

Abschlussbericht

Forschungsprojekt: Integration von ärztlicher Erfahrung, Studienevidenz und Patientenpräferenz anhand eines auf maschinellem Lernen basierenden Therapieempfehlungssystems.

Kurzbezeichnung: Therapieempfehlungssystem

Förderung durch den Etat der **Roland Ernst Stiftung für Gesundheitswesen** für Forschungsprojekte an sächsischen Hochschulen und Fachhochschulen.

Sperrvermerk: Bitte noch nicht veröffentlichen

Antragsteller

- Malberg, Hagen: Professor, Institutsdirektor (Institut für Biomedizinische Technik, Fakultät Elektrotechnik und Informationstechnik, TU Dresden)
- Zaunseder, Sebastian: Wissenschaftlicher Mitarbeiter, Gruppenleiter (Institut für Biomedizinische Technik, Fakultät Elektrotechnik und Informationstechnik, TU Dresden)
- Schmitt, Jochen: Professor, Zentrumsdirektor (Zentrum für evidenzbasierte Gesundheitsversorgung, TU Dresden)

Projektleiter

- Zaunseder, Sebastian, Wissenschaftlicher Mitarbeiter, Gruppenleiter (Institut für Biomedizinische Technik, Fakultät Elektrotechnik und Informationstechnik, TU Dresden)
Telefon: 0351 463 33786
Mail: sebastian.zaunseder@tu-dresden.de
- Schmitt, Jochen, Professor, Zentrumsdirektor (Zentrum für evidenzbasierte Gesundheitsversorgung, TU Dresden)
Telefon: 0351 458 6495
Mail: Jochen.Schmitt@uniklinikum-dresden.de

Projektlaufzeit: 01.01.2015 - 31.03.2017

Autoren des Abschlussberichtes (alphabetisch):

- Abraham, Susanne
- Beckert, Stefanie
- Gräßer, Felix
- Haufe, Eva
- Küster, Denise
- Schmitt, Jochen
- Zaunseder, Sebastian

Inhalt

1	Hintergrund	4
1.1	Ausgangssituation	4
1.1	Grundlagen des maschinellen Lernens und der automatisierten Klassifikation	6
1.1.1	Allgemeine und anwendungsspezifische Vor- und Nachteile des maschinellen Lernens	7
1.1.2	Ensemble-basierten Methoden zur Entscheidungsfindung	8
1.2	Zielsetzung des Projektes	10
1.3	Lösungsansatz.....	11
2	Arbeitspakete	13
2.1	Arbeitspaket 1 - Technisches Management, Projektkoordination, Organisation.....	13
2.2	Arbeitspaket 2 - Medizinische Spezifikation	13
2.3	Arbeitspaket 3 - Datenauswahl und -aufbereitung.....	13
2.4	Arbeitspaket 4 - Grundlagenuntersuchung Klassifikation	13
2.5	Arbeitspaket 5 - Entwicklung eines automatisierten Klassifikationssystems	14
2.6	Arbeitspaket 6 - Interpretation und Überführung in das klinische Studienprotokoll.....	14
3	Ergebnisse.....	15
3.1	Methodische Betrachtungen für die Umsetzung des Therapieempfehlungssystems.....	15
3.2	Funktionsweise der Software für das Therapieempfehlungssystem.....	16
3.3	Training des Prototypen des Therapieempfehlungssystems.....	17
3.4	Überprüfung der Therapieentscheidung durch Fachärzte der Dermatologie	18
3.5	Ergebnisse der klinische Validierung des entwickelten Prototyps.....	19
4	Diskussion.....	20
4.1	Limitationen und Stärken.....	20
4.2	Schlussfolgerungen	21
5	Literatur.....	22

1 Hintergrund

1.1 Ausgangssituation

Medizinische Entscheidungen sollten anhand wissenschaftlicher Kriterien getroffen werden. Zur Überprüfung des Nutzens von Therapien dienen Studien, die den Maßgaben der Evidence-Based-Medicine (EBM) folgen.

Die höchste Güte bieten multizentrische, doppelblinde, placebokontrollierte Studien. Für diese Studien erfolgt der Ausschluss von Störvariablen durch Selektion von Patienten (z.B. nach Alter, Geschlecht, Begleiterkrankungen, Begleitmedikation etc.). Jedoch erfüllt nur ein geringer Anteil aller Patienten (ca. 20%) die strengen Einschlusskriterien, die in klinischen Studien gelten. Patienten in der medizinischen Routineversorgung unterscheiden sich von Studienpatienten in Alter (Kinder, Frauen im gebärfähigen Alter, Senioren sind in Studien stark unterrepräsentiert), Komorbiditäten, Begleittherapien und der Adhärenz. Das ärztliche Monitoring ist zudem im Rahmen von Studien typischerweise aufwendiger als im klinischen Alltag.

Durch diese Strukturungleichheit zwischen dem Setting „klinische Studie“ und dem Setting „Versorgungsalltag“, sowohl seitens der Patienten als auch der Versorgung, sind die Studienergebnisse nicht ohne weiteres auf die medizinische Praxis übertragbar. Dadurch kann die Behandlung eines *individuellen* Patienten nach Studienlage im Einzelfall nicht ausreichend und zweckmäßig, sondern sogar schlechter sein als eine alternative Therapie.

Den Goldstandard bei der Umsetzung der EBM bilden derzeit so genannte S3-Leitlinien (<http://www.awmf.org/>). Die Erstellung von S3-Leitlinien beinhaltet eine systematische Evaluation aller verfügbaren Therapiestudien zu einem Krankheitsbild und auf dieser externen Evidenz basierende, durch Expertenkonsensus verabschiedete Therapieempfehlungen.

Ein genereller, häufig aus den Reihen von im ambulanten Versorgungsalltag tätigen Ärzten geäußelter Kritikpunkt an diesen Leitlinien sind Einschränkungen in der Übertragbarkeit der in die Leitlinienempfehlung eingehenden ärztlichen Expertise, die gegenüber der Studienevidenz eine deutlich untergeordnete Rolle spielt. Nach der gängigen Praxis lässt sich die interne Evidenz (*individuelle klinische Expertise* [1]) folglich nur indirekt, eingeschränkt und möglicherweise unzureichend im Versorgungsalltag nutzen.

Diese Vermutung konnte der Antragsteller anhand einer Reihe von Routinedatenanalysen sächsischer AOK PLUS Versicherter stützen [2]–[6]: Sowohl für die Neurodermitis, als auch für die Psoriasis, die beiden häufigsten chronischen dermatologischen Erkrankungen, erfolgt die Therapie in klinischen Praxis vielfach nicht leitliniengerecht. Zudem konnte eine deutliche Heterogenität in der Versorgung zwischen den einzelnen Ärzten gezeigt werden: Während einzelne Ärzte bestimmte Therapien bei fast allen betroffenen Patienten anwenden, wird die gleiche Therapie von anderen Ärzten der glei-

chen Fachrichtung (Dermatologie bzw. Pädiatrie) nur sehr selten genutzt. Dieser Befund legt nahe, dass die individuelle ärztliche Expertise und Erfahrung wesentliche Determinanten der Therapieempfehlung sind und durch eine systematische Nutzung der (kollektiven) Expertise ein deutlicher Zusatznutzen für die medizinische Versorgung erreicht werden kann.

Anhand einer weiteren Voruntersuchung konnten die Antragsteller zeigen, dass eine standardisierte Behandlung von Kindern und Erwachsenen mit Neurodermitis gemäß einem evidenzbasierten Algorithmus basierend auf klinischen Studien praktikabel ist und zu einer effektiven Kontrolle der Krankheitsaktivität führt, dass jedoch vielfach ein individuelles Abweichen von der Studienempfehlung notwendig ist und eine optimale Therapieempfehlung ausschließlich auf der Basis von Studiendaten bzw. ausschließlich auf Basis ärztlicher Erfahrung nicht erreichbar ist. [7], [8]

Vor diesem Hintergrund hat bereits 1996 David Sackett, der Begründer der EBM, gefordert, dass bei der Behandlung individueller Patienten eine „*Integration individueller klinischer Expertise (des Behandlers) mit der best-verfügbaren externen Evidenz aus systematischer Forschung.*“ [1], [9] anzustreben ist. Neben den beiden Aspekten „Studienevidenz“ und „ärztliche Expertise“ ist als dritte wesentliche Einflussgröße auf die Therapieentscheidung die „Präferenz des Patienten“ [10] anzusehen.

