EMERGENCE OF MACHINE LEARNING TOOLS IN ORTHODONTIC PRACTICE AND RESEARCH

Maxime Gillot, Baptiste Baquero, Mathieu Leclercq, Marcela Gurgel, Najla Al Turkestani, Jonas Bianchi, Lucia Cevidanes, Juan Carlos Prieto

ABSTRACT

After the emergence of artificial intelligence (AI) in the 1950s, the technology went through periods of optimism ("AI spring"), promising life-changing applications followed by periods of disappointment ("AI winter"), where the promises seemed unattainable. During the past decade, fast paced breakthroughs in machine learning (ML) reached the point where AI actually started to have a direct and increasing impact on the quality of everyday life. It seemed as though we were in a long and promising "spring" in which not a single sector was spared, including the field of orthodontics. This chapter offers a quick overview of the roles that machine learning is currently playing in assisting in orthodontics and orthodontic research. ML is able to process images more quickly and with greater consistency via automated clinical decision support systems than what can be accomplished by humans when performing basic steps such as landmark identification and segmentation.

KEY WORDS: Automated Tools, Segmentation, Landmark Identification, Deep Learning, Machine Learning

INTRODUCTION

After many breakthroughs in computing power and AI research over the past decade, the development of machine learning tools has had a direct and increasing impact on everyday life. Currently, most social networks offer personalized recommendations to consumers through AI algorithms [1, 2]. In agriculture, ML applications in various areas have shown outstanding results, from detecting weeds and diseases, to offering predictions regarding livestock production and quality of crops [3]. Automobile companies get closer each day to a fully self-driving car [4]. In research, Google DeepMind recently made AlphaFold, a solution to a 50-year-old grand challenge in biology, available for free to the whole world [5]. DeepMind has already been used to predict protein structures of SARS-CoV-2 [6]. The list of examples is endless, not a single field is spared by the emergence of ML tools, including orthodontics.

Accurate segmentation, registration, multimodality fusion, and quantification of multi-facial structures, including the maxilla, mandible, mandibular nerve canal, dental crowns, roots, and their root canals are vital to determine safety margins in the bones, adjacent teeth, and gingival margins in dental, oral, and craniofacial (DOC) care. The current tools such as ITK-SNAP [7] and 3D Slicer [8] are set up for individual and manual scan registration. It is a time-consuming and challenging task, due to the low signal/noise ratio, particularly in full face cone-beam computed tomography (CBCT) scans. Developing AI to perform fast and robust registration, segmentation, and landmark identification on a large set of CBCT or intraoral scans (IOS) could also allow experts to reach high stability and agreement on orthodontic decision-making procedures and treatment effect evaluation.

Recent ML models have been found to perform at or above the accuracy levels achieved by humans in landmark identification, skeletal classification, bone age prediction, and tooth segmentation. Commercial companies have recently marketed AI-based segmentation and registration tools for CBCT and IOS scans, but they are expensive and the precision of their algorithms is not yet generalizable and they do not provide quantitative decision support tools. Our lab worked on improving this generalizability and developing free, open-source tools to perform segmentation and landmark identification tasks by using large databases. This goal was made possible thanks to a collaboration of 15 clinician centers from all over the world: University of Michigan, University of Pacific, University of Minnesota, University of Medellin (Colombia), University Científica del Sur (Peru), University of Firenze and Genoa (Italy), State University of Sao Paulo, Federal University of Ceara, Federal University of Goias and Federal University of Rio de Janeiro, Hospital for Rehabilitation of Craniofacial Anomalies in Bauru (Brazil), and two private practices in the US and in Egypt. Through common effort and collaboration, we have developed four machine learning tools ready to be used by clinicians on 3D Slicer and our web-based system, the Data Storage for Computation and Integration (DSCI). FiboSeg and AMASSS were developed to automatically segment IOS and CBCT, respectively, while Automatic Landmark Identification (ALI) in CBCT ALI-CBCT and in IOS (ALI-IOS) specialize in landmark identification on both supports.

