Diplomarbeit

ALKOHOLSTATUSEXTRAKTION UND STANDARDISIERUNG MIT HILFE MASCHINELLER LERNVERFAHREN AUS FREITEXTLICH-KLINISCHER ROUTINEDOKUMENTATION

eingereicht von

Leon Magnus Potjan

zur Erlangung des akademischen Grades

Doktor(in) der gesamten Heilkunde

(Dr. med. univ.)

an der

Medizinischen Universität Graz

ausgeführt am

Institut für Medizinische Informatik, Statistik und Dokumentation

unter der Anleitung von

Ass.-Prof. Dipl.-Ing. Dr. scient. med. Markus Eduard Kreuzthaler

Univ.-Prof. Dr. med. Stefan Schulz

Graz, 18.06.2023

Eidesstattliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am 18.06.2023 Leon Potjan eh.

Danksagung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich während der Anfertigung dieser Diplomarbeit unterstützt und motiviert haben.

Mein besonderer Dank geht an Ass.-Prof. DI Dr. Markus Kreuzthaler, der meine Diplomarbeit betreut und begutachtet hat. Seine ständige Erreichbarkeit, der großzügige zeitliche Freiraum und die konstruktive Kritik hat mir bei der Erstellung dieser Arbeit sehr geholfen.

Ich bedanke mich zudem bei Prof. Dr. med. Stefan Schulz, der mir gerade zu Beginn dieser Diplomarbeit beratend zur Seite stand. Ihm ist zudem der „letzte Schliff“ dieser Arbeit zu verdanken.

Ein besonderer Dank gilt außerdem Dr. med. univ. Larissa Hammer, die mich bei der Durchführung des praktischen Teils dieser Arbeit mit ihrer Expertise tatkräftig unterstützt hat.

Abschließend möchte ich mich bei meinen Eltern bedanken, die mir stets mit Rat und Tat zur Seite stehen.

Leon Potjan

Graz, 18.06.2023

Inhaltsverzeichnis

Abkürzungen und deren Erklärungen

cNLP:	clinical natural language processing
AUDIT:	alcohol use disorder identification test
FAST:	fast alcohol screening test
CAGE:	cutting down, annoyance by criticism, guilty feeling, and eye-openers
DSM:	Diagnostic and Statistical Manual of Mental Disorders
ICD:	International Statistical Classification of Diseases and Related Health Problems
SNOMED CT:	Systematized Nomenclature of Medicine Clinical Terms

Tabellenverzeichnis

Zusammenfassung

Das geschriebene Wort nimmt im medizinischen Sektor besonders in der klinischen Dokumentation einen hohen Stellenwert ein. Dokumentieren dient dem Zusammentragen, Erschließen und Nutzbarmachen von Informationen. Es erfüllt im klinischen Alltag verschiedenste Aufgaben, und hat im Gesundheitssektor zur Akkumulation gigantischer Datenmengen geführt (auch *Big Data* genannt). Diese liegen größtenteils unstrukturiert in Form von Freitexten vor, und lassen sich somit schwer von traditionellen, computergestützten Verfahren auswerten. Clinical Natural Language Processing (cNLP) als Schnittpunkt der Fachgebiete Künstliche Intelligenz und Computerlinguistik versucht an dieser Stelle mittels maschineller Lernverfahren den Informationsgehalt dieser Datenberge zu erschließen.

Ziel dieser Diplomarbeit ist es, Unterschiede zwischen einer manuellen und einer cNLP-basierten Analyse, sowie Klassifikation von unstrukturierten Daten aufzuzeigen bzw. zu bewerten. Hierfür werden Patientinnen und Patienten hinsichtlich ihres dokumentierten Alkoholkonsums unterschieden. Der Alkoholkonsum als Klassenvariable wurde aufgrund seiner standardisierten Erhebung während der Anamnese und seinen assoziierten gesundheitlichen Risiken gewählt.

Ein initialer Datensatz, bestehend aus 47.600 de-identifizierten Arztbriefen der Fachbereiche Kardiologie, Dermatologie und Onkologie eines österreichischen Krankenhauses, wird hierfür zunächst in 1429 Arztbriefausschnitte (Snippets) umgewandelt. Hierbei kommen simple, regelbasierte NLP-Verfahren, wie beispielsweise die Schlüsselwortsuche, zum Einsatz. Anschließend wird mittels passender Veröffentlichung und bestehender Kodiersysteme für Alkoholkonsum (DSM 5, ICD-10, SNOMED CT) ein eigenes Einteilungsschema erstellt. Es folgt die, dem erstellten Schema entsprechende, manuelle Klassifizierung der Snippets durch den Autor. 20% der Snippets werden erneut durch eine zweite wissenschaftliche Mitarbeiterin annotiert. Ein Cohen’s Kappa von 0,9 zeigt eine gute Übereinstimmung der beiden Annotationsdurchgänge. In weiterer Folge werden zwei maschinelle Lernverfahren aus dem Bereich des cNLP miteinander verglichen: fastText (basierend auf einem neuronalen Netzwerk) und Support Vector Machines (SVM). Des Weiteren wurde bei fastText untersucht, welchen Einfluss ein vorab trainiertes

Sprachmodell auf die Qualität der Klassifikationsaufgabe nimmt. Die parameteroptimierte SVM erzielt einen gemittelten Makro-F_1-Score von 0,83 auf den Testdatensatz. Im Gegensatz dazu erreicht fastText bei Verwendung eines vorab trainierten Sprachmodelles einen gemittelten Makro-F_1-Score von 0,78 und ohne Verwendung eines Sprachmodells einen gemittelten Makro-F_1-Score von 0,80. Diese Werte entsprechen Ergebnissen von Veröffentlichungen mit ähnlicher Forschungsfrage, wobei das hier untersuchte Verfahren mit technisch weniger Aufwand auskommt als bereits untersuchte Methoden.

Abstract

The written word plays a particularly vital role in healthcare. Documentation is intended to collect, index and mine information. It fulfills a variety of tasks in everyday clinical practice and causes the accumulation of datasets in the healthcare sector (also known as Big Data). This data exists largely in the form of unstructured text and is therefore difficult to analyze using traditional computer-based methods. Clinical Natural Language Processing (cNLP), as the intersection of artificial intelligence and computational linguistics, attempts to extract the information of these datasets using machine learning.

This thesis aims to demonstrate and evaluate the differences between manual and cNLP-based selection and classification of unstructured data. For this, patients are differentiated based on their documented alcohol consumption. Alcohol consumption as a class variable was chosen due to its standardized assessment during patient history taking and its associated health risks.

An initial dataset consisting of 47,600 de-identified discharge letters from the departments of cardiology, dermatology and oncology of an Austrian hospital is first transformed into 1429 text snippets. Rule-based NLP methods, such as keyword search, are used for this task. Subsequently, a classification scheme for alcohol consumption is created using suitable publications and existing coding systems (DSM 5, ICD-10, SNOMED CT). The snippets are then manually annotated, according to the created scheme. 20% of the snippets are annotated again by a second research assistant. A Cohen's Kappa of 0.9 shows good agreement between the annotations. Subsequently, two machine learning methods from the field of cNLP are compared: fastText (based on a neural network) and SMV (support vector machine). For fastText, the influence of a pre-trained language model on the quality of the classification task was investigated. The parameter-optimized SVM achieves an averaged macro F_1-score of 0.83 on the test data set. In contrast, fastText achieves an averaged macro F_1-score of 0.78 when using a pre-trained language model and 0.80 without using a language model. These values correspond to the results of publications with a similar research question, although the method investigated here involves less technical effort than previous ones.

Bereits erfolgte Veröffentlichungen

Kugic A, Potjan LM, Hammer LM, Schulz S, Kreuzthaler M. Alcohol Status Standardization from Clinical Real World Data with Transformer Architectures. In: 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI). 2022. S. 233–8.

Einleitung

Als ältestes Schriftsystem der Menschheit gilt die sumerische Keilschrift, die Ende des vierten Jahrtausends vor Christus in Südmesopotamien entstand (1 [S. 33], 2).

Das Aufkommen der Schrift stellte für die Entwicklung der menschlichen Kultur einen entscheidenden Meilenstein dar. Durch das Festhalten von Informationen konnten Erkenntnisse besser vermittelt und längerfristig fixiert werden. Die begrenzte Speicherfähigkeit des menschlichen Gedächtnisses erwies sich somit nicht mehr als limitierender Faktor der Wissensübermittlung (3 [S. 9 f.]).

Auch nach 5000 Jahren hat die Schrift nicht an Bedeutung verloren.

Gerade im Bereich der medizinischen Versorgung besitzt das geschriebene Wort eine besondere Bedeutung. Hier fungiert die Schrift als wichtiges Werkzeug zur Kontrolle von Krankheitsverläufen, z.B. in Form von Pflegeberichten, als Medium für den interdisziplinären Informationsaustausch zwischen verschiedenen Fachrichtungen, oder als rechtliche Absicherung in Form von Aufklärungsbögen. Es gilt „wer schreibt, der bleibt“, was die Alternativlosigkeit der schriftlichen Dokumentation im medizinischen Alltag hervorhebt.

Im medizinischen Bereich generierte, schriftliche Dokumente sind auch in wissenschaftlicher Hinsicht von größter Bedeutung. Retrospektive Studien in der Medizin basieren auf selektierten dokumentierten Krankheits- und Behandlungsverläufen. Sie sind somit direkt abhängig von der zur Verfügung stehenden medizinischen Dokumentation. Diese liegt meist in Form von Freitexten, also schriftlichen Dokumenten ohne klare Struktur vor.

Die Analyse dieser Texte wird traditionell von wissenschaftlichen Mitarbeitern vollzogen und ist, je nach Studienumfang, mit enormen Zeit- und Kostenaufwand verbunden.

Clinical Natural Language Processing, kurz cNLP, kann diesen Aufwand verringern (4). Durch maschinelles Verarbeiten textueller Daten wird hierbei die Extraktion relevanter Informationen unterstützt und durch den Benutzer im Kontext der zu unterstützenden Aufgabenstellung evaluiert.

In dieser Arbeit wird anhand einer konkreten Aufgabenstellung untersucht, inwieweit eine manuell durchgeführte Klassifikation mit der eines cNLP-Systems übereinstimmt. Hierfür werden Angaben zum Alkoholkonsum, die sich in Arztbriefen finden lassen, sowohl manuell, als auch mittels cNLP, ausgelesen und in vorgegebene Klassen eingeteilt.

Die vorliegende Diplomarbeit strukturiert sich wie folgt:

Im ersten Kapitel werden zunächst die verschiedenen Gründe der medizinischen Datenerhebung aufgezeigt, bevor erklärt wird, wie sich medizinische Daten einteilen lassen. Schließlich wird erläutert, worum es sich bei *Big Data* handelt, und welches Potential *Big Data* im Gesundheitswesen mit sich bringt. Es folgt die Definition wichtiger Begriffe des Themengebietes Natural Language Processing (NLP) und die Beschreibung der Funktionsweise eines klassischen NLP-Systems. In weiterer Folge wird definiert, was Clinical NLP (cNLP) darstellt, und es werden Beispiele angeführt, die zeigen, wie und wo cNLP bereits heute im klinischen Alltag Anwendung findet. Auf die Dokumentation des Alkoholkonsums als Teil der standardisierten Anamnese wird als nächstes eingegangen. Verschiedene Möglichkeiten der Erhebung des Alkoholstatus werden vorgestellt und die bekanntesten Kodiersysteme des Alkoholkonsums beschrieben. Anschließend wird der leitliniengerechte Standard der Alkoholanamnese präsentiert. Es folgt die Zusammenfassung der aktuellen Literatur zur NLP-basierten Analyse von medizinischen Freitexten bezüglich des Alkoholkonsums und motiviert die Zielsetzung der Arbeit am Ende dieses Kapitels.

Im Kapitel „Material und Methoden“ werden in weiterer Folge die Datenbasis, die zur Klärung der Forschungsfrage verwendet wurde, und deren Modifikationen beschrieben. Anschließend wird aufgezeigt, welche Einflüsse zur Entstehung des Klassifikationsschemas dieser Diplomarbeit beigetragen haben, und welche Klassen dieses Schema umfasst. Nachdem die manuelle Annotation in ihren Details beschrieben wurde, folgt die Erläuterung der verwendeten Methoden. Es schließt sich eine kurze Erklärung der Vorprozessierung des Datensatzes, sowie die Beschreibung der verwendeten maschinellen Lernverfahren, an. Das Kapitel endet mit den Definitionen der Validierungsparameter der Klassifikationsergebnisse. Hier werden jeweils Beispiele zur Veranschaulichung angegeben.

Anschließend werden die Ergebnisse der Diplomarbeit angeführt und diskutiert, und abschließend in der Konklusion die Limitationen und mögliche Ausblicke besprochen.

Daten und ihre Formen in der Medizin

Medizinische Dokumentation

Man schätzt, dass im Jahr 2012 die Menge an vorliegenden Daten in der Gesundheitsversorgung weltweit etwa 500 Petabyte betrug. Diese Menge soll sich im Jahr 2020 auf 25000 Petabyte verfünfzigfacht haben (5,6).

Wenn davon ausgegangen wird, dass das gesamte menschliche Genom etwa 3 Gigabyte Speicherkapazität benötigt (7), entsprechen 25000 Petabyte dem Erbgut von etwa 8,3 Milliarden Menschen. Dies entspricht also dem Erbgut der gesamten aktuellen Menschheit.

Andere Quellen führen sogar noch größere Zahlen auf und schätzen die Datenmenge des amerikanischen Gesundheitssektors im Jahr 2011 allein auf 150 Exabytes (entspricht 150.000 Petabytes). Sie prophezeien ein Wachstum bis hin zu Yottabyte (10^{24} Gigabyte) (8,9).

Ein solches Wachstums an medizinischen Daten erscheint nicht realistisch vor dem Hintergrund, dass heutzutage, allein in den USA, etwa eine Milliarde Patientenbesuche pro Jahr in Form von elektrischen Dokumenten festgehalten werden (10).

Um nun zu verstehen, wie und warum es zur Akkumulation solcher Datenmengen kommen kann, muss zunächst die Sinnhaftigkeit der Dokumentation im Allgemeinen begriffen werden.

Dokumentieren bedeutet das Zusammentragen, Erschließen und (geordnetes) Aufbewahren von Informationen, mit dem Ziel, diese zu einem späteren Zeitpunkt für ein gegebenes Ziel nutzbar zu machen, bzw. aufgrund relevanter Informationen aus der Dokumentation die Entscheidungsunterstützung des ärztlichen Handelns qualitativ hochwertig zu gestalten.
Gründe für die medizinische Dokumentation können unterschiedlichster Natur sein.

