

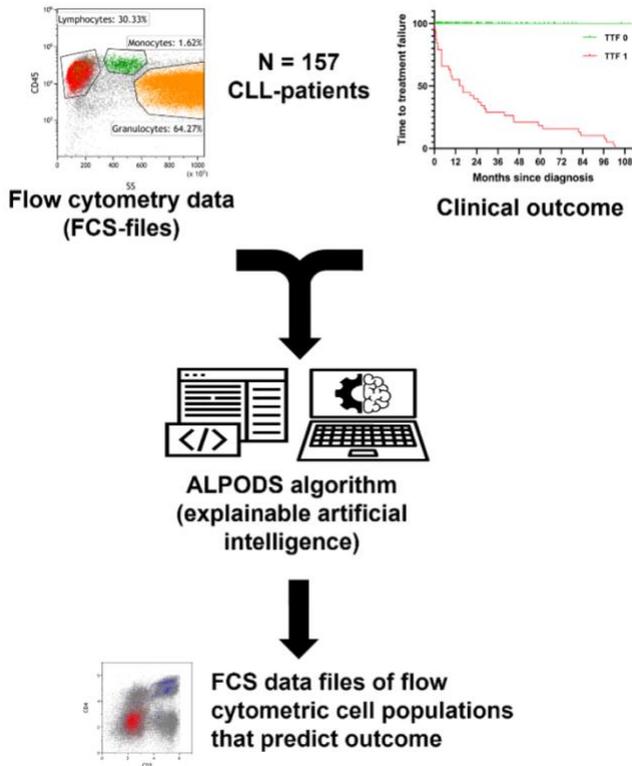


Abbildung 6: Arbeitsablauf und Datenverarbeitung mit ALPODS

Die Rohdaten der MPFC wurden standardisiert und der Ergebnisgruppe (TTF 1, TTF 0) zugeordnet. Der ALPODS Algorithmus wurde verwendet, um charakteristische Zellpopulationen mit

unterschiedlicher Häufigkeit in TTF 1 und TTF 0 zu identifizieren. Die für die Bestimmung wichtigsten Populationen wurden in durchflusszytometrischen Punktdiagrammen visualisiert und ihren biologischen Entsprechungen zugeordnet.

Mittels Mann-Whitney-U-Tests und ROC-Analysen konnten die XAI-Populationen ermittelt werden, die am besten prädiktiv für ein schlechteres bzw. inferiores Ergebnis sind. Die Ergebnisse sind in *Tabelle 11* sowie *Tabelle 12* aufgeführt.

Die 17 von ALPODS identifizierten XAI-Populationen mit einer signifikanten Vorhersagekraft für das Ergebnis (TTF 1 vs. TTF 0) wurden auf ihren prognostischen Wert bei Patienten mit hohem IPI (≥ 4) im Vergleich zu Patienten mit niedrigem IPI (≤ 1) überprüft (siehe *Tabelle 11*). XAI-Populationen mit einem signifikanten Vorhersagewert für den vorliegenden CLL-IPI-Score (assumed-CLL-IPI-Score: ≥ 4 vs. ≤ 1) waren T1C0011, T1C0016, T1C0020, T1C0023, T2C0004

und T2C0018 (fettgedruckt in *Tabelle 11*). Unter diesen Populationen hatte nur T1C0016 eine höhere Häufigkeit bei den Patienten mit niedrigerem CLL-IPI-Score (CLL-IPI-Score ≤ 1 ; Mittelwert 12,50% vs. 5,37%; Standardabweichung der Differenz 2,42). Die restlichen Populationen mit statistisch signifikanten Vorhersagewerten zeigten sich allesamt häufiger bei Patienten mit hohem CLL-IPI-Score und damit verbundener schlechterer Prognose.

Tabelle 11: Vorhersagewerte der XAI-Populationen bezüglich des IPI-Scores

	IPI		SF der Differe nz	p-Wert (MWU- Test)	ROC		
	≥4 (hoch)	≤1 (niedrig)			AUC	95%KI	p-Wert
CD38+ <i>XAI-</i> <i>Populationen</i>	41,35	21,36	7,55	0,0056	0,66	0,57-0,76	0,0018
T1C0011	2,49	1,69	0,93	0,0212	0,67	0,53-0,80	0,0219
T1C0012	1,69	1,44	0,54	0,5321	0,55	0,40-0,69	0,5281
T1C0016	5,37	12,50	2,42	0,0008	0,74	0,63-0,84	0,0010
T1C0017	7,03	4,74	1,92	0,4278	0,56	0,41-0,07	0,4237
T1C0019	5,17	6,48	1,52	0,1839	0,60	0,45-0,74	0,1815
T1C0020	0,42	0,15	0,14	0,0080	0,69	0,55-0,83	0,0086
T1C0023	4,30	0,48	0,77	0,0046	0,70	0,55-0,86	0,0051
T2C0001	4,01	5,01	1,98	0,9892	0,50	0,36-0,65	0,9868
T2C0002	3,05	0,93	0,97	0,3466	0,57	0,41-0,73	0,3429
T2C0004	8,12	1,78	1,84	0,0029	0,71	0,58-0,85	0,0033
T2C0009	0,57	1,17	0,42	0,1602	0,60	0,47-0,73	0,1584
T2C0010	0,28	0,32	0,19	0,9169	0,51	0,36-0,66	0,9146
T2C0014	11,27	10,34	2,71	0,8283	0,52	0,37-0,66	0,8238
T2C0018	4,71	2,27	0,99	0,0030	0,71	0,60-0,82	0,0034
T2C0020	4,34	2,00	0,89	0,2346	0,59	0,44-0,73	0,2317
T2C0021	3,16	3,20	1,06	0,9586	0,50	0,36-0,65	0,9572
T2C0028	1,24	0,42	0,22	0,4348	0,56	0,39-0,72	0,4309

Abkürzungen: IPI = international prognostic index; ROC = receiver operation characteristics; SF = Standardfehler; MWU = Mann-Whitney-U; AUC = area under curve; KI = Konfidenzintervall, IPI = international prognostic index; T1 = Tube 1 (zweite Tube des Durchflusszytometrie-Panels); T2 = Tube 2 (zweite Tube des Durchflusszytometrie-Panels).

Eine AUC von 0,74 der Zellpopulation T1C0016 in *Tabelle 11* bedeutet, dass alleine durch Betrachten dieser XAI-Zellpopulation in 74% aller Fälle der

betrachteten Patientenkohorte korrekt differenziert werden kann, ob ein Patient einen CLL-IPI-Score von ≥ 4 oder ≤ 1 besitzt.

Veranschaulicht man sich in *Abbildung 7 A-F* die Verteilung der sechs relevantesten Zellpopulationen hinsichtlich ihres Vorkommens bei Patienten mit verschiedenen CLL-IPI-Scores, so wird deutlich, dass die Zellpopulationen T1C0011 (AUC 0,67; 95%CI 0,53-0,80, $p=0,0219$), T1C0020 (AUC 0,69; 95%CI 0,55-0,83, $p=0,0086$), T1C0023 (AUC 0,70; 95%CI 0,55-0,86, $p=0,0051$), T2C004 (AUC 0,71; 95%CI 0,58-0,85, $p=0,0033$) sowie T2C0018 (AUC 0,71; 95%CI 0,60-0,82, $p=0,0034$) vor allem bei Patienten mit höheren CLL-IPI-Scores überrepräsentiert sind und lediglich T1C0016 (AUC 0,74; 95%CI 0,63-0,84, $p=0,0010$) bei Patienten mit niedrigerem Assumed CLL-IPI-Score überrepräsentiert ist.

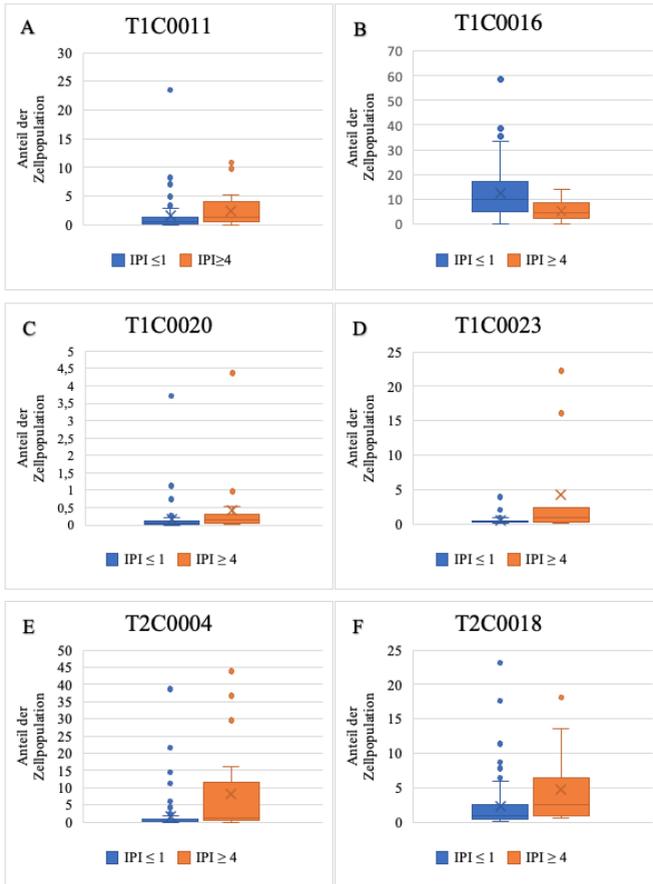


Abbildung 7A-F: Box-Plots zur Verteilung der Zellpopulationen T1C0011, T1C0016, T1C0020, T1C0023, T2C0004 sowie T2C0018 zu IPI-Scores ≤ 1 und ≥ 4

In dieser Arbeit wurde außerdem untersucht, ob die Zellpopulationen einen statistisch signifikanten Vorhersagewert für den Therapieausgang (TTF 0 vs. TTF 1) liefern können und somit einen Marker für die Prognose der Patienten stellen können. Anschließend folgt ein Vergleich mit den klinisch etablierten Prognosemarkern wie dem CLL-IPI-Score sowie CD38+.

