Journal of Dentistry 148 (2024) 105228

Contents lists available at ScienceDirect

Journal of Dentistry

journal homepage: www.elsevier.com/locate/jdent

External validation of an artificial intelligence-based method for the detection and classification of molar incisor hypomineralisation in dental photographs

Julia Neumayr [a], Elisabeth Frenkel [a], Julia Schwarzmaier [a], Nour Ammar [a], Andreas Kessler [a,b], Falk Schwendicke [a], Jan Kühnisch [a,*], Helena Dujic [a]

[a] *Department of Conservative Dentistry and Periodontology, LMU University Hospital, Klinikum der Ludwig-Maximilians-Universität München, Klinik für Zahnerhaltung und Parodontologie, LMU, Goethestraße 70, Munich 80336, Germany*

[b] *Department of Prosthetic Dentistry, Faculty of Medicine, Center for Dental Medicine, Medical Center-University of Freiburg, University of Freiburg, Freiburg, Germany*

ARTICLE INFO

Keywords:
Enamel hypomineralisation
Diagnosis
Validation study
Artificial intelligence

ABSTRACT

Objectives: This ex vivo diagnostic study aimed to externally validate an open-access artificial intelligence (AI)-based model for the detection, classification, localisation and segmentation of enamel/molar incisor hypomineralisation (EH/MIH).

Methods: An independent sample of web images showing teeth with (n = 277) and without (n = 178) EH/MIH was evaluated by a workgroup of dentists whose consensus served as the reference standard. Then, an AI-based model was used for the detection of EH/MIH, followed by automated classification and segmentation of the findings (test method). The accuracy (ACC), sensitivity (SE), specificity (SP) and area under the curve (AUC) were determined. Furthermore, the correctness of EH/MIH lesion localisation and segmentation was evaluated.

Results: An overall ACC of 94.3 % was achieved for image-based detection of EH/MIH. Cross-classification of the AI-based class prediction and the reference standard resulted in an agreement of 89.2 % for all diagnostic decisions (n = 594), with an ACC between 91.4 % and 97.8 %. The corresponding SE and SP values ranged from 81.7 % to 92.8 % and 91.9 % to 98.7 %, respectively. The AUC varied between 0.894 and 0.945. Image size had only a limited impact on diagnostic performance. The AI-based model correctly predicted EH/MIH localisation in 97.3 % of cases. For the detected lesions, segmentation was fully correct in 63.4 % of all cases and partially correct in 33.9 %.

Conclusions: This study documented the promising diagnostic performance of an open-access AI tool in the detection and classification of EH/MIH in external images.

Clinical significance: Externally validated AI-based diagnostic methods could facilitate the detection of EH/MIH lesions in dental photographs.

1. Introduction

Molar incisor hypomineralisation (MIH) is considered a multifactorial condition, meaning it is possibly genetically determined and is most likely influenced by environmental factors [1]. The global prevalence of MIH was estimated by Schwendicke et al. [2] and Zhao et al. [3] to be approximately 14 %, which indicates that it is a prevalent condition in young populations. The main clinical findings in teeth with MIH are demarcated opacity, enamel breakdown or atypical restoration, which may further be associated with hypersensitivity of the affected teeth [4]. As a result, there are not only functional limitations regarding occlusal function or maintaining oral hygiene for those affected [4], but also difficulties in anesthetizing the affected teeth prior to dental treatment. EH/MIH findings can be detected and classified by visual examination predominantly in the first permanent molars and incisors [4–6]. However, other deciduous and permanent teeth may also be affected, although with a lower prevalence [7]. The correct diagnostic assessment of enamel hypomineralisation (EH), including MIH, remains a challenge in daily dental practice because other hard tissue defects, e.g., caries or fluorosis, have similar appearances [8–10]. Therefore, independent

* Corresponding author.
E-mail address: Jan.Kuehnisch@med.uni-muenchen.de (J. Kühnisch).

https://doi.org/10.1016/j.jdent.2024.105228
Received 27 May 2024; Received in revised form 29 June 2024; Accepted 4 July 2024
Available online 5 July 2024

diagnostic methods that can differentiate between EH/MIH and other hard tissue lesions should be welcomed by the dental community [8].