Zunächst muss die wirkungsvolle und angemessene Patientenversorgung gewährleistet werden. Diese ist nur möglich, wenn den berechtigten Personen alle relevanten Informationen (z.B. zu Patienten/Patientinnen oder Behandlungen) zur richtigen Zeit, am richtigen Ort und in der richtigen Form zur Verfügung gestellt werden können.

Neben der Patientenversorgung erfüllt die Dokumentation im medizinischen Bereich noch administrative und rechtliche Aufgaben. So ist die Leistungsabrechnung im medizinischen Sektor von der Dokumentation durchgeführter Maßnahmen abhängig, und auch die rechtliche Absicherung der Versorgungseinrichtung ist durch zahlreiche Dokumentations- und Meldepflichten geregelt. Zusätzlich ist die medizinische Dokumentation auch wichtige Ressource für die Qualitätssicherung, sowie für Aus-, Fort- und Weiterbildung in klinischen Berufen. Durch das Festhalten von Krankheitsverläufen wird eine nachträgliche, kritische Selbstreflexion sowie ein Qualitätsmonitoring möglich. Außerdem kann anhand der dokumentierten Verläufe eine Bewertung der Fähigkeiten von Auszubildenden und Fachpersonal erfolgen. Auch die Schulung anhand echter Verlaufsbeispiele ist nur durch eine ausreichende medizinische Dokumentation möglich. Schließlich ist das Dokumentieren in Versorgungseinrichtungen auch essenzieller Bestandteil klinisch-wissenschaftlicher Forschung. So werden im Verlauf retrospektiver Studien große Mengen von Krankheitsverläufen ausgewertet, um zur Beantwortung der jeweiligen Forschungsfragen beizutragen (11 [S. 1-6]).

Formate medizinischer Dokumentation

Bevor die Einteilung medizinischer Daten ergründet werden kann, empfiehlt es sich, die Aufteilung von Daten im Allgemeinen zu beleuchten.

Grundsätzlich werden Daten sowohl hinsichtlich ihrer Struktur als auch ihrer Standardisierung unterschieden.

Strukturell werden Daten in drei Klassen eingeteilt: strukturierte, semistrukturierte, und unstrukturierte Daten (12 [S. 31], 13 [S. 146]).

Strukturierte Daten werden durch Metadaten (13 [S. 146]) und insbesondere auch durch die Dokumentenstruktur (bspw. XML-Format) (14) festgelegt. Die Metadaten können sich beispielsweise auf die Einheit und/oder auf die erlaubten Werte der

Daten beziehen. Beispiele für strukturierte Daten in der Medizin sind Vitalparameter wie Herz- und Atemfrequenz, Blutdruck und Temperatur.

Semistrukturierte Daten können strukturierte Bestandteile enthalten, weisen in ihrer Gesamtheit aber keine eindeutige Struktur auf (13 [S. 146]). Die Dokumentenstruktur von XML-Daten, die häufig in der medizinischen Dokumentation und Kommunikation vorkommen, sowie mit ICD-10 versehene Fließtexte, erfüllen diese Kriterien (15).

Bei unstrukturierten Daten ist der Informationsgewinn sehr stark vom Informationsempfänger abhängig (13 [S. 146]). Beispiele für unstrukturierte Daten in der Gesundheitsversorgung sind Arztbriefe und Pflegeberichte, aber auch Bild- oder Videodateien.

Man geht davon aus, dass medizinische Daten zu etwa 80-90% unstrukturiert vorliegen (16,17).

Standardisierte Daten zeigen einen einheitlichen Aufbau hinsichtlich der dokumentierten Merkmale von Datenobjekten in einem Objekttyp. Standardisiert wird dabei der Datenobjekttyp (z.B. Text-, Bild- oder Video-Dateien), die Merkmalsarten (z.B. der Ernährungszustand) und deren Merkmalausprägungen (z.B. in diesem Fall kachektisch, mager, normalgewichtig, übergewichtig, oder adipös).
Eine Standardisierung ermöglicht in diesem Fall eine Vergleichbarkeit von Datenobjekten auf zwei Ebenen, nämlich formal und inhaltlich.
Dabei bezieht sich die formale Ebene auf die Vergleichsmerkmale und ihre spezifischen Bezeichnungen. Sie gewährleistet das Festhalten aller Vergleichsmerkmale jedes Objekts und die Verwendung der gleichen Bezeichnungen.
Die inhaltliche Ebene beschreibt die jeweiligen Merkmalsausprägungen und bildet somit den Kontext der erhobenen Daten ab.

Nicht-standardisierte und vollständig standardisierte Daten sind in der medizinischen Dokumentation eher selten. Viele medizinische Dokumente weisen einen teilstandardisierten Aufbau auf, mit standardisierten Teilaspekten (z.B. Datum oder Fallnummer) und nicht-standardisierten Abschnitten, wobei Letztere vor allem

für die feingranuläre Dokumentation von Besonderheiten genutzt werden. (11 [S. 25 f.])

An dieser Stelle sei der Vollständigkeit halber der Begriff der Semantik, bzw. des semantischen Datenmodells kurz dargestellt.

Das semantische Datenmodell nutzt eben jenen oben beschriebenen Kontext, der sich aus der inhaltlichen Analyse standarisierter Daten ergibt, um Datenobjekte miteinander in Beziehung treten zu lassen. Somit lassen sich Abhängigkeiten zwischen den Datenobjekten der „echten Welt“ (*external level*) durch ein bestimmtes Informationsmodel (*conceptual level*), wie beispielsweise das Entity-Relationship Model, in ein abstraktes Model der physischen Datenbank *(internal level*) überführen. (18)

Nicht standardisierte sowie unstrukturierte Daten müssen vor ihrer Verarbeitung durch Maschinen in ein von Computern bearbeitbares Format gebracht werden. Sie stellen eine große Herausforderung für die computergestützte Datenanalyse z.B. textueller Daten durch NLP dar.

Big Data in der Medizin

Wie bereits erwähnt, generiert das Gesundheitswesen weltweit riesige Mengen an Daten. Eine rein manuelle Verwertung dieser „*Big Data*“ ist längst unmöglich geworden. So entsteht zunehmend Bedarf an Informationstechnologien, um den Informationsgehalt dieser Daten zugänglich zu machen (19).

Zum tieferen Verständnis und zur Vervollständigung des Themengebiets *Daten und ihre Formen in der Medizin* werden nun einige Grundbegriffe aus dem Themengebiet *Big Data* erläutert und einige Anwendungsgebiete aufgeführt, in denen *Big Data* im Gesundheitswesen eine Rolle spielt.

Definition von Big Data

Big Data zu definieren, erweist sich als schwieriger als erwartet. Die Literatur weist dabei unterschiedliche Ansätze auf, wobei De Mauro et al. (20) erstmals versuchen, einen einheitlichen Konsens zu schaffen. Sie definieren *Big Data* als eine Datenmenge, die sich durch vier Eigenschaften auszeichnet, und die spezielle Technologien und Analysemethoden für ihre Umwandlung in einen Mehrwert

erfordert. Diese vier Eigenschaften („großen V's") von *Big Data* sind in Anlehnung an (20):

Als Volumen (**V**olume) bezeichnet man die gewaltige Menge an Daten, die *Big Data* ihren Namen gibt, und von Terabytes bis Zetabytes reichen kann.

Die Geschwindigkeit (**V**elocity) bezieht sich auf den Anteil der generierten Daten, der zeitabhängig ist und in Echtzeit prozessiert und aktualisiert werden muss. (Solche Daten entstammen häufig Sensoren.)

Die Vielfalt (**V**ariety) beschreibt die Heterogenität der Daten. Damit wird auf die mögliche Strukturierung (unstrukturiert, semistrukturiert, strukturiert), Standardisierung, aber auch auf die verschiedenen Datenformate (Text, Bild oder Video) Bezug genommen.

Mit Mehrwert (**V**alue) wird schließlich das Potential angesprochen, das die Analyse von *Big Data* mit sich bringt. Im medizinischen Sektor besteht der Mehrwert der Big-Data-Analyse in verbesserter Patientenversorgung, Qualitätssicherung und Kostenersparnis (8,20).

Basierend auf dieser Beschreibung sind heute zahlreiche Neuauflagen der ursprünglichen Definition zu finden, die *Big Data* um zusätzliche Eigenschaften ergänzen. Bis zu 10 verschiedene „V"s werden in manchen Fällen angeführt, um *Big Data* zu beschreiben (Value, Velocity, Verification, Variability, Validity, Viscosity, Volatility, Visualization, Virility, Valence) (8,21).

Big Data im Gesundheitswesen

Die primären Ziele der Nutzung von *Big Data* im Gesundheitssektor lassen sich grob in zwei Kategorien aufteilen: Verbesserung der Patient:innenversorgung und Zufriedenheit sowie die Steigerung der Effizienz im Gesundheitssektor (mit einhergehenden Kostenersparnissen).

Zur Umsetzung dieser Ziele können verschiedene Anwendungsgebiete beitragen:

Bei *Healthcare Monitoring* wird über die Einbindung von autonomer Monitorisierung des Patienten / der Patientin in den Therapieplan (beispielsweise durch Smartwatches, intelligente Blutmessgeräte oder Tablettenspender) eine

Verbesserung der Therapie, der Prognose, sowie der Theapieadherence angestrebt.

Werkzeuge, die wie „*MedAware*“ KI-Methoden verwenden, nutzen große Datenmengen, um Medikamenteninteraktionen frühzeitig zu erkennen und möglicherweise gefährliche, unerwünschte Arzneimittelwirkungen zu verhindern.

Bei *Healthcare Prediction* wird versucht, eine Vorhersage über den gesundheitlichen Zustand des Patienten / der Patientin sowie mögliche Krankheitstrigger zu treffen (22). In diesem Zusammenhang werden bestehende medizinische Daten zusammengetragen (z.B. Daten des Genoms), eventuell sogar mit durch den Patienten / die Patientin selbst geposteten Social-Media-Beiträgen kombiniert, und schließlich hinsichtlich möglicher Muster analysiert. So soll zur primären Prävention oder zur rechtzeitigen Therapie von Erkrankungen beigetragen werden.

Bei *Performance Enhancements* und smarten *Healthcare Management Systems* steht die Effizienz des Gesundheitssektors im Vordergrund. Basierend auf der automatisierten Analyse von Krankenhausdaten (beispielsweise einer Klinik-Notaufnahme oder eines Arzt-Patienten-Gesprächs) werden hierbei neue Arbeitsmodelle sowie -abläufe vorgeschlagen, um in weiterer Folge zu verkürzten Wartezeiten, verbessertem Zeitmanagement bis zur Behandlung oder, allgemein, zu effizienteren Abläufen im Krankenhaus beizutragen. Somit kann die Zufriedenheit von Patienten / Patientinnen wie auch von Mitarbeitern/ Mitarbeiterinnen gesteigert werden.

Bei *Recommendation Systems* und *Healthcare Knowledge Systems* handelt es sich um Informationstechnologien und -programme, die medizinisches Personal bei der Entscheidungsfindung unterstützen. Basierend auf der Analyse von *Big Data* können solche Systeme Diagnosen, Medikationen und andere Therapien vorschlagen (8).

Zusammenfassend ergibt sich: Möglichkeiten zur Anwendung von *Big Data* im Gesundheitsbereich sind zahlreich. Das spiegelt sich in einer vermehrten internationalen Fokussierung auf die Entwicklung von Methoden und Techniken zur Analyse von medizinischen Daten wider. Zu den vielversprechenden Methoden, die zunehmend ins Rampenlicht der internationalen Forschung rücken, zählt hierbei

unter anderem Natural Language Processing (NLP), das einen besseren Zugang zu relevanter Information in unstrukturierten Textdokumenten anstrebt (23,24).

Natural Language Processing (NLP)

NLP befasst sich mit der Interaktion zwischen Computern und Menschen auf natürlichsprachlicher Ebene und stellt heutzutage eine interdisziplinäre Kombination aus Computerlinguistik und künstlicher Intelligenz dar. Unter natürlicher Sprache versteht man menschliche Sprache (z.B. Deutsch oder Englisch) im Gegensatz zu künstlichen Sprachen wie Programmiersprachen.

Ursprünglich aus dem Wunsch entstanden, das Übersetzten von Sprache zu automatisieren, versucht NLP sowohl die Struktur (Syntax), als auch die Bedeutung (Semantik) eines natürlichsprachlichen Inputs zu erschließen. In weiterer Folge sollen die Ziele und Überzeugungen der Kommunizierenden, aber auch der Sprachvorgang und die Gesprächsstruktur im Allgemeinen, erschlossen werden.

Somit gilt als primäre Zielsetzung des NLP die Schaffung von intelligenten Systemen, die idealerweise ein menschliches Verständnis von natürlicher Sprache (*natural language understanding*) aufweisen.

Praktische Ziele bzw. Anwendungsgebiete solcher Systeme lassen sich wie folgt unterteilen:

- *Information Retrieval* - das Wiederauffinden von bestimmten Textdaten aus einer Datenmenge.

- *Information Extraction* - das Erkennen, Annotieren und Extrahieren bestimmter Schlüsselelemente eines Textdokuments und die anschließende strukturierte Repräsentation der Suchergebnisse.

- *Question-Answering* - die direkte Beantwortung einer Userfrage mittels Text oder die Präsentation des Textabschnittes, der die Antwort beinhaltet.

- *Summarization* - das Zusammenfassen eines umfangreichen Textes in eine abgekürzte, aber hochgradig strukturierte, narrative Darstellung des Originaldokuments.

- *Machine Translation* - das automatisierte Übersetzen von natursprachlichen Textdokumenten.

- *Dialogue Systems* - die Dialogführung mit einem NLP-System.

Neben den informationstechnologischen Bestrebungen des NLP soll uns dieser Fachbereich aber auch dabei helfen zu verstehen, wie Menschen unter Verwendung natürlicher Sprache miteinander kommunizieren.

Hierbei müssen einige Schwierigkeiten überwunden werden, die sich im Zusammenhang mit der Mehrdeutigkeit von natürlicher Sprache ergeben. Die Linguistik unterscheidet mehrere Formen der Mehrdeutigkeit:

Lexikalische Mehrdeutigkeit. Ein Lexem besitzt mehrere Bedeutungen, wie beispielsweise „Bank“. In der Medizinsprache kommt lexikalische Mehrdeutigkeit vor allem bei Abkürzungen zum Tragen: „HWI“ als „Hinterwandinfarkt“ oder „Harnwegsinfekt“).

Syntaktische Mehrdeutigkeit. Ganze Sätze können unterschiedlich verstanden werden. (So lässt der Satz „Der Arzt untersuchte den Patienten mit Tinnitus.“ auch die Lesart zu, dass der Arzt unter Tinnitus leidet).