XAI-Populationen mit einem signifikanten Vorhersagewert für das Ergebnis (TTF 1 vs. TTF 0) und für die Prognose (IPI niedrig vs. IPI hoch) waren T1C0011, T1C0016, T2C0004 und T2C0018 (fettgedruckt in *Tabelle 12*). Unter diesen Populationen hatte wie auch bei der Prognose (siehe *Tabelle 11*) ebenfalls nur T1C0016 eine höhere Häufigkeit bei den Patienten mit gutem Therapieausgang (Therapieausgang: TTF 0 vs. TTF 1 → 13,51% vs. 4,91% (Mittelwerte); Standardfehler der Differenz 1,83; Prognose: IPI ≤ 1 vs. IPI ≥ 4 → 12,50% vs. 5,37% (Mittelwerte); Standardfehler der Differenz 2,42).

Tabelle 12: Vorhersagewerte der XAI-Populationen bezüglich TTF 1 und TTF 0

	TTF 1		TTF 0		ROC		
	Mittel-Wert (%)	Mittel-Wert (%)	SF der Differenz	p-Wert (MWU-test)	AUC	95%KI	p-Wert
CLL-IPI	2,845*	1,322*	0,2712	<0,0001	0,78	0,70-0,86	<0,0001
CD38+	37,05	22,51	5,414	0,0016	0,66	0,57-0,76	0,0018
<i>XAI-populations</i>							
Total					0,95	0,91-0,98	<0,0001
T1C0011	5,36	1,39	0,81	<0,0001	0,76	0,68-0,84	<0,0001
T1C0012	1,91	1,61	0,40	0,2650	0,56	0,45-0,66	0,2635
T1C0016	4,91	13,51	1,83	<0,0001	0,78	0,70-0,86	<0,0001
T1C0017	6,46	4,15	1,32	0,2136	0,57	0,46-0,67	0,2124
T1C0019	5,19	6,38	1,01	0,0970	0,59	0,47-0,70	0,0966
T1C0020	0,35	0,11	0,08	0,1514	0,58	0,47-0,68	0,1506
T1C0023	2,25	0,57	0,49	0,5593	0,53	0,41-0,65	0,5573
T2C0001	2,59	5,50	1,31	0,0011	0,67	0,57-0,76	0,0012
T2C0002	2,85	0,73	0,61	0,0052	0,64	0,54-0,75	0,0055
T2C0004	9,64	1,23	1,29	0,0002	0,69	0,58-0,80	0,0002
T2C0009	1,74	0,84	0,49	0,6877	0,52	0,41-0,63	0,6859
T2C0010	0,55	0,21	0,15	0,5369	0,53	0,43-0,64	0,5349
T2C0014	8,19	12,14	1,97	0,0078	0,64	0,54-0,74	0,0081
T2C0018	5,37	1,66	0,72	0,0001	0,73	0,63-0,82	<0,0001
T2C0020	4,38	1,75	0,63	<0,0020	0,66	0,55-0,77	0,0022
T2C0021	3,25	2,74	0,73	0,4286	0,04	0,44-0,64	0,4266
T2C0028	0,85	0,36	0,14	0,2021	0,57	0,45-0,68	0,2010

*Punkte (≠Prozentangabe) Abkürzungen: TTF = time to first line treatment failure (Zeit bis zum Versagen der Erstlinientherapie); ROC = receiver operation characteristics; SF = Standardfehler; MWU = Mann-Whitney U; AUC = area under curve; KI = Konfidenzintervall, IPI = international prognostic index; T1 = Tube 1 (zweite Tube des Durchflusszytometrie-Panels); T2 = Tube 2 (zweite Tube des Durchflusszytometrie-Panels)

In der ROC-Kurvenanalyse hatte T1C0016 von allen XAI-Populationen die höchste Vorhersagekraft für das Ergebnis (TTF 0 vs. TTF 1). Es ist zu beachten, dass nur die T1C0016-Population den gleichen Vorhersagewert wie der CLL-IPI-Score hatte (*Tabelle 12*; beide: AUC 0,78; 95%CI 0,70-0,86).

Darüber hinaus hatte die Häufigkeit von CD38-positiven CLL-Zellen eine geringere Vorhersagekraft (AUC 0,66; 95%CI 0,57-0,76, $p=0,0018$) im Vergleich zu allen signifikanten XAI-Populationen T1C0011 (AUC 0,76; 95%CI 0,68-0,84, $p<0,0001$), T1C0016 (AUC 0,78; 95%CI 0,70-0,86, $p<0,0001$), T2C0004 (AUC 0,69; 95%CI 0,58-0,80, $p=0,0002$), und T2C0018 (AUC 0,72; 95%CI 0,63-0,82, $p<0,0001$).

Die 17 XAI-Populationen in Kombination hatten eine Vorhersagefähigkeit von AUC 0,95 (95%CI 0,91-0,98; $p<0,0001$) für TTF unter Verwendung der multiplen logistischen Regressionsanalyse (*Abbildung 8A*). Die Beschränkung auf die vier

Populationen von XAI (d. h. T1C0011, T1C0016, T2C0004, T2C0018), die sich auch als prädiktiv für IPI erwiesen haben, führte zu einer geringeren diagnostischen Fähigkeit von AUC 0,87 (95%CI 0,80-0,93; $p < 0,0001$) (Abbildung 8B).

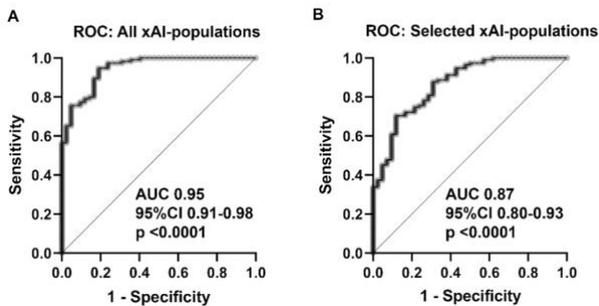


Abbildung 8A und 8B: Multiple logistische Regression mit ROC-Kurvenanalyse

8A Die multiple logistische Regression mit ROC-Kurvenanalyse für alle XAI-Populationen ergab eine Vorhersagekraft von AUC 0,95 (95%CI 0,91-0,98; $p < 0,0001$). 8B Die Einschränkung auf die vier XAI-Populationen mit der höchsten Prädiktion für TTF und IPI führte zu einer geringeren Vorhersagekraft (AUC 0,87; 95%CI 0,80-0,93; $p < 0,0001$). Abkürzungen: ROC = Receiver-Operations-Charakteristik; XAI = explainable

artificial intelligence; AUC = Area under the Curve;
CI = Confidence Intervall.

3.3 Identifizierung der XAI-Populationen

Der ALPODS-Algorithmus generiert FCS-Dateien, die mit konventionellen zweidimensionalen MPFC-Punktdiagrammen dargestellt werden können. Daher können XAI-Populationen von einem menschlichen MPFC-Experten erfasst und analysiert werden.

Die Population T1C0011 wurde innerhalb der CLL-Zellen lokalisiert, während T1C0016 aus CD4+ T-Zellen bestand (*Abbildung 9A*). T2C0004 repräsentierte fast ausschließlich eine Untergruppe von CLL-Zellen (*Abbildung 9B*) und T2C0018 war eine Mischung aus einer Untergruppe von CLL-Zellen (höherer Anteil) und einer Untergruppe von

T- und Natürlichen Killer-Zellen (NK-Zellen) (niedrigerer Anteil).

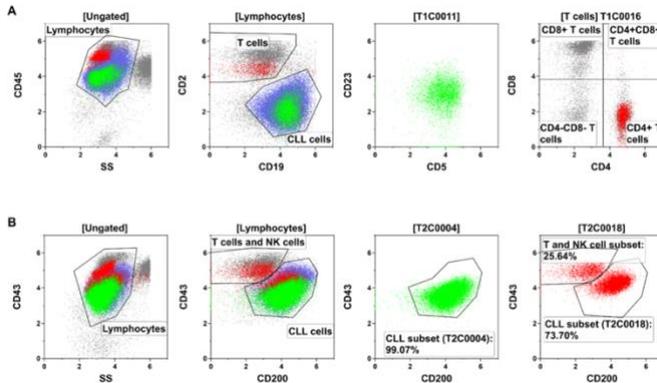


Abbildung 9A und 9B: Identifizierung und Beschreibung der XAI-Populationen in zweidimensionalen durchflusszytometrischen Punktwolkendiagrammen

9A Die Populationen T1C0011 (grün) und T1C0016 (rot) waren die relevantesten Populationen für das Ergebnis und die Prognose in Röhrcchen 1 (T1) des analysierten B-Zell-Panels. T1C0011 (grün) konnte innerhalb der CLL-Zellen (blau) lokalisiert werden. Die Population von T1C0016 (rot) entsprach CD4+ T-Zellen (d. h. T-Helferzellen).

9B T2C0004 (grün) und T2C0018 (rot) waren die wichtigsten Populationen für das Ergebnis und die Prognose in Röhren 2 (T2). Beide befanden sich hauptsächlich innerhalb der CLL-Zellen (blau), aber T2C0018 war eine Mischung aus verschiedenen biologischen Populationen (CLL-Zellen und T/NK-Zellen).

Die Zellpopulation (T1C0016), die mit der höchsten Wahrscheinlichkeit den Therapieausgang einer CLL vorhersagen konnte, lag nicht innerhalb der malignen CLL-Zellen, sondern bestand aus T-Helferzellen, und war bei Patienten mit günstigem Verlauf (TTF 0) überrepräsentiert. Dies führte entsprechend zu der Frage, ob CD8+ T-Zellen für ein schlechteres Ergebnis prädiktiv sind. Es stellte sich heraus, dass die XAI-Population T1C0023 aus CD8+ T-Zellen bestand. T1C0023 war signifikant häufiger bei Patienten mit schlechterer Prognose vorhanden (IPI \geq 4 Mittelwert 4,30% vs. IPI \leq 1 Mittelwert 0,48%; SE der Differenz 0,77; p<0,00461). In der ROC-Analyse konnte T1C0023

jedoch nicht alleine zwischen TTF 1 und TTF 0 differenzieren (AUC 0,53; 95%CI 0,41-0,65; p=0,5573).

