
Auf dem Weg zu interpretierbaren, KI-basierten,
explorativen und
entscheidungsunterstützenden Systemen
im Kontext biomechanischer Daten

Vom Fachbereich Sozialwissenschaften
der Technischen Universität Kaiserslautern
zur Verleihung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
genehmigte

Dissertation

vorgelegt von
Carlo Dindorf

Tag der Disputation: 07.06.2022
Dekan: Prof. Dr. Michael Fröhlich
Vorsitzende: Prof. Dr. Schiefner-Rohs
1. Gutachter: Prof. Dr. Michael Fröhlich
2. Gutachterin: Prof. Dr. Gabriele Bleser

Dekan: Prof. Dr. Michael Fröhlich

Kommissionsvorsitz: Prof. Dr. Schiefner-Rohs

1. Gutachter: Prof. Dr. Michael Fröhlich

2. Gutachterin: Prof. Dr. Gabriele Bleser

1. Diskutant: Prof. Dr. Allen

2. Diskutant: Prof. Dr. Lachmann

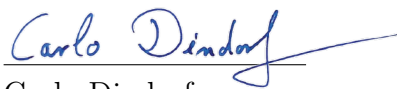
Tag der Einreichung: 30.03.2022

Tag der Disputation: 07.06.2022

Eidesstattliche Erklärung

Ich versichere, dass ich die vorliegende kumulative Dissertationsschrift selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen.

Kaiserslautern, den 30.03.2022



Carlo Dindorf

Gewidmet meinen Eltern
Ellen und Ernst Dindorf

Danksagung

DAS Promotionsvorhaben stellt für mich einen wichtigen und ereignisreichen Lebensabschnitt dar. Rückblickend auf die vergangenen Jahre ist es an der Zeit mich bei denjenigen zu bedanken, die mich in dieser herausfordernden Lebensphase, aber auch ungemein lohnendem Abschnitt meiner akademischen Laufbahn, unterstützt und begleitet haben. An dieser Stelle möchte ich meinen herauszustellenden Dank aussprechen.

Mein besonderer Dank gilt zunächst meinem Doktorvater, Univ.-Prof. Dr. Michael Fröhlich, für den Zeitaufwand der Betreuung dieser Arbeit, der freundlichen Hilfe und der mannigfachen Ideengebung, die mir einen kritischen Zugang zu dieser Thematik eröffnete. Vielen Dank, für die Möglichkeiten meinen Interessen zu folgen, der Chance zur Verwirklichung meines Vorhabens, die vielen Anregungen und konstruktive Kritik und das mir entgegengebrachte Vertrauen sowie die außergewöhnlichen Rahmenbedingungen und Freiheitsgrade. Die zahlreichen Gespräche auf intellektueller und persönlicher Ebene werden mir immer als bereichernder und konstruktiver Austausch in Erinnerung bleiben. Ich habe unsere Dialoge stets als Ermutigung und Motivation empfunden. Weiterhin gebührt ein großer Dank Prof. Dr. Gabriele Bleser und Dr. Bertram Taetz, die mich von Anfang an unterstützt haben. Vielen Dank für die tollen Ideen und Inspirationen zur Umsetzung der gesamten Arbeit.

Ein großes Dankeschön geht zudem an das ganze (teilweise ehemalige) Kollegium der Sportwissenschaft. Stellvertretend möchte ich namentlich besonders Dr. Joshua Berger, Dr. Marco Backfisch, Dr. Stephan Becker, Dr. Oliver Ludwig, Dr. Anna Thomas und Eva Bartaguiz an dieser Stelle erwähnen. Vielen Dank für die tolle Zusammenarbeit, die vielen sowohl beruflichen als auch privaten Gespräche, die herausragende Unterstützung sowie insgesamt sehr schöne Zeit in den letzten drei Jahren.

In der Schnittmenge zwischen beruflich und privat gilt mein ganz großer Dank dem Team rund um das MotionLab der Universitätsmedizin der Johannes Gutenberg-Universität Mainz. Namentlich möchte ich mich hier noch einmal besonders bei Dr. Jürgen Konradi und Claudia Wolf für die sehr enge Zusammenarbeit und schöne Zeit bedanken. Zudem möchte ich mich im Rahmen dieser Schnittmenge bei dem Unisport der Technischen Universität Kaiserslautern bedanken. Danke für die Zeit, die vielen Eindrücke und Lektionen, die ich durch die Kongressorganisation mitnehmen konnte. Danke Dir, Dr. Max Sprenger, u.a. für die Exkursionen sowie vielen privaten und hochschulpolitischen Gespräche.

Privat möchte ich einen großen Dank an meine Familie und meine Partnerin Hannah richten. Danke, an meine Eltern Ellen und Ernst, ohne euch wäre mein ganzer Bildungsweg so nicht möglich gewesen. Vielen Dank, für eure Geduld, die Ermutigungen und Zusprüche und dass ihr insbesondere während des Promotionsvorhabens alles getan habt, um mir möglichst viel Zeit für die Arbeit an dieser Dissertation und dem Sport freizuhalten.

Danke Ihnen / Euch allen!

Inhaltsverzeichnis

1	Einleitung	12
1.1	Motivation	12
1.2	Forschungsfragen	16
1.3	Publikationen	16
1.4	Besonderer Forschungsbeitrag	19
1.5	Aufbau	19
2	Forschungsdesiderat und Hintergrund	20
2.1	Grundlagen Künstliche Intelligenz (KI), Maschinelles Lernen (ML) und Deep Learning	20
2.1.1	Begriffsklärung und aktuelle Relevanz	20
2.1.2	Arten des Lernens und Einsatz von Maschinellem Lernen (ML) im Rahmen von Knowledge Discovery in Databases (KDD)	23
2.1.3	Relevante Algorithmen	24
2.1.4	Explainable Artificial Intelligence (XAI)	26
2.2	E-Health die Anwendung von Künstlicher Intelligenz (KI) in Medizin und Biomechanik	27
2.2.1	E-Health und der Einsatz von KI-Anwendungen in der menschlichen Lebensspanne	27
2.2.2	Exemplarischer Einsatz von KI bei klinischen Daten	28
2.2.3	Exemplarischer Einsatz von KI bei biomechanischen Daten	29
2.3	Vertiefende Messmethodische Grundlagen mit Fokus auf die Wirbelsäule	32
2.3.1	Klinische und sozioökonomische Bedeutung	32
2.3.2	Wirbelsäulenanalyse mit Fokus auf das DIERS Messsystem	34
2.3.3	Quantitative Güte des Messsystems	36
3	Die Publikationen	37
3.1	1. Publikation	38
3.2	2. Publikation	50
3.3	3. Publikation	81
3.4	4. Publikation	97
3.5	Vorbereitende Publikation für Publikation 5	131
3.6	5. Publikation	133
4	Resümee der Befunde	153
4.1	Ausnutzung vorhandener Daten und Feature Extraktion	157
4.2	Feature Extraktion und Stabilität	161
4.3	Einflussfaktoren Interpretierbarkeit	162
4.4	Längs- und Querschnittvergleiche via Siamese Neural Networks (SNN)	168
4.5	Pathologieunabhängige Klassifikation und Interpretation	171
4.6	Weitere methodische Limitationen und kritische Anmerkungen	174
4.7	Zukünftige Forschungsrichtungen und Aufgaben	178

4.8	Ausblick - die Zukunft von Maschinellem Lernen (ML) bei der Analyse biomechanischer Daten	181
5	Fazit	184
6	Literaturverzeichnis	186
7	Anhang	208
8	Curriculum Vitae	210

Abbildungsverzeichnis

2.1	Zusammenhänge Data Mining, Künstliche Intelligenz (KI), Maschinelles Lernen (ML), Deep Learning sowie weitere Bereiche. Eigene Darstellung in Anlehnung an [69].	21
2.2	Abgrenzung Maschinelles Lernen (ML) von klassischem Programmieren. Eigene Darstellung.	22
2.3	Jährliche Anzahl an Publikationen zu dem Suchbegriff „machine learning“. Stand: 22.02.2022. Suche mit https://www.dimensions.ai/ . Eigene Darstellung.	22
2.4	Einsatz von KI in der Lebensspanne des Menschen in alltäglichen, klinischen und sportlichen Bereichen. Eigene Darstellung. Die klinischen Aspekte sind angelehnt an [9] und wurden um die biomechanische Perspektive erweitert.	28
2.5	Datenquellen in der Biomechanik. Für ein holistisches Verständnis muss die Komplexität des Menschen beachtet werden und multimodale Daten berücksichtigt werden. Eigene Darstellung.	31
2.6	DIERS 4D motion® Lab eingesetzt im MotionLab der Universitätsmedizin der Johannes Gutenberg-Universität Mainz. Abbildung basiert auf den freundlicherweise zur Verfügung gestellten Visualisierungsmedien der Firma DIERS International GmbH.	35
4.1	Mittelwert und Standardabweichung der einzelnen Wirbelkörperpositionen sowie des Beckens bei Gesunden (blaue Linie) sowie Einzelfallbetrachtung einer gesunden Person (schwarze Linie). Die Daten basieren auf Publikation 5. Eigene Darstellung.	165
4.2	SHAP-Abhängigkeitsdiagramm (engl. SHAP dependence plot) für das global relevanteste Feature (T3 Flexion Extension) und dessen Interaktion mit einem automatisch ausgewählten Feature (T10 Flexion Extension). Das zugrunde liegende Modell klassifiziert in gesund und pathologisch aufgrund einer Wirbelsäulenversteifung. Niedrige Merkmalsausprägungen des Features T3 Flexion Extension indizieren entsprechend der SHAP Werte eine Zugehörigkeit zur Klasse der pathologischen Subjekte. Umgekehrt indizieren hohe Merkmalsausprägungen eine Zugehörigkeit zur Klasse der Gesunden. Interaktionseffekte der beiden Feature erscheinen für das angeführte Beispiel gering bis nicht vorhanden. Bei der Interpretation von möglichen Interaktionen darf nicht außer Acht gelassen werden, dass es sich möglicherweise nur um Korrelationen ohne Kausalität handelt.	167

4.3	Trainingsdaten und 15 der Best-Fit-Regressionenlinien mit ihren jeweiligen Konfidenzintervallen (CI). Das Diagramm stellt die Schätzung der aleatorischen (Konfidenzintervallen) und epistemischen (Best-Fit-Regressionenlinien) Unsicherheit dar im Rahmen der Prediktion der Position von T4 in Frontalebene auf Basis der Rotation von T5 in Transversalebene. Als Trainingsdaten wurde der Datensatz aus Publikation 4, unter alleiniger Verwendung der statischen Wirbelsäuledaten, verwendet. Die Modellierung erfolgte mittels TensorFlow Probability [231]. Die standardisierten Werte sind graphisch aufgetragen. Eigene Darstellung angelehnt an die Ausführungen von [233].	177
4.4	Visualisierung des Doctor in the Loop Paradigmas. Eigene Darstellung angelehnt und erweitert auf Basis von [268]	183

Tabellenverzeichnis

1.1	Übersicht der zentralen Inhalte der Publikationen	17
2.1	Übersicht über verwendete ML Algorithmen. Die Art bezieht sich auf die Art der Verwendung in den Publikationen. SVM = Support Vector Machine; RF = Random Forest; KNN = Künstliche neuronale Netze; LR = Logistische Regression; k-NN = k-Nearest-Neighbours Algorithmus; PCA = Principal Component Analysis; LDA = Linear Discriminant Analysis; OCSVM = One Class Support Vector Machine; IF = Isolation Forest; Klf = Klassifikation; Dimr = Dimensionsreduktion; Anod = Anomaliedetektion	25
4.1	Zusammenfassung der Publikationen hinsichtlich des Ausmaßes der Adressierung zentraler Aspekte der Forschungsfragen.	156

Abkürzungs- und Begriffsverzeichnis

AutoML	Automatisiertes maschinelles Lernen
DeepLIFT	Deep Learning Important FeaTures
engl.	englisch
IMU	Inertiale Messeinheit (engl. Inertial Measurement Unit)
JMIM	Joint Mutual Information Maximization Algorithmus
KDD	Wissensentdeckung in Datenbanken (engl. Knowledge Discovery in Databases)
KI	Künstliche Intelligenz (engl. Artificial Intelligence)
KNN	künstliche neuronale Netzwerke (engl. Artificial Neural Network)
LIME	Local Interpretable Model-Agnostic Explanations
LRP	Layerwise Relevance Propagation
MCC	Matthews Korrelationskoeffizient
ML	Maschinelles Lernen
mRMR	Minimum Redundancy Maximum Relevance
OCSVM	One Class Support Vector Machine
RF	Random Forest
ROM	Bewegungsumfang (engl. range of motion)
SG	Smart Gait
SHAP	SHapley Additive exPlanations
SNN	Siamese Neural Networks
SpRAy	Spectral Relevance Analysis
TEP	Totalendoprothetik
t-SNE	t-distributed Stochastic Neighbor Embedding
XAI	Explainable Artificial Intelligence

Die Arbeit übernimmt an einzelnen Stellen englischsprachige Termini, da teilweise kein deutschsprachiges Pendant verfügbar ist oder die englischsprachigen Begriffe in der disziplinspezifischen Sprache geläufiger sind. Nachfolgend sind die wichtigsten Begriffe zu finden. Nähere Erläuterungen zu einzelnen Aspekten sind im Text zu finden.

Continual Learning	effiziente Adaption eines bestehenden Klassifikators mit neuen Daten, ohne dass bei sequentieller Präsentation frühere Infos vergessen werden
Deep Learning	Methode des maschinellen Lernens, die künstliche neuronale Netze mit mehreren Zwischenschichten einsetzt
Data Mining	Suche nach bedeutungsvollen, nicht-trivialen Informationen in (zumeist großen) Datenmengen
Data Science	Lehre von der Extraktion von verallgemeinerbarem Wissen aus Daten
Feature	unabhängige Variable bzw. Merkmal
Feature Engineering	Merkmalskonstruktion
Feature Extraktion	Extraktion von Merkmalen
Feature Learning	auch engl. representation learning genannt; Technik der automatischen Entdeckung von aufgabenrelevanten Datenrepräsentationen
Feature Selektion	Auswahl von Merkmalen
Ensemble Feature Selektion	Kombinieren der Ausgaben mehrerer Modelle zur Auswahl von Merkmalen
Knowledge Discovery	Wissensentdeckung
Metric Learning	Erlernen einer Repräsentationsfunktion, die Objekte in einem latenten Raum abbildet
One Shot Learning	One Shot Learning zielt darauf ab Objekte auf Basis von einem bzw. wenigen Fällen zu klassifizieren
Transfer Learning	Anwendung von Wissen auf Basis der Lösungen eines bestimmten Problems auf ein anderes verwandtes Problem

1.1 Motivation

SOWOHL Künstliche Intelligenz (KI)^{1,2} als auch das Teilgebiet des Maschinellen Lernens (ML) haben historisch betrachtet einen Boom erfahren. In vielen Lebenslagen hat sich KI bereits etabliert und ist mittlerweile nicht mehr wegzudenken. Bei der Gesichtserkennung auf dem Smartphone [1] sowie bei der Sprachverarbeitung, -synthese und -übersetzung (z.B. Google Translate, Alexa, Google Assistant [2]) hat ML eine große Erfolgsgeschichte erlebt und ist längst Routine im Alltag geworden. ML wird in zahlreichen weiteren Bereichen, wie der Objekterkennung [3], dem autonomen Fahren [4] oder bei Spamfiltern [5] erfolgreich eingesetzt. Auch in den Sektoren Sport [6], [7] und Gesundheit [8], [9] hat KI zunehmend an Interesse gewonnen. Im Jahr 2016 haben KI-Projekte im Gesundheitswesen sogar mehr Investitionen angezogen, als KI-Projekte in anderen Sektoren der Weltwirtschaft [10].

KI basierte Anwendungen können dabei bestimmte Aufgaben ähnlich effektiv bzw. effektiver und effizienter als Menschen lösen. Beispielsweise können annähernd menschenähnliche Leistungen bei der Erkennung handgeschriebener Ziffern bzw. eine Übertreffung des Menschen bei Verkehrszeichenerkennung [11] gezeigt werden. Bei der Klassifikation von Hautkrebs konnten mittels KI Ansätze vergleichbare Kompetenzen wie die von erfahrenen Dermatologen erzielt werden [12].

Der KI Boom und die fortschreitende Entwicklung der Ansätze wurde bzw. wird durch verschiedene Aspekte bedingt. Hierzu zählt die verfügbare Datenlage (Qualität und Quantität), welche entscheidend den Erfolg von KI Modellen beeinflusst [13]. Früher lag oftmals eine begrenzte Datenbasis vor, heute hingegen werden durch moderne Technologien im Vergleich zu früher Unmengen an großen und komplexen Datenbeständen generiert [14]. Vom ursprünglichen Problem der Verfügbarkeit der Daten geht es über in ein Problem der Bewältigung der Datenmassen [15].

Parallel zu dem Anstieg der verfügbaren Datenmenge haben immense Entwicklungen im Bereich der Modelle stattgefunden, die u.a. dazu beigetragen haben, dass Modelle immer genauer werden [16]. Durch benutzerfreundliche Open-Source-Programmierbibliotheken wie scikit-learn [17] oder Pytorch [18] wird zudem die Implementierung von ML Modellen immer einfacher und zugänglicher für eine breite Masse von Menschen. Weiterhin kommt Hardwareentwicklungen und der

Aspekte

¹ Gender-Hinweis: Aus Gründen der besseren Lesbarkeit wird teilweise auf die gleichzeitige Verwendung der Sprachformen männlich, weiblich und divers (m/w/d) verzichtet. Sämtliche Personenbezeichnungen in dem Dokument gelten gleichermaßen für alle Geschlechter.

² Für einzelne Termini werden aus dem englischen Sprachgebrauch geprägte Begrifflichkeiten verwendet, da kein geläufiges deutschsprachiges Pendant vorhanden ist.

damit verbundene Anstieg der verfügbaren Rechenleistung und der einhergehenden Beschleunigung von Berechnungen eine entscheidende Rolle zu [19].

Diese Treiber haben dafür gesorgt, dass KI viele neue Möglichkeiten erschlossen und auch in bewegungswissenschaftlichen und klinischen Kontexten nach und nach Einzug gefunden hat. Neuerdings wurde so der Begriff Smart Gait (SG) geprägt, der für integrierte Analysesysteme steht, welche KI bei menschlichen Gangdaten verwenden [20]. Bewegung und Haltung sind dabei komplexe, multivariate Phänomene. Moderne Technologien ermöglichen es diese durch Erhebung großer Datenmengen abzubilden bzw. ein umfassendes Gesundheitsbild der Personen/Patienten zu erstellen. Es lässt sich ein Anstieg der verfügbaren Datenmenge bis hin zu einer Big Data Domäne im Bereich der Bewegungswissenschaft beobachten [21], [22].

Dabei überschreitet die Datenlage zunehmend die Menge an Einzelinformationen, die ein Mensch für eine zusammenfassende Betrachtung gleichzeitig interpretieren kann. Zudem wird das Verständnis dieser (teilweise temporalen) Daten durch Hochdimensionalität und Multikollinearität oftmals zusätzlich erschwert [23]–[25]. Die eigentlichen Potentiale, welche neuartige Messverfahren bieten, können somit oftmals nicht vollständig ausgeschöpft werden. Es bestehen also einerseits Potentiale, neues Wissen zu generieren und positiv auf die Praxis einzuwirken, andererseits wird durch diese Entwicklung die Datenlage zunehmend komplexer. Bewegungswissenschaftler, Therapeuten und Ärzte stehen vor der zentralen Herausforderung, die Daten adäquat zu analysieren und deren Potentiale bestmöglich zu nutzen.

Klassische inferenzstatische Methoden sind bei derartigen Datenbeständen oft limitiert und ML zeigt sich vorteilhaft [21], [22], [26]. Viele inferenzstatistische Methoden erfordern zudem eine Reihe starker Annahmen (z.B. keine Multikollinearität bei der logistischen Regression), wohingegen ML oft ohne viele Annahmen der zugrundeliegenden Daten verwendet wird [27]. Klassische inferenzstatistische Werkzeuge werden zwar häufig im Bereich der Ganganalyse angewendet, allerdings fehlt ihnen oft die Vorhersagekraft, um auf nicht gesehene Daten zu verallgemeinern [28]. Hingegen ermöglicht ML nichtlinear Approximation auch bei komplexer, hochdimensionaler Datenlage [22].

Verschiedenste Arbeiten konnten Potentiale von KI gestützten Analysen im Gesundheitssektor aufzeigen. Zahlreiche Arbeiten verwenden dabei (klinische) Bilddaten als Ausgangspunkt für die Anwendung von KI-Methoden und nutzen diese zur Interpretation der Daten, z.B. in Kontext der Lokalisation von Glaukomen [29] oder bei der Erkennung von Hautläsionen [12]. So konnten Potentiale aufgezeigt werden, objektive, datenbasierte Einschätzungen (von Krankheitsbildern) zu generieren [30], [31], Entdeckungsraten von Krankheitsbildern zu erhöhen und Fehldiagnosen zu reduzieren [32], die Früherkennung von Auffälligkeiten zu verbessern [33], Kompetenzunterschiede abzuschwächen [34], sowie insgesamt durch den Einsatz computergestützter Methoden Prozesse zu beschleunigen und zu automatisieren (für genauere Ausführungen siehe Abschnitt 2.2.2).

Zusammenfassend gewinnen datenbasierte und KI-unterstützte Verfahren damit als Unterstützung bzw. zur Festlegung personalisierter Therapiepfade immer mehr an Bedeutung um Ärzte und Therapeuten dabei zu unterstützen, medizinische Abläufe effizienter und patientenbezogener zu gestalten, in dem die Datenflut, die mit jedem Patienten einhergeht, strukturiert und möglichst optimal analysiert wird. Zudem können diese dazu beitragen Gesundheitskosten zu senken, da beispielsweise frühzeitige Diagnosen Mobilitätsverluste verhindern und damit Folgekosten vermie-

*komplexe
Datenlage*

Potentiale

den werden können [35]. Neben positiven Effekten für das Patientenwohl besteht damit insgesamt ein großes gesellschaftliches und wirtschaftliches Innovationspotential. Hinzu kommen Chancen im Bereich der Exploration und Identifikation von potentiell bedeutsamen (multivariaten) Korrelaten und damit zur Hypothesenbildung im Wissenschaftsparadigma.

Im Bereich der klinischen Ganganalyse konnten ML und KI zur Entscheidungsunterstützung bei verschiedensten Krankheitsbildern eingesetzt werden (z.B. Schlaganfall [36], Parkinson [37], Osteoarthritis [38]). Trotz der hohen Potentiale gehen jedoch bestehende Arbeiten bei bewegungswissenschaftlichen Daten nur selten über eine reine Klassifikation hinaus [39]. Interpretationen werden nur bei wenigen Arbeiten angedeutet und der entscheidende Interpretationsgrad wie bei klinischen Bilddaten wird nicht erreicht. Ausgangspunkt für die Arbeit sollen in diesem Zusammenhang fünf Defizite bilden, welche als zentrale Limitationen und Herausforderungen charakterisiert werden können und den Einsatz von KI als exploratives bzw. unterscheidungsunterstützendes System stark einschränken. Vor praktischer Transferierung der Methoden besteht die Notwendigkeit die Aspekte zu evaluieren und möglichst optimal und kontextspezifisch zu adressieren.³

Bewegungswissenschaftliche Daten liegen oft als multivariate Zeitreihen vor. Außerdem bestehen die Datenbestände oft aus mehreren in Beziehung stehenden Datentabellen (relationale Datenbestände), da verschiedenste Parameter und Modalitäten zur Erfassung der Komplexität des Menschen herangezogen werden müssen. In der Praxis werden häufig Gelenkwinkel zu bestimmten Zeitpunkten und einfache deskriptive Statistiken wie der Bewegungsumfang (ROM) oder Maximalwerte für Klassifikationsaufgaben herangezogen [40], [41]. Es ist jedoch fraglich, ob es durch diese simple Art der Datennutzung zu einem a priori Informationsverlust kommt und die Potentiale der Daten daher nicht optimal genutzt werden. Es gilt Methoden zu entwickeln, um unabhängig⁴ von Vorwissen, Annahmen oder subjektiven Entscheidungen die Potentiale der vorliegenden Daten ausschöpfen zu können. **Defizit 1**

Die Stabilität bzw. Robustheit, also das Ausmaß, in dem Variationen der Trainingsdaten zu unterschiedlichen Merkmalspräferenzen führen [42], ist bei der Bestimmung von bedeutsamen Parametern aus Datenbeständen (Feature Selektion) im Rahmen der Knowledge Discovery von besonders hoher Relevanz [43]. Stabilität im Rahmen der Feature Selektion wird bei einer Vielzahl an Studien im biomechanischen Kontext allerdings nicht beachtet (z.B. [44], [45]). Geringe Stabilität beeinträchtigt jedoch negativ die Reproduzierbarkeit und damit die Vertrauenswürdigkeit der Ergebnisse [46]. Es gilt daher Pipelines zu entwickeln und zu evaluieren, welche die Bestimmung robuster Parameter ermöglichen. **Defizit 2**

Gangklassifizierung stellt ein wichtiges Tool und Ausgangspunkt für verschiedene Fragestellungen dar [47], [48]. Mit komplexen Modellen (z.B. künstliche neuronale Netzwerke, KNN) können im Vergleich zu einfachen Modellen (z.B. eine lineare Regression) oft höhere Genauigkeiten erzielt werden. Auf der anderen Seite erschwert **Defizit 3**

³ Die Herleitung der einzelnen Forschungsdefizite kann hier nur angerissen werden. Es sei im Detail auf die einzelnen Publikationen und weiteren Darstellungen dieser kumulativen Dissertationsschrift verwiesen.

⁴ Unabhängig soll an dieser Stelle nicht als Verwerfung bestehenden Vorwissens missgedeutet werden. Stattdessen ist eine Erweiterung bestehenden Vorwissens bzw. Integration dessen anvisiert. Siehe hierzu Sektion 4.1 und 4.3.

die Komplexität der Modelle die Interpretierbarkeit und Transparenz (Trade-off zwischen Genauigkeit des Modells und der Interpretierbarkeit [49]), weshalb den komplexen Modellen ein Black Box Charakter zugeschrieben wird [50]. Da intransparent ist, warum bestimmte Entscheidungen vom Modell getroffen wurden und was das Modell wirklich gelernt hat, ist es schwer derartigen Modellentscheidungen zu vertrauen [51]. Diese Opazität entspricht nicht den Anforderungen der europäischen Datenschutzgrundverordnung (GDPR, EU 2016/679) [52] und schränkt praktische Anwendungen, beispielsweise im klinischen Kontext, stark ein [53]. Auch [20] konstatiert den Black Box Charakter als zentrale Herausforderung, welche durch eine Vielzahl von Arbeiten, welche KI im Kontext biomechanischer Gangdaten einsetzen, nicht überwunden wird. Der Einfluss unterschiedlicher Datenrepräsentationen auf sowohl die Klassifikationsgenauigkeit als auch Interpretierbarkeit ist dabei nach aktuellem Forschungsstand weitgehend unklar.

Die vorhandene Datenlage spielt oft eine große Rolle für den Erfolg der Modelle. Komplexe Modelle benötigen oft größere Datenmengen im Vergleich zu einfachen Modellen für das Modelltraining [13]. Konträr zu bestimmten Aufgaben aus dem Bereich der Bildklassifikation, bei denen oft immense Datenbestände vorliegen (z.B. ImageNet Datensatz [54]), ist insbesondere in der Bewegungs- und Trainingswissenschaft die Datenlage oft limitiert. Testpersonen, wie beispielsweise Patienten bestimmter Krankheitsbilder sind oft begrenzt verfügbar oder Wettkampfleistungen sowie Zeitreihen über Fitnesszustände liegen oft nur über wenige Wochen vor. Im Kontext von Personenvergleichen bzw. Identifikationsaufgaben, welche u.a. hohe Relevanz zum Monitoring von Rehabilitationsfortschritten oder pathologischen Veränderungen hohe Relevanz erfahren, stellt zudem die intra-individuelle Varianz der Personen eine Herausforderung dar, welche die Modellgüte negativ beeinflussen kann [55]. *Defizit 4*

Bisherige Arbeiten basieren zumeist auf Klassifikatoren, welche auf Basis von Gesunden und Personen einer bestimmten Pathologie trainiert wurden (z.B. [56]–[58]). Die Modelle sind daher nur auf eine bestimmte Pathologie anwendbar - sie sind pathologieabhängig. Um ein Modell für eine andere interessierende Pathologie zu erhalten, muss ein komplett neues Training mit Probanden der jeweiligen Pathologie durchgeführt werden, auch wenn die Trainingsdaten gesunder Probanden gleich bleiben. Es ist unklar, inwieweit ein pathologieunabhängiger Klassifikator realisierbar ist, welcher zudem in Anlehnung an Defizit 3 durch Offenlegung von Entscheidungsgrundlagen interpretierbar ist. *Defizit 5*

Zusammenfassend ist die praktische Transferierung vieler bestehender, in anderen Settings erfolgreich eingesetzter, Methoden in den Bereich der biomechanischen Datenanalyse daher in weiten Teilen noch ausstehend. Die Forschungslage kann subsumiert in weitreichenden Teilen als defizitär angesehen werden. Was bislang gemacht wurde ist, aufgrund der bestehenden Defizite, oftmals wenig praxisrelevant. Bestehende Methoden und Anwendungen gilt es praxisnah anzupassen bzw. weiterzuentwickeln, um die Verfahren in das praktische Geschehen effektiv zu implementieren und Outcome-orientiert zu bewerten.

1.2 Forschungsfragen

Aufbauend auf den genannten Defiziten setzt sich die Arbeit mit den nachfolgenden übergeordneten Forschungsfragen auseinander. Die Fragestellungen stellen dabei einen Auszug der wichtigsten Inhalte dar, um einen Gesamtüberblick zu geben. Sie verfolgen dabei methodische und beispielhafte explorative (Knowledge Discovery) Ziele. Der inhaltlich-methodische Zusammenhang wird durch die methodischen Forschungsfragen und Entwicklungen gebildet. Die konkreten explorativen Forschungsfragen werden zur Bearbeitung der methodischen Fragestellungen herangezogen und sind zur Übersicht in Tabelle 1.1 mit angeführt. Einzelne Publikationen bearbeiten die hier dargestellten übergeordneten Fragen zum Teil in spezifischeren Teilfragen. Hierzu sei auf Tabelle 4.1 verwiesen.

1. Wie können möglichst ohne Verlust von potentiell bedeutsamen Informationen bestehende Daten optimal genutzt und besonders relevante Parameter ohne Annahmen, subjektive Auswahl und Vorkenntnisse objektiv bestimmt werden?
2. Eignen sich Ensemble Feature Selektion Methoden im Kontext biomechanischer Daten, um die Robustheit bzw. Stabilität und damit die Vertrauenswürdigkeit bei Bestimmung besonders relevanter Parameter positiv zu beeinflussen?
3. Inwieweit erweisen sich Methoden der Explainable Artificial Intelligence (XAI) zur Interpretation biomechanischer Daten als sinnvoll?
4. Eignen sich Siamese Neural Networks (SNN) um bei geringer Datenlage, Datenrepräsentationen automatisch zu erzeugen, welche Personenvergleiche und -identifikationen ermöglichen und intra-individuelle Varianzen minimieren?
5. Eignet sich eine One Class Support Vector Machine (OCSVM) im Vergleich zu einem klassischen binären Klassifikator als pathologieunabhängiger, interpretierbarer Klassifikator?

Im Rahmen dieser Arbeit erfolgt die Fokussierung auf drei Datenquellen (markerbasierte 3D- sowie Inertial Measurement Units (IMU) basierte Bewegungsdaten des unteren Körpers; rasterstereographisch bestimmte statische und dynamische Daten der Wirbelsäule), für die sich hohe Potentiale im Bereich der Gesundheitsversorgung konstatieren lassen und respektiv die genannten Defizite repräsentieren.⁵ Konkret werden die Potentiale von KI-Ansätzen bei Patienten nach Hüft-Totalendoprothetik (TEP) (Asymmetrien und Unterschiede zu Gesunden), nach prothetischer Versorgung nach Oberschenkelamputation und im Bereich der Wirbelsäule (biologische Geschlechtsunterschiede, individuelle Bewegungsmuster und pathologische Unterschiede der Krankheitsbilder Rückenschmerz, Wirbelsäulenversteifung und Arthrose im Vergleich zu gesunden Subjekten) erprobt.

1.3 Publikationen

Nachfolgend soll in Kürze der Zusammenhang der einzelnen Publikationen beschrieben werden, um ein besseres Verständnis der folgenden Inhalte zu fördern. Die hier dargestellte Reihenfolge bezieht sich dabei auf den Entwicklungsprozess während

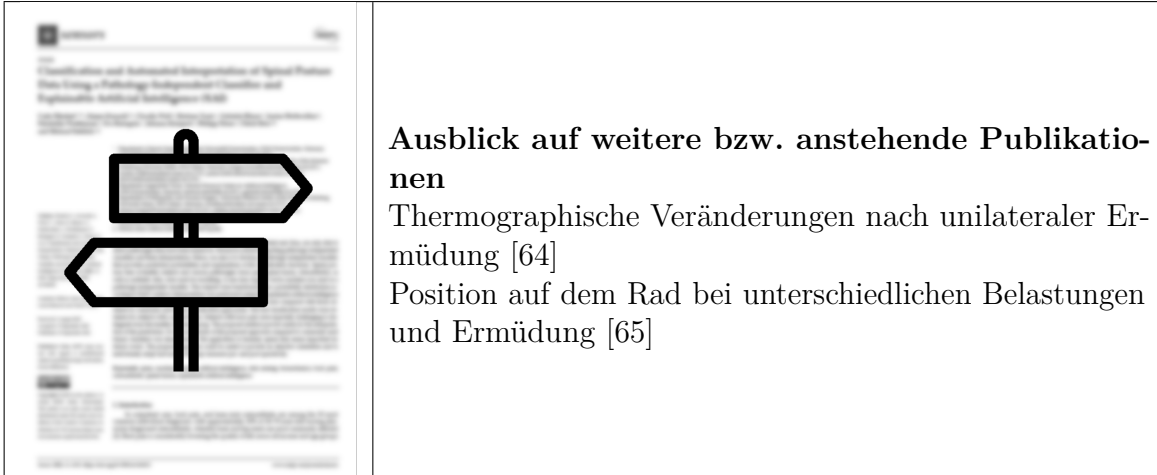
⁵ Für eine Erläuterung der Relevanz der Wirbelsäuledaten sei auf Sektion 2.3.1 verwiesen.

der Promotionsphase und kann in Teilen von den eigentlichen Publikationszeiten abweichen.

Tabelle 1.1 stellt hierzu einen Auszug der Inhalte und Hintergründe der Publikationen im Hinblick auf die Forschungsfragen dar. Die logisch aufeinander aufbauende Struktur ergibt sich dabei auf methodischer Ebene. Für detaillierte Angaben zur jeweils verwendeten Datenbasis sowie den eingesetzten Methoden sei auf das Resümee und Tabelle 4.1 verwiesen. Die vorbereitende Studie für Publikation 5 wird im Rahmen dieser Schrift nicht als Volltext angeführt, da der Autor aufgrund seines Ausmaßes der Mitwirkung nicht der Erstautor dieses Artikels ist. Aufbauende Arbeitsschritte zur Erweiterung der Datenbasis, Schaffung der Voraussetzungen zur multimodalen Analyse (siehe Sektion 4.7) sowie Transfer der beschriebenen Methoden in weitere sportwissenschaftliche Kontexte sind als Ausblick auf weitere Arbeiten jedoch im Folgenden nicht intensiv diskutiert, da diese eine Erweiterung der beschriebenen Ansätze perspektivisch anvisieren.

Tabelle 1.1: Übersicht der zentralen Inhalte der Publikationen

<p>sciendo Original Paper</p> <p>Biomedical Human Kinetics, 13, 177–186, 2021 DOI: 10.2478/bhm-2021-0022</p> <p>Feature extraction and gait classification in hip replacement patients on the basis of kinematic waveform data Carlo Dieder¹, Wolfgang Isenhardt¹, Bertman Taatz¹, Stephan Becker¹, Gabriele Bleser¹, Michael Fritsch²</p> <p>¹Department of Sports Science, Technische Universität Kaiserslautern, Kaiserslautern, Germany; ²First Sport and Movement Science, Paris Lodron University Salzburg, Austria; ³Department Angewandte Vision, German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany; ⁴Center Research Group WOODKAISER, Technische Universität Kaiserslautern, Kaiserslautern, Germany</p> <p>Abstract Study aim: To find an efficient robust gait-specific recognition or gait knowledge, which possesses an easy requirement for the description of conventional gait patterns, also used by arthroplasty (TEP). Abstract and methods: The gait of 22 patients after TEP was recorded using an optical motion capture system. The kinematic data of the normal position, inclination, and acceleration, as well as joint and segment angles, were used as initial feature. The random forest (RF) and support vector machine (SVM) algorithms were chosen for feature selection. The results were compared with the obtained data on the different dimensional kinematic features. Results: Hip movement in the sagittal plane, knee kinematics in the frontal and sagittal planes, ankle joint position data of the stance and pressure sensor data and inclination data of the center point of the pelvis and the flexion in the right hip segment for classification (accuracy: 91.0%). With feature selection, better results were obtained (accuracy: 85.00%) in relation. Conclusion: The proposed approach can be used to identify and individualize abnormal gait patterns during the rehabilitation process via waveform data. The results indicate that precise and consistent data also allow an efficient administration for the user.</p> <p>Keywords: Classification – Total hip arthroplasty – Feature selection – Dimensionality reduction</p> <p>Introduction The total hip arthroplasty (TEP) is the most important reason for the treatment of degenerative hip osteoarthritis [1]. The frequency of hip joint replacement surgery is increasing year after year and is expected to continue to increase in the coming years. In addition, progress in the past decades often displays an increasing absolute number [2]. The results of conventional gait analysis are often limited, which makes it difficult to detect subtle differences in gait patterns, which could be used for further targeted interventions. Gait classification is an important tool for clinical diagnosis and fitness identification [3]. The classification of the gait pattern can be used to derive knowledge about the user's characteristics and to identify of</p> <p>clinical relevance. With modern movement tracking systems, large amounts of data are available (e.g., 3D and motion learning models) have gained importance compared to classical motion capture [4]. Feature selection and dimensionality reduction are important steps toward improving a model's accuracy and interpretability, preventing overfitting, and reducing the associated computing power [5]. Just angles of specific points in time and descriptive statistics, such as the frequency of steps or peak values, are not sufficient for the classification [2]. However, if the questionnaire whether such features actually help gait pattern differences or if kinematic information is described a priori. For example, kinematic data have been given relatively little consideration, although it has been shown that they may represent information about gait patterns [14]. The dependence on gait knowledge, kinematics,</p> <p>Author's address: Carlo Dieder, Department of Sports Science, Technische Universität Kaiserslautern, Erwin-Schicklerstraße 53, 67663 Kaiserslautern, Germany; carlo.dieder@tu-kl.de</p>	<p>Publikation 1 [59] Einfache robuste Feature Selektion zur objektiven Bestimmung relevanter Gangparameter unter möglichst optimaler Ausnutzung der Datenbasis. Welche Parameter sind am wichtigsten zur Beschreibung von asymmetrischen Gangmustern bei Patienten nach einer Hüft-TEP?</p>
<p>ONLINE FIRST: https://doi.org/10.1002/ajpa.24260 https://onlinelibrary.wiley.com/doi/10.1002/ajpa.24260</p> <p>General method for automated feature extraction and selection and its application for gender classification and biomechanical knowledge discovery of sex differences in spinal posture during stance and gait Carlo Dieder¹, Jürgen Kopp¹, Claudia Wolf¹, Bertman Taatz¹, Gabriele Bleser¹, Janine Hoffmann¹, Philipp Dietz¹, Michael Fritsch² and Stefan Mielke³</p> <p>¹Department of Sports Science, Technische Universität Kaiserslautern, Kaiserslautern, Germany; ²Institute of Physical Therapy, Hochschule Rhein-Waal, Unkenburg, Germany; ³Department of Sports Science, Technische Universität Kaiserslautern, Kaiserslautern, Germany; ⁴Center Research Group WOODKAISER, Technische Universität Kaiserslautern, Kaiserslautern, Germany; ⁵Department of Orthopedics and Sports Injury, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany</p> <p>Abstract Human kinematics enables to capture subtle biomechanical patterns often resulting in subtle differences. These can be used to improve the understanding of the complexity of human motion and to optimize the design of prostheses and orthoses. However, feature extraction methods are often not suitable for the analysis of large amounts of data. This paper presents a general method for automated feature extraction and selection and its application for gender classification and biomechanical knowledge discovery of sex differences in spinal posture during stance and gait. The proposed method is based on a combination of feature extraction and selection methods. The results of the proposed method are compared with the results of the classical methods. The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features. The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features. The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features.</p> <p>1. Introduction Human biomechanics is a complex and multidisciplinary field. To approach an understanding of its complexity, multiple parameters and variables have to be taken into account, which often results in multiple related tables of data (collected data). Modern movement tracking systems enable the collection of large amounts of data (collected data). Modern movement tracking systems enable the collection of large amounts of data and to select the most relevant features. The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features. The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features.</p> <p>2. Methods The proposed method is based on a combination of feature extraction and selection methods. The results of the proposed method are compared with the results of the classical methods. The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features. The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features.</p> <p>3. Results The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features. The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features. The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features.</p> <p>4. Discussion The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features. The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features. The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features.</p> <p>5. Conclusion The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features. The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features. The proposed method is able to extract relevant features from large amounts of data and to select the most relevant features.</p> <p>Author's address: Carlo Dieder, Department of Sports Science, Technische Universität Kaiserslautern, Erwin-Schicklerstraße 53, 67663 Kaiserslautern, Germany; carlo.dieder@tu-kl.de</p>	<p>Publikation 2 [60] Erweiterung von Publikation 1 um relationale Daten sowie fortgeschrittene robuste Feature Selektion. Gibt es Unterschiede im Bereich der Wirbelsäule zwischen den biologischen Geschlechtern? Welche komplexen Parameter sind für die Unterscheidung besonders relevant?</p>



1.4 Besonderer Forschungsbeitrag

Der besondere Forschungsbeitrag der vorliegenden Arbeit ergibt sich insbesondere durch Erkenntnisse 1) im Bereich des Einsatzes von ML bei biomechanischen Daten und 2) auf Basis bislang wenig betrachteter rasterstereographisch gemessener Wirbelsäuledaten.

1) Methodisch wird das übergeordnete Ziel verfolgt, die Basis zu erarbeiten, um hybride Architekturen aus Expertensystemen und ML zu ermöglichen und KI anwendbar in praktischen Settings zu machen. Zentral adressiert die Arbeit damit auf bestehende Limitationen und soll aufzeigen, wie diese potentiell überwunden werden können bzw. Perspektiven erarbeiten, welche Ansätze vielversprechend erscheinen können. Ferner soll weiterer Forschungsbedarf aufgezeigt werden. Des Weiteren ergibt sich der Forschungsbeitrag durch Offenlegung notwendiger Transferschritte bestehender allgemeiner ML Methoden für die Anwendung in biomechanischen bzw. klinischen Settings und die wissenschaftliche Evaluation von Potentialen und Limitationen von ML im biomechanischen Kontext.

2) Messmethodisch ermöglicht das Verfahren der Rasterstereographie mit dem DIERS 4D motion® Lab (DIERS International GmbH, Schlangenbad, Deutschland) neuartige Aspekte der Wirbelsäule zu betrachten, welche zuvor messmethodisch nur schwer bzw. unvollständig analysiert werden konnten. Ein besonderer Beitrag wird durch die teilweise Verwendung von dynamischen Wirbelsäuledaten geleistet, welche bislang aufgrund messmethodischer Aspekte kaum betrachtet werden konnten. Ferner wird ein Beitrag geleistet, ML Methoden besonders für den vorliegenden Bereich anzupassen.

1.5 Aufbau

Nachfolgend beschreibt Kapitel 2 den Hintergrund der Thematik. In Kapitel 3 werden die 5 relevanten Publikationen in Originalform (die vorbereitende Publikation für Publikation 5 wird ausgeklammert) angeführt und jeweils um eine kurze deutschsprachige Zusammenfassung ergänzt. Die Erkenntnisse der Publikationen werden im Anschluss in Kapitel 4 additiv zusammengeführt und diskutiert. Aufbauend auf den Erkenntnissen der Synopsis wird zuletzt ein Fazit und ein Ausblick gegeben.

Forschungsdesiderat und Hintergrund

ZUNÄCHST werden nachfolgend Grundlagen von KI und ML angerissen, um ein besseres Verständnis des Forschungsstandes sowie der in den Publikationen dargestellten Inhalte zu ermöglichen. Nach einer Begriffsklärung und Darstellung der Aktualität der Thematik werden anhand von Studien die Potentiale von KI bzw. ML zur Entscheidungsunterstützung und Knowledge Discovery dargestellt. Hierzu soll auch eine Einordnung in das Feld der E-Health erfolgen¹. Aufbauend wird der Forschungsstand im Bereich der Applikation im klinischen sowie explizit im biomechanischen Kontext dargestellt, wobei eine Erweiterung der Perspektiven der einzelnen Publikationen angestrebt wird. Im Anschluss werden messmethodische Grundlagen erläutert, wobei der Fokus auf die Wirbelsäulenanalyse gelegt wird.

2.1 Grundlagen Künstliche Intelligenz (KI), Maschinelles Lernen (ML) und Deep Learning

2.1.1 Begriffsklärung und aktuelle Relevanz

Unter Künstlicher Intelligenz (KI) versteht man im Allgemeinen eine Art menschliche Intelligenz, welche von Maschinen ausgeführt wird [66]. Oft wird zwischen starker (allgemeine Intelligenz, die der des Menschen entspricht oder diese übertrifft) und schwacher KI (Fokus auf konkrete Anwendungsprobleme) unterschieden, wobei generell fraglich ist, inwieweit eine starke KI überhaupt realisierbar ist² [67]. Maschinelles Lernen (ML) ist ein Ansatz, um KI zu erreichen. Man versteht darunter Computersysteme, welche in die Lage versetzt werden, aus Beispielen bzw. Erfahrungen zu lernen, ohne explizit programmiert zu sein. Als wichtige Technik zur Implementierung von ML hat sich wiederum Deep Learning abgesetzt [16] (siehe Abbildung 2.1). Im Rahmen dieser Arbeit geht es um die lernende KI, also das ML. KI geht aber auch ohne Lernvorgänge, wie bei der symbolischen oder regelbasierten KI [68]. Derartige Ansätze sollen für den weiteren Verlauf jedoch ausgeklammert werden.

Begriffsklärung

Zur Verdeutlichung werden bei einer Klassifikationsaufgabe im Falle des klassischen Programmierens von Experten bestimmte Merkmale für die einzelnen Zielklassen definiert. Anschließend wird ein Modell von Menschen programmiert, welches

¹ Das Vorhaben zielt nicht darauf ab konkrete Anwendungen im Bereich der E-Health zu entwickeln. Stattdessen sollen Perspektiven aufgezeigt werden, wie Grundlagen für diese auf methodischer Ebene geschaffen werden können. Neben der Nähe zum E-Health Feld lässt sich das Vorhaben methodisch dem Bereich der Knowledge Discovery zuordnen.

² Bei dem Einsatz von KI bei der Auswertung von biomechanischen Daten reden wir über eine schwache KI und nicht über eine starke.

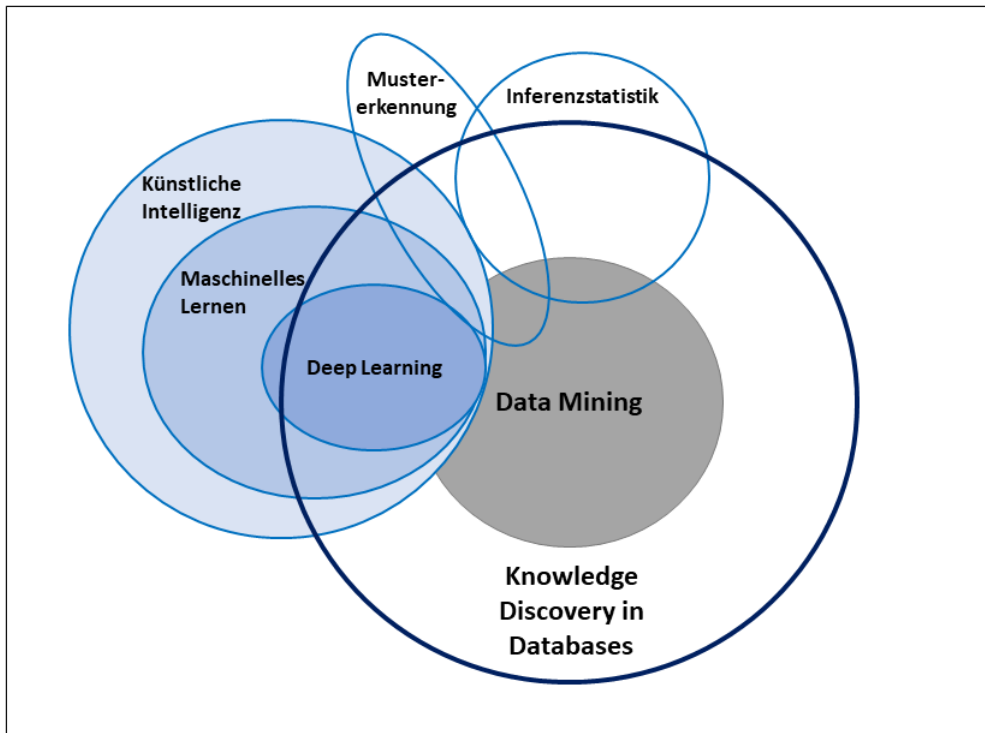


Abbildung 2.1: Zusammenhänge Data Mining, Künstliche Intelligenz (KI), Maschinelles Lernen (ML), Deep Learning sowie weitere Bereiche. Eigene Darstellung in Anlehnung an [69].

wenn-dann Entscheidungen auf Basis dieser Merkmale trifft. Um die Klassifikationsaufgabe mittels ML zu lösen wird hingegen ein Modell auf Grundlage von gelabelten Beispieldaten (den Trainingsdaten) angelernt. Dieses Modell kann dann genutzt werden, um neue Daten zu klassifizieren (siehe Abbildung 2.2).

Der Begriff ML wird oft in Zusammenhang mit dem Begriff Data Mining verwendet. Data Mining ist die Suche nach bedeutungsvollen, nicht-trivialen Informationen in (zumeist großen) Datenmengen [70]. Data Mining boomt und findet immer mehr Beachtung sowohl in Wirtschaft als auch Wissenschaft [15]. Unter Verwendung u.a. von Technologien der Mustererkennung sowie statistischen und mathematischen Techniken beschreibt Data Mining einen Prozess der Extraktion impliziter, potentiell nützlicher Informationen (z.B. Wissensregeln, Beschränkungen, Regelmäßigkeiten) aus Daten [71]. Ein Teil der Methoden des Data Minings beinhaltet somit das ML (siehe auch Abbildung 2.1). Data Mining lässt sich im Gesamtprozess der Wissensentdeckung in Datenbanken (engl. Knowledge Discovery in Databases; KDD) verorten, wobei in Abgrenzung zum Data Mining KDD auch die Vorbereitung der Daten sowie die Bewertung der Ergebnisse umfasst [72]. Generell lassen sich die angesprochenen Themenfelder unter dem interdisziplinären Wissenschaftsfeld der *Data Science* subsumieren, welche allgemein die Lehre von der Extraktion von verallgemeinerbarem Wissen aus Daten beschreibt [73].

**Data
Mining &
KDD**

ML hat sich in zahlreichen realen Anwendungsfeldern erfolgreich bewährt. Dies beinhaltet u.a. Gesichtserkennung [74], Spracherkennung [75], Objekterkennung [3], Malware Detektion [76] und Spamfilter [5]. Zudem lassen sich Einsatzschwerpunkte in sicherheitskritischen Anwendungsfeldern finden, bei denen Fehler besonders große Auswirkungen haben (z.B. autonomes Fahren [4], oder wie schon mehrfach angeführt im Gesundheitswesen [77], [78]). Die Relevanz der Thematik wird auch ersichtlich

**Anwendun-
gen**

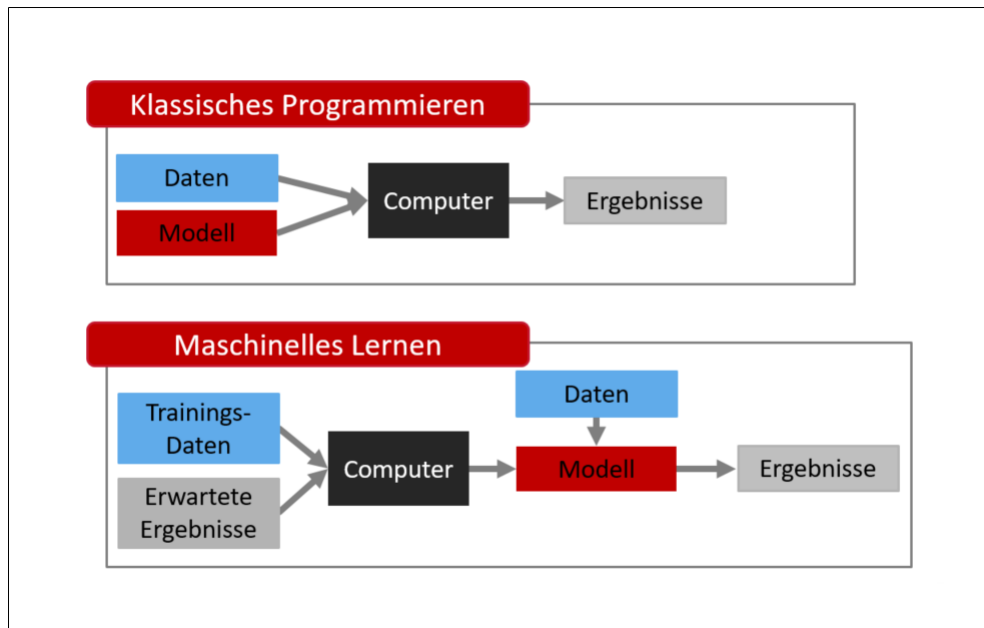


Abbildung 2.2: Abgrenzung Maschinelles Lernen (ML) von klassischem Programmieren. Eigene Darstellung.

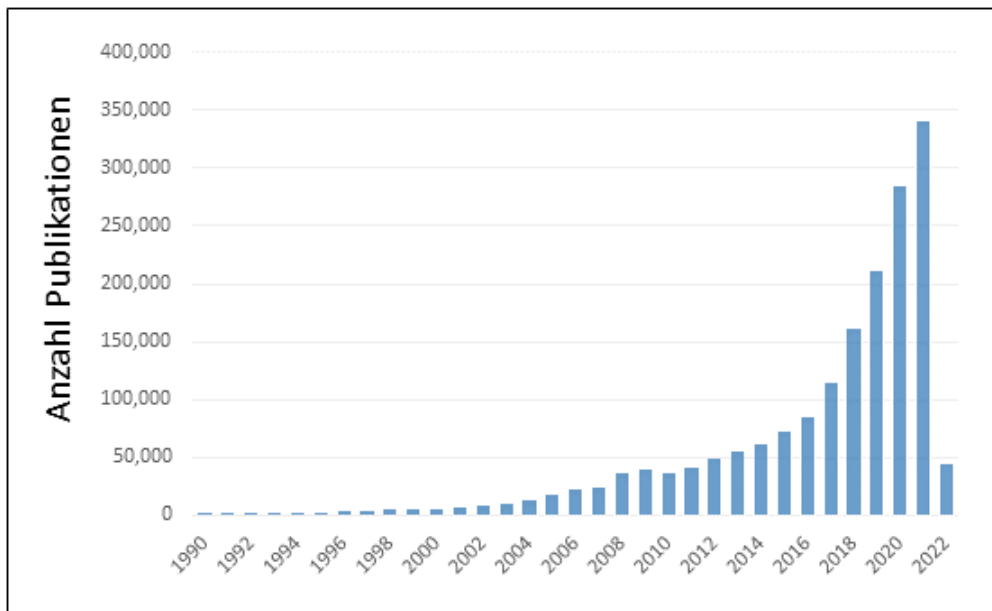


Abbildung 2.3: Jährliche Anzahl an Publikationen zu dem Suchbegriff „machine learning“. Stand: 22.02.2022. Suche mit <https://www.dimensions.ai/>. Eigene Darstellung.

bei Betrachtung der jährlichen Anzahl an Publikationen zur Thematik. Unter Verwendung des Suchbegriffs „machine learning“ kann ein starker und noch immer andauernder Anstieg erkannt werden (siehe Abbildung 2.3). Unter der Prämisse, dass „Künstliche Intelligenz [...] ein Schlüssel zur Welt von morgen“ [79] ist lässt sich vermuten, dass die Relevanz der Thematik in Zukunft noch weiter ansteigen wird.

2.1.2 Arten des Lernens und Einsatz von Maschinellern Lernen (ML) im Rahmen von Knowledge Discovery in Databases (KDD)

Es können grundsätzlich in Anlehnung an [80] drei Arten des Lernens bei Algorithmen des ML kategorisiert werden:

- **Überwachtes Lernen (engl. Supervised Learning):** Ein Algorithmus lernt auf Basis gelabelter Daten und liefert ein Ergebnis, welches zur Beurteilung der Genauigkeit auf dem Trainingsdatensatz verwendet werden kann.
- **Unüberwachtes Lernen (engl. Unsupervised Learning):** Ein Algorithmus versucht auf Basis ungelabelter Daten Merkmale und Muster zu extrahieren.
- **Bestärkendes Lernen (engl. Reinforcement Learning):** Selbständiges Erlernen einer Strategie, bei der ein Agent versucht die zu erhaltenden Belohnungen zu maximieren.

Zudem existiert als Mischform das halb-überwachte Lernen (engl. Semi-Supervised Learning), bei dem eine (kleine) Menge gelabelter Daten eine größere Menge an ungelabelten Daten unterstützt [81]. Grundsätzlich ist in den Hauptfeldern Klassifikation und Regression überwachtes Lernen nützlich. Unter Klassifikation versteht man dabei die Vorhersage eines diskreten Wertes zur Identifizierung der Eingangsdaten als Mitglied einer bestimmten Gruppe bzw. Klasse. Regressionsprobleme betrachten hingegen die Vorhersage kontinuierlicher Daten. Als ein weiteres wichtiges Feld soll an dieser Stelle das Sequenzlernen (engl. Sequence Learning) angeführt werden, welches sich mit der Anwendung von ML Methoden bei sequentiellen Daten beschäftigt, wie die Vorhersage eines bzw. mehrerer nächster Werte einer gegebenen Eingabesequenz oder der Klassifikation dieser Eingabesequenzen [82]. Für unüberwachtes Lernen lassen sich verschiedene Aufgabenfelder wie Clustern, Anomalieerkennung, Assoziation und Autoencoder charakterisieren, wobei die Messung der Genauigkeit aufgrund der fehlenden Grundwahrheit oftmals erschwert ist [16].

Klassifikationsaufgaben mittels ML können genutzt werden um Informationen zu generieren, wobei die Klassifikationsaufgabe das Mittel zum Zweck darstellen kann. Wird ML zur Exploration eingesetzt, lassen sich nach [72], [83] im Detail die folgenden Schritte im Rahmen von KDD charakterisieren. Im ersten Schritt (i) geht es um die Beschaffung bzw. die Vorauswahl von Daten (im einfachsten Fall durch Rückgriff auf eine Datenbank), welche für die aufbauenden Schritte verwendet werden sollen. Im Anschluss ist der Schritt der Datenvorverarbeitung (ii) anzusiedeln, unter der sich u.a. Arbeitsschritte wie der Umgang mit fehlenden

Workflow

Werten, Ausreißerdetektion und Behandlung sowie die Integration unterschiedlicher Datenquellen einordnen lässt.

Die (iii) Datentransformation folgt im Anschluss, welche auch den Schritt der Feature Extraktion, also die Reduktion der Anzahl an Eingangsvariablen für die Entwicklung eines Modells, beinhaltet. Diese ist nicht nur wichtig um die Genauigkeit eines Modells zu verbessern, eine Überanpassung zu verhindern und die erforderliche Rechenleistung zu reduzieren, sondern spielt auch eine große Rolle dabei die Interpretierbarkeit zu erhöhen und bedeutsame Informationen zu extrahieren [84]–[86]. Für genauere Ausführungen der folgenden Inhalte sei beispielsweise auf [86]–[88] verwiesen. In Kürze kann demnach im Allgemeinen unterschieden werden zwischen Dimensionsreduktion und Feature Selektion. Dimensionsreduktion zielt darauf ab, die Eingabedaten auf einen niedrigerdimensionalen Feature-Raum zu projizieren. Feature Selektion bezeichnet dagegen die Auswahl einer Teilmenge an Features aus dem ursprünglichen Datensatz. Die Feature Selektion lässt sich erneut strukturieren in unüberwachte (ohne Verwendung der Zielvariable) und überwachte (unter Verwendung der Zielvariable) Selektion. Zur unüberwachten Selektion zählt bspw. der Ausschluss von redundanten Features durch eine Korrelationsanalyse. Zu den überwachten Methoden zählen wiederum Wrapper (Auswahl auf Basis Modellperformanz), Filter (Auswahl basierend auf der Beziehung der Features mit der Zielvariable) und Intrinsic / Embedded Methoden (Verwendung von Algorithmen, die über integrierte Feature Selektionsmethoden verfügen).

Erst nach den vorausgehend geschilderten Schritten erfolgt das eigentliche Data Mining (iv), dessen Ergebnisse abschließend interpretiert und evaluiert werden müssen (v), wobei auch geeignete Visualisierungsformen zu beachten sind. Bei den genannten Schritten können dabei mehrere Iterationen notwendig sein sowie Schleifen zwischen den einzelnen Schritten auftreten [72].

2.1.3 Relevante Algorithmen

Eine ständig wachsende Menge an Algorithmen existiert. Tabelle 2.1 gibt einen Überblick über die wichtigsten verwendeten ML Algorithmen in den einzelnen Publikationen.

Tabelle 2.1: Übersicht über verwendete ML Algorithmen. Die Art bezieht sich auf die Art der Verwendung in den Publikationen. SVM = Support Vector Machine; RF = Random Forest; KNN = Künstliche neuronale Netze; LR = Logistische Regression; k-NN = k-Nearest-Neighbours Algorithmus; PCA = Principal Component Analysis; LDA = Linear Discriminant Analysis; OCSVM = One Class Support Vector Machine; IF = Isolation Forest; Klf = Klassifikation; Dimr = Dimensionsreduktion; Anod = Anomaliedetektion

Algorithmus	Art	Beschreibung	Referenz
SVM	Klf	Breiter-Rand-Klassifikator (engl. Large Margin Classifier) welcher versucht eine Menge von Objekten in Klassen zu unterteilen, dass um die Trennungsfläche herum ein möglichst breiter Bereich frei von Objekten bleibt	[89]
RF	Klf	Randomisierung von Entscheidungsbäumen auf Basis von Bootstrap-Datensätzen (Bagging) und der randomisierten Auswahl von Features (Feature Bagging)	[90]
KNN	Klf	Miteinander verbundene Knoten (Neuronen), die Informationen von außen oder von anderen Neuronen aufnehmen, modifizieren, weiterleiten bzw. als Ergebnis ausgeben	[91]
LR	Klf	Erweiterung des linearen Regressionsmodells für Klassifikationsprobleme durch Modellierung der Wahrscheinlichkeiten für Klassenzugehörigkeiten mittels Logistischer Funktion (Sigmoidfunktion)	[92]
k-NN	Klf	Bestimmung der Klassenzugehörigkeit von Punkten basierend auf ähnlichen, naheliegenden Punkten (nächste Nachbarn)	[93]
PCA	Dimr	Unüberwachte Suche nach unkorrelierten Linearkombinationen, welche die maximale Varianz in Daten erfassen	[94]
LDA	Dimr	Überwachte Findung eines Feature-Unterraums, der die Klassentrennbarkeit optimiert bei gleichzeitiger Reduktion der Dimensionalität	[95]
OCSVM	Anod	Variation der SVM welche versucht Inlier auf Basis einer angelernten Entscheidungsfunktion zu umhüllen	[96]
IF	Anod	Erkennung von Anomalien durch Isolierung auf Basis von Entscheidungsbäumen. Zufällige Partitionierung von Merkmalen erzeugt kürzere Pfade in Bäumen für anomale Datenpunkte	[97]

2.1.4 Explainable Artificial Intelligence (XAI)

Aufgrund des hohen Fokus der vorliegenden Arbeit auf Interpretierbarkeit der Modelle soll eine intensivere Betrachtung von Methoden der Explainable Artificial Intelligence (Erklärbaren Künstlichen Intelligenz; XAI) abschließend erfolgen. Viele Modelle des ML weisen einen Black Box Charakter und mangelnde Transparenz auf. Das entspricht beispielsweise nicht den Anforderungen der europäischen Datenschutzgrundverordnung (GDPR, EU 2016/679) [52] und schränkt die praktische Anwendung beispielsweise im klinischen Kontext stark ein (siehe auch Einleitung). XAI hat in den letzten Jahren großes Interesse gefunden und bietet Methoden zur Erhöhung der Transparenz, Interpretierbarkeit und Erklärbarkeit und damit Vertrauenswürdigkeit von Black Box Modellen [50]. Unter Transparenz, Interpretierbarkeit und Erklärbarkeit versteht man dabei:

- *Transparenz*: Ein ML Modell ist transparent, wenn die Bestimmung der Modellparameter auf Basis der Trainingsdaten sowie die Generation der Outputs auf Basis der Testdaten beschrieben und begründet werden können [98].
- *Interpretierbarkeit*: Interpretierbarkeit beschreibt die Möglichkeit ein ML-Modell nachzuvollziehen und die zugrundeliegenden Entscheidungsgrundlagen für Menschen verständlich darzustellen [98]. Ein ML Ansatz ist umso interpretierbarer, desto einfacher Ursache-Wirkungs-Beziehungen zwischen Input und Output identifiziert werden können [99].
- *Erklärbarkeit*: Interpretierbarkeit impliziert nicht unbedingt Erklärbarkeit vice versa, da der Mensch nicht zwingend die interne Logik oder zugrunde liegenden Prozesse der Modelle verstehen kann [100]. Ein Modell ist umso erklärbarer, je tiefer das Verständnis des Menschen über interne Abläufe (während Modelltraining bzw. Modellentscheidungen) ist [99]. Domänenwissen stellt einen wesentlichen Bestandteil von Erklärbarkeit dar, da Erklärbarkeit meistens nicht rein algorithmisch erreicht werden kann [98].

Für die weitere Strukturierung von XAI Methoden hinsichtlich der Arten von Erklärungen soll eine Einteilung angelehnt an [101] in (i) Datenexploration (engl. data exploration), (ii) Modellerklärung (engl. model explanation) und (iii) Vorhersageerklärungen (engl. prediction explanation) vorgestellt werden. **Strukturierung XAI Methoden**

Methoden der (i) Datenexploration versuchen die Daten in einen Raum zu projizieren um zu ermöglichen, sinnvolle Strukturen zu finden. Als Beispiel ist hier der Algorithmus t-distributed Stochastic Neighbor Embedding (t-SNE) [102] zu nennen, welcher in Publikation 4 zur Visualisierung der einzelnen Personen im zweidimensionalen Raum verwendet wurde.

Unter (ii) Modellerklärung lassen sich Methoden einordnen, welche Interpretationen ermöglichen, was ein trainiertes Modell gelernt hat (charakteristische Repräsentationen einer ganzen Klasse). Als wichtige Methode lässt sich Spectral Relevance Analysis (SpRAy) [103] benennen.

Methoden der (iii) Vorhersageerklärungen versuchen das lokale Verhalten eines Modells (für eine einzelne Eingabeinstanz) zu erklären, z.B. indem aufgezeigt wird welche Feature die Vorhersagen eines Klassifikators am stärksten beeinflussen. Eine weitere Unterscheidung kann hier in (a) selbsterklärende Modelle und (b) post-hoc Methoden vorgenommen werden, wobei die eingesetzten Methoden sich (b) zuordnen lassen. Post-hoc Methoden können weiter in (i) propagation-basierte (Deep Learning

Important FeaTures (DeepLIFT) [104]; Layer-wise Relevance Propagation (LRP) [105]), (ii) perturbation-basierte (Local Interpretable Model-Agnostic Explanations (LIME) [106]), und (iii) Shapley-value-basierte Methoden (SHapley Additive exPlanations (SHAP) [107]) unterschieden werden.

LIME hat im Rahmen der Publikationen besonderen Fokus erfahren. LIME führt eine Approximation einer lokalen Vorhersage eines Black Box Modells mit einem einfacher interpretierbaren Modell durch. Dafür werden zunächst durch Störung (Perturbation) Datenpunkte um eine Instanz von Interesse erzeugt. Anschließend werden diese Datenpunkte mit dem Black Box Modell vorhergesagt und nach ihrer Nähe zur ausgewählten Instanz gewichtet. Schließlich wird ein interpretierbares Modell an den gewichteten Datenpunkten angelern und zur Erklärung der Vorhersage verwendet [106].

2.2 E-Health die Anwendung von Künstlicher Intelligenz (KI) in Medizin und Biomechanik

2.2.1 E-Health und der Einsatz von KI-Anwendungen in der menschlichen Lebensspanne

„Für die erfolgreiche Weiterentwicklung unserer Gesundheitsversorgung ist das Vorantreiben der Digitalisierung die zentrale Voraussetzung“, so das Bundesministerium für Gesundheit [108]. Damit wird die Relevanz von E-Health Anwendungen, welche auf Basis moderner Kommunikations- und Informationstechnologien unterstützend bei der Behandlung und Betreuung von Patienten:innen sind, besonders betont. Hierunter fallen zum einen auch Entwicklungen wie die elektronische Patientenakte, das elektronische Rezept und Telemedizin. Zum anderen erfahren jedoch auch KI und Big Data Anwendungen als Teil der E-Health Bewegung an Relevanz. Entsprechende Fördermaßnahmen sind eingebettet in die KI-Strategie der Bundesregierung unter der Prämisse, dass „Künstliche Intelligenz [...] ein Schlüssel zur Welt von morgen“ [79] ist. [9] führt ferner an, dass KI eine Rolle bei der Verbesserung vielfältiger Probleme im Gesundheitswesen (Diagnose- und Behandlungsfehler, Ressourcenverschwendung, Ineffizienz und unzureichende Zeit) spielen kann.

E-Health findet auch im sportlichen Bereich Einzug. Im Rahmen der Digitalisierung findet KI Methodik generell schon Einsatz in sportlichen Kontexten (z.B. bei der Vorhersage der Sauerstoffaufnahme bei körperlicher Aktivität [109] oder bei der Sportergebnisvorhersage [110]). Im gesundheitlichen sportlichen Bereich erfahren beispielsweise Anwendungen im Bereich der Verletzungsüberwachung [111] oder Bewertung des Verletzungsrisikos bei Fußballspielern [112] Relevanz.

Es ist auffällig, wie auch aus den dargestellten Arbeiten ersichtlich wird, wie KI sich in vielfältige Anwendungen der menschlichen Lebensspanne integriert. [9] führt dies für den klinischen Einsatzbereich aus. Abbildung 2.4 erweitert diese Perspektive um biomechanische Aspekte sowie Einsatzschwerpunkte im Rahmen des Sports und Aspekte des alltäglichen Lebens, um visuell einen Überblick über die breitgefächerten Einsatzschwerpunkte zu geben.

*sportliche
Anwendungen*

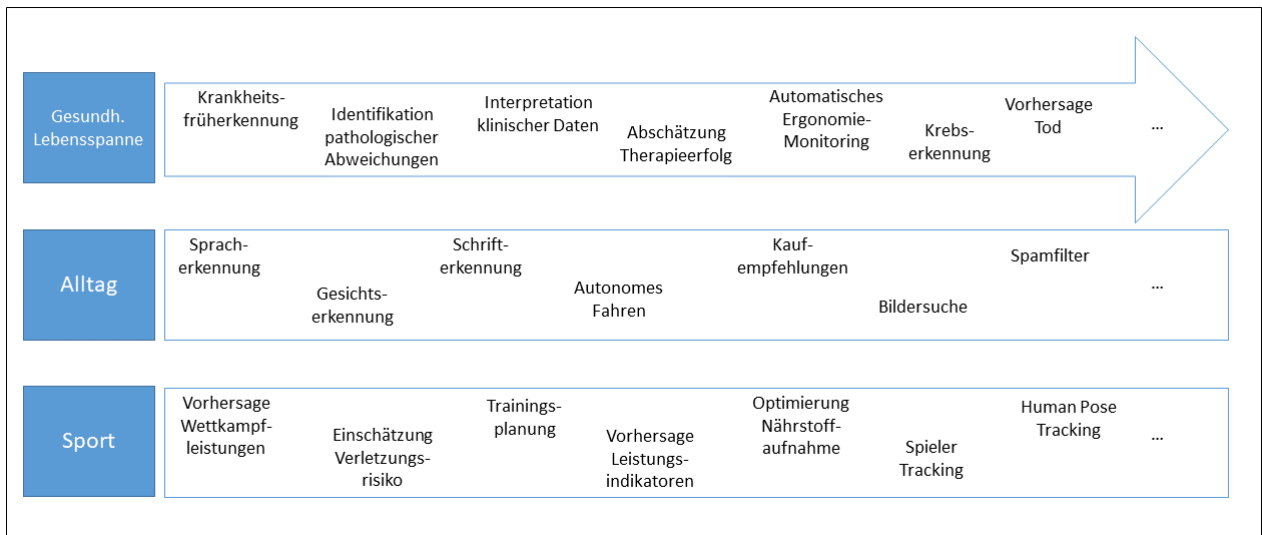


Abbildung 2.4: Einsatz von KI in der Lebensspanne des Menschen in alltäglichen, klinischen und sportlichen Bereichen. Eigene Darstellung. Die klinischen Aspekte sind angelehnt an [9] und wurden um die biomechanische Perspektive erweitert.

2.2.2 Exemplarischer Einsatz von KI bei klinischen Daten

Wie anfänglich in der Einleitung angeführt lassen sich eine Vielzahl an Arbeiten mit KI-Systemen im klinischen Kontext finden, wobei die direkte Anwendung bei biomechanischen Daten unterrepräsentiert ist. Daher sollen einige beispielhafte generelle Arbeiten betrachtet werden, um dahingehend mögliche Potentiale der Methodik im Bereich der Biomechanik abzuleiten, welche einen generellen Mehrwert begründen. Im Anschluss an diese generelle Betrachtung wird konkret der Bereich der biomechanischen Datenanalyse beleuchtet.

Im Bereich der medizinischen Bildgebung lässt sich eine große Vielzahl an Forschungsarbeiten finden. Die medizinische Bildgebung befasst sich dabei mit der Analyse, Visualisierung und Exploration medizinischer Bilddaten [113]. KI konnte Erfolge bei der Interpretation von Bilddaten erzielen, wie im Kontext von Netzhautscans [29], Röntgenbildern [114] und Ultraschall [115]. Neben Bilddaten werden auch eine Vielzahl anderer Datenquellen im klinischen Kontext verwendet (z.B. KI in der Kardiologie [27]). Der KI wird eine Unterstützungsrolle zugeschrieben und so kann beispielsweise durch Kombination klassischer Bildverarbeitung und Deep Learning eine Unterstützung radiologischer Befunde gezeigt werden [116]. KI werden Aufgaben anvertraut, bei denen Experten oft uneinheitliche Meinung haben, wie die Identifizierung von Lungentuberkulose auf Röntgenaufnahmen des Brustkorbs [114]. Einige KI Anwendungen wurden bereits von der Food and Drug Administration (FDA) genehmigt (z.B. MaxQ-AI - Diagnose von Hirnblutungen auf Basis CT [117]; Imagen OsteoDetect - Erkennungs- und Diagnosesoftware zur Erkennung von Handgelenksfrakturen [118]).

