



UNIVERSITÄT ZU LÜBECK

Aus dem Institut für Gesundheitswissenschaften
Fachbereich Physiotherapie
Universität zu Lübeck
Direktorin Prof. Dr. Kerstin Lütke

**VALIDIERUNG UND EVALUATION EINER
KÜNSTLICHEN INTELLIGENZ FÜR ÜBUNGSEMPFEHLUNGEN
BEI PATIENT*INNEN MIT MUSKULOSKELETTALEN ERKRANKUNGEN**

Inauguraldissertation
zur Erlangung der Doktorwürde
der Universität zu Lübeck
– **Aus der Sektion Medizin** –

vorgelegt von
Annika Griefahn
aus Hanau

Lübeck 2025

1. Berichterstatter*in: Prof. Dr. Kerstin Lüdtko

Ko-Betreuer: Prof. Dr. med. Christoff Zalpour

2. Berichterstatter*in: Prof. Dr. med. Alexander Münchau

Tag der mündlichen Prüfung: 17.03.2026

Zum Druck genehmigt. Lübeck, den 24.03.2026

- Promotionskommission der Sektion Medizin -

INHALTSVERZEICHNIS

Abbildungsverzeichnis	IV
Tabellenverzeichnis	IV
Abkürzungsverzeichnis	IV
1. Einführung	5
1.1. Muskuloskelettale Erkrankungen	5
1.2. Digitale Lösungen	6
1.3. Künstliche Intelligenz im Gesundheitswesen.....	7
1.4. medicalmotion	8
1.4.1.medicalmotion App	9
1.4.2.Die künstliche Intelligenz medico	10
1.5. Forschungsziele	12
2. Darstellung der Publikationen	13
2.1. Publikation I	13
2.2. Publikation II	16
2.3. Publikation III.....	19
3. Gesamtkonklusion	21
4. Literaturverzeichnis	25
5. Anhang (Publikationen).....	30
6. Danksagung.....	36

ABBILDUNGSVERZEICHNIS

Abbildung 1: Screenshots aus der App medicalmotion.....	10
Abbildung 2: Zyklus der Übungsauswahl und -optimierung	11
Abbildung 3: Patientenbeispiel für die Bewertung von Kontraindikationen	14
Abbildung 4: Studienablauf für die Risikobewertung	17

TABELLENVERZEICHNIS

Tabelle 1: Einsatzmöglichkeiten von künstlicher Intelligenz in der Medizin: Aufgaben und zugehörige Datentypen	8
Tabelle 2: Überblick über die Anzahl der Übungen, die von Physiotherapeut*innen als Risikoübungen eingestuft werden	18

ABKÜRZUNGSVERZEICHNIS

ANOVA	einfaktorielle Varianzanalyse
App	mobile Anwendung
CDSS	klinisches Entscheidungsunterstützungssystem
DiGA	digitale Gesundheitsanwendung
KI	künstliche Intelligenz
MSK	muskuloskelettale Erkrankungen
NRS	numerische Ratingskala
TRS	Therapieempfehlungssystem

1. EINFÜHRUNG

"The digital health revolution is the most important health care innovation of the 21st century" – The Lancet Digital Health Commission

1.1. Muskuloskelettale Erkrankungen

Muskuloskelettale Erkrankungen (MSK) sind weltweit die häufigste Ursache für chronische Schmerzen und funktionelle Einschränkungen – sie führen zu erheblichen Beeinträchtigungen der Lebensqualität (Gill et al., 2023). Der Begriff MSK bezeichnet ein breites Spektrum von Erkrankungen des Bewegungsapparates, das Strukturen wie Skelettmuskulatur, Sehnen, Bänder, Knorpel und Nerven einschließt. MSK umfassen unter anderem degenerative Gelenkerkrankungen wie Arthrose, entzündlich-rheumatische Krankheitsbilder sowie systemische Skeletterkrankungen wie Osteoporose (Gill et al., 2023).

In Deutschland zählen MSK zu den häufigsten gesundheitlichen Problemen mit erheblichen sozioökonomischen Konsequenzen. Diese reichen von kostenintensiven Behandlungen über Arbeitsunfähigkeit bis hin zur Frühverrentung (Heidemann et al., 2021). Im Jahr 2020 beliefen sich die Krankheitskosten für MSK auf 41,7 Milliarden Euro (Statistisches Bundesamt, 2022). Darüber hinaus verursachten MSK in Deutschland im Jahr 2016 Produktionsausfälle in Höhe von 17,2 Milliarden Euro sowie Verluste der Bruttowertschöpfung in Höhe von 30,4 Milliarden Euro (Brenscheidt et al., 2018).

Die Ätiologie für MSK ist meist multifaktoriell und umfasst körperliche, psychische und arbeitsbezogene Einflussfaktoren (Cieza et al., 2021). Bewegungsmangel stellt einen weiteren signifikanten Risikofaktor für die Entwicklung von MSK dar (Holth et al., 2008). Aktuelle Studien zeigen, dass regelmäßige körperliche Aktivität sowohl bei der Prävention als auch bei der Therapie, insbesondere bei Adipositas, chronischen Schmerzen und psychischen Erkrankungen, als essenzieller Bestandteil gilt (Jakicic et al., 2018; Martinsen, 2008; Sluka et al., 2013). Die Wirksamkeit regelmäßiger körperlicher Aktivität basiert unter anderem auf antiinflammatorischen Effekten, verbesserter neuromuskulärer Kontrolle und der Förderung des psychischen Wohlbefindens (Vincent et al., 2015). In einem Cochrane-Review von (Jordan et al., 2010) konnten signifikant positive Effekte von Bewegungstherapien bei Kniearthrose und Rückenschmerzen in 18 von 42 analysierten Studien belegt werden. Um MSK jedoch langfristig zu verbessern, ist eine kontinuierliche Durchführung der Übungen erforderlich – ein Aspekt, der eng mit der sogenannten Adhärenz

verknüpft ist. Unter Adhärenz ist die aktive und eigenverantwortliche Mitwirkung von Patient*innen an gemeinsam vereinbarten therapeutischen Maßnahmen zu verstehen. Sie gilt als zentraler Erfolgsfaktor in der Behandlung von MSK (Jack et al., 2010). Zur Förderung der Übungsadhärenz werden unterschiedliche Strategien, wie bspw. angeleitete Einheiten, individualisierte Übungspläne, Follow-ups, multimediale Unterstützungen, Selbstmanagement-Programme und verhaltensorientierte Interventionen, diskutiert (Jordan et al., 2010). Dennoch bleibt die Umsetzung im Alltag herausfordernd: Wie Heisig et al. (2025) betonen, hängt die Wirksamkeit bewegungstherapeutischer Interventionen maßgeblich von ihrer konsequenten Durchführung ab – in der Praxis wird dies häufig durch fehlende Motivation oder externe Barrieren eingeschränkt.

Darüber hinaus bestehen erhebliche strukturelle Versorgungslücken. Lange Wartezeiten auf physiotherapeutische Termine, regionale Unterschiede in der Verfügbarkeit und begrenzte Behandlungskapazitäten erschweren den Zugang zu adäquater Therapie (Ärztezeitung, 2023; Bundesverband selbstständiger Physiotherapeut*innen – IFK e. V., 2023; Physio Deutschland, 2018). In Kombination mit der häufig eingeschränkten Adhärenz mindert dies die Effektivität etablierter Versorgungskonzepte. Diese Herausforderungen verdeutlichen den Bedarf an innovativen Ansätzen, die sowohl die Individualisierung therapeutischer Maßnahmen als auch deren nachhaltige Umsetzung im Alltag stärker in den Vordergrund rücken.

1.2. Digitale Lösungen

Das 2019 in Kraft getretene Digitale-Versorgung-Gesetz schuf erstmals einen rechtlichen Rahmen für die Digitalisierung des deutschen Gesundheitswesens (Bundesministerium für Gesundheit, 2019). Seitdem dürfen Ärzt*innen digitale Gesundheitsanwendungen (DiGA) verordnen, sofern diese vom Bundesinstitut für Arzneimittel und Medizinprodukte zugelassen sind (Albrecht & Jungmann, 2016). Mit der zunehmenden Verbreitung von DiGAs erweitert sich das Spektrum an therapeutischen Optionen für Patient*innen und Ärzt*innen kontinuierlich. Gleichzeitig erschwert die hohe Dynamik des Marktes von mobilen Anwendungen (Apps) sowie das Fehlen standardisierter Qualitätssicherungsmechanismen eine fundierte Bewertung einzelner Anwendungen. Vor diesem Hintergrund wird in der aktuellen Literatur die Notwendigkeit betont, evidenzbasierte Bewertungsansätze zu etablieren, die sowohl die Sicherheit als auch die Wirksamkeit von digitalen Gesundheitsangeboten systematisch erfassen und in den klinischen Alltag übertragbar machen (Merolli et al., 2024).

Aktuell sind mehr als 325.000 Apps mit Gesundheitsbezug, deren Spektrum von digitalen

Übungsprogrammen und Schmerzmanagementlösungen bis hin zu Monitoring-Tools reicht, in kommerziellen App-Stores verfügbar (Grundy, 2022). Dabei zeigen aktuelle Studien, dass die Anwendung dieser Apps bei der Behandlung chronischer MSK zu positiven Ergebnissen geführt hat. Anzuführen sind hier signifikante Verbesserungen in Bezug auf die Schmerzintensität, das psychische Wohlbefinden sowie die gesundheitsbezogene Lebensqualität (Nowell et al., 2021; Wang et al., 2022; Weise et al., 2022). Weitere Studien unterstreichen, dass individualisierte Selbstmanagementstrategien die Adhärenz und die Wirksamkeit verbessern können (Hussain et al., 2022). Zudem führten kombinierte Interventionen mit DiGAs zu signifikanten gesundheitsökonomischen Einsparungen (Walter et al., 2022). Gleichzeitig bestehen aber auch Limitationen bei der Anwendung digitaler Lösungen mit Gesundheitsbezug, die es zu beachten gilt. So beruhen viele Anwendungen auf standardisierten Programmen und berücksichtigen individuelle Bedürfnisse nur unzureichend (Schlieter et al., 2024). Weiterhin ist die Orientierung an den Bedürfnissen der Nutzer*innen und die nachhaltige Förderung der Adhärenz nur unzureichend umgesetzt. Letztlich bleibt zu erwähnen, dass der Einsatz von künstlicher Intelligenz (KI) in den meisten Apps bislang begrenzt ist, wodurch eine adaptive Reaktion auf individuelles Feedback erschwert wird.