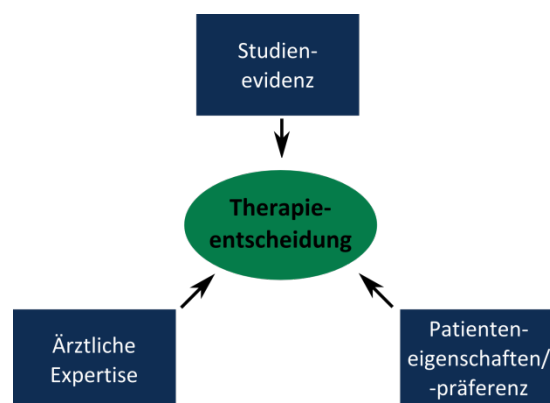


Abbildung 1 Drei Einflussgrößen auf die individuelle Therapie

Praktisch ist die Verknüpfung von EBM mit der internen Evidenz aus der Erfahrung der Behandler sowie Patienteneigenschaften und Patientenpräferenzen aufgrund der hohen Dimension (u.a. Faktoren wie Lebensqualität, Patientenpräferenz, bisherige Therapien, Komorbidität, Behandlungszufriedenheit sollten zusätzlich einfließen) und der Heterogenität der Eingangsdaten allerdings äußerst komplex. Methoden, die die EBM um die kollektive interne Evidenz sowie Patienteneigenschaften/-präferenzen erweitern und in der klinischen Praxis systematisch und transparent in individuelle Therapieempfehlungen überführen, sind heute nicht verfügbar.

Verfahren des maschinellen Lernens, speziell automatisierte Klassifikatoren, sind in diesem Kontext von Interesse, da sie prinzipiell in der Lage sind, eine Bewertung durch assoziative Betrachtung der unterschiedlichsten Faktoren hochdimensionaler Entscheidungsräume vorzunehmen.

1.1 Grundlagen des maschinellen Lernens und der automatisierten Klassifikation

Maschinelles Lernen beinhaltet Algorithmen, die Zusammenhänge und Regeln aus Datensätzen ableiten. Die im Rahmen des maschinellen Lernens zum Einsatz kommenden Algorithmen lassen sich in drei Kategorien einteilen:

- Überwachtes Lernen: ein Algorithmus erlernt anhand von Paaren bestehend aus Merkmalsvektoren¹ X und jeweils gewünschter Ausgabe Y eine Abbildungsvorschrift $X \rightarrow Y$.
- Unüberwachtes Lernen: ein Algorithmus erschließt anhand von Merkmalsvektoren Zusammenhänge innerhalb eines Datenmaterials (ohne dass die jeweils gewünschte Ausgabe vorhanden ist).
- Bestärkendes Lernen: ein Algorithmus erlernt anhand einer zu optimierenden Funktion des Ausgangssystems durch Belohnung und Bestrafung seine Funktion.

Von besonderem Interesse für das Projekt und im Kontext bisheriger medizinischer Forschung, Entwicklung und Anwendungen am verbreitetsten ist der Einsatz des überwachten Lernens.

Sobald Daten gemäß der Struktur aus Abbildung 2 (d.h. Merkmalsvektoren mit zugehöriger Ausgabe) vorhanden sind, bieten Algorithmen des überwachten Lernens (sog. Induktoren) die Möglichkeit, eine Abbildungsvorschrift $X \rightarrow Y$ zu erzeugen. Nach abgeschlossenem Lernvorgang kann diese Abbildungsvorschrift auf unbekannte Eingaben (d.h. Merkmalsvektoren der entsprechend Anlernen erwarteten Form) angewandt werden. So gewonnene Aussagen können sowohl diagnostischer Art (z.B. Zuordnung eines Merkmalsvektors zu gesund/krank) als auch direkt therapeutischer Art sein (Entscheidung für Medikament A/B in Abhängigkeit eines Merkmalsvektors). Handelt es sich bei Y um eine Menge diskreter Elemente, so spricht man von automatisierter Klassifikation (die Abbildungsvorschrift wird entsprechend als Klassifikator bezeichnet; ein Klassifikator wird zusammen mit dem ihn erzeugenden Lernalgorithmus im Folgenden als (Klassifikations-)Paradigma bezeichnet).

Aufgrund der weit gefassten Definition kommen Konzepte die dem maschinellen Lernen zuzuordnen sind in nahezu allen Gebieten der Medizin zum Einsatz. Exemplarisch für typische Anwendungsgebiete automatisierter Klassifikatoren sei hier lediglich auf die Bildverarbeitung [11] und intensivmedizinische Anwendungen [12] verwiesen. Für die automatisierte Klassifikation kommen dabei fachübergreifend verschiedenste Methoden zum Einsatz, u.a. Entscheidungsbäume, Fuzzy Klassifikatoren, Bayes Klassifikatoren, neuronale Netze verschiedener Ausprägung und Support Vector Machines.

¹ Als Merkmalsvektor ist im Kontext des Antrags die Sammlung der Eigenschaften zur Beschreibung eines Patienten zu verstehen

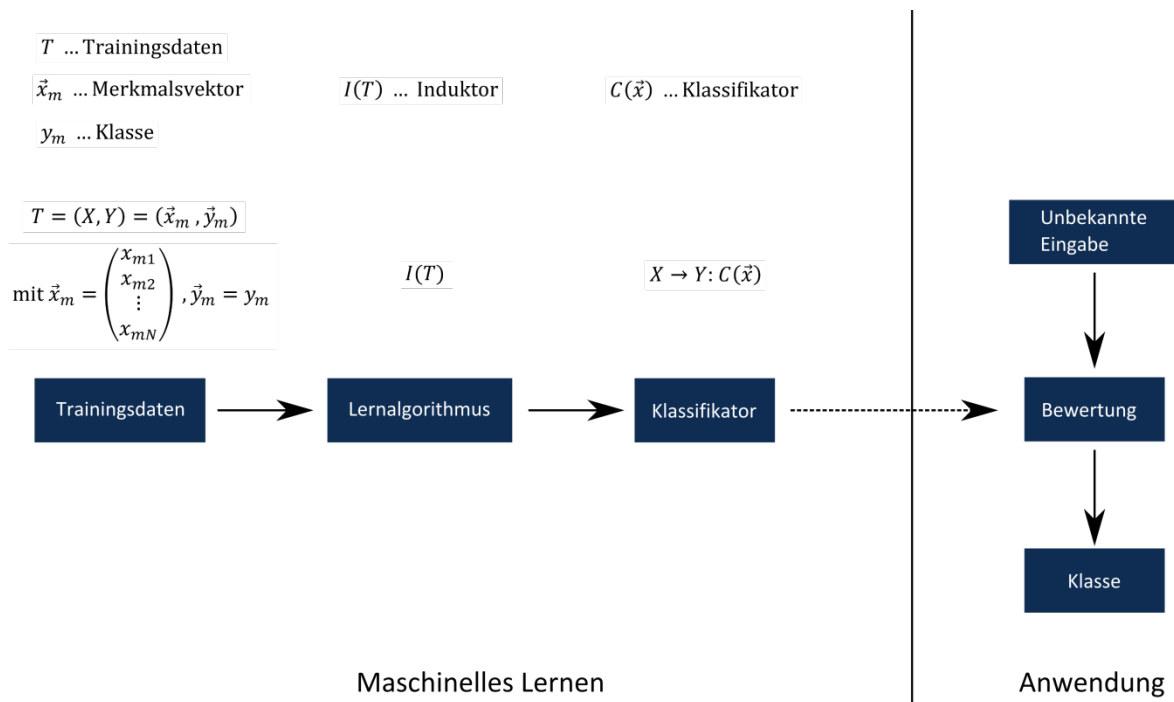


Abbildung 2 Schematische Darstellung der Anwendung maschineller Lernverfahren zur automatisierten Klassifikation

1.1.1 Allgemeine und anwendungsspezifische Vor- und Nachteile des maschinellen Lernens

Neben der Möglichkeit zur Verwendung im Falle sehr großer Datenmengen besteht der entscheidende Vorteil des Einsatzes von maschinellem Lernen und automatisierten Klassifikatoren gegenüber einer menschlichen Bewertung in der Möglichkeit, hochdimensionale Eingaberäume durch (ggf. nicht-lineare Operationen) effizient analysieren zu können. Automatisierte Klassifikatoren sind in der Lage, eine Bewertung durch eine assoziative Betrachtung der unterschiedlichsten Faktoren hochdimensionaler Entscheidungsräume vorzunehmen, wobei wesentlich komplexere Zusammenhänge als dies durch den Menschen möglich ist erfasst werden können. So kann trotz der für Krankheitsbilder typisch unscharfen Voraussetzungen eine wissenschaftlich fundierte Entscheidung getroffen werden, was die automatisierte Klassifikation für die Medizin höchst interessant macht.

