Technische Universität München

TUM School of Life Sciences

Leveraging policy setting, impact measurement and privacy technology for an increased implementation of Artificial Intelligence in healthcare

Justus Richard Wolff

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitzender:

Prof. Dr. Bernhard Küster

Prüfer der Dissertation:

1. **Prof. Dr. Jan Baumbach**
2. **Prof. Dr. Dimitri Frischmann**

Die Dissertation wurde am 16.08.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 25.11.2022 angenommen.

Table of Content

Acknowledgements

I would like to thank Prof. Dr. Jan Baumbach for the opportunity to pursue a doctoral degree, the provision of the topic and my supervision. In addition to that, I would like to thank Dr. Josch Pauling for my mentorship and the close collaboration throughout this time. Furthermore, I would like to thank the committee for the evaluation of this thesis.

The interaction with the colleagues at the chair has been very valuable and I would like to mention in particular the collaboration with Julian Matschinske during the FeatureCloud project. The guidance of Prof. Dr. Jan Baumbach, Dr. Josch Pauling and the chair members over the years was not only academically shaping but also personally very rewarding.

I also would like to thank my colleagues, in particular Dr. Andreas Keck, for their continuous support, encouragement and advice over the years. My family, friends and relationship were my foundation throughout these years.

List of Tables

List of Figures

List of Abbreviations

AI	Artificial intelligence
CAC	Coronary artery calcification
CACS	Coronary artery calcification score
CEO	Chief Executive Officer
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CE (in figures)	Conformité Européenne
CT	Computerized tomography
DH	Digital health
DiGa	Digitale Gesundheitsanwendung (English: Digital health application)
DoC	Department of Commerce
DVG	Digitale-Versorgung-Gesetz (English: German Digital Healthcare Act)
e.g.	Exempli gratia
EHR	Electronic Healthcare Record
EQ-5D	European Quality of Life 5 Dimensions 3 Level Version
et al.	Et alii
EU	European Union
FL	Federated learning

GMLP	Good Machine Learning Practices
i.e.	Id est
IoT	Internet of Things
IT	Information technology
LQT	Long-QT syndrome
ML	Machine learning
NAIAC	National Artificial Intelligence Advisory Committee
NIST	National Institute of Standards and Technology
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QALY	Quality Adjusted Life Year
R&D	Research and development
RS (in figures)	Research stream
SF (in figures)	Success factor
TTC	Trade and Technology Council
UK	United Kingdom
US	United States

Abstract

Leveraging policy setting, impact measurement and privacy technology for an increased implementation of Artificial Intelligence in healthcare

The decision of whether and how to implement Artificial Intelligence (AI) in healthcare depends on an interplay of medical, process, economic and further factors and can also differ significantly between medical indications. Since there is a significant gap between the very comprehensive and promising academic results for AI in healthcare and the low level of practical implementation, this thesis focuses on the identification of success factors and the deduction of measures for an increased implementation of AI in healthcare.

To achieve that, systematic reviews of academic research and real-world AI use cases, assessments of policy frameworks for Digital Health (DH) and AI as well as an empirical study in terms of an implementation of AI in healthcare were conducted.

The analysis of academic research and real-world AI use cases revealed three key success factor categories, namely policy setting, medical and economic impact measurement, and technological implementation, which were subsequently analyzed in more detail. First, successful policy frameworks such as the German Digital Healthcare Act can serve as a blueprint by providing clear medical and structural guidelines as well as evidence generation and reimbursement processes, which can serve as an international benchmark for DH and AI. Second, medical and economic impact assessments have not been conducted to a sufficient extent. In addition to the quantity, also the quality of such assessments needs to be improved. In particular, medical impact needs to be measured, for example, through Quality Adjusted Life Years, taking into account the CHEERS and PRISMA reporting criteria, while economic impact assessments should include, for example, Net Present Values and Cost Alternative Scenarios. Third, the technological implementation of AI should take place in a privacy-preserving manner to circumvent data availability and accessibility issues. In this respect, federated machine learning (FL) has proven to be one highly useful tool when analyzing the coronary artery calcification (CAC) scores of c.1,500 patients as a basis for the development of a predictive AI model for CAC risk. Privacy-preserving technologies such as FL can provide significant benefits in several medical fields, in this case potentially reduced radiation exposure for patients and cost savings due to a reduction of CT scans.

The identified success factors shall contribute to an overall increased application of AI in healthcare: Governments can continuously support this development via the first success factor of facilitating policy frameworks, e.g., through AI regulatory frameworks, reimbursement models and evidence generation standards. Furthermore, private and public institutions can leverage the two latter success factor categories, namely standardized medical and economic impact measurement as well as privacy-preserving technology like FL, to actively steer the implementation of AI. A close collaboration between governmental, medical and industry stakeholders as well as further academic research shall support this development to leverage the full potential of AI in healthcare and to ultimately achieve significant healthcare benefits globally.

Regulatorische Rahmenbedingungen, Wirkungsmessung und datenschutzkonforme Technologien als Erfolgsfaktoren für eine erhöhte Anwendung von Künstlicher Intelligenz im Gesundheitswesen

Die Entscheidung, ob und wie Künstliche Intelligenz (KI) im Gesundheitswesen implementiert werden soll, hängt von einem komplexen Zusammenspiel verschiedener Faktoren ab, die in hohem Maße kontextspezifisch sind und sich somit auch je nach Anwendungsfall unterscheiden können. Da eine offensichtliche Lücke zwischen den sehr umfangreichen und vielversprechenden akademischen Forschungsergebnissen zu KI im Gesundheitswesen und der geringen praktischen Umsetzung besteht, wurden Erfolgsfaktoren identifiziert und Maßnahmen für eine erhöhte Anwendung von KI im Gesundheitswesen abgeleitet.

Diesbezüglich wurden u.a. systematische Literaturübersichten von wissenschaftlichen Publikationen und realen KI-Anwendungsfällen im Gesundheitswesen erstellt sowie Analysen von regulatorischen Rahmenbedingungen für Digital Health (DH) und KI sowie eine empirische Studie in Form einer eigenständigen KI-Implementierung in medizinischen Einrichtungen durchgeführt.

Die Analyse von wissenschaftlichen Publikationen sowie realen KI-Anwendungsfälle ergab drei zentrale Kategorien von Erfolgsfaktoren: Regulatorische Rahmenbedingungen, medizinische und ökonomische Wirkungsmessungen und datenschutzkonforme Technologien, welche jeweils im Anschluss detaillierter analysiert wurden. Erstens können regulatorische Rahmenbedingungen wie das Digitale-Versorgung-Gesetz in Deutschland als internationaler Maßstab für DH und KI dienen, da sie klare medizinische und strukturelle Endpunkte für die Evidenzgenerierung und entsprechende Kostenerstattungsprozesse aufzeigen. Zweitens sind medizinische und ökonomische Effekte bisher nicht in ausreichendem Maße gemessen worden. Neben der Quantität muss auch die Qualität solcher Untersuchungen erhöht werden. Medizinischer Mehrwert sollte z.B. auf der Grundlage von sog. Quality Adjusted Life Years gemessen werden und die CHEERS- und PRISMA-Kriterien berücksichtigt werden, während ökonomische Bewertungen zum Beispiel um Net Present Values und den Kostenvergleich mit Alternativen ergänzt werden sollten. Drittens sollte die technologische Umsetzung von KI datenschutzkonform erfolgen, um gängige Probleme der Datenverfügbarkeit und -zugänglichkeit zu umgehen. In diesem Kontext hat sich z.B. Federated Machine Learning (FL) bei der Analyse der Verkalkung von Herzkranzgefäßen von ca. 1,500 Patienten als ein essentielles Instrument erwiesen. Algorithmen auf Basis von “Privacy-by-design”-Technologien wie FL können signifikanten Mehrwert erzeugen, in diesem konkreten Anwendungsfall potenziell geringere Strahlenbelastung für die Patienten sowie signifikante Kosteneinsparungen für das Gesundheitssystem.

Die identifizierten Erfolgsfaktoren und Maßnahmen sollen zu einer vermehrten Anwendung von KI im Gesundheitswesen beitragen. Die Politik kann dies durch die Bereitstellung geeigneter regulatorischer Rahmenbedingungen kontinuierlich unterstützen, z.B. durch Kostenerstattungsmodelle und Standards für die Evidenzgenerierung. Des Weiteren können private und öffentliche Einrichtungen die Erfolgsfaktoren standardisierter medizinischer und ökonomischer Bewertung und datenschutzkonformer Technologien nutzen, um die Anwendung von KI aktiv zu steuern. Eine enge Zusammenarbeit zwischen staatlichen Organisationen, medizinischen Einrichtungen und Industrie sowie weitere akademische Forschung sollen diese Entwicklung unterstützen, um das gesamte Potenzial von KI im Gesundheitswesen auszuschöpfen und damit schlussendlich weltweit erhebliche Verbesserungen im Gesundheitswesen zu ermöglichen.

1. Introduction

Prior research has demonstrated that Artificial Intelligence (AI) is highly promising regarding its improvement potential in healthcare, for example through higher accuracy levels in diagnostic assessments or time savings in therapeutic processes. At the same time, researchers also found that AI is not yet implemented to a significant extent in the day-to-day healthcare processes (e.g., Topol 2019, Kelly et al. 2019, Panch, Mattie, and Celi 2019).

The European Union (EU) also clearly states in one of its latest research reports that despite a number of initiatives, healthcare organizations are slow in implementing AI technologies and that the level of adoption is overall low. To promote the development and adoption of AI technologies, the European Commission names a variety of challenges that need to be addressed such as the lack of policy and access to healthcare data, low investments, the need to upskill healthcare professionals as well as to educate AI developers on current clinical practices (European Union 2021).

Even when analyzing the situation in specific EU countries, for example, regarding AI policies in healthcare, the report reveals that most initiatives focus on the research and innovation area with little activity or initiatives to promote actual adoption within the healthcare sector itself (European Union 2021).

This gap between the promising academic results on AI in healthcare and its low practical implementation is based on various circumstances and therefore corresponding success factors needed to be identified and subsequently leveraged to increase the real-world implementation of AI in healthcare. In order to achieve that, systematic reviews of academic research and real-world AI use cases in healthcare, analysis of governmental DH policies as well as an empirical study in terms of an actual implementation of AI in healthcare were conducted.

Looking at the market more broadly, namely across different industries, AI research and real-world implementation has drastically increased over the last years and AI has already transformed several industries, markets, and business models. One illustrative example is the retail market in which AI-powered functionalities have become widely used. Among others, retailers commonly incorporate chatbots into their websites and online marketplaces, which respond to customers' inquiries and provide assistance to their claims. Also, through AI-based customer profiling, companies commonly provide personalized shopping preference recommendations to their customers (Ameen et al. 2021).

In a similar vein, AI has already significantly transformed the manufacturing industry. For example, AI is commonly applied in the context of demand prediction to improve manufacturing planning and logistics. Further, by means of AI-powered optical quality assurance mechanisms, defects and deviations from standards can be identified at lower cost. Also, AI has been used for predictive maintenance which aims at maximizing the useful life of machines and avoiding disruptions in operations (Fahle 2020). The latter is not only commonly applied in factories, but also to avoid, for example, the failure of rail infrastructure (e.g., by Konux).

In the healthcare industry, AI has also a significant transformation potential and this is strongly supported by both, academic research (e.g., Triantafyllidis and Tsanas 2019) and economic figures revealed by market studies (e.g., Grand View Research, 2019). Also specific physician categories can already be seen as potential frontrunners and for example research by Lin highlights that primary care providers could be early AI adopters due to their dominant role in the overall healthcare structure (Lin 2022).

Interviews with industry leaders even suggest that the potential of AI may actually be most significant in the healthcare industry. For example, in early 2020, Sundar Pichai, the CEO of Alphabet and its subsidiary Google, announced that healthcare offers the largest potential over the next five to ten years for using AI to improve outcomes (Reuters 2020).

Indeed, benefits from AI may arise in numerous different forms and contexts in healthcare and can provide significant medical value through enhanced procedures in research, prevention, diagnosis, and therapy. Examples are a higher accuracy and pace in drug discovery and clinical trials, improved diagnostic decision-making regarding patient treatment based on automatized data analysis, and decision support in the context of medical imaging. Further, AI in healthcare may also create significant economic value in the form of saved time resources on the side of the medical professionals, less costly clinical trials, more efficient treatment procedures (e.g., through AI-powered self-management set-ups), and reduced interventions due to more accurate disease detection (Ilan 2020).

Against this background, it seems surprising that AI has not already "conquered" healthcare to a significant and comparable extent, as opposed to other industries and markets such as retail, manufacturing or finance. In particular, prior research showed that there are only a few large-scale real-world cases of AI application in healthcare (He et al. 2019). The number of real-world and large-scale AI applications lags behind considering the high number of AI start-ups in healthcare and record numbers in their funding, as well as the large number of academic studies and their positive predictions on AI's value-added in healthcare (Pifer 2019).

This is even more surprising as the right to health and respective equality have moved more and more into the focus of today's society (World Health Organization 2008), e.g., as part of the 2030 Global Agenda for Sustainable Development.

In addition to that, the Corona pandemic has clearly revealed healthcare systems' insufficiencies and the importance of digitization of healthcare procedures. Given the urgent need for timely action in this context, the pandemic has considerably increased people's comfort with Digital Health (DH) and AI applications for the purpose of safety and health (MedTech Innovation 2021).

This “controversy“, namely the gap between promising academic research and the low real-world AI implementation in healthcare, raised the question, which hurdles exist and which success factors can be leveraged to overcome them to increase the implementation of AI in healthcare. While the real-world implementation gap can only be closed over time as a collaborative effort between researchers, medical institutions, governments and further stakeholders, this research aims to be one first step in this direction and provide concrete recommendations in this segment.

Several research projects were undertaken which resulted in the four publications of this thesis:

- “Success Factors of Artificial Intelligence Implementation in Healthcare” highlighting key success factor categories that can be leveraged to achieve an increase in real-world AI implementation (published in Frontiers in Digital Health on 16.06.2021, 8 citations)
- “The Impact of Artificial Intelligence on the Healthcare Economy” elaborating on the current status of AI research, real-world implementation in healthcare, reimbursement structures and market trends (under review in Elsevier “Artificial Intelligence and Machine Learning in Healthcare”)
- “The Economic Impact of Artificial Intelligence in Healthcare: A Systematic Review” analyzing the quantity and quality of existing medical and economic impact assessments and deriving respective areas for improvement (published in Journal of Medical Internet Research on 20.02.2020, 61 citations)
- “Federated machine learning for a facilitated implementation of Artificial Intelligence in healthcare – a proof of concept study for the prediction of Coronary Artery Calcification Scores” representing an actual real-world implementation case of AI in healthcare (accepted with minor revisions in Journal of Integrative Bioinformatics on 22.07.2022)

The thesis is structured as follows: The Chapter 2 contains an introduction outlining the status of the current scientific research. Chapter 3 elaborates on the contribution of the described research in this landscape. Chapter 4 summarizes the key results and interconnection between the four publications forming part of this cumulative thesis. Chapter 5 contains the four individual publications and finally, a discussion and conclusion are presented in Chapters 6 and 7, respectively.

2. State of the art

There have been various prior academic research projects about AI in healthcare and the broad existing publication landscape can be segmented into two research fields: On the one hand healthcare as a market and which medical and economic advancements are needed, and on the other hand AI as a technology in healthcare and which medical and economic advancements could be possible. In the middle between those two segments can the success factors be placed, that shall "unlock" the potential for advancements through AI technology. An overview about the segments can be found in Figure 1:

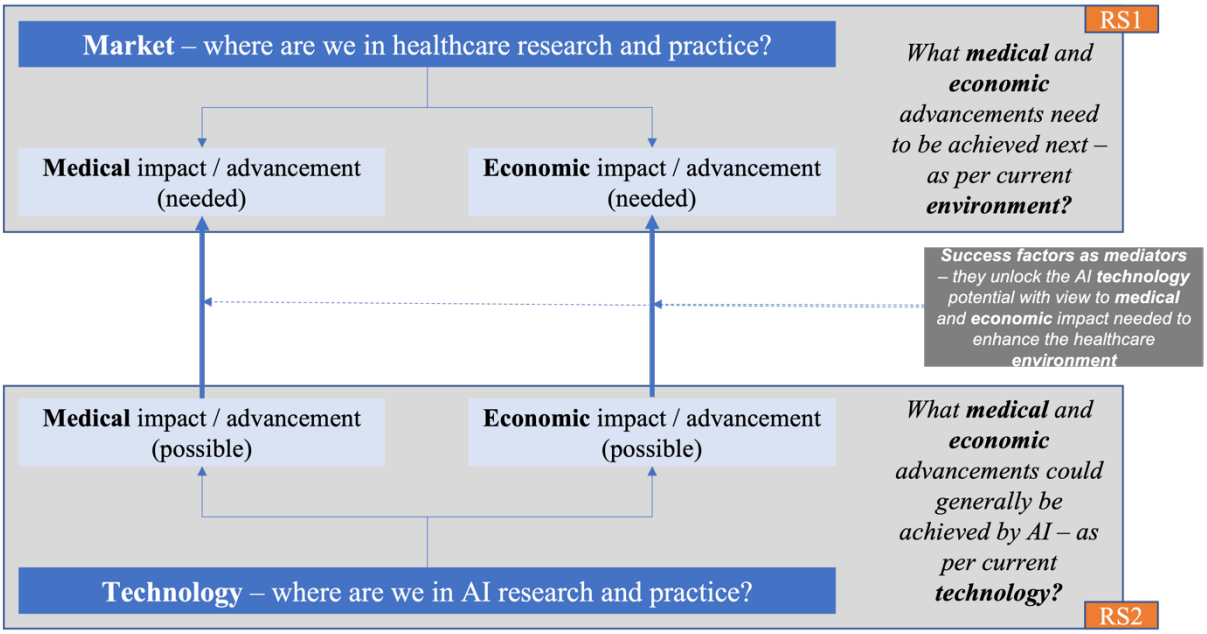


Figure 1: Overview of key research streams

Market

Accordingly, the first research stream consists of the following two research areas: "Market – Medical Impact", which relates to the question of which medical advancements are needed based on the current healthcare environment, and "Market

– Economic Impact", which relates to the question of which economic advancements are needed based on the current healthcare environment.

Overall, this research stream reveals a significant need for both medical and economic advancements in healthcare, considering the current status quo of the healthcare environment. As to medical advancements, undersupply for certain groups and / or in certain regions as well as different demographic developments (e.g., aging population, increase in chronic diseases) actually require an expansion of both the access to and the scope of healthcare services. At the same time, in many countries, medical facilities are facing a shortage of labor, causing delays and pressure on existing medical staff (e.g., Kumar 2019). With regard to economic advancements, global healthcare spending has not only grown steadily, but also represents a considerable burden for economies worldwide, e.g. ca. 18% of the annual GDP are spent on healthcare in the US and ca. 12% in Germany (World Health Organization 2019), (Wolff et al., 2020).

As such, taken together, there is an urgent need to to cope with increased demand for healthcare services and, in many cases, simultaneous labor shortages while containing the worldwide rise in healthcare spending. In addition, specific needs for medical and economic advancements have emerged as a consequence of the global corona pandemic such as to achieve efficient infection chain tracing or to organize safe vaccination campaigns globally.

Technology

AI can play an important role in order to meet these different needs, especially when considering how AI solutions have already transformed many other markets and industries and their respective business models, products and services (Budd et al. 2020). As outlined above, some common examples are AI-powered customer service offerings such as chatbots, personalized advertising based on each customer's search

and purchase history in e-commerce, and the predictive maintenance of machines in manufacturing.

The second research stream also consists of two research areas in parallel to the above presented first research stream: “Technology – Medical Impact”, which relates to the question of which medical advancements are possible based on the current AI and related technology, and “Technology – Economic Impact”, which relates to the question of which economic advancements are possible based on the current AI technology.

The potential for medical impact of AI is broad and ranges from drug discovery over AI driven symptom checkers and decreasing non-adherence costs to shortening of recovery time (Garbuio and Lin 2019). Additionally, a completely new scope of diagnostic and therapeutic approaches is possible and, for example, the application of machine learning for genome analysis is a promising reference case that shall lead to better medical outcomes with e.g., LQT cardiac rhythm patients (Horizon 2020).

As to economic advancements, previous analyses revealed that AI comes along with significant cost saving potential in healthcare, e.g., through time savings via procedural optimizations or the involvement of new diagnostic tools within medical institutions (Accenture 2017) and also by process improvements like insurance claim approvals (Accenture 2018). In addition to that, the economic burden of an ageing population that faces a global labour shortage of over 9,9 Million physician, nurse and midwives until 2030 shall be encountered through significant AI driven increases in productivity and efficiency in chronic care management, clinical decision support etc. (McKinsey 2020).

The potential of AI in healthcare also greatly benefits from improvements in general technological capabilities. These have been achieved over the last years with regard to data storage capacities, processing power and cloud computing for AI applications (Mordor Intelligence 2020), (Aguis 2019). Overall, the academic landscape indicates

that the technical status of AI could be very promising to achieve the medical and economic advancements that are needed as per the current healthcare environment.

Yet, several authors confirmed that it became visible in their research on the real-world implementation of AI that, despite the significant growth potential and promising academic research findings, there are still relatively few actual real-world AI applications in routine healthcare processes (e.g., He et al. 2019) and that those can only be found in a limited number of healthcare segments (e.g., MarketsAndMarkets 2020) and regions (e.g., Grand View Research 2019). Further, in the past, the medical and economic impact of AI has not been measured to a sufficient extent and quality (Wolff et al. 2020), which makes it difficult to conclusively evaluate or precisely estimate the actual potential of AI in healthcare (Sanyal et al. 2018).

Several authors also elaborated on the reason for the gap between research and implementation and the various challenges regarding the real-world use of AI in healthcare, such as patient consent issues, transparency and ownership of data or privacy and discrimination regulations (Racine, Boehlen, and Sample 2019). He et al. (2019), for example, discussed various concrete and practical improvement areas that would be required related to data sharing, transparency of algorithms, data standardization and interoperability. Also, Alhashmi et al. surveyed 53 health and IT specialists and highlighted the importance of managerial, organizational, operational and IT infrastructure-related factors for applying AI in healthcare (Alhashmi, Salloum, and Abdallah 2020). A further recent qualitative interview study regarding AI implementation challenges with 26 healthcare leaders like politicians or hospital managers in Sweden additionally highlighted, that 1) conditions external to the healthcare system like liability or quality standards, 2) the internal capacity for change management like strategy setting or resource allocation, and 3) transformation of healthcare professions and practice like managing new roles are key hurdles (Petersson et al. 2022).

3. Own contribution

Despite the previous recognition of the gap between real-world AI implementation in healthcare, both by the promising academic literature and economic figures in this field, and analyses of potential improvement areas, this thesis is one of the first attempts to approach this gap systematically. The aim was to examine both, AI's potential in healthcare from a multi-stakeholder perspective and common hurdles to AI implementation, for the deduction of concrete success factor categories and respective measures. To this end, the chosen research approach is based on systematic reviews of previous academic publications, market research, assessments of real-world AI applications in healthcare and existing facilitator frameworks, as well as an empirical analysis which has been performed through the implementation of an AI diagnostic application. These success factors and respective recommendations shall unlock the potential of AI in healthcare, as illustrated in Figure 2:

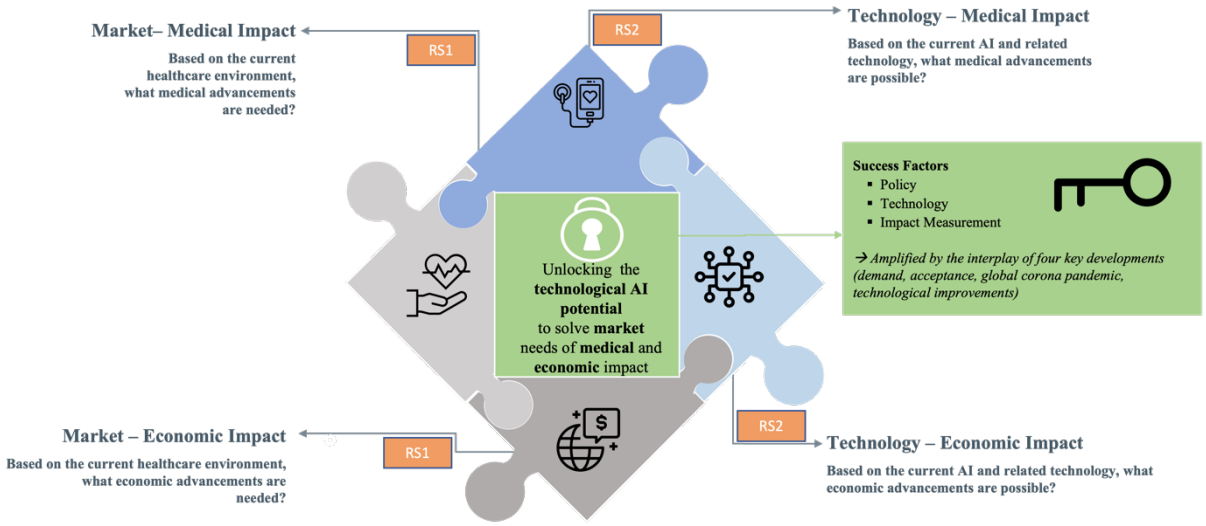


Figure 2: Overview of interlinked research streams

In the following, the research approaches of the four publications are presented:
The first paper represents an analysis of academic studies and real-world AI use cases to systematically identify success factors for AI implementation in healthcare. The possible success factor categories were derived from a prior World Health Organization survey about barriers of adoption of Big Data within 125 countries in which categories such as “Lack of integration”, “Privacy and security”, “Missing national policy” were mentioned as particularly important for the adoption of Big Data applications in healthcare. The research was conducted as a systematic literature review covering the literature databases Scopus and Opac Plus as well as a Google advanced search query for real-world AI use cases. The inclusion and exclusion criteria were, among others, a publication date between 2015-2020, a comprehensive description of an AI functionality as well as its efficiency and outcomes. Following this systematic search approach, out of 1,494 identified academic studies, 51 were included, while out of 237 identified real-world AI use cases, 30 were included.

The second paper represents a market analysis of the impact of AI on the healthcare economy as well as recommendations on how common hurdles to AI implementation in healthcare can be overcome. The research approach combines a non-systematic literature review and market research with two assessments of real-world facilitator framework for AI implementation, namely of the German Digital Healthcare Act and the EU funded FeatureCloud platform. The paper includes academic, market as well as real-world case and governmental sources in order to define policy and technology reference cases.

The third paper elaborates in more detail on medical and economic impact measurement. As such measurements are crucial for decision-making on AI implementation, an in-depth evaluation of the quantity and quality of existing studies has been performed. The research was conducted as a systematic literature review covering the literature database PubMed using combinations of the most frequently used search terms related to AI in healthcare based on a 1-year Google Trends analysis. The inclusion and exclusion criteria were, among others, a publication date

in the year 2015-2020, the description of concrete outcomes (e.g., cost savings per patient per year) and at least one of the following content sectors: a) a comprehensive description of an AI functionality, b) an evaluation of the economic efficiency and outcomes of the AI functionality, and c) quantitative outcomes of the AI functionality in at least one health care system. Following this systematic search approach, out of 66 identified academic studies, 6 were included.

Finally, the fourth paper represents a real-world AI implementation in a privacy-preserving medical setting to assess whether the success factor category technology and the abovementioned facilitator framework FeatureCloud for privacy-preserving data access can successfully be leveraged to circumvent common data-related hurdles to AI implementation. The paper represents a novel empirical research study as federated machine learning (FL) was implemented to predict coronary artery calcification (CAC) scores (CACSs) as a risk indicator for subsequent CT screening. The prediction model takes the following independent risk factor areas into account: Age and biological sex, obesity, dyslipidemia, and diabetes mellitus. For the actual implementation, the FeatureCloud platform was applied in two medical units in Germany and the model was trained based on medical data of 1,450 patients. The results were analyzed with regard to sensitivity and specificity in a comparison between a traditional centralized approach and the FL approach.

Altogether, the thesis is based on different methods and research approaches to provide answers for the question of how to increase real-world AI implementation in healthcare. In particular, the aim was a combination of academic and practical insights through a translational research approach, as both types of insights contribute to an increased AI implementation in healthcare.

4. Results

The main results and interconnection of the four papers forming part of this thesis are described in what follows.

The first paper, entitled “Success Factors of Artificial Intelligence Implementation in Healthcare”, assessed the key success factor categories for increased real-world implementation of AI in healthcare. Three key success factor categories, namely (1) policy setting, (2) technological implementation, and (3) medical and economic impact measurement could be identified and for each of them a set of specific recommendations was deducted: First, a risk-adjusted policy framework is required that distinguishes between precautionary and permissionless principles, and differentiates among accountability, liability, and culpability. Second, a “privacy by design” technological infrastructure is particularly promising to overcome common hurdles to AI implementation in healthcare as it enables practical and legally compliant data access. Third, medical and economic impact assessments need to be conducted at higher frequency and quality as respective evidence represents a key prerequisite for strategic decision-making, both from a medical and economic perspective. Overall, the analysis revealed that private and public institutions can already today actively leverage these success factor categories and follow the provided recommendations and thereby drive the translation from academic research to real-world application. There are likely additional key success factors for AI implementation in healthcare (e.g., trust-building measures), the identified success factors are interlinked, and different success factors can be relevant under different circumstances. Thus, future research should elaborate on further success factors, their context-specific relevance, and how they can be leveraged together to exploit the full potential of AI in real-world applications.

The second paper, entitled “The Impact of AI on the Healthcare Economy”, assessed the economic impact from a market perspective. This assessment shows that the interplay of four key drivers is likely to trigger a transformation of the healthcare market

through AI. First, there is an urgent need for economies worldwide to limit the rise in healthcare spending while responding to an increased demand for healthcare services, which in many cases is particularly challenging due to labor shortages. Second, significant technological improvements in recent years enable simplified and scaled AI implementation today more than ever. Third, for the first time, noteworthy awareness and acceptance levels can be observed for AI applications in healthcare. Finally, the corona pandemic has significantly increased the need for DH structures, more generally, and AI, in particular. Indeed, it has put a significant strain on hospitals and medical staff across the globe, frequently bringing them to the edge of their resources. Therefore, for example, since the outbreak of the corona pandemic, EU-backed artificial intelligence has been used to analyze over 20,000 CT scans (European Commission, 2021). Furthermore, the paper also elaborated on the above mentioned success factors categories in more depth.

As to categories (1) and (2), the paper presents two concrete facilitator frameworks for DH and AI implementation. The first facilitator framework is a policy framework that has recently been adopted in Germany, namely the German Digital Healthcare Act, which has significantly facilitated reimbursement for DH and AI healthcare services (so-called ‘Digitale Gesundheitsanwendung’ or ‘DiGa’). In order to prescribe a DiGA, a comprehensive application submission, validation and reimbursement scheme has been developed and implemented by the German authorities. Accepted solutions must fulfill the general requirements of safety, quality, functionality, privacy and data security as well as demonstrate a so-called “positive care effect”. The latter consists of a medical benefit and/or structural and procedural effects with clearly defined endpoints. For all solutions, there are two application forms, differentiating between solutions with priorly collected data for permanent listing and solutions without priorly collected data and, thus, only for a preliminary “trial period” in a provisional listing. It was shown that this reimbursement system, which is open for national and international applications, and its underlying evidence generation structure, enables market and reimbursement access even for comparatively new solutions, and will, thus, very likely lead to a significant increase of real-world DH and AI applications. After one year of existence

of the law, so far ca. 50,000 DiGAs have been prescribed by physicians to publicly insured patients in Germany (HIH Presentation, 2021).

The second facilitator framework is the FeatureCloud platform which implements a software toolkit for privacy preserving data access and model training. The technological set-up has two key features: (A) no sensitive raw data is exchanged and (B) data by several institutions are aggregated in a meta model. This technological infrastructure allows for privacy-preserving data access and therefore has the potential to lead to a significant increase of real-world AI applications in healthcare. Besides the direct positive impact of such facilitator frameworks on real-world AI implementation, it is likely that they will further induce an indirect positive impact by serving as role models for similar initiatives. As to success factor category (3), the paper elaborated on the fact that in order to support an AI solution, high-quality and comprehensive impact measurement is indispensable, which is also assessed in detail in the third paper.

The third publication, entitled: “The Economic Impact of AI in Health Care: Systematic Review”, covered the success factor category (3), namely medical and economic impact measurement, by systematically reviewing existing impact assessments of AI applications. It revealed that there are only few medical and economic impact assessments and that these are commonly subject to methodological flaws. The systematic literature review revealed that only 6 out of 66 publications could be included in the analysis based on the determined inclusion criteria. Out of these 6 studies, none comprised a methodologically complete cost impact analysis. Thus, to date, decisions for or against AI implementation often lack a suitable foundation. Due to the corresponding uncertainty and lack of scientific justifications, decision-makers likely abstain from implementing AI. To counteract this, the study presents concrete levers for improvement when conducting impact assessments. In particular, the impact should be defined based on Quality Adjusted Life Years (QALYs) while applying the CHEERS and PRISMA quality criteria. In addition to that, the initial investment and operational costs for the AI solution need to be considered and alternatives to achieve

a similar impact must be evaluated to allow for a comprehensive comparison as a basis for strategic decision-making.

