Assessment of Automatic Cephalometric Landmarks Identification Using Artificial Intelligence

BY

GALINA V. BULATOVA
D.M.D., First Moscow Medical University, 2008

THESIS

Submitted as a partial fulfillment of the requirements
for the degree of Master of Science in Oral Sciences
in the Graduate College of the
University of Illinois at Chicago, 2021

Chicago, Illinois

Defense committee:

Flavio Jose Castelli Sanchez, Chair and advisor
Budi Kusnoto, Orthodontics
Maria Grace Costa Viana, Orthodontics
T. Peter Tsay, Orthodontics
David M. Avenetti, Pediatric Dentistry

To my mother, Nataliya Bulatova, for her endless love and efforts to give her children best education despite all difficulties she had to go through.

AKNOWLEDGMENTS

I would like to thank my committee member for their support and guidance. I am very grateful to Dr. Sanchez for the bright idea of the project itself and all the encouragement along the way. I would like to thank Dr. Kusnoto for his wisdom and creative solution to many difficult obstacles we met throughout the process. I am very grateful to Mrs. Viana for her tremendous work and many hours dedicates to this project. Her reassuring support and kind words inspired me to keep working and never give up. I would like to acknowledge Dr. Tsay for being one of the greatest teachers in the world and laying the foundation of orthodontic discipline in my mind and in my heart. Words are not enough to say how grateful I am for his knowledge and feedback regarding this project, that helped me to detect some issues, that I was able to correct. In addition, I would like to thank Dr. Avenetti for his participation in this project and for bringing his perspective.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE PAGE

LIST OF ABBREVIATIONS

AAO	American Association of Orthodontists
MT	Manual Tracing
AI	Artificial Intelligence
CNN	Convolutional Neural Networks
YOLOv3	“You Look Only Once” 3d generation algorithm

SUMMARY

Cephalometrics is an essential part of orthodontic diagnosis and treatment planning. However, difficulties in identification of several cephalometric landmarks, as well as questionable reliability, make the cephalometric analysis a time-consuming process. To make research of huge datasets affordable and to simplify the routine use of cephalometric analysis in clinical practice Artificial Intelligence (AI) software, which allows automatic point identification, gained a vast application in recent five years.

The current study aims to compare the accuracy of cephalometric landmark identification between AI Deep Learning Convolutional Neural Networks YOLOv3 algorithm and Manually Traced cephalometric landmarks with AAOF collection as a control group. 110 images from Denver AAOF Legacy Collection were traced manually in Dolphin Imaging. The same images were uploaded to AI software Ceppro DDH Inc. For control group, coordinates for the same 110 images were extracted from AAOF collection. The mean distances were assessed relative to the reference value of 2mms. SPSS (IBM-vs. 27.0) software was use for the data analysis. The results showed that there is no statistical difference for 12 out of 16 points when analyzing absolute difference between MT and AI group. Successful detection rate for AI within 2 mm of accuracy while comparing MT and AI group was 75 % and 93% within 4 mm. The findings of our research are consistent with existing literature. Relative difference analysis revealed that AI tends to underestimate in vertical direction for about 1 mm for most of the points. The comparison between AI and AAOF group as well as MT and AAOF showed statistical differences, however, most of the points were within 4 mm range.

Therefore, we can conclude, that, AI could be considered a promising tool to facilitate cephalometric tracing process in routine clinical practice and in research settings.

1. INTRODUCTION

1.1 Background

Cephalometric analysis is an integral part of establishing proper diagnosis and treatment planning in orthodontics. Landmark identification and its reliability are important aspects of cephalometric analysis. Studies have found that cephalometric analysis using lateral cephalograms can be somewhat time consuming, particularly when digital data were not properly calibrated (in size or resolution). For clinicians, despite the use of the analysis as part of the overall diagnostic framework, the accuracy of cephalometric landmark identification can also be unreliable. More experienced clinicians/researchers who routinely use cephalometric analysis may be more consistent in identifying the cephalometric landmarks, whereas others may not be as reliable. Digitizing cephalometric landmarks, if done properly, can be a difficult and time-consuming process, and that could be the reason why clinicians abandoning its routine use. It is also a challenge for researchers, who often utilize cephalometric numbers/values as exclusion or inclusion criteria in their sample selection. Vast cephalometric collections, such as the AAOF Legacy collection, have been obtained through the outstanding effort of organizations. Unfortunately, there is a vast amount of valuable data that could be collected and analyzed but is not part of the AAOF Legacy collection. The lack of properly digitized, classified, and catalogued data makes evidence-based study, such as systematic review and meta-analyses, difficult to conduct.

Rapid development of artificial intelligence (AI) in recent years has penetrated many aspects of daily life, including the analyses of extensive available datasets. With respect to orthodontic diagnostic and treatment planning, developers of AI technology aim to auto-identify cephalometric landmarks from cephalometric radiograph images with little human interaction.

AI can be trained to analyze and find new patterns, which can be considered monumental in diagnostic and treatment planning. The aim of this study is to establish the accuracy of landmark detection using a fully automated AI-based machine learning and convolutional neural network by DDH Inc. and compare its reliability to that of well-calibrated, experienced orthodontists conducting landmark identification, using the AAOF Legacy collection data as a control. In this study, reliability of 16 standard lateral cephalometric automatic detected landmark locations derived from 110 samples of the AAOF Legacy collection will be compared to the expert orthodontist panels. The primary goal of the study is to investigate whether there will be difference in accuracy between the AI fully automated processes when compared to expert orthodontists and to the pre-detected AAOF collection data when identifying locations of the 16 most commonly used cephalometric landmarks; This study represents a significant step toward subsequent phases leading to detecting anomalies and establishing more robust diagnostics and assessment of treatment outcomes in both 2D and 3D diagnostics.

1.2 **<u>Objectives</u>**

- Compare accuracy of 16 cephalometric points between AI and manual tracing.
- Compare accuracy of 16 cephalometric points between AI and pre-detected AAOF group.

1.3 **<u>Null hypothesis</u>**

- There is less than 2 mm variation in mean accuracy of AI fully automated processes when compared to expert orthodontists and to the pre-detected AAOF data when identifying locations of each of the most used 16 cephalometric landmarks.

2. BACKGROUND

2.1 **Difficulties associated with manual cephalometric tracing**

Despite the advanced technology in this orthodontic era, which includes innovation(s) of imaging systems and software, the tools used in diagnostics and treatment planning have not experienced similar advances during the past century.[1,2] For instance, most clinicians use cephalometrics for orthodontic treatment diagnostics and planning. In 2002, 90 percent of orthodontists in the United States routinely obtained cephalograms,[3] even though many of those who take these radiographs routinely do not even trace them to get the measurements but simply use them as part of what is considered the standard of care.[4,5,6] Several systematic reviews and prospective studies have argued that cephalograms are not routinely needed for orthodontic treatment and have no significant impact on treatment planning decisions.[4,5,6,7] These studies also stated that lateral cephalograms are time consuming and only taken for other uses, such as medico-legal reasons, in a teaching environment, or because of the practitioner's lack of experience.[4,5,6] There is an interesting scenario that has developed worldwide within the orthodontic and craniofacial fields within the last 10 years, wherein the difficulties of properly using digital data and its associated time requirements, with special respect to digital imaging, are somehow demotivating users to access real measurements and cephalograms and to fail to compare superimpositions at different timepoints during treatment. This unfortunate practice leads to imprecise diagnostics and treatment plans that are not optimized. Using cephalometric imaging software (such as Dolphin Imaging, QuickCeph, etc), an experienced clinician spends on average 10-15 minutes to place landmarks manually, which makes the procedure time-consuming and subject to errors.[7,8]

2.2 AI implementation for medical imaging analysis

Rapid development of artificial intelligence (AI) in recent years has penetrated many aspects of daily life, including the analyses of extensive available datasets. The accumulation of data in many formats by search engines such as Google and social media (Twitter, Facebook, and Instagram) has great potential for enhancement and improvement of all aspects of our lives. For example, in conjunction with the current Oral Health 2020 government program (https://www.healthypeople.gov/2020/topics- objectives/topic/oral-health), there are multiple aspects in which current and future AI technology can be implemented.[30] These include: (1) broadening access and quality of care; and (2) implementing disruptive technology (in this case AI) for the analyses of extensive datasets, such as in population and demographic studies, leading to improved evidence-based clinical care. With orthodontic diagnostic and treatment planning, this AI technology could deliver not only an easy, practical, and precise tool for the practicing clinician, but also would be capable of significantly improving the amount of available labeled data.