In recent years, most of the published diagnostic models based on artificial intelligence (AI) methods have focused on the detection of caries [11], and there are only a few publications on the automated detection of EH/MIH in clinical photographs of teeth [8,12,13]. Here, the AI-based model that was recently introduced by Felsch et al. [12] enables simultaneous detection, classification, localisation and segmentation of caries and EH/MIH. Felsch et al. [12] achieved promising accuracy in the detection of EH/MIH classes and the segmentation and localisation of specific findings using an AI-based model on their own test data. These authors made this AI-based model openly accessible as a web application. Therefore, external validation is possible, and the generalisability of the model can be investigated.

This ex vivo diagnostic study aimed to validate the diagnostic performance of this model in the detection, classification, localisation and segmentation of EH/MIH on independent image samples. It was hypothesised that the external validity would be similar to the previously published internal validity [12].

2. Materials & methods

The study was performed following the guidelines of the Standards for Reporting of Diagnostic Accuracy Studies (STARD) Steering Committee [14], and the recommendations for standardising the planning, conduct and publication of dental studies using AI methods were considered [15].

2.1. Collection of dental images from the web

To independently validate the AI-based model (http://demo.dental-ai.de) that was recently published by Felsch et al. [12], it was necessary to use images that were not involved in the development of the model. To fulfil this requirement, no images from our own image databases were used; rather, images of teeth with or without EH/MIH were obtained from freely accessible internet sources. For the web-based search the following keywords were entered into a web browser (www.google.com): molar-incisor hypomineralisation, MIH, enamel hypomineralisation, chalky teeth, and developmental disorder in teeth. A total of 455 images were identified and included in the image sample. Of these, 277 images showed teeth with EH/MIH on occlusal and/or smooth surfaces, and 178 images showed teeth without EH/MIH. The included images showed an image resolution of at least 72 pixel per inch with the image size ranging between 12 kB and 5100 kB. Blurred or distorted images were excluded. Images showing carious lesions, fluorosis or rare dental diseases such as amelogenesis imperfecta or dentinogenesis imperfecta were not considered in this investigation. All of the images that were found were saved for later use together with information on the image features. It is important to note that no image was altered through the use of an editing software.

2.2. EH/MIH detection and classification by the dental workgroup (reference standard)

Discussion of the findings, classification and concordance of all included images ($n = 455$) occurred simultaneously within the workgroup consisting of one principal investigator (JK; >20 years of clinical experience including paediatric dentistry) and four general dentists (JN, EF, JS, HD; ≤2 years of clinical experience). All images were projected on a television screen. First, the images were checked for the presence of EH/MIH. Teeth with EH/MIH were classified based on the EAPD criteria for MIH detection by visual assessment [4–6]. The degree of EH/MIH was classified as "demarcated opacity" if the change in enamel translucency was clearly visible, the estimated size was >1 mm and the opacity colour was white, creamy, yellow or brown [4]. "Enamel breakdown" was assigned to images showing loss of the naturally formed tooth surface, resulting in exposed dentin and porous, hypomineralised remaining enamel [4]. The EH/MIH class "atypical restoration" was given to those images in which the size and shape of the shown restoration corresponded more to the EH/MIH extent on molars and incisors and not to the typical caries pattern [4]. The classes included in the AI-based model were also based on the EAPD criteria [4–6]. When an image showed several classes of findings, each pathology was documented separately. Furthermore, the image quality was evaluated based on the criteria "good" and "acceptable". This subjective assessment of the image quality was made considering lighting, resolution and angle of the images. For example, minimal motion blur, slight lighting deficiencies that did not impair the visibility of the relevant surfaces or a slightly unfavourable angle were considered acceptable. The absence of the above-mentioned defects was accepted as good. When there were differing opinions on a finding or image quality, the corresponding image was discussed within the workgroup (JN, EF, JS, HD, JK) until a consensus was reached. These visual evaluations of all images by the dentists subsequently served as the basis for the reference standard. The detection and classification of the image-related reference standard always occurred independently of the AI-based evaluation.

2.3. EH/MIH detection, classification, localisation and segmentation by the AI-based model (test method)

The AI-based model that was used (http://demo.dental-ai.de) offers the possibility of detecting, classifying, localising and segmenting EH/MIH [12]. The detection is determined by the ability of the tested AI-based model to recognize a finding at pixel level for the analysed image. The detection is automatically associated with a classification, also at pixel level, which assigns the detected findings to the following categories: demarcated opacity, enamel breakdown and atypical restoration [4–6]. Healthy tooth surfaces are not listed separately or marked by the AI-based model.