Referentielle Mehrdeutigkeit. Pronomen, die im Verlauf eines Diskurses anstelle von Nominalphrasen benutzt werden, können zu diesem Typ der Mehrdeutigkeit führen. (z.B.: „Als der Arzt den Patienten untersuchte, erlitt er einen plötzlichen Herztod.“)

Pragmatische Mehrdeutigkeit. Viele Sätze in der natürlichen Sprache werden mit einer gewissen Intention kommuniziert. Das fehlende Verständnis dieser Intention führt zur pragmatischen Mehrdeutigkeit. (z.B.: „Können sie Blut sehen?“ bei der Blutentnahme). (25 [S. 4-9,15,16], 26)

Definitionen

Bevor Aufbau und Funktionsweise eines allgemeinen, sowie eines cNLP-Systems veranschaulicht werden kann, werden in diesem Abschnitt essenzielle Fachtermini wie Algorithmus, Parser, maschinelles Lernen und neuronales Netzwerk erläutert.

Algorithmus

Ein Algorithmus besteht aus einer Folge von definierten Rechenschritten, die sich automatisieren lassen und dabei eine Eingabe (Input) zu einer Ausgabe (Output) verarbeiten. Diese Prozessierung muss sich in endlich vielen Schritten vollziehen

lassen und jeweils in endlicher Zeit abgeschlossen werden können (27 [S. 4], 28 [S. 5]).

Parser

Parser sind eine Klasse von Algorithmen, die vor allem für die Syntax- bzw. Strukturanalyse natürlicher Sprache entwickelt und verwendet werden. Hierbei kann das „Parsing“ wieder als Prozess mit Ein- und Ausgabe verstanden werden. Die Eingabe, die in natürlich- (z.B. jede gesprochene Sprache) oder formalsprachlicher Form (z.B. eine Programmiersprache) vorliegt, wird anhand vordefinierter grammatikalischer Regeln analysiert, die sich meist von Chomskys Hierarchie der Grammatik ableiten.
Die für die syntaktische Analyse von natürlicher Sprache am häufigsten verwendeten Grammatiken sind die *Context-free Grammar* und die *Definite Clause Grammar*. Als Ausgabe wird eine syntaktische Strukturbeschreibung wiedergegeben, die der Parser in Form von Baumdiagrammen oder Attribut-Wert-Matrize ausgibt (25 [S. 46,47,64-66], 29 [S. 4 f.]).

Maschinelles Lernen

Maschinelles Lernen gehört zum Fachgebiet Künstliche Intelligenz und beschäftigt sich mit der Entwicklung von informationstechnologischen Modellen, die es Computern erlauben, zu lernen. Dabei wird Lernen als Fähigkeit definiert, durch Repetition die eigene Architektur anzupassen, und somit aufgrund vorheriger Ergebnisse und „Erfahrungen“ die eigene Leistung (Performance) zu verbessern (30,31).
Um jenen gewünschten Lerneffekt zu erreichen, wird der jeweilige Lernalgorithmus erst mit einem Trainingsdatensatz adaptiert und schließlich mit einem separaten Datensatz validiert (Validierungsdatensatz). Dabei ist wichtig, dass beide Datensätze keine überschneidenden Anteile besitzen, um die spätere Validierung nicht zu gefährden.

Zu den Lernalgorithmen, die beim maschinellen Lernen Verwendung finden, zählen neben dem Entscheidungsbaum (*Decision Trees*), der induktiven logischen Programmierung (*Inductive Logic Programming*), dem verstärkenden Lernen

(*Reinforcement Learning*) und vielen anderen, auch die neuronalen Netzwerke (*Neural Networks*) (31 [S. 488]).

Neuronale Netzwerke

Neuronale Netzwerke sind Lernalgorithmen, die sich in ihrer Architektur an dem Aufbau und an der Funktionsweise einfacher biologischer Neuronen orientieren und durch hohe Lernfähigkeit, Robustheit, Fehlertoleranz, Generalisierungsfähigkeit und Performance überzeugen.

Sie können als verknüpfte Ebenen (*Layers*) einfacher Prozessoren bzw. Rechenvorgänge (Neuronen) verstanden werden, wobei die Verknüpfung der Prozessoren (Synapsen) jeweils mit einer Gewichtung versehen wird. Über die Anpassung dieser Gewichtungen kann die gewünschte Datenverarbeitung sowie der begehrte Lerneffekt erreicht werden. Neuronale Netzwerke wandeln eine definierte Menge von Inputs mittels Vektorkodierung in eine definierte Menge von Outputs um, die ebenfalls im Vektorformat ausgegeben wird (31 [S. 488], 32 [S. 13], 33 [S. 3-5]).

Funktionsweise eines NLP-Systems

Nachdem die für das Verständnis dieser Diplomarbeit wichtigsten Grundbegriffe erläutert wurden, wird im Folgenden beschrieben, wie sich ein klassisches NLP-System zusammensetzt und welche Verarbeitungsstufen hier üblicherweise genutzt werden, wobei die Eingabe in natürlicher Sprache erfolgt.

Phonetik und Phonologik. Die meisten Eingaben eines NLP-Systems liegen bereits in geschriebener digitaler Form vor, doch prinzipiell kann eine Sprachprozessierung bereits mit der computergestützten phonologischen Analyse gesprochener Sprache beginnen. Diese wird erkannt und in geschriebene Sprache überführt (Spracherkennung und -generierung).

Syntaktische Analyse. Die syntaktische Analyse teilt sich in eine morphologische und eine lexikalische Analyse auf.

Bei der morphologischen Beurteilung (auch Tokenisierung genannt) werden mittels Parsingalgorithmen die verschiedenen Worte der Eingabe von Leer- und Satzzeichen unterschieden und anschließend hinsichtlich ihrer morphologischen

Grundbausteine (Morpheme) analysiert (auch *Morphological Parsing* genannt). Morpheme stellen die kleinsten grammatischen Bausteine eines Wortes, bzw. die kleinste sinnhafte Einheit einer Sprache dar. Das Wort „Tisch“ besteht beispielsweise aus einem Morphem, wohingegen sich „Tische“ aus zwei dieser Einheiten zusammensetzt, nämlich „Tisch“ und „e“. Hierbei vermittelt das Morphem „e“ die Mehrzahl. Man spricht dabei von freien und gebundenen Morphemen. Freie Morpheme können selbständig als Wort fungieren (z.B. „Tisch“), wohingegen Gebundene (z.B. „e“), wie Präfixe und Suffixe, nur als Teil anderer Worte („Tisch“-„e“) auftreten.

Das sinnhafte Verständnis eines Wortes ist bei diesem Schritt essenziell, da ansonsten ein Wort seine Bedeutung verliert (z. „Tische“ – „Tis“, „che“) und geht daher mit einer lexikalischen Analyse einher. Hierbei wird die Gültigkeit der verschiedenen Worte mittels einer Lexikon-Datenbank überprüft, die eine Sammlung aller möglichen und gültigen Wörter einer Sprache einschließlich ihrer linguistischen Informationen (Nomen, Verb, Adverb…) enthält. WordNet stellt hierbei die meistbenutzte Datenbank für die lexikalische Analyse in der englischen Sprache dar (34).

Grammatikalische Analyse. Es folgt die Analyse der formellen bzw. grammatikalischen Zusammenhänge zwischen den verschiedenen Wörtern und Sätzen. Auch hier kommen wieder Parsingalgorithmen zum Einsatz, die die Sätze der Eingabe durch im Voraus definierte Grammatikregeln analysieren. Die am häufigsten verwendeten Regelwerke für diesen Schritt sind, wie bereits oben beschrieben, die *Context free Grammar* und *Definite Clause Grammar*. Hinsichtlich der Parsingalgorithmen wird entweder *Top-Down-Parsing* oder *Bottom-Up-Parsing* verwendet. *Top-Down-Parsing* analysiert hierbei erst grobe Zusammenhänge im Satz, bevor jedes einzelne Wort hinsichtlich seiner grammatikalischen Bedeutung (Nomen, Verb, Artikel etc.) eingeordnet wird. Beim *Bottom-Up-Parsing* wird hingegen zuerst jedes einzelne Wort kategorisiert, bevor der gröbere Satzaufbau erschlossen wird.

Semantische Analyse. Es schließt sich die semantische Analyse an, die mittels eines weiteren Algorithmus die Bedeutung der natürlichsprachlichen Eingabe erschließt und sie mit einer für Maschinen verständlichen Repräsentation der

Bedeutung versieht. Solch eine Repräsentation kann in weiterer Folge logische Schlussfolgerungen zulassen.

Pragmatik und Diskursanalyse. Wichtiger Teil der semantischen Analyse stellt die Diskursanalyse dar. Die Pragmatik befasst sich mit dem Zusammenhang zwischen Sprache und dem Kontext von Inhalt und Sprecher. Hierbei wird, nachdem die Bedeutung der Eingabe bereits erfasst wurde, erschlossen, wie und wann sich die natursprachliche Eingabe auf Personen oder Objekte bezieht und wann im Verlauf eines Diskurses Bezug auf vorherige Aussagen genommen wird.

Die semantische Information wird anschließend als tiefere Satzstruktur ausgegeben. Die Konvertierungsregeln akzeptieren schließlich diese tiefere Satzstruktur und übersetzten sie für den Datenbankbearbeiter. Dieser führt abschließend den in der Eingabe enthaltenen Befehl aus und sucht beispielsweise die entsprechenden Daten in Form von Datenbankeinträgen. Diese Schritte beschreiben die Eingabeverarbeitung von Sätzen in natürlicher Sprache. Der umgekehrte Prozess wird bei der Erzeugung natürlicher Sprache verfolgt (25).

Nachdem die Funktionsweise eines allgemeinen NLP-Systems nun in ihren wichtigsten Zügen beschrieben wurde, soll im Folgenden kurz auf NLP-Systeme eingegangen werden, die im medizinischen Sektor Anwendung finden.

Clinical Natural Language Processing (cNLP)

Wie bereits beschrieben, generiert der medizinische Sektor enorme Mengen an größtenteils unstrukturiert vorliegenden und somit schlecht analysierbaren Daten. cNLP-Systeme versuchen, durch die Strukturierung von geschriebenen medizinischen Dokumenten (Freitexten), den enormen Informationsgehalt dieser hochspezifischen Daten zu erschließen und am Ende nutzbar zu machen (35).

Schließlich sollen somit bspw. die Qualität der Pflege verbessert, Kosten gespart, und im besten Fall Leben gerettet werden (23).

Beispiele für cNLP

Nach der Definition und Zielsetzung wird nun anhand einiger Beispiele das Potential sowie mögliche Anwendungsgebiete von cNLP aufgezeigt.

Sabra et al. haben das SESARF-NLP-System entwickelt (*Semantic Extraction and Sentimental Assessment of Risk Factors*), das unbekannte bzw. „versteckte“ Risikofaktoren einer bestimmten, zu wählenden Erkrankung automatisiert aus den zur Verfügung stehen medizinischen, geschriebenen Patientendaten ausliest und somit zur Diagnosestellung beiträgt (36).

Auch Doan et al. stellen einen NLP-basierten Ansatz vor, der mittels automatisierter Verarbeitung von Freitexten einer Notaufnahme das Risiko von Kindern hinsichtlich eines Kawasaki-Syndroms ausgibt (37).

Wi et al. entwickelten einen NLP-Algorithmus (NLP-PAC) zur Detektion von asthmaassoziierten Vorfällen, basierend auf textuellen Inhalten der Patientenakte. Dieser Algorithmus wurde später durch Sohn et al. aufgegriffen und erweitert, um zusätzlich die Prognose von an Asthma erkrankten Kindern zu evaluieren (38,39).

Auch Izquierdo et al. beschäftigten sich mit der NLP-unterstützen Ermittlung prognostischer Faktoren einer Lungenerkrankung. Sie ermittelten klinische Charakteristika und prädiktive Faktoren, die eine Wahrscheinlichkeitsvorhersage hinsichtlich der Intensivversorgung von COVID-positiven Patientinnen und Patienten erlauben. Alter, Fieber und Tachypnoe stellten sich hierbei als effizienteste Faktoren der Vorhersage heraus (40).

Lingeman et al. detektierten unter Verwendung NLP-gestützter Textanalyse aus Patientendaten auffälliges Verhalten, das mit Opioid-Missbrauch verbunden ist (41).

Zhong et al. verfolgten einen ähnlichen Ansatz. Sie ermittelten selbstmordassoziiertes Verhalten mittels NLP anhand dokumentierter Daten von schwangeren Frauen (42).

Als letztes Beispiel soll noch das Review von Warrer et al. vorgestellt werden. Hierbei wurden 7 Studien der USA verglichen, die unter anderem mittels NLP unerwünschte Arzneimittelwirkungen (UAWs) aus klinischen Entlassungsbriefen extrahieren. Es zeigte sich, dass die Performance der NLP-basierten Modelle jenen anderer Analyseverfahren überlegen ist (43).

cNLP ist also vielseitig einsetzbar und kann zur Beantwortung zahlreicher wissenschaftlicher Fragen beitragen. Es ist in der Lage, den medizinischen Alltag

zu bereichern, denn je spezifischer die Informationen für ärztliches und pflegerisches Personal sind, desto individueller fällt die Therapie aus.

Entscheidende Faktoren für eine effiziente Therapie sind beispielsweise frühere Krankheitsverläufe, die Vorgeschichte der gegenwärtigen Erkrankung, der Allergiestatus, die Medikamentenanamnese, aber auch Lifestyle- und Risikofaktoren.

Alkoholstatus als Teil der standardisierten Anamnese

Erhebung des Alkoholstatus

Da der Alkoholstatus einer Person auf verschiedene Weise erfasst werden kann, werden im Folgenden drei gängige Alkohol-Screening-Tests vorgestellt.

AUDIT, AUDIT-C

Der *Alcohol Use Disorder Identification Test* (AUDIT) ist ein Fragebogen mit 10 Fragen. Diese beziehen sich auf den jüngsten Alkoholkonsum, Anzeichen von Alkoholabhängigkeit und andere alkoholbedingte Probleme. Er wurde 1982 von der WHO entwickelt und 1989 erstmals veröffentlicht. Die Auswertung des Tests ist sehr einfach. Für jede der 10 Fragen gibt es festgelegte Antwortmöglichkeiten, die jeweils mit einem bis vier Punkten bewertet werden. Nachdem alle Fragen beantwortet worden sind, werden die erreichten Punkte addiert. Liegt die Gesamtpunktzahl bei 8 (bzw. 7 bei Personen über 65 Jahren) oder höher, muss von einem gefährlichen bzw. schädlichen Alkoholkonsum, sowie einer möglichen Alkoholabhängigkeit ausgegangen werden (44). Der AUDIT-C ist eine kürzere und damit praktischere Form des ursprünglichen Fragebogens von 1998. Er besteht aus nur drei Fragen und ist dennoch in seiner Aussagekraft mit dem AUDIT vergleichbar (45).

