Evaluation of AI model for cephalometric landmark classification (TG Dental)

Tanne Johannes (✉ tannejohannes@gmail.com)
Johannes Tanne Consulting

Chaurasia Akhilanand
King George's Medical University

Krois Joachim
Charité-Universitätsmedizin Berlin

Vinayahalingam Shankeeth
Radboud University Medical Centre (Radboudumc)

Haiat Anahita
ITU/WHO Focus Group AI on Health

Schwendicke Falk
Charité-Universitätsmedizin Berlin

Saeed Reza Motamedian
Shahid Beheshti University of Medical Sciences

Behnaz Mohammad
Shahid Beheshti University of Medical Sciences

Mohammad-Rahimi Hossein
Shahid Beheshti University of Medical Sciences

Research Article

Keywords: Cephalometric landmarks, AI-Audit, Regulation

Posted Date: October 26th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2147359/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Evaluation of AI model for cephalometric landmark classification (TG Dental)

Tanne Johannes\textsuperscript{1*}, Chaurasia Akhilanand\textsuperscript{2†}, Krois Joachim\textsuperscript{3†}, Vinayahalingam Shankeeth\textsuperscript{4†}, Haiat Anahita\textsuperscript{5†}, Schwendicke Falk\textsuperscript{6†}, Motamedian Saeed Reza\textsuperscript{7†}, Behnaz Mohammad\textsuperscript{8†} and Mohammad-Rahimi Hossein\textsuperscript{9†}

\textsuperscript{1*}Regulatory Affairs, Johannes Tanne Consulting, Waldstraße 73A, Königs Wusterhausen, 15711, Brandenburg, Germany.
\textsuperscript{2}Oral Medicine and Radiology, King George’s Medical University, Shahmina Road, Lucknow, 226003, Uttar-Pradesh, India.
\textsuperscript{3}Oral Diagnostics, Digital Health and Health Services Research, Charité-Universitätsmedizin Berlin, Aßmannshausener Straße 4-6, Berlin, 14197, Berlin, Germany.
\textsuperscript{4}Oral and Maxillofacial surgery, Radboud University Medical Centre (Radboudumc), Geert Grooteplein Zuid 10, Nijmegen, 6525 GA, Nijmegen, Netherlands.
\textsuperscript{5}Topic Group Dental Diagnostics and Digital Dentistry, ITU/WHO Focus Group AI on Health, Berlin, Berlin, Germany.
\textsuperscript{6}Oral Diagnostics, Digital Health and Health Services Research, Charité-Universitätsmedizin Berlin, Aßmannshausener Straße 4-6, Berlin, 14197, Berlin, Germany.
\textsuperscript{7}Department of Orthodontics, Dentofacial Deformities Research Center, Research Institute of Dental Sciences, and School of Dentistry, Shahid Beheshti University of Medical Sciences, Daneshjoo Blvd, Evin, Shahid Chamran Highway, Tehran, 1983963113, Tehran, Iran.
\textsuperscript{8}Department of Orthodontics, Dentofacial Deformities Research Center, Research Institute of Dental Sciences, and School of Dentistry, Shahid Beheshti University of Medical Sciences, Daneshjoo Blvd, Evin, Shahid Chamran Highway, Tehran, 1983963113, Tehran, Iran.
Abstract

**Purpose:** The accuracy of cephalometric landmark identification for malocclusion classification is essential for the diagnosis and treatment planning. The identification of these landmarks is often complex and time consuming for the orthodontists. An AI model for the classification was recently developed. Due to the strict regulation on software systems but few content on AI requirements in this publication the model was investigated under consideration of current regulatory considerations.

**Methods:** The AI platform of the ITU/WHO is used to allow the assessment of the models of the application. An Audit was performed assessing the development process with regard to medical device regulations, data protection regulation and ethical consideration. Upon that the major task during the development were evaluated such as qualification, annotation procedure and data set attributes.

**Results:** The AI models were investigated under consideration of technical, clinical, regulatory and ethical consideration. The risk to the health of the patient and user can be considered as low according to the IMDRF definition.

**Conclusion:** The application is useful to aid the decision and treatment planning for malocclusion classification on lateral cephalograms without cephalometric landmarks. It is comparable with common standards in orthodontic diagnosis.

**Keywords:** Cephalometric landmarks, AI-Audit, Regulation
1 Introduction

It is an essential part of orthodontic diagnosis and treatment planning to classify malocclusions based on their skeletal and dental components. A imaging technique commonly used for malocclusion classification is the lateral cephalogram. Through the use of landmarks, cephalometric analysis can quantify the vertical and anteroposterior position of the jaw. Hence, landmark identification and analysis have a decisive impact on diagnosis and treatment planning [6], [5]. ANB angle is one of the most important cephalometric analyses evaluating sagittal skeletal relationship. This angle is between point A, point N (nasion), and point B on a lateral cephalogram. There is a close relationship between the ANB angle and the occlusal relationship of the teeth. Accordingly, patients may be classified into skeletal classes I, II, and III, which affect treatment planning [4]. It is challenging to identify cephalometric landmarks in lateral cephalograms for a variety of reasons. Anatomical variations, radiographic distortions, and pathological cases inhibit clinicians from making accurate interpretations. Furthermore, manual or semi-manual landmark plotting is time-consuming and prone to inconsistencies [2][1][3]. With an automated assistance system, time constraints and interprofessional variability may be reduced, allowing less experienced professionals to assess lateral cephalograms more accurately and reliably. In this study, we aim to classify sagittal skeletal without cephalometric landmarks using an end-to-end approach.

