3D MORPHOMETRIC QUANTIFICATION OF MAXILLAE AND PALATAL DEFECTS FOR PATIENTS WITH UNILATERAL CLEFT LIP AND PALATE VIA AUTO-SEGMENTATION

Matthew Pastewait

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the School of Dentistry (Orthodontics).

Chapel Hill
2020

Approved by:

Ching-Chang Ko

Feng-Chang Lin

Tai-Hsien Wu

ABSTRACT

Matthew Pastewait: 3D Morphometric Quantification of Maxillae and Palatal Defects for Patients with Unilateral Cleft Lip and Palate via Auto-segmentation
(Under the direction of Ching-Chang Ko)

The accurate quantification of the complex 3D cleft defect structure is key for optimal treatment planning and patient outcomes. Furthermore, very little is known about the morphometric differences between the affected versus the unaffected maxillary halves. The aim of this study is to characterize the 3D morphometry of the maxillae and cleft defects in non-syndromic patients with unilateral cleft lip and palate. To test the hypothesis that the defect size is positively correlated with the affected maxillary half, CBCT images were acquired from 60 patients presenting with unilateral cleft lip and palate. The machine learning program LINKS was used to segment the maxilla and defect. The height, width, and length of the maxilla and defect were measured from the segmented images. To fully characterize the defect, the distribution probability was mapped from superimposed 3D models, paired t tests were performed for statistical analysis, and a multiple linear regression was completed. The defect side demonstrated a significant decrease in maxillary length, anterior width, and volume with mean measurements of 34.31±2.56mm, 17.83±2.06mm, and 18.02±3.24x$10^3$$mm^3$, respectively, and an increased maxillary anterior height with a mean of 25.91±4.12mm as compared to the non-defect side. Defect superimposition displayed a concentrated distribution near the alveolar bone region and anterior maxillary structures appeared to contribute to defect variability.

ACKNOWLEDGEMENTS

Thank you to my committee members, Dr. Ko, Dr. Lin, and Dr. Wu, for your expertise, guidance, and advice throughout my project. Thank you to Dr. Xiaoyu Wang, Beatriz Tejera, Dr. Li Wang, Dr. Dinggang Shen, and Dr. Song Li for your research assistance. Thank you to Dr. Feng-Chang Lin for the statistical analysis of the study data. Thank you to the National Institutes of Health for the generous research grant.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
ABG	Alveolar Bone Graft
AI	Artificial Intelligence
ANN	Artificial Neural Network
AS	Automatic Segmentation
BoW	Bag-of-Words
CAD/CAM	Computer-aided Design/Computer-aided Manufacturing
CBCT	Cone Beam Computed Tomography
CG	Control Group
CMF	Craniomaxillofacial
CT	Computed Tomography
DL	Deep Learning
DSC	Dice Similarity Coefficient
GPU	Graphics Processing Unit
HAC	Hierarchical Agglomerative Clustering
ICC	Intra-class Correlation Coefficient
LINKS	Learning-based multi-source IntegratioN frameworK for Segmentation
ML	Machine Learning
MR	Magnetic Resonance
MS	Manual Segmentation

NLP	Natural Language Processing
OoK	One-of-K
PPV	Positive Prediction Value
PRM	Probabilistic Roadmap
RF	Random Forest
RL	Reinforcement Learning
RME	Rapid Maxillary Expansion
RNN	Recurrent Neural Network
SD	Standard Deviation
SEN	Sensitivity
SG	Study Group
SVM	Support Vector Machine
UCLP	Unilateral Cleft Lip and Palate
USE	Universal Sentence Encoder

LIST OF SYMBOLS

©	Copyright Symbol

REVIEW OF THE LITERATURE

Artificial Intelligence (AI) is a branch of computer science that has been fruitful in its utilization in many fields ranging from marketing and entertainment to medicine. The general understanding of AI is that machines can work and react like humans. It is a very broad field that covers applications from automatic light timers to a computer program that knows how to play chess. One domain of AI that has been widely discussed recently is machine learning (ML).

ML, a subset of AI, is a data analysis method that allows machines to learn how to accomplish a particular task without being explicitly programmed. Generally, ML can be divided into three major categories: supervised learning, unsupervised learning, and reinforcement learning.[1]

Supervised learning is learning from labeled training data. A supervised learning algorithm analyzes the training data based on the label and proposes an inferred model, which can be used to predict the output of new data. Common supervised learning tasks include classification and regression. Cases such as tumor differentiation, in which the aim is to determine whether a tumor is benign or malignant based on the features of the tumor (e.g., size and shape), are called classification problems. In the field of orthodontics, the decision to extract teeth based on intraoral findings is another classification example.[2] Conversely, an example of a regression problem is the prediction of the yield in a chemical manufacturing process in which the inputs are the concentrations of reactants, temperature, and pressure, and the output is a continuous, non-discrete value that is produced based on fitting the inputs to an inferred model.[1]

Unsupervised learning is finding hidden patterns and structures in unlabeled data. The primary goal of unsupervised learning is clustering, which involves discovering reoccurring patterns and grouping similar examples within the data based on the specific characteristics. This method also determines the distribution of data within the input space, known as density estimation, or projects the data from a high-dimensional space down to two or three dimensions for the purpose of visualization.[1] For instance, a hierarchical agglomerative clustering (HAC) method was used to classify morphological variants of temporomandibular joint osteoarthritis among 169 patients.[3]

Reinforcement learning means learning by interacting with an environment and changing behavior to maximize reward.[4] Reinforcement learning involves a series of steps where each step is associated with some reward. The machine develops an algorithm that maximizes the reward. With the development of advanced machines, more involved learning has become feasible. As a result, the machine can utilize more than one learning method to design an algorithm. AlphaGo is a well-known recent example of ML that incorporates both supervised and reinforcement learning to play Go.[5] In medicine, the Probabilistic Roadmap (PRM) method combined with reinforcement learning (RL) has been applied to a surgery robot to perform minimally invasive laparoscopic surgery.[6]

Deep learning (DL) is another term that has been popularly discussed in relation to ML. DL is a sub-domain of ML, in which the machine itself calculates specific features of a given input. The precursor of DL is the artificial neural network (ANN), which was developed in the 1900s. Due to its need for extreme computing power to calculate numerous artificial neurons/weights inside the networks, the network structure could only be simple, and the practical applications were limited. However, graphics processing unit (GPU) computing, a very

powerful parallel computing technology, has undergone rapid developments since 2010, and it has become possible to fulfill the need of computing power for neural network algorithms. As a result, researchers have been able to design more complicated and "deeper" neural networks to solve more complicated practical problems, and the neural network has become known as "deep learning."

Although DL is a powerful and accurate tool in many applications, the criteria for such precise output is extensive datasets or "big data," which is difficult to achieve in the field of medicine. Therefore, some conventional ML methods, such as random forest (RF) or support vector machine (SVM), still play an important role in practical problems. Specifically, according to the "No Free Lunch" theorem, there does not exist an ML algorithm that is best for all problems.[7] The selection of algorithm depends on numerous properties of data such as size and structure.

Despite ML and DL's popularity, there are not many related studies in the field of orthodontics. Nevertheless, those that do exist demonstrate great potential for AI in this area. For example, the research by Lee et al. demonstrates the use of a neural network to automate landmark identification on lateral cephalograms.[8] In the following sections, we review additional ML orthodontic articles, covering two-dimensional (2D) and three-dimensional (3D) image processing, from object detection to volumetric segmentation, as well as natural language processing. In the first case, AI is used to identify facial traits to aid in orthodontic diagnosis.[9] The second case involves the use of the random forest classifier for volumetric analysis of maxillae with and without impacted canines.[10] In the third case, the PointNet deep learning algorithm is used to automatically segment teeth on maxillary models. The fourth case involves

the use of natural language processing to analyze a patient's clinical findings and develop a prioritized problem list and related treatment plan.[11]

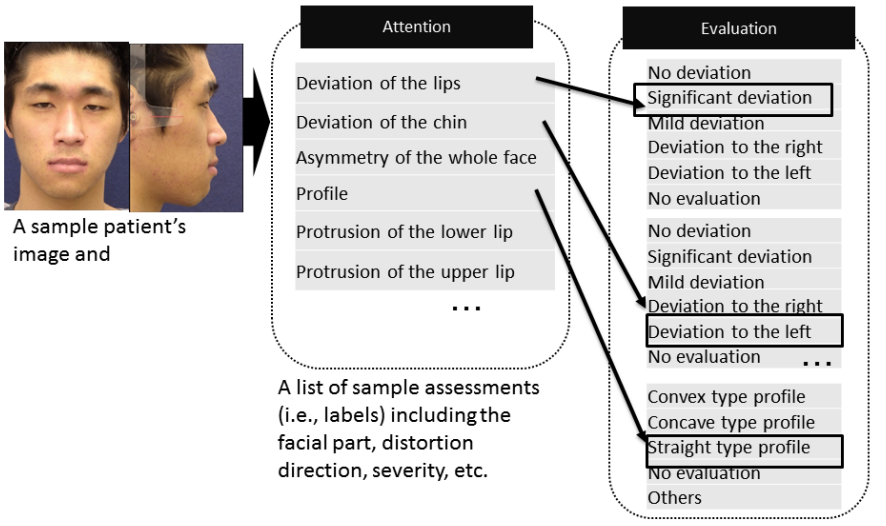
Case I: An AI System for the Orthodontic Examination of Facial Images

The face plays an important role as a means of nonverbal communication in the transmission of emotions and thoughts during our social lives. Thus, facial topography influences the social acceptability and self-image of individuals. With this in mind, orthodontic diagnosis and treatment planning has been changed from the "Angle" paradigm, which is based on hard tissue relationships, to the "soft-tissue" paradigm, which focuses on the facial outcome. Therefore, when making treatment plans in orthodontics, observing patients' faces in the clinical setting is important. Recently, AI has emerged in the examination of faces using deep learning algorithms. For example, the Face2Gene software[12] combines computational facial recognition with a clinical knowledge database in order to prioritize possible genetic syndromes for a given patient. More recently, systems that automatically provide clinical descriptions of oral or facial images for orthodontic diagnostic purposes have been reported.[9,13]

In general, there are two traditional models for the deep learning of images, the convolutional neural network (CNN) and the recurrent neural network (RNN). The CNN is a traditional neural network model that is generally composed of convolutional layers (where filters extract the target features), pooling layers (where the spatial sizes of features and the amount of model parameters are reduced), and fully-connected layers (linear combinations of the features of the previous layer, which make up the next layer). While CNN is a feed forward neural network that is generally used for image recognition and classification, RNN works on the principle of saving the output of a layer and feeding this back to the input in order to predict the output of the layer. RNN is useful when a sequence of data is being processed to make a

classification decision, such as with time series data. Orthodontic facial diagnosis is also a time series examination as orthodontists diagnose a patient comprehensively by looking at the entire face while assessing multiple aspects of the face from different angles, rather than by simply targeting one part of the face. For instance, an orthodontist must first look at the frontal face of a patient and examine the patient for asymmetry, including inclination of the eyelids and/or distortion of the nose. The orthodontist must also confirm maxillary protrusion and/or prognathism from the side of the face, check the tooth alignment while smiling, and finally give the patient a facial diagnosis. This complex assessment process generates variation in the diagnosis among different orthodontists. To mimic an orthodontist's comprehensive analysis using an AI system, Murata et al. [9] developed a hybrid CNN/RNN model with an attention mechanism. This model can provide an objective facial morphological assessment by identifying clinically relevant facial traits (e.g., concave profile, upper lip retrusion, presence of scars). The automation considerably reduces the assessment workload for dentists and also reduces variation in diagnosis.

In Murata et al.'s research, lateral and frontal facial images of 352 patients who visited the Department of Orthodontics at Osaka University Dental Hospital were employed as the training and evaluation data. An experienced orthodontist examined all of the facial images and identified as many relevant facial traits as possible (e.g., deviation of the lips, deviation of the mouth, asymmetry of the face, concave profile, upper lip retrusion, presence of scars). A sample patient image, a list of sample assessments (i.e., labels), and the multi-label data are shown in Figure 1.



