ITUPublications

International Telecommunication Union
Standardization Sector

ITU-T Focus Group Deliverable

(09/2023)

Focus Group on Artificial Intelligence for Health (FG-AI4H)

DEL10.17 – FG-AI4H Topic Description Document for the Topic group on dental diagnostics and digital dentistry (TG-Dental)

ITU-T FG-AI4H Deliverable

DEL10.17 – FG-AI4H Topic Description Document for the Topic group on dental diagnostics and digital dentistry (TG-Dental)

Summary

This deliverable detailing the work of the Topic Group on Dental Diagnostics and Digital Dentistry (TG-Dental). It provides a comprehensive overview of the group's activities from 2019 to 2023, outlining the challenges and opportunities of AI in various dental specialties, including diagnostics, treatment planning, and digital dentistry. The text emphasizes the need for standardized benchmarking of AI systems to ensure their robustness and generalizability across diverse populations and clinical settings, highlighting the ethical considerations, such as data diversity and privacy, that are crucial for the responsible development and implementation of dental AI solutions.

Keywords

Artificial intelligence, checklist, deep learning, dental research, dentistry, guidelines, machine learning, teeth.

Note

This is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

Change Log

This document contains Version 1 of the Deliverable DEL10.17 on "*FG-AI4H Topic Description Document for the Topic group on dental diagnostics and digital dentistry (TG-Dental)*" approved on 15 September 2023 via the online approval process for the ITU-T Focus Group on AI for Health (FG-AI4H).

Editor:	Falk Schwendicke TG-Dental Charité – Universitätsmedizin Berlin, Germany	E-mail:	falk.schwendicke@charite.de
	Joachim Krois TG-Dental dentalXrai GmbH Berlin, Germany	E-mail:	joachim.krois@gmail.com
	Tarry Singh TG-Dental deepkapha.ai, Netherlands	E-mail:	tarry.singh@deepkapha.ai

Contributors:

	Falk Schwendicke Charité – Universitätsmedizin Berlin Berlin, Germany	E-mail:	falk.schwendicke@charite.de

Joachim Krois
dentalXrai GmbH
Berlin, Germany
E-mail: joachim.krois@gmail.com

Tarry Singh
deepkapha.ai
Netherlands
E-mail: tarry.singh@deepkapha.ai

Prof. Jae-Hong Lee
Wonkwang University College of Dentistry,
Daejeon, Korea
E-mail: ljaehong@gmail.com

Dr Akhilanand Chaurasia
King George Medical University,
Lucknow, India
E-mail: chaurasiaakhilanand49@gmail.com

Dr Robert André Gaudin
Charité – Universitätsmedizin Berlin
Berlin, Germany
E-mail: robert-andre.gaudin@charite.de

Sergio Uribe
Riga Stradins University
Riga, Latvia
Universidad Austral de Chile
Valdivia, Chile
E-mail: sergio.uribe@rsu.lv

Hossein Mohammad-Rahimi
Research Assistant, Computer Engineering
Department, Sharif University of Technology
Tehran, Islamic Republic of Iran
E-mail: ramtin.rhm@gmail.com

Janet Brinz
University Hospital LMU Munich
Munich, Germany
E-mail: janet.brinz@med.uni-muenchen.de

Anahita Haiat
University of Western Australia
Perth, Australia
E-mail: anahita.haiat@gmail.com

Gürkan Ünsal
Assistant Professor, Department of
Dentomaxillofacial Radiology, Faculty of
Dentistry, Near East University
Nicosia, Cyprus
E-mail: gurkanunsal@aol.com

Nielsen Santos Pereira
Private Sector
Brazil
E-mail: nielsen.spereira@gmail.com

Ulrike Kuchler
Associate Professor, Oral Surgery
Medical University of Vienna
Vienna, Austria
E-mail: ulrike.kuchler@meduniwien.ac.at

Shankeeth Vinayahalingam
Radboud University Medical Centre
Nijmegen, Kingdom of the Netherlands
E-mail: shankeeth.vinayahalingam@radboudumc.nl

Balazs Feher
Medical University of Vienna,
Vienna, Austria
Harvard School of Dental Medicine,
Cambridge, United States of America
E-mail: balazs.feher@meduniwien.ac.at, balazs_feher@hsdm.harvard.edu

Francesc Perez Pastor
Collaborating Prof. Master Orthodontics UB, IDISBA,
Palma, Spain
E-mail: nidrell@gmail.com

Lisa Schneider
Charité – Universitätsmedizin Berlin
Berlin, Germany
E-mail: schneider.lisa@charite.de

Chen Nadler
Oral-Maxillofacial Imaging Unit, Hadassah Medical Center, Faculty of Dental Medicine, Hebrew University of Jerusalem,
Jerusalem, Israel
E-mail: Nadler@hadassah.org.il

Sahel Hassanzadeh-Samani
Dentofacial Deformities Research Center, Shahid Beheshti University of Medical Sciences
Tehran, Islamic Republic of Iran
E-mail: sahelsamani@sbmu.ac.ir

Parisa Motie
Student Research Committee, school of dentistry, Shahid Beheshti University of Medical Sciences
Tehran, Islamic Republic of Iran
E-mail: pm.motie@gmail.com

Ragda Abdalla-Aslan
Oral Medicine Specialist
Department of Oral and Maxillofacial Surgery, Rambam Health Care Campus, Haifa
Department of Oral Medicine, Sedation and Maxillofacial Imaging, Hebrew University-Hadassah School of Dental MedicineJerusalem, Israel
E-mail: ragdaa@gmail.com

Teodora Karteva
Operative Dentistry & Endodontics
Department of Operative Dentistry & Endodontics, Medical University of Plovdiv
Plovdiv, Bulgaria
E-mail: teodora.karteva@mu-plovdiv.bg

Jelena Roganovic
Associate professor
Department of Pharmacology in Dentistry, University of Belgrade,
Belgrade, Serbia
E-mail: jelena.roganovic@stomf.bg.ac.rs

Kunaal Dhingra
Periodontics Division, Centre for Dental Education and Research, All India Institute of Medical Sciences
New Delhi, India
E-mail: kunaaldhingra@yahoo.co.in

Dr Prabhat Kumar Chaudhari
Division of Orthodontics and Dentofacial Deformities, Centre for Dental Education and Research (CDER), All India Institute of Medical Sciences
New Delhi, India.
E-mail: dr.prabhatkc@gmail.com

Olga Tryfonos
Periodontology Department
Academisch Centrum Tandheelkunde
Amsterdam (ACTA)
Amsterdam, Kingdom of the Netherlands
E-mail: tryfonosolga@yahoo.com

Prof. Marja Laine
Academisch Centrum Tandheelkunde
Amsterdam (ACTA)
Amsterdam, Kingdom of the Netherlands
Chair Periodontology Department
ACTA
The Netherlands
E-mail: m.laine@acta.nl

Rata Rokhshad
Boston medical Campus
University School of Medicine
Boston, United States of America
E-mail: Ratarokhshad@gmail.com

Fatemeh Sohrabniya
Research Committee, Shiraz University of Medical Science
Iran
E-mail: dr.sohrabniya@gmail.com

Zeynab Pirayesh
Department of Orthodontics, School of Dentistry, Zanjan University of Medical Sciences
Zanjan, Islamic Republic of Iran.
E-mail: zeinabpirayesh@gmail.com

Shada Alsalamah
Technical Officer, Digital Health and Innovations Department, World Health Organization
Geneva, Switzerland;
Assistant Professor, College of Computer and Information Sciences, King Saud University
Riyadh, Saudi Arabia
E-mail: alsalamahs@who.int

Sakher AlQahtani
Associate Professor, College of Dentistry, King Saud University
Riyadh, Saudi Arabia
Honorary Professor, University of Queensland
Brisbane, Australia
E-mail: asakher@ksu.edu.sa

Revan Birke Koca-Ünsal
Associate Professor, Department of Periodontology, Faculty of Dentistry, University of Kyrenia
Kyrenia, Cyprus
E-mail: revanbirke.koca@kyrenia.edu.tr

Lubaina T. Arsiwala
Associate researcher
Charité – Universitätsmedizin Berlin
Germany
E-mail: lubaina.arsiwala@charite.de

Parul Khare
Assistant Professor, Dept Oral & Maxillofacial Pathology and Oral Microbiology, SDS, Sharda University
Noida, India
Assistant Professor (adjunct) Saraswati Dental College
Lucknow, India
Assistant Professor (adjunct) Saveetha Dental College
Chennai, India
E-mail: parulsi02@gmail.com

Amit Punj
Associate Professor
MontefioreEinstein Medical Centre
New York, United States of America
E-mail: apunj@montefiore.org

Manal Hamdan
Assistant Professor of Oral and Maxillofacial Radiology
Department of General Dental Sciences
Marquette University School of Dentistry
Milwalkee, United States of America
E-mail: hamdanai4h@gmail.com

Zaid Badr
University of Nebraska Medical Center
Lincoln, United States of America
E-mail: zaidbadr@hotmail.com

Tamara Peric
Associate Professor
Clinic for Pediatric and Preventive Dentistry, School of Dental Medicine, University of Belgrade,
Belgrade, Serbia
E-mail: tamara.peric@stomf.bg.ac.rs

Dr Mihiri Silva
Senior Lecturer Paediatric Dentistry | Division Lead Cariology, Population Health & Oral Health, Melbourne Dental School, University of Melbourne,
Melbourne, Australia
E-mail: mihiri.silva@unimelb.edu.au

Consultant Paediatric Dentist & Clinician Science Fellow, Murdoch Childrens Research Institute, Royal Children's Hospital
Melbourne, Australia

Ms. Bree Jones
Lecturer and PhD candidate
Melbourne Dental School, University of Melbourne,
Victoria, Australia
E-mail: bree.jones@unimelb.edu.au

Prof. Miroslav Radenkovic
Department of Pharmacology, Clinical Pharmacology and Toxicology
Faculty of Medicine, University of Belgrade,
Belgrade, Serbia
E-mail: miroslav.radenkovic@med.bg.ac.rs

Martha Büttner Charité – Universitätsmedizin Berlin Oral Diagnostics, Digital Health & Health Services Research Berlin, Germany	E-mail: martha.buettner@charite.de
Mohammed Omar University of Iowa Iowa City, United States of America	E-mail: mohammed.omar.2000@hotmail.com
Gowri Sivaramakrishnan Primary Health Care Centers Bahrain	E-mail: gowri.sivaramakrishnan@gmail.com
Jaisri Thoppay President, Center for Integrative Oral Health Winter Park, Fl, United States of America	E-mail: jthoppay@inoralhealth.com drjthoppay@gmail.com
Saujanya Karki Postdoctoral Researcher, Research Unit of Population Health, University of Oulu Oulu, Finland	E-mail: saujanya.karki@oulu.fi
Tarja Tanner Associate Professor, Research Unit of Population Health, University of Oulu Oulu, Finland	E-mail: tarja.tanner@oulu.fi
Marja-Liisa Laitala Professor, Department of Cariology, Endodontics and Pediatric Dentistry, Research Unit of Population Health, University of Oulu Oulu, Finland	E-mail: marja-liisa.laitala@oulu.fi
Sanaa Chala Professor, Department of endodontics and restorative dentistry, Faculty of dental medicine, Mohammed V University Rabat, Morocco	E-mail: s.chala@um5r.ac.ma
Raphaël Richert Assistant, Department of Endodontics and Restorative Dentistry, Faculty of Dental Surgery, Claude Bernard University Lyon 1 Lyon, France	E-mail: raphael.richert@chu-lyon.fr
Pierre Lahoud Clinical Assistant | Ph. D Researcher Department of Periodontology & Oral Microbiology OMFS-IMPATH Research Group Department of Oral and MaxilloFacial Surgery KU Leuven Leuven, Belgium	E-mail: pierre.lahoud@outlook.com
Julien Issa PhD candidate, Department of Diagnostics Faculty of Dentsitry Poznań University of Medical Sciences, Poznań, Poland	E-mail: julien21issa@gmail.com

Nisha Manila
Assistant Professor, Oral and Maxillofacial Radiology, Department of Diagnostic Sciences
Louisiana State University, School of Dentistry
New Orleans, United States of America
E-mail: manila.nisha@gmail.com

Fahad Umer
Assistant Professor, Operative Dentistry and Endodontics
Dentistry, Department of Surgery
Aga Khan University Hospital
Karachi, Pakistan
E-mail: fahad.umer@aku.edu

Sharon Tan
Saw Swee Hock School of Public Health,
National University of Singapore,
Singapore
E-mail: tan.sharon@nus.edu.sg

Gauthier Dot
Assistant Professor of Orthodontics,
UFR Odontologie, Université Paris Cité
Paris, France
E-mail: gauthier.dot@aphp.fr

Sara Haghighat
Dentist/Researcher
Shiraz University of Medical Sciences
Shiraz, Islamic Republic of Iran
E-mail: Haghighatsd@gmail.com

Nora Saif
Assoc. Prof. Oral & Maxillofacial Radiology
Cairo University
Cairo, Egypt
E-mail: nora.taha@dentistry.cu.edu.eg

Ali Rahbar
Research Committee, Shahid Beheshti University of Medical Sciences
Tehran, Islamic Republic of Iran
E-mail: ali.rahbar.taramsari@gmail.com

Marwa Baraka
Lecturer in Pediatric Dentistry and Dental Public Health Department,
Faculty of Dentistry,
Alexandria University
Alexandria, Egypt
E-mail: marwa.baraka@alexu.edu.eg
marwa.baraka87@gmail.com
mbaraka@umich.edu

Aleksander Krasowski
Charite - Universitätsmedizin Berlin
Berlin, Germany
E-mail: aleksander.krasowski@charite.de

Nermin Morgan
Ass. Professor. Oral and maxillofacial radiology,
Mansoura University,
E-mail: nermin_morgan@yahoo.com

Mansoura, Egypt
KU Leuven
Leuven, Belgium

Sarina Azimian
Dentist, Graduate School of Dentistry, Shahid Beheshti University of Medical SciencesTehran, Islamic Republic of Iran
E-mail: azimiansarina@gmail.com

Lamya El Fadaly Lecturer at October University for Modern Science and Arts 6th of October City, Egypt	E-mail:	lzaki@msa.edu.eg
Johannes Tanne Regulatory Affairs and Dental AI, Consultant	E-mail:	tannejohannes@gmail.com
Yunpeng Li Senior Lecturer in Artificial Intelligence, Department of Computer Science, University of Surrey Guildford, Surrey, UK	E-mail:	yunpeng.li@surrey.ac.uk

Table of Contents

ITU-T FG-AI4H Deliverable

DEL10.17 – FG-AI4H Topic Description Document for the Topic group on dental diagnostics and digital dentistry (TG-Dental)

1 Introduction

The Topic Group Dental Diagnostics and Digital Dentistry (TG-Dental) focuses on diagnostics in dentistry, dental and oral medicine, and associated disciplines (oral and maxillofacial surgery, orthodontics, dental and oral traumatology). Specifically, it deals with the following (not exhaustive) ICD-10 code headings:

- K00: Disorders of tooth development and eruption
- K01: Embedded and impacted teeth
- K02: Dental caries
- K03: Other diseases of hard tissues of teeth
- K04: Diseases of the pulp and periapical tissues
- K05: Gingivitis and periodontal diseases
- K06: Other disorders of the gingiva and edentulous alveolar ridges
- K07: Dento-facial anomalies including malocclusion
- K08: Other disorders of the teeth and supporting structures
- K09: Cysts of oral origin, not elsewhere classified
- K10: Other diseases of jaws
- K11: Diseases of the salivary glands
- K12: Stomatitis and related lesions
- K13: Other diseases of the lip and oral mucosa
- K14: Diseases of the tongue
- J01: Acute sinusitis
- J32: Chronic sinusitis.

Diagnostics include the detection, assessment and prognosis of and on these conditions, as well as diagnostics on patients' risk to experience a condition or its progression. Diagnostics also includes the (image or non-image) assessment of anatomical structures or physiological functions for the purpose of detection of conditions and also treatment planning for various therapies (e.g., implantology).

Dental conditions, such as caries, periodontitis or tooth loss, are among the most prevalent diseases of humankind, affecting up to 98% of the population. Direct treatment costs due to dental diseases globally were estimated at US $298 billion annually, corresponding to an average of 4.6% of global health expenditure. The burden emanating from oral diseases is comparable to that from diabetes or cardiovascular diseases. The majority of the world's 1.6 million dentists are based in Europe and the Americas, such that 69% of the world's dentists serve 27% of the global population. Africa has only 1% of the global workforce. The overall workforce in dentistry exceeds 10 million worldwide. Diagnostics in dentistry largely relies on dentists diagnosing diseases via a combination of dental history taking, clinical investigation and imaging as well as, if required, further physical, (bio)chemical or microbiological assessments.

AI will help to (1) improve the accuracy of each of these individual tasks, (2) allow the integration of different data at a higher level effectiveness than an individual can achieve, (3) without ease also longitudinally assessing these data, compare them over time, and hence allow predictions, (4) reduce the reliance of diagnosis making on the dentist, expanding the scope of dental auxiliary staff, thereby increasing the access and efficiency of dental services, and (5) enable patients and healthy individuals to better participate in their dental health experience and management. AI will pave the way to a more personalized, precise, preventive and participatory dentistry for more people worldwide. It has the potential to aid in overcoming current ineffective, expensive care models.

