

Aus der Poliklinik für Zahnerhaltung und Parodontologie
der Ludwig-Maximilians-Universität München



Automatisierte Erkennung und Kategorisierung der Molaren-Inzisiven-Hypomineralisation mit Hilfe künstlicher Intelligenz auf Fotografien von Zähnen

Dissertation

zum Erwerb des Doktorgrades der Zahnmedizin

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität München

vorgelegt von

Jule Schönewolf

aus

Bad Hersfeld

Jahr

2023

Mit Genehmigung der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

Erster Gutachter: *Prof. Dr. med. dent. Jan Kühnisch*

Zweiter Gutachter: Prof. Dr. Roswitha Heinrich-Weltzien

Dritter Gutachter: Priv. Doz. Dr. Felicitas Mayinger

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 17.05.2023

Affidavit



Eidesstattliche Versicherung

Schönewolf, Jule

Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel:

„Automatisierte Erkennung und Kategorisierung der Molaren-Inzisiven-Hypomineralisation mit Hilfe künstlicher Intelligenz auf Fotografien von Zähnen“

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Frankfurt am Main, 17.05.2022
Ort, Datum

Jule Schönewolf
Unterschrift Doktorandin bzw. Doktorand

Inhaltsverzeichnis

Affidavit	3
Inhaltsverzeichnis	4
Abkürzungsverzeichnis	5
Publikationsliste	6
1. Ihr Beitrag zu der Veröffentlichung	7
2. Einleitung	8
3. Ziel des Projektes	12
4. Material und Methodik	13
5. Ergebnisse	15
6. Zusammenfassung:	16
7. Abstract (English):	18
8. Paper I	19
9. Literaturverzeichnis	27
Danksagung	33

Abkürzungsverzeichnis

MIH	Molaren-Inzisiven-Hypomineralisation
KI	Künstliche Intelligenz
EAPD	European Academy of Paediatric Dentistry
DMS V	fünfte deutsche Mundgesundheitsstudie
MIH-TNI Index	MIH-Treatment-Need Index
AUC	Area under curve
ROC	receiver operating characteristic curve
CNN	Convolutional Neural Network

Publikationsliste

Englischsprachige Originalarbeit mit Erstautorenschaft

Schönewolf, J., Meyer, O., Engels, P., Schlickerrieder, A., Hickel, R., Gruhn, V., Hesenius, M., & Kühnisch, J. (2022). *Artificial intelligence- based diagnostics of molar-incisor-hypomineralisation (MIH) on intraoral photographs*. Clin Oral Investig **26**, 5923–5930 (2022). <https://doi.org/10.1007/s00784-022-04552-4>

Englischsprachige Originalarbeit mit Koautorenschaft

Engels, P., Meyer, O., Schönewolf, J., Schlickerrieder, A., Hickel, R., Hesenius, M., Gruhn, V., Kühnisch, J. *Automated detection of posterior restorations in permanent teeth using artificial intelligence on intraoral photographs*. J Dent. 2022 Jun;121:104124. doi: 10.1016/j.jdent.2022.104124. Epub 2022 Apr 5. PMID: 35395346.

Schlickerrieder, A., Meyer, O., Schönewolf, J., Engels, P., Hickel, R., Gruhn, V., Hesenius, M., Kühnisch, J. *Automatized Detection and Categorization of Fissure Sealants from Intraoral Digital Photographs Using Artificial Intelligence*. Diagnostics (Basel). 2021 Sep 3;11(9):1608. doi: 10.3390/diagnostics11091608. PMID: 34573949; PMCID: PMC8469974.

1. Ihr Beitrag zu den Veröffentlichungen

Im Rahmen des Dissertation- und Publikationsprojekts waren die nachfolgenden Arbeitsschritte notwendig. Der eigene Beitrag, als auch jeweils der Projektbeteiligten können der nachstehenden Übersicht entnommen werden.

	Jule Schönewolf	Prof. Dr. Jan Kühnisch	Co-Autoren
Ethikantrag	-	100 %	-
Literaturrecherche	100 %	-	-
Publikation "Artificial intelligence-based diagnostics of molar-incisor-hypomineralisation (MIH) on intraoral photographs"			
Projektidee	50%	50%	-
<i>Schritt 1:</i> Training und Kalibrierung	50 %	50 %	-
Schritt 2: Bilderselektion	100 %	-	-
<i>Schritt 3:</i> Bilderauswertung und Bearbeitung	100 %	-	-
<i>Schritt 4:</i> Einteilung anhand des Klassifizierungssystems	80 %	-	10% P. Engels; 10% A. Schlick- enrieder
<i>Schritt 5:</i> Programmierung der KI und Datenanalyse	-	-	Projektpartner an der Universi- tät Duiseburg- Essen
<i>Schritt 6:</i> Datenselektion und grafische Aufarbeitung	90 %	10 %	-
<i>Schritt 7:</i> Erstellen und Einreichen des Manuskriptes	80 %	10 %	10% Ko-Autoren
	100 %	%	-
Schritt 8: Verfassen der kumulativen Dissertation	100 %	-	-

2. Einleitung

Künstliche Intelligenz (KI) kann bestimmte Verhaltensweisen und Entscheidungsstrukturen von Menschen nachahmen. Dabei werden neuronale Netzwerke erstellt, die dem Nervensystem des Menschen ähneln. KI-Systeme erkennen Bilder und Muster und sind aufgrund dessen fähig, logische Schlüsse zu ziehen. Sie können trainiert werden und erweitern im Rahmen der Nutzung ihren Erfahrungsschatz, wodurch die Qualität der KI-basierten Entscheidungen stetig verbessert wird. In der Zahnmedizin werden KI-basierte Systeme aktuell im Rahmen der Röntgenbild-Diagnostik eingesetzt [Khanagar et al. 2021]. Die detaillierte Untersuchung eines Röntgenbildes durch den Zahnarzt bezüglich pathologischer Befunde ist mitunter zeitintensiv, weshalb der KI-Einsatz mit einer Zeitersparnis und möglicherweise auch mit präziseren Befunden einhergeht [Schwendicke et al. 2020].

In den letzten Jahren wurden Studien veröffentlicht, welche zahnärztliche Befunde auf unterschiedlichen Bildtypen mit Hilfe von KI-basierten Algorithmen analysierten. Als Bildtyp kamen hier bevorzugt Röntgenaufnahmen in verschiedenen Formaten zum Einsatz. Beispielsweise wurden Bissflügelaufnahmen zur automatisierten Kariesdetektion [Lee et al. 2021, Moran et al. 2021, Bayrakdar et al. 2022, Bayraktar und Ayan 2022, Estani et al. 2022, Chen et al. 2022], oder Panoramaschichtaufnahmen zur Erkennung von Karies an dritten Molaren [Vinayahalingam et al., 2021], sowie zur Erkennung und Nummerierung von Zähnen [Bilgir et al. 2021, Yaren et al. 2022, Basaran et al. 2022] als auch zur Diagnostik und Kategorisierung von Zahnrestorationen [Abdalla-Aslan et al. 2020, Çelik et al. 2022, Mao et al. 2021] genutzt. Die KI-basierte Erkennung des periimplantären Knochenniveaus [Cha et al. 2021], parodontaler Knochenverluste [Krois et al. 2019] oder apikaler Parodontitiden [Li et al. 2021, Ekert et al. 2019] auf Zahnfilmen waren Gegenstand weiterer Studien. Gegenüber der KI-basierten Auswertung von Röntgenbildern waren klinische Fotografien von Zähnen bislang nur vereinzelt Gegenstand wissenschaftlicher Betrachtungen, obwohl diese als das digitale Äquivalent zur klinischen Untersuchung anzusehen sind. Diesbezüglich publizierten Kühnisch et al. [2021] sowie Park et al. [2022] und Moutselos et al. [2019] vielversprechende Daten zur Kariesdetektion. Kühnisch et al. erlangten eine Genauigkeit in der Erkennung von Kavitäten verursacht durch Karies von bis zu 93,3%.

Neben der automatisierten Erkennung von Biofilm auf intraoralen Zahnbildern [You et al. 2020 und You et al. 2021] gelang die erfolgreiche Bestimmung von Fissurenversiegelungen mit 98,7% [Schlickerrieder et al., 2021] sowie verschiedenen Restaurationen im Seitenzahnbereich [Engels et al. 2022] mit Ergebniswerten von durchschnittlich 97,1%.

Jedoch liegen etwa zur automatisierten Detektion von Strukturstörungen an den Zähnen bislang nur wenige Studienergebnisse vor. Dies erscheint insofern von Bedeutung, da die Molaren-Inzisiven-Hypomineralisation (MIH) aktuell ein häufiges Erkrankungsbild bei Kindern und Jugendlichen darstellt, welches auch in der Öffentlichkeit vermehrt wahrgenommen wird [Kühnisch et al. 2021].