	Deviation of lips				Deviation of the nose				Profile				...
	No deviation	Significant	Mild	To right ...	No deviation	Significant	Mild	To right ...	Convex	Concave	Straight	Others	...
Face A	1	0	0	0	0	1	0	0	0	0	1	0	...
Face B	0	0	0	0	0	0	0	0	1	0	0	0	...
Face C	1	0	0	1	1	0	0	0	1	0	0	1	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
Face X	1	0	0	0	0	0	0	0	0	0	0	0	...

Figure 1. Sample patient's image and a list of sample assessments (i.e., labels) including the region of interest, evaluation, etc. In a previous study by Murata et al., they employed labels representing only the facial part (mouth, chin, and whole face), distorted direction (right and left), and its severity (severe, mild, no deviation).

In general, medical images may contain multiple regions of interest to be evaluated. An orthodontist diagnoses a patient based on the assessment results of various facial regions from several different facial images. Thus, for automated diagnostic imaging, a typical single-label (binary or multi-class) image classification model was extended to solve the problem of multilabel image classification. Murata et al. proposed a hybrid model using CNN and RNN that sequentially focuses on multiple regions without pre-processing, such as cropping. For each region, the model produces a label (i.e., an assessment) consisting of words, each of which is predicted based on the previously predicted words (at earlier layers in the RNN). The model is trained on patients' facial images to predict a set of assessments of facial attention parts, such as the eyes, nose, lips, chin, and profile. In practice, the assessment of different facial attention parts has mutual dependencies. As a result, Murata et al. designed a model that learns these dependencies. The attention mechanism in the RNN tells the network which sub-area of the

image impacts the prediction of each particular label. This mechanism helps to reduce the computational cost by selecting and learning the most relevant parts of the image for the predicted labels.

Figure 2 shows some patients' facial images with the visual attention. The white regions in the images represent the attention. Patient A has heavy distortion around his jaw while Patient B has only mild asymmetry. The model was able to predict both of the samples correctly. Patient C has only mild distortions around his mouth and jaw, but the model predicted the wrong labels. Table 1 shows the classification accuracy (%) for each facial part, averaged in a 10-fold cross validation. Although the accuracy itself is still low, the proposed model has better mechanisms to learn visual attention. In addition, label dependency contributes to an increase in accuracy.

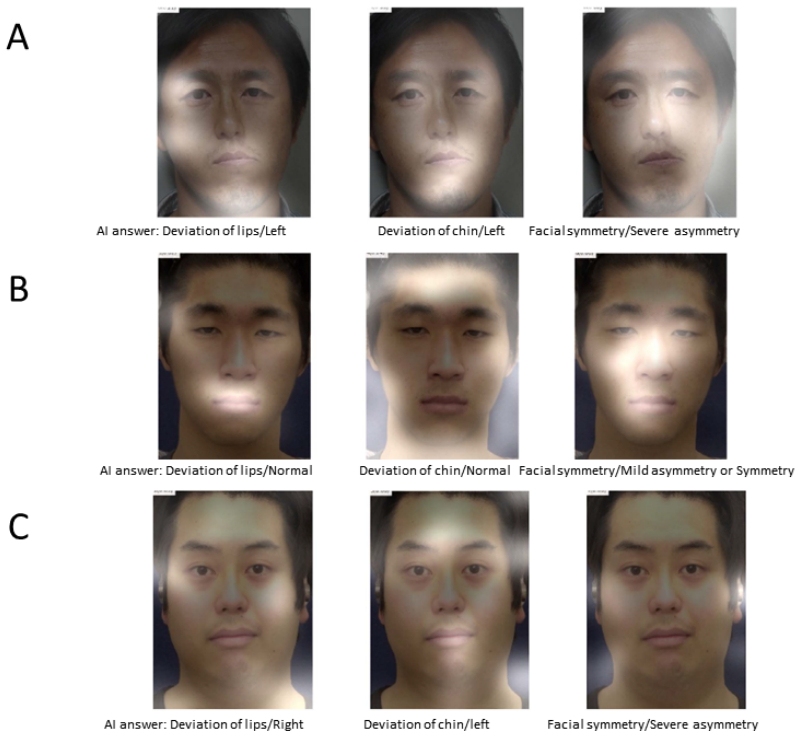


Figure 2. Resultant figures of the AI system. A. Images with visual attention of Patient A who has severe asymmetry around his mouth and chin. The model correctly predicts the labels; B. Images with visual attention of Patient B who does not have any severe problems. The model correctly predicts the labels; C. Images with visual attention of Patient C who has mild distortions around his mouth and chin. The model predicts incorrect labels.

Table 1. The average classification accuracy (%) of the results of a 10-fold cross validation.[9] The numbers in parentheses indicate the standard deviation.

	Traditional model (Multiple CNNs)	Their Model (a deep CNN/RNN model with an attention mechanism)
Lips	64.0 (±7.5)	65.7 (±7.4)
Chin	57.9 (±16.5)	61.3 (±12.7)
Whole face	67.1 (±9.7)	67.4 (±9.1)
Average	63.0 (±9.6)	64.8 (±7.7)
Worst, Best	40.0, 74.3	49.5, 74.3

Case II: Automated Volumetric Segmentation of Dental CBCT Images Using Prior-guided Sequential Random Forests

Segmentation of cone-beam computed tomography (CBCT) images has been widely used in orthodontics, which is a crucial step for generating 3D models for the advanced diagnosis and treatment planning of patients. Since manual segmentation is tedious, several automatic volumetric segmentation methods of CT/CBCT have been presented over the last decade, based on different ML algorithms, like random forest (RF)[14], support vector machine (SVM)[15], and CNN[16].

Chen et al.[10] used a random forest algorithm to assess maxillary constriction in unilaterally impacted canine patients. Their results demonstrated that the maxilla of patients with unilaterally impacted canines had significantly smaller volumes than those without impacted teeth. In their work, 60 CBCT images were successfully auto-segmented. Typically, this number of samples would be difficult to segment manually, suggesting an advantage of machine learning in orthodontics.

Chen et al. used the random forest-based algorithm known as Learning-based multi-source IntegratioN frameworK for Segmentation (LINKS), which was first presented by Wang et al. for the segmentation of infant brain magnetic resonance (MR) images,[17] and then applied to the segmentation of patients with craniomaxillofacial (CMF) deformities[14]. Using this method, the volumetric segmentation is formulated as a classification problem, and the random forest serves as the classifier algorithm. The detailed implementation is referenced in the work by Wang.[14] In general, there are three stages in a machine learning application: a preparation stage, a training and testing stage, and an application stage. In the preparation stage, 36 CBCT images were manually segmented using the software ITK-SNAP to isolate the maxilla and mandible. The training stage was divided into the four steps listed below.

Step 1. Estimation of initial probability maps with majority voting

All the expert-segmented CBCT scans were used as training samples and further aligned onto every subject image by affine registration. Then, a majority voting method was employed to count the votes for each label at every voxel for estimating the initial probability maps of all labels. The initial probability maps provided rough localizations of every label, serving as spatial priors that were important for guiding the segmentation.

Step 2. Extraction of CBCT appearance and context features

The 3D Haar-like features[18] were used in this method. Superficially, for each voxel x in the original CBCT image or probability maps, its Haar-like features f were computed by:

$$f(x, I) = \frac{1}{|R_1|}\sum_{u \in R_1} I(u) - b\frac{1}{|R_2|}\sum_{u \in R_2} I(v),$$

$$R_1 \in R, R_2 \in R, b \in \{0, 1\},$$

where I is the intensity of an original CBCT image or a probability map; R is the patch centered at voxel x; R_1 and R_2 are randomly displaced cubical regions in the patch R, as shown in Figure 3; and the parameter b is either 0 or 1, indicating whether one or two cubical regions are used. The features extracted from the original CBCT images and probability maps were called appearance features and context features, respectively. They were used to coordinate the segmentations in different parts of the CBCT image. These context features have been shown to be effective in both computer vision and medical image analysis fields.[19] It is important to note that the extraction of context features is recursively conducted on the iteratively updated probability maps, whereas the extraction of appearance features is performed only on the original CBCT images.

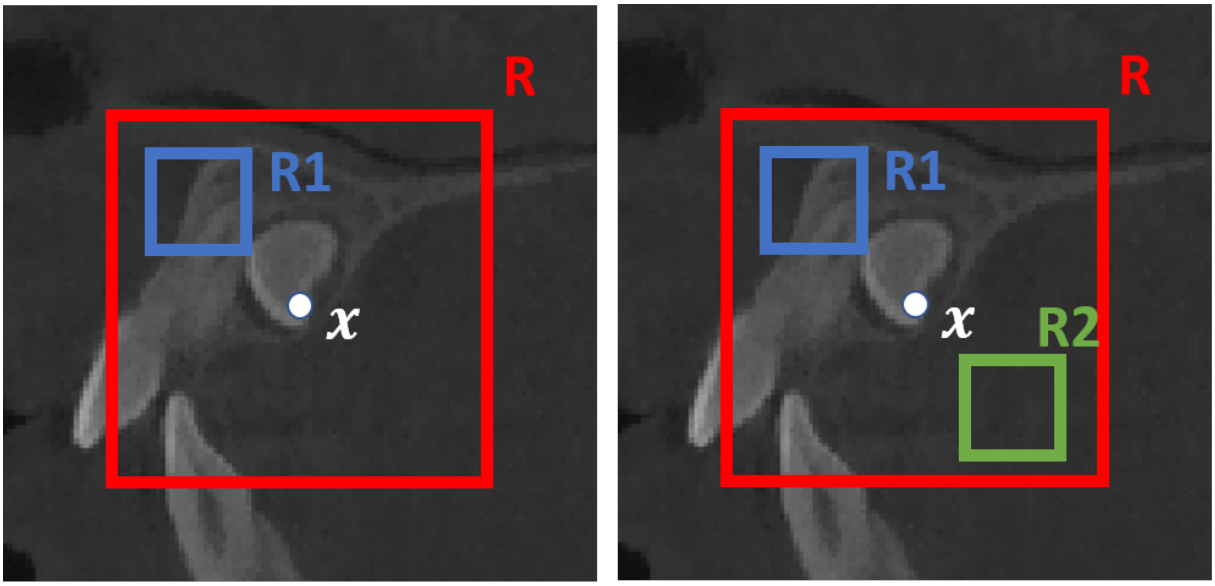


Figure 3. 3D Haar-like features.

Step 3. Training of random-forest-based classifiers

To refine the segmentations (probability maps), a random forest classifier was trained to learn the complex relationship between local appearance/context features and the corresponding manual segmentation labels on all voxels of the training atlases.

Step 4. Repeating Steps 2 and 3 until convergence

In this final step, the classifiers were trained in a sequential manner. Specifically, the segmentation probability maps were updated based on the classifier trained in Step 3. Then, according to Step 2, the context features were extracted from the updated segmentation probability maps and further used with the original CBCT appearance features to train the next classifier. Eventually, a sequence of classifiers was obtained for CBCT segmentation. The flowchart of the entire training stage is shown in Figure 4.

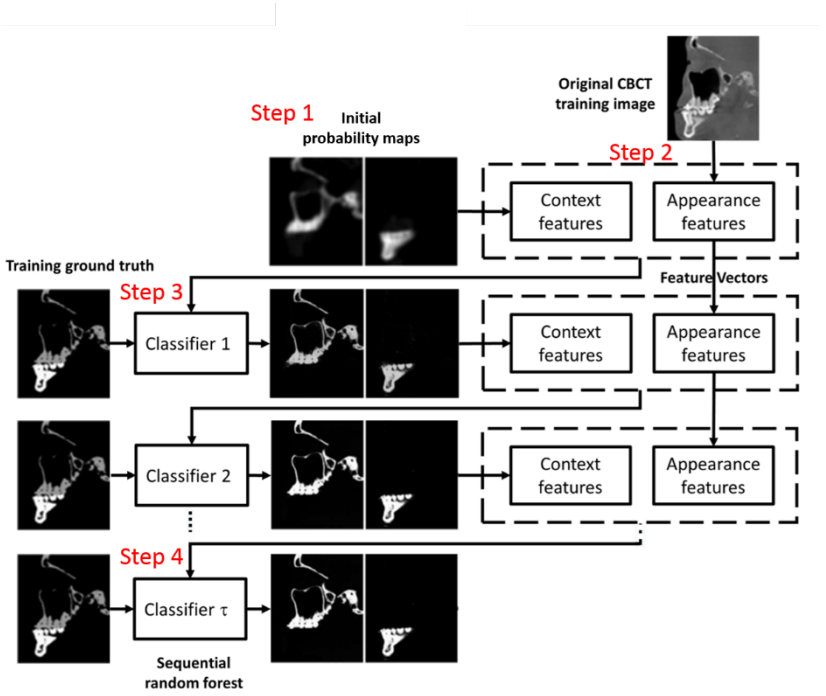


Figure 4. Random Forest classifier flowchart.

