A NOVEL FULLY AUTOMATED CEPHALOMETRIC LANDMARK DETECTION SOFTWARE MODEL COMPARED WITH MANUAL ANNOTATION METHOD

By

Dr JISHNU S

Dissertation submitted to the

Kerala University of Health Sciences, Trissur

In partial fulfilment of the requirements of the degree of

MASTER OF DENTAL SURGERY

IN

ORTHODONTICS AND DENTOFACIAL ORTHOPEDICS

Under the guidance of

Dr Tony Michael

Professor

Department of Orthodontics & Dentofacial Orthopedics

St. Gregorios Dental College, Kothamangalam,

2021

ST. GREGORIOS DENTAL COLLEGE

UNDER THE MANAGEMENT OF MJSCE TRUST, PUTHENCRUZ

CHELAD, KOTHAMANGALAM, ERNAKULAM DIST, KERALA - 686681

ETHICAL CLEARANCE CERTIFICATE

SGDC/152/2019/3729

15/11/2019

To,

Dr.Jishnu S
St.Gregorios Dental College
Chelad,Kothamangalam

Dear Dr.Jishnu S,

Subject: Ethics Committee Clearance Reg.

Protocol: Evaluation of accuracy of novel Fully Automated Cephalometric Landmark Detection software by comparing with standard manual landmark annotation method.

After the Institutional Ethics Committee (IEC) held on 15th of November 2019, this study was examined and discussed. After the consideration, the committee had decided to approve and grant clearance for the aforementioned study.

The members who attended the meeting at which the protocol was discussed were:

1) Dr.C.K.K Nair - Former BARC Scientist.
2) Dr.Cinu Thomas A - Scientist, Senior lecturer, Department of Pharmaceutical Sciences Centre for Professional and Advanced Studies.
3) Dr.Lissy Jose - Former member Women's Welfare Association.
4) Adv. Jose Aranjam - Advocate.
5) Dr.Sangeeth Paul - Reader, Department of Biochemistry, St.Gregorios Dental College.
6) Dr.Eapen Cherian - Secretary.
7) Dr.Jain Mathew - Principal and Head of the Department, Department of Conservative Dentistry and Endodontics.
8) Dr.George Francis - Head of the Department, Department of Prosthodontics and Crown& Bridge.
9) Dr.Binoy Kurian- Head of the Department, Department of Orthodontics &Dentofacial Orthopaedics.

Dr. C.K.K Nair
Chairman Institutional Ethics Committee
St.Gregorios Dental College,Chelad

Dr.Eapen Cherian
Secretary

Phone : 0485-2572529 530 531 2571428 Fax : 0485-2572530
Email : sgdc@rediffmail.com, Web : sgdc.ac.in

To

My Beloved Parents

(Sri. V Sivaraman and Smt. Suseela P)

Who Are Always in My Heart

DECLARATION BY THE CANDIDATE

I hereby declare that this dissertation entitled **"Evaluation of accuracy of a novel fully automated cephalometric landmark detection software model by comparing with standard manual landmark annotation method"** is a bonafide and genuine research work carried out by me under the guidance of **Dr Tony Michael,** Professor, Department of Orthodontics and Dentofacial Orthopaedics, St. Gregorios dental college, Chelad, Kothamangalam.

Date: 23/01/2022

Place: Kothamangalam

St.Gregorios Dental College
Chelad - 686 681

Dr. Jishnu S

CERTIFICATE BY THE GUIDE

This is to certify that the dissertation entitled **"Evaluation of accuracy of a novel fully automated cephalometric landmark detection software model by comparing with standard manual landmark annotation method"** is a bonafide research work done by **Dr Jishnu S**, in partial fulfilment of the requirement for the Degree of Master of Dental Surgery.

Date: 13/01/2022

Place: Kothamangalam

Dr Tony Michael

Professor

Department of Orthodontics

St. Gregorios Dental College, Cheald

St. Gregorios Dental College Chelad - 686 681

ST. GREGORIOS DENTAL COLLEGE Dept. of Orthodontics CHELAD - 686 681

ENDORSEMENT BY THE HOD AND PRINCIPAL OF THE INSTITUTION

This is to certify that the dissertation entitled **"Evaluation of accuracy of a novel fully automated cephalometric landmark detection software model by comparing with standard manual landmark annotation method"** is a bonafide research work done by **Dr Jishnu S** under the guidance of **Dr Tony Michael**, Professor, Department of Orthodontics and Dentofacial Orthopaedics.

PRINCIPAL
St.Gregorios Dental College
Chelad, Kerala - 686 681

13/1/22

Dr Binnoy Kurian	**Prof Dr Jain Mathew**
Professor and HOD	Principal
Department of Orthodontics	St Gregorios Dental College, Chelad

ST. GREGORIOS DENTAL COLLEGE
Dept. of Orthodontics
CHELAD - 686 681

St. Gregorios Dental College
Chelad - 686 681

Date: 13/01/2022	Date: 13/01/2022
Place: Kothamangalam	Place: Kothamangalam

COPY RIGHT

DECLARATION BY THE CANDIDATE

I hereby declare that the Kerala University of Health Sciences, Thrissur shall have the right to preserve, use and disseminate in print or electronic format for academic/research purposes.

St. Gregorios Dental College
Chelad - 686 681

Date: 13/01/2022

Place: Kothamangalam.

Dr Jishnu S

TABLE OF CONTENTS

ABSTRACT

Introduction: Precise programmed quantitative Cephalometry is fundamental for orthodontics. However, manual labelling of cephalometric landmarks is tedious and subjective, which the clinician must perform. In recent years, deep learning has gained attention for its success in the computer vision field. It has achieved enormous progress in resolving image classification or image segmentation.

Aim: This paper proposes a two-step method that automatically detects cephalometric landmarks on the X-ray images and compares these values with the manual annotation method.

Methodology: Initially a patch or area from the region of interest pertaining to each landmark is extracted from the Cephalometric image by registering the testing image to training image with annotated landmarks. Then, we utilize pre-trained networks with a backbone of EfficientNetB7, which is the best in the class Convolutional Neural Network, to detect each landmark in each ROI patch. The Network directly detects the coordinates of the landmarks. The method was assessed on two datasets: 1) ISBI 2015 Grand Challenge in Dental X-ray Image Analysis, 2) Dataset created in association with St Gregorios Dental College.

Result: The EfficientNetB7 obtains detection accuracies similar to the manual annotation method in the R2 Score.

Conclusion: The new method outperformed other benchmarks' results of previous models, which proves that the proposed method is effective for cephalometric landmark detection. The proposed method could be used for landmark detection in clinical practice under the supervision of clinicians.

Keywords: Dental X-ray images, Cephalometric landmarks, Convolutional Neural Networks, Deep learning, EfficientNetB7.

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

One of the significant goals of Orthodontics is to resolve craniofacial discrepancies and to establish functional and esthetic demands. Orthodontists have adopted cephalometric measurement to assess the deviated skeletal parameters. Since it was introduced into orthodontics during the 1930s.

Cephalometry had its beginnings in craniometry. Craniometry is defined in the Edinburgh encyclopedia of 1813 as “the art of measuring skulls of animals to discover their specific differences.” For many years anatomists and anthropologists were confined to measuring craniofacial dimensions using the skull of long-dead individuals. Although precise measurements were possible, Craniometry had limitations for growth studies. Cephalometry is concerned with the measurement of hard and soft tissues. However, this procedure had its limitations due to the inaccuracies that resulted from measuring the skull through the varying thickness of soft tissues. With the discovery of X-rays by Roentgen in 1895, radiographic Cephalometry came into being. It is defined as the cranial measurements from the hard and soft tissue landmarks on the radiographic image (Krogman & Sassouni 1957). This approach combines the advantages of Craniometry and anthropometry. The disadvantage is that it produces a two-dimensional image of a three-dimensional structure.

Cephalometric radiographs are widely used in orthodontics, orthopaedics, and maxillofacial surgery to assess and predict craniofacial growth, plan treatment and evaluate treatment effects. Cephalometric analysis is widely considered a critical diagnostic tool in determining the treatment outcome. (1)

A detailed analysis should be done using a high-resolution two-dimensional x-ray image of the head taken from the side called Lateral Cephalogram to achieve these goals. Cephalometric analysis is done using a group of approved points named craniofacial landmarks. In orthodontics, there are around 90 landmarks, of which 30 are commonly used. The position of the landmarks is decided by a group of predetermined geometrical shapes, lines, intersections, and exterior boundaries.

Once the landmarks are located, the measurement and analysis of various angular and linear parameters can be performed. Measurements obtained based on the landmarks provide supportive information for the operator to determine the optimal treatment plan. The more information collected, the better will be the treatment outcome. (2)

In the manual landmarking method, Orthodontists usually trace out the craniofacial contours first on the X-ray images and then extract the landmarks from corners, line intersections, and other geometrical line and shapes. This process is tedious and requires much time. The fatigue level of the clinician will invariably influence the accuracy of the values. Moreover, high intrapersonal and interpersonal variations of landmark tracing are other problems that can lead to errors in orthodontic problem diagnosis and consequently the treatment planning and decision making. (3) Therefore, a stable and consistently automated end-to-end analytic method is required for precise evaluation if Cephalograms.

In the current scenario, the manual landmarking technique is time-consuming, and the fatigue level of the clinician will invariably triger intra-observer errors. The computer-assisted cephalometric analysis has certain shortcomings. The calibration and identification of landmarks got to be done manually. Auto-identification of anatomical landmarks is complex and poorly explored due to the complexity and variability of cephalometric images. Challenges in auto-identifying anatomical landmarks of cephalometric images include variations on individual skeletal structures, the image blurs caused by projection magnifications and, image complexity due to the overlapping contralateral structures. Many types of research are underway to digitalize the cephalometric analysis, as their accuracy remains an issue compared to the manual gold standard method. The dataset available for the deep learning method is another limitation.

