

Aus der Poliklinik für Zahnärztliche Prothetik und Werkstoffkunde
der Universitätsmedizin der Johannes Gutenberg-Universität Mainz

Erstellung einer Datensatzgrundlage für eine automatisierte KI-gestützte
Röntgenidentifizierung eines Original- und eines Replika-Zahnimplantates: Weisen
menschlicher Experte und künstliche Intelligenz gleiche Resultate in der Diagnostik
auf?

Inauguraldissertation
zur Erlangung des Doktorgrades der
Zahnmedizin
der Universitätsmedizin
der Johannes Gutenberg-Universität Mainz

Vorgelegt von

Mark Kian Bremer
aus Wiesbaden

Mainz, 2024

Wissenschaftlicher Vorstand: Univ.-Prof. Dr. Hansjörg Schild

1. Gutachter:

2. Gutachter:

Tag der Promotion: 16.12.2024

Meiner Familie

Inhaltsverzeichnis

Abkürzungsverzeichnis	I
Abbildungsverzeichnis	II
Tabellenverzeichnis	III
1 Einleitung	1
2 Literaturdiskussion	3
2.1 Künstliche Intelligenz.....	3
2.1.1 Grundlagen von künstlicher Intelligenz und maschinellem Lernen	3
2.1.2 Technologien und Algorithmen für die automatisierte Implantatidentifizierung.....	6
2.1.2.1 CNN – Convolutional Neural Network.....	6
2.1.2.2 MLP – Multilayer Perceptrons	11
2.1.2.3 PNN – Probabilistic Neural Network	11
2.1.2.4 Random Forest	12
2.1.3 Aktuelle Studienlage zu KI-gestützter automatisierter Implantatidentifizierung anhand von Röntgenaufnahmen	14
3 Material und Methoden	23
3.3 Datenerhebung.....	26
3.3.1 Röntgenbilder	26
3.4 Programmierung der künstlichen Intelligenz	31
3.5 Studienablauf im Detail	32
3.5.1 Studienteilnehmer	32
3.5.2 Studiendurchführung.....	32
4 Ergebnisse	34
4.1 Ergebnisse des Künstlichen Neuronalen Netzwerks.....	34
4.2 Ergebnisse der Studienteilnehmer	36
5 Diskussion	38
5.1 Ergebnisse im Kontext der Hypothese	38

5.2 Relevanz der inkludierten Implantate	39
5.3 Diskussion der Auswahl der Population	42
5.4 Kritische Betrachtung der Materialien und Methoden	45
5.5 Ergebnisse der Arbeit im Kontext der aktuellen Literatur	48
5.5 Ausblick	50
6 Zusammenfassung	52
7 Literaturverzeichnis	53
8 Anhang	61
8.1 Beispielbilder zur Vorbereitung der Befragung für die menschlichen Experten	61
8.1.1 Original-Implantat mit inserierter Schraube	61
8.1.2 Original-Implantat ohne inserierte Schraube	61
8.1.3 Replikat-Implantat mit inserierter Schraube	62
8.1.4 Replikat-Implantat ohne inserierte Schraube	62
9 Danksagung	63
10 Tabellarischer Lebenslauf	64

Abkürzungsverzeichnis

CNN	<i>convolutional neural network</i>
ConvNeXt	<i>convolutional networks with external transformations</i>
DL	<i>deep learning</i>
DVT	Digitale Volumentomographie
FC	<i>fully connected layer</i>
KI	Künstliche Intelligenz
KNN	Künstliches Neuronales Netzwerk
MLP	<i>multilayer perceptron</i>
PNN	<i>probalistic neural network</i>
ResNet	<i>residual neural network</i>
VGG	<i>visual geometry group</i>
Y-TZP	Yttriumstabilisiertes tetragonales polykristallines Zirkoniumoxid

Abbildungsverzeichnis

Abbildung 1 Architektur eines Single Layer Perceptrons	3
Abbildung 2 Aufbau eines mehrschichtigen künstlichen neuronalen Netzwerkes	5
Abbildung 3 Exemplarische Darstellung des Skalarproduktes in einem CNN aus einem Pixelfeld und einem Filter	7
Abbildung 4 Architektur eines CNNs mit mehreren unterschiedlichen Schichten	9
Abbildung 5 Architektur eines PNNs	12
Abbildung 6 Aufbau eines Random Forest-Algorithmus	13
Abbildung 7 Original- (jeweils links) und Replikat-Implantat (jeweils rechts) mit passendem manuellem Schraubendreher; links inserierte Schraube, rechts ohne inserierte Schraube	24
Abbildung 8 Nahaufnahme des Implantats vor dem Detektor.....	26
Abbildung 9 Untersuchte Implantate	27
Abbildung 10 Kontstruktionselemente.....	28
Abbildung 11 Gesamtkonstruktion für die standardisierte Bildgebung.....	29
Abbildung 12 Visuelle Orientierung zum Einordnen der Konfidenzniveaus der Teilnehmer.....	33
Abbildung 13 Trainingsdatensatz = 674	35
Abbildung 14 Trainingsdatensatz = 906	35
Abbildung 15 Konfidenzniveaus der männlichen und weiblichen Studienteilnehmer	37

Tabellenverzeichnis

Tabelle 1 Straumann Roxolid® Implantat.....	24
Tabelle 2 Meisinger Implantat	25
Tabelle 3 Trainingsdatensatz = 488.....	34
Tabelle 4 Genauigkeitswerte der menschlichen Experten	36
Tabelle 5 Genauigkeitswerte der weiblichen Experten.....	36
Tabelle 6 Genauigkeitswerte der männlichen Experten	37

1 Einleitung

Die rasante Entwicklung von künstlicher Intelligenz (KI) und maschinellem Lernen hat in den letzten Jahren auch in der Zahnmedizin zu bedeutenden Fortschritten geführt [1], insbesondere im Bereich der automatisierten Bildanalyse. Künstliche neuronale Netze (KNN), die auf den Prinzipien der Informationsverarbeitung des menschlichen Gehirns basieren [2], werden in der Bildverarbeitung genutzt, um präzise Diagnosen und Klassifikationen durchzuführen [3]. Besonders im Fokus steht dabei die Fähigkeit von Deep Learning-basierten Modellen, anhand von Röntgenbildern zahnmedizinische Implantate zu identifizieren. Dieser Prozess erfolgt, indem ein Netzwerk, wie zum Beispiel ein Convolutional Neural Network (CNN), durch die Analyse von Bilddaten trainiert wird, um Muster und Merkmale zu erkennen und daraus eine fundierte Klassifikation abzuleiten [4, 5, 6].

Ein bisher noch vernachlässigter Bereich der Forschung liegt in der Differenzierung zwischen Original- und Replikat-Implantaten anhand von Röntgenaufnahmen mittels KI. Studien zeigen, dass KI-Modelle imstande sind, Implantate auf der Grundlage subtiler Bildmerkmale, die für das menschliche Auge oft schwer erkennbar sind, zu unterscheiden [7, 8, 9, 10, 11]. In der vorliegenden Studie wurde ein Residual Neural Network (ResNet) verwendet, um zwei sehr ähnliche Implantatmodelle zu klassifizieren – ein Original-Implantat von Straumann und ein Replikat-Implantat von Meisinger. Dabei wurden Röntgenbilder ex situ unter standardisierten Bedingungen aufgenommen, um die KI zu trainieren und zu testen. Die Ergebnisse der KI wurden mit den Einschätzungen menschlicher Experten¹ verglichen, um die Leistungsfähigkeit der KI zu evaluieren.

Das Ziel der vorliegenden Studie ist es, die Hypothese zu testen, dass eine künstliche Intelligenz in der Lage ist, zwischen einem Original- und einem Replikat-Implantat mit einer höheren Genauigkeit zu differenzieren als menschliche Experten. Die Hypothese basiert auf den bisherigen Erkenntnissen zur Effizienz von KI-gestützten Bildverarbeitungssystemen, die in mehreren Studien bereits hohe Klassifikationsgenauigkeiten unter Beweis stellen konnten [1, 3].

¹ In der folgenden Arbeit wird aus Gründen der besseren Lesbarkeit ausschließlich die männliche Form verwendet. Es wird an dieser Stelle darauf hingewiesen, dass die ausschließliche Verwendung der männlichen Form geschlechtsunabhängig verstanden werden soll.

Um einen realitätsnahen Vergleich zu ermöglichen, wurden neben den Ergebnissen der KI auch die Konfidenzniveaus der teilnehmenden Zahnmediziner erfasst, um deren diagnostische Sicherheit zu bewerten. Darüber hinaus wurden mit unterschiedlich großen Datensätzen trainierte KI-Modelle verwendet, um die Leistungsfähigkeit der KI zu maximieren. Die Studiendurchführung und Datenerhebung erfolgten unter kontrollierten Bedingungen, um die Reproduzierbarkeit und Aussagekraft der Ergebnisse zu gewährleisten.

Diese Studie legt somit dar, inwiefern KI-Systeme in der zahnmedizinischen Diagnostik effektiv eingesetzt werden können, um menschliche Experten auch bei diffizilen Klassifikationsproblemen zu unterstützen.

2 Literaturdiskussion

2.1 Künstliche Intelligenz

Künstliche neuronale Netze (KNN) sind – wie im Namen schon bereits enthalten – natürlich vorkommenden neuronalen Netzen nachempfunden. Es handelt sich dabei um mathematische Nachbildungen der Reizverarbeitung im Gehirn, wobei anstelle elektrischer oder chemischer Signale bei biologischen Systemen Zahlenwerte von Algorithmen verarbeitet werden [2].

2.1.1 Grundlagen von künstlicher Intelligenz und maschinellem Lernen

Für einen allgemeinen Überblick bietet sich zunächst eine Differenzierung an, welche Methoden bzw. grundsätzlichen Ideen angewendet werden können, künstliche Intelligenz zu programmieren und zu trainieren.

Zunächst soll ein klassisches künstliches neuronales Netzwerk in seinen Grundzügen umrissen werden. Die Netze haben alle gemeinsam, dass Input-Signale eingespeist werden (z. B. Daten von einer binär klassifizierbaren Problemstellung), diese dann eine Netzstruktur aus einer oder mehreren Neuronenschichten durchlaufen und später als Output-Signale und damit als Entscheidungsfindung vom System herausgegeben werden. In der folgenden Abbildung ist ein sogenanntes „Single Layer Perceptron“ dargestellt. Dabei handelt es sich um ein KNN mit einer Schicht Neuronen – speziell in diesem Fall ein Neuron – zwischen der Input- und der Output-Schicht [12].

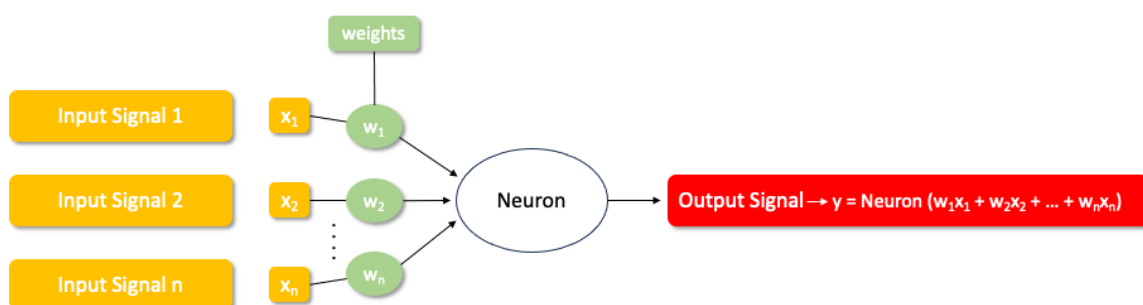


Abbildung 1 Architektur eines Single Layer Perceptrons

Die Input-Signale (x_1, x_2, x_n) werden noch auf ihrem Weg zum Neuron durch „weights“ modifiziert. Diese Gewichtungen sind die Parameter, in denen der Lerneffekt einer künstlichen Intelligenz begründet ist. Denn sie werden fortlaufend angepasst, je mehr Daten dem KNN zugespielt werden. Genau diese Anpassung trägt schlussendlich dazu bei, dass die KI stets genauer wird, was die entsprechenden Zielwerte (Output-Signale) betrifft. Im Neuron wird im Anschluss die gewichtete Summe aller Eingangswerte gebildet. Bei dem Neuron selbst handelt es sich um eine mathematische Funktion, die dann auf die entstandene gewichtete Summe angewendet wird. Man spricht von einer sogenannten Aktivierungsfunktion. Abhängig vom Ergebnis gibt das Neuron entweder ein Signal weiter oder nicht. Das Ergebnis beträgt immer 0 oder 1. Mithilfe der Aktivierungsfunktion werden die gewichteten Summen letzten Endes einer der beiden Ziffern zugewiesen. Ab einem bestimmten Schwellenwert nimmt das Ergebnis den Wert 1 an, wodurch die Entscheidung zugunsten desjenigen Ergebnisses getroffen wird, dem dieser Schwellenwert zugeordnet ist. Die Aktivierungsfunktion muss also vorher im Sinne des jeweiligen Problems festgelegt worden sein [13, 14].

Im Folgenden soll auf die Einteilung verschiedener KI-Konzepte eingegangen werden. Ein wichtiges Gebiet im Zusammenhang mit der Zahnmedizin ist hierbei das maschinelle Lernen mit seinen konzentrischen Teilmengen, wozu das sogenannte Deep Learning und die künstlichen neuronalen Netze gezählt werden müssen.

Maschinelles Lernen umfasst Algorithmen, die häufig für das Klassifizieren von Objekten, zur Mustererkennung oder zur Vorhersage von Ergebnissen ausgelegt sind. Ihr Training kann auf unterschiedliche Weise erfolgen. Grundsätzlich läuft dies entweder überwacht, unüberwacht, teilüberwacht oder bestärkend ab. Allgemein gesprochen wird bei allen vier Varianten eine sehr große Datenmenge dem Algorithmus zugespielt. Im überwachten und teilüberwachten Bereich wird dem Programm vorgegeben, wie korrekte Ergebnisse aussehen, woraus es mit der Zeit selbst lernt, diese richtig zu erkennen. Bei der unüberwachten Methode teilt die KI die Rohdaten selbst ein und wird stets besser mit der Identifizierung von Mustern und Korrelationen. Bestärkendes Lernen bedeutet, dass in dem Programm Regeln und potenzielle Ergebnisse vordefiniert worden sind und eine damit einhergehende Belohnung. Diese ist als etwas programmiert, das das System gerne erfassen möchte [15].

Deep Learning-Modelle stützen sich auf Schichten von neuronalen Netzen. Die Funktionsweise ist dem oben erklärten „Single Layer Perceptron“ zu entnehmen mit dem Unterschied, dass die meisten Deep Learning-Modelle aus mehreren solcher Schichten, sogenannten „hidden layers“, und mehreren Neuronen pro Schicht bestehen (Abbildung 2) [16].

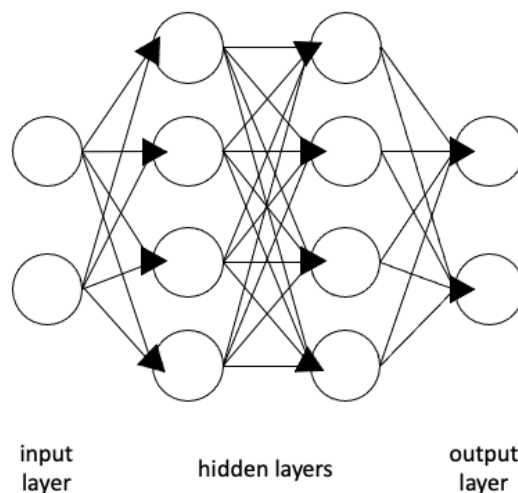


Abbildung 2 Aufbau eines mehrschichtigen künstlichen neuronalen Netzwerkes

Maschinelles Lernen lässt sich davon abgrenzen, dass hierbei der Mensch in die Analyse der Daten und den eigentlichen Entscheidungsprozess eingreift. Deep Learning hingegen beruht auf dem vollkommen eigenständigen Lernen des Programms, wobei aus den vorhandenen Daten Muster extrahiert und klassifiziert werden. Dies geschieht kleinschrittig pro Schicht, sodass am Ende die Informationen zusammengeführt werden und zu einem bestimmten Output-Signal führen [17].

2.1.2 Technologien und Algorithmen für die automatisierte Implantatidentifizierung

Im Folgenden soll nun auf die Netzwerktypen eingegangen werden, die zurzeit die populärsten in der zahnmedizinischen Studienlage zu der Thematik „KI“ darstellen. Dazu zählen CNNs, MLPs, PNNs und random forests.

2.1.2.1 CNN – Convolutional Neural Network

CNNs sind neuronale Netze, die vor allem in der „Computer Vision“ Anwendung finden [18, 19, 20]. Sie extrahieren anhand von Filtern charakteristische Merkmale aus Bilddaten, wodurch letzten Endes eine automatisierte Mustererkennung ermöglicht wird. Konvolutionale neuronale Netze (CNNs) weisen strukturelle Ähnlichkeiten zu traditionellen künstlichen neuronalen Netzen (KNNs) auf, da auch sie aus Neuronen bestehen, die durch Lernprozesse optimiert werden. Dabei erhält jedes Neuron eine Eingangsinformation und führt eine Operation durch, die in der Regel in Form eines Skalarprodukts sowie einer nachfolgenden nichtlinearen Funktion erfolgt. Diese Verarbeitung bildet ebenfalls das Fundament vieler weiterer neuronaler Netzwerke. Von den Eingabedaten bis zur finalen Klassifikation bleibt das Netzwerk darauf ausgerichtet, eine konsistente Bewertungsfunktion zu formulieren, die durch die Gewichtung der Neuronen festgelegt wird. In der letzten Schicht des CNNs kommen Verlustfunktionen zum Einsatz, die den Klassenzuweisungen zugeordnet sind [12].

Das grundlegende Prinzip eines CNNs in Bezug auf die Objektidentifizierung besteht darin, dass das Programm eigenständig Bilder lesen und Objekte erkennen kann, indem mehrere verschiedene Filter über das Pixelfeld des jeweiligen Bildes laufen. Ein Filter kann z. B. ein 3x3 Pixelfeld sein. Dieser fährt nun das gesamte Bild ab, indem das Skalarprodukt jedes 3x3 Pixelfeldes des Bildes mit dem des Filters gebildet wird. Pro Pixel besitzt das Bild einen bestimmten Farbwert, dargestellt durch eine Zahl. Die einzelnen Pixel des Filters bekommen ebenfalls Zahlenwerte durch den Programmierer zugewiesen. Als Ergebnis erhält man eine Abbildung des zu untersuchenden Objektes.

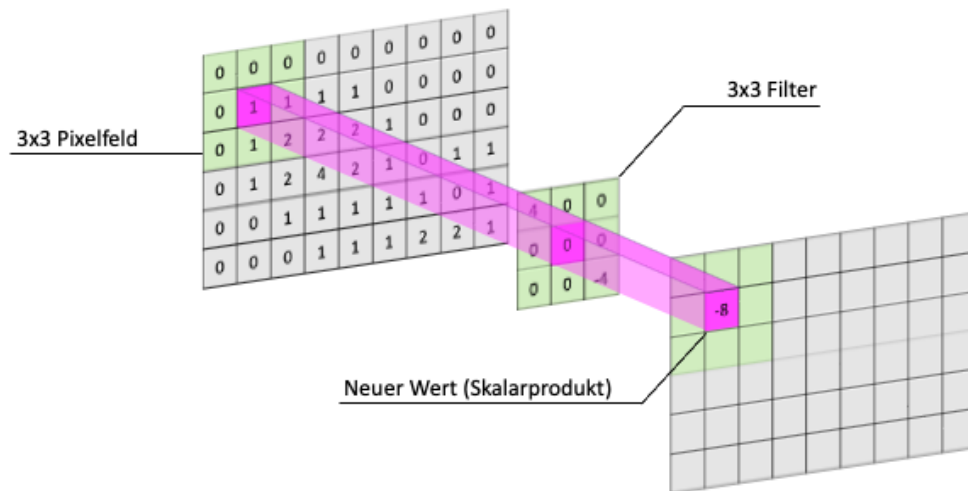


Abbildung 3 Exemplarische Darstellung des Skalarproduktes in einem CNN aus einem Pixelfeld und einem Filter

Je nach festgelegtem Filter unterscheidet sich die Abbildung jedoch. Durch die Kombination der Ergebnisse verschiedener Filter entsteht eine detaillierte und differenzierte Abbildung des Objekts, die es dem Netzwerk ermöglicht, eine Erkennung und Klassifikation vorzunehmen. Denn die verschiedenen Ansichten des Bildes werden kombiniert, um eine umfassende Repräsentation des Bildes zu erstellen, die für die Objekterkennung genutzt werden kann [4, 5].

Im Folgenden wird detailliert beschrieben, wie die verschiedenen Schichten innerhalb eines CNNs funktionieren und in welcher Weise die Verlustfunktionen der finalen Schicht zur Entscheidungsfindung bei der Klassifikation beitragen. Dabei wird die Funktionsweise jeder Schicht erläutert, beginnend mit den ersten Faltungsschichten, die grundlegende Merkmale aus den Eingangsdaten extrahieren, bis hin zu den komplexeren Schichten, die zunehmend abstraktere Merkmale erkennen und interpretieren. Besondere Bedeutung kommt dabei der letzten Schicht des CNNs zu, in der die Verlustfunktion implementiert ist.

Die Architektur eines typischen CNN besteht im Wesentlichen aus mehreren Schichten: der Eingabeschicht, den Faltungsschichten, den Pooling-Schichten, den vollständig verbundenen Schichten (Fully Connected Layer, FC) und der

Ausgabeschicht. In der Regel sind mehrere Wechsel zwischen den Faltungsschichten und Pooling-Schichten vorhanden.

Zu Beginn wird das Eingabebild durch die Faltungsschichten geleitet, die, wie bereits erwähnt, einen vordefinierten Filter über das Bild verschieben und dabei relevante Merkmalsinformationen, wie Kanten oder Texturen, schichtweise extrahieren. Es entstehen sogenannte Merkmalskarten (Feature Maps). Im Anschluss daran folgen die Pooling-Schichten, deren Hauptaufgabe darin besteht, die Größe der Merkmalskarten zu reduzieren. Dies wird erreicht, indem nur die wichtigsten Merkmale – wie der maximale Farbwert in einem bestimmten Bildbereich (Max Pooling) – weitergegeben werden. Dieser Prozess, bekannt als Downsampling, hilft dabei, die Komplexität des Modells zu verringern und die Gefahr einer Überanpassung („Overfitting“) zu minimieren [5]. Durch den Wechsel der beiden Schichten wird erreicht, dass das Downsampling öfter wiederholt wird. Die regelmäßige Anwendung des Downsampling-Prozesses ermöglicht es, die Effizienz des Netzwerks zu wahren, indem übermäßige Datenkomplexität verringert und gleichzeitig die wesentlichen Merkmale des Bildes beibehalten werden, damit schlussendlich eine effiziente Rechenleistung gewährleistet wird [21].

Demzufolge entspricht ein CNN nicht einem klassischen „feed-forward“-Netzwerk, bei dem die Daten in der Netzstruktur chronologisch verarbeitet werden. Es bezieht sich nämlich im Zuge der Erstellung seiner Karten und Muster mehrfach auf dieselben Daten unter Anwendung diverser Filter [22]. Am Ende des CNNs befindet sich die vollständig verbundene Schicht, die dazu dient, die extrahierten zweidimensionalen Merkmale in eine eindimensionale Darstellung zu überführen, die für die finale Klassifikation verwendet wird. Diese Schicht ähnelt den versteckten Schichten traditioneller mehrschichtiger Perzeptronen (MLPs) und kombiniert die Merkmale der vorhergehenden Schichten, um eine genaue Vorhersage zu ermöglichen (Abbildung 4) [23].

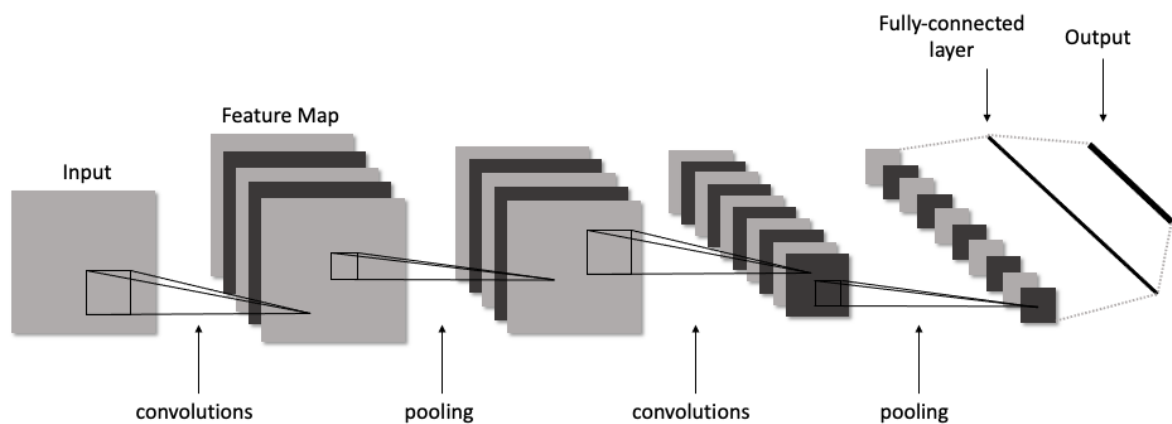


Abbildung 4 Architektur eines CNNs mit mehreren unterschiedlichen Schichten

Nachfolgend wird erklärt, wie das Lernen eines solchen Netzwerkes stattfindet. Dabei muss auf die in Kapitel 2.1.1 erwähnten Aktivierungsfunktionen, die einzelnen individuellen Gewichte in den Neuronen und die Verlustfunktion eingegangen werden. Die Aktivierungsfunktionen, die nach jeder Faltungs- und vollständig verbundenen Schicht angewendet werden, tragen dazu bei, dass das Netzwerk komplexe, nichtlineare Muster erlernt. Ohne diese Funktionen könnte das Netzwerk nicht die notwendigen Transformationen durchführen, die für eine erfolgreiche Klassifikation erforderlich sind, da durch sie die Aktivierung eines Neurons festgelegt wird [23]. Die Gewichte werden im Hinblick auf die Ausgabe des Netzwerkes beeinflusst. Sie befinden sich zwischen den Neuronen im CNN und zeigen an, wie stark deren Verknüpfung miteinander ist. Das Ziel ist es im Rahmen des Trainings die Gewichte derart anzupassen, um den Fehler zu minimieren, der durch die Verlustfunktion gemessen wird. Die Gewichte bestimmen, wie stark ein bestimmtes Merkmal (z.B. ein Pixelwert in einem Bild) die Ausgabe eines Neurons des CNNs beeinflusst. Das Netzwerk passt diese Gewichte so an, dass das Erlernen von Merkmalen in den zugrundeliegenden Datensätzen ermöglicht wird und diese immer besser klassifiziert werden können [6]. Die Verlustfunktion dient der Beurteilung der Vorhersagegenauigkeit des neuronalen Netzwerkes. Sie quantifiziert die Abweichung zwischen den vom Netzwerk ausgegebenen Vorhersagen und den tatsächlichen, in den Trainingsdaten enthaltenen Werten. Der Lernprozess in einem CNN besteht darin, die Aktivierungen (durch die Anpassung der Gewichte in den Neuronen) so zu

verändern, dass die durch die Verlustfunktion gemessenen Fehler minimiert werden. Entsprechend bewertet sie das Gesamtergebnis aller Aktivierungen in Bezug auf das Ziel, das das Netzwerk zu erreichen versucht (z.B. eine bestimmte Klasse korrekt zu identifizieren) [5]. Das Minimieren der Fehler erfolgt über einen Rückwärtslauf, auch als Backpropagation bekannt. Dieser beschreibt den Prozess der Rückwärtsausbreitung von Fehlern innerhalb eines neuronalen Netzwerks. Dabei wird der Gradientenabstieg eingesetzt, um den Gradienten der Verlustfunktion in Bezug auf die Netzwerkparameter wie die Gewichte zu berechnen. Der Gradient ist im Wesentlichen eine mathematische Größe, die angibt, in welche Richtung und wie stark die Verlustfunktion ansteigen würde, wenn man die Gewichte verändert. Beim Gradientenabstieg wird der Gradient genutzt, um die Parameter so anzupassen, dass die Verlustfunktion möglichst klein wird. Während der Backpropagation wird zuerst der Gradient der Verlustfunktion in den letzten Schichten des Netzwerks berechnet, wobei die Gradienten der früheren Schichten sukzessive zuletzt ermittelt werden. Ein geringerer Wert der Verlustfunktion ist dabei ein Indikator für die verbesserte Leistungsfähigkeit respektive Genauigkeit des Modells [24].

Ein zentraler Aspekt, der zu erkennen gibt, weshalb die Genauigkeit und damit die Leistungsfähigkeit eines Modells beeinträchtigt werden kann, ist das Phänomen des Overfittings, welches beim Prozess des Programmierens minimiert werden soll. Overfitting ist ein häufig auftretendes Problem bei der unsachgemäßen Anwendung von ML-Algorithmen. Es tritt auf, wenn ein KI-Modell eine übermäßige Passgenauigkeit auf die Trainingsdaten aufweist, jedoch aufgrund einer zu hohen Modellkomplexität bei neuen, unberührten Daten eine unzureichende Leistung erbringt. Ein zu stark vereinfachtes Modell ist möglicherweise nicht in der Lage, wesentliche Vorhersagebeziehungen zu erfassen und führt daher zu Underfitting. Im Gegensatz dazu kann eine übermäßig hohe Modellkomplexität dazu führen, dass das Modell sich zu stark an das Rauschen in den Trainingsdaten anpasst und dadurch überangepasst wird. Wenn also beispielsweise Trainingsdaten nicht ausreichend vielfältig oder zu klein sind, lernt das Modell Muster, die spezifisch für diese Daten sind. Diese speziellen Bildbereiche sind allerdings nicht repräsentativ für die allgemeinen Muster, die in neuen Daten auftreten, was dazu führt, dass das Modell bei neuen Daten nicht gut generalisiert [25, 26].

2.1.2.2 MLP – Multilayer Perceptrons

MLPs bestehen aus mehreren vollständig miteinander verbundenen Neuronenschichten und lernen Probleme zu lösen, indem sie Beziehungen zwischen den Eingabedaten und den Zielwerten durch eine Vielzahl nichtlinearer Transformationen herstellen. Der Aufbau und die Funktion sind mit den Single Layer Perceptrons gleichzusetzen, nur dass mehrere Neuronen genutzt werden, wodurch komplexere Sachverhalte gelöst werden können (Abbildung 2). Vor allem nicht-lineare Probleme werden mit MLPs bewältigt [27].

2.1.2.3 PNN – Probabilistic Neural Network

Ein PNN nutzt Wahrscheinlichkeitsverteilungen, um die Eingabedaten in Kategorien zu klassifizieren, wobei es die Unsicherheit in den Daten berücksichtigt. Dies unterscheidet es von anderen neuronalen Netzen, die deterministische Vorhersagen treffen. Diese geben für denselben Eingabewert immer denselben Ausgabewert an. Demzufolge ist ein PNN besonders vorteilhaft für Anwendungen, bei denen die Unsicherheit von hoher Relevanz ist, wo eine klare Einschätzung der Risiken und Unsicherheiten entscheidend sein kann.

PNNs setzen sich aus vier Schichten zusammen, der input, pattern, summation und decision layer. Die Besonderheit liegt hier darin begründet, dass für jede Kategorie der Zielvariablen ein Musterneuron in der „summation layer“ vorhanden ist. Dieses bekommt nur Werte von Neuronen aus der „pattern layer“ eingespeist, die der tatsächlichen Zielkategorie entsprechen. Die Ausgabeschicht vergleicht die gewichteten Stimmen für jede Zielkategorie und verwendet die größte Stimme, um das Ergebnis vorherzusagen.

Ein wesentlicher Vorteil gegenüber einem MLP besteht darin, dass die zweite und dritte Schicht durch weniger Verknüpfungen miteinander verbunden sind. Das macht PNNs schneller und dennoch können sie in der Klassifikation akkurater sein. Es können genauere Werte für die prognostizierte Zielwahrscheinlichkeit erzeugt werden. Nachteilig erweist sich der größere Speicherplatz, der vonnöten ist, um das Netzwerk abzuspeichern. Zudem läuft das Klassifizieren gänzlich neuer Fälle langsamer ab [28].

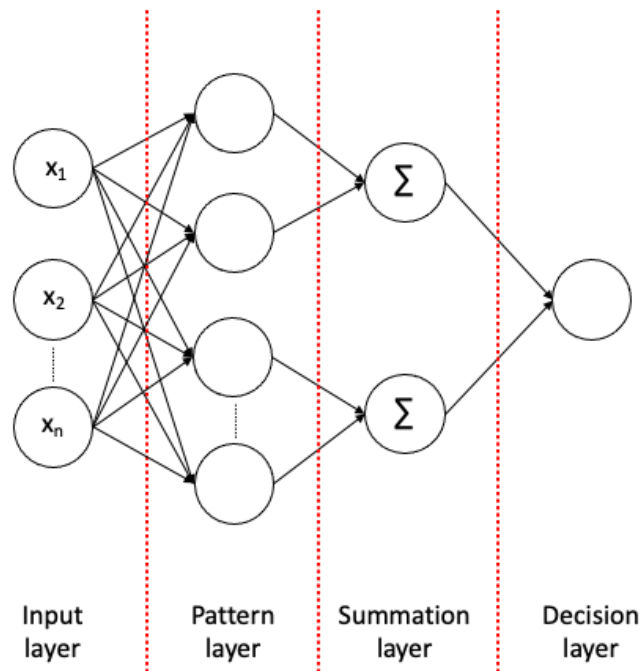


Abbildung 5 Architektur eines PNNs

2.1.2.4 Random Forest

Bei einem „Random Forest“ entsteht eine Vorhersage dadurch, dass viele Entscheidungsbäume zusammenarbeiten. Jeder Baum trifft eine eigene Vorhersage und die Mehrheit der Bäume entscheidet am Ende, wie das Ergebnis aussieht. „Random Forests“ finden im Bereich des maschinellen Lernens Anwendung. Sie sind der Klasse des überwachten Lernens zuzuschreiben. Die Entscheidungsbäume mit den einzelnen Verzweigungen entstehen aufgrund der Zuordnung von Daten zu einer Klasse mithilfe vordefinierter Regeln. Die Eigenschaften der Daten und Entscheidungskriterien unterscheiden sich von Baum zu Baum. Die Vorhersagegenauigkeit des random forest – Algorithmus wird durch die Varianz innerhalb der Einzelvorhersagen jedes Baumes erhöht [29]. Wichtig im Zusammenhang mit der Entscheidungsfindung ist das „Averaging/Majority Voting“. Es setzt sich aus den Resultaten der einzelnen Entscheidungsbäume zusammen, wobei die jeweiligen Ergebnisse mit einer festgelegten Gewichtung in das Endresultat einfließen [30].

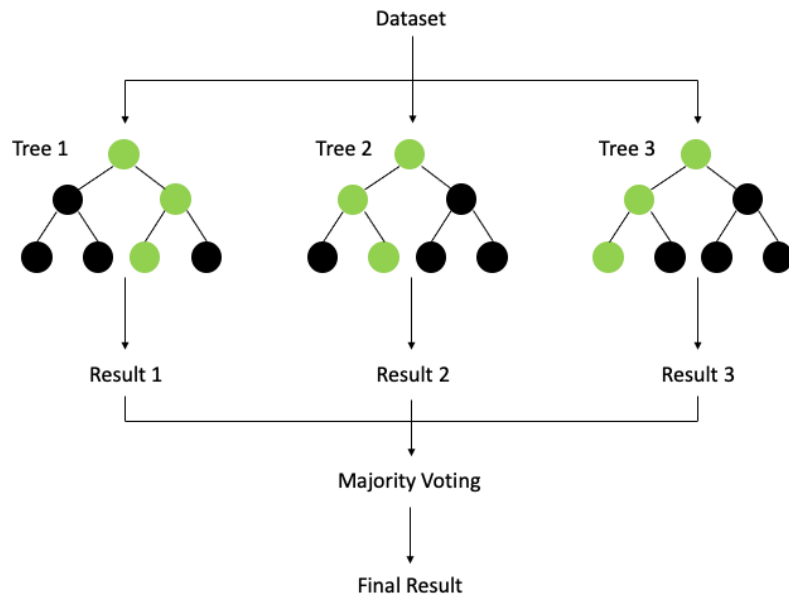


Abbildung 6 Aufbau eines Random Forest-Algorithmus

Random-Forest-Algorithmen erweisen sich als besonders effektiv, wenn umfangreiche Datensätze mit einer großen Anzahl von Klassen, Merkmalen und Trainingsdaten trainiert werden sollen. Sie benötigen außerdem geringere Rechenleistungen als künstliche neuronale Netze. Andererseits büßen sie an Effizienz bei Echtzeitanwendungen ein, was gerade für die Anwendung im medizinischen Bereich eine wichtige Thematik ist [31].

2.1.3 Aktuelle Studienlage zu KI-gestützter automatisierter Implantatidentifizierung anhand von Röntgenaufnahmen

In den letzten Jahren hat die automatisierte, KI-gestützte Implantatidentifizierung anhand von Röntgenaufnahmen bedeutende Fortschritte gemacht [1]. Es gibt eine zunehmende Anzahl von Studien, die mit größeren und vielfältigeren Datensätzen arbeiten. Jedoch bleibt die Herausforderung bestehen, diese Studien miteinander zu vergleichen. Insbesondere die Heterogenität der Datensätze, die für das Training der KI-Programme verwendet wurden, erschwert eine objektive Bewertung der Leistungsfähigkeit der verschiedenen Ansätze. Diese Vielfalt an Datensätzen führt dazu, dass es schwierig ist, allgemeingültige Aussagen über die Effektivität und Zuverlässigkeit der eingesetzten KI-Modelle zu treffen. Jedoch lassen sich bestimmte Tendenzen erkennen, die auf eine positive Gesamtentwicklung hinweisen und das Potenzial von der Anwendung künstlicher Intelligenz in der zahnmedizinischen Diagnostik hervorheben.

Diese Entwicklungen zeigen, dass KI-basierte Systeme zunehmend dazu fähig sind, komplexe diagnostische Aufgaben zu bewältigen, die traditionell von menschlichen Experten durchgeführt wurden. Die Fähigkeit dieser Systeme, große Mengen an Bilddaten zu verarbeiten und dabei präzise Ergebnisse zu liefern, könnte zu einer bedeutenden Verbesserung der diagnostischen Genauigkeit und Effizienz in der Zahnmedizin führen [32]. Um ein Verständnis für die Heterogenität der einzelnen Studien und Daten zu entwickeln, muss zunächst auf die diversen Kriterien eingegangen werden, die zu selbiger führen.

Maßgeblich werden die Unterschiede in den Ergebnissen der Studien durch die Menge der Trainingsdaten beeinflusst. Diese variieren in der Literatur zum Teil stark, sowohl von der Anzahl der verwendeten Röntgenbilder als auch von den berücksichtigten diversen Implantatsystemen [7]. Die Stichprobengrößen der Studien umfassen Datensätze von 801 [8] bis zu 156.965 Röntgenbildern [33]. Die Varianz innerhalb der Datensätze wird erheblich durch die Anzahl der eingeschlossenen Implantatsysteme beeinflusst, wobei in einigen Studien lediglich drei Systeme einbezogen werden [34, 35, 36, 37], während in anderen bis zu 130 unterschiedliche Systeme [38] analysiert werden. Die Verteilung der Daten zwischen Trainings- und Testbildern zeigt ebenfalls eine erhebliche Variabilität. Diese reicht von einer Aufteilung im Verhältnis 80:20 [8, 9, 10, 11, 36, 39] bis hin zu einer Verteilung von 97,5:2,5 [40]. Diese Disparität führt zu einer erheblichen Verzerrung der Daten, was die Vergleichbarkeit und Generalisierbarkeit der Studienergebnisse beeinträchtigt [3].

Es zeigt sich, dass die Genauigkeit der Identifizierung und Klassifikation tendenziell mit der Anzahl der verwendeten Trainingsdaten steigt. Entsprechend gegenteilig führt eine geringere Anzahl eingesetzter Implantatsysteme in der Regel zu einer höheren Genauigkeitsrate [7].

Des Weiteren haben bestehende Studien keine wesentlichen Parameter (einschließlich Kontrast, Intensitätsniveau, Schärfe, Seitenverhältnis, Orientierung und Auflösung) berücksichtigt, die sich auf das Qualitätsmanagement und die Standardisierung zahnmedizinischer Röntgenbilder beziehen [3]. Mehrere Studien [10, 34, 36, 40] legen ein Augenmerk auf die Bewertung der Bildqualität und analysieren dabei Aspekte wie Bildrauschen, Unschärfe, Fehler bei der Positionierung sowie Verzerrungen der eingesetzten Bildaufnahmen. Allerdings wurde in keiner dieser Studien eine Validierung durch einen ausgewiesenen Experten durchgeführt, der direkt in die Verwaltung des Datensatzes involviert war [3]. Im Falle einer Validierung der verwendeten Bilddaten durch Experten zeigte sich eine heterogene Zusammensetzung der beteiligten Fachgruppen in den entsprechenden Studien [7]. Die zahnärztlichen Experten, die an diesem Validierungsprozess teilnahmen, stammten aus unterschiedlichen Fachrichtungen, darunter die Parodontologie, Prothetik sowie die Oral- und Kieferchirurgie [10, 33, 36, 41, 42, 43]. Insgesamt besteht das Problem darin, dass es kaum möglich ist, einen Goldstandard in Bezug auf die Datensätze zu definieren. Die Einschätzungen von Experten können unterschiedliche Interpretationen liefern. Es erweist sich als komplex, eine genaue Anzahl von Experten festzulegen, die notwendig ist, um einen zuverlässigen und einheitlichen Referenzstandard zu erstellen und wie dabei die einzelnen, voneinander abweichenden Einschätzungen zusammengeführt werden müssen [32, 44].

Ein weiterer Aspekt, der den Vergleich der Daten erschwert und zu Inkonsistenzen bei der Bewertung der Ergebnisse führt, ist die Verwendung unterschiedlicher Bildformate in den Studien. Einige Datensätze greifen ausschließlich auf periapikale Röntgenbilder zurück [9, 11, 39, 40, 45], während andere nur Panoramaröntgenaufnahmen verwenden [8, 34]. In manchen Fällen werden beide Bildformate kombiniert und innerhalb desselben Trainingsdatensatzes genutzt [10, 36]. Die Verwendung eines vielschichtigen Datensatzes fördert die Zuverlässigkeit und die Generalisierbarkeit der Ergebnisse, indem sie eine fundierte Grundlage für die Analyse bietet. Entsprechend zeigt sich, dass die Kombination im Datensatz aus sowohl panoramischen als auch periapikalen Röntgenaufnahmen zuverlässig zur

Identifikation von Implantatsystemen herangezogen werden kann [1, 3]. Die Einbeziehung von dreidimensionalen Röntgenaufnahmen, insbesondere der digitalen Volumentomographie (DVT), in die Datensätze ist bislang unzureichend berücksichtigt worden [3, 7]. DVT-Bilder, die in der dentalen Implantologie weit verbreitet sind, bieten im Vergleich zu zweidimensionalen Bildgebungsverfahren den Vorteil, dass sie verzerrungsarme Darstellungen und dreidimensionale Volumeninformationen liefern können [46]. Aus diesem Grund ist die Entwicklung von Deep Learning-Modellen, die DVT-Daten als Eingabe verarbeiten können, von entscheidender Bedeutung [47]. Ebenso unerlässlich ist eine vergleichende Analyse dieser Modelle mit solchen, die auf der Auswertung von zweidimensionalen Bildern basieren, um deren relative Leistungsfähigkeit und potenziellen Vorteile in der klinischen Anwendung zu bewerten [3].

Die Heterogenität der Studien spiegelt sich ebenfalls in den DL-Algorithmen wider, welche unterschiedliche Architekturen aufweisen. Obwohl allgemein erwartet wird, dass neuere oder modifizierte DL-Algorithmen eine höhere Genauigkeit aufweisen, konnten bislang keine signifikanten Unterschiede in der Genauigkeit identifiziert werden, die direkt auf Unterschiede zwischen den Algorithmen zurückzuführen wären. Diese Beobachtung legt nahe, dass die Quantität und Qualität der zugrunde liegenden Datensätze gegenwärtig von noch größerer Bedeutung sind als das verwendete DL-Modell [3]. Zwei Studien, bei denen verschiedene DL-Algorithmen auf identischen Datensätzen angewendet wurden, zeigen keine signifikanten Unterschiede in der Genauigkeit, die auf die Wahl des spezifischen DL-Algorithmus zurückzuführen wären [10, 40].

Es zeigt sich dementsprechend eine signifikante Heterogenität in den Ergebnissen der einzelnen Studien. Die berichtete Genauigkeit variiert von einem Minimum von 67% [35] bis zu einem Maximum von 98,51% [11]. Mehrere Studien in diesem Forschungsbereich weisen Genauigkeitswerte von über 90 % auf [7, 8, 9, 10, 11, 36, 37, 41, 42, 43, 45, 48]. In systematischen Übersichtsarbeiten werden mittlere Gesamtgenauigkeitswerte der jeweils berücksichtigten Studien von 92,16 % [3] bzw. 92,56 % [1] genannt. Des Weiteren wird darauf hingewiesen, dass die Genauigkeit mit zunehmender Anzahl an Schichten des verwendeten CNN tendenziell steigt. In diesem Kontext wird von einem „Deep CNN“ gesprochen, wenn die Architektur des CNN durch eine erhöhte Anzahl von Schichten gekennzeichnet ist. Die Unterschiede

in der Genauigkeit sind jedoch gering und bewegen sich im Zehntel-Prozent-Bereich [11]. Eine weitere Studie zeigt, dass selbst bei Verwendung von Testdatensätzen mit qualitativ minderwertigen und verzerrten Röntgenbildern eine hohe Genauigkeit von 95,1 % erreicht werden konnte. Die mindere Qualität der Bilder wurde auf mangelnde orthogonale Ausrichtung zur Implantatachse, übermäßige Strahlenexposition, unvollständige Darstellung der Implantatspitze und das Vorhandensein von Fremdkörpern zurückgeführt [49].

Für den zugrundeliegenden Ansatz der vorliegenden Arbeit ist wichtig zu beachten, dass die Heterogenität der Datensätze und Studienmethoden in der KI-gestützten Implantatidentifizierung eine objektive Beurteilung der Leistungsfähigkeit der Modelle erschwert. Besonders entscheidend sind die Größe und Vielfalt der Trainingsdaten sowie die Standardisierung der Bildqualität, da diese die Generalisierbarkeit und Genauigkeit maßgeblich beeinflussen. Zudem zeigt sich, dass die zugeführten Daten mehr Einfluss auf die Ergebnisse haben als die Wahl des spezifischen Deep Learning-Algorithmus.

2.2 Implantate

2.2.1 Material und Klassifikation von dentalen Implantaten

Die Grundlage der modernen Implantologie wurde durch die Arbeit von Prof. Per-Ingvar Brånemark in den 1950er und 1960er Jahren gelegt, als er bei seinen Untersuchungen zur Mikrozirkulation bei der Knochenheilung das feste Verwachsen von Titan und Knochen feststellte, was er als Osseointegration beschrieb [50]. Diese wegweisende Entdeckung führte schließlich zur Publikation seiner Ergebnisse im Jahr 1977, in der er den erfolgreichen Einsatz von Titanimplantaten im menschlichen Kieferknochen beschrieb. Damit legte Brånemark den Grundstein für die moderne Implantologie und etablierte Titanimplantate als Standard in der zahnmedizinischen Praxis [51]. Um Kriterien und Standards für objektive Nachsorgeprotokolle von Implantaten festzulegen, veranstalteten die National Institutes of Health 1978 die „Dental Implant Consensus Conference“ [52]. Die Ergebnisse dieser Konferenz in Kombination mit der Arbeit von Brånemark et al. [53] setzten den aktuellen Standard für die Bewertung von Implantaten [54].

Heutzutage stehen verschiedene Implantattypen für unterschiedliche Indikationen zur Verfügung. Diese Implantate lassen sich in verschiedenen strukturellen Ebenen klassifizieren. Aus makroskopischer Perspektive betrachtet, erfolgt die Einteilung nach Kriterien wie der Implantatgeometrie, der Anzahl der Implantatkomponenten, der Relation zur Gingiva während der Einheilphase sowie dem Gewindedesign [55]. Hinsichtlich der Geometrie wird im Wesentlichen zwischen parallelwandigen, teilkonischen und konischen Implantaten differenziert [56]. Letztere zeichnen sich durch ihre Verjüngung von koronal nach apikal aus, wodurch sie die Form einer natürlichen Zahnwurzel nachahmen.

Einteilige Implantate werden als solche beschrieben, bei denen das Implantat fest mit der Mesostruktur (Abutment) verbunden ist, sodass eine Einheit bestehend aus Implantatschraube und der Suprakonstruktion vorliegt. Bei zweiteiligen Implantaten hingegen bestehen das Implantat und das Abutment aus separaten Komponenten, die mittels einer Schraubenverbindung fixiert werden [57].

Auf mikroskopischer Ebene betrachtet, erfolgt die Klassifikation von Implantaten primär nach dem verwendeten Material und der Oberflächenbeschaffenheit. Die am häufigsten eingesetzten Materialien sind Titan und Keramik, wobei diese in unterschiedlichen Zusammensetzungen verwendet werden. Titanimplantate weisen je nach Sauerstoffgehalt in der Legierung unterschiedliche mechanische Eigenschaften auf: Während Implantate aus reinem Titan mit einem geringen Sauerstoffanteil (Grad 1) eine vergleichsweise geringere mechanische Festigkeit aufweisen, sind Implantate aus höhergradigem Titan (Grad 4) deutlich bruchfester. Diese erhöhte Bruchfestigkeit geht jedoch zulasten der Wechselbiegebelastung [58]. Darüber hinaus kommen zunehmend Implantate aus Titan-Zirkon-Legierungen zum Einsatz, die eine verbesserte Festigkeit mit sich bringen.

Keramikimplantate bestehen heute überwiegend aus yttriumstabilisiertem tetragonalem polykristallinem Zirkoniumoxid (Y-TZP). Dieses Material zeichnet sich durch seine hohe Härte und Biegefestigkeit aus, welche durch die Beimischung von Aluminiumoxid weiter gesteigert werden kann. In aktuellen Studien konnte gezeigt werden, dass die klinischen Erfolgsraten von Zirkonimplantaten mit denen von Titanimplantaten vergleichbar sind [59, 60]. Allerdings bleibt die wissenschaftliche Evidenz zu zweiteiligen Zirkonimplantaten bislang begrenzt und es sind weitere Studien erforderlich, um deren Langzeitergebnisse umfassend zu beurteilen [59]. Folglich kann der Einsatz von Titanimplantaten zum aktuellen Zeitpunkt denen der Keramikimplantate als überlegen bewertet werden und stellt in der Implantologie den Goldstandard dar.

Die Modifikation von Implantatoberflächen erfolgt mittels diverser technischer Verfahren, die insbesondere bei Titanimplantaten Anwendung finden. Eine häufig verwendete Methode zur Oberflächenmodifikation ist das Aufrauen der Implantatoberfläche, das durch den Einsatz mineralischer Säuren oder durch das Abstrahlen mit verschiedenen Medien wie Aluminiumoxid erreicht werden kann. Häufig wird auch eine Kombination dieser Verfahren eingesetzt, um die Oberfläche zu optimieren. Durch die dabei entstehenden Mikrovertiefungen wird das Einwachsen von Osteoblasten begünstigt, wodurch die Osseointegration des Implantats gefördert wird. Eine weitere Technik zur Oberflächenmodifikation besteht in der Erzeugung einer anodischen Oxidschicht auf dem Implantatkörper durch Funkentladung. Alternativ kann Titanpulver direkt auf die Oberfläche des Implantats

aufgeschmolzen werden. Diese beiden Verfahren, die als additive Techniken bekannt sind, bringen jedoch das Risiko der Ablösung von Titanpartikeln mit sich, insbesondere während der Implantatinsertion, was zu potenziellen Komplikationen führen kann [58, 61].

Die Osseointegration von Implantaten kann zu einem großen Teil durch das Makrodesign beeinflusst werden. Hinsichtlich der Frage, ob konische Implantate im Vergleich zu parallelwandigen eine höhere Primärstabilität aufweisen, besteht jedoch in der Literatur noch Uneinigkeit [56]. Während einige Autoren darauf hinweisen, dass konische Implantate in dieser Hinsicht Vorteile bieten könnten [62], gibt es auch Studien, die keinen signifikanten Unterschied feststellen konnten [63]. Eine Studie berichtet allerdings von einer erhöhten Stabilität konischer Implantate insbesondere in Knochen mit geringer Dichte, da das Design eine verstärkte Kontaktosteogenese fördert [64].

Neuere Implantatdesigns zielen zunehmend darauf ab, die Osseointegration zu beschleunigen. Hierbei wird oft empfohlen, spezielle Bohr- und Insertionsprotokolle zu befolgen, um eine übermäßige Kompression des Knochens während der Implantatinsertion zu vermeiden. Bei einigen Implantattypen wird das Makrodesign durch das Einfügen spezieller Kammern modifiziert, die als Vertiefungen auf der Implantatoberfläche fungieren. Die Dekompression des Knochengewebes während der Insertion erfolgt automatisch, da das Implantat aufgrund der Kammern ein reduziertes Einsetzdrehmoment aufweist [65, 66, 67, 68].

2.2.2 Replikat dentaler Implantate

Replikat und preisgünstigere Alternativen, die laut den jeweiligen Herstellern mit den Originalprodukten kompatibel sein sollen, gewinnen zunehmend an Marktanteil im Bereich der dentalen Implantologie [69]. Eine Analyse aus dem Jahr 2019 prognostiziert, dass bis 2025 nahezu dieselbe Anzahl verkaufter Einheiten von kostengünstigeren Herstellern erreicht werden könnte wie bei den etablierten Originalherstellern, die ein Oligopol aus fünf bis sechs Unternehmen bilden [70].

Der stetige Zuwachs dieses Marktes birgt jedoch gewisse Risiken. Es besteht die Gefahr, dass kostengünstigere Hersteller aufgrund des starken Wettbewerbs aus

dem Markt auch schnell wieder ausscheiden könnten. Dies hätte zur Folge, dass die Beschaffung von Komponenten für zukünftige Reparaturen oder Nachbehandlungen erschwert wird. Zudem besteht das Risiko eingeschränkter Verfügbarkeiten, da entsprechende Hersteller möglicherweise nicht über ausreichende Kapazitäten verfügen, um auf schnelle Nachfragen in großen Mengen zu reagieren. Ein weiteres Problem ergibt sich aus der zunehmenden Vielfalt der Implantatsysteme und Hersteller. Es wird immer schwieriger, die einzelnen Systeme eindeutig zu identifizieren, was vor allem bei der Analyse von Röntgenaufnahmen von Bedeutung ist. Diese Identifikation ist jedoch notwendig, wenn Unsicherheit über das verwendete Implantatsystem besteht und spezifische Komponenten für Reparaturen oder Weiterbehandlungen bestellt werden müssen [69].

Obwohl einige Hersteller angeben, dass ihre Komponenten mit Implantatsystemen anderer Marken kompatibel seien, kann diese Kompatibilität nicht immer hinreichend gewährleistet werden. In einigen Ländern können dentale Implantate ohne strenge Einhaltung von Qualitätsstandards produziert und vertrieben werden, was das Risiko von Fehlpassungen erhöht [71]. Insbesondere nachgeahmte Produkte, wie etwa Abutments, weisen häufig signifikante Abweichungen vom Original auf. Der Einsatz solcher Replikat führt zudem zum Erlöschen der Herstellergarantie, da diese nur bei der Verwendung von Originalkomponenten bestehen bleibt [72].

Eine systematische Übersichtsarbeit hat aufgezeigt, dass es bei der Verwendung von nachgeahmten Abutments in Verbindung mit den entsprechenden Originalimplantaten zu signifikanten Passungsabweichungen kommt [73]. Insbesondere wurde eine erhöhte Anzahl an Mikrobewegungen bei nachgeahmten Abutments im Vergleich zu den Originalkomponenten festgestellt [74]. Darüber hinaus weisen nachgeahmte Abutments eine signifikant höhere Rate an Rotationsfehlpassungen auf [73].

Mehrere Studien konnten außerdem nachweisen, dass das Risiko einer Schraubenlockerung bei der Verwendung von Originalabutments deutlich geringer ist [75, 76, 77, 78, 79]. Fehlpassungen, die im Rahmen standardisierter Ermüdungstests als maßgeblicher Einflussfaktor identifiziert wurden, tragen wesentlich zu Komplikationen wie Schraubenlockerungen oder Schraubenbrüchen bei [80, 81].

Es lässt sich mit großer Sicherheit annehmen, dass kompatible Alternativanbieter, deren Produktionsprozesse nicht denselben strengen Qualitätsstandards unterliegen

wie die der Originalhersteller, größere Abweichungen in der Passgenauigkeit aufweisen könnten [80].

Derzeit liegen nur unzureichende Studienergebnisse in Bezug auf Mikroleckage und In-vivo-Ergebnisse vor [73]. Dennoch wird weiterhin die Verwendung von Originalabutments empfohlen, was im Einklang mit den Ergebnissen mehrerer Studien steht, die für Originalabutments günstigere Resultate in der Langzeitüberlebensrate der prothetischen Suprakonstruktion zeigen [81, 82, 83, 84].

Implantatprothetische Komplikationen wie Schraubenlockerung und -fraktur stellen neben Keramikfrakturen (Chipping) die häufigsten Komplikationen dar [85, 86]. Entsprechend wird ein stärkerer Anstieg der Komplikationen beim Verwenden von Replikat-Produkten in Kombination mit Original-Implantaten erwartet, da Replikate häufig nicht die gleiche Präzision in der Passgenauigkeit aufweisen wie Originalprodukte. Folglich sollten Zahnmediziner wissen, um welches System es sich handelt, um ein unnötig gesteigertes Komplikationsrisiko zu vermeiden. Die Garantie des Herstellers ist ausschließlich unter der Voraussetzung gültig, dass keine Kombination des jeweiligen Original-Produktes mit Produkten anderer Hersteller stattgefunden hat und dass die entsprechende Gebrauchsanweisung befolgt worden ist. Die Hersteller gewähren sogar eine lebenslange Garantie, sofern Originalteile verwendet wurden. Diese Garantie umfasst nicht nur das Implantat selbst, sondern die gesamte Konstruktion, sodass bei Bedarf auch das Sekundärteil ersetzt werden kann [87]. Dies deckt sich mit den Interessen des Behandlers, da sonst dieser im Garantiefall für den Schaden aufkommen müsste, da eine gesetzliche Gewährleistungspflicht laut § 634a BGB von zwei Jahren vorgesehen ist, sofern Mängel auf unsachgemäßes Arbeiten oder fehlerhaftes Material zurückzuführen sind. Außerdem wird so die langfristige Gesundheit der Patienten sowie eine erfolgreiche Nachsorge gewährleistet.

3 Material und Methoden

3.1 Studiendesign

Ziel dieser Studie war die Klassifizierung von zwei sehr ähnlichen Implantaten zweier verschiedener Hersteller anhand von Datensätzen, die aus extraoralen Röntgenaufnahmen dieser Implantate segmentiert wurden. Diese in vitro Versuche wurden unter Verwendung mehrerer CNNs (ResNet - Residual Neural Network) bewertet. In einem weiteren Schritt bewerteten menschliche Experten denselben Testdatensatz, der auch zur Überprüfung der Genauigkeit der KI verwendet wurde. Die erzielten Ergebnisse wurden anschließend miteinander verglichen. Die Studie erhielt die Zustimmung der lokalen Ethikkommission des Landes Rheinland-Pfalz (Nr.: 2023-16979) und wurde gemäß dem festgelegten Protokoll durchgeführt. Dabei wurden die moralischen, ethischen und wissenschaftlichen Grundsätze für klinische Forschung beachtet, wie sie in der Deklaration von Helsinki von 1975, in der überarbeiteten Fassung von 1983, festgelegt sind.

3.2 Implantate

Die Implantate, die dieser Arbeit als Grundlage für den Vergleich dienten, sind Produkte der Firmen Straumann® und Meisinger®. Dabei handelt es sich bei dem Straumann-Implantat (Implantattyp Standard Plus) um das Originalprodukt und das entsprechende Replikat stammt von Meisinger (Abbildung 7). Sie weisen identische Werte in Form, Größe und Durchmesser auf. Die Implantatmodelle verfügen jeweils über eine enossale Länge von 10 mm. Sie umfasst den Bereich des Implantates, der gänzlich in den Knochen im Zuge der Implantatinsertion eingebracht wird. Die Implantatschulter ist bei beiden Implantaten 1,8 mm lang und endet auf Höhe des Weichgewebes (Tissue-Level). Sie stellt die Zone des Implantates dar, in der die Verbindung zwischen Implantat und Abutment erfolgt, auf dem wiederum eine entsprechende Suprakonstruktion angebracht werden kann (z. B. eine künstliche Zahnkrone). Der enossale Durchmesser beträgt 4,1 mm und der der Implantatplattform 4,8 mm. Gemessen wird jener an der breitesten Stelle der Implantatschulter.

Aufgrund der beworbenen Kompatibilität des Replikat-Implantates und der einzelnen Komponenten mit dem Original-Produkt, sei es dem Behandler möglich, unter

anderem Abdeckschrauben, Abformpfosten, Laboranaloge und Abutments des jeweils anderen Herstellers kongruent und risikofrei zu verwenden.



Abbildung 7 Original- (jeweils links) und Replikat-Implantat (jeweils rechts) mit passendem manuellem Schraubendreher; links inserierte Schraube, rechts ohne inserierte Schraube

In ihrer chemischen Zusammensetzung zeichnen sich Unterschiede ab. Während das Straumann®-Implantat aus einer Hochleistungslegierung namens Roxolid® besteht, ist das Meisinger®-Implantat aus Rein-Titan Grade 4 gefertigt. Die prozentualen Anteile der entsprechenden einzelnen Komponenten sind in den Tabellen 1 und 2 aufgeführt.

Titan-Zirkonium-Legierung:	
Chemische Komponenten	Zusammensetzung % (Masse/Masse)
Zirkonium	15
Reststoffe (Fe, O, C, N, H) insgesamt	≤ 1,05
Titan	Differenz

Tabelle 1 Straumann Roxolid® Implantat

Rein-Titan Grade 4:	
Chemische Komponenten	Zusammensetzung % (Masse/Masse)
Stickstoff (N)	0,05
Kohlenstoff (C)	0,08
Wasserstoff (H)	0,015
Eisen (Fe)	0,5
Sauerstoff (O)	0,4
Titan	Differenz

Tabelle 2 Meisinger® Implantat

3.3 Datenerhebung

3.3.1 Röntgenbilder

Als Datensatzgrundlage für die Programmierung einer auf künstlichen Intelligenz beruhenden Erkennungssoftware, speziell dieser zwei beschriebenen Implantate, diente ein Datensatz bestehend aus 906 Röntgenbildern. Diese wurden ex situ angefertigt, ohne vorherige klinische Verwendung der Implantate. Bis zum Startzeitpunkt der ersten Aufnahme waren sie jeweils originalverpackt. Es wurde ein standardisiertes Verfahren angewandt, sodass die Aufnahmen eine gleichbleibende Qualität aufwiesen. Dabei wurden die Implantate in verschiedenen Abständen, Neigungen und Rotationen zum Detektor in einem für diesen Versuchsaufbau angefertigten Arretierungssystem platziert. Die Abstände variierten von 1 cm bis 15 cm von Objekt zu Detektor, sodass sich unterschiedliche Bildgrößen ergaben. Die Neigungen der Implantate erstreckten sich von 0° bis 45° in 5° Schritten. Diese beziehen sich auf die Neigung des Implantates von der Implantatschulter hin zum Detektor. In jeder Kombination aus Abstand und Neigung wurde das Implantat zusätzlich um 45° , 90° und 135° um die eigene Achse rotiert (Abbildung 8).

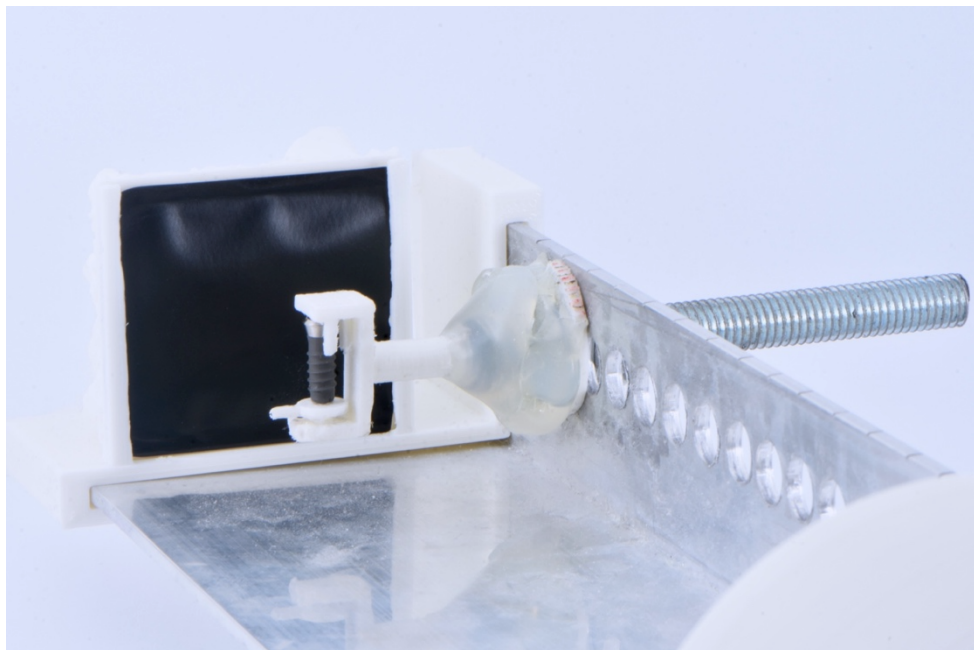


Abbildung 8 Nahaufnahme des Implantats vor dem Detektor

Zur Anfertigung der Röntgenbilder wurde ein Dental-Röntgentubusgerät von der Firma Soredex (Modell MinRay®) mit einem Brennpunkt von 0,7 mm und einer Leistung von 60/70 kV bei 7 mA verwendet. Die eingesetzten Speicherfolien waren vom Typ Dürr VistaScan Plus® mit einer Größe von 3 x 4 cm (Größe 2), welche nach Belichtung mit dem Dürr VistaScan Mini View® Scanner digital in die Bildgebungssoftware Sidexis (Version XG 2.65) eingepflegt wurden.

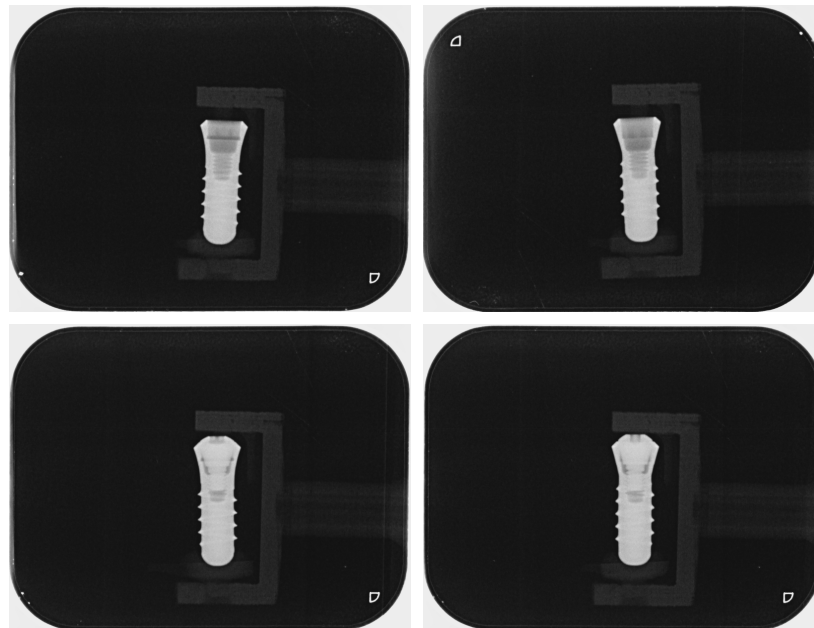
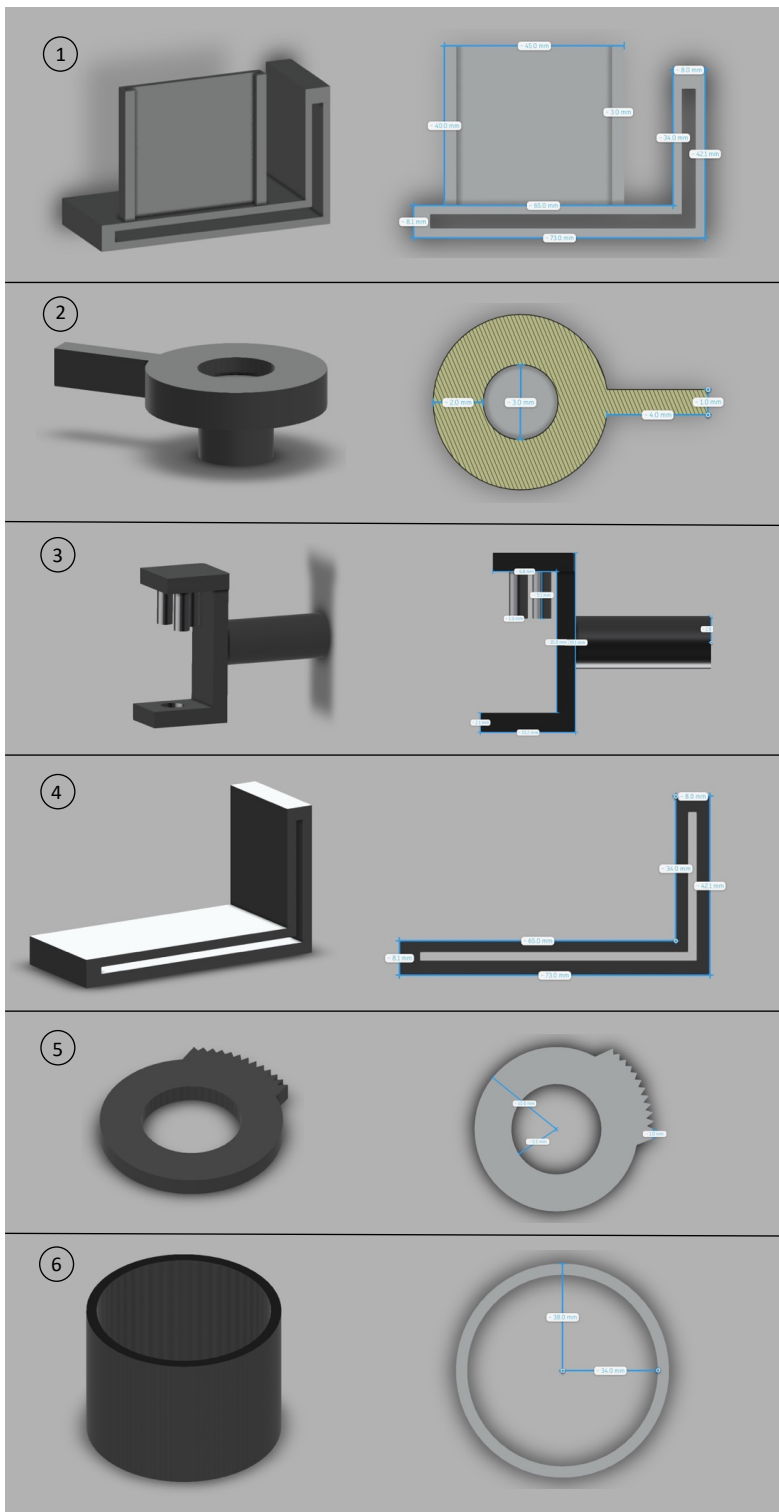


Abbildung 9 Untersuchte Implantate

- Oben links: Straumann Implantat ohne Schraube
- Unten links: Straumann Implantat mit Schraube
- Oben rechts: Meisinger Implantat ohne Schraube
- Unten rechts: Meisinger Implantat mit Schraube

3.3.2 Experimenteller Aufbau

Das standardisierte Verfahren zur Bildgebung basiert auf einer Konstruktion, die es ermöglicht, die Implantate in den verschiedenen beschriebenen Bildtypen in einer konstanten Geometrie zu fixieren. Diese Vorrichtung wurde größtenteils mit einem 3D-Drucker (i3 Mega S, Anycubic®) hergestellt, ergänzt durch ein Winkelprofil aus Edelstahl, in das gleichmäßig verteilte Löcher gefräst wurden (Abbildung 11). In Abbildung 10 sind die detaillierten Zeichnungen und Maße der einzelnen Komponenten dargestellt.



1. Speicherfolienhalter
2. Rotationsmesser
3. Implantathalter
4. Verbinder für Röntgen-tubushalterung
5. Winkelmesser
6. Röntgentubushalterung

Abbildung 10 Konstruktionselemente

Die Planung der Konstruktion erfolgte mit der Software Autodesk® Fusion 360™. Die Erstellung der STL-Dateien und deren Umwandlung in einen G-Code, welcher die Steuerbefehle für den 3D-Drucker enthält, wurde mit der Software UltiMaker Cura™ (Version 4.11.0, Ultimaker B.V.®) durchgeführt. Anschließend wurden die Druckbefehle an einen 3D-Drucker übertragen, der die Komponenten aus PLA-

Filament (Geeetech® 3D Printer Filament, PLA White, 1,75 mm/1000 g) fertigte. Nach der Fertigung wurden die gedruckten Teile mit dem Edelstahlwinkelprofil verbunden.

Der Implantathalter wurde mittels thermoplastischen Klebstoffes an einer Schraube fixiert, welche in die gefrästen Löcher des Winkelprofils eingesetzt und individuell abhängig vom Abstand zum starr positionierten Detektor (Speicherfolie) mit einer Schraubenmutter befestigt werden konnte. Über jedem dieser Löcher befindet sich eine zentrale Einkerbung, die in Verbindung mit einem Winkelmesser die korrekte Neigung des Implantathalters entsprechend der vorgegebenen Bildgeometrie ermöglichte. Der Winkelmesser selbst ist über eine weitere Schraubenmutter an der Schraube befestigt, die zur Stabilisierung dient und ebenso mit thermoplastischem Klebstoff fixiert ist (Abbildung 8). Das jeweilige Implantat wurde mithilfe von Weichwachs an einem speziellen Bauteil fixiert (Rotationsmesser), das sowohl eine präzise Positionierung im Implantathalter ermöglicht als auch die Rotation des Implantates um die eigene Achse erlaubt. Für jedes Implantat war ein Rotationsmesser vorgesehen, an dem es dauerhaft fixiert war, um eine gleichbleibende Ausrichtung des Implantates sicherzustellen. Am gegenüberliegenden Ende des Winkelprofils befindet sich ein Halter für den Röntgentubus, der ebenfalls eine feste Positionierung gewährleistete, welcher in Abbildung 11 mit der Gesamtkonstruktion zu sehen ist.

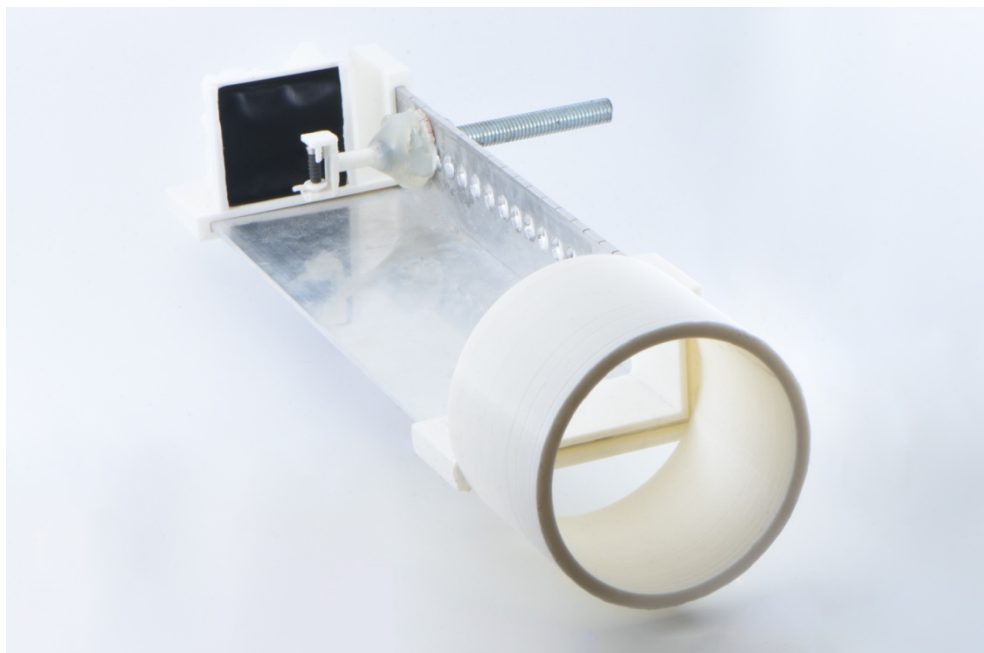


Abbildung 11 Gesamtkonstruktion für die standardisierte Bildgebung

Diese Konstruktion stellte sicher, dass die Implantate stets im gleichen Abstand zum Detektor positioniert waren und dabei eine reproduzierbare, definierte Geometrie einhielten. Die Bildgebung erfolgte dadurch unter standardisierten Bedingungen und war konsistent.

3.4 Programmierung der künstlichen Intelligenz

In dieser Studie wurde das Convolutional Neural Network (CNN)-Standardmodell ResNet verwendet. Die genaue Bezeichnung des Modells in diesem Fall ist ResNet-50-V2. Generell liegen verschiedene Modelltypen bei der Programmierung eines residuellen Netzwerks (ResNet) zugrunde. Die für diese Studie verwendeten drei Modelle nutzten die sogenannte deterministische Programmierung. Dabei werden die Gewichte des neuronalen Netzwerkes als ein einziger skalarer Wert für jeden Parameter geschätzt. Sobald diese Gewichte erlernt sind, verhält sich das Modell deterministisch in dem Sinne, dass es jedes Mal, sobald eine identische Eingabe erfolgt, die gleichen Ergebnisse liefert, da die Gewichte als Konstanten feststehen.

Die Modellschulung erfolgte unter Verwendung der K-Fold Cross Validation im Modellschulungsalgorithmus. Der Trainingsalgorithmus für die Röntgenbilder ausschließlich mit inserierter Implantatschraube nutzte $K = 5$ für die K-Fold Cross Validation, der Algorithmus für die Bilder mit und ohne inserierte Schraube nutzte ebenfalls $K = 5$. Die Trainingsdaten für das Modell, welches ausschließlich Röntgenbilder mit inserierter Schraube inkludiert, bestanden aus 488 Bildern. 244 Bilder lagen jeweils pro Implantat vor. Die beiden weiteren 5-Fold Cross Validation-Modelle wurden einmal mit einem Datensatz bestehend aus 674 Bildern und einmal bestehend aus 906 Bildern trainiert. Die 674 Bilder bestanden aus den bereits vorhandenen 488 Bildern sowie aus weiteren 186 Aufnahmen ohne inserierte Schraube. Beim Datensatz aus insgesamt 906 Bildern wurden die Aufnahmen ohne Schraube um weitere 232 Bilder auf 418 erweitert.

Zur Evaluierung der Genauigkeit der verschiedenen Modelle wurde ein gänzlich neuer Datensatz (Testdatensatz) mit 48 Röntgenaufnahmen der Implantate sowohl mit als auch ohne Schraube erstellt.

3.5 Studienablauf im Detail

3.5.1 Studienteilnehmer

Die Teilnehmergruppe setzte sich aus Personen, die im Bereich der Zahnmedizin tätig sind, zusammen, darunter sowohl Studierende aus dem zweiten klinischen bis zum fünften klinischen Semester als auch praktizierende Zahnärzte aus einer Universitätsklinik. Ihre Erfahrung in der Zahnmedizin variierte von mindestens dreieinhalb bis zu maximal neun Jahren (Durchschnitt = 5,29 Jahre). Insgesamt nahmen 28 zahnmedizinisch tätige Experten an der Befragung teil. Hinsichtlich der Geschlechterverteilung setzte sich die Gruppe aus 15 weiblichen und 13 männlichen Personen zusammen.

3.5.2 Studiendurchführung

Vor der Befragung betrachteten die Teilnehmer zunächst vier analoge Beispielbilder aus dem Trainingsdatensatz (Anhang 8.1). Jedes Bild zeigte das Implantat in einem Abstand von 1 cm zum Detektor, bei einer Neigung von 0° und einer Rotation um 45° um die eigene Achse. Die Beispielbilder repräsentierten jeweils eine der vier in dieser Studie einbezogenen Bildkategorien (siehe Kapitel 3.3.1). Für diese erste Betrachtung hatten die Teilnehmer vier Minuten Zeit, um sich mit den Eigenschaften der Implantate vertraut zu machen. Ein diagnostisches Display sowie die Befundungssituation befanden sich in einer Umgebung, die einem zahnmedizinischen Bildarbeitsplatz entsprach und somit der Raumklasse 5 zugeordnet werden konnte. Vor Beginn der Befragung wurde der Raum abgedunkelt, sodass die Raumbeleuchtungsstärke auf unter 100 Lux reduziert wurde, um die radiologische Befundung zu gewährleisten.

Anschließend wurden den Teilnehmern die Röntgenbilder des Testdatensatzes in zufälliger Reihenfolge präsentiert. Während dieser Phase konnten die Teilnehmer die Beispielbilder nicht zur Orientierung heranziehen. Jedes Röntgenbild wurde für die Dauer von einer Minute angezeigt. Den Teilnehmern war es gestattet, eigenständig durch einen Mausklick zum nächsten Bild fortzuschreiten, jedoch konnten sie nicht mehr zurück. Es wurde ein Betrachtungsmonitor von der Marke Fujitsu mit der Modellbezeichnung B22W-6 und einer Bildschirmauflösung 1680 x 1050 (WSXGA+) verwendet.

Bei jedem Bild mussten die Teilnehmer entscheiden, ob es sich um ein Original- oder um ein Replikat-Implantat handelt. Zudem wurden sie gebeten, ihr Konfidenzniveau anzugeben, um es mit den Ergebnissen der künstlichen Intelligenz vergleichen zu

können. Dazu nutzten sie eine visuelle Skala von 0 bis 100 %, unterteilt in 5 %-Schritte, um ihre Einschätzung zur Sicherheit ihrer Entscheidung anzugeben. Eine Einschätzung von 0 % sagte aus, dass der Studienteilnehmer keinerlei Sicherheit in Bezug auf die Korrektheit seiner Entscheidung empfand, während eine Bewertung von 100 % eine vollkommene Gewissheit hinsichtlich der Richtigkeit der getroffenen Entscheidung ausdrückte (Abbildung 12).

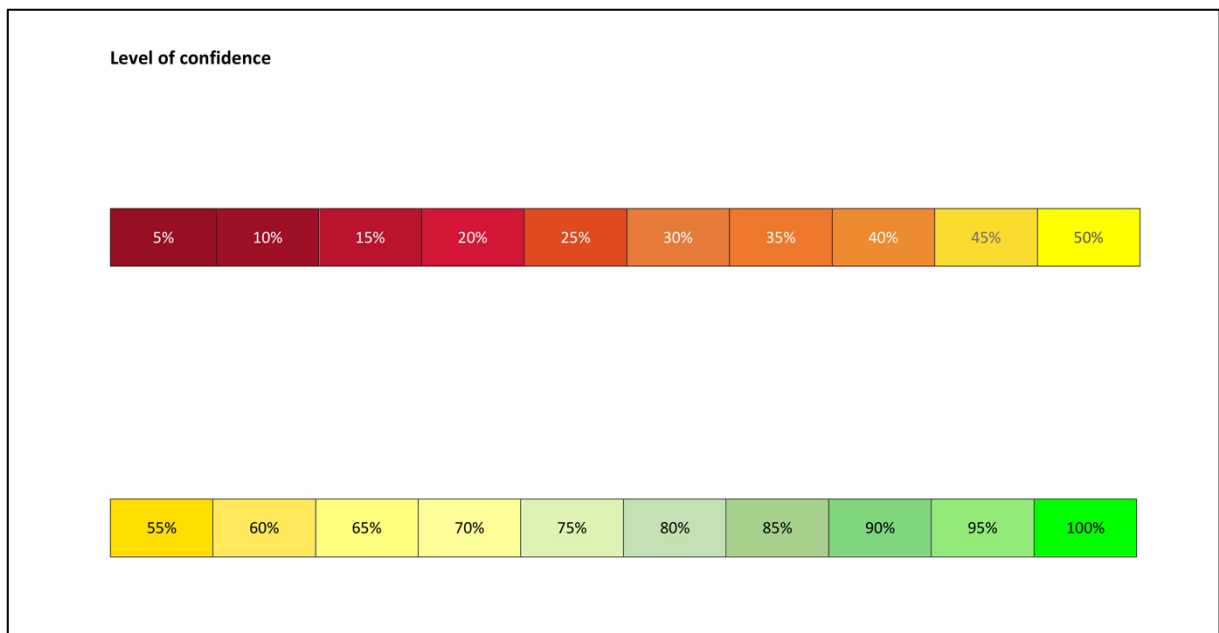


Abbildung 12 Visuelle Orientierung zum Einordnen der Konfidenzniveaus der Teilnehmer

Die Antworten sowie die angegebenen Konfidenzniveaus wurden tabellarisch erfasst. Zusätzlich wurde das Geschlecht der Teilnehmer, die Anzahl der Tätigkeitsjahre in der Zahnmedizin und die Angabe, ob sie mehr Zeit benötigt hätten, dokumentiert.

4 Ergebnisse

4.1 Ergebnisse des Künstlichen Neuronales Netzwerks

Die Genauigkeit des deterministischen 5-Fold-Cross-Validation-Modells, welches rein mit 488 Aufnahmen von Implantaten mit inserierter Schraube trainiert wurde, beträgt im Gesamten 57,9 % bezogen auf den Testdatensatz. Für die Röntgenaufnahmen des Originalimplantates beträgt sie 54,2 %. Das Replikat wird mit einer Genauigkeit von 61,7 % korrekt erkannt (Tabelle 3).

	% Korrekt	χ^2	df	p
Total	57.92			
Original	54.17			
Replikat	61.67	6.05	1	.014

Tabelle 3 (false classifier = 5): Trainingsdatensatz = 488

Bilder mit Schraube = 488, Bilder ohne Schraube = 0

Das zweite deterministische Modell, welches den Datensatz aus 674 Bildern umfasste, ordnete mit einer Genauigkeit von >90 % bei allen fünf Klassifikatoren die Röntgenaufnahmen des Testdatensatzes korrekt zu (Abbildung 13). Die fünf Scores in den Abbildungen 13 und 14 stellen die Ergebnisse der unterschiedlichen Klassifikatoren dar, die pro KI Modell die Testdatensätze jeweils evaluiert haben. Jeder Klassifikator gibt für eine bestimmte Eingabe eine Wahrscheinlichkeit (Score) an, die darlegt, wie sicher sich das Modell ist, dass die Eingabe zu einer bestimmten Klasse gehört.

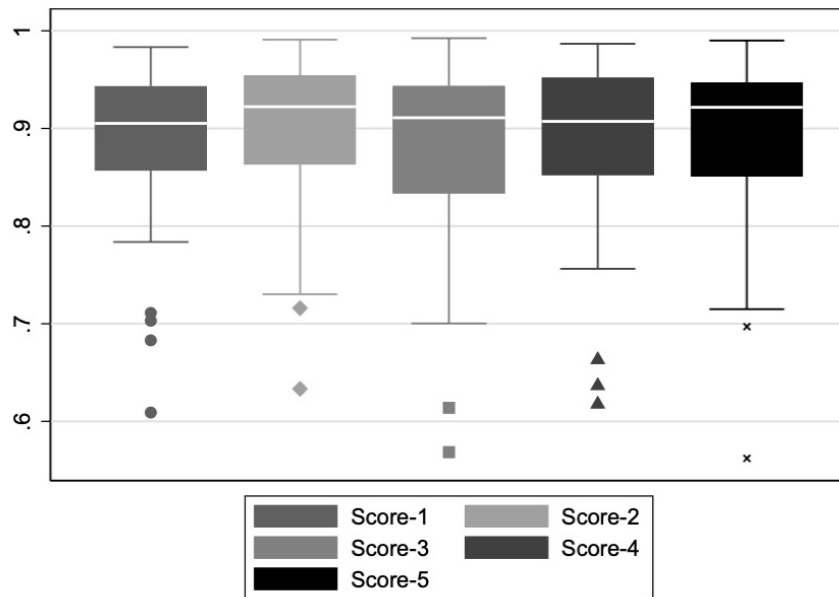


Abbildung 13 Trainingsdatensatz = 674

Bilder mit Schraube = 488, Bilder ohne Schraube = 186

Im Gegensatz dazu konnte beim dritten Modell die Genauigkeit signifikant gesteigert werden, indem die Anzahl der Röntgenbilder ohne inserierte Schraube um mehr als das doppelte auf 418 Stück erhöht wurde. Dadurch ergab sich beim zweiten und dritten Klassifikator eine Genauigkeit von 100 % (Abbildung 14).

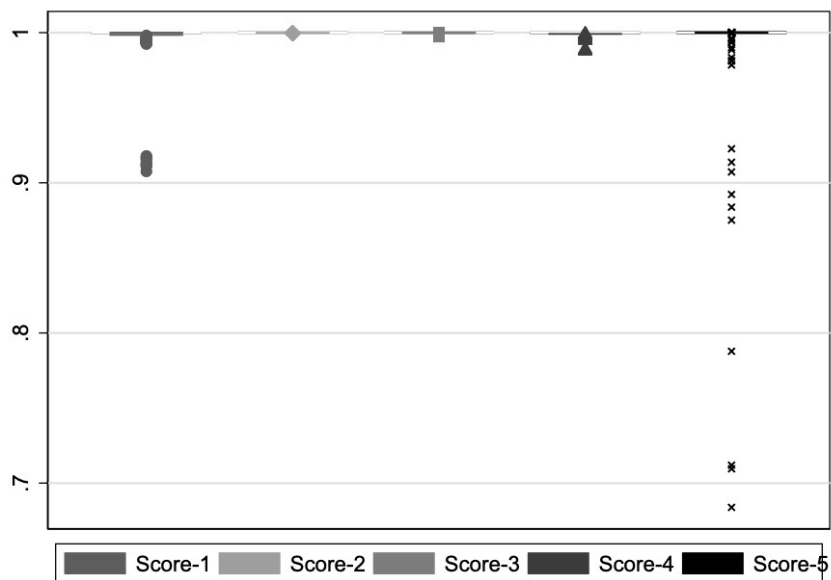


Abbildung 14 Trainingsdatensatz = 906

Bilder mit Schraube = 488, Bilder ohne Schraube = 418

4.2 Ergebnisse der Studienteilnehmer

Die Studienteilnehmer zeigten eine Gesamtgenauigkeit von 86,16 % bezogen auf den Testdatensatz. Beim Original-Implantat lag eine höhere Trefferquote vor, da eine Genauigkeit von 88,69 % erreicht wurde. Das Replikat wurde im Gegensatz dazu mit einer Genauigkeit von 83,63 % erkannt. Die einzelnen Ergebnisse sind in Tabelle 4 dargestellt.

	% Korrekt	χ^2	df	p
Total	86.16			
Original	88.69			
Replikat	83.63	705	1	<.001

Tabelle 4 Genauigkeitswerte der menschlichen Experten

Zwischen den Ergebnissen der weiblichen und der männlichen Studienteilnehmer gab es Unterschiede. Die weiblichen Studienteilnehmer erzielten eine Gesamtgenauigkeit von 84,86 %, wobei eine um ca. 8 % höhere Genauigkeit des Original-Implantats verglichen zum Replikat-Implantat vorlag (Tabelle 5). Die männlichen Teilnehmer wiesen eine Gesamtgenauigkeit von 87,66 % auf. Das Original-Implantat wurde bei dieser Gruppe mit einer 1,6 %igen höheren Genauigkeit erkannt (Tabelle 6).

	% Korrekt	χ^2	df	p
Total	84.86			
Original	88.89			
Replikat	80.83	352	1	<.001

Tabelle 5 Genauigkeitswerte der weiblichen Experten

	% Korrekt	χ^2	df	p
Total	87.66			
Original	88.46			
Replikat	86.86	354	1	<.001

Tabelle 6 Genauigkeitswerte der männlichen Experten

Zudem zeigte sich, dass die einzelnen Konfidenzniveaus der Gruppen unterschiedlich waren. Die Verteilung der Konfidenzniveaus der weiblichen Teilnehmer zeigte eine breite Streuung über die Skala (Abbildung 15), was anzeigt, dass viele Frauen mittlere bis hohe Konfidenz in ihren Urteilen hatten. Es gibt jedoch relativ wenige Frauen, die eine maximale Sicherheit (100 %) angaben.

Im Vergleich dazu zeigte die Verteilung der Konfidenzniveaus bei den männlichen Teilnehmern eine stärkere Konzentration bei höheren Konfidenzniveaus. Besonders bei einem Konfidenzwert von 100 % gab es eine hohe Dichte. Dadurch wird deutlich, dass einige Teilnehmer ihre Urteile mit maximaler Sicherheit getroffen hatten.

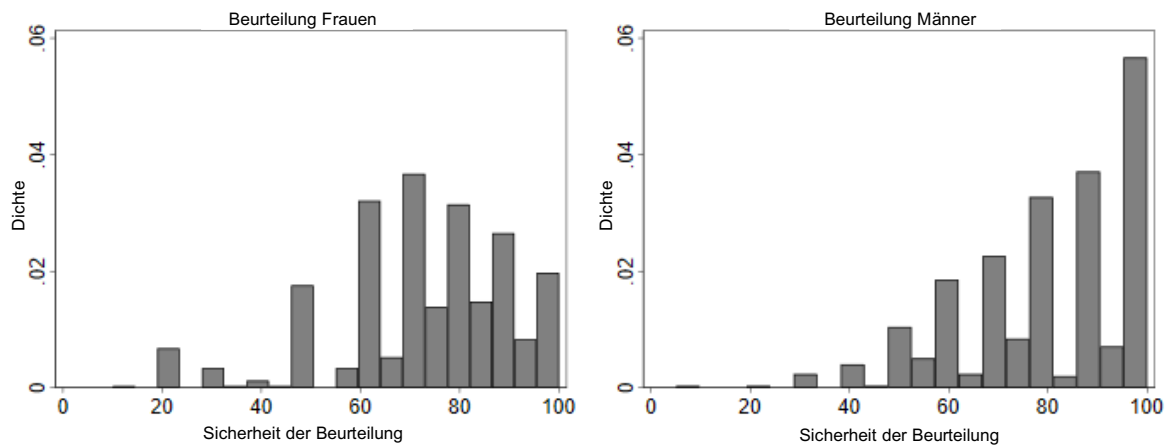


Abbildung 15 Konfidenzniveaus der männlichen und weiblichen Studienteilnehmer

5 Diskussion

In den vergangenen Jahren hat die künstliche Intelligenz eine exponentielle Entwicklung erlebt und zeigt auch in der zahnmedizinischen Implantologie immer weiter verbesserte Vorhersagewahrscheinlichkeiten in der Erkennung von Implantatsystemen verschiedener Hersteller. Einige Studien belegten, dass bereits aussagekräftige Ergebnisse mit kleinen Trainingsdatensätzen mittels einer KI erzielt werden konnten [9, 48]. Insbesondere in spezifischen Fragestellungen, die ein besonderes Maß an Expertenwissen erfordern, kann sie eine wertvolle Unterstützung in der Entscheidungsfindung darstellen [88, 89].

Dies war der Ausgangspunkt für die vorliegende Hypothese, da selbst erfahrene Experten oft vor der Herausforderung stehen, ein Originalimplantat von einem Replikat zu unterscheiden. In der aktuellen Literatur gibt es bislang keine vergleichbare Studie, die den Einsatz von KI zur Identifizierung zweier nahezu identischer Implantate untersucht.

5.1 Ergebnisse im Kontext der Hypothese

Die Ergebnisse der vorliegenden Studie können die aufgestellte Hypothese, dass die KI besser als der menschliche zahnmedizinisch tätige Experte die Implantate röntgenologisch identifizieren kann, widerlegen, sofern die KI nur mit einem Teil des gesamten Datensatzes trainiert wurde. Es zeigte sich, dass bei einem Trainingsdatensatz von nur 488 Bildern, aus dem Bilder ohne inserierte Schraube exkludiert waren, die Genauigkeit der KI geringer war als die des menschlichen Experten. Im Gegensatz dazu kann die Hypothese bestätigt werden, wenn die KI mit größeren und umfassenderen Datensätzen trainiert wurde, die 674 bzw. 906 Bilder umfassten und auch solche mit inserierter Schraube beinhalteten. Diese Ergebnisse verdeutlichen den erheblichen Einfluss der Datensatzgröße und -vielfalt auf die Leistungsfähigkeit eines auf Deep Learning basierenden Modells. Je umfangreicher und repräsentativer der Datensatz ist, desto besser kann die KI entsprechende Muster erkennen und eine vergleichbare oder sogar überlegene Leistung gegenüber dem menschlichen Experten erbringen. Dies unterstreicht die Notwendigkeit, bei der Entwicklung KI-gestützter Diagnosesysteme auf ausreichend große und diversifizierte Trainingsdatensätze zurückzugreifen, um eine valide und klinisch relevante Anwendung zu gewährleisten.

5.2 Relevanz der inkludierten Implantate

Das Originalprodukt (Straumann-Implantat) gehört weltweit zu den am häufigsten verwendeten Implantatsystemen, was anhand der Deklaration der Straumann AG dieses Implantats als „Bestseller“ [90] und der führenden Position im globalen dentalen Implantatmarkt abzuleiten ist [91]. Diese Marktdominanz verdeutlicht die klinische Relevanz, speziell dieses Implantatsystems in einer ersten Pilotstudie mit einem entsprechenden Replikat zu vergleichen. Aufgrund der Verbreitung des Straumann-Systems ist es häufig der Fall, dass zahlreiche Drittanbieter ihre Produkte als kompatibel zu diesem System anbieten. In der Praxis führt dies jedoch häufig zu Passungsungenauigkeiten und daraus resultierenden Komplikationen, insbesondere bei der kritischen Schnittstelle zur Implantatsuprakonstruktion.

Vor diesem Hintergrund ist es von großer Bedeutung, die Realisierbarkeit zu prüfen, ob zwei röntgenologisch nahezu identische Implantate mithilfe von KI zuverlässig unterschieden werden können. Ein solcher Ansatz könnte langfristig dazu beitragen, die potenziellen Risiken von Nachahmerprodukten zu minimieren und den Behandlern mehr Sicherheit in Bezug auf die verwendeten Komponenten und ihre Behandlung zu bieten. Dies wäre ein bedeutender Schritt zur Verbesserung der klinischen Kontrolle über die eingesetzten Implantate und könnte langfristig die Anzahl der Fehlpassungen und zu erwartenden häufigen technischen Komplikationen wie beispielsweise Schraubenlockerungen verringern. Ein solcher Einsatz von KI könnte dazu beitragen, den wachsenden Markt der Drittanbieter kritisch zu überwachen und die Sicherheit und Qualität in der Implantologie weiter zu erhöhen. Insgesamt ist durch die fortschreitende Digitalisierung und auslaufende analoge Dokumentationsform von einer suffizienten Datenübermittlung auch bei Praxisinhaberwechseln auszugehen, jedoch stehen Zahnmediziner regelmäßig vor dem Problem der gesicherten Implantatidentifikation bei nicht vorliegendem Implantatpass [92, 93, 94].

Von den Replikat-Herstellern wird eine Kompatibilität mit den Instrumenten, Abdeckschrauben und Abutments ihrer Systeme mit denen des entsprechenden Originals teilweise beworben, was in der Realität allerdings nicht zuverlässig der Fall ist [74, 82]. Es kommt zu vermehrten Bildungen von Mikrospalten und daraus entstehenden Mikrobewegungen. Dieser Effekt wird verstärkt, je nachdem welcher Implantat-Abutment-Verbindungstyp gewählt wird bzw. welcher bei dem vorliegenden System vorgesehen ist. Es zeigte sich, dass Mikrobewegungen bei einem plattformbasierten Verbindungstyp verstärkter auftreten als bei einer konischen

Verbindung [95]. Ebenso erlaubt ein Originalabutment sehr wahrscheinlich keine präzise Passung auf einem Replikat-Implantat aufgrund der unterschiedlichen Fertigungstoleranzen der Hersteller.

Die Garantie des Herstellers erlischt, wenn nicht-originale Mesostrukturen verwendet werden oder eine falsche Wahl der Suprakonstruktion getroffen wird. Technische Fehler wie Lockerung oder im schlimmsten Fall Bruch der Schraube können auftreten [75, 96] und neben einem hohen medizinischen Risiko wie der Implantatfraktur, ebenfalls mit einem wirtschaftlichen Schaden einhergehen. Zudem kann ein vergrößerter Mikrospace zu verstärkten Pumpbewegungen führen, wodurch mehr Flüssigkeit, welches sich in den Hohlräumen des Implantats befindet, in Kontakt mit dem krestalen Knochen kommt, was zu einer Kontamination desselbigen führen könnte [97, 98, 99]. Diese möglichen Fehler können sich negativ auf die Gesundheit und Zufriedenheit des Patienten auswirken, insbesondere wenn Arbeitsvorgänge wiederholt werden müssen, um eine ausreichende Passgenauigkeit zu erreichen. Darüber hinaus wird die Arbeit des behandelnden Zahnarztes beeinträchtigt, was zu zusätzlicher und zeitaufwändiger Mehrarbeit sowie möglicher Unzufriedenheit seitens des Patienten führen kann.

Entsprechend kann bei etwaigen Unsicherheiten keine ausreichende medizinische Versorgung garantiert werden. Die Unsicherheiten im Umgang mit Replikaten bestehen vor allem darin, dass diese auf Röntgenbildern nur schwer eindeutig zu identifizieren sind, wodurch oft unklar bleibt, um welches Implantatsystem es sich konkret handelt [45]. Dies liegt insbesondere daran, dass es mittlerweile über mehrere tausend international auftretender Implantatmodelle gibt, was die Rückverfolgung des Herstellers anhand radiologischer Aufnahmen erheblich erschwert [100].

Ein zusätzlicher, entscheidender Faktor ist die erschwerte Kontaktaufnahme mit dem ursprünglichen Behandler, da dieser sich oft im Ausland befindet [100] oder nicht mehr verfügbar ist, beispielsweise aufgrund von Ruhestand. Diese Gegebenheiten erschweren die Versorgung der Patienten erheblich, da der Zugang zu den notwendigen Informationen über das verwendete Implantatsystem fehlt.

Zudem ist die Verfügbarkeit von Komponenten des Replikatherstellers nicht immer gewährleistet. Es besteht die Gefahr, dass solche Hersteller in kleinen Stückzahlen produzieren, was die Lieferung verzögern kann, oder dass das Unternehmen nicht mehr existiert [69]. In solchen Fällen ist es von großer Bedeutung, schnell zu handeln

und die entsprechenden Komponenten zeitnah zu bestellen, um eine längere unversorgte Situation des Patienten zu vermeiden.

Vor diesem Hintergrund wird die Bedeutung eines Implantatpasses besonders deutlich. In Anbetracht der zunehmenden Vielfalt an Implantatherstellern ist es heutzutage unerlässlich, dass jeder Patient nach der Implantation einen zugehörigen Implantatpass erhält. Dieser sollte wesentliche Informationen wie den Implantathersteller, den Ort der Behandlung, das Datum der Insertion und die Details zum verwendeten Implantat sowie zu den benötigten prothetischen Komponenten enthalten [69]. Es ist jedoch zu diskutieren, dass dentale Implantate gemäß § 3 in Artikel 18 der europäischen Medizinprodukteverordnung (EU 2017/745) lediglich als zertifizierte Medizinprodukte der Klasse IIb klassifiziert sind, anstatt der Klasse III (außer Implantate mit bioaktiver Oberfläche) zugeordnet zu werden, wie es beispielsweise bei Stents der Fall ist. Diese Einstufung birgt Gefahren, da die Klasse IIb von der Verpflichtung zur Aushändigung eines Implantationsausweises ausgenommen ist. Ein solcher Implantatpass würde allerdings sicherstellen, dass im Falle einer Nachsorge oder Reparatur schnell auf die notwendigen Informationen zugegriffen werden kann, wodurch die Behandlung effizienter und sicherer gestaltet werden kann. Bei einem Verlust des benachbarten Zahnes kann auf diese Weise die Erweiterung der prothetischen Versorgung gewährleistet werden. Im Fall einer Re-Implantation oder einer neu benötigten Suprakonstruktion können in kurzer Zeit entsprechende prothetische und chirurgische Komponenten beschafft bzw. nachfolgende Schritte wie Überweisungen unmittelbar an Fachzahnärzte mit allen notwendigen Daten eingeleitet werden.

Allerdings ist die Chargennummer der bei einem Patienten verwendeten Implantate aus Gründen der Rückverfolgbarkeit in der Dokumentation des Behandlers zu erfassen. Auf diese Weise können über die einzelnen Hersteller die Implantate in die jeweiligen Auslieferungsgebiete und Zahnarztpraxen rückverfolgt werden. Die Dokumentation sollte in digitaler Form erfolgen, da die damit einhergehende Infrastruktur einen effizienten Abgleich mit den Datenbanken der Hersteller zulässt. Es zeichnet sich ab, dass immer mehr Zahnärzte digital arbeiten, wodurch die digitale Dokumentation in Zukunft sehr wahrscheinlich zum Standard werden wird und Patientendaten, wie zum Beispiel Scans, effektiv und leicht zugänglich archiviert werden können [101].

5.3 Diskussion der Auswahl der Population

Die Zusammensetzung der Expertengruppe in dieser Studie war heterogen, da die Teilnehmer unterschiedliche Erfahrungsstufen in der Zahnmedizin aufwiesen. Während einige Teilnehmer erst seit kurzer Zeit im zahnmedizinischen Bereich tätig waren, verfügten andere über eine Berufserfahrung von bis zu neun Jahren. Diese Varianz in der Berufserfahrung beeinflusste maßgeblich die individuellen Analysen und die Sicherheit der Entscheidungen der Teilnehmenden. Weiterhin ist zu diskutieren, dass keiner der Teilnehmer als ausgewiesener Spezialist in der Implantologie galt oder entsprechende Zertifizierungen in diesem Bereich vorweisen konnte. Der Einfluss der Berufserfahrung ist maßgeblich für die Bewertungen der Teilnehmer. Bei der Analyse der vorliegenden Bewertungen ist es von großer Bedeutung, eine Vielzahl an Einflussfaktoren zu berücksichtigen, die das Ergebnis in entscheidendem Maße prägen. Mecler et al. beschrieben als wesentlichen Faktor die Berufserfahrung der Behandler. Sie untersuchten die Faktoren, die die Entscheidungsfindung in Bezug auf die Extraktion oder den Erhalt von parodontal geschädigten Zähnen beeinflussen können. Die Untersuchung ergab, dass sich erfahrenerer Behandler häufiger für eine Extraktion entschieden haben. Desweiteren machte es einen Einfluss, ob die Teilnehmer spezialisiert waren in Parodontologie oder in anderen zahnmedizinischen Bereichen. Parodontologen entschieden sich häufiger für den Zahnerhalt [102]. Dies ist dem Umstand entsprechend, dass es eine intuitive Tendenz gibt, neue Informationen zu suchen und zu interpretieren, die bestehende Überzeugungen bestätigen. Demzufolge steht für spezialisierte Parodontologen der Zahnerhalt im Vordergrund und entsprechend suchen Parodontologen intuitiv nach Informationen, die dem Zahnerhalt dienlich sind. Ein weiterer Einflussfaktor ist, dass erfahrene Zahnärzte komplexe Aufgaben mit weniger fokussierter Aufmerksamkeit und kognitivem Aufwand bewältigen können. Durch ihr höheres Fachwissen können sie ihre Aufmerksamkeit effektiver auf relevante Details des Problems richten. Erfahrene Behandler können schneller Muster erkennen und so bei routinemäßigen Fällen unter Verwendung einer automatischen, intuitiven Mustererkennung und nicht-analytischer Denkprozesse entscheiden. Dahingegen treten bei unerfahrenen Zahnärzten Ermüdungserscheinungen auf, da schnell mentale Ressourcen aufgebraucht werden, um neue Probleme zu lösen, woraus ein gemindertem Urteilsvermögen resultiert. Allerdings neigt man auch dazu, Entscheidungen eher auf Intuition (schnelles, fast automatisches und müheloses,

nicht-analytisches Denken) als auf zeitaufwändiges, analytisches und absichtlich aufmerksames Denken zu stützen. Dies führt bei unerfahrenen Zahnärzten dazu, dass sie tendenziell mehr Fehler machen könnten, da sie mit der Materie und den entsprechenden Mustern noch nicht so vertraut sind wie erfahrene Zahnärzte [103].

Ein weiterer limitierender Faktor dieser Studie war, dass alle Teilnehmer entweder in studentischer oder beruflicher Funktion an derselben zahnmedizinischen Universitätsklinik tätig waren. Um die Repräsentativität der Ergebnisse zu erhöhen, wäre es sinnvoll, eine diversifiziertere Gruppe von Teilnehmern in zukünftigen Studien einzubeziehen. Dies würde eine größere Varianz sowohl hinsichtlich der Dauer der zahnmedizinischen Tätigkeit als auch im Hinblick auf den Grad der Spezialisierung sicherstellen. Eine breitere Expertise könnte die Validität und Übertragbarkeit der Ergebnisse deutlich verbessern. Außerdem sollte die Verteilung der Geschlechter in der Stichprobe gleichmäßig erfolgen, sodass die Anzahl der weiblichen und männlichen Teilnehmer ausgeglichen ist. Die Erhöhung der Repräsentativität durch das Einbeziehen vieler verschiedener Experten zeigt sich sehr gut bei der ersten Umfrage im Rahmen von Konsensus Methoden, bei denen schlussendlich eine wissenschaftlich fundierte Einigkeit gefunden werden soll. Eine bekannte Methode ist die Delphi Technik. Lazarou et al. befragten in ihrer Studie Internisten und Neurologen aus universitären Krankenhäusern in ganz Kanada bezüglich eines Lehrplans im Rahmen der neurologischen Weiterbildung von Assistenzärzten der inneren Medizin [104]. Dies gewährleistete eine Repräsentativität sowie einen validierten Konsens nach weiteren Frage- und Feedbackrunden [105].

Es lässt sich allerdings feststellen, dass bei beiden Geschlechtern eine höhere Erkennungsgenauigkeit für das Original-Implantat im intergruppalen Vergleich festzustellen war. Die zusammengeführten unterschiedlichen Erkennungsraten von Original-Implantaten (88,69%) im Vergleich zu Replik-Implantaten (83,63%) könnten auf markante visuelle Unterscheidungsmerkmale zurückzuführen sein, die den Studienteilnehmern beim Original besser zugänglich waren. Darüber hinaus könnte die Verteilung der Konfidenzniveaus bei den weiblichen Teilnehmerinnen, die eine breite Streuung über die Skala zeigte, ein Hinweis darauf sein, dass diese Gruppe vorsichtiger und reflektierter bei der Entscheidungsfindung vorgeht, was sich in den niedrigeren maximalen Konfidenzstufen im Vergleich zur männlichen

Teilnehmergruppe widerspiegelt. Dies deckt sich mit verschiedenen Studien über geschlechtsspezifische Psychologie. Es konnte festgestellt werden, dass Frauen insgesamt ein gemäßigteres und vorsichtigeres Verhalten aufweisen als Männer, die generell risikofreudiger sind [106]. Beispielsweise zeigen Frauen weniger Risikobereitschaft in Bezug auf wirtschaftliche Entscheidungen [107]. Darüber hinaus neigen sie dazu, einen detaillierten und umfassenden Ansatz in Bezug auf eine Problemlösung zu wählen. Männer fixieren sich dahingegen auf weniger Daten und weniger spezifische Aspekte [106]. Wenn beide Geschlechter allerdings dieselbe Ausbildung erhalten haben und gleiche Expertise in einer Fachrichtung aufweisen, sind keine Unterschiede in der Vorgehensweise hinsichtlich des Lösens neuer Probleme nachzuweisen [108].

Ein Vergleich der Leistungsfähigkeit von KI-Systemen mit der Expertise spezialisierter Implantologen ist unerlässlich, um die tatsächliche Zuverlässigkeit einer KI-gestützten Erkennungssoftware im klinischen Kontext zu bewerten. Dabei ist es von entscheidender Bedeutung, dass die Heterogenität der Expertengruppe, die als Vergleich herangezogen wird, möglichst gering ist. Es sollten ausschließlich Implantologen mit umfassender Erfahrung und Spezialisierung in der Implantologie berücksichtigt werden, die mit den spezifischen Implantatsystemen, die im Vergleich untersucht werden, in ihrer täglichen Praxis vertraut sind [109]. Alle Teilnehmer sollten eine ähnliche Berufserfahrung und vergleichbare Qualifikationen aufweisen, um eine realistische und faire Gegenüberstellung zu ermöglichen.

Da es weltweit über 300 verschiedene Implantathersteller gibt, die jeweils unterschiedliche Systeme anbieten [69], ist es essenziell, den Vergleich unter streng kontrollierten Bedingungen durchzuführen. Eine KI kann über Wochen hinweg intensiv auf zwei oder mehr Implantatsysteme trainiert werden, während menschliche Experten nur eine begrenzte Zeit zur Verfügung haben, um sich mit den entsprechenden Röntgenbildern vertraut zu machen. Diese ungleichen Voraussetzungen könnten dazu führen, dass die KI den Menschen scheinbar leicht übertrifft. Daher bedarf es einer Gegenüberstellung mit erfahrenen und spezialisierten Implantologen, um ein aussagekräftiges Bild davon zu erhalten, wie gut der Mensch im Vergleich zur KI tatsächlich abschneidet.

Die Herausforderung, einen klaren Goldstandard für die Auswahl der Bilder zu treffen, mit der die KI trainiert wird, bleibt bestehen [110, 111]. Demzufolge ist es schwierig, ein objektives Maß zu finden, das die Fähigkeiten beider Parteien fair bewertet. In diesem Zusammenhang wäre es sinnvoll zu untersuchen, ob Zahnärzte in der Lage sind, unter Zuhilfenahme einer KI bessere Vorhersagen zu treffen als ohne deren Unterstützung [40, 112, 113]. Dies würde den realen klinischen Bedingungen näherkommen, da in der Praxis niemals ausschließlich auf die Empfehlungen einer Computersoftware vertraut wird, sondern stets das Urteil des behandelnden Zahnarztes den entscheidenden Einfluss auf die finale Entscheidung hat. Ein solches Studiendesign könnte Aufschluss darüber geben, wie KI-Systeme im klinischen Alltag als unterstützende Entscheidungshilfe eingesetzt werden können, um die Diagnose- und Behandlungsqualität zu verbessern [114].

5.4 Kritische Betrachtung der Materialien und Methoden

Dieser Arbeit liegen diverse Einschränkungen zugrunde.

1) Der vorliegende Gesamtdatensatz, bestehend aus 906 Bildern, ist als relativ begrenzt anzusehen. Es ist weithin anerkannt, dass die Größe eines Datensatzes einen signifikanten Einfluss auf die Effektivität und Zuverlässigkeit des Trainings einer künstlichen Intelligenz hat [7]. Wie bereits in vergleichbaren Publikationen dokumentiert wurde [3, 115], können umfangreichere Datensätze zu einer verbesserten Leistungsfähigkeit der KI-Modelle führen. In Anbetracht dieses Erkenntnis ist es anzustreben, in Zukunft größere und umfangreichere Datensätze zu generieren, um die Zuverlässigkeit und Genauigkeit der KI-basierten Bilderkennung weiter zu verbessern.

2) Es erfolgte keine Qualitätskontrolle der für diese Studie angefertigten Röntgenaufnahmen mit bestimmten Parametern. Die Kontrolle geschah lediglich über die durch die Halterungskonstruktion der Implantate standardisierte vorgegebene Geometrie. Parameter wie Bildschärfe, Intensität, Rauschen und Kontrast wurden nicht einer spezifischen Qualitätskontrolle unterzogen, wobei anzunehmen ist, dass sie sich aufgrund der konstanten Verwendung desselben Röntgengeräts und Röntgenraumes und der gleichen Speicherfolie nur minimal unterscheiden sollten. Es sei jedoch darauf hingewiesen, dass eine detaillierte Untersuchung dieser Parameter einen zusätzlichen Einblick in die Qualität der

erfassten Röntgenbilder bieten könnte und somit eine etwaige Verbesserung der Ergebnisgenauigkeit ermöglichen würde [116].

Allerdings erweist es sich aus wissenschaftlicher Sicht als sehr vorteilhaft, dass ein standardisierter Datensatz vorlag, da dadurch ein mögliches Bias geringgehalten wurde. Ohne eine solche Standardisierung bestünde das Risiko, dass von einem Implantattyp qualitativ minderwertige Aufnahmen vorliegen, was die Vergleichbarkeit der Ergebnisse beeinträchtigen könnte. Durch die Standardisierung des Datensatzes wird jedoch sichergestellt, dass alle Bildaufnahmen in nahezu gleichbleibender Qualität vorliegen und somit eine verlässliche Analyse der Daten ermöglicht wird.

3) Die Bilddaten bestehen nur aus zweidimensionalen Einzelzahnaufnahmen, nicht aber aus dreidimensionalen Röntgenbildern wie sie von DVT-Geräten aufgenommen werden. Solche Röntgenbilder haben allerdings einen sehr hohen Stellenwert vor allem in der Disziplin der oralen Implantologie. Sie geben dem Implantologen exakten Aufschluss darüber, wie die Knochenverhältnisse in Bezug auf die Dichte, Dicke und Ausmaße und wie die Nervverläufe sind. Somit sind dreidimensionale Aufnahmen ein wichtiges Instrument zur Planung eines Implantates als auch zur Begutachtung eines oder mehrerer bereits gesetzter Implantate [46].

Außerdem müsste der Datensatz durch zweidimensionale Panoramaschichtaufnahmen ergänzt werden. Sie stellen ebenfalls eine Standardaufnahme in der Zahnmedizin dar und kommen regelmäßig bei Routineuntersuchungen zum Einsatz.

4) Des Weiteren ist es von großer Bedeutung, dass ebenfalls Bilder miteinbezogen werden, die bereits eingebrachte Abutments und Suprakonstruktionen einschließen, da diese bereits in der Mehrzahl der Fälle bei Erstvorstellung eines Neupatienten mit einem inserierten Implantat in einer zahnärztlichen Praxis vorhanden sind. Die zum Teil sehr guten Ergebnisse der vorliegenden Untersuchung bei einer vergleichsweise kleinen Trainingsdatensatzgröße könnten unter anderem darauf zurückzuführen sein, dass die verwendeten Bilder unter standardisierten und ex situ aufgenommenen Bedingungen erstellt wurden, was ideale Rahmenbedingungen darstellt. Diese sind jedoch in der Realität in der Form fast nie aufzufinden, wie z. B., dass das Implantat sich ohne inserierte Abdeckschraube oder Mesostruktur im Kiefer befindet. Dieser Umstand muss bei der Bewertung der Resultate unbedingt berücksichtigt werden.

Eine zentrale Fragestellung, die in zukünftigen Untersuchungen aufgegriffen werden sollte, betrifft das Verhalten der KI bei in situ Röntgenaufnahmen. Diese Aufnahmen zeichnen sich durch eine größere Variabilität und mögliche Verzerrungen aus, die in

der klinischen Praxis aufgrund der patientenindividuellen Bildgeometrie teilweise unvermeidlich sind. Bei einer klinischen Anwendung müsste die KI dazu fähig sein mit diesen Herausforderungen umzugehen, um zuverlässige Ergebnisse zu erzielen. Es wäre daher von entscheidender Bedeutung, die Leistungsfähigkeit der KI unter diesen weniger idealen, praxisnäheren Bedingungen zu evaluieren [49].

5) Um den Versuchsaufbau realistischer zu gestalten, hätte die Nutzung eines künstlichen Knochenphantoms bei der Bildgebung bzw. bei der Erstellung des Trainings- und Testdatensatzes in Erwägung gezogen werden können. Dies könnte durch ein dreidimensional gedrucktes Modell hergestellt werden. Hatamikia et al. beschrieben dazu ein Vorgehen, bei dem mittels additiver Filamentdrucktechnologie Zylindermodelle mit verschiedenen Dichten erzeugt wurden, um so verschiedene Radiodichten zu reproduzieren. Die jeweilige Füllung der internen Struktur (Infill) des dreidimensionalen Modells führten zu unterschiedlichen Grauwerten, sodass ein Knochenphantom für radiologische Anwendungen erzeugt werden konnte [117].

Eine weitere Methode besteht in der Verwendung von humanem Knochen. Eine Arbeit berichtet von einer Verwendung getrockneter Unterkiefersegmente, in denen noch Zähne inseriert waren. Mit diesen wurden Röntgenaufnahmen durchgeführt, um anhand derer die Verzerrung der Zähne und ihre tatsächliche Länge zu bestimmen [118]. In einer weiteren Studie wurden ebenfalls humane Unterkieferproben verwendet und segmentiert. Diese wurden zusätzlich in Wasser eingetaucht, um eine Weichgewebsabsorption und -streuung zu simulieren [119]. Außerdem berichten Studien davon, dass komplette menschliche Schädel verwendet wurden, um Bewegungen vom Schädel und Unterkiefer während einer DVT-Aufnahme zu untersuchen und zu korrigieren [120] oder um mit CT und DVT-Aufnahmen erzeugte dreidimensionale Schädelmodelle auf ihre Präzision hin zu überprüfen [121]

Alternativ wäre die digitale Umsetzung mithilfe berechneter Phantommodelle denkbar. Auf Basis der Oberflächendefinitionen eines Knochens wird ein Algorithmus entwickelt, der stochastische, bikontinuierliche Mikrostrukturen generiert. Diese Strukturen bilden ein Netzwerk, welches die trabekuläre Knochenarchitektur mit den für Knochen typischen geometrischen und topologischen Eigenschaften nachbildet. Durch diese Methode wird die komplexe innere Struktur des Knochens präzise erfasst und kann somit in der Modellierung rekombiniert wiedergegeben werden [122, 123]. Beide Ansätze würden dazu beitragen, die Bedingungen der klinischen Praxis besser abzubilden und somit die Aussagekraft der Ergebnisse zu erhöhen.

6) Ein weiterer zu diskutierender Punkt dieser Studie besteht darin, dass lediglich ein einzelnes KI-Modell berücksichtigt wurde. Um ein umfassenderes Bild hinsichtlich der Effizienz und Genauigkeit zu erhalten, wäre es sinnvoll, den Datensatz mit mehreren Modellen zu testen. Dadurch könnte man in Bezug auf die Heterogenität der Studienteilnehmer auch eine entsprechende Vielfalt innerhalb der KI-Architekturen schaffen, um einen Vergleich der verschiedenen Modelle anzustellen und herauszufinden, welche Architektur sich für diese spezifische Problemstellung am besten eignet. Dabei sollten unterschiedliche Kriterien, wie Schnelligkeit, Genauigkeit und Kosten berücksichtigt werden. Einige der in Betracht kommenden CNN-Modelle könnten Inception-ResNet und Inception-V3/V4 [124, 125, 126], die Visual Geometry Group (VGG) [4], das Residual Attention Network (RAN) [127] sowie weitere Varianten des Residual Network (ResNet) sein. Diese Modelle unterscheiden sich hauptsächlich in der Filtergröße und der Schichttiefe, was jeweils zu unterschiedlichen Stärken und Schwächen führt.

Auch innerhalb der Modellarchitekturen variiert die Schichttiefe. So existiert bei der VGG-Architektur beispielsweise ein Modell mit 16 (VGG-16) [128] und eines mit 19 Schichten (VGG-19) [129]. Die ResNet-Architektur, die auch in dieser Studie verwendet wurde, kann mit 18, 34, 50, 101 oder 152 Schichten trainiert werden [11]. Ein Vergleich dieser verschiedenen Varianten könnte bereits aufschlussreiche Erkenntnisse bezüglich der Effizienz und der erforderlichen Ressourcen liefern. Eine zentrale Frage wäre, ob der zusätzliche Aufwand für eine möglicherweise nur marginale Verbesserung der Klassifizierungsgenauigkeit tatsächlich gerechtfertigt ist [130].

5.5 Ergebnisse der Arbeit im Kontext der aktuellen Literatur

Wie bereits in den vergangenen Kapiteln deutlich wurde, ist ein für den Vergleich schwieriger Umstand, dass die Studien zum Teil sehr heterogen sind in Bezug auf die verwendeten Daten zum Training der KI hinsichtlich der Qualität und Quantität, die Studienteilnehmer und die KI-Modelle.

Um dennoch einen möglichst realistischen Vergleich mit der vorliegenden Studie zu ermöglichen, sollten vorrangig Arbeiten herangezogen werden, die ebenfalls mit kleinen Trainingsdatensätzen und nur einem einzelnen KI-Modell durchgeführt wurden. Hadj Saïd et al. erzielten beispielsweise eine Genauigkeit von 93,8 % mit einem Trainingsdatensatz, der aus 1206 Bildern und sechs Implantatmodellen

bestand. Kim et al. erreichten eine Genauigkeit von 96,7% bei einem Datensatzumfang von 355 Bildern, die auf drei Implantatmodelle verteilt waren. Leblebicioglu et al. konnten unter Einsatz eines ConvNeXt-Modells (Convolutional Networks with External Transformations) eine maximale Genauigkeit von 95,74 % bei einem Trainingsdatensatz von 1258 Bildern und sechs verschiedenen Implantattypen verzeichnen. In einer weiteren Studie von Kim et al. wurde mit verschiedenen KI-Architekturen eine Genauigkeit von 96-98 % bei einem Trainingsdatensatz von 801 Bildern erreicht, wobei die ResNet-Architekturen mit 98 % am besten abschnitten. Hierbei wurden vier verschiedene Implantate einbezogen. Benakatti et al. erreichten mit einem Datensatz, der ausschließlich Panoramaschichtaufnahmen enthielt, eine Genauigkeit von 67 %. Das in der vorliegenden Studie erzielte Ergebnis von bis zu 100 % Genauigkeit lässt sich vermutlich darauf zurückführen, dass der verwendete Datensatz ex situ hergestellt wurde, dabei äußerst standardisiert war und lediglich zwei Implantatmodelle miteinander verglichen wurden. Dies sorgte für ideale Trainingsdaten und verhinderte Verzerrungen zwischen den Bildgruppen, sodass selbst mit einem kleineren Datensatz eine maximale Genauigkeit erreicht werden konnte.

Alle Studien, die einen Vergleich zwischen menschlichen Experten und KI-gestützten Systemen durchgeführt haben, zeigen übereinstimmend, dass die KI in der Regel eine bessere Leistung bei der Identifizierung der jeweils untersuchten Implantate erbringt [10, 36, 40]. In einer Studie wurde ein besonders deutlicher Unterschied festgestellt, wobei die Genauigkeit der KI im Vergleich zu menschlichen Experten um bis zu 58,5 % höher lag [115]. In manchen Fällen konnte jedoch eine Teilgruppe der befragten Experten bei einem bestimmten Implantat eine höhere Genauigkeit erzielen [10]. In einer anderen Studie konnten zertifizierte Fachzahnärzte für Parodontologie unter Zuhilfenahme der entsprechenden KI bessere Ergebnisse erzielen als die KI allein [40]. Dies verdeutlicht, dass die Vorhersagen einer KI in der klinischen Praxis stets durch den behandelnden Zahnarzt oder einen ausgewiesenen Experten validiert werden sollten. Die Ergebnisse dieser Studien stehen im Einklang mit den vorliegenden Studienergebnissen, die ebenfalls belegen, dass die KI im Durchschnitt bessere Resultate liefert als Zahnärzte mit einer allgemeinen zahnärztlichen Tätigkeit.

5.5 Ausblick

Insgesamt lässt sich feststellen, dass zwei sowohl makroskopisch als auch röntgenologisch nahezu identische Implantate mithilfe eines KI-Modells voneinander differenziert werden können. KI-Softwares zur Identifizierung von sehr ähnlichen Implantaten im Sinne eines Originalproduktes und eines entsprechenden Replikates könnten sich als sehr aussagekräftige potenzielle Hilfen in der klinischen Praxis erweisen. Sie ermöglichen, das vorliegende Implantatsystem zu erkennen und die entsprechenden herstellereigenen Komponenten zu verwenden. Auf diese Weise wird die Patientengesundheit sowie auch die medizinische Qualität der Arbeit des behandelnden Arztes maßgeblich verbessert. Die gegenwärtige Dynamik innerhalb der oralen Implantologie deutet auf eine zunehmende Verbreitung und Nutzung von Replikaten hin. Dieser Trend spiegelt sich in der steigenden Vielfalt von Replikat-Implantaten wider, die auf dem Markt verfügbar sind und von Patienten und Zahnärzten gleichermaßen in Betracht gezogen werden. Diese Entwicklung ist durch mehrere Faktoren bedingt, darunter potenzielle Kosteneinsparungen, verbesserte Verfügbarkeit und vielfältigere Optionen für Patienten. In Zukunft wird eine für diesen Fall spezialisierte Erkennungssoftware von immer größerer Bedeutung sein. Zudem lässt sie sich ohne großen finanziellen und organisatorischen Aufwand in den Praxisalltag integrieren. Dies könnte im Sinne einer in die praxiseigenen Röntgensoftware zusätzlich integrierten Erweiterung realisiert werden, vorausgesetzt es handelt sich dabei um ein digitales Röntgensystem.

Ein weiterer Vorteil einer solchen Lösung liegt in der kontinuierlichen Verbesserung der künstlichen Intelligenz, indem sie fortlaufend mit neuen Daten versorgt wird, die von den geröntgten Patienten stammen. Die rasante Entwicklung in der KI-Technologie erlaubt in Zukunft die Programmierung immer präziserer Modelle, die eine noch genauere und schnellere Identifizierung ermöglichen [33].

Zukünftig muss die für diese Studie programmierte KI-Software mit mehr Daten trainiert werden. Wie bereits erwähnt, sollten diese um dreidimensionale Röntgenaufnahmen und Panoramaschichtaufnahmen erweitert werden. Dies erfordert zugleich die Integration von in situ Röntgenaufnahmen in den Trainings- und Testdatensatz. Auf diese Weise lassen sich die Vorhersagewahrscheinlichkeiten bestätigen und immer weiter verfeinern. Ebenfalls muss eine Vergleichsgruppe bestehend aus spezialisierten Implantologen herangezogen werden. Des Weiteren ist es von großer Bedeutung, dass ebenfalls Bilder miteinbezogen werden, die

bereits eingebrachte Abutments und Suprakonstruktionen einschließen, da diese bereits häufig bei Erstvorstellung eines Neupatienten in einer zahnärztlichen Praxis vorhanden sind.

6 Zusammenfassung

Diese Pilotstudie soll dazu dienen, einen ersten Ausblick zu geben, wie sicher KI zwischen einem Original- und einem Replikat-Implantat auf Röntgenaufnahmen unterscheiden kann. Es wurde das Ziel verfolgt, die Hypothese zu überprüfen, ob eine KI-basierte Bilderkennung von zwei nahezu identischen dentalen Implantaten menschlichen Experten überlegen ist. Um dies zu untersuchen, wurden zwei entsprechende Implantate – eines vom Originalhersteller Straumann und ein Replikat von Meisinger – anhand von extraoralen Röntgenaufnahmen durch verschieden umfangreich trainierte CNNs klassifiziert. Dabei wurde auch die Leistung menschlicher Zahnmediziner einbezogen, deren Ergebnisse mit denen der KI verglichen wurden.

Der für diese Studie erhobene Datensatz bestand aus 906 standardisierten ex situ Einzelzahnaufnahmen der beiden Implantatmodelle. Der Testdatensatz bestand aus weiteren 48 ebensolcher Röntgenbilder. Das deterministische KI-Modell, das mit 488 Röntgenbildern der insgesamt 906 vorhandenen trainiert wurde, wies eine Gesamtgenauigkeit von 57,9 % auf. Die 488 Röntgenbilder inkludierten nur solche mit inserierter Verschlusschraube. Die beiden weiteren KI-Modelle wiesen eine Genauigkeit von >90 % und 100 % auf. Zwei Klassifikatoren des dritten Modells (2+3) erkannten somit fehlerlos alle Bilder des Testdatensatzes. Im Vergleich zur KI konnte die menschliche Expertengruppe eine Gesamtgenauigkeit von 86,16 % erzielen. Die Ergebnisse der KI zeigten, dass mit einem kleinen Trainingsdatensatz von 488 Bildern eine deutlich geringere Genauigkeit erzielt wurde, während größere Datensätze (674 bzw. 906 Bilder) die Vorhersagewahrscheinlichkeit signifikant erhöhten. Dies verdeutlicht die Bedeutung der Datensatzgröße und -qualität für die Leistungsfähigkeit von KI-Algorithmen [25, 26, 40].

Die Studie belegt, dass KI-basierte Systeme, insbesondere wenn sie mit größeren und standardisierten Datensätzen trainiert werden, eine hohe Genauigkeit bei der Identifizierung von zwei sehr ähnlichen Implantaten aufweisen können [100]. Dennoch bleibt festzuhalten, dass menschliche Expertise weiterhin eine entscheidende Rolle spielt, insbesondere in klinischen Kontexten, wo das Urteil von Fachärzten unerlässlich ist [37, 99]. Die fortlaufende Optimierung der KI und deren Integration in den Praxisalltag könnte zukünftig eine wertvolle Unterstützung für zahnmedizinische Behandler darstellen, da somit viel Zeit und Aufwand eingespart sowie Komplikationsraten im Sinne des Patientenwohls vermindert werden können.

7 Literaturverzeichnis

1. Alqutaibi AY, Algabri RS, Elawady D, Ibrahim WI. Advancements in artificial intelligence algorithms for dental implant identification: A systematic review with meta-analysis. *J Prosthet Dent.* 2023.
2. Schilling A, Maier A, Gerum R, Metzner C, Krauss P. Quantifying the separability of data classes in neural networks. *Neural Netw.* 2021;139:278-93.
3. Chaurasia A, Namachivayam A, Koca-Ünsal RB, Lee JH. Deep-learning performance in identifying and classifying dental implant systems from dental imaging: a systematic review and meta-analysis. *J Periodontal Implant Sci.* 2024;54(1):3-12.
4. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:14091556.* 2014.
5. Wu J. Introduction to convolutional neural networks. National Key Lab for Novel Software Technology Nanjing University China. 2017;5(23):495.
6. Narkhede MV, Bartakke PP, Sutaone MS. A review on weight initialization strategies for neural networks. *Artificial Intelligence Review.* 2022;55(1):291-322.
7. Ibraheem WI. Accuracy of Artificial Intelligence Models in Dental Implant Fixture Identification and Classification from Radiographs: A Systematic Review. *Diagnostics (Basel).* 2024;14(8).
8. Kim JE, Nam NE, Shim JS, Jung YH, Cho BH, Hwang JJ. Transfer Learning via Deep Neural Networks for Implant Fixture System Classification Using Periapical Radiographs. *J Clin Med.* 2020;9(4).
9. Hadj Saïd M, Le Roux MK, Catherine JH, Lan R. Development of an Artificial Intelligence Model to Identify a Dental Implant from a Radiograph. *Int J Oral Maxillofac Implants.* 2020;36(6):1077-82.
10. Lee JH, Kim YT, Lee JB, Jeong SN. A Performance Comparison between Automated Deep Learning and Dental Professionals in Classification of Dental Implant Systems from Dental Imaging: A Multi-Center Study. *Diagnostics (Basel).* 2020;10(11).
11. Sukegawa S, Yoshii K, Hara T, Matsuyama T, Yamashita K, Nakano K, et al. Multi-Task Deep Learning Model for Classification of Dental Implant Brand and Treatment Stage Using Dental Panoramic Radiograph Images. *Biomolecules.* 2021;11(6).
12. Gallant SI. Perceptron-based learning algorithms. *IEEE Trans Neural Netw.* 1990;1(2):179-91.
13. Raudys Š. Evolution and generalization of a single neurone: I. Single-layer perceptron as seven statistical classifiers. *Neural Networks.* 1998;11(2):283-96.
14. Selle S. Künstliche Neuronale Netzwerke und Deep Learning 2018.
15. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. *Translational vision science & technology.* 2020;9(2):14-.
16. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. *J Med Imaging Radiat Sci.* 2019;50(4):477-87.
17. Bini SA. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *J Arthroplasty.* 2018;33(8):2358-61.

18. Lee JH, Kim DH, Jeong SN, Choi SH. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J Periodontal Implant Sci.* 2018;48(2):114-23.
19. Schwendicke F, Chaurasia A, Arsiwala L, Lee JH, Elhennawy K, Jost-Brinkmann PG, et al. Deep learning for cephalometric landmark detection: systematic review and meta-analysis. *Clin Oral Investig.* 2021;25(7):4299-309.
20. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent.* 2018;77:106-11.
21. Zhou D-X. Theory of deep convolutional neural networks: Downsampling. *Neural Networks.* 2020;124:319-27.
22. Murugan P. Feed forward and backward run in deep convolution neural network. *arXiv preprint arXiv:171103278.* 2017.
23. Cong S, Zhou Y. A review of convolutional neural network architectures and their optimizations. *Artificial Intelligence Review.* 2023;56(3):1905-69.
24. Murugan P. Feed Forward and Backward Run in Deep Convolution Neural Network. *CoRR.* 2017;abs/1711.03278.
25. Gygi JP, Kleinstein SH, Guan L. Predictive overfitting in immunological applications: Pitfalls and solutions. *Hum Vaccin Immunother.* 2023;19(2):2251830.
26. Bishop CM, Nasrabadi, N.M. *Pattern recognition and machine learning.* : Springer New York, NY; 2006. ISBN: 978-0387-31073-2
27. Castro W, Oblitas J, Santa-Cruz R, Avila-George H. Multilayer perceptron architecture optimization using parallel computing techniques. *PLoS One.* 2017;12(12):e0189369.
28. Mohebbali B, Tahmassebi A, Meyer-Baese A, Gandomi AH. Chapter 14 - Probabilistic neural networks: a brief overview of theory, implementation, and application. In: Samui P, Tien Bui D, Chakraborty S, Deo RC, editors. *Handbook of Probabilistic Models: Butterworth-Heinemann;* 2020. p. 347-67.
29. Biau G, Scornet E. A random forest guided tour. *Test.* 2016;25:197-227.
30. Hu J, Szymczak S. A review on longitudinal data analysis with random forest. *Brief Bioinform.* 2023;24(2).
31. Rigatti SJ. Random forest. *Journal of Insurance Medicine.* 2017;47(1):31-9.
32. Schwendicke F, Samek W, Krois J. Artificial Intelligence in Dentistry: Chances and Challenges. *J Dent Res.* 2020;99(7):769-74.
33. Park WS, Huh JK, Lee JH. Automated deep learning for classification of dental implant radiographs using a large multi-center dataset. *Sci Rep.* 2023;13(1):4862.
34. da Mata Santos RP, Vieira Oliveira Prado HE, Soares Aranha Neto I, Alves de Oliveira GA, Vespasiano Silva AI, Zenóbio EG, et al. Automated Identification of Dental Implants Using Artificial Intelligence. *Int J Oral Maxillofac Implants.* 2021;36(5):918-23.
35. Benakatti VB, Nayakar RP, Anandhalli M. Machine learning for identification of dental implant systems based on shape - A descriptive study. *J Indian Prosthodont Soc.* 2021;21(4):405-11.
36. Lee JH, Jeong SN. Efficacy of deep convolutional neural network algorithm for the identification and classification of dental implant systems, using panoramic and periapical radiographs: A pilot study. *Medicine (Baltimore).* 2020;99(26):e20787.
37. Kim HS, Ha EG, Kim YH, Jeon KJ, Lee C, Han SS. Transfer learning in a deep convolutional neural network for implant fixture classification: A pilot study. *Imaging Sci Dent.* 2022;52(2):219-24.

38. Kong HJ, Eom SH, Yoo JY, Lee JH. Identification of 130 Dental Implant Types Using Ensemble Deep Learning. *Int J Oral Maxillofac Implants.* 2023;38(1):150-6.
39. Takahashi T, Nozaki K, Gonda T, Mameno T, Wada M, Ikebe K. Identification of dental implants using deep learning-pilot study. *Int J Implant Dent.* 2020;6(1):53.
40. Lee JH, Kim YT, Lee JB, Jeong SN. Deep learning improves implant classification by dental professionals: a multi-center evaluation of accuracy and efficiency. *J Periodontal Implant Sci.* 2022;52(3):220-9.
41. Sukegawa S, Yoshii K, Hara T, Tanaka F, Yamashita K, Kagaya T, et al. Is attention branch network effective in classifying dental implants from panoramic radiograph images by deep learning? *PLoS One.* 2022;17(7):e0269016.
42. Lee DW, Kim SY, Jeong SN, Lee JH. Artificial Intelligence in Fractured Dental Implant Detection and Classification: Evaluation Using Dataset from Two Dental Hospitals. *Diagnostics (Basel).* 2021;11(2).
43. Hsiao CY, Bai H, Ling H, Yang J. Artificial Intelligence in Identifying Dental Implant Systems on Radiographs. *Int J Periodontics Restorative Dent.* 2023;43(3):363-8.
44. Walsh T. Fuzzy gold standards: Approaches to handling an imperfect reference standard. *J Dent.* 2018;74 Suppl 1:S47-s9.
45. Sukegawa S, Yoshii K, Hara T, Yamashita K, Nakano K, Yamamoto N, et al. Deep Neural Networks for Dental Implant System Classification. *Biomolecules.* 2020;10(7).
46. Shahbazian M, Vandewoude C, Wyatt J, Jacobs R. Comparative assessment of panoramic radiography and CBCT imaging for radiodiagnostics in the posterior maxilla. *Clin Oral Investig.* 2014;18(1):293-300.
47. Hartoonian S, Hosseini M, Yousefi I, Mahdian M, Ghazizadeh Ahsaie M. Applications of artificial intelligence in dentomaxillofacial imaging-a systematic review. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2024.
48. Kong HJ. Classification of dental implant systems using cloud-based deep learning algorithm: an experimental study. *J Yeungnam Med Sci.* 2023;40(Suppl):S29-s36.
49. Lee JH, Kim YT, Lee JB. Identification of dental implant systems from low-quality and distorted dental radiographs using AI trained on a large multi-center dataset. *Sci Rep.* 2024;14(1):12606.
50. Branemark P-I. Osseointegration and its experimental background. *The Journal of Prosthetic Dentistry.* 1983;50(3):399-410.
51. Brånemark PI, Hansson BO, Adell R, Breine U, Lindström J, Hallén O, et al. Osseointegrated implants in the treatment of the edentulous jaw. Experience from a 10-year period. *Scand J Plast Reconstr Surg Suppl.* 1977;16:1-132.
52. Schnitman PA, Shulman LB. Recommendations of the consensus development conference on dental implants. *J Am Dent Assoc.* 1979;98(3):373-7.
53. Brånemark P-I, Zarb GA, Albrektsson T. *Tissue-integrated prostheses : osseointegration in clinical dentistry.* Chicago: Quintessence; 1985.
54. Block MS. *Dental Implants: The Last 100 Years.* *Journal of Oral and Maxillofacial Surgery.* 2018;76(1):11-26.
55. Jackowski J, Peters H, Hölzle F. *Zahnärztliche Chirurgie.* 1 ed: Springer Berlin, Heidelberg; 2017. ISBN: 978-3-642-54754-6

56. Sergio Alexandre G. Characteristics of Implant Systems That Can Accelerate and Improve the Osseointegration Process. In: Dragana G, Marko V, editors. *Current Concepts in Dental Implantology*. Rijeka: IntechOpen; 2021. p. Ch. 1.
57. Gutwald R, Gellrich N-C, Schmelzeisen R. *Zahnärztliche Chirurgie und Implantologie 3. A.: Für Studium, Examen und Beruf*. 3 ed. Köln: Deutscher Zahnärzte Verlag; 2018. ISBN: 978-3-7691-3686-9
58. Cacaci C, Neugebauer J, Schlegel K-A, Seidel F. *Orale Implantologie*. 1 ed. Stuttgart: Thieme 2006. ISBN: 978-3-13-143231-5
59. Neugebauer J, Schoenbaum TR, Pi-Anfruns J, Yang M, Lander B, Blatz MB, et al. Ceramic Dental Implants: A Systematic Review and Meta-analysis. *Int J Oral Maxillofac Implants*. 2023;38(suppl):30-6.
60. Buser D, Sennerby L, De Bruyn H. Modern implant dentistry based on osseointegration: 50 years of progress, current trends and open questions. *Periodontology 2000*. 2017;73(1):7-21.
61. Addison O, Davenport AJ, Newport RJ, Kalra S, Monir M, Mosselmans JF, et al. Do 'passive' medical titanium surfaces deteriorate in service in the absence of wear? *J R Soc Interface*. 2012;9(76):3161-4.
62. Torroella-Saura G, Mareque-Bueno J, Cabratosa-Termes J, Hernández-Alfaro F, Ferrés-Padró E, Calvo-Guirado JL. Effect of implant design in immediate loading. A randomized, controlled, split-mouth, prospective clinical trial. *Clin Oral Implants Res*. 2015;26(3):240-4.
63. Sakoh J, Wahlmann U, Stender E, Nat R, Al-Nawas B, Wagner W. Primary stability of a conical implant and a hybrid, cylindrical screw-type implant in vitro. *Int J Oral Maxillofac Implants*. 2006;21(4):560-6.
64. Wilson TG, Jr., Miller RJ, Trushkowsky R, Dard M. Tapered Implants in Dentistry: Revitalizing Concepts with Technology: A Review. *Adv Dent Res*. 2016;28(1):4-9.
65. Gehrke SA, Aramburú JJ, Pérez-Díaz L, do Prado TD, Dedavid BA, Mazon P, et al. Can changes in implant macrogeometry accelerate the osseointegration process?: An in vivo experimental biomechanical and histological evaluations. *PLoS One*. 2020;15(5):e0233304.
66. Gehrke SA, Aramburú Júnior J, Pérez-Díaz L, Treichel TLE, Dedavid BA, De Aza PN, et al. New Implant Macrogeometry to Improve and Accelerate the Osseointegration: An In Vivo Experimental Study. *Applied Sciences*. 2019;9(15):3181.
67. Gehrke SA, Pérez-Díaz L, Mazón P, De Aza PN. Biomechanical Effects of a New Macrogeometry Design of Dental Implants: An In Vitro Experimental Analysis. *J Funct Biomater*. 2019;10(4).
68. Gehrke SA, Tumedei M, Aramburú Júnior J, Treichel TLE, Kolerman R, Lepore S, et al. Histological and Histomorphometrical Evaluation of a New Implant Macrogeometry. A Sheep Study. *Int J Environ Res Public Health*. 2020;17(10).
69. Misch CM. Editorial: The global dental implant market: Everything has a price. *Int J Oral Implantol (Berl)*. 2020;13(4):311-2.
70. Burnaby B. *Dental Implants Market Analysis, Size, Trends, Global, 2019-2025.*: iData Research; 2019.
71. Jokstad A, Braegger U, Brunski JB, Carr AB, Naert I, Wennerberg A. Quality of dental implants. *Int Dent J*. 2003;53(6 Suppl 2):409-43.
72. Karl M, Irastorza-Landa A. In Vitro Characterization of Original and Nonoriginal Implant Abutments. *Int J Oral Maxillofac Implants*. 2018;33(6):1229-39.
73. Rizvi N, Alyahya Y, Rizvi A, Narvekar U, Petridis H. Accuracy of Original vs. Non-Original Abutments Using Various Connection Geometries for Single Unit Restorations: A Systematic Review. *J Prosthodont*. 2022;31(7):e21-e52.

74. Berberi A, Maroun D, Kanj W, Amine EZ, Philippe A. Micromovement Evaluation of Original and Compatible Abutments at the Implant-abutment Interface. *J Contemp Dent Pract.* 2016;17(11):907-13.
75. Alonso-Pérez R, Bartolomé JF, Ferreiroa A, Salido MP, Pradíes G. Original vs. non-original abutments for screw-retained single implant crowns: An in vitro evaluation of internal fit, mechanical behaviour and screw loosening. *Clin Oral Implants Res.* 2018;29(12):1230-8.
76. Kim ES, Shin SY. Influence of the implant abutment types and the dynamic loading on initial screw loosening. *J Adv Prosthodont.* 2013;5(1):21-8.
77. Park JM, Baek CH, Heo SJ, Kim SK, Koak JY, Kim SK, et al. An In Vitro Evaluation of the Loosening of Different Interchangeable Abutments in Internal-Connection-Type Implants. *Int J Oral Maxillofac Implants.* 2017;32(2):350-5.
78. Kim SK, Koak JY, Heo SJ, Taylor TD, Ryoo S, Lee SY. Screw loosening with interchangeable abutments in internally connected implants after cyclic loading. *Int J Oral Maxillofac Implants.* 2012;27(1):42-7.
79. Cashman PM, Schneider RL, Schneider GB, Stanford CM, Clancy JM, Qian F. In vitro analysis of post-fatigue reverse-torque values at the dental abutment/implant interface for a unitarian abutment design. *J Prosthodont.* 2011;20(7):503-9.
80. Pol CWP, Cune MS, Raghoobar GM, Naves LZ, Meijer HJA. Mechanical strength of stock and custom abutments as original and aftermarket components after thermomechanical aging. *Clin Exp Dent Res.* 2024;10(4):e892.
81. Alonso-Pérez R, Bartolomé JF, Pradíes G. Original vs compatible stock abutment- implant connection: An in vitro analysis of the internal accuracy and mechanical fatigue behaviour. *J Prosthodont Res.* 2022;66(3):476-83.
82. Berberi A, Tehini G, Hjeij B, Aoun G. Evaluation of Marginal and Internal Fit at Implant-Abutment Interface of Original and Compatible Nonoriginal Abutments. *J Long Term Eff Med Implants.* 2022;32(3):21-8.
83. Jarman JM, Hamalian T, Randi AP. Comparing the Fracture Resistance of Alternatively Engineered Zirconia Abutments with Original Equipment Manufactured Abutments with Different Implant Connection Designs. *Int J Oral Maxillofac Implants.* 2017;32(5):992–1000.
84. Tallarico M, Fiorellini J, Nakajima Y, Omori Y, Takahisa I, Canullo L. Mechanical Outcomes, Microleakage, and Marginal Accuracy at the Implant-Abutment Interface of Original versus Nonoriginal Implant Abutments: A Systematic Review of In Vitro Studies. *Biomed Res Int.* 2018;2018:2958982.
85. Morton D, Gallucci G, Lin WS, Pjetursson B, Polido W, Roehling S, et al. Group 2 ITI Consensus Report: Prosthodontics and implant dentistry. *Clin Oral Implants Res.* 2018;29 Suppl 16:215-23.
86. Pjetursson BE, Sailer I, Merino-Higuera E, Spies BC, Burkhardt F, Karasan D. Systematic review evaluating the influence of the prosthetic material and prosthetic design on the clinical outcomes of implant-supported multi-unit fixed dental prosthesis in the posterior area. *Clinical Oral Implants Research.* 2023;34(S26):86-103.
87. Straumann AG. Straumann® Garantie 2018 [zitiert 27.09.2024] Verfügbar unter: https://www.straumann.com/content/dam/media-center/straumann/de-ch/documents/brochure/services/ch_151.360_de_straumann_garantie.pdf.
88. Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. *Seminars in Cancer Biology.* 2021;72:214-25.

89. Sechopoulos I, Mann RM. Stand-alone artificial intelligence - The future of breast cancer screening? *The Breast*. 2020;49:254-60.
90. Straumann AG. 4,1 mm Standard Plus Implantate (RN) [zitiert 09.08.2024] Verfügbar unter: https://shop.straumann.com/de/de_de/Implantate/Tissue-Level-Implantate/Tissue-Level-Implantate-%28TL%29/%C3%98-4-1mm-Standard-Plus-Implants-RN/p/base_stmn_implant_sp41/.
91. iData Research. Dental Implant and Final Abutment Market Size, Share and Trends Analysis, Global, 2024-2030 2024.
92. Lustig LL, Gozalo-Diaz D. Creating QR-coded identification cards using CardExchange for straightforward identification of implant systems. *J Prosthet Dent*. 2020;124(5):623-4.
93. Naemi R, Jangi M, Barikani HR, Shahmoradi L. Design and Evaluation of Web-Based Dental Implant Registry (DIR) for Better Clinical Outcomes. *Int J Biomater*. 2022;2022:7162645.
94. Sohn L. IMPLANT ID CARDS. *The Journal of the American Dental Association*. 1998;129(5):532-3.
95. Zipprich H, Weigl P, Ratka C, Lange B, Lauer HC. The micromechanical behavior of implant-abutment connections under a dynamic load protocol. *Clin Implant Dent Relat Res*. 2018;20(5):814-23.
96. Wittneben JG, Millen C, Brägger U. Clinical performance of screw- versus cement-retained fixed implant-supported reconstructions--a systematic review. *Int J Oral Maxillofac Implants*. 2014;29 Suppl:84-98.
97. Zipprich H WP, Lange B, Lauer H-C. Erfassung, Ursachen und Folgen von Mikrobewegungen am Implantat-Abutment-Interface. *Implant Dent*. 2007;15:31-46.
98. Broggin N, McManus LM, Hermann JS, Medina RU, Oates TW, Schenk RK, et al. Persistent acute inflammation at the implant-abutment interface. *J Dent Res*. 2003;82(3):232-7.
99. Scarano A, Assenza B, Piattelli M, Iezzi G, Leghissa GC, Quaranta A, et al. A 16-year study of the microgap between 272 human titanium implants and their abutments. *J Oral Implantol*. 2005;31(6):269-75.
100. Mattheos N, Janda MS. Exotic encounters with dental implants: managing complications with unidentified systems. *Aust Dent J*. 2012;57(2):236-42.
101. Schlenz M, Schulz-Weidner N, Olbrich M, Buchmann D, Wöstmann B. Insights on the digitalisation of dental practices: A cross-sectional pilot study in Hesse. *Int J Comput Dent*. 9999;0(0):0-.
102. Mecler N, Rodrigues RM, Silva AMP, Fernandes AC, Almeida BM, Gonçalves LS, et al. Role of professional expertise in the decision-making process for periodontally compromised teeth. *Gen Dent*. 2022;70(2):62-8.
103. Feller L, Lemmer J, Nemutandani MS, Ballyram R, Khammissa RAG. Judgment and decision-making in clinical dentistry. *J Int Med Res*. 2020;48(11):300060520972877.
104. Lazarou J, Hopyan J, Panisko D, Tai P. Neurology for internal medicine residents: working towards a national Canadian curriculum consensus. *Med Teach*. 2011;33(2):e65-8.
105. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm*. 2016;38(3):655-62.
106. Meyers-Levy J, Loken B. Revisiting gender differences: What we know and what lies ahead. *Journal of Consumer Psychology*. 2015;25(1):129-49.
107. Dwyer PD, Gilkeson JH, List JA. Gender differences in revealed risk taking: evidence from mutual fund investors. *Economics Letters*. 2002;76(2):151-8.

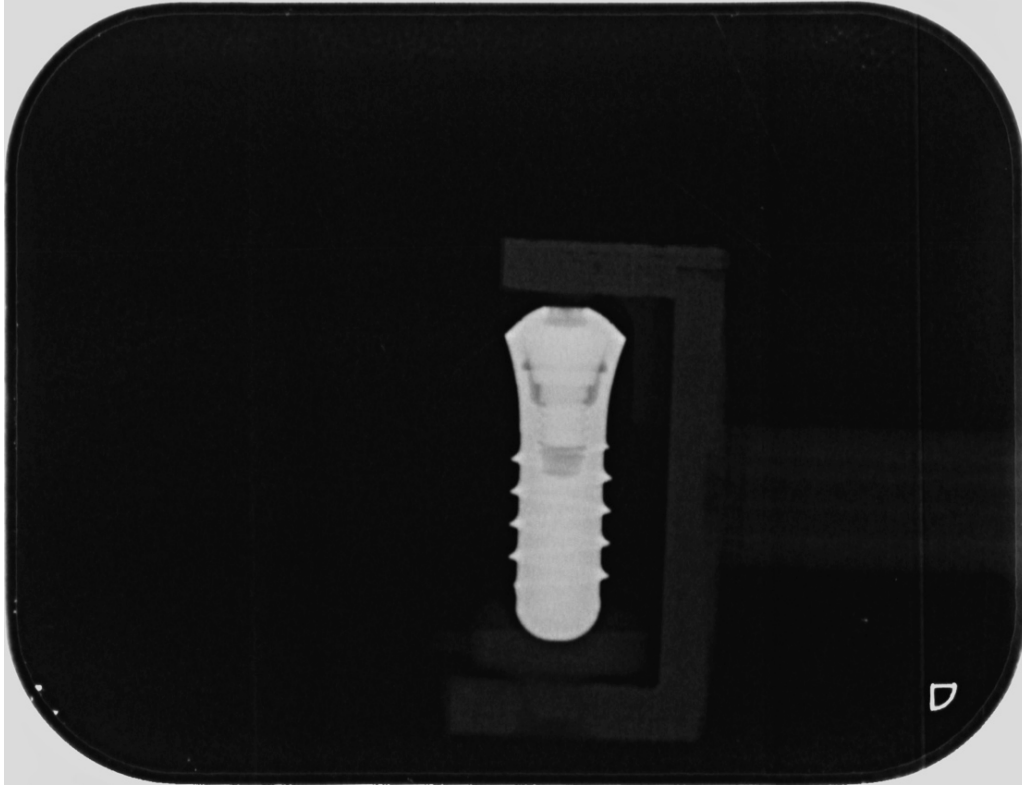
108. Johnson JEV, Powell PL. Decision Making, Risk and Gender: Are Managers Different? *British Journal of Management*. 1994;5(2):123-38.
109. Lops D, Palazzolo A, Calza S, Proietto L, Sordillo A, Mensi M, et al. Guided versus freehand single implant placement: A 3-year parallel randomized clinical trial. *J Dent*. 2024;149:105317.
110. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-8.
111. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *Jama*. 2016;316(22):2402-10.
112. Li S, Liu J, Zhou Z, Zhou Z, Wu X, Li Y, et al. Artificial intelligence for caries and periapical periodontitis detection. *J Dent*. 2022;122:104107.
113. Han SS, Park I, Eun Chang S, Lim W, Kim MS, Park GH, et al. Augmented Intelligence Dermatology: Deep Neural Networks Empower Medical Professionals in Diagnosing Skin Cancer and Predicting Treatment Options for 134 Skin Disorders. *Journal of Investigative Dermatology*. 2020;140(9):1753-61.
114. Schaffter T, Buist DSM, Lee CI, Nikulin Y, Ribli D, Guan Y, et al. Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms. *JAMA Netw Open*. 2020;3(3):e200265.
115. Park W, Schwendicke F, Krois J, Huh JK, Lee JH. Identification of Dental Implant Systems Using a Large-Scale Multicenter Data Set. *J Dent Res*. 2023;102(7):727-33.
116. Wright AI, Dunn CM, Hale M, Hutchins GGA, Treanor DE. The Effect of Quality Control on Accuracy of Digital Pathology Image Analysis. *IEEE J Biomed Health Inform*. 2021;25(2):307-14.
117. Hatamikia S, Kronreif G, Unger A, Oberoi G, Jaksa L, Unger E, et al. 3D printed patient-specific thorax phantom with realistic heterogeneous bone radiopacity using filament printer technology. *Z Med Phys*. 2022;32(4):438-52.
118. Roeder F, Brüllmann D, d'Hoedt B, Schulze R. Ex vivo radiographic tooth length measurements with the reference sphere method (RSM). *Clin Oral Investig*. 2010;14(6):645-51.
119. Wolf TG, Castañeda-López F, Gleißner L, Schulze R, Kuchen R, Briseño-Marroquín B. Detectability of simulated apical lesions on mandibular premolars and molars between radiographic intraoral and cone-beam computed tomography images: an ex vivo study. *Sci Rep*. 2022;12(1):14032.
120. Birklein L, Niebler S, Schömer E, Brylka R, Schwanecke U, Schulze R. Motion correction for separate mandibular and cranial movements in cone beam CT reconstructions. *Med Phys*. 2023;50(6):3511-25.
121. Ghamri M, Dritsas K, Probst J, Jäggi M, Psomiadis S, Schulze R, et al. Accuracy of facial skeletal surfaces segmented from CT and CBCT radiographs. *Sci Rep*. 2023;13(1):21002.
122. Abadi E, Segars WP, Sturgeon GM, Harrawood B, Kapadia A, Samei E. Modeling "Textured" Bones in Virtual Human Phantoms. *IEEE Trans Radiat Plasma Med Sci*. 2019;3(1):47-53.
123. Sauer TJ, McCabe C, Abadi E, Samei E, Segars WP. Surface-based anthropomorphic bone structures for use in high-resolution simulated medical imaging. *Phys Med Biol*. 2023;69(1).

124. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al., editors. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015.
125. Szegedy C, Ioffe S, Vanhoucke V, Alemi A, editors. Inception-v4, inception-resnet and the impact of residual connections on learning. Proceedings of the AAAI conference on artificial intelligence; 2017.
126. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z, editors. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
127. Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, et al., editors. Residual attention network for image classification. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.
128. Ghosh S, Chaki A, Santosh KC. Improved U-Net architecture with VGG-16 for brain tumor segmentation. *Physical and Engineering Sciences in Medicine*. 2021;44(3):703-12.
129. Wen L, Li X, Li X, Gao L, editors. A New Transfer Learning Based on VGG-19 Network for Fault Diagnosis. 2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD); 2019 6-8 May 2019.
130. Tamburrini G. The AI Carbon Footprint and Responsibilities of AI Scientists. *Philosophies*. 2022;7(1):4.

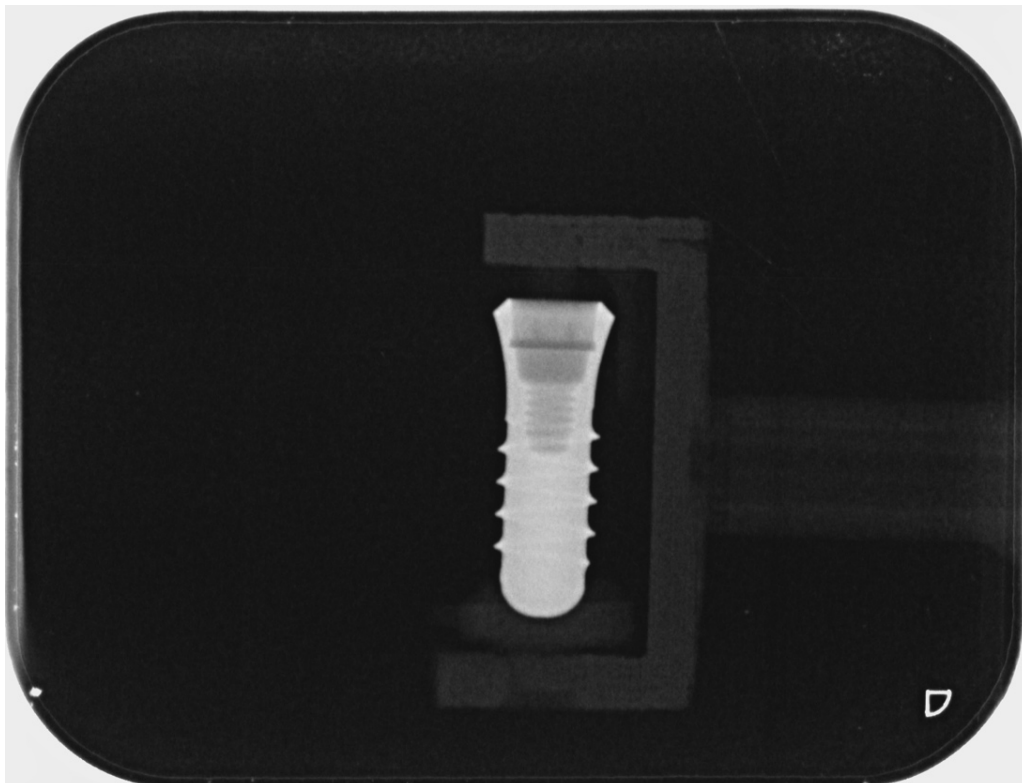
8 Anhang

8.1 Beispielbilder zur Vorbereitung der Befragung für die menschlichen Experten

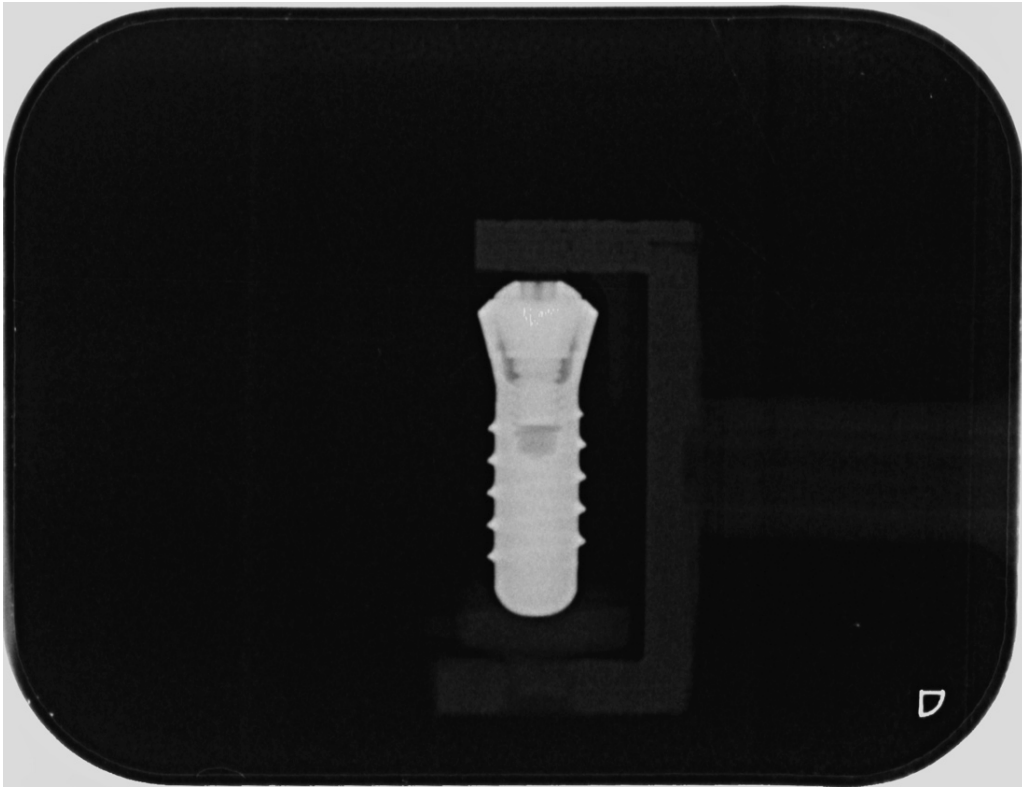
8.1.1 Original-Implantat mit inserierter Schraube



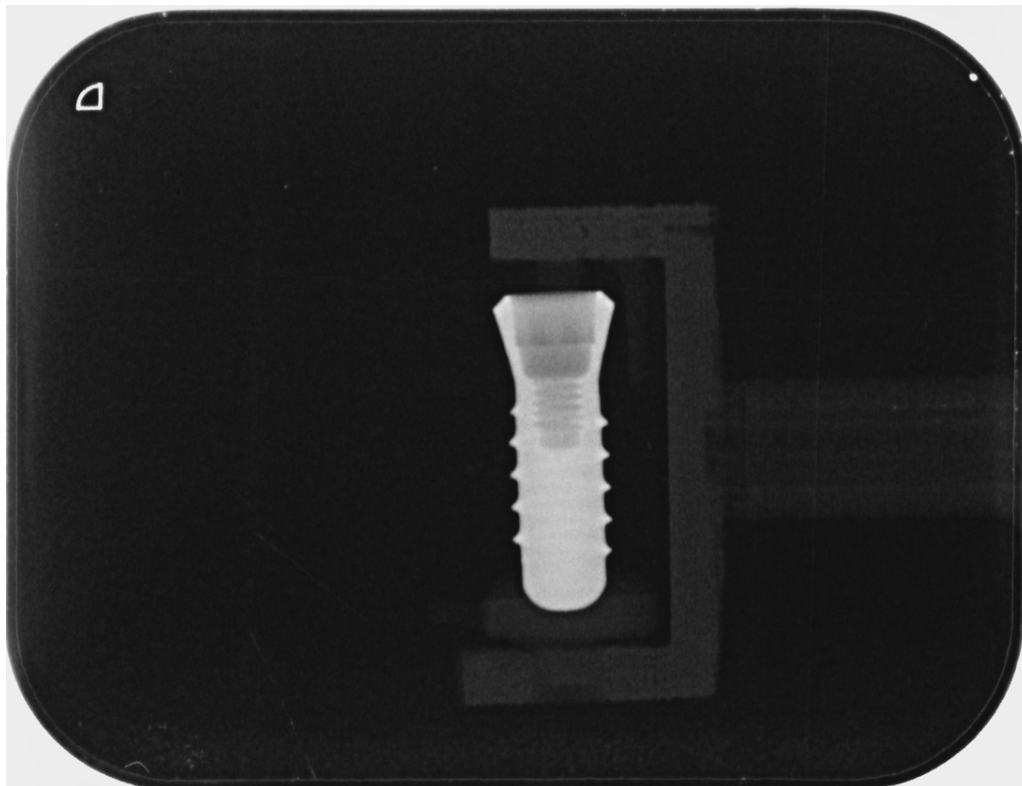
8.1.2 Original-Implantat ohne inserierte Schraube



8.1.3 Replikat-Implantat mit inserierter Schraube



8.1.4 Replik-Implantat ohne inserierte Schraube



9 Danksagung

An dieser Stelle möchte ich mich herzlich bei allen Personen bedanken, die mich auf meinem Weg zur Promotion jederzeit unterstützt haben.

Mein besonderer Dank gilt meiner Betreuerin und Doktormutter, Frau _____, die mir mit vielen Inspirationen zur Seite stand und mir stets geholfen hat, das Projekt zu realisieren und voranzutreiben. Sie war für mich eine Quelle unermüdlicher Motivation, woraus ich immer meine Kraft und den Willen für den Abschluss der Promotionsschrift schöpfen konnte.

Außerdem möchte ich mich bei Frau _____ für das Zweitgutachten und ihre hilfsbereite und wissenschaftliche Mitbetreuung bedanken.

Ein weiteres Dankeschön möchte ich an Herrn _____ aussprechen, der es mir jederzeit und ganz unkompliziert ermöglicht hat, in seinen Praxisräumen das Röntgentubusgerät zu nutzen, um die vielen notwendigen Röntgenbilder anzufertigen. Darüber hinaus hat er mir seinen 3D-Drucker anvertraut und mich tatkräftig unterstützt mit zahlreichen wertvollen Tipps und Gedanken für die Erstellung eines standardisierten Datensatzes.

Ohne die Mitwirkung der Zahnärzte und Kommilitonen aus den verschiedenen Abteilungen der Klinik und Poliklinik für Zahn-, Mund- und Kieferkrankheiten der Universitätsmedizin Mainz, die sich bereit erklärt haben, an meiner Studie teilzunehmen, hätte ich diese nicht realisieren können. Für euren Einsatz und eure Unterstützung möchte ich euch herzlich danken.

Für die wertvolle Unterstützung, die ich während der Entstehung dieser Arbeit und meines gesamten Studiums erfahren durfte, möchte ich meinen Freunden, meiner Familie und besonders meinen Eltern von ganzem Herzen danken. Ihre stete Fürsorge, ihr Rückhalt und ihre Ermutigung haben mich in allen Phasen begleitet und ohne sie wäre diese Arbeit in der Form nicht möglich gewesen.

10 Tabellarischer Lebenslauf