

Aus dem Institut für Tierpathologie
des Fachbereichs Veterinärmedizin
der Freien Universität Berlin

**Erforschung und Validierung der digitalen Mikroskopie und
automatisierten Bildanalysen für die Diagnostik und Lehre
in der Veterinärpathologie und -zytologie**

Inaugural-Dissertation
zur Erlangung des Grades eines
PhD of Biomedical Science
an der
Freien Universität Berlin

vorgelegt von
Christof Albert Bertram
Tierarzt
aus Nideggen

Berlin, 2021

Journal-Nr.: 4297

Bibliografische Informationen der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet abrufbar über
<http://dnb.ddb.de>

© 2021 by Verlag:

Deutsche Veterinärmedizinische Gesellschaft Service GmbH, Gießen
Printed in Germany

ISBN 978-3-86345-612-2

1. Auflage 2021

Verlag:

DVG Service GmbH
An der Alten Post 2
35390 Gießen
Tel.: 0641 984446-0
info@dvg.de
www.dvg.de

Aus dem Institut für Tierpathologie
des Fachbereichs Veterinärmedizin
der Freien Universität Berlin

**Erforschung und Validierung der digitalen Mikroskopie und
automatisierten Bildanalysen für die Diagnostik und Lehre
in der Veterinärpathologie und -zytologie**

Inaugural-Dissertation
zur Erlangung des Grades eines
PhD of Biomedical Science
an der
Freien Universität Berlin

vorgelegt von
Christof Albert Bertram
Tierarzt
aus Nideggen

Berlin, 2021

Journal-Nr.: 4297

Gedruckt mit Genehmigung
des Fachbereichs Veterinärmedizin
der Freien Universität Berlin

Dekan:	Univ.-Prof. Dr. Uwe Rösler
Erster Gutachter:	Univ.-Prof. Dr. Robert Klopffleisch
Zweiter Gutachter:	Prof. Dr. Andreas Maier
Dritter Gutachter:	PD Dr. Kerstin Müller
Vierter Gutachter:	Prof. Dr. Vitaly Belik

Deskriptoren (nach CAB-Thesaurus): pathology, cytology, microscopy, digital technology, neoplasms, data analysis, evaluation

Tag der Promotion: 03.12.2021

Für Beate, Carl-Benedict, Carlotta, Heidi und Horst

Inhaltsverzeichnis

Inhaltsverzeichnis.....	V
Abbildungsverzeichnis	VII
Tabellenverzeichnis.....	IX
1 Einleitung	1
2 Literaturübersicht	2
2.1 Der Weg des Pixels	4
2.1.1 Herstellung von Gewebeproben	4
2.1.2 Whole Slide Imaging	4
2.1.3 Aufbewahrung und Bereitstellung der digitalen Bilder.....	5
2.1.4 Computerarbeitsplatz	6
2.2 Digitale Mikroskopie	7
2.2.1 Potential und Limitationen für die Diagnostik	8
2.2.2 Vor- und Nachteile für die Lehre.....	11
2.3 Automatisierte Bildanalyse	13
2.3.1 Algorithmische Herangehensweisen	13
2.3.2 Deep Learning für die Objekterkennung.....	21
3 Eigene Untersuchungen	30
3.1 Validierung der digitalen Mikroskopie für die Diagnose von kaninen Hauttumoren.	30
3.2 Meinungsumfrage über den Einsatz der digitalen Mikroskopie in der Lehre	42
3.3 Mitosedatensatz von kaninen kutanen Mastzelltumoren.....	49
3.4 Untersuchung zur Variabilität des Mitotic Count.....	59
3.5 Mitosedatensatz von kaninen Mammakarzinomen.....	73
3.6 Untersuchung zur Variabilität der Datensatzentwicklung	84
3.7 Untersuchung zur Datenbankqualität von Multi-Experten-Konsensus.....	95
3.8 Datensatz über zwei- und mehrkernige Tumorzellen in kaninen Mastzelltumoren	102
4 Diskussion.....	109
4.1 Erkenntnisse zur diagnostischen Leistungsfähigkeit der digitalen Mikroskopie.....	111
4.1.1 Methoden für die Validierung der digitalen Mikroskopie	111

4.1.2	Aktueller Stand der Validierung in der Tiermedizin	117
4.1.3	Aktueller Stand der Validierung in Humanmedizin	118
4.2	Erkenntnisse zum Meinungsbild der Studierenden über den Einsatz der digitalen Mikroskopie in der Lehre	119
4.3	Erkenntnisse über das Potential und Herausforderungen von Deep learning- basierten Algorithmen.....	121
4.3.1	Mitosefiguren in histologischen Tumorpräparaten	121
4.3.2	Zwei- und mehrkernige Tumorzellen	129
4.3.3	Zytologie der equinen bronchoalveolären Lavageflüssigkeit	131
4.3.4	Methoden zur Erstellung von Datensätzen.....	136
5	Zusammenfassung	142
6	Summary.....	144
7	Literaturverzeichnis.....	146
8	Publikationsverzeichnis	XI
8.1	Eigene Untersuchungen	XI
8.2	Weitere themenbezogene Studien.....	XII
8.3	Themenbezogene Übersichtsartikel.....	XIII
8.4	Themenbezogene Kongressbeiträge (inklusive Originalartikel- oder Abstrakt - Veröffentlichung).....	XIII
9	Danksagung	XX
10	Finanzierungsquelle	XXI
	Selbstständigkeitserklärung	XXII

Abbildungsverzeichnis

Abbildung 1: Arbeitsablauf der digitalen Pathologie.....	2
Abbildung 2: Arbeitsplatz mit Whole-Slide-Image-Scannern (Aperio ScanScope CS2, Leica) im Institut für Tierpathologie, Freie Universität Berlin.....	5
Abbildung 3: Vergleich relevanter algorithmischer Herangehensweisen für die automatisierte Bildanalyse in der digitalen Pathologie.....	14
Abbildung 4: Graphische Darstellung der Untergruppen der künstlichen Intelligenz.	15
Abbildung 5: Beispielhafter Vergleich der Leistungsfähigkeit abhängig von der Trainingsdatenquantität von daten-basierten Modellen.	16
Abbildung 6: Schematische Darstellung des Ablaufs von Training und Evaluation eines Deep learning-basierten Modells unter Verwendung der verschiedenen Datensatzanteile...	22
Abbildung 7: Schematische Darstellung der Trainingsschleife für Deep Learning.....	23
Abbildung 8: Schematische Darstellung einer möglichen Fehlerkurve für die Trainings- und Validierungsfehler über eine zunehmende Anzahl an Trainingsiterationen/epochen.	24
Abbildung 9: Beispiele für die Entscheidung der Übereinstimmung von Ground truth Annotationen und algorithmischen Prädiktionen (richtig positive Prädiktion) anhand von Mitosefiguren in histologischen Bildausschnitten	26
Abbildung 10: Beispiele von richtig positiven, falsch negativen und falsch positiven Prädiktionen von Mitosefiguren. Die Betrachtung richtig negativer Prädiktionen ist nicht sinnvoll bei der Objekterkennung (im Unterschied zur Bildklassifikation).....	27
Abbildung 11: Vergleich der Untersuchungsmodalitäten und -Zeitpunkte für eine Validierung der digitalen Mikroskopie mittels einfacher Konkordanzstudie (A, rote Kasten), Genauigkeitsstudie (B; grüne Kasten) und Nichtunterlegenheitsstudie (C, blaue Kasten)..	112
Abbildung 12: Darstellung beispielhafter Ergebnisse einer einfachen Konkordanzstudie .	113
Abbildung 13: Darstellung beispielhafter Ergebnisse einer Genauigkeitsstudie	114
Abbildung 14: Darstellung beispielhafter Ergebnisse einer Nichtunterlegenheitsstudie	115
Abbildung 15: Histologischer Bildausschnitt eines kaninen Mastzelltumors mit Deep learning-basierter Detektionen von Mitosen (grüne Kästchen).	122
Abbildung 16: Beispiel eines vielversprechenden Labelingansatzes für das Annotieren von Mitosefiguren mittels eines immunhistochemischen Goldstandards (Phospho-Histone H3).	125
Abbildung 17: Mögliches computer-assistiertes Diagnosesystem für die Bestimmung des Mitotic Counts in Whole Slide Images von kaninen Mastzelltumoren	128
Abbildung 18: Histologischer Bildausschnitt eines kaninen Mastzelltumors mit algorithmischen Detektionen von zweikernigen (rote Kästchen) und mehrkernigen (grüne	

Kästchen) Tumorzellen basieren auf dem Deep learning-basierten Algorithmus von BERTRAM et al. (2021)	130
Abbildung 19: Zytologischer Bildausschnitt einer Lungenspülprobe eines Pferdes mit algorithmischen Detektionen von Hämosiderophagen. Eisenspezialfärbung (blaues Pigment).	133
Abbildung 20: Vier zytologische Bildausschnitte von Lungenspülproben des Pferdes mit algorithmischen Detektionen der verschiedenen Zelltypen. May-Grünwald-Giemsa-Färbung.	135
Abbildung 21: Möglicher Arbeitsablauf einer multi-Experten-basierten Labelingmethode von Mitosefiguren in histologischen Bildern. Labelingmethode nach VETA et al. (2015).	138
Abbildung 22: Möglicher Arbeitsablauf einer algorithmisch-assistierten Labelingmethode zur Erstellung qualitativ hochwertiger Datensätze von Mitosefiguren in histologischen Bildern.	139

Tabellenverzeichnis

Tabelle 1: Definitionen von relevanten Begriffen der digitalen Pathologie.....	3
Tabelle 2: Gegenüberstellung der möglichen Verbesserungen und Herausforderungen der Arbeitsabläufe für die Diagnostik mit Verwendung der digitalen Pathologie	9
Tabelle 3: Vor- und Nachteile der digitalen Mikroskopie für die Ausbildung von Studierenden	12
Tabelle 4: Definitionen von relevanten Begriffen der automatisierten Bildanalyse.....	18
Tabelle 5: Relevante Metriken für die Bewertung von automatisierten Bildanalysealgorithmen für die Objekterkennung.....	27
Tabelle 6: Übersicht über die öffentlich verfügbaren Datensätze über Mitosefiguren in histologischen Gewebeschnitten von Tumoren.....	124
Tabelle 7: Übersicht über die öffentlich verfügbaren Datensätze über zwei- und mehrkernige Tumorzellen in Gewebeschnitten von Tumoren.....	130
Tabelle 8: Übersicht über den entwickelten zytologischen Datensatz mit den verschiedenen Klassen an Hämosiderophagen aus equinen bronchoalveolären Lavage-Flüssigkeiten.....	132
Tabelle 9: Übersicht über die entwickelten zytologischen Datensätze mit den verschiedenen Zelltypen aus equinen bronchoalveolären Lavageflüssigkeiten.....	134
Tabelle 10: Vergleich zwischen verschiedenen Labelingmethoden für die Erstellung von Datensätzen aus histologischen und zytologischen Bildern	141

1 Einleitung

Die mikroskopische Untersuchung von Gewebeproben ist eine seit Jahrhunderten bewährte Methode zur Bestimmung von pathologischen Veränderungen. Die Histopathologie und Zytologie stellt eine elementare Grundlage für die Therapieentscheidung von Patienten dar und ist ein unverzichtbares Messinstrument für die Erforschung von zahlreichen Prozessen in menschlichen und tierischen Geweben. Traditionell wurde für die mikroskopische Untersuchung von Gewebeproben der physische Glasobjektträger und ein Lichtmikroskop verwendet. Ende des 20. Jahrhunderts wurden die ersten Geräte entwickelt, die eine Digitalisierung ganzer mikroskopischer Gewebeproben ermöglichen. Erstmals konnten histologische Präparate an einem lokalen oder fern gelegenen Computer und Monitor untersucht werden. Aufgrund der rapiden Entwicklung der Informationstechnologie findet in den Arbeitsweisen der (Tier-)Pathologie – wie auch in anderen Aspekten unseres alltäglichen Lebens – gerade ein gewaltiger Wandel statt. Eine stetige Weiterentwicklung der digitalen Pathologie hat dazu geführt, dass in den letzten 15 Jahren eine immer größere Nachfrage nach einer routinemäßigen Anwendung für die akademische Lehre, diagnostische Labore und Forschungseinrichtungen entstanden ist. Mit der Verfügbarkeit von digitalen Bildern wurden revolutionäre Möglichkeiten entwickelt, wie Gewebeproben mithilfe von Computerprogrammen dargestellt, bearbeitet und automatisiert ausgewertet werden können. Die neuen Möglichkeiten, die sich aus der digitalen Pathologie ergeben, haben unter anderem das Potential den Lernerfolg der Studierenden zu verbessern, die diagnostische Effizienz und Arbeitsflexibilität in (tier)pathologischen Laboren zu steigern und hohe Genauigkeit und Reproduzierbarkeit von Forschungsergebnissen sicherzustellen. Innovative Entwicklungen auf dem Gebiet der Informatik, wie die künstliche Intelligenz, stellen die Pathologie vor ganz neue Herausforderungen. Um das Potential, die Leistungsfähigkeit und die Limitationen der digitalen Pathologie zu ergründen bedarf es noch einer Vielzahl von Studien.

Das Ziel der eigenen Untersuchungen war es die neuartigen Möglichkeiten und aktuellen Hindernisse der digitalen Pathologie für die Diagnostik und Lehre in der Veterinärpathologie und -zytologie zu erforschen. Die verschiedenen Projekte umfassten eine Validierung (Nichtunterlegenheit) von der digitalen Mikroskopie im Vergleich zur Lichtmikroskopie, eine Meinungsumfrage von Studierenden über die Anwendung der digitalen Mikroskopie in der Lehre und zahlreiche Untersuchungen über die Leistungsfähigkeit und Anforderungen von daten-basierten Bildanalysealgorithmen (künstliche Intelligenz) für die automatisierte beziehungsweise computer-assistierte Diagnose von mehreren histologischen und zytologischen Fragestellungen.

2 Literaturübersicht

Ende des 20. Jahrhunderts war es erstmals technisch möglich, gesamte Gewebeproben von Glasobjektträgern zu digitalisieren (FARAHANI *et al.*, 2015; STATHONIKOS *et al.*, 2020a). Diese digitalen Bilder von gesamten Gewebeproben (Whole-Slide Images, WSIs) stellt die Grundlage für die digitale Pathologie dar (BERTRAM und KLOPFLEISCH, 2017). Mit der Entwicklung von leistungsstarken WSI-Scanner (siehe Tabelle 1 für Definition), ist in den letzten Jahrzehnten ein große Nachfrage nach Implementierung dieser Informationstechnologie für verschiedene Anwendungsgebiete entstanden, wie der akademischen Lehre (MILLS *et al.*, 2007; NEEL *et al.*, 2007; SIMS *et al.*, 2007; DEE, 2009), die mikroskopische Diagnostik (MOLNAR *et al.*, 2003; HO *et al.*, 2006; STATHONIKOS *et al.*, 2013; THORSTENSON *et al.*, 2014) und Forschung (WEBSTER und DUNSTAN, 2014) sowohl in der Humanmedizin als auch in der Tiermedizin. Relevante Teilaspekte der digitalen Pathologie sind unter anderem die Digitalisierung der Gewebeproben und dessen Metainformationen, die Speicherung und Bereitstellung der WSIs, die IT-Infrastruktur, und die Anwendung (Abbildung 1). Für die folgende Arbeit werden die Anwendungen der digitalen Pathologie in die digitale Mikroskopie und die automatisierte Bildanalyse unterteilt. Während die digitale Mikroskopie die Untersuchung von WSIs durch einen Menschen (zum Beispiel diagnostische Pathologen/innen oder Studierenden) beschreibt, bezieht sich die automatisierte Bildanalyse auf die Extraktion von Informationen aus WSIs durch eine Computersoftware.

In den nächsten Kapiteln wird kurz auf die relevanten Aspekte der verschiedenen Arbeitsabläufe (Kapitel 2.1) und nachfolgend auf die Anwendungen – die digitale Mikroskopie (Kapitel 2.2) und die automatisierte Bildanalyse (Kapitel 2.3) – der digitalen Pathologie eingegangen.

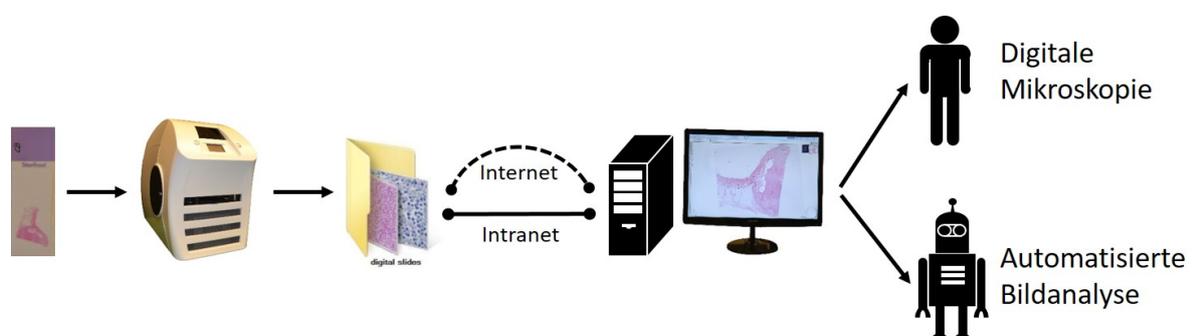


Abbildung 1: Arbeitsablauf der digitalen Pathologie.

Abbildung modifiziert von BERTRAM und KLOPFLEISCH (2017)

Tabelle 1: Definitionen von relevanten Begriffen der digitalen Pathologie

Begriff	Definition
Automatisierte Bildanalyse:	Beschreibt die Extraktion von Informationen aus digitalen Bildern mithilfe einer Computersoftware. Definitionen für relevante Begriffe der automatisierten Bildanalysen finden sich in Tabelle 4.
Digitale Mikroskopie:	Beschreibt die „dynamische“ Begutachtung von WSIs an einem computerisierten Arbeitsplatz. Vergleichbar zur Lichtmikroskopie können die Gewebeproben in verschiedenen Vergrößerungen untersucht und das Betrachtungsfeld bewegt werden. Zumeist hat das WSI nur eine Fokusebene und das Navigieren ist auf die x- und y-Achse (horizontal) beschränkt. Z-Stacking (Digitalisierung mit mehreren Fokusebene) ermöglicht hingegen auch das vertikale Navigieren auf dem WSI entlang der z-Achse.
Digitale Pathologie:	Ist ein Sammelbegriff für das Teilgebiet der Pathologie an denen WSIs beteiligt sind. Die digitale Pathologie umfasst unter anderem die Digitalisierung der Gewebeproben und dessen Meta-Informationen, die Speicherung, Bearbeitung und Bereitstellung der WSIs, die IT-Infrastruktur, die digitale Mikroskopie und die automatisierte Bildanalyse.
Whole-Slide Images (WSIs):	Ein digitales Replikat („digitaler Slide“) eines gesamten Glasobjektträger (bzw. dessen Anteile mit Gewebe) mit hoher Bildauflösung, sodass mikroskopische Strukturen, wie Zellen, gut erkennbar sind. Abhängig von der Scan-Vergrößerung und -Fläche ist ein WSI bis zu mehreren Gigabyte groß.
Whole Slide Imaging:	Beschreibt den Herstellungsprozess von WSIs mittels WSI-Scannern. Zumeist entstehen WSIs mit einer x-y-Bildachse; seltener werden pro Gewebeschnitt mehrere Fokusebenen (z-Bildachsen) erstellt (z-stacking).
Whole-Slide-Image-Scanner:	Hardware und assoziierte Software, welche für die Digitalisierung der Glasobjektträger zu WSIs notwendig sind. Die Hardware besteht aus einer Lichtquelle, Robotik, mindestens einem Objektiv und einer Kamera.

Definitionen in Anlehnung an BERTRAM und KLOPFLEISCH (2017) und DIGITAL PATHOLOGY ASSOCIATION (2021)

2.1 Der Weg des Pixels

Der "Weg des Pixels" (pixel pathway) beschreibt alle Schritte vom Glasobjektträger bis zur Untersuchung der WSIs (ZARELLA *et al.*, 2019). In den nachfolgenden Unterkapiteln werden relevante Aspekte für die Herstellung von Gewebeproben (Kapitel 2.1.1), des Whole-Slide Imaging (Kapitel 2.1.2), der Aufbewahrung und Bereitstellung von WSIs (Kapitel 2.1.3) und des Computerarbeitsplatzes (Kapitel 2.1.4) zusammengefasst.

2.1.1 Herstellung von Gewebeproben

Für die Erzeugung von WSIs sind weiterhin Glasobjektträger notwendig. Obwohl sich die Gewebepreparierung und Objektträger-Herstellung für die digitale Pathologie grundlegend nicht von den herkömmlichen Arbeitsschritten für die Lichtmikroskopie unterscheidet, ist es gegebenenfalls notwendig diesen Arbeitsschritt zu optimieren um eine ideale Qualität der WSIs, eine uneingeschränkte Funktionsfähigkeit der WSI-Scanner und einen schnellen Arbeitsablauf zu gewährleisten (STATHONIKOS *et al.*, 2013; BERTRAM und KLOPFLEISCH, 2017; RETAMERO *et al.*, 2020; STATHONIKOS *et al.*, 2020a). Beispielsweise kann eine besonders dünne Gewebedicke die erfolgreiche automatisierte Auswahl der Fokuspunkte des Scanners verbessern (STATHONIKOS *et al.*, 2013). Weiterhin sollte das Gewebe möglichst mittig auf den Objektträger gelegt werden sowie überschüssiges Eindeckmedium am Rand des Deckglases und Luftblasen unter dem Deckglas vermieden werden (STATHONIKOS *et al.*, 2013; BERTRAM und KLOPFLEISCH, 2017; RETAMERO *et al.*, 2020). Um ein Verkleben der Apparaturen des WSI-Scanners zu verhindern, sollte das Eindeckmedium getrocknet sein. Eine Beschleunigung des Trocknungsschrittes (Luftstromtrocknung) und ein Wechsel des Eindeckmediums können eine Verzögerung des Digitalisierungsschrittes verringern (STATHONIKOS *et al.*, 2019). Die Standardisierung der Präparatherstellung (z.B. Verringerung der Gewebeartefakte, dünne Schnittdicke, und gleichmäßige Färbung) kann sich ebenfalls positiv auf automatisierte Bildanalysen auswirken (AEFFNER *et al.*, 2018).

2.1.2 Whole Slide Imaging

WSI-Scanner bestehen aus der Hardware und Software, die für die Digitalisierung der Glasobjektträger zuständig sind (BERTRAM und KLOPFLEISCH, 2017). Die Hardware besteht in der Regel aus einer Hellfeldlichtquelle, einem Objektträgerhalter (mit einer Ladekapazität von einem bis mehreren hundert Objektträgern), mindestens einer Objektivlinse (je nach gewünschter maximaler Vergrößerung) und einer Charge-coupled Device (CCD) Kamera (BERTRAM und KLOPFLEISCH, 2017). Vor dem Scannen werden die Gewebefläche

und die Fokusebenen mittels einer Software automatisiert erkannt. Beim Scanprozess nimmt die Kamera zahlreiche Bilder auf, während der Objektträgerhalter gleichmäßig unter der Kamera bewegt wird. Es ist möglich, das Gewebe in mehreren aufeinanderfolgenden Fokusebenen zu digitalisieren (z-stacking) (ZARELLA *et al.*, 2019). Abschließend wird das WSI aus den einzelnen Bildern zusammengesetzt (Stitching) und nachbearbeitet (zum Beispiel Bildkomprimierung). Die Zeitdauer des Scanprozesses ist unter anderem vom WSI-Scanner, der Größe der Gewebefläche, der Scanvergrößerung und der Anzahl der Fokusebenen abhängig (FARAHANI *et al.*, 2015; BERTRAM und KLOPFLEISCH, 2017). Das Datenformat des WSI ist oft spezifisch für die Firma des WSI-Scanners und erlaubt nicht zwangsweise eine Interoperabilität zwischen verschiedenen Geräten und Software unterschiedlicher Hersteller (EVANS *et al.*, 2017). FARAHANI *et al.* (2015) geben einen ausführlichen Überblick über verschiedene kommerzielle WSI-Scanner.



Abbildung 2: Arbeitsplatz mit Whole-Slide-Image-Scannern (Aperio ScanScope CS2, Leica) im Institut für Tierpathologie, Freie Universität Berlin

2.1.3 Aufbewahrung und Bereitstellung der digitalen Bilder

Die kurzfristige, langfristige oder dauerhafte Speicherung von WSIs auf einem digitalen Slide Archiv (lokaler Server oder Cloud) ermöglicht einen einfachen und schnellen Zugriff auf die Proben von Arbeitsplätzen innerhalb oder außerhalb des pathologischen Labors via Intra- oder

Internet Verbindung (BERTRAM und KLOPFLEISCH, 2017; ZARELLA *et al.*, 2019; STATHONIKOS *et al.*, 2020a). Für eine verzögerungsfreie Betrachtung von WSIs ist eine konstante Netzwerkverbindung mit ausreichend hoher Bandbreite (mindestens 20-100 Mbps) unerlässlich (HUISMAN *et al.*, 2010; VODOVNIK, 2015). Ferner können die WSIs auch auf portablen Speicherplätzen (USB-Stick, externe Festplatte) aufbewahrt oder versendet werden.

Die Herausforderungen für ein großes dauerhaftes WSI-Archiv inkludieren die Datengröße, -Backup, -Schutz, -Management und -Identifizierung sowie ein schneller Zugriff auf langfristige Speicherplätze (HUISMAN *et al.*, 2010). Aufgrund der hohen Datengröße von WSIs ist die langfristige Aufbewahrung (Archivierung) einer großen Anzahl an WSIs sehr kostenintensiv (HUISMAN *et al.*, 2010). Eine hohe Bildauflösung/Scanvergrößerung sowie eine lange Dauer der Archivierung muss gegen die hohen Kosten für die Aufbewahrung abgewogen werden.

2.1.4 Computerarbeitsplatz

Der Computerarbeitsplatz besteht in der Regel aus einem Computer oder Laptop mit entsprechender Software, mindestens einem Bildschirm und einem Eingabegerät. Der ideal ausgestattete Arbeitsplatz hängt entscheidend von den persönlichen Präferenzen des/der Pathologen/in ab (BERTRAM und KLOPFLEISCH, 2017). Unter anderem wird der Wahl des „richtigen“ Bildschirms eine große Bedeutung zugesprochen (AEFFNER *et al.*, 2018), jedoch liegen aktuell nur wenige Studien vor, die den Einfluss von Bildschirmeigenschaften (Auflösung, Größe und Format, Aktualisierungsrate, Reaktionszeit usw.) und -Einstellungen (zum Beispiel Farbkalibrierung) auf die Leistungsfähigkeit und Effizienz der digitalen Mikroskopie untersucht haben (CLARKE und TREANOR, 2017; ABEL *et al.*, 2020). Die Größe des Betrachtungssichtfeldes („high power field“) des WSI hängt größtenteils von der Größe und Auflösung des Monitors ab (KIM *et al.*, 2020). Ein großer Bildschirm mit hoher Auflösung ist empfehlenswert für ein effizientes Arbeiten (RANDELL *et al.*, 2015; BERTRAM und KLOPFLEISCH, 2017). Als Eingabegerät kommen eine Vielzahl von unterschiedlichen Geräten (z.B. konventionelle Computermaus, Trackball, 6 degree-of-freedom Navigatoren, Touchscreens usw.) in Frage (BERTRAM und KLOPFLEISCH, 2017).

Für die Betrachtung der WSIs ist eine spezielle Untersuchungssoftware (kompatibel mit dem jeweiligen WSI-Datenformat) notwendig, welche in der Regel von den entsprechenden WSI-Scanner-Firmen bereitgestellt werden (BERTRAM und KLOPFLEISCH, 2017). Neben dem einfachen Betrachten der WSIs sind fast immer weitere Tools, wie das Annotieren der Bilder, verfügbar (siehe Kapitel 2.2). Für die Generierung von Datensätzen und automatisiertem Analysieren der WSIs sind spezialisierte Annotations- und Analysesoftware notwendig (siehe Kapitel 2.3).

2.2 Digitale Mikroskopie

Die digitale Mikroskopie ermöglicht das Untersuchen einer mikroskopischen Gewebeprobe an einem computerisierten Arbeitsplatz analog zur traditionellen Lichtmikroskopie (BERTRAM und KLOPFLEISCH, 2017). Bilddateien mit einer Größe von mehreren Gigabyte Speichervolumen sind die Folge von einer extrem hohen Bildauflösung, welche erforderlich ist, um wenige Mikrometer kleine Strukturen deutlich darstellen zu können. Im Unterschied zu statischen Fotografien von mikroskopischen Bildausschnitten, ist es bei der Betrachtung der WSIs möglich, das Sichtfeld zu bewegen und stufenlos zu vergrößern und zu verkleinern. Über die Möglichkeiten der traditionellen Lichtmikroskopie hinausgehend erlaubt eine spezialisierte Untersuchungssoftware das Darstellen von Miniaturansichten zur besseren Orientierung, das Rotieren der WSIs, das schnelle Erstellen von Schnappschüssen, genaue Größenmaßstäbe, Messwerkzeuge, vielfältige Annotationen auf mikroskopischem Level und vieles mehr (BERTRAM und KLOPFLEISCH, 2017; EVANS *et al.*, 2017; FRAGGETTA *et al.*, 2018; ECCHER und GIROLAMI, 2020). Eine der größten Vorteile ist die Unabhängigkeit von einem physischen Glasobjektträger, sodass ein schneller und zumeist problemloser Zugriff auf die digitalisierte Gewebeproben (gespeichert in einem digitalen Archiv) von jedem Ort möglich ist (BERTRAM und KLOPFLEISCH, 2017; ECCHER und GIROLAMI, 2020). Diese und viele weitere Vorteile (siehe unten) haben den Einsatz der digitalen Mikroskopie für viele Felder der Veterinärpathologie lukrativ gemacht. Jedoch weist die digitale Mikroskopie auch potentielle Limitationen auf, welche eine sorgfältige Überprüfung für das jeweilige Einsatzgebiet notwendig macht. In den nachfolgenden Unterkapiteln wird speziell auf den Einsatz der digitalen Mikroskopie in der histopathologischen Diagnostik und Lehre eingegangen.

Unabhängig vom Anwendungsgebiet der digitalen Mikroskopie wird die digitale Zytologie – im Vergleich zur digitalen Histo(patho)logie – deutlich kontroverser betrachtet (BERTRAM und KLOPFLEISCH, 2017; STATHONIKOS *et al.*, 2019; ECCHER und GIROLAMI, 2020). Dies liegt daran, dass für die zytologische Untersuchung häufig sehr feine zelluläre Details entscheidend sind. Obwohl sehr hohe Scanvergrößerungen (bis zur 1000er Vergrößerung) und das Scannen von mehreren Fokusebenen (z-stacking, für das „Feinfokussieren“ der WSIs) technisch möglich sind, wird dies nur selten routinemäßig angewendet aufgrund der sehr langen Scanzeit und der sehr hohen Datengröße der WSIs. Weiterhin ist die Herstellung von zytologischen Präparaten nur schwer standardisierbar. Insbesondere Gewebe, welche mittels Feinnadelaspiration oder dem Abklatschverfahren gewonnen werden, führen häufig zu einer sehr variablen Gewebedicke und Zelldichte. Dies kann zu Problemen bei der Bildschärfe führen. Im Gegensatz dazu eignen sich Zytospinpräparate von Körperhöhlenflüssigkeiten (zum Beispiel der bronchoalveolären Lavage-Flüssigkeit) nach eigener Erfahrung recht gut für die digitale Mikroskopie.

2.2.1 Potential und Limitationen für die Diagnostik

Moderne WSI-Scanner haben einen sehr hohen Probendurchsatz und ermöglichen somit das routinemäßige Digitalisieren von großen Probenmengen (BERTRAM und KLOPFLEISCH, 2017). Dies ist die Grundvoraussetzung für die Umstellung des Arbeitsablaufes von der traditionellen Lichtmikroskopie zur digitalen Pathologie von großen diagnostischen Laboren. Aus der Humanmedizin liegen mehrere Erfahrungsberichte über eine erfolgreiche Integration der digitalen Pathologie in den Routinearbeitsablauf vor (STATHONIKOS *et al.*, 2013; THORSTENSON *et al.*, 2014; CHENG *et al.*, 2016; EVANS *et al.*, 2017; STATHONIKOS *et al.*, 2019; RETAMERO *et al.*, 2020; STATHONIKOS *et al.*, 2020a). Auch die Anzahl der tiermedizinischen Labore, welche auf die digitale Mikroskopie umgestiegen sind, nimmt zu (LUONG, 2020). Eine der großen Vorteile des Whole-Slide Imaging ist die sehr leistungsstarke Telepathologie mit sofortigem Zugriff auf WSIs von jedem Ort mit einer ausreichenden Internetverbindung (BERTRAM und KLOPFLEISCH, 2017; LUONG, 2020). Andere verfügbare Modalitäten für eine Telepathologie, wie die statische Photographien oder die Region-of-Interest digitale Mikroskopie, sind in der Regel deutlich zeitaufwendiger und fehleranfälliger aufgrund der manuellen Selektion von bestimmten Bildausschnitten (MAIOLINO *et al.*, 2006; BLANCHET *et al.*, 2019; BROOKER *et al.*, 2019). Insbesondere in Situationen mit einem hohen Bedarf an flexiblen Arbeitsmöglichkeiten – wie die COVID-19-Pandemie – kann die Telepathologie eine elementare Hilfestellung sein, um weiterhin ein großes Arbeitsaufkommen ohne Verzögerung zu bewältigen (HANNA *et al.*, 2020; STATHONIKOS *et al.*, 2020b). Weiterhin konnte eine erhebliche Zeiteinsparung im Arbeitsablauf mittels digitaler Pathologie eines großen diagnostischen Labors nachgewiesen werden (BAIDOSHVILI *et al.*, 2018).

Neben den oben erwähnten Vorteilen der digitalen Mikroskopie für diagnostische Labore gibt es zahlreiche Ungewissheiten und Herausforderungen, die durch systematische Forschungsprojekte und Optimierungen der Arbeitsabläufe geklärt werden müssen. Vor allem birgt der geringe Kenntnisstand zur Leistungsfähigkeit der digitalen Mikroskopie für die Tiermedizin ein Risiko für den Patienten. Validierungsstudien über die verschiedenen pathologischen Fragestellungen sind unbedingt notwendig (siehe Unterkapitel 2.2.1.1). Fehlerquellen in der Diagnose, die sich aus den technischen Gegebenheiten des digitalen Pathologiesystems ergeben (wie zum Beispiel eine zu geringe Bildauflösung), können vermutlich mit zukünftigen Entwicklungen behoben werden. Weiterhin erscheint es essentiell, dass Pathologen/innen zukünftig für die digitale Pathologie ausgebildet werden, um mit der neuen Technologie umzugehen und auf diesem Weg die Vorteile sowie derzeitigen Limitationen kennenzulernen. Die automatisierte Bildanalyse ist eine bisher nur wenig

erforschte, jedoch sehr vielversprechende Methode für die zukünftige Verbesserung der diagnostischen Genauigkeit und Reproduzierbarkeit und wird in Kapitel 2.3 besprochen.

Tabelle 2: Gegenüberstellung der möglichen Verbesserungen und Herausforderungen der Arbeitsabläufe für die Diagnostik mit Verwendung der digitalen Pathologie

Verbesserungen des Arbeitsablaufes	Herausforderungen des Arbeitsablaufes
Telepathologie für die primäre Diagnose oder Zweitmeinung	Nur wenige Validierungsstudien aus tiermedizinischen Laboren verfügbar
Räumliche Flexibilität des computerisierten Arbeitsplatzes	Technische Störungen / Defekte des WSI-Scanners, des Archives, des Netzwerkes, des Computerarbeitsplatzes
Schneller Zugriff auf aktuelle und alte Fälle (digitales Archiv)	Ausreichende Netzwerkverbindung und Schnelle IT-Unterstützung notwendig
Annotieren und Ausmessen von Strukturen im WSI	Einschränkungen der WSI-Qualität (Fokusebenen, Auflösung, Kompression); besonders relevant für Zytologie
Paralleles Betrachten von mehreren WSIs (Vergleich von zwei Fällen oder zwei unterschiedlichen Färbungen)	Scanartefakte
Schnelle Orientierung auf dem Gewebeschnitt aufgrund geringer Vergrößerung (<< 4x) und Miniaturansicht	Möglicherweise Optimierung der Herstellung der Gewebeprobe notwendig
Automatisierte Bildanalysen zur Unterstützung des/der Pathologen/in (computer-assistierte Diagnose)	Entwicklung, Evaluation und Anwendung von automatisierten Bildanalysesoftware notwendig
Möglicherweise verbesserte Ergonomie	Teilweise niedrige Akzeptanz der Pathologen/innen; Training der Pathologen/innen sinnvoll
Möglicherweise Zeiteinsparung im gesamten Arbeitsablauf	Große Speicherkapazitäten des digitales Archiv notwendig
Vereinfachtes Management der Arbeitsschritte in großen Laboren	

Tabelle modifiziert von BERTRAM und KLOPFLEISCH (2017)

2.2.1.1 Notwendigkeit der Validierung der digitalen Mikroskopie

Mit modernen WSI-Scannern ist die Qualität von WSIs enorm hoch und ermöglicht einen vielfältigen Einsatz der digitalen Mikroskopie für die Diagnostik (BERTRAM und KLOPFLEISCH, 2017). Jedoch sind WSIs „lediglich“ Kopien von Glasobjektträgern. In den verschiedenen Schritten des digitalen Arbeitsablaufes liegen mögliche Fehlerquellen, die einen negativen Einfluss auf die histologische beziehungsweise zytologische Diagnose und schlussendlich auf die Therapie des Patienten nehmen kann. Dieser Aspekt verursacht nicht nur Skepsis bei Pathologen/innen, sondern verlangt eine genaue Überprüfung des digitalen Arbeitsablaufes. WSIs und dessen Darstellung am Bildschirm können eine Vielzahl von Fehlern und Artefakten mit möglicherweise klinischer Relevanz verursachen (BERTRAM und KLOPFLEISCH, 2017; FRAGGETTA *et al.*, 2018). Es ist erforderlich mittels einer Validierungsstudie sicherzustellen, dass der digitale Arbeitsablauf nicht zur Verschlechterung der diagnostischen Fähigkeiten der/des Pathologin/en führt (PANTANOWITZ *et al.*, 2013). Validierungsstudien vergleichen die diagnostische Leistungsfähigkeit der/des Pathologin/en mittels digitaler Mikroskopie (Testmodalität) zur Lichtmikroskopie als Goldstandard. Lichtmikroskopie ist traditionell die „ideale“ Untersuchungsmodalität der/des Pathologin/en für die mikroskopische Diagnose von Gewebeveränderungen. Zum Startzeitpunkt dieses PhD-Projekts (in 2016) war für die Tiermedizin noch keine Validierungsstudie veröffentlicht.

Für die Sicherstellung einer hohen diagnostischen Qualität der digitalen Mikroskopie sind Validierungsstudien sowohl aus dem akademischen als auch dem klinischen Kontext erforderlich (HANNA *et al.*, 2015). Akademische Validierungsstudien (veröffentlicht in Fachzeitschriften) untersuchen die generelle Anwendbarkeit der digitalen Mikroskopie für eine zumeist spezifische pathologische Fragestellung. Neben den Vorteilen der digitalen Mikroskopie sollten hier vor allem auch die möglichen Limitationen und Mindestanforderungen (zum Beispiel Bildauflösung) für den Einsatz der digitalen Mikroskopie bestimmt werden. Klinische Validierungsstudien untersuchen die korrekte Implementierung der digitalen Mikroskopie (basierend auf den Erkenntnissen der akademischen Validierungsstudien) in einem spezifischen Labor. In der Humanpathologie wird empfohlen, dass jedes Labor den individuellen digitalen Mikroskopie-Arbeitsablauf anhand von Richtlinien validiert (PANTANOWITZ *et al.*, 2013). Hierzu hat das College of American Pathologists in 2013 erste wissenschaftlich basierte Empfehlungen formuliert (PANTANOWITZ *et al.*, 2013). Relevante Aspekte (zum Beispiel technische Mindestanforderungen) für die Implementierung der digitalen Mikroskopie (insbesondere für die Telepathologie) wurden in verschiedenen Leitfäden beschrieben (BERNARD *et al.*, 2014; PANTANOWITZ *et al.*, 2014; GARCÍA-ROJO, 2016; HUFNAGL *et al.*, 2018; CHONG *et al.*, 2020). Für die Tierpathologie stehen zurzeit keine eigenen Richtlinien für Validierungsstudien zur Verfügung.