1.3. Künstliche Intelligenz im Gesundheitswesen

Der Einsatz von KI zur Unterstützung medizinischer Entscheidungen ist nicht neu. Bereits im Jahr 1990 thematisierte Mazoué die zentralen Herausforderungen der computergestützten Diagnostik (Mazoué, 1990). Aufgrund mangelnder Rechenleistung und begrenzter Datenverfügbarkeit war eine praxistaugliche Umsetzung damals jedoch nicht realisierbar. Erst durch technologische Fortschritte der letzten zwei Jahrzehnte wurde der breite Einsatz von KI im Gesundheitswesen, insbesondere durch Big Data, leistungsfähige Algorithmen und Methoden wie Deep Learning oder Natural Language Processing, ermöglicht (Jiang et al., 2017). Heute findet KI unter anderem in der Neurologie, Kardiologie, Onkologie und Ophthalmologie Anwendung (Jiang et al., 2017). Die eingesetzten Verfahren reichen von Support Vector Machines bis hin zu neuronalen Netzen (Davenport & Kalakota, 2019). Die Anwendungsfelder der KI umfassen unter anderem die Bildanalyse, Diagnostik, Prozesssteuerung und Therapieplanung (Esteva et al., 2017; Jansson et al., 2022; Schlieter et al., 2024). Tabelle 1 illustriert exemplarische Aufgabenbereiche von KI in Abhängigkeit der zugrunde liegenden Datentypen.

Tabelle 1: Einsatzmöglichkeiten von künstlicher Intelligenz in der Medizin: Aufgaben und zugehörige Datentypen

Aufgabenstellung	Datengrundlage
Detektion einer Retinopathie	Bilder des Augenhintergrunds
Zählen und Erkennen bestimmter Zelltypen	Histologische Schnittbilder
Diagnose von Infarkten, Alzheimer, Krebs	Radiologische Bildgebung
Erkennung von Depression	Sprache, Bewegungsmuster
Auswahl und Dosierung von Medikamenten	Diagnosen, Gen-Daten
Erkennen von Epidemien	Internet-Suchverhalten
Prognose von Krankheitsverläufen	Laborwerte, Umweltfaktoren
Prognose des Todeszeitpunkts von Intensivpatienten	Vitalparameter, elektronische Patientenakten

(Quelle: Eigene Darstellung, 2025)

Eine besondere Form von KI im Gesundheitswesen sind regelbasierte Expertensysteme. Diese Systeme verknüpfen formalisierte Entscheidungsregeln mit einer Wissensbasis aus empirischen Daten, Leitlinien und Fachwissen. Sie bestehen aus zwei zentralen Komponenten: der Wissensbasis (z. B. empirische Daten, medizinische Leitlinien) und einem Regelsatz (Wenn-Dann-Beziehungen), der die Reaktion des Systems auf Eingaben definiert. Die Logik basiert auf deduktiven Ableitungen zur Identifikation relevanter Zusammenhänge zwischen Fakten und Annahmen (Grosan & Abraham, 2011).

1.4. medicalmotion

Die medicalmotion GmbH¹ ist ein 2017 in München gegründetes Digital-Health-Unternehmen, das eine innovative Lösung in Form einer App zur digitalen Schmerztherapie entwickelt hat. Im Fokus steht die multimodale Behandlung und Prävention von MSK unter Einsatz von KI. Ziel ist es, Patient*innen mit akuten oder chronischen Schmerzen durch ein ganzheitliches und individualisiertes Therapieangebot langfristig zu unterstützen (Duran, 2021; medicalmotion GmbH, 2024). Die medicalmotion App ist als Medizinprodukt der Klasse I CE-zertifiziert (medicalmotion GmbH, 2024).

Zentrale technologische Komponente ist die entwickelte KI *medico*, welche auf Basis gesundheitsbezogener Eingaben von Nutzer*innen individualisierte Übungsempfehlungen generiert. Während die App medicalmotion als Oberfläche die Inhalte bereitstellt, übernimmt *medico* im Hintergrund die datenbasierte Steuerung des therapeutischen Verlaufs.

¹ <https://www.medicalmotion.com/de>

Im Zusammenspiel ermöglichen App und KI eine evidenzbasierte und dynamisch angepasste Therapie, die kontinuierlich an die Bedürfnisse der Nutzer*innen angepasst wird. Die folgenden Abschnitte beschreiben zunächst die Struktur und Funktionalität der App (Kapitel 1.4.1) und anschließend das zugrunde liegende Expertensystem *medico* (Kapitel 1.4.2).

1.4.1. medicalmotion App

Die App *medicalmotion* stellt das zentrale Interaktionsmedium dar, bei dem die Nutzer*innen täglich personalisierte Übungseinheiten erhalten. Die Auswahl der Übungen berücksichtigt Angaben zur Krankengeschichte, zu Schmerzcharakteristika, dem Lebensstil und dem aktuellen subjektiven Wohlbefinden. Dabei fließen Rückmeldungen aus vorangegangenen Übungseinheiten in die dynamische Anpassung der Übungsauswahl ein, sodass auch akute Veränderungen im Schmerzgeschehen berücksichtigt werden können. Die Nutzer*innen können pro Tag zwischen drei und fünf Übungseinheiten, die als audiogestützte Echtzeitvideos bereitgestellt werden, wählen. Das Expertensystem identifiziert zudem potenzielle Schmerzursachen und passt die Übungsempfehlungen entsprechend an. Die Übungen werden drei Kategorien zugeordnet: Entspannungsübungen zur Lockerung muskulärer Strukturen, Beweglichkeitsübungen zur Verbesserung endgradiger Gelenkbewegungen und Kräftigungsübungen zur Stabilisierung des muskuloskelettalen Systems (Griefahn et al., 2025).

Ergänzend zu der Bewegungstherapie bietet die App Inhalte zur Förderung des Selbstmanagements. Dazu zählen geführte Meditationen, Podcasts sowie edukative Beiträge zu Themen wie Schmerzphysiologie, Stressregulation und gesundheitsförderndem Verhalten. Ein integriertes Gesundheits-Cockpit erfasst kontinuierlich relevante Parameter, wie bspw. den Verlauf der Schmerzintensität. Die erfassten Daten können mit behandelnden Ärzt*innen oder Therapeut*innen geteilt werden, um eine gezielte Verlaufskontrolle im ambulanten Versorgungskontext zu unterstützen. Die Benutzeroberfläche der App ist in Abbildung 1 dargestellt.

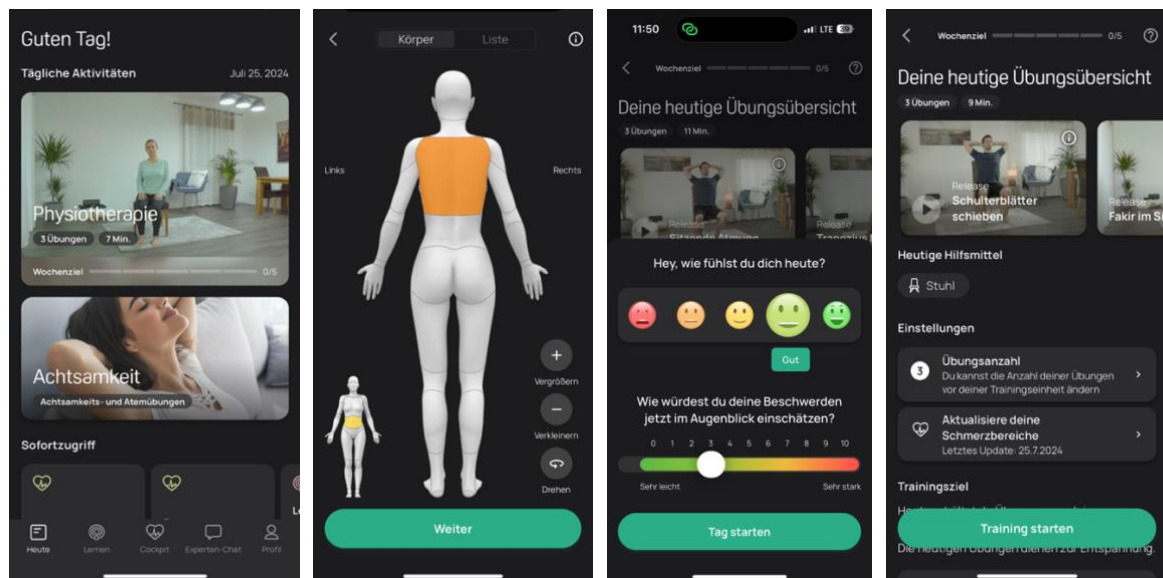


Abbildung 1: Screenshots aus der App medicalmotion
(Quelle: medicalmotion GmbH, 25.07.2024)

1.4.2. Die künstliche Intelligenz *medico*

Das Expertensystem *medico* bildet die zentrale Entscheidungsinstanz zur Auswahl der Übungen innerhalb der medicalmotion App. In der vorliegenden Forschungsarbeit wird die Version 1.2 validiert. Das Expertensystem stellt evidenzbasierte und personalisierte Übungsempfehlungen zur Behandlung von MSK bereit, wobei die Generierung dieser Empfehlungen auf den Angaben von Nutzer*innen basiert. Diese Daten werden mittels regelbasierter Logik ausgewertet, um eine täglich individuelle Übungsauswahl zu ermöglichen.

Die Wissensbasis von *medico* basiert auf der Integration heuristischer und faktischer Informationen. Das heuristische Wissen speist sich dabei aus dem Erfahrungswissen klinischer Expert*innen und umfasst typische Schmerzverläufe, die Wirkungsweise spezifischer Übungen sowie die Einschätzung physiotherapeutischer Reaktionen. Somit bildet das Expertensystem ein digitales Abbild physiotherapeutischer Entscheidungsprozesse, welche für die Behandlung von Patient*innen im klinischen Kontext entscheidend sind, ab. Das faktische Wissen hingegen bezieht sich auf medizinisch-anatomische Grundlagen, evidenzbasierte Leitlinien, bekannte Kontraindikationen sowie wissenschaftlich fundierte Erkenntnisse zu Wirkmechanismen, Zielstrukturen und Belastungsparametern einzelner Übungen (Griefahn et al., 2024b).

Der Entscheidungsprozess zur Auswahl einer individualisierten Übungsempfehlung folgt einem zweiphasigen Ablauf:

In der ersten Phase werden auf Basis individueller Eingaben der Nutzer*innen spezifische

Therapieziele definiert. Zu den Eingabemöglichkeiten zählen beispielsweise Angaben zur Schmerzintensität, zu einer oder mehreren gestellter Diagnosen, zur Arbeitsbelastung und auch Angaben zur sportlichen Aktivität. Diese Informationen werden regelmäßig aktualisiert und bilden die Grundlage für eine adaptive Priorisierung geeigneter Übungskategorien. Die zweite Phase umfasst die gezielte Selektion einzelner Übungen, die die definierten Therapieziele unter Berücksichtigung klinisch relevanter Faktoren bestmöglich berücksichtigen. In diese Auswahl fließen potenzielle Kontraindikationen, vorherige Interaktionen (z. B. ausgelassene oder abgebrochene Übungen) und das tägliche Feedback der Nutzer*innen mit ein. Zusammengefasst erfolgt diese Auswahl also einem iterativen und zyklischen Prozess, der aus Strukturierung, Selektion, Feedbackverarbeitung und Optimierung besteht (siehe Abbildung 2). Ziel dieses iterativen und zyklisch durchgeführten Prozesses ist die kontinuierliche Optimierung der Übungsempfehlungen und die Förderung der langfristigen Therapieadhärenz. Die Auswahl der Übungen erfolgt adaptiv und reagiert flexibel auf kurzfristige Veränderungen des subjektiven Schmerzgeschehens, um eine optimale Anpassung an die individuellen Bedürfnisse der Patient*innen zu gewährleisten.

