



HOCHSCHULE RUHR WEST
UNIVERSITY OF APPLIED SCIENCES

Eine bibliometrische Analyse relevanter Anwendungsgebiete von Big Data in der Medizin

Bachelorarbeit

Studiengang Wirtschaftsinformatik
der Hochschule Ruhr West

Samir Güngör

10011683

Erstprüferin:

Prof. Dr. Susanne Winter

Zweitprüfer:

Michael Schellenbach

Bottrop, 28. März 2022

Abstract

In recent years, the healthcare industry has increasingly relied on modern technologies. Conventional methods are supported by Big Data methods or are being investigated in the research. Has Big Data become more relevant in medicine in recent years? What does the future look like? In which medical subject area is Big Data being applied? These questions will be clarified during the thesis. In the first part, the usage of Big Data in medicine is shown and then, by using a bibliometric analysis, the importance and development of Big Data in medicine is presented. Afterwards, there is a discussion of the results followed with a summary and the future perspective. This thesis gives an overview about the currently technological possibilities and the potentials of Big Data in healthcare and medicine.

Inhaltsverzeichnis

| | |
|--|-----------|
| ABSTRACT | II |
| ABBILDUNGSVERZEICHNIS | V |
| ABKÜRZUNGSVERZEICHNIS | VI |
| 1 EINLEITUNG | 1 |
| 1.1 ALLGEMEINE BESCHREIBUNG DES THEMAS | 1 |
| 1.2 BIG DATA..... | 2 |
| 1.3 DIE 7 V'S VON BIG DATA | 4 |
| 1.3.1 <i>Volume</i> | 5 |
| 1.3.2 <i>Velocity</i> | 7 |
| 1.3.3 <i>Variety</i> | 8 |
| 1.3.4 <i>Veracity</i> | 9 |
| 1.3.5 <i>Variability</i> | 9 |
| 1.3.6 <i>Value</i> | 9 |
| 1.3.7 <i>Visualization</i> | 10 |
| 1.4 BIG DATA IN DER MEDIZIN | 10 |
| 1.4.1 <i>Herausforderungen</i> | 10 |
| 1.4.2 <i>Lernendes Gesundheitssystem</i> | 11 |
| 1.4.3 <i>Datenstromtechnologie</i> | 11 |
| 1.5 IT-SYSTEME UND IT-STANDARDS IN DER MEDIZIN..... | 12 |
| 1.5.1 <i>Die Anwendungen im Gesundheitswesen</i> | 12 |
| 1.5.2 <i>Digitalisierung</i> | 13 |
| 1.5.3 <i>Das heterogene und monolithische Informationssystemmodell</i> | 15 |
| 1.6 FRAGESTELLUNG DIESER ARBEIT..... | 17 |
| 2 MATERIAL UND METHODEN | 18 |
| 2.1 GOOGLE SCHOLAR | 18 |
| 2.2 BIBLIOMETRISCHE ANALYSE | 18 |
| 3 ERGEBNISSE | 21 |
| 3.1 AUSWAHL DER ANWENDUNGSGEBIETE..... | 21 |
| 3.2 IDENTIFIKATION DER ANWENDUNGSGEBIETE | 22 |
| 3.3 BIBLIOMETRISCHE ANALYSE VON BIG DATA | 26 |
| 3.4 BIBLIOMETRISCHE ANALYSE VON BIG DATA IN DER MEDIZIN | 28 |
| 3.5 GENOMDATEN | 30 |
| 3.5.1 <i>DNA-Sequenzierung</i> | 30 |
| 3.5.2 <i>Bibliometrische Analyse</i> | 31 |
| 3.6 RADIOLOGIE | 32 |
| 3.6.1 <i>Computertomographie mit iterativer Bildrekonstruktion</i> | 32 |
| 3.6.2 <i>Bibliometrische Analyse</i> | 33 |
| 3.7 PROTEINDATEN | 34 |
| 3.7.1 <i>Krebstherapie mittels Proteinanalyse</i> | 34 |
| 3.7.2 <i>Bibliometrische Analyse</i> | 35 |

| | | |
|----------|--|-----------|
| 3.8 | PATIENTENAKTE | 35 |
| 3.8.1 | <i>Elektronische Patientenakte</i> | 36 |
| 3.8.2 | <i>Bibliometrische Analyse</i> | 37 |
| 3.9 | MESSDATENANALYSE IN DER MEDIZIN | 38 |
| 3.9.1 | <i>Alltagsgerät: Wearables</i> | 38 |
| 3.9.2 | <i>Bibliometrische Analyse</i> | 38 |
| 3.10 | BIG DATA IN DER MOLEKULAREN DIAGNOSTIK/RHEUMATOLOGIE | 39 |
| 3.10.1 | <i>Hochdurchsatztechnologie</i> | 39 |
| 3.10.2 | <i>Bibliometrische Analyse</i> | 41 |
| 4 | DISKUSSION | 42 |
| 5 | ZUSAMMENFASSUNG UND AUSBLICK | 45 |
| | LITERATURVERZEICHNIS | 47 |
| | ERKLÄRUNG | 52 |

Abbildungsverzeichnis

| | | |
|---------------|--|----|
| ABBILDUNG 1: | REIFEGRADMODELL NACH GARTNER [5]..... | 4 |
| ABBILDUNG 2: | EIGENSCHAFTEN VON BIG DATA [9]..... | 5 |
| ABBILDUNG 3: | DATENVOLUMEN VON IP-TRAFFICS WELTWEIT VON 2014 - 2017 MIT EINER PROGNOSE BIS 2022 [12] | 6 |
| ABBILDUNG 4: | ENTWICKLUNG DES ÜBERTRAGENEN DATENVOLUMENS IM MOBILFUNK VON 2005 BIS 2020 [13] | 7 |
| ABBILDUNG 5: | DER AUFBAU VON E-HEALTH-ANWENDUNGEN | 12 |
| ABBILDUNG 6: | MEDIZINISCHE NUTZUNG VON GENOM-DATEN [27]..... | 14 |
| ABBILDUNG 7: | HETEROGENES KRANKENHAUSINFORMATIONSSYSTEMMODELL..... | 16 |
| ABBILDUNG 8: | MONOLITHISCHE KRANKENHAUSINFORMATIONSSYSTEMMODELL..... | 17 |
| ABBILDUNG 9: | ANFORDERUNGEN UND LEISTUNGEN DER RADIOLOGIE SEIT 2014 MIT PROGNOSE BIS 2023 [32]..... | 23 |
| ABBILDUNG 10: | ANZAHL ZU PROBENDATEN ZU GENEXPRESSIONSANLAGEN UND SEQUENZIERUNGEN [38]..... | 26 |
| ABBILDUNG 11: | ANZAHL DER PUBLIKATIONEN PRO JAHR VON „BIG DATA“ IM TITEL..... | 27 |
| ABBILDUNG 12: | ANZAHL DER PUBLIKATIONEN PRO JAHR VON „BIG DATA“ IM TEXT | 27 |
| ABBILDUNG 13: | ANZAHL DER PUBLIKATIONEN PRO JAHR VON „BIG DATA“ UND „MEDICINE“ IM TITEL | 28 |
| ABBILDUNG 14: | ANZAHL DER PUBLIKATIONEN PRO JAHR VON „BIG DATA“ UND „MEDICINE“ IM TEXT..... | 28 |
| ABBILDUNG 15: | ANZAHL DER PUBLIKATIONEN PRO JAHR VON „BIG DATA“ UND „MEDIZIN“ IM TITEL | 29 |
| ABBILDUNG 16: | ANZAHL DER PUBLIKATIONEN PRO JAHR VON „BIG DATA“ UND „MEDIZIN“ IM TEXT..... | 29 |
| ABBILDUNG 17: | PREIS PRO GENOMANALYSE [41]..... | 30 |
| ABBILDUNG 18: | ANZAHL DER PUBLIKATIONEN PRO JAHR VON „BIG DATA“ UND „GENE“ IM TEXT | 31 |
| ABBILDUNG 19: | CT MIT MITTELBASIERTER BILDREKONSTRUKTION (RECHTS) [47] | 33 |
| ABBILDUNG 20: | ANZAHL DER PUBLIKATIONEN PRO JAHR VON „BIG DATA“ UND „RADIOLOGY“ IM TEXT | 34 |
| ABBILDUNG 21: | ANZAHL DER PUBLIKATIONEN PRO JAHR VON „BIG DATA“ UND „MACROMOLECULAR“ IM TEXT | 35 |
| ABBILDUNG 22: | ANZAHL DER PUBLIKATIONEN PRO JAHR VON „BIG DATA“ UND „ELECTRONIC HEALTH RECORD“ IM TEXT .. | 37 |
| ABBILDUNG 23: | ANZAHL DER PUBLIKATIONEN PRO JAHR VON „BIG DATA“ UND „MEDICAL MEASUREMENT“ IM TEXT | 39 |
| ABBILDUNG 24: | EFFEKTIVITÄT DER ALLGEMEINEN „ONE-FITS-ALL“ MEDIZIN MIT DER PERSONALISIERTEN MEDIZIN | 40 |
| ABBILDUNG 25: | ANZAHL DER PUBLIKATIONEN PRO JAHR VON „BIG DATA“ UND „RHEUMATOLOGY“ IM TEXT | 41 |

Abkürzungsverzeichnis

| | |
|------|--|
| CRP | C-reaktives Protein |
| CT | Computertomographie |
| ePA | elektronische Patientenakte |
| HTML | Hypertext Markup Language |
| IP | Internet Protocol |
| IT | Information Technology |
| JSON | Javascript Object Notation |
| KI | Künstliche Intelligenz |
| KIS | Krankenhausinformationssysteme |
| MRT | Magnetresonanztomographie |
| PACS | Picture Archiving & Communications Systems |
| RIS | Radiologieinformationssysteme |
| SQL | Structure Query Language |

1 Einleitung

Die voranschreitende Entwicklung der IT sowie der Technik, bringt neue Technologien hervor. Während der 1980er Jahre konnten die ersten Datenbanksysteme bis zu einem Terabyte an Daten speichern. Im Jahr 2012 stieg die tägliche Datenproduktion auf 2,5 Exabytes. 90 % aller Daten, die Nutzer¹ produzieren, wurden in den letzten zwei Jahren erzeugt. Die Definition der Datenverarbeitungstechnik „Big Data“ wird seit den 1990er Jahren stets verändert. Heutzutage wird mit dem Begriff „Big Data“ der Einsatz von Tools, wie smarte Datenbanken, für große Datenmengen verstanden. Big Data wird in der Wissenschaft und in vielen Anwendungsbereichen, wie in der Medizin, angewendet. Der Datenverkehr wird mit zunehmendem Grad an Digitalisierung stets höher, wodurch eine alleinige Aufnahme von Daten komplizierter wird. Big Data wird bereits in der Medizin angewendet und soll sich in naher Zukunft ausbreiten. In der Medizin hat Big Data eine der größten Ausschöpfungspotentiale. Die Wichtigkeit von Big Data zeichnet sich durch die 3 Punkte - elektronische Geschäftsprozesse, mehr Daten von Patienten, sowie die Digitalisierung des Gesundheitswesens - aus.

1.1 Allgemeine Beschreibung des Themas

Big Data hat in der Medizin ein breites Spektrum an Anwendungsgebieten. Dazu zählen genetische Analysen über die klinische Laborpraxis, fortgeschrittene medizinische Analysen und die Verbesserung der Qualität der Gesundheitsversorgung. Eine erfolgreiche Technik von Big Data in der Medizin sind die Streaming-Daten. Diese Daten kommen von verschiedenen Quellen, wie unter anderem Sensoren, die laufend Daten erzeugen. Mithilfe von Stream-Processing-Techniken werden Daten automatisiert verarbeitet. Nach dem Mooreschen-Gesetz wird die Anzahl an Transistoren in den Schaltkreisen alle zwei Jahre verdoppelt. Aus diesem Grund sind immer leistungsfähigere Systeme nötig. Komponenten wie die Datenleistung und die Datenübertragung nehmen mehr zu und die Clouds sind die gängigsten Speicherorte geworden. Die Big-Data-Technik ist heutzutage im Praxisalltag vorhanden, dies ist in beruflichen und auch privaten Lebenssituationen vorhanden. Für die Gesundheitsbranche ist eine effektive Datenanalyse und Datennutzung vorteilhaft, um eine Prävention vor Krankheiten oder gesundheitlichen Schädigungen zu ermöglichen.

¹ In der vorliegenden Abschlussarbeit wird aus Gründen der besseren Lesbarkeit das generische Maskulinum verwendet. Sämtliche Personenbezeichnungen in verschiedenen Formen gelten für alle Geschlechteridentitäten (m/w/d).

Das Ziel dieser wissenschaftlichen Arbeit ist es, zu erkennen, welchen Trendverlauf verschiedene medizinische Anwendungsgebiete in Verbindung mit Big Data in der Literatur haben und wie Big Data in der Medizin genutzt wird. Wird Big Data in der Medizin in der Fachliteratur zunehmend präsenter? Wie sieht die Prognose zu Big Data in der Medizin aus?

Die vorliegende Arbeit besteht aus den vier Teilen Einleitung, Material und Methoden, Ergebnisse und Diskussion mit einem abschließenden Fazit. In der Einleitung wird auf die Definition von Big Data und in der Medizin eingegangen, hierbei werden die 7 V's von Big Data und die Entwicklung hiervon dargestellt. Big Data in der Medizin sowie die IT-Systeme und die IT-Standards werden thematisiert. Im Anschluss ist eine Fragestellung der Arbeit vorhanden.

1.2 Big Data

Laut dem technischen Philosophen Brian Ballsun-Stanton gibt es in Big Data drei Datenkonzepte.

1. Data as facts

Das erste Datenkonzept sind faktenbasierte Daten (data as facts), diese sind Messdaten, die objektiv und reproduzierbar sind. Die Messdaten beinhalten wahre Aussagen [1].

2. Data as observations

Das nächste Datenkonzept sind beobachtungsbasierte Daten (data as observations), die die aufgenommenen faktenbasierten Daten beinhalten und kontextbezogene Wissensdaten benötigen. Durch Filterungen relevanter Daten ist ein optimaler Datengewinn möglich [1].

3. Data as communications

Daten als binäre Nachrichten (data as bits / data as communications) dienen als Kommunikationsmittel für Zeichensätze. Im Gegensatz zu data as facts, die auf wahren Aussagen aufbauen, bauen data as communications auf bedeutungstragenden Charakter von Daten im Vordergrund auf [1].

Die Definition von Big Data ist, dass Datenbestände gemeint sind, die eine große Menge an Daten speichern können, jedoch vielfältig in der Anwendung sind [1]. Es handelt sich um Daten, die mit konventionellen Softwaresystemen nicht zu bewältigen sind. Big Data wird als eine schnell verarbeitende und vielfältige Informationskapital angesehen, die eine innovative Informationsverarbeitung besitzt, um eine gute Informationsgrundlage für Entscheidungen zu bieten [2].

Die größten Herausforderungen für Unternehmen in Bezug auf Big Data sind die Analysen großer Datenvolumen, bessere oder neue Analysemöglichkeiten, Aufbau von Vorhersagemodellen sowie die Analyse von Informationen aus polystrukturierten Datenquellen. Die Vorteile der Nutzung von Big Data sind die besseren strategischen Entscheidungen, bessere Steuerung der operativen Prozesse und besseres Kundenverständnis beziehungsweise Verbesserung der Kundenerfahrung. Durch die Nutzung von Datenanalysen kann bis zu 10 % Kosten gespart und bis zu 8 % mehr Umsatz generiert werden [3]. Diese Data-Science-Gedanken sind jedoch nicht nur in der Wirtschaft vorhanden, sondern auch vermehrt im Gesundheitswesen zu beobachten. Biostatistiker und Medizincontroller nutzen in Ansätzen ebenfalls Datenanalysen [4].

Die Abbildung 1 zeigt das Reifemodell nach Gardner. Anhand dieses Modells kann aufgezeigt werden, welche Strategien in der Datenanalyse angewendet werden. Die vier Analysemethoden werden von der Beschreibung bis zur Vorhersage dargestellt. In der deskriptiven Analyse werden gängige Rückblickmetriken wie Volumen- und Effizienzmetriken angewendet. Eine Digitalisierung der nötigen Daten ist nötig, um den Arbeitsaufwand zu minimieren. Für Unternehmen zahlt sich eine Investition in die deskriptive Analyse aus, um Erkenntnisse aus vergangenen, nicht messbaren Informationen zu gewinnen. Die erste Analysemethode unterstützt mit ihrer Analyse die nächsten Stufen. In der diagnostischen Analyse geht es um die Erkennung von bestimmten Leistungsmustern. Weitere Handlungen sind die frühzeitige Erkennung von Fehlern in den Prozessen sowie die Findung einer Lösung zur Verhinderung der kritischen Situation. Es beinhaltet somit eine Korrelationsanalyse von Massendaten, womit Abhängigkeiten erkannt werden, und leitet entsprechende Maßnahmen vor.

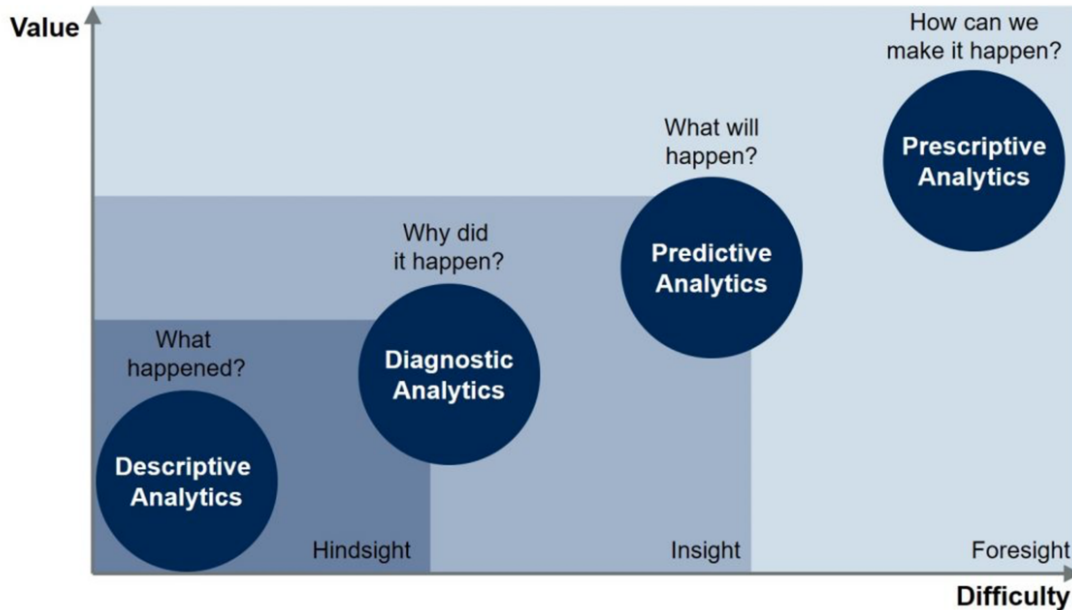


Abbildung 1: Reifegradmodell nach Gartner [5]

Im Gegensatz zur deskriptiven und diagnostischen Analyse, die mit Daten aus der Vergangenheit arbeiten und einen rückblickenden Ansatz verfolgen, kümmert sich die prädiktive Analyse um Vorhersagen. Um genaue Vorhersagen durchzuführen, sind Einblicke nötig, die mit Daten und Faktoren verglichen werden. Einige Faktoren sind zum Beispiel Geografie, Zeitpunkt, Produktkauf, Nutzungsmuster und Kundensegmente. Durch die eingehende Vorhersage von Mustern durch die Massendaten kann als Ergebnis der Analyse ein Prognosemodell möglicher zukünftiger Ergebnisse dargestellt werden. Diese leistungsfähige Analyse ist durch das maschinelle Lernen möglich [6].

In der letzten Methode des Reifegradmodells nach Gartner Inc. steht die präskriptive Analytik. Diese kann nach der Identifizierung von möglichen Ereignissen und Kennzahlen, die aus den Massendaten die Entwicklung der Muster beeinflussen können, erst vollständig durchgeführt werden. Die präskriptive Analytik zeigt dabei verschiedenste Simulationsmodelle als Ergebnis. Somit ist es möglich, genauere Entscheidungen zu treffen und Risiken beim Vorgehen zu reduzieren.