The AI-based evaluation of all the included dental images was performed as follows. First, each image ($n = 455$) was uploaded individually to the above-mentioned AI-based model. The area of interest was centred, and irrelevant parts at the image margins were cropped out. Second, following the image analysis, the AI-based findings visualized at pixel level were saved as screenshots for later evaluation.

Dental evaluations of the AI-based analysis occurred several weeks after the initial consensus on the reference standard and was also performed within the workgroup (JN, EF, JS, HD, JK) to achieve consensus and exclude individual misclassifications. The accuracy of the AI-based evaluation in terms of detection, classification, localisation and segmentation were assessed, discussed, and documented. Here, AI-based detection and classification were considered and documented according to previously described procedures, followed by an assessment of AI-based localisation and segmentation. For localisation to be considered correct, it was crucial that at least one pixel of the marked segment was located in the corresponding EH/MIH lesion. Conversely, the localisation was considered incorrect if no pixel was located within the corresponding EH/MIH class. The dentists assessed the quality of segmentation by comparing the segment predicted at pixel level with the actual EH/MIH lesion. A distinction was made between fully (the predicted area was >90 % identical to the actual extent of the EH/MIH lesion), partially (the predicted area was <90 % identical to the actual extent of the EH/MIH lesion) or not matched. The percentage agreement of the segment with the actual lesion area could only be estimated or given as a category.

2.4. Training and calibration

Prior to the study, an experienced principal investigator (JK) conducted a training on detecting and classifying EH/MIH for all participating investigators (JN, EF, JS, HD). The training included details about the study design and discussion of numerous images showing EH/MIH, caries and potential differential diagnoses. Each examiner then assessed sixty images to assess the presence of EH/MIH on occlusal and smooth surfaces and repeated the procedure two weeks later to determine intra-/inter-examiner reliability. The calculated kappa values (JN: 0.846/0.492–0.755, EF: 0.853/0.411–0.721, JS: 0.814/0.466–0.872, HD: 0.870/0.428–0.893, JK: 1.000/0.464–0.956) showed a moderate to nearly perfect agreement [16].

2.5. Data management and statistical analysis

Using the freely available data processing system Epidata Manager, a digital data sheet was created, and in this sheet, all the image information and diagnostic decisions for the reference standard and the test method were recorded using the Epidata Entry Client (EpiData version v4.6.0.6, EpiData Association, Odense, Denmark, http://www.epidata.dk). The findings were subsequently exported and validated using an Excel spreadsheet (Excel 2019, Microsoft, Redmond, WA, USA). Data analysis was performed using Python (version 3.8.5, http://www.python.org). To determine the diagnostic performance, the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN), positive and negative predictive values (PPV and NPV, respectively), sensitivity (SE), specificity (SP), accuracy (ACC) and area under the receiver operating characteristic (ROC) curve (AUC) were calculated [17].

3. Results

The diagnostic performance of the AI-based image analysis (test method) was compared to the dental workgroup consensus (reference standard). In the first step, the image-related detection of EH/MIH was determined, and this detection had a diagnostic accuracy of 94.3 %. The SE and SP were 94.4 % and 94.2 %, respectively (Table 1). In the next step, all the recorded EH/MIH decisions (n = 594) were analysed from the entire image sample (n = 455). Demarcated opacity (n = 262) and enamel breakdown (n = 115) were found most frequently in the image sample (Table 2). The cross-classification of the diagnostic results from the AI-based prediction (test method) and the reference standard (Table 2) resulted in a match for 89.2 % of all diagnostic decisions (n = 530/594). False-positive findings occurred in 31 cases (5.2 %) and false-negative findings occurred in 33 cases (5.6 %) (Table 2). The diagnostic performance of the test method in relation to each EH/MIH class is shown in Table 3. The overall diagnostic accuracy ranged from 91.4 % (demarcated opacity) to 97.8 % (atypical restoration). The SE ranged from 81.7 % (enamel breakdown) to 92.8 % (no EH/MIH), and the SP ranged from 91.9 % (demarcated opacity) to 98.7 % (atypical restoration). The AUC ranged from 0.894 (enamel breakdown) to 0.945 (no EH/MIH). The corresponding receiver operating curves are shown in Fig. 1.

