



Forschungsanwendungen in der digitalen Radiologie

„Big data“ und Co

Einführung

Bilddaten werden in immer größeren Mengen produziert. In [3] wird geschätzt, dass 30 % der weltweiten Speicherkapazität medizinische Bilddaten sind (digital und Film). Tomographische Geräte produzieren mehr und dünnere Schichten in höherer Auflösung, aber das radiologische Personal nimmt nicht in der gleichen Größenordnung wie die Bildgebung zu, sodass de facto immer weniger Zeit zur Befundung zur Verfügung steht. „Incidental findings“ oder zufällig gefundene Anomalien [4] sind häufiger, und es ist nicht immer klar, bis auf welchen Level alles im Befund beschrieben werden muss. Automatische Bildanalyse kann bei der Befundung helfen und Routineaufgaben abnehmen (z. B. durch automatisches Ausfüllen von „digital imaging and communications in medicine (DICOM) structured reporting“) und ggf. Zusatzinformationen zur Entscheidungsunterstützung liefern, z. B. Fälle mit ähnlichen Anomalien. „Big data“ ist dabei ein wichtiges Keyword, denn je mehr Daten aus der Vergangenheit analysiert werden können, um so generalisierbarer und anwendbarer können die Ergebnisse werden. Es ist aber nicht nur die Menge der Daten, sondern auch die Qualität entscheidend, und natürlich müssen bei der Bearbeitung medizinischer Daten Datensicherheit und Vertraulichkeit gewahrt bleiben.

Dieser Artikel hat als Ziel, aktuelle Forschungsergebnisse v. a. von 2 Forschungsprojekten vorzustellen, die potenziell einen wichtigen Einfluss auf die

Zukunft der digitalen Radiologie haben könnten.

VISCERAL-Projekt

Das VISCERAL¹-Projekt hat 2 Hauptziele. Zum einen soll eine Architektur aufgebaut werden, die es erlaubt, große Mengen an Forschungsdaten zentral auf sicheren Servern zu installieren und so Forschungsgruppen einen Rahmen für den Vergleich von Prototypen zu liefern. Da Daten nicht mehr bewegt/dupliziert werden müssen, kann auch mit sehr großen Datenmengen gearbeitet werden, und die Sicherheit der Daten ist in dieser Struktur wesentlich höher, da keine Daten verteilt gespeichert werden. Zweites Ziel von VISCERAL ist das Benchmarking von Forschungsprototypen für 3 Aufgaben: Bildsegmentierung, Erkennung von Läsionen und das Auffinden ähnlicher Fälle. Alle 3 Aufgaben können Bausteine für die Automatisierung oder Entscheidungsunterstützung sein. Durch das Einbeziehen vieler Forschungsgruppen kann so auch die Weisheit der Crowd („wisdom of the crowd“ [5]) benutzt werden, also nicht eine Gruppe arbeitet an einer Aufgabe, sondern jeder interessierte Forscher kann daran teilhaben, und die besten Ergebnisse werden am Ende übernommen. Die Plattform verbindet damit Forscher aus verschiedenen Gruppen weltweit und erlaubt es, die besten Resultate und Tools zu identifizieren und diese auch zu kombinieren, um die Qualität weiter zu steigern.

¹ <http://visceral.eu/>

Khresmoi-Projekt

Das Khresmoi²-Projekt befasst sich v. a. mit der Informationssuche in medizinischen Datenquellen für unterschiedliche Nutzergruppen (Patienten, Hausärzte, Radiologen). In Bezug auf die Radiologie geht es in einem Teilprojekt um die Suche in Bilddaten. Dies beinhaltet sowohl die Suche in Texten, wie den Radiologieberichten, aber auch eine Suche direkt in den Bildpixeln, auch inhaltsbasierte Suche genannt. Dies erlaubt es, z. B. unbekannte Strukturen oder Läsionen zu markieren und dann nach ähnlichen Läsionen im PACS („picture archival and communication system“) oder der medizinischen Literatur zu suchen. Dabei hilft, dass PubMedCentral³ über 800.000 biomedizinische Open-access-Artikel indiziert und diese mit allen über 2 Mio. Bildern kostenlos zur Verfügung stellt.