Forschungsarbeiten von KI im klinischen Kontext haben oft den Fokus auf die Leistung im Vergleich zu menschlichen Ärzten gelegt. Im Allgemeinen liegen bei vielen Aufgaben klar definierte Input-Features und binäre Outputs vor, welche relativ leicht zu evaluieren sind (Klassifikationsaufgaben). Mittels KI konnten so ähnliche Genauigkeiten, z.B. bei Erkennung von Hautläsionen auf Basis von Bildern, wie durch zertifizierte Dermatologen erzielt werden [12]. Teilweise konnte auch gezeigt werden, dass ML bei der Erkennung von Krankheiten den Menschen übertreffen kann, wie bei angeborenen Herzfehlern [30] oder bei Cervix Bildern im Rahmen

der Krebsvorsorge [31]. Mittels ML können dabei Auffälligkeiten erkannt werden, welche vom Menschen noch nicht oder nur sehr schwer ersichtlich sind (z.B. erste Auffälligkeiten einer Krebserkrankung [31], [33]). Zudem konnten die Methoden dazu beitragen, falsch-positive Fehler und Unterschiede in der Krankheitserkennung aufgrund der unterschiedlichen Erfahrungsstufen von Ärzten zu reduzieren [34]. Im Detail konnte beispielsweise das Lungenkrebs-Screening durch Reduktion falsch-positiver und falsch-negativer Raten sowie Interrater-Reliabilität optimiert werden [32]. Damit einhergehend kann es durch die datengetriebene, objektive Orientierung zu einer Reduktion von Fehlentscheidungen kommen, da Entscheidungen nicht nur auf subjektiven Experteneinschätzungen beruhen müssen. Als Sonderfall können KI Chatbots genutzt werden um medizinische Fragen zu beantworten, Wartezeit zu reduzieren und zu beurteilen wie dringend Fälle sind [119]. Bei all dem können Computer meist schneller Entscheidungen treffen als Menschen, was gerade im Bereich der Notfallmedizin von hoher Bedeutung ist [9].

Highlights:

Der Mehrwert sowie die Potentiale von KI ergeben sich entsprechend der obigen Ausführungen auf folgenden Ebenen, welche sich auf individuelle, gesellschaftliche und wirtschaftliche Ebenen auswirken können:

- Reduktion von Fehlentscheidungen
- Früherkennung von Auffälligkeiten
- Verbesserung Entdeckungsraten
- Kompensation unterschiedlicher Erfahrungsstufen bei Ärzten
- Automatisierung und Beschleunigung
- Exploration und Identifikation neuer Korrelate
- Medizinische Abläufe effektiver und effizienter gestalten
- Senkung von Gesundheitskosten
- Verbesserung personalisierte Medizin

2.2.3 Exemplarischer Einsatz von KI bei biomechanischen Daten

Gang- und Haltungsanalysen sind hoch-relevant in wissenschaftlichen, sportlichen und klinischen Kontexten. (Klinische) Ganganalysen fokussieren dabei auf die quantitative Beschreibung und Analyse des menschlichen Gangs aus kinetischer (Bodenreaktionskräfte und Gelenkmomente) und kinematischer (Gelenkwinkel) Sicht [120]. Aufgrund der aktuellen Entwicklungen in der Biomechanik und der Fülle an entstehenden Daten spricht [21] von Big Data Charakteristika in der Biomechanik, welche er durch die vorherrschenden Merkmale (a) Volumen (Datenmenge), (b) Vielfalt (verschiedene Datenkategorien), (c) Geschwindigkeit (schnelle Generierung neuer Daten), (d) Genauigkeit (Qualität der Daten) und (e) Wert (in den Big Data Beständen) begründet. Eine Übersicht der Datenlage, welche Ausgangspunkt für

Big Data

die Big Data Charakteristika bildet, wird in Abbildung 2.5 gegeben.

Mehr und mehr Daten werden pro Subjekt erhoben (*Volumen*). Für ein holistisches Verständnis des Menschen muss dessen Komplexität beachtet werden. Daher darf die biomechanische Analyse nicht nur auf kinetische und kinematische Daten beschränkt sein. Stattdessen müssen verschiedenste Datenquellen berücksichtigt werden. Neben den primären kinematischen und kinetischen Datenquellen kommen daher noch (*Vielfalt*) sekundäre Datenquellen, wie Selbstberichte zu Schmerzen hinzu. Erweitert wird das Ganze durch unterschiedliche Untersuchungsmodalitäten (z.B. unterschiedliche Gehgeschwindigkeiten, Laufen in der Ebene und in der Schräge, Treppensteigen) sowie die wachsende Relevanz von Felduntersuchungen, beispielsweise während alltäglicher Bewegungen [9]. Dies führt auch zu multimodalen und relationalen Datenbeständen.

Im Bereich der kinematischen Analyse sind markerbasierte 3D-Bewegungsanalysen als Goldstandard bekannt. Gerade in den letzten Jahren boomen jedoch andere Technologien, welche Einsatzmöglichkeiten im Feld zulassen und den Durchführungsaufwand reduzieren. Die Anwendung von IMUs bei der Analyse des menschlichen Ganges wächst so rapide [121]–[124]. Die Systeme ermöglichen es schnell und relativ leicht Bewegungen zu erfassen. Auch im klinischen Kontext haben daher die IMU Systeme in unterschiedlicher Form Einzug gefunden [57], [125]. Durch derartige Entwicklungen steigt die Geschwindigkeit mit der neue Daten entstehen dramatisch an (*Geschwindigkeit*). Besonders bei Felddaten, allerdings nicht nur darauf beschränkt, kommt der Einfluss von verrauschten, fehlerhaften oder unvollständigen Daten hinzu (*Genauigkeit*). Vielfältigen Fehlerquellen kommt hier eine Rolle zu (Weichteilartefakte, Ghost-Marker, inakkurate automatische Erkennung, usw.).

Theoretisch gesehen könnte der Wert (*Wert*)³ solcher biomechanischer Big-Data-Bestände, beispielsweise im Sinne der Verbesserung von Diagnosen oder der Identifikation neuer Korrelate, sehr hoch sein. [21] sagt aus, dass der wahre Wert erst noch durch Einsatz entsprechender Analysemethoden bewiesen werden muss. Gerade die ausgeführten Big Data Charakteristika heben daher noch einmal die Relevanz von ML Methoden zur Bewältigung dieser hervor.

Anwendungsbereiche von ML in Ganganwendungen stellen u.a. die Diagnose von Gangstörungen [36], [37], [126], [127], die Erkennung menschlicher Aktivitäten [128], altersbedingte Einschätzung [129]–[131] oder die Optimierung der Rehabilitationsphase [132] dar (für die Darstellung einzelner weiterer Arbeiten sei auf die einzelnen Publikationen verwiesen). Zur Einordnung der einzelnen Arbeiten in übergeordnete Einsatzfelder sollen nachfolgend zwei Strukturierungskonzepte vorgestellt werden.

[28] nimmt eine Kategorisierung der Einsatzschwerpunkte von ML im Bereich der Ganganalyse in Gangaktivitätserkennung, Gangereigniserkennung, Gangstörungserkennung (a) aufgrund von Pathologie und b) aufgrund von Alterung sowie Erkennung von Gangasymmetrie und neurologische Auswirkungen auf den Gang vor. Die meisten Studien lassen sich nach [28] im Bereich der Gangstörungserkennung aufgrund von Pathologie gefolgt von Gangereigniserkennung finden, wobei die Mehrzahl der Studien auf Sensordaten beruht. Arbeiten beispielsweise bezogen auf die Schätzung biomechanischer Daten (z.B. die Schätzung von Gelenkwinkeln und -momenten [133]) werden bei dieser Einteilung ausgelassen. Die Erkennung von Gangasymmetrien und die Gangstörungserkennung aufgrund von Pathologie nimmt in der Arbeit eine zentrale Rolle ein.

³ In der vorliegenden Schrift wird hier häufig von „Potential“ gesprochen

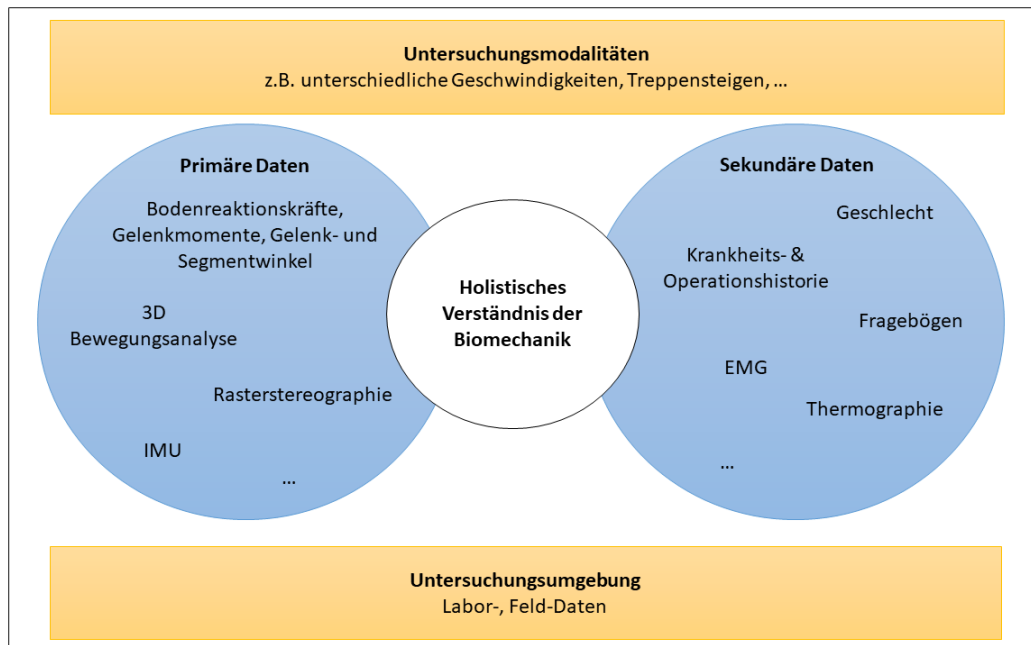


Abbildung 2.5: Datenquellen in der Biomechanik. Für ein holistisches Verständnis muss die Komplexität des Menschen beachtet werden und multimodale Daten berücksichtigt werden. Eigene Darstellung.

Eine Ausweitung der Kategorien erfolgt bei [20], welcher auf Basis eines aktuellen Reviews die folgenden auf den menschlichen Gang basierenden Anwendungen von KI charakterisiert, wobei nach eigenen Angaben die Kategorien teilweise überlappen und nicht klar abgrenzbar sind:

1. Verbesserung der Analysetechnik durch KI (optimale Feature Extraktion, automatische Eventdetektion)
2. Gesundheit und Wellness. Anwendungen im Bereich des Sports und der Gangüberwachung zur Identifikation abnormalem Gangs und menschlicher Aktivitäten sowie der Sturzerkennung
3. Human Pose Tracking
4. Biometrie im Bereich Personenidentifikation, Authentifizierung und Geschlechts-erkennung
5. Smart Gait (SG) Anwendungen (intelligente Kleidung und Wearables)
6. Animationen inklusive Gangmodellierung und Simulation

Entsprechend dieser Einteilung lässt sich der Schwerpunkt dieser Arbeit in den Bereichen 2. und 4. verorten.

Zusätzlich bzw. teilweise in Übereinstimmung mit den anfänglich herausgearbeiteten Defiziten lassen sich angelehnt an [20] als zentrale Herausforderungen beim Einsatz von KI bei Gangdaten das Black Box Problem (1), das Problem des Labelns der Daten (2) und die Lücke zwischen Forschung und Kommerzialisierung (3) charakterisieren. Für die bereits erfolgten Ausführungen zum Black Box Problem sei auf Sektion 2.1.4 sowie die Einleitung verwiesen.

*weitere
Defizite*

Das Problem des Labeln der Daten (2) bezieht sich auf mehrere Dimensionen. Manuelles Labeln der Daten durch den Menschen ist ein fehleranfälliger und äußerst zeitaufwendiger Prozess, insbesondere wenn große Datenbestände z.B. durch Messungen von Subjekten in alltäglichen Settings durchgeführt werden sollen [20]. Daher wurden automatische Techniken zum Datenlabeln mittels Deep Learning Ansätzen vorgestellt [134]. Für alternative Bewältigungsstrategien wie halb-überwachtes Lernen sei auf die Ausführungen im Resümee verwiesen.

Hinsichtlich der Lücke zwischen Forschung und Kommerzialisierung besteht die Hürde der Zulassung als Medizinprodukt inklusive der dafür anfallenden Kosten. Zwar wurden, wie im vorherigen Abschnitt angeführt, bereits generelle KI Anwendungen im medizinischen Kontext als Medizinprodukt zugelassen, jedoch lassen sich nach bestem Wissen des Autors keine zugelassenen Anwendungen im Bereich der Biomechanik finden.

Ein wichtiger Schritt für die Zulassung als Medizinprodukt stellt dabei die Offenlegung von Entscheidungsgrundlagen mittels XAI dar. Durch Fortschritte bei der Anwendung von erklärbaren Methoden der künstlichen Intelligenz (XAI) im biomechanischen (klinischen) Bereich zur Überwindung des Black Box Charakters wird ML in letzter Zeit immer mehr in praktischen Situationen anwendbar. Hier sind die Arbeiten von [39], [58], [101], [135] zu nennen, welche durch Einsatz von XAI Methoden auf Basis von Bodenreaktionskräften einen ersten wichtigen Beitrag zu Einsatzmöglichkeiten von XAI im biomechanischen Kontext geleistet haben. **XAI**

2.3 Vertiefende Messmethodische Grundlagen mit Fokus auf die Wirbelsäule

Diese Sektion beschäftigt sich mit messmethodischen Grundlagen, wobei der Fokus auf der Wirbelsäulenanalyse liegen soll. Dies begründet sich auf Basis der hohen Potentiale, welche nachfolgend zuerst dargestellt und herausgearbeitet werden sollen (Erläuterung der Relevanz der verwendeten Datensätze). Drei der aufgeführten Publikationen bauen so auf den Wirbelsäuledaten auf, wobei zwei dynamische Wirbelsäuledaten (Publikation 2 und 4) und zwei statische Daten (Publikation 2 und 5) verwendet werden. Anschließend sollen messmethodische Grundlagen der Analyse der Wirbelsäule angerissen und die Güte der verwendeten rasterstereographischen Messmethodik erläutert werden.⁴

2.3.1 Klinische und sozioökonomische Bedeutung

Der besondere Fokus auf die Wirbelsäuledaten begründet sich durch hohe gesellschaftliche und wirtschaftliche Relevanz. Bezugnehmend auf Wirbelsäuleprobleme sind exemplarisch Rückenschmerzen ein relevantes Gesundheitsproblem und verursachen allein in Deutschland jährlich Kosten in Milliardenhöhe. Die Volkskrankheit hat immense Auswirkung auf die Lebensqualität der Betroffenen und führt zugleich auch zu hohen volkswirtschaftlichen Folgekosten. Die Jahresprävalenz von Rückenschmerzen in der deutschen Bevölkerung liegt bei 61,3% [138]. Im Jahr 2016 waren muskuloskelettale Erkrankungen bei Frauen und Männern der zweit- bzw. **Ausgangslage**

⁴ Für die Gütekriterien zu den anderen, in den Publikationen verwendeten Datenquellen, sei bspw. für die markerbasierte 3D-Analyse auf [136] und für die IMUs auf [137] verwiesen.

dritthäufigste Grund (13,5% bzw. 12,7%) für Rentenneuzugänge wegen verminderter Erwerbsfähigkeit [139]. Ein Viertel aller Befragten, nach einer die deutsche Bevölkerung abbildenden Studie, waren wegen Rückenschmerzen im Jahre 2003 in ärztlicher Behandlung [140]. Bei einem Bevölkerungsstand von 83,2 Millionen in Deutschland im Jahr 2021 kann daher von ca. 20 Millionen ärztlich behandelten Rückenschmerzen-Patienten ausgegangen werden. Der aktuelle Trend zeigt, dass Rückenschmerzen, repräsentativ für die allgemeine Wirbelsäulengesundheit stehend, verstärkt negativ beeinflusst wird. Die meisten Fehltage DAK-Versicherter im ersten Halbjahr 2021 waren auf Grund von Rückenschmerzen und im Vergleich zum Vorjahreszeitraum konnte ein Anstieg von 7% festgestellt werden [141]. Es ist damit zu rechnen, dass dadurch erhebliche Mehrkosten im Gesundheitssystem entstehen werden, welche bedingt durch die Pandemiesituation noch verstärkt erscheinen, und die Lebensqualität bei einer wachsenden Anzahl an Betroffenen negativ beeinflusst wird.

Aufgrund der hohen Relevanz von Gesundheitsproblemen der Wirbelsäule und den humanitären und finanziellen negativen Folgeeffekten besteht Bedarf an einer möglichst optimalen Diagnose und Therapie von Krankheitsbildern der Wirbelsäule. In allen Fällen hängt allerdings die Diagnosestellung bzw. Befundung stark von den individuellen Erfahrungen des durchführenden Diagnostikers ab, was auch zu Fehldiagnosen führen kann. Eine durch die Techniker Krankenkasse durchgeführte Zweitbegutachtung von operativen Eingriffen der Wirbelsäule konnte so zeigen, dass womöglich für eine Vielzahl der Patienten auch konservative Therapieformen hätten geeignet sein können [142]. Diagnosestellungen basieren zudem oft auf Teilbetrachtungen, da multivariate Daten aufgrund ihrer Komplexität häufig eine direkte Interpretation durch den Menschen ausschließen. Hinzu kommt beispielsweise bei Rückenschmerzen, dass diese teilweise auf spezifische vertebrale Ursache zurückzuführen sind, wie Spondylitis, Bandscheibenvorfall oder Spinalkanalstenose [143]. Allerdings werden 90 % der Rückenschmerzen als unspezifisch angegeben, wobei kein anatomisches Korrelat als spezifische behandlungsbedürftige Ursache nachgewiesen werden kann [144]. Data Mining auf Basis eines umfassenden, multimodalen Datensatzes könnte mittels der entwickelten Methoden einen Beitrag zur Exploration und Identifikation potentieller (multivariater) Korrelate leisten.

Folgeeffekte

Auf Basis dieser klinischen und sozioökonomischen Bedeutung soll durch die vorliegende Arbeit ein Beitrag geleistet werden, perspektivisch Diagnosen zu optimieren und eine objektive, datenbasierte Einschätzung (Zweitmeinung) zu geben.⁵ Es sollen erste Schritte zur Objektivierung und Verbesserung der Wirbelsäule-Behandlungsmöglichkeiten durch Integration von ML zur Interpretation und Bewertung von statischen und dynamischen Wirbelsäuledaten aufgezeigt werden. Durch die KI-basierte Orientierung könnte es zu einer Vermeidung von Fehldiagnosen kommen. Dahingehend könnten Entdeckungsraten von Krankheitsbildern bzw. pathologischen Unterschieden (insbesondere in frühen Stadien oder bei funktionellen Dysfunktionen) verbessert werden, indem das Framework bereits kleine Auffälligkeiten identifizieren kann, noch bevor der Mensch diese wahrnimmt. Dadurch könnten weniger falsch-negative Diagnosen mittels Unterstützung durch KI generiert werden. Im Hinblick auf die vorherrschende Praxis könnte es weiterhin zu einer Verbesserung von Entscheidungen kommen indem individuelle pathologische Abweichungen / Einzigartigkeiten besonderen Fokus erhalten. Insgesamt könnte das auf KI-basis neu entstandene Wissen, verbunden mit bereits vorhandenem Expertenwissen, dabei

Beiträge

⁵ Für die studienbasierte Herleitung der nachfolgenden Punkte sei auf Sektion 2.2.2 verwiesen.

helfen neue Einsichten zu generieren und die klinische Praxis nachhaltig positiv beeinflussen.

2.3.2 Wirbelsäulenanalyse mit Fokus auf das DIERS Messsystem

Statische und dynamische Verfahren der Messung der Wirbelsäule werden grund-sätzlich in der Literatur beschrieben. Die Bestimmung von dynamischen Wirbelsäu-leparametern wie der ROM der Lendenwirbelsäule wird dabei eine hohe Bedeutung zur Ätiologie, Diagnose und Behandlung vieler Wirbelsäulenerkrankungen zuge-schrieben, wobei sowohl invasive als auch nichtinvasive Methoden für die in vivo Messung beschrieben werden [145].

*generelle
Messver-
fahren*

Zu den nichtinvasiven dynamischen Methoden zählen beispielsweise der modi-fizierte-modifizierte Schober Test (engl. Modified-Modified Schober Test), welcher moderate Validität und sehr gute Reliabilität zeigt [146] und der Fingerspitzen-zu-Boden Test, für den sehr gute Validität und Reliabilität berichtet wird [147]. Es muss jedoch angemerkt werden, dass externe Messungen möglicherweise nicht die tatsächliche intervertebrale Bewegung abbilden und die Aussagekraft auf Ebene einzelner Segmente bzw. Wirbelkörper stark eingeschränkt ist [145].

Zu den invasiven Methoden zählen bspw. in den Dornfortsatz eingebrachte Kirschnerdrähte [148], perkutane transpedikulär-externe Fixationsstifte [149], [150], die Verwendung der offenen interventionellen Magnetresonanztomographie [151] oder dynamische radiographische Messungen, beispielsweise mittels der häufiger verwendeten Superimposition oder Cobb-Methode, wobei selbst unter möglichst optimalen Bedingungen (gleicher Beobachter, gleiche Methode) ROM-Messungen der Lendenwirbelsäule in Sagittalebene Messfehler im Bereich $\pm 4^\circ$ aufweisen und unerfahrene Beobachter zu größeren Fehlern führen [145].

Einen alternativen Ansatz zur sowohl statischen als auch dynamischen Vermes-sung der Wirbelsäule bietet das DIERS 4D motion® Lab (DIERS International GmbH, Schlangenbad, Deutschland). Das System ermöglicht es die Wirbelsäule umfassend abzubilden (von Vertebra prominens bis L4 und Becken in allen 3 Ebe-nen). Durch die Weiterentwicklung des Systems mit der Herstellerfirma DIERS International GmbH wurde es möglich die messbaren Wirbelkörper und das Becken auch während des Gangs auf einem Laufband zu erfassen und zu analysieren. Das System arbeitet dabei ohne Strahlenbelastung was besonders vorteilhaft ist, da so wiederholte Scanvorgänge ohne unerwünschte Auswirkungen auf die menschliche Gesundheit möglich sind [152], [153]. Ferner kann es relativ zeiteffizient eingesetzt und potentielle Fehlerquellen reduziert werden, da nur wenige Marker gesetzt werden müssen. Zudem ist keine besonders aufwändige Schulung des Personals erforder-lich, da die Messungen relativ leicht durchzuführen sind. Außer der Notwendigkeit für dynamische Aufnahmen diese auf dem Laufband durchzuführen, werden die Testpersonen in ihrem natürlichem Gangverhalten bei den Messungen nicht direkt beeinflusst.

*DIERS
System*

Die lichtoptische Vermessung des Rückens mit dem System basiert auf (Video-) Rasterstereographie.⁶ Rasterstereographie ist dabei ein Verfahren der stereophoto-grammetrischen Oberflächenvermessung. Abbildung 2.6 stellt das System und den groben Ablauf einer Messung dar. Für eine Messung wird zunächst ein Lichtraster

⁶ Im Englischen werden in diesem Zusammenhang sowohl die Begriffe „surface topography“ als auch „rasterstereography“ verwendet.



Abbildung 2.6: DIERS 4D motion® Lab eingesetzt im MotionLab der Universitätsmedizin der Johannes Gutenberg-Universität Mainz. Abbildung basiert auf den freundlicherweise zur Verfügung gestellten Visualisierungsmedien der Firma DIERS International GmbH.

mittels Projektor auf den Rücken der Testperson geworfen und mit einer Kameraeinheit aufgezeichnet. Das Rasterdiapositiv ist dabei im Projektor eingebaut und bildet zusammen mit dem Kamerabild das Stereobildpaar. Anhand der Linienkrümmung wird entsprechend Photogrammetrie ein dreidimensionales Abbild der Oberflächentopographie des Rückens erzeugt. Die rekonstruierte Rückenoberfläche wird zunächst als 3D Punktkoordinaten (x, y, z) beschrieben, welche von Stellung der Person relativ zur Kamera abhängen. Daher werden aufbauend invariante Formparameter (konvex, konkav, sattelförmig) bestimmt, welche Unabhängig von der Relativlage im Raum sind. Die Aufstellung zur Kamera oder Drehung um die longitudinale Körperachse sollten daher keinen Einfluss auf die Ergebnisse haben [154]. Darauf aufbauend werden anatomische Strukturen (Vertebra prominens, Lumbalgrübchen) bestimmt auf denen rechnerisch weitere Punkte abgeleitet werden können (bspw. die Mitte zwischen den Lubargrübchen). Auf Basis dieser Informationen kann abschließend auf die Skelettgeometrie geschlossen werden. Dies erfolgt auf Basis eines Korrelationsmodells nach Turner-Smith [155] sowie Drerup [156], das die Abhängigkeit zwischen den Oberflächenkrümmungen des Körpers und der Ausrichtung der Wirbel beschreibt.

Generell wurden rasterstereographische Vermessungen der Wirbelsäule in einer Vielzahl von Arbeiten verwendet. Beispielsweise wurde von [152] ein Vergleich der stehenden und sitzenden Positionen mittels Rasterstereographie durchgeführt. [153] untersuchten die Wirbelsäulenverkrümmungen in der Sagittalebene beim Sitzen auf einem Stuhl ohne Rückenlehne und beim Sitzen auf einem Rehaball.

Auch das System der Firma DIERS wurde bereits in vielfältigen Forschungsarbeiten eingesetzt, wobei diese Arbeiten sich vorwiegend mit statischen Aufnahmen auseinandersetzten. Nur wenige Arbeiten haben sich mit dynamischen Parametern befasst. Statische Messungen wurden beispielsweise bei [157] verwendet, welche das

*Anwendun-
gen*

System zur Analyse unterschiedlicher Sitzpositionen eingesetzt haben. Korrelationen von unspezifischem Rückenschmerz der unteren Wirbelsäule mit der Wirbelsäulenkrümmung wurden von [158] mit dem System untersucht. [159] untersuchte Unterschiede zwischen beschwerdefreien Personen und Personen mit Beschwerden des unteren Rückens.

Neben den eigenen Arbeiten wurde das System beispielsweise bei dynamischen Messungen von [160] verwendet, welche Evaluation des Systems mit einem Goldstandardsystem der Bewegungsanalyse (VICON-System) unter statischen und dynamischen Bedingungen durchführten.

2.3.3 Quantitative Güte des Messsystems

Ergänzend zu den einzelnen Publikationen soll hier eine Erweiterung der Güteaspekte des Messverfahrens erfolgen. In der Theorie sollte eine verbesserte statistische Zuverlässigkeit durch den Einsatz des DIERS Messsystems möglich sein, da eine größere Zahl an Einzelmessungen möglich ist. Daher sollte der Einfluss zufälliger Schwankungen von beispielsweise tatsächlichen pathologischen Unterschieden besser unterschieden werden können [154].

Insgesamt wurde die Evaluation der Güte des Messsystems überwiegend für statische Messungen durchgeführt. Eine Metaanalyse konnte die Reliabilität und Validität der Messmethodik zur Beurteilung der Wirbelsäulenhaltung für globale Parameter (Skoliose, thorakale Kyphose, Lendenlordose) im Vergleich zu traditionellen radiographischen Untersuchungen zeigen. Die Validität wird teilweise als besser bei Skoliosepatienten als bei Gesunden berichtet, was aber auch an der geringeren Stichprobengröße liegen könnte [161]. Bei Patienten mit Skoliose konnte ferner sowohl präoperativ als auch nach chirurgischer Korrektur und Fusion gute Reliabilität und Validität, gemessen am Goldstandard der Radiographie, für die Methode gezeigt werden [162], [163]. Die Validität bei bestimmten Ätiologien mit einem hohen Body Mass Index (BMI) erscheint jedoch eingeschränkt [164]. Herstellerinterne Angaben deuten zudem an, dass sehr muskulöse Personen sowie das Vorliegen einer sehr starken Skoliose die Schätzung der Wirbelkörperpositionen stark negativ beeinträchtigen. Zudem sei darauf hingewiesen, dass die Berechnung des Wirbelkörpermodells bei Personen, die beispielsweise einen 6. Lendenwirbel aufweisen (Lumbalisation [165]), scheitern.

Statik

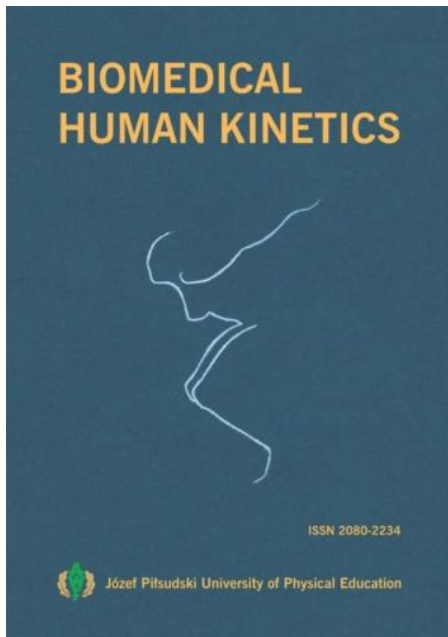
Im Hinblick auf dynamische Messungen führte [160] eine Evaluation des Systems mit einem Goldstandardsystem (VICON) durch und konnte gute Übereinstimmungen zwischen dem System für statische und dynamische Parameter (Markerdetektion und Seitenlängen definierter Dreiecke) feststellen. Subsumierend muss jedoch festgehalten werden, dass insbesondere die Empirie zur Validität auf Ebene der einzelnen Wirbelkörper in Dynamik eingeschränkt ist. Messmethodisch erscheint die ganzheitliche Validierung der Wirbelsäule auf Ebene einzelner Wirbelkörper in Dynamik problematisch, da es keinen Goldstandard gibt, um Erkenntnisse und Annahmen bezogen auf alle gemessenen Wirbelkörperparameter zu validieren. Zumindest können hohe Reliabilität für den Einsatz in der dynamischen Ganganalyse nachgewiesen werden [166]. Noch unveröffentlichte Ergebnisse des MotionLab der Universitätsmedizin der Johannes Gutenberg-Universität Mainz sprechen zudem für sehr gute Reliabilität bei Betrachtung dynamischer Wirbelsäuledaten auf Ebene der einzelnen Wirbelkörper. Inwieweit sich die Ergebnisse auf verschiedene Messmethoden verallgemeinern lassen, ist gegenwärtig nach bestem Wissen des Autors unbekannt.

Dynamik

KAPITEL 3

Die Publikationen

3.1 1. Publikation



Feature extraction and gait classification in hip replacement patients on the basis of kinematic waveform data

C. Dindorf, W. Teufl, B. Taetz u. a., “Feature extraction and gait classification in hip replacement patients on the basis of kinematic waveform data,” *Biomedical Human Kinetics*, Jg. 13, Nr. 1, S. 177–186, 2021. DOI: 10.2478/bhk-2021-0022

Zusammenfassung

Hintergrund: Postoperative asymmetrische Gangmuster nach Hüft-TEP sind ein potenzieller Risikofaktor. Die Klassifizierung der operierten Hüfte könnte Informationen über diskriminierende Variablen liefern, die für die Beschreibung des asymmetrischen Gangs am wichtigsten sind und ist daher von klinischer Relevanz. Vergleiche zwischen verschiedenen Klassifizierungs- und Feature Extraktionsansätzen in diesem Zusammenhang sind defizitär. Daher wird in dieser Arbeit die Feature Extraktion auf Basis von Gangwellenformdaten von TEP-Patienten angewendet, um, ohne sich auf gangspezifische Annahmen oder Vorkenntnisse zu stützen und möglichst wenig Informationen a priori zu verwerfen, herauszufinden, welche Parameter für die Beschreibung des asymmetrischen Gangs bei Patienten nach Hüft-TEP am wichtigsten sind. Verschiedene Klassifikations- und Featureextraktionsansätze werden verglichen.

Methodische Forschungsfragen: Ist die Feature Extraktion ausgehend von einer breiten Datenbasis von Wellenformdaten möglich? Führt Feature Selektion oder die Reduzierung der Dimensionalität zu besseren Ergebnissen? Können Ensemble Feature Selektionsmethoden in dem Kontext erfolgreich eingesetzt werden, um die Stabilität zu erhöhen?

Fachspezifische Forschungsfragen: Ist die Klassifizierung der operierten Seite nach TEP anhand von Gangwellenformdaten möglich? Welche globalen Parameter sind für die Beschreibung des asymmetrischen Gangs bei Patienten nach Hüft-TEP am bedeutendsten?

Methodik: Der Gang von 22 Patienten nach Hüft-TEP wurde mit einem optischen Bewegungserfassungssystem aufgezeichnet. Als Eingangsmerkmale wurden die Wellenformdaten der Markerpositionen, Geschwindigkeiten und Beschleunigungen sowie Gelenk- und Segmentwinkel verwendet. Für die Feature Selektion wurden die Algorithmen RF und mRMR gewählt. Stabilität wurde durch Ensemble Feature Selektion adressiert. Die Ergebnisse wurden mit denen verglichen, die durch die Verwendung verschiedener Methoden zur Reduktion der Dimensionalität erhalten wurden.

Ergebnisse: Hüftbewegung in der Sagittalebene, Kniekinematik in Frontal- und Sagittalebene, Markerpositionsdaten der Spinae iliacae anteriores und posteriores superiores sowie Beschleunigungsdaten für Marker am proximalen Ende der Fibula sind für die Klassifizierung von großer Bedeutung (Genauigkeit: 91,09 %). Mit der Merkmalsauswahl wurden im Vergleich zur Dimensionsreduktion bessere Ergebnisse erzielt. Die Stabilität konnte durch den verwendeten Ansatz erhöht werden.

Fazit: Die vorgeschlagenen Ansätze können verwendet werden, um anormale Gangmuster während des Rehabilitationsprozesses über Wellenformdaten zu identifizieren und individuell zu adressieren. Die Ergebnisse zeigen, dass Positions- und Beschleunigungsdaten auch für diese Aufgabe wichtige Informationen enthalten.

Feature extraction and gait classification in hip replacement patients on the basis of kinematic waveform data

Carlo Dindorf¹, Wolfgang Teuffl², Bertram Taetz³, Stephan Becker¹, Gabriele Bleser⁴, Michael Fröhlich¹

¹ Department of Sports Science, Technische Universität Kaiserslautern, Kaiserslautern, Germany; ² IFFB Sport and Movement Science, Paris Lodron University Salzburg, Austria; ³ Department Augmented Vision, German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany; ⁴ Junior Research Group wearHEALTH, Technische Universität Kaiserslautern, Kaiserslautern, Germany

Abstract

Study aim: To find out, without relying on gait-specific assumptions or prior knowledge, which parameters are most important for the description of asymmetrical gait in patients after total hip arthroplasty (THA).

Material and methods: The gait of 22 patients after THA was recorded using an optical motion capture system. The waveform data of the marker positions, velocities, and accelerations, as well as joint and segment angles, were used as initial features. The random forest (RF) and minimum-redundancy maximum-relevance (mRMR) algorithms were chosen for feature selection. The results were compared with those obtained from the use of different dimensionality reduction methods.

Results: Hip movement in the sagittal plane, knee kinematics in the frontal and sagittal planes, marker position data of the anterior and posterior superior iliac spine, and acceleration data for markers placed at the proximal end of the fibula are highly important for classification (accuracy: 91.09%). With feature selection, better results were obtained compared to dimensionality reduction.

Conclusion: The proposed approaches can be used to identify and individually address abnormal gait patterns during the rehabilitation process via waveform data. The results indicate that position and acceleration data also provide significant information for this task.

Keywords: Classification – Total hip arthroplasty – Feature selection – Dimensionality reduction

Introduction

The total hip arthroplasty (THA) is the most important surgery for the treatment of degenerative hip osteoarthritis [26]. Postoperative changes in gait can not only affect the operated joint and the surrounding structures, but also influence the contralateral side. In addition, postoperative gait patterns often display an asymmetric character [2, 17]. The possible consequences include increased joint loadings, which could lead to injuries and a failure of the implant, resulting in the need for further surgical interventions [21].

Gait classification is an important tool for clinical diagnostic and illness identification [13, 14]. The classification of the operated hip side could deliver knowledge about the best discriminative variables and is, therefore, of

clinical relevance. With modern movement tracking systems, a huge amount of data are available (big data) [30], and machine learning models have gained importance compared to classical statistical approaches [6]. Feature selection and dimensionality reduction are important steps toward improving a model's accuracy and interpretability, preventing overfitting, and reducing the necessary computing power [25].

Joint angles at specific points in time and descriptive statistics, such as the total range of motion or peak values, are commonly used for classification [3]. However, it is questionable whether such features optimally map group differences or if meaningful information is discarded a priori. For example, acceleration data have been given relatively little consideration, although it has been shown that they map important information about gait patterns [16]. The dependence on prior knowledge, assumptions,

or subjective decisions is a further limitation. The use of entire gait waveforms that have been transformed using principal component analysis (PCA) could be a potential alternative [13, 30]. Comparative studies regarding different dimensionality reduction methods and classification algorithms have rarely been reported [14]. To the best of our knowledge, the use of feature selection on gait waveform data and classification in the context of THA (for classification of THA patients and healthy subjects, see [12, 34]) has hardly been investigated so far. Therefore, this work applies feature selection to gait waveform data of THA patients to objectively deduce discriminative variables (including discriminative time points in a gait cycle), which are most important for the description of asymmetrical gait. In order to do this, patients are classified according to the operation side after unilateral THA. The classification results were compared with those obtained through the use of different dimensionality reduction methods for feature extraction.

Materials and methods

Subjects and data acquisition

Motion data were collected from 22 subjects (P1–P22; 7 males, 15 females; age: 56.90 ± 8.20 ; height: 1.74 ± 0.08 ; weight: 82.90 ± 18.85) after THA (11 left side, 11 right side). The adequate sample size was derived from previous gait classification studies [3, 10]. All subjects were recruited from among patients of the Klinik Lindenplatz (Bad Sassendorf, NRW, Germany). All measurements were conducted in the biomechanics laboratory of the institute of biomechanics of the Klinik Lindenplatz.

Patients were only considered fit to participate in the study if they were at least 14 days, but not more than 21 days, after surgery. The subjects had to be able to walk steadily for at least six minutes without support on even ground to be included in the examination. All included patients received a standard cemented THA using a direct anterior or lateral approach [36]. The patients were allowed to bear their full weight by their attending physicians at the time of measurement.

After receiving all relevant study information, the patients signed an a form indicating their informed consent to the study, which included permission to publish the data. The study was approved by the responsible ethics commission and meets the criteria of the Declaration of Helsinki.

The gait data of the lower bodies of the THA patients were recorded during a 6 minute walking test using an OptiTrack® stereophotogrammetric motion capture system (NaturalPoint, Inc., Oregon, USA). The accuracy of the OptiTrack motion capture system was shown in recent publications [9, 35].

The subjects were equipped with retroreflective markers attached to anatomical landmarks according to the recommendations of Leardini et al. [23]. For hygienic reasons, the subjects had to wear shoes. Therefore, the markers corresponding to the foot had to be positioned on the surface of the shoe, approximating the underlying anatomical landmarks. The 3D positions of the markers during movement were recorded with a frame rate of 60 Hz. The walking direction was along the x-axis according to the defined global coordinate system, with the z-axis drawn vertically.

Data preparation

The following steps described below were performed separately for each subject: In addition to marker position data, the marker velocity and acceleration, as well as joint and segment angles, were calculated to check if those variables have higher importance. Missing data points were interpolated using spline interpolation (maximum gap = 5 frames). Noise attenuation was performed using a low-pass Butterworth filter (cutoff frequency = 6 Hz). With the marker position data, the marker velocity and acceleration were calculated. The joint and segment angles were calculated using Visual3D (V6 professional, Germantown, Maryland, USA).

The turning movement at the end of the gait path as well as steps form gait initiation and stopping were removed from the data. For direct comparison of the operated and non-operated legs, the raw data were split into gait cycles for each leg separately. The beginning of a gait cycle was defined as the point of the heel marker (LCA for the left body side, RCA for the right body side; all patients started contact for both legs with the heel) with the lowest z-axis position (initial contact). Incomplete cycles were dropped. Marker trajectories were rotated so that all movements followed one direction along the x-axis. Trajectories along the y-axis were inverted for the right body side so that they fit to the trajectories of the contralateral side. The mean position for every determined cycle determined along each marker and its three axes was subtracted from the respective sequences for centering the position data and for elimination of anthropometric differences between subjects (see, e.g., [13]). Afterwards, the respective variables of each gait cycle were individually time-normalized to 101 time steps (from 0% to 100%) using cubic spline interpolation. The isolation forest algorithm was used for outlier removal. Outliers are defined as minority of all instances with very different feature values compared to normal instances. With the algorithm, the respective instances are isolated and not detected by using distance or density measurements [24]. In the current case, most of the detected outlying cycles showed very different gait patterns compared to the majority of cycles (e.g., due to visually visible compensatory movements to maintain the balance or abnormally shortened swing phases due to early contact

with the ground). Asymmetric sequences were calculated by subtracting the gait cycles (separately for each of the x, y and z movement directions of the marker trajectories, the marker velocities and accelerations as well as the joint and segment angles in each plane) of the right body side from those of the left. The corresponding sequences are denoted as a w-feature in the following. The resulting sequences were concatenated into a vector of 8,383 dimensions, where each time step maps one dimension (Table 1). The dimensions were achieved through:

- x, y and z – movement directions of the 13 markers ($3 \times 13 \times 101$ time steps = 3939);
- marker-velocities as well as – accelerations of the 13 markers ($2 \times 13 \times 101$ time steps = 2626);
- joint and segment angles in three planes of the six joint and segment angles ($3 \times 6 \times 101$ time steps = 1818).

Finally, 50 random samples (without replacement) were extracted from 70 to 120 gait cycles for every subject, resulting in a final vector with a size of $1,100 \times 8,383$ (cycles \times features), where one sample represents the

difference between the left and right side during one gait cycle.

Feature extraction and classification

Feature extraction and classification were integrated into a leave-one-group-out cross-validation (LOGO, 11 folds) process to obtain an unbiased accuracy score (test data were not used for feature extraction) and to check if classification can individual differences take into account. The basic principal behind LOGO is that for each hold-out group, model fitting is performed, using the data of all the subjects except the hold out, which is only used for model evaluation [37]. A group of two subjects was formed – one on whom the operation was on the right side, and the other on the left side. Each resulting fold consists of 1000 training samples of 20 patients and 100 test samples of 2 patients. The initial input features were independently standardized (z-transform) based on the respective training sets before application of the different extraction methods. The methods are described below.

Table 1. Initial input features. The naming convention of the markers is according to [8]

	Location/Description	Name	Type	Steps
Marker	anterior superior iliac spine	ASIS		
	posterior superior iliac spine	PSIS		
	lateral prominence of greater trochanter	GT		
	midline of the thigh	TH		
	lateral prominence of lateral femoral epicondyle	LE		
	proximal tip of the head of the fibula	HF	x-axis pos. (X)	
	anterior border tibial tuberosity	TT	y-axis pos. (Y)	
	midline of the shin	SK	z-axis pos. (Z)	
	lateral prominence of the lateral malleolus	LM	velocity (v) acceleration (a)	
	Achilles tendon insertion	CA		$\times 101$ time steps
	dorsal margin of the fifth metatarsal head	VM		
	dorsal margin of the first metatarsal head	FM		
	distal phalanx hallux	DPI		
Joint angle	ankle angle	ANKLE_ANGLE		
	hip angle	HIP_ANGLE	flexion and extension in sagittal plane (X)	
	knee angle	KNEE_ANGLE	abduction and adduction in frontal plane (Y)	
Segment angle	foot angle	FOOT_ANGLE	axial rotation in transverse plane (Z)	
	shank angle	SHANK_ANGLE		
	thigh angle	THIGH_ANGLE		

Terminology: Feature labels are composed of marker/joint name _ type _ frame number, e.g., VM_a_46

Two different ranking approaches were applied to find relevant features for classification: a) random forest (RF) feature ranking and b) a minimum-redundancy maximum-relevance (mRMR) filter.

a) Random forest is a powerful algorithm for both regression and classification tasks. It is based on the combination of many weak learners to create one strong learner (ensemble method) [5]. The algorithm enables the calculation of a feature importance (FI) score and was applied in the following with 100 trees. Gini importance – normalized so that the total sum for all features used is one – was used as an importance score.

The following procedure was repeated until no features were left: Features were ranked with the RF algorithm according to their median feature importance based on 10 runs on bootstrap samples (80% of the dataset size). The feature with the highest importance was selected and ranked according to the time at which it was included. Afterward, features related to the selected one (a whole w-feature corresponding to the selected feature), as well as features with zero FI, were excluded. With the reduced feature set, the process was repeated. The proposed approach should be able to reduce redundancy by only selecting one point in time (the most important feature) per w-feature.

Variations in the training data often lead to different feature preferences. The robustness to these changes is described as stability [20]. Ensemble feature selection is reported as a promising approach for reducing instability [1]. The basic principal is that a combination of different feature selection processes leads to increased stability [1]. Therefore, feature ranking was performed with the proposed approach on different data samples (homogeneous feature selection) of each training fold during the LOGO process. Instead of bootstrap samples, which are commonly used for data variation, data of one group were excluded for each iteration. Due to the small number of subjects, this should lead to highly different data samples. Afterward, the different ranking lists were aggregated into one ensemble ranking. Minimum union (Min) aggregation was used to combine the ranking lists. Based on the ensemble rankings, final feature subsets were obtained with the following cut-off thresholds: 1%, 5%, 10%, 20%, 50%, and 100%.

b) The proposed RF approach does not capture redundancy between different w-features. In addition, redundancy increases through the aggregation step [32]. To obtain a mutual information-based ranking of the most relevant and minimally redundant features, an mRMR filter [28] was applied to reduce the impact of redundant features. Therefore, the features of the final ensemble list of each fold were ranked according to the mRMR filter. The classification results were then compared with the results of the proposed RF approach.

Finally, the results obtained using feature selection were compared with the following dimensionality

reduction methods: PCA, kernel PCA (kPCA), factor analysis (FA), and linear discriminant analysis (LDA). Due to the lack of comparative studies, various dimensionalities of input features and combinations with the support vector classifier (SVC) and RF were considered. Due to collinearity, PCA (99% variance threshold to select the number of components) was applied before LDA. The evaluation was performed using LOGO.

Classification algorithm and further calculations

The support vector classifier (SVC) is a commonly used algorithm for gait classification [22, 34], and was applied with a linear kernel in the current study. Stability was evaluated using the mean Jaccard similarity of the pairwise comparison between the different subsets and threshold values during the LOGO procedure (see e.g. [7]). The more similar the different subsets are, the higher the stability of the selection approach is. Calculations were made with the Python scikit-learn package (machine learning, data mining) [27].

Results

Feature selection

The classification results for the feature selection with different threshold values and the use of the SVC with a linear kernel are presented in Table 2. The best classification performance was obtained without the application of a threshold value (threshold = 100%). The corresponding subsets also showed the highest stability. Using other classification algorithms with the features of the 100% threshold also led to high classification performance (SVC with a radial basis function (rbf) kernel: $87.64 \pm 18.81\%$, $M_{AUC} = 0.96 \pm 0.07$; logistic regression: $90.45 \pm 16.78\%$, $M_{AUC} = 0.91 \pm 0.25$). In comparison with the use of all initial input features, higher accuracy was found for the use of the 100% and 50% thresholds. The receiver operating characteristic (ROC) curve for the classification using the 100% threshold value is presented in Figure 1a. The classification accuracy for one test set – consisting of P4 and P7 – was below average ($M_{acc} = 44.00\%$). Further analysis showed that the incorrect classification was mainly relate to P4, where the wrong operation side was predicted for all cycles. For the remaining test sets, the classification accuracy was between 87.00% and 100.00% ($M_{acc} = 95.80 \pm 4.58\%$).

The occurrence of the ten highest ranked features of every fold is presented in Figure 1b. Overall, ten features occurred in the top ten rankings in ten out of the eleven folds (from HIP_ANGLE_X_54 to HF_a_80). The feature HIP_ANGLE_X_54 occurred in every fold in the top ten rankings. The correlations of those features are presented in Figure 1c. High to very high correlations were present (mean correlation = 0.63 ± 0.23).

Gait classification in hip replacement patients

181

Table 2. Model performance as mean accuracy (M_{acc}) and mean area under the curve (M_{AUC}) for the different approaches evaluated during leave-one-group-out cross-validation (LOGO). The 100% threshold refers to the use of all selected features. The true positives and true negatives are highlighted with a gray background color in the confusion matrix (CM). Similarity is reported as the mean Jaccard similarity (M_{Jacc}): 0 = no similarity, 1 = completely similar

Threshold	n feature	RF				mRMR					
		Macc [%]	MAUC	MJacc	CM	Macc [%]	MAUC	MJacc	CM		
1%	3.00 ±	87.36 ±	0.87 ±	0.64 ±	474	76	77.36 ±	0.79 ±	0.21 ±	456	94
	0.00	0.20	0.26	0.31	63	487	25.51	0.31	0.13	155	395
5%	15.82 ±	83.27 ±	0.82 ±	0.78 ±	453	97	76.73 ±	0.83 ±	0.23 ±	397	153
	0.40	24.27	0.36	0.24	87	463	25.80	0.30	0.08	103	447
10%	31.91 ±	83.82 ±	0.87 ±	0.78 ±	428	122	86.27 ±	0.89 ±	0.34 ±	461	89
	0.54	22.34	0.25	0.13	56	494	16.73	0.28	0.06	62	488
20%	64.55 ±	85.55 ±	0.90 ±	0.84 ±	446	104	88.73 ±	0.92 ±	0.46 ±	492	58
	1.13	18.83	0.19	0.06	55	495	15.55	0.21	0.05	66	484
50%	162.36 ±	89.55 ±	0.91 ±	0.85 ±	494	56	90.73 ±	0.92 ±	0.61 ±	512	38
	2.98	15.02	0.22	0.04	59	491	15.57	0.23	0.04	64	486
100%	325.45 ±	91.09 ±	0.91 ±	0.87 ±	511	39					
	5.04	16.28	0.25	0.02	59	491					
All initial features	8383	88.55 ±	0.91 ±	–	486	64					
		18.13	0.25	–	62	488					

The rankings of all features selected by RF with the mRMR filter are presented in Figure 1d. The ten highest ranked features showed little correlation (mean correlation = 0.35 ± 0.34). Using the mRMR ranking for classification with the different thresholds resulted in lower classification performance compared to that of the RF approach for the 1% and 5% thresholds. For the 10%, 20%, and 50% thresholds, the classification performance increased.

The gait waveforms of the highest ranked features and the selected discriminative time points are shown in Figure 2. Different gait patterns are prevalent for the operated and non-operated sides. Highly overlapping areas are noticeable.

Dimensionality reduction

The results for the different dimensionality reduction techniques are presented in Figure 3. Only slightly higher accuracy compared to the use of all initial input features could be reached by using LDA after applying PCA ($M_{acc} = 89.00 \pm 15.85\%$, $M_{AUC} = 0.93 \pm 0.18$).

Discussion

The current study demonstrates that the data-driven determination of parameters, which are most important for the description of asymmetrical gait in patients after THA, is possible. The basis for this is the classification of

THA patients according to the operation side. The operated side was predicted with high accuracy (91.09%) using the features selected with the proposed approach and a linear SVC. A higher accuracy was obtained compared to modeling with all initial input features (88.55%). The elimination of noisy, meaningless, and redundant variables could be a reason for the difference, which underlines the quality of the selected predictors. Using different classification algorithms (SVC with an rbf kernel, logistic regression) also leads to good classification results. Biases towards a specific classification algorithm were, therefore, low, which highlights the generalizability of the selected features.

Using 325.45 ± 5.04 features led to the best classification results. However, for practical applications, it should be decided if a less complex model could be sufficient. The use of only a few input variables could be especially beneficial for easier interpretation and, therefore, enhance the relevance in clinical contexts. In this regard, the results show that 87.36% of all cycles could be correctly classified with only three features.

Then, the predominantly occurring features in the top ten rankings were interpreted and validated using domain knowledge. Due to the initial centering of the marker position data, the direct interpretation through a comparison between the operated and non-operated sides for the marker-position-based features PSIS_Z_97, PSIS_Z_96, and ASIS_Z_14 is misleading.

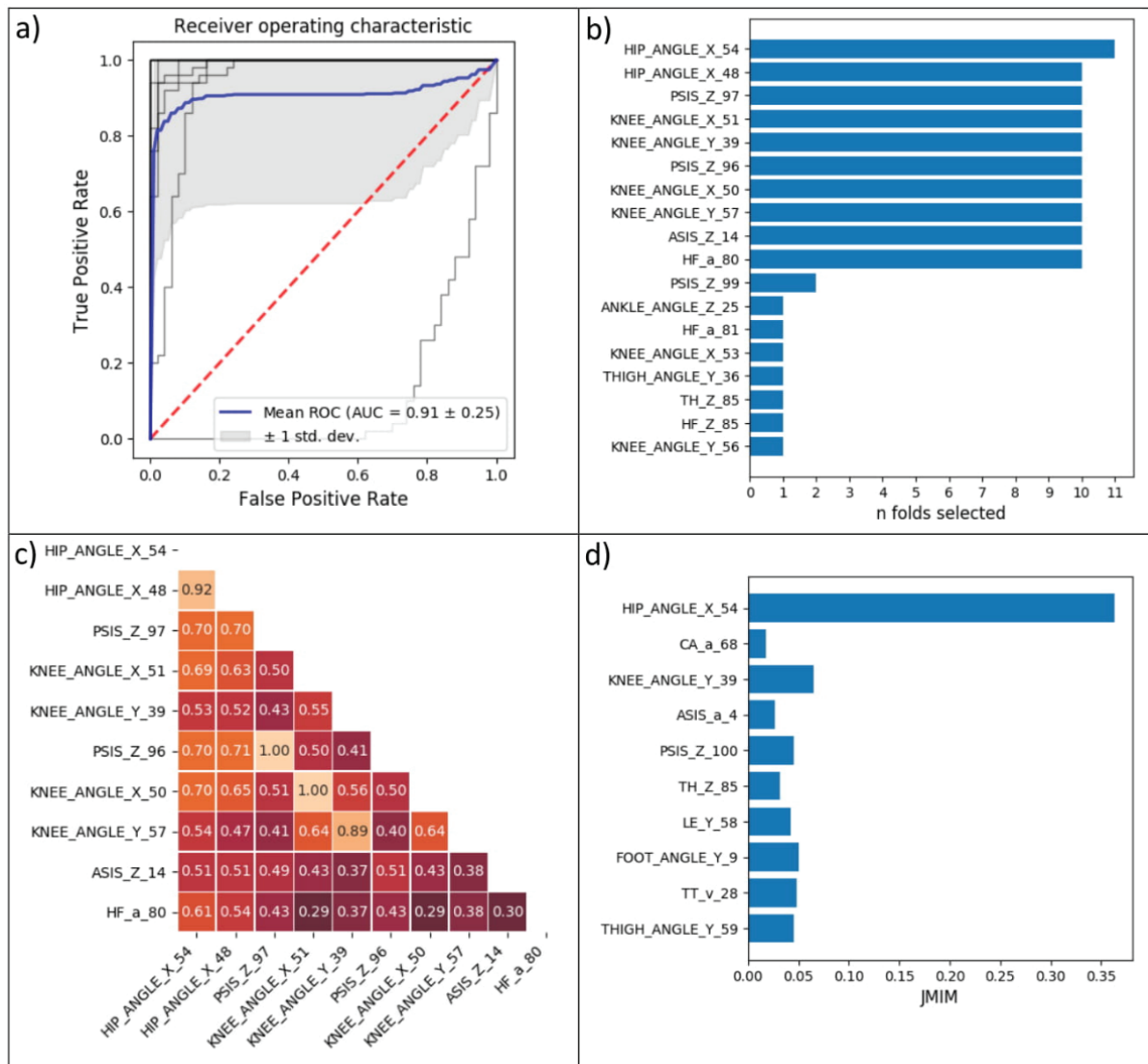


Figure 1. a) Receiver operating characteristic for each fold (gray lines) and for the mean (blue line). b) Numbers of times that features occurred in the top ten rankings of the LOGO folds (eleven folds in total). c) Correlation head map of the ten most stable features selected. d) Ranking of all selected features with the minimum-redundancy maximum-relevance filter, starting with the highest ranked feature at the top. The respective joint mutual information maximization (JMIM) scores [4] are plotted on the x-axis. Labelling: for the features representing joint and segment angles, X refers to movement in sagittal plane, Y to movement in frontal plane, and Z to movement in transverse plane

HIP_ANGLE_X_54 and HIP_ANGLE_X_48 highlight the importance of hip movement in the sagittal plane. HIP_ANGLE_X_54 was selected in the top ten rankings of every fold, and was also ranked as the feature with the highest relevance by the mRMR algorithm, which additionally highlights the discriminative power of the feature. Indirectly, due to the reduction of the minimal value for hip flexion/extension, a lower range of motion is relevant for classification. The effect has also been mentioned in previous studies [11, 29].

KNEE_ANGLE_X_51 and KNEE_ANGLE_X_50 represent knee movement in the sagittal plane. The knee shows increased flexion during the late stance phase. The predictors KNEE_ANGLE_Y_39 and KNEE_ANGLE_Y_57 show that, for class discrimination, knee abduction and adduction are important (movement in frontal plane). An altered varus has been reported in the literature [11]. The data revealed a bilateral valgus (abduction) for most patients, which was significantly increased for the operated side. A possible reason for the lateral difference

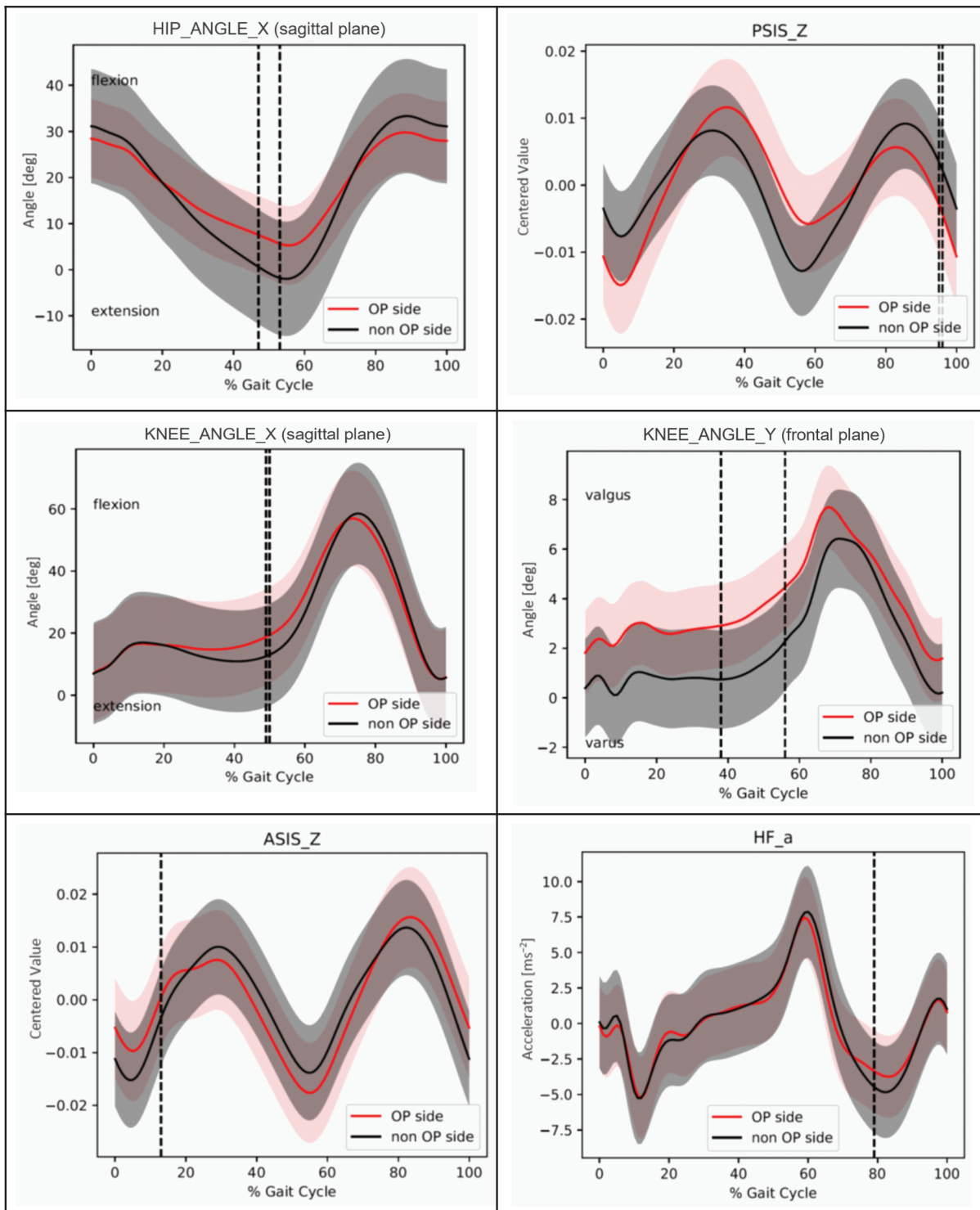


Figure 2. Asymmetric gait patterns (mean and standard deviation) in total hip arthroplasty (THA) patients. The points in time marked with a vertical line represent the most stable and highest ranked features. Note: Marker position data are centered

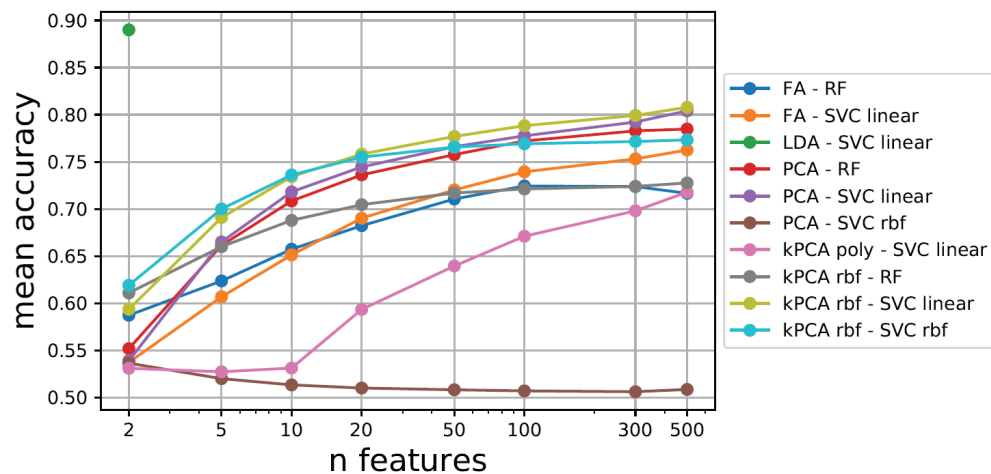


Figure 3. Model performance after dimensionality reduction; comparison of different techniques and different numbers of input features. Algorithms: FA = factor analysis, LDA = linear discriminant analysis, PCA = principal component analysis, kPCA = kernel PCA, RF = random forest, SVC = support vector classifier; kernels: poly = polynomial, rbf = radial basis function

is the reduced hip muscle strength on the operated side [31]. Furthermore, a general limitation of postoperative gait analysis is the potential influence of an abnormal pre-operative status, which has to be kept in mind as a possible bias. Moreover, the use of optical markers depends on the experience of the investigator and is therefore not free of errors when being placed on the anatomical landmarks through palpation [18]. Therefore, future studies should aim to verify these results with a higher sample size. Looking at feature HF_a_80, altered acceleration for the operated side is noticeable. Detailed biomechanical backgrounds for the marker-position – and acceleration-based features are not present and should be evaluated in the future. The rankings with the proposed RF approach as well as the rankings with the mRMR algorithm indicate that not only classical features, like joint angles, but also alternatives, like marker-position – or acceleration-data-based features, carry large amounts of information and show discriminative power.

The consideration above focuses on average characteristics, but individual patterns sometimes diverge (e.g., increased varus in patient 15). In line with the work of Chopra and Kaufman [11], the current study indicates a high variability in gait patterns (also see standard deviation) and possible subject-related differences and adaptations in patients after THA. Furthermore, the results for the single-subject analysis show that for P4, the wrong side was predicted as the operated side. The single-case analysis for P4 points out that when several variables were used as input features, altered inverted gait characteristics (e.g., for HIP_ANGLE_X and KNEE_ANGLE_Y) were prevalent compared to those of the remaining subjects. Therefore,

application of the classification algorithm could also be useful in identifying and individually addressing abnormal gait patterns in groups of patients after THA.

Several investigations demonstrated the potential of dimensionality reduction before gait classification [13, 14, 19]. In the present case, the combination of PCA and LDA showed similar classification results compared to the use of all initial inputs. The other methods applied showed considerably lower performance. However, compared to dimensionality reduction, better results were obtained with feature selection. Another advantage of the selection approach is that features can be immediately interpreted. The interpretation of principal components would require an additional interpretation step to analyze the components themselves. Nevertheless, dimensionality reduction allows the determination of dependencies between different parameters [30], which may be additionally relevant in clinical contexts. If only classification is important, application of the present combination of PCA and LDA seems especially promising because it is independent of a specification of a number of features to use.

Limitations arise from each patient wearing his or her own shoes instead of walking barefoot during gait analysis. It was shown, that different types of shoes influence the gait [33] and therefore introduce a bias during gait analysis. Moreover, placement of markers on the shoes instead of anatomical landmarks as well as relative motion of the shoes in respect to the feet might have influenced the study. Barefoot gait analysis should be considered for future analysis. To further evaluate the results, future research should also focus on a replication using different and increased numbers of subjects. The additional use of

kinetic data, as well as consideration of the sex – due to sex-specific differences in gait recovery after THA [15] – should also be analyzed. High variability in gait patterns and highly overlapping areas between the groups for the determined predictors may be limiting factors for accuracy. The development of an optimal experimental design that better maps group differences may be useful for increasing classification performance. With the use of waveforms as a starting point for feature selection, promising results could be obtained. However, further feature engineering could be profitable for an improvement of performance and should, therefore, be evaluated in future works.

Conclusions

The aim of the study was to find out, without relying on gait-specific assumptions or prior knowledge, which parameters are most important for the description of asymmetrical gait in patients after THA. In order to do this, meaningful features for classification of the side of the body that received a hip replacement were extracted from gait waveform data using a data-driven approach. The proposed RF approach seems promising for finding relevant features and their corresponding time points on the basis of high-dimensionality waveform data. This might be useful in identifying and individually addressing abnormal gait patterns during the rehabilitation process. For the present THA patients, the importance of features that represent hip movement in the sagittal plane and knee kinematics in the frontal and sagittal planes is emphasized. The use of marker position data of the anterior and posterior superior iliac spine, as well as acceleration data for markers placed at the proximal end of the fibula, is of further importance. The results indicate that studies should not only focus on variables like joint and segment angles because position and acceleration data also provide significant information.

Ethics Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Universität Paderborn. Informed consent was obtained from all subjects involved in the study.

Conflict of interest: Authors state no conflict of interest.

References

1. Afef B.B., Limam M. (2018) Ensemble feature selection for high dimensional data: a new method and a comparative study. *Adv. Data Anal. Classif.* 12(4): 937–952.
2. Beaulieu M.L., Lamontagne M., Beaulé P.E. (2010) Low-er limb biomechanics during gait do not return to normal following total hip arthroplasty. *Gait & Posture* 32(2): 269–273.
3. Begg R., Kamruzzaman J. (2005) A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *J. Biomech.*, 38(3): 401–408.
4. Bennasar M., Hicks Y., Setchi R. (2015) Feature selection using Joint Mutual Information Maximisation. *Expert Syst. Appl.*, 42(22): 8520–8532.
5. Breiman L. (2001) Random Forests. *Machine Learning* 45: 5–32.
6. Bzdok D., Altman N., Krzywinski M. (2018) Statistics versus machine learning. *Nature Methods* 15(4): 233–234.
7. Cannas L.M., Dessi N., Pes B. (2013) Assessing similarity of feature selection techniques in high-dimensional domains. *Pattern Recognit. Lett.*, 34(12): 1446–1453.
8. Cappozzo A., Catani F., Della Croce U., Leardini A. (1995) Position and orientation in space of bones during movement: anatomical frame definition and determination. *Clin. Biomech.*, (Bristol, Avon) 10(4): 171–178.
9. Carse B., Meadows B., Bowers R., Rowe P. (2013) Affordable clinical gait analysis: an assessment of the marker tracking accuracy of a new low-cost optical 3D motion analysis system. *Physiother.*, 99(4): 347–351.
10. Chan H., Yang M., Wang H., Zheng H., McClean S., Sterritt R., Mayagoitia R.E. (2013) Assessing Gait Patterns of Healthy Adults Climbing Stairs Employing Machine Learning Techniques. *Int. J. Intell. Syst.*, 28(3): 257–270.
11. Chopra S., Kaufman K.R. (2018) Effects of Total Hip Arthroplasty on Gait. In: Müller B. and Wolf S. (eds.) *Handbook of Human Motion*, Springer, Cham, pp. 1–15. DOI: 10.1007/978-3-319-30808-1_81-1.
12. Dindorf C., Teufl W., Taetz B., Bleser G., Fröhlich M. (2020) Interpretability of Input Representations for Gait Classification in Patients after Total Hip Arthroplasty. *Sensors*, 20: 16.
13. Eskofier B.M., Federolf P., Kugler P.F., Nigg B.M. (2013) Marker-based classification of young-elderly gait pattern differences via direct PCA feature extraction and SVMs. *Comput. Methods Biomech. Biomed. Engin.*, 16(4): 435–442.
14. Figueiredo J., Santos C.P., Moreno J.C. (2018) Automatic recognition of gait patterns in human motor disorders using machine learning: A review. *Med. Eng. Phys.*, 53: 1–12.
15. Foucher K.C. (2016) Gait abnormalities before and after total hip arthroplasty differ in men and women. *J. Biomech.*, 49(14): 3582–3586.
16. Głowiński S., Losiński K., Kowiański P., Waśkow M., Bryndal A., Grochulska A. (2020) Inertial Sensors as a Tool for Diagnosing Discopathy Lumbosacral Pathologic Gait: A Preliminary Research. *Diagnostics*, (Basel, Switzerland) 10(6): 342.

17. Horstmann, T., Listringhaus, R., Haase, G.-B., Grau, S., Mündermann, A. (2013) Changes in gait patterns and muscle activity following total hip arthroplasty: A six-month follow-up. *Clinical biomechanics (Bristol, Avon)* 28, 7: 762–769.
18. Hutchinson L., Schwartz J.B., Morton A.M., Davis I.S., Deluzio K.J., Rainbow M.J. (2018) Operator Bias Errors Are Reduced Using Standing Marker Alignment Device for Repeated Visit Studies. *J. Biomech. Eng.*, 140(4).
19. Ilias S., Tahir N.M., Jailani R., Hasan C.Z.C. (2017) Linear Discriminant Analysis in Classifying Walking Gait of Autistic Children. In: *2017 European Modelling Symposium (EMS)*. IEEE, pp. 67–72. DOI: 10.1109/EMS.2017.22.
20. Kalousis A., Prados J., Hilario M. (2007) Stability of feature selection algorithms: a study on high-dimensional spaces. *Knowl. Inf. Syst.*, 12(1): 95–116.
21. Kleemann R.U., Heller M.O., Stoeckle U., Taylor W.R., Duda G.N. (2003) THA loading arising from increased femoral anteversion and offset may lead to critical cement stresses. *J. Orthop. Res.*, 21(5): 767–774.
22. Laroche D., Tolambiya A., Morisset C., Maillefert J.F., French R.M., Ornetti P., Thomas E. (2014) A classification study of kinematic gait trajectories in hip osteoarthritis. *Comput. Biol. Med.*, 55: 42–48.
23. Leardini A., Sawacha Z., Paolini G., Ingrosso S., Nativio R., Benedetti M.G. (2007) A new anatomically based protocol for gait analysis in children. *Gait & Posture* 26(4): 560–571.
24. Liu F.T., Ting K.M., Zhou Z.-H. (2012) Isolation-Based Anomaly Detection. *ACM Trans. Knowl. Discov. Data*, 6(1): 1–39.
25. Liu H., Yu L. (2005) Toward Integrating Feature Selection Algorithms for Classification and Clustering. *IEEE Transactions on Knowledge and Data Engineering* 17(4): 491–502.
26. OECD. (2016) *Health at a Glance: Europe 2016. State of Health in the EU Cycle*. OECD Publishing, Paris.
27. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D. (2011) Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, 12: 2825–2830.
28. Peng H., Long F., Ding C. (2005) Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8): 1226–1238.
29. Perron M., Malouin F., Moffet H., McFadyen B.J. (2000) Three-dimensional gait analysis in women with a total hiparthroplasty. *Clin. Biomech.*, (Bristol, Avon), 15(7): 504–515.
30. Phinyomark A., Petri G., Ibáñez-Marcelo E., Osis S.T., Ferber R. (2018) Analysis of Big Data in Gait Biomechanics: Current Trends and Future Directions. *J. Med. Biol. Eng.*, 38(2): 244–260.
31. Rasch A., Dalén N., Berg H.E. (2010) Muscle strength, gait, and balance in 20 patients with hip osteoarthritis followed for 2 years after THA. *Acta Orthop.*, 81(2): 183–188.
32. Seijo-Pardo B., Bolón-Canedo V., Alonso-Betanzos A. (2019) On developing an automatic threshold applied to feature selection ensembles. *Inf. Fusion*, 45: 227–245.
33. Shakoor N., Sengupta M., Foucher K.C., Wimmer M.A., Fogg L.F., Block J.A. (2010) Effects of common footwear on joint loading in osteoarthritis of the knee. *Arthritis Care Res.*, 62(7): 917–923.
34. Teuff W., Taetz B., Miezal M., Lorenz M., Pietschmann J., Jöllenbeck T., Fröhlich M., Bleser G. (2019) Towards an Inertial Sensor-Based Wearable Feedback System for Patients after Total Hip Arthroplasty: Validity and Applicability for Gait Classification with Gait Kinematics-Based Features. *Sensors*, 19(22): 5006.
35. Thewlis D., Bishop C., Daniell N., Paul G. (2013) Next-Generation Low-Cost Motion Capture Systems Can Provide Comparable Spatial Accuracy to High-End Systems. *J. Appl. Biomech.*, 29: 112–117.
36. Wang Z., Bao H.-W., Hou J.-Z. (2019) Direct anterior versus lateral approaches for clinical outcomes after total hip arthroplasty: a meta-analysis. *J. Orthop. Surg. Res.*, 14(1): 1–11.
37. Wong T.-T. (2015) Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit.*, 48(9): 2839–2846.