The correlation between diagnostic performance and image size was further evaluated (Table 4). The image size only slightly influenced the diagnostic performance for the detection and classification of EH/MIH, and no serious differences were documented. In addition to assessment of EH/MIH detection and classification, the evaluation included considerations of EH/MIH localisation and segmentation (Table 5). In principle, the AI-based model correctly predicted EH/MIH localisation in 97.3 % of cases (n = 399). The segmentation was fully correct in 63.4 % of all cases and was partially correct in 33.9 % (Table 5). No localisation or segmentation of the EH/MIH lesions on the dental images was possible in only 2.7 % of all cases.

Table 1
Diagnostic performance of the AI-based model (test method) in the detection of EH/MIH compared to the assessment of the dental workgroup (reference standard). The analysis was image-based (n = 455).

EH/MIH Detection		Visual evaluation (Reference standard)			
		No EH/MIH	EH/MIH	∑	
AI-based evaluation (Test method)	No EH/MIH	168	16	184	NPV= 91.3 %
	EH/MIH	10	261	271	PPV= 96.3 %
	∑	178	277	455	
		SP= 94.4 %	SE= 94.2 %		ACC= 94.3 %

4. Discussion

The present ex vivo diagnostic study validated a recently introduced AI-based method for the detection, classification, localisation and segmentation of EH/MIH [12] on an independent sample of digital images of teeth. The overall image-based diagnostic accuracy was markedly high at 94.3 % (Table 1), with a low proportion of false-positive findings (Table 2). The diagnostic accuracy for all EH/MIH classes ranged from 91.4 % (demarcated opacity) to 97.8 % (atypical restoration) and was rated high (Table 3). In addition, the AI-based model correctly localised most EH/MIH lesions (97.3 %) (Table 5).

The discussion on the documented results of the external validity of this new AI-based method is limited because no comparable external validation studies of AI-based models for EH/MIH detection or classification are available, and the number of AI-based models for the automated detection of EH/MIH is low [8,12,13]. A comparison between the internal [12] and external validity of the AI-based model revealed slightly lower values (ACC 94.3 %, Table 1) for the overall diagnostic accuracy of EH/MIH detection compared to the internal validation (ACC 97.8 %). Therefore, our initial hypothesis that the internal and external validity would be identical must be rejected. However, the results of the external validation are promising because the differences were limited. Although the included dental images were heterogeneous in image size, this did not appear to significantly affect diagnostic performance (Table 4). Specifically, the diagnostic performance achieved with small (<100 kB) images was comparable to that achieved with larger (>300 kB) images. Here, the ACC values for the detection of different EH/MIH classes in images <100 kB ranged between 93.5 % and 97.8 %, and the ACC values for images >300 kB ranged between 94.1 % and 97.9 %. Overall, the documented diagnostic performance (Tables 1–4, Fig. 1) supports the quality of the AI-based model for EH/MIH detection because previously unknown images were correctly assessed with an accuracy of 94.3 %.

We found slightly lower or comparable quality parameters for diagnostic performance compared to other studies on EH/MIH detection. Alevizakos et al. [8] achieved accuracies ranging from 84.0 % to 92.9 % using different convolutional neural networks. Schönewolf et al. [13] also presented an AI-based model and reported an overall accuracy of 95.2 %. To the best of our knowledge, these are currently the only data on the AI-based detection of MIH in clinical images of teeth. Moreover, the validity of visual EH/MIH diagnostics, compared to caries diagnostics, has not been investigated. Therefore, further comparisons of the validity of clinical EH/MIH diagnostics are currently not possible. Notably, Amarante et al. [7] and Restrepo et al. [10] published data on intra- and inter-examiner reliability in the diagnosis of EH/MIH on clinical images of teeth with EH/MIH. Both studies considered educational effects in the calibration of students and dentists in the field of EH/MIH diagnostics.

An analysis of the results of EH/MIH classification from the external validation (Tables 2 and 3) compared to the internal validation data, as presented by Felsch et al. [12], revealed a slight decrease in diagnostic performance in the external validation. This result was particularly evident for the most frequently found "demarcated opacity" (Table 2). Here, an ACC of 91.4 %, SE of 90.8 % and SP of 91.9 % (Table 3) were observed, but Felsch et al. [12] reported an ACC of 96.9 %, SE of 94.0 %

Table 2
AI-based image evaluation (test method) in relation to visual evaluation (reference standard) cross-tabulated for all EH/MIH diagnoses (n = 594) from all images (n = 455); multiple findings per image were possible.