In the testing stage, given a new CBCT image, the initial segmentation probability maps of every label were first estimated using majority voting. Then, based on the estimated

probability maps, the context features were extracted and, together with CBCT appearance features, served as the input to the sequential classifiers for iteratively updating the segmentation probability maps. The refined probability maps, the output of the sequential classifiers, were the final segmentation results. The accuracy of the segmentation classifiers was evaluated using the Dice Similarity Coefficient (DSC). The average DSC of the maxilla was 0.800±0.029, ranging from 0.742 to 0.830, based on 30 training and 6 test samples.

In the application stage, Chen et al. used this random forest-based method to segment the maxillae of 60 CBCT images, with 30 scans of unilaterally impacted canines, i.e., the Study Group (SG) and 30 scans without impacted teeth, i.e., the Control Group (CG). Their clinical outcomes are listed in Table 2.

Table 2. The clinical outcomes of Study Group and Control Group in Ref.[10].

	Study Group	Control Group
Bone volume Impacted side for study group Left side for control group Mean ± SD[1] (10^4 mm^3)	2.36±0.35	2.57±0.30
Bone volume Non-impacted side for study group Right side for control group Mean ± SD[1] (10^4 mm^3)	2.37±0.34	2.65±0.38
Maxillary width (mm)	64.3±5.3	66.6±3.6
Maxillary height (mm)	65.1±3.6	67.0±3.5
Maxillary depth (mm)	47.7±3.6	49.6±3.3

Generally, Chen et al. found that the difference between SG ($4.73\pm0.67\times10^4$ mm^3) and CG ($5.22\pm0.65\times10^4$ mm^3) was significant in volume, even after adjusting for gender and age,

[1] SD=Standard Deviation

showing that the SG tended to have a smaller maxillary volume (roughly 5000 mm^3 less) than the CG. In addition, in the SG, the average volumes of the non-impacted ($(2.36\pm0.35)\times10^4\ mm^3$) and impacted sides ($(2.37\pm0.34)\times10^4\ mm^3$) were not significantly different. Their data might explain O'Neill's finding, in which the use of rapid maxillary expansion (RME) in the early mixed dentition effectively increased the rate of eruption of palatally displaced maxillary canines compared to an untreated control group.[20]

In conclusion, the work done by Chen et al. is an excellent example to illustrate the advantages of machine learning in orthodontics. The amount of data is almost impossible to collect manually, particularly the tedious work of volumetric segmentation.

Case III: Automatic Segmentation of Dental Surface Images Using POINTNET

Computer-aided design/computer-aided manufacturing (CAD/CAM) technology has been widely used in orthodontics. For example, 3D dental surface images can be easily obtained through an intraoral scanner in orthodontic practices. As a fundamental part of CAD/CAM-based treatment plans, labeling teeth accurately on 3D dental surfaces is a crucial step for advanced diagnosis. However, automation of this labeling/segmentation task is challenging on the raw dental surface image acquired from the intraoral scanners for multiple reasons, including large non-tooth and irregular parts (e.g., gingival tissues) in the image.

Since volumetric CNNs have demonstrated their outstanding image processing abilities, many researchers transform the surface images to regular 3D voxel grids and feed them into a CNN-based deep learning network. However, this method is not efficient in terms of computing and storage and also results in blurring natural images. For this reason, Qi et al. reported a deep learning architecture, called PointNet, on a point/mesh set for 3D segmentation.[21]

In this section, we demonstrate a tooth-segmentation task using PointNet and discuss its performance. The PointNet architecture shown in Figure 5 is slightly different than the original one, which only takes 3D position (i.e., x, y, and z) of vertices as input features.[21] Instead, we consider N triangles in the surface mesh as input and use a total of 15 features for each triangle. The 15 features are 3D positions of three vertices (9 features) and the normal vector (3 features) of each triangle as well as the relative position (3 features) of each triangle with respect to the whole surface. The raw dataset here consists of 20 maxillary dental surfaces from a 3D intraoral scanner (iTero Element). All surfaces were down-sampled to 10,000 triangles while preserving the original topologies and augmented by random rotation, translation, and rescaling of each surface image in reasonable ranges. The number of categories (M) is 15, including 14 teeth between the left and right second molars as well as the remaining part in the surface image (i.e., background).

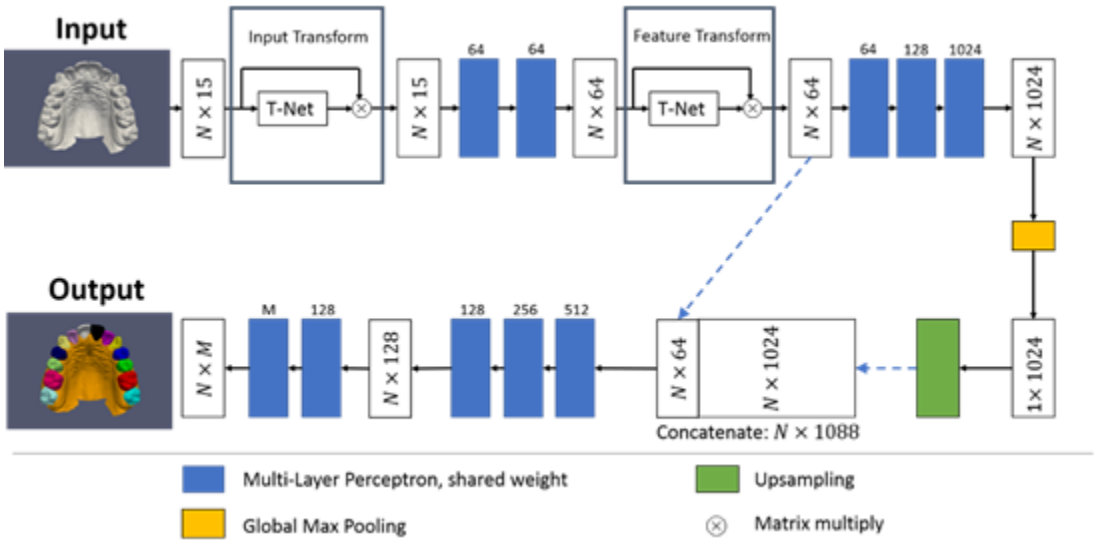


Figure 5. PointNet architecture.

The implementation was carried out using Keras with Tensorflow as the backend. Three-fold cross-validation was performed for this demonstration. For each training/validation sample, 6,000 triangles were randomly selected from each surface as the network input data with a 50:50

ratio of triangles from each tooth and gingiva. Specifically, the input array ($N \times 15$) and output array ($N \times M$) were both $6{,}000 \times 15$. The Adam optimizer was used with a mini-batch size of 10; the number of epochs was 100 for the training; and the loss function was generalized Dice loss.[22] The segmentation results were evaluated by three metrics: DSC, sensitivity (SEN), and positive prediction value (PPV).

The segmentation results of DSC, SEN, and PPV (mean$\pm$standard deviation) for all teeth were 0.781$\pm$0.134, 0.828$\pm$0.167, and 0.766$\pm$0.163, respectively, as shown in Table 3. A screenshot of the results is shown in Figure 6. Although the accuracy still needs to be improved for clinical purposes, the results of this demonstration reveal the promise of automatically labeling teeth in the surface image, creating a new path for digesting surface mesh directly instead of converting to 3D voxel grids. A better result could be expected using a more sophisticated deep learning network and a larger dataset in the future.

Table 3. Results of the labelled teeth using PointNet.

Metric	Mean$\pm$SD	Min	Max
DSC	0.781$\pm$0.134	0.695	0.882
SEN	0.828$\pm$0.167	0.702	0.960
PPV	0.766$\pm$0.163	0.575	0.855

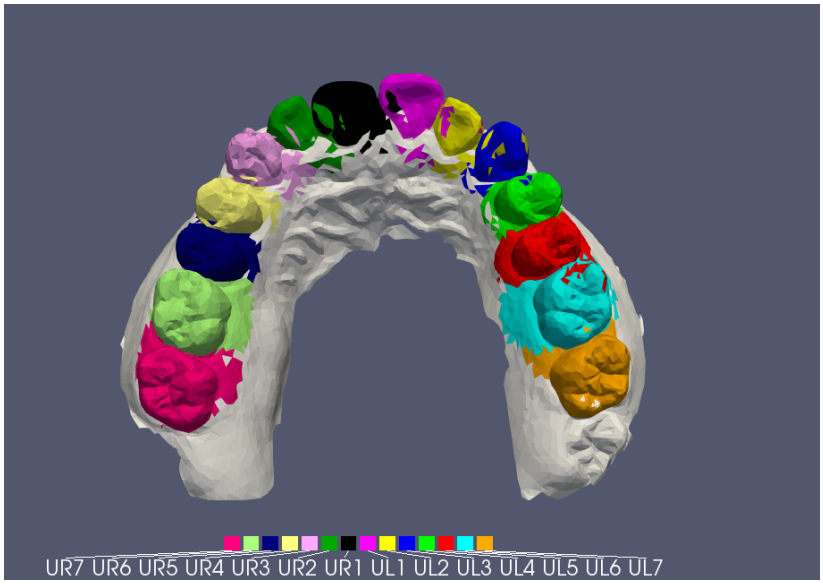


Figure 6. Screenshot of the labelled teeth using PointNet.

To conclude, we have demonstrated a promising path to automatically segment each tooth on a surface image. This task is a critical step for advanced diagnosis and customized treatment planning. It also serves as the first step in a series of orthodontic AI systems to obtain further patient data and information.

CASE IV: USING NATURAL LANGUAGE PROCESSING TO DEVELOP AN AUTOMATED ORTHODONTIC DIAGNOSTIC SYSTEM

An orthodontic diagnosis and treatment plan involves predicting the course of action that a dentist should take to obtain the optimal treatment results at the lowest possible risk.[23] Making such an assessment requires years of knowledge and experience. As a result, inexperienced dentists can easily make judgment errors or misunderstand the parameters of a case. An AI system that can automatically summarize orthodontic diagnoses or present necessary examinations would be very helpful in helping these dentists avoid such oversights as well as reduce the heavy workload of practitioners.

In general, the process of modern orthodontic diagnosis and treatment planning consists of the following three steps: (1) collect and itemize patient information regarding the problem(s), (2) contemplate solutions for each problem, and (3) determine the course of action and its implementation. Since these steps following a logical structure, attempts have been made to automate orthodontic diagnosis and treatment planning. An orthodontic diagnosis support system using fuzzy logic[24], a system for the selection of orthodontic appliances[25], and a mathematical model that simulates whether or not to extract teeth in optimizing orthodontic treatment outcomes[26] have previously been developed. However, total orthodontic diagnosis and treatment planning support systems for use in clinics have yet to be established.

To put the orthodontic diagnostic process in mathematical terms, if patient information regarding the problem is thought of as a set of feature values (or measurable characteristics), then Step (1) from above is comparable to representing medical conditions based on the individual weight of each feature value. Step (2) is analogous to learning how to treat each medical condition. At this point, a natural language processing (NLP) AI system could be used to find a solution (Kajiwara et al.[11]). The aim of the study by Kajiwara et al.[11] was to develop an AI system that uses NLP on various clinical text evaluations and their accompanying treatment protocols in order to create an automated process of diagnosis and treatment planning.

In their study, Kajiwara et al. worked on the task of automatically designing a treatment plan from the findings included in the medical certificate written by the dentist. They developed an AI system where the input was a document that described the patient's medical findings and the output was a problem list in the order of treatment priority. The researchers employed a natural language processing approach to extract features from the text because medical findings are typically written as free-form descriptions in natural language.