There have been various studies for lateral cephalometric analysis. In particular, the International Symposium on Biomedical Imaging (ISBI) held in 2014 and 2015 challenged this problem and several approaches were published(4,5). Ibragimov et al. used Haar-like features to express the landmark’s intensity and linked it to a point detector using a random forest; the landmark is found by using random forest regression(6). In the second stage, the landmark is modified by a sparse shape composition model. The Model of Chen et al. learned the visual characteristics of the image patches and the distance from the landmarks and made a prediction model by voting the landmarks obtained from each patch(7). Vandaele et al. solved this problem with each of the 19 landmarks binary classification problems(8). They used highly randomized trees as pixel classifiers. Despite the wide variety of studies, no accurate model has yet been developed for use in clinics with less than 2mm(4).

A fully automated cephalometric landmark detection software enables us to do accurate calculations in minimum time. The software evaluated during this study uses Artificial Intelligence technology (Deep learning) to execute the cephalometric analysis. Thanks to its deep brain algorithm, it simplifies and fastens the landmark identification process. The software improves after every analysis performed, thereby acquiring perfection over time. In recent years, deep learning has outperformed existing algorithms in various areas. Especially since the AlexNet in ILSVRC in 2012 (9), Convolutional Neural Network (CNN) has been developed rapidly in image processing. CNN is a multi-layered perceptron model inspired by animal visual systems (10). The color images given as input to the image processing problem are represented in a three-dimensional array inside the computer. For high-resolution images, one image is represented by many numbers. CNN has the characteristics of local connections and shared variables. This property allows spatial properties from images with few parameters. Therefore, CNN enables us to get specific information efficiently from images.

CNN has been widely applied to medical imaging (11), image segmentation (12,13), object/lesion detection (14,15), image/exam classification (16), and registration. Some papers find landmarks in medical images. Payer et al. used CNN to find multiple landmark points. They first defined the location of the landmark as a heatmap using Gaussian (17). Then, the landmark was estimated in the learning process by learning the heatmap from the input image. Arik et al. solved the problem of cephalometric landmarks detection using CNN. Their idea is to find intensity appearance patterns for each landmark (18). This method showed that the CNN-based method is better than the random forest-based methods.

This study builds and evaluates a State-of-the-Art Artificial Intelligence(AI) -based landmark detection Convolutional Neural Network (CNN) model software, The proposed Model’s performance is tested and compared with the manual landmark annotation method. Six hundred publicly available images and well-defined landmarks were used for the study.

RELEVANCE OF THE RESEARCH

The landmark detection is done manually by the clinician currently. This method has many variables, which can affect the accuracy of the analysis. These problems suggest the requirement of an accurate and predictable system for landmark detection in cephalometric analysis.

The CNN model used in this study is more efficient than the currently available architectures, which use less time for processing and reduce the amount of dataset required for precisely detecting the Cephalometric landmarks.

AIM OF THE STUDY

To evaluate the accuracy of a novel Fully Automated Cephalometric Landmark Detection Software compared to the manual landmark annotation method.

OBJECTIVES OF THE STUDY

1. To annotate the 19 anatomical landmarks on 600 Cephalograms manually by two orthodontists having a minimum clinical experience of 10 years.
2. To test the cephalometric landmark detection software model using the same Lateral Cephalograms that are manually annotated.
3. To compare the landmarks obtained by the software and the landmarks annotated manually.

BACKGROUND AND REVIEW OF LITERATURE

Cephalometry:

The name cephalometry is the morphological study of all the structures present in the head. This age, race, and sex are valuable for diagnosis, treatment monitoring and predicting orthodontic treatment results.

Broadbent(19) and later **Brodie**(20) applied a method based on landmarks to quantify malocclusions. Landmarks were defined using cephalometric radiographs, where some bony or soft tissue structures had to be identified. The power of this approach comes when standard values are known for a specified measurement of an individual at known age, race, and sex, So that differences can be quantified and used for diagnosis.

Downs(21) in 1948 introduced the first cephalometric analysis method. He selected ten angular measurements on Lateral Cephalograms from a group of selected individuals, taking average values, giving them clinical relevance. Downs analysis has been the basis for most methods used at present, such as Steiner's and Rickett's (22) methods.

To identify cephalometric landmarks on a particular patient, lateral head radiographs are taken. Lately, some work has been carried out to study potential advantages of 3D imaging methods, such as computed tomography for analysis, and concluded that 3D landmarking present a benefit only on patients with severe asymmetric craniofacial syndrome. No substantial benefits are obtained that compensate for the high costs and problems of 3D imaging methods in clinical routine, The conventional radiographs and few analysis still remains the gold standard for imaging analysis in Orthodontic analysis.

The automation of lateral cephalogram for landmark detection was started late back in the 1980s; the first step towards an automated extraction of these points was presented in work by **Levy-Mandel A D (1986)** by a knowledge-based technique with edge tracking. This technique uses a global line following /line extracting technique to locate all the existing lines and edges of an image and then uses a knowledge base to select the relevant landmarks. (23) Later, in 1989, Parthasarathy et al. created an algorithm that uses digital image processing and feature recognition techniques to detect 27 different landmarks in the lateral Cephalogram. (24)

Forsyth D B et al. (**1996)** compared the diagnostic quality of conventional cephalometric with that of digital image counterparts. In this, they suggested that, for digital imaging of cephalometric radiographs, a pixel matrix larger than 512x512 with more than 64 Gray levels is required to maintain the diagnostic quality of the original radiograph. (25)

B Trpkova (1997) done a meta analysis. According to this investigation, for the landmark detection, they recommend that 0.59 mm of total error for the X coordinate and 0.56 mm for the Y coordinate are acceptable levels of accuracy. (26)

Rudolph D J et al. (1998) shows a model-based approach for automated computer identification of landmarks using Spatial spectroscopy (SS) is a computerized method that identifies image structure based on a convolution of the image with a set of filters followed by a decision method using statistical pattern recognition technique concludes that SS shows potential for the automatic detection of landmarks. (27)

Jia-Kuang Liu (2000) evaluated the accuracy of a computerized automatic landmark identification system that used an edge-based technique. They divided the image into eight rectangular sub-image regions, reduced resolutions of the sub-image, and performed the landmark detection. Only selected landmarks show a comparable result with the manual method. (28)

Hutton T J et al. **in 2000** evaluated active shape models (ASM) to cephalometric landmarking. The study showed that the tools could be used as a time-saving tool to provide a first-estimate location of the landmarks. The method provided a framework for a range of future improvements. (29)

Grau V et al. (2001) utilized a pattern-matching technique; Landmark detection is carried out in two steps: a line detection module searches for significant, well-contrasted lines of the image, such as the jawline or the nasal spine. The landmark detection module uses the lines in the first module to determine the search areas and then applies a pattern detection algorithm based on mathematical morphology techniques. Relations between landmarks and lines are determined utilizing a training process. The system has been tested to detect 17 landmarks on 20 images: more than 90% of the landmarks are accurately identified. (30)

El-Feghi.I et al. (2003) tried the soft computing pattern-matching method in two steps: deriving a smaller expectation window for each landmark using a trained neuro-fuzzy system (NFS) then applying a template-matching algorithm to pinpoint the exact location of the landmark. The system is trained to locate 20 landmarks on a database of 565 images which shows improved results than the previous works. (31)

Yi-Jane Chen et al. (2004) evaluated the effect of differences in landmarking on traditional versus digitalized cephalometry. The results show significant differences but within the clinically acceptable level and substantiated the benefits of digital cephalometry in terms of reliability on analysis. (32)

Mohseni and Hadis (2007) used image processing and pattern matching techniques to locate the three prominent reference landmarks on each image which are then used to form an affine matrix. This matrix is used to estimate the initial location of the other landmarks. The proposed method finds proper initial estimations, limits the search regions, and precisely obtains the location of landmarks. More than 90% accuracy ensures the efficiency of the proposed method. (33)

Rosalia Leonardi et al. (2008) concluded a Systemic review that the systems described in the literature, such as image filtering plus knowledge-based landmark search, model-based approaches, soft-computing approaches, and hybrid approaches are not accurate enough to allow their use for clinical purposes as errors in landmarking detection were more significant than those expected with manual tracing. (34)

Ibragimov B (2014) proposed a novel framework method using Random forests (RFs) and uses the concept of game theory to determine the optimum landmark points. The results indicate that the proposed framework can be used for computerized cephalometry. Moreover, the framework is universal and can be applied to images of various anatomical structures acquired by different modalities. (6)

Mirzaalian.H.et al. (2014), Random decision forest-based likelihoods model is used using 200 sample sizes lateral cephalograms for locating 19 landmarks automated, got a success detection rate of 65.26%. (35)

Wang et al. (2015) evaluated the methods submitted for the Automatic Cephalometric X-ray landmark detection challenge held at the IEEE International Symposium on Biomedical Imaging 2014 with an on-site competition. The experimental results show that three methods can achieve detection rates greater than 80% using the 4 mm precision range. However, only one method achieves a detection rate greater than 70% using the 2 mm precision range, The acceptable precision range in clinical practice. The study provides insights into the performance of different landmark detection approaches under real-world conditions and highlights the achievements and limitations of current image analysis techniques. (36)

Linder. et al. (2016), used Modified Fully automated landmark annotation (FALA), which follows a machine learning approach using Random Forest regression-voting and Constrained local model framework (RFRV-CLM) to locate individual landmarks with 400 sample size images. They showed a state-of-the-art result with four-fold cross-validation on all images in the dataset with an accuracy of 84.7%. (37)

Arik. et al. (2017), used a custom deep CNN combined with a shape model for refinement for landmark detection of 19 points from the 2D lateral cephalometric image with a sample size of 400 lateral cephalometric images; the overall framework demonstrates high anatomical landmark detection accuracy of 75% and high anatomical type classification accuracy. (18)

O'Neil. et al. (2018), Custom Fully Conventional Neural Network (FCN) and Atlas Correction with a sample size of 22 images. Compared with the Decision Forest model, and outperforms that model without additional engineering and attains similar agreement to human observers with landmark detection. (38)

Wang. et al. 2018, used Multiresolution Decision tree Regression Voting created an accuracy of 73.37% landmark detection rate within the range of 2.0mm. (36)

Dai. et al. (2019), Adversarial encoder-decoder networks model is used with Cropping and template matching done for data processing got a 35-40% landmark detection rate for each landmark. (39)