Die MIH wurde erstmals im Jahre 1987 von einer schwedischen Arbeitsgruppe als „idiopathische Schmelzstörung an bleibenden Zähnen von Kindern“ beschrieben [Koch et al. 1987]. Um die Jahrtausendwende wurde dann der Begriff ‚Molar-Incisor-Hypomineralisation‘ von Weerheijm et al. [2001] geprägt. MIH wird dabei als ein „systemisch bedingter, qualitativer Schmelzdefekt definiert, welcher zumindest an einem ersten bleibenden Molaren auftritt und optional an bleibenden Inzisiven vorliegen kann-“ [Weerheijm et al. 2001, Lygidakis et al. 2010]. Darauf bauend wurden die diagnostischen Kriterien durch die European Academy of Paediatric Dentistry (EAPD) im Jahr 2003 formuliert und im Jahr 2010 aktualisiert [Lygidakis et al. 2010].

Von einer MIH betroffene Zähne zeichnen sich - abhängig von dem Schweregrad - nicht nur durch den Verlust von Zahnhartsubstanz, sondern auch durch Hypersensibilitäten und ein erhöhtes Kariesrisiko aus [Silva et al. 2016]. Dies begründet sich in dem Mineralisationsgrad, welcher mit einem höheren Proteingehalt einhergeht und letztlich die qualitative Strukturstörung beschreibt [Jälevik et al. 2001, Mahoney et al. 2004, Fagrell et al. 2010, Farah et al. 2010, Crombie et al. 2013]. Anatomisch zeichnen sich die fehlstrukturierten Zähne durch eine Veränderung in der Transluzenz des Zahnschmelzes aus [Weerheijm et al. 2001]. Klinisch zeigen die von einer MIH betroffenen Zähne unterschiedlich stark ausgeprägte Opazitäten bis hin zu Schmelzverlusten. Die Mineralisationsstörung ist im Bereich der Inzisiven oft bukkal zu finden, bei Molaren kann sie auf einzelne

Bereiche, die Fissuren oder sogar die gesamte Glattfläche erstreckt sein [Koch et al. 1987]. Ist ein Zahn nur geringfügig betroffen, sind meist weiß-gelbliche oder gelb-bräunliche Verfärbungen zu sehen. Bei schweren Hypomineralisationen kommen fehlende Schmelz- und/oder Dentinareale hinzu. Das Ausprägungsmuster kann innerhalb einer Dentition variieren.

Aktuelle epidemiologische Daten aus der fünften deutschen Mundgesundheitsstudie (DMS V; Kriterium: mindestens ein bleibender Zahn ist betroffen) dokumentierten eine MIH-Häufigkeit von 28,7% [Jordan 2014]. Weltweit wird eine mittlere Prävalenz von 13,1% beschrieben, wobei Unterschiede zwischen verschiedenen Erkrankungsdefinitionen und Regionen dokumentiert sind [Schwendicke et al. 2018]. Obwohl die publizierten Häufigkeitsraten ein vermehrtes Vorkommen weltweit signalisieren, ist demgegenüber die Ätiologie als ungeklärt zu beurteilen [Alaluusua 2010, Crombie et al. 2009]. Es wird aktuell davon ausgegangen, dass es sich um eine multifaktorielle Erkrankung systemischen Ursprungs handelt, wobei die potenziellen Einflussfaktoren offensichtlich vielfältig sein können [Bandeira et al. 2021, Silva et al. 2016].

Die Diagnose von MIH wird ausschließlich anhand des klinischen Erscheinungsbildes im Rahmen einer visuellen Untersuchung gestellt.

Dazu haben die Kriterien der European Academy of Pediatric Dentistry ihre Brauchbarkeit national und international in einer Vielzahl von Studien unter Beweis gestellt. Als Kriterien sind hierbei klar umschriebene Opazitäten, posteruptive Schmelzabbrüche, atypische Restaurationen und Extraktionen zu nennen [Lygidakis et al. 2010].

Darüber hinaus wurde 2016 der MIH-Treatment-Need Index (MIH-TNI) vorgestellt, welcher sowohl die Größe und den Schweregrad des Defekts als auch das Vorhandensein von Hypersensibilitäten mit in Betracht zieht [Bekes und Steffen et al. 2016]. Die Graduierung basiert auf zwei klinischen Merkmalen, welche als wichtigste Leitsymptome benannt werden: vorhandene Hypersensibilitäten und Oberflächeneinbrüche. Zudem bietet der MIH-TNI-Index Handlungsempfehlungen für den behandelnden Zahnarzt [Bekes et al. 2016].

Während die klinische MIH-Diagnostik mit Hilfe der vorhandenen Kriterien gut gelingen kann, bestehen jedoch immer noch diagnostische Unsicherheiten in der Zahnärzteschaft. Daher erscheint es sinnvoll, die Objektivierung der

zahnärztlichen Diagnostik voranzubringen, welche mit Hilfe von Algorithmen der Künstlichen Intelligenz unterstützt werden kann.

3. Ziel dieser Arbeit

Das Hauptziel des vorliegenden Dissertationsprojektes war es, die Möglichkeit zu überprüfen, ob MIH-bedingte Zahnhartsubstanzstörungen automatisiert mit Hilfe von Algorithmen der Künstlichen Intelligenz erkannt werden können. Dazu sollten in der zahnärztlichen Arbeitsgruppe einerseits die Voraussetzungen geschaffen werden und andererseits der entwickelte Algorithmus einer Evaluierung unterzogen werden. Als Hypothese wurde formuliert, dass eine diagnostische Akkuratheit von 90% erreicht werden sollte.

4. Material und Methodik

In einem ersten Arbeitsschritt wurden 3.241 anonymisierte, intraorale, digitale Fotografien, auf denen u.a. Zähne mit einer MIH abgebildet waren, ausgewählt. Alle Fotografien wurden mit professionellen Spiegelreflexkameras erstellt. Zur Qualitätssicherung des Bilddatensatzes wurden Duplikate, über- oder unterbelichtete, oder unscharfe Bilder exkludiert. Darüber hinaus wurden intraorale Fotografien, auf denen karies-bedingte Restaurationen oder Entwicklungsstörungen wie eine Amelogenesis imperfecta diagnostiziert wurden, nicht mit in den Datensatz einbezogen.

Die diagnostische Beurteilung der intraoralen Fotografien erfolgte mit Hilfe der EAPD-Kriterien zur Detektion und Diagnostik einer MIH. Im Detail wurden abgegrenzte Opazitäten, Schmelzeinbrüche und atypische Restaurationen unterschieden. Zusätzlich wurden mögliche zahnärztliche Versorgungen, wie z.B. Fissurenversiegelungen dokumentiert. Zur Sicherstellung der korrekten Befunderhebung wurden alle Diagnostik-Kategorien innerhalb der Arbeitsgruppe überprüft und anschließend vom Studienleiter kontrolliert und korrigiert. Bei Unstimmigkeiten wurden die betreffenden Bilder in der Gruppe diskutiert und ein Konsens hergestellt. Die diagnostischen Entscheidungen stellten die Grundlage für das spätere zyklische Training des KI-Algorithmus dar. Die Intra- und Inter-Untersucher-Reproduzierbarkeit aller mitwirkenden Zahnärzte und Zahnärztinnen lag im Bereich einer hohen Übereinstimmung mit Kappa-Werten zwischen 0,836 und 1,000 (Intra) und 0,693 bis 1,000 (Inter).

Die Gesamtheit aller intraoralen Fotografien (N = 3.241) wurde für das Training und die Validierung des KI-Algorithmus in ein Trainingsset (N = 2.596) und ein unabhängiges Testset (N=649) geteilt. Letzteres wurde dem KI-Algorithmus nur zur Überprüfung des Lernerfolges vorgelegt und keinesfalls für Trainingszwecke verwendet. Die Entwicklung des KI-Algorithmus erfolgte durch den Projektpartner an der Universität Duisburg und Essen und nutzte dazu etablierte Vorgehensweisen. Die statistische Datenanalyse umfasste die Berechnung der diagnostischen Genauigkeiten unter Einbeziehung üblicher Maßzahlen, wie Akkuratheit, Sensitivität oder Spezifität. Im Weiteren umfasste die statistische Analyse die

Berechnung von Kontingenztabelle, Flächen unter der Receiver Operating Characteristic Curve (ROC), Area under curve (AUCs) und Saliency Maps.