Kajiwara et al. developed two subtasks to efficiently address the task of generating a treatment plan from clinical findings: (1) list the problems and (2) prioritize treatment. First, subtask (1) summarized the findings and listed the orthodontic problems of each patient. Then, subtask (2) ranked each problem listed in subtask (1) in terms of treatment priority. After accomplishing the subtasks, the system automatically created a treatment plan from the findings contained in the medical certificate.

Kajiwara et al. examined 990 dental certificates. According to the dataset, each patient had an average of 15.4 orthodontic problems. These 990 documents were randomly divided into 810 for training, 90 for validation, and 90 for evaluation.

The researchers developed an NLP system that solved the two aforementioned subtasks using the given dataset. Subtask (1) was regarded as a text generation task that generated a summary of medical findings. Compared to the millions of datasets commonly used in text generation tasks, such as machine translation and automatic summarization, the researchers' dataset of 990 documents was significantly smaller. Therefore, Kajiwara et al. added an annotation (shown in Figure 7(c)) and tackled subtask (1) as a multi-label classification problem. In the annotation, one dentist organized orthodontic problems into 423 classes. Each patient problem corresponded to one class. Subtask (2) ranked each problem listed in subtask (1) in terms of treatment priority. After solving the two subtasks, the AI system generated a treatment plan from the findings in the medical certificate.

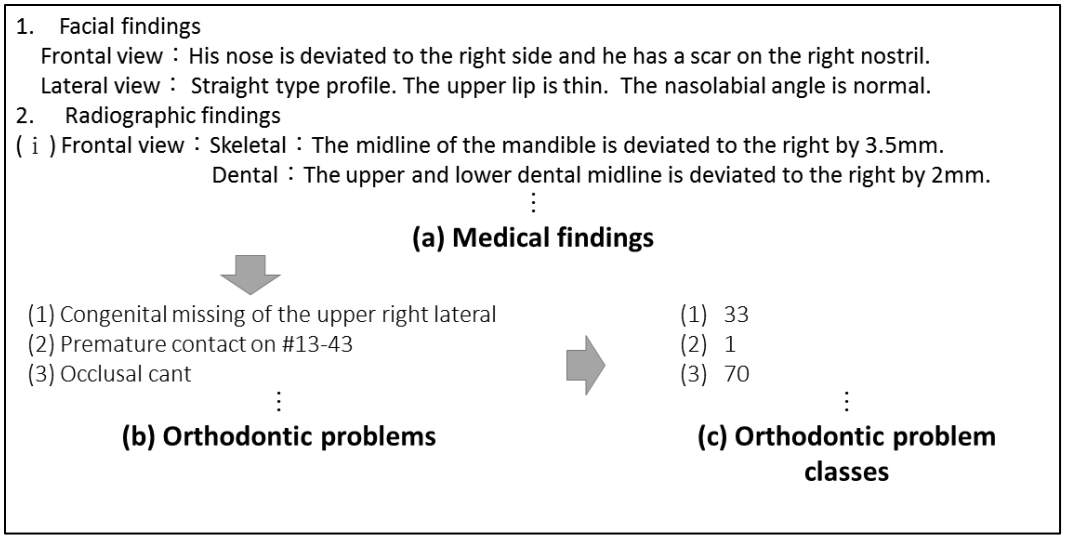


Figure 7. Annotation of orthodontic problem classes to the medical certificate.

The researchers developed classification models of orthodontic problems on the datasets shown in Figure 7(a) and (c). These models recorded the text of the findings as inputs and generated a list of class labels as outputs (shown in Figure 7(c)). Kajiwara et al. converted the inputs into vector representation using a natural language processing approach and performed supervised learning of multi-label classification with them as features. The following methods were used to vectorize each document.

- BoW (Bag-of-Words): The BoW representations have dimensions corresponding to the number of vocabulary words in the training dataset, and each dimension has a value of 1 if the corresponding word appears in the input text or 0 if it does not. These are high-dimensional sparse vectors.
- USE (Universal Sentence Encoder): The researchers construct feature vectors from the text using the cross-lingual version of the universal sentence encoder.[27] These are 512-dimensional dense vectors.

For the BoW model, each sentence was divided into words using the text segmentation library MeCab.[28] The vocabulary size was 2,075 because only words appearing five or more times in the training dataset were used. Semantically equivalent classes were grouped, and 151 class labels were used. In machine learning, the perceptron is an algorithm for the supervised

learning of binary classifiers. For this research, a multi-layer perceptron was implemented on the deep learning framework Chainer[29] for a multi-label classifier. In the output layer, a sigmoid function was used instead of the softmax function in the single-label classification.

The performance of each model was automatically evaluated using the F-measure, which is the harmonic average of the precision and the recall. The experimental results showed that the simple BoW model achieved higher performance (0.59) than the USE model (0.54). Unique structures, such as bullets and incomplete sentences, may have had a negative impact on the sentence encoder. BoW models, on the other hand, treated documents as a set of words, so they were not affected by these unique sentence structures.

Kajiwara et al. developed a prioritization model of treatment on the datasets shown in Figure 7(b) and (c). This model recorded lists of text representing the orthodontic problems or classes as inputs and generated lists of treatment priority for each problem as outputs.

Again, the researchers converted each problem into vector representation using a natural language processing approach and performed learning-to-rank with them as features. The following three methods were used to vectorize each problem:

- BoW (Bag-of-Words): They constructed feature vectors from the text of the problem. These vector representations had dimensions corresponding to the number of vocabulary words in the training dataset, and each dimension had a value of 1 if the corresponding word appeared in the input text, or 0 if it did not.

- OoK (One-of-K): They constructed feature vectors from the labels that represented the problem. These vector representations had dimensions corresponding to numbers of labels appearing in the training dataset, and only one of the dimensions corresponding to the input class had a value of 1 and the other dimensions had a value of 0.

- USE (Universal Sentence Encoder): They constructed feature vectors from the text of the problem using the cross-lingual version of universal sentence encoder.[27] These were 512-dimensional dense vectors.

In the BoW model, each sentence was divided into words using MeCab.[28] Since the training dataset had 146 vocabulary words and 320 classes, the feature vectors of each model were 146 dimensions for BoW, 320 dimensions for OoK, and 512 dimensions for USE.

The researchers used the machine learning technique known as learning-to-rank in order to rank the orthodontic problems or classes. Specifically, they employed SVM-rank[30] with a linear kernel, a standard toolkit. For a given set of features, they examined a hyper-parameter among $C \in \{1, 5, 10, 50, 100, 500, 1000, 5000\}$ on the validation dataset. Spearman's rank correlation coefficient was used to evaluate the performance of each model. When the correlation coefficient between the human ranking and the estimated ranking exceeded 0.4, the estimation result of the model was interpreted as having a positive correlation with human evaluation.

The experimental results are shown in Figure 8. Each method has a Spearman's rank correlation coefficient exceeding 0.4. Therefore, it can be interpreted that these estimation results have a positive correlation with human evaluation. Unlike the BoW model, the OoK model can obtain feature vectors that reflect an annotator's expertise, so it is considered that high performance has been achieved. The USE model achieved the highest performance because the dense vectors obtained by deep learning can represent rich information.

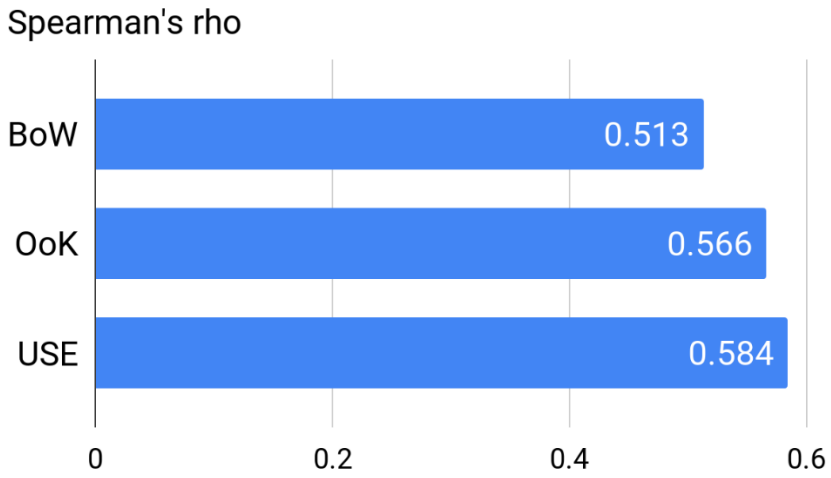


Figure 8. Automatic evaluation of treatment ranking models by Spearman's rho.

The above algorithm can also be applied to the automated planning of treatment protocols (Figure 9). The system can train a series of transformation models using pairs of relevant treatment protocol summaries and consent form documents. First, relevant sentence pairs would be automatically extracted from the relevant documents using an NLP method of sentence alignment. Next, statistical machine translation or neural machine translation techniques would be used to automatically translate experts' language into simpler language that can be understood by patients.

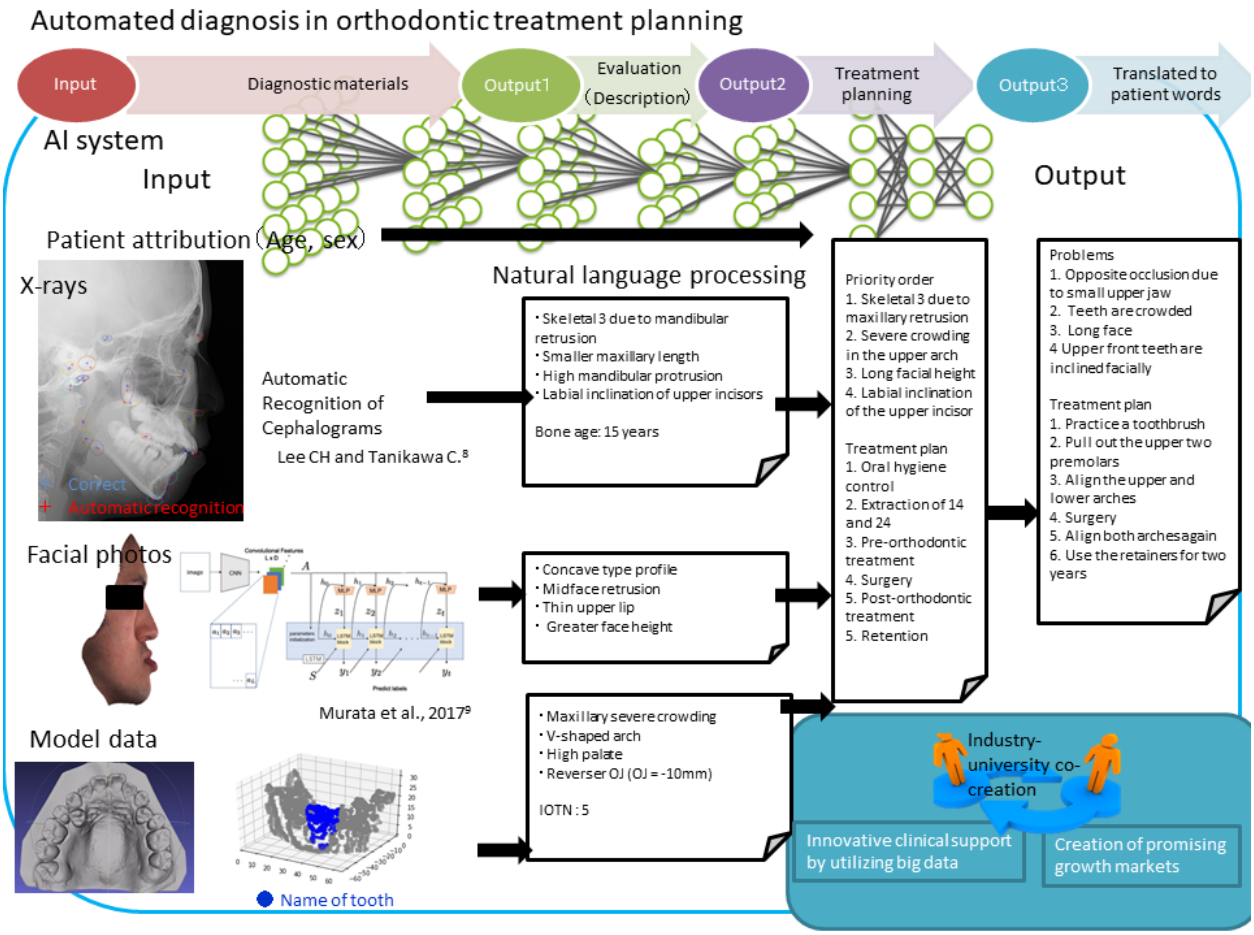


Figure 9. Overview of the fully automated orthodontic diagnosis system.