FAST

Der FAST ALCOHOL SCREENING TEST (FAST) von Hodgson et al., der 2001 veröffentlicht wurde, basiert auf dem AUDIT-Test, besteht aber lediglich aus vier Fragen und ist geschlechterspezifisch. Er wurde, ähnlich dem AUDIT-C, entwickelt, um ein schnelles Screening des Alkoholkonsums in Abteilungen mit hohem Zeitdruck, wie z.B. Notaufnahmen, zu ermöglichen. Wieder sind verschiedene Antwortmöglichkeiten gegeben, die jeweils mit einem bis vier Punkten bewertet

werden. Eine Endpunktzahl von drei oder höher deutet auf einen wahrscheinlich riskanten Alkoholkonsum hin (46).

CAGE

Der CAGE (Cutting down, Annoyance by criticism, Guilty feeling, and Eye-openers) Fragebogen, 1984 von John A. Ewing veröffentlicht, enthält ebenfalls einen Fragebogen mit vier Fragen. Hier wird jede positiv beantwortete Frage mit einem Punkt bewertet. Zwei bis drei Punkte weisen auf hohe Wahrscheinlichkeit einer Alkoholsucht hin. Vier Punkte sind gleichbedeutend mit der Diagnose von Alkoholismus (47,48).

Klassifikationen des Alkoholkonsums

Nachdem eine Person zu ihrem Alkoholkonsum befragt und ein möglicher Alkoholmissbrauch festgestellt wurde, folgt eine genauere Diagnose und eine endgültige Klassifizierung. Die gebräuchlichsten Klassifikationen für Alkoholkonsum sind:

- Die internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (ICD),
  und
- Das *Diagnostic and Statistical Manual of Mental Disorders* (DSM).

Das DSM wird hauptsächlich in den USA verwendet, die ICD außerhalb der Vereinigten Staaten (49). Beide Klassifizierungen werden im Folgenden kurz beschrieben, bevor auf SNOMED CT eingegangen wird.

DSM-4, DSM-5

DSM wurde erstmals 1952 vom *American Psychiatric Association Committee of Nomenclature and Statistics* veröffentlicht und liegt seit Mai 2013 in seiner fünften Fassung vor. Im Gegensatz zu den Vorgängerversionen (Alkoholmissbrauch, Alkoholabhängigkeit) wird Alkoholmissbrauch in dieser Version in drei Klassen eingeteilt: leichte, mittelschwere und schwere Alkoholmissbrauchsstörung (AUD) (50).

ICD-10

Die erste Version der ICD wurde 1948 von der WHO entwickelt. Seit 1992 befindet sich die ICD in ihrer zehnten Version und enthält etwa 155.000 Codes (51). Eine elfte Version existiert bereits, ihr Einsatz wird in vielen Ländern vorbereitet.

Die ICD-10 zählt ebenfalls drei Klassen von Alkoholmissbrauch auf, betrachtet aber dabei lediglich den Alkoholkonsum und schließt nicht seine Folgen ein: akute Intoxikation, schädlicher Gebrauch und Abhängigkeitssyndrom. Diese Klassen werden durch Unterklassen ergänzt, die eine genauere Klassifizierung der Patienten / Patientinnen je nach Diagnose ermöglichen, zum Beispiel beim Abhängigkeitssyndrom:

- Gegenwärtig abstinent
- Gegenwärtig abstinent, aber in einem geschützten Umfeld
- Gegenwärtig in einem klinisch überwachten Erhaltungs- oder Ersatzregime [kontrollierte Abhängigkeit]
- Gegenwärtig abstinent, aber in Behandlung mit aversiven oder blockierenden Medikamenten
- Gegenwärtiger Konsum der Substanz [aktive Abhängigkeit]
- Kontinuierlicher Konsum
- Episodischer Konsum [Dipsomanie] (52).

SNOMED CT

SNOMED CT, ursprünglich *Systematized Nomenclature of Medicine Clinical Terms*, stellt ein internationales, mehrsprachiges Terminologiesystem im Gesundheitssektor zur standardisierten Strukturierung von Patientendaten dar. Es ermöglicht die automatisierte Verarbeitung dieser Daten entsprechend der FAIR-Kriterien (Findable, Accessible, Interoperable, Reusable) (53,54).

Dieses seit 2002 stetig weiterentwickelte Terminologiesystem umfasst insgesamt mehr als 300.000 bedeutungstragende Einheiten (*Concepts*), die sich in einer komplexen Hierarchie aus etwa 1.360.000 semantischen Beziehungen befinden. Diese definieren u.a. Ober- und Unterbegriffe (*Parent* / *Children*), auslösende Faktoren (*causative agent*) oder betroffene anatomische Strukturen (*finding site*) (55,56).

SNOMED CT listet unter dem Konzept: *Finding relating to alcohol drinking behavior (finding)* (SCTID: 228273003) 24 Kindkonzepte auf. Davon weisen elf wiederum Unterteilungen auf. Diese hierarchische Struktur lässt eine feingranulierte Klassifizierung des Alkoholkonsums zu. Welche Klassen für die zu untersuchende Forschungsfrage dieser Diplomarbeit gewählt wurden, wird im Kapitel 2.2 beschrieben.

Geforderter Standard der Alkoholanamnese

Die genaue Erhebung der Alkoholanamnese ist für Arzt / Ärztin und Patient / Patientin potenziell unangenehm, und wird daher in vielen Fällen vernachlässigt. Aufgrund der Neigung, den eigenen Alkoholkonsum zu rationalisieren, zu bagatellisieren oder zu negieren, besteht das Risiko, einen schädlichen Alkoholkonsum bei der Anamneseerhebung zu übersehen.

Die Alkoholanamnese stellt jedoch aufgrund der hohen Prävalenz von Alkoholkonsum und den zahlreichen, alkoholassoziierten Erkrankungen (wie bereits erwähnt) einen wichtigen Teil der klinischen Anamnese dar und sollte deshalb konsequent durchgeführt werden (57–59).

Die S3-Leitlinien der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), der Deutschen Gesellschaft für Psychiatrie und Psychotherapie, der Psychosomatik und Nervenheilkunde (DGPPN) und der Deutschen Gesellschaft für Suchtforschung und Suchttherapie e.V. (DG-SUCHT) empfehlen in diesem Zusammenhang das Screening des Alkoholkonsums mittels:

- Fragebogen AUDIT (AUDIT-c in Settings mit hohem Zeitdruck),
- der Erhebung eines Mengen-Frequenz-Index (mittels getrennter Fragen zu Frequenz und Alkoholmenge des üblichen Konsums, sowie die Befragung hinsichtlich Frequenz und Menge vermehrten Alkoholkonsums)
- mittels tageweise rückblickender Anamnesen (Timeline-Followback) (60).

Gerade die frühe Phase problematischen Alkoholkonsums lässt sich meist nur anamnestisch diagnostizieren. Da hier die Heilungsprognose am besten ist und irreversible Organschädigungen noch fehlen, wird erneut klar, welche Bedeutung einer konsequenten Alkoholanamnese zukommt (57).

NLP kann in diesem Zusammenhang zwar nicht die Anamnese selbst erleichtern, aber durch die automatisierte Analyse klinischer Texte dazu beitragen, schädlichen Alkoholkonsum zu identifizieren, sofern er in einem Dokument erwähnt ist. Somit kann rechtzeitig eine Diagnose gestellt und eine zeitnahe Intervention eingeleitet werden.

Literatur hinsichtlich NLP und Alkoholstatus

Alzoubi et al. präsentieren ein System, das maschinelles Lernen und regelbasierte Methoden verwendet, um den Alkoholkonsum von Patientinnen und Patienten automatisch zu identifizieren. Dieses System extrahiert relevante Sätze durch Anwendung von NLP mittels eines Bag-Of-Words- und Keyword-Suchansatzes. Anschließend werden die Patientinnen und Patienten mithilfe von maschinellem Lernen hinsichtlich Identifizierung, Annotation von Negation und dem zeitlichen Status des Alkoholkonsums ihrer Patientendaten klassifiziert. Die zur Entwicklung des Systems verwendeten Daten stammen aus dem *Multiparameter Intelligent Monitoring in Intensive Care* (MIMIC) III (Version 1.3) (61), einer kostenlosen, frei zugänglichen und umfassenden klinischen Datensammlung. Für die Systementwicklung klassifizierten Alzoubi et al. 5000 Entlassungsbriefe manuell hinsichtlich des Alkoholkonsums. Das endgültige Modell zeigt eine hohe Leistung bei einem F_1-Score von bis zu 0,95 für die Klassifizierung des Alkoholstatus, und einem F_1-Score von 0,87 für die Klassifizierung der zeitlichen Komponente des Alkoholkonsums (62).

Topaz et al. nutzten ebenfalls MIMIC, um ein ähnliches System zur Klassifizierung des Alkoholstatus zu entwickeln. Sie setzen das Open-Source-NLP-System "NimbleMiner" (63) zur automatischen Identifizierung von Dokumenten ein, in denen Alkohol und Drogenmissbrauch erwähnt werden, gefolgt von der Textklassifizierung mittels maschinellen Lernens (Random-Forest-Klassifikator). Das System erzielt einen F_1-Score von 0,84 sowohl für Drogen- als auch für Alkoholmissbrauch. Es zeigt sich zudem, dass ein Lernalgorithmus, der mit ärztlich verfassten Entlassungsberichten UND Pflegeberichten trainiert wurde, 79 % mehr Bezeichnungen für Drogenmissbrauch und 97 % mehr für Alkoholmissbrauch erkennt (64).

Lix et al. (2017) entwickeln ein System zur automatischen Extraktion und Klassifizierung des Alkoholkonsums von Patientinnen und Patienten, basierend auf unstrukturierten oder frei formulierten medizinischen Daten. Sie wurden entnommen aus elektronischen Krankenakten des regionalen Netzwerks Manitoba des *Canadian Primary Care Sentinel Surveillance Network* (CPCSSN). 2.000 Dokumente wurden für die Studie manuell annotiert und anschließend zum Trainieren des Lernalgorithmus verwendet. In dieser Studie wurde ein SVM-Klassifikator mit Unigrammextraktion mit einem n-Gramm-Extraktionsmodell verglichen, das einen Bag-of-Words-Ansatz verwendet. Der SVM-Klassifikator mit Unigrammextraktion schnitt besser ab und erreichte einen F_1-Score von 0,89 für die Klassifizierung von aktuellen Trinkern und 0,98 für die Klassifizierung von unbekanntem Alkoholkonsum (65).

Afshar et al. stellen ein System vor, das NLP-Klassifikatormethoden zur Identifizierung von Patientinnen und Patienten mit und ohne Alkoholmissbrauch verwendet. Die verwendeten Daten stammen aus der Notaufnahme eines Traumazentrums der Stufe I. Inkludiert wurden nur Daten, die in den ersten 24 Stunden nach der Patientenaufnahme generiert wurden. Die linguistische Verarbeitung der klinischen Daten erfolgt mit Hilfe des klinischen Textanalyse- und Wissensextraktionssystems (*Clinical Text Analysis and Knowledge Extraction System* - cTAKES). In der Validierungskohorte mit 285 Personen ergab ihr NLP-Klassifikator eine AUC-ROC-Kurve von 0,78 (95% CI, 0,72 bis 0,85) (66).

To et al. validieren den zuvor erwähnten Alkohol-NLP-Klassifikator von Afshar et al. anhand von Daten erwachsener Patientinnen und Patienten aus einem Krankenhaus der Maximalversorgung. Ziel der Studie ist es, herauszufinden, inwieweit der Klassifikator auf die gesamte Patientenpopulation eines Krankenhauses angewendet werden kann. Ihr neuer Klassifikator für Alkoholmissbrauch weist eine AUC-ROC-Kurve von 0,91 (95% CI, 0,90 bis 0,93) auf (67).

Die Studie von Phillips et al. bewertet die Anwendbarkeit eines Systems, das NLP und maschinelles Lernen kombiniert, um zwischen Personen mit und ohne Alkoholmissbrauch zu unterscheiden. Für die Vorverarbeitung der Daten

verwenden sie ebenfalls cTAKES zusammen mit einem Modell, das auf *word2vec*[1] basiert. Ein logistischer Regressionsklassifikator zeigt schließlich die beste Performance und weist einen F_1-Score von 54,1 % und einer AUC ROC von 80,4 % auf (68).

Basierend auf der vorliegenden Literaturrecherche lässt sich der mögliche Erkenntnisgewinn dieser Diplomarbeit einschätzen. Zunächst stellt diese Abhandlung nach bestem Wissen und Gewissen des Autors die erste wissenschaftliche Arbeit dar, die eine Alkoholstatusextraktion und -standardisierung mit Hilfe maschineller Lernverfahren auf Basis *deutschsprachiger* Arztbriefe untersucht. Außerdem unterscheidet sich das verwendete NLP-Verfahren dieser Diplomarbeit grundsätzlich von den bereits publizierten Methoden. So wird im Rahmen dieser Arbeit der freitextliche Input nicht vor-tokenisiert, ein Teilschritt, der in anderen Publikationen beispielweise mittels cTAKES vollzogen werden musste. Die verwendete NLP-Methodik dieser Diplomarbeit unterscheidet sich somit vor allem durch ihren geringeren technischen Aufwand von den angeführten Publikationen mit ähnlicher wissenschaftlicher Fragestellung.

Zielsetzung der Diplomarbeit

Im Jahr 2016 führte schädlicher Alkoholkonsum weltweit zu etwa 3 Millionen Toten. Das entspricht 5,3% aller und 7,3% der vorzeitigen (bis zu einem Alter von 69 Jahren) Todesfälle weltweit. Die Mortalität von Alkohol übersteigt somit die von Tuberkulose, AIDS/HIV und Diabetes (69). Dabei wird Alkoholkonsum mit mindestens 18 chronischen Gesundheitseinschränkungen in Zusammenhang gebracht (70). Es zeigt sich: Ausmaß und Art des Alkoholkonsums sind wichtige Lifestyle-Faktoren, die bei der Patientenanamnese erhoben werden sollten.

Fast ausschließlich wird der Alkoholstatus nur in freitextlicher Form in der klinischen Routinedokumentation erfasst und ist daher an kein standardisiertes Format gebunden. Freie sprachliche Formulierungen erschweren, wie bereits beschrieben, eine strukturierte Erhebung dieser Lifestyleausprägung für retrospektive und vergleichende Auswertungen. Eine strukturierte und standardisierte Klassifikation der freitextlichen Ausprägungen ist daher wünschenswert.