1.3 Die 7 V's von Big Data

Big Data wird in den meisten Beschreibungen mit drei oder fünf V's definiert. Die drei bekanntesten Eigenschaften von Big Data sind Volume, Velocity und Variety. Diese drei Definitionen werden seit dem Jahr 2001 mit Datenmanagementsystemen assoziiert [7]. Im Nachhinein wurden die Eigenschaften um Veracity, Variability, Value und Visualization ergänzt [8]. Diese Definitionen sind die

etabliertesten Eigenschaften von Big Data. Weitere Eigenschaften wurden in verschiedenen Berichten oder Studien erwähnt [9].

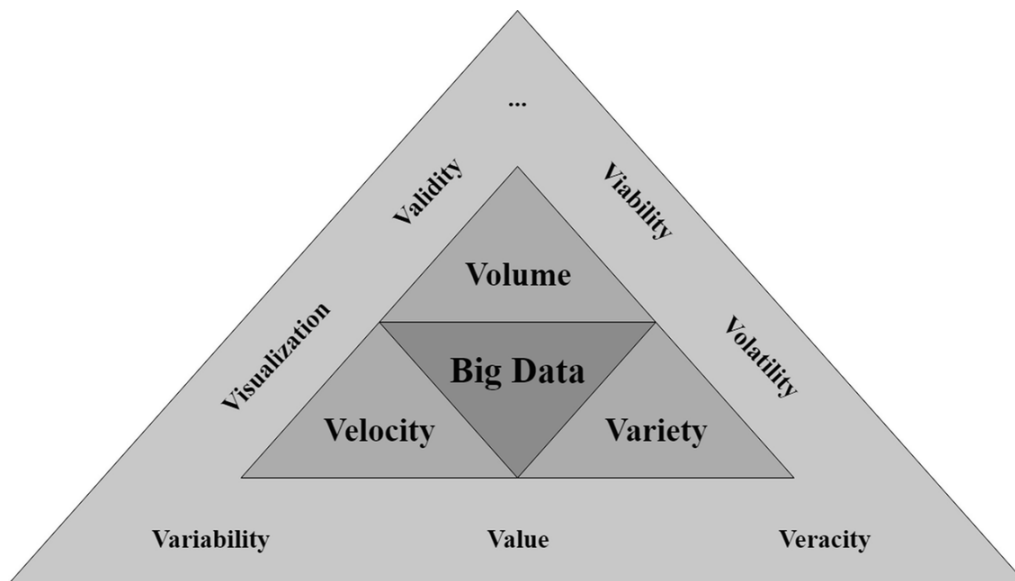


Abbildung 2: Eigenschaften von Big Data [9]

1.3.1 Volume

Eines der wichtigsten Eigenschaften von Big Data ist Volume. Es handelt sich hierbei um eine Datengröße, die jederzeit wächst und umfangreich ist. Der Datenbestand liegt im Tera- bis Zettabytebereich [10]. Zum Datenbestand gehören alle existierenden Daten auf der Welt. Einige Beispiele sind die Kunden-, Verkehrs-, Stamm-, Bewegungs- und Social Mediadaten, ebenso sind die Daten aus allen Geräten gespeichert, die Daten sammeln können wie Smartphones, Wearables, Navigationssysteme, Smart Home Geräte und Patientendaten in der Medizin [11]. In der nachfolgenden Grafik werden die privaten und geschäftlichen IP-Traffics weltweit von 2014 bis 2017 dargestellt sowie eine Prognose des Datenverkehrs bis zum Jahr 2022 gestellt. In der x-Achse werden die Jahre dargestellt und die y-Achse zeigt das Datenvolumen in Exabytes pro Monat.

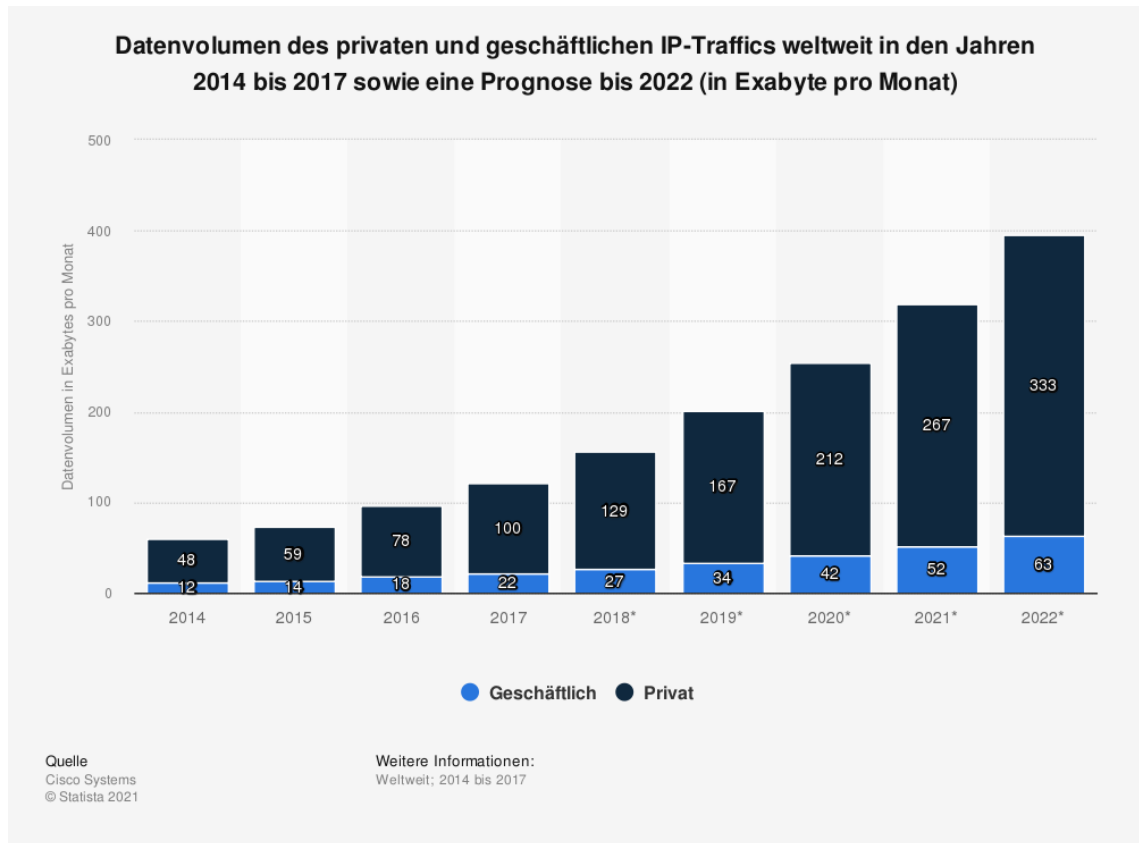


Abbildung 3: Datenvolumen von IP-Traffics weltweit von 2014 - 2017 mit einer Prognose bis 2022 [12]

Der weltweite Datenverkehr ist in Exabyte pro Monat dargestellt. Innerhalb von vier Jahren haben sich die IP-Traffics um mehr als verdoppelt. Privater Datenverkehr hat einen höheren Anstieg als der geschäftliche Datenverkehr, da er sich um 52 Exabytes pro Monat erhöht und der geschäftliche Datenverkehr um 10 Exabytes pro Monat. Somit hat der private Datenverkehr einen fünffach höheren Anstieg als der geschäftliche Datenverkehr. Laut der Prognose soll der weltweite Datenverkehr im privaten Bereich bis 2022 auf 333 Exabytes pro Monat steigen, während der geschäftliche Datenverkehrsanstieg um 41 Exabytes pro Monat auf 63 Exabytes pro Monat steigen wird.

Im nachfolgenden Diagramm wird der mobile Datenverkehr von 2005 bis 2020 in Deutschland veranschaulicht. Die x-Achse stellt hierbei das Jahr dar und die y-Achse zeigt das Datenvolumen in Millionen Gigabyte an.

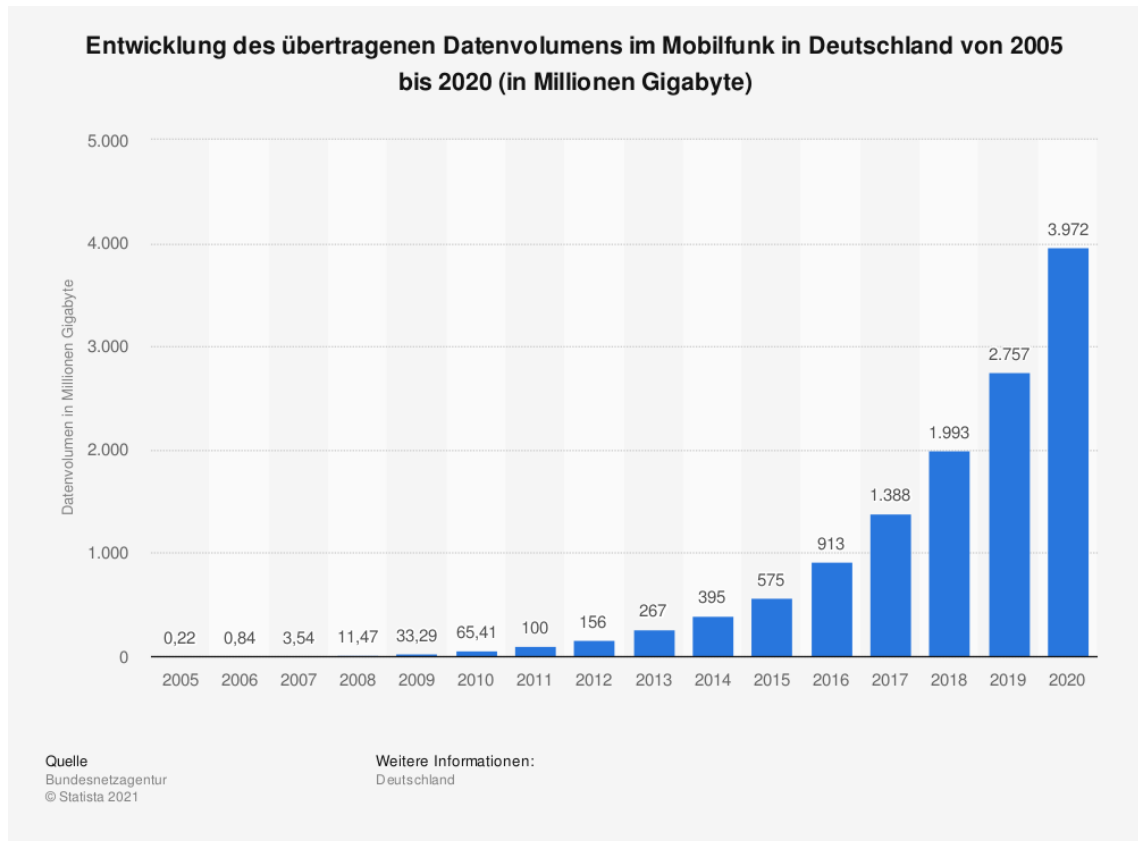


Abbildung 4: Entwicklung des übertragenen Datenvolumens im Mobilfunk von 2005 bis 2020 [13]

Im Jahre 2005 war das übertragene Datenvolumen im Mobilfunk gering. Nach der massenhaften Nutzung des Smartphones ab dem Jahr 2007 steigt das Datenvolumen rasant auf. Während 2006 noch 0,84 Millionen Gigabyte übertragen wurden, war es bis zum Jahr 2007 das Vierfache mit insgesamt 3,54 Millionen Gigabyte. Im Jahre 2007 kam der Release von Smartphones wie das Iphone raus, womit die Nutzung von Social Media wie Facebook möglich waren, das ebenfalls hohe Datenvolumen verbraucht, ist der weitere Aufstieg ab dem Jahr 2007 zu erkennen [14]. Innerhalb von sieben Jahren ist der Datenverkehr im Mobilfunk von 267 Millionen Gigabyte auf 3972 Millionen Gigabyte im Jahr 2020 gestiegen. Da der Datenverkehr im Mobilfunk zum Großteil aus privatem Datenverkehr besteht, ist ein größerer Zuwachs als im geschäftlichen Datenverkehr weiterhin zu erwarten.

1.3.2 Velocity

Eine weitere Definition von Big Datas 7 V's ist Velocity. Dieser gibt an, mit welcher Schnelligkeit die Daten erfasst und der Zugang für weitere Analysen erstellt werden können [15].

Eine Anforderung von Unternehmen für die Geschwindigkeit von Big Data ist idealerweise eine (nahezu) Echtzeit-Übertragung. Einen Richtwert für die

Geschwindigkeit aller Übertragungen zu definieren, wäre hierbei nicht möglich. Je nach Übertragungsgröße, sollte der Richtwert von Millisekunden bis hin zu Stunden gehen können beziehungsweise eine nach Batch, periodisch, nahezu Echtzeit oder Echtzeitübertragung haben [16].

1.3.3 Variety

Variety definiert die Vielfalt der Big Data Daten. Aufgrund der Zunahme der verschiedenen Variationen von Datenquellen werden ebenso die Datenmengen stetig größer, diese Daten werden von Geräten wie Smartphones, Fotokameras und Applikationen wie bei den sozialen Medien sowie mit Datenübertragungen durch das Internet produziert. Anhand der Vielfalt der Daten kann eine dreistufige Datenstruktur identifiziert werden.

In der ersten Stufe wären die strukturierten Daten in einem festgelegten Datenbankmodell, die eine gleichmäßige Struktur haben, und auf einer bestimmten Art erfolgt ihre Ordnung und Verknüpfung. In der Datenbank werden Datenelemente und Relationen verdeutlicht, hierdurch ist eine bessere Verwaltung sowie der vereinfachte Zugang auf die Daten möglich. Die SQL-Datenbank (Structured Query Language) ist ein Datenbankmodell für strukturierte Daten.

In der mittleren Stufe sind die semistrukturierten Daten. Diese enthalten meistens die Datenstrukturinformationen, enthalten jedoch kein einheitliches Schema. Für flexiblere Datenbanken sind semistrukturierte Daten geeignet, da diese durch ihre tiefen, unregelmäßigen und volatilen Strukturen oft ohne wiederkehrende Komponenten besser darstellbar sind. Eine hohe Flexibilität stellt einen Nachteil beim Auslesen und Verarbeiten der Daten dar. Beispiele für die Verwendung von semistrukturierten Daten sind XML-, HTML- und JSON-Daten, ebenfalls sind E-Mails mit einem strukturierten Header geeignet.

Die automatische Verarbeitung ist bei unstrukturierten Daten aufgrund der fehlenden Struktur erschwert. Eine Modellierung, um strukturierte Daten zu gewinnen, führt oft zu einem Datenverlust. Protokolldateien wären ein Beispiel von unstrukturierten Daten [17].

Zur Aufbereitung der Daten werden manuelle Strukturierungen und Verfahren wie Textanalysen und Text Mining, neuronale Netze, statistische Bayes-Klassifikationen, machine Learning Systeme, latent semantische Analysen und linguistische Verfahren eingesetzt. Mittels dieser Verfahren werden unter anderem die Stimmungslagen durch Sentimentanalysen analysiert [18].

1.3.4 Veracity

Veracity (z. dt. Wahrhaftigkeit) bezieht sich auf den Grad des Vertrauens auf die Daten und das hohe Niveau, bezüglich der Qualität und der Genauigkeit der Daten. Es überprüft, ob die Daten einheitlich, konsistent und kontrollierbar sind [19]. Daten, die gesammelt wurden, können unvollständig, ungenau oder nicht von Nutzen für die Datenanalyse sein. Für die Datenanalyse ist es wichtig, dass gesammelte Daten ebenfalls effektiv sind, unnütze Informationen würden keine Erkenntnisse bringen. Wenn zum Beispiel die Patientenakte unvollständig ist, können die Ärzte nicht bestens über die gesundheitliche Lage des Patienten urteilen, diese fehlenden Informationen können unter Umständen die Gesundheit des Patienten gefährden [20].

1.3.5 Variability

Variability bezieht sich auf veränderbare Daten. Der Kernpunkt ist es, die Rohdaten zu erkennen und zu definieren. Wenn zum Beispiel die Kombination einzelner physiologischer Parameter nicht die gleiche klinische Aussagekraft hat, haben diese Daten in der Big Data eine hohe Variabilität [21].

1.3.6 Value

In Big Data bedeutet Value (z. dt. Wert), dass die gesammelten Datenmengen für das Unternehmen in einem Mehrwert definiert werden. Hierbei können die Datenmengen in der Hinsicht auf das monetäre oder den gesellschaftlichen Nutzen unterstellt werden. Durch die vielfältigen Einsatzmöglichkeiten und der Menge an unterschiedlichen Anwendungsbereichen hat Big Data einen hohen Stellenwert sowie ein hohes Wertschöpfungspotential für Unternehmen aus den verschiedensten Branchen, die individuell den Stellenwert durch verschiedene Methoden erfassen können.

1. Schaffung von Transparenz

Die Nutzung von Daten aus Organisationen für zukünftige Entscheidungen erschafft einen Aufbruch der Hindernisse zwischen den Abteilungen, wodurch das Unternehmen sowie die Prozesse transparenter werden. Der positive Effekt ist eine kürzere Zeit zur Markteinführung von Produkten und Dienstleistungen.

2. Datengesteuerte Entdeckung

Je größere Datenmengen gespeichert werden, desto mehr Leistungsdaten sind für ein Unternehmen verfügbar. Somit können unerkannte Potenziale, Muster und Trends erkannt werden.

3. Individualisierung und Segmentierung

Eine Anpassung von Angeboten steigert den generierten Umsatz. Durch Big Data können Kundenprofile durch das Kundenverhalten erstellt werden, um passende Produktangebote zu erstellen.

4. Automatisierung

Durch Algorithmen können Geschäftsprozesse automatisiert werden. Hierbei werden Kosten und Zeit gespart.

5. Innovation neuer Produkte

Durch die Datenanalyse kann der Wunsch nach neuen Produkten durch vorhandene Produkte erkannt werden. Hierbei wird die Nachfrage prognostiziert, die ohne Big Data nicht erkannt werden könnte.

Neben den oben aufgeführten Methoden haben Unternehmen dennoch andere Herausforderungen, die durch Big Data gelöst werden können. Die Vorteile von Big Data helfen Unternehmen und der Gesundheitsbranche, Prognosen und Herausforderungen zu erkennen und Vorteile hierdurch zu schaffen [22].

1.3.7 Visualization

Die graphische Darstellung in Big Data kommt von Visualization (z. dt. Visualisierung). Tabellenkalkulationen und Berichte werden durch Diagramme und Schaubilder veranschaulicht. Die Veranschaulichung bietet den Nutzern, schnellere Erkenntnisse aus den gewonnenen Daten zu sammeln, zu verstehen und umzusetzen [21].

1.4 Big Data in der Medizin

1.4.1 Herausforderungen

Durch die effektive Nutzung von massenhaften Daten sowie der Speicherung und ihrer Verknüpfung verspricht Big Data, in der Medizin große Fortschritte zu ermöglichen. Ebenfalls kommen jedoch neue ethische und datenschutzrechtliche Fragen auf. Die Mischung aus der Statistik, Mustererkennung sowie der KI ermöglichen neue Wissenszugänge, um eine Prävention oder Früherkennung von Krankheiten oder Risikofaktoren zu erlangen. Prognosen, die durch Algorithmen in der Big Data ausgewertet werden, sind nicht immer geeignet, hierbei sind die Auswertungen von der Tiefe der Algorithmen sowie ebenfalls der nötigen Daten abhängig. Ebenfalls ist eventuell ein menschliches Urteilsvermögen nötig. Hierbei ist die datenschutzrechtliche Frage zu stellen, ob dies erlaubt, wäre [23].

Erfolgreiche KI-Lernmethoden und Mechanismen aus anderen Branchen sind Szenarien, in denen tausende von Daten von Menschen ausgewertet werden. Durch die manuelle Auswertung von Daten kann die KI ihren Algorithmus verfeinern und eine genauere Datenanalyse durchführen. Dies ist, in Anbetracht auf Menschenleben, jedoch eine Herausforderung, da für die Probeläufe qualifizierte Ärzte dabei sein müssen, was nicht zu bewerkstelligen ist. Ebenfalls wäre die praktische Nutzung in der Medizin nicht vorteilhaft, da ebenso mehrere „false Positive“ Werte herausgegeben werden können, die den Gesundheitszustand des Patienten täuschen können [24].