Received 22.02.2021

Accepted 16.04.2021

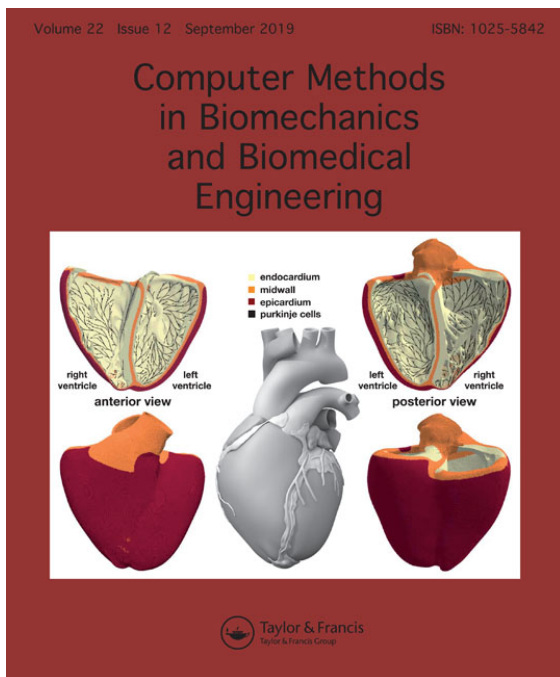
© University of Physical Education, Warsaw, Poland

Acknowledgments

This research was funded by Offene Digitalisierungsallianz Pfalz, BMBF [03IHS075B].

We would like to thank the staff of the Klinik Lindenplatz, Bad Sassendorf, Germany for their support concerning the subject and data acquisition. We would especially like to thank Juliane Pietschmann and Thomas Jöllenbeck, who supported us with the recruitment and measurements.

3.2 2. Publikation



General method for automated feature extraction and selection and its application for gender classification and biomechanical knowledge discovery of sex differences in spinal posture during stance and gait

C. Dindorf, J. Konradi, C. Wolf u. a., "General method for automated feature extraction and selection and its application for gender classification and biomechanical knowledge discovery of sex differences in spinal posture during stance and gait," *Computer Methods in Biomechanics and Biomedical Engineering*, Jg. 24, Nr. 3, S. 1–9, 2020. DOI: 10.1080/10255842.2020.1828375

Zusammenfassung

Hintergrund: Moderne Technologien ermöglichen die Erfassung vielfältiger biomechanischer Parameter, was oft zu relationalen Datenbeständen führt. Die Arbeit schlägt eine allgemein anwendbare Methode vor, die eine automatisierte Ensemble Feature Extraktion und Klassifizierung umfasst, um das Potenzial der Daten auch für die Generierung neuer biomechanischer Erkenntnisse bestmöglich zu erfassen. Der Nutzen wird an einem konkreten biomechanisch und medizinisch relevanten Beispiel, der Geschlechterklassifizierung auf Basis von Wirbelsäulendaten für Stand und Gang, evaluiert.

Methodische Forschungsfragen: Wie können relationale Datenbestände im biomechanischen Kontext möglichst mit geringem Verlust an bedeutsamen Informationen genutzt werden? Erweist sich die Kombination von homogener und heterogener Ensemble Feature Selektion als vielversprechend?

Fachspezifische Forschungsfragen: Gibt es Unterschiede der Wirbelsäule zwischen den biologischen Geschlechtern? Welche globalen, komplexen Parameter sind für die Unterscheidung besonders relevant?

Methodik: Alter, statische und dynamische Wirbelsäuledaten von 201 Gesunden bilden die Ausgangsdatenbasis. Getrennt für die statischen und dynamischen Daten wurde die vorgestellte Pipeline zur Ausnutzung der relationalen Datenbestände, Ensemble Feature Selektion und Klassifikation angewendet und die Ergebnisse miteinander verglichen.

Ergebnisse: Unter Verwendung von Gangdaten wurden sehr gute Ergebnisse für die Genauigkeit erzielt (mittlerer MCC = 0,97). Die Verwendung der statischen Daten führt zu reduzierter Klassifikationsgenauigkeit (mittlerer MCC = 0.62). Dynamische Bewegungen der Lendenwirbelsäule in sagittaler und frontaler Ebene und des Beckens in frontaler Ebene bilden Geschlechtsunterschiede am besten ab. Die Pipeline, welche homogene und heterogene Ensemble Feature Selektion kombiniert, zeigt sich überlegen hinsichtlich alleiniger heterogener Ensemble Feature Selektion.

Fazit: Automatische Feature Extraktion auf Basis relationaler Datenbestände kann zu abstrakten Features führen. Interpretierbares Feature Engineering sowie die Evaluation der Notwendigkeit der Verwendung komplexer Feature erscheint daher zukünftig wichtig.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

SCHOLARONE™
Manuscripts

This is an Accepted Manuscript of an article published by Taylor & Francis in
Computer Methods in Biomechanics and Biomedical Engineering on 02 Nov 2020,
available at: <https://doi.org/10.1080/10255842.2020.1828375>.

URL: <http://mc.manuscriptcentral.com/gcmb>

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

General method for automated feature extraction and selection and its application for gender classification and biomechanical knowledge discovery of sex differences in spinal posture during stance and gait

Please refer to the cover letter for the contribution / order of the authors.

Modern technologies enable to capture multiple biomechanical parameters often resulting in relational data. The current work proposes a generally applicable method comprising automated feature extraction, ensemble feature selection and classification to best capture the potentials of the data also for generating new biomechanical knowledge. Its benefits are demonstrated in the concrete biomechanically and medically relevant use case of gender classification based on spinal data for stance and gait. Very good results for accuracy were obtained using gait data. Dynamic movements of the lumbar spine in sagittal and frontal plane and of the pelvis in frontal plane best map gender differences.

Keywords: ensemble feature selection, classification, surface topography, spine, motion, gender

1. Introduction

Human biomechanics is a complex and multivariate matter. To approach an understanding of its complexity, multiple parameters and modalities may have to be taken into account, which often results in multiple related tables of data (relational data). Modern movement tracking systems enable the collection of large, multivariate datasets (big data) (Phinyomark et al. 2018). This opens up the potential of generating new knowledge and a deeper understanding of human biomechanics. However, classical inferences based static and hypothesis testing methods show limited capabilities in properly analyzing these huge amounts of data and included parameters (Bzdok et al. 2018). In order to generate actual benefit from the data that we can now capture, exploratory approaches, including machine learning methods such as automated feature extraction, selection and pattern recognition that can simultaneously consider a large

1
2
3 amount of data and parameters, including their interactions, are gaining importance.
4

5
6 Therefore, we propose a generally applicable ensemble feature selection approach in
7
8 this work.

9
10 In terms of applications, the biomechanical analysis of the human spine during
11
12 stance and gait is a specific case that received a lot of attention (Mohan and Huynh
13
14 2019), and huge innovations concerning data capturing capabilities took place in the last
15
16 decades. Rasterstereography, lately called Surface Topography (ST) (Betsch et al.
17
18 2013), is one important innovation. In its latest development, gait specific ST (Betz et
19
20 al. 2017) which also enables the precise determination of spinal movement in direct
21
22 relation to phases of gait via timely synchronization of spinal position and foot pressure
23
24 measuring data (Betz et al. 2018) can serve as a valuable tool for measuring spinal
25
26 motion in all body planes without extensive preparation or the usage of invasive or
27
28 radiation-based approaches.
29
30
31

32
33 Moreover, in biomechanical analysis of the human spine, gender aspects play an
34
35 important role in medicine. For example, spinal sex differences can be found in various
36
37 pathologies (e.g. spondylolysis, disc degeneration, scoliosis: (Mohan and Huynh 2019);
38
39 and low back pain: (Bailey 2009; Fillingim et al. 2009; Bagheri et al. 2018)). The ability
40
41 to understand and measure parameters that are different between males and females
42
43 might help to detect early stages of potentially pathologic gait (Ko et al. 2011) as well
44
45 as pathologic aspects of posture contributing to, for instance, severe back pain among
46
47 females (Araujo et al. 2014). Regarding general sex differences of the spine results of
48
49 different measuring approaches and varying angle definitions do not converge to a
50
51 consistent picture for alignment (Schröder et al. 2011; Endo et al. 2012; Hamm 2016;
52
53 Kinel et al. 2018), as well as motion in stance (Dvorak et al. 1995; Dreischarf et al.
54
55 2014). As for stance measures of healthy individuals, varying marker based approaches
56
57
58
59
60

1
2
3 focusing on range of motion (ROM) during gait do not lead to a coherent picture either
4
5 (Kadaba et al. 1990; Crosbie et al. 1997; Chockalingam et al. 2002; Smith et al. 2002;
6
7 Cho et al. 2004; Chumanov et al. 2008; Goutier et al. 2010; Bruening et al. 2015).
8
9 Converging results were only seen when using the same methodic approach, i.e. ST,
10
11 whereas ST research mainly focused on stance (Schröder et al. 2011; Hamm 2016). Gait
12
13 was considered little (Gipsman et al. 2014) but different walking speeds not at all.
14
15

16
17 ST has mainly focused on transversal movement for it shows the best accuracy
18
19 in measuring rotation angels resulting in the lowest standard error of the mean for
20
21 transversal plane values (Gipsman et al. 2014). Additionally, based on the assumption
22
23 that human bipedal gait has developed from the quadrupedal diagonal-sequence gait
24
25 (Schmitt et al. 2006; Niemitz 2010) leading to an expectable interaction between the
26
27 upper and lower extremity movements, previous research (Gregersen and Lucas 1967;
28
29 Stokes et al. 1989; Needham et al. 2016) investigated spinal motion in different planes
30
31 and revealed characteristic types in the transversal plane (i.e. amplitudes of different
32
33 amounts along the spine or opposite rotation between shoulder girdle and pelvis) that
34
35 can be used to detect specific movement patterns.
36
37
38

39
40 Previous measuring approaches have shown only limited capabilities in
41
42 adequately analyzing spinal biomechanics. Further, they have not been able to
43
44 comprehensively measure movements of the spinal segments in all body planes.
45
46 Consequently, the recently developed Ggait specific ST may provide a promising
47
48 approach for capturing parameters, which were little or not regarded in the analysis until
49
50 now. Moreover, as mentioned above, aAn exploratory, data driven analysis approach
51
52 may be useful to better exploit the extended database for expanding prior knowledge
53
54 and gaining new insights. Consequently, the current work proposes a generally
55
56 applicable method comprising automated feature extraction, ensemble feature selection
57
58
59
60

1
2
3 and classification and then demonstrates its benefits in the concrete biomechanically
4
5 and medically relevant use case of gender classification based on gait kinematics. This
6
7 demonstration comprises three aspects: Due to the focus of previous research on stance
8
9 data, we first apply the proposed method to show that the use of gait data is more
10
11 appropriate for gender classification compared to stance data. Second, our attempt is to
12
13 apply the method to demonstrate that the use of movement data from all planes of
14
15 motion not separate but simultaneously outperforms the pure use of transversal
16
17 movement data for gender classification. Third, with the goal of increasing the
18
19 understanding of gender differences in gait kinematics, we apply the method to analyze,
20
21 which variables during gait have the highest discriminative power for the given task of
22
23 gender classification.
24
25
26
27
28
29

30 **2. Material and methods**

31 ***2.1 Proposed feature extraction methodology***

32 *2.1.1 Automated feature extraction*

33
34
35
36 For feature calculation, deep feature synthesis is performed with the automated
37
38 feature engineering framework featuretools (Feature Labs 2019). Therefore, relational
39
40 data frames are defined. Aggregation primitives are used to calculate new features and
41
42 to summarize the data frames to one single feature vector. They are computations which
43
44 are applied across datasets according to parent-child relationships. Prominent examples
45
46 are arithmetic mean, median, minima, maxima and standard deviation. In the case of
47
48 gait waveform data, range of motion, the time points for the maxima and minima during
49
50 a cycle and the length of each gait cycle (before interpolation) should also be applied as
51
52 primitives to summarize the main characteristics of the waveforms.
53
54
55
56
57
58
59
60

2.1.2 Ensemble feature selection

Feature selection is an important step towards improving a model's accuracy, interpretability, prevent overfitting and reduce the required computing power (Liu and Yu 2005; Tang et al. 2014). As part of knowledge discovery, new insights can be gained by selecting the most relevant features for a specific task. However, not only predictive performance but also stability / robustness to changes in the input data are important aspects of feature selection. In this context, stability quantifies the extent to which variations in training data lead to different feature preferences (Kalousis et al. 2007). This aspect is especially important, when the selected features are used for knowledge discovery (Victo Sudha George and Cyril Raj 2015). For example does a low stability lead to reduced reproducibility and thus a reduced confidence in the results (Pes 2019). Recently, *ensemble feature selection* approaches have been reported as a promising approach for dealing with those limitations (Afef and Limam 2018; Shahrjooihaghighi et al. 2018). The basic idea is that a combination of multiple feature selection processes leads to a higher stability and generalizability, particularly in high dimensionality, low sample size domains (Bolón-Canedo and Alonso-Betanzos 2019). In general, ensemble selection can be divided in *homogeneous* (same selection methods with different subsets) and *heterogeneous* (different selection methods with same dataset) approaches. Furthermore, both approaches can be combined (Afef and Limam 2018).

In the proposed methodological approach, a combination of homogeneous and heterogeneous selection is carried out to address three main objectives:

- a) Increasing stability and confidence in selected features
 - b) Independence from deciding which single selection approach is the best
- (Seijo-Pardo et al. 2017)

- 1
2
3 c) Independence from specific classification algorithms and therefore
4
5 generalizability
6
7

8 The approach is based on algorithms which return a ranking of the features according to
9 their relevance. The ranking algorithms are described in the following.
10
11

12 13 14 *2.1.3 Ranking algorithms* 15

16 In general, feature selection methods can be divided into three groups: filter, wrapper
17 and embedded methods (for a detailed description see e.g. Chandrashekar and Sahin
18 2014 or Bolón-Canedo et al. 2013). Filter methods are performing a variable ranking
19 based on statistical criteria. They are independent of any machine learning algorithm.
20 Wrapper methods are using a predictive model to compare and evaluate different
21 subsets of features. Finally, embedded methods perform a variable selection during the
22 training process. To ensure diversity, one filter, one embedded and one wrapper method
23 was chosen for feature ranking. Features with the highest relevance got the highest
24 ranking score. Not selected features were scored with zero.
25
26
27
28
29
30
31
32
33
34
35
36
37

- 38 • Joint Mutual Information Maximization (JMIM) (filter): Nonlinear, classifier
39 independent feature selection method which uses mutual information and the
40 maximum of the minimum strategy. Promising results for accuracy and stability
41 were reported which outperform similar methods (e.g. Maximum Relevancy
42 Minimum Redundancy) (Bennasar et al. 2015).
43
44
45
46
47
48
- 49 • Random Forest (RF) (embedded): Ensemble / collection of tree predictors
50 (Breiman 2001). For variable ranking the average FI (Gini importance) score
51 from 50 runs is suggested (see Genuer et al. 2010).
52
53
54
55
56
57
58
59
60

- Recursive Feature Elimination with Support Vector Machine (SVM-RFE) (wrapper): Forward selection of relevant features with ranking according to the time they were included.

The proposed ensemble selection design is depicted in Figure 1. First, the training data is split into bootstrap samples. For each bootstrap sample, feature ranking is performed using the three different algorithms. To reduce the number of redundant features and to lower the computational costs (especially for the wrapper), a preselection and ranking with the JMIM filter is suggested, before application of the other selection approaches (green arrow). Therefore, a threshold is applied in order to preselect and simultaneously rank a specific number of features with the JMIM filter. The other ranking algorithms were then applied with the preselected features of the JMIM filter.

The rankings for each algorithm are combined into one ranking per algorithm (aggregation 1). In a second aggregation step (aggregation 2), the ensemble lists of the different algorithms are combined to one final ranking. For aggregation, different methods like the median, arithmetic mean or more sophisticated methods were proposed (Bolón-Canedo and Alonso-Betanzos 2019). Arithmetic mean aggregation showed the best performance on various data sets (Wang et al. 2019) and is proposed for combining the rankings.

In order to obtain a practical subset of features based on the ensemble ranking, it is necessary to set a cut-off threshold. Before application of the threshold values, features with the rank zero are dropped. The resulting feature set is then used for classification. Beside, classification performance allows to evaluate the quality of the extracted features.

Please insert figure 1 here.

2.1.4 Proposed evaluation methodology (classification and feature stability)

The ranking approach is integrated into a cross-validation (CV) procedure (Stratified Shuffle Split CV for unbalanced classes; preserves the same percentage of each class in the test set as in the original dataset), to prevent the use of test data for feature selection and therefore to obtain an unbiased accuracy score. Further, to evaluate stability of the feature sets through comparing the rankings of the different training datasets. In order to compensate for possible imbalanced class distribution, synthetic minority oversampling (Chawla et al. 2002), each time performed on the respective training set, is proposed. Training and test data are standardized based on the respective training set before classification.

2.2 Use case application of the proposed approach: gender differences

2.2.1 Measuring method

As a means for ST we used the DIERS formetric III 4D™, DICAM v3.7.1.7 analyzing system. Structured light that is projected onto the textile free back of the individual person gets recorded by a camera unit in defined position with a frequency of 60 Hz. Based on up to 150.000 measuring points (depending on body size) a 3D image of the surface gets created by the software. For each spinal segment starting at spinous process of C7 and ending at the pelvis the system then calculates the corresponding 3D movements. L5 is not measured, for it is embedded in the structure of the pelvis leading to less correlation accuracy between the surface structure and this spinal position. Additional video signals from a rear axis and two lateral cameras enable consecutive visual inspection for various purposes. The system also features a Zebris™ foot pressure measuring plate (5376 sensors, scanning frequency 120 Hz, sensitivity 1

1
2
3 N/cm², accuracy 5 % FS) that measures ground reaction forces. In a further
4
5 development of the system (Betz et al. 2017) we use the foot pressure measuring data to
6
7 encode for step and swing phases as well as for complete gait cycles. A modification of
8
9 the systems export functions timely synchronizes the spinal motion data with foot
10
11 pressure measuring data and combines both measurement results into the same raw data
12
13 export. Since all exports are separate for each measurement, we merge single export
14
15 files to create a complete raw data table using Statistical Analysis System (SAS, v 9.1).
16
17 This was the basis for the feature engineering.
18
19

20 21 22 *2.2.2 Subjects and data acquisition*

23
24 We used ST data from a reference data pool that was collected using the described
25
26 measurement method (Betz et al. 2017). The responsible ethics committee of the
27
28 medical chamber Rhineland-Palatinate (837.194.16) approved the study. It is registered
29
30 with the WHO (INT: DRKS00010834). Informed consent was obtained from all
31
32 participants. The data pool contains measurement values of 201 healthy subjects (132
33
34 female, 69 male) during standing position and while walking at 2, 3, 4 and 5 km/h. A
35
36 stance measure contains 12 single frames. For each walking speed, three gait cycles
37
38 were recorded, leading to 300-500 frames for each spinal parameter. More than 100
39
40 spinal parameters were recorded in total. For a detailed description of the measured
41
42 parameters used for feature engineering please refer to the supplemental online material.
43
44 Originally recorded measures of length and BMI were excluded. Missing data points for
45
46 the gait data were interpolated using spline interpolation (maximum gap = 5 frames).
47
48 Cycles with remaining gaps were dropped. The gait cycles were then individually time-
49
50 normalized to 101 time steps (from 0 % to 100 %) using cubic spline interpolation.
51
52
53
54
55

56 57 *2.2.3 Ensemble design configuration*

58
59 For automated feature extraction mean, max and min as well as the waveform
60

1
2
3 specific computations were used as primitives. The relational data frames were defined
4
5 the following: Every subject has demographic data (age) and recordings for a static and
6
7 a dynamic trial. The dynamic trial consists of the different walking speeds. Every
8
9 walking speed is represented through the aggregations of each of the three gait cycles
10
11 per speed. In total 202 stance respectively 2021 gait features were obtained. For each
12
13 calculated feature, outliers were separately detected using interquartile range.
14
15 Anomalies were imputed with the mean value of each class. To compare the single use
16
17 of transversal movements for classification, the respective features ($n_{\text{feat}} = 570$) were
18
19 extracted from the initial gait features.
20
21
22

23
24 The proposed ensemble design is applied with the following parameters: feature
25
26 ranking was integrated into 5 folds Stratified Shuffle Split CV (test size 20 %). On each
27
28 training set, 50 bootstrap samples (size = 80 % training data) were drawn for ensemble
29
30 ranking. 10 % of the initial gait features ($n_{\text{feat}} = 202$; no improvement for preselection of
31
32 more features) were preselected and ranked with the JMIM filter. Stance data was not
33
34 preselected and directly ranked with JMIM filter due to the relatively low amount of
35
36 available features ($n_{\text{feat}} = 202$). For ranking, the RF algorithm is applied with 100 trees.
37
38 Radial basis function kernel (RBF) is used for SVM-RFE. The following threshold
39
40 values are used to generate the final ensemble subsets: 5, 10, 20, 50, 100 %.
41
42
43
44

45
46 Support Vector Machine (SVM) with RBF kernel is mainly used for
47
48 classification. To check classifier independency and generalizability, Logistic
49
50 Regression and RF algorithm were additionally applied. Due to the imbalanced classes
51
52 in the test set, Matthews correlation coefficient and F1 score were chosen as accuracy
53
54 metrics. Further, the output quality is reported with confusion matrix and Precision-
55
56 Recall Area Under the Curve (AUC) metrics. To check usefulness of the proposed
57
58 combination of homogeneous and heterogeneous feature selection with regard to
59
60

1
2
3 stability, heterogeneous feature selection (without the bootstrap samples) is separately
4
5 performed. The feature occurrence is then counted and compared with the results of the
6
7 proposed approach.
8
9

10 11 **3. Results**

12
13 Classification results for stance and gait data are presented in table 1. Using the gait
14
15 data leads to noticeably higher classification performance than only using data for the
16
17 stance posture. Using motion of all body planes ($n_{\text{feat}} = 2021$, $M_{\text{MCC}} = 0.96$,
18
19 $M_{\text{AUC}} = 1.00$) leads to better classification results compared to the single use of
20
21 transversal movement ($n_{\text{feat}} = 570$, $M_{\text{MCC}} = 0.69$, $M_{\text{AUC}} = 0.90$). For the stance data, the
22
23 best classification is obtained using the 20 % threshold. The results for the gait data are
24
25 analyzed in more detail below.
26
27
28
29
30
31
32

33 **Please insert table 1 here.**
34
35
36
37

38 With the use of the 100 % threshold and a mean of 1256 features, the best classification
39
40 for the gait data is achieved. Compared to the use of all initial input features accuracy is
41
42 little higher. Application of the 10, 20 or 50 % thresholds results in a minor decrease in
43
44 classification performance but a major reduction in subset size. Similar results are
45
46 present using different classification algorithms (e.g. for 20 % threshold, RF:
47
48 $M_{\text{MCC}} = 0.96 \pm 0.05$, Logistic Regression: $M_{\text{MCC}} = 0.96 \pm 0.04$).
49
50
51
52
53

54 **Please insert figure 2 here.**
55
56
57
58
59
60

1
2
3 The feature rankings of each CV fold are presented in figure 2. The highest
4 ranked features show little variations in the ranking scores between the different folds.
5
6 Of the top 50 highest ranked features, 30 features occurred in all folds, 36 at least in
7 four out of five (21, respectively 28 without bootstrapping / heterogeneous feature
8 selection). Corresponding movements for the highest ranked features during a gait
9 cycle, separate for the 2 and 5 km/h walking speeds, are exemplary shown in figure 3.
10
11
12
13
14
15
16
17
18
19

20 **Please insert figure 3 here.**
21
22

23 **4. Discussion**

24
25 High classification performances and improved stability compared to application of
26 purely heterogeneous feature selection indicate that the proposed ensemble approach is
27 able to identify discriminative features and increase stability of the final subsets. Using
28 data of different walking speeds, automated feature engineering and ensemble feature
29 selection, meaningful features could be objectively extracted without prior knowledge
30 using a data driven strategy. According to Seijo-Pardo et al. (2019), ensemble
31 approaches are able to capture relevance rather than redundancy. In the present case, the
32 ability of the JMIM algorithm to remove redundant features gets lost during the
33 aggregation steps. Thus, a further reduction of the feature set size while maintaining the
34 accuracy may be possible through reduction of redundant features.
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 The use of gait data of different walking speeds results in higher classification accuracy
50 compared to the use of stance data alone. In literature, gender differences regarding the
51 spine were mainly analyzed for stance data (Schröder et al. 2011; Endo et al. 2012;
52 Hamm 2016; Kinel et al. 2018). The current results show, that gait data better maps sex
53 differences for classification and therefore important information were previously not
54
55
56
57
58
59
60

1
2
3 regarded.
4
5

6 The use of transversal movement leads to a major reduction of classification
7
8 performance compared to the use of data of all planes of motion. According to
9
10 literature, gender effects in transversal plane are likely to be seen for ROM of the lower
11
12 thoracic spine, and the pelvis (Crosbie et al. 1997; Bruening et al. 2015). Nevertheless,
13
14 transversal movement does also not occur as a feature in the most stable and most
15
16 relevant features of the top 50 ensemble rankings. Frontal and sagittal plane movement
17
18 show the highest relevancy for classification.
19
20
21
22

23 The following characteristics and gender differences can be identified regarding the
24
25 gait data: Overall, the most stable and highest ranked features regarding the top 50
26
27 rankings are based on movement of the lumbar spine (frontal, sagittal plane) and the
28
29 pelvis (frontal plane). Regarding sagittal plane, range of motion for L1, L2, L3 is
30
31 increased for females. In frontal plane, the features indicate the importance of L2, L3,
32
33 L4 and pelvic (pelvic obliquity) movement. Increased L3 and L4 ROM for females is
34
35 important for classification. Further, the point in time of the maximal respectively
36
37 minimal deflection of the variables L2, L3 and pelvic obliquity is highly relevant. Males
38
39 and females show diverging movement for the respective variables. One of few
40
41 consistent findings across existing studies is an increased ROM of female's pelvic
42
43 obliquity (Chockalingam et al. 2002; Smith et al. 2002; Cho et al. 2004; Chumanov et
44
45 al. 2008; Bruening et al. 2015). However, the current study indicates, that the point in
46
47 gait cycle of the maximal deflection for pelvic obliquity may have higher discriminative
48
49 power.
50
51
52
53
54

55 The aim of the present work was to provide a generally applicable feature extraction,
56
57 selection and classification method and demonstrate its use in the concrete case of
58
59 gender classification based on spinal data for stance and gait. As a result, an initial
60

1
2
3 overview of ST based gender-specific differences in relation to the spine was provided,
4
5 i.e. general information where important gender differences can be found. To quantify
6
7 the uncertainty of the predictions, future studies should focus more on predicting
8
9 probabilities in addition to class memberships. Through the use of complex, nested
10
11 features, which are based on relation data, there is little loss of potentially meaningful
12
13 information which results in high classification accuracy. However, the complexity of
14
15 the calculated features can also make interpretation more difficult. Following studies
16
17 should therefore also focus on interpretability in addition to classification performance,
18
19 for example using explainable artificial intelligence approaches (Adadi and Berrada
20
21 2018). Further, different dimensionality reduction methods should be applied in order to
22
23 reveal and analyze dependencies between different parameters. Based on literature (see
24
25 e.g. Teufl et al. 2019) we normalized the gait cycle to 101 time steps. However, the
26
27 influence of the resolution of the gait data is unclear, generally little addressed in the
28
29 literature and should therefore be compared and analyzed in future studies.
30
31
32
33
34

35
36 Limitations arise from the ST methodology. First, it is not clear to which extend our
37
38 results can be generalized to different measuring methods. Second, validity was only
39
40 shown for dynamic stance measures (Mohokum et al. 2015). For the use in dynamic gait
41
42 analysis there is at least good evidence for strong reliability and reproducibility
43
44 (Gipsman et al. 2014). To further validate this assumption, as a next step, we will apply
45
46 repeated measurement testing to gait analysis for the same individual to search for intra-
47
48 individual patterns in the sense of a personal spinal ‘footprint’. This would also provide
49
50 an indication for the reliability of the ST approach. Additionally, we want to transfer the
51
52 proposed method to other clinically relevant classification tasks cases, such as the
53
54 classification into back pain, hip and knee osteoarthritis, as well as spondylodosis.
55
56
57
58
59
60

5. Conclusion

Using the proposed, generally applicable, feature extraction, ensemble feature selection and classification method, it was demonstrated that gender classification based on gait ST using parameters from all planes of motion, provides superior results compared to stance data and data purely from the transversal plane. Moreover, the method proved useful for identifying features that are most relevant in characterizing gender differences, i.e. the results indicate that dynamic movements of the lumbar spine in sagittal and frontal plane and of the pelvis in frontal plane best map gender differences. A deeper understanding of the latter could aid e.g. the detection of aspects contributing to severe back pain among females (Ko et al. 2011; Araujo et al. 2014).

Acknowledgements

Most of all we would like to thank the participants of this study. The DIERS Company is acknowledged for donating the measuring device the University Medical Center of the Johannes Gutenberg University Mainz for the purpose of research. Finally, we express our gratitude to Kjell Heitmann and Amira Basic for technological support and to Helmut Diers for manifold contributions in this project.

This work was supported by the BMBF under grant numbers 16SV7115 and 03IHS075B.

References

- Adadi A, Berrada M. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*. 6:52138–52160.
- Afef BB, Limam M. 2018. Ensemble feature selection for high dimensional data: a new method and a comparative study. *Adv Data Anal Classif*. 12:937–952.
- Araujo F, Lucas R, Alegrete N, Azevedo A, Barros H. 2014. Sagittal standing posture, back pain, and quality of life among adults from the general population: a sex-specific association. *Spine (Phila Pa 1976)*. 39(13):E782-794.
- Bagheri R, Ebrahimi Takamjani I, Dadgoo M, Ahmadi A, Sarrafzadeh J, Pourahmadi MR, Jafarpisheh AS. 2018. Gender-Related Differences in Reliability of Thorax, Lumbar, and Pelvis Kinematics During Gait in Patients With Non-specific Chronic Low Back Pain. *Annals of rehabilitation medicine*. 42(2):239-249.
- Bailey A. 2009. Risk factors for low back pain in women: still more questions to be answered. *Menopause (New York, NY)*. 16(1):3-4.
- Bennasar M, Hicks Y, Setchi R. 2015. Feature selection using Joint Mutual Information Maximisation. *Expert Systems with Applications*. 42:8520–8532.
- Betsch M, Wild M, Johnstone B, Jungbluth P, Hakimi M, Kuhlmann B, Rapp W. 2013. Evaluation of a novel spine and surface topography system for dynamic spinal curvature analysis during gait. *PloS one*. 8(7):e70581.
- Betz U, Huthwelker J, Konradi J, Schmidtman I, Bodem F, Heil J, Drees P. 2017. First results of spinal segment-related motion analysis in human gait using rasterstereography. *Gait & Posture*. 57(Supplement 1):345-346.
- Betz U, Konradi J, Huthwelker J, Wolf C, Diers H, Heitmann K, Westphal R, Drees P. 2018. Quantitative and qualitative parameters to characterize segmental spine

1
2
3 movement during gait. World Congress of Biomechanics; 2018.07.08-12.; Dublin,
4
5 Ireland.

6
7
8 Bolón-Canedo V, Alonso-Betanzos A. 2019. Ensembles for feature selection: A review
9
10 and future trends. *Information Fusion*. 52:1–12.

11
12 Bolón-Canedo V, Sánchez-Marño N, Alonso-Betanzos A. 2013. A review of feature
13
14 selection methods on synthetic data. *Knowl Inf Syst*. 34:483–519.

15
16
17 Breiman L. 2001. Random Forests. *Machine Learning*. 45:5–32.

18
19 Bruening DA, Frimenko RE, Goodyear CD, Bowden DR, Fullenkamp AM. 2015. Sex
20
21 differences in whole body gait kinematics at preferred speeds. *Gait & Posture*.
22
23 41(2):540-545.

24
25
26 Bzdok D, Altman N, Krzywinski M. 2018. Statistics versus machine learning. *Nature*
27
28 *Methods*. 15:233–234.

29
30 Chandrashekar G, Sahin F. 2014. A survey on feature selection methods. *Computers &*
31
32 *Electrical Engineering*. 40:16–28.

33
34
35 Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. 2002. SMOTE: Synthetic
36
37 Minority Over-sampling Technique. *jair*. 16:321–357.

38
39
40 Cho SH, Park JM, Kwon OY. 2004. Gender differences in three dimensional gait
41
42 analysis data from 98 healthy Korean adults. *Clin Biomech (Bristol, Avon)*.
43
44 19(2):145-152.

45
46
47 Chockalingam N, Dangerfield PH, Giakas G, Cochrane T. 2002. Study of marker
48
49 placements in the back for opto-electronic motion analysis. *Stud Health Technol*
50
51 *Inform*. 88:105-109.

52
53
54 Chumanov ES, Wall-Scheffler C, Heiderscheit BC. 2008. Gender differences in
55
56 walking and running on level and inclined surfaces. *Clinical Biomechanics*.
57
58 23(10):1260-1268.

- 1
2
3 Crosbie J, Vachalathiti R, Smith R. 1997. Age, gender and speed effects on spinal
4 kinematics during walking. *Gait & Posture*. 5(1):13-20.
5
6
7
8 Dreischarf M, Albiol L, Rohlmann A, Pries E, Bashkuev M, Zander T, Duda G,
9
10 Druschel C, Strube P, Putzier M et al. 2014. Age-related loss of lumbar spinal
11 lordosis and mobility--a study of 323 asymptomatic volunteers. *PloS one*.
12
13 9(12):e116186.
14
15
16 Dvorak J, Vajda EG, Grob D, Panjabi MM. 1995. Normal motion of the lumbar spine as
17 related to age and gender. *Eur Spine J*. 4(1):18-23.
18
19
20
21 Endo K, Suzuki H, Nishimura H, Tanaka H, Shishido T, Yamamoto K. 2012. Sagittal
22 lumbar and pelvic alignment in the standing and sitting positions. *Journal of*
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Feature Labs I. 2019. Featuretools: automated feature engineering. [accessed 2020 Feb 05]. <https://www.featuretools.com/>.
- Fillingim RB, King CD, Ribeiro-Dasilva MC, Rahim-Williams B, Riley JL. 2009. Sex, Gender, and Pain: A Review of Recent Clinical and Experimental Findings. *The Journal of Pain*. 10(5):447-485.
- Genuer R, Poggi J-M, Tuleau-Malot C. 2010. Variable selection using random forests. *Pattern Recognition Lett*. 31:2225–2236.
- Gipsman A, Rauschert L, Daneshvar M, Knott P. 2014. Evaluating the Reproducibility of Motion Analysis Scanning of the Spine during Walking. *Adv Med*. 2014:721829.
- Goutier KM, Jansen SL, Horlings CG, Kung UM, Allum JH. 2010. The influence of walking speed and gender on trunk sway for the healthy young and older adults. *Age and ageing*. 39(5):647-650.

- 1
2
3 Gregersen GG, Lucas DB. 1967. An in vivo study of the axial rotation of the human
4 thoracolumbar spine. *The Journal of Bone & Joint Surgery*. 49(2):247-262.
5
6
7
8 Hamm J. 2016. Etablierung und Evaluation eines neuartigen lichtoptischen
9
10 Messsystems zur Untersuchung der Haltung des Beckens und der Wirbelsäule unter
11
12 dynamischen Bedingungen [Dissertation]. Düsseldorf: Heinrich-Heine-Universität
13
14 Düsseldorf.
15
16
17 Kadaba MP, Ramakrishnan HK, Wootten ME. 1990. Measurement of lower extremity
18
19 kinematics during level walking. *Journal of orthopaedic research : official*
20
21 *publication of the Orthopaedic Research Society*. 8(3):383-392.
22
23
24 Kalousis A, Prados J, Hilario M. 2007. Stability of feature selection algorithms: a study
25
26 on high-dimensional spaces. *Knowl Inf Syst*. 12:95–116.
27
28
29 Kinel E, D'Amico M, Roncoletta P. 2018. Normative 3D opto-electronic stereo-
30
31 photogrammetric sagittal alignment parameters in a young healthy adult population.
32
33 *PloS one*. 13(9):e0203679.
34
35
36 Ko SU, Tolea MI, Hausdorff JM, Ferrucci L. 2011. Sex-specific differences in gait
37
38 patterns of healthy older adults: Results from the Baltimore Longitudinal Study of
39
40 Aging. *Journal of Biomechanics*. 44(10):1974-1979.
41
42
43 Liu H, Yu L. 2005. Toward Integrating Feature Selection Algorithms for Classification
44
45 and Clustering. *IEEE Transactions on Knowledge and Data Engineering*. 17:491–
46
47 502.
48
49
50 Mohan M, Huynh L. 2019. Sex Differences in the Spine. *Current Physical Medicine and*
51
52 *Rehabilitation Reports*. 7(3):246-252.
53
54
55 Mohokum M, Schulein S, Skwara A. 2015. The validity of rasterstereography: a
56
57 systematic review. *Orthopedic reviews*. 7(3):68-73.
58
59
60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Needham R, Naemi R, Healy A, Chockalingam N. 2016. Multi-segment kinematic model to assess three-dimensional movement of the spine and back during gait. *Prosthetics and orthotics international*. 40(5):624-635.
- Niemitz C. 2010. The evolution of the upright posture and gait--a review and a new synthesis. *Naturwissenschaften*. 97(3):241-263.
- Pes B. 2019. Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains. *Neural Computing & Applications*. 3:1157.
- Phinyomark A, Petri G, Ibáñez-Marcelo E, Osis ST, Ferber R. 2018. Analysis of Big Data in Gait Biomechanics: Current Trends and Future Directions. *J. Med. Biol. Eng.* 38:244–260.
- Schmitt D, Cartmill M, Griffin TM, Hanna JB, Lemelin P. 2006. Adaptive value of ambling gaits in primates and other mammals. *The Journal of experimental biology*. 209(Pt 11):2042-2049.
- Schröder J, Stiller T, Mattes K. 2011. Referenzdaten in der Wirbelsäulenformanalyse. *Manuelle Medizin*. 49(3):161-166.
- Seijo-Pardo B, Bolón-Canedo V, Alonso-Betanzos A. 2019. On developing an automatic threshold applied to feature selection ensembles. *Information Fusion*. 45:227–245.
- Seijo-Pardo B, Porto-Díaz I, Bolón-Canedo V, Alonso-Betanzos A. 2017. Ensemble feature selection: Homogeneous and heterogeneous approaches. *Knowledge-Based Systems*. 118:124–139.
- Shahrjooihaghighi A, Frigui H, Zhang X, Wei X, Shi B, Trabelsi A. 2018. An Ensemble Feature Selection Method for Biomarker Discovery. *Proc IEEE Int Symp Signal Proc Inf Tech*. 2017:416–421.

- 1
2
3 Smith LK, Lelas JL, Kerrigan DC. 2002. Gender differences in pelvic motions and
4 center of mass displacement during walking: stereotypes quantified. *Journal of*
5 *women's health & gender-based medicine.* 11(5):453-458.
6
7
8
9
10 Stokes VP, Andersson C, Forsberg H. 1989. Rotational and translational movement
11 features of the pelvis and thorax during adult human locomotion. *J Biomech.*
12 22(1):43-50.
13
14
15
16
17 Tang J, Alelyani S, Liu H. 2014. Feature Selection for Classification: A Review. In:
18 Aggarwal CC, editor. *Data Classification: Algorithms and Applications.* New York:
19 Chapman and Hall/CRC; p. 37–64.
20
21
22
23
24 Teufl W, Taetz B, Miezal M, Lorenz M, Pietschmann J, Jöllenbeck T, Fröhlich M,
25 Bleser G. 2019. Towards an Inertial Sensor-Based Wearable Feedback System for
26 Patients after Total Hip Arthroplasty: Validity and Applicability for Gait
27 Classification with Gait Kinematics-Based Features. *Sensors.* 19.
28
29
30
31
32
33
34
35
36 Victo Sudha George G, Cyril Raj V. 2015. Accurate and Stable Feature Selection
37 Powered by Iterative Backward Selection and Cumulative Ranking Score of
38 Features. *Indian Journal of Science and Technology.* 8:1–6.
39
40
41
42 Wang J, Xu J, Zhao C, Peng Y, Wang H. 2019. An ensemble feature selection method
43 for high-dimensional data based on sort aggregation. *Systems Science & Control*
44 *Engineering.* 7:32–39.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 **Tables**
5

6 **Table 1.** Results for stance and gait. RBF SVC is used for classification. 100 % threshold corresponds to all
7 preselected features. Number of features n_{feat} , Mean Matthews correlation coefficient M_{MCC} , M_{F1} mean F1
8 score, mean Precision-Recall Area Under the Curve M_{AUC} . The true positives and true negatives are
9 highlighted with gray back color in the Confusion Matrix (CM).
10

11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Threshold	Stance						Gait					
	n_{feat}	M_{MCC}	M_{F1}	M_{AUC}	CM		n_{feat}	M_{MCC}	M_{F1}	M_{AUC}	CM	
5 %	10.00 ± 0	0.48 \pm 0.13	0.66 \pm 0.09	0.73 \pm 0.11	99 17	36 53	62.20 \pm 0.84	0.92 \pm 0.03	0.95 \pm 0.02	0.98 \pm 0.02	132 4	3 66
10 %	20.00 ± 0	0.58 \pm 0.09	0.71 \pm 0.09	0.83 \pm 0.09	112 18	23 52	125.00 ± 2.12	0.95 \pm 0.04	0.96 \pm 0.03	0.99 \pm 0.01	133 3	2 67
20 %	40.00 ± 0	0.62 \pm 0.07	0.74 \pm 0.04	0.82 \pm 0.10	117 18	18 52	250.80 ± 4.02	0.96 \pm 0.04	0.97 \pm 0.03	0.99 \pm 0.01	133 2	2 68
50 %	101.0 0 \pm 0	0.61 \pm 0.09	0.73 \pm 0.08	0.84 \pm 0.05	120 21	15 49	627.60 ± 9.89	0.96 \pm 0.02	0.97 \pm 0.02	0.99 \pm 0.01	134 3	1 67
100 %							1256.00 ± 19.60	0.97 \pm 0.03	0.98 \pm 0.02	1.00 \pm 0.00	134 2	1 68
all initial	202	0.61 \pm 0.06	0.72 \pm 0.07	0.84 \pm 0.06	122 23	13 47	2021	0.96 \pm 0.04	0.97 \pm 0.03	1.00 \pm 0.01	134 3	1 67

Figures

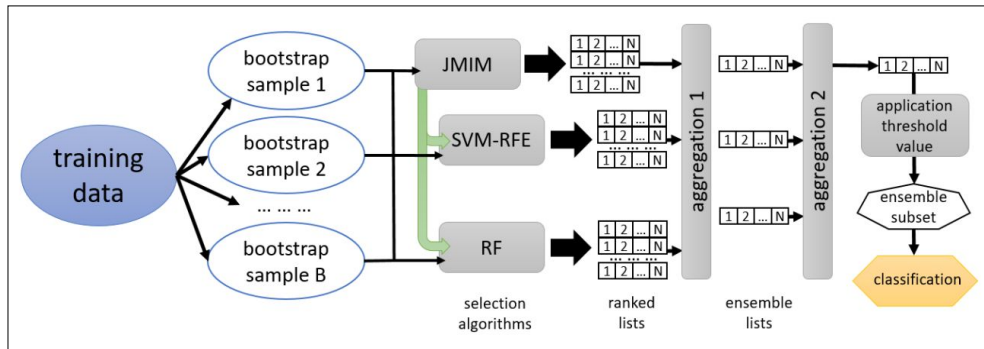


Figure 1. Ensemble design. JMIM = Joint Mutual Information Maximization filter; SVM-RFE = Recursive Feature Elimination with Support Vector Machine; RF = Random Forest

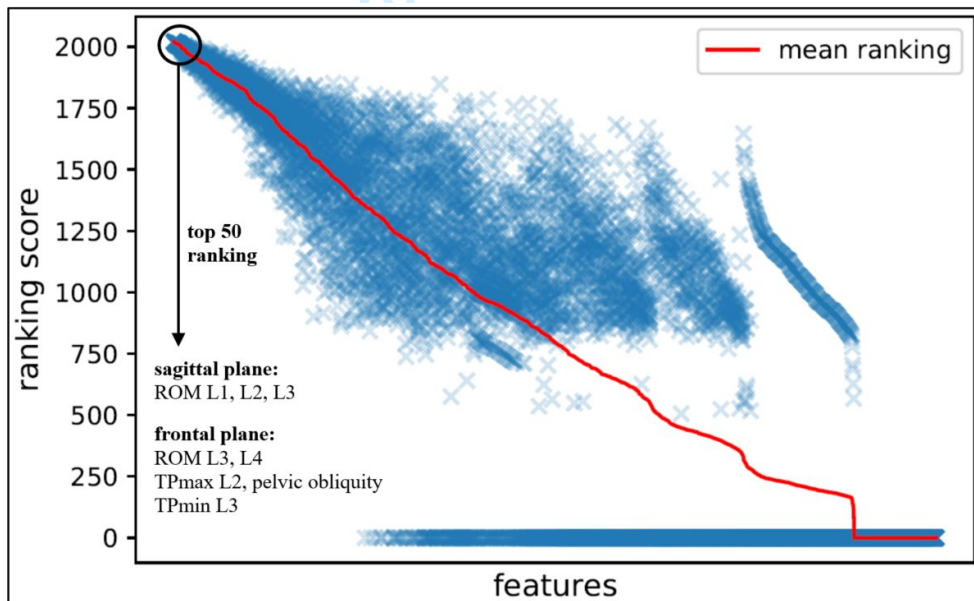


Figure 2. Feature rankings of each cross validation fold and the most stable selected features (selected in all folds) of the top 50 rankings for the gait data. ROM = range of motion; TPmax, TPmin = point in time of the maximal respectively minimal deflection

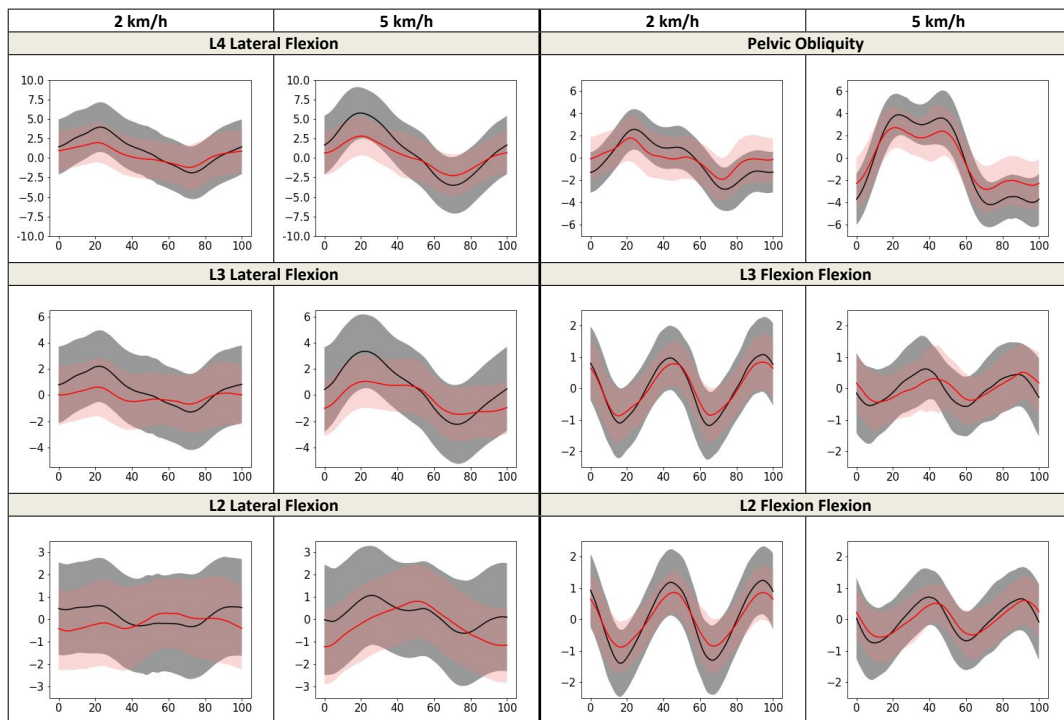


Figure 3. Gait waveforms for the different walking speeds; mean values and standard deviations are presented; L2 movement in sagittal plane is centered to visualize range of motion differences; black = female, red = male

URL: <http://mc.manuscriptcentral.com/gcmb>

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review Only

Supplemental online material

Table. Used variables for automated feature extraction. Variables were measured in sagittal, frontal and transversal movement direction. Age was additionally included.

Variable	Plane	Description
VP/C7 (Vertebra Prominens / 7. cervical vertebral body) T1 – T12 (Thoracic spine) L1 – L4 (Lumbar spine)	sagittal: Vertebral sagittal Flexion and Extension [°]	sagittal: The parameter describes the inclination of the calculated vertebra in space (relative to a plumb / gravity line) as seen from a left view. The angle (in degrees) is calculated from the projection of the vertebra in a sagittal plane (rotation and lateral flexion are ignored). A positive value means a forward tilt of the vertebra (flexion).
	frontal: Vertebral Lateral Flexion [°]	frontal: The parameter describes the lateral inclination of the vertebra in space (relative to a plumb / gravity line) as seen from a posterior-anterior view. The angle (in degrees) is calculated from the projection of the vertebra in the coronal plane (rotation and sagittal extension / flexion are ignored). A positive value means a tilt of the vertebra to the left (lateral flexion left).
	transversal: Vertebral Rotation [°]	transversal: The vertebral rotation describes the rotation of a vertebra in the transversal plane (relative to the neutral pelvis). A positive value means a vertebra is rotated to the left (counterclockwise) when seen from behind. The rotation of vertebral bodies happens in situ, therefore the direction of rotation between surface and vertebral rotation changes. Hence, a surface rotation to the right, mathematically represented with a +, becomes a vertebral body rotation to the left. This is due to the calculation process in which a vector is used that points from Processus spinosus towards the middle of the vertebral body meaning that the surface rotation changes its direction within the vertebral body.
Pelvis	Pelvic Obliquity [°]	pelvic obliquity: A line is drawn from DL to DR (left and right dimple) and is compared to a horizontal line representing the horizon. The angle (in degrees) between them is measured. A positive value means that the right pelvis is elevated
	Pelvic Torsion (dimples) [°]	The parameter describes the torsion of the surface normals on the two lumbar dimples.
	Pelvic Inclination (dimples) [°]	The parameter describes the mean vertical torsion of the two surface normals on DL and DR
	Pelvic Inclination (symmetry line) [°]	The parameter describes the angle of the vertical components of the surface normals on point DM (Dimple Midpoint) based on the horizontal
	Pelvic Rotation [°]	The pelvic rotation is the rotation in the transversal plane of the right dimple relative to a reference coronal plane that is defined from the system setup, perpendicular to the camera-projection axis. A positive value means the pelvis is rotated to the left when seen from behind (the value is corrected * (-1)).

Thoracic Kyphosis and Lumbar Lordosis Angle	Thoracic Kyphosis Angle (ICT-ITLmax) [°]	The thoracic kyphosis is calculated by the angle formed between the two surface tangent lines of the ICT (inflection point between cervical and thoracic spine) and ITL (inflection point between thoracic and lumbar spine), based on the spinous process line
	Thoracic Kyphosis Angle (VP-ITL) [°]	- calculated by VP and ITL
	Thoracic Kyphosis Angle (VP-T12) [°]	- calculated by VP and T12
	Lumbar Lordosis Angle (ITL-ILSmax) [°]	The lumbar lordosis is calculated by the angle formed between the two surface tangent lines of the ITL and ILS (inflection point between lumbar spine and sacrum), based on the spinous process line.
	Lumbar Lordosis Angle (ITL-DM) [°]	- calculated by ITL and DM
	Lumbar Lordosis Angle (T12-DM) [°]	- calculated by T12 and DM
Other	Trunk Inclination (VP-DM) [°]	The trunk inclination is defined by the angle between the plumb line dropped from VP and the line between VP and DM (in degrees). A positive value means VP is more ventral than DM.
	Trunk Imbalance (VP-DM) [°]	The trunk imbalance in the coronal plane is defined by the angle between the plumb line dropped from VP and the line between VP and DM (in degrees). A positive value means that the patient is leaning to the right.
	Trunk Torsion [°]	The maximum value of the horizontal components on VP compared with the value of the horizontal components of the surface normals on the symmetry line on DM. The degree is the unit for the trunk torsion. A positive value for trunk torsion signifies that the body at the height of the VP is turned more sharply to the right than at the height of the DM.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

	Shoulder Obliquity [°]	shoulder obliquity: A line between those two points is drawn and compared with the horizontal line representing the horizon (perpendicular to the gravity line). The angle (in degrees) between them is measured. A positive value means that the right shoulder is elevated.
	Shoulder Rotation (SL/SR-Pelvis) [°]	The shoulder rotation is the rotation in the transversal plane of the right shoulder point (SR) relative to a reference coronal plane through the left shoulder point (SL) that is defined from the system setup, perpendicular to the camera-projection axis. The pelvic rotation is then subtracted from this value, to get shoulder rotation relative to pelvic rotation.

3.3 3. Publikation



Interpretability of Input Representations for Gait Classification in Patients after Total Hip Arthroplasty

C. Dindorf, W. Teuffl, B. Taetz u. a., “Interpretability of Input Representations for Gait Classification in Patients after Total Hip Arthroplasty,” *Sensors*, Jg. 20, Nr. 16, S. 1–14, 2020. DOI: [10.3390/s20164385](https://doi.org/10.3390/s20164385)

Zusammenfassung

Hintergrund: Viele ML-Modelle weisen Black Box Charakteristiken und damit einen Mangel an Transparenz, Interpretierbarkeit und Vertrauenswürdigkeit auf. Dies schränkt ihre praktische Anwendung im klinischen Kontext stark ein. Zur Überwindung dieser Einschränkungen hat XAI vielversprechende Ergebnisse gezeigt.

Methodische Forschungsfragen: Welchen Einfluss haben unterschiedliche Eingabedarstellungen auf die Genauigkeit, Interpretierbarkeit und klinische Relevanz eines trainierten Modells unter Verwendung von XAI-Methoden? Führt die Anwendung von XAI-Methoden auf Modelle, die auf unterschiedliche Input-Repräsentationen trainiert wurden, zu kongruenten Ergebnissen und daher insgesamt zu einer besseren Interpretierbarkeit?

Fachspezifische Forschungsfragen: Welche Gangmuster (lokal und global) sind am bedeutendsten für die Klassifikation als Gesunder bzw. als Patient nach Hüft-TEP?

Methodik: Der Gang von 27 gesunden Probanden und 20 Probanden nach Hüft-TEP wurde mit einem IMU-basierten System aufgezeichnet. Zur Klassifikation wurden drei verschiedene Eingabedarstellungen (Wellenformdaten, automatisch extrahierten Features und einfachen deskriptiven Features) verwendet. LIME wurde für die Modellinterpretation verwendet.

Ergebnisse: Die beste Genauigkeit wurde mit automatisch extrahierten Merkmalen erreicht (mittlere Genauigkeit $\text{Macc} = 100\%$), gefolgt von Merkmalen, die auf einfachen deskriptiven Statistiken ($\text{Macc} = 97,38\%$) und Wellenformdaten ($\text{Macc} = 95,88\%$) basieren. Die Interpretationen für die einzelnen Input-Repräsentationen stimmen überwiegend überein. Global gesehen waren für diese spezifische Klassifikationsaufgabe insbesondere die sagittale Bewegung von Hüfte, Knie und Becken sowie die transversale Bewegung des Sprunggelenks wichtig.

Fazit: Die aktuelle Arbeit zeigt, dass die Art der Eingangsrepräsentation maßgeblich die Interpretierbarkeit sowie die klinische Relevanz bestimmt. Vorteilhaft erscheint ein kombinierter Ansatz mit unterschiedlichen Darstellungsformen. Die Ergebnisse könnten Ärzten und Therapeuten helfen, individuelle pathologische Gangmuster zu finden und zu behandeln.



Article

Interpretability of Input Representations for Gait Classification in Patients after Total Hip Arthroplasty

Carlo Dindorf ^{1,*} , Wolfgang Teufl ² , Bertram Taetz ² , Gabriele Bleser ²
and Michael Fröhlich ¹

¹ Department of Sports Science, Technische Universität Kaiserslautern, Erwin-Schrödinger-Str. 57, 67663 Kaiserslautern, Germany; michael.froehlich@sowi.uni-kl.de

² Junior Research Group wearHEALTH, Technische Universität Kaiserslautern, Gottlieb-Daimler-Str. 48, 67663 Kaiserslautern, Germany; teufl@cs.uni-kl.de (W.T.); taetz@cs.uni-kl.de (B.T.); bleser@cs.uni-kl.de (G.B.)

* Correspondence: carlo.dindorf@sowi.uni-kl.de; Tel.: +49-631-205-5172

Received: 1 July 2020; Accepted: 4 August 2020; Published: 6 August 2020



Abstract: Many machine learning models show black box characteristics and, therefore, a lack of transparency, interpretability, and trustworthiness. This strongly limits their practical application in clinical contexts. For overcoming these limitations, Explainable Artificial Intelligence (XAI) has shown promising results. The current study examined the influence of different input representations on a trained model's accuracy, interpretability, as well as clinical relevancy using XAI methods. The gait of 27 healthy subjects and 20 subjects after total hip arthroplasty (THA) was recorded with an inertial measurement unit (IMU)-based system. Three different input representations were used for classification. Local Interpretable Model-Agnostic Explanations (LIME) was used for model interpretation. The best accuracy was achieved with automatically extracted features (mean accuracy $M_{acc} = 100\%$), followed by features based on simple descriptive statistics ($M_{acc} = 97.38\%$) and waveform data ($M_{acc} = 95.88\%$). Globally seen, sagittal movement of the hip, knee, and pelvis as well as transversal movement of the ankle were especially important for this specific classification task. The current work shows that the type of input representation crucially determines interpretability as well as clinical relevance. A combined approach using different forms of representations seems advantageous. The results might assist physicians and therapists finding and addressing individual pathologic gait patterns.

Keywords: explainable artificial intelligence; inertial measurement unit; machine learning; biomechanics; gait; total hip replacement

1. Introduction

Identification and discrimination of group differences are important aspects of biomechanical research [1,2]. With modern movement tracking systems, a huge amount of data is available (big data) [3,4]. The progressive development of motion analysis systems based on inertial measurement units (IMUs) contributes in particular to the generation of large amounts of data, because they make valid and reliable biomechanical data easily accessible [5]. This provides the potential to generate new knowledge and a better understanding of human biomechanics. However, classical inference-based statistical methods show limited capabilities in analyzing the emerging—often complex and multivariate—amounts of data and, thus, machine learning models have gained importance [3,4,6].

Many studies have obtained promising results for the classification of pathological movements (e.g., hip osteoarthritis [7], after stroke [8], and Parkinson's Disease [9]). However, many of the machine learning models show black box characteristics and a lack of transparency [10]. For example, this does not comply with the requirements of the European General Data Protection Regulation (GDPR,

EU 2016/679) [11] and strongly limits practical application in clinical contexts. In order to improve interpretability, transparency, and clinical relevancy, the models themselves (model interpretability) as well as the practical usefulness and interpretability of the input variables used for modeling (feature interpretability) must be taken into account.

The input representation plays an important role in classification accuracy [12], as well as in interpretability, clinical relevancy, and comparability with previous research. Different feature extraction approaches can be found in the literature:

- (i) Many studies have used simple descriptive statistics of the gait waveforms such as peak values, range of motion, or respective side differences [13,14]. They are straight forward to interpret and are often mentioned in the literature for describing gait characteristics. However, it is unclear if important information would be a priori discarded, and model performance is consequently negatively affected. A further limitation is the dependence on expert or prior knowledge.
- (ii) An alternative approach, which is independent of prior knowledge, is the use of entire concatenated waveforms as input features [7,15,16]. This allows for an interpretation of group differences through the determination of important areas of the waveforms. However, it is unclear if this shows better discriminative power compared with the abovementioned extracted statistical features and, therefore, enhances classification performance. Correlations and redundancy of the inputs may further be problematic.
- (iii) Lastly, automated feature extraction using a vast amount of possibly meaningful statistics can be applied [17]. Feature extraction algorithms such as *tsfresh* [18] or *featuretools* [19] can be used for this. However, the extracted features are often nested, complex, and hard to interpret, therefore showing limited comparability with the literature, which results in questionable clinical relevance.

Dimensionality reduction methods can be applied before classification to compress the data onto a new feature space, with the aim of capturing as much variance as possible of the original data [15,20]. Nevertheless, this makes interpretation harder because the components of the new feature space must be interpreted first. For all mentioned approaches, it should be noted that feature selection may further improve a model's accuracy, reduce computing power, prevent overfitting, and improve interpretability [21].

Regarding model interpretability, on the one hand, complex machine learning models (e.g., deep neural networks) often achieve more accurate results compared with simpler models (e.g., decision trees). On the other hand, complexity leads to reduced transparency and interpretability (trade-off between model accuracy and interpretability) [22,23], which results in the black box characteristic of many models [10]. For the user, it is therefore hard to trust in the model and its decision because it is opaque as to what the model has really learned and why it makes certain decisions [24]. These factors currently limit practical application in clinical contexts [25].

Explainable Artificial Intelligence (XAI) has gained great interest in recent years and offers methods for increasing the transparency and trustworthiness of black box models [10]. Local Interpretable Model-Agnostic Explanations (LIME) [26], SHapley Additive exPlanations (SHAP) [27], and Deep Learning Important Features (DeepLIFT) [28] should be mentioned as prominent interpretation tools. For example, LIME performs an approximation of a single prediction of a black box model with a simpler interpretable model to explain how a black box model makes a single prediction [26]. Initial studies applying XAI methods with clinical data have produced promising results [16]. Therefore, XAI methods seem promising for making machine learning models more useable in practical clinical applications.

The application of XAI methods in the context of biomechanical data analysis is a young field of research. Practical applications in clinical contexts are still very rare. There is need for research in order to assess potentials, show limitations, and identify further research directions in the biomechanical and clinical domain. As a step towards practical clinical application, the present work focuses on the input representation. Different kinds of input representations take different perspectives on the data and provide different insights. Yet, they all show benefits and limitations regarding model accuracy,

interpretability, and clinical relevancy. The application of XAI methods based on different input representations could, therefore, lead to new insights and an even better understanding of the data. To the best of our knowledge, no similar comparison has been performed so far. Therefore, we wanted to check if the application of XAI methods on models, trained on different input representations, leads to congruent results and provides, taking them all together, better interpretability. For this reason, we used a highly relevant use case example and compared the gait kinematics, measured by means of an IMU system, of patients after total hip arthroplasty (THA) (most important surgery for the treatment of degenerative hip osteoarthritis [29]) with that of a group of healthy subjects.

2. Materials and Methods

2.1. Subjects, Data Acquisition, and Data Preprocessing

For the present study, the IMU-based gait data of a healthy sample from [30] and a sample of patients after THA from [13] were employed. The studies were approved by the ethical committee of the Technische Universität Kaiserslautern and the Universität Paderborn and met the criteria of the Declaration of Helsinki. After receiving all relevant study information, the participants signed an informed consent form for the study, including permission to publish the data.

In the mentioned studies, 27 healthy subjects (14 females, 13 males; age: 24.63 ± 2.80 years; weight: 70.44 ± 12.56 kg; height: 1.76 ± 0.09 m) and 20 subjects approximately 2 weeks after THA (13 females, 7 males; age: 57.79 ± 7.41 years; weight: 83.89 ± 17.22 kg; height: 1.73 ± 0.08 m) performed a 6 min walking test. The accelerometer and gyroscope raw data during the gait were recorded by means of seven MTw Awinda IMUs (Xsens Technologies BV, Enschede, The Netherlands) attached to the segments of the lower extremities according to [13].

The IMU raw data were then processed using a recently developed sensor fusion algorithm based on an extended Kalman filter approach [31,32]. Using this algorithm, relative segment orientations were estimated exploiting the knowledge of a biomechanical model (i.e., segment lengths, joint centers, virtual anatomical landmarks, and an IMU to segment calibration). From this, it was then possible to interpret these angles as anatomically meaningful joint angles and further estimate gait-specific events (i.e., initial contact and terminal contact), and calculate based on those spatiotemporal parameters. The event detection and the estimation of the joint angles of the lower body using the mentioned algorithm were validated in recent publications [30,33,34].

Consequently, the following parameters were calculated based on the IMU data: the hip, knee, and ankle joint angle waveforms as well as the global pelvic motion in the sagittal, frontal, and transversal plane.

The joint angle waveforms of all subjects were divided into gait cycles (GCs) using the initial contact information. The initial contact was detected using a kinematics-based approach according to [34]. The gait cycles were then checked for outliers using the mean gait cycle duration (i.e., the stride time) of a subject ± 2 times standard deviation. Gait cycles displaying a duration higher or lower than these thresholds were excluded from the following evaluation. The remaining gait cycles of all subjects were normalized to 100-time steps using cubic spline interpolation. Twenty gait cycles were extracted for every subject. The original sidewise (left, right) consideration was transformed into a differentiation between affected and unaffected sides.

Three different forms of input representations were calculated and used as input vectors: entire waveforms (V_waves), discrete features based on simple descriptive statistics (V_simple), and automatically extracted features (V_tsfresh) (see Table 1).

To analyze the influence of different data preprocessing, training and test data were standardized based on the respective training set, with three different scaling approaches:

- without data scaling,
- removal of the mean and scaling to unit variance (StandardScaler),
- scaling to a feature range between 0 and 1 (MinMaxScaler).

The processing of the IMU raw data and the calculation of the joint angle waveforms were conducted in C++. The segmentation of the joint angle waveforms as well as the outlier detection were conducted in Matlab 2019b (Mathworks, Inc., Natick, MA, USA). Further calculations were performed in Python (Python Software Foundation, Wilmington, DE, USA).

Table 1. Input vectors used for modeling. Twenty gait cycles (GCs) per subject were extracted. To reduce the number of irrelevant features of V_simple and V_tsfresh, the fresh-algorithm (FeatuRe Extraction based on Scalable Hypothesis tests) [18] was applied as a filter. ROM = range of motion

Abbreviation	Description	Size (GC × Feature)
V_waves	Concatenated time-normalized GC for the measured variables.	940 × 2100
V_simple	Calculated features based on simple descriptive statistics which are commonly mentioned in the literature [35,36]. Maxima, minima, and ROM for every variable as well as the difference between affected and unaffected sides for the respective variables were calculated.	940 × 74
V_tsfresh	Automated feature extraction with the tsfresh algorithm [18].	940 × 8349

2.2. Model Training and Classification

The following procedure was applied for all input vectors: fivefold cross-validation was performed with each test set consisting of the gait cycles respective to the extracted features of four subjects of each class (8 subjects in total with 8×20 GCs). Every subject of the THA group was used once for testing. Due to imbalanced class distribution in the training set, synthetic minority oversampling [37] was for each fold performed on the respective training set. Random Forest (RF), linear Support Vector Machine (SVM), SVM with radial basis function kernel (rbf), and a neural network (multilayer perceptron, MLP) were applied for classification with the standard parameters of Scikit-learn [38].

2.3. Model Interpretation

An a priori specification of the best performing classification algorithm for a certain task is in most cases not possible. For practical application, interpretation tools should, therefore, be generally applicable (model agnostic) and not dependent on a specific algorithm (model specific). Further, subjects show individual gait characteristics [39,40]. In the context of personalized medical treatment, it becomes, therefore, especially important to understand why a model made a specific decision for a single instance/GC or for a single subject, respectively. Therefore, local interpretability gains importance.

For that reason, the model-agnostic interpretation tool LIME [26] was used for model interpretation. To explain a black box model, LIME performs an approximation of a single prediction of a black box model with a simpler interpretable model (e.g., decision tree, linear model). The simpler model will probably not perform a globally faithful approximation of the complex model but will perform well locally. Therefore, LIME is based on local interpretations of single instances of interest. To explain a single prediction of a black box model, the instance of interest is chosen and data points around it are generated through perturbation. These data points are predicted with the black box model and weighted by their proximity to the selected instance. Finally, a simpler model is learned on the weighted data points and used for explaining the prediction [26].

For the special case of serial data (gait waveforms), LIME for Time was developed [41]. It is based on the idea of using LIME in the context of image data. For image data, Superpixels (connected pixels of one color) are used for data variation because variations of individual pixels would hardly change a model's prediction. As an equivalent to Superpixels, LIME For Time performs data variation of parts of a series by replacing them, for example, by noise or the entire mean of the series. The approach makes it possible to identify areas of serial data that are important for the classifier in its prediction. The algorithm was applied for interpretation of the concatenated waveforms (V waves). Total mean was used for perturbation.

For each input vector, the best performing model was used for interpretation with LIME. Local interpretations are presented as exemplars. To determine if effects due to group differences were important for class selections across multiple instances, global interpretations are presented (indication of generality). They were calculated by mean aggregation of the absolute weight values of the local results (5 GCs of each subject to reduce computing power) of the LIME algorithm (see, e.g., [42]). Statistical Parameter Mapping (SPM) [43] was used to verify the results from a statistical perspective.

3. Results

3.1. Classification Results

Classification results are presented in Figure 1. The overall best accuracy was obtained for the use of the V_tsfresh input vector, SVM linear (MinMaxScaler, StandardScaler), and MLP (MinMaxScaler) (mean accuracy $M_{acc} = 100\%$). The best result for the waveform data V_waves was obtained with SVM linear without normalization ($M_{acc} = 95.88\%$). SVM linear with MinMaxScaler performed best for the gait-specific data ($M_{acc} = 97.38\%$). The best performing models are used in the following subsections for interpretation.

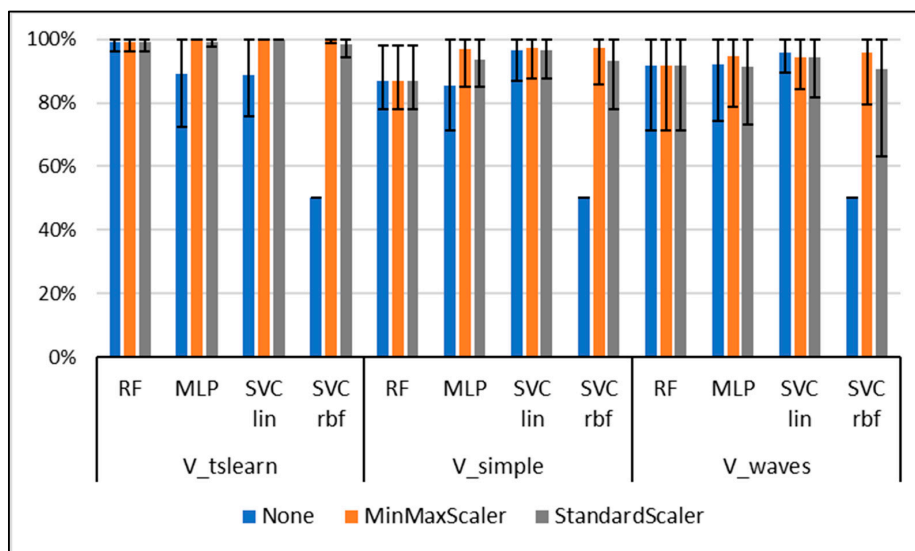


Figure 1. Classification accuracy (mean, min, max value) for the different input vectors, classification algorithms, and normalization approaches over the 5-fold cross-validation. RF = Random Forest; MLP = multilayer perceptron; SVM lin = Support Vector Machine with linear kernel; SVM rbf = Support Vector Machine with radial basis function kernel.

3.2. Model Interpretation Based on Waveforms

The global results for the comparison of patients and healthy subjects based on the concatenated waveforms (V waves) are presented in Figure 2. SPM indicates statistical group differences for most of the movements ($p < 0.05$). LIME indicates a high global effect for slices representing the ankle rotation as well as knee and hip movement in the sagittal plane. The ankle rotation of the affected side, especially during initial contact and midstance, plays an important role. For the knee joint of the affected as well as unaffected side, the important slices represent the maximal flexion and the maximal extension. Maximal flexion and maximal extension are further relevant for the hip movement in the sagittal plane of the affected side. Group differences for the respective slices are all significant.

Model explanations for single instances (gait cycles) regarding the waveform data are presented in Figure 3. It is noticeable that for both correctly classified GCs, few slices indicate an effect towards the other class. Additionally, one misclassified GC of a patient is shown. Most of the slices with the highest effect indicate an effect towards the class of healthy subjects (e.g., ankle rotation, knee flexion). Similar movement patterns compared to healthy subjects are noticeable for the respective slices.

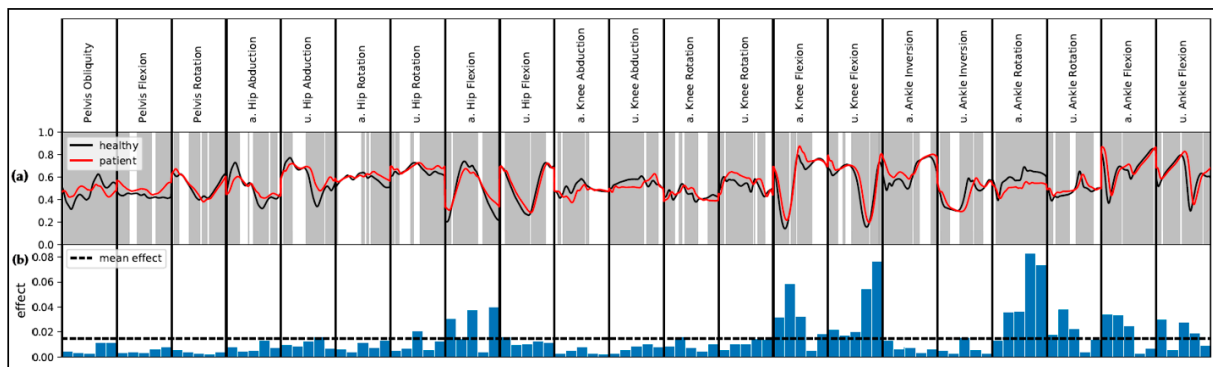


Figure 2. Global results for the use of the waveform data (V_waves). For visualization reasons, data are separately scaled to a range from 0 to 1 for each variable. (a) Mean movements for healthy subjects and patients after total hip arthroplasty (THA). The grey areas indicate statistical difference ($\alpha = 0.05$) according to Statistical Parameter Mapping. (b) Aggregated Local Interpretable Model-Agnostic Explanations (LIME) results as mean absolute effect. Abbreviations: a. = affected side; u. = unaffected side.

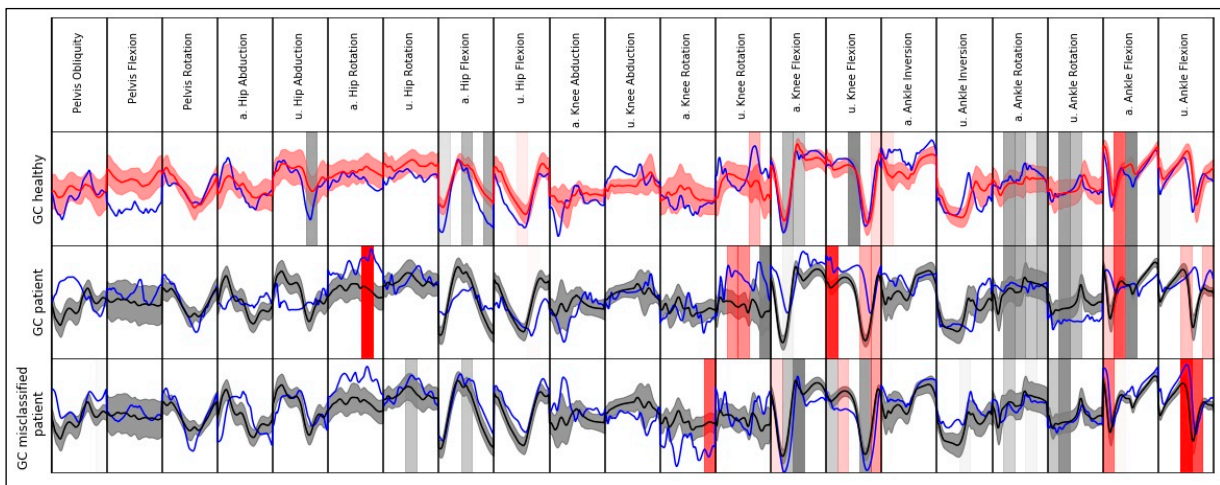


Figure 3. Exemplary local model interpretation for single gait cycle (GC) (instance) of a patient with THA, a healthy subject, and a GC of a patient classified as healthy (blue lines). The instances are plotted against the mean value and standard deviation of the other class (black = healthy, red = patient). For better visualization, the top 20 slices with the highest absolute effect are displayed for each instance (grey vertical span = effect towards class of healthy subjects, red vertical span = effect towards class of patients with THA). The color saturation indicates the effect size. For visualization reasons, data are separately scaled to a range from 0 to 1 for each variable. Abbreviations: a. = affected side; u. = unaffected side.