Despite readily available studies demonstrating different process to auto-detect craniofacial landmarks, most clinicians use approaches based on image-processing techniques where images of cephalometric radiographs require intense human preparation, such as re-scaling, calibration, and labeling. Calibration and other image preparations are time consuming and, if not done properly, often generate landmark outliers[9,10,11] as they strongly rely on the quality and size of the cephalometric radiographic images. Many other studies propose different novel frameworks for landmark detection in cephalometric radiographs and demonstrate results with accuracy of 72% but, again, these are not fully automated procedures.[12,13] Current advances in this technology have, in turn, provided hardware and software development that is sufficiently robust to support the large computational requirements of complex AI algorithms and their application to machine learning. Applications of a variety of deep learning architectures, such as convolutional deep neural networks, deep belief networks, and recurrent neural networks, to the creation of algorithms in important fields such as natural language processing, computer vision, speech recognition, and bioinformatics have resulted in efficient and accurate automation of many pragmatic tasks.[14,15,16] However, the developed methods were unable to compete with manual landmark identification. In recent years, several IEEE International Symposium on Biomedical Imaging (ISBI) Grand Challenges were organized on this topic to encourage the development of better algorithms. The results were described as providing a benchmark for any future development.[17]

2.3 **Convolutional Neural Networks**

Computer vision, a part of AI, that enable machines to perceive the world as human beings, and use the knowledge for image recognition, analysis, and classification, has been constructed and tremendously improved with time, mostly over one particular algorithm – a Convolutional Neural Network.[29] A Convolutional Neural Network (CNN) - is a deep learning algorithm which can take an input image, assign importance to various aspect/objects in the image and differentiate one from the other. Deep learning means that CNN can learn different characteristics of the image, or, in other words, to be trained to understand the sophistication of the image better than traditional classification algorithms.[29] The role of the CNN is to reduce the images into a form which is easier to process, without losing critical features. Deep learning CNN has a vast application in dentistry and includes developing programs capable of detection of pathologies, automatic identification of cephalometric landmarks, and segmentation of teeth and other structures.[21]

The architecture of a CNN is like that of the connectivity pattern of neurons in the human brain and was inspired by the organization of the visual cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. When these fields overlap, they cover the entire visual area. The "neurons" of CNNs are filters. Filters have input parameters (the features they are searching for) and generate a value as an output.[29] The input size is a fixed square called a receptive field.[29] The output of one filter applied to the previous layer is called a feature map.[29] A given filter is drawn across the entire previous layer, moved one pixel at a time.[29] Each position results in an activation of the "neuron" and the outputs are collected in the feature map. (Figure 1). If the filter is designed to detect a specific type of feature in the input, then the application of that filter systematically across the entire image allows the filter to discover that feature anywhere in the input.[29] The application of a filter to an input that results in an activation is called "convolution". Together filters and feature maps make up a convolutional layer. Convolutional layer is central to CNN and it gives the network its name.[22]

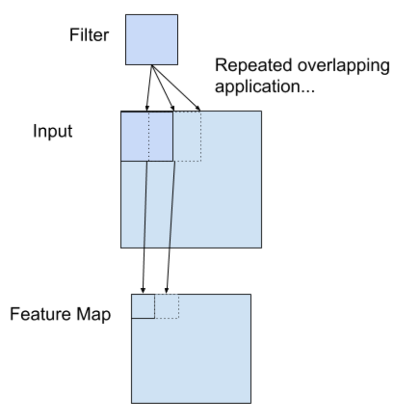


Figure 1. Filter scanning two-dimensional image to create a feature map. (https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/)[29]

In summary, there is an input, which basically an image, and there is a filter, which is looking for specific parameters. The filter is systematically applied to the input data to create a feature map.[29] The feature map represents the detected features in the input.[22]

CNNs learn multiple features in parallel for a given input. In average, a convolutional layer learns from 32 to 512 filters.[29] This diversity allows specialization – each filter looking for a specific pattern with complexity gradually increasing. For example, the first convolutional layer is responsible for capturing the low-level features such as edges, color, or gradient orientation. With added layers, the architecture adapts to the high-level features, allowing the network to have wholesome understanding of images in the dataset, like how human beings would (Figure 2).

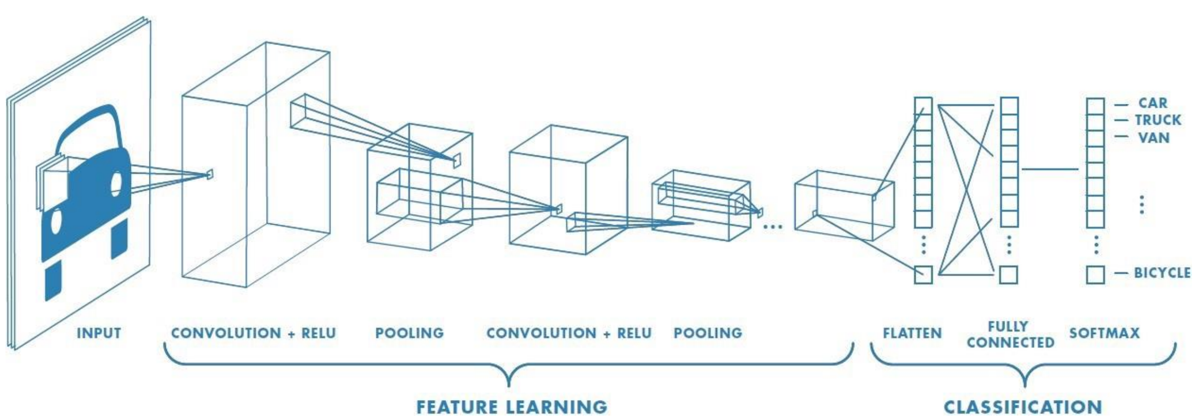


Figure 2. A CNN sequence to assess a visual object.
(https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/)[29]

Besides convolutional layers, CNNs also contain pooling layers and fully connected layers. The role of pooling layers is to compress or generalize feature representations. Fully connected layers are used at the end of the network after feature extraction and consolidation has been completed by the convolutional and pooling layers. Fully connected layers create final combinations of features and aid in making predictions by the network.[22]

In contract to traditional classification algorithms, that demand a perfectly resized and oriented image, CNN allowing for the objects in the images to be tilted or repositioned in the scene and still be detectable by the network. It is this reason why CNNs are so useful for object recognition in photographs, faces and other objects with varying orientation.[23] In orthodontics the example of traditional machine vision algorithm utilization is Planmecas Romexis®Cephalometric Analysis Software. It provides automatic cephalometric points detection and tracing in seconds, however, the software requires a lateral radiograph to be obtained only on Planmeca Cephalometric Imaging Unit, where it is initially automatically resized, properly oriented and calibrated. Whereas, CEPPRO software used in a current study, utilizes a CNN (YOLOv3) algorithm, which makes it invariant to object position and distortion, as well as open source, which means that the lateral radiograph could be taken on any cephalometric machine and be automatically traced in CEPPRO software.

2.4 **YOLOv3 algorithm**

The *"You Only Look Once,"* or YOLO, family of CNNs designed for fast object detection, developed by Joseph Redmon, et al. and first described in the 2015 paper titled "You Only Look Once: Unified, Real-Time Object Detection".[24] The approach involves a single deep CNN (originally a version of GoogLeNet, later updated and called DarkNet) that splits the input into a grid of cells and each cell directly predicts an object classification.[24]

There are three versions of the approach: YOLOv1, YOLOv2, and YOLOv3. The first generation proposed the general architecture, whereas the second version refined the design, and the final version further refined the model architecture and training process. According J. Park and et al, YOLOv3 algorithm was faster and more accurate than another latest widely used deep-learning algorithm – Single Short Multibox Detector (SSD).[25] YOLOv3 outperformed SSD in accuracy for 38 of 80 landmarks tested on 283 images. The mean computational time spent per image was 0.05 sec for YOLOv3 and 2.89 sec for SSD. Therefore, YOLOv3 seemed to be more promising than SSD algorithm for automated cephalometric landmark identification in orthodontic clinical practice.[25]

2.5 CEPPRO as Automatic Cephalo-Diagnostic Solution by DDH Inc

DDH Inc is an innovated A.I-based digital healthcare company that develops imaging software for medical and dental fields. DDH Inc was found in April, 2017 with a headquarters in Seoul Korea. In 2018 A.I. academic-industry laboratory based in Seoul National University Biomaterial Research Building was launched. In 2019 DDH office was opened in Chicago, USA. Over 3 years DDH Inc has attracted the investments of 5.3 million dollars and currently has partnership with NeoBiotech and Daewon Corporations. DDH Inc claims that it’s area of expertise is deep learning based medical image reading and diagnoses, and the mission is to reduce the doctor’s burden in performing intensive monotonous tasks in order to allow more patient’s engagement. Screening and diagnostic solutions are supposed to automatically infer insights from medical images and provide a quotative outlook of health or decease. In the long run AI is believed to contribute to improve clinical efficiencies, lower the doctors’ ambiguity and errors.

The range of DDH Inc software includes DDHAIM Medical Solution: A.I. based DDHAIM MRI Brain Volumetry for early detection of Alzheimer's disease; DDHAIM CHEST CT&X-ray for detection of lung cancer/other pathology; DDHAIM Breast Cancer, and DDHAIM MR Body Markers. DDHAIM Dental Solutions consists of DDHAIM CEPPRO software which is A.I. based orthodontic cephalogram analysis; DDHAIM PANO for auto detection of tooth caries, peri-apical lesions and peri-implantitis; DDHAIM CEPAIR for cephalometric analysis of airway for sleep apnea risks evaluation. In addition to it DDH Inc also provides services in Orthodontic Consulting, CAD/CAM orthodontic appliance manufacturing, patient management, and oral&maxillofacial image interpretation. All software programs are cloud service and require membership.