Diese beiden Projekte sind beispielhaft für eine große Anzahl an Projekten, die aktuell versuchen, aus großen Datenmengen wichtige Informationen zu extrahieren, sowohl in der Radiologie als auch in anderen Bereichen. Diese Projekte sind mit der aktuellen Computergeneration erst möglich geworden, denn bei guter Benutzung können hunderte von Terabytes mittlerweile gespeichert und auch ausgewertet werden, wie z. B. auch erste Firmen versuchen, im Medizinbereich zu zeigen, wie z. B. Zebra Medical Vision⁴ in der medizinischen Bildgebung oder IBM mit Watson.

² <http://khresmoi.eu/>

³ <http://www.ncbi.nlm.nih.gov/pmc/>

⁴ <http://www.zebra-med.com/>

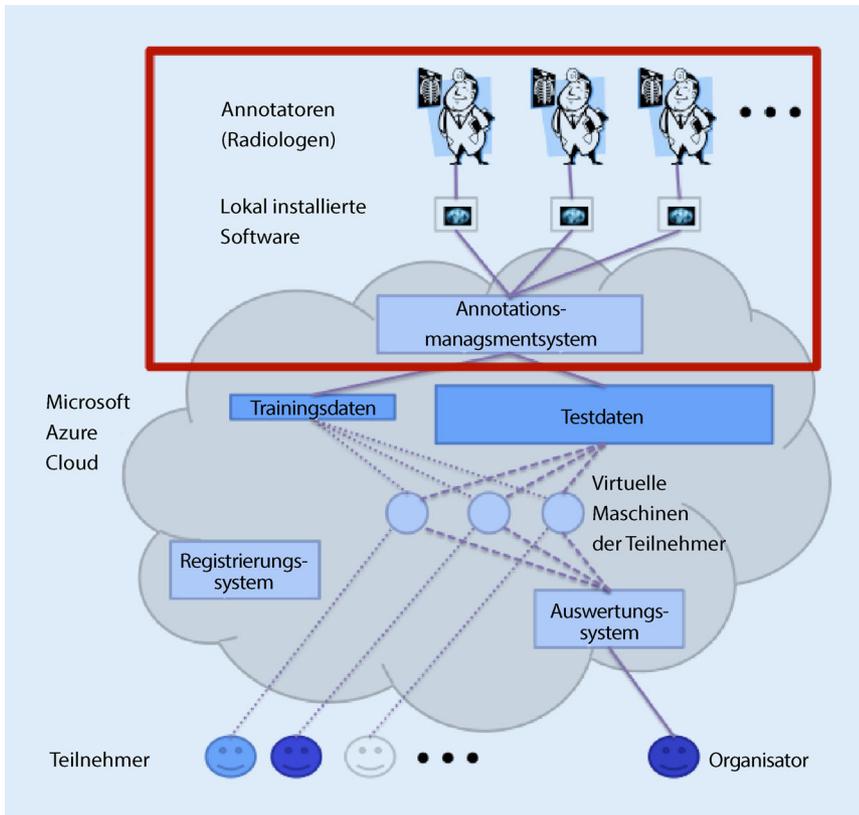


Abb. 1 ▲ Architektur für die Datenannotation und -analyse in VISCERAL mit allen Daten in der Cloud (in diesem Beispiel in Azure) und einem verteilten Zugriff auf diese Daten auch für die Datenannotation

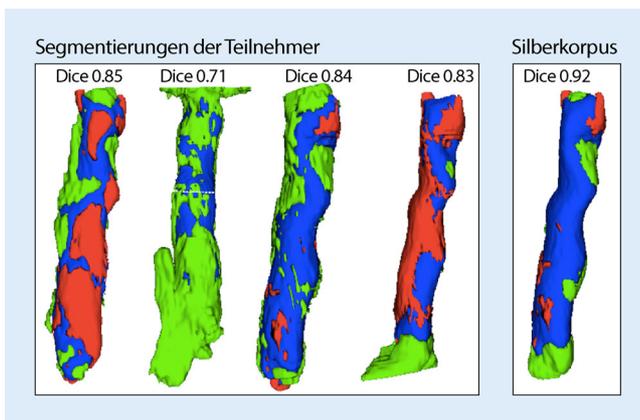


Abb. 2 ▲ Segmentierungsergebnisse verschiedener Forschungsgruppen können kombiniert und damit normalerweise durch Fusion eine deutlich bessere Qualität erzielt werden als mit jedem einzelnen Algorithmus (blau korrekte Segmentierung, grün zu viel segmentiert, rot fehlende Segmentierung). Diese Daten können dann auch wieder als Silberkorpus für das Trainieren von Algorithmen benutzt werden [6]. Dice Metrik für die Auswertung der Segmentierung