3.3. Model Interpretation: Discrete Features

LIME results for the use of the V simple input vector are presented in Figure 4. The analyzed instances correspond to the interpreted waveform instances displayed in Figure 3. Even though different input features were used, the model performed the same misclassification. The features with the highest effect globally seen are based on hip, knee, and pelvic sagittal motion and ankle rotation in the transversal plane.

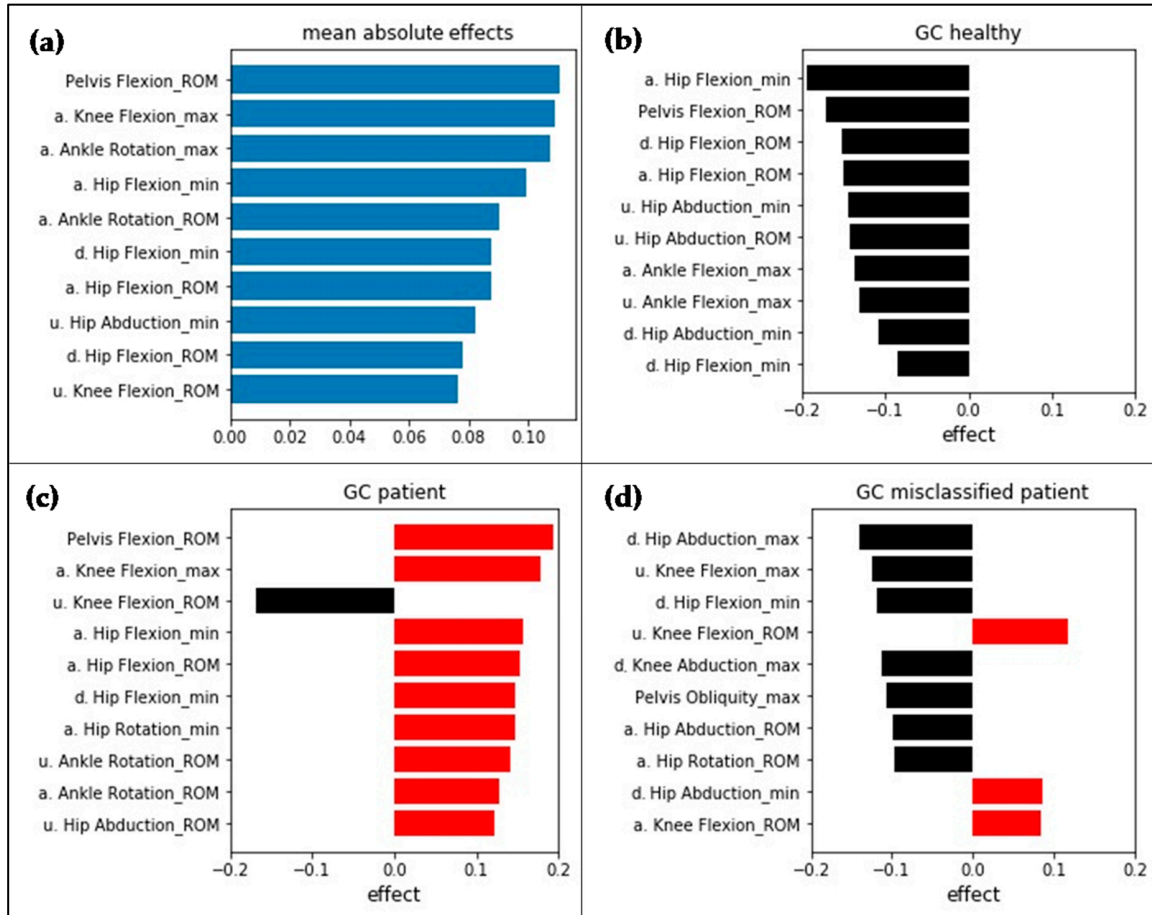


Figure 4. (a) Global LIME results as mean absolute effects. The analyzed instances (b–d) correspond to the instances displayed in Figure 3. Negative values indicate an effect towards the class of healthy subjects (black), positive effects towards the class of patients after THA (red). Feature abbreviations: a. = affected side; u. = unaffected side; d. = difference affected, unaffected side.

The mean absolute effects for the use of the $V_{tsfresh}$ input vector are shown in Figure 5. Features based on knee and ankle sagittal movement as well as hip, knee, and pelvic transversal movement show the highest effect. For calculation of the features with the highest effect, the operation “large_standard_deviation” was used. The result of the operation is a Boolean variable, denoting if the standard deviation of the series (in this case, knee flexion) is higher than “r” times the range of motion (ROM).

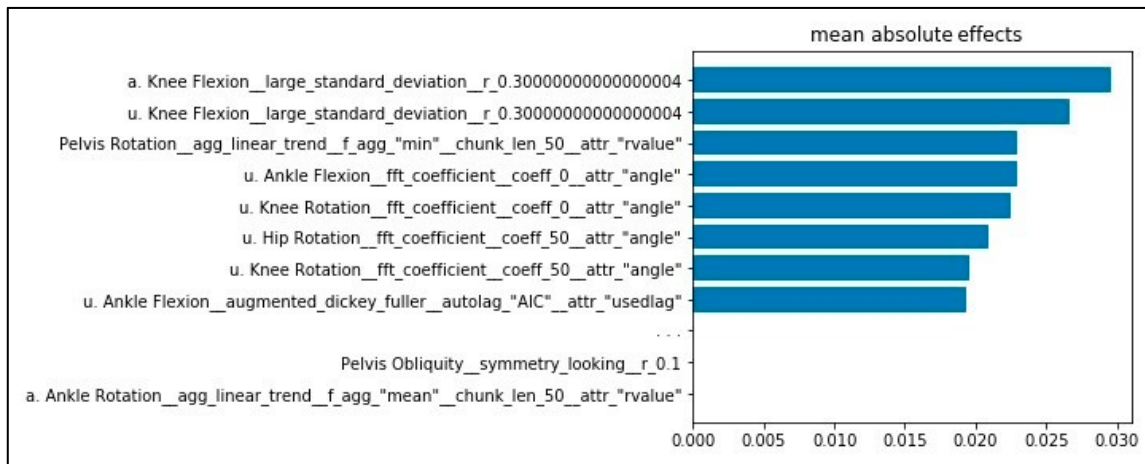


Figure 5. Mean absolute effects for the $V_{tsfresh}$ input vector determined using LIME. Results are based on SVM linear with MinMaxScaler. Feature labels are according to the automated feature extraction algorithm $tsfresh$ [18].

4. Discussion

Very good results regarding accuracy were obtained for the classification of patients with THA and healthy subjects. Three different input vectors were used. The best classification results were obtained when using automatically extracted features ($V_{tslearn}$). However, with the use of simple descriptive statistics (V_{simple}) or the waveform data (V_{waves}), very good classification performances with only a slight reduction in accuracy were obtained. It can be assumed, in line with the previous study [13], that even simple descriptive statistics have high discriminative power, appropriately map the classes, and, therefore, slightly outperform the pure use of the waveforms.

In line with other works [8], we can demonstrate the superior performance of linear SVM in the context of gait classification. In most cases, linear SVM showed the best results and outperformed the more complex and computationally expensive models (e.g., MLP). The main focus of the current work was on interpretability; therefore, no extensive parameter tuning of the models was performed. It might be possible to obtain the same or even better results using the more complex models. However, the cost would be an extensive and time-consuming parameter tuning.

With the use of the model agnostic method LIME, it was possible to gain insight into how the models made their decisions and to increase interpretability. Through the selection of a maximum number of features to be displayed by LIME, it is possible to provide different interpretation levels for different contexts. Using complex features ($V_{tsfresh}$) leads to slightly higher performance with the cost of interpretability and comparability with previous research. The reason for this is that the operations used for feature calculation often make it difficult to attribute class differences to the direct movements because they describe the original movements on a more abstract level compared with simpler and, in a biomechanical context, more commonly used descriptive statistics. Therefore, their usage in clinical contexts is questionable. Thus, the following interpretation mainly focuses on the input representations V_{waves} and V_{simple} .

To evaluate the validity/plausibility of the effects and the identified group differences, three aspects are addressed as follows: (i) asymmetric gait patterns are often prevalent after THA [15,44]. Not only are the operated joint and surrounding structures affected but also the contralateral side [36]. However, the main difference of the gait characteristics between healthy subjects and patients after THA should be seen for the affected side. LIME results should therefore generally emphasize an effect for features regarding the affected side. For both input vectors, most of the time, this holds true. Yet, some features with the highest effect map gait differences for the unaffected side (e.g., V_{simple} : unaffected hip abduction) or for both affected and unaffected sides (e.g., waveform data: unaffected/affected knee flexion/extension). A possible reason for this might be that the trained model compares the affected

and unaffected sides. Regarding the sample of subjects, age-related differences in gait, which possibly led to a group separation with regard to the unaffected side (see, e.g., classification young, old [45]), should not go unnoticed and possibly have further influence.

(ii) Another aspect which speaks for the validity of the results is that most results are congruent between $V_{tsfresh}$ and V_{wave} . For both input vectors, ankle rotation as well as sagittal knee and hip movement plays an important role. Nevertheless, pelvic flexion (ROM) only shows a high effect for the use of $V_{tsfresh}$. Therefore, application of different input representations can provide new insights and show effects, which were possibly not detected with only the use of a single input vector.

Using only the waveform data, interactions between different slices are opaque. For example, maximal flexion and maximal extension of hip movement in the sagittal plane is highly relevant regarding the waveform data. Looking at $V_{tsfresh}$, the ROM of the hip movement in the sagittal plane shows a high effect. This might indicate that the slices mapping maximal flexion and maximal extension interact. Comparing the results for different input vectors might, therefore, be useful for a better understanding of possible interactions.

(iii) Finally, LIME results are discussed taking into account previous research. Overall, a ground truth of the automatically determined explanations is missing, and it therefore becomes more difficult to evaluate if the results are meaningful and appropriately map gait characteristics. In the literature, gait patterns of patients with THA are mostly described using simple descriptive statistics comparable with the V_{simple} input vector [36,46,47]. Consequently, it is harder to directly compare the effects respective to the determined relevant regions regarding the waveform data (V_{wave}) with previous works. Further, it is not possible to evaluate all our findings with previous research because the literature often focuses on a few gait characteristics to describe group differences. In agreement with our findings, previous research reports a reduced ROM for knee and hip movement of the operated side in the sagittal plane compared with healthy subjects [36,44,46–48]. Further, altered postoperative ankle rotation [46] and increased sagittal pelvic movement compared with healthy subjects was found [46,47].

Regarding the current state of research, there are no objective criteria to evaluate interpretability [42]. Subjective ratings from end-users or task performance might be possible ways for evaluation [22]. Regarding interpretability and the goal of making machine learning applicable in practical clinical contexts, it is important to make a distinction between the interpretability of the model itself and clinical interpretability. Interpretability of a model describes the degree to which humans can understand a decision from a model by its causes and, hence, why a model made a certain decision [49]. The usage of XAI tools in this study made it possible to understand why a model made a certain decision through revealing the effects for features or areas of waveforms which were important for a certain prediction. However, with this alone, it is unclear if the identified effects are interpretable, relevant, and usable in clinical contexts. In this respect, the current work emphasizes the importance of the input representation because it highly influences the interpretability and usability in clinical contexts. In the current case, automatically extracted features show no significantly better classification accuracy, which could possibly justify a loss of interpretability. Consequently, waveform data and simple descriptive statistics should be used for practical applications because effects can be traced back to variables that are mentioned in the literature and already used in clinical contexts for describing biomechanical differences.

In cases where classification is harder, a combination of expert-knowledge-based features extended by the best performing automated extracted features might be an appropriate compromise to increase classification performance and ensure best possible clinical interpretability. In settings where only classification performance without the need for biomechanical interpretability is important, automated feature extraction approaches can be suggested.

Simple metrics are often used for describing pathologic movements (e.g., symmetry index [48]) in clinical contexts because the full consideration of all movements is too complex for human beings. Further, many clinical decisions are influenced by expert knowledge and the experiences of the physicians and therapists. As the current study demonstrates, the benefits of data-driven approaches

based on machine learning and XAI are their ability to take into account the full complexity of the motion data, and they therefore may provide objective orientations and assistance for physicians in their decisions. Further, through focusing on local model interpretability, they are able to take into account the individual gait characteristics that were important for classification and class discrimination and, therefore, play an important role in the context of personalized medicine.

Local interpretability additionally helps one to understand why a model wrongly classified single instances. As previously presented as an example, a GC of a patient was wrongly classified as a GC of a healthy person. A possible explanation is that the subject showed fewer pathologic gait patterns compared with the remaining patients and was therefore classified as a healthy person. In this regard, the algorithm could provide information about the rehabilitation status of a patient and an objective orientation for physicians and therapists. However, it cannot be excluded that the model was not able to handle and correctly classify the instance, because the regarded instance was an outlier. In this regard, the usage of automatic systems is dependent on experts, because in such cases, expert knowledge and experience is crucial for the right decision-making.

The current work emphasizes the importance of data preprocessing for a model's accuracy and interpretability and shows that it is not possible to give general recommendations. In the present case, the best results were obtained without scaling for the use of gait waveform data. This might also be promising for interpretation because features can be directly interpreted without the need of scaling the data back to its original representation.

Another important aspect is data aggregation. In the present case, in line with various other studies [50,51], single gait cycles were used for input feature calculation. However, in the literature, averaged waveforms were also reported for gait classification [20,52], which might be useful for elimination of the intraindividual variance between different gait cycles. Depending on the requirements in possible fields of application, this might be an alternative approach. Various other methods should not go unnoticed for dealing with imbalanced data (see, e.g., [53]) and should be further considered and compared in future works.

Previous research showed that IMU-based systems show higher errors for movements in the transversal plane [54]. Therefore, it is uncertain if the findings for transversal movements are meaningful and interpretable. The exclusion of transverse movements due to great inaccuracies when using IMU-based systems might, therefore, be considered in future works.

As noted previously, there are limitations regarding the sample of subjects. In particular, the large age and weight difference between the groups of healthy subjects and patients after THA is striking. At this point, it cannot be excluded that the corresponding effects influenced the classification. In consecutive studies, the analysis should be repeated with a matched group of healthy subjects. Further, it worth mentioning the possible strong correlations between the features which are a priori present and lead to redundancies in the different input vectors. To reduce the impact of redundant features, application of the minimum redundancy maximum relevance (mRMR) filter [15,55] might be a promising approach.

The present work used the LIME algorithm and checked if its application would lead to coherent results from different input representations. The utility and generalizability of the mentioned methodology should be evaluated through application on data of different subjects and groups of patients. Moreover, future works should compare different explanation methods (e.g., LIME, SHAP, DeepLIFT) and check whether results are consistent.

5. Conclusions

XAI is promising for making the decisions of machine learning models more transparent, interpretable, and, therefore, more trustworthy. It is, therefore, a promising step towards the practical applicability of machine learning in clinical contexts. However, the research is still very young in this domain. Before practical clinical application, further research is necessary. The current study shows that the type of input representation crucially determines interpretability as well as clinical

relevancy. A combined approach using different forms of representations seems advantageous, because it can provide a better understanding of the underlying group differences and discriminative effects that were important for classification. Based on the current findings, waveform data as well as features based on simple descriptive statistics can be suggested. In the context of personalized medicine, XAI approaches focusing on local interpretability enable the identification of individual gait characteristics, which are important for classification and class discrimination. Thus, the results might assist physicians and therapists in finding and addressing individual pathologic gait patterns by offering an objective orientation.

Author Contributions: Conceptualization, C.D., W.T., B.T., G.B., and M.F.; methodology, C.D. and B.T.; software, C.D. and W.T.; validation, C.D.; formal analysis, C.D.; investigation, C.D. and W.T.; resources, M.F.; data curation, W.T.; writing—original draft preparation, C.D. and W.T.; writing—review and editing, C.D., W.T., B.T., G.B., and M.F.; visualization, C.D.; supervision, M.F.; project administration, M.F.; funding acquisition, M.F. and G.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Offene Digitalisierungsallianz Pfalz, BMBF, grant number 03IHS075B.

Acknowledgments: We would like to thank the staff of the Klinik Lindenplatz, Bad Sassendorf, Germany for their support concerning the subject acquisition and data acquisition.

Conflicts of Interest: The authors declare no conflict of interest. The authors alone are responsible for the content and writing of this paper. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

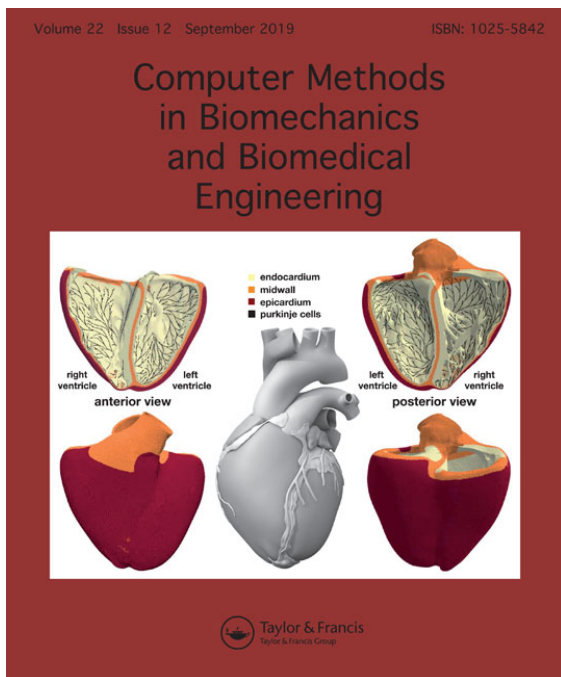
1. Eskofier, B.M.; Kraus, M.; Worobets, J.T.; Stefanyshyn, D.J.; Nigg, B.M. Pattern classification of kinematic and kinetic running data to distinguish gender, shod/barefoot and injury groups with feature ranking. *Comput. Methods Biomech. Biomed. Eng.* **2012**, *15*, 467–474. [[CrossRef](#)] [[PubMed](#)]
2. Ferber, R.; McClay Davis, I.; Williams, D.S., III. Gender differences in lower extremity mechanics during running. *Clin. Biomech.* **2003**, *18*, 350–357. [[CrossRef](#)]
3. Phinyomark, A.; Petri, G.; Ibáñez-Marcelo, E.; Osis, S.T.; Ferber, R. Analysis of Big Data in Gait Biomechanics: Current Trends and Future Directions. *J. Med. Biol. Eng.* **2018**, *38*, 244–260. [[CrossRef](#)]
4. Halilaj, E.; Rajagopal, A.; Fiterau, M.; Hicks, J.L.; Hastie, T.J.; Delp, S.L. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. *J. Biomech.* **2018**, *81*, 1–11. [[CrossRef](#)] [[PubMed](#)]
5. Kobsar, D.; Charlton, J.M.; Tse, C.T.F.; Esculier, J.-F.; Graffos, A.; Krowchuk, N.M.; Thatcher, D.; Hunt, M.A. Validity and reliability of wearable inertial sensors in healthy adult walking: A systematic review and meta-analysis. *J. Neuroeng. Rehabil.* **2020**, *17*, 62. [[CrossRef](#)]
6. Bzdok, D.; Altman, N.; Krzywinski, M. Statistics versus machine learning. *Nat. Methods* **2018**, *15*, 233–234. [[CrossRef](#)]
7. Laroche, D.; Tolambiya, A.; Morisset, C.; Mailliefert, J.F.; French, R.M.; Ornetti, P.; Thomas, E. A classification study of kinematic gait trajectories in hip osteoarthritis. *Comput. Biol. Med.* **2014**, *55*, 42–48. [[CrossRef](#)]
8. Lau, H.-Y.; Tong, K.-Y.; Zhu, H. Support vector machine for classification of walking conditions of persons after stroke with dropped foot. *Hum. Mov. Sci.* **2009**, *28*, 504–514. [[CrossRef](#)]
9. Wahid, F.; Begg, R.K.; Hass, C.J.; Halgamuge, S.; Ackland, D.C. Classification of Parkinson's Disease Gait Using Spatial-Temporal Gait Features. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 1794–1802. [[CrossRef](#)]
10. Adadi, A.; Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160. [[CrossRef](#)]
11. European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Off. J. Eur. Union* **2016**, *L119*, 1–88.
12. Slijepcevic, D.; Zeppelzauer, M.; Schwab, C.; Raberger, A.-M.; Breiteneder, C.; Horsak, B. Input representations and classification strategies for automated human gait analysis. *Gait Posture* **2020**, *76*, 198–203. [[CrossRef](#)] [[PubMed](#)]

13. Teufl, W.; Taetz, B.; Miezal, M.; Lorenz, M.; Pietschmann, J.; Jöllenbeck, T.; Fröhlich, M.; Bleser, G. Towards an Inertial Sensor-Based Wearable Feedback System for Patients after Total Hip Arthroplasty: Validity and Applicability for Gait Classification with Gait Kinematics-Based Features. *Sensors* **2019**, *19*, 5006. [CrossRef]
14. Begg, R.; Kamruzzaman, J. A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *J. Biomech.* **2005**, *38*, 401–408. [CrossRef] [PubMed]
15. Dindorf, C.; Teufl, W.; Taetz, B.; Becker, S.; Bleser, G.; Fröhlich, M. Feature extraction and gait classification in hip replacement patients on basis of kinematic waveform data. (under review).
16. Horst, F.; Slijepcevic, D.; Lapuschkin, S.; Raberger, A.-M.; Zeppelzauer, M.; Samek, W.; Breiteneder, C.; Schöllhorn, W.I.; Horsak, B. On the Understanding and Interpretation of Machine Learning Predictions in Clinical Gait Analysis Using Explainable Artificial Intelligence. Available online: <http://arxiv.org/pdf/1912a.07737v1> (accessed on 10 March 2020).
17. Dindorf, C.; Konradi, J.; Wolf, C.; Taetz, B.; Bleser, G.; Huthwelker, J.; Drees, P.; Fröhlich, M.; Betz, U. General method for automated feature extraction and selection and its application for gender classification and biomechanical knowledge discovery of sex differences in spinal posture during stance and gait. (under review).
18. Christ, M.; Kempa-Liehr, A.W.; Feindt, M. Distributed and Parallel Time Series Feature Extraction for Industrial Big Data Applications. Available online: <http://arxiv.org/pdf/1610.07717v3> (accessed on 10 January 2020).
19. Feature Labs, I. Featuretools: Automated Feature Engineering. Available online: <https://www.featuretools.com/> (accessed on 30 May 2020).
20. Eskofier, B.M.; Federolf, P.; Kugler, P.F.; Nigg, B.M. Marker-based classification of young-elderly gait pattern differences via direct PCA feature extraction and SVMs. *Comput. Methods Biomech. Biomed. Eng.* **2013**, *16*, 435–442. [CrossRef]
21. Liu, H.; Yu, L. Toward Integrating Feature Selection Algorithms for Classification and Clustering. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 491–502. [CrossRef]
22. Gunning, D.; Stefik, M.; Choi, J.; Miller, T.; Stumpf, S.; Yang, G.-Z. XAI—Explainable artificial intelligence. *Sci. Robot.* **2019**, *4*, eaay7120. [CrossRef]
23. Bologna, G.; Hayashi, Y. Characterization of Symbolic Rules Embedded in Deep DIMLP Networks: A Challenge to Transparency of Deep Learning. *J. Artif. Intell. Soft Comput. Res.* **2017**, *7*, 265–286. [CrossRef]
24. Samek, W.; Müller, K.-R. Towards explainable artificial intelligence. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, 1st ed.; Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.-R., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 5–22.
25. Holzinger, A.; Biemann, C.; Pattichis, C.S.; Kell, D.B. What do We Need to Build Explainable AI Systems for the Medical Domain? Available online: <http://arxiv.org/pdf/1712.09923v1> (accessed on 20 February 2020).
26. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.
27. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
28. Shrikumar, A.; Greenside, P.; Kundaje, A. Learning important features through propagating activation differences. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 3145–3153.
29. OECD. *Health at a Glance: Europe 2016. State of Health in the EU Cycle*; OECD Publishing: Paris, France, 2016; ISBN 978-92-64-26559-2.
30. Teufl, W.; Miezal, M.; Taetz, B.; Fröhlich, M.; Bleser, G. Validity, Test-Retest Reliability and Long-Term Stability of Magnetometer Free Inertial Sensor Based 3D Joint Kinematics. *Sensors* **2018**, *18*, 1980. [CrossRef]
31. Miezal, M.; Taetz, B.; Bleser, G. On Inertial Body Tracking in the Presence of Model Calibration Errors. *Sensors* **2016**, *16*, 1132. [CrossRef]
32. Miezal, M.; Taetz, B.; Bleser, G. Real-time inertial lower body kinematics and ground contact estimation at anatomical foot points for agile human locomotion. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 3256–3263.
33. Teufl, W.; Miezal, M.; Taetz, B.; Fröhlich, M.; Bleser, G. Validity of inertial sensor based 3D joint kinematics of static and dynamic sport and physiotherapy specific movements. *PLoS ONE* **2019**, *14*, e0213064. [CrossRef] [PubMed]

34. Teufl, W.; Lorenz, M.; Miezal, M.; Taetz, B.; Fröhlich, M.; Bleser, G. Towards Inertial Sensor Based Mobile Gait Analysis: Event-Detection and Spatio-Temporal Parameters. *Sensors* **2018**, *19*, 38. [[CrossRef](#)] [[PubMed](#)]
35. Ewen, A.M.; Stewart, S.; St Clair Gibson, A.; Kashyap, S.N.; Caplan, N. Post-operative gait analysis in total hip replacement patients—a review of current literature and meta-analysis. *Gait Posture* **2012**, *36*, 1–6. [[CrossRef](#)] [[PubMed](#)]
36. Beaulieu, M.L.; Lamontagne, M.; Beulé, P.E. Lower limb biomechanics during gait do not return to normal following total hip arthroplasty. *Gait Posture* **2010**, *32*, 269–273. [[CrossRef](#)]
37. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *JAIR* **2002**, *16*, 321–357. [[CrossRef](#)]
38. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
39. Schöllhorn, W.I.; Nigg, B.M.; Stefanyshyn, D.J.; Liu, W. Identification of individual walking patterns using time discrete and time continuous data sets. *Gait Posture* **2002**, *15*, 180–186. [[CrossRef](#)]
40. Horst, F.; Kramer, F.; Schäfer, B.; Eekhoff, A.; Hegen, P.; Nigg, B.M.; Schöllhorn, W.I. Daily changes of individual gait patterns identified by means of support vector machines. *Gait Posture* **2016**, *49*, 309–314. [[CrossRef](#)]
41. Hering, J.; Metzenthin, E.; Zenner, A. LIME For Time. Available online: <https://github.com/emanuel-metzenthin/Lime-For-Time> (accessed on 20 February 2020).
42. Molnar, C. Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. 2018. Available online: <https://christophm.github.io/interpretable-ml-book/index.html> (accessed on 20 February 2020).
43. Ataky, T.C. One-dimensional statistical parametric mapping in Python. *Comput. Methods Biomech. Biomed. Eng.* **2012**, *15*, 295–301. [[CrossRef](#)]
44. Horstmann, T.; Listringhaus, R.; Haase, G.-B.; Grau, S.; Mündermann, A. Changes in gait patterns and muscle activity following total hip arthroplasty: A six-month follow-up. *Clin. Biomech.* **2013**, *28*, 762–769. [[CrossRef](#)]
45. Begg, R.K.; Palaniswami, M.; Owen, B. Support vector machines for automated gait classification. *IEEE Trans. Biomed. Eng.* **2005**, *52*, 828–838. [[CrossRef](#)] [[PubMed](#)]
46. Chopra, S.; Kaufman, K.R. Effects of total hip arthroplasty on gait. In *Handbook of Human Motion*; Müller, B., Wolf, S., Eds.; Springer: Cham, Germany, 2018; pp. 1–15, ISBN 978-3-319-30808-1.
47. Perron, M.; Malouin, F.; Moffet, H.; McFadyen, B.J. Three-dimensional gait analysis in women with a total hip arthroplasty. *Clin. Biomech.* **2000**, *15*, 504–515. [[CrossRef](#)]
48. Madsen, M.S.; Ritter, M.A.; Morris, H.H.; Meding, J.B.; Berend, M.E.; Faris, P.M.; Vardaxis, V.G. The effect of total hip arthroplasty surgical approach on gait. *J. Orthop. Res.* **2004**, *22*, 44–50. [[CrossRef](#)]
49. Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.* **2019**, *267*, 1–38. [[CrossRef](#)]
50. Levinger, P.; Lai, D.T.H.; Begg, R.K.; Webster, K.E.; Feller, J.A. The application of support vector machines for detecting recovery from knee replacement surgery using spatio-temporal gait parameters. *Gait Posture* **2009**, *29*, 91–96. [[CrossRef](#)]
51. Nüesch, C.; Valderrabano, V.; Huber, C.; von Tscharnar, V.; Pagenstert, G. Gait patterns of asymmetric ankle osteoarthritis patients. *Clin. Biomech.* **2012**, *27*, 613–618. [[CrossRef](#)] [[PubMed](#)]
52. Soares, D.P.; de Castro, M.P.; Mendes, E.A.; Machado, L. Principal component analysis in ground reaction forces and center of pressure gait waveforms of people with transfemoral amputation. *Prosthet. Orthot. Int.* **2016**, *40*, 729–738. [[CrossRef](#)]
53. Barlow, H.; Mao, S.; Khushi, M. Predicting High-Risk Prostate Cancer Using Machine Learning Methods. *Data* **2019**, *4*, 129. [[CrossRef](#)]
54. Poitras, I.; Dupuis, F.; Biellmann, M.; Campeau-Lecours, A.; Mercier, C.; Bouyer, L.J.; Roy, J.-S. Validity and Reliability of Wearable Sensors for Joint Angle Estimation: A Systematic Review. *Sensors* **2019**, *19*, 1555. [[CrossRef](#)]
55. Peng, H.; Long, F.; Ding, C. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [[CrossRef](#)]



3.4 4. Publikation



Machine learning techniques demonstrate individual movement patterns of the vertebral column: The fingerprint of spinal motion

C. Dindorf, J. Konradi, C. Wolf u. a., "Machine learning techniques demonstrating individual movement patterns of the vertebral column: The fingerprint of spinal motion," *Computer Methods in Biomechanics and Biomedical Engineering*, S. 1–11, 2021 (published online). DOI: 10.1080/10255842.2021.1981884

Zusammenfassung

Hintergrund: Oberflächentopographiesysteme ermöglichen es, dynamische Bewegungen der Wirbelsäule zu erfassen. Es ist unklar, ob die Wirbeldynamik einzigartig genug ist, um Individuen zu erkennen. Das Vorliegen limitierter Datenbestände stellt zudem ein Hindernis dar. Daher analysiert diese Studie auf der Grundlage dynamischer Wirbelsäulendaten, ob eine Identifizierung von Personen möglich ist.

Methodische Forschungsfragen: Eignen sich SNN zum Quer- und Längsschnittvergleich sowie zur Identifikation von Subjekten bei limitierter Ausgangsdatenlage und Vorhandensein von intra-individuellen Varianzen?

Fachspezifische Forschungsfragen: Sind dynamische Wirbelsäulemuster einzigartig genug um Subjekte wiederzuerkennen? Für welche Parameter bzw. Parameterinteraktionen lassen sich besonders individuelle Charakteristika feststellen?

Methodik: Dynamische Wirbelsäuledaten von 226 Gesunden bilden die Ausgangsdatenbasis. Es werden drei verschiedene Datendarstellungen zur Subjektidentifikation verglichen (automatisierte extrahierte Feature via SNN und Contrastive Loss und Triplet Loss Funktionen sowie extrahierte Feature auf Basis einfacher deskriptiver Statistiken).

Ergebnisse: Hohe Identifikationsgenauigkeit (96,00 - 100,00 %) liegt vor. Der Einsatz der extrahierten Feature via Contrastive Loss ist überlegen gegenüber den zwei anderen Arten von extrahierten Features.

Fazit: Insgesamt erscheint der vorgestellte Ansatz vielversprechend bei Vorliegen einer intra-individuellen Variabilität zwischen bei unterschiedlichen Geschwindigkeiten aufgezeichneten Gangzyklen eines Tages sowie der Tag zu Tag Variabilität um die intra-individuellen Varianzen zu minimieren und Personenvergleiche und -identifikationen zu verbessern. Hohe Genauigkeiten weisen auf das Vorhandensein eines individuellen „Fingerabdrucks“ der Wirbelsäule hin. Die vorliegende Arbeit bildet die Grundlage für einen objektiven Personenvergleich und die Übertragung der Methode auf klinische Anwendungsfälle.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

SCHOLARONE™
Manuscripts

This is an Accepted Manuscript of an article published by Taylor & Francis in
Computer Methods in Biomechanics and Biomedical Engineering on 30 Sep 2021,
available at: <https://doi.org/10.1080/10255842.2021.1981884>.

URL: <http://mc.manuscriptcentral.com/gcmb>

1
2
3
4
5 **Machine learning techniques demonstrating individual movement**
6 **patterns of the vertebral column: The fingerprint of spinal motion**
7
8

9
10 Carlo Dindorf^{ao}, Jürgen Konradi^{bo}, Claudia Wolf^b, Bertram Taetz^c, Gabriele
11 Bleser^d, Janine Huthwelker^b, Friederike Werthmann^e, Philipp Drees^e,
12 Michael Fröhlich^{a*} and Ulrich Betz^{b*}
13
14

15
16 *^aDepartment of Sports Science, Technische Universität Kaiserslautern, Kaiserslautern,*
17 *Germany; ^bInstitute of Physical Therapy, Prevention and Rehabilitation, University*
18 *Medical Centre of the Johannes Gutenberg University Mainz, Mainz, Germany;*
19 *^cDepartment Augmented Vision, German Research Center for Artificial Intelligence,*
20 *Kaiserslautern, Germany; ^dJunior Research Group wear HEALTH, Technische*
21 *Universität Kaiserslautern, Kaiserslautern, Germany; ^eDepartment of Orthopedics and*
22 *Trauma Surgery, University Medical Centre of the Johannes Gutenberg University*
23 *Mainz, Mainz, Germany*
24
25
26
27
28
29

30
31 *^omutual first authors / authors contributed equally.*
32

33
34 **mutual senior authors / authors contributed equally.*
35

36
37 CONTACT Carlo Dindorf carlo.dindorf@sowi.uni-kl.de
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Machine learning techniques demonstrating individual movement patterns of the vertebral column: The fingerprint of spinal motion

Surface topography systems enable the capture of spinal dynamic movement; however, it is unclear whether vertebral dynamics are unique enough to identify individuals. Therefore, in this study, we investigated whether the identification of individuals is possible based on dynamic spinal data. Three different data representations were compared (automated extracted features using contrastive loss and triplet loss functions, as well as simple descriptive statistics). High accuracies indicated the possible existence of a personal spinal ‘fingerprint’, therefore enabling subject recognition. The present work forms the basis for an objective comparison of subjects and the transfer of the method to clinical use cases.

Keywords: Siamese neural networks; triplet loss, contrastive loss; surface topography; subject identification

1. Introduction

Biometric person identification is an important research field that has many practical applications, with the general aim of identifying persons based on unique biological characteristics. Several works have demonstrated the utility of human gait for subject identification (Zhang et al. 2016; Liu et al. 2018; Kececi et al. 2020). Regarding health-related sectors, subject identification and the determination of similarities between subjects are of high research interest, as are longitudinal comparisons of subjects over time (Zhao et al. 2018; Li et al. 2020). However, large intra-class and small inter-class variations of the human gait pose significant challenges (Tong et al. 2017; Liu et al. 2018).

In previous studies, researchers have been able to identify subjects based, for example, on silhouette shapes (Ekinici 2006) or three-dimensional motion capture data (Nguyen et al. 2018). Recently, systems based on surface topography (ST) have enabled the capture of dynamic spinal movement. The human skeleton appears to be unique

1
2
3 enough to allow for specific attributions of an individual (Valenzuela 1997; Mundorff et
4 al. 2006; Ognard et al. 2020). A prior study demonstrated that there are sex-specific
5 differences in the dynamic movement of the spine, which are suitable for classification
6 (Dindorf and Konradi et al. 2020a). However, it is unclear whether the vertebral
7 dynamics are unique enough, as fingerprints are, to recognize individuals.
8
9

10
11
12
13
14
15 As a methodological approach for the identification and comparison of data
16 samples, deep metric learning via a Siamese neural network (SNN), a special type of
17 neural network, has gained increasing interest. The general aim of this network is to put
18 similar samples nearer to each other and different samples further apart by learning
19 discriminative embeddings using neural networks (Hadsell et al. 2006). Two commonly
20 used loss functions for optimization are contrastive loss (Hadsell et al. 2006) and triplet
21 loss (Weinberger and Saul 2009). An SNN showed promising results in that it was able
22 to learn from only a few or even only one training sample (one-shot learning) of each
23 class (Chicco 2021). This is an important aspect, as the amount of available clinical data
24 is often limited. Furthermore, previous studies have demonstrated that SNNs were able
25 to extract meaningful, robust, and discriminative gait features (Zhang et al. 2016; Wu et
26 al. 2017; Liu et al. 2018), which forms the basis for both identification as well as the
27 determination of similarities.
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

44 SNNs have mainly been applied in the field of computer vision (e.g., Figueroa-
45 Mata and Mata-Montero 2020). Several clinical- and health-related studies can be found
46 in the literature (Zhao et al. 2018; Li et al. 2020; Bhagwat et al. 2018; Liu et al. 2019).
47 To the best of our knowledge, the utility of SNNs in the context of biomechanical
48 waveform data has not been analyzed so far.
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Therefore, using the ST data of healthy subjects¹, our primary goal was to
4 investigate whether individual dynamic movement patterns of the spine are unique
5
6 enough that the identification of individuals is possible. Second, we wanted to compare
7
8 automated extracted features using an SNN with contrastive loss and triplet loss
9
10 functions with simple descriptive statistics and determine which data representation is
11
12 better suited for subject identification.
13
14
15

16
17 Predictive performance and interpretability are important aspects of practical
18
19 clinical applications. There is a need to reveal what an individual has in common with
20
21 others as well as what distinguishes them. Model interpretability and the interpretability
22
23 of the features themselves become important aspects in clinical contexts. In this way,
24
25 simple descriptive statistics are of high relevancy for describing biomechanical
26
27 characteristics (Dindorf and Teufl et al. 2020b). In this regard, our third and final aim
28
29 was to determine the features that best map similarities within and differences between
30
31 the subjects overall, with a focus on simple descriptive statistics.
32
33
34
35
36

37 2. Method

38
39 An overview of the research methodology is depicted in Figure 1a. A detailed
40
41 description of the displayed steps is provided in Sections 2.1–2.3.
42
43
44

45 2.1 Measuring method, subjects, and data

46
47 The DIERS formetric III 4D™, DICAM v3.7.1.7 was used as a means to collect ST
48
49 data. Light projected onto a textile-free back of an individual person was recorded by a
50
51 camera (60°Hz), resulting in a software-based 3D image of the back's surface, which
52
53 was used to calculate corresponding 3D movements from the vertebra prominens (C7)
54
55

56
57 ¹ The present evaluation of the general methodology based on the dynamic spinal data
58
59 of healthy subjects forms the first step towards the authors' greater aim—the
60
development of a practical application for the use in a clinical context with patient data.

1
2
3 down to the pelvis. An integrated Zebris™ foot pressure measuring plate (5376 sensors,
4
5 100 Hz, sensitivity 1 N/cm², accuracy 5% FS) was used to measure gait cycles, as
6
7 previously described by Betz et al. (2017) and Dindorf and Konradi et al. (2020a).
8

9
10 Two reference data pools were used as datasets (see Table 1²); they were both
11
12 approved by the responsible ethics committee of the Medical Chamber of Rhineland-
13
14 Palatinate (837.194.16, 2018-13607-Klinische Forschung) and registered with WHO
15
16 (INT: DRKS00010834, DRKS00014325). All participants provided written informed
17
18 consent. The data pools contained measurements from 201/25 healthy subjects (132/13
19
20 female, 69/12 male) while standing and walking (Data A: 2, 3, 4, and 5 km/h; Data B: 2
21
22 and 4 km/h). The Data B group was measured on three different days (Day 1, Day 2,
23
24 and 30 ± 7 days later). A gait measurement, which lasts for three gait cycles, consisted
25
26 of 300–500 single images. More than 100 spine parameters were recorded. For the
27
28 current study, 55 parameters were used (pelvic obliquity (°), pelvic torsion (dimples)
29
30 (°), pelvic inclination (dimples) (°), pelvic rotation (°), and movement of the VP, T1–
31
32 12, and L1–L4 in all planes of motion (°); please refer to the online Supplementary
33
34 Materials and to Dindorf and Konradi et al. (2020a)). Missing data points were
35
36 interpolated using spline interpolation (maximum gap = 5 frames). Cycles with
37
38 remaining gaps were dropped. Afterwards, the gait cycles were individually time-
39
40 normalized to 101 time steps (from 0% to 100%) using cubic spline interpolation (time-
41
42 normalized sequences are shown in Figure 1 (a); exemplary visualizations are provided
43
44 in the online Supplementary Materials).
45
46
47
48
49
50
51
52
53
54
55
56
57

58 ² Dataset B is part of the dissertation project of Friederike Werthmann.
59
60

2.2 Training, validation, and testing procedure

For both Data A and Data B, the data of two different speeds were used (2 and 4 km/h). Feature extraction using the training and validation data was performed using Data A. The training data for the contrastive loss function was generated by random sampling comprising 50% of the data samples that mapped the same persons and 50% of the data samples that mapped different persons. For the triplet loss, random sampling was performed in triplets of anchor, positive, and negative samples. To check the generalizability of the extracted features, test samples from classes (subjects) that were not observed during the training were classified (zero-shot learning). Therefore, only Data B were used for testing (see Figure 1). To make the task for testing more difficult, Data A were combined with the reference data of Data B, resulting in a reference data pool of 226 subjects (25 target subjects) with 1496 gait cycles (150 target gait cycles).

2.3 Feature extraction and subject identification

To check, if the identification of individuals is possible based on dynamic movement patterns of the spine (i), as well as for comparing different data representations for subject identification (ii), three feature extraction methods were applied based on the waveform data, as follows:

- (1) Automatically extracted features using the contrastive loss function;
- (2) Automatically extracted features using the triplet loss function;
- (3) Features extracted using simple descriptive statistics.

Similar to Horst et al. (2019) and Dindorf and Teufl et al. (2020b), concatenated waveform data was used as input features for the automatic feature extraction with the SNN. Thus, the time-normalized sequences for each gait cycle were concatenated into a vector of 5555 dimensions. The raw data of two exemplary subjects are presented in

1
2
3 **Figure 1b.** The initial input features for the use of simple descriptive statistics **were**
4 **based on operations** commonly applied in **the** literature (e.g., Dindorf and Teufl et al.,
5
6 2020b) and calculated using the range of motion (ROM), minima, and maxima. Both
7
8 input feature sets were separately scaled to a range between zero and one based on the
9
10 training data.
11
12

13
14
15 Figure 2 shows the general architecture for the use of **the** contrastive loss and
16 triplet loss functions. In a first step, the features **were** automatically extracted **by** passing
17 each input through one of the identical weight-sharing subnetworks (two for contrastive
18 loss, three for triplet loss). Then, during the optimization process, the error of the model
19
20 **was** calculated based on the resulting feature vectors and using the respective loss
21 function. For identification, **the** SNN performed **a** classification of the inputs based on
22 their similarity instead of directly classifying the inputs. Based on the output feature
23 vectors, a similarity score **was** calculated between the subject to be identified and the
24 data sample in the reference data using, for example, the Euclidian distance.
25
26

27
28
29 The SNN with contrastive loss was chosen as **the** starting point for the
30 identification task. A multilayer perceptron (MLP) feedforward neural network was
31 selected with two hidden layers. The network architecture and hyperparameters were
32 determined manually. The output layer size was gradually increased and the best
33 performing model (performance on validation set) with the lowest complexity was
34 selected. The following architecture was chosen:
35

- 36 ● 5555-neuron input layer;
 - 37 ● 1000-neuron hidden layer 1;
 - 38 ● 100-neuron hidden layer 2;
 - 39 ● 5-neuron output layer.
- 40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 The rectified linear activation function (ReLU) was selected. Using contrastive
4 loss, the learning rate was set to 0.0005, with 315 epochs and a minibatch size of 32. To
5 minimize distortion, the same network architecture was used for comparing the use of
6 the contrastive loss to the triplet loss function. After manually tuning, the learning rate
7 was set to 0.001, with 500 epochs and a minibatch size of 100 for the triplet loss
8 function. For both approaches, the Adam optimizer was used.

9
10 For comparing feature subsets of the same size, sequential forward selection
11 using Euclidean distance (see below) for subject identification was applied for ranking
12 and selecting a feature subset of the features based on simple descriptive statistics. The
13 ranking was also used as the global feature importance. The number of selected features
14 was set as equal to the number of automatically extracted features ($n = 5$).

15
16 Based on the extracted features, subject identification was performed using
17 Euclidean distance. The Euclidean distance was calculated between the subject to be
18 identified and each data sample in the reference data. The data sample in the reference
19 pool with the lowest distance was predicted as the person to be identified. This matches
20 with the classification using the k-nearest neighbor classifier with the number of
21 neighbors equal to one.

22
23 The test data was held out for the entire training and search process.
24 Calculations were performed in Python (Python Software Foundation, Wilmington, DE,
25 USA), and PyTorch (Paszke et al. 2017) was used for building the neural networks. For
26 visualization, the test subjects were plotted in a 2D space after doing dimensionality
27 reduction on the extracted features with t-SNE (van der Maaten and Hinton 2008).

28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

3. Results

For validation (True: 402, False: 0) and test data (True: 150, False: 0), 100%
accuracy was obtained using the concatenated waveform data without performing

1
2
3 feature extraction. Figure 3 shows the corresponding visualization of the test subjects in
4
5 a 2D space. A separation between the subjects is **observable. The** calculated features
6
7 based on ROM, Min, and Max ($n = 165$) **resulted** in a validation set accuracy of 99.25%
8
9 (True: 399, False: 3) and a test set accuracy of 99.33% (True: 149, False: 1).
10
11

12 The results for the extracted features are presented in Table 2. The highest
13
14 accuracy **was** present with the SNN using the contrastive loss function, followed by the
15
16 SNN with triplet loss and the features based on simple descriptive statistics. The test set
17
18 accuracy **was** (slightly) reduced after **the** feature extraction.
19
20

21 Figure 5 **shows the comparison between** the 2D representations of the extracted
22
23 features. The best separation of the different subjects is visible for **the automatically**
24
25 extracted features using contrastive loss. For example, for the subjects C, G, and S, the
26
27 respective gait cycles **were** closer together compared to the t-SNE visualization of the
28
29 concatenated waveform data before **the** feature extraction.
30
31

32 Figure 4a shows the validation set accuracy during the sequential feature
33
34 selection of the features calculated using ROM, Min, and Max. It is noticeable that an
35
36 increase in the subset size led to higher accuracy. With nine features, 100% accuracy for
37
38 the validation set (True: 402, False: 0) and 95.33% accuracy for the test set (True: 143,
39
40 False: 7) **were** achieved. According to **when they were** included during **the** sequential
41
42 feature selection, the features based on T3 followed by L4 and T8 movement in **the**
43
44 sagittal plane show the highest feature importance (**visualizations are additionally**
45
46 **presented in the online Supplementary Materials**). Most of the features with the highest
47
48 importance relate to movement **in the** sagittal plane (except for L3 movement in **the**
49
50 frontal plane and T5 movement **in the** transversal plane). Figure 4b uses the top two
51
52 features for plotting the test subjects in a 2D space. Using these features, **the accuracy of**
53
54 the validation set is 33.08% (True: 133, False: 269) and on the test set 25.33% (True:
55
56
57
58
59
60

1
2
3 38, False: 112). Overlapping subjects are present in the respective 2D space. However,
4
5 some subjects are well separated from the others (e.g., B, H, J, P, and Q), although only
6
7 two features were used.
8
9

10 11 **4. Discussion**

12
13 The current study shows that the identification of individuals based on dynamic
14
15 movement patterns of the spine is possible. The overall best results were obtained
16
17 without feature extraction and the use of concatenated waveform data as input features
18
19 (100% accuracy). The results of the comparisons between the different feature
20
21 extraction approaches demonstrate that the SNN with the contrastive loss function led to
22
23 the best results when extracting a subset of five features (96.00%). Overall, very good
24
25 generalizability of the extracted features was present, as the test samples of the subjects
26
27 were not observed during the training phase. Contrastive loss was used as a starting
28
29 point for determining the subset size. However, different subset sizes of the extracted
30
31 features may lead to better results with the use of triplet loss and features based on
32
33 simple descriptive statistics. In this way, an increase in accuracy was present when
34
35 using nine instead of five features for identification, with the features based on ROM,
36
37 Min, and Max (a 95.33% accuracy compared to 82.00% using five features). Similar to
38
39 Horst et al. (2019) and Dindorf and Teufl et al. (2020b), we used concatenated
40
41 waveform data as the input features for an MLP neural network. The use of different
42
43 network architectures might have influenced the results. In the context of waveform
44
45 data, 1D convolutional neural network models (Kiranyaz et al. 2021) and recurrent
46
47 neural networks (Yu et al. 2019) may also be promising network architectures.
48
49 Therefore, future research should also consider and compare these network
50
51 architectures.
52
53
54
55
56
57
58
59
60

1
2
3 Previous research showed that an SNN with triplet loss could learn better
4 representations compared to an SNN with contrastive loss (Hoffer and Ailon 2018).
5
6 However, in the present case, contrastive loss showed better results. Possible reasons for
7
8 this might be that the network architecture was predetermined using contrastive loss
9 and, therefore, was not optimal for use with triplet loss. Furthermore, through random
10 sampling of the training data in triplets of anchor, positive, and negative samples, the
11 training data might have been too easy to learn. Choosing a difficult triplet could be an
12 essential step for further increasing the model's performance (Gordo et al. 2016) and
13 should be considered in future works.
14
15
16
17
18
19
20
21
22

23
24 The results highlight the uniqueness of individual dynamic movement patterns
25 of the spine. The visualizations of the subjects in a 2D space using t-SNE showed that,
26 depending on the feature extraction approach, a clear separation of the subjects was
27 visible. Furthermore, different gait cycles of the subjects were plotted closely together
28 and, therefore, seemed relatively similar. This is surprising, as the GCs of each subject
29 were recorded at two different speeds. Overall, the intra-individual variability between
30 the gait cycles of one day and the day-to-day variability of the present subjects might be
31 low overall and did not appear to significantly affect the task. The extracted features,
32 especially for use with an SNN with contrastive loss, seemed to further reduce intra-
33 individual variability. For example, the GCs of subjects C, G, and S were closer
34 together compared to the t-SNE visualization of the concatenated waveform data before
35 the feature extraction. In this way, the extracted features should map movement patterns
36 of the spine that show little variance or are invariant over different gait cycles, speeds,
37 and measuring days. This results in a higher similarity between the gait cycles per
38 subject as well as better separation between the subjects.
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

However, looking at the extracted feature using contrastive loss, few subjects showed higher variability (e.g., subject K), because the gait cycles were plotted further apart compared to other subjects (e.g., subject Q), which was also shown by Horst et al. (2016). One possible reason for this may be that, in general, these subjects showed higher variability in the movement patterns of the spine. Other possible factors could be the influence of an unfamiliar (slow) walking speed (England and Granata 2007; Kang and Dingwell 2008), age (Kang and Dingwell 2008), or psycho-emotional factors (Randhavane et al. 2019). Thus, it became hard for the algorithm to best map invariant features.

Using extracted features, the accuracy was slightly reduced compared to the use of the entire concatenated waveforms. However, feature extraction is an important step toward improving interpretability and the model's accuracy, and for preventing overfitting and reducing the computing power (Liu and Yu 2005). Based on the huge amount of data, direct interpretation of the dynamic spinal data is hardly possible for humans. In line with a previous study (Dindorf and Konradi et al. 2020a), a clinically relevant interpretation is only adequately possible through the extraction of meaningful features.

Looking at the highest-ranked features based on simple descriptive statistics is useful for finding characteristics that are highly similar and dissimilar between subjects. Almost all of the top nine features with the highest feature importance, except for the ROM of the T5 movement in the transverse plane and the Min L3 in the frontal plane, mapped movement in the sagittal plane and were based on the Min and Max operations (vertebra T2, T3, T8, T10, L1, L3, and L4). Hence, the sagittal plane contributed the most to the differentiation between individuals and the recognition of the individual. This could be explained by the s-curvature of the spine, which enables spinal functions

1
2
3 such as shock-absorption and weight distribution, and as such, is the foundation for
4
5 balance and stability (Institute for Quality and Efficiency in Health Care 2019). All of
6
7 this is particularly relevant when walking on a treadmill. In addition, a person's posture
8
9 (i.e., the base position of the vertebral bodies prior to walking) that has more
10
11 longitudinal inter-individual variation compared to the two other dimensions may also
12
13 lead to the prominent role of the sagittal plane in the recognition process. No primary
14
15 determining spinal section was apparent, as the two most important features are located
16
17 on nearly opposite sides of the spine under the scrutiny of ST. Nevertheless, the lumbar
18
19 spine seemed to be more prominently represented amongst the features and had the
20
21 highest importance. Adjacent or nearby vertebral bodies contributed with mostly the
22
23 same direction of movement (e.g., Min L3/L4) to their feature importance. Beyond the
24
25 sagittal plane, L3 also played an important role in the frontal plane, thereby contributing
26
27 to the recognition of the individual as well as to sex differentiation (Dindorf and
28
29 Konradi et al. 2020a). Surprisingly, none of the pelvic parameters belonged to the most
30
31 relevant features as one would expect, considering the crucial roles of pelvic
32
33 functioning during gait in the transverse plane, where it increases the functional limb
34
35 length and moves the hips closer to the midline, both reducing the limb obliquity that is
36
37 necessary to achieve the needed step length (Perry and Burnfield 1992). This knowledge
38
39 about prevalent unique spinal movement patterns might be of interest in the context of
40
41 personalized medicine.
42
43
44
45
46
47

48
49 Pairwise plotting of the top two features with the highest feature importance (the
50
51 Max of T3 and the Min of L4 movement in the sagittal plane) revealed that, even on this
52
53 level, some subjects were clearly separable from others (e.g., subjects B, J, P, Q).
54
55 However, for most of the test subjects, no clear separation was visible in the respective
56
57
58
59
60

1
2
3 2D visual space, which might explain the low accuracy (25.33 %) with the use of these
4
5 two features.

6
7
8 **Knowledge about these** features forms the first step in comparing subjects. On
9
10 the one hand, visual inspection using t-SNE and comparing the variable values can be
11
12 used for comparisons. On the other hand, **the** Euclidean distance metric helps to
13
14 compare subjects using an objective and continuous metric. As a next step, we wanted
15
16 to perform clustering to gain knowledge about possible groups that show similar
17
18 movement patterns of the spine.

19
20
21 In clinical contexts, longitudinal changes of the similarity score in subjects
22
23 might be of interest **for the detection of** early change (e.g., approaching pathologies) by
24
25 supporting experts with an objective metric. Furthermore, finding similarities between
26
27 different subjects may be useful for prospectively **estimating** treatment outcomes.
28
29 Regarding the comparison of healthy subjects and patients, absolute group assignments
30
31 often shorten the facts. **The** application of the approach and the determination of
32
33 similarity scores in health and patient data could be useful for **creating** a continuous
34
35 value, which seems more appropriate and useful for monitoring progress (e.g., during
36
37 rehabilitation) or continual changes.
38
39
40

41
42 In the present work, good results were obtained using the Euclidean distance
43
44 metric. However, future works should also consider other distance metrics (e.g.,
45
46 Manhattan distance, Cosine distance) and compare them, especially when classification
47
48 is harder. In the current study, **we** used the lowest distances of the samples in the
49
50 reference pool for recognizing the subjects. The selection of a threshold distance value
51
52 might be important for future works, especially when outliers are present in the data. In
53
54 this way, if the gait cycle of the test subject is an outlier, the highest similarity would
55
56
57
58
59
60

1
2
3 fall under the threshold value. Setting a threshold value might also be interesting for
4
5 automatic monitoring and detecting abnormal longitudinal changes in subjects.
6
7

8 Overall, it is questionable if the use of automatically extracted features using an
9
10 SNN makes sense and is appropriate in clinical contexts. The interpretation of the
11
12 automatically extracted features using an SNN with contrastive loss and triplet loss is
13
14 not directly possible because the extracted features describe the original movements on
15
16 an abstract level. The description of movements based on ROM, Min, and Max is not
17
18 very abstract, and therefore, interpretation is quite straightforward and findings can be
19
20 directly traced back to the original movement patterns. When using nine instead of five
21
22 features, similar accuracy was reached compared to the use of the automatically
23
24 extracted features. Therefore, accuracy alone cannot justify the use of an SNN in
25
26 clinical contexts.
27
28
29

30 This is a common challenge in the context of machine learning models. Many of
31
32 the models, including the applied SNN, show black-box characteristics and a lack of
33
34 transparency and interpretability because the reasons the models make certain decisions
35
36 and what they specifically learn are not transparent (Adadi and Berrada 2018). These
37
38 factors strongly limit practical application in clinical settings (Holzinger et al. 2017).
39
40
41

42 Overall, the current task of subject identification indicates that features based on
43
44 simple descriptive statistics seem more appropriate in clinical contexts. However, the
45
46 results also show that with the use of an SNN, data representations can be generated,
47
48 reducing intra-individual variability and making subjects more distinct. Therefore, the
49
50 approach may be superior when identification tasks are more difficult. Furthermore,
51
52 understanding which waveform components were used for generating features that are
53
54 invariant or show little variance might be of interest. Hence, the use of explainable
55
56 artificial intelligence (XAI) is a promising approach, and it has gained much interest in
57
58
59
60

1
2
3 recent years for making machine learning models interpretable and, thus, useable in
4
5 practical clinical applications (Dindorf and Teufl et al. 2020b; Horst et al. 2019).
6

7
8 ST methodology might limit our findings, as validity has only been shown for
9
10 dynamic stance measures (Mohokum et al. 2015) and also seems compromised in
11
12 certain etiologies with a patient's BMI of up to 35 (Wanke-Jellinek et al. 2019).
13

14 However, not less than strong reliability and reproducibility were demonstrated for its
15
16 use in dynamic gait analysis (Gipsman et al. 2014). Hence, the transfer from validated
17
18 stance measurements to dynamic gait analysis appears to be feasible. Furthermore, the
19
20 extent to which our results can be generalized to different measuring methods is still
21
22 unknown. As there is no standard to validate our findings and assumptions, we
23
24 considered aspects of our results that indicate internal validity, for example, the intra-
25
26 individual consistency as well as the inter-individual variation, as demonstrated by the
27
28 spinal fingerprint.
29
30
31
32

33 34 35 **5. Conclusion**

36 Using an SNN with the contrastive loss and triplet loss functions, as well as
37
38 simple descriptive statistics, the described method proved useful for identifying subjects
39
40 based on biomechanical spinal waveform data. We found intra-individual movement
41
42 patterns of the spine that made subject recognition possible. Therefore, the results
43
44 indicate a possible existence of a personal spinal 'fingerprint'. Using a data-driven
45
46 approach, it was possible to gain insight into meaningful task-relevant and
47
48 discriminative gait features. The present work forms the basis for an objective
49
50 comparison of subjects and the transfer of the proposed method to clinically relevant
51
52 use cases. A possible field of application is the monitoring or detection of longitudinal
53
54 changes (in physiology) in subjects.
55
56
57
58
59
60

Acknowledgments

Foremostly, we would like to thank all the participants of this study. We also acknowledge our colleagues for their support in the recruitment process and Theresa Dersch for visual inspection of the captured data. Finally, we want to express our gratitude to Kjell Heitmann, Amira Basic, and Helmut Diers for multifarious endorsements and technological support during this project.

Disclosure statement

No potential conflict of interest is reported by the authors.

Funding

This work was supported by the BMBF, under grant numbers 16SV7115 and 03IHS075B, and from funds of the intramural research funding for third-party funded projects of the University Medicine of the Johannes Gutenberg University Mainz.

References

- Adadi A, Berrada M. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*. 6:52138–52160. doi:10.1109/ACCESS.2018.2870052.
- Bhagwat N, Viviano JD, Voineskos AN, Chakravarty MM. 2018. Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data. *PLoS Comput Biol*. 14(9):e1006376. doi:10.1371/journal.pcbi.1006376.
- Chicco D. 2021. Siamese Neural Networks: An Overview. *Methods Mol Biol*. 2190:73–94. doi:10.1007/978-1-0716-0826-5_3.
- Dindorf C, Konradi J, Wolf C, Taetz B, Bleser G, Huthwelker J, Drees P, Fröhlich M, Betz U. 2020a. General method for automated feature extraction and selection and

- 1
2
3 its application for gender classification and biomechanical knowledge discovery of
4
5 sex differences in spinal posture during stance and gait. Computer methods in
6
7 biomechanics and biomedical engineering.1-9.
8
9 doi:10.1080/10255842.2020.1828375
10
11
12 Dindorf C, Teufl W, Taetz B, Bleser G, Fröhlich M. 2020b. Interpretability of Input
13
14 Representations for Gait Classification in Patients after Total Hip Arthroplasty.
15
16 Sensors. 20(16):4385. doi:10.3390/s20164385.
17
18
19 Ekinci, M. 2006. A New Approach for Human Identification Using Gait Recognition.
20
21 In: Gavrilova M. et al. (eds) Computational Science and Its Applications - ICCSA
22
23 2006. ICCSA 2006. Lecture Notes in Computer Science. Springer, Berlin,
24
25 Heidelberg. 3982. https://doi.org/10.1007/11751595_128
26
27
28 England, SA, Granata, KP. 2007. The influence of gait speed on local dynamic stability
29
30 of walking. Gait Posture. 25(2). doi:10.1016/j.gaitpost.2006.03.003.
31
32
33 Figueroa-Mata G, Mata-Montero E. 2020. Using a Convolutional Siamese Network for
34
35 Image-Based Plant Species Identification with Small Datasets. Biomimetics (Basel).
36
37 5(1):8. doi:10.3390/biomimetics5010008.
38
39
40 Gipsman A, Rauschert L, Daneshvar M, Knott P. 2014. Evaluating the Reproducibility
41
42 of Motion Analysis Scanning of the Spine during Walking. Adv Med. 2014:721829.
43
44
45 Gordo A, Almazan J, Revaud J, Larlus D. 2016. Deep Image Retrieval: Learning global
46
47 representations for image search. In: Leibe B., Matas J., Sebe N., Welling M. (eds)
48
49 Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science.
50
51 Springer, Cham. 9910:241-257. doi.org/10.1007/978-3-319-46466-4_15.
52
53
54 Hadsell R, Chopra S, LeCun Y. 2006. Dimensionality Reduction by Learning an
55
56 Invariant Mapping. In: IEEE 2006. Computer Society Conference on Computer
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Vision and Pattern Recognition (CVPR'06); June 17-22; New York, NY, USA. New York: IEEE; 2:1735–1742. doi: 10.1109/CVPR.2006.100.

Hoffer E, Ailon N. 2018. Deep metric learning using Triplet network. arXiv pre-print server. [Submitted on 20 Dec 2014 (v1), last revised 4 Dec 2018 (this version, v4); accessed 2021 Feb 15]. <https://arxiv.org/pdf/1412.6622.pdf>.

Holzinger A, Biemann C, Pattichis CS, Kell DB. 2017. What do we need to build explainable AI systems for the medical domain? arXiv pre-print server. [updated 2017 Dec 28; accessed 2021 Feb 15]. <https://arxiv.org/pdf/1712.09923.pdf>

Horst F, Kramer F, Schafer B, Eekhoff A, Hegen P, Nigg BM, Schollhorn WI. 2016. Daily changes of individual gait patterns identified by means of support vector machines. *Gait Posture*. 49:309-314. doi:10.1016/j.gaitpost.2016.07.073.

Horst F, Slijepcevic D, Lapuschkin S, Raberger A-M, Zeppelzauer M, Samek W, Breiteneder C, Schöllhorn WI, Horsak B. 2019. On the Understanding and Interpretation of Machine Learning Predictions in Clinical Gait Analysis Using Explainable Artificial Intelligence. [updated 2019 Dec 16; accessed 2020 Mar 10]. <https://arxiv.org/pdf/1912.07737v1.pdf>.

Institute for Quality and Efficiency in Health Care (IQWiG, Germany). 2019. How does the spine work?. [Updated 2019 Feb 14; accessed 2021 Feb 15]. <https://www.informedhealth.org/how-does-the-spine-work.html>.

Kang HG, Dingwell JB. 2008. Separating the effects of age and walking speed on gait variability. *Gait Posture*. doi:10.1016/j.gaitpost.2007.07.009.

Kececi A, Yildirak A, Ozyazici K, Ayluctarhan G, Agbulut O, Zincir I. 2020. Implementation of machine learning algorithms for gait recognition. *Engineering Science and Technology, an International Journal*. 23(4):931-937. doi:10.1016/j.jestch.2020.01.005.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman DJ. 2021. 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*. 151:107398. doi:10.1016/j.ymssp.2020.107398.
- Li MD, Chang K, Bearce B, Chang CY, Huang AJ, Campbell JP, Brown JM, Singh P, Hoebel KV, Erdoğan D, et al. 2020. Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging. *NPJ Digit Med*. 3:48. doi:10.1038/s41746-020-0255-1.
- Liu C-F, Padhy S, Ramachandran S, Wang VX, Efimov A, Bernal A, Shi L, Vaillant M, Ratnanather JT, Faria AV, et al. 2019. Using deep Siamese neural networks for detection of brain asymmetries associated with Alzheimer's Disease and Mild Cognitive Impairment. *Magn Reson Imaging*. 64:190–199. doi:10.1016/j.mri.2019.07.003.
- Liu H, Yu L. 2005. Toward Integrating Feature Selection Algorithms for Classification and Clustering. *IEEE Transactions on Knowledge and Data Engineering*. 17:491–502. doi:10.1109/TKDE.2005.66.
- Liu W, Zhang C, Ma H, Li S. 2018. Learning Efficient Spatial-Temporal Gait Features with Deep Learning for Human Identification. *Neuroinformatics*. 16:457–471. doi:10.1007/s12021-018-9362-4.
- Liu Y, Sui X, Liu C, Kuang X, Hu Y. 2020. Automatic Lumbar Spine Tracking Based on Siamese Convolutional Network. *J Digit Imaging*. 33:423–430. doi:10.1007/s10278-019-00273-5.
- Mohokum M, Schulein S, Skwara A. 2015. The validity of rasterstereography: a systematic review. *Orthopedic reviews*. 7(3):68-73.

- 1
2
3 Mundorff AZ, Vidoli G, Melinek J. 2006. Anthropological and radiographic
4 comparison of vertebrae for identification of decomposed human remains. *J Forensic*
5 *Sci.* 51(5):1002-1004.
6
7
8
9
10 Nguyen D-P, Phan C-B, Koo S. 2018. Predicting body movements for person
11 identification under different walking conditions. *Forensic Sci Int.* 290:303–309.
12 doi:10.1016/j.forsciint.2018.07.022.
13
14
15
16 Ognard J, Deloire L, Saccardy C, Burdin V, Ben Salem D. 2020. Automated contour
17 detection in spine radiographs and computed tomography reconstructions for
18 forensic comparative identification. *Forensic Science, Medicine and Pathology.*
19 16(1):99-106.
20
21
22
23
24
25
26 Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A,
27 Antiga L, Lerer A. 2017. Automatic differentiation in PyTorch. NIPS 2017.
28 Proceedings of the 31st Conference on Neural Information Processing Systems;
29 December 4–9; Long Beach, CA, USA. [accessed 2021 Feb 15].
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Perry J, Burnfield JM. 1992. *Gait Analysis: Normal and Pathological Function.*
Thorofare, New Jersey: SLACK Incorporated.
- Randhavane T, Bera A, Kapsaskis K, Bhattacharya U, Gray K, Manocha D. 2019.
Identifying Emotions from Walking using Affective and Deep Features. [accessed
2021 Feb 15]. <https://arxiv.org/pdf/1906.11884.pdf>.
- Tong S, Fu Y, Ling H. 2017. Verification-based pairwise gait identification. 2017 IEEE
International Conference on Multimedia & Expo workshops (ICMEW), Hong Kong.
669–673. doi: 10.1109/ICMEW.2017.8026299.

- 1
2
3 Valenzuela A. 1997. Radiographic comparison of the lumbar spine for positive
4 identification of human remains. A case report. *Am J Forensic Med Pathol.*
5 18(2):215-217.
6
7
8
9
10 van der Maaten L, Hinton G. 2008. Visualizing Data using t-SNE. *Journal of Machine*
11 *Learning Research.* 9:2579–2605.
12
13
14 Wanke-Jellinek L, Heese O, Krenauer A, Wurtinger C, Siepe CJ, Wiechert K, Mehren
15 C. 2019. Is there any use? Validity of 4D rasterstereography compared to EOS 3D
16 X-ray imaging in patients with degenerative disk disease. *Eur Spine J.* 28(9):2162-
17 2168.
18
19
20
21
22
23 Weinberger KQ, Saul LK. 2009. Distance Metric Learning for Large Margin Nearest
24 Neighbor Classification. *Journal of Machine Learning Research.* 10:207–244.
25
26
27
28 Wu Z, Huang Y, Wang L, Wang X, Tan T. 2017. A Comprehensive Study on Cross-
29 View Gait Based Human Identification with Deep CNNs. *IEEE Transactions on*
30 *Pattern Analysis and Machine Intelligence.* 39:209–226.
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Zhang C, Liu W, Ma H, Fu H. 2016. Siamese neural network based gait recognition for human identification. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 3/20/2016 - 3/25/2016; Shanghai, China. 2832–2836. doi: 10.1109/ICASSP.2016.7472194.
- Zhao F, Xu J, Lin Y. 2018. Similarity Measure for Patients via A Siamese CNN Network. In: Vaidya J, Li J, editors. *Algorithms and Architectures for Parallel Processing.* Cham: Springer International Publishing. 11335:319–328.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review Only

Tables

Table 1. Descriptive statistics of the sample of subjects for the two data pools. Data 2 was recorded three times for every subject (Day 1, Day 2, 30 ± 7 days afterwards).

	Data A	Data B
Measurement days	1	3
Subjects (n)	201	25
Sex (n; female/male)	132/69	13/12
Age (years)	41.28 ± 13.42 18 – 70	34.68 ± 12.07 22 – 64
BMI (kg/m²)	23.49 ± 3.21 17.5 – 29.9	24.01 ± 3.45 18.6 – 29.9

Table 2. Results of the top five extracted features using the three approaches. Range of motion (ROM), minima (Min), and maxima (Max) represent the feature extraction using simple descriptive statistics.

	Contrastive Loss	Triplet Loss	ROM, Min, Max
Validation set accuracy	100 % True: 402, False: 0	96.52 % True: 388, False: 14	95.02 % True: 382, False: 20
Test set accuracy	96.00 % True: 144, False: 6	85.33 % True: 128, False: 22	82.00 % True: 123, False: 27

Figures

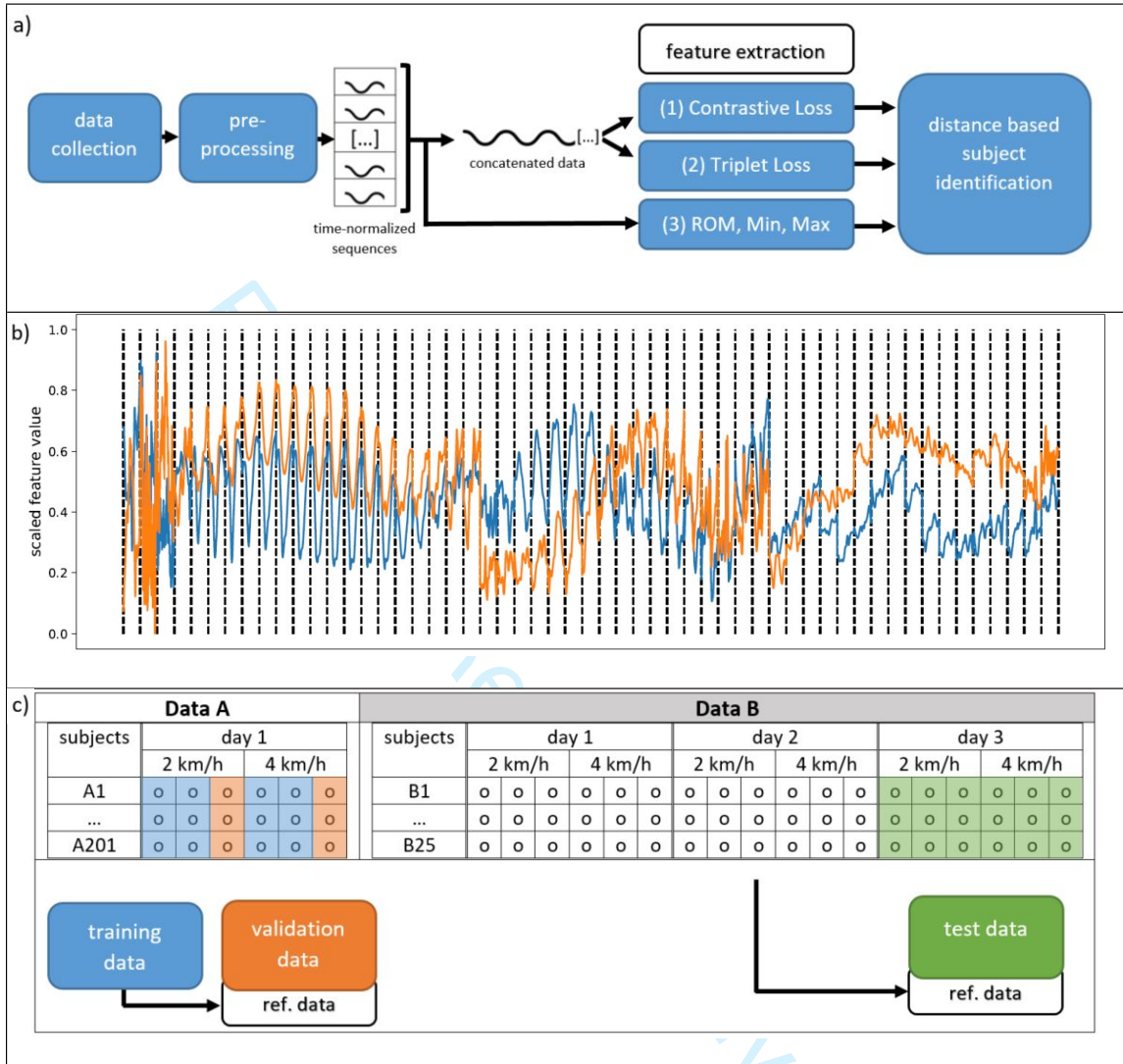


Figure 1. (a) Overview of the feature extraction procedure. The time-normalized sequences are directly used for features extraction using simple descriptive statistics (3). The concatenated sequences are used for the SNN approaches (1) and (2). ROM = range of motion, Min = minimal value, Max = maximal value. (b) Raw concatenated data of two exemplary subjects. The data were used for feature extraction using contrastive loss and triplet loss. Vertical dashed lines highlight the individual time-normalized sequences of the 55 parameters used. (c) Data used for training, validation, and testing. o = data of a single gait cycle; ref. data = reference data used for comparison.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

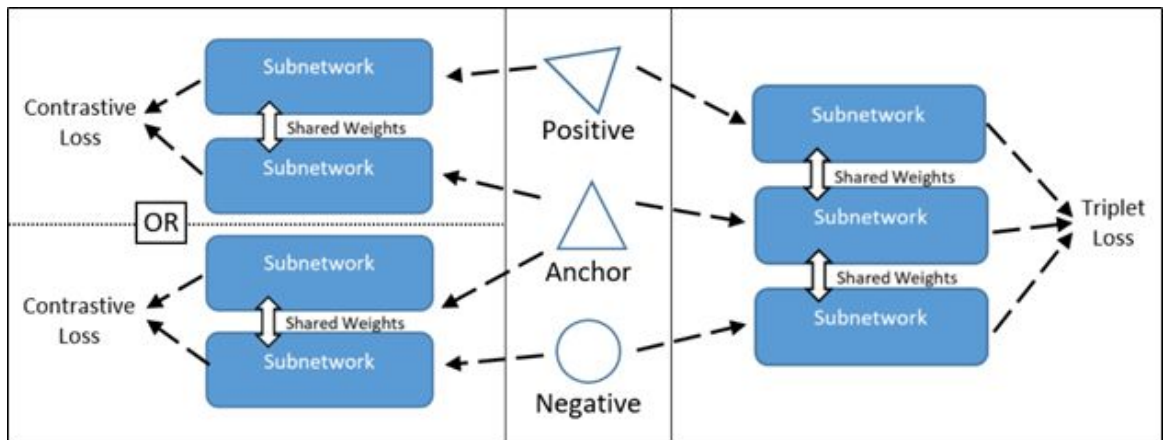


Figure 2. Siamese network architecture for contrastive loss (left) and triplet loss (right) functions.