The fourth paper, entitled: “Federated machine learning for a facilitated implementation of Artificial Intelligence in healthcare – a proof of concept study for the prediction of Coronary Artery Calcification Scores”, describes the success factor category (2) in more detail, namely technological implementation. It covers the implementation of FL in a real-world medical context, using the above described FeatureCloud platform. In particular, the paper addresses with that a way to overcome one key hurdle to AI implementation, which is the necessity of accessing large, private and scattered amounts of data. The study is based on real patient data of two medical institutions in Germany and provides insights on the accuracy of a privacy-preserving FL approach and the according benefits as compared to a traditional, i.e., centralized, AI approach. The FL approach slightly outperforms the centralized approach with a sensitivity of 67% (compared with 69%) while slightly underperforming it with a specificity of 69% (compared with 70%). Overall, it could be demonstrated that AI-based prediction of CACSs is feasible via both a centralized and a FL approach, since their accuracy is very comparable. In order to increase the prediction accuracy and, thus, enable real clinical value, further patient data are required and FL can be utterly necessary for that, since these data are otherwise in most cases not accessible. The developed FL approach “CACulator” serves as proof of concept, and is available as an open research tool and shall support future research internationally to facilitate AI implementation.

Taken together, Fig. 3 provides an overview of the unifying conceptual framework and shows how the four papers are interlinked and contribute to the question of how to increase real-world AI implementation in healthcare.

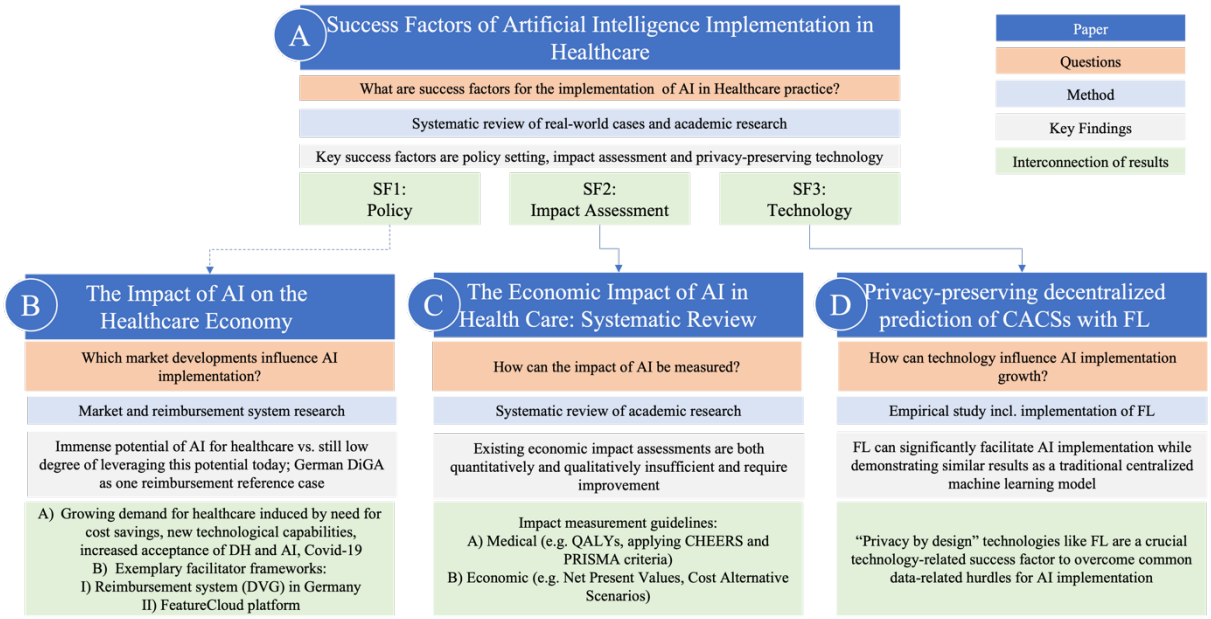


Figure 3: Overview about publications, research questions, provided answers and interconnection between publications

The contribution of this thesis is of both academic and practical nature:
As to the academic contribution, a systematic analysis of success factors for real-world AI implementation in healthcare based on academic literature and real-world AI use cases was conducted. In this context, conclusions regarding policy setting, technological implementation as well as medical and economic impact measurement could be drawn. Furthermore, the quantity and quality of medical and economic impact measurement as one central lever and research area to support AI scaling were analyzed and respective improvement areas identified and described. In addition to that, translational research was applied since FL as a privacy-preserving AI technology was implemented in the real-world context of two medical institutions in Germany.

As to the practical contribution, for the identified success factor categories, specific recommendations for governmental stakeholders, healthcare professionals and business management have been derived in order to increase real-world application of AI in healthcare (Details of these recommendations can be found in chapter 7). Related to this, existing facilitator frameworks and role model use cases, such as a

successful reimbursement policy framework and reference cases for AI implementation, were presented.

Taken together, the research aimed at systematically and comprehensively examining the gap between the promising academic results about AI and the low real-world implementation, and deducted success factor categories with respective recommendations for multiple stakeholder groups. In the following, key success factors and recommendations are summarized:

- Policy settings that contain clear standards with regard to regulatory requirements such as medical product classes, evidence generation endpoints and according evidence generation pathways as well as clearly formulated reimbursement models for DH and AI solutions
- Application of AI technologies that allow for privacy-preserving data access and data sharing for AI model training in order to circumvent common data-related hurdles while still maintaining data privacy standards
- High-quality medical and economic impact assessments of AI applications in order to analyze the benefits of AI and, thereby, enabling more comprehensive strategic decision-making on medical and business management level

5. Publications

Paper A: Success Factors of Artificial Intelligence Implementation in Healthcare

The paper entitled “Success Factors of Artificial Intelligence Implementation in Healthcare” represents a systematic review of academic studies and real-world AI use cases and aims at identifying the key success factors for increased real-world implementation of AI in healthcare.

Title	Success Factors of Artificial Intelligence implementation in healthcare
Research question	What are the success factors to achieve a higher level of real-world AI implementation in healthcare?
Background	Low number of real-world AI implementations in healthcare in general due to various reasons and significant gap between recent years’ academic advancements and reality
Contribution	Shows that there are only very few real-world AI implementations in healthcare, and presents both barriers to and benefits of AI in healthcare as well as success factors to increase real-world implementation, especially from three categories, namely policy, technology, and medical and economic impact measurement
Method	Systematic review of academic studies and real-world use cases as a basis for the identification of success factors for real-world implementation; studies for in-depth analysis are identified and assessed for inclusion via a systematic search and inclusion process
Take-away	In light of the key potential of AI in healthcare, on the one hand, and the low number of real-world AI implementations in healthcare, on the other hand, in the future, governments, scientists, medical practitioners, industry and further stakeholders should consider and improve on a range of success factors that facilitate AI implementation. In this regard, success factors from the following three categories could be identified: a) a facilitating policy setting (especially with view to risk allowance), b) a non-restrictive technological infrastructure (especially with view to data privacy

	preservation), and c) a high-quality medical and economic impact assessment

Table 1: Overview about the publication “Success Factors of Artificial Intelligence implementation in healthcare”

Contribution of the doctoral candidate: First authorship including the planning of the publication structure, a systematic review of academic studies and real-world use cases and deduction of consequences for governments, scientists, medical practitioners and industry stakeholders. Manuscript: Writing, review and editing.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 16 June 2021
doi: 10.3389/fdgth.2021.594971

Success Factors of Artificial Intelligence Implementation in Healthcare

Justus Wolff[1,2]*, *Josch Pauling*[3], *Andreas Keck*[2] and *Jan Baumbach*[1,4]

[1] Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Munich, Germany, [2] Syte - Strategy Institute for Digital Health, Hamburg, Germany, [3] LipiTUM, Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Munich, Germany, [4] Chair of Computational Systems Biology, University of Hamburg, Hamburg, Germany

Background: Artificial Intelligence (AI) in healthcare has demonstrated high efficiency in academic research, while only few, and predominantly small, real-world AI applications exist in the preventive, diagnostic and therapeutic contexts. Our identification and analysis of success factors for the implementation of AI aims to close the gap between recent years' significant academic AI advancements and the comparably low level of practical application in healthcare.

Methods: A literature and real life cases analysis was conducted in Scopus and OpacPlus as well as the Google advanced search database. The according search queries have been defined based on success factor categories for AI implementation derived from a prior World Health Organization survey about barriers of adoption of Big Data within 125 countries. The eligible publications and real life cases were identified through a catalog of in- and exclusion criteria focused on concrete AI application cases. These were then analyzed to deduct and discuss success factors that facilitate or inhibit a broad-scale implementation of AI in healthcare.

Results: The analysis revealed three categories of success factors, namely (1) policy setting, (2) technological implementation, and (3) medical and economic impact measurement. For each of them a set of recommendations has been deducted: First, a risk adjusted policy frame is required that distinguishes between precautionary and permissionless principles, and differentiates among accountability, liability, and culpability. Second, a "privacy by design" centered technology infrastructure shall be applied that enables practical and legally compliant data access. Third, the medical and economic impact need to be quantified, e.g., through the measurement of quality-adjusted life years while applying the CHEERS and PRISMA reporting criteria.

Conclusions: Private and public institutions can already today leverage AI implementation based on the identified results and thus drive the translation from scientific development to real world application. Additional success factors could include trust-building measures, data categorization guidelines, and risk level assessments and as the success factors are interlinked, future research should elaborate on their optimal interaction to utilize the full potential of AI in real world application.

Keywords: artificial intelligence, digital health, technology assessment, impact measurement, policy framework, success factor, public health

OPEN ACCESS

Edited by:
Raghav Sundar, National University Health System, Singapore

Reviewed by:
Bee Luan Khoo, City University of Hong Kong, Hong Kong
Dean Ho, National University of Singapore, Singapore

***Correspondence:**
Justus Wolff
justus.wolff@syte-institute.com

Specialty section:
This article was submitted to Original Research Article, a section of the journal Frontiers in Digital Health

Received: *14 August 2020*
Accepted: *22 April 2021*
Published: *16 June 2021*

Citation:
Wolff J, Pauling J, Keck A and Baumbach J (2021) Success Factors of Artificial Intelligence Implementation in Healthcare. Front. Digit. Health 3:594971. doi: 10.3389/fdgth.2021.594971

INTRODUCTION

Artificial Intelligence (AI) is having the potential for a significant impact on the entire healthcare industry. Consequently, first governmental structures for Digital Health and subsequent AI scaling are currently being defined. For instance, the German government has published a national law for the reimbursement of registered Digital Health services by public health insurances (1, 2). Based on the growing amount of digital health applications, the high expectations related to medical, social, and economic improvements, as well as the need for digital health routines triggered by COVID-19, the success factors for AI implementation need to be defined now.

The academic literature elaborated in detail on the benefits and challenges of AI in healthcare. Already in 2015, Deo reported that "although there are thousands of papers applying machine learning algorithms to medical data, very few have contributed to clinical care" and potential obstacles for machine learning implementation require further research (3). In 2018, Park and Han provided methodological guidelines to evaluate the clinical performance of AI for medical diagnosis and prediction (4). In the same year, Yu et al. described different potential applications of AI and the clinical integration at different AI development stages (5).

In 2019, Triantafyllidis and Tsanas noted that still only few real world Digital Health intervention studies could be identified for their review of machine learning applications. However, the ones identified and analyzed were useful and effective (6). In the same year, Racine et al. highlighted substantial challenges concerning the use of AI, including dynamic information and consent, transparency and ownership, and privacy and discrimination (7). Furthermore, He et al. confirmed there are limited real-world AI applications, and the authors discussed various concrete and practical improvement areas related to data sharing, transparency of algorithms, data standardization and interoperability (8).

In 2020, Alhashmi et al. surveyed 53 health and IT specialists and highlighted the importance of managerial, organizational, operational and IT infrastructure related factors for AI applications (9).

Despite the substantial ongoing research regarding the benefits and improvement of AI in healthcare, there are only a few real-world application cases covered in academic research or openly published. These include, among others, major initiatives such as IBM's investment of over 4 billion USD into IBM Watson (10), and Amazon, which agreed with Cerner to establish a range of AI in healthcare services under Amazon Web Services (11). In addition, start-ups have also brought successful AI applications to the market. For example, the FDA approved deep learning platform Arterys or Babylon Health, which performs ∼4,000 clinical consultations on their platform per day (8, 12).

From our perspective, a gap between the promising and comprehensive academic research on the high potential of AI in healthcare and the comparably low level of actual practical implementation can be observed. Despite previous recognition of this gap and isolated analyses of potential areas of improvement, this is the first attempt to systematically identify success factors that significantly facilitate the implementation of AI in healthcare based on previous academic research and real-world AI applications.

Abbreviations: TUM, Technical University Munich; OPAC, Online public access catalog.

MATERIALS AND METHODS

First, the success factor categories and according database search queries have been defined and there are several success factors, that had already been researched in prior publications. For example in 2016, Ross et al. identified factors that influence the implementation of eHealth and found that the individual e-health technology, the outer setting, the inner setting, the individual health professionals as well as the process of implementation are key success factors (13).

In our case we derived the success factor categories from the Big Data section results of the "Global diffusion of eHealth: Making universal health achievable" report of the World Health Organization (WHO), as displayed in **Figure 1**. In this global survey with 125 WHO member countries the following results with regard to adoption barriers of Big Data were revealed (14).

Roughly 70% of countries mentioned "lack of integration" (72%; $n = 81$) and "privacy and security" (68%; $n = 78$) as very or extremely important barriers to adoption. Furthermore, about 60% of countries considered "information sharing" (61%; $n = 70$), "promotion of standards" (61%; $n = 70$), and "building capacity" (59%; $n = 68$) in the same category. In addition to that, "new analytical methods" were mentioned (55%; $n =$ NA). Furthermore, only less than a fifth of all countries (17%; $n = 21$) reported to have a national policy or strategy regulating the use of big data in the health sector and thus from our perspective "Strategy setting" based on consequent impact measurement is also a key barrier for the adoption.

Based on these results three improvement categories have been deducted:

1) Technology ("Lack of integration," "Privacy and security," and "Information sharing")
2) Policy ("Promotion of standards" and "Building capacity")
3) Medical and economic impact ("New analytical methods" and "Strategy setting")

Thus, in this paper success factors are defined as facilitators for AI implementation based on recommendations across the segments technology, policy as well as medical and economic impact.

Academic Literature

Academic literature was accessed and identified via a research of the data base "Scopus" with the search terms "Artificial Intelligence," "Healthcare," "Health care," "Success factor," "Technology," "Policy," "Medical Impact," and "Economic Impact" (Search term query: "artificial intelligence" AND "healthcare" OR "health care" AND "success factor" AND "technology" OR "policy" OR "medical impact" OR "economic impact"). Furthermore, since not every journal is included in Scopus and the defined success factor categories are covering a broad spectrum of journal types, additionally also the online

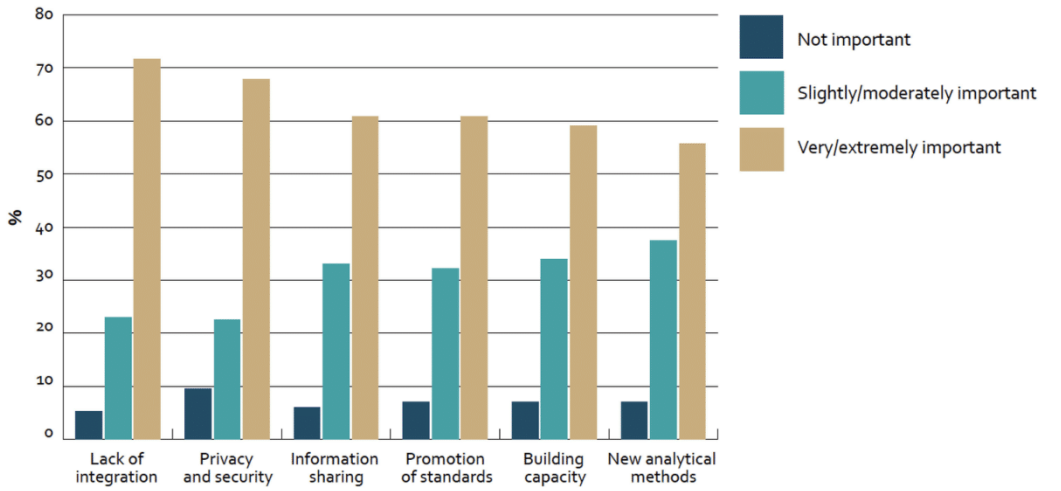


FIGURE 1 | Barriers to adoption of big data for health globally—survey of 125 countries by the WHO (14).

Category	Academic literature	Real world cases	
Sources	1) Scopus and 2) Online public access catalogue of the Technical University Munich	1) Scopus and 2) Online public access catalogue of the Technical University Munich	Press articles and website texts of Google advanced search results
Source prerequirements (all)	The content has been published in journal articles, in English or German language and the publishing age was not more than 5 years ago (latest from the year 2015)		Publicly available website texts Published between April 1, 2019 - April 1, 2020 (cases up to 2015)
Article inclusion critera (Left: At least one, Right: all)	A comprehensive description of an AI functionality An evaluation of the efficiency and outcomes Concrete implementation cases (e.g. on hospital, insurance, other level)		Naming of the AI provider Description of the technology Implementation location or institution
Article exclusion critera (at least one)	The title or abstract did not mention a topic related to AI The abstract did not contain a description of the AI application The full text did not elaborate on the implementation process		Inclusion criteria were not met Content from tweet and blogs

FIGURE 2 | Methodology for the identification of academic literature and real-world AI application cases in healthcare (authors).

public access catalog OPACplus of the Technical University Munich was used as second database.

The search term "Artificial Intelligence" has not been exchanged with other options like "Machine Learning" or "Neural Networks" as the term "Artificial Intelligence" has been used by far the most, according to the results of a Google Trend Analysis comparing the most frequently used search terms regarding AI in healthcare (15).

The following further inclusion and exclusion criteria were applied:

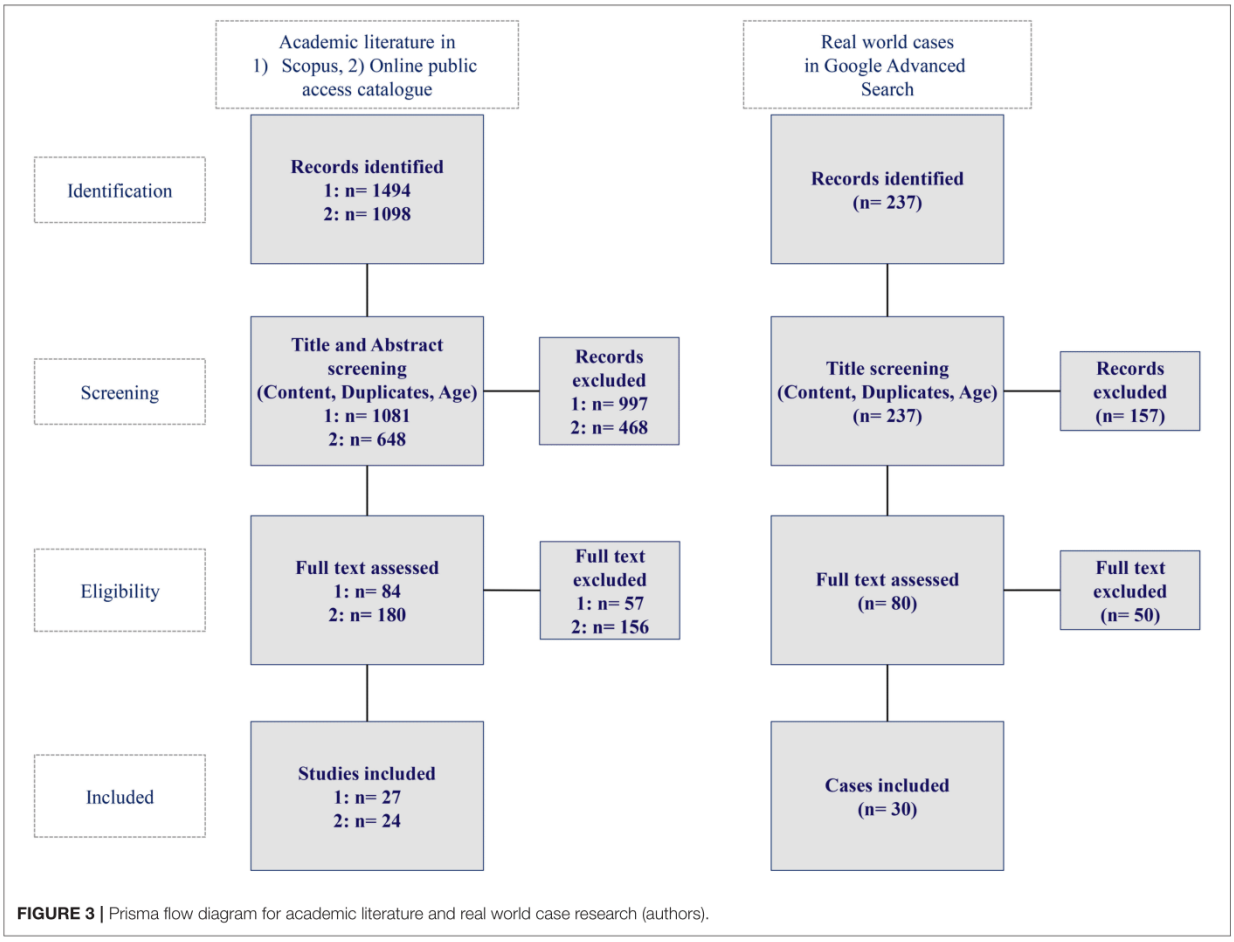


FIGURE 3 | Prisma flow diagram for academic literature and real world case research (authors).

1) The research is published in a journal article.
2) The publication is written in the English or German language.
3) The publication date was between the years 2015 and 2020.

Further, in terms of content, they were included if at least one of the following content-related criteria were met:

1) Comprehensive description of an AI application.
2) Evaluation of the efficiency and outcomes of an AI application.
3) Description of a concrete real-world AI application.

Subsequently, publications were excluded from the analysis if they met any of the following criteria:

1) The title or abstract did not mention a topic related to AI.
2) The abstract did not contain a description of the AI application.
3) The full text did not elaborate on the implementation process of an AI application.

The query returned 1,494 hits out of which 1,081 were published between the years 2015 and 2020 in Scopus as well as 1,098 hits out of which 648 were published in the mentioned time frame in OpacPlus. Applying the listed in- and exclusion criteria, 26 publications qualified as a basis for the academic literature review in Scopus and 24 publications in OpacPlus.

Real-World Cases

We identified real-world AI applications covered in academic literature using the abovementioned search approach. However, since only a small fraction of the practical AI implementation cases is covered by academic research, further real-world cases were identified through a Google-based advanced search for listings using the following search terms: "Artificial Intelligence," "Healthcare," and "Implementation." Google listings were included if they fulfilled all of the following criteria:

1) The AI implementation description was uploaded within the last year (i.e., results between 1 April 2019 and 1 April 2020), and the described practical case was not implemented before 2015.
2) The AI implementation is written in English or German language.

3) The AI implementation has a clear identification of the real-word AI application (i.e., cited the name of the AI provider, the technology, and the implementation location or institution).

AI applications originating from tweets or blogs were excluded. The query yielded 237 hits in the Google advanced search, of which 30 hits qualified as a basis for our analysis of real-world AI applications in healthcare. **Figure 2** depicts the methodology for the identification of academic literature and real-world AI application cases in healthcare while **Figure 3** shows the Prisma flow diagram.

RESULTS

Barriers to AI Implementation in Healthcare

Based on the academic literature and real-world case analysis, various barriers to AI implementation were identified. Given the need to access large amounts of data under strict privacy regulations and the dependence on managerial acceptance, it became evident that AI implementation needs to be tailored further to fit into existing healthcare routines. An illustrative example of how AI can be integrated into routine healthcare processes is shown in **Figure 4**.

As described above, the key identified barriers for AI implementation relate to the following fundamental issues: (1) non-privacy focused technological implementation, (2) shortcomings in current policy settings, and (3) the lack of medical and economic impact measurements. As comparison, in a framework about the success factors for AI implementation in the telecommunication industry in China, the author concluded that three success factors apply, namely the external environment, e.g., government involvement or vendor partnerships, organizational capabilities, e.g., managerial or technical skills, and innovation attributes, e.g., compatibility or relative advantage (16).

Our first barrier consists of major technological limitations that constrain AI implementation. Notably, access to medical data is commonly too fragmented and limited to Electronic Health Record (EHR) data and the existing data silos in the healthcare provider context do not enable complete access for AI applications (17). Furthermore, some data material, though available and accessible, may not be useable because of a lack of precise data requirements. For instance, in medical image analysis, edges of pictures may be unclearly defined, or high noise may inhibit the analyses (5). Further examples show that AI for breast, lung and liver cancer detection would require significantly enhanced data preprocessing and image processing or that in general a much more facilitated integration into existing workflows of EHRs is required to foster the use of clinical decision support systems (18, 19).

The second barrier shows, that there are major policy deficiencies that inhibit AI implementation. In numerous countries, it is neither clear who the regulatory authority for AI in healthcare is, nor how the ever-changing black box of AI will be assessed from a policy perspective (13). The General Data Protection Regulation (GDPR) in the EU and the Food and Drug Administration (FDA) regulations in the US for general data handling are very specific. However, there are no overarching policies, reporting standards, or recommendations concerning AI in healthcare. It could even be argued that no specific regulatory authority would be needed, as for example there is also no dedicated authority for decision support systems or treatment algorithms. Still, due to the potential risks of applying black box AI algorithms, it can be expected that clinicians will request clear and comprehensive regulations for increased application.

The third barrier in form of the lack of clinical and economic impact measurement further contributes to the low level of practical implementation. Although performance metrics on the outcomes of AI, such as levels of accuracy of preventive care or recommendations for therapeutic decisions are abundant, medical and economic benefits are often not measured, or the measurement approach is not clearly defined (4). The strategy, business models and, especially, reimbursement as a core element for AI application in healthcare are thus, often still unclear (3).

Success Factors for AI Implementation in Healthcare

Technological Implementation

The academic literature describes in detail the different technological categories of AI applications, ranging from natural language processing up to expert systems (20). In certain medical sectors, specific types of AI applications are more commonly applied, such as image analysis in radiology or dermatology (21). Most of the real-world AI application types face the challenge of combining practicality with privacy since they require complete data access.

This challenge could successfully be mitigated by several indication-focused practical cases of real-world AI applications. For instance, a "Persuasive Communication Tailoring" AI tool has been implemented to send motivational smoking cessation messages to adults. The machine learning version of the anti-smoking application significantly outperformed the prior rule-based system, and the algorithm was trained using data from messages, feedback databases, and user profiles (22). Another example is the pharmaceutical company MSD, which created an AI-driven communication channel based on the Facebook messenger for a chatbot about urgent matters in immune-oncology. The underlying conversational relationship between the physician and the chatbot is not bound to the data of EHRs, but is a stand-alone tool focused on the concrete problem-solution data access (23).

Furthermore, "privacy-by-design" technologies that aim to integrate privacy concepts in the design phase of an AI application, are increasingly being used (24). For example, at the institutional level, a health insurance system in Romania developed a GDPR compliant cloud-based AI application using a "SwarmESB-based" architecture with advanced data protection features. In the cloud infrastructure, multiple small entities are established, which possess one specific function for each task, such as ID copying, check of employment status, or retirement agency verification (25). Another reference case

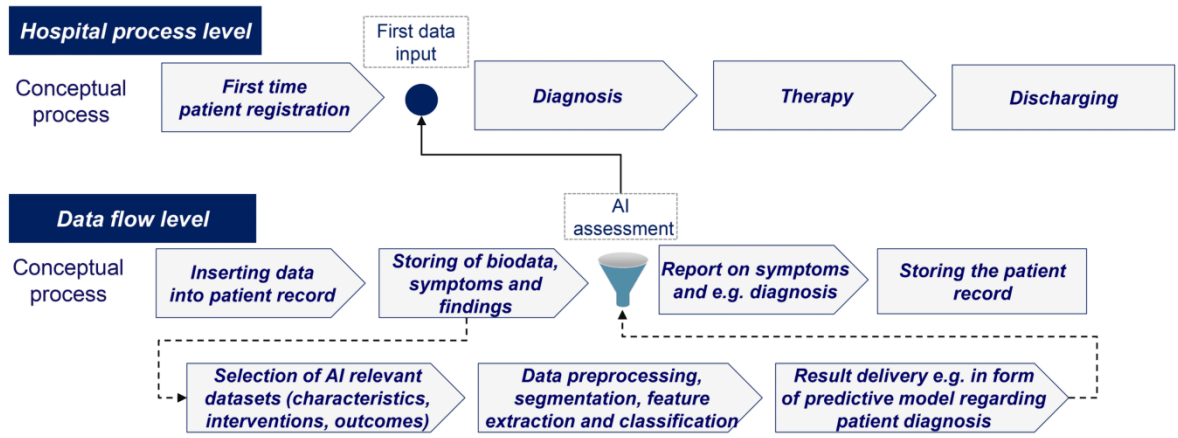


FIGURE 4 | AI integration into routine healthcare processes (authors).

TABLE 1 | Comparison of a US and an EU framework related to AI (29, 30).

Title	Proposed regulatory framework for modifications to Artificial Intelligence/Machine Learning (AI/ML)—Based Software as Medical Device (SaMD) Discussion paper and request for feedback (29)	Ethics guidelines for trustworthy AI (30)
Key content (excerpts)	- Establishment of quality systems and Good Machine Learning Practices (GMLP), including usage of only relevant data, the separation between training, tuning and test datasets or transparency of the output - Conduction of initial pre-market reviews to assure safety and effectiveness - Monitoring of the AI devices based on development, validation, and execution of algorithm changes such as "Algorithm Change Protocol" - Post-market real-world evidence performance reporting for maximized safety and effectiveness	- Independent high-level expert group on artificial intelligence set up by the European Commission/April 8, 2019 - Ethical principles as foundations of trustworthy AI (respect for human autonomy, prevention of harm, fairness, and explicability) - Seven key requirements of realizing reliable AI [(1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination, and fairness, (6) environmental and societal well-being and (7) accountability] - Assessing trustworthy AI (assessment list when developing, deploying or using AI systems)

for privacy by design is "FeatureCloud," a platform for the exchange of model parameters instead of raw data in a combined federated AI model (26). The technological implementation should consider the recommendations illustrated in **Table 1**.

Policy Setting

Previous publications cover a wide range of policy topics ranging from the dangers of so-called "black box" AI decisions to the paradigm shift from almost absolute protection of patient data to an economy of patient data sharing (27, 28). Nevertheless, there are almost no laws or standards that comprehensively regulate the use of AI in healthcare and there are significant geographical differences as shown in the US and EU propositions in **Table 2**.

The European Commission published also a risk-based legal adoption plan in the "White book for Artificial Intelligence" regarding training data, data storage, and human supervision (31).

In addition to the analysis of various regulatory frameworks, we also examined geographically independent policy factors.

First, it is expected, that AI, more generally, will evolve over several stages from the "Artificial Narrow Intelligence" to the "Artificial General Intelligence" up to "Artificial Super Intelligence," and the according use cases will develop from stand-alone problem-solving over strategic decision-making up to independent strategy execution (32). To support this evolution of AI, one should differentiate between a permissionless approach, where innovation can be tested and problems are solved as they occur, and a precautionary approach, where AI applications are banned from the beginning if they impose a distinct risk (33). Therefore when defining policy principles, one can build on a "form follows function" (permissionless) and a "first frame then function" (precautionary) approach, where the permissionless approach is less restrictive for AI implementations.

Second, it should be taken into account that AI decision-making processes are different from human decision-making

TABLE 2 | Key factors for AI technology development planning (authors).

Application scenario differentiation	Data processing structure definition	Privacy by design and product class setting
Indication-focused, e.g., smoking	Data access, e.g., EHR, wearables	AI technology implementation with a "privacy-by-design" structure
Institution-focused, e.g., health insurance	Data exchange pathways, e.g., connected vs. stand-alone	Compliance with medical product classification
Other	Data confidentiality measures, e.g., cloud infrastructure	Adaptability for changing AI regulatory requirements

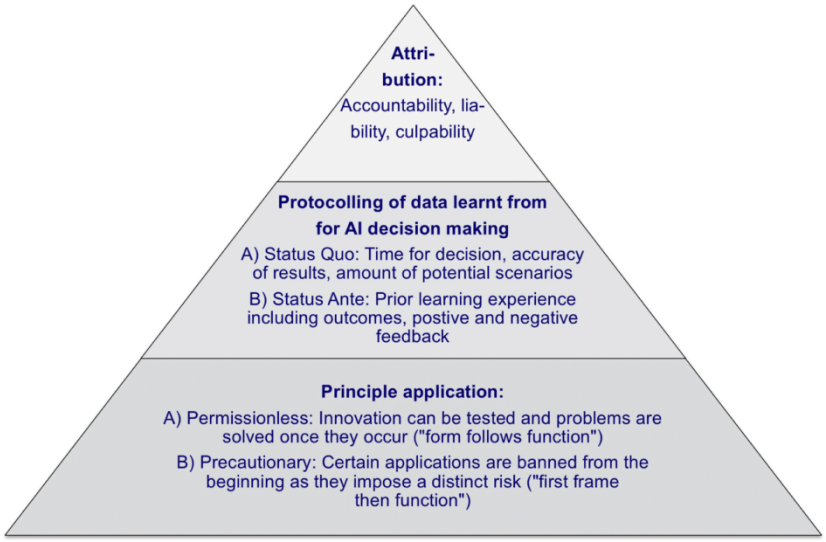


FIGURE 5 | Policy framework pyramid (authors).

processes. AI is able to infer answers more quickly and accurately and to consider a significantly larger number of scenarios simultaneously, and can, thus, reach different decision outcomes. Furthermore, AI learns from "wrong" behavior, and the severity of such adverse experiences and failures varies from case to case. Consequently, AI decision outcomes can also differ from that of human (34). To assess the reasoning process, protocols are required for the status ante, the status quo concerning the time taken for a decision, the number of scenarios considered, and the accuracy of the result obtained by AI.