Eine Architektur für Forschungsevaluation und Benchmarks

Ein generelles Problem für Datenanalyseprojekte war oft die fehlende Verfügbar-

keit großer Datenmengen, und so wurde häufig an kleinen und privaten Datensätzen gearbeitet, was die Resultate aber nicht reproduzierbar macht und teilweise auch zu falschen Resultaten führte, weil ein Bias in den Daten vorlag. Bench-

marks oder Challenges auf Konferenzen wie der MICCAI (Medical Image Computing and Computer Assisted Interventions) versuchen gemeinsame und offene Datensätze zu erzeugen, wie auch das NCI (National Cancer Institute) mit dem Cancer Imaging Archive⁵. Das führte zu einer Vergleichbarkeit der Algorithmen und half, wirklichen Fortschritt in der Forschung zu erzielen und zu messen. Sehr große Datenmengen können allerdings nicht mehr einfach heruntergeladen werden und so versandten einige Benchmarks physische Festplatten, was aber auch Nachteile hat wie hohe Kosten und das Risiko von Defekten. Auch diese Benchmarks haben Nachteile, denn große medizinische Datenmengen haben immer das Risiko einer Re-Identifikation des Patienten und es steigt potenziell, wenn mehr Daten verfügbar sind oder longitudinale Daten desselben Patienten, was aber gerade der für die Forschung interessanteste Bereich ist. So können Forschungsdaten häufig nur schwer geteilt werden, da Ethikkommissionen sich dieser Probleme bewusst sind.

Die VISCERAL-Architektur für Forschungsdaten versucht, diese Probleme zu lösen und Daten auf sicheren Servern zu speichern, anstelle die Daten zu duplizieren und zu versenden (Abb. 1). In dieser Architektur installieren die teilnehmenden Forscher ihre Algorithmen in einer virtuellen Maschine (VM) in der Cloud und erhalten Zugriff auf einen kleinen Datensatz mit Trainingsdaten, nachdem sie sich registriert und einen Datennutzungsvertrag unterschrieben haben. Mit diesen Trainingsdaten kann das System getestet und optimiert werden.

Die Architektur enthält ebenfalls die Annotierung der Daten (Segmentierung von Organen, Markierung von Läsionen usw.) durch Radiologen, die direkt in der Cloud an den Datensätzen arbeiten. Die Qualität der Annotierung wird automatisch analysiert und ein Inter-rater disagreement berechnet, auch um die Subjektivität einer Aufgabe zu messen. Die Forscher haben dabei nie direkten Zugriff auf die große Menge der Testdaten. Dieser Zugriff wird von den Organisatoren durchgeführt, welche die Forschungspro-

⁵ <http://www.cancerimagingarchive.net/>

totypen auf den Testdaten ausführen. Damit kann ein maximaler Schutz der Daten erreicht werden und eine volle Reproduzierbarkeit der Resultate, da Daten und ausführbare Programme in den virtuellen Maschinen gespeichert werden und so langfristig verfügbar sind. Die Programme können somit auch direkt auf neuen Daten getestet werden, um die Qualität zu verifizieren. Die Teilnehmer können auch nicht manuell optimieren, da sie nur Zugriff auf die Trainingsdaten haben. Die Testdaten können auch innerhalb eines Krankenhauses liegen, denn virtuelle Maschinen können einfach dupliziert und auf einen anderen Computer übertragen werden. Die sicheren Server können sich also überall befinden, und nur die Algorithmen werden zu den Daten gebracht.

VISCERAL hat in diesem Modell 3 Benchmarks durchgeführt. Ein *Segmentierungsbenchmark* hat verschiedene Algorithmen zur Segmentierung von 20 Organen in CT- und MR-Bildern jeweils mit und ohne Kontrastmittel getestet. Die Organsegmentierung ist generell der erste Schritt einer Automatisierung, denn basierend auf Segmentierung weiß der Computer, in welchem Organ ein Radiologe analysiert oder misst und kann das direkt in das „structured reporting“ schreiben oder weitere Algorithmen zur Klassifizierung von Gewebe aufrufen.