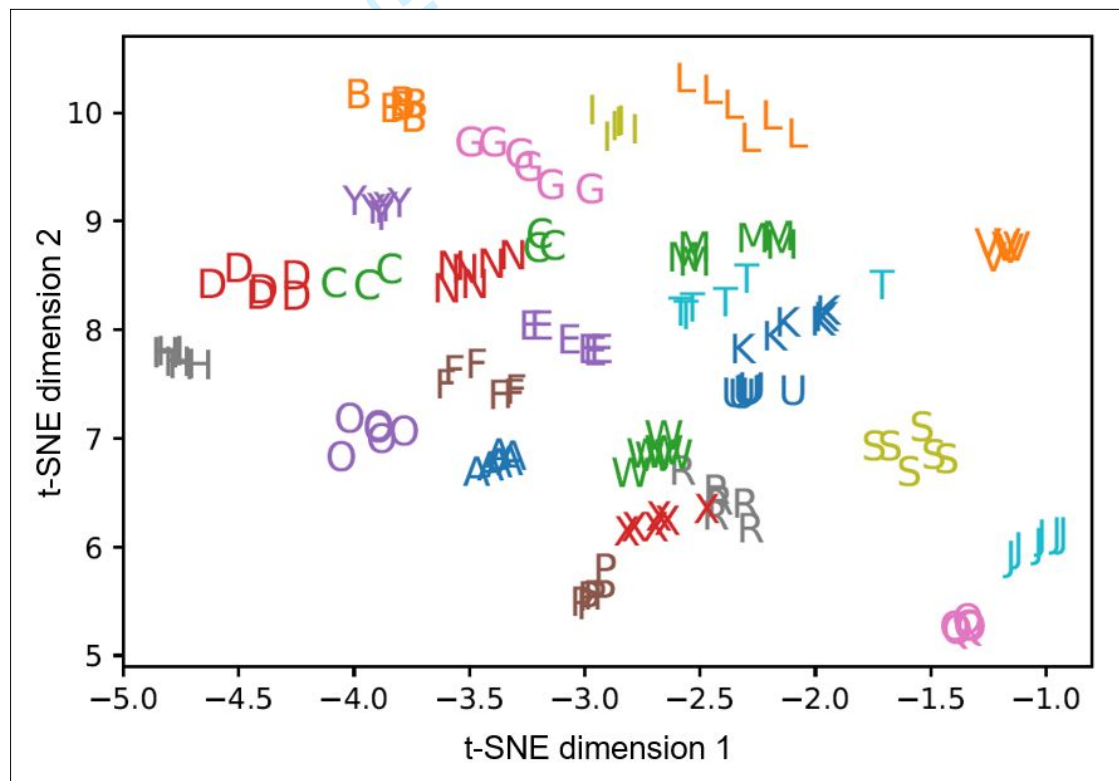


Figure 3. t-SNE visualization of the test subjects for the concatenated waveform data before feature extraction.

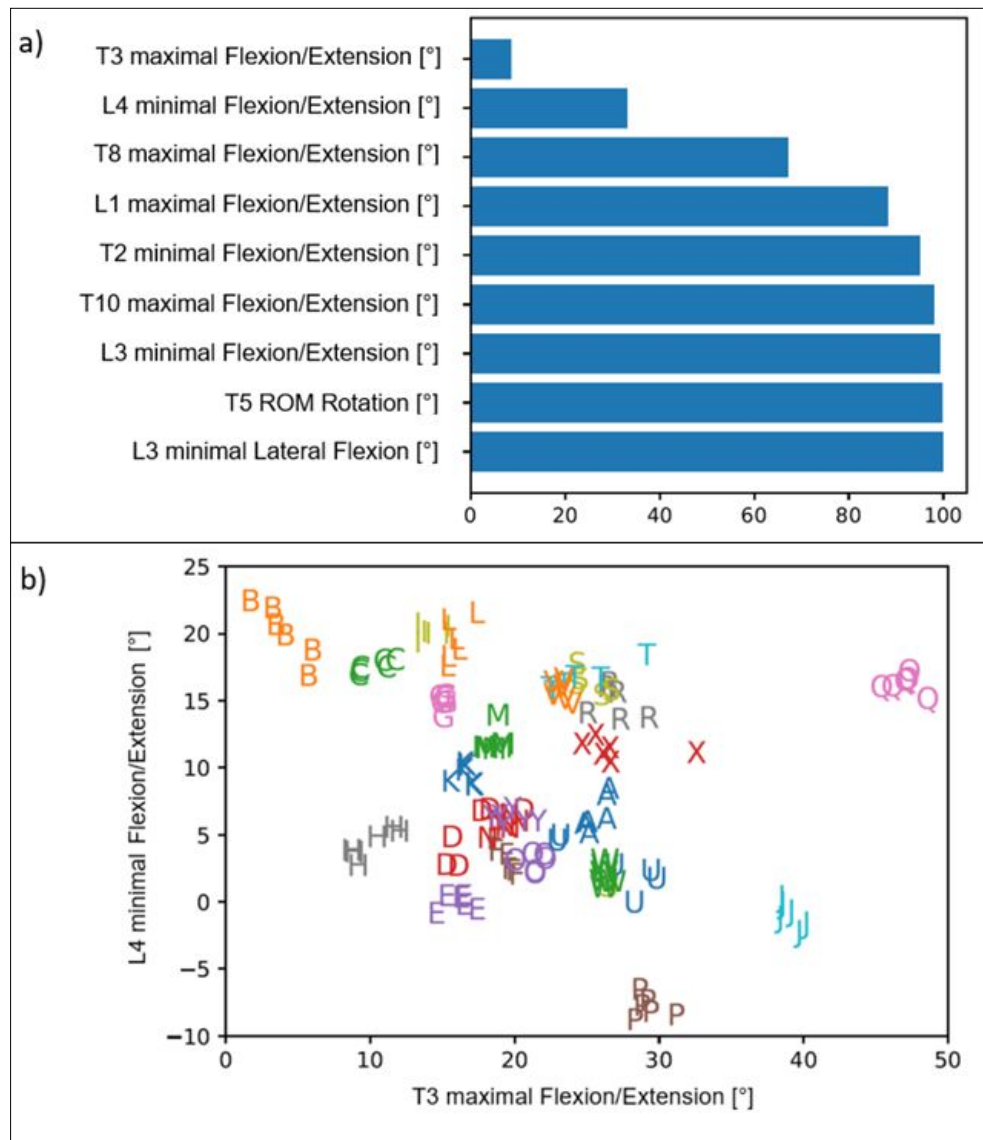


Figure 4. (a) Accuracy of the validation set during sequential feature selection of the features calculated using range of motion (ROM), minima, and maxima. The top feature in the bar plot was selected first. Gradually, from the top to the bottom, the features were then selected and included in the subset. (b) 2D visualization of the test subjects using the top two ranked features during sequential selection. For plotting and better interpretation, the features were rescaled to their original feature space.

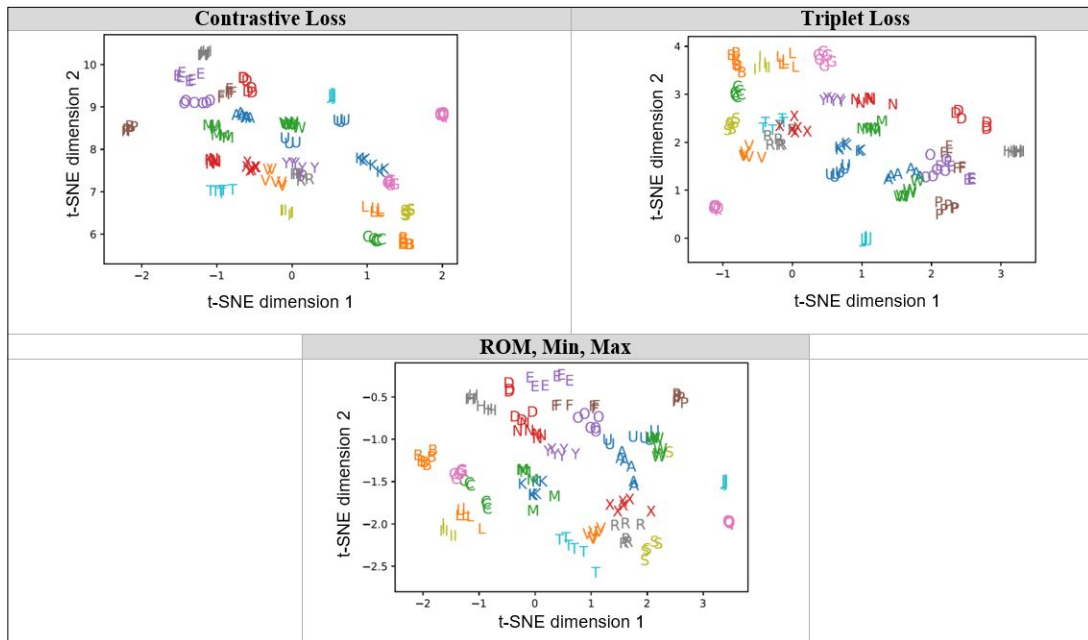


Figure 5. Test set subjects in a 2D space after performing t-SNE on the representations of the three different feature extraction approaches. Range of motion (ROM), minima (Min), and maxima (Max) represent the feature extraction using simple descriptive statistics.

URL: <http://mc.manuscriptcentral.com/gcmb>

Supplementary Materials

Table. Used variables. Variables were measured in sagittal, frontal, and transversal movement directions. Definitions were adapted from the DIERS manual, and the table was adapted from Dindorf and Konradi et al. (2020a).

Variable	Plane	Description
VP/C7 (Vertebra prominens / 7. cervical vertebral body)	Sagittal: Vertebral sagittal Flexion and Extension [°]	The parameter describes the inclination of the calculated vertebra in space (relative to a plumb/gravity line) as seen from a left view. The angle (in degrees) is calculated from the projection of the vertebra in a sagittal plane (rotation and lateral flexion are ignored). A positive value means a forward tilt of the vertebra (flexion).
T1 – T12 (Thoracic spine)		
L1 – L4 (Lumbar spine)	frontal: Vertebral Lateral Flexion [°]	The parameter describes the lateral inclination of the vertebra in space (relative to a plumb/gravity line) as seen from a posterior–anterior view. The angle (in degrees) is calculated from the projection of the vertebra in the coronal plane (rotation and sagittal extension/flexion are ignored). A positive value means a tilt of the vertebra to the left (lateral flexion left).
	transversal: Vertebral Rotation [°]	The vertebral rotation describes the rotation of a vertebra in the transversal plane (relative to the neutral pelvis). A positive value means a vertebra is rotated to the left (counterclockwise) when seen from behind. The rotation of vertebral bodies happens in situ, and therefore, the direction of rotation between the surface and vertebral rotation changes. Hence, a surface rotation to the right, mathematically represented with a +, becomes a vertebral body rotation to the left.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

		This is due to the calculation process in which a vector is used that points from the Processus spinosus towards the middle of the vertebral body, meaning that the surface rotation changes its direction within the vertebral body.
Pelvis	Pelvic Obliquity [°]	A line is drawn from DL to DR (left and right dimple) and is compared to a horizontal line representing the horizon. The angle (in degrees) between them is measured. A positive value means that the right pelvis is elevated.
	Pelvic Torsion (dimples) [°]	The parameter describes the torsion of the surface normals on the two lumbar dimples.
	Pelvic Inclination (dimples) [°]	The parameter describes the mean vertical torsion of the two surface normals on DL and DR.
	Pelvic Inclination (symmetry line) [°]	The parameter describes the angle of the vertical components of the surface normals on point DM (dimple midpoint) based on the horizontal.
	Pelvic Rotation [°]	The pelvic rotation is the rotation in the transversal plane of the right dimple relative to a reference coronal plane that is defined from the system setup, perpendicular to the camera-projection axis. A positive value means the pelvis is rotated to the left when seen from behind (the value is corrected * (-1)).

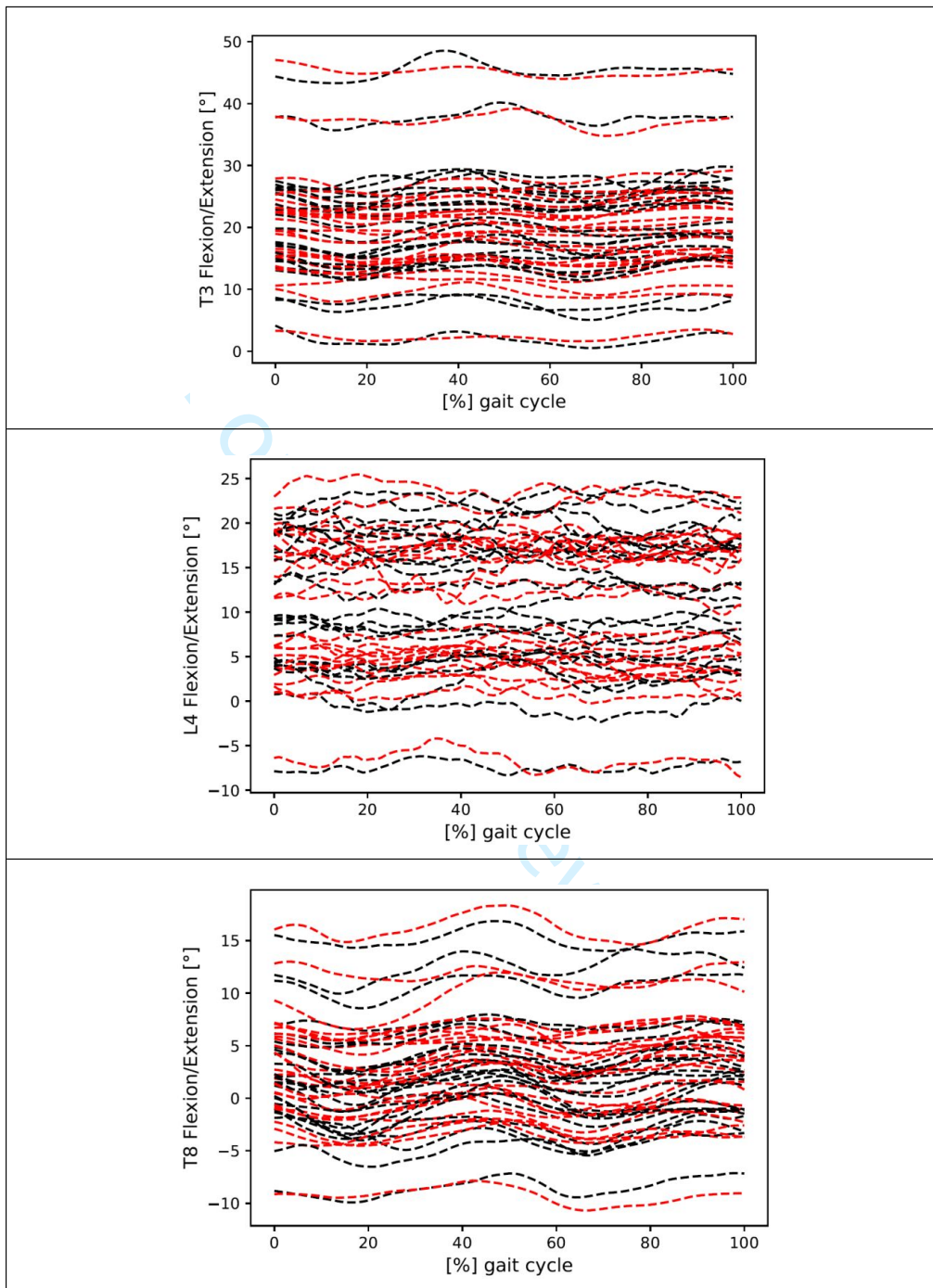


Figure. Exemplary visualization of the time-normalized sequence data for T3, T8, and L4 movement in the sagittal plane of the 25 test subjects. For every test subject, one measurement of each walking speed is displayed (black = 2 km/h, red = 4 km/h).

3.5 Vorbereitende Publikation für Publikation 5



Automated detection and explainability of pathological gait patterns using a one-class support vector machine trained on inertial measurement unit based gait data

W. Teufl, B. Taetz, M. Miezal u. a., “Automated detection of pathological gait patterns using a one-class support vector machine trained on discrete parameters of IMU based gait data,” *Clinical Biomechanics*, Jg. 89, S. 1–7, 2021. DOI: 10.1016/j.clinbiomech.2021.105452

Zusammenfassung

Hintergrund: ML Ansätze zur Klassifikation von pathologischen Gangbildern werden üblicherweise im Sinne eines Mehrklassen-Klassifizierungsproblems verwendet. Es fehlt jedoch an Forschung zu Ein-Klassen-Klassifikatoren, die unabhängig von bestimmten Pathologien sind. Daher war es das Ziel dieser Arbeit, einen Ein-Klassen-Klassifikator basierend auf gesunden Normdaten zu entwerfen, der nicht nur eine Vorhersagewahrscheinlichkeit, sondern vielmehr eine Erklärung der Klassifizierungsentscheidung liefert, um die Akzeptanz dieses maschinellen Lernansatzes zu erhöhen.

Methodische Forschungsfragen: Eignet sich eine OCSVM als pathologieunabhängiger, interpretierbarer Klassifikator?

Fachspezifische Forschungsfragen: Welche Unterschiede im Gang ermöglichen eine Klassifikation als gesund bzw. anormal bei Personen nach Hüft-TEP bzw. bei einer Person mit verschiedenen Prothesen nach Oberschenkelamputation?

Methodik: Auf Basis von IMU-Gangkinematik-Daten wurden 30 diskrete Merkmale berechnet und 25 gesunde Probanden zum Training einer einklassen Support-Vector-Maschine verwendet. 25 gesunde Probanden, 20 Patienten nach Hüft-TEP und eine Oberschenkelamputierte Person dienten der Validierung des Klassifikators. Vorhersagewahrscheinlichkeiten und Featurerelevanzen (Permutation Feature Importance) wurden für jedes Subjekt geschätzt.

Ergebnisse: 100 % der Patienten wurden als Ausreißer aus der gesunden Gruppe korrekt klassifiziert. Als Ausreißer wurden drei gesunde Probanden fehlklassifiziert. Die Hüfte in der Sagittalebene stellt das relevanteste Merkmal für die Gruppe nach Hüft-TEP dar. Für die falsch klassifizierte gesunde Person mit dem niedrigsten Wahrscheinlichkeitswert wurden die Kniebeugung und die Beckenschiefelage identifiziert.

Fazit: Die OCSVM scheint ein nützliches Werkzeug zu sein, um Ausreißer aus einer gesunden Normgruppe zu identifizieren. Die Featurerelevanz lieferte wertvolle Informationen über den muskuloskelettalen Status der Probanden. In dieser Kombination könnte der vorliegende Ansatz in verschiedenen Disziplinen eingesetzt werden, um Gangstörungen zu erkennen und darauf aufbauend ein Training vorzuschlagen sowie Therapiemaßnahmen prä- und postoperativ individuell anzupassen und zu überwachen. Die Ergebnisse deuten eine Anschlussfähigkeit für fortgeschrittene XAI Ansätze an.

3.6 5. Publikation



Classification and Automated Interpretation of Spinal Posture Data Using a Pathology-Independent Classifier and Explainable Artificial Intelligence (XAI)

C. Dindorf, J. Konradi, C. Wolf u. a., “Classification and Automated Interpretation of Spinal Posture Data Using a Pathology-Independent Classifier and Explainable Artificial Intelligence (XAI),” *Sensors*, Jg. 21, Nr. 18, S. 1–18, 2021. DOI: [10.3390/s21186323](https://doi.org/10.3390/s21186323)

Zusammenfassung

Hintergrund: Klinische Klassifikationsmodelle sind meistens pathologieabhängig und können daher nur Pathologien erkennen, für die sie trainiert wurden. Forschungsbedarf besteht bezüglich pathologieunabhängiger Klassifikatoren und deren Interpretation. Ziel ist es daher, einen pathologieunabhängigen Klassifikator zu entwickeln, der Vorhersagewahrscheinlichkeiten und Erklärungen für die Klassifikationsentscheidungen liefert.

Methodische Forschungsfragen: Eignet sich eine OCSVM im Vergleich zu klassischen binären Klassifikatoren als pathologieunabhängiger, interpretierbarer Klassifikator?

Fachspezifische Forschungsfragen: Welche pathologischen Abweichungen bei den Krankheitsbildern Rückenschmerz, Wirbelsäulenversteifung und Arthrose lassen sich im Vergleich zu Gesunden bei statischer Wirbelsäulenbetrachtung finden? Welche Pathologien sind am besten abgrenzbar bei Gesunden?

Methodik: Für die Modellierung wurden Wirbelsäulenhaltungsdaten von gesunden Probanden ($n = 25$) und verschiedene Pathologien (Rückenschmerzen $n = 32$, Wirbelsäulenversteifung $n = 34$, Arthrose $n = 60$) sowie synthetische Daten verwendet. Als pathologieunabhängiger Klassifikator wurde eine OCSVM verwendet. Die Ausgaben wurden nach Platt's Methode in eine Wahrscheinlichkeitsverteilung transformiert. Die Interpretation wurde unter Verwendung des XAI Tools Local Interpretable Model-Agnostic Explanations (LIME) durchgeführt. Die Ergebnisse wurden mit denen verglichen, die durch allgemein verwendete binäre Klassifikationsansätze erhalten wurden.

Ergebnisse: Die besten Klassifikationsergebnisse wurden für Probanden mit einer Wirbelsäulenversteifung erzielt ($MCC = 0,80$). Probanden mit Rückenschmerzen waren besonders schwierig von der gesunden Referenzgruppe zu unterscheiden ($MCC = 0,54$). Die vorgeschlagene Methode erwies sich als nützlich für die Interpretation der Vorhersagen. Es wurde keine klare Unterlegenheit des vorgeschlagenen Ansatzes im Vergleich zu allgemein verwendeten binären Klassifikatoren nachgewiesen.

Fazit: Die Anwendung dynamischer Wirbelsäulendaten scheint für zukünftige Arbeiten wichtig zu sein. Das vorgeschlagene Vorgehen könnte hilfreich sein, um eine objektive Orientierung zu geben und Therapiemaßnahmen prä- und postoperativ individuell anzupassen und zu überwachen.

Article

Classification and Automated Interpretation of Spinal Posture Data Using a Pathology-Independent Classifier and Explainable Artificial Intelligence (XAI)

Carlo Dindorf ^{1,*†}, Jürgen Konradi ^{2,†}, Claudia Wolf ², Bertram Taetz ³, Gabriele Bleser ³, Janine Huthwelker ², Friederike Werthmann ⁴, Eva Bartaguiz ¹, Johanna Kniepert ⁴, Philipp Drees ⁴, Ulrich Betz ^{2,‡} and Michael Fröhlich ^{1,‡}

Citation: Dindorf, C.; Konradi, J.; Wolf, C.; Taetz, B.; Bleser, G.; Huthwelker, J.; Werthmann, F.; Bartaguiz, E.; Kniepert, J.; Drees, P.; et al. Classification and Automated Interpretation of Spinal Posture Data Using a Pathology-Independent Classifier and Explainable Artificial Intelligence (XAI). *Sensors* **2021**, *21*, 6323. <https://doi.org/10.3390/s21186323>

Academic Editors: Italo Zoppis, Sara Lucia Manzoni and Giulia Cisotto

Received: 2 August 2021

Accepted: 17 September 2021

Published: 21 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

- ¹ Department of Sports Science, Technische Universität Kaiserslautern, 67663 Kaiserslautern, Germany; eva.bartaguiz@sowi.uni-kl.de (E.B.); michael.froehlich@sowi.uni-kl.de (M.F.)
 - ² Institute of Physical Therapy, Prevention and Rehabilitation, University Medical Centre of the Johannes Gutenberg University Mainz, 55122 Mainz, Germany; juergen.konradi@unimedizin-mainz.de (J.K.); claudia.wolf@unimedizin-mainz.de (C.W.); janine.huthwelker@unimedizin-mainz.de (J.H.); ulrich.betz@unimedizin-mainz.de (U.B.)
 - ³ Department Augmented Vision, German Research Center for Artificial Intelligence, 67663 Kaiserslautern, Germany; bertram.taetz@dfki.de (B.T.); gabriele.bleser@dfki.de (G.B.)
 - ⁴ Department of Orthopedics and Trauma Surgery, University Medical Centre of the Johannes Gutenberg University Mainz, 55122 Mainz, Germany; fwerthma@students.uni-mainz.de (F.W.); johanna.kniepert@unimedizin-mainz.de (J.K.); philipp.drees@unimedizin-mainz.de (P.D.)
- * Correspondence: carlo.dindorf@sowi.uni-kl.de; Tel.: +49-631-205-5172
 † Mutual first authors/authors contributed equally.
 ‡ Mutual senior authors/authors contributed equally.

Abstract: Clinical classification models are mostly pathology-dependent and, thus, are only able to detect pathologies they have been trained for. Research is needed regarding pathology-independent classifiers and their interpretation. Hence, our aim is to develop a pathology-independent classifier that provides prediction probabilities and explanations of the classification decisions. Spinal posture data of healthy subjects and various pathologies (back pain, spinal fusion, osteoarthritis), as well as synthetic data, were used for modeling. A one-class support vector machine was used as a pathology-independent classifier. The outputs were transformed into a probability distribution according to Platt's method. Interpretation was performed using the explainable artificial intelligence tool Local Interpretable Model-Agnostic Explanations. The results were compared with those obtained by commonly used binary classification approaches. The best classification results were obtained for subjects with a spinal fusion. Subjects with back pain were especially challenging to distinguish from the healthy reference group. The proposed method proved useful for the interpretation of the predictions. No clear inferiority of the proposed approach compared to commonly used binary classifiers was demonstrated. The application of dynamic spinal data seems important for future works. The proposed approach could be useful to provide an objective orientation and to individually adapt and monitor therapy measures pre- and post-operatively.

Keywords: spine; machine learning; artificial intelligence; data mining; biomechanics; back pain; osteoarthritis; spinal fusion; explainable artificial intelligence

1. Introduction

In outpatient care, back pain, and knee joint osteoarthritis are among the 30 most common individual diagnoses, with approximately 20% of 18–79-year-olds having physician-diagnosed osteoarthritis, whereby knee and hip joints are most commonly affected [1]. Back pain is considerably lowering the quality of life across all income and age groups

and is now the leading cause of disability worldwide [2] with a point prevalence of 25–40% [3], a 12-month prevalence of approximately 61% [4], for example in the German population, and a lifetime prevalence in the American population of up to 85% [5]. Hence, more emphasis should be put on the back. Back pain may be due to a specific vertebro-genic cause, such as spondylitis, herniated disc, or spinal stenosis [6]. However, 90% of back pain is stated to be nonspecific, whereby no anatomic correlate can be detected as the specific cause that requires treatment [7]. Non-specific back pain can be caused by altered function; moreover, pathologies, such as knee or hip osteoarthritis, may cause postural changes or back pain. This is because the body reacts to a pain stimulus with avoidance in both stance and gait, which leads to poor or relieved postures [8]. For those patients diagnosed with a specific cause of their back pain, such as spondylolisthesis, surgery is often required. In the field of spondylodesis, in particular, there has been a large increase in surgery rates in recent years [9,10], even though patients do not always clearly benefit from this type of surgery—most notably, those with persisting pain [11,12]. Hence, objectifiable data, such as radiography or magnetic resonance imaging, are not able to explain this dissatisfying postoperative outcome alone. Besides daily-life monitoring data of wearables [13,14], a posture and motion analysis, combined with artificial intelligence, might be able to provide useful insights.

Artificial intelligence (AI) and machine learning approaches are gaining increasing interest in the field of biomechanical clinical data analysis and have obtained promising results (e.g., after stroke [15] or in Parkinson's disease [16]). They have proven to be useful in analyzing complex and multivariate data, giving objective orientation and finding discriminative group-specific differences [17,18]. Furthermore, they have even shown advantages over commonly used inference-based statistical analysis methods in those databases [19–21]. In the clinical context, they are able to identify pathologic characteristics and even surpass human guidance in the detection of diseases [22,23]. Additionally, they might be able to reduce false-positive mistakes and differences in disease detection based on the different experience levels of the medical practitioners [24]. However, regarding the application of machine learning methods on spinal data for the mentioned pathologies, there is a lack of research. Using machine learning approaches, existing classification studies have investigated pathologies such as osteoarthritis [25] or total hip arthroplasty [17,26]. To the best of the authors' knowledge, no studies regarding back pain and spinal fusion (spondylodesis) are currently available.

Many machine learning models often show black box characteristics and a lack of transparency [27]. For the user, it is, therefore, hard to trust in the model and its decision because it is opaque—why does the model make certain decisions and what has the model really learned [28]? This opacity does not comply with the requirements of the European General Data Protection Regulation (GDPR, EU 2016/679) [29] and strongly limits practical applications in clinical contexts [30]. Recently, through advances in the application of explainable artificial intelligence (XAI) methods in the biomechanical clinical domain, machine learning is becoming more and more applicable in practical clinical settings [31,32]. XAI offers methods for increasing the trustworthiness and transparency of black box models [27]. The prominent interpretation tools are, for example, Local Interpretable Model-Agnostic Explanations (LIME) [33], SHapley Additive exPlanations (SHAP) [34], and Deep Learning Important Features (DeepLIFT) [35]. The usage of XAI has shown to be particularly useful in understanding individual pathologic differences and is, therefore, of high relevance in the context of personalized medicine, e.g., to monitor therapy measures pre- and post-operatively [31,32].

Most works are based on binary classification tasks for knowledge discovery through distinguishing between healthy subjects and subjects with a certain pathology (e.g., [17,25,26,31,32]). For model training, both the data of healthy subjects and of the pathology of interest are often used. The resulting models are, therefore, only applicable for a certain pathology—they are pathology-dependent; hence, they are only able to detect the certain pathology for which they have been trained. To obtain a model for another pathology of

interest, completely new training with subjects of the respective pathology has to be performed, even if the training data of healthy subjects remain the same.

Differences in comparison to a healthy group are usually of interest in the clinical context. Therefore, the question arises if it is possible to develop a pathology-independent classifier by only learning the characteristics of healthy subjects in order to recognize any (pathological) deviations. Unsupervised outlier detection methods may have the potential to perform this. A previous work [36] used a one-class support vector machine (OCSVM) and obtained the first promising results. Transferred to the terminology of an outlier detection task, a decision function was learned based on regular observations (healthy subjects) and outliers (subjects with a pathology; in the quoted case, total hip arthroplasty, and one transfemoral amputee) detected due to their location outside the learned frontier.

Overall, there is a lack of research regarding those classifiers and their interpretation. According to current practice, they are treated as black boxes. This does not comply with GDPR and strongly limits applications in the clinical context. Furthermore, it is questionable whether XAI tools are suitable for interpretation in this context. Therefore, the aim of this research is to design and evaluate a pathology-independent classifier that provides explanations of classification decisions using XAI. Using clinical spinal posture data of the clinically highly relevant pathologies back pain, spinal fusion, and osteoarthritis, we wanted to check whether it is possible to adequately explain pathologic differences compared to healthy controls with a pathology-independent classifier and to generate clinically relevant insights. The classification results are compared with those obtained using commonly used binary classification approaches.

2. Materials and Methods

2.1. Measuring Method, Data, and Subjects

Across four studies, we collected data of 151 subjects. Depending on the study design of the respective study (see Table 1), for every subject, on one or on three separate days, 12 postural data measurements of the spine averaging from 12 to 14 individual images were obtained for each day (36 total measurements for each healthy subject which were recorded on three days). Data were collected of healthy subjects, as well as subjects with various pathologies (back pain, spinal fusion, and osteoarthritis), using the DIERS formetric III 4D™, DICAM v3.7Beta analyzing system as a means for rasterstereography, also called surface topography (ST). All subjects received the same marker placement (vertebra prominens, both dimples, and shoulders). The subjects' characteristics are described in Table 1. The method allowed to measure the spine in all body planes without the use of invasive or radiation-based approaches or extensive preparation. Recently, in addition to static measurements, this method has proven useful in measuring dynamic spinal data [37,38]. (¹ The dataset is part of the dissertation project of Friederike Werthmann; ² The dataset is part of the dissertation project of Claudia Wolf.)

Table 1. Subject characteristics and related trials.

	Subjects (<i>n</i>)	Male (<i>n</i>); Female (<i>n</i>)	Age (SD)	Hight cm (SD)	BMI (SD)	Further Information
Healthy ¹ (asymptomatic)	25	12; 13	34.68 (12.07)	176.28 (8.83)	24.01 (3.45)	Repeated measurements at three points in time; walking without walking aids and pain; no acute or chronic diseases; no pregnancy; BMI < 30; WHO register (INT: DRKS00014325)
Back pain	32	14; 18	44.53 (14.84)	174.00 (11.00)	26.01 (4.79)	Area of pain: 6% thoracic spine (TS), 72% lumbar spine (LS), and 22% TS + LS; no acute fractures, walking aids, or acute/chronic illnesses that prevent safe walking; WHO register (INT: DRKS00013145)

Spinal fusion	34	20; 14	56.26 (15.40)	171.00 (11.00)	26.95 (4.43)	Spinal fusion somewhere between C7 and L5; no acute fractures, walking aids, or acute/chronic illnesses that prevent safe walking; WHO register (INT: DRKS00013145)
Osteoarthritis ²	60	29; 31	64.00 (11.27)	171.00 (9.15)	25.68 (2.35)	30 knee osteoarthritis and 30 hip osteoarthritis; walking without walking aids; no walking impairments that prevent safe walking; no acute or chronic diseases; no pelvic or spinal surgery; no pregnancy; BMI < 30; WHO register (INT: DRKS00017240)

Fifty-five static parameters provided by the system were used (pelvic obliquity (°), pelvic torsion (dimples) (°), pelvic inclination (dimples) (°), pelvic rotation (°), and orientation of VP, T1–12, and L1–L4 in all planes (°)) for modeling. For a detailed description of the parameters, see [37,39] and Appendix A. Data of healthy subjects were used for training the OCSVM. Consequently, only the outliers in the group of healthy subjects influenced the learning of a decision function. Therefore, outliers were removed for the data of the healthy subjects using isolation forest [40] that also pay attention to multivariate outliers. Of the 900 samples, 150 were removed (one subject was completely removed).

For evaluation of the classification performance in terms of the dependence of the class separation, synthetic subject data ($n = 24$ subjects) of known separation to the healthy reference subjects were created. The use of the synthetic data is intended to ensure that groups of different class separations are present to check the performance of the proposed approach in settings with well separated groups as well as settings with little separated groups. To maintain similar spinal characteristics compared to real-world data for the synthetic posture data, data creation was based on the principal components (PCs) of the healthy subjects. After removing the mean and scaling to unit variance, principal component analysis (PCA) was applied. Four different synthetic classes were created, each of which was based on a modification of one PC. Therefore, separately for the first four PCs, data were created by adding random samples from a normal Gaussian distribution. The center of the distribution was set according to the desired cluster separation (silhouette scores of synthetic data compared to healthy data—class S1: 0.49; class S2: 0.41; class S3: 0.33; class S4: 0.16) with a standard deviation equal to 1. Afterward, the synthetic PCA data were transformed back into the original space.

For every real and synthetic subject, 10 samples without replacement were drawn for further calculations.

2.2. Data Preprocessing and Model Evaluation

For both the pathology-independent classifier, as well as the binary classifiers, grouped K-fold cross-validation (KFold) with five folds (25–29 healthy subjects were used each time in the training, while 6–9 healthy subjects were used each time in the test set) was used for model evaluation to check if the model was able to generalize to new subjects. Therefore, the data were split considering the subjects to ensure that the samples of each subject were not present in both the training and the testing data. The data of each training fold were split into data for actual training and data for validation (validation size = 20%) to allow hyperparameter searching and probability calibration/calculation without introducing a bias. Standardization was performed based on each respective training set by removing the mean and scaling to unit variance.

2.3. Pathology-Independent and Binary Classifier

OCSVM [41] was used for building a pathology-independent classifier. Initially, the SVM algorithm was developed for binary classification tasks. In the case of one-class classification, the algorithm tries to capture the density of the majority class. Outliers are detected as extremes of the learned density function. The OCSVM was trained using training

samples of healthy subjects in the training set only. A random hyperparameter search was performed with the parameter “kernel” (linear, radial basis function), “nu,” and “gamma” using the samples of healthy subjects in the validation set for performance evaluation. The remaining hyperparameters were set to the default parameters of Scikit-learn [42]. Validation set performance was measured as the ratio of misclassified healthy subjects to the total number of healthy subjects in the validation set.

The classification output was transformed into a probability distribution over the classes according to Platt’s method [43]. Therefore, a logistic regression model was fitted to the output scores of the model with respect to the actual class labels. In the present case, logistic regression using five-fold CV for hyperparameter selection was fitted to the validation set for each group of subjects.

2.4. Validation

For comparing the results of the proposed methodology with those obtained using a classical binary classification method, the random forest classifier (RF) [44] was used (a preliminary study using spinal posture data of another group of subjects showed that the algorithm yielded similar or slightly better classification results compared to other regarded classifiers). A random hyperparameter search was performed with the parameters “n_estimators” and “max_depth” using the validation set performance. The remaining hyperparameters were set to the default parameters of Scikit-learn. The F1 score was used to measure the validation set performance. Probabilities for the RF model were calibrated according to Platt’s method described above.

2.5. XAI Interpretations

The model-agnostic interpretation tool Local Interpretable Model-Agnostic Explanations (LIME) [33] was used for explaining the predictions of the OCSVM or the binary classifier using probabilities calculated/calibrated with the described Platt’s method. To explain how a black box model makes a single prediction, LIME performs an approximation of a local prediction of a black box model with a simpler interpretable model. Therefore, data points around an instance of interest of a black box model were generated through perturbation. Afterward, these data points were predicted with the black box model and weighted by their proximity to the selected instance. Finally, an interpretable model was learned on the weighted data points and used for explaining the prediction [33]. Each trained model during KFold was used to explain the predictions of the respective test set. Detailed analyses of the XAI results were performed for the pathology with the highest classification performance. Subject-specific results were exemplarily presented and discussed.

2.6. Evaluation Metrics and Calculations

Uncertainties were analyzed using classification probabilities. Due to partly imbalanced classes, predicted probabilities were evaluated using the Brier Skill Score (BSS; normalized by the naive score [45]). Classification performance was reported using the Matthews correlation coefficient (MCC), the F1 score, and a confusion matrix (CM) based on the classification founded on prediction probabilities. Calculations were performed in Python (Python Software Foundation, Wilmington, DE, USA) using Scikit-learn [42] and LIME [33].

3. Results

The results for the OCSVM, as well as the binary classification approach, are presented in Table 2. For the pathologies, the best classification performance was obtained for the spinal fusion data. The proposed OCSVM showed improved performance compared to the binary RF classifier for the respective data and the F1 and MCC scores. The

BSS was slightly reduced compared to the RF classifier. The worst classification performance was obtained for both approaches for the back pain data. Performance for the osteoarthritis data was between the classifications of spinal fusion and back pain.

Table 2. Classification results (mean and standard deviation) for the pathology-independent one-class classifier (OCSVM) and the binary classification approach using a random forest classifier (RF). Absolute class predictions were conducted according to the calculated/calibrated prediction probabilities. MCC, Matthews correlation coefficient; F1, F1 score; CM, confusion matrix.

	Data	One Class SVM				Binary RF Classifier					
		F1	MCC	BSS	CM	F1	MCC	BSS	CM		
Syn- thetic	S1	0.96 ± 0.05	0.92 ± 0.1	0.84 ± 0.12	220	2	1.0 ± 0.0	1.0 ± 0.01	0.93 ± 0.02	239	0
					20	238				1	240
	S2	0.99 ± 0.01	0.99 ± 0.02	0.95 ± 0.02	237	0	−0.98 ± 0.03	0.96 ± 0.07	0.88 ± 0.05	230	1
					3	240				10	239
S3	0.89 ± 0.03	0.77 ± 0.05	0.61 ± 0.12	203	19	−0.95 ± 0.04	0.90 ± 0.09	0.79 ± 0.12	223	6	
				37	221				17	234	
S4	0.82 ± 0.09	0.65 ± 0.19	0.46 ± 0.23	192	38	−0.90 ± 0.04	0.82 ± 0.06	0.65 ± 0.09	217	23	
				48	202				23	217	
Real	BP	0.54 ± 0.13	0.13 ± 0.19	0.02 ± 0.10	149	165	−0.62 ± 0.17	0.08 ± 0.34	−0.08 ±	98	113
					91	155			0.35	142	207
	Spinal fu- sion	0.80 ± 0.12	0.57 ± 0.23	0.33 ± 0.28	194	78	−0.74 ± 0.25	0.45 ± 0.25	0.36 ± 0.31	171	86
					46	262				69	254
Osteoar- thritis	0.69 ± 0.04	0.21 ± 0.12	0.35 ± 0.30	138	230	−0.78 ± 0.09	0.19 ± 0.21	41.28 ±	73	107	
				102	370			0.35	167	493	

For the synthetic data, the binary classifier performed better than the OCSVM (except S2). The difference was especially visible for the data with the lowest separation from the group of healthy subjects (MCC OCSVM: 0.65 ± 0.19; MCC RF: 0.82 ± 0.06). Surprisingly, OCSVM performance was reduced for S1 compared to S2, and even the class separation measured by means of the silhouette score increased for S1.

The local results for the pathology with the highest classification performance (subjects with spinal fusion) for the OCSVM, as well as the binary classifier, are presented in Table 3. Examples for one correctly and one falsely classified subject are given in Figure 1.

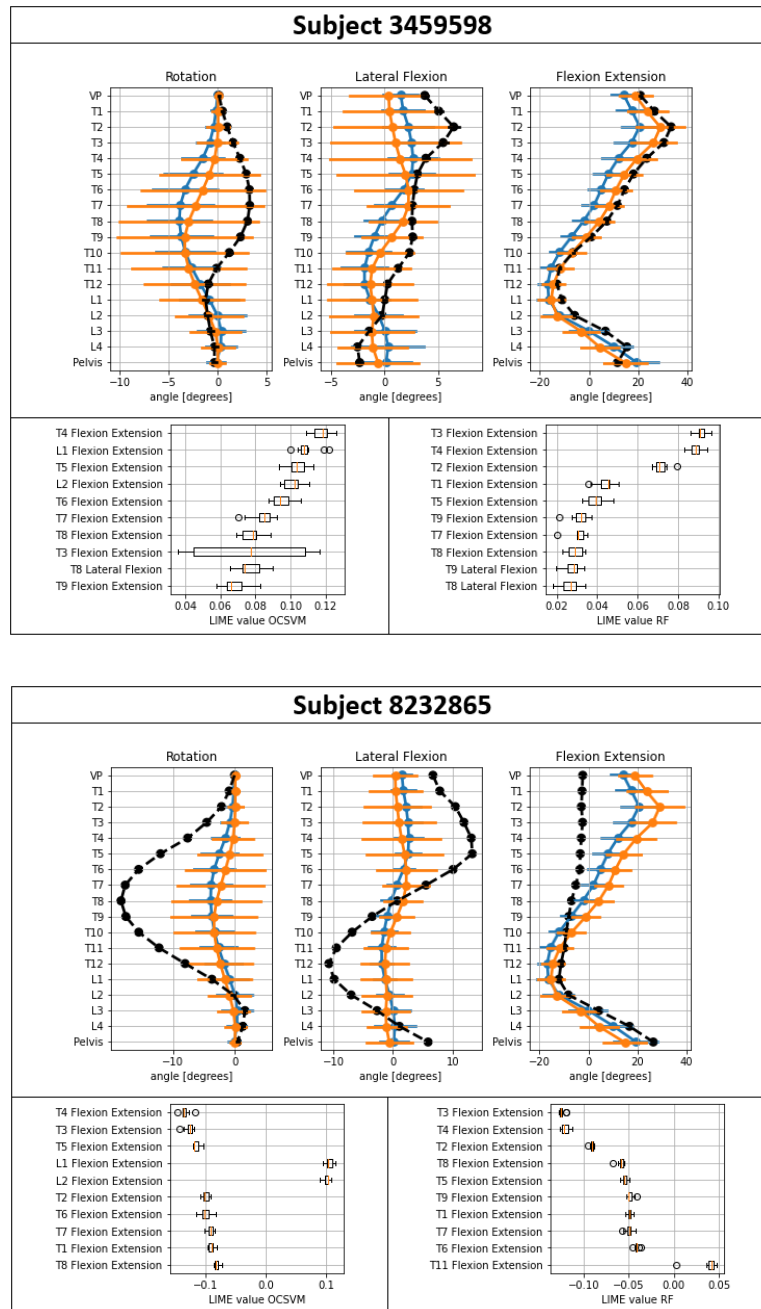


Figure 1. Exemplary posture of one correctly (3459598) and one falsely (8232865) classified subject. Bottom: Displayed LIME values show the effect for the 10 most important features. Negative values represent an effect toward the group of healthy subjects, with positive values indicating an effect that indicates an outlier (patient). Top: Vertebral body positions in the transversal (rotation), coronal (lateral flexion), and sagittal (flexion extension) planes. Positive values indicate a rotation/tilt to the left or ventral (toward flexion), while negative values indicate a rotation/tilt to the right or dorsal (toward extension). Blue = mean and standard deviation (SD) of healthy reference group; orange = mean and SD of group of subjects with the respective pathology; black = mean and SD of the 10 measurements of the subjects of interest.

Table 3. Results for the spinal fusion subjects (mean and standard deviation of the 10 measurements per subject), as well as localization of the spinal fusion and the features with the highest relevance according to the LIME results. The probabilities are for classification as an outlier/patient. P, pelvis.

Subject ID	OCSVM	RF	Location of Spinal Fusion	LIME	LIME
	Mean Prediction Probability	Mean Prediction Probability		OCSVM	RF
1962247	1.00 ± 0.00	0.92 ± 0.00	L1–S1	T8, T9, P	T9, T10, T7
3459598	1.00 ± 0.00	0.93 ± 0.00	T10–L2	T4, L1, T5	T3, T4, T2
7741511	0.99 ± 0.00	0.90 ± 0.01	L5–S1	T9, P, T3	T9, T10, T7
5777016	0.98 ± 0.01	0.92 ± 0.00	T10–L3	L4, P, P	T9, T7, T5
7475130	0.96 ± 0.01	0.73 ± 0.01	L3–S1	T4, T5, T6	T4, T3, T2
9342653	0.96 ± 0.01	0.80 ± 0.04	L4–L5	L4, P, L1	T9, T7, T8
3729138	0.90 ± 0.01	0.87 ± 0.02	L2–L3	P, P, T8	T3, T4, T7
5536002	0.87 ± 0.02	0.71 ± 0.03	T6–L3	L4, P, T2	T3, T4, T7
6705867	0.87 ± 0.02	0.84 ± 0.01	L5–S1	P, T12, L4	T3, T4, T12
5247355	0.83 ± 0.03	0.88 ± 0.02	L5–S1	P, L4, T8	T4, T5, T8
5297873	0.81 ± 0.03	0.80 ± 0.02	L4–L5	P, P, T7	T8, T7, T5
5408449	0.80 ± 0.02	0.71 ± 0.02	T4–L1	T7, T8, T12	T5, T4, L1
3336746	0.78 ± 0.02	0.93 ± 0.01	L3–L5	P, T8, P	T4, T5, T8
9747703	0.77 ± 0.01	0.94 ± 0.01	L3–L5	P, L4, T3	T4, T5, T8
2324908	0.76 ± 0.06	0.90 ± 0.01	T11–L3	P, L4, L3	T3, T4, T7
8398276	0.75 ± 0.05	0.43 ± 0.03	T10–L2	L4, P, T11	T3, T4, T7
3012624	0.70 ± 0.07	0.83 ± 0.07	L4–L5	P, T8, T12	T3, T4, T7
7767875	0.62 ± 0.02	0.82 ± 0.02	C6–T2	P, P, L4	T5, T4, T6
5815929	0.60 ± 0.03	0.92 ± 0.00	T12–L2	P, L4, P	T4, T5, T8
9621669	0.56 ± 0.09	0.34 ± 0.05	T12–L2	T4, T3, T5	T3, T4, T2
1082776	0.55 ± 0.00	0.43 ± 0.05	L2–L4	L4, T4, T5	T3, T11, L4
649887	0.53 ± 0.01	0.82 ± 0.01	L4–L5	VP, T3, T9	T3, T3, T4
7550216	0.53 ± 0.00	0.82 ± 0.02	T10–L5	VP, T3, T9	T3, T9, T6
3943929	0.51 ± 0.00	0.88 ± 0.01	L3–L4	L4, T9, L3	T9, T3, T4
5584179	0.51 ± 0.01	0.89 ± 0.01	L4–L5	T9, T4, VP	T9, T3, T8
6777530	0.51 ± 0.01	0.86 ± 0.01	L2–L4	VP, T3, L3	T3, T9, T6
9299446	0.50 ± 0.01	0.32 ± 0.03	T5–T10	L3, VP, T12	T3, T7, T3
632814	0.45 ± 0.01	0.76 ± 0.06	L4–S1	L4, VP, T3	T3, T9, T6
3035442	0.42 ± 0.00	0.71 ± 0.02	T11–L2	L4, VP, T3	T3, T9, T8
6683738	0.41 ± 0.09	0.11 ± 0.01	L4–L5	T3, L1, T2	T3, T2, T10
2064644	0.31 ± 0.06	0.19 ± 0.05	L4–S1	P, T7, T11	T10, T9, T7
9664225	0.08 ± 0.04	0.44 ± 0.13	T6–T11	T12, L4, T11	T3, T4, T12
1084868	0.03 ± 0.00	0.08 ± 0.01	T6–T10	T8, T9, P	T9, T10, T8
8232865	0.00 ± 0.00	0.37 ± 0.02	T1–L1	T4, T3, T5	T3, T4, T2

Using the mean probability values of the measurements of each subject, 27 and 7 subjects were correctly and falsely classified, respectively, out of the 34 subjects with spinal fusion. For 8 of the 27 correctly classified subjects, the class probability difference was below 0.2. Using the RF approach, 9 and 25 subjects were falsely and correctly classified, respectively. For the majority of the misclassified subjects, the results for the OCSVM and the binary classification approach were congruent.

The point plots in Figure 1 showed, in general, the presence of highly overlapping areas between the classes. Subject 3459598 was correctly classified for both OCSVM and RF with a high probability (probability for classification as an outlier—OCSVM: 1.00; RF: 0.93). For subject 8232865, pathologic differences were clearly visible in the point plot.

However, both approaches failed to classify the subject correctly (probability for classification as an outlier—OCSVM: 0.00; RF: 0.37). Looking at the LIME interpretations for both classifiers showed similar features in the top 10 listings. For the correctly classified subject, the LIME values of the top 10 features showed an effect that indicated an outlier. For the misclassified subject, the majority of LIME values for the top 10 features showed an effect toward the class of healthy subjects.

4. Discussion

For the discussion of validity/plausibility of the found effects, as well as the proposed pathology-independent classifier, five aspects were addressed in the following: (a) classification performance of the pathology-independent and binary classification approaches, (b) previous research results and biomechanical characteristics in relation to classification performance, (c) LIME effects between the pathology-independent and binary classification approaches, (d) LIME effects in relation to the location of the spinal fusion, and (e) expert knowledge-based interpretation of the exemplary subject results.

(a) Looking at the real data, the best performance was achieved using the pathology-independent classifier for the spinal fusion data ($MCC = 0.57 \pm 0.23$). These results also matched the use of the binary classifier using the RF algorithm trained on both classes ($MCC = 0.45 \pm 0.25$). OCSVM performed better using the spinal fusion data. Therefore, the current study indicated prevalent spinal differences between healthy subjects and subjects after spinal fusion that were suitable for classification. The classification performance of the subjects with osteoarthritis might have indicated potential spinal differences with a limited ability for classification. For the osteoarthritis data, the predicted and actual classes were only weakly correlated when using the stance data (OCSVM: $MCC = 0.21 \pm 0.12$; RF: 0.19 ± 0.21). The results indicated that subjects with osteoarthritis seemed to stand differently compared to healthy subjects. However, the differences seemed relatively small, which resulted in a low discriminative power of the features. An interpretation of the results should, therefore, be carried out with caution. For the back pain data, no good model solution was found for either approach (OCSVM: $MCC = 0.13 \pm 0.19$; RF: 0.08 ± 0.34).

The BSS as a probability scoring metric also quantified the poor distinguishability of subjects with back pain from healthy subjects. Due to the poor results for both the pathology-independent, as well as the binary classifier, it can be ruled out that the proposed approach alone was the cause of the poor results. This was in line with a previous work that demonstrated that even the use of different complex classifiers and metric learning approaches is not able to lead to a significant improvement in the classification performance for subjects with back pain [46].

For the synthetic data, the binary classification approach seemed to show superior performance when the class separation was reduced. This may highlight the importance of highly discriminant data for optimal performance of the pathology-independent approach. Surprisingly, the OCSVM approach showed a lower performance on S1 compared to S2, which was contrary to this statement. It cannot be fully clarified at this point whether this was a random event and, consequently, if further research is necessary.

There were diverging indications that, on the one hand, suggested no clear, systematic performance difference and, on the other hand, that binary classifiers performed better when the class separation was low. This study was intended as an initial investigation, but in order to clarify the mentioned aspects, further research is necessary. However, overall, the results seemed relatively congruent for the OCSVM, as well as for the binary RF approach.

(b) It can be assumed that non-specific back pain and osteoarthritis affect the spine more in terms of dynamic function than in terms of posture. Therefore, these pathologies are less detectable in static measurements. Back pain often originates in the muscular system [47]. Due to an altered movement sequence, other muscle activities, angular courses, and greater joint moments are conceivable and probable [7]. Muscular imbalances are also

easier to recognize in dynamic function. Osteoarthritis patients, on the other hand, have postural changes due to contractures in the affected joint, which should also be recognizable in static measurements of the cranially located vertebrae [48,49], which would explain why their detection was more successful. However, the entire spine was demonstrated to compensate for the flexed hip joint in the sagittal plane, but since this is limited, a strongly unbalanced spine–pelvis alignment was found [48]. This compensatory phenomenon should actually be identifiable by the proposed methodology. Nevertheless, since our individual vertebral body positions did not necessarily reflect the spinal position as a whole, our classification approach might appear as not sensitive enough. In spondylodesis patients, the underlying pathologies were directly in the measurement area. As a result, the vertebral bodies to be measured usually no longer contribute to a physiological position, since the affected segments were fixed. After spondylodesis surgery, structural alterations such as a slight shortening of the spine have been found [50], but only after one year, and none of them in the frontal plane. The authors hypothesized that these measurable changes were due to other structural changes, such as muscle shortening and deformed vertebrae. Years of pain with functional alterations of movement patterns and the resulting pathological posture, thus, appeared to have a major lasting impact on anatomical structure [51]. In addition, it can be assumed that neighboring segments compensated for the stiffened area, which is why stronger vertebral body rotations could be recognizable here. Existing surgical scars, as well as spinal hardware such as rods and screws, could also have an influence on the measurement results. Besides the influence of different pathologies on posture, advancing age could have also negatively affected the postural control [52].

(c) The following discussion of the LIME effects focuses on the results for the spinal fusion data due to the best classification performance. The reason is that, on the one hand, local LIME effects may indicate characteristics of a healthy or pathologic subject for the respective vertebrae and, on the other hand, it is also possible that the model learned a wrong relationship between the feature value and the class membership. For the identification of subject-specific characteristics, it is, therefore, important to reduce model mistakes.

Looking at the LIME values of the OCSVM and the binary RF approaches with the highest effects, showed that partly congruent results were present for the features with the highest effects for both approaches. However, there were also diverging results; for example, the pelvis more often had a higher relevance for the OCSVM. In general, only a little research addresses the agreement of XAI results for different classifiers, different XAI approaches [32], and different extracted features [31] in the context of biomechanical data and, consequently, more research is necessary. The exemplary LIME results for each of the presented subjects often showed effects for both classes (healthy and outlier). Therefore, the present study's patients also showed spinal characteristics, which seemed to be similar to those of the healthy subjects.

Regarding the partly different/instable XAI results for both approaches, the aggregation of the results of different XAI approaches and different models on the same data might be an interesting approach to increase the robustness of XAI interpretations. In the context of feature selection, the ensemble method has already proven useful in generating more robust results [53,54]. The use of similar methods in the context of XAI could, thus, be useful and could help to increase the stability and, consequently, the trust in the XAI interpretations. The inclusion of global interpretations, e.g., through permutation importance or partial dependence plots [55], may be complementary and add more insights into the data. For a practical use in clinical settings, the appropriateness of an exclusive focus on LIME appears questionable. Therefore, the combination of different approaches seems necessary.

Looking at the LIME effects, the boxplots showed little variation of the LIME values between the 10 measurements of each subject most of the time. Therefore, in general, the same effects were identified for the measurements of each subject, which also speaks for

the quality of the interpretation. These results can also be justified by the fact that previous works showed a high accordance between multiple static and dynamic measurements of the same subjects [38].

(d) The LIME values did not always show high effects for the vertebrae related to the direct location of the spinal fusion. A possible reason for this might be that spinal fusion causes a higher pathologic deviation in other regions of the spine. This could also be explained by the connection instability that often develops after surgery [56,57]. However, more distant vertebral bodies also showed large effects, and they are usually in the regions of reverse curvature. They appeared to deviate from the physiological positions in the sense of maintaining balance to compensate for the altered statics of the fused vertebral bodies.

(e) Looking at the exemplarily results for the correctly classified subject 3459598 showed that the LIME values with the highest effect (OCSVM: Flexion extensions T4, T5, and L1; RF: Flexion extensions T2, T3, and T4) indicated an effect toward an outlier. The effects were similar for both approaches. In line with the above-mentioned aspects, the features with the highest effects did not directly map the vertebrae where the spinal fusion was located (T10–L2). For the misclassified subject 8232865, the LIME values (OCSVM: Flexion extensions T3, T4, and T5; RF: Flexion extensions T2, T3, and T4; location of the spinal fusion: T1–L1) indicated an effect for the class of healthy subjects. However, the subject belonged to the class of spinal fusion. The misclassified subject showed very different patterns compared to the other subjects of the spinal fusion group. In addition to the still obvious vertebral body position data deviating from physiological posture, the very long fusion (long-axis fusions with more than four vertebral bodies are more common in scoliosis patients) also suggested that subject 8232865 suffered from severe scoliosis. In general, the mean values for the flexion extensions T3, T4, and T5 of the group of spinal fusion were above the mean values for the healthy. For the misclassified sample, this was the other way around. This might be the reason why the binary classifier failed to correctly classify the subject, because it learned an oversimplified relationship of the feature values to the class membership. The training data for the respective pathology should, therefore, be expanded, especially for subjects that show similar characteristics to the misclassified subject, so that the model is able to map the respective characteristics during the training phase. Further feature engineering through an automated feature extraction [37] or the inclusion of global spinal parameters (e.g., lordosis and kyphosis angle) might also be promising. For the OCSVM, a further possible reason for misclassification was that hyperparameters were not optimally chosen due to the hyperparameter search, because it was only based on the validation set performance related to the healthy subjects. Hyperparameter searching using a scoring metric that captures the classification results of healthy and patient data might be a promising alternative approach for improving performance.

Overall, the results highlighted the usefulness of the proposed XAI approach for explaining the predictions of the pathology-independent classifier. With the proposed XAI approach, it was possible to understand why subjects were classified (including why they were misclassified) and to reduce the black box character of the machine learning model. Therefore, the current study formed an important step for making OCSVM classifiers more applicable in clinical contexts.

Common inference-based statistical analysis methods often aim to find global effects between different groups of subjects. However, global interpretations are misleading in the case of highly individual subject characteristics, which are potentially relevant for class memberships. For spinal data, a previous work showed that highly individual patterns are present, which even enables the recognition of subjects [38]. For example, for the present groups of pathologies, the location of the spinal fusion was highly individual. Therefore, local instead of global interpretations gained high relevance. This highlights the clinical importance of the use of XAI tools, such as in the present study, for obtaining local interpretations.

Overall, the present results showed a high potential for the proposed pathology-independent classifier and no clear superiority of commonly used binary classification approaches. Limitations are to be mentioned in connection with the validity of the measurements while standing. Although a meta-analysis was able to confirm an overall reliable and valid measurement method for the assessment of spinal posture, this was particularly the case for global parameters such as thoracic kyphosis or lumbar lordosis. Pelvic parameters, such as obliquity or torsion, were less reliable and showed higher reliability in scoliosis patients than in healthy individuals, but one reason for this may be the smaller sample size [58]. However, further research has also shown that similar vertebral body deviations in the transverse plane of healthy persons could be measured with the help of rasterstereography [39], as another research group was able to perform with the help of CT and MRI images [59,60].

The sample of healthy subjects for training the pathology-independent classifier was relatively small. The discovered differences could, therefore, also be due to the sample and not due to actual differences in the subjects with pathologies. An expansion of the sample is, therefore, necessary for future studies. There were limitations regarding the sample of subjects and the large age difference between the groups. At this point, it cannot be excluded that the corresponding effects influenced the classification task. In following studies, the analysis should be repeated with matched groups of subjects. Further, the present study evaluated the posture while standing. Other positions (e.g., sitting), as demonstrated in [61], may influence spinal parameters and should also be considered in future works.

Interacting features might influence the dependency between the feature and, thus, the LIME values. Hence, possible interactions should be considered and analyzed in future works. For different contexts, different interpretation levels can be provided through selecting a maximum number of features to be displayed by LIME. However, there is no ground truth for the evaluation of the interpretation results. It is, therefore, difficult to evaluate if they were meaningful and appropriately mapped posture characteristics. Furthermore, according to the current state of research, there were no objective criteria to evaluate interpretability [55]. For the evaluation of the interpretation results, task performance might be a useful approach [62]. A comparison of clinical expert-based decisions with the XAI results might also be relevant for the additional evaluation of automated XAI interpretations.

The present study used static spinal data for the application of the methods. However, in the context of biologic gender classification, for example, dynamic data have proven superior for classification compared to the use of static data [37]. Furthermore, there are biomechanical reasons why dynamic data might better map pathologic differences [63,64]. Previous studies have shown that chronic back pain patients, in addition to altered thoracic–pelvic or lumbar–pelvic coordination in the transverse and frontal planes, also had different muscular control of the back muscles than healthy subjects [63,64]. In the transverse plane, coordination was less variable and more rigid, and the pelvis, lumbar spine, and thorax moved in phase; meanwhile, in the frontal plane, all three body parts showed looser and more variable coordination, especially at higher walking speeds, while rotational amplitudes were not affected [64]. The ipsilateral back muscles (on the side of the pain) showed increased muscle activity during the actual rather inactive swing leg phase and decreased activation during the double-supported stance phase [63,64]. A dynamic adaptation in gait to existing muscular or capsular deficits can also be assumed in osteoarthritis patients. These altered biomechanical parameters probably only become apparent through an asymmetrical and less-coordinated movement sequence. This is why the application of dynamic spinal data seems to have the potential to improve classification performance and should be evaluated in future works. In addition, studies have reported improved classification performance for feature extraction or mapping objects into an embedded space through learning a representation function (metric learning) [65].

Consequently, feature extraction and metric learning approaches should also be considered for future works. Regarding the used outlier detection algorithm, various other unsupervised approaches (e.g., isolation forest algorithm [40]), as well as semi-supervised outlier detection approaches [66], are proposed in literature. Regarding the used XAI algorithm for explaining the predictions, various other approaches have been reported (e.g., SHAP [34] and DeepLIFT [35]). As a next step, the comparison of the proposed approach with other algorithms and other XAI tools seems to be important.

As a possible field of application, except of the use in the clinical analysis of spinal data, the proposed approach might also be interesting in the field of ergonomics. Trunk posture classifiers are an emerging application for estimating spinal loads during manual lifting tasks that may lead to low back pain [67]. Many of the most-used and simple biomechanical models applied for estimating spinal loads utilize only a few parameters related to the trunk posture for estimating spinal loads [68]. The present study's pathology-independent approach could possibly be expanded to automatically identify incorrect positions in real time by only learning the patterns of correct ergonomic positions.

5. Conclusions

The results suggested the potential suitability of the proposed pathology-independent approach. In the present study, no clear superiority of the commonly used binary classifiers compared to the proposed approach could be demonstrated. Static data did not seem to be optimally discriminant for mapping differences between the currently studied patient groups (especially for the subjects with back pain) and the healthy subjects. As a next step, spinal movement data should be used for classification to check if the dynamic data better map group differences and lead to increased classification performance. Moreover, metric learning approaches should be evaluated.

The proposed pathology-independent data-driven approach could be helpful for providing clinicians and therapists an objective orientation and to individually adapt and monitor therapy measures pre- and post-operatively. Overall, the approach might be beneficial for finding and addressing individual spinal characteristics. In the context of personalized medicine, the most relevant characteristics for each subject classification might be useful as an objective orientation and for an individual adaptation of therapy measures.

Author Contributions: Conceptualization, C.D., B.T., J.K. (Jürgen Konradi), G.B. and C.W.; methodology, C.D. and B.T.; software, C.D.; validation, C.D., J.K. (Jürgen Konradi), J.H. and C.W.; formal analysis, C.W. and J.K. (Jürgen Konradi); investigation, C.W., J.H., F.W. and J.K. (Johanna Kniepert); resources, M.F. and P.D.; data curation, C.W., J.H., F.W. and J.K. (Johanna Kniepert); writing—original draft preparation, C.D., C.W., E.B. and J.K. (Jürgen Konradi); writing—review and editing, C.D., G.B., B.T., E.B., M.F., F.W. and J.K. (Jürgen Konradi); visualization, C.D.; supervision, M.F. and P.D.; project administration, U.B., P.D. and M.F.; funding acquisition, P.D., U.B. and M.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the BMBF under grant numbers 16SV7115 and 03IHS075B and from funds of the intramural research funding for third-party funded projects of the University Medicine of the Johannes Gutenberg University Mainz.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the ethics committee of the medical chamber Rhineland-Palatinate and registered with the WHO (INT: DRKS00014325, DRKS00013145, DRKS00013145, DRKS00017240).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are not publicly available due to restrictions on the use of clinical patient data.

Acknowledgments: Foremost, the authors would like to thank all of the participants of this study. Their colleagues also acknowledge for their support in the recruitment process. Finally, the authors

want to express their gratitude to Kjell Heitmann, Amira Basic, as well as Helmut Diers for multi-farious endorsements and technological support during this project.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

Table A1. Variables used for modeling. Variables were measured in sagittal, frontal, and transversal plane. The descriptions were adapted from the DIERS manual.

Variable	Plane	Description
	Sagittal: Vertebral sagittal Flexion and Extension (°)	The parameter describes the inclination of the calculated vertebra in space (relative to a plumb/gravity line) as seen from a left view. The angle (in degrees) is calculated from the projection of the vertebra in a sagittal plane (rotation and lateral flexion are ignored). A positive value means a forward tilt of the vertebra (flexion).
VP/C7 (Vertebra prominens/7. cervical vertebral body) T1–T12 (Thoracic spine) L1–L4 (Lumbar spine)	Frontal: Vertebral Lateral Flexion (°)	The parameter describes the lateral inclination of the vertebra in space (relative to a plumb/gravity line) as seen from a posterior–anterior view. The angle (in degrees) is calculated from the projection of the vertebra in the coronal plane (rotation and sagittal extension/flexion are ignored). A positive value means a tilt of the vertebra to the left (lateral flexion left).
	Transversal: Vertebral Rotation (°)	The vertebral rotation describes the rotation of a vertebra in the transversal plane (relative to the neutral pelvis). A positive value means a vertebra is rotated to the left (counterclockwise) when seen from behind. The rotation of vertebral bodies happens in situ and, therefore, the direction of rotation between the surface and vertebral rotation changes. Hence, a surface rotation to the right, mathematically represented with a +, becomes a vertebral body rotation to the left. This is due to the calculation process in which a vector is used that points from the Processus spinosus towards the

		middle of the vertebral body, meaning that the surface rotation changes its direction within the vertebral body.
	Pelvic Obliquity (°)	A line is drawn from DL to DR (left and right dimple) and is compared to a horizontal line representing the horizon. The angle (in degrees) between them is measured. A positive value means that the right pelvis is elevated.
	Pelvic Torsion (dimples) (°)	The parameter describes the torsion of the surface normals on the two lumbar dimples.
	Pelvic Inclination (dimples) (°)	The parameter describes the mean vertical torsion of the two surface normals on DL and DR.
Pelvis	Pelvic Inclination (symmetry line) (°)	The parameter describes the angle of the vertical components of the surface normals on point DM (dimple midpoint) based on the horizontal.
	Pelvic Rotation (°)	The pelvic rotation is the rotation in the transversal plane of the right dimple relative to a reference coronal plane that is defined from the system setup, perpendicular to the camera-projection axis. A positive value means the pelvis is rotated to the left when seen from behind (the value is corrected * (-1)).

References

1. Robert Koch Institut. *Gesundheit in Deutschland. Gesundheitsberichterstattung des Bundes Gemeinsam Getragen von RKI und Destatis*; Robert Koch Institut: Berlin, Germany, 2015.
2. Hartvigsen, J.; Hancock, M.; Kongsted, A.; Louw, Q.; Ferreira, M.L.; Genevay, S.; Hoy, D.; Karppinen, J.; Pransky, G.; Sieper, J.; et al. What low back pain is and why we need to pay attention. *Lancet* **2018**, *391*, 2356–2367, doi:10.1016/s0140-6736(18)30480-x.
3. Schmidt, C.O.; Raspe, H.; Pflugsten, M.; Hasenbring, M.I.; Basler, H.D.; Eich, W.; Kohlmann, T. Back Pain in the German Adult Population. *Spine* **2007**, *32*, 2005–2011, doi:10.1097/brs.0b013e318133fad8.
4. von der Lippe, E.; Krause, L.; Prost, M.; Wengler, A.; Leddin, J.; Müller, A.; Zeisler, M.-L.; Anton, A.; Rommel, A.; BURDEN Study Group. Prävalenz von Rücken- und Nackenschmerzen in Deutschland. Ergebnisse der Krankheitslast-Studie BURDEN 2020. *J. Health Monit.* **2021**, *6*, 1–14, doi:10.25646/7854.
5. Urits, I.; Burshtein, A.; Sharma, M.; Testa, L.; Gold, P.A.; Orhurhu, V.; Viswanath, O.; Jones, M.; Sidransky, M.A.; Spektor, B.; et al. Low Back Pain, a Comprehensive Review: Pathophysiology, Diagnosis, and Treatment. *Curr. Pain Headache Rep.* **2019**, *23*, 23, doi:10.1007/s11916-019-0757-1.
6. Casser, H.-R.; Seddigh, S.; Rauschmann, M. Acute Lumbar Back Pain: Investigation, Differential Diagnosis, and Treatment. *Dtsch. Arzteblatt Online* **2016**, *113*, 223–234, doi:10.3238/ARZTEBL.2016.0223.
7. Koes, B.W.; van Tulder, M.; Thomas, S. Diagnosis and treatment of low back pain. *BMJ* **2006**, *332*, 1430–1434, doi:10.1136/bmj.332.7555.1430.
8. Götz-Neumann, K. *Gehen Verstehen. Ganganalyse in der Physiotherapie*; Thieme: Stuttgart, Germany, 2016. ISBN 9783131323743.
9. Deyo, R.A. Fusion surgery for lumbar degenerative disc disease: Still more questions than answers. *Spine J.* **2015**, *15*, 272–274, doi:10.1016/j.spinee.2014.11.004.
10. Rajaei, S.S.; Bae, H.W.; Kanim, L.E.; Delamarter, R.B. Spinal Fusion in the United States. *Spine* **2012**, *37*, 67–76, doi:10.1097/brs.0b013e31820cccfb.