CONCLUSIONS

Artificial intelligence, including ML and DL, is rapidly expanding into multiple facets of society. Orthodontics may very well be one of the fastest branches of dentistry to adapt AI for three reasons. First, patient encounters during treatment generate many types of data. Cephalometric landmarks, digital photographs, intraoral and extraoral features are just a few types of data generated in the dental clinic. AI can perform analytics to decipher this information and aid in efficient diagnosis and treatment planning. Second, the standardization in the field of dentistry is low compared to other areas of healthcare. A range of valid treatment options exist for any given case. Using AI and large datasets (that include diagnostic results, treatments, and outcomes), one can now empirically measure the effectiveness of different treatment modalities given very specific clinical findings and conditions. Third, orthodontics is largely practiced by independent dentists in their own clinics. These dentists have the autonomy to adopt beneficial technologies without the bureaucracy often found in large healthcare organizations. In order to remain competitive in the modern dental market, orthodontists must be proactive in seeking innovation and adopting various technologies. Despite the promise of AI, the volume of orthodontic research in this field is relatively low. Further, the clinical accuracy of AI must be improved with an increased number and variety of cases. Before AI can take on a more important role in making diagnostic recommendations, the volume and quality of research data will need to increase.

REFERENCES

1. Bishop CM. *Pattern Recognition and Machine Learning*. springer; 2006.

2. Jung S-K, Kim T-W. New approach for the diagnosis of extractions with neural network machine learning. *Am J Orthod Dentofac Orthop*. 2016;149(1):127-133. doi:https://doi.org/10.1016/j.ajodo.2015.07.030

3. Gomes LR, Gomes M, Jung B, Paniaqua B, Ruellas AC, Goncalves JR, Styner MA, Wolford L, Cevidanes L. Diagnostic index of 3D osteoarthritic changes in TMJ condylar morphology. In: *Proc.SPIE*. 2015;9414. https://doi.org/10.1117/12.2082226.

4. Sutton RS, Barto AG. *Reinforcement Learning: An Introduction*. MIT press; 2018.

5. Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. *Nature*. 2016;529:484. https://doi.org/10.1038/nature16961.

6. Baek D, Hwang M, Kim H, Kwon D. Path Planning for Automation of Surgery Robot based on Probabilistic Roadmap and Reinforcement Learning. In: *2018 15th International Conference on Ubiquitous Robots (UR)*. 2018:342-347. doi:10.1109/URAI.2018.8441801

7. Wolpert DH, Macready WG. No free lunch theorems for optimization. *IEEE Trans Evol Comput*. 1997;1(1):67-82. doi:10.1109/4235.585893

8. Lee Chonho, Tanikawa C. Deep Neural Network based Cephalometric Landmark Identification using Multi-scale Patches. *arXiv*. 2019;in submiss.

9. Murata S, Lee C, Tanikawa C, Date S. Towards a Fully Automated Diagnostic System for Orthodontic Treatment in Dentistry. In: *2017 IEEE 13th International Conference on E-Science (e-Science)*. 2017:1-8. doi:10.1109/eScience.2017.12

10. Chen S, Wang L, Li G, Wu T-H, Diachina S, Tejera B, Kwon J, Lin F-C, Lee Y-T, Xu T, Shen D; Ko C-C. Machine learning in Orthodontics: Introducing a 3D auto-segmentation and auto-landmark finder of CBCT images to assess maxillary constriction in unilateral impacted canine patients. *Angle Orthod*. 2019;under review.

11. Kajiwara T, Tanikawa C, Shimizu Y, Chenhui C, Yamashiro T, Nagahara H. Using natural language processing to develop an automated orthodontic diagnostic system. *arXiv*. 2019;in submiss.

12. No Title. https://www.face2gene.com/.

13. Murata S, Ishigaki K, Lee C, Tanikawa C, Date S, Yoshikawa T. Towards a Smart Dental Healthcare: An Automated Assessment of Orthodontic Treatment Need. *Healthinfo*. 2017;(c):35-39.

14. Wang L, Gao Y, Shi F, et al. Automated segmentation of dental CBCT image with prior-guided sequential random forests. *Med Phys*. 2016;43(1):336-346. doi:10.1118/1.4938267

15. Pei Y, Ai X, Zha H, Xu T, Ma G. 3D exemplar-based random walks for tooth segmentation from cone-beam computed tomography images. *Med Phys*. 2016;43(9):5040-5050. doi:10.1118/1.4960364

16. Qiu B, Guo J, Kraeima J, Borra RJH, Witjes MJH, van Ooijen PMA. 3D segmentation of mandible from multisectional CT scans by convolutional neural networks. *arXiv Prepr arXiv180906752*. 2018.

17. Wang L, Gao Y, Shi F, et al. LINKS: Learning-based multi-source Integration frameworK for Segmentation of infant brain images. *Neuroimage*. 2015;108:160-172. doi:https://doi.org/10.1016/j.neuroimage.2014.12.042

18. Viola P, Jones MJ. Robust Real-Time Face Detection. *Int J Comput Vis*. 2004;57(2):137-154. doi:10.1023/B:VISI.0000013087.49260.fb

19. Sutton C, McCallum A, Rohanimanesh K. Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data. *J Mach Learn Res*. 2007;8(Mar):693-723.

20. O'Neill J. Maxillary expansion as an interceptive treatment for impacted canines. *Evid Based Dent*. 2010;11:86. https://doi.org/10.1038/sj.ebd.6400742.

21. Qi CR, Su H, Mo K, Guibas LJ. Pointnet: Deep learning on point sets for 3d classification and segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017:652-660.

22. Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. 2017:240-248.

23. Takada K. *Elements of Orthodontics*. I-Cube Co. Ltd. 2017.

24. Sims-Williams JH, Brown ID, Matthewman A, Stephens CD. A computer-controlled expert system for orthodontic advice. *Br Dent J*. 1987;163(5):161-166.

25. Stephens CD, Mackin N, Sims-Williams JH. The Development and Validation of an Orthodontic Expert System. *Br J Orthod*. 1996;23(1):1-9. doi:10.1179/bjo.23.1.1

26. Takada K, Yagi M, Horiguchi E. Computational Formulation of Orthodontic Tooth-Extraction Decisions. *Angle Orthod*. 2009;79(5):885-891. doi:10.2319/081908-436.1

27. Chidambaram M, Yang Y, Cer D, et al. Learning Cross-Lingual Sentence Representations

via a Multi-task Dual-Encoder Model. *arXiv Prepr arXiv181012836*. 2018.

28. Kudo T, Yamamoto K, Matsumoto Y. Applying conditional random fields to Japanese morphological analysis. In: *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. 2004:230-237.

29. Tokui S, Oono K, Hido S, Clayton J. Chainer: a next-generation open source framework for deep learning. In: *Proceedings of Workshop on Machine Learning Systems (LearningSys) in the Twenty-Ninth Annual Conference on Neural Information Processing Systems (NIPS)*. 2015:5:1-6.

30. Joachims T. Training linear SVMs in linear time. In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2006:217-226.

3D MORPHOMETRIC QUANTIFICATION OF MAXILLAE AND PALATAL DEFECTS FOR PATIENTS WITH UNILATERAL CLEFT LIP AND PALATE VIA IMAGE AUTO-SEGMENTATION

Introduction

Cleft lip and palate is a common congenital maxillofacial hypoplasia that exhibits multifactorial inheritance. The resultant defect, caused by genetic and/or environmental influences, is characterized by the incomplete formation of the lip, alveolar bone, hard palate, and soft palate. The deficient soft tissue and skeletal formation can result in perinatal death. For survivors, it often results in some combination of feeding, deglutition, speaking, hearing, and/or cognitive difficulties.[1] Unilateral cleft lip and palate (UCLP), the focus of this research, can present as an isolated anomaly or as part of a sequence of congenital anomalies associated with different syndromes.[2] Data for 7.5 million births from the international perinatal database of typical oral clefts (IPDTOC) showed that the prevalence of cleft lip and palate was 6.64 per 10,000 births worldwide.[1] The Chinese national birth defects monitoring network reported an even higher prevalence of 7.62 per 10,000 births in the Chinese population, based on 4.9 million newborn records.[3]

The use of cone-beam computed tomography (CBCT) for preoperative volumetric assessment of alveolar bone grafts (ABGs) has been well established, and its accuracy has been demonstrated extensively in the literature. ABGs are considered necessary for the treatment and improvement of life quality for patients affected by cleft defects. With 75% of cleft patients presenting alveolar bone involvement, both esthetics and function are targeted by correctional ABG procedures. Through 3D analysis, surgeons can better understand the dental and bony

conditions around the defect, estimate the amount of bone needed for grafting, and observe the locations and quantities of bony bridges formed postoperatively. A study by de Rezende Barbosa et al.[4] indicated that CBCT volumes proved reliable for the volumetric assessment of alveolar cleft defects using different methods. However, the volumetric measurement could not illustrate the complex 3D structure variation of the defect. Currently, there are conflicting findings regarding preoperative cleft morphology and ABG success. Feng et al.[5], Linderup et al.[6], and Oberoi et al.[7] found no correlation between initial cleft size and the extent of graft filling. Conversely, van der Meij et al.[8] and Long et al.[9] found significant correlations between initial cleft width and residual bone one year postoperatively. In addition, the use of 3D data to calculate grafting volume has been confirmed to be reliable. Albuquerque et al. [10] and Amirlak et al.[11] performed in vitro research on dry skulls and found that data-derived volumes of both cleft and graft bone were accurate compared with actual volumes calculated using a water displacement technique. In an in vivo study, Shirota et al.[12] found a positive correlation between graft bone volume estimated by CBCT data and actual bone volume measured by syringe during surgery.

Given the potential value of preoperative 3D analysis in estimating grafting success as well as required grafting volume, coupled with the lack of 3D structure data, further study of maxillae and associated clefts is warranted. It is obvious to observe that irregular shapes of maxillae usually appear on patients suffering alveolar defects. However, the morphological relationship between defects and maxillae is still not clear. Based on this observation, we hypothesize that the defect size is positively correlated with the affected maxillary half. In this study, a machine learning segmentation method[13,14] (LINKS) was adopted to isolate the alveolar cleft defects and maxillae from CBCT images to test our hypothesis. This study aims to (1) quantify the 3D structural parameters of the maxilla and defect based on CBCT imaging segmentation; (2) investigate the

morphological relationship between the maxilla and defect; and (3) present a probability map to characterize the cleft defect.

Methods

Subjects

This study was approved by the institutional review board (IRB: KY2017-072-01) of Beijing Tiantan Hospital, Capital Medical University. All CBCT data was obtained from the Department of Orthodontics, Beijing Stomatological Hospital, Capital Medical University. The inclusion criteria were non-syndromic Chinese origin subjects diagnosed with UCLP who received initial lip and palate repair. The exclusion criteria were previous orthodontic treatment and maxillary dental/skeletal trauma or surgery. All of these CBCT images were routinely acquired for dental/skeletal evaluation during orthodontic/surgical planning. A total of 60 subjects were included in this study, consisting of 39 men (65%) and 21 women (35%), with a mean age of 12.08 years (SD=4.83 years; range of 8-33 years), presenting with 41 left-side defects (68%) and 19 right-side defects (32%).

All images were acquired on the same CBCT scanner (NewTom, Verona, Italy) (110 kV, 1-20 mA, 15×15 cm field of view, 0.250 mm voxel size). The subjects were in an upright position with the Frankfort plane parallel to the floor. The patients were asked to bite in intercuspation position, breathe slowly, and not swallow during the radiograph procedure.

Maxilla and Defect Segmentation

The 60 CBCT image sets were exported in DICOM format and imported into the 3D imaging analysis software ITK-SNAP[15] (Version 3.6.0; www.itksnap.org). To expedite the computer algorithm processing, the orientation of each CBCT image was adjusted to have every defect on the left side of the maxilla. The maxilla was labeled and reconstructed by selectively identifying the suture where it articulated with the adjacent cranial and facial bones, including the

frontal, ethmoid, zygomatic, and pterygoid plate of the sphenoid. The defect was marked and reconstructed following the contralateral shape to obtain a continuation of the alveolar ridge and hard palate.

