Artificial intelligence-assisted assessment for incidental findings on lateral cephalograms

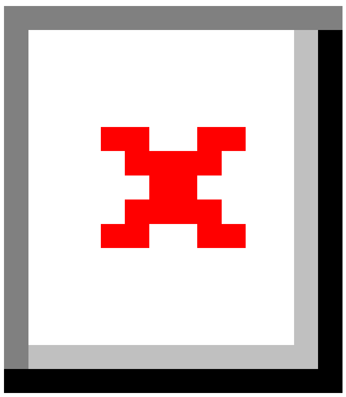
2022 Biomedical Research Awards (BRA)

Dr. Bingshuang Zou

drzou@dentistry.ubc.ca
O: 604-822-4158

FollowUp Form

Award Information



In an attempt to make things a little easier for the reviewer who will read this report, please consider these two questions before this is sent for review:

- *Is this an example of your very best work, in that it provides sufficient explanation and justification, and is something otherwise worthy of publication? (We do publish the Final Report on our website, so this does need to be complete and polished.)*
- *Does this Final Report provide the level of detail, etc. that you would expect, if you were the reviewer?*

Title of Project:*
Artificial intelligence-assisted assessment for incidental findings on lateral cephalograms

Award Type
Biomedical Research Award (BRA)

Period of AAOF Support

July 1, 2022 through June 30, 2023

Institution

Faculty of Dentistry, University of British Columbia

Names of principal advisor(s) / mentor(s), co-investigator(s) and consultant(s)

Edwin Yen, Yankai Cao, David MacDonald

Amount of Funding

$4,000.00

Abstract

(add specific directions for each type here)

Artificial intelligence (AI) and machine learning (ML) have integrated with many aspects of modern society already. Orthodontic researchers and practitioners must join this wave to steer and advance our specialty. AI has been widely utilized in orthodontic diagnosis and treatment planning, automated landmark detection and cephalometric analyses, assessment of growth and development, and evaluation of treatment outcomes in recent years.1-7

Lateral cephalometric radiographs (LCs) have long been taken as a standard and routine in orthodontic patients for assessment and prediction of craniofacial growth, orthodontic diagnosis, and oromaxillofacial treatment planning through cephalometric analysis. LCs also contain information beyond the maxillofacial complex such as cranium, cervical spine etc. Craniofacial diseases or anatomical variations without any symptoms of discomfort or deformity may be detected unexpectedly. Thus, our overall objectives are to build a fully automatic cephalometric analysis system with landmark detection and cephalometric analyses, establish a AI-based model to detect any incidental findings on LCs, and expand the application even wider onto 3D images, as CBCT and 3D facial images, with the aim to reduce the cost, time and need for human knowledge, and the number of human errors.

To fulfill the objectives, we have successfully developed a fully automatic system of cephalometric analysis, including cephalometric landmark detection and cephalometric measurement by using a machine learning technique of multiresolution decision tree regression voting. Experimental results show that the performance of our proposed method is satisfactory for landmark detection and measurement analysis in LCs.8

The next step, also indicated as the specific aim of this project, is to build an AI-based model to detect any incidental findings on LCs; such vital information as Os odontoideum can easily be overlooked without proper knowledge of this region, and might lead to potential lethality even with minor trauma. In this proposal, we plan to answer the research question that if AI and machine learning-based model will show better performance than classical human detection ability in incidental findings on LCs. The hypothesis to be testes is there is no difference between AI and human in performance of incidental finding detection.

The overall project's clinical implications or significance are: 1. to relieve orthodontists from the tedious works including landmark identification, cephalometric tracing, and radiograph reading or interpreting with

the aid of AI and machine learning, 2. to reduce the cost, time, the need for human knowledge, and the number of human error, and 3. to increase the quality of life among orthodontic patients.

Respond to the following questions:

Detailed results and inferences:*

If the work has been published, please attach a pdf of manuscript below by clicking "Upload a file".
OR
Use the text box below to describe in detail the results of your study. The intent is to share the knowledge you have generated with the AAOF and orthodontic community specifically and other who may benefit from your study. Table, Figures, Statistical Analysis, and interpretation of results should also be attached by clicking "Upload a file".

Final report.pdf
This retrospective study was approved by the ethics board at the University of British Columbia, Vancouver, Canada (H22-01253).