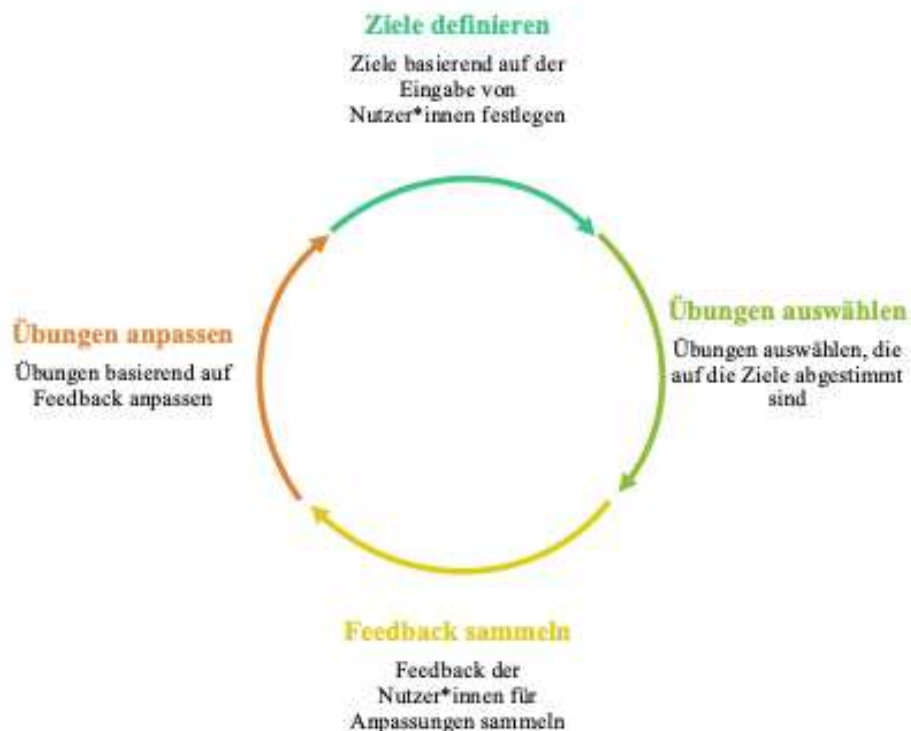


Abbildung 2: Zyklus der Übungsauswahl und -optimierung
(Quelle: Eigene Darstellung, 2025)

Ein wesentliches Merkmal von *medico* ist dessen Doppelfunktionalität: Es dient sowohl als Therapieempfehlungssystem (Therapy Recommendation System, TRS) für Patient*innen als auch als klinisches Entscheidungsunterstützungssystem (Clinical Decision Support System, CDSS) für Physiotherapeut*innen.

1.5. Forschungsziele

Die Anwendung von KI zur Generierung individualisierter Übungsempfehlungen bei MSK stellt ein bislang wenig erforschtes Einsatzgebiet dar. Während KI zunehmend in der medizinischen Diagnostik und Entscheidungsunterstützung eingesetzt werden, mangelt es an empirischer Evidenz für ihre sichere und wirksame Integration in bewegungstherapeutische Interventionen. Diese Forschungslücke wurde im Rahmen dieses Projekts durch drei aufeinander aufbauende Studien systematisch untersucht.

Studie I untersuchte, inwieweit die Einschätzungen der KI zu Kontraindikationen für eine alleinige Übungstherapie mit den Bewertungen von Physiotherapeut*innen übereinstimmten. Anhand von Leistungsparametern wie der Sensitivität, der Spezifität, der Genauigkeit und dem F1-Score wurde die diagnostische Validität der KI bewertet. Zudem wurden potenzielle Unterschiede in den klinischen Merkmalen von Patient*innen, bei denen Kontraindikationen festgestellt oder nicht festgestellt wurden, untersucht.

Studie II analysierte die Fähigkeit der KI, risikofreie Übungen zu empfehlen – solche sind definiert als Übungen, die keine symptomatische Verschlechterung hervorrufen. Darüber hinaus wurde untersucht, in welchem Maße die Risikobewertung KI-generierter Übungsempfehlungen zwischen Physiotherapeut*innen übereinstimmte. Zudem wurde das Übungsfeedback der Patient*innen ausgewertet, um Unterschiede in der subjektiven Wahrnehmung in Abhängigkeit von der physiotherapeutischen Risikoeinschätzung zu ermitteln.

Studie III untersuchte die Auswirkungen einer achtwöchigen, appbasierten und KI-generierten Übungsempfehlung auf Schmerzintensität und subjektives Wohlbefinden von Patient*innen mit Rückenschmerzen. Der Zusammenhang zwischen der Frequenz der Übungen und den Veränderungen der Schmerzintensität bzw. des subjektiven Wohlbefindens wurde ebenfalls statistisch untersucht. Das Nutzungsverhalten in der App wurde im Hinblick auf Sitzungsdauer und -frequenz während des Interventionszeitraums analysiert.

2. DARSTELLUNG DER PUBLIKATIONEN

Die drei in dieser Dissertation enthaltenen Originalarbeiten bauen thematisch aufeinander auf und widmen sich jeweils einem spezifischen Aspekt des Expertensystems: der diagnostischen Validität hinsichtlich Kontraindikationen für eine Übungstherapie (Studie I), der Risikoeinschätzung (Studie II) und der Wirksamkeit im Versorgungskontext (Studie III). Mit dem Expertensystem *medico*, das über die App *medicalmotion* bereitgestellt wird, liegt allen Studien dasselbe technische System zu Grunde. Trotz unterschiedlicher Studiendesigns verbindet die Arbeiten das gemeinsame Ziel, die klinische Relevanz und Umsetzbarkeit KI-generierter Übungsempfehlungen im bewegungstherapeutischen Setting zu bewerten. Die Ergebnisse bilden eine empirisch fundierte Grundlage für die Diskussion über Potenziale, Limitationen und praxisnahe Anwendungsmöglichkeiten KI-gestützter Therapieansätze bei Patient*innen mit MSK.

2.1. Publikation I

Griefahn A, Zalpour C, Luedtke K (2024) Evaluation of the accuracy of an artificial intelligence in identifying contraindications to exercise therapy - Comparison with and interrater reliability of physical therapists judgments. *Health and Technology*, 14, 513–522. <https://doi.org/10.1007/s12553-024-00827-w>

Hintergrund und Zielsetzung

Die Integration von KI in die physiotherapeutische Entscheidungsfindung eröffnet neue Möglichkeiten für eine standardisierte und objektivierte Ableitung von Therapieempfehlungen. Ein relevantes Anwendungsfeld ist die Identifikation potenzieller Kontraindikationen für Bewegungstherapien bei Patient*innen mit MSK. Die Studie befasste sich mit der Erkennung von Kontraindikationen für eine Übungstherapie. Hierbei wurde die Übereinstimmung zwischen der KI und physiotherapeutischen Bewertungen analysiert. Zudem erfolgte die Bewertung von KI-spezifischen Leistungsparametern.

Methode

Im Rahmen einer standardisierten Online-Befragung bewerteten 20 examinierte Physiotherapeut*innen 80 anonymisierte Fälle von Patient*innen hinsichtlich des Vorliegens einer Kontraindikation für eine alleinige Übungstherapie. Die Bewertungen der examinierten Physiotherapeut*innen basierten auf den Fällen der Patient*innen mit Angaben

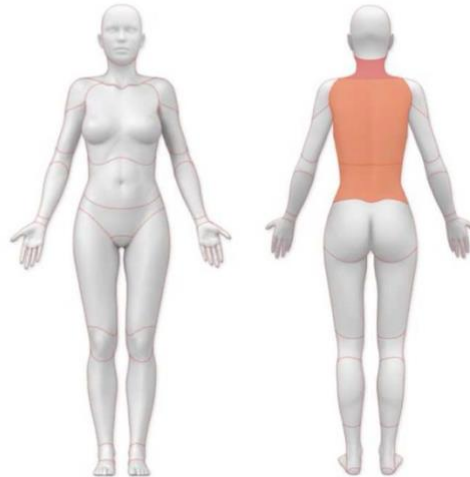
zu Schmerzcharakteristik, Diagnosen, Alter, Geschlecht, Operationen und weiteren relevanten Gesundheitsdaten. Abbildung 3 zeigt einen der dargestellten Fälle.

Alter: 38 Jahre

Diagnosen: E66 - Adipositas, M54.5 - Kreuzschmerzen, M54.4 - Lumboischialgie, R51 - Kopfschmerz

Operationen: keine

Übersicht Schmerzbereiche:



VORDERANSICHT

Rücken

RÜCKANSICHT

Lendenbereich

8 SEIT Mehr als 6 Monate
 TYP Stechend, Ziehend, Brennend, Drückend
 WANN Bei sportlicher Aktivität, Dauerhaft

8 SEIT Mehr als 6 Monate
 TYP Stechend
 WANN Bei sportlicher Aktivität

Nacken

9 SEIT Mehr als 6 Monate
 TYP Ziehend, Stechend
 WANN Dauerhaft

Abbildung 3: Patientenbeispiel für die Bewertung von Kontraindikationen (Quelle: in Anlehnung an (Griefahn et al., 2024a))

Es wurde die prozentuale Übereinstimmung sowie der Fleiss-Kappa-Koeffizient (κ_{π}) zur Bewertung der Übereinstimmung hinsichtlich des Vorliegens einer Kontraindikation zwischen den Physiotherapeut*innen herangezogen. Die Übereinstimmung zwischen der KI und den einzelnen Physiotherapeut*innen wurde zusätzlich mittels gepooltem Cohen's Kappa (κ_{pooled}) analysiert.

Die Ableitung eines Konsens über das Vorliegen oder Nichtvorliegen einer Kontraindikation erfolgte anhand des folgenden Klassifikationsschemas:

- **Kontraindikation vorhanden:** Wenn mindestens 67 % der Physiotherapeut*innen angaben, dass eine Kontraindikation besteht, wurde das Vorliegen einer Kontraindikation angenommen.
- **Keine Kontraindikation vorhanden:** Wenn mehr als 75 % der Physiotherapeut*innen angaben, dass **keine** Kontraindikation besteht, wurde angenommen, dass keine Kontraindikation vorliegt.
- **Kein Konsens:** Wenn beide Kriterien nicht erfüllt waren – das heißt, wenn der Anteil der zustimmenden bzw. ablehnenden Bewertungen zwischen diesen Schwellenwerten lag – wurde das Ergebnis als **kein Konsens** gewertet.