Teilnehmer am Benchmark konnten ihre Algorithmen auf über 100 manuell annotierten Volumina testen. Die Algorithmen in den virtuellen Maschinen wurden aber noch auf tausenden weiteren Volumina ausgeführt. Zwar liegt auf diesen Volumina keine manuelle Segmentierung vor, aber man kann die Resultate aller ausgeführten Algorithmen kombinieren und damit eine Art Silberstandard erzeugen. Dies liefert deutlich bessere Ergebnisse als sogar die besten Algorithmen. Ein Beispiel für diese Label-Fusion und den Vergleich mit einer manuellen Annotierung zeigt **Abb. 2**. Dadurch wird manuelle Annotierung, die sehr zeitaufwendig und teuer ist, limitiert, aber es können der Forschung trotzdem tausende annotierter Volumina zur Verfügung gestellt werden, um aktuelle Algorithmen zu verbessern.

Radiologe DOI 10.1007/s00117-015-0042-1
© Springer-Verlag Berlin Heidelberg 2015

H. Müller · A. Hanbury

Forschungsanwendungen in der digitalen Radiologie. „Big data“ und Co

Zusammenfassung

Medizinische Bildgebung produziert immer komplexere Bilder (höhere Auflösung, dünnere Schichten) mit einer zunehmenden Zahl unterschiedlicher Protokolle, sodass die Befundung schwieriger wird und mehr Informationen verarbeitet werden müssen, da die Zahl der Radiologen bei weitem nicht im selben Umfang wächst wie die zu analysierenden Daten. Ziel dieses Artikels ist es, aktuelle Forschungsergebnisse von Projekten vorzustellen, die medizinische Bilddaten zur Entscheidungsunterstützung benutzen. Eine Infrastruktur, die es Forschern erlaubt, Bilder zu analysieren und beste Algorithmen zu erkennen, ohne dass die Daten einen sicheren Server verlassen müssen, wird auch erklärt. Der Text stellt Resultate der EU-finanzierten Khresmoi- und VISCERAL-Projekte vor. Diese erlauben es, Daten vergangener Fälle in einem Bildarchiv zu benutzen, um Entscheidungen zu unterstützen und Aufgaben zu automatisieren. Die Resultate stellen eine Evaluationsumgebung für medizinische Bilddaten und andere Datensätze vor. Diese erlauben es, in einer sicheren

Umgebung aus vergangenen Fällen, Daten für zukünftige Befundungen zu extrahieren. Die vorgestellten Prototypen erlauben es, aus Bilddaten direkt Information zu extrahieren und als Entscheidungsunterstützung zu benutzen. Die Forschungsprototypen sind allerdings noch nicht in klinischem Einsatz evaluiert worden, nur subjektive Benutzertests wurden unternommen, welche sehr positiv waren. Die Zukunft der Radiologie wird sicherlich z. T. davon abhängen, Prozesse zu automatisieren und Daten („big data“) aus institutionellen Bildarchiven zu benutzen, um bestmöglich Wissen zu extrahieren und damit die Arbeit der Radiologen auf die wichtigen, entscheidungsrelevanten Bereiche zu konzentrieren, und Routineaufgaben zu automatisieren, wo dies möglich ist.

Schlüsselwörter

„Big data“ · Digitale Bildanalyse · Benchmarking · Segmentierung · Entscheidungsunterstützung

Research applications in digital radiology. Big data and co

Abstract

Medical imaging produces increasingly complex images (e.g. thinner slices and higher resolution) with more protocols, so that image reading has also become much more complex. More information needs to be processed and usually the number of radiologists available for these tasks has not increased to the same extent. The objective of this article is to present current research results from projects on the use of image data for clinical decision support. An infrastructure that can allow large volumes of data to be accessed is presented. In this way the best performing tools can be identified without the medical data having to leave secure servers. The text presents the results of the VISCERAL and Khresmoi EU-funded projects, which allow the analysis of previous cases from institutional archives to support decision-making and for process automation. The results also represent a secure evaluation environment for medical image analysis. This

allows the use of data extracted from past cases to solve information needs occurring when diagnosing new cases. The presented research prototypes allow direct extraction of knowledge from the visual data of the images and to use this for decision support or process automation. Real clinical use has not been tested but several subjective user tests showed the effectiveness and efficiency of the process. The future in radiology will clearly depend on better use of the important knowledge in clinical image archives to automate processes and aid decision-making via big data analysis. This can help concentrate the work of radiologists towards the most important parts of diagnostics.