		Visual evaluation (Reference standard)				
	EH/MIH Classification	No EH/MIH	Demarcated opacity	Enamel breakdown	Atypical restoration	∑
AI-based evaluation (Test method)	No EH/MIH	168	10	1	5	184
	Demarcated opacity	10	238	16	1	265
	Enamel breakdown	3	11	94	0	108
	Atypical restoration	0	3	4	30	37
	∑	181	262	115	36	594

Table 3
Summary of the diagnostic performance in the classification of all EH/MIH findings (n = 594) from all images (n = 455) by the AI-based model.

EH/MIH Classification	True Positives		True Negatives		False Positives		False Negatives		Diagnostic performance (%)					
	n	%	n	%	n	%	n	%	ACC	SE	SP	PPV	NPV	AUC
No EH/MIH	168	28.3	397	66.8	16	2.7	13	2.2	95.1	92.8	96.1	91.3	96.8	0.945
Demarcated opacity	238	40.1	305	51.3	27	4.5	24	4.0	91.4	90.8	91.9	89.8	92.7	0.914
Enamel breakdown	94	15.8	465	78.3	14	2.4	21	3.5	94.1	81.7	97.1	87.0	95.7	0.894
Atypical restoration	30	5.1	551	92.8	7	1.2	6	1.0	97.8	83.3	98.7	81.1	98.9	0.910

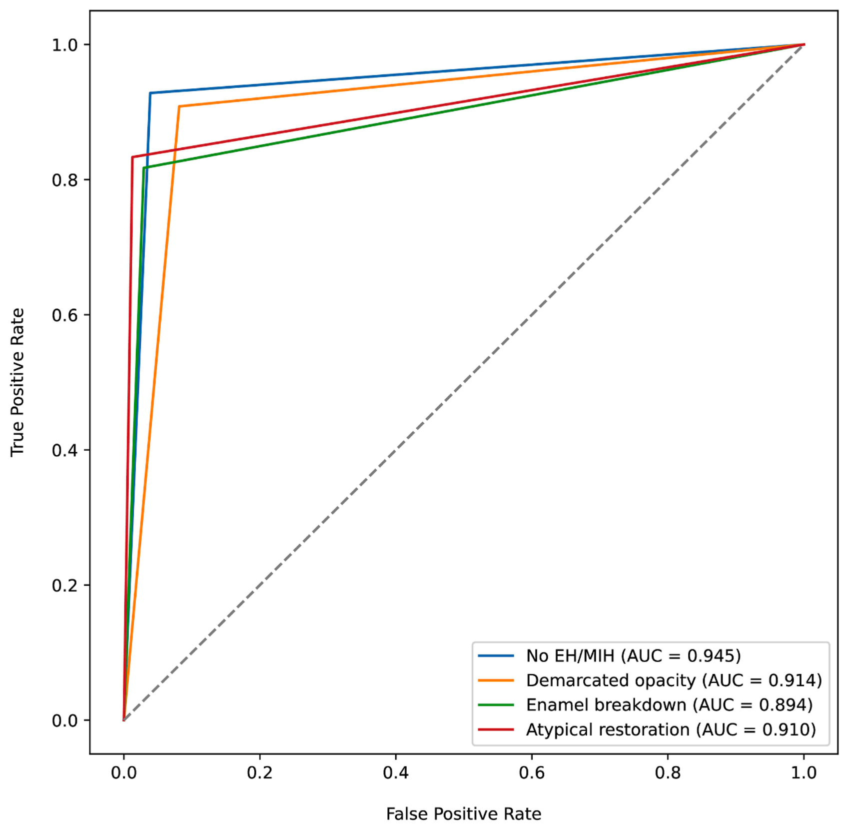

Fig. 1. Receiver operating characteristic (ROC) curves with the respective areas under the curve (AUCs) for all EH/MIH findings.