11. Chan, C.-W.; Peng, P. Failed Back Surgery Syndrome. *Pain Med.* **2011**, *12*, 577–606, doi:10.1111/j.1526-4637.2011.01089.x.
12. Brox, J.I.; Reikerås, O.; Nygaard, Ø.; Sørensen, R.; Indahl, A.; Holm, I.; Keller, A.; Ingebrigtsen, T.; Grundnes, O.; Lange, J.E.; et al. Lumbar instrumented fusion compared with cognitive intervention and exercises in patients with chronic back pain after previous surgery for disc herniation: A prospective randomized controlled study. *Pain* **2006**, *122*, 145–155, doi:10.1016/j.pain.2006.01.027.
13. Papi, E.; Koh, W.S.; McGregor, A.H. Wearable technology for spine movement assessment: A systematic review. *J. Biomech.* **2017**, *64*, 186–197, doi:10.1016/j.jbiomech.2017.09.037.
14. Rast, F.M.; Labruyère, R. Systematic review on the application of wearable inertial sensors to quantify everyday life motor activity in people with mobility impairments. *J. Neuroeng. Rehabil.* **2020**, *17*, 1–19, doi:10.1186/s12984-020-00779-y.
15. Lau, H.-Y.; Tong, K.-Y.; Zhu, H. Support vector machine for classification of walking conditions of persons after stroke with dropped foot. *Hum. Mov. Sci.* **2009**, *28*, 504–514, doi:10.1016/j.humov.2008.12.003.
16. Wahid, F.; Begg, R.K.; Hass, C.J.; Halgamuge, S.; Ackland, D. Classification of Parkinson’s Disease Gait Using Spatial-Temporal Gait Features. *IEEE J. Biomed. Heal. Inform.* **2015**, *19*, 1794–1802, doi:10.1109/jbhi.2015.2450232.
17. Dindorf, C.; Teufl, W.; Taetz, B.; Becker, S.; Bleser, G.; Fröhlich, U. Feature extraction and gait classification in hip replacement patients on the basis of kinematic waveform data. *Biomed. Hum. Kinet.* **2021**, *13*, 177–186, doi:10.2478/bhk-2021-0022.
18. Horst, F.; Lapuschkin, S.; Samek, W.; Müller, K.-R.; Schöllhorn, W.I. Explaining the unique nature of individual gait patterns with deep learning. *Sci. Rep.* **2019**, *9*, 1–13, doi:10.1038/s41598-019-38748-8.
19. Phinyomark, A.; Petri, G.; Ibáñez-Marcelo, E.; Osis, S.T.; Ferber, R. Analysis of Big Data in Gait Biomechanics: Current Trends and Future Directions. *J. Med. Biol. Eng.* **2018**, *38*, 244–260, doi:10.1007/s40846-017-0297-2.
20. Halilaj, E.; Rajagopal, A.; Fiterau, M.; Hicks, J.L.; Hastie, T.J.; Delp, S.L. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. *J. Biomech.* **2018**, *81*, 1–11, doi:10.1016/j.jbiomech.2018.09.009.
21. Bzdok, D.; Altman, N.; Krzywinski, M. Statistics versus machine learning. *Nat. Methods* **2018**, *15*, 233–234, doi:10.1038/nmeth.4642.
22. Arnaut, R.; Curran, L.; Zhao, Y.; Levine, J.C.; Chinn, E.; Moon-Grady, A.J. An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. *Nat. Med.* **2021**, *27*, 882–891, doi:10.1038/s41591-021-01342-5.
23. Hu, L.; Bell, D.; Antani, S.; Xue, Z.; Yu, K.; Horning, M.P.; Gachuhi, N.; Wilson, B.; Jaiswal, M.S.; Befano, B.; et al. An Observational Study of Deep Learning and Automated Evaluation of Cervical Images for Cancer Screening. *J. Natl. Cancer Inst.* **2019**, *111*, 923–932, doi:10.1093/jnci/djy225.
24. Luo, H.; Xu, G.; Li, C.; He, L.; Luo, L.; Wang, Z.-X.; Jing, B.; Deng, Y.; Jin, Y.; Li, Y.; et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: A multicentre, case-control, diagnostic study. *Lancet Oncol.* **2019**, *20*, 1645–1654, doi:10.1016/s1470-2045(19)30637-0.
25. Laroche, D.; Tolamiya, A.; Morisset, C.; Maillefert, J.; French, R.; Ornetti, P.; Thomas, E. A classification study of kinematic gait trajectories in hip osteoarthritis. *Comput. Biol. Med.* **2014**, *55*, 42–48, doi:10.1016/j.compbiomed.2014.09.012.
26. Teufl, W.; Taetz, B.; Miezal, M.; Lorenz, M.; Pietschmann, J.; Jöllenbeck, T.; Fröhlich, M.; Bleser, G. Towards an Inertial Sensor-Based Wearable Feedback System for Patients after Total Hip Arthroplasty: Validity and Applicability for Gait Classification with Gait Kinematics-Based Features. *Sensors* **2019**, *19*, 5006, doi:10.3390/s19225006.
27. Adadi, A.; Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160, doi:10.1109/access.2018.2870052.
28. Samek, W.; Müller, K.-R. Towards Explainable Artificial Intelligence. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, 1st ed.; Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., Müller, K.-R., Eds.; Springer: Cham, Switzerland, 2019; pp. 5–22.
29. European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Off. J. Eur. Union* **2016**, *L 119*, 1–88.
30. Holzinger, A.; Biemann, C.; Pattichis, C.S.; Kell, D.B. What Do We Need to Build Explainable AI Systems for The Medical Domain? Available online: <http://arxiv.org/pdf/1712.09923v1> (accessed on 20 February 2020).
31. Dindorf, C.; Teufl, W.; Taetz, B.; Bleser, G.; Fröhlich, M. Interpretability of Input Representations for Gait Classification in Patients after Total Hip Arthroplasty. *Sensors* **2020**, *20*, 4385, doi:10.3390/s20164385.
32. Horst, F.; Slijepcevic, D.; Lapuschkin, S.; Raberger, A.-M.; Zeppelzauer, M.; Samek, W.; Breiteneder, C.; Schöllhorn, W.I.; Horsak, B. On the Understanding and Interpretation of Machine Learning Predictions in Clinical Gait Analysis Using Explainable Artificial Intelligence. Available online: <http://arxiv.org/pdf/1912a.07737v1> (accessed on 10 March 2020).
33. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’16), San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.
34. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
35. Shrikumar, A.; Greenside, P.; Kundaje, A. Learning important features through propagating activation differences. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 3145–3153.

36. Teufl, W.; Taetz, B.; Miezal, M.; Dindorf, C.; Fröhlich, M.; Trinler, U.; Hogam, A.; Bleser, G. Automated detection of pathological gait patterns using a one-class support vector machine trained on discrete parameters of IMU based gait data. *Clin. Biomech.* **2021**, *89*, 105452.
37. Dindorf, C.; Konradi, J.; Wolf, C.; Taetz, B.; Bleser, G.; Huthwelker, J.; Drees, P.; Fröhlich, M.; Betz, U. General method for automated feature extraction and selection and its application for gender classification and biomechanical knowledge discovery of sex differences in spinal posture during stance and gait. *Comput. Methods Biomech. Biomed. Eng.* **2021**, *24*, 299–307, doi:10.1080/10255842.2020.1828375.
38. Dindorf, C.; Konradi, J.; Wolf, C.; Taetz, B.; Bleser, G.; Huthwelker, J.; Werthmann, F.; Drees, P.; Fröhlich, M.; Betz, U. Machine learning techniques demonstrating individual movement patterns of the vertebral column: The fingerprint of spinal motion. *Comput. Methods Biomech. Biomed. Eng.* **2021**, *24*, doi: 10.1080/10255842.2021.1981884.
39. Wolf, C.; Betz, U.; Huthwelker, J.; Konradi, J.; Westphal, R.; Cerpa, M.; Lenke, L.; Drees, P. Evaluation of 3D Vertebral and Pelvic Position by Surface Topography in Asymptomatic Females: Presentation of Normative Reference Data. Available online: <https://www.researchsquare.com/article/rs-612767/v1> (accessed on 18 September 2021).
40. Liu, F.T.; Ting, K.M.; Zhou, Z.-H. Isolation-Based Anomaly Detection. *ACM Trans. Knowl. Discov. Data (TKDD)* **2012**, *6*, 1–39, doi:10.1145/2133360.2133363.
41. Dreiseitl, S.; Osl, M.; Scheibböck, C.; Binder, M. Outlier Detection with One-Class SVMs: An Application to Melanoma Prognosis. *AMIA Annu. Symp. Proc.* **2010**, 172–176.
42. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
43. Platt, J.C. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In *Advances in Large Margin Classifiers*; MIT Press: Cambridge, MA, USA, 1999; pp. 61–74.
44. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32.
45. Ferro, C.A.T.; Fricker, T.E. A bias-corrected decomposition of the Brier score. *Q. J. R. Meteorol. Soc.* **2012**, *138*, 1954–1960, doi:10.1002/qj.1924.
46. Dindorf, C.; Konradi, J.; Wolf, C.; Taetz, B.; Bleser, G.; Kniepert, J.; Drees, P.; Fröhlich, M.; Betz, U. Towards a better understanding of spinal differences between healthy subjects and subjects with back pain using Explainable Artificial Intelligence (XAI). In Proceedings of the 9th International Performance Analysis Workshop and Conference & 5th IACSS Conference, Vienna, Austria, 30–31 August 2021; pp. 1–3.
47. Ract, I.; Meadeb, J.-M.; Mercy, G.; Cueff, F.; Husson, J.-L.; Guillin, R. A review of the value of MRI signs in low back pain. *Diagn. Interv. Imaging* **2015**, *96*, 239–249, doi:10.1016/j.diii.2014.02.019.
48. Weng, W.-J.; Wang, W.-J.; Wu, M.-D.; Xu, Z.-H.; Xu, L.-L.; Qiu, Y. Characteristics of sagittal spine–pelvis–leg alignment in patients with severe hip osteoarthritis. *Eur. Spine J.* **2014**, *24*, 1228–1236, doi:10.1007/s00586-014-3700-5.
49. Kechagias VA, Grivas TB, Papagelopoulos PJ, Kontogeorgakos VA, Vlasis K. Truncal Changes in Patients Suffering Severe Hip or Knee Osteoarthritis: A Surface Topography Study. *Clin Orthop Surg.* **2021**, *13*, 185–95, doi:10.4055/cios20123.
50. Scheidt, S.; Endres, S.; Gesicki, M.; Hofmann, U.K. Using video rasterstereography and treadmill gait analysis as a tool for evaluating postoperative outcome after lumbar spinal fusion. *Gait Posture* **2018**, *64*, 18–24, doi:10.1016/j.gaitpost.2018.05.019.
51. Hackenberg, L.; Hierholzer, E.; Pözl, W.; Götz, C.; Liljenqvist, U. Rasterstereographic back shape analysis in idiopathic scoliosis after anterior correction and fusion. *Clin. Biomech.* **2003**, *18*, 1–8, doi:10.1016/s0268-0033(02)00165-1.
52. Leirós-Rodríguez, R.; Arce-Fariña, M.E.; Álvarez, C.M.M.; Garcia-Soidan, J.L. Definition of the proper placement point for balance assessment with accelerometers in older women. *Rev. Andal. Med. Deport.* **2016**, *9*, 1–6, doi:10.1016/j.ram.2016.09.001.
53. Ben Brahim, A.; Limam, M. Ensemble feature selection for high dimensional data: A new method and a comparative study. *Adv. Data Anal. Classif.* **2018**, *12*, 937–952, doi:10.1007/s11634-017-0285-y.
54. Shahrjooihaghighi, A.; Frigui, H.; Zhang, X.; Wei, X.; Shi, B.; Trabelsi, A. An ensemble feature selection method for biomarker discovery. In Proceedings of the 2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT); Bilbao, Spain, 18–20 December 2017; volume 2017, pp. 416–421.
55. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*; Leanpub: Victoria, British Columbia, Canada, 2018.
56. Schlegel, J.D.; Smith, J.A.; Schlausener, R.L. Lumbar motion segment pathology adjacent to thoracolumbar, lumbar, and lumbosacral fusions. *Spine* **1996**, *21*, 970–981, doi:10.1097/00007632-199604150-00013.
57. Bredow, J.; Löhner, L.; Oppermann, J.; Scheyerer, M.J.; Sobottke, R.; Eysel, P.; Siewe, J. Pathoanatomic Risk Factors for Instability and Adjacent Segment Disease in Lumbar Spine: How to Use Topping Off? *Biomed. Res. Int.* **2017**, *2017*, 2964529, doi:10.1155/2017/2964529.
58. Krott, N.L.; Wild, M.; Betsch, M. Meta-analysis of the validity and reliability of rasterstereographic measurements of spinal posture. *Eur. Spine J.* **2020**, *29*, 2392–2401, doi:10.1007/s00586-020-06402-x.
59. Janssen, M.M.A.; Vincken, K.L.; Kemp, B.; Obradov, M.; Kleuver, M. de; Viergever, M.A.; Castelein, R.M.; Bartels, L.W. Pre-existent vertebral rotation in the human spine is influenced by body position. *Eur. Spine J.* **2010**, *19*, 1728–1734, doi:10.1007/s00586-010-1400-3.
60. Kouwenhoven, J.-W.M.; Vincken, K.L.; Bartels, L.W.; Castelein, R.M. Analysis of preexistent vertebral rotation in the normal spine. *Spine* **2006**, *31*, 1467–1472, doi:10.1097/01.brs.0000219938.14686.b3.

61. Chevillotte, T.; Coudert, P.; Cawley, D.; Bouloussa, H.; Mazas, S.; Boissière, L.; Gille, O. Influence of posture on relationships between pelvic parameters and lumbar lordosis: Comparison of the standing, seated, and supine positions. A preliminary study. *Orthop. Traumatol. Surg. Res.* **2018**, *104*, 565–568, doi:10.1016/j.otsr.2018.06.005.
62. Gunning, D.; Stefik, M.; Choi, J.; Miller, T.; Stumpf, S.; Yang, G.-Z. XAI—Explainable artificial intelligence. *Sci. Robot.* **2019**, *4*, eaay7120, doi:10.1126/scirobotics.aay7120.
63. Arendt-Nielsen, L.; Graven-Nielsen, T.; Svarrer, H.; Svensson, P. The influence of low back pain on muscle activity and coordination during gait: A clinical and experimental study. *Pain* **1996**, *64*, 231–240, doi:10.1016/0304-3959(95)00115-8.
64. Lamothe, C.J.C.; Meijer, O.G.; Daffertshofer, A.; Wuisman, P.I.J.M.; Beek, P.J. Effects of chronic low back pain on trunk coordination and back muscle activity during walking: Changes in motor control. *Eur. Spine J.* **2006**, *15*, 23–40, doi:10.1007/s00586-004-0825-y.
65. Weinberger, K.Q.; Saul, L.K. Distance Metric Learning for Large Margin Nearest Neighbor Classification. *J. Mach. Learn. Res.* **2009**, *10*, 207–244.
66. Gao, J.; Cheng, H.; Tan, P.-N. Semi-supervised outlier detection. In Proceedings of the 2006 ACM Symposium on Applied Computing—SAC'06, Dijon, France, 27 April 2006; Haddad, H.M., Ed.; ACM Press: New York, NY, USA, 2006; p. 635, ISBN 1595931082.
67. Conforti, I.; Mileti, I.; Del Prete, Z.; Palermo, E. Measuring Biomechanical Risk in Lifting Load Tasks Through Wearable System and Machine-Learning Approach. *Sensors* **2020**, *20*, 1557, doi:10.3390/s20061557.
68. Picerno, P. An Enhanced Planar Linked Segment Model for Predicting Lumbar Spine Loads during Symmetric Lifting Tasks. *Appl. Sci.* **2020**, *10*, 6700, doi:10.3390/app10196700.

Resümee der Befunde

ZIEL war es anhand der aufgeführten Publikationen den Einsatz von ML Ansätzen im biomechanischen Kontext zu evaluieren und ausgewählte Limitationen zu adressieren. Zusammenfassend soll damit ein Beitrag geleistet werden, um die Akzeptanz und Einsatzfähigkeit von ML bei biomechanischen Daten zu erhöhen. Im Nachfolgenden werden die zentralen Ergebnisse der Publikationen diskutiert. Die hier dargestellten Punkte sind additiv zu sehen und werden die bereits in den einzelnen Publikationen durchgeführten zentralen methodischen Diskussionen ausweiten und vernetzen. Zukunftspotentiale, weiterhin bestehende Limitationen und Herausforderungen werden kumulativ herausgearbeitet. Ferner werden durch sozioinformatische Aspekte die bestehenden Überlegungen ausgeweitet. Abschließend wird ein Ausblick auf die Zukunft sowie weitere Forschungsarbeiten gegeben. Nachfolgend wird ein zusammenfassender Überblick über die Inhalte der einzelnen Publikationen dargestellt.

Publikation 1

Methodik:

- Markerrohdaten, Gelenk- und Segmentwinkel von 22 Patienten nach Hüft-TEP
- Robuste Feature Selektion unter Ausnutzung breiter Datenbasis von Wellenformdaten

Ergebnisse / Interpretation:

- Hohe Klassifikationsgenauigkeit operierte Seite (91,09 %)
- Robuste Feature Selektion auf Basis Wellenformdaten möglich

Publikation 2

Methodik:

- Alter, statische und dynamische Wirbelsäuledaten von 201 Gesunden
- Pipeline zur Ausnutzung relationaler Datenbestände & robuster Feature Selektion

Ergebnisse / Interpretation:

- Hohe Klassifikationsgenauigkeit Geschlecht (mittlerer MCC = 0,97)
- Hohe Stabilität durch kombinierte Ensemble Feature Selektion
- Resultierte Features teilweise abstrakt → interpretierbares Feature Engineering

Publikation 3

Methodik:

- IMU-Daten von 27 Gesunden, 20 Patienten nach Hüft-TEP
- Interpretation der Prediktionen mittels XAI auf Basis Wellenformdaten, automatisch extrahierten Features und einfachen deskriptiven Features

Ergebnisse / Interpretation:

- Hohe Klassifikationsgenauigkeit Gesund bzw. Patient (95,88 - 100,00 %)
- Interpretationen auf Basis der drei verwendeten Eingangsdatentypen möglich

Publikation 4

Methodik:

- dynamische Wirbelsäuledaten von 226 Gesunden
- automatische Feature Extraktion via SNN zur Subjektidentifikation

Ergebnisse / Interpretation:

- hohe Identifikationsgenauigkeit (96,00 - 100,00 %) Indiz für Einzigartigkeit Wirbelsäulendynamik
- SNN nützlich zur Reduktion intra-individueller Varianz

Vorbereitende Publikation für Publikation 5

Methodik:

- 30 diskrete Merkmale von 25 Gesunden, 20 Patienten nach Hüft-TEP, 1 Oberschenkelamputierter
- Pathologieunabhängiger Klassifikator der Wahrscheinlichkeiten und Erklärungen liefert

Ergebnisse / Interpretation:

- 3 Gesunde fehlassifiziert, 100 % Patienten korrekt klassifiziert
- Ergebnisse deuten Anschlussfähigkeit für fortgeschrittene XAI Ansätze an

Publikation 5

Methodik:

- Daten: synthetische Daten, statische Wirbelsäuledaten von Gesunden ($n = 25$) und Pathologien Rückenschmerzen ($n = 32$), Spondylodese ($n = 34$), Arthrose ($n = 60$)
- Pathologieunabhängiger Klassifikator der Wahrscheinlichkeiten und Erklärungen mittels XAI liefert
- Vergleich mit klassischem binären Klassifikator

Ergebnisse / Interpretation:

- beste Klassifikation für Spondylodese ($MCC = 0,57$), Rückenschmerzen besonders herausfordernd ($MCC = 0,13$)
- pathologieunabhängiger Klassifikator ist realisierbar
- keine klare Unterlegenheit gegenüber einem klassischen binären Klassifikator

Die Publikationen bauen inhaltlich aufeinander auf. Spezifische Forschungsfragen werden zwar durch einzelne Publikationen akzentuiert betrachtet, allerdings werden zur Beantwortung der Forschungsfragen diese in unterschiedlicher Ausprägung in mehreren Publikationen adressiert. Jede Publikation hat somit nach und nach zusätzliche Informationen und damit zu einem umfassenden Überblick über Potentiale und Herausforderungen beigetragen. Eine Übersicht über jeweilige Schwerpunkte entsprechend der Forschungsfragen ist in Tabelle 4.1 gegeben.

Tabelle 4.1: Zusammenfassung der Publikationen hinsichtlich des Ausmaßes der Adressierung zentraler Aspekte der Forschungsfragen.

Publika- tion	Ausnut- zung Daten	Globale Interp- retation	Lokale Interp- retation	Stabili- tät	Wenige Daten, Verglei- che	Unab- hängig- keit
1	+++	+++		++	+	
2	+++	+++		+++	+	
3	++	++	+++			
4	++	+	+		+++	
vor. 5			++			+++
5	+		+++			+++

Die Wahl einer geeigneten Datenrepräsentation und die Interpretierbarkeit wird dabei von allen Publikationen adressiert, wobei unterschiedliche Schwerpunkte und Ziele verfolgt werden. *Publikation 1* analysiert generell, wie objektiv, ohne Vorwissen Informationen aus Wellenformdaten generiert werden können. Stabilität bei der Feature Selektion wird in einfacher Weise adressiert und globale Interpretation der Ergebnisse durchgeführt. *Publikation 2* erweitert den Ansatz von *Publikation 1* um die Verwendung von relationalen Daten. Zudem wird der Ansatz der stabilen Feature-Selektion weiterentwickelt.

Publikation 3 erweitert die globalen Interpretationen von *Publikation 1* und *2* um lokale. Schwerpunkt wird auf die Überwindung des Black Box Charakters und das Verstehen von (Fehl-)Entscheidungen gelegt. Der Einfluss der Input-Repräsentationen auf Genauigkeit, Interpretierbarkeit und klinische Relevanz wird zudem betrachtet. Basis bilden Wellenformdaten wie in *Publikation 1* und eine ausgeweitete Betrachtung von extrahierten Features in Anlehnung an *Publikation 2*. Interpretationen werden unter Verwendung der Wellenformdaten nicht nur punktuell wie bei *Publikation 1* durchgeführt, sondern Wellenbereiche herangezogen.

Publikation 4 verwendet in Abgrenzung zu den übrigen Publikationen Ansätze des Metric Learnings, um Feature zu extrahieren und intra-individuelle Varianzen zu minimieren. Möglichkeiten zum Umgang mit wenig Daten werden erprobt und die Querschnittbetrachtung der anderen Publikationen um eine Längsschnittbetrachtung und den Vergleich von Personen erweitert.

Die vorbereitende Publikation für Publikation 5 bildet die Grundlagen für die Entwicklung eines pathologieunabhängigen Klassifikators. Erste Ansätze zur Interpretierbarkeit eines solchen Modells werden erprobt. *Publikation 5* baut auf den vorangegangenen Publikationen auf und vergleicht die pathologieunabhängige Klassifikation mit der eines konventionellen binären Klassifikators. Ansätze zur Interpretierbarkeit werden weiterentwickelt und lokale Interpretationen mittels XAI durchgeführt.

4.1 Ausnutzung vorhandener Daten und Feature Extraktion

In der Praxis wird oftmals nur auf eine geringe Anzahl an Variablen zurückgegriffen [167]. Um ein holistisches Verständnis der Biomechanik bzw. der Patientengesundheit zu erhalten ist es jedoch notwendig, eine breite Datenbasis zu betrachten [28]. Die nachfolgenden kumulativen Diskussionspunkte beschäftigen sich schwerpunktmäßig mit den evaluierten Ansätzen, um eine breite Datenbasis mit möglichst geringem Verlust an potentiell bedeutsamen Informationen nutzen zu können. Dies soll die Basis bilden, um bedeutsame Parameter darauf aufbauend auszuwählen. Für Aspekte der Stabilität bei der Auswahl bedeutsamer Parameter sei im Detail auf die nächste Sektion verwiesen. Feature Extraktion im Sinne des Meric Learnings wird später in Sektion 4.4 betrachtet.

Biomechanische Daten liegen oft als Zeitreihen vor. Für die weiteren Analyse-schritte birgt die Auswahl bestimmter Zeitpunkte sowie deskriptiver Beschreibungsmerkmale die Gefahr Informationen zu verlieren. Zudem besteht eine Abhängigkeit von Vorwissen oder subjektiven Entscheidungen. Besonders vielversprechend erscheint die direkte Verwendung von konkatenierten Wellenformdaten als Eingangsvariablen. Damit kann ein Verlust an Informationen reduziert werden, da keine Beschränkung auf einzelne Parameter der Zeitreihen erfolgen muss. [39], [58], [101], [135] konnten bereits in ihren Arbeiten vielversprechende Ergebnisse auf Basis reduzierter Datenmengen (nur Bodenreaktionskräfte) mit konkatenierten Wellenformdaten erzielen. Durch die deutlich größeren Längen der konkatenierten Wellenformdaten in den durchgeführten Studien sowie alternativen Ansätzen konnten diese ersten Ergebnisse bestätigt und ausgeweitet werden. So konnte gezeigt werden, dass diese sowohl geeignet sind um aufbauend darauf Feature zu extrahieren (Selektion und Dimensionsreduktion) (Publikation 1) als auch insgesamt mit diesen konkatenierten Wellenformdaten gute Klassifikationsergebnisse (Publikation 3 und 4) und aufbauend XAI basierte Interpretationen zu erzielen (Publikation 4).

Wellenformdaten

An dieser Stelle sei darauf hingewiesen, dass für die Verwendung von Wellenformdaten als Ausgangsbasis auch andere Ansätze interessant sein könnten. Hier sind beispielsweise Ansätze aus dem Bereich der Functional Data Analysis (FDA) zu nennen, welche zudem zu einer besseren Datenausnutzung beitragen können [168].

Ein Aspekt der durch die durchgeführten Arbeiten nicht geklärt werden kann ist die Samplingrate. In der Literatur werden häufig 101 Zeitpunkte zur Beschreibung der Zeitreihe eines normierten Gangzyklus für bspw. einen Gelenkwinkel verwendet [39], [47], [169]. Für die Publikationen wurde diese Samplingrate übernommen. Es kann an dieser Stelle nicht sicher gesagt werden, ob ggf. durch dieses Sampling Informationen verworfen wurden bzw. ob weniger Zeitpunkte auch eine adäquate Repräsentation der einzelnen Zeitreihen darstellen. Dies sollte zukünftig weiter untersucht werden.

Zudem gilt es anzuführen, dass nicht generell gesagt werden kann, dass Wellenformdaten auch für die weiteren Analyseschritte ohne Feature Extraktion (Publikation 1, 2, 3) oder Feature Learning (Publikation 4) genutzt werden sollten wie bei Publikation 3. Feature Extraktion ist wichtig um die Genauigkeit eines Modells zu verbessern, eine Überanpassung zu verhindern (insbesondere wenn das Verhältnis der Anzahl der Probanden zur Anzahl der Variablen gering ist), die Interpretier-

barkeit zu erhöhen und die erforderliche Rechenleistung zu reduzieren [84]–[86]. In Publikation 3 und 4 wurde ersichtlich, dass aufgabenspezifisch auch einfache extrahierte Merkmale gute Klassifikationsgenauigkeiten ermöglichen und zugleich die Interpretation verbessern können. Auf der anderen Seite kann Feature Extraktion bei lokaler Interpretation ggf. auch zum Verlust von bedeutsamen Informationen führen, da Feature Extraktion häufig auf Basis globaler Parameter durchgeführt wird. Daher sollte geprüft werden, ob extrahierte Feature kontextspezifisch der Verwendung der vollständigen Wellenformdaten überlegen sind. Zusammenfassend kann jedoch gesagt werden, dass die Wellenformdaten einen guten Ausgangspunkt darstellen, um ausgehend von einer breiten Datenbasis möglichst objektiv, ohne Vorwissen und mit möglichst geringem Verlust an bedeutsamen Informationen die Potentiale der Daten auszunutzen.

Häufig werden Gelenk- und Segmentwinkel für Klassifikationsaufgaben herangezogen (siehe auch Defizit 1). Gelenkwinkel sind definiert als Orientierung eines Segmentes relativ zu einem anderen. Die Berechnung dieser ist daher abhängig von einem zuvor kalibrierten biomechanischen Modell. Markerpositionsdaten sowie Markergeschwindigkeiten und -beschleunigungen, welche darauf aufbauend berechnet werden können, werden in der Literatur kaum verwendet [47]. Die Arbeit konnte jedoch aufzeigen, dass neben kinematischen Gelenk- und Segmentwinkeln auch Rohdaten (Markertrajektorien) und Beschleunigungsdaten wichtige Informationen enthalten (Publikation 1). Die Feature Selektion auf Basis der umfassenden Eingangsvariablen (Gelenk- und Segmentwinkeln, Markerpositions-, -beschleunigungs- und -geschwindigkeitsdaten) zeigt, dass eine bessere Klassenseparation durch diese ermöglicht und daher die Klassifikation verbessert wird. Es sollte sich also nicht nur, wie so oft in der Praxis, auf Gelenk- und Segmentwinkel alleine konzentriert werden, da die Integration von Rohdaten eine Möglichkeit sein kann, einen a priori Informationsverlust zu vermindern und die vorhandene Datenlage besser auszunutzen.

Rohdaten sind allerdings nicht so gut geeignet für die biomechanische bzw. klinische Interpretation, da es teilweise schwierig ist die Ergebnisse in für den Menschen verständliche und interpretierbare Bewegungsmuster / Konzepte einzuordnen. Zudem ist die Anschlussfähigkeit mit vorherigen Forschungsarbeiten erschwert, da diese sich oftmals auf Gelenk- und Segmentwinkel konzentrieren. Hier lässt sich eine klare Überlegenheit von Gelenk- und Segmentwinkeln feststellen, welche auf Basis von Rohdaten berechnet werden und diese besser interpretierbar machen. Weitere Ausführungen zur Interpretierbarkeit erfolgen in Sektion 4.3.

Die Integration von Rohdaten erscheint als Möglichkeit zusätzlich verfügbare Informationen in die Analyseschritte aufzunehmen und einen Informationsverlust zu reduzieren. Eine alleinige Verwendung dieser sollte jedoch nur für Anwendungen angestrebt werden, welche die Modelle als Black Box und alleine die Klassifikationsergebnisse verwenden. Zudem ist dies vorteilhaft, da die Verwendung dieser unabhängig von einem kalibrierten biomechanischen Modell ist und somit Analysepipelines deutlich vereinfacht werden können.

Neben der Verwendung von Rohdaten erscheint die umfassende Verwendung von relationalen Datenbeständen (z.B. Mehrfachmessungen, Messungen in unterschiedlichen Geschwindigkeiten und Modalitäten, Einbeziehung weiterer Personencharakteristika) als eine weitere Möglichkeit eine breite Datenbasis mit möglichst geringem Informationsverlust als Ausgangsdatenbasis zu nutzen. Automatische Extraktions-

Rohdaten

***Relational
Datenbestände***

verfahren auf Basis dieser Datenbestände bieten die Möglichkeit eine Vielzahl von Rechenoperationen für die Extraktion bei kontextspezifischen Aufgabenstellungen zu erproben und den Einfluss subjektiver Auswahlentscheidungen vorab zu reduzieren. Die automatische Verarbeitung wie in Publikation 2 kann jedoch auch zu schwer interpretierbaren Features führen. Eine Auswahl besonders gut interpretierbarer Operationen für die Feature Extraktion kann hier ein guter Kompromiss sein, falls eine biomechanische Interpretation angestrebt wird.

Als weiteres Problem, welches sich besonders durch die Verwendung von Wellenformdaten ergibt (und weniger stark bei anderen extrahierten Features), sind die zum Teil starken (Auto-) Korrelationen zu nennen. Neben entstehenden Redundanzen bei der Interpretation der Ergebnisse können diese zudem negativ die Leistung der Modelle beeinflussen [84]. Zur Begegnung des Problems wurden zwei Ansätze erprobt: Minimum-Redundancy Maximum-Relevance [170] (mRMR; Publikation 1) und Joint Mutual Information Maximization Algorithmus [171] (JMIM; Publikation 2). Die Ansätze haben durch Reduktion redundanter Feature bei gleichzeitiger Beibehaltung (bzw. Verbesserung) der Klassifikationsgenauigkeit zu einer optimierten Ausnutzung vorhandener Daten beigetragen.

Als weitere Möglichkeit der Verwendung einer möglichst breiten Ausgangsdatenbasis, bei gleichzeitiger Reduktion von Redundanzen und gleichzeitiger Auswertung von Zusammenhängen sowie besonders varianzaufklärenden Merkmalen, können Dimensionalitätsreduktionsverfahren wie die Hauptkomponentenanalyse eingesetzt werden. Viele Forschungsarbeiten konnten sehr vielversprechende Ergebnisse mit diesen erzielen und durch Verwendung einer breiten Ausgangsdatenbasis a priori Informationsverluste durch Vorauswahl von Merkmalen vermindern [47], [172]. Entgegen diesen Ergebnissen konnten hier nur unterlegene Ergebnisse im Vergleich zur direkten Feature Selektion auf Basis von Wellenformdaten gezeigt werden. Als mögliche Erklärung kann angeführt werden, dass die Dimensionsreduktion unüberwacht erfolgt ist und damit nur schlechter diskriminierende Feature im Vergleich zur überwachten Feature Selektion erzeugt werden konnten.

Ziel war es Methoden zu evaluieren, um die Ausnutzung vorhandener Daten zu optimieren, einen a priori Informationsverlust zu reduzieren und darauf aufbauend bedeutsame Parameter zu bestimmen. Trotzdem wurden bei den Publikationen vorab aufgrund von ökonomischen und zielsetzungsbedingten Gründen Daten verworfen, die potentiell bedeutsam sein könnten. Hierzu zählen a) Rohdaten, z.B. bei den IMU Ansätzen, b) zusätzliche anthropometrische und Gesundheitsdaten und c) noch nicht prozessierte dynamische Daten.

a) IMU-Systemmessungen liefern so auch verschiedene Parameter, die von Sensorrohdaten, d.h. 3D-Beschleunigung, 3D-Winkelgeschwindigkeit, bis hin zu biomechanisch aussagekräftigen Gelenkwinkeln und räumlich-zeitlichen Parametern reichen [137], [173]. Auch für die verwendeten IMUs aus Publikation 3 könnten daher theoretisch Rohdaten bestehend aus Gyroskop und Beschleunigung verwendet werden. Im Bereich IMUs wird sehr häufig sogar auf diese Daten zurückgegriffen und darauf aufbauend ML Berechnungen durchgeführt [174]. Der Schwerpunkt dieser Arbeit lag allerdings auf der biomechanischen Interpretierbarkeit, weshalb die Verwendung von Rohdaten gerechtfertigt wäre, wenn die Klassifikationsergebnisse bei alleiniger Verwendung der kinematischen Daten unzureichend wären, was allerdings nicht der Fall war.

b) Relationale Daten wurden bei Publikation 2 nur auf Basis der biomechanischen

**(Auto-)
Korrelationen**

**A priori
Informationsverlust**

Daten verwendet. Im Rahmen einer elektronischen Gesundheitsakte hätten hier auch weitere Daten aufgenommen werden können. Die vorgestellten Ansätze erscheinen jedoch auch anwendbar für die Suche von Korrelationen in solchen erweiterten Datenbeständen.

c) Die Verwendung von dynamischen Wirbelsäuledaten war aufgrund noch nicht vollständig abgeschlossener Prozessierung nicht in allen Publikationen möglich. Insbesondere bei Publikation 5 ist davon auszugehen, dass eine große Menge an bedeutsamen Informationen dadurch nicht beachtet wurde. Folglich sollten dynamische Daten in den folgenden Arbeiten berücksichtigt werden.

Zusammengefasst kann festgehalten werden, dass, wenn die biomechanische Interpretation der Ergebnisse angestrebt wird, zuerst eine Prüfung der Modellgüte auf Basis (bewährter) interpretierbarer Eingangsmerkmale erfolgen sollte, um die Anschlussfähigkeit zu bestehendem Wissen zu gewährleisten und dieses zu nutzen. Falls keine kontextspezifisch ausreichende Modellgüte erreicht wird, sollte eine Integration von Rohdaten evaluiert werden. Damit sollte geprüft werden, ob eine Erweiterung der interpretierbaren Feature um Feature auf Basis von Rohdaten die Modellgüte verbessert. Falls eine zufriedenstellende Verbesserung erreicht wird, sollte zusätzlich versucht werden, durch interpretierbares Feature Engineering (siehe Sektion 4.3) die bestimmten Feature unter Inklusion von Expertenwissen in besser interpretierbare Feature umzuwandeln. Die Inklusion von Rohdaten soll damit Ausgangspunkt für die Erweiterung bestehenden Wissens durch Identifikation zusätzlicher diskriminierender Variablen liefern. Bei alledem erscheint die Verwendung von Wellenformdaten als Ausgangspunkt für eine Feature Extraktion vielversprechend. Es wird damit ermöglicht relevante Zeitpunkte im Gangzyklus zu bestimmen und ohne abhängig, von wenigen deskriptiven, oft subjektiv bestimmten Parametern zur Beschreibung der Charakteristika, zu sein.

*Zwischenfa-
zit*

Highlights:

- Ausgangspunkt Wellenformdaten um Potentiale einer breiten Datenbasis objektiv, ohne Vorwissen und mit möglichst geringem a priori Informationsverlust zu nutzen
- Wellenformdaten vielversprechend für Feature Extraktion, direkte Klassifikation und aufbauende XAI basierte Interpretation
- Integration von Rohdaten als Möglichkeit a priori Informationsverlust zu vermindern und Datenlage besser auszunutzen → Ausgangspunkt Erweiterung bestehenden Wissens
- Entwickelte Pipeline zur Nutzung relationaler Daten vielversprechend zur Ausnutzung der Datenlage
- Rohdaten und automatisch extrahierte Feature teilweise abstrakt → Anschlussfähigkeit an bestehendes Wissen limitiert → interpretierbares Feature Engineering relevant
- mRMR und JMIM vielversprechend zur Reduktion korrelierter Feature

4.2 Feature Extraktion und Stabilität

Wie vorangehend diskutiert können Data Science Methoden große Datenmengen verarbeiten, ohne ausschließlich abhängig von a priori Wissen prädiktiver Variablen zu sein. Stattdessen können diese genutzt werden um zu schauen, was über das a priori Wissen hinaus noch relevant sein könnte. Feature Extraktion wurde als Art der globalen Interpretation eingesetzt zur Bestimmung besonders relevanter Parameter. Vertrauenswürdigkeit und Übertragbarkeit auf andere Datensätze sind jedoch stark eingeschränkt, wenn Stabilität bei der Feature Extraktion nicht beachtet wird, da hochgradig unterschiedliche Ergebnisse bei geringfügiger Modifikation der Datensätze entstehen können (siehe Defizit 2). Insbesondere wenn die Feature Selektion im Rahmen der Wissensentdeckung eingesetzt wird, ist der Aspekt der Stabilität besonders wichtig [43].

Durch die eingesetzten Ensemble Feature Selektion Methoden (homogene und heterogene Methode sowie Kombination aus homogener und heterogener Methode) in Publikation 1 und 2 konnte die Stabilität der ausgewählten Subsets deutlich verbessert werden. In Publikation 1 wurde homogene Ensemble Feature Selektion zur erfolgreichen Erhöhung der Stabilität eingesetzt. Aufbauend konnte in Publikation 2 ein fortgeschrittener Ansatz verwendet werden, welcher durch Kombination von homogener und heterogener Ensemble Feature Selektion unabhängig von der Auswahl einer bestimmten Ensemble Methode ist. Es konnte zudem durch die Ergebnisse von Publikation 2 gezeigt werden, dass mittels Kombination beider Methoden eine bessere Stabilität erreicht werden kann. Zudem gibt es Anzeichen dafür (ähnliche Genauigkeiten bei Anwendung unterschiedlicher Klassifikatoren in Publikation 2), dass die Kombination unterschiedlicher Feature Selektionsmethoden eine verbesserte Unabhängigkeit von bestimmten Klassifikationsalgorithmen ermöglicht, was jedoch durch weitere Arbeiten weiter untersucht werden sollte. Als Nachteil lässt sich feststellen, dass diese kombinierten Ansätze deutlich rechenaufwendiger im Vergleich zur homogener oder heterogener Ensemble Feature Selektion sind. An dieser Stelle gilt jedoch kritisch zu hinterfragen, inwieweit dieser Nachteil in der Praxis tatsächlich Relevanz erfährt. Werden die Methoden im Bereich von KDD eingesetzt, erscheint ein größerer Rechenaufwand kein limitierender Faktor zu sein.

Trotz der verwendeten Ensemble Ansätze und nachgewiesener Erhöhung der Stabilität sind weiterhin geringere Instabilitäten vorherrschend, welche durch die Unterschiede der Feature-Subsets der unterschiedlichen Kreuzvalidierungs-Datenteilmengen ersichtlich werden. Dies wirft die Frage auf, wie damit umgegangen werde sollte. Bei den vorgestellten Ergebnissen der Publikationen zeigen sich Überschneidungen der relevantesten Feature in den Kreuzvalidierungs-Datenteilmengen. Eine Möglichkeit könnte sein, für spätere Arbeiten nur die Feature zu verwenden, welche in allen Datenteilmengen auch ausgewählt wurden. Eine Überprüfung der Klassifikationsgenauigkeit sollte dann jedoch erneut auf einem unabhängigen Testset erfolgen.

Instabilitäten

Ein weiterer anzumerkender Punkt ist, dass zwar einzelne Feature Selektionsmethoden in einer Ensemble Feature Selektion Pipeline (z.B. JMIM Algorithmus [171]) Redundanzen erniedrigen, jedoch durch den Aggregationsprozess erneut Redundanzen entstehen [175]. Ein post-hoc Einsatz von redundanzreduzierenden Feature Extraktionsmethoden wie mRMR [170] scheint nur teilweise eine geeignete Lösung, da diese erneut zu einer Erhöhung der Instabilität führen, wie auch in Publikation 1 gezeigt. Daher scheint es notwendig vorab zu entscheiden, ob möglichst geringe

Redundanzen

ge Redundanzen oder eine möglichst stabile Lösung anzustreben ist. Es gilt zu betonen, dass bei den Ergebnissen die Redundanzen in den Subsets nach Ensemble Feature Selektion um ein vielfaches niedriger als in den Eingangsvariablen waren.

Die vorgestellten Ensemble Methoden erlauben ein Ranking der Feature entsprechend ihrer Relevanz für die Klassifikationsaufgabe durchzuführen. Damit wird es, durch Setzung unterschiedlicher Schwellenwerte zur Bestimmung der Subset-Größe, möglich, in Abhängigkeit der kontextspezifischen Anforderungen, unterschiedliche Klassifikationsgenauigkeiten sowie Interpretationslevel zu bedienen. Die Ergebnisse zeigen insgesamt an, dass im biomechanischen Kontext die angewendeten Ensemble Feature Selektion Ansätze äußerst vielversprechend erscheinen, da diese eine erhebliche Verbesserung der Stabilität und Reproduzierbarkeit ermöglichen und damit die Ergebnisse vertrauenswürdiger machen. In Kontexten, bei denen Daten oft hochdimensional bei geringer Stichprobengröße vorliegen, wie es oftmals im Feld der Biomechanik der Fall ist, erscheinen die Ansätze dabei besonders relevant [176]. Die Anwendung im Rahmen von sowohl Wellenformdaten (Publikation 1) als auch automatisch extrahierten Features auf Basis relationaler Datenbestände (Publikation 2) erscheint im Rahmen der vorstellten Pipelines möglich und vielversprechend. Angeregt durch die Ensemble Feature Selektion wurden erste Ideen hinsichtlich Ensemble XAI Methoden ausgeführt, deren nähere Erläuterung in der nachfolgenden Sektion 4.3 zu finden sind.

**Zwischenfa-
zit**

Highlights:

- Homogene und heterogene Ensemble Feature Selektion Methode sowie deren Kombination zur deutlichen Erhöhung Stabilität und damit Vertrauenswürdigkeit erfolgreich eingesetzt
- Anwendung bei Wellenformdaten und automatisch extrahierten Features auf Basis relationaler Datenbestände möglich und vielversprechend
- Trotz Ensemble Feature Selektion weiterhin geringere Instabilitäten vorherrschend
- Aggregationsprozess bei Ensemble Feature Selektion erhöht teilweise Redundanzen → Tradeoff Redundanz vs. Stabilität
- Ensemble Feature Selektion ermöglicht Ranking der Feature entsprechend globaler Relevanz → Bedienung unterschiedlicher Klassifikationsgenauigkeiten & Interpretationslevel möglich

4.3 Einflussfaktoren Interpretierbarkeit

Einzelne Aspekte zum Thema Interpretierbarkeit wurden in Zusammenhang mit der Ausgangsdatenlage und deren Ausnutzung in Sektion 4.1 ausgeführt. Die weiteren bestehenden Diskussionspunkte sollen weiter ausgebaut, in ein übergeordnetes Strukturkonzept gebracht und vernetzend diskutiert werden.

Automatische Interpretationen von Daten können auf unterschiedlichen Ebenen durchgeführt werden. Im Folgenden soll eine Unterscheidung von *direkten Interpretationen* und *indirekten Interpretationen* die Ausgangslage bilden. Von direkten Interpretationen von Daten soll gesprochen werden, wenn die Vorhersage der Al-

**direkte &
indirekte
Interpretationen**

gorithmen genau den Labels der Trainingsdaten entsprechen. Hierunter fällt die Klassifikation in Krank und Gesund, aber auch die Prediktion von Bildregionen, welche in Zusammenhang mit Krankheitsbildern stehen (z.B. voxelweise Vorhersage für 3D-Bilder [177]). Die Evaluation dieser direkten Interpretationen ist eindeutig anhand der gegebenen Grundwahrheit im Rahmen der überwachten Lernvorgänge ermöglicht.

Indirekte automatische Interpretationen können auf Basis von Modellen, welche direkte Interpretationen liefern, erzeugt werden. Sie unterscheiden sich von direkten Interpretationen dahingehend, dass die Ergebnisse sich nicht an vorab klaren Labels orientieren, sondern die Entscheidungsgrundlagen der trainierten Modelle aufgedeckt werden. Der Einsatz von XAI-Ansätzen lässt sich hier zuordnen.¹ Indirekte Interpretationen werden auch häufig im Bereich klinischer Bilddaten angewendet [100] und in Einzelfällen bei biomechanischen Daten [39], [101]. Die Evaluation dieser Ergebnisse gestaltet sich schwieriger, da keine Grundwahrheit (engl. ground truth) vorliegt.

In den vorliegenden Publikationen wurde der Fokus auf beide Arten von Interpretationen gelegt. Die direkten Interpretationen haben jedoch immer nur auf eine Einschätzung in gesund oder pathologisch abgezielt (Ausnahme Publikation 1) und es wurden keine pathologischen Korrelate auf direktem Wege klassifiziert. Man hätte im Bezug auf die Wirbelsäuledaten auch gelabelte Daten verwenden können, um die Lokalisation einer Auffälligkeit direkt zu prognostizieren (direkte Interpretation). Hierzu hätte anhand gelabelter Trainingsdaten vorab klar sein müssen, wo eine Auffälligkeit zu finden ist. Durch dieses Vorgehen könnten Aussagen getroffen werden, dass mit einer bestimmten Wahrscheinlichkeit an der prognostizierten Stelle eine pathologische Auffälligkeit vorliegt. Am Beispiel Rückenschmerz wird jedoch deutlich, dass dieses Vorgehen nicht möglich ist. Rückenschmerzen können eine spezifische vertebrale Ursache haben, wie zum Beispiel Spondylitis, Bandscheibenvorfall oder Spinalkanalstenose [143]. In diesem Fall sollten gelabelte Daten mehr oder weniger gut erzeugt werden können. Allerdings werden 90 Prozent der Rückenschmerzen als unspezifisch angegeben, wobei kein anatomisches Korrelat als spezifische behandlungsbedürftige Ursache nachgewiesen werden kann [144]. Gelabelte Daten können also gar nicht erst vorliegen bzw. es fehlt aktuell das Wissen, um eine Datenlabelung durchzuführen. Auch [20] weist auf dieses Problem hin. Daher wurden statt direkten Interpretationen für die Lokalisation besonders bedeutsamer Regionen indirekte Interpretationen herangezogen.

Die Klassifikationsgenauigkeiten, welche eine zufällige Klassifikation überschreitet (z.B. für Arthrose und Wirbelsäulenfusion in Publikation 5), sind ein Indiz dafür, dass es anatomische Korrelate geben könnte, die die Krankheitsbilder charakterisieren. Zugleich stellt eine gute Modellgüte eine wichtige Voraussetzung dar, um aufbauende Interpretationen durchzuführen. Anderenfalls besteht die starke Annahme, dass das Modell eine falsche Beziehung zwischen den Featureausprägungen und der Klassenzugehörigkeit gelernt hat. Diese Korrelate lassen sich womöglich aber nicht unter getrennter Betrachtung einzelner Variablen aufspüren. Der neue Ansatz unter Verwendung von Modellinterpretation (indirekte Interpretationen) könnte nun nützlich sein, um ggf. Einsichten zu (komplexen) anatomischen Korrelaten zu

***komplexe
Korrelate***

¹ Entsprechend für den Fall des pathologieunabhängigen Klassifikators lassen sich die XAI-Interpretationen den indirekten zuordnen. Hingegen ist die Klassifikation in Gesund und Ausreißer (pathologisch) eine direkte Interpretation, da gelabelte Daten für die Gesunden vorliegen deren Klassenzugehörigkeit auch direkt vorhergesagt wird.

gewinnen. Mit komplexen anatomischen Korrelaten ist dabei zum einen gemeint, dass dynamischen Daten bzw. relationalen Daten (siehe Publikation 2) Rechnung getragen werden kann, zum anderen Variablen erzeugt werden können, welche diese Korrelate besser abbilden als die Ursprungsvariablen (siehe z.B. auch Kongressbeitrag [178]).

Direkte Interpretationen zeigen Überlegenheit im Hinblick der Validität der Ergebnisse. Zukünftig gilt es den Übergang von indirekten zu direkten Interpretationen anzustreben. Indirekte Interpretationen sollen dazu dienen notwendiges Wissen aufzudecken und könnten die Basis schaffen, direkte Interpretationen zu ermöglichen.

**Vorteil
lokale Interpretationen**

Warum macht die indirekte Interpretation neben der Aufdeckung potentieller Korrelate noch Sinn? Neben der Notwendigkeit der indirekten Interpretation aufgrund der Ausgangslage spricht noch die Möglichkeit, lokale Interpretationen zu erzeugen, besonders für die Verwendung der XAI Methoden. (Wirbelsäulen-) Gesundheit ist ein komplexer Zustand. Entgegen der herkömmlichen Meinung steht der Mensch nicht gerade (siehe Abbildung 4.1). Gesundheit ist nicht einfach eine gerade Wirbelsäule! Stattdessen gibt es oftmals eine physiologische Rotation sowie laterale Abweichung [179]. Dies könnte mit der Lage der Organe (eine invertierte Anordnung aller inneren Organe liegt im Falle des Situs inversus totalis vor) zusammenhängen [180]. Zusätzlich existieren eine große Bandbreite individueller Haltungs- und Bewegungsmuster der Wirbelsäule (wie in Publikation 4 gezeigt). Auch Krankheitsbilder und operative Maßnahmen, wie beispielsweise der Bereich der Wirbelsäulenfusion, sind sehr individuell. Es ist daher sehr komplex zu bestimmen, was im physiologischen Spektrum ist. Dies erschwert es auch anatomische Korrelate zu finden, um gelabelte Daten zu erzeugen. Gängige inferenzbasierte statistische Analysemethoden zielen jedoch oft darauf ab, globale Effekte zwischen verschiedenen Personengruppen zu finden. Globale Interpretationen sind jedoch bei hoch-individuellen Merkmalen (z.B. Lokalisation der Wirbelkörperfusion), die potenziell für Klassenzugehörigkeiten relevant sind, irreführend. Daher gewinnen lokale statt globale Interpretationen eine hohe Relevanz. Die indirekte Interpretation bietet die Möglichkeit auf individuelle Besonderheiten Rücksicht zu nehmen und lokale Interpretationen durchzuführen. Sie können damit dem komplexen Zustand von Gesund besser Rechnung tragen.

Generell muss sich gefragt werden ob durch Verwendung von komplexen, weniger gut interpretierbaren Modellen die Modellgüte so viel besser ist, dass deren Verwendung gerechtfertigt ist. Ähnlich stellt sich auch die Frage, ob durch abstraktere und komplexere Features die Genauigkeit der Modelle so stark erhöht werden kann, dass deren Verwendung gerechtfertigt ist, da deren Abstraktheit und Komplexität auf Kosten der (biomechanischen) Interpretierbarkeit geht. Solche Feature können beispielsweise durch automatische Feature Extraktion von relationalen Daten (Publikation 2), automatische Extraktion auf Basis von Wellenformdaten (Publikation 2 & 4) oder generell Metric Learning (Publikation 4 & Kongressbeitrag [178]) entstehen. Datenrepräsentationen werden allerdings nicht gleichermaßen gut von Algorithmen und Experten verstanden. Ein Tradeoff zwischen optimaler Interpretierbarkeit für die Praxis und hoher Modellgüte kann im Hinblick auf diese (komplexen) Korrelate existieren, da sie auf abstrakter Ebene biomechanische Charakteristika beschreiben. Sie müssen demnach für die Praxis ggf. durch Repräsentations-Matching nutzbar gemacht werden, um hybride Architekturen aus Expertenwissen und datengetriebenen

Tradeoff

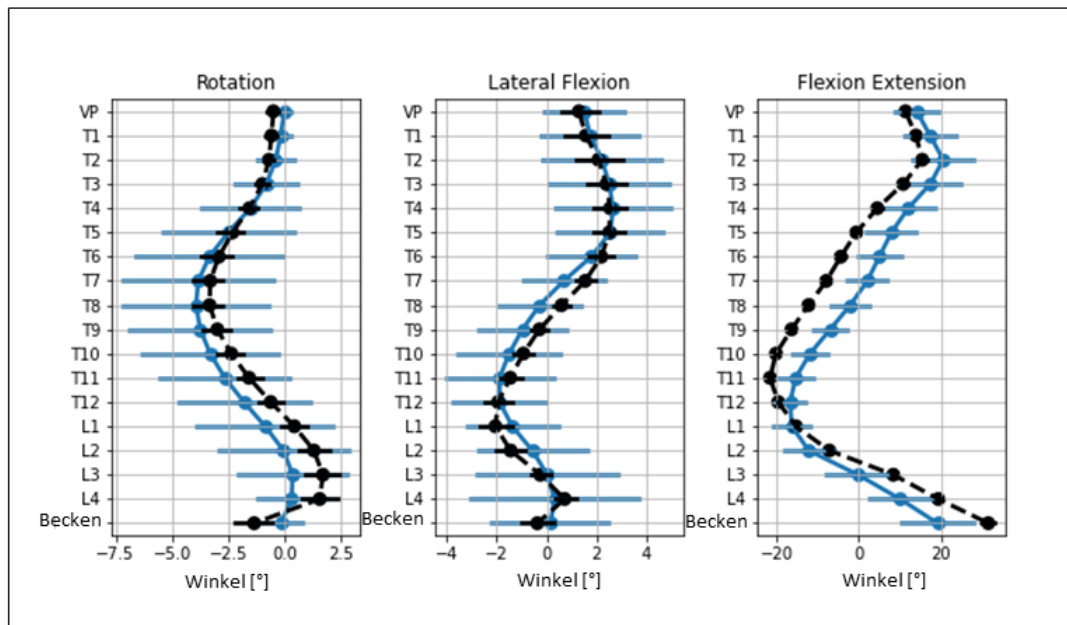


Abbildung 4.1: Mittelwert und Standardabweichung der einzelnen Wirbelkörperpositionen sowie des Beckens bei Gesunden (blaue Linie) sowie Einzelfallbetrachtung einer gesunden Person (schwarze Linie). Die Daten basieren auf Publikation 5. Eigene Darstellung.

Verfahren zu ermöglichen.

Nicht immer sind komplexe Features notwendig, um gute Ergebnisse erzielen zu können. Neben der erfolgreichen Verwendung einfacher Feature auf Basis deskriptiver Statistiken wurden in mehreren der Publikationen erfolgreich Wellenformdaten als Eingangsmerkmale verwendet. Durch Hervorhebung von Bereichen, die für die Klassifikation besonders relevant sind, konnten Interpretationen erzeugt werden, welche visuell direkt auf biomechanische Bewegungsmuster zurückgeführt werden können. Das Verfahren erscheint dadurch vielversprechend für die Zukunft und könnte in Kombination mit diskreten Features eine verbesserte Interpretierbarkeit für die Praxis ermöglichen. Wenn bereits einfache Repräsentationen gute Ergebnisse liefern, sollten diese aufgrund der guten Andockmöglichkeiten an Repräsentationen von Experten präferiert werden (Repräsentations-Matching). Als denkbarer Kompromiss sollte eine Erweiterung von einfachen Features um komplexe angestrebt werden, wenn einfache Features keine zufriedenstellenden Ergebnisse ermöglichen. In diesem Sinne sollte auch interpretierbares Feature Engineering erfolgen, indem relevante und gleichzeitig für den Kontext gut interpretierbare Feature unter Einbeziehung von domänenspezifischem Expertenwissen und automatischen ML Methoden versucht werden zu erzeugen.

Zuletzt soll noch angemerkt werden, dass vorab durchgeführte Feature Selektion die Interpretation, welche auf die Eingangsvariablen zurückgeführt wird, erleichtern könnte, da vorab eine Reduktion auf besonders relevante Parameter erfolgt. Oftmals liegen korrelierte Eingangsvariablen vor, deren Abhängigkeiten zwar auch interessant sein können, allerdings durch die erhöhte Menge an Informationen die Interpretierbarkeit für den Menschen erschwert (bzw. auch Prediktionsgenauigkeit stört [85]). Der in Publikation 1 und 2 vorgestellte Ansatz unter Nutzung von mRMR und JMIM erscheint hier eine vielversprechende Lösung (siehe vorherige Sektion 4.2).

Nach aktuellem Forschungsstand gibt es keine objektiven Kriterien zur Beurteilung der Interpretierbarkeit [181]. Für die Auswertung der Interpretationsergebnisse könnte ein Maß der Aufgabenerfüllung ein sinnvoller Ansatz sein [49]. Da es keine Grundwahrheit für automatisch generierte Interpretationen gibt, sollte eine Bewertung dieser durch Vorwissen sowie durch klinische Experten erfolgen. Eine zusätzliche statistische Verifizierung der Ergebnisse kann zudem, wie in Publikation 3 sowie durch [101], [182] vorgeschlagen, mittels Statistical Parametric Mapping (SPM) erfolgen. Ein weiterer Ansatz um die XAI Ergebnisse zu evaluieren könnte die Verwendung von künstlichen Daten sein, bei denen bewusst bekannte Variationen durchgeführt werden, um diese mit XAI Interpretationen abzugleichen. Eine erste Verwendung von künstlichen Daten sowie die Generierung dieser, unter Sicherstellung ähnlicher biomechanischer Charakteristika, wurde bereits in Publikation 5 durchgeführt. Hier sollten weitere Arbeiten ansetzen.

**Evaluation
Ergebnisse**

Die verwendete Methodik, XAI Interpretationen für mehrere GC der einzelnen Subjekte zu erstellen, ermöglicht es Unsicherheiten der Interpretationen zu quantifizieren und zu evaluieren. Zusammenfassend konnten die Arbeiten aufzeigen, dass die Variationen der LIME Effekte (visualisiert mittels Boxplots) für die unterschiedlichen Gangzyklen der jeweils betrachteten Subjekte in den meisten Fällen gering ausfallen. Zudem indiziert dies auch, dass die XAI Interpretationen eine gewisse Robustheit gegenüber der intra-individuellen Varianz aufweisen.

**Unsicher-
heiten**

In Zusammenhang mit diesen geschilderten Unsicherheiten sowie den teilweise vorhandenen Unterschieden verschiedener XAI Ansätze hinsichtlich der Interpretationen (siehe Publikation 5) könnte die Aggregation der Ergebnisse verschiedener XAI-Ansätze und verschiedener Modelle ein interessanter Ansatz sein, um die Robustheit von XAI-Interpretationen zu erhöhen (Ensemble XAI Interpretationen). In Zusammenhang mit der Feature Selektion hat sich bereits die Ensemble Methode als nützlich erwiesen, um robustere Ergebnisse zu generieren [183], [184] (siehe auch Sektion 4.2). Der Einsatz ähnlicher Methoden im Kontext von XAI könnte daher sinnvoll sein und helfen, die Stabilität und damit das Vertrauen in die XAI-Interpretationen zu erhöhen.

Feature Interaktionen führen dazu, dass eine Vorhersage nicht als Summe der einzelnen Featureeffekte ausgedrückt werden kann, da der Effekt eines Features in Abhängigkeit zu der Merkmalsausprägung eines anderen Features steht.² Interaktionseffekten kommt daher eine Bedeutung bei der (biomechanischen) Interpretation der Ergebnisse zu. Die bisherigen Arbeiten haben auf diese Effekte aufmerksam gemacht, allerdings diese noch nicht ausführlich analysiert. Neben Partial Dependence Plots und der H-Statistik [181] könnte eine Analyse mittels SHAP Interaction Values [107] ansetzen. In Abbildung 4.2 wird eine beispielhafte Analyse aufgezeigt, welche Ausgangspunkt für zukünftige Arbeiten bilden soll.

**Feature In-
teraktionen**

Die vorliegende Arbeit verwendete LIME als XAI Methoden schwerpunktmäßig. Weitere Methoden sollten einbezogen und verglichen werden. [100] konnte so zeigen, dass die Contextual Importance and Utility (CIU) Methode [185] in ihrer Studie besser den menschlichen Entscheidungsprozess unterstützen kann, indem diese bessere kontextspezifische Transparenz und Verstehbarkeit für die Nutzer bereitstellte. Zudem ist ein Übergang in Richtung Ensemble XAI zu evaluieren.

**weitere
Erweiterun-
gen**

² Hierzu passt die Aussage von Aristoteles, dass „das Ganze ist mehr als die Summe seiner Teile“ ist.

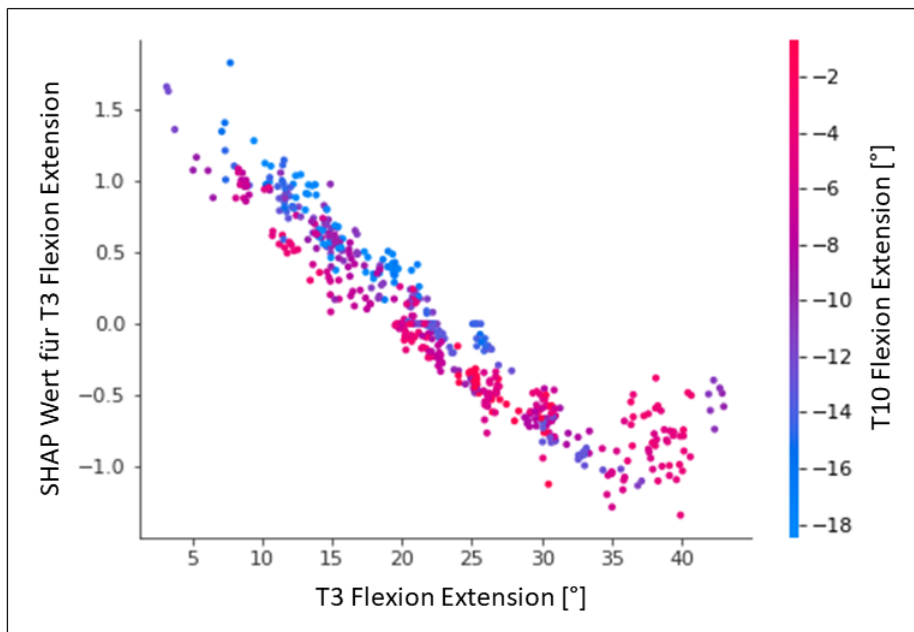


Abbildung 4.2: SHAP-Abhängigkeitsdiagramm (engl. SHAP dependence plot) für das global relevanteste Feature (T3 Flexion Extension) und dessen Interaktion mit einem automatisch ausgewählten Feature (T10 Flexion Extension). Das zugrunde liegende Modell klassifiziert in gesund und pathologisch aufgrund einer Wirbelsäulenversteifung. Niedrige Merkmalsausprägungen des Features T3 Flexion Extension indizieren entsprechend der SHAP Werte eine Zugehörigkeit zur Klasse der pathologischen Subjekte. Umgekehrt indizieren hohe Merkmalsausprägungen eine Zugehörigkeit zur Klasse der Gesunden. Interaktionseffekte der beiden Feature erscheinen für das angeführte Beispiel gering bis nicht vorhanden. Bei der Interpretation von möglichen Interaktionen darf nicht außer Acht gelassen werden, dass es sich möglicherweise nur um Korrelationen ohne Kausalität handelt.

Die Erweiterung der lokalen Ansätze um globale Interpretationen, z. B. durch Permutation Importance oder Partial Dependence Plots [181] könnte ergänzend sein und weitere Einblicke in die Daten hinzufügen. Neben den XAI Methoden zur Interpretation sollte der Aspekt der Datenvisualisierung nicht außen vor gelassen werden. Geeignete Visualisierungsformen könnten ergänzend zu den hier beschriebenen Ansätzen zu weiteren Verbesserungen der Interpretierbarkeit führen und neue Interpretationsmöglichkeiten eröffnen (siehe auch Konferenzbeitrag [186]).

In der Fachliteratur werden teilweise kontroverse Meinungen hinsichtlich der Existenz des Black Box Charakters von Modellen dargestellt [187]–[189]. In den Publikationen konnten erfolgreich Perspektiven aufgezeigt werden, wie Modellentscheidungen nachvollziehbar gemacht und zugleich eine anwendungsspezifische Praxisrelevanz der Informationen gewährleistet werden kann. Selbst wenn entsprechend der Kritik keine vollständige Erhellung der Black Box Modelle ermöglicht wird, soll an dieser Stelle die Parallele zum Einsatz von bestimmten Arzneimitteln betont werden, deren Wirkmechanismen zum Teil auch nicht eindeutig bekannt sind (z.B. auch bei Paracetamol [190]), diese jedoch sich in der Praxis bewährt und daher fortlaufend eingesetzt werden. Alles in allem erscheint der Einsatz der XAI Methoden als ein wichtiger Schritt um ML im biomechanischen Setting erfolgreich einsetzbar zu machen, wobei weiterhin wie dargestellt Herausforderungen bewältigt und in Zusammenhang mit dem Einsatz der XAI Methoden die Datenrepräsentationen sowie kontextspezifische Anforderungen an die Interpretierbarkeit beachtet werden müssen.

*Zwischenfa-
zit*

Highlights:

- Methoden der automatischen direkten & indirekten Interpretation konnten erfolgreich in Kontext Biomechanik transferiert werden → Verbesserung Nachvollziehbarkeit und Interpretierbarkeit
- Ansatz nützlich um Einsichten in (komplexen) anatomischen Korrelaten zu gewinnen
- Verschiedene Datenrepräsentationen haben erheblichen Einfluss auf Genauigkeit, Interpretierbarkeit und klinische Relevanz → Repräsentations-Matching für hybride Architekturen aus Expertenwissen und datengetriebenen Verfahren
- Verifizierung der Ergebnisse durch Experten, Statistical Parametric Mapping & künstlicher Daten erscheint vielversprechend → Übergang von indirekten zu direkten Interpretationen anzustreben
- Feature Interaktionen bei Interpretationen müssen beachtet und zukünftig stärker adressiert werden

4.4 Längs- und Querschnittvergleiche via Siamese Neural Networks (SNN)

In gesundheitlichen Kontexten spielen Personenvergleiche eine entscheidende Rolle, in der Praxis jedoch werden Patienten oftmals einzeln analysiert, da Werkzeuge

oder die Zeit fehlen, neue Patienten statistisch mit Tausenden von anderen Patienten zu vergleichen [167]. Im biomechanischen Kontext werden Subjektvergleiche zusätzlich erschwert, da große Variationen innerhalb der Klasse und kleine Variationen zwischen den Klassen für menschliche Gangbilder berichtet werden [55], [191]. Hinzu kommt, dass die Variabilität zwischen definierten Testsituationen (z.B. unterschiedliche Schuhe, muskuläre Ermüdung) hinsichtlich bestimmter Gelenkwinkel oft größer als zwischen den einzelnen Probanden sein kann [192] (siehe auch Ausführungen Defizit 4). Publikation 4 konnte Ansätze vorstellen, mit denen diese Punkte erfolgreich adressiert werden konnten. Es konnten Datenrepräsentationen durch Metric Learning via SNN generiert werden, welche die intra-individuelle Varianz minimiert und die inter-individuelle Varianz erhöht haben. Es kann daher eine Bedeutung der Methodik zur Identifikation von biomechanischen Invarianzen von Personen über mehrere Gangzyklen, Geschwindigkeiten und Messzeitpunkte konstatiert werden. Die Querschnittsbetrachtung der anderen Publikationen wurde zudem damit um eine Längsschnittbetrachtung erweitert. Schwerpunkt soll nachfolgend die Erweiterung der Ansätze sowie die Darstellung möglicher Potentiale im Kontext pathologischer Krankheitsbilder sein. Für die Gegenüberstellung von Publikation 4 mit dem pathologieunabhängigen Klassifikator sei auf den nächsten Diskussionsabschnitt 4.5 verwiesen.

Im aktuellen Kontext wurden die Methoden nicht für die Identifikation bzw. objektive Beurteilung von Pathologien eingesetzt. Dies könnte ein besonders bedeutender praktischer Einsatzschwerpunkt sein, insbesondere da klinische Daten oftmals nicht in großen Mengen verfügbar sind und so das eingesetzte One Shot Learning besondere Relevanz erfährt. In diesem Sinne wurden SNN bereits in mehreren klinischen und gesundheitsbezogenen Studien eingesetzt (Ähnlichkeiten von Patienten aus elektronischen Patientenakten [193]; Beurteilung des Krankheitsschweregrads und Erkennung von Veränderungen Frühgeborenenretinopathie in Netzhautfotos und Arthrose in Knieröntgenaufnahmen [194], Vorhersage des symptomatischen Verlaufs der Alzheimer-Krankheit [195], automatisches Tracking der Lendenwirbelsäule [196], Erkennung von Gehirnasymmetrien [197]).

*Transfer
Pathologien*

Ein wichtiger Schritt für diesen Transfer in praktische Settings wäre zudem die Verbesserung der Interpretierbarkeit der Ansätze aus Publikation 4. In Publikation 4 wurde bislang keine lokale Interpretation durchgeführt, sondern nur eine globale Relevanzmessung auf Basis der deskriptiv extrahierten Parameter. Die Anwendung der in den anderen Publikationen verwendeten XAI Methoden zur lokalen Interpretation der Ergebnisse ist daher noch ausstehend. Passend zu diesem ausstehenden Projekt konnten [198] bereits erfolgreich einen Autoencoder einsetzen und Interpretationen der Ergebnisse mittels SHAP im Kontext hämatopoetischer Krebssubtypen erzeugen.

Wie bereits durch [199] im Kontext der Lendenwirbelsäulen-Operation angedeutet, ist ein Umdenken und die Erstellung neuer Metriken zur Beurteilung von klinischen Veränderungen notwendig. Das durch die Methodik von Publikation 4 ermöglichte Vorgehen, Personenvergleiche auf Basis Distanzmetriken durchzuführen ist dabei besonders vorteilhaft, da dieses eine Einteilung in absolute Klassen überschreiten kann. Dies ist besonders relevant, da Veränderungen, z.B. während eines Rehabilitationsprozesses mehr oder weniger kontinuierlich verlaufen und das (automatische) Monitoring des Prozesses anstatt einer absoluten Klassifikation in Krank und Gesund Relevanz erfährt. Das Erlernen der Distanzmetrik oder die Verwendung alternativer Metriken könnten weitere Forschungsperspektiven sein.

*kontinuierliche
Metriken*

Mit dem Ansatz sollte es zudem möglich sein, eine Annäherung bzw. Entfernung an ein komplexes Verständnis von Gesundheit zu monitoren. Wie auch in Sektion 4.3 ausgeführt ist es schwer zu bestimmen, was im physiologischen Spektrum liegt, da die Wirbelsäulenstatik und -dynamik durch hochgradige Individualität (und intra-individuelle Varianz) in der initialen Datenrepräsentation geprägt ist. Durch das Metric Learning können Datenrepräsentationen erzeugt werden, welche eine Bewertung auf Ebene eines latenten Raumes ermöglichen. Durch weitere Arbeiten und Anwendung auf Krankheitsbilder gilt zu prüfen, inwieweit die Beurteilung, ob etwas im physiologischen Bereich liegt, auf Ebene eines latenten Raumes tatsächlich geeignet ist. Vielversprechende Ergebnisse konnten hierzu bereits im Kontext eines Hautläsionen Klassifikators gezeigt werden, welcher latente Repräsentationen erlernte, die menschlich verständlichen Konzepten entsprechen [200].

Zudem sollte es auf Basis eines großen multizentrischen Datenpools möglich sein, durch die Personenvergleiche digitale Zwillinge identifizieren zu können (Personen, welche im latenten Raum besonders hohe Ähnlichkeiten aufzeigen) und damit die Möglichkeit geboten werden, bessere Abschätzungen z.B. von Therapieerfolgen zu ermöglichen. Dies lässt sich im Kontext der personalisierten Medizin durch hochgradig individuelle Bewegungsmuster und Adaptionen im Rahmen von Therapien festmachen [201].

Es ist generell schwierig genaue Aussagen zu treffen, welche Größe Datensätze aufweisen müssen, um erfolgreich Modelle anzulernen, jedoch werden in den aller meisten Fällen große Datenbestände zum Training von ML Modellen eingesetzt [202]. Die eingesetzten Methoden von Publikation 4 haben sich als vielversprechend erwiesen, um auch bei geringer Datenlage Modelle erfolgreich zu erstellen. An dieser Stelle soll betont werden, dass die notwendige Ausgangsdatenmenge von verschiedenen Faktoren abhängt. Allgemein kann gesagt werden, dass, wenn Klassen gut abgegrenzt sind und eine gute Datenqualität gegeben ist, auch weniger Daten ausreichen um erfolgreich Modelle zu trainieren. Falls das Vorhandensein gelabelter Daten die Ausgangsdatenmenge einschränkt (nur wenige gelabelte Daten sind vorhanden) sei darauf hingewiesen, dass dann Ansätze des halb-überwachten Lernens besonders vielversprechend erscheinen, wobei jüngst Generative Adversarial Networks (GANs) besonders an Interesse gewonnen haben und im medizinischen Kontext auf Basis von Bilddaten bereits eingesetzt werden [203].

Als andere Möglichkeit zur Bewältigung von ML Aufgaben bei geringer vorherrschender Datenlagen ist das Transfer Learning zu nennen. In diesem Bereich konnte [204] bereits erste Erfolge erzielen, indem durch Verwendung eines bereits vortrainierten Convolutional Neural Network (CNN) Klassifikators bereits wenige Schritte ausreichten, um neue Gangarten zu lernen und zu identifizieren.

Es gibt zum Teil starke Klassenüberlagerungen, welche eine Klassifikation erschweren, insbesondere wenn wenige Daten vorhanden sind. Diese Arbeit möchte hier eine ganzheitliche Perspektive einnehmen und betonen, dass die reine Fokussierung auf die Algorithmik die Betrachtung verkürzt.³ An dieser Stelle soll die Gestaltung der Experimente / Messbedingungen hervorgehoben werden. Viele bestehende Studien fokussieren auf den Gang in der Ebene oder auf dem Laufband. Bewegungen wie Stufen gehen, springen oder Gegenstände heben könnten aber

Datenmenge

Optimales Versuchsdesign

³ Andere Netzwerkstrukturen oder alternative Metric Learning Ansätze könnten nichtsdestotrotz für zukünftige Forschungsarbeiten interessant sein. Im Kontext Rückenschmerz haben eigene Arbeiten diese teilweise schon erprobt (siehe hierzu Konferenzbeitrag [178]).

zu viel besser diskriminierenderen Variablen führen. Daher sollte nicht nur an der Verbesserung der Modelle gearbeitet werden, sondern auch ein starker Fokus auf ein möglichst optimales Versuchsdesign gelegt werden. In dieser Hinsicht ist daher eine transdisziplinäre Zusammenarbeit von Data Science und Sportwissenschaft von hoher Bedeutung!

Zusammenfassend konnte gezeigt werden, dass SNN, welche bereits erfolgreich bei Bilddaten eingesetzt werden konnten (z.B. [205]), erfolgreich in den Kontext biomechanischer Daten transferiert werden können. Insgesamt erscheinen die vorgestellten Ansätze vielversprechend bei Vorliegen einer intra-individuellen Variabilität zwischen den bei unterschiedlichen Geschwindigkeiten aufgezeichneten Gangzyklen eines Tages sowie der Tag zu Tag Variabilität, um die intra-individuellen Varianzen zu minimieren und Personenvergleiche und -identifikationen zu verbessern. Wie [206] im Bereich der Verarbeitung von Elektromyografie-Daten festgestellt hat, könnte die Arbeit ein erster Schritt in Richtung einer Verschiebung von klassischem Feature Engineering (mit oft handgefertigten Features) in Richtung Feature Learning im Bereich der Biomechanik sein.