Im Kontext der Anwendung von Klassifikatoren existieren allerdings auch allgemeine Limitationen. Zunächst betrifft dies die Auswahl eines Klassifikationsparadigmas. Pauschal ist nicht zu sagen, welcher, aus der Vielzahl möglicher Klassifikatoren, problemspezifisch die optimale Wahl darstellt [13]. Ebenfalls ist nicht gewährleistet, dass ein Klassifikator nach abgeschlossenem Lernprozess die Fähigkeit besitzt, zu Generalisieren (d.h. von den während des Trainings präsentierten Mustern auf unbekannte Muster zu schließen). Des Weiteren sorgt die Arbeitsweise von Klassifikatoren dazu, dass oftmals keine Transparenz bei der Entscheidungsfindung vorhanden ist (Black-Box Verhalten). Exemplarisch sei beispielsweise auf SVMs verwiesen, welche Eingabedaten in der Regel zunächst mittels Kerneltransformation [14] in einen höherdimensionalen Raum bringen, um diese dort separieren zu können. Die Nachvollziehbarkeit der letztendlich getroffenen Entscheidung kommt dabei

vollkommen abhanden. Einfachere Klassifikatoren, die auf derartige Schritte und nicht-lineare Operationen verzichten, besitzen wiederum oftmals nicht die nötige Leistungsfähigkeit. Gerade im Kontext medizinischer Anwendungen besitzt die mangelnde Transparenz entscheidende Bedeutung und ist maßgeblich verantwortlich dafür, dass maschinelles Lernen und automatisierte Klassifikation in der Medizin noch keine weitere Verbreitung gefunden haben.

Neben den Limitationen allgemeiner Art besteht im Kontext des Forschungsvorhabens insbesondere die anwendungsspezifische Limitation, dass ein System wie es angestrebt wird, eine von klassischen Klassifikationsparadigmen abweichende Auslegung erfordert. Die Kombination von EBM mit der kollektiven Evidenz, Patienteneigenschaften und Patientenpräferenzen – typischerweise heterogene und teilweise unvollständige Daten – erfordert bereits beim Anlernen besondere Flexibilität. Eine definierte Berücksichtigung einzelner Faktoren, um Eingangsdaten systematisch und transparent in individuelle Therapieempfehlungen zu überführen, wird durch klassische Methoden ebenfalls nicht unterstützt.

1.1.2 Ensemble-basierten Methoden zur Entscheidungsfindung

Ensemble-basierten Methoden zur Entscheidungsfindung (im Englischen u.a. auch als multiple classifier systems, committee of classifiers oder mixture of experts bezeichnet) werden heute typischerweise als Erweiterung klassischer Klassifikatoren betrachtet und eingesetzt. Die im Kontext von Ensembleklassifikatoren angestellten theoretischen Überlegungen und erzielten methodischen Entwicklungen definieren allerdings gleichzeitig ein Framework, welches die strukturellen Voraussetzungen einer interpretierbaren Kombination im oben angesprochenen Sinn bietet.

Ensemble-basierten Methoden zur Entscheidungsfindung, im Folgenden als Ensembleklassifikatoren bezeichnet, legen ihrer Entscheidung keine einzelne Abbildungsvorschrift zu Grunde, vielmehr wird eine Zuordnung auf Basis einer Vielzahl einzelner Klassifikatoren vorgenommen (siehe Abbildung 3).

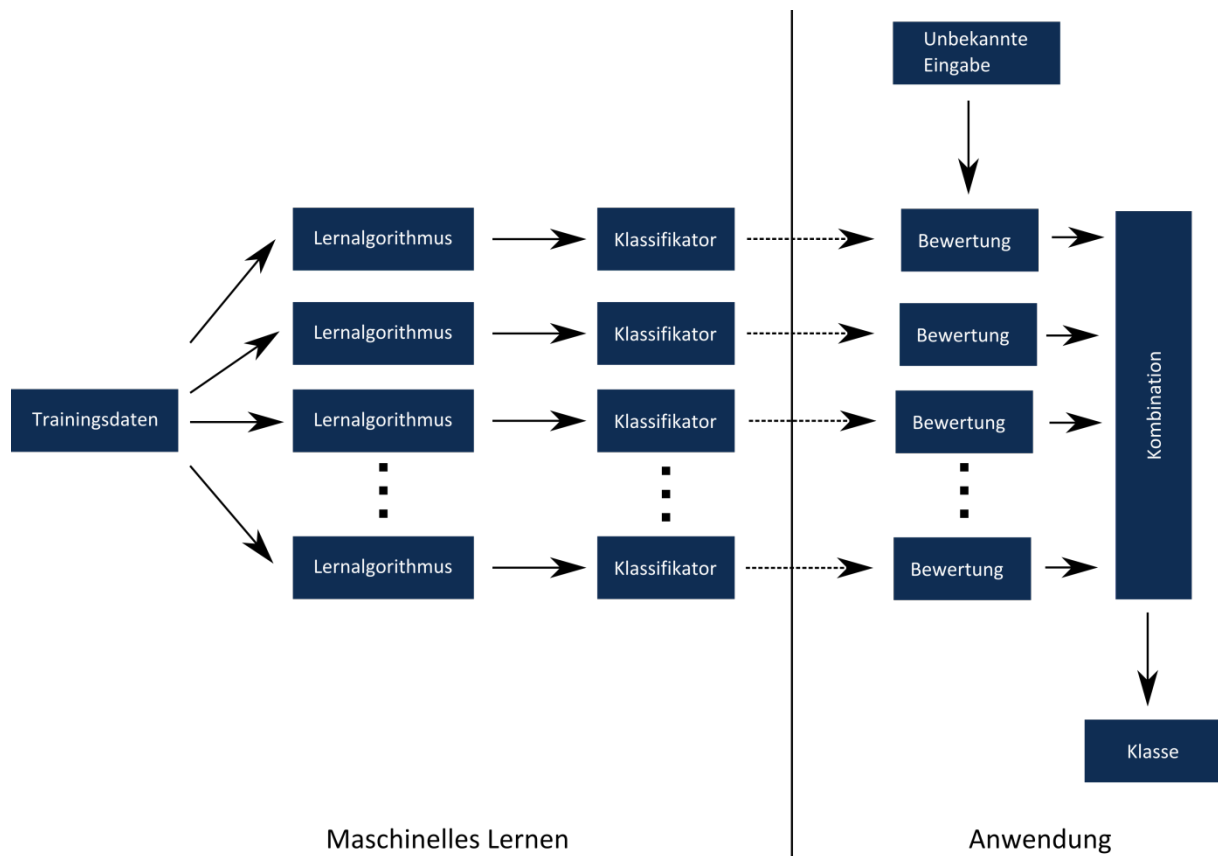


Abbildung 3 Schematische Darstellung der Anwendung maschineller Lernverfahren zur Ensemble-basierten Klassifikation

Das Vorgehen weist damit einen engen Zusammenhang zu dem oftmals auch durch Menschen, gerade im Kontext medizinischer Entscheidungen, angewandtem Vorgehen auf. Statistisch lässt sich einfach zeigen, dass ein Ensemble-Klassifikator dabei bereits dann perfekte Ergebnisse erzielt, wenn²

- die Anzahl der beinhalteten Klassifikatoren groß genug ist,
- sich die einzelnen Klassifikatoren bezüglich der produzierten Resultate unterscheiden und
- einzelne Klassifikationsgüten von mindestens 50% erreicht werden.

Da es sich bei den genannten Voraussetzungen um relativ milde Vorgaben handelt und Ensembleklassifikatoren ihre hohe Leistungsfähigkeit bewiesen haben, besitzen sie heute eine zunehmende Bedeutung in verschiedenen Anwendungsgebieten. [15]

Aus technischer Sicht sind für Anlernen und Einsatz eines Ensembleklassifikators zwei Charakteristika prägend: das Anlernen der einzelnen Klassifikatoren und die Kombination dieser.