Subsequently, the responsibilities of different stakeholders in AI processes should be addressed. For instance, in the real-world case of AI-based automatic robotic surgery, it is required to differentiate between accountability, liability, and culpability (35). A clear task differentiation is necessary, so that accountability can be clearly defined based on the process steps (e.g., x-ray image analysis), liability can be limited (e.g., manufacturer, operator, maintenance) and culpability can be exclusively attributed (e.g., an obligatory second human check of a decision obtained by an AI application).

A practical case of a real-world AI application that follows a permissionless approach is the collaboration between Philips, Salesforce, and Radboud University Medical Center. In this context, the involved parties extracted specific medical datasets, such as cancer research or COPD, and established the cloud software "HealthSuite" as a database on which patients and physicians can store health data for authorized access (36, 37). The case complies with the regulatory requirements via data protection measures, and available data is currently used by ca. 40 deep learning researchers focusing on various topics like medical image analysis (38).

In an environment of continually evolving national and international recommendations that lack concrete implementation guidance, a comprehensive policy is needed. An overview about a potential policy framework structure is displayed in **Figure 5**.

Medical and Economic Impact Measurement

AI strategy setting and implementation is a decision that is based on medical and economic decisions. Previous research has demonstrated that there are generally too few economic impact evaluations and, that many available ones lack critical components such as a net present value calculation or a comparison of alternative AI applications

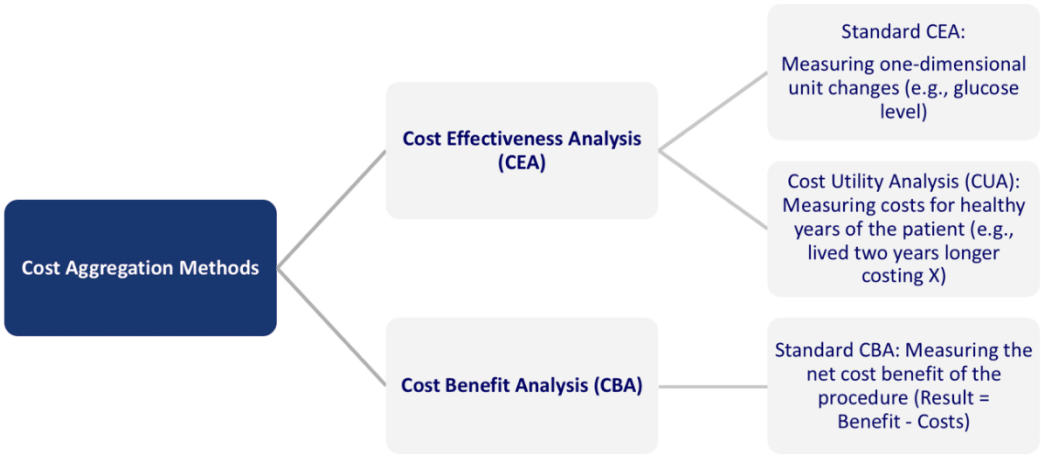


FIGURE 6 | Overview of cost aggregation methods in healthcare (39).

(15). This is particularly relevant in light of the meaningful investment volumes in the area of AI in healthcare, especially by large corporate entities, and the difficult economic impact measurement led to the application of industry-specific evaluation methods (40). Consequently, precise, accurate and internationally applicable medical and economic impact measurements are required.

The approaches to measure the outcomes of Digital Health, in general, and AI, in particular, can be classified into two categories: Cost Effectiveness Analysis (CEA) and Cost Benefit Analysis (CBA) (39). The first category can be further divided into standard CEA and Cost Utility Analysis (CUA). The CEA analysis refers to a cost comparison of a new vs. an old method, for example, regarding blood glucose measurement, wound size, or symptom-free days. In CUA, the outcome is measured in healthy years, for example, measured as quality-adjusted life years (QALYs). Specifically, QALYs provide an estimate of how many extra months or years of life, a person might gain by undergoing a specific treatment. Under a cost-minimization approach and the precondition of an equal medical outcome, different treatments can be compared. The difference between the approaches is that while the CBA can answer whether a new digital service is worthwhile, the CEA can answer the question of which of the alternative services is less costly to reach the equivalent outcome. **Figure 6** provides an overview of the different categories.

For a large-scale implementation of AI in healthcare and to qualify for reimbursement on a broad scale across insurance systems, the methods to measure medical and economic outcomes of AI applications have to follow standardized established procedures. The QALY analysis can be conducted based on different questionnaires to fulfill these requirements, and most studies follow the EQ-5D and the SF-6D format (see Appendix in **Supplementary Material**) (41).

Still, for existing studies, the quality of the respective impact measurements was often too low to produce reliable and valid results that could serve as basis for a well-founded decision about an AI implementation. This quality can be assessed through the so-called Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and Consolidated Health Economic Evaluation Reporting Standards (CHEERS) (42). The PRISMA guidelines should be used to identify the result report as a systematic review, meta-analysis, or both. The CHEERS criteria support the assessment, as the most common mistakes include items that are not reported in the study. This is of particular relevance as Iribarren et al. outlined that distinct items were missing in up to three-quarters of the publications about the impact of AI applications (43).

The medication selection and dosing company CURATE.AI reported in a cutting edge publication that, based on individually collected data, the adequate drug and respective dosing could be determined with limited side effects. An additional validation of the medical and economic impact of this solution using QALYs-based measurement, could significantly benefit the roll-out process with institutional payors like insurances and healthcare providers, even internationally (44).

Although further approaches such as comparator evaluation, multi-stakeholder analysis or organizational impact were discussed within prior research, a concrete approach with QALYs and quality criteria is needed immediately in order to generate short term results (45).

Recommendations for Increased Implementation of AI in the Success Factor Categories

As a starting point, concrete measures have been identified regarding the set-up of the technological infrastructure. First, it shall be tailored to the application segment, differentiating

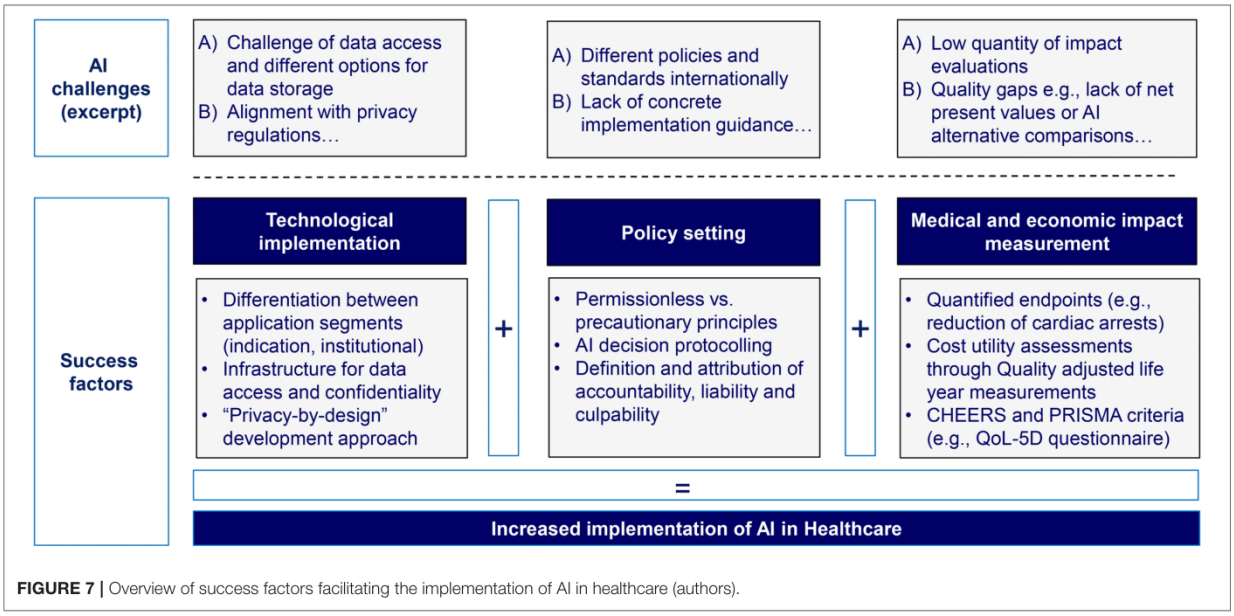


FIGURE 7 | Overview of success factors facilitating the implementation of AI in healthcare (authors).

between indication-focused and institutional-focused applications. Second, the data processing structure needs to focus on data access and exchange pathways as well as confidentiality measures. Third, a "privacy-by-design" approach shall be implemented and, the overall technological infrastructure should feature a high degree of adaptability in order to also be able to fulfill changing or upcoming regulatory requirements.

In addition to that, a clear and comprehensive AI policy framework is required. This should distinguish between permissionless and precautionary principles, namely between a risk-allowing "fast response" approach and a more cautious "safety first" approach. Furthermore, it should contain principles for AI decision-making protocolling in terms of the time taken for a decision, the number of scenarios considered, and the accuracy of the result obtained by AI to assess AI decisions ex-post. Finally, it must be possible to attribute accountability, liability, and culpability between the involved stakeholders, both human and AI, within the framework.

Furthermore, methodologies and metrics for assessing the medical and economic impact of AI applications must be refined and medical and economic impact assessments have to be intensified significantly. Such assessments should rely on cost-utility estimates and, in particular, on QALYs. Furthermore, we believe that it is indispensable that standardized quality criteria such as the CHEERS and PRISMA criteria (e.g., using a EuroQol-5D questionnaire) are applied so that the results can be evaluated not only by physicians, but also by institutional players.

An overview of the policy, technology, and impact measurement success factors is shown in **Figure 7**.

DISCUSSION

We systematically identified success factors that significantly facilitate the implementation of AI in healthcare based on existing academic research and real-world AI applications. In the following, we highlight some limitations.

First, an analysis of additional real-life AI application cases would have provided further relevant insights for the analysis. However, there is no open-access information or there are confidentiality clauses about technological features and economic impact independently of the databases used. Second, academic publications sometimes provide research results with a significant time delay due to the elaborated research process, such as data collection and analysis. Thus, research on very recent developments such as AI policy frameworks, frequently has not yet been conducted or published. Third, there are significant differences across categories. For example, an AI-supported medication adherence system and an AI-driven robotic surgery software are subject to different policy, technological and medical as well-economic impact measurement requirements. As a consequence, success factors will have to be weighted according to the Digital Health and AI conditions in each healthcare system.

Due to these limitations, several further success factors could not be included in the model, but should be a focus of further research and are here briefly discussed.

First, it is important to build trust and confidence among health professionals and patients. This can be seen, for example, in the discussions on COVID-19 tracking solutions. There are different approaches, e.g., for centralized or decentralized data storage, and in many countries intense political debates took place on data storage and tracking. Therefore, trust-building through open communication with easy to understand and

well-presented lines of argument is required, and this would also positively influence the acceptance of physicians as "gatekeepers" for AI.

Second, although the categories for "learned from," "training," "testing," or "validation" data are clearly defined in machine learning, in reality often processes are substantially changed or shortened e.g., no model validation takes place with independent datasets. This significantly affects the underlying specificity and sensitivity of AI solutions. Consequently, a clear set of recommended actions for each category would simplify the planning, programming and review processes. Furthermore, continuous reporting also facilitates ex-post verification processes due to the continuous AI learning process.

Third, the different levels of risk associated with AI need to be more clearly differentiated and for instance, the existing medical product classes in Europe could be tailored to AI solutions. Accordingly, AI solutions associated with higher risk will face more stringent regulation. Similarly, more stringent regulations will also be associated with higher costs for registration, documentation, and regulatory compliance. Thus, the market size must be reasonably large, and common market standards for AI risk levels should be established across all states in the US or all EU countries to provide still convincing arguments for AI development.

In summary, there are various barriers to AI implementation, which are likely to significantly have contributed to the considerable gap between the comprehensive and promising academic research on the high potential of artificial intelligence and the comparably low level of its actual practical implementation. Nevertheless, AI has already been applied in different healthcare sectors and is likely to have a meaningful impact on the entire healthcare industry. In particular, due to intense and steadily growing technological developments, current political developments, as well as the fast-evolving industry landscape, we expect a significant AI-driven transformation of healthcare delivery in the future.

The success factors identified in this paper (1) risk adjusted policy frame with clear accountability, liability, and culpability, (2) application scenario specific data processing structures on the basis of legally compliant and still practical privacy by design infrastructures, (3) comprehensive quantification of the medical and economic impact of AI on the basis of QALYs) can significantly facilitate the implementation of AI in routine healthcare processes. While some of the success factors require input from public institutions, private companies can use the success factor analysis already today to build and scale AI services e.g., through high-quality economic measurements and comprehensive technological planning regarding data processing and privacy-by-design structures. However, the current and upcoming success factors should not be perceived as stand-alone measures. Instead, they are strongly interlinked, and their effectiveness is, thus, interdependent to a certain extent. As such, future research needs to elaborate further on the interaction between optimal policy as well as technological, medical, and economic frameworks.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

JB was grateful for financial support from Villum Young Investigator Grant No. 13154. In addition, some of the work from JB was funded by H2020 projects RepoTrial and FeatureCloud (Nos. 777111 and 826078). Contributions by JP are funded by the Bavarian State Ministry of Science and the Arts within the framework of the Bavarian Research Institute for Digital Transformation (bidt).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2021.594971/full#supplementary-material

REFERENCES

1. German Bundestag. Available online at: https://dip21.bundestag.de/dip21/btd/19/134/1913438.pdf
2. Bundesanzeiger Verlag www.bundesanzeiger-verlag.de. Das Bundesgesetzblatt (BGBl.) - Bundesanzeiger Verlag. Bundesanzeiger Verlagsgesellschaft mbH. (2020). Available online at: https://www.bgbl.de/xaver/bgbl/text.xav?SID=&tf=xaver.component.Text_0&tocf=&qmf=&hlf=xaver.component.Hitlist_0&bk=bgbl&start=%2F%2F*%5B%40node_id%3D%27449138%27%5D&skin=pdf&tlevel=-2&nohist=1
3. Deo RC. Machine learning in medicine. *Circulation*. (2015) 132:1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593
4. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*. (2018) 286:800–9. doi: 10.1148/radiol.2017171920
5. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng*. (2018) 2:719–31. doi: 10.1038/s41551-018-0305-z
6. Triantafyllidis AK, Tsanas A. Applications of machine learning in real-life digital health interventions: review of the literature. *J Med Internet Res*. (2019) 21:e12286. doi: 10.2196/12286
7. Racine E, Boehlen W, Sample M. Healthcare uses of artificial intelligence: challenges and opportunities for growth. *Healthcare Manage Forum*. (2019) 32:272–5. doi: 10.1177/0840470419843831
8. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K, et al. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. (2019) 25:30–6. doi: 10.1038/s41591-018-0307-0
9. Alhashmi SFS, Salloum SA, Abdallah S. Critical success factors for implementing artificial intelligence (AI) projects in Dubai Government United Arab Emirates (UAE) Health Sector: applying the extended technology acceptance model (TAM). *Adv Intelligent Syst Comp*. (2020) 61:393–405. doi: 10.1007/978-3-030-31129-2_36
10. Lohr S. *IBM is Counting on Its Bet on Watson, and Paying Big Money For It*. The New York Times 2016. (2019). Available online at: https://

www.nytimes.com/2016/10/17/technology/ibm-is-counting-on-its-bet-on-watson-and-paying-big-money-for-it.html (accessed August 20, 2019).
11. Pifer R. *Amazon, Cerner Team Up on AI, machine Learning*. Healthcare Dive 2019. (2019). Available online at: https://www.healthcaredive.com/news/amazon-cerner-team-up-on-ai-machine-learning/559847/ (accessed August 20, 2019).
12. Lunden I. *Babylon Health Confirms $550M Raise at $2B+ Valuation to Expand its AI-Based Health Services*. TechCrunch 2019. (2019). Available online at: https://techcrunch.com/2019/08/02/babylon-health-confirms-550m-raise-to-expand-its-ai-based-health-services-to-the-us-and-asia/?guccounter=1&guce_referrer_us=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_cs=fHIZQ2SyG35__3eSBDOjoQ (accessed August 20, 2019).
13. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). *Implement Sci*. (2016) 11:146. doi: 10.1186/s13012-016-0510-7
14. World Health Organization. *Global Diffusion of EHealth: Making Universal Health Coverage Achievable, Report of the Third Global Survey on EHealth*. Geneva: World Health Organization (2017).
15. Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: systematic review. *J Med Internet Res*. (2020) 22:e16866. doi: 10.2196/16866
16. Chen H. Success Factors Impacting Artificial Intelligence Adoption — Perspective From the Telecom Industry in China. Doctor of Philosophy (PhD), Dissertation. Old Dominion University (2019).
17. Dimitrov DV. Medical internet of things and big data in healthcare. *Healthcare Inform Res*. (2016) 22:156. doi: 10.4258/hir.2016.22.3.156
18. Patel D, Shah Y, Thakkar N, Shah K, Shah M. Implementation of artificial intelligence techniques for cancer detection. *Augmented Human Res*. (2020) 5:6. doi: 10.1007/s41133-019-0024-3
19. Greenes RA, Bates DW, Kawamoto K, Middleton B, Osheroff J, Shahar Y. Clinical decision support models and frameworks: seeking to address research issues underlying implementation successes and failures. *J Biomed Inform*. (2018) 78:134–43. doi: 10.1016/j.jbi.2017.12.005
20. Hadley TD, Pettit RW, Malik T, Khoei AA, Salihu HM. Artificial intelligence in global health—a framework and strategy for adoption and sustainability. *Int J Mater Child Health AIDS*. (2020) 9:121–7. doi: 10.21106/ijma.296
21. Nakata N. Recent technical development of artificial intelligence for diagnostic medical imaging. *Japan J Radiol*. (2019) 37:103–8. doi: 10.1007/s11604-018-0804-6
22. Sadasivam RS, Borglund EM, Adams R, Marlin BM, Houston TK. Impact of a collective intelligence tailored messaging system on smoking cessation: the perspect randomized experiment. *J Med Internet Res*. (2016) 18:e285. doi: 10.2196/jmir.6465
23. Murtha J. *How the 1st Pharma Chat Bot Came to Life*. Inside Digital Health. (2017). Available online at: https://www.idigitalhealth.com/news/how-the-1st-pharma-chat-bot-came-to-life (accessed July 18, 2019).
24. Semantha FH, Azam S, Yeo KC, Shanmugam B. A systematic literature review on privacy by design in the healthcare sector. *Electronics*. (2020) 9:452. doi: 10.3390/electronics9030452
25. Calancea CG, Alboaie L, Panu A. A SwarmESB based architecture for an European healthcare insurance system in compliance with GDPR. *Parallel Distribut Comp Appl Technol*. (2019) 261–7. doi: 10.1007/978-981-13-5907-1_27
26. European Commission: Featurecloud. https://featurecloud.eu/wp-content/uploads/2019/04/CORDIS_project_220225_en-1.pdf.
27. Sullivan HR, Schweikart SJ. Are current tort liability doctrines adequate for addressing injury caused by AI? *JAMA J Ethics*. (2019) 21:E160–6. doi: 10.1001/amajethics.2019.160
28. Jaremko JL, Azar M, Bromwich R, Lum A, Alicia Cheong LH, Gibert M, et al. Canadian association of radiologists white paper on ethical and legal issues related to artificial intelligence in radiology. *Can Assoc Radiol J*. (2019) 70:107–18. doi: 10.1016/j.carj.2019.03.001
29. Center for Devices & Radiological Health. *Artificial Intelligence and Machine Learning in Software*. Montgomery, MD: U.S. Food and Drug Administration (2019). Available online at: https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf
30. Anonymous. Ethics guidelines for trustworthy AI - Digital Single Market - European Commission. *Digital Single Market – European Commission*. (2019). Available online at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (accessed August 14, 2020).
31. European Commission. (2020). *WEISSBUCH Zur Künstlichen Intelligenz - ein europäisches Konzept für Exzellenz und Vertrauen, COM/2020/65 final*. (2020). Available online at: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
32. Wolff J, Keck A, König A, Graf-Vlachy L, Menacher J. *Künstliche Intelligenz: Strategische Herausforderungen für etablierte Unternehmen*. Wiesbaden: Handbuch Industrie 4.0 und Digitale Transformation (2019). doi: 10.1007/978-3-658-24576-4_21
33. Thierer AD, O'Sullivan A, Russell R. *Artificial Intelligence and Public Policy*. Arlington, VA: SSRN Electronic Journal (2017). doi: 10.2139/ssrn.3046799
34. Pesapane F, Volonté C, Codari M, Sardanelli F. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. *Insights Imaging*. (2018) 9:745–53. doi: 10.1007/s13244-018-0645-y
35. O'Sullivan S, Nevejans N, Allen C, Blyth A, Leonard S, Pagallo U, et al. Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. *Int J Med Robot Computer Assisted Surg*. (2019) 15:e1968. doi: 10.1002/rcs.1968
36. HITC Staff. *Salesforce Launches Salesforce Health Cloud to Build Patient Relationships Beyond EHRs*. (2015). Available online at: https://hitconsultant.net/2015/09/03/salesforce-launches-salesforce-health1-to-build-patient-relationships-beyond-ehrs/ (accessed August 14, 2020).
37. Philips. *Philips and Dutch Radboud University Medical Center Introduce First Diabetes Prototype App With Integrated Online Community to Empower Patients and Enhance Continuity of Care*. (2015). Available online at: https://www.prnewswire.com/news-releases/philips-and-dutch-radboud-university-medical-center-introduce-first-diabetes-prototype-app-with-integrated-online-community-to-empower-patients-and-enhance-continuity-of-care-300141815.html (accessed July 18, 2019).
38. Deep learning for Prostate MRI Intervention. *Diagnostic Image Analysis Group*. Available online at: http://diag.radboudimaging.nl/index.php/Deep_learning_for_Prostate_MRI_Intervention
39. Bergmo TS. How to measure costs and benefits of eHealth interventions: an overview of methods and frameworks. *J Med Internet Res*. (2015) 17:e254. doi: 10.2196/jmir.4521
40. Wolff J, Kocher C, Menacher J, Keck A. Too big to ignore? how to define the enterprise value of Ai and big data firms. *Int J Strategic Manage*. (2018) 18:13–24. doi: 10.18374/IJSM181.2
41. Bergmo TS. Using QALYs in telehealth evaluations: a systematic review of methodology and transparency. *BMC Health Serv Res*. (2014) 14:332. doi: 10.1186/1472-6963-14-332
42. Website. See http://prisma-statement.org/documents/PRISMA%202009%20checklist.pdf for further information (accessed July 18, 2019).
43. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. *PLoS ONE*. (2017) 12:e0170581. doi: 10.1371/journal.pone.0170581
44. Blasiak A, Khong J, Kee T. CURATE.AI: optimizing personalized medicine with artificial intelligence. *SLAS Technol*. (2020) 25:95–105. doi: 10.1177/2472630319890316
45. Kolasa K, Kozinski G. How to value digital health interventions? A systematic literature review. *Int J Environ Res Public Health*. (2020) 17:2119. doi: 10.3390/ijerph17062119

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Paper B: The Impact of Artificial Intelligence on the Healthcare Economy

The paper entitled “The Impact of Artificial Intelligence on the Healthcare Economy” identifies which trends and circumstances as well as policy frameworks positively influence AI implementation in healthcare. The paper combines a non-systematic literature review and market research with two assessments of real-world facilitator frameworks for AI implementation from the policy and technology area, respectively: The German Digital Healthcare Act and the FeatureCloud platform. Furthermore, the paper shows that the interplay of four key drivers is likely to trigger a transformation of the healthcare market through AI.

Title	**The impact of Artificial Intelligence on the healthcare economy**
Research question	What is the impact of AI on the healthcare market today and in the future?
Background	Although various market evaluations clearly point towards a meaningful impact of AI, the actual real-world AI implementations stem from a limited number of regions and healthcare segments
Contribution	Presents two concrete facilitator frameworks from the policy and technology areas (e.g., reflecting on the importance of regulation, evidence generation and reimbursement to scale DH and AI in healthcare) and describes four key drivers that support AI implementation in healthcare
Method	Non-systematic review of academic literature and market research about the impact of AI on the healthcare economy and analysis of existing AI-related facilitator frameworks

Take-away	Four interlinked developments (i.e. demand, technological improvements, acceptance, global corona pandemic) and existing facilitator frameworks influence AI growth. In order to support this transformation, governments, scientists and practitioners need to demonstrate the impact of AI, leverage privacy-preserving technologies, and support policy formation for data access, evidence generation and reimbursement guidelines.

Table 2: Overview about the publication “The impact of AI on the healthcare economy”

Contribution of the doctoral candidate: First authorship including the planning of the publication structure, a non-systematic review of academic literature and conduction of market research as well as discussions with the chair members about potential reference cases like the FeatureCloud platform and the German Digital Healthcare Act. Manuscript: Writing, review and editing.

The Impact of Artificial Intelligence on the Healthcare Economy

[Author Name(s)]: Justus R. Wolff MSc[1,3], Julia Menacher PhD[3], Josch K. Pauling PhD[1,2], Andreas J. Keck MD[3], Jan Baumbach PhD[1]

[Author Affiliation(s)]:

1 Chair of Experimental Bioinformatics, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Maximus-von-Imhof-Forum 3, 85354 Freising, Germany

2 Computational Lipidomics research group LipiTUM, Chair of Experimental Bioinformatics, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Maximus-von-Imhof-Forum 3, 85354 Freising, Germany

3 Syte – Strategy Institute for Digital Health, Hohe Bleichen 8, 20354 Hamburg, Germany

[Corresponding Author]: Justus Wolff, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Maximus-von-Imhof-Forum 3, Freising, 85354, Germany, Telephone: +49 (0) 40 5555360, Email: justus.wolff@syte-institute.com

[Non-print Items]

Abstract: Artificial Intelligence (“AI”) AI in healthcare has become a billion-dollar market and will become even more important as there is an urgent need to limit healthcare spending globally and, often, to simultaneously cope with shortage of labor. At the same time, AI applications benefit from technological improvements, increased acceptance levels and a substantial need for Digital Health structures as a consequence of the corona pandemic. Nonetheless, real-world application is still in its early phase and can predominantly be found in only a few regions and healthcare segments. To close the gap between the considerable amount of promising academic research results on AI and the comparatively small number of actual real-world applications, in this chapter, we elaborated on improved economic impact measurement, reference cases for reimbursement and pathways for privacy-aware data access. With these measures, an increased translation from academia to real-world application and, thus, also economic expansion of AI in healthcare shall be facilitated.

Keywords: Artificial Intelligence, Artificial Intelligence economy, Medical impact measurement, Economic impact measurement, Reimbursement, Data access

Acknowledgements: Contributions by Josch K. Pauling are funded by the Bavarian State Ministry of Science and the Arts in a framework coordinated by the Bavarian Research Institute for Digital Transformation (bidt). Jan Baumbach is grateful for financial support from H2020 project FeatureCloud (nr. 826078) and BMBF grant Sys_Care (nr. 5103388) as well as his VILLUM Young Investigator grant nr. 13154.

[Chapter Starts Here]

1. Growth history of Artificial Intelligence in healthcare – more than 4bn USD in funding in 2019

Artificial Intelligence ("AI") in healthcare is comprehensively analyzed in academic research and real-world applications are also increasingly covered in the media landscape. Although precise industry growth figures are not available, it can be assumed that the AI in healthcare market has grown significantly over the last few years from approx. USD 0.6bn in 2014 (Accenture 2017) to approx. USD 0.7bn in 2016 (DigitalMR 2017) to approx. USD 1.3bn in 2018 (Global Market Insights 2019), implying a CAGR of 21.3% over this time period. In other words, the AI in healthcare market has more than doubled and has become a billion-dollar market within a few years only.

Other economic figures such as the significant increase in the number of respective start-ups over the last few years (Grand View Research 2019) and the significant upward trend in funding activities also clearly reveal the promising prospects of AI in healthcare. As to the latter, not only has healthcare consistently represented the top industry for AI investment deals over the last years (AI in Healthcare 2017; Pifer 2020). Also, while AI start-ups in healthcare raised funding of approx. USD 2.1bn across 323 deals between 2012 to mid-2017 (AI in Healthcare 2017), they raised almost double this amount, namely approx. USD 4bn, across 367 deals in 2019 alone (Pifer 2020). Even during the global corona pandemic, which considerably slowed down financing activities, almost 80 AI start-ups in healthcare raised approx. USD 1bn in the first quarter of 2020 (CB Insights 2020). Furthermore, tech giants such as IBM and Amazon have gotten increasingly involved into developing AI algorithms for healthcare applications (Davenport and Kalakota 2019).

Despite this economic growth, we can currently observe relatively few actual implementations in terms of real-world AI applications in routine healthcare processes (He et al. 2019; Wolff et al. 2020). In particular, while enormous research efforts have been made to produce and analyze algorithms and AI-powered solutions, only a fraction of the respective results has so far been implemented in the healthcare systems (Lindsell, Stead, and Johnson 2020). Panch, Mattie, and Celi even went as far as to postulate that prominently featured algorithms "are in fact not, for the most part, executable at the frontlines of clinical practice" (Panch, Mattie, and Celi 2019).

Also, the real-world AI applications to date predominantly stem from a limited number of healthcare segments. In global healthcare research, AI applications have been classified into categories including diagnosis, patient morbidity, disease outbreak prediction and surveillance, as well as health policy and planning (Schwalbe and Wahl 2020). Various market analyses show that AI can mainly be encountered in the context of inpatient care and hospital workflow management, medical imaging and diagnostics, patient data and risk analysis, virtual care, and drug discovery (MarketsAndMarkets 2020). For instance, studying over 90 AI start-ups in healthcare, CB Insights revealed that the areas medical imaging and diagnostics as well as drug discovery jointly accounted for 52% of the sample start-ups and 45% of disclosed funding (CB Insights 2019). In a similar vein, a recent scientific review showed that AI and,

especially, deep learning, has attracted particular attention in medical imaging and diagnostics (Nakata 2019).

A similar pattern can be observed when analyzing the market across regions. To date, North America accounts by far for the largest market share (Market Research Future 2019; Grand View Research 2019), whereof approx. 90% of revenues stem from the U.S. (MarketWatch 2020). The fact that economic growth so far mainly has been driven by one country leaves room for tremendous growth in other regions. While even in the U.S., further exponential growth of AI in healthcare is expected (ReportLinker 2020), market analysts largely agree that the APAC region is likely to become the fastest growing region in the coming years (Meticulous Research 2019; Market Research Future 2019; Grand View Research 2019). There is also considerable potential for AI application in low- and middle-income regions as AI-powered health interventions can lead to better health outcomes, bearing in mind the continued severe lack of healthcare resources in these regions (Carrillo-Larco et al. 2020).

This moderate level of AI penetration in healthcare will, however, soon be a matter of the past. The global AI in healthcare market is anticipated to grow at a CAGR of c. 40% over the next years (Meticulous Research 2019; Mordor Intelligence 2020; Grand View Research 2019), where the average CAGR may even exceed 50% from 2019 to 2030 (Market Research Future 2019). Also, market studies suggest that, already by 2025, a growing number of AI applications will especially be found in the areas of chronic disease management, drug discovery, delivery of healthcare services, and detection of disease (Market Study Report 2019).

Considering the bigger picture, AI will also play a key role in the transformation of the healthcare market in its very essence. In particular, the healthcare industry is expected to experience a shift from fragmented and location-driven healthcare to consumer-oriented, virtual and convenient healthcare, entailing significant changes in value chains and business models and affecting all actors involved (Cognizant 2019).

INFO BOX 1 (ONLY TEXT) – TO BE FORMATTED AS INFO BOX:

Taken together, significant research efforts have been made to produce algorithms and AI-powered software and hardware products, and researchers have devoted high attention to this topic demonstrating promising results on the potential of AI to revolutionize healthcare. Nonetheless, the current level of practical implementation and economic impact measurement of AI in healthcare is still relatively low (Wolff et al. 2020). Furthermore, at this stage, only a limited number of healthcare segments and regions dominate the market and there is considerable global growth potential over the next years.

2. Drivers of Artificial Intelligence in healthcare – Demand, technology, acceptance and corona pandemic

While the above-mentioned economic factors such as extensive start-up growth as well as investment activities will continue to make a significant contribution to economic growth in the future, the interplay of four key drivers is likely to trigger a transformation of the healthcare market through AI.