Keywords

Big data · Digital image analysis · Benchmarking · Segmentation · Decision support

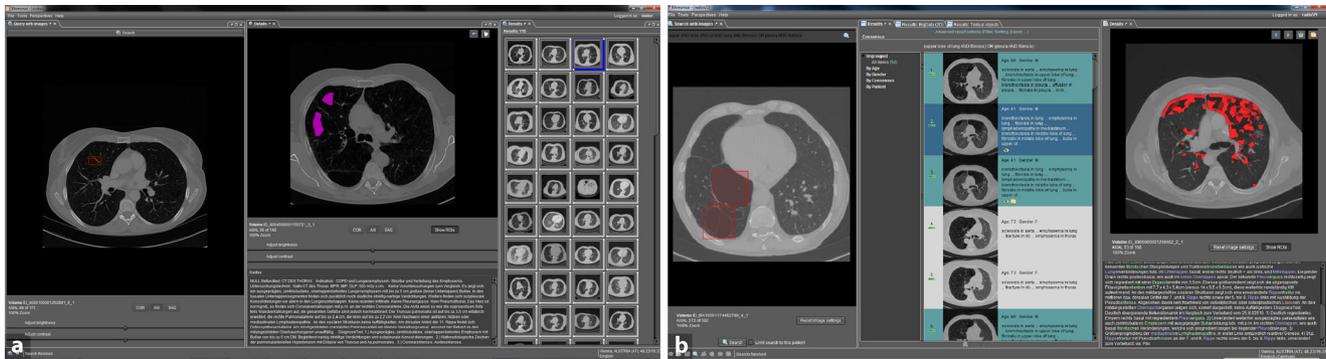


Abb. 3 ▲ Radiologiesuchanwendung KhresmoiRadiology, die es erlaubt, mit Beispielregionen in Bildern nach visuell ähnlichen Regionen in Bildern des „picture archival and communication system“ (PACS) oder der medizinischen Literatur zu suchen. Die rechte Seite zeigt, dass die häufigsten Diagnosen in den ersten 100 Resultaten farblich anders hervorgehoben werden

Im „*lesion detection benchmark*“ wurden über 1600 Läsionen in Knochen, Leber, Gehirn, Lunge und Lymphknoten manuell annotiert. Ziel war es wieder, automatisch in neuen Bildern solche Läsionen zu finden. Dies kann ebenfalls wieder für die Entscheidungsunterstützung benutzt werden, denn in neuen Fällen können mögliche Läsionen farblich hervorgehoben werden und so die Befundung vereinfachen. Wie bei der Segmentierung können dann die automatischen Algorithmen in sehr viel größere Datenmengen eingesetzt werden und so einen Silberstandard kreieren.

Beim „*retrieval benchmark*“ ging es um das Auffinden ähnlicher Fälle basierend auf Bilddaten und RadLex(Radiology-lexicon)-Termen, die aus den Radiologieberichten extrahiert wurden.

Auch hier wurden manuell in mehreren 1000 Fällen die relevanten Fälle herausgesucht, sodass Algorithmen objektiv verglichen werden können. Die Fälle, die als Anfragen dienten, hatten das entsprechende Organ automatisch segmentiert, und die Fälle der Datenbank auch dank des Silberstandards, der automatisch kreiert wurde.

Alle VISCERAL-Datensätze sind auch noch für mehrere Jahre nach Projektende für die Forschung verfügbar und es wird eine häufige Benutzung dieser Ressourcen erwartet.

In **Abb. 3** wird ein Beispielsystem für die Suche in der Radiologie mit dem Khresmoi-Prototypen gezeigt. Das System erlaubt es, ganze Fälle wie in einem normalen Viewer zu öffnen und zu be-

gutachten. Zusätzlich kann der Benutzer aber Regionen im Bild markieren und diese als Eingabe für eine Informationssuche benutzen. Es werden dann Volumen mit visuell ähnlichen Regionen gesucht, bei den auch sonstige Parameter des Patienten (Alter, Geschlecht etc.) ähnlich sind. Die Fälle können dann angeschaut werden, ähnliche Regionen sind markiert, und der Zugriff auf die (anonymisierten) Radiologieberichte ist möglich. Dabei werden RadLex-Terme hervorgehoben, in anderer Farbe falls verneint. Dies erlaubt es, schnell ein Maximum an Informationen zu bekommen, die relevant für die Befundung sind.

Der Prototyp erlaubt es auch, die Informationen der ersten 100 ähnlichen Resultate zu synthetisieren und in einer Art Differenzialdiagnose die wahrscheinlichsten Diagnosen zu vergleichen (s. rechte Seite in **Abb. 3**).