Benchmarking is expected to yield more robust models and algorithms, with initially lower accuracy compared with current validation strategies (largely in-sample). Benchmarking is further expected to allow transparent comparisons of different models and algorithms. This topic description document specifies the standardized benchmarking for dental diagnostics and digital dentistry systems. It serves as deliverable No. 10.17 of the ITU/WHO Focus Group on AI for Health (FG-AI4H). This document further describes activities of the TG during the period since 2019 and its deliverables over this period.

2 About the FG-AI4H topic group on Dental Diagnostics and Digital Dentistry

The introduction describes the chances and challenges of AI in dental diagnostics, specifically in different subtopics, and highlights the potential of a standardized benchmarking of AI systems for Dental Diagnostics and Digital Dentistry to help to solve important health issues and provide decision-makers with the necessary insight to successfully address these challenges. To develop this benchmarking framework, FG-AI4H decided to hold the TG-Dental at their meeting G in New Delhi, on 13–15 November 2019.

FG-AI4H assigns a topic driver to each topic group (similar to a moderator) who coordinates the collaboration of all topic group members. During FG-AI4H Meeting G in New Delhi, Prof. Dr Falk Schwendicke and Dr Joachim Krois from Charité – Universitätsmedizin Berlin, Germany, were nominated as the topic drivers for the TG-Dental.

2.1 Documentation

This document is the TDD for the TG-Dental. It introduces the health topic including the AI task, outlines its relevance and the potential impact that AI and benchmarking will have on the health system and patient outcome, and provides an overview of the existing AI solutions for Dental Diagnostics and Digital Dentistry. It describes the existing approaches for assessing the quality of Dental Diagnostics and Digital Dentistry systems and provides the details that are likely relevant for setting up a new standardized benchmarking. Finally, it summarizes the results of the topic group's benchmarking initiative and its further activities. In addition, the TDD addresses ethical and regulatory aspects.[14] The TDD has been developed cooperatively by all members of the topic group over time and updated TDD iterations were presented at each FG-AI4H meeting.

The final version of this TDD is released as a deliverable "DEL 10.17 Dental Diagnostics and Digital Dentistry (TG-Dental)". The Topic Group was expected to submit input documents reflecting updates to the work on this deliverable (Table 1) to each FG-AI4H meeting.

Table 1 – Topic group output documents

Number	Title
FGAI4H-S-010-A01	Latest update of the Topic Description Document of the TG-Dental
FGAI4H-S-010-A02	Latest update of the Call for Topic Group Participation (CfTGP)
FGAI4H-S-010-A03	The presentation summarizing the latest update of the Topic Description Document of the TG-Dental

2.2 Status of this topic group

The following subsections describe the update of the collaboration within the TG-Dental for the official focus group meetings.

2.2.1 Status update for meeting H

After the establishment of the TG-Dental at meeting G the "Call for Topic Group Participation Document (CfTGP)" was drafted and the "Topic Description Document (TDD)" was updated with a focus on the sections 1. Introduction and 3. Methods.

- Introduction
  - Relevance
  - Impact
  - Existing work.
- Method
  - Anatomical structures
  - Pathologies
  - Data sets and format
  - AI output data structure
  - Metrics.

Further, other academic groups in dentistry that focus not only on machine learning and deep learning but also on ethics, public health and economics were contacted.

2.2.2 Status update for meeting I

The TDD abstract was updated and the TDD was reviewed to emphasize sections not yet populated with text. In addition, the list of contributors was expanded and Tarry Singh, from deepkapha.ai, Netherlands, as well as Prof Jae-Hong Lee from Wonkwang University, Korea, joined the TG-Dental. The CfTGP was updated accordingly. The contributors of the TG-Dental defined tools for communication and set up biweekly calls via the Zoom chat room of FGAI4H.

Prof. Schwendicke initiated consultations with key opinion leaders from SPIRIT/CONSORT-AI in order to prepare the dissemination in key dental journals. Further, the contributors of the TG-Dental started the drafting of a series of scientific articles on dental AI methods, opportunities and challenges in leading dental journals.

2.2.3 Status update for meeting J

Two new contributors joined the TG-Dental, Dr Robert Gaudin from Charite – Universitätsmedizin Berlin, Germany and Dr Akhilanand Chaurasia from King George Medical University, Lucknow, India. The TG-Dental contributors drafted a manuscript on *Artificial intelligence for dental image analysis: A Guide for Authors and Reviewers* and shared it for discussion with FGAI4H members.

The metrics description in the TDD was updated, where the concept of computing the tooth-based confusion matrix for dental image segmentation tasks was described. This allows the calculation of tooth level metrics for segmentation tasks, which in turn is useful to compare computer vision metrics to clinical data that is widely accessible in clinical studies and trials.

2.2.4 Status update for meeting K

For meeting K, the TDD structure was adapted as defined in document FGAI4H-J-105. The contributors of TG-Dental as well as Thomas Wiegand (Chair of the ITU/WHO FGAI4H) and Sergio Uribe (President of the e-Oral Health Network of the IADR) drafted the manuscript *Artificial intelligence in dental research: A Checklist for Authors and Reviewers.* This manuscript highlights the challenge that the number of studies employing artificial intelligence in dentistry is growing fast but that the majority of studies suffer from limitations in planning, conduct and reporting, resulting in low robustness and applicability. The manuscript provides a consented checklist for authors, reviewers and readers of AI studies in dental research. The initial draft of the checklist and an explanatory document were derived and discussed among the members of IADR's e-oral network and the ITU/WHO focus group "Artificial Intelligence for Health (AI4H)". Consensus on the checklist was built by 27 group members via an e-Delphi process.

In addition, members of TG-Dental contributed to the systematic review and meta-analysis of deep learning for cephalometric landmark detection. From 318 identified records, 19 studies (published 2017–2020), all employing convolutional neural networks, mainly on 2-D-lateral cephalometric radiographs (n = 15), using data from publicly available datasets (n = 12) and testing the detection of a mean of 30 (SD: 25; median: 19; min–max: 7–93) landmarks were included. The reference test was established by two (n = 11), 1 (n = 4), 3 (n = 3) or a set of experts (n = 1). The risk of bias was high and applicability concerns were detected for most studies. The mean (95% confidence interval) deviation of predicted from reference landmarks was 2.1 (95% CI: 1.2–3.0) mm, while the mean % of landmarks were detected with 0.05). The analysis concluded that deep learning (DL) shows relatively high accuracy for detecting landmarks on cephalometric imagery; however, only a few studies directly compared DL with clinicians or sufficiently reflected on the dimension of deviation (vertical versus sagittal). The overall body of evidence is consistent, but of limited robustness.

2.2.5 Status update for meeting L

Three new contributors joined the TG-Dental, Janet Brinz, final year dentistry student at Universität Regensburg, Germany; Sergio Uribe, PhD, DDS, a specialist in oral and maxillofacial Radiology, leading researcher, Bioinformatics Research Unit Riga Stradins University, Riga, Latvia and Associate Professor Universidad Austral de Chile, Valdivia, Chile; and Hossein Mohammad-Rahimi a general dentist (DDS) graduated from Shahid Beheshti University of Medical Sciences, Tehran, Islamic Republic of Iran, Research Assistant, Computer Engineering Department, Sharif University of Technology, Tehran, Islamic Republic of Iran and founder of DeepBites. The topic group members initiated a discussion on the structure of AI in dentistry and developed a figure that thematically structures AI subtopics along the clinical care pathway and subfields of dentistry (see Figure 3). In addition, a systematic review of AI applications in dentistry was initialized resulting in over eligible 190 studies, with nearly all of them focusing on either image analysis or shallow machine learning. Many of these studies showed a high risk of bias. Various accuracy estimates and more than 25 different use cases were found. A preliminary conclusion indicates limited consistency and comparability of the studies but reporting quality increased over time. Notably, outcomes beyond accuracy score are scarce and robustness and generalizability remain unclear.

2.2.6 Status update for meeting M

Seven new contributors joined the TG-Dental: Anahita Haiat, a final year dental student at the University of Western Australia, Perth, Australia holding a Master of Engineering and a Bachelor of Mathematics; Jaisri Thoppay, President of the Center for Integrative Oral Health Inc., Winter Park, Florida, USA; Nielsen Santos Pereira, a dentist and entrepreneur; Gürkan Ünsal, a dentomaxillofacial radiologist at Near East University, Faculty of Dentistry and member of the Artificial Intelligence in Medicine Research Group at Near East University, DESAM Institute; Ulrike Kuchler, an Associate Professor at the Department of Oral Surgery, Medical University of Vienna, Vienna, Austria; Balazs Feher, a junior scientist and lecturer in oral surgery at the Medical University of Vienna, Vienna, Austria; Shankeeth Vinayahalingam, a PhD student in oral and maxillofacial Surgery at Radboud University Medical Centre, Nijmegen, Kingdom of the Netherlands and a Master’s AI student at Radboud University, Nijmegen, Kingdom of the Netherlands. Figure 1 was revised and updated according to the new members joining the group. These new members were assigned to different subdomains according to their expertise and interest. In addition, an infrastructure team was added. In this team, members that are interested in technical aspects of the collaboration in particular are included .

Section 3 and in particular the subtopic descriptions have been significantly advanced.

The TG-Dental successfully published the FGAI4H output document "Artificial intelligence in dental research: A checklist for authors and reviewers (FGAI4H-M-004-A01)". The checklist for dental AI studies was produced as a collaboration of experts from the International Association for Dental Research (IADR), the E-oral Health Network and the ITU/WHO Focus Group on AI for Health. The TG-Dental further conceptualized its first multicentre benchmark study. The goal was to include approximately 10 centres/clinics/practices from all over the world. It was expected that differences in cohort characteristics (age and gender), dental status (e.g., DMFT) and technical conditions (X-ray machine, image characteristics, etc.), among others, would be present. In a collaborative effort, different DL models should be trained on the data of different centres (or different subsets, etc.) and predicted on hold-out test sets from the centre the model was trained on as well as on centres representing other populations. This benchmark study was accepted as one more use case for the FG-AI4H Trial Audits 2.0 project. (see Section 3.B under Trial Audit 2.0 Playbook).

2.2.7 Status update for meeting N

Nineteen new contributors joined the TG-Dental:

- Chen Nadler from the Oral-Maxillofacial Imaging Unit, Hadassah Medical Center, Faculty of Dental Medicine, Hebrew University of Jerusalem, Israel
- Sahel Hassanzadeh-Samani from the Dentofacial Deformities Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Islamic Republic of Iran
- Parisa Motie, Student Research Committee, School of Dentistry, Shahid Beheshti University of Medical Sciences, Tehran, Islamic Republic of Iran
- Ragda Abdalla-Aslan, Oral Medicine Specialist, Department of Oral and Maxillofacial Surgery, Rambam Health Care Campus, Haifa and Department of Oral Medicine, Sedation and Maxillofacial Imaging, Hebrew University-Hadassah School of Dental Medicine, Jerusalem, Israel
- Teodora Karteva, Operative Dentistry & Endodontics, Department of Operative Dentistry and Endodontics, Medical University of Plovdiv, Bulgaria• Jelena Roganovic, Associate Professor, Department of Pharmacology in Dentistry, University of Belgrade, Serbia
- Kunaal Dhingra, Periodontics Division, Centre for Dental Education and Research, All India Institute of Medical Sciences, New Delhi, India
- Dr Prabhat Kumar Chaudhari, Division of Orthodontics and Dentofacial Deformities, Centre for Dental Education and Research (CDER), All India Institute of Medical Sciences (AIIMS), New Delhi, India
- Olga Tryfonos, Periodontology Department, Academisch Centrum Tandheelkunde Amsterdam (ACTA), Amsterdam, Kingdom of the Netherlands
- Prof. Marja Laine, Chair Periodontology Department, Academisch Centrum Tandheelkunde Amsterdam (ACTA), Amsterdam, Kingdom of the Netherlands
- Rata Rokhshad, Dental Material Center, Azad Tehran University, Islamic Republic of Iran
- Fatemeh Sohrabniya, Research Committee, Shiraz University of Medical Science, Islamic Republic of Iran
- Zeynab Pirayesh, Department of Orthodontics, School of Dentistry, Zanjan University of Medical Sciences, Zanjan, Islamic Republic of Iran
- Shada Alsalamah, Technical Officer, Digital Health and Innovations Department, World Health Organization, Geneva, Switzerland and Assistant Professor, College of Computer and Information Sciences, King Saud University, Saudi Arabia, Riyadh
- Sakher AlQahtani, Associate Professor, College of Dentistry, King Saud University and Honorary Professor, University of Queensland
- Revan Birke Koca-Ünsal, Assistant Professor, Department of Periodontology, Faculty of Dentistry, University of Kyrenia, Kyrenia, Cyprus
- Lubaina T. Arsiwala, Associate researcher, Charité – Universitätsmedizin Berlin, Germany

- Francesc Perez Pastor, Collaborating Prof. Master Orthodontics UB, IDISBA Palma, Mallorca, Spain
- Lisa Schneider, Charité – Universitätsmedizin Berlin, Germany.

A subtopic group for **Ethics** was established under the leadership of Rata Rokhshad. As AI applications in the dental field raise a wide range of potential ethical concerns and due to the limitations of dental publications about ethics, the main goal of the group is awareness. Further goals are acknowledging and applying privacy and data protection, the communication between the patient and the clinician, data diversity and protecting autonomy. The guidelines for the diversity and quality of data are going to be applied to the current projects. Currently, the group is working on ethical issues in the field of aesthetics.

The subtopic group **Periodontal** updated the TDD. The subtopic group is continuously updating and searching for periodontal-related deep learning-based solutions. In addition, a benchmarking was explored about AI for the identification and classification of dental implant systems.

The **Surgical** subtopic group expanded from 4 to 12 members in total. Multiple research collaborations within the subtopic were initialized. The strategy of this subtopic group is to review the current literature on AI in oral surgery and implantology and to provide an update regarding the current status of the field. The aim is to update the clinical knowledge of AI in this field and to evaluate the protocols in diagnostics based on AI. Besides this screening, the group has started with collaborative research projects collecting datasets for diagnostic reasons from different centres to test and compare available AI models as well as train new ones. The application of AI in oral surgery diagnostics (e.g., cystic lesions and tumours) is still limited to research and needs further development and evaluation before clinical utilization. The subtopic group **Endodontics** was established under the leadership of Teodora Karteva. Endodontics is the science of retaining a natural tooth using root-canal treatment. Machine learning algorithms can improve the diagnosis, treatment, and prognosis of pulpal and periapical diseases as they assist clinicians in their exploration of the winding pathways of the pulp.

2.2.8 Status update for meeting O

Nine new collaborators joined the TG-Dental:

- Parul Khare, GP, Alpha Dental Clinic, Shanghai, China, Associate Professor (adjunct) Saraswati dental College Lucknow, India, Associate Professor (adjunct) Saveetha Dental College, Chennai, India
- Amit Punj, Associate Professor, Montefiore- Einstein, New York, USA
- Manal Hamdan, Oral and Maxillofacial Radiologist, Lincoln, NE, USA
- Zaid Badr, Assistant Professor, University of Nebraska Medical Center, Lincoln, NE, USA
- Tamara Peric, Associate Professor, Clinic for Pediatric and Preventive Dentistry, School of Dental Medicine, University of Belgrade, Serbia
- Dr Mihiri Silva, Senior Lecturer Paediatric Dentistry | Division Lead Cariology, Population Health and Oral Health, Melbourne Dental School, University of Melbourne, Australia, Consultant Paediatric Dentist and Clinician Science Fellow, Murdoch Childrens Research Institute, Royal Children’s Hospital Melbourne
- Ms. Bree Jones, Lecturer and PhD candidate, Melbourne Dental School, University of Melbourne, Victoria, Australia
- Prof. Miroslav Radenkovic, Department of Pharmacology, Clinical, Pharmacology and Toxicology, Faculty of Medicine, University of Belgrade, Serbia
- Martha Duchrau, Charité – Universitätsmedizin Berlin, Oral Diagnostics, Digital Health & Health Services Research Germany.

The subtopic group **Oral and Maxillofacial Oncology** was established under the leadership of Dr Parul Khare Sinha. The first initiative was to obtain a diverse dataset of oral histopathological slides in order to classify benign, dysplastic and carcinomatous pathologies. Therefore, 4000 slides of benign, dysplastic and squamous cell carcinoma were to be collected from 10 different centres worldwide. The subtopic group **Cariology** worked on the development of an annotated caries lesion dataset to independently evaluate different machine learning models for radiographic caries detection and the evaluation of the quality of annotations on dental radiographic datasets. Further, the group worked on knowledge dissemination and published two manuscripts and one white paper in conjunction with the FDI World Dental Federation to inform and provide recommendations on the use of AI in dentistry.

The group on **Ethics** prepared a checklist for raising awareness and overcoming ethical issues in dental AI research and applications. Relevant terms and concepts such as prudence, equity, autonomy, privacy, intimacy, responsibility, democratic participation, solidarity, data diversity, data protection, well-being, development and patient–clinician relationship, among others, were identified. The efforts concerning benchmarking continued. During the ongoing Magnitude of Clinical Benefit Scale (**MCBS**) more data was collected and an analytical team under the leadership of Lisa Schneider was established to start with the first experiments and model development.

2.2.9 Status update for meeting P

Nine new collaborators joined the TG-Dental:

- Mohammed Omar, University of Iowa, Iowa City, IA, USA
- Gowri Sivaramakrishnan, Dental Postgraduate Training Department, Ministry of Health, Bahrain
- Saujanya Karki, Postdoctoral Researcher, Research Unit of Population Health, University of Oulu
- Tarja Tanner, Associate Professor, Research Unit of Population Health, University of Oulu
- Marja-Liisa Laitala, Professor, Department of Cariology, Endodontics and Pediatric Dentistry, Research Unit of Population Health, University of Oulu
- Johannes Tanne, Consultant for Regulatory Affairs and Dental AI

In this period the Group mainly worked on the preparation and organization of the first TG-Dental Symposium, which took place on 19 September 2022 (Figure 1).