2.2.2 Vor- und Nachteile für die Lehre

Die digitale Mikroskopie wird zunehmend in der studentischen Ausbildung der Histologie, Histopathologie und Zytologie eingebunden aufgrund zahlreicher Vorteile für Studierende und Lehrinrichtungen (SIMS *et al.*, 2007; DEE, 2009; BERTRAM und KLOPFLEISCH, 2017). Eine der größten Vorteile ist die Unabhängigkeit der Übungsmöglichkeiten von den offiziellen Lehrveranstaltungen mit Zugriff auf digitale Übungspräparate zu jedem Zeitpunkt und von jedem Ort mit Internetverbindung aus (Tabelle 3). Studierende nehmen diese Möglichkeit insbesondere in den Nachmittags- und Abendstunden sowie kurz vor den Prüfungen wahr (PLENDL *et al.*, 2009; PANTANOWITZ *et al.*, 2012; ORDI *et al.*, 2015; SACO *et al.*, 2016a). Die digitale Mikroskopie kann auch problemlos in den Distanzunterricht (als Ergänzung oder Ersatz von Präsenzveranstaltungen) eingebunden werden. Außerdem wird die Interaktion zwischen Studierenden und ihren Dozenten vereinfacht, da die Präparate komplikationslos gleichzeitig betrachtet werden und relevante oder komplizierte Strukturen im Bild markiert werden können. Vorherige Umfragen von Studierenden aus der Human- und Tiermedizin haben durchweg ein positives Meinungsbild zur digitalen Mikroskopie ergeben (MILLS *et al.*, 2007; NEEL *et al.*, 2007; SIMS *et al.*, 2007; BRAUN und KEARNS, 2008; PLENDL *et al.*, 2009; PANTANOWITZ *et al.*, 2012; KOGAN *et al.*, 2014; ORDI *et al.*, 2015; EVANS *et al.*, 2020; RODRIGUES-FERNANDES *et al.*, 2020). Jedoch wird von vielen Studierenden eine Kombination von Licht- und digitaler Mikroskopie bevorzugt (NEEL *et al.*, 2007; BRAUN und KEARNS, 2008; KOGAN *et al.*, 2014). Einige Studien konnten eine gleichbleibende oder verbesserte Lerneffizienz – gemessen an der Prüfungsleistung – für die Histopathologie und Zytologie nachweisen (BRAUN und KEARNS, 2008; PANTANOWITZ *et al.*, 2012; ORDI *et al.*, 2015; BROWN *et al.*, 2016; EVANS *et al.*, 2020; RODRIGUES-FERNANDES *et al.*, 2020). Eine Prüfung mittels digitaler Mikroskopie mit vergleichbaren Ergebnissen erscheint ebenfalls machbar, wenn vorherige Übungsmöglichkeiten mit der digitalen Mikroskopie angeboten werden (NEEL *et al.*, 2007; BRICK *et al.*, 2013; RODRIGUES-FERNANDES *et al.*, 2020).

Limitationen für die Einbindung der digitalen Mikroskopie in die Lehre sind die initial hohen Anschaffungskosten für WSI-Scanner und IT-Infrastruktur (SACO *et al.*, 2016a). Jedoch sind langfristig auch deutliche Einsparungen möglich, da Lichtmikroskope und qualitativ hochwertige histologische Präparate möglicherweise nicht mehr für große Gruppen bereitgestellt werden müssen (DEE, 2009). Ein möglicher Kritikpunkt in der Studierenden-Ausbildung ist die reduzierte Erfahrung im Umgang mit Lichtmikroskopen, insbesondere für die Untersuchung von zytologischen Präparaten, welche als Kernkompetenz im Beruf von praktischen Tierärzten angesehen wird (STEWART *et al.*, 2014).

Tabelle 3: Vor- und Nachteile der digitalen Mikroskopie für die Ausbildung von Studierenden

Für Studierende	
Vorteile	Nachteile
Ständiger Zugang zu WSIs für das selbstorganisierte Lernen	Technische Ausstattung (z.B. Laptop) und stabile Internetverbindung notwendig
Zumeist einfache Bedienung (stufenlose Vergrößerung, kein Einstellen des Mikroskops notwendig)	Verringerte Aneignung von Grundfähigkeiten der Lichtmikroskopie
Untersuchen eines einheitlichen Gewebeschnittes (WSI) mit sehr hoher und konstanter Qualität	Feinfokus und sehr hohe Auflösungen (1000er Vergrößerung) oft nicht verfügbar (besonders relevant für die Zytologie)
Verbesserte Übersicht aufgrund geringer Vergrößerung (<< 4x) und Miniaturansicht	
Gleichzeitiges Betrachten des gleichen Präparates (interaktives Lernen)	
Markierungen im Bild, um auf relevante Strukturen hinzuweisen oder unbekannte Strukturen zu erfragen	
Einbindung von Quiz und Lernkontrollen z.B. durch Annotationen im Bild	
Für Dozenten und Lehreinrichtungen	
Vorteile	Nachteile
Einfaches Demonstrieren von mikroskopischen Strukturen (Studierenden-Dozenten-Interaktion)	Mögliche technische Störungen während der Vorlesung bzw. praktischen Übung
Anschaffung und Wartung von zahlreichen Lichtmikroskopen sowie Vervielfachung und Erneuerung von histologischen Präparaten entfällt möglicherweise	Zugang zu einem WSI-Scanner, Archivmöglichkeiten für Präparate, online-basierte Mikroskopiesoftware und IT-Betreuung notwendig
Einfache Integration von histologischen Bildern in Vorlesungen (z.B. durch Internetlink zum WSIs oder Screenshots)	
Vereinheitlichung und vereinfachte Organisation der Leistungsüberprüfung von Studierenden möglich	

Tabelle modifiziert von BERTRAM und KLOPFLEISCH (2017) und DEE (2009)

2.3 Automatisierte Bildanalyse

Automatisierte Bildanalysen haben ein großes Potential die Reproduzierbarkeit, Genauigkeit und Effizienz von pathologischen Diagnosen zu verbessern. Für diagnostische Labore stellen automatisierte Bildanalysen daher einen sehr vielversprechenden Lösungsansatz für zahlreiche mikroskopische Fragestellungen dar (LUONG, 2020; STATHONIKOS *et al.*, 2020a). Allerdings sind noch wesentliche Entwicklungen und Überprüfungen der algorithmischen Ansätze notwendig, um eine sichere Anwendung für die Routinediagnose zu gewährleisten (STATHONIKOS *et al.*, 2020a). Um diese Entwicklung in der digitalen Pathologie voranzutreiben, müssen sich Pathologen/innen aktiv an jedem Schritt der Entwicklung von Bildanalyse-Systemen beteiligen und ein grundlegendes Verständnis von der Entwicklung, Überprüfung und Anwendung von Algorithmen aufbringen (AEFFNER *et al.*, 2016; MCALPINE *et al.*, 2020). Die erfolgreichsten Bildanalyse-Systeme werden von einem multidisziplinären Team mit umfangreichen Kompetenzen zur pathologische Fragestellung, Mustererkennung, Biostatistik und vielem mehr erzielt (RASHIDI *et al.*, 2019; MCALPINE und MICHELOW, 2020). In diesem Kapitel werden die relevanten algorithmischen Ansätze (Algorithmen mit manueller Parameteranpassung und daten-basierte Ansätze) verglichen und detaillierter auf das Deep Learning im voll überwachten Fall (fully supervised) eingegangen.

2.3.1 Algorithmische Herangehensweisen

Aktuell spielen drei algorithmische Herangehensweisen eine große Bedeutung für die digitale Pathologie (AEFFNER *et al.*, 2018; ABELS *et al.*, 2019):

- Algorithmen mit manueller Parameteranpassung (insbesondere thresholding-basiert)
- Daten-basierte Methoden (künstliche Intelligenz; zumeist überwachtes Lernen)
 - (Traditionelles) maschinelles Lernen
 - Deep Learning

Bei der Auswahl der optimalen Herangehensweise spielen unter anderem die Art und Komplexität der Fragestellung, die verfügbare Rechenleistung, die verfügbare Datensatzqualität und -quantität sowie die gewünschte algorithmische Leistungsfähigkeit, Flexibilität und Nachvollziehbarkeit eine Rolle (RASHIDI *et al.*, 2019; MCALPINE *et al.*, 2020; ZURAW *et al.*, 2020). Algorithmen zur manuellen Anpassung bestehen aus einer programmierten Software mit einer Kombination von vergleichsweise einfachen Bildverarbeitungsverfahren (auch als Filter bezeichnet). Bei der Anwendung dieser Software ist es möglich, bestimmte Parameter auf eine spezifische Anwendung anzupassen, indem Schwellwerte durch visuelle Kontrolle der Prädiktionsergebnisse von einem Anwender optimiert werden (Thresholding). Im Gegensatz zu den anderen Herangehensweisen, wird bei

diesem Ansatz der Algorithmus nicht durch einen Optimierungsprozess unter Zuhilfenahme von Beispielen generiert (Abbildung 3). Eine sehr häufige Anwendung der Algorithmen zur manuellen Anpassung ist die Quantifizierung von immunhistochemischen Färbungen oder anderen Spezialfärbungen basierend auf Farbwerten von Pixeln, die einer Fläche oder einzelnen Zelle zugeordnet werden (Segmentierung) (CENTER *et al.*, 2013; AEFNER *et al.*, 2018). Allerdings ist mittels dieser Methoden keine genaue Prädiktion von komplexen strukturellen Mustern, wie Mitosefiguren, möglich.

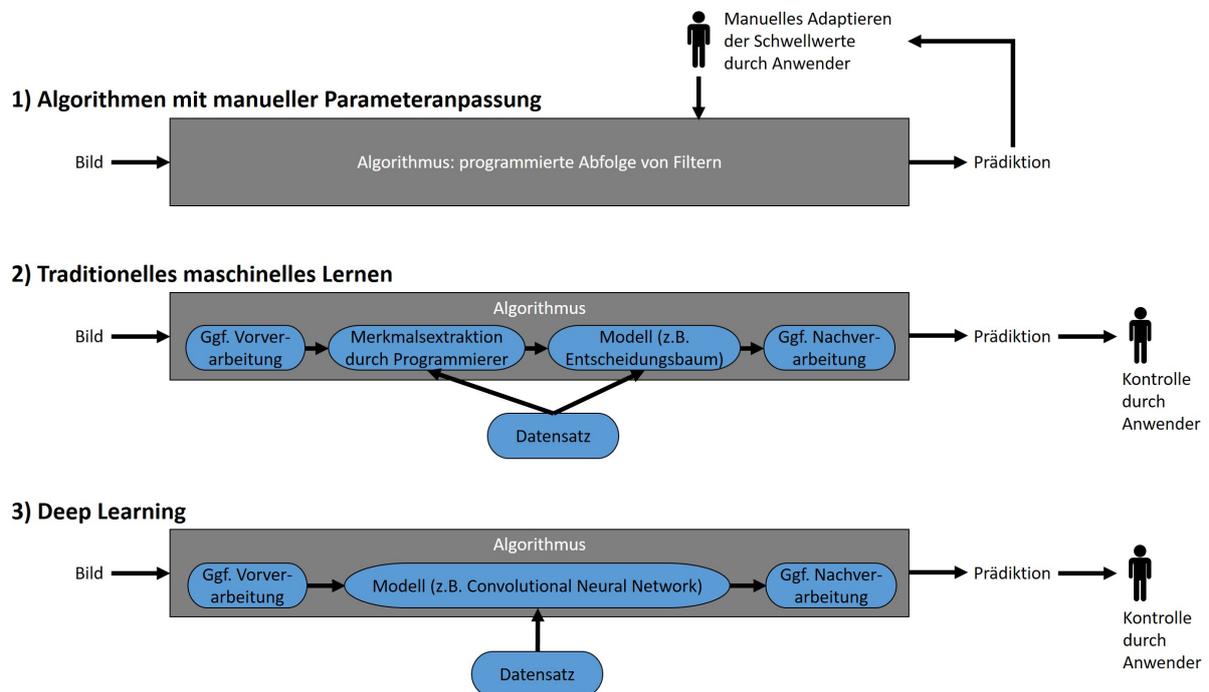


Abbildung 3: Vergleich relevanter algorithmischer Herangehensweisen für die automatisierte Bildanalyse in der digitalen Pathologie

Daten-basierte Herangehensweisen nutzen Methoden der künstlichen Intelligenz (siehe Tabelle 4 für Definition), welche das maschinelle Lernen und insbesondere das Deep Learning einschließt (Abbildung 4). Die Modelle bestehen hierbei aus einer Verschaltung einfacher mathematischer Mechanismen (beispielsweise Multiplikation mit einem Parameter, Addition eines Parameters, nichtlineare Operationen), die so strukturell aufgebaut sind, dass am Ende prinzipiell Daten im erwünschten Wertebereich und mit der erwünschten Dimensionalität (z.B. Vektorlänge) zurückgeliefert werden. Die Zielsetzung der Optimierung dieser Parameter (überwachtes Lernen, supervised learning) muss im Vorfeld festgelegt werden (beispielsweise eine möglichst genaue Schätzung eines bekannten Labels) und wird dann in einem iterativen Prozess anhand von Eingabedaten (Bildern) und zusätzlich mit Ausgabedaten (assoziierte Labels) durchgeführt (AWAYSHEH *et al.*, 2019; MCALPINE *et al.*, 2020). Die Ausgabedaten

dienen dem Modell als Zielsetzung der richtigen Prädiktionen während des Trainings. Davon abzugrenzen ist das unüberwachte Lernen (unsupervised learning): Hierbei versucht man die Eingabedaten aufgrund statistischer Eigenschaften eigenständig zuzuordnen, beispielsweise in eine bestimmte Anzahl an Kategorien zu unterteilen, ohne dass die Art der Gruppierung von Ausgabedaten beeinflusst wird. Diese Art des Lernens spielt bisher in der Bildgebung im Allgemeinen und in der digitalen Pathologie im Speziellen eine untergeordnete Rolle (RASHIDI *et al.*, 2019) und wird daher hier nicht weiter besprochen. Für die daten-basierten Ansätze ist die Verfügbarkeit von adäquaten Datensätzen zum Training und zur Evaluation (Testen) der Modelle eine unverzichtbare Grundvoraussetzung. Allerdings weisen die entwickelten Algorithmen eine hohe Abhängigkeit zur Datenverteilung auf und eine Übertragbarkeit auf andere Daten mit einer abweichenden Verteilung kann im Unterschied zum Thresholding nicht ohne Weiteres (d.h. durch eine einfache Anpassung von wenigen Parametern) vorgenommen werden.

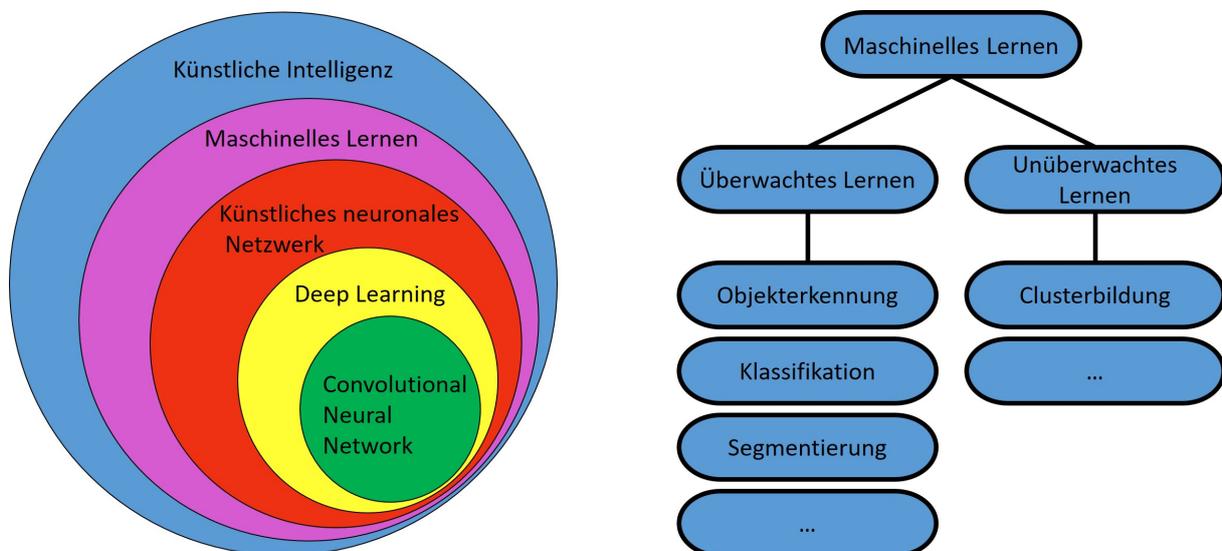


Abbildung 4: Graphische Darstellung der Untergruppen der künstlichen Intelligenz.

Abbildungen modifiziert nach TURNER *et al.* (2020)

Für das Trainieren der Modelle basierend auf dem traditionellen maschinellen Lernen werden umfangreiche domainspezifische Information über relevante Merkmalkriterien des gesuchten Musters für die Eingabe in das Modell benötigt (AWAYSHEH *et al.*, 2019). Die Merkmalkriterien sind jedoch oft in ihrer vollen Komplexität nicht bekannt (AUBREVILLE, 2020). Im Gegensatz dazu sind Deep learning-basierte Modelle aufgrund der Verwendung von komplexen Netzwerkarchitekturen und Millionen von Parametern in der Lage relevante und sehr komplexe Bildeigenschaften eigenständig zu extrahieren. Die Verbindungen von zahlreichen Neuronen (angeordnet in Schichten) in komplexen Netzwerken ermöglicht eine

hohe Leistungsfähigkeit (MAIER *et al.*, 2019). Es ist daher nicht erstaunlich, dass komplex morphologische Muster, wie Mitosefiguren, mit einer deutlich höheren Genauigkeit prädiziert werden können, wenn Deep learning-basierte Ansätze – anstelle von dem traditionellen maschinellen Lernen – verwendet werden (SOMMER *et al.*, 2012; IRSHAD, 2013; VETA *et al.*, 2015; LI *et al.*, 2018; AUBREVILLE, 2020; AUBREVILLE *et al.*, 2020c; MATHEW *et al.*, 2020). Obwohl Deep learning-basierte Algorithmen für die meisten Fragestellungen der digitalen Pathologie den aktuellen Stand der Technik darstellen, sind sie extrem „daten-hungrig“ und können ihr volles Potential erst bei umfangreichen Datenbanken und hoher Rechenleistung entfalten (ABELS *et al.*, 2019; BERTRAM *et al.*, 2019b) (Abbildung 5). Eines der relevantesten Modelle für die digitale Pathologie sind Convolutional Neural Networks (ein spezifischer Typ des künstlichen neuronalen Netzwerks), welche neben vollverbundenen (also jedes Neuron mit jedem anderen Neuron) Schichten auch Schichten mit Faltungsoperatoren einsetzen (BERA *et al.*, 2019; MAIER *et al.*, 2019). Die Faltungsoperatoren werden genutzt, um Muster aus benachbarten Pixelregionen zu extrahieren. Die Filtergewichte der Faltungsoperation werden hierbei, wie alle anderen Gewichte im Netzwerk, anhand den Daten erlernt. Über mehrere Stufen, kombiniert mit einer räumlichen Verdichtung, gelingt es so zunächst einfache Muster (wie Kanten) zu extrahieren und diese im weiteren Verlauf des Netzwerks zur Erkennung komplexer Muster anzureichern.

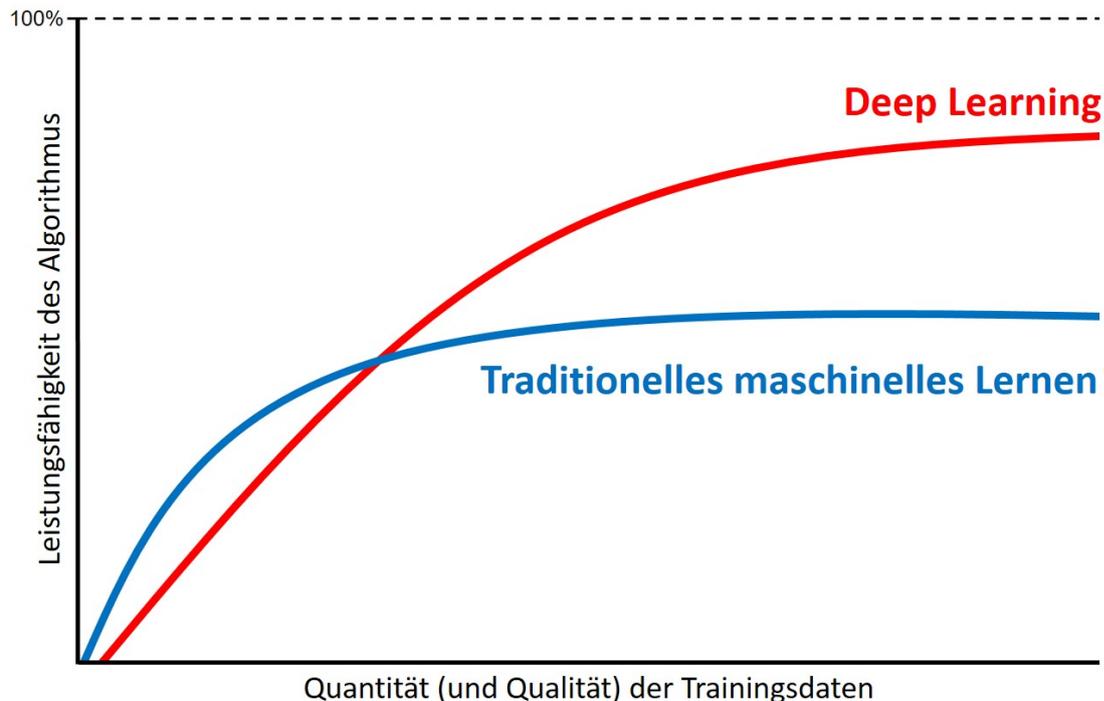


Abbildung 5: Beispielhafter Vergleich der Leistungsfähigkeit von daten-basierten Modellen in Abhängigkeit von der Trainingsdatenquantität.

Bild modifiziert nach (ZURAW *et al.*, 2020)

Daten-basierte Methoden haben ein sehr breites Anwendungsgebiet und können je nach Aufgabenstellung stark veränderte Ansprüche an die Modelle und Datensätze erzeugen. Überwachte daten-basierte Algorithmen können unter anderem folgende Aufgabenstellungen für die Mustererkennung aufweisen: 1) Objekterkennung, 2) Segmentierung und 3) Bildklassifikation (AUBREVILLE, 2020). Bei der Objekterkennung werden bestimmte Muster (einzelne Zellen oder gewebliche Strukturen) im Bild lokalisiert (x- und y-Koordinate) und gleichzeitig einer Labelklasse zugewiesen (Klassifikation) (AUBREVILLE, 2020). Häufig untersuchte Fragestellungen im Bereich der digitalen Pathologie sind Mitosefiguren in histologischen Tumorschnitten (MATHEW *et al.*, 2020). Ein Beispiel für die Objekterkennung einer Gewebestruktur ist die Identifikation von Tumormetastasen (humane Mammakarzinom) in regionalen Lymphknoten (BEJNORDI *et al.*, 2017). Eine wesentliche Limitation von Objekterkennungsansätzen ist die Notwendigkeit von lokalen Annotationen der gesuchten Objekte im Bild für das Training und Evaluation der Modelle. Die manuelle Erstellung dieser Labels durch Experten/innen (Annotatoren/innen) ist extrem zeitaufwendig, insbesondere bei zellulären Strukturen. Deshalb steht oft nur eine eher geringe Anzahl an Annotationen zur Verfügung, welche die Generalisierungsleistung der Modelle entscheidend einschränken kann (BERTRAM *et al.*, 2019b) (siehe auch Kapitel 2.3.2.2).

Die Segmentierung ist die genaue Demarkierung der Grenzen eines Objektes oder einer Struktur (einzelne Zellen oder gewebliche Strukturen) zu den angrenzenden Strukturen auf Pixelebene (AUBREVILLE, 2020). Anstelle einer einzelnen Bildkoordinate (Objekterkennung) werden hier alle Pixel, die dem Objekt zugesprochen werden, zusammengefasst. Beispiele einer möglichen Fragestellung für die Segmentierung sind ebenfalls Mitosefiguren (LI *et al.*, 2019) und Flächen von Tumoranteilen in einem Gewebeschnitt (GUO *et al.*, 2019; PAN *et al.*, 2020; KIM *et al.*, 2021). Für diese Mustererkennungsfragestellung sind ebenfalls lokale (pixel-basierte) Annotationen der jeweiligen Strukturen im WSIs notwendig (WANG *et al.*, 2019).

Bei der Bildklassifikation werden jedem Bild genau eine Kategorie (Labelklasse) von zwei (binäre Klassifikation) oder mehreren möglichen Klassen zugewiesen (AUBREVILLE, 2020). Dementsprechend werden für das Training und Testen des Modells Annotationen auf der Bildebene (ein Bild hat genau ein globales Label) zugewiesen. Diese globalen Labels können in aller Regel deutlich einfacher und mit geringerem Zeitaufwand als lokale Labels generiert werden und einige aktuelle Studien haben auf eine enorme Bildmenge mit bis zu > 40.000 WSIs zurückgreifen können (CAMPANELLA *et al.*, 2019). Anwendungen der Bildklassifikation sind beispielsweise die Einteilung der Gewebeproben in bestimmte Diagnosen (ARESTA *et al.*, 2019), Tumorgrade (BULTEN *et al.*, 2020) oder sogar Muster, die nicht direkt aus dem histologischen Bild für Pathologen/innen ersichtlich sind, wie Patientenoutcome (COURTIOL *et al.*, 2019) oder genetische Eigenschaften von Tumoren (KATHER *et al.*, 2020).

Tabelle 4: Definitionen von relevanten Begriffen der automatisierten Bildanalyse

Begriff	Definition
Algorithmus:	Eine Durchführungsvorschrift zur Lösung eines bestimmten Problems. Im Kontext dieser Arbeit: Eine einsatzbereite Computersoftware die mittels einer automatisierten Analyse von mikroskopischen Bildern (z.B. WSIs) eine Prädiktion zu einer bestimmten Aufgabenstellung (z.B. Objekterkennung von Mitosefiguren) erzeugt. Bei der Verwendung von maschinellem Lernen basiert ein Algorithmus auf einem Modell, gegebenenfalls unterstützt durch Vor- und Nachverarbeitung.
Annotation (für die Objekterkennung):	In der Regel vom Experten/innen erzeugte Interpretation eines Bildobjektes. Ein Datensatz besteht normalerweise aus vielen einzelnen Annotationen, die jeweils das Bild, die Position im Bild und die Labelklasse einer Struktur (z.B. Mitosefigur) beinhalten.
Datensatz (maschinelles Lernen):	Besteht aus mikroskopischen Bildern (z.B. WSIs) und assoziierten Annotationen (gespeichert z.B. in einer Datenbank), welche für das Training (Trainingsdatensatz), Validieren (Validierungsdatensatz) und Testen (Testdatensatz) des Modells genutzt werden können.
Deep Learning:	Eine Untergruppe des maschinellen Lernens mit besonders tiefen, künstlichen neuronalen Netzen (d.h. mit vielen Schichten). Im Unterschied zum traditionellen maschinellen Lernen, sind Deep learning-basierte Ansätze fähig, relevante Bildmerkmale für eine bestimmte Fragestellung eigenständig (ohne manuellem Input eines/r Programmierers/in) zu extrahieren und für die Fragestellung zu nutzen. Entsprechend entfällt beim Deep Learning die Merkmalsextraktion, da diese im Netzwerk enthalten ist.
Convolutional Neural Network (CNN):	Ein spezifischer Typ des künstlichen neuronalen Netzwerks mit hoher Relevanz für das Deep Learning in der digitalen Pathologie. CNNs beinhalten mindestens eine Netzwerkschicht mit einem Faltungsoperator, welcher für die Extraktion von räumlichen Mustern im Bild verantwortlich ist.

Begriff	Definition
Goldstandard (maschinelles Lernen):	Die Methode, welche als besten geeignet für die Generierung von Ground truth Labels eines Datensatzes angesehen wird. Für das Annotieren von Objekten in histologische und zytologische Bilder werden zumeist Annotationen von erfahrenen Pathologen/innen als Goldstandard verwendet. Allerdings haben selbst sehr erfahrene Annotatoren/innen visuelle und kognitive Limitationen, die eine gewisse Variabilität und Ungenauigkeit der Annotationen bedingen.
Ground truth (maschinelles Lernen):	Eine Information in der Datenbank über das als „wahr“ angenommene Label bzw. Annotation basierend auf der Bestimmung mittels einer Goldstandardmethode.
Künstliche Intelligenz (KI):	Ein Teilgebiet der Computerwissenschaften, in dem ein intelligentes Verhalten vom Computer simuliert wird. Eine Untergruppe der KI ist das maschinelle Lernen.
Künstliches neuronales Netzwerk (artificial neural network, ANN):	Ein Modell des maschinellen Lernens, bei dem die künstlichen Neurone (ein Regressionsmodell mit nachgeschalteter Nichtlinearität) in verschiedenen Schichten (Eingabe-, versteckte und Ausgabeschichten) angeordnet und vernetzt sind. Während des Trainings werden die Parameter innerhalb der Neuronen sowie die Gewichte ihrer Verbindungen angepasst.
Label:	Ein Label ist eine definierte morphologische Kategorie / Klasse, welche einer Struktur zugewiesen wird. Zusammen mit der Position dieser Struktur im Bild macht es die Annotation aus.
Maschinelles Lernen (automatisierte Bildanalyse):	Eine Untergruppe der künstlichen Intelligenz, bei der ein beobachteter Prozess modelliert wird. Das Modell wird hierbei anhand von (annotierten) Daten dazu optimiert, bestimmte Muster in Bildern zu identifizieren. Die Netzwerkstruktur ist dabei häufig von der Aufgabenstellung nur zum kleinen Teil beeinflusst. Es werden traditionelle maschinelle Lernansätze und die potenziell leistungsstärkeren Deep-Learning-Ansätze unterschieden.

Begriff	Definition
Überwachtes maschinelles Lernen (automatisierte Bildanalyse):	Eine Form des maschinellen Lernens, bei dem die Modelle von annotierten Daten lernen. Häufige Mustererkennungsaufgaben sind Objekterkennung, Bildklassifikation und Segmentierung. Im Unterschied zum unüberwachten Lernen (in dieser Dissertation nicht weiter besprochen) ist zu jedem Bild des Trainingsdatensatzes der komplette vorherzusagende Inhalt bekannt und kann während des Trainings zur Optimierung eingesetzt werden.
Thresholding (Schwellwertverfahren):	Algorithmus, welcher basierend auf einem Schwellwert ein Objekt oder einen Bildpunkt einer Klasse zuweist. Schwellwertbasierte Verfahren haben häufig die Möglichkeit der manuellen Optimierung des Schwellwerts durch den Anwender in Abhängigkeit vom tatsächlichen Eingangsbild und der erzeugten Prädiktion. Im Vergleich zu den Methoden der KI ist bei diesen Methoden jeder Einzelschritt leicht nachvollziehbar.

Definitionen in Anlehnung an AMERICAN COLLEGE OF VETERINARY PATHOLOGISTS (2021) und ABEL *et al.* (2020).

2.3.2 Deep Learning für die Objekterkennung

Deep learning-basierte Modelle können Bildanalysen mit einer zuvor unerreichten Leistungsfähigkeit durchführen und stellen für viele pathologischen Fragestellungen den aktuellen Stand der Technik dar (AUBREVILLE, 2020). Sie weisen jedoch auch eine sehr hohe Komplexität der Modelle auf, welche beträchtliche Anforderungen an die Entwicklung sowie eine eingeschränkte Nachvollziehbarkeit der Ergebnisse bedingt. Die daraus resultierende Skepsis, Unwissenheit und Reserviertheit von Pathologen/innen behindert entscheidend die Weiterentwicklung der digitalen Pathologie in der Tiermedizin. Erste wissenschaftliche Erkenntnisse für die Tierpathologie hinsichtlich der Leistungsfähigkeit im Vergleich zu erfahrenen Pathologen/innen und den möglichen Fehlerquellen für die Patientensicherheit sind entscheidende Beiträge, um das Potential und Limitationen von Deep learning-basierten Algorithmen für tiermedizinische Diagnostiklabore zu ergründen. Mit den eigenen Untersuchungen wurden deshalb erste Deep learning-basierte Objekterkennungsansätze von verschiedenen relevanten diagnostischen Fragestellungen der Veterinärpathologie und -Zytologie untersucht. Im den nachfolgenden Abschnitten wird ein Überblick über den generellen Entwicklungsprozess von Deep learning-basierten Modellen gegeben.

2.3.2.1 Erstellung von Ground truth Datensätzen

Die sorgfältige Erstellung von umfangreichen Datensätzen ist sehr zeitaufwendig (MARZAHN *et al.*, 2020a; MARZAHN *et al.*, 2020d) aber elementar für die Entwicklung von Deep learning-basierten Modellen (MARÉE, 2017). Der Datensatz muss ausreichend informativ für das Training des Modells sein (MCALPINE und MICHELOW, 2020). Mithilfe einer speziellen Annotierungssoftware werden die gesuchten Objekte in den Bildern annotiert (AUBREVILLE *et al.*, 2018). Dabei werden die Annotationen mit einer definierten Goldstandardmethode durchgeführt. Für morphologische Strukturen in mikroskopischen Bildern stellt zumeist ein/e (oder mehrere) erfahrene/r Pathologe/in den Goldstandard dar (ABELS *et al.*, 2019). Allerdings kann zwischen verschiedenen Labelingmethoden und individuellen Annotatoren/innen erhebliche Variabilität auftreten (BERTRAM *et al.*, 2020b; MARZAHN *et al.*, 2020a; MARZAHN *et al.*, 2020d; WILM *et al.*, 2021). Neben der hohen Qualität der Annotationen sollte sowohl eine ausreichende Quantität der Annotationen als auch eine realistische Variabilität der Bilder vorhanden sein, damit leistungsstarke und robuste Algorithmen entwickelt werden können (siehe Kapitel 4.3.4). Die erstellten Datensätze sollten stets in zwei Teile unterteilt werden, welche ausschließlich für das Training (inklusive Optimierung der Gewichte und Validierung des Modells) beziehungsweise die Evaluation verwendet werden (Abbildung 6): 1A) Trainingsdatensatz, 1B) Validierungsdatensatz und 2) Testdatensatz (LIU *et al.*, 2019; MAIER *et al.*, 2019; MCALPINE und MICHELOW, 2020).

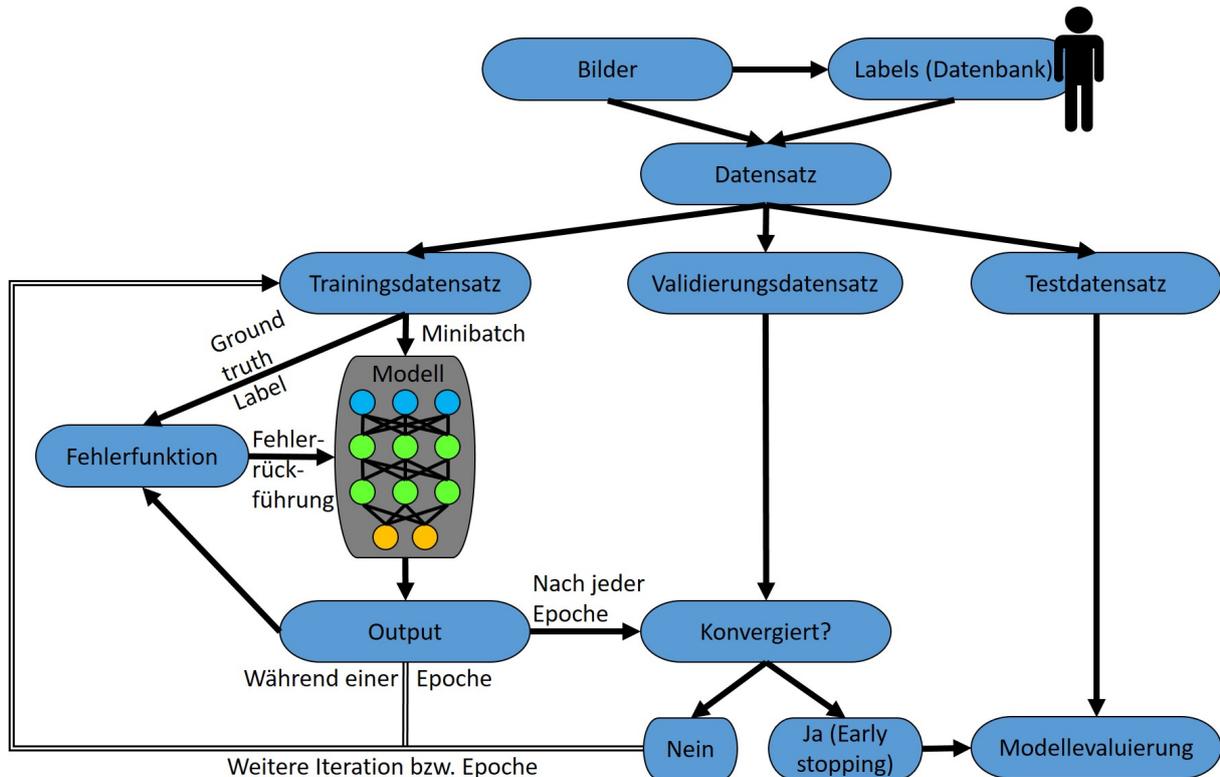


Abbildung 6: Schematische Darstellung des Ablaufs von Training und Evaluation eines Deep learning-basierten Modells unter Verwendung der verschiedenen Datensatzanteile

2.3.2.2 Modell Training und Validierung

Das Training von neuronalen Netzwerken ist ein iterativer Prozess mit zahlreichen Schritten und dem Ziel, die Fehler (Loss) des daten-basierten Modells zunehmend zu reduzieren (beispielsweise durch das Gradientenabstiegsverfahren) (MAIER *et al.*, 2019). In der Regel werden Modelle verwendet, die schon auf großen Bilddatensätzen vortrainiert wurden und für die Aufgabenstellung in einer neuen Domäne weitertrainiert werden (RASHIDI *et al.*, 2019). Während des Trainings wird pro Iteration jeweils eine bestimmte Trainingschleife durchlaufen (Abbildung 6 und 7). Für eine Iteration werden dem Modell eine bestimmte Anzahl an Eingabedaten (Minibatches), die zufällig aus dem Datensatz gezogen werden, präsentiert und die entsprechende Modellantwort berechnet (forward pass). Die Eingabedaten werden im Modell von den verschiedenen Neuronen, welche untereinander mit einer bestimmten Gewichtung verbunden sind, mittels mathematischer Operatoren modifiziert und an die nächsten Neuronen weitergeleitet. In der Ausgabeschicht werden die generierten Informationen zusammengeführt. Nach der Analyse dieser Eingabedaten wird die Ausgabedaten (Prädiktion) des Modells mit den erwarteten Werten des Ground truth Datensatzes verglichen und mittels einer Fehlerfunktion auf eine einzelne Zahl überführt. Ziel der Optimierung ist dabei die Reduktion des skalaren Fehlers. Basierend darauf werden dann

die Netzwerkparameter (Gewichte der Neuronenverbindungen und sonstige Parameter) mittels Fehlerrückführung (backpropagation) optimiert (MAIER *et al.*, 2019). Die Fähigkeit des Modells ein bestimmtes Muster zu erkennen, liegt letztendlich in den spezifischen Gewichten der Verbindungen zwischen den Neuronen. Anschließend findet jeweils eine weitere Iteration mit einem weiteren Trainingsbatch statt. Im Verlauf des Trainings sollte der Fehler tendenziell stetig abnehmen. Aufgrund der zufälligen Auswahl der Trainingsbatches ist dies aber nur gemittelt über mehrere Minibatches der Fall. Auch ist der Trainingsverlauf nicht deterministisch (d.h. nicht 100% reproduzierbar).