*Zwischenfa-
zit*

Highlights:

- Erfolgreich bei Bilddaten eingesetzte SNN konnten in den Kontext biomechanischer Daten transferiert werden
- Reduktion intra-individuelle Variabilität durch SNN → Verbesserung Personenvergleiche und -identifikationen
- Vielversprechende Methodik bei geringer Datenlage
- Bedeutung der Methodik zur Identifikation von biomechanischen Invarianzen
- Anwendung bei Krankheitsbildern und Beurteilung mittels kontinuierlicher Beurteilungsmetriken ausstehend
- Schritt zur Verschiebung von klassischem Feature Engineering in Richtung Feature Learning im Bereich der Biomechanik

4.5 Pathologieunabhängige Klassifikation und Interpretation

Bei der Identifikation anormaler Fälle im Kontext malaginer Melanome [96] sowie der Detektion von anormalen Segmenten bei Fingertippen-Wellenformdaten [207] konnten bereits vielversprechende Klassifikationsergebnisse mittels Einklassen-Klassifikatoren aufgezeigt werden. Mit der hier dargestellten vorbereitenden Publikation sowie Publikation 5 wurden die ersten Arbeiten im Bereich der biomechanischen Datenanalyse mittels Einklassen-Klassifikatoren publiziert. Die angeführten Arbeiten haben dabei die Klassifikatoren als Black Boxen verwendet. Aufbauend auf den Ansätzen schwerpunktmäßig aus Publikation 3 wurde mit den beiden Publikationen nach bestem Wissen des Autors erstmalig die lokale Interpretierbarkeit von Einklassen-Klassifikatoren adressiert. Damit eröffnen sich Potentiale, den Ansatz im klinischen Kontext einsatzfähig zu machen und ihn nicht nur zur Klassifikation, son-

dern auch zur datenbasierten, objektiven Entscheidungsunterstützung zu verwenden. Da es sich bei den vorgestellten Arbeiten um erste Arbeiten in dem Feld handelt und wenige bis keine Vergleichsarbeiten vorliegen, ergeben sich weitreichende Diskussionspunkte und Forschungsperspektiven. Aspekte die die Interpretierbarkeit betreffen sollen nur für spezifische Punkte mit Bezug zum pathologieunabhängigen Klassifikator aufgeführt werden. Für die generelle Diskussion der Interpretierbarkeit des Klassifikators sei auf die Sektion 4.3 verwiesen.

Neben der allgemeinen Realisierbarkeit eines interpretierbaren Einklassen-Klassifikators war zu klären, ob überhaupt ausreichend gute Klassifikationsergebnisse mit diesem erzielt werden können oder kein Weg an einem klassischen binären Klassifikationsansatz, und den damit einhergehenden Nachteilen, vorbeiführt. Es gab divergierende Hinweise darauf, dass einerseits kein klarer systematischer Leistungsunterschied erkennbar war und andererseits binäre Klassifikatoren bei geringer Klassentrennung besser abschneiden. Um die genannten Aspekte zu klären, sind weitere Forschungen notwendig. Zusammengefasst konnte jedoch keine klare Unterlegenheit des pathologieunabhängigen Klassifikators festgestellt werden.

Bei genauerer Betrachtung der Wahrscheinlichkeiten für die Klassenzugehörigkeit einzelner Subjekte fällt auf, dass bei mehreren Subjekten keine stark polarisierende Klassenzugehörigkeit vorhanden ist. Wenn die Ergebnisse tatsächlich in der Praxis genutzt werden sollen, sollten nur Fälle betrachtet werden, bei denen sich das Modell besonders sicher ist. Ein bestimmter Grenzwert könnte hier nützlich sein. Demnach gilt es nicht nur die absolute Klassifikationsgenauigkeit zu verbessern, sondern vielmehr unter Verwendung der entsprechenden Metriken (z.B. Brier Skill Score [208]) die Qualität der Vorhersagewahrscheinlichkeiten zu verbessern. Platt Scaling [209] wurde in beiden Publikationen eingesetzt, um die Klassifikationsausgabe in eine Wahrscheinlichkeitsverteilung über die Klassen zu transformieren und eine Interpretation mittels XAI zu ermöglichen. Isotonic Regression könnte für zukünftige Arbeiten hierfür zudem nützlich sein [210].

Hyperparameter-Optimierungsprobleme sind hauptsächlich NP-schwere Probleme [211]. Daher sollten fortgeschrittenere Optimierungsmethoden der Hyperparametersuche (siehe z.B. [212]) zunächst ansetzen, da bislang diese nicht vollständig (Standardwerte der Bibliotheken wurde zum Teil verwendet) sowie nur durch zufällige Hyperparametersuche optimiert wurden. Ferner wurde die Leistung des Validierungssets bislang als das Verhältnis der falsch klassifizierten gesunden Probanden zur Gesamtzahl der gesunden Probanden im Validierungsset gemessen. Die Hyperparametersuche mit einer Bewertungsmetrik, die die Klassifizierungsergebnisse von Gesunden und Patientendaten erfasst, könnte ein vielversprechender alternativer Ansatz zur weiteren Verbesserung des Modells sein.

Mit dem Ansatz war es möglich, mehrere Pathologien auf Basis eines bestehenden trainierten Modelles zu klassifizieren, ohne ein erneutes Training für jede Pathologie durchzuführen. Ein alternatives Vorgehen für die Entwicklung eines Klassifikators, der auf mehrere Pathologien anwendbar ist, wäre ein Mehrklassen-Klassifikator, welcher beispielsweise durch Zerlegung in binäre Klassifikation via one-versus-all Klassifikation Vorhersagen für mehrere Klassen trifft [213]. Allerdings würde durch eine derartige Realisierung keine Pathologieunabhängigkeit gegeben sein. Es konnte so gezeigt werden, dass eine Einklassen-Klassifikation einem Zwei- oder Mehrklassen-Klassifikator vorzuziehen ist, wenn zum Zeitpunkt

Realisierbarkeit

Hyperparameter

Methodische Zugänge

des Trainings nur eine begrenzte Anzahl von Testgruppen zur Verfügung steht, aber weitere Gruppen, abweichend von der Normgruppe, zu erwarten sind [214]. Genau dies ist im klinischen und generell biomechanischen Kontext zu erwarten, wo es kaum realisierbar erscheint, alle Arten beispielsweise von Pathologien in den Trainingsdaten eines Mehrklassen-Klassifikationsmodells zu repräsentieren. Ein Einklassen-Klassifikationsansatz scheint damit entscheidend.

Aufgrund der Pathologieunabhängigkeit eignet sich der Klassifikator dabei auch für das automatische Monitoring bei unklarer Krankheitslage. Dies könnte, wie in der vorbereitenden Publikation gezeigt, beispielsweise auf die bestmögliche Wahl der prothetischen Versorgung hinweisen. Bei Personen die keine normalen Bewegungsmuster bzw. Haltung erreichen können, könnte es das Ziel sein, der Normgruppe möglichst nahe zu kommen, d.h. den höchstmöglichen Wahrscheinlichkeitswert für diese zu erreichen.

Zur Realisierung des pathologieunabhängigen Klassifikators wurde die OCSVM aus dem Bereich der Ausreißerdetektion eingesetzt. Andere Methoden aus dem Bereich der Ausreißerdetektion sollten jedoch auch übertragbar zur Realisierung eines pathologieunabhängigen Klassifikators sein. Vielversprechend erscheint so beispielsweise der Isolation Forest Algorithmus, welcher bei klassischen Ausreißerdetektionsaufgaben sogar in Studien der OCSVM überlegen war [97].

Neben der OCSVM oder dem Isolation Forest, welche sich dem Bereich der unüberwachten Ausreißerdetektion zuordnen lassen können, erfährt das Feld der halb-überwachten Ausreißerdetektion immer mehr Aufmerksamkeit [215]. Generell lassen sich drei Arten von Ausreißerdetektionsmethoden unterscheiden: unüberwachte, überwachte und halb-überwachte Ausreißerdetektion. Die Ausreißerererkennung wurde im Kontext des unüberwachten Lernens intensiv erforscht [216] [217]. Deren Lernergebnisse sind jedoch nicht immer zufriedenstellend und die Einbeziehung einiger weniger gelabelter Daten (halb-überwachte Ausreißerdetektion, z.B. bei [81], [218]) erscheint deutliche Verbesserungen hinsichtlich der Identifikationsleistung zu bringen [215]. Der Einsatz von halb-überwachten Ansätzen könnte daher dazu beitragen, die Klassifikationsergebnisse weiter zu verbessern. Ferner, Expertenwissen besser zu integrieren mit dem finalen Ziel eine hybride Architektur durch Integration von Expertenwissen in Data Science zu erzeugen.

Die Methodik von Publikation 4 ermöglicht in der Theorie auch ein Monitoring ohne vorab Training mit Daten von neu hinzukommenden Pathologien. Für die Entwicklung des Modells muss jedoch anfänglich ein Training auf Basis verschiedenster Pathologien erfolgen um Features zu erlernen, welche eine gute Abgrenzung zwischen gesund und pathologisch ermöglichen. Denkbar erscheint, dass der Erfolg eines solchen Modells auf ungesehenen neuen Pathologien umso besser ausfällt, je vielfältiger die initialen Pathologien zur Modellerstellung gewählt sind. Den tatsächliche Erfolg eines solchen Modelles gilt es jedoch noch in Zukunft zu evaluieren. Folglich unterscheiden sich die Ansätze von Publikation 4 und 5 dahingehend, dass für die Übertragung der Ansätze aus Publikation 4 (zumindest teilweise) Daten von Pathologien zum Modelltraining notwendig sind, für die Ansätze von Publikation 5 jedoch nicht zwingend (trotzdem können diese vorteilhaft für z.B. die Bestimmung der Hyperparameter sein). Ein weiterer Unterschied ist, dass Publikation 4 mittels Distanzen Einschätzungen vornimmt, während für Publikation 5 absolute Klassenvorhersagen bzw. Wahrscheinlichkeiten für Klassenzugehörigkeiten die Basis bilden. Hinzu kommt, dass der Ansatz aus Publikation 4 voraussichtlich besser im Kontext intra-individueller Varianzen performt. Ein kombinierter Einsatz beider Methoden

***Publikation
4 vs. 5***

im Kontext eines pathologieunabhängigen Klassifikators könnte in der Praxis, durch Ermöglichung eines gegenseitigen Vergleichs der Ergebnisse, vorteilhaft sein. Inwieweit beide Ansätze zu übereinstimmenden Einschätzungen führen, gilt es in Zukunft zu evaluieren.

Generell kann der entwickelte Klassifikator zur Identifizierung von Probanden mit erhöhtem Verletzungs- oder Erkrankungsrisiko sowie der Verfolgung des Rehabilitationsfortschritts eingesetzt werden. Zudem eignet sich der Ansatz um Ausreißer der gesunden Population zu identifizieren. In Publikation 5 wurden beispielsweise Ausreißer mittels Isolation Forest [97] aus dem Datensatz der Gesunden entfernt, wobei eine Person komplett entfernt wurde. Da der entwickelte Klassifikator Transparenz für die Entscheidungen hat, könnte dieser alternativ auch dafür genutzt werden, Ausreißer aus der gesunden Population zu identifizieren und die biomechanischen Gründe dafür offenzulegen. Dies könnte ein Indiz dafür sein, warum diese Person später Probleme bekommen könnte. Kritisch muss jedoch auch angemerkt werden, dass eventuell zu wenige Trainingsdaten von Gesunden vorliegen und so die Person als Ausreißer erscheint, obwohl ihre Haltung entsprechend der hohen Individualität auf eine größere Population gesehen, nicht mehr auffallend ist.

*Anwendung
bei
Gesunden*

Die vorliegenden Ergebnisse deuten an, dass hohe Zukunftspotentiale in der pathologieunabhängigen Klassifikation liegen. Zusammengefasst konnte keine klare Unterlegenheit des interpretierbaren pathologieunabhängigen Klassifikators im Vergleich zu einem klassischen binären Klassifikator festgestellt werden. Weitere Forschungsarbeiten sind jedoch noch zwingend notwendig. Zudem wurden bislang lediglich statische Wirbelsäuledaten verwendet, sodass verhältnismäßig wenige Eingangsvariablen vorlagen. Die nächste Aufgabe wird es sein, den Klassifikator bei einer deutlich größeren Menge Eingangsvariablen einzusetzen und zu erproben, wie beispielsweise der Fall bei den dynamischen Wirbelsäuledaten.

*Zwischenfa-
zit*

Highlights:

- Potentielle Eignung der OVSVM als interpretierbarer pathologieunabhängiger Klassifikator kann durch Publikationen bestätigt werden
- Keine klare Unterlegenheit des interpretierbaren, pathologieunabhängigen Klassifikators im Vergleich zu einem klassischen binären Klassifikator
- Ermöglicht objektive Orientierung sowie automatisches Monitoring auch bei unklarer Krankheitslage
- Einfluss bei erweiterter Datenbasis sowie Verwendung anderer Ausreißer Detektionsmethoden als zukünftige Forschungsperspektiven

4.6 Weitere methodische Limitationen und kritische Anmerkungen

Die hier geschilderten Limitationen sollen die in den Publikationen sowie bereits in den obigen Diskussionpunkten benannten Punkte erweitern. Besonders soll hier die Perspektive um eine sozioinformatische Betrachtung erweitert werden. Bevor im

Heuristiken

Detail auf einzelne Limitationen eingegangen wird, soll vorab noch eine generelle Limitation beim Einsatz von Methoden, die sich dem Bereich KI zuordnen lassen, erwähnt werden: KI basiert meist auf Heuristiken und nicht auf Algorithmen [219]. Ein Algorithmus ist eine festgelegte Handlungsanweisung, welche garantieren kann, die beste Lösung zu finden. Heuristiken können dies nicht garantieren und es ist meist nicht überprüfbar, ob eine optimale Lösung gefunden wurde.

Generell lassen sich nach [9] Limitationen in die Bereiche Bias, Privatsphäre und Sicherheit sowie Undurchsichtigkeit für den Einsatz von KI in klinischen Kontexten konstatieren. Für die Limitationen sowie die genannten Adressierungsstrategien zur Undurchsichtigkeit der Entscheidungen, sei auf obige Ausführungen verwiesen. Auch der Punkt Privatsphäre und Sicherheit soll an dieser Stelle nicht ausgiebig diskutiert werden, da nicht auf eine direkte Anwendung für die Praxis abgezielt wird.⁴ Zusätzlich sollen hier kritisch die verwendeten Evaluationsmetriken und generelle Unsicherheiten betrachtet und abschließend sollen weitere Limitationen der Messmethodik diskutiert werden.

Obwohl es Argumente dafür gibt, dass das menschliche Bias den Bias der Algorithmen übertrifft, erscheint wichtiger Aufholbedarf in der medizinischen Forschung den wirklich repräsentativen Querschnitt der Bevölkerung zu generieren [9]. Diese Art der Diskriminierung durch Algorithmen ist beispielsweise in Forschungsarbeiten vorzufinden, welche Algorithmen in der Dermatologie einsetzen um Melanome zu diagnostizieren, aber die Hautfarbe nicht berücksichtigen [12] oder den Korpus genomischer Datennutzung, welcher Minderheiten bisher jedoch stark unterrepräsentiert [220]. Kritisch müssen auch die verwendeten Ausgangsdaten der Studien diskutiert werden. Es ist davon auszugehen, dass diese nicht repräsentativ sind. Dies birgt eine große Gefahr, da einzelne Personengruppen bzw. Haltungs- und Bewegungscharakteristika in den Trainingsdaten nicht repräsentiert sind und daher die Anwendung der bestehenden Modelle auf andere Personengruppen verstärkt zu Fehlentscheidungen führen könnten. Die Bedeutung von weltweitem Datenaustausch wird hierdurch hervorgehoben, um repräsentative Ausgangsdatenlagen bereitzustellen, welche die Diskriminierung reduzieren bzw. aufheben. Ein erneutes Training der Modelle mit einer optimierten Ausgangsdatenlage ist für die Zukunft daher anzustreben.

Ausgangsdatenlage

In Zusammenhang mit der Ausgangsdatenlage sei auch der Aspekt des Overfittings angesprochen. Overfitting verringert die Generalisierbarkeit der Ergebnisse und führt zu Fehlern bei der Auswahl der statistischen Modelle, welche die Populationen am besten charakterisieren [221]. Durch Testung der Modelle auf unabhängigen Datensätzen (Cross Validation) wurde dahingehend versucht eine gute Balance zwischen Bias und Varianz zu finden, welche den totalen Fehler minimiert (engl. Bias-Variance Tradeoff) [222]. Im Rahmen KNN wurde gezeigt, dass insbesondere große Netze, welche auf relativ kleinen Datensätzen trainiert werden, anfällig für Overfitting sind [223]. Eine Möglichkeit Overfitting zu adressieren ist die Vergrößerung der Datensätze [224]. Da die vorliegenden Datenmengen für das Training der

Overfitting

⁴ Es soll jedoch nicht unangesprochen sein, dass die Datensicherheit im Spannungsfeld mit Datenaustausch und Datenverfügbarkeit steht, insbesondere wenn Cloud Lösungen, welche heutzutage eine entscheidende Rolle einnehmen, zum Einsatz kommen. Im Folgenden wird sich mehrfach für einen Datenaustausch umfassender Datenbestände ausgesprochen. An dieser Stelle möchte der Autor betonen, dass hierbei die Perspektive, welche primär v.a. förderlich zur Weiterentwicklung der Ansätze sowie zur Knowledge Discovery ist, eingenommen wird. Die Relevanz des Datenschutzes soll in keinster Weise dabei missachtet werden.

Modelle verhältnismäßig klein ausfallen und Patientendaten schwierig zu generieren sind, sollten parallel zur Vergrößerung der Datensätze andere Methoden der Adressierung herangezogen werden. Zu nennen sind hier Regulierungstechniken wie Dropout, L1 und L2 Regulierung [225].

Der Auswahl einer Evaluationsmetrik kommt eine entscheidende Bedeutung bei der kontextspezifischen Beurteilung eines geeigneten Modelles zu. Für z.B. die Verwendung der Area Under the Curve (AUC) als Schlüsselleistungsmetrik, welche oft bei der Modellierung klinischer Daten herangezogen wird, gibt es allerdings erhebliche Diskussionen, da sie tatsächliche Wahrscheinlichkeitswerte nicht beachtet und in Bezug auf die Sensitivitäts- und Spezifitätswerte, die von klinischem Interesse sind, irreführend sein kann [226]. Um diese Kritik zu adressieren wurden in den Publikationen mehrere Metriken berichtet, um die tatsächliche Modellgüte besser einschätzen zu können und auch die Problematik ungleicher Klassenverteilung auf die Performance Metrik zu beachten. Anzumerken gilt, dass für das Hypermetertuning und die Modellauswahl selbst, immer nur eine im Kontext als geeignetste Metrik verwendet wurde. Kontextspezifisch könnte daher die Auswahl anderer Metriken unter Umständen geeigneter sein.⁵ Auch die kombinierte Betrachtung mehrerer im Kontext als sinnvoll erscheinender Metriken während der Modellauswahl könnte lohnenswert sein.

Evaluationsmetriken

Unsicherheiten beeinflussen bedeutend den Entscheidungsprozess von Praktikern [227]. Zu bedenken gilt, dass jeder weitere Schritt in einer Auswertungspipeline zusätzliche Unsicherheiten einbringen kann. Zudem gilt hervorzuheben, dass Unsicherheiten nicht alleine auf der Modellierungsebene auftreten, sondern auch z.B. im Rahmen der Datenerhebung durch Messungenauigkeiten oder Bewegungen der Testsubjekte während statischer Aufnahmen. Sie können zu Fehlinterpretationen führen, falls sie nicht adäquat dargestellt werden [228] und insbesondere im klinischen Kontext bedeutende negative Folgen haben. Daher ist es wichtig diese zu betrachten und geeignet zu visualisieren (unsicherheitsbewusste Visualisierung) [229], [230]. Ein wichtiger zukünftiger Schritt erscheint daher die Modellierung von Unsicherheiten im Rahmen der eingesetzten ML Modelle, z.B. durch die Verwendung von TensorFlow Probability [231], um zusätzliche Informationen der Modelleistung offenzulegen.

Unsicherheiten

Dies gewinnt insbesondere an Bedeutung, da wie bereits mehrfach angeführt Datensätze mit Patienten oft begrenzt in der Personenzahl sind, was Overfitting begünstigt und höhere Unsicherheiten der Modelleistung (Leistungs- / Modellunsicherheit) bedingt. In diesem Fall spricht man von *epistemischer Unsicherheit*, welche durch Hinzufügen weiterer Daten verringert werden kann. Die Erklärung ist, dass es schwieriger für ein Modell ist, stabile Parameterschätzungen auf einem kleineren Datensatz zu finden [232].

Neben der epistemischen Unsicherheit existiert noch die *aleatorische Unsicherheit*, welche mit dem Datengenerierungsprozess verbunden ist und nicht durch Hinzunahme weiterer Daten reduziert werden kann [232]. Eine Quantifizierung kann mittels Konfidenzintervallen erfolgen. Für beispielhafte Ausführungen beider Unsicherheiten sei auf Abbildung 4.3 verwiesen. Zusammenfassend sollten daher Unsicherheiten der Modelle in Zukunft zusätzlich durch geeignete Methoden quanti-

⁵ Beispielsweise könnte ein Tools darauf abzielen, Hochrisikopatienten für bestimmte Interventionen zu identifizieren. Eine geeignete Metrik könnte hier z.B. der Prozentsatz der wahren Fälle im Perzentil mit dem höchsten Risiko sein.

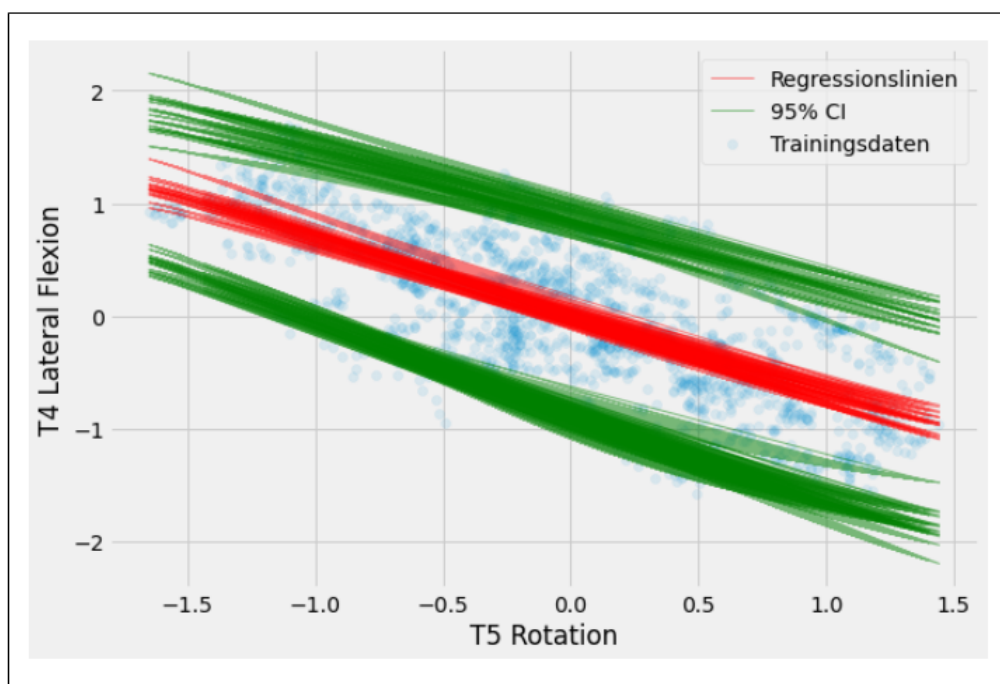


Abbildung 4.3: Trainingsdaten und 15 der Best-Fit-Regressionslinien mit ihren jeweiligen Konfidenzintervallen (CI). Das Diagramm stellt die Schätzung der aleatorischen (Konfidenzintervallen) und epistemischen (Best-Fit-Regressionslinien) Unsicherheit dar im Rahmen der Prediktion der Position von T4 in Frontalebene auf Basis der Rotation von T5 in Transversalebene. Als Trainingsdaten wurde der Datensatz aus Publikation 4, unter alleiniger Verwendung der statischen Wirbelsäuledaten, verwendet. Die Modellierung erfolgte mittels TensorFlow Probability [231]. Die standardisierten Werte sind graphisch aufgetragen. Eigene Darstellung angelehnt an die Ausführungen von [233].

fiziert werden, um Fehlinterpretationen durch zu starke vereinfachte Betrachtungen der Modelleistung zu vermeiden.

Wie bereits aus den vergangenen Ausführungen ersichtlich ist der stärkste Kritikpunkt an der dynamischen rasterstereographischen Vermessung der Wirbelsäule die nicht vollständig überprüfbare Validität an einem Goldstandard. Aber auch andere Methodiken der Vermessung der Wirbelsäule haben Nachteile. Es konnte so z.B. gezeigt werden, dass Frauen mit Skoliose durch die Mehrfachuntersuchungen mit Röntgenstrahlungen ein erhöhtes Brustkrebsrisiko aufweisen [234]. Ein strahlenfreies rasterstereographisches Monitoring könnte so diese Auswirkungen abschwächen bzw. vermeiden. Die zusammenfassende Betrachtung deutet daher an, dass es beim Einsatzschwerpunkt der Rasterstereographie weniger um harte Entscheidungen (z.B. OP-Indikation) gehen sollte. Hierfür erscheint der Rückgriff z.B. auf Röntgen sinnvoll, da hier die Genauigkeit des Systems genau bekannt ist. Schwerpunkt scheint stattdessen eher das Monitoring zu sein, durch die Möglichkeit eine Vielzahl von Messungen durchzuführen.

*weitere
messmetho-
dische
Unsicher-
heiten*

Es konnte zudem gezeigt werden, dass das Gangbild auf dem Laufband sich von dem Gangbild auf festem Untergrund insbesondere hinsichtlich Variabilität unterscheidet [235]. Dies könnte auch Einfluss auf die Variabilität der Wirbelsäulendynamik haben. Die Messungen mit dem DIERS System sind jedoch auf die Verwendung eines Laufbandes angewiesen. Dieser Aspekt sollte mitbedacht werden bei der Übertragung der Ergebnisse in die natürliche Umgebung der Testpersonen.

Highlights:

- Erweiterung Datenbasis unter Beachtung Repräsentativität zukunftsweisend
- Verwendete Evaluationsmetriken müssen kontextspezifisch hinterfragt werden
- Unsicherheiten müssen modelliert, besser quantifiziert und offengelegt werden
- Überprüfung der Übertragbarkeit durch Datenerhebungen im Feld ausstehend

4.7 Zukünftige Forschungsrichtungen und Aufgaben

Eine Vielzahl an Arbeiten im biomechanischen Kontext beschäftigt sich mit überwachten Ansätzen. Auch bei den zugrunde liegenden Publikationen wurde viel in diesem Bereich gemacht. In Anlehnung an beispielsweise [236] könnte die unüberwachte Identifizierung von Bewegungsmustern der Wirbelsäule via Clusteranalyse ein nächster Schritt sein, um ein besseres Verständnis der Wirbelsäulendynamik zu ermöglichen. Hier könnten auch selbstorganisierende Karten Einsatz finden [237]. Ein weiterer vielversprechender Ansatz erscheint die Verwendung von rekurrenten neuronalen Netzen (z.B. Long short-term memory (LSTM)[237], Gated recurrent unit (GRU)[238]), welche im Gegensatz zu den hier verwendeten Feedforward-Netzen eventuell besser zeitlich codierte Informationen in den Daten entdecken [239], [240].

weitere Ansätze

Im klinischen Kontext wurde, wie bereits mehrfach erwähnt, viel im Bereich Bilddaten geforscht. Bei den vorliegenden biomechanischen Daten handelt es sich allerdings zumeist um Wellenformdaten. Als alternativer Zugang, um Methoden im Kontext von Bilddaten direkt anwenden zu können, sollte daher auch ein Image Encoding, also die Umwandlung der vorliegenden Daten in Bilder, z.B. mittels Gramian Angular Fields oder Markov Transition Fields, in Betracht gezogen werden [241], [242] um Methoden, welche bereits erfolgreich im Bereich von Bilddaten eingesetzt werden, auf biomechanischen Daten anwendbar zu machen.

Viele Erfolge beim maschinellen Lernen hängen entscheidend von menschlichen Experten ab, um die geeigneten ML-Architekturen und deren Hyperparameter manuell auszuwählen. Dieses Problem zu adressieren versucht der Bereich des automatisierten maschinellen Lernens (AutoML), welcher auf eine fortschreitende Automatisierung des maschinellen Lernens abzielt [243], [244]. Um ein fortlaufendes Re-Training bei wachsender Datenlage zu ermöglichen (falls kein Continual Learning erfolgt) erscheint AutoML notwendig. Zudem sollten in Zukunft die Modelle, deren Architekturen und Hyperparameter manuell durch Experten bestimmt wurden, mit AutoML Lösungen verglichen werden, um sicherzustellen, dass eine möglichst gute Lösung gefunden wird.

AutoML

Die vorliegenden Arbeiten haben vor allem einen Schwerpunkt auf die Offenlegung von Modellentscheidungen sowie Feature Selektion und Relevanzbestimmung gelegt, um die (medizinischen) Bewegungsdaten verstehbarer und interpretierba-

Visualisierung

rer zu machen. Ergänzend kommt im Bereich Data Science auch der Visualisierung eine hohe Bedeutung zu [245]. Erste Ideen hinsichtlich potentiell geeigneter Visualisierungsformen wurden bereits in einem Konferenzbeitrag ausgeführt [186]. An dieser Stelle sollte aufbauend weitergemacht und Methoden, wie z.B. 2D-Zeitreibencodierung [242] in Betracht gezogen werden.

Bislang wurden einzelne Datenquellen getrennt voneinander betrachtet. Es ist fraglich, inwieweit die menschliche Komplexität damit ausreichend abgedeckt werden kann. Technologieintegration ist daher heute entscheidend. Durch Integration oder Kombination bereits bewährter Technologien können neue Innovationen entstehen. Die Kombination bzw. Integration verschiedener Datenquellen erscheint unabdingbar, um sich der hohen Komplexität des Menschen anzunähern. Die vorgestellten Methoden sollten auch bei Aufnahme weiterer Datenquellen anwendbar sein und noch zusätzliche Relevanz erfahren, da aufgrund der vergrößerten Ausgangsdatenlage die Selektion bedeutsamer Informationen noch wichtiger wird. Für die Erweiterung kinematischer Analysen auf multimodale Analysen erscheint beispielsweise die Hinzunahme von Kinetik, Thermographie sowie Muskelaktivierung interessant. Thermographische Daten ermöglichen beispielsweise Aspekte der Wirbelsäule abzubilden [246], wobei die Auswertung von thermographischer Daten mittels Methoden des ML als Forschungsfeld in letzter Zeit viel Aufmerksamkeit gefunden hat [247]. Eigene Arbeiten bestätigen zudem die Relevanz von thermographischen Daten im Bereich des Monitorings [64]. Zusammenfassend könnte damit das Erlernen von Korrelation von interpretierbaren Features der multimodalen Daten sowie das mögliche Erlernen von Relationen einen wichtigen Schritt für weitere Arbeiten darstellen.

**Integration
weiterer
Datenquellen**

In den meisten realen Anwendungen sind die gesammelten Daten selten von hoher Qualität, verrauscht, fehleranfällig oder anfällig für Manipulationen. Standardmäßige ML- und Datenanalyse-Methoden versagen in solchen Szenarien oft. Selbst geringfügige Störungen der Eingabedaten können zu dramatisch unterschiedlichen Ergebnissen der Modelle des maschinellen Lernens führen [248]. Solche negativen Ergebnisse behindern die Anwendbarkeit dieser Modelle erheblich, führen zu nicht intuitiven und unzuverlässigen Ergebnissen und öffnen zusätzlich die Tür für Angreifer, die diese Schwachstellen ausnutzen können [249]. Gerade in kritischen Einsatzorten, wie z.B. in Gesundheitssektoren, kann dies große Konsequenzen haben [250]. Adversarial Examples nutzen beispielsweise Wissen über Parameter, Algorithmen und die Struktur des Zielmodells um Fehlklassifikationen durch Manipulation der Inputs vorzunehmen, ohne dass diese Veränderungen für den Menschen erkennbar sind [251]. Im Bereich klinischer Bilddaten könnte die große Gefahr dieser Attacken gezeigt werden [250], [252].

Robustheit

Die eingesetzten Methoden wurden bislang nicht hinsichtlich deren Robustheit evaluiert. Die Evaluation und Anpassung dieser erscheint zukünftig substantiell für den Einsatz in klinischen Settings, insbesondere wenn diese im Rahmen eines Smart Cyber-Physical System [253] (CPS; verbindet die Cyber- und die physische Umgebung mit intelligenter Steuerung) zusammen mit dem Internet-of-Things (bietet einen Fernzugriff auf das CPS zum Steuern und Überwachen verbundener Computergeräte) eingesetzt werden sollen.

Ziel sollte es daher für zukünftige Arbeiten sein, robuste ML-Verfahren zu entwickeln, die verschiedene Formen von Fehlern / Beschädigungen sowie Änderungen in der zugrunde liegenden Datenverteilung automatisch handhaben. Insgesamt könnte dies zu zuverlässigen einsetzbaren Modellen führen, welche auch in sensiblen Anwen-

dungsdomänen anwendbar sind. Um dies zu erreichen sollte in Anlehnung an [248] 1) die Wahrnehmbarkeit der Angriffe verbessert werden, um sicherzustellen, dass die sauberen und böartigen Eingaben unterscheidbar sind, oder 2) die Auswirkungen eines Angriffs reduziert werden, indem die Robustheit des Modells demgegenüber erhöht wird (z.B. via Adversal Training [254] oder einer geeigneten Eingangsdatenvorbereitung [255]). Die genannten Punkte sollten hierbei auch im Rahmen biomechanischer Daten umsetzbar sein und in Folgearbeiten evaluiert werden.

Die Art und Weise des Anlernens, insbesondere die Effizienz wenn neue Daten dazukommen, ist ein Thema mit hoher aktueller Relevanz und Ansätze des Continual Learnings werden hier umfassend in der Literatur diskutiert [256]–[258]. Continual Learning ohne katastrophales Vergessen (engl. catastrophic forgetting) beschreibt dabei das effiziente Adaptieren eines bestehenden Klassifikators mit neuen Daten, ohne dass man alle Daten auf einmal präsentieren muss und dann eine sehr lange Trainingszeit hat und auch ohne dass man bei sequentieller Präsentation frühere Infos wieder vergisst [256]. Der Einsatz im klinischen Kontext ist bereits ein diskutiertes Thema [259], [260]. Um ein effektives Training sowie die ständige Weiterentwicklung der hier vorgestellten Modelle bei stetig wachsender Datenlage zu ermöglichen, sollte eine Erweiterung um diese Ansätze erprobt werden.

Continual Learning

Im Rahmen der eingesetzten Wirbelsäuledaten wurde bereits in den Publikationen angeführt, dass der Einsatz von dynamischen Daten vielversprechend erscheint und möglicherweise zu verbesserten Klassifikationsleistungen⁶, z.B. bei den Patienten aus Publikation 5 führen kann. In Kürze wurde argumentiert, dass bei der biologischen Geschlechterklassifikation sich bereits dynamische Daten für die Klassifikation gegenüber der Verwendung statischer Daten als überlegen erwiesen haben (Publikation 2). Zudem wurde argumentiert, dass muskuläre Dysbalancen in der dynamischen Funktion leichter zu erkennen sein sollten. Rückenschmerzen haben so oft ihren Ursprung in der Muskulatur [261]. Durch einen veränderten Bewegungsablauf sind andere Muskelaktivitäten, Winkelverläufe und größere Gelenkmomente daher denkbar [144]. Zusammenfassend sollte zukünftig damit eine kombinierte Betrachtung von sowohl statischen, als auch dynamischen Wirbelsäuledaten erfolgen, um sicherzustellen, dass keine bedeutsamen Informationen a priori verworfen und mögliche Interaktionen beachtet werden.

Dynamische Wirbelsäuledaten

Eine Überführung der Methodik zur Wirbelsäulenanalyse in ein mögliches Medizinprodukt, unter Schaffung einer Schnittstelle, welche eine multizentrische Sammlung und Zusammenführung von Daten, die Anwendung KI-basierter Methoden zur personalisierten Diagnose- und Therapieentscheidungsunterstützung sowie den interdisziplinären Austausch ermöglicht und verbessert, könnte perspektivisch denkbar sein. Da das Messsystem auf Laborsettings beschränkt ist, sollte der Einsatz von IMUs in Betracht gezogen werden. Erste Arbeiten könnten die Einsatzmöglichkeiten bereits bestätigen [262]–[264]. Der mobile Einsatz der Wirbelsäulenanalyse würde beispielsweise auch Ergonomieanalysen im Alltag sowie ein tägliches Monitoring (z.B. erlernter Korrekturhaltung Skoliose) ermöglichen.

Messsystem

⁶ An dieser Stelle soll zudem angeführt werden, dass nicht nur an den Auswertungspipelines gearbeitet werden sollte, um weitere Optimierungen hinsichtlich Prädiktionsgenauigkeit zu erreichen. Um bessere Klassenseparationen zu erreichen ist die Entwicklung eines optimalen experimentellen Designs unabdingbar (siehe Ausführungen Sektion 4.4). Im Kontext der Wirbelsäuledaten könnte eine weitere Variation der Gehgeschwindigkeit oder eine alternative standardisierte statische Haltung eine verbesserte Ausgangslage bieten.

Highlights:

- Erprobung weiterer, in anderen Kontexten erfolgreich eingesetzter Ansätze
- Integration weiterer Datenquellen → Erweiterung kinematischer auf multimodale Analysen
- Verbesserung Robustheit der Verfahren gegenüber Störfaktoren
- Erprobung Continual Learning zur effizienten Adaption bestehender Klassifikatoren mit neuen Daten
- Anwendung der Pipelines bei dynamischen Wirbelsäuledaten
- Verbesserung kontextspezifische Klassenseparationen durch Entwicklung optimaler experimenteller Designs

4.8 Ausblick - die Zukunft von Maschinellem Lernen (ML) bei der Analyse biomechanischer Daten

Der Einfluss von ML kann umrissen auf zwei Ebenen, der Extraktion bedeutungsvoller Informationen aus großen Datenbeständen und die Unterstützung des Menschen gesehen werden [265]. Generell lässt sich methodisch und objektseitig (KI ermöglicht Zugang zu Objektsystemen, welche bislang wissenschaftlich schwer zugänglich waren, z.B. aufgrund ihrer Komplexität) ein induzierter Wandel von Wissenschaft durch KI- und ML beobachten [266].

Gerade für den Einsatz in medizinischen Kontexten besteht jedoch eine hohe **Kritikalität**. Öffentliche bzw. unabhängige Evaluation bei Verwendung von ML Methoden entsprechend der Kritikalität der Systeme ist daher unabdingbar. Daher sollte die Notwendigkeit und Sinnhaftigkeit des Einsatzes von ML für eine Problemstellung immer hinterfragt werden. Nach [219] sollte, wenn es möglich ist, Entscheidungsregeln für ein Problem eindeutig und strukturiert zu definieren, Expertensysteme bevorzugt eingesetzt werden. Zudem sollte der Einsatz von Algorithmen zur optimalen Lösung eines Problems bevorzugt werden, wenn die Problemstellung sich als mathematisches Problem darstellen lässt. Der Einsatz von KI sollte hingegen als Plan B verfolgt werden, wenn große Datenmengen vorherrschen⁷, Hochdimensionalität oder Multikollinearität existieren und auch kleine Korrelationen einbezogen werden müssen. Wie gezeigt war der Einsatz von KI für die Ziele der einzelnen Publikationen notwendig. Durch die Vielzahl an Arbeiten, welche den Einsatz der Methoden zudem im biomechanischen Kontext bekräftigen [20], [21], [101], kann ferner abgeleitet werden, dass bei vielfältigen Problemen im Kontext biomechanischer Daten die Anwendung von KI Methoden gerechtfertigt erscheint. Eine Prüfung des Mehrwertes der Verwendung von ML Ansätzen ist jedoch zwingend notwendig, da diese nicht bei jeder Aufgabenstellung tatsächlich bessere Leistungen erzielen [92].

⁷ Wie in den Publikationen gezeigt, können auch bei kleineren Datenbeständen Erfolge erzielt werden, jedoch besteht durch erhöhte Unsicherheiten die Notwendigkeit besondere Vorsicht bei der Evaluation der Modelle zu zeigen.

Zudem sollte, wie durch die Ausführungen ersichtlich, zusätzlich KI im Rahmen der Erweiterung bestehenden Wissens durch Identifikation (komplexer) Korrelate in Betracht gezogen werden.⁸ Für Einsatzschwerpunkte in Bereichen mit hoher Kritikalität, wie in der Medizin, muss eine möglichst hohe Transparenz der Entscheidungen sowie die Entscheidungshoheit und Kontrolle durch Ärzte Voraussetzung sein. Möglichkeiten der Offenlegung von Entscheidungen im biomechanischen Kontext konnten durch die Publikationen sowie Arbeiten von [39], [58], [101], [135] aufgezeigt werden, weshalb von einem ersten vielversprechenden Schritt in die praktische Applikation ausgegangen werden kann. Der Einsatz von Black Box Modellen kann in weniger kritischen Sektoren gerechtfertigt sein (z.B. automatische Erkennung von Bewegungstypen im Rahmen einer automatischen Trainingsprotokollierung).

Das ständige Wachstum der Rechenleistung und beispielsweise immer kleiner werdende tragbare Sensoren werden auch Wege für allgegenwärtige Echtzeitanwendungen und eine Fülle von verfügbaren Daten im Feld eröffnen. Sie könnte eine Verschiebung in Richtung allgegenwärtigen Gesundheitssysteme (engl. pervasive healthcare systems) ermöglichen [267]. Einsatzfelder werden sich von Trainingsüberwachung und Trainingsunterstützung bis hin zu Rehabilitationsmonitoring oder Einsatzfelder zuhause erstrecken. Mit ML werden wir einen Schritt näher kommen, diese allgegenwärtigen Gesundheitssysteme umzusetzen. Grundsätzlich kann in Anlehnung an die zuvor geschilderten Punkte durch Integration von ML mit Ärzten Einfluss auf drei Ebenen charakterisiert werden. In Kürze können Auswirkungen auf Ebene der Kliniker (schnelle, genaue Interpretationen), Gesundheitssystem (Verbesserung Workflow, Reduktion Fehlentscheidungen, Kosteneinsparungen), Patienten (eigene Daten selbst zu prozessieren, Menschen mehr in eigene Hand geben, Verbesserung personalisierte Medizin, kontinuierliche Einstufung und Monitoring) entstehen (genauere Ausführungen siehe Sektion 2.2.2).

*Gesund-
heitswesen*

Doch welche Maßnahmen müssen im Gesundheitswesen ergriffen werden, damit KI wirklich da helfen kann, wo sie sinnvoll ist? Oft arbeiten klinische BewegungsanalySELABORE meistens getrennt voneinander [167]. Datenaustausch und multizentrische Datennutzung werden in Zukunft hohe Potentiale zur weiteren Verbesserung bieten. Ein kultureller Wandel hin zur gemeinsamen Nutzung von Daten ist notwendig, um die erforderliche Allgemeingültigkeit sowie ein kontinuierliches Weiterlernen der Systeme zu erreichen und diese Systeme in die klinische Praxis einzubringen [22], [167]. Trends lassen sich durch verschiedenste sich entwickelnde Plattformen bereits erkennen (z.B. Simtk.org, ODHSI.org, CrowdSignals.io). Hierfür müssen in Zukunft Infrastrukturen geschaffen und Interoperabilität ermöglicht werden. Die Einführung bzw. der Ausbau der elektronischen Patientenakte könnte hier ein wichtiger Schritt sein. Zudem müssen (digitale) Gesundheitskompetenz aller Akteure berücksichtigt und gefördert werden.

⁸ Methoden des ML sind Methoden der Korrelationssuche und sollten im Rahmen der wissenschaftlichen Methode zunächst nur den ersten Schritt der Beobachtung bilden, um zur Hypothesenbildung beizutragen, welche aufbauend einer Prüfung unterzogen werden müssen.

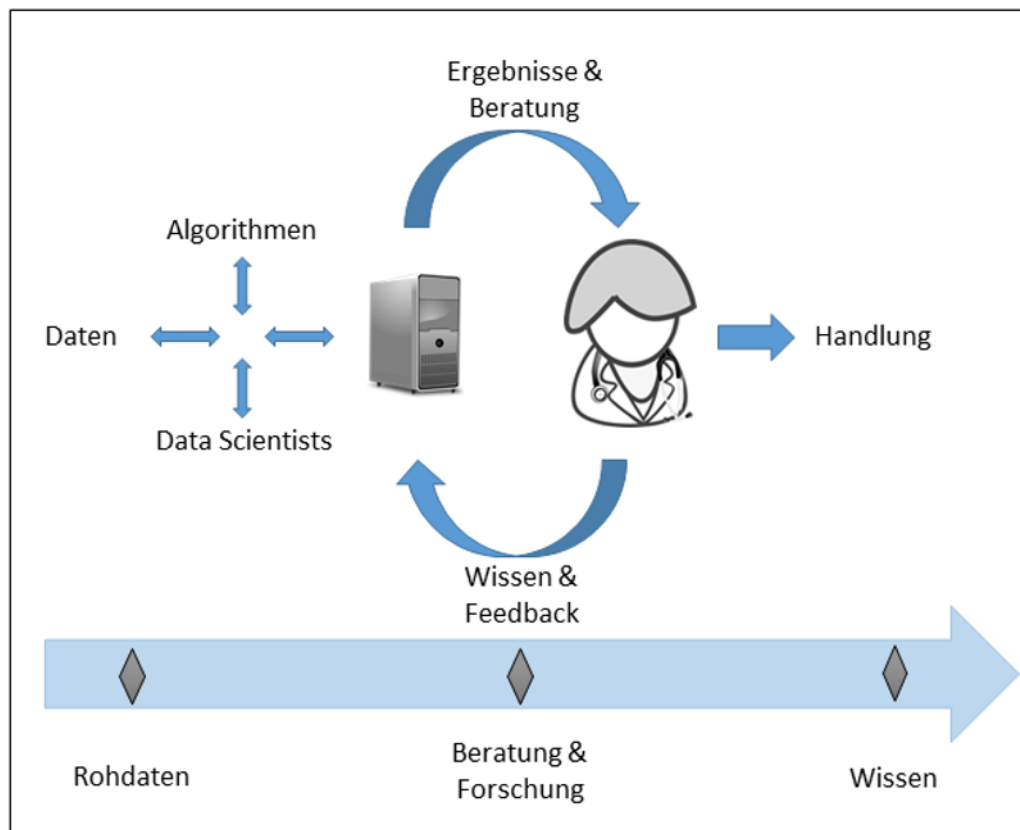


Abbildung 4.4: Visualisierung des Doctor in the Loop Paradigmas. Eigene Darstellung angelehnt und erweitert auf Basis von [268]

Aufgrund der oben genannten Punkte sollten grundsätzlich Symbiosen zwischen medizinischen Praktikern und KI Systemen angestrebt werden. Die Kombination von klinischer Expertise mit Werkzeugen für ML erscheint für die Zukunft am vielversprechendsten [202], [269]. So spricht [66] davon, dass die Unterstützung menschlicher Entscheidungen zu den einflussreichsten Fortschritten zählen wird, welche Data Science und Bewegungswissenschaft der Medizin im nächsten Jahrzehnt gemeinsam bieten können. Es geht also nicht um eine Entmündigung! KI unterstützt Ärzte aber ersetzt sie nicht⁹ und soll ermöglichen, schneller und präziser zu behandeln, um am Ende vielleicht mehr Zeit für ein persönliches Gespräch mit Patient:innen zu haben. Diese geschilderten Punkte werden im sogenannten Doctor in the Loop Paradigma ersichtlich (siehe Abbildung 4.4). Beispielsweise kann KI Zielzonen für die Kopf-Hals-Strahlentherapie genauer und viel schneller berechnen als ein Mensch. Die KI spielt eine wichtige Hintergrundrolle beim Schutz des Patienten vor schädlicher Strahlung, jedoch ist ein Radiologe letztendlich immer noch für die Durchführung der Therapie verantwortlich [271]. Ähnliches könnte in biomechanischen Kontexten denkbar sein.

**Verhältnis
KI -
Mensch**

Die Zielvorstellung könnte also sein, dass sinnvolle KI sich wie t-Test in der Wissenschaft und Stethoskop und Blutdruckmanschette im klinischen Kontext als Tool im Instrumentenkoffer der beiden Einsatzfelder wiederfinden wird.

⁹ Eine teilweise Ausnahme könnte KI global gesehen für Regionen in denen Ärzte unterrepräsentiert sind (siehe [270] für Aspekte des Ärztemangels in Afrika) haben und dort eine Schlüsselrolle einnehmen, um lokal das Gesundheitswesen zu verbessern.

Fazit

IM Jahre 2006 wurde noch gefragt, ob ML in der Biomechanik a „new dawn or false hope“ [272] ist. In der Zwischenzeit hat sich jedoch einiges getan und vorliegende Projekt sowie zahlreiche Arbeiten [20] konnten aufzeigen, dass ML in der Biomechanik vielversprechende neue Perspektiven ermöglicht. Die Fähigkeit moderner Data Science Methoden Erkenntnisse aus großen, komplexen, heterogenen und verrauschten Datensätzen zu gewinnen beginnt, die traditionellen Ansätze der experimentellen Bewegungserfassung und biomechanischen Modellierung zu ergänzen. Mit dieser Arbeit sollte deutlich geworden sein, dass mittels KI bzw. ML Potentiale für eine gewinnbringende Betrachtung von Haltung und Bewegung ermöglicht wird und neue früher bislang wenig bis gar nicht betrachtete Aspekte nun analysiert werden können. Passend dazu sagt [273]:

„To date, artificial intelligence does not simply provide new tools to study human motion. Rather, the way we study human motion is evolving thanks to artificial intelligence.“

[21] spricht davon, dass der zugeschriebene Wert großer biomechanischer Datenbestände hoch erscheint, jedoch der tatsächliche Wert dieser, welcher sich durch Analyse mittels geeigneter Tools ergeben kann, noch bewiesen werden muss. Hierfür müssen smarte Lösungen gefunden werden, welche die Ressourcen der Daten nicht verschwenden (siehe auch Ausführungen zu Defizit 1, 2, 4), indem bspw. wie so oft in der Praxis nur einzelne Variablen betrachtet werden. Stattdessen müssen umfassend die Datenbestände analysiert und Interaktionen betrachtet werden, sonst gehen Informationen verloren, denn das Ganze ist oftmals mehr als die Summe der Einzelteile. Solche smarten Ansätze konnten aufgezeigt und ein erster Schritt in Richtung Ausnutzung der tatsächlichen Potentiale der Datenbestände erbracht werden.

Der Zweck dieser Arbeit war, Perspektiven aufzuzeigen, wie datenwissenschaftliche Methoden in dieses Feld integriert und bestehende Limitationen reduziert werden können. Der Black Box Charakter (siehe auch Ausführungen zu Defizit 3, 5), welcher nicht der europäischen Datenschutzgrundverordnung (GDPR, EU 2016/679) [52] entspricht, sollte also nicht als ärgerliches Hemmnis gesehen werden. Stattdessen erscheint die Betrachtung der GDPR als Innovationstreiber, welche ganz neue Perspektiven eröffnet und Forschungsnotwendigkeit begründet, ein angemessener Perspektivenwechsel, welcher stark im Zusammenhang der hier präsentierten Inhalte steht.

Trotz der vielversprechenden Perspektiven soll abschließend eine Warnung ausgesprochen werden. Das Risiko fehlerhafter Algorithmen insbesondere in kritischen Sektoren ist exponentiell höher als bei einzelnen Ärzteentscheidungen [9]. Als dramatisches Beispiel sei hier der IBM Watson Health’s cancer AI Algorithmus (Watson

for Oncology) genannt, welcher von hunderten Krankenhäusern in der ganzen Welt verwendet wird und teilweise fehlerhafte Entscheidungen empfiehlt, welche sogar divergierend zu Kontraindikationen stehen [274], [275]. Zudem gibt es zahlreiche Anwendungsfälle, bei denen beispielsweise klassische inferenzstatistische Methoden oder Expertensysteme, KI-basierten Systemen vorzuziehen sind [219]. Es sollte daher betont werden, dass der KI Boom und der unreflektierte, pauschale Einsatz der Methodik deutliche Gefahren mit sich bringt. *Der KI Hype darf daher nicht der KI Forschung zuvorkommen!*

Der Einsatz von KI bzw. generell Data Science Methoden wird im Bereich der Biomechanik am erfolgreichsten sein, wenn Ausbildung und Methodenentwicklung in enger Zusammenarbeit der unterschiedlichen Disziplinen erfolgt [20]. Diese Zusammenarbeit und gegenseitige Befruchtung wird auch im Rahmen der KI-Strategie der Bundesregierung [79] gefördert. Genau unter dieser Prämisse lässt sich auch diese Arbeit einordnen. Die dargestellten Projekte sind geprägt durch enge Kooperationen zwischen Sportwissenschaft, Informatik, Klinik und KMUs. Zukünftig braucht man mehr davon!

Bereits mit dem Titel der Arbeit „Auf dem Weg zu interpretierbaren, KI-basierten, explorativen und entscheidungsunterstützenden Systemen im Kontext biomechanischer Daten“ wurde auf den Prozess verwiesen, der in Zukunft womöglich die biomechanische Datenanalyse revolutionieren wird. Die hier beschriebenen Ansätze stellen vielleicht einen kleinen Stein des langen Weges dar, auf dem viele weitere Steine in Zukunft aufbauen können. In Anlehnung daran sowie den Ausführungen von [273] erscheint generell, unter sorgfältiger Vermeidung eines blinden Vertrauens in die Daten, der Weg für eine schnelle und unvermeidbare (Re-)Evolution in der Biomechanik geebnet zu sein.¹

¹ englisches Original: „While carefully avoiding falling into a simplistic (and potentially dangerous) ‚idolatry of data,‘ we believe that the road is paved for rapid and inevitable (re)volutions“ [273].

Literaturverzeichnis

- [1] P. Wasnik, K. B. Raja, R. Ramachandra und C. Busch, “Assessing face image quality for smartphone based face recognition system,” in *5th International Workshop on Biometrics and Forensics (IWBF)*, Coventry, UK: IEEE, 2017, S. 1–6. DOI: 10.1109/IWBF.2017.7935089.
- [2] M. B. Hoy, “Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants,” *Medical Reference Services Quarterly*, Jg. 37, Nr. 1, S. 81–88, 2018. DOI: 10.1080/02763869.2018.1404391.
- [3] M. Fink, Y. Liu, A. Engstle und S.-A. Schneider, “Deep Learning-Based Multi-scale Multi-object Detection and Classification for Autonomous Driving,” in *Fahrerassistenzsysteme 2018*, T. Bertram, Hrsg., Wiesbaden, Germany: Springer, 2019, S. 233–242. DOI: 10.1007/978-3-658-23751-6_20.
- [4] M. Bojarski, D. Del Testa, D. Dworakowski u. a., *End to End Learning for Self-Driving Cars*, 25.04.2016. Adresse: <https://arxiv.org/pdf/1604.07316>.
- [5] T. S. Guzella und W. M. Caminhas, “A review of machine learning approaches to Spam filtering,” *Expert Systems with Applications*, Jg. 36, Nr. 7, S. 10 206–10 222, 2009. DOI: 10.1016/j.eswa.2009.02.037.
- [6] J. G. Claudino, D. d. O. Capanema, T. V. de Souza, J. C. Serrão, A. C. Machado Pereira und G. P. Nassis, “Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a Systematic Review,” *Sports Medicine - Open*, Jg. 5, Nr. 1, S. 1–12, 2019. DOI: 10.1186/s40798-019-0202-3.
- [7] T. Horvat und J. Job, “The use of machine learning in sport outcome prediction: A review,” *WIREs Data Mining and Knowledge Discovery*, Jg. 10, Nr. 5, S. 1–28, 2020. DOI: 10.1002/widm.1380.
- [8] V. H. Buch, I. Ahmed und M. Maruthappu, “Artificial intelligence in medicine: current trends and future possibilities,” *British Journal of General Practice*, Jg. 68, Nr. 668, S. 143–144, 2018. DOI: 10.3399/bjgp18X695213.
- [9] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” *Nature Medicine*, Jg. 25, Nr. 1, S. 44–56, 2019. DOI: 10.1038/s41591-018-0300-7.
- [10] CB Insights Research, *Healthcare Remains The Hottest AI Category For Deals*, 12.04.2017. Adresse: <https://www.cbinsights.com/research/artificial-intelligence-healthcare-startups-investors/>.
- [11] D. Cireşan, U. Meier und J. Schmidhuber, *Multi-column Deep Neural Networks for Image Classification*, 13.02.2012. Adresse: <https://arxiv.org/pdf/1202.2745>.

- [12] A. Esteva, B. Kuprel, R. A. Novoa u. a., “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, Jg. 542, Nr. 7639, S. 115–118, 2017. DOI: 10.1038/nature21056.
- [13] A. Vabalas, E. Gowen, E. Poliakoff und A. J. Casson, “Machine learning algorithm validation with a limited sample size,” *PLoS ONE*, Jg. 14, Nr. 11, S. 1–20, 2019. DOI: 10.1371/journal.pone.0224365.
- [14] D. Helbing, B. S. Frey, G. Gigerenzer u. a., “Will Democracy Survive Big Data and Artificial Intelligence?” In *Towards digital enlightenment*, D. Helbing, Hrsg., Cham: Springer, 2018, S. 73–98. DOI: 10.1007/978-3-319-90869-4_7.
- [15] S. Jun Lee und K. Siau, “A review of data mining techniques,” *Industrial Management & Data Systems*, Jg. 101, Nr. 1, S. 41–46, 2001. DOI: 10.1108/02635570110365989.
- [16] G. Nguyen, S. Dlugolinsky, M. Bobák u. a., “Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey,” *Artificial Intelligence Review*, Jg. 52, Nr. 1, S. 77–124, 2019. DOI: 10.1007/s10462-018-09679-z.
- [17] F. Pedregosa, G. Varoquaux, A. Gramfort u. a., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, Jg. 12, S. 2825–2830, 2011.
- [18] A. Paszke, S. Gross, S. Chintala u. a., “Automatic differentiation in PyTorch,” in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, U. von Luxburg, I. Guyon, S. Bengio, H. Wallach und R. Fergus, Hrsg., Red Hook, NY, United States: Curran Associates Inc., 2017, S. 1–4.
- [19] O. Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, G. K. Karagiannidis und K. Taha, “Efficient Machine Learning for Big Data: A Review,” *Big Data Research*, Jg. 2, Nr. 3, S. 87–93, 2015. DOI: 10.1016/j.bdr.2015.04.001.
- [20] E. J. Harris, I.-H. Khoo und E. Demircan, “A Survey of Human Gait-Based Artificial Intelligence Applications,” *Frontiers in Robotics and AI*, Jg. 8, Nr. 749274, 2022. DOI: 10.3389/frobt.2021.749274.
- [21] A. Phinyomark, G. Petri, E. Ibáñez-Marcelo, S. T. Osis und R. Ferber, “Analysis of Big Data in Gait Biomechanics: Current Trends and Future Directions,” *Journal of Medical and Biological Engineering*, Jg. 38, Nr. 2, S. 244–260, 2018. DOI: 10.1007/s40846-017-0297-2.
- [22] E. Halilaj, A. Rajagopal, M. Fiterau, J. L. Hicks, T. J. Hastie und S. L. Delp, “Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities,” *Journal of Biomechanics*, Jg. 81, S. 1–11, 2018. DOI: 10.1016/j.jbiomech.2018.09.009.
- [23] D. T. H. Lai, R. K. Begg und M. Palaniswami, “Computational intelligence in gait research: a perspective on current applications and future challenges,” *IEEE Transactions on Information Technology in Biomedicine*, Jg. 13, Nr. 5, S. 687–702, 2009. DOI: 10.1109/TITB.2009.2022913.
- [24] T. Chau, “A review of analytical techniques for gait data. Part 1: fuzzy, statistical and fractal methods,” *Gait & Posture*, Jg. 13, Nr. 1, S. 49–66, 2001. DOI: 10.1016/S0966-6362(00)00094-1.
- [25] S. Wolf, T. Loose, M. Schablowski u. a., “Automated feature assessment in instrumented gait analysis,” *Gait & Posture*, Jg. 23, Nr. 3, S. 331–338, 2006. DOI: 10.1016/j.gaitpost.2005.04.004.

- [26] D. Bzdok, N. Altman und M. Krzywinski, “Statistics versus machine learning,” *Nature Methods*, Jg. 15, Nr. 4, S. 233–234, 2018. DOI: 10.1038/nmeth.4642.
- [27] K. W. Johnson, J. Torres Soto, B. S. Glicksberg u. a., “Artificial Intelligence in Cardiology,” *Journal of the American College of Cardiology*, Jg. 71, Nr. 23, S. 2668–2679, 2018. DOI: 10.1016/j.jacc.2018.03.521.
- [28] P. Khera und N. Kumar, “Role of machine learning in gait analysis: a review,” *Journal of Medical Engineering & Technology*, Jg. 44, Nr. 8, S. 441–467, 2020. DOI: 10.1080/03091902.2020.1822940.
- [29] M. N. Bajwa, M. I. Malik, S. A. Siddiqui u. a., “Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning,” *BMC Medical Informatics and Decision Making*, Jg. 19, Nr. 136, S. 1–16, 2019. DOI: 10.1186/s12911-019-0842-8.
- [30] R. Arnaout, L. Curran, Y. Zhao, J. C. Levine, E. Chinn und A. J. Moon-Grady, “An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease,” *Nature Medicine*, Jg. 27, Nr. 5, S. 882–891, 2021. DOI: 10.1038/s41591-021-01342-5.
- [31] L. Hu, D. Bell, S. Antani u. a., “An Observational Study of Deep Learning and Automated Evaluation of Cervical Images for Cancer Screening,” *Journal of the National Cancer Institute*, Jg. 111, Nr. 9, S. 923–932, 2019. DOI: 10.1093/jnci/djy225.
- [32] D. Ardila, A. P. Kiraly, S. Bharadwaj u. a., “End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography,” *Nature Medicine*, Jg. 25, Nr. 6, S. 954–961, 2019. DOI: 10.1038/s41591-019-0447-x.
- [33] A. Yala, C. Lehman, T. Schuster, T. Portnoi und R. Barzilay, “A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction,” *Radiology*, Jg. 292, Nr. 1, S. 60–66, 2019. DOI: 10.1148/radiol.2019182716.
- [34] H. Luo, G. Xu, C. Li u. a., “Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study,” *The Lancet Oncology*, Jg. 20, Nr. 12, S. 1645–1654, 2019. DOI: 10.1016/S1470-2045(19)30637-0.
- [35] E. Dolatabadi, B. Taati und A. Mihailidis, “An Automated Classification of Pathological Gait Using Unobtrusive Sensing Technology,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Jg. 25, Nr. 12, S. 2336–2346, 2017. DOI: 10.1109/TNSRE.2017.2736939.
- [36] H.-y. Lau, K.-y. Tong und H. Zhu, “Support vector machine for classification of walking conditions of persons after stroke with dropped foot,” *Human Movement Science*, Jg. 28, Nr. 4, S. 504–514, 2009. DOI: 10.1016/j.humov.2008.12.003.
- [37] F. Wahid, R. K. Begg, C. J. Hass, S. Halgamuge und D. C. Ackland, “Classification of Parkinson’s Disease Gait Using Spatial-Temporal Gait Features,” *IEEE Journal of Biomedical and Health Informatics*, Jg. 19, Nr. 6, S. 1794–1802, 2015. DOI: 10.1109/JBHI.2015.2450232.

- [38] C. Nüesch, V. Valderrabano, C. Huber, V. von Tscharner und G. Pagenstert, “Gait patterns of asymmetric ankle osteoarthritis patients,” *Clinical Biomechanics (Bristol, Avon)*, Jg. 27, Nr. 6, S. 613–618, 2012. DOI: 10.1016/j.clinbiomech.2011.12.016.
- [39] F. Horst, S. Lapuschkin, W. Samek, K.-R. Müller und W. I. Schöllhorn, “Explaining the unique nature of individual gait patterns with deep learning,” *Scientific Reports*, Jg. 9, Nr. 1, S. 1–13, 2019. DOI: 10.1038/s41598-019-38748-8.
- [40] R. Begg und J. Kamruzzaman, “A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data,” *Journal of Biomechanics*, Jg. 38, Nr. 3, S. 401–408, 2005. DOI: 10.1016/j.jbiomech.2004.05.002.
- [41] W. Teufl, B. Taetz, M. Miezal u. a., “Automated detection of pathological gait patterns using a one-class support vector machine trained on discrete parameters of IMU based gait data,” *Clinical Biomechanics*, Jg. 89, S. 1–7, 2021. DOI: 10.1016/j.clinbiomech.2021.105452.
- [42] A. Kalousis, J. Prados und M. Hilario, “Stability of feature selection algorithms: a study on high-dimensional spaces,” *Knowledge and Information Systems*, Jg. 12, Nr. 1, S. 95–116, 2007. DOI: 10.1007/s10115-006-0040-8.
- [43] G. Victo Sudha George und V. Cyril Raj, “Accurate and Stable Feature Selection Powered by Iterative Backward Selection and Cumulative Ranking Score of Features,” *Indian Journal of Science and Technology*, Jg. 8, Nr. 11, S. 1–6, 2015. DOI: 10.17485/ijst/2015/v8i11/71766.
- [44] P. Levinger, D. T. H. Lai, K. Webster, R. K. Begg und J. Feller, “Support Vector Machines for detecting recovery from knee replacement surgery using quantitative gait measures,” in *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, NY.: IEEE, 2007, S. 4875–4878. DOI: 10.1109/IEMBS.2007.4353432.
- [45] R. K. Fukuchi, B. M. Eskofier, M. Duarte und R. Ferber, “Support vector machines for detecting age-related changes in running kinematics,” *Journal of Biomechanics*, Jg. 44, Nr. 3, S. 540–542, 2011. DOI: 10.1016/j.jbiomech.2010.09.031.
- [46] B. Pes, “Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains,” *Neural Computing & Applications*, Jg. 3, Nr. 3, S. 1–23, 2019. DOI: 10.1007/s00521-019-04082-3.
- [47] B. M. Eskofier, P. Federolf, P. F. Kugler und B. M. Nigg, “Marker-based classification of young-elderly gait pattern differences via direct PCA feature extraction and SVMs,” *Computer Methods in Biomechanics and Biomedical Engineering*, Jg. 16, Nr. 4, S. 435–442, 2013. DOI: 10.1080/10255842.2011.624515.
- [48] J. Figueiredo, C. P. Santos und J. C. Moreno, “Automatic recognition of gait patterns in human motor disorders using machine learning: A review,” *Medical Engineering & Physics*, Jg. 53, S. 1–12, 2018. DOI: 10.1016/j.medengphy.2017.12.006.
- [49] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf und G.-Z. Yang, “XAI—Explainable artificial intelligence,” *Science Robotics*, Jg. 4, Nr. 37, S. 1–2, 2019. DOI: 10.1126/scirobotics.aay7120.

- [50] A. Adadi und M. Berrada, “Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI),” *IEEE Access*, Jg. 6, S. 52 138–52 160, 2018. DOI: 10.1109/ACCESS.2018.2870052.
- [51] W. Samek und K.-R. Müller, “Towards Explainable Artificial Intelligence,” in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen und K.-R. Müller, Hrsg., Cham: Springer, 2019, S. 5–22.
- [52] European Union, “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation),” *Official Journal of the European Union*, Jg. L 119, S. 1–88, 2016.
- [53] A. Holzinger, C. Biemann, C. S. Pattichis und D. B. Kell, *What do we need to build explainable AI systems for the medical domain?* 28.12.2017. Adresse: <http://arxiv.org/pdf/1712.09923v1>.
- [54] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li und L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, New York, NY.: IEEE, 2009, S. 248–255. DOI: 10.1109/CVPR.2009.5206848.
- [55] W. Liu, C. Zhang, H. Ma und S. Li, “Learning Efficient Spatial-Temporal Gait Features with Deep Learning for Human Identification,” *Neuroinformatics*, Jg. 16, S. 457–471, 2018. DOI: 10.1007/s12021-018-9362-4.
- [56] D. Laroche, A. Tolambiya, C. Morisset u. a., “A classification study of kinematic gait trajectories in hip osteoarthritis,” *Computers in Biology and Medicine*, Jg. 55, S. 42–48, 2014. DOI: 10.1016/j.combiomed.2014.09.012.
- [57] W. Teuffl, B. Taetz, M. Miezal u. a., “Towards an Inertial Sensor-Based Wearable Feedback System for Patients after Total Hip Arthroplasty: Validity and Applicability for Gait Classification with Gait Kinematics-Based Features,” *Sensors*, Jg. 19, Nr. 22, 2019. DOI: 10.3390/s19225006.
- [58] F. Horst, D. Slijepcevic, S. Lapuschkin u. a., *On the Understanding and Interpretation of Machine Learning Predictions in Clinical Gait Analysis Using Explainable Artificial Intelligence*, 16.12.2019. Adresse: <http://arxiv.org/pdf/1912a.07737v1>.
- [59] C. Dindorf, W. Teuffl, B. Taetz, S. Becker, G. Bleser und M. Fröhlich, “Feature extraction and gait classification in hip replacement patients on the basis of kinematic waveform data,” *Biomedical Human Kinetics*, Jg. 13, Nr. 1, S. 177–186, 2021. DOI: 10.2478/bhk-2021-0022.
- [60] C. Dindorf, J. Konradi, C. Wolf u. a., “General method for automated feature extraction and selection and its application for gender classification and biomechanical knowledge discovery of sex differences in spinal posture during stance and gait,” *Computer Methods in Biomechanics and Biomedical Engineering*, Jg. 24, Nr. 3, S. 1–9, 2020. DOI: 10.1080/10255842.2020.1828375.
- [61] C. Dindorf, W. Teuffl, B. Taetz, G. Bleser und M. Fröhlich, “Interpretability of Input Representations for Gait Classification in Patients after Total Hip Arthroplasty,” *Sensors*, Jg. 20, Nr. 16, S. 1–14, 2020. DOI: 10.3390/s20164385.

- [62] C. Dindorf, J. Konradi, C. Wolf u. a., “Machine learning techniques demonstrating individual movement patterns of the vertebral column: The fingerprint of spinal motion,” *Computer Methods in Biomechanics and Biomedical Engineering*, S. 1–11, 2021 (published online). DOI: 10.1080/10255842.2021.1981884.
- [63] C. Dindorf, J. Konradi, C. Wolf u. a., “Classification and Automated Interpretation of Spinal Posture Data Using a Pathology-Independent Classifier and Explainable Artificial Intelligence (XAI),” *Sensors*, Jg. 21, Nr. 18, S. 1–18, 2021. DOI: 10.3390/s21186323.
- [64] C. Dindorf, E. Bartaguiz, E. Janowicz, M. Fröhlich und O. Ludwig, “Effects of Unilateral Muscle Fatigue on Thermographic Skin Surface Temperature of Back and Abdominal Muscles—A Pilot Study,” *Sports (Basel, Switzerland)*, Jg. 10, Nr. 3, S. 1–10, 2022. DOI: 10.3390/sports10030041.
- [65] E. Bartaguiz, C. Dindorf, J. Dully, S. Becker und M. Fröhlich, “Is it necessary to reconsider bike fitting practice? On the influence of different intensity levels and treatment-introduced fatigue on the biomechanics of elite cyclists,” under review.
- [66] L. D. Jones, D. Golan, S. A. Hanna und M. Ramachandran, “Artificial intelligence, machine learning and the evolution of healthcare: A bright future or cause for concern?” *Bone & Joint Research*, Jg. 7, Nr. 3, S. 223–225, 2018. DOI: 10.1302/2046-3758.73.BJR-2017-0147.R1.
- [67] R. Fjelland, “Why general artificial intelligence will not be realized,” *Humanities and Social Sciences Communications*, Jg. 7, Nr. 1, 2020. DOI: 10.1057/s41599-020-0494-4.
- [68] M. Flasiński, *Introduction to artificial intelligence*. Cham: Springer, 2016. DOI: 10.1007/978-3-319-40022-8.
- [69] C. Yu, K. Li, Y. Zhang u. a., “A survey on machine learning based light curve analysis for variable astronomical sources,” *WIREs Data Mining and Knowledge Discovery*, Jg. 11, Nr. 5, S. 1–25, 2021. DOI: 10.1002/widm.1425.
- [70] S. M. Weiss und N. Indurkha, *Predictive data mining: A practical guide*. San Francisco: Morgan Kaufmann Publishers, 1998.
- [71] G. E. Piatetsky-Shapiro und W. J. E. Frawley, *Knowledge discovery in Databases*. Menlo Park, CA: MIT Press, 1991.
- [72] U. Fayyad, G. E. Piatetsky-Shapiro und P. Smyth, “From Data Mining to Knowledge Discovery in Databases,” *AI Magazine*, Jg. 17, Nr. 3, S. 37–54, 1996. DOI: 10.1609/aimag.v17i3.1230.
- [73] V. Dhar, “Data science and prediction,” *Communications of the ACM*, Jg. 56, Nr. 12, S. 64–73, 2013. DOI: 10.1145/2500499.
- [74] G. Hu, Y. Yang, D. Yi u. a., *When Face Recognition Meets with Deep Learning: an Evaluation of Convolutional Neural Networks for Face Recognition*, 9.04.2015. Adresse: <https://arxiv.org/pdf/1504.02351>.
- [75] G. Hinton, L. Deng, D. Yu u. a., “Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups,” *IEEE Signal Processing Magazine*, Jg. 29, Nr. 6, S. 82–97, 2012. DOI: 10.1109/MSP.2012.2205597.

- [76] A. Naway und Y. LI, “A Review on The Use of Deep Learning in Android Malware Detection,” *International Journal of Computer Science and Mobile Computing*, Jg. 7, Nr. 12, 2018. DOI: 10.48550/arXiv.1812.10360.
- [77] A. Esteva, A. Robicquet, B. Ramsundar u. a., “A guide to deep learning in healthcare,” *Nature Medicine*, Jg. 25, Nr. 1, S. 24–29, 2019. DOI: 10.1038/s41591-018-0316-z.
- [78] F. Amato, A. López, E. M. Peña-Méndez, P. Vañhara, A. Hampl und J. Havel, “Artificial neural networks in medical diagnosis,” *Journal of Applied Biomedicine*, Jg. 11, Nr. 2, S. 47–58, 2013. DOI: 10.2478/v10136-012-0031-x.
- [79] Bundesministerium für Bildung und Forschung, *Strategie Künstliche Intelligenz der Bundesregierung*, 1.11.2018. Adresse: <https://www.bundesregierung.de/resource/blob/975226/1550276/3f7d3c41c6e05695741273e78b8039f2/2018-11-15-ki-strategie-data.pdf?download=1>.
- [80] W. Ertel, *Grundkurs Künstliche Intelligenz: Eine praxisorientierte Einführung*. Wiesbaden: Springer, 2016.
- [81] A. Daneshpazhouh und A. Sami, “Semi-Supervised Outlier Detection with Only Positive and Unlabeled Data Based on Fuzzy Clustering,” *International Journal on Artificial Intelligence Tools*, Jg. 24, Nr. 3, S. 1–16, 2015. DOI: 10.1142/s0218213015500037.
- [82] Z. C. Lipton und J. Berkowitz, *A Critical Review of Recurrent Neural Networks for Sequence Learning*, 2015. Adresse: <https://arxiv.org/pdf/1506.00019v4.pdf>.
- [83] M. Ester und J. Sander, *Knowledge Discovery in Databases: Techniken und Anwendungen*. Berlin, Heidelberg: Springer, 2000. DOI: 10.1007/978-3-642-58331-5.
- [84] H. Liu und L. Yu, “Toward Integrating Feature Selection Algorithms for Classification and Clustering,” *IEEE Transactions on Knowledge and Data Engineering*, Jg. 17, Nr. 4, S. 491–502, 2005. DOI: 10.1109/TKDE.2005.66.
- [85] S. Khalid, T. Khalil und S. Nasreen, “A survey of feature selection and feature extraction techniques in machine learning,” in *2014 Science and Information Conference*, New York, NY: IEEE, 2014, S. 372–378. DOI: 10.1109/SAI.2014.6918213.
- [86] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari und J. Saeed, “A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction,” *Journal of Applied Science and Technology Trends*, Jg. 1, Nr. 2, S. 56–70, 2020. DOI: 10.38094/jastt1224.
- [87] V. Bolón-Canedo, N. Sánchez-Marño und A. Alonso-Betanzos, “A review of feature selection methods on synthetic data,” *Knowledge and Information Systems*, Jg. 34, Nr. 3, S. 483–519, 2013. DOI: 10.1007/s10115-012-0487-8.
- [88] G. Chandrashekar und F. Sahin, “A survey on feature selection methods,” *Computers & Electrical Engineering*, Jg. 40, Nr. 1, S. 16–28, 2014. DOI: 10.1016/j.compeleceng.2013.11.024.
- [89] W. S. Noble, “What is a support vector machine?” *Nature Biotechnology*, Jg. 24, Nr. 12, S. 1565–1567, 2006. DOI: 10.1038/nbt1206-1565.
- [90] L. Breiman, “Random Forests,” *Machine Learning*, Jg. 45, S. 5–32, 2001.