Figure 1 – Flyer for the TG-Dental Symposium

2.2.10 Status update for meeting Q

Seven new collaborators joined the TG-Dental:

- Sanaa Chala, Professor, Department of endodontics and restorative dentistry, Faculty of dental medicine, Mohammed V University in Rabat, Morocco
- Raphaël Richert, Assistant, Department of Endodontics and Restorative Dentistry, Faculty of Dental Surgery, Claude Bernard University Lyon 1, Lyon, France
- Pierre Lahoud, Clinical Assistant and Ph. D Researcher, Department of Periodontology & Oral Microbiology, OMFS-IMPATH Research Group, Department of Oral and MaxilloFacial Surgery, KU Leuven, Leuven, Belgium
- Julien Issa, PhD candidate, Department of Diagnostics, Faculty of Dentistry, Poznań University of Medical Sciences, Poznań, Poland
- Nisha Manila, Assistant Professor, Oral and Maxillofacial Radiology, Department of Diagnostic Sciences, Louisiana State University, School of Dentistry, New Orleans, LA, USA
- Fahad Umer, Assistant Professor, Operative Dentistry and Endodontics, Dentistry, Department of Surgery, Aga Khan University Hospital, Karachi, Pakistan
- Yunpeng Li, Senior Lecturer in Artificial Intelligence, Department of Computer Science, University of Surrey, Guildford, Surrey, UK.

The TG started with a dataset review, wherein multiple databases (including PubMed, GitHub, and IEEE) were reviewed to collect dental datasets that could be used for AI applications. The protocol is available at https://osf.io/mf897/. Further, the paper "Artificial intelligence for oral and dental health care: Core education curriculum" was published in the Journal of Dentistry, on behalf of the FG-AI4H (https://www.sciencedirect.com/science/article/abs/pii/S0300571222004158). The TG aimed to define a core curriculum for both undergraduate and postgraduate education, establishing a minimum set of outcomes learners should achieve when taught about oral and dental AI. Therefore, existing curricula and other documents focusing on the literacy of medical professionals around AI were screened and relevant items were extracted. Items were scoped and adapted using expert interviews with members of the IADR's e-oral health and education group and the ITU/WHO's Focus Group AI for Health. Learning outcome levels were defined and each item was assigned to a level. Items were systematized into domains and a curricular structure was defined. The resulting curriculum was consented using an online Delphi process. Finally, four domains of learning outcomes emerged, with most outcomes being on the "knowledge" level:

1 Basic definitions and terms, the reasoning behind AI and the principle of machine learning, the idea of training, validating and testing models, the definition of reference tests, the contrast between dynamic and static AI, and the problem of AI being a black box and requiring explainability should be known.

2 Use cases, the required types of AI to address them, and the typical set-up of AI software for dental purposes should be taught.

3 Evaluation metrics, their interpretation, the relevant impact of AI on patient or societal health outcomes and associated examples should be considered.

4 Issues around generalizability and representativeness, explainability, autonomy and accountability and the need for governance should be highlighted.

In addition, a study on the feasibility of federated learning in dentistry was submitted for expert review. In this study, the ML-task of tooth segmentation/classification on dental radiographs was used to study the differences between local, central and federated learning.

2.2.11 Status update for meeting R

Three new collaborators joined the TG-Dental:

- Sharon Tan from Saw Swee Hock School of Public Health at the National University of Singapore, Singapore
- Gauthier Dot, Assistant Professor of Orthodontics, UFR Odontologie, Université Paris Cité, Paris France
- Sara Haghighat, dentist and researcher at Shiraz University of Medical Sciences, Islamic Republic of Iran.

The TG continued the annotation process and conducted a annotation, repeated five times, of periapical radiolucencies on panoramic radiographs. Further, the TG prepared and organized the second TG-Dental Symposium, which took place on 21 March 2023 (Figure 2).

Figure 2 – Flyer for the second TG-dental symposium

2.2.12 Status update for meeting S

New collaborators joined the TG-Dental:

- Nermin Morgan, Ass. Professor. Oral and maxillofacial radiology, Mansoura University, Egypt, Collaborative researcher. OMFS-IMPATH group, KU Leuven, Belgium
- Marwa Baraka, Lecturer in Pediatric Dentistry and Dental Public Health Department, Faculty of Dentistry, Alexandria University, Egypt
- Nora Saif, Nora Saif, Assoc. Prof. Oral and Maxillofacial Radiology, Cairo University, Egypt
- Ali Rahbar, Research Committee, Shahid Beheshti University of Medical Sciences, Tehran, Islamic Republic of Iran
- Aleksander Krasowski, Charite - Universitätsmedizin Berlin, Germany.

A current analysis applied federated learning (FL) on a dataset of 4,177 panoramic radiographs collected by the TG-Dental covering nine different centres (n = 143 to n = 1881 per centre) across the globe and used FL to train a machine learning model for tooth segmentation. The paper was one of the first to apply FL for dentistry and showed that FL is a useful alternative to train performant and, more importantly, generalizable deep learning models in dentistry, where data protection barriers are high [1].

The TG further prepared and organized the third TG-Dental Symposium for 5 July 2023.

2.3 Topic group participation

The participation in both, the Focus Group on AI for Health and in a TG is generally open to anyone. Each topic group also has a corresponding subpage on the ITU collaboration site. For participation in this topic group, interested parties can also join the regular online meetings.

All relevant administrative information about FG-AI4H — such as upcoming meetings or document deadlines — will be announced via the general FG-AI4H mailing list fgai4h@lists.itu.int.

All TG members should subscribe to this mailing list as part of the registration process for their ITU user account by following the instructions in the 'Call for Topic Group participation' and this link: https://itu.int/go/fgai4h/join

Regular FG-AI4H workshops and meetings are held about every two months at changing locations around the globe or remotely. More information can be found on the official FG-AI4H website: https://itu.int/go/fgai4h

3 Topic description

The topic group is structured in a range of subtopic groups led by principal investigators. Moreover, there is a cross-sectional working group on ethics and one on infrastructure (Figure 3).

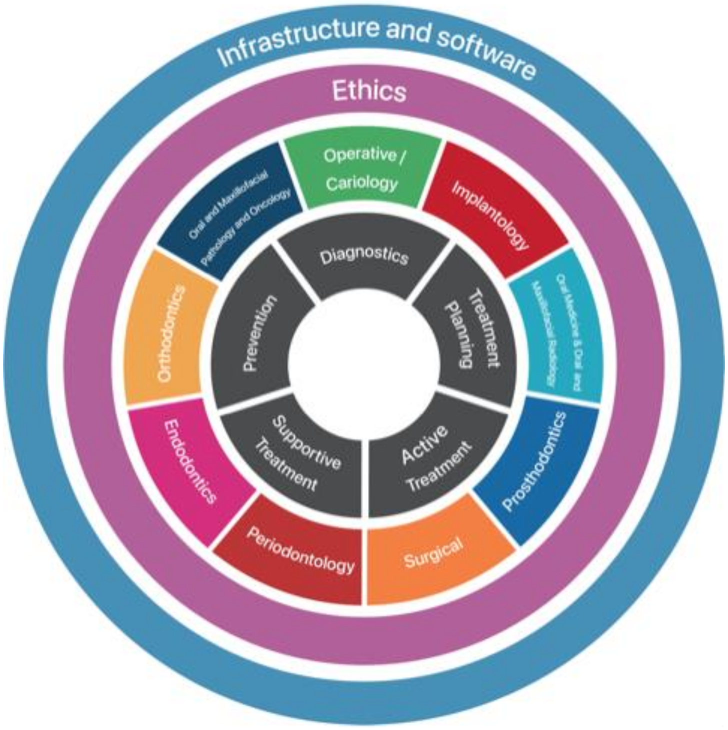

Figure 3 – Thematic structure of subtopics along the clinical care pathway and along subfields of dentistry

3.1 Operative and cariology

3.1.1 Diagnostic and treatment pathways

Dental caries is the most prevalent and ubiquitous non-communicable disease affecting humankind today. It was first understood to be an infectious disease, requiring the removal of all plaque (biofilm) from the teeth or from affected carious hard tissues (specific plaque hypothesis). This concept, while always debated (i.e., each biofilm being cariogenic under certain conditions; non-specific plaque hypothesis), was later modified, suggesting that the mere presence of biofilm is not sufficient for the pathogenesis of caries, but that an overlapping interaction between the host/teeth, substrate and microbiota is needed. Despite being 'infected' (or rather, contaminated) with cariogenic bacteria, a cavitated carious lesion will not develop without a cariogenic diet. Marsh introduced the ecological plaque theory. The microbial composition of the biofilm is stable unless 'environmental perturbations' occur that can affect microbial homeostasis, leading to dysbiosis. For caries, diet (mainly free sugars), oral hygiene and salivary factors are the contributing drivers of dysbiosis, leading to a shift in the microbiota towards acidogenic and aciduric microorganisms. Both hereditary and environmental factors influence the initial (naïve) dental biofilm, but, as time goes on, the types and proportions of micro-organisms acquired early in life are modified by environmental influences. Currently, the extended ecological plaque hypothesis is accepted to explain the pathogenesis of caries.

This pathogenesis of dental caries involves organic acids, the by-product of microbial metabolism of dietary free sugars. As the pH of the biofilm decreases, it reaches a point where the biofilm fluid at the surface of the tooth is under-saturated concerning tooth minerals and dissolution occurs to maintain equilibrium. Initially, the dissolution occurs at the tooth's surface, but, if conditions persist and the lesion becomes more extensive, minerals from deeper in the enamel (and subsequently dentine) will be lost. While environmental factors mainly explain the pathogenesis of dental caries, it also has a genetic component, with influencing factors including enamel quality and quantity, immune response, dietary preferences and salivary characteristics. In summary, dental caries is a disease characterized by a demineralization process of the dental hard tissues, caused by frequent free sugars exposure to the dental biofilm, shifting the ecological balance towards a cariogenic dysbiosis. For dentin and root caries, cleavage of collagen by bacterial or dentinal enzymes follows early mineral loss and contributes to hard tissue loss.

The former "traditional" management of the caries process and carious lesions was influenced by an understanding that caries was a purely infectious disease and could be managed invasively/restoratively by removing all demineralized and 'contaminated' tissue. This was grounded in (1) A lack of understanding that the caries process and carious lesions are separate, but related; (2) the incorrect understanding that once a lesion had established and the tooth was "infected", eradication of microbiota was needed; (3) the erroneous concept that lesion progression was inevitable; and, (4) the fact that the majority of carious lesions dentists encountered in the past were truly "decayed", i.e., cavitated dentine lesions. Therefore, the professional education of dental *surgeons* concentrated on mechanistic surgical procedures instead of that of dental *physicians* who manage the disease, and remuneration systems incentivized such invasive/restorative therapies.

Nowadays, building on evidence accrued over several decades, it is clear that (1) The caries process can be controlled by modifying the patient's caries risk/susceptibility, depending on his/her adherence to behavioural modifications and not only by intervening operatively on carious lesions, yet success/understanding of behavioural interventions on caries control has been limited; (2) The caries process and carious lesions can be managed without removing microorganisms, but by rebalancing the dysbiosis within the tooth surface biofilm and arresting those within the depths of the tissues; (3) Active (progressing) lesions can be inactivated, and progression is usually slow for most lesions; (4) In many high-income countries the spectrum of carious lesions has been and is shifting, especially in younger people, as there are now more non-cavitated lesions being detected.

Hence, the conventional restorative/invasive approach towards managing the caries process and carious lesions is not grounded in current understanding of the disease and it is also not appropriate for managing the broad spectrum of lesions found in many individuals (from very early to large cavitated). It should also be considered that subsequent interventions on restorations are often necessary. This is classically known as the "restorative death spiral". Given these alternative arguments, there is consensus that invasive/restorative interventions alone are not beneficial for managing the caries process and lesions in all situations. Instead, invasive/restorative interventions represent a late stage in the management puzzle, repairing the gross tissue damage and restoring form, function, aesthetics and cleansability, thereby allowing to control the risk of future loss of function. Invasive strategies may also be used to approach acute caries lesions. Invasive/restorative interventions are an important and relevant tool, but they should be complemented by other (non- or microinvasive) management strategies.

Three levels of invasiveness to classify intervention strategies for treating existing carious lesions can be distinguished. They are based on the degree of tissue removal associated with each strategy: *Non-invasive* strategies do not remove dental hard tissue and involve, for example, fluorides and other chemical strategies for controlling mineral balance, biofilm control measures and dietary control. *Micro-invasive* strategies remove the dental hard tissue surface at the micrometre level, usually during an etching step, such as sealing or infiltration techniques. *Invasive* strategies remove gross dental hard tissue, such as using hand excavators, rotary instruments or other devices. In most cases, this process is associated with the placement of restorations.

Caries detection is needed for each pathway of management, regardless of whether it is non-, micro- or invasive. Moreover, agreed principles for when to use different intervention strategies have been outlined: Lesion activity, cavitation and cleansability are the main factors to be considered to determine intervention thresholds. Inactive lesions do not usually require any treatment (in some cases, restorations might be placed for form, function or aesthetics); active lesions do. Non-cavitated carious lesions should be managed non- or micro-invasively, as should cavitated carious lesions, which are cleansable. Cavitated carious lesions which are not cleansable usually require invasive/restorative management, also to restore form, function and aesthetics. In specific circumstances, mixed interventions may be applicable. On occlusal surfaces, cavitated lesions confined to enamel and non-cavitated lesions radiographically extending deep into dentine (middle or inner dentine third, D2/3) may be exceptions to that rule. On proximal surfaces, cavitation is usually hard to assess tactile-visually. Hence, radiographic lesion depth is used to determine the likelihood of cavitation. Lesions extending radiographically into the middle or inner third of the dentine (D2/3) can be assumed to be cavitated, while those restricted to the enamel (external or middle enamel third, E1/2) are usually not cavitated. For lesions radiographically extending into the outer third of the dentine (D1), cavitation status is unclear. These lesions should be managed as if they were non-cavitated unless otherwise indicated. Individual decisions and clinical judgment should consider factors modifying the described intervention thresholds. Comprehensive diagnostics are the basis for systematic decision-making on when to intervene in the caries process and on existing carious lesions. Patients should be comprehensively informed about treatment options and should provide informed consent accordingly.

3.1.2 Definition of the AI task

A wide range of AI tasks along these pathways of risk assessment, detection, diagnostics, treatment planning, conduct and re-evaluation are conceivable. Currently, these entail

Caries risk assessment:

- Individual caries risk factors: Is X a risk factor for future caries experience?
- Risk factors models: (Factor Z x.3) + (Factor Y x.3) + (Factor Z x.3) =caries risk 90%
- Predictors of treatment effect: Patients with the X gene undergoing Z treatment have better outcomes than patients without the X gene.
- Caries detection and screening on
- Photos or videos (real-time or asynchronous)
- Bitewing radiographs
- Periapical radiographs
- Panoramic radiographs
- Cone beam CT (CBCT)
- Less frequently used imagery such as OCT, NILT
- Caries lesion depth assessment and activity/progression prediction

- Treatment decision support (decision between management levels and options)
- Personalized patient advice or mentoring.

3.1.3 Current gold standard

It is important to differentiate between a diagnosis validated against a system capable of assessing the actual health status, in what is known as ground truth (i.e., an independent diagnostic confirmation) and a system validated against a consensus or gold standard, which represents an approximation to the true diagnosis. For the imaging diagnosis of caries, no test currently provides a ground truth against which to evaluate the performance of a new test. In vitro, the gold standard is established using destructive preparation and analytic techniques such as histology, microradiography or (non-destructively) micro-CT. For clinically gathered data, this is not available and clinical assessment or, for imagery, expert opinion is needed instead; if possible, triangulation is used.

If only imagery is available, one or a range of experts annotate each image independently, providing a classification task, i.e., instance label (e.g., caries present), a detection task (e.g., a bounding box around an area) or a segmentation task (e.g., a pixel-wise area) of interest. Unifying such fuzzy labels is now a matter of choice: (1) For instance-based annotations, majority voting schemes are often applied (e.g., three or more out of five experts need to agree that pathology is present. Otherwise it is assumed not to be present; in some cases, images where the uncertainty is especially high are discarded). (2) It is also possible to combine these annotations with a "master" review process where one or more master experts review the labels created by a range of other human annotators to correct them. Hence, the annotations can be unionized based on a state of higher information (as the master reviewer can see all annotations) and on the underlying assumption that the master's experience leads to high-quality final labels. (3) Also, it is possible to discuss uncertain cases by an expert board, which would not be able to jointly assess all images from a dataset given time and resource constraints but can assess the small minority of unclear ("edge") cases to come to a consensus there.

3.1.4 Relevance and impact of an AI solution

Caries risk assessment and lesion detection and activity assessment suffer from poor accuracy, low reliability, grounded in either a limited set of predictor variables with poor predictive performance being available, existing prediction models being insufficiently calibrated to individual settings, populations or people, or experts' limited diagnostics accuracy in routine settings being further modified by their experience.

AI support for these tasks promises to increase the accuracy and reliability of assessment, prediction and detection/localization or achieve an expert-level status for a range of tasks. It may further reduce the effort for doing so (efficiency gain), increase safety and contain costs. Notably, as AI will remain a support tool for a long time, the final impact (benefit/harm) will largely depend on the user decision after employing AI support.