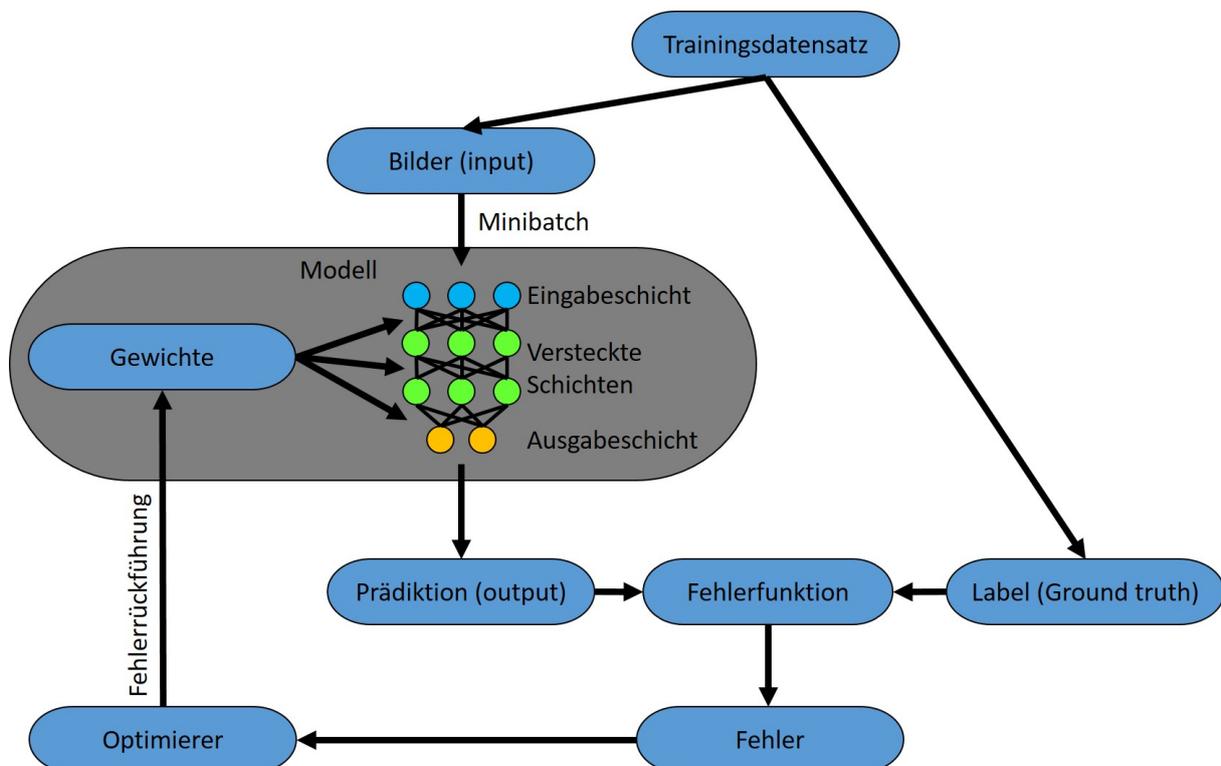


Abbildung 7: Schematische Darstellung der Trainingsschleife für Deep Learning.

Abbildung modifiziert von MCALPINE und MICHELOW (2020)

Anhand des Validierungsdatensatzes wird in regelmäßigen Abständen, das heißt nach jeder Epoche (Trainingszyklus eines vollständigen Datensatzes), überprüft, wie hoch die Generalisierungsleistung (reziprok zum Validierungsverlust) des Modells zum aktuellen Zeitpunkt auf ungesehenen Validierungsdaten ist (MAIER *et al.*, 2019). Die Generalisierungsleistung beschreibt wie gut die extrahierten Muster für andere Daten, die nicht zum Training verwendet wurden, zutreffend sind. Generell verläuft bei guten Modellen und hinreichend großer Anzahl repräsentativer Trainingsdaten der Validierungsverlust und der Trainingsverlust mehr oder weniger parallel zueinander. Das heißt das Modell lernt im zunehmenden Maß das gesuchte Muster gut zu generalisieren. Ein zu frühes Abbrechen der

Trainingsschleife würde dazu führen, dass das Modell eine geringere Leistungsfähigkeit aufweist als es hätte erreichen können (Underfitting; Abbildung 8). Da das Modell mit zunehmenden Iterationen/Epochen des Trainings auch zunehmend spezifischere Muster des Trainingsdatensatzes extrahiert, kommt es allerdings auch vor, dass ab einer bestimmten Anzahl der Iterationen/Epochen (unter anderem abhängig von der Variabilität des Trainingsdatensatzes) der Vorhersagefehler auf dem Trainings- und Validierungsdatensatz divergieren (MAIER *et al.*, 2019). Ab diesem Zeitpunkt erlernt das Modell Muster, die zu spezifisch für den Trainingsdatensatz sind und nicht mehr auf anderen Daten generalisieren (Overfitting). Anhand des Validierungsfehlers kann der optimale Zeitpunkt zum Beenden des Trainings (Early Stopping) bestimmt werden (Abbildung 8) (AUBREVILLE, 2020). Das Optimieren der Hyperparameter für das Training und die Bestimmung der idealen Objekterkennungs- und -Klassifikationsgrenzwerte erfolgt ebenfalls anhand des Validierungsdatensatzes (AUBREVILLE, 2020).

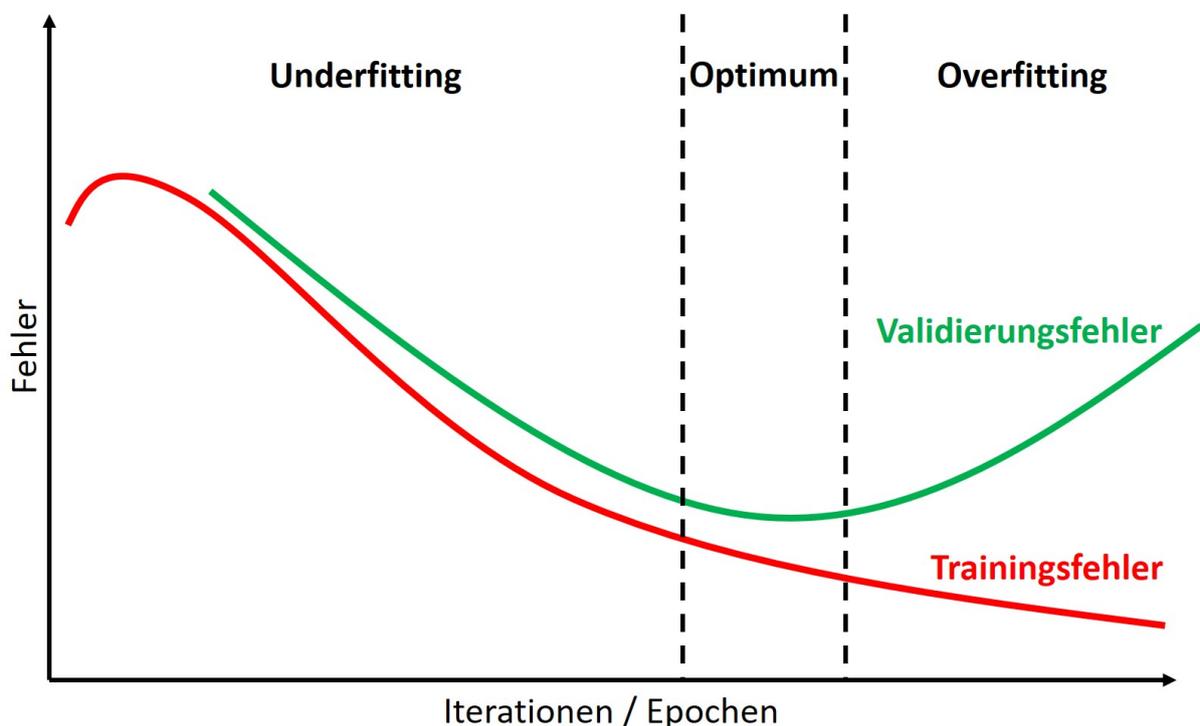


Abbildung 8: Schematische Darstellung einer möglichen Fehlerkurve für die Trainings- und Validierungsfehler über eine zunehmende Anzahl an Trainingsiterationen/epochen. Der optimale Zeitpunkt für das Early Stopping ist der geringste Validierungsfehler.

Abbildung modifiziert von RASHIDI *et al.* (2019)

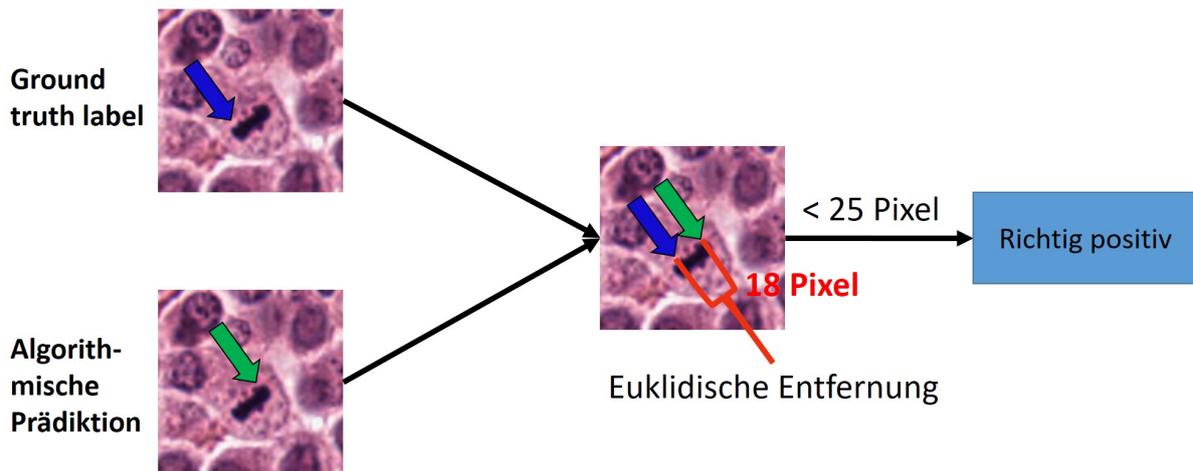
2.3.2.3 Modell Evaluation (Testen)

Erst die Evaluation des finalen Modells erfolgt auf den unabhängigen Testdatensatz von dem die Bilder nicht für Training und Validierung verwendet wurden (MAIER *et al.*, 2019). In der Regel werden die Testdatensätze im selben Labor (gleiche Präparatherstellungsmethode), Spezies und Präparattyp wie der Trainingsdatensatz hergestellt. Die Unabhängigkeit des Testdatensatzes ist unter anderem entscheidend um sicherzustellen, dass das Modell nicht durch unerwünschte Bildeigenschaften der Trainingsdaten, die nicht generell anwendbar für das gesuchte Objekt sind, negativ beeinflusst wurde (BERA *et al.*, 2019). Für eine Überprüfung der Robustheit (Generalisierungsfähigkeit des Modells unter variablen Bildbedingungen) können Bilder aus anderen Laboren, Spezies oder/und Präparattypen (zum Beispiel andere Tumortyp) verwendet werden. Die Ground truth Annotationen sollten jedoch, wenn möglich, mit den gleichen Methoden und Annotatoren/innen hergestellt werden wie im Trainingsdatensatz, um eine möglichst äquivalente Interpretation zu ermöglichen. Eine vorherige Studie hat gezeigt, dass Datenbanken von verschiedenen Arbeitsgruppen (mit unterschiedlichen Labelingmethoden) für die gleichen Bilder erhebliche Abweichungen zueinander aufweisen können (BERTRAM *et al.*, 2020b). Variabilität zwischen Datensätzen (zum Beispiel in der Qualität und Quantität der Labels, Labelingmethode, Komplexität der Bilder und vielem mehr) hat einen offensichtlichen Einfluss auf das Training und Testen der Modelle (BERTRAM *et al.*, 2020b; MATHEW *et al.*, 2020). Ein direkter Vergleich der Leistungsfähigkeit von Modellen, die mit unterschiedlichen Datensätzen evaluiert wurden, sollte möglichst vermieden werden.

Deep learning-basierte Algorithmen für die Objekterkennung präzisieren eine bestimmte Labelklasse in einer bestimmten Lokalisation im Bild. Die Lokalisation des Objekts kann dabei sowohl als Flächenschwerpunktannotation (ermittelter Mittelpunkt des Objekts) oder mittels Bounding Boxes (rechteckige Annotation variabler Größe, die das Objekt idealerweise vollständig einschließt) angegeben werden. Fehler in der Objekterkennung können aufgrund einer falschen Lokalisation oder Klassenzuordnung entstehen. Durch den Vergleich der Prädiktionen und der Ground truth Annotationen kann die Anzahl der richtig positiven, falsch positiven und falsch negativen (nicht aber der richtig negativen) Prädiktionen bestimmt werden (Abbildung 9 und 10). Die Entscheidung, ob der Abstand zwischen der Prädiktion und die Ground truth Annotation ausreichend klein ist, um als übereinstimmend angesehen zu werden, kann entweder über die euklidische Entfernung oder ein Flächenverhältnis (Intersection over Union, IoU) erfolgen (Abbildung 9). Häufig verwendete Bewertungsmaßeinheiten für ein zwei-Klassen-Problem (gesuchtes Objekt versus Hintergrund; das heißt, das Erkennen einer bestimmten Objektklasse in einem Bild) ist der F_1 -Wert (harmonische Mittel von Recall und Precision; Tabelle 5) oder die Average Precision (AP, Mittelwert von mehreren Precision-

Messungen für unterschiedliche Recall-Werte; Tabelle 5). Beide Werte unterscheiden sich vor allem dadurch, dass beim F_1 -Wert eine binäre Erkennung vorliegt, also ein einziger Schwellwert für die Wahrscheinlichkeit der Erkennung angewandt wird. Die richtige Wahl dieses Schwellwerts kann dabei die Qualität der Prädiktionen entscheidend beeinflussen. Im Gegensatz dazu kommt die AP-Metrik ohne einen solchen Schwellwert aus und berechnet eine generelle Erkennungsstärke eines Modells. Sollen in diesem Fall mehrere Labelklassen bewertet werden, wird häufig der Mittelwert aller AP-Werte der einzelnen Klassen (mean Average Precision, mAP) als Metrik genutzt (Tabelle 5).

Flächenschwerpunkts-annotation



Bounding Box Annotation

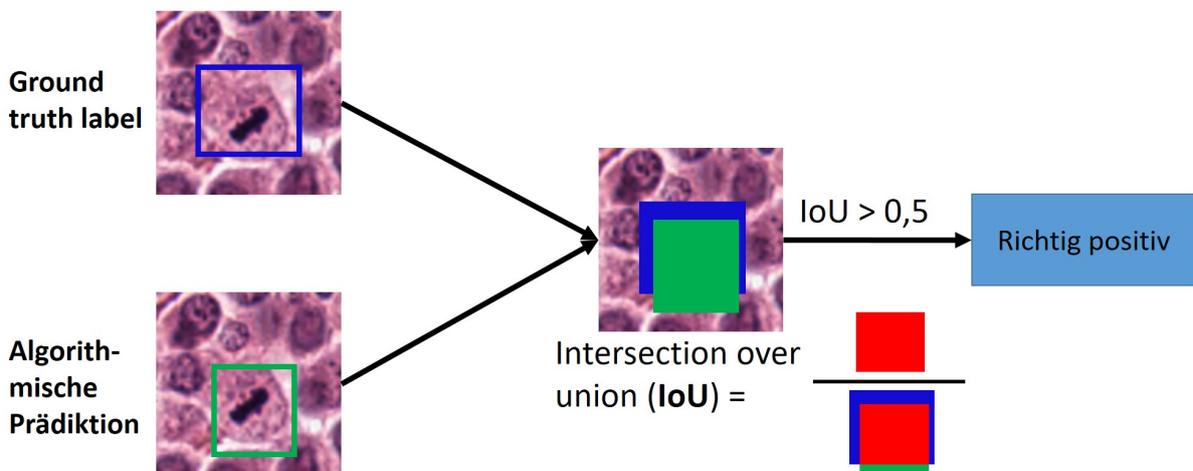


Abbildung 9: Beispiele für die Entscheidung der Übereinstimmung von Ground truth Annotationen und algorithmischen Prädiktionen (richtig positive Prädiktion) anhand von Mitosefiguren in histologischen Bildausschnitten

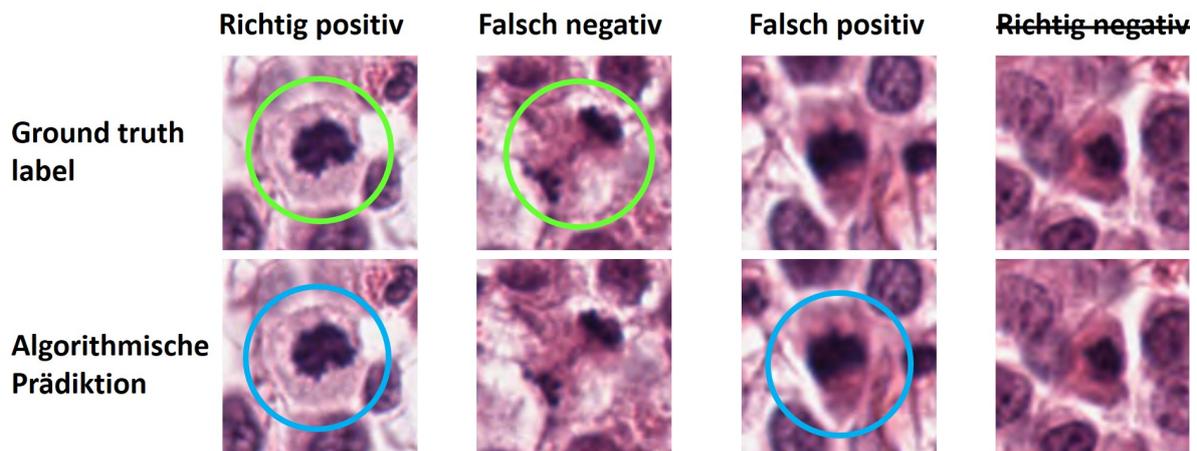


Abbildung 10: Beispiele von richtig positiven, falsch negativen und falsch positiven Prädiktionen von Mitosefiguren. Die Betrachtung richtig negativer Prädiktionen ist nicht sinnvoll bei der Objekterkennung (im Unterschied zur Bildklassifikation)

Tabelle 5: Relevante Metriken für die Bewertung von automatisierten Bildanalysealgorithmen für die Objekterkennung

Metriken	Definition
Average Precision (AP):	Ist der Mittelwert von mehreren Precision-Messungen für unterschiedliche (meistens N=11) Recall-Werte (Schätzung der Recall-Precision-Kurve). Die AP kann für eine binäre Objekterkennung (gesuchtes Objekt versus Hintergrund) bestimmt werden und reicht von 0 (schlechte Leistung) bis 1 (perfekte Leistung). Für die Leistungsevaluation von multi-Klassen-Objekterkennungs-Fragestellungen kann die mAP aus mehreren AP-Werten für die einzelnen Klassen (Query) bestimmt werden (siehe unten).
Falsch negative (FN) Prädiktion:	Die gesuchte Struktur wurde nicht erkannt. Es liegt keine Prädiktion vom Modell mit einem maximalen Pixelabstand oder IoU-Wert zur Ground truth Annotation vor (siehe Abbildung 10).
Falsch positive (FP) Prädiktion:	Das Modell hat fälschlicherweise ein Objekt erkannt. Die Prädiktion des Modells liegt außerhalb des maximalen Pixelabstands oder IoU-Wert zu einer Ground truth Annotation (siehe Abbildung 10).

Metriken	Definition
F ₁ -Wert:	$F_1 = \frac{2TP}{2TP+FP+FN}$ <p>Oder</p> $F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$ <p>Ist das harmonische Mittel vom Recall und Precision und reicht von 0 (schlechte Leistung) bis 1 (perfekte Leistung). Im Unterschied zu AP stellt der F₁-Wert einen einzigen Punkt der Precision-Recall-Kurve dar mit einem sinnvollen Precision-Recall-Kompromiss. Dieser Punkt entsteht durch Festlegung eines Schwellwerts.</p>
Intersection over Union (IoU)	$IOU = \frac{\text{Überlappende Fläche (intersection)}}{\text{Gemeinsame Fläche (union)}}$ <p>Misst die Überlappung von Bounding boxes der Ground truth Annotation und der algorithmischen Objekterkennung. Ein definierter IoU-Grenzwert kann zur Entscheidung der TP, FP, FN Prädiktionen verwendet werden (siehe Abbildung 9).</p>
mean Average Precision (mAP):	$mAP = \frac{AP_{Q1} + AP_{Q2} + \dots + AP_{QN}}{N}$ <p>Der Mittelwert aller bestimmten AP-Werte (N) für verschiedene Klassen (Query, Q). Eignet sich für die Leistungsevaluation von Multi-Klassen-Objekterkennungs-Fragestellungen. Dafür werden AP-Werte der einzelnen Klassen (binäre Klassifikation) bestimmt und der Mittelwert von allen (N) Klassen (Q) berechnet.</p>
Precision (positiver prädiktiver Wert):	$Precision = \frac{TP}{TP + FP}$ <p>Gibt an, wie viele Prädiktionen des Modells wahrhaftig (laut Ground truth) gesuchte Strukturen darstellen.</p>

Metriken	Definition
Recall (Sensitivität):	$\text{Recall} = \frac{TP}{TP + FN}$ <p>Gibt an, wie viele gesuchten Strukturen (laut Ground truth) tatsächlich vom Modell erkannt wurden.</p>
Richtig positive (TP) Prädiktion:	Das Modell hat die gesuchte Struktur (Ground truth Annotation) richtig entdeckt. Die Prädiktion des Modells liegt innerhalb des maximalen Pixelabstands oder IoU-Wert zu einer Ground truth Annotation (siehe Abbildung 10).

Definitionen in Anlehnung an AUBREVILLE (2020) und MCALPINE und MICHELOW (2020)

2.3.2.4 Implementierung und Anwendung des Modells

Die entwickelten und getesteten Modelle werden abschließend in einen gebrauchsfertigen und benutzerfreundlichen Algorithmus implementiert. Beispielsweise kann in einem Algorithmus – zusätzlich zu einer Deep learning-basierten Objekterkennung – eine computerisierte Heatmap-Berechnung oder Regionvorauswahl erfolgen (AUBREVILLE *et al.*, 2020c). Ein weiteres Beispiel eines einsatzbereiten Algorithmus ist die Implementierung einer Deep learning-basierten Einteilungen von Zellen in quantitative Klassen mit nachfolgender automatisierter Berechnung (Nachverarbeitung) eines Scores (MARZAHN *et al.*, 2020c). Bisher haben nur wenige Studien untersucht wie Algorithmen im diagnostischen Alltag sinnvollerweise eingesetzt werden sollen. Aufgrund der geringen Nachvollziehbarkeit der Prädiktionen („Blackbox“) und der hohen Abhängigkeit der Leistungsfähigkeit vom Trainingsdatensatz erscheint eine Überprüfung der algorithmischen Ergebnisse durch einen/r erfahrenen Pathologen/in sinnvoll, um eine hohe Zuverlässigkeit der Diagnose sicherzustellen. Dabei sollte der Algorithmus den/die Pathologen/in in den kritischen Schritten der Diagnosestellung bestmöglich unterstützen und einfach in der Anwendung sein (MCALPINE *et al.*, 2020). So kann beispielsweise ein Algorithmus mit einer hohen Sensitivität für das Auffinden seltener und unregelmäßig verteilter Strukturen mit einer hohen Spezifität von erfahrenen Pathologen/innen kombiniert werden (STEINER *et al.*, 2018; AUBREVILLE *et al.*, 2020c). Computer-assistierte Entscheidungs-Systeme haben das Potential, die Genauigkeit, Reproduzierbarkeit und Effizienz der Bewertung durch einen/eine Pathologen/in zu verbessern, wenn die verwendeten Algorithmen insgesamt eine hohe Leistungsfähigkeit aufweisen und Pathologen/innen in der Lage sind mögliche Fehler zu identifizieren und korrigieren (STEINER *et al.*, 2018; AUBREVILLE *et al.*, 2020c; PANTANOWITZ *et al.*, 2020).

3 Eigene Untersuchungen

3.1 Validierung der digitalen Mikroskopie für die Diagnose von kaninen Hauttumoren

- Titel:** Validation of Digital Microscopy Compared With Light Microscopy for the Diagnosis of Canine Cutaneous Tumors
- Autoren:** Bertram CA, Gurtner C, Dettwiler M, Kershaw O, Dietert K, Pieper L, Pischon H, Gruber AD, Klopffleisch R
- Jahr:** 2018
- Zeitschrift:** Journal of Veterinary Pathology (Vet Pathol), 55.4, 490-500
- DOI:** <https://doi.org/10.1177/0300985818755254>

Wissenschaftlicher Beitrag: Ich habe größtenteils die Planung, Umsetzung und Auswertung der Studie durchgeführt. Das Manuskript und die verwendeten Graphen und Tabellen wurden von mir angefertigt. Prof. Dr. Robert Klopffleisch hat maßgeblich bei der Planung und Auswertung der Studie sowie Anfertigung des Manuskriptes mitgewirkt. Dr. Laura Pieper und Hannah Pischon haben maßgeblich bei der Auswertung der Studiendaten geholfen. Dr. Corinne Gurtner, Dr. Martina Dettwiler, Dr. Oliver Kershaw, Dr. Kristina Dietert, Prof. Dr. Achim D. Gruber und Prof. Dr. Robert Klopffleisch waren Studienteilnehmer/innen. Alle Autoren haben sich an der Korrektur des Manuskriptes beteiligt.

Wissenschaftliche Erkenntnisse: Die Validierung der diagnostischen Leistungsfähigkeit der digitalen Mikroskopie (im Vergleich zur Lichtmikroskopie) ist eine elementare Grundvoraussetzung für die Einbindung dieser Technologie in die histopathologische Diagnostik. Die eigene Validierungsstudie war die erste veröffentlichte Untersuchung auf dem Gebiet der Veterinärpathologie. Dabei konnten wir nachweisen, dass die Diagnose von kaninen Hauttumoren mittels digitaler Mikroskopie insgesamt nicht signifikant niedriger ist als die Diagnose mittels Lichtmikroskopie. Eine geringe Verschlechterung der diagnostischen Übereinstimmung wurde lediglich für Rundzelltumore und das Tumorigradung von Mastzelltumoren festgestellt. Diese erste Studie zeigt, dass eine Verwendung der digitalen Mikroskopie für die routinemäßige Diagnose möglich sein kann. Weitere Validierungsstudien für die Veterinärpathologie sind unbedingt erforderlich.

3.2 Meinungsumfrage über den Einsatz der digitalen Mikroskopie in der Lehre

Titel: Virtual Microscopy in Histopathology Training: Changing Student Attitudes in 3 Successive Academic Years

Autoren: Bertram CA, Firsching T, Klopffleisch R

Jahr: 2018

Zeitschrift: Journal of Veterinary Medical Education (J Vet Med Educ), 45.2, 241-249

DOI: <https://doi.org/10.3138/jvme.1216-194r1>

Wissenschaftlicher Beitrag:

Ich habe größtenteils die Planung, Umsetzung und Auswertung der Studie durchgeführt. Das Manuskript und die verwendeten Graphen und Tabellen wurden von mir angefertigt. Theresa Firsching hat sich bei der Datenauswertung und Korrektur des Manuskriptes beteiligt. Prof. Dr. Robert Klopffleisch hat maßgeblich bei der Planung und Auswertung der Studie sowie Anfertigung des Manuskriptes geholfen.

Wissenschaftliche Erkenntnisse:

Für diese Studie wurde das Meinungsbild über die Anwendung der digitalen Mikroskopie im Histopathologie-Curriculum mittels eines Fragebogens bestimmt. Dabei wurden erstmals die Antworten von Studierende aus drei aufeinanderfolgenden Jahrgängen (mit zunehmendem Erfahrungslevel) vergleichend ausgewertet. Während sich das Meinungsbild der drei Jahrgänge für einen Großteil der Fragen ähnlich darstellte, waren sehr deutliche Unterschiede hinsichtlich der wahrgenommenen Nachteile der digitalen Mikroskopie erkennbar. Die jüngeren Jahrgänge befürchteten vor allem eine mangelnde Ausbildung in der Lichtmikroskopie, bei den höheren Semestern standen technische Limitationen im Vordergrund. Die Meinungsumfrage hat gezeigt, dass die digitale Mikroskopie von Studierenden als ein geeignetes Instrument für die Ausbildung angesehen wird. Jedoch scheint das Erfahrungslevel mit dieser Technologie einen relevanten Einfluss auf die wahrgenommenen Nachteile zu haben.

3.3 Mitosedatensatz von kaninen kutanen Mastzelltumoren

Titel: A large-scale dataset for mitotic figure assessment on whole slide images of canine cutaneous mast cell tumor

Autoren: Bertram CA*, Aubreville M*, Marzahl C, Maier A, Klopffleisch R
* im gleichen Maß beigetragen

Jahr: 2019

Zeitschrift: Scientific Data (Sci Data), 6.1, 1-9

DOI: <https://doi.org/10.1038/s41597-019-0290-4>

Wissenschaftlicher Beitrag: Prof. Dr. Marc Aubreville und ich haben im gleichen Maß zur Planung, Umsetzung und Auswertung der Studie beigetragen. Prof. Dr. Robert Klopffleisch, Prof. Dr. Andreas Maier und Christian Marzahl haben bei der Planung und Auswertung geholfen. Ich habe als erster Annotator und Prof. Dr. Robert Klopffleisch als zweiter Annotator fungiert. Prof. Dr. Marc Aubreville hat bei der algorithmisch-assistierten Datenbankerstellung geholfen und war für das Training und Testen der Modelle verantwortlich. Das Manuskript wurde gemeinsam von Prof. Dr. Marc Aubreville und mir verfasst. Alle Autoren haben sich an der Korrektur des Manuskriptes beteiligt. Dieses Forschungsprojekt war ebenfalls Bestandteil der Dissertation von Prof. Dr. Marc Aubreville (AUBREVILLE, 2020).

Wissenschaftliche Erkenntnisse: Die Anzahl der Mitosefiguren in histologischen Schnitten ist eines der relevantesten Prognosekriterien von kaninen Mastzelltumoren; weist jedoch eine hohe inter- und intra-observer Variabilität auf. Daher besteht eine große Nachfrage nach einer Entwicklung von daten-basierte Algorithmen. Allerdings ist das Vorhandensein von qualitativ und quantitativ hochwertigen Datensätzen eine unverzichtbare Grundvoraussetzung für die Entwicklung. Mit der vorliegenden Studie wurde der erste öffentlich verfügbare Mitosedatensatz von einem kaninen Tumor veröffentlicht. Die Anzahl der Labels für Mitosefiguren (N = 44.880) hat dabei vorherige Mitosedatensätze von humanen Tumoren um einen Faktor von mindestens 29 übertroffen. Erstmals wurden gesamte WSIs (N = 32) vollständig annotiert, welches einen entscheidenden Vorteil bezüglich der Gewebevariabilität darstellt und die Berechnung der algorithmischen Leistung (zum Beispiel für die Regionauswahl der höchsten Mitosedichte) auf den gesamten WSIs ermöglicht. Das Deep learning-basierte Modell hat einen F_1 -Wert von 0,82 auf den Testbildern erreicht. Dieser Wert ist unerreicht von vorherigen Studien und spricht für die ausgesprochen hohe Quantität und Qualität des Datensatzes.

3.4 Untersuchung zur Variabilität des Mitotic Count

- Titel:** Computerized Calculation of Ground Truth Mitotic Distribution in Canine Cutaneous Mast Cell Tumor Sections: Mitotic count is Area Dependent
- Autoren:** Bertram CA*, Aubreville M*, Gurtner C, Bartel A, Corner SM, Dettwiler M, Kershaw O, Noland EL, Schmidt A, Sledge DG, Smedley RC, Thaiwong T, Kiupel M, Maier A, Klopffleisch R
* im gleichen Maß beigetragen
- Jahr:** 2020
- Zeitschrift:** Journal of Veterinary Pathology (Vet Pathol), 57.2, 214-226
- DOI:** <https://doi.org/10.1177/0300985819890686>

Wissenschaftlicher Beitrag: Prof. Dr. Marc Aubreville und ich haben im gleichen Maß zu Planung, Umsetzung und Auswertung der Studie und Erstellung des Manuskriptes beigetragen. Prof. Dr. Robert Klopffleisch, Prof. Dr. Andreas Maier und Alexander Bartel haben bei der Planung und Auswertung geholfen. Corinne Gurtner, Sarah M. Corner, Martina Dettwiler, Olivia Kershaw, Erica L. Noland, Anja Schmidt, Dodd G. Sledge, Rebecca C. Smedley, Tuddow Thaiwong, Matti Kiupel und Robert Klopffleisch waren Studienteilnehmer/innen. Alle Autoren haben sich an der Korrektur des Manuskriptes beteiligt. Dieses Forschungsprojekt war ebenfalls Bestandteil der Dissertation von Prof. Dr. Marc Aubreville (AUBREVILLE, 2020).

Wissenschaftliche Erkenntnisse: Der Mitotic Count ist einer der relevantesten histologischen Prognoseparameter von kaninen Mastzelltumoren, welcher jedoch eine sehr hohe inter- und intra-observer Variabilität aufweist. Laut aktuellen Empfehlungen soll der Mitotic Count in der Tumorregion mit der höchsten Dichte bestimmt werden. Allerdings ist die Regionauswahl durch Pathologen/innen sehr variabel. In dieser Studie haben wir untersucht, wie die Mitosedichte in Gewebeschnitten von kaninen Mastzelltumoren verteilt ist und wie groß der Einfluss der Regionauswahl auf die Reproduzierbarkeit des Mitotic Count sein kann. Anhand eines Datensatzes mit 28 vollständig annotierten WSIs (veröffentlicht in BERTRAM *et al.* (2019b)) konnten wir zeigen, dass die Mitosedichte regional sehr unterschiedlich ist und zumindest bei einem Teil der Tumore die Bestimmung der Prognose entscheidend beeinflussen kann. Nützliche Kriterien, die eine Verbesserung der manuellen Regionauswahl erleichtern könnten (wie z.B. die Bevorzugung der Tumorperipherie), waren nicht erkennbar. Schlussfolgernd erscheint die Entwicklung von Bildanalysealgorithmen für die computer-assistierte Regionauswahl mit der höchsten Mitosedichte sinnvoll.

3.5 Mitosedatensatz von kaninen Mammakarzinomen

Titel: A completely annotated whole slide image dataset of canine breast cancer to aid human breast cancer research

Autoren: Aubreville M*, Bertram CA*, Donovan TA, Marzahl C, Maier A, Klopffleisch R
* im gleichen Maß beigetragen

Jahr: 2020

Zeitschrift: Scientific Data (Sci Data), 7, 417

DOI: <https://doi.org/10.1038/s41597-020-00756-z>

Wissenschaftlicher Beitrag:

Prof. Dr. Marc Aubreville und ich haben im gleichen Maß zur Planung, Umsetzung und Auswertung der Studie beigetragen. Prof. Dr. Robert Klopffleisch, Prof. Dr. Andreas Maier und Christian Marzahl haben bei der Planung und Auswertung geholfen. Ich habe als erster Annotator, Prof. Dr. Robert Klopffleisch als zweiter Annotator und Dr. Taryn A. Donovan als dritte Annotatorin fungiert. Prof. Dr. Marc Aubreville hat bei der algorithmisch-assistierten Datenbankerstellung geholfen und war für das Training und Testen der Modelle verantwortlich. Das Manuskript wurde gemeinsam von Prof. Dr. Marc Aubreville und mir verfasst. Alle Autoren haben sich an der Korrektur des Manuskriptes beteiligt.

Wissenschaftliche Erkenntnisse:

Die Anzahl der Mitosefiguren in histologischen Schnitten ist eines der relevantesten Prognosekriterien von kaninen und humanen Mammatumoren. Für diese Studie haben wir einen Datensatz mit 13.907 Labels für Mitosefiguren und 36.379 Labels für streng Negative (Strukturen mit Verwechslungsgefahr zu Mitosefiguren) in 21 WSIs von kaninen Mammatumoren erzeugt. Die Anzahl der Mitoselabels übertrifft Mitosedatensätze von humanen Mammatumoren anderer Arbeitsgruppen um einen Faktor von mindestens neun. Außerdem haben wir erstmals für Mammatumore eine vollständige Annotation vom gesamten WSIs vorgenommen. Unser vorgeschlagenes Deep learning-basierte Modell hat einen F_1 -Wert von 0,791 auf dem Testdatensatz erreicht. Eine Anwendung unseres Modells auf einen veröffentlichten humanen Mammatumordatensatz (AMIDA13; (VETA *et al.*, 2015)) brachte einen F_1 -Wert von 0,696. Dieser Datensatz ist ein erster Schritt für akkurate Mitosealgorithmen für kanine Mammatumore und ist eine sinnvolle Ergänzung für humane Datensätze, welche nur vorausgewählte Tumorregionen abdecken.

3.6 Untersuchung zur Variabilität der Datensatzentwicklung

- Titel:** Are Pathologist-Defined Labels Reproducible? Comparison of the TUPAC16 Mitotic Figure Dataset with an Alternative Set of Labels
- Autoren:** Bertram CA, Veta M, Marzahl C, Stathonikos N, Maier A, Klopffleisch R, Aubreville M
- Jahr:** 2020
- Zeitschrift:** Proceedings des Labels workshop der MICCAI conference; LABELS 2020 MICCAI Workshop on Large-scale Annotations of Biomedical data and Expert Label Synthesis, 08.10.2020. Originalartikel in J. Cardoso et al. (Eds.): iMIMIC 2020/MIL3iD 2020/LABELS 2020, LNCS 12446, pp. 204–213, 2020, Springer Nature Switzerland AG 2020
- DOI:** https://doi.org/10.1007/978-3-030-61166-8_22

Wissenschaftlicher Beitrag: Gemeinsam mit Prof. Dr. Marc Aubreville habe ich die Planung, Umsetzung und Auswertung der Studie sowie Erstellung des Manuskriptes durchgeführt. Prof. Dr. Mitko Veta, Prof. Dr. Robert Klopffleisch, Prof. Dr. Andreas Maier, Christian Marzahl und Nikolas Stathonikos haben bei der Planung und Auswertung geholfen. Ich habe als erster Annotator, Prof. Dr. Robert Klopffleisch als zweiter Annotator fungiert. Prof. Dr. Marc Aubreville hat bei der algorithmisch-assistierten Datenbankerstellung geholfen und war für das Training und Testen der Modelle verantwortlich. Alle Autoren haben sich an der Korrektur des Manuskriptes beteiligt.

Wissenschaftliche Erkenntnisse: Unterschiedliche Arbeitsgruppen haben öffentlich zugängliche Mitosedatensätze von humanen Mammatumoren erstellt, welche häufig gegenseitig zum Training von Modellen und Testen der Leistungsfähigkeit (Robustheit) verwendet werden. Es war jedoch unbekannt, wie stark sich die unterschiedlichen Labelmethoden auf die algorithmische Leistungsfähigkeit auswirken kann. Wir haben daher zu den Bildern von dem öffentlich zugänglichen TUPAC16 Datensatz eine neue Mitosedatenbank mit der eigenen etablierten Labelingmethode erstellt. Wir konnten feststellen, dass sich die Anzahl und Objekte der Annotationen deutlich unterscheiden. Der verwendete algorithmisch-assistierte Labelansatz scheint die Sensitivität des Annotierens entscheidend erhöht zu haben. Während mit dem originalen Datensatz ein F_1 -Wert von 0,587 erzielt wurde, konnte mit dem alternativen Datensatz ein F_1 -Wert von 0,735 erreicht werden. Die Variabilität zwischen Labelansätzen hat einen hochrelevanten Einfluss auf das Training und Evaluieren von datenbasierten Modellen.

3.7 Untersuchung zur Datenbankqualität von Multi-Experten-Konsensus

Titel: Inter-observer Variability in Automatic Mitotic Figure Assessment – How Many Annotators Do We Need?

Autoren: Wilm F*, Bertram CA*, Marzahl C, Bartel A, Donovan TA, Assenmacher C-A, Becker K, Bennett M, Corner S, Cossic B, Denk D, Dettwiler M, Garcia Gonzalez B, Gurtner C, Heier A, Lehmbecker A, Merz S, Plog S, Schmidt A, Sebastian F, Smedley RC, Tecilla M, Thaiwong T, Breininger K, Kiupel M, Maier A, Klopffleisch R und Aubreville M * im gleichen Maß beigetragen

Jahr: 2021

Veröffentlichung: Proceedings, Bildverarbeitung für die Medizin 2021

DOI: https://doi.org/10.1007/978-3-658-33198-6_56

Wissenschaftlicher Beitrag: Frauke Wilm und ich haben im gleichen Maß zu der Planung, Umsetzung und Auswertung der Studie und Erstellung des Manuskriptes beigetragen. Prof. Dr. Marc Aubreville, Prof. Dr. Robert Klopffleisch, Prof. Dr. Andreas Maier, Prof. Dr. Katharina Breininger, Prof. Dr. Matti Kiupel, Christian Marzahl und Alexander Bartel haben bei der Planung und/oder Auswertung geholfen. Datenbanken wurden von Taryn A. Donovan, Charles-Antoine Assenmacher, Kathrin Becker, Mark Bennett, Sarah Corner, Brieuc Cossic, Daniela Denk, Martina Dettwiler, Beatrice Garcia Gonzalez, Corinne Gurtner, Annabelle Heier, Annika Lehmbecker, Sophie Merz, Stephanie Plog, Anja Schmidt, Franziska Sebastian, Rebecca R Smedley, Marco Tecilla und Tuddow Thaiwong erstellt. Ich habe die Erstellung der Datenbanken koordiniert und Frauke Wilm war für das Training und Testen der Modelle verantwortlich. Alle Autoren haben sich an der Korrektur des Manuskriptes beteiligt.