Bezüglich des Anlernens existieren grundsätzliche ähnliche Möglichkeiten und Bedingungen wie generell beim überwachten Lernen. Entscheidend für den Erfolg eines Ensembles ist neben dem erfolgreichen Anlernen jedes einzelnen Klassifikators die Diversität (Unterschiede in den durch die Klassifikatoren erzielten Resultaten) des Ensembles. Diversität wird in der Regel dadurch erreicht, dass Paradigmen gewählt werden, welche durch Variationen der Trainingsdaten und/oder genutzten Merk-

² Die Angaben beziehen sich hier auf ein Zweiklassenproblem

male in stark abweichenden Klassifikatoren resultieren. Bagging [16], Random Subspace Methoden [17] und Boosting [18] sind bekannte Strategien, um die Diversität zu erzielen. Die im Rahmen von Ensembles zum Einsatz kommenden Klassifikatorparadigmen zeichnen sich typischerweise durch ihren relativ einfachen Aufbau und einer einzeln betrachtet niedrigen Leistungsfähigkeit aus (daher als *WeakLearner* bezeichnet). Oft genutzt werden beispielsweise Entscheidungsbäume und neuronale Netze.

Die zweite Charakteristik von Ensemble-Klassifikatoren ist die Strategie zur Kombination der einzelnen Klassifikatoren. Dazu wurden verschiedene Varianten erfolgreich angewandt, wobei das Spektrum von sehr einfachen Ansätzen (Mehrheitsentscheidung, Maximumentscheidung) bis hin zu komplexen Kombinationen der Einzelentscheidungen geht (gewichtete Mehrheitsentscheidung unter Einbeziehung von Informationen aus dem Lernprozess, ggf. adaptiv an Eingabedaten angepasst) [15]. Beide Elemente sind für die Funktion eines Ensembles gleichermaßen bedeutend. Gerade die Kombination und die dieser zu Grunde liegenden theoretischen Überlegungen erlauben es aber sehr gezielt, Einfluss auf die Funktion von Ensembles zu nehmen.

1.2 Zielsetzung des Projektes

Ziel des beantragten Vorhabens ist die Entwicklung einer lernenden **informationstechnischen Lösung** (sog. *Therapieempfehlungssystem*), die dem Arzt auf Basis

- der Evidenz aus Studien
- der (kollektiven) Erfahrung aus der Versorgungspraxis
- individueller Patienteneigenschaften und Patientenpräferenzen

die **individuell optimale Therapie** für einen Patienten vorschlägt und anhand **nachvollziehbarer Auswahlkriterien** begründet.

Inhaltlich strebt das Projekt dazu **methodische Entwicklungen** im Bereich des maschinellen Lernens, speziell im Zusammenhang mit automatisierten Klassifikatoren und **Ensembleklassifikatoren**, an. Den Arbeiten liegt die Hypothese zu Grunde, dass Ensembleklassifikatoren aufgrund ihrer Struktur einerseits die definierte Integration heterogener Informationen erlauben, andererseits aufgrund der Einfachheit der einzelnen Klassifikatoren unter Anwendung entsprechender Methoden die Möglichkeit einer Interpretation automatisiert generierter Therapieempfehlungen bieten. Ziel ist somit, die Ensemble-basierten Methoden dahingehend zu erweitern/einzusetzen, dass eine systematische und transparente Integration der oben genannten Informationen mit definiertem Einfluss für ein Therapieempfehlungssystem möglich wird. Gleichzeitig werden im Rahmen des Forschungsvorhabens Methoden geschaffen, die die Eigenschaften angelernter Ensembles nutzen, um der mangelnden Transparenz, einem generellen Problem automatisierter Klassifikatoren, entgegenzuwirken und so ein bezüglich der realen Anwendung maßgebliches Hindernis auszuräumen.

1.3 Lösungsansatz

Die angestrebten methodischen Entwicklungen sollen im Rahmen des Forschungsvorhabens exemplarisch anhand eines Therapieempfehlungssystems für ein dermatologisches Krankheitsbild (Schuppenflechte/Psoriasis) entwickelt und evaluiert werden.

Zur Lösung der Aufgabe werden folgende inhaltliche Teilziele verfolgt:

- **Datenauswahl/-aufbereitung:**

Selektion und Aufbereitung von Daten aus Patientenakten des Universitätsklinikums Dresden.

- **Auswahl und Umsetzung von Methoden zur automatisierten Klassifikation:**

Einbezogen in den Auswahlprozess werden verschiedene Paradigmen zur Klassifikation, u.a. Support Vector Machines, Fuzzy-Klassifikatoren, Entscheidungsbäume, Bayes-Klassifikator. Ebenfalls wird die Möglichkeit der Klassifikation mittels Ensembles betrachtet, wobei individuelle Patientencharakteristika (Alter, Geschlecht, aktuelle Erkrankungsschwere, Lebensqualität, Vorbehandlungen, Komorbiditäten und Begleittherapien) und Patientenpräferenzen explizit berücksichtigt werden. Auf der Basis dieser Vorbetrachtungen sollen Klassifikatoren umgesetzt und vergleichend bewertet werden.

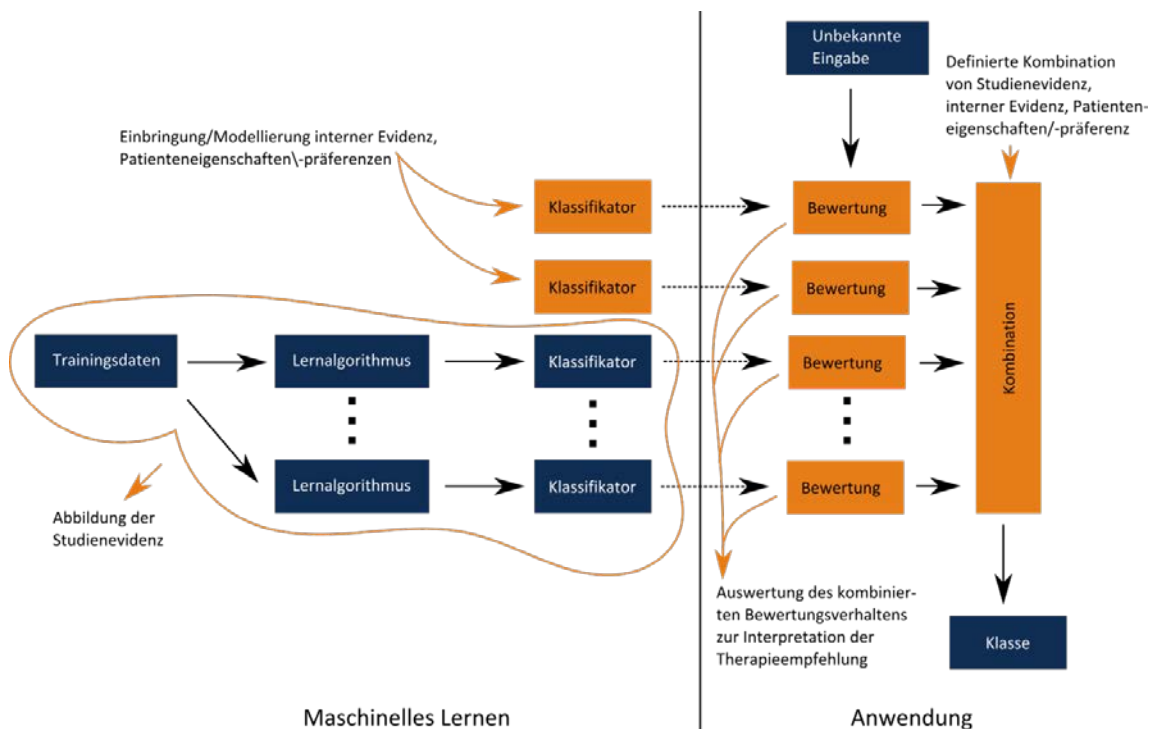


Abbildung 4 Schematische Darstellung zu Aspekten eines Ensembleklassifikators und Ansatzpunkten, anhand derer im Forschungsvorhaben Einfluss auf die Therapieempfehlung genommen wird bzw. weiterreichende Informationen abgeleitet werden (in orange)

- **Interpretation hochdimensionaler Entscheidungsstrukturen:**

Um Entscheidungen des Therapieempfehlungssystems transparent zu gestalten werden Methoden analysiert/geschaffen, die die jeweils zu Grunde liegenden Auswahlkriterien zugänglich machen. Ausgenutzt werdend dabei die in der Struktur und Funktion angelernter Klassifikatoren/Ensembles beinhalteten Informationen. Verfahren wie sie etwa im Bereich der Merkmalsselektion zum Einsatz kommen, können hierbei zum Einsatz kommen.