Tabelle 13: Zuordnung wichtiger XAI-Zellpopulationen

XAI-Zellpopulation	Biologische Zellpopulationen
T1C0011	Untergruppe von CLL-Zellen
T1C0016	CD4+ T-Zellen
T1C0023	CD8+ T-Zellen
T2C0004	Fast ausschließlich Untergruppe von CLL-Zellen
T2C0018	Untergruppe von CLL-Zellen (73,70%) + Untergruppe von T- und NK-Zellen (25,46%)

3.4 Charakterisierung der prädiktiven Untergruppen innerhalb der CLL-Zellen

Neben T1C0016 (CD4+ T-Zellen) und T1C0023 (CD8+ T-Zellen) waren die meisten XAI-Populationen CLL-Untergruppen (Tube 1: T1C0011, T1C0012, T1C0017, T1C0019, T1C0020; Tube 2: T2C0002, T2C0004, T2C0009, T2C0010, T2C0014, T2C0018, TC0020). Die wichtigsten CLL-Untergruppen T1C0011, T2C0004 sowie T2C0018 sind in den *Abbildungen 10A-C* dargestellt. Darüber hinaus wurden die medianen Antigenexpressionsniveaus und die Strehöhe jeder relevanten XAI-Population innerhalb der CLL-Zellen mittels Heatmap mit denen der medianen Antigenexpression der CLL-Zellen eines durchschnittlichen Patienten aus der Patientenkohorte verglichen (*Abbildungen 10A und 10B*).

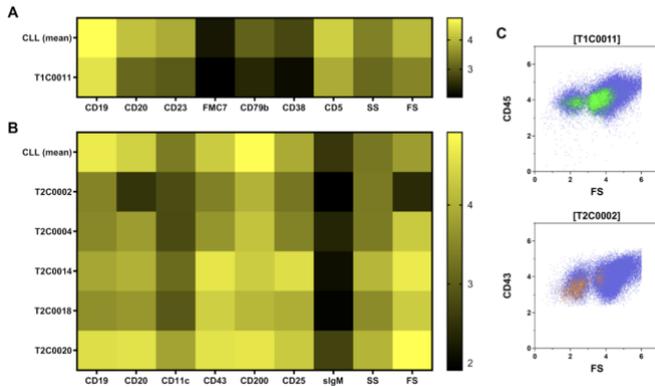


Abbildung 10A, 10B und 10C: Vergleich der medianen Antigenexpression und Streuungseigenschaften der für TTF prädiktiven XAI-Populationen mit den CLL-Zellen eines durchschnittlichen Patienten

Je dunkler ein Feld, desto geringer ausgeprägt ist das Merkmal. Die Population T1C0011 (A), die in Röhren 1 (T1) identifiziert wurde, zeigte eine geringere Vorwärtsstreuung (FS) und eine verringerte Antigenexpression. Die Population T2C0002 (B) in Röhren 2 (T2) hatte die gleiche geringere FS und Antigenexpression wie T1C0011. In durchflusszytometrischen Punktdiagrammen (C) befanden sich Anteile von T1C0011 und T2C0002

in der sehr niedrigen FS-Region von apoptotischen/abgestorbenen CLL-Zellen.

Die Populationen T1C0011 und T2C0002 wiesen im Vergleich zu durchschnittlichen CLL-Zellen also eine verringerte Vorwärtsstreuungsexpression sowie eine geringere Antigenexpression auf. In der MPFC wurden diese Populationen teilweise im Bereich der toten und apoptotischen CLL-Zellen lokalisiert (*Abbildung 10C*). Dieser Befund deutet darauf hin, dass eine erhöhte Anzahl von toten und apoptotischen CLL-Zellen bei der Messung mit einer schlechteren Prognose und einem schlechteren Therapieverlauf verbunden sein könnte. Im Gegensatz dazu zeigten die CLL-Untergruppen T2C0004, T2C0014, T2C0018 und T2C0020 eine höhere Antigenexpression und ein höheres Streulichtprofil als die durchschnittlichen CLL-Zellen.

Zusammenfassend lässt sich sagen, dass CLL-Zellen mit geringer Vorwärtsstreuungsexpression (T1C0011, T2C0002) sowie CLL-Zellen mit hoher Vorwärtsstreuungsexpression (T2C0004, T2C0014,

T2C0018 und T2C0020) beide auf einen schlechteren klinischen Verlauf schließen lassen können. Es ist also besonders interessant auf CLL-Zellen mit Streulichteigenschaften in den Randbereichen zu achten (besonders hoch oder besonders niedrig im Vergleich zur durchschnittlichen CLL-Hauptpopulation), da diese als prognostisch ungünstig identifiziert werden konnten.

3.5 Klinische Anwendbarkeit

Mit ALPODS konnten 17 XAI-Populationen identifiziert werden, die das Ergebnis (TTF 0 vs. TTF 1) in Kombination statistisch signifikant vorhersagen konnten (AUC 0,95 [95%CI 0,91-098; $p < 0,0001$]). Ein manuelles Gating dieser Populationen und insbesondere der CLL-Untergruppen ist ohne Verwendung von ALPODS nicht gut möglich. Ausnahmen waren T1C0016 und T1C0023, bestehend aus CD4+ T-Zellen bzw.

CD8+ T-Zellen, denn beide Populationen ließen sich manuell leicht in durchflusszytometrischen Punktdiagrammen erfassen. Dementsprechend erfolgte die Untersuchung, ob T1C0016 (CD4+ T-Zellen) und T1C0023 (CD8+ T-Zellen) einen zusätzlichen prädiktiven Wert für den CLL-IPI-Score sowie CD38+ CLL-Zellen haben. Dafür wurde der Vorhersagewert von verschiedenen Modellen untersucht, die sich gut auf andere durchflusszytometrische Panels übertragen lassen.

In praktisch gut anwendbaren Vier- sowie Zwei-Faktor-Modellen wurde eine multiple logistische Regression durchgeführt (*Tabelle 14*). Odds Ratio (OR) >1 begünstigte ein schlechteres Ergebnis und <1 begünstigte ein besseres Ergebnis.

Tabelle 14: Vorhersagefähigkeit für das Ergebnis (TTF 0 vs. TTF 1) von drei Modellen

	OR	95% KI	p-Wert	AUC	95% KI	p-Wert
Alle XAI-Populationen				0,95	0,91-0,98	<0,0001
CLL-IPI				0,78	0,70-0,86	<0,0001
Vier-Faktoren-Modell						
CLL-IPI	1,53	1,19-2,04	0,0018			
T1C0016 (CD4+ T-Zellen)	0,86	0,78-0,93	0,0003	0,84	0,77-0,91	<0,0001
T1C0023 (CD8+ T-Zellen)	1,14	0,95-1,77	0,3305			
CD38	1,01	1,00-1,02	0,1790			
Zwei-Faktoren-Modell mit IPI						
CLL-IPI	1,64	1,29-2,16	0,0001	0,83	0,77-0,90	<0,0001
T1C0016 (CD4+ T-Zellen)	0,85	0,78-0,92	0,0002			
Zwei-Faktoren-Modell ohne IPI						
T1C0016 (CD4+ T-Zellen)	0,85	0,78-0,91	<0,0001	0,79	0,71-0,87	<0,0001
T1C0023 (CD8+ T-Zellen)	1,29	1,05-2,14	0,1390			

Abkürzungen: OR = odds ratio; TTF = time to first line treatment failure (Zeit bis zum Versagen der Erstlinientherapie); AUC = area under curve; KI = Konfidenzintervall, IPI = international prognostic index.

In dem Vier-Faktoren-Modell zeigten sich T1C0023 (OR 1,14; 95%CI 0,95-1,77; p=0,3305) und CD38-positive CLL-Zellen (OR 1,01; 95%CI 1,00-1,02; p=0,1790) als entbehrlich. CLL-IPI und T1C0016 (Zwei-Faktoren-Modell mit CLL-IPI) zeigten nur einen geringfügig niedrigeren prädiktiven Wert als das Vier-Faktoren-Modell (Vier-Faktoren-Modell: 0,84 AUC; 95%CI 0,77-0,91; p <0,0001; vs. Zwei-Faktoren-Modell: 0,83 AUC; 95%CI 0,77-0,90; p<0,0001). Im Zwei-Faktoren-Modell (IPI und T1C0016) kann CLL-IPI durch T1C0023 mit

akzeptabler diagnostischer Fähigkeit ersetzt werden (0,79 AUC; 95%CI 0,71-0,87; $p < 0,0001$).

Alleine mit der Bestimmung der CD4+- und CD8+-Zellen konnte also in fast 80% der Fälle der betrachteten Patientenkohorte korrekt zwischen TTF 0 und TTF 1 unterschieden werden.

4. Diskussion

Im Rahmen dieser Arbeit konnten aus MPFC-Daten 17 XAI-Populationen identifiziert werden, die die klinische Prognose der CLL besser vorhersagen konnten als der etablierte Prognosescore CLL-IPI oder die Häufigkeit von CD38-positiven CLL-Zellen.

Ein Großteil der XAI-Populationen befand sich teilweise oder auch vollständig innerhalb der CLL-Population. Jedoch gab es auch einzelne XAI-Populationen, die nicht als bösartig einzuordnen sind: die T1C0016-Zellpopulation bestand beispielweise ausschließlich aus CD4+ T-Zellen und war bei Patienten mit schlechterem klinischem Ausgang unterrepräsentiert. Eben diese Zellpopulation T1C0016 (CD4+ T-Zellen) war von allen 17 XAI-Populationen auch der beste einzelne Klassifikator für die Vorhersage des klinischen Ausgangs. Im Gegensatz dazu war die Zellpopulation T1C0023 bestehend aus CD8+ T-

Zellen bei Patienten mit schlechterem Ergebnis überrepräsentiert.

4.1 Diskussion klinischer Ergebnisse

In der Literatur sind T-Zellen bei CLL bereits als dysreguliert beschrieben worden. CD4⁺ T-Zellen und CD8⁺ T-Zellen bei CLL-Patienten unterscheiden sich von denen bei gesunden Personen u.a. durch die Anhäufung von Gedächtnis-T-Zellen bei Verlust von naiven T-Zellen, der verstärkten Expression von Immun-Checkpoint-Rezeptoren (z.B. PD1, TIGIT, CTLA-4) und einer erhöhten Aktivierung.^{16,29,87,126,90,108} Außerdem zeigt sich bei der CLL typischerweise eine Inversion des CD4/CD8-Verhältnisses mit Werten <1 .^{71,77,78,91} Elston et al. konnten zeigen, dass Patienten mit einem CD4/CD8-Verhältnis >1 ein besseres Gesamtüberleben und progressionsfreies Überleben haben.²⁹ Eine erhöhte Anzahl von CD8⁺ Zellen konnte mit einer Krankheitsprogression, kürzerer Time-to-First-Treatment und Progression-Free-

Survival assoziiert werden, möglicherweise aufgrund der Koexpression des PD-1-Rezeptors.⁸³ Diese Ergebnisse decken sich mit denen dieser Arbeit, wonach CD4+ T-Zellen auf einen guten und CD8+ T-Zellen auf einen schlechten Ausgang hinweisen. Hieran anschließend wäre es sehr interessant zu untersuchen, welche Untergruppen von CD4+ T-Zellen einen besonderen Einfluss auf einen günstigen Krankheitsverlauf nehmen.