Chen. et al. (2019), used Visual Geometric Group (VGG)-19, ResNet20 and inception; Custom attentive feature pyramid fusion module with a sample size of 400 lateral cephalometric images. In their framework, The AFPF module gets high resolution and semantically enhanced fusion feature to improve prediction accuracy. The pixel-wise regression-voting technique based on heat maps and offset maps also benefits the performance. (40)

Lee. et al. (2019), Custom CNN for Region of Interest (ROI) and custom Bayesian CNN for landmark detection based on 400 sample size images. The proposed model successfully identified hard tissue landmarks within the error range of 1.32±3.5mm and soft tissue landmarks with a mean success rate of 1.16±4mm and with a mean success rate of 75.2%. This model reduces the landmarking time from 5-7min by an orthodontist to 21 sec for 33 landmarks. (41)

Noothout. et al. (2019), Custom FCNs based on ResNet34, the proposed method can localize landmarks in 2D and 3D medical images of arbitrary size, acquired with three different imaging modalities and depicting different anatomical coverage. The method localizes multiple or single landmarks with high accuracy and speed, making it suitable for application in studies including many images or real-time localization. (42)

Zhong. et al. (2019), used 2-stage (global and local), U-Net models, with 400 sample size images. The attention-guide and comprehensive exploitation strategy ensure that the searching scopes are smaller and data resolution is higher with minimum information redundancy. They achieved a state-of-the-art result of 86.74% accuracy on landmark detection in cephalometric radiography. (43)

Park. et al. (2019), Comparative study on You-Only-Look-Once version 3 (YOLO V3) and Single Shot Multibox (SSD) using 1311 sample size images. YOLO v3 seems to be more promising as a fully automated landmark identification system for use in clinical practice with an accuracy of 80.4%. (44)

Gilmour. et al. (2020), used Modified ResNet34 combined with a custom image pyramids approach (spatialized features) using 400 sample size images. Using this model, they got an accuracy of 88.32% in test 1 and 77.05% in test 2. (45)

Kim. et al. (2020) They tried 2 Stage DNN using a stacked hourglass network model with 2500 sample size images detected 23 landmarks achieved a landmark detection success rate of 82-84%. (46)

METHODOLOGY

Materials:

1. Dataset 1: includes 300 lateral cephalograms collected from 2015 ISBI Grand Challenge conducted by IEEE.
2. Dataset 2: includes 300 lateral cephalograms collected from St Gregorios Dental College, Ernakulam, Kerala.
3. Digital Cephalometric X-ray Machine- Orthophos XG 3D
4. Dentsply Sirona Sidexis ver.4.0 software.
5. Marked dataset containing 600 files with 19 landmarks annotations.
6. Unmarked dataset containing 600 files
7. Pytorch Framework for CNN building with Torch Vision and OpenCV using python language.
8. Asus Rog laptop, 32GB ram, 8GB RTX 2070 Super Max-Q, Intel 10^{th} Gen CPU

Description of Datasets:

The CNN needs a considerable amount of data to work efficiently; otherwise, there will be an over-fitting issue. The Data used in this study was provided in the 2015 ISBI Grand Challenges in Dental X-ray Image Analysis of IEEE International Symposium on Biomedical Imaging 2015 (website: http://www-o.ntust.edu.tw/~cweiwang/ISBI2015/challenge1/). This was combined with the dataset created from the St Gregorios Dental College itself. The challenge dataset contains 300 images, with a resolution of 1935 x 2400 pixels in TIFF format, each pixel's size is 0.1 x 0.1 mm. The cephalograms were acquired with Soredex CRANEXr Excel Ceph machine (Tuusula, Finland) and Soredex SorCom software (3.1.5, version 2.0). The downloaded file contained 3 folders. The first folder had x-rays with landmarks already identified. The second folder had x-rays without having the landmarks. The third folder had the coordinates of the marked landmarks for every 300 images.
The custom-created dataset containing 300 patient images between the ages of 6 and 60 years were collected with Orthophos XG 3D Digital X-ray machine and Dentsply Sirona Sidexis (version 4.0) software from the Department of Orthodontics, St Gregorios dental college (Kothamangalam, Kerala, India). Two Orthodontists manually annotated the collected

images, and the coordinates were listed for every 19 landmarks in the same way as the Grand challenge dataset.

Thus, a total of 600 images were taken for this study after exclusion criteria.

Inclusion criteria

- Digital lateral cephalograms of patients having Angle's Class I, II, III skeletal pattern.

Exclusion criteria

- Digital lateral cephalograms of patients with a history of facial trauma
- Digital lateral cephalograms of patients with Syndromes and or other craniofacial anomalies.
- Digital lateral cephalograms of patients with gross asymmetry
- Digital lateral cephalograms of patients undergone surgical corrections.
- Digital lateral cephalograms of poor quality that does not allow manual tracing.

Proposed method:

As described above, the dataset collected contained six hundred X-ray images, and each image contained Nineteen landmarks. The dataset collected from the Grand challenge contained 300 X-rays which 2 experienced doctors had manually annotated and, the coordinates are listed in tables for 19 landmarks. The Orthodontists with a clinical and academic experience of more than 10 years were selected to augment the database of the present study. They analysed the 300 cephalograms collected from St Gregorios Dental College, and manually marked the 19 Cephalometric landmarks listed in Figure 1. After obtaining the marked x-rays, the landmarks' x,y coordinates were automatically extracted by the computer. Once the Dataset was created, the Cleaning of the dataset was done i.e., the dataset was analyzed for errors of different types and were removed. The mean error between the two Orthodontists was calculated by the RMSE score listed in Table 1 to analyze the inter-examiner reliability, and find the ground truth data for each landmark. After making sure that the dataset is clear without any error, the training of the model was done. 80 per cent of the dataset were used for training purposes and the remaining dataset was used for testing the model. During the training time, the images were studied by the model. After adequate training, the testing dataset is given to the model for identifying the landmarks. This method is a two-step method: 1) ROI extraction, 2) Landmark detection. The cropped patches were by registering the test image to training images for ROI extraction, with annotated landmarks. Later the pre-trained CNN models with the backbone of EfficientNetB7, a state-of-the-art CNN, were used to detect the landmarks in the extracted ROI patches. Once the model detected the landmarks, the RMSE score for each landmark is evaluated to obtain the amount of variation in detecting the landmark from the ground truth data, listed in Table 2.

The final comparison of manual and automated models was done using the R2 score.

Region of Interest Creation

Due to the large size of the input image, it was decided to extract small ROI patches automatically. ROI patches that include landmarks were extracted. The entire image was cropped into areas around each landmark. To extract the ROI automatically, the registration to locate a course location of the landmark was used. Next, the ROI patches centred on the coarse location was extracted. Thus, the landmark location to be detected was included within the ROI patch.

After registration, the landmark locations of the training images was copied to test images. The landmark location of the reference image was considered as the centre of the ROI patch to extract a 512 x 512 resolution patch image on the test image. The extracted images was treated as input to the trained CNN; thus, could detect the corresponding landmark within the patch image.

Every person's head shape varies; therefore, randomly choosing one image from the training images as a reference image was insufficient. Therefore, the landmark to be detected was not included in the ROI. To avoid this situation, All the training images were titrated, which means registration were done 300 times for one test image (the total number of training images was 300). Then, as the standard image, the training image was chosen with the minor square error with test images. This enabled to seek out the closest training sample to the test images. For computation, the trained and test images vary a lot. The shortest only took a couple of seconds for one registration, while the longest to do more than one minute. In short, the average time for registering one image to every training sample was around twenty minutes.

CNN Model

The Model program was written using Pytorch framework CNN building with Torch Vision and OpenCV using Python language. Convolutional Neural Networks (CNN) are commonly developed at a fixed resource budget and then scaled up for better accuracy if more resources are available. In CNNs, have more layers, have deeper networks, and if keeps on increasing the layers in our network, had to scale the depth of the Network. Thus, if needed to increase the accuracy, need to add more layers to create a robust architecture. Nevertheless, if keeps on increasing the layers and the saturation point is attained to some extent, the algorithm will not perform and will face a vanishing gradient problem. Later on, in further studies, have overcomes the vanishing gradient problem by inventing ResNets architecture widely studied in landmark detection. These ResNet's uses Skip Connection which reduces the vanishing gradient problem. However, as these problems are resolved, the architecture layers keep increasing, leading to a need for more powerful computer and computation power. The processing becomes time-consuming for training and testing these models. E.g., ResNet is a type of CNN model that can be scaled up from ResNet-18 to ResNet-200 by using more layers. Since the start of CNN development, we have kept increasing the layers, which means we only increase the depth of the network.

In this study, the EfficientNetB7 architecture is used as the backbone, which Google developed, and these models are used for computer vision applications. They can be effectively used to find features in ROI patches. The EfficientNetB7 architecture uses mobile inverted bottleneck convolution (MBConv) as a baseline network. This CNN can find valuable features automatically for different computer vision tasks. This model can perform scaling on depth, width and resolution. Like earlier said, the depth means increasing the number of layers in the Network. The width simply means increasing the number of channels or feature maps. The increase in the number of feature maps in a specific picture will increase the accuracy. Resolution Scaling means that if the algorithm is trained using a low-resolution image dataset, the model will capture less information, leading to less accuracy. Therefore, to get more information from the image, so it is mandatory to use a high-resolution image to learn more complex features by the algorithm. Since the high-resolution images are used, the architecture needs more depth scaling, i.e., a Deep learning neural network is required for processing complex pieces of information. To pursue better accuracy and efficiency, it is critical to balance all dimensions of networks, i.e., width, depth and resolution, during scaling; otherwise, if keep on increasing, can face the vanishing gradient problem. For the balancing stage, used a technique called Compound Scaling was used in the study.

The compound scaling formula is $f=\alpha.\beta^{\phi}.\gamma^{\phi}$, where f is the network scaling factor, α is d: depth scaling factor, β is w: width scaling factor, γ is r: resolution scaling factor. Where α,β,γ are constants that a small grid search can determine to find out the coefficient value (ϕ). Based on these ϕ values, the remaining values are founded. The compound scaling suggests that the scaling of the network should be performed using a constant ratio in all the dimensions.