3.1.5 Existing AI solutions

In research, a recent review identified 48 studies on caries detection using AI in PubMed [2]. AI for caries lesion localization or progression prediction has not been found; the same applies to decision support. The detection studies mainly employed radiographs; the minority used photographs or other imagery. Most reports focus on detection (caries yes/no) and classification (healthy/enamel caries/dentin caries). The main algorithm used was CNN. In 10 studies, general dentists were used as annotators, in two studies, radiologists were used, in two other studies non-radiologist specialists were used, and in five reports it is not clear who annotated the training, test and internal validation datasets. The main strategy used to train and test the algorithms was to divide the dataset into a training and test set (28 studies), and in two studies an internal validation was considered. No study reported an independent external validation.

Commercially, to the knowledge of the TG, only a few applications are available for caries detection. In October 2022 there were three applications confirmed to have received certifications from regulatory bodies for the caries detection task: Overjet (Boston, USA) and Second Opinion (Pearl, CA, USA) by FDA, and dentalXrai Pro (dentalXrai GmbH, Berlin) by Conformité Européenne (CE). No solutions for other use cases are known to the TG at this point.

Given that dental caries is primarily influenced by individual patient habits and behaviours, the use of a personalized approach is crucial in empowering patients to effectively control risk factors such as diet and oral hygiene. As such, any available AI technology capable of providing tailored recommendations and mentoring could hold significant potential in caries control. One such technology that has garnered attention is conversational chatbots, which have the ability to offer personalized approaches to patient care. However, the extent of their value in this context remains limited and not yet well-documented. Further research and development of personalized AI technologies are needed to effectively combat the prevalence of dental caries.

3.2 Prosthodontics

Prosthodontics involves the diagnosis, treatment planning, and rehabilitation of clinical conditions related to missing or deficient teeth or maxillofacial tissue by using prosthetic substitutes to maintain the oral health, function, comfort and appearance of patients with these conditions. The scope of prosthodontic treatment can range from a single indirect restoration to full mouth rehabilitation using fixed or removable, tooth or implant-supported prosthesis, and guidelines such as the Prosthodontic Diagnostic Index (PDI) have been established to assess the complexity of clinical cases and aid in formulating appropriate treatment plans.

Improvements in the oral health of populations over the last century have led to a significant reduction in edentulism rates and an increasing number of people retaining most of their teeth throughout their life.

3.2.1 Definition of the AI task

AI can be envisaged to enhance the processes of prosthodontic diagnostics, treatment planning and procedures, broadly categorized as follows:

- Diagnosis and treatment-planning support
- Prediction of treatment outcome and prognosis of teeth and prostheses
- Support systems for tooth preparation
- Fabrication of prostheses.

3.2.2 Current gold standard

At present, planning for prosthodontic treatment draws on the clinical judgment and experience of the dental clinician. However, the treatment-planning process in this field of dentistry is highly complex, and there is no gold standard as there is often disagreement on the majority of clinical cases, even among specialists in the field. The introduction of the digital workflow has changed the face of prosthodontics and fabrication of most types of prostheses can now be done via the CAD/CAM digital workflow. However, this technology is not available at many dental practices, and improvements still need to be made to digital to increase the precision of conducted treatments. Preparation of teeth is done by dentists; however, even after years of training and clinical practice, it remains challenging and error-prone, and the resulting imprecision can compromise the final outcome.

3.2.3 Relevance and impact of an AI solution

Errors during diagnosis, treatment planning and execution of treatment can lead to poor outcomes such as occlusal imbalances, damage to the stomatognathic system, early fracture or failure of restorations, and suboptimal aesthetics. The long-term success of prosthodontic treatment relies on the good prognosis of abutment teeth and implants and both tooth level prognostic factors of periodontal, endodontic, reconstructive and patient-specific factors. Additionally, prosthodontics treatment is costly and not easily available to the entirety of the population. By implementing AI through the various stages of prosthodontic treatment, it can be possible to improve treatment outcomes and reduce inequalities in access to prosthodontic treatment.

3.2.4 Existing AI solutions

AI research in prosthodontics is still in its early stages and has been far less explored compared with other fields of dentistry due to the complex diagnostics and treatment required in prosthodontics. A recent literature review on AI in prosthodontics identified the following studies. To the best of the TG's knowledge, these applications have not yet undergone regulatory approval and are not available for commercial use.

I. Prosthesis shade selection of maxillofacial prosthesis. The traditional approach to colour matching for fabrication of maxillofacial prosthesis is a very difficult and time-consuming task and requires the presence of the patient to validate the colour match. Hence, this study used a CNN to indicate the compounding amount of different pigments to reconstruct the colour of skin tone [3].

II. Decision support for extraction of compromised teeth. In this study electronic dental records were used as data to train five different machine learning algorithms to solve a classification problem of extraction versus non-extraction. Out of the five algorithms, extreme gradient boost (XGBoost) performed best and achieved high accuracy in correct prediction of decision-making for tooth extraction [4].

III. Classification of partially edentulous arches. In this study, a CNN was employed for the classification of dental arches into four arch types of edentulous, intact dentition, arches with posterior tooth loss, and arches with bounded edentulous areas [5].

IV. Predicting the debonding of CAD/CAM composite resin crowns with AI. This retrospective study utilized 2D images generated from 3D models of dies scanned by a 3D scanner to build and train a CNN for predicting debonding of CAD-CAM fabricated composite resin crowns. The model was able to successfully predict the debonding of composite resin crowns from the shape of the prepared tooth [6].

3.3 Periodontology

3.3.1 Diagnostic and treatment pathways

For periodontal disease, especially periodontitis, the severity, complexity of management, extent, rate of progression, and risk factors are classified based on staging (I–IV) and grading (A–C) systems (Figure 4). Due to the highly heterogeneous clinical manifestations and inheritance patterns, dental professionals tend to conduct several diagnostic analyses based on both stage and grade to determine the accurate diagnosis. Several risk factors that may influence the periodontal condition, such as smoking, diabetes mellitus, obesity, specific genetic factors, physical activity, nutrition and stress, are also systematically and comprehensively examined for a more detailed analysis.

A broad range of periodontitis-related treatment modalities is available, including chemotherapy, resective treatment, periodontal regenerative treatment, periodontal plastic surgery for gingival augmentation, occlusal treatment, pre-prosthetic periodontal treatment and extraction and implant surgery. There is no single effective treatment approach for periodontitis. One treatment procedure may be appropriate and cost-effective for one site of the mouth, while another approach may be suitable for other sites.

STAGE		Stage I	Stage II	Stage III	Stage IV
Severity	Interdental CAL at site greatest loss	1-2mm	3-4mm	≥5mm	≥5mm
	Radiographic bone loss	Coronal third (<15%)	Coronal third (15-33%)	Extending to middle or apical third of the root	Extending to middle or apical third of the root
	Tooth loss	No tooth loss due to periodontitis		Tooth loss due to periodontitis of 5< teeth	Tooth loss due to periodontitis of ≥5 teeth
Complexity	Local	Maximum probing depth ≥4mm Most horizontal bone loss	Maximum probing depth ≥5mm Most horizontal bone loss	In addition to state II complexity: Probing depth ≥6mm Vertical bone loss ≥3mm Furcation involvement class II or III Moderate ridge defect	In addition to stage III complexity: Need for complex rehabilitation due to: Masticatory dysfunction Secondary occlusal trauma Severe ridge defect Bite collapse, drifting, flaring Less than 20 remaining teeth (10 opposing paris)
Extent & Distribution	Add to stage as descriptor	For each stage, describe extent as localized (<30% of teeth involved), generalized, or molar/incisor pattern			

GRADE			Grade A Slow rate of progression	Grade B Moderate rate of progression	Grade C Rapid rate of progression
Primary Criteria	Direct evidence of progression	Longitudinal data (Radiographic bone loss or CAL)	Evidence of no loss over 5 years	<2mm over 5 years	≥2mm over 5 years
	Indirect evidence of progression	% bone loss/age	<0.25	0.25-1.0	>1.0
		Case phonotype	Heavy biofilm deposits with low levels of destruction	Destruction commensurate with biofilm deposits	Destruction exceeds expectation given biofilm deposits; specific clinical patterns suggestive of periods of rapid progression and/or early onset disease
Grade modifiers	Risk factors	Smoking	Non-smoker	Smoker <10 cigarettes/day	Smoker ≥10 cigarettes/day
		Diabetes	No diagnostic of diabetes	HbA1c <7% in patients with diabetes	HbA1c ≥7% in patients with diabetes

Figure 4 – Staging and grading of periodontal diseases

3.3.2 Definition of the AI task

AI systems play a key role in the management of periodontal diseases. AI systems provide a variety of diagnostic and detection methods for periodontal diseases, including dental plaque, calculus, gingivitis and periodontitis. The use of unsupervised or supervised AI systems will aid in the accurate diagnosis and prediction of periodontal diseases. Periodontal disease-related AI can be applied in the following areas: 1) dental plaque and calculus detection, 2) gingivitis detection, and 3) periodontitis detection.

AI can serve as a second opinion. AI interpretations are based on the experience of professional radiologists and periodontists who participated in setting the AI algorithms. AI training relies on deep learning, especially convolutional neural network algorithms based on thousands of dental periapical and panoramic radiographic images, and continuous training occurs as more data become available. Moreover, the classification task is mainly used for periapical images, and the segmentation task is used for panoramic images. Automatic detection of regions of interest for pathologies such as alveolar bone loss is also used to provide a detailed diagnosis of the periodontal condition.

3.3.3 Current gold standard

Clinical diagnostic analyses are generally based on the signs and symptoms of gingival inflammation and periodontal tissue destruction. Common diagnostic tools are used, including clinical and dental radiographic examinations. Typically, the presence of dental plaque, calculus, clinical attachment loss, tooth mobility, bleeding and pus discharge is detected by clinicians using a periodontal probe, and sometimes, with the aid of panoramic and periapical radiographic images. However, these assessment methods are not only inconvenient but also unreliable and inconsistent. Although several attempts have been made to address or reduce their disadvantages, they still do not provide substantial advantages over existing conventional diagnostic methods, in terms of cost, time and standardization.

3.3.4 Relevance and impact of an AI solution

Periodontal disease is a major cause of tooth loss in adults and is one of the most important oral diseases contributing to the global chronic disease burden. More than half of the world's population has periodontal disease. Periodontal disease is the sixth most prevalent inflammatory disease. Globally, 11% of adults have severe and advanced periodontitis.

Periodontal disease is closely related to various sociodemographic factors (including age, sex, household income, insurance status, and residence area) and comorbidities (including diabetes mellitus, obesity, cardiovascular disease, rheumatoid arthritis, metabolic syndrome and adverse pregnancy outcomes). Therefore, early detection and accurate diagnosis are very important, not only for personal health but also for reducing the social burden.

AI is particularly useful for the standardized diagnosis of periodontal disease, which is relevant to board-certified periodontists and general dentists. In many cases, periodontal disease can be underdiagnosed by the general dentist due to a lack of experience, skill, time and cost. A more rapid, accurate, and reliable periodontal diagnosis based on AI algorithms will enable early intervention, enhanced prevention and favourable treatment outcomes. In particular, due to the visualizations of the AI second opinion, the acceptance of earlier detection and treatment or additional examinations will be improved among periodontally compromised patients.

3.3.5 Existing AI solutions

Currently, the direct use of most AI systems in clinical practice is limited. Most systems are still being used for research and investigational purposes only. Recently, several companies and related research institutes have begun developing automated diagnostics for dental radiographs based on AI. Some AI tools or solutions available for the diagnosis of periodontal diseases are as follows:

Diagnocat: Diagnocat is an AI solution that uses CBCT scans, full mouth series or panoramic and periapical radiographic images. Diagnocat assists dental professionals in the diagnosis of various oral conditions: periodontal, conservative, endodontic, prosthetic, implant and surgical. The limitation of this tool is that it mainly focuses on three-dimensional radiographic image analyses.

dentalXrai: DentalXrai measures the periodontal bone loss of each tooth on panoramics and periapical radiographs in % of the root length. It identifies the root surface with the highest bone loss and also displays bone loss levels along EFP/AAP stage definitions. The findings are displayed as overlays and as summary report. The tool has CE approval (Figure 5).

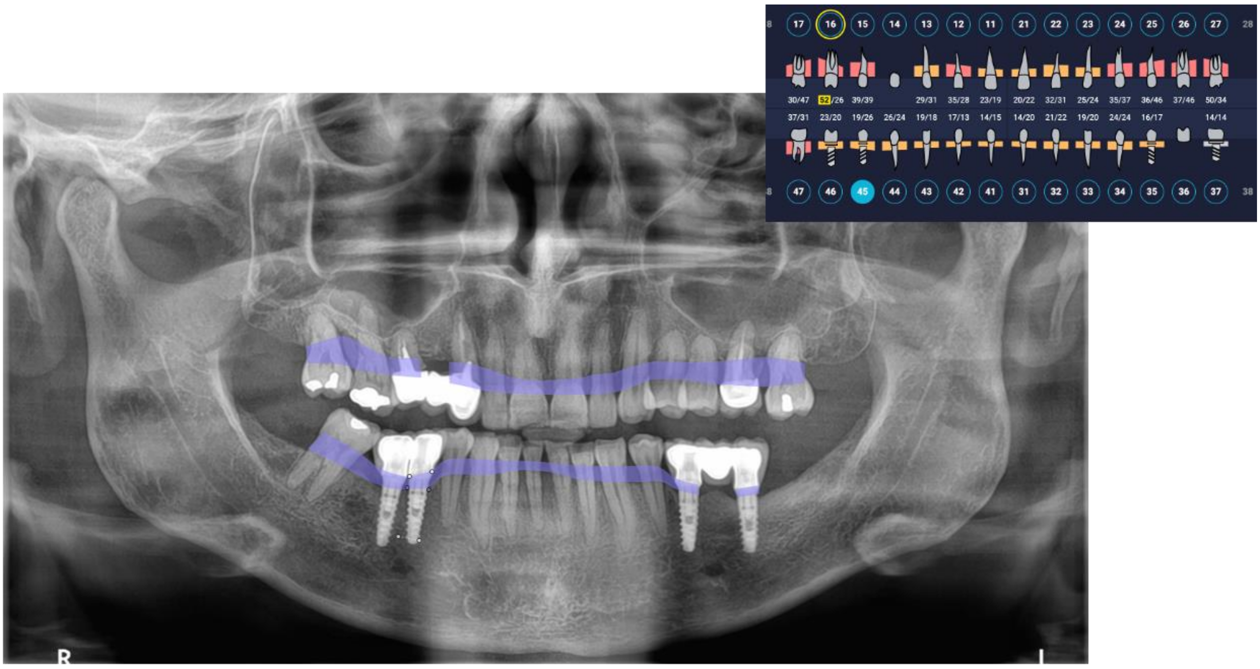

Figure 5 – DentalXrai front-end displaying AI-detected periodontal bone loss

OralCam: OralCam is a smartphone application that helps users examine common oral conditions (including dental caries, periodontal disease and dental calculus). Although it has the advantage of simple self-diagnosis, its accuracy is too low for use in clinical practice. Moreover, diagnosis is mainly confined to the labial surfaces of teeth.

Overjet: Overjet has obtained FDA approval as a deep learning solution for periodontal disease diagnosis (Figure 6). The major feature of Overjet is its deep learning-based technology that can analyse dental radiographic images in real time. Overjet is trained to measure tooth-related bone loss from these radiographic images, making it easy for dental professionals to diagnose periodontitis.

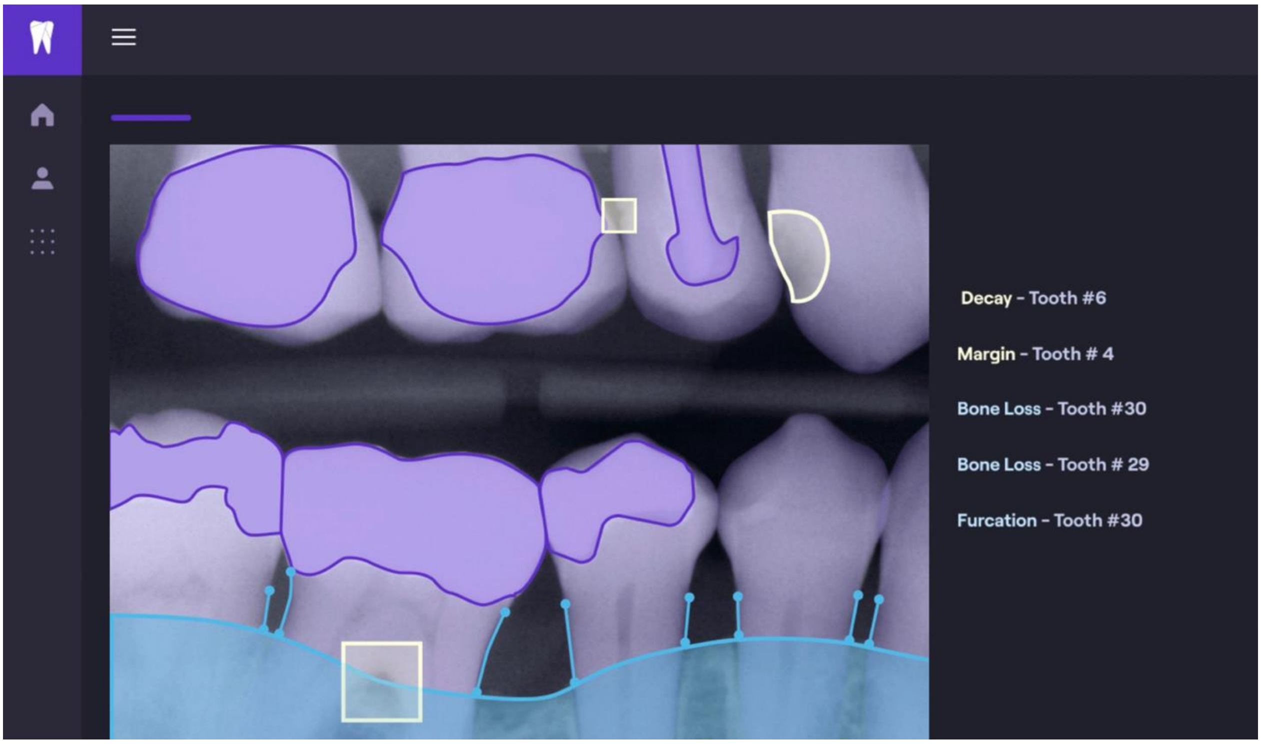

Figure 6 – Overjet displaying AI-detected periodontal bone loss

3.3.6 AI for identification and classification of dental implant systems

Dental implants are one of the most widely used and commonly accepted treatment modalities for oral rehabilitation of partially and fully edentulous patients. Nevertheless, various major mechanical (e.g., loosening or fractures of screw or fixture) and biological (e.g., peri-implantitis and peri-implant mucositis) complications of dental implants, affecting long-term survival and reintervention outcomes, are a growing concern in the dental community worldwide and a public health problem associated with a high socioeconomic burden.