and SP of 97.6 %. The values for the clinically relevant diagnosis of "enamel breakdown" were an ACC of 94.1 %, SE of 81.7 % and SP of 97.1 %; in contrast, Felsch et al. [12] reported an ACC of 98.8 %, SE of 80.5 % and SP of 99.6 %. The diagnosis of "atypical restoration", for which very good performance was achieved in the internal and external validation, also achieved slightly lower results in the external validation, with an ACC of 97.8 %, SE of 83.3 % and SP of 98.7 % compared to values for the ACC, SE and SE of 99.5 %, 90.5 % and 99.9 %, respectively, in Felsch et al. [12]. A closer look at the available results of EH/MIH classification (Table 3) revealed that the accuracy values for "enamel breakdown" and "atypical restoration" were based on very high SP values of 97.1 % and 98.7 %, respectively, but slightly lower SE values of 81.7 % and 83.3 %, respectively. Felsch et al. [12] obtained similar SE and SP values for these categories, but the SE and SP values for the "atypical restoration" class did not diverge as much as in the external validation. This difference suggests that the AI-based model was better at recognising EH/MIH-free surfaces than EH/MIH-affected tooth surfaces in the external and internal validation. These values describe distinct uncertainties of the AI-based model in the correct classification of EH/MIH findings. This difference is likely related to the known and previously discussed problem [12,18] that EH/MIH findings were unevenly distributed in the available image sample, and rarer or more variable

Table 4
Summary of the diagnostic performance in the classification of EH/MIH findings by the AI-based model in relation to image size.

EH/MIH classification	Image size (kB)	True Positives (n)	True Negatives (n)	False Positives (n)	False Negatives (n)	Diagnostic performance (%)					
						ACC	SE	SP	PPV	NPV	AUC
No EH/MIH	<100	22	156	5	3	95.7	88.0	96.9	81.5	98.1	0.924
	100–300	47	158	9	7	92.8	87.0	94.6	83.9	95.8	0.908
	>300	99	83	2	3	97.3	97.1	97.6	98.0	96.5	0.974
Demarcated opacity	<100	108	66	7	5	93.5	95.6	90.4	93.9	93.0	0.930
	100–300	88	105	12	16	87.3	84.6	89.7	88.0	86.8	0.872
	>300	42	134	8	3	94.1	93.3	94.4	84.0	97.8	0.938
Enamel breakdown	<100	31	147	3	5	95.7	86.1	98.0	91.2	96.7	0.921
	100–300	40	160	11	10	90.5	80.0	93.6	78.4	94.1	0.868
	>300	23	158	0	6	96.8	79.3	100.0	100.0	96.3	0.897
Atypical restoration	<100	9	173	1	3	97.8	75.0	99.4	90.0	98.3	0.872
	100–300	11	205	3	2	97.7	84.6	98.6	78.6	99.0	0.916
	>300	10	173	3	1	97.9	90.9	98.3	76.9	99.4	0.946

Table 5
Overview of the correctness of localisation and segmentation by the AI-based model for EH/MIH findings ($n = 410$).

EH/MIH Classification	Localisation		Segmentation		
	Incorrect	Correct	Incorrect	Partially	Fully
Demarcated opacity	7	258	7	75	183
Enamel breakdown	3	105	3	47	58
Atypical restoration	1	36	1	17	19
Σ	11	399	11	139	260
	410*		410*		

* 184 segments were classified by the AI-based model as "No EH/MIH" and localisation and segmentation could not be assessed.

findings may be recognised with less accuracy. This finding highlights the need to continuously improve developed models and, in particular, to increase the number of images of rare EH/MIH findings.

As part of the external validation, the correctness of the localisation and segmentation of the AI-based model was evaluated. The recorded localisation was classified as correct in 97.3 % of all cases. In contrast, segmentation was assessed as fully correct in less than two-thirds of cases (63.4 %). Conversely, the AI-based model achieved only partially satisfactory segmentation of the existing EH/MIH lesions in 33.9 % of cases. However, these assessments and figures should be viewed critically for several reasons. (1) They are subjective estimates of the participating dentists. (2) There are no current recommendations for the qualitative and quantitative assessment of segmented areas. (3) The selected cut-off of ~90 % of the lesion area may be considered strict. Therefore, these values should not be overinterpreted at this time. To date, no studies have evaluated localisation and segmentation in the context of external validation. Therefore, further studies are needed to develop methodological standards for the assessment of segmentation.