First, there is a significant demand for AI in healthcare. In particular, global healthcare spending has not only grown steadily, but also generally represents a considerable burden for economies worldwide. In 2017, it amounted to approx. USD 7.8tn, namely 10% of global GDP and USD 1,080 per capita, up from approx. USD 7.6tn in 2016 (World Health Organization 2019). In the U.S., healthcare spending even accounted for 18% of GDP in 2018 (Hartman et al. 2020).

However, developments such as the aging population as well as the significant increase in chronic diseases actually require an expansion of both the access to and the scope of healthcare services. As to the former development, for instance, in Japan, 38% of the population is expected to be aged 65 or older by 2050 (Kumar 2019). Thus, for example, oncology already today represents one of the major application fields of AI (Agrawal and Prabakaran 2020). As to the second development, it was projected that chronic diseases will account for almost 75% of global deaths by 2020 (World Health Organization 2002).

At the same time, medical facilities are facing a shortage of labor, causing delays and pressure on existing medical staff. For instance, the NHS in the U.K. reported almost 100k job vacancies in 2018 (Kumar 2019). Taken together, there is considerable pressure to reduce healthcare spending without limiting actual care, which creates a significant demand for AI applications. For instance, it was forecasted that in the U.S., the top ten AI applications can potentially create USD 150bn in annual savings and that AI could meet 20% of unmet clinical demand by 2026 (Accenture 2017). In addition, U.S. health insurers could save approx. USD 7bn by implementing AI, among others, in customer interaction (Accenture 2018b).

Overall, AI applications come along with significant cost saving potential, among others, by optimizing revenue cycle management (Accenture 2018b; RevCycleIntelligence 2020), hastening drug discovery, decreasing non-adherence costs and shortening recovery time (Garbuio and Lin 2019). According to Frost & Sullivan, AI could reduce treatment costs by up to 50% while improving patient outcomes by 30% to 40% (Ahuja 2019).

As a second driver, AI in healthcare greatly benefits from improvements in technological possibilities. These have been achieved over years, and sometimes decades, e.g., related to data storage capacities and processing power, such as cloud computing (Mordor Intelligence 2020), as well as related to AI methods (Aguis 2019). In the healthcare sector, the need for, but lack of comprehensive data access has proven to be a main obstacle, especially as a consequence of high fragmentation as well as data protection regulations and data security concerns (Panch, Mattie, and Celi 2019).

Recently, however, a wide range of promising initiatives that enhance access to comprehensive data could be observed. In particular, numerous aggregate health datasets that can be used by AI researchers and developers for research and programming purposes have been produced such as by the Radboud University in the Netherlands, (Radboudumc 2019) or the U.S. Department of Veterans Affairs (U.S. Department of Veterans Affairs 2014). Also, it is generally expected that the amount of health data will grow remarkably, namely at an average CAGR of 36% between 2018-2025 according to the IDC Data Age 2025 report (Reinsel, Gantz, and Rydning 2018), paving the way for further AI research and development.

As a third driver, for the first time, there are signs of noteworthy acceptance of AI in healthcare, both by key decision-makers such as governmental institutions as well as by patients. As to the former group, AI in healthcare policy frameworks and initiatives have gained considerable momentum in recent years, both on national and international level. Respective activities include, but are not limited to, the definition of abstract AI strategies such as the German Federal Government's "AI strategy" (European Commission 2019), the definition of sector-specific policy frameworks such the U.K.'s "Code of conduct for data-driven health and care technology" (U. K. Department of Health & Social Care 2019) or the provision of budgets for R&D activities for AI in healthcare such as in the U.S. (OECD 2019).

In addition, governmental institutions have adopted a wide range of targeted legislative measures that aim to facilitate real-world AI applications in healthcare. One example is Germany's Digital Healthcare Act, which paves the way for the reimbursement of Digital Health ("DH") applications for over 70mn insured citizens (see section 4 for a detailed description) (Matthies 2019). Similarly, the U.S. FDA is working with high priority on a regulatory framework to evaluate AI-powered software as a medical device (U.S. Food and Drug Administration 2020).

In a similar vein, we can observe first meaningful acceptance of AI in healthcare among (potential) patients. For instance, a consumer survey showed that, as early as 2018, consumers were on average more likely than not to use a wide range of AI-powered healthcare services such as to get medical advice or a diagnosis of their symptoms (Accenture 2018a).

As the final driver, there is an accelerated implementation of AI in healthcare settings as a consequence of the corona pandemic. While the long-standing digital consumer trend has not been able to trigger a break-through of AI in healthcare until today, this development has just recently gained new sentiment due to the pandemic. In particular, numerous recent research results revealed that the pandemic has considerably increased people's comfort with AI applications for the purpose of safety and health.

With respect to AI applications applied directly in patient interaction, for instance, a survey uncovered that 33% of respondents have become more comfortable with the idea of AI helping doctors and nurses and that 30% of respondents have become more comfortable with the idea of relying on AI-powered tools such as chatbots to determine their medical needs (Interactions 2020).

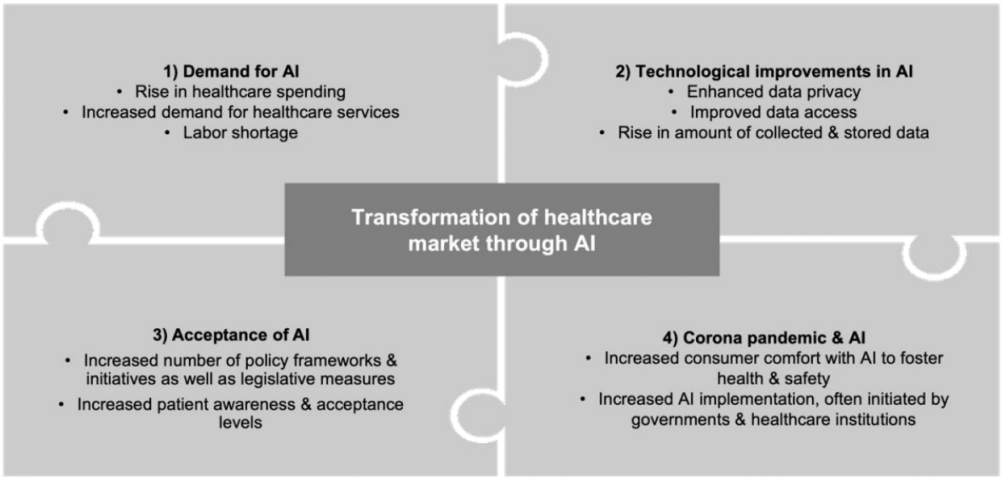
This pattern becomes even clearer when looking at a more aggregate societal level. In particular, initiatives by governments and healthcare institutions have been significantly augmented since the hit of the crisis. The application areas comprise, among others, vaccine development as well as screening, analyzing and predicting disease patterns (Vaishya et al.

2020). First results regarding the diagnosis of COVID-19 via data mining and machine learning ("ML") algorithms revealed that with additional sources, e.g., sensor technologies in outdoor scenarios, governmental COVID-19 measures can be supported (Albahri et al. 2020) A further concrete example of AI for COVID-19 is the drug repurposing platform Covex, which integrates experimentally validated virus-human protein interactions, human protein-protein interactions and drug-target interactions (Sadegh et al. 2020).

This widespread use of AI in connection with the corona pandemic can be observed globally. For example, in South Korea, AI is used to trace the spread of COVID-19 via mobile data (Hunt 2020). In Israel, AI-powered contact tracing algorithms are used to send text alerts to citizens who have been near someone who was diagnosed with COVID-19. Applications of AI focused on the pandemic are early detection and diagnosis of the infection, monitoring the treatment, contact tracing of the individuals, projection of cases and mortality, development of drugs and vaccines, reducing the workload of healthcare workers and prevention of the disease (Vaishya et al. 2020).

INFO BOX 2 (TEXT + FIGURE 1) – TO BE FORMATTED AS INFO BOX:

Text: The interplay of four key drivers is likely to trigger a transformation of the healthcare market through AI in the short term, as illustrated in Fig. 1. First, there is an urgent need for economies worldwide to limit the rise in healthcare spending while responding to an increased demand for healthcare services, which in many cases is further impeded by labor shortages. Second, significant technological improvements in recent years enable simplified and scaled AI implementation today more than ever. Third, for the first time, noteworthy awareness and acceptance levels can be observed for AI applications in healthcare. Finally, the corona pandemic has significantly increased the need for DH structures, more generally, and AI, in particular.



Caption: Transformation of the healthcare market through AI

Credit: Authors

3. Impact of Artificial Intelligence on the healthcare economy – quantity and quality of measurements still insufficient

AI can significantly impact the healthcare economy, and benefits have become apparent in the preventive, diagnostic and therapeutic contexts.

The economic significance of AI in healthcare becomes visible in the start-up and industry landscape, and there are already several young companies that have positioned themselves as market niche leaders, such as IDx-DR, now called ”Digital Diagnostics”, for diabetic retinopathy screening in the U.S. or mySugr for diabetes management in Europe. Furthermore, also large corporates and tech giants (Meticulous Research 2019), among others, IBM Watson (MarketsAndMarkets 2020) and Amazon and their respective activities clearly demonstrate the importance of AI for the healthcare economy.

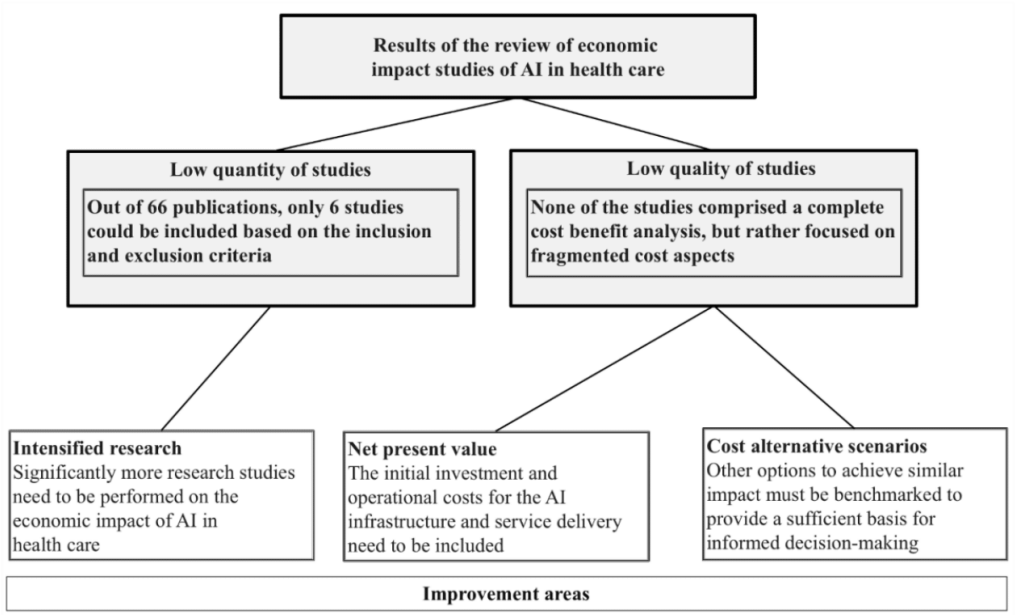
AI also has a significant impact on the health insurance industry (McKinsey & Company 2017). In particular, AI applications entail changes in the product portfolio and insurance coverage, in the use of data, e.g., regarding price discrimination, and in the necessity of collaborations. The fact that insurers play a key role for the scalability of AI applications may lead to a stronger interlinkage between healthcare providers and insurers, which is particularly challenging for traditional insurers without access to tele-medical infrastructures as well as comparably less dynamic business models.

Overall, there are different impact areas of AI such as social change, e.g., the interaction between humans and AI, a disruption in the competitive environment, e.g., advantages through data access and algorithms, or new business models, e.g., analytics for health data.

Nonetheless, for a sustainable and large-scale economic impact of AI, standardized integration into healthcare processes and according reimbursement structures in healthcare systems are required. The lack of these structures is a key inhibitor of real-world AI applications and hampers respective implementation and scalability.

There is an underlying complexity of impact validation for these implementation processes as many aspects need to be considered when measuring the value of AI in healthcare. Examples for parameters include safety, clinical effectiveness, usability as well as a set of ethical and legal aspects (Kolasa and Kozinski 2020). Out of this broad spectrum, medical and economic benefits of AI are particularly crucial to justify the human resource and financial investments that are typically necessary to implement AI.

However, in the past, the medical and economic impact of AI has not been measured to a sufficient extent and quality. The significant shortcomings in respective assessments lead to a plethora of performance measures, such as the accuracy of a solution, without providing a framework for strategic decision-making. In particular, prior research showed that there are too few and too incomprehensive impact measurements for AI applications in healthcare (Wolff et al. 2020). Key takeaways of this research were that in addition to intensified research on medical and economic impact, a net present value calculation and cost alternative scenarios are required, as displayed in Fig. 2.



Caption: Result of the review of economic impact studies of AI in healthcare

Credit: Wolff et al., 2020, p. e16866 (figure 2)

Standardized economic impact measurement requires structured data input about the result of the digital intervention and this shall be collected based on quality-adjusted life years ("QALYs") while applying both the Consolidated Health Economic Evaluation Reporting Standards ("CHEERS") statement and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses ("PRISMA").

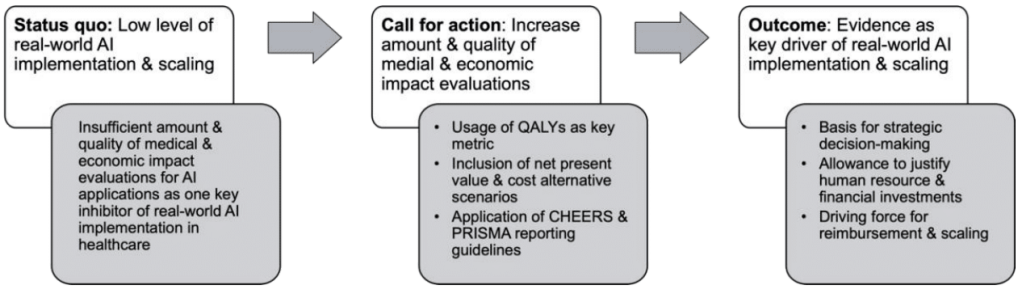
An example of the result of a digital intervention would be that an AI-powered software enables three additional years of healthy living, costing USD 20k per year, based on data collection via an EQ-5D questionnaire. The questionnaire could be split into "Mobility, Self Care, Usual Activities, Pain/Discomfort, Anxiety/Depression" criteria, scoring each criterion on a scale from 0 to 100. The sum of costs needs to encompass the initial investment and operating costs and, if possible, an alternative solution targeting the same result should also be assessed.

Such an assessment could encompass the following steps:

- Collection of patient data based on EQ-5D QALY values and deduction of additional healthy living years gained via the AI solution
- Cost calculation considering initial investment and operating costs of the AI solution
- Comparison of the AI solution with existing treatment methods and potential alternative solutions (e.g., individualized treatment based on genetic profile)
- Overall comparison of costs for the additional healthy living years gained via the AI solution as compared to the existing treatment methods and/or potential alternative solutions
- Decision about implementation and tracking of results

INFO BOX 3 (TEXT + FIGURE 3) – TO BE FORMATTED AS INFO BOX:

Text: In order to support an AI solution, in particular, and overall AI implementation, more generally, it can be summarized that the impact should be defined based on QALYs while applying the CHEERS and PRISMA quality criteria. In addition, the initial investment and operational costs for the AI solution need to be considered and alternatives to achieve a similar impact must be evaluated to allow for a comprehensive comparison as a basis for strategic decision-making. The respective line of reasoning is illustrated in Fig. 3.



Caption: Status quo, call for action and outcome of economic impact measurement of AI in healthcare

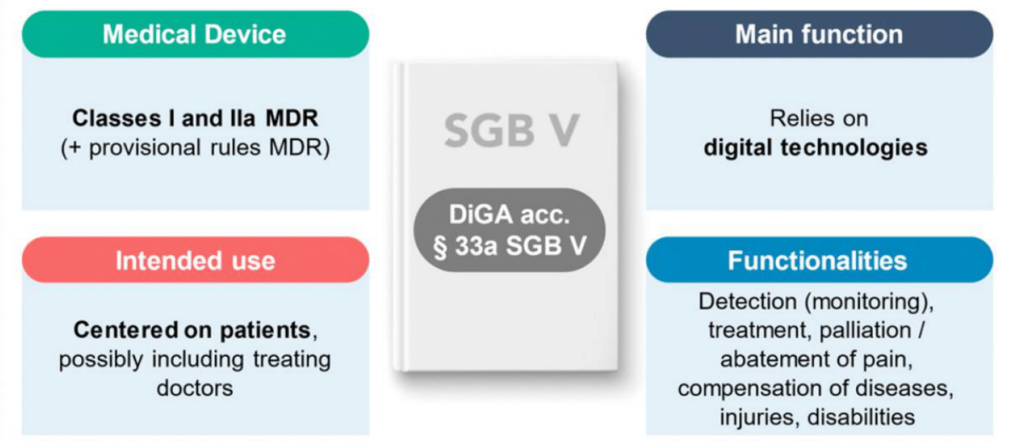
Credit: Authors

4. Selected frameworks to demonstrate facilitation of Artificial Intelligence implementation – pathways to reimbursement and data access

The economic growth of AI in healthcare can be fostered through various measures and, in the following, several international facilitator frameworks are outlined.

One example is the reimbursement scheme established by the German government via the Digital Healthcare Act (in German: "Digitales Versorgungsgesetz"), a comprehensive framework effective since May 2020 to foster DH structures. It aims at an easy use of tele-consultation, access to secure healthcare data networks, and apps on prescription (German Federal Ministry of Health 2019). The last aspect, in particular, will lead to an increased use of AI in healthcare, as every physician in Germany is now entitled to prescribe DH solutions in the same way as a remedy, and the digital apps are reimbursed by the public health insurance system, covering 90% of the population, i.e., over 70mn citizens.

The law targets DH applications, so-called DiGAs (abbreviation to German term "Digitale Gesundheitsanwendung"), which are then listed on an official DiGA list, so that doctors can prescribe them. The definition for such a DiGA is provided in Fig. 4.



Caption: DiGA definition as by the German government

Credit: Health Innovation Hub, 2021, p. 7

In order to prescribe a DiGA, a comprehensive application submission, validation and reimbursement scheme has been developed. All solutions must fulfil the general requirements of safety, quality, functionality, privacy and data security as well as demonstrate a so-called "positive care effect". The latter consists of a medical benefit and/or structural and procedural effects. For all solutions, there are two application forms, differentiating between DH solutions with priorly collected data for final listing and solutions without priorly collected data and, thus, for provisional listing. Both application systems are administered by the Federal Institute for

Drugs and Medical Devices (in German: “Bundesinstitut für Arzneimittel und Medizinprodukte” or “BfArM”).

The first system enables application submissions for DH solutions with an already existing evidence data basis with regard to a medical benefit and/or structural and procedural effects. This evidence is then assessed by the BfArM in a 3-month time period before making a decision about final listing into the reimbursement system. In case acceptance is granted, the DH solution is reimbursed during a 12-month time period with a predefined pricing scheme which is agreed upon between the manufacturer and statutory health insurance (“SHI”) associations. After the 12-month period, the price is renegotiated and continues to be placed on the market with potential annual pricing adaptations.

The second alternative features a key difference, namely in the form of a “probation period” in the application process. The respective process differs from the first system as only limited prior evidence of a medical benefit and/or structural and procedural effects of the DH solution has been generated upfront. Thus, after the 3-month application submission verification, the solution is listed as DiGA under probation. In this trial period, a plausible hypothesis needs to be validated based on an evaluation concept which is conducted by an independent scientific institution. During this trial period, the solution is already reimbursed based on a preliminary pricing agreement between the above-mentioned associations. After the 12-month period, the BfArM decides on a final DiGA listing, which, in case of acceptance, is followed by annual price renegotiations.

It becomes visible that this reimbursement system, which is open for national and international applications and according evidence demonstration, enables market and reimbursement access even for comparatively new solutions, and will, thus, very likely lead to a significant increase of real-world AI applications in healthcare.

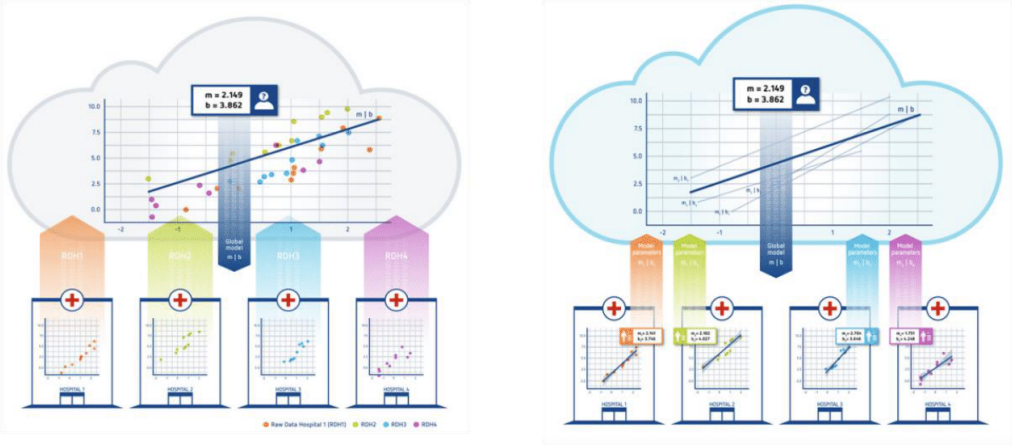
In a similar vein, the economic impact of AI will increase significantly throughout technologies which are able to overcome data access barriers while still maintaining data security. Since comprehensive data access is crucial for the growth of AI implementation, yet has proven particularly challenging in the healthcare sector, several approaches have been initiated to provide data-sharing economies, thereby also creating financial impact.

For instance, the MyHealthMyData (MHMD) project, which was set up to link hospitals and research centres in Europe and enable the exchange of health data, is based on a blockchain network and complies with GDPR (Manset et al. 2019).

Another framework which facilitates AI implementation and, thus, increases economic impact of AI in healthcare is the platform FeatureCloud. This platform simplifies privacy-aware access to comprehensive data for AI model training instead of relying on repetitive implementation of machine learning methods in each medical unit. The EU-funded technology targets to provide a platform for comprehensive data access across hospitals in Europe. Based on a cloud infrastructure, the local AI model parameters from each hospital can be shared without the need for any transfer of primary medical data, thus, being anonymous by default. A simplified procedure overview is listed in what follows and the respective methodology is graphically displayed in Fig. 5:

- Installation of FeatureCloud on hospital server
- Patient consent to use data or at least part of data in anonymous format

- Data workflow process mapping (depending on ML type, e.g., neural networks for pictures or decision trees for clinical decision support)
- Training of data set in each participating hospital locally
- Sharing of the model parameters, i.e., not the primary data but the trained model data (e.g., weights, decision trees), with one coordinator hospital
- Generation of one new trained model based on the model data of all participating hospitals



Caption: Illustrative example of hospital data upload in classical cloud setting with the upload of primary data (left side) vs. federated machine learning setting, i.e., data are not in cloud but inside the hospital and only model parameter are uploaded (right side)[1]

Credit: FeatureCloud, 2020

As visible in the figures, only trained model representatives and not the primary hospital data are exchanged between hospitals in one cloud (FeatureCloud 2020). Based on this privacy-preserving technology, the AI models can be trained with model parameters and then shared again with e.g., health care providers and payors. The system contributes significantly to improved security of data infrastructures as well as to patient trust (FeatureCloud 2020).

INFO BOX 4 (ONLY TEXT) – TO BE FORMATTED AS INFO BOX:

Governmental frameworks and platforms, such as the German Digital Healthcare Act which facilitates reimbursement of DH solutions, incl. AI-powered healthcare services, or the FeatureCloud platform which enhances data access while ensuring data privacy, will significantly contribute to increased real-world AI implementation in healthcare. Besides their direct positive impact on real-world AI implementation, it is likely that they will further induce an indirect positive impact by serving as role models for similar initiatives as well as by raising awareness and increasing acceptance of AI in healthcare.

[1] FeatureCloud ©.

5. Summary and outlook – the transformation of the healthcare market through Artificial Intelligence has started

The AI in healthcare market has more than doubled and became a billion-dollar market within a few years only. The promising prospects are, for instance, also reflected in the significant upward trend in funding activities which have amounted to approx. USD 4bn across 367 deals in 2019 (Pifer 2020). Nonetheless, real-world application is still in its early phase and only in some regions and healthcare segments first market traction can already be observed.

In particular, real-world AI applications to date predominantly stem from a limited number of healthcare segments such as medical imaging, diagnostics and drug discovery (MarketsAndMarkets 2020). In a similar vein, North America, by far accounts for the largest market share (Market Research Future 2019; Grand View Research 2019), where approx. 90% of revenues stem from the U.S. (MarketWatch 2020). However, there is significant growth potential for AI applications globally, including in low- and middle-income regions (Carrillo-Larco et al. 2020).

Specifically, the interplay of four key developments is likely to trigger a transformation of the healthcare market through AI in the short term. First, there is an urgent need to contain the worldwide rise in healthcare spending and to cope with increased demand for healthcare services and these challenges are, in many countries, further exacerbated by labor shortages. Second, improvements with regard to technological capabilities in recent years enable simplified and scaled AI implementation today more than ever. Third, for the first time, noteworthy awareness and acceptance levels can be observed for AI application in healthcare, both among key decision-makers such as governmental institutions as well as among patients. Finally, the corona pandemic has significantly increased the need for DH structures, more generally, and, in particular, AI has already been subject to large-scale testing in a range of different settings.

In order to close the significant gap between the considerable amount of promising academic research results on AI and the comparatively small number of actual real-world applications as well as to foster economic growth, intensified research on the medical and economic impact of AI is needed. In the past, this has neither been investigated to a sufficient extent nor at a high quality. While medical impact assessments have to some extent already yielded valid and reliable results, the economic impact has commonly been neglected in assessments of AI solutions, leading to a plethora of performance measures without providing a framework for strategic decision-making (Wolff et al. 2020). In fact, standardized economic impact measurement requires structured data input about the result of the digital intervention and this can be collected for example based on QALYs while applying the CHEERS and PRISMA quality criteria.

In addition to increased and improved economic impact measurement, reference cases for reimbursement, such as the Digital Healthcare Act in Germany, and for privacy-aware data access, such as the FeatureCloud platform, can serve as facilitators for increased implementation.

References

Accenture. 2017. “Artificial Intelligence: Healthcare’s New Nervous System.” https://www.accenture.com/_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf#zoom=50 (accessed 25.03.2021).

Accenture. 2018a. “2018 Consumer Survey on Digital Health – US Results.” https://www.accenture.com/t20180306t103559z__w__/us-en/_acnmedia/pdf-71/accenture-health-2018-consumer-survey-digital-health.pdf (accessed 25.03.2021).

Accenture. 2018b. “The Intelligent Payer: A Survival Guide.” https://www.accenture.com/_acnmedia/pdf-82/accenture-intelligent-payer-survivor-guide.pdf (accessed 25.03.2021).

Agrawal, Raag, and Sudhakaran Prabakaran. 2020. “Big Data in Digital Healthcare: Lessons Learnt and Recommendations for General Practice.” *Heredity* 124 (4): 525–34.

Aguis, Christina. 2019. “Evolution of AI: Past, Present, Future.” Data Driven Investor. 2019. https://medium.com/datadriveninvestor/evolution-of-ai-past-present-future-6f995d5f964a (accessed 25.03.2021).

Ahuja, Abhimanyu S. 2019. “The Impact of Artificial Intelligence in Medicine on the Future Role of the Physician.” *PeerJ* 7: e7702.

AI in Healthcare. 2017. “Healthcare AI Startups See Record Deals.” https://www.aiin.healthcare/sponsored/9667/topics/artificial-intelligence/healthcare-ai-startups-see-record-deals (accessed 25.03.2021).

Albahri, A. S., Rula A. Hamid, Jwan K. Alwan, Z. T. Al-Qays, A. A. Zaidan, B. B. Zaidan, A. O. S. Albahri, et al. 2020. “Role of Biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19): A Systematic Review.” *Journal of Medical Systems* 44 (7): 122.

Bundesinstitut für Arzneimittel und Medizinprodukte. 2020. “Das Fast-Track-Verfahren Für Digitale Gesundheitsanwendungen (DiGA) Nach § 139e SGB V – Ein Leitfaden Für Hersteller, Leistungserbringer und Anwender.” https://www.bfarm.de/SharedDocs/Downloads/DE/Service/Beratungsverfahren/DiGA-Leitfaden.pdf?__blob=publicationFile&v=4 (accessed 25.03.2021).

Carrillo-Larco, Rodrigo M., Lorainne Tudor Car, Jonathan Pearson-Stuttard, Trishan Panch, J. Jaime Miranda, and Rifat Atun. 2020. “Machine Learning Health-Related Applications in Low-Income and Middle-Income Countries: A Scoping Review Protocol.” *BMJ Open* 10 (5): e035983.

CB Insights. 2019. “From Drug R&D To Diagnostics: 90+ Artificial Intelligence Startups In Healthcare.” https://www.cbinsights.com/research/artificial-intelligence-startups-healthcare/ (accessed 25.03.2021).

CB Insights. 2020. “Healthcare AI In Numbers Q1’20: The Impact Of Covid-19 On Global Funding, Exits, Valuations, R&D, And More.” https://www.cbinsights.com/research/report/healthcare-ai-in-numbers-q1-2020/ (accessed 25.03.2021).

Cognizant. 2019. “New Provider Business Models for Healthcare on Demand.” https://www.cognizant.com/whitepapers/new-provider-business-models-for-the-digital-economy-codex4969.pdf (accessed 25.03.2021).

Davenport, Thomas, and Ravi Kalakota. 2019. “The Potential for Artificial Intelligence in Healthcare.” *Future Healthcare Journal* 6 (2): 94–98.

DigitalMR. 2017. “AI in UK Healthcare Landscape Overview.” http://analytics.dkv.global/data/pdf/AIUKHealthcare.pdf (accessed 25.03.2021).

European Commission. 2019. “Germany: Artificial Intelligence Strategy.” https://ec.europa.eu/knowledge4policy/publication/germany-artificial-intelligence-strategy_en (accessed 25.03.2021).

FeatureCloud. 2020. “FeatureCloud Publishable Summary First Reporting Period 1 St January 2019 – 30th April 2020.” https://featurecloud.eu/wp-content/uploads/2020/07/FeatureCloud_Publishable-summary_2020.pdf (accessed 25.03.2021). © 2021 FeatureCloud

Garbuio, Massimo, and Nidthida Lin. 2019. “Artificial Intelligence as a Growth Engine for Health Care Startups: Emerging Business Models.” *California Management Review* 61 (2): 59–83.
German Federal Ministry of Health. 2019. “Driving the Digital Transformation of Germany’s Healthcare System for the Good of Patients.” https://www.bundesgesundheitsministerium.de/digital-healthcare-act.html (accessed 25.03.2021).
Global Market Insights. 2019. “Healthcare Artificial Intelligence Market Size By Application (Medical Imaging & Diagnosis, Drug Discovery, Therapy Planning, Hospital Workflow, Wearables, Virtual Assistants) Industry Analysis Report, Regional Outlook, Application Potential, Competitive Market Share & Forecast, 2019 - 2025.” https://www.gminsights.com/industry-analysis/healthcare-artificial-intelligence-market (accessed 25.03.2021).
Grand View Research. 2019. “Artificial Intelligence In Healthcare Market Size, Share & Trends Analysis Report By Component (Hardware, Software, Services), By Application, By Region, Competitive Insights, And Segment Forecasts, 2019 - 2025.” https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-healthcare-market?utm_source=prnewswire&utm_medium=referral&utm_campaign=hc_16-dec-19&utm_term=ai-healthcare-market&utm_content=rd1 (accessed 25.03.2021).
Greenman, Simon. 2020. “Governments Must Build Trust in AI to Fight COVID-19 – Here’s How They Can Do It.” World Economic Forum. 2020. https://www.weforum.org/agenda/2020/04/governments-must-build-trust-in-ai-to-fight-covid-19-here-s-how-they-can-do-it/ (accessed 25.03.2021).
Hartman, Micah, Anne B. Martin, Joseph Benson, and Aaron Catlin. 2020. “National Health Care Spending In 2018: Growth Driven By Accelerations In Medicare And Private Insurance Spending.” *Health Affairs* 39 (1): 8–17.
Health Innovation Hub. 2019. DVG – Germany’s Jumpstart into Digital Health. https://hih-2025.de/wp-content/uploads/2019/11/2019-11-11_Frontiers-Health_DVG-english1.pdf (accessed 25.03.2021).
He, Jianxing, Sally L. Baxter, Jie Xu, Jiming Xu, Xingtao Zhou, and Kang Zhang. 2019. “The Practical Implementation of Artificial Intelligence Technologies in Medicine.” *Nature Medicine* 25 (1): 30–36.
Hunt, Mia. 2020. “Governments Turn to AI in Fight against Coronavirus.” Global Government Forum. 2020. https://www.globalgovernmentforum.com/governments-turn-to-ai-in-fight-against-coronavirus/ (accessed 25.03.2021).
Interactions. 2020. “Changing Consumer Comfort With AI.” https://www.interactions.com/press-releases/the-coronavirus-is-changing-consumer-comfort-with-ai/ (accessed 25.03.2021).
Kolasa, Katarzyna, and Grzegorz Kozinski. 2020. “How to Value Digital Health Interventions? A Systematic Literature Review.” *International Journal of Environmental Research and Public Health* 17 (6): 2119.
Kumar, Sudhesh. 2019. “Capitalising on Japan’s Digital Healthcare Economy during this Era of Aging Societies.” https://www.himss.eu/sites/himsseu/files/education/whitepapers/191007_HIMSS_Japan_White_Paper_WEB.pdf (accessed 25.03.2021).
Lindsell, Christopher J., William W. Stead, and Kevin B. Johnson. 2020. “Action-Informed Artificial Intelligence – Matching the Algorithm to the Problem.” *The Journal of the American Medical Association* 323 (21): 2141–42.
Manset, David, Laura Bernal, Mirko Koscina, and Octavio P. Kempner. 2019. “Blockchain and GDPR Compliance for the Healthcare Industry.” *HealthManagement.org* 19 (1): 42–44.
Market Research Future. 2019. “Healthcare Artificial Intelligence (AI) Market Research Report: Information by Component (Hardware, Software, Services), Application (Robot-Assisted Surgery, Virtual Nursing Assistant, Dosage Error Reduction, Clinical Trial,

Preliminary Diagnosis, Automated Image Diagnosis), Technology (Machine Learning, Querying Method, Natural Language Processing), End User (Hospital & Diagnostic Centers, Pharmaceutical & Biotechnology Companies Academic & Research Laboratories) and Region – Global Forecast till 2023." https://www.marketresearchfuture.com/reports/healthcare-artificial-intelligence-market-5681 (accessed 25.03.2021).