- **Bewertung und Generalisierung:**

Abschließend sollen die durch das Therapieempfehlungssystem getroffenen Aussagen umfassend charakterisiert werden. Dabei sind sowohl die Leistungsfähigkeit insgesamt, als auch Abweichungen gegenüber einer EBM-basierter Therapieempfehlung von Fachexperten und Korrelationen zu einer erfahrungsbasierten Therapie Gegenstand der Betrachtungen.

2 Arbeitspakete

2.1 Arbeitspaket 1 - Technisches Management, Projektkoordination, Organisation

Bearbeiter: ZEGV, IBMT

- Während der gesamten Laufzeit des Forschungsvorhabens fand ein kontinuierliches Projektmanagement durch Studienmitarbeiter des IBMT und ZEGV statt.
- Die Projektergebnisse beider Partner wurden nach Projektabschluss zu einem gemeinsamen Abschlussbericht zusammengefasst.
- Berichterstattung und Koordinationsaktivitäten mit dem Projektträger wurden durch Studienmitarbeiter des IBMT und ZEGV erbracht.
- In regelmäßigen Abständen fanden interne Workshops der Projektpartner statt.

2.2 Arbeitspaket 2 - Medizinische Spezifikation

Bearbeiter: ZEGV, IBMT

- Am ZEGV wurden durch den Projektleiter und die Studienmitarbeiter Patienten- und Erkrankungscharakteristika ausgewählt, die in die Therapieempfehlung eingeflossen sind. Vorgaben aus den jeweiligen Fachinformation der verfügbaren Therapieoptionen (z. B. Kontraindikationen) wurden berücksichtigt.
- Der Entwurf der Spezifikationen wurde mit 6 externen, klinischen Experten konsentiert
- Am IBMT wurde im Folgenden eine Datenbank zur Erfassung der festgelegten Patienten- und Erkrankungscharakteristika aus dem Klinikinformationssystem programmiert und in Abstimmung mit dem ZEGV getestet.

2.3 Arbeitspaket 3 - Datenauswahl und -aufbereitung

Bearbeiter: ZEGV

- Eingabe der anonymisierten, retrospektiven Daten von 213 Psoriasis Patienten aus der Poliklinik für Dermatologie des Universitätsklinikum Carl Gustav Carus durch eine Studienmitarbeiterin des ZEGV.

2.4 Arbeitspaket 4 - Grundlagenuntersuchung Klassifikation

Bearbeiter: IBMT

- Mitarbeiter des IBMT führten die Bewertung und Recherche zu Klassifikationsparadigmen (Betrachtung der Eigenschaften verschiedener Ansätze, u.a. Fuzzy-Systeme, Support Vector Machines, Entscheidungsbäume, Recommender Methoden) durch.

- Es wurden Möglichkeiten der Vereinfachung hochdimensionaler Klassifikatorsysteme erarbeitet.
- Die Bewertung von Klassifikationsmethoden und die Auswahl des Schwerpunktes Recommender Systeme (→ Recommender Systeme sind prinzipiell für den Umgang mit lückenhaften und heterogenen Daten geeignet und erlauben durch den ähnlichkeitsbasierten Klassifikationsansatz zudem die Interpretation von Klassifikationsentscheidungen) wurde von Mitarbeitern des IBMT durchgeführt.

2.5 Arbeitspaket 5 - Entwicklung eines automatisierten Klassifikationssystems

Bearbeiter: IBMT, ZEGV

- Anlernen der automatisierter Klassifikatoren gemäß AP 4 mit den anonymisierten, retrospektiven Daten von 213 Psoriasis Patienten durch eine Studienmitarbeiterin des ZEGV
- Bereitstellung des ersten Testprogramms eines automatisierten Unterstützungssystems für den klinischen Betrieb durch IBMT
- Im Anschluss erfolgte die Evaluation des Testprogramms und Bewertung der klinischen Daten durch 6 externe klinische Experten

2.6 Arbeitspaket 6 - Interpretation und Überführung in das klinische Studienprotokoll

Bearbeiter: ZEGV, IMBT

- Die klinische Validierung des entwickelten Prototyps des Therapieempfehlungssystems erfolgte in der ambulanten Psoriasis-Sprechstunde der Poliklinik für Dermatologie des Universitätsklinikums Carl Gustav Carus Dresden auf Praktikabilität und Funktionalität.
- Die Validierung erfolgte in Form einer Querschnittserhebung durch eine Studienmitarbeiterin des ZEGVs an 67 Patienten, die in der Psoriasis-Sprechstunde der Poliklinik für Dermatologie des Universitätsklinikum Carl Gustav Carus mit einer diagnostisch gesicherten Psoriasis zwischen vorstellig werden.

3 Ergebnisse

3.1 Methodische Betrachtungen für die Umsetzung des Therapieempfehlungssystems

Im Rahmen des Vorhabens wurden umfangreiche Recherchen zu Möglichkeiten des maschinellen Lernens für die Therapieempfehlung durchgeführt und verschiedene Varianten implementiert. Das betrifft nachbarschaftsbasierte Klassifikationen, auf denen ein Schwerpunkt lag, allerdings auch herkömmliche Klassifikatoren, wie Entscheidungsbäume, und Ensembleklassifikatoren, wobei speziell die Kombination von demografischen Informationen und Outcome-Informationen im Sinne kollaborativer Filter im Mittelpunkt stand. Die Entwicklungen sind in drei methodischen Veröffentlichungen beschrieben [19]–[21]. Abbildung 5 und Abbildung 6 verdeutlichen exemplarisch die erzielten Ergebnisse. Es wird deutlich, dass maschinelle Lernverfahren in ähnlicher Weise wie menschliche Experten in der Lage sind, den „Goldstandard“ (gegeben durch die Therapieentscheidung des Uniklinikums Dresden) wiederzugeben. Speziell Abbildung 6 verdeutlicht zudem, dass gerade für mit dem Goldstandard übereinstimmenden Therapien, das Outcome oftmals gut ist. Im Umkehrschluss heißt das, dass die Empfehlungen des automatisierten Systems oftmals sinnvoll, und gerade auch im Falle von Abweichungen gegenüber der Einschätzung des menschlichen Bewerter, eine zu prüfende Option darstellen.

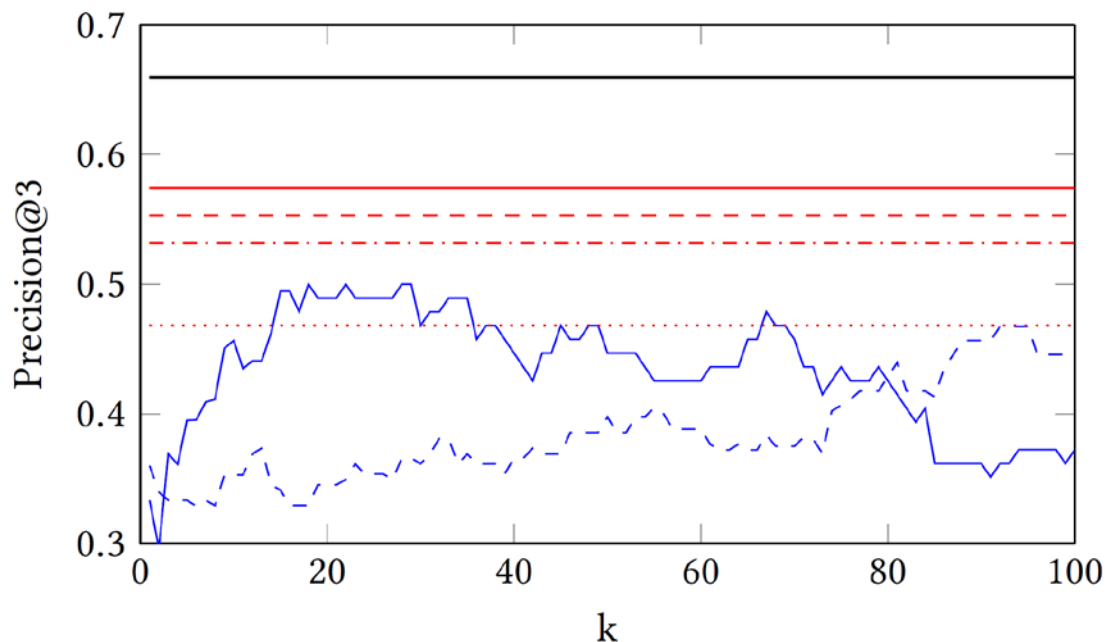


Abbildung 5 Exemplarische Darstellung quantitativer Ergebnisse bei der Therapieempfehlung jeweils gegenüber dem Goldstandard des Uniklinikums Dresden (Rot: Vier Expertenbewertungen; Schwarz: Anwendung von Entscheidungsbäumen; Blau durchgezogen: Demografischer Recommender; Blau gestrichelt: Kollaborativer Filter)

