

Bernd Christian Lahrmann
Dr. sc. hum.

Quantitative zytologische und histologische Hochdurchsatzanalyse mittels Whole Slide Imaging: Anwendung in Diagnostik und Forschung

Promotionsfach: DKFZ (Deutsches Krebsforschungszentrum)
Doktorvater: Priv.-Doz. Dr.-Ing. Niels Grabe

Das digitale Zeitalter hat eine immer größere Bedeutung in vielen Feldern der medizinischen und biologischen Forschung sowie der Routinediagnostik eingenommen. Die Einführung der "WSI-Scansysteme" zur kompletten Digitalisierung von Objektträgern ermöglicht neue Anwendungsmöglichkeiten im Umfeld der Forschung und Routinediagnostik. Die Forschung an solch hochauflösenden virtuellen Slides erfordert die Analyse einer derart großen Menge an Bildinformationen, dass diese manuell kaum noch zu bewerkstelligen sind. Für die Extraktion von klinisch relevanten Parametern zur weiteren Verarbeitung müssen automatisierte Lösungen implementiert werden. Innerhalb dieser Arbeit wurden Methoden erarbeitet, die in der Lage sind, automatisiert Parameter aus virtuellen Slides zu extrahieren, sodass diese anschließend mittels automatisierter Lernverfahren zur Klassifizierung angewandt werden konnten. Des Weiteren konnte allgemein die diagnostische Performance (Verarbeitungsgeschwindigkeit, Fehlerrate) verbessert werden. In dieser Arbeit wurden Algorithmen der Bildverarbeitung in Kombination mit der virtuellen Mikroskopie angewandt, um vollautomatische Workflows zur Analyse von Geweben und zytologischen Proben zu etablieren. Dies konnte exemplarisch an zwei verschiedenen Entitäten der pathologischen Diagnostik gezeigt werden:

1. Methoden zur Diagnostik der Zervixkarzinoms.

Zervixkarzinome sind weltweit die zweithäufigste Krebserkrankung bei Frauen. Werden die Krebsvorstufen rechtzeitig identifiziert, so liegen die Chancen einer Heilung bei annähernd 100%. Die kürzlich eingefügten Doppelfärbungen versprechen eine hohe diagnostische Performance. Bei diesen Verfahren werden abnormale Zellen mithilfe von zwei Biomarkern detektiert. Das Verfahren ist optimal für eine automatische Analyse geeignet, da hier keine morphologischen Parameter interpretiert werden müssen. Die zytopathologische Diagnostik stützt sich heute weitgehend auf eine visuelle Inspektion unter dem Mikroskop. Der Trend geht jedoch hin zur virtuellen digitalen Diagnose der Zellpräparate. Die digitalisierten Slides, welche bis zu 10 GB groß sein und über 150.000 Zellen enthalten können, lassen sich manuell gut begutachten. Hierbei besteht jedoch aufgrund der Fülle der Zellen die Gefahr, eine abnormale Zelle zu übersehen, die jedoch entscheidend für die Diagnose sein kann. Das Ziel in diesem Teil der Arbeit war es, ein komplett automatisiertes Screening für die Analyse von zervikalen Abstrichen zu implementieren. Für die Digitalisierung der Slides wurde eine vollautomatische Lösung implementiert. Dabei wurde eine semantische Fokussiermethode etabliert, die erkennt, ob Zellen im Fokus sind. Eine anschließende Analyse der Bildqualität garantiert qualitativ hochwertige virtuelle Slides. Die hier implementierte semantische Fokussiermethode und die Implementierung einer automatischen Qualitätskontrolle der Slides wurden in dieser Arbeit erstmalig beschrieben. Durch Kombination beider Algorithmen war es möglich, vollautomatisch zytologische Präparate mit nur einem Fokuslayer unter 13,5 Minuten in höchster Qualität zu digitalisieren. Im nächsten Teil der Arbeit wurde eine Methodik zur Klassifikation aller Zellen eines Slides entwickelt. Die Implementierung der Zellklassifikation besteht aus zwei größeren Teilen: zum einen aus der Segmentierung aller Objekte eines Slides und zum anderen der anschließenden Einzelzellklassifikation. Implementiert wurde diese mithilfe einer Kombination aus Support Vector Machines und regelbasierten Klassifikatoren. Der wichtigste Schritt ist hierbei die Detektion von Doppelfärbungen, also Zellen, die gleichzeitig Ki-67 und p16^{INK4a} exprimieren. Es konnte dabei eine Sensitivität und Spezifität von 95,47% und 99,84% erreicht werden. Der letzte Teil des Workflows diente zur Klassifikation des Schweregrades der Dysplasie der Slides. Für die Klassifikation der Slides wurden insgesamt 48 globale Slidefeatures extrahiert und mittels mRMR oder F-Score-Featureselektion nach