Die Rolle der CD4+ und CD8+ T-Zellen bei der CLL wird trotz obiger Studienergebnisse weiterhin kritisch diskutiert, denn noch immer ist nicht klar, ob die Zellen das Wachstum bösartiger B-Zellen eher unterstützen oder der Tumorkontrolle dienen.⁹⁹ Gonzalez-Rodriguez et al. zeigten beispielsweise, dass ein höherer Quotient aus CD8+ T-Zellen und CLL-Zellen mit einem besseren Gesamtüberleben korrelierte.³⁶ Dies deutet darauf hin, dass CD8+ T-Zellen an der CLL-Kontrolle beteiligt sein könnten. Weiterhin zeigte sich in dieser Studie eine höhere relative Anzahl an CD4+ Zellen mit besserem

Überleben assoziiert³⁶, was wiederum mit den Ergebnissen dieser Arbeit einhergeht.

Im Gegensatz zu den XAI-Populationen von CLL-Untergruppen ist das Gating der CD4+ und CD8+ T-Zellen einfach und insbesondere gut auf andere MPFC-Panels übertragbar, die Antikörper gegen CD4 und CD8 enthalten. Auf Basis dessen wurde in dieser Arbeit ein vereinfachter Zwei-Faktoren-Ansatz entwickelt, um das Ergebnis bei CLL durch die Kombination des CLL-IPI-Scores und den CD4+ T-Zellen oder der Kombination der CD4+ T-Zellen und CD8+ T-Zellen vorherzusagen. Beide Zwei-Faktoren-Modelle unterscheiden in 83% bzw. 79% der CLL-Fälle korrekt zwischen einem schlechteren und einem besseren klinischen Verlauf.

Darüber hinaus konnten mit Hilfe der XAI-Populationen auch immunpathologische Erkenntnisse der CLL gewonnen werden. So wurde beispielsweise deutlich, dass es sich bei einzelnen CLL-Untergruppen, die ein schlechteres Ergebnis

vorhersagen, um kleine apoptotische/abgestorbene CLL-Ereignisse (Zellpopulationen T1C0011, T2C0002) mit geringer Vorwärtsstreuung handelt. Dies steht im Einklang mit den Ergebnissen von Witkowska et al.¹⁴⁰ und Jahrsdörfer et al.⁵², die beschreiben, dass die spontane In-vitro-Apoptose von CLL-Zellen mit einem Krankheitsprogress sowie einer mit schlechter Prognose assoziierten Zytogenetik korreliert.

In dieser Arbeit wurden CLL-Untergruppen mit hoher Vorwärts- und Seitenstreuung (Zellpopulationen T2C0004, T2C0014, T2C0018, T2C0020) ebenfalls mit einem ungünstigen Ergebnis in Verbindung gebracht. Die Vorwärts- und Seitenstreuung korreliert mit einer größeren Zellgröße und höherer interner Komplexität, was auf die prognostische Bedeutung von den größeren Prolymphozyten bei CLL hindeutet. Oscier et al.⁸⁶ beschrieben, dass ein Anteil von Prolymphozyten >10% bei CLL mit einem kürzeren Gesamtüberleben und progressionsfreiem

Überleben assoziiert ist. Außerdem konnte eine Assoziation zwischen der Zellgröße und der prognostisch ungünstigen CD38+-Expression nachgewiesen werden.⁶⁶ Zusätzliche Untersuchungen weiterer CLL-Prognosemarker wie beispielsweise CD49d und ZAP-70 würden die Schlussfolgerungen dieser Arbeit untermauern.¹⁴

Diese Dissertation liefert zusätzlich zu den etablierten CLL-Prognosemarkern neue Erkenntnisse über prognostische Faktoren, die mit der Immunologie der CLL und dem nicht-malignen, reaktiven Immunsystem zusammenhängen.

4.2 Diskussion des Patientenkollektivs

Bei einigen Patienten lagen nicht für alle Parameter des CLL-IPI-Scores vollständige Daten vor. In diesen Fällen wurde ein halber Punktwert zugewiesen, was zu einer vermehrten Kategorisierung in den mittleren Prognosebereichen

(mittleres Risiko mit 79,3% Gesamtüberleben nach 5 Jahren bei IPI-Score 2-3) führen könnte. Darüber hinaus hätte eine größere Patientenkohorte eine multivariate Cox-Regression ermöglicht. Dies würde einen besseren Anhaltspunkt dafür liefern, ob die Ergebnisse von verschiedenen Behandlungsstrategien unabhängig sind, denn in der betrachteten Patientenkohorte wurden unterschiedliche Therapieregimes verwendet. In *Tabelle 5* und *6* zur Erst- und Zweitlinientherapie wird deutlich, dass die aktuellen Therapieleitlinien zumeist eine Therapie mit BTK-Inhibitoren, monoklonalen Antikörpern oder Venetoclax vorsehen; die Anwendung der klassischen zytostatischen Therapie erfolgt zwar in bestimmten Fällen weiterhin, der Anteil der chemotherapeutisch u.a. mit R-Bendamustin behandelten Patienten in der hier betrachteten Patientenkohorte ist jedoch relativ hoch. Die Therapie beeinflusst die Prognose, sodass auch hier in weiteren Studien eine größere Kohorte mit möglichst aktuellen Therapieschemata betrachtet werden sollte.

Eine weitere Limitation dieser Arbeit sind die Unterschiede der Follow-Up-Zeiten zwischen TTF 0 und TTF 1 (Median in Monaten: 24 (7-46) bei TTF 0 vs. 71 (38-114) bei TTF 1), sodass möglicherweise Informationen bezüglich des weiteren Krankheitsverlaufs der als TTF 0 eingestuften Patienten fehlen. Auch hier könnte eine größere Patientenkohorte zu einer Angleichung der Follow-Up-Zeiten führen.

4.3 Diskussion von ALPODS und weiterer Clusteralgorithmen

Ein Problem des manuellen Gatings im Zuge der Analyse der durchflusszytometrischen Daten ist der potenzielle Informationsverlust durch die lediglich zweidimensionale Darstellung, welche die komplexe hochdimensionale Datenstruktur nicht ausreichend widerspiegelt. Darüber hinaus zeigt es eine Reihe weiterer Nachteile: insbesondere im klinischen Alltag besteht eine Anfälligkeit für eine mögliche menschliche Verzerrung, Fehler bei der

Datenanalyse sowie eine geringe Verarbeitungsgeschwindigkeit.⁴⁹ Diese Einschränkungen sind insbesondere bei der Untersuchung von Tumorzellen bedeutend, da deren Segmentierung in zelluläre Subpopulationen im Allgemeinen weniger definiert ist als die von gesunden Zellen, welche anhand der Expression von Zelloberflächenmarkern oftmals relativ einfach in Gruppen unterteilt werden können.⁵³ Diesen Limitationen kann mit Machine-Learning-Algorithmen entgegengesteuert werden.

Keyes et al.⁵³ haben vier verschiedene Clustering-Algorithmen auf einen relativ einfachen Durchflusszytometrie-Datensatz (von Good et al.³⁷) mit insgesamt 30.000 Zellereignissen (je 10.000 Zellen von drei Patienten mit einer Hochrisiko-Vorläufer-B-Zell akuten lymphatischen Leukämie (BCP-ALL)) angewandt. Der PhenoGraph-Algorithmus⁵⁹ erkannte automatisch vier Cluster, sodass diese Anzahl als Vorgabe für die drei weiteren Algorithmen K-means Clustering⁴³, Spanning-tree Analysis of Density-normalized

Events (SPADE)¹⁴² und FlowSOM¹³³ genutzt wurde. Für alle Clustering-Algorithmen wurden die gleichen Marker genutzt. Trotz der möglichst gleichen Voraussetzungen und des relativ einfachen Datensatzes zeigte sich sehr eindrücklich, dass alle vier Algorithmen unterschiedliche Zellcluster identifiziert haben. Um den Algorithmus zu finden der am besten für den eigenen Datensatz passt, sollte man verschiedene Clustering-Ansätze iterativ testen, die Ergebnisse visualisieren und die Hyperparameter anpassen.⁵³ Vor diesem Hintergrund wäre es interessant ALPODS mit weiteren Algorithmen zu vergleichen.

Die prognostisch relevanten Zellpopulationen konnte ALPODS durch die Analyse der MPFC-Daten der gesamten Patientenkohorte detektieren. Die gefundenen Cluster decken also die Zellen aller Patienten der Kohorte ab. Ein individuelleres, populationsspezifisches Gating wäre jedoch auch wünschenswert, sodass man prognostisch relevante Zellpopulationen für einzelne Patienten finden kann.

Der ALPODS-XAI-Algorithmus identifizierte prädiktive Immunzellpopulationen, die eine Vorhersage des klinischen Verlaufs bei CLL ermöglichen. CD4+ T-Zellen waren der beste Einzelklassifikator, die Kombination mit dem CLL-IPI-Score konnte die Vorhersagegenauigkeit weiter verbessern. Zudem konnte gezeigt werden, dass mit dem klinisch relativ simpel durchführbarem Gating von CD4+ T-Zellen in Kombination mit CD8+ T-Zellen ebenfalls eine klinische Vorhersage getroffen werden kann. Diese Arbeit zeigt dass der ALPODS-XAI-Algorithmus das Potenzial hat, konventionelle Prognosemodelle wie den CLL-IPI zu präzisieren.

5. Zusammenfassung

Einleitung: Die Chronische lymphatische Leukämie (CLL) ist die häufigste leukämische Erkrankung in den westlichen Ländern. Ein etabliertes Prognosemodell für CLL-Patienten ist der International Prognostic Index (CLL-IPI), der auf fünf Prognosefaktoren basiert (TP53-Status, IGHV-Mutationsstatus, beta2-Mikroglobulin im Serum, klinisches Stadium und das Alter). In dieser Arbeit wird untersucht, ob eine CLL-Prognose möglich ist, indem ausschließlich Multiparameter-Durchflusszytometriedaten (MPFC) von CLL-Patienten mit Hilfe von einer erklärbaren künstlichen Intelligenz (XAI) analysiert werden. Dabei sollte die XAI die Ergebnisse basierend auf der Darstellung von unterschiedlichen, für die Prognose relevanten Zellpopulationen in MPFC-Punktdiagrammen erklären können.