[1] https://code.google.com/p/word2vec/

Im Rahmen dieser Diplomarbeit wird beurteilt, ob man mit Methoden des maschinellen Lernens eine standardisierte Kategorisierung des Alkoholstatus erzielen kann. Ein Fokus liegt dabei auf der Ausarbeitung des Klassifikationsschemas, basierend auf der zugrundeliegenden Datenlage der klinischen Routinedokumentation dieser Lifestyleausprägung und einer sinnvollen Zuordnung zu bestehenden SNOMED-CT-Codes.

Material und Methoden

Beschreibung der Datenbasis

Für die Beantwortung der wissenschaftlichen Fragestellung dieser Diplomarbeit werden Daten aufbereitet und schließlich analysiert, die ursprünglich aus den elektronischen Krankenakten der steiermärkische Krankenanstaltengesellschaft (KAGes) stammen. Die Daten umfassen anonymisierte Arztbriefe, Beschreibungen der Krankengeschichte und Entlassungsbriefe von sowohl ambulanten wie auch stationären Patienten-Aufenthalten der Fachgebiete Kardiologie, Dermatologie und Onkologie. Insgesamt enthält die initiale Datenbasis etwa 32.000 kardiologische, 1.700 dermatologische, und 13.000 onkologische Entlassungsbriefe, wobei sich der onkologische Anteil vor allem auf Personen mit Kolonkarzinom konzentriert.

Um bei der automatisierten Klassifikation des Alkoholstatus eine möglichst hohe Performance zu erreichen, müssen zunächst die relevantesten Teile der Dokumente ermittelt werden. Hierfür werden mittels einem regelbasiertem Ansatz[2] relevante Textzeilen extrahiert. Durchgeführt wird dieser Teilschritt von einer erfahrenen Datenmanagerin, die in der Extraktion von Informationen aus klinischen Informationssystemen spezialisiert ist. Ist ein Suchstring in einem Entlassungsbrief gefunden, wird ein Teilausschnitt/Snippet erstellt. Jedes Snippet selbst setzt sich aus dem Suchstring und 100 Zeichen vor und nach dem Suchstring zusammen, um die kontextuelle Einbettung des gesuchten Begriffs zu erfassen. Insgesamt werden 1429 Snippets aus der verwendeten Datenbasis erstellt, die jedoch noch weiter pro-

[2] [aA]l[ck]ohol|[cC]2(H5(O|0)H){0,1}[\-\s]{0,3}(a|A)|[äÄeE]thyl|F10|K70|G31\.2|T51|P04\.3|Q86\.0|K29\.2|K86\.0|I42\.6|Z71

zessiert werden müssen, bevor eine Analyse durch das NLP-System möglich ist.

Als nächster Schritt wird das Einteilungsschema der Snippets erstellt. Eine überschneidungsarme Klassenerstellung ist essenziell für eine leistungsstarke automatisierte Klassifikation. Sie wird deshalb mit größter Sorgfalt vollzogen und im folgenden Kapitel genauer beschrieben.

Verwendetes Klassifikationsschema

Die Forschungsfrage dieser Diplomarbeit bezieht sich auf die automatisierte Klassifizierung des Alkoholstatus von Patientinnen und Patienten basierend auf Ausschnitten unstrukturierter Textdaten. Bei der Definition dieser disjunkten Klassen wurden zunächst die Studien der Literaturreche herangezogen und hinsichtlich der alkoholkonsumassoziierten Klasseneinteilung analysiert. Es zeigten sich folgende Varianten:

Phillips (68):

- alcohol misuse - yes
- alcohol misuse - no
- unknown alcohol status

Alzoubi et al. (62):

- current alcohol consumption
- past alcohol consumption
- non-drinker
- unknown.

Lix et al. (65):

- current drinker,
- not a current drinker,
- unknown

Basierend auf diesen Einteilungen, und unter Verwendung des Terminologiesystems SNOMED CT wurden folgende Klassen (als *Fully Specified Names* + SNOMED-IDs) identifiziert:

- Current drinker of alcohol (finding) (SCTID: 219006)
- Current non-drinker of alcohol (finding) (SCTID: 105542008)
- Alcoholic/ Problem drinker (finding) (SCTID 228281002)
- Ex-problem drinker (finding) (SCTID: 286857004)
- Disorder caused by alcohol (disorder) (SCTID 719848005)
- Alcohol consumption unknown (finding) (SCTID: 160580001)

Im Folgenden werden die Klassen definiert und jeweils einige Beispiel-Snippets aufgeführt.

Current drinker of alcohol (SCT-ID: 219006)

Diese Klasse bezeichnet das Verhalten jeder Person, die angibt, Alkohol zu konsumieren, aber keinen Hinweis auf schädlichen Alkoholgebrauch bzw. einen Alkoholabusus aufweist. Hinweise auf schädlichen Alkoholkonsum und Alkoholabusus werden in weiterer Folge definiert.

- (Ileumblasenanlagen ca. 2000, keine bekannten Allergien. Vegetative Anamnese: Appetit unauffällig, ***Alkohol: abends 1/8 L Rotwein***, Nikotin: Gelegenheitsraucher, Stuhl: normalerweise regelmäßig (siehe)
- (tand nach Appendektomie und Varizen-OP bds., keine Dauermedikation. Nikotinanamnese: ca. 10 pro Tag ***Alkoholanamnese: gelegentlich*** Allergien: keine bekannt. Aufnahmesetatus: Unauffälliger AZ, Psy: wac)
- (sher wurde noch nie eine Gastroskopie bzw. Colonoskopie bei dem Pat. durchgeführt. Rauchen: negativ ***Alkohol: selten*** Allergien: keine bekannt. Status AZ normal, EZ adipös, Caput und Collum: Zunge leic)

Current non-drinker of alcohol (SCT-ID: 105542008)

Die Klasse umfasst alle Snippets, die das Verhalten von Personen bezeichnen, die aktuell keinen Alkohol trinken, also Alkoholkonsum verneinen und in ihrer Vergangenheit keine überstandene Alkoholsucht aufweisen. Patientinnen und Patienten, bei denen „nur“ kein Alkoholabusus festgestellt werden konnte, fallen NICHT in diese Gruppe, da ein fehlender Alkoholabusus nicht mit einer Abstinenz einhergehen muss. Sie werden in die Klasse *Alcohol consumption unknown* eingeordnet.

- (im Rahmen der Peritonealkarzinose, stationär. Harnverhalten sei unauffällig, der Appetit reduziert. ***Alkohol- und Nikotinabusus werden verneint***. Untersuchung/Befunde PHYSIKALISCHER STATUS: Es präsenti)
- (ert und heute bei der Aufnahme wieder selbst entfernt. Allergien: keine bekannt. Nikotin: negativ. ***Alkohol: negativ***. Aufnahmestatus **********)
- (e li. US. Nachtragsschmerzen US re. bei St.p. Poliomyelitis. Rauchen: Nichtraucher seit 40 Jahren; ***Alkohol: seit 10 Jahren nicht mehr (davor 2-3 Bier/Tag)***; Schlafmittel: Trittico u. Adjuvin bei Schla)
- (ällig. Mittlerweile befindet sich die Patientin in palliativmedizinischer Betreuung. Nikotin: nihl **Alkohol: nihil** Allergien: ACE-Hemmerunverträglichkeit, Aufnahmestatus: AZ: reduziert, EZ: normal Ps)

Problem drinker (SCT-ID: 228281002)

Die Klasse bezeichnet das Verhalten von Personen, die einen diagnostizierten Alkoholabusus aufweisen. Snippets mit folgenden Begriffen und ICD-Codes werden dieser Klasse zugeordnet: C2H5OH-Abusus, C2-Abusus, Alkoholkrankheit, Alhohol-/C2-Abhängigkeit, Äthylismus, ICD Code F10.x (meist F10.1 oder F10.2). Bei Konflikten zwischen ICD-10 Codes und Beschreibung, zählt das geschriebene Wort. Patientinnen und Patienten, die überzeugende Hinweise auf eine Alkoholsucht bieten, aber keine klare Diagnose aufweisen, werden NICHT in diese Klasse integriert, sondern finden sich in der Klasse „*Disorder caused by alcohol*“.

- (ch am ******2017 ********** 8 A-**** ***** Ambulanter Arztbrief ****, ******2017/WT ***Diagnosen: Alkoholkrankheit*** Z.n. Colonkarzinom mit Operation und Radiochemotherapie Sehr geehrte Frau Kollegin)
- (her Seite dzt. keine Beschwerden an. Einziges Problem sei dzt., dass er kein Bier bekomme (***der Pat. Alkoholiker***). Chron. Husten wird verneint. Körpergewicht stabil bzw. zugenommen. Kein Nachtschweiß.)
- (ehr schlechte Schallbedingungen. LPS, geringe Splenomegalie. PSYCHIATRISCHES KONSILIUM: ***DIAGNOSEN: Alkoholkrankheit*** (mit Folgeschäden), Benzodiazepindependenz. THERAPIEVORSCHLAG: Empfehle Fortführung)
- (ierenschmerzen. Die Thoraxschmerzen beim Husten. Relevante Vorerkrankungen: Bronchitis, Pleuritis, ***C2-Abusus***, Hepatopathie, Hyperlipidaemie, degenerative Wirbelsäulenveränderungen, Amaurosis posttrau)
- ((seit Jahren). Med: TASS (wurde angiolog.seits 2009 empfohlen, OAK lt. AB bei deutl. red. AZ sowie ***C2-Abusus*** zu risikoreich). ****** *** **** ******** Seite 1/2 AT-**** **********, **********)

- (19 Arztbrief Stat.Aufenthalt: ******2007 bis ******2007 Diagnosen Depression F32.9 ***Alkoholabusus F10.1*** Anamnese Patient kommt am 4.7 zur Aufnahme, da er aufgrund des depressiven ZB vermehrt Alkoho)
- (filtration Gallenblase, C78.7 Diabetes mellitus Typ II, E11.9 Adipositas, E66.0 ***C2-Abhängigkeit, F10.1*** Arterielle Hypertonie, I10 AkutesNierenversagenAKI I,N17.9 ****, am ******2017 / Ad Sehr g)
- (ch mit Pat. absolvierbar Anamnese: Pat, kam am 3.7. mit der Rettung über EBA; bek. Epileptiker und ***Alkoholiker***, Verdacht auf Krampfanfall, laut Station wäre Verlegung geplant, der Pat. jedoch sei zun)

Ex-problem drinker (SCT-ID: 286857004)

Diese Klasse befasst sich mit allen Snippets, in denen von einem „Zustand nach (Z.n.)“ bzw. „Status post (St.p.)“ – Alkoholabusus (und Ähnlichem) oder einer Karenz bei vorliegenden alkoholassoziierten Folgen sprechen. Personen mit Entzugssymptomatiken (z.B. Delirium tremens), die also aktuell keinen Alkohol mehr konsumieren, aber keinen St.p. / Z.n. Alkoholabusus oder Ähnliches aufweisen, werden NICHT in dieser Klasse festgehalten, sondern finden sich in der Klasse „*Disorder caused by alcohol*“ wieder. Zu *Ex-problem drinker* zählen auch Personen, die in ihrer Vergangenheit Hinweise auf schädlichen Alkoholgebrauch aufweisen.

- (e seit 2009 in 1. Linie sek. von Willebrand.Syndrom im Rahmen der ETH art.HT Eisenmangelanämie ***St.p C2 Abusus*** axonale sensible sensomotorische Polyneuropathie St.p Insult 2012 Port-a-cath-Implantation)
- (Der Patient kommt zur geplanten Coloskopie mit Stenteinlage bei bekanntem Sigma-CA. Nikotin: neg. ***Alkohol: St.p. Abusus***, dzt. neg. Allergien: neg. Aufnhamestatus: AZ und EZ unauffällig, Psyche: wac)
- (lonkarzinom (rechte Flexur), (***12), C18.3 Nebendiagnose(n), ICD-10 ***Alkoholismus chronisch,St.p.*** , F10.2 alkoholtoxische Polyneuropathie, G62.1 Demenz, F02.8 St.p. Subarachnoidalblutung traumatisch,)
- (le Hypertonie E78.4 Hyperlipidämie E01.8 Subklinische Hypothyreose M51.2 Diskusprolaps L5/S1 ***St.p. C2 Abusus*** HWS-Syndrom ---------- RELEVANTE BEFUNDE ---------- * EKG 11.05.2010 SR,)
- ((Nicht-ST-Hebungsinfarkt) 05/2013 St.p. PLA-DES 14.5.12 schwere Jod-Kontrastmittelallergie ***Z.n. Alkoholabusus*** Glaukom re. Auge ---------- RELEVANTE BEFUNDE ---------- * EKG SR, 5)

- (0 St.p. Rectumresektion bei St.p. N. recti 2011 St.p. pAVK IV - Oberschenkelamputation rechts ***St.p. C2-Abusus*** Sehr geehrte Frau Kollegin! Sehr geehrter Herr Kollege! Anamnese Wir berichten über Ihre)
- (htig mit *** Kilogramm, wobei er in den letzten 4 Wochen bereits bewusst 12 Kilo abgenommen hat. An Alkohol wurden ***5 bis 6 Bier tgl***. getrunken, ***seit 1 Monat*** zwecks Gewichtsreduktion ***kein Alkoholkonsum***)
- (linken Leberlappen. Anamnese Chron. tachykarde VHFa (ED 1995) V.a. äthylische CMP Art. Hypertonie Äthylische Steatohepatitis, Cirr. hep. Cholezystolithiasis Mikrozytäre Anämie ***Z.n. chron. Alkoholabu***)
- (Hg * BEURTEILUNG Bei Herrn <-LN_10538-> besteht eine dilat. Kardiomyopathie, höchstwahrscheinlich äthylischer Genese. **Zur Zeit der Pat. (seit August) in Karenz**. Wir empfehlen vorerst die Fortführun)
- (III. * Klinisch-physikalische Symptomatik RR: 124/89mmHg; Gewicht: 90kg; * BEURTEILUNG ***Dil CMP äthyl.-tox.-Genese; Seit Dez. 2012 abstinent***. Card. Dekompensation mit massiven Beinödemen, Aszites)

Disorder caused by alcohol (SCT-ID: 719848005)

Diese Klasse ist in der Einteilung der Snippets besonders wichtig, denn hier werden einige Grauzonen zwischen den anderen Klassen abgedeckt.
In diese Klasse wird eingeteilt:

- wer mehr als 3 Bier oder 0,75 Liter Wein (3 Gläser) pro Tag trinkt
- wer alkoholspezifische Organveränderungen (z.B. ICD-10: K70.x, K86.0 oder I42.6), aber keinen diagnostizierten Alkoholabusus (oder passenden ICD Code) aufweist
- Personen mit Entzugssymptomatiken OHNE Diagnose einer Alkoholkrankheit (oder Ähnliches; siehe *Problem drinker*)
- akute Alkoholintoxikationen (F10.0), die klinisch behandelt werden müssen

Die Grenzwerte des schädlichen Alkoholkonsums sind dem *Handbuch Alkohol Österreich (Band 3: Kapitel 4.2)* des Bundesministeriums für Soziales, Gesundheit, Pflege und Konsumentenschutz entnommen (71).