MACHINE LEARNING

An AI is the simulation of human intelligence processes by machines. Simply, it can be considered as a function that intelligently processes the input to give output as if computer systems were smart. Machine Learning is a subcategory of artificial intelligence, that refers to the process by which computers learn from a given dataset to generalize and make accurate predictions when exposed to new data. In ML the function is a mathematical equation made of adjustable parameters represented by what we call a neural network, or artificial neural network (ANN) [9]. It requires an input layer and an output layer separated by one or more hidden layers. We speak about deep learning when a neural network consists of more than three layers. Today's deepest neural Network is GPT-3 from Open AI [10]. It is a model with over 175 billion machine learning parameters specialized to produce human-like text. This number might seem large, but it stays rather small compared to over 100 trillion synapses (up to 1,000 trillion, by some estimates) connecting around 100 billion neurons in the human brain [11].

For an AI, learning means adjusting the parameters of the neural network to return an expected output from a given input. The machine learning field is made of 4 main branches used to perform different tasks [12]. They differ in the strategy used to modify the network parameters. The most common are supervised and unsupervised learning followed by semi-supervised learning and reinforcement learning.

Supervised learning

As the name suggests, supervised learning involves learning in the presence of a supervisor. For machine learning algorithms, it means providing feedback on the performances during the training process. To do so, the given dataset is made of pre-labeled data. The parameters of the neural network are slightly updated at each training iteration to reduce the error between the network prediction and the labeled output. It is used for tasks such as image recognition, medical diagnosis, statistical arbitrage, predictive analysis, and data extraction. A well-known and widespread example of machine learning applications is facial recognition, which can be used to unlock smartphones [13].

Unsupervised learning

Unsupervised learning, in contrast to supervised learning, is a type of algorithm that learns patterns from unlabeled data. The machine is fed raw data and is forced to build a compact internal representation of its world. It is mainly developed to do clustering, anomaly detection, market segmentation, or dimensionality reduction. Financial organizations utilize the technique to spot fraudulent transactions [14]. It is also widely used for more efficient marketing and targeting campaigns by sorting customers' information.

Semi-supervised learning

Semi-supervised machine learning combines supervised and unsupervised learning. It uses a large amount of unlabeled data mixed with some labeled data. This strategy provides the benefits of both previous learning strategies while avoiding the challenges of finding a large amount of labeled data. It is used to perform speech recognition, web content classification, and text document classification.

Reinforcement learning

Reinforcement learning is a special and complex subfield of ML. The AI is trained with a system of reward and punishment. The developers create a virtual environment that gives rewards when the ML algorithm performs desired behaviors and punishment for negative behaviors. In this context, AI evolves to maximize the reward. It is a very promising field of ML used to train robots and computers to interact with the real world. A good example is the Google DeepMind Challenge Match where AlphaGo [15], an AI trained to play Go, won 4 of the 5 games against the top human Go player Lee Sedol in 2016. More recently, the robot company Boston Dynamics released a video of their robots' ability to evolve and dance with outstanding smoothness in a challenging real-life environment [16].

ORTHODONTIC TOOLS

The machine learning tools we developed belong to the supervised learning branch. We used large datasets pre-labeled by expert clinicians. We specifically focused on developing the following tools to perform important tasks which are segmentation and landmark identification on cone beam computed tomography (CBCT) and intraoral optical scan IOS data.

AMASSS - CBCT (Automatic multi-anatomical skull structure segmentation in CBCT)