Step 1

To train the AI model, a consecutive of 4514 LCs was selected from UBC Graduate Orthodontic Clinic and screened. Each LC was divided into 3 zones and examined in a systematic way: cranium; neck and cervical spine; and dentofacial complex (Figure 1). Incidental findings were mainly categorized into 7 groups: 1. Sellar bridging; 2. Ponticulus posticus; 3. Occipital spur; 4. Vertebral fusion; 5. Enlarged parietal foramen; 6. Antrolith/Tonsilolith/carotid artery calcification; 7. Stylohyoid ligament calcification. One researcher did all the inspections and repeated them after 4 weeks in a subsample of 100 to test intra-examiner reliability. The examiner was also validated by a radiologist using this subsample. Finally, we labeled them by different categories: 3717 samples, 166 samples, 411 samples, 172 samples, 26 samples, 4 samples, 15 samples, 3 samples for Group1-NSF, Group2-Bridging, Group3-PontPost, Group4-OccSpur, Group5-VertFusion, Group6-EnlParFora, Group7-AtlOccAssiCalc, and Group8-SHLCal, respectively.

Step 2

Since the distribution of the incidental findings in each group were imbalanced, especially, the images of the last 4 groups are seriously insufficient for effective training of an AI model. As a perfect data distribution, the numbers of samples belonging to different classes are expected to be almost the same. In this way, an AI model would be able to fully exploit the characteristics of these 4 groups with fewer samples and avoid biased predictions.

Alleviating this imbalance problem, we discarded the last four classes and randomly selected 250 samples from Group 1 and the other 3 groups to train our algorithm. In other words, we apply 250+166+411+172 samples to train a 4-classes classification model. The accuracy is 65-70%, which means our model has a chance of 65-70% to correctly predict the label of a given LC.

The two most common challenges in the application of intelligent image processing systems involve the limited number of annotated samples and the less discriminative medical images. In addition, the problem in the application of cephalometric diagnosis is more serious. These two difficulties entangle and form a combined problem: how to effectively transfer knowledge gained on a set of very few examples with less rich representations, assumed to be available in large quantities, to another set of unseen samples.

We propose a deep neural network to handle the challenge. Compared with the common practice that only applies a single feature, we employ a deep feature extractor that learns to convert an input image into sets of discriminative feature vectors that can be easily and efficiently transferred to unseen images. Concretely, the extractor decomposes the representation into a couple of independent components, allowing the capture of several distinctive aspects and properties of images. Furthermore, to intensively explore the information of limited images, we fully utilize the potential intra-class similarity and inter-class difference among different categories of images. As a result, a series of set-to-set metrics are proposed to quantify the distance between samples and effectively guide the training process of the feature extractor.

Were the original, specific aims of the proposal realized?*

Yes, we have partially realized the original aim: to build a fully automatic cephalometric analysis system with landmark detection and cephalometric analyses, establish a AI-based model to detect any incidental findings on LCs, and expand the application even wider onto 3D images, as CBCT and 3D facial images, with the aim to reduce the cost, time and need for human knowledge, and the number of human errors.
For the specific aim, to build an AI-based model to detect any incidental findings on LCs (such vital information as Os odontoideum can easily be overlooked without proper knowledge of this region, and might lead to potential lethality even with minor trauma), we have also partially realized.
In the future study, we will try to increase the sample size and balance the case numbers in each category to train a more efficient AI-model and expand this application even further to CBCTs.

Were the results published?*

No

Have the results of this proposal been presented?*

No

To what extent have you used, or how do you intend to use, AAOF funding to further your career?*

We have used the funding in full to purchase the software license and attend relevant conferences.

Accounting: Were there any leftover funds?

$0.00

Not Published

Are there plans to publish? If not, why not?*

Yes. We plan to publish our study once we rebuild the AI model with an increased sample size.

Not Presented

Are there plans to present? If not, why not?*

Yes. We will present in the relevant conference once we rebuild the AI model with an increased sample size.

Internal Review

Reviewer comments

Reviewer Status*

File Attachment Summary

Applicant File Uploads

- Final report.pdf

Artificial intelligence-assisted assessment for incidental findings on lateral cephalograms

Introduction

Artificial intelligence (AI) and machine learning (ML) have integrated with many aspects of modern society already. Orthodontic researchers and practitioners must join this wave to steer and advance our specialty. AI has been widely utilized in orthodontic diagnosis and treatment planning, automated landmark detection and cephalometric analyses, assessment of growth and development, and evaluation of treatment outcomes in recent years.

Lateral cephalometric radiographs (LCs) have long been taken as a standard and routine in orthodontic patients to assess and predict craniofacial growth, orthodontic diagnosis, and oral-maxillofacial treatment planning through cephalometric analysis. LCs also contain information beyond the maxillofacial complex such as the cranium, cervical spine etc. Craniofacial diseases or anatomical variations without any symptoms of discomfort or deformity may be detected unexpectedly. Thus, this study aims to establish a fully automatic AI-based model to detect any incidental findings on LCs to reduce the cost, time and need for human knowledge and the number of human errors.