- (des Patienten! --------- DIAGNOSEN --------- G40.8 Krampfanfall bei bek. posttraumat. Epilepsie F10.7 ***Chronisches alkoholisches organisches Psychosyndrom*** F05.8 Delirantes Syndrom multifaktorieller)

- (0 1-2 Temesta 1 - 2,5 mg bei Unruhe Procedere: Strikte **********! Bezüglich einer stationären ***Alkohol-Entwöhnungstherapie*** sollte die ambulante Vorstellung am ******* für Suchtmedizin an der Land)
- (z.AZ, adipösem EZ. Der Pat. ist orientiert, im Ductus verlangsamt, kooperativ. Nikotin: 35 ***/Tag. ***Alkohol: 5 Bier/Tag*** bei reduziertem AZ. und erhöhten Durstgefühl. Caput/Collum: die Pupillen rund, m)
- (kommt selbstständig ohne Terminvereinbarung. Anamnese Pat. kommt um etwas gegen sein exazerbiertes Alkoholproblem zu unternehmen, er trinke aktuell bis zu ***24 kleinen Bier pro Tag*** und dies bereits sei)
- (bei Unruhe Procedere: Dem Patient wird eine strikte Alkoholkarenz nahegelegt sowie eine stationäre ***Alkohol-Entwöhungstherapie*** empfohlen. Ein ausführlicher Arztbrief folgt. Mit kollegialen Grüßen D)
- (: Appetit: normal, Nikotin: seit drei Wochen Nichtraucher, Miktion: dreimal Nykturie, kein Brennen, ***Alkohol: 4 - 5 Bier pro Tag***, Stuhl: regelmäßig, eher weich, ******, Stimmung unauffällig. Keine rele)
- (008 und LAD DES Art.Hypertonus HLP positiver Familienanamnese Adipositas Nikotinabusus ***Verdacht auf C2 Abusus*** Angststörung, Depression * Physikalischer Status reduz. AZ, adipöser EZ, 168cm, 84kg. RR)

Alcohol consumption unknown (SCT-ID: 160580001)

Diese Klasse beinhaltet alle Snippets, die keine klare Klassifikation des Alkoholkonsums zulassen. Besonders häufig sind in dieser Klasse Anweisung zur Alkoholkarenz, aber auch Behandlungen mit Alkohol, wie beispielsweise die Alkoholablation zu finden.

- (all sprechen für eine Arthritis urica. Gichtattacken können sowohl durch exzessive Nahrungsaufnahme/***Alkoholgenuss*** oder hungern auftreten und verursachen CRP-Erhöhungen auch über 300mg. Empfehlung: - F)
- (er Obstipation, ansonsten unauffällig. Harn unauffällig. Keine Allergien bekannt. ***Kein Nikotin- und Alkoholabusus***. ****** verminderter AZ, normaler EZ. Haut und Schleimhaut unauffällig. Atmung: gering)
- () Stolpersturz mit leichter Benommenheit, aber ohne Bewusstseinsverlust 2) Kollaps bei erhöhtem ***Blutalkoholspiegel*** nach Feier, keine Bewusstseinsverlust 3) Soweit heute durch den Pat. erinnerlich ein)
- (**** in **********, seine Frau sei gestorben. Seine Lebensgefährtin lebe in **********, ***diese wäre Alkoholikerin***. Er fahre immer wieder zwischen ********** und seinem Bruder hin und her, zeitweise a)

- (der Patientin erfolgte bei bekannter hypertroph- obstruktiver Kardiomyopathie (HOCM) zur geplanten ***Alkoholablation*** (PTSMA). In der am 19.05.2011 durchgeführten Koronarangiographie wurde eine Alkohola)
- (4 Wochen, danach 1x tgl. für weitere 2 Monate Hirudoid-Gel re. Unterarm lokal bis 2x tgl. kalte und Alkoholumschläge re. Unterarm 2x tgl. Mit freundlichen Grüßen Dr. ******* ****** OA Dr. ***** ****)
- (.9 postinfektiöse Begleitmyocarditis E66.8 Adipositas permagna I10 Arterielle Hypertonie K76.0 ***Nichtalkoholische*** Fettleber ---------- RELEVANTE BEFUNDE ---------- * EKG SR, 73/min, IT)
- (Allgemeinzustand. Therapie Stationäre Aufnahme an der Chirurg. Abteilung des *** Graz-West, ***Stat. C2*** am *******, den *****12 um 12.00 Uhr. Bzgl. der Hautläsion an der Unterlippe re. wird bei fehlen)
- (eine Suizidversuche in der Vorgeschichte. Der Patient ***verneint Missbrauch und/oder Abhängigkeit von Alkohol*** und/oder psychotrop wirksamen Substanzen. Status psychicus Zum Zeitpunkt der fachärztlichen)

Goldstandard-Erstellung und Interrater Agreement

Nachdem das Einteilungsschema für die Annotation der Snippets erstellt wurde, folgt die Klassifikation. Hierbei wird durch den Autor manuell jedem Snippet eine der erstellten Klassen zugeordnet. Dabei kommt es zu folgender Verteilung:

Klasse	SCT-ID	Klassenbezeichnung	Anzahl
0	219006	Current drinker of alcohol	197
1	105542008	Current non-drinker of alcohol	259
2	286857004	Ex-problem drinker	100
3	719848005	Disorder caused by alcohol	249
4	228281002	Problem drinker	374
5	160580001	Alcohol consumption unknown	250

Tabelle 1. Klassenverteilung der Textsnippets.

Nach der initialen manuellen Annotation werden 20% der Snippets unabhängig durch eine zweite medizinische Expertin, basierend auf dem gleichen Klassifikationsschema, annotiert. Anschließend wird das *Interrater Agreement* berechnet (72). Ein *Cohen's Kappa* von 0,9 zeigt eine hohe Übereinstimmung der beiden Annotationen (73) und damit indirekt eine suffiziente Wahl der verschiedenen Einteilungsklassen.

Textvorverarbeitung

Bei der anschließenden Vorprozessierung der Snippets werden Groß- in Kleinbuchstaben umgewandelt, Leerzeichen entfernt und deutsche Umlaute in die entsprechenden Umschreibungen (z.B. „ü“ → "ue“) überführt.

Maschinelles Lernverfahren

Aufbau des cNLP Systems

Das NLP-System dieser Diplomarbeit nutzt im Gegensatz zu den beschriebenen NLP-Systemen die Möglichkeiten von maschinellem Lernen, insbesondere durch neuronale Netzwerke, um die beschriebenen Schwierigkeiten bei klassischen NLP-Ansätzen (siehe 1.2) zu überwinden.

Im Gegensatz zu einer vollständigen NLP-Pipeline wird in dieser Diplomarbeit *eine* Komponente entwickelt, die in eine Verarbeitungspipeline integriert werden kann. Um die Qualität des Standardisierungsprozesses hinsichtlich der gewählten Terminologie SNOMED CT mit Hilfe von Methoden des maschinellen Lernens zu gewährleisten, ist die quantitative Betrachtung der Klassifizierungsqualität unabdingbar. Die gewählten Methoden aus dem Bereich des maschinellen Lernens sind in Abschnitt 2.6 beschrieben. Stellt sich die Komponente hinsichtlich Qualität der standardisierten Zuordnung des Alkoholstatus als brauchbar heraus, soll diese als UIMA-Komponente (74) implementiert werden. Die Kapselung in dieses Framework ist nicht Teil dieser Arbeit.

Folgende Systeme kommen in dieser Abhandlung zum Einsatz:

Support-Vektor-Maschinen

Support-Vektor-Maschinen (SVMs) eignen sich in Kombination mit der hochdimensionalen Darstellung von Texten sehr gut zur Textklassifikation (75). Die Trainingsdaten werden dabei nichtlinear in einem höherdimensionalen Merkmalsraum abgebildet. Dort wird eine trennende Hyperebene mit maximalem Abstand zu den sogenannten „*Support Vektoren*“ errechnet. Dies führt zu einer nichtlinearen Entscheidungsgrenze im Eingaberaum. Durch die Verwendung einer Kernel-Funktion ist es möglich, die trennende Hyperebene zu berechnen, ohne die

Abbildung in den Merkmalsraum explizit vorzunehmen. Dies wird auch als *Kernel Trick* bezeichnet (76). Verschiedene Arten von Kernelfunktionen können dabei angewandt werden. Häufig wird z.B. ein polynomiales Kernel verwendet. Die genaue Parametrierung wird im Kontext der Ergebnisse beschrieben.

fastText

fastText (77,78), entwickelt vom *Facebook AI Research Lab*, zählt zu den Systemen, die mit nicht-kontextualisierten Word-Embeddings arbeiten. Hier bekommen, im Gegensatz zu kontextualisierten Word-Embeddings, homonyme Begriffe eine gleiche Vektorrepräsentation (79). fastText unterstützt dabei die Generierung von Word-Embeddings basierend auf einer Substring-Ebene, wodurch das *Out Of Vocabulary* (OOV) Problem minimiert wird. In Erweiterung der Substring-Betrachtung können aber auch n-Gramm-Ausprägungen für das Trainieren des Modells parametriert werden. Eine vollständige Auflistung der Parametrierungsoptionen ist auf dem öffentlich zugänglichen Git-Repository zu finden[3]. In dieser Arbeit werden Textausschnitte von 200 Zeichen über einen Embedding-Vektor der Dimensionalität 300 repräsentiert. Die genaue Parametrierung wird im Kontext der Ergebnisse beschrieben.

Evaluierungskennzahlen

Die Performance der Klassifikation wird anhand der folgenden drei statistischen Messwerte bewertet: Precision, Recall und F_1-Score. Die Definitionen dieser Messwerte werden im Folgenden veranschaulicht.

Precision

Die Präzision (*Precision*) beschreibt den Anteil der Datenobjekte, die durch den Algorithmus, im Vergleich zur manuellen Annotation, korrekt einer Klasse zugeordnet worden sind (80). Somit beschreibt eine Präzision von 0,9, dass neun von zehn der automatisiert zugeteilten Datenobjekte *einer* Klasse der manuellen Klassifikation entsprechen. Die entsprechende Konfusionsmatrix einer bestimmten Klasse könnte bei 25 Datenobjekten wie folgt aussehen:

[3] https://github.com/facebookresearch/fastText

Konfusionsmatrix		Manuelle Klassifikation	
		POSITIV	NEGATIV
Automatische Klassifikation	POSITIV	9	1
	NEGATIV	11	4

Tabelle 2. Beispiel einer Klasseneinteilung bei einer Präzision von 0,9.

Recall

Die Ausbeute (*Recall*) gibt (pro Klasse) den Anteil der Übereinstimmung zwischen NLP-basierter und manuell annotierte Klasseneinteilung an (80). Ein Recall von 0,9 beschreibt also, dass (unter Betrachtung *aller* manuell klassifizierten Datenobjekte einer Klasse) neun von zehn Datenobjekten vom Algorithmus richtig klassifiziert wurden. Eine passende, klassenbezogene Konfusionsmatrix mit 25 Datenobjekten könnte sich in diesem Fall folgendermaßen präsentieren:

Konfusionsmatrix		Manuelle Klassifikation	
		POSITIV	NEGATIV
Automatische Klassifikation	POSITIV	9	11
	NEGATIV	1	4

Tabelle 3. Beispiel einer Klasseneinteilung bei einem Recall von 0,9.

F_1-Score

Als gewichtetes harmonisches Mittel von Präzision und Recall kombiniert der F_1-Score (81) beide Parameter und eignet sich somit gut zum Vergleich unterschiedlicher Klassifikationsmodelle. Der Maximalwert von 1 entspricht einer perfekten Klassifikation (82 [S. 136]). Bei dieser Arbeit wird bei mehreren Klassen der ungewichtet gemittelte Makro-F_1-Score angeben.

Ergebnisse

Nachdem die wichtigsten statistischen Messwerte nun erklärt wurden, werden im Folgenden die Resultate der automatisierten Klassifikation vorgestellt.

Support-Vector-Maschinen

Trainings- und Testdatensatz wurden im Verhältnis 90:10 aufgeteilt. Die beste Parametrierung für das Modell wurde über eine 10-fache Kreuzvalidierung in Kombination mit der Variierung der Hyperparameter 'tfidf__use_idf': (True, False) und 'clf__alpha': (1e-2, 1e-3, 1e-4, 1e-5), auf dem Trainingsdatensatz eruiert. 'tfidf__use_idf' ist dabei das Gewichtungsschema der Vektorrepräsentation, ‘clf__alpha' beeinflusst die Lage der trennenden Hyperebene und der korrespondierenden Supportvektoren.

Zu erwähnen ist, dass bei der Vorverarbeitung des Textes eine Character-n-Gramm-Zerlegung in Kombination mit einer Token-n-Gramm-Zerlegung im Bereich $2 \leq n < 6$ durchgeführt wurde.

Class	Precision	Recall	F_1-Score	Support
Current drinker of alcohol	1,00	0,68	0,81	19
Current non-drinker of alcohol	0,84	1,00	0,91	31
Ex-problem drinker	0,83	0,56	0,67	9
Disorder caused by alcohol	0,81	0,88	0,85	25
Problem drinker	0,84	0,95	0,89	38
Alcohol consumption unknown	0,94	0,76	0,84	21
macro avg	0,88	0,80	0,83	143

Tabelle 4. Evaluierungsergebnisse des Testdatensatzes für die beste Parameterkombination 'tfidf__use_idf': (True), 'clf__alpha': 1e-4

fastText

Trainings, Validierungs- und Testdatensatz wurden im Verhältnis 80:10:10 aufgeteilt. Der Validierungsdatensatz wurde verwendet, um folgende Hyperparameter zu optimieren: 'wordNgrams', 'maxn'. 'wordNgrams' steht dabei für eine Token-n-Gramm-Zerlegung und 'maxn' für eine Character n-Gramm Aufteilung. Die Zerlegung wurde jeweils im Bereich $2 \leq n < 6$ durchgeführt bei einer Lernrate von 'lr = 1.0'. Für die restlichen Parameter wurden die Grundeinstellungen belassen. Der Parameter n wurde programmatisch variiert und die beste Performance über das Validierungsset ermittelt. Das trainierte Modell bei dieser Parameterkombination wurde auf den Testdatensatz angewandt. Abschnitt 3.2.1

zeigt dabei die Ergebnisse bei Verwendung eines vorab trainierten Sprachmodells (83) und Abschnitt 3.2.2 die Ergebnisse ohne Verwendung eines vortrainierten Sprachmodells bei einer Embedding-Dimensionalität von 300.