The segmentation of medical and dental images is a fundamental step in automated clinical decision support systems. It supports the entire clinical workflow from diagnosis, therapy planning, intervention, and follow-up. With medical and dental images being acquired at multiple scales and/or with multiple imaging modalities, automated image analysis techniques are required to integrate patient data across scales of observation. Anatomic image segmentation, registration, and multimodality fusion using current open-source tools, such as ITK-SNAP and 3D Slicer are time-consuming and challenging for clinicians and researchers due to the low signal/noise ratio. Particularly in the large field of view CBCT images that are commonly used for orthodontics and oral maxillofacial surgery clinical applications. To perform a manual full-face segmentation, 7 hours of work on average is required by experienced clinicians, including approximately 1.5h for the mandible, 2h for the maxilla and 2h for the cranial base (CB), 1h for the cervical vertebra (CV), and 0.5h for the skin. The accurate and robust automatic anatomical segmentation of medical imaging data is a challenging problem due to the rich variety of anatomical structures and the difference in scan acquisition protocols from one center to another. Patients with facial bone defects may pose additional challenges for automatic segmentation. For this reason, in our study, we also included gold standard (ground truth) segmentations of CBCT images from patients with craniofacial large bone defects such as cleft lip and palate (CLP). Being able to accurately segment those deformities in the maxilla is challenging but necessary to generate 3D models for the diagnosis and treatment planning of patients with craniomaxillofacial anomalies (Figure 1). To answer this request, our lab developed a tool to automatically and accurately process a full-face segmentation in about 5 minutes using the state-of-the-art UNEt TRansformers (UNETR), a neural network of the Medical Open Network for Artificial Intelligence (MONAI) framework. We trained and tested our models using 618 de-identified CBCT volumetric images of the head acquired with several parameters from different centers for a generalized clinical application. Our results showed high accuracy and robustness. This tool can be used to segment the mandible, the maxilla, the cranial base, the cervical vertebra, and the airways in large CT as well as the crown, root, and mandible canal in a small field of view.

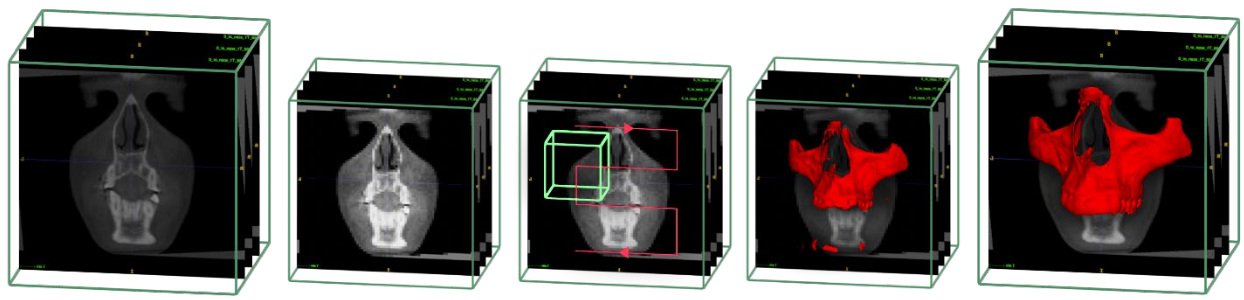

Figure 1. Visualization of the automatic maxilla segmentation steps. Re-sample and contrast adjustment of the input image, segmentation with the sliding window using UNETR, and finally, re-sampling of the cleaned-up segmentation to the input size.

ALI - CBCT (Automatic Landmark Identification in CBCT)

Robust and accurate solutions for anatomical landmark detection is a fundamental task to perform image-to-image registration, structure tracking, and simulations. We developed a new approach that reformulates landmark detection as a classification problem through a virtual agent placed in a 3D CBCT scan. This agent is trained to navigate in a multi-scale volumetric space to reach the estimated landmark position. The landmark detection task is set up as a behavior classification problem for an artificial agent that navigates through the voxel grid of the image at different spatial resolutions (the agent's environments). The detection starts at a low-resolution image with a global context and continues at the higher-resolution image, thereby capturing increased levels of detail. The image features are used as indicators to guide the landmark search. To adapt the feature extraction, we train different neural networks at each resolution. After the feature extraction, the network takes the image features as input and decides in which direction the agent should move to get closer to its target landmark position. The agent's movement decision relies on a combination of Densely Connected Convolutional Networks (DCCN) and fully connected layers. To predict the landmark positions in a CBCT, we rescale it to the resolutions used during training (here 1mm and 0.3mm voxel size). For each landmark to predict, an agent is generated with its corresponding network.

The landmark prediction is then made in 3 steps (Figure 2):

- **Step 1:** The prediction begins at the low-resolution level. The agent is placed in the middle of the scan to optimize the search time. Once the agent reaches a confident zone, it goes to the high-resolution layer.
- **Step 2:** The agent starts moving in the high-resolution from the confident zone. A preliminary estimation is set where the agent stops moving.
- **Step 3:** Now, a verification step is applied. This step consists of searching again in the high resolution scan starting from 6 positions (in each direction) in a small radius from the predicted point in **Step 2**. The result is an average of the 6 predicted positions.