Separierungsstärke sortiert. Es wurden insgesamt sechs verschiedene Klassifikatoren an einem 100 Slide großen Trainingsset trainiert. Die sechs Klassifikatoren wurden an einem Testset von 539 Slides evaluiert. Es konnte eine maximale Accuracy von 75,7% erreicht werden, welche im Vergleich zu einer manuellen Auswertung höher ist. Die mittlere Scanzeit pro Slide beträgt ca. 13,5 Minuten. Die Verarbeitungsgeschwindigkeit pro Slide bei der Zellklassifikation betrug im Schnitt 7,2(\pm 5.0) Minuten, die Verarbeitungszeit des globalen Klassifikators ca. 1 Minute. Die Verarbeitungsgeschwindigkeit der Zellklassifikation und der globalen Slide-Klassifikation sind kombiniert also immer noch deutlich geringer als die Scanzeit eines Slides. Mit dem hier präsentierten Workflow können zytologische Abstriche hinsichtlich des CIN-Grades voll automatisiert innerhalb von 13,5 Minuten mit einer Accuracy von 75,7% klassifiziert werden.

2. Analyse von Tissue Microarrays.

Das zweite Ziel dieser Arbeit war es, einen komplett automatisierten Workflow für die Analyse von Tissue Microarrays zu implementieren. Die TMA-Technologie ermöglicht die simultane immunohistochemische Analyse von mehreren hundert verschiedenen Gewebeproben auf nur einem Objektträger. TMAs werden hergestellt, indem mehrere verschiedene Gewebeproben (Tissue-Cores) aus konventionellen histologischen Paraffinblöcken herausgestanzt werden und in einen anderen Empfängerparaffinblock übertragen werden. Die Kombination der Tissue-Microarray-Technologie mit der Whole-Slide-Imaging-Technologie setzt die Grundlage für eine automatische Hochdurchsatzanalyse. Um beide Technologien für die Hochdurchsatzanalyse zu kombinieren, wurden die folgenden Schritte implementiert: Nach Präparation wurden die TMAs digitalisiert. Auftretende Intensitätsschwankungen wurden automatisch identifiziert und anschließend aus den Slides herausgefiltert. Dieser Schritt erhöht die Bildqualität der TMAs und bereitet diese für nachfolgende Bildverarbeitungsschritte vor. Im nächsten Schritt des Workflows wurden Methoden entwickelt, welche den einzelnen Cores der TMAs ihre aktuellen Positionen im Array zuweisen. Die Zuweisung der richtigen Arraypositionen ist ein unverzichtbarer Schritt für die spätere Analyse und Auswertung der Tissue-Cores. Eine fehlerhafte Zuordnung kann fatal sein, wenn basierend auf der Auswertung der Tissue-Cores patientenbasierte Diagnosen oder Therapieentscheidungen gefällt werden. Mit der hier präsentierten Methode konnten 8.864 Cores von insgesamt 8.900 Cores richtig zugewiesen werden (Acc. 99,59%). Der letzte Teil des Workflows behandelt die räumliche Trennung von verschiedenen Gewebetypen innerhalb der TMAs. Ziel war es, zwischen Tumor- und Stromagewebe automatisch zu unterscheiden, sodass in den klassifizierten Regionen Proteinlevel quantifiziert werden können. Dieser Schritt ist also eine Alternative zur Verwendung von Biomarkern zur Proteinlevelquantifizierung mittels Kolo-kalisation. Für diesen Schritt wurden zuerst die einzelnen Zellkerne der TMA-Cores segmentiert (Accuracy 94,1%). Basierend auf den segmentierten Zellkernen wurden Zellgraphen zur Klassifikation generiert. Die Tumorgewebe konnten vom Stroma mit einer Accuracy von 86,1% separiert werden.

Als Gesamtergebnis dieser Arbeit wurden zwei vollautomatisierte Workflows etabliert, die innerhalb der pathologischen Routinediagnostik und Forschung verwendet werden können. Es wurden somit wichtige Schritte innerhalb eines pathologischen Arbeitsablaufs automatisiert.