2 Use Case

The skeletal and dental classification of malocclusion is an essential part of orthodontic/orthognathic diagnosis and treatment planning. One commonly performed imaging technique for malocclusion classification is the lateral cephalogram. Cephalometric landmarks are used to quantify the vertical and anteroposterior positions of the jaw. Therefore, the accuracy of landmark identification and analysis has a decisive impact on the diagnosis and treatment planning. The identification of cephalometric landmarks in lateral cephalograms is challenging for numerous reasons. The lack of three dimensionality, radiographic distortion, anatomical variations and pathological cases impede an accurate interpretation by clinicians. Furthermore, manual or semi-manual landmark plotting are prone to inconsistencies. An automated assistance system may reduce the time-constraints and interrater variability, allowing a more reliable and accurate assessment of lateral cephalograms, especially in the hands of less experienced professionals. Thus, this study aims for an end-to-end malocclusion classification on lateral cephalograms without cephalometric landmarks.

2.1 Data collection and preparation

A total of 800 lateral cephalograms (Fig. 1) gathered from all the patients visited oral and maxillofacial radiology department of Shahid Beheshti University
ML4H Auditing

of Medical Sciences from January 2016 to December 2019. A ProMax Dimax 3 Digital Pan/Ceph device was used to acquire all of the images (Planmeca, Helsinki, Finland). In the next step, the images were exported as JPEGs with a resolution of $2,143 \times 2,300$ pixels. Finally, based on inclusion and exclusion criteria, 614 lateral cephalograms were selected for the purpose of this study. Exclusion criteria were as follows:

1. Non-standard images (including incorrect head positions or with signs of patient movement during radiography).
2. Low-quality images (blurry or noisy images).
3. Difficulties in locating any of A, N, and B landmarks.
4. If there was more than one image from a patient, one of them were selected randomly.

Data annotation procedure was done using LabelMe tool (MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, Massachusetts, USA)[1]. Three landmarks were selected for the purpose of this study including A, N, B. The definition of each landmark is presented in Table 1.
<table>
<thead>
<tr>
<th>#</th>
<th>Landmark</th>
<th>Abb</th>
<th>Definition and Guide</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A (Subspinale)</td>
<td>A</td>
<td>The deepest (most posterior) midline point on the curvature between the ANS and prosthion.</td>
</tr>
<tr>
<td>2</td>
<td>Nasion</td>
<td>N</td>
<td>The intersection of the internasal and frontonasal sutures, in the midsagittal plane.</td>
</tr>
<tr>
<td>3</td>
<td>B(Supermentale)</td>
<td>B</td>
<td>The deepest (most posterior) midline point on the bony curvature of the anterior mandible, between infradentale and pogonion.</td>
</tr>
</tbody>
</table>

Table 1  Classes

Then, for the purpose of Angle classification, the angle between these three landmarks were calculated. Then based on the mean and standard deviations of the data, various classes were defined as follows:

1. Class I: ANB angle from 2.11° to 5.00° (mean ± 0.5 standard deviation).
2. Class II: ANB angle more than 5.00°.
3. Class III: ANB angle less than 2.11°.

For the annotation, first, a calibration session was held with three orthodontists for reaching consensus in defining landmarks. Two orthodontist annotated the landmarks, independently. Then, the mean of their X- and Y-coordinates were calculated and exported to new JSON files. Then, these outcomes were double-checked by the third orthodontist, which had more experience. Other than the main dataset, we create another version by removing borderline samples. For this purpose sample with ANB angle between 2.11° to 2.40° and 4.70° to 5.00° were removed. In this version, a total of 577 lateral cephalograms were included.

3 Planning and Scope of the Audit

The audit is performed with the help of a predefined audit verification checklist. Different phases during the model development were considered starting from planning to data collection and preparation to model training, tuning, evaluation and maintenance. It has to be mentioned here that the most important consideration is focus on the data labeling and the model training since these steps are the crucial parts for the evaluation of the AI.

4 Audit Methods

The audit is performed with the help of an AI Audit platform from the ITU/WHO Focus Group on artificial Intelligence for Health. The platform was used to assess the models on different levels of observation (technical, clinical, regulatory) based on quantitative or qualitative methods.

5 Audit Report

The audit report include the descriptive assessment of the model.