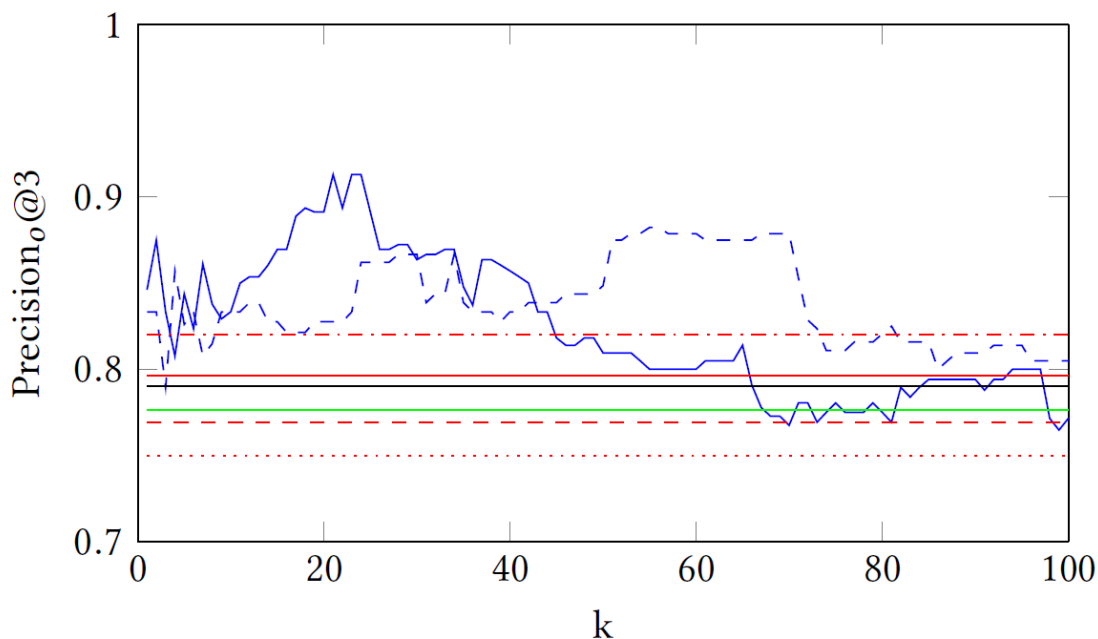


Abbildung 6 Anteil der erfolgreichen Therapien (definiert über positives Outcome) wenn Empfehlung mit Goldstandard übereinstimmt (Rot: Expertenbewertungen; Schwarz: Anwendung von Entscheidungsbäumen; Blau durchgezogen: Demografischer Recommender; Blau gestrichelt: Kollaborativer Filter; Grün: Anteil der erfolgreichen Therapien im Goldstandard)

3.2 Funktionsweise der Software für das Therapieempfehlungssystem

Auf Basis der methodischen Entwicklungen wurde eine Variante des Therapieempfehlungssystems für die klinische Evaluation umgesetzt. Bei der entwickelten Software handelt es sich um einen Prototyp für ein intelligentes klinisches Unterstützungssystem (CDSS), mit dem exemplarisch die Einsatzfähigkeit derartiger Systeme in der klinischen Praxis untersucht wurde. Die methodische Basis sind sog. Recommender Methoden, die im Bereich des e-Commerces sehr erfolgreich zum Einsatz kommen (vgl. AP 4). Ziel des im Rahmen dieser Arbeit entwickelten Therapieempfehlungssystems ist es, basierend auf der kollektiven klinischen Erfahrung und unter Berücksichtigung von evidenzbasierten Leitlinien und Vorgaben aus der Zulassung der unterschiedlichen Therapieoptionen, individualisierte Therapien vorzuschlagen. Die kollektive klinische Erfahrung liegt dabei in Form von umfangreichen anonymisierten Daten zu Krankheits- und Therapieverlauf von 213 Patienten vor, auf deren Grundlage der Erfolg einer Therapie geschätzt und eine Therapieempfehlung abgeleitet wird. Hierfür wird ein Ähnlichkeitsmaß zwischen Patienten bzw. Visiten berechnet, welches deren Einfluss auf die Vorhersage des Erfolges für die eingesetzten Therapien bestimmt. Basierend auf evidenzbasierten Ausschlusskriterien können Therapien nachfolgend von der Empfehlung ausgeschlossen werden.

Die Software umfasst den im Folgenden aufgeführten Funktionsumfang: Patientendaten können über eine Benutzeroberfläche eingegeben, angezeigt und editiert werden, wobei für einen Patienten mehrere Visiten angelegt werden können. Sämtliche Daten werden in einer lokalen MySQL Datenbank gespeichert und organisiert. Für alle in der Datenbank gespeicherten Patienten können bereits

vorhandene Visiten ausgewählt und verordnete bzw. angewandte Therapien sowie vom System generierte Therapieempfehlungen angezeigt werden. Wird eine neue Visite für einen bereits vorhandenen oder neuen Patienten angelegt, so werden ebenfalls die vom System generierten Therapieempfehlungen angezeigt und die vom Arzt verordnete Therapie kann ausgewählt und gespeichert werden. Therapieempfehlungen werden dabei als Liste von vorgeschlagenen Therapien angezeigt, die nach der durch die Software angenommenen Relevanz geordnet ist.

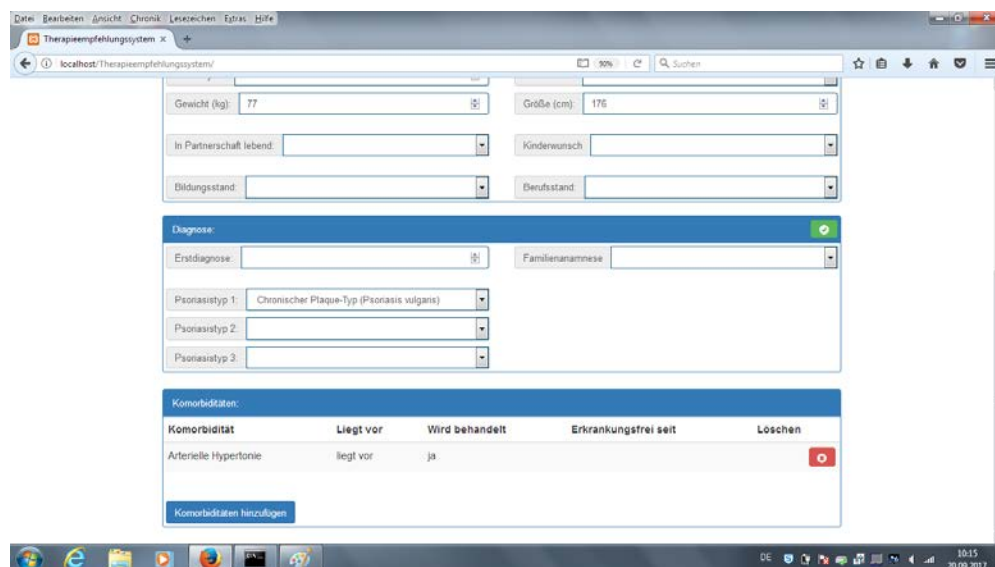
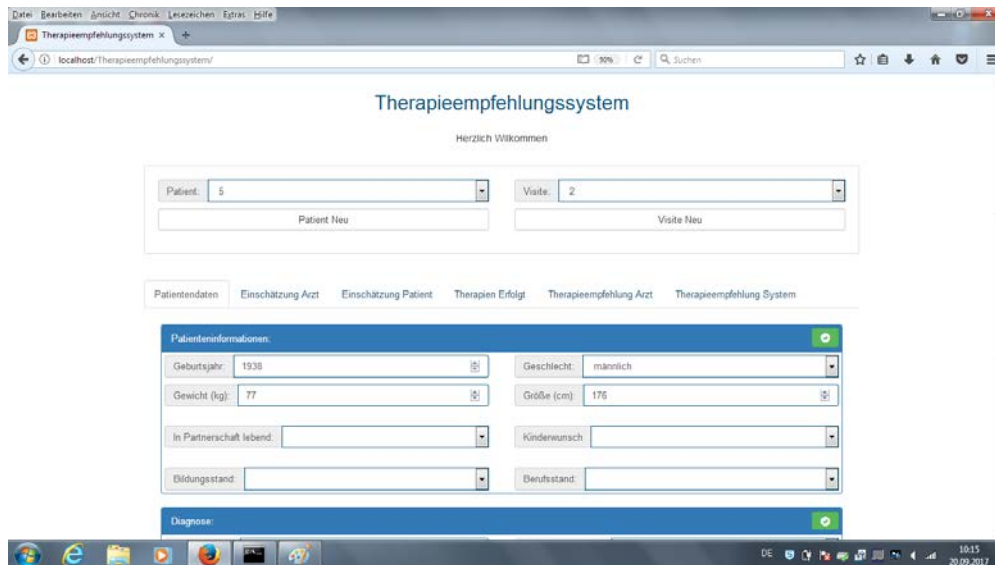