5. Ergebnisse

Im Anschluss an das KI-Training gelang es der KI in acht von neun diagnostischen Kategorien, eine diagnostische Genauigkeit von über 90% abzubilden. Im Detail wurde eine diagnostische Genauigkeit von 95,2% dokumentiert. Die korrekte Beurteilung der diagnostischen Kategorien variierte und lag bei 91,5% (Schmelzeinbruch/keine Restauration) und 99,1% (Schmelzeinbruch/Fissurenversiegelung). In der Kategorie diffuse Opazität/ keine Intervention wies die KI mit einem Ergebniswert von 88,4% hingegen die niedrigste diagnostische Akkuratheit auf. Mithin konnte einzig in dieser Kategorie nicht das gewünschte Ziel einer Übereinstimmung in Höhe von mindestens 90% erreicht werden.

Die korrekte Erkennung und Beurteilung von MIH-Zähnen unter Zuhilfenahme von Deep-Learning-basierter CNN (Convolutional Neural Network) zeigte sich in der Gesamtheit mit einer diagnostischen Genauigkeit von über 90% erfolgreich.

6. Zusammenfassung

Das Ziel des vorliegenden Studienprojektes, einen auf KI-basierten Algorithmus zur automatisierten Erkennung von MIH zu entwickeln, wurde erfolgreich erreicht. Die erreichte diagnostische Akkuratheit lag mit durchschnittlich 95.2% etwas über den hypothetisch erwarteten 90% und ist für den gewählten Studienansatz als auch die weiterführenden Arbeiten als erfolgreich zu beurteilen.

Da es bislang kaum vergleichbare Studien zur KI-basierten bzw. automatisierten MIH-Erkennung gibt, sind keine umfangreichen Vergleiche möglich. Einzig publizierten Alevizakos et al. [2022] weitgehend zeitgleich Daten zur Thematik. In diesem Studienprojekt fanden insgesamt 463 klinische Bilder, auf denen MIH, Amelogenesis imperfecta, Fluorose und Kariesvorstufen abgebildet waren, Verwendung. Unter den fünf verwendeten Netzwerken wurde die niedrigste diagnostische Genauigkeit mit dem vgg-16 Netzwerk und einem Ergebnis von 80% erzielt. Die höchste Genauigkeit lag bei 92,9% für das dense121-Netzwerk. Die durchschnittliche Trefferquote betrug 88,4% [Alevizakos et al. 2022]. Keine Ergebnisse lieferte die Studie zur Detektionsgenauigkeit von einzelnen MIH-Kategorien.

In einer weiteren, kürzlich veröffentlichten Studie von Kühnisch et al. [2022] über die Erkennung von Karies auf Grundlage klinischer Bilder wurde eine Genauigkeit von 85,7% erzielt. Ähnliche Ergebnisse wurden bei einer Studie über die Erkennung und Klassifizierung von Restaurationen im Seitenzahnbereich von Engels et al. [2022] mit einer Genauigkeit von durchschnittlich 97,1% erzielt. Bei der Kategorisierung von intakten, unzureichenden, ungenügenden und nicht vorhandenen Fissurenversiegelungen erreichte die Studie von Schlickerrieder et al. [2021] ein Ergebnis von 91%. Ebenso gelang You et al. [2020] die Plaque-Erkennung auf Milchzähnen mit einem KI-basierten System zu diagnostizieren. Das Ergebnis wies hierbei eine Genauigkeit von durchschnittlich 72,6% auf.

Zusammenfassend lässt sich sagen, dass die Stärken dieses Studienprojektes in der Verwendung einer vergleichsweise hohen Anzahl an Bildern und eines leistungsfähigen Netzwerkes liegen. Zur Erhöhung der Ergebnisgenauigkeit sollten in einem nächsten Schritt Bilder mit pixelgenauen Markierungen genutzt werden. Zur Umsetzung dessen bedarf es einer qualifizierten Arbeitsgruppe sowie leistungsfähiger Rechenressourcen.

Zwar ist die Anwendung KI-basierter Systeme noch nicht in den zahnmedizinischen Arbeitsalltag implementiert, auch ist als wesentlicher Grund hierfür die aktuell eingeschränkte Verfügbarkeit anzuführen. Außerdem sollten Fragen der Ethik und Verantwortung vorab verbindlich im Konsens geklärt werden [Schwendicke et al. 2020].

Umgekehrt aber zeigen die Studienergebnisse, dass KI-basierte Diagnostiksysteme dem klinischen Praktiker erhebliche Vorteile bieten könnten. Ähnlich einer «zweiten Meinung» könnte hierdurch nicht nur dem Behandler, sondern auch dem Patienten eine Befundobjektivierung ermöglicht werden. Das KI-basierte System würde somit die Diagnostik des Behandlers unterstützen und dem Patienten eine Visualisierung der Erkrankung ermöglichen [Kosan et al. 2022].

7. Abstract (English):

A deep learning-based convolutional neural network (CNN) could improve dental diagnostic accuracy by automated detection and categorization of molar-incisor-hypomineralisation (MIH) on intra-oral photographs.

For the purpose of this study on artificial intelligence (AI), an image set consisting of 3.241 intraoral images was split into training (N = 2.596) and test set (N = 649). The overall dataset was classified into the following categories: teeth with no signs of hypomineralisation and no dental intervention (N = 767), teeth with no signs of hypomineralisation and a MIH-related “atypical” restoration (N = 76), teeth with no signs of hypomineralisation and presence of pit and fissure sealant (N = 742), teeth with hypomineralisation and intervention (N = 815), teeth with hypomineralisation and an atypical restoration (N = 158), teeth with hypomineralisation and fissure sealing (N = 181), teeth with enamel disintegration and no intervention (n = 290), teeth with enamel disintegration and an atypical restoration (N = 169) and teeth with enamel disintegration and presence of sealing material (N = 43). After the cyclic training of the convolutional neural network most dental photographs could be automatically classified with an acceptable diagnostic accuracy. Here, an overall diagnostic accuracy of 95.2% was achieved. AUC values ranged from 0.873 (enamel breakdown with a sealant) to 0.994 (atypical restoration with no MIH).

It can be concluded that AI powered MIH detection and diagnostics showed promising results in the diagnostic study. However, there is a substantial need for further improvements.

8. Veröffentlichung I

Clinical Oral Investigations

2022 Journal Impact Factor 3,573

Volume 26, Issue 9

24 May 2022

Pages 5923-5930

Artificial intelligence-based diagnostics of molar-incisor-hypomineralization (MIH) on intraoral photographs.

Jule Schönewolf, Ole Meyer, Paula Engels, Anne Schlickerrieder, Reinhard Hickel, Mar Hesenius, Jan Kühnisch



Artificial intelligence-based diagnostics of molar-incisor-hypomineralization (MIH) on intraoral photographs

Jule Schönewolf¹ · Ole Meyer² · Paula Engels¹ · Anne Schlickerrieder¹ · Reinhard Hickel¹ · Volker Gruhn² · Marc Hesenius² · Jan Kühnisch¹

Received: 11 February 2022 / Accepted: 13 May 2022
 © The Author(s) 2022

Abstract

Objective The aim of this study was to develop and validate a deep learning–based convolutional neural network (CNN) for the automated detection and categorization of teeth affected by molar-incisor-hypomineralization (MIH) on intraoral photographs.

Materials and methods The data set consisted of 3241 intraoral images (767 teeth with no MIH/no intervention, 76 with no MIH/atypical restoration, 742 with no MIH/sealant, 815 with demarcated opacity/no intervention, 158 with demarcated opacity/atypical restoration, 181 with demarcated opacity/sealant, 290 with enamel breakdown/no intervention, 169 with enamel breakdown/atypical restoration, and 43 with enamel breakdown/sealant). These images were divided into a training ($N=2596$) and a test sample ($N=649$). All images were evaluated by an expert group, and each diagnosis served as a reference standard for cyclic training and evaluation of the CNN (ResNeXt-101–32×8d). Statistical analysis included the calculation of contingency tables, areas under the receiver operating characteristic curve (AUCs) and saliency maps.

Results The developed CNN was able to categorize teeth with MIH correctly with an overall diagnostic accuracy of 95.2%. The overall SE and SP amounted to 78.6% and 97.3%, respectively, which indicate that the CNN performed better in healthy teeth compared to those with MIH. The AUC values ranging from 0.873 (enamel breakdown/sealant) to 0.994 (atypical restoration/no MIH).

Conclusion It was possible to categorize the majority of clinical photographs automatically by using a trained deep learning–based CNN with an acceptably high diagnostic accuracy.

Clinical relevance Artificial intelligence-based dental diagnostics may support dental diagnostics in the future regardless of the need to improve accuracy.

Keywords Chalky teeth · Automated image analysis · Convolutional neural networks · Deep learning · Transfer learning

Introduction

Visual examination is the method of choice for screening, monitoring, detecting, and diagnosing dental pathologies of teeth, and the corresponding diagnostic indices and methodological procedures have been described by the researchers [e.g., 1, 2]. However, the knowledge transfer from scientists

to dental practitioners might sometimes be lacking, which is especially true for detecting and diagnosing individuals or teeth with molar-incisor-hypomineralization (MIH). Here, families notified diagnostic uncertainties by dental professionals which potentially results in conflicting positions, diverging recommendations and additional dental consultations [3, 4]. It might be beneficial to develop diagnostic methods to verify suspected dental hard tissue findings independently from the investigating dentist. In addition, this aim might be supported by the documented MIH prevalence rates. The mean global MIH prevalence was estimated recently at 13.1% by Schwendicke et al. [5]. In Germany, 28.7% of all 12-year-olds were found to have hypomineralizations [6, 7]. Both numbers indicate that a relevant proportion of adolescents is affected by this developmental

✉ Jan Kühnisch
 jkuehn@dent.med.uni-muenchen.de

¹ Department of Conservative Dentistry and Periodontology, University Hospital, Ludwig-Maximilians University Munich, Goethestraße 70, 80336 Munich, Germany

² Institute for Software Engineering, University of Duisburg-Essen, Essen, Germany

disorder. Therefore, diagnosing and managing MIH is a frequent challenge in daily dental practice.

The aim of establishing independent diagnostic methods might become feasible by the availability of smart image analysis methods. Artificial intelligence (AI) currently offers the potential for the automated detection and evaluation of diagnostic information in medicine and dentistry [6–9]. The aim to digitalise medical and dental workflows must be understood as an emerging topic, and interest in this area has recently increased in dental research as well. Meanwhile, different workgroups have started to analyze all available types of dental radiographs [10–14] by using deep learning with convolutional neural networks (CNNs) for the detection of caries [15], apical pathologies [16], or periodontitis [17]. In contrast, only a few projects using AI-based algorithms for the automated identification of pathologies on intraoral clinical photographs have been reported [18–26]. When considering recently published reports and the latest software developments, it can be stated that, to the best of our knowledge, no application for the automated detection of MIH on intraoral photographs has been developed and/or evaluated thus far. Therefore, this diagnostic study aimed to train a CNN for MIH detection (test method); this CNN was then compared in its final

stage to the expert evaluation (reference standard). The aim was to reach a diagnostic accuracy of at least 90% for the test method.

Materials and methods

Study design

This diagnostic study used anonymized intraoral clinical photographs (Fig. 1) from clinical situations in which photographs were captured for educational purposes as well as from previously conducted clinical trials. The Ethics Committee of the Medical Faculty of the Ludwig-Maximilians University of Munich reviewed and approved the study concept (project number 020–798). This investigation was reported in accordance with the recommendations of the Standard for Reporting of Diagnostic Accuracy Studies (STARD) steering committee [27] and recently published recommendations for the reporting of AI studies in dentistry [28]. The pipeline of methods, mentioned below, was applied and described in previously published reports [19, 20].

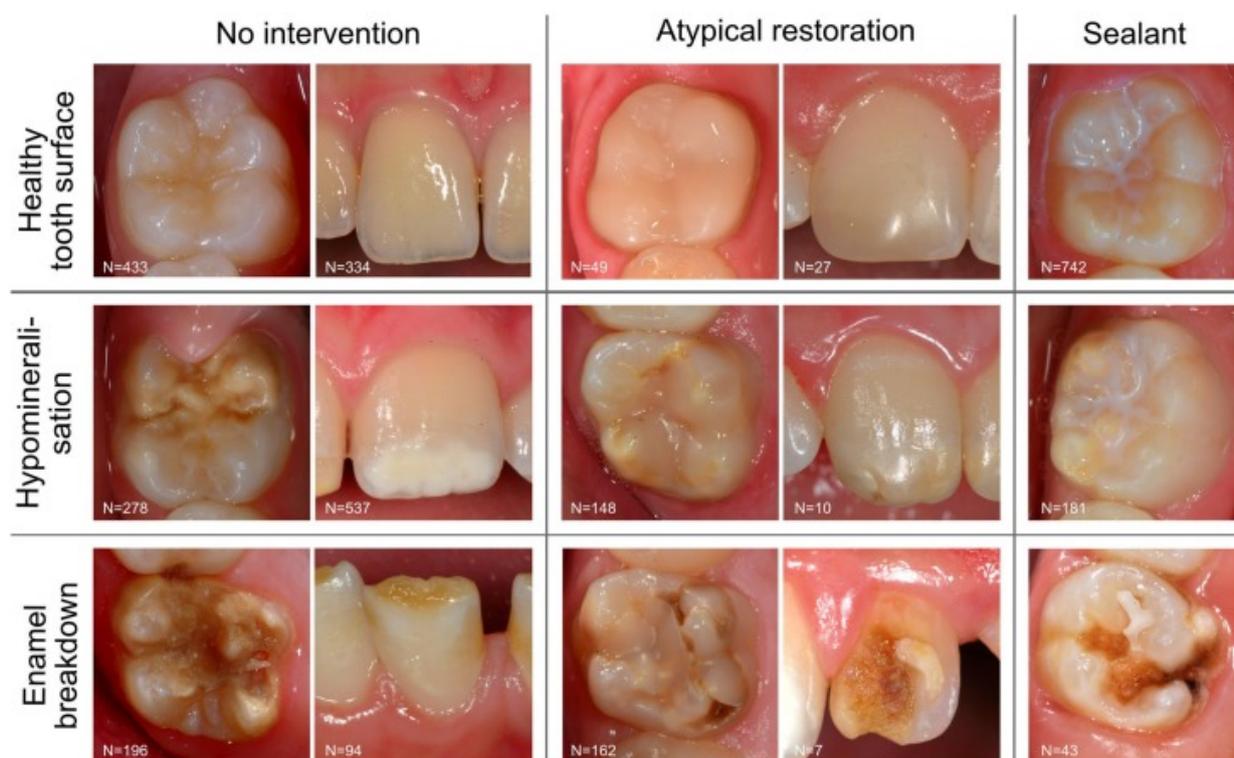


Fig. 1 Overview of the chosen diagnostic categories based on the criteria provided by the European Academy of Paediatric Dentistry [3] and frequent intervention modalities

Intraoral photographs

Dental photographs were consistently taken with professional single-reflex cameras equipped with a 105-mm macro lens and a macro flash after tooth cleaning and drying [19, 20]. All images were stored (jpeg format, RGB colors, aspect ratio of 1:1) and selected for this study project. To ensure high data quality, duplicate or inadequate photographs, such as out-of-focus images, under- or overexposed pictures and photographs with saliva contamination, were excluded. Clinical photographs showing additional caries cavities and any other developmental disorders, e.g., amelogenesis or dentinogenesis imperfect or hypoplasia, were omitted. Caries-related restorations were also excluded to rule out potential evaluation bias. Finally, 3241 anonymized, high-quality clinical photographs from anterior and posterior permanent teeth with MIH (test group) and without any pathology/restoration (control group) were included in the study.

Classification of teeth with MIH (reference standard)

Each photograph was classified with the goal of detecting and categorizing teeth with MIH in relation to the diagnostic classification system of the European Academy of Paediatric Dentistry [3] and possible dental interventions, such as restorations or fissure sealants. In detail, characteristics indicating the well-established MIH categories of demarcated opacities and enamel breakdowns are prevalent and can appear clinically in combination without any dental restoration, with an MIH-related—so called atypical—restoration or sealant (Fig. 1). Each image was pre-categorized by three graduated dentists (JS, PE, and AS) according to the given cross classification; afterwards, images were independently counterchecked by an experienced examiner (JK, > 20 years of clinical practice and scientific experience). In the case of divergent findings, each intraoral photograph was re-evaluated and discussed until consensus was reached. Every diagnostic decision—one per image—served as a reference standard for cyclic training and repeated evaluation of the deep learning-based CNN.

All the annotators were trained and calibrated before the study. During a 2-day theoretical and practical workshop guided by the principal investigator (JK), all annotators (JS, PE, and AS) were educated. Finally, 140 photographs were evaluated by all participating dentists to determine intra/interexaminer reproducibility for MIH classifications. Statistically, kappa values were computed for all coder pairs using Excel (Excel 2016, Microsoft, Redmond, WA, USA) and SPSS (SPSS Statistics 27, 2020, IBM corporation, Armonk, NY, USA). Intra/interexaminer reproducibility was calculated as 0.964/0.840–0.712 (JS), 0.982/0.747–0.727 (PE), 1.000/0.774–0.693 (AS), and 0.836/0.749–0.693 (JK),

respectively. The documented kappa values indicated substantial to perfect agreement [29].

Training of the deep learning-based CNN (test method)

In the following, the used pipeline of methods for developing the AI-based algorithm is described. Before training, the whole set of images ($N=3241$) was divided into a training sample ($N=2596$) and a test sample ($N=649$); the CNN had no knowledge of the latter during training; it served as an independent test set only. The distribution of all images in relation to the diagnostic classification can be taken from Table 1.

To increase variability within the images, the underlying training set was augmented. For this purpose, the randomly selected images (batch size = 16) were multiplied by a factor of ~5, altered by different transformations (random center and margin cropping by up to 30% each; random deletion removing up to 30%; random affine transformation up to 180°; random perspective transformation up to a distortion of 0.5; and random changes in brightness, contrast, and saturation up to 10%) and resized (300 × 300 pixels). In addition, to compensate for under- and overexposure, all images were normalized [19, 20]. Torchvision (version 0.9.1, <https://pytorch.org>) in conjunction with the PyTorch library (version 1.8.1, <https://pytorch.org>) was used. ResNeXt-101–32 × 8d [30] was selected as the basis for the continuous adaptation of the CNN for MIH detection and categorization. The CNN was trained using backpropagation to determine the gradient for learning. Backpropagation was repeated iteratively for images and labels using the abovementioned batch size

Table 1 Description of the image set in relation to the diagnostic classification

Restoration status	MIH classification	Training sample	Test sample	Sum
No intervention	No MIH	627	140	767
	Demarcated opacity	659	156	815
	Enamel breakdown	232	58	290
Atypical restoration	No MIH	59	17	76
	Demarcated opacity	127	31	158
	Enamel breakdown	123	46	169
Sealant	No MIH	585	157	742
	Demarcated opacity	147	34	181
	Enamel breakdown	33	10	43
	Sum	2596	649	3241

and parameters. Overfitting was prevented by two measures: selecting a low learning rate (0.0001) and performing dropout (at a rate of 0.5) on the final linear layers as a regularization technique. CNN training was repeated over 15 epochs with cross entropy loss as an error function and the application of the Adam optimizer (betas 0.9 and 0.999, epsilon $1e-8$). With an open-source neural network employing pretrained weights (ResNeXt-101–32×8d pretrained on ImageNet, Stanford Vision and Learning Laboratory, Stanford University, Palo Alto, CA, USA), CNN training was accelerated. Existing learning results regarding the recognition of basic structures in the existing image set could thus be reused and skipped in the initial training. Training was performed on a university-based computer with the following specifications: RTX A6000 48 GB (Nvidia, Santa Clara, CA, USA); i9 10850 K 10×3.60 GHz (Intel Corp., Santa Clara, CA, USA) and 64 GB RAM [19, 20].

Statistical analysis

The data were analyzed using Python (<http://www.python.org>, version 3.8). The overall diagnostic accuracy ($ACC = (TNs + TP) / (TNs + TP + FN + FP)$) was determined by calculating the number of true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs). The sensitivity (SE), specificity (SP), positive and negative predictive values (PPVs and NPVs, respectively), and the area under the receiver operating characteristic (ROC) curve (AUC) were computed for the chosen MIH categorization [31]. Saliency maps were plotted to illustrate image areas that were used by the CNN to make individual decisions. The saliency maps were calculated by

back propagating the CNN prediction and visualizing the gradient of the input of the resized images [19, 20, 32].

Results

After the deep learning–based CNN was trained, the CNN was able to detect MIH and correlated interventions correctly in eight out of nine MIH categories with a diagnostic accuracy higher than 90% (Table 2). The overall diagnostic accuracy was determined at 95.2%. The SE and SP amounted to 78.6% and 97.3%, respectively. In detail, the accuracy values ranged from 91.5% (enamel breakdown/no intervention) to 99.1% (enamel breakdown/sealant). The lowest diagnostic accuracy of 88.4% was found for demarcated opacities with no intervention (Table 2). This was the only category—one out of nine—where the target accuracy of 90% was not reached (Table 2).

When considering the diagnostic parameters of SE and SP in detail (Table 2), it is important to note that SP values were found to be consistently high, ranging from 92.9% (no intervention/demarcated opacity) to 100.0% (sealant/enamel breakdown) in comparison to the SE. The latter ranged from 40.0% (enamel breakdown/sealant) to 96.2% (sealant/no MIH). The AUC values varied from 0.873 (enamel breakdown/sealant) to 0.994 (sealant/no MIH). With respect to the overall high AUC values, no ROC curves were plotted.

The confusion matrix (Fig. 2) illustrates the case distribution in the test set. Here, it also became obvious that the majority of diagnostic predictions by the AI-based algorithm (test method) were made in accordance with the expert decision in the test set. However, a distinct number of cases were

Table 2 Overview of the diagnostic performance of the developed convolutional neuronal network (CNN), where the independent test set ($n = 649$ images) was evaluated by the AI-based algorithm for the detection of MIH-related enamel disturbances and related interven-

tions. The overall diagnostic accuracy (ACC, including the sensitivity (SE), the specificity (SP), the negative predictive value (NPV), the positive predictive value (PPV) and the area under the receiver operating characteristic curve (AUC)) was computed

Category		True positives (TPs)		True negatives (TNs)		False positives (FPs)		False negatives (FNs)		Diagnostic performance					
		N	%	N	%	N	%	N	%	ACC	SE	SP	PPV	NPV	AUC
No intervention	No MIH	128	19.7	485	74.7	24	3.7	12	1.9	94.5	91.4	95.3	84.2	97.6	0.985
	Demarcated opacity	116	17.9	458	70.6	35	5.4	40	6.1	88.4	74.4	92.9	76.8	92.0	0.922
	Enamel breakdown	37	5.7	557	85.9	34	5.2	21	3.2	91.5	63.8	94.3	52.1	96.4	0.901
Atypical restoration	No MIH	12	1.9	630	97.1	2	0.3	5	0.7	98.9	70.6	99.7	85.7	99.2	0.987
	Demarcated opacity	15	2.3	611	94.1	7	1.1	16	2.5	96.5	48.4	98.9	68.2	97.5	0.953
	Enamel breakdown	30	4.6	584	90.0	19	2.9	16	2.5	94.6	65.2	96.9	61.2	97.3	0.938
Sealant	No MIH	151	23.3	480	74.0	12	1.9	6	0.8	97.2	96.2	97.6	92.6	98.8	0.994
	Demarcated opacity	17	2.6	609	93.9	6	0.9	17	2.6	96.5	50.0	99.0	73.9	97.3	0.916
	Enamel breakdown	4	0.6	639	98.5	0	0	6	0.9	99.1	40.0	100.0	100.0	99.1	0.873
	Σ	510	8.7	5053	86.5	139	2.4	139	2.4	95.2	78.6	97.3	78.6	97.3	n.c

n.c., not calculable

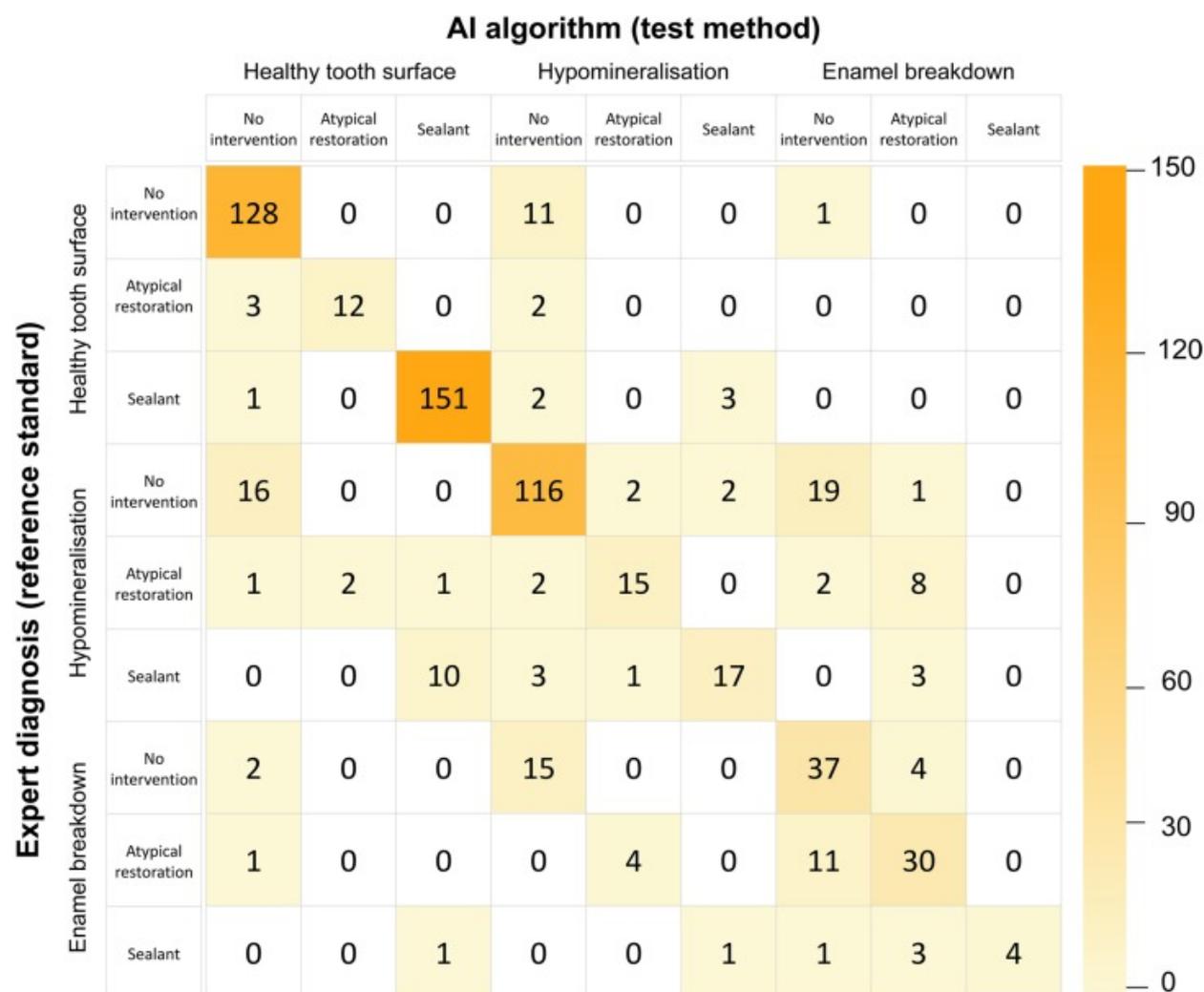


Fig. 2 The confusion matrix shows the case distribution between the convolutional neuronal network (CNN, test method) and expert diagnosis for MIH assessment in the independent test set ($n = 649$ images)

not categorized correctly, especially if multiple characteristics were present on one photograph. In addition to the explorative data analysis, exemplary saliency maps (Fig. 3) are shown to illustrate areas on each intraoral photograph that the CNN used for decision-making.

Discussion

The present diagnostic study demonstrated that an AI-based algorithm is able to detect MIH on intraoral photographs with a moderately high diagnostic accuracy (Table 2). With respect to the fact that accuracy > 90% was achieved in eight out of nine categories, the initially formulated hypothesis was accepted. When considering the documented accuracy and AUC values (Table 2), it could be further concluded

that on the one hand, the overall diagnostic performance appears to be satisfactory but on the other hand, the partially low SE and high SP values indicate that the reported data need to be interpreted with caution. In detail, SP played a far more important role in this image sample probably because of the higher number of teeth without MIH. Therefore, the diagnostic accuracy is mainly driven by the SP rather than the SE and it could be argued that the AI-based algorithm is better in scoring sound teeth compared to MIH teeth. In this context, the complex clinical appearance of teeth with MIH, especially molars, needs to be highlighted. In addition to the fact that multiple findings can be present in teeth with MIH, this information will be further enhanced on intraoral photographs, which currently have a good resolution and can be thoroughly evaluated by the study team. Here, several demarcated opacities were found to have more

detection on primary teeth, where an accuracy of 86.0% was reached. Noncavitated and cavitated caries lesions were detected with accuracies of 92.5% and 93.3%, respectively [19]. In other recently published diagnostic studies, white spot lesions were registered automatically with 81–84% accuracy [26] and caries lesions were classified and located with a mean AUC of 85.6% [33]. When also considering the available diagnostic performance data for various dental findings on different types of X-ray images [10–13, 15, 16, 18–20, 34, 35], it can be emphasized that the documented diagnostic accuracies in this trial are on the same order of magnitude compared to those of several other dental reports.

When summarizing the methodological strengths of this study project, it can be concluded that it was technically feasible to develop CNNs with substantial precision by using the described pipeline for software development. Therefore, it can be predicted that AI-based diagnostics will gain increasing attention in dentistry in the near future. However, further developments are needed before they can be used in a clinical setting [35, 36]. Moreover, it is crucial to assess the necessity of numerical extensive and qualitative image material to further improve the performance of the developed CNN for MIH categorization. Simultaneously, less frequent diagnostic categories should be included in appropriate numbers as well. Independently from this, it should further be noted that AI-based algorithms need to be also developed for rare developmental disorders, e.g., dentinogenesis imperfecta or amelogenesis imperfecta. The chosen methodology primarily presents a simple approach to handle dental diagnoses and is typically linked with diagnostic accuracy values of approximately 90% (Table 2, Fig. 2). Aiming at increasing diagnostic performance up to 100%, the methodological requirements for consistent improvement of the data set and detailed image annotation by pixelwise labeling have been expressed. Another aim might be to perform CNN training on high-performance computers to reach a higher degree of neuronal connectivity. However, all these requirements will necessitate more time and personal and computing resources.

Conclusion

It was possible in the present study to automatically categorize clinical photographs from teeth with MIH by using a trained deep learning-based CNN with an overall diagnostic accuracy of 95.2%. The higher NPV and SP values in comparison to PPV and SE indicate that the CNN performed better in healthy teeth compared to those with MIH. Future improvements are necessary to increase the diagnostic performance.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The developed AI-based algorithm can be made available as a web application. In case of interest in our software solutions, please contact the author group or visit: <https://dental-ai.de>.

Declarations

Ethics approval This study was approved by the Ethics Committee of the Medical Faculty of the Ludwig-Maximilians-University of Munich (project number 020–798). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Consent to participate Authorization was obtained from all participating patients.

Conflict of interest The authors declare no competing interests.

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

References

1. Lygidakis NA, Garot E, Somani C, Taylor GD, Rouas P, Wong FSL (2022) Best clinical practice guidance for clinicians dealing with children presenting with molar-incisor-hypomineralisation (MIH): an updated European Academy of Paediatric Dentistry policy document. *Eur Arch Paediatr Dent* 23:3–21
2. Lygidakis NA, Wong F, Jälevik B, Vierrou AM, Alaluusua S, Espelid I (2010) Best clinical practice guidance for clinicians dealing with children presenting with molar-incisor-hypomineralisation (MIH): an EAPD policy document. *Eur Arch Paediatr Dent* 11:75–81
3. Moreno T, Sanz JL, Melo M, Llena C (2021) Overtreatment in restorative dentistry: decision making by last-year dental students. *Int J Environ Res Public Health* 18(23):12585
4. Gupta P, Gupta M, Koul N (2020) Overdiagnosis and overtreatment; how to deal with too much medicine. *J Family Med Prim Care* 9(8):3815–3819
5. Schwendicke F, Elhennawy K, Reda S, Bekes K, Manton DJ, Krois J (2018) Global burden of molar incisor hypomineralization. *J Dent* 68:10–18
6. Jordan AR, Micheelis W (2016) Fünfte Deutsche Mundgesundheitsstudie. Deutscher Zahnärzte Verlag, Cologne
7. Schwendicke F, Samek W, Krois J (2020) Artificial intelligence in dentistry: chances and challenges. *J Dent Res* 99:769–774

8. Khanagar SB, Al-Ehaideb A, Maganur PC, Vishwanathiah S, Patil S, Baeshen HA, Sarode SC, Bhandi S (2021) Developments, application, and performance of artificial intelligence in dentistry - a systematic review. *J Dent Sci* 16:508–522
9. Grischke J, Johannsmeier L, Eich L, Griga L, Haddadin S (2020) Dentronics: towards robotics and artificial intelligence in dentistry. *Dent Mater* 36:765–778
10. Cantu AG, Gehrung S, Krois J, Chaurasia A, Rossi JG, Gaudin R, Elhennawy K, Schwendicke F (2020) Detecting caries lesions of different radiographic extension on bitewings using deep learning. *J Dent* 100:103425
11. Krois J, Ekert T, Meinhold L, Golla T, Kharbot B, Wittemeier A, Dörfler C, Schwendicke F (2020) Deep learning for the radiographic detection of periodontal bone loss. *Sci Rep* 9:8495
12. Kim JE, Nam NE, Shim JS, Jung YH, Cho BH, Hwang JJ (2020) Transfer learning via deep neural networks for implant fixture system classification using periapical radiographs. *J Clin Med* 9:1117
13. Abdalla-Aslan R, Yeshua T, Kabla D, Leichter I, Nadler C (2020) An artificial intelligence system using machine-learning for automatic detection and classification of dental restorations in panoramic radiography. *Oral Surg Oral Med Oral Pathol Oral Radiol* 130:593–602
14. Schwendicke F, Elhennawy K, Paris S, Friebertshäuser P, Krois J (2020) Deep learning for caries lesion detection in near-infrared light transillumination images: a pilot study. *J Dent* 92:103260
15. Lee JH, Kim DH, Jeong SN, Choi SH (2018) Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* 77:106–111
16. Cha JY, Yoon HI, Yeo IS, Hun KH, Han JS (2021) Peri-implant bone loss measurement using a region-based convolutional neural network on dental periapical radiographs. *J Clin Med* 10:1009
17. Lee JH, Kim DH, Jeong SN, Choi SH (2018) Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. *J Periodontol Implant Sci* 48:114–123
18. Li RZ, Zhu JX, Wang YY, Zhao SY, Peng CF, Sun Q, Hao AM, Li S, Wang Y, Xia B (2021) Development of a deep learning based prototype artificial intelligence system for the detection of dental caries in children. *Zhonghua Kou Qiang Yi Xue Za Zhi* 56(12):1253–1260
19. Kühnisch J, Meyer O, Hesenius M, Hickel R, Gruhn V (2022) Caries detection on intraoral images using artificial intelligence. *J Dent Res* 101:158–165
20. Schlickerieder A, Meyer O, Schönewolf J, Engels P, Hickel R, Gruhn V, Hesenius M, Kühnisch J (2021) Automated detection and categorization of fissure sealants from intraoral digital photographs using artificial intelligence. *Diagnostics (Basel)* 11(9):1608
21. You W, Hao A, Li S, Wang Y, Xia B (2020) Deep learning-based dental plaque detection on primary teeth: a comparison with clinical assessments. *BMC Oral Health* 20(1):141
22. Pauwels R (2021) A brief introduction to concepts and applications of artificial intelligence in dental imaging. *Oral Radiol* 37(1):153–160
23. You WZ, Hao AM, Li S, Zhang ZY, Li RZ, Sun RQ, Wang Y, Xia B (2021) Deep learning based dental plaque detection on permanent teeth and the influenced factors. *Zhonghua Kou Qiang Yi Xue Za Zhi* 56(7):665–671
24. Schwendicke F, Golla T, Dreher M, Krois J (2019) Convolutional neural network for dental image diagnostics: A scoping review. *J Dent* 91:103226
25. Takahashi T, Nozaki K, Gonda T, Mameno T, Ikebe K (2021) Deep learning-based detection of dental prostheses and restorations. *Sci Rep* 11(1):1960.
26. Askar H, Krois J, Rohrer C, Mertens S, Elhennawy K, Ottolenghi L, Mazur M, Paris S, Schwendicke F (2021) Detecting white spot lesions on dental photography using deep learning: a pilot study. *J Dent* 107:103615
27. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, de Vet HC, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L, Korevaar DA, Cohen JF (2015) STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 351:h5527
28. Schwendicke F, Singh T, Lee JH, Gaudin R, Chaurasia A, Wiegand T, Uribe S, Krois J (2021) Artificial intelligence in dental research: checklist for authors, reviewers, readers. *J Dent* 107:103610
29. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1):159–174
30. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. *JMLR* 15:1929–1958
31. Matthews DE, Farewell VT (2015) *Using and understanding medical statistics*, 5th edn. Karger, Basel
32. Simonyan K, Vedaldi A, Zisserman A (2021) Deep inside convolutional networks: visualising image classification models and saliency maps. In: *In workshop at international conference on learning representations 2014*. Accessed 31 January 2022. <https://arxiv.org/pdf/1312.6034.pdf>.
33. Zhang X, Liang Y, Li W, Liu C, Gu D, Sun W (2022) Miao L (2022) Development and evaluation of deep learning for screening dental caries from oral photographs. *Oral Dis* 28:173–181
34. Bayraktar Y, Ayan E (2022) Diagnosis of interproximal caries lesions with deep convolutional neural network in digital bitewing radiographs. *Clin Oral Investig* 26(1):623–632
35. Lee S, Oh SI, Jo J, Kang S, Shin Y, Park JW (2021) Deep learning for early dental caries detection in bitewing radiographs. *Sci Rep* 11:16807
36. Engels P, Meyer O, Schönewolf J, Schlickerieder A, Hickel R, Hesenius M, Gruhn V, Kühnisch J (2022) Automated detection of posterior restorations in permanent teeth on intraoral photographs using artificial intelligence. *J Dent* 121:104124

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

9. Literaturverzeichnis

Abdalla-Aslan, R., Yeshua, T., Kabla, D., Leichter, I., & Nadler, C. (2020). *An artificial intelligence system using machine-learning for automatic detection and classification of dental restorations in panoramic radiography*. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 130(5), 593–602.

Alaluusua, S. (2010). *Aetiology of Molar-Incisor Hypomineralisation: A systematic review*. *European Archives of Paediatric Dentistry: Official Journal of the European Academy of Paediatric Dentistry*, 11(2), 53–58.

Alevizakos, V., Bekes, K., Steffen, R., & von See, C. (2022). *Artificial intelligence system for training diagnosis and differentiation with molar incisor hypomineralization (MIH) and similar pathologies*. *Clinical Oral Investigations*, 26(12), 6917–6923.

Bandeira Lopes, L., Machado, V., Botelho, J., & Haubek, D. (2021). *Molar-incisor hypomineralization: an umbrella review*. *Acta Odontologica Scandinavica*, 79(5), 359–369.

Başaran, M., Çelik, Ö., Bayrakdar, I. S., Bilgir, E., Orhan, K., Odabaş, A., Aslan, A. F., & Jagtap, R. (2022). *Diagnostic charting of panoramic radiography using deep-learning artificial intelligence system*. *Oral Radiology*, 38(3), 363–369.

Bayrakdar, I. S., Orhan, K., Akarsu, S., Çelik, Ö., Atasoy, S., Pekince, A., Yasa, Y., Bilgir, E., Sağlam, H., Aslan, A. F., & Odabaş, A. (2022). *Deep-learning approach for caries detection and segmentation on dental bitewing radiographs*. *Oral Radiology*, 38(4), 468–479.

Bayraktar, Y., & Ayan, E. (2022). *Diagnosis of interproximal caries lesions with deep convolutional neural network in digital bitewing radiographs*. *Clinical Oral Investigations*, 26(1), 623–632.

Bekes, K., Krämer, N., van Waes, H., & Steffen, R. (2016). *Das Würzburger MIH-Konzept: Teil 2. Der Therapieplan*. *Oralprophylaxe & Kinderzahnheilkunde*, 38(4), 171–175.

Bekes, K., & Steffen, R. (2016). *Das Würzburger MIH-Konzept: Teil 1. Der MIH-Treatment Need Index (MIH-TNI): Ein neuer Index zur Befunderhebung und Therapieplanung bei Patienten mit Molaren-Inzisiven-Hypomineralisation (MIH)*. *Oralprophylaxe & Kinderzahnheilkunde*, 38(4), 165–170.

- Bilgir, E., Bayrakdar, İ. Ş., Çelik, Ö., Orhan, K., Akkoca, F., Sağlam, H., Odabaş, A., Aslan, A. F., Ozcetin, C., Kılı, M., & Rozylo-Kalinowska, I. (2021). *An artificial intelligence approach to automatic tooth detection and numbering in panoramic radiographs*. BMC Medical Imaging, 21(1), 124.
- Çelik, B., & Çelik, M. E. (2022). *Automated detection of dental restorations using deep learning on panoramic radiographs*. Dento Maxillo Facial Radiology, 51(8), 20220244.
- Cha, J.-Y., Yoon, H.-I., Yeo, I.-S., Huh, K.-H., & Han, J.-S. (2021). *Peri-implant bone loss measurement using a region-based convolutional neural network on dental periapical radiographs*. Journal of Clinical Medicine, 10(5), 1009.
- Chen, X., Guo, J., Ye, J., Zhang, M., & Liang, Y. (2022). *Detection of proximal caries lesions on bitewing radiographs using deep learning method*. Caries Research.
- Crombie, F. A., Manton, D. J., Palamara, J. E. A., Zaluzniak, I., Cochrane, N. J., & Reynolds, E. C. (2013). *Characterisation of developmentally hypomineralised human enamel*. Journal of Dentistry, 41(7), 611–618.
- Crombie, F., Manton, D., & Kilpatrick, N. (2009). *Aetiology of molar-incisor hypomineralization: a critical review*. International Journal of Paediatric Dentistry, 19(2), 73–83.
- Ekert, T., Krois, J., Meinhold, L., Elhennawy, K., Emara, R., Golla, T., & Schwendicke, F. (2019). *Deep learning for the radiographic detection of apical lesions*. Journal of Endodontics, 45(7), 917-922.e5.
- Engels, P., Meyer, O., Schönewolf, J., Schlickerrieder, A., Hickel, R., Hesenius, M., Gruhn, V., & Kühnisch, J. (2022). *Automated detection of posterior restorations in permanent teeth using artificial intelligence on intraoral photographs*. Journal of Dentistry, 121(104124), 104124.
- Estai, M., Tennant, M., Gebauer, D., Brostek, A., Vignarajan, J., Mehdizadeh, M., & Saha, S. (2022). *Evaluation of a deep learning system for automatic detection of proximal surface dental caries on bitewing radiographs*. Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology, 134(2), 262–270.

- Fagrell, T. G., Dietz, W., Jälevik, B., & Norén, J. G. (2010). *Chemical, mechanical and morphological properties of hypomineralized enamel of permanent first molars*. *Acta Odontologica Scandinavica*, 68(4), 215–222.
- Farah, R. A., Monk, B. C., Swain, M. V., & Drummond, B. K. (2010). *Protein content of molar-incisor hypomineralisation enamel*. *Journal of Dentistry*, 38(7), 591–596.
- Jälevik, B., Odellius, H., Dietz, W., & Norén, J. (2001). *Secondary ion mass spectrometry and X-ray microanalysis of hypomineralized enamel in human permanent first molars*. *Archives of Oral Biology*, 46(3), 239–247.
- Jordan, R. A., Bodechtel, C., Hertrampf, K., Hoffmann, T., Kocher, T., Nitschke, I., Schiffner, U., Stark, H., Zimmer, S., Micheelis, W., & DMS V Surveillance Investigators' Group. (2014). *The Fifth German Oral Health Study (Fünfte Deutsche Mundgesundheitsstudie, DMS V) – rationale, design, and methods*. *BMC Oral Health*, 14(1), 161.
- Khanagar, S. B., Al-Ehaideb, A., Maganur, P. C., Vishwanathaiah, S., Patil, S., Baeshen, H. A., Sarode, S. C., & Bhandi, S. (2021). *Developments, application, and performance of artificial intelligence in dentistry - A systematic review*. *Journal of Dental Sciences*, 16(1), 508–522.
- Koch, G., Hallonsten, A. L., Ludvigsson, N., Hansson, B. O., Holst, A., & Ullbro, C. (1987). *Epidemiologic study of idiopathic enamel hypomineralization in permanent teeth of Swedish children*. *Community Dentistry and Oral Epidemiology*, 15(5), 279–285.
- Kosan, E., Krois, J., Wingefeld, K., Deuter, C. E., Gaudin, R., & Schwendicke, F. (2022). *Patients' perspectives on artificial intelligence in dentistry: A controlled study*. *Journal of Clinical Medicine*, 11(8), 2143.
- Krois, J., Ekert, T., Meinhold, L., Golla, T., Kharbot, B., Wittemeier, A., Dörfer, C., & Schwendicke, F. (2019). *Deep learning for the radiographic detection of periodontal bone loss*. *Scientific Reports*, 9(1), 8495.
- Kühnisch, J., Meyer, O., Hesenius, M., Hickel, R., & Gruhn, V. (2022). *Caries detection on intraoral images using artificial intelligence*. *Journal of Dental Research*, 101(2), 158–165.

- Kühnisch, Jan, Standl, M., Hickel, R., & Heinrich, J. (2021). *Molaren-Inzisiven-Hypomineralisation (MIH). Häufigkeit und mögliche Ursachen unter besonderer Berücksichtigung der Ergebnisse aus den Münchner Geburtskohorten GINIplus und LISA*. Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz, 64(8), 924–930.
- Lee, S., Oh, S.-I., Jo, J., Kang, S., Shin, Y., & Park, J.-W. (2021). *Deep learning for early dental caries detection in bitewing radiographs*. Scientific Reports, 11(1), 16807.
- Li, C.-W., Lin, S.-Y., Chou, H.-S., Chen, T.-Y., Chen, Y.-A., Liu, S.-Y., Liu, Y.-L., Chen, C.-A., Huang, Y.-C., Chen, S.-L., Mao, Y.-C., Abu, P. A. R., Chiang, W.-Y., & Lo, W.-S. (2021). *Detection of dental apical lesions using CNNs on periapical radiograph*. Sensors (Basel, Switzerland), 21(21), 7049.
- Lygidakis, N. A., Wong, F., Jälevik, B., Vierrou, A.-M., Alaluusua, S., & Espelid, I. (2010). *Best Clinical Practice Guidance for clinicians dealing with children presenting with Molar-Incisor-Hypomineralisation (MIH): An EAPD Policy Document*. European Archives of Paediatric Dentistry: Official Journal of the European Academy of Paediatric Dentistry, 11(2), 75–81.
- Mahoney, E. K., Rohanizadeh, R., Ismail, F. S. M., Kilpatrick, N. M., & Swain, M. V. (2004). *Mechanical properties and microstructure of hypomineralised enamel of permanent teeth*. Biomaterials, 25(20), 5091–5100.
- Mao, Y.-C., Chen, T.-Y., Chou, H.-S., Lin, S.-Y., Liu, S.-Y., Chen, Y.-A., Liu, Y.-L., Chen, C.-A., Huang, Y.-C., Chen, S.-L., Li, C.-W., Abu, P. A. R., & Chiang, W.-Y. (2021). *Caries and restoration detection using bitewing film based on transfer learning with CNNs*. Sensors (Basel, Switzerland), 21(13), 4613. <https://doi.org/10.3390/s21134613>
- Moran, M., Faria, M., Giraldo, G., Bastos, L., Oliveira, L., & Conci, A. (2021). *Classification of approximal caries in bitewing radiographs using convolutional neural networks*. Sensors (Basel, Switzerland), 21(15), 5192.

- Moutselos, K., Berdouses, E., Oulis, C., & Maglogiannis, I. (2019). *Recognizing occlusal caries in dental intraoral images using Deep Learning*. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2019, 1617–1620. <https://doi.org/10.1109/EMBC.2019.8856553>
- Park, E. Y., Cho, H., Kang, S., Jeong, S., & Kim, E.-K. (2022). *Caries detection with tooth surface segmentation on intraoral photographic images using deep learning*. BMC Oral Health, 22(1), 573. <https://doi.org/10.1186/s12903-022-02589-1>
- Schlickenrieder, A., Meyer, O., Schönewolf, J., Engels, P., Hickel, R., Gruhn, V., Hesenius, M., & Kühnisch, J. (2021). *Automatized detection and categorization of fissure sealants from intraoral digital photographs using artificial intelligence*. Diagnostics (Basel, Switzerland), 11(9).
- Schwendicke, F., Samek, W., & Krois, J. (2020). *Artificial intelligence in dentistry: Chances and challenges*. Journal of Dental Research, 99(7), 769–774.
- Schwendicke, Falk, Elhennawy, K., Reda, S., Bekes, K., Manton, D. J., & Krois, J. (2018). *Global burden of molar incisor hypomineralization*. Journal of Dentistry, 68, 10–18.
- Silva, M. J., Scurrah, K. J., Craig, J. M., Manton, D. J., & Kilpatrick, N. (2016). *Etiology of molar incisor hypomineralization - A systematic review*. Community Dentistry and Oral Epidemiology, 44(4), 342–353.
- Vinayahalingam, S., Kempers, S., Limon, L., Deibel, D., Maal, T., Hanisch, M., Bergé, S., & Xi, T. (2021). *Classification of caries in third molars on panoramic radiographs using deep learning*. Scientific Reports, 11(1), 12609.
- Weerheijm, K. L., Jälevik, B., & Alaluusua, S. (2001). *Molar-incisor hypomineralisation*. Caries Research, 35(5), 390–391.
- Yaren Tekin, B., Ozcan, C., Pekince, A., & Yasa, Y. (2022). *An enhanced tooth segmentation and numbering according to FDI notation in bitewing radiographs*. Computers in Biology and Medicine, 146(105547), 105547.
- You, W., Hao, A., Li, S., Wang, Y., & Xia, B. (2020). *Deep learning-based dental plaque detection on primary teeth: a comparison with clinical assessments*. BMC Oral Health, 20(1), 141.

You W. Z., Hao A. M., Li S., Zhang Z. Y., Li R. Z., Sun R. Q., Wang Y., & Xia B. (2021). *Deep learning-based dental plaque detection on permanent teeth and the influenced factors*. *Zhonghua kou qiang yi xue za zhi [Chinese journal of stomatology]*, 56(7), 665–671.

Danksagung

An dieser Stelle möchte ich allen beteiligten Personen meinen Dank aussprechen, die mich bei der Anfertigung meiner Dissertation unterstützt haben.

Zunächst gebührt mein besonderer Dank meinem Doktorvater, Herrn Prof. Dr. Kühnisch für die Betreuung dieser Arbeit und der mannigfachen Ideengebung, die mir einen kritischen Zugang zu dieser Thematik eröffnet hat. Ich hätte mir das besser nicht wünschen können.

Ebenso geht mein Dank an meine beiden Kommilitoninnen, Promotionskolleginnen und Freundinnen Paula Engels und Anne Schlickerrieder. Sie waren mir sowohl während der Studienzeit als auch im Rahmen der Promotion eine große fachliche und moralische Stütze.

Mein besonderer Dank aber gilt meiner Familie und meinem Partner, die mir meinen bisherigen Lebensweg ermöglichten und denen ich diese Arbeit widme.