CEPPRO is A.I. based software that automatically provides detection of landmarks, cephalometric tracing, measurements, and cephalometric analysis. The program was developed in Seoul National University Dental Hospital (SNUDH) by the professor group after uploading 15.000 cephalometric images to the A.I engine. The manufacturer claims that CEPPRO has a powerful accuracy of detecting cephalometric landmarks. The error between AI algorithm utilized in CEPPRO (YOLOv3) software and human examiners was 0.9 mm based on evaluation of 80 landmarks in 253 consecutive digital lateral cephalometric radiographs.[20]

2.6 AAOF Collection

American Association of Orthodontist Foundation Craniofacial Growth Legacy Collection is an open-source web-site https://www.aaoflegacycollection.org/aaof_home.html, where nine of eleven longitudinal growth records from USA and Canada are presented.[28] The main purpose of AAOF database, which is mainly comprised of lateral cephalometric images, is to provide material for further investigation by clinicians, students, and researchers. AAOF collection presents a unique source of longitudinal records of craniofacial growth among children who did not receive orthodontic treatment, and, therefore, could be used as a control group in many studies. The pixel resolution of the images is considered sufficient for most practical uses and equals or exceeds resolution of digital lateral cephalometric radiographs. The entire content of AAOF collection site is available without cost for on-line use and downloading by all members of the orthodontic community. The University of Oklahoma Denver Growth Study, used in our study, includes the growth data from untreated children from 1927 to 1967. Denver Growth pilot study was initially focused on the lateral cephalometric radiographs between 8 years and 18 years with at least 4 films per person. Denver Collection is made up of images that belongs to 57 male and 56 female subjects. The main reason why we chose Denver Collection is that 123 images have been traced by expert orthodontist, Diplomates of American Orthodontic Board. The coordinates were extracted when images were superimposed by Sella-Nasion line with Sella as 0:0, and available for download and further comparison. The second reason is that all lateral cephalometric images available from Denver collection has either a 30 mm ruler on the image or 150 mm distance between upper right and lower right fiducial points, that could be used for calibration purposes and coordinate of cephalometric points comparison.

3. **METHODS AND MATERIALS**

3.1 **Subjects**

Each of the 125 lateral radiographs was downloaded twice from the AAOF website Denver collection sample (https://www.aaoflegacycollection.org).[28] The first download was with the landmarks already identified. The second download file did not have the landmarks on the image (Figure 3). To avoid possible association with the original files, which are available online for public consultation, each file was de-identified and relabeled with the same number in a simple numeric sequence and divided in two different folders.

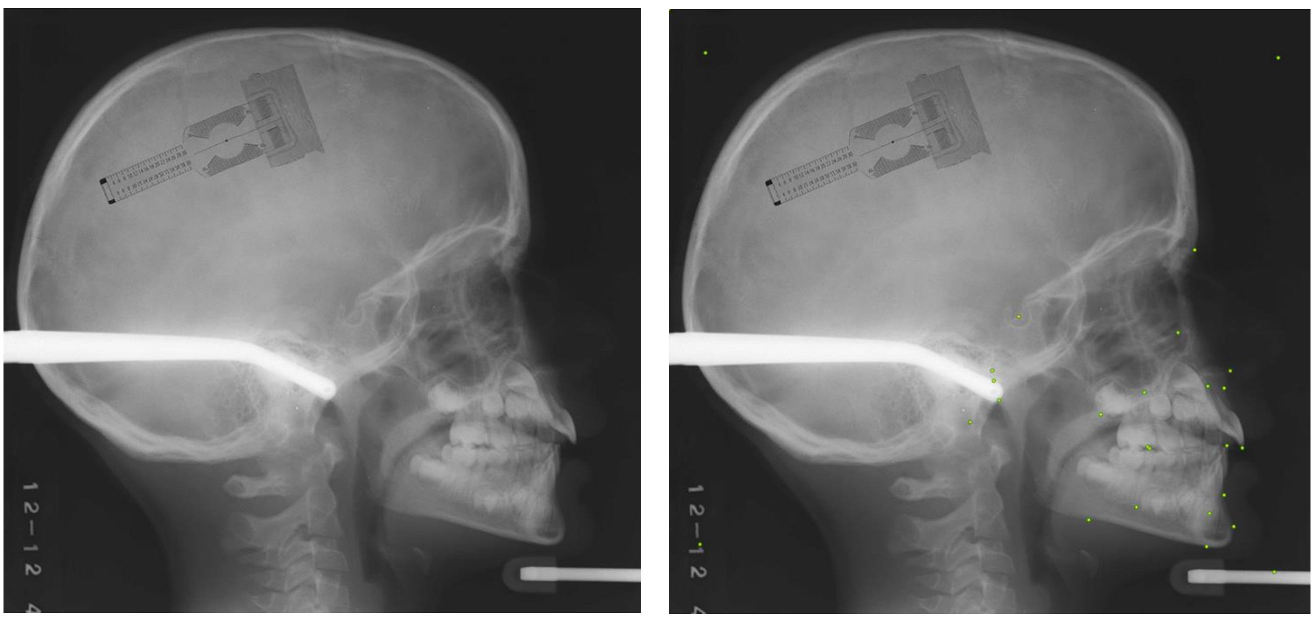


Figure 3. Sample lateral cephalometric radiograph downloaded from AAOF website Denver collection twice: without and with landmarks.

Since data is publicly available from the web, an IRB exception was applied for with this project (IRB: 2019-1180).

The criteria for inclusion and exclusion in this study were as follows:

3.2 **Inclusion criteria**

- Digital lateral radiographs that have a ruler for calibration or fiducial points on the Computer Imaging System.
- Digital lateral radiographs from the same subjects were done at least with 3 years interval to avoid images to be too much similar to each other.

3.3 **Exclusion criteria**

- Digital lateral radiographs that do not satisfy the 2 inclusion criteria.
- Digital lateral radiographs of patients with syndromes and or other craniofacial anomalies that do not represent a normal craniofacial configuration.
- Digital lateral radiographs of poor quality that does not allow manual tracing.

110 subjects from Denver collection AAOF website remained after exclusion criteria. Based on previous studies and calculations including power tables, 110 subjects were considered enough to do this study. Out of 110 subjects, 25 had 30 mm ruler on their images and remaining 85 subjects had 150 mm distance between two fiducial points (upper right and lower right) for calibration purposes.

3.4 **Data processing**

Before digitizing the images intra- and inter reliability check was performed. For intra reliability check 20 lateral cephalometric images were traced twice by first investigator. For the inter reliability, another trained investigator traced the same images to be compared with the first investigator tracing, in each of the 20 subjects. The Intraclass Correlation Coefficient (ICC) is a reliability index that reflects both degree of correlation and agreement between measurements. A high ICC (close to 1) indicates high similarity between values from the same group. A low ICC (ICC close to zero) means that values from the same group are not similar. In our study ICCs were approximately at least 0.80, which indicated a good intra- and inter reliability. We had to exclude Porion point from the study since on majority of radiographs from AAOF Denver collection Porion was covered by cephalostat contours.

The landmarks that were evaluated and further used in this study included: Nasion (Na), A point (A), B point (B), Menton (Me), Gonion (Go), Upper incisor tip, Lower incisor tip, Upper incisor apex, Lower incisor apex, Anterior Nasal Spine (ANS), Posterior Nasal Spine (PNS), Pogonion (Pg), Pterigomaxillary fissure point (Pt), Basion (Ba), Articulare (Art) and Orbitale (Or).[19] Exact definitions of points are presented in Table I.

TABLE I

SELECTED LANDMARKS AND DEFINITION

	Dolphin Imaging	AAOF collection	AI Ceppro software	Description
1.	Nasion	NASION	Nasion	The most anterior point on the frontonasal suture
2.	A-point	POINT A	A-Point	The innermost point on the contour of the maxilla between the anterior nasal spine and the incisor
3.	B-point	POINT B	B-point	The innermost point on the contour of the mandible between the incisor and the bony chin
4.	Menton (Me)	MENTON	Menton	The most inferior point on the mandibular symphysis in the midline
5.	Gonion (Go)	GONION L	Gonion	The most inferior point on the curvature of the angle of the mandible
6.	Upper incisor tip	U I EDGE	Maxilla1crown	The incisal edge of maxillary central incisors
7.	Lower incisors tip	L I EDGE	Mandible1crown	The incisal edge of mandibular incisor
8.	Upper incisor root apex	U I APEX	Maxilla1root	The root apex of maxillary central incisor
9.	Lower incisor root apex	L I APEX	Mandible1root	The root apex of mandibular incisor
10.	Anterior Nasal Spine	ANS	AnteriorNasalSpine	The most anterior point of the maxilla at the nasal base
11.	Posterior Nasal spine	PNS	PosteriorNasalSpine	Tip of the posterior spine of the palatine bone of the hard palate
12.	Pogonion (Pg)	POGONION	Pogonion	The most anterior point on the chin
13.	PT point	N/A	Pterygoid	The most posterior and superior point on the pterygomaxillary fissure
14.	Basion	BASION	Basion	Most inferior point on the anterior margin of the foramen magnum in the median plane
15.	Articulare	ARTICULAR	Articulare	A point midway between the two posterior borders of the left and right mandibular rami at the intersection with the basilar portion of the occipital bone.
16.	Orbitale	ORBITALE	Orbitale	Lowest point of the floor of the right orbit, the most inferior point of the external border of the orbital cavity

3.4.1 **Manually Detected Landmarks group**

From the folder that did not include the landmarks on the image, the radiographs were given to the first investigator (orthodontic resident), who identified and traced landmarks using the Dolphin Imaging System software with an ABO analysis sequence. The images were not changed in size or rotated. Every image was uploaded to Dolphin imaging as it is, without realigning/reorienting. After 110 radiographs were traced, coordinates in absolute values in mm with Sella as 0:0 point in the Cartesian system were extracted for each cephalometric image, copied to Excel sheet and saved as MDL (Manually Detect Landmark) group (Figure 4).