1.4.2 Lernendes Gesundheitssystem

Um in der Medizin große Mengen an Daten gezielt zu nutzen, ist das Paradigma lernendes Gesundheitssystems, zu betrachten. Das Konzept wurde 2007 in den USA durch das Institut of Medicine (IOM) veröffentlicht. Hierbei wurde ein Entwicklungsplan zur Lösung der Probleme vorgestellt. Die Ausrichtung von dem lernenden Gesundheitssystem basiert auf wissenschaftliche Anreize, Informatik, Wissenschaft und Arbeitsformen auf stetige Besserungen sowie Innovationen. Das neue Datenwissen wird hierbei als Nebenprodukt betrachtet und qualifizierte Verfahren werden in den Versorgungsprozess eingebettet. Durchlaufende Zyklen im System stellen die Daten aus verschiedenen Quellen zusammen, adäquate Verfahren können Daten analysieren, Befunde werden objektiv interpretiert und neues Wissen wird durch Rückführung von Ergebnissen in verschiedene Formate des Systems implementiert [1].

Daten werden im lernenden Gesundheitssystem in ein Netzwerk aus Institutionen, Technologien und wirtschaftlichen Interessen übertragen, um mit Werkzeugen in gesundheitliche Praxen integriert zu werden. Die Wissenschaft und die Gesellschaft gehen eine Symbiose ein, um das Gesundheitswesen zu bestimmen [1].

1.4.3 Datenstromtechnologie

Überwachungsgeräte, wie der Überwachungsmonitor auf der Intensivstation, nutzen die Streaming-Data-Technologie. Früher wurden aus den Überwachungsgeräten die Messwertdaten periodisch in Minutentakten oder jede halbe Stunde zusammengefasst und der Durchschnitt berechnet, obwohl Tausende von Messwerten pro Sekunde erzeugt wurden. Diese Daten wurden überwacht und gespeichert, jedoch gab es zum damaligen Zeitpunkt keine technische Datenanalyse, um immense Datenmengen zu analysieren. Heutzutage kann mit der Big-Data-Technologie (Streaming-Data) der Datenfluss überwacht, gespeichert und

durch verschiedene Algorithmen vom medizinischen Team verarbeitet werden, um zum Beispiel Frühwarnzeichen für physiologische Veränderungen im Organismus eines Patienten zu erkennen. Da Streaming-Daten Echtzeitdaten sind, können Ärzte jederzeit frühzeitig das Problem erkennen und präventieren. Mit Hilfe der Datenstromtechnologie können Ärzte dem Patienten 24 bis 72 Stunden eher helfen beziehungsweise ihn warnen als mit den klassischen Mitteln [21].

1.5 IT-Systeme und IT-Standards in der Medizin

1.5.1 Die Anwendungen im Gesundheitswesen

Eine jederzeit aktuelle Patientenübersicht über die Krankheitsgeschichte des Patienten, elektronische Anforderungen und die Darstellung von Untersuchungen, problemorientiertes Bereitstellen von Expertenwissen, eine durchlaufend elektronische Unterstützung von Arbeitsprozessen in der Medizin sowie die automatische Übernahme von Vorinformationen des Patienten für die Berichte ist einzuhalten. Solche Informationssysteme stellen kontextbezogene Daten in der Diagnose, Dokumentation, Behandlung und Abrechnung von Leistungen bereit. Diese Anwendungen werden in verschiedene Bereiche im Gesundheitswesen, wie in der folgenden Abbildung ersichtlich, gegliedert [25].

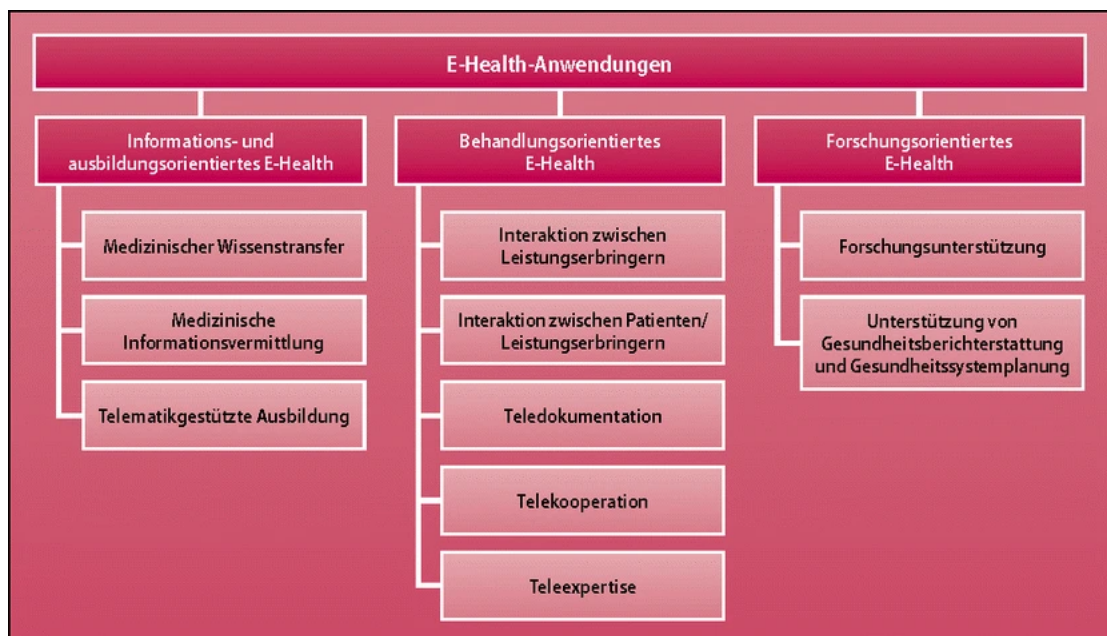


Abbildung 5: Der Aufbau von E-Health-Anwendungen

- **Behandlungsbezogene Anwendungen**

Zu den behandlungsbezogenen Anwendungen gehören eArztbrief, eRezept, eÜberweisung, eAbrechnung, die elektronische Gesundheitskarte, elektronische Patienten- und Gesundheitsakte, Monitoring-Systeme und Einweiseportale. Diese haben eine Interaktion zwischen Leistungserbringer, Teledokumentation, Telekooperation und Teleexpertise.

- **Informations- und ausbildungsbezogene Anwendungen**

Der medizinische Wissenstransfer, die Informationsvermittlung und die telematische Ausbildung gehören zum Anwendungsgebiet. Zum Beispiel die Krankenhausinformationssysteme (KIS), Radiologieinformationssysteme (RIS) und Picture Archiving & Communications Systems (PACS) sind informations- und ausbildungsbezogene Anwendungen.

- **Forschungsbezogene Anwendungen**

Zu den forschungsbezogenen Anwendungen zählen die Unterstützung der Forschung sowie die Gesundheitssystemplanung und die Berichterstattung [26].

1.5.2 Digitalisierung

Seit den 2000er Jahren hat sich, in Bezug auf Digitalisierung, in der Arbeitsweise der Medizin viel verändert. Nach der ersten digitalen Röntgenfotografie kamen immer mehr Geräte und Systeme in der Gesundheitsbranche zum Einsatz. Einige Beispiele von der Digitalisierung sind, dass die alte Gesundheitskarte von jedem Patienten durch eine elektronische Gesundheitskarte beziehungsweise Versicherungskarte ersetzt wurde. Ebenfalls werden keine physischen Patientenakten mehr verwendet, diese wurden von elektronischen Patientenakten ersetzt. In diesen Patientenakten sind individuelle Daten von jedem Patienten gespeichert. Der Vorteil hierbei ist der unverzügliche Zugang zu allen Daten eines Patienten in jeder Arztpraxis. Die ersten Röntgenbilder, die auf Film aufgenommen wurden, wurden durch das PACS/RIS mit Digitalisaten ersetzt. In einigen Krankenhäusern ist nun anstatt der klassischen OP-Tische mit Feldbesteck eine videogestützte Chirurgie vorhanden. Andere Systeme, die nun analoge Hilfsmittel ersetzt haben, sind zum Beispiel das computergestützte Bettenmanagementsystem, das das klassische Buch zum Notieren von Informationen ersetzt hat, oder die Spektrometrie mit digitaler Datenausgabe anstatt einer klassischen Laboruntersuchung. In dem Big Data Bereich würde das durch KI unterstützte Untersuchen und Analysieren von Moleküldatenbanken und die Erforschung mittels künstlicher Intelligenz von Wirkmechanismen sein, die die experimentelle

Erforschung von Wirksubstanzen ersetzen. Eine simplere Digitalisierung ist der Wegfall von Briefen. Die darin enthaltenen Informationen können ebenfalls per Mail schneller gesendet werden [25].

Im nachfolgenden Diagramm wird der aktuelle Stand dargestellt, wie viele Krankenhausärzte prozentual die Genomdaten in ihrer Behandlung als Unterstützung nutzen.

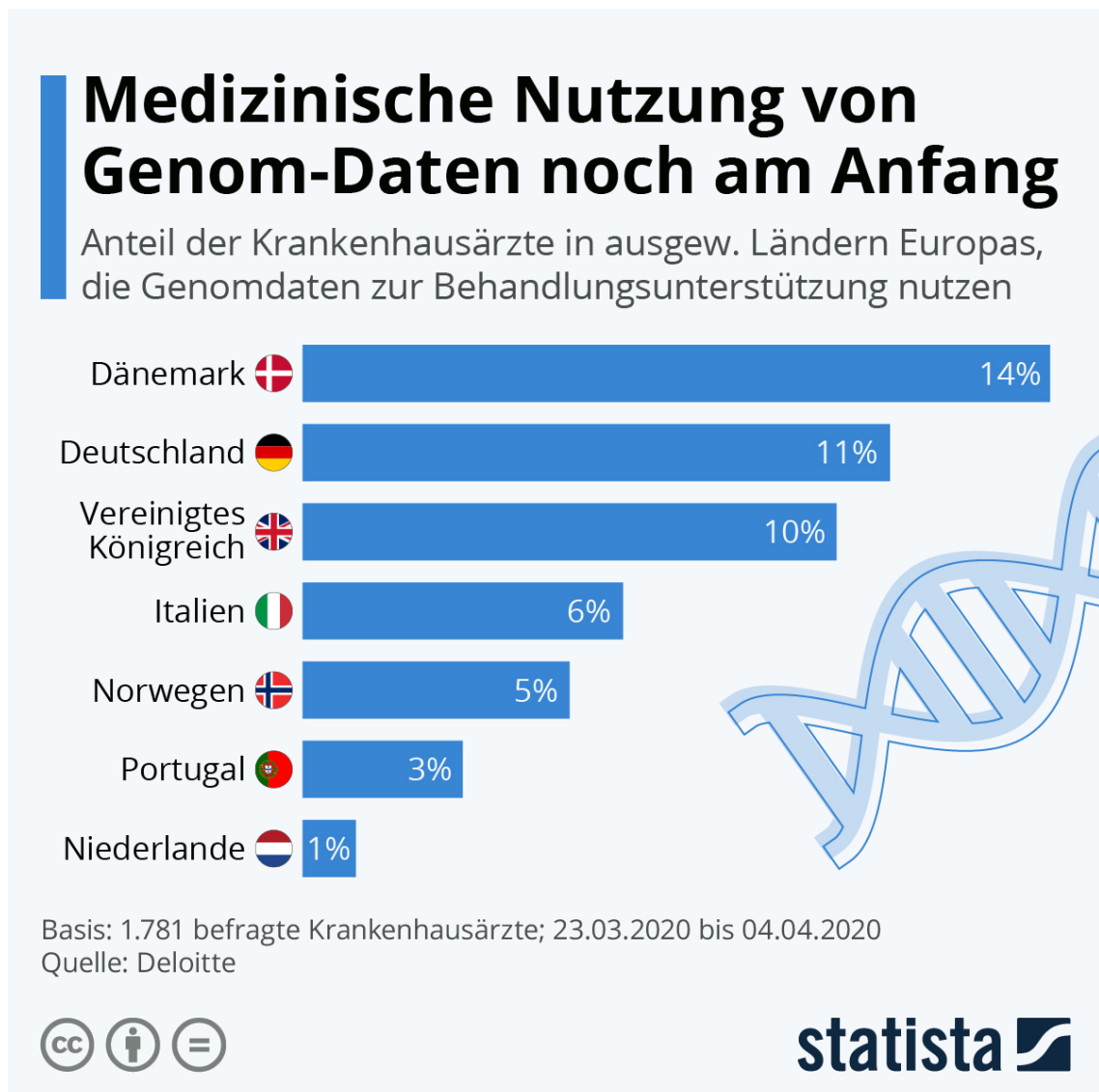


Abbildung 6: Medizinische Nutzung von Genom-Daten [27]

Hierbei ist zu erkennen, dass der höchste Anteil an Krankenhausärzten, die Genomdaten als Unterstützung nutzen, in Dänemark mit 14 % liegt. Darauf folgend greifen deutsche Krankenhausärzte mit einem Anteil von 11 % europaweit am zweithäufigsten auf Genomdaten als Unterstützung zurück. Als siebthöchster Anteil an Krankenhausärzten liegt die Niederlande mit einem Prozent.

1.5.3 Das heterogene und monolithische Informationssystemmodell

Da die verschiedenen Informationssysteme zahlreicher Anbieter in verschiedenen Gesundheitseinrichtungen in ihren Funktionen ähneln müssen, gibt es einige Anwendungen, die jedes Informationssystem speziell für Arztpraxen, die standardisierte Systeme nutzen, beherrschen müssen. Krankenhäuser haben für jeden Fachbereich spezialisierte Systeme wie zum Beispiel Radiologieinformationssysteme, Laborinformationssysteme, Apothekeninformationssysteme oder Blutbankinformationssysteme. Jedoch sollten alle Systeme diese aufgelisteten Anwendungen beinhalten.

- Verwaltung von Stammdaten
- Patientenverwaltung
- Gesetzliche Fallverwaltung
- Dokumentation der Behandlung (Elektronische Patientenakte)
- Organisation sowie die der Ablauf der Patientenbehandlung bei Terminen, Ablauf, Arbeitslisten etc.
- Unterstützung bei der Abrechnung von Leistungen mit einer automatischen Erstellung von Rechnungen aus der vorhandenen Dokumentation
- Controlling und Qualitätsmanagement in der Gesundheitsvorsorgeeinrichtung mit medizinischer und ökonomischer Auswertung
- Unterstützung in der externen und internen Kommunikation [28]

In der folgenden Abbildung werden die verschiedenen Informationssysteme in einer Gesundheitsvorsorgeeinrichtung miteinander kommuniziert und die Arbeitsstruktur dargestellt.

Hierbei ist zu beachten, dass im heterogenen Krankenhausinformationssystem eine zentrale Datenbank existiert, die mit dem Kommunikationsserver verbunden ist. Dieser Server empfängt und versendet Daten und Informationen an die verschiedenen Informationssysteme weiter.

Im Gegensatz zum heterogenen Informationssystem besitzt das monolithische Krankenhausinformationssystemmodell keinen eigenständigen Zentraldatenbestand. Alle Daten werden im Kommunikationsserver gelagert.

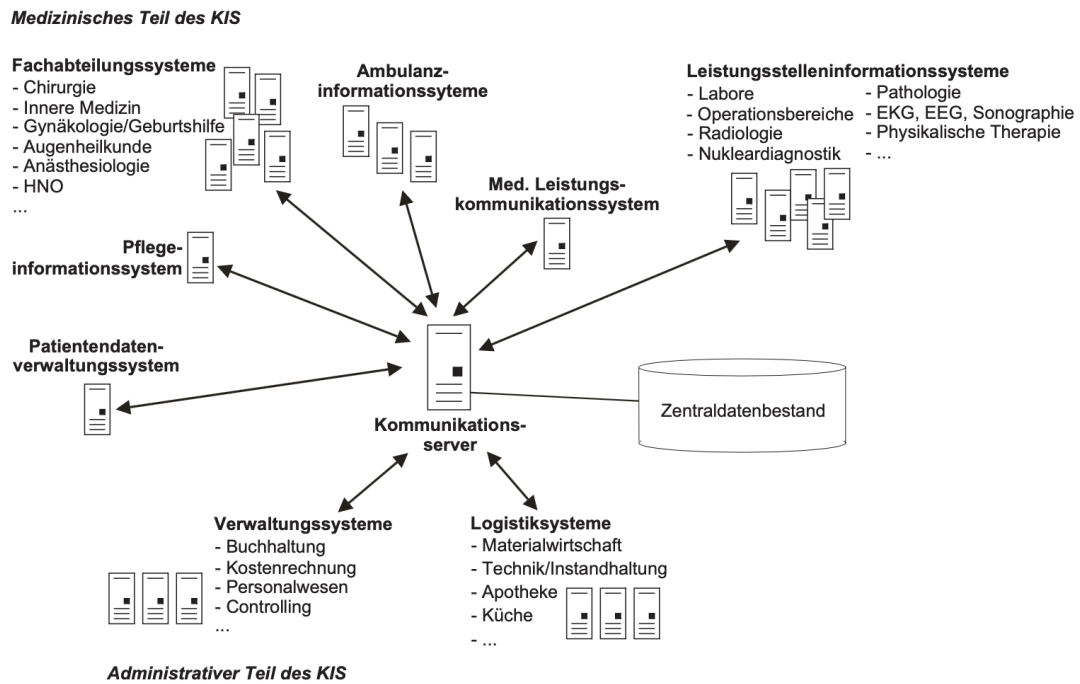


Abbildung 7: Heterogenes Krankenhausinformationssystemmodell

Im heterogenen Informationssystem sind mehrere Key-User in den Fachabteilungen nötig, um eine Weiterentwicklung sowie Wartung der Systeme zu ermöglichen. Eine zentrale IT würde dies in dem Ausmaß nicht bewerkstelligen können, da es sehr aufwendig wäre.

Zu dem medizinischen Teil der KIS gehören:

- Fachabteilungssysteme (Chirurgie, HNO, Gynäkologie)
- Ambulanzinformationssysteme
- Med. Leistungskommunikationssystem
- Leistungsstelleninformationssysteme
- Pflegeinformationssystem
- Patientendatenverwaltungssystem

Zu dem administrativen Teil der KIS gehören:

- Verwaltungssysteme
- Logistiksysteme

Eine Alternative zum heterogenen Krankenhausinformationssystem ist das monolithische Krankenhausinformationssystem, das viele Anwendungssysteme in einem Anwendungssystem gruppiert. Hierbei können jedoch weiterhin spezielle Informationssysteme wie RIS oder PACS extern verwaltet werden, die über einen internen Kommunikationsserver verbunden sind. Im monolithischen Ansatz kann

die IT-Abteilung, die Betreuung der Systeme selbst übernehmen, da der Aufwand nicht so groß ist [29].