Material und Methoden: Insgesamt erfolgte die Analyse von MPFC-Daten aus dem peripheren Blut

von 157 Patienten mit CLL und der Abgleich mit klinischen Daten (Parameter des CLL-IPI, Geschlecht, ECOG-Stadium, Richter-Transformation ja/nein, Behandlung ja/nein, Todesdatum, Behandlungsversagen und Datum der letzten Nachuntersuchung). Die MPFC-Daten wurden erhoben für die routinemäßige diagnostische Analyse am Universitätsklinikum Marburg von 2014 bis 2020. Die Patienten wurden in eine Gruppe mit schlechterem und eine Gruppe mit besserem klinischen Verlauf eingeteilt. Patienten, die während der Nachbeobachtung verstarben sowie diejenigen bei denen die systemische Erstlinientherapie nicht angesprochen hat, wurden als TTF 1 (Time-to-First-Line-Treatment-Failure bzw. Zeit bis zum Versagen der Erstlinientherapie) kategorisiert, d. h. der Gruppe mit schlechterem klinischen Verlauf zugewiesen. Alle anderen Patienten wurden mit TTF 0 einem besseren Ergebnis zugeordnet. Anschließend wurde der XAI-Algorithmus ALPODS verwendet, um Zellpopulationen zu identifizieren, die bei Patienten mit schlechterem

(TTF 1) oder besserem klinischen Verlauf (TTF 0) über- oder unterrepräsentiert waren und damit eine Vorhersage zur Prognose treffen können. Dabei wurde die prädiktive Fähigkeit jeder einzelnen XAI-Population mit Hilfe von Receiver-Operating-Characteristic-Kurven (ROC-Kurven) ausgewertet. Zur Überprüfung wurde dieser prädiktive Wert der XAI-Populationen mit der Häufigkeit von CD38-positiven CLL-Zellen und CLL-IPI auf den ROC-Kurven verglichen.

Ergebnisse: ALPODS definierte 17 XAI-Populationen, die zusammengenommen eine genauere Prognose als der CLL-IPI-Score treffen können (AUC der XAI-Populationen: 0,95 [95%KI 0,91-0,98; $p < 0,0001$] vs. AUC von CLL-IPI: 0,78 [95%KI 0,70-0,86; $p < 0,0001$]). Der beste Einzelklassifikator war eine XAI-Population bestehend aus CD4+ T-Zellen (AUC 0,78; 95%KI 0,70–0,86; $p < 0,0001$). Patienten mit weniger CD4+ T-Zellen zeigten einen schlechteren klinischen

Verlauf. Bei Hinzufügen der CD4+ T-Zellpopulation steigerte sich die Vorhersagefähigkeit des CLL-IPI-Scores (AUC 0,83; [95%KI 0,77–0,90; $p < 0,0001$]). Alleine mit einer Kombination der CD4+ mit der CD8+ T-Zellpopulation ließ sich in 79% der Fälle der betrachteten Patientenkohorte korrekt unterscheiden zwischen TTF 0 und TTF 1 (AUC 0,79; [95%KI 0,71–0,87; $p < 0,0001$]). Dies ist insbesondere vor dem Hintergrund wichtig, dass sich diese beiden Zellpopulationen manuell leicht in durchflusszytometrischen Punktdiagrammen erfassen ließen, gerade im Vergleich zu anderen von ALPODS definierten Zellpopulationen, deren korrektes manuelles Gating zum Teil sehr anspruchsvoll ist (insbesondere das der CLL-Untergruppen).

Schlussfolgerung:

Der ALPODS-XAI-Algorithmus war in der Lage prädiktive Immunzellpopulationen zu identifizieren, die den klinischen Verlauf einer CLL signifikant

vorhersagen können. CD4+ T-Zellen wurden als bester Einzelklassifikator identifiziert, wobei sie in Kombination mit dem CLL-IPI-Score die Vorhersagefähigkeit nochmal verbessern konnten. Darüber hinaus konnte mit dieser Arbeit gezeigt werden, dass mit dem in der Klinik relativ einfach durchführbarem Gating von CD4+ T-Zellen in Kombination mit den CD8+ T-Zellen ebenfalls eine klinische Prognose getroffen werden kann. Die Ergebnisse müssen an einer wesentlich größeren Patientenkohorte validiert werden, sind aber sehr vielversprechend was die Verfeinerung konventioneller Prognosemodelle wie den CLL-IPI betrifft.

6. Summary

Introduction: Chronic lymphocytic leukemia (CLL) is the most common leukemic disease in Western countries. An established prognostic model for CLL patients is the International Prognostic Index (CLL-IPI), which is based on five prognostic factors (TP53 status, IGHV mutation status, serum beta2-microglobulin, clinical stage, and age). In this work, we investigate whether CLL prognosis is possible by exclusively analyzing multiparameter flow cytometry (MPFC) data from CLL patients using an explainable artificial intelligence (XAI). Here, the XAI should be able to explain the results based on the representation of different cell populations relevant for prognosis in MPFC dot plots.

Material and Methods: Analysis of MPFC data from peripheral blood of 157 patients with CLL was performed in alignment with matching clinical data (parameters of CLL IPI, sex, ECOG stage, Richter

transformation yes/no, treatment yes/no, date of death, treatment failure, and date of last follow-up). MPFC data were collected for routine diagnostic analysis at Marburg University Hospital from 2014 to 2020, and patients were divided into a worse outcome group and a better outcome group. Patients who died during follow-up and those who failed first-line systemic therapy were categorized as TTF 1 (time-to-first-line treatment failure), i.e., assigned to the worse outcome group. All other patients were assigned to a better outcome with TTF 0. Subsequently, the XAI algorithm ALPODS was used to identify cell populations that were over- or underrepresented in patients with worse outcome (TTF 1) or better outcome (TTF 0) and thus predictive of prognosis. Here, the predictive ability of each XAI population was evaluated using receiver operating characteristic (ROC) curves. For verification, this predictive value of the XAI populations was compared with the frequency of CD38-positive CLL cells and CLL-IPI on the ROC curves.

Results: ALPODS defined 17 XAI populations that collectively can predict more accurately than the CLL-IPI score (AUC of XAI populations: 0.95 [95%CI 0.91-0.98; $p < 0.0001$] vs. AUC of CLL-IPI: 0.78 [95%CI 0.70-0.86; $p < 0.0001$]). The best single classifier was an XAI population consisting of CD4+ T cells (AUC 0.78; 95% CI 0.70-0.86; $p < 0.0001$). Patients with fewer CD4+ T cells showed a worse clinical outcome. When the CD4+ T cell population was added, the predictive ability of the CLL-IPI score increased (AUC 0.83; [95%CI 0.77-0.90; $p < 0.0001$]). Combining the CD4+ with the CD8+ T cell population alone correctly distinguished TTF 0 from TTF 1 in 79% of the cases in the patient cohort considered (AUC 0.79; [95%CI 0.71-0.87; $p < 0.0001$]). This is particularly important in light of the fact that these two cell populations could be easily gated manually in flow cytometric dot plots, especially compared to other cell populations defined by ALPODS, some of which are

very challenging to correctly gate manually (especially that of CLL subgroups).

Conclusion: The ALPODS-XAI algorithm was able to identify predictive immune cell populations that significantly predicted the clinical course of CLL. CD4⁺ T cells were identified as the best single classifier, further improving predictive ability when combined with the CLL-IPI score. In addition, this work demonstrated that gating CD4⁺ T cells in combination with the CD8⁺ T cells, which is relatively easy to perform in the clinic, can also be used to make a clinical prediction. The results need to be validated in a much larger cohort of patients, but are very promising in terms of refining conventional prognostic models such as the CLL-IPI.

Literaturverzeichnis

1. Abdelaal T, van Unen, V., Höllt, T., Koning, F., Reinders, M. J., & Mahfouz, A. . Predicting Cell Populations in Single Cell Mass Cytometry Data - PubMed. Cytometry Part A : the journal of the International Society for Analytical Cytology. 2019 Jul;95(7).
2. Abdi H, Williams LJ. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics. 2010/07/01;2(4).
3. Adadi A, & Berrada, M. . Explainable AI for healthcare: from black box to interpretable models. Embedded systems and artificial intelligence: proceedings of ESAI 2019, Fez, Morocco Springer Singapore. 2020:327-37.
4. Adadi A, Berrada M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI) | IEEE Journals & Magazine | IEEE Xplore. IEEE Access. 2018;6.
5. Alaggio R, Amador C, Anagnostopoulos I, Attygalle AD, Araujo IBdO, Berti E, et al. The 5th edition of the World Health Organization

Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia* 2022 36:7. 2022-06-22;36(7).

6. Arvaniti E, Claassen M. Sensitive detection of rare disease-associated cell subsets via representation learning. *Nature Communications*. 2017;8(1).

7. Balakrishnama S, Ganapathiraju A. Linear discriminant analysis-a brief tutorial. Institute for Signal and information Processing. 1998;18(1998):1-8.

8. Baliakas P, Agathangelidis A, Hadzidimitriou A, Sutton L-A, Minga E, Tsanousa A, et al. Not all IGHV3-21 chronic lymphocytic leukemias are equal: prognostic considerations. *Blood*. 2015/01/29;125(5).

9. Baliakas P, Hadzidimitriou A, Sutton L-A, Rossi D, Minga E, Villamor N, et al. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia* 2015 29:2. 2014-06-19;29(2).

10. Berrar D. Cross-Validation. Encyclopedia of Bioinformatics and Computational Biology 2019. p. 542-5.
11. Binet JL, Auquier A, Dighiero G, Chastang C, Piguet H, Goasguen J, et al. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer*. 1981/07/01;48(1).
12. Breiman L, Breiman L. Random Forests. *Machine Learning* 2001 45:1. 2001/10;45(1).
13. Bruggner RV, Bodenmiller B, Dill DL, Tibshirani RJ, Nolan GP. Automated identification of stratifying signatures in cellular subpopulations. *Proceedings of the National Academy of Sciences*. 2014-7-1;111(26).
14. Bulian P, Shanafelt TD, Fegan C, Zucchetto A, Cro L, Nüchel H, et al. CD49d Is the Strongest Flow Cytometry–Based Predictor of Overall Survival in Chronic Lymphocytic Leukemia. *Journal of Clinical Oncology*. 2014/03/03;32(9).
15. Cabitza F, Campagner, A., Malgieri, G., Natali, C., Schneeberger, D., Stoeger, K., &

Holzinger, A. Quod erat demonstrandum? - Towards a typology of the concept of explanation for the design of explainable AI. *Expert Systems with Applications*. 2023/03/01;213.