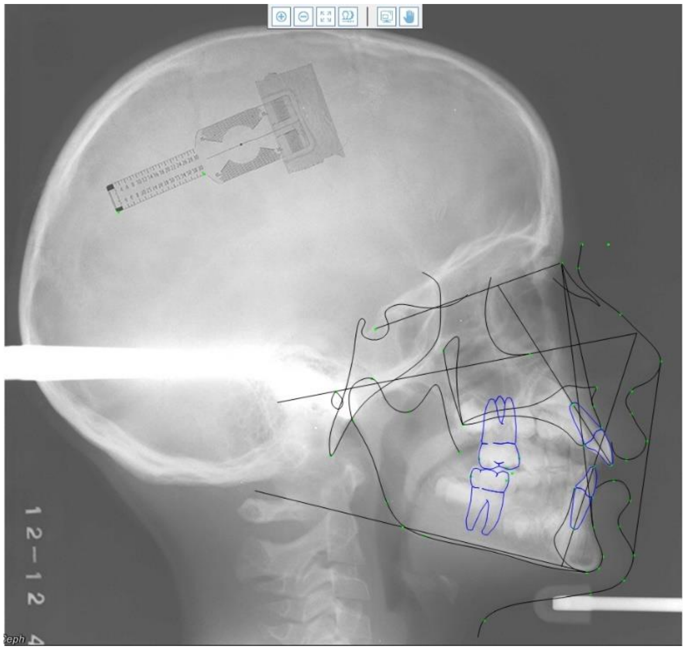


Figure 4. Sample lateral cephalometric radiograph from Manual tracing group with digital landmarks in Dolphin imaging.

Cartesian system consists of two perpendicular directed lines: x-axis and y-axis with specified unit length. Each point is presented in the system by two numbers: x-coordinate and y-coordinate. Both coordinates were extracted for each landmark.

The average time for manual tracing of lateral cephalometric image was 6.5 minutes. Overall, it took 11 hours 54 minutes to trace 110 images.

3.4.2 **AAOF control group**

From the folder that included the green dot landmarks on the image, coordinates with Sella as 0:0 point in the Cartesian system were extracted from Denver Collection and saved in Excel sheet as AOF control group (Figure 5). Since the images in Denver collection were aligned along Sella-Nasion line by default settings before coordinates were extracted, it was necessary to "derotate" them for future comparison with Manually detected landmark group and Auto-detected landmark group. First, the customized rotational angle was calculated based on Matrix Formula for rotation. Then, all coordinates for 16 points from 110 subjects were transformed in a separate Excel sheet and saved as AAOF control group. Since Pt point coordinates were not available in Denver collection, we had to exclude this point from our control group.

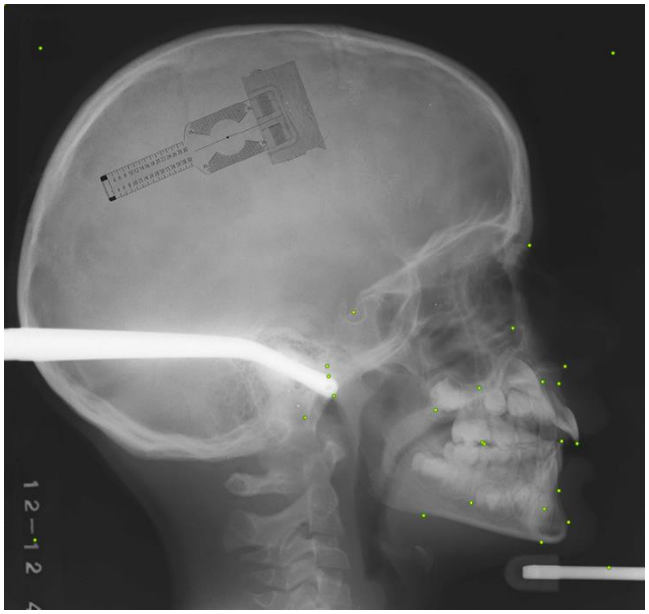


Figure 5. Sample lateral cephalometric radiograph from AAOF control group with digital landmarks marked with green dots.

3.4.3 **Auto-detected landmarks group**

Before uploading radiographs to AI, several issues had to be solved. Initially, coordinate system in AI program (Ceppro software) was set with lower left corner as a 0:0 point. As a result, all quadrants exhibited “+” sign in front of coordinate values. To make the comparison with MDL and ADL possible, we asked software engineer from DDH company to change coordinate system into in the Cartesian system with Sella as 0:0 point. The next issue was that AI software used to calculate coordinates in pixels. To transform pixels into mm, we asked engineer to put a virtual ruler into AI software, that allowed us to determine mm-pixels ratio (how many pixels in one mm) and, therefore, present extracted coordinates in mm (Figure 6).

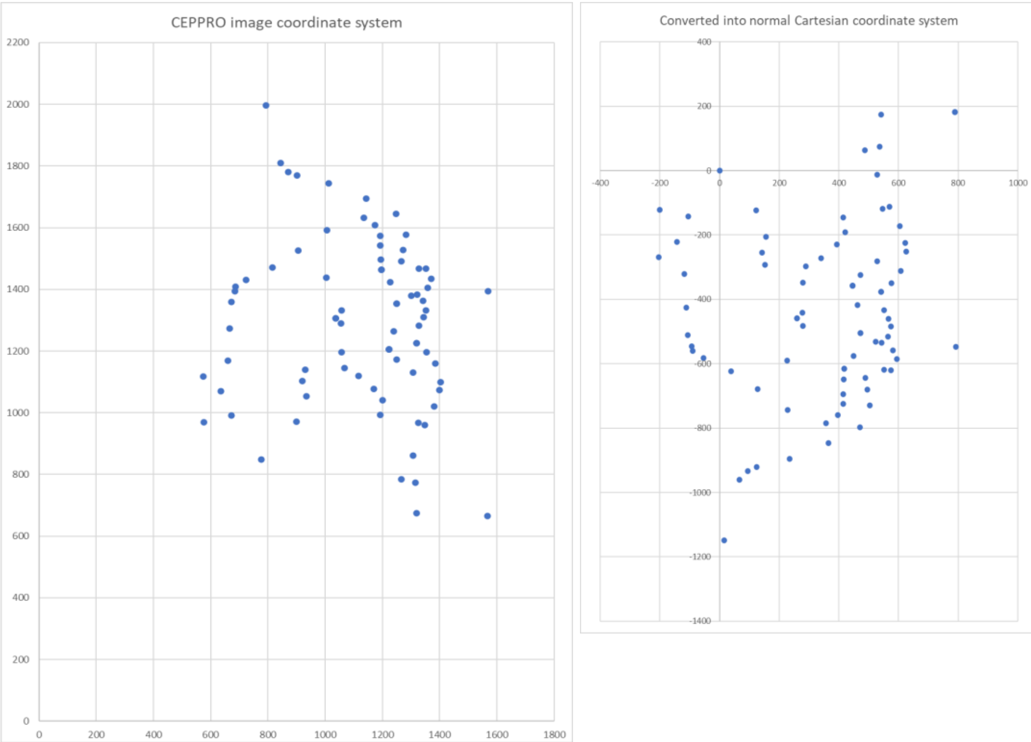


Figure 6. Converting CEPPRO coordinate system into Cartesian coordinate system with Sella as 0:0.

The new software features were performed on 2 subjects and the implemented changes were successful: the accuracy of majority of AI landmark identification was within 2 mm in comparison with manual tracing. In this pilot study, 2 out of 150 lateral radiographs were downloaded from AAOF website collection sample. Radiographs went through the process of manually re-digitizing into the Dolphin Imaging™ System to get coordinates. The same radiographs were also uploaded to the AI engine. The landmark coordinates from each group, represented by 2 Cartesian coordinates (horizontal and vertical), were exported, and saved as an Excel file.

Initial Observations: the figure 7 below shows comparison of 18 landmark coordinates between AI and humans for 2 subjects. The mean difference was equal or less than 2 mm for 15 out of 18 landmarks.

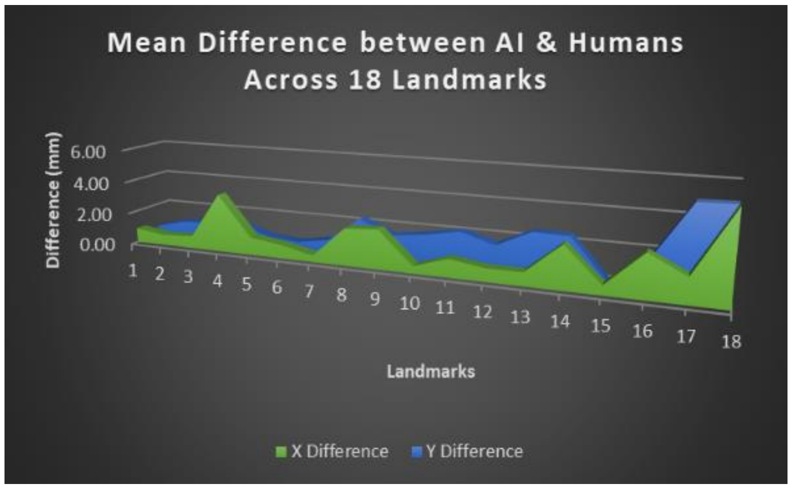


Figure 7. Pilot study results.

