

Schlagwortbasierte und KI-unterstützte Erstellung von Operationsberichten am Beispiel der FESS

Dissertation

zur Erlangung des akademischen Grades
Dr. med.

an der Medizinischen Fakultät
der Universität Leipzig

eingereicht von:

Frau Valentina Wildfeuer

Geburtsdatum / Geburtsort:

26.11.1998 / Berlin

angefertigt an:

Universität Leipzig, Universitätsklinikum Leipzig AöR, Klinik und Poliklinik für Hals-, Nasen-,
Ohrenheilkunde

Direktor: Prof. Dr. med. Andreas Dietz

Betreuer:

Prof. Dr. med. Andreas Dietz

Dr. med. Markus Pirlich

Richard Bieck

Beschluss über die Verleihung des Doktorgrades vom: 19/09/2023

1	EINFÜHRUNG	3
1.1	KÜNSTLICHE INTELLIGENZ (KI)	3
1.1.1	<i>Künstliche Intelligenz – Allgemein</i>	3
1.1.2	<i>Was ist künstliche Intelligenz?</i>	3
1.1.3	<i>Wie beeinflusst künstliche Intelligenz den Markt des Gesundheitswesens?</i>	4
1.1.4	<i>Mediziner und KI im Gesundheitssystem</i>	6
1.2	HALS-, NASEN-, OHRENHEILKUNDE (HNO)	6
1.2.1	<i>KI in der Hals-, Nasen-, Ohrenheilkunde</i>	6
1.2.2	<i>Nasennebenhöhlen-Operationen</i>	7
1.3.	DOKUMENTATION IN DER MEDIZIN	7
1.3.1	<i>KI-basierte Dokumentation</i>	7
1.3.2	<i>Dokumentation von Nasennebenhöhlen-Operationen und unser Projekt</i>	8
1.4	KÜNSTLICHE INTELLIGENZ – ZUKUNFTSAUSSICHT	10
1.5	FRAGEBOGEN	10
1.6	ZIELSETZUNG	11
1.7	ZUSAMMENFASSUNG DER ERGEBNISSE	11
2	PUBLIKATION	13
3	ZUSAMMENFASSUNG	22
4	LITERATURVERZEICHNIS	27
5	ANLAGEN	29
1	AUFGABENANLEITUNG FÜR DIE STUDIENTEILNEHMER	29
2	AUSGEWÄHLTE OP-BERICHTE FÜR DIE EVALUATION	30
3	EVALUATIONSBOGEN	39

1 Einführung

1.1 Künstliche Intelligenz (KI)

1.1.1 Künstliche Intelligenz – Allgemein

„Künftig kann man sich aussuchen, ob man überhaupt noch manuell arbeiten will, wenn es sich um langweilige, wiederkehrende Tätigkeiten handelt.“ – Elon Musk, am 19.08.2021 in Palo Alto über einen humanoiden Roboter, welcher die KI-basierten Technologien der Tesla Fahrcomputer nutzen soll um sich zu orientieren und zu bewegen (Hohensee 2021).

Technologien, welche auf künstlicher Intelligenz (KI) basieren, werden tagtäglich von uns genutzt (Xiaoxuan Liu MBChB et al.; Bur et al. 2019; Liu et al. 2019). Meist unbemerkt und komplett in den Alltag integriert erleichtern sie diesen in Form von Gesichtserkennung, wenn man das Smartphone entriegelt oder abends nach dem Feierabend Filme und Serien von seinem Streamingdienst empfohlen bekommt (Xu 2021).

Mittlerweile leben wir in einer Welt, in der künstliche Intelligenz nicht mehr wegzudenken ist. KI-Programme in Unternehmen stehen kurz vor dem Durchbruch, die wichtigsten Technologien sind so weit ausgereift, um Anwendungen zu ermöglichen, die vor ein paar Jahren noch undenkbar waren. Somit werden Prozesse automatisiert, effizienter und nachhaltiger (Xu 2021).

1.1.2 Was ist künstliche Intelligenz?

Künstliche Intelligenz ermöglicht es Computerprogrammen Probleme zu lösen und Aufgaben in einer Form zu bewältigen, die so bisher nur dem menschlichen Gehirn vorbehalten war (McCarthy 2007). Seit den 50er Jahren des 20. Jahrhunderts konkretisiert sich die Forschung und Entwicklung von künstlicher Intelligenz und dem enormen Potenzial dieser Anwendungssysteme (Eberl und Ulrich 2016; Ahmad et al. 2021). Die Innovation solcher neuen Technologien besteht darin, dass diese sich kontinuierlich weiterentwickeln, indem sie ihre eigenen Erfolge und Niederlagen analysieren und aus diesen selbstständig lernen (Bur et al. 2019; Hashimoto et al. 2018; McCarthy 2007; Sharma et al. 2021). KI-Programme können mit Hilfe von visueller Wahrnehmung und Interpretation, Objekt- und Worterkennung sowie komplexer Entscheidungsfindung Prozesse optimieren und effizienter gestalten (Hashimoto et al. 2018; Letourneau-Guillon et al. 2020).

Grob lässt sich das weite Feld der künstlichen Intelligenz in das „Machine Learning“ und das „Natural Language Processing“ (NLP) aufteilen (Bur et al. 2019). Das NLP

lässt sich wiederum in das „Deep Learning“ und das „Un-“ und „Supervised Learning“ unterteilen (Sharma et al. 2021; Lorkowski et al. 2021; Bur et al. 2019).

Unter „Machine Learning“ werden Algorithmen zusammengefasst, die es ermöglichen Vorhersagen aus riesigen unstrukturierten Datenmengen zu treffen und so komplexe Probleme zu lösen oder zum Beispiel Mediziner* bei der Entscheidungsfindung zu unterstützen (Bur et al. 2019; Hashimoto et al. 2018). Sogenannte „Artificial Neural Networks“ (ANNs) basieren auf der Struktur eines menschlichen Gehirns und bestehen aus multiplen Ebenen und Vernetzung, wie die Synapsen im zentralen Nervensystem (Bur et al. 2019; Hashimoto et al. 2018). Mithilfe dieser können nicht-lineare Kalkulationen, das sogenannte „Deep Learning“, die komplexesten Aufgaben, wie Sprach- und Bilderkennung, lösen (Bur et al. 2019; Monshi et al. 2020; Bieck et al. 2020).

Es gibt „Training Data Sets“, welche genutzt werden, um maschinelle Lernmodelle zu konstruieren und „Test Data Sets“, die zur Evaluation der Performanz des Modells dienen (Bur et al. 2019).

„Natural Language Processing“ ermöglicht es Computersystemen die natürliche Sprache zu verstehen und zu interpretieren, um so unstrukturierte Informationen aus einfacher Sprache in strukturierte, maschinenverständliche Daten zu überführen und in „Machine Learning“-Algorithmen einzuspeisen (Bur et al. 2019; Letourneau-Guillon et al. 2020; Hashimoto et al. 2018). Diese Algorithmen können bei der Ergebnisfindung und Vorhersage von Ereignissen den Menschen unterstützen (Bur et al. 2019).

Die Leistung dieser komplexen Prozesse hängt stark von der Qualität und Größe der zur Verfügung stehenden Daten ab, je umfangreicher diese sind, desto präzisere Vorhersagen können getroffen werden (Bur et al. 2019; Letourneau-Guillon et al. 2020; Hashimoto et al. 2018).

1.1.3 Wie beeinflusst künstliche Intelligenz den Markt des Gesundheitswesens?

Vor allem im Gesundheitswesen ist die Integration von künstlicher Intelligenz revolutionierend (Bur et al. 2019; Ahmad et al. 2021). Insbesondere für die individuelle Medizin, die eine immer bedeutsamere Rolle einnimmt, sind die KI-Technologien ein unabdingbarer Grundbaustein (Grand View Research 2021). Auf dem Gesundheitsmarkt steigt die Anzahl der KI-Anwendungen bereits enorm (Grand View Research 2021; Bur et al. 2019). Insbesondere während der Covid-19 Pandemie gab es große Fortschritte in der Entwicklung und Implementierung von KI-basierten Applikationen im medizinischen Versorgungsbereich, um riesige Datenmengen zu analysieren und zu verarbeiten, beispielsweise in der Impfstoffherstellung oder dem Monitoring von kritisch erkrankten Patienten (Lorkowski et al. 2021). Der Marktwert

* Aus Gründen der besseren Lesbarkeit wird im folgenden Text auf die gleichzeitige Verwendung der Sprachformen männlich, weiblich und divers verzichtet. Das generische Maskulinum bezieht gleichermaßen alle Geschlechter mit ein.

beträgt über 15.4 Milliarden US-Dollar im Jahr 2022 und es wird prognostiziert, dass dieser von 2022 bis 2030 mit einer jährlichen Wachstumsrate von 38.40% ansteigen wird (Grand View Research 2022). Dieser Wachstumstrend wird durch junge Unternehmen, wie Start-Ups, die immer mehr in KI-Systeme investieren, noch stärker vorangetrieben (Grand View Research 2021).

Nordamerika ist Vorreiter und dominiert den Markt dank Faktoren wie der lokalen Präsenz von wichtigen technologischen und medizinischen Unternehmen, des wachsenden geriatrischen Anteils an der Bevölkerung und des immer präsenteren Bewusstseins über das Potenzial, welches KI-Technologien im medizinischen Sektor mit sich bringen (Grand View Research 2021; Bieck et al. 2020).

Die neuen Technologien werden möglicherweise ganze Disziplinen revolutionieren, jedoch werden sie die Mediziner nur sehr unwahrscheinlich komplett ersetzen können (Hashimoto et al. 2018; Bur et al. 2019; Ahmad et al. 2021; Topol 2019; Lorkowski et al. 2021). Vielmehr werden sie diese bei aufwendigen, repetitiven Aufgaben, welche zeitintensiv und ökonomisch ineffizient sind unterstützen und entlasten (Bur et al. 2019; Letourneau-Guillon et al. 2020; Lorkowski et al. 2021; Lorowski und Malinowska 2020). Somit werden neue Rollen und ärztliche Tätigkeitsgebiete sowie Raum und Zeit für andere, wichtigere Aufgaben geschaffen (Liu et al. 2019). Durch diese Entlastung können Behandlungsergebnisse verbessert, der Kostenaufwand reduziert und der klinikinterne und globale Austausch von Informationen und Wissen erleichtert werden (Grand View Research 2021; Lorowski und Malinowska 2020).

Insbesondere in der Diagnostik gibt es bereits große Fortschritte, da KI-basierte Technologien in diesem Bereich mithilfe von Bilderkennungssystemen die Ärzte bei der Entscheidungsfindung und Diagnosestellung unterstützen (Grand View Research 2021; Xiaoxuan Liu MBChB et al.; Liu et al. 2019; Topol 2019; Ahmad et al. 2021). Es ist, insbesondere zukünftig, möglich schon bevor Patienten erste Symptome zeigen Diagnosen zuverlässig und frühzeitig zu prognostizieren und auf den Patienten zugeschnittene, individuelle Risiko-Scores vollautomatisch zu konstruieren (Hashimoto et al. 2018; Grand View Research 2021). So können Krankheiten rechtzeitig detektiert und therapiert, mögliche Symptome reduziert und diesen präventiv entgegengewirkt werden.

Ein anschauliches Beispiel hier ist ein dermatologisches Diagnostiksystem, 3DermSpot, welches dank KI-basiertem Bilderkennungsprogramm fast vollständig autonom Hautkrebsscreenings zuverlässig durchführen kann und sich im Zulassungsverfahren der FDA befindet (Whooley 2020; 3Derm Systems, Inc. 2020). Dieses KI-System hat als erstes seiner Art in der Dermatologie und als zweites KI-Device überhaupt in der Medizin eine bahnbrechende Auszeichnung der FDA erhalten (Whooley 2020).

Der Großteil der bereits existierenden KI-Systeme in der Medizin werden im Bereich der diagnostischen Radiologie entwickelt und implementiert, da diese mithilfe von Bilderkennungssoftwares enorme Fortschritte in der Befundung von Tomographien und anderen bildgebenden Verfahren erzielen können (Liu et al. 2019; Monshi et al. 2020; Ahmad et al. 2021).

1.1.4 Mediziner und KI im Gesundheitssystem

Bis jetzt sind KI-Technologien noch nicht vollständig im Gesundheitssystem etabliert (Bur et al. 2019). Deshalb ist es von sehr großer Bedeutung sich als Mediziner mit diesen innovativen Computersystemen vertraut zu machen und die Ängste und Sorgen, welche meist mit neuen Anwendungen und Technologien einhergehen, zu reduzieren (Eryigit et al. 2019; Xu 2021). Einige Ärzte befürchten, dass sich die neuen Anwendungen und Funktionen negativ auf die Arbeitslast auswirken könnten, weil diese erst verstanden und erlernt werden müssen (Eryigit et al. 2019; Xu 2021). Diese Sorgen sind jedoch meist unbegründet, da in Wahrheit der Aufwand minimal ist im Gegensatz zu dem Nutzen und der Zeitersparnis, welche solche Technologien mit sich bringen (Eryigit et al. 2019; Bur et al. 2019; Xu 2021).