Khresmoi Radiology funktioniert im Moment mit einem Datensatz der Medizinischen Universität Wien als Forschungsprototyp, weitere Projekte zur Verbesserung des Prototyps sind geplant.

Fazit für die Praxis

Die Analyse großer Datenmengen, auch im Bereich der Bildgebung, kann Entscheidungsunterstützung in der Radiologie weiterbringen. Auch scheinbar einfache Aufgaben wie die vollautomatische Segmentierung von Organen in sehr unterschiedlichen Bilddaten können helfen, eine Basis für Entscheidungsunterstützung und Prozessautomatisierung zu erstellen. Dazu ist Input

von Experten notwendig, aber die Interaktion kann relativ niedrig gehalten werden, wenn Daten automatisch analysiert werden können und von ihnen gelernt werden kann.

Große Datenmengen können nicht mehr per Harddisk verschickt oder heruntergeladen werden, sondern verteilte Datenstrukturen, wie in der Cloud, sind notwendig. Die Algorithmen müssen zu den Daten gebracht werden und nicht andersherum. Dies kann auch für einen besseren Datenschutz sorgen. Das beschriebene Modell zentral organisierter Daten kann es auch erlauben, Daten in den Krankenhäusern zu belassen und nur virtuelle Maschinen für die Datenanalyse zu duplizieren und zu importieren. Da diese VMs nach der Analyse gelöscht werden können, besteht auch keine Gefahr, dass Daten heimlich exportiert werden. Das Modell erlaubt eine hohe Reproduzierbarkeit der Studien und möglicherweise den Zugriff auf sehr viel mehr Daten, als jetzt in den Studien üblich verwendet werden. Eine fehlende Reproduzierbarkeit ist schon häufig als sehr problematisch beschrieben worden, v. a. in den Bereichen der Informatik aber auch in Studien mit kleinen Fallzahlen.

In verschiedenen industriellen Bereichen hat die Analyse großer Datenmengen schon zu wichtigen Ergebnissen geführt, dies könnte auch in der Radiologie der Zukunft der Fall sein.

Korrespondenzadresse



H. Müller

HES-SO
Rue du TechnoPôle 3,
3960 Sierre, Schweiz
henning.mueller@hevs.ch

Einhaltung ethischer Richtlinien

Interessenkonflikt. H. Müller ist im Advisory Board der Firma ZebraMed und hat ein finanziertes Projekt mit der Firma ContextVision, beides aber nicht direkt mit dem Inhalt des Artikels verbunden. A. Hanbury gibt an, dass kein Interessenkonflikt besteht. Das VISCERAL Projekt hat kostenlose Cloud Ressourcen der Firma Microsoft bekommen.

Alle Studien der beschriebenen Projekte sind mit Bestätigung der Ethikkommissionen der beteiligter Krankenhäusern durchgeführt wurden und folgen den Richtlinien der Ethikkommission.

Literatur

1. Hanbury A, Müller H, Langs G, Weber MA, Menze BH, Salas Fernandez T (2012) Bringing the algorithms to the data: cloud-based benchmarking for medical image analysis. In: Catarci (Hrsg) Information access evaluation. Multilinguality, multimodality, and visual analytics. Springer, Heidelberg, 524–29
2. Langs G, Müller H, Menze B, Hanbury A (2013) Visceral: Towards large data in medical imaging – challenges and directions. In: Medical Content-Based Retrieval for Clinical Decision Support. Springer, Heidelberg
3. High Level Expert Group on Scientific Data (2010) How Europe can gain from the rising tide of scientific data – final report of the high level expert group on scientific data. <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>. Zugegriffen: 27. Oktober 2015
4. Kaplan DA (2013) What Radiologists Should Do About Incidental Findings, Diagnostic Imaging. <http://www.diagnosticsimaging.com/ct/what-radiologists-should-do-about-incidentalfindings>. Zugegriffen: 27. Oktober 2015
5. Yi SKM, Steyvers M, Lee MD, Dry MJ (2012) The wisdom of the crowd in combinatorial problems. *Cognitive Sci* 36(3):452–470
6. Krenn M, Dorfer M (2015) Oscar Alfonso Jimenez del Toro, Henning Müller, Bjoern Menze, Marc-Andre Weber, Allan Hanbury, Georg Langs, Creating a Large-Scale Silver Corpus from Multiple Algorithmic Segmentations, MICCAI medical computer vision workshop at MICCAI. In: MICCAI medical computer vision workshop at MICCAI, Munich, Germany