Identifying and classifying the dental implant systems after post-implant surgery is one of the most important factors in preventing or treating mechanical or biological complications, but there is a fundamental and practical limit to classifying the dental implant systems using post-operative two-dimensional images including periapical and panoramic radiographs.

In the last few years, AI-based deep learning technologies, in particular deep and CNN, have rapidly become a methodology of choice for two-dimensional and three-dimensional dental image analysis. Recent studies reported that deep learning algorithms emerged as the state of the art in terms of accuracy performance for identifying and classifying various types of dental implant systems, often outperforming the dental professionals specialized in implantology (Figure 7).

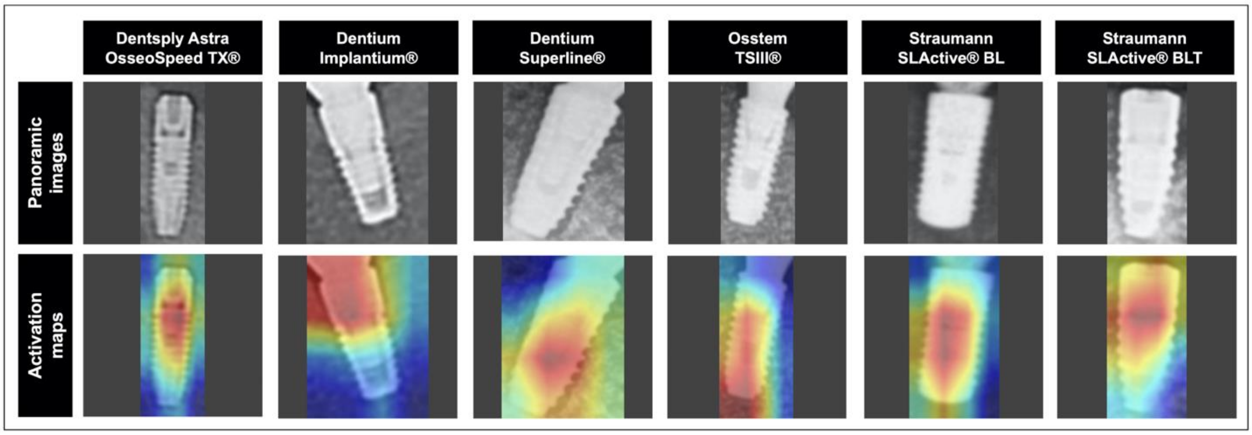

Figure 7 – Example of AI-based deep learning in 2D dental image analysis*

* Image from Lee et al, Deep learning improves implant classification by dental professionals: a multi-centre evaluation of accuracy and efficiency. Journal of Periodontal & Implant Science, 2022. Reproduced with permission.

3.4 Surgery

For the purposes of this Document, oral and maxillofacial surgery is recognized as a surgical specialty of both medicine and dentistry; nonetheless, due to the scope of this TG, the dental aspects of oral and maxillofacial surgery are considered to a larger extent. Wherever applicable, the term oral implantology is used for clarity to refer to the subfield concerning the surgical use of intraosseous dental implants.

3.4.1 Definition of the AI task

The goals of the implementation of AI in oral and maxillofacial surgery as well as oral implantology are to aid the clinician in the processes of establishing diagnoses, planning treatments and assessing individual risks associated with the proposed treatment plans. The AI tasks used to achieve these goals can be divided into regression/classification tasks (for risk prediction) and diverse computer vision tasks (for automated diagnosis and treatment planning). Preoperative virtual planning (e.g., of dental implants ideally positioned using ML) and 3D printing of surgical templates could take place in-house. To this aim, computer vision models must be able to differentiate between not only hard and soft tissue but also between compact and spongy bone, determine the dimensions of the available bone, detect relevant anatomical structures (e.g., inferior alveolar nerve, maxillary sinus) and assess the width of the gingival tissue and calculate optimal length and position of the dental implant. Completing all these tasks is considerably challenging due to the large size of the input 3D image dataset. Further, prediction models could be used to approximate implant survival and success rates while taking into consideration a large number of possible predictors, including bone structure, bone density and systemic patient-level factors. Based on preoperative diagnostics, prediction models could further be used to estimate the healing of an osseous defect following cyst or tumour removal. These approximations could contribute to obtaining a truly informed consent of the patient as well as aid the preoperative surgical decision-making process.

3.4.2 Current diagnostic and treatment pathways

In general, diagnostics and treatment planning in oral surgery are dependent on the knowledge, experience and skills of the oral surgeon. Besides the problem of an inherent bias in all non-SOP human decisions and procedures, the risk of human error is evident. The application of AI and deep learning in diagnostics, disease risk evaluation, treatment planning, and personalized risk prediction has enormous potential in the future of oral pathology, oral surgery and implantology, as well as maxillofacial surgery. The implementation of these new technologies raises the chances for further improvement in all fields of dentistry. ML models are very likely to improve accuracy in detection and diagnostics and may become an effective tool in risk prediction and individual planning. They may help to prevent human errors and lead to objective treatment decisions, which are standardized and reproducible.

Surgical tooth extraction: The extraction of teeth is a common surgical procedure with a wide spectrum of indications ranging from pathological conditions to lack of space (e.g., wisdom teeth, premolars in orthodontic indications). The procedure is associated with a number of complications which can primarily be derived from the position of neighbouring anatomical structures. In the posterior maxilla, the surgical creation of an oroantral communication is fairly common due to the proximity of the dental roots to the maxillary sinus. In the posterior mandible and especially with regards to mandibular third molars, one of the most serious complications is neuropathy due to the proximity of multiple branches of the mandibular nerve, including the inferior alveolar nerve and the lingual nerve. Since the surgical treatment of actual nerve damage is prohibitively difficult, prevention is of the utmost importance. Thus, the classification of lower third molars with the aim of neuropathy prediction has been a focus of research for almost a century. Recently, computer vision models have been used to categorize lower wisdom teeth based on existing surgical classifications. Nonetheless, the majority of current ML approaches rely on single-centre datasets. Contributors to this subtopic are currently developing advanced computer vision models using large multicentre datasets.

Cyst and tumour surgery: Cysts and tumours of the jaw are often incidental findings in a dental setting (often through a general practice dentist) which are then referred to oral surgeons. Jaw cysts and tumours can grow slowly and painlessly, meaning that, when undiagnosed for a long time, they can reach a size at which their removal necessitates radical surgery. Therefore, any tool that aids the dentist in making a primary diagnosis which they can refer to an oral surgeon is of high clinical relevance. Computer vision models have been demonstrated to detect cysts and tumours of the jaw with high accuracy. The classification into different histological subtypes has also shown promising results. Contributors to this subtopic have further published results showing that a combined computer vision workflow of an object detection model and multiple segmentation models is able to identify whether a cyst is odontogenic. This could be of additional benefit to both general practice dentists and oral surgeons as patients with cysts which penetrate the facial bone can present with severe symptoms (e.g., pain, swelling). Having a reliable diagnostic aid when deciding whether a root-canal treatment needs to be performed is of high clinical relevance. Further research could focus on automatic three-dimensional segmentation of jaw cysts and tumours. This would decrease the surgical risk of damaging neighbouring structures and provide a more accurate baseline for better approximation of post-operative bone regeneration.

Implantology: Radiographic examinations are the base for modern implant planning. In addition to clinical examinations (assessment of remaining teeth, height and width of the jaw, mucosal assessment), radiographic imaging in the form of 2D (panoramic radiograph) or 3D imaging (computed tomography [CT] or cone beam CT [CBCT]) is crucial to provide information about anatomical structures and the actual bone volume. Based on this information right now in the most advanced practice, the surgeon can determine the length, width and the angle of the implant through an implant planning software. Importantly the quality of this human treatment planning depends on the knowledge, experience and skills of the oral surgeon. Especially in borderline cases, this treatment planning can prove difficult because some implant positions might not be realistic during the surgical procedures and the prosthetic reconstruction might be impossible or not in line with aesthetic considerations. After the planning phase, the dataset will be transferred to the dental laboratory, which produces a 3D-printed drilling template.

Bone grafting: Based on the inherent regenerative capacity of the bone tissue, grafting is a routine treatment to either reconstruct lost bone (e.g., to inflammatory osteolysis) or increase volume (e.g., to enable dental implant placement. Depending on the quantitative as well as geometric aspects of osseous defects, there are different strategies available for grafting. In every case, the main goal of modern bone grafting procedures is to increase bone volume at the grafting site while avoiding morbidity at the donor site (if there is one). The material used for grafting can be autologous (i.e., from the same individual), allogenic (i.e., from a different individual of the same species), xenogenic (i.e., from a different species) or synthetic. These have different properties which influence the outcome following grafting. Furthermore, while there is a general agreement that autologous bone is the gold standard, there is a considerable international disagreement in which bone grafting materials are preferred to increase the volume of a combined graft, with European clinicians generally preferring xenogenic material and American clinicians generally preferring allogenic material. The need for more objectivity in bone grafting is apparent. Various regression models are already being used to approximate long-term quantitative outcomes following bone grafting. Nonetheless, the potential for AI remains high. Computer vision models could be highly beneficial in assessing the grafting site from both a quantitative and a geometric point of view. The choice of bone graft material and volume would then be made based on an evaluation arguably more objective than clinicians' personal preference.

3.4.3 Relevance and impact of an AI solution

Dental radiography (both two- and three-dimensional) is commonly used in daily practices and provides an incredibly rich resource for AI development. This attracts many researchers to develop its application for various purposes. The application of AI in oral surgery, dental implantology and maxillofacial surgery is very heterogeneous and needs further development and evaluation before clinical utilization. Tasks include object detection, object classification, as well as image segmentation models. Further AI applications in oral surgery include models to predict individual risk and therapy success. Another important task is the treatment of patients, especially the task in designing patient-specific objects and optimizing the fit, length and diameter of dental implants. With the assistance of novel AI models and a strong collaboration of dental professionals worldwide, clinical effectiveness in oral surgical and maxillofacial treatments can be optimized.

3.4.4 Existing AI solutions in oral surgery

In dentistry, and especially oral surgery, radiographic examination is a standard of care for the evaluation of physiological and pathological structures as well as treatment planning. Successful treatment of complex surgeries in patients of varying ages and health conditions depends on the examiner's education and training, which forms the basis of treatment planning. At present, much of the preoperative planning is still done manually, but digital, semi-automatic planning is becoming increasingly common in oral surgery, especially in dental implantology.

In oral and maxillofacial surgery, automated planning has been introduced to simplify the preoperative surgical workflow. Currently, bony reconstruction of the face after resection of extensive tumours is a particular challenge in facial surgery. Besides these complex reconstruction cases, AI and ML will lead to the automation of aesthetic evaluation, smile design and treatment-planning processes. There are already commercially available software tools for the virtual planning of these procedures [15]. However, most of these tools are not automated and so complicated to use that they are rarely used directly by surgeons and require cumbersome and lengthy communication with service providers.

Currently, many AI models have been developed for automatic diagnosis, the detection of pathologies and the prediction of disease risk. AI in this field has the potential to revolutionize and simplify these complicated processes.

3.5 Oral Medicine and maxillofacial radiology

3.5.1 Scope of AI and current gold standard

Diagnosis of oral lesions is of crucial importance in dental practices because early detection significantly improves prognosis. Diagnosis and classification of bone diseases (cysts and tumours) are carried out. To date, the majority of oral diagnostics are based on manual workouts and personalized perceptions. AI has the potential to revolutionize the diagnostic accuracy and prediction of disease course, outcomes and prognosis.

1 Diagnosis and classification of systemic diseases based on oral manifestations.

2 Diagnosis and classification of periodontal disease.

3 Use of immunologic parameters to diagnose aggressive periodontitis.

4 Classification of halitosis based on the identification of periodontal pathogens in saliva.

5 Predicting recurrence of aphthous ulcers.

6 Diagnosis and classification of head and neck syndromes

3.5.1.1 Temporomandibular joint disorders (TMD)

1 Use of magnetic resonance imaging to determine the prognosis of TMD.

2 Use of screening questions to aid in treatment of TMD.

3 Identification of subgroups of internal derangement of the temporomandibular joint.

4 Classification of TMD based on clinical signs and symptoms.

5 Automated TMD diagnostics.

3.5.1.2 Oral cancer

1 Diagnosis and classification of oral precancers and cancer.

2 Prognosis of oral cancer considering lesion histology and genetic characteristics.

3 Assessment of hypernasal speech following treatment of oral/or pharyngeal cancer.

4 Risk assessment for oral cancer (prediction).

5 Prediction of cervical lymph node metastasis of oral squamous cell carcinoma.

3.5.1.3 Masticatory muscle diseases

1 AI-based classification of masticator muscle disorders.

2 Prediction of masticator muscle disorders based on chewing pattern.

3.5.2 Relevance and impact of an AI solution

AI is being studied concerning various topics in the field of maxillofacial radiology. Radiographic imaging is very useful for diagnostic purposes and to ensure proper treatment. Maxillofacial radiologists as professionals understand the basic principles and characteristics of radiographic imaging and interpret them in terms of various diseases and continue to play an important role in artificial intelligence-related research.

Radiographs of interest for AI-based diagnostics in oral and maxillofacial radiology are:

a Conventional radiographs

- Periapical
- Bitewings
- Panoramic radiograph
- Lateral cephalogram
- Hand and wrist radiograph

b Advanced imaging

- Plane CBCT, 3D CBCT scans (Figure 8).
- CT scans
- MR scans

In the field of oral and maxillofacial radiology, AI can be used for the diagnosis of a broad range of oral diseases, dental caries, periodontal disease, osteosclerosis, odontogenic cysts and tumours, diseases of the maxillary sinus or temporomandibular joints**,** vertical root fractures, periapical pathosis and cysts and tumours of the jawbone. It can also be used for the detection of the filling or filled teeth.

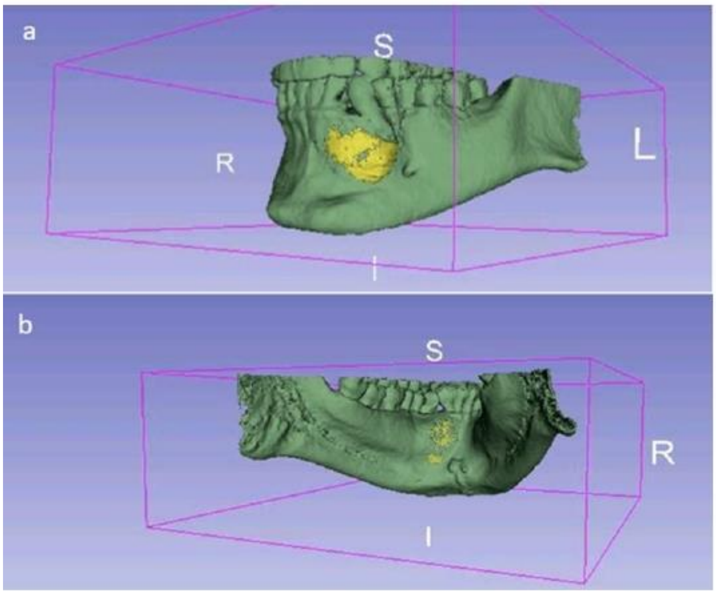

Figure 8 – 3D segmentation of a bone tumour

For age and sex determination in forensic odontology, dental radiology-based methods being non-invasive and accurate are beneficial over conventional methods such as sectioning or extracting teeth. They can therefore be used for both living and deceased individuals.

- Automated age estimation based on anatomical landmarks and tooth structures.
- Automated sex estimation based on anatomical landmarks and tooth structures.
- Identification of deceased based on dentition and tooth structure.

3.5.3 Existing AI solutions

Artificial intelligence holds a promising future in the field of oral medicine and maxillofacial radiology. AI-based diagnostics for detection of oral precancer and cancer, soft tissue lesions and odontogenic tumours are highly needed in context of current conventional tedious diagnostic criteria. Though much research has been published related to this, most systems are still being used for research and investigational purposes only. Recently, several companies and related research institutes have begun developing automated diagnostics from clinical photographs and histopathological features based on AI. However, AI-based prediction and treatment tasks are still in a naïve stage.