MarketsAndMarkets. 2020. "Artificial Intelligence in Healthcare Market with Covid-19 Impact Analysis by Offering (Hardware, Software, Services), Technology (Machine Learning, NLP, Context-Aware Computing, Computer Vision), End-Use Application, End User and Region - Global Forecast to 2026." https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-healthcare-market-54679303.html?gclid=Cj0KCQjwjer4BRCZARIsABK4QeWYH8M5LX-zbfM2UeF1Nal9XNCtcbmSLXx7SaFdwpUKNO229faOvt4aAoQuEALw_wcB (accessed 25.03.2021).

Market Study Report. 2019. "Healthcare Artificial Intelligence Market Size By Application (Medical Imaging & Diagnosis, Drug Discovery, Therapy Planning, Hospital Workflow, Wearables, Virtual Assistants) Industry Analysis Report, Regional Outlook (U.S. Canada, UK, Germany, Spain, Italy, France, Russia, Netherlands, Sweden, China, India, Japan, Australia, Singapore, Brazil, Mexico, South Africa, Saudi Arabia), Application Potential, Competitive Market Share & Forecast, 2019 - 2025." https://www.marketstudyreport.com/reports/healthcare-artificial-intelligence-market-size-by-application-medical-imaging-diagnosis-drug-discovery-therapy-planning-hospital-workflow-industry-analysis-report-regional-outlook-u-s-canada-germany-uk-france-spain-italy-china-india-japan-australia-brazil-mexico-south-africa-growth-potential-competitive-market-share-forecast-2017-2024 (accessed 25.03.2021).

MarketWatch. 2020. "Healthcare Artificial Intelligence Market Global Growth Opportunities Industry Analysis Forecast to 2025." https://www.marketwatch.com/press-release/artificial-intelligence-in-healthcare-market-growth-leading-players-and-forecast-to-2025-2021-03-12 (accessed 25.03.2021).

Matthies, Henrik. 2019. "DVG – A Summary of Germany's New Law for Digital Health Applications." HIH 2025. https://hih-2025.de/dvg-a-summary-of-germanys-new-law-for-digital-health-applications/ (accessed 25.03.2021).

McKinsey & Company. 2017. "Artificial Intelligence in Health Insurance - Smart Claims Management with Self-Learning Software." https://healthcare.mckinsey.com/wp-content/uploads/2020/02/Artificial-intelligence-in-Health-Insurance.pdf (accessed 25.03.2021).

Meticulous Research. 2019. "Artificial Intelligence In Healthcare Market By Product (Hardware, Software, Services), Technology (Machine Learning, Context-Aware Computing, NLP), Application (Drug Discovery, Precision Medicine), End User, And Geography - Global Forecast To 2025." https://www.meticulousresearch.com/product/artificial-intelligence-healthcare-market/?utm_source=wiredrelease&utm_medium=pharmiweb&utm_campaign=kavita-17-02-2020 (accessed 25.03.2021).

Mordor Intelligence. 2020. "Artificial Intelligence in Medicine Market - Growth, Trends, and Forecast (2020-2025)." https://www.mordorintelligence.com/industry-reports/artificial-intelligence-in-medicine-market (accessed 25.03.2021).

Nakata, Norio. 2019. "Recent Technical Development of Artificial Intelligence for Diagnostic Medical Imaging." *Japanese Journal of Radiology* 37 (2): 103–8.

OECD. 2019. *Artificial Intelligence in Society*. OECD Publishing.

Panch, Trishan, Heather Mattie, and Leo Anthony Celi. 2019. "The 'Inconvenient Truth' about AI in Healthcare." *Npj Digital Medicine* 2: 77.

Pifer, Rebecca. 2020. "Health AI Startups Netted a Record $4B in Funding Last Year." Healthcare Dive. https://www.healthcaredive.com/news/health-ai-startups-netted-a-record-4b-in-funding-last-year/570957/ (accessed 25.03.2021).

Radboudumc. 2019. "Deep Learning for Prostate MRI Intervention – Diagnostic Image

Analysis Group.” http://diag.radboudimaging.nl/index.php/Deep_learning_for_Prostate_MRI_Intervention (accessed 25.03.2021).
Reinsel, David, John Gantz, and John Rydning. 2018. “Data Age 2025: The Digitization of the World – From Edge to Core.” https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf (accessed 25.03.2021).
ReportLinker. 2020. “Artificial Intelligence in Healthcare Market with Covid-19 Impact Analysis by Offering, Technology, End-Use Application, End User and Region – Global Forecast to 2026.” https://www.reportlinker.com/p04897122/Artificial-Intelligence-in-Healthcare-Market-by-Offering-Technology-Application-End-User-Industry-and-Geography-Global-Forecast-to.html (accessed 25.03.2021).
RevCycleIntelligence. 2020. “How Artificial Intelligence Is Optimizing Revenue Cycle Management.” https://revcycleintelligence.com/features/how-artificial-intelligence-is-optimizing-revenue-cycle-management (accessed 25.03.2021).
Sadegh, Sepideh, Julian Matschinske, David B. Blumenthal, Gihanna Galindez, Tim Kacprowski, Markus List, Reza Nasirigerdeh, et al. 2020. “Exploring the SARS-CoV-2 Virus-Host-Drug Interactome for Drug Repurposing.” *Nature Communications* 11 (1): 3518.
Schwalbe, Nina, and Brian Wahl. 2020. “Artificial Intelligence and the Future of Global Health.” *The Lancet* 395 (10236): 1579–86.
U. K. Department of Health & Social Care. 2019. “Code of Conduct for Data-Driven Health and Care Technology.” https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology (accessed 25.03.2021).
U.S. Department of Veterans Affairs. 2014. “Corporate Data Warehouse (CDW).” https://www.hsrd.research.va.gov/for_researchers/vinci/cdw.cfm (accessed 25.03.2021).
U.S. Food and Drug Administration. 2020. “Artificial Intelligence and Machine Learning in Software.” 2020. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device (accessed 25.03.2021).
Vaishya, Raju, Mohd Javaid, Ibrahim Haleem Khan, and Abid Haleem. 2020. “Artificial Intelligence (AI) Applications for COVID-19 Pandemic.” *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 14 (4): 337–39.
Wolff, Justus, Josch Pauling, Andreas Keck, and Jan Baumbach. 2020. “The Economic Impact of Artificial Intelligence in Health Care: Systematic Review.” *Journal of Medical Internet Research* 22 (2): e16866.
World Health Organization. 2002. “The World Health Report 2002: "Reducing Risks, Promoting Healthy Life".” https://www.who.int/whr/2002/en/whr02_en.pdf?ua=1 (accessed 25.03.2021).
World Health Organization. 2019. “Global Spending on Health: A World in Transition.” https://www.who.int/health_financing/documents/health-expenditure-report-2019.pdf?ua=1 (accessed 25.03.2021).

Paper C: The Economic Impact of Artificial Intelligence in Health Care: Systematic Review

The paper entitled “The Economic Impact of Artificial Intelligence in Health Care: Systematic Review” represents a systematic literature review of existing economic impact assessments of AI applications in healthcare and assesses how economic impact is and should be measured. The study reveals that there are only few economic impact assessments and that these are commonly subject to methodological flaws. Therefore, it presents concrete levers for improvement when conducting economic impact assessments.

Title	**The economic impact of Artificial Intelligence in healthcare: A systematic review**
Research question	To what degree and in what quality has the medical and economic impact of AI in healthcare been assessed and which areas could be improved?
Background	Evidence on benefits of AI is not sufficiently measured and this hinders the implementation in medical routine on various levels
Contribution	Shows that there are only few studies analyzing the impact of AI in healthcare and that these studies lack quality and consistency in their evaluation procedures
Method	Systematic review of academic studies analyzing the economic impact of AI by benchmarking them against a predefined set of quality criteria for cost-effectiveness assessments; studies for in-depth analysis are identified and assessed for inclusion via a systematic search and inclusion process
Take-away	In light of the high relevance of medical and economic impact assessments, on the one hand, and the low number of impact assessments for AI and significant methodological deficits, on the other hand, scientists and practitioners should undertake more and higher quality impact assessments, In this context they should, for example apply QALYs based on EQ-5D questionnaires as well as the CHEERS and PRISMA criteria and include a Net Present Value calculation as well as Cost Alternative Scenarios into their assessments

Table 3: Overview about the publication “The economic impact of AI in healthcare: A systematic review”

Contribution of the doctoral candidate: First authorship including the planning of the publication structure (in particular the designing and application of in- and exclusion criteria as well as analysis of academic studies) and deduction of consequences for improvements of medical and economic impact measurements. Manuscript: Writing, review and editing.

Review

The Economic Impact of Artificial Intelligence in Health Care: Systematic Review

Justus Wolff[1,2*], MSc; Josch Pauling[1*], PhD; Andreas Keck[2*], MD; Jan Baumbach[1*], PhD

[1]TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany

[2]Strategy Institute for Digital Health, Hamburg, Germany

[*]all authors contributed equally

Corresponding Author:
Justus Wolff, MSc
TUM School of Life Sciences Weihenstephan
Technical University of Munich
Maximus-von-Imhof-Forum 3
Freising, 85354
Germany
Phone: 49 40329012 0
Email: justus.wolff@syte-institute.com

Abstract

Background: Positive economic impact is a key decision factor in making the case for or against investing in an artificial intelligence (AI) solution in the health care industry. It is most relevant for the care provider and insurer as well as for the pharmaceutical and medical technology sectors. Although the broad economic impact of digital health solutions in general has been assessed many times in literature and the benefit for patients and society has also been analyzed, the specific economic impact of AI in health care has been addressed only sporadically.

Objective: This study aimed to systematically review and summarize the cost-effectiveness studies dedicated to AI in health care and to assess whether they meet the established quality criteria.

Methods: In a first step, the quality criteria for economic impact studies were defined based on the established and adapted criteria schemes for cost impact assessments. In a second step, a systematic literature review based on qualitative and quantitative inclusion and exclusion criteria was conducted to identify relevant publications for an in-depth analysis of the economic impact assessment. In a final step, the quality of the identified economic impact studies was evaluated based on the defined quality criteria for cost-effectiveness studies.

Results: Very few publications have thoroughly addressed the economic impact assessment, and the economic assessment quality of the reviewed publications on AI shows severe methodological deficits. Only 6 out of 66 publications could be included in the second step of the analysis based on the inclusion criteria. Out of these 6 studies, none comprised a methodologically complete cost impact analysis. There are two areas for improvement in future studies. First, the initial investment and operational costs for the AI infrastructure and service need to be included. Second, alternatives to achieve similar impact must be evaluated to provide a comprehensive comparison.

Conclusions: This systematic literature analysis proved that the existing impact assessments show methodological deficits and that upcoming evaluations require more comprehensive economic analyses to enable economic decisions for or against implementing AI technology in health care.

(J Med Internet Res 2020;22(2):e16866) doi: 10.2196/16866

KEYWORDS

telemedicine; artificial intelligence; machine learning; cost-benefit analysis

Introduction

Background

In times of value-based health care and also because of the high share of the health care industry in the overall economy, economic impact assessment is of increasing importance. For instance, health care expenditures account for approximately US $3.5 trillion out of US $19.4 trillion (18%) of the overall gross domestic product (GDP) in the United States and for approximately US $0.4 trillion out of US $3.7 trillion (11.5%) of the overall GDP in Germany [1,2]. Accordingly, the cost impact of digital health applications has also been analyzed in several studies.

In 2002, in a review of cost-effectiveness studies in the context of telemedicine interventions, Whitten et al [3] revealed that only 55 out of 612 identified articles presented actual cost-benefit data, which were required to be included in a detailed review. In addition, after analyzing these articles, the authors concluded that the provided evidence was not sufficient to assess whether telemedicine represents a cost-effective mean of delivering health care [3].

More than a decade later, in 2014, Elbert et al [4] described in a review of systematic reviews and meta-analyses regarding electronic health (eHealth) interventions in somatic diseases that out of 31 reviews, 7 papers concluded that digital health is effective or cost-effective, 13 underlined that evidence is promising, and the other 11 found only limited or inconsistent proof. They also highlighted that the development and evaluation of strategies to implement effective or cost-effective eHealth initiatives in daily practice needed to be significantly enhanced [4].

In another systematic review study on the economic evaluations of eHealth technologies from 2018, Sanyal et al [5] analyzed multiple databases with publications between 2010 and 2016. On the basis of 11 studies that fulfilled the inclusion criteria, the authors found that most of the studies demonstrated efficacy and cost-effectiveness of an intervention using a randomized control trial and statistical modeling. However, there was insufficient information provided on the feasibility of adopting these modeling technologies. Thus, the paper emphasizes that the current level of evidence is inconclusive and that more research is needed to evaluate possible long-term cost benefits [5].

Research in this segment has been continuously intensified, and in several studies, the digital health cost-effectiveness, for example, of telemedicine for remote orthopedic consultations [6], digital behavioral interventions for type 2 diabetes and hypertension [7], and internet-based interventions for mental health [8] was analyzed in detail.

As significant medical quality enhancements and cost-saving improvements through artificial intelligence (AI) as one of the key emerging technologies in digital health are expected, the economic impact assessment of AI in health care has a crucial role for all stakeholders in health care and, thus, needs to be analyzed in detail.

Objective

It was systematically investigated whether the existing cost-effectiveness evaluations meet the established quality criteria to enable comprehensive decision making regarding the implementation of AI in health care. On the basis of these thorough economic assessments, the necessary information to decide for or against the application of AI in hospitals, industry, and payer context will be provided.

Methods

A systematic literature review was performed as described in the following sections.

Search Strategies

A literature search was conducted utilizing the PubMed database and using the search terms provided in Table 1.

Table 1. Search terms (title and abstract) in the PubMed analysis (conducted on July 29, 2019).

Components	Syntax	Hits, n
Artificial intelligence OR machine learning AND cost effectiveness	(Artificial intelligence [title/abstract] OR machine learning [title/abstract]) AND cost effectiveness [title/abstract]	54
Artificial intelligence OR machine learning AND economic impact	(Artificial intelligence [title/abstract] OR machine learning [title/abstract]) AND economic impact [title/abstract]	9
Artificial intelligence OR machine learning AND cost saving	(Artificial intelligence [title/abstract] OR machine learning [title/abstract]) AND cost saving [title/abstract]	3

The search terms *Artificial Intelligence* and *Machine Learning* for the overall segment are not exhaustive as eg, *Decision trees*, *Support vector machines*, or *Deep neural networks* could also have been used as search terms for the database queries. Nonetheless, as strategic decisions based on economic impact are mostly made on a strategic managerial and medical level without a specific technological background, the most frequently used search terms regarding AI in health care have been used. In addition, it is highly probable that papers about, for example, *deep neural networks* would also include such terms as *artificial intelligence*, *support vector machines*, and *machine learning* at least in the abstract. Finally, it was decided to use a Google Trends analysis comparing the most frequently used search terms regarding AI in health care over the last 12 months globally [9]: The terms *Artificial Intelligence* and *Machine Learning* have been used the most by far, as illustrated in Multimedia Appendix 1.

Inclusion Criteria

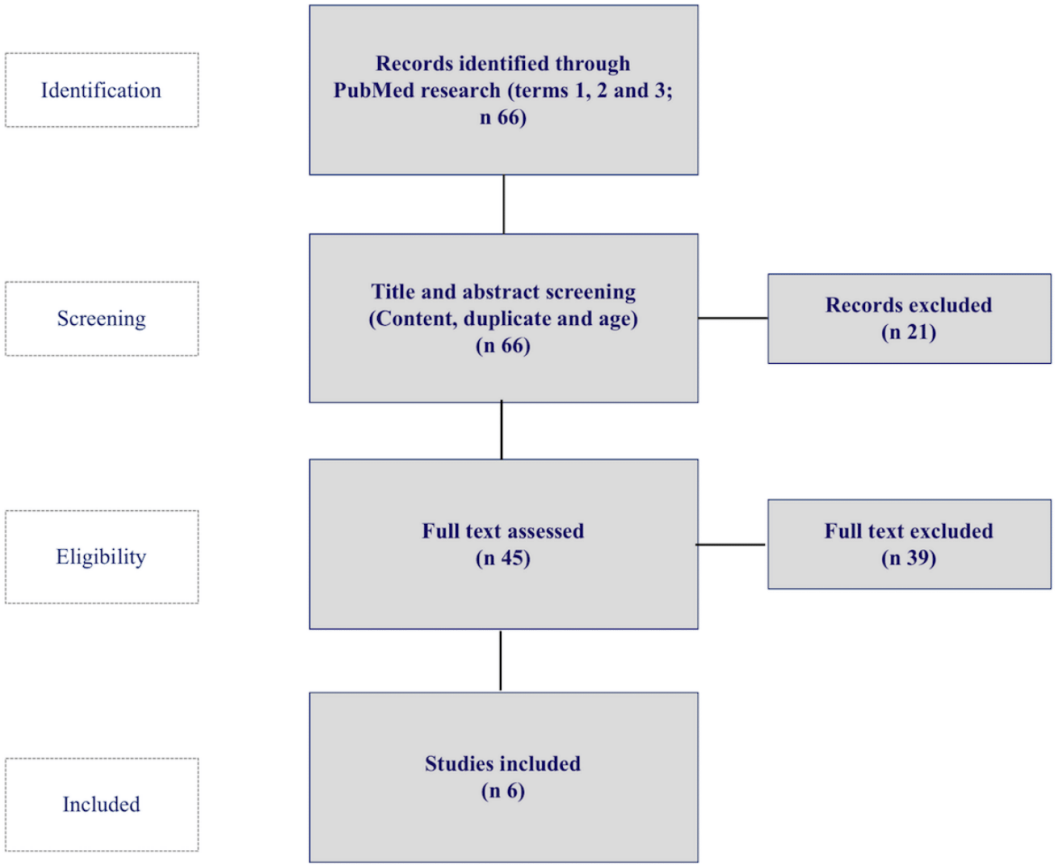
For the publications identified through the PubMed searches, the titles, abstracts, and full texts have been reviewed. Publications were included into the subsequent analysis if they were (1) published journal articles, (2) written in English language, and (3) published no more than 5 years ago. With regard to the content, the publications were included if they focused on at least one of the following content sectors: (1) a comprehensive description of an AI functionality, (2) an evaluation of the economic efficiency and outcomes of the AI functionality, and (3) quantitative outcomes of the AI functionality in at least one health care system. Furthermore, only publications describing concrete medical and economic outcomes, such as cost savings per patient per year, and reviews or meta-analyses comparing AI solutions have been included.

Exclusion Criteria

Exclusion criteria for an article were defined as follows: (1) the title did not cover a topic related to AI in health care; (2) neither the title nor the abstract contained a description of an AI application in health care; or (3) the title, abstract, or full text did not elaborate on the quantitative economic outcome of AI in health care application in any health care system. In contrast to other previous research review approaches, such as those chosen by Elbert et al [4] or Ekeland et al [10], the third exclusion criterion was covered. Although this significantly limited the number of cost-effectiveness studies included, it was applied to compare the different cost-effectiveness analysis approaches and not only the health- or process-related outcomes without quantified economic impact from a national or international health care perspective.

After identifying potential studies for inclusion via the PubMed search, as previously described, the evaluation took place in two steps (Figure 1). First, all titles, abstracts, and full texts were screened for the fulfillment of the inclusion and exclusion criteria. Second, publications viable for inclusion were assessed with a quality criteria catalog, which is explained in section Quality Criteria for Economic Impact Assessment.

Figure 1. Study selection and identification flowchart.



Quality Criteria for Economic Impact Assessment

A combined criteria catalog for cost-effectiveness studies was designed. Besides own criteria, additional evaluation aspects from classical health care effectiveness studies and digital health assessments were considered [5,11]. The quality criteria are summarized in Table 2.

Table 2. Quality criteria for economic impact assessment.

Criteria	Explanation	Source
Description of cost-effectiveness of AI[a] solution	Level of detail of cost-effectiveness explanation	Authors
Hypothesis formulation	Analysis if a comprehensive question has been formulated that allows AI cost-effectiveness evaluation (eg, comparing the AI approach with the recommended guideline routine)	Study by Haycox and Walley [11]
Cost-effectiveness perspective	Impact of change in the cost of stand-alone functionality vs overall reduction of burden of care	Study by Haycox and Walley [11]
Consideration of cost alternative	Analysis if the cost-saving results could also have been achieved with an alternative strategy	Study by Haycox and Walley [11]
Benefit today	Net present value of the AI service, including upfront investments and running costs	Study by Haycox and Walley [11]
Verification of base case	Analysis of cost-effectiveness of the AI solution based on benchmarking with base case data	Study by Sanyal et al [5]

[a]AI: artificial intelligence.

Results

Quality Criteria Evaluation

Quality criteria have been applied to assess the economic impact assessments on a scale of 1 to 3 (1=superficial coverage, 2=solid coverage, and 3=detailed explanation). As outlined above, 6 publications have been assessed regarding the described quality criteria for economic impact evaluation. An overview of the analysis of the publications [12-17] is given in Multimedia Appendix 2.

Quality Assessment Results

We first conclude that the level of detail of description of the cost-effectiveness measurement was overall high as the descriptions were for the most part precise and detailed, for instance, "for an incremental cost effectiveness threshold of €25,000/quality-adjusted life year, it was demonstrated that the AI tool would have led to slightly worse outcomes (1.98%), but with decreased cost (5.42%)" [14]. Overall, 5 out of the 6 publications had a very high level of detail, and only 1 study had a medium level of detail in the general description (only a positive/negative cost-saving impact description and no further outcome explanations have been provided [13]).

Second, the hypothesis formulation (eg, cost saving through machine learning–based prediction models to identify optimal heart failure patients for disease management programs to avoid 30-day readmissions [17]) was clear and accurate across all publications. All comprised well-explained and coherent hypothesis formulations.

Third, the cost-effectiveness perspective had in all cases a *health care system* context, although additional perspectives could have been included, such as ambulant or nurse perspectives. Furthermore, 5 studies demonstrated a comprehensive health care system perspective, whereas 1 could have been extended from a hospital to an overall system view [13].

Fourth, the cost alternative consideration, that is, the analysis of whether the cost-saving results could also have been achieved alternatively, was mostly missing. Only 2 papers elaborated on the different alternatives in detail, for example, differentiating on the levels of risks of the respective patient groups or different treatment options. Besides these 2 publications [12,16] that covered various alternatives to achieve a similar cost saving, the remaining 4 publications did not elaborate on such cost alternative considerations at all.

Fifth, the benefit achieved today, that is, in terms of a net present value (NPV) including not only the benefits but also the necessary investment for the AI implementation and the operational costs of an AI service delivery, was not covered in any of the 6 studies. Only 1 study compared AI vs non-AI scenarios but without providing a NPV calculation. Hence, all 6 studies included a quantification of economic outcomes but failed to calculate an overall NPV.

Finally, the verification of the base case was conducted using different approaches across the 6 studies. Mostly solid data sources have been collected in dedicated AI-focused studies based on, for example, comparison of cost with/without the algorithm, reimbursement code analysis, or benchmarking of the result with the reported performance of other clinics. All papers presented a cost-effectiveness measurement based on a comprehensive comparison dataset.

One additional aspect that emerged throughout the analysis was the measurement of resource usage, which was (almost) in all papers conducted via a top-down approach, meaning from an overall health care perspective but not from a single cost split per task. In this way, important cost drivers of potentially *hidden* stakeholders could have been missed (eg, additional workload for ambulatory care if a hospital treatment is altered).

Discussion

Principal Findings

Overall, the outcomes of the analysis described above can be split into two result categories, namely, general feedback from the analysis and detailed assessment of the studies that have been included in the review process based on the study's inclusion and exclusion criteria.

Generally, only a few publications can be found for the economic impact assessment of AI in health care. On the basis

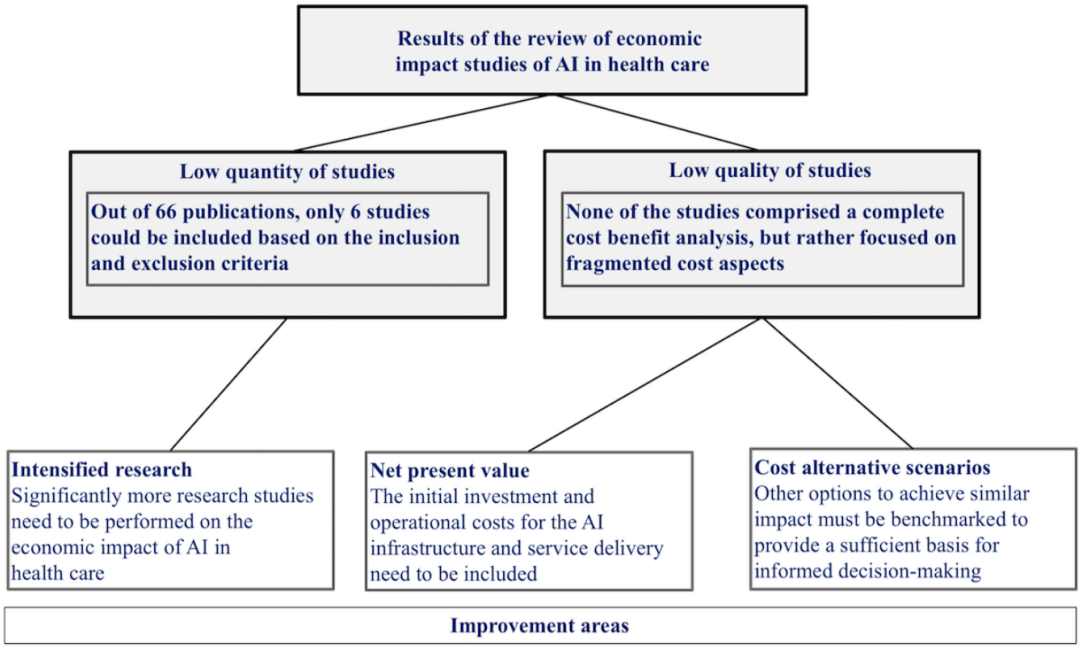
of the different search terms that include the most frequently searched phrases by far in this segment (*Artificial Intelligence* and *Machine Learning*) in combination with the economic impact (*Cost effectiveness*, *Economic impact*, *Cost saving*), there were only 66 PubMed hits. As AI strategies and consequent decision-making processes depend on solid data as the basis for decision making, this is a significant challenge for both the management and medical staff, for example, when general pro and contra decisions and specific implementations regarding AI are discussed.

When accounting for the details given in the identified AI in health care publications, the economic assessment quality shows several deficits that need to be overcome in the future. Only 6 out of the 66 publications (9%) could be included in the detailed assessment. Out of these 6 studies, none comprised a complete cost-benefit analysis; rather, they all focused on fragmented cost or cost-saving aspects.

Room for improvement (Figure 2) has been identified in two main areas:

- First, initial investment and operational costs for the AI infrastructure and service need to be included in the assessment. This is a core element for any strategic decision-making process, and the complete initial and operational investment costs for an AI solution must be compared with the expected economic benefits to provide concrete decision-making support.
- Second, further options to achieve similar impact must be evaluated to reach a sufficient basis for comprehensive and transparent decision-making, allowing comparisons among different strategic and investment options (eg, a genetic sequencing process or different medical expertise allocation for a diagnosis and treatment outcome improvement could also be applied instead of an AI-driven patient screening).

Figure 2. Result of the literature review and improvement areas for economic impact assessment of artificial intelligence (AI) in health care.



The conducted review has a rather narrow focus on economics and business perspectives of AI in health care. However, the literature review revealed further significant success factors for AI, for example, regarding the legal framework, such as compliance with data security, protection, and privacy policies, and also universally accepted technological requirements to enable comprehensive data collection and to analyze content while complying with data privacy requirements. Despite the benefits in assisting diagnostic and therapeutic decisions, so far, no standards for these legal and technological issues have been defined, and these aspects should be analyzed in future research with a broader focus.

Furthermore, aside from the sole economic quantitative aspects, the qualitative aspects of AI in health care for patients and the society require further research. For instance, in rural areas where the availability of primary care physicians is limited, AI can replace processes through focused test support, for example, for type 2 diabetes, thus addressing the challenges of demographic change [18]. The comparison between AI and physicians with regard to diagnosis performance demonstrated that AI can deliver equal results, for example, in image recognition–related fields [19]. This can, among others, also support a reallocation of medical capacities. In addition, AI can also enable a shift from a generalized to a more personalized treatment. AI-steered outcome prediction and clinical decision support processes are already used today, for instance, for patients in radiation therapy [20].

Prior reviews in the digital health segment categorized the results into groups, for example, computerized decision support system, Web-based physical activity intervention, internet-delivered cognitive behavioral therapy, and telehealth. In addition, user's age was differentiated (eg, children vs old patients), and shortcomings such as a missing difference between short- and long-term cost savings were highlighted [5]. They also covered challenges that go beyond the cost-effectiveness aspect and mentioned, for instance, that the way to implement digital health in daily practice is still unclear [4] or that patient perspectives and collaborative approaches among a variety of stakeholders are needed [10].

Note that the focus on AI in health care required considering novel factors and a refined search strategy as compared with typical reviews on digital health resulting in differential results. First, in contrast to other reviews, Google Trends has proven to be an effective tool to narrow the search space for a representative collection of results. On the basis of the Google Trends analysis, the key phrases *Artificial Intelligence* and *Machine Learning* could be identified as the most frequently used terms by far. Second, the review covered a higher percentage of included studies after applying the defined inclusion and exclusion criteria (9% of the analyzed papers were included), whereas prior reviews had much lower inclusion rates—8% (55/612) in the study by Whitten et al [3], 2% (31/1657) in the study by Elbert et al [4], or 0.1% (11/1625) in the study by Sanyal et al [5]). This was because of two reasons: (1) AI as a subsegment of digital health in business and industry is still not covered well in scientific publications and (2) the high importance of quantitatively reported outcomes required as inclusion criterion. Third, the evaluation of cost-effectiveness studies has been conducted with a quality criteria catalog from a management perspective. As AI implementation is cost- and labor-intensive and decisions are not exclusively driven by medical improvement rates, the business management decision making basis has been chosen as crucial for positive implementation decisions and subsequent widescale applications. The addition of the business management view includes classical cost factors (onetime and running expenses) as well as decisions among different strategies to deliver cutting edge health services.

Conclusions

Current research covers impact assessments of AI in health care rather moderately and shows qualitative deficits in methodology. Future cost-effectiveness analyses need to increase in number and quality. They should include initial investment and running costs as well as the comparison with alternative technologies. This way a comprehensive and clearly segmented cost-benefit evaluation can be provided, which will serve as a sufficient basis for decision making regarding AI implementations.