Wissenschaftliche Erkenntnisse: Daten-basierte Algorithmen für die Erkennung von Mitosefiguren in histologischen Schnitten haben ein sehr hohes Forschungsinteresse. Eine entscheidende Limitation von aktuellen Datensätzen ist die ausgeprägte inter- und intra-observer Variabilität der Annotationen/innen von Mitosefiguren, weshalb in der Regel auf einen Konsensus von mehreren Experten/innen zurückgegriffen wird. In dieser Studie haben wir die notwendige Anzahl an Annotatoren/innen für eine optimale Leistung des trainierten Algorithmus untersucht. Während Modelle, welche mit einzelnen Datenbanken trainiert wurden, eine sehr hohe Schwankung der Leistung aufwiesen, konnte ab einer Anzahl von drei Annotatoren/innen eine konstant hohe Leistung erzielt werden. Diese Erkenntnis trägt maßgeblich zur Erstellung von qualitativ hochwertigen Datensätzen bei.

3.8 Datensatz über zwei- und mehrkernige Tumorzellen in kaninen Mastzelltumoren

Titel: Dataset on Bi- and Multi-Nucleated Tumor Cells in Canine Cutaneous Mast Cell Tumors

Autoren: Bertram CA, Donovan TA, Tecilla M, Bartenschlager F, Fragoso M, Wilm F, Marzahl C, Breininger K, Maier A, Klopfeisch R und Aubreville M

Jahr: 2021

Veröffentlichung: Proceedings, Bildverarbeitung für die Medizin 2021

DOI: https://doi.org/10.1007/978-3-658-33198-6_33

Wissenschaftlicher Beitrag: Gemeinsam mit Prof. Dr. Marc Aubreville habe ich die Planung, Umsetzung und Auswertung der Studie durchgeführt. Prof. Dr. Robert Klopfeisch, Prof. Dr. Andreas Maier, Prof. Dr. Katharina Breininger, Christian Marzahl und Frauke Wilm haben bei der Planung und Auswertung geholfen. Der Datensatz wurde von mir erstellt. Prof. Dr. Marc Aubreville hat bei der algorithmisch-assistierte Datenbankerstellung geholfen und war für das Training und Testen der Modelle verantwortlich. Das Manuskript wurden von mir angefertigt. Dr. Taryn A. Donovan, Marco Tecilla, Florian Bartenschlager, Marco Fragoso und Prof. Dr. Robert Klopfeisch waren Studienteilnehmer/innen. Alle Autoren haben sich an der Korrektur des Manuskriptes beteiligt.

Wissenschaftliche Erkenntnisse: Die Polyploidie (das Vorkommen von mehr als zwei Chromosomensätzen pro Zelle) von Tumorzellen wird als relevantes Kriterium der Tumormalignität angesehen. Tumorzellen mit mehr als einem Zellkern weisen eine Polyploidie auf. Mit dieser Studie haben wir den ersten öffentlich zugänglichen histologischen Datensatz für zwei- und mehrkernige Tumorzellen erstellt, welches eine Grundvoraussetzung für die Entwicklung von daten-basierte Algorithmen darstellt. In 32 WSIs von kaninen Mastzelltumoren wurden 19.983 zweikernige und 1.416 mehrkernige Tumorzellen annotiert. Unser Deep learning-basiertes Modell hatte einen F_1 -Wert von 0,675 für zweikernige Tumorzellen und einen F_1 -Wert von 0,623 für mehrkernige Tumorzellen auf dem Testdatensatz erzielt. Auf Bildausschnitten des Testdatensatzes konnte der Algorithmus die Leistung aller Pathologen/innen übertreffen. Wir konnten zeigen, dass dieses Prognosekriterium eine hohe inter-observer Variabilität aufweist und Deep learning-basierte Algorithmen ein vielversprechendes Hilfsmittel darstellen. Der öffentlich zugängliche Datensatz ermöglicht die Erforschung von optimalen Netzwerkarchitekturen.

4 Diskussion

Jüngste Entwicklungen auf dem Gebiet der digitalen Pathologie ermöglichen einen breiten Einsatz dieser Technologie für die Diagnostik, Lehre und Forschung (BERTRAM und KLOPFLEISCH, 2017). Allerdings liegen momentan nur wenige wissenschaftliche Erkenntnisse über die diagnostische Sicherheit der digitalen Mikroskopie, dem Meinungsbild von Studierenden sowie dem Potential und Anforderungen von daten-basierten Algorithmen vor. Die eigenen Untersuchungen umfassen daher ein breites Anwendungsspektrum der digitalen Mikroskopie und automatisierten Bildanalyse für die Veterinärpathologie und -zytologie. In Übersichtsartikeln haben wir den aktuellen Stand der digitalen Pathologie für die Tiermedizin zusammengefasst (BERTRAM und KLOPFLEISCH, 2017; KLOPFLEISCH und BERTRAM, 2018; DONOVAN *et al.*, 2020).

Die ersten Studien des PhD-Projektes befassten sich mit der digitalen Mikroskopie. Wir konnten in der ersten akademischen Validierungsstudie für die Veterinärpathologie zeigen, dass die digitale Mikroskopie weitestgehend eine adäquate Modalität für die histologische Untersuchung von kaninen Hauttumoren ist (BERTRAM *et al.*, 2018c). Während eine aktuelle Richtlinie von PANTANOWITZ *et al.* (2013) lediglich einen einfachen Vergleich zwischen der digitalen Mikroskopie und Lichtmikroskopie fordert, haben wir uns für ein aufwendigeres Studiendesign (Nichtunterlegenheitsstudie) mit Bestimmung einer tatsächlichen Bezugsgröße für die Leistungsfähigkeit der Lichtmikroskopie entschieden (siehe Kapitel 4.1). Nachfolgend konnten wir in einer Meinungsumfrage von Studierenden ein vorwiegend sehr positives Meinungsbild über die digitale Mikroskopie als Bestandteil des Histopathologie-Curriculums ermitteln (BERTRAM *et al.*, 2018b). Dabei hatte das Erfahrungslevel der Studierenden (gemessen anhand der Jahrgangsstufen) einen erkennbaren Einfluss auf die wahrgenommenen Nachteile der digitalen Mikroskopie (siehe Kapitel 4.2).

In darauffolgenden Projekten haben wir ausführlich das Potential und die Herausforderungen von Deep learning-basierten Algorithmen für die histopathologische und zytologische Diagnostik erforscht (siehe Kapitel 4.3). Die erste Fragestellung war das Erkennen von Mitosefiguren in histologischen Präparaten von Tumoren des Hundes (AUBREVILLE *et al.*, 2017; AUBREVILLE *et al.*, 2019a; AUBREVILLE *et al.*, 2019b; BERTRAM *et al.*, 2019b; AUBREVILLE *et al.*, 2020a; AUBREVILLE *et al.*, 2020b; AUBREVILLE *et al.*, 2020c; BERTRAM *et al.*, 2021; WILM *et al.*, 2021). Eine deutliche Limitation vorheriger Datensätze von humanen Mammakarzinomen ist die Beschränkung der Annotationen auf kleine vorausgewählte Tumorregionen mit unzureichender Repräsentierbarkeit für den gesamten WSI, welche wir mit unseren eigenen Datensätzen erstmals überwinden konnten. Daraus resultierend konnten wir sehr hohe F_1 -Werte für die Analyse vom gesamten WSIs erreichen

und erstmals eine algorithmische Regionauswahl mit der höchsten Mitosedichte ermöglichen (siehe Kapitel 4.3.1). Mittels computerisierten Analysen der Mitoseverteilung in den eigenen Datensätze konnten wir auch zur Grundlagenforschung der Tumorproliferation beitragen (AUBREVILLE *et al.*, 2020a; BERTRAM *et al.*, 2020a; BERTRAM *et al.*, 2021). Weiterhin konnten wir den ersten öffentlich zugänglichen histologischen Datensatz und einen ersten Deep learning-basierten Algorithmus für zwei- und mehrkernige Tumorzellen erstellen (BERTRAM *et al.*, 2021). Die Ergebnisse sind sehr vielversprechend für eine verbesserte Standardisierung dieses prognostischen Parameters mittels Deep learning-basierten Algorithmen (siehe Kapitel 4.3.2). Auf dem Gebiet der Zytologie wurden erstmals Datensätze und Deep learning-basierte Bildanalysesoftware von equinen Lungenspülproben (bronchoalveoläre Lavage) für die Diagnose von equinen Lungenbluten (exercise-induced pulmonary hemorrhage) (MARZ AHL *et al.*, 2020c) und entzündliche Veränderungen der tiefen Atemwege (MARZ AHL *et al.*, 2021) erstellt. In beiden Fällen konnten wir mit großen Datensätzen genaue Algorithmen entwickeln mit einer vergleichbaren Leistungsfähigkeit wie sie erfahrene Pathologen/innen erreicht haben (siehe Kapitel 4.3.3). Ein Forschungsschwerpunkt für die oben erwähnten Projekte war stets die Frage, wie die Qualität und Quantität von Datensätzen effizient optimiert werden kann (AUBREVILLE *et al.*, 2018; BERTRAM *et al.*, 2019b; AUBREVILLE *et al.*, 2020a; AUBREVILLE *et al.*, 2020c; BERTRAM *et al.*, 2020b; MARZ AHL *et al.*, 2020a; MARZ AHL *et al.*, 2020b; MARZ AHL *et al.*, 2020d; BERTRAM *et al.*, 2021; WILM *et al.*, 2021). Computer-assistierte Labelingmethoden wurden als eine sinnvolle Ergänzung zu sorgfältigen experten-basierten Annotationen in repräsentativen Bildern erkannt (siehe Kapitel 4.3.4).

4.1 Erkenntnisse zur diagnostischen Leistungsfähigkeit der digitalen Mikroskopie

Verschiedene tiermedizinischen Labore verwenden die digitale Mikroskopie bereits ergänzend oder ausschließlich zur Diagnose von Gewebeproben. Die Vorteile der digitalen Mikroskopie sind hinreichend bekannt (siehe Literaturübersicht), jedoch ist die diagnostische Leistungsfähigkeit dieser Untersuchungsmodalität für die Tiermedizin bisher unzureichend untersucht (siehe Kapitel 4.1.2). In der Humanmedizin liegen deutlich umfangreichere Erkenntnisse aus akademischen Validierungsstudien vor (siehe Kapitel 4.1.3). Allerdings unterscheidet sich das Studiendesign von diesen Veröffentlichungen erheblich und bringt verschiedene Vor- und Nachteile mit sich, wie nachfolgend im Kapitel 4.1.1 zusammengefasst wird. Viele akademische Studien orientieren sich an den Richtlinien von PANTANOWITZ *et al.* (2013), welche jedoch vor allem die Minimalanforderungen für klinische Validierungsstudien abdecken. Weitere Richtlinien für die Validierung der digitalen Mikroskopie in der Tiermedizin sind unbedingt erforderlich. Daher werden in Kapitel 4.1.1.4 relevanten Erkenntnisse für die Durchführung von (zukünftigen) Validierungsstudien, welche aus der eigenen Untersuchung gewonnen wurden, aufgeführt.

4.1.1 Methoden für die Validierung der digitalen Mikroskopie

Das Ziel einer Validierungsstudie ist es sicherzustellen, dass die Verwendung der digitalen Mikroskopie keinen negativen Einfluss auf die diagnostische Leistung (gemessen an der gestellten Diagnose) von Pathologen/innen hat. Es wird in der Regel das gesamte digitale Pathologie-System (inklusive WSI-Scanner, Netzwerk, computerisierter Arbeitsplatz) und dessen Rahmenbedingungen (zum Beispiel Scanauflösung der WSIs) als eine Einheit gleichzeitig validiert (PANTANOWITZ *et al.*, 2013). Dafür werden die gleichen Proben vom den gleichen Untersuchern/innen mindestens einmal mittels Lichtmikroskopie („Goldstandard“) und einmal mittels digitaler Mikroskopie (Testmodalität) befundet (PANTANOWITZ *et al.*, 2013). Die relevanteste Messgröße einer Validierungsstudie ist der (prozentuale) Anteil der vollkommen oder überwiegend übereinstimmenden Diagnosen (Konkordanzrate) zwischen den beiden Untersuchungsmodalitäten (PANTANOWITZ *et al.*, 2013). In einer Validierungsstudie wird also die Angemessenheit der technischen Gegebenheiten („der Weg des Pixels“) für die Diagnosestellung im Vergleich zur herkömmlichen Methode (der Lichtmikroskopie) überprüft. Die diagnostischen Fähigkeiten des/der Pathologen/in und die generelle Komplexität der Gewebeprobe stellen hingegen Verzerrungseffekte dar, da diese nicht nur einen Einfluss auf die digitale Mikroskopie, sondern auch auf die Lichtmikroskopie haben (BERTRAM *et al.*, 2018c; BERTRAM und KLOPFESCH, 2019). Auch die Diagnose

eines/r sehr erfahrenen Pathologen/in hat weniger als eine 100%ige Reproduzierbarkeit aufgrund von verschiedenen visuellen und kognitiven Limitationen (AEFFNER *et al.*, 2017). Allerdings ist es – je nach Studiendesign – nicht immer einfach das Ausmaß dieses Verzerrungseffekts einzuschätzen. Wie bereits von AZAM *et al.* (2020) festgestellt wurde, unterscheidet sich das Studiendesign der veröffentlichten Validierungsstudien in vielen Aspekten erheblich. Nachfolgend sollen die Vor- und Nachteile – insbesondere in Hinsicht der Dateninterpretation – der verschiedenen Studiendesigns besprochen werden. Dafür wurde eine grobe Einteilung der veröffentlichten Studien in drei Studiendesigns – nach Kriterien die erstmals für einen eigenen Kongressbeitrag erarbeitet wurden (BERTRAM und KLOPFESCH, 2019) – vorgenommen: Konkordanz-, Genauigkeits- und Nichtunterlegenheitsstudie. Die Einteilung basiert auf der angewendeten Methodik, den Verzerrungseffekt einzuschätzen beziehungsweise zu reduzieren. Verbunden mit dem Studiendesign ändert sich jedoch nicht nur die Interpretierbarkeit der Daten, sondern auch die Studienhypothese und der Zeitaufwand (siehe nachfolgende Unterkapitel). Während für einfache Konkordanzstudie der Verzerrungseffekt nur indirekt eingeschätzt werden kann, stehen für Genauigkeitsstudien und Nichtunterlegenheitsstudien ein Vergleichswert beziehungsweise ein Bezugswert zur Verfügung (siehe nachfolgende Kapitel). Allerdings ist der Zeitaufwand für einfache Konkordanzstudien am geringsten, sodass ein Großteil der Validierungsstudien dieses Studiendesign verwendet haben. Für einfache Konkordanzstudien sind lediglich zwei Untersuchungszeiträume notwendig, wohingegen Genauigkeitsstudien eine weitere Goldstandarddiagnose und Nichtunterlegenheitsstudien ein oder idealerweise zwei weitere Untersuchungszeitpunkte benötigen (Abbildung 11).

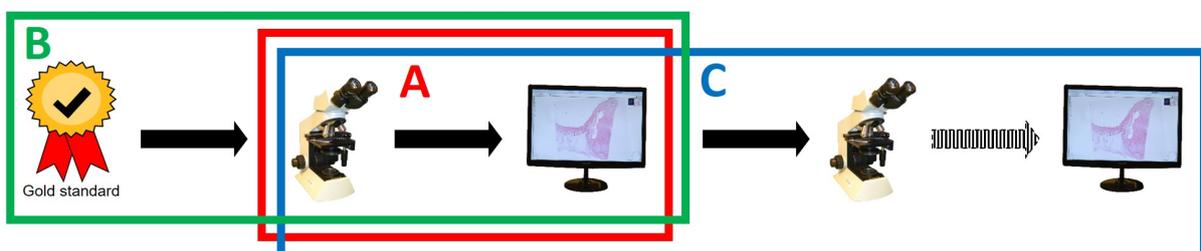


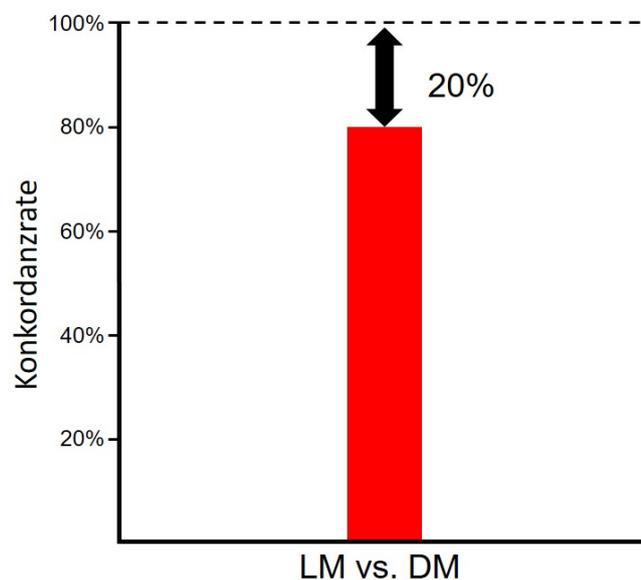
Abbildung 11: Vergleich der Untersuchungsmodalitäten und -Zeitpunkte für eine Validierung der digitalen Mikroskopie mittels einfacher Konkordanzstudie (A; mittlerer, roter Kasten), Genauigkeitsstudie (B; linker, grüner Kasten) und Nichtunterlegenheitsstudie (C; rechter, blauer Kasten)

4.1.1.1 Einfache Konkordanzstudie

Dieses Studiendesign wurde von den meisten Validierungsstudien gewählt (AL-JANABI *et al.*, 2012a; AL-JANABI *et al.*, 2012b; AL-JANABI *et al.*, 2012c; AL-JANABI *et al.*, 2012d; AL-JANABI *et al.*, 2013; AL-JANABI *et al.*, 2014; THRALL *et al.*, 2015; HANNA *et al.*, 2019; HANNA *et al.*, 2020). Im Vergleich zu den anderen beiden Studiendesigns (siehe Kapitel 4.1.1.2 und 4.1.1.3) ist der Zeitaufwand deutlich geringer, da die Proben nur einmal mittels Lichtmikroskopie und einmal mittels digitaler Mikroskopie untersucht werden müssen (Abbildung 11). Das Ergebnis ist eine einfache Konkordanzrate zwischen den beiden Untersuchungsmodalitäten (Abbildung 12). Neben dem möglichen Einfluss der digitalen Modalität, wird die einfache Konkordanzrate stark von den diagnostischen Fähigkeiten des/der Pathologen/innen und der Komplexität der Fragestellung beeinflusst (BERTRAM *et al.*, 2018c). Dementsprechend ist eine perfekte (100%) Konkordanzrate nicht zu erwarten (PANTANOWITZ *et al.*, 2013). Wissenschaftlich fundierte Grenzwerte liegen nicht vor und sind auch nicht unbedingt zwischen verschiedenen Studien (mit unterschiedlichen Studienpathologen/in und Probenkomplexität) vergleichbar. PANTANOWITZ *et al.* (2013) empfehlen eine minimal akzeptable Konkordanzrate für jede Studie einzeln anhand eines „guten medizinischen Urteilsvermögens“ der Studienorganisatoren anzusetzen. Nach eigener Erfahrung ist es jedoch äußerst schwierig den Verzerrungseffekt (siehe Kapitel 4.1.1) einzuschätzen und daher weist dieses Studiendesign eine deutliche Limitation bei der Dateninterpretation auf. Die meisten Studien versuchen daher den Einfluss der Untersuchungsmodalität besser zu bewerten, indem die vermuteten Fehlerquellen von den abweichenden Diagnosepaaren nachträglich bestimmt werden (AL-JANABI *et al.*, 2012c; AL-JANABI *et al.*, 2012d; AL-JANABI *et al.*, 2014; SNEAD *et al.*, 2016).

Abbildung 12: Darstellung beispielhafter Ergebnisse einer einfachen Konkordanzstudie

Abkürzungen: LM, Lichtmikroskopie; DM, digitale Mikroskopie; vs., versus

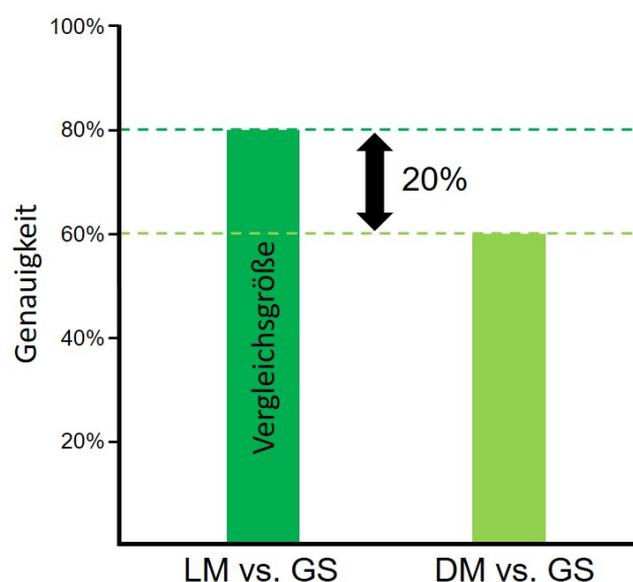


4.1.1.2 Genauigkeitsstudie

Das Ziel einer Genauigkeitsstudie ist die Überprüfung der Gleichwertigkeit (oder Überlegenheit) der digitalen Mikroskopie im Vergleich zur Lichtmikroskopie als Vergleichsgröße. Einige Studien haben dieses Studiendesign gewählt (NIELSEN *et al.*, 2010; CAMPBELL *et al.*, 2014; ELMORE *et al.*, 2017; KENT *et al.*, 2017; BONSEMBIANTE *et al.*, 2018; MILLS *et al.*, 2018; MUKHOPADHYAY *et al.*, 2018; BONSEMBIANTE *et al.*, 2019). Die Diagnosen mit beiden Untersuchungsmodalitäten (Licht- und digitale Mikroskopie) werden unabhängig voneinander mit einer Goldstandarddiagnose verglichen. Anstelle der bzw. zusätzlich zur Konkordanz wird so die Fehlerfreiheit der Diagnosen zu den „richtigen“ Diagnose bestimmt. Die Genauigkeitswerte für die beiden Untersuchungsmodalitäten können leicht verglichen werden (Abbildung 13) und die Signifikanz des Unterschiedes kann statistisch (zum Beispiel mittels eines exakten Tests nach Fisher) bestimmt werden (KENT *et al.*, 2017; BONSEMBIANTE *et al.*, 2018; BONSEMBIANTE *et al.*, 2019). Die Limitation dieses Studiendesigns ist die Schwierigkeit eine echte Goldstandarddiagnose für mikroskopische Präparate zu erstellen. Lediglich BONSEMBIANTE *et al.* (2018) und (2019) konnten für zytologische Präparate eine höherwertige Grundwahrheit mittels Histologie oder Durchflusszytometrie erstellen. Andere Validierungsstudien für histologische Präparate greifen auf die lichtmikroskopischen Diagnosen vom ursprünglichen Untersuchungsbefund (CAMPBELL *et al.*, 2014; MILLS *et al.*, 2018; MUKHOPADHYAY *et al.*, 2018), von einem/r erfahrenen Pathologen/in (NIELSEN *et al.*, 2010) oder einer Expertengruppe (ELMORE *et al.*, 2017; KENT *et al.*, 2017) zurück, welches eine Verzerrung für die Datenauswertung bedingen kann.

Abbildung 13: Darstellung beispielhafter Ergebnisse einer Genauigkeitsstudie

Abkürzungen: LM, Lichtmikroskopie; GS, Goldstandarduntersuchung; DM, digitale Mikroskopie; vs., versus

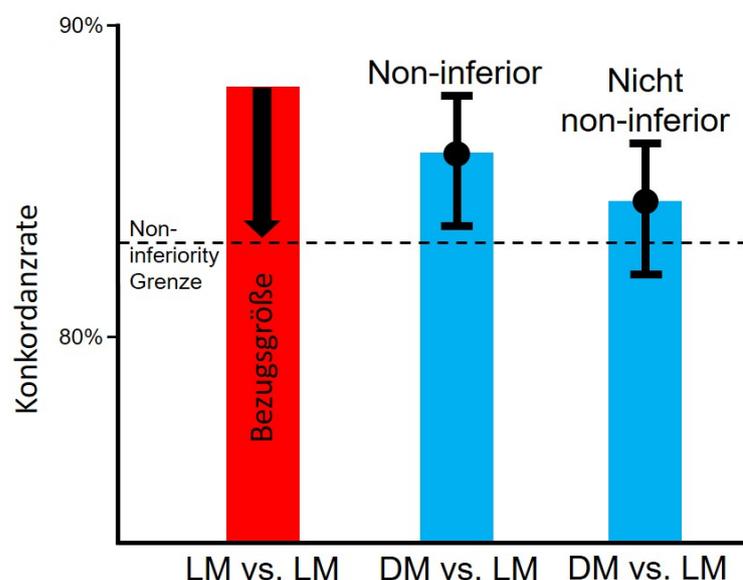


4.1.1.3 Nichtunterlegenheitsstudie

Nichtunterlegenheitsstudien haben das Ziel eine Nichtunterlegenheit (non-inferiority) der digitalen Mikroskopie im Vergleich zur Lichtmikroskopie als Bezugsgröße nachzuweisen. Dafür werden die Proben ein weiteres Mal mittels Lichtmikroskopie und idealerweise ein weiteres Mal mittels digitaler Mikroskopie untersucht (Abbildung 11). Dadurch kann eine Konkordanzrate für zwei Untersuchungszeitpunkte mittels Lichtmikroskopie bestimmt werden, welche die intra-observer Reproduzierbarkeit der Referenzmodalität darstellt. Dieser Wert kann als Bezugsgröße (Nicht-unterlegenheitsgrenze nach Abzug eines akzeptablen Toleranzbereiches) für die Bestimmung der Leistungsfähigkeit der digitalen Mikroskopie genutzt werden (Abbildung 14). Ein einseitiger Binominaltest kann angewendet werden, um zu bestimmen, ob die untere Grenze des 95%-Konfidenzintervalls der Testmodalität innerhalb der Nichtunterlegenheitsgrenze liegt (BAUER *et al.*, 2013; BERTRAM *et al.*, 2018c). Wenn ein zweiter Untersuchungszeitpunkt mittels digitaler Mikroskopie erfasst wird, kann die Lernkurve über den Studienverlauf (früherer versus späterer Untersuchungszeitpunkt) für beide Modalitäten bestimmt werden. Das Nichtunterlegenheits-studiendesign wurde für wenigen veröffentlichten Validierungsstudien verwendet (OZLUK *et al.*, 2012; BAUER *et al.*, 2013; SHAH *et al.*, 2016). Auch in der eigenen Validierungsstudie haben wir uns für dieses Studiendesign entschieden aufgrund der sehr hohen Interpretierbarkeit der Ergebnisse (BERTRAM *et al.*, 2018c). Limitationen dieses Studiendesigns sind der deutlich höhere Zeitaufwand aufgrund der abermaligen (idealerweise viermalige) Untersuchungen der Proben. Wegen der mehrfachen Wiederholungen liegt auch eine höhere Erinnerungsverzerrung vor, welche mit ausreichend langen Zeitabständen zwischen den Untersuchungszeiträumen und alternierenden Probenreihenfolgen reduziert werden sollte (BERTRAM *et al.*, 2018c).

Abbildung 14: Darstellung beispielhafter Ergebnisse einer Nichtunterlegenheitsstudie

Abkürzungen: LM, Lichtmikroskopie; DM, digitale Mikroskopie; vs., versus



4.1.1.4 Erkenntnisgewinn für zukünftige Validierungsstudien

Die veröffentlichten Validierungsstudien weisen ein sehr vielfältiges Studiendesign auf, welche teilweise mit deutlichen Limitationen behaftet sind. Die größte Limitation vieler Studien ist die ausschließliche Berechnung einer einfachen Konkordanzrate zwischen Licht- und digitaler Mikroskopie. In der eigenen Untersuchung konnten wir feststellen, dass sich die Konkordanzrate unabhängig von der Untersuchungsmodalität zwischen verschiedenen Studienpathologen/innen und zwischen verschiedenen Probenspektren (epitheliale Tumore und Spindelzelltumore versus Rundzelltumore) erheblich unterscheidet (BERTRAM *et al.*, 2018c). Ein allgemeingültiger minimaler Grenzwert für die Konkordanzrate unabhängig vom individuellen Pathologen/in, Probenspektrum und individuellem digitalem Pathologiesystem ist daher aus unserer Sicht undenkbar. Das Nichtunterlegenheitsstudiendesign hat sich als sehr hilfreich erwiesen die diagnostische Leistungsfähigkeit der digitalen Mikroskopie unter Berücksichtigung der Verzerrungseffekte zu bestimmen. Die Studien von BONSEMBIANTE *et al.* (2018) und (2019) haben gezeigt, dass die Genauigkeitsstudie ein ideales Studiendesign für zytologische Fragestellungen darstellen kann; jedoch ist das Studiendesign für histologische Fragestellungen in der Regel weniger geeignet aufgrund des Fehlens einer echten Goldstandarddiagnose.

Weiterhin konnten wir feststellen, dass die Konkordanzrate und diagnostische Geschwindigkeit bei manchen Pathologen/innen für die digitale Mikroskopie und bei anderen Pathologen/innen für die Lichtmikroskopie höher war (BERTRAM *et al.*, 2018c), welches sich aus persönlichen Präferenzen oder Erfahrung mit der jeweiligen Untersuchungsmodalität ergeben könnte. Zur Berücksichtigung dieses Effektes erscheint es daher empfehlenswert mehrere Studienteilnehmer/innen einzuschließen. Sollte eine geringe Vorerfahrung mit der digitalen Mikroskopie bestehen, ist ein vorheriges Training der Pathologen/innen sinnvoll.

Ebenfalls konnten wir, wie auch in anderen Studien, eine deutliche Steigerung der diagnostischen Leistungsfähigkeit (Lernkurve) zwischen frühen und späten Untersuchungszeitpunkten feststellen (KRISHNAMURTHY *et al.*, 2013; ORDI *et al.*, 2015; BERTRAM *et al.*, 2018c; MILLS *et al.*, 2018). Dies könnte mit einem Trainingseffekt (zunehmende Erfahrung mit der digitalen Mikroskopie), einem Gewöhnungseffekt an das Studiendesign oder/und einem Erinnerungseffekt an die inkludierten Proben zusammenhängen. Insbesondere bei Studiendesigns mit mehreren Untersuchungswiederholungen (Nichtunterlegenheitsstudie) sollte daher der Studienablauf sorgfältig überdacht werden. Dies betrifft vor allem die Untersuchungsreihenfolge der Proben und den Zeitabstand zwischen den Untersuchungszeitpunkten. PANTANOWITZ *et al.* (2013) empfehlen einen minimalen Zeitabstand für einfache Konkordanzstudien von mindestens zwei Wochen. Allerdings konnten CAMPBELL *et al.* (2015) feststellen, dass Pathologen/innen bis

zu 40% der Proben über einen Zeitraum von zwei Wochen wiedererkennen können. Der ideale Zeitabstand für Nichtunterlegenheitsstudien mit drei beziehungsweise idealerweise vier Untersuchungszeitpunkten wurde bisher nicht untersucht.

Eine Schwierigkeit, die beim Vergleich der Diagnosepaare aufgetreten ist, war die Entscheidung, ob die Diagnosen übereinstimmend zueinander sind. Pathologische Diagnosen bestehen aus verschiedenen deskriptiven Elementen und variablen, teilweise überlappenden Terminologien. Eindeutige und konsistente Entscheidungskriterien für eine (vollständige) Übereinstimmung sind daher nicht in jedem Fall einfach. Daher ist es möglicherweise empfehlenswert, vor der Studiendurchführung die Diagnosemöglichkeiten zu standardisieren und für die Auswertung der Diagnosen vordefinierte Entscheidungskriterien zu verwenden. Bisher haben nur wenige Studien standardisierte Diagnosekriterien, vereinheitlichte Terminologien und/oder Referenztabellen für synonyme Begriffe berücksichtigt (NIELSEN *et al.*, 2010; KRISHNAMURTHY *et al.*, 2013; WACK *et al.*, 2016; BOROWSKY *et al.*, 2020).

4.1.2 Aktueller Stand der Validierung in der Tiermedizin

Für die Tiermedizin liegen aktuell nur wenige akademische Validierungsstudien vor und zusätzliche Validierungsstudien über weitere diagnostische Fragestellungen sind unbedingt erforderlich. Die erste Validierungsstudie in der Tiermedizin war die eigene Untersuchung aus dem Jahr 2018 (BERTRAM *et al.*, 2018c). Diese Nichtunterlegenheitsstudie konnte für kanine Hauttumore eine ähnliche hohe diagnostische Übereinstimmung mittels digitaler Mikroskopie im Vergleich zur Lichtmikroskopie nachweisen. Lediglich die Diagnose von Rundzelltumoren und das Tumorigradung von Mastzelltumoren hatten eine etwas geringere Konkordanzrate mittels digitaler Mikroskopie. Obwohl ein etwas niedrigeres diagnostisches Vertrauen und empfundene Präparatqualität für die digitale Mikroskopie nachgewiesen wurde, war die Diagnosezeit mit der digitalen Mikroskopie durchschnittlich kürzer. Der Mitotic Count von Rundzelltumoren war vergleichbar zwischen den beiden Untersuchungsmodalitäten.

BONSEMBIANTE *et al.* (2018) und (2019) haben in zwei Genauigkeitsstudien die digitale Mikroskopie und Lichtmikroskopie von zytologischen Präparaten gegen einen echten Goldstandard (histologische oder durchflusszytometrische Diagnose) verglichen. Es wurden keine signifikanten Unterschiede zur zytologischen Diagnose und dem Grading von Lymphomen zwischen den beiden Untersuchungsmodalitäten festgestellt.

WEI *et al.* (2019) untersuchten die diagnostische Übereinstimmung des Mitotic Counts von kaninen Melanomen zwischen der Lichtmikroskopie und der digitalen Mikroskopie (200x und 400x Vergrößerung). Insgesamt wurde eine vergleichbare Leistungsfähigkeit zwischen den beiden Untersuchungsmodalitäten nachgewiesen.

4.1.3 Aktueller Stand der Validierung in Humanmedizin

Der größte Anteil der veröffentlichten Validierungsstudien aus humanmedizinischen Laboren befasst sich mit chirurgischen Gewebeproben. Während die vorhandenen Validierungsstudien bereits ein großes Spektrum an pathologischen Fachgebieten abdecken, liegen für einige Fachgebiete (wie die Hämatopathologie oder ophthalmologische Pathologie) noch unzureichende Erkenntnisse vor (SACO *et al.*, 2016b; ARAÚJO *et al.*, 2019; AZAM *et al.*, 2020). Systematische Übersichtsartikel über 13 bis 38 veröffentlichte Studien zeigen eine akzeptable diagnostische Übereinstimmung (GOACHER *et al.*, 2017; ARAÚJO *et al.*, 2019; AZAM *et al.*, 2020). Obwohl einige aktuelle Validierungsstudien eine sehr große Anzahl an Diagnosepaaren ($N \geq 1000$) untersucht haben (THRALL *et al.*, 2015; SNEAD *et al.*, 2016; TABATA *et al.*, 2017; MUKHOPADHYAY *et al.*, 2018; BOROWSKY *et al.*, 2020), wird die durchschnittlich geringe Probenanzahl der Validierungsstudien als deutliche Limitation angesehen (GOACHER *et al.*, 2017). Weiterhin wird kritisiert, dass die diagnostische Variabilität innerhalb der Lichtmikroskopie zumeist nicht bekannt ist und daher nicht hinreichend bestimmt werden kann, ob die Diskrepanzen tatsächlich durch die digitale Untersuchungsmodalität bedingt waren (AZAM *et al.*, 2020). Derartige Erkenntnisse können nur aus Genauigkeitsstudien oder Nichtunterlegenheitsstudien gewonnen werden; jedoch wurden diese Studiendesigns aufgrund des aufwendigeren Studienablaufs bisher selten durchgeführt. Die meisten Studien bewerten die abweichenden Diagnosepaare mittels einer Expertengruppe, um eine mögliche Fehlerquelle der Diskrepanz nachträglich abschätzen zu können. Dabei wird ein Großteil der Fehler anderen Ursachen als der Untersuchungsmodalität (zum Beispiel generell hohe Schwierigkeit der Probe, geringe Gewebequalität, fehlende klinische Informationen oder Nichtvorhandensein von Spezialfärbungen) zugesprochen (ARAÚJO *et al.*, 2019). Es ist jedoch bemerkenswert, dass in den meisten Fällen der abweichenden Diagnosepaare (85%) die lichtmikroskopische Diagnose und nur selten (13%) die Diagnose mittels digitaler Mikroskopie bevorzugt wurde (WILLIAMS *et al.*, 2017). Es ist daher sehr naheliegend, dass zumindest ein Teil der Diskrepanzen auf den technischen Limitationen oder möglicherweise einer geringen Erfahrung der Pathologen/innen mit der digitalen Mikroskopie beruhen. Die relevanteste Fehlerquelle der digitalen Mikroskopie scheint das sichere Erkennen von kleinen Objekten wie Mitosefiguren, zellulärer Dysplasien, Entzündungszellen und Mikroorganismen zu sein (WILLIAMS *et al.*, 2017; AZAM *et al.*, 2020).

Ein systematischer Übersichtsartikel hat den aktuellen Stand der Validierung für die Zytopathologie untersucht (GIROLAMI *et al.*, 2020). Insgesamt entsprachen 19 Studien den Such- und Inklusionskriterien. Obwohl diese Studien insgesamt eine akzeptable Übereinstimmung zwischen digitaler Mikroskopie und Lichtmikroskopie aufgewiesen haben, schlussfolgern die Autoren, dass aktuell nur eine limitierte Beweislage vorliegt.

4.2 Erkenntnisse zum Meinungsbild der Studierenden über den Einsatz der digitalen Mikroskopie in der Lehre

Die digitale Mikroskopie hat viele Vorteile für Studierende in der Histopathologie-, Histologie- und Zytologie-ausbildung (siehe Literaturübersicht). Eine steigende Anzahl an tierpathologischen Instituten hat Zugang zu einem WSI-Scanner und notwendiger IT-Infrastruktur, welches eine vermehrte Integration dieser vielfältig einsetzbaren Technologie in den Lehrplan ermöglicht. Vorherige Meinungsumfragen von Studierenden der Tier- und Humanmedizin haben ein vorwiegend positives Meinungsbild zur digitalen Mikroskopie als Lehr- und Lerntechnologie ergeben (MILLS *et al.*, 2007; NEEL *et al.*, 2007; SIMS *et al.*, 2007; BRAUN und KEARNS, 2008; PANTANOWITZ *et al.*, 2012; KOGAN *et al.*, 2014; ORDI *et al.*, 2015; RODRIGUES-FERNANDES *et al.*, 2020). In Kapitel 4.1 ist beschrieben, dass die Leistungsfähigkeit der digitalen Mikroskopie in der Diagnostik vermutlich von der Erfahrung der Pathologen/innen mit dieser Untersuchungsmodalität abhängt. Auch für die Lehre ist beschrieben, dass für eine Mikroskopieprüfung mittels digitaler Mikroskopie von Dermatologie-Residents die Vorerfahrung eine wichtige Grundlage für gute Prüfungsergebnisse darstellt hat (BRICK *et al.*, 2013). Inwiefern das Erfahrungslevel der Studierenden mit der digitalen Mikroskopie einen Einfluss auf das Meinungsbild hat, wurde in den vorherigen Meinungsumfragen nicht untersucht.