First, a baseline neural network called Efficientnet-B0 is formed using Neural Architectural Search (NAS) using the machine learning technique. Once baseline network is created, further scaling of the network is done in terms of depth, width and resolution to generate a more significant model B7 to provide better accuracy. According to the studies, with a smaller number of parameters, the EfficientNets can provide better accuracy than the other currently available models. In particular, the EfficientNet-B7 achieves state-of-the-art 84.3% top-1 accuracy on ImageNet, while being 8.4x smaller and 6.1x faster on inference than the best existing CNNs. In this study, used a fully connected layer to estimate landmark location as a regression problem. First of all, flatten all the features. Then, input one fully connected layer, which directly outputs the coordinate of the landmark in the patch.

Evaluation of the Result:

The output data are analysed using RMSE and R2 Score to detect the model's accuracy and compare it with the manual annotation method respectively. The Root Mean Squared Error (RMSE) is a common metric for assessing the performance of machine learning models. It is often used to provide a metric that is related to the unit being measured. The study uses the RMSE score for comparing the results obtained in the manual annotation method and automated method. The results would give the error between the actual point and marked landmark. This way, rather than a percentage, helps the readers to understand the error better.

Calculation of RMSE Score:

For each landmark prediction, the difference between respected landmarks and the actual ground truth data is to be found and they have to square these values. After that, the mean of these squared values are found and the Square root of this mean gives the RMSE score. The output is a non-negative value, and it would be better if it is brought near zero.

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$$

'Σ' represents sum, 'yi', the predicted value for the ith observation, 'ẏ' the observed value for the ith observation, and 'N' is the sample size. The score will give us an idea of the average distance between the ground truth data values and the predicted data values. The RMSE results are shown in pixel difference which will give a clear idea about how much deviation is present from the actual landmark point to the predicted landmark point. As the pixels difference decreases, the accuracy of the predicted landmark increase which can come up to zero value.

Calculation of R2 Score:

R-squared (R2) is a statistical measure representing the proportion of the difference or variance for a dependent variable that an independent variable or variables can explain. It shows how well the dataset will fit the model. For the calculation of R squared, the correlation coefficient is obtained and the square is the result.

$$\mathbf{r} = \frac{n\,(\sum xy) - (\sum x)\,(\sum y)}{\sqrt{[n^*\,(\sum x^2 - (\sum x)^2)]\; ^*\; [n^*\,(\sum y^2 - (\sum y)^2)]}}$$

Where 'r' represents the Correlation coefficient, n is the number in a given dataset, 'x' the first variable and 'y' the second variable. The square root of the result gives an R2 score.

R2 Score shows how much information can be gathered from the images to accurately predict the landmark. The score can be ranged from values -1 to 1. As the values are closer to 1, we can say the dataset is best fitted for the model for predicting the landmarks and vice versa.

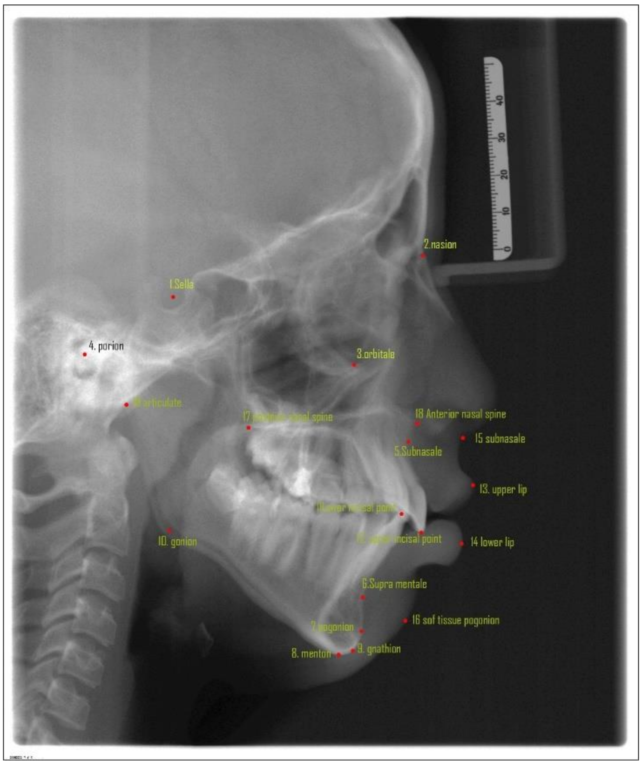

Figure 1. Landmarks selected for detection

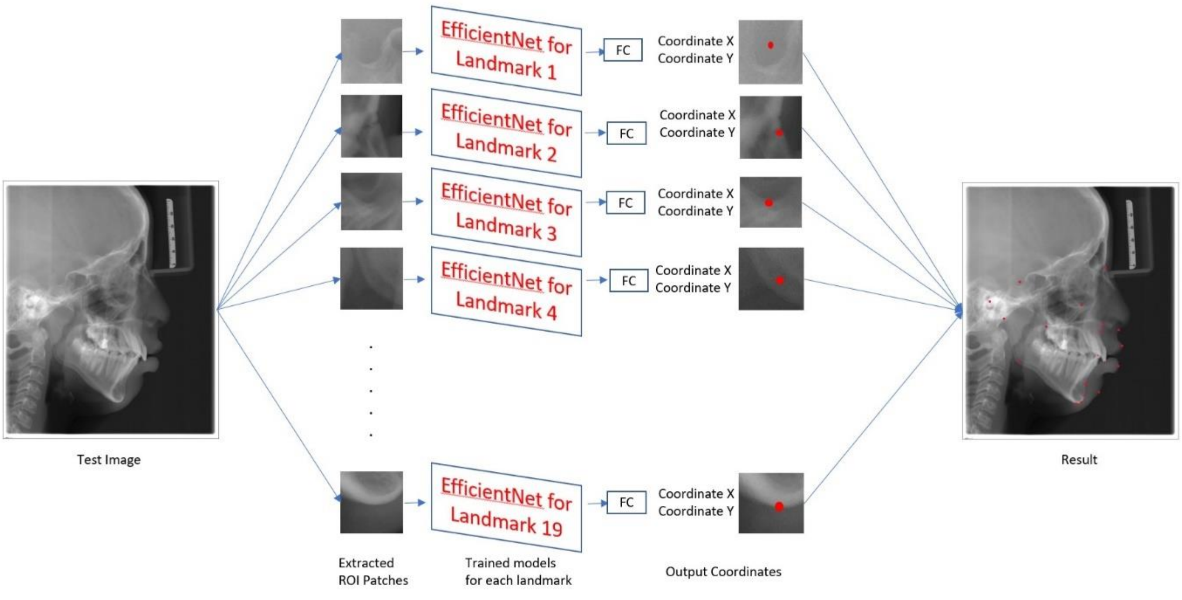

Figure 2. Trained 19 models with the same architecture but different weights

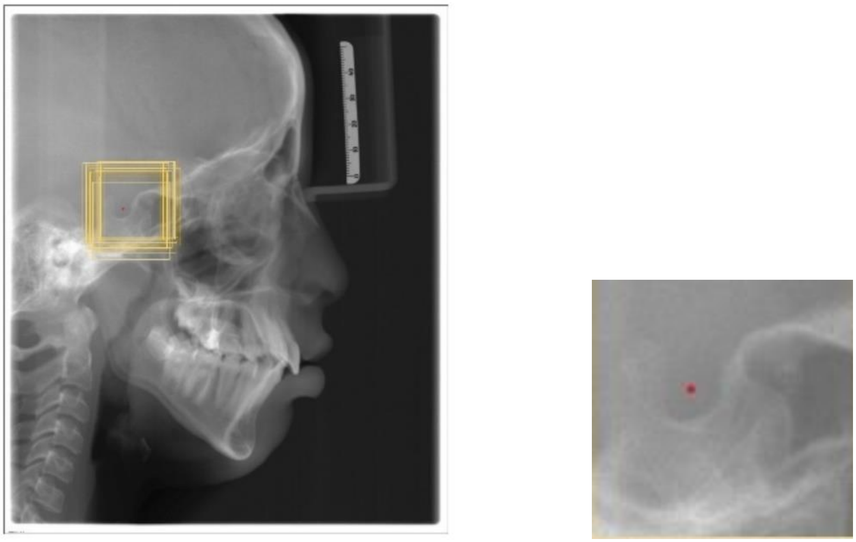

Figure 3(a) Example of our training image, yellow boxes are cropped training images, 3(b) The cropped 512x512 patches; Shows how we cropped ROI patches for each landmark: Red dot is the target landmark, the yellow boxes are 512x512 ROI patches.

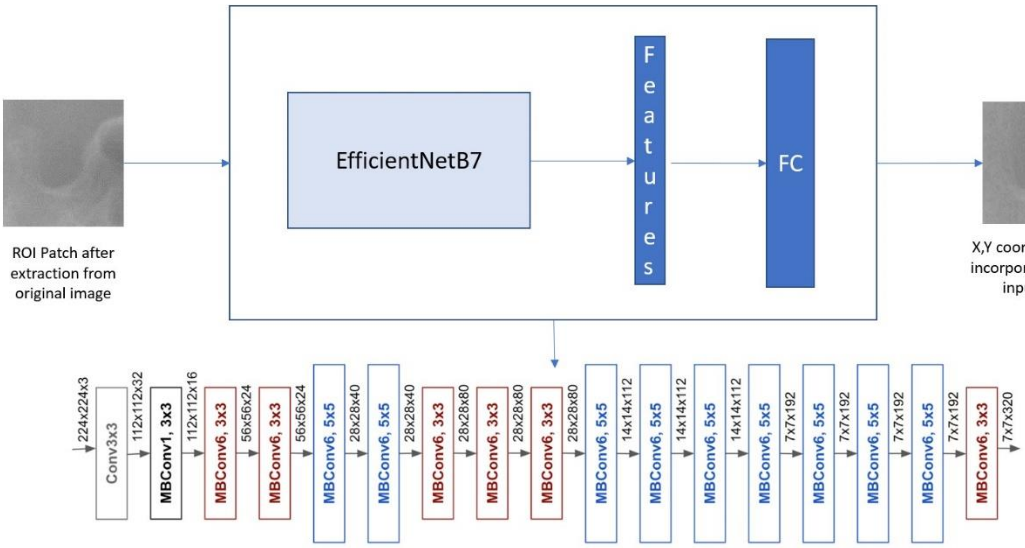

Figure 4. The architecture of the EfficientNetB7 shows the conventional layers present in the network.