The present study has strengths and limitations. As a strength, it should be emphasised that for the first time an AI-based model for automated detection of EH/MIH in dental photographs at pixel level was subjected to external validation. However, the following points should be discussed as limitations. To create an independent image database, images from the internet were used, which was ultimately associated with heterogeneous image quality. While the assessment of image quality includes several aspects such as resolution, lighting and sharpness, this study only documented subjective assessments in terms of "acceptable" or "good", considering the above-mentioned factors. Merely the image size was recorded from the image metadata, since this information could be consistently obtained. At present, the absence of generally accepted and applicable imaging criteria potentially limits the replication of such a study design. Considering that more AI-based models may be available for external validation in the future, a consensus on the implementation of such study types and imaging standards would be beneficial. Moreover, images of occlusal and smooth tooth surfaces were considered in this study. The focus on a specific tooth surface may result in EH/MIH findings on other tooth surfaces not being optimally displayed and thus not being recognised. Similarly, different imaging angles and lighting conditions during imaging may also affect the image quality and thus the visualization of EH/MIH findings. Nevertheless, it should be mentioned that the workgroup had no influence on the photographic technique or focus for web-based images. In addition, the ability of the AI-based model to analyse multiple findings simultaneously at a pixel level proved to be useful, as just a few pixels gave an indication of the possible presence of a finding. By evaluating the localisation and segmentation, we were able to further assess the quality of this analysis. However, it remains unclear whether images of the same tooth with different angulations would lead to divergent findings by the AI-based model. This can be seen as a further limitation and should be considered in future studies. Given that the AI-based model provided pixelwise diagnoses for each image and that in diagnostic studies usually only one diagnostic decision is made per tooth surface and method, the prediction of multiple findings per surface and/or tooth required a different evaluation approach. Therefore, a diagnostic decision for each image was derived first (Table 1), followed by an analysis of the segments in terms of classification (Tables 2–4), localisation and segmentation (Table 5). With this in mind, the decisions were made via group consensus because it was not possible to completely analyse the segments independently. Furthermore, the EH/MIH findings segmented by the AI-based model could not be blinded, which resulted in potential verification bias. A further limitation is that the current model is unable to recognise other structural disorders, such as fluorosis or amelogenesis imperfecta, which makes a comprehensive diagnostic approach impossible. These limitations highlight the need to further develop future AI-based models to recognise other developmental disorders. The integration of this technology into intraoral cameras and scanners or smartphones appears feasible and could contribute to the effective early detection of dental diseases. However, AI-based diagnostics have not been implemented in dental practice. The benefits of AI-based models and ethical issues have also not yet been conclusively clarified, which currently limits their use in dentistry [19].

5. Conclusion

The present ex vivo diagnostic study validated a recently introduced AI-based method and documented its promising diagnostic performance in EH/MIH detection, classification, localisation and segmentation using an external sample of digital photographs of teeth. The diagnostic parameters were slightly lower than the earlier published internal validation data. Future studies are recommended to investigate the diagnostic validity and practicability of these findings in different settings.