Abbildung 7 Screenshots des umgesetzten Prototypischen Therapieempfehlungssystems

3.3 Training des Prototypen des Therapieempfehlungssystems

Aus den Patientenakten von 213 Patienten der Psoriasis-Sprechstunde der Poliklinik für Dermatologie des Universitätsklinikums Carl Gustav Carus Dresden wurden anonymisierte und rückblickende Daten zu Befunden, Therapie, unerwünschte Ereigniss, Erkrankungsverlauf und –ergebnisse (wie zum Bei-

spiel der PASI (Psoriasis Area and Severity Index), der DLQI (Dermatology Quality of Life Index), die momentanen Lokalisationen oder die Behandlungszufriedenheit)) extrahiert und das Therapieempfehlungssystem damit „angelernt“. Dabei konnten Daten von 1089 Visiten in das System eingegeben werden.

Ebenfalls wurden Ausschlusskriterien bei der Auswahl von Systemtherapien zur Behandlung der Psoriasis aus der aktuell gültigen S3-Leitlinie in das System eingearbeitet, wie beispielsweise Komorbiditäten oder Kinderwunsch.

Eine detaillierte Beschreibung zu Aufbau und Inhalt des Prototypen wird im Dezember 2017 im Rahmen der Promotion der Studienmitarbeiterin und Koautorin des Berichts, Frau Stefanie Beckert veröffentlicht.

3.4 Überprüfung der Therapieentscheidung durch Fachärzte der Dermatologie

Wie im Antrag vorgesehen wurden nach dem Training des Systems sechs Experten auf dem Gebiet der Psoriasis zur Überprüfung der Therapieentscheidung eingeladen. Ziel dieses Arbeitspaketes war es die Datengrundlage aufgrund derer das entwickelte Therapieempfehlungssystem seine Therapieentscheidungen vornimmt zu validieren.

Aus den unter Punkt 3.3 beschriebenen Patienten der der Psoriasis-Sprechstunde der Poliklinik für Dermatologie des Universitätsklinikums Carl Gustav Carus wurden 100 Patienten zum Zweck der Validierung selektiert. Dabei wurden 75% an Patienten mit Therapieumstellung und 25% ohne Therapieumstellung für die Überprüfung herangezogen und den sechs Fachärzten der Dermatologie vorgelegt. Diese Experten sind an geographisch gleichmäßig auf Deutschland verteilten dermatologischen Kliniken in Deutschland tätig. Dazu wurden den Experten online ein Zugang zum Therapieempfehlungssystem bereitgestellt.

Den Experten wurden für ihre Therapieempfehlung alle Informationen aus dem System zugänglich gemacht (z. B. wie Alter, Geschlecht, Größe und Gewicht, Jahr der Erstdiagnose, Familienanamnese und Komorbiditäten). Zusätzlich erhielten die Experten in Form einer Tabelle die Informationen zu allen jemals erfolgten Therapien, Therapieverlauf, die jeweils aktuelle Therapie, unerwünschten Arzneimittelwirkungen, PASI, DLQI, der Patientenbewertung der Therapie und die betroffenen Körperregion.

Die Experten waren daraufhin aufgefordert pro Patient zu entscheiden, wie sie die Therapie fortführen würden und ob, bei Abweichung von der Empfehlung des Systems, diese eine nachvollziehbare Alternative darstellt.

Die Übereinstimmung der Experten mit dem Goldstandard (ursprüngliche Therapieempfehlung laut Krankenakte) wurde mit statistischen Verfahren ermittelt. Diese Ergebnisse verdeutlichen die Schwierigkeit und Heterogenität der Therapieempfehlung auch für menschliche Bewerter. Die detail-

lierten Ergebnisse der Überprüfung der Therapieentscheidung durch fünf Fachärzte der Dermatologie (Experten) wird im Dezember 2017 im Rahmen der Promotion der Studienmitarbeiterin und Koautorin des Berichts, Frau Stefanie Beckert veröffentlicht.

3.5 Ergebnisse der klinische Validierung des entwickelten Prototyps

Nach der Überprüfung durch die Fachexperten erfolgte die klinische Validierung des entwickelten Prototyps des Therapieempfehlungssystems in der ambulanten Psoriasis-Sprechstunde der Poliklinik für Dermatologie des Universitätsklinikums Carl Gustav Carus Dresden. Ziel war es die Praktikabilität und Funktionalität des Systems zu ermitteln. Für die Validierungsstudie lag ein positives Ethikvotum vor der Ethikkommission der TU Dresden vor.

Die behandelnde Ärztin Frau Dr. Abraham oder ein/e Stellvertreter/in erhoben im Rahmen der ambulanten Psoriasis-Sprechstunde der Poliklinik für Dermatologie des Universitätsklinikum Carl Gustav Carus alle zur Anamnese und Krankheitsaktivität nötigen Daten und Scores der Patienten, die der Studienteilnahme eingewilligt hatten. Parallel erfolgte durch die Studienmitarbeiterin Frau Beckert die Eingabe der relevanten Informationen in das Therapieempfehlungssystem. Der Verlauf der Sprechstunde und die Empfehlung zur weiteren Therapie durch die behandelnde Ärztin wurden dabei vom Therapieempfehlungssystem nicht beeinflusst. Erst nach Entlassung des Patienten nach Beendigung der Konsultation wurde die Therapieempfehlung des Systems generiert und mit der tatsächlich begonnenen Therapie abgeglichen.

Zur Validierung der generierten Therapieempfehlungssystems wurde die tatsächliche Entscheidung der behandelnden Ärztin mit der Empfehlung des Systems verglichen. Bei Abweichungen der Systemempfehlung wurde von der behandelnden Ärztin geprüft, ob die alternative Therapie des Empfehlungssystems plausibel ist.

Die Übereinstimmung der Software mit der behandelnden Ärztin lag bei 64%.

Zur Beurteilung der Praktikabilität wurde der zeitliche Aufwand für die Eingabe gemessen, um zu ermitteln, ob aufgrund der guten Handhabung eine Nutzbarkeit im Praxisalltag zu erwarten ist. Die Eingabe der Daten in das Therapieempfehlungssystem dauerte je nach Umfang der Nebendiagnosen und Vortherapien etwa 3 bis 10 Minuten, die Generierung der entsprechenden Therapieempfehlung erfolgte binnen weniger Sekunden. Bis auf wenige Ausnahmen lief das System stabil mit konstanten Empfehlungen.

Die detaillierten Ergebnisse der klinischen Validierung des entwickelten Prototyps werden im Dezember 2017 im Rahmen der Promotion der Studienmitarbeiterin und Koautorin des Berichts, Frau Stefanie Beckert veröffentlicht.