From the folder that did not include the landmarks on the image, the radiographs were uploaded to the AI engine with no further labeling and/or changes, and the landmarks were auto identified. The upload was performed by DDH company software engineer, Mr. Thomas Kim.

During the upload the AI engine sent back the image with the auto-detected landmarks on it in an average time of 7 seconds for each image. Coordinates with Sella as 0:0 point in the Cartesian system were exported for each cephalometric image and saved in Excel file as Auto-detected landmarks (ADL) first upload.

In a group of radiographs with 30 mm ruler, the beginning, and the end of virtual ruler in AI, that also equals 30 mm, were manually aligned with the ruler on radiographs (Figure 8).

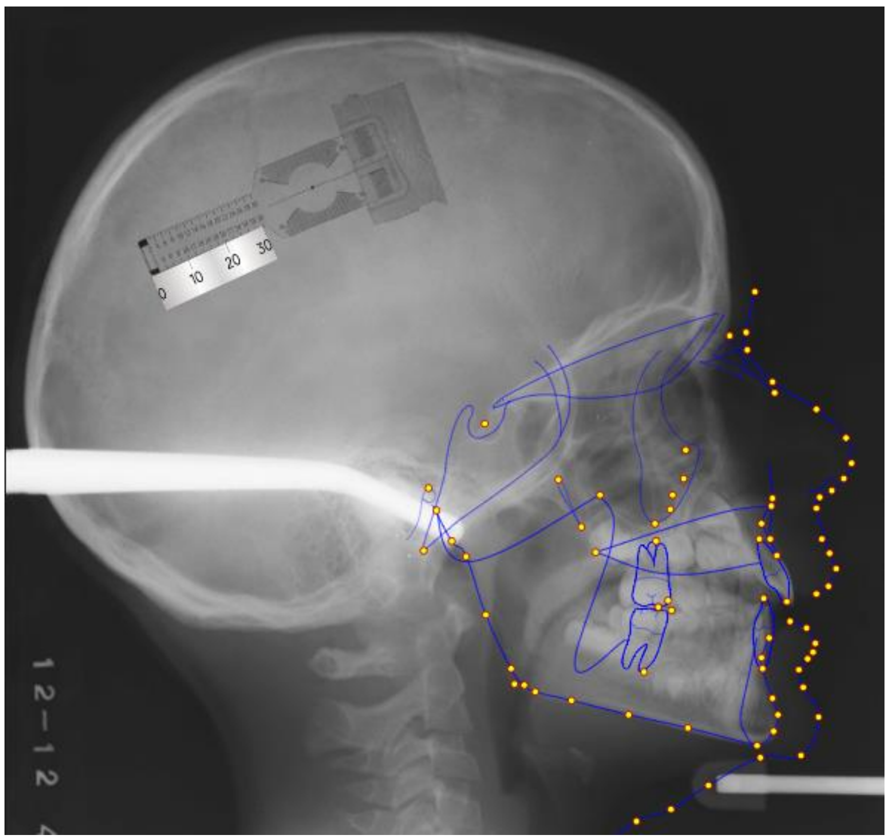


Figure 8. Sample lateral cephalometric radiograph with 30 mm ruler uploaded to CEPPRO software

This procedure took additional 12 second for each image. In a group of radiographs with 150 mm distance between two fiducial points, the virtual ruler was stretched between two fiducial points. Then, the extracted coordinates, both x and y values, were divided by 5 with the help of formula in the Excel sheet.

During the upload, AI software indicated “ERROR” on 37 images and 1 image was regarded as “non-identifiable”. Coordinates from the upload of radiographic images to AI was sent by DDH engineer Thomas Kin in separate Excel workbooks within 64 minutes after he received the 110 images from me.

After data collection all subjects were reassigned as follows: Manually Digitized Group, AI Test Group, AAOF Control Groups (Figure 9).

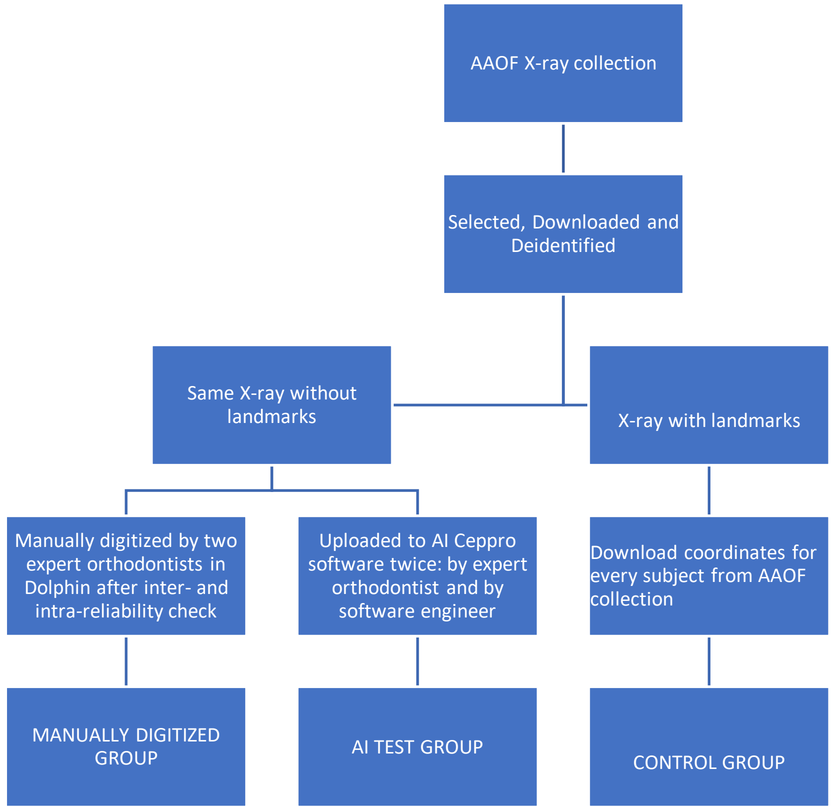


Figure. 9. Data preparation flow chart.

4. RESULTS

4.1. Inter and intra reliability check

For inter reliability check 20 images were traced twice in Dolphin Imaging by one examiner – orthodontic resident (Galina Bulatova). Both times coordinates were extracted, put in the Excel table and analyzed. For Intra-reliability check the same 20 images were traced in Dolphin Imaging by Orthodontic Faculty – Prof Flavio Sanchez. In our study ICCs (Interclass Correlation Coefficients) were approximately 0.80, which indicated a good intra- and inter reliability.

4.2. Comparison Manual Tracing vs AI group

Two parameters were used for comparison between groups: absolute and relative difference. The absolute difference shows how far was deviation on average. Relative difference shows a tendency for AI to "put" landmarks vertically lower or higher, and horizontally closer or further away from the point of origin, in comparison to Manual Tracing. The way how average and relative difference were calculates is presented in Figure 10.

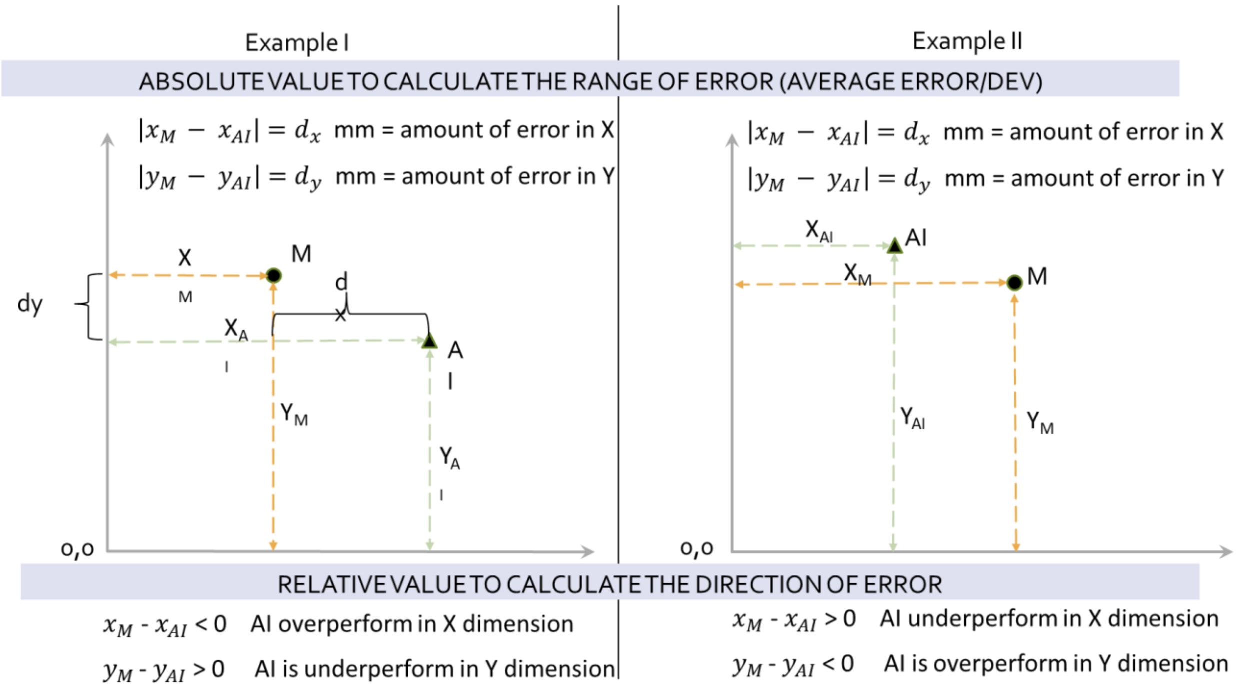


Figure 10. Absolute and relative differences comparison.