Sprachmodell in Verwendung

Class	Precision	Recall	F_1-Score	Support
Current drinker of alcohol	0,75	0,63	0,69	19
Current non-drinker of alcohol	0,82	0,87	0,84	31
Ex-problem drinker	0,70	0,78	0,74	9
Disorder caused by alcohol	0,76	0,76	0,76	25
Problem drinker	0,90	0,95	0,92	38
Alcohol consumption unknown	0,79	0,71	0,75	21
macro avg	0,79	0,78	0,78	143

Tabelle 5. Evaluierungsergebnisse des Testdatensatzes mit Verwendung eines Sprachmodells ('wordNgrams=1', 'maxn'=5').

Sprachmodell nicht in Verwendung

Class	Precision	Recall	F_1-Score	Support
Current drinker of alcohol	0,75	0,63	0,69	19
Current non-drinker of alcohol	0,85	0,90	0,88	31
Ex-problem drinker	0,88	0,78	0,82	9
Disorder caused by alcohol	0,71	0,80	0,75	25
Problem drinker	0,90	0,97	0,94	38
Alcohol consumption unknown	0,82	0,67	0,74	21
makro avg	0,82	0,79	0,80	143

Tabelle 6. Evaluierungsergebnisse des Testdatensatzes ohne Verwendung eines Sprachmodells ('wordNgrams=2', 'maxn'=5').

Diskussion

Bevor die gewonnenen Erkenntnisse dieser Diplomarbeit und die daraus resultierenden Möglichkeiten und Fragestellungen genauer behandelt werden, soll

mittels Zusammenfassung der vorliegenden Arbeit das Verständnis des Lesers nochmal unterstützt werden.

Zu Beginn dieser Diplomarbeit wurde die Bedeutung des geschriebenen Wortes im medizinischen Sektor aufgezeigt und anhand ausgewählter Beispiele veranschaulicht.

Anschließend wurde erläutert, aus welchen Gründen Daten im Gesundheitssektor erhoben werden, wie sich diese einteilen lassen, und welches Potential medizinische Daten aufweisen. Hier konnte erkannt werden, dass die medizinische Dokumentation verschiedene Zielsetzungen verfolgt und zu riesigen Datenmengen führt. Diese sind trotz des Vorliegens von herausfordernden Struktur-Standardisierungsverhältnissen bereits heute als Ressource für zahlreiche, informationstechnologische Anwendungen auch in Kontext von Sekundärnutzungsszenarien im medizinischen Sektor nutzbar.

Es folgte eine allgemeine Einführung in den Bereich des NLP (Natural Language Processing). Hierbei wurden die wichtigsten Begriffe des Fachgebietes erklärt, und der Aufbau eines allgemeinen NLP-Systems beschrieben.

In weiterer Folge wurde klinisches NLP (cNLP) genauer beleuchtet und beispielhaft anhand gewählter Veröffentlichungen aufgezeigt, dass cNLP bereits heute vielseitig eingesetzt wird, um die Diagnostik und Behandlung von Patienten zu optimieren.

Ziel dieser Diplomarbeit war es, die Unterschiede zwischen einer manuellen und einer cNLP-basierten Analyse und Klassifikation von unstrukturierten Daten anhand von gewählten Methoden des maschinellen Lernens aufzuzeigen bzw. zu bewerten. Hierfür wurden Personen hinsichtlich ihres dokumentierten Alkoholkonsums unterschieden. Der Alkoholkonsum als Klassenvariable wurde aufgrund seiner standardisierten Erhebung während der Anamnese und seinen zahlreichen assoziierten gesundheitlichen Risiken gewählt.

In diesem Zusammenhang wurde zunächst ergründet, wie der Alkoholkonsum von Patientinnen und Patienten standardisiert aufgenommen und klassifiziert werden kann, bevor der S3-leitliniengeforderte Standard der Alkoholanamnese angeführt wurde. Es zeigte sich, dass zahlreiche Verfahren für die Erhebung des Alkoholkonsums zur Verfügung stehen, und dass für die Klassifikation verschiedene

Systeme verwendet werden können (DSM 5, ICD-10, SNOMED CT). Anschließend wurde aktuelle Literatur angeführt, die veranschaulichte, dass die cNLP-basierte Analyse und Klassifikation von Personen hinsichtlich ihres Alkoholkonsums bereits mehrfach auf Basis der Analyse freitextlicher klinischer Routinedokumentation, im englischsprachigen Raum erfolgreich durchgeführt werden konnte. Eine automatische cNLP-basierte Analyse von klinischen Texten aus dem deutschsprachigen Raum, im Kontext der Beurteilung des Alkoholstatus, ist hingegen nach aktueller Erkenntnis und Wissensstand noch nicht durchgeführt worden.

Im praktischen Teil dieser Diplomarbeit wurde zunächst die initiale Datenbasis, bestehend aus etwa 46.700 de-identifizierten Arztbriefen der Steiermärkische Krankenanstaltengesellschaft m. b. H. (KAGes) modifiziert und vorprozessiert. Hierbei entstanden 1429 relevante, in ihrer Struktur festgelegte Arztbriefausschnitte (Snippets). Nun wurde, basierend auf einer Literaturrecherche und den aufgeführten Kodiersystemen (v.a. SNOMED CT) und unter Betrachtung der neu erstellten Snippets, ein Einteilungssystem des Alkoholkonsums in sechs Klassen entwickelt und anhand einiger Beispiele veranschaulicht und motiviert.

Im Anschluss wurden die Snippets unter Verwendung der neu erstellten Klasseneinteilung manuell durch den Autor dieser Diplomarbeit klassifiziert. Eine zweite wissenschaftliche Mitarbeiterin annotierte gleichzeitig 20% der Arztbriefausschnitte. Hierbei kam es zu einem Cohen’s Kappa von 0,9, was für eine sehr gute Übereinstimmung der beiden Annotationen und gleichzeitig für eine hohe Effizienz und eine geringe Überschneidung der erstellten Klassen spricht. Daraufhin wurden die zu untersuchenden cNLP-Systeme dieser Diplomarbeit mit 90% der annotierten Snippets trainiert bzw. validiert. Die verbliebenen 10% wurden für das Testen der Klassifikationskomponente verwendet.

Die optimierte Komponente über eine Support-Vektor-Maschine erreicht einen durchschnittlichen, über alle sechs Klassen verteilten F_1-Score von 0,83. fastText, basierend auf einem neuronalen Netzwerk, konnte einerseits bei einer Verwendung eines Sprachmodells einen durchschnittlichen F_1-Score von 0,78 und andererseits ohne Verwendung eines vorab trainierten Sprachmodells einen F_1-Score von 0,80 erzielen. Das Interrater-Agreement einer Stichprobe des Datensatzes konnte dabei nicht erzielt werden.

Im Vergleich der Kennzahlen mit denen ähnlicher Publikationen (siehe Kapitel 1.4.4) weist diese Diplomarbeit sehr gute Resultate auf (F_1-Scores über 0,9, ähnlich Alzoubi et al. (62), für ausgewählte Klassen und durchschnittliche F_1-Scores im Bereich von 0,8, ähnlich Topaz et al. (64) über alle Klassen).

Hervorzuheben ist, dass die Klasse *problem drinker* bei beiden untersuchten Varianten des maschinellen Lernes (SVM, fastText) den zweithöchsten (SVM) bzw. höchsten F_1-Score (fastText) erreichte.

Wie bereits weiter oben beschrieben (siehe Kapitel 1.4.3) weist die Behandlung der frühen Phase des problematischen Alkoholkonsums die beste Heilungsprognose durch das Fehlen von irreversible Organschäden auf (57). Unabdingbare Voraussetzung der Erkennung einer solchen Frühphase ist das sichere Ermitteln problematischen Alkoholkonsums als solchem aus anamnestischen Daten. Hier setzten die Stärken der von mir untersuchten Verfahren, auch im Vergleich zu anderen Methoden, an. In der Folge ist es dann möglich in der Gruppe der *problem drinker* Frühphasen zu erkennen und daraus einen echten therapeutischen Nutzen zu ziehen.

Basierend auf den gewonnenen Resultaten dieser Arbeit, stellt die Ausweitung der computergestützten Analyse von Faktoren der Lebensführung, wie beispielsweise die Inklusion des Rauch- oder Ernährungsverhaltens, sicherlich den logischen nächsten Schritt in der Erprobung der untersuchten cNLP-Verfahren dar. So könnte in Zukunft eine umfassende, automatisierte Analyse der Lebensumstände eines Patienten / einer Patientin zu einer noch individuelleren Behandlung und somit zu besseren Therapieergebnissen führen.

Bevor die hier analysierten cNLP-Komponenten zur Beantwortung neuer Forschungsfragen herangezogen werden können, und ein nächster Schritt in Richtung Anwendbarkeit im klinischen Alltag gegangen werden kann, müssen an dieser Stelle auch die einhergehenden Limitationen der Verfahren sowie dieser Arbeit beachtet werden.

Limitationen

Datenqualität. Vorrangig muss das bekannte Problem der Datenqualität in Bezug auf den Alkoholstatus erwähnt werden. Da diese Studie ausschließlich auf klinischen Dokumenten beruht, kann keinerlei Aussage bezüglich der Validität der darin enthaltenen Angaben zum Alkoholstatus getroffen werden. Es ist nach der vorhandenen Literatur jedoch anzunehmen, dass auch in unseren Kliniken die Angaben zum Alkoholstatus nur bedingt der Realität entsprechen, insbesondere dort, wo die klinischen Probleme mit diesem wenig oder gar nicht in Verbindung stehen.

Datenumfang. Im Rahmen dieser Diplomarbeit wurde nicht erforscht, inwieweit Datensatzumfang und Leistung des NLP-Systems zusammenhängen. Eine klare Beantwortung der Frage, ab welcher Datensatzgröße eine maximale Performance des Systems erreicht werden kann, ist sicherlich für die weitere Verwendung des untersuchten NLP-Models maßgeblich und sollte deshalb in Zukunft genauer untersucht werden.

Generalisierbarkeit. Auch zur Robustheit des verwendeten NLP-Modells konnte im Laufe dieser Diplomarbeit keine Stellung genommen werden. Die Robustheit eines NLP-Systems beschreibt, inwieweit die Performance des Verfahrens vom Ursprung der verwendeten Datenbasis abhängt, also inwieweit eine Abhängigkeit von der Datendomain besteht. Es stellt sich also die Frage, ob das bereits trainierte System vergleichbar gut klassifiziert, wenn die zu annotierenden Daten aus einer anderen Quelle stammen, als diejenige, die zum Trainieren des Systems benutzt wurden. Es gilt zu überprüfen, ob das untersuchte NLP-Modell beispielweise Daten eines anderen Krankenhauses in Österreich gleich erfolgreich klassifizieren würde. Ist diese Frage geklärt, könnte anschließend z.B. untersucht werden, ob Unterschiede zwischen deutschen und österreichischen Datenbasen auszumachen sind.

Nachvollziehbarkeit. Ein Problem von Ansätzen, die auf maschinellem Lernen beruhen, ist die fehlende Nachvollziehbarkeit (auch Explainability genannt). So ist der Benutzter des Systems zu keinem Zeitpunkt der automatisierten Klassifikation in der Lage im Einzelnen nachzuvollziehen, anhand welcher Kriterien die Maschine ihre Datenobjekte klassifiziert. Man spricht an dieser Stelle von einer *Black Box*.

Gerade im medizinischen Bereich, wo Individualentscheidungen im schlimmsten Fall über Leben und Tod bestimmen, sollte eine solche Nachvollziehbarkeit, allein schon aus juristischen Gesichtspunkten, zu jedem Zeitpunkt gegeben sein. Dieses Problem stellt eine allgemeine Limitation maschineller Lernverfahren dar und hat sich bereits zu einem zentralen Aspekt der Forschung hinsichtlich künstlicher Intelligenz entwickelt (84,85).

Ausblick

Zusammenfassend zeigt sich, dass der Einsatz von NLP im klinischen Setting, trotz vorhandener Limitationen, bezüglich des hier untersuchten Teilaspektes der Analyse und Einteilung von medizinischen Daten eine hilfreiche Ergänzung der rein manuellen Analyse bietet. In diesem Zusammenhang sollte in Zukunft bei der Planung von retrospektiven Studien, vor allem aber auch bei der direkten Patientenversorgung, cNLP als Werkzeug dort eingesetzt werden, wo Datenverarbeitungsprozesse damit sinnvoll unterstützt werden können. So ließe sich der finanzielle wie auch personelle Aufwand bei der Analyse von textbasierten Datenmengen (gerade in Bezug auf Vorbefunde, die aufgrund von Zeitmangel oft ignoriert werden) möglicherweise reduzieren. Eine individuellere Therapieanpassung durch detaillierte, entscheidungsrelevante Information könnte zudem die medizinische Versorgung verbessern.

Literaturverzeichnis

1. Daniels PT, Bright W. The World’s Writing Systems. Oxford University Press; 1996. 970 S.

2. Trigger BG. Writing systems: A case study in cultural evolution. Norwegian Archaeological Review. 1. Januar 1998;31(1):39–62.

3. Haarmann H. Geschichte der Schrift. C.H.Beck; 2002. 132 S.

4. Li I, Pan J, Goldwasser J, Verma N, Wong WP, Nuzumlalı MY, u. a. Neural Natural Language Processing for Unstructured Data in Electronic Health Records: a Review. arXiv:210702975 [cs] [Internet]. 6. Juli 2021 [zitiert 04. Juni 2023]; Verfügbar unter: http://arxiv.org/abs/2107.02975

5. Piai S, Claps M. Bigger Data for Better Healthcare. September 2013;

6. Charitha M, Cholli NG. Big Data Analysis and Management in Healthcare. 2021;03(07):12.

7. Wandelt S, Rheinländer A, Bux M, Thalheim L, Haldemann B, Leser U. Data Management Challenges in Next Generation Sequencing. Datenbank-Spektrum. 1. November 2012;12.

8. Bahri S, Zoghlami N, Abed M, Tavares JMRS. BIG DATA for Healthcare: A Survey. IEEE Access. 2019;7:7397–408.

9. Shafqat S, Kishwer S, Rasool RU, Qadir J, Amjad T, Ahmad HF. Big data analytics enhanced healthcare systems: a review. J Supercomput. 1. März 2020;76(3):1754–99.

10. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. J Am Med Inform Assoc. 2013;20(1):117–21.

11. Leiner F. Medizinische Dokumentation: Grundlagen einer qualitätsgesicherten integrierten Krankenversorgung ; Lehrbuch und Leitfaden ; mit 24 Tabellen. Schattauer Verlag; 2012. 258 S.

12. Cleve J, Lämmel U. Data Mining. Walter de Gruyter GmbH & Co KG; 2014. 377 S.

13. Hildebrand K, Gebauer M, Hinrichs H, Mielke M. Daten- und Informationsqualität: Auf dem Weg zur Information Excellence. Springer-Verlag; 2009. 419 S.