Acknowledgments

The authors would like to thank Julia Menacher for her editorial support in creating this publication. JB is grateful for financial support from Villum Young Investigator grant number 13154. In addition, some of the work from JB was funded by H2020 project RepoTrial (number 777111). Contributions by JP are funded by the Bavarian State Ministry of Science and the Arts in the framework of the Center Digitisation.Bavaria (Zentrum Digitalisierung.Bayern).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshot of a Google Trends analysis of search terms related to artificial intelligence in health care globally over the last 12 months (conducted on October 9, 2019).
[PNG File , 176 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Analysis of the included economic impact studies.
[PNG File , 396 KB-Multimedia Appendix 2]

References

1. Phelps CE. Health Economics. Abingdon, Oxfordshire: Routledge; 2017.
2. Statistisches Bundesamt. Health Expenditure in 2017: +4.7% URL: https://www.destatis.de/DE/Presse/Pressemitteilungen/2019/03/PD19_109_23611.html [accessed 2019-10-03]
3. Whitten PS, Mair FS, Haycox A, May CR, Williams TL, Hellmich S. Systematic review of cost effectiveness studies of telemedicine interventions. Br Med J 2002 Jun 15;324(7351):1434-1437 [FREE Full text] [doi: 10.1136/bmj.324.7351.1434] [Medline: 12065269]
4. Elbert N, van Os-Medendorp H, van Renselaar W, Ekeland A, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of eHealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
5. Sanyal C, Stolee P, Juzwishin D, Husereau D. Economic evaluations of eHealth technologies: a systematic review. PLoS One 2018;13(6):e0198112 [FREE Full text] [doi: 10.1371/journal.pone.0198112] [Medline: 29897921]

6. Buvik A, Bergmo TS, Bugge E, Smaabrekke A, Wilsgaard T, Olsen JA. Cost-effectiveness of Telemedicine in remote orthopedic consultations: randomized controlled trial. J Med Internet Res 2019 Feb 19;21(2):e11330 [FREE Full text] [doi: 10.2196/11330] [Medline: 30777845]
7. Nordyke RJ, Appelbaum K, Berman MA. Estimating the impact of novel digital therapeutics in Type 2 diabetes and hypertension: health economic analysis. J Med Internet Res 2019 Oct 9;21(10):e15814 [FREE Full text] [doi: 10.2196/15814] [Medline: 31599740]
8. Le LK, Sanci L, Chatterton ML, Kauer S, Buhagiar K, Mihalopoulos C. The cost-effectiveness of an internet intervention to facilitate mental health help-seeking by young adults: randomized controlled trial. J Med Internet Res 2019 Jul 22;21(7):e13065 [FREE Full text] [doi: 10.2196/13065] [Medline: 31333199]
9. Google Trends. Compare URL: https://trends.google.com/trends/explore?cat=45&q=Artificial%20Intelligence,Machine%20learning,Neural%20network,Linear%20regression,Support%20vector%20%20machine [accessed 2019-10-09]
10. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
11. Haycox A, Walley T. Pharmacoeconomics: evaluating the evaluators. Br J Clin Pharmacol 1997 May;43(5):451-456 [FREE Full text] [doi: 10.1046/j.1365-2125.1997.00575.x] [Medline: 9159558]
12. Lee HK, Jin R, Feng Y, Bain PA, Goffinet J, Baker C, et al. An analytical framework for TJR readmission prediction and cost-effective intervention. IEEE J Biomed Health Inform 2019 Jul;23(4):1760-1772. [doi: 10.1109/JBHI.2018.2859581] [Medline: 30047916]
13. Gönel A. Clinical biochemistry test eliminator providing cost-effectiveness with five algorithms. Acta Clin Belg 2018 Dec 25:1-5. [doi: 10.1080/17843286.2018.1563324] [Medline: 30585134]
14. Bremer V, Becker D, Kolovos S, Funk B, van Breda W, Hoogendoorn M, et al. Predicting therapy success and costs for personalized treatment recommendations using baseline characteristics: Data-driven analysis. J Med Internet Res 2018 Aug 21;20(8):e10275 [FREE Full text] [doi: 10.2196/10275] [Medline: 30131318]
15. Grover D, Bauhoff S, Friedman J. Using supervised learning to select audit targets in performance-based financing in health: an example from Zambia. PLoS One 2019;14(1):e0211262 [FREE Full text] [doi: 10.1371/journal.pone.0211262] [Medline: 30695057]
16. Lee I, Monahan S, Serban N, Griffin P, Tomar S. Estimating the cost savings of preventive dental services delivered to Medicaid-enrolled children in six southeastern states. Health Serv Res 2018 Oct;53(5):3592-3616 [FREE Full text] [doi: 10.1111/1475-6773.12811] [Medline: 29194610]
17. Golas SB, Shibahara T, Agboola S, Otaki H, Sato J, Nakae T, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. BMC Med Inform Decis Mak 2018 Jun 22;18(1):44 [FREE Full text] [doi: 10.1186/s12911-018-0620-z] [Medline: 29929496]
18. Spänig S, Emberger-Klein A, Sowa J, Canbay A, Menrad K, Heider D. The virtual doctor: an interactive clinical-decision-support system based on deep learning for non-invasive prediction of diabetes. Artif Intell Med 2019 Sep;100:101706. [doi: 10.1016/j.artmed.2019.101706] [Medline: 31607340]
19. Shen J, Zhang CJ, Jiang B, Chen J, Song J, Liu Z, et al. Artificial Intelligence versus clinicians in disease diagnosis: systematic review. JMIR Med Inform 2019 Aug 16;7(3):e10010 [FREE Full text] [doi: 10.2196/10010] [Medline: 31420959]
20. Pillai M, Adapa K, Das SK, Mazur L, Dooley J, Marks LB, et al. Using Artificial Intelligence to improve the quality and safety of radiation therapy. J Am Coll Radiol 2019 Sep;16(9 Pt B):1267-1272. [doi: 10.1016/j.jacr.2019.06.001] [Medline: 31492404]

Abbreviations

AI: artificial intelligence
eHealth: electronic health
GDP: gross domestic product
NPV: net present value

Edited by G Eysenbach; submitted 01.11.19; peer-reviewed by D Heider, N Przulj, E van der Velde; comments to author 07.12.19; revised version received 18.12.19; accepted 19.12.19; published 20.02.20

Please cite as:
Wolff J, Pauling J, Keck A, Baumbach J
The Economic Impact of Artificial Intelligence in Health Care: Systematic Review
J Med Internet Res 2020;22(2):e16866
URL: http://www.jmir.org/2020/2/e16866/
doi: 10.2196/16866
PMID: 32130134

Paper D: Federated machine learning for a facilitated implementation of Artificial Intelligence in healthcare – a proof of concept study for the prediction of Coronary Artery Calcification Scores

The paper entitled “Federated machine learning for a facilitated implementation of Artificial Intelligence in healthcare – a proof of concept study for the prediction of Coronary Artery Calcification Scores” represents a real-world AI implementation in a privacy-preserving medical setting and addresses the question of how a rather novel, yet promising “privacy by design” technology can influence AI implementation growth in healthcare. The study is based on real patient data of two medical institutions in Germany and provides insights on the accuracy of a privacy-preserving FL approach compared with a traditional, i.e., centralized, AI approach.

Title	**Federated machine learning for a facilitated implementation of Artificial Intelligence in healthcare - a proof of concept study for the prediction of Coronary Artery Calcification Scores**
Research question	Is FL, as a privacy-preserving approach to overcome data access issues for AI implementation in healthcare, a potential pathway for increased AI implementation in healthcare?
Background	Technological limitations as one key reason for the low number of real-world AI applications; FL can be a potential pathway for privacy-preserving data access and, thus, increased AI implementation
Contribution	Shows that FL yields similar accuracy levels as a traditional, i.e., centralized, AI algorithm, thus indicating that FL can be a very valuable pathway for AI implementation due to facilitated data access across different medical institutions
Method	Empirical study comparing the results of a centralized locally trained random forest model with a FL random forest model based on the data of two medical facilities, using the FeatureCloud platform

Take-away	Access to privacy-sensitive and fragmented data is commonly required for AI implementation and scaling in real-world settings. In particular, practitioners need to comply with regulations (even more than scientific researchers since there is no clinical trial/research environment) and can, to this end, apply “privacy by design” technologies. The FL accuracy was very comparable to a centralized model and can serve as a reference case for future AI implementations in healthcare

Table 4: Overview about the publication “Federated machine learning for a facilitated implementation of AI in healthcare - a proof of concept study for the prediction of Coronary Artery Calcification Scores”

Contribution of the doctoral candidate: First authorship including the collaboration planning with the medical institution and a close collaboration with the members of the chair for the planning of the implementation of FeatureCloud (in particular Julian Matschinske) and the data analysis. Manuscript: Writing, review and editing.

Federated machine learning for a facilitated implementation of Artificial Intelligence in healthcare – a proof of concept study for the prediction of Coronary Artery Calcification Scores

Justus Wolff[1,2], Julian Matschinske[3], Dietrich Baumgart[4], Anne Pytlik[4], Andreas Keck[2], Arunakiry Natarajan[5] , Claudio E. von Schacky[6], Josch K. Pauling[1,7], Jan Baumbach[3,8]

[1]Chair of Experimental Bioinformatics, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Maximus-von-Imhof-Forum 3, 85354 Freising, Germany

[2]Syte – Strategy Institute for Digital Health, Hohe Bleichen 8, 20354 Hamburg, Germany

[3]Chair of Computational Systems Biology, University of Hamburg, Notkestreet 9-11, 22607 Hamburg, Germany

[4]Preventicum Essen, Theodor-Althoff-Strasse 47 45133 Essen and Preventicum Duesseldorf, Koenigsallee 11, 40212 Duesseldorf, Germany

[5]Independent researcher, Digital Health, Informatics and Data Science, Germany

[6]Department of Diagnostic and Interventional Radiology, Klinikum rechts der Isar, Technical University of Munich, Ismaningerstr. 22, 81675 Munich, Germany

[7]LipiTUM, Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Maximus-von-Imhof-Forum 3, 85354 Freising, Germany

[8]Computational BioMedicine Lab, Institute of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark

*** Correspondence:**

Justus Wolff

Justus.wolff@syte-institute.com

Abstract

The implementation of Artificial Intelligence (AI) still faces significant hurdles and one key factor is the access to data. One approach that could support that is federated machine learning (FL) since it allows for privacy preserving data access. For this proof of concept, a prediction model for coronary artery calcification scores (CACS) has been applied. The FL was trained based on the data in the different institutions, while the centralized machine learning model was trained on one allocation of data. Both algorithms predict patients with risk scores >=5 based on age, biological sex, waist circumference, dyslipidemia and HbA1c. The centralized model yields a sensitivity of c. 66% and a specificity of c. 70%. The FL slightly outperforms that with a sensitivity of 67% while slightly underperforming it with a specificity of 69%. It could be demonstrated that CACS prediction is feasible via both, a centralized and a FL approach, and that both show very comparable accuracy. In order to increase accuracy, additional and a higher volume of patient data is required and for that FL is utterly necessary. The developed “CACulator” serves as proof of concept, is available as research tool and shall support future research to facilitate AI implementation.

Keywords: Artificial intelligence, Privacy-preserving data processing, Federated machine learning, Coronary artery calcification

1 Introduction

Benefits and objectives

The real world usage of Artificial intelligence in healthcare routines is still in its beginning and several success factors for a facilitated implementation have been defined in the past.(1) One key success factor is the technological implementation, as the data access and algorithm training is crucial in order to generate reliable results. In our study a federated and centralized machine learning model approach for the early detection of coronary artery diseases (CADs) were compared with each other.

Physicians commonly assess the CAD risk of patients based on coronary artery calcification scores (CACSs), obtained via non-contrast computed tomography (CT) scans. However, CT produces high costs and exposes patients to radiation and, hence, unnecessary scans must be avoided. Therefore, an effective decision support tool predicting coronary artery calcification (CAC) risks and, thus, CT necessity can provide significant benefits.

In this study we aim to analyse 1) whether it is possible to predict CACSs with Artificial Intelligence (AI) and 2) how a privacy-preserving machine learning (ML) approach based on federated data sources, namely federated machine learning (FL), performs as compared to a conventional ML approach that is based on centralized data. Based on these results, improvement options for the particular indication and FL in general are discussed.

Coronary artery calcification scores (CACSs)

Cardiovascular diseases are the leading cause of death worldwide and coronary artery diseases (CADs) represent the leading cause of cardiovascular mortality.(2),(3) In order to counteract its significant mortality rates, treatment regimens should be adapted as early as possible.(4) For individuals with no symptoms or known pre-existing conditions, various factors (e.g. age, behavioral characteristics) can be used to determine the risk of a cardiovascular disease and increased CAC levels have been found to be significantly positively associated with cardiovascular diseases.(4) In this regard, prior research revealed that 27% of radiologists in the US already rely on CAC CT scans on a regular basis, making it the most common type of CT scan in the

country.(5) This does not seem surprising in light of the fact that there are guidelines recommending CAC CT screening for asymptomatic men aged 45-75 and woman aged 55-75 (except for those with very low risk).(5),(6),(7) This translates into c. 30 million citizens being generally eligible for CAC CT screening in the US.(8),(9) However, many of them may not actually have elevated CACSs, which causes the discussion on whether too many CACS CT scans are conducted.(10)

Application of Artificial Intelligence for CACSs prediction

Considering this risk/benefit trade-off, we assessed whether an Artificial Intelligence (AI)-based decision support tool could potentially contribute to reducing cases of CAC CT screening where patients do not actually have elevated CACSs. To this end, we used datasets from two medical institutions with a total of 1,450 patients and analyzed the following four independent CAC risk factor areas - those are also in detail explained in the supporting information section:

I) Age and biological sex

II) Obesity (measured through waist circumference)

III) Dyslipidemia (measured through cholesterol, triglycerides, high-density lipoprotein HDL, low-density lipoprotein LDL)

IV) Diabetes mellitus (measured through HbA1c)

Further potential risk factors were not included, as the selected factors are A) reflecting some of the major named risk factors in literature and B) the algorithm should serve as a tool for physicians in daily routine. (11)

We used AI for the purpose of our study as AI can lead to significant benefits in healthcare and has, for example, already proven to be valuable in the area of cardiology, but also in the context of other medical indications, such as Covid-19 as well as of personalized treatment.(12),(13),(14)

However, conventional AI approaches generally require access to large amounts of data to achieve precise prediction and are, therefore, often not easy to be implemented. To circumvent this issue, the concept of FL can be applied as this enables data access across different units without requiring the exchange of raw data,

thus, maintaining data privacy. Therefore, besides assessing whether AI can generally be useful to predict CACSs, we assessed how FL as a privacy-preserving ML approach performs as compared to a conventional ML approach.

The concept of FL goes back many decades, where it did not initially relate to AI, but was supposed to assist in the discovery and access of learning content from the diverse collection of content repositories.(15) Only in 2017, Google proposed to apply the concept of FL in the ML context where the main idea was to generate ML models based on datasets that are distributed across multiple devices, while at the same time preventing data leakage.(16)

Today, the term FL refers to AI model training based on multiple local datasets, yet without the exchange of the participating units' primary data but instead only model parameters. The key goal is to enable multiple actors to jointly build an aggregated ML model without facing the difficulties of data sharing.(17)

Altogether, an FL model represents a decentralized ML approach where the participating units' respective ML models are locally trained in a first step, and the model parameters are subsequently shared by all units, commonly via a central coordinating server, and merged into an aggregated model.(18),(19) Consequently, FL can provide significant advantages as compared to conventional ML approaches as the privacy-sensitive medical patient data must not be shared across participating units, enabling FL applications even where data-related restrictions prevent the application of conventional ML approaches.

There are already some first FL applications in healthcare, for example, in the context of the prediction of mortality and hospital stays, the medical diagnosis classification of diabetes and heart failure, and the prediction of the hospital readmission risk.(20),(21),(22),(23). One major real world example is Owkin, a startup which has raised > 300 Million USD in funding for their federated learning platform and which works for example in their EU project Melloddy in a collaboration with 10 pharmaceutical companies on a federated learning driven drug discovery process with molecule data (24). For the described CACS prediction, an AI model based on risk factors for elevated CACSs could serve as a valuable decision support tool and in order to circumvent data privacy challenges, the benefits of FL shall be leveraged.

Materials and Methods

To apply FL, we used the FeatureCloud platform, which targets to simplify privacy-preserving access to comprehensive medical data across medical institutions.(25) FeatureCloud, which is publicly available under http://featurecloud.ai, is supported by the European Union Horizon 2020 program and was developed as a joint effort between several European universities and companies, including the Technical University of Munich, the Philipps University of Marburg, the University of Maastricht, the Medical University of Graz, the University of Hamburg, the University of Southern Denmark and Gnome Design SRL, Romania.

After the FeatureCloud app has been installed by all participating medical institutions, the procedure is as follows - as also illustrated in Figure 1:

1. Obtaining patients' consent for the usage of their anonymized data
2. Creation of the FL project in FeatureCloud and assembling the workflow
3. Inviting other medical institutions to participate and starting the project

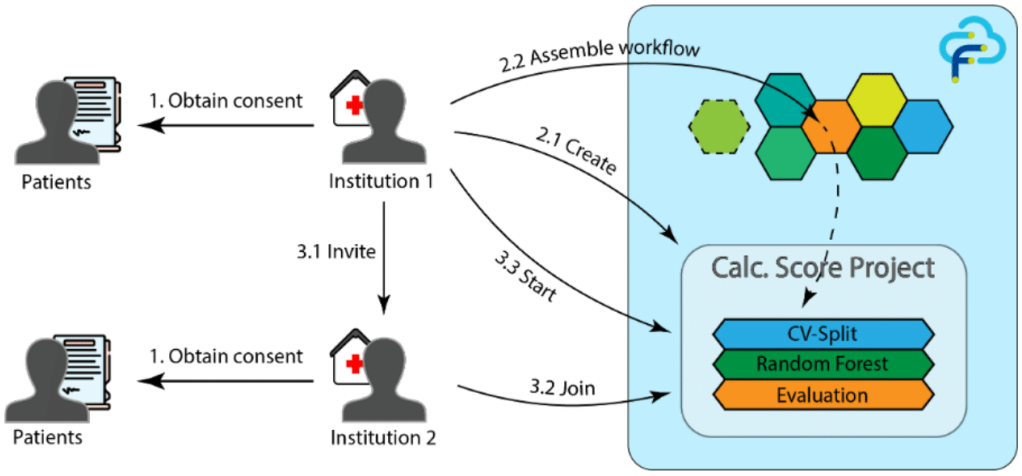


Figure 1. Two institutions used the FeatureCloud platform and Institution 1 has been the FL project leader (Authors)

Since we aimed at shedding light on whether it is possible to predict CAC with conventional AI and FL as a privacy-preserving ML approach performs as compared to a conventional ML approach, an according FeatureCloud app has been developed and can be accessed via https://featurecloud.ai/ai-store/71. FeatureCloud is an

integrative solution since multiple medical institutions can join the platform to share their data for AI model training and thus by each new institution the ML training data set is increased and higher prediction accuracies or additional prediction usecases can be defined – currently there are already more than 20 applications available on the FeatureCloud application store.

Study design - Conventional AI for CACS prediction

As to the first research question, we trained the algorithm for the decision support tool using an anonymized dataset of the above mentioned two medical institutions that in total comprises 1,450 patients. The pre-processing resulted in 680 remaining patient samples. For the centralized ML approach, the AI model was trained based on the entire dataset, i.e. on all patient samples of the two medical institutions, as if the data stemmed from the same source in the first place. Further details can be found in the supporting information section.

Study design - FL for CACS prediction

In addition, we analysed how FL performs as compared to a conventional ML approach, namely whether a similar performance in terms of sensitivity and specificity can be achieved when the AI model is trained in a FL (rather than a centralized) setting.

The local data is pre-processed by each medical institution before the local model training takes place. The model parameters are subsequently shared by both medical institutions and merged into an aggregated ML model. This aggregated model is then shared with both institutions for their own use. During the entire process, FeatureCloud does not require any transfer of the medical institutions' underlying primary medical data, but only the exchange of model parameters. Accordingly, in our particular use case, participating physicians would not need very extensive data of their own, but could use the aggregated ML model to predict patients' CACSs and, based on this, assess the need for CAC CT screening.

In order to demonstrate the likelihood analysis of a CACS >=5, we trained two random forest models locally, i.e. based on each of the medical institution's own local dataset. The resulting decision trees are visualized in Figure 2.

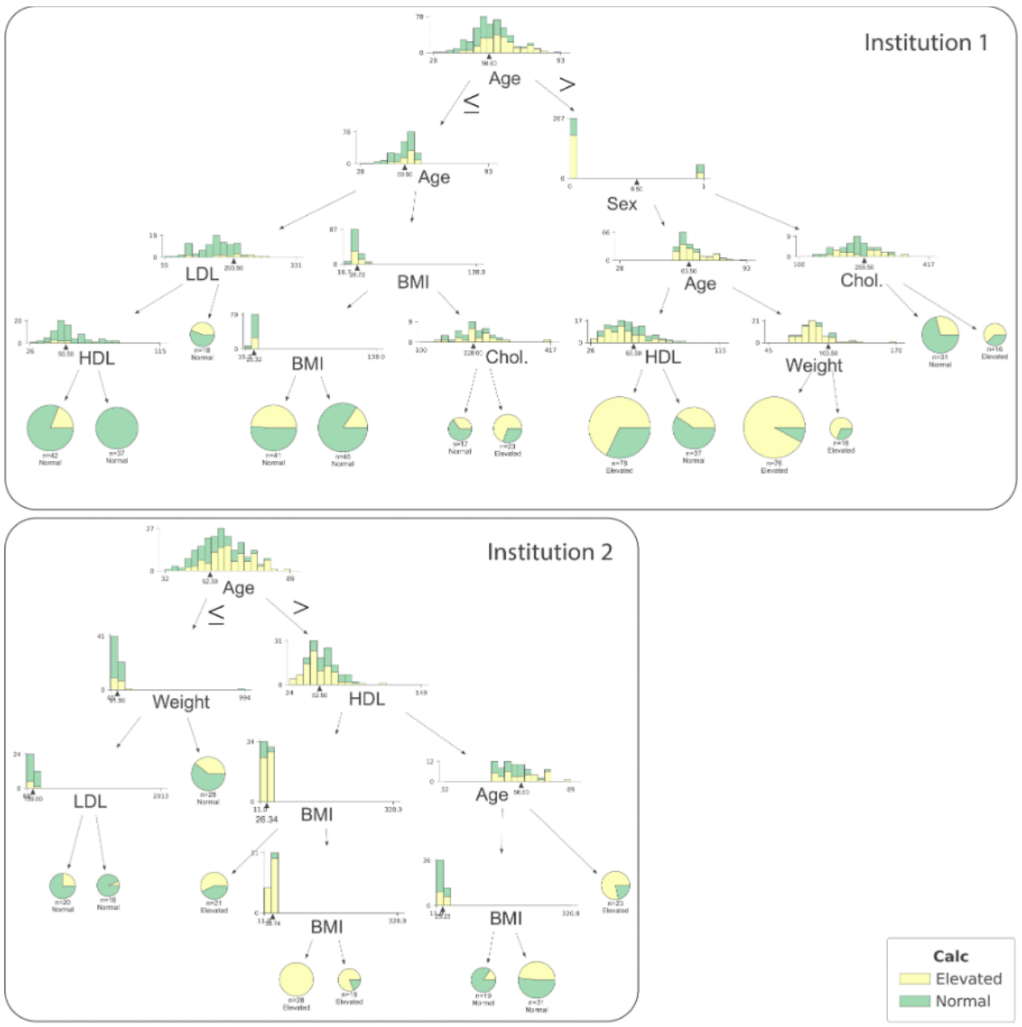


Figure 2. Decision trees trained with the data from Institution 1 and 2; both trees are first split by age, but already diverge at the next level (Authors)

The parameters of the locally trained AI models from both medical institutions are aggregated into one joint FL model, via sharing of the local model parameters rather than the medical institutions' primary data as it would be required when applying the conventional ML approach. Additional information regarding the training process is described in the following section.

For the conventional centralized ML model, a random forest model with all 680 patient samples (i.e. from both medical institutions) was trained. For the FL model, an aggregated random forest model was computed based on the model parameters of the locally trained models of the two medical institutions with 477 and 202 samples, respectively.

We pre-processed our data in terms of removing samples with missing values and transforming the continuous CACSs into two categories: Elevated and normal, representing values above or equal (CACS =>5) and below a CACS threshold of 5 (CACS <5), respectively. We investigated how the results change when replacing missing values with their mean or median values to allow for using incomplete samples during the training of the random forest model without distorting the results too heavily. This led to no improvement and was, therefore, not pursued for the subsequent steps. After removing the incomplete samples, 680 patients remained. The score threshold of 5 was chosen because it both indicates a health risk and divides the dataset into two equally large groups of 340 patients each.

For the classification, we selected a random forest model for our analysis as such models are frequently used as classifiers due to their ability to cope with categorical and continuous features alike and due to their general versatility. To conduct a FL analysis with the two participating medical institutions, we first trained a random forest at each institution locally using only their local data. To obtain the aggregated model, the decision trees of both random forests were merged into one random forest model. Institution 1 and 2 contributed 67 and 33 decision trees, respectively, proportional to their number of samples. During training, gini index minimization was used for splitting and the number of selected random features for each tree was 4.

To evaluate the models, we developed a tailored cross-validation which takes all possible combinations of splits into consideration (see figure 3). This type of cross-validation was performed ten times to further reduce random effects in the evaluation, leading to 1,000 test runs in total.

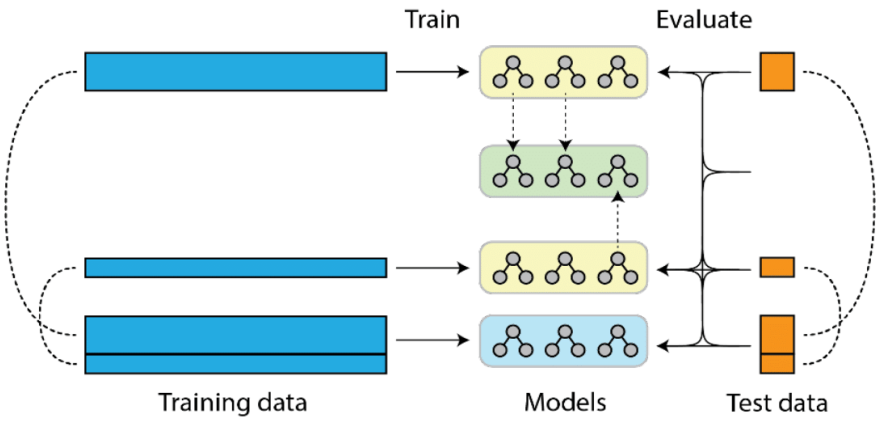


Figure 3. Training and evaluation of the FL model

The yellow random forests have been trained on the local data only. The green random forest model emerges from aggregation of the local random forest models. To reflect the different size of the training data, Institution 1 contributes twice as many trees to the aggregated model as Institution 2. The blue reference model is the conventional centralized ML model, which was trained on the whole dataset, i.e. of both institutions.

Each institution performed ten even splits on their data, as illustrated in Figure 4. To assess the performance of the aggregated model, each combination of splits from both institutions was used for validation, mirroring the conventional cross-validation approach in a FL federated setting, leading to 100 cross-validation steps.

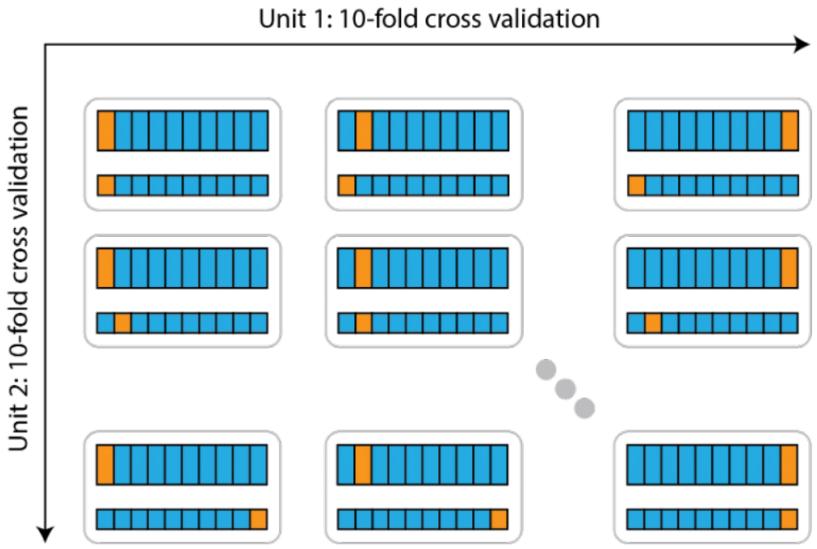


Figure 4. Cross-validation in the FL setting

The application is available on the FeatureCloud AI Store and allows for running the analysis for an arbitrary number of participants. Each participant contributes a number of decision trees proportional to their share of the total data. A web frontend allows for checking the number of valid samples and monitoring progress. The final accuracy, obtained through cross-validation, is displayed as well. The global random forest, consisting of all decision trees, can be downloaded by each participant for further evaluation or to predict CAC classes on unseen data.

Results

It has been assessed if and to what extend the existing "classical" analogue CACS assessments via CT screening could be improved through an AI based prediction model and the according processes as well as potential benefits are displayed in Figure 5:

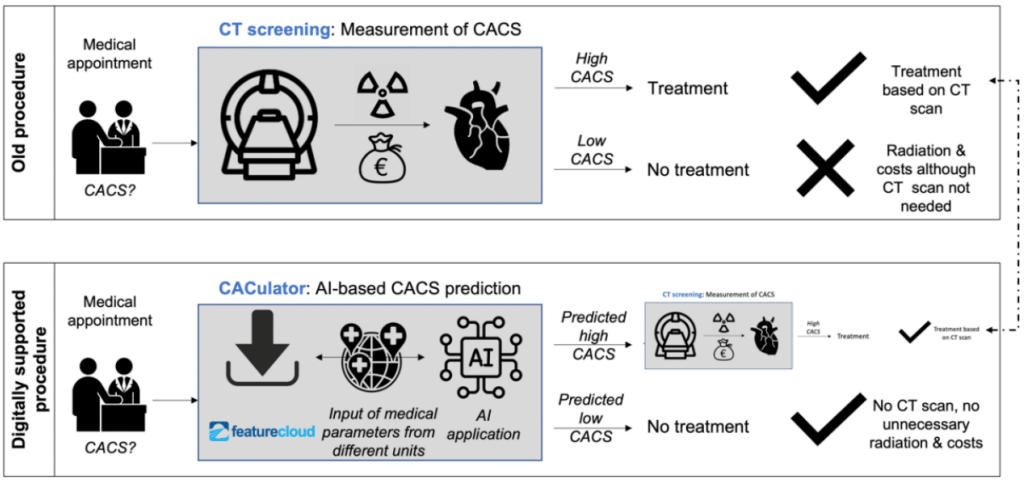


Figure 5: Comparison of the "analogue" and "digitally" supported CACS patient assessment and according benefits (Authors)

The results of both, the centralized ML and the FL approach and the respective sensitivity and specificity values are displayed in Table 1 and the median values of all patient characteristics are tabulated in Table 2:

[Insert Table 1 and 2 here: All tables below main text]

11

As visible, it can be derived that the centralized ML and the FL random forest models showed similar sensitivity of 65.5% and 66.7% as well as specificity of 69.7% and 68.6%, respectively.

With regard to the main research questions, the following conclusions can be drawn: 1) It is technically possible that an AI model is used for the prediction of CACS, which itself is based on the five risk factors age, sex, obesity, dyslipidemia and diabetes. Although the sensitivity and specificity is with ca. 70% not very high and further factors for improvement need to be identified and included, such an algorithm can already provide some value add, since existing CAC CT scan assessments entail a significant risk of erroneous patient risk group classification.

2) The FL-based decision support tool would essentially be comparably useful as both of the approaches, the centralized and FL approach, deliver very similar outcomes in terms of sensitivity and specificity. However, given FL's significant advantage of allowing privacy-preserving data access without actual primary data sharing across different institutions, FL might often represent the much more feasible approach in daily practice. Thus, FL can provide a significant benefit for future AI model training, especially where data is privacy-sensitive and scattered across medical institutions.

Discussion and conclusion

This publication focused on the data access and algorithm training through federated machine learning using the example of CACS prediction. While our results suggested that an AI-based decision support tool can generally be constructed and that the performance of the FL approach is similar to the one of the conventional ML approach, it should be highlighted that we did neither aim at creating a completely new risk indicator for the measurement of CACSs, i.e. choosing the optimal combination of risk factors, nor at replacing CAC CT scans or other assessment processes currently undertaken by physicians. Especially for the second one, a significantly higher accuracy of a prediction model would be required.

The obtained results with sensitivity and specificity levels of close to 70% still require further improvement in order to actually enable true benefits in clinical practice; suggestions to achieve such improvements are presented as part of this discussion.

Thus, in the following, 1) an embedding of the work in the research landscape, 2) improvement areas for the developed FL CACS prediction model, and 3) overarching facilitators for FL application are discussed and the results are aggregated in a conclusion and outlook.

Research landscape

Prior studies also reflected on FL and for example Chamikara et Al. highlighted that FL can require additional measures to guarantee data privacy or that computational IT bottlenecks can occur and thus supporting algorithms inside the FL platform can be applied (26).

One of these privacy measures can be so called “Differential Privacy”. Adnan et Al. have analyzed The Cancer Genomce Atlas dataset, which is particularly interesting since the images are derived through diverse imaging methods as well as devices and are marked differently. Since, as mentioned, FL can not provide a privacy guarantee because private information could in theory also be traced based on the shared model parameters, they applied differentially private FL. This framework quantifies the privacy of provided protocols and thus focusses on extracting as much information as possible while consuming the least privacy. (27)

Also the accuracy difference between centrally trained models and FL has been analyzed and Kirienko et Al. described in their literature research covering 26 publications that the prediction accuracy between both approaches is “equal” (28).