16. Catakovic K, Gassner FJ, Ratswohl C, Zaborsky N, Rebhandl S, Schubert M, et al. TIGIT expressing CD4+T cells represent a tumor-supportive T cell subset in chronic lymphocytic leukemia. *OncoImmunology*. 2018-1-2;7(1).

17. Chaddad A, Peng J, Xu J, Bouridane A. Survey of Explainable AI Techniques in Healthcare. *Sensors* 2023, Vol 23, Page 634. 2023-01-05;23(2).

18. Cheng L, Karkhanis P, Gokbag B, Liu Y, Li L. DGCytoF: Deep learning with graphic cluster visualization to predict cell types of single cell mass cytometry data. *PLOS Computational Biology*. 11.04.2022;18(4).

19. Cohen J. *Statistical power analysis for the behavioral sciences.*: Routledge; 2013.

20. Cramer P, Hallek M. Prognostic factors in chronic lymphocytic leukemia—what do we need to

know? Nature Reviews Clinical Oncology 2011 8:1.
2010-10-19;8(1).

21. Dal Bo M, Tissino, E., Benedetti, D.,
Caldana, C., Bomben, R., Del Poeta, G., ... & Gattei,
V. . Functional and clinical significance of the
integrin alpha chain CD49d expression in chronic
lymphocytic leukemia. Current cancer drug targets.
2016;16(8):659-68.

22. Damle RN, Wasil T, Fais F, Ghiotto F,
Valetto A, Allen SL, et al. Ig V Gene Mutation
Status and CD38 Expression As Novel Prognostic
Indicators in Chronic Lymphocytic
Leukemia Presented in part at the 40th Annual
Meeting of The American Society of Hematology,
held in Miami Beach, FL, December 4-8, 1998.
Blood. 1999/09/15;94(6).

23. Das A, & Rad, P. . Opportunities and
challenges in explainable artificial intelligence (xai):
A survey. arXiv preprint arXiv:200611371. 2020.

24. Delgado J, Doubek M, Baumann T,
Kotaskova J, Molica S, Mozas P, et al. Chronic
lymphocytic leukemia: A prognostic model

comprising only two biomarkers (IGHV mutational status and FISH cytogenetics) separates patients with different outcome and simplifies the CLL-IPI. *American Journal of Hematology*. 2017/04/01;92(4).

25. Döhner H, Stilgenbauer S, Benner A, Leupolt E, Kröber A, Bullinger L, et al. Genomic Aberrations and Survival in Chronic Lymphocytic Leukemia.

<https://doi.org/10.1056/NEJM200012283432602>.

2000-12-28;343(26).

26. Dorfman DM, LaPlante CD, Li B. FLOCK cluster analysis of plasma cell flow cytometry data predicts bone marrow involvement by plasma cell neoplasia - PubMed. *Leukemia research*. 2016 Sep;48.

27. Du M, Liu, N., & Hu, X. . Techniques for interpretable machine learning. *Communications of the ACM*. 2019;63(1):68-77.

28. Eilers PH, Goeman JJ. Enhancing scatterplots with smoothed densities. *Bioinformatics*. 2004;20(5):623-8.

29. Elston L, Fegan C, Hills R, Hashimdeen SS, Walsby E, Henley P, et al. Increased frequency of CD4+PD-1+HLA-DR+ T cells is associated with disease progression in CLL. *British Journal of Haematology*. 2020/03/01;188(6).
30. Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves | Canadian Journal of Emergency Medicine | Cambridge Core. *Canadian Journal of Emergency Medicine*. 2006/01;8(1).
31. Faraggi D, Reiser B. Estimation of the area under the ROC curve. *Statistics in Medicine*. 2002/10/30;21(20).
32. Fischer K, Bahlo J, Fink AM, Goede V, Herling CD, Cramer P, et al. Long-term remissions after FCR chemoimmunotherapy in previously untreated patients with CLL: updated results of the CLL8 trial. *Blood*. 2016/01/14;127(2).
33. Fisher RA. The Use of multiple measurements in taxonomic problems. *Annals of Eugenics*. 1936/09/01;7(2).

34. Ginaldi L, Martinis MD, Matutes E, Farahat N, Morilla R, Catovsky D. Levels of expression of CD19 and CD20 in chronic B cell leukaemias. *Journal of Clinical Pathology*. 1998-05-01;51(5).
35. Goebel R, Chander A, Holzinger K, Lecue F, Akata Z, Stumpf S, et al. Explainable AI: the new 42? Machine Learning and Knowledge Extraction: Second IFIP TC 5, TC 8/WG 84, 89, TC 12/WG 129 International Cross-Domain Conference, CD-MAKE 2018 (eds Holzinger, A, Kieseberg, P, Tjoa, A, & Weippl, E), Springer. 2018:295–303.
36. Gonzalez-Rodriguez AP, Contesti J, Huergo-Zapico L, Lopez-Soto A, Fernández-Guizán A, Acebes-Huerta A, et al. Prognostic significance of CD8 and CD4 T cells in chronic lymphocytic leukemia. *Leukemia & Lymphoma*. 2010-10-1.
37. Good Z, Sarno J, Jager A, Samusik N, Aghaeepour N, Simonds EF, et al. Single-cell developmental classification of B cell precursor acute lymphoblastic leukemia at diagnosis reveals predictors of relapse. *Nature Medicine* 2018 24:4. 2018-03-05;24(4).

38. Hallek M, Al-Sawaf O. Chronic lymphocytic leukemia: 2022 update on diagnostic and therapeutic procedures. *American Journal of Hematology*. 2021/12/01;96(12).
39. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood*. 2018/06/21;131(25).
40. Hallek M, Langenmayer I, Nerl C, Knauf W, Dietzfelbinger H, Adorf D, et al. Elevated Serum Thymidine Kinase Levels Identify a Subgroup at High Risk of Disease Progression in Early, Nonsmoldering Chronic Lymphocytic Leukemia. *Blood*. 1999/03/01;93(5).
41. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig VH Genes Are Associated With a More Aggressive Form of Chronic Lymphocytic Leukemia. *Blood*. 1999/09/15;94(6).
42. Hamblin TJ, Orchard JA, Ibbotson RE, Davis Z, Thomas PW, Stevenson FK, et al. CD38

expression and immunoglobulin variable region mutations are independent prognostic variables in chronic lymphocytic leukemia, but CD38 expression may vary during the course of the disease. *Blood*. 2002/02/01;99(3).

43. Hartigan JA, Wong MA. Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society series c (applied statistics)*, 28(1), 100-108. 1979.

44. Heintel D, Schwarzinger I, Chizzali-Bonfadin C, Thalhammer R, Schwarzmeier J, Fritzer-Szekeres M, et al. Association of CD38 antigen expression with other prognostic parameters in early stages of chronic lymphocytic leukemia. *Leuk Lymphoma*. 2001;42(6):1315-21.

45. Hoffman RR, Mueller ST, Klein G, Litman J. Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv:181204608*. 2018.

46. Hoffmann J, Rother M, Kaiser U, Thrun MC, Wilhelm C, Gruen A, et al. Determination of CD43 and CD200 surface expression improves accuracy of

B-cell lymphoma immunophenotyping. *Cytometry Part B: Clinical Cytometry*. 2020/11/01;98(6).

47. Hoffmann J, Thrun MC, Röhnert MA, Bonin Mv, Oelschlägel U, Neubauer A, et al. Identification of critical hemodilution by artificial intelligence in bone marrow assessed for minimal residual disease analysis in acute myeloid leukemia: The Cinderella method. *Cytometry Part A*. 2023/04/01;103(4).

48. Holzinger A, Biemann, C., Pattichis, C. S. & Kell, D. B. . What do we need to build explainable AI systems for the medical domain? arXiv preprint arXiv:171209923. 2017.

49. Hu Z, Bhattacharya S, Butte AJ. *Frontiers | Application of Machine Learning for Cytometry Data*. *Frontiers in Immunology*. 2022/01/03;12.

50. Hu Z, Tang A, Singh J, Bhattacharya S, Butte AJ, Hu Z, et al. A robust and interpretable end-to-end deep learning model for cytometry data. *Proceedings of the National Academy of Sciences*. 2020-9-1;117(35).

51. Hurlbert SH. The Nonconcept of Species Diversity: A Critique and Alternative Parameters. *Ecology*. 1971/07/01;52(4).
52. Jahrsdörfer B, Wooldridge JE, Blackwell SE, Taylor CM, Link BK, Weiner GJ, et al. Good prognosis cytogenetics in B-cell chronic lymphocytic leukemia is associated in vitro with low susceptibility to apoptosis and enhanced immunogenicity. *Leukemia* 2005 19:5. 2005-03-10;19(5).
53. Keyes TJ, Domizi P, Lo Y-C, Nolan GP, Davis KL. A Cancer Biologist's Primer on Machine Learning Applications in High-Dimensional Cytometry. *Cytometry Part A*. 2020/08/01;97(8).
54. Kleinbaum DG, Dietz, K., Gail, M., Klein, M., & Klein, M. . Logistic regression. New York: Springer-Verlag. 2002:536.
55. Kohonen T. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 1982 43:1. 1982/01;43(1).
56. Kohonen T. The self-organizing map | IEEE Journals & Magazine | IEEE Xplore.

57. Kröber A, Seiler T, Benner A, Bullinger L, Brückle E, Lichter P, et al. VH mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. *Blood*. 2002/08/15;100(4).
58. Lee DK. Alternatives to P value: confidence interval and effect size. *Korean Journal of Anesthesiology*. 2016/10/25;69(6).
59. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir E-aD, Tadmor M, et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell*. 2015/07/07;162(1).
60. Lipton ZC. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*. 2018;16(3):31-57.
61. Litjens G, Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. A survey on deep learning in medical image analysis. *Medical Image Analysis*. 2017/12/01;42.

62. Luan S, Schooler, L. J., & Gigerenzer, G. . A signal-detection analysis of fast-and-frugal trees. *Psychological review*. 2011;118 (2):316-38.
63. Lundberg SM, & Lee, S. I. A unified approach to interpreting model predictions. *Advances in neural information processing systems*. 2017;30.
64. Magnac C, Porcher R, Davi F, Nataf J, Payelle-Brogard B, Tang RP, et al. Predictive value of serum thymidine kinase level for Ig-V mutational status in B-CLL. *Leukemia* 2003 17:1. 2003-01-03;17(1).
65. Mamdani EH, & Assilian, S. . An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*. 1975;7(1-13).
66. Manocha S, Matrai Z, Osthoff M, Carter A, Pettitt AR. Correlation Between Cell Size and CD38 Expression in Chronic Lymphocytic Leukaemia. *Leukemia & Lymphoma*. 2003-1-1.
67. Mansfield AS, Rabe KG, Slager SL, Schwager SM, Call TG, Brewer JD, et al. Skin

Cancer Surveillance and Malignancies of the Skin in a Community-Dwelling Cohort of Patients With Newly Diagnosed Chronic Lymphocytic Leukemia. *Journal of Oncology Practice*. 2013-08-27;10(1).