Daher haben wir in den eigenen Untersuchungen drei separate Studierendengruppen (aus drei Studienjahren) mit einem zunehmendem Level an Vorerfahrung mit der digitalen Mikroskopie befragt (BERTRAM *et al.*, 2018b). Die letzte Gruppe hatte ihre Histopathologieprüfung bereits abgeschlossen und somit das verfügbare digitale Mikroskopie-Angebot für die Prüfungsvorbereitungen umfangreich genutzt. Es ist bekannt, dass Studierende auf das digitale Mikroskopie-Selbststudium vorwiegend erst kurz vor der Prüfung zurückgreifen (PLENDL *et al.*, 2009; PANTANOWITZ *et al.*, 2012; ORDI *et al.*, 2015; SACO *et al.*, 2016a). Unsere Umfrage hat ergeben, dass sich die Meinungsbilder von Studierenden unterschiedlicher Erfahrungslevel vor allem hinsichtlich der wahrgenommenen Nachteile unterschiedlich darstellt (BERTRAM *et al.*, 2018b). Während der dritte und vierte Jahrgang mehrheitlich das unzureichende Training mit Lichtmikroskopen befürchtete, wurde vom fünften Jahrgang mehrheitlich eine unzureichende Ladegeschwindigkeit der WSIs bemängelt. Als größter Vorteil der digitalen Mikroskopie wurde von allen drei Jahrgängen die hohe Flexibilität bezüglich des Ortes und Uhrzeit des Selbststudiums angeführt und die große Mehrzahl der Studierenden gab an, die digitale Mikroskopie vorwiegend für das Selbststudium nach der Vorlesung oder vor der Prüfung zu verwenden. Für eine zufriedenstellende Ladegeschwindigkeit erscheint es daher besonders wichtig, dass eine ausreichende Internetverbindung vor allem am (flexiblen) Lernort des Selbststudiums verfügbar ist.

Die Qualität der WSIs, der Umgang mit der Betrachtungssoftware, der Lernerfolg, den Nutzen für Selbststudium und Vorlesungen und die Nützlichkeit von Markierungen in WSIs wurden von allen drei Jahrgängen ähnlich positiv betrachtet (BERTRAM *et al.*, 2018b). Trotz der Vorteile der digitalen Mikroskopie haben alle drei Jahrgänge eine Verwendung beider Modalitäten mit ungefähr gleichgroßem Anteil gewünscht. Die alleinige Ausbildung mittels Lichtmikroskopie oder digitaler Mikroskopie wurde nur von einer sehr geringen Anzahl an Studierenden bevorzugt. Einer Prüfung mittels digitaler Mikroskopie standen allerdings alle Jahrgänge vorwiegend offen gegenüber.

Die Umfrage hat bestätigt, dass das modernisierte Histopathologie-Curriculum mit digitaler Mikroskopie viele Vorteile für die Studierenden mit sich bringt (BERTRAM *et al.*, 2018b). Nichtsdestotrotz stellt die Lichtmikroskopie aus Sicht der Studierenden aller Jahrgänge weiterhin einen wichtigen Bestandteil des Curriculums dar.

4.3 Erkenntnisse über das Potential und Herausforderungen von Deep learning-basierten Algorithmen

Im Rahmen meines PhD-Projektes wurde gemeinsam mit meinen Forschungspartnern von der Friedrich-Alexander-Universität Erlangen-Nürnberg und der Technischen Hochschule Ingolstadt verschiedene Deep learning-basierte Algorithmen für histologische und zytologische Fragestellungen entwickelt und veröffentlicht. Die Forschungsergebnisse haben folgende Anwendungsgebiete umfasst: 1) Mitoseerkennung in histologischen Tumorschnitten, 2) Erkennung von zwei- und mehrkernigen Tumorzellen in histologischen Tumorschnitten, 3) Quantifizieren von Hämosiderophagen in zytologischen Präparaten von Pferden mit Lungenbluten und 4) Erkennung der verschiedenen Zelltypen in zytologischen Präparaten von Lungenspülproben des Pferdes. In den folgenden Unterkapiteln werden unsere Erfahrungen zu den einzelnen Anwendungsgebieten sowie generelle Erkenntnisse zur Erstellung von hochwertigen Datensätzen diskutiert.

4.3.1 Mitosefiguren in histologischen Tumorpräparaten

Mitosefiguren sind das histologische Erscheinungsbild von teilenden Zellen (Mitosephase des Zellzyklus) in histologischen Präparaten (DONOVAN *et al.*, 2020). Die Anzahl von Mitosefiguren pro Tumorfläche (Mitotic Count) ist ein relevantes Maß für die Tumorpheriferation (growth fraction) (SLEDGE *et al.*, 2016) und stellt somit eines der wichtigsten Prognoseparameter für Tumore, inklusive kaniner Mastzelltumore (SLEDGE *et al.*, 2016; KIUPEL und CAMUS, 2019) und Mammatumore (GOLDSCHMIDT *et al.*, 2011), dar. Gewöhnlich wird der Mitotic Count von einem/r Pathologen/in in zehn aufeinanderfolgenden Gesichtsfeldern (high power fields) in der Tumorregion mit der höchsten Mitosedichte bestimmt (ROMANSIK *et al.*, 2007; ELSTON *et al.*, 2009; KIUPEL *et al.*, 2011; CANADAS *et al.*, 2019; DONOVAN *et al.*, 2020). Allerdings ist aus zahlreichen Studien ersichtlich, dass der Mitotic Count eine hohe inter- und intra-observer Variabilität aufweist (VETA *et al.*, 2015; BERTRAM *et al.*, 2018c; BERTRAM *et al.*, 2020a; MARZAHN *et al.*, 2020d). Mögliche Fehlerquellen des Mitotic Count liegen in jedem der einzelnen Arbeitsschritte, welche wie folgt aufgeteilt werden können: 1) Tumorregionauswahl mit der höchsten Mitosedichte und 2) Zählen der Mitosefiguren innerhalb dieser Region basierend auf 2a) der Identifizierung von möglichen Mitosefiguren und 2b) Klassifizierung von Mitosefiguren versus anderer Zellen mit Verwechslungsgefahr (DONOVAN *et al.*, 2020). Aktuelle Studien haben gezeigt, dass das Bestimmen des Mitotic Counts im WSIs prinzipiell möglich ist (AL-JANABI *et al.*, 2013; BERTRAM *et al.*, 2018c; BALKENHOL *et al.*, 2019; WEI *et al.*, 2019), jedoch je nach Qualität des digitalen Bildes gegebenenfalls ein höherer Schwierigkeitsgrad für die Erkennung von

Mitosefiguren liegen kann (TABATA *et al.*, 2019). Ein sehr vielversprechender Ansatz, um die Reproduzierbarkeit und Genauigkeit des Mitotic Counts zu verbessern, ist die automatisierte Bildanalyse (Abbildung 15) (MATHEW *et al.*, 2020). Aktuelle Veröffentlichungen haben eine potenziell sehr hohe Leistungsfähigkeit von Deep learning-basierten Algorithmen nachgewiesen (VETA *et al.*, 2016; LI *et al.*, 2018; BERTRAM *et al.*, 2019b; AUBREVILLE *et al.*, 2020a; AUBREVILLE *et al.*, 2020c). Es wurde gezeigt, dass diese Algorithmen Mitosefiguren mit der gleichen Leistungsfähigkeit wie Pathologen/innen erkennen können (VETA *et al.*, 2016). Der Vorteil von Algorithmen ist die 100% Reproduzierbarkeit und die Fähigkeit den gesamten WSI innerhalb kurzer Zeit zu analysieren (AUBREVILLE *et al.*, 2020c). Wir konnten zeigen, dass Deep learning-basierte Algorithmen eine Regionauswahl für den Mitotic Count (höchste mitotische Aktivität) mit einer durchschnittlich höheren Leistung trafen als erfahrene Pathologen/innen (AUBREVILLE *et al.*, 2020c).

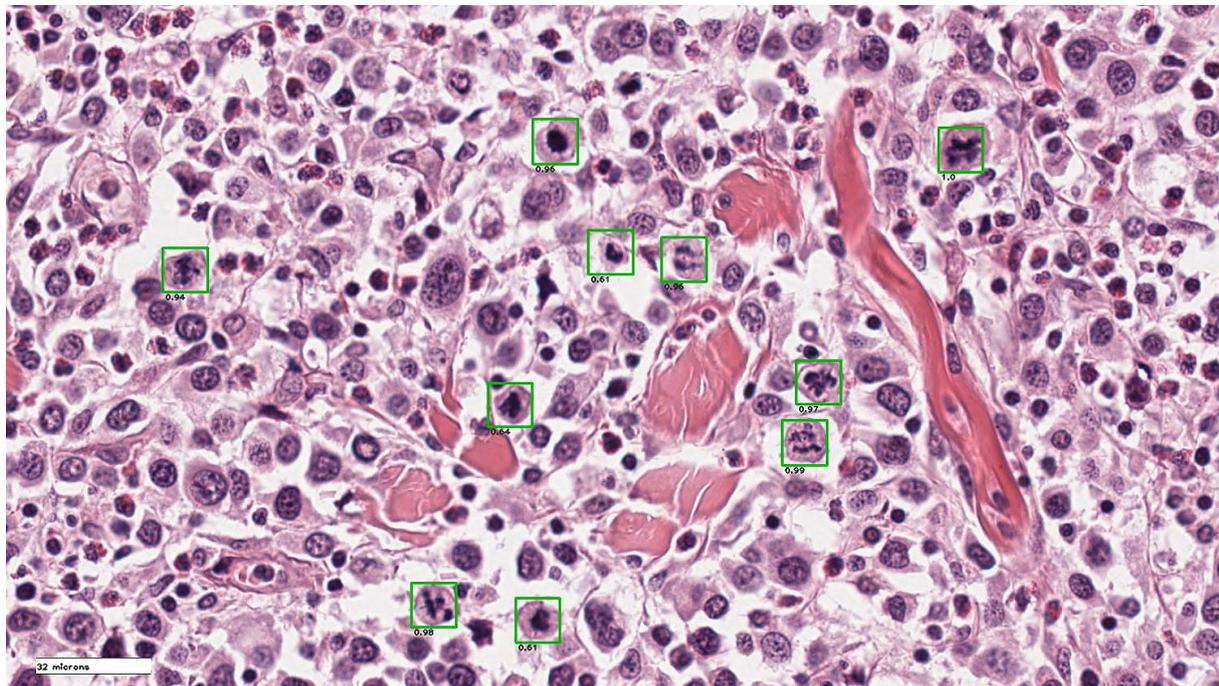


Abbildung 15: Histologischer Bildausschnitt eines kaninen Mastzelltumors mit Deep learning-basierter Detektionen von Mitosen (grüne Kästchen). Detektionen basieren auf dem Algorithmus von BERTRAM *et al.* (2020a)

4.3.1.1 Datensätze und Deep learning-basierte Algorithmen

Mittlerweile stehen einige öffentlich zugängliche Mitosedatensätze über Mitosefiguren von der eigenen und von anderen Arbeitsgruppen zur Verfügung (Tabelle 6). Die Datensätze von anderen Arbeitsgruppen wurden für Mustererkennungs-Challenges bereitgestellt und haben die Entwicklung von algorithmischen Ansätzen enorm vorangetrieben (ROUX *et al.*, 2013; ROUX *et al.*, 2014; VETA *et al.*, 2015; VETA *et al.*, 2019). Die eigenen Datensätze haben erstmals Tumore (Mastzelltumore und Mammatumore) von Hunden verwendet (AUBREVILLE *et al.*, 2020a; BERTRAM *et al.*, 2020a; WILM *et al.*, 2021). Dabei haben wir die Anzahl der Labels vorheriger Datensätze bei Weitem übertroffen und einen großen Aufwand erbracht, um die Qualität (Labelkonsistenz) zu optimieren (siehe Kapitel 4.3.4). Eine Limitation der Datensätze anderer Arbeitsgruppen ist, dass lediglich ausgewählte Tumorregionen annotiert und somit die Variabilität des gesamten Gewebeschnittes nicht abgedeckt wurde (AUBREVILLE, 2020). In einer eigenen Studie konnten wir zeigen, dass das Fehlen der Informationen anderer Geweberegionen zu einer erheblichen Fehlerrate (vor allem falsch positive Detektionen) führen kann, wenn gesamte WSIs analysiert werden (BERTRAM *et al.*, 2019b; AUBREVILLE *et al.*, 2020c). Daher weisen die eigenen Datensätze mit vollständig annotierten WSIs ein bisher unerreichtes Potential für eine hohe algorithmische Leistungsfähigkeit auf gesamten WSIs auf. In verschiedenen Studien können wir zeigen, wie stark der Einfluss der Datenbankqualität auf die Leistungsfähigkeit der Algorithmen ist (BERTRAM *et al.*, 2020b; MARZAHN *et al.*, 2020d; WILM *et al.*, 2021). Weitere Studien zu konsistenteren Labelmethoden sind unbedingt erforderlich um die Datenbankqualität weiter zu verbessern. TELLEZ *et al.* (2018) haben beispielsweise einen sehr interessanten Ansatz entwickelt, bei dem ein spezifischer immunhistochemischer Marker für Mitosefiguren (Phosphohiston H3) zum Einsatz gekommen ist. Dieser verbesserte Goldstandard hat ein genaueres Annotieren ermöglicht und die Labels konnten mittels Bildregistrierung auf histologische WSIs (H&E-Färbung) übertragen werden (Abbildung 16).

Die Mustererkennungs-Challenges und die öffentliche Verfügbarkeit der Datensätze haben eine Vielzahl von algorithmischen Ansätzen hervorgebracht (ROUX *et al.*, 2013; ROUX *et al.*, 2014; VETA *et al.*, 2015; AUBREVILLE *et al.*, 2017; KRAPPMANN *et al.*, 2018; LI *et al.*, 2018; LI *et al.*, 2019; VETA *et al.*, 2019; AUBREVILLE *et al.*, 2020c; MATHEW *et al.*, 2020). Die aktuellen Algorithmen aus der eigenen Arbeitsgruppe (BERTRAM *et al.*, 2019b; AUBREVILLE *et al.*, 2020a; AUBREVILLE *et al.*, 2020c; BERTRAM *et al.*, 2020b) haben sich an dem Ansatz von LI *et al.* (2018) orientiert, welcher ein Objekterkennungs-Netzwerk (erste Stufe) und ein Patchklassifikations-Netzwerk (zweite Stufe) vereint. Dieser zweistufige Ansatz hat den Vorteil, dass sehr hohe algorithmische Leistungen erzielt werden können und trotzdem ein relativ geringer Rechenaufwand und eine hohe Nachvollziehbarkeit der Ergebnisse besteht

(AUBREVILLE *et al.*, 2020c). Die erste Stage hat eine relativ (im Vergleich zu anderen leistungsstarken Modellen) geringe Komplexität und kann daher schnell ganze WSIs analysieren. Allerdings ist die Leistung nur mäßig und die Objekterkennungsergebnisse werden bewusst zugunsten des Recalls (hoher Recall, niedrige Precision) angepasst. Die Zweite Stufe analysiert mit einer hohen Precision die Bildausschnitte der identifizierten Objekte und kann so die Anzahl der falsch positiven Detektionen entscheidend reduzieren.

Tabelle 6: Übersicht über die öffentlich verfügbaren Datensätze über Mitosefiguren in histologischen Gewebeschnitten von Tumoren

Referenz	Open Data?	Tumortyp	Annotierte Bilder	Labels		Annotatoren
				MF	SN	
Datensätze der eigenen Arbeitsgruppe						
BERTRAM <i>et al.</i> (2019b)	Ja	Kanine Mastzelltumore	32 WSIs	44.880	27.965	2
AUBREVILLE <i>et al.</i> (2020a)	Ja	Kanine Mammatumore	21 WSIs	13.937	36.349	2 (+1)
BERTRAM <i>et al.</i> (2020b)	Ja	Humane Mammakarzinome	107 RVI	1.999	10.484	2
WILM <i>et al.</i> (2021)	Nein *	Kanine Mastzelltumore	50 RVI	1.324 – 4.412	0	12
Datensätze anderer Arbeitsgruppen						
ROUX <i>et al.</i> (2013)	Ja	Humane Mammakarzinome	5 RVI	226	0	1
VETA <i>et al.</i> (2015)	Ja	Humane Mammakarzinome	23 RVI	1.083	0	2 (+2)
ROUX <i>et al.</i> (2014)	Ja	Humane Mammakarzinome	11 RVI	749	2.884	2 (+1)
VETA <i>et al.</i> (2019)	Ja	Humane Mammakarzinome	73 RVI	1.552	0	2 (+2)

Abkürzungen: MF, Mitosefiguren; SN, Streng Negative (Strukturen mit Verwechslungsgefahr zu MF); n. v., nicht verfügbar; RVI, Region von Interesse; WSIs, Whole Slide Images

* Eine zukünftige Open Data Veröffentlichung ist vorgesehen

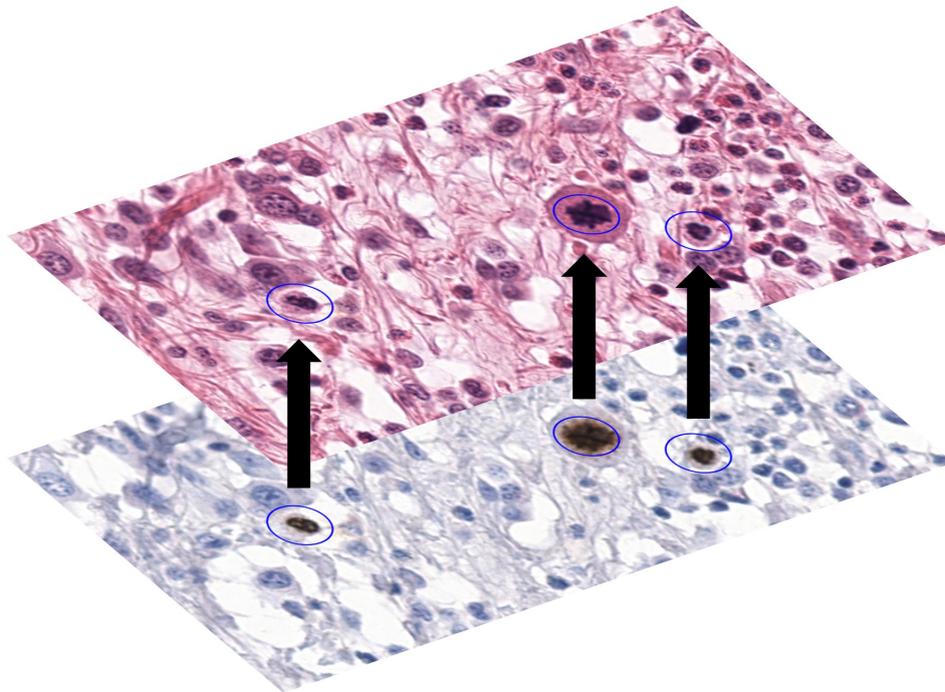


Abbildung 16: Beispiel eines vielversprechenden Labelingansatzes für das Annotieren von Mitosefiguren mittels eines immunhistochemischen Goldstandards (Phospho-Histone H3). Die positiven (braunen) Strukturen im immunhistochemischen Bild (unten) können mit hoher Genauigkeit annotiert werden und auf das histologische Bild (oben, H&E-Färbung) übertragen werden. Labelansatz nach TELLEZ et al. (2018)

4.3.1.2 Bestehende Herausforderungen

Qualitativ und quantitativ hochwertige Datensätze zusammen mit innovativen Deep learning-basierten Netzwerkarchitekturen haben das Potential, die Genauigkeit und Reproduzierbarkeit des Mitotic Counts entscheidend zu verbessern (AUBREVILLE et al., 2020c; DONOVAN et al., 2020). Jedoch liegt aktuell noch eine Vielzahl von Fehlerquellen vor, die eine weitere Entwicklung und Validierung benötigen. Eine der größten aktuellen Limitationen ist die sehr hohe Bildvariabilität der WSIs aufgrund der verschiedenen Gewebearten und -morphologien, Präparatherstellungs- und Färbemethoden, Gewebequalität und Scanner-basierte Bildeigenschaften (KOMURA und ISHIKAWA, 2018; AUBREVILLE et al., 2020b). Wenn zwei Datensätze relevante Unterschiede in diesen Eigenschaften (Domains) aufweisen, kann es zu einem sogenannten Domainshift kommen, welcher zu einer signifikanten Verschlechterung der algorithmischen Leistung führen kann (MARÉE, 2017; AUBREVILLE, 2020). Damit datenbasierte Algorithmen eine hohe Leistung (Robustheit) auf verschiedenen Domains erbringen können, muss der Trainingsdatensatz diese Variabilität abdecken (MARÉE, 2017). So konnten wir beispielsweise zeigen, dass vollständig annotierte Trainings-WSIs entscheidend für eine zuverlässige Analyse gesamter WSIs ist (AUBREVILLE et al., 2020c). Weiterhin konnten wir

zeigen, dass der Scannertyp einen relevanten Einfluss auf die algorithmische Leistung haben kann (AUBREVILLE *et al.*, 2020a). Die Schlussfolgerung daraus ist, dass Deep learning-basierte Algorithmen nicht ohne Weiteres auf WSIs von anderen Domains, wie WSIs von andere Laboren, übertragen werden können. Eine hohe Übertragbarkeit ist eine Grundvoraussetzung für einen ubiquitären Einsatz von Algorithmen. Vielversprechende Ansätze für robuste Algorithmen sind unter anderem Datensätze mit einer hohen Bildvariabilität (AUBREVILLE *et al.*, 2020a), die Normalisierung oder Augmentierung von (Trainings-)Bildern (ABEL *et al.*, 2020) und algorithmische Ansätze (Domainadaptation oder Transfer Learning) (AUBREVILLE *et al.*, 2020a; AUBREVILLE *et al.*, 2020b).

Eine weitere Limitation aktueller Algorithmen besteht darin, dass die Mitoseerkennung nicht spezifisch für Tumorzellen ist. Dies ist besonders relevant für Analysen von gesamtem WSIs mit diversen Gewebetypen (BALKENHOL *et al.*, 2019; BERTRAM *et al.*, 2019b; AUBREVILLE *et al.*, 2020a). In einer Studie, die eine algorithmische Selektion der Tumorregion für den Mitotic Count vorgenommen hat, wurde beispielsweise berichtet, dass in 58% der Proben diese Regionauswahl außerhalb der Tumorregion gelegen ist (BALKENHOL *et al.*, 2019). Ein möglicher Lösungsansatz kann eine Beschränkung der algorithmischen Analyse auf die Tumorregion durch manuelle oder algorithmische Segmentierung der relevanten Bildausschnitte sein (KHAN *et al.*, 2013; BALKENHOL *et al.*, 2019; GUO *et al.*, 2019).

Eine große Herausforderung für die Implementierung in den diagnostischen Alltag ist die Frage, wie die algorithmischen Ergebnisse in die Diagnosestellung eingebunden werden sollen. Ein Kritikpunkt von Deep learning-basierten Ansätzen ist die geringe Nachvollziehbarkeit der Entscheidungskriterien der Modelle („Blackbox“) (ABELS *et al.*, 2019). Wie oben bereits erwähnt sind Algorithmen bei Weitem nicht fehlerfrei und können zu falschen Ergebnissen (zum Beispiel falsch positive Mitosedetektion oder Regionauswahl außerhalb des Tumors) führen. Aktuelle Deep learning-basierte Algorithmen haben in der Regel nicht die Fähigkeit ihre Entscheidungskriterien und –Grenzwerte an den individuellen Fall anzupassen wie es von Pathologen/innen routinemäßig gemacht wird (KLOPFLEISCH und BERTRAM, 2018). Die Situationen, in denen ungenaue Ergebnisse vom Algorithmus erzielt werden, sind zumeist ohne gründliche Überprüfung nicht ersichtlich und unterscheiden sich möglicherweise zwischen den individuellen Modellen (inter-algorithmische Variabilität) insbesondere basierend auf den verwendeten Trainingsdaten und Netzwerkarchitekturen. Obwohl datenbasierte Algorithmen sehr gut statistisch bewertet werden können, ist jedoch das Vorhandensein von allumfassenden Testdatensätzen, welche alle möglichen Fehlerquellen umfangreich abdecken, oft nicht gegeben. Ein sinnvoller Ansatz, um das Risiko für den Patienten von fehlerhaften algorithmischen Ergebnissen zu verringern, ist die computerassistierte Diagnose (siehe Kapitel 4.3.1.3).

4.3.1.3 Computer-assistierter Mitotic Count

Bei dem computer-assistierten Diagnoseansatz bindet der/die Pathologe/in die algorithmischen Ergebnisse als Hilfestellung für die Entscheidungsfindung ein. Es ist das Ziel, die Effizienz, Genauigkeit und Reproduzierbarkeit der Pathologen/innen bei zeitaufwendigen und/oder komplizierte Fragestellungen mithilfe von Algorithmen zu verbessern. Gleichzeitig soll die Zuverlässigkeit der abschließenden Diagnose, infolge einer Überprüfung oder wenn notwendig Korrektur durch eine/n Pathologin/en, gewährleistet werden (KLOPFLEISCH und BERTRAM, 2018). Aktuell liegen zwei Veröffentlichungen über computer-assistierte Mitosezählungen aus der Humanmedizin vor (BALKENHOL *et al.*, 2019; PANTANOWITZ *et al.*, 2020). PANTANOWITZ *et al.* (2020) konnten feststellen, dass das Anzeigen von möglichen Mitosefiguren in Bildausschnitten die Genauigkeit (im Vergleich zum Mehrheitsbeschluss) und die Effizienz verbessert. BALKENHOL *et al.* (2019) haben hingegen den Mitotic Count von Pathologen/innen mittels digitaler Mikroskopie mit algorithmischer Hilfestellung (Regionauswahl und Anzeige von Mitosekandidaten) mit dem Mitotic Count von Pathologen/innen mittels Lichtmikroskopie verglichen und eine hohe Übereinstimmung zwischen den beiden Methoden festgestellt.

Wir haben für kanine Mastzelltumore erste Ansätze entwickelt, wie die Tumorregionauswahl mittels Deep learning-basierten Algorithmen verbessert werden kann (AUBREVILLE *et al.*, 2019a; AUBREVILLE *et al.*, 2019b; AUBREVILLE *et al.*, 2020c). Die Mitosedichte in kaninen Mastzelltumoren ist extrem heterogen verteilt (BERTRAM *et al.*, 2020a) und Pathologen/innen sind häufig nicht in der Lage, die Tumorregion mit der höchsten Mitosedichte aufzufinden (AUBREVILLE *et al.*, 2020c). Algorithmen haben den großen Vorteil, dass diese das gesamte WSI problemlos analysieren können (Abbildung 17A) und somit eine Heatmap für die Mitosedichte berechnen können (Abbildung 17B). Daher haben Algorithmen eine deutlich höhere Sensitivität, die Tumorregionen mit der höchsten mitotischen Aktivität aufzufinden. Eine fehlerhafte Regionauswahl (z.B. außerhalb der Tumorregion) kann möglicherweise durch eine/n Pathologin/en mit hoher Spezifität erkannt werden und entsprechend mithilfe verschiedener algorithmischer Hilfestellungen (z.B. Heatmaps) korrigieren kann. Zukünftige Studien müssen zeigen, inwiefern diese automatisierte Regionauswahl zur Standardisierung des Mitotic Counts beitragen kann und welche Maßnahmen getroffen werden müssen, um eine hohe Zuverlässigkeit der computer-assistierten Mitosezählungen zu gewährleisten.

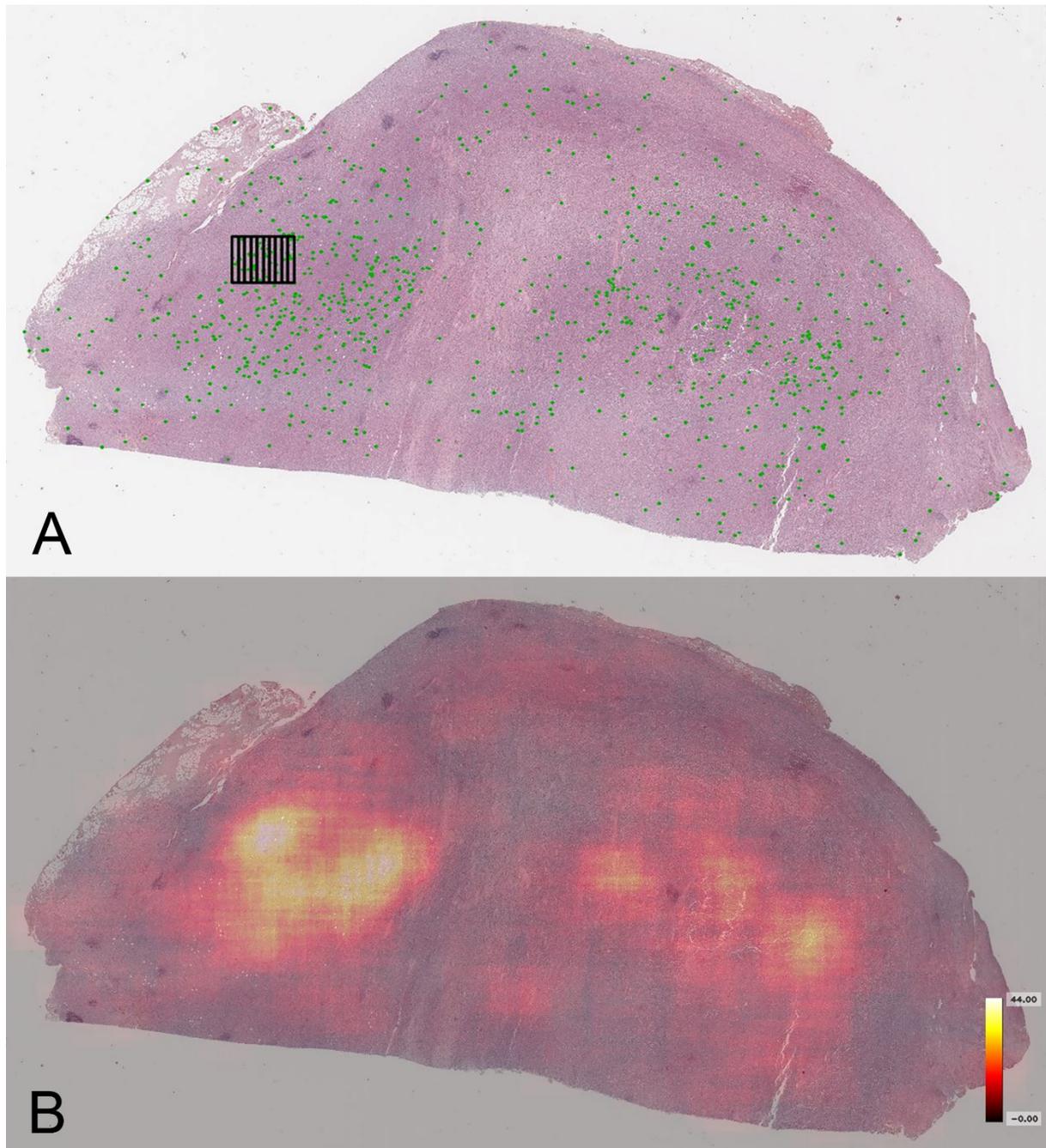


Abbildung 17: Mögliches computer-assistiertes Diagnosesystem für die Bestimmung des Mitotic Counts in Whole Slide Images von kaninen Mastzelltumoren

A) Mitosedetektionen (grüne Punkte) und Region of Interest Selektion (schwarzer Kasten; Tumorregion mit der höchsten mitotischen Aktivität) mit einem Deep learning-basierten Algorithmus von AUBREVILLE et al. (2020c)

B) Heatmap-Darstellung der Mitosedichte (Anzahl der Mitosefiguren pro $2,37 \text{ mm}^2$ Gewebefläche) basierend auf den Mitosedetektionen von Abbildung 17A)

4.3.2 Zwei- und mehrkernige Tumorzellen

Die erhöhte Anzahl an Zellkernen pro Tumorzelle weist auf einen vermehrten Gehalt an Chromosomen (Polyploidie) hin. Zwei- und mehrkernige Tumorzellen können auf zwei Weisen entstehen: 1) wenn der letzte Schritt der Zellteilung (Zytokinese) abgebrochen wird und somit keine neue Tochterzelle entsteht (Endoreplikation) oder 2) wenn zwei Tumorzellen fusionieren (Synzytium) (CHEN *et al.*, 2019). Aufgrund des vermehrten genetischen Materials weisen polyploide Tumorzellen eine erhöhte metabolische Aktivität auf und können somit den zunehmend steigenden Ansprüchen für das Tumorstadium besser bewältigen (AMEND *et al.*, 2019; CHEN *et al.*, 2019). Die Polyploidie wird allgemein als einer der relevanten Schlüsselantreiber der Onkogenese, Tumorstadium und Resistenzentwicklung gegen Chemotherapeutika angesehen (AMEND *et al.*, 2019; CHEN *et al.*, 2019). Bei Mastzelltumoren des Hundes wird die Anzahl der mehrkernigen Tumorzellen pro histologischer Tumorstadium als ein prognostisch relevantes Kriterium für das histologische Tumorstadium verwendet (KIUPEL *et al.*, 2011). Weiterhin konnte eine Korrelation zwischen der Anzahl von zweikernigen Tumorzellen mit dem klinischen Outcome in zytologischen Präparaten von Hunden mit Mastzelltumoren nachgewiesen werden (CAMUS *et al.*, 2016). Wir konnten zeigen, dass Pathologen/innen eine sehr hohe inter-observer Variabilität für die Objekterkennung von zwei- und mehrkernigen Tumorzellen aufweisen (BERTRAM *et al.*, 2021). Um die Genauigkeit und Reproduzierbarkeit der Bestimmung von zwei- und mehrkernigen Zellen zu verbessern, haben wir erste Ansätze für die automatisierte Bildanalyse entwickelt.

Als Grundvoraussetzung für die Entwicklung von daten-basierten Algorithmen haben wir den ersten öffentlich zugänglichen Datensatz mit einer großen Anzahl an Annotationen in der gesamten Gewebefläche von 32 WSIs erstellt (Tabelle 7). Ein zweistufiger Deep learning-basierter Algorithmus (in Anlehnung an unseren Mitosealgorithmus von AUBREVILLE *et al.* (2020c)) wurde trainiert, welcher eine hohe Objekterkennungsleistung erzielen hat (Abbildung 18) mit einem F_1 -Wert von 0,675 für zweikernige Tumorzellen und einen F_1 -Wert von 0,623 für mehrkernige Tumorzellen auf dem Testdatensatz (BERTRAM *et al.*, 2021). Zusätzlich haben wir die Leistung des Algorithmus mit der Objekterkennungsfähigkeit von fünf Pathologen/innen auf kleineren Bildausschnitten des Testdatensatzes verglichen. Dabei konnten wir zeigen, dass der Algorithmus höhere F_1 -Werte als die Pathologen/innen erzielen konnte. Diese ersten Ergebnisse sind sehr vielversprechend, dass computer-assistierte Diagnosesysteme die Bestimmung von zwei- und mehrkernigen Tumorzellen und somit die Tumorstadium verbessern kann. Unser öffentlich zugänglicher Datensatz ermöglicht eine weiterführende Erforschung optimaler algorithmischer Ansätze und Netzwerkarchitekturen.

Tabelle 7: Übersicht über die öffentlich verfügbaren Datensätze über zwei- und mehrkernige Tumorzellen in Gewebeschnitten von Tumoren

Referenz	Open Data?	Tumortyp	Annotierte Bilder	Labels	Annotatoren
Datensätze der eigenen Arbeitsgruppe					
BERTRAM <i>et al.</i> (2021)	Ja	Kanine Mastzelltumore	32 WSIs	Zweikernige: 19.983 Mehrkernige: 1.416	1
Datensätze anderer Arbeitsgruppen					
Nicht verfügbar					

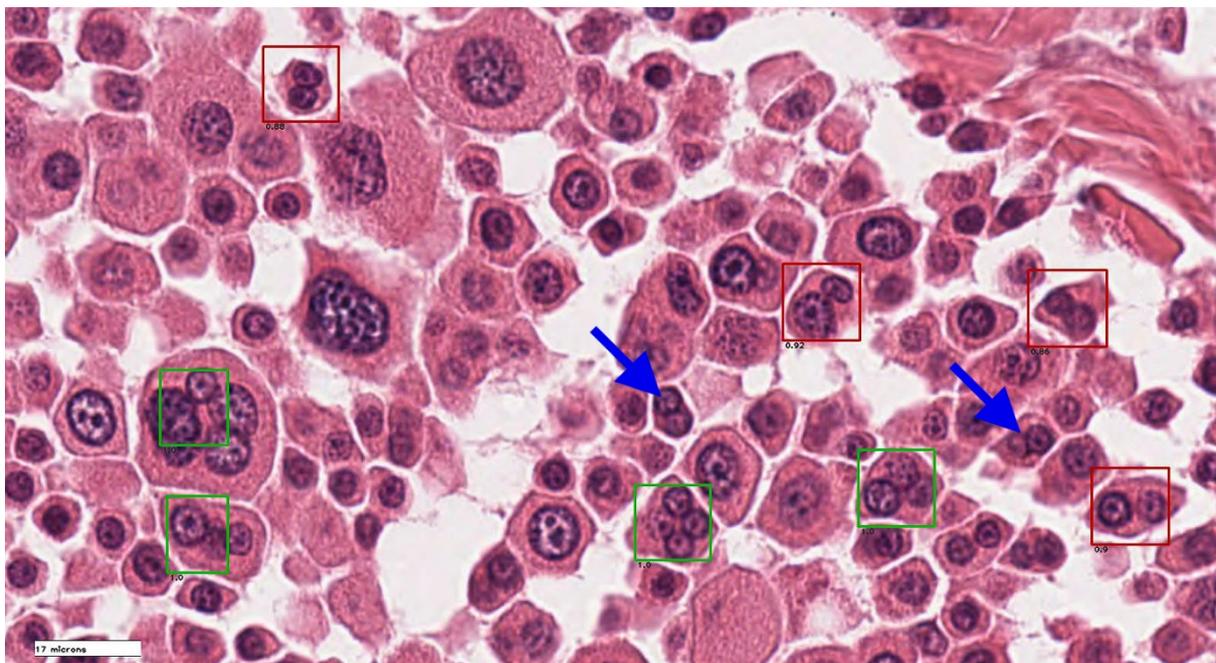


Abbildung 18: Histologischer Bildausschnitt eines kaninen Mastzelltumors mit algorithmischen Detektionen von zweikernigen (rote Kästchen) und mehrkernigen (grüne Kästchen) Tumorzellen basieren auf dem Deep learning-basierten Algorithmus von BERTRAM *et al.* (2021)

Legende: Blaue Pfeile, falsch negative Detektionen

4.3.3 Zytologie der equinen bronchoalveolären Lavageflüssigkeit

Die Anwendung der digitalen Pathologie für die Zytologie wird als umstritten angesehen, da die Bewertung von zytologischen Präparaten häufig auf sehr hohe Auflösung (1.000er Vergrößerung) und z-stacking (mehrere Fokusebenen zum Feinfokussieren) angewiesen ist (BERTRAM und KLOPFLEISCH, 2017). Es ist beschrieben, dass eine standardmäßige Bildqualität möglicherweise für die Beurteilung durch eine/n Pathologen/in (fungiert auch als Annotator/in) unzureichend sein kann und ebenfalls eine automatisierte Bildanalysen behindern kann. Im Gegensatz zu Feinnadel- und Abklatschpräparaten von soliden Geweben, ermöglicht das Zytospinverfahren von Körperflüssigkeiten eine vergleichsweise standardisierte und zellschonende Herstellung von qualitativ hochwertigen zytologischen Präparaten (BERTRAM *et al.*, 2018a; KLOPFLEISCH und BERTRAM, 2019). Die Untersuchung der bronchoalveolären Lavageflüssigkeit (Lungenspülprobe der tiefen Atemwege) ist eine häufige Methode für die Diagnose von verschiedenen Erkrankungen der tiefen Atemwege beim Pferd (KLOPFLEISCH und BERTRAM, 2019). Nach eigener Einschätzung haben WSIs (400er Vergrößerung, eine Fokus-ebene) von Zytospinpräparaten eine ausreichende diagnostische Qualität. Die Entwicklung von automatisierten Bildanalyse-Systeme für die Bewertung von WSIs equiner Lungenspülproben ist daher ein sehr interessantes Forschungsgebiet.