This has also been confirmed when FL was specifically applied for data access of electronic health records. Dang et. Al could show that it was possible to apply FL for in-hospital mortality prediction by analysing gender, ethnicity, lab tests for blood urea nitrogen, the Glasgow Coma Scale, surgery type, heart rate, blood pressure and others out of in total 82 parameters for their model training (29)

In addition to that, Zerka et Al. mentioned in their research that the convincing of medical centers to join a FL approach can be a hurdle and that also regulators need to be actively involved as suitable safeguards, since also with FL it remains a significant task to collect enough patient samples for ML training (30).

Furthermore, the analysis of CAC systems with support of technology has been researched in the past. Rogers and Aikawa for example focussed on both, the advances in molecular imaging and big data technology to map the disease more comprehensively. They described that AI models can improve diagnostics and risk assessment and that they could include additional biological data like vesicle release, mineral deposition or inflammation data as well as the analysis of omics data, which can contribute to a CAC development (31).

In addition to that, Sandstedt et Al. evaluated an AI based CACS software for the scanning of tomography scans and came to the conclusion that calcified lesions could be accurately detected by the AI compared with conventional methods. They also highlighted that AI could be a crucial success factor for this global disease burden and that it should be applied in the real world clinical setting (32).

Our publication focussed not on a tool for the assessment of diagnostic scans, but medical parameters as input factors that could even be analyzed before a CT scan is conducted and thus potentially the algorithm can enable patient screening without exposure to radiation and its according costs. The prediction accuracy of 70% is not yet high enough for a clinical day to day implementation and improvement areas for an increased prediction capability are described in the following paragraph.

Improvement areas for FL-based CACS prediction

The accurate FL prediction of elevated CACSs could provide a very meaningful value-add, as this would allow to adopt treatment regimens at an early stage.

There are various means by which the model's prediction power could be further improved, both quantitatively and qualitatively:

1) Quantitative enhancement measure: An extension of the 1,450 patient collective, out of which only 680 patient samples could be included after data pre-processing, could improve accuracy. This could be achieved through the inclusion of additional medical institutions which grant access to their respective patient data.
2) Qualitative enhancement measure: Several medical parameters can influence CACS and the prediction accuracy of the trained AI algorithms

demonstrates, that further parameters are required e.g. from medical practice and laboratories. The prediction model's accuracy could be improved through the inclusion of these additional CAC risk factor areas and extension examples could be family history, diet, hypertension, smoking, chronic kidney disease, psychosocial factors, elevated lipoprotein and elevated apolipoprotein.(33),(34)

Both of these measures shall be addressed through open access to the CACulator app on the FeatureCloud website (https://featurecloud.ai/ai-store/71). The input of additional and new model parameters by various medical institutions shall be fostered through this international research collaboration.

Furthermore, the developed FL model could be compared with other scores for the identification of cardiovascular risk, but the American College of Cardiology and the American Heart Association published in their review, that current risk scores were found to vary widely with regard to the populations from which they were derived, risk marker inputs/covariates, and outcomes of interest (35). One example is the ASCVD score – a calculation of the patient's 10-year risk of having a cardiovascular problem, such as a heart attack or stroke, but also this scores has pitfalls, like overestimation of 10 year risk (36).

Overarching facilitators for FL application

In addition to that, there are more general improvement areas that could increase the real-world application of FL in healthcare, irrespective of this particular use case:

First, easy access to the software should be ensured by implementing it into daily medical routines, e.g. in the form of a smartphone application or by integration into the electronic health record (EHR) system of the participating medical institutions. To this end, several practical challenges need to be overcome. For example, the software would need to comply with the respective regulatory requirements, such as the GDPR in Europe or the FDA guidelines in the US, as well as medical product class guidelines. Furthermore, the solution would need to fulfil the requirements for app downloads (e.g. in the Google Playstore or Apple's App Store) and the requirements of EHR providers (e.g. Cerner or Siemens).

An additional practical challenge for the application of FL is that its implementation requires human and financial resources on the clinical side and is, thus, also a business / economic decision. In order to also provide a compelling case from this perspective, the economic benefits and costs of implementing the FL solution should be assessed and measured. Such quantitative evidence would often contribute to a faster scaling of the respective solution and additionally pave the way for public grants and for better access to external resources.(37) This commercial aspect is also reflected in the fact that FL research is currently mostly driven by large and primarily tech-oriented industry players and not academia.(38) An increased amount of research by industry-independent scientific institutions would likely entail a number of benefits, including increased trust and application levels.

Furthermore, when applying FL in this research project, we also identified several administrative steps that would significantly contribute to achieving higher levels of FL application in real-world contexts in the future:

1) Implementation: The first installation of FeatureCloud or a similar solution has to comply with legal standards, e.g. in the form of a contract between the medical institution and the software provider. This process should follow an automatized legal compliance process adapted to every country's legal framework. It could even differentiate between research and operational implementation.(39) Furthermore, the software installation needs to be compatible with the local IT infrastructure specifications (e.g. firewalls) of the participating medical institutions. To avoid technical issues and lengthy and complex adaptation processes as well as mitigate the risk of delays, respective guidelines are required and should be made available upfront.

2) Understandability: The processes for the application of the different ML models should be easy to understand and designed in a way for them to be initiated as well as steered by the participating medical institutions' staff. To this end, explanations and training materials need to be provided by the FL provider. Potentially, even governmental institutions may get involved and support this process as this could further increase the trust level as to the application of AI in healthcare.

3) Scaling: AI can and should be applied in the context of different medical indications. Patient data could be collected and, at best, be evaluated automatically for different medical indications (e.g. asthma, diabetes and electric implant monitoring). This requires the input of key decision-makers on medical and business management level and, as a basis for their decision-making, an objective and comprehensive overview of the opportunities and risks of AI in healthcare is needed. This could be enabled by a simple-to-use “entry point”, such as a functionality that evaluates existing datasets with regard to their applicability for an AI use case.

Progress in these improvement areas, namely the practical IT implementation in existing EHR structures, measurement of medical and economic impact as well as administrative support tools for implementation, understandability and scaling could significantly contribute to an increased usage of FL.

Conclusion and outlook

An overall facilitated implementation of AI in healthcare depends on several factors, but with regard to technological implementation, the access to data for the application of algorithms is crucial. This publication focused on the data access and algorithm training through federated machine learning using the example of CACS prediction.

It can be concluded that, AI can be used for CACSs prediction with a moderate accuracy using the mentioned five data points (age, sex, obesity, dyslipidemia, diabetes) and that a decentralized FL approach, which allows for privacy-preserving data access across medical institutions, demonstrated a similar performance as a conventional ML approach.

The current model performance of the proof of concept is still too limited for a clinical setting and further improvements are needed to allow for clinical implementation. The prediction accuracy of the CACS model could be improved through a) a more comprehensive patient collective and b) an extension to further CAC risk factor areas. These improvements require data access across institutions and are only possible in a privacy preserving context, thus federated machine learning can be a very important success factor regarding AI application for CACS prediction.

In addition to the use case-specific improvement areas, we also identified several general improvement areas that could increase the real-world application of FL in healthcare: First, medical institutions should be able to access the FL software in the context of daily medical routines and electronic patient records. Second, a reliable economic impact assessment is needed to support the strategic decision to apply FL, especially given the required medical team and financial resources. Third, administrative support in terms of legal and technological implementation standards, training material for the participating medical institutions' staff and simple data validation mechanisms for the verification of AI suitability can facilitate implementation.

Overall, it can be concluded that FL can provide significant value across different indications as it allows to exploit data from different sources in AI model training via its privacy-preserving design and can thus support an overall increased implementation of AI in healthcare.

Tables:

Metric	Institution 1	Institution 2	Centralized	Federated
Accuracy	68.12 %	64.71 %	67.65 %	67.65 %
ROC AUC	75.22 %	71.14 %	75.52 %	75.09 %
Sensitivity	64.86 %	66.67 %	65.52 %	66.67 %
Specificity	71.61 %	63.64 %	69.70 %	68.57 %
Samples	477	202	680	477 + 202

Table 1. Results of the centralized and FL analysis, showing the mean accuracy, ROC AUC, sensitivity, specificity and number of patient samples for Institution 1 and 2 individually, the FL model and the centralized ML model

	All	CACS < 5	CACS >= 5	Inst. 1	Inst. 2
Age	57.0	54.0	61.0	57.0	57.0
Height	179.0	179.0	178.0	179.0	178.0
Weight	87.0	86.0	87.0	87.0	86.0
Waist	98.0	96.0	101.0	99.0	98.0
Chol.	228.0	230.0	224.0	228.0	224.5
Tri.	129.0	122.0	136.0	127.0	131.0
HDL	51.0	53.0	49.0	50.0	53.0
LDL	156.0	158.0	152.0	160.0	149.5
HBA	5.5	5.4	5.6	5.5	5.5

Table 2. Median values for age, height, weight, waist, cholesterol, triglycerides, HDL, LDL, HBA and BMI in the complete group, the group of patients with a CACS below and above 5, and the patients from Institution 1 and Institution 2 individually

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 826078 (J.B., J.M.) and No 777111 (J.B.). This publication reflects only the authors' view, and the European Commission is not responsible for any use that may be made of the information it contains. J.B. was partially funded by his VILLUM Young Investigator Grant no.13154. Contributions by J.K.P. are funded by the Bavarian State Ministry of Science and the Arts within the framework of the Bavarian Research Institute for Digital Transformation (bidt).

Data availability

The non-confidential data underlying this article will be shared on reasonable request to the corresponding author. The code of the FeatureCloud app can be accessed via https://github.com/featurecloud/cacs_forest.

Conflict of Interest

Authors state no conflict of interest. All authors have read the journal's publication ethics and publication malpractice statement available at the journal's website and hereby confirm that they comply with all its parts applicable to the present scientific work.

Ethical Committee

Ärztekammer Nordrhein

Tersteegenstraße 9, 40474 Duesseldorf

Germany

References

1. Wolff J, Pauling J, Keck A, Baumbach J. Success Factors of Artificial Intelligence Implementation in Healthcare. Front Digit Health. 2021 Jun 16;3:594971.

2. Cardiovascular diseases (CVDs) [Internet]. [cited 2021 Feb 20]. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

3. Okrainec K, Banerjee DK, Eisenberg MJ. Coronary artery disease in the developing world. Am Heart J. 2004 Jul;148(1):7–15.

4. Correction to: Use of Risk Assessment Tools to Guide Decision-Making in the Primary Prevention of Atherosclerotic Cardiovascular Disease: A Special Report From the American Heart Association and American College of Cardiology [Internet]. Vol. 139, Circulation. 2019. Available from: http://dx.doi.org/10.1161/cir.0000000000000701

5. Burger IM, Kass NE, Sunshine JH, Siegelman SS. The Use of CT for Screening: A National Survey of Radiologists' Activities and Attitudes [Internet]. Vol. 248, Radiology. 2008. p. 160–8. Available from: http://dx.doi.org/10.1148/radiol.2481071369

6. Naghavi M, Falk E, Hecht HS, Jamieson MJ, Kaul S, Berman D, et al. From Vulnerable Plaque to Vulnerable Patient—Part III: Executive Summary of the Screening for Heart Attack Prevention and Education (SHAPE) Task Force Report [Internet]. Vol. 98, The American Journal of Cardiology. 2006. p. 2–15. Available from: http://dx.doi.org/10.1016/j.amjcard.2006.03.002

7. Kim KP, Einstein AJ, Berrington de González A. Coronary artery calcification screening: estimated radiation dose and cancer risk. Arch Intern Med. 2009 Jul 13;169(13):1188–94.

8. Elias-Smale SE, Proença RV, Koller MT, Kavousi M, van Rooij FJA, Hunink MG, et al. Coronary calcium score improves classification of coronary heart disease risk in the elderly: the Rotterdam study. J Am Coll Cardiol. 2010 Oct 19;56(17):1407–14.

9. Brugts JJ, Yetgin T, Hoeks SE, Gotto AM, Shepherd J, Westendorp RGJ, et al. The benefits of statins in people without established cardiovascular disease but with cardiovascular risk factors: meta-analysis of randomised controlled trials [Internet]. Vol. 338, BMJ. 2009. p. b2376–b2376. Available from: http://dx.doi.org/10.1136/bmj.b2376

10. Coronary artery calcium score: Are we doing too many or too few? - Mayo Clinic [Internet]. [cited 2021 Apr 21]. Available from: https://www.mayoclinic.org/medical-professionals/cardiovascular-diseases/news/coronary-artery-calcium-score-are-we-doing-too-many-or-too-few/mcc-20438011

11. Kronmal RA, McClelland RL, Detrano R, Shea S, Lima JA, Cushman M, et al. Risk factors for the progression of coronary artery calcification in asymptomatic subjects: results from the Multi-Ethnic Study of Atherosclerosis (MESA). Circulation. 2007 May 29;115(21):2722–30.

12. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial Intelligence in Cardiology. J Am Coll Cardiol. 2018 Jun 12;71(23):2668–79.

13. Panicacci S, Donati M, Profili F, Francesconi P, Fanucci L. Trading-Off Machine Learning Algorithms towards Data-Driven Administrative-Socio-Economic Population Health Management [Internet]. Vol. 10, Computers. 2020. p. 4. Available from: http://dx.doi.org/10.3390/computers10010004

14. van der Schaar M, Alaa AM, Floto A, Gimson A, Scholtes S, Wood A, et al. How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. Mach Learn. 2020 Dec 9;1–14.

15. Rehak D Dodds P Lannom. A model and infrastructure for federated learning content repositories. Interoperability of web-based educational systems workshop, vol 143. 2005;

16. Blog [Internet]. [cited 2021 Jan 4]. Available from: http://ai.googleblog.com/2017/04/federated-learning-collaborative.html

17. Asad M, Moustafa A, Yu C. A Critical Evaluation of Privacy and Security Threats in Federated Learning [Internet]. Vol. 20, Sensors. 2020. p. 7182. Available from: http://dx.doi.org/10.3390/s20247182

18. Konečný J, Brendan McMahan H, Ramage D, Richtárik P. Federated Optimization: Distributed Machine Learning for On-Device Intelligence [Internet]. arXiv [cs.LG]. 2016. Available from: http://arxiv.org/abs/1610.02527

19. Konečný J, Brendan McMahan H, Yu FX, Richtárik P, Suresh AT, Bacon D. Federated Learning: Strategies for Improving Communication Efficiency [Internet]. arXiv [cs.LG]. 2016. Available from: http://arxiv.org/abs/1610.05492

20. Huang L, Shea AL, Qian H, Masurkar A, Deng H, Liu D. Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. J Biomed Inform. 2019 Nov;99:103291.

21. Grama M, Musat M, Muñoz-González L, Passerat-Palmbach J, Rueckert D, Alansary A. Robust Aggregation for Adaptive Privacy Preserving Federated Learning in Healthcare [Internet]. arXiv [cs.CR]. 2020. Available from: http://arxiv.org/abs/2009.08294

22. Min X, Yu B, Wang F. Predictive Modeling of the Hospital Readmission Risk from Patients' Claims Data Using Machine Learning: A Case Study on COPD [Internet]. Vol. 9, Scientific Reports. 2019. Available from: http://dx.doi.org/10.1038/s41598-019-39071-y

23. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. N Engl J Med. 2019 Nov 14;381(20):1909–17.

24. European Commission: Available on https://www.melloddy.eu;

https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5d78d81a9&appId=PPGMS

25. Matschinske J, Späth J, Nasirigerdeh R, Torkzadehmahani R, Hartebrodt A, Orbán B, et al. The FeatureCloud AI Store for Federated Learning in Biomedicine and Beyond [Internet]. arXiv [cs.LG]. 2021. Available from: http://arxiv.org/abs/2105.05734

26. Chamikara MAP, Bertok P, Khalil I, Liu D, Camtepe S. Privacy preserving distributed machine learning with federated learning. Comput Commun. 2021 Apr;171:112–25.).

27. Adnan M, Kalra S, Cresswell JC, Taylor GW, Tizhoosh HR. Federated learning and differential privacy for medical image analysis. Sci Rep. 2022 Feb 4;12(1):1953.

28. Kirienko M, Sollini M, Ninatti G, Loiacono D, Giacomello E, Gozzi N, et al. Distributed learning: a reliable privacy-preserving strategy to change multicenter collaborations using AI. Eur J Nucl Med Mol Imaging. 2021 Nov;48(12):3791–804.)

29. Dang TK, Lan X, Weng J, Feng M. Federated learning for electronic health records. ACM Trans Intell Syst Technol. 2022 Oct 31;13(5):1–17.

30. Zerka F, Barakat S, Walsh S, Bogowicz M, Leijenaar RTH, Jochems A, et al. Systematic Review of Privacy-Preserving Distributed Machine Learning From Federated Databases in Health Care. JCO Clin Cancer Inform. 2020 Mar;4:184–200.)

31. Rogers MA, Aikawa E. Cardiovascular calcification: artificial intelligence and big data accelerate mechanistic discovery. Nat Rev Cardiol. 2019 May;16(5):261–74.

32. Sandstedt M, Henriksson L, Janzon M, Nyberg G, Engvall J, De Geer J, et al. Evaluation of an AI-based, automatic coronary artery calcium scoring software. Eur Radiol. 2020 Mar;30(3):1671–8.)

33. Fruchart J-C. New Risk Factors for Atherosclerosis and Patient Risk Assessment [Internet]. Vol. 109, Circulation. 2004. p. III – 15. Available from: http://dx.doi.org/10.1161/01.cir.0000131513.33892.5b

34. New aspects of the risk assessment guidelines: Practical highlights, scientific evidence and future goals - American college of cardiology [Internet]. [cited 2021 Jan 4]. Available from: https://www.acc.org/latest-in-cardiology/articles/2018/11/14/07/10/new-aspects-of-the-risk-assessment-guidelines

35. Strengths and Limitations of the ASCVD Risk Score and What Should Go in the Risk Discussion - American College of Cardiology [Internet]. [cited 2021 Nov 22]. Available from: http%3a%2f%2fwww.acc.org%2flatest-in-cardiology%2farticles%2f2014%2f08%2f25%2f14%2f48%2fstrengths-and-limitations-of-the-ascvd-risk-score-and-what-should-go-in-the-risk-discussion

36. ASCVD (Atherosclerotic Cardiovascular Disease) 2013 Risk Calculator from AHA/ACC [Internet]. [cited 2021 Nov 22]. Available from:

https://www.mdcalc.com/ascvd-atherosclerotic-cardiovascular-disease-2013-risk-calculator-aha-acc

37. Wolff J, Baumbach J, Pauling J, Keck A. A Systematic Review of Economic Impact Studies of Artificial Intelligence in Healthcare (Preprint) [Internet]. Available from: http://dx.doi.org/10.2196/preprints.16866

38. Lo SK, Lu Q, Wang C, Paik H-Y, Zhu L. A Systematic Literature Review on Federated Machine Learning: From A Software Engineering Perspective [Internet]. arXiv [cs.SE]. 2020. Available from: http://arxiv.org/abs/2007.11354

39. Zerka F, Barakat S, Walsh S, Bogowicz M, Leijenaar RTH, Jochems A, et al. Systematic Review of Privacy-Preserving Distributed Machine Learning From Federated Databases in Health Care. JCO Clin Cancer Inform. 2020 Mar;4:184–200.

40. Raggi P. Prognostic implications of absolute and relative calcium scores. Herz. 2001 Jun;26(4):252–9.

41. Williams JK, Adams MR, Klopfenstein HS. Estrogen modulates responses of atherosclerotic coronary arteries. Circulation. 1990 May;81(5):1680–7.

42. Manson JE, Allison MA, Rossouw JE, Carr JJ, Langer RD, Hsia J, et al. Estrogen therapy and coronary-artery calcification. N Engl J Med. 2007 Jun 21;356(25):2591–602.

43. Coutinho T, Goel K, Corrêa de Sá D, Kragelund C, Kanaya AM, Zeller M, et al. Central obesity and survival in subjects with coronary artery disease: a systematic review of the literature and collaborative analysis with individual subject data. J Am Coll Cardiol. 2011 May 10;57(19):1877–86.

44. Flint AJ, Rexrode KM, Hu FB, Glynn RJ, Caspard H, Manson JE, et al. Body mass index, waist circumference, and risk of coronary heart disease: A prospective study among men and women [Internet]. Vol. 4, Obesity Research & Clinical Practice. 2010. p. e171–81. Available from: http://dx.doi.org/10.1016/j.orcp.2010.01.001

45. Li TY, Rana JS, Manson JE, Willett WC, Stampfer MJ, Colditz GA, et al. Obesity as compared with physical activity in predicting risk of coronary heart disease in women. Circulation. 2006 Jan 31;113(4):499–506.

46. Logue J, Murray HM, Welsh P, Shepherd J, Packard C, Macfarlane P, et al. Obesity is associated with fatal coronary heart disease independently of traditional risk factors and deprivation. Heart. 2011 Apr;97(7):564–8.

47. Barry Levine T, Levine AB. Metabolic Syndrome and Cardiovascular Disease. John Wiley & Sons; 2012. 512 p.

48. De Schutter A, Lavie CJ, Milani RV. The impact of obesity on risk factors and prevalence and prognosis of coronary heart disease-the obesity paradox. Prog Cardiovasc Dis. 2014 Jan;56(4):401–8.

49. Lofgren I, Herron K, Zern T, West K, Patalay M, Shachter NS, et al. Waist circumference is a better predictor than body mass index of coronary heart disease risk in overweight premenopausal women. J Nutr. 2004 May;134(5):1071–6.

50. Chen Q, Li L, Yi J, Huang K, Shen R, Wu R, et al. Waist circumference increases risk of coronary heart disease: Evidence from a Mendelian randomization study. Mol Genet Genomic Med. 2020 Apr;8(4):e1186.

51. van Setten J, Išgum I, Pechlivanis S, Tragante V, de Jong PA, Smolonska J, et al. Serum lipid levels, body mass index, and their role in coronary artery calcification: a polygenic analysis. Circ Cardiovasc Genet. 2015 Apr;8(2):327–33.

52. Alamir MA, Goyfman M, Chaus A, Dabbous F, Tamura L, Sandfort V, et al. The Correlation of Dyslipidemia with the Extent of Coronary Artery Disease in the Multiethnic Study of Atherosclerosis [Internet]. Vol. 2018, Journal of Lipids. 2018. p. 1–9. Available from: http://dx.doi.org/10.1155/2018/5607349

53. Raggi P, Shaw LJ, Berman DS, Callister TQ. Prognostic value of coronary artery calcium screening in subjects with and without diabetes [Internet]. Vol. 43, Journal of the American College of Cardiology. 2004. p. 1663–9. Available from: http://dx.doi.org/10.1016/j.jacc.2003.09.068

54. Rumberger JA, Simons DB, Fitzpatrick LA, Sheedy PF, Schwartz RS. Coronary artery calcium area by electron-beam computed tomography and coronary atherosclerotic plaque area. A histopathologic correlative study. Circulation. 1995 Oct 15;92(8):2157–62.

Supporting information

Medical parameters

This section covers the relationship between the included models parameters and the CACS and outlines our rationale for including them in our prediction model:

I) Age and biological sex

The extent of CAC is strongly associated with age in men and women. Calcification is first detected in most men at around age 40, whereas women first show calcification around age 50.(40) The incidence in women is delayed by 10 to 15 years as compared to men, likely attributable to the protective effects of estrogen.(41)

The effect of estrogen and, also, according therapy on CAC has been examined, for example, for women aged 50 to 59 years during a 7-year period of treatment. The CAC

was significantly lower for women assigned to estrogen substitution as compared to those receiving placebo.(42)

II) Obesity (measured through waist circumference)

Further, obesity was shown to be strongly associated with an elevated risk of chronic heart diseases across several clinical studies.(43),(44),(45),(46),(47),(48) In this regard, the waist circumference, rather than the body mass index, is more likely directly associated with mortality.(49),(50)

III) Dyslipidemia (measured through cholesterol, triglycerides, HDL, LDL)

A prior study assessing CAD risk scores for 2,599 participants of the Dutch-Belgian Lung Cancer Screening Trial revealed nominally significant associations for genetic risk scores of low-density lipoprotein-cholesterol, total cholesterol, and obesity.(49)

Furthermore, in a multi-ethic cross-sectional analysis, 4,917 atherosclerosis participants were classified into six groups defined by specific LDL-c, HDL-c, or triglyceride cut-off points. Multivessel CAC was defined as the involvement of at least two coronary arteries. This study revealed that all groups except for the one exhibiting hypertriglyceridemia had statistically significant prevalence ratios of having multivessel CAC as compared to the normo lipidemia group.(52)

IV) Diabetes mellitus (measured through Hba1c)

CAC tends to be higher in diabetic patients and represents an independent risk factor for adverse outcomes.(53),(55) In different studies, the average CACS for subjects with and for those without diabetes was 281 +/- 567 and 119 +/- 341, respectively. Also, the death rate was 3.5% and 2.0% for subjects with and without diabetes, respectively. In a risk-factor-adjusted model, there was a significant interaction of CACSs with diabetes, indicating that, for every increase in CACS, there was a greater increase in mortality for diabetic than for non-diabetic subjects. However, patients suffering from diabetes with no CAC demonstrated a survival rate similar to the one of individuals without diabetes and no detectable CAC (98.8% and 99.4%, respectively).

6. Discussion

This chapter is split into a discussion of (1) academic research about AI implementation in healthcare, (2) the identified success factor categories, i.e., risk-allowing policy framework, privacy-preserving data access, evidence on impact, and (3) limitations of this thesis.

Academic research

The implementation of AI in healthcare can contribute to solve significant medical and economic challenges, for example to cope with the growing need for healthcare services of an ageing population, the shortage of medical staff and the increasing amount of healthcare expenses. It has also been the subject of numerous academic research projects in the past.

For example, Park et al. argue that for a range of AI applications, from digital secretaries over voice recognition to predictive modeling, further action is needed in terms of a) a better utilization of healthcare data, especially by tackling the privacy issue, b) adequate policies for new devices, and c) the prevention of safety and liability issues (Park et al. 2020). Also, concrete roadmaps for building effective, reliable and safe AI systems have been developed in previous publications. For example, Bajwa et al. propose a process that consists of the following steps/phases: Design and development, stakeholder engagement, human-centered AI approach, experimentation, validation, scaling and maintenance (Bajwa et al. 2021). Another approach is the design-thinking mixed methods approach by Smith et al. The authors propose for the process to be broken down into four steps, namely plan, do, study and adjust, where each AI implementation team shall conduct as many cycles as necessary to refine the workflow and model in order to successfully implement an AI solution (Smith et al. 2021).

While these studies also presume a high potential for AI in healthcare if the necessary actions are undertaken to unlock its full potential, other authors are rather critical about the potential scope and value-added of AI. For example, Shaw et al. consider a breakthrough of AI in healthcare in the short term only likely in the form of machine learning applications for the purpose of decision-making support. However, the authors argue that even for that, first, an appropriate environment, for example, with view to privacy and scalability, needs to be established (Shaw et al. 2019). Also, Gama et al. analyzed in a review implementation frameworks for AI and came to the conclusion that many existing AI implementation frameworks do not fully include the unique requirements that AI require and they propose to leverage existing knowledge from implementation science as well as significantly increase empirical research in this area for implementation uptake (Gama et. al 2022).

The existing research shows product focussed improvement areas, workflow suggestions and expected limitations and this thesis tries to approach the gap between the promising academic research and the low practical implementation systematically by examining both, AI's potential from a multi-stakeholder perspective and real world cases. The different success factor categories which have been identified are discussed in the following.

Risk-allowing policy framework

The healthcare industry is generally highly regulated and, thus, policy frameworks play a key role when developing healthcare hardware and software solutions. As one example, medical product class certifications like the MDR in Europe affect numerous steps from R&D to admission including clinical trial procedures, reimbursement as well as data safety and interoperability.

Therefore, it is not surprising that first attempts towards AI policies can already be observed, for example through US and EU frameworks. While both have an AI policy framework, it must be noted that these differ significantly:

The EU published for example two key documents: The Commission Whitepaper “On Artificial Intelligence – A European approach to excellence and trust” from 19 February 2020 and the “Proposal for a regulation of the European Parliament and of the Council - Laying down harmonized rules on Artificial Intelligence and amending certain Union Legislative Acts” from 21 April 2021 (EU Commission 2020, EU Commission 2021). With regard to the first document, the aim of the EU was adequate compensation in the event of damage, fair distribution of liability and no worse treatment of victims of AI systems compared to those of other products. The EU Commission plans its own proposal to initiate the legislative process from 2022. The objective of the second document was to create a legal framework for trustworthy intelligence while enhancing European competitiveness. There is a discussion of the draft and opinions in the European Parliament and the European Council, although the entry into force is not expected before 2023 (CMS 2021).

The EU applies in these documents a hazard/risk-based approach in which a differentiation of AI applications takes place between A) low or minimal risk, B) high risk and C) unacceptable risk. Possible penalties for serious infringements (use of prohibited AI system practices and non-compliance with the quality criteria for data used) can be up to EUR 30 million or 6% of the last year’s global turnover of a company.

The following differentiation of AI systems shall take place:

- Low or minimal risk: Can be applications like chatbots, where the users themselves still can take decisions on their own
- High risk:
  - A) Risk to the health or safety of natural persons, e.g., through biomedical identification, access to critical infrastructure, HR applications, law enforcement or administration of democratic processes
  - B) AI systems as safety components of products covered by certain sectoral Union legislation and as such products themselves, e.g., medical devices or in vitro diagnostics
- Unacceptable risk: A use is considered unacceptable as contravening Union values, for instance, by violating fundamental rights, e.g., manipulation of persons beyond consciousness or exploitation of vulnerabilities of specific groups such as children

In order to assess the risk, a conformity assessment shall take place with strict and binding requirements regarding: Quality of the data sets used, technical documentation and other records, transparency and provision of user information and human oversight as well as robustness, accuracy and cybersecurity.

The process is displayed as followed:

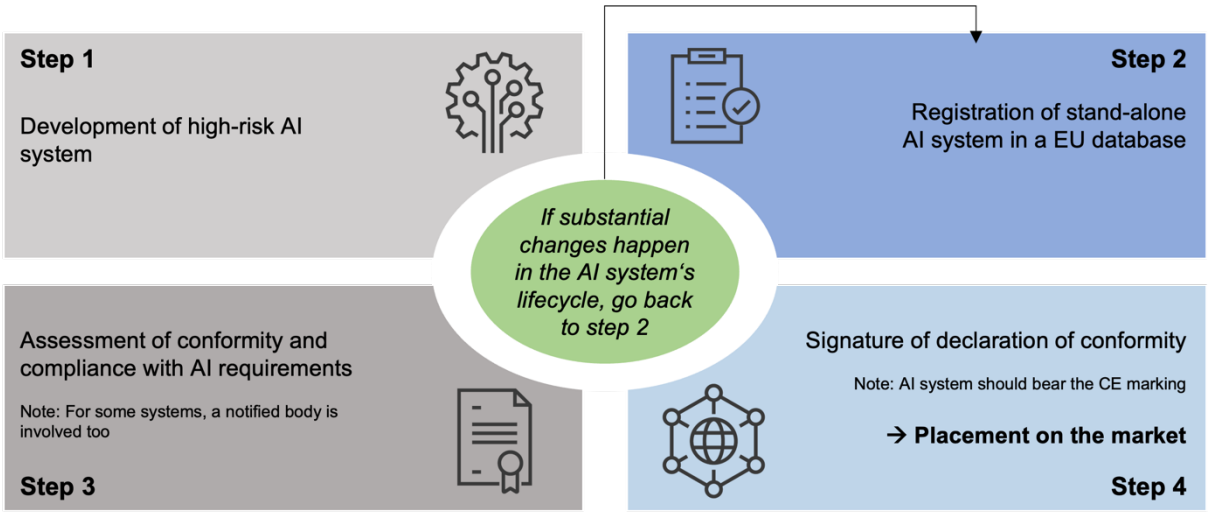


Figure 4: Process of market placement of a high-risk AI system by the EU (CMS 2021, adapted by author)

The process raises concerns of challenging requirements, especially for Small and Medium Sized Enterprises, as well as its wide room for interpretation/judgment in terms of classification.

In the US, there is currently no federal regulation on AI, but there is an agency-by-agency approach, with several institutions that publish guidelines and documents. For example, the US Food and Drug Administration published the "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning - Based Software as Medical Device" in 2019 (U.S. Food and Drug Administration 2019). This document refers to:

- Establishment of Quality Systems and Good Machine Learning Practices (GMLP), including usage of only relevant data, the separation between training, tuning and test datasets or transparency of the output
- Conduction of initial pre-market reviews to assure safety and effectiveness
- Monitoring of the AI devices based on development, validation, and execution of algorithm changes such as “Algorithm Change Protocol”
- Post-market real-world evidence performance reporting for maximized safety and effectiveness

The White House (via the Trade and Technology Council (TTC)) stated, that it is committed to cooperate on developing “AI systems that are innovative and trustworthy and that respect universal human rights and shared democratic values” as well as to “uphold and implement the OECD Recommendation on Artificial Intelligence” and to discuss “measurement and evaluation tools (...) to assess the technical requirements for trustworthy AI” (Orric 2021).