Die Schwellenwerte wurden auf Basis methodischer Standards aus Delphi-Studien festgelegt, um eine möglichst klare Entscheidungsgrundlage für die Bewertung der KI zu schaffen.

Die Beurteilung der Leistungsparameter der KI erfolgte anhand der Sensitivität, der Spezifität, der Genauigkeit und dem F1-Score. Darüber hinaus wurde mittels einfaktorieller Varianzanalyse (ANOVA) mit Bonferroni-Post-hoc-Korrektur geprüft, ob sich die klinischen Merkmale der Patient*innen je nach Konsensgruppe (Kontraindikation vorhanden, keine Kontraindikation vorhanden, kein Konsens) signifikant unterschieden (Signifikanzniveau $p = 0,05$).

Ergebnisse

Die Physiotherapeut*innen bewerteten 43,75 % der Fälle mit „Kontraindikation vorhanden“ und 36,25 % mit „keine Kontraindikation vorhanden“. In 20 % der Fälle ergab sich kein Konsens zwischen den Physiotherapeut*innen. Die Übereinstimmung innerhalb der Gruppe der Physiotherapeut*innen lag bei 87,5 % ($\kappa_{\pi} = 0,43$), dies entspricht einer mittleren Übereinstimmung. Die Anzahl der als kontraindiziert bewerteten Fälle reichte von 18 bis 64. Die Übereinstimmung zwischen der KI und dem Konsens der Physiotherapeut*innen betrug $\kappa_{\text{pooled}} = 0,63$ und einer prozentualen Übereinstimmungsrate zwischen 53,75 % und 100 %. Die Leistungsparameter der KI erreichten optimale Werte (Sensitivität, Spezifität, Genauigkeit, F1-Score = 1,0). Darüber hinaus zeigten sich signifikante Unterschiede in bestimmten klinischen Merkmalen zwischen den Gruppen „Kontraindikation vorhanden“, „keine Kontraindikation vorhanden“ und „kein Konsens“ ($p < 0,05$).

Schlussfolgerung

Aus den oben aufgeführten Ergebnissen lässt sich schlussfolgern, dass die KI zuverlässig zur standardisierten Bewertung potenzieller Kontraindikationen für eine alleinige Übungstherapie bei Patient*innen mit MSK beitragen kann. Die beobachtete Übereinstimmung zwischen KI und Physiotherapeut*innen verdeutlicht zusätzlich ihr Potenzial, klinische Entscheidungsprozesse zu unterstützen. Gleichzeitig zeigen die Unterschiede in den Bewertungen, dass physiotherapeutische Diagnosen in hohem Maße subjektiv geprägt sind und von individuellen Erfahrungswerten abhängen.

Die hohe interindividuelle Streuung innerhalb der Gruppe der Physiotherapeut*innen unterstreicht die Variabilität klinischer Entscheidungen. Insbesondere im Umgang mit ambivalenten Fällen und bei der Weiterentwicklung von Algorithmen, die komplexe Entscheidungslogiken abbilden sollen, bestehen weiterhin Herausforderungen.

Die Ergebnisse bestätigen Erkenntnisse aus anderen Fachbereichen, die eine hohe Übereinstimmung zwischen KI und Fachpersonal dokumentieren. Für den Praxiseinsatz ist jedoch eine kontinuierliche Validierung solcher Systeme notwendig. Zudem erfordert die Einbindung von KI in den Praxisalltag eine Begleitung durch erfahrene Fachkräfte und eine kritische Reflexion.

2.2. Publikation II

Griefahn A, Zalpour C & Luedtke K (2024) Identifying the risk of exercises, recommended by an artificial intelligence for patients with musculoskeletal disorders. *Scientific Reports*, 14, 14472. <https://doi.org/10.1038/s41598-024-65016-1>.

Hintergrund und Zielsetzung

Der Einsatz von KI in der muskuloskelettalen Rehabilitation gewinnt an Bedeutung. Aufbauend auf den Erkenntnissen aus Publikation I zur Identifikation von Kontraindikationen stellt sich die Frage, inwieweit die KI risikofreie Übungen empfehlen kann. Die Studie erfasste physiotherapeutische Einschätzungen zum Risiko der von der KI generierten Übungen und setzte sie mit dem subjektiven Übungsfeedback von Patient*innen in Beziehung.

Methodik

Die Untersuchung basierte auf einer Stichprobe von 80 Patient*innen mit MSK, die die medicalmotion App zwischen Januar 2020 und Juni 2021 nutzten. Für jeden Fall dieser

Stichprobe lagen standardisierte Angaben zur Schmerzlokalisierung, Schmerzart, Schmerzintensität, Schmerzdauer und Schmerzqualität vor. Soziodemografische Daten sowie Angaben zu Diagnosen und zu Operationen wurden ebenfalls berücksichtigt. Die KI generierte für diese Patient*innen 944 individualisierte Übungsempfehlungen.

Eine Gruppe von 20 Physiotherapeut*innen mit Deutschkenntnissen, mindestens fünf Jahren Berufserfahrung und überwiegender Tätigkeit im MSK-Bereich (über 80 % der Arbeitszeit) bewertete alle empfohlenen Übungen auf potenzielle Risiken. Als Risiko wurde eine mögliche Schmerzverstärkung oder der Anstieg des empfundenen Unwohlseins definiert. Die Einschätzung erfolgte binär („Risiko“ vs. „kein Risiko“). Abbildung 4 zeigt den Studienablauf als Flussdiagramm.

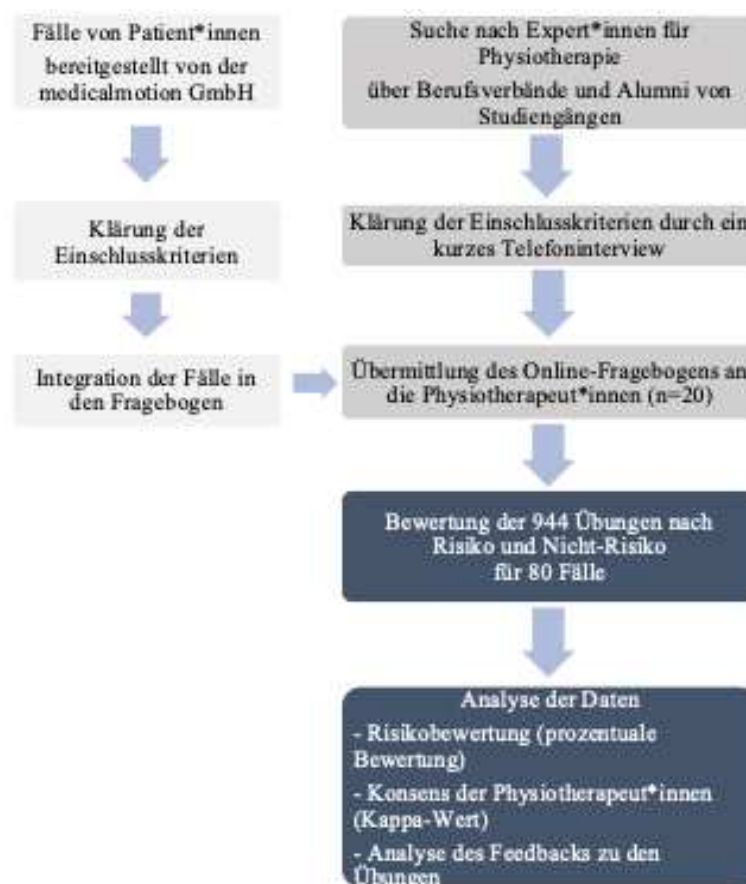


Abbildung 4: Studienablauf für die Risikobewertung
(Quelle: in Anlehnung an (Griefahn et al., 2024b))

Neben der qualitativen Risikoeinschätzung wurde die Interrater-Reliabilität anhand von Fleiss-Kappa (κ_{π}), Cohen's-Kappa (κ_d) und der prozentualen Übereinstimmung berechnet. Bei einer großen Abweichung zwischen der prozentualen Übereinstimmung und den berechneten Kappa-Werten wurde der Brennan & Prediger Koeffizient (κ_q) zusätzlich berechnet. In diesen Fällen war eine große Abweichung definiert mit Kappa-Werten von $\kappa < 0,00$ oder $0,00 \leq \kappa \leq 0,20$, aber einer prozentualen Übereinstimmung von $> 80,0 \%$.

Ein unabhängiger t-Test wurde durchgeführt, um – abhängig von der Risikobewertung – Unterschiede im Übungsfeedback der Patient*innen zu prüfen.

Ergebnisse

Es konnte ein maximaler Konsens zwischen vier Physiotherapeut*innen über das Vorliegen eines Risikos bei drei Übungen für die Fälle erreicht werden. Dies bedeutet, dass bei dem maximal erzielten Konsens zwischen vier Physiotherapeut*innen 99,9 % der 944 bewerteten Übungen als risikofrei eingestuft wurden. Bei der individuellen Bewertung durch eine*n Physiotherapeut*in wurden 247 Übungen als risikoreich eingestuft. In Tabelle 2 wird die Bewertung der Physiotherapeut*innen detailliert aufgezeigt.

Tabelle 2: Überblick über die Anzahl der Übungen, die von Physiotherapeut*innen als Risikoübungen eingestuft werden

Übereinstimmung pro Fall der Patient*innen zwischen	Gesamtzahl der Risikobewertung pro Übung in den Fällen der Patient*innen (absolut)	Gesamtzahl der Risikobewertung pro Übung in den Fällen der Patient*innen (%)
Eine*n Physiotherapeut*in	247	1.31 %
Zwei Physiotherapeut*innen	48	0.25 %
Drei Physiotherapeut*innen	14	0.08 %
Vier Physiotherapeut*innen	3	0.03 %

(Quelle: in Anlehnung an Griefahn et al. (2024b))

Die durchschnittliche Übereinstimmung über die Gesamtzahl der Fälle für risikofreie Übungen zwischen den Physiotherapeut*innen lag bei 90,5 %. Der Fleiss-Kappa zeigte eine geringe Übereinstimmung mit $\kappa_{\pi} = 0,017$, wohingegen der Brennan & Prediger Koeffizient eine substanzielle Übereinstimmung mit $\kappa_q = 0,643$ aufwies.

Das Übungsfeedback der Patient*innen zeigte signifikante Unterschiede zwischen risikofreien und risikoreichen Übungen.