RESULTS

This study trains and validates EfficientNetB7 on the public benchmark dataset from the cephalometric landmark detection challenges at IEEE ISBI 2014 and 2015 combined with a custom dataset created later. There are 600 cephalometric X-rays images collected from 600 patients. For each image, 19 landmarks are manually marked by two experienced orthodontists, and their mean deviation between two doctors is evaluated in RMSE Score listed in Table 1. The dataset contains 520 training data and 80 testing data. The resolution of an image was 1935 x 2400 pixels. The detection accuracy to evaluate the performance, EfficientNetB7 was used. If the distance between a detected landmark and its ground truth becomes less, this landmark can be classified as accurate. In table 1, points like Orbitale, porion, gonion, articulare, Point A and point B shows a moderate difference between the two orthodontists. At the same time, some points show a high degree of acceptance between the two. It is noted that the landmark, soft tissue pogonion, shows a significant difference between the 2 orthodontists.

After training the proposed model with the dataset of 520 images, the learned model was tested with 80 images. The red landmark is the predicted results of the proposed efficientNetB7. For all 19 landmarks, it is clear that the predicted results highly agree with the ground truths, and it demonstrated that EfficientNet could obtain better results than current architectures. The Mean Error calculated on CNN automated landmarking model is obtained from the established ground truth data listed in Table 2. The results were obtained using RMSE score. It will provide information about deviation of the actual landmarks from the predicted landmark in pixels. The table shows that the points marked by the model have acceptable accuracy with the manual annotation method. For points like pogonion, gnathion, menton, upper incisor point, lower incisor point; the manual method was found to have better predictability.

As for detecting the accuracy of the test data, compared EfficientNetB7 with the manual annotation method. To prove the model’s effectiveness, the results in Test data are shown in Table 3. As noted, EfficientNetB7 obtains detection accuracies similar to the manual annotation method in the R2 Score. The landmarks like, Porion, articulare, soft tissue pogonion, the model outperformed the human annotation method and provides a consistent better result. As the same time, the points like Point A, pogonion, gnathion and menton, the manual methods show more accurate results.

Table 1: Mean Deviation calculated on Observers Manual landmarking done by experienced Orthodontists

Anatomical Landmarks	RMSE Score- Mean Error pixel
Sella	8.158992995
Nasion	13.12710301
Orbitale	18.89700153
Porion	20.56398551
Pogonion	7.834805
Menton	8.064738
Gnathion	6.3035045
Gonion	18.8819135
Incisal point Upper	4.5414755
Incisal point Lower	5.1604265
Upper lip	17.1833495
Lower lip	13.8348595
Subnasale	10.2873305
PNS	9.973005
ANS	13.0483715
Articulare	18.3700165
Soft tissue pogonion	42.107541
Point A	20.8449435
Point B	18.9083315

Table 2: Error calculated on CNN automated landmarking from the established ground truth.

Anatomical Landmarks	Detection Rate	CNN Model – Mean Error (RMSE)
Sella	99.6	8.775
Nasion	95.6	14.16
Orbitale	99.6	14.22
Porion	92.3	15.975
Pogonion	93.5	9.36
Menton	95.5	10.32
Gnathion	93.3	11.31
Gonion	90.7	15.48
Incisal point Upper incisor	96.7	8.7
Incisal point Lower incisor	94.6	8.01
Upper lip	95.5	14.31
Lower lip	99.9	10.17
Subnasale	96.3	9.375
PNS	97.6	9.315
ANS	94.8	13.455
Articulare	91.5	15.165
Soft tissue pogonion	93.8	13.08
Point A	92.9	14.97
Point B	95.8	11.91

Table 3: Performance comparison between manual and automated landmark identification based on R2-Score.

Anatomical Landmarks	**Manual Landmark Accuracy (R2-Score)**	**Automated landmark Accuracy (R2-Score)**
Sella	0.944790597	0.95381
Nasion	0.928399311	0.90111
Orbitale	0.772685175	0.86632
Porion	0.442685418	0.78339
Pogonion	0.985662897	0.95729
Menton	0.987690519	0.94617
Gnathion	0.991861202	0.92341
Gonion	0.840970697	0.90562
Incisal point Upper	0.993262713	0.99738
Incisal point Lower	0.991278187	0.98637
Upper lip	0.914773372	0.98565
Lower lip	0.947982437	0.94655
Subnasale	0.963626204	0.95022
PNS	0.917825451	0.90707
ANS	0.921108837	0.93311
Articulare	0.655885858	0.75922
Soft tissue Pogonion	0.621971525	0.807231
Point A	0.834820777	0.80375
Point B	0.898344184	0.96541

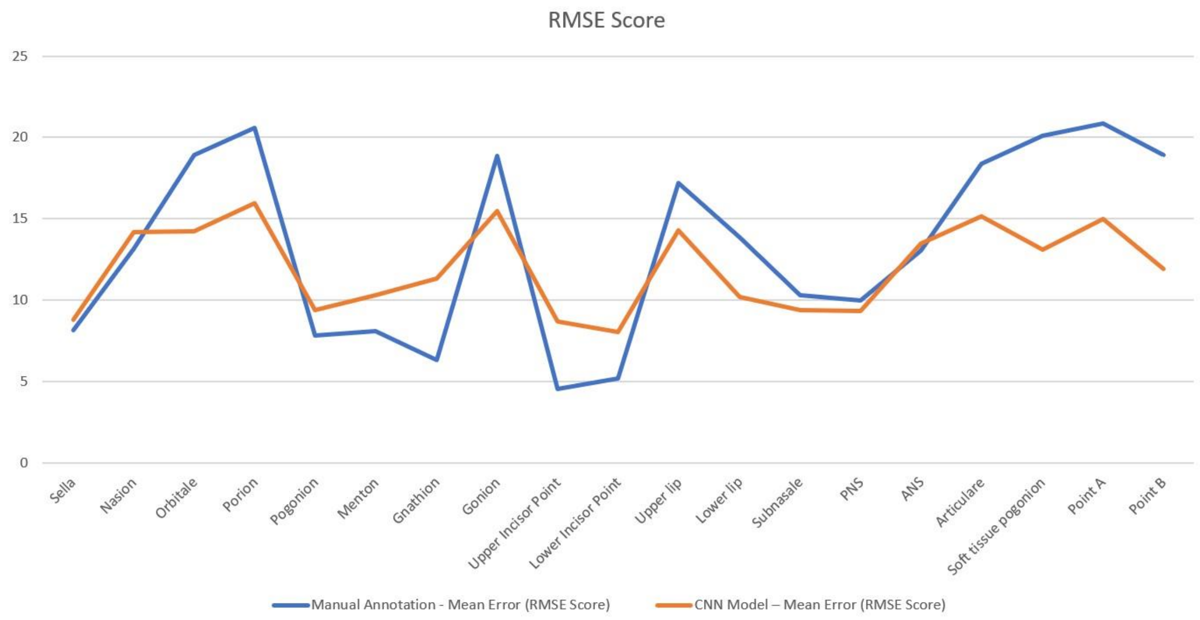

Figure 5. RMSE Score for 19 landmarks. The blue line shows the mean error that occurred during the manual annotation method; the red line shows the mean error that occurred during Automated detection of landmarks.

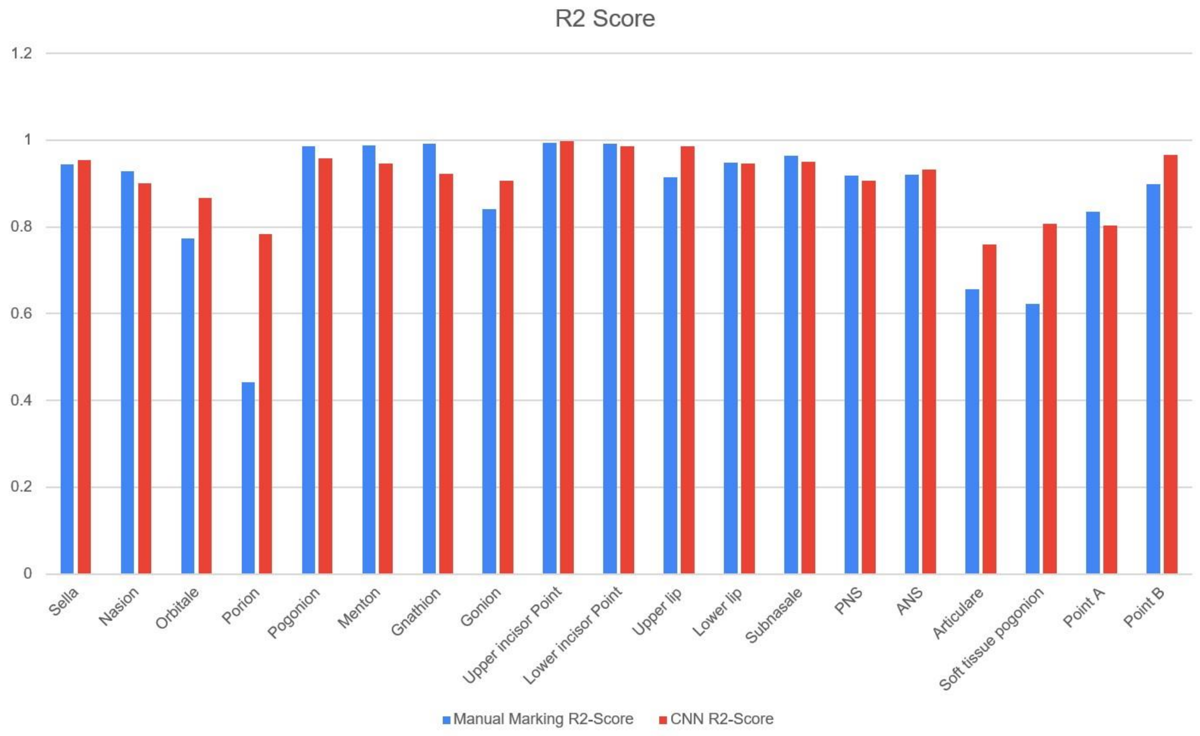

Figure 6. R2 Score for Manual and Automated landmark annotation method, the blue bar shows the performance of manual annotation, red bar shows the performance of the automated model.