Technical
The development was done on a front-end web application. For the audit, the data collection was placed on a GitHub repository (https://github.com/aiaudit-org/trial-audits-team-i-tg-dental/tree/main/data). Data split of 60:20:20 (train:validate:test) is a valid ratio for partitioning the dataset. The model algorithm of CNN (ResNet-101, pretrained on ImageNet) was used, which is acceptable for this kind of application.

Clinical
The model uses the sagittal skeletal classification for orthodontic (pediatric) patient images (based on A-point-Nasion-point-B-point (ANB) angle). All lateral cephalometric images taken from the described time period were included, except those which were a localization for relevant landmarks were not possible. For data bias and variance minimization, histogram equalization was used. It has to be mentioned here that this caused a worsened model output.

Regulatory
The model was evaluated under the regulatory aspects. The application can be defined as a medical device software which is used as service for diagnostic purpose. The software belongs to low risk category according to IMDRF document about risk categorization of SaMD. Intended users are orthodontists, general dentists and dental professionals. The combination of case randomization, definition of inclusion and exclusion criteria and choosing a variability of field data is intended to control for any bias effect and any specific findings. With that we reduce the effect of borderline cases. The labeling was done by two professional orthodontists with a open source annotation tool (LabelMe). A sanity check of the annotations was done by a review of the coordinates of the landmarks by a third professional orthodontist. Also the annotation tool was not validated the quality on the labeling process is ensured by the evaluation of qualified experts. Risk management was done on a simplified FMEA (Failure mode and effect analysis). We consider the risk related to the diagnostic and treatment planning by false positive and false negative as the highest risk. This leads two hazardous cases. The first consider that the user could act on a false detection (error of commission) or does not act on a missing detection (error of omission). In our case misclassification could be caused by missing specifications of the training data. One example is the potential effect of (foreseeable) variances due to limited consideration during training (e.g. hardware information). Due to the nature of dental radiographs a anonymization of the images is not possible. Not additional data from the patients were collected. All samples were de-identified as much as possible.

Ethical
Ethics committee of the Shahid Beheshti University of Medical Sciences approved the initial study (IRB number: IR.SBMU.DRC.REC.1400.007). No exemptions were obtained. Data were anonymized so that the de-identified data could be used for health care operations in such a way that they were suitable to train the algorithm.
6 Discussion

The tool is useful to aid the decision and treatment planning of the intended user group. It is useful for diagnostics, especially for malocclusion classification on lateral cephalograms without cephalometric landmarks. The model is used for the angle classification in lateral cephalogram for sagittal skeletal relationships, a common standard in orthodontic diagnosis. We consider the risk related to the diagnostic and treatment planning by false positive and false negative as the highest risk. This leads to two hazardous cases. The first considers that the user could act on a false detection (error of commission) or does not act on a missing detection (error of omission). In our case, misclassification could be caused by missing specifications of the training data. One example is the potential effect of (foreseeable) variances due to limited consideration during training (e.g., hardware information such as filters). There is no patient safety severity impact in case of an erroneous outcome. The software is used as a supportive tool, not as a replacement for the orthodontic decision. Even if the orthodontist decides on the ML outcome. The outcome is not related to a pathological indication. The model allows the accelerated classification due to its supportive diagnosis. This allows the speed up the decision by the orthodontist. Currently, the model is developed on a limited dataset. Potential bias can affect the result, such as the limit of hardware settings (filters, machines) or the missing consideration of borderline cases. Validation on real-world data, including patient groups with different demographic backgrounds from the location where the ML is used is currently not applied. The current version of the model is developed on a single-centered data set. It is recommended to include the mentioned borderline cases in the dataset. Include hardware specifications such as X-Ray model, version and filters applied. A sufficient front-end application for efficient usability evaluation should be applied.

7 Actionable Recommendations

It is recommended to include the mentioned borderline cases for the data set. Include hardware specifications such as X-Ray model, version and filters applied. A sufficient front-end application for efficient usability evaluation should be applied.

8 References

Supplementary information

See appendix.
Declarations

8.1 Ethical Approval

No ethical approval was necessary for this kind of investigation. The ethics committee of the Shahid Beheshti University of Medical Sciences approved the initial study (IRB number: IR.SBMU.DRC.REC.1400.007).

8.2 Competing interests

The authors declare not conflict of interest.

8.3 Authors’ contributions

Behnaz Mohammad, Saeed Reza Motamedian conceived of the presented medical background and developed the performed the computations. Falk Schwendicke and Akhilanand Chaurasia performed the clinical evaluation. Shankeeth Vinayahalingam and Joachim Krois evaluated the technical aspects of the model. Anahita Haiat also contributed to the technical evaluation and supported in the configuration. Johannes Tanne did the regulatory assessment, supported in the configuration. Hossein Mohammad-Rahimi did the ethical assessment, shared the data and did the supervision of the publication. All authors discussed the results and contributed to the final manuscript.

8.4 Funding

This publication was not financially supported.

Appendix A Appendix

https://health.aiaudit.org/web/challenges/challenge-page/383/overview

References