Damit sich künstliche Intelligenz entfalten und auf dem großen Markt des Gesundheitswesens integriert werden kann, sind Mediziner unabdingbar (Bur et al. 2019; Hashimoto et al. 2018; Eryigit et al. 2019; Xu 2021). Die Implementierung KI-basierter Technologien in der Medizin wird immer bedeutsamer, um eine Optimierung der Patientensicherheit und ökonomisches Arbeiten zu realisieren und die Zufriedenheit der Ärzte zu steigern (Letourneau-Guillon et al. 2020; Hashimoto et al. 2018; Bur et al. 2019; Liu et al. 2019; Monshi et al. 2020; Anteby et al. 2021; Topol 2019).

Für die Entwicklung und das Training dieser Anwendungssysteme ist das Sammeln von großen Mengen an standardisierten Daten über Patienten und klinikinterne Prozesse notwendig (Bur et al. 2019; Letourneau-Guillon et al. 2020; Hashimoto et al. 2018). Mediziner müssen mit Ingenieuren und Informatikern in enger Kooperation arbeiten, damit die klinischen Daten und Informationen korrekt im medizinischen Kontext interpretiert werden können (Hashimoto et al. 2018; Eryigit et al. 2019). Nur so sind Programmierer in der Lage, die richtigen Ansätze zu finden, passende Algorithmen zu entwickeln und Bias zu minimieren, um Mediziner bestmöglich unterstützen und entlasten zu können (Hashimoto et al. 2018).

In Zukunft könnte ein großes weltweites Netzwerk an Daten geknüpft und so globales Wissen geteilt werden (Hashimoto et al. 2018). Dies würde eine Innovation in der Gründung und Validierung eines weltweit einheitlichen Qualitätsstandards und evidenzbasierter Medizin bedeuten (Hashimoto et al. 2018; Eryigit et al. 2019; Letourneau-Guillon et al. 2020).

1.2 Hals-, Nasen-, Ohrenheilkunde (HNO)

1.2.1 KI in der Hals-, Nasen-, Ohrenheilkunde

Auch in der Hals-, Nasen-, Ohrenheilkunde (HNO) gibt es viele Bereiche, insbesondere die operativen (Hashimoto et al. 2018), die großes Potenzial bergen von KI-Systemen revolutioniert zu werden (Bur et al. 2019). Dies zeigt sich allein schon in

dem enormen Anstieg der publizierten Artikel über KI Themen in der HNO (Bur et al. 2019). Beispielsweise wurden bereits erste Ansätze für KI-basierte Systeme zur Detektion von Risikostrukturen während der endoskopischen Sinus-Operation entwickelt, um die Patientensicherheit und das Behandlungsergebnis zu optimieren (Bur et al. 2019; Bieck et al. 2020).

1.2.2 Nasennebenhöhlen-Operationen

Die endoskopisch geführte Nasennebenhöhlen-Operation (NNH-OP) ist eine der häufigsten Standardoperationen, die in der Hals-, Nasen-, Ohrenheilkunde durchgeführt werden.

Bei dieser Methode ist es möglich minimalinvasiv zu operieren, wodurch das Trauma und Blutungsrisiko minimiert wird und so die Wundheilungszeit beachtlich reduziert werden kann (Bieck et al. 2020). Mithilfe von Navigationssystemen, welche sich an den Strukturen im Endoskopvideo mithilfe von Bilderkennungssystemen orientieren, können noch bessere Behandlungsergebnisse erzielt werden (Bieck et al. 2020).

Die NNH-OP dient zur Wiederherstellung der Drainage der Nasennebenhöhlen, um eine verbesserte Belüftung zu gewährleisten. Hierfür wird polypös erkranktes beziehungsweise entzündetes Gewebe entfernt und gesunde Schleimhaut belassen sowie die Rima olfactoria geschont, um das Riechvermögen zu erhalten (Bieck et al. 2020; Simmen und Jones 2005).

Indikationen stellen insbesondere chronisch-entzündliche Erkrankungen der Nasennebenhöhlen dar, aber auch starke Beeinträchtigungen der Nasenatmung beispielsweise aufgrund einer ausgeprägten Polyposis nasi (Nasenpolypen) oder auch einer zystischen Fibrose (Leunig 2010).

Ziel dieser Operation ist es, die Belüftung und das Riechvermögen zu verbessern und eine erleichterte Applikationsmöglichkeit sowie eine daraus resultierende bessere Wirksamkeit von topischen Medikamenten zu schaffen (Leunig 2010).

1.3. Dokumentation in der Medizin

1.3.1 KI-basierte Dokumentation

Die Dokumentation wird immer bedeutsamer und ist essenziell für die Patientensicherheit und Qualitätssicherung sowie für die Abrechnung von Leistungen und die Klärung rechtlicher Fragen (Eryigit et al. 2019; Letourneau-Guillon et al. 2020). Wir konzentrieren uns in dieser Arbeit auf die Dokumentation beziehungsweise Berichterstattung von Operationen (OP) der Nasennebenhöhlen (NNH).

Die Erstellung von OP-Berichten geschieht meist erst Stunden nach der eigentlichen Operation und die Prozedur muss aus dem Gedächtnis reproduziert werden. Dies ist sehr fehleranfällig und für den Operateur mit großem Zeitaufwand verbunden

(Letourneau-Guillon et al. 2020; Eryigit et al. 2019; Jing et al. 2018; Lorkowski et al. 2021). Auch für erfahrenere Ärzte ist es eine sehr repetitive und ermüdende Aufgabe, wodurch ihre Effizienz und Produktivität stark reduziert wird (Lin et al. 2018; Jing et al. 2018). Allein aus ökonomischer Sicht hat ein unnötig großer Aufwand der Dokumentation enorme finanzielle Auswirkungen (Lorkowski et al. 2021; Lorkowski und Malinowska 2020). Ein Arzt investiert durchschnittlich 34% bis 55% seiner Arbeitszeit pro Tag in die Erstellung, Überprüfung und Korrektur von medizinischen Berichten (Lin et al. 2018). Dies bedeutet, dass sich zum Beispiel allein im Jahr 2015 in den USA mit 861.000 Medizinern und einem Jahresgehalt von durchschnittlich 294.000 US-Dollar die Zeitkosten für die Dokumentation auf beeindruckende 90 bis 140 Milliarden US-Dollar belaufen haben (Lin et al. 2018). Wenn allein nur ein Bruchteil dieses enormen Zeitaufwandes und den damit verbundenen ökonomischen Einbußen reduziert werden könnte, wäre es möglich, einen signifikanten Gewinn für Patienten, Ärzte und das Gesundheitssystem zu schaffen (Lin et al. 2018; Eryigit et al. 2019; Letourneau-Guillon et al. 2020).

Deshalb ist es von enormer Bedeutsamkeit, dass in diesem Feld KI-Technologien eingesetzt werden, um die Mediziner zu unterstützen und zu entlasten. Dadurch kann eine Steigerung der Produktivität und Zufriedenheit der Mediziner erreicht und die Fehlerrate in der Dokumentation verringert werden (Eryigit et al. 2019; Lin et al. 2018; Letourneau-Guillon et al. 2020; Finley et al. 2018; Lorkowski et al. 2021). Durch die Abnahme dieser zeitaufwendigen Aufgaben können sich die Ärzte auf andere Tätigkeiten konzentrieren und ökonomisch effizienter arbeiten (Eryigit et al. 2019; Finley et al. 2018; Letourneau-Guillon et al. 2020; Lorkowski et al. 2021).

Mithilfe einer fehlerfreien und automatisierten Dokumentation ist ferner eine bessere rechtliche Absicherung, optimierte Kommunikation und erleichterter Austausch über Qualitätssicherung und Patientensicherheit möglich. Eine Standardisierung der OP-Berichterstellung würde außerdem eine bessere Vergleichbarkeit der Berichte und einen allumfassenden Überblick über verschiedene chirurgische Prozeduren ermöglichen (Eryigit et al. 2019). Des Weiteren würde dies einen enormen Vorteil für die Ausbildung und das Training unerfahrener Operateure bieten (Eryigit et al. 2019; Letourneau-Guillon et al. 2020).

1.3.2 Dokumentation von Nasennebenhöhlen-Operationen und unser Projekt

Wir haben im Rahmen unseres Projektes ein Tool entwickelt, welches in Zukunft schlagwortbasiert die OP-Berichterstellung intraoperativ durchführen soll. So kann die Fehlerrate und die Arbeitslast stark reduziert werden (Bur et al. 2019; Hashimoto et al. 2018; Letourneau-Guillon et al. 2020; Lorkowski et al. 2021; Topol 2019).

Hierfür entwickelte unsere Arbeitsgruppe in enger Kooperation mit dem ICCAS (Innovation Center Computer Assisted Surgery) Leipzig ein neuronales Sprachmodell, welches mithilfe von „Natural Language Processing“ und dem sogenannten „Deep

Learning“, in der Lage ist eingesprochene Schlagwörter während der OP zu verarbeiten. Anhand der Wortreihenfolge und der Schlagwörter wird ein sinnvoll zusammenhängender OP-Bericht unmittelbar während der laufenden Operation erstellt. An dieser Stelle sei auf das BMT Paper 2021 „Generation of surgical reports using keyword-augmented next sequence prediction“ (Bieck et al. 2021) verwiesen, welches ebenfalls im Rahmen dieses Projektes entstand.

Im Zuge unseres Projektes erstellten wir zunächst eine Datenbasis von 1500 Sätzen aus 48 Berichten von Nasennebenhöhlen-Operationen (functional endoscopic sinus surgery, FESS). Diese speisten wir mit 150 spezifischen Schlagwörtern, welche von medizinisch erfahrener Personal zu unterschiedlichen Schritten einer solchen Operation und zu bestimmten Berichtssätzen zugeordnet wurden, für das Training der künstlichen Intelligenz in das neuronale Modell ein. Zugehörige Satzpaare wurden anschließend tokenisiert und mit einer Encoder-Decoder-Architektur im ersten Schritt durch den Encoder in sogenannte „latent features“ umgewandelt. Auf Basis des vorhergehenden Wortes und der „latent features“ generierte anschließend der Decoder einen passenden OP-Berichtssatz. Das Sprachmodell orientierte sich hierbei an der typischen Wortverteilung von FESS-Berichten, um so die Möglichkeit zu maximieren, einen korrekten und sinnvoll zusammenhängenden Berichtssatz zu generieren. Für das Training des Tools wendeten wir eine „k-fold cross-validation“ auf die existierenden 48 OP-Berichte an. Die Berichtssätze der vorhandenen OP-Berichte wurden in ein gemeinsames Datenset integriert und per Zufallsprinzip in Trainings- und Test-Sätze mit einem Verhältnis von 85:15 aufgeteilt. Um die Datenmengen künstlich zu erweitern, wurden mehrere Sequenzen an Satzaugmentationen, wie das Ergänzen, Austauschen oder Löschen von Sätzen durchgeführt. Den Encoder trainierten wir vorher mithilfe des German CC100 Datensets, um Einschränkungen der Encoder-Performanz durch die geringe Menge an spezifischen Schlagwörtern zu vermeiden.

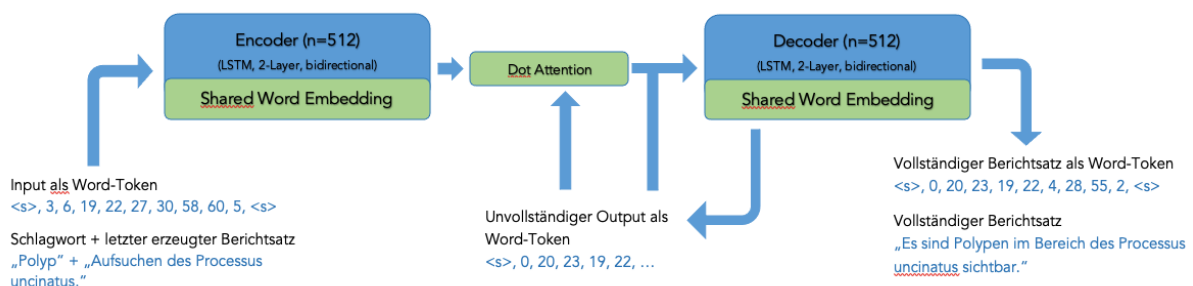


Abbildung 1: vereinfachte schematische Darstellung der Encoder-Decoder-Architektur

Mithilfe von vier verschiedenen Metriken, die bereits in der Literatur beschrieben wurden (Lin und Och 2004), konnten wir die Optimierung der Performanz durch das Training des Tools objektivieren. Wir verwendeten BLEU für die Satzspezifität, ROUGE für die Satzsensitivität, COSS als eine Art höherdimensionales Ähnlichkeitsmaß zwischen den generierten und den originalen Berichtssätzen und METEOR zur

Erfassung des F1-Werts der Sätze mit Gewicht auf Spezifität und Semantik. Diese Parameter werden auf einer Skala von 0-1 gemessen, wobei wir die konventionell erstellten OP-Berichte als Standard mit einem Wert von 1.0 festlegten.

Für die Evaluation unseres Tools wurden anschließend drei zufällig ausgewählte, künstlich generierte OP-Berichte von HNO-Ärzten korrekturgelesen und mit den konventionell nach der OP, erstellten Berichten verglichen. Zur Erfassung der subjektiven Zufriedenheit und möglichen Arbeitslastreduktion füllten die Teilnehmenden abschließend einen Evaluationsbogen aus. Die Studie wurde nach positivem Votum der Ethikkommission Leipzig mit der Ethikantragsnummer 455/20-ek durchgeführt.

1.4 Künstliche Intelligenz – Zukunftsaussicht

In den kommenden Jahren werden die künstliche Intelligenz und Systeme, welche auf ihr basieren, einen enormen Impact in der Medizin haben und dieses Feld in vielerlei Hinsicht positiv beeinflussen, optimieren und revolutionieren (Bur et al. 2019; Ahmad et al. 2021; Topol 2019).

Es ist daher wichtig Schritt zu halten und diese Entwicklung zu unterstützen, damit wir selbst in Zukunft von dieser profitieren können (Eryigit et al. 2019; Hashimoto et al. 2018). Nicht zuletzt gehört die Abnahme der ermüdenden, fehleranfälligen, aber sehr wichtigen OP-Dokumentation dazu. Somit wird mehr Zeit für andere, wichtige ärztliche Aufgaben geschaffen, die Arbeitslast reduziert und die Zufriedenheit der Mediziner gesteigert (Letourneau-Guillon et al. 2020; Hashimoto et al. 2018; Lorkowski et al. 2021).

1.5 Fragebogen

Um neben den technischen Daten auch subjektive Einschätzungen zu erheben, haben wir einen Evaluationsbogen mit 13 Fragen in fünf unterschiedlichen Kategorien entwickelt. Diesen sollten die Teilnehmer ausfüllen, nachdem sie drei zufällig ausgewählte, künstlich erzeugte Berichte korrekturgelesen und nach eigenem Empfinden bearbeitet haben. Die Anzahl und Art der Korrekturen wurden hierbei für jeden Bericht einzeln erfasst und ausgewertet.

In dem Evaluationsbogen wurden zu Beginn der Fachbereich und Spezialisierungsgrad (Kategorie F.1) erfragt sowie die Beurteilung der Qualität der von dem Tool erstellten Berichte in Grammatik, Inhalt und Lesefluss (Kategorie F.2). Anschließend wurden diese mit den konventionell erstellten OP-Berichten der jeweils gleichen Operation in Inhalt und Form verglichen (Kategorie F.3). Die konventionellen OP-Berichte wurden als Goldstandard verwendet, da kein offizieller Standard für die Form und den Inhalt eines OP-Berichtes existiert. Der Detailreichtum kann so zum Teil

von Bericht zu Bericht stark variieren, weil dies den individuellen Präferenzen des Operateurs überlassen ist.

In der Kategorie F.4 wurde der subjektiv eingeschätzte Zeitaufwand zur täglichen OP-Berichterstellung und die mögliche Zeitersparnis durch solch ein Tool erfasst. In einer ersten Version des Fragebogens wurde die Antwortmöglichkeit „keine Zeitersparnis“ nicht gegeben. Zur Vermeidung einer Ergebnisverzerrung wurde die Frage der Zeitersparnis um detailreichere Antwortoptionen ergänzt und anschließend von den Teilnehmern reevaluiert.

Letztendlich wurde der selbst empfundene Mehrwert und Nutzen und die mögliche Arbeitslastreduktion eingeschätzt und die Teilnehmer sollten angeben, ob sie selbst in Zukunft diese KI-basierte Technologie zur Unterstützung bei der Berichterstellung nutzen würden (Kategorie F.5)

1.6 Zielsetzung

Primärer Endpunkt dieses Projektes ist die Objektivierung der Performanzentwicklung durch Training des KI-basierten Sprachmodells zur intraoperativen OP-Berichterstellung anhand objektiver Metriken (BLEU, ROUGE, COSS, METEOR), die in der „Natural Language Processing“-Literatur verwendet werden (E.1). Außerdem wurde eine Evaluation der von unserem Tool erzeugten Berichte von NNH-Operationen im Hinblick auf die Qualität (E.2, E.3), die mögliche Zeitersparnis (E.4) sowie dem daraus resultierenden klinischen Nutzen (E.5) durchgeführt.

1.7 Zusammenfassung der Ergebnisse

Die vier Metriken, welche in unserem BMT Paper 2021 (Bieck et al. 2021) ausführlicher beschrieben werden, wurden auf einer Skala von 0-1 dargestellt, wobei den konventionell erstellten Berichten als Standard der Wert 1 zugeordnet wurde. Die Ergebnisse vor und nach dem Training des Tools wurden mit dem Wilcoxon-Test im SPSS (*Statistical Package for the Social Sciences, IBM, Version 23.0.0*) ausgewertet. In der Evaluation wurden zwei der 18 teilnehmenden Ärzte der HNO aus der Studie ausgeschlossen, da die Ergebnisse dieser Teilnehmer nicht auswertbar waren. Die Hälfte der 16 Teilnehmer waren Fachärzte, die anderen acht Assistenzärzte.

E.1 Schlagwortbasiertes Training des Sprachmodells:

Insgesamt wurden 48 künstlich erstellte mit den dazugehörigen konventionellen Berichten anhand von objektiven Parametern verglichen. Die Ergebnisse zeigten eine signifikante Verbesserung ($p < .001$) aller objektiven Parameter durch das schlagwortbasierte Training des KI-Sprachmodells.

E.2 Korrekturen:

Die Anzahl und Art der Korrekturen, welche die Teilnehmer in den künstlich erzeugten OP-Berichten vornahmen, wurden über *Microsoft Word*® aufgezeichnet und anschließend ausgewertet. Durchschnittlich 23.25 Korrekturen waren pro Bericht notwendig, um einen subjektiv zufriedenstellenden Bericht zu erhalten.

E.3 Grammatik, Inhalt und Lesefluss:

Im Schnitt wurde die Grammatik der künstlich erstellten Berichte von den Teilnehmern mit einer 3.1, der Inhalt mit einer 2.8 und der Lesefluss mit einer 2.6 bewertet.

E.4 Täglicher Zeitaufwand und Zeitersparnis:

66.67% der Teilnehmer gaben einen täglichen Zeitaufwand von 30 bis 60 Minuten für die Erstellung von OP-Berichten an. In einer detaillierteren Reevaluation schätzten 61.11% der Teilnehmer eine mögliche Zeitersparnis von 16 bis 30 Minuten und 27.78% eine Ersparnis von 1 bis 15 Minuten pro Tag durch die Unterstützung eines solchen Tools.

E.5 Klinischer Nutzen und Mehrwert:

61.11% der Teilnehmer erwarteten eine Arbeitslastreduktion und 66.67% sahen einen klinischen Mehrwert und Nutzen in der teilautomatisierten Erstellung von OP-Berichten. Die Mehrheit mit 66.67% würden dieses Tool zur OP-Berichterstellung in Zukunft selbst nutzen. 33.33% der Teilnehmer gaben eine inhaltliche Ähnlichkeit und 27.78% eine formale Ähnlichkeit zwischen den künstlich und konventionell erstellten Berichten an.

2 Publikation

Erschienen in:

International Journal of Computer Assisted Radiology and Surgery, online publiziert (17.11.2022).

Keyword-augmented and semi-automatic generation of FESS reports: a proof-of-concept study

V. Kunz^{1*}, V. Wildfeuer^{1*}, R. Bieck², M. Sorge¹, V. Zebralla¹, A. Dietz¹,
T. Neumuth², M. Pirlich¹

*(*geteilte Erstautorenschaft)*

¹ *Klinik und Poliklinik für Hals-, Nasen- und Ohrenheilkunde Universität Leipzig
Direktor: Prof. Dr. med. A. Dietz*

² *Innovation Center Computer Assisted Surgery (ICCAS) Universität Leipzig*



Keyword-augmented and semi-automatic generation of FESS reports: a proof-of-concept study

V. Kunz¹ · V. Wildfeuer¹ · R. Bieck² · M. Sorge¹ · V. Zebralla¹ · A. Dietz¹ · T. Neumuth² · M. Pirlich¹

Received: 7 January 2022 / Accepted: 2 November 2022
© The Author(s) 2022

Abstract

Introduction Surgical reports are usually written after a procedure and must often be reproduced from memory. Thus, this is an error-prone, and time-consuming task which increases the workload of physicians. In this proof-of-concept study, we developed and evaluated a software tool using Artificial Intelligence (AI) for semi-automatic intraoperative generation of surgical reports for functional endoscopic sinus surgery (FESS).

Materials and methods A vocabulary of keywords for developing a neural language model was created. With an encoder-decoder-architecture, artificially coherent sentence structures, as they would be expected in general operation reports, were generated. A first set of 48 conventional operation reports were used for model training. After training, the reports were generated again and compared to those before training. Established metrics were used to measure optimization of the model objectively. A cohort of 16 physicians corrected and evaluated three randomly selected, generated reports in four categories: “quality of the generated operation reports,” “time-saving,” “clinical benefits” and “comparison with the conventional reports.” The corrections of the generated reports were counted and categorized.

Results Objective parameters showed improvement in performance after training the language model ($p < 0.001$). 27.78% estimated a timesaving of 1–15 and 61.11% of 16–30 min per day. 66.66% claimed to see a clinical benefit and 61.11% a relevant workload reduction. Similarity in content between generated and conventional reports was seen by 33.33%, similarity in form by 27.78%. 66.67% would use this tool in the future. An average of 23.25 ± 12.5 corrections was needed for a subjectively appropriate surgery report.

Conclusion The results indicate existing limitations of applying deep learning to text generation of operation reports and show a high acceptance by the physicians. By taking over this time-consuming task, the tool could reduce workload, optimize clinical workflows and improve the quality of patient care. Further training of the language model is needed.

Keywords Artificial intelligence · Deep learning · Natural language processing · Neural language models · FESS · Surgical reports

Abbreviations

AI Artificial intelligence
BLEU Bilingual evaluation understudy
COSS Cosine similarity

FESS Functional endoscopic sinus surgery
GPU Graphics processing unit
IBM International business machines corporation
LSTM Long short-term memory
METEOR Metric for evaluation of translation with explicit ordering
ML Machine learning
NLP Natural language processing
ORL Otorhinolaryngology
p P-value
ROUGE Recall-oriented understudy for gisting evaluation
SD Standard deviation
SPSS Statistical package for the social sciences

V. Kunz and V. Wildfeuer have contributed equally to this work.

✉ V. Kunz
Viktor.Kunz@medizin.uni-leipzig.de

¹ Department of Otolaryngology, Head and Neck Surgery, University Hospital Leipzig, Liebigstraße 10-14, 04103 Leipzig, Germany

² Innovation Center Computer Assisted Surgery (ICCAS), University of Leipzig, Leipzig, Germany

Introduction

Artificial Intelligence (AI) represents the continuous digital development of our environment, especially in the health sector, where the number of applications is expected to increase in the future. In 2020, the AI market within the healthcare system had an estimated value of around 5.9 billion US-Dollar [1]. Still controversial remains the question whether AI developments will fundamentally revolutionize the healthcare system or even replace entire areas of it. It is known that these applications can solve complex and time-consuming problems in the clinical routine. So far, AI is not broadly established and consistently implemented in the healthcare sector [2]. However, a large spectrum of AI-based applications exists in development and clinical trials that are closely related to medical imaging, e.g., in diagnostic radiology [3, 4]. Conversely, there is still far-reaching potential for the specific field of surgery and, respectively, surgical procedures—as explained in the following.

One sub-genre of AI is machine learning, which currently receives the most attention in research. Specifically, computer vision (CV) and Natural Language Processing (NLP) are core algorithm domains that solve complex tasks such as object detection and text analysis. NLP allows technologies to understand and interpret natural language for transferring unstructured data into structured and machine-understandable information [2, 5, 12]. The implementation of CV and NLP applications could change and improve the clinical routine of many fields within the healthcare system, especially in otorhinolaryngology (ORL) [2]. It was shown that AI applications, in general, could decrease the workload and optimize the workflow of physicians by overtaking complex tasks, e.g., automatically generated radiology reports by analyzing medical images with connected natural language descriptions [5].

Usually, surgical reports are written in free text form after the actual procedure by the performing surgeons is finished. Some authors reported advantages of standard text templates for the creation of reports compared to conventionally using free-form reports [6]. Due to the difficulty of writing reports from memory—at some points hours after the actual procedure—this task is error-prone and time-consuming. Considering this, an application producing detailed, individual reports during the procedure might be a relevant contribution to optimizing the clinical workflow. NLP-based approaches where text keywords relate to a specific part of the procedure to produce coherent sentence structures already exist. These enable the generation of a complete and semantically logical surgical report. Regarding that topic, improvements in the field of image- and text-based production of reports based on NLP and deep learning processes were shown [4, 7]. Those previous works set the foundation for generating radiology reports (4) and prediction of

navigation steps during endoscopic operations by automatically analyzing image structures and datasets (7). The ability of “intelligent” technologies to generate reports by analyzing images or interpreting natural language like keywords is used in this work.

In this proof-of-concept study, an NLP tool was implemented to generate semi-automatically surgical reports based on existing reports of functional endoscopic sinus surgery (FESS) and keywords recorded during surgery. This study highlights the potential that a neural language model offers to produce complete and coherent surgical reports along with the actual procedure. Using this development, the daily workload of physicians could be reduced, and clinical workflows might be optimized. Furthermore, this AI application might increase physicians’ time for patient care and treatment.

Materials and methods

Neural language model and generation of surgical reports

In this proof-of-concept study, a machine-interpretable vocabulary was created for documented FESS procedures and enclosed in a neural language model. Specific keywords were defined and assigned to steps of the FESS procedure so that the model could create coherent sentence structures as they would be expected in conventional surgical reports. Therefore, 48 anonymized conventional reports of FESS, written by ten experienced surgeons, were retrospectively included in the model’s training. These reports had a length of 10 to 45 sentences and provided the baseline training data. Thus, a tool with the ability to artificially generate a coherent report from conventional text modules was created.

Subsequently, a database with 1500 sentences out of the 48 reports and an average number of 8 words per sentence was set up. In the first step, a vocabulary of 150 keywords relating to a specific part of the surgical reports was defined with the help of medical experts to create a sentence of anchor words, which are likely to be used as comments by the operating surgeons during the procedure. These words were enclosed in the model for training. Furthermore, the keywords or keyword combinations were matched with the corresponding sentences within the surgical reports through an annotation step by two experts. In this way, 2 to 4 keywords were assigned to each sentence. Respectively, two consecutive sentences were paired to create individual training patterns. These patterns consist of an initial sentence, a target sentence, and the specific keywords of the target sentence, which were assigned to the associated initial sentence.

The prepared sentence pairs were tokenized for obtaining a numeric representation for model training. We decided to train an unsupervised word-tokenizer with *SentencePiece*

[8], as it provides several benefits for generating surgical reports. Our tokenizer uses a fixed vocabulary size of $s = 500$, which is set before training the language model. This assures that the trained model is independent of language-specific aspects like the openness of the vocabulary. Additionally, an augmentation step is integrated to virtually increase database variability and improve the robustness of the model in the context of imbalanced word occurrence.

For the language model, we used an *encoder-decoder-architecture* which consists of bidirectional 2-layer-LSTMs (long short-term memory) for both coders. The encoder transforms the tokenized input sentences into a continuous intermediate representation, which is also known as *latent features*. The decoder further generates an output sentence word by word based on the previously generated word and these latent features of the input sentence. The decoder model is autoregressive and uses the last generated output word as an input word for generating the following word in the sentence so that the conditional probability can be maximized. The primary assumption of the language model is that surgical report sentences follow an intrinsic word distribution typical for reports of the ORL-Department. The language model training maximizes the probability of generating the correct word combination for the new reports by emulating this specific word distribution.

Additionally, we used the attention pattern by Chan et al. [9], which calculates a weighted context between the latent features of the encoder and the already generated output words for each step of the word generation. This enables the possibility that the decoder can respond to individual, word-specific dependencies between the input and the output sentence more accurately. Both the encoder and the decoder have a word embedding size of 512, which translates our numeric word representation into a vector representation optimized for GPU-based training.

For the model training, k-fold cross-validation with $k = 10$ was used for our 48 datasets of surgical reports. The report sentences of the left-in-datasets were combined into one dataset, shuffled, and divided into training and validation sentences with a ratio of 85:15. In the first step of the training pipeline, we artificially enlarged the data with a sequence of random sentence augmentations by swapping, deletion, addition, synonym and insertion operations in a sentence. Our sentence encoder was pretrained on the German CC100 dataset to offset the limitations of the currently small corpus of procedure-specific keywords.

To outline the effectiveness of the training process, model results were generated and compared before and after the training with objective machine translation metrics used in NLP literature, ROUGE, COSS, and METEOR [10]. All metrics were calculated by comparing generated and original report sentences with each other. BLEU (bilingual evaluation understudy) describes the sentences' specificity within

a range of 0–1. We employed BLEU-2 as the metric variation that allows us to compare how many two-word pairs in the generated reports match with the original sentences. Additionally, ROUGE (recall-oriented understudy for gisting evaluation) has been chosen to describe the sentences' sensitivity within a range of 0–1. This metric also identifies how many words in the generated sentences are also found in the original sentence. The cosine similarity metric (COSS) is a custom metric that describes the similarity based on angular distance between two vectors in space. We used the vectors generated in our word embedding layers of the model representing a word in a vector with 500 values). The COSS is the mean angular distance between each word of the generated sentence and the original one with a range of 0–1. A value of 0 indicates an orthogonal relationship with the lowest semantic similarity, while a value of 1 would indicate maximum word similarity. Lastly, we chose the METEOR metric (metric for evaluation of translation with explicit ordering) that describes the F1-value of the sentences with more focus on specificity and semantics. The metric generates values in a range between 0 and 1. The conventional surgical reports were validated as the standard with a value of 1.0. Figure 1 displays a summary of the AI-assisted tool for generating surgical reports.

Study participants and data acquisition

For the evaluation of the artificially generated reports, 18 Otorhinolaryngologists, trainees and consultants in equal parts, were recruited. Thereby, only those who at least had clinical-theoretic knowledge of FESS were included, ensuring they were able to understand the possible benefits and optimization of clinical workflows by using this tool. Each physician received three AI-augmented, artificially generated reports with the related conventional one, as well as an accompanying task explanation and a link for the online evaluation via e-mail. For this first evaluation, three representative conventional reports were disassembled into specific keywords. On this basis, the tool generated new reports like it would perform during an operation with speaking in these keywords.

The surgical reports, including the editing information of the participants, were sent back to us via e-mail. Before analyzing the data, the documents were anonymized. This study was conducted under the approval granted by the University ethics committee (ethics approval: 455/20-ek).

Evaluation of the artificially generated reports

First, the participants were asked to correct three artificially generated reports until they were considered appropriate surgical reports. For the subsequent analysis of the corrections,

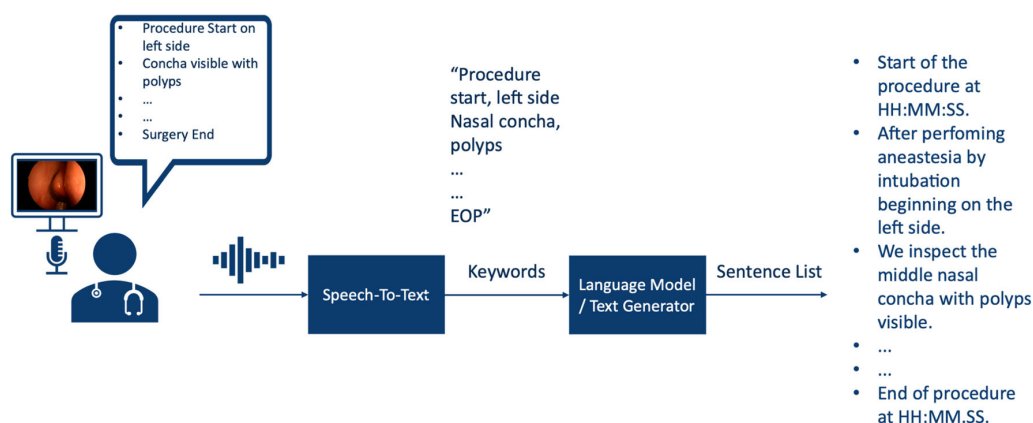


Fig. 1 Operating scheme of the NLP-assisted tool for generating surgical reports

all edits in the documents were counted and categorized using *Microsoft Word*. Then, the evaluation was accomplished online using a questionnaire with 13 questions created for this study, which was divided into the following four categories: (1) “quality of the generated surgical reports”, (2) “time-saving”, (3) “clinical benefits” and (4) “comparison with the conventional surgical reports”. In the first category, grammar, content, and reading flow of the artificially generated surgical reports had to be evaluated using school grades according to the German grading system. In the second category, the participants had to estimate their daily time consumption by writing conventional surgical reports after the procedure and the possible time saved by using this tool. To avoid possible distortions of the results, we reevaluated the question regarding daily time saving, as the option to select “no time saving” was not explicitly provided in a first version of the questionnaire. We replaced the original question by the reevaluated one and updated the online questionnaire. Within the third category, participants had to evaluate the “clinical benefits”, workload reduction, and their willingness to use this tool in the future. In the fourth and last category, the artificially generated and corrected reports were compared with the related conventional surgical reports concerning their form and content.

Statistical analysis

Statistical analysis was performed using *SPSS (Statistical Package for the Social Sciences, IBM, Version 23.0.0)*. Categorical data are reported as absolute and relative frequencies (proportions and percentages). Continuous data are expressed as mean (\pm SD). Wilcoxon’s test was used to analyze differences in objective parameters before and after the training of the language model.

Results

In total, 18 physicians participated in this study, out of which two had to be excluded because the corrections and recorded editing of the generated surgical reports from these participants were unavailable. Out of the 16 remaining participants, 50% were ORL-consultants, and 50% were physicians in training.

Corrections of artificially generated reports

The type and number of corrections done by the participants were recorded and evaluated (Table 1). An average of 23.25 corrections were necessary until the reports were subjectively considered appropriate surgical reports by the participants.

Grading of grammar, content, and reading flow

The participants gave an average grade of 3.1 for the category “grammar”, an average grade of 2.8 for “content,” and an average grade of 2.6 for “reading flow”.

Daily expenditure and possible saving of time

66.67% of physicians estimated a time-expenditure of about 30–60 min per day for conventional surgical reports after the actual procedure was performed. In a more detailed reevaluation, a majority of 61.11% estimated that 16–30 min per day can be saved by using this AI-assisted tool for generating surgical reports (Table 2). 27.78% guessed a timesaving of 1–15 min and 5.56% of 31–45 min per day (Table 2).

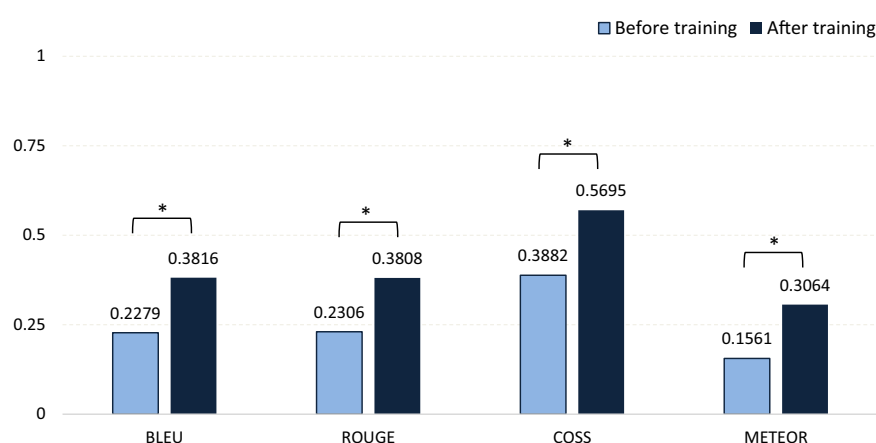
Table 1 Number of corrections after editing the artificially generated surgical reports

	Elements added	Elements removed	Grammar mistakes	Sentences added	Sentences removed	Corrections total
Corrections	9.3 ± 6.5	13.9 ± 7.2	13.4 ± 5.9	5.4 ± 4.8	8.1 ± 5.3	23.3 ± 12.5
Minimum	1	6	4	0	3	8
Maximum	26	32	23	14	18	58

Table 2 Estimated time generating conventional reports and time saved* using the AI-assisted tool

Expenditure of time for surgical reports (minutes/day)	0–30		30–60		
Participant's choice (%)	33.33		66.67		
Time saved using the AI-assisted tool (minutes/day)	0	1–15	16–30	31–45	46–60
Participant's choice* (%)	5.56	27.78	61.11	5.56	0.00

*Reevaluated question

Fig. 2 Wilcoxon's test for objective parameters of the neural language model

Keyword-augmented training of the language model

Objective parameters from 48 FESS reports were compared before and after training. The first metric BLEU increased with a Z_{BLEU} of -4.09 , and ROUGE increased with a Z_{ROUGE} of -3.84 , COSS with a Z_{COSS} of -4.61 , and METEOR with a Z_{METEOR} of -3.90 . The results of Wilcoxon's test show that the keyword-augmented training and optimization of the language model significantly ($p < 0.001$) increased all objective parameters (Fig. 2).

Clinical evaluation

The results show that 61.11% of the study participants expected a workload reduction by using this tool for generating surgical reports. Clinical benefits were seen by 66.66% of the participating physicians. 33.33% claimed that there is a similarity in the content between the generated and corrected reports compared to the conventionally generated

ones. 27.78% of the participants agreed that there is a formal similarity between the generated and corrected reports compared to the conventionally generated ones (Table 3 and Fig. 3).

Discussion

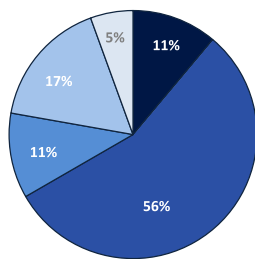
Implementing AI-based applications in clinical work is becoming more important [2–4]. Those technologies partially adopt repetitive and time-consuming tasks to reduce the physicians' workload and standardize and optimize the quality of surgical reports. Within this study, we developed and evaluated a neural language model for the semi-automatic generation of surgical reports using the example of FESS.

Number of corrections

The extent of corrections among the study participants varied considerably. This might either be caused by different levels

Table 3 Evaluation of advantages and content-related parameters of this tool

Parameters	Strongly agree (%)	Agree (%)	Undecided (%)	Disagree (%)	Strongly disagree (%)
Workload reduction	27.78	33.33	16.67	16.67	5.56
Clinical benefits	22.22	44.44	16.67	11.11	5.56
Would use tool in future	11.11	55.56	11.11	16.67	5.56
Strongly similar in content	0.00	33.33	50	11.11	5.56
Strongly similar in form	0.00	27.78	33.33	38.89	0.00



■ Strongly agree ■ Agree ■ Undecided ■ Disagree ■ Strongly disagree

Fig. 3 Pie chart of willingness to use the AI-assisted tool in the future

of training between residents and consultants, or due to individual adjustments and preferences of the participants. More studies are needed to further investigate the reason behind these findings. The total number of corrections indicates that there is still a need of editing the generated reports after the surgical procedure. Further training with more reports is necessary to optimize the generated reports' quality to reduce the time of correcting the reports after the procedure [2].

Daily expenditure and possible saving of time

The results of the reevaluation show that 27.78% of the participants are expecting a daily time saving of 1–15 min and 61.11% of 16–30 min, which indicates an important aspect for reducing the workload and optimizing the clinical workflow [5]. Through taking over repetitive and time-consuming tasks with these applications, such as writing reports [11], physicians can potentially shift valuable time to patient treatment [5, 12–14]. Thus, the quality and safety of patient care might improve as well, while mortality may decrease in this context [5]. Furthermore, these results indicate, that there is an acceptance of our tool by the participants, which increases the chance for implementation in clinical routine.

Metrics

For the objective evaluation of our model, we used metrics to assess neural machine translation. These parameters were used to validate the reports before and after the training. The results prove that the generated reports can improve significantly ($p < 0.001$) through the training model. This language model generates the reports based on the keywords and is also oriented to word order. Consequently, the most likely sentence is generated as it would be expected in a surgical report. By enlarging the training data set of more conventional reports, the model may generate more accurate and high-quality surgical reports before integration into a daily clinical routine [2].

Evaluation

About 66.66% of the participants see benefits, while 61.11% see a relevant workload reduction through this tool. Furthermore, 66.67% of the attended surgeons would use this tool for generating surgical reports in the future. On the one hand, these results show a great acceptance for implementing those new technological applications in the clinical workflow [6]. On the other hand, this emphasizes the requirement to revolutionize the repetitive and time-consuming task of writing surgical reports. This is in line with the review by Bur et al., who describe the future major impact of AI and the need to understand its principles and scopes, especially in ORL practice and documentation [2]. The similarity of form and content between the generated and the related conventional reports (considered as "gold-standard"), seen by about 30% of the participants, indicates a satisfying performance of the tool, which is continuously improving. Thus, the generated reports will soon be brought progressively in line with the conventional reports concerning form and content.

Limitations

There are some obstacles to overcome before this AI tool can be used in clinical practice in the future. First, the surgeons must be taught how to handle this keyword-based application to avoid misconceptions. Second, the surgeons using this tool must, at the moment, wear a microphone during the operation procedure to speak in their keywords. This has already been realized for education purposes. Furthermore, the language model must be able to distinguish between the keywords and outer-operative dialogs, as this could distort the content of the generated reports significantly. The conventional reports written by experienced surgeons were used as a gold standard in clinical routine and may vary in detail, which is left up to each individual surgeon. With the help of AI-based applications, it could be possible to create a certain standard of detail and information for operation reports. Moreover, we only recruited a small first cohort of 16 physicians with clinical occupations in ORL to evaluate this proof-of-concept study. In further evaluations, larger cohorts must prove the interface and handling of this tool for facilitating a user-friendly application. In this study, trainees and consultants were included in equal parts. The consultants are more experienced in documentation, so they might take less time, but are also performing more operations per day than trainees do. Thus, the time needed for writing surgical reports should be balanced between the trainees and the consultants. In a first version of the online questionnaire, within the question regarding possible time saved per day by using this tool, the option of “no time saving” was not explicitly provided. To avoid distortions of the results, we contacted the participants again and reevaluated this question and provided more detailed options to choose for the participants. We replaced the original question by the reevaluated question and updated the online questionnaire. It is relevant to reevaluate this item in further studies under clinical conditions after optimizing the tool with enlarged datasets. The number of corrections needed to gather an acceptable operation report also shows that optimizing the quality is necessary to reduce the postoperative editing expenditure time. Therefore, well-structured conventional reports are indispensable for enlarging the training data set and optimizing the quality of the artificially generated reports [5].

Prospects

Due to the ability of AI-based technologies to enhance themselves by learning from their successes and defeats as a feedback loop [2], it is relevant to generate a huge number of reports to improve the model’s performance continuously. Further optimization and training of this AI-based application are indispensable and a continuous process. Through this, the quality and accuracy of the generated reports can

and will be continuously improved. It is possible to upgrade the tool with other AI-based technologies to achieve a higher quality of reports. Therefore, we are simultaneously developing an image-based tool that can recognize and analyze endonasal endoscopic videos with the help of a pattern recognition software. That software was already reported in ORL procedures [2, 7]. Google® already uses similar technologies to provide video content information [2] automatically. Based on the images within the endoscopic video, this tool can generate a detailed operation report during the actual procedure.

Moreover, it is possible to include screenshots in the created report for accurate documentation of, for example, suspect findings. These highly-detailed reports can also have advantages for legal purposes. By combining both tools, the model can receive more information from the endoscopic video and the spoken keywords to generate more accurate and higher-quality surgical reports in the future. A further consideration might be to create several versions of one report with different levels of detail for different purposes. For example, highly detailed reports could be saved for legal purposes, and moderately detailed reports for briefly reconstructing a procedure. Another important aim of those applications could be the standardization of surgical reports for better interdisciplinary exchange and quality management [1, 5, 6, 12].

First studies describing the regular use of artificial intelligence in surgery [12], as well as automatic report generation in radiology [5] already exist. This proof-of-concept study showed that there is a need and acceptance of new AI-based applications for generating (ORL-) surgical reports. Therefore, it is important that physicians actively support this technological development [6, 12]. Finally, our AI tool can be transferred to other clinical departments.

Conclusion

The digital revolution in healthcare is inevitable [1–3]. AI-based technologies will become an indispensable part of the clinical workflow for reducing physicians’ workload and optimizing the quality and patient safety [1, 2, 5, 6, 12, 14]. The developed tool is an example and, therefore, part of the healthcare sector’s digital revolution. An estimated time saving of 16–30 min per day and a relevant workload reduction was seen by most of the participants. These results point out the need of optimizing the repetitive and time-consuming task of generating surgical reports. Further training of the model is needed to improve its performance, emphasizing the necessary process of advancement in AI-based technologies for clinical application. With this proof-of-concept study, a first step in the indispensable development of medical AI technologies

was made to optimize clinical workflows, possibly contributing to enhance the quality and safety of patient care in future.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11548-022-02791-0>.

Acknowledgements The authors thank all participating physicians in the study, as this study would not have been possible without their contribution.

Author contributions VK, VW, and MP designed and coordinated the study. VK and VW collected and analyzed the data. VK and RB performed the model training and optimization. VK, VW, RB, MS, VZ, TN, AD and MP interpreted the data. VK, VW and MP wrote the first version of the paper. All authors approved the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. Not applicable.

Availability of data and materials The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Artificial Intelligence In Healthcare Market Size Report (2021), 2019–2025. <https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-healthcare-market>. Accessed 28 Feb 2021
- Bur AM, Shew M, New J (2019) Artificial intelligence for the otolaryngologist. A state of the art review. *Otolaryngol-Head Neck Surg* 160(4):603–611. <https://doi.org/10.1177/0194599819827507>
- Liu X, Faes L, Kale AU et al (2019) A comparison of deep learning performance against healthcare professionals in detecting diseases from medical imaging. A systematic review and meta-analysis. *Lancet Digit Health* 1(6):e271–e297. [https://doi.org/10.1016/S2589-7500\(19\)30123-2](https://doi.org/10.1016/S2589-7500(19)30123-2)
- Monshi MMA, Poon J, Chung V (2020) Deep learning in generating radiology reports. A survey. *Artif Intell Med* 106:101878. <https://doi.org/10.1016/j.artmed.2020.101878>
- Letourneau-Guillon L, Camirand D, Guilbert F et al (2020) Artificial intelligence applications for workflow, process optimization and predictive analytics. *Neuroimaging Clin N Am* 30(4):e1–e15. <https://doi.org/10.1016/j.nic.2020.08.008>
- Eryigit Ö, van de Graaf FW, Lange JF (2019) A systematic review on the synoptic operative report versus the narrative operative report in surgery. *World J Surg* 43(9):2175–2185. <https://doi.org/10.1007/s00268-019-05017-8>
- Bieck R, Heuermann K, Pirlich M et al (2020) Language-based translation and prediction of surgical navigation steps for endoscopic wayfinding assistance in minimally invasive surgery. *Int J CARS* 15(12):2089–2100. <https://doi.org/10.1007/s11548-020-02264-2>
- Kudo T (2018) Subword regularization. Improving neural network translation models with multiple subword candidates. In: Gurevych I, Miyao Y (eds) Proceedings of the 56th annual meeting of the association for computational linguistics, vol 1: long papers. Association for computational linguistics, Stroudsburg, PA, USA, pp 66–75
- Chan W, Jaitly N, Le QV et al. (2015) Listen, attend and spell
- Lin C-Y, Och FJ (2004) Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. In: Scott D (ed) Proceedings of the 42nd annual meeting on association for computational linguistics—ACL '04. Association for computational linguistics, Morristown, NJ, USA, 605–es
- Jing B, Xie P, Xing E (2018) On the automatic generation of medical imaging reports. In: Gurevych I, Miyao Y (eds) Proceedings of the 56th annual meeting of the association for computational linguistics, vol 1: long papers. Association for computational linguistics, Stroudsburg, PA, USA, pp 2577–2586
- Hashimoto DA, Rosman G, Rus D et al (2018) Artificial intelligence in surgery promises and perils. *Ann Surg* 268(1):70–76. <https://doi.org/10.1097/SLA.0000000000002693>
- Lin SY, Shanafelt TD, Asch SM (2018) Reimagining clinical documentation with artificial intelligence. *Mayo Clin Proc* 93(5):563–565. <https://doi.org/10.1016/j.mayocp.2018.02.016>
- Quiroz JC, Laranjo L, Kocaballi AB et al (2019) Challenges of developing a digital scribe to reduce clinical documentation burden. *NPJ Digit Med* 2:114. <https://doi.org/10.1038/s41746-019-0190-1>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

3 Zusammenfassung

Dissertation zur Erlangung des akademischen Grades

Dr. med. an der Medizinischen Fakultät der Universität Leipzig

Schlagwortbasierte und KI-unterstützte Erstellung von Operationsberichten am Beispiel der FESS

eingereicht von Valentina Wildfeuer (01/2023)

angefertigt am Universitätsklinikum Leipzig AÖR,

Klinik und Poliklinik für Hals-, Nasen-, Ohrenheilkunde

Direktor: Prof. Dr. med. Andreas Dietz

betreut von

Prof. Dr. med. Andreas Dietz

Dr. med. Markus Pirlich

Richard Bieck

Die Bedeutung von Technologien, welche auf künstlicher Intelligenz basieren, nimmt stetig zu. Vor allem im Gesundheitswesen spielen solche Anwendungen eine zunehmend größere Rolle (Monshi et al. 2020; Bur et al. 2019; Liu et al. 2019), denn mithilfe dieser können Qualität- und Patientensicherheit enorm verbessert werden (Letourneau-Guillon et al. 2020; Anteby et al. 2021; Ahmad et al. 2021).

Ein bedeutendes Beispiel ist hier die Dokumentation von Operationen, welche sehr zeitaufwendig und nicht zuletzt auch fehleranfällig ist, da die Prozedur meist Stunden nach der eigentlichen Operation (OP) aus dem Gedächtnis reproduziert werden muss (Jing et al. 2018; Lorowski und Malinowska 2020; Lorkowski et al. 2021). Die Abnahme solcher zeitintensiven und repetitiven Aufgaben durch KI-basierte Anwendungssysteme ermöglicht es, dass mehr Zeit für die Arbeit mit und am Patienten geschaffen und die Arbeitslast der Ärzte reduziert werden kann

(Letourneau-Guillon et al. 2020; Hashimoto et al. 2018; Lin et al. 2018; Quiroz et al. 2019; Jing et al. 2018; Lorkowski et al. 2021).

Hierfür wurde mithilfe von „Natural Language Processing“ und einem neuronalen Sprachmodell, dem sogenannten „Deep Learning“, ein schlagwortbasiertes KI-Tool entwickelt, welches Schlagwörter, die während der Prozedur ausgesprochen werden, erkennt und verarbeitet. Anhand dieser und der Wortreihenfolge werden durch eine Encoder-Decoder-Architektur sinnvoll zusammenhängende OP-Berichtssätze erzeugt, sodass parallel zur laufenden Operation ein vollständiger OP-Bericht teilautomatisiert konstruiert wird. Das Tool wurde vorerst auf der Basis von 48 konventionell erstellten Berichten von Nasennebenhöhlenoperationen mithilfe einer „k-fold cross-validation“ trainiert und weiterentwickelt, kann aber in Zukunft in verschiedenen Fachdisziplinen und Prozeduren zur Dokumentation integriert werden.

Das Ziel der vorliegenden Arbeit ist es, die Performanzentwicklung durch das Training des KI-basierten Programms zur OP-Dokumentation von NNH-Operationen anhand von bestimmten Metriken zu objektivieren.

Hierfür wurden die erzeugten Berichte vor und nach Training des Tools anhand von vier metrischen Parametern (BLEU, ROUGE, COSS, METEOR) verglichen und die Optimierung der Performanz analysiert und ausgewertet.

Anschließend erfolgte eine Evaluation mit der Fragestellung, inwiefern solche Anwendungen die Mediziner im Klinikalltag entlasten und die Patientensicherheit optimieren können.

Hierfür wurde eine Studie mit 16 Probanden durchgeführt, für welche gleich viele Fachärzte wie Assistenzärzte der Hals-, Nasen-, Ohrenheilkunde rekrutiert wurden. Die Teilnehmenden sollten drei zuvor zufällig ausgewählte, von dem Tool erstellte OP-Berichte Korrekturlesen und diese nach eigenem Empfinden bearbeiten. Dabei wurden die Anzahl und Art der Korrekturen erfasst und ausgewertet.

Anschließend wurden die Studienteilnehmer aufgefordert, einen Evaluationsbogen auszufüllen (s. Anlage 3). In diesem sollten neben der Beantwortung des Fachbereiches und des Spezialisierungsgrades, die Qualität der Berichte, die mögliche Zeitersparnis und der künftige Mehrwert und Nutzen eines solches Tools eingeschätzt werden. In einer ersten Version des Fragebogens wurde die Antwortoption „keine Zeitersparnis“ nicht gegeben, sodass diese Frage mit detaillierteren Antwortmöglichkeiten reevaluiert wurde, um eine Ergebnisverzerrung zu vermeiden. Des Weiteren sollten die künstlich erzeugten mit den konventionell erstellten Berichten der jeweils gleichen Operation verglichen werden.

Schlagwortbasiertes Training des Sprachmodells: Insgesamt wurden 48 künstlich erstellte mit den dazugehörigen konventionellen Berichten anhand von vier objektiven Parametern (BLEU, ROUGE, COSS, METEOR) verglichen. Die Ergebnisse des Wilcoxon-Tests zeigten eine signifikante Verbesserung ($p < .001$) aller objektiver Parameter durch das schlagwortbasierte Training des KI-Sprachmodells.

Dies verdeutlicht die Lernfähigkeit KI-basierter Tools, die aus eigenen Erfolgen und Niederlagen lernen und sich selbstständig weiterentwickeln (Bur et al. 2019; Hashimoto et al. 2018; Sharma et al. 2021).

Alle KI-Tools haben gemein, dass je umfangreicher die Datenmengen sind, desto präziser können Vorhersagen getroffen werden (Lorkowski et al. 2021). Mithilfe von großen Datenmengen, beispielsweise in Form von hochwertigen, ausführlichen Operationsberichten, könnte unser Tool zukünftig noch qualitativere und detailliertere Berichte erstellen, die unseren Erwartungen gerecht werden, bevor dieses in den klinischen Alltag integriert wird.

Korrekturen: Die Anzahl und Art der Korrekturen, welche die Teilnehmer in den Dateien der KI-gestützten OP-Berichte vornahmen, wurden über *Microsoft Word®* aufgezeichnet und anschließend ausgewertet. Die Korrekturanzahl variierte zwischen den Teilnehmern stark. Dies lässt sich vermutlich auf den unterschiedlichen Ausbildungs- und Erfahrungsgrad der Teilnehmer zurückführen. Es wären jedoch Folgestudien notwendig, um die Ursache dieser Diskrepanz genauer zu untersuchen. Durchschnittlich waren 23.25 Korrekturen pro Bericht erforderlich, um einen subjektiv zufriedenstellenden Bericht zu erhalten.

Dieses Ergebnis zeigt, dass noch weiteres Training zur Optimierung des Tools notwendig ist, damit ein angemessener Bericht, der nur wenig Korrekturen im Nachhinein bedarf, konstruiert werden kann (Bur et al. 2019).

Grammatik, Inhalt und Lesefluss: Im Schnitt wurde die Grammatik der künstlich erstellten Berichte von den Teilnehmern mit einer 3.1, der Inhalt mit einer 2.8 und der Lesefluss mit einer 2.6 im Schulnotensystem bewertet. Dies zeigt eine bereits befriedigende Performanz des Tools, welche weiterhin kontinuierlich verbessert wird.

Täglicher Zeitaufwand und Zeitersparnis: 66.67% der Teilnehmer schätzten einen täglichen Zeitaufwand von 30 bis 60 Minuten für die Erstellung von OP-Berichten. In einer detaillierteren Reevaluation schätzten 61.11% der Teilnehmer eine mögliche Zeitersparnis von 16 bis 30 Minuten und 27.78% von 1 bis 15 Minuten pro Tag durch die Unterstützung eines solchen Tools.

Diese Ergebnisse zeigen die Möglichkeit einer wertvollen Zeitersparnis, die die Arbeitslast der Mediziner wesentlich reduzieren könnte und somit Freiraum für wichtige ärztliche Tätigkeiten mit und am Patienten schafft (Letourneau-Guillon et al. 2020; Hashimoto et al. 2018; Lin et al. 2018; Quiroz et al. 2019). Mithilfe der KI-unterstützten Übernahme solcher repetitiven Aufgaben kann die Patientensicherheit und Behandlungsqualität maßgeblich verbessert und folglich die Mortalitätsrate gesenkt werden (Letourneau-Guillon et al. 2020; Lorkowski und Malinowska 2020; Ahmad et al. 2021).

Klinischer Nutzen und Mehrwert: 61.11% der Teilnehmer erwarteten eine Arbeitslastreduktion und 66.67% sahen einen klinischen Mehrwert und Nutzen in der teilautomatisierten Erstellung von OP-Berichten. Die Mehrheit mit 66.67% würden dieses Tool zur OP-Berichterstellung in Zukunft selbst nutzen.

Diese Ergebnisse zeigen, dass zum einen bereits eine große Akzeptanz gegenüber einer solchen computergestützten Anwendung vorhanden ist (Eryigit et al. 2019) und zum anderen, dass Mediziner es notwendig finden von repetitiven Tätigkeiten im klinischen Alltag entlastet zu werden, um weiterhin den qualitativ hohen Standard der Patientenversorgung gewährleisten zu können (Bur et al. 2019; Letourneau-Guillon et al. 2020). Dieser wichtige Einfluss der KI sowie die Bedeutsamkeit des Verständnisses für die Tragweite dieser revolutionierenden Tools, auch in der HNO-Dokumentation, wird ebenfalls von Bur et al. beschrieben.

33.33% der Teilnehmer gaben eine inhaltliche Ähnlichkeit und 27.78% eine formale Ähnlichkeit zwischen den künstlich und konventionell erstellten Berichten an.

Diese bereits bestehende Ähnlichkeit zwischen den generierten und konventionell erstellten Berichten verdeutlicht die zufriedenstellende Performanz des Tools, welche fortlaufend verbessert wird, um noch akkuratere Berichte zu erhalten. Die konventionellen OP-Berichte wurden als Goldstandard verwendet, da kein offizieller Standard für die Form und den Inhalt eines OP-Berichtes existiert. Die Detailtiefe kann so zum Teil von Bericht zu Bericht stark variieren, weil dies den individuellen Präferenzen des Operateurs überlassen ist. In Zukunft könnte mithilfe dieses Tools möglicherweise ein Standard als Berichtvorlage für verschiedene Operationen konstruiert werden. Dies würde große Vorteile für den globalen Austausch und die Ausbildung junger Chirurgen bieten.

Dieses Projekt stellt den Grundbaustein dar, auf welchem weitere Entwicklungen und Optimierungen folgen müssen, um eine Integrierung von KI im Gesundheitswesen realisieren zu können. Es wäre auch möglich, zukünftig das Tool um andere KI-Anwendungen zu ergänzen, um beispielsweise mithilfe von Bilderkennungsprogrammen intraoperative Endoskopvideos zu analysieren und die Berichterstellung weiter zu präzisieren. Diese Studie zeigt die enorme Bedeutsamkeit dieser Technologien für die Zukunft der Medizin, welche immer sicherer, individueller und erfolgsversprechender wird.

Die Evaluationsergebnisse verdeutlichen, dass sich bereits eine große Akzeptanz der Mediziner gegenüber solchen neuen Anwendungssystemen abzeichnet. Angesichts einer mehrheitlich geschätzten Zeitersparnis von 16 bis 30 Minuten pro Tag und einer vermuteten Arbeitslastreduktion, erscheint eine Optimierung der repetitiven und zeitintensiven OP-Berichterstellung notwendig. Vor Implementierung des vorgestellten Tools zur OP-Berichterstellung ist es wichtig, dass dieses mit großen Datenmengen weiter trainiert wird, damit es sich zukünftig selbstständig fortlaufend optimieren kann, um die Leistung und Effizienz kontinuierlich zu verbessern (Bur et al. 2019; Anteby et al. 2021; Sharma et al. 2021).

Durch die Nutzung solcher KI-basierter Systeme kann zukünftig eine Steigerung der Zufriedenheit der Ärzte erreicht und somit gleichzeitig die Qualität- und die Patientensicherheit optimiert werden (Bur et al. 2019; Eryigit et al. 2019; Quiroz et al. 2019; Hashimoto et al. 2018; Letourneau-Guillon et al. 2020; Lorowski und Malinowska 2020).

4 Literaturverzeichnis

1. 3Derm Systems, Inc. (2020): 3Derm gibt zwei FDA Breakthrough Device-Bezeichnungen für autonome Hautkrebs-KI bekannt. Online verfügbar unter <https://www.prnewswire.com/news-releases/3derm-announces-two-fda-breakthrough-device-designations-for-autonomous-skin-cancer-ai-300982072.html>.
2. Ahmad, Zubair; Rahim, Shabina; Zubair, Maha; Abdul-Ghafar, Jamshid (2021): Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review (1).
3. Anteby, Roi; Horesh, Nir; Soffer, Shelly; Zager, Yaniv; Barash, Yiftach; Amiel, Imri et al. (2021): Deep learning visual analysis in laparoscopic surgery: a systematic review and diagnostic test accuracy meta-analysis (4).
4. Bieck, Richard; Heuermann, Katharina; Pirlich, Markus; Neumann, Juliane; Neumuth, Thomas (2020): Language-based translation and prediction of surgical navigation steps for endoscopic wayfinding assistance in minimally invasive surgery. In: *International journal of computer assisted radiology and surgery* 15 (12), S. 2089–2100. DOI: 10.1007/s11548-020-02264-2.
5. Bieck, Richard; Wildfeuer, Valentina; Kunz, Viktor; Sorge, Martin; Pirlich, Markus; Rockstroh, Max; Neumuth, Thomas (2021): Generation of surgical reports using keyword-augmented next sequence prediction.
6. Bur, Andrés M.; Shew, Matthew; New, Jacob (2019): Artificial Intelligence for the Otolaryngologist: A State of the Art Review. In: *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery* 160 (4), S. 603–611. DOI: 10.1177/0194599819827507.
7. Eberl; Ulrich (2016): Smarte Maschinen. Wie Künstliche Intelligenz unser Leben verändert.
8. Eryigit, Özgür; van de Graaf, Floyd W.; Lange, Johan F. (2019): A Systematic Review on the Synoptic Operative Report Versus the Narrative Operative Report in Surgery. In: *World journal of surgery* 43 (9), S. 2175–2185. DOI: 10.1007/s00268-019-05017-8.
9. Finley, Gregory; Edwards, Erik; Robinson, Amanda; Brenndoerfer, Michael; Sadoughi, Najmeh; Fone, James et al. (2018): An automated medical scribe for documenting clinical encounters. In: Yang Liu, Tim Paek und Manasi Patwardhan (Hg.): Proceedings of the 2018 Conference of the North American Chapter of. Proceedings of the 2018 Conference of the North American Chapter of. New Orleans, Louisiana. Stroudsburg, PA, USA: Association for Computational Linguistics, S. 11–15.
10. Grand View Research (2021): Artificial Intelligence In Healthcare Market Report, 2021-2028. Online verfügbar unter <https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-healthcare-market>.
11. Grand View Research (2022): Artificial Intelligence In Healthcare Market Size Report, 2030. Online verfügbar unter <https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-healthcare-market>.
12. Hashimoto, Daniel A.; Rosman, Guy; Rus, Daniela; Meireles, Ozanan R. (2018): Artificial Intelligence in Surgery: Promises and Perils. In: *Annals of surgery* 268 (1), S. 70–76. DOI: 10.1097/SLA.0000000000002693.
13. Hohensee, Matthias (2021): Jetzt wittert Elon Musk seine nächste Milliardenchance. In: *Wirtschaftswoche*. Online verfügbar unter <https://www.wiwo.de/my/unternehmen/auto/tesla-roboter-jetzt-wittert-elon-musk-seine-naechste-milliardenchance/27533898.html>.
14. Jing, Baoyu; Xie, Pengtao; Xing, Eric (2018): On the Automatic Generation of Medical Imaging Reports, S. 2577–2586. DOI: 10.18653/v1/P18-1240.

15. Letourneau-Guillon, Laurent; Camirand, David; Guilbert, Francois; Forghani, Reza (2020): Artificial Intelligence Applications for Workflow, Process Optimization and Predictive Analytics. In: *Neuroimaging clinics of North America* 30 (4), e1-e15. DOI: 10.1016/j.nic.2020.08.008.
16. Leunig, Andreas (2010): Wann Und Wie Sollte Operiert Werden.
17. Lin, Chin-Yew; Och, Franz Josef (2004): Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. In: Donia Scott (Hg.): Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics - ACL '04. the 42nd Annual Meeting. Barcelona, Spain, 7/21/2004 - 7/26/2004. Morristown, NJ, USA: Association for Computational Linguistics, 605-es.
18. Lin, Steven Y.; Shanafelt, Tait D.; Asch, Steven M. (2018): Reimagining Clinical Documentation With Artificial Intelligence. In: *Mayo Clinic proceedings* 93 (5), S. 563-565. DOI: 10.1016/j.mayocp.2018.02.016.
19. Liu, Xiaoxuan; Faes, Livia; Kale, Aditya U.; Wagner, Siegfried K.; Fu, Dun Jack; Bruynseels, Alice et al. (2019): A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. In: *The Lancet Digital Health* 1 (6), e271-e297.
20. Lorkowski, Jacek; Grzegorowska, Oliwia; Pokorski, Mieczysław (2021): Artificial Intelligence in the Healthcare System: An Overview.
21. Lorowski, Jacek; Malinowska, Maria (2020): The future of medical documentation - review of selected literature.
22. McCarthy, John (2007): What is artificial intelligence.
23. Monshi, Maram Mahmoud A.; Poon, Josiah; Chung, Vera (2020): Deep learning in generating radiology reports: A survey. In: *Artificial Intelligence in Medicine* 106, S. 101878.
24. Quiroz, Juan C.; Laranjo, Liliana; Kocaballi, Ahmet Baki; Berkovsky, Shlomo; Rezazadegan, Dana; Coiera, Enrico (2019): Challenges of developing a digital scribe to reduce clinical documentation burden. In: *NPJ digital medicine* 2, S. 114. DOI: 10.1038/s41746-019-0190-1.
25. Sharma, Neha; Sharma, Reecha; Jindal, Neeru (2021): Machine Learning and Deep Learning Applications-A Vision (1).
26. Simmen, Daniel; Jones, Nick (2005): Chirurgie der Nasennebenhöhlen und der vorderen Schädelbasis. Kartonierte Sonderausg. Stuttgart: Thieme.
27. Topol, Eric J. (2019): High-performance medicine: the convergence of human and artificial intelligence (1).
28. Whooley, Sean (2020): 3Derm lands FDA breakthrough device designations for autonomous skin cancer AI - MassDevice. In: *MASSDEVICE MEDICAL NETWORK*.
29. Xiaoxuan Liu MBChB; Livia Faes MD; Aditya U Kale MBChB; Siegfried K Wagner MBChB; Dun Jack Fu PhD; Alice Bruynseels MBChB et al.: A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis.
30. Xu, Feiyu (2021): Die Zukunft gehört der Künstlichen Intelligenz. In: *Handelsblatt*. Online verfügbar unter <https://www.handelsblatt.com/meinung/gastbeitraege/gastkommentar-die-zukunft-gehört-der-kuenstlichen-intelligenz/27189264.html>. Accessed 09 Aug 2021

5 Anlagen

1 Aufgabenanleitung für die Studienteilnehmer

Aufgabe 1:

In Aufgabe 1 sollen die drei beigefügten, künstlich erzeugten OP-Berichte von Ihnen überarbeitet werden bis Sie der Meinung sind, dass es sich um einen angemessenen OP- Bericht handelt. Ihre Bearbeitungszeit und die Anzahl Ihrer Korrekturen werden dabei automatisch erfasst. Es ist wichtig, dass Sie für die Bearbeitung von Aufgabe 1 das **Word- Dokument** zu den jeweiligen OP-Berichten mit dem Titel **Aufgabe1_OPBericht1-3** öffnen, Ihre Bearbeitung ohne Verzögerungen vornehmen, die Datei abspeichern und **umgehend wieder schließen**, da es ansonsten zu einer Verfälschung der aufgezeichneten Bearbeitungszeit kommt. **Bitte öffnen Sie das bearbeitete Word-Dokument anschließend nicht erneut.**

Aufgabe 2:

Für Aufgabe 2 bitten wir Sie das jeweilige **PDF-Dokument** zu den OP-Berichten mit dem Titel **Aufgabe2_Evaluation_OPBericht1-3** zu öffnen. Anschließend nehmen Sie über den Link: <https://forms.gle/DjCsaVDjXum7cRKb9> die Evaluation der künstlich erzeugten OP- Berichte, verglichen mit den konventionell erzeugten OP-Berichten, vor.

Bitte senden Sie uns die von Ihnen bearbeiteten und abgespeicherten Word-Dokumente **bis zum 30.06.2021** an die E-Mail-Adresse: valentina.wildfeuer@medizin.uni-leipzig.de zurück.

Wir danken Ihnen recht herzlich für Ihre Teilnahme und verbleiben

mit kollegialen Grüßen,

Dr. med. Markus Pirlich

Oberarzt

Klinik und Poliklinik für Hals-, Nasen- und Ohrenheilkunde/ Plastische Operationen
Universitätsklinikum Leipzig – AÖR

Dr. med. Viktor Kunz

Arzt in Weiterbildung

Klinik und Poliklinik für Hals-, Nasen- und Ohrenheilkunde/ Plastische Operationen
Universitätsklinikum Leipzig – AÖR

Cand. med. Valentina Wildfeuer

Doktorandin

Klinik und Poliklinik für Hals-, Nasen- und Ohrenheilkunde/ Plastische Operationen
Universitätsklinikum Leipzig – AÖR

2 Ausgewählte OP-Berichte für die Evaluation

OP-Bericht 1

Beispiel eines automatisch erzeugten OP-Berichts anhand einer Schlagwortliste.

Künstlich erzeugter OP-Bericht

OP in Intubationsnarkose, Team-time-out, korrekter Patient/in, Aufklärung vorliegend.

Hohe Einlage (Naphazolin/Tetracain) bds., folgend steriles Abwaschen und Abdecken.

Steriles Abwaschen und Abdecken.

Medialisieren der mittleren Muschel mit dem Freer (stumpf).

Schlitzung am Vorderrand des Proc uncinatus an dessen Vorderrand bis auf Knochenniveau.

Nun eingehen mit dem geraden Blakesley.

Resektion von Processus uncinatus.

Zudem wird die Bulla ethmoidalis eröffnet und das vordere Siebbein ausgeräumt, ohne dass hier Polypen oder chronisch entzündliche Schleimhaut sich darstellt.

Erweiterung des Kieferhöhlenostiums mit der Antrumstanze im Sinne einer Sinusotomie Typ I frei präpariert und erweitert, unter Verwendung der rückwärtsgreifende Stanze und der Antrumstanze sowie der 45° Optik.

Insgesamt imponiert der Situs sehr entzündet und Entfernung einer großen bodenständigen und Entfernung einer großen bodenständigen Zyste, sonst reizlos.

Präparation etagenweise nach oben bis zum Erreichen der Schädelbasis unter videoendoskopischer Sicht.

Anschließend Präparation des vorderen Siebbeins mit Ausräumung von polypösen Massen.

Nun noch Darstellung der Keilbeinhöhlenvorderwand paraseptal links, genau wie auf der Gegenseite.

Weiterhin Ausschluss von polypösem Restgewebe im mittleren und unteren Nasengang.

Anschließend endoskopische Kontrolle. Die Schädelbasis und die Lamina papyracea sind intakt.

Dieser wird problemlos sondiert und zeigt sich frei.

Gleiches Vorgehen und Verhältnisse auf der rechten Seite.

Hierbei entleert sich ebenfalls eine polypöse Verlegung des mittleren Nasenganges/Infundibulum.

Beginn links. Medialisieren der mittleren Muschel mit dem Freer.

Nun Eingehen mit dem geraden Blakesley.

Resektion von Proc. uncinatus und reichlich polypösem Gewebe im mittleren Nasengang.

Jetzt kann das Septum submuköse.

Die Siebbeinendoskopie zeigt eine verdickte Schleimhaut im vorderen Siebbein.

Die Keilbeinhöhlenvorderwand wird dargestellt und anschließend das Ostium mit dem geraden Sauger, die keilbeinhöhle stellt sich frei dar..

Anschließend endoskopische Kontrolle. Die Schädelbasis und die Lamina papyracea sind intakt.

Anschließend endoskopische Kontrolle. Die Schädelbasis und die Lamina papyracea sind intakt.

Die Keilbeinhöhle ist frei.

End_of_Document.

Bei OP-Ende keine Blutung.

Originalbericht

OP in ITN, hohe Einlage (Naphazolin/Tetracain) bds., folgend steriles Abwaschen und Abdecken.

Infiltrationsanästhesie mit Xylocitin/Adrenalin unter endoskopischer Sicht im Bereich des Proc uncinatus beidseits.

Beginn auf der linken Seite.

Medialisieren der linken mittleren Muschel mit dem Freer (stumpf).

Schlitzung am Vorderrand des Proc. uncinatus an dessen Vorderrand bis auf Knochenniveau (spitze Freerseite).

Nun Eingehen mit dem geraden Blakesley.

Resektion von Proc uncinatus.

Die Bulla ethmoidalis wird dargestellt.

Mit dem Kieferhöhlensauger wird das Kieferhöhlenostium im Sinne einer Sinusotomie Typ I freipräpariert und erweitert, unter Verwendung der rückwärtsgreifenden Stanze und der Antrumstanze sowie der 45° Optik.

Zwischendurch immer wieder Fotodokumentation.

Weitere vorsichtige Präparation etagenweise nach oben bis zum Erreichen der Schädelbasis unter videoendoskopischer Sicht.

Präparation bis zum Übergang in das hintere Siebbein unter kontinuierlicher Exstirpation von polypösem Gewebe.

Darstellen der Keilbeinhöhlevorderwand mit anschließender Eröffnung.

Resektionskontrolle und Ausschluss von Verletzung und polypösem Restgewebe: Keilbeinhöhlevorderwand, Schädelbasis, Lamina papyracea.

Weiterhin Ausschluss von polypösem Restgewebe im mittleren und unteren Nasengang.

Aufsuchen des Rec frontalis, der funktionell frei dargestellt wird..

Zuwenden zur rechten Seite und Vorgehen in gleicher Weise.

Medialisieren der mittleren Muschel rechts mit dem Freer.

Schlitzung am Vorderrand des Proc. uncinatus an dessen Vorderrand bis auf Knochenniveau.

Nun Eingehen mit dem geraden Blakesley.

Resektion von Proc uncinatus und ebenfalls reichlich polypösem Gewebe im mittleren Nasengang.

Weitere vorsichtige Präparation etagenweise nach oben bis zum Erreichen der Schädelbasis unter videoendoskopischer Sicht.

Präparation bis zum Übergang in das hintere Siebbein unter kontinuierlicher Exstirpation von polypösem Gewebe.

Darstellen der Keilbeinhöhlevorderwand mit anschließender Eröffnung.

Resektionskontrolle und Ausschluss von Verletzung und polypösem Restgewebe: Keilbeinhöhlevorderwand, Schädelbasis, Lamina papyracea.

Weiterhin Ausschluss von polypösem Restgewebe im mittleren und unteren Nasengang.

Aufsuchen des Rec frontalis, der funktionell frei dargestellt wird.

Folgend Einlage von je einer Fingerlingstamponade im mittleren Nasengang beidseits.

End_of_Document.

OP-Bericht 2

Beispiel eines automatisch erzeugten OP-Berichts anhand einer Schlagwortliste.

Künstlich erzeugter OP-Bericht

OP in Intubationsnarkose, Team-time-out, korrekter Patient/in, Aufklärung vorliegend.

Hohe Einlage (Naphazolin/Tetracain) bds., folgend steriles Abwaschen und Abdecken.

Steriles Abwaschen und Abdecken.

Beginn links. Mit dem Blakesly Abtragung von etwas polypösem Gewebe aus dem Siebeinschacht und Entfernung eines Restes des Proc. uncinatus.

Die Uncinektomie in gleicher Weise.

Resektion reichlich polypösem Gewebe im mittleren Nasengang wird schrittweise Durchführung einer hinteren Ethmoidektomie mit entsprechender Darstellung der Schädelbasis sowie der Lamina papyracea.

Darstellen der Keilbeinhöhlenvorderwand, die linksseitig sehr klein ist - hier keine weiteren Maßnahmen.

Darstellung der Kieferhöhlenöffnung und Erweiterung des Kieferhöhlenostiums mittels Kieferhöhlensauger, rückwärtsgreifender Stanze und Antrumstanze unter Resektion von reichlich Polypen aus der Kieferhöhle mit Kieferhöhlensauger und Heuwieser.

Aufsuchen des Rec frontalis, der funktionell frei dargestellt wird.

Es entleert sich massic Pus_Abstrich ad MIBI.

Nun Resektionskontrolle, Ausschluss von Verletzung und polypösem Restgewebe. Keilbeinhöhle, Schädelbasis, Lamina papyracea..

Anschließend Zuwenden zur rechten Seite und Vorgehen in gleicher Weise..

Hier stellt sich die Keilbeinhöhle soweit unangetastet dar.

Nach Eröffnung Entfernung von Polypengewebe.

Eröffnung der Bulla ethmoidalis und des Processus uncinatus und reichlich polypösem Gewebe im mittleren Nasengang.

Die Kieferhöhle weist lediglich eine Schleimhauthypertrophie auf, ausreichende und freie Öffnung.

Hier keine weiteren Maßnahmen.

Abschließend NaCl-Spülung beidseits und Einbringen von Predni-Ophthalgel beidseits.

Jeweils eine Fingerlingstamponade beidseits unter Sicht in den entstandenen Siebbeinschacht.

End_of_Document.

Originalbericht

Team timeout. Richtiger Patient.

Hohe Einlage bds., steriles Abwaschen und Abdecken.

Beginn links: Mit dem Blakesly Abtragung von etwas polypösem Gewebe aus dem Siebeinschacht und Entfernung des Proc. uncinatus.

Weitere Extirpation von Polypengewebe bis zum Erreichen der Schädelbasis unter videoendoskopischer Sicht.

Präparation des hinteren Siebbeins unter kontinuierlicher Exstirpation von wandständigem polypösem Gewebe.

Darstellen der Keilbeinhöhlenvorderwand, die linksseitig sehr klein ist - hier keine weiteren Maßnahmen.

Darstellung der Kieferhöhlenöffnung und Erweiterung des Kieferhöhlenostiums mittels Kieferhöhlensauger, rückwärtsgreifender Stanze und Antrumstanze unter Resektion von reichlich Polypen aus der Kieferhöhle mit Kieferhöhlensauger und Heuwieser.

Aufsuchen des Rec. frontalis.

Dieser wird nicht forciert, da CT-morphologisch beide Stirnhöhlen aplastisch sind.

Nun Resektionskontrolle, Ausschluss von Verletzung und polypösem Restgewebe: Keilbeinhöhle, Schädelbasis, Lamina papyracea..

Zuwenden zur rechten Seite und Vorgehen in gleicher Weise.

Hier stellt sich die Keilbeinhöhle soweit unangetastet dar.

Nach Eröffnung Entfernung von Polypengewebe.

Außerdem Ausräumung von Polypengewebe aus dem Siebbein und Stirnhöhleneingang sowie dem dorsalen mittleren Nasengang.

Die Kieferhöhle weist lediglich eine Schleimhauthypertrophie auf, ausreichende und freie Öffnung.

Hier keine weiteren Maßnahmen.

Abschließend NaCl-Spülung beidseits und Einbringen von Predni-Ophthalgel beidseits.

Je eine Fingerlingstamponade beidseits unter Sicht in den entstandenen Siebbeinschacht.

End_of_Document.

OP-Bericht 3

Beispiel eines automatisch erzeugten OP-Berichts anhand einer Schlagwortliste.

Künstlich erzeugter OP-Bericht

Rückenlagerung.

Intubationsnarkose.

Anbringen des Stirnbandes für die Navigation.

Medialisieren der linken mittleren Muschel bei Zustand nach bereits stattgehabter Siebbeinausräumung, jetzt Vernarbungen, die sukzessive gelöst werden.

Schlitzung am Vorderrand des Processus uncinatus an dessen Vorderrand bis auf Knochenniveau.

Nun eingehen mit dem geraden Blakesley.

Weitere vorsichtige Präparation etagenweise nach oben bis zum Erreichen der Schädelbasis unter videoendoskopischer Sicht.

Präparation bis zum Übergang in das hintere Siebbein unter kontinuierlicher Exstirpation von polypösem Gewebe.

Resektionskontrolle und Ausschluss von Verletzung und polypösem Restgewebe. Keilbeinhöhlenvorderwand, Schädelbasis, Lamina papyracea.

Erweiterung im Sinne einer Sinusotomie mit Hilfe der rückwärtsgreifenden Knochenstanze und der Antrumstanze erweitert.

Dieses wird nach posterior mit dem scharfen Blakesley nach anterior mit der Rückwärtsstanze erweitert.

Fotodokumentation.

Einbringen von Prednisolon-Ophtalgel-Salbe in den Siebbeinschacht beidseits und Fingerlingstamponade jeweils eine rechts und links.

End_of_Document.

Originalbericht

OP in Intubationsnarkose, Team-time-out, korrekter Patient/in, Aufklärung vorliegend.

Hohe Einlage (Naphazolin/Tetracain) bds., folgend steriles Abwaschen und Abdecken.

Infiltrationsanästhesie mit Xylocitin/Adrenalin unter endoskopischer Sicht in das polypöse Gewebe und den Processus uncinatus links.

Medialisieren der mittleren Muschel mit dem Freer.

Schlitzung am Vorderrand des Proc uncinatus an dessen Vorderrand bis auf Knochenniveau.

Nun Eingehen mit dem geraden Blakesley und Resektion des Proc. uncinatus.

Weiter vorsichtige Präparation etagenweise nach oben bis zum Erreichen der Schädelbasis unter videoendoskopischer Sicht.

Präparation bis zum Übergang in das hintere Siebbein unter kontinuierlicher Exstirpation von leicht polypösem Gewebe (insgesamt jedoch geringe Pathologie).

Weiterhin Ausschluss von polypösem Restgewebe im mittleren und unteren Nasengang.

Anschließend Aufsuchen des Kieferhöhlenostium mit dem gebogenen Sauger, welches anschließend mit dem gebogenen Blackesley, der Antrum- sowie rückwärtsgreifenden Stanze erweitert wird.

Die Kieferhöhle links imponiert weitgehend frei.

Am Ende nochmals Kontrolle des Situs, und mehrfache Fotodokumentation.

Einlage von insgesamt 2 Fingerlingstamponaden.

End_of_Document.

3 Evaluationsbogen

Schlagwortbasierte OP-Berichterstellung mit Hilfe künstlicher Intelligenz

Vergleichende Studie zur Evaluation von schlagwortbasierten, durch ein Computerprogramm erstellten Berichten von Nasennebenhöhlen-Operationen

* Erforderlich

1. Liebe Teilnehmer*innen, bei der Umfrage zum Thema "Schlagwortbasierte OP- Berichterstellung mit Hilfe künstlicher Intelligenz" handelt es sich um eine Befragung, in der Valentina Wildfeuer verantwortlich für den Inhalt sowie die Auswertung der Umfrage ist. Fragen zur Umfrage richten Sie gerne an: valentina.wildfeuer@medizin.uni-leipzig.de. Der Datensatz wird anonym ausgewertet und im Rahmen der gesetzlichen Aufbewahrungsfrist digital und für unbefugte unzugänglich gespeichert. Ich versichere, dass die von Ihnen erhobenen und nachfolgend benannten personenbezogenen Daten zu jeder Zeit ausreichend vor unberechtigtem Zugriff gesichert sind. Für die Teilnahme an dieser Umfrage brauchen sie sich weder registrieren, noch an anderer Stelle Ihren Namen angeben. Nehmen Sie sich nach dem Ausfüllen des Fragebogens noch einmal die Zeit, Ihre Angaben zu überprüfen und ggf. zu ändern, bevor Sie den Fragebogen absenden. Bitte beachten Sie, dass Ihnen jederzeit ein Auskunfts- und Berichtigungsrecht in Bezug auf die zu ihrer Person gespeicherten Daten zusteht. Außerdem gibt Ihnen das Datenschutzrecht die Möglichkeit, Ihre Einwilligung zur Teilnahme an dieser Befragung jederzeit zu widerrufen und die Löschung Ihrer personenbezogenen Daten zu verlangen. Nach erfolgtem Widerruf werden Ihre Daten umgehend gelöscht. Sofern Sie einen Widerruf in Erwägung ziehen, merken Sie sich bitte Datum und Uhrzeit der Fragebogenausfüllung, damit wir sicherstellen können, den richtigen Fragebogen aus dem Datensatz zu entfernen. Dasselbe gilt für die Daten (Bearbeitungszeit und Korrekturen) der Word-Dateien mit den von Ihnen bearbeiteten OP- Berichten, welche Sie per Mail an valentina.wildfeuer@medizin.uni-leipzig.de zurückschicken. *

Markieren Sie nur ein Oval.

- Ja, ich bin einverstanden.

Daten des Evaluierenden

2. In welchem Fachbereich arbeiten Sie? *

Markieren Sie nur ein Oval.

- Hals-Nasen-Ohrenheilkunde
- Sonstiges: _____

3. Was trifft auf Sie zu? *

Markieren Sie nur ein Oval.

- Assistenzärztin/-arzt
- Oberärztin/-arzt
- Chefärztin/-arzt
- Sonstiges: _____

Qualität der Berichte

Bitte beziehen Sie sich auf den Gesamteindruck der OP-Berichte nach Ihrer Korrektur.

4. Grammatik (Rechtschreibung, Zeichensetzung, Satzbau) in Bezug auf einen OP-Bericht *
Markieren Sie nur ein Oval.

- sehr gut
- gut
- befriedigend
- ausreichend
- mangelhaft

5. Inhalt (keine überflüssigen Details, enthält alle wichtigen Informationen, ist prägnant und umfasst das Wesentliche) *

Markieren Sie nur ein Oval.

- sehr gut
- gut
- befriedigend
- ausreichend
- mangelhaft

6. Lesefluss (lässt sich flüssig lesen, keine langen Schachtelsätze, leicht verständlich) *

Markieren Sie nur ein Oval.

- sehr gut
- gut
- befriedigend
- ausreichend
- mangelhaft

Zeitaufwand

7. Wie viel Zeit (min) wenden Sie ungefähr pro Tag für die Erstellung von OP- Berichten auf? *

Markieren Sie nur ein Oval.

- 0-30min /Tag
- 30-60min /Tag
- 60-120min /Tag
- >120min /Tag

8. Reevaluation: Zeitersparnis (Wie viel Zeit (min) würden Sie im Schnitt mit dieser Art von schlagwortbasierter, intraoperativer Berichterstellung pro Tag sparen?) *

Markieren Sie nur ein Oval.

- Keine Zeitersparnis (0 Minuten pro Tag)
- 1-15 Minuten pro Tag
- 16-30 Minuten pro Tag
- 31-45 Minuten pro Tag
- 46-60 Minuten pro Tag
- 61-90 Minuten pro Tag

- 91-120 Minuten pro Tag
- mehr als 120 Minuten pro Tag

Mehrwert und Nutzen

Bitte beziehen Sie sich auf den Gesamteindruck der OP-Berichte nach Ihrer Korrektur.

9. Arbeitslastreduktion (Workload-Entlastung, Stressreduktion, Zeitersparnis – Sehen Sie in dem Tool eine Reduktion der Arbeitslast?) *

Markieren Sie nur ein Oval.

- Ich stimme voll und ganz zu
- Ich stimme eher zu
- Ich stimme teilweise zu
- Ich stimme eher nicht zu
- Ich stimme überhaupt nicht zu

10. Mehrwert und Nutzen (Sehen Sie einen Nutzen in dieser neuen Art der Berichterstellung?) *

Markieren Sie nur ein Oval.

- Ich stimme voll und ganz zu
- Ich stimme eher zu
- Ich stimme teilweise zu
- Ich stimme eher nicht zu
- Ich stimme überhaupt nicht zu

11. Würden Sie in Zukunft dieses Tool für die OP-Berichterstellung nutzen? *

- *Markieren Sie nur ein Oval.*
- Ich stimme voll und ganz zu
- Ich stimme eher zu
- Ich stimme teilweise zu
- Ich stimme eher nicht zu
- Ich stimme überhaupt nicht zu

Vergleich zu den konventionellen OP-Berichten

Bitte beziehen Sie sich auf den Gesamteindruck aller drei OP-Berichte nach Ihrer Korrektur.

12. Inhalt (Die Originalberichte und die erzeugten Berichte ähneln sich stark in der inhaltlichen Qualität) *

Markieren Sie nur ein Oval.

- Ich stimme voll und ganz zu
- Ich stimme eher zu
- Ich stimme teilweise zu
- Ich stimme eher nicht zu
- Ich stimme überhaupt nicht zu

13. Form/ Grammatik (Die Originalberichte und die erzeugten Berichte ähneln sich stark in der formalen Qualität) *

Markieren Sie nur ein Oval.

- Ich stimme voll und ganz zu
- Ich stimme eher zu
- Ich stimme teilweise zu
- Ich stimme eher nicht zu
- Ich stimme überhaupt nicht