Schlussfolgerung

Die Ergebnisse belegen, dass die Einschätzung potenzieller Risiken durch Physiotherapeut*innen für Übungen subjektiv ist – Risiken wurden je nach Physiotherapeut*in unterschiedlich bewertet. Die Streuung in der Bewertung der einzelnen Übungen hinsichtlich eines Risikos durch die Physiotherapeut*innen, trotz formaler Qualifikationen und Spezialisierungen, verdeutlicht die Komplexität und Individualität klinischer Entscheidungsprozesse im Bereich der Bewegungstherapie. Die signifikanten

Ergebnisse zwischen der Risikoeinschätzung der Physiotherapeut*innen und dem Übungsfeedback der Patient*innen zeigt, dass die subjektive Wahrnehmung der Patient*innen eine wichtige Rolle bei der Bewertung KI-generierter Übungsempfehlungen spielt. Dies unterstützt die Forderung, technologische Systeme nicht ausschließlich anhand formaler Kriterien zu bewerten, sondern auch patientenzentrierte Rückmeldungen systematisch zu integrieren. Abschließend bleibt zu erwähnen, dass die Ergebnisse das Potenzial und die Herausforderungen bei der Implementierung von KI im physiotherapeutischen Kontext aufzeigen.

2.3. Publikation III

Griefahn A, Avermann F, Zalpour C, Marshall RP, Cordon Morillas I, Luedtke K (2025) Exploring the effect of an 8-week AI-composed exercise program on pain intensity and well-being in patients with spinal pain: Retrospective cohort analysis. *JMIR Formative Research*, 9, e57826. <https://doi.org/10.2196/57826>

Hintergrund und Zielsetzung

Zahlreiche Studien belegen eine grundsätzliche Wirksamkeit digital unterstützter Übungsprogramme. Dennoch wurde bisher unzureichend untersucht, welche Effekte die Anwendung KI-generierter Übungsempfehlungen bei Patient*innen mit Rückenschmerzen haben. Folglich war es das Ziel der Studie, die Wirkung eines achtwöchigen Programms mit KI-generierten Übungsempfehlungen über eine App auf Schmerzintensität und subjektives Wohlbefinden zu evaluieren. Dabei wurde ebenfalls geprüft, ob ein Zusammenhang zwischen Nutzungsverhalten, Übungshäufigkeit, Schmerzintensität und Wohlbefinden besteht.

Methodik

Im Rahmen einer retrospektiven Analyse wurden die Nutzungsdaten der medicalmotion App ausgewertet. In die Analyse wurden Nutzer*innen eingeschlossen, die sich zwischen Januar 2020 und Juni 2023 registriert und das Vorhandensein von Rückenschmerzen im Eingangsfragebogen bejaht haben, zwischen 18 und 65 Jahre alt waren und mindestens acht dokumentierte App-Interaktionen innerhalb von acht Wochen aufwiesen. Nutzer*innen mit neurologischen oder psychischen Erkrankungen, systemischen Infektionen oder psychopharmakologischer Medikation wurden ausgeschlossen.

Erfasst wurden demografische Variablen wie Alter, Geschlecht und die initiale Schmerzregion sowie anwendungsbezogene Parameter wie Nutzungsdauer und Übungszeit.

Die Schmerzintensität wurde mithilfe einer elfstufigen numerischen Ratingskala (NRS) zu drei Zeitpunkten erhoben (t0 zu Beginn, t1 nach vier Wochen und t2 nach acht Wochen). Das subjektive Wohlbefinden wurde mit einer elfstufigen Likert-Skala dokumentiert. Zur statistischen Auswertung wurden eine ANOVA mit Messwiederholung sowie Pearson- und Spearman-Korrelationen eingesetzt.

Ergebnisse

Die analysierte Stichprobe umfasste 379 Nutzer*innen mit einem Durchschnittsalter von 50,96 Jahren (SD = 12,22). Die durchschnittliche Schmerzintensität lag zu t0 bei 6,08 Punkten (SD = 2,16) auf der NRS. Zu t2 zeigte sich eine Reduktion der Schmerzintensität um 1,78 Punkte (SD = 2,05). Der Faktor „Zeit“ zeigte ein signifikantes Ergebnis ($F_{1,88, 710,96} = 179,86$; $p < 0,001$; $\eta^2 = 0,32$). Ebenso zeigte die Post-hoc-Analyse mit Bonferroni-Korrektur signifikante Unterschiede zu jedem Messzeitpunkt ($p < 0,001$).

Das Wohlbefinden, auf einer elfstufigen Likert-Skala, lag zu t0 bei 4,95 Punkten (SD = 1,65). Zu t2 stieg es auf 8,06 Punkte (SD = 1,38). In diesem Fall zeigte der Faktor „Zeit“ ein signifikantes Ergebnis ($F_{1,94, 732,57} = 671,97$; $p < 0,001$; $\eta^2 = 0,64$). Die Post-hoc-Analyse mit Bonferroni-Korrektur zeigte zu jedem Messzeitpunkt einen signifikanten Unterschied ($p < 0,001$).

Die App wurde im Mittel an 49,2 % der insgesamt 56 verfügbaren Tage genutzt, wobei die durchschnittliche Übungsdauer pro Einheit 597,86 Sekunden (ca. zehn Minuten) betrug. Zwischen der täglichen Übungshäufigkeit und der Schmerzintensität zeigten sich signifikant negative Korrelationen, sowohl zu t1 ($p < 0,001$) als auch zu t2 ($p = 0,004$). Auch die Gesamtzahl absolvierter Übungen stand zu t2 in einem signifikant negativen Zusammenhang mit der Schmerzintensität ($p = 0,04$). Darüber hinaus konnte ein signifikant positiver Zusammenhang zwischen der täglichen Übungshäufigkeit und dem subjektiven Wohlbefinden festgestellt werden ($p = 0,01$).

Diskussion

Die Ergebnisse deuten darauf hin, dass KI-generierte Übungsempfehlungen die Schmerzintensität und das subjektive Wohlbefinden von Nutzer*innen mit Rückenschmerzen positiv beeinflussen. Die Studienergebnisse verdeutlichen das Potenzial digitaler Interventionen bei dieser Gruppe von Nutzer*innen.

Die Heterogene Zusammensetzung der Nutzer*innen hinsichtlich Diagnose und Schmerzcharakteristika erschwert eine generalisierbare Übertragbarkeit der oben aufgeführten Ergebnisse. Auch wurde keine differenzierte Analyse der Effektivität einzelner

Übungsarten durchgeführt. Zuletzt ist anzuführen, dass im Rahmen der Studie nicht bewertet wurde, ob die gemessene Wirksamkeit durch den individualisierten Algorithmus oder die spezifischen Übungsformen begründet war.

Schlussfolgerung

Die Ergebnisse der Studie zeigen eine signifikante Verbesserung von Schmerzintensität und subjektivem Wohlbefinden durch individuell angepasste und KI-generierte Übungsempfehlungen über eine App – so führte eine intensivere Nutzung zu besseren Ergebnissen. Zur Absicherung dieser Ergebnisse sind jedoch weitere Studien mit homogeneren Stichproben, standardisierter Diagnostik und längeren Beobachtungszeiträumen notwendig.

3. GESAMTKONKLUSION

Die vorliegende Dissertation widmet sich dem Einsatz von KI zur Generierung individualisierter Übungsempfehlungen für Patient*innen mit MSK. In drei empirischen Studien wurde untersucht, inwiefern eine KI eine valide, risikofreie und wirksame Bewegungstherapie empfehlen kann – ein bisher wenig erforschter Bereich an der Schnittstelle von Digitalisierung und Bewegungstherapie.

Als Ergebnis der ersten Studie lässt sich festhalten, dass die eingesetzte KI klinisch relevante Kontraindikationen für eine eigenständige Übungstherapie zuverlässig identifizieren kann. Die hohe Übereinstimmung mit den Einschätzungen erfahrener Physiotherapeut*innen weist darauf hin, dass die KI ohne Risiko für die Patient*innen als eigenständige Übungstherapie anwendbar ist. Für Indikationen aus dem MSK-Bereich könnte eine ergänzende fachliche Prüfung somit entbehrlich sein. Gleichzeitig unterstreicht die Studie die Notwendigkeit weiterer Forschung zur Implementierung von KI in den Versorgungsalltag.

Die zweite Studie untersuchte, ob die KI risikofreie Übungen empfehlen kann. Bei dem maximal erzielten Konsens – zwischen vier Physiotherapeut*innen – wurden 99,9 % der Übungen als risikofrei eingestuft. Gleichzeitig offenbarte sich jedoch eine erhebliche Heterogenität in den Bewertungen: Während einzelne Physiotherapeut*innen bis zu 247 Übungen als risikobehaftet klassifizierten, bewerteten andere hingegen keine einzige Übung als risikoreich. Diese Diskrepanz verdeutlicht die Subjektivität klinischer Entscheidungsprozesse und unterstreicht die Notwendigkeit einer integrativen Bewertungsperspektive, die sowohl die professionelle Expertise als auch das Übungsfeedback der Patient*innen berücksichtigt.

Die dritte Studie analysierte die Wirksamkeit einer appbasierten Intervention mit KI-generierten Übungsempfehlungen nach einer Anwendungsdauer von acht Wochen. Dabei zeigte sich eine signifikante Reduktion der Schmerzintensität und eine Verbesserung des subjektiven Wohlbefindens. Die statistisch signifikanten negativen Korrelationen zwischen Adhärenz und den benannten Veränderungen in der Schmerzintensität bzw. dem subjektiven Wohlbefinden deuten darauf hin, dass das individuelle Nutzungsverhalten einen relevanten Einfluss auf den Therapieerfolg nimmt. Aufgrund der Heterogenität der Stichprobe und der fehlenden Kontrollgruppe lässt sich der spezifische Beitrag der KI zur Wirksamkeit jedoch nicht abschließend bewerten – diese methodische Limitation retrospektiver Beobachtungsdesigns erfordert eine differenzierte Betrachtung der Ergebnisse.

Grundsätzlich schränkt das Fehlen einer Kontrollgruppe die Interpretierbarkeit solcher Ergebnisse erheblich ein, weshalb alternative Erklärungen für die beobachteten Verbesserungen nicht ausgeschlossen werden können. Die Symptomreduktion könnte auf natürliche Krankheitsverläufe zurückzuführen sein: Spontane Remissionen treten bei Rückenschmerzen – insbesondere bei akuten und subakuten Verläufen – häufig auf (Pengel et al., 2003). Zusätzlich ist das statistische Phänomen der Regression zur Mitte zu berücksichtigen: Patient*innen suchen typischerweise zu Zeitpunkten besonders hoher Symptomlast Unterstützung, sodass eine Rückkehr zu weniger extremen Werten auch ohne therapeutische Intervention zu erwarten ist (Barnett et al., 2005). Darüber hinaus können psychologische Mechanismen wie der Hawthorne-Effekt, Erwartungseffekte gegenüber digitalen Innovationen oder die Selbstselektion besonders motivierter Patient*innen zur Verbesserung beigetragen haben (Grimes & Schulz, 2002; Rossetini et al., 2020; Sedgwick & Greenwood, 2015). Letztlich ist nicht auszuschließen, dass parallel erfolgte Interventionen wie physiotherapeutische Behandlungen oder die Einnahme von Analgetika die Ergebnisse beeinflusst haben, ohne dass deren spezifischer Beitrag quantifiziert werden kann (Grimes & Schulz, 2002).