14. Müller ML, Ückert F, Bürkle T, Prokosch HU. Cross-institutional data exchange using the clinical document architecture (CDA). International Journal of Medical Informatics. März 2005;74(2–4):245–56.

15. Mohr MTJ, Lange T, Schall T, Nerlich M. XML in der (tele-)medizinischen Kommunikation: Sektorübergeifende Interoperabilität. In: Siewert JR, Hartel

W, Herausgeber. Digitale Revolution in der Chirurgie. Berlin, Heidelberg: Springer; 2002. S. 904–904. (Deutsche Gesellschaft für Chirurgie).

16. Ilyasova N, Kupriyanov A, Paringer R, Kirsh D. Particular Use of BIG DATA in Medical Diagnostic Tasks. Pattern Recognit Image Anal. Januar 2018;28(1):114–21.

17. Bhardwaj R, Nambiar AR, Dutta D. A Study of Machine Learning in Healthcare. In: 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC) [Internet]. Turin: IEEE; 2017 [zitiert 04. Juni 2023]. S. 236–41. Verfügbar unter: http://ieeexplore.ieee.org/document/8029924/

18. Peckham J, Maryanski F. Semantic data models. ACM Comput Surv. September 1988;20(3):153–89.

19. Ross MK, Wei W, Ohno-Machado L. “Big Data” and the Electronic Health Record. Yearb Med Inform. August 2014;23(01):97–104.

20. De Mauro A, Greco M, Grimaldi M. A formal definition of Big Data based on its essential features. Library Review. 1. Januar 2016;65(3):122–35.

21. Manogaran G, Lopez D, Thota C, Abbas KM, Pyne S, Sundarasekar R. Big Data Analytics in Healthcare Internet of Things. In: Qudrat-Ullah H, Tsasis P, Herausgeber. Innovative Healthcare Systems for the 21st Century [Internet]. Cham: Springer International Publishing; 2017 [zitiert 04. Juni 2023]. S. 263–84. (Understanding Complex Systems). Verfügbar unter: https://doi.org/10.1007/978-3-319-55774-8_10

22. Jauk S, Kramer D, Großauer B, Rienmüller S, Avian A, Berghold A, u. a. Risk prediction of delirium in hospitalized patients using machine learning: An implementation and prospective evaluation study. Journal of the American Medical Informatics Association. 1. September 2020;27(9):1383–92.

23. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst. Dezember 2014;2(1):3.

24. Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, u. a. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. Journal of Biomedical Informatics. 1. September 2017;73:14–29.

25. Kumar E. Natural Language Processing. I. K. International Pvt Ltd; 2019. 220 S.

26. Liddy E. Natural Language Processing. School of Information Studies - Faculty Scholarship [Internet]. 1. Januar 2001; Verfügbar unter: https://surface.syr.edu/istpub/63

27. Barth AP. Algorithmik für Einsteiger: Für Studierende, Lehrer und Schüler in den Fächern Mathematik und Informatik. Springer-Verlag; 2013. 207 S.

28. Cormen TH, Leiserson CE, Rivest R, Stein C. Algorithmen - Eine Einführung. Walter de Gruyter GmbH & Co KG; 2017. 1339 S.

29. Hohagen F. Parsing: Eine Einführung in die maschinelle Analyse natürlicher Sprache. Springer-Verlag; 2013. 324 S.

30. El Naqa I, Murphy MJ. What Is Machine Learning? In: El Naqa I, Li R, Murphy MJ, Herausgeber. Machine Learning in Radiation Oncology [Internet]. Cham: Springer International Publishing; 2015 [zitiert 04. Juni 2023]. S. 3–11. Verfügbar unter: http://link.springer.com/10.1007/978-3-319-18305-3_1

31. Yao X, Liu Y. Machine Learning. In: Burke EK, Kendall G, Herausgeber. Search Methodologies [Internet]. Boston, MA: Springer US; 2014 [zitiert 04. Juni 2023]. S. 477–517. Verfügbar unter: http://link.springer.com/10.1007/978-1-4614-6940-7_17

32. Müller B, Reinhardt J, Strickland MT. Neural Networks: An Introduction. Springer Science & Business Media; 1995. 358 S.

33. Scherer A. Neuronale Netze: Grundlagen und Anwendungen. Springer-Verlag; 2013. 259 S.

34. McCrae JP, Rademaker A, Rudnicka E, Bond F. English WordNet 2020: Improving and Extending a WordNet for English using an Open-Source Methodology. 16. Mai 2020;6.

35. Névéol A, Zweigenbaum P, Section Editors for the IMIA Yearbook Section on Natural Language Processing. Clinical Natural Language Processing in 2015: Leveraging the Variety of Texts of Clinical Interest. Yearb Med Inform. August 2016;25(01):234–9.

36. Sabra S, Mahmood K, Alobaidi M. A Semantic Extraction and Sentimental Assessment of Risk Factors (SESARF): An NLP Approach for Precision Medicine: A Medical Decision Support Tool for Early Diagnosis from Clinical Notes. In: 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC). 2017. S. 131–6.

37. Doan S, Maehara CK, Chaparro JD, Lu S, Liu R, Graham A, u. a. Building a Natural Language Processing Tool to Identify Patients With High Clinical Suspicion for Kawasaki Disease from Emergency Department Notes. Academic Emergency Medicine. 2016;23(5):628–36.

38. Sohn S, Wi CI, Wu ST, Liu H, Ryu E, Krusemark E, u. a. Ascertainment of asthma prognosis using natural language processing from electronic medical records. Journal of Allergy and Clinical Immunology. 1. Juni 2018;141(6):2292-2294.e3.

39. Wi CI, Sohn S, Ali M, Krusemark E, Ryu E, Liu H, u. a. Natural Language Processing for Asthma Ascertainment in Different Practice Settings. J Allergy Clin Immunol Pract. Februar 2018;6(1):126–31.

40. Izquierdo JL, Ancochea J, Group SC 19 R, Soriano JB. Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing. Journal of Medical Internet Research. 28. Oktober 2020;22(10):e21801.

41. Lingeman JM, Wang P, Becker W, Yu H. Detecting Opioid-Related Aberrant Behavior using Natural Language Processing. AMIA Annu Symp Proc. 16. April 2018;2017:1179–85.

42. Zhong QY, Mittal LP, Nathan MD, Brown KM, Knudson González D, Cai T, u. a. Use of natural language processing in electronic medical records to identify pregnant women with suicidal behavior: towards a solution to the complex classification problem. Eur J Epidemiol. 1. Februar 2019;34(2):153–62.

43. Warrer P, Hansen EH, Juhl-Jensen L, Aagaard L. Using text-mining techniques in electronic patient records to identify ADRs from medicine use. British Journal of Clinical Pharmacology. 2012;73(5):674–84.

44. Saunders JB, Aasland OG, Babor TF, De La Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. Addiction. Juni 1993;88(6):791–804.

45. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA, for the Ambulatory Care Quality Improvement Project (ACQUIP). The AUDIT Alcohol Consumption Questions (AUDIT-C): An Effective Brief Screening Test for Problem Drinking. Archives of Internal Medicine. 14. September 1998;158(16):1789–95.

46. Hodgson R, Alwyn T, John B, Thom B, Smith A. THE FAST ALCOHOL SCREENING TEST. Alcohol and Alcoholism. 1. Januar 2002;37(1):61–6.

47. Ewing JA. Detecting Alcoholism: The CAGE Questionnaire. JAMA. 12. Oktober 1984;252(14):1905–7.

48. O'Brien CP. The CAGE Questionnaire for Detection of Alcoholism. JAMA. 5. November 2008;300(17):2054–6.

49. Hasin D. Classification of Alcohol Use Disorders. Alcohol Res Health. 2003;27(1):5–17.

50. National Institute on Alcohol Abuse and Alcoholism, Herausgeber. Alcohol Use Disorder: A Comparison Between DSM-IV and DSM-5. Oktober 2021;

51. Hirsch JA, Nicola G, McGinty G, Liu RW, Barr RM, Chittle MD, u. a. ICD-10: History and Context. American Journal of Neuroradiology. 1. April 2016;37(4):596–9.

52. World Health Organization. The ICD-10 classification of mental and behavioural disorders : clinical descriptions and diagnostic guidelines

[Internet]. World Health Organization; 1992 [zitiert 04. Juni 2023]. Verfügbar unter: https://apps.who.int/iris/handle/10665/37958

53. Jacobsen A, de Miranda Azevedo R, Juty N, Batista D, Coles S, Cornet R, u. a. FAIR Principles: Interpretations and Implementation Considerations. 31. Januar 2020;10–29.

54. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, u. a. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 15. März 2016;3(1):160018.

55. Lee D, de Keizer N, Lau F, Cornet R. Literature review of SNOMED CT use. Journal of the American Medical Informatics Association. 1. Februar 2014;21(e1):e11–9.

56. Schulz S. Wozu benötigen wir standardisierte Terminologien wie SNOMED CT? Swiss Medical Informatics. 1. Januar 2011;27–32.

57. Füeßl H, Middeke M. Anamnese und klinische Untersuchung. Georg Thieme Verlag; 2010. 520 S.

58. Manthey J, Kilian C, Schomerus G, Kraus L, Rehm J, Schulte B. Alkoholkonsum in Deutschland und Europa während der SARS-CoV-2 Pandemie. SUCHT. 1. Oktober 2020;66(5):247–58.

59. Soyka M. Diagnostik und Therapie der Alkoholabhängigkeit. DNP. 1. Oktober 2018;19(5):53–9.

60. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und, Nervenheilkunde (DGPPN), Deutsche Gesellschaft für Suchtforschung und Suchttherapie e.V. (DG-SUCHT), Herausgeber. S3-Leitlinie „Screening, Diagnose und Behandlung alkoholbezogener Störungen" AWMF-Register Nr. 076-001. 2020.

61. Johnson AEW, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, u. a. MIMIC-III, a freely accessible critical care database. Sci Data. 24. Mai 2016;3:160035.

62. Alzoubi H, Ramzan N, Alzubi R, Mesbahi E. An Automated System for Identifying Alcohol Use Status from Clinical Text. In: 2018 International Conference on Computing, Electronics Communications Engineering (iCCECE). 2018. S. 41–6.

63. Topaz M, Murga L, Bar-Bachar O, McDonald M, Bowles K. NimbleMiner: An Open-Source Nursing-Sensitive Natural Language Processing System Based on Word Embedding. CIN: Computers, Informatics, Nursing. November 2019;37(11):583–90.

64. Topaz M, Murga L, Bar-Bachar O, Cato K, Collins S. Extracting Alcohol and Substance Abuse Status from Clinical Notes: The Added Value of Nursing Data. 2019;5.

65. Lix L, Munakala SN, Singer A. Automated Classification of Alcohol Use by Text Mining of Electronic Medical Records. Online J Public Health Inform [Internet]. 1. Mai 2017 [zitiert 04. Juni 2023];9(1). Verfügbar unter: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5462225/

66. Afshar M, Phillips A, Karnik N, Mueller J, To D, Gonzalez R, u. a. Natural language processing and machine learning to identify alcohol misuse from the electronic health record in trauma patients: development and internal validation. Journal of the American Medical Informatics Association. 1. März 2019;26(3):254–61.

67. To D, Sharma B, Karnik N, Joyce C, Dligach D, Afshar M. Validation of an alcohol misuse classifier in hospitalized patients. Alcohol. Mai 2020;84:49–55.

68. Phillips A. A Study Into the Feasibility of Using Natural Language Processing and Machine Learning for the Identification of Alcohol Misuse in Trauma Patients [MASTER THESIS]. [CHICAGO, IL]: LOYOLA UNIVERSITY CHICAGO; 2018.

69. World Health Organization. Global Status Report on Alcohol and Health 2018. World Health Organization; 2019. 472 S.

70. Gutjahr E, Gmel G, Rehm J. Relation between average alcohol consumption and disease: an overview. Eur Addict Res. August 2001;7(3):117–27.

71. Alfred Uhl, Sylvia Gaiswinkler, Markus Hojni, Alexandra Puhm, Julian Strizek. Handbuch Alkohol Österreich Band 3: Ausgewählte Themen. 2021.

72. O'Connor C, Joffe H. Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. International Journal of Qualitative Methods. 1. Januar 2020;19:1609406919899220.

73. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276–82.

74. Ogren P, Bethard S. Building Test Suites for UIMA Components. In: Proceedings of the Workshop on Software Engineering, Testing, and Quality Assurance for Natural Language Processing (SETQA-NLP 2009) [Internet]. Boulder, Colorado: Association for Computational Linguistics; 2009 [zitiert 04. Juni 2023]. S. 1–4. Verfügbar unter: https://aclanthology.org/W09-1501

75. Joachims T. Text categorization with Support Vector Machines: learning with many relevant features. In: Proceedings of the 10th European Conference on Machine Learning [Internet]. Berlin, Heidelberg: Springer-Verlag; 1998 [zitiert 04. Juni 2023]. S. 137–42. (ECML'98). Verfügbar unter: https://doi.org/10.1007/BFb0026683

76. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intelligent Systems and their Applications. Juli 1998;13(4):18–28.

77. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching Word Vectors with Subword Information [Internet]. arXiv; 2017 [zitiert 04. Juni 2023]. Verfügbar unter: http://arxiv.org/abs/1607.04606

78. Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of Tricks for Efficient Text Classification [Internet]. arXiv; 2016 [zitiert 04. Juni 2023]. Verfügbar unter: http://arxiv.org/abs/1607.01759

79. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:181004805 [cs] [Internet]. 24. Mai 2019 [zitiert 04. Juni 2023]; Verfügbar unter: http://arxiv.org/abs/1810.04805

80. Maleki F, Ovens K, Najafian K, Forghani B, Reinhold C, Forghani R. Overview of Machine Learning Part 1. Neuroimaging Clinics of North America. November 2020;30(4):e17–32.

81. Sasaki Y. The truth of the F-measure. 26. Oktober 2007;6.

82. Anandarajan M, Hill C, Nolan T. Practical Text Analytics: Maximizing the Value of Text Data. Springer; 2018. 294 S.

83. Grave E, Bojanowski P, Gupta P, Joulin A, Mikolov T. Learning Word Vectors for 157 Languages. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) [Internet]. Miyazaki, Japan: European Language Resources Association (ELRA); 2018 [zitiert 04. Juni 2023]. Verfügbar unter: https://aclanthology.org/L18-1550

84. Bibal A, Lognoul M, de Streel A, Frénay B. Legal requirements on explainability in machine learning. Artif Intell Law. Juni 2021;29(2):149–69.

85. Burkart N, Huber MF. A Survey on the Explainability of Supervised Machine Learning. jair. 19. Januar 2021;70:245–317.