The automatic segmentation protocol used in this study is based on an advanced machine learning technique, sequential random forest classifiers with prior guidance, which Wang et al. have described in detail in their previous studies.[13,14,16] Since this technique is a supervised learning method, 30 randomly selected CBCT images were manually segmented, serving as the ground truth. These 30 ground-truth CBCT images were further split into training and test samples. The training stage began with acquiring the initial segmentation probability maps of the maxilla and defect based on the training samples. Then, the subsequent random forest classifier was iteratively trained using the original CBCT features and the segmentation probability maps. In the test stage, sequentially trained classifiers were used to estimate corresponding probability maps of the maxilla and defect of the test samples. The probability maps (i.e., automatic segmentation) of the test samples were compared to the manually segmented samples to evaluate the accuracy of the model. The accuracy of the segmentation model was evaluated by the Dice similarity coefficient (DSC), defined as follows:

$$DSC = \frac{2|A \cap B|}{|A| + |B|},$$

where $|A|$ and $|B|$ represent the cardinalities of the learned and manual sets, and $|A \cap B|$ represents the intersection of the two sets. A value of 0 indicates no similarity, whereas a value of 1 indicates perfect agreement.

A three-fold cross-validation was performed to evaluate the accuracy of the model. The 30 ground-truth CBCT images were split into three groups with each group containing ten images. One unique group was considered the test group, and the remaining two groups served as training groups to create the models. This process was repeated until each group was selected as a test group. The accuracy of this method was found by calculating the average DSC of all the testing samples. By performing three-fold cross-validation, we determined a reliable set of hyper-parameters of the random forest algorithm. Then, all 30 ground-truth images were used with these hyper-parameters to re-train a set of classifiers. In this way, we obtained the best model possible, based on the efficient use of all the ground-truth images. These final, well-trained classifiers were then applied to another set of 30 CBCT images to obtain the automatic segmentation results. Finally, all 60 pairs of segmented maxilla and defect 3D models were used for regression analysis.

Description of Measurement

A coordinate system was established to standardize the orientation of the maxilla and defect for the purpose of quantitative analysis. The Frankfort horizontal plane (FH plane) was the plane that passed through bilateral Porion (Po) and Orbitale (Or) on the non-defect side. The midsagittal plane (MS plane) was the plane perpendicular to the FH plane passing through Nasion (N) and Sella (S). The coronal plane (CR plane) was the plane perpendicular to the FH and MS planes passing through Nasion (N). The parameters of the maxilla and defect were measured in horizontal, midsagittal, and coronal plane projections as well as 3D segmentations, including the length, width, height, and volume. The maxilla was separated by the midsagittal plane for measurements on the defect and non-defect sides. The six maxillary parameters were as follows: maxillary length (L_{max}), maxillary anterior width ($AntW_{max}$), maxillary posterior width ($PosW_{max}$), maxillary anterior height ($AntH_{max}$), maxillary posterior height ($PosH_{max}$), and maxillary volume (V_{max}). The maxillary parameters were measured on the defect and non-defect sides (denoted by superscript ^{def}

and ^{nor}), respectively. All parameters are defined in Table 4, while various maxilla and defect measurements are shown in Figures 10 and 11.

Table 4. Definitions and abbreviations of reference planes, maxilla distances, and defect distances used for quantitative analysis standardization is shown.

	Parameter	**Abbreviation**	**Definition**
Reference plane	Horizontal plane	FH plane	Plane that passes through the bilateral Porion (Po) and Orbitale (Or) on the non-defect side
	Midsagittal plane	MS plane	Plane perpendicular to the FH plane passing through the Nasion (N) and Sella (S)
	Coronal plane	CR plane	Plane perpendicular to the FH and MS plane passing through the Nasion (N)
Maxilla	Maxillary length	L_{max}	Maximum posteroanterior (PA) distance from the maxillary tuberosity to the anterior contour of the maxilla in lateral view
	Maxillary anterior width	$AntW_{max}$	Minimum transverse distance from the outer cortices of the root apices of canines to the MS plane
	Maxillary posterior width	$PosW_{max}$	Minimum transverse distance from the outer cortices of the first molars to the MS plane
	Maxillary anterior height	$AntH_{max}$	Vertical distance between the inferior most extent of the continuous nasal floor and the inferior orbital rim
	Maxillary posterior height	$PosH_{max}$	Vertical distance from the alveolar crest of the first molar to the superior most extent of the maxillary sinus
	Maxillary volume	V_{max}	Volume of the segmented individual maxilla
Defect	Defect length	L_{def}	Maximum posteroanterior (PA) distance of the defect
	Defect width	W_{def}	Maximum transverse distance of the defect
	Defect height	H_{def}	Maximum vertical distance of the defect
	Defect volume	V_{def}	Volume of the segmented defect

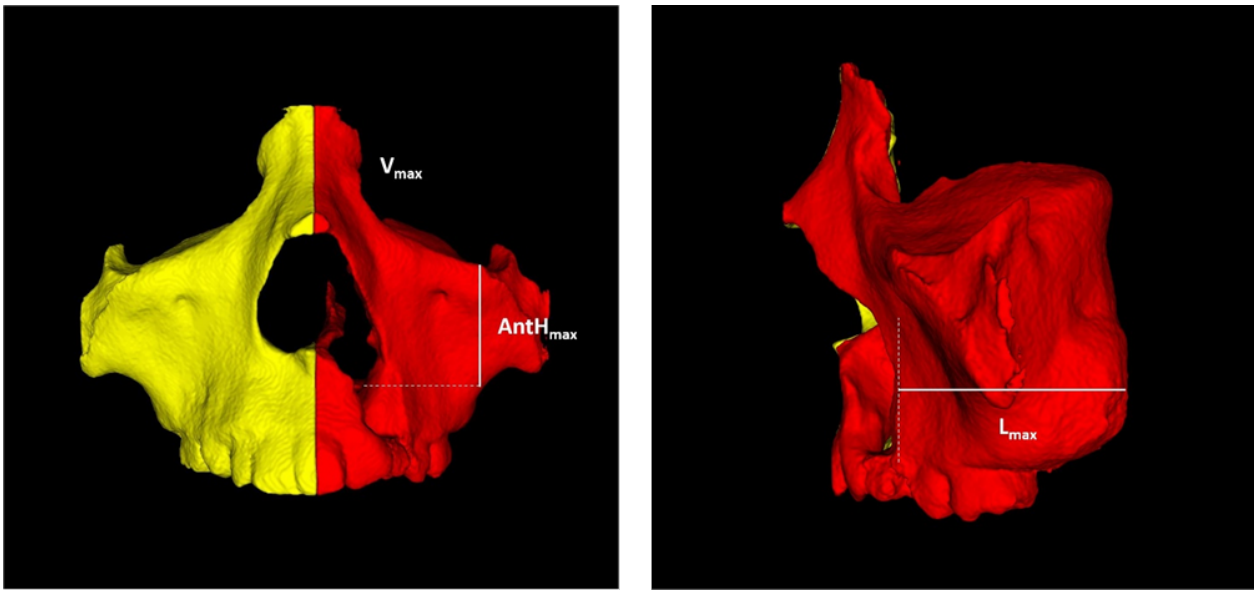


Figure 10. Maxillary volume, anterior height, and length. The measurement of maxillary anterior height ($AntH_{max}$) is defined by the vertical distance between the inferior most extent of the continuous nasal floor and the inferior orbital rim. The measurement of maxillary length (L_{max}) is defined by the maximum posteroanterior (PA) distance from the posterior most extend of the anterior contour of the maxilla to the posterior most extent of the maxillary tuberosity.

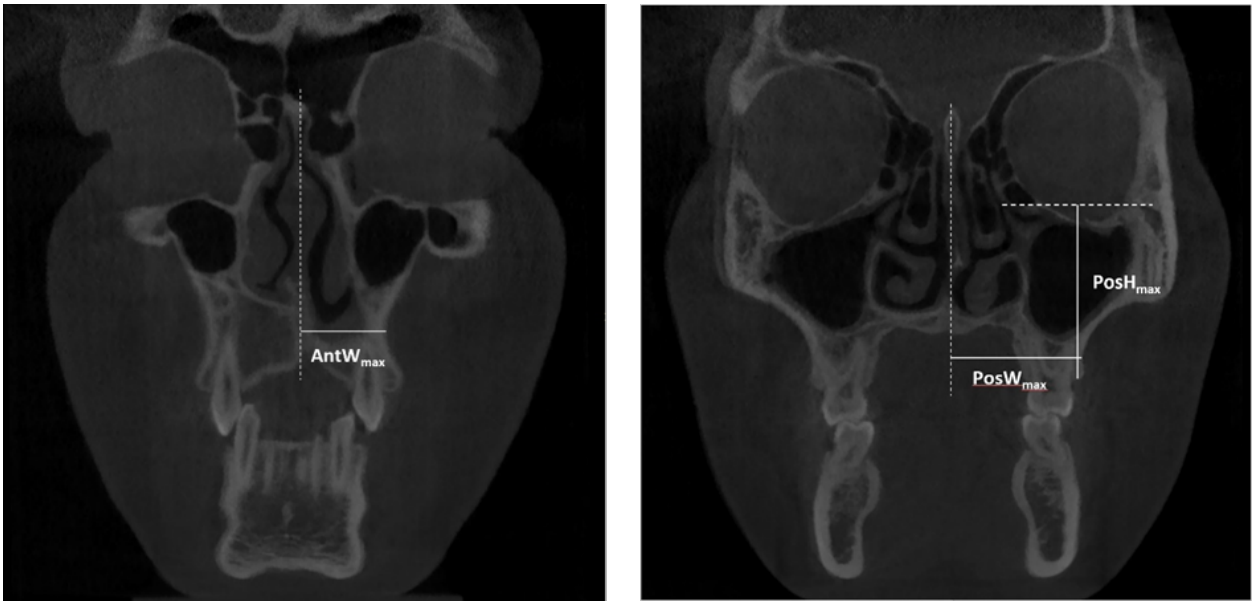


Figure 11. Maxillary anterior width, posterior width, and posterior height. To the left, $AntW_{max}$ indicates maxillary anterior width (defect side). To the right, $PosW_{max}$ indicates maxillary posterior width (defect side) and $PosH_{max}$ indicates maxillary posterior height (defect side).

The length, width, and height were measured by calculating the distance between the position (voxel coordinates) of landmarks, and the volume of the maxilla and defect were measured based on the segmentation voxel counting. These measurements were carried out using ITK-SNAP.

All distances were measured twice (i.e., landmarks were selected twice) on two separate occasions two weeks apart by one observer. The mean values were used for statistical analysis. The intra-observer reliability was assessed using the intra-class correlation coefficient (ICC), yielding an ICC greater than 0.88 (95% confidence interval).

Defect Distribution

From the superimposed 3D models, the distribution map of the defects was presented visually and multi-directionally, allowing for visualization of the distribution probability. The defect distribution probability is shown in Figure 12. The superimposition of all 60 segmentations of the defect were completed with each voxel.

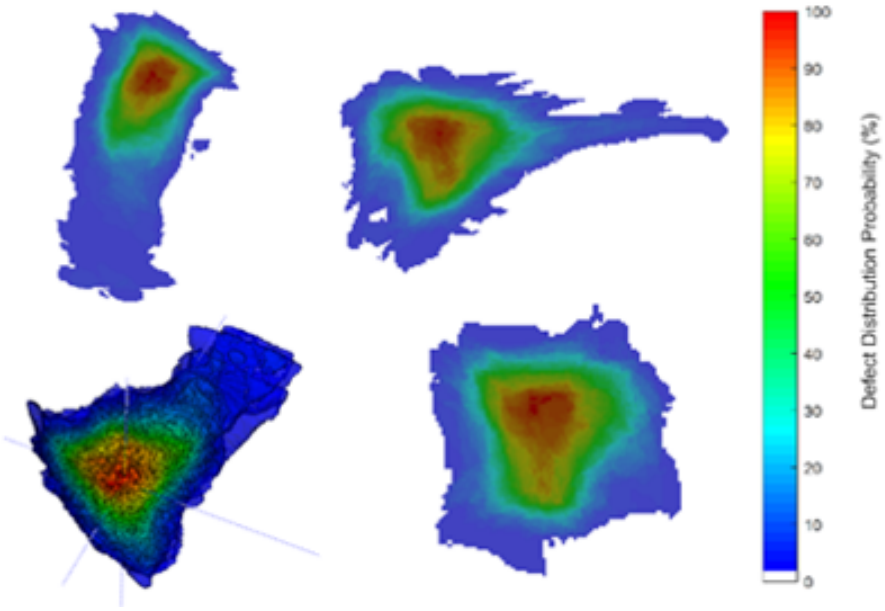


Figure 12. Defect distribution probability maps from superimposed 3D models.