Trotz dieser methodischen Einschränkungen liefert das retrospektive Studiendesign Erkenntnisse für die Weiterentwicklung der Forschung in diesem Bereich. Die Ergebnisse bilden eine empirische Grundlage für die Planung prospektiver, randomisiert-kontrollierter Studien, etwa durch die Bereitstellung von Daten für Fallzahlberechnungen, die Identifikation relevanter Outcome-Parameter und die Abschätzung realistischer Effektgrößen. Künftige Studien sollten durch geeignete Kontrollgruppen den spezifischen Wirkmechanismus der KI-basierten Übungsempfehlungen systematisch untersuchen. Wartelisten-Kontrollen würden die Abgrenzung von einem natürlichen Verlauf und unspezifischen Effekten ermöglichen, während aktive Kontrollen mit standardisierten

Übungsprogrammen den Mehrwert der KI-basierten Personalisierung quantifizierbar machen würden.

Vor dem Hintergrund struktureller Versorgungsengpässe in der Physiotherapie, etwa begrenzter Behandlungskapazitäten und langer Wartezeiten, erscheint der Einsatz KI-gestützter Systeme als sinnvoller Beitrag zur Steigerung von Effizienz und Adhärenz. Digitale Anwendungen wurden bereits als Pilotprojekte im Rahmen politischer Initiativen wie der „Zukunftsregion Digitale Gesundheit“, die vom Bundesministerium für Gesundheit gefördert wurde, erfolgreich in den Versorgungsalltag integriert (Sulzer et al., 2022). Auch die Ergebnisse der Umfrage des Verbands Physio Deutschland verdeutlichen den Bedarf an digitalen Lösungen für die Integration in die Regelversorgung: So beträgt die durchschnittliche Wartezeit auf eine Erstbehandlung mindestens drei Wochen (Physio Deutschland, 2018). Des Weiteren tragen unregelmäßige Terminvergaben zu einer Beeinträchtigung der Kontinuität physiotherapeutischer Maßnahmen bei. Der Einsatz von KI-basierten Apps könnte in diesem Kontext durch adaptive Programme und intelligente Zeitsteuerung zur Optimierung der Planbarkeit und Versorgung beitragen. Die stetige Implementierung digitaler Anwendungen in den Versorgungsalltag könnte weitergehend dazu führen, die Physiotherapie zu unterstützen und zu entlasten und langfristig die allgemeine Akzeptanz von KI zu steigern. Um den Übergang von der Forschung in die Praxis zu erleichtern, bedarf es konkreter Implementierungsstrategien. Dazu gehören standardisierte Schnittstellen zu digitalen Akten der Patient*innen und Praxissoftware, Integration in digitale Therapieplanung sowie qualifizierende Fortbildungsformate, die sowohl technologische als auch ethische und kommunikative Kompetenzen fördern (Ramachandran et al., 2024; Regan, 2022). Der reflektierte Umgang mit algorithmischen Empfehlungen sollte dabei ein zentrales Element physiotherapeutischer Bildung sein.

Der Einsatz technischer Systeme darf nicht zur Marginalisierung physiotherapeutischer Fachkompetenz führen, sondern sollte diese sinnvoll unterstützen. Unersetzliche Elemente wie Empathie, patientenzentrierte Kommunikation und Clinical Reasoning bleiben zentrale Bestandteile qualitativ hochwertiger physiotherapeutischer Versorgung (Davenport & Kalakota, 2019; Moudatsou et al., 2020). Auch im Umgang mit psychologischen Einflussfaktoren wie Motivation oder Antriebsschwäche zeigen sich Limitationen seitens KI-basierter Systeme, da diese oft nicht hinreichend durch standardisierte Eingabemasken abgebildet werden können und relevante Faktoren für Adhärenz und Verhaltensänderung sind. Eine gezielte Progression im Trainingsverlauf sowie adaptive Reaktionen auf sich verändernde gesundheitliche Rahmenbedingungen stellen bislang eine Herausforderung dar. Die World Confederation for Physical Therapy und International Network of Physiotherapy

Regulatory Authorities (2020) definieren die digitale Praxis als Bereitstellung physiotherapeutischer Leistungen über digitale Kommunikationsmittel mit dem Ziel, Zugang, Effizienz und Ressourcennutzung zu optimieren. Gleichzeitig betonen sie jedoch, dass diese Systeme ausschließlich als Ergänzung zur physiotherapeutischen Expertise zu verstehen sind.

Darüber hinaus bestehen weiterhin grundlegende ethische und haftungsrechtliche Herausforderungen im Zusammenhang mit der Integration KI-gestützter Systeme in die physiotherapeutische Versorgung (Hafeez et al., 2024; Laurijssen et al., 2025). Die Übertragung klinischer Entscheidungsprozesse auf algorithmisch arbeitende Systeme macht verbindliche Regelungen zur Verantwortungszuweisung, zur Nachvollziehbarkeit von Entscheidungen sowie zur Revisionsfähigkeit erforderlich. Insbesondere im Hinblick auf fehlerhafte Empfehlungen müssen haftungsrechtliche Zuständigkeiten eindeutig definiert sein, um die Sicherheit der Patient*innen und rechtliche Absicherung für alle Beteiligten zu gewährleisten (Saenz et al., 2023). Eine transparente Systemarchitektur und dokumentierte Entscheidungslogik gelten dabei als zentrale Voraussetzungen für eine verantwortungsvolle Nutzung (Bleher & Braun, 2022).

Die Ergebnisse dieser Arbeit verdeutlichen das Spannungsfeld zwischen dem innovativen Potenzial KI-gestützter Systeme und den komplexen Anforderungen an deren verantwortungsvolle Implementierung in die physiotherapeutische Praxis. Unter geeigneten Rahmenbedingungen können solche Systeme valide, risikofreie und wirksame Empfehlungen generieren. Ihre zukünftige Relevanz wird jedoch maßgeblich davon abhängen, inwiefern es gelingt, technische Innovationen mit klinischer Praxis, ethischer Verantwortung und gesundheitsökonomischer Nachhaltigkeit zu verbinden. Eine nachhaltige Integration erfordert daher nicht nur eine belastbare Evidenzbasis, sondern auch interprofessionelle Entwicklungsprozesse, eine regulatorische Einbettung und ethische Reflexion. Hierfür wird es notwendig sein, geeignete Studiendesigns zu entwickeln, die die Wirksamkeit und Sicherheit KI-gestützter Systeme unter Alltagsbedingungen überprüfbar machen. Diese Forschungsarbeit leistet hierzu einen Beitrag, indem sie die Validierung und Wirkung von KI in der Bewegungstherapie systematisch untersucht und empirisch fundiert. Sie erweitert das bestehende Forschungsfundament zur digitalen Gesundheitsversorgung. Dabei beteiligt sie sich am internationalen Diskurs über die Rolle digitaler Technologien im therapeutischen Kontext, besonders im deutschsprachigen Raum, indem empirische Untersuchungen bislang begrenzt sind.

4. LITERATURVERZEICHNIS

- Albrecht, U.-V., & Jungmann, S. (2016). Kapitel 3. Gesundheits-Apps und politische Rahmenbedingungen. In U.-V. Albrecht (Hrsg.), *Chancen und Risiken von Gesundheits-Apps (CHARISMHA)* (S. 84–99). Medizinische Hochschule Hannover. <http://www.digibib.tu-bs.de/?docid=60008>
- Ärztezeitung (2023). Physiotherapeuten: Wartezeiten auf Termin immer länger. <https://www.aerztezeitung.de/Nachrichten/Physiotherapeuten-Wartezeiten-auf-Termin-immer-laenger-439183.html> (Tag des Zugriffs: 25.08.2025)
- Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*, 34(1), 215–220. <https://doi.org/10.1093/ije/dyh299>
- Bleher, H., & Braun, M. (2022). Diffused responsibility: attributions of responsibility in the use of AI-driven clinical decision support systems. *AI and Ethics*, 2(4), 747–761. <https://doi.org/10.1007/s43681-022-00135-x>
- Brenscheidt S., Siefer A., Hinnenkamp H. & Hünefeld L. (2018). Arbeitswelt im Wandel – Zahlen, Daten, Fakten. *Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)*. <https://doi.org/10.21934/baua:praxis20180131>
- Bundesministerium für Gesundheit. (2019). Gesetz für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale-Versorgung-Gesetz – DVG). *Bundesgesetzblatt Teil I, 2019(45)*, 2562–2576. https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBI&start=//%5B@attr_id='bgbl119s2562.pdf'%5D
- Bundesverband selbstständiger Physiotherapeut*innen – IFK e. V. (2023). Rechtzeitig Physiotherapietermin vereinbaren. <https://ifk.de/print/pdf/node/14183> (Tag des Zugriffs: 25.08.2025)
- Cieza, A., Causey, K., Kamenov, K., Hanson, S. W., Chatterji, S., & Vos, T. (2021). Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, 396(10267), 2006–2017. [https://doi.org/10.1016/S0140-6736\(20\)32340-0](https://doi.org/10.1016/S0140-6736(20)32340-0)
- Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2), 94–98. <https://doi.org/10.7861/futurehosp.6-2-94>
- Duran, H. (2021). medicalmotion: 7 Fragen - Munich Startup. *Munich Startup*. <https://www.munich-startup.de/68127/medicalmotion-7-fragen/> (Tag des Zugriffs: 25.08.2025)
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>