4.3.3.1 Zytologische Quantifizierung des equinen Lungenblutens

Das Lungenbluten (exercise-induced pulmonary hemorrhage) bei Pferden entsteht zumeist im Zusammenhang mit hoher sportlicher Leistung (SULLIVAN und HINCHCLIFF, 2015). Bei Sportpferden kann diese Erkrankung mit sehr hoher Prävalenz nachgewiesen werden (SULLIVAN und HINCHCLIFF, 2015). Es wird angenommen, dass diese Erkrankung eine deutliche Leistungsminderung bei Pferden bewirken kann und diese Erkrankung hat daher für den Rennsport eine hohe ökonomische Bedeutung (SULLIVAN und HINCHCLIFF, 2015). Die zytologische Untersuchung der bronchoalveolären Lavageflüssigkeit mittels Spezialfärbungen für Eisen wird als die sensitivste diagnostische Methode angesehen (KLOPFLEISCH und BERTRAM, 2019). DOUCET und VIEL (2002) haben für das Pferd ein Gradingssystem entwickelt. Laut dieses Systems sollen 300 alveoläre Makrophagen in fünf Grade (0 - 4) anhand des intrazytoplasmatischen Hämosideringehaltes eingestuft werden. Der Total Hemosiderin Score wird anschließend für 100 Zellen berechnet (Messbereich: 0 - 400) und die Diagnose „Lungenbluten“ wird ab einen Grenzwert von 75 gestellt. Die Durchführung des Grading-Systems ist jedoch sehr zeitaufwendig (erfahrungsgemäß circa 10 Minuten) und das Klassifizieren von Hämosiderophagen weist eine sehr hohe intra- und inter-observer Variabilität auf (MARZAHN *et al.*, 2020c; MARZAHN *et al.*, 2020d).

Zusammen mit unseren Forschungspartnern von der Friedrich-Alexander-Universität Erlangen-Nürnberg habe wir einen ersten Deep learning-basierten Algorithmus entwickelt (MARZAHN *et al.*, 2020c). Dafür wurde einen Datensatz aus 17 WSIs mit insgesamt 78.047 Annotationen erzeugt (Tabelle 8). Das Modell besteht aus einem RetinaNet und erzielte für die Objekterkennung auf den drei Test-WSIs ein mean Average Precision (mAP) von 0,66 für alle fünf Klassen (Abbildung 19). Bei der Klassifizierung von einzelnen Zellausschnitten konnte eine Übereinstimmung zu den Ground truth Klassen in 85% erreicht werden, während neun Experten/innen eine Übereinstimmung zwischen 68 und 86% aufwiesen.

Tabelle 8: Übersicht über den entwickelten zytologischen Datensatz mit den verschiedenen Klassen an Hämosiderophagen aus equinen bronchoalveolären Lavage-Flüssigkeiten

Referenz	Open Data?	Annotierte Bilder	Labels	Annotatoren
Datensätze der eigenen Arbeitsgruppe				
MARZAHN <i>et al.</i> (2020c)	Nein *	17 WSIs	Insgesamt: 78.047 Klasse 0: 29.055 Klasse 1: 26.925 Klasse 2: 13.524 Klasse 3: 7.070 Klasse 4: 1.473	1
Datensätze anderer Arbeitsgruppen				
Nicht verfügbar				

* Eine Open Data Veröffentlichung ist vorgesehen

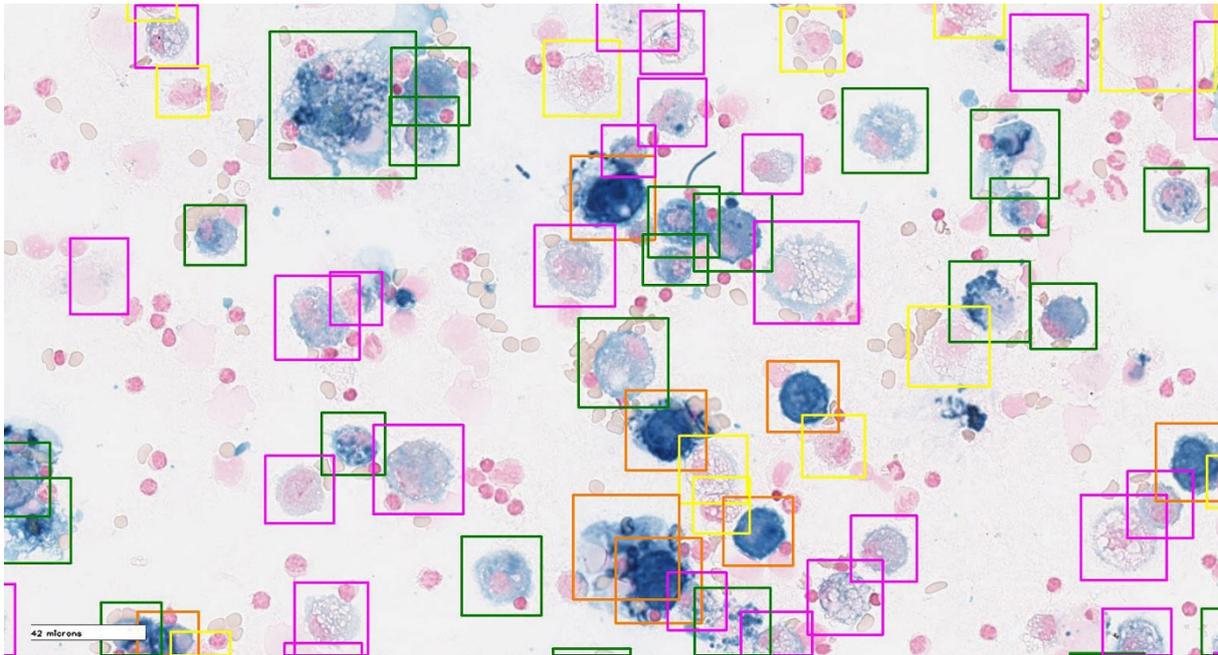


Abbildung 19: Zytologischer Bildausschnitt einer Lungenspülprobe eines Pferdes mit algorithmischen Detektionen von Hämosiderophagen. Eisenspezialfärbung (blaues Pigment). Legende: Klassifizierung der Hämosiderophagen nach DOUCET und VIEL (2002); Klasse 0: gelbe Kästchen, Klasse 1: pinke Kästchen, Klasse 2: grüne Kästchen, Klasse 3, orange Kästchen; Klasse 4, nicht vorhanden auf dem Bildausschnitt; Detektionen basieren auf dem Deep learning-basierten Algorithmus von MARZAHL et al. (2020c)

4.3.3.2 Automatisierte Bestimmung der Zellzusammensetzung

Die zytologische Bestimmung der Zellzusammensetzung von bronchoalveolären Lavageflüssigkeiten ist die Goldstandardmethode für die Diagnose von entzündlichen Veränderungen der tiefen Atemwege des Pferdes (KLOPFLEISCH und BERTRAM, 2019). Es werden mehrere hundert Zellen in die verschiedenen Zelltypen (Lymphozyten, Makrophagen, neutrophile und eosinophile Granulozyten und Mastzellen) eingeteilt. Je nach prozentualem Anteil der Granulozyten und der Mastzellen an der ermittelten Gesamtzellzahl kann auf eine chronische Atemwegsentzündung zurückgeschlossen werden (KLOPFLEISCH und BERTRAM, 2019). Das Auszählen von mehreren hundert Zellen im zytologischen Präparat stellt dabei eine recht zeitaufwendige Arbeit dar und die Klassifizierung der Zelltypen kann eine gewisse Variabilität aufweisen (MARZAHL et al., 2020d). Eine automatisierte Bildanalyse für die Bestimmung der Zellzusammensetzung kann daher eine deutliche Zeitersparnis und möglicherweise eine Verbesserung der diagnostischen Genauigkeit ermöglichen.

Eine Vorstudie für dieses Projekt hat sich mit der Optimierung der Probenqualität durch Zytospray-Fixierung des Zytospinpräparates beschäftigt (BERTRAM *et al.*, 2018a). Wir haben uns für dieses Experiment entschieden, da eine suboptimale Probenqualität durchaus die Fähigkeiten der Differenzierung der verschiedenen Zelltypen negativ beeinflussen kann und somit die Leistungsfähigkeit sowohl des/der Experten/in (Annotator/in von Datensätzen) als auch die Bildanalysesoftware reduzieren kann.

Für die Entwicklung eines daten-basierten Algorithmus wurde eine Datenbank mit insgesamt 87.099 Annotationen in 6 WSIs erstellt (Tabelle 9). Unsere Forschungspartner von der Friedrich-Alexander-Universität Erlangen-Nürnberg haben eine RetinaNet Architektur verwendet, um einen ersten Deep Learning-basierten Algorithmus zu entwickeln (MARZAHL *et al.*, 2021). Auf dem Testdatensatz zeigte der Algorithmus eine sehr hohe Leistung mit einer mean Average Precision (mAP) von 0,87. Auf einem zweiten Datensatz mit Annotationen von mehreren Experten/innen (MARZAHL *et al.*, 2020d) konnte eine mAP von 0,76 für den Algorithmus und eine durchschnittliche mAP von 0,73 für die Experten/innen bestimmt werden. Diese ersten Ergebnisse sind sehr vielversprechend für eine automatisierte Bestimmung der Zellzusammensetzung zur Standardisierung und Effizienzsteigerung dieser diagnostischen Aufgabe. Insbesondere sollten größere Datensätze mit einer größeren Anzahl an inkludierten Bildern entwickelt werden.

Tabelle 9: Übersicht über die entwickelten zytologischen Datensätze mit den verschiedenen Zelltypen aus equinen bronchoalveolären Lavageflüssigkeiten

Referenz	Open Data?	Annotierte Bilder	Labels	Annotatoren
Datensätze von der eigenen Arbeitsgruppe				
MARZAHL <i>et al.</i> (2020b)	Nein *	6 WSIs (unvollständig annotiert)	Insgesamt: 87.099 Lymphozyten: 46.398 Makrophagen: 26.501 Neutrophile Granulozyten: 12.557 Mastzellen: 1.535 Eosinophile Granulozyten: 108	1
Datensätze von anderen Arbeitsgruppen				
Nicht verfügbar				

* Eine Open Data Veröffentlichung ist vorgesehen

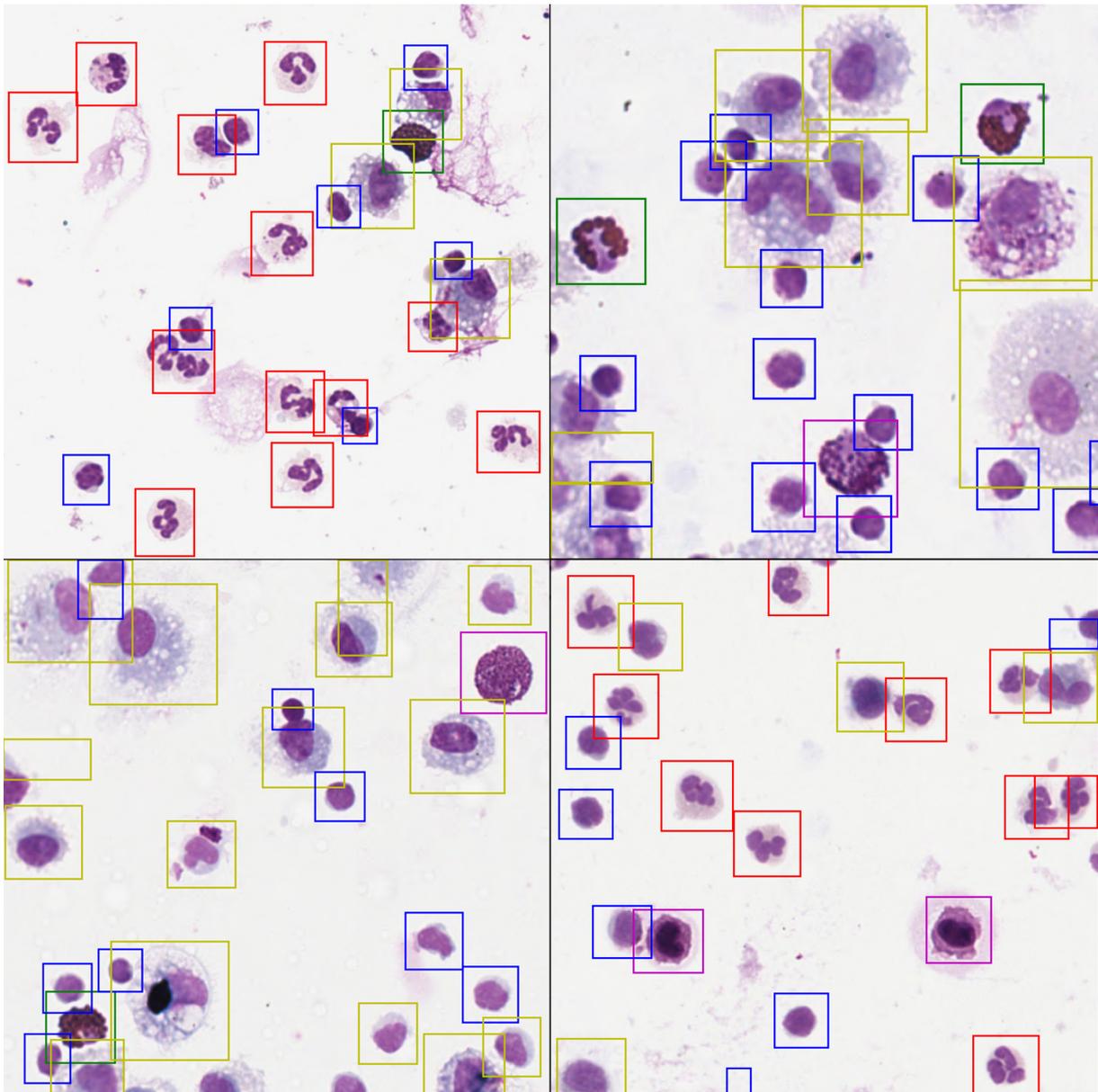


Abbildung 20: Vier zytologische Bildausschnitte von Lungenspülproben des Pferdes mit algorithmischen Detektionen der verschiedenen Zelltypen. May-Grünwald-Giemsa-Färbung. Legende: Gelbe Kästchen, Makrophagen; blaue Kästchen, Lymphozyten; rote Kästchen, neutrophile Granulozyten; pinke Kästchen, Mastzellen; grüne Kästchen, eosinophile Granulozyten; Detektionen basieren auf den Deep learning-basierter Algorithmus von MARZ AHL et al. (2021)

4.3.4 Methoden zur Erstellung von Datensätzen

Eines der am meisten kritisierten Limitationen von Deep learning-basierten Algorithmen ist die fehlende Verfügbarkeit von „guten“ Datensätzen (VETA *et al.*, 2015; MARÉE, 2017; KOMURA und ISHIKAWA, 2018; BERA *et al.*, 2019; RASHIDI *et al.*, 2019; ABEL *et al.*, 2020; BERTRAM *et al.*, 2020b; MARZAHN *et al.*, 2020d; MCALPINE und MICHELOW, 2020; MCALPINE *et al.*, 2020; SULTAN *et al.*, 2020; WILM *et al.*, 2021). Komplexe artifizielle Netzwerkarchitekturen haben ein hohes Potential für leistungsstarke Modelle, jedoch besteht – im Vergleich zu anderen algorithmischen Ansätzen – ein besonders hoher Bedarf an ausreichenden Trainingsdaten (AEFFNER *et al.*, 2018; KLOPFLEISCH und BERTRAM, 2018). Mit dem steigenden Anspruch an hoher Leistungsfähigkeit und Robustheit der Algorithmen, sollen auch die Anforderungen an die verfügbaren Datensätze zunehmen, um eine zufriedenstellende Generalisierungsleistung der Modelle zu bewirken (MARÉE, 2017; BERTRAM *et al.*, 2019a; RASHIDI *et al.*, 2019; AUBREVILLE, 2020). Weiterhin wird eine umfassende Leistungsüberprüfung von den generierten Modellen gefordert, da potentielle Fehlerquellen anderenfalls nicht erkannt werden können („Blackbox“) (ABELS *et al.*, 2019; MCALPINE und MICHELOW, 2020). Daher wird auch bei der Evaluation von Deep learning-basierten Algorithmen ein sehr hoher Anspruch an die Testdatensätze gestellt (BERTRAM *et al.*, 2020b). Allerdings hat eine eigene Untersuchung gezeigt, dass große Unterschiede zwischen experten-basierten Annotationen für die gleichen Bilder entstehen können, welche einen großen Einfluss auf das Trainieren und die Evaluation von Modellen hat (BERTRAM *et al.*, 2020b). In den eigenen Untersuchungen konnten wir feststellen, dass „gute“ Datensätze (sowohl Training- als auch Testdatensätze) vor allem folgende Eigenschaften aufweisen:

- Hohe Quantität der Annotationen
- Hohe Qualität der Annotationen (unerlässlich für Testdatensatz)
 - Hohe Richtigkeit der Annotationen im Vergleich zur echten Wahrheit
 - Hohe Konsistenz / Reproduzierbarkeit der Labelingkriterien
 - Hohe Vollständigkeit der Annotationen (wenige falsch negative)
- Hohe Repräsentierbarkeit der annotierten Bilder
 - Bilder/Bildausschnitte entsprechen denen der vorgesehenen Anwendung
 - Realistisches Maß an Bildvariabilität (Gewebemorphologien, Gewebetypen, Färbevariabilität, Präparatherstellung, WSI-Scannertypen usw.)

Histologische Datensätze bestehen aus Bildern (idealerweise WSIs) und assoziierten Ground truth Labels. Diese Labels können auf verschiedenen Ebenen des Bildes zugewiesen werden: unter anderem der 1) globalen WSI/Bild-Ebene (zum Beispiel Klassifikation der Malignität eines Tumors (ARESTA *et al.*, 2019)), 2) oder lokalen Mustern im Bild entweder auf der 2A) gewebestrukturelle Ebene (zum Beispiel Segmentierung von metastatische Herde in einem

Lymphknoten (BEJNORDI *et al.*, 2017)) und der 2B) Objekt/Zell-Ebene (zum Beispiel Objekterkennung von Mitosefiguren im Gewebeschnitt (BERTRAM *et al.*, 2019b)). Abhängig von dieser Labelbenen unterscheiden sich die optimalen Methoden der Datensatzentwicklung teilweise erheblich. Bei den eigenen Datensätzen konnten wir vor allem Erfahrung mit Annotationen auf dem zellulären Level sammeln (BERTRAM *et al.*, 2019b; AUBREVILLE *et al.*, 2020a; AUBREVILLE *et al.*, 2020b; BERTRAM *et al.*, 2020b; MARZAHN *et al.*, 2020a; MARZAHN *et al.*, 2020c; MARZAHN *et al.*, 2020d; BERTRAM *et al.*, 2021; MARZAHN *et al.*, 2021; WILM *et al.*, 2021).

Bei der Erstellung der Datenbank mit lokalen Annotationen werden die Bilder von Annotatoren/innen visuell untersucht und alle gefundenen Objekte der gesuchten Klasse(n) manuell annotiert, welches mit einem sehr großen Zeitaufwand verbunden ist. Die Verwendung von spezialisierter Annotationssoftware (AUBREVILLE *et al.*, 2018; MARZAHN *et al.*, 2020b) ist dabei relevant für einen maßgerechten, effizienten und hochwertigen Labelingprozess. Beispielweise sind Softwaretools wie das Guided Screening essentiell für das vollständige Annotieren gesamter WSI. Neben dem Labelingverfahren (siehe unten) ist entscheidend, dass Datensätze mit repräsentativen Bildern/Bildausschnitten für die vorgesehene Anwendung generiert werden. Wenn beispielsweise Algorithmen mit Annotationen nur aus gewissen Geweberegionen trainiert werden, kann es vorkommen, dass zahlreiche falsch positive Detektionen bei einer Analyse von gesamten WSI auftreten (BERTRAM *et al.*, 2019b; AUBREVILLE *et al.*, 2020c). Wir können zeigen, dass die Annotationen aus allen Bildabschnitten eines WSIs eine wesentliche Verbesserung der Genauigkeit bei der WSI-Analyse ermöglicht.

Für das Labeling der meisten morphologischen Muster stellen erfahrene Pathologen/innen den Goldstandard dar. Allerdings ist es hinreichend bekannt, dass erfahrene Pathologen/innen für zahlreiche Fragestellungen eine hohe inter- und intra-observer Variabilität aufweisen aufgrund verschiedener visueller und kognitiver Limitationen (AEFFNER *et al.*, 2017). Das heißt, dass eine perfekte Labelqualität (fehlerfreie Datensätze) infolge der manuellen Labelmethode nicht erwartet werden kann, selbst wenn diese von Experten/innen erstellt wurden (BERTRAM *et al.*, 2020b). Häufig wird ein Multi-Expert-Konsensus für jedes Label erstellt, um die Fehlerquelle individueller Annotatoren/innen zu reduzieren (Abbildung 21; Tabelle 10). In einer eigenen Untersuchung wurde beispielsweise gezeigt, dass Trainingsdatensätze von mindestens drei Pathologen/innen eine deutlich höhere und konsistentere Leistungsfähigkeit der trainierten Algorithmen ermöglichen als ein Datensatz eines individuellen Pathologen/in im Vergleich zur mehrheitsmeinungs-basierten Ground truth (WILM *et al.*, 2021). Für Situationen in denen keine Experten/innen für die Erstellung von Datensätzen verfügbar sind, hat das Crowd-Sourcing (Mehrheitsmeinung von möglichst vielen Nicht-Experten) eine große

Bedeutung gewonnen (MARZAHN *et al.*, 2020a). Besonders bei speziellen und komplizierten pathologischen Fragestellungen kann eine hohe Erfahrung jedoch enorm wichtig sein.

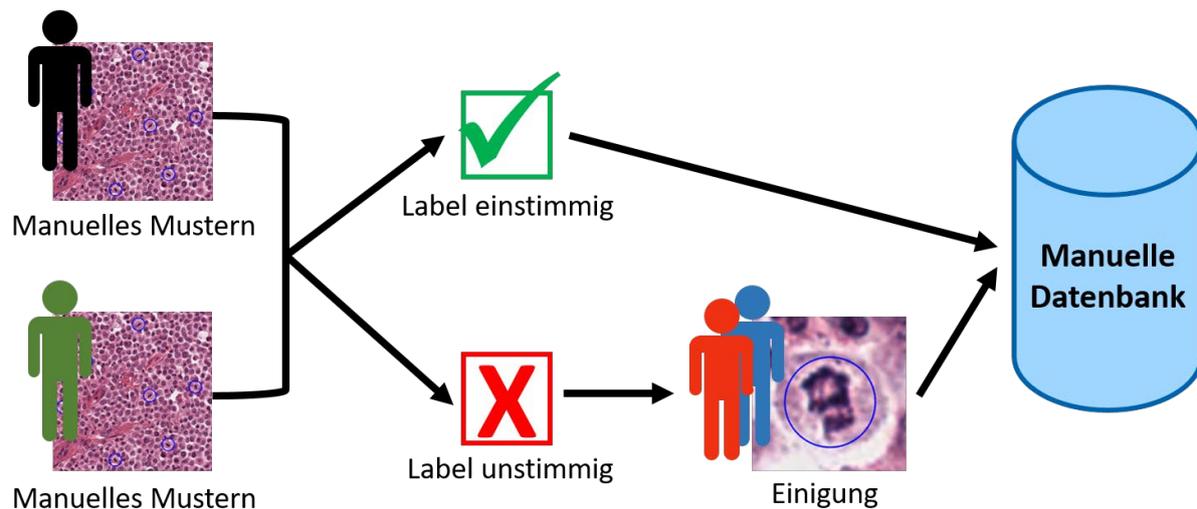


Abbildung 21: Möglicher Arbeitsablauf einer multi-Experten-basierten Labelingmethode von Mitosefiguren in histologischen Bildern. Labelingmethode nach VETA *et al.* (2015).
Abbildung modifiziert von BERTRAM *et al.* (2020b)

Eine sinnvolle Ergänzung zu einer manuell erstellten Datenbank kann das algorithmisch-augmentierte Labeling sein. Hierbei können sowohl zuvor manuell annotierte Labels mit computerisierter Hilfe korrigiert werden (Erhöhung der Datensatzqualität) oder neue Bilder mithilfe einer algorithmischen Vorannotierung effizient ergänzt werden (Tabelle 10). Für beide Labelingmethoden ist die Verfügbarkeit von manuell erstellten, vorläufigen Datenbanken mit einer annehmbaren Qualität und Quantität notwendig, um ausreichend leistungsstarke datenbasierte Algorithmen für die Vorannotierung zu generieren. Alle algorithmisch-augmentierten Labelingmethoden bedürfen einer sehr sorgfältigen Überprüfung durch (idealerweise erfahrene) Annotatoren/innen, um eine hohe Labelqualität sicherzustellen. Nichtsdestotrotz muss beachtet werden, dass die Verwendung von algorithmischen Prädiktionen ein gewisses Risiko birgt, dass der/die Annotator/in beeinflusst wird. Je nach Leistungsfähigkeit des Algorithmus und Kompetenz der Annotatoren/innen kann sich im Einzelfall dieser Bestätigungsbias negativ oder sogar positiv auf die Labelqualität auswirken (MARZAHN *et al.*, 2020a; MARZAHN *et al.*, 2020d). Letzteres ist jedoch nur der Fall, wenn der Algorithmus bereits eine sehr hohe Leistung hat und somit der initiale Datensatz schon relativ gut ist. Den Risiken der algorithmisch-augmentierten Datenbankherstellung stehen die visuellen und kognitiven Limitationen der Annotatoren/innen und/oder der reduzierte Zeitaufwand gegenüber.

Ein Anwendungsgebiet für die Erhöhung der Labelvollständigkeit ist beispielsweise das algorithmische Auffinden von übersehenen Objekten (Abbildung 22). Bei der Erstellung von vorherigen Datensätzen konnte festgestellt werden, dass eine hohe Anzahl an gesuchten Objekten (zum Beispiel Mitosefiguren) durch den manuellen Labelingprozess übersehen werden, trotz der Verwendung mehrerer Experten/innen oder wiederholter Screeningvorgängen (VETA *et al.*, 2015; BERTRAM *et al.*, 2019b; AUBREVILLE *et al.*, 2020c; BERTRAM *et al.*, 2020b; BERTRAM *et al.*, 2021). In eigenen Untersuchungen konnten wir den positive Effekt derartig algorithmisch augmentierter Datenbanken auf die Leistungsfähigkeit des finalen Algorithmus nachweisen (BERTRAM *et al.*, 2019b; AUBREVILLE *et al.*, 2020a; BERTRAM *et al.*, 2020b). Der angewendete (vorläufige) Algorithmus für diese Labelmethode sollte vor allem einen hohe Recall – unter Einbußen der Precision – aufweisen (BERTRAM *et al.*, 2020b). Die hohe Sensitivität gewährleistet, dass nahezu alle übersehenen Kandidaten detektiert werden und eine hohe Expertise der Annotationen garantiert eine hohe Richtigkeit der Labels. Der hohe Anteil an falsch positiven Detektionen vermag möglicherweise ebenfalls den Bestätigungsbias der Annotatoren/innen zu reduzieren. Allerdings führt eine übermäßig niedrige Precision (und somit eine unverhältnismäßige Anzahl von falsch positiven) auch zu einem erheblichen Zeitaufwand, sodass die Sensitivität und der Zeitaufwand dieser Methode gegeneinander abgewogen werden müssen.

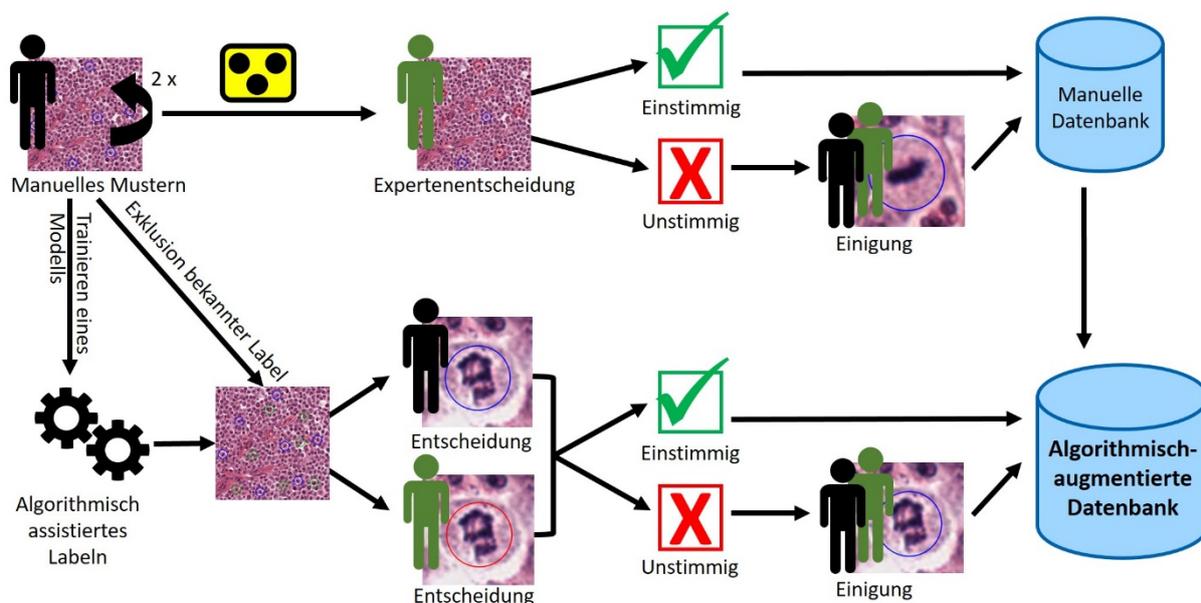


Abbildung 22: Möglicher Arbeitsablauf einer algorithmisch-assistierten Labelingmethode zur Erstellung qualitativ hochwertiger Datensätze von Mitosefiguren in histologischen Bildern.

Labelingmethode nach BERTRAM *et al.* (2019b), BERTRAM *et al.* (2020b) und AUBREVILLE *et al.* (2020a). Abbildung modifiziert von BERTRAM *et al.* (2020b)

Eine Möglichkeit, die Labelkonsistenz von bestehenden Labels zu verbessern, ist die Visualisierung von bestimmten Clusterbildungen der Annotationen basierend auf Merkmalsvektoren. Mit dieser Methode konnten wir beispielsweise fehlerhafte Annotationen in einem Mitosedatensatz sehr schnell darstellen und eliminieren (AUBREVILLE *et al.*, 2020a). Bei Objekten, die einen kontinuierlichen Übergang zwischen den Labelklassen aufweisen, stellt dieser Ansatz eine ideale Methode zum einheitlichen Anwenden Kriterien im Grenzbereich dar. So können beispielsweise Hämosiderophagen mit einem unterschiedlichen Gehalt an Eisenpigment anhand eines algorithmischen Regressionswertes aufsteigend sortiert und inkonsistente Klassifizierungen korrigiert werden (MARZAHN *et al.*, 2020b).

Eine zeiteffiziente Erhöhung der annotierten Bilder kann anhand von algorithmischen Vorannotationen der neuen Bilder erfolgen. Wenn die Überprüfung der Labels von einem Experten/innen erfolgt, wird diese Methode „Expert-Algorithmus-Kollaboration“ genannt (MARZAHN *et al.*, 2020d), und wenn die Kontrolle von Nicht-Experten/innen durchgeführt wird, handelt es sich um Crowd-Algorithmus-Kollaboration (MARZAHN *et al.*, 2020a). In eigenen Untersuchungen konnten wir zeigen, dass diese Methoden durchaus eine enorme Zeitersparnis ermöglichen und – bei der Verwendung von leistungsstarken Algorithmen – ebenfalls eine Verbesserung der Labelkonsistenz bewirken können (MARZAHN *et al.*, 2020a; MARZAHN *et al.*, 2020d). Allerdings haben wir auch festgestellt, dass eingeführte Fehler häufig von den Annotatoren/innen nicht gefunden wurden. Ein interessantes Anwendungsgebiet für diese Methode (idealerweise Expert-Algorithmus-Kollaboration) ist die effiziente Erschließung neuer Domains (zum Beispiel Proben von einer anderen Spezies oder anderem Gewebetyp) für die gleiche pathologische Fragestellung.

Manche Arbeitsgruppen haben vollständig algorithmisch erstellte Annotationen ohne Kontrolle durch Annotatoren/innen (Pseudolabels) generiert (AKRAM *et al.*, 2018; PATI *et al.*, 2020). Diese Methode dient dazu, die kleine Anzahl an experten-basierten Datensätze ohne weiteren Zeitaufwand und Pathologie-Expertise mit weiteren Annotationen zu ergänzen, um die Limitationen von kleinen Datensätzen zu reduzieren. Tatsächlich vermag die Methode die Leistungsfähigkeit von Deep learning-basierten Algorithmen zu verbessern, wie es von AKRAM *et al.* (2018) für einen Mitosealgorithmus gezeigt wurde. Die Algorithmen zur Erstellung von Pseudolabels sollten eine möglichst hohe Leistungsfähigkeit haben, was jedoch bei der Verfügbarkeit von nur kleinen Trainingsdatensätzen problematisch sein kann. Das Generieren von Pseudolabels birgt daher ein sehr hohes Risiko für falsch positive und falsch negative Annotationen. Nach eigener Meinung sollten Pseudolabels ausschließlich für den Trainingsdatensatz verwendet werden. Die Verwendung von Pseudolabels für Testdatensätze sollte unter allen Umständen vermieden werden, da die tatsächliche Leistungsfähigkeit des Algorithmus nicht anhand von Pseudolabels bewertet werden kann.

Tabelle 10: Vergleich zwischen verschiedenen Labelingmethoden für die Erstellung von Datensätzen aus histologischen und zytologischen Bildern

Vergleichende Aspekte	Manuelles Labeling	Algorithmisch-augmentiertes Labeling	
		Von bereits annotierten Bildern	Von zuvor nicht annotierten Bildern
Einsatzzweck	Erstellen eines (initialen) Datensatzes	Erhöhung der Datensatzqualität	Erhöhung der Datensatzquantität
Angewendete Methoden	<ul style="list-style-type: none"> Manuelles Screening durch Experten/innen Crowd-sourcing 	<ul style="list-style-type: none"> Auffinden von verpassten Kandidaten (Label-vollständigkeit) Vergleichende Visualisierung von Labelklassen (Label-konsistenz) 	<ul style="list-style-type: none"> Expert-Algorithmus-Kollaboration Crowd-Algorithmus-Kollaboration
Gewünschte Pathologie-Expertise des/der Annotator/in	Hoch	Hoch	Hoch (Expert-Algorithmus-Kollaboration)
Benötigte Mustererkennungs-Expertise	Gering	Hoch	Hoch
Vorteile	<ul style="list-style-type: none"> Expertenmeinung ist der „Goldstandard“ 	<ul style="list-style-type: none"> Abhängig von der angewendeten Methode (siehe oben) 	<ul style="list-style-type: none"> Schnelles Erstellen von neuen Labels (Quantität) Gegebenenfalls verbesserte Qualität (Label-konsistenz, Label-vollständigkeit)
Nachteile	<ul style="list-style-type: none"> Inter- & intra-rater Variabilität (Label-Konsistenz) Visuelle und kognitive Limitationen (Label-Richtigkeit) Ermüdung der Konzentration (Label-Vollständigkeit) Hoher Zeitaufwand 	<ul style="list-style-type: none"> Bestätigungsbias von falschen algorithmischen Prädiktionen (Label-richtigkeit?) Zusätzlicher Arbeitsschritt (Zeitaufwand) 	<ul style="list-style-type: none"> Bestätigungsbias von falschen algorithmischen Prädiktionen (Label-richtigkeit? Label-vollständigkeit?) Möglicherweise geringe Expertise der Annotatoren/innen (Crowd-sourcing)
Maßnahmen zur Verbesserung des entstandenen Datensatzes	<ul style="list-style-type: none"> Adäquate Software (Screening-Modus etc.) Multi-Experten-Konsensus Wiederholtes Screening der Bilder Hohe Sorgfältigkeit und Konzentration Standardisierte Labelingkriterien 	<ul style="list-style-type: none"> Hohe Sensitivität des Algorithmus für das Auffinden von verpassten Kandidaten Sorgfältige Überprüfung durch einen/multiple erfahrenen Experten/innen (hohe Spezifität) Multi-Experten Konsensus 	<ul style="list-style-type: none"> Sorgfältige Überprüfung durch einen/multiple Experten/innen (hohe Spezifität) Hohe Leistung der verwendeten Algorithmen Verwendung von Crowd-Sourcing ausschließlich für „einfache“ Fragestellungen

5 Zusammenfassung

Erforschung und Validierung der digitalen Mikroskopie und automatisierten Bildanalysen für die Diagnostik und Lehre in der Veterinärpathologie und -zytologie

Moderne Whole-Slide-Image (WSI) Scanner sind in der Lage eine Digitalisierung von gesamten mikroskopischen Gewebeproben mit hoher Auflösung durchzuführen und ermöglichen somit eine breite Anwendung der digitalen Pathologie für die Diagnostik, Lehre und Forschung. Die Betrachtung des WSI am Bildschirm durch einen Menschen (digitale Mikroskopie) findet bereits routinemäßige Anwendung für die Untersuchung von Gewebeproben in Laboren und in der Ausbildung von (Tier-)Medizinstudierenden unter anderem aufgrund einer verbesserten Arbeits- beziehungsweise Lernflexibilität. Hingegen werden automatisierte Analysen von WSIs durch Computersoftware (automatisierte Bildanalyse) bisher vorwiegend in Forschungsinstitutionen eingesetzt, obwohl Bildanalysealgorithmen unter Verwendung von künstlicher Intelligenz – insbesondere Deep Learning – auch für die Diagnostik ein sehr großes Potential haben, die Effizienz, Reproduzierbarkeit und Genauigkeit der Befunde zu verbessern. Allerdings lagen vor meinem PhD-Projekt unzureichende Erkenntnisse für die Tiermedizin hinsichtlich der diagnostischen Sicherheit der digitalen Mikroskopie, des Meinungsbildes von Studierenden in Abhängigkeit vom Erfahrungslevel und der Anwendungsmöglichkeiten der automatisierten Bildanalyse für diagnostische Fragestellungen in der Histopathologie und Zytologie vor. Daher haben sich die eigenen Untersuchungen mit einem breiten Anwendungsspektrum der digitalen Pathologie für die Veterinärpathologie und -zytologie beschäftigt: 1) Validierung der digitalen Mikroskopie für die Diagnose von Hauttumoren des Hundes, 2) Erfassung des Meinungsbildes von Studierenden drei verschiedener Jahrgangsstufen und 3) Entwicklung und Evaluation von Bildanalysesoftware unter Verwendung von Deep Learning für die Objekterkennung von histologischen Prognoseparametern in kaninen Tumoren sowie von verschiedenen Zelltypen in zytologischen Präparaten von Lungenspülproben des Pferdes.

Unsere Validierungsstudie war die erste für die Tierpathologie. Wir konnten zeigen, dass die Untersuchung von diversen kaninen Hauttumoren mittels der digitalen Mikroskopie nicht unterlegen ist im Vergleich zu der Untersuchung mittels der Lichtmikroskopie. Allerdings wurde in der Tumorguppe der Rundzelltumore und dem Tumorgrading der Mastzelltumore eine geringfügig niedrigere Konkordanzrate für die digitale Mikroskopie festgestellt. Die durchschnittliche Zeit für die Diagnosestellung war hingegen signifikant kürzer mittels digitaler

Mikroskopie. Zusätzliche Validierungsstudien für weitere Anwendungsgebiete sind notwendig, um eine generelle Aussage über die diagnostische Sicherheit der digitalen Mikroskopie treffen zu können.

Die eigene Meinungsumfrage über die Anwendung der digitalen Mikroskopie im Histopathologie-Curriculum hat erstmals das Meinungsbild der Studierenden aus drei aufeinanderfolgenden Jahrgängen erfasst. Unabhängig vom Semester wurde die digitale Mikroskopie als sehr positiv für das Selbststudium und als Demonstrationstool in Vorlesungen bewertet. Allerdings ergaben sich Unterschiede im Meinungsbild hinsichtlich der größten Nachteile: Studierende aus den früheren Semestern befürchteten vor allem eine mangelnde Ausbildung in der Lichtmikroskopie, während Studierende aus dem höchsten Semester vorwiegend eine unzureichende Ladegeschwindigkeit der WSIs (bedingt durch eine unzureichende Internetverbindung) kritisierten. Die digitale Mikroskopie ist ein geschätztes Instrument für die Histopathologieausbildung von Tiermedizinierenden aller Jahrgangsstufen mit bestehenden technischen Limitationen. Allerdings wird das Erlernen der Lichtmikroskopie weiterhin als wichtiger Bestandteil des Curriculums angesehen.