AI-based research in maxillofacial radiology has travelled far. There are many AI-based solutions based on 2D and 3D radiographs are available and commercialized for use in clinical practice. Available solutions are:

Diagnocat: Diagnocat is an AI solution that uses CBCT scans, full mouth series or panoramic and periapical radiographic images. Diagnocat assists dental professionals in the diagnosis of various oral conditions: periodontal, conservative, endodontic, prosthetic, implant and surgical. The limitation of this tool is that it mainly focuses on three-dimensional radiographic image analyses.

Relu: Relu is an AI solution which creates 3D anatomical models in minutes using AI-assisted segmentation software as well 3D printing anatomical models. It is highly efficient in automated nerve tracing, incredible bone segmentation as well panoramic arch tracings. This software has high compatibility for use in implantology, craniomaxillofacial surgery, orthodontics and treatment planning.

dentalXrai: DentalXrai measures the periodontal bone loss of each tooth on panoramic and periapical radiographs in % of the root length. It identifies the root surface with the highest bone loss and also displays bone loss levels along EFP/AAP stage definitions. The findings are displayed as overlays and as summary reports. The tool has CE approval.

OralCam: OralCam is a smartphone application that helps users examine common oral conditions (including dental caries, periodontal disease and dental calculus). Although it has the advantage of simple self-diagnosis, its accuracy is too low for use in clinical practice. Moreover, diagnosis is mainly confined to the labial surfaces of teeth.

Overjet: Overjet has obtained FDA approval as a deep learning solution for periodontal disease diagnosis. The major feature of Overjet is its deep learning-based technology that can analyse dental radiographic images in real-time. The Overjet is trained to measure tooth-related bone loss from these radiographic images, making it easy for dental professionals to diagnose periodontitis.

Many more AI-based software with advanced features are still evolving, having advanced technologies with high accuracy and precision.

3.6 Endodontics

3.6.1 Definition of the AI task

AI algorithms learn the relationship between the characteristics of the given data and the ground truth. In endodontics, their current use encompasses classification, prediction and segmentation tasks:

Screening and diagnosis

- Periapical pathosis detection
- Vertical root fracture
- Determination of working length
- Treatment planning
- Apical foramen determination

- Tooth-root morphology determination
- Prognosis
- Prognosis prediction in endodontic microsurgery
- Predicting treatment failure.

The data utilized by the algorithms encompass:

- Numeric data — patient age, exam scores, etc.
- Categorical data — disease labels
- Image data — radiographic analysis (OPG, PA, CBCT).

The ground truth used by the studies so far is:

- Specialist estimation (single operator / a panel of two or more operators)
- Semi-automated (supervised computer-aided segmentation of pathologies on 2D and 3D images)
- Simulated lesions.

Scope for future AI applications in endodontics:

- AI-assisted patient management software
- Comprehensive diagnostic and prediction algorithms based on medical, dental, clinical and paraclinical patient data
- AI-assisted treatment planning and personalized care.

The current goal of the use of AI in endodontics is:

- Establishing adequate ground truth for data labelling with the notion that trained CNNs can reach or surpass the diagnostic performance of experienced clinicians
- Establishing adequate generalizability of the trained models so that they can adequately predict on data which sources differ from those involved in the model training [16].

3.6.2 Current gold standard

The current gold standard in endodontics:

- Diagnosis
- A blend of information gathered from the medical and dental history, the clinical examination, the clinical and paraclinical tests
- Dental and medical history
- Extra and intra-oral examination
- Clinical test: percussion, palpation, mobility, periodontal examination, pulp tests (vitality/vascularity), staining and transillumination, selective anaesthesia
- Radiographic examinations (PA, OPG, CBCT)
- Clinical classification of pulpal and periapical diseases
- Treatment planning
- A decision-making process on the necessity and modality of treatment is executed by the clinician based on the collected data based on best practices
- Prognosis
- Acquisition of clinical and paraclinical data during the follow-up period and an educated prediction based on the published cases in the literature and the clinician's own experience.

The current gold standard of the diagnostic and treatment pathways in endodontics is limited by the country-specific discrepancies in patient data collecting protocols, the sensitivity and specificity of the devices, disease classification variations, access to dental materials, variations in treatment modalities and materials, the patient's payment capacity and the expertise and experience of the clinician interpreting the data and performing the treatment.

3.6.3 Relevance and impact of an AI solution

The application of AI algorithms in endodontics has the following potential:

- Diagnosis
- Improve time efficiency
- Improve accuracy and precision
- Empower patient and clinician education
- Treatment planning
- Automate laborious tasks (working length determination)
- Referral suggestions
- Suggest alternative treatment protocols
- Highlight potential difficulties during the treatment
- Prognosis
- Improved prognosis prediction based on objective data.

The successful overall implementation of AI algorithms in the daily dental practice will improve quality of care, efficiency in time, cost and labour.

As AI algorithms have already found their way into the market, benchmarking grows even more paramount to ensure:

- Robustness of the AI model
- Generalizability of the AI model.

3.6.4 Existing AI solutions

An expanding range of AI applications are already available in the research setting:

- Diagnosis
- Periapical pathosis detection
- Vertical root fracture
- Determination of working length
- Treatment planning
- Determination of working length
- Apical foramen determination
- Tooth-root morphology determination
- Prognosis
- Prognosis prediction in endodontic microsurgery
- Predicting root-canal failure.

The number of solutions available for clinical care and regulated as medical products is much smaller and the current scope mainly focuses on the detection of apical lesions on 2D and 3D radiographs as well as the segmentation of the root-canals on 3D radiographs.

3.7 Orthodontic

3.7.1 Current diagnostic and treatment pathways

Orthodontic treatments aim at correcting the malocclusion and repositioning the dentition concerning the craniofacial structures in the most harmonious way. To accomplish a successful orthodontic treatment, it is necessary to conduct a patient interview, an examination and gather proper records with the purpose of a comprehensive diagnosis. Clinical evaluations have to be systematic and organized so the clinician does not override any of the diagnostic items that include patient questionnaires and interviews, as well as extra-oral and intra-oral clinical examinations and paraclinical assessments. This can then result in an appropriate treatment plan and desirable treatment outcomes. The following data is required before designing a comprehensive treatment plan:

1 Personal details
2 Chief complaint
3 Medical history
4 Dental history
5 TMJ assessment
6 Clinical examination
7 Paraclinical assessment
    a. Facial/Intra-oral Photographs
    b. OPG Radiographs
    c. Lateral cephalometry
    d. Study casts
    e. Cone beam computed tomography
    f. Hand and wrist radiograph.

Due to the limitations in the length of this Document, the paraclinical data modalities will be discussed here since they are more applicable in the AI context.

A) Facial/intra-oral photographs

The following items can be evaluated on facial or intra-oral photographs:

1 The shape of the head (mesocephalic, dolichocephalic and brachycephalic)
2 Facial form (mesoprosopic, euryprosopic, leptoprosopic)
3 Facial symmetry
4 Facial proportions
5 Facial profile
6 Facial divergence
7 Anterior-posterior tooth/jaw relations
8 Lips examination
9 Nose examination
10 Chin examination
11 Nasolabial angle
12 Palatal evaluation
13 Frenal attachments
14 Overjet
15 Overbite

16 Transverse mal relations
17 Arch form
18 Smile analysis
19 General aesthetic evaluation.

B) Panoramic radiographs

It is necessary to assess general conditions regarding patients' dentition, before treatment planning. The following items can be evaluated on panoramic radiographs:

1 Absence and presence of teeth, supernumerary teeth
2 Evaluation of dentition age
3 Evaluation of root formation
4 Status of deciduous teeth
5 General oral health (caries, previous restoration, etc.)
6 Presence of any pathologies.

C) Lateral cephalograms

The main data modality used in orthodontics is lateral cephalograms. The main dental and skeletal orthodontics problems can be diagnosed through lateral cephalograms. Diagnosis in lateral cephalograms is based on the anatomic landmarks and their quantitative relations. The main landmarks are displayed in Figure 9.

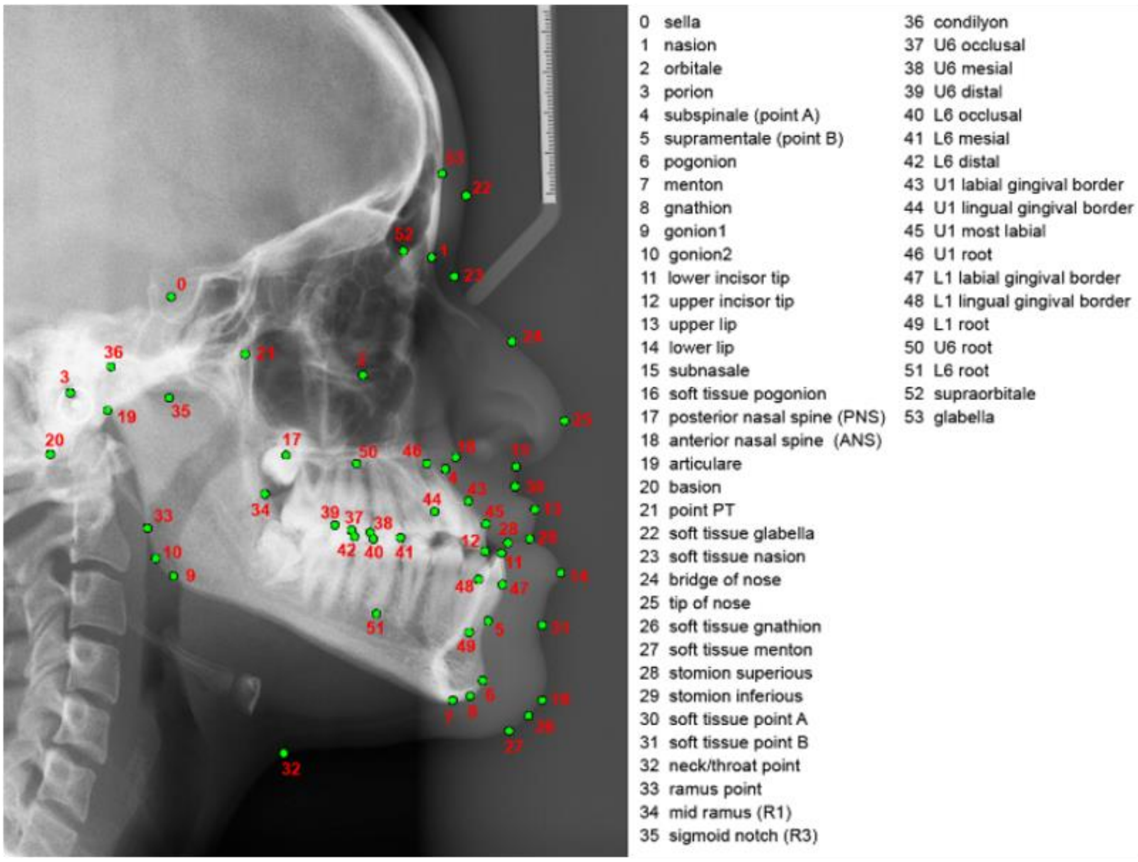

Figure 9 – Automated cephalometric landmarking*

* Image from https://personalpages.manchester.ac.uk/staff/claudia.lindner/default_files/curr_project_ceph.htm [AH6] [FS7]

Other than the mentioned items, skeletal maturation degree can be evaluated on lateral cephalograms. Determination of skeletal maturation degree is necessary for treatment timing and reaching a desirable treatment outcome. The most common classification approach is based on the cervical vertebrae maturation degree and is displayed in Figure 10.

Schematic representation	CS 1	CS 2	CS 3	CS 4	CS 5	CS 6

Inferior borders of C2, C3, and C4[a]	F, F, F	C, F, F	C, C, F	C, C, C	C, C, C	C, C, C
C3 morphology[a]	T	T	T	RH	S/RH	RV/RH
C4 morphology[a]	T	T	T/RH	RH	S/RH	RV/RH
Clinical implication	Prepubertal stage	Prepubertal (“get-ready”) stage	Circumpubertal stage	Circumpubertal stage	Postpubertal stage	Postpubertal stage

[a] F= Flat; C= Concavity; T= Trapezoid; RH=Rectangular Horizontal; S=Square; RV=Rectangular Vertical

Figure 10 – Vertebrae maturation*

* Image from "McNamara et al, The cervical vertebral maturation method: A user's guide, The Angle Orthodontist, 2018 [AH8] [FS9] "

D) Study casts:

Following items can be evaluated on orthodontics study casts:

1 Occlusion
2 Space analysis
3 Arch form
4 Arch symmetry
5 The curve of Spee
6 Arch width.

E) Cone beam computed tomography (CBCT):

Nowadays, CBCTs are becoming more popular in orthodontics. In clinical orthodontics, CBCT can provide volumetric information in three-dimensional form. As a result of advancements in CBCT technology, compared with conventional radiography, CBCT provides much more accurate and reliable information regarding diagnosis, treatment and follow-up. CBCTs can be used in the following cases:

1 Evaluation of impacted and transposed teeth
2 Evaluation of root resorption of adjacent teeth
3 Evaluation of affected temporomandibular joints contributing to malocclusion
4 Airway morphology analysis
5 Midpalatal suture maturation evaluation
6 Orthognathic surgery treatment planning

7 Locating the best temporary anchorage site for skeletal anchorage.

F) Hand and wrist radiograph

Hand and wrist radiographs have been used to predict the correct bone age (skeletal maturity) that can be compared with the patient's chronological age.

3.7.2 Definition of the AI task

Various AI tasks can be defined in each modality for the purpose of enhancing orthodontic diagnosis and treatment planning. Due to the limitations in the length of this document, it only mentions those which already been addressed in the literature:

1) Facial/intra-oral photographs
   a) Landmark detection
   b) Facial attractiveness score (as regression task)
   c) Facial asymmetry and distortion (as a classification task)*
2) Panoramic radiographs
   - All the AI tasks in dentistry can be beneficial in case of orthodontics, e.g., tooth numbering, caries detection, etc.
3) Lateral cephalometrics
   a) Landmark detection
   b) Lateral cephalometric analysis (as a classification task) *
   c) Cervical maturation degree (as a classification task) *
4) Study Casts
   a) Arch form predication (as a classification task) *
5) CBCT
   a) Maxillary suture assessment (as a segmentation task).

These tasks can be defined in the context of AI in three different approaches:

1. <u>One-stage automatic approach</u>

The AI model decides on its prediction as an end-to-end model, the model will be fed with the raw image and the diagnostic class will be the model's output.

2. <u>Two-stage automatic approach</u>

The AI model is trained to detect landmarks. Then through some quantitative calculations between landmarks, it predicts the diagnostic class.

3. <u>Semi-automatic approach</u>

The AI model is trained on manually extracted data to predict the diagnostic class, e.g., some studies defined landmarks around cervical vertebrae and fed their models with features based on the relations between these landmarks. However, this approach has many shortcomings, and practitioners cannot use all the advantages of AI in their practice.

3.7.3 Current gold standard

As can be seen, most orthodontics diagnoses are based on clinical and paraclinical evaluation. Almost in all of the AI tasks mentioned, there are no definitive gold standards and diagnoses are made by orthodontists. There is even very low inter/intra-observer agreement in some cases. To encounter this challenge, most diagnostic criteria are defined based on quantitative relations based on landmarks (angle, proportions, distance, etc.). However, this approach is very time-consuming and prone to error itself.

3.7.4 Relevance and impact of an AI solution

In the diagnosis and evaluation of orthodontic problems, AI can be effective. Orthodontic diagnosis and treatment planning for malocclusion can be a very difficult and time-consuming procedure. New practitioners often have trouble making these decisions since they are traditionally made on the basis of clinical experience. AI approaches might be helpful for solving these problems. Cephalometric analysis, for example, is frequently used to detect skeletal and dental anomalies in orthodontics. Manually locating cephalometric landmarks requires a lot of time and is prone to error. The use of deep learning models for landmark detection has achieved outstanding results in recent studies.

Precision medicine is one of the most important applications for AI-based systems in orthodontics. To enhance treatment outcomes, researchers are considering customized treatment approaches based on the characteristics of each patient. Orthodontic and precision medicine advancements have led to breakthroughs in the development of customized treatment approaches. Precision orthodontics has been reported to be the next paradigm shift in orthodontic treatments. AI is capable of enabling and enriching precision medicine approaches based on data.

3.7.5 Existing AI solutions

In research, a recently published scoping review identified 49 studies that used AI approaches in orthodontics. It has been reported that the number of publications in this field is growing (Figure 11).