68. Marti GE, Rawstron AC, Ghia P, Hillmen P, Houlston RS, Kay N, et al. Diagnostic criteria for monoclonal B-cell lymphocytosis. *British Journal of Haematology*. 2005/08/01;130(3).

69. Matthews C, Catherwood MA, Morris TCM, Kettle PJ, Drake MB, Gilmore WS, et al. Serum TK levels in CLL identify Binet stage A patients within biologically defined prognostic subgroups most likely to undergo disease progression. *European Journal of Haematology*. 2006/10/01;77(4).

70. Matutes E, Owusu-Ankomah K, Morilla R, Garcia Marco J, Houlihan A, Que TH, et al. The immunological profile of B-cell disorders and proposal of a scoring system for the diagnosis of CLL. *Leukemia*. 1994;8(10):1640-5.

71. Matutes E, Wechsler A, Gomez R, Cherchi M, Catovsky D. Unusual T-Cell Phenotype in

- Advanced B-Chronic Lymphocytic Leukaemia. *British Journal of Haematology*. 1981/12/01;49(4).
72. Maurer C, Hallek M. Chronische lymphatische Leukämie. *DMW - Deutsche Medizinische Wochenschrift*. 2013/10;138(42).
73. McGrayne SB. *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, & Emerged Triumphant from Two Centuries of Controversy.*: Yale University Press.
74. McInnes L, Healy, J., & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv preprint arXiv:180203426*. 2020.
75. Miller GA. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*. 1956;63(2):81–97.
76. Milligan GW, Cooper MC. A study of standardization of variables in cluster analysis. *Journal of Classification* 1988 5:2. 1988/09;5(2).

77. Mills KHGC, J.C. Suppressor t cells in b-cell chronic lymphocytic leukaemia: Relationship to clinical stage. *Leukemia Research*. 1982;6(5):653–7.
78. Mittelman A, Denny T, Gebhard D, Cirrincione C, Kurland E, Koziner B. Analysis of T-cell subsets in B-cell chronic lymphocytic leukemia: A correlation with the stage of disease. *American Journal of Hematology*. 1984/01/01;16(1).
79. Molica S. Chronic lymphocytic leukemia prognostic models in real life: still a long way off. *Expert Review of Hematology*. 2021-02-01.
80. Moreau EJ, Matutes E, A'Hern RP, Morilla AM, Morilla RM, Owusu-Ankomah KA, et al. Improvement of the Chronic Lymphocytic Leukemia Scoring System With the Monoclonal Antibody SN8(CD79b). *American Journal of Clinical Pathology*. 1997/10/01;108(4).
81. Morice WG, Kurtin, P. J., Hodnefield, J. M., Shanafelt, T. D., Hoyer, J. D., Remstein, E. D., & Hanson, C. A. Predictive Value of Blood and Bone Marrow Flow Cytometry in B-Cell Lymphoma

Classification: Comparative Analysis of Flow Cytometry and Tissue Biopsy in 252 Patients. Mayo Clinic Proceedings. 2008/07/01;83(7).

82. Ng DP, Zuromski LM. Augmented Human Intelligence and Automated Diagnosis in Flow Cytometry for Hematologic Malignancies. American Journal of Clinical Pathology. 2021/03/15;155(4).

83. Nunes C, Wong R, Mason M, Fegan C, Man S, Pepper C. Expansion of a CD8+PD-1+ Replicative Senescence Phenotype in Early Stage CLL Patients Is Associated with Inverted CD4:CD8 Ratios and Disease Progression. Clinical Cancer Research. 2012/02/01;18(3).

84. Oberle V, Soßdorf M, Lösche W. Durchflusszytometrie. 2010.

85. Oken MM, Creech, R. H., Tormey, D. C., Horton, J., Davis, T. E., McFadden, E. T., & Carbone, P. P. Toxicity and response criteria of the Eastern Cooperative Oncology Group. American journal of clinical oncology,. 1982;5(6):649-56.

86. Oscier D, Else M, Matutes E, Morilla R, Strefford JC, Catovsky D. The morphology of CLL revisited: the clinical significance of prolymphocytes and correlations with prognostic/molecular markers in the LRF CLL4 trial. *British Journal of Haematology*. 2016/09/01;174(5).
87. Palma M, Gentilcore G, Heimersson K, Mozaffari F, Näsman-Glaser B, Young E, et al. T cells in chronic lymphocytic leukemia display dysregulated expression of immune checkpoints and activation markers. *Haematologica*. 2017/03;102(3).
88. Parikh SA, Kay NE, Shanafelt TD. How we treat Richter syndrome. *Blood*. 2014/03/13;123(11).
89. Pearson F.R.S. K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 1901-11-1;2(11).
90. Peller SK, S. Decreased CD45RA T Cells in B-Cell Chronic Lymphatic Leukemia Patients: Correlation With Disease Stage. *Blood*. 1991;78(6):1569–73.

91. Platsoucas CD, Galinski M, Kempin S, Reich L, Clarkson B, Good RA. Abnormal T lymphocyte subpopulations in patients with B cell chronic lymphocytic leukemia: an analysis by monoclonal antibodies. *The Journal of Immunology*. 1982/11/01;129(5).
92. Puente XS, López-Otín C, Puente XS, López-Otín C. The evolutionary biography of chronic lymphocytic leukemia. *Nature Genetics* 2013 45:3. 2013-02-24;45(3).
93. Rai K, Sawitsky A, Cronkite E, Chanana A, Levy R, Pasternack B. Clinical staging of chronic lymphocytic leukemia. *Blood*. 1975/08/01;46(2).
94. Rassenti LZ, Jain S, Keating MJ, Wierda WG, Grever MR, Byrd JC, et al. Relative value of ZAP-70, CD38, and immunoglobulin mutation status in predicting aggressive disease in chronic lymphocytic leukemia. *Blood*. 2008/09/01;112(5).
95. Rawstron AC, Bennett FL, O'Connor SJM, Kwok M, Fenton JAL, Plummer M, et al. Monoclonal B-Cell Lymphocytosis and Chronic Lymphocytic Leukemia.

<https://doi.org/10.1056/NEJMoa075290>. 2008-08-07;359(6).

96. Rawstron AC, Kreuzer K-A, Soosapilla A, Spacek M, Stehlikova O, Gambell P, et al. Reproducible diagnosis of chronic lymphocytic leukemia by flow cytometry: An European Research Initiative on CLL (ERIC) & European Society for Clinical Cell Analysis (ESCCA) Harmonisation project. *Cytometry Part B: Clinical Cytometry*. 2018/01/01;94(1).

97. Reinisch W, Willheim M, Hilgarth M, Gasché C, Mader R, Szepfalusi S, et al. Soluble CD23 reliably reflects disease activity in B-cell chronic lymphocytic leukemia. *Journal of Clinical Oncology*. 2016-09-21.

98. Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?". *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016.

99. Roessner PM, Seiffert M. T-cells in chronic lymphocytic leukemia: Guardians or drivers of disease? *Leukemia* 2020 34:8. 2020-05-26;34(8).

100. Rossi D, Rasi S, Fabbri G, Spina V, Fangazio M, Forconi F, et al. Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood*. 2012/01/12;119(2).
101. Rossi D, Spina V, Gaidano G. Biology and treatment of Richter syndrome. *Blood*. 2018/06/21;131(25).
102. Rossi D, Terzi-di-Bergamo L, De Paoli L, Cerri M, Ghilardi G, Chiarenza A, et al. Molecular prediction of durable remission after first-line fludarabine-cyclophosphamide-rituximab in chronic lymphocytic leukemia. *Blood*. 2015/10/15;126(16).
103. Rothe G. Technische und methodische Grundlagen der Durchflusszytometrie. *Zelluläre Diagnostik Grundlagen, Methoden und klinische Anwendungen der Durchflusszytometrie: Karger*; 2007. p. 27-70.
104. Ruck DW, Rogers SK, Kabrisky M, Oxley ME, Suter BW. The multilayer perceptron as an approximation to a Bayes optimal discriminant function. *IEEE Trans Neural Netw*. 1990;1(4):296-8.

105. Saeys Y, Van Gassen S, Lambrecht BN. Computational flow cytometry: helping to make sense of high-dimensional immunology data. *Nature Reviews Immunology* 2016 16:7. 2016-06-20;16(7).
106. Sarfati M, Chevret S, Chastang C, Biron G, Stryckmans P, Delespesse G, et al. Prognostic importance of serum soluble CD23 level in chronic lymphocytic leukemia. *Blood*. 1996/12/01;88(11).
107. Schroers R, Griesinger F, Trümper L, Haase D, Kulle B, Klein-Hitpass L, et al. Combined analysis of ZAP-70 and CD38 expression as a predictor of disease progression in B-cell chronic lymphocytic leukemia. *Leukemia* 2005 19:5. 2005-03-10;19(5).
108. Serrano D, Monteiro J, Allen SL, Kolitz J, Schulman P, Lichtman SM, et al. Clonal expansion within the CD4+CD57+ and CD8+CD57+ T cell subsets in chronic lymphocytic leukemia. *The Journal of Immunology*. 1997/02/01;158(3).
109. Simonyan K, Vedaldi, A., & Zisserman, A. . Deep inside convolutional networks: Visualising

image classification models and saliency maps. arXiv preprint arXiv:13126034. 2013.

110. Smuha NA. The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence. *Computer Law Review International*. 2019-08-01;20(4).

111. Solomon BM, Rabe KG, Slager SL, Brewer JD, Cerhan JR, Shanafelt TD. Overall and cancer-specific survival of patients with breast, colon, kidney, and lung cancers with and without chronic lymphocytic leukemia: a SEER population-based study. *J Clin Oncol*. 2013;31(7):930-7.

112. Soumerai JD, Ni A, Darif M, Londhe A, Xing G, Mun Y, et al. Prognostic risk score for patients with relapsed or refractory chronic lymphocytic leukaemia treated with targeted therapies or chemoimmunotherapy: a retrospective, pooled cohort study with external validations. *The Lancet Haematology*. 2019/07/01;6(7).