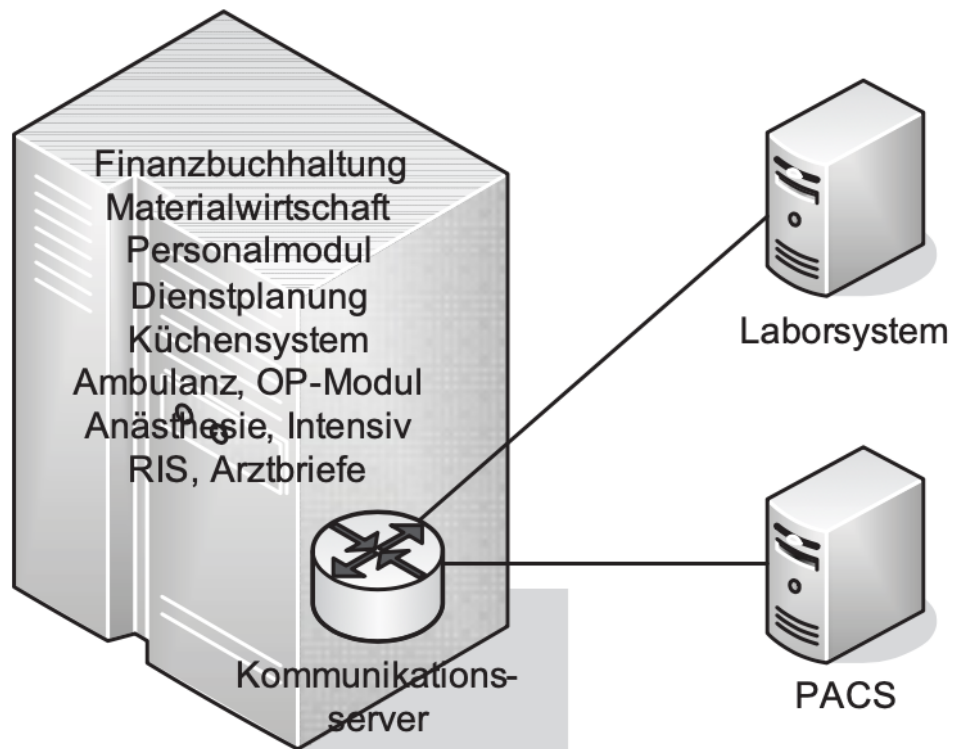


Abbildung 8: Monolithische Krankenhausinformationssystemmodell

1.6 Fragestellung dieser Arbeit

Die Fragestellung dieser Arbeit besteht darin, ob anhand der Auswertungen in der Bibliometrie eine Trendanalyse bezüglich Big Data in der Medizin möglich ist. Hierbei wird die Entwicklung der verschiedenen Literaturveröffentlichungen in den letzten Jahren beobachtet, in welcher Höhe die Anzahl der Veröffentlichungen jährlich steigen.

Ebenfalls werden die Auswahlkriterien für die Wahl der relevanten Anwendungsgebiete im Gesundheitswesen definiert und ausgewertet.

Die Leitfragen, die in Bezug dieser Abschlussarbeit gestellt werden, sind:

Wie hoch ist die Häufigkeit der Publikationsveröffentlichungen?

Welche relevanten Anwendungsgebiete gibt es in Big Data in der Medizin?

Was sagt die Trendanalyse in der bibliometrischen Analyse bei relevanten Anwendungsgebieten im Gesundheitswesen aus?

2 Material und Methoden

Im folgenden Kapitel werden die Materialien und Methoden dargestellt, die für den nachfolgenden Teil relevant sind.

Für die Trendanalyse werden im jährlichen Abstand die Anzahl der Literaturveröffentlichungen im Zeitraum von 2000 bis 2021 berücksichtigt. Die Angaben wurden der Literatursuchmaschine Google Scholar entnommen und ausgewertet. Hierbei wird hauptsächlich Literatur mit der deutschen oder englischen Sprache entnommen. Die Daten wurden am 17.02.2022 der Suchmaschine Google Scholar entnommen.

2.1 Google Scholar

Google Scholar ist eine kostenlose Suchmaschine explizit für Literatur, die von dem gleichnamigen Unternehmen Google bereitgestellt wird. Dieser hat kostenpflichtige Bezahlschranken. Durch die Bezahlung kann der Leseinhalt erweitert werden, ansonsten ist ein kostenfreier Abschnitt zum Lesen freigegeben. Als größte akademische Suchmaschine beinhaltet Google Scholar über 400 Millionen Dokumente in ihrer Datenbank. Durch die Eingabe von den Schlüsselwörtern beziehungsweise Suchbegriffen wie „Big Data“, „Big Data in der Medizin“ etcetera können genaue Angaben in der Literatursuche angegeben werden. Dieser zeigt die Anzahl der gefundenen Literaturergebnisse und dem Durchschnitt der jährlichen Publikation im Jahr an. Ebenfalls sind eine Filtrierung und Optimierung der erforderlichen Suchen möglich. In der „erweiterten Suche“ können mehrere Begriffe einzeln gesucht werden, nach dem Booleschen Operator OR (Oder-Gatter). Eine Suche der genauen Wortgruppe aus mehreren Worten ist ebenfalls möglich. Das Ausschließen von bestimmten Wörtern ist durch die Optionen im Filter möglich. Die Filtrierung ist auf Artikel oder auf den Text möglich. Der jährliche Zeitraum kann in der erweiterten Suche nach Belieben definiert werden.

2.2 Bibliometrische Analyse

Um eine bibliometrische Analyse durchzuführen, müssen vorher einige Regularien festgelegt werden.

Datenquelle:

Zuerst sollten die möglichen verfügbaren Datenquellen in Betracht gezogen werden, die für eine bibliometrische Analyse in Frage kommen könnten.

Als mögliche Datenquellen gelten Jahresberichte von Unternehmen, wissenschaftliche Berichte oder Veröffentlichungsdatenbanken.

Indikatoren zur Festlegung von passenden Datenquellen sind andererseits, ob das Fachgebiet zum Ziel der bibliometrischen Analyse passt, und es wird auf die Vollständigkeit der Datenquelle sowie die Aktualität geachtet. Es sollten hierbei Datenquellen mit vielen zur Verfügung stehenden Daten sowie aktuellen Quellen beinhaltet werden. Je aktueller die Quelle, desto höher ist die Aussagekraft der bibliometrischen Analyse [30].

Beobachtungszeitraum festlegen:

Der Beobachtungszeitraum legt fest, in welchem Zeitabstand die bibliometrische Analyse erfolgt. In der IT ist eine Obergrenze zwischen zehn und 20 Jahren am sinnvollsten, um keine veralteten Veröffentlichungen, die nicht mehr auf dem aktuellen Stand der Technik sind, zu benutzen.

Bei der Festlegung des Beobachtungszeitraumes ist jedoch auch zu betrachten, dass einige Themengebiete schnelllebiger sind als andere Themengebiete.

Als Beispiel ist hierfür das Themengebiet „Informatik“ mit dem Themengebiet „Germanistik“ zu vergleichen. Während die IT einen großen Wandel in den letzten 20 Jahren hatte, bleiben in der Germanistik auch Veröffentlichungen vor 20 Jahren meist aktuell, da sich die Sprache nicht so schnell ändert wie die Technik in der IT [30].

Indikatoren festlegen:

Indikatoren dienen zur Festlegung von bestimmten Regularien, die bei der Suche an Veröffentlichungen für die bibliometrische Analyse gelten sollen, um nicht das Bild zu verzerren.

Als Erstes sollte der oberste Indikator als die „Spitze einer Pyramide“ definiert werden. In dieser wissenschaftlichen Arbeit ist der oberste Indikator „Big Data in der Medizin“. Dieser Indikator wird darauffolgend in verschiedene kleinere Indikatoren aufgeteilt, um eine genauere Filtrierung der Veröffentlichungen zu ermöglichen.

Einige Indikatoren sind zum Beispiel die Veröffentlichungstypen, das Auffinden der gesuchten Begriffe im Titel oder im Text sowie auch die Literaturart (Primär oder Sekundärliteratur) [30].

Zeitliche Entwicklung wissenschaftlicher Veröffentlichungen

Unter dem Begriff „Zeitliche Entwicklung wissenschaftlicher Veröffentlichungen“ ist zu verstehen, dass dabei geachtet wird, wie der Verlauf an Literaturveröffentlichungen zu einem bestimmten Thema innerhalb mehrerer Jahre verläuft. Hierbei geht es um die reine Anzahl an Veröffentlichungen innerhalb eines Jahres.

Hierbei werden periodische Zeitabstände wie zum Beispiel jährlich, alle 2 Jahre oder Jahrzehnte als Beobachtungszeitraum genommen.

Mithilfe der Veröffentlichungsdatenbank „Google Scholar“ wird die Anzahl der Literatur in einer Periode entnommen und vermerkt.

Unter der zeitlichen Entwicklung des wissenschaftlichen Veröffentlichens wird die Relevanz der Themengebiete im jeweiligen Jahr erkannt [30].

3 Ergebnisse

Im folgenden Kapitel werden die Ergebnisse vorgestellt. Die Ergebnisse wurden für eine bessere Darstellung in mehrere Abschnitte aufgeteilt.

Im Folgenden wird die Identifizierung der einzelnen relevanten Anwendungsgebiete in der Medizin erläutert. Anschließend werden exemplarische Beispiele zu jedem Anwendungsgebiet genannt. Hiernach folgt die bibliometrische Analyse, indem die Häufigkeit, Relevanz sowie die zeitliche Entwicklung betrachtet wird.

3.1 Auswahl der Anwendungsgebiete

Relevante Anwendungsgebiete in Big Data in der Medizin haben die Eigenschaft, dass sie fortgeschrittene Informationssysteme im Fachbereich besitzen, diese somit mittel- und langfristig im jeweiligen Gebiet im Einsatz sind, Big Data als Technologie benutzen sowie viele digitale Daten aufnehmen und verarbeiten.

Eine ausreichend vorhandene Anzahl an Publikationen über die genannten Anwendungsgebiete ist ebenfalls relevant, um eine genaue Analyse in der Bibliometrie durchführen zu können.

In dieser Arbeit werden die relevanten Anwendungsgebiete Genomdaten, Radiologie/Bildverarbeitung, Proteindaten, Patientenakten, Messdatenanalyse in der Medizin sowie der Einsatz von Big Data in der molekularen Diagnostik und Rheumatologie erläutert. Es sind ausgewählte Anwendungsgebiete, um es prototypisch darzustellen.

Auswahlkriterien von den Anwendungsgebieten sind:

- Relevanz in der Medizin
- Anzahl des Vorkommens der Anwendungsgebiete in der Literatur
- Gewinn bei der Entscheidung der Behandlungsmethoden durch Big Data
- Zukunftsvoraussichten der Technologien im Anwendungsgebiet
- Aktueller Stand von Big Data Methoden in den jeweiligen Fachgebieten
- Relevanz der Big Data Technologien während der Behandlung

Aufgrund der vielfältigen und zahlreichen Fachgebiete in der Gesundheitsbranche wurden durch die Auswahlkriterien die nennenswerten Anwendungsgebiete ausgewählt.

Die spezifische Filtrierung der Anwendungsgebiete in der Gesundheitsbranche dient zur besseren Erläuterung und Darstellung der Möglichkeiten von Big Data in diesen Fachgebieten.

3.2 Identifikation der Anwendungsgebiete

Relevante Anwendungsgebiete in der Medizin, wie Genomik, Radiologie, Proteine, Patientenakte, künstliche Intelligenz und machine Learning in der Medizin und die molekulare Diagnostik haben einen besonderen Stellenwert in der Gesundheitsbranche. Dies sind entweder Anwendungsgebiete, die erste Big Data Methodiken im Praxisalltag nutzen, wie zum Beispiel die Computertomographie. Es können auch Methodiken in der Forschung sein, die erste oder vollständige Ergebnisse, wie in der DNA-Sequenzierung erzielt haben.

Ebenfalls wurden die Anwendungsgebiete in den letzten Jahren in der Literatur mit Big Data in Verbindung gebracht. Inwieweit die Fachgebiete in der Literatur veröffentlicht wurden und ob ein Trend zu erkennen ist, wird in der folgenden bibliometrischen Analyse näher erläutert. Hierbei wird die bibliometrische Analyse das Vorkommen der Anwendungsgebiete in Titel sowie im Text darstellen.

Eine zukünftige Nutzung von Big Data Methodiken zur Verbesserung der Therapiebehandlungen oder Datenanalysen ist in den genannten Anwendungsgebieten zu erwarten.

Genomdaten:

Die Sequenzierung und Entzifferung von Gendaten werden immer häufiger für verschiedene Zwecke wie zum Beispiel für die personalisierte Medizin oder für diverse Erbgutanalysen angewendet. Hierbei wird davon ausgegangen, dass die weltweite Anzahl der Daten an Genomik vielfältig steigen wird. Im Jahre 2020 ist Deutschland dem europäischen Großprojekt „1+ Million Genomes Initiative“ beigetreten und unterstützt fortan die Forschung und Entwicklung des Großprojektes. Ziel dieser Initiative ist, eine internationale und länderübergreifende Datenbank zu erstellen, die durch einen sicheren und geregelten Zugang zu mindestens einer Million kompletter Genomsequenzen und anderer Gesundheitsdaten den Zugriff ermöglichen lässt. Ein solches Großprojekt unterstreicht die Wichtigkeit der Genomdaten in der Zukunft [31].

Ein umfassender Zugang zu einer Genomdatenbank beziehungsweise anderen medizinischen Daten fördert die Erforschung sowie die Behandlung von Krankheiten, dies kann die positive Entwicklung und Heilung der Krankheit voranschreiten lassen.

Ebenfalls werden Gendaten immer mehr von Pflanzen, Tieren und anderen Organismen erkannt, aufgenommen, weiterverbreitet und eine Analyse von diesen Genträgern durchgeführt.

Radiologie:

Der medizinische Fachbereich Radiologie wird mittels KI sowie ebenfalls Algorithmen stetig in der Technik verbessert.

Im nachfolgenden Diagramm ist zu sehen, dass die Anzahl an aufgenommenen Daten in der Radiologie stetig ansteigt. Von 2014 bis 2018 gab es ein Anforderungswachstum an Daten von circa 36,31 %. Im Gegensatz dazu stieg die Anzahl an Personal in der Radiologie um circa 18,83 %. Laut der Prognose wird das Wachstum an Daten in den folgenden Jahren stetig höher werden, während die Anzahl an Personal nicht mit einer gleich hohen Wachstumsrate steigen wird. Dies hat zur Folge, dass das Personal immer mehr unterbesetzt ist und die Aufgaben in der vorgegebenen Zeit nicht schaffen könnte. Durch diesen Zeitdruck können Radiologen weniger Zeit pro Behandlung aufbringen. Ein Nachteil hierbei ist, dass je weniger Zeit pro Behandlung auch eine schlechtere Behandlung des Patienten durch den Radiologen zur Folge haben kann, zum Beispiel durch eine falsche Beratung sowie auch durch die falsche Beurteilung der Radiologiedaten vom Radiologen.

Hierbei ist eine Unterstützung der Radiologie durch die Technologien möglich, die Big Data bietet, die ebenfalls die Daten selbstständig auswerten können. Die Unterstützung durch die Verwendung von Algorithmen sowie der künstlichen Intelligenz.

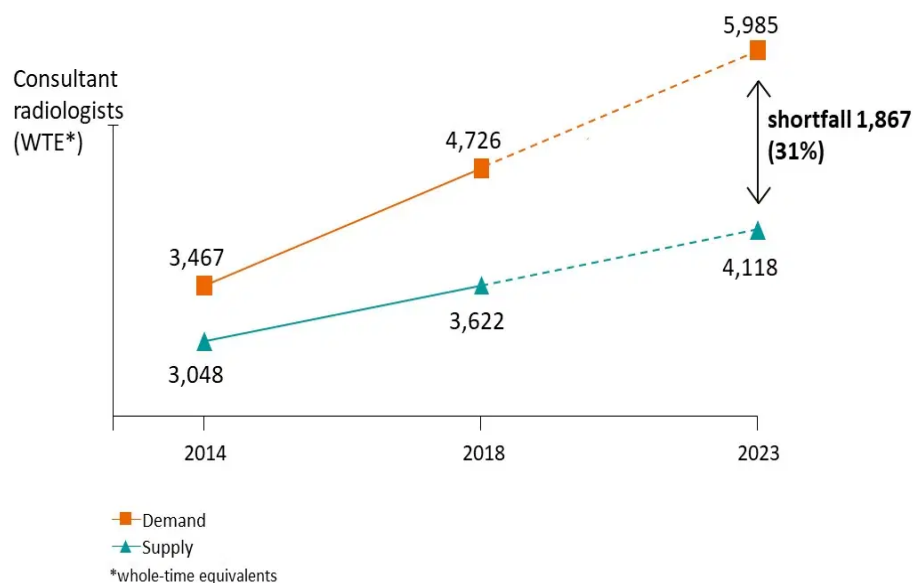


Abbildung 9: Anforderungen und Leistungen der Radiologie seit 2014 mit Prognose bis 2023 [32]

Proteindaten:

Die Gesamtheit aller Proteine, die im menschlichen Körper enthalten sind, werden Proteom genannt. Dieser besteht aus verschiedenen Proteintypen, die Millionen verschiedene Aufgaben und Funktionen erfüllen [33].

Die Zusammensetzung von Proteinen im menschlichen Körper ist den Wissenschaftlern gemeinhin bekannt. Die Funktionsweisen und das Nutzen von Proteindaten sind jedoch bislang größtenteils unbekannt. Neue Erkenntnisse über Proteine werden durch verschiedene Studien gewonnen. Durch das technische Verfahren „machine Learning“ werden alte Proteindaten neu sortiert und analysiert. Mit der Nutzung von speziellen Algorithmen werden Datensätze vom menschlichen Körper und dem Proteom erstellt [33].

Die Forschung sowie die technische Nutzung von Proteindaten sind derzeit in den Anfangsstufen. Es wird davon ausgegangen, dass eine personalisierte Krebstherapie mit Fokus auf diverse Proteine den Krebs besser bekämpfen könnte. Durch Proteinanalysen sollen akkuratere Behandlungen empfohlen werden können.

Die Bekämpfung der Krebstherapie ist eine äußerst relevante Methodik in der Big Data Anwendung in der Medizin [33].

Patientenakte:

Eine zentrale Rolle im Themengebiet „Big Data in der Medizin“ ist die effektive Nutzung von der Patientenakte mit einer großen Datenbank, auf die alle Praxen und Krankenhäuser Zugriff haben. Diese Daten steigen dabei jährlich stetig an.

Managementinformationssysteme, die Big Data Anwendungen nutzen, können Datenmengen analysieren und Abhängigkeiten zu finden, um eine personalisierte Behandlung zu ermöglichen. Hierbei gilt, je größer die Datenmenge, desto besser können die zur Verfügung stehenden Daten analysiert werden, um Ansätze zu finden.

In Deutschland hat die stufenweise Einführung von elektronischen Patientenakten (ePA) begonnen. Hiermit sind hohe Erwartung in der Telematik zu erwarten, um eine verbesserte Patientenversorgung zu bieten. Die Kernaufgabe der persönlichen elektronischen Patientenakte ist es, die vorliegenden Dokumente an verschiedenen Orten digital bereitstellen zu können. Hierbei zählen die Dokumente zu Behandlungen, Therapien, anamnestische Informationen und Befunde als relevante Dokumente, die digital zusammengeführt, verwaltet und für die darauffolgende Behandlung verfügbar sein sollen.

Die dreistufige Einführung der ePA hat am 01.01.2021 begonnen und soll bis zum 01.01.2023 gesetzlich geregelt werden [34].

Mithilfe der elektronischen Patientenakte wird eine verbesserte Patientenbehandlung sowie die Komfortabilität durch die Digitalisierung bei der Kommunikation zwischen verschiedenen Gesundheitseinrichtungen erwartet.

Messdatenanalyse in der Medizin

Die Behandlung von chronischen Erkrankungen und älteren Patienten stellt das Gesundheitssystem vor hohe Herausforderungen, die zu bewältigen sind. Solche Erkrankungen zu erkennen, analysieren und auszuwerten erfordert Langzeitdaten eines Patienten, um eine geeignete Behandlungsmethode zu finden. Hierbei eignet sich remote-basierte Pflege, die durch technologische Geräte an Daten vom Patienten verfügen können. Technologien wie die Telemedizin oder auch Wearables können Daten sammeln und diese an Ärzte weiterleiten [35]. Daten, die von Geräten wie Wearables, Smartphones oder Fitness-Tracker gesammelt werden, messen EKG-Werte, die Messung der Sauerstoffsättigung, Schrittzähler oder Schlafphasen der Nutzer. Bei eigenständiger Erkennung von schlechten Messwerten wird der Nutzer alarmiert und auf die schlechten Werte hingewiesen [36].

Big Data in der molekularen Diagnostik/Rheumatologie:

Seit Jahrhunderten wurde die Lehre der Rheumatologie weiterentwickelt. Die Behandlungsmöglichkeiten sind fortgeschritten und das große Fachwissen ist gestiegen.

Rheumatologie nutzt in Bezug auf Big Data folgende Techniken:

- Bildgebungsverfahren
- Funktionelle Genomik
- Post-Genomik [37]

Eine frühzeitige Behandlung von Rheumapatienten kann einen Stillstand der Krankheit führen, weswegen eine Vorbeugung der Krankheit in der Rheumatologie essenziell ist.