DISCUSSION

Unlike other dental radiographs, Cephalograms are not limited to diagnose qualitatively. Within the relation to the reference plane (considered as stable structures), angular and linear measurements are also assessed quantitatively. The reference points may be a skeletal landmark, a virtual point or can be a constructed point like gonion (crossing point between two lines)[(47)]. For the precise diagnosis of a condition, the accurate marking of the landmarks is important[(48)]. Although Graber way back in 1956, warned of definite limitations of cephalometrics, many in this field still swear by it and its utility in orthodontic diagnostic treatment planning is widely accepted.

Landmark detection is the most vulnerable area in Cephalometrics. In many situations, the very definition of the landmark has been criticized, leading to location errors and reproducibility. Orbitale, for example, is defined as “the lowest point of the orbital rim”. Sir Martin criticized this definition stating that it does not serve as a landmark since it is not well defined and is unreliable since it is entirely an aesthetic judgement[(49)]. There is some inappropriateness in landmark detection. Each landmark has its characteristic distribution of errors too. Even when the same head film is assessed, errors in landmark identification occur and that to an extend cannot be ignored. With the lack of clarity in landmark definition and identification, the actual problem arises when these same landmarks are used to construct planes and angles. A most common example would be the angle formed between the long axis of the mandibular incisor and the mandibular plane. The apex of the lower central incisor is the most difficult point be to accurately located. Likewise, Point A is also very hard to identify and locate.

In spite of the advanced technology in this era, which includes innovations of imaging systems and software, the tools used in diagnosis and treatment planning have not experienced similar advances during the past decades. For instance, most clinicians use Cephalometrics for orthodontic diagnosis and treatment planning. But orthodontists were hesitant to do the cephalometric analysis, even after getting a lateral cephalogram because of the tedious work and time to be spend to do so. 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. These studies also stated that lateral cephalograms are time-consuming and are taken for other reasons, such as medico-legal issues, for academic purposes, or due to lack of experience of practitioners (50–52). 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 time points during treatment. This unfortunate practise 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 an average of 10-15 minutes to place landmarks manually, which makes the procedure time-consuming and subjects to errors (5,53).

The rapid development of artificial intelligence (AI) in recent years has penetrated many aspects of daily life, including the analysis of extensively 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. With orthodontic diagnostic and treatment planning, this AI technology could deliver not only an easy, practical, and precise tool for the practising clinician, but also significantly improves the amount of available labelled data. Despite of the readily available studies demonstrating the different processes to auto-detect craniofacial landmarks, most clinicians use approaches based on image-processing techniques where the image of cephalometric radiographs require intense human preparation, such as re-scaling, calibration, and labelling. Calibration and other image preparations are time-consuming and, if not done properly, often generate landmark outliners (23,54,55) as they strongly rely on the quantity and size of the cephalometric images. Many other studies propose different novel frameworks for landmark detection in cephalometric radiographs and demonstrate results with an accuracy of 72 per cent but again, these are not fully automated procedures (6,56). 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 (57–59). However, the developed methods are unable to compete with manual landmark identification. In recent years, the Institute of Electrical and Electronics Engineers (IEEE) and the International Symposium on Biomedical Imaging (ISBI) had organised 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 (4).

The study was formulated to investigate whether AI might be a viable option for the repetitive and arduous task of identifying multiple cephalometric landmarks for use in clinical orthodontic practice. The null hypothesis that there will be no difference between the manual and automated methods could not be rejected. The mean error between the automated and manual did not exceed 0.9mm. In all landmarks, the Model demonstrated accurate identification as an orthodontist. All those mean differences showing less than 2mm would not seem to be a clinically significant error. However, since the computers always detected identical positions, the reproducibility by the model upon repeated detection tasks was better than that associated with the human counterparts.

Computer vision, a part of AI, that enables machines to perceive the world similar to 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. A Convolutional Neural Network (CNN) – is a deep learning algorithm that can take an input image, assign importance to various aspects/objects in the image and differentiate one from the other. Among the machine learning methods, deep-learning methods have demonstrated superiority in automatically recognizing anatomical landmarks on diagnostic images. Deep learning means that CNN can learn different characteristics of the image, or, in other words, can be trained to understand the sophistication of the image better than traditional classification algorithms. Studies on related topics in various fields have also gained more popularity. Although three-dimensional images have gained popularity these days (60–64), two-dimensional cephalometric analysis is still a vital tool in orthodontic diagnosis and treatment planning since it provides information regarding a patient’s skeletal and soft tissue condition. Currently, computer-assisted cephalometric analysis eliminates human-induced mechanical errors. Fully automated cephalometric analysis has been long attempted to reduce the time required to obtain a cephalometric analysis, Also these long attempts to improve the accuracy of landmark identification, have reduced the errors caused by a clinician subjectively. The current studies detected less than 20 landmarks and the accuracy results were not satisfactory for use in clinical orthodontic practice. For example, in 2009, 10 landmarks on 41 digital images were identified. (65) In 2013, 16 landmarks were identified on 40 cephalometric radiographs, and the mean error from automatically identified landmarks were 2.59mm. (66) The accuracy of those automated methods were not as good as those associated with manual identification. In addition, cephalometric landmarks need not be limited to simply obtaining patients skeletal characteristics but could also be applied to plan treatment and to predict treatment outcomes, including soft tissue changes. For those purposes, hundreds of variables of anatomical landmarks, are needed. (67–69)

In the present study, the manual and automated models were compared to find whether the automated model is able to attain a landmark detection accuracy as same as humans. For that, the study was conducted using 600 raw X-ray images which were randomly collected from an age group of 6-60 years. The 300 images were collected from Open-sourced dataset from the internet and the remaining images were collected and the dataset was made indegenously and made publicly available for further studies. These 300 lateral cephalogram images that are custom created were manually annotated by two Orthodontists who had a clinical and academic experience of over 10 years. These 600 images were randomly mixed for increasing the complexity of the detection of landmarks and were categorised into two sections: 1) Training (520 X-rays), 2) testing image (80 images) datasets. The training images were used to build an architectural model from a base model called EfficientNetB7 for detecting the landmarks. The EfficientNetB7 architecture was found by Google which was far more superior than other currently available architectures for computer vision applications. It has the advantage of faster processing with less amount of data over the other currently available superior models. So, these EfficientNetB7 models are a suitable opponent for the other methods which used automated landmark detection. Once the model is made, the efficiency of this model was analysed by testing the remaining dataset. The result was obtained in pixels. The mean detection score for each of the 19 landmarks was formed for the manual annotation method and automated model using RMSE Score. By comparing the testing dataset result with the manually annotated landmarks, the accuracy of detecting the landmarks were evaluated and compared. For comparing the result, the R2 score is calculated, which shows how much variation is present for detecting a landmark by the model. In this way, The accuracy of the model can be assessed better

The learning and testing data included images from various malocclusion patients in the present study. From the first formulation of the current study, the selection of these mixed images were intended to test the model’s performance in a more complex condition, rather than identifying landmarks on images from good-looking subjects.

For registration, since people's heads vary in shape, even though the closet image to the training data was selected as the reference image for each test image, there were still missed situations. This means that after the registration, the patch we created for the test does not include the ground-truth landmark. For the ISBI dataset, there is only one missed patch, and the rate is about 0.0002. Overall, it has little impact on the results. For Testset2 of the ISBI Grand Challenge, it is seen that Landmark 3, Landmark 6, Landmark 13 and Landmark 16 have relatively low accuracy. However, the process works fine on Test1. After visualizing the testing result, it was clear that the anatomy of those failed cases is very different from the successfully detected ones.

In general, the pattern of differences between AI and orthodontists demonstrated that AI acts like an orthodontist. For example, when doctors had difficulty in identifying landmarks on poor quality images, so did AI. This might be the reason why image factors did not meaningfully affect the accuracy of AI in finding landmarks. In those subjects with fixed orthodontic appliances, massive prostheses, and surgical bone plates, it was initially anticipated that there would be difficulties in identifying the landmarks because of the multiple metallic artefacts. However, metal artefacts did not appear to have a clinically significant impact on identifying landmarks either.

As a limitation of the present study, the way AI learned during the training session and identified landmarks later in the test step are not explainable without describing computer science jargon. Although some technical details were necessary, this present study intended to focus on showcasing the results from AI. Upon repeated trials, AI always found identical positions. How much learning data might be sufficient enough to teach AI is currently unknown. Furthermore, it could be conjectured that the number of target landmarks might also be contributing factor in deciding a sufficient number of learning data. A study to elucidate the sufficient quantity of data for deep-learning of AI might be necessary for the future.

From the clinical perspective, however, AI would never replace trained specialists in orthodontics, nor might AI intend to replace a comprehensive orthodontic training program. Rather it could supplement, augment, and amplify diagnostic performance by objectively evaluating each patient seeking orthodontic treatment. The AI proposed in the present study can be compatible with the current clinical environment and would retain its validity under the constant supervision of experts in orthodontics.

LIMITATIONS OF THE RESEARCH

The deep learning models in CNN have some shortcomings like memory leakage, overfitting, which need a considerable amount of data set to improve the accuracy. Vanishing gradient problem due to overfilling of the data beyond a threshold level of the model, variation in the lateral cephalometric images leading to poor accuracy of the landmarks.

CONCLUSION

The study proposed an approach to automatically predict landmark location and used a deep learning method with very small training data, with 600 X-ray training images. The results outperformed other benchmarks' results of previous models, which proves that the proposed method is effective for cephalometric landmark detection. The proposed method could be used for landmark detection in clinical practice under the supervision of doctors.