Absolute differences comparison demonstrated that the following point did NOT indicate statistically significant mean difference between MT and AI: Nasion (X,Y), Point A (X,Y), Point B (X,Y), Menton (X,Y), L1 tip (X,Y), ANS (X,Y), PNS (X,Y), Pogonion (X,Y), PT (X,Y), Articulare (X,Y) and Orbitale (Y). Therefore, successful detection rate for AI within 2 mm is 75 %. Points U1 apex (X,Y), L1 apex (X,Y), Basion (X,Y), Gonion (X,Y) and Orbitale (X) showed significant difference. Possible explanation for root apexes lies in the fact, some lateral cephalometric images belonged to growing children and upper central incisors apex could not be formed completely. Having in consideration, that AI and Dolphin imaging software has template for upper and lower incisors, open root apexes could be identified according to the template rather than real position. The bigger difference in vertical direction supports this assumption. As for Gonion, confusion can come from the fact that this is point is usually an average between two mandibular angle contours. Basion and Orbitale are generally considered hard to detect and not reliable points in cephalometric analysis.

As for relative differences, most of the variables (85%) show to be normally distributed when evaluated with the Shapiro-Wilk test. Student paired t-tests at significance level of 5 % were used to compare the mean differences for each pair of measurement techniques in each of the X- and Y-components. SPSS (IBM-vs. 27.0) software was use for the data analysis. Relative difference analysis showed that AI has a tendency to identify vertically lower points Nasion, Gonion, U1 apex, L1 apex, PNS and Articulare. AI was rather accurate for most of the points in horizontal direction except for Gonion (8.4 mm).

4.3 **Comparison AAOF vs AI**

Absolute difference comparison showed that only point Orbitale (Y) did not have statistical difference within 2 mm between AAOF and AI. The remaining parameters, for the exception of Articulare (X), Gonion (X, Y) and ANS (X,Y), showed statistical difference, however it mostly remained within 4 mm. Relative difference again demonstrated AI tendency to put most of the points in horizontal and vertical dimension closer to 0:0 Sella point.

4.4 **Comparison AAOF vs MT**

Four points: Basion (X), Point A (Y), PNS (Y) and Orbitale (Y) did not show statistically significant difference within 2 mm between AAOF and MT. However, most of other points with a small exception of Articulare (X) and ANS (X), were within 4 mm. Relative difference revealed that for the exception of point ANS (X), MT was prone to underestimate in comparison to AAOF. Most of the points in vertical direction demonstrated the similar tendency.

4.5. **Overall comparison between groups in X and Y directions**

The results of statistical analysis for Absolute differences are presented in Table II, and the Results for Relative differences could be found in Table III. Based on relative differences, AI tends to underestimate approximately 1 mm in vertical direction in comparison with MT group. Figure 11 and Figure 12 illustrate absolute and relative mean differences in X and Y.

TABLE II

ABSOLUTE DIFFERENCE

	MT vs AI						AAOF vs AI						MT vs AAOF					
	V_x			V_y			V_x			V_y			V_x			V_y		
	Landmark	$\bar{x}$	SD	Landmark	$\bar{x}$	SD	Landmark	$\bar{x}$	SD	Landmark	$\bar{x}$	SD	Landmark	$\bar{x}$	SD	Landmark	$\bar{x}$	SD
No significant difference $p>0.05$	Nasion	1.4	3.9	Nasion	1.3	2.7				Orbitale	2	2.5	Basion	2.3	2.3	Point A	2	1.4
	Point A	2.4	2.1	Points A	2.2	3.6										PNS	1.8	1.4
	Point B	1.9	3.2	Point B	2.7	3.7										Orbitale	2	1.5
	Menton	2.4	3.6	Menton	1.9	3.5												
	U tip	2.8	6.3	U tip	1.9	3.6												
	L tip	1.7	2.5	L tip	2.1	3.8												
	ANS	1.7	1.7	ANS	2.3	3.3												
	PNS	2.1	1.9	PNS	2.1	1.7												
	Pogonion	1.9	3.6	Pogonion	2.5	4.1												
	PT	1.7	1.6	PT	2.1	1.6												
	Articulare	2	2.5	Articulare	2	1.7												
				Orbitale	2.4	2.4												
	Mean	2.0	3.0	Mean	2.1	3.0				Mean	2.0	2.5	Mean	2.3	2.3	Mean	1.9	1.4
Significant difference $p<0.05$	U apex	3	2	U apex	3.4	3.6	Nasion	7.9	26.2	Nasion	1.8	3.3	Nasion	2.6	1.2	Nasion	0.9	1.1
	L apex	2.6	3	L apex	5.4	2.9	Point A	4.1	2.5	Points A	2.6	3.2	Point A	3.3	1.6			
	Basion	3.1	2.4	Basion	4.3	3	Point B	4.2	3.6	Point B	4.2	3.4	Point B	3.3	1.7	Point B	3.5	2.9
	Orbitale	3	2.3				Menton	4.7	4.5	Menton	5.7	2.5	Menton	5.1	2.6	Menton	5.6	1.4
	Gonion	8.7	9	Gonion	4.9	4.7	Utip	5.3	6.8	U tip	3.8	3	U tip	3.7	1.8	U tip	3.4	1.4
							L tip	3.8	2.9	L tip	3.9	3.5	L tip	3.5	1.7	L tip	3.3	1.6
							U apex	2.3	1.9	U apex	2.9	3.2	U apex	4	2	U apex	3.4	3.2
							L apex	3	3.4	L apex	6.2	3.1	L apex	3.6	2.1	L apex	3.4	2.6
							Pogonion	4.2	4.2	Pogonion	6.5	3	ANS	15.1	8.2	ANS	2.4	1.6
							Articulare	65.2	32.9	Articulare	6.4	5.3	PNS	2.4	1.6			
							Orbitale	2.4	1.6				Pogonion	3.5	2	Pogonion	6.8	2.1
							Gonion	9.9	5.2	Gonion	7.3	3.8	Articulare	64	32	Articulare	6.2	5
							ANS	15	8.5	ANS	23.6	12.3	Orbitale	4.2	2.2	Basion	2.4	1.9
							PNS	2.5	2.3	PNS	2.6	1.6	Gonion	4.4	8.1	Gonion	5.6	5.2
	Mean	4.1	3.7	Mean	4.5	3.6	Mean	9.6	7.6	Mean	6.0	3.9	Mean	8.8	4.9	Mean	3.9	2.5

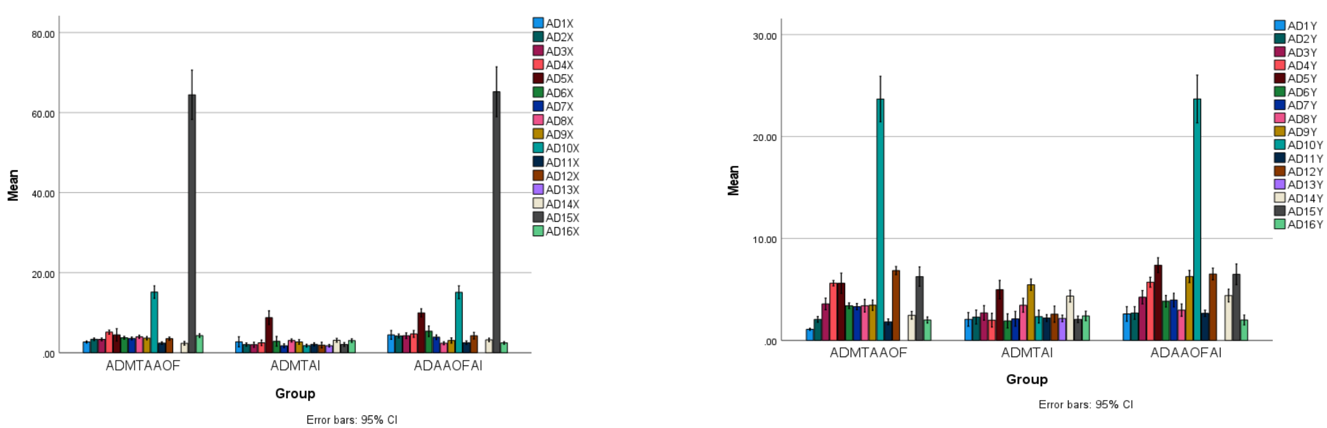
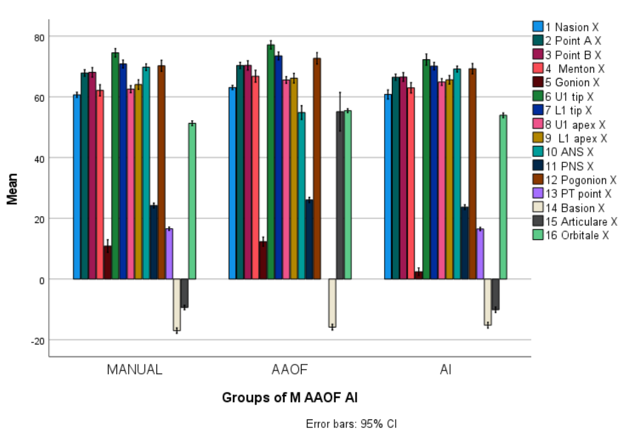


Figure 11. Absolute Distances Mean Differences.