Die Wichtigkeit von Big Data Techniken in der molekularen Diagnostik zeigt die folgende Abbildung.

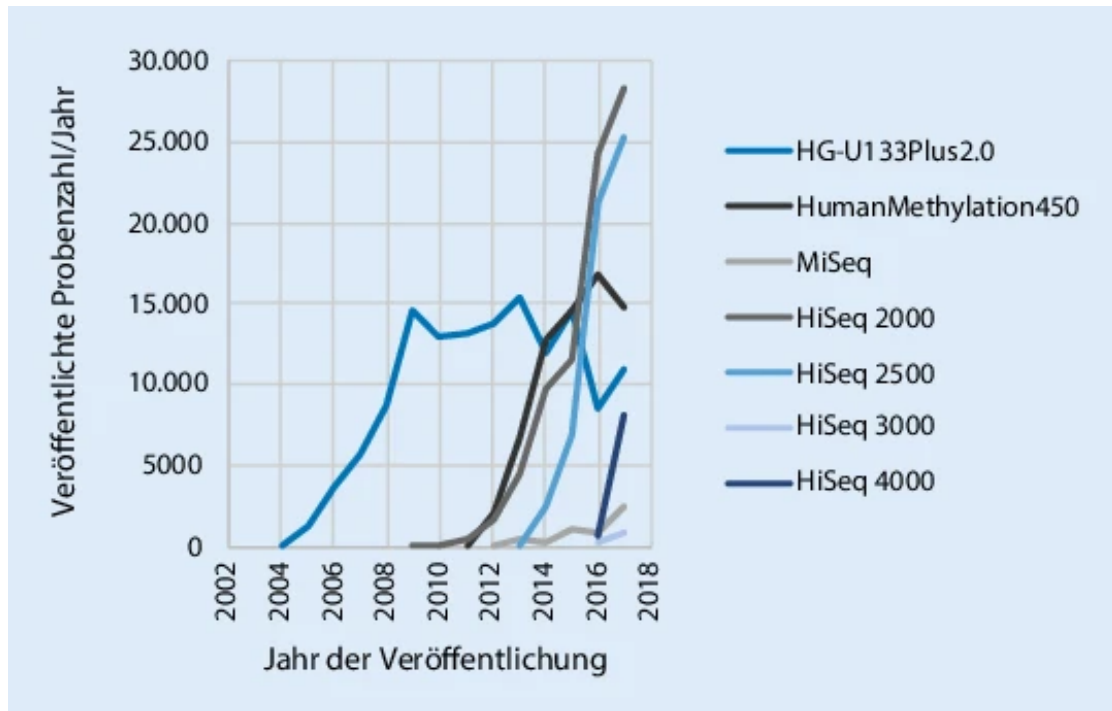


Abbildung 10: Anzahl zu Proben Daten zu Genexpressionsanlagen und Sequenzierungen [38]

In Abbildung 10 ist zu sehen, dass die Proben Daten für Sequenzierungen der DNA seit 2014 rapide gestiegen sind. Dies liegt einerseits daran, dass die Sequenzierung der DNA wichtiger in der Forschung wird, sowie die Möglichkeit einer Sequenzierung durch die gesunkenen Preise. Eine Sequenzierung der DNA ist außerhalb der Wissenschaft heutzutage ebenfalls durch die gesunkenen Preise möglich und wird genutzt.

3.3 Bibliometrische Analyse von Big Data

Der Begriff „Big Data“ existiert seit dem 20. Jahrhundert und wurde für große Datenmengen bezeichnet. Die heutige Bezeichnung von „Big Data“ kam nach der Jahrtausendwende. Durch die Bestimmung der ursprünglich 3 V's wurde der Begriff „Big Data“ genauer definiert. Eine große Bekanntheit erlangte Big Data als moderne Datenverwaltung ab den 2010er Jahren, zum Beispiel durch Investitionen der US-Regierung im Jahr 2012 in Big Data Technologien mit über 200 Millionen \$ [39].

In der nachfolgenden Abbildung wird eine Literatursuche über Google Scholar für „Big Data“ im Zeitraum von 2010 bis 2021 angewendet. Die Grafik stellt die jährlichen gefundenen Suchergebnisse von Erwähnung des Begriffes „Big Data“ im Titel dar. Die x-Achse zeigt das Jahr der Veröffentlichungen an und die y-Achse stellt die Anzahl der Suchergebnisse in Google Scholar dar. Wie zu erkennen, steigen die Titel, die den Begriff „Big Data“ beinhalten, ab dem Jahr 2012, das

über 2000 Suchergebnisse aufweist, schnell an. Vor 2012 sind die Treffer nahe null, beziehungsweise in einem niedrigen Bereich von unter 500 Treffern pro Jahr. Aufgrund der Tatsache, dass vor dem Jahr 2010 keine Daten zum Auswerten existieren, fangen die Grafiken ab dem Jahr 2010 an. Ein Höhepunkt der Grafik ist im Jahr 2018 zu erkennen. In dem Zeitraum sind circa 15.000 Ergebnisse zu finden. Nach 2018 sinken die Treffer jährlich und sind im Jahr 2021 mit 11.500 Suchergebnissen vorhanden. Es ist zu bemerken, dass „Big Data“ für einige Jahre ein Trendwachstum hatte und seit 2018 einen Abwärtstrend hat.

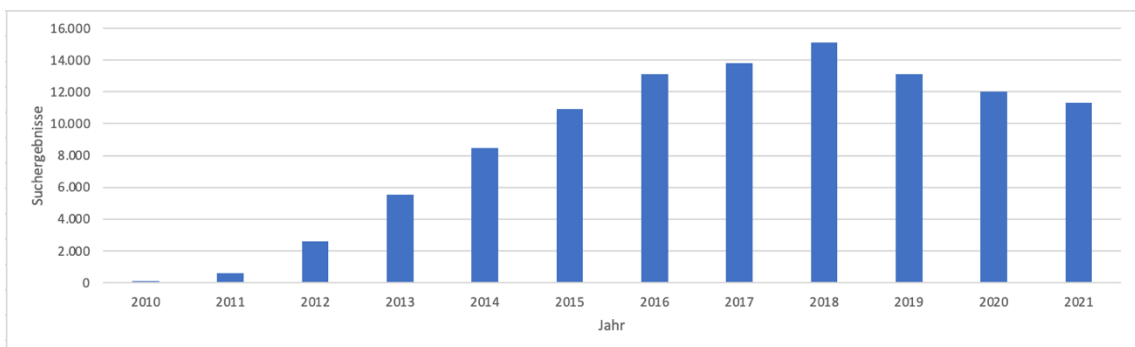


Abbildung 11: Anzahl der Publikationen pro Jahr von „Big Data“ im Titel

Der Verlauf in Abbildung 12 ist identisch wie von der vorherigen Abbildung. Bemerkenswerte Trendzuwächse sind ab 2012 mit circa 200.000 Ergebnissen zu erkennen. Von 2014 bis 2018 steigen die Suchergebnisse im Jahr um circa 50.000 Treffern jährlich. Wie bei der Titelsuche ist der höchste Peak bei der Textsuche im Jahr 2018 mit circa 240.000 Treffern. Nach 2019 gibt es einen höheren Abwärtstrend als in Abbildung 11. Während im Jahr 2019 noch der hohe Stand von 230.000 Treffern vorzuweisen war, sind es im Jahr 2020 noch 150.000 Treffer und im Jahr 2021 knapp 130.000 Treffer. Die Abwärtstrends in beiden Abbildungen können durch den kürzeren Zeitabstand zu heute passieren. In der Suchmaschinen Datenbank von Google Scholar können aktuellere Literaturen fehlen.

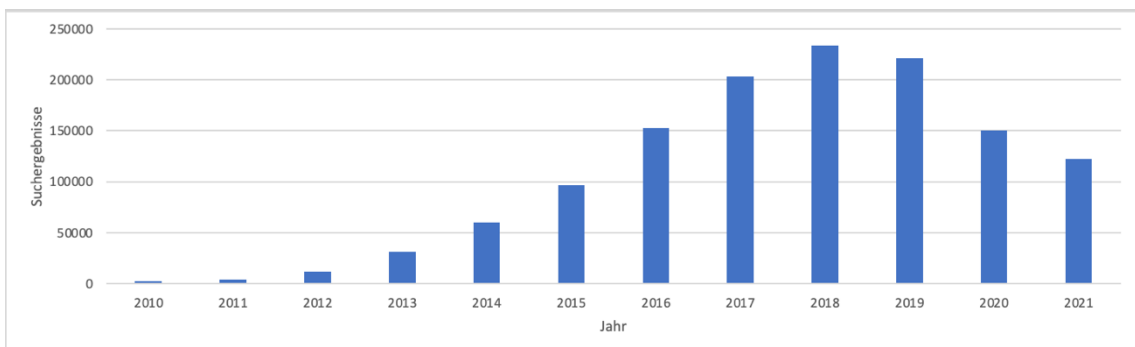


Abbildung 12: Anzahl der Publikationen pro Jahr von „Big Data“ im Text

3.4 Bibliometrische Analyse von Big Data in der Medizin

In der folgenden Abbildung werden „Big Data“ und das englische Wort „Medicine“ als Begriffe für die Titelsuche verwendet. Hierbei ist zu erkennen, dass in den Anfangsjahren 2010 und 2011 keine Werte vorhanden sind. Dies liegt daran, dass der Trend von Big Data, wie in den Abbildungen 11 und 12 zu erkennen ist, erst ab 2012 anfängt. Vorher wurde Big Data in Verbindung mit der Medizin eher gering in der Literatur erwähnt. Nachdem 2012 und 2013 in der englischen Literatur circa zehn Suchergebnisse nachzuweisen sind, steigt die Suchergebnisanzahl pro Jahr im Jahr 2013 auf ca. 25.000 Treffer an. Im Jahr 2014 sind circa 50 Suchergebnisse vorliegend, dieser verdoppelt sich in den folgenden Jahren bis 2018. Seitdem ist ein Abwärtstrend zu erkennen. Im letzten Jahr sind ungefähr 50 Suchergebnisse vorhanden.

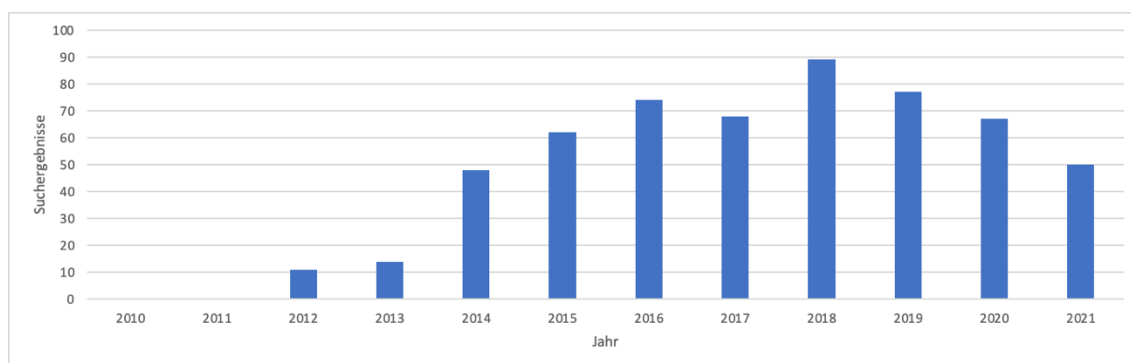


Abbildung 13: Anzahl der Publikationen pro Jahr von „Big Data“ und „Medicine“ im Titel

Bei der Textsuche von „Big Data“ und Medicine ist der stetig starke Anstieg pro Jahr zu erkennen. Während die ersten Jahre der Grafik von 2010 bis 2012 es ungefähr 2500 Treffer pro Jahr waren, steigt die Anzahl der Treffer ab 2013 um 4500 Ergebnisse jährlich an. Der Höchststand ist im Jahr 2020 mit über 40.000 Suchergebnissen aufzuweisen. Im darauffolgenden Jahr sind es ungefähr 35.000 Suchergebnisse.

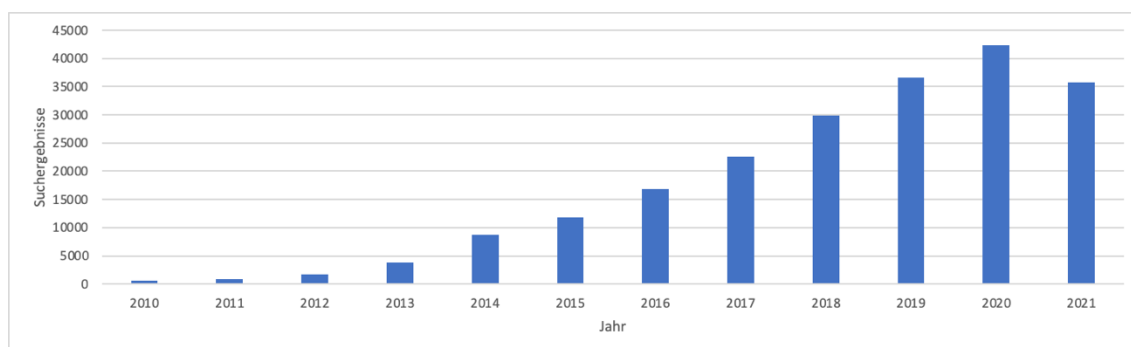


Abbildung 14: Anzahl der Publikationen pro Jahr von „Big Data“ und „Medicine“ im Text

Die folgenden Grafiken 15 & 16 zeigen die Suchergebnisse von „Big Data“ und dem deutschen Begriff „Medizin“ in der Titel- und Textsuche. Ein großer Unterschied zu den Suchergebnissen der englischen Begriffe sind die geringeren Treffer pro Jahr. Während bei der Titelsuche im Jahr 2016 ein Maximum von neun Ergebnissen vorliegen, liegen mit dem englischen Begriff „Medicine“ im Jahr 2016 ungefähr 75 Ergebnisse vor. Ebenfalls sind Schwankungen im Zeitraum von 2010 bis 2021 zu erkennen. In den Jahren 2016 und 2019 sind die Suchergebnisse pro Jahr am höchsten, jedoch sind die Suchergebnisse in den darauffolgenden Jahren 2017 und 2020 am niedrigsten.

Die Textsuche zeigt einen sehr hohen Anstieg im Jahr 2014 an. Im vorherigen Jahr waren ungefähr 200 Suchergebnisse vorhanden, während im Jahr 2014 das Siebenfache an Suchergebnissen mit circa 1500 Datensätzen vorliegt. Nach dem Rückgang der Suchergebnisse auf 500 im Jahr 2015 steigen die Suchergebnisse pro Jahr stetig weiter an. Im Jahr 2020 liegen circa 1900 Suchergebnisse vor, dass ebenfalls den Hochpunkt darstellt.

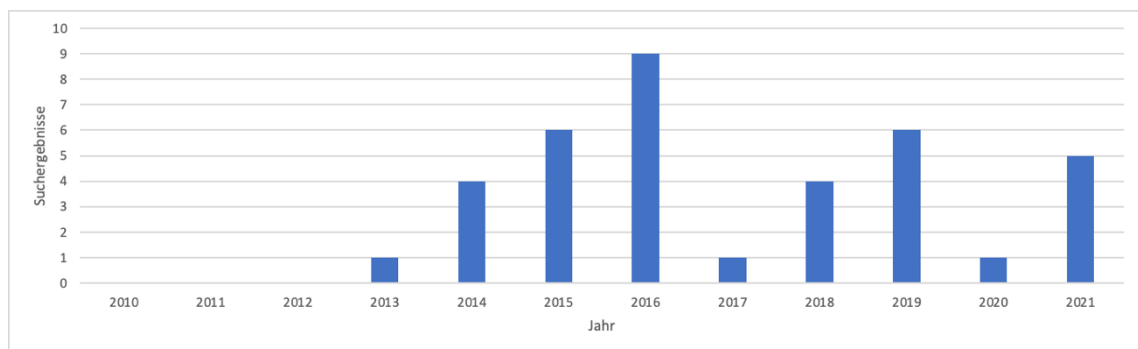


Abbildung 15: Anzahl der Publikationen pro Jahr von „Big Data“ und „Medizin“ im Titel

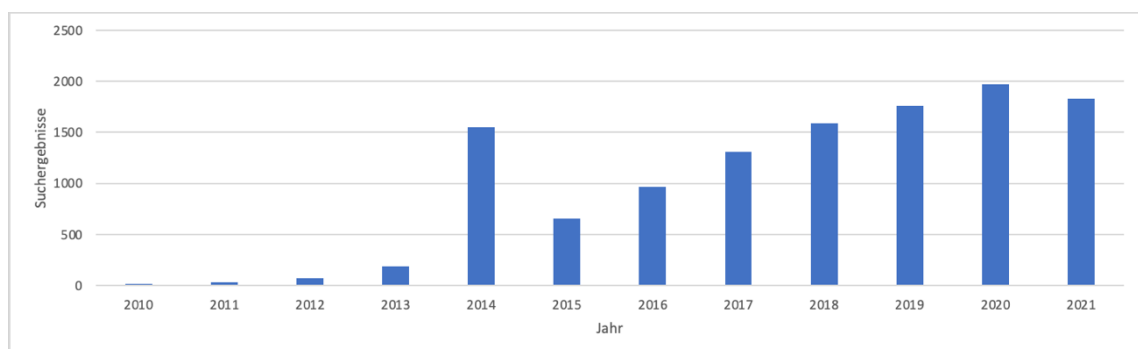


Abbildung 16: Anzahl der Publikationen pro Jahr von „Big Data“ und „Medizin“ im Text

3.5 Genomdaten

Das Genom ist ein Erbgut eines Lebewesens, das als Gesamtgruppe von allen Genen definiert wird. Es beinhaltet das Molekül DNA und Chromosomen, die ebenfalls wie die Gene weitervererbt werden können.

In der Genomik gibt es reichlich viele Erfolge in der Wissenschaft und die Genomdaten werden zum Teil auch im Praxisalltag verwendet.

3.5.1 DNA-Sequenzierung

Durch die DNA-Sequenzierung wurde seit 1995 von über 50.000 Organismen das Genom analysiert. Ein großer Meilenstein in der Geschichte der Sequenzierung ist die vollständige Entschlüsselung der menschlichen Gene im Mai 2021 mit 19.969 Genen. Dies erfolgte im Humangenomprojekt, das seit 1990 die Ziele verfolgt, die komplette Basensequenz eines Menschen zu analysieren, die Sequenz von drei Milliarden Basenpaaren zu entdecken, die Entwicklung der Datenanalyse und ethische, rechtliche und soziale Aspekte aufzugreifen [40].

In der folgenden Abbildung ist der Preis einer DNA-Sequenzierung pro Genom zu erkennen. Während der Preis um die Jahrtausendwende circa 100 Millionen Dollar betrug, ist der Preis seitdem rapide auf circa 1000 Dollar gesunken.

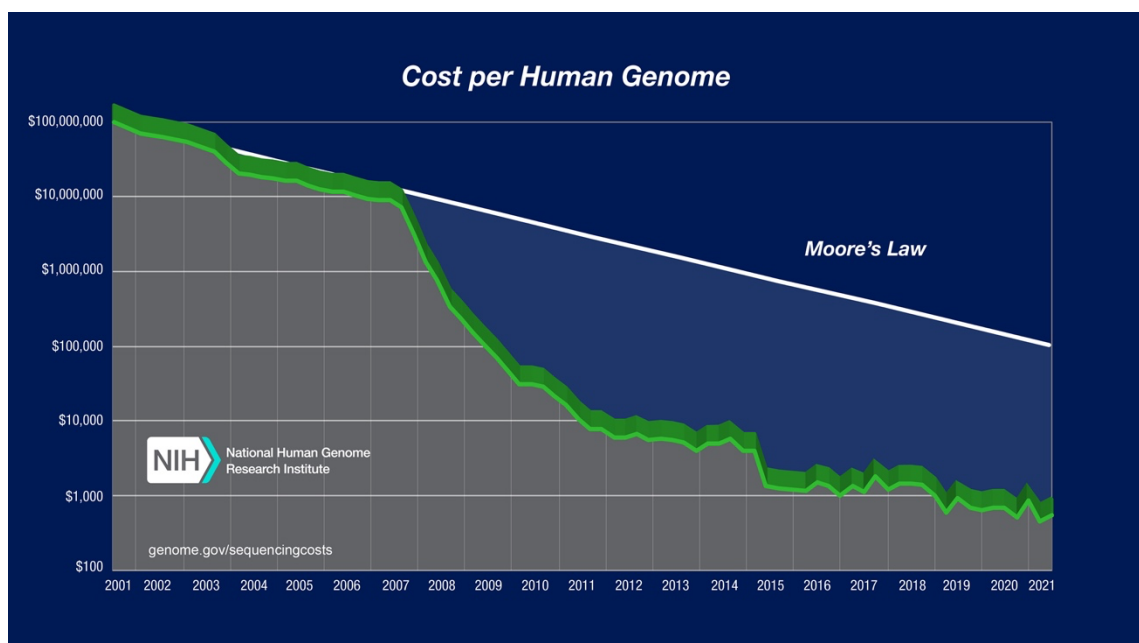


Abbildung 17: Preis pro Genomanalyse [41]

Die Gründe für den Preisfall liegen daran, dass die Methoden der DNA-Sequenzierung geändert wurden. Während früher die Genome in großen Fragmenten

analysiert wurden, werden heutzutage die Genome in kleinere Fragmente zur besseren Datenanalyse unterteilt [41].