For future work, study can be used for developing a Graphical User Interface (GUI) for doctors to use. The system offers either automatic or semi-automatic landmark detection. In the automatic mode, the system will automatically extract an ROI (based on registration) and select a proper model for each landmark. In the semi-automatic mode, the doctor needs to give a bounding box extract ROI manually and select a corresponding model for each landmark detection, which can reduce computational time. Since used simple registration in this study and it took a long time to register one test image with all the training images, around twenty minutes as we mentioned before. Under such a case, aims to design a better method that reduces the registration time in the testing phase, to make the automatic detection more efficient. For example, maybe we will also utilize the deep learning method, either to do registration or just simply to regress a coarse landmark region, to replace the rigid registration we used in this paper. The Dataset which was made in our institution made available as Open source, helps the future studies to utilize this Dataset. Moreover, because Model detected all the landmarks in patch images, and did not take global-context information (i.e., the relationship among all the landmarks) into consideration. In future work, need to utilize this global-context information to check whether this model can achieve better performance. For example, maybe will try to change the network architecture to allow the network to utilize the whole image's features to better and faster locate the coarse locations of landmarks.

BIBLIOGRAPHIC REFERENCES

1. McNamara JA. A method of cephalometric evaluation. Am J Orthod. 1984 Dec;86(6):449–69.

2. Oh K, Oh I-S, Le VNT, Lee D-W. Deep Anatomical Context Feature Learning for Cephalometric Landmark Detection. IEEE J Biomed Health Inform. 2021 Mar;25(3):806–17.

3. Kamoen A, Dermaut L, Verbeeck R. The clinical significance of error measurement in the interpretation of treatment results. Eur J Orthod. 2001 Oct;23(5):569–78.

4. Wang C-W, Huang C-T, Lee J-H, Li C-H, Chang S-W, Siao M-J, et al. A benchmark for comparison of dental radiography analysis algorithms. Med Image Anal. 2016 Jul 1;31:63–76.

5. Wang C-W, Huang C-T, Hsieh M-C, Li C-H, Chang S-W, Li W-C, et al. Evaluation and Comparison of Anatomical Landmark Detection Methods for Cephalometric X-Ray Images: A Grand Challenge. IEEE Trans Med Imaging. 2015 Sep;34(9):1890–900.

6. Ibragimov B. Automatic Cephalometric X-Ray Landmark Detection by Applying Game Theory and Random Forests. undefined [Internet]. 2014 [cited 2021 Oct 29]; Available from: https://www.semanticscholar.org/paper/Automatic-Cephalometric-X-Ray-Landmark-Detection-by-Ibragimov/952fc812851ce58f165d723d0a3e9f8f935293ed

7. Chen C, Xie W, Franke J, Grutzner PA, Nolte L-P, Zheng G. Automatic X-ray landmark detection and shape segmentation via data-driven joint estimation of image displacements. Med Image Anal. 2014 Apr 1;18(3):487–99.

8. Vandaele R, Marée R, Jodogne S, Geurts P. Automatic Cephalometric X-Ray Landmark Detection Challenge 2014: A tree-based algorithm. undefined [Internet]. 2014 [cited 2021 Nov 25]; Available from: https://www.semanticscholar.org/paper/Automatic-Cephalometric-X-Ray-Landmark-Detection-A-Vandaele-Mar%C3%A9e/adf9c236d4cd9dc1a9fb7b01d113f79bf3f5ede5

9. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017 May 24;60(6):84–90.

10. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May;521(7553):436–44.

11. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017 Dec;42:60–88.

12. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. ArXiv150504597 Cs [Internet]. 2015 May 18 [cited 2021 Nov 25]; Available from: http://arxiv.org/abs/1505.04597

13. Milletari F, Navab N, Ahmadi S-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). 2016. p. 565–71.

14. Shen W, Zhou M, Yang F, Yang C, Tian J. Multi-scale Convolutional Neural Networks for Lung Nodule Classification. Inf Process Med Imaging Proc Conf. 2015;24:588–99.

15. Kawahara J, BenTaieb A, Hamarneh G. Deep features to classify skin lesions. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI) [Internet]. Prague, Czech Republic: IEEE;

2016 [cited 2021 Nov 25]. p. 1397–400. Available from: http://ieeexplore.ieee.org/document/7493528/

16. Antony J, McGuinness K, Connor NEO, Moran K. Quantifying Radiographic Knee Osteoarthritis Severity using Deep Convolutional Neural Networks. ArXiv160902469 Cs [Internet]. 2016 Sep 8 [cited 2021 Nov 25]; Available from: http://arxiv.org/abs/1609.02469

17. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 978-3-319-46723-8, 3319467239, 978-3-319-46722-1 [Internet]. ebin.pub. 2019 [cited 2021 Nov 25]. Available from: https://ebin.pub/medical-image-computing-and-computer-assisted-intervention-miccai-2016-19th-international-conference-athens-greece-october-17-21-2016-proceedings-part-ii-978-3-319-46723-8-3319467239-978-3-319-46722-1.html

18. Arık SÖ, Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. J Med Imaging. 2017 Jan;4(1):014501.

19. Broadbent B. A NEW X-RAY TECHNIQUE and ITS APPLICATION TO ORTHODONTIA. 2009;

20. Brodie AG. On the growth pattern of the human head. From the third month to the eighth year of life. Am J Anat. 1941;68(2):209–62.

21. Downs WB. Variations in facial relationships; their significance in treatment and prognosis. Am J Orthod. 1948 Oct;34(10):812–40.

22. Steiner CC. Cephalometrics In Clinical Practice. Angle Orthod. 1959 Jan 1;29(1):8–29.

23. Lévy-Mandel AD, Venetsanopoulos AN, Tsotsos JK. Knowledge-based landmarking of cephalograms. Comput Biomed Res. 1986 Jun 1;19(3):282–309.

24. Parthasarathy S, Nugent ST, Gregson PG, Fay DF. Automatic landmarking of cephalograms. Comput Biomed Res. 1989 Jun 1;22(3):248–69.

25. Forsyth DB, Shaw WC, Richmond S, Roberts CT. Digital imaging of cephalometric radiographs, Part 2: Image quality. Angle Orthod. 1996;66(1):43–50.

26. B T, P M, N P, B N. Cephalometric landmarks identification and reproducibility: a meta analysis. Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod [Internet]. 1997 Aug [cited 2021 Oct 27];112(2). Available from: https://pubmed.ncbi.nlm.nih.gov/9267228/

27. Rudolph DJ, Sinclair PM, Coggins JM. Automatic computerized radiographic identification of cephalometric landmarks. Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod. 1998 Feb;113(2):173–9.

28. Liu JK, Chen YT, Cheng KS. Accuracy of computerized automatic identification of cephalometric landmarks. Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod. 2000 Nov;118(5):535–40.

29. Hutton TJ, Cunningham S, Hammond P. An evaluation of active shape models for the automatic identification of cephalometric landmarks. Eur J Orthod. 2000 Oct;22(5):499–508.

30. Grau V, Alcañiz M, Juan MC, Monserrat C, Knoll C. Automatic localization of cephalometric Landmarks. J Biomed Inform. 2001 Jun;34(3):146–56.

31. El-Feghi I, Sid-Ahmed MA, Ahmadi M. Automatic localization of craniofacial landmarks for assisted cephalometry. Pattern Recognit. 2004 Mar 1;37(3):609–21.

32. Yj C, Sk C, Jc Y, Hf C. The effects of differences in landmark identification on the cephalometric measurements in traditional versus digitized cephalometry. Angle Orthod [Internet]. 2004 Apr [cited 2021 Oct 27];74(2). Available from: https://pubmed.ncbi.nlm.nih.gov/15132440/

33. Mohseni H, Kasaei S. Automatic Localization of Cephalometric Landmarks. In: 2007 IEEE International Symposium on Signal Processing and Information Technology [Internet]. Giza, Egypt: IEEE; 2007 [cited 2021 Oct 29]. p. 396–401. Available from: http://ieeexplore.ieee.org/document/4458132/

34. Leonardi R, Giordano D, Maiorana F, Spampinato C. Automatic cephalometric analysis. Angle Orthod. 2008 Jan;78(1):145–51.

35. Mirzaalian H, Hamarneh G. Automatic Globally-Optimal Pictorial Structures with Random Decision Forest Based Likelihoods For Cephalometric X-Ray Landmark Detection. 2014. p. 1–12.

36. Wang S, Li H, Li J, Zhang Y, Zou B. Automatic Analysis of Lateral Cephalograms Based on Multiresolution Decision Tree Regression Voting. J Healthc Eng. 2018;

37. Lindner C, Wang C-W, Huang C-T, Li C-H, Chang S-W, Cootes TF. Fully Automatic System for Accurate Localisation and Analysis of Cephalometric Landmarks in Lateral Cephalograms. Sci Rep. 2016 Sep 20;6(1):33581.

38. O'Neil AQ, Kascenas A, Henry J, Wyeth D, Shepherd M, Beveridge E, et al. Attaining human-level performance with atlas location autocontext for anatomical landmark detection in 3D CT data. ArXiv180508687 Cs [Internet]. 2018 Sep 30 [cited 2021 Oct 12]; Available from: http://arxiv.org/abs/1805.08687

39. Dai X, Zhao H, Liu T, Cao D, Xie L. Locating Anatomical Landmarks on 2D Lateral Cephalograms Through Adversarial Encoder-Decoder Networks. IEEE Access. 2019;

40. Chen R, Ma Y, Chen N, Lee D, Wang W. Cephalometric Landmark Detection by AttentiveFeature Pyramid Fusion and Regression-Voting. ArXiv190808841 Cs [Internet]. 2019 Aug 23 [cited 2021 Oct 12]; Available from: http://arxiv.org/abs/1908.08841

41. Lee C, Tanikawa C, Lim J-Y, Yamashiro T. Deep Learning based Cephalometric Landmark Identification using Landmark-dependent Multi-scale Patches. ArXiv190602961 Cs Eess [Internet]. 2019 Jun 7 [cited 2021 Oct 12]; Available from: http://arxiv.org/abs/1906.02961

42. Noothout JMH, De Vos BD, Wolterink JM, Postma EM, Smeets PAM, Takx RAP, et al. Deep Learning-Based Regression and Classification for Automatic Landmark Localization in Medical Images. IEEE Trans Med Imaging. 2020 Dec;39(12):4011–22.