TABLE III

RELATIVE DIFFERENCE

	MT vs AI						AAOF vs AI						MT vs AAOF					
	V_x			V_y			V_x			V_y			V_x			V_y		
	Landmark	$\bar{x}$	SD	Landmark	$\bar{x}$	SD	Landmark	$\bar{x}$	SD	Landmark	$\bar{x}$	SD	Landmark	$\bar{x}$	SD	Landmark	$\bar{x}$	SD
No significant difference $p>0.05$	Nasion	-0.12	7.2	Point A	0.5	4.4	L1 apex	0.5	4.6	Point A	0.2	4.2	Gonion	-1.4	9.2	Point A	0.3	2.5
	PNS	0.5	2.8	Point B	-0.3	4.6	Basion	-0.6	3.9	Point B	-0.2	5.4				Point B	-0.1	4.6
	PT	0.08	2.3	Menton	0.4	4				Articulare	0.03	8.4				PNS	0.04	2.3
				U1 tip	0.4	4.1				Orbitale	-0.3	3.2				Articulare	-0.9	7.9
				L1 tip	0.35	4.3												
				ANS	-0.5	4												
				Pogonion	0.8	4.8												
				PT	0.2	2.7												
	Mean	0.2	4.1		0.2	4.1		-0.1	4.3		-0.1	5.3		-1.4	9.2		-0.2	4.3
Significant difference $p<0.05$	Point A	1.3	2.6	Nasion	0.9	4	Nasion	2.2	6.9	Nasion	1.4	4.3	Nasion	-2.4	1.7	Nasion	-0.4	1.1
	Point B	1.5	3.4				Point A	3.8	3				Point A	-2.4	2.7	Gonion	3.5	6.8
	Menton	-0.8	4.3				Point B	3.8	4.1				Point B	-2.3	3			
	Gonion	8.4	9.3	Gonion	-3.6	5.8	Menton	3.8	5.2	Menton	-2.2	5.8	Menton	-4.6	3.4	Menton	2.6	5.1
	U1 tip	2.2	6.3				Gonion	9.9	5.4	Gonion	-7.1	4.2	U1 tip	-2.6	3.2	U1 tip	1.7	3.2
	L1 tip	0.6	3				U1 tip	4.8	7.2	U1 tip	-1.2	4.7	L1 tip	-2.6	2.9	L1 tip	1.2	3.5
				U1 apex	2.5	4.3	L1 tip	3.3	3.5	L1 tip	-0.8	5.2	U1 apex	-3.3	3.2	U1 apex	1.3	4.5
	L1 apex	-1.5	3.7	L1 apex	-4	4.7	U1 apex	0.7	3	U1 apex	1.2	4.2	L1 apex	-2.1	4.6	L1 apex	1.3	4.1
	ANS	0.6	2.4	PNS	-1.3	2.4				L1 apex	-5.3	4.5	ANS	14.9	8.6	ANS	22.3	14.2
	Basion	-1.7	3.5				ANS	-14.3	9.7	ANS	-22.8	13.8	PNS	-1.8	2.2			
	Articulare	0.7	3.1	Articulare	-0.9	2.5	PNS	2.3	2.4	PNS	-1.3	2.8	Pogonion	-2.4	3.2	Pogonion	3.9	6
	Orbitale	-2.6	2.8	Orbitale	-1.2	3.2	Pogonion	3.4	4.9	Pogonion	-3.1	6.4	Basion	-1.1	3.1			
	Pogonion	0.9	4				Articulare	65	32				Articulare	-64	32			
							Orbitale	1.5	2.5				Orbitale	-4.1	2.5	Orbitale	-1.6	1.9
	Mean	0.8	4.0	Mean	-1.1	3.8	Mean			Mean	-4.1	5.6	Mean	-5.8	5.5	Mean	3.6	5.0



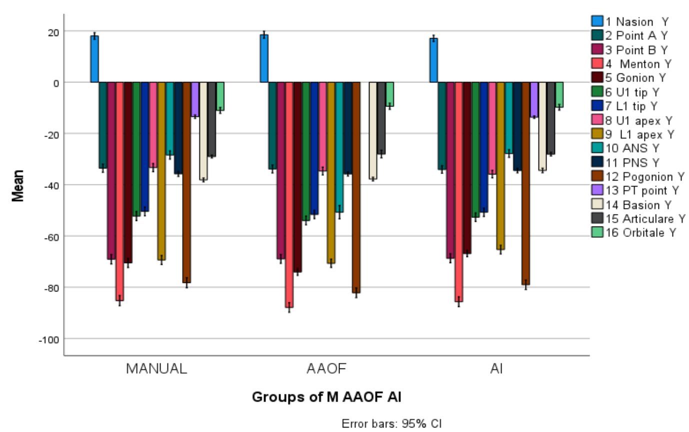


Figure 12. Relative differences. Paired t-test mean differences.

5. **DISCUSSION**

The results of our study are consistent with the existing literature in terms of AI accuracy in identification of cephalometric landmarks within 2 mm. Therefore, we accept Null Hypothesis that "There is less than 2 mm variation in mean accuracy of AI fully automated processes when compared to expert orthodontists."

However, we reject the Null Hypothesis in terms of comparison AI with AAOF group. Partially, it could be explained by the fact that some of the samples in AAOF have inconsistent data entry. When external data is used, caution must be taken. Probably, it would be better for future research to use raw images from AAOF with predetermined points and trace them in Dolphin Imaging to get coordinates.

According to literature, the accuracy of AI landmark identification can be calculated in several ways: by mean radial errors or by absolute differences or relative differences in x and y directions. According to Wang et al, the radial error is formulated as R = $\sqrt{\Delta x^2 + \Delta y^2}$, where Δx is the absolute distance in the x-direction between the obtained landmark and the referenced landmark, and Δy is the absolute distance in the y-direction between the obtained landmark and reference landmark. The mean radial error is then calculated as sum of all radial errors divided by the number of subjects. In previous studies, the accuracy of AI determined with this method varied from 1.5 to 1.7 mm, with a successful detection rate 71% within 2 mm for 400 lateral radiographs and 19 cephalometric points.[7]

Absolute difference is absolute distance in mm Δx in the x-direction between the obtained landmark and the referenced landmark, and Δy is the absolute distance in the y-direction between the obtained landmark and reference landmark. Mean absolute difference in x-direction is defined as sum of all absolute differences in x-direction divided by the number of subjects. Similarly, mean absolute difference in y-direction is calculated as sum of all absolute differences for y divided by the number of subjects. This approach to estimate accuracy of AI while detecting cephalometric landmarks is more widespread in literature.[9,20,25] Moreover, it was used while training and evaluating the performance of YOLOv3 algorithms in the study of Ji-Hoon Park and et al., where he mentioned it as "*absolute distance value between the ground truth position and the corresponding automatically identified landmarks*", and calculated separately for x- and y-direction.[20] The reason why we prefer to use these parameter is that it shows whether some AI placed landmarks are prone to mistakes in vertical or horizontal direction. This information could be helpful for further AI software improvement. In addition to it, using x- and y-dimensions absolute differences is well established practice for accuracy evaluation and comparison of cephalometric points in other orthodontic fields like growth prediction studies.[27]

In Park et al study, 283 lateral cephalometric images with 19 anthropometric points were evaluated, and YOLOv3 algorithm demonstrated accuracy within 2 mm in 80.4% cases, and 3 mm for 92.0% cases.[25] Similar results were found in Hwang et al work: the mean error between AI and human examiners while tracing 283 images and 80 points were 1.46 ±2.97 mm.[20] Some other authors also consider 2 mm range as clinically acceptable error.[7,26] In our study success detection rate for AI within 2 mm of accuracy while comparing MT and AI group was 75 % and 93% within 4 mm, which is could be considered clinically acceptable.

We could identify at least two main sources of possible mistakes and inaccuracies for the performance of AI. First, the database of scanned cephalometric images that we used in the study. The reason why we preferred to study AI on “old” cephalometric images is that AI could be very helpful to save human resources and efficiently trace big datasets of available images accumulated in the Department of Orthodontics for purpose of future research. However, AI engine operates based on the collection of images that it was trained. CEPPRO software was trained on 15 000 digital contemporary lateral cephalometric images with 1:1 scale obtained from 1 cephalometric machine at Seoul National University Dental Hospital (Seoul, Korea). Therefore, using analogue scanned images with non-uniform image quality, improper photo angle, non-skeletal objects (metal ear rods from cephalostat machine) and unnecessary markings (fiducial points) can affect AI accuracy, since AI was not trained initially to recognize such objects on the image. In our study AI made unintended detection results and marked such data with "ERROR" in 38 images out of 110. We still used the coordinates from the marked “ERROR” images in our study, because after visual inspection the inaccuracies were among soft tissue points, rather than skeletal points.