Die Methoden der Sequenzierung werden in klassische und moderne Methoden (zweite und dritte Generation) unterschieden [42].

Zu den klassischen Methoden zählen:

- Sequenziermethode von Maxam und Gilbert
- Didesoxymethode

Zu den modernen Methoden gehören:

- Pyrosequenzierung
- Sequenzierung durch Hybridisierung
- Halbleitersequenzierung
- Sequenzierung mit Brückensynthese
- Zwei-Basen-Sequenzierung
- Sequenzierung mit gepaarten Enden
- Nanoporen-Sequenzierung

3.5.2 Bibliometrische Analyse

In Abbildung 18 sind die Begriffe „Big Data“ und „Gene“ als Suchbegriffe eingegeben worden. Der Anfangswert im Jahr 2010 fängt mit circa 1000 Suchergebnissen an. In den darauffolgenden Jahren steigen die Ergebnisse pro Jahr weiter an. Über 5000 Ergebnisse in einem Jahr sind erst im Jahr 2014 zu erfassen, die weiterhin jährlich weiter ansteigen. Es ist zu erkennen, dass pro Jahr ungefähr 3000 Datensätze dazukommen. 2017 sind bei der Literatursuche ungefähr 15.000 Daten zu erkennen. Die Suchergebnisse steigen bis in das Jahr 2021 auf circa 21.000 Treffer an. Das Wachstum hält bis zum Jahr 2021 weiter an und es ist keine Stagnation festzustellen.

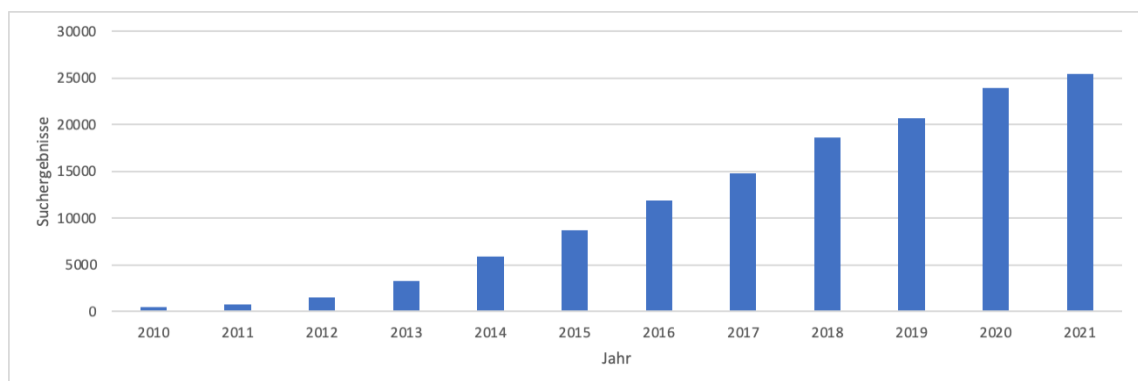


Abbildung 18: Anzahl der Publikationen pro Jahr von „Big Data“ und „Gene“ im Text

3.6 Radiologie

Der Fachbereich „Radiologie“ beinhaltet die Therapie mit der Hilfe von bildgebenden Verfahren und Systemen, um Krankheiten zu erkennen.

Die gängigsten Systeme in der Radiologie sind:

- Röntgenstrahlung
- MRT
- Ultraschall
- CT [43]

Zur Klärung der Krankheit wird in gebrauchtem Maße eine Radiologie durchgeführt. Die Hauptaufgabe der Radiologie ist die Aufnahme und Befundung von Bilddaten in den fachübergreifenden Themen [44].

Neben den diagnostischen Verfahren existieren noch Radiologieinformationssysteme, die früher für die Archivierung und den Transport der Röntgenaufnahmen zwischen den Praxen zuständig waren. Bildgebende Verfahren werden in kürzeren Zeitabständen neu dimensioniert. Das klinische Aufgabenspektrum breitet sich immer weiter aus. Eine stetige Weiterentwicklung und Implementierung von neuen Techniken verändern beziehungsweise erweitern die vorhandenen Indikatoren der Verfahren. Neue Konzepte unterstützen die Optimierung der medizinischen Prozesse und erreichen eine höhere Effizienz in Krankenhäusern [45].

3.6.1 Computertomographie mit iterativer Bildrekonstruktion

Die Computertomographie kann durch Techniken der künstlichen Intelligenz erweitert werden. KI-basierte Bildrekonstruktionen vom CT können mittels Teils iterativen Verfahrens eine reduzierte Strahlendosis und verbesserte Detailauflösungen erhalten. Mit den Algorithmen wird automatisiert das CT-Bild rekonstruiert. Durch mittelbasierte Rekonstruktionen ist eine höhere Bildqualität vorhanden, die eine Erkennung von Tumoren oder ähnliches erleichtert [46].

Auf der linken Seite von Abbildung 19 ist zu erkennen, wie eine Computertomographie ohne eine mittelbasierte Bildrekonstruktion aussieht. Hierbei ist zu beobachten, dass das Bild viele kleine weiße Flächen aufweist, die das Bild verfälschen können und somit eine gute Diagnose erschwert wird. In der mittleren Ebene ist die Fotografie von demselben CT zu erkennen. Hierbei sind die Störfächen zum Teil retuschiert worden. Auf der rechten Seite ist ein CT mit mittelbasierter Bildrekonstruktion, die das Bild schärfer als die vorherigen Fotos darstellt und somit eine bessere Beurteilung der Aufnahme ermöglicht.

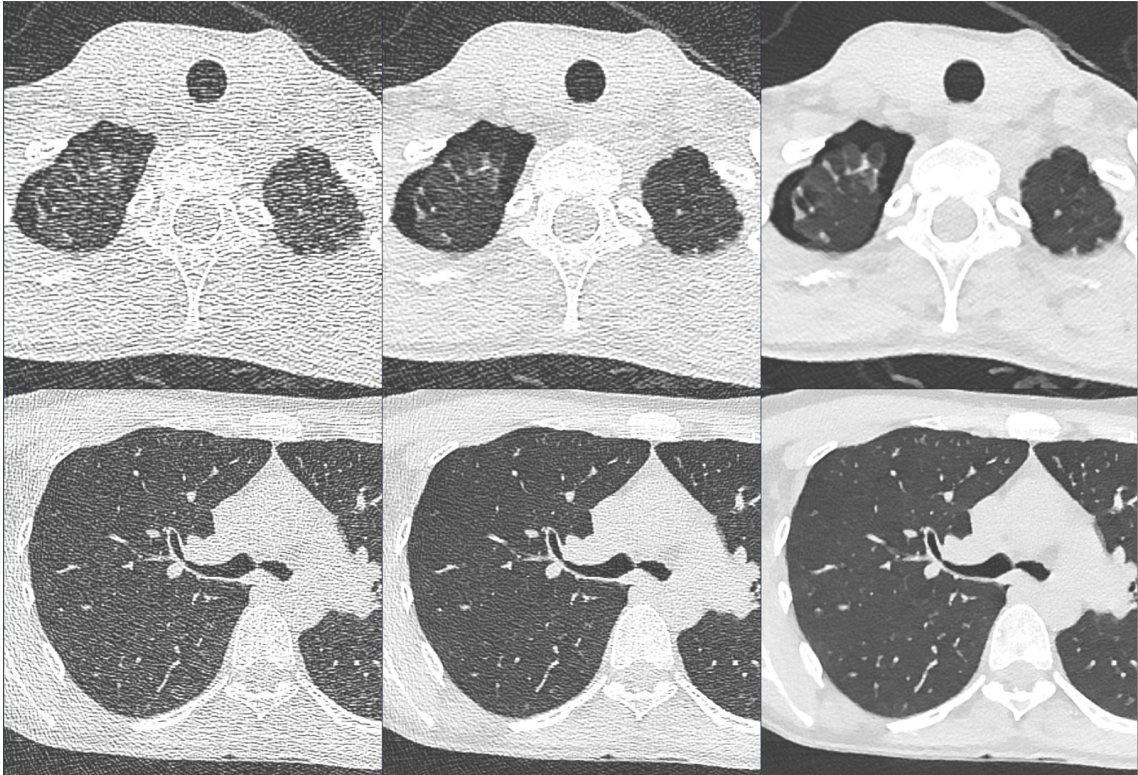


Abbildung 19: CT mit mittelbasierter Bildrekonstruktion (rechts) [47]

3.6.2 Bibliometrische Analyse

In der nachfolgenden Abbildung handelt es sich um eine Grafik, die die Literatursuche von „Big Data“ mit „Radiology“ umfasst. Von 2010 bis 2013 sind wenige Daten in Bezug von Big Data mit Radiology vorhanden. Mehr Erwähnungen der beiden Begriffe in der Literatur sind ab dem Jahr 2014 mit über 1000 Erwähnungen zu finden. Nach einer Abnahme der Ergebnisse im Jahr 2015 auf 900, steigen die jährlichen Treffer stetig bis zum Jahr 2018 auf 2500 Veröffentlichungen an. Mit 2500 Suchergebnissen haben sich die Erwähnungen von Radiology und Big Data in der Literatur innerhalb von vier Jahren um mehr als verdoppelt. Ab 2019 ist ein Wachstum zu erkennen. Während die Literaturdatenbank für Jahr 2019 ungefähr 3800 Ergebnisse herausgibt, sind es im Jahr 2020 über 5500 Ergebnisse. Ein Wachstum von über 2000 Ergebnissen innerhalb eines Jahres ist zu bemerken. Im Jahr 2021 sind über 7000 Suchergebnisse vorhanden.

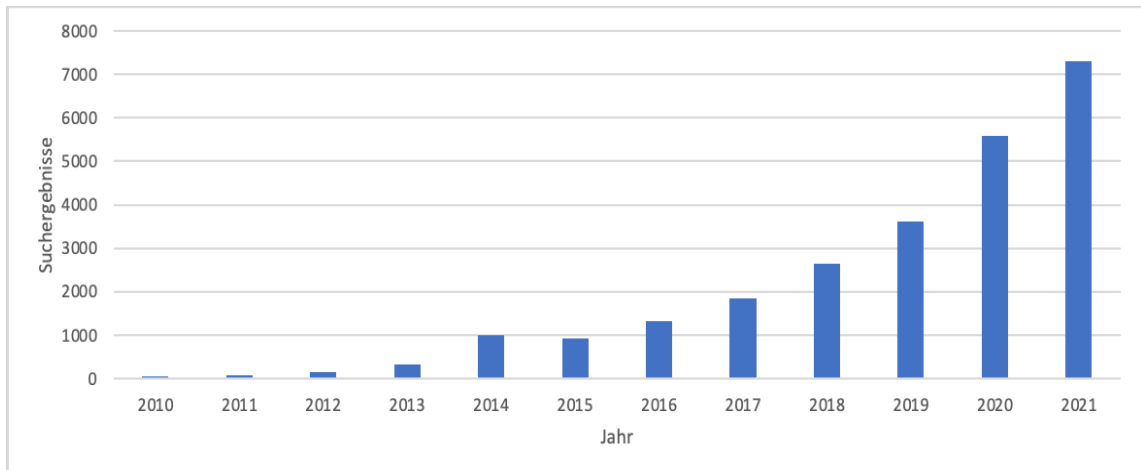


Abbildung 20: Anzahl der Publikationen pro Jahr von „Big Data“ und „Radiology“ im Text

3.7 Proteindaten

Proteindaten werden in einer Proteindatenbank gesammelt, in denen verschiedene Proteinsequenzen in den großen Datenbanken wie GenBank oder RefSeq gespeichert sind.

Proteinsequenzen dienen als Grundlage von Determinanten der biologischen Struktur und Funktion.

3.7.1 Krebstherapie mittels Proteinanalyse

Krebstherapie

Mittels einer Proteinanalyse ist zu verstehen, wie sich die Proteine im menschlichen Körper ändern, um gegen Krebs vorzubeugen. Beim Auftreten von Krebs ist es für Ärzte wichtig, zu betrachten, wie die Gene und die Proteine reagieren, um individuelle Krebstherapiemethoden zu empfehlen. Hierbei wird darauf geachtet, welche Gene aktiv sind und in RNA umgeschrieben werden, die für Proteine als Bauplan gelten [47].

Um genügend Informationen der fortgeschrittenen Behandlung zu erhalten, sind Daten in Echtzeit essenziell, um eine individuelle Behandlung auszuführen. In der Präzisionsonkologie werden Patienten mithilfe ihres eigenen Tumorprofils behandelt. Hierbei werden für die Behandlung vergangene Fallberichte als Muster der Behandlung genommen. Durch die Datenerfassung ist der Verlauf der positiven oder negativen Behandlung zu erkennen. Softwaresysteme können hierbei auf anonymisierte Daten zugreifen, um zum Beispiel die Behandlungsmethoden wie bei einer Strahlentherapie oder Medikamenteneinnahme entsprechend zu dosieren [48].

3.7.2 Bibliometrische Analyse

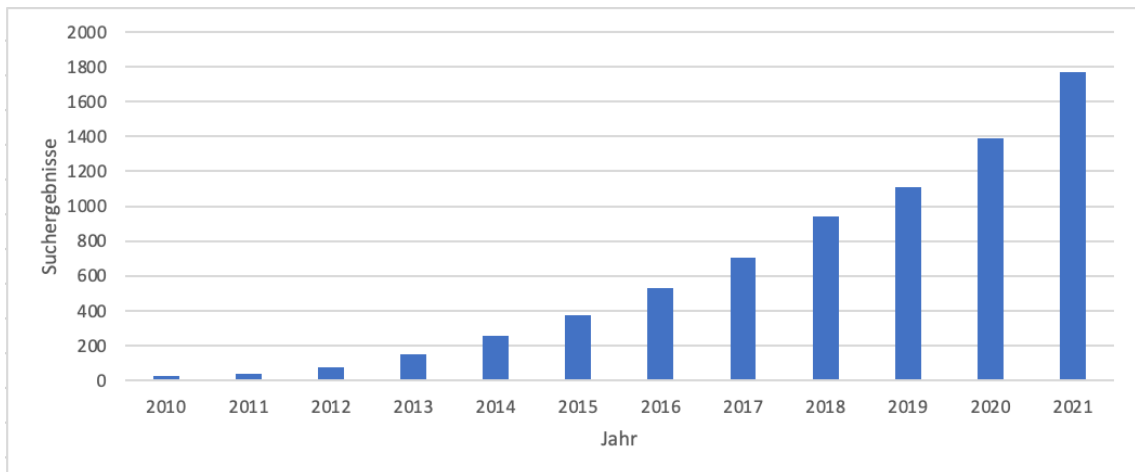


Abbildung 21: Anzahl der Publikationen pro Jahr von „Big Data“ und „Macromolecular“ im Text

Proteindaten gehören zu den Makromolekülen, weswegen „Macromolecular“ in der Literatursuche als Begriff mit „Big Data“ genommen wird. Hierbei ist zu erkennen, dass der Verlauf der Grafik kontinuierlich von 2010 bis 2021 ansteigt. Im Jahr 2010 sind wenige Literaturdaten in Google Scholar zu dem genannten Anwendungsgebiet vorhanden. Mindestens 100 Suchergebnisse sind zum ersten Mal im Jahr 2013 zu erkennen. Im darauffolgenden Jahr sind 200 Suchergebnisse wiedergegeben worden und für das folgende Jahr hat sich die Anzahl der Suche auf circa 400 verdoppelt. Nach 2016 steigt der Zuwachs an Suchergebnissen pro Jahr weiterhin bis zum Jahr 2019 leicht an. Ab 2019 sind mindestens 1000 Suchergebnisse im Jahr vorzufinden, die jährlich bis 2021 ungefähr um 400 Suchergebnisse jeweils ansteigen. Im Jahr 2021 sind knapp 1800 Suchergebnisse vorhanden.

3.8 Patientenakte

Die Patientenakte dient als Informationsaustausch zwischen Leistungserbringern und den Versicherten. Hierbei haben Ärzte sowie die Patienten Zugriff auf die gleichen Daten. Die Kontrolle sowie die Bewahrung und Löschung gewisser Daten kann einzig der Versicherte entscheiden.

In einer Patientenakte werden medizinische Dokumente meist als eine PDF-Datei gespeichert, mit den Dokumenteninformationen Dokumententyp, Dokumentengröße, Dokumententitel, Einstellzeit, Zugriffsberechtigungen sowie die Patienteninformationen PatientenID und die demographischen Daten.

3.8.1 Elektronische Patientenakte

Die elektronische Patientenakte wird seit dem 01.01.2021 in Deutschland von den Versicherungen ihren Versicherten angeboten.

Gesetzlich geregelt wurden die Grundlagen im Patientendaten-Schutz-Gesetz. Die Rechte an den Daten in der ePA besitzen allein nur die Versicherten (§ 341 ff. SGB V). Die ePA soll die Primärdokumentation in den Datenbanken der Krankenhäuser und Arztpraxen nicht ersetzen.

Die vollständige gesetzliche Einführung vom ePA, um alle Daten aufzunehmen, folgt in drei Stufen

Stufe 1

Ab 01. Januar 2021

In der ersten Stufe werden medizinische Informationen von Versicherten ausgestellt. Die Informationen können für einrichtungs-, fach- und sektorübergreifende Stellen genutzt werden.

Zur Verfügung stehen:

- Daten zu Diagnosen, Befunden, durchgeführte und geplante Therapiemaßnahmen
- Früherkennungsuntersuchungen
- Behandlungsberichte
- Sonstige medizinische untersuchungs- und behandlungsbezogene Informationen
- Elektronischer Medikationsplan
- Notfalldaten
- Elektronische Arztbriefe
- Gesundheitsdaten, die vom Versicherten freigestellt werden

Stufe 2

Ab 01. Januar 2022 wird die ePA ergänzt um

- Zahn-Bonusheft
- Mutterpass und Versorgung mit Hebammenhilfe
- Elektronische Untersuchungsheft für Kinder
- Impfpass
- Daten über die in Anspruch genommenen Leistungen von Krankenkassen

Ab 01. März 22 wird die ePA ergänzt mit der elektronischen Gesundheitsakte von Krankenkassen

Stufe 3

Ab 01. Januar 2023 wird die ePA ergänzt um

- Digitale Gesundheitsanwendungen (DiGA)
- Elektronische Rezepte
- Arzneimittelhistorie
- Pflegerische Versorgung
- Elektronische Arbeitsunfähigkeitsbescheinigungen (eAU) [49]

3.8.2 Bibliometrische Analyse

In der nachfolgenden Abbildung wird „Big Data“ mit dem Thema „Electronic Health Record“ (in Deutschland als elektronische Patientenakte bekannt) in Verbindung gebracht und darauf bauend eine bibliometrische Analyse erstellt. In den ersten beiden Jahren 2010 und 2011 sind keine bis wenige Datensätze für dieses Anwendungsgebiet in der Literatur. Im Jahr 2013 sind ungefähr 250 Suchergebnisse vorhanden, die nach einem Jahr auf über 600 Literaturdaten erhöht wurden. Im Jahr 2015 gab es ungefähr 1000 Literaturveröffentlichungen in Bezug zu „Electronic Health Record“ mit „Big Data“. Seitdem steigen die Literaturveröffentlichungen pro Jahr stark an. Von 2015 bis 2018 ist ein Zuwachs von jährlichen Veröffentlichungen als Gesamtzahl mit circa 500 Treffern zu nennen. Ab dem Jahr 2018 steigen die Literaturveröffentlichungen pro Jahr um ungefähr 1000 Veröffentlichungen mehr an. Im Jahr 2020 sind ungefähr 3700 Ergebnisse zu finden und im darauffolgenden Jahr sind es über 4600 Suchergebnisse.

