

Aus dem Institut für Zahn-, Mund- und Kieferheilkunde
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Eine Machine Learning Analyse von patienten-, zahn- und be-
handlungsspezifischen Faktoren in Bezug auf den Therapieer-
folg einer Wurzelkanalbehandlung

A machine learning study analyzing the association between
patient-, tooth- and treatment-level factors on the outcome of
root canal therapy

zur Erlangung des akademischen Grades
Doctor medicinae dentariae (Dr. med. dent.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Chantal Sophie Herbst

Datum der Promotion: 25.06.2023

Inhaltsverzeichnis

Tabellenverzeichnis	iii
Abbildungsverzeichnis	iv
Abkürzungsverzeichnis	v
Zusammenfassung	1
1. Einleitung	4
1.1 Das Prinzip der künstlichen Intelligenz und des maschinellen Lernens.....	4
1.2 Anwendung der KI im Gesundheitswesen und in der Zahnmedizin.....	6
1.3 Die Prognose einer Wurzelkanalbehandlung.....	7
1.3.1 <i>Patientenbezogene Faktoren</i>	7
1.3.2 <i>Zahnbezogene Faktoren</i>	9
1.3.3 <i>Behandlungsspezifische Faktoren</i>	10
1.4 Zielstellung	10
2. Methodik	11
2.1 Datenakquise und Studiendesign	11
2.2 Auswertungskriterien und Einordnung der Variablen.....	13
2.3 Bewertung des Risikos für Bias	15
2.4 Statistische Analyse.....	15
3. Ergebnisse.....	17
3.1 Studienteilnehmer*innen und deskriptive Daten	17
3.2 Einfluss der Faktoren auf Misserfolge einer Wurzelkanalbehandlung	21
3.3 Vorhersagemodelle und Faktorgewichtungsranking	24
4. Diskussion	28
4.1 Interpretation der Ergebnisse	28
4.2 Evaluation der ML Modelle und Bewertung von Vorhersagemöglichkeiten	30
4.3 Stärken und Schwächen der Studie.....	31
4.4 Ausblick für die Praxis und zukünftige Forschung	32

5. Schlussfolgerungen	34
Literaturverzeichnis.....	35
Eidesstattliche Versicherung.....	39
Anteilsklärung an den erfolgten Publikationen	40
Auszug aus der Journal Summary List	41
Druckexemplar der ausgewählten Publikation	42
Lebenslauf	50
Komplette Publikationsliste	51
Danksagung.....	52

Tabellenverzeichnis

Tabelle 1 Zusammenfassung der Auswertungskriterien.....	14
Tabelle 2 Zusammenhänge zwischen den untersuchten Faktoren und dem Zahntyp..	20
Tabelle 3 Ergebnisse der logistischen Regression (logR)	22
Tabelle 4 Leistung der verschiedenen ML Modelle auf dem Trainingsdatensatz	24
Tabelle 5 Leistung der verschiedenen ML Modelle auf dem Testdatensatz.....	25
Tabelle 6 Faktorgewichtungsranking	27

Abbildungsverzeichnis

Abbildung 1 Übersicht Pfade der künstlichen Intelligenz.....	6
Abbildung 2 Veranschaulichung der röntgenologischen Auswertung.....	13
Abbildung 3 Patientenselektionsprozess.....	17
Abbildung 4 Heatmap Darstellung von Risikofaktoren einer Wurzelkanalbehandlung	23
Abbildung 5 ROC-Kurven der Modelle	25

Abkürzungsverzeichnis

ABL	Alveolar bone loss
AUC	Area under the curve
bspw.	beispielsweise
CI	Confidence interval
DL	Deep Learning
DM	Diabetes mellitus
eD	eigene Darstellung
GBM	Gradient Boosting Machine
HbA1c	Hämoglobin A1c
IL-1 β	Interleukin 1 beta
KI	Künstliche Intelligenz
logR	Logistische Regression
ML	Maschinelles Lernen
mod.	modifiziert
NIR	No-Information-Rate
NN	Neuronale Netzwerke
OPG	Osteoprotegerin
OR	Odds-Ratio
PAI	Periapikaler Index
RANKL	Receptor Activator of NF- κ B Ligand
RCT	Root canal treatment
RF	Random Forest
ROC	Receiver Operating Characteristic
ROCAUC	Receiver Operating Characteristic Area under the curve
S.	Seite
SD	Standardabweichung
TNF- α	Tumornekrosefaktor alpha
u.a.	unter anderem
WF	Wurzelfüllung
WKB	Wurzelkanalbehandlung
WSR	Wurzelspitzenresektion
XGB	Extreme Gradient Boosting
z.B.	zum Beispiel

Zusammenfassung

Deutsche Version:

Zielsetzung: Das Erkennen und Bewerten von Risikofaktoren einer Wurzelkanalbehandlung (WKB) stellt einen entscheidenden Schritt bei der Therapieplanung zahnärztlicher Behandler*innen dar. Das maschinelle Lernen (ML) fand innerhalb der letzten Dekade vermehrt auch Anwendung in der Zahnmedizin. Das Ziel unserer Studie war es zum einen, Zusammenhänge zwischen zahn-, patienten- und behandlungsspezifischen Faktoren und der Prognose einer WKB zu detektieren und folgend ein Faktorgewichtungsranking zu erstellen; zum anderen sollte analysiert werden, inwiefern durch die Anwendung komplexer ML Modelle Vorhersagen über den Erfolg oder Misserfolg einer WKB möglich werden.

Methodik: Es wurden Patientenfälle untersucht, die im Zeitraum von 2016 bis 2020 eine WKB am CharitéCentrum 03 für Zahn-, Mund- und Kieferheilkunde erhalten hatten und mindestens sechs Monate nachuntersucht worden sind. Bei der Analyse wurden sowohl Patientendaten als auch Röntgenbilder bewertet und nach zuvor festgelegten Kriterien ausgewertet. Ein Misserfolg der Behandlung war definiert durch das Bestehen klinischer Symptome und/oder radiologischer Auffälligkeiten. Mithilfe einer logistischen Regression (logR) wurden Zusammenhänge zwischen den einzelnen Faktoren detektiert. Neben der logR sollten komplexere ML Modelle (Random Forest (RF), Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGB)) eingesetzt werden, um Vorhersagen über den Erfolg oder Misserfolg einer WKB in einem Testdatensatz zu treffen. Zur Bewertung der Vorhersagegüte wurden ROC-Kurven (Receiver Operating Characteristic) und AUC-Werte (Area under the curve) eingesetzt. Anschließend wurde auf Basis der relativen Faktorgewichtung der einzelnen Modelle eine modellübergreifende Rangfolge der Faktoren erstellt, um deren Wichtigkeit für die Vorhersage zu beschreiben.

Ergebnisse: Insgesamt wurden 591 Zähne von 458 Patient*innen (weiblich n = 216 (47,2%), männlich n = 242 (52,8%)) analysiert. Die Gesamterfolgsrate aller Behandlungen

gen betrug 79,5%. LogR zeigte, dass vor allem zahnbezogene Faktoren einen signifikanten Einfluss auf den Ausgang einer WKB hatten. Die wichtigsten Variablen waren hierbei ein schwerer alveolärer Knochenabbau von 66-100% (OR 6,48; 95% CI [2,86; 14,89], $p < 0,001$) und ein erhöhter Periapikaler Index von ≥ 4 (OR 4,59 [2,44; 8,79], $p < 0,001$). Misserfolge waren auch für Revisionstherapien signifikant häufiger (OR 1,77 [1,01; 2,86], $p < 0,01$). Bei den patientenbezogenen Faktoren war lediglich das Rauchen mit einem Misserfolg einer WKB assoziiert (OR 2,05 [1,18; 3,53], $p < 0,05$). Die Vorhersagegüte der verschiedenen ML Modelle blieb insgesamt stark begrenzt (ROCAUC: logR 0,63 [0,53; 0,73]; GBM 0,59 [0,50; 0,68]; RF 0,59 [0,50; 0,68]; XGB 0,60 [0,50; 0,70]).

Schlussfolgerungen: Misserfolge einer WKB waren primär mit zahnbezogenen Faktoren assoziiert. Vorhersagen über den Ausgang einer Behandlung waren auch mit komplexeren ML Modellen nur eingeschränkt möglich.

English Abstract:

Objective: Identifying potential risk factors of a root canal treatment (RCT) is a crucial step in endodontic treatment planning. Machine learning (ML) was found beneficial for health care applications in recent years; it has also been applied in dentistry. We intended to detect tooth-, patient- and treatment-level covariates associated with the outcome of endodontic therapy, and rank them according to their importance. Additionally, we aimed to apply ML for predicting the outcome of RCT.

Methods: We analyzed patients who received one or more RCT with at least six months follow-up at the Charité Dental Clinic between 2016 and 2020. To derive covariates, patient data including medical history and treatment protocols as well as periapical radiographs were employed. Failure was defined as persistent clinical symptoms and/or radiographical signs of persisting or progressing apical periodontitis. By using logistic regression (logR) on the full data set we analyzed associations between covariates and outcomes. LogR and more complex ML models (Random Forest (RF), Gradient Boosting

Machine (GBM), Extreme Gradient Boosting (XGB)) were then trained and their performance for predicting success or failure of root canal therapy assessed on a separate test data. ROC (Receiver Operating Characteristic) curves and AUC (Area under the curve) values were employed to evaluate the predictive performance. Mean rank values were calculated to construct a ranking showing the relative importance of each factor.

Results: A total of 591 teeth from 458 patients (female n = 216 (47.2%), male n = 242 (52.8%)) were examined. The overall success rate of root canal treatments was 79.5%. LogR showed that tooth-related covariates were significantly associated with the outcome of root canal therapy, with severe alveolar bone loss (ABL 66-100%) (OR 6.48, 95% CI [2.86, 14.89], $p < 0.001$) and a PAI-Score ≥ 4 (OR 4.59, 95% CI [2.44, 8.79], $p < 0.001$) increasing the risk of failure. Retreatments showed similarly increased risks (OR 1.77, 95% CI [1.01, 2.86], $p < 0.01$) and smoking was significantly associated with failure on patient-level (OR 2.05, 95% CI [1.18, 3.53], $p < 0.05$). The predictive performance of all ML models was limited (ROCAUC: logR 0.63 [0.53, 0.73]; GBM 0.59 [0.50, 0.68]; RF 0.59 [0.50, 0.68]; XGB 0.60 [0.50, 0.70]).

Conclusions: Failure of root canal therapy was primarily associated with tooth-related factors. In general, predicting the outcome of RCT with ML models was only limitedly possible.

1. Einleitung

1.1 Das Prinzip der künstlichen Intelligenz und des maschinellen Lernens

Künstliche Intelligenz (KI, engl. Artificial Intelligence) ist ein Zweig der angewandten Informatik und wurde als Begrifflichkeit erstmalig 1956 von John McCarthy beschrieben [1, 2]. Die Überlegung, ob Maschinen menschliches Denken nachahmen können, wurde jedoch bereits 1947 von Alan Turing formuliert [3]. Als Teilbereich von KI (Abb. 1) verwendet das maschinelle Lernen (ML, engl. Machine Learning) Algorithmen zur Analyse von Daten und lernt anhand von spezifischen Mustern in Datensätzen. Die humane Intelligenz gilt als Inspiration für KI [4], indem Algorithmen menschliche Entscheidungsstrukturen nachbilden und Probleme eigenständig lösen. Die zugrundeliegenden mathematischen Modelle sind hierbei der menschlichen Neurophysiologie zum Teil angelehnt [5].

ML Algorithmen prozessieren wiederholt Daten (aus dem Trainingsdatensatz) und können im Anschluss dasselbe Problem anhand zuvor ungesehener Daten (aus dem Testdatensatz) lösen – im Idealfall ähnlich gut oder besser als das menschliche Individuum. ML lernt, durch Wiederholung des Lernprozesses die Fehlerrate hinsichtlich einer bestimmten Aufgabe zu minimieren (Beispiel: Zeigt das Röntgenbild eine apikale Läsion - ja oder nein?). Die Algorithmen lernen demnach aus Erfahrung, können Informationen aus neuen Daten abstrahieren und die Genauigkeit als auch Effizienz der Datenprozessierung und -ausgabe optimieren. Somit benötigt die Maschine nicht zwingend explizite Instruktionen der Programmierer*innen, sondern der Algorithmus legt anhand der von ihm evaluierten Daten sein eigenes Regelwerk fest [6].

Zur Verarbeitung sehr komplexer Daten werden vermehrt sogenannte Neuronale Netzwerke (NN) verwendet [7], deren mathematische Architektur dem menschlichen Gehirn nachempfunden ist. NN bestehen aus verschiedenen Schichten – (1) Eingabeschicht: Input Layer, (2) Ausgabeschicht: Output Layer, (3) verborgene Schichten: Hidden Layers – die mathematisch miteinander verbunden sind. NN mit tiefen, mehrschichtigen Netzstrukturen gehören dem Deep Learning (DL) an und werden beispielsweise in der Bilderkennung genutzt [7].

Der Lernprozess von Algorithmen kann auf unterschiedliche Weise stattfinden [7]. Das überwachte Lernen (engl. Supervised Learning) ist eine maschinelle Lernmethode, bei der die Zielvariable bereits bekannt ist. Das System wird aktiv mit Datensätzen aus Eingabe und erwarteter Ausgabe trainiert [8], das heißt, Daten (z.B. aus Röntgenbildern) werden mit einer bestimmten Information („label“) verknüpft (bspw. Vorhandensein einer apikalen Läsion). Die erlernten Muster aus den Beispieldaten werden anschließend auf ungesehenen Daten angewandt, um Vorhersagen zu treffen. Methoden des überwachten Lernens werden unter anderem für Klassifikationsaufgaben genutzt. Zu den gängigsten Algorithmen gehören hierbei z.B. die logistische Regression und baumbasierte Verfahren [9] wie die ML Modelle „Random Forest“ oder „Gradient Boosting“.

Im Gegensatz dazu steht das unüberwachte Lernen (engl. Unsupervised Learning), das keine Vorgaben der richtigen Lösung enthält, sondern mittels sogenanntem „Clustering-Verfahren“ funktioniert. Die Algorithmen suchen explorativ nach Ähnlichkeiten zwischen Datenstücken, die anschließend kategorisiert und in zusammenhängende Gruppen („Cluster“) sortiert werden [6]. Bei dieser maschinellen Lernmethode sind die Eingangsdaten nicht beschriftet („labeled“); der Algorithmus trainiert somit anhand von Beispieldaten ohne Zielvariable. Wenn spezifische Cluster identifiziert wurden, erfordert es häufig wieder menschliches Zutun diese weiter zu analysieren oder zu visualisieren [8, 10]. Ein Beispiel in der Medizin ist das Erkennen genetischer Muster in der Evolutionsbiologie [7].

Als dritte Lernstrategie ist das bestärkende oder verstärkte Lernen (engl. Reinforcement Learning) anzufügen. Hierbei lernt das System im Sinne von „Versuch und Irrtum“ (engl. „trial and error“) [7], das heißt, es wird so lange nach einer Lösung des Problems gesucht bis sie schließlich gefunden wurde, wobei bewusst Fehler hingenommen werden und für den Lernprozess sogar von Nöten sind.

Alle angewandten Modelle in dieser Studie fallen in die Kategorie des überwachten Lernens.

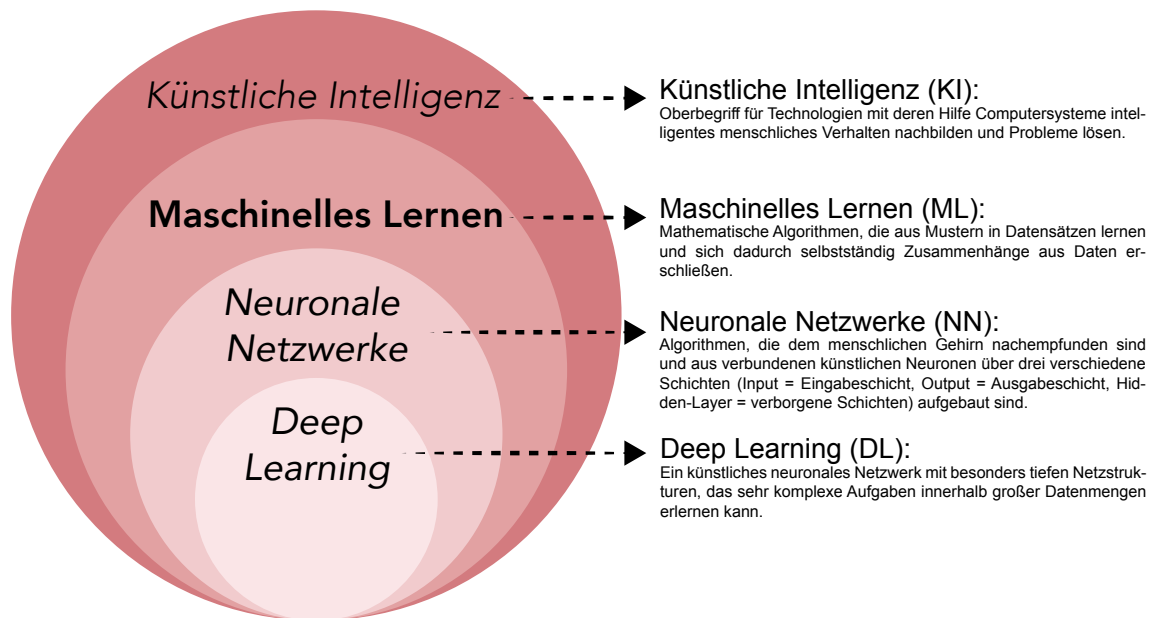


Abbildung 1 Übersicht Pfade der künstlichen Intelligenz (eD)

1.2 Anwendung der KI im Gesundheitswesen und in der Zahnmedizin

Innerhalb der letzten Dekade fand KI vermehrt auch Anwendung im Gesundheitswesen mit der Zielsetzung, komplexe Fragestellungen zu lösen und die Effizienz von Arbeitsprozessen zu steigern, was langfristig zu mehr Effektivität und Präzision in Behandlung und Diagnostik führen kann. In der Medizin werden zum einen mathematische Algorithmen u.a. zur Diagnostik (z.B. Erkennung von Hautkrebs in der Dermatologie [11] oder diabetischer Retinopathie in der Ophthalmologie [12]) und Prognosestellung, zur automatisierten Terminplanung oder zur Dosierung von Medikamenten angewandt; ebenso kommen KI-gestützte Roboter zum Einsatz, die zum Beispiel assistierend bei Operationen genutzt werden [1].

In der Diagnostik von Erkrankungen konnte KI teils gleichwertige oder häufiger korrekte Aussagen als das medizinische Fachpersonal treffen. Somit könnten möglicherweise diagnostische und in der Konsequenz auch therapeutische Fehler der Behandler*innen in der Zukunft vermieden werden [13]. Im Bereich der Zahnmedizin findet KI Anwendung in der Detektion von Pathologien in Röntgenbildern, zum Beispiel zur Erkennung von apikaler Parodontitis [14] und Parodontalerkrankungen [15] oder kariösen Läsionen [16] mit-

tels DL Netzwerken. Auch in der Prothetik werden KI Modelle angewandt, um beispielsweise die Langlebigkeit von Implantaten zu prognostizieren [17]. KI unterstützt die zahnärztlichen Behandler*innen hierbei zum einen in der Diagnosestellung und zum anderen in der Auswahl einer geeigneten Therapie.

1.3 Die Prognose einer Wurzelkanalbehandlung

Als Wurzelkanalbehandlung wird die Aufbereitung, Desinfektion und Obturation des Endodonts bezeichnet. Meist ist eine WKB bei ausgeprägten Entzündungszuständen der Pulpa und/oder chronischen Entzündungsprozessen an der Wurzelspitze indiziert [18]. Jede Behandlung ist individuell und von patienten- und zahnspezifischen Aspekten abhängig. Die präoperative Bewertung des betroffenen Zahnes und die Schwierigkeitseinschätzung der WKB sind relevant für eine adäquate klinische Behandlung. Hierbei ist der Erfolg oder Misserfolg wurzelkanalbehandelter Zähne von verschiedenen Faktoren abhängig, die die Prognose negativ oder positiv beeinflussen.

1.3.1 Patientenbezogene Faktoren

Nicht allein die präoperative röntgenologische und zahnbezogene Diagnostik, sondern auch eine vollständige Patientenanamnese enthält wichtige Informationen für die zahnärztlichen Behandler*innen. So konnte in der Literatur gezeigt werden, dass patientenbezogen unter anderem systemische Erkrankungen wie Diabetes mellitus (DM) einen signifikanten Einfluss auf den Therapieerfolg nahmen. Diabetespatient*innen wiesen in Studien im Allgemeinen eine höhere Prävalenz für apikale Läsionen auf [19-21] und infolgedessen kam es zu einem erhöhten Verlust an wurzelkanalbehandelten Zähnen [22, 23]. Bei Patient*innen mit DM Typ 2 zeigten sich zudem deutlich geringere Ausheilungsraten periapikaler Entzündungszustände [24]. Diabetespatient*innen hatten häufiger Symptome, im Allgemeinen größere periradikuläre Läsionen [25] und ein verändertes bakterielles Spektrum im nekrotischen Wurzelkanal im Vergleich mit gesunden Patient*innen [26]. Auch rezurrente endodontische Infektionen waren bei Diabetespatient*innen signifikant häufiger zu beobachten [27]. Generell war die Erfolgsrate einer WKB in Studien bei Patient*innen mit DM geringer als bei Patient*innen ohne Diabeteserkrankung [28]. Darüber hinaus scheint ein schlecht eingestellter Diabetes orale Infektionen zu

begünstigen. So kam es häufiger zu Pulpaerkrankungen und apikaler Parodontitis bei Patient*innen mit hohen HbA1c Werten. Patient*innen mit gut eingestellten HbA1c Werten zeigten hingegen höhere Erfolgsraten einer WKB, was diese Kreuz-Abhängigkeit nochmals unterstützt [29]. DM führt aufgrund erhöhter Blutglukosewerte und konsekutiver Hyperglykämie zu einer gestörten Immunantwort [30], sodass endodontische Pathologien häufiger und stärker ausgeprägt sind. Infektionen können schneller voranschreiten und die Heilungskapazität der periapikalen Gewebe kann herabgesetzt sein [31].

In Bezug auf kardiovaskuläre Erkrankungen zeigten Studien unterschiedliche Ergebnisse, sodass kein eindeutiger Zusammenhang mit dem Misserfolg einer WKB formuliert werden kann [32]. Einige Autor*innen berichteten von einem bis zu 67% erhöhten Risiko für Misserfolge endodontisch behandelter Zähne bei Patient*innen mit kardiovaskulären Erkrankungen [33]. Auch die Extraktionsraten waren bei Patient*innen mit Erkrankungen wie Hypertonie oder koronarer Herzkrankheit erhöht [34]. Außerdem konnte eine Assoziation zwischen dem vermehrten Auftreten einer apikalen Parodontitis und kardiovaskulären Erkrankungen festgestellt werden [35]. Weitere Daten zeigten auf, dass nach erfolgreicher WKB das C-reaktive Protein im Serum (high-sensitivity C-reactive protein, hsCRP) sank und somit auch das Risiko für kardiovaskuläre Erkrankungen reduziert wurde [36]. Andere Studien konnten jedoch keinen signifikanten Unterschied in der Prognose von Wurzelkanalbehandlungen zwischen gesunden Patient*innen und Patient*innen mit kardiovaskulären Erkrankungen feststellen [37, 38].

Weiterhin gibt es Hinweise, dass bei Raucher*innen ein erhöhtes Risiko für Misserfolge vorliegt, wenngleich die Evidenz durch Studien noch recht gering ist. Der negative Einfluss des Rauchens auf Wundheilungsprozesse ist in der Medizin seit langem nachgewiesen [39, 40], jedoch konnten die einzelnen Mechanismen bisher nicht vollständig erschlossen werden. Es wird angenommen, dass regelmäßiger Tabakkonsum die Mikrovaskularisation stört und somit zu einer erniedrigten Sauerstoffkonzentration im Blut führt [41]; freie Radikale [42] schädigen zusätzlich die Endothelzellen. Entzündetes periapikales Gewebe könnte durch diese Effekte weniger perfundiert werden und es resultiert eine Mangelversorgung, unter anderem mit Sauerstoff. Außerdem zeigen Raucher*innen zum einen eine verzögerte Einwanderung von Fibroblasten in Wunden und zum anderen eine gestörte Funktion dieser Zellen, sodass die Kollagensynthese zur Reparatur des Gewebes nicht ausreichend stattfinden kann [43]. Raucher*innen haben ein erhöhtes

RANKL/OPG Verhältnis im Speichel, was durch ein Übergewicht von RANKL und eine Suppression der Osteoprotegerinproduktion [44] in der Folge knochenabbauende Prozesse fördern kann. Rauchen unterdrückt zusätzlich die Funktion von Abwehrzellen wie polymorphkernigen Leukozyten, Makrophagen und T-Zellen, was in der Konsequenz zu einem reduzierten Level an Antikörpern führt [45]. Auf der anderen Seite findet durch das Rauchen eine stärkere systemische Entzündungsreaktion durch den Anstieg des C-reaktiven Proteins im Serum und pro-inflammatorischer Zytokine wie IL-1 β und TNF- α statt [45]. Dieses immunologische Ungleichgewicht kann zu einem gestörten und weniger effizienten Abwehrmechanismus in der Pulpa führen und somit die Prognose einer WKB negativ beeinflussen. In der Literatur finden sich eine erhöhte Prävalenz wurzelkanalbehandelter Zähne und apikaler Läsionen [46] bei Raucher*innen. Es wurde eine bis zu dreifach erhöhte Extraktionsrate von endodontisch behandelten Zähnen in rauchenden Patient*innen beschrieben [47]. Ein besonders hohes Risiko für Misserfolge bei Raucher*innen ergab sich bei zusätzlich vorliegendem Knochenabbau im Sinne einer Parodontitis [48].

Bestanden mehrere systemische Erkrankungen gleichzeitig, stieg in Studien auch das Risiko für einen Misserfolg der WKB [49], wobei ursächlich nicht geklärt ist, ob die Kombination der Erkrankungen oder eine höhere Gewichtung einer einzelnen Erkrankung die Prognose der WKB letztendlich verschlechterte.

1.3.2 Zahnbezogene Faktoren

Bei den zahnbezogenen Faktoren sind vor allem das Vorliegen einer apikalen Parodontitis und ein vorhandener Knochenabbau Risikofaktoren für eine WKB [50]. Insbesondere präoperative apikale Läsionen resultierten in einer signifikant höheren Misserfolgsrate endodontischer Behandlungen. In der Literatur wurden Erfolgsraten von 93% bei Zähnen ohne apikale Parodontitis und 80% bei Zähnen mit apikaler Parodontitis beschrieben [51]. Ursächlich hierfür ist eine verstärkte Keimbesiedelung in den Dentintubuli und Verzweigungen des Kanalsystems, wodurch eine suffiziente Reinigung und Desinfektion des Endodonts erschwert ist [52]. Zähne mit erhöhten parodontalen Taschentiefen von über 5 mm wiesen teilweise ein doppelt so hohes Risiko für eine spätere Extraktion nach erfolgter WKB auf [23]. Durch den vorliegenden Knochenabbau werden zusätzliche Eintrittspforten für Keime über freiliegende Dentinoberflächen geschaffen [53]. Umgekehrt zeigt

auch die intrakanaläre Infektion des Endodonts einen synergetischen Effekt auf das Parodont und fördert somit eine weitere Progression der Parodontitis [54, 55].

1.3.3 Behandlungsspezifische Faktoren

Behandlungsspezifische Faktoren wie eine suffiziente koronale Restauration oder eine optimale Wurzelkanalfüllung werden im Zusammenhang mit erhöhten Erfolgsraten diskutiert. Endodontisch behandelte Zähne, die mit einer indirekten höckerüberkuppelnden Versorgung restauriert wurden, wiesen in der Literatur signifikant höhere Langzeitprognosen auf als nicht restaurierte Zähne [23, 49]. Die koronale Undichtigkeit der postendodontischen Restauration ist eine der häufigsten Ursachen für eine erneute Besiedelung des Wurzelkanalsystems mit Keimen und für ein Wiederaufflammen von endodontischen Infektionen [56]. Der Einfluss der Qualität einer Wurzelfüllung auf den Therapieerfolg wird in Studien kontrovers diskutiert. Einige Veröffentlichungen zeigten eine signifikante Verbesserung der Prognose bei optimaler Wurzelfüllung [23, 28, 57], andere Arbeiten stellten keinen Zusammenhang fest [51, 58].

1.4 Zielstellung

In vorherigen wissenschaftlichen Arbeiten wurden die Zusammenhänge zwischen den einzelnen Faktoren und der Prognose einer WKB meist anhand von uni- und bivariaten Statistiken oder reinen Assoziationsanalysen aufgezeigt. Bisher wurden für diese Fragestellungen keine Vorhersagemodelle angewendet. Ziel der vorliegenden Dissertation war es zum einen, Assoziationen von patienten-, zahn- und behandlungsspezifischen Faktoren mit dem Erfolg oder Misserfolg einer WKB abzuleiten. Zum anderen sollten mittels komplexer ML Modelle (Random Forest (RF), Gradient Boosting Machine (GBM) und Extreme Gradient Boosting (XGB)) Vorhersagen über die Prognose einer WKB unternommen werden. Zuletzt sollten aus den Modellen Rangfolgen erstellt werden, die die relative Gewichtung (Relevanz) der einzelnen Variablen für den Behandlungserfolg oder -misserfolg abbilden.

2. Methodik

2.1 Datenakquise und Studiendesign

Das Reporting unserer Studie orientierte sich an den Richtlinien nach STROBE (Strengthening the Reporting of Observational studies in Epidemiology) [59] und TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) [60] und wurde von der Ethikkommission der Charité - Universitätsmedizin Berlin (EA4/080/18) genehmigt. Es handelt sich um eine retrospektive longitudinale Analyse von Wurzelkanalbehandlungen (Primärbehandlungen und Revisionen), die im Zeitraum vom 1. Januar 2016 bis zum 31. März 2020 am CharitéCentrum 03 für Zahn-, Mund- und Kieferheilkunde mit einer Mindestnachuntersuchungszeit von sechs Monaten durchgeführt wurden. Alle begutachteten Röntgenbilder wurden mittels digitaler Sensortechnologie oder Speicherfolie mit einem Röntgengerät (CS 2200, Carestream Dental LLC, Atlanta, USA) erstellt und anschließend in einer diagnostischen Software (Merlin, Phönix-PACS, Freiburg im Breisgau, Deutschland) auf hochauflösenden Bildschirmen (MultiSync EA275WMi, NEC, Tokyo, Japan) analysiert.

Alle patientenbezogenen Dokumente (Anamnese, Patientenakte, Röntgenbilder) wurden vor der Analyse in Form eines spezifischen Kodierungssystems anonymisiert. Eine Bewerterin (CH) traf anschließend eine Vorauswahl aus allen Patientendaten nach zuvor festgelegten Kriterien. Nach Kalibrierung mit einem anderen Bewerter (SH) wurden alle inkludierten Röntgenaufnahmen und Patientendokumente von CH analysiert. Zur Bestimmung der Urteilsübereinstimmung klassifizierten beide Bewerter*innen den Periapikalen Index (PAI-Score) der ersten 50 Patient*innen unabhängig voneinander (Cohens Kappa Test, R Studio Version 1.3.1073, Package „psych“).

Die Inklusionskriterien lauteten wie folgt: Vollständige Dokumentation von Anamnese und Patientendaten, Vorhandensein von prä- und postoperativen Röntgenbildern sowie mindestens einer Röntgenaufnahme im Nachuntersuchungszeitraum, eine Mindestnachuntersuchungszeit von sechs Monaten und eine unterschriebene Zustimmung der Patient*innen an der Teilnahme an einer Studie. Exklusionskriterien umfassten behandelte Milchzähne, Pulpotomien und nicht beendete Wurzelkanalbehandlungen.

Wie zuvor dargelegt, wurden die einbezogenen Fälle zunächst nach vorgegebenen Inklusionskriterien sortiert. Darüber hinaus wurde eine Fallzahlplanung für die logistische Regression durchgeführt, die auf einer Evaluation realer Patientendaten aus einer Beobachtungsstudie basiert [61]. Hierbei wurde beurteilt, inwieweit sich unterschiedliche Stichprobengrößen auf die Diskrepanz zwischen den Stichprobenstatistiken und den tatsächlichen Parametern der Zielpopulation auswirken. Dieser Vergleich diente dazu, eine Mindeststichprobengröße für die logistische Regression zu schätzen, die für eine Forschungsarbeit erforderlich ist und auf die Grundgesamtheit übertragen werden kann. Der somit empfohlene Mindeststichprobenumfang für eine logR ergab eine Mindestfallzahl von $n = 500$. Unsere Fallzahlgröße betrug letztlich $n = 591$ und erfüllte dieses Kriterium.

Die Behandlung der eingeschlossenen Fälle wurde von Zahnmedizinstudent*innen unter Supervision oder approbierten Zahnärzt*innen nach weitgehend standardisierten Schemata durchgeführt. Alle Zähne wurden zunächst auf die anschließende Restaurierbarkeit geprüft und erhielten zu diagnostischen Zwecken ein präoperatives Röntgenbild. Die Isolation der Zähne erfolgte durch eine absolute Trockenlegung mit Kofferdam. Die Arbeitslänge wurde mithilfe der elektrometrischen Längenbestimmung festgelegt und anschließend durch eine Kontrastaufnahme verifiziert. Das Behandlungsprotokoll umfasste eine wiederholte intraoperative Spülung mit Natriumhypochlorit (NaOCl). Die Präparation der Wurzelkanäle wurde indikationsgerecht mit verschiedenen Feilensystemen (vorwiegend ProTaper Next (Dentsply Sirona, York, USA), Reciproc (VDW, München, Deutschland) oder WaveOne Gold (Dentsply)) durchgeführt. Zwischen den einzelnen Behandlungssitzungen erfolgte eine medikamentöse Einlage mit Kalziumhydroxid (UltraCal XS, Ultradent, South Jordan, USA). Vor der abschließenden Obturation der Wurzelkanäle mit Gut-tapercha und AH Plus (Dentsply Sirona) wurden die Kanäle mit NaOCl (in einer Konzentration von 1% oder 3%) und Ethylendiamintetraessigsäure (EDTA) 17% (CanalPro EDTA 17%, Coltene, Altstätten, Schweiz) schallaktiviert gespült. Unterschiede in der Behandlung zwischen Zahnmedizinstudent*innen und approbierten Zahnärzt*innen bestanden zum einen in der Konzentration des Natriumhypochlorits (1% NaOCl bei Student*innen, 3% NaOCl bei Zahnärzt*innen) und zum anderen in der Wurzelfülltechnik (modifizierte Zentralstifttechnik bei Student*innen, warm-vertikale Obturation nach Schilder bei Zahnärzt*innen).

2.2 Auswertungskriterien und Einordnung der Variablen

Ein Misserfolg der WKB wurde anhand klinischer und radiologischer Parameter bestimmt. Klinische Kriterien für Misserfolge waren definiert über das Bestehen von Symptomen (z.B. Vorhandensein von Fisteln oder Schwellungen, Perkussionsempfindlichkeit des Zahnes, auffällige Palpationsbefunde, bestehende Schmerzsymptomatik der Patient*innen) und die Notwendigkeit weiterführender Behandlungen wie Revisionen, Wurzelspitzenresektionen (WSR) oder Extraktionen.

Radiologisch wurde der „Periapikale Index“ (PAI-Score) nach Ørstavik 1986 [62] zur Bewertung eines Erfolges oder Misserfolges verwendet. Als gesund wurden Zähne mit einem PAI-Punktwert von 1 oder 2 bewertet. Ein Erfolg der Behandlung im Sinne einer Heilungstendenz wurde bei einem reduzierten PAI-Punktwert im Vergleich zum Ausgangsbild definiert (Abb. 2a/b). Als Misserfolg galten Zähne mit einem konstanten PAI-Punktwert von 3 bis 5 oder einem im Vergleich zum Ausgangsbild erhöhten PAI-Punktwert größer oder gleich 3 im Sinne eines progredienten oder konstanten apikalen Entzündungsgeschehens (Abb. 2c/d).



Abbildung 2 Veranschaulichung der röntgenologischen Auswertung, die roten Pfeile zeigen die äußere Begrenzung der periapikalen Läsion. (a) Ausgangsbild 47 mit einem PAI-Punktwert von 5, (b) Recallbild 47 mit einem PAI-Punktwert von 3 = Erfolg; (c) Ausgangsbild 47 mit einem PAI-Punktwert von 5, (d) Recallbild 47 mit gleichbleibendem Höchstscore von 5, aber deutlich progredientem apikalen Entzündungsgeschehen = Misserfolg (eD)

Aus statistischen Gründen erfolgte eine Dichotomisierung in erfolgreiche Behandlung („gesund“ oder „erkennbare Heilungstendenz“) und Misserfolg („Ausbleiben einer Heilung/Heilungstendenz“). Alle Auswertungskriterien sind in Tabelle 1 zusammengefasst.

Tabelle 1 Zusammenfassung der Auswertungskriterien (eD)

	Erfolg	Misserfolg
Klinisch	Ausbleiben von Symptomen	Bestehende Schmerzsymptomatik der Patient*innen Vorhandensein von Fisteln oder Schwellungen Perkussionsempfindlichkeit des Zahnes Auffällige Palpationsbefunde Weiterführende Behandlungen (Revision, WSR, Extraktion)
Radiologisch	Gesund: PAI-Punktwert 1-2 Reduzierter PAI-Punktwert im Vergleich zum Ausgangsbild	Konstanter PAI-Punktwert von 3-5 Erhöhter PAI-Punktwert bei Wiedervorstellung

Im Anschluss wurden alle gemessenen Variablen entsprechend kategorisiert. Die zahnbezogenen Faktoren umfassten den Zahntyp (Frontzahn, Prämolare oder Molare), die Behandlungsart (Primärbehandlung oder Revision) und das Knochenniveau, gemessen anhand der Distanz von der Schmelz-Zement-Grenze zum Alveolarkamm im Röntgenbild (kein oder milder Knochenabbau: 0-33%, moderater Knochenabbau: 33-66%, schwerer Knochenabbau: > 66%). Der periapikale Zustand wurde anhand des PAI-Scores auf dem prä- und postoperativen Röntgenbild bewertet. Für die statistische Auswertung erfolgte eine Gruppierung der PAI-Scores: PAI-Punktwerte 1 und 2 galten als „physiologisch“, PAI-Punktwert 3 als „apikale Läsion präsent“ und PAI-Punktwerte 4 und 5 als „schwere apikale Parodontitis“.

Bei den patientenbezogenen Faktoren wurde zum einen nach Alter gruppiert (< 30 Jahre, 30-50 Jahre und > 50 Jahre), zum anderen wurden systemische Erkrankungen (kardiovaskuläre Erkrankungen, Diabetes mellitus, andere immunsupprimierende Zustände, andere systemische Erkrankungen) erfasst und die Patient*innen in Raucher*innen beziehungsweise Nichtraucher*innen eingeteilt.

Die behandlungsspezifischen Variablen unterschieden sich in der Art der Behandler*innen (Zahnmedizinstudent*innen oder approbierte Zahnärzt*innen), der Länge der Wurzelfüllung (WF) auf den postoperativen Röntgenbildern (optimal = 0-2 mm Abstand zum radiologischen Apex, suboptimal = mehr als 2 mm Abstand zum radiologischen Apex im Sinne einer zu kurzen oder überextendierten Wurzelfüllung; auch eine Sealerextrusion wurde dokumentiert) und der koronalen Restauration bei Wiedervorstellung der Patient*innen (vorhanden beziehungsweise nicht vorhanden).

2.3 Bewertung des Risikos für Bias

Die strikten Inklusionskriterien (Mindestnachuntersuchungszeitraum sechs Monate sowie vollständige klinische und röntgenologische Dokumentation) führten möglicherweise zu Selektions- und Nachuntersuchungsbias, da von anfänglich 2288 Patient*innen mit 3036 Zähnen nur 458 Patient*innen mit 591 Zähnen eingeschlossen wurden.

Das Risiko von Informationsbias wurde bei der Konzeptualisierung der Studie reduziert, indem die auswertende Autorin (CH) grundsätzlich die Ausgangsröntgenbilder zeitlich versetzt von den Nachuntersuchungsröntgenbildern bewertete. Der Zeitraum betrug in diesem Fall drei Monate. Dadurch konnte eine Verblindung innerhalb der einzelnen Behandlungsfälle erreicht werden.

2.4 Statistische Analyse

Die gesamte statistische Auswertung erfolgte mit R via RStudio (Version 1.3.1073-1). Für die univariate statistische Analyse der Gruppen wurde der Chi-Quadrat-Test und Exakte Fisher-Test durchgeführt. Multivariate Analysen erfolgten mithilfe von vier verschiedenen

ML Modellen im Caret Package (Classification And REgression Training, Version 6.0-78): Logistische Regression (logR), Random Forest (RF), Gradient Boosting Machine (GBM) und Extreme Gradient Boosting (XGB).

Zunächst wurde eine logR auf dem gesamten Datensatz vorgenommen, um die Odds-Ratio (OR) und die jeweiligen 95%-Konfidenzintervalle (95% CI) zu bestimmen. Für die Vorhersagemodelle wurde der Datensatz im Verhältnis 60/40 aufgeteilt. Anhand von 60% der Daten wurden die Modelle wiederholt mittels dreifacher Kreuzvalidierung trainiert und 40% der Daten wurden dem Algorithmus anschließend ungesehen präsentiert. Als Referenz wurden dem Algorithmus nach dem Training ebenfalls die bekannten Trainingsdaten als „neue“ Testdaten zur Prädiktion offeriert. ROC-Kurven dienten als Grundlage für die Optimierung der Modelle. Die Ergebnisse der einzelnen Modelle evaluierten wir im Anschluss anhand einer Wahrheitsmatrix (Treffsicherheit, F1 Score, Sensitivität, Spezifität, positive und negative Vorhersagewerte). Die Qualität der Modelle und statistisch signifikante Unterschiede zwischen den Modellen wurden zusätzlich mit der Fläche unter der Grenzwertoptimierungskurve (ROC Area under the curve = ROCAUC, R package pROC Version 1.16.2) bestimmt. Für die Treffsicherheit und AUC-Werte wurden 95%-Konfidenzintervalle kalkuliert, um statistisch signifikante Unterschiede zwischen den ML Modellen zu erkennen.

Darüber hinaus erfolgte die Bestimmung der No-Information-Rate (NIR), die anschließend mit den Werten für die Treffsicherheit mittels einseitigem binomialen Hypothesentest verglichen wurde. Ziel hierbei war es herauszufinden, ob ein Unterschied zwischen den Treffsicherheiten der Vorhersagemodelle und der konsequenten Annahme der Mehrheitsklasse (in diesem Fall „erfolgreiche Behandlung“) vorlag.

Des Weiteren wurde bei jedem Modell dem jeweiligen Einflussfaktor ein entsprechender Rang auf Basis der Gewichtung (in Prozent) zugeordnet. Diese Ränge wurden zwischen den Modellen verglichen und für jeden Einflussfaktor die Ränge addiert sowie der Mittelwert der Rangsummen gebildet. Final erstellten wir ein modellübergreifendes Faktorgewichtungsranking. Das Ziel hierbei war, die Konsistenz der Faktorgewichtung zwischen den Prädiktionsmodellen zu evaluieren.

3. Ergebnisse

3.1 Studienteilnehmer*innen und deskriptive Daten

Die der Datenanalyse vorangegangene Urteilsübereinstimmung der beiden Bewerter*innen (CH und SH) wurde mittels Cohens Kappa Gleichung bestimmt. Der ungewichtete Übereinstimmungswert betrug 0,89 (95% CI [0,79; 0,99]) und der gewichtete Wert 0,97 (95% CI [0,97; 0,97]).

In unserer Analyse wurden insgesamt 2288 Patient*innen mit 3036 wurzelkanalbehandelten Zähnen geprüft, von denen letztlich 458 Patient*innen (weiblich n = 216 (47,2%), männlich n = 242 (52,8%); Altersdurchschnitt 53,4 (median/min.-max./SD: 56,0/10,0-90,0/17,0) Jahre) mit 591 bleibenden Zähnen (Frontzähne: 141 (23,8%), Prämolaren: 160 (27,1%), Molaren: 290 (49,1%)) in den Datensatz eingeschlossen wurden (Abb. 3). Die verbliebenen 1830 Patient*innen mit 2445 Zähnen wurden aufgrund fehlender beziehungsweise unzureichender Nachuntersuchungszeiten, unvollständiger Patientendaten und Röntgenbilder oder inkompletter Anamnese exkludiert. Der größte Anteil (64,8%) fiel hierbei der nicht erfüllten Nachuntersuchungszeit von sechs Monaten zu.

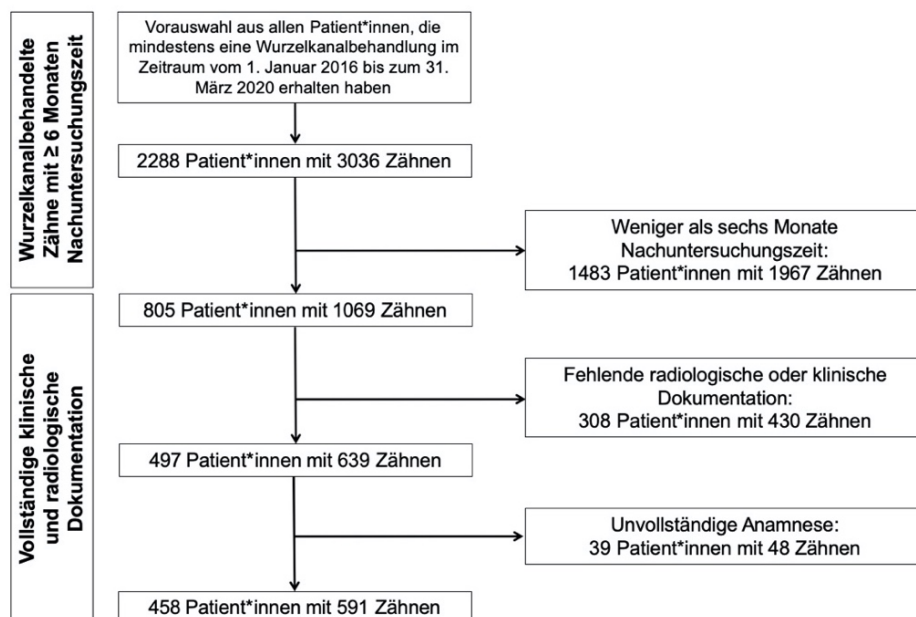


Abbildung 3 Patientenselektionsprozess (mod. nach Herbst CS, Schwendicke F, Krois J, Herbst SR. Association between patient-, tooth- and treatment-level factors and root canal treatment failure: A retrospective longitudinal and machine learning study. J Dent. 2022; S. 2)

Die durchschnittliche Nachuntersuchungszeit der Fälle betrug 15,2 (median/min.-max./SD: 12,4/6,0-48,0/8,8) Monate. 48,7% der Fälle wurden im Zeitraum von 6-12 Monaten klinisch als auch radiologisch nachuntersucht. 35,2% der Zähne lagen innerhalb eines Nachuntersuchungszeitraums von 12-24 Monaten und 16,1% wiesen mehr als 24 Monate Nachuntersuchungszeit auf.

Die Wurzelkanalbehandlungen wurden zu näherungsweise gleichen Anteilen entweder von Zahnmedizinstudent*innen unter Supervision (n = 282 (52,3%)) oder approbierten Zahnärzt*innen (n = 309 (47,7%)) durchgeführt.

48,2% der untersuchten Zähne waren gesunden Patient*innen zuzuordnen. Innerhalb der verschiedenen Zahngruppen (Frontzähne, Prämolaren, Molaren) fanden sich hierbei statistisch keine signifikanten Unterschiede ($p=0,057$). Am häufigsten waren kardiovaskuläre Erkrankungen unter den Patient*innen prävalent (35,9%), gefolgt von anderen systemischen Erkrankungen (31,6%). DM trat in 5,6% der Patientenfälle auf, mit statistisch signifikanten Unterschieden innerhalb der Zahngruppen ($p=0,001$). Unter den systemischen Erkrankungen waren andere immunsupprimierende Zustände am seltensten (4,2%) prävalent und 15,6% der Zähne bei Raucher*innen vorkommend, wobei es statistisch keine signifikanten Unterschiede der Zahngruppen ($p=0,735$) gab.

Auf der Ebene zahnbezogener Parameter wiesen die meisten Zähne (76%) keinen oder einen milden (< 33%) alveolären Knochenabbau auf, währenddessen bei 18,4% ein moderater (33-66%) und bei 5,6% ein schwerer Knochenabbau (> 66%) vorlag. Hierbei konnte ein statistisch signifikanter Unterschied innerhalb der Zahngruppen festgestellt werden ($p=0,003$). Eine optimale Wurzelkanalfüllung fand sich in 82,1% der Fälle mit statistisch signifikanten Unterschieden zwischen den Zahngruppen ($p<0,001$). 41% der Zähne erhielten nach abgeschlossener WKB eine höckerüberkuppelnde indirekte Restauration, die restlichen 59% blieben provisorisch oder definitiv mit direkten Restaurationen versorgt. Bei 63,3% der wurzelkanalbehandelten Zähne konnte eine apikale Läsion detektiert werden. In 45,0% der Fälle lag ein PAI-Punktwert von 3 vor und in 18,3% der Fälle von über 3 mit statistisch signifikanten Unterschieden innerhalb der Zahngruppen ($p=0,048$). Ein PAI-Punktwert von 5 trat bei Patient*innen mit DM signifikant häufiger auf

(33,3%) als bei Patient*innen ohne Diabeteserkrankung (16,4%, $p=0,034$). Auch kardiovaskuläre Erkrankungen waren häufiger mit einem erhöhten PAI-Score von 5 assoziiert (in 78,8% der Fälle) als bei gesunden Patient*innen (33,3%, $p=0,001$).

Die Gesamtmisserfolgsrate der Wurzelkanalbehandlungen lag bei 20,5% mit einer jährlichen Misserfolgsquote von 6,4%. Hierbei fand sich statistisch kein signifikanter Unterschied zwischen gesunden Patient*innen (79,7%) und Patient*innen mit Vorerkrankungen (79,3%, $p=0,976$). Auch Patient*innen mit DM wiesen insgesamt keine höheren Misserfolgsraten (in 78,8% der Fälle Behandlung erfolgreich) als Patient*innen ohne Diabeteserkrankung (79,6% Erfolgsquote, $p=0,828$) auf.

Alle Ergebnisse der deskriptiven Statistik sind in Tabelle 2 zusammengefasst dargestellt.

Tabelle 2 Zusammenhänge zwischen patienten-, zahn- und behandlungsspezifischen Faktoren und dem Zahntyp, bezogen auf die Gesamtanzahl (n = 591) an Wurzelkanalbehandlungen (Chi-Quadrat-Test und Exakter Fisher-Test). (mod. nach Herbst CS, Schwendicke F, Krois J, Herbst SR. Association between patient-, tooth- and treatment-level factors and root canal treatment failure: A retrospective longitudinal and machine learning study. J Dent. 2022; S. 4)

Faktoren	Fälle gesamt n = 591 % (n)	Frontzähne n = 141 % (n)	Prämolaren n = 160 % (n)	Molaren n = 290 % (n)	p-Wert
Anamnese					
Gesund	48,2 (285)	41,8 (59)	45,0 (72)	53,1 (154)	0,057
Kardiovaskuläre Erkrankungen	35,9 (212)	40,4 (57)	39,8 (63)	31,7 (92)	0,117
Diabetes mellitus	5,6 (33)	6,4 (9)	10,6 (17)	2,4 (7)	0,001
Andere immunsupprimierende Zustände	4,2 (25)	5,0 (7)	5,6 (9)	3,1 (9)	0,393
Andere systemische Erkrankungen	31,6 (187)	39,7 (56)	31,9 (51)	27,6 (80)	0,040
Rauchen	15,6 (92)	13,5 (19)	16,3 (26)	16,2 (47)	0,7347
Alveolärer Knochenabbau					
Physiologisch oder mild (< 33%)	76,0 (449)	64,5 (91)	81,9 (131)	78,3 (227)	0,003
Moderat (33-66%)	18,4 (109)	25,5 (36)	13,1 (21)	17,9 (52)	
Schwer (> 66%)	5,6 (33)	9,9 (14)	5,0 (8)	3,8 (11)	
Abstand der WF zum Apex					
Optimal (0-2 mm)	82,1 (485)	90,8 (128)	86,3 (138)	75,5 (219)	<0,001
Zu kurz (> 2 mm)	15,7 (93)	6,4 (9)	12,5 (20)	22,1 (64)	
Überextendiert	2,2 (13)	2,8 (4)	1,3 (2)	2,4 (7)	
Sealereextrusion	32,5 (192)	22,0 (31)	30,6 (49)	38,6 (112)	0,002
Dichte der WF					
Optimal	81,6 (482)	83,7 (118)	89,4 (143)	76,2 (221)	0,002
Suboptimal	18,4 (109)	16,3 (23)	10,6 (17)	23,8 (69)	
Qualität der WF zusammengefasst					
Optimal	69,0 (408)	78,7 (111)	76,9	60,0 (174)	<0,001
Suboptimal	31,0 (183)	21,2 (30)	23,1	40,0 (116)	
Indirekte höckerüberkuppelnde Restauration					
Vorhanden	41,0 (242)	34,0 (48)	38,7 (62)	45,5 (132)	0,061
Nicht vorhanden	59,0 (349)	66,0 (93)	61,3 (98)	54,5 (158)	
Periapikaler Zustand					
PAI > 3	18,3 (108)	22,7 (32)	14,4 (23)	18,3 (53)	0,048
PAI = 3	45,0 (266)	37,6 (53)	43,1 (69)	49,7 (144)	
PAI < 3	36,7 (217)	39,7 (56)	42,5 (68)	32,1 (93)	
Behandler*innen					
Zahnärzt*innen	47,7 (309)	22,3 (69)	21,4 (66)	56,3 (174)	<0,001
Student*innen	52,3 (282)	25,5 (72)	33,3 (94)	41,1 (116)	
Ergebnis der WKB					
Erfolg	79,5 (470)	77,3 (109)	81,3 (130)	79,7 (231)	0,697
Misserfolg	20,5 (121)	22,7 (32)	18,7 (30)	20,3 (59)	

3.2 Einfluss der Faktoren auf Misserfolge einer Wurzelkanalbehandlung

Mithilfe einer logR konnte gezeigt werden, dass zahnbezogene Faktoren die stärkste Assoziation mit Misserfolgen aufwiesen (Tabelle 3). Zähne mit schwerem alveolären Knochenabbau zeigten ein 6,5-fach erhöhtes Risiko für Misserfolge (OR 6,48; 95% CI [2,86; 14,89], $p < 0,001$) und Zähne mit moderatem Knochenabbau ein 2,5-fach erhöhtes Risiko (OR 2,59; 95% CI [1,49; 4,49], $p < 0,001$) im Vergleich zu Zähnen mit mildem Knochenabbau oder physiologischem Knochenniveau. Auch präsente und fortschreitende apikale Läsionen erhöhten das Risiko für einen Misserfolg deutlich. Ein PAI-Punktwert ≥ 4 war in unserer Analyse mit einem 4,5-fach erhöhten Risiko (OR 4,59; 95% CI [2,44; 8,79], $p < 0,001$) und ein PAI-Punktwert = 3 mit einem 2,5-fach erhöhten Risiko (OR 2,45; 95% CI [1,43; 4,34], $p < 0,01$) für einen Misserfolg der WKB assoziiert. Darüber hinaus nahm die Art der Behandlung einen statistisch signifikanten Einfluss auf den Ausgang der WKB. Revisionen zeigten ein um 77% erhöhtes Risiko für einen Misserfolg verglichen mit Primärbehandlungen (OR 1,77; 95% CI [1,01; 2,86]). Der Zahntyp nahm hingegen statistisch keinen signifikanten Einfluss auf die Prognose der Behandlung, auch Molaren zeigten kein erhöhtes Risiko.

Auf Patientenebene war vor allem der Faktor „Rauchen“ mit der Prognose einer WKB assoziiert. So zeigte sich in unserer Analyse ein zweifach erhöhtes Risiko für Misserfolge bei Raucher*innen im Vergleich zu Nichtraucher*innen (OR 2,05; 95% CI [1,18; 3,53], $p < 0,05$). Weitere systemische Erkrankungen (DM, kardiovaskuläre Erkrankungen, andere immunsupprimierende Zustände, andere systemische Erkrankungen) nahmen statistisch keinen signifikanten Einfluss auf das Ergebnis der WKB; auch das Alter der Patient*innen spielte hierbei keine Rolle.

Keine der untersuchten behandlungsspezifischen Faktoren (Länge der Wurzelfüllung, Behandler*innen, postendodontische Restauration) zeigte einen statistisch signifikanten Einfluss auf das Risiko für einen Misserfolg (Tabelle 3).

Tabelle 3 Ergebnisse der logistischen Regression (logR) (mod. nach Herbst CS, Schwendicke F, Krois J, Herbst SR. Association between patient-, tooth- and treatment-level factors and root canal treatment failure: A retrospective longitudinal and machine learning study. J Dent. 2022; S. 4)

Risikofaktor	Odds-Ratio	95% CI
Patientenbezogen		
Alter		
< 30 Jahre	Ref.	
30 - 50 Jahre	0,98	[0,44; 2,32]
> 50 Jahre	0,96	[0,42; 2,32]
Anamnese		
Raucher*innen (Ref.: Nichtraucher*innen)	2,05	[1,18; 3,53]*
Diabetes mellitus (Ref.: Kein Diabetes mellitus)	1,20	[0,41; 3,13]
Kardiovaskuläre Erkrankungen (Ref.: Keine kardiovaskulären Erkrankungen)	0,82	[0,49; 1,39]
Andere immunsupprimierende Zustände (Ref.: Keine anderen immunsupprimierenden Zustände)	2,62	[0,97; 6,80]
Andere systemische Erkrankungen (Ref.: Keine anderen systemischen Erkrankungen)	0,98	[0,60; 1,60]
Zahnbezogen		
Zahntyp		
Frontzähne	Ref.	
Prämolaren	0,98	[0,52; 1,85]
Molaren	1,02	[0,59; 1,81]
Behandlungsart		
Primärbehandlung	Ref.	
Revision	1,77	[1,01; 2,86]*
Alveolärer Knochenabbau		
Physiologisch oder mild (< 33%)	Ref.	
Moderat (33-66%)	2,59	[1,49; 4,49]*
Schwer (> 66%)	6,48	[2,86; 14,89]**
Periapikaler Zustand		
PAI-Punktwert 0-2	Ref.	
PAI-Punktwert = 3	2,45	[1,43; 4,34]*
PAI-Punktwert ≥ 4	4,59	[2,44; 8,79]**
Behandlungsspezifisch		
Länge der Wurzelfüllung		
Optimal	Ref.	
Suboptimal	1,53	[0,86; 2,66]
Behandler*innen		
Approbierte Zahnärzt*innen	Ref.	
Student*innen	1,57	[0,99; 2,50]
Indirekte höckerüberkuppelnde Restauration		
Vorhanden	Ref.	
Nicht vorhanden	0,82	[0,50; 1,33]

*p <0,05; **p <0,001

Vor allem ein kombiniertes Vorliegen verschiedener Faktoren erhöhte das Risiko deutlich, wie grafisch dargestellt in Abbildung 4. Als Referenz wurden der Periapikale Index (PAI-Score 1-5) und der alveoläre Knochenabbau (1: physiologisch oder mild; 2: moderat; 3: schwer) als größte Risikofaktoren unserer Analyse verwendet. So zeigte sich, dass z.B. bei Raucher*innen deutlich höhere Misserfolgsraten vorlagen; auch bei einem im Vergleich zu Nichtraucher*innen niedrigeren PAI-Punktwert oder geringeren Knochenabbau. Ähnliches konnte für Revisionen verglichen mit Primärbehandlungen visualisiert werden: Ein PAI-Punktwert von 4 mit gleichzeitig vorliegendem moderaten alveolären Knochenabbau war bei Revisionen bereits mit einer Misserfolgsrate von > 50% assoziiert, währenddessen für selbige Kombination der Risikofaktoren bei Primärbehandlungen ein Risiko von lediglich < 25% vorlag.

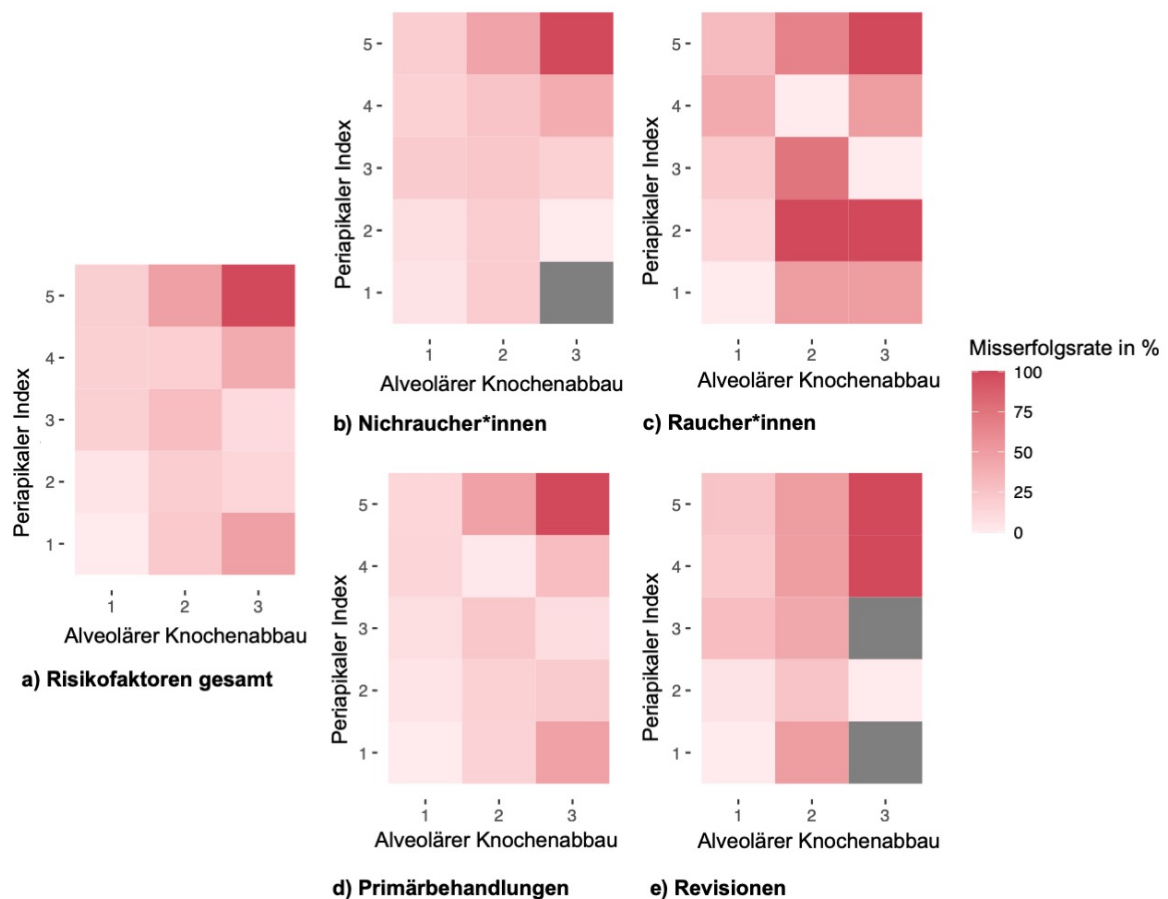


Abbildung 4 Heatmap Darstellung von Risikofaktoren einer Wurzelkanalbehandlung. (a) zeigt alle Risikofaktoren zusammengeführt (b) Nichtraucher*innen (c) Raucher*innen (d) Primärbehandlungen (e) Revisionen; graue Felder: keine Fälle/Daten (mod. nach Herbst CS, Schwendicke F, Krois J, Herbst SR. Association between patient-, tooth- and treatment-level factors and root canal treatment failure: A retrospective longitudinal and machine learning study. *J Dent.* 2022; S. 6)

3.3 Vorhersagemodelle und Faktorgewichtungsranking

Alle Modelle wiesen eine konsistent hohe Spezifität (logR: 96,5%; GBM: 93,5%; RF: 100%; XGB: 96,0%) auf dem Trainingsdatensatz, jedoch variierende Sensitivitäten auf. Hierbei zeigte die logR die niedrigste (29,5%) und der RF die höchste Sensitivität (79,5%). Dementsprechend wies die logR mit 0,77 (95% CI [0,71; 0,83]) den niedrigsten AUC-Wert und der RF den höchsten (0,99; 95% CI [0,98; 1,00]) auf. Basierend auf der Auswertung der Konfidenzintervalle eigneten sich RF, XGB und GBM auf dem Trainingsdatensatz statistisch signifikant besser als die logistische Regression (Tabelle 4; Abb. 5a). Die Treffsicherheit der logR war mit 81,7% (95% CI [77,3; 85,6]) geringfügig, jedoch statistisch nicht signifikant ($p=0,05$) höher als die NIR mit 78,0%. Dagegen zeigten die anderen ML Modelle RF, GBM und XGB auf dem Trainingsdatensatz statistisch signifikant höhere Treffsicherheiten als die NIR ($p<0,01$; Tabelle 4).

Tabelle 4 Leistung der verschiedenen ML Modelle (Logistische Regression (logR), Gradient Boosting Machine (GBM), Random Forest (RF), Extreme Gradient Boosting (XGB)) auf dem Trainingsdatensatz. Die No-Information-Rate (NIR) betrug in diesem Fall 78,0%. (mod. nach Herbst CS, Schwendicke F, Krois J, Herbst SR. Association between patient-, tooth- and treatment-level factors and root canal treatment failure: A retrospective longitudinal and machine learning study. *J Dent.* 2022; S. 5)

Modell	Treffsicherheit (95% CI)	p-Wert (NIR)	AUC (95% CI)	F1-Score	Sensitivität	Spezifität	Positiver Vorher- sagewert	Negativer Vorher- sagewert
logR	81,7% [77,3; 85,6] ^a	0,05	0,77 [0,71; 0,83] ^a	41,4%	29,5%	96,5%	69,7%	82,9%
GBM	86,8% [82,8; 90,1] ^a	<0,01	0,89 [0,85; 0,93] ^b	67,6%	62,8%	93,5%	73,1%	89,9%
RF	95,5% [92,8; 97,4] ^b	<0,01	0,99 [0,98; 1,00] ^c	88,6%	79,5%	100%	100%	94,5%
XGB	85,4% [81,2; 88,9] ^a	<0,01	0,88 [0,84; 0,92] ^b	58,7%	47,4%	96,0%	77,1%	86,6%

^{a,b,c} die Kombination verschiedener Buchstaben zeigt einen statistisch signifikanten Unterschied ($p<0,05$)

Alle ML Modelle zeigten ähnliche, aber insgesamt limitierte Vorhersagemöglichkeiten auf dem Testdatensatz. Hierbei ergaben sich statistisch keine signifikanten Unterschiede in der Treffsicherheit oder den AUC-Werten zwischen den verschiedenen Modellen (Tabelle 5; Abb. 5b). Im Allgemeinen zeigten die Modelle hohe Spezifitäten (logR: 94,3%; GBM:

87,6%; RF: 90,2%; XGB: 91,7%), jedoch geringe Sensitivitäten (11,6-21,0%). Keines der Modelle konnte auf den ungesehenen Daten eine bessere Treffsicherheit als die korrespondierende NIR ($p > 0,05$) erreichen. Die konsequente Annahme der Mehrheitsklasse („erfolgreiche Behandlung“) war somit akkurater als die Vorhersageergebnisse der jeweiligen Modelle.

Tabelle 5 Leistung der verschiedenen ML Modelle (Logistische Regression (logR), Gradient Boosting Machine (GBM), Random Forest (RF), Extreme Gradient Boosting (XGB)) auf dem Testdatensatz. Die No-Information-Rate (NIR) betrug in diesem Fall 81,8%. (mod. nach Herbst CS, Schwendicke F, Krois J, Herbst SR. Association between patient-, tooth- and treatment-level factors and root canal treatment failure: A retrospective longitudinal and machine learning study. J Dent. 2022; S. 6)

Modell	Treffsicherheit (95% CI)	p-Wert (NIR)	AUC (95% CI)	F1-Score	Sensitivität	Spezifität	Positiver Vorher- sagewert	Negativer Vorher- sagewert
logR	80,5% [74,9; 85,4] ^a	0,73	0,63 [0,53; 0,73] ^a	25,8%	18,6%	94,3%	42,1%	83,9%
GBM	75,4% [69,4; 80,8] ^a	0,99	0,59 [0,50; 0,68] ^a	23,7%	21,0%	87,6%	27,3%	83,3%
RF	75,9% [69,9; 81,1] ^a	0,99	0,59 [0,50; 0,68] ^a	14,9%	11,6%	90,2%	20,8%	82,1%
XGB	78,4% [72,6; 83,5] ^a	0,92	0,60 [0,50; 0,70] ^a	23,9%	18,6%	91,7%	33,3%	83,5%

a,b,c die Kombination verschiedener Buchstaben zeigt einen statistisch signifikanten Unterschied ($p < 0,05$)

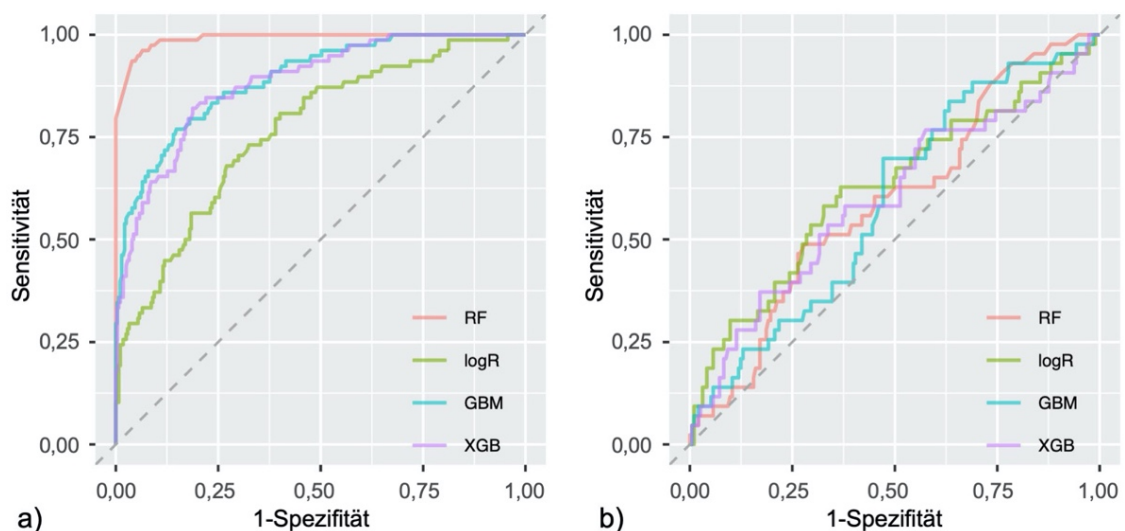


Abbildung 5 ROC-Kurven der Modelle auf dem (a) Trainingsdatensatz (b) Testdatensatz (mod. nach Herbst CS, Schwendicke F, Krois J, Herbst SR. Association between patient-, tooth- and treatment-level factors and root canal treatment failure: A retrospective longitudinal and machine learning study. J Dent. 2022; S. 6)

Das anschließend durchgeführte kalkulierte Faktorgewichtungsranking (Tabelle 6) ergab folgende Ergebnisse: Ein schwerer alveolärer Knochenabbau zeigte die höchste relative Gewichtung, gefolgt von einem PAI-Punktwert ≥ 4 . An dritter Stelle ordnete sich ein moderater alveolärer Knochenabbau ein, gefolgt von einem PAI-Punktwert = 3 auf Platz vier und dem Behandlungstyp „Revision“ auf Platz fünf. Der patientenbezogene Faktor „Rauchen“ belegte den sechsten Platz. Im Mittelfeld des Rankings platzierten sich der Zahntyp „Molar“ sowie kardiovaskuläre Erkrankungen, gefolgt von den behandlungsspezifischen Faktoren „suboptimale Wurzelfüllung“ und dem Fehlen einer höckerüberkuppelnden indirekten Restauration. Am Ende der Rangfolge waren das Patientenalter und die systemische Erkrankung „Diabetes mellitus“ aufgeführt. Alle kalkulierten Rangfolgen der einzelnen ML Modelle sind in Tabelle 6 zusammengefasst.

Table 6 Kalkulierte mittlere Rangwerte (aus GBM, RF, XGB) und daraus resultierende Rangfolgen der Faktoren aus den verschiedenen Modellen mit standardisierter Faktorgewichtung in %. Der wichtigste Faktor des jeweiligen Modells wurde gleich 100% gesetzt, sodass sich die Faktoren mit geringerer Wichtigkeit relativ und anteilig daran orientieren. (mod. nach Herbst CS, Schwendicke F, Krois J, Herbst SR. Association between patient-, tooth- and treatment-level factors and root canal treatment failure: A retrospective longitudinal and machine learning study. J Dent. 2022; S. 6)

Faktor	Rang	Kalkulierte mittlere Rangwerte	Gradient Boosting Ma- chine (GBM)	Random Forest (RF)	Extreme Gra- dient Boosting (XGB)	Logistische Regression (logR)
Schwerer alveolärer Knochenabbau	1	1,0	100%	100%	100%	89,8%
Periapikaler Index ≥ 4	2	2,0	76,7%	81,2%	86,6%	100%
Moderater alveolärer Knochenabbau	3	3,3	54,9%	59,2%	56,9%	64,2%
Periapikaler Index = 3	4	3,7	72,9%	45,6%	37,9%	72,0%
Revision	5	6,7	20,4%	37,6%	20,6%	40,2%
Rauchen	6	7,7	20,9%	34,7%	24,2%	48,2%
Molaren	7	8,0	34,7%	36,4%	25,8%	31,4%
Kardiovaskuläre Erkrankungen	8	8,0	37,4%	34,3%	14,7%	0%
Zahnmedizinstudent*innen	9	9,7	43,1%	30,8%	10,0%	1,1%
Suboptimale Länge der Wurzelfüllung	10	10,3	19,2%	28,1%	14,8	28,0%
Prämolaren	11	10,3	17,9%	21,8%	11,5%	10,4%
Andere systemische Erkrankungen	12	10,7	21,4%	31,7%	10,2%	20,8%
Keine höckerüberkuppelnde indirekte Restauration	13	11	13,7%	37,2%	7,0%	17,2%
Andere immunsupprimi- erende Zustände	14	13	9,3%	15,7%	12,11%	28,9%
Alter > 50 Jahre	15	15	8,1%	21,0%	2,8%	29,8%
Alter 30-50 Jahre	16	15,7	1,8%	12,0%	3,2%	22,0%
Diabetes mellitus	17	17	0%	0%	0%	9,0%

4. Diskussion

4.1 Interpretation der Ergebnisse

Durch die Anwendung einer logR auf dem gesamten Datensatz konnte gezeigt werden, dass die zahnbezogenen Faktoren den stärksten Einfluss auf das Behandlungsergebnis hatten. Insbesondere ein alveolärer Knochenabbau und das Vorhandensein und der Ausprägungsgrad apikaler Läsionen nahmen hierbei Einfluss [63]. Diese Ergebnisse decken sich mit der nachfolgend kalkulierten Rangfolge aus allen ML Modellen und auch mit anderen Forschungsarbeiten. So war vor allem das Vorhandensein einer apikalen Parodontitis eines der häufigsten Ursachen für den Misserfolg einer WKB in weiteren Studien [28, 50, 51]. Auch der zahnbezogene Faktor „Revision“ platzierte sich im oberen Drittel des Rankings und war mit mehr Misserfolgen assoziiert als eine Primärbehandlung [63]. Dieser Zusammenhang wird durch vorangegangene Literatur gestützt [64], die geringere Erfolgsraten bei Revisionszähnen verzeichnete.

Der patientenbezogene Faktor „Rauchen“ war in der logR die fünftstärkste und im Gesamtranking die sechststärkste Variable. Somit zeigten Raucher*innen im Vergleich ein hohes Risiko für einen Misserfolg [63]. In vorhandenen Studien wurde das Rauchen ebenso als negativer prognostischer Faktor beschrieben [46, 65, 66], wobei die Evidenz insgesamt recht gering ist. Zudem konnte in unserer Arbeit gezeigt werden, dass ein kombiniertes Vorliegen verschiedener Risikofaktoren (z.B. Patient*innen mit einer apikalen Läsion, zusätzlichem Knochenabbau sowie einer positiven Raucheranamnese) zu einem höheren Risiko eines Misserfolges führt [63].

In unserer Analyse hatten systemische Erkrankungen keinen statistisch signifikanten Einfluss auf die Prognose einer WKB. Der Faktor „Diabetes mellitus“ zeigte verglichen mit den anderen Faktoren im Ranking sogar die geringste Signifikanz [63]. Andere Studien stehen hierbei teilweise im Kontrast zu unseren Ergebnissen, insbesondere in Bezug auf Diabeteserkrankungen. Einige vorherige Untersuchungen kamen zu dem Schluss, dass mit DM vorerkrankte Patient*innen zum einen eine höhere Prävalenz von periapikalen Läsionen [19-21] aufweisen und zum anderen mit einem höheren Risiko für einen Misserfolg der WKB assoziiert sind [22-24, 28]. Hervorzuheben ist, dass Patient*innen mit DM in unserer Studie häufiger einen PAI-Punktwert von 5 ($p < 0,05$) und zusätzlich häufiger

kardiovaskuläre Erkrankungen ($p < 0,001$) aufwiesen als Patient*innen ohne Diabeteserkrankung, sodass diese Faktoren einzeln oder in Kombination womöglich einen stärkeren Einfluss nahmen als der Diabetes an sich. Zudem waren die meisten Patient*innen mit systemischen Erkrankungen wie DM medikamentös eingestellt, wodurch die Effekte der jeweiligen Erkrankung auf eine WKB abgeschwächt werden können. Zuletzt war die Anzahl an Diabetespatient*innen in unserer Studie im Verhältnis gering ($n = 33$), was möglicherweise zu einer verzerrten Darstellung führen kann.

Auch die meisten behandlungsspezifischen Faktoren hatten in unserer Analyse statistisch keinen signifikanten Einfluss auf den Ausgang einer WKB. Dementsprechend ordneten sich die Behandlungsvariablen auf dem unteren Ende der Rankingskala an [63]. Diese Beobachtung steht im Kontrast zu den Ergebnissen anderer Autor*innen. Vor allem eine indirekte höckerüberkuppelnde Versorgung war in der Literatur häufig mit einer verbesserten Langzeitprognose assoziiert [23]. Jedoch muss betont werden, dass der behandlungsbezogene Faktor „suboptimale Wurzelfüllung“ in unserer Studie eine relative Wichtigkeit von 28% aufweist [63] und somit der Einfluss auf die Vorhersage nicht zu vernachlässigen ist. Die relative Wichtigkeit bezieht sich hierbei auf einen PAI-Punktwert von 4, der als 100%-Referenz bei der Angabe von Wichtigkeiten gewertet wurde.

Interessanterweise zeigte sich im Vergleich zwischen approbierten Zahnärzt*innen und Zahnmedizinstudierenden kein statistisch signifikanter Unterschied in der Erfolgsrate der durchgeführten Wurzelkanalbehandlungen [63]. Dies lässt sich vor allem darauf zurückführen, dass Patientenfälle im Vorhinein basierend auf dem erwarteten Schwierigkeitsgrad selektiert wurden. Zahnmedizinstudierende erhielten in der Regel weniger komplexe Behandlungsfälle als approbierte Zahnärzt*innen. Dies spiegelt sich unter anderem in der Verteilung von Zahngruppen wider: Approbierte Zahnärzt*innen behandelten deutlich mehr Molaren als Student*innen ($p < 0,001$). Zudem findet die Studentenbehandlung unter Supervision von Zahnärzt*innen statt, sodass die einzelnen Behandlungsschritte kontrolliert und somit potenzielle Behandlungsfehler möglicherweise vermieden werden können.

4.2 Evaluation der ML Modelle und Bewertung von Vorhersagemöglichkeiten

Der Erfolg oder Misserfolg einer WKB hängt - wie dargestellt - von verschiedenen patienten- und zahnbezogenen sowie behandlungsspezifischen Faktoren ab. Für die Behandlungsplanung als auch Patientenaufklärung ist es von entscheidender Bedeutung, welche dieser Faktoren in der Gewichtung den stärksten Einfluss auf die Prognose einer WKB nehmen können. Mithilfe von Assoziationsanalysen können Aussagen über die Gewichtung von einzelnen Faktoren getroffen werden. Für die Prognoseeinschätzung ist es jedoch wichtig, multiple Faktoren in Relation zu setzen, um die Präzision der Aussage zu erhöhen. Durch die Anwendung komplexer ML Modelle wäre es denkbar, reliable Vorhersagen für den Ausgang einer Behandlung unter Einbeziehung vielfältiger Variablen treffen zu können. Das Ziel dieser Studie war es dementsprechend, ML Modelle für die Vorhersage des Behandlungsergebnisses von Wurzelkanalbehandlungen zu entwickeln.

Hierbei sollten die Vorhersagemodelle unter anderem eine hohe Genauigkeit und Reliabilität aufweisen, vor allem bezogen auf ungesehene Daten. Die Modelle in unserer Studie wurden anhand verschiedener Metriken auf einem ungesehenen Datensatz evaluiert. Dabei galt vorab die Annahme, dass die Modelle auf ungesehenen Daten im Vergleich zum Trainingsdatensatz potenziell weniger aussagekräftige Ergebnisse generieren werden. Die logistische Regression zeigte auf dem Trainingsdatensatz das geringste Leistungspotenzial, sodass die Treffsicherheit geringer ausfiel als die No-Information-Rate [63]. Das heißt, die konsequente Annahme der Mehrheitsklasse („Erfolg“) war akkurater als die Klassifizierung durch das logR Modell. Komplexere Modelle wie RF, GBM und XGB neigten zu einer sogenannten Überanpassung (engl. Overfitting) im Trainingsdatensatz mit einem AUC-Wert von bis zu 0,99 (95% CI [0,98; 1,00]) [63]. Überanpassung beschreibt eine Spezifizierung des Algorithmus an die prozessierten Daten im Sinne eines Auswendiglernens. Das eigentliche zugrundeliegende Muster wird dabei nicht ausreichend gelernt, was zu unbrauchbaren Vorhersagen auf ungesehenen Datensätzen führt [67]. Es ist anzunehmen, dass die herausgearbeiteten Korrelationen daher nur auf den zuvor gesehenen Daten valide sind aufgrund des insgesamt geringen Leistungspotenzials der Modelle auf dem ungesehenen Testdatensatz. Die von den Algorithmen identifizierten Muster auf den Trainingsdatensätzen scheinen entsprechend auch nur für die Trainingsdaten zu gelten. Durch die beschriebene Überanpassung haben die Algorithmen die eigentlichen Muster nicht identifiziert und somit waren auch komplexere Modelle

nicht in der Lage, sinnvolle Vorhersagen zum Ergebnis der WKB zu liefern oder einfachere Modelle wie die logR in Bezug auf die Qualität der Vorhersagen zu übertreffen.

Eine weitere Ursache für die limitierten Vorhersagemöglichkeiten stellt ein vorhandenes Ungleichgewicht der resultierenden Gruppen Erfolg/Misserfolg dar. So gab es in unserem Datensatz deutlich mehr Erfolge als Misserfolge, wobei gerade die Vorhersage von Misserfolgen die größte klinische Relevanz hat. Dieser Effekt könnte kompensiert werden, indem die Stichprobengröße insgesamt erhöht und der Trainingsdatensatz entsprechend erweitert wird mit einer erhöhten Anzahl an Minderheitsklassen (Misserfolge). Eine weitere Möglichkeit bestände darin, erfolgreiche Fälle aus dem Datensatz zu entnehmen, sodass sich das Erfolg-Misserfolg-Verhältnis annäherungsweise zu einem Gleichgewicht entwickelt. Wenn die Stichprobengröße nicht erhöht werden kann, bietet andererseits das sogenannte synthetische Oversampling (SMOTE) [68] eine Möglichkeit, neue Einträge synthetisch zu generieren, ohne dabei nur die vorhandenen Werte zu duplizieren. In unserer Studie war es eines der Ziele, ein Faktorgewichtungsranking zu erstellen, das durch synthetisch generierte Patientenfälle verzerrt worden wäre. Die Anwendung von Oversampling-Methoden wäre mit dem Ziel eines reinen Vorhersagemodells sinnvoll gewesen, jedoch musste hiervon in Kombination mit einem Faktorgewichtungsranking abgesehen werden.

4.3 Stärken und Schwächen der Studie

Diese Studie weist sowohl einige Stärken als auch Limitationen auf. Zunächst wurden vier ML Modelle auf einem adäquat dimensionierten Datensatz aus zahn- und patientenbezogenen sowie behandlungsassoziierten Variablen angewandt, woraus eine statistisch komplexe Methodik resultiert.

Als Limitation kann formuliert werden, dass die Bewertung der radiologischen und klinischen Daten durch nur eine Autorin (CH) nach Kalibrierung von 50 Fällen mit einem anderen Autor (SH) erfolgte. Jedoch zeigte die Interrater-Reliabilität beider Bewerter*innen mittels Cohens Kappa Gleichung eine hohe Übereinstimmung (0,89; 95% CI [0,79; 0,99]) [63]. Zudem konnte durch zuvor strikt festgelegte Kriterien auch die nachfolgende Analyse der Patientenfälle zielgerichtet und reproduzierbar erfolgen.

Des Weiteren enthielten unsere Daten basierend auf dem Studiendesign Fälle, die entweder von approbierten Zahnärzt*innen oder Zahnmedizinstudent*innen behandelt wurden. Somit konnte evaluiert werden, ob die Erfahrung der Behandler*innen die Prognose einer WKB beeinflussen kann. Jedoch waren die Behandlungsprotokolle hierbei nicht identisch, wodurch eine erhöhte Heterogenität im Datensatz resultierte. Erwähnenswert ist, dass die Art der Behandler*innen in unserer Analyse nur einen geringen Einfluss zeigte [63]. In der Konsequenz kann formuliert werden, dass moderate Unterschiede im Behandlungsprotokoll den Ausgang einer WKB in unserer Studie nicht signifikant beeinflussten.

Zuletzt bestand unser Datensatz aus 79,5% erfolgreichen Behandlungen (Mehrheitsklasse) und 20,5% Misserfolgen (Minderheitsklasse) und ist per definitionem unbalanciert. Ein Datensatz wird dann als unausgewogen bezeichnet, wenn eine Klasse des Datensatzes gegenüber der anderen Klasse deutlich über- beziehungsweise unterrepräsentiert ist [68]. Statistische Klassifikationsmodelle können Muster, die der unterrepräsentierten Klasse zugrunde liegen nur schwer erlernen, weshalb primär die Muster der Mehrheitsklasse erkannt und infolgedessen für die Prädiktion verwendet werden. Dementsprechend kann eine Verzerrung zugunsten der Mehrheitsklasse vorliegen [69]. Unsere angewandten Modelle zeigten auf dem unausgewogenen Datensatz entsprechend kein ausreichendes Leistungspotenzial [63]. Wir versuchten dies zu berücksichtigen, indem wir die Modelle anhand der ROC-Werte anstelle der Treffsicherheit optimierten, da die ROC-Werte auf unbalancierten Daten validere Ergebnisse und eine höhere Aussagekraft liefern können [68]. Die Treffsicherheit eignet sich umgekehrt für balancierte Datensätze mit gleichen Gewichtungen der Mehrheits- und Minderheitsklassen. In der Konsequenz kann formuliert werden, dass die Wahl der entsprechenden Leistungsmetriken je nach Beschaffenheit des Datensatzes gewählt werden sollte, um die Resultate aus den Modellen insgesamt zu verbessern.

4.4 Ausblick für die Praxis und zukünftige Forschung

Anwendungen von KI und ML finden vermehrt praktische Anwendung im Gesundheitswesen und somit auch in der Zahnmedizin. Die ML Modelle in unserer Studie konnten

insgesamt keine valideren Vorhersagen treffen als die konsequente Annahme der Mehrheitsklasse. Dies impliziert jedoch keine fehlende Relevanz für die Klinik, sondern vielmehr eine Notwendigkeit für weitere Forschungsarbeiten mit einer erhöhten Anzahl an zu prozessierenden Daten.

Nützliche Vorhersagemodelle könnten eingesetzt werden, um Patient*innen bereits vor Behandlungsbeginn eine objektive und trotzdem patientenindividuelle Orientierung für die Langzeitprognose der WKB aufzuzeigen. Ebenso könnten Vorhersagen angewandt werden, um klinische Protokolle bereits a priori anzupassen und den Therapieerfolg individuell zu optimieren. Ebenso ist es denkbar, auf der Basis von Erfolgsvorhersagen grundsätzliche Therapieentscheide (Therapie ja/nein) zu treffen oder beispielsweise Zuschussentscheidungen von Versicherungen entlang des zu erwartenden Therapieerfolgs zu treffen. Die hierfür eingesetzten Modelle sollten jedoch robust und an heterogenen Datensätzen auf Generalisierbarkeit getestet worden sein.

Zu beachten gilt jedoch, dass Komplikationen wie eine Zahnfraktur oder exazerbierende Entzündungsgeschehen durch rekurrierende Infektionen selbst bei günstigster Prognose weiterhin möglich sind. Dies erfordert eine umfassende Patientenaufklärung, da die Prognoseeinschätzung nicht als absolut zu verstehen sein wird.

5. Schlussfolgerungen

Unter Einbezug der Limitationen unserer Studie kann abschließend geschlussfolgert werden, dass die Prognose einer Wurzelkanalbehandlung im Wesentlichen von zahnbezogenen Faktoren abhängig ist. Vorhersagen über den Erfolg oder Misserfolg einer WKB waren nur limitiert möglich. Auch komplexe ML Modelle konnten keine nützlichen Vorhersagen treffen.

Literaturverzeichnis

- [1] Aminoshariae A, Kulild J, Nagendrababu V. Artificial Intelligence in Endodontics: Current Applications and Future Directions. *Journal of endodontics*. 2021;47(9):1352-7.
- [2] Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. *J Family Med Prim Care*. 2019;8(7):2328-31.
- [3] Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. *Minim Invasive Ther Allied Technol*. 2019;28(2):73-81.
- [4] McCarthy J. What is artificial intelligence. 2007.
- [5] Hassabis D, Kumaran D, Summerfield C, Botvinick M. Neuroscience-Inspired Artificial Intelligence. *Neuron*. 2017;95(2):245-58.
- [6] Nasteski V. An overview of the supervised machine learning methods. *HORIZONSB*. 2017;4:51-62.
- [7] Rodrigues JA, Krois J, Schwendicke F. Demystifying artificial intelligence and deep learning in dentistry. *Brazilian oral research*. 2021;35:e094.
- [8] Nguyen CN, Zeigermann O. *Machine Learning kurz & gut*. O'Reilly 2 Auflage. 2021:3-6.
- [9] Kingsford C, Salzberg SL. What are decision trees? *Nat Biotechnol*. 2008;26(9):1011-3.
- [10] MacEachern SJ, Forkert ND. Machine learning for precision medicine. *Genome*. 2021;64(4):416-25.
- [11] Soenksen LR, Kassis T, Conover ST, Marti-Fuster B, Birkenfeld JS, Tucker-Schwartz J, Naseem A, Stavert RR, Kim CC, Senna MM, Aviles-Izquierdo J, Collins JJ, Barzilay R, Gray ML. Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images. *Sci Transl Med*. 2021;13(581).
- [12] Islam MM, Yang HC, Poly TN, Jian WS, Jack Li YC. Deep learning algorithms for detection of diabetic retinopathy in retinal fundus photographs: A systematic review and meta-analysis. *Comput Methods Programs Biomed*. 2020;191:105320.
- [13] Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2(4):230-43.
- [14] Ekert T, Krois J, Meinhold L, Elhennawy K, Emara R, Golla T, Schwendicke F. Deep Learning for the Radiographic Detection of Apical Lesions. *Journal of endodontics*. 2019;45(7):917-22 e5.
- [15] Krois J, Ekert T, Meinhold L, Golla T, Kharbot B, Wittemeier A, Dorfer C, Schwendicke F. Deep Learning for the Radiographic Detection of Periodontal Bone Loss. *Scientific reports*. 2019;9(1):8495.
- [16] Cantu AG, Gehrung S, Krois J, Chaurasia A, Rossi JG, Gaudin R, Elhennawy K, Schwendicke F. Detecting caries lesions of different radiographic extension on bitewings using deep learning. *J Dent*. 2020;100:103425.
- [17] Ha SR, Park HS, Kim EH, Kim HK, Yang JY, Heo J, Yeo IL. A pilot study using machine learning methods about factors influencing prognosis of dental implants. *The journal of advanced prosthodontics*. 2018;10(6):395-400.
- [18] El-Ma'aita AM, Qualtrough AJ, Darcey J, Hunter MJ. Rethinking root canal treatment: understanding the difference between preventing and treating endodontic infection. *British dental journal*. 2015;219(1):13-6.
- [19] Gupta A, Aggarwal V, Mehta N, Abraham D, Singh A. Diabetes mellitus and the healing of periapical lesions in root filled teeth: a systematic review and meta-analysis. *International endodontic journal*. 2020;53(11):1472-84.

- [20] Jakovljevic A, Duncan HF. Diabetes Mellitus May Negatively Influence the Outcome of Conventional Nonsurgical Root Canal Treatment. *The journal of evidence-based dental practice*. 2020;20(3):101467.
- [21] Segura-Egea JJ, Martín-González J, Cabanillas-Balsera D, Fouad AF, Velasco-Ortega E, López-López J. Association between diabetes and the prevalence of radiolucent periapical lesions in root-filled teeth: systematic review and meta-analysis. *Clinical oral investigations*. 2016;20(6):1133-41.
- [22] Cabanillas-Balsera D, Martín-González J, Montero-Miralles P, Sánchez-Domínguez B, Jiménez-Sánchez MC, Segura-Egea JJ. Association between diabetes and nonretention of root filled teeth: a systematic review and meta-analysis. *International endodontic journal*. 2019;52(3):297-306.
- [23] Ng YL, Mann V, Gulabivala K. A prospective study of the factors affecting outcomes of non-surgical root canal treatment: part 2: tooth survival. *International endodontic journal*. 2011;44(7):610-25.
- [24] Arya S, Duhan J, Tewari S, Sangwan P, Ghalaut V, Aggarwal S. Healing of Apical Periodontitis after Nonsurgical Treatment in Patients with Type 2 Diabetes. *Journal of endodontics*. 2017;43(10):1623-7.
- [25] Rudranaik S, Nayak M, Babshet M. Periapical healing outcome following single visit endodontic treatment in patients with type 2 diabetes mellitus. *Journal of clinical and experimental dentistry*. 2016;8(5):e498-e504.
- [26] Fouad AF. Diabetes mellitus as a modulating factor of endodontic infections. *J Dent Educ*. 2003;67(4):459-67.
- [27] Fouad AF, Burleson J. The effect of diabetes mellitus on endodontic treatment outcome: data from an electronic patient record. *Journal of the American Dental Association*. 2003;134(1):43-51; quiz 117-8.
- [28] Laukkanen E, Vehkalahti MM, Kotiranta AK. Impact of systemic diseases and tooth-based factors on outcome of root canal treatment. *International endodontic journal*. 2019;52(10):1417-26.
- [29] Lima SM, Grisi DC, Kogawa EM, Franco OL, Peixoto VC, Gonçalves-Júnior JF, Arruda MP, Rezende TM. Diabetes mellitus and inflammatory pulpal and periapical disease: a review. *International endodontic journal*. 2013;46(8):700-9.
- [30] Berbudi A, Rahmadika N, Tjahjadi AI, Ruslami R. Type 2 Diabetes and its Impact on the Immune System. *Curr Diabetes Rev*. 2020;16(5):442-9.
- [31] Ríos-Osorio N, Muñoz-Alvear HD, Montoya Cañón S, Restrepo-Mendez S, Aguilera-Rojas SE, Jiménez-Peña O, García-Perdomo HA. Association between type 2 diabetes mellitus and the evolution of endodontic pathology. *Quintessence international*. 2020;51(2):100-7.
- [32] Aloutaibi YA, Alkarim AS, Qumri EM, Almansour LA, Alghamdi FT. Chronic Endodontic Infections and Cardiovascular Diseases: Does the Evidence Support an Independent Association? *Cureus*. 2021;13(11):e19864.
- [33] Aminoshariae A, Kulild J, Fouad AF. The impact of cardiovascular disease and endodontic outcome: a systematic review of longitudinal studies. *Clinical oral investigations*. 2020;24(11):3813-9.
- [34] Wang CH, Chueh LH, Chen SC, Feng YC, Hsiao CK, Chiang CP. Impact of diabetes mellitus, hypertension, and coronary artery disease on tooth extraction after nonsurgical endodontic treatment. *Journal of endodontics*. 2011;37(1):1-5.
- [35] Virtanen E, Nurmi T, Soder PO, Airila-Mansson S, Soder B, Meurman JH. Apical periodontitis associates with cardiovascular diseases: a cross-sectional study from Sweden. *BMC oral health*. 2017;17(1):107.

- [36] Poornima L, Ravishankar P, Abbott PV, Subbiya A, PradeepKumar AR. Impact of root canal treatment on high-sensitivity C-reactive protein levels in systemically healthy adults with apical periodontitis - a preliminary prospective, longitudinal interventional study. *International endodontic journal*. 2020.
- [37] Cowan LT, Lakshminarayan K, Lutsey PL, Beck J, Offenbacher S, Pankow JS. Endodontic therapy and incident cardiovascular disease: The Atherosclerosis Risk in Communities (ARIC) study. *J Public Health Dent*. 2020;80(1):79-91.
- [38] Jimenez-Sanchez MC, Cabanillas-Balsera D, Areal-Quecuty V, Velasco-Ortega E, Martin-Gonzalez J, Segura-Egea JJ. Cardiovascular diseases and apical periodontitis: association not always implies causality. *Medicina oral, patologia oral y cirugia bucal*. 2020;25(5):e652-e9.
- [39] Kean J. The effects of smoking on the wound healing process. *J Wound Care*. 2010;19(1):5-8.
- [40] Liu D, Zhu L, Yang C. The effect of preoperative smoking and smoke cessation on wound healing and infection in post-surgery subjects: A meta-analysis. *Int Wound J*. 2022.
- [41] Ijzerman RG, Serne EH, van Weissenbruch MM, de Jongh RT, Stehouwer CD. Cigarette smoking is associated with an acute impairment of microvascular function in humans. *Clin Sci (Lond)*. 2003;104(3):247-52.
- [42] Wang Y, Liu M, Zhu Y, Cheng K, Da W, Liu B, Li F. Identifying the tobacco related free radicals by UPCC-QTOF-MS with radical trapping method in mainstream cigarette smoke. *Talanta*. 2016;160:106-12.
- [43] Faarvang AS, Rordam Preil SA, Nielsen PS, Beck HC, Kristensen LP, Rasmussen LM. Smoking is associated with lower amounts of arterial type I collagen and decorin. *Atherosclerosis*. 2016;247:201-6.
- [44] Lappin DF, Sherrabeh S, Jenkins WM, Macpherson LM. Effect of smoking on serum RANKL and OPG in sex, age and clinically matched supportive-therapy periodontitis patients. *Journal of clinical periodontology*. 2007;34(4):271-7.
- [45] Johannsen A, Susin C, Gustafsson A. Smoking and inflammation: evidence for a synergistic role in chronic disease. *Periodontology 2000*. 2014;64(1):111-26.
- [46] Cabanillas-Balsera D, Segura-Egea JJ, Bermudo-Fuenmayor M, Martín-González J, Jiménez-Sánchez MC, Areal-Quecuty V, Sánchez-Domínguez B, Montero-Miralles P, Velasco-Ortega E. Smoking and Radiolucent Periapical Lesions in Root Filled Teeth: Systematic Review and Meta-Analysis. *J Clin Med*. 2020;9(11).
- [47] Cabanillas-Balsera D, Segura-Egea JJ, Jiménez-Sánchez MC, Areal-Quecuty V, Sánchez-Domínguez B, Montero-Miralles P, Saúco-Márquez JJ, Martín-González J. Cigarette Smoking and Root Filled Teeth Extraction: Systematic Review and Meta-Analysis. *J Clin Med*. 2020;9(10).
- [48] Mahmood AA, AbdulAzeez AR, Hussein HM. The Effect of Smoking Habit on Apical Status of Adequate Endodontically Treated Teeth with and Without Periodontal Involvement. *Clin Cosmet Investig Dent*. 2019;11:419-28.
- [49] Mindiola MJ, Mickel AK, Sami C, Jones JJ, Lalumandier JA, Nelson SS. Endodontic treatment in an American Indian population: a 10-year retrospective study. *Journal of endodontics*. 2006;32(9):828-32.
- [50] Jahreis M, Soliman S, Schubert A, Connert T, Schlagenhaut U, Krastl G, Krug R. Outcome of non-surgical root canal treatment related to periodontitis and chronic disease medication among adults in age group of 60 years or more. *Gerodontology*. 2019;36(3):267-75.
- [51] Maslamani M, Khalaf M, Mitra AK. Association of Quality of Coronal Filling with the Outcome of Endodontic Treatment: A Follow-up Study. *Dent J (Basel)*. 2017;5(1).

- [52] Wong J, Manoil D, Nasman P, Belibasakis GN, Neelakantan P. Microbiological Aspects of Root Canal Infections and Disinfection Strategies: An Update Review on the Current Knowledge and Challenges. *Front Oral Health*. 2021;2:672887.
- [53] Schmidt JC, Walter C, Amato M, Weiger R. Treatment of periodontal-endodontic lesions--a systematic review. *Journal of clinical periodontology*. 2014;41(8):779-90.
- [54] Jansson L, Ehnevid H, Lindskog S, Blomlof L. The influence of endodontic infection on progression of marginal bone loss in periodontitis. *Journal of clinical periodontology*. 1995;22(10):729-34.
- [55] Jansson LE, Ehnevid H. The influence of endodontic infection on periodontal status in mandibular molars. *Journal of periodontology*. 1998;69(12):1392-6.
- [56] Srivastava PK, Nagpal A, Setya G, Kumar S, Chaudhary A, Dhanker K. Assessment of Coronal Leakage of Temporary Restorations in Root Canal-treated Teeth: An in vitro Study. *The journal of contemporary dental practice*. 2017;18(2):126-30.
- [57] Doyle SL, Hodges JS, Pesun IJ, Baisden MK, Bowles WR. Factors affecting outcomes for single-tooth implants and endodontic restorations. *Journal of endodontics*. 2007;33(4):399-402.
- [58] Almohareb R, Ahlquist M, Englund GS. The association between root canal filling and treatment outcome. *Swedish dental journal*. 2016;40(2):133-42.
- [59] Cuschieri S. The STROBE guidelines. *Saudi J Anaesth*. 2019;13(Suppl 1):S31-S4.
- [60] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine*. 2015;162(1):W1-73.
- [61] Bujang MA, Sa'at N, Sidik T, Joo LC. Sample Size Guidelines for Logistic Regression from Observational Studies with Large Population: Emphasis on the Accuracy Between Statistics and Parameters Based on Real Life Clinical Data. *Malays J Med Sci*. 2018;25(4):122-30.
- [62] Orstavik D, Kerekes K, Eriksen HM. The periapical index: a scoring system for radiographic assessment of apical periodontitis. *Endodontics & dental traumatology*. 1986;2(1):20-34.
- [63] Herbst CS, Schwendicke F, Krois J, Herbst SR. Association between patient-, tooth- and treatment-level factors and root canal treatment failure: A retrospective longitudinal and machine learning study. *J Dent*. 2022;117:103937.
- [64] Imura N, Pinheiro ET, Gomes BP, Zaia AA, Ferraz CC, Souza-Filho FJ. The outcome of endodontic treatment: a retrospective study of 2000 cases performed by a specialist. *Journal of endodontics*. 2007;33(11):1278-82.
- [65] Duncan HF, Pitt Ford TR. The potential association between smoking and endodontic disease. *International endodontic journal*. 2006;39(11):843-54.
- [66] Kirkevang LL, Wenzel A. Risk indicators for apical periodontitis. *Community Dent Oral Epidemiol*. 2003;31(1):59-67.
- [67] Kernbach JM, Staartjes VE. Foundations of Machine Learning-Based Clinical Prediction Modeling: Part II-Generalization and Overfitting. *Acta Neurochir Suppl*. 2022;134:15-21.
- [68] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Int Res*. 2002;16(1):321-57.
- [69] Stoll A. Supervised Machine Learning mit Nutzergenerierten Inhalten: Oversampling für nicht balancierte Trainingsdaten. *Publizistik*. 2020;65(2):233-51.

Eidesstattliche Versicherung

„Ich, Chantal Sophie Herbst, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Eine Machine Learning Analyse von patienten-, zahn- und behandlungsspezifischen Faktoren in Bezug auf den Therapieerfolg einer Wurzelkanalbehandlung“ (Engl.: „A machine learning study analyzing the association between patient-, tooth- and treatment-level factors on the outcome of root canal therapy“) selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.org) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum

Unterschrift

Anteilerklärung an den erfolgten Publikationen

Chantal Sophie Herbst hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1: Chantal S Herbst, Falk Schwendicke, Joachim Krois, Sascha R Herbst; „Association between patient-, tooth- and treatment-level factors and root canal treatment failure: A retrospective longitudinal and machine learning study“; Journal of Dentistry, 2022; IF: 4,379

Beitrag im Einzelnen:

Ich habe die Studie von der Konzeptionierung bis zur statistischen Umsetzung und Ausarbeitung in Form einer Publikation und vorliegender Dissertation begleitet.

Zunächst habe ich mich umfassend in den aktuellen Stand der Literatur eingelesen hinsichtlich Prinzipien der künstlichen Intelligenz und verschiedener Einflussfaktoren der Prognose einer Wurzelkanalbehandlung. Anschließend wurden die zu untersuchenden Faktoren gemeinsam mit den Co-Autoren (Falk Schwendicke, FS; Joachim Krois, JK; Sascha Herbst, SH) ausgewählt und tabellarisch von mir dargestellt. Alle Patientendaten (Anamnese, Akten) und Röntgenbilder wurden von mir anhand zuvor festgelegter Inklusionskriterien in Abstimmung mit den Co-Autoren (FS, JK, SH) ausgewertet. Die Bewertung der Röntgenbilder fand nach einem von mir mitentwickelten standardisierten Protokoll statt.

In Zusammenarbeit mit Datenwissenschaftler*innen habe ich die ML Modelle zur statistischen Analyse ausgewählt. Die Ergebnisse der Statistik wurden von mir interpretiert und in einen klinischen Kontext gestellt. Anhand der statistischen Daten schrieb ich die zugrundeliegende Publikation dieser Arbeit. Ich erstellte alle vorhandenen Abbildungen (Abb. 1, 2, 3) und Tabellen (Tab. 1, 2, 3, 4, 5) der Publikation selbstständig. Folgende Abbildungen und Tabellen aus der Publikation fanden in modifizierter Form auch Verwendung im Manteltext: Abb. 3, 4, 5 (entspricht Abb. 1, 2, 3 der Publikation) und Tab. 3, 4, 5, 6 (entspricht Tab. 1, 2, 3, 4 der Publikation).

Unterschrift, Datum und Stempel des/der erstbetreuenden Hochschullehrers/in

Unterschrift des Doktoranden/der Doktorandin

Auszug aus der Journal Summary List

Journal Data Filtered By: **Selected JCR Year: 2020** Selected Editions: SCIE,SSCI
 Selected Categories: **"DENTISTRY, ORAL SURGERY and MEDICINE"**
 Selected Category Scheme: WoS
Gesamtanzahl: 91 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	JOURNAL OF CLINICAL PERIODONTOLOGY	19,562	8.728	0.013470
2	PERIODONTOLOGY 2000	6,873	7.589	0.005670
3	JOURNAL OF PERIODONTOLOGY	20,398	6.993	0.010000
4	International Journal of Oral Science	2,890	6.344	0.002260
5	JOURNAL OF DENTAL RESEARCH	26,197	6.116	0.016830
6	CLINICAL ORAL IMPLANTS RESEARCH	15,673	5.977	0.012460
7	ORAL ONCOLOGY	13,860	5.337	0.014090
8	DENTAL MATERIALS	17,990	5.304	0.011490
9	Journal of Evidence-Based Dental Practice	899	5.267	0.001050
10	INTERNATIONAL ENDODONTIC JOURNAL	10,290	5.264	0.006960
11	Japanese Dental Science Review	618	5.093	0.000650
12	Journal of Prosthodontic Research	1,950	4.642	0.002870
13	JOURNAL OF PERIODONTAL RESEARCH	5,801	4.419	0.004260
14	JOURNAL OF DENTISTRY	12,037	4.379	0.010250
15	JOURNAL OF ORAL PATHOLOGY & MEDICINE	6,375	4.253	0.004060
16	JOURNAL OF ENDODONTICS	22,265	4.171	0.010770
17	CARIES RESEARCH	5,362	4.056	0.002940
18	Clinical Implant Dentistry and Related Research	4,918	3.932	0.006570
19	JOURNAL OF ORAL REHABILITATION	7,769	3.837	0.004710
20	JOURNAL OF THE AMERICAN DENTAL ASSOCIATION	9,187	3.634	0.005040

Druckexemplar der ausgewählten Publikation

Quellenangabe:

Chantal S Herbst, Falk Schwendicke, Joachim Krois, Sascha R Herbst

„Association between patient-, tooth- and treatment-level factors and root canal treatment failure: A retrospective longitudinal and machine learning study.” J Dent 2022

DOI Verlinkung:

<https://doi.org/10.1016/j.jdent.2021.103937>

Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Komplette Publikationsliste

„Association between patient-, tooth- and treatment-level factors and root canal treatment failure: A retrospective longitudinal and machine learning study.”

Chantal S Herbst, Falk Schwendicke, Joachim Krois, Sascha R Herbst

Journal of Dentistry, 2022; IF: 4,379; DOI: 10.1016/j.jdent.2021.103937

Danksagung

Zunächst möchte ich mich bei meinem Doktorvater Herrn Univ.-Prof. Dr. Falk Schwendicke bedanken für die Möglichkeit dieser Promotionsarbeit und seine beispiellose Expertise als auch Förderung während des gesamten Projektes.

Außerdem danke ich Herrn Dr. Joachim Krois für die statistische Beratung und die vielen Gespräche für ein tieferes Verständnis des maschinellen Lernens.

Insbesondere möchte ich meinem Ehemann Herrn Dr. Sascha Herbst danken für seine Geduld und Unterstützung bei der Umsetzung der vorangegangenen Publikation und letztlich dieser Dissertation. Lieber Sascha, du inspirierst mich täglich mit deiner niemals endenden Wissbegierde und Motivation, Großes zu erreichen.

Auch meiner Familie möchte ich einen besonderen Dank aussprechen. Liebe Mama, lieber Papa, ihr habt mich auf allen Wegen meines Lebens bedingungslos unterstützt und mir Rückhalt gegeben, dafür möchte ich euch von Herzen danken.