- [91] I. A. Basheer und M. Hajmeer, “Artificial neural networks: fundamentals, computing, design, and application,” *Journal of Microbiological Methods*, Jg. 43, S. 3–31, 2000. DOI: 10.1016/S0167-7012(00)00201-3.
- [92] E. Christodoulou, J. Ma, G. S. Collins, E. W. Steyerberg, J. Y. Verbakel und B. van Calster, “A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models,” *Journal of Clinical Epidemiology*, Jg. 110, S. 12–22, 2019. DOI: 10.1016/j.jclinepi.2019.02.004.
- [93] L. Peterson, “K-nearest neighbor,” *Scholarpedia*, Jg. 4, Nr. 2, S. 1883, 2009. DOI: 10.4249/scholarpedia.1883.
- [94] F. Kherif und A. Latypova, “Principal component analysis,” in *Machine learning. Methods and applications to brain disorders*, A. Mechelli und S. Vieira, Hrsg., Amsterdam: Academic Press, 2019, S. 209–225. DOI: 10.1016/B978-0-12-815739-8.00012-2.
- [95] A. Tharwat, T. Gaber, A. Ibrahim und A. E. Hassanien, “Linear discriminant analysis: A detailed tutorial,” *AI Communications*, Jg. 30, Nr. 2, S. 169–190, 2017. DOI: 10.3233/AIC-170729.
- [96] S. Dreiseitl, M. Osl, C. Scheibböck und M. Binder, “Outlier Detection with One-Class SVMs: An Application to Melanoma Prognosis,” in *AMIA Annual Symposium Proceedings*, Washington, DC: AMIA, 2010, S. 172–176.
- [97] F. T. Liu, K. M. Ting und Z.-H. Zhou, “Isolation-Based Anomaly Detection,” *ACM Transactions on Knowledge Discovery from Data*, Jg. 6, Nr. 1, S. 1–39, 2012. DOI: 10.1145/2133360.2133363.
- [98] R. Roscher, B. Bohn, M. F. Duarte und J. Garcke, “Explainable Machine Learning for Scientific Insights and Discoveries,” *IEEE Access*, Jg. 8, S. 42 200–42 216, 2020. DOI: 10.1109/ACCESS.2020.2976199.
- [99] P. Linardatos, V. Papastefanopoulos und S. Kotsiantis, “Explainable AI: A Review of Machine Learning Interpretability Methods,” *Entropy (Basel, Switzerland)*, Jg. 23, Nr. 1, 2020. DOI: 10.3390/e23010018.
- [100] S. Knapič, A. Malhi, R. Saluja und K. Främling, “Explainable Artificial Intelligence for Human Decision Support System in the Medical Domain,” *Machine Learning and Knowledge Extraction*, Jg. 3, Nr. 3, S. 740–770, 2021. DOI: 10.3390/make3030037.
- [101] D. Slijepcevic, F. Horst, S. Lapuschkin u. a., “Explaining Machine Learning Models for Clinical Gait Analysis,” *ACM Transactions on Computing for Healthcare*, Jg. 3, Nr. 2, S. 1–27, 2021. DOI: 10.1145/3474121.
- [102] L. van der Maaten und G. Hinton, “Visualizing Data using t-SNE,” *Journal of Machine Learning Research*, Jg. 9, S. 2579–2605, 2008.
- [103] S. Lapuschkin, S. Wäldchen, A. Binder, G. Montavon, W. Samek und K.-R. Müller, “Unmasking Clever Hans predictors and assessing what machines really learn,” *Nature Communications*, Jg. 10, Nr. 1, S. 1–8, 2019. DOI: 10.1038/s41467-019-08987-4.
- [104] A. Shrikumar, P. Greenside und A. Kundaje, “Learning important features through propagating activation differences,” in *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia: ICML, 2017, S. 3145–3153.

- [105] G. Montavon, A. Binder, S. Lapuschkin, W. Samek und K.-R. Müller, “Layer-Wise Relevance Propagation: An Overview,” in *Explainable AI. Interpreting, Explaining and Visualizing Deep Learning*, W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen und K.-R. Müller, Hrsg., Cham: Springer, 2019, S. 193–209. DOI: 10.1007/978-3-030-28954-6_10.
- [106] M. T. Ribeiro, S. Singh und C. Guestrin, “Why Should I Trust You? Explaining the Predictions of Any Classifier,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, B. Krishnapuram, M. Shah, A. Smola, C. Aggarwal, D. Shen und R. Rastogi, Hrsg., New York, 2016, S. 1135–1144. DOI: 10.1145/2939672.2939778.
- [107] S. M. Lundberg und S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Red Hook, NY, United States: Curran Associates Inc., 2017, S. 1–10.
- [108] Bundesministerium für Gesundheit, *E-Health – Digitalisierung im Gesundheitswesen*, 14.12.2021. Adresse: <https://www.bundesgesundheitsministerium.de/e-health-initiative.html>.
- [109] H. Alzamer, T. Abuhmed und K. Hamad, “A Short Review on the Machine Learning-Guided Oxygen Uptake Prediction for Sport Science Applications,” *Electronics*, Jg. 10, Nr. 1956, S. 1–16, 2021. DOI: 10.3390/electronics10161956.
- [110] R. P. Bunker und F. Thabtah, “A machine learning framework for sport result prediction,” *Applied Computing and Informatics*, Jg. 15, Nr. 1, S. 27–33, 2019. DOI: 10.1016/j.aci.2017.09.005.
- [111] A.-W. de Leeuw, S. van der Zwaard, R. van Baar und A. Knobbe, “Personalized machine learning approach to injury monitoring in elite volleyball players,” *European Journal of Sport Science*, S. 1–10, 2021 (published online). DOI: 10.1080/17461391.2021.1887369.
- [112] N. Rommers, R. Rössler, E. Verhagen u. a., “A Machine Learning Approach to Assess Injury Risk in Elite Youth Football Players,” *Medicine and Science in Sports and Exercise*, Jg. 52, Nr. 8, S. 1745–1751, 2020. DOI: 10.1249/MSS.0000000000002305.
- [113] B. Preim und C. Botha, *Visual computing for medicine. Theory, Algorithms, and Applications*. Amsterdam und Boston: Elsevier, 2014.
- [114] P. Lakhani und B. Sundaram, “Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks,” *Radiology*, Jg. 284, Nr. 2, S. 574–582, 2017. DOI: 10.1148/radiol.2017162326.
- [115] H. Chen, L. Wu, Q. Dou u. a., “Ultrasound Standard Plane Detection Using a Composite Neural Network Framework,” *IEEE Transactions on Cybernetics*, Jg. 47, Nr. 6, S. 1576–1586, 2017. DOI: 10.1109/TCYB.2017.2685080.
- [116] A. Iannessi, P.-Y. Marcy, O. Clatz, A.-S. Bertrand und M. Sugimoto, “A review of existing and potential computer user interfaces for modern radiology,” *Insights into Imaging*, Jg. 9, Nr. 4, S. 599–609, 2018. DOI: 10.1007/s13244-018-0620-7.
- [117] MaxQ AI, *MaxQ AI*. Adresse: <https://www.maxq.ai/products>.

- [118] U.S. Food & Drug Administration, *FDA permits marketing of artificial intelligence algorithm for aiding providers in detecting wrist fractures*, 24.05.2018. Adresse: <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-algorithm-aiding-providers-detecting-wrist-fractures>.
- [119] M. Burgess, *The NHS is trialling an AI chatbot to answer your medical questions*, 5.01.2017. Adresse: <https://www.wired.co.uk/article/babylon-nhs-chatbot-app>.
- [120] O. Ludwig, *Ganganalyse in der Praxis: Anwendung in Prävention, Therapie und Versorgung*, 2., erweiterte Auflage. Geislingen: C. Maurer Druck und Verlag, 2015.
- [121] R. Caldas, M. Mundt, W. Potthast, F. Buarque de Lima Neto und B. Markert, “A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms,” *Gait & Posture*, Jg. 57, S. 204–210, 2017. DOI: 10.1016/j.gaitpost.2017.06.019.
- [122] I. H. Lopez-Nava und A. Munoz-Melendez, “Wearable Inertial Sensors for Human Motion Analysis: A Review,” *IEEE Sensors Journal*, Jg. 16, Nr. 22, S. 7821–7834, 2016. DOI: 10.1109/JSEN.2016.2609392.
- [123] P. Picerno, “25 years of lower limb joint kinematics by using inertial and magnetic sensors: A review of methodological approaches,” *Gait & Posture*, Jg. 51, S. 239–246, 2017. DOI: 10.1016/j.gaitpost.2016.11.008.
- [124] I. Weygers, M. Kok, M. Konings, H. Hallez, H. de Vroey und K. Claeys, “Inertial Sensor-Based Lower Limb Joint Kinematics: A Methodological Systematic Review,” *Sensors (Basel, Switzerland)*, Jg. 20, Nr. 3, S. 1–23, 2020. DOI: 10.3390/s20030673.
- [125] M. Bertoli, A. Cereatti, D. Trojaniello u. a., “Estimation of spatio-temporal parameters of gait from magneto-inertial measurement units: multicenter validation among Parkinson, mildly cognitively impaired and healthy older adults,” *BioMedical Engineering OnLine*, Jg. 17, Nr. 1, S. 1–14, 2018. DOI: 10.1186/s12938-018-0488-2.
- [126] M. Alaqtash, T. Sarkodie-Gyan, H. Yu, O. Fuentes, R. Brower und A. Abdelgawad, “Automatic classification of pathological gait patterns using ground reaction forces and machine learning algorithms,” in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, NY: IEEE Engineering in Medicine and Biology Society, 2011, S. 453–457. DOI: 10.1109/IEMBS.2011.6090063.
- [127] Y. Guo und X. Wang, “Applying TS-DBN model into sports behavior recognition with deep learning approach,” *The Journal of Supercomputing*, Nr. 77, S. 12 192–12 208, 2021. DOI: 10.1007/s11227-021-03772-x.
- [128] E. Mitchell, D. Monaghan und N. E. O’Connor, “Classification of sporting activities using smartphone accelerometers,” *Sensors (Basel, Switzerland)*, Jg. 13, Nr. 4, S. 5317–5337, 2013. DOI: 10.3390/s130405317.
- [129] R. Begg und J. Kamruzzaman, “Neural networks for detection and classification of walking pattern changes due to ageing,” *Australasian Physical & Engineering Sciences in Medicine*, Jg. 29, Nr. 2, S. 188–195, 2006. DOI: 10.1007/BF03178892.

- [130] R. K. Begg, M. Palaniswami und B. Owen, “Support vector machines for automated gait classification,” *IEEE Transactions on Biomedical Engineering*, Jg. 52, Nr. 5, S. 828–838, 2005. DOI: 10.1109/TBME.2005.845241.
- [131] E. Khodabandehloo, D. Riboni und A. Alimohammadi, “HealthXAI: Collaborative and explainable AI for supporting early diagnosis of cognitive decline,” *Future Generation Computer Systems*, Jg. 116, S. 168–189, 2021. DOI: 10.1016/j.future.2020.10.030.
- [132] J. Paulo, P. Peixoto und U. J. Nunes, “ISR-AIWALKER: Robotic Walker for Intuitive and Safe Mobility Assistance and Gait Analysis,” *IEEE Transactions on Human-Machine Systems*, Jg. 47, Nr. 6, S. 1110–1122, 2017. DOI: 10.1109/THMS.2017.2759807.
- [133] M. Mundt, W. R. Johnson, W. Potthast, B. Markert, A. Mian und J. Alderson, “A Comparison of Three Neural Network Approaches for Estimating Joint Angles and Moments from Inertial Measurement Units,” *Sensors (Basel, Switzerland)*, Jg. 21, Nr. 13, S. 1–14, 2021. DOI: 10.3390/s21134535.
- [134] N. J. Cronin, T. Rantalainen, J. P. Ahtiainen, E. Hynynen und B. Waller, “Markerless 2D kinematic analysis of underwater running: A deep learning approach,” *Journal of Biomechanics*, Jg. 87, S. 75–82, 2019. DOI: 10.1016/j.jbiomech.2019.02.021.
- [135] F. Horst, S. Lopuschkin, W. Samek, K.-R. Müller und W. I. Schöllhorn, “Explaining the unique nature of individual gait patterns with deep learning,” *Scientific Reports*, Jg. 9, Nr. 1, S. 1–33, 2019. DOI: 10.1038/s41598-019-38748-8.
- [136] L. Hutchinson, J. B. Schwartz, A. M. Morton, I. S. Davis, K. J. Deluzio und M. J. Rainbow, “Operator Bias Errors Are Reduced Using Standing Marker Alignment Device for Repeated Visit Studies,” *Journal of Biomechanical Engineering*, Jg. 140, Nr. 4, S. 1–7, 2018. DOI: 10.1115/1.4038358.
- [137] W. Teufel, M. Miezal, B. Taetz, M. Fröhlich und G. Bleser, “Validity, Test-Retest Reliability and Long-Term Stability of Magnetometer Free Inertial Sensor Based 3D Joint Kinematics,” *Sensors (Basel, Switzerland)*, Jg. 18, Nr. 7, S. 1–22, 2018. DOI: 10.3390/s18071980.
- [138] E. von der Lippe, L. Krause, M. Prost u. a., “Prävalenz von Rücken- und Nackenschmerzen in Deutschland. Ergebnisse der Krankheitslast-Studie BURDEN 2020,” *Journal of Health Monitoring*, Jg. 6, Nr. S3, S. 1–14, 2021. DOI: 10.25646/7854.
- [139] Deutsche Rentenversicherung, *Rentenversicherung in Zahlen 2017*. Berlin: Deutsche Rentenversicherung, 2017.
- [140] H. Raspe, *Rückenschmerzen: Gesundheitsberichterstattung des Bundes Heft 53*. Berlin: Robert Koch-Institut, 2012.
- [141] DAK, *DAK-Studie zeigt im 1. Halbjahr 2021 12 Prozent weniger Fehltag*, 5.08.2021. Adresse: <https://www.dak.de/dak/landesthemen/gesundheitsreport-krankenstand-der-beschaeftigten-sinkt-auf-rekordtief-2473714.html#/>.
- [142] Techniker Krankenkasse, *Operation oder Turn-Matte- Mit ärztlicher Zweitmeinung OP vermeiden*, 14.01.2021. Adresse: <https://www.tk.de/presse/themen/medizinische-versorgung/krankenhausversorgung/mit-aerztlicher-zweitmeinung-operationen-vermeiden-2098948>.

- [143] H.-R. Casser, S. Seddigh und M. Rauschmann, “Acute Lumbar Back Pain,” *Deutsches Ärzteblatt International*, Jg. 113, Nr. 13, S. 223–234, 2016. DOI: 10.3238/arztebl.2016.0223.
- [144] B. W. Koes, M. W. van Tulder und S. Thomas, “Diagnosis and treatment of low back pain,” *BMJ*, Jg. 332, Nr. 7555, S. 1430–1434, 2006. DOI: 10.1136/bmj.332.7555.1430.
- [145] B. Cakir, M. Richter, W. Käfer, M. Wieser, W. Puhl und R. Schmidt, “Evaluation of lumbar spine motion with dynamic X-ray—a reliability analysis,” *Spine*, Jg. 31, Nr. 11, S. 1258–1264, 2006. DOI: 10.1097/01.brs.0000217763.80593.50.
- [146] M. Tousignant, L. Poulin, S. Marchand, A. Viau und C. Place, “The Modified-Modified Schober Test for range of motion assessment of lumbar flexion in patients with low back pain: a study of criterion validity, intra- and inter-rater reliability and minimum metrically detectable change,” *Disability and Rehabilitation*, Jg. 27, Nr. 10, S. 553–559, 2005. DOI: 10.1080/09638280400018411.
- [147] C. Perret, S. Poiraudou, J. Fermanian, M. M. Colau, M. A. Benhamou und M. Revel, “Validity, reliability, and responsiveness of the fingertip-to-floor test,” *Archives of Physical Medicine and Rehabilitation*, Jg. 82, Nr. 11, S. 1566–1570, 2001. DOI: 10.1053/apmr.2001.26064.
- [148] T. Steffen, R. K. Rubin, H. G. Baramki, J. Antoniou, D. Marchesi und M. Aebi, “A new technique for measuring lumbar segmental motion in vivo. Method, accuracy, and preliminary results,” *Spine*, Jg. 22, Nr. 2, S. 156–166, 1997. DOI: 10.1097/00007632-199701150-00006.
- [149] J. P. Dickey, M. R. Pierrynowski, D. A. Bednar und S. X. Yang, “Relationship between pain and vertebral motion in chronic low-back pain subjects,” *Clinical Biomechanics*, Jg. 17, Nr. 5, S. 345–352, 2002. DOI: 10.1016/s0268-0033(02)00032-3.
- [150] I. K. Sahni, J. A. Hipp, B. C. Kirking, J. W. Alexander und S. I. Esses, “Use of percutaneous transpedicular external fixation pins to measure intervertebral motion,” *Spine*, Jg. 24, Nr. 18, S. 1890–1893, 1999. DOI: 10.1097/00007632-199909150-00005.
- [151] A. H. McGregor, L. Anderton, W. M. Gedroyc, J. Johnson und S. P. Hughes, “Assessment of spinal kinematics using open interventional magnetic resonance imaging,” *Clinical Orthopaedics and Related Research*, Nr. 392, S. 341–348, 2001. DOI: 10.1097/00003086-200111000-00044.
- [152] J. Chowańska, T. Kotwicki und K. Rosadziński, “Comparison of standing and sitting position used in surface topography trunk assessment,” *Postępy Nauk Medycznych*, Jg. 25, Nr. 6, S. 476–483, 2012.
- [153] A. Truszczynska, “Measurement of spinal curvatures during sitting on a rehabilitation ball versus stool,” *Türkiye Fiziksel Tıp ve Rehabilitasyon Dergisi*, Jg. 62, Nr. 2, S. 148–155, 2016. DOI: 10.5606/tftrd.2016.10734.
- [154] B. Drerup, B. Ellger, F. M. Meyer zu Bentrup und E. Hierholzer, “Rasterstereographische Funktionsaufnahmen. Eine neue Methode zur biomechanischen Analyse der Skelettgeometrie,” *Der Orthopäde*, Jg. 30, Nr. 4, S. 242–250, 2001. DOI: 10.1007/s001320050603.

- [155] A. R. Turner-Smith, J. Harris, G. R. Houghton und R. J. Jefferson, “A method for analysis of back shape in scoliosis,” *Journal of Biomechanics*, Jg. 21, Nr. 6, S. 497–509, 1988. DOI: 10.1016/0021-9290(88)90242-4.
- [156] B. Drerup und E. Hierholzer, “Back shape measurement using video rasterstereography and three-dimensional reconstruction of spinal shape,” *Clinical Biomechanics*, Jg. 9, Nr. 1, S. 28–36, 1994. DOI: 10.1016/0268-0033(94)90055-8.
- [157] E. Dulina, A. Gola, M. Gašo u. a., “Influence of Various Types of Office Desk Chair for Dynamizing the Operation Assessed by Raster Stereography,” *Applied Sciences*, Jg. 11, Nr. 4910, S. 1–12, 2021. DOI: 10.3390/app11114910.
- [158] I. Roman, M. Luyten, H. Croonenborghs u. a., “Relating the Diers formetric measurements with the subjective severity of acute and chronic low back pain,” *Medical Hypotheses*, Jg. 133, Nr. 109390, S. 1–10, 2019. DOI: 10.1016/j.mehy.2019.109390.
- [159] J. Schröder, K. Strübing und K. Mattes, “Rückenbeschwerden und Wirbelsäulenform,” *Manuelle Medizin*, Jg. 48, Nr. 6, S. 454–459, 2010. DOI: 10.1007/s00337-010-0800-0.
- [160] M. Betsch, M. Wild, B. Johnstone u. a., “Evaluation of a novel spine and surface topography system for dynamic spinal curvature analysis during gait,” *PLoS ONE*, Jg. 8, Nr. 7, S. 1–8, 2013. DOI: 10.1371/journal.pone.0070581.
- [161] N. L. Krott, M. Wild und M. Betsch, “Meta-analysis of the validity and reliability of rasterstereographic measurements of spinal posture,” *European Spine Journal*, Jg. 29, Nr. 9, S. 2392–2401, 2020. DOI: 10.1007/s00586-020-06402-x.
- [162] T. L. Schulte, E. Hierholzer, A. Boerke u. a., “Raster stereography versus radiography in the long-term follow-up of idiopathic scoliosis,” *Journal of Spinal Disorders & Techniques*, Jg. 21, Nr. 1, S. 23–28, 2008. DOI: 10.1097/BSD.0b013e318057529b.
- [163] L. Hackenberg, E. Hierholzer und U. Liljenqvist, “Accuracy of rasterstereography versus radiography in idiopathic scoliosis after anterior correction and fusion,” *Studies in Health Technology and Informatics*, Jg. 91, S. 241–245, 2002.
- [164] L. Wanke-Jellinek, O. Heese, A. Krenauer u. a., “Is there any use? Validity of 4D rasterstereography compared to EOS 3D X-ray imaging in patients with degenerative disk disease,” *European Spine Journal*, Jg. 28, Nr. 9, S. 2162–2168, 2019. DOI: 10.1007/s00586-019-06082-2.
- [165] E. Schrade, “Zur klinischen und unfallmedizinische Bedeutung der Sakralisation und Lumbalisation,” *Archives of Orthopaedic and Trauma Surgery*, Jg. 30, S. 351–361, 1931.
- [166] A. Gipsman, L. Rauschert, M. Daneshvar und P. Knott, “Evaluating the Reproducibility of Motion Analysis Scanning of the Spine during Walking,” *Advances in Medicine*, Jg. 2014, S. 1–10, 2014. DOI: 10.1155/2014/721829.
- [167] R. Ferber, S. T. Osis, J. L. Hicks und S. L. Delp, “Gait biomechanics in the era of data science,” *Journal of Biomechanics*, Jg. 49, Nr. 16, S. 3759–3761, 2016. DOI: 10.1016/j.jbiomech.2016.10.033.

- [168] S. Ullah und C. F. Finch, “Applications of functional data analysis: A systematic review,” *BMC Medical Research Methodology*, Jg. 13, Nr. 43, S. 1–12, 2013. DOI: 10.1186/1471-2288-13-43.
- [169] B. M. Eskofier, M. Kraus, J. T. Worobets, D. J. Stefanyshyn und B. M. Nigg, “Pattern classification of kinematic and kinetic running data to distinguish gender, shod/barefoot and injury groups with feature ranking,” *Computer Methods in Biomechanics and Biomedical Engineering*, Jg. 15, Nr. 5, S. 467–474, 2012. DOI: 10.1080/10255842.2010.542153.
- [170] H. Peng, F. Long und C. Ding, “Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Jg. 27, Nr. 8, S. 1226–1238, 2005.
- [171] M. Bennisar, Y. Hicks und R. Setchi, “Feature selection using Joint Mutual Information Maximisation,” *Expert Systems with Applications*, Jg. 42, Nr. 22, S. 8520–8532, 2015. DOI: 10.1016/j.eswa.2015.07.007.
- [172] J. Wu, J. Wang und L. Liu, “Feature extraction via KPCA for classification of gait patterns,” *Human Movement Science*, Jg. 26, Nr. 3, S. 393–411, 2007. DOI: 10.1016/j.humov.2007.01.015.
- [173] W. Teufl, M. Lorenz, M. Miezal, B. Taetz, M. Fröhlich und G. Bleser, “Towards Inertial Sensor Based Mobile Gait Analysis: Event-Detection and Spatio-Temporal Parameters,” *Sensors (Basel, Switzerland)*, Jg. 19, Nr. 1, S. 1–20, 2018. DOI: 10.3390/s19010038.
- [174] S. Ranakoti, S. Arora, S. Chaudhary u. a., “Human Fall Detection System over IMU Sensors Using Triaxial Accelerometer,” in *Computational intelligence. Theories, applications and future directions*. N. K. Verma und A. K. Ghosh, Hrsg., Singapore: Springer, 2018, S. 495–507. DOI: 10.1007/978-981-13-1132-1_39.
- [175] B. Seijo-Pardo, V. Bolón-Canedo und A. Alonso-Betanzos, “On developing an automatic threshold applied to feature selection ensembles,” *Information Fusion*, Jg. 45, S. 227–245, 2019. DOI: 10.1016/j.inffus.2018.02.007.
- [176] V. Bolón-Canedo und A. Alonso-Betanzos, “Ensembles for feature selection: A review and future trends,” *Information Fusion*, Jg. 52, S. 1–12, 2019. DOI: 10.1016/j.inffus.2018.11.008.
- [177] D. Robben, J. Bertels, S. Willems, D. Vandermeulen, F. Maes und P. Suetens, *DeepVoxNet: voxel-wise prediction for 3D images*. 2018. Adresse: <https://lirias.kuleuven.be/2129622?limo=0>.
- [178] C. Dindorf, J. Konradi, C. Wolf u. a., “Towards a better understanding of spinal differences between healthy subjects and subjects with back pain using Explainable Artificial Intelligence (XAI),” in *9th International Performance Analysis Workshop and Conference & 5th IACSS Conference*, Universität Wien, Hrsg., Wien: Universität Wien, 2021.
- [179] C. Wolf, U. Betz, J. Huthwelker u. a., “Evaluation of 3D vertebral and pelvic position by surface topography in asymptomatic females: presentation of normative reference data,” *Journal of Orthopaedic Surgery and Research*, Jg. 16, Nr. 703, S. 1–10, 2021. DOI: 10.1186/s13018-021-02843-2.

- [180] J.-W. M. Kouwenhoven, L. W. Bartels, K. L. Vincken u. a., “The relation between organ anatomy and pre-existent vertebral rotation in the normal spine: magnetic resonance imaging study in humans with situs inversus totalis,” *Spine*, Jg. 32, Nr. 10, S. 1123–1128, 2007. DOI: 10.1097/01.brs.0000261563.75469.b0.
- [181] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. n.p.: Leanpub, 2018.
- [182] D. Slijepcevic, A.-M. Raberger, M. Zeppelzauer, B. Dumphart und B. Horsak, “On the usefulness of statistical parameter mapping for feature selection in automated gait classification,” in *25th Conference of the European Society of Biomechanics*, Wien: European Society of Biomechanics, 2019, S. 1–2.
- [183] B. B. Afef und M. Limam, “Ensemble feature selection for high dimensional data: a new method and a comparative study,” *Advances in Data Analysis and Classification*, Jg. 12, Nr. 4, S. 937–952, 2018. DOI: 10.1007/s11634-017-0285-y.
- [184] A. Shahrjooihighi, H. Frigui, X. Zhang, X. Wei, B. Shi und A. Trabelsi, “An Ensemble Feature Selection Method for Biomarker Discovery,” in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, New York, NY: IEEE, 2017, S. 416–421. DOI: 10.1109/ISSPIT.2017.8388679.
- [185] S. Anjomshoae, T. Kampik und K. Främling, “Py-CIU: A Python Library for Explaining Machine Learning Predictions Using Contextual Importance and Utility,” in *IJCAI-PRICAI 2020 Workshop on Explainable Artificial Intelligence (XAI)*, Yokohama, Japan: IJCAI, 2020, S. 1–5.
- [186] C. Dindorf, J. Konradi, C. Wolf u. a., “Visualization of interindividual differences in spinal dynamics in the presence of intraindividual variabilities,” in *LEVIA ’22: Leipzig Symposium on Visualization in Applications 2022*, C. Gillmann, J. Schmidt, S. Jänicke und D. Wiegrefe, Hrsg., Leipzig: Universität Leipzig, in press.
- [187] D. Castelvechi, “Can we open the black box of AI?” *Nature*, Jg. 538, Nr. 7623, S. 20–23, 2016. DOI: 10.1038/538020a.
- [188] W. Knight, *The Dark Secret at the Heart of AI*, 11. 04. 2017. Adresse: <https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/>.
- [189] D. Weinberger, *Our Machines Now Have Knowledge We’ll Never Understand*, 4.10.2017. Adresse: <https://www.wired.com/story/our-machines-now-have-knowledge-well-never-understand/>.
- [190] G. G. Graham und K. F. Scott, “Mechanism of action of paracetamol,” *American journal of therapeutics*, Jg. 12, Nr. 1, S. 46–55, 2005. DOI: 10.1097/00045391-200501000-00008.
- [191] S. Tong, Y. Fu und H. Ling, “Verification-based pairwise gait identification,” in *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Piscataway, NJ: IEEE, 2017, S. 669–673. DOI: 10.1109/ICMEW.2017.8026299.

- [192] O. Ludwig, S. Dillinger und F. Marschall, “Intra-individual gait pattern variability in specific situations: Implications for forensic gait analysis,” *Forensic Science International*, Jg. 264, S. 15–23, 2016. DOI: 10.1016/j.forsciint.2016.02.043.
- [193] F. Zhao, J. Xu und Y. Lin, “Similarity Measure for Patients via A Siamese CNN Network,” in *Algorithms and Architectures for Parallel Processing*, J. Vaidya und J. Li, Hrsg., Cham: Springer, 2018, S. 319–328. DOI: 10.1007/978-3-030-05054-2_25.
- [194] M. D. Li, K. Chang, B. Bearce u. a., “Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging,” *npj Digital Medicine*, Jg. 3, Nr. 48, S. 1–9, 2020. DOI: 10.1038/s41746-020-0255-1.
- [195] N. Bhagwat, J. D. Viviano, A. N. Voineskos und M. M. Chakravarty, “Modeling and prediction of clinical symptom trajectories in Alzheimer’s disease using longitudinal data,” *PLOS Computational Biology*, Jg. 14, Nr. 9, S. 1–25, 2018. DOI: 10.1371/journal.pcbi.1006376.
- [196] Y. Liu, X. Sui, C. Liu, X. Kuang und Y. Hu, “Automatic Lumbar Spine Tracking Based on Siamese Convolutional Network,” *Journal of Digital Imaging*, Jg. 33, Nr. 2, S. 423–430, 2020. DOI: 10.1007/s10278-019-00273-5.
- [197] C.-F. Liu, S. Padhy, S. Ramachandran u. a., “Using deep Siamese neural networks for detection of brain asymmetries associated with Alzheimer’s Disease and Mild Cognitive Impairment,” *Magnetic Resonance Imaging*, Jg. 64, S. 190–199, 2019. DOI: 10.1016/j.mri.2019.07.003.
- [198] K. H. Park, E. Batbaatar, Y. Piao, N. Theera-Umpon und K. H. Ryu, “Deep Learning Feature Extraction Approach for Hematopoietic Cancer Subtype Classification,” *International Journal of Environmental Research and Public Health*, Jg. 18, Nr. 4, S. 1–24, 2021. DOI: 10.3390/ijerph18042197.
- [199] A. M. Asher, E. R. Oleisky, J. S. Pennings u. a., “Measuring clinically relevant improvement after lumbar spine surgery: is it time for something new?” *The Spine Journal*, Jg. 20, Nr. 6, S. 847–856, 2020. DOI: 10.1016/j.spinee.2020.01.010.
- [200] A. Lucieri, M. N. Bajwa, S. Alexander Braun, M. I. Malik, A. Dengel und S. Ahmed, “On Interpretability of Deep Learning based Skin Lesion Classifiers using Concept Activation Vectors,” in *2020 International Joint Conference on Neural Networks (IJCNN)*, New York, NY: IEEE, 2020, S. 1–10. DOI: 10.1109/IJCNN48605.2020.9206946.
- [201] S. Chopra und K. R. Kaufman, “Effects of Total Hip Arthroplasty on Gait,” in *Handbook of Human Motion*, B. Müller und S. Wolf, Hrsg., Cham: Springer, 2018, S. 1–15. DOI: 10.1007/978-3-319-30808-1_81-1.
- [202] A. L. Beam und I. S. Kohane, “Big Data and Machine Learning in Health Care,” *JAMA*, Jg. 319, Nr. 13, S. 1317–1318, 2018. DOI: 10.1001/jama.2017.18391.
- [203] X. Yi, E. Walia und P. Babyn, “Generative adversarial network in medical imaging: A review,” *Medical Image Analysis*, Jg. 58, S. 1–24, 2019. DOI: 10.1016/j.media.2019.101552.
- [204] P. Terrier, “Gait Recognition via Deep Learning of the Center-of-Pressure Trajectory,” *Applied Sciences*, Jg. 10, Nr. 3, S. 1–20, 2020. DOI: 10.3390/app10030774.

- [205] G. Figueroa-Mata und E. Mata-Montero, “Using a Convolutional Siamese Network for Image-Based Plant Species Identification with Small Datasets,” *Biomimetics (Basel, Switzerland)*, Jg. 5, Nr. 1, S. 1–17, 2020. DOI: 10.3390/biomimetics5010008.
- [206] U. Côté-Allard, E. Campbell, A. Phinyomark, F. Laviolette, B. Gosselin und E. Scheme, “Interpreting Deep Learning Features for Myoelectric Control: A Comparison With Handcrafted Features,” *Frontiers in Bioengineering and Biotechnology*, Jg. 8, S. 1–22, 2020. DOI: 10.3389/fbioe.2020.00158.
- [207] Y. Sano, Y. Yin, T. Mizuguchi und A. Kandori, “Detection of Abnormal Segments in Finger Tapping Waveform using One-class SVM,” in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, New York, NY: IEEE, 2019, S. 1378–1381. DOI: 10.1109/EMBC.2019.8856598.
- [208] C. A. T. Ferro und T. E. Fricker, “A bias-corrected decomposition of the Brier score,” *Quarterly Journal of the Royal Meteorological Society*, Jg. 138, Nr. 668, S. 1954–1960, 2012. DOI: 10.1002/qj.1924.
- [209] J. C. Platt, “Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods,” in *Advances in Large Margin Classifiers*, A. Smola, P. Bartlett, B. Schölkopf und D. Schuurmans, Hrsg., Cambridge, MA: MIT Press, 1999, S. 61–74.
- [210] B. Böken, “On the appropriateness of Platt scaling in classifier calibration,” *Information Systems*, Jg. 95, S. 1–16, 2021. DOI: 10.1016/j.is.2020.101641.
- [211] L. Yang und A. Shami, “On hyperparameter optimization of machine learning algorithms: Theory and practice,” *Neurocomputing*, Jg. 415, S. 295–316, 2020. DOI: 10.1016/j.neucom.2020.07.061.
- [212] M. Claesen und B. de Moor, *Hyperparameter Search in Machine Learning*, 6.04.2015. Adresse: <https://arxiv.org/pdf/1502.02127>.
- [213] M. Aly, “Survey on multiclass classification methods,” *Technical Report*, Jg. 11, S. 1–9, 2005.
- [214] K. Hempstalk und E. Frank, “Discriminating Against New Classes: One-class versus Multi-class Classification,” in *AI 2008. Advances in artificial intelligence : 21st Australasian Joint Conference on Artificial Intelligence*, W. Wobcke und M. Zhang, Hrsg., Berlin: Springer, 2008, S. 325–336. DOI: 10.1007/978-3-540-89378-3_32.
- [215] J. Gao, H. Cheng und P.-N. Tan, “Semi-supervised outlier detection,” in *Proceedings of the 2006 ACM Symposium on Applied Computing*, H. M. Haddad, Hrsg., New York, NY: ACM Press, 2006, S. 635–636. DOI: 10.1145/1141277.1141421.
- [216] V. J. Hodge und J. Austin, “A Survey of Outlier Detection Methodologies,” *Artificial Intelligence Review*, Jg. 22, Nr. 2, S. 85–126, 2004. DOI: 10.1007/s10462-004-4304-y.
- [217] J. Zhang, “Advancements of Outlier Detection: A Survey,” *ICST Transactions on Scalable Information Systems*, Jg. 13, Nr. 1, S. 1–26, 2013. DOI: 10.4108/trans.sis.2013.01-03.e2.

- [218] M. E. Villa-Pérez, M. Á. Álvarez-Carmona, O. Loyola-González, M. A. Medina-Pérez, J. C. Velazco-Rossell und K.-K. R. Choo, “Semi-supervised anomaly detection algorithms: A comparative summary and future research directions,” *Knowledge-Based Systems*, Jg. 218, S. 1–18, 2021. DOI: 10.1016/j.knosys.2021.106878.
- [219] K. A. Zweig, *Ein Algorithmus hat kein Taktgefühl: Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können*, 1. Auflage. München: ABOD Verlag, 2019.
- [220] J. Wapner, *Cancer Scientists Have Ignored African DNA in the Search for Cures*, 18.07.18. Adresse: <https://www.newsweek.com/2018/07/27/cancer-cure-genome-cancer-treatment-africa-genetic-charles-rotimi-dna-human-1024630.html>.
- [221] T. Dietterich, “Overfitting and undercomputing in machine learning,” *ACM Computing Surveys*, Jg. 27, Nr. 3, S. 326–327, 1995. DOI: 10.1145/212094.212114.
- [222] B. Neal, S. Mittal, A. Baratin u. a., *A Modern Take on the Bias-Variance Tradeoff in Neural Networks*, 18.12.2019. Adresse: <https://arxiv.org/pdf/1810.08591>.
- [223] R. Caruana, S. Lawrence und L. Giles, “Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping,” in *NIPS’00: Proceedings of the 13th International Conference on Neural Information Processing Systems*, T. K. Leen, T. G. Dietterich und V. Tresp, Hrsg., Cambridge, MA: MIT Press, 2000, S. 1–7.
- [224] X. Ying, “An Overview of Overfitting and its Solutions,” *Journal of Physics: Conference Series*, Jg. 1168, S. 1–7, 2019. DOI: 10.1088/1742-6596/1168/2/022022.
- [225] R. Moradi, R. Berangi und B. Minaei, “A survey of regularization strategies for deep models,” *Artificial Intelligence Review*, Jg. 53, Nr. 6, S. 3947–3986, 2020. DOI: 10.1007/s10462-019-09784-7.
- [226] S. Rose, “Machine Learning for Prediction in Electronic Health Data,” *JAMA Network Open*, Jg. 1, Nr. 4, S. 1–3, 2018. DOI: 10.1001/jamanetworkopen.2018.1404.
- [227] T. Torsney-Weir, M. Sedlmair und T. Möller, “Decision making in uncertainty visualization,” in *Workshop on Visualization for decision making under uncertainty 2015*, Chicago, IL: IEEE, 2015, S. 1–4. DOI: 10.6084/M9.FIGSHARE.1585848.
- [228] P. Levontin und J. L. Walton, *Visualising Uncertainty: A short introduction*. London, UK: Sad Press, 2020.
- [229] G. Ristovski, T. Preusser, H. K. Hahn und L. Linsen, “Uncertainty in medical visualization: Towards a taxonomy,” *Computers & Graphics*, Jg. 39, S. 60–73, 2014. DOI: 10.1016/j.cag.2013.10.015.
- [230] C. Gillmann, D. Saur, T. Wischgoll und G. Scheuermann, “Uncertainty-aware Visualization in Medical Imaging – A Survey,” *Computer Graphics Forum*, Jg. 40, Nr. 3, S. 665–689, 2021. DOI: 10.1111/cgf.14333.
- [231] B. Pang, E. Nijkamp und Y. N. Wu, “Deep Learning With TensorFlow: A Review,” *Journal of Educational and Behavioral Statistics*, Jg. 45, Nr. 2, S. 227–248, 2020. DOI: 10.3102/1076998619872761.

- [232] A. D. Kiureghian und O. Ditlevsen, “Aleatory or epistemic? Does it matter?” *Structural Safety*, Jg. 31, Nr. 2, S. 105–112, 2009. DOI: 10.1016/j.strusafe.2008.06.020.
- [233] A. Molak, *Modeling uncertainty in neural networks with TensorFlow Probability*, 26.11.2021. Adresse: <https://towardsdatascience.com/modeling-uncertainty-in-neural-networks-with-tensorflow-probability-391b29538a7a>.
- [234] D. A. Hoffman, J. E. Lonstein, M. M. Morin, W. Visscher, B. S. Harris und J. D. Boice, “Breast cancer in women with scoliosis exposed to multiple diagnostic x rays,” *Journal of the National Cancer Institute*, Jg. 81, Nr. 17, S. 1307–1312, 1989. DOI: 10.1093/jnci/81.17.1307.
- [235] J. H. Hollman, M. K. Watkins, A. C. Imhoff, C. E. Braun, K. A. Akervik und D. K. Ness, “A comparison of variability in spatiotemporal gait parameters between treadmill and overground walking conditions,” *Gait & Posture*, Jg. 43, S. 204–209, 2016. DOI: 10.1016/j.gaitpost.2015.09.024.
- [236] A. Phinyomark, S. Osis, B. A. Hettinga und R. Ferber, “Kinematic gait patterns in healthy runners: A hierarchical cluster analysis,” *Journal of Biomechanics*, Jg. 48, Nr. 14, S. 3897–3904, 2015. DOI: 10.1016/j.jbiomech.2015.09.025.
- [237] S. Hochreiter und J. Schmidhuber, “Long short-term memroy,” *Neural Computation*, Jg. 9, Nr. 8, S. 1735–1780, 1997.
- [238] K. Cho, van Merriënboer, Bart, C. Gulcehre u. a., *Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation*, 2014. Adresse: <https://arxiv.org/pdf/1406.1078.pdf>.
- [239] R. Kruse, C. Borgelt, C. Braune, F. Klawonn, C. Moewes und M. Steinbrecher, *Computational Intelligence. Eine methodische Einführung in Künstliche Neuronale Netze, Evolutionäre Algorithmen, Fuzzy-Systeme und Bayes-Netze*. Wiesbaden: Springer, 2015. DOI: 10.1007/978-3-658-10904-2.
- [240] Y. Yu, X. Si, C. Hu und J. Zhang, “A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures,” *Neural Computation*, Jg. 31, Nr. 7, S. 1235–1270, 2019. DOI: 10.1162/neco_a_01199.
- [241] T. O. Zhiguang Wang, “Encoding Time Series as Images for Visual Inspection and Classification Using Tiled Convolutional Neural Networks,” in *29th AAAI Conference on Artificial Intelligence*, Palo Alto, California: AAAI Press, 2015, S. 40–47.
- [242] C.-L. Yang, Z.-X. Chen und C.-Y. Yang, “Sensor Classification Using Convolutional Neural Network by Encoding Multivariate Time Series as Two-Dimensional Colored Images,” *Sensors (Basel, Switzerland)*, Jg. 20, Nr. 1, S. 1–15, 2019. DOI: 10.3390/s20010168.
- [243] F. Hutter, L. Kotthoff und J. Vanschoren, *Automated Machine Learning. Methods, Systems, Challenges*. Cham, Switzerland: Springer, 2019. DOI: 10.1007/978-3-030-05318-5.
- [244] K. Chauhan, S. Jani, D. Thakkar u. a., “Automated Machine Learning: The New Wave of Machine Learning,” in *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, New York, NY: IEEE, 2020, S. 205–212. DOI: 10.1109/ICIMIA48430.2020.9074859.

- [245] B. Preim und D. Bartz, *Visualisation in medicine: Theory, algorithms, and applications*. San Francisco, Calif.: Morgan Kaufmann, 2007.
- [246] R. Koprowski, “Automatic analysis of the trunk thermal images from healthy subjects and patients with faulty posture,” *Computers in Biology and Medicine*, Jg. 62, S. 110–118, 2015. DOI: 10.1016/j.compbimed.2015.04.017.
- [247] C. Magalhaes, J. Mendes und R. Vardasca, “Meta-Analysis and Systematic Review of the Application of Machine Learning Classifiers in Biomedical Applications of Infrared Thermography,” *Applied Sciences*, Jg. 11, Nr. 2, S. 1–18, 2021. DOI: 10.3390/app11020842.
- [248] M. Shafique, M. Naseer, T. Theocharides u. a., “Robust Machine Learning Systems: Challenges, Current Trends, Perspectives, and the Road Ahead,” *IEEE Design & Test*, Jg. 37, Nr. 2, S. 30–57, 2020. DOI: 10.1109/MDAT.2020.2971217.
- [249] E. Wong und J. Z. Kolter, *Learning perturbation sets for robust machine learning*, 8.10.2020. Adresse: <https://arxiv.org/pdf/2007.08450>.
- [250] A. Qayyum, J. Qadir, M. Bilal und A. Al-Fuqaha, “Secure and Robust Machine Learning for Healthcare: A Survey,” *IEEE Reviews in Biomedical Engineering*, Jg. 14, S. 156–180, 2021. DOI: 10.1109/RBME.2020.3013489.
- [251] S. Qiu, Q. Liu, S. Zhou und C. Wu, “Review of Artificial Intelligence Adversarial Attack and Defense Technologies,” *Applied Sciences*, Jg. 9, Nr. 5, S. 1–29, 2019. DOI: 10.3390/app9050909.
- [252] S. G. Finlayson, H. W. Chung, I. S. Kohane und A. L. Beam, *Adversarial Attacks Against Medical Deep Learning Systems*, 4.02.2019. Adresse: <https://arxiv.org/pdf/1804.05296>.
- [253] G. R. Karpagam, B. V. Kumar, J. U. Maheswari und X.-Z. Gao, *Smart cyber physical systems: Advances, challenges and opportunities*. Boca Raton: Chapman & Hall, 2020.
- [254] U. Shaham, Y. Yamada und S. Negahban, “Understanding Adversarial Training: Increasing Local Stability of Neural Nets through Robust Optimization,” *Neurocomputing*, Jg. 307, S. 195–204, DOI: 10.1016/j.neucom.2018.04.027.
- [255] F. Khalid, M. A. Hanif, S. Rehman, J. Qadir und M. Shafique, *FAdeML: Understanding the Impact of Pre-Processing Noise Filtering on Adversarial Machine Learning*, 4.11.2018. Adresse: <https://arxiv.org/pdf/1811.01444>.
- [256] M. Delange, R. Aljundi, M. Masana u. a., “A continual learning survey: Defying forgetting in classification tasks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, S. 1–20, 2021 (published online). DOI: 10.1109/TPAMI.2021.3057446.
- [257] F. Zenke, B. Poole und S. Ganguli, *Continual Learning Through Synaptic Intelligence*, 12.06.2017. Adresse: <https://arxiv.org/pdf/1703.04200>.
- [258] T. Diethe, T. Borchert, E. Thereska, B. Balle und N. Lawrence, *Continual Learning in Practice*, 2019. Adresse: <https://arxiv.org/pdf/1903.05202>.
- [259] C. S. Lee und A. Y. Lee, “Clinical applications of continual learning machine learning,” *The Lancet Digital Health*, Jg. 2, Nr. 6, e279–e281, 2020. DOI: 10.1016/S2589-7500(20)30102-3.

- [260] K. N. Vokinger, S. Feuerriegel und A. S. Kesselheim, “Continual learning in medical devices: FDA’s action plan and beyond,” *The Lancet Digital Health*, Jg. 3, Nr. 6, e337–e338, 2021. DOI: 10.1016/S2589-7500(21)00076-5.
- [261] I. Ract, J.-M. Meadeb, G. Mercy, F. Cuff, J.-L. Husson und R. Guillin, “A review of the value of MRI signs in low back pain,” *Diagnostic and Interventional Imaging*, Jg. 96, Nr. 3, S. 239–249, 2015. DOI: 10.1016/j.diii.2014.02.019.
- [262] K. H. E. Beange, A. D. C. Chan, S. M. Beaudette und R. B. Graham, “Concurrent validity of a wearable IMU for objective assessments of functional movement quality and control of the lumbar spine,” *Journal of Biomechanics*, Jg. 97, S. 1–25, 2019. DOI: 10.1016/j.jbiomech.2019.109356.
- [263] L. Franco, R. Sengupta, L. Wade und D. Cazzola, “A novel IMU-based clinical assessment protocol for Axial Spondyloarthritis: a protocol validation study,” *PeerJ*, Jg. 9, S. 1–29, 2021. DOI: 10.7717/peerj.10623.
- [264] L. Simpson, M. M. Maharaj und R. J. Mobbs, “The role of wearables in spinal posture analysis: a systematic review,” *BMC Musculoskeletal Disorders*, Jg. 20, Nr. 55, S. 1–14, 2019. DOI: 10.1186/s12891-019-2430-6.
- [265] A. Sheikhtaheri, F. Sadoughi und Z. Hashemi Dehaghi, “Developing and using expert systems and neural networks in medicine: a review on benefits and challenges,” *Journal of Medical Systems*, Jg. 38, Nr. 110, S. 1–6, 2014. DOI: 10.1007/s10916-014-0110-5.
- [266] J. C. Schmidt, “Wandel und Kontinuität von Wissenschaft durch KI. Zur aktuellen Veränderung des Wissenschafts- und Technikverständnisses,” in *Künstliche Intelligenz in der Forschung*, C. F. Gethmann, P. Buxmann, J. Distelrath u. a., Hrsg., Berlin, Germany: Springer, 2022, S. 79–125.
- [267] L. Zhou, E. Fischer, C. Tunca u. a., “How We Found Our IMU: Guidelines to IMU Selection and a Comparison of Seven IMUs for Pervasive Healthcare Applications,” *Sensors (Basel, Switzerland)*, Jg. 20, Nr. 15, S. 1–29, 2020. DOI: 10.3390/s20154090.
- [268] P. Kieseberg, E. Weippl und A. Holzinger, “Trust for the “Doctor in the Loop”,” *ERCIM News*, Jg. 104, S. 32–33, 2016.
- [269] A. Verghese, N. H. Shah und R. A. Harrington, “What This Computer Needs Is a Physician: Humanism and Artificial Intelligence,” *JAMA*, Jg. 319, Nr. 1, S. 19–20, 2018. DOI: 10.1001/jama.2017.19198.
- [270] A. George, D. Blaauw, J. Thompson und L. Green-Thompson, “Doctor retention and distribution in post-apartheid South Africa: tracking medical graduates (2007-2011) from one university,” *Human Resources for Health*, Jg. 17, Nr. 1, S. 1–9, 2019. DOI: 10.1186/s12960-019-0439-4.
- [271] C. Chu, J. de Fauw, N. Tomasev u. a., “Applying machine learning to automated segmentation of head and neck tumour volumes and organs at risk on radiotherapy planning CT and MRI scans,” *F1000Research*, Jg. 5, S. 1–8, 2016. DOI: 10.12688/f1000research.9525.1.
- [272] R. Bartlett, “Artificial Intelligence in Sports Biomechanics: New Dawn or False Hope?” *Journal of Sports Science & Medicine*, Jg. 5, Nr. 4, S. 474–479, 2006.

- [273] M. Zago, A. F. R. Kleiner und P. A. Federolf, “Editorial: Machine Learning Approaches to Human Movement Analysis,” *Frontiers in Bioengineering and Biotechnology*, Jg. 8, S. 1–3, 2020. DOI: 10.3389/fbioe.2020.638793.
- [274] F.-W. Zou, Y.-F. Tang, C.-Y. Liu, J.-A. Ma und C.-H. Hu, “Concordance Study Between IBM Watson for Oncology and Real Clinical Practice for Cervical Cancer Patients in China: A Retrospective Analysis,” *Frontiers in Genetics*, Jg. 11, S. 1–8, 2020. DOI: 10.3389/fgene.2020.00200.
- [275] C. Ross und I. Swetlitz, *IBM’s Watson supercomputer recommended ‘unsafe and incorrect’ cancer treatments, internal documents show*, 25.07.2018. Adresse: <https://www.statnews.com/wp-content/uploads/2018/09/ibms-watson-recommended-unsafe-and-incorrect-cancer-treatments-stat.pdf>.
- [276] C. Dindorf, O. Ludwig, J. Konradi u. a., “Modellierung von Beckenstellungen auf Basis von Wirbelsäulenparametern,” *German Journal of Sports Medicine*, Jg. 72, Nr. 3, S. 128, 2021.
- [277] C. Dindorf, W. Teufl, B. Taetz, G. Bleser und M. Fröhlich, “Interpretation von Modellen des Maschinellen Lernens mittels Explainable Artificial Intelligence bei Patienten nach Hüft-Total-Endoprothetik,” in *#Sport #Gesundheit #Digital*, M. Sprenger, C. Dindorf, S. Defren, B. Steinke und M. Fröhlich, Hrsg., Kaiserslautern: Technische Universität Kaiserslautern, 2020, S. 214–217.
- [278] C. Dindorf, W. Teufl, B. Taetz, S. Becker, G. Bleser und M. Fröhlich, “Vergleichende Feature-Extraktion und Gangklassifizierung bei Hüftgelenkersatz-Patienten auf Basis kinematischer Trajektorien,” in *„Bewegungsanalyse: Klinik – Alltag – Sport“ III. Gamma-Kongress 2020 München*, Technische Universität München, Hrsg., München: Technische Universität München, 2020, S. 41–42.

Liste der Publikationen der Dissertation

1. **Publikation 1:** C. Dindorf, W. Teufl, B. Taetz u. a., “Feature extraction and gait classification in hip replacement patients on the basis of kinematic waveform data,” *Biomedical Human Kinetics*, Jg. 13, Nr. 1, S. 177–186, 2021. DOI: 10.2478/bhk-2021-0022
2. **Publikation 2:** C. Dindorf, J. Konradi, C. Wolf u. a., “General method for automated feature extraction and selection and its application for gender classification and biomechanical knowledge discovery of sex differences in spinal posture during stance and gait,” *Computer Methods in Biomechanics and Biomedical Engineering*, Jg. 24, Nr. 3, S. 1–9, 2020. DOI: 10.1080/10255842.2020.1828375
3. **Publikation 3:** C. Dindorf, W. Teufl, B. Taetz u. a., “Interpretability of Input Representations for Gait Classification in Patients after Total Hip Arthroplasty,” *Sensors*, Jg. 20, Nr. 16, S. 1–14, 2020. DOI: 10.3390/s20164385
4. **Publikation 4:** C. Dindorf, J. Konradi, C. Wolf u. a., “Machine learning techniques demonstrating individual movement patterns of the vertebral column: The fingerprint of spinal motion,” *Computer Methods in Biomechanics and Biomedical Engineering*, S. 1–11, 2021 (published online). DOI: 10.1080/10255842.2021.1981884
5. **Vorbereitende Publikation für Publikation 5:** W. Teufl, B. Taetz, M. Miezal u. a., “Automated detection of pathological gait patterns using a one-class support vector machine trained on discrete parameters of IMU based gait data,” *Clinical Biomechanics*, Jg. 89, S. 1–7, 2021. DOI: 10.1016/j.clinbiomech.2021.105452¹
6. **Publikation 5:** C. Dindorf, J. Konradi, C. Wolf u. a., “Classification and Automated Interpretation of Spinal Posture Data Using a Pathology-Independent Classifier and Explainable Artificial Intelligence (XAI),” *Sensors*, Jg. 21, Nr. 18, S. 1–18, 2021. DOI: 10.3390/s21186323

¹ Der Autor hat hier mitgewirkt und steht an vierter Stelle. Auf dieser Publikation baut Publikation 5 auf und erweitert diese.

Ausblick auf weitere bzw. anstehende Publikationen in Zusammenhang mit der Dissertationsthematik

1. C. Dindorf, E. Bartaguiz, E. Janowicz u. a., “Effects of Unilateral Muscle Fatigue on Thermographic Skin Surface Temperature of Back and Abdominal Muscles—A Pilot Study,” *Sports (Basel, Switzerland)*, Jg. 10, Nr. 3, S. 1–10, 2022. DOI: 10.3390/sports10030041
2. E. Bartaguiz, C. Dindorf, J. Dully u. a., “Is it necessary to reconsider bike fitting practice? On the influence of different intensity levels and treatment-introduced fatigue on the biomechanics of elite cyclists,” under review

Liste der Kongressbeiträge der Dissertation

1. C. Dindorf, J. Konradi, C. Wolf u. a., “Visualization of interindividual differences in spinal dynamics in the presence of intraindividual variabilities,” in *LEVIA '22: Leipzig Symposium on Visualization in Applications 2022*, C. Gillmann, J. Schmidt, S. Jänicke u. a., Hrsg., Leipzig: Universität Leipzig, in press
2. C. Dindorf, J. Konradi, C. Wolf u. a., “Towards a better understanding of spinal differences between healthy subjects and subjects with back pain using Explainable Artificial Intelligence (XAI),” in *9th International Performance Analysis Workshop and Conference & 5th IACSS Conference*, Universität Wien, Hrsg., Wien: Universität Wien, 2021
3. C. Dindorf, O. Ludwig, J. Konradi u. a., “Modellierung von Beckenstellungen auf Basis von Wirbelsäulenparametern,” *German Journal of Sports Medicine*, Jg. 72, Nr. 3, S. 128, 2021
4. C. Dindorf, W. Teufl, B. Taetz u. a., “Interpretation von Modellen des Maschinellen Lernens mittels Explainable Artificial Intelligence bei Patienten nach Hüft-Total-Endoprothetik,” in *#Sport #Gesundheit #Digital*, M. Sprenger, C. Dindorf, S. Defren u. a., Hrsg., Kaiserslautern: Technische Universität Kaiserslautern, 2020, S. 214–217
5. C. Dindorf, W. Teufl, B. Taetz u. a., “Vergleichende Feature-Extraktion und Gangklassifizierung bei Hüftgelenkersatz-Patienten auf Basis kinematischer Trajektorien,” in *„Bewegungsanalyse: Klinik – Alltag – Sport“ III. Gamma-Kongress 2020 München*, Technische Universität München, Hrsg., München: Technische Universität München, 2020, S. 41–42

KAPITEL 8

Curriculum Vitae

Carlo Dindorf

Geboren: 08.03.1996 in Bad Kreuznach
 Adresse: Erwin-Schrödinger-Straße 57/532,
 67663 Kaiserslautern
 Telefon: +49 631 205 5172
 E-Mail: carlo.dindorf@sowi.uni-kl.de



Vita	
2019 bis 2022	Promotion unter dem Titel „Auf dem Weg zu interpretierbaren, KI-basierten, explorativen und entscheidungsunterstützenden Systemen im Kontext biomechanischer Daten“, Abschluss Dr. rer. nat.
Seit 2021	Leitung Ergometrie Labor, Arbeitsbereich Bewegungs- und Trainingswissenschaft, TU Kaiserslautern
Seit 2020	Co-Leitung Ergometrie Labor, Arbeitsbereich Bewegungs- und Trainingswissenschaft, TU Kaiserslautern
2020	Kongressdurchführung und -organisation des digitalen Kongresses #Sport #Gesundheit #Digital an der TUK
Seit 2019	Wissenschaftlicher Mitarbeiter an der TU Kaiserslautern im Fachgebiet Sportwissenschaft im Arbeitsbereich Bewegungs- und Trainingswissenschaft
2016 bis 2019	Studentische/Wissenschaftliche Hilfskraft im Fachgebiet Sportwissenschaft der TU Kaiserslautern
2017 bis 2018	Masterstudium in den Fächern Sport, Chemie und Bildungswissenschaften an der TU Kaiserslautern, Abschluss M.Ed.
2014 bis 2017	Bachelorstudium in den Fächern Sport, Chemie und Bildungswissenschaften an der TU Kaiserslautern, Abschluss B.Ed.

Auszeichnungen	
2020	1. Platz Wissenschaftspreis Sportbund Pfalz 2020, Kategorie Abschlussarbeiten

Monographien und Herausgeberwerke	
1.	Dindorf, C. (2022). <i>Auf dem Weg zu interpretierbaren, KI-basierten, explorativen und entscheidungsunterstützenden Systemen im Kontext biomechanischer Daten. Kumulative Dissertationsschrift zur Erlangung des Grades eines Doktors der Naturwissenschaften.</i> Technische Universität Kaiserslautern.
2.	Sprenger, M., Dindorf, C., Defren, S., Steinke, B., & Fröhlich, M. (Hrsg.) (2020). <i>#Sport #Gesundheit #Digital: Der Kongress zu Chancen und Risiken der Digitalisierung in Sport und Gesundheit.</i> Technische Universität Kaiserslautern.

Peer-Reviewed Journal Artikel	
	Bartaguiz, E., Dindorf, C., Janowicz, E., Fröhlich, M., & Ludwig, O. (in preparation). Effects of asymmetric muscle fatigue on pelvic inclination and rotation.
	Dindorf, C., Thomas, A., Bartaguiz, E., Sprenger, M., & Fröhlich, M. (under review). Effekte einer Kletterexkursion im hochschulischen Ausbildungskonzept während der Corona-Pandemie auf das psychische Wohlbefinden.
	Bartaguiz, E., Dindorf, C., Dully, J., Becker, S., & Fröhlich, M. (under review). Is it necessary to reconsider bike fitting practice? On the influence of different intensity levels and treatment-introduced fatigue on the biomechanics of elite cyclists.
1.	Dindorf, C., Bartaguiz, E., Janowicz, E., Fröhlich, M., & Ludwig, O. (2022). Effects of Unilateral Muscle Fatigue on Thermographic Skin Surface Temperature of Back and Abdominal Muscles—A Pilot Study. <i>Sports (Basel, Switzerland)</i> , 10(3), 1–10. https://doi.org/10.3390/sports10030041
2.	Dindorf, C., Konradi, J., Wolf, C., Taetz, B., Bleser, G., Huthwelker, J., Werthmann, F., Bartaguiz, E., Kniepert, J., Drees, P., Betz, U., & Fröhlich, M. (2021). Classification and Automated Interpretation of Spinal Posture Data Using a Pathology-Independent Classifier and Explainable Artificial Intelligence (XAI). <i>Sensors</i> , 21(18), 1–18. https://doi.org/10.3390/s21186323
3.	Dindorf, C., Konradi, J., Wolf, C., Taetz, B., Bleser, G., Huthwelker, J., Werthmann, F., Drees, P., Fröhlich, M., & Betz, U. (2021, published online). Machine learning techniques demonstrating individual movement patterns of the vertebral column: The fingerprint of spinal motion. <i>Computer Methods in Biomechanics and Biomedical Engineering</i> , 1–11. https://doi.org/10.1080/10255842.2021.1981884
4.	Teufl, W., Taetz, B., Miezal, M., Dindorf, C., Fröhlich, M., Trinler, U., Hogam, A., & Bleser, G. (2021). Automated detection of pathological gait patterns using a one-class support vector machine trained on discrete parameters of IMU based gait data. <i>Clinical Biomechanics</i> , 89, 1–7. https://doi.org/10.1016/j.clinbiomech.2021.105452
5.	Dindorf, C., Teufl, W., Taetz, B., Becker, S., Bleser, G., & Fröhlich, M. (2021). Feature extraction and gait classification in hip replacement patients on the basis of kinematic waveform data. <i>Biomedical Human Kinetics</i> , 13(1), 177–186. https://doi.org/10.2478/bhk-2021-0022
6.	Dindorf, C., Konradi, J., Wolf, C., Taetz, B., Bleser, G., Huthwelker, J., Drees, P., Fröhlich, M., & Betz, U. (2020). General method for automated feature extraction and selection and its application for gender classification and biomechanical knowledge discovery of sex differences in spinal posture during stance and gait. <i>Computer Methods in Biomechanics and Biomedical Engineering</i> , 24(3), 1–9. https://doi.org/10.1080/10255842.2020.1828375
7.	Dindorf, C., Teufl, W., Taetz, B., Bleser, G., & Fröhlich, M. (2020). Interpretability of Input Representations for Gait Classification in Patients after Total Hip Arthroplasty. <i>Sensors</i> , 20(16), 1–14. https://doi.org/10.3390/s20164385
8.	Dindorf, C., & Fröhlich, M. (2020). Zur Konnotation und Denotation des Trainingsbegriffs in Theorie und Praxis des Sports. <i>German Journal of Exercise and Sport Research</i> , 50(2), 297–307. https://doi.org/10.1007/s12662-019-00641-x

Abstracts, Proceedings, Tagungs-/Symposiumsbeiträge (mit Peer Review-Verfahren)	
	Obentheuer, M., Harant, M., Bartaguiz, E., Dindorf, C., Linn, J., & Fröhlich, M. (in press). Muscle torque generators for digital human model control - measurement protocol for data acquisition. In <i>27th Congress of the European Society of Biomechanics</i> , Porto, Portugal.
1.	Dindorf, C., Konradi, J., Wolf, C., Taetz, B., Bleser, G., Huthwelker, J., Werthmann, F., Bartaguiz, E., Drees, P., Betz, U., & Fröhlich, M. (in press). Visualization of interindividual differences in spinal dynamics in the presence of intraindividual variabilities. In C. Gillmann, J. Schmidt, S. Jänicke & D. Wiegrefe (Hrsg.), <i>LEVIA'22: Leipzig Symposium on Visualization in Applications 2022</i> . Universität Leipzig.
2.	Dindorf, C., Konradi, J., Wolf, C., Taetz, B., Bleser, G., Kniepert, J., Drees, P., Fröhlich, M., & Betz, U. (2021). Towards a better understanding of spinal differences between healthy subjects and subjects with back pain

	using Explainable Artificial Intelligence (XAI). In Universität Wien (Hrsg.), <i>9th International Performance Analysis Workshop and Conference & 5th IACSS Conference</i> . Universität Wien.
3.	Dindorf, C., Ludwig, O., Konradi, J., Betz, U., Huthwelker, J., Wolf, C., Bleser, G., Taetz, B., Drees, P., & Fröhlich, M. (2021). Modellierung von Beckenstellungen auf Basis von Wirbelsäulenparametern. <i>German Journal of Sports Medicine</i> , 72(3), 128.
4.	Dindorf, C., Teufl, W., Taetz, B., Bleser, G., & Fröhlich, M. (2020). Interpretation von Modellen des Maschinellen Lernens mittels Explainable Artificial Intelligence bei Patienten nach Hüft-Total-Endoprothetik. In M. Sprenger, C. Dindorf, S. Defren, B. Steinke & M. Fröhlich (Hrsg.), <i>#Sport #Gesundheit #Digital: Der Kongress zu Chancen und Risiken der Digitalisierung in Sport und Gesundheit</i> (S. 214-217). Technische Universität Kaiserslautern.
5.	Dindorf, C., Teufl, W., Taetz, B., Becker, S., Bleser, G., & Fröhlich, M. (2020). Vergleichende Feature-Extraktion und Gangklassifizierung bei Hüftgelenkersatz-Patienten auf Basis kinematischer Trajektorien. In Technische Universität München (Hrsg.), <i>„Bewegungsanalyse: Klinik – Alltag – Sport“ III. Gamma-Kongress 2020 München</i> (S. 41-42). Technische Universität München.
6.	Dindorf, C., Fröhlich, M., & Ludwig, O. (2019). Denotation, connotation and association of the term ‚training‘. <i>German Journal of Exercise and Sport Research</i> , 49(Suppl 1), 14. https://doi.org/10.1007/s12662-019-00567-4

Wissenschaftliche und eingeladene Vorträge (erster Vortragende)	
1.	Dindorf, C., Konradi, J., Wolf, C., Taetz, B., Bleser, G., Huthwelker, J., Werthmann, F., Bartaguiz, E., Drees, P., Betz, U., & Fröhlich, M. (6. April 2022-7. April 2022). <i>Visualization of interindividual differences in spinal dynamics in the presence of intraindividual variabilities</i> . LEVIA'22: Leipzig Symposium on Visualization in Applications 2022, Leipzig.
2.	Dindorf, C., & Bartaguiz, E. (4. November 2021-5. November 2021). <i>Aktuelle Entwicklungen in der rad-sportspezifischen Leistungsdiagnostik. Implikationen für die Praxis</i> . Radsportverband Rheinland-Pfalz, Kaiserslautern.
3.	Dindorf, C., Konradi, J., Wolf, C., Taetz, B., Bleser, G., Bartaguiz, E., Kniepert, J., Drees, P., Fröhlich, M., & Betz, U. (30. August 2021-31. August 2021). <i>Towards a better understanding of spinal differences between healthy subjects and subjects with back pain using Explainable Artificial Intelligence (XAI)</i> . 9th International Performance Analysis Workshop and Conference & 5th IACSS Conference, Wien.
4.	Dindorf, C., Ludwig, O., Konradi, J., Betz, U., Huthwelker, J., Wolf, C., Bleser, G., Taetz, B., Drees, P., & Fröhlich, M. (20. April 2021-24. April 2021). <i>Modellierung von Beckenstellungen auf Basis von Wirbelsäulenparametern</i> . Sports, Medicine and Health Summit 2021, Hamburg.
5.	Dindorf, C., & Defren, S. (18. März 2021). <i>#SGD-Kongress. Die Lessons Learned</i> . Kongress Armut und Gesundheit 2021, Berlin.
6.	Dindorf, C., Teufl, W., Taetz, B., Bleser, G., & Fröhlich, M. (26. November 2020-27. November 2020). <i>Interpretation von Modellen des Maschinellen Lernens mittels Explainable Artificial Intelligence bei Patienten nach Hüft-Total-Endoprothetik</i> . #Sport #Gesundheit #Digital: Der Kongress zu Chancen und Risiken der Digitalisierung in Sport und Gesundheit, Kaiserslautern.
7.	Dindorf, C. (1. Oktober 2020-31. Dezember 2020). <i>Data Mining und Maschinelles Lernen in der Bewegungs- und Trainingswissenschaft</i> . Aufstiegskongress 2020, Saarbrücken.
8.	Dindorf, C., Teufl, W., Taetz, B., Becker, S., Bleser, G., & Fröhlich, M. (6. März 2020-7. März 2020). <i>Vergleichende Feature-Extraktion und Gangklassifizierung bei Hüftgelenkersatz-Patienten auf Basis kinematischer Trajektorien</i> . 3. GAMMA-Kongress, München.
9.	Dindorf, C., Fröhlich, M., & Ludwig, O. (20. Februar 2019-22. Februar 2019). <i>Denotation, connotation and association of the term ‚training‘</i> . German Exercise Science & Training Conference (GEST19) of the German Society of Sport Science (dvs), Würzburg.

Transferaktivitäten - Veranstaltungen	
1.	MEDICA 2021 Fachmesse: Präsentation der Entwicklungen unter dem Titel „Künstliche Intelligenz in der Wirbelsäulendiagnostik (KIWI)“. Düsseldorf, 15.-18.11.2021.
2.	Digitaltag 2021: Interaktiver virtueller Publikumsdialog zum Thema „Digitalisierung in Sport und Gesundheit – ein Blick auf aktuelle Entwicklungen in der digitalen Welt“. Kaiserslautern, 18.06.2021.
3.	8. medtech Rheinland-Pfalz: Vorstellung der entwickelten KI-Ansätze zur Analyse von Wirbelsäuledaten. Vorstellung der Forschungskooperation der Universitätsmedizin der JGU Mainz, Technische Universität Kaiserslautern und DIERS International GmbH. Mainz, 16.06.2021.
4.	Kongress #Sport #Gesundheit #Digital: Organisation und Durchführung des online Kongresses. Kaiserslautern, 26.-27. 11.2020
5.	7. medtech Rheinland-Pfalz: Vorstellung der Forschungskooperation der Universitätsmedizin der JGU Mainz, Technische Universität Kaiserslautern und DIERS International GmbH im Bereich des Einsatzes von künstlicher Intelligenz in der Gesundheitsbranche. Mainz, 12.05.2020.
6.	interaktive Ausstellung „Markt der digitalen Möglichkeiten“: Informationsstand, Ausstellung der Demonstratoren jumpBall und Stressmentor. Kaiserslautern, 12.11.2019.
7.	Zukunftskongress „Gesundheit digital“ der Zukunftsinitiative Rheinland-Pfalz (ZIRP): Informationsstand, Ausstellung der Demonstratoren jumpBall und Stressmentor. Koblenz, 03.-04.09.2019.
8.	Sport- und Fitnesstag der Technischen Universität Kaiserslautern: Mitgestaltung und Publikumsdialog. Kaiserslautern, 25.06.2019.

Transferaktivitäten - Radiointerviews	
1.	mdr Wissen (5. November 2021). <i>Ich hab' Rücken: Da hilft in Zukunft die KI.</i> Abruf unter https://www.mdr.de/wissen/kuenstliche-intelligenz-hilfe-bei-rueckenschmerzen-100.html
2.	Antenne Kaiserslautern (2. Juni 2021). <i>Mit KI gegen Rückenschmerzen.</i> Abruf unter https://soundcloud.com/antenne-kaiserslautern/mit-ki-gegen-ruckenschmerzen

Transferaktivitäten - Ausgewählte Pressemitteilungen	
1.	Computerworld. (17. November 2021). <i>Diagnose bei Rückenproblemen, KI hilft</i> [Press release]. Abruf unter https://www.computerworld.ch/technik/ki/diagnose-rueckenproblemen-ki-hilft-2713503.html
2.	medizin & technik. (10. November 2021). <i>KI für medizinische Diagnose bei Rückenproblemen</i> [Press release]. Abruf unter https://medizin-und-technik.industrie.de/medizin/news-medizin/ki-fuer-medizinische-diagnose-bei-rueckenproblemen/
3.	idw - Informationsdienst Wissenschaft. (2. November 2021). <i>Medica 2021: Verfahren der Künstlichen Intelligenz liefert künftig personalisierte Diagnose bei Rückenproblemen</i> [Press release]. Abruf unter https://idw-online.de/de/news778539
4.	Healthcare Computing. (27. Mai 2021). <i>„Ich habe Rücken“: KI ermöglicht personalisierte Diagnose</i> [Press release]. Abruf unter https://www.healthcare-computing.de/ich-habe-ruecken-ki-ermoeglicht-personalisierte-diagnose-a-1026983/
5.	esatum. (26. Mai 2021). <i>Rückenschmerzen: Künstliche Intelligenz ermöglicht personalisierte Diagnose</i> [Press release]. Abruf unter https://www.esatum.de/today/posts/rueckenschmerzen-kuenstliche-intelligenz-ermoeglicht-personalisierte-diagnose
6.	idw - Informationsdienst Wissenschaft. (26. Mai 2021). <i>Volkskrankheit Rücken: Verfahren der Künstlichen Intelligenz ermöglicht künftig eine personalisierte Diagnose</i> [Press release]. Abruf unter https://idw-online.de/de/news769433
7.	fitness MANAGEMENT. (16. September 2020). <i>Aufstiegskongress 2020 – ONLINE ONLY: Data Mining und Maschinelles Lernen in der Bewegungs- und Trainingswissenschaft</i> [Press release]. Abruf unter https://www.fitnessmanagement.de/digital/aufstiegskongress-online-only-data-mining-dindorf