Second factor that influences AI inaccuracy is operator’s mistake while calibrating images in AI software. To extract the coordinates of the AI placed cephalometric points, the images should be calibrated in the CEPPRO software by human operator. The operator is supposed to put a digital ruler, showing AI how many pixels are in 1 mm. The operator can also introduce a mistake while stretching out the ruler, and as a result change coordinates for the landmarks. Even the smallest mistake in putting digital ruler alters the number of pixels in one mm for the computer and can influence the coordinates for all points.

1. CONCLUSION

Successful detection rate for AI within 2 mm of accuracy while comparing MT and AI group was 75 %. Thus, AI is a promising tool to facilitate cephalometric tracing process in routine clinical practice and analyzing big databases for research purposes to make it more affordable. However, cation must be taken in assessing quality of images, specifically the presence of different artifacts, like cephalostat or chin-cup contours. AI can analyze only something that it was trained before. Therefore, no new elements should not be introduced to a machine to avoid mistakes or the machine should be trained to recognize these elements on the lateral cephalometric images from the very beginning.

CITED LITERATURE

1. Proffit W, Fields H. Contemporary orthodontics. St. Louis: Mosby, Inc. 2000.

2. Graber T, Vanarsdall R. Orthodontics: current principles and techniques. St. Louis: Mosby, Inc. 2000.

3. Keim RG, Gottlieb EL, Nelson AH, Vogels III DS. Study of orthodontic diagnosis and treatment procedures. Part 1. Results and trends. Journal of Clinical Orthodontics. 2002; 36: 553–568.

4. Rischen RJ, Breuning KH, Bronkhorst EM, Kuijpers-Jagtman AM. Records needed for orthodontic diagnosis and treatment planning: a systematic review. PLoS One. 2013 Nov 12; 8(11): e74186.

5. Durão AR, Pittayapat P, Rockenbach MI, Olszewski R, Ng S, Ferreira AP, Jacobs R. Validity of 2D lateral cephalometry in orthodontics: A systematic review. Prog Orthod. 2013 Sep 20; 14: 31.

6. Manosudprasit, Amornrut & Haghi, Arshan & Allareddy, Veerasathpurush & Masoud, Mohamed. (2017). Diagnosis and treatment planning of orthodontic patients with 3-dimensional dentofacial records. American Journal of Orthodontics and Dentofacial Orthopedics. 151. 1083-1091. 10.1016/j.ajodo.2016.10.037.

7. Evaluation and Comparison of Anatomical Landmark Detection Methods for Cephalometric X-Ray Images: A Grand Challenge. Wang CW, Huang CT, Hsieh MC, Li CH, Chang SW, Li WC, Vandaele R, Marée R, Jodogne S, Geurts P, Chen C, Zheng G, Chu C, Mirzaalian H, Hamarneh G, Vrtovec T, Ibragimov B IEEE Trans Med Imaging. 2015 Sep; 34(9):1890-900.

8. El-Fegh, I., Galhood, M., Sid-Ahmed, M., Ahmadi, M.: Automated 2-D cephalometric analysis of X-ray by image registration approach based on least square approximator. In: 30th International Conference of the IEEE Engineering in Medicine and Biology Society - EMBS 2008. 3949–3952 (2008).

9. Cardillo, J., Sid-Ahmed, M.A.: An image processing system for locating craniofacial land- marks. IEEE Trans. Med. Imaging 13, 275–289 (1994).

10. Forsyth, D.B., Davis, D.N.: Assessment of an automated cephalometric analysis system. Eur. J. Orthod. 18, 471–478 (1996).

11. Lévy-Mandel, A.D., Venetsanopoulos, A.N., Tsotsos, J.K.: Knowledge-based landmarking of cephalograms. Comput. Biomed. Res. Int. J. 19, 282–309 (1986).

12. Bulat Ibragimov, Boštjan Likar, Franjo Pernuš, and Tomaž Vrtovec Ibragimov B.

“Automatic cephalometric x-ray landmark detection by applying game theory and random forests,” in Proc. ISBI Int. Symp. on Biomedical Imaging. adfa, p. 1, 2014. © Springer-Verlag Berlin Heidelberg 2014.

13. Shantanu Chakrabartty and Masakazu Yagi and Tadashi Shibata and Gert Cauwenberghs Robust cephalometric landmark identification using support vector machines, ICASSP, 2003.

14. Collobert, R., Weston, J. (2008, July). A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning (pp. 160-167). ACM.

15. Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., Kingsbury, (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine, 29(6), 82-97.

16. Alipanahi, B., Delong, A., Weirauch, M. T., Frey, B. J. (2015). Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning. Nature biotechnology.

17. A benchmark for comparison of dental radiography analysis algorithms.Wang CW, Huang CT, Lee JH, Li CH, Chang SW, Siao MJ, Lai TM, Ibragimov B, Vrtovec T, Ronneberger O, Fischer P, Cootes TF, Lindner C Med Image Anal. 2016 Jul; 31():63-76.

18. AAOF Legacy Collection.org [Internet]. St. Louis: American Association of Orthodontists Foundation, Inc; c2013 [updated 2013]. Available from: http://www.aaoflegacycollection.org.

19. Baumrind S, Curry S. American Association of Orthodontists Foundation Craniofacial Growth Legacy Collection: Overview of a powerful tool for orthodontic research and teaching. Am J Orthod Dentofacial Orthop 2015; 148:217e25. Available from: http://dx.doi.org/10.1016/ j. ajodo.2015.06.

20. Hwang, H. W., Park, J. H., Moon, J. H., Yu, Y., Kim, H., Her, S. B., ... & Lee, S. J. (2020). Automated identification of cephalometric landmarks: Part 2-Might it be better than human? *The Angle Orthodontist*, *90*(1), 69-76.

21. Talaat, S., Kaboudan, A., Talaat, W., Kusnoto, B., Sanchez, F., Elnagar, M. H., ... & Bourauel, C. (2020). Improving the accuracy of publicly available search engines in recognizing and classifying dental visual assets using convolutional neural networks. *International Journal of Computerized Dentistry*, *23*(3), 211-218.

22. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning (Adaptive Computation and Machine Learning series).

23. Szeliski, R. (2010). *Computer vision: algorithms and applications*. Springer Science & Business Media.

24. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

25. Park, J. H., Hwang, H. W., Moon, J. H., Yu, Y., Kim, H., Her, S. B., & Lee, S. J. (2019). Automated identification of cephalometric landmarks: Part 1- Comparisons between the latest deep-learning methods YOLOV3 and SSD. *The Angle Orthodontist*, *89*(6), 903-909.

26. Arik, S. Ö., Ibragimov, B., & Xing, L. (2017). Fully automated quantitative cephalometry using convolutional neural networks. *Journal of Medical Imaging*, *4*(1), 014501.

27. Sagun, M., Kusnoto, B., Evans, C. A., Galang-Boquiren, M. T., Viana, G., & Obrez, A. (2015). Evaluation of Ricketts' and Bolton's growth prediction algorithms embedded in two diagnostic imaging and cephalometric software. *Journal of the World Federation of Orthodontists*, *4*(4), 146-150.

28. *AAOF Legacy Collection Home Page*. (n.d.). Retrieved February 23, 2021, from https://www.aaoflegacycollection.org/aaof_home.html

29. Brownlee, J. (2019, April 16). How Do Convolutional Layers Work in Deep Learning Neural Networks? *Machine Learning Mastery*. https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/

30. *Oral Health | Healthy People 2020*. (n.d.). Retrieved February 23, 2021, from https://www.healthypeople.gov/2020/topics-objectives/topic/oral-health

APPENDIX

Notice of Determination
Activity Does Not Represent Human Subjects Research

October 29, 2019

Galina Bulatova
Orthodontics

RE: **Protocol # 2019-1180**
"Assessment of Automatic Cephalomertic Landmarks Identification Using Artificial Intelligence Deep Learning Convolutional Neural Networks"

Sponsor: **None**

Dear Dr. Bulatova:

The UIC Office for the Protection of Research Subjects received your application, and has determined that this activity **DOES NOT meet the definition of human subject research** as defined by 45 CFR 46.102(e)/ 21 CFR 50.3(g) and 21 CFR 56.102(e).

Specifically, comparative analysis of publicly available lateral radiographs from the American Association of Orthodontists Foundation collection with cephalometric landmarks identified, un-identified and analyzed by orthodontist experts (with inter- and intra-rater reliability calibrations), and un-identified and analyzed by an automated AI engine made available by DDH Inc.

You may conduct your activity without further submission to the IRB.

Please note:

- If this activity is used in conjunction with any other research involving human subjects, prospective IRB approval or a Claim of Exemption is required.
- If this activity is altered in such a manner that may result in the activity representing human subject research, a NEW Determination application must be submitted.

Sincerely,
Sandra Costello
Assistant Director, IRB # 7
Office for the Protection of Research Subjects

cc: Veerasathpurush Allareddy, Orthodontics, M/C 841
Flavio J. Sanchez (faculty advisor), Orthodontics, M/C 841

VITA

NAME:	Galina V. Bulatova
EDUCATION:	DMD, First Moscow Medical University, 2008. Graduated summa cum laude. Internship in General Dentistry, First Moscow Medical University, 2009 Orthodontic Residency Program, Federal University of Postgraduate Education, Moscow, 2012
HONORS:	Medal as the best dental student for academic achievements. Department of Orthodontics UIC Scholarship for clinical excellence and academic achievements.
PROFESSIONAL MEMBERSHIP:	American Association of Orthodontists