- Gill, T. K., Mittinty, M. M., March, L. M., Steinmetz, J. D., Culbreth, G. T., Cross, M., Kopec, J. A., Woolf, A. D., Haile, L. M., Hagins, H., Ong, K. L., Kopansky-Giles, D. R., Dreinhofer, K. E., Betteridge, N., Abbasian, M., Abbasifard, M., Abedi, K., Adesina, M. A., Aithala, J. P., ... Brooks, P. M. (2023). Global, regional, and national burden of other musculoskeletal disorders, 1990–2020, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. *The Lancet. Rheumatology*, 5(11), e670–e682. [https://doi.org/10.1016/s2665-9913\(23\)00232-1](https://doi.org/10.1016/s2665-9913(23)00232-1)
- Griefahn, A., Avermann, F., Zalpour, C., Marshall, R. P., Cordon Morillas, I., & Luedtke, K. (2025). Exploring the effect of an 8-week AI-composed exercise program on pain intensity and well-being in patients with spinal pain: Retrospective cohort analysis. *JMIR Formative Research*, 9, e57826–e57826. <https://doi.org/10.2196/57826>
- Griefahn, A., Zalpour, C., & Luedtke, K. (2024a). Evaluation of the accuracy of an artificial intelligence in identifying contraindications to exercise therapy - Comparison with and interrater reliability of physical therapists judgments. *Health and Technology*, 14, 513–522. <https://doi.org/10.1007/s12553-024-00827-w>
- Griefahn, A., Zalpour, C., & Luedtke, K. (2024b). Identifying the risk of exercises, recommended by an artificial intelligence for patients with musculoskeletal disorders. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-65016-1>
- Grimes, D. A., & Schulz, K. F. (2002). Bias and causal associations in observational research. *The Lancet*, 359(9302), 248–252. [https://doi.org/10.1016/S0140-6736\(02\)07451-2](https://doi.org/10.1016/S0140-6736(02)07451-2)
- Grosan, C., & Abraham, A. (2011). Rule-Based Expert Systems. In C. Grosan & A. Abraham (Hrsg), *Intelligent Systems: A Modern Approach* (S. 149–185). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21004-4_7
- Grundy, Q. (2022). A Review of the Quality and Impact of Mobile Health Apps. *Annual Review of Public Health*, 43, 117–134. <https://doi.org/10.1146/annurev-publhealth-052020-103738>
- Hafeez, M., Haq, M. Z. U., Tahzeem, Z., Rahim, S., & Rao, N. (2024). AI associated challenges to physical therapy profession. *Insights-Journal of Health and Rehabilitation*, 2(2), 531–538. <https://doi.org/10.71000/jj1x96329a>
- Heidemann, C., Scheidt-Nave, C., Beyer, A.-K., Baumert, J., Thamm, R., Maier, B., Neuhauser, H., Fuchs, J., Kuhnert, R., & Hapke, U. (2021). Gesundheitliche Lage von Erwachsenen in Deutschland – Ergebnisse zu ausgewählten Indikatoren der Studie GEDA 2019/2020-EHIS. *Journal of Health Monitoring*, 6(1), 1–27. <https://doi.org/10.25646/8456>
- Heisig, J., Lindner, N., Kornder, N., Reichert, W., Becker, A., Haasenritter, J., Viniol, A., & van der Wardt, V. (2025). Adherence support strategies for physical activity interventions in people with chronic musculoskeletal pain-A systematic review and meta-analysis. *Journal of Physical Activity & Health*, 22(1), 4–52. <https://doi.org/10.1123/jpah.2024-0099>

- Holth, H. S., Werpen, H. K. B., Zwart, J.-A., & Hagen, K. (2008). Physical inactivity is associated with chronic musculoskeletal complaints 11 years later: results from the Nord-Trøndelag Health Study. *BMC Musculoskeletal Disorders*, 9(1), 159. <https://doi.org/10.1186/1471-2474-9-159>
- Hussain, A., Haroon, H., Ahmed, A., & Gilani, S. A. (2022). Digital technologies in management of chronic pain- a systematic review. *Journal of the Pakistan Medical Association*, 72(6), 1158–1183. <https://doi.org/10.47391/jpma.3885>
- Jack, K., McLean, S. M., Moffett, J. K., & Gardiner, E. (2010). Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. *Manual Therapy*, 15(3), 220–228. <https://doi.org/10.1016/j.math.2009.12.004>
- Jakicic, J. M., Rogers, R. J., Davis, K. K., & Collins, K. A. (2018). Role of physical activity and exercise in treating patients with Overweight and obesity. *Clinical Chemistry*, 64(1), 99–107. <https://doi.org/10.1373/clinchem.2017.272443>
- Jansson, M., Ohtonen, P., Alalääkkölä, T., Heikkinen, J., Mäkinen, M., Lahtinen, S., Lahtela, R., Ahonen, M., Jämsä, S., & Liisanti, J. (2022). Artificial intelligence-enhanced care pathway planning and scheduling system: content validity assessment of required functionalities. *BMC Health Services Research*, 22, 1355. <https://doi.org/10.1186/s12913-022-08780-y>
- Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, 2(4), 230–243. <https://doi.org/10.1136/svn-2017-000101>
- Jordan, J. L., Holden, M. A., Mason, E. E., & Foster, N. E. (2010). Interventions to improve adherence to exercise for chronic musculoskeletal pain in adults. *Cochrane Database of Systematic Reviews 2010*, 1, CD005956. <https://doi.org/10.1002/14651858.CD005956.pub2>
- Laurijssen, S., Six-Dijkstra, M., van den Berg, G., Hanskamp, M., & Rijken, N. (2025). Guest editorial: Ethics for artificial intelligence: A game based- framework for physiotherapists. *Journal of Back and Musculoskeletal Rehabilitation*, 38(2), 205–207. <https://doi.org/10.1177/10538127251316945>
- Martinsen, E. W. (2008). Physical activity in the prevention and treatment of anxiety and depression. *Nordic Journal of Psychiatry*, 62(sup47), 25–29. <https://doi.org/10.1080/08039480802315640>
- Mazoué, J. G. (1990). Diagnosis without doctors. *The Journal of Medicine and Philosophy*, 15(6), 559–579. <https://doi.org/10.1093/jmp/15.6.559>
- medicalmotion GmbH (2025). *Über uns - medicalmotion*. <https://www.medicalmotion.com/de/uber-uns> (Tag des Zugriffs: 25.08.2025)
- Merolli, M., Francis, J. J., Vallance, P., Bennell, K. L., Malliaras, P., & Hinman, R. S. (2024). Evaluation of patient-facing mobile apps to support physiotherapy care: Systematic review. *JMIR MHealth and UHealth*, 12, e55003. <https://doi.org/10.2196/55003>

- Moudatsou, M., Stavropoulou, A., Philalithis, A., & Koukouli, S. (2020). The role of empathy in health and social care professionals. *Healthcare*, 8(1), 26. <https://doi.org/10.3390/healthcare8010026>
- Nowell, W. B., Gavigan, K., Kannowski, C. L., Cai, Z., Hunter, T., Venkatachalam, S., Birt, J., Workman, J., & Curtis, J. R. (2021). Which patient-reported outcomes do rheumatology patients find important to track digitally? A real-world longitudinal study in ArthritisPower. *Arthritis Research & Therapy*, 23, 53. <https://doi.org/10.1186/s13075-021-02430-0>
- Pengel, L. H., Herbert, R. D., Maher, C. G., & Refshauge, K. M. (2003). Acute low back pain: systematic review of its prognosis. *BMJ*, 327(7410), 323. <https://doi.org/10.1136/bmj.327.7410.323>
- Physio Deutschland (2018). Wartezeitenbarometer Physiotherapie: Sprunghafter Anstieg der Wartezeiten im Dezember 2018. Physio Deutschland. <https://www.physio-deutschland.de/fachkreise/news-bundesweit/einzelansicht/artikel/wartezeitenbarometer-physiotherapie-sprunghafter-anstieg-der-wartezeiten-im-dezember-2018-1.html#:~:text=Wartezeitenbarometer%20Physiotherapie%3A%20Sprunghafter%20Anstieg%20der,den%20Start%20ihrer%20Behandlungsserie%20warten.> (Tag des Zugriffs: 25.08.2025)
- Ramachandran, S., Chang, H.-J., Worthington, C., Kushniruk, A., Ibáñez-Carrasco, F., Davies, H., McKee, G., Brown, A., Gilbert, M., & Iyamu, I. (2024). Digital competencies and training approaches to enhance the capacity of practitioners to support the digital transformation of public health: Rapid review of current recommendations. *JMIR Public Health and Surveillance*, 10, e52798. <https://doi.org/10.2196/52798>
- Regan, E. A. (2022). Changing the research paradigm for digital transformation in healthcare delivery. *Frontiers in Digital Health*, 4, 911634. <https://doi.org/10.3389/fdgth.2022.911634>
- Rossetini, G., Camerone, E. M., Carlino, E., Benedetti, F., & Testa, M. (2020). Context matters: the psychoneurobiological determinants of placebo, nocebo and context-related effects in physiotherapy. *Archives of Physiotherapy*, 10(1), 11. <https://doi.org/10.1186/s40945-020-00082-y>
- Saenz, A. D., Harned, Z., Banerjee, O., Abramoff, M. D., & Rajpurkar, P. (2023). Autonomous AI systems in the face of liability, regulations and costs. *NPJ Digital Medicine*, 6(1), 185. <https://doi.org/10.1038/s41746-023-00929-1>
- Schlieter, H., Kählig, M., Hickmann, E., Fürstenau, D., Sunyaev, A., Richter, P., Breitschwerdt, R., Thielscher, C., Gersch, M., Maaß, W., Reuter-Oppermann, M., & Wiese, L. (2024). Digital health applications (DiGA) in the area of tension between progress and criticism. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 67(1), 107–114. <https://doi.org/10.1007/s00103-023-03804-2>
- Sedgwick, P., & Greenwood, N. (2015). Understanding the Hawthorne effect. *BMJ*, 351, h4672. <https://doi.org/10.1136/bmj.h4672>

- Sluka, K. A., O'Donnell, J. M., Danielson, J., & Rasmussen, L. A. (2013). Regular physical activity prevents development of chronic pain and activation of central neurons. *Journal of Applied Physiology*, *114*(6), 725–733. <https://doi.org/10.1152/jappphysiol.01317.2012>
- Statistisches Bundesamt (Destatis) (2022). Krankheitskosten: Deutschland, 2020, Krankheitsdiagnosen M00–M99 (Krankheiten des Muskel-Skelett-Systems), alle Geschlechter, alle Altersgruppen. <https://www-genesis.destatis.de/genesis/online?operation=table&code=23631-0003> (Tag des Zugriffs: 25.08.2025)
- Sulzer, L., Willer, E., Moog, S., & Ehrentraut, O. (2022). Testung digitaler Versorgungsangebote im Praxisalltag - Rückenschmerzen. Prognos AG (Hrsg) https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Z/ZDG/DiVA2_Testungsbericht.pdf (Tag des Zugriffs: 25.08.2025)
- Vincent, H., Zdziarski, L., & Wasser, J. (2015). Chronic pain management in the obese patient: a focused review of key challenges and potential exercise solutions. *Journal of Pain Research*, *8*, 63-77. <https://doi.org/10.2147/jpr.s55360>
- Walter, E., Wegner, S., & Burziwoda, D. (2022). EE378 cost-effectiveness analysis of a digital health application (DIGA) plus best supportive care (BSC) compared to BSC in German episodic migraine patients. *Value in Health*, *25*(12), S129. <https://doi.org/10.1016/j.jval.2022.09.624>
- Wang, G., Yang, M., Hong, M., Krauss, J., & Bailey, J. F. (2022). Clinical outcomes one year after a digital musculoskeletal (MSK) program: an observational, longitudinal study with nonparticipant comparison group. *BMC Musculoskeletal Disorders*, *23*(1), 237. <https://doi.org/10.1186/s12891-022-05188-x>
- Weise, H., Zenner, B., Schmiedchen, B., Benning, L., Bulitta, M., Schmitz, D., & Weise, K. (2022). The effect of an app-based home exercise program on self-reported pain intensity in unspecific and degenerative back pain: Pragmatic open-label randomized controlled trial. *Journal of Medical Internet Research*, *24*(10), e41899. <https://doi.org/10.2196/41899>
- World Confederation for Physical Therapy (WHO) und International Network of Physiotherapy Regulatory Authorities (INPTRA) (2020). Report of the WCPT/INPTRA Digital Physical Therapy Practice Task Force. <https://world.physio/sites/default/files/2020-06/WCPT-INPTRA-Digital-Physical-Therapy-Practice-Task-force-March2020.pdf> (Tag des Zugriffs: 25.08.2025)

5. ANHANG (PUBLIKATIONEN)

Publikation I:

Griefahn A, Zalpour C, Luedtke K (2024) Evaluation of the accuracy of an artificial intelligence in identifying contraindications to exercise therapy - Comparison with and interrater reliability of physical therapists judgments. *Health and Technology*, 14, 513–522. <https://doi.org/10.1007/s12553-024-00827-w>

Abstract

Purpose: The study validates a rule-based system for identifying contraindications to exercise therapy in a medical context. It evaluates accuracy and performance by comparing it with physical therapists' assessments and patients' characteristics.

Method: The dataset included 80 patient cases with clinical characteristics assessed by 20 physical therapists for contraindications to exercise therapy. Fleiss kappa and pooled kappa values measured agreement between physical therapists and AI. AI performance was assessed by sensitivity, specificity, accuracy and F1 score. Clinical characteristics were compared between therapists' votes using ANOVA and Bonferroni post-hoc test.

Results: The physical therapists had a mean age of 40.85 (8.23) years and a mean experience of 14.53 (8.20) years. Out of 64 patient cases, there was consensus on 35 cases with no contraindication and 29 cases with a consensus on "contraindication exists" for exercise therapy. In 16 cases there was no consensus between therapists. Overall, therapists had 87.5% agreement with Fleiss Kappa $\kappa = .43$. The pooled kappa value between therapists and AI was $\kappa_{\text{pooled}} = .63$. AI achieved perfect values (1) for sensitivity, specificity, accuracy and F1 score. Statistically, consensus-based comparisons by therapists revealed significant differences in pain intensity, duration, timing, and quality.

Conclusion: The study shows significant agreement between physical therapists and the AI, consistent with similar musculoskeletal studies. Various clinical characteristics highlight the importance of clinical reasoning and contraindication detection. In conclusion, advanced technologies such as decision support and expert systems could have a profound impact on clinical practice, improving accuracy, personalized exercises and telemedicine referrals for efficient care and improved patient decisions.

Trial registration: 30.12.2021 via OSF Registries, <https://doi.org/10.17605/OSF.IO/YCNJQ>.

Keywords Artificial intelligence . Accuracy . Contraindication . Exercise therapy .
Validation

Publikation II:

Griefahn A, Zalpour C & Luedtke K (2024) Identifying the risk of exercises, recommended by an artificial intelligence for patients with musculoskeletal disorders. *Scientific Reports*, **14**, 14472 (2024). <https://doi.org/10.1038/s41598-024-65016-1>.

Abstract

Musculoskeletal disorders (MSDs) impact people globally, cause occupational illness and reduce productivity. Exercise therapy is the gold standard treatment for MSDs and can be provided by physiotherapists and/or also via mobile apps. Apart from the obvious differences between physiotherapists and mobile apps regarding communication, empathy and physical touch, mobile apps potentially offer less personalized exercises. The use of artificial intelligence (AI) may overcome this issue by processing different pain parameters, comorbidities and patient-specific lifestyle factors and thereby enabling individually adapted exercise therapy. The aim of this study is to investigate the risks of AI-recommended strength, mobility and release exercises for people with MSDs, using physiotherapist risk assessment and retrospective consideration of patient feedback on risk and non-risk exercises. 80 patients with various MSDs received exercise recommendations from the AI-system. Physiotherapists rated exercises as risk or non-risk, based on patient information, e.g. pain intensity (NRS), pain quality, pain location, work type. The analysis of physiotherapists' agreement was based on the frequencies of mentioned risk, the percentage distribution and the Fleiss- or Cohens-Kappa. After completion of the exercises, the patients provided feedback for each exercise on an 11-point Likert scale, e.g. the feedback question for release exercises was "How did the stretch feel to you?" with the answer options ranging from "painful (0 points)" to "not noticeable (10 points)". The statistical analysis was carried out separately for the three types of exercises. For this, an independent t-test was performed. 20 physiotherapists assessed 80 patient examples, receiving a total of 944 exercises. In a three-way agreement of the physiotherapists, 0.08% of the exercises were judged as having a potential risk of increasing patients' pain. The evaluation showed 90.5% agreement that exercises had no risk. Exercises that were considered by physiotherapists to be potentially risky for patients also received lower feedback ratings from patients. For the 'release' exercise type, risk exercises received lower feedback, indicating that the patient felt more pain (risk: 4.65 (1.88), non-risk: 5.56 (1.88)). The study shows that AI can recommend almost risk-free exercises for patients with MSDs, which is an effective way to create individualized exercise plans without putting patients at risk for higher pain intensity or discomfort. In addition, the study shows significant agreement between physiotherapists in

the risk assessment of AI-recommended exercises and highlights the importance of considering individual patient perspectives for treatment planning. The extent to which other aspects of face-to-face physiotherapy, such as communication and education, provide additional benefits beyond the individualization of exercises compared to AI and app-based exercises should be further investigated.

Trial registration: 30.12.2021 via OSF Registries, <https://doi.org/10.17605/OSF.IO/YCNJQ>.

Keywords: Artificial intelligence, Risk analysis, Exercise, Musculoskeletal disorders

Publikation III:

Griefahn A, Avermann F, Zalpour C, Marshall RP, Cordon Morillas I, Luedtke K (2025) Exploring the effect of an 8-week AI-composed exercise program on pain intensity and well-being in patients with spinal pain: Retrospective cohort analysis. *JMIR Formative Research*, 9:e57826–e57826. <https://doi.org/10.2196/57826>

Abstract

Background: Spinal pain, one of the most common musculoskeletal disorders (MSDs), significantly impacts the quality of life due to chronic pain and disability. Physical activity has shown promise in managing spinal pain, although optimizing adherence to exercise remains a challenge. The digital development of artificial intelligence (AI)-driven applications offers a possibility for guiding and supporting patients with MSDs in their daily lives.

Objective: The trial aimed to investigate the effect of an 8-week AI-composed exercise program on pain intensity and well-being in patients with spinal pain. It also examined the relationship between exercise frequency, pain intensity, and well-being. In addition, app usage frequency was examined as a proxy for app engagement.

Methods: Data from users who met the inclusion criteria were collected retrospectively from the medicalmotion app between January 1, 2020, and June 30, 2023. The intervention involved the use of the medicalmotion app, which provides 3–5 personalized exercises for each session based on individual user data. The primary outcomes assessed pain intensity and well-being using the numeric rating scale (NRS) and the Likert scale. Data were collected at baseline (t0), 4 weeks (t1), and 8 weeks (t2). The correlation between exercise frequency, pain intensity, and well-being was analyzed as a secondary outcome. In addition, average session length and frequency were measured to determine app engagement. Statistical analysis included ANOVA and Spearman correlation analysis.

Results: The study included 379 participants with a mean age of 50.96 (SD 12.22) years. At t2, there was a significant reduction of 1.78 points on the NRS ($P < .001$). The score on the Likert scale for well-being improved by 3.11 points after 8 weeks. Pain intensity showed a negative correlation with the number of daily exercises performed at t1 and t2. Well-being had a small negative correlation with the average number of exercises performed per day. The average number of exercises performed per day was 3.58. The average session length

was approximately 10 minutes, and the average interaction with the app was 49.2% (n=27.6 days) of the 56 available days.

Conclusions: Overall, the study demonstrates that an app-based intervention program can substantially reduce pain intensity and increase well-being in patients with spinal pain. This retrospective study showed that an app that digitizes multidisciplinary rehabilitation for the self-management of spinal pain significantly reduced user-reported pain intensity in a preselected population of app users.

Trial registration: OSF Registries, <https://osf.io/kjhcf>

Keywords: exercise, mHealth, app engagement, spinal pain, artificial intelligence, AI, intensity, well-being, mobile health, apps, applications, retrospective analysis, physical activity, adults, questionnaire

6. DANKSAGUNG

In erster Linie danke ich meiner Doktormutter Prof. Dr. Kerstin Lüdtker für die zuverlässige und engagierte Betreuung während der gesamten Promotionszeit. Ihre fachliche Expertise, ihr konstruktives Feedback und ihre Geduld haben mein Vorhaben entscheidend unterstützt und mir immer wieder neue Impulse gegeben. Ebenso danke ich meinem Ko-Betreuer Prof. Dr. med. Christoff Zalpour, der das Projekt über die gesamte Zeit aufmerksam begleitet und mich mit Interesse und hilfreichen Anregungen unterstützt hat.

Kerstin und Christoff standen mir nicht nur in fachlichen Fragen, sondern auch durch ihre offene und ermutigende Art zur Seite. Dafür, dass sie jederzeit ein offenes Ohr hatten und meine Arbeit mit großem Einsatz begleitet haben, bin ich sehr dankbar.

Mein besonderer Dank gilt meinem Lebensgefährten Sascha. Mit seiner Geduld, Zuversicht und beständigem Rückhalt hat er mich durch die Höhen und Tiefen dieser Zeit getragen. Unser Sohn Levi hat mir zudem eine wertvolle neue Perspektive eröffnet und diese Zeit mit Freude und Gelassenheit bereichert. Auch meinen Eltern und meiner Familie danke ich herzlich für ihre Unterstützung und ihr Verständnis, die mir den notwendigen Freiraum für diese Arbeit ermöglicht haben.

Ein weiterer Dank gilt CK, mit der ich diese Promotionsreise teilen durfte. Ihr positiver Blick, die gemeinsamen Diskussionen und ihr stetiger Zuspruch haben vieles leichter und schöner gemacht.

Zur Transparenz wird festgehalten, dass für Teile des Manteltextes das Sprachmodell ChatGPT (Version GPT-5, OpenAI) ausschließlich zur sprachlichen Glättung und Optimierung von Formulierungen eingesetzt wurde. Inhaltliche Entscheidungen, wissenschaftliche Bewertungen, Analysen und Interpretationen wurden vollständig eigenständig durch die Autorin vorgenommen.