Steps of studies

This retrospective study was approved by the ethics board at the University of British Columbia, Vancouver, Canada (H22-01253).

Step 1

To train the AI model, a consecutive of 4514 LCs was selected from UBC Graduate Orthodontic Clinic and screened. Each LC was divided into 3 zones and examined in a systematic way: cranium; neck and cervical spine; and dentofacial complex (Figure 1). Incidental findings were mainly categorized into 7 groups: 1. Sellar bridging; 2. Ponticulus posticus; 3. Occipital spur; 4. Vertebral fusion; 5. Enlarged parietal foramen; 6. Antrolith/Tonsilolith/carotid artery calcification; 7. Stylohyoid ligament calcification. One researcher did all the inspections and repeated them after 4 weeks in a subsample of 100 to test intra-examiner reliability. The examiner was also validated by a radiologist using this subsample. Finally, we labeled them by different categories: **3717** samples, **166** samples, **411** samples, **172** samples, **26** samples, **4** samples, **15** samples, **3** samples for Group1-NSF, Group2-Bridging, Group3-PontPost, Group4-OccSpur, Group5-VertFusion, Group6-EnlParFora, Group7-AtlOccAssiCalc, and Group8-SHLCal, respectively.

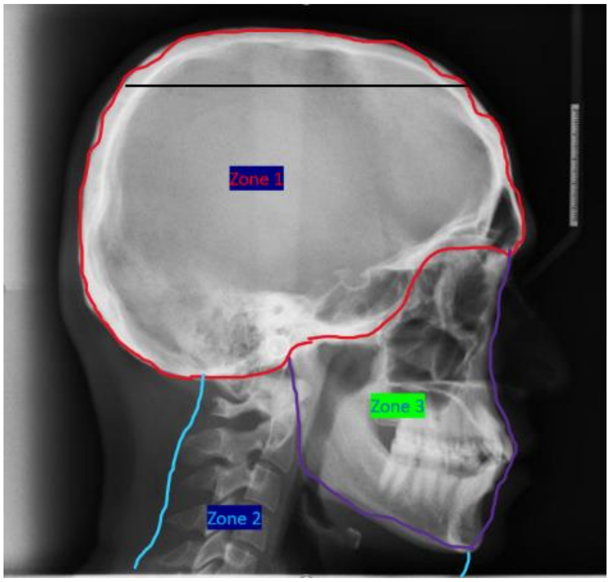


Figure 1: Zones of reading LC

Step 2

Since the distribution of the incidental findings in each group were imbalanced, especially, the images of the last 4 groups are seriously insufficient for effective training of an AI model. As a perfect data distribution, the numbers of samples belonging to different classes are expected to be almost the same. In this way, an AI model would be able to fully exploit the characteristics of these 4 groups with fewer samples and avoid biased predictions.

Alleviating this imbalance problem, we discarded the last four classes and randomly selected 250 samples from Group 1 and the other 3 groups to train our algorithm. In other words, we apply 250+166+411+172 samples to train a 4-classes classification model. The accuracy is 65-70%, which means our model has a chance of 65-70% to correctly predict the label of a given LC.

The two most common challenges in the application of intelligent image processing systems involve the limited number of annotated samples and the less discriminative medical images. In addition, the problem in the application of cephalometric diagnosis is more serious. These two difficulties entangle and form a combined problem: how to effectively transfer knowledge gained on a set of very few examples with less rich representations, assumed to be available in large quantities, to another set of unseen samples.

We propose a deep neural network to handle the challenge. Compared with the common practice that only applies a single feature, we employ a deep feature extractor that learns to convert an input image into sets of discriminative feature vectors that can be easily and efficiently transferred to unseen images. Concretely, the extractor decomposes the representation into a couple of independent components, allowing the capture of several distinctive aspects and properties of images. Furthermore, to intensively explore the information of limited images, we fully utilize the potential intra-class similarity and inter-class difference among different categories of images. As a result, a series of set-to-set metrics are proposed to quantify the distance between samples and effectively guide the training process of the feature extractor.

Primary conclusion

Benefiting from the design, the neural network is qualified to detect distinct characteristics of medical images at different scales and offer diagnosis, even though the given training samples are not sufficient.

Future direction

We will increase the sample size and try to make the distribution of different categories more even. A similar deep neural network will be used to train the AI model, which will be tested with another set of LCs.