113. Stamatopoulos K, Agathangelidis A, Rosenquist R, Ghia P. Antigen receptor stereotypy

in chronic lymphocytic leukemia. *Leukemia* 2017 31:2. 2016-11-04;31(2).

114. Stöger K, Schneeberger D, Holzinger A. Medical artificial intelligence. *Communications of the ACM*. 2021-10-25.

115. Strati P, Shanafelt TD. Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: diagnosis, natural history, and risk stratification. *Blood*. 2015/07/23;126(4).

116. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*. 2016/05/19;127(20).

117. TheInternationalCLL-IPIworkinggroup. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *The Lancet Oncology*. 2016/06/01;17(6).

118. Thompson PA, Tam CS, O'Brien SM, Wierda WG, Stingo F, Plunkett W, et al. Fludarabine, cyclophosphamide, and rituximab treatment achieves long-term disease-free survival

in IGHV-mutated chronic lymphocytic leukemia. *Blood*. 2016/01/21;127(3).

119. Thorsélius M, Kröber A, Murray F, Thunberg U, Tobin G, Bühler A, et al. Strikingly homologous immunoglobulin gene rearrangements and poor outcome in VH3-21-using chronic lymphocytic leukemia patients independent of geographic origin and mutational status. *Blood*. 2006/04/01;107(7).

120. Thrun MC. Exploiting Distance-Based Structures in Data Using an Explainable AI for Stock Picking. *Information* 2022, Vol 13, Page 51. 2022-01-21;13(2).

121. Thrun MC. Identification of Explainable Structures in Data with a Human-in-the-Loop. *KI - Künstliche Intelligenz* 2022 36:3. 2022-11-21;36(3).

122. Thrun MC. Projection-Based Clustering through Self-Organization and Swarm Intelligence: Combining Cluster Analysis with the Visualization of High-Dimensional Data. 2018.

123. Thrun MC, Gehlert T, Ultsch A. Analyzing the fine structure of distributions. PLOS ONE. 14.10.2020;15(10).
124. Thrun MC, Ultsch A, Breuer L. Explainable AI Framework for Multivariate Hydrochemical Time Series. Machine Learning and Knowledge Extraction 2021, Vol 3, Pages 170-204. 2021-02-04;3(1).
125. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology. 1996/01/01;58(1).
126. Tötterman THC, M.; Simonsson, B.; Bengtsson, M.; Nilsson, K. T-Cell Activation and Subset Patterns Are Altered in B-CLL and Correlate With the Stage of the Disease. Blood. 1989;74(2):786–92.
127. Tresckow Jv, Eichhorst B, Bahlo J, Hallek M. The Treatment of Chronic Lymphatic Leukemia. Deutsches Ärzteblatt International. 2019/01;116(4).
128. Ultsch A. Data mining and knowledge discovery with emergent self-organizing feature

maps for multivariate time series. Kohonen maps
Elsevier science BV. 1999:33-45.

129. Ultsch A. U*-matrix: a tool to visualize
clusters in high dimensional data. 2003.

130. Ultsch A, Hoffmann J, Röhnert MA, Bonin
Mv, Oelschlägel U, Brendel C, et al. An Explainable
AI System for the Diagnosis of High-Dimensional
Biomedical Data. *BioMedInformatics 2024, Vol 4,*
Pages 197-218. 2024-01-11;4(1).

131. Ultsch A, Lötsch J. Computed ABC Analysis
for Rational Selection of Most Informative
Variables in Multivariate Data. *PLOS ONE.*
10.06.2015;10(6).

132. Van der Maaten L, Hinten G. Visualizing
data using t-SNE. *Journal of machine learning
research.* 2008;9(11).

133. Van Gassen S, Callebaut B, Van Helden MJ,
Lambrecht BN, Demeester P, Dhaene T, et al.
FlowSOM: Using self-organizing maps for
visualization and interpretation of cytometry data.
Cytometry Part A. 2015/07/01;87(7).

134. Van Gassen S, Vens C, Dhaene T, Lambrecht BN, Saeys Y. FloReMi: Flow density survival regression using minimal feature redundancy. *Cytometry Part A*. 2016/01/01;89(1).
135. Von Eschenbach WJ. Transparency and the Black Box Problem: Why We Do Not Trust AI. *Philosophy & Technology* 2021 34:4. 2021-09-01;34(4).
136. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, et al. SF3B1 and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia. <https://doi.org/10.1056/NEJMoa1109016>. 2011-12-29;365(26).
137. Wickham H. *ggplot2*. Wiley Interdisciplinary Reviews: Computational Statistics. 2011/03/01;3(2).
138. Wierda WG, O'Brien S, Wang X, Faderl S, Ferrajoli A, Do K-A, et al. Prognostic nomogram and index for overall survival in previously untreated patients with chronic lymphocytic leukemia. *Blood*. 2007/06/01;109(11).

139. Wilson EJ, Sherrell DL. Source effects in communication and persuasion research: A meta-analysis of effect size. *Journal of the Academy of Marketing Science* 1993 21:2. 1993/03;21(2).
140. Witkowska M, Nowak W, Cebula-Obrzut B, Majchrzak A, Medra A, Robak T, et al. Spontaneous in vitro apoptosis of de novo chronic lymphocytic leukemia cells correlates with risk of the disease progression. *Cytometry Part B: Clinical Cytometry*. 2014/11/01;86(6).
141. Yang G, Ye Q, Xia J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *Information Fusion*. 2022/01/01;77.
142. Zaki MJ, Zaki MJ. SPADE: An Efficient Algorithm for Mining Frequent Sequences. *Machine Learning* 2001 42:1. 2001/01;42(1).
143. Zeiler MD, Fergus R. Visualizing and Understanding Convolutional Networks. 2014.
144. Zenz T, Benner A, Döhner H, Stilgenbauer S. Chronic lymphocytic leukemia and treatment

resistance in cancer: The role of the p53 pathway. *Cell Cycle*. 2008-12-15;7(24).

145. Zenz T, Eichhorst B, Busch R, Denzel T, Habe S, Winkler D, et al. TP53 mutation and survival in chronic lymphocytic leukemia. *J Clin Oncol*. 2010;28(29):4473-9.

146. Zhang Y, Weng Y, Lund J. Applications of Explainable Artificial Intelligence in Diagnosis and Surgery. *Diagnostics* 2022, Vol 12, Page 237. 2022-01-19;12(2).

147. Zhao M, Mallesh N, Höllein A, Schabath R, Haferlach C, Haferlach T, et al. Hematologist-Level Classification of Mature B-Cell Neoplasm Using Deep Learning on Multiparameter Flow Cytometry Data. *Cytometry Part A*. 2020/10/01;97(10).

148. Zhou B, Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016:2921-9.

Anhang

I. Lebenslauf

Aus datenschutzrechtlichen Gründen wird der Lebenslauf in der elektronischen Fassung der Dissertation nicht veröffentlicht.

II. Verzeichnis der akademischen Lehrer/-innen

Meine akademischen Lehrenden an der Philipps-Universität Marburg waren:

Adamkiewicz, Aigner, Alter, Andrasi-Wensauer, Baranovski, Barth, Bartsch, Bauer, Baumann, Becker, Bender, Bette, Best, Bertoune, Beutel, Bien, Birk, Bliemel, Bogdan, Bonaterra, Bösner, Breder, Brehm, Buchholz, Burchert, Buttgereit, Cabanel, Cetin, Chatzis, Cordes, Czubayko, Damanakis, De Cruppé, Decher, Denkert, Dettmeyer, Divchev, Donner-Banzhoff, Duda, Dworschak, Ehlenz, Eggers, Eggert, El-Zayat, Eming, Eschbach, Falkenberg, Fendrich, Feuser, Frink, Fritz, Fuchs-Winkelmann, Geisthoff, Geks, Geraedts, Gebhardt, Gjorgjevskil, Görg, Gress, Greulich, Grimm, Halaszovich, Hertl, Hildebrandt, Hoch, Hofmann, Holland, Holzer, Homberg, Hoyer, Huber, Jansen, Jerrentrup, Josephs, Kann, Kalder, Kanngießer, Keber, Kemmling, Kerwat, Kill, Kinscherf, Kiper, Kirchner, Kirschbaum, Klemmer, Kluge, Knipper,

Koczulla, Koehler, Köhler, Kühnert, Lill, Lohoff,
Lüsebrink, Mahnken, Malcherczyk, Maisner,
Maurer, Meisgeier-Moosdorf, Menzler, Mey,
Miehling, Milani, Moll, Morin, Mutters, Nenadic,
Neubauer, Neumüller, Nimsky, Oberkircher,
Oberwinkler, Oertel, Oliver, Opitz, Pagenstecher,
Parahuleva, Pedrosa, Peterlein, Pfaar, Pfützner,
Plant, Pöttgen, Preisig-Müller, Rastan, Reese,
Reiman, Renke, Renz, Riera-Knorrenschild, Rost,
Ruchholtz, Rupa, Rust, Rüsç, Saß, Schachtner,
Schieffer, Schmeck, Schu, Schulze, Schulze,
Schütz, Schwarz, Seitz, Sekundo, Sevinc,
Sieveking, Sommer, Stahl, Stathopoulos, Stibane,
Steiniger, Stuck, Swaid, Tackenberg, Thangavelu,
Thieme, Thum, Timmermann, Torossian, Trebels,
Verburg, Vogelmeier, Vogt, Vojnar, Wächter,
Wack, Wagner, Weihe, Weiß, Westermann,
Wiesmann, Wilhelm, Worzfeld, Wrocklage, Wulf,
Zacharis

III. Danksagung

Ich bedanke mich aufrichtig bei Herrn Priv.-Doz. Dr. med. Jörg Hoffmann für sein Vertrauen und die Möglichkeit diese wissenschaftliche Arbeit zu verfassen sowie für die hervorragende Betreuung, auf die man sich während der gesamten Zeit bedingungslos verlassen konnte. Weiterer Dank gilt außerdem dem Team der hämatologischen, onkologischen und immunologischen Klinik für ihre Unterstützung.

Ein besonderer Dank gebührt meiner Mutter Ismeta, meinem Vater Ujkan, meinem Bruder Edis und meiner Schwester Dzeneta für ihre unermüdliche Unterstützung während meiner Studienzeit und der Dissertation.

IV. Ehrenwörtliche Erklärung

Entnahme vor Drucklegung.