Statistical Analysis

The data is presented as mean values and standard deviations. The maxillary asymmetry between the defect and non-defect sides was compared using the Paired-Samples *t* Test with a significance level of $p<0.05$. A multiple linear regression was carried out to analyze the relationship between the parameters of the defect and parameters of the defect side of the maxilla. The age and gender of subjects were also included in the regression. All statistical analyses were done with SPSS (Version 19.0; IBM Co., Armonk, NY, USA).

Results

Maxilla and Defect Segmentation

The Dice similarity coefficients of the maxilla and defect between the manual and automatic segmentations in the test samples were 0.852±0.02 and 0.75±0.07, respectively, suggesting that the model was accurate. The manual segmentation of both the maxilla and defect in test samples and their corresponding automatic segmentations were compared and shown in Figures 13 and 14, where the auto-segmented maxilla and defect were observed to have features morphologically similar to those obtained via manual segmentation.

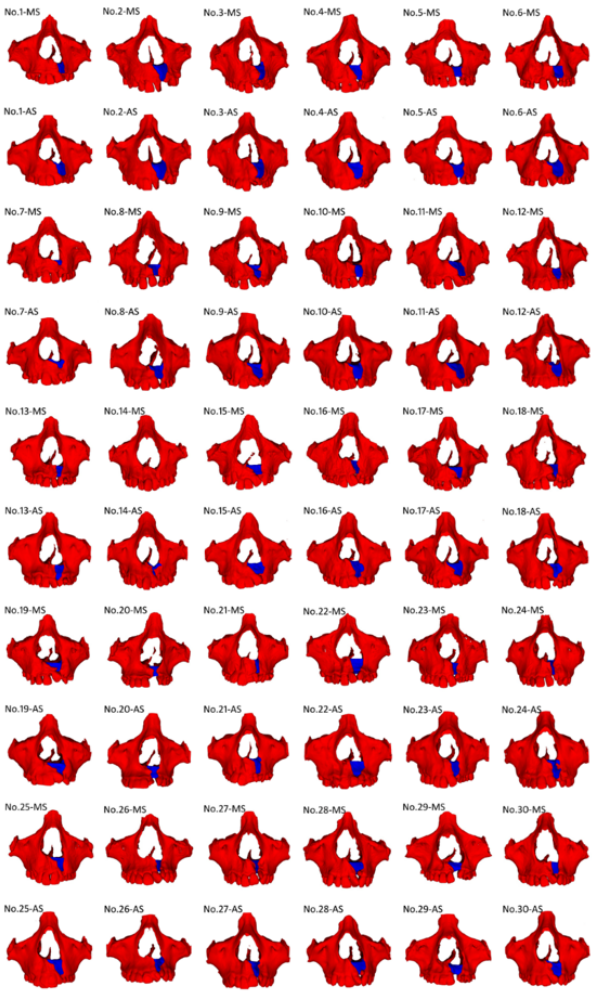


Figure 13. Maxilla (red) and defect (blue) renderings using 3D imaging software. Both manual segmentation (MS) and automatic segmentation (AS) results are shown.

Figure 14. Defect renderings using 3D imaging software. Both manual segmentation (MS) and automatic segmentation (AS) results are shown.

Statistically significant differences were observed upon maxillary asymmetry analysis, shown in Table 5. There were several differences in the measurements between the defect and non-defect sides, mostly concerning the anterior maxilla around the defect region. The defect side demonstrated a significant decrease in maxillary length L_{max}, anterior width $AntW_{max}$, and volume V_{max} with mean measurements of 34.31±2.56mm, 17.83±2.06mm, and

18.02±3.24x$10^3$$mm^3$, respectively. In comparison, the non-defect side had mean measurements of 41.27±3.72mm, 20.31±2.48mm, and 19.32±3.53x$10^3$$mm^3$, respectively for these parameters. A significant increase in maxillary anterior height $AntH_{max}$ was demonstrated for the defect side, with a mean of 25.91±4.12mm as compared to the non-defect side at 21.50±3.45mm. The maxillary posterior width $PosW_{max}$ and height $PosH_{max}$ demonstrated no statistically significant differences between the defect and non-defect sides.

Table 5. Measurement and analysis of the defect and non-defect sides of the maxilla. Measurements were obtained as the mean and standard deviation (SD). Significance of $p<0.05$ is denoted with an asterisk.

Parameter	Defect side		Non-defect side		P
	Mean	SD	Mean	SD	
L_{max} (mm)	34.31	2.56	41.27	3.72	0.000*
$AntW_{max}$ (mm)	17.83	2.06	20.31	2.48	0.000*
$PosW_{max}$ (mm)	29.08	2.47	29.22	2.29	0.652
$AntH_{max}$ (mm)	25.91	4.12	21.50	3.45	0.000*
$PosH_{max}$ (mm)	42.92	5.13	42.72	5.15	0.380
V_{max} (×10^3 mm^3)	18.02	3.24	19.32	3.53	0.000*

Defect Distribution

Superimposition of the defect illustrated a tendency toward concentrated distribution. The defect distribution probability was high at the alveolar bone region and gradually decreased toward the margin of the defect, with the extended hard palate portion of the defect illustrating a low distribution probability, as shown in Figure 15. The defect structure parameters of length, width, height, and volume are shown in Table 6.

Table 6. Defect structure parameters and measurements. Measurements were obtained as the mean and standard deviation (SD).

Parameter	Mean	SD
L_{def} (mm)	22.84	5.07
W_{def} (mm)	18.01	2.08
H_{def} (mm)	16.94	2.59
V_{def} ($\times 10^3$ mm^3)	1.24	0.29

Multiple Linear Regression

Since cleft lip and palate can be caused by genetics, we considered defect parameters as independent variables and maxillary parameters on the defect side as dependent variables. Each analysis was adjusted for age and gender. After performing a multiple linear regression, it was found that the defect-side maxillary length (L_{max}^{def}), posterior height ($PosH_{max}^{def}$), and volume (V_{max}^{def}) were significantly related to the defect height (H_{def}) with p=0.022, 0.001, and 0.029, respectively. Moreover, it was also found that the defect-side maxillary posterior width ($PosW_{max}^{def}$) was significantly related to the defect length (L_{def}) with p=0.040, and that the maxillary anterior width on the defect side ($AntW_{max}^{def}$) was slightly related to the defect width (W_{def}) with p=0.054. Furthermore, no significant relationship was found between the maxillary anterior height ($AntH_{max}^{def}$) and any defect variables. The linear regression analysis results between the maxillary and defect parameters are presented in Table 7. The volume of the defect was also significantly related to the defect length, width, and height with p=0.006, 0.000, and 0.000, respectively (see Table 8).

Table 7. Results of multiple linear regression analysis regarding the defect and relationship to the maxilla with adjusted age and gender. Significance of p<0.05 is denoted with an asterisk.

Dependent variable	Independent variable	Coefficient	Standard error	P	R^2 adjusted

L_{max}^{def} (mm)	L_{def} (mm)	-0.007	0.069	0.920	0.191
	W_{def} (mm)	-0.125	0.194	0.522	--
	H_{def} (mm)	0.348	0.147	0.022*	--
	V_{def} ($\times 10^3$ mm^3)	-0.945	1.674	0.575	--
$AntW_{max}^{def}$ (mm)	L_{def} (mm)	-0.058	0.060	0.333	0.059
	W_{def} (mm)	0.332	0.169	0.054	--
	H_{def} (mm)	0.097	0.127	0.450	--
	V_{def} ($\times 10^3$ mm^3)	-0.906	1.451	0.535	--
$PosW_{max}^{def}$ (mm)	L_{def} (mm)	-0.1471	0.070	0.040*	0.112
	W_{def} (mm)	0.289	0.197	0.148	--
	H_{def} (mm)	0.183	0.149	0.224	--
	V_{def} ($\times 10^3$ mm^3)	0.156	1.695	0.927	--
$AntH_{max}^{def}$ (mm)	L_{def} (mm)	-0.035	0.094	0.711	0.415
	W_{def} (mm)	0.163	0.266	0.542	--
	H_{def} (mm)	0.089	0.201	0.661	--
	V_{def} ($\times 10^3$ mm^3)	3.078	2.290	0.185	--
$PosH_{max}^{def}$ (mm)	L_{def} (mm)	-0.093	0.111	0.405	0.479
	W_{def} (mm)	-0.317	0.313	0.316	--
	H_{def} (mm)	0.842	0.237	0.001*	--
	V_{def} ($\times 10^3$ mm^3)	3.081	2.697	0.258	--
V_{max}^{def} ($\times 10^3$ mm^3)	L_{def} (mm)	-0.099	0.075	0.191	0.399

	W_{def} (mm)	0.022	0.212	0.916	--
	H_{def} (mm)	0.359	0.160	0.029*	--
	V_{def} ($\times 10^3$ mm^3)	1.976	1.826	0.284	--

Table 8. Results of multiple linear regression analysis regarding the defect volume and relationship to the defect length, width, and height. Significance of $p<0.05$ is denoted with an asterisk.

Dependent variable	Independent variable	Coefficient	Standard error	P	R^2 adjusted
V_{def} ($\times 10^3$ mm^3)	L_{def} (mm)	0.015	0.005	0.006*	0.579
	W_{def} (mm)	0.060	0.014	0.000*	--
	H_{def} (mm)	0.040	0.011	0.000*	--

Discussion

This study conducted automatic segmentation of the maxilla and cleft defect using CBCT images, quantified the 3D structural parameters of the maxilla and defect, and investigated the morphological relationship between them. Achieving a deeper understanding of defect variability and the defect's anatomical relationship with the maxilla was the major goal, contributing to the limited body of information regarding 3D assessment of craniofacial anomalies, namely non-syndromic unilateral cleft lip and palate (UCLP). This study had a relatively large sample size, with 60 CBCT image sets. The subjects consisted of 65% men and 35% women, of which 68% presented with left-side defects and 32% with right-side defects. Sex and location preferences were observed from the composition of the sample studied. Similar results have also been reported in several other large scale surveys. For example, Xiong et al.[17] and Wenzhe et al.[18] reported in their

retrospective reviews that both the sex ratio of men and women and the ratio of left- and right-side defects were approximately 2:1. Although many genetic and environmental factors related to the cleft lip and palate were identified,[19] the mechanism underlying these conditions is still unknown.

The use of CBCT for preoperative volumetric assessment of ABGs has been well established, and its accuracy has been demonstrated extensively in the literature.[11] De Rezende et al.[4] indicated that CBCT volumes proved reliable for the volumetric assessment of alveolar cleft defects using different methods. However, these methods involved the threshold-selected auto-segmentation of CBCT images, which could not easily separate the maxilla from other bony structures of the skull. One possible method to overcome this gap is a machine-learning based segmentation algorithm because it takes into account not only intensity but also morphological characteristics from each voxel.