Die eigenen Deep learning-basierten Bildanalysealgorithmen, die zusammen mit Forschern der Friedrich-Alexander-Universität Erlangen-Nürnberg und der Technischen Hochschule Ingolstadt entwickelt wurden, haben durchweg eine sehr hohe Leistungsfähigkeit (vergleichbar oder besser als zahlreiche Experten/innen) erreicht. Für derartige Ergebnisse sind qualitativ und quantitativ hochwertige Datensätze notwendig, welche wir für die verschiedenen histopathologischen und zytologischen Fragestellungen generieren konnten. Die meisten unserer Datensätze wurden öffentlich zugänglich gemacht, sodass diese für die weitere Entwicklung der künstlichen Intelligenz durch andere Arbeitsgruppen zur Verfügung stehen. Weiterhin haben wir erforscht, wie die Qualität und Quantität von Datensätzen durch algorithmisch-assistierte Labelingmethoden und dem Multi-Experten-Konsensus verbessert werden kann. Neben Algorithmen für die Quantifizierung von Hämosiderophagen und Zelltypen in zytologischen Präparaten von Lungenspülflüssigkeiten des Pferdes sowie die Erkennung von zwei- und mehrkernigen Tumorzellen in histologischen Schnitten war unser Forschungsschwerpunkt die Erkennung von Mitosefiguren in histologischen Tumorschnitten. Alle Ansätze haben das Potential gezeigt, die Diagnosefähigkeiten bzw. die diagnostische Effizienz der/des Pathologin/en sinnvoll zu unterstützen. Auf dem Gebiet der Mitosefiguren konnten wir beispielweise zeigen, dass unsere Algorithmen durchschnittlich eine bessere Auswahl an Regionen mit der höchsten Mitosedichte im Gewebeschnitt treffen konnten. Bildanalysen können – im Unterschied zu Pathologen/innen – die Informationen des gesamten WSI innerhalb kurzer Zeit verarbeiten und bieten somit Vorteile für eine Regionauswahl.

6 Summary

Exploration and Validation of Digital Pathology and Automated Image Analysis for Diagnostic and Teaching Purposes in Veterinary Pathology and Cytology

Modern whole slide image (WSI) scanners are capable of quickly digitizing entire microscopic tissue samples with high resolution, thereby enabling a wide field of application including diagnostic service, university curriculum and research. Interactive viewing of WSIs at a computerized workstation by a person (digital microscopy) is already routinely used for examination of tissue samples at diagnostic laboratories and for teaching of veterinary students due to the improved workflow and learning flexibility (respectively). On the contrary, computerized analysis of WSIs with computer software (automated image analysis) is currently almost exclusively restricted to research, regardless of its high potential to improve efficiency, reproducibility and accuracy of obtaining diagnoses, especially if artificial intelligence (deep learning) is used. Review of the published veterinary literature revealed a dearth of information for veterinary medicine regarding the diagnostic performance of digital microscopy, utilization for teaching students with different experience levels and possible applications of automated image analysis for histopathologic and cytologic diagnostic tasks. Therefore, the investigations of my PhD project included a wide scope of digital pathology for veterinary pathology and cytology: 1) Validation of digital microscopy for diagnosis of canine skin tumors, 2) capturing the trends of student opinions from three consecutive academic years and 3) development and evaluation of image analysis software using deep learning for object detection of histologic prognostic parameters in canine tumors as well as different cell types in cytologic specimens of equine bronchoalveolar fluids.

Our validation study was the first of its kind in veterinary pathology. It showed that digital microscopy is an acceptable modality for the diagnosis of numerous canine skin tumor types in comparison to examination with traditional light microscopy. However, diagnosis of round cell tumors and two-tiered grading of mast cell tumors had slightly lower concordance rates if examined with digital microscopy. The average diagnostic time was significantly shorter with examination of WSIs. Further validation studies are necessary for additional applications of digital microscopy in order to obtain sufficient evidence of the general diagnostic capability of digital microscopy for routine use in veterinary diagnostic laboratories.

Our survey regarding use of digital microscopy in the histopathology curriculum was the first to evaluate the trends of student opinions during three consecutive academic years.

Independent of the academic year, digital microscopy was perceived very positively overall for self-study purposes and for demonstrating concepts during lectures. However, a significant difference in opinion between the three academic years was noted regarding the perceived disadvantages. While students from the two earlier academic years were mostly concerned about deficiencies in obtaining light microscopy skills, students of the highest academic year mostly criticized insufficient transmission speeds (related to the available internet connection). Digital microscopy was perceived as a valuable tool for teaching histopathology with some technical limitations. Regardless, veterinary students considered learning of light microscopy skills to be an important aspect of their curriculum.

Our deep learning-based image analysis algorithms, which were created in collaboration with the Pattern Recognition Lab from the Friedrich-Alexander-Universität Erlangen-Nürnberg and the Faculty of Computer Science of the Technische Hochschule Ingolstadt, were able to achieve a very high performance (comparable to or even better than human experts). For such a high performance it is necessary to have access to high quality and high quantity datasets, which we were able to create for the histopathologic and cytologic tasks investigated. Most of our datasets are publicly available in order to foster further development of artificial intelligence approaches by other research groups. Furthermore, we investigated how the quality and quantity of datasets can be improved by algorithmic-assisted labeling approaches and multi-expert consensus on labels. Our main research focus was to use these algorithms for object detection of mitotic figures in histologic tumor sections. We also developed algorithms for quantification of hemosiderophages and cell types in cytologic specimens of bronchoalveolar lavage fluid of horses and for the detection of bi- and multinucleated tumor cells in histologic specimens of canine mast cell tumors. All developed algorithms showed the potential to support and improve diagnostic capabilities and/or efficiency of pathologists. For example, we proved that our deep learning-based algorithms typically outperform pathologists in finding the tumor area with the highest mitotic activity (an important part of tumor evaluation). As opposed to pathologists, algorithms have the capability to quickly and efficiently process the information captured within an entire WSIs and therefore offer an advantage for detecting regions of interest.

7 Literaturverzeichnis

- Abel JT, Ouillette P, Williams CL, Blau J, Cheng J, Yao K, Lee WY, Cornish TC, Balis UGJ und McClintock DS (2020): Display Characteristics and Their Impact on Digital Pathology: A Current Review of Pathologists' Future "Microscope". *J Pathol Inform*, 11, 23
- Abels E, Pantanowitz L, Aeffner F, Zarella MD, van der Laak J, Bui MM, Vemuri VN, Parwani AV, Gibbs J, Agosto-Arroyo E, Beck AH und Kozlowski C (2019): Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *J Pathol*, 249(3), 286-294
- Aeffner F, Wilson K, Bolon B, Kanaly S, Mahrt CR, Rudmann D, Charles E und Young GD (2016): Commentary: Roles for Pathologists in a High-throughput Image Analysis Team. *Toxicol Pathol*, 44(6), 825-834
- Aeffner F, Wilson K, Martin NT, Black JC, Hendriks CLL, Bolon B, Rudmann DG, Gianani R, Koegler SR, Krueger J und Young GD (2017): The Gold Standard Paradox in Digital Image Analysis: Manual Versus Automated Scoring as Ground Truth. *Arch Pathol Lab Med*, 141(9), 1267-1275
- Aeffner F, Adissu HA, Boyle MC, Cardiff RD, Hagendorn E, Hoenerhoff MJ, Klopfleisch R, Newbigging S, Schaudien D, Turner O und Wilson K (2018): Digital Microscopy, Image Analysis, and Virtual Slide Repository. *Ilar j*, 59(1), 66-79
- Akram SU, Qaiser T, Graham S, Kannala J, Heikkilä J und Rajpoot N (2018): Leveraging unlabeled whole-slide-images for mitosis detection. In: *Computational Pathology and Ophthalmic Medical Image Analysis*, Springer, pp. 69-77
- Al-Janabi S, Huisman A, Nap M, Clarijs R und van Diest PJ (2012a): Whole slide images as a platform for initial diagnostics in histopathology in a medium-sized routine laboratory. *J Clin Pathol*, 65(12), 1107-1111
- Al-Janabi S, Huisman A, Vink A, Leguit RJ, Offerhaus GJ, ten Kate FJ und van Diest PJ (2012b): Whole slide images for primary diagnostics of gastrointestinal tract pathology: a feasibility study. *Hum Pathol*, 43(5), 702-707
- Al-Janabi S, Huisman A, Vink A, Leguit RJ, Offerhaus GJ, Ten Kate FJ, van Dijk MR und van Diest PJ (2012c): Whole slide images for primary diagnostics in dermatopathology: a feasibility study. *J Clin Pathol*, 65(2), 152-158
- Al-Janabi S, Huisman A, Willems SM und Van Diest PJ (2012d): Digital slide images for primary diagnostics in breast pathology: a feasibility study. *Hum Pathol*, 43(12), 2318-2325
- Al-Janabi S, van Slooten HJ, Visser M, van der Ploeg T, van Diest PJ und Jiwa M (2013): Evaluation of mitotic activity index in breast cancer using whole slide digital images. *PLoS One*, 8(12), e82576

- Al-Janabi S, Huisman A, Jonges GN, Ten Kate FJ, Goldschmeding R und van Diest PJ (2014): Whole slide images for primary diagnostics of urinary system pathology: a feasibility study. *J Renal Inj Prev*, 3(4), 91-96
- Amend SR, Torga G, Lin KC, KostECKa LG, de Marzo A, Austin RH und Pienta KJ (2019): Polyploid giant cancer cells: Unrecognized actuators of tumorigenesis, metastasis, and resistance. *Prostate*, 79(13), 1489-1497
- American College of Veterinary Pathologists (2021): Pathology Informatics Terminology Reference. Abgerufen am: 31.01.2021, von https://www.acvp.org/page/pathology_informatics
- Araújo ALD, Arboleda LPA, Palmier NR, Fonsêca JM, de Pauli Paglioni M, Gomes-Silva W, Ribeiro ACP, Brandão TB, Simonato LE, Speight PM, Fonseca FP, Lopes MA, de Almeida OP, Vargas PA, Madrid Troconis CC und Santos-Silva AR (2019): The performance of digital microscopy for primary diagnosis in human pathology: a systematic review. *Virchows Arch*, 474(3), 269-287
- Aresta G, Araújo T, Kwok S, Chennamsetty SS, Safwan M, Alex V, Marami B, Prastawa M, Chan M, Donovan M, Fernandez G, Zeineh J, Kohl M, Walz C, Ludwig F, Braunewell S, Baust M, Vu QD, To MNN, Kim E, Kwak JT, Galal S, Sanchez-Freire V, Brancati N, Frucci M, Riccio D, Wang Y, Sun L, Ma K, Fang J, Kone I, Boulmane L, Campilho A, Eloy C, Polónia A und Aguiar P (2019): BACH: Grand challenge on breast cancer histology images. *Med Image Anal*, 56, 122-139
- Aubreville M, Krappmann M, Bertram C, Klopffleisch R und Maier A (2017): A guided spatial transformer network for histology cell differentiation. In: *Eurographics Workshop on Visual Computing for Biology and Medicine*, I. Hotz, D. Merhof and C. Rieder, Eds, Eurographics Proceedings, pp. 21-25
- Aubreville M, Bertram C, Klopffleisch R und Maier A (2018): SlideRunner - A Tool for Massive Cell Annotations in Whole Slide Images. In: *Bildverarbeitung für die Medizin 2018*, A. Maier, T. Deserno, H. Handels, K. Maier-Hein, C. Palm and T. Tolxdorff, Eds, Springer Vieweg, pp. 309-314
- Aubreville M, Bertram CA, Klopffleisch R und Maier A (2019a): Augmented mitotic cell count using field of interest proposal. In: *Bildverarbeitung für die Medizin 2019*, H. Handels, T. Deserno, A. Maier, K. Maier-Hein, C. Palm and T. Tolxdorff, Eds, Springer Vieweg, pp. 321-326
- Aubreville M, Bertram CA, Klopffleisch R und Maier A (2019b): Field Of Interest Proposal for Augmented Mitotic Cell Count: A Comparison of Two Networks. In: *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies* pp. 30-37
- Aubreville M (2020): Computer-Aided Tumor Diagnosis of Microscopy Images; Dissertation an der Technischen Fakultät der Friedrich-Alexander-Universität Erlangen-Nürnberg.

Abgerufen am: 31.01.2021, von <https://opus4.kobv.de/opus4-fau/frontdoor/index/index/docId/13755>

- Aubreville M, Bertram CA, Donovan TA, Marzahl C, Maier A und Klopffleisch R (2020a): A completely annotated whole slide image dataset of canine breast cancer to aid human breast cancer research. *Sci Data*, 7(1), 417
- Aubreville M, Bertram CA, Jabari S, Marzahl C, Klopffleisch R und Maier A (2020b): Inter-Species, Inter-Tissue Domain Adaptation for Mitotic Figure Assessment - Learning New Tricks from Old Dogs. In: *Bildverarbeitung für die Medizin 2020*, T. Tolxdorff, T. Deserno, H. Handels, A. Maier, K. Maier-Hein and C. Palm, Eds, Springer Vieweg, pp. 1-7
- Aubreville M, Bertram CA, Marzahl C, Gurtner C, Dettwiler M, Schmidt A, Bartenschlager F, Merz S, Fragoso M, Kershaw O, Klopffleisch R und Maier A (2020c): Deep learning algorithms out-perform veterinary pathologists in detecting the mitotically most active tumor region. *Sci Rep*, 10(1), 16447
- Awaysheh A, Wilcke J, Elvinger F, Rees L, Fan W und Zimmerman KL (2019): Review of Medical Decision Support and Machine-Learning Methods. *Vet Pathol*, 56(4), 512-525
- Azam AS, Miligy IM, Kimani PK, Maqbool H, Hewitt K, Rajpoot NM und Snead DRJ (2020): Diagnostic concordance and discordance in digital pathology: a systematic review and meta-analysis. *J Clin Pathol*,
- Baidoshvili A, Bucur A, van Leeuwen J, van der Laak J, Kluin P und van Diest PJ (2018): Evaluating the benefits of digital pathology implementation: time savings in laboratory logistics. *Histopathology*, 73(5), 784-794
- Balkenhol MCA, Tellez D, Vreuls W, Clahsen PC, Pinckaers H, Ciompi F, Bult P und van der Laak J (2019): Deep learning assisted mitotic counting for breast cancer. *Lab Invest*, 99(11), 1596-1606
- Bauer TW, Schoenfield L, Slaw RJ, Yerian L, Sun Z und Henricks WH (2013): Validation of whole slide imaging for primary diagnosis in surgical pathology. *Arch Pathol Lab Med*, 137(4), 518-524
- Bejnordi BE, Veta M, van Diest PJ, van Ginneken B, Karssemeijer N, Litjens G, van der Laak J, Hermsen M, Manson QF, Balkenhol M, Geessink O, Stathonikos N, van Dijk MC, Bult P, Beca F, Beck AH, Wang D, Khosla A, Gargeya R, Irshad H, Zhong A, Dou Q, Li Q, Chen H, Lin HJ, Heng PA, Haß C, Bruni E, Wong Q, Halici U, Öner M, Cetin-Atalay R, Berseth M, Khvatkov V, Vylegzhanin A, Kraus O, Shaban M, Rajpoot N, Awan R, Sirinukunwattana K, Qaiser T, Tsang YW, Tellez D, Annuschein J, Hufnagl P, Valkonen M, Kartasalo K, Latonen L, Ruusuvoori P, Liimatainen K, Albarqouni S, Mungal B, George A, Demirci S, Navab N, Watanabe S, Seno S, Takenaka Y, Matsuda H, Phoulady HA, Kovalev V, Kalinovsky A, Liauchuk V, Bueno G, Fernandez-Carrobles MM, Serrano I, Deniz O, Racoceanu D und Venâncio R (2017):

- Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *Jama*, 318(22), 2199-2210
- Bera K, Schalper KA, Rimm DL, Velcheti V und Madabhushi A (2019): Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol*, 16(11), 703-715
- Bernard C, Chandrakanth SA, Cornell IS, Dalton J, Evans A, Garcia BM, Godin C, Godlewski M, Jansen GH, Kabani A, Louahia S, Manning L, Maung R, Moore L, Philley J, Slatnik J, Srigley J, Thibault A, Picard DD, Cracower H und Tetu B (2014): Guidelines from the Canadian Association of Pathologists for establishing a telepathology service for anatomic pathology using whole-slide imaging. *J Pathol Inform*, 5(1), 15
- Bertram CA und Klopffleisch R (2017): The Pathologist 2.0: An Update on Digital Pathology in Veterinary Medicine. *Vet Pathol*, 54(5), 756-766
- Bertram CA, Dietert K, Pieper L, Erickson NA, Barton AK und Klopffleisch R (2018a): Effects of on-slide fixation on the cell quality of cytocentrifuged equine bronchioalveolar lavage fluid. *Vet Clin Pathol*, 47(3), 513-519
- Bertram CA, Firsching T und Klopffleisch R (2018b): Virtual Microscopy in Histopathology Training: Changing Student Attitudes in 3 Successive Academic Years. *J Vet Med Educ*, 45(2), 241-249
- Bertram CA, Gurtner C, Dettwiler M, Kershaw O, Dietert K, Pieper L, Pischon H, Gruber AD und Klopffleisch R (2018c): Validation of Digital Microscopy Compared With Light Microscopy for the Diagnosis of Canine Cutaneous Tumors. *Vet Pathol*, 55(4), 490-500
- Bertram CA, Aubreville M, Maier A und Klopffleisch R. (2019a). Histopathological Datasets – An Elementary Step in the Development of Deep Learning Based Algorithms for Automated Image Analysis. *17th European Congress of Toxicologic Pathology*, Cologne
- Bertram CA, Aubreville M, Marzahl C, Maier A und Klopffleisch R (2019b): A large-scale dataset for mitotic figure assessment on whole slide images of canine cutaneous mast cell tumor. *Sci Data*, 6(1), 274
- Bertram CA und Klopffleisch R. (2019). Glass versus Digital –Validation of Digital Microscopy in Veterinary Pathology. *17th European Congress of Toxicologic Pathology*, Cologne
- Bertram CA, Aubreville M, Gurtner C, Bartel A, Corner SM, Dettwiler M, Kershaw O, Noland EL, Schmidt A, Sledge DG, Smedley RC, Thaiwong T, Kiupel M, Maier A und Klopffleisch R (2020a): Computerized Calculation of Mitotic Count Distribution in Canine Cutaneous Mast Cell Tumor Sections: Mitotic Count Is Area Dependent. *Vet Pathol*, 57(2), 214-226

- Bertram CA, Veta M, Marzahl C, Stathonikos N, Maier A, Klopffleisch R und Aubreville M (2020b): Are pathologist-defined labels reproducible? Comparison of the TUPAC16 mitotic figure dataset with an alternative set of labels. In: *iMIMIC 2020/MIL3iD 2020/LABELS 2020, LNCS 12446. Interpretable and Annotation-Efficient Learning for Medical Image Computing*, J. C. e. al., Ed, Springer Nature Switzerland AG, pp. 204-213
- Bertram CA, Donovan TA, Tecilla M, Bartenschlager F, Fragoso M, Wilm F, Marzahl C, Breininger K, Maier A, Klopffleisch R und Aubreville M (2021): Dataset on Bi- and Multi-Nucleated Tumor Cells in Canine Cutaneous Mast Cell Tumors. In: *Bildverarbeitung für die Medizin 2021*, P. e. al., Ed, Springer Fachmedien Wiesbaden GmbH, pp. 134-139
- Blanchet CJK, Fish EJ, Miller AG, Snyder LA, Labadie JD und Avery PR (2019): Evaluation of Region of Interest Digital Cytology Compared to Light Microscopy for Veterinary Medicine. *Vet Pathol*, 56(5), 725-731
- Bonsembiante F, Martini V, Bonfanti U, Casarin G, Trez D und Gelain ME (2018): Cytomorphological description and intra-observer agreement in whole slide imaging for canine lymphoma. *Vet J*, 236, 96-101
- Bonsembiante F, Bonfanti U, Cian F, Cavicchioli L, Zattoni B und Gelain ME (2019): Diagnostic Validation of a Whole-Slide Imaging Scanner in Cytological Samples: Diagnostic Accuracy and Comparison With Light Microscopy. *Vet Pathol*, 56(3), 429-434
- Borowsky AD, Glassy EF, Wallace WD, Kallichanda NS, Behling CA, Miller DV, Oswal HN, Feddersen RM, Bakhtar OR, Mendoza AE, Molden DP, Saffer HL, Wixom CR, Albro JE, Cessna MH, Hall BJ, Lloyd IE, Bishop JW, Darrow MA, Gui D, Jen KY, Walby JAS, Bauer SM, Cortez DA, Gandhi P, Rodgers MM, Rodriguez RA, Martin DR, McConnell TG, Reynolds SJ, Spigel JH, Stepenaskie SA, Viktorova E, Magari R, Wharton KA, Qiu J und Bauer TW (2020): Digital Whole Slide Imaging Compared With Light Microscopy for Primary Diagnosis in Surgical Pathology. *Arch Pathol Lab Med*, 144(10), 1245-1253
- Braun MW und Kearns KD (2008): Improved learning efficiency and increased student collaboration through use of virtual microscopy in the teaching of human pathology. *Anat Sci Educ*, 1(6), 240-246
- Brick KE, Sluzevich JC, Cappel MA, DiCaudo DJ, Comfere NI und Wieland CN (2013): Comparison of virtual microscopy and glass slide microscopy among dermatology residents during a simulated in-training examination. *J Cutan Pathol*, 40(9), 807-811
- Brooker AJ, Krimer PM, Meichner K und Garner BC (2019): Impact of photographer experience and number of images on telecytology accuracy. *Vet Clin Pathol*, 48(3), 419-424
- Brown PJ, Fews D und Bell NJ (2016): Teaching Veterinary Histopathology: A Comparison of Microscopy and Digital Slides. *J Vet Med Educ*, 43(1), 13-20

- Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, van der Laak J, Hulsbergen-van de Kaa C und Litjens G (2020): Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol*, 21(2), 233-241
- Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, Brogi E, Reuter VE, Klimstra DS und Fuchs TJ (2019): Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med*, 25(8), 1301-1309
- Campbell WS, Hinrichs SH, Lele SM, Baker JJ, Lazenby AJ, Talmon GA, Smith LM und West WW (2014): Whole slide imaging diagnostic concordance with light microscopy for breast needle biopsies. *Hum Pathol*, 45(8), 1713-1721
- Campbell WS, Talmon GA, Foster KW, Baker JJ, Smith LM und Hinrichs SH (2015): Visual memory effects on intraoperator study design: determining a minimum time gap between case reviews to reduce recall bias. *Am J Clin Pathol*, 143(3), 412-418
- Camus MS, Priest HL, Koehler JW, Driskell EA, Rakich PM, Ilha MR und Krimer PM (2016): Cytologic Criteria for Mast Cell Tumor Grading in Dogs With Evaluation of Clinical Outcome. *Vet Pathol*, 53(6), 1117-1123
- Canadas A, França M, Pereira C, Vilaça R, Vilhena H, Tinoco F, Silva MJ, Ribeiro J, Medeiros R, Oliveira P, Dias-Pereira P und Santos M (2019): Canine Mammary Tumors: Comparison of Classification and Grading Methods in a Survival Study. *Vet Pathol*, 56(2), 208-219
- Center SA, McDonough SP und Bogdanovic L (2013): Digital image analysis of rhodanine-stained liver biopsy specimens for calculation of hepatic copper concentrations in dogs. *Am J Vet Res*, 74(12), 1474-1480
- Chen J, Niu N, Zhang J, Qi L, Shen W, Donkena KV, Feng Z und Liu J (2019): Polyploid Giant Cancer Cells (PGCCs): The Evil Roots of Cancer. *Curr Cancer Drug Targets*, 19(5), 360-367
- Cheng CL, Azhar R, Sng SH, Chua YQ, Hwang JS, Chin JP, Seah WK, Loke JC, Ang RH und Tan PH (2016): Enabling digital pathology in the diagnostic setting: navigating through the implementation journey in an academic medical centre. *J Clin Pathol*, 69(9), 784-792
- Chong Y, Kim DC, Jung CK, Kim DC, Song SY, Joo HJ und Yi SY (2020): Recommendations for pathologic practice using digital pathology: consensus report of the Korean Society of Pathologists. *J Pathol Transl Med*, 54(6), 437-452
- Clarke EL und Treanor D (2017): Colour in digital pathology: a review. *Histopathology*, 70(2), 153-163

- Courtiol P, Maussion C, Moarii M, Pronier E, Pilcer S, Sefta M, Manceron P, Toldo S, Zaslavskiy M, Le Stang N, Girard N, Elemento O, Nicholson AG, Blay JY, Galateau-Sallé F, Wainrib G und Clozel T (2019): Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nat Med*, 25(10), 1519-1525
- Dee FR (2009): Virtual microscopy in pathology education. *Hum Pathol*, 40(8), 1112-1121
- Digital Pathology Association (2021): Glossary of Terms. Abgerufen am: 31.01.2021, von https://digitalpathologyassociation.org/glossary-of-terms_1
- Donovan TA, Moore FM, Bertram CA, Luong R, Bolfa P, Klopffleisch R, Tvedten H, Salas EN, Whitley DB, Aubreville M und Meuten DJ (2020): Mitotic Figures-Normal, Atypical, and Imposters: A Guide to Identification. *Vet Pathol*, 58(2), 243-257
- Doucet MY und Viel L (2002): Alveolar macrophage graded hemosiderin score from bronchoalveolar lavage in horses with exercise-induced pulmonary hemorrhage and controls. *J Vet Intern Med*, 16(3), 281-286
- Eccher A und Girolami I (2020): Current state of whole slide imaging use in cytopathology: Pros and pitfalls. *Cytopathology*, 31(5), 372-378
- Elmore JG, Longton GM, Pepe MS, Carney PA, Nelson HD, Allison KH, Geller BM, Onega T, Tosteson AN, Mercan E, Shapiro LG, Brunyé TT, Morgan TR und Weaver DL (2017): A Randomized Study Comparing Digital Imaging to Traditional Glass Slide Microscopy for Breast Biopsy and Cancer Diagnosis. *J Pathol Inform*, 8, 12
- Elston LB, Sueiro FA, Cavalcanti JN und Metze K (2009): The importance of the mitotic index as a prognostic factor for survival of canine cutaneous mast cell tumors: a validation study. *Vet Pathol*, 46(2), 362-364, author reply 364-365
- Evans AJ, Salama ME, Henricks WH und Pantanowitz L (2017): Implementation of Whole Slide Imaging for Clinical Purposes: Issues to Consider From the Perspective of Early Adopters. *Arch Pathol Lab Med*, 141(7), 944-959
- Evans SJM, Moore AR, Olver CS, Avery PR und West AB (2020): Virtual Microscopy Is More Effective Than Conventional Microscopy for Teaching Cytology to Veterinary Students: A Randomized Controlled Trial. *J Vet Med Educ*, 47(4), 475-481
- Farahani N, Parwani AV und Pantanowitz L (2015): Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. *Pathol Lab Med Int*, 7(23-33), 4321
- Fraggetta F, Yagi Y, Garcia-Rojo M, Evans AJ, Tuthill JM, Baidoshvili A, Hartman DJ, Fukuoka J und Pantanowitz L (2018): The Importance of eSlide Macro Images for Primary Diagnosis with Whole Slide Imaging. *J Pathol Inform*, 9, 46

- García-Rojo M (2016): International Clinical Guidelines for the Adoption of Digital Pathology: A Review of Technical Aspects. *Pathobiology*, 83(2-3), 99-109
- Girolami I, Pantanowitz L, Marletta S, Brunelli M, Mescoli C, Parisi A, Barresi V, Parwani A, Neil D, Scarpa A, Rossi ED und Eccher A (2020): Diagnostic concordance between whole slide imaging and conventional light microscopy in cytopathology: A systematic review. *Cancer Cytopathol*, 128(1), 17-28
- Goacher E, Randell R, Williams B und Treanor D (2017): The Diagnostic Concordance of Whole Slide Imaging and Light Microscopy: A Systematic Review. *Arch Pathol Lab Med*, 141(1), 151-161
- Goldschmidt M, Peña L, Rasotto R und Zappulli V (2011): Classification and grading of canine mammary tumors. *Vet Pathol*, 48(1), 117-131
- Guo Z, Liu H, Ni H, Wang X, Su M, Guo W, Wang K, Jiang T und Qian Y (2019): A Fast and Refined Cancer Regions Segmentation Framework in Whole-slide Breast Pathological Images. *Sci Rep*, 9(1), 882
- Hanna MG, Pantanowitz L und Evans AJ (2015): Overview of contemporary guidelines in digital pathology: what is available in 2015 and what still needs to be addressed? *J Clin Pathol*, 68(7), 499-505
- Hanna MG, Reuter VE, Hameed MR, Tan LK, Chiang S, Sigel C, Hollmann T, Giri D, Samboy J, Moradel C, Rosado A, Otilano JR, 3rd, England C, Corsale L, Stamelos E, Yagi Y, Schüffler PJ, Fuchs T, Klimstra DS und Sirintrapun SJ (2019): Whole slide imaging equivalency and efficiency study: experience at a large academic center. *Mod Pathol*, 32(7), 916-928
- Hanna MG, Reuter VE, Ardon O, Kim D, Sirintrapun SJ, Schüffler PJ, Busam KJ, Sauter JL, Brogi E, Tan LK, Xu B, Bale T, Agaram NP, Tang LH, Ellenson LH, Philip J, Corsale L, Stamelos E, Friedlander MA, Ntiamoah P, Labasin M, England C, Klimstra DS und Hameed M (2020): Validation of a digital pathology system including remote review during the COVID-19 pandemic. *Mod Pathol*, 33(11), 2115-2127
- Ho J, Parwani AV, Jukic DM, Yagi Y, Anthony L und Gilbertson JR (2006): Use of whole slide imaging in surgical pathology quality assurance: design and pilot validation studies. *Hum Pathol*, 37(3), 322-331
- Hufnagl P, Zwönitzer R und Haroske G (2018): Guidelines Digital Pathology for Diagnosis on (and Reports of) Digital Images Version 1.0 Bundesverband deutscher Pathologen eV (Federal Association of German Pathologist). *Diagn Pathol*, 4(1)
- Huisman A, Looijen A, van den Brink SM und van Diest PJ (2010): Creation of a fully digital pathology slide archive by high-volume tissue slide scanning. *Hum Pathol*, 41(5), 751-757

- Irshad H (2013): Automated mitosis detection in histopathology using morphological and multi-channel statistics features. *J Pathol Inform*, 4, 10
- Kather JN, Heij LR, Grabsch HI, Loeffler C, Echle A, Muti HS, Krause J, Niehues JM, Sommer KA und Bankhead P (2020): Pan-cancer image-based detection of clinically actionable genetic alterations. *Nature Cancer*, 1(8), 789-799
- Kent MN, Olsen TG, Feeser TA, Tesno KC, Moad JC, Conroy MP, Kendrick MJ, Stephenson SR, Murchland MR, Khan AU, Peacock EA, Brumfiel A und Bottomley MA (2017): Diagnostic Accuracy of Virtual Pathology vs Traditional Microscopy in a Large Dermatopathology Study. *JAMA Dermatol*, 153(12), 1285-1291
- Khan AM, El-Daly H, Simmons E und Rajpoot NM (2013): HyMaP: A hybrid magnitude-phase approach to unsupervised segmentation of tumor areas in breast cancer histology images. *J Pathol Inform*, 4(Suppl), S1
- Kim D, Pantanowitz L, Schüffler P, Yarlagadda DVK, Ardon O, Reuter VE, Hameed M, Klimstra DS und Hanna MG (2020): (Re) Defining the High-Power Field for Digital Pathology. *J Pathol Inform*, 11, 33
- Kim YJ, Jang H, Lee K, Park S, Min SG, Hong C, Park JH, Lee K, Kim J, Hong W, Jung H, Liu Y, Rajkumar H, Khened M, Krishnamurthi G, Yang S, Wang X, Han CH, Kwak JT, Ma J, Tang Z, Marami B, Zeineh J, Zhao Z, Heng PA, Schmitz R, Madesta F, Rösch T, Werner R, Tian J, Puybareau E, Bovio M, Zhang X, Zhu Y, Chun SY, Jeong WK, Park P und Choi J (2021): PAIP 2019: Liver cancer segmentation challenge. *Med Image Anal*, 67, 101854
- Kiupel M, Webster JD, Bailey KL, Best S, DeLay J, Detrisac CJ, Fitzgerald SD, Gamble D, Ginn PE, Goldschmidt MH, Hendrick MJ, Howerth EW, Janovitz EB, Langohr I, Lenz SD, Lipscomb TP, Miller MA, Misdorp W, Moroff S, Mullaney TP, Neyens I, O'Toole D, Ramos-Vara J, Scase TJ, Schulman FY, Sledge D, Smedley RC, Smith K, P WS, Southorn E, Stedman NL, Steficek BA, Stromberg PC, Valli VE, Weisbrode SE, Yager J, Heller J und Miller R (2011): Proposal of a 2-tier histologic grading system for canine cutaneous mast cell tumors to more accurately predict biological behavior. *Vet Pathol*, 48(1), 147-155
- Kiupel M und Camus M (2019): Diagnosis and Prognosis of Canine Cutaneous Mast Cell Tumors. *The Veterinary Clinics of North America. Small Animal Practice*, 49(5), 819-836
- Klopfleisch R und Bertram C (2018): Gefahr oder Chance für den tierärztlichen Beruf? *Deutsches Tierärzteblatt*, 66(8), 1078-1083
- Klopfleisch R und Bertram CA (2019): Die Zytologie von Tracheobronchialsekret (TBS) und bronchoalveolärer Lavage-Flüssigkeit (BALF) beim Pferd mit equinem Asthma - Methoden und Möglichkeiten. *Der Praktische Tierarzt*, 100(1), 56-68

- Kogan LR, Dowers KL, Cerda JR, Schoenfeld-Tacher RM und Stewart SM (2014): Virtual microscopy: a useful tool for meeting evolving challenges in the veterinary medical curriculum. *J Sci Educ Technol*, 23, 756-762
- Komura D und Ishikawa S (2018): Machine Learning Methods for Histopathological Image Analysis. *Comput Struct Biotechnol J*, 16, 34-42
- Krappmann M, Aubreville M, Maier A, Bertram C und Klopffleisch R (2018): Classification of Mitotic Cells. In: *Bildverarbeitung für die Medizin 2018*, A. Maier, T. Deserno, H. Handels, K. Maier-Hein, C. Palm and T. Tolxdorff, Eds, Springer Vieweg, pp. 245-250
- Krishnamurthy S, Mathews K, McClure S, Murray M, Gilcrease M, Albarracin C, Spinosa J, Chang B, Ho J, Holt J, Cohen A, Giri D, Garg K, Bassett RL, Jr. und Liang K (2013): Multi-institutional comparison of whole slide digital imaging and optical microscopy for interpretation of hematoxylin-eosin-stained breast tissue sections. *Arch Pathol Lab Med*, 137(12), 1733-1739
- Li C, Wang X, Liu W und Latecki LJ (2018): DeepMitosis: Mitosis detection via deep detection, verification and segmentation networks. *Med Image Anal*, 45, 121-133
- Li C, Wang X, Liu W, Latecki LJ, Wang B und Huang J (2019): Weakly supervised mitosis detection in breast histopathology images using concentric loss. *Med Image Anal*, 53, 165-178
- Liu Y, Chen PC, Krause J und Peng L (2019): How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. *Jama*, 322(18), 1806-1816
- Luong RH (2020): Commentary: Digital histopathology in a private or commercial diagnostic veterinary laboratory. *J Vet Diagn Invest*, 32(3), 353-355
- Maier A, Syben C, Lasser T und Riess C (2019): A gentle introduction to deep learning in medical image processing. *Z Med Phys*, 29(2), 86-101
- Maiolino P, Restucci B, Papparella S und De Vico G (2006): Evaluation of static telepathology in veterinary diagnostic cytology. *Vet Clin Pathol*, 35(3), 303-306
- Marée R (2017): The Need for Careful Data Collection for Pattern Recognition in Digital Pathology. *J Pathol Inform*, 8, 19
- Marzahl C, Aubreville M, Bertram CA, Gerlach S, Maier J, Voigt J, Hill J, Klopffleisch R und Maier A (2020a): Is Crowd-Algorithm Collaboration an Advanced Alternative to Crowd-Sourcing on Cytology Slides? In: *Bildverarbeitung für die Medizin 2020*, T. Tolxdorff, T. Deserno, H. Handels, A. Maier, K. Maier-Hein and C. Palm, Eds, Springer Vieweg, pp. 26-31

- Marzahl C, Aubreville M, Bertram CA, Maier J, Bergler C, Kröger C, Voigt J, Klopffleisch R und Maier A (2020b): EXACT: A collaboration toolset for algorithm-aided annotation of almost everything. *Sci Rep*, 11, 4343
- Marzahl C, Aubreville M, Bertram CA, Stayt J, Jasensky AK, Bartenschlager F, Fragoso-Garcia M, Barton AK, Elsemann S, Jabari S, Krauth J, Madhu P, Voigt J, Hill J, Klopffleisch R und Maier A (2020c): Deep Learning-Based Quantification of Pulmonary Hemosiderophages in Cytology Slides. *Sci Rep*, 10(1), 9795
- Marzahl C, Bertram CA, Aubreville M, Petrick A, Weiler K, Gläsel AC, Fragoso M, Merz S, Bartenschlager F und Hoppe J (2020d): Are fast labeling methods reliable? A case study of computer-aided expert annotations on microscopy slides. In: *MICCAI 2020, LNCS 12261*, A. Martel, Ed, Springer Nature Switzerland AG, pp. 24-32
- Marzahl C, Bertram CA, Wilm F, Voigt J, Barton AK, Klopffleisch R, Breininger K, Maier A und Aubreville M (2021): Cell Detection for Asthma on Partially Annotated Whole Slide Images: Learning to be EXACT. In: *Bildverarbeitung für die Medizin 2021*, P. e. al., Ed, Springer Fachmedien Wiesbaden GmbH, pp. 147-152
- Mathew T, Kini JR und Rajan J (2020): Computational methods for automated mitosis detection in histopathology images: A review. *Biocybern Biomed Eng*,
- McAlpine ED und Michelow P (2020): The cytopathologist's role in developing and evaluating artificial intelligence in cytopathology practice. *Cytopathology*, 31(5), 385-392
- McAlpine ED, Pantanowitz L und Michelow PM (2020): Challenges Developing Deep Learning Algorithms in Cytology. *Acta Cytol*, 1-9
- Mills AM, Gradecki SE, Horton BJ, Blackwell R, Moskaluk CA, Mandell JW, Mills SE und Cathro HP (2018): Diagnostic Efficiency in Digital Pathology: A Comparison of Optical Versus Digital Assessment in 510 Surgical Pathology Cases. *Am J Surg Pathol*, 42(1), 53-59
- Mills PC, Bradley AP, Woodall PF und Wildermoth M (2007): Teaching histology to first-year veterinary science students using virtual microscopy and traditional microscopy: a comparison of student responses. *J Vet Med Educ*, 34(2), 177-182
- Molnar B, Berczi L, Diczhazy C, Tagscherer A, Varga SV, Szende B und Tulassay Z (2003): Digital slide and virtual microscopy based routine and telepathology evaluation of routine gastrointestinal biopsy specimens. *J Clin Pathol*, 56(6), 433-438
- Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Beltaifa S, Cacciabeve NG, Cathro HP, Cheng L, Cooper K, Dickey GE, Gill RM, Heaton RP, Jr., Kerstens R, Lindberg GM, Malhotra RK, Mandell JW, Manlucu ED, Mills AM, Mills SE, Moskaluk CA, Nelis M, Patil DT, Przybycin CG, Reynolds JP, Rubin BP, Saboorian MH, Salicru M, Samols MA, Sturgis CD, Turner KO, Wick MR, Yoon JY, Zhao P und Taylor CR (2018): Whole Slide Imaging Versus Microscopy for Primary Diagnosis in Surgical Pathology:

- A Multicenter Blinded Randomized Noninferiority Study of 1992 Cases (Pivotal Study). *Am J Surg Pathol*, 42(1), 39-52
- Neel JA, Grindem CB und Bristol DG (2007): Introduction and evaluation of virtual microscopy in teaching veterinary cytopathology. *J Vet Med Educ*, 34(4), 437-444
- Nielsen PS, Lindebjerg J, Rasmussen J, Starklint H, Waldstrøm M und Nielsen B (2010): Virtual microscopy: an evaluation of its validity and diagnostic performance in routine histologic diagnosis of skin tumors. *Hum Pathol*, 41(12), 1770-1776
- Ordi O, Bombí JA, Martínez A, Ramírez J, Alòs L, Saco A, Ribalta T, Fernández PL, Campo E und Ordi J (2015): Virtual microscopy in the undergraduate teaching of pathology. *J Pathol Inform*, 6, 1
- Ozlu Y, Blanco PL, Mengel M, Solez K, Halloran PF und Sis B (2012): Superiority of virtual microscopy versus light microscopy in transplantation pathology. *Clin Transplant*, 26(2), 336-344
- Pan Y, Sun Z, Wang W, Yang Z, Jia J, Feng X, Wang Y, Fang Q, Li J, Dai H, Ku C, Wang S, Liu C, Xue L, Lyu N und Zou S (2020): Automatic detection of squamous cell carcinoma metastasis in esophageal lymph nodes using semantic segmentation. *Clin Transl Med*, 10(3), e129
- Pantanowitz L, Szymas J, Yagi Y und Wilbur D (2012): Whole slide imaging for educational purposes. *J Pathol Inform*, 3, 46
- Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L, Beckwith BA, Evans AJ, Lal A und Parwani AV (2013): Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med*, 137(12), 1710-1722
- Pantanowitz L, Dickinson K, Evans AJ, Hassell LA, Henricks WH, Lennerz JK, Lowe A, Parwani AV, Riben M, Smith CD, Tuthill JM, Weinstein RS, Wilbur DC, Krupinski EA und Bernard J (2014): American Telemedicine Association clinical guidelines for telepathology. *J Pathol Inform*, 5(1), 39
- Pantanowitz L, Hartman D, Qi Y, Cho EY, Suh B, Paeng K, Dhir R, Michelow P, Hazelhurst S, Song SY und Cho SY (2020): Accuracy and efficiency of an artificial intelligence tool when counting breast mitoses. *Diagn Pathol*, 15(1), 80
- Pati P, Foncubierta-Rodríguez A, Goksel O und Gabrani M. (2020). Mitosis Detection Under Limited Annotation: A Joint Learning Approach. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*
- Plendl J, Bahramsoltani M, Gemeinhardt O, Hünigen H, Käsmeyer S und Janczyk P (2009): Active participation instead of passive behaviour opens up new vistas in education of veterinary anatomy and histology. *Anat Histol Embryol*, 38(5), 355-360

- Randell R, Ambepitiya T, Mello-Thoms C, Ruddle RA, Brettle D, Thomas RG und Treanor D (2015): Effect of display resolution on time to diagnosis with virtual pathology slides in a systematic search task. *J Digit Imaging*, 28(1), 68-76
- Rashidi HH, Tran NK, Betts EV, Howell LP und Green R (2019): Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods. *Acad Pathol*, 6, 2374289519873088
- Retamero JA, Aneiros-Fernandez J und Del Moral RG (2020): Complete Digital Pathology for Routine Histopathology Diagnosis in a Multicenter Hospital Network. *Arch Pathol Lab Med*, 144(2), 221-228
- Rodrigues-Fernandes CI, Speight PM, Khurram SA, Araújo ALD, Perez D, Fonseca FP, Lopes MA, de Almeida OP, Vargas PA und Santos-Silva AR (2020): The use of digital microscopy as a teaching method for human pathology: a systematic review. *Virchows Arch*, 477(4), 475-486
- Romansik EM, Reilly CM, Kass PH, Moore PF und London CA (2007): Mitotic index is predictive for survival for canine cutaneous mast cell tumors. *Vet Pathol*, 44(3), 335-341
- Roux L, Racoceanu D, Loménie N, Kulikova M, Irshad H, Klossa J, Capron F, Genestie C, Le Naour G und Gurcan MN (2013): Mitosis detection in breast cancer histological images An ICPR 2012 contest. *J Pathol Inform*, 4, 8
- Roux L, Racoceanu D, Capron F, Calvo J, Attieh E, Le Naour G und Gloaguen A (2014): MITOS & ATYPIA-Detection of mitosis and evaluation of nuclear atypia score in breast cancer histological images. IPAL, Agency Sci, Technol Res Inst Infocom Res, Singapore, Tech. Rep. Abgerufen am: 31.01.2021, von <https://mitos-atypia-14.grand-challenge.org/>
- Saco A, Bombi JA, Garcia A, Ramírez J und Ordi J (2016a): Current Status of Whole-Slide Imaging in Education. *Pathobiology*, 83(2-3), 79-88
- Saco A, Ramírez J, Rakislova N, Mira A und Ordi J (2016b): Validation of Whole-Slide Imaging for Histopathological Diagnosis: Current State. *Pathobiology*, 83(2-3), 89-98
- Shah KK, Lehman JS, Gibson LE, Lohse CM, Comfere NI und Wieland CN (2016): Validation of diagnostic accuracy with whole-slide imaging compared with glass slide review in dermatopathology. *J Am Acad Dermatol*, 75(6), 1229-1237
- Sims MH, Mendis-Handagama C und Moore RN (2007): Virtual microscopy in a veterinary curriculum. *J Vet Med Educ*, 34(4), 416-422
- Sledge DG, Webster J und Kiupel M (2016): Canine cutaneous mast cell tumors: A combined clinical and pathologic approach to diagnosis, prognosis, and treatment selection. *Vet J*, 215, 43-54

- Snead DR, Tsang YW, Meskiri A, Kimani PK, Crossman R, Rajpoot NM, Blessing E, Chen K, Gopalakrishnan K, Matthews P, Momtahan N, Read-Jones S, Sah S, Simmons E, Sinha B, Suortamo S, Yeo Y, El Daly H und Cree IA (2016): Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology*, 68(7), 1063-1072
- Sommer C, Fiaschi L, Hamprecht FA und Gerlich DW. (2012). Learning-based mitotic cell detection in histopathological images. *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*
- Stathonikos N, Veta M, Huisman A und van Diest PJ (2013): Going fully digital: Perspective of a Dutch academic pathology lab. *J Pathol Inform*, 4, 15
- Stathonikos N, Nguyen TQ, Spoto CP, Verdaasdonk MAM und van Diest PJ (2019): Being fully digital: perspective of a Dutch academic pathology laboratory. *Histopathology*, 75(5), 621-635
- Stathonikos N, Nguyen TQ und van Diest PJ (2020a): Rocky road to digital diagnostics: implementation issues and exhilarating experiences. *J Clin Pathol*,
- Stathonikos N, van Varsseveld NC, Vink A, van Dijk MR, Nguyen TQ, Leng WWJ, Lacle MM, Goldschmeding R, Vreuls CPH und van Diest PJ (2020b): Digital pathology in the time of corona. *J Clin Pathol*, 73(11), 706-712
- Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammage C, Thng F, Peng L und Stumpe MC (2018): Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. *Am J Surg Pathol*, 42(12), 1636-1646
- Stewart SM, Dowers KL, Cerda JR, Schoenfeld-Tacher RM und Kogan LR (2014): Microscope use in clinical veterinary practice and potential implications for veterinary school curricula. *J Vet Med Educ*, 41(4), 331-336
- Sullivan S und Hinchcliff K (2015): Update on exercise-induced pulmonary hemorrhage. *Vet Clin North Am Equine Pract*, 31(1), 187-198
- Sultan AS, Elgharib MA, Tavares T, Jessri M und Basile JR (2020): The use of artificial intelligence, machine learning and deep learning in oncologic histopathology. *J Oral Pathol Med*, 49(9), 849-856
- Tabata K, Mori I, Sasaki T, Itoh T, Shiraishi T, Yoshimi N, Maeda I, Harada O, Taniyama K, Taniyama D, Watanabe M, Mikami Y, Sato S, Kashima Y, Fujimura S und Fukuoka J (2017): Whole-slide imaging at primary pathological diagnosis: Validation of whole-slide imaging-based primary pathological diagnosis at twelve Japanese academic institutes. *Pathol Int*, 67(11), 547-554

- Tabata K, Uraoka N, Benhamida J, Hanna MG, Sirintrapun SJ, Gallas BD, Gong Q, Aly RG, Emoto K, Matsuda KM, Hameed MR, Klimstra DS und Yagi Y (2019): Validation of mitotic cell quantification via microscopy and multiple whole-slide scanners. *Diagn Pathol*, 14(1), 65
- Tellez D, Balkenhol M, Otte-Höller I, van de Loo R, Vogels R, Bult P, Wauters C, Vreuls W, Mol S und Karssemeijer N (2018): Whole-slide mitosis detection in H&E breast histology using PHH3 as a reference to train distilled stain-invariant convolutional networks. *IEEE transactions on medical imaging*, 37(9), 2126-2136
- Thorstenson S, Molin J und Lundström C (2014): Implementation of large-scale routine diagnostics using whole slide imaging in Sweden: Digital pathology experiences 2006-2013. *J Pathol Inform*, 5(1), 14
- Thrall MJ, Wimmer JL und Schwartz MR (2015): Validation of multiple whole slide imaging scanners based on the guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med*, 139(5), 656-664
- Turner OC, Aeffner F, Bangari DS, High W, Knight B, Forest T, Cossic B, Himmel LE, Rudmann DG, Bawa B, Muthuswamy A, Aina OH, Edmondson EF, Saravanan C, Brown DL, Sing T und Sebastian MM (2020): Society of Toxicologic Pathology Digital Pathology and Image Analysis Special Interest Group Article*: Opinion on the Application of Artificial Intelligence and Machine Learning to Digital Toxicologic Pathology. *Toxicol Pathol*, 48(2), 277-294
- Veta M, van Diest PJ, Willems SM, Wang H, Madabhushi A, Cruz-Roa A, Gonzalez F, Larsen AB, Vestergaard JS, Dahl AB, Cireşan DC, Schmidhuber J, Giusti A, Gambardella LM, Tek FB, Walter T, Wang CW, Kondo S, Matuszewski BJ, Precioso F, Snell V, Kittler J, de Campos TE, Khan AM, Rajpoot NM, Arkoumani E, Lacle MM, Viergever MA und Pluim JP (2015): Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Med Image Anal*, 20(1), 237-248
- Veta M, van Diest PJ, Jiwa M, Al-Janabi S und Pluim JP (2016): Mitosis Counting in Breast Cancer: Object-Level Interobserver Agreement and Comparison to an Automatic Method. *PLoS One*, 11(8), e0161286
- Veta M, Heng YJ, Stathonikos N, Bejnordi BE, Beca F, Wollmann T, Rohr K, Shah MA, Wang D, Rousson M, Hedlund M, Tellez D, Ciompi F, Zerhouni E, Lanyi D, Viana M, Kovalev V, Liauchuk V, Phoulady HA, Qaiser T, Graham S, Rajpoot N, Sjöblom E, Molin J, Paeng K, Hwang S, Park S, Jia Z, Chang EI, Xu Y, Beck AH, van Diest PJ und Pluim JPW (2019): Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge. *Med Image Anal*, 54, 111-121
- Vodovnik A (2015): Distance reporting in digital pathology: A study on 950 cases. *J Pathol Inform*, 6, 18
- Wack K, Drogowski L, Treloar M, Evans A, Ho J, Parwani A und Montalto MC (2016): A multisite validation of whole slide imaging for primary diagnosis using standardized data collection and analysis. *J Pathol Inform*, 7, 49

- Wang S, Yang DM, Rong R, Zhan X und Xiao G (2019): Pathology Image Analysis Using Segmentation Deep Learning Algorithms. *Am J Pathol*, 189(9), 1686-1698
- Webster JD und Dunstan RW (2014): Whole-slide imaging and automated image analysis: considerations and opportunities in the practice of pathology. *Vet Pathol*, 51(1), 211-223
- Wei BR, Halsey CH, Hoover SB, Puri M, Yang HH, Gallas BD, Lee MP, Chen W, Durham AC, Dwyer JE, Sánchez MD, Traslavina RP, Frank C, Bradley C, McGill LD, Esplin DG, Schaffer PA, Cramer SD, Lyle LT, Beck J, Buza E, Gong Q, Hewitt SM und Simpson RM (2019): Agreement in Histological Assessment of Mitotic Activity Between Microscopy and Digital Whole Slide Images Informs Conversion for Clinical Diagnosis. *Acad Pathol*, 6, 2374289519859841
- Williams BJ, DaCosta P, Goacher E und Treanor D (2017): A Systematic Analysis of Discordant Diagnoses in Digital Pathology Compared With Light Microscopy. *Arch Pathol Lab Med*, 141(12), 1712-1718
- Wilm F, Bertram CA, Marzahl C, Bartel A, Donovan TA, Assenmacher C-A, Becker K, Bennett M, Corner S, Cossic B, Denk D, Dettwiler M, Garcia Gonzalez B, Gurtner C, Heier A, Lehmbecker A, Merz S, Plog S, Schmidt A, Sebastian F, Smedley RC, Tecilla M, Thaiwong T, Breininger K, Kiupel M, Maier A, Klopfleisch R und Aubreville M (2021): Influence of Inter-Annotator Variability on Automatic Mitotic Figure Assessment. In: *Bildverarbeitung für die Medizin 2021*, P. e. al., Ed, Springer Fachmedien Wiesbaden GmbH, pp. 241-246
- Zarella MD, Bowman D, Aeffner F, Farahani N, Xthona A, Absar SF, Parwani A, Bui M und Hartman DJ (2019): A practical guide to whole slide imaging: a white paper from the digital pathology association. *Archives of pathology & laboratory medicine*, 143(2), 222-234
- Zuraw A, Staup M, Klopfleisch R, Aeffner F, Brown D, Westerling-Bui T und Rudmann D (2020): Developing a Qualification and Verification Strategy for Digital Tissue Image Analysis in Toxicological Pathology. *Toxicol Pathol*, 192623320980310

8 Publikationsverzeichnis

8.1 Eigene Untersuchungen

- Bertram CA, Donovan TA, Tecilla M, Bartenschlager F, Fragoso M, Wilm F, Marzahl C, Breininger K, Maier A, Klopfleisch R, Aubreville M (2021): Dataset on Bi- and Multi-Nucleated Tumor Cells in Canine Cutaneous Mast Cell Tumors. Vortrag: BVM Workshop, Regensburg, 07-09.03.2021. Originalartikel: Bildverarbeitung für die Medizin 2021, Eds. Palm et al., Springer Fachmedien Wiesbaden GmbH, 134-139. https://doi.org/10.1007/978-3-658-33198-6_33
- Wilm F, Bertram CA, Marzahl C, Bartel A, Donovan TA, Assenmacher CA, Becker K, Bennett M, Corner S, Cossic B, Denk D, Dettwiler M, Garcia Gonzalez B, Gurtner C, Heier A, Lehmbecker A, Merz S, Plog S, Schmidt A, Sebastian F, Smedley RC, Tecilla M, Thaiwong T, Breininger K, Kiupel M, Maier A, Klopfleisch R, Aubreville M (2021): Influence of Inter-Annotator Variability on Automatic Mitotic Figure Assessment. Vortrag: BVM Workshop, Regensburg, 07-09.03.2021. Originalartikel: Bildverarbeitung für die Medizin 2021, Eds. Palm et al., Springer Fachmedien Wiesbaden GmbH, 241-246. https://doi.org/10.1007/978-3-658-33198-6_56
* im gleichen Maß beigetragen
- Bertram CA, Veta M, Marzahl C, Stathonikos N, Maier A, Klopfleisch R, Aubreville M (2020): Are Pathologist-Defined Labels Reproducible? Comparison of the TUPAC16 Mitotic Figure Dataset with an Alternative Set of Labels. Vortrag und Poster: LABELS 2020 MICCAI Workshop on Large-scale Annotations of Biomedical data and Expert Label Synthesis, 08.10.2020. Originalartikel: in J. Cardoso et al. (Eds.): iMIMIC 2020/MIL3iD 2020/LABELS 2020, LNCS 12446, pp. 204–213, 2020, Springer Nature Switzerland AG 2020. https://doi.org/10.1007/978-3-030-61166-8_22
- Bertram CA*, Aubreville M*, Gurtner C, Bartel A, Corner SM, Dettwiler M, Kershaw O, Noland EL, Schmidt A, Sledge DG, Smedley RC, Thaiwong T, Kiupel M, Maier A, Klopfleisch R (2020): Computerized Calculation of Ground Truth Mitotic Distribution in Canine Cutaneous Mast Cell Tumor Sections: Mitotic count is Area Dependent. *Vet Pathol*, 57.2, 214-226. <https://doi.org/10.1177/0300985819890686>
* im gleichen Maß beigetragen

- Aubreville M*, Bertram CA*, Donovan TA, Marzahl C, Maier A, Klopfleisch R (2020): A completely annotated whole slide image dataset of canine breast cancer to aid human breast cancer research. *Sci Data*, 7, 417. <https://doi.org/10.1038/s41597-020-00756-z>
* im gleichen Maß beigetragen
- Bertram CA*, Aubreville M*, Marzahl C, Maier A, Klopfleisch R (2019): A large-scale dataset for mitotic figure assessment on whole slide images of canine cutaneous mast cell tumor. *Sci Data*, 6.1, 1-9. <https://doi.org/10.1038/s41597-019-0290-4>
* im gleichen Maß beigetragen
- Bertram CA, Gurtner C, Dettwiler M, Kershaw O, Dietert K, Pieper L, Pischon H, Gruber AD, Klopfleisch R (2018): Validation of Digital Microscopy Compared With Light Microscopy for the Diagnosis of Canine Cutaneous Tumors. *Vet Pathol*, 55.4, 490-500. <https://doi.org/10.1177/0300985818755254>
- Bertram CA, Firsching T, Klopfleisch R (2018): Virtual Microscopy in Histopathology Training: Changing Student Attitudes in 3 Successive Academic Years. *J Vet Med Educ*, 45.2, 241-249. <https://doi.org/10.3138/jvme.1216-194r1>

8.2 Weitere themenbezogene Studien

- Marzahl C, Aubreville M, Bertram CA, Maier J, Bergler C, Kröger C, Voigt J, Breininger K, Klopfleisch R, Maier A (2021): EXACT: A collaboration toolset for algorithm-aided annotation of images with version control. *Sci Rep*, 11, 4343. <https://doi.org/10.1038/s41598-021-83827-4>
- Aubreville M, Bertram CA, Marzahl C, Gurtner C, Dettwiler M, Schmidt A, Bartenschlager F, Merz S, Fragoso M, Kershaw O, Klopfleisch R, Maier A (2020): Deep learning algorithms out-perform veterinary pathologists in detecting the mitotically most active tumor region. *Sci Rep*, 10, 16447. <https://doi.org/10.1038/s41598-020-73246-2>
- Marzahl C, Aubreville M, Bertram CA, Stayt J, Jasensky AK, Bartenschlager F, Fragoso-Garcia M, Barton AK, Elsemann S, Jabari S, Krauth J, Madhu P, Voigt J, Hill J, Klopfleisch R, Maier A (2020): Deep learning-Based Quantification of Pulmonary Hemosideriophages in Cytology Slides. *Sci Rep*, 10, 9795. <https://doi.org/10.1038/s41598-020-65958-2>
- Bertram CA, Dietert K, Pieper L, Erickson NA, Barton AK, Klopfleisch R (2018): Effects of On-Slide Fixation Methods on Cell Quality of Equine Bronchioalveolar Lavage Fluid Prepared by Cytospin. *Vet Clin Pathol*, 47, 513-519. <https://doi.org/10.1111/vcp.12623>

8.3 Themenbezogene Übersichtsartikel

- Bertram CA, Stathonikos N, Donovan TA, Bartel A, Fuchs-Baumgartinger A, Lipnik K, van Diest PJ, Bonsembiante F, Klopfleisch R (2021): Validation of Digital Microscopy: Review of Validation Methods and Sources of Bias. *Vet Pathol*, Online First. <https://doi.org/10.1177/03009858211040476>
- Meuten DJ, Moore FM, Donovan TA, Bertram CA, Klopfleisch R, Foster RA, Smedley RC, Dark MJ, Milovancev M, Stromberg P, Williams BH, Aubreville M, Avallone G, Bolfa P, Cullen J, Dennis MM, Goldschmidt M, Luong R, Miller AD, Miller MA, Munday JS, Roccabianca P, Salas EN, Schulman FY, Laufer-Amorim R, Asakawa MG, Craig L, Dervis N, Esplin DG, George JW, Hauck M, Kagawa Y, Kiupel M, Linder K, Meichner K, Marconato L, Oblak ML, Santos RL, Simpson RM, Tvedten H, Whitley D (2021): International Guidelines for Veterinary Tumor Pathology: A Call to Action. *Vet Pathol*, 58(5):766-794; <http://doi/10.1177/03009858211013712>
- Donovan TA, Moore F, Bertram CA, Luong R, Bolfa P, Klopfleisch R, Tvedten H, Salas E, Whitley DB, Aubreville M, Meuten DJ (2020): Mitotic Figures - Normal, Atypical and Imposters: A Guide to Identification. *Vet Pathol*, 58.2, 243-257. <https://doi.org/10.1177/0300985820980049>
- Klopfleisch R, Bertram CA (2019): Die Zytologie von Tracheobronchialsekret (TBS) und bronchoalveolärer Lavage-Flüssigkeit (BALF) beim Pferd mit equinem Asthma – Methoden und Möglichkeiten. *Prakt Tierarzt* 100.1, 56 – 68
- Klopfleisch R, Bertram CA (2018): Gefahr oder Chance für den tierärztlichen Beruf? Digitalisierung und künstliche Intelligenz. *Deutsches Tierärzteblatt*, 66.8, 1078-1083
- Bertram CA, Klopfleisch R (2017): The Pathologist 2.0: An Update on Digital Pathology in Veterinary Medicine. *Vet Pathol*, 54.5, 756-766. <https://doi.org/10.1177/0300985817709888>

8.4 Themenbezogene Kongressbeiträge (inklusive Originalartikel- oder Abstrakt -Veröffentlichung)

- Aubreville A, Bertram CA, Veta M, Klopfleisch R, Stathonikos N, Breining K, ter Hoeve N, Ciompi F, Maier A (2021): Generalization Challenge: Structured description of the challenge design. 24th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2021), <https://doi.org/10.5281/zenodo.4573978>

- Marzahl C, Wilm F, Dressler FF, Tharun L, Perner S, Bertram CA, Kröger C, Voigt J, Klopffleisch R, Maier A, Aubreville M, Breininger K (2021): Robust Quad-Tree based Registration on Whole Slide Images. MICCAI Computational Pathology (COMPAY) Workshop, 27.09.2021
- Ganz J, Kirsch T, Hoffmann L, Bertram CA, Hoffmann C, Maier A, Breininger K, Blümcke I, Jabari S, Aubreville M (2021): Automatic and explainable grading of meningiomas from histopathology images. MICCAI Computational Pathology (COMPAY) Workshop, 27.09.2021
- Theelke L, Wilm F, Marzahl C, Bertram CA, Klopffleisch R, Maier A, Aubreville M, Breininger K (2021): Iterative Cross-Scanner Registration for Whole Slide Images. Workshop on Computational Challenges in Digital Pathology, ICCV 2021, 11.11.2021
- Bertram CA (2021): Mitotic Count: Standardization Through Artificial Intelligence. ACVP/ASVCP Annual Meeting 2021, 30.10 – 02.11.2021
- Bertram CA, Marzahl C, Bartel A, Stayt J, Bonsembiante F, Beeler-Marfis, J, Barton AK, Brocca G, Gelain ME, Gläsel A, du Preez K, Weiler K, Breininger K, Aubreville M, Maier A, Klopffleisch R, Hill, J (2021): Deep learning-based algorithms can be used to predict cytologic total hemosiderin scores. ACVP/ASVCP Annual Meeting 2021, 30.10 – 02.11.2021
- Bertram, CA; Aubreville, M; Donovan, TA; Bartel, A; Wilm, F; Marzahl, C; Assenmacher, CA; Becker, K; Bennett, M; Corner, S; Cossic, B; Denk, D; Dettwiler, M; Garcia Gonzalez, B; Gurtner, C; Haverkamp, AK; Heier, A; Lehmbecker, A; Merz, S; Noland, EL; Plog, S; Schmidt, A; Sebastian, F; Sledge, DG; Smedley, RC; Tecilla, M; Thaiwong, T; Fuchs-Baumgartinger AM; Meuten, DJ; Breininger, K; Kiupel, M; Maier, A; Klopffleisch, R (2021): Computer assistance: A new Gold standard for the mitotic count? 4th ESVP, ECVP and ESTP Cutting Edge Pathology Congress, 15. – 17.09.2021
- Bertram CA, Klopffleisch R, Bartel A, Donovan TA, Fuchs-Baumgartinger A, Breininger K, Kiupel M, Aubreville M (2021): Expert review of algorithmic mitotic count predictions ensures high reliability. 4th ESVP, ECVP and ESTP Cutting Edge Pathology Congress, 15. – 17.09.2021
- Aubreville M, Bertram CA, Veta M, Klopffleisch R, Stathonikos N, Breininger K, ter Hoeve N, Ciompi F, Maier A (2021): Quantifying the Scanner-Induced Domain Gap in Mitosis Detection. Medical Imaging with Deep Learning (MIDL) 2021, Lübeck, 09.07.2021; <https://arxiv.org/abs/2103.16515>

- Klopffleisch R, Bertram C (2021): Künstliche Intelligenz / Deep Learning zur Unterstützung der diagnostischen Patholog*in im Alltag – Der Anfang ist gemacht. Übersichtsvortrag: 64- Jahrestagung der Fachgruppe Pathologie der Deutschen Veterinärmedizinischen Gesellschaft 2021, Fulda, 07.03.2021
- Aubreville M, Bertram CA, Veta; M, Klopffleisch R, Stathonikos N, Breininger K, ter Hoeve N, Ciompi F, Maier A (2021): Mitosis DDomain Generalization Challenge. Challengebeschreibung für die 24th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2021). <https://doi.org/10.5281/zenodo.4573977>
- Marzahl C, Bertram CA, Wilm F, Voigt J, Barton AK, Klopffleisch R, Breininger K, Maier A, Aubreville M (2021): Cell Detection for Asthma on Partially Annotated Whole Slide Images - Learning to be EXACT. Poster und Vortrag: BVM Workshop, Regensburg, 07-09.03.2021. Originalartikel: Bildverarbeitung für die Medizin 2021, Eds. Palm et al., Springer Fachmedien Wiesbaden GmbH, 147-152. https://doi.org/10.1007/978-3-658-33198-6_36
- Marzahl C, Aubreville M, Bertram CA, Stayt J, Jasensky AK, Bartenschlager F, Fragoso M, Barton AK, Elsemann S, Jabari S, Krauth J, Madhu P, Voigt J, Hill J, Klopffleisch R, Maier A (2021): Abstract: Deep Learning-based Quantification of Pulmonary Hemosiderophages in Cytology Slides. Poster und Vortrag: BVM Workshop, Regensburg, 07-09.03.2021. Abstraktveröffentlichung: Bildverarbeitung für die Medizin 2021, Eds. Palm et al., Springer Fachmedien Wiesbaden GmbH, 48. https://doi.org/10.1007/978-3-658-33198-6_13
- Aubreville M, Bertram CA, Marzahl C, Gurtner C, Dettwiler M, Schmidt A, Bartenschlager F, Merz S, Fragoso M, Kershaw O, Klopffleisch R, Maier A (2021): Abstract: Deep Learning Algorithms Out-perform Veterinary Pathologists in Detecting the Mitotically Most Active Tumor Region. Poster und Vortrag: BVM Workshop, Regensburg, 07-09.03.2021. Abstraktveröffentlichung: Bildverarbeitung für die Medizin 2021, Eds. Palm et al., Springer Fachmedien Wiesbaden GmbH, 260. https://doi.org/10.1007/978-3-658-33198-6_60

- Aubreville M, Bertram CA, Donovan TA, Marzahl C, Maier A, Klopffleisch R (2021): Abstract: A Completely Annotated Whole Slide Image Dataset of Canine Breast Cancer to Aid Human Breast Cancer Research. Poster und Vortrag: BVM Workshop, Regensburg, 07-09.03.2021. Abstraktveröffentlichung: Bildverarbeitung für die Medizin 2021, Eds. Palm et al., Springer Fachmedien Wiesbaden GmbH, 198. https://doi.org/10.1007/978-3-658-33198-6_48
- Marzahl C, Bertram CA, Aubreville M, Petrick A, Weiler K, Gläsel AC, Fragoso M, Merz S, Bartenschlager B, Hoppe J, Langenhagen A, Jasensky AK, Voigt J, Klopffleisch R, Maier A (2021): Abstract: Are Fast Labeling Methods Reliable? A Case Study of Computer-aided Expert Annotations on Microscopy Slides. Poster und Vortrag: BVM Workshop, Regensburg, 07-09.03.2021. Abstraktveröffentlichung: Bildverarbeitung für die Medizin 2021, Eds. Palm et al., Springer Fachmedien Wiesbaden GmbH, 296. https://doi.org/10.1007/978-3-658-33198-6_71
- Bertram CA, Aubreville M, Marzahl C, Maier A, Klopffleisch R (2020): Abstract: How Big is Big Enough? A Large-Scale Histological Dataset of Mitotic Figures. Poster und Abstraktveröffentlichung: BVM Workshop, Berlin, 15-17.3.2020. Abstraktveröffentlichung: Bildverarbeitung für die Medizin 2020, Eds. Tolxdorff et al., Springer Vieweg, 293. https://doi.org/10.1007/978-3-658-29267-6_65
- Aubreville M, Bertram CA, Jabari S, Marzahl C, Klopffleisch R, Maier A (2020): Inter-spezies, Inter-Tissue Domain Adaptation for Mitotic Figure Assessment - Learning New Tricks from old Dogs. Vortrag: BVM Workshop, 15-17.3.2020, Berlin. Originalartikel: Bildverarbeitung für die Medizin 2020, Eds. Tolxdorff et al., Springer, 1-7. https://doi.org/10.1007/978-3-658-29267-6_1
- Marzahl C, Aubreville M, Bertram CA, Gerlach S, Maier J, Voigt J, Hill J, Klopffleisch R, Maier A (2020): Is Crowd-Algorithm Collaboration an Advanced Alternative to Crowd-Sourcing on Cytology Slides. Vortrag: BVM Workshop, 15-17.3.2020, Berlin. Originalartikel: Bildverarbeitung für die Medizin 2020, Eds. Tolxdorff et al., Springer, 26-31. https://doi.org/10.1007/978-3-658-29267-6_5
- Marzahl C, Bertram CA, Aubreville M, Petrick A, Weiler K, Gläsel AC, Fragoso M, Merz S, Bartenschlager F, Hoppe J, Langenhagen A, Jasensky AK, Voigt J, Klopffleisch R, Maier A (2020): Are Fast Labeling Methods Reliable? A Case Study of Computer-Aided Expert Annotations on Microscopy Slides. Vortrag: MICCAI 2020, 04-08.10.2020. Originalartikel: A. L. Martel et al. (Eds.): MICCAI 2020, LNCS 12261, Springer Nature Switzerland AG 2020, pp. 24–32, 2020; https://doi.org/10.1007/978-3-030-59710-8_3

- Bertram CA, Aubreville M, Gurtner C, Bartel A, Corner SM, Dettwiler M, Kershaw O, Noland EL, Schmidt A, Sledge DG, Smedley RC, Thaiwong T, Kiupel M, Maier A, Klopfleisch R (2019): MITOTIC COUNT IN CANINE CUTANEOUS MAST CELL TUMORS: MANUAL COUNTS CONSISTENT WITH HIGHEST MITOTIC DENSITY? Vortrag: ACVP & ASVCP 2019 Concurrent Annual Meeting, San Antonio, USA, 9.-13.11.2019.
Abstraktveröffentlichung: https://cdn.ymaws.com/www.acvp.org/resource/resmgr/meetings_&_events/2019/2019_acvp_and_asvcp_final_ab.pdf
- Bertram CA, Aubreville M, Gurtner C, Bartel A, Corner SM, Dettwiler M, Kershaw O, Noland EL, Schmidt A, Sledge DG, Smedley RC, Thaiwong T, Kiupel M, Maier A, Klopfleisch R (2019): Mitotic Count in Canine Cutaneous Mast Cell Tumours – Not Accurate but Reproducible. Vortrag: Joint Congress of Veterinary Pathology and Veterinary Clinical Pathology (ECVP, ESVP; ECVCP), Arnhem, 25.9-28.9.2019. Abstraktveröffentlichung: J Comp Pathol 2020, 174, 143-156. <https://doi.org/10.1016/j.jcpa.2019.10.015>
- Bertram, CA, M Aubreville, Gurtner C, Kershaw O, Maier A, Klopfleisch R (2019): Mastzelltumorigradung: Wo zähle ich den Mitotic count? Vortrag: 62. Jahrestagung der Fachgruppe Pathologie der Deutschen Veterinärmedizinischen Gesellschaft, 2-3.03.2019 in Fulda. Abstraktveröffentlichung: Tierarztl Prax Ausg K, 2019, 47.3, 209. <https://doi.org/10.1055/s-0039-1688607>
- Bertram CA, Mahrzahl C; Aubreville A, Barton AK, Maier A, Hill J, Klopfleisch K (2019): Automated Scoring of Exercise-Induced Pulmonary Hemorrhage (EIPH) in Equine Bronchoalveolar Lavage Cytology. Vortrag: Joint Congress of Veterinary Pathology and Veterinary Clinical Pathology, Arnhem, 25-28.9.2019. Abstraktveröffentlichung: Vet Clin Pathol, 2019, 48, 797
- Bertram CA, Marzahl C, Aubreville M, Stayt J, Jasensky AK, Bartenschlager F, Fragoso-Garcia M, Jabari S, Elsemann S, Barton AK, Maier A, Hill J, Klopfleisch R (2019): SCORING EXERCISE-INDUCED PULMONARY HEMORRHAGE: ACCURACY OF HUMAN OBSERVER AND DEEP-LEARNING BASED ALGORITHMS. Vortrag: ACVP & ASVCP 2019 Concurrent Annual Meeting, San Antonio, USA, 9-13.11.2019. Abstraktveröffentlichung: https://cdn.ymaws.com/www.acvp.org/resource/resmgr/meetings_&_events/2019/2019_acvp_and_asvcp_final_ab.pdf

- Bertram CA, Aubreville M, Maier A, Klopfleisch R (2019): Histopathological Datasets – An Elementary Step in the Development of Deep Learning Based Algorithms for Automated Image Analysis. Poster: 17th European Congress of Toxicologic Pathology, Köln, 17-20.9.2019. Abstraktveröffentlichung: Proceedings
- Bertram CA, Klopfleisch R (2019): Glass versus Digital –Validation of Digital Microscopy in Veterinary Pathology. Poster: 17th European Congress of Toxicologic Pathology, Köln, 17-20.9.2019. Abstraktveröffentlichung: Proceedings
- Bertram CA, Dietert K, Pieper L, Kühnel S, Erickson NA, Barton AK, Klopfleisch R (2019): Effects of wet spray fixation on cell quality of equine bronchoalveolar lavage and tracheal wash fluid. Poster: Joint Congress of Veterinary Pathology and Veterinary Clinical Pathology, Arnhem, 25-28.9.2019. Abstraktveröffentlichung: Vet Clin Pathol, 2019, 48, 804. <https://doi.org/10.1111/vcp.12798>
- Bertram CA (2019): Ok google: Was ist die Diagnose? – Digitaler Pathologie Assistent. Science Slam: Niedersächsischen Tierärztetag 24-26.01.2019, Hannover
- Aubreville M, Bertram CA, Klopfleisch R, Maier A (2019): Augmented Mitotic Cell Count Using Field of Interest Proposal. Vortrag: BVM Workshop, 17-19.03.2019, Lübeck. Originalartikel: Bildverarbeitung für die Medizin 2019. Eds. Handels H et al., Springer Vieweg, 321-326, https://doi.org/10.1007/978-3-658-25326-4_71
- Aubreville M, Bertram CA, Klopfleisch R, Maier A (2019): Field Of Interest Proposal for Augmented Mitotic Cell Count: A Comparison of Two Networks. Vortrag: INSTICC BIOIMAGING/BIOSTEC 2019 Conference in Prague, Czech Republic, 23.02.2019. Originalartikel: Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 2, BIOIMAGING (BIOIMAGING/BIOSTEC), Prague, Czech Republic, 30-37, 2019, ISBN 978-989-758-353-7
- Bertram CA, Gurtner C, Dettwiler M, Kershaw O, Dietert K, Pieper L, Pischon H, Gruber AD, Klopfleisch R (2018): Validierung von digitaler Mikroskopie im Vergleich zur konventionellen Lichtmikroskopie für die primäre Diagnose von kaninen Hauttumoren. Vortrag: 61. Jahrestagung der Fachgruppe Pathologie der Deutschen Veterinärmedizinischen Gesellschaft, 3-04.03.2018, Fulda. Abstraktveröffentlichung: Tierarzt Prax Ausgabe K, 3/2018, A21. <https://doi.org/10.1055/s-0038-1665513>

- Bertram CA, Dietert K, Pieper L, Kühnel S, Barton AK, Klopffleisch R (2018): Effects of Wet Spray Fixation on the Cell Quality of Equine Bronchoalveolar Lavage and Tracheal Wash Fluids. Poster: 3. Internationaler Kongress zur Pferdemedizin, 5-6.10.2018, Berlin. Abstraktveröffentlichung: Proceedings
- Krappmann M, Aubreville M, Maier A, Bertram C, Klopffleisch R (2018): Classification of Mitotic Cells – Potentials Beyond the Limits of Small Data Sets. Poster: BVM Workshop, 11-13.03.2018, Erlangen. Originalartikel: Bildverarbeitung für die Medizin 2018. Eds. A. Maier, et al. Berlin, Heidelberg, Springer Vieweg, 245-250. https://doi.org/10.1007/978-3-662-56537-7_66
- Aubreville M, Bertram C, Klopffleisch R, Maier A (2018): SlideRunner – A Tool for Massive Cell Annotations in Whole Slide Images. Softwaredemonstration: BVM Workshop, 11-13.03.2018, Erlangen. Originalartikel: Bildverarbeitung für die Medizin 2018. Eds. A. Maier, et al. Berlin, Heidelberg, Springer Vieweg, 309-314. https://doi.org/10.1007/978-3-662-56537-7_81
- Aubreville M, Krappmann M, Bertram CA, Klopffleisch R, Maier A (2017): A guided spatial transformer network for histology cell differentiation. Vortrag: Eurographics Workshop on Visual Computing for Biology and Medicine, 7-8.09.2017. Originalartikel: Proceedings. <https://doi.org/10.2312/vcbm.20171233>

9 Danksagung

Meine ganz besondere Dankbarkeit möchte ich meinem PhD-Betreuer Prof. Dr. Robert Klopffleisch ausdrücken für die uneingeschränkte Unterstützung und wissenschaftliche Anleitung meiner Forschungsprojekte. Vielen Dank für das große Vertrauen, das du mir gegenüber gebracht hast. Meine Dankbarkeit gilt auch meinen PhD-Mentoren Prof. Dr. Matti Kiupel und PD Dr. Ann Kristin Barton, die mich in meinem PhD-Projekt begleitet und beraten haben. Weiterhin möchte ich einen ganz besonderen Dank an meine Forschungskollaboration aus der Arbeitsgruppe von Prof. Dr. Andreas Maier (Lehrstuhl für Mustererkennung der Friedrich-Alexander-Universität Erlangen-Nürnberg) und Prof Dr. Marc Aubreville (Fakultät Informatik der Technischen Hochschule Ingolstadt) richten, die mit mir zusammen die zahlreichen Datensätze und Bildanalysealgorithmen entwickelt und evaluiert haben. Dabei möchte ich mich ausdrücklich bei Prof. Dr. Marc Aubreville, Prof. Dr. Andreas Maier, Christian Marzahl, Frauke Wilm und Prof. Dr. Katharina Breininger bedanken für die stets konstruktive und lehrreiche Zusammenarbeit. Weiterhin möchte ich mich bei den zahlreichen Pathologen/innen, die sich an meinen Projekten als Kooperationspartner und Studienteilnehmer/innen beteiligt haben, bedanken. Bezüglich der Probenaufbereitung möchte ich die exzellente Arbeit der medizinisch-technischen Assistentinnen aus dem Histologielabor honorieren.

Für die inhaltlichen und sprachlichen Korrekturen dieser Dissertation möchte ich mich bei Prof. Dr. Marc Aubreville, Heidi Bertram und Dr. Taryn Donovan bedanken.

Weiterhin möchte ich mich ganz besonders bei der Dres. Jutta & Georg Bruns-Stiftung für innovative Veterinärmedizin bedanken für die Vergabe eines zweijährigen Stipendiums, welches mir ermöglicht hat, meine Forschungsprojekte weiterzuführen.

10 Finanzierungsquelle

Die Arbeiten wurden finanziell unterstützt durch ein Stipendium der Dres. Jutta & Georg Bruns-Stiftung für innovative Veterinärmedizin. Es besteht kein Interessenskonflikt durch die finanzielle Unterstützung der Arbeiten.

Selbstständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbstständig angefertigt habe. Ich versichere, dass ich ausschließlich die angegebenen Quellen und Hilfen in Anspruch genommen habe.

Berlin, 03.12.2021

Christof Albert Bertram

ISBN: 978-3-86345-612-2



Verlag der DVG Service GmbH

An der Alten Post 2 • 35390 Gießen

Tel.: 0641 984446-0 • Fax: 0641 984446-25

E-Mail: info@dvg.de • Web: www.dvg.de