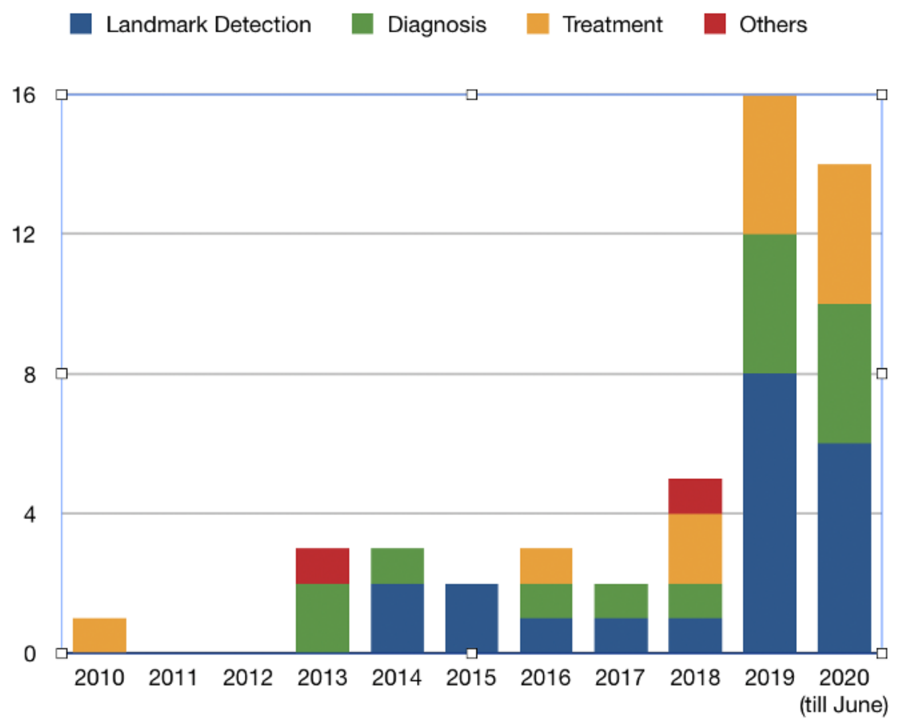

Figure 11 – Publications in orthodontics using AI approaches*

* Image from "Mohammad-Rahimi et al, Machine learning and orthodontics, current trends and the future opportunities: A scoping review. American Journal of Orthodontics and Dentofacial Orthopedics, 2021[AH10] [FS11] "

Most of the studies reported the AI model's application on the lateral cephalometry landmark detection and they reached a good performance here. However, orthodontics utilizing AI diagnosis and decision-making is in its very beginnings.

In industry, there are a few companies trying to provide AI solutions in orthodontics. They can be divided into these categories:

1. Diagnostic assistant tools
   - WeDoCeph: This company provides a tool for cephalometric tracing analysis on 2D lateral or PA X-rays
   - CephX: This company provides a tool for cephalometric tracing analysis on 2D lateral cephalometry and 3D CBCTs. It can also do airway volume analysis on 3D CBCTs.
2. Patient management and monitoring
   - Dental monitoring: Engage new patients using AI-generated reporting and smile simulations.
3. AI-based aligners
   - 3D Predict
   - Soft Smile.

3.8 Oral and maxillofacial oncology

3.8.1 Definition of the AI task

Head and neck squamous cell carcinoma (HNSCC) is the most common cancer of the oral cavity. Treatment for HNSCC has a risk for significant morbidity and outcomes for mid- to late-stage disease demonstrates a higher risk for death. Frequently, HNSCC is preceded by a precursor lesion, oral epithelial dysplasia. Epithelial dysplasia is generally graded to estimate its potential for transformation to HNSCC (Figures 12, 13).

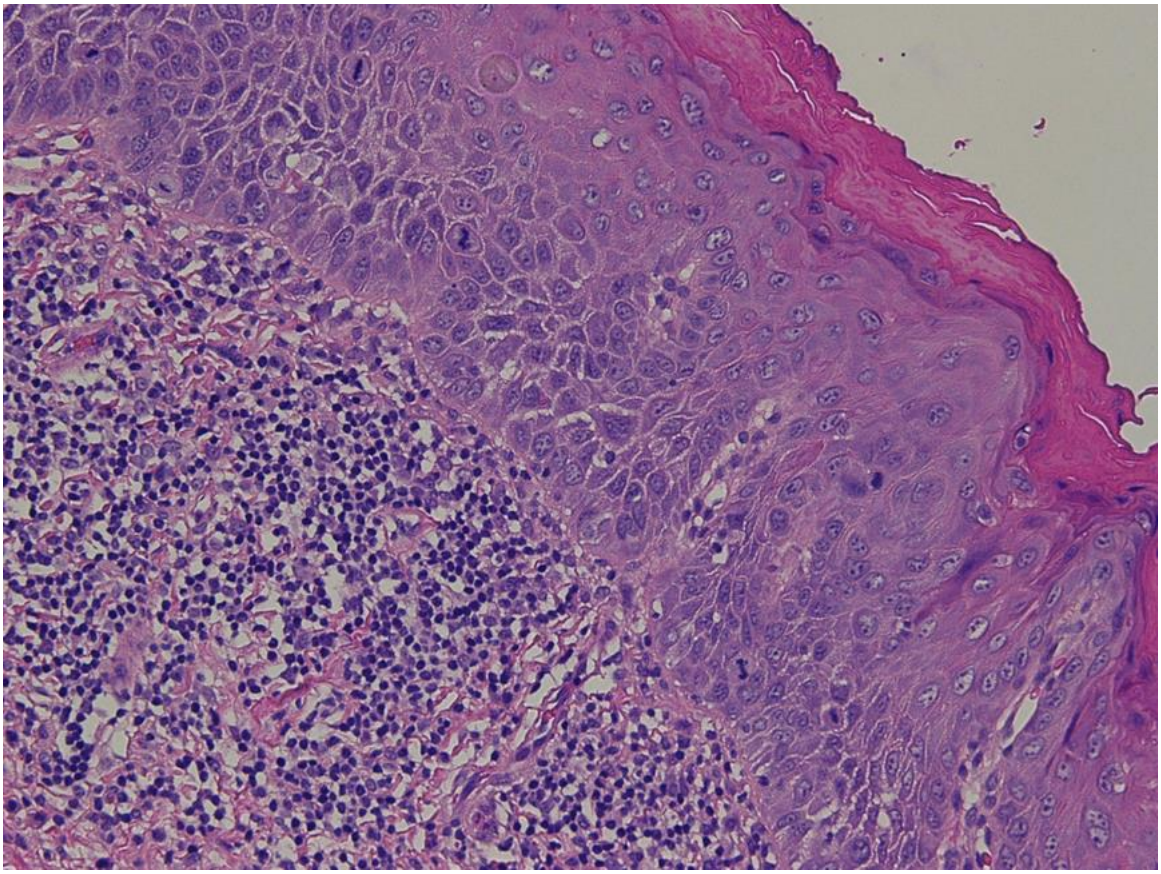

Figure 12 – Example of a dysplastic lesion

Figure 12 is an example of a dysplastic lesion. There are numerous mitotic figures, including some above the parabasal region. The basal and parabasal nuclei demonstrate hyperchromasia and pleomorphism. Most pathologists would probably agree that this represents a dysplastic lesion, though there would likely be some disagreement as to if this represents a mild dysplasia or moderate dysplasia, or if this represents a low-grade dysplasia or high-grade dysplasia. Epithelial cells are not seen beneath this surface epithelium; thus, it is not yet an invasive lesion.

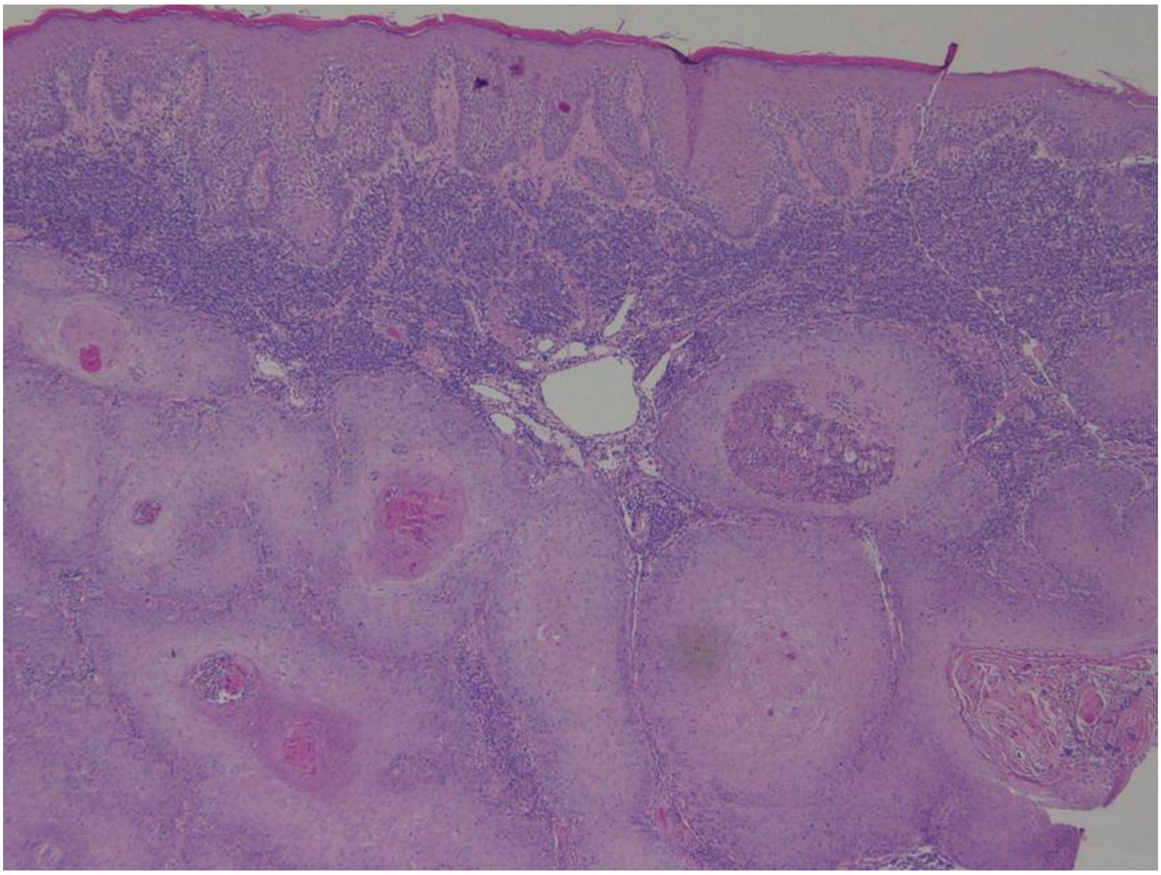

Figure 13 – **Example of invasive squamous cell carcinoma**

Figure 13 is an example of invasive squamous cell carcinoma. The overlying epithelium is at the top of the image; the rounded nests of cells containing keratin in the middle to bottom of the image represent invasive disease. Pathologists should be able to recognize that this lesion is invasive squamous cell carcinoma.

However, this grading involves the assessment of numerous architectural and cytological features that individually are somewhat subjective, and as a result there is substantial interrater disagreement. This is highlighted by the fact that several grading systems exist in an attempt to improve interrater observability and the determination of the risk for transformation. One reported method to improve interrater disagreement in this context is to obtain a consensus agreement between multiple pathologists, but institutions do not always have a large number of content experts to review these cases. This project aims to train computer vision tasks to classify oral epithelial dysplasia based on a consensus of pathologist opinions on the dataset.

The AI is making a classification between five classes: benign, mild epithelial dysplasia, moderate epithelial dysplasia, severe epithelial dysplasia, and invasive squamous cell carcinoma (SCC). A change to this classification is undergoing validation at multiple institutions – renaming mild dysplasia to low-grade dysplasia, and combining moderate and severe dysplasia into a category of high-grade dysplasia. This binary classification of dysplasia has not yet been completely validated in the oral cavity.

This will be a classification task for the aforementioned five categories (as well as testing the performance of the binary dysplasia classification system). After the development of this classification algorithm, the TG will also re-evaluate the quality of ancillary data available to evaluate the possibility of exploring a predictive regression to establish risk of transformation to invasive squamous cell carcinoma; preliminary data suggests that obtaining the quality of outcomes data will be very challenging.

Minimally, the data will consist of 2 000 whole slide histopathology images (WSI) scanned at 40× magnification as well as the original diagnosis for the slide. Those institutions that do not have access to whole slide imaging will obtain histomicrographs rather than WSI. After receipt of the images, 12 oral and/or head and neck pathologists will make their diagnosis; the target classification value will be the majority opinion of these 12 pathologists. Ancillary data will also be collected for the possibility of future use, including the diagnosis initially given for that image, patient age (maxed at 90 to comply with USA HIPAA regulations), whether or not a patient had SCC in the future, and whether or not the patient had a recurrence, metastasis or death.

The output is a classification per the five classes as defined above as well as an indicator of the algorithm's confidence in that diagnosis (this may be a percentage confidence interval or a rank order list of the likelihood that this could represent any category). Intermediate tasks may include various segmentation tasks, such as identification of epithelium or nuclei; features of these segmentations may be features of the overall classification model.

In addition, inter-institutional generalizability analysis will be performed. Iteratively, institutions will be included and excluded from the AI training, which is then tested on data from the hold-out institutions.

3.8.2 Current gold standard

This section provides a description of the established gold standard of the addressed health topic.

The oral cavity is subject to continuous trauma that results in many reactive lesions. In addition, immunological and systemic conditions can occur in the oral cavity. Most of these lesions can clinically mimic true leukoplakia. Leukoplakia appears clinically as a thick white patch and histopathologically ranges from benign hyperkeratosis to epithelial dysplasia to invasive squamous cell carcinoma (the most common cancer affecting the oral cavity). Epithelial dysplasia is considered a premalignant condition that should be recognized early; treatment with excision helps to prevent progression into invasive squamous cell carcinoma. Because leukoplakia can histologically present as multiple different pathologic entities of variable severity, biopsies are done to provide a definitive diagnosis to direct clinical management.

The histopathological distinction between clinically white and/or red lesions is not always simple because some reactive epithelial atypias can be indistinguishable from true epithelial dysplasia. Further complicating the histologic diagnostic process, epithelial dysplasia is classified according to its severity as mild, moderate and severe; as the severity increases the likelihood of transformation into an invasive squamous cell carcinoma also increases. Thus, these categories modify the management of disease, as more severe dysplasias receive more aggressive treatment. Despite the importance of this classification, the histopathological classification of epithelial dysplasia is highly subjective between pathologists and demonstrates a high interrater observability.

Depending on the study, the rate of interrater observability when using the WHO criteria for grading oral epithelial dysplasia (as published in the WHO Classification of Head and Neck Tumours) rarely exceeds a Cohen's kappa of 0.5; most studies show an interrater observability of 0.2–0.4.

3.8.3 Relevance and impact of an AI solution

This section addresses the relevance and impact of the AI solution (e.g., on the health system or the patient outcome) and describes how solving the task with AI improves a health issue.

Because of the high interrater observability with gold standard histopathologic diagnosis, a standardized tool that lacks subjectivity with the ability to classify normal and/or reactive epithelium, epithelial dysplasia (with grading as mild, moderate or severe), and invasive squamous cell carcinoma would improve the diagnosis of these lesions and assist in guiding clinical decision-making in the management of these processes. Recent studies show promising results using AI in classifying mucosal epithelium as invasive versus non-invasive. Therefore, the use of AI may act as an aiding tool that helps pathologists achieve a precise diagnosis.

If the algorithm learned how to accurately differentiate between normal, reactive, dysplastic and neoplastic epithelium, this will result in a tremendous improvement in diagnosing one of the most challenging and clinically significant use cases of the oral cavity, resulting in receiving the appropriate management for the patient. For example, if a pathologist misdiagnoses early epithelial dysplasia and interprets it as reactive changes secondary to inflammation, the patient will not have the lesion removed, allowing a dysplasia to possibly progress a more severe dysplasia or SCC. If this progress is also not identified, it is more likely that the patient would require adjuvant treatment such as neck dissection, chemotherapy and/or radiotherapy. Therefore, early diagnosis and early intervention can improve patient outcomes as well as reduce health system costs by avoiding adjuvant treatments required in advanced stage disease.

Benchmarking is important because it facilitates the assessment of the existing work in the context of AI in dental diagnostics and digital dentistry. It evaluates different aspects of the existing work on benchmarking of AI systems (e.g., relevant scientific publications, benchmarking frameworks, software, results, annotation and histopathological evaluation attempts). Evaluating and collecting all relevant data from previous benchmarking will help implement the most appropriate process needed in this topic group. Additionally, other disease entities throughout the body have a similar pattern of disease progression (e.g., SCC of the skin) and may benefit from the experiences of this group.

3.8.4 Existing AI solutions

This section provides an overview of existing AI solutions for the same health topic that are already in operation. It should contain details of the operations, limitations, robustness and the scope of the available AI solutions. The details on performance and existing benchmarking procedures will be covered in Section 6.

A recent overview of studies assessing the classification and prognostication found that some early studies using clinicopathologic variables (and fewer using histology images, usually whole slide images) highlighted promising results when AI is applied to a binary classification of invasive SCC versus non-invasive disease. For those that used histology images, different ML approaches were used to demarcate specific histological features of interest to compare differences in spatial architectural patterns for differentiation between benign and malignant lesions. However, the vast majority of these studies use datasets comprising fewer than 1 000 patients, raising the question of the ability of these algorithms to respond reliably to data from other institutions.

More importantly, there is extremely limited data on the use of AI to diagnose oral dysplasias based on histology alone. Few reports exist of using AI in the context of cytology as well as various ancillary diagnostic techniques such as infrared spectroscopy. The TG was only able to find a single in-press publication using 203 whole slide images from a single institution that was able to have high agreement (Cohen's kappa of ~0.8) with humans. Because the dataset is larger, more diverse and with a greater number of human interpretations, this project would be able to offer unique diagnostic assistance to pathologists while also providing that assistance with materials conventionally created in the context of patient care and without the assistance of other costly devices.

To the knowledge of the TG, there are no validated medical tests addressing the classification of oral malignant and/or premalignant conditions.

Studies for identifying SCC use clinicopathologic data (e.g., staging criteria) and/or whole slide images of SCC and non-SCC disease entities. Some studies use a convolutional neural network or fully connected network approach on the whole slide images, while others may use subfeatures (e.g., tumour-infiltrating lymphocytes, nuclear features such as pleomorphism and clinicopathologic data) to create a final classification or regression prediction. An overview and table providing a summary of current progress can be found in Sultan et al. (2020) [17].