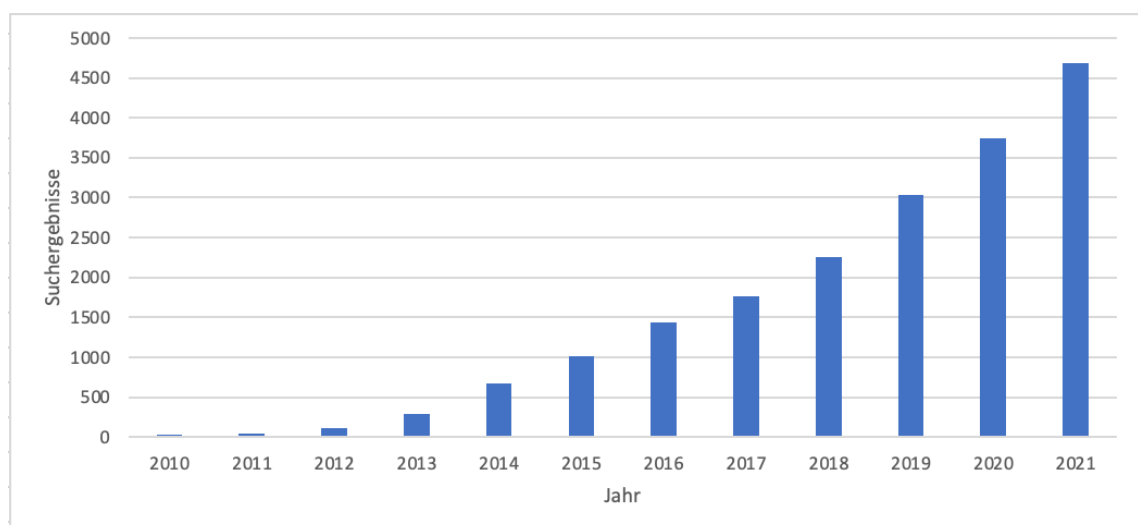


Abbildung 22: Anzahl der Publikationen pro Jahr von „Big Data“ und „Electronic Health Record“ im Text

3.9 Messdatenanalyse in der Medizin

Daten fallen in jeder medizinischen Behandlung an. Diese Daten dienen für die Krankheitsgeschichte sowie auch für die große Datenanalyse mit immer größer werdenden Daten.

Unter den Begriff medizinische Daten fallen Daten aus drei Kategorien.

- Daten aus der Bioinformatik
- Medizinische Daten aus Krankenhäusern, Arztpraxen und Versicherungen
- Lifestyle-Daten wie Sport-, Essen- und Schlafgewohnheiten

Die Verknüpfung von Daten aus den genannten Kategorien bietet entsprechende Vorteile in der Gesundheitsversorgung. Durch die verfügbaren Daten ist eine schnellere und sichere Diagnose mit individuellen Behandlungsmethoden möglich [50].

3.9.1 Alltagsgerät: Wearables

Die Verwaltung großer Datenmengen hat bei der Entscheidungsfindung eine große Rolle. Hierzu zählen auch die Wearable-Geräte hinzu, die immer mehr Daten sammeln und weitergeben. Es entsteht der Bedarf von Speicher-, Verarbeitungs- und Analysetools. Wearables werden in verschiedenen Bereichen wie Fitness, Gesundheit und Unterhaltung eingesetzt [51].

Bei chronischen Erkrankungen sowie auch bei plötzlichen Erkrankungen sind Alltagsgeräte wie eine Smartwatch für die Behandlung des Patienten hilfreich. Durch die 24-stündige Überwachung des Patienten ist es möglich, die Blutwerte, körperliche Aktivitäten, Pulsmessung, Blutsauerstoff oder Ähnliches durch die Datensammlung des Wearables zu kontrollieren und auszuwerten [52].

3.9.2 Bibliometrische Analyse

Wie in der Abbildung 23 zu erkennen, existiert im Jahr 2010 keine Veröffentlichungen in Bezug auf Messdatenanalysen in der Medizin (Medical Measurement) mit „Big Data“. Ein kontinuierliches Wachstum der Veröffentlichungen pro Jahr ist über den gesamten Beobachtungszeitraum zu erkennen. Von 2011 bis 2013 sind Daten mit unter 500 Veröffentlichungen pro Jahr zu finden. Im Jahr 2014 wurden circa 1000 schriftliche Texte in Bezug zu Medical Measurement und Big Data veröffentlicht. Seitdem steigen die jährlichen Veröffentlichungen pro Jahr stetig an. Für das Jahr 2019 sind über 5000 Suchergebnisse in der Literaturlatenbank Google Scholar vorhanden. Über 6300 Suchergebnisse sind im Jahr 2020 zu finden und im darauffolgenden Jahr gibt es einen Zuwachs von über

2000 Veröffentlichungen pro Jahr, mit insgesamt ungefähr 8300 Suchergebnissen für das Jahr 2021.

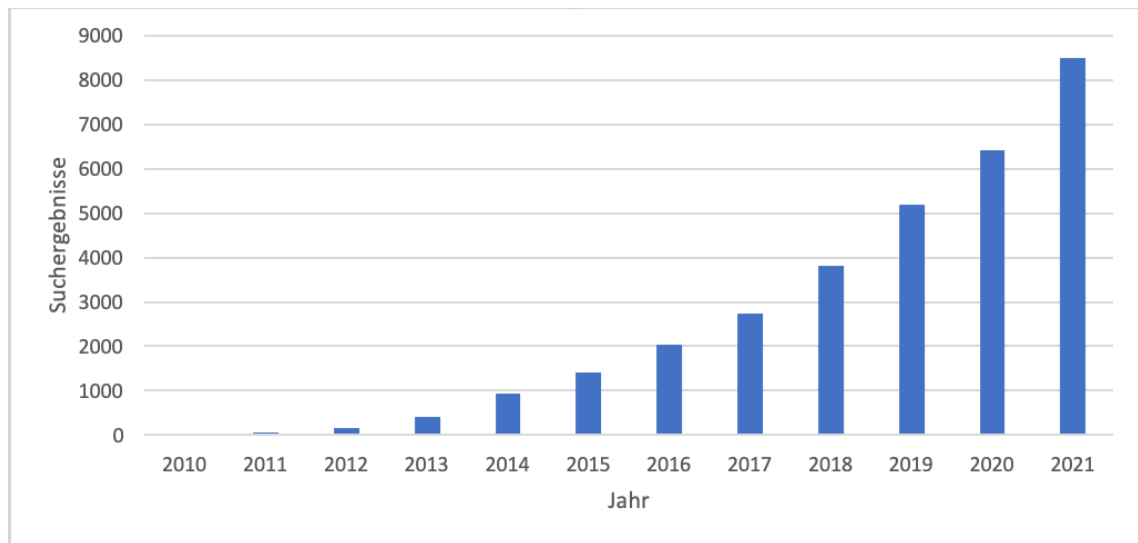


Abbildung 23: Anzahl der Publikationen pro Jahr von „Big Data“ und „Medical Measurement“ im Text

3.10 Big Data in der molekularen Diagnostik/Rheumatologie

In der molekularen Diagnostik werden genetische Änderungen im Erbgut nachgewiesen, die krankheitsrelevant sein können. Eine Änderung der Erbanlagen kann zu einem Tumor führen, weswegen ein Nachweis der Genveränderungen für die weitere Therapiebehandlung relevant ist. Durch Beeinflussung der genetischen Veränderungen wird in der Praxis versucht, den Tumor zu stoppen und die Behandlung erfolgreich abzuschließen [53]. In der Rheumatologie werden zum Teil Entzündungen im Gelenkbereich behandelt.

3.10.1 Hochdurchsatztechnologie

Für die Datenanalyse in der molekularen Diagnostik entstehen große und komplexe Datenmengen, die KI-gestützte Methoden benötigt, um den Umfang, Strukturvielfalt sowie die Geschwindigkeit der Anzahl an Daten zu bewältigen. Dies kann durch die Hochdurchsatztechnologie, das gängigste Verfahren in der molekularen Diagnostik, bewältigt werden.

Molekulare Daten beinhalten eine von der Messtechnik vorgegebene Struktur, mit der die Daten ausgewertet werden. Durch die Datenanalysen in der Hochdurchsatztechnologie können die biologischen Funktionen der molekularen Daten interpretiert und ausgewertet werden.

Durch das Verstehen der molekularen Daten soll der Erfolg der Therapiebehandlungen in der Rheumatologie wie zum Beispiel bei entzündungsbedingten

Gelenken besser erfolgen können. Aktuell werden Methoden wie CRP und Blut-senkung bei Entzündungen angewendet, die jedoch keine individuellen Aussagen über den Krankheitsverlauf des einzelnen Patienten berichten können

Im Praxisalltag der Rheumatologie werden Big Data Methoden noch nicht angewendet, diese Methodiken sind noch größtenteils in der Forschung [38].

Die Abbildung 24 zeigt die Unterschiede der „One fits all“ Medizin mit der personalisierten Medizin, die unter anderem die Hochdurchsatztechnologie in der molekularen Diagnostik oder auch die Biomarker-Analyse anwenden.

Hierbei ist zu erkennen, dass in der „One fits all“ Medizin jede Person die gleiche Behandlung beziehungsweise die gleichen Medikamente für die Behandlung erhält. Es sind keine individuellen Unterschiede zu erkennen. In der x-Achse werden die Nebenwirkungen und die Wirksamkeit der Behandlung und der Medikamente dargestellt. Die y-Achse zeigt die Häufigkeit der Nebenwirkungen sowie die Wirksamkeit. Es ist festzustellen, dass Nebenwirkungen vermehrt auftreten. Ebenfalls kann es aber bei der gleichen Häufigkeit gute Wirkungen (Responder) oder auch keine bis schlechte Wirksamkeit (Non-Responder) zeigen. Die Häufigkeiten sind gleichmäßig aufgeteilt.

Die personalisierte Medizin hat eine individuelle Behandlung der Patienten. In der Grafik ist zu erkennen, dass wenige Nebenwirkungen auftreten können. Die Wahrscheinlichkeit der guten Wirksamkeit der Behandlung ist jedoch sehr hoch, während die Behandlung in selteneren Fällen nicht wirkt.

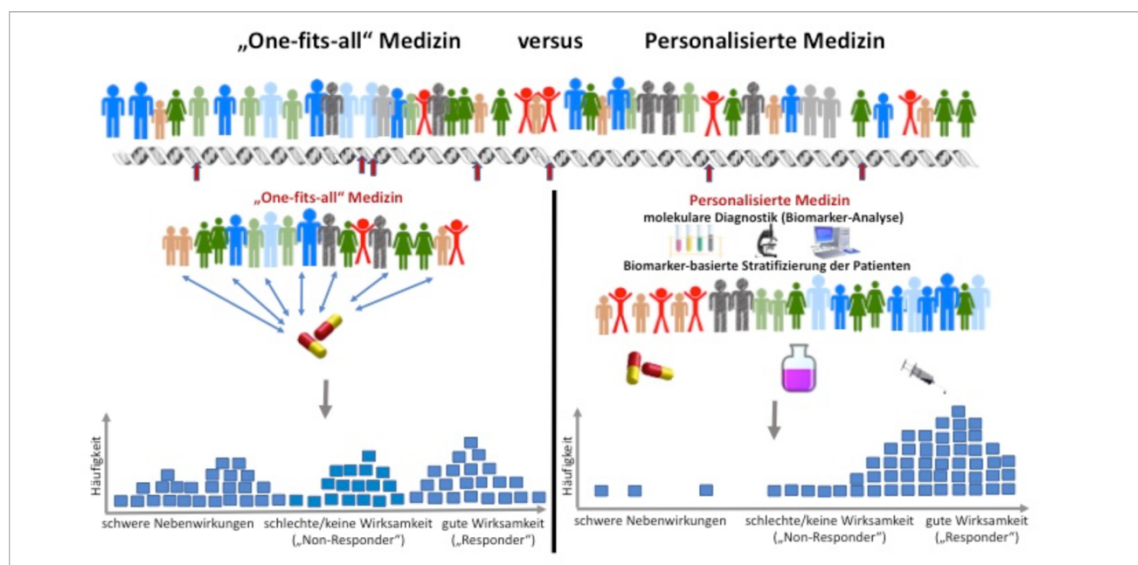


Abbildung 24: Effektivität der allgemeinen „One-Fits-All“ Medizin mit der personalisierten Medizin

3.10.2 Bibliometrische Analyse

Bei der Abbildung 25 wurden die Begriffe „Big Data“ und „Rheumatology“ als Textsuche in Google Scholar gesucht. Hierbei wurde der Zeitraum von 2010 bis 2021 ausgewählt. Von 2010 bis 2012 steigen die Veröffentlichungen pro Jahr nicht an, diese sind konstant bei ungefähr 20 Veröffentlichungen pro Jahr. Nachdem es einen leichten Anstieg im Jahr 2013 mit circa 70 Veröffentlichungen existieren, kam es zu einem hohen Wachstum im darauffolgenden Jahr. Über 570 Lektüren wurden innerhalb des Jahres 2014 in englischer Sprache veröffentlicht. Seit 2015 ist ein kontinuierlicher Anstieg von Literaturveröffentlichungen pro Jahr zu kennzeichnen. Für das Jahr 2015 werden 200 Suchergebnisse in Google Scholar angegeben. Dies verdoppelt sich fast im Jahr 2016 und im Jahr 2017 werden 500 Treffer angegeben. Die meisten Veröffentlichungen folgten in den Jahren 2020 und 2021 mit circa 1100 neu veröffentlichten Schriften pro Jahr.

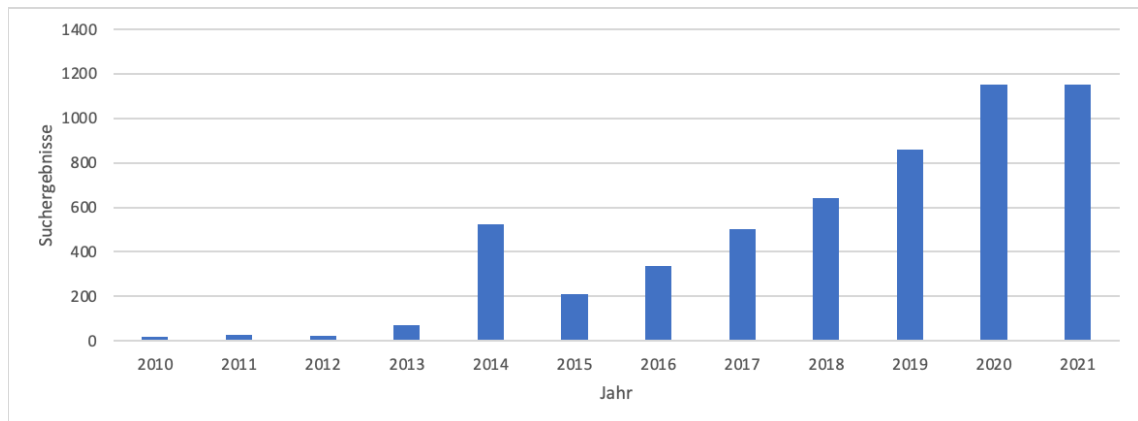


Abbildung 25: Anzahl der Publikationen pro Jahr von „Big Data“ und „Rheumatology“ im Text

4 Diskussion

Im folgenden Kapitel werden die Ergebnisse, die im vorherigen Kapitel dargestellt worden sind, erörtert und erläutert. Es werden die Ergebnisse der bibliometrischen Analysen ausgewertet. Hierbei werden ebenfalls positive sowie negative Kritikpunkte bei der Anwendung der bibliometrischen Analyse genannt.

Die Daten, die im Rahmen dieser Abschlussarbeit verwendet wurden, stammen von der Literaturdatenbank Google Scholar. Es ist anzumerken, dass Google Scholar als Einstieg respektive im Rahmen einer Abschlussarbeit für den Bachelor ausreichend ist. Jedoch existieren kostenpflichtige Literaturdatenbanken, welche für eine größere wissenschaftliche Arbeit geeignet sind. Jedoch ist die verwendete Literaturdatenbank eine der größten existierenden Literaturdatenbanken weltweit, weswegen die Nutzung von Google Scholar sinnvoll ist und diese Analysen dennoch aussagekräftig genug sind.

Der Zeitraum von 2010 bis 2021 wurde gewählt, weil Big Data in der heutigen Definition vor 2010 nicht verwendet wurde. Ebenfalls sind nicht genug Veröffentlichungen in den vorherigen Jahren vorhanden, weswegen eine Auswertung erschwert werden würde.

Für die Suche wurden englische Begriffe in der Literaturdatenbank Google Scholar verwendet, da hierbei ein größerer Umfang an Veröffentlichungen als in der deutschen Sprache existieren. Der Unterschied von der Anzahl der Veröffentlichungen an deutscher und englischer Literatur ist in den Abbildungen 13 und 15 zu erkennen. Während die englische Literatur bis zu 100 Veröffentlichungen im Jahr beinhaltet und ein konstanter Anstieg über die Jahre vorzuweisen ist, gibt es in der deutschen Literatur bis zu zehn Veröffentlichungen pro Jahr, mit einem Maximalpunkt im Jahr 2014, mit einem abschließenden Abstieg der darauffolgenden Jahre. Die deutsche Version ist aufgrund der mangelnden Daten nicht auswertbar. Dies liegt daran, dass die deutsche Sprache nicht so viele Sprecher wie die englische Sprache hat und somit weniger Veröffentlichungen existieren. Ebenfalls werden wissenschaftliche Arbeiten meistens in englischer Sprache verfasst und veröffentlicht.

Die Anwendungsgebiete wurde an den Kriterien, wie die Relevanz des Anwendungsgebietes in der Medizin, Anzahl der Publikationen, aktuelle Situation und Zukunftsaussichten gewählt. Alle Anwendungsgebiete hatten genug Publikationen um eine Auswertung in einem Zeitraum von mehr als 10 Jahren zu ermöglichen. Anwendungsgebiete wie Genomdaten, Messdatenanalyse und elektronische Patientenakten haben auch einen aktuelleren Bezug in der Praxis, die jedoch auch wie die anderen Anwendungsgebiete weiterhin in der Zukunft erweitert

werden können, um mehr Vorteile zu schaffen. Viele Forschungen sind bereits fortgeschritten und können in der Zukunft auch in der Praxis angewendet werden.

In den Abbildungen 11 und 12 handelt es sich um die Erwähnung von „Big Data“ in der Literatur, die jeweils im Titel beziehungsweise im Text vorzufinden sind. Es ist zu erkennen, dass diese Abbildungen die meisten Veröffentlichungen pro Jahr beinhalten als in den nachfolgenden Abbildungen. Dies liegt daran, dass der Begriff „Big Data“ sehr vielfältig in verschiedenen Bereichen benutzt wird. „Big Data“ im Allgemeinen wird hauptsächlich nicht nur mit der Medizin in Verbindung gebracht. Die Definition von Big Data wurde im Laufe des letzten Jahrzehnts genauer definiert. Ansonsten wurde dies auch als ein Synonym für konventionelle großen Datenbanken erwähnt. Der Begriff „Big Data“ in konventionellem Sinn, wurde erstmals in den 1990er Jahren von John Mashey, einem Chief Scientist von Silicon Graphics, einem ehemaligen Computerhersteller, erwähnt [54]. Die Definition hat sich seitdem geändert. Es ist in den meisten Abbildungen ersichtlich, dass die Anzahl der jährlichen Veröffentlichungen von 2010 bis 2018 stets gestiegen ist, jedoch ab 2018 bis 2021 gesunken sind. Dies liegt daran, dass nicht in der Realität weniger Veröffentlichungen mit Big Data seit 2018 herrschen und somit ein Abwärtstrend zu erkennen ist, sondern, dass die Daten von 2019 bis 2021 sich eher in den aktuelleren Zeiträumen befinden und somit nicht alle Veröffentlichungen der letzten drei Jahre in der digitalisierten Literaturdatenbank von Google vorzufinden sind. Ein Abwärtstrend ist somit nicht zu bemerken. Diese Betrachtung des Grafikverlaufes ist in der Abbildung 13 und Abbildung 14 ebenfalls zu erkennen. Ebenso ist zu beachten, dass COVID-19 ende 2019/Anfang 2020 aufgetreten ist und der Hauptfokus in der Pandemie liegen kann. Literarische Veröffentlichungen können während einer Pandemie weniger auftreten, weswegen in den Jahren 2020 und 2021 weniger schriftliche Verfassungen publiziert wurden als üblich.