4 Diskussion

4.1 Limitationen und Stärken

Die eigenen Arbeiten belegen, dass Therapieempfehlungssysteme möglich sind. Der Ansatz zur Verwendung von Recommender Systemen und zur Einbeziehung von Demografischen Daten und Outcome-Daten hat sich insofern bewährt, da beide Nutzinformationen enthält und insbesondere im Falle von erfolgreichen Therapien eine gute Übereinstimmung zum Goldstandard erzielt wurde (siehe Abbildung 6). Aus wissenschaftlicher Sicht von Interesse ist, dass Recommender Systeme im Kontext der Therapieempfehlung noch nicht weit verbreitet sind. Das Vorhaben hat diesbezüglich zu einer maßgeblichen Innovation geführt. Da Recommender Systeme zahlreiche Variationen erlauben, ergeben sich aus dem nutzbar gemachten Ansatz auch für die Zukunft noch Möglichkeiten für Verbesserungen auf Basis des verfolgten Ansatzes.

Als sehr kritisch im Hinblick auf die erzielten Ergebnisse und die Therapieempfehlung im Allgemeinen sind die hohen Abweichungen zwischen verschiedenen Experten zu sehen. Es ist davon auszugehen, dass ein automatisiertes System oftmals von der Meinung des bewertenden Experten abweicht. In diesen Fällen ist die Plausibilisierung der automatisierten Entscheidung von großer Bedeutung. Der gewählte Ansatz bietet dazu die Möglichkeit (durch die Vorstellung der Fälle, auf deren Grundlage die Entscheidung getroffen wurde), bedarf allerdings weiterer Verbesserungen, da im Status quo zahlreiche andere Patienten die Entscheidung beeinflussen und eine umfassende Vorstellung kaum möglich ist.

Aus klinischer Sicht ist an den vorliegenden Prototypen noch die Detailgenauigkeit der Informationen zu Patientencharakteristika und Vortherapien zu optimieren, damit das System noch präzisere Therapieempfehlungen abgeben kann. Hinzu kommt, dass in der Zwischenzeit zwei neue Substanzgruppen zur Therapie der Psoriasis zugelassen wurden, welche bisher nicht im System vorkamen. Weiterhin wurden in der Zwischenzeit für einzelne Präparate sog. „first-line-Zulassungen“ erteilt, so dass die formulierten Kriterien die als Grundbedingung zur Therapieempfehlung hinzugezogen werden, nun nicht mehr komplett zutreffen. Da auch weiterhin ein großer Wandel in der Zulassung weiter Medikamente zu erwarten ist und auch neue Erkenntnisse bezüglich eingesetzter Therapien hinzukommen, wäre es erforderlich, dass das System stets mit neuen wissenschaftlichen Erkenntnissen gespeist wird.

Im klinischen Alltag ist aufgrund der hohen Patientendichte jeder zusätzliche Zeitaufwand eine große Hürde. Aus diesem Grund ist eine zusätzliche Eingabe der erforderlichen Daten in dem Empfehlungssystem in der Praxis schwer umsetzbar. Durch die Meidung der Doppeldokumentation durch eine Integration der erforderlichen Angaben in die Praxissoftware könnte eine höhere Akzeptanz im Praxisalltag zu erwarten sein.

4.2 Schlussfolgerungen

Das Projekt Therapieempfehlungssystem für Psoriasis konnte erfolgreich in der vorgesehenen Projektlaufzeit durchgeführt werden. Wie im Antrag vorgesehen wurde der Prototyp einer Software entwickelt und im Anschluss anhand 67 Patienten der Poliklinik für Dermatologie des Universitätsklinikum Carl Gustav Carus getestet. Die Ergebnisse zeigen, dass der Prototyp bereits plausible Therapieempfehlungen generieren kann. Eine Weiterentwicklung der Software insbesondere im Hinblick auf die den Einbezug externer Evidenz wird jedoch empfohlen. Zudem gelangen mehrere wissenschaftliche Veröffentlichungen zum Thema [19]–[21], die als Grundlage für zukünftige Arbeiten/Antragstellungen genutzt werden können.

5 Literatur

- [1] D. L. Sackett, W. M. Rosenberg, J. A. Gray, R. B. Haynes, and W. S. Richardson, "Evidence based medicine: what it is and what it isn't.," *BMJ*, vol. 312, no. 7023, pp. 71–2, Jan. 1996.
- [2] J. Schmitt, "Behandlungsziele und Zufriedenheit mit der medizinischen Versorgung von Erwachsenen mit Neurodermitis.," in *Prävention und Versorgungsforschung*, W. Kirch, B. Badura, and H. Pfaff, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 819–32.
- [3] M. M. Schmitt J, Schmitt NM, Maywald U, Kirch W, "Health services utilization and treatment of elderly patients with atopic eczema.," *J Invest Dermatol*, vol. 128 (Suppl, p. 2555, 2008.
- [4] J. Schmitt, N. M. Schmitt, W. Kirch, and M. Meurer, "Ambulante Versorgung und medikamentöse Behandlung von Kindern und Erwachsenen mit atopischem Ekzem," *J. der Dtsch. Dermatologischen Gesellschaft*, vol. 7, no. 4, pp. 345–352, Apr. 2009.
- [5] J. Schmitt, N. M. Schmitt, W. Kirch, and M. Meurer, "Outpatient care and medical treatment of children and adults with atopic eczema.," *J. Dtsch. Dermatol. Ges.*, vol. 7, no. 4, pp. 345–51, Apr. 2009.
- [6] J. Schmitt, N. M. Schmitt, W. Kirch, and M. Meurer, "[Significance of atopic dermatitis in outpatient medical care. Analysis of health care data from Saxony].," *Hautarzt.*, vol. 60, no. 4, pp. 320–7, Apr. 2009.
- [7] J. Schmitt, M. Meurer, U. Schwanebeck, X. Grahlert, and K. Schakel, "Treatment following an evidence-based algorithm versus individualised symptom-oriented treatment for atopic eczema. A randomised controlled trial.," *Dermatology*, vol. 217, no. 4, pp. 299–308, Jan. 2008.
- [8] S. K. Schmitt J, Meurer M, Schwanebeck U, Grählert X, "Effektivität einer standardisierten, zeitlich determinierten und score-orientierten Behandlung des moderaten bis schweren atopischen Ekzems verglichen mit individueller symptom-orientierter Behandlung - Eine randomisierte kontrollierte Studie.," *Allergo J.*, vol. 16[Suppl.], p. P66, 2007.
- [9] D. L. Sackett, "Evidence-based medicine and treatment choices.," *Lancet*, vol. 349, no. 9051, p. 570; author reply 572-3, Feb. 1997.
- [10] M. Ryan and S. Farrar, "Using conjoint analysis to elicit preferences for health care.," *BMJ*, vol. 320, no. 7248, pp. 1530–3, Jun. 2000.
- [11] G. Meyfroidt, F. Güiza, J. Ramon, and M. Bruynooghe, "Machine learning techniques to examine large patient databases.," *Best Pract. Res. Clin. Anaesthesiol.*, vol. 23, no. 1, pp. 127–43, Mar. 2009.
- [12] P. Gueth *et al.*, "Machine learning-based patient specific prompt-gamma dose monitoring in proton therapy.," *Phys. Med. Biol.*, vol. 58, no. 13, pp. 4563–77, Jul. 2013.
- [13] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*,

- vol. 7, pp. 1–30, 2006.
- [14] T. Hofmann, B. Schölkopf, and A. J. Smola, “Kernel methods in machine learning,” *Ann. Stat.*, vol. 36, no. 3, pp. 1171–1220, Jun. 2008.
 - [15] R. Polikar, “Ensemble based systems in decision making,” *IEEE Circuits Syst. Magaine*, vol. 6, no. 3, pp. 21–45, 2006.
 - [16] L. Breiman, “Bagging Predictors,” *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.
 - [17] T. K. Ho, “The random subspace method for constructing decision forests,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, 1998.
 - [18] Y. Freund and R. E. Schapire, “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, ” *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, 1997.
 - [19] F. Gräßer *et al.*, “Application of recommender system methods for therapy decision support,” in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 2016, pp. 1–6.
 - [20] F. Gräßer *et al.*, “Therapy Decision Support Based on Recommender System Methods,” *J. Healthc. Eng.*, vol. 2017, pp. 1–11, 2017.
 - [21] F. Gräßer *et al.*, “Neighborhood-based Collaborative Filtering for Therapy Decision Support,” in *International Workshop on Health Recommender Systems*, 2017.