Many benchmarks that would benefit this project would also benefit other projects relating to cancer, including nuclear features (pleomorphism, hyperchromasia), mitotic activity and architectural patterns suggesting invasive disease. The WHO Classification of Head and Neck Tumours (5th edition) [7] contains a table more broadly delineating features of dysplasia that pathologists interpret during the diagnosis of dysplasia.

In order to be broadly available to a wide audience, the TG suggests that a front-end website be made available allowing for an image upload with a return of the AI output. Translation would be made available in this context.

As discussed above, the scope of AI systems in this domain typically involves either a classification of invasive versus non-invasive disease, prognostication of invasive SCC or uses machine learning in combination with ancillary testing to determine severity of dysplasia. These are all relatively small studies with limited validation testing on a broad variety of data to ensure that their findings are compatible with data from other institutions.

3.9 Ethics

The ethics group has discussed extensively about the ethical aspects of AI, and, specifically, benchmarking. It provided a list of issues facing users of AI in dentistry:

- Prudence
- Equity
- Autonomy
- Privacy and intimacy
- Responsibility
- Democratic participation
- Solidarity
- Data diversity
- Data protection
- Well-being
- Development
- Patient–clinician relationship
- Law and governance

The Merriam-Webster Dictionary defines wellness as "the quality or state of being in good health, especially as an actively sought goal". In that way, hardware and software breakthroughs, especially in mobile and wearable devices have been combined with computational advances in AI to scale wellness coaching and automate promotion of health. Moreover, these technologies are now tightly converging with the rapid development of powerful AI systems. AI systems are now a tool for supplementing clinicians' decision-making, for providing customized and tailored health management plans, for predicting the next health crisis, and for designing personalized treatments. Still, how AI algorithms can improve wellness assessment, aid in personalizing intervention strategies to promote healthier lifestyle behaviours and uncover previously unknown disease risk factors is still under investigation. Moreover, AI may impact organizational wellness, for instance, safety, performance and employee wellness. As AI can facilitate work processes, organizations will have the opportunity to focus more on activities they consider more vital.

One of the most important issues is data diversity. Data diversity is often necessary in order to increase the diversity of AI technologies. For minimizing and identifying potential biases, software developers need to include a broad set of experts in the design and development process who know about bias, contexts and regulations, as well as consultation with stakeholders, data labelling and testing. Moreover, AI algorithms are often developed on non-representative samples evaluating one ethnic or socioeconomic group and a lack of diversity in training data could have serious implications on decision-making. There are generally four different types of diversity: internal (ethnicity, age, nationality, gender, cultural identity), external (education, socioeconomic status or religious beliefs), organizational (employment, financial status) and worldview (political or moral beliefs). If an AI algorithm fails to account for this diverse data, it can perpetuate human biases that could leave out underrepresented populations or put them at a health risk.

Also, patients must give informed consent for the harvesting of their data. Next, this confidential data should be deidentified and protected from the bodies collecting them and from cyberattacks. Furthermore, data collection should meet local ethical, regulatory and legal standards. In addition, data storage should be decentralized with FL schemes applied.

Equity in dentistry means providing equal access and quality of care to everyone, regardless of gender, age, ethnicity, social, cultural or economic factors. When applying AI technology in dental practice, three aspects of equity should be considered: algorithmic fairness, availability of AI technology to all interested groups, and using AI to improve health equity in public and population health. Algorithmic fairness addresses bias in data collection and algorithm design, availability ensures equal access to AI technology, and using AI for public and population health can mitigate human bias and improve outcomes.

Future regulations need to close the gap in the current laws around AI as more products are being developed. Recent advances in deep learning allow for the rapid screening of diversity of data and possible discrimination by race, sex, ethnicity or other parameters. Assessing the diversity using the deep neural networks and eliminating the human factor may lead to development of unbiased AI technology. Ensuring diversity in algorithms' data is a necessity in order AI technology to broadly benefit diverse populations from around the world. Otherwise, AI could lead to exacerbation of existing health disparities, existing within one country or between developed and developing countries.

To increase transparency and participation, AI can be designed with open-source software or source code available to the public. Open-source software offers the benefit of users contributing and providing feedback, allowing them to understand how the system works, to identify potential problems and to extend and adapt it. Open-source software must be accessible and transparent as well as engaging. The subgroup is providing a checklist for ethical challenges for study design and clinical use for AI applications in dentistry. Furthermore, the subgroup is working on a study regarding using AI tools for smile design and its ethical challenges.

4 Existing work on benchmarking

This section focuses on the existing benchmarking processes in the context of AI and dental diagnostics and digital dentistry for quality assessment. It addresses different aspects of the existing work on benchmarking of AI systems (e.g., relevant scientific publications, benchmarking frameworks, scores and metrics and clinical evaluation attempts). The goal was to collect all relevant learnings from previous benchmarking that could help to implement the benchmarking process in this topic group. Notably, the TG’s activities so far have not reached full maturity towards establishing a benchmarking dataset for one or more dental AI tasks; this is planned for further activities of the group.

4.1 Existing work on benchmarking

To evaluate clinical performance, metrics are required to estimate the effect of the application at the patient (or, for caries, on tooth or surface) level and its effect (harm, benefit, costs) at the societal level, according to the framework described by Fryback and Thornbury (1991) [8]. These refer to changes in the patient's diagnosis (% of diagnostic changes) and at the therapeutic level (% of changes in treatment decisions). Relevant metrics include the percentage of patients who improve with the use of AI compared to without AI, morbidity or procedures avoided with the use of AI, and cost per effect or utility gained or lost with the AI-test information. The TG-Dental has generated an initial set of high-quality annotated dataset of approximately 350 radiographs to be released on a public platform to be defined (e.g., EvalAI, Grand-challenge, Kaggle). This dataset will be divided into training and validation and testing on a test set. One part will be retained for the independent and developer-blinded evaluation of AI models. Annotator information, detailed diagnostic criteria and demographic metrics of the dataset composition will be provided. Some recommendations for benchmark reporting of AI models for the imaging diagnosis of dental caries are shown in Table 2.

Table 2 – Benchmark reporting of AI Models: Possible metrics for clinical performance and societal impact

Area	Description	Examples
Annotators	Detailed description of the annotators and how they perform the annotations	Number Expertise Diagnostic criteria Disagreement resolution
Annotations	Ideally, annotations should be made with the ITU/WHO annotation tool or an open access tool	https://github.com/FG-AI4H/annotation-tool
Dataset fairness		Description of ethnicity, gender, marital status, age, education, insurance type.
Compliance with FAIR principles		Compliance with the core FAIR object assessment metrics [9]
Tasks definition	Classification Detection Segmentation	Presence of a pathology pixel-wise annotation
Metrics oriented to technical diagnostic efficiency	Usual diagnostic performance metrics	Sensitivity (Recall) Specificity Accuracy Jaccard index Dice's coefficient Positive predicted value (precision) Negative predicted values F1 score Area-under-the-ROC curve (if several diagnostic thresholds are being tested)
	Usual diagnostic reliability metrics	Discrete values: Kappa (normal, weighted, Fleiss) Continuous values: ICC
Metric oriented to diagnostic clinical performance	Diagnostic thinking efficacy metrics	Number (percentage) of cases in a series in which image judged 'helpful' to making the diagnosis.

**Table 2 – Benchmark reporting of AI Models:
Possible metrics for clinical performance and societal impact**

Area	Description	Examples
		Difference in clinicians' subjectively estimated diagnosis probabilities pre- to post-test information. Empirical subjective log-likelihood ratio for caries positive and negative in a case series.
	Therapeutic efficacy metrics	Number (percentage) of times image judged helpful in planning management of the patient in a case series. Percentage of times procedure avoided due to image information. Number or percentage of times therapy planned pre-test changed after the image information was obtained (retrospectively inferred from clinical records). Number of percentage of times clinicians' prospectively stated therapeutic choices changed after test information.
Patient-level diagnostic efficacy	Patient outcome efficacy	Percentage of patients with pathology with test compared with/without test. Morbidity (or invasive procedures) avoided when using AI. Change in quality-adjusted life expectancy. Expected value of AI in quality-adjusted life years (QALYs). Cost per QALY saved with AI.
Societal level diagnostic efficacy	Societal efficacy	Benefit — cost analysis from societal perspective. Cost-effectiveness analysis from societal perspective. Decrease in population morbidity attributable to the test (WHO goal –10%) Increase in population health coverage attributable to the test (WHO goal +75%)

Ideally, a large, heterogeneous dataset of a clearly defined population should be available and annotated transparently, with an accepted criterion, ideally against a ground truth or consensus (gold standard of an appropriate number of experts).

To benchmark **technical diagnostic metrics**, the characteristics of the reports should adhere to the recommendations of Schwendicke et al [10], with an emphasis on:

1 The characteristics of the dataset used for training and testing, ideally available publicly;

2 The demographic characteristics of the persons who provided their information for the generation of the training and testing dataset;

3 The method of validation, ideally external and independent;

4 The diagnostic criteria used for annotation;

5 The characteristics and training of the annotators;

6 The annotation disagreement resolution methods.

For general caries diagnostic AI-based systems, the STARD-AI reporting guideline is recommended [11].

To benchmark the **clinical performance of AI models**, clinical studies should adhere to the recommendations of Schwendicke et al. (2021) and the methodological items required by the CONSORT-AI reporting guideline [12]. For any clinical claim, it is required to show how the use of AI systems allows changing the diagnostic or therapeutic decision at the patient level. Additionally, it is necessary to report any potential adverse effects of the clinical use of an AI system, including potential overdiagnosis and overtreatment.

Available economic analyses [13] [AH12] [FS13] show that the use of AI affects the cost of treatment. Further explorations are needed to generalize these findings and to determine health economic implications of AI-based on their evaluation on a benchmarking dataset.

4.2 Publications on benchmarking systems

Currently, there are no high-quality annotated datasets publicly available for imaging diagnosis in dentistry. Ultimately, any benchmarking of AI models in dentistry should be measured according to how much they help achieve the WHO goal of (1) reducing oral disease morbidity by 10% and (2) increasing health coverage to 75% of the population.

In 2015 a competition sponsored by the IEEE International Symposium on Biomedical Imaging was held. One of the datasets contained 120 annotated bitewing intra-oral radiographs, divided into sets of 40 for training, 40 for testing and 40 for on-site testing within the competition, with annotations for dental caries. The dataset is no longer available. The caries-related annotators corresponded to lesions at the enamel and dentin levels separately. The information available about the annotators was that they were two experienced clinicians. The metrics evaluated were sensitivity, specificity and F-score. Of nine registered investigator groups, two submitted models Ronnenberg et al. [18] and Lee et al. [19], are detailed in Table 3 for caries detection.

Table 3 – Performance evaluation of caries detection models on bitewing intra-oral radiographs*

Author	Ronnenberg et al.	Lee et al.
Model	u-net	random forest
Implementation	Caffe-Framework (C++)	Java
CPU and OS	Nvidia-Titan, i7, 32GB Ram, Nvidia GTX980 with 8GB RAM	2xIntel Xeon E5-2650, 128GB RAM for training and 2xIntel Xeon E5-2687, 16GB RAM
Processing time per image	1.5 s	150 s
Test 1		
Precision	0.073	0.022
Sensitivity	0.120	0.060
Specificity	0.998	0.989
F-Score	0.119	0.042
Test 2		
Precision	0.078	0.032
Sensitivity	0.086	0.050
Specificity	0.999	0.991
F-Score	0.131	0.061

* Modified from Wang et al. (2016) [19]. A benchmark for comparison of dental radiography analysis algorithms. *Med. Image Anal. 31*, 63–76.

Notably, these low performances differ from what is reported by a range of studies (see above). At the same time, it is not clear why that is given those datasets (not necessarily model architectures, U-net is widely used for segmentation of caries lesions) are no longer available.

Similar datasets of cephalometrics have been used in the past and employed for evaluating AI models for landmark detection. Again, the datasets are not available any longer, nor are they representative and annotated with a quality assurance.

Bibliography

[1] Schneider et al. (2023) Federated vs Local vs Central Deep Learning of Tooth Segmentation on Panoramic Radiographs, *Journal of Dentistry*, Vol. 135;104556.

[2] Mohammad-Rahimi H, Motamedian SR, Rohban MH, Krois J, Uribe SE, Mahmoudinia E, et al. (2022), Deep Learning for Caries Detection: A systematic review, *Journal of Dentistry*, Vol. 122:104115

[3] Mine et al. (2020) Applying Deep Artificial Neural Network Approach to Maxillofacial Prostheses Coloration, Journal of Prosthodontic Research, Vol. 64, No. 3, pp. 296–300.

[4] Cui et al. (2020), Clinical Decision Support Model for Tooth Extraction Therapy Derived from Electronic Dental Records, The Journal of Prosthetic Dentistry. Vol. July.

[5] Takahashi et al. (2021), A System for Designing Removable Partial Dentures Using Artificial Intelligence. Part 1. Classification of Partially Edentulous Arches Using a Convolutional Neural Network, Journal of Prosthodontic Research, Vol. 65, No. 1.

[6] Yamaguchi et al. (2019), Predicting the Debonding of CAD/CAM Composite Resin Crowns with AI, Journal of Dental Research, Vol. 98, No. 11.

[7] World Health Organization, WHO Classification of Tumours Online. https://tumourclassification.iarc.who.int/welcome/

[8] Fryback DG, Thornbury JR (1991), The Efficacy of Diagnostic Imaging, Medical Decision Making, Vol. 11, pp. 88–94.

[9] Devaraju et al. (2022), FAIRsFAIR Data Object Assessment Metrics (0.4), Zenodo, Vol. 12;4081213.

[10] Schwendicke F, Singh T, Lee J-H, Gaudin R, Chaurasia A, Wiegand T, et al. (2021), Artificial Intelligence in Dental Research: Checklist for Authors, Reviewers, Readers, *Journal of Dentistry*, Vol. 107:103610.

[11] Sounderajah et al. (2021), Developing a Reporting Guideline for Artificial Intelligence-Centred Diagnostic Test Accuracy Studies: The STARD-AI Protocol, BMJ Open, Vol. 11, No. 6.

[12] Liu et al. (2020), Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension, National Library of Medicine, Vol. 26, No. 9.

[13] JMIR Med Inform. 2022;10:e33703; *Journal of Dentistry*, 2022;119:104080; JAMA Netw Open. 2022;5:e220269; J Dent Res. 2020;:22034520972335

[14] FGAI4H DEL1, Focus Group on AI for Health Deliverable 1 (2022), *Ethics and governance of artificial intelligence for health*.

[15] NAEINI, E. N. et al. Narrative review regarding the applicability, accuracy, and clinical outcome of flapless implant surgery with or without computer guidance. Clinical Implant Dentistry and Related Research, v. 22, n. 4, p. 454–467, 13 May 2020.

[16] Krois, J., Garcia Cantu, A., Chaurasia, A. *et al.* Generalizability of deep learning models for dental image analysis. *Sci Rep* **11**, 6102 (2021). https://doi.org/10.1038/s41598-021-85454-5

[17] Sultan, Ahmed & Elgharib, Mohamed & Tavares, Tiffany & Jessri, Maryam. (2020). The Use of Artificial Intelligence and Deep Machine Learning in Oncologic Histopathology. Journal of Oral Pathology & Medicine. 49. 10.1111/jop.13042.

[18] Ronneberger, O., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI).accepted, url = http://arxiv.org/abs/1505.04597

[19] Wang, Ching-Wei & Huang, Cheng-Ta & Lee, Jia-Hong & Li, Chung-Hsing & Chang, Sheng-Wei & Siao, Ming-Jhih & Lai, Tat-Ming & Ibragimov, Bulat & Vrtovec, Tomaž & Ronneberger, Olaf & Fischer, Philipp & Cootes, Tim & Lindner, Claudia. (2016). A benchmark for comparison of dental radiography analysis algorithms. Medical Image Analysis. 31. 10.1016/j.media.2016.02.004.

Annex A

Glossary

This section lists all the relevant abbreviations, acronyms and uncommon terms used in the document.

Abbreviation/term	Expansion	Comment
AI	Artificial intelligence	
AI4H	Artificial Intelligence for Health	
AI-MD	AI-based Medical Device	
API	Application Programming Interface	
CfTGP	Call for Topic Group Participation	
DEL	Deliverable	
FDA	Food and Drug Administration	
FGAI4H	Focus Group on AI for Health	
GDP	Gross domestic product	
GDPR	General Data Protection Regulation	
IMDRF	International Medical Device Regulators Forum	
IP	Intellectual Property	
ISO	International Organization for Standardization	
ITU	International Telecommunication Union	
LMIC	Low-and Middle-Income Countries	
MCBS	Magnitude of Clinical Benefit Scale	
MDR	Medical Device Regulation	
PII	Personally Identifiable Information	
SaMD	Software as a Medical Device	
TDD	Topic Description Document	Document specifying the standardized benchmarking for a topic on which the FG AI4H Topic Group works. This document is the TDD for the Topic Group Dental Diagnostics and Digital Dentistry (TG-Dental)
TG	Topic Group	
WG	Working Group	
WHO	World Health Organization	

Annex B

Declarations of conflicts of interest

In accordance with the ITU transparency rules, the following conflict of interest is declared. JK and FS, both authors of this document, are cofounders of dentalXrai, a spin-off of Charité Universitätsmedizin Berlin focusing on dental radiograph analysis using AI.