The stopping criteria is active at prediction time and is implemented using a visitation map. If the agent tries to reach a previously visited voxel, it stops. The third step increased the prediction accuracy and compensated for a portion of the error caused by the discrete aspect of the space.

We evaluated our approach on 60 CBCT scans from teenagers to older patients. For each CBCT, 31 ground truth landmark positions were identified by clinicians. Our method achieved accuracy with an average of 1.54±0.87 mm error on the 31 landmark positions with rare failures. It takes an average of 4.2s computation time for the algorithm to identify each landmark on one large 3D CBCT scan. After a successful proof of concept, we trained new agents to reach a total of 119 landmarks that could be automatically identified using a 3D slicer module or the DSCI.

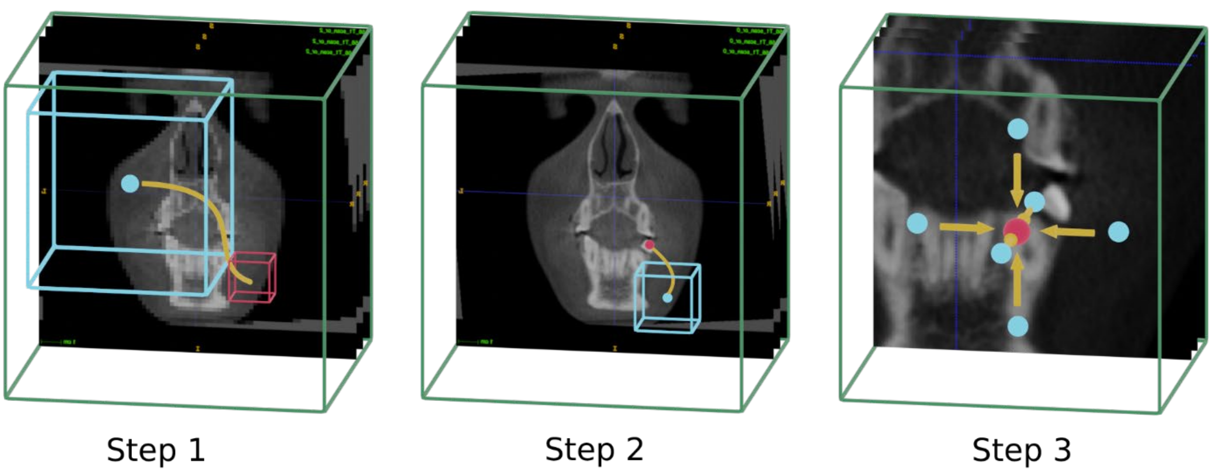

Figure 2. Visualization of the agent (blue) in the multi-scale environment (green) searching for the target (red).

FiboSeg - IOS (Fully automated segmentation of upper and lower jaws from 3D IOS.)

Developments in dentistry led to an improved application of 3D technologies such as IOS scanners which are used to design ceramic crowns, veneers, inlays, and occlusal guards, as well as to assist with implants. IOS is being used more often for automated diagnoses such as caries detection, analysis of risk factors of tooth movement, and treatment planning. These 3D surface models require shape analysis techniques for analyzing and understanding the geometry and they achieve state-of-the-art performance for tasks such as segmentation, classification, and retrieval. In this paper, we present a novel method for 3D surface segmentation based on a multi-view approach. Fast and accurate segmentation of the IOS remains a challenge due to various geometric shapes of teeth, complex tooth arrangements, different dental model qualities, and varying degrees of crowding problems. Our target application is the multi-class segmentation following the Universal Numbering System proposed by the American Dental Association (ADA), the dental notation system used in the US.

The multi-view approach consists of generating 2D images of the 3D surface from different viewpoints. The generated images serve as a training set for a neural network. We used Pytorch3D (open-source, https://pytorch.org/) to generate images of the surface on the fly during training and a one-to-one mapping that relates faces in the 3D model and pixels in the generated images (Figure 3). This is useful in inference time when we have to put the resulting labels from the images back into the 3D model.

This new method for