43. Zhong Z, Li J, Zhang Z, Jiao Z, Gao X. An Attention-Guided Deep Regression Model for Landmark Detection in Cephalograms. MICCAI. 2019;

44. Park J-H, Hwang H-W, Moon J-H, Yu Y, Kim H, Her S-B, et al. Automated identification of cephalometric landmarks: Part 1-Comparisons between the latest deep-learning methods YOLOV3 and SSD. Angle Orthod. 2019 Nov;89(6):903–9.

45. Gilmour L, Ray N. Locating Cephalometric X-Ray Landmarks with Foveated Pyramid Attention. ArXiv200804428 Cs [Internet]. 2020 Aug 10 [cited 2021 Oct 12]; Available from: http://arxiv.org/abs/2008.04428

46. Kim H, Shim E, Park J, Kim Y-J, Lee U, Kim Y. Web-based fully automated cephalometric analysis by deep learning. Comput Methods Programs Biomed. 2020 Oct;194:105513.

47. Athanasiou AE, editor. Orthodontic cephalometry. London: Mosby-Wolfe; 1995. 296 p.

48. Baumrind S, Frantz RC. The reliability of head film measurements: 1. Landmark identification. Am J Orthod. 1971 Aug 1;60(2):111–27.

49. Bookstein FL. Reconsidering "The inappropriateness of conventional cephalometrics." Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod. 2016 Jun;149(6):784–97.

50. Manosudprasit A, Haghi A, Allareddy V, Masoud MI. Diagnosis and treatment planning of orthodontic patients with 3-dimensional dentofacial records. Am J Orthod Dentofacial Orthop. 2017 Jun 1;151(6):1083–91.

51. Durão AR, Pittayapat P, Rockenbach MIB, Olszewski R, Ng S, Ferreira AP, et al. Validity of 2D lateral cephalometry in orthodontics: a systematic review. Prog Orthod. 2013 Sep 20;14(1):31.

52. 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.

53. 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. Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Int Conf. 2008;2008:3949–52.

54. Forsyth DB, Davis DN. Assessment of an automated cephalometric analysis system. Eur J Orthod. 1996 Oct 1;18(5):471–8.

55. Cardillo J, Sid-Ahmed MA. An image processing system for locating craniofacial landmarks. IEEE Trans Med Imaging. 1994 Jun;13(2):275–89.

56. Chakrabartty S, Yagi M, Shibata T, Cauwenberghs G. Robust cephalometric landmark identification using support vector machines. In: 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003 Proceedings (ICASSP '03). 2003. p. II–825.

57. Collobert R, Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning. In: Proceedings of the 25th international conference on Machine learning [Internet]. New York, NY, USA: Association for Computing Machinery; 2008 [cited 2021 Dec 10]. p. 160–7. (ICML '08). Available from: https://doi.org/10.1145/1390156.1390177

58. Hinton G, Deng L, Yu D, Dahl GE, Mohamed A, Jaitly N, et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. IEEE Signal Process Mag. 2012 Nov;29(6):82–97.

59. Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol. 2015 Aug;33(8):831–8.

60. Sam A, Currie K, Oh H, Flores-Mir C, Lagravére-Vich M. Reliability of different three-dimensional cephalometric landmarks in cone-beam computed tomography : A systematic review. Angle Orthod. 2019 Mar;89(2):317–32.

61. Castillo JC, Gianneschi G, Azer D, Manosudprasit A, Haghi A, Bansal N, et al. The relationship between 3D dentofacial photogrammetry measurements and traditional cephalometric measurements. Angle Orthod. 2019 Mar;89(2):275–83.

62. Isidor S, Di Carlo G, Cornelis MA, Isidor F, Cattaneo PM. Three-dimensional evaluation of changes in upper airway volume in growing skeletal Class II patients following mandibular advancement treatment with functional orthopedic appliances. Angle Orthod. 2018 Sep;88(5):552–9.

63. Tanikawa C, Takada K. Test-retest reliability of smile tasks using three-dimensional facial topography. Angle Orthod. 2018 May;88(3):319–28.

64. Feng J, Yu H, Yin Y, Yan Y, Wang Z, Bai D, et al. Esthetic evaluation of facial cheek volume: A study using 3D stereophotogrammetry. Angle Orthod. 2019 Jan;89(1):129–37.

65. Leonardi R, Giordano D, Maiorana F. An evaluation of cellular neural networks for the automatic identification of cephalometric landmarks on digital images. J Biomed Biotechnol. 2009;2009:717102.

66. Shahidi S, Shahidi S, Oshagh M, Gozin F, Salehi P, Danaei SM. Accuracy of computerized automatic identification of cephalometric landmarks by a designed software. Dento Maxillo Facial Radiol. 2013;42(1):20110187.

67. Lee Y-S, Suh H-Y, Lee S-J, Donatelli RE. A more accurate soft-tissue prediction model for Class III 2-jaw surgeries. Am J Orthod Dentofac Orthop Off Publ Am Assoc Orthod Its Const Soc Am Board Orthod. 2014 Dec;146(6):724–33.

68. Suh H-Y, Lee H-J, Lee Y-S, Eo S-H, Donatelli RE, Lee S-J. Predicting soft tissue changes after orthognathic surgery: The sparse partial least squares method. Angle Orthod. 2019 Nov;89(6):910–6.

69. Yoon K-S, Lee H-J, Lee S-J, Donatelli RE. Testing a better method of predicting postsurgery soft tissue response in Class II patients: A prospective study and validity assessment. Angle Orthod. 2015 Jul;85(4):597–603.

ANNEXURES

DESCRIPTION OF TERMS

1. **Artificial Intelligence**: According to John McCarthy in 2004, “It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to biologically observable methods”.

2. **Machine Learning**: Machine learning may be a branch of Artificial Intelligence (AI) and computing that focuses on utilizing knowledge and algorithms to imitate how humans learn, gradually improving accuracy.

3. **Deep Learning**: Deep learning may be a subset of machine learning, a neural network with three or more layers. These neural networks plan to simulate the behaviour of the human brain—albeit far away from matching its ability—allowing it to “learn” from large amounts of knowledge.

4. **Convolutional Neural Network (CNN**): CNN is a type of deep learning model for processing data that has a grid pattern, such as images, which is inspired by the organization of the animal visual cortex and designed to automatically and adaptively learn spatial hierarchies of features, from low- to high-level patterns.

5. **Dataset**: In machine learning, it is, quite simply, a set of knowledge pieces that a computer will treat as one unit for analytic and prediction purposes. So the info collected should be made uniform and understandable for a machine that does not equivalently see data as humans do.

6. **Model**: A machine learning model is a file that has to be trained to acknowledge certain sorts of patterns. The Model is trained over a set of knowledge, providing an algorithm to reason over and learn from those data. Once the model has been trained,

It can be used to reason over data that it has not seen before and make predictions about those data.

7. **Back-propagation**: It is the essence of neural net training. It is the practice of fine-tuning the weights of a neural net based on the error rate (i.e., loss) obtained in the previous epoch (i.e., iteration). Proper tuning of the weights ensures lower errors in results.
8. **Euclidean distance**: Distance between two points in either the plane or 3-dimensional space measures the length of a segment connecting the two points. It is the most obvious way of representing the distance between two points. The Pythagorean Theorem can be used to calculate the distance between two points, as shown in the figure below. If the points (x1,y1)(x1,y1) and (x2,y2)(x2,y2) are in 2-dimensional space, then the Euclidean distance between them is $\sqrt{(x2-x1)^2+(y2-y1)^2}$.

9. **Root Mean Square Error (RMSE)**: Root Mean Squared Error is the square root of Mean Squared error. It measures the standard deviation of residuals. Mean Squared Error represents the average squared difference between the original and predicted values in the data set. It measures the variance of the residuals.

10. **R2 Score:** This is a fundamental metric that is used to evaluate the performance of a regression-based machine learning model. It is pronounced as R squared and is also known as the coefficient of determination. It works by measuring the amount of variance in the predictions explained by the dataset. Simply put, it is the difference between the samples in the dataset and the predictions made by the model.

LIST OF ABBREVIATIONS USED

ABBREVIATIONS	**EXPLANATION**
CNN	Convolutional Neural Network
AI	Artificial Intelligence
ROI	Region of Interest
IEEE	Institute of Electrical and Electronics Engineers
ISBI	International Symposium on Biomedical Imaging
NAS	Neural Architectural Search
2D	Two Dimensional
FIG	Figure
GUI	Graphical User Interface
RMSE	Root Mean Square Error
R2 Score	Coefficient of Determination

ACKNOWLEDGMENT

This is high time for me to take this opportunity to express gratitude and thanks from the core of my heart to those without whose cooperation it would not at all be possible to complete this task. At first, I want to express profuse thanks to my research guide Prof Dr Tony Michael and my HOD and Prof Dr Binnoy Kurian. They rendered expert guidance and generous help in the successful turnover of this work and assisted me in different ways by exploiting different facilities on their behalf.

I am highly obliged to my academic mentors Dr Abraham George, Dr Renji K Paul, Dr Mathew Jain, Dr Deby for their valuable discussions and suggestions about my topic.

I am highly thankful to prof. Dr Ajith V V, Prof Dr George Jose for guiding and helping me in the completion of my thesis. They have also extended their friendly supports while doing the manual landmarking process.

I am expressing hearty thanks to my Co-PGs Dr Jose Nelson and Dr Kareena Miriyam Kafeel Reynold, who stretched their warm hands at some critical junctures to accelerate the progress of this project.

My extreme gratitude finds no horizon to express thankful acknowledgement to my senior and junior colleagues in St.Gregorios Dental College, They continuously encouraged me from different dimensions to complete this research work.

This is the right time for me to remember my respected parents, my beloved brother (Sreeraj S), who are undoubtedly the happiest persons at such a joyful moment of my life.

Date: **Signature of the Candidate**

Place: **Dr Jishnu S**