With the aid of CBCT, it seems simple to conclude that the defect side of the maxilla is obviously hypoplastic, although comprehensive analysis is still deficient in the literature. In this study, the quantitative assessment indicated that the maxillary volume, length, and anterior width were significantly smaller on the defect side. In addition, the maxillary posterior width and height demonstrated no statistically significant differences on either side. These findings are not surprising as the defect is located in the anterior portion of the maxilla, and there is no evidence that the posterior portion of the maxilla is affected. Suri et al.[20] and Li et al.[21] compared the asymmetry of the defect and non-defect sides using spiral CT. They both concluded that significant differences between the defect and non-defect sides were displayed in the area near the defect and not in the deeper regions of the maxilla. The maxilla is an irregular bone with one body and four processes. Thus, it is difficult to thoroughly evaluate the complex 3D maxillary structure with conventional linear measurements. Volume is, therefore, the best index with which to assess the

amount of overall maxillary hypoplasia. Agarwal et al.[22] calculated maxillary volume by manual segmentation of the individual maxilla and found reduced values on the defect side, which mirrors our results in this study. Furthermore, it is interesting to note that the value of the maxillary anterior height on the defect side is significantly larger than that on the non-defect side. The difference is most likely due to the nasal floor, which exhibits tilting toward the defect side (see Figure 10). A similar phenomenon was also described in a previous study by Schneiderman et al.[23]

To the best of our knowledge, ours is the first study to evaluate the defect distribution probability. The results (see Figure 12) showed that the distribution of the defect had certain predictable characteristics, specifically regarding a concentrated distribution at the alveolar bone region which decreased toward the defect margins, with the extended hard palate portion of the defect illustrating a low distribution probability. Many studies of various designs have been used to investigate defects with the common goal of evaluating defect volume, regardless of the defect shape, alveolar bone graft, or simulant volume.[10,24] This limitation is likely due to the fact that the defect lacks anatomical features. So far, there is no widely accepted method to study the cleft defect due to its varying size and shape. For treatment planning, previous studies have mainly focused on the accuracy of the graft volume. The powdered or small granule forms of bone, or its substitutes, were most commonly used in ABG.[25,26] Thus, compared with its shape, the defect volume was deemed as the key factor associated with optimal surgical outcomes. With 3D printed biomaterial scaffolds increasing in popularity, there is a growing need for new methods of morphological assessment. To help meet this need, the defect distribution probability map used in this study could provide relatively detailed information regarding two different aspects. The first is visualization of the map's distribution probability per voxel. The second aspect is the visual and multi-directional morphology and volume of the defect at a certain percentage of the distribution

probability. In the future, defect distribution probability mapping could provide a solid foundation for comprehensive morphological analysis.

Determining which defect factors are responsible for maxillary variability has proved challenging. Barbosa et al.[2] revealed some clues indicating that the defect volume is related to the gap, arch form, nasal base defect, and dental parameters to a certain degree. Their study presented a classification system for UCLP based on CBCT images and evaluated the relationship between the maxillary volume and gap (or cleft size) as well as the defect size in the nasal base region/piriform margin. Although a statistically significant relationship was not observed, the researchers did find a growing trend demonstrating that the larger the defect size, the larger the maxillary volume.

A multiple linear regression was used in this study to further investigate the underlying relationship between the maxilla and the defect, with the results demonstrating a complicated relationship (see Table 7). In general, the maxilla on the cleft side has decreased width, height, length, and volume. Among the results, the cleft height likely has a pivotal role in the variability of the maxilla. The lowered nasal floor on the defect side is a structure that is characteristically deformed in almost all UCLP subjects in this study. Cleft patients typically have a defect in the area of the nasal floor caudal to the pyriform fossa, which correpsonds to the cleft itself. Since the projection of the maxilla is also deficient in this area, the result is an alar base depression. The decreased length of the maxilla on the cleft side indicates that the anterolateral surface of the maxilla and area around the pyriform margin are hypoplastic. The defect length, width, and height exhibited a positive correlation with the cleft volume (see Table 8). However, it should be noted that the low adjusted R^2 values suggest a relatively weak relationship between the maxilla and

defect, as shown in Tables 7 and 8. Thus, more variables should be taken into account in future studies to identify factors responsible for the variability of the maxilla.

One major limitation of this study is automatic segmentation accuracy. Although our accuracy is comparable to that found in other medical fields (i.e., averaged DSC is roughly 0.75 for the defect), it still needs improvement, particularly for the defects characterized by a long, extended hard palate (see No. 23 and 30 in Figures 13 and 14) due to their relatively low proportion. In addition, as UCLP is commonly concomitant with malocclusion resulting from crowding or teeth dislocations, the difficulty of accurate segmentation is increased. We have also performed another automatic segmentation using a deep learning method, U-Net[27,28], for comparison. However, the results revealed that U-Net provided comparable accuracy. To further improve the accuracy of segmentation, the inclusion of more training subjects and more advanced machine learning algorithms are suggested in future studies.

Conclusions

We conducted a 3D automatic segmentation for CBCT images of non-syndromic unilateral cleft lip and palate (UCLP). Based on the resultant auto-segmented models, the anterior portion of the maxilla was found to be significantly smaller on the defect side as compared to the non-defect side. The defect distribution probability map showed a concentrated distribution in the alveolar bone region and a low distribution probability toward the margin of the defect. Furthermore, certain structures of the anterior maxilla may likely contribute to the variability of defect, though additional studies are needed. This study accurately quantified each unique cleft defect in 3D to increase UCLP understanding and demonstrated the potential for large scale clinical applications.

The auto-segmented 3D model and morphometric quantification of the defect allows a quick estimation of the extent and complexity of ABG surgery. The models have the potential to

serve as templates for 3D printed biomaterial scaffolds. Furthermore, the maxillary parameters and their relationship with the defect parameters help with estimating the magnitude of the maxillary asymmetry and whether additional preoperative preparation is necessary, such as orthodontic expansion or nasal floor augmentation.

REFERENCES

1. Anon. Prevalence at Birth of Cleft Lip with or without Cleft Palate: Data from the International Perinatal Database of Typical Oral Clefts (IPDTOC). *Cleft Palate-Craniofacial J.* 2011;48(1):66–81.

2. Barbosa GL deR., Emodi O, Pretti H, van Aalst JA, de Almeida SM, Tyndall DA, Pimenta LA. GAND classification and volumetric assessment of unilateral cleft lip and palate malformations using cone beam computed tomography. *Int. J. Oral Maxillofac. Surg.* 2018;45(11):1333–40.

3. Li D, Jun Z, Meng M, Yanhua L, Ying D, Yanping W, Juan L, Liu T, et al. Time trends in oral clefts in Chinese newborns: Data from the Chinese National Birth Defects Monitoring Network. *Birth Defects Res. Part A Clin. Mol. Teratol.* 2010;88(1):41–7.

4. de Rezende Barbosa GL, Wood JS, Pimenta LA, Maria de Almeida S, Tyndall DA. Comparison of different methods to assess alveolar cleft defects in cone beam CT images. *Dentomaxillofacial Radiol*. 2016;45(2):20150332.

5. Feng B, Jiang M, Xu X, Li J. A new method of volumetric assessment of alveolar bone grafting for cleft patients using cone beam computed tomography. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol*. 2017;124(2):e171–82.

6. Linderup BW, Cattaneo PM, Jensen J, Kuseler A. Mandibular Symphyseal Bone Graft for Reconstruction of Alveolar Cleft Defects: Volumetric Assessment with Cone Beam Computed Tomography 1-Year Postsurgery. *Cleft Palate-Craniofacial J.* 2016;53(1):64–72.

7. Oberoi S, Chigurupati R, Gill P, Hoffman WY, Vargervik K. Volumetric Assessment of Secondary Alveolar Bone Grafting Using Cone Beam Computed Tomography. *Cleft Palate-Craniofacial J.* 2009;46(5):503–11.

8. van dr Meij AJW, Baart JA, Prahl-Andersen B, Kostense PJ, van der Sijp JR, Tuinzing DB. Outcome of bone grafting in relation to cleft width in unilateral cleft lip and palate patients. *Oral Surgery, Oral Med. Oral Pathol. Oral Radiol. Endodontology* 2003;96(1):19–25.

9. Long RE, Spangler BE, Yow M. Cleft Width and Secondary Alveolar Bone Graft Success. *Cleft Palate-Craniofacial J.* 1995;32(5):420–7.

10. Albuquerque MA, Gaia BF, Cavalcanti MGP. Comparison between multislice and cone-beam computerized tomography in the volumetric assessment of cleft palate. *Oral Surgery, Oral Med. Oral Pathol. Oral Radiol. Endod.* 2011;112(2):249–57.

11. Amirlak B, Tang CJ, Becker D, Palomo JM, Gosain AK. Volumetric Analysis of Simulated Alveolar Cleft Defects and Bone Grafts Using Cone Beam Computed Tomography. *Plast. Reconstr. Surg.* 2013;131(4).

12. Shirota T, Kurabayashi H, Ogura H, Seki K, Maki K, Shintani S. Analysis of bone volume using computer simulation system for secondary bone graft in alveolar cleft. *Int. J. Oral Maxillofac. Surg.* 2010;39(9):904–8.

13. Wang L, Chen KC, Gao Y, Shi F, Liao S, Li G, Shen SGF, Yan J, et al. Automated bone segmentation from dental CBCT images using patch-based sparse representation and convex optimization. *Med. Phys.* 2014;41(4):043503.

14. Wang L, Gao Y, Shi F, Li G, Chen KC, Tang Z, Xia JJ, Shen D. Automated segmentation of CBCT image with prior-guided sequential random forest. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 2016;9601 LNCS(1):72–82.

15. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 2006;31(3):1116–28.

16. Chen S, Wang L, Li G, Wu T-H, Diachina S, Tejera B, Kwon JJ, Lin F-C, et al. Machine learning in orthodontics: Introducing a 3D auto-segmentation and auto-landmark finder of CBCT images to assess maxillary constriction in unilateral impacted canine patients. *Angle Orthod.* 2019.

17. Xiong B, Zhao M, Cheng T, Gao P. [Analysis of 5459 cleft lip and palate cases]. *Zhonghua Zheng Xing Wai Ke Za Zhi* 2002;18(5):294–6.

18. Hua W, Guo J, Li N, Hou D, Liu R. Cleft Lip and Cleft Palate in Gansu Province：A Retrospective Review of 9720 Cases. *J. Oral Maxillofac. Surg.* 2014;(5):356–9.

19. Leslie EJ, Marazita ML. Genetics of Cleft Lip and Cleft Palate. *Am. J. Med. Genet. C. Semin. Med. Genet.* 2013;163(4):246–58.

20. Suri S, Utreja A, Khandelwal N, Mago SK. Craniofacial Computerized Tomography Analysis of the midface of patients with repaired complete unilateral cleft lip and palate. *Am. J. Orthod. Dentofac. Orthop.* 2008;134(3):418–29.

21. Li H, Yang Y, Chen Y, Wu Y, Zhang Y, Wu D, Liang Y. Three-Dimensional Reconstruction of Maxillae Using Spiral Computed Tomography and Its Application in Postoperative Adult Patients With Unilateral Complete Cleft Lip and Palate. *J. Oral Maxillofac. Surg.* 2011;69(12):e549–57.

22. Agarwal R, Parihar A, Mandhani PA, Chandra R. Three-Dimensional Computed Tomographic Analysis of the Maxilla in Unilateral Cleft Lip and Palate. *J. Craniofac. Surg.* 2012;23(5):1338–42.

23. Schneiderman ED, Xu H, Salyer KE. Characterization of the maxillary complex in unilateral cleft lip and palate using cone-beam computed tomography: A preliminary study. *J.*

Craniofac. Surg. 2009;20(8 SUPPL. 2):1699–710.

24. Kasaven CP, McIntyre GT, Mossey PA. Accuracy of both virtual and printed 3-dimensional models for volumetric measurement of alveolar clefts before grafting with alveolar bone compared with a validated algorithm: a preliminary investigation. *Br. J. Oral Maxillofac. Surg.* 2017;55(1):31–6.

25. Nyberg E, Rindone A, Dorafshar A, Grayson WL. Comparison of 3D-Printed Poly-ε-Caprolactone Scaffolds Functionalized with Tricalcium Phosphate, Hydroxyapatite, Bio-Oss, or Decellularized Bone Matrix. *Tissue Eng. Part A* 2016;23(11–12):503–14.

26. Choi HS, Choi HG, Kim SH, Park HJ, Shin DH, Jo DI, Kim CK, Uhm K Il. Influence of the Alveolar Cleft Type on Preoperative Estimation Using 3D CT Assessment for Alveolar Cleft. *Arch. Plast. Surg.* 2012;39(5):477–82.

27. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention* Springer; 2015:234–41.

28. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation BT - Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, eds. Cham: Springer International Publishing; 2016:424–32.