CRediT authorship contribution statement

Julia Neumayr: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation. **Elisabeth Frenkel:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Data curation. **Julia Schwarzmaier:** Writing – review & editing, Validation. **Nour Ammar:** Writing – review & editing, Formal analysis. **Andreas Kessler:** Writing – review & editing, Methodology, Conceptualization. **Falk Schwendicke:** Writing – review & editing. **Jan Kühnisch:** Writing – original draft, Writing – review & editing, Methodology, Formal analysis, Supervision, Project administration, Conceptualization. **Helena Dujic:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A.R. Vieira, D.J. Manton, On the variable clinical presentation of molar-incisor hypomineralization, Caries Res. 53 (4) (2019) 482–488.
[2] F. Schwendicke, K. Elhennawy, S. Reda, K. Bekes, D.J. Manton, J. Krois, Global burden of molar incisor hypomineralization, J. Dent. 68 (2018) 10–18.
[3] D. Zhao, B. Dong, D. Yu, Q. Ren, Y. Sun, The prevalence of molar incisor hypomineralization: evidence from 70 studies, Int. J. Paediatr. Dent. 28 (2) (2018) 170–179.
[4] N.A. Lygidakis, E. Garot, C. Somani, G.D. Taylor, P. Rouas, F.S.L. Wong, Best clinical practice guidance for clinicians dealing with children presenting with molar-incisor-hypomineralisation (MIH): an updated European Academy of Paediatric Dentistry policy document, Eur. Arch. Paediatr. Dent. 23 (1) (2022) 3–21.
[5] N.A. Lygidakis, F. Wong, B. Jälevik, A.M. Vierrou, S. Alaluusua, I. Espelid, Best clinical practice guidance for clinicians dealing with children presenting with molar-incisor-hypomineralisation (MIH): an EAPD policy document, Eur. Arch. Paediatr. Dent. 11 (2) (2010) 75–81.
[6] K.L. Weerheijm, I. Mejàre, Molar incisor hypomineralization: a questionnaire inventory of its occurrence in member countries of the European Academy of Paediatric Dentistry (EAPD), Int. J. Paediatr. Dent. 13 (6) (2003) 411–416.
[7] B.C. Amarante, L.Y. Arima, E. Pinheiro, P. Carvalho, E. Michel-Crosato, M. Bönecker, Diagnosis training and calibration for epidemiological studies on primary and permanent teeth with hypomineralization, Eur. Arch. Paediatr. Dent. 23 (1) (2022) 169–177.
[8] V. Alevizakos, K. Bekes, R. Steffen, C. von See, Artificial intelligence system for training diagnosis and differentiation with molar incisor hypomineralization (MIH) and similar pathologies, Clin. Oral Investig. 26 (12) (2022) 6917–6923.
[9] A. Gunay, Knowledge and attitudes of a group of dental students in Turkey about molar incisor hypomineralization, Med. Sci. Monit. 29 (2023) e941824.
[10] M. Restrepo, D.F. Rojas-Gualdron, A.L. de Farias, A. Escobar, L.F. Velez, D. G. Bussaneli, L. Santos-Pinto, Development of undergraduate students' diagnostic accuracy for the classification of molar incisor hypomineralization, Eur. J. Dent. Educ. 28 (1) (2024) 154–160.
[11] M. Moharrami, J. Farmer, S. Singhal, E. Watson, M. Glogauer, A.E.W. Johnson, F. Schwendicke, C. Quinonez, Detecting dental caries on oral photographs using artificial intelligence: a systematic review, Oral Dis. (2023).
[12] M. Felsch, O. Meyer, A. Schlickenrieder, P. Engels, J. Schönewolf, F. Zöllner, R. Heinrich-Weltzien, M. Hesenius, R. Hickel, V. Gruhn, J. Kühnisch, Detection and localization of caries and hypomineralization on dental photographs with a vision transformer model, NPJ Digit. Med. 6 (1) (2023) 198.
[13] J. Schonewolf, O. Meyer, P. Engels, A. Schlickenrieder, R. Hickel, V. Gruhn, M. Hesenius, J. Kühnisch, Artificial intelligence-based diagnostics of molar-incisor-hypomineralization (MIH) on intraoral photographs, Clin. Oral Investig. 26 (9) (2022) 5923–5930.
[14] P.M. Bossuyt, J.B. Reitsma, D.E. Bruns, C.A. Gatsonis, P.P. Glasziou, L. Irwig, J. G. Lijmer, D. Moher, D. Rennie, H.C. de Vet, H.Y. Kressel, N. Rifai, R.M. Golub, D. G. Altman, L. Hooft, D.A. Korevaar, J.F. Cohen, S. Group, STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies, BMJ 351 (2015) h5527.
[15] F. Schwendicke, T. Singh, J.H. Lee, R. Gaudin, A. Chaurasia, T. Wiegand, S. Uribe, J. Krois, IADR e-oral health network and the ITU WHO focus group AI for Health, Artificial intelligence in dental research: checklist for authors, reviewers, readers, J. Dent. 107 (2021) 103610.
[16] J.L. Fleiss, J. Cohen, The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability, Educ. Psychol. Meas. 33 (3) (1973) 613–619.
[17] D.E. Matthews, V.T. Farewell, Using and understanding medical statistics, S.Karger AG, 2015.
[18] J. Kühnisch, O. Meyer, M. Hesenius, R. Hickel, V. Gruhn, Caries detection on intraoral images using artificial intelligence, J. Dent. Res. 101 (2) (2022) 158–165.
[19] F. Schwendicke, W. Samek, J. Krois, Artificial intelligence in dentistry: chances and challenges, J. Dent. Res. 99 (7) (2020) 769–774.