Es ist in allen Abbildungen zu erkennen, dass die Popularität von Big Data sowie auch in Verbindung mit der Medizin und ihren Anwendungsgebieten ab 2012 stärker zunimmt. Seitdem gibt es einen Zuwachs bei den Erwähnungen von Big Data in der Medizin, weswegen Big Data ab dem Jahr 2012 stark an Popularität gewonnen hat, kann viele Gründe haben. Es kann unter anderem davon ausgegangen werden, dass Big Data an Popularität gewann, weil um das Jahr 2011 herum Smartphones sehr beliebt wurden. Durch Aktivitäten mit Smartphones und Social Media Plattformen fließen große Mengen an Daten auf der Welt. Die Menge an Datenverkehr war bis zum Jahr 2010 im privaten und geschäftlichen Umfeld nahezu gleich groß, jedoch zeigt die Abbildung 3, dass der private Datenverkehr seit dem Jahr 2014 am Steigen ist, während der geschäftliche Datenverkehr nicht signifikant steigt. Es ist auch in Abbildung 4 zu erkennen, dass ab

dem Jahr 2012 der Datenverkehr über Mobilfunk enorm angestiegen ist. Der Anstieg des privaten Datenverkehrs und die Popularität von Big Data ist ähnlich hoch. Die Wichtigkeit von Big Data in der heutigen Zeit hängt unter anderem mit den Massendaten zusammen, die jederzeit im Datenverkehr liegen. Das Streaming im privaten und geschäftlichen Nutzen sowie die künstliche Intelligenz, um Datensätze zu analysieren, wurden im letzten Jahrzehnt ebenfalls immer wichtiger.

In den Grafiken der Anwendungsgebiete ist der Verlauf nahezu identisch. Es gibt bis zum Jahr 2021 einen stetigen Anstieg der jährlichen Veröffentlichungen. Der Trendzuwachs ist hierbei zu erkennen, jedoch ist zu beachten, dass bei einigen Anwendungsgebieten wie zum Beispiel in der Abbildung 21 („Big Data und „Macromolecular“), weniger Datensätze vorhanden sind als in anderen Anwendungsgebieten. Dies liegt daran, dass einige Fachbereiche in Bezug zu Big Data fortgeschrittener in der Forschung sind beziehungsweise diese sogar in der Praxis Anwendung finden.

Eine Titelsuche bei den Anwendungsgebieten ist aufgrund mangelnder Veröffentlichungen nicht möglich. Eine Analyse mit zu wenigen Daten enthält keine Aussagekraft, weswegen in den Anwendungsgebieten allein die Textsuche angewendet wurde und in den allgemeineren Themenbereichen (Abbildung 11 bis 16) auch die Titelsuche beinhaltet ist.

Ein Kritikpunkt an der Literaturdatenbank Google Scholar ist, dass diese nicht unterscheidet, ob die vorgeschlagenen Veröffentlichungen als Suchergebnisse auch wissenschaftliche Arbeiten sind. Dies kann die Auswertung fälschen, falls Literaturen vorgeschlagen werden, die für die Analyse nicht vorgesehen sind. Ebenfalls werden Zitierungen von wissenschaftlichen Arbeiten, die in der Datenbank vermerkt sind, als Suchergebnisse wiedergegeben. Es ist zu beachten, dass bei der Recherche nur nach den Begriffen im Text gesucht wird. Ein Zusammenhang der gesuchten Begriffe mit der eigentlichen Arbeit muss nicht bestehen. Eine kurze Erwähnung in einem Abschnitt reicht für die Literaturdatenbank aus, um die Literatur als entsprechendes Suchergebnis wiederzugeben.

Big Data in der Medizin befindet sich zurzeit in vielen Fachbereichen in der Forschung und Entwicklung, jedoch gibt es hier positive Beispiele, die einen Mehrwert im Praxisalltag darstellen. Abschließend kann gesagt werden, dass Big Data in der Wissenschaft aber ebenfalls in der Praxis, sowie in den genannten Anwendungsgebieten seit 2010 weiterhin populärer sein wird und eine Trendabnahme mittelfristig nicht zu erwarten wird. Es ist zu erwarten, dass der Nutzen von Big Data steigen wird und dies viele Vorteile im Gesundheitswesen bringen wird.

5 Zusammenfassung und Ausblick

Die vorliegende wissenschaftliche Arbeit befasst sich mit der Hauptfrage, ob ein Trend in Bezug zu Big Data in der Medizin vorliegt. Hierfür wurde die Methode der bibliometrischen Analyse angewendet, um die Popularität von Big Data in der Literatur zu identifizieren. Dazu wurde in der Einleitung Big Data im allgemeinen Sinne sowie auch die Anwendung von Big Data in der Medizin beschrieben. Ebenso wurden theoretische sowie Beispiele aus der Praxis in der Gesundheitsbranche dargestellt. Im Anschluss wurden die Veröffentlichungen pro Jahr von Big Data mit Bezug zur Medizin und zu den verschiedenen Anwendungsgebieten (Genomdaten, Proteindaten, Messdatenanalyse in der Medizin, elektronische Patientenakte, Radiologie und Rheumatologie) dargestellt. Die Frage, ob ein Trendzuwachs von Big Data mit den verschiedenen Themenbereichen vorliegt, konnte somit aufgezeigt werden. Ein Trendzuwachs liegt seit Jahren vor und eine Stagnation ist zurzeit nicht zu erwarten. Die Veröffentlichungen pro Jahr sind in den verschiedenen Bereichen im Tausenderbereich und steigen weiterhin an. Ebenfalls wurden Beispiele aus der Forschung beziehungsweise aus der Praxis in den Anwendungsgebieten beschrieben und dargestellt. Es ist jedoch auch anzumerken, dass Big Data Technologien, obwohl einige bereits für die Anwendung in Krankenhäusern existieren, nicht ausreichend benutzt werden. Wie in Abbildung 6 zu erkennen ist, gibt es in den wenigsten europäischen Staaten Ärzte, die Big Data verwenden. Big Data Technologien werden laut der Umfrage bei maximal 10 % der Ärzte angewendet. Dies kann daran liegen, dass die Akzeptanz von neuen Technologien wie zum Beispiel Wearables als medizinische Hilfsttechnologien bei den Ärzten nicht vorhanden ist. Ärzte setzen bei der Behandlung eher auf Erfahrung als auf neue Methodiken. Jedoch ist die Nutzung von Big Data Technologien eine große Hilfe in der Medizin, da Datenbanken mehr Informationen aufnehmen und speichern können als Menschen. Ärzte können in naher Zukunft das Wissen von großen Datenbanken nutzen, um eine bessere Behandlung zu ermöglichen. Ebenfalls können unerfahrenere Ärzte bei scheinbar unlösbaren Problemen sich auf die Technik verlassen. Es ist davon auszugehen, dass Big Data in der Medizin in vielen Fachbereichen noch in der Entwicklungsstufe ist und dies intensiv erforscht wird. Mittelfristig wird die Hilfe von Big Data Technologien die Gesundheitsbranche weiterentwickeln und viele Prozesse ändern. Eine Entlastung der Fachkräfte in der Medizin wie Ärzte oder Krankenpfleger durch zum Beispiel automatisierte Auswertungen oder automatisierte Routineplanung ist zu erwarten. Trotz vieler Vorteile von neuen Big Data Technologien ist der Datenschutz nicht zu vernachlässigen. Aufgrund der hohen Anzahl an Daten, die von Patienten gesammelt werden, sind die Patienten transparenter als zuvor. Für die IT-Sicherheit muss hierbei vorgesorgt werden, dass nur

berechtigte Nutzer auf die Daten zugreifen können. Sobald der Datenschutz vorgesorgt wurde und dieser verlässlich und datenschutzkonform ist, sollten in naher Zukunft Big Data Technologien in der Medizin angewendet werden, jedoch sollten die Fachkräfte mit Bedacht vorgehen und auch auf Erfahrung setzen.

Literaturverzeichnis

- [1] L. Summa, U. Mansmann, B. Buchner und M. Schnebbe, *Big Data in der Medizin*, Bd. 22, Freiburg/ München: Verlag Karl Alber, 2021.
- [2] A. Maier und M. Kaufmann, *SQL- & NoSQL-Datenbanken*, Heidelberg: Springer, 2016.
- [3] M. Brandt, „Statista,“ 18 August 2015. [Online]. Available: <https://de-statista-com.hs-ruhrwest.idm.oclc.org/infografik/3729/big-data-use-cases/>. [Zugriff am 28 Dezember 2021].
- [4] S. Rüping, „Big Data in Medizin und Gesundheitswesen,“ *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, Nr. 58, pp. 794-798, 2015.
- [5] J. McNellis, „You’re likely investing a lot in marketing analytics, but are you getting the right insights?,“ Gartner, 2019.
- [6] D. Eiduzzis, „Computerwoche,“ initions AG, 29 März 2018. [Online]. Available: <https://www.computerwoche.de/a/auf-der-suche-nach-dem-use-case,3544628>. [Zugriff am 05 Januar 2022].
- [7] D. Laney, „3D Data Management: Controlling Data Volume, Velocity, and Variety,“ META Group Inc., Connecticut, Stamford, 2001.
- [8] L. Fernando, „blogsofdatawarehousing,“ 17 Januar 2017. [Online]. Available: <http://blogsofdatawarehousing.blogspot.com/2017/01/7-vs-of-big-data.html>. [Zugriff am 23 Dezember 2021].
- [9] M. Volk, D. Staegemann und K. Turowski, „Big Data,“ in *Handbuch Digitale Wirtschaft*, Wiesbaden, Springer Gabler, 2020, pp. 1037-1053.
- [10] S. D’Onofrio und A. Meier, *Big Data Analytics: Grundlagen, Fallbeispiele und Nutzungspotenziale*, Wiesbaden: Springer Fachmedien Wiesbaden, 2021.
- [11] C. König, J. Schröder und E. Wiegand, *Big Data: Chancen, Risiken, Entwicklungstendenzen*, Wiesbaden: Springer VS, 2018.
- [12] L. Rabe, „Statista,“ 10 November 2021. [Online]. Available: <https://de-statista-com.hs-ruhrwest.idm.oclc.org/statistik/daten/studie/266885/umfrage/prognose-zum-datenvolumen-des-privaten-und-geschaefentlichen-ip-traffics-weltweit/>. [Zugriff am 10 Januar 2022].

- [13] F. Tenzer, „Statista,“ 19 Mai 2021. [Online]. Available: <https://de-statista-com.hs-ruhrwest.idm.oclc.org/statistik/daten/studie/172798/umfrage/datenvolumen-im-deutschen-mobilfunkmarkt-seit-2005/>. [Zugriff am 10 Januar 2022].
- [14] Apple, „Apple erfindet mit dem iPhone das Mobiltelefon neu,“ 9 Januar 2007. [Online]. Available: <https://www.apple.com/de/newsroom/2007/01/09Apple-Reinvents-the-Phone-with-iPhone/>. [Zugriff am 26 März 2022].
- [15] S. Dash, S. K. Shakyawar, M. Sharma und S. Kaushik, „Big data in healthcare: management, analysis and future prospects,“ *Journal of Big Data*, Bd. 6, Nr. 54, pp. 1-25, 2019.
- [16] R. Klausnitzer, *Das Ende des Zufalls: wie Big Data uns und unser Leben vorhersagbar macht*, Salzburg: Ecowin, 2013.
- [17] H. Holland, *Big Data – Chancen und Herausforderungen*, Wiesbaden: Springer Fachmedien Wiesbaden GmbH, 2020.
- [18] G. D. Rey und K. F. Wender, *Neuronale Netze*, Bern: Huber, 2008.
- [19] T. L. Nguyen, „A Framework for Five Big V’s of Big Data and Organizational Culture in Firms,“ IEEE, Denton, 2018.
- [20] A. S. Gillis, „Computerweekly,“ Mai 2021. [Online]. Available: <https://www.computerweekly.com/de/definition/Die-sechs-Vs-von-Big-Data>. [Zugriff am 26 Dezember 2021].
- [21] B. K. Hackenberger, „Data by Data,“ *Croatian Medical Journal*, Bd. 60, Nr. 3, pp. 290-292, 2019.
- [22] „Enterprise Big Data Framework,“ 31 Juli 2018. [Online]. Available: <https://www.bigdataframework.org/value-of-big-data/>. [Zugriff am 12 Dezember 2021].
- [23] „aerzteblatt,“ 22 Mai 2015. [Online]. Available: <https://www.aerzteblatt.de/nachrichten/62914/Ethikrat-Viele-Herausforderungen-durch-Big-Data-in-der-Medizin>. [Zugriff am 3 Januar 2022].
- [24] H. Landrock, „manage it,“ Experton Group, 8 Mai 2017. [Online]. Available: <https://ap-verlag.de/big-data-fuer-das-gesundheitswesen-huerden-und-chancen/33852/>. [Zugriff am 3 Januar 2022].
- [25] A. Gadatsch und H. Landrock, *Big Data im Gesundheitswesen*, Wiesbaden: Springer Fachmedien Wiesbaden, 2018.

- [26] A. Frodl, „E-Health-Economics,“ *Wirtschaftsinformatik & Management*, pp. 32-41, 25 März 2011.
- [27] M. Janson, „Statista,“ 09 Februar 2021. [Online]. Available: <https://de-statista-com.hs-ruhrwest.idm.oclc.org/infografik/24134/anteil-der-krankenhausaerzte-die-genomdaten-nutzen/>. [Zugriff am 10 Januar 2022].
- [28] P. Haas und R. Röhrig, „Institutionelle Informationssysteme im Gesundheitswesen,“ in *Praxishandbuch IT im Gesundheitswesen*, München, Carl Hanser Fachbuchverlag, 2009, pp. 201-232.
- [29] W. Stefan und B. Kurt, „Betrieb von IT-Systemen im Gesundheitswesen,“ in *Praxishandbuch IT im Gesundheitswesen*, München, Carl Hanser Fachbuchverlag, 2009, pp. 373-402.
- [30] R. Ball und D. Tunger, *Bibliometrische Analysen - Daten, Fakten und Methoden*, Jülich: Forschungszentrum Jülich, 2005.
- [31] Bundesministerium für Bildung und Forschung, „Bundesministerium für Bildung und Forschung,“ 16 Januar 2020. [Online]. Available: <https://www.bmbf.de/bmbf/shareddocs/kurzmeldungen/de/deutschland-tritt-genomprojekt-der-eu-bei.html>. [Zugriff am 18 01 2022].
- [32] The Royal College of Radiologists, „Clinical radiology UK workforce census report 2018,“ London, 2019.
- [33] N. Suttorp, M. Möckel, B. Siegmund und M. Dietel, *Harrisons Innere Medizin*, Berlin: ABW Wissenschaftsverlag, 2020.
- [34] Bundesärztekammer, „Elektronische Patientenakte,“ Berlin, 2021.
- [35] S. S. Abdul, L. F. Luque und P.-Y. Hsueh, *Data Analytics and Applications of the Wearable Sensors in Healthcare*, Basel: MDPI - Multidisciplinary Digital Publishing Institute, 2020.
- [36] H. Freistedt, *Die Pflegeversicherung - die Sicherung der Pflege durch Wandel*, Marosi Verlag, 2021.
- [37] N. L. Bragazzi, G. Damiani und M. Martini, „From Rheumatology 1.0 to Rheumatology 4.0 and beyond: the contributions of Big Data to the field of rheumatology,“ *Mediterranean Journal of Rheumatology*, Bd. 30, Nr. 1, pp. 3-6, 2019.
- [38] M. Bonin-Andresen, B. Smiljanovic, B. Stuhlmüller, T. Sörensen, A. Grützkau und T. Häupl, *Bedeutung von Big Data für die molekularen Diagnostik*, Berlin: Springer Medizin Verlag GmbH, 2018.

- [39] T. Kalil, „Obamawhitehouse,“ Office of Science and Technology, 29 März 2012. [Online]. Available: <https://obamawhitehouse.archives.gov/blog/2012/03/29/big-data-big-deal>. [Zugriff am 25 Februar 2022].
- [40] J. L. Fridovich-Keil, „Human Genome Project,“ *Encyclopedia Britannica*, 2020.
- [41] National Human Genome Research Institute, „The Cost of Sequencing a Human Genome,“ Bethesda, 2021.
- [42] Nature Methods, Method of the Year, New York: Nat Methods, 2008.
- [43] H. W. Weuck, Radiologie Träger des Fortschritts, Berlin: Springer Berlin Heidelberg, 1996.
- [44] T. Hackländer, T. Baumann und E. Kotter, Befundung in der Radiologie heute, Berlin: Springer Verlag, 2014.
- [45] R. Kramme, Medizintechnik: Verfahren – Systeme – Informationsverarbeitung, Berlin: Springer Verlag, 2011.
- [46] R. Singh, Tabari Azadeh und Y. Hoi, „Image Quality and Lesion Detection on Deep Learning Reconstruction and Iterative Reconstruction of Submillisievert Chest and Abdominal CT,“ *American Journal of Roentgenology*, Bd. 214, Nr. 3, pp. 566-573, 2020.
- [47] Hellerhoff, „CT mit mittelbasierter Bildrekonstruktion,“ 2021 November 2021. [Online]. Available: <https://de.wikipedia.org/wiki/Computertomographie>. [Zugriff am 27 März 2022].
- [48] L. Wahjudi, S. Bernhardt und K. Abnaof, „Integrating proteomics into precision oncology,“ *International Journey of Cancer*, Bd. 148, Nr. 6, pp. 1438-1451, 2021.
- [49] B. f. B. u. Forschung, „Bundesministerium für Bildung und Forschung,“ 29 Juli 2019. [Online]. Available: <https://www.bmbf.de/bmbf/shareddocs/kurzmeldungen/de/mit-big-data-gegen-den-krebs.html>. [Zugriff am 4 März 2022].
- [50] Bundesministerium, „Bundesgesundheitsministerium,“ 05 November 2021. [Online]. Available: <https://www.bundesgesundheitsministerium.de/elektronische-patientenakte.html>. [Zugriff am 24 Januar 2022].
- [51] G. Richter, W. Loh, A. Buyx und S. Graf von Kielmansegg, Datenreiche Medizin und das Problem der Einwilligung, Berlin: Springer Berlin Heidelberg, 2021.

- [52] B. Seref und E. Bostanci, „Opportunities, threats and future directions in big data for medical wearables,“ in *Proceedings of the International Conference on Big Data and Advanced Wireless Technologies*, New York, ACM, 2016, pp. 1-5.
- [53] R. Huss, *Künstliche Intelligenz, Robotik und Big Data in der Medizin*, Berlin: Springer Verlag, 2019.
- [54] S. Gröbner, B. Worst und J. Weischenfeldt, „The landscape of genomic alterations across childhood cancers,“ *Nature*, pp. 321-327, 28 Februar 2018.
- [55] F. X. Diebold, „On the Origin(s) and Development of “Big Data”: The Phenomenon, the Term, and the Discipline,“ University of Pennsylvania, Pennsylvania, 2012.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Bachelorarbeit selbstständig angefertigt habe. Es wurden nur die in der Arbeit ausdrücklich benannten Quellen und Hilfsmittel benutzt. Wörtlich oder sinngemäß übernommenes Gedankengut habe ich als solches kenntlich gemacht. Die vorgelegte Arbeit hat weder in der gegenwärtigen noch in einer anderen Fassung schon einem anderen Fachbereich der Hochschule Ruhr West oder einer anderen wissenschaftlichen Hochschule vorgelegen.

Bottrop, 28. März 2022

Ort, Datum

S. Güngör

Unterschrift