The Department of Commerce (DoC) demonstrated through the National Institute of Standards and Technology (NIST) a risk management framework: They announced the development of the “AI Risk Management Framework”, which could influence how companies and organizations approach AI-related risks, including avoiding bias and promoting accuracy, privacy, and security. Furthermore, the NIST established a National Artificial Intelligence Advisory Committee (NAIAC) and will “advise the President and other federal agencies on a range of issues related to artificial intelligence” and will offer recommendations on the “current state of U.S. AI competitiveness, the state of science around AI, issues related to the AI workforce” (US Department of Commerce 2021).

As demonstrated in the approaches of the EU and US, the risk-allowing policy frameworks are still under development and are vital components for implementation, because they influence the deployment process. Policy frameworks need to address specificities of the healthcare industry with a crucial balance between data security and privacy as well as feasibility of technological progress and AI application. In a similar vein, especially in healthcare systems with statutory health insurance,

appropriate organizational structures e.g., for market access and reimbursement need to be established.

While industrial institutions have some freedom with regard to the policy environment (e.g., when choosing where to locate their research centers or choosing the market for a healthcare solution), there is a need for action by governments to facilitate AI implementation in healthcare internationally. Indeed, looking at the population growth development and the need for healthcare globally, AI will play a crucial role in public policy. Based on the current trends and needs of the global population, by 2030, there will be ca. 10 million fewer healthcare professionals, including ca. 5 million fewer doctors than society will require (World Health Organization 2016). As one country example, until 2030, the gap between supply of and demand for staff employed by National Health Service trusts in the United Kingdom (UK) could increase to almost 250,000 full-time equivalent positions (The Health Foundation 2018). Furthermore, the World Health Organization has predicted that by 2030, 30% of global death will be caused by lifestyle diseases and this can be prevented with an appropriate identification of associated risk factors and intervention plans, where especially behavioral change plays a major role (Chatterjee, Gerdes, and Martinez 2020). Key associated indications are cardiovascular diseases, obesity and diabetes type II and for each of those, AI can be vital for their early diagnosis and respective treatment. In this context and against the background of the clear and global need for AI in healthcare, it is essential that governments create frameworks that facilitate AI implementation across the medical value chain, i.e., from R&D over market access to scaling.

Privacy-preserving technology

As to the technological implementation, it has become evident that a key challenge in healthcare is that on the one hand AI requires large data sets and on the other hand medical data is of utmost privacy-sensitivity and also scattered across institutions. Therefore, a privacy-preserving technological infrastructure is required to enable large-scale data access for the purpose of high-quality AI model training.

This thesis presented FL as one promising approach in this respect. Yet, the extent to which FL can be applied across medical data types as well as combinations of different data types has not been assessed within this study and should be assessed going forward. For example, one study already showed that a FL approach for medical images showed the same classification performance as a centralized AI approach (Chatterjee, Gerdes, and Martinez 2020; Kaissis et al. 2021).

In another study, COVID-19 case data were analyzed in a FL setting to predict infectious cases and recovery rates using chest x-ray data (Abdul Salam, Taha, and Ramadan 2021). The FL model demonstrated a better prediction accuracy and loss while requiring higher performance time than the traditional machine learning model - this parameter has not been considered in the conducted publication of this thesis. Such a specific trade-off of requirements is therefore suggested for future studies.

In addition to that, there are alternative approaches that may be more appropriate than the proposed FL approach, in general or under specific circumstances. Some first examples from the area of precision medicine are based on the “Swarm Learning” methodology, a different decentralized machine-learning approach that unites edge computing and blockchain-based peer-to-peer networking. First results for COVID-19, tuberculosis, leukemia and lung pathologies showed positive results while maintaining privacy laws (Warnat-Herresthal et al. 2021).

Future research should compare different privacy-preserving approaches to shed more light on their suitability in different medical environments. In this context, different dimensions should be considered such as data types, costs, quality from a methodological perspective (e.g., accuracy), process complexity (e.g., in terms of feasibility for the medical staff involved at the participating medical institutions) and capacity (e.g., in terms of feasibility with “regular” servers at medical institutions).

Medical and economic impact measurement

Regarding impact measurement, the thesis showed that there are only few medical and economic impact assessments and that these are commonly subject to methodological flaws. This implies that to-date, strategic decisions for or against AI implementation by different institutions commonly lack evidence on the impact consequences of applying AI. Given this uncertainty, decision-makers could likely abstain from implementing AI to avoid financial downside risk, competitive disadvantage, investor discontent etc. However, considering the promising evidence as to the benefits of AI in the context of the global corona pandemic in the last years, impact assessments would likely often indicate that there is actually a high financial upside potential (Wang et al. 2021).

The overall limited evidence on the economic benefits of AI in healthcare may also more generally disincentivize actors from engaging in costly innovation activities in the first place. As such, the quantity and quality of impact assessments should be increased not only to support decision-makers with view to the concrete solution under analysis, but also to generally improve knowledge on the medical and economic benefits of AI. Public institutions, governments, academic researchers, medical institutions and also the industry could contribute to such a knowledge base by according data collection in reference cases.

In particular, appropriate quantitative approaches, e.g., in terms of uniform and well-established outcome measures, and qualitative approaches with standardized reporting processes should be applied. A higher amount of reliable evidence on the value-added of AI would likely also increase patients' trust in AI solutions in healthcare and, thus, market acceptance. Although medical impact is likely more relevant for patients, some positive economic impacts are also rewarding for them. For example, an AI chatbot for healthcare advice can be efficient from the doctor's perspective and simultaneously generates time savings (e.g., no waiting or travel time) and reduced

costs (e.g., transport costs) for patients when no medical intervention is needed or when it is simply unavailable at the point of care.

As some decision-makers such as private institutions could prefer to keep their impact assessment results private, academic researchers can play a particularly valuable role to improve the state of knowledge in this area. In addition, governments could support by setting incentives for publication of impact data of private institutions or by publishing the results, where available to them, in an anonymous, less granular or aggregated form.

Limitations of this thesis

While this thesis identified and elaborated on three key success factor categories to increase AI implementation in healthcare, there are of course additional success factors that are not addressed or could be further elaborated on.

First, further success factor categories could be researched. One option could be “responsible” AI as well as according ethical standards. In this regard, key aspects comprise strong ethical practices, information security, well-being of the society, workers’ skills, and organizations’ AI-culture (Fosso Wamba and Queiroz 2021). Responsible AI could contribute to the likelihood of successful AI implementation, both in terms of compliance, but also patient acceptance and support by governments and payors. A further parameter could be social or environmental dimensions (e.g., contribution to lower emissions through reduced traveling).

Second, this thesis elaborated on how AI implementation in healthcare can be increased overall, yet does not focus in particular on the perspective of patients or physicians, which could be further highlighted. In particular, a prior study showed that patients have multiple concerns regarding AI in healthcare, for example regarding the safety of AI, threats to patient choice, potential increases in healthcare costs, data-source bias, and data security and the authors highlighted, that patient acceptance of

AI need to be improved (Richardson et al. 2021). These patient concerns are also affecting physicians and their role as “gatekeepers” for healthcare access. There are significant confidence building measures required which should be elaborated further. Governments and national public authorities could for instance help to reduce such concerns and increase trust by establishing appropriate policy frameworks and by publicly promoting successful AI application cases. Furthermore, medical institutions likely need to take this into account as well and, next to high-quality impact assessments, they could also develop communication strategies for patients, e.g., with regard to AI benefits in their treatment regimen. In any case, given the need to achieve AI acceptance, future research should assess the patient and physician perspective in more detail.

Third, reimbursement frameworks need to be analyzed further. There is a constantly developing reimbursement landscape for DH and AI, for example with the DiGA reimbursement in Germany or the mHealthBelgium M1-M3 reimbursement system. These requirements also significantly shape the regulatory frameworks, for example the German law requires a medical product class I or IIA classification as obligatory requirement for approval as also elaborated in the publication “The impact of AI on the Healthcare Economy”. Still there could be further research on the overall development of DH reimbursement frameworks and also the interconnection between country specific systems, e.g., in order to conduct one AI clinical trial for internationally accepted evidence generation.

In addition to these three points relating to success factors, future research could elaborate on a) how the assessed success factors are interlinked and b) their context-specificity, namely which success factors are particularly relevant under which circumstances. Exemplary dimensions are preventive care vs. chronic diseases or low-resource vs. high-resource healthcare countries. This could provide further context-specific guidance and recommendations to medical institutions and thereby likely further contribute to closing the gap of real-world AI implementation. In addition to that, AI in healthcare is a very dynamic field and recent developments show already

some hints about future development in this sector: The scientific landscape is fastly developing and, for example, health services management, predictive medicine, patient data and diagnostics, and clinical decision-making are key research areas, while in the US, China, and the UK so far the highest number of academic studies have been published (Secinaro et al. 2021). Also the technological development is ongoing and, for example, first architectures that contain blockchain-based IoT platforms and use FL have been introduced, which shall enable faster scaling of solutions (Singh et al. 2022). Long et. al also give a broader outlook by stating that FL can enable a new chapter of “Open Innovation” since it can be the next general AI model training framework within the research community, but also with external partners (Long et al. 2022). Overall, additional success dimensions, the patient and physician perspective, future reimbursement models, the case specific interconnection of success factors categories as well as key technological trends could be areas of future research.

7. Conclusions

This thesis examines how real-world AI implementation in healthcare can be increased. It sheds light on the current status quo of AI implementation within the healthcare industry, showing that there are currently only few large-scale real-world AI use cases and it elaborates on key success factors and measures for increased implementation. While various barriers have hindered such a transformation to take place in the healthcare market yet, this thesis suggests that this development can be actively steered.

In particular, four current developments are likely to trigger the transformation of the healthcare market: First, there is an urgent demand for AI to support in coping with the worldwide healthcare challenges of rising costs, increased demand for services and labor shortages in this sector. Second, improvements with regard to technological capabilities in recent years have simplified access to and scaling of AI applications. Third, for the first time, noteworthy awareness and acceptance levels can be observed for AI application in healthcare, both among key decision-makers such as governmental institutions as well as among patients. Finally, the global corona pandemic has demonstrated the significant need for DH and AI structures and accelerated the development of reference cases.

Yet, to facilitate the transition, concrete steps need to be undertaken by governments, researchers, medical practitioners and industry key decision-makers. In this regard, this thesis lays out three key success factors for AI implementation: 1) A facilitating policy setting 2) A privacy preserving technological infrastructure and 3) A high-quality medical and economic impact assessment.

The first success factor category, namely policy setting, can be influenced through continuous support by governments in the form of facilitating policy frameworks, such as the German Digital Healthcare Act, providing clear structures for the approval of DH services (e.g., regarding safety, medical product class, etc.), medical and

structural endpoints for evidence generation processes, and the reimbursement through public health systems in order to support the growth of DH and AI.

The latter two success factor categories, privacy preserving technological implementation as well as medical and economic impact measurement, can be actively steered by researchers and practitioners to increase the likelihood of success of AI implementation.

As to technological implementation, data access across medical institutions is a key factor for the real-world use of AI and given its privacy preservation characteristic, FL is a promising potential pathway for AI application. This could be demonstrated in a concrete AI use case in the context of CACSs prediction. Accordingly, the thesis suggests that privacy preserving technology should be applied for AI implementation and a suitability differentiation between medical contexts would contribute to actual real-world use.

With regard to impact measurement, both medical and economic impact represent key factors in strategic AI decision-making. The thesis shows that impact has so far been assessed rarely and with insufficient quality. Since high-quality evidence measurement is crucial for increased real-world scalability this should be collected in standardized assessments.

In the following figure a short summary is provided:

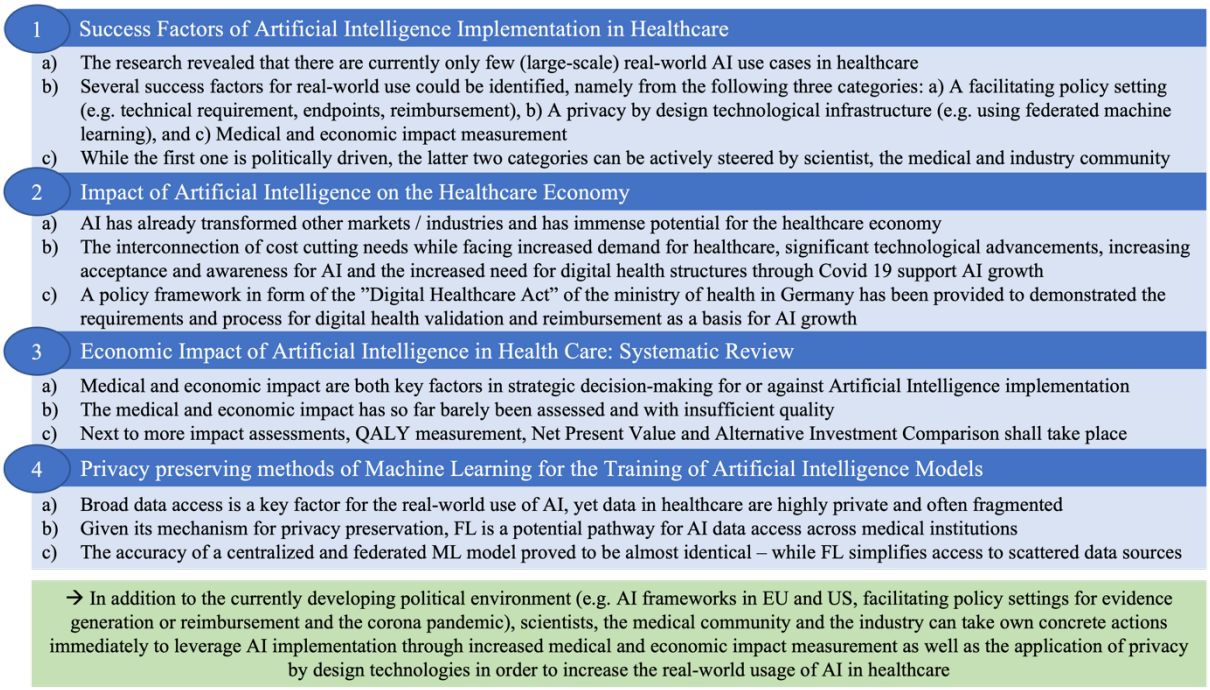


Figure 5: Overview about overarching conclusions in the publications

With view to the above-described conclusions, this thesis provides suggestions for governments, researchers, medical practitioners and the healthcare industry. The specific recommendations for increased real-world AI implementation in healthcare were derived directly from the research results and I recommend the following actions:

Governments/regulators:

- Governments shall set concrete and transparent frameworks for the implementation of DH and AI, including regulatory requirements with view to e.g., medical products classes, evidence generation requirements (split into medical and economic data) as well as guidelines for the reimbursement in the same form as for classical healthcare services like drugs/ medtech devices
- Clear risk assessments and evaluation should be implemented to enhance the trust levels of physicians and patients
- The existing AI regulatory frameworks in the US and Europe will likely serve as a benchmark or even a reference case for other regions and, thus,

require extensive research in their currently ongoing completion process; the real-world insights from medical practitioners, industry and academics about AI should be integrated and systematically evaluated (one example is the collaboration with notified bodies)

Academic research:

- Through increased and open access translational research on AI driven diagnosis and treatment processes, researchers can significantly contribute to closing the current implementation gap
- Medical and economic impact measurement of DH and AI needs to be extended both from a quantitative perspective and a qualitative perspective e.g., via the inclusion of reporting standards inside publications or even requirements of publishers for assessments
- Existing datasets should be used to assess the accuracy of AI technologies and their respective value-added in terms of a medical decision-making support tool; a reference case is the decentralized AI-based CACS prediction model that contains > 1.500 patient data
- FL may also represent a valuable tool in other medical contexts and future research should analyze the benefits of FL in different applications

Medical practitioners:

- DH and AI offer significant improvement potential and should be an integral part of the innovation strategy within the medical unit and contain concrete implementation goals
- Assessments of the medical and economic impact of DH and AI applications should serve as evidence basis for budget decisions, e.g. with the hospital management
- Medical impact should be measured and standardized, for example, by applying QALYs based on EQ-5D as well as the CHEERS and PRISMA criteria

- Economic impact should be measured through Net Present Value calculations and respective assessments should include Cost Alternative Scenarios
- The medical and economic impact is also a crucial requirement for reimbursement, like in the DVG example from Germany, and provides room for external collaboration with the industry. Therefore these impact assessments should be integrated in a way to facilitate engagement with other healthcare stakeholders such as insurance providers and government agencies.

Industry:

- Although there is significant market potential, only very few real-world cases of AI in healthcare exist yet - the significant business opportunity is also visible in a comparison with other industries that have used AI applications, such as e-commerce, and measurable growth goals for AI solutions shall be integrated in the company's growth strategy
- There are concrete market needs such as the need for cost savings in healthcare, shortage of medical staff and increased need for medical care due to population growth that AI can at least partially solve - these are opportunities for large-scale industrial application
- FL can be even more important in industrial applications due to stricter regulations (esp. with regard to data privacy) as compared to a merely academic or hospital research context - thus the application should be fostered early and with a comprehensive roll out plan
- Industry providers should structure AI applications in the most user-friendly way to allow for easy and efficient integration into the existing day-to-day business e.g., via smartphone apps, integration into EHR systems etc. while complying with regulatory frameworks

Altogether, it can be stated that many of the abovementioned recommendations could be directly implemented and there is of course also an interconnection between the stakeholders and their interests: For example the data about medical and economic impact assessments and the real-world implementation processes inside the medical facilities are urgently needed by the regulators to provide according frameworks that in turn support real-world implementation. The necessity to take account of this interconnection can be seen in yet unsuccessful digital infrastructure projects that have been imposed by governments or the slow development of AI regulatory frameworks as well as the generally low level of large-scale AI use cases induced by a lack of government policy.

Considering the significant potential to positively impact people's lives and the economies worldwide, a fast acceleration of AI in healthcare should be enabled as quickly as possible. A close collaboration between governmental institutions, academic research, medical practitioners and industry stakeholders could actively support this development in order to leverage the full potential of AI in healthcare and to ultimately obtain significant medical and economic benefits globally.

8. References

Abdul Salam, Mustafa, Sanaa Taha, and Mohamed Ramadan. 2021. “COVID-19 Detection Using Federated Machine Learning”. *PloS One* 16 (6): e0252573.

Accenture. 2017. “Artificial Intelligence: Healthcare’s New Nervous System”. Accessed December 31, 2021. https://www.accenture.com/_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf#zoom=50.

Accenture. 2018. “Consumer Survey on Digital Health”. Accessed December 31, 2021. https://www.accenture.com/t20180306t103559z__w__/us-en/_acnmedia/pdf-71/accenture-health-2018-consumer-survey-digital-health.pdf.

Accenture. 2018. “The Intelligent Payer: A Survival Guide”. Accessed December 31, 2021. https://www.accenture.com/_acnmedia/pdf-82/accenture-intelligent-payer-survivorguide.pdf.

Aguis, Christina. 2019. “Evolution of AI: Past, Present, Future”. *Data Driven Investor*. Accessed December 31, 2021. https://medium.com/datadriveninvestor/evolution-of-ai-past-present-future-6f995d5f964a.

Albahri, O. S., A. A. Zaidan, A. S. Albahri, B. B. Zaidan, Karrar Hameed Abdulkareem, Z. T. Al-Qaysi, A. H. Alamoodi, et al. 2020. “Systematic Review of Artificial Intelligence Techniques in the Detection and Classification of COVID-19 Medical Images in Terms of Evaluation and Benchmarking: Taxonomy Analysis, Challenges, Future Solutions and Methodological Aspects”. *Journal of Infection and Public Health* 13 (10): 1381–96.

Alhashmi, Shaikha F. S., Said A. Salloum, and Sherief Abdallah. 2020. “Critical Success Factors for Implementing Artificial Intelligence (AI) Projects in Dubai Government United Arab Emirates (UAE) Health Sector: Applying the Extended Technology Acceptance Model (TAM)”. *Advances in Intelligent Systems and Computing*, 393–405. Cham: Springer International Publishing.

Ameen, Nisreen, Ali Tarhini, Alexander Reppel, and Amitabh Anand. 2021. “Customer Experiences in the Age of Artificial Intelligence”. *Computers in Human Behavior* 114 (January): 106548.

Bajwa, Junaid, Usman Munir, Aditya Nori, and Bryan Williams. 2021. “Artificial Intelligence in Healthcare: Transforming the Practice of Medicine”. *Future Healthcare Journal* 8 (2): e188–94.

Budd, Jobie, Benjamin S. Miller, Erin M. Manning, Vasileios Lampos, Mengdie Zhuang, Michael Edelstein, Geraint Rees, et al. 2020. “Digital Technologies in the Public-Health Response to COVID-19”. *Nature Medicine* 26 (8): 1183–92.

Chatterjee, Ayan, Martin W. Gerdes, and Santiago G. Martinez. 2020. “Identification of Risk Factors Associated with Obesity and Overweight-A Machine Learning Overview”. *Sensors* 20 (9): 2734.

CMS. 2021. “CMS Global Lifescience & Healthcare Festival” (Presentation by Justus Wolff, Dr. Roland Wiring and Jean-Baptiste Tienot on 29 November .2021). Presentation available on request.

EU Commission. 2020. “Whitepaper on AI - A European Approach to Excellence and Trust”. Accessed December 31, 2021. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

EU Commission. 2021. “Proposal for a regulation of the European Parliament and the Council - Laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts”. Accessed December 31, 2021. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

EU Commission. 2021. “Artificial Intelligence can help us battle the Corona Virus” Accessed 01.05.2022. https://ec.europa.eu/info/strategy/recovery-plan-europe/recovery-coronavirus-success-stories/health/artificial-intelligence-can-help-us-combat-coronavirus_en

European Union. 2021. “Study on eHealth, Interoperability of Health Data and Artificial Intelligence for Health and Care in the European Union”. Accessed December 27, 2021. https://euagenda.eu/publications/study-on-ehealth-interoperability-of-health-data-and-

artificial-intelligence-for-health-and-care-in-the-european-union.
Fahle. 2020. “Systematic Review on Machine Learning (ML) Methods for Manufacturing Processes – Identifying Artificial Intelligence (AI) Methods for Field Application”. *Procedia CIRP* 93 (January): 413–18.
Fosso Wamba, Samuel, and Maciel M. Queiroz. 2021. “Responsible Artificial Intelligence as a Secret Ingredient for Digital Health: Bibliometric Analysis, Insights, and Research Directions”. *Information Systems Frontiers* (May): 1–16.
Garbuio, Massimo and Nidthida Lin. 2019. “Artificial Intelligence as a Growth Engine for Health Care Startups: Emerging Business Models”. *California Management Review* 61 (2): 59–83.
Grand View Research. 2019. “Artificial Intelligence In Healthcare Market Size, Share & Trends Analysis Report By Component (Hardware, Software, Services), By Application, By Region, Competitive Insights, And Segment Forecasts, 2019 - 2025”. Accessed December 27, 2021. https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-aihealthcaremarket?utm_source=prnewswire&utm_medium=referral&utm_campaign=hc_16-dec-19&utm_term=ai-healthcare-market&utm_content=rd1.
He, Jianxing, Sally L. Baxter, Jie Xu, Jiming Xu, Xingtao Zhou, and Kang Zhang. 2019. “The Practical Implementation of Artificial Intelligence Technologies in Medicine”. *Nature Medicine* 25 (1): 30–36.
FeatureCloud. 2020. ”FeatureCloud Publishable Summary First Reporting Period January 2019 – 30th April 2020”. Accessed December 31, 2021. https://featurecloud.eu/wpcontent/uploads/2020/07/FeatureCloud_Publishable-summary_2020.pdf.
Gama, Fábio, Daniel Tyskbo, Jens Nygren, James Barlow, Julie Reed, and Petra Svedberg. 2022. “Implementation Frameworks for Artificial Intelligence Translation Into Health Care Practice: Scoping Review.” Journal of Medical Internet Research 24 (1): e32215.
HIH. 2021. Source not available online.
Horizon. 2020. ”To the NEXT level of risk prediction in patients with Long QT Syndrome”. Accessed December 31, 2021. https://www.era-learn.eu/network-information/networks/sc1-bhc-04-2018/1st-ejp-rd-joint-transnational-call-for-rare-diseases-research-project-jtc-2019/to-the-next-level-of-risk-prediction-in-patients-with-long-qt-syndrome.
Ilan, Yaron. 2020. “Second-Generation Digital Health Platforms: Placing the Patient at the Center and Focusing on Clinical Outcomes”. *Frontiers in Digital Health* 2 (December): 569178.
Kaissis, Georgios, Alexander Ziller, Jonathan Passerat-Palmbach, Théo Ryffel, Dmitrii Usynin, Andrew Trask, Ionésio Lima Jr, et al. 2021. “End-to-End Privacy Preserving Deep Learning on Multi-Institutional Medical Imaging”. *Nature Machine Intelligence* 3 (6): 473–84.
Kelly, Christopher J., Alan Karthikesalingam, Mustafa Suleyman, Greg Corrado, and Dominic King. 2019. “Key Challenges for Delivering Clinical Impact with Artificial Intelligence”. *BMC Medicine* 17 (1): 195.
Kumar, Sudhesh. 2019. “Capitalising on Japan’s Digital Healthcare Economy during This Era of Aging Societies”. Accessed December 27, 2021. https://www.himss.eu/sites/himsseu/files/education/whitepapers/191007_HIMSS_Japan_White_Paper_WEB.pdf.
Lin, Steven. 2022. “A Clinician’s Guide to Artificial Intelligence (AI): Why and How Primary Care Should Lead the Health Care AI Revolution.” Journal of the American Board of Family Medicine: JABFM 35 (1): 175–84.
Long, Guodong, Tao Shen, Yue Tan, Leah Gerrard, Allison Clarke, and Jing Jiang. 2022. “Federated Learning for Privacy-Preserving Open Innovation Future on Digital Health”. In Humanity Driven AI, 113–33. Cham: Springer International Publishing.
MarketsAndMarkets. 2020. “Artificial Intelligence in Healthcare Market with Covid-19 Impact, Analysis by Offering (Hardware, Software, Services), Technology (Machine Learning,

NLP, Context-Aware Computing, Computer Vision), End-Use Application, End User and Region - Global Forecast to 2026." Accessed December 31, 2021. https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-healthcare-market-54679303.html?gclid=.

McKinsey. 2020. "Transforming healthcare with AI: The impact on the workforce and organisations". Accessed December 31, 2021. https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/transforming-healthcare-with-ai.

MedTech Innovation. 2021. "The future of digital health technologies in the changing Covid 19 landscape". Accessed December 31, 2021. https://www.med-technews.com/medtech-insights/digital-in-healthcare-insights/the-future-of-digital-health-technologies-in-the-changing-co/.

Mordor Intelligence. 2020. "Artificial Intelligence in Medicine Market - Growth, Trends, and Forecast (2020-2025)". Accessed December 27, 2021.https://www.mordorintelligence.com/industry-reports/artificial-intelligence-in-medicine-market.

NIST. 2021. "NIST Provisions in the National Defense Authorization Act of 2021". Accessed December 31, 2021. https://www.nist.gov/system/files/documents/2021/02/01/6a.%20Schufreider%20NIST%20Provisions%20in%20the%20National%20Defense%20Authorization%20Act%20of%202021.pdf.

Orric. 2021. U.S. "Artificial Intelligence Regulation takes shape". Accessed December 31, 2021. https://www.orrick.com/en/Insights/2021/11/US-Artificial-Intelligence-Regulation-Takes-Shape.

Panch, Trishan, Heather Mattie, and Leo Anthony Celi. 2019. "The 'Inconvenient Truth' about AI in Healthcare". *NPJ Digital Medicine* 2 (August): 77.

Park, Chan Woo, Sung Wook Seo, Noeul Kang, Beomseok Ko, Byung Wook Choi, Chang Min Park, Dong Kyung Chang, et al. 2020. "Artificial Intelligence in Health Care: Current Applications and Issues". *Journal of Korean Medical Science* 35 (42): e379.

Petersson, Lena, Ingrid Larsson, Jens M. Nygren, Per Nilsen, Margit Neher, Julie E. Reed, Daniel Tyskbo, and Petra Svedberg. 2022. "Challenges to Implementing Artificial Intelligence in Healthcare: A Qualitative Interview Study with Healthcare Leaders in Sweden." BMC Health Services Research 22 (1): 850.

Pifer, Rebecca. 2019. "Amazon, Cerner Team up on AI, Machine Learning." Accessed July 31, 2019. https://www.healthcaredive.com/news/amazon-cerner-team-up-on-ai-machine-learning/559847/.

Racine, Eric, Wren Boehlen, and Matthew Sample. 2019. "Healthcare Uses of Artificial Intelligence: Challenges and Opportunities for Growth". *Healthcare Management Forum / Canadian College of Health Service Executives = Forum Gestion Des Soins de Sante / College Canadien Des Directeurs de Services de Sante* 32 (5): 272–75.

Radboudumc. 2019. "Deep Learning for Prostate MRI Intervention - Diagnostic Image Analysis Group". Accessed December 27, 2021. http://diagnijmegen.nl/index.php/Deep_learning_for_Prostate_MRI_Intervention.

Reuters. 2020. "Google CEO Eyes Major Opportunity in Health Care, Says It Will Protect Privacy". *CNBC*. Accessed January 22, 2020. https://www.cnbc.com/2020/01/22/google-ceo-eyes-major-opportunity-in-health-care-says-it-will-protect-privacy.html.

Richardson, Jordan P., Cambray Smith, Susan Curtis, Sara Watson, Xuan Zhu, Barbara Barry, and Richard R. Sharp. 2021. "Patient Apprehensions about the Use of Artificial Intelligence in Healthcare". *NPJ Digital Medicine* 4 (1): 140.

Sanyal, Chiranjeev, Paul Stolee, Don Juzwishin, and Don Husereau. 2018. "Economic Evaluations of eHealth Technologies: A Systematic Review". *PloS One* 13 (6): e0198112.

Secinaro, Silvana, Davide Calandra, Aurelio Secinaro, Vivek Muthurangu, and Paolo Biancone. 2021. "The Role of Artificial Intelligence in Healthcare: A Structured Literature Review". *BMC Medical Informatics and Decision Making* 21 (1): 125.

Shaw, James, Frank Rudzicz, Trevor Jamieson, and Avi Goldfarb. 2019. "Artificial Intelligence

and the Implementation Challenge". *Journal of Medical Internet Research* 21 (7): e13659.
Singh, Saurabh, Shailendra Rathore, Osama Alfarraj, Amr Tolba, and Byungun Yoon. 2022. "A Framework for Privacy-Preservation of IoT Healthcare Data Using Federated Learning and Blockchain Technology". *Future Generations Computer Systems: FGCS* 129 (April): 380–88.
Smith, Margaret, Amelia Sattler, Grace Hong, and Steven Lin. 2021. "From Code to Bedside: Implementing Artificial Intelligence Using Quality Improvement Methods". *Journal of General Internal Medicine* 36 (4): 1061–66.
The Health Foundation, Nuffield Trust, and The King's Fund. 2018. "The health care workforce in England: make or break?". Accessed December 31, 2021. https://www.kingsfund.org.uk/sites/default/files/2018-11/The%20health%20care%20workforce%20in%20England.pdf.
Topol, Eric J. 2019. "High-Performance Medicine: The Convergence of Human and Artificial Intelligence". *Nature Medicine* 25 (1): 44–56.
Triantafyllidis, Andreas K., and Athanasios Tsanas. 2019. "Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature". *Journal of Medical Internet Research* 21 (4): e12286.
U.S. Food and Drug Administration. 2019. "Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Software as a Medical Service (SaMD)". Accessed December 31, 2021. https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf.
US Department of Commerce. 2021. "Department of Commerce Establishes National Artificial Intelligence Advisory Committee". Accessed December 31, 2021. https://www.commerce.gov/news/press-releases/2021/09/department-commerce-establishes-national-artificial-intelligence.
Vaishya, Raju, Mohd Javaid, Ibrahim Haleem Khan, and Abid Haleem. 2020. "Artificial Intelligence (AI) Applications for COVID-19 Pandemic". *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 14 (4): 337–39.
Wang, Lian, Yonggang Zhang, Dongguang Wang, Xiang Tong, Tao Liu, Shijie Zhang, Jizhen Huang, et al. 2021. "Artificial Intelligence for COVID-19: A Systematic Review". *Frontiers of Medicine* 8 (September): 704256.
Warnat-Herresthal, Stefanie, Hartmut Schultze, Krishnaprasad Lingadahalli Shastry, Sathyanarayanan Manamohan, Saikat Mukherjee, Vishesh Garg, Ravi Sarveswara, et al. 2021. "Swarm Learning for Decentralized and Confidential Clinical Machine Learning". *Nature* 594 (7862): 265–70.
Wolff, Justus, Josch Pauling, Andreas Keck, and Jan Baumbach. 2020. "The Economic Impact of Artificial Intelligence in Health Care: Systematic Review". *Journal of Medical Internet Research* 22 (2): e16866.
World Health Organization. 2008. "Right to Health and Human Rights". Accessed December 31, 2021. https://www.ohchr.org/documents/publications/factsheet31.pdf.
World Health Organization. 2016. "Working for health and growth: Investing in the health workforce". Accessed December 31, 2021. http://apps.who.int/iris/bitstream/10665/250047/1/9789241511308-eng.pdf.
World Health Organization. 2019. "Global Spending on Health: A World in Transition". Accessed December 27, 2021. https://www.who.int/health_financing/documents/health-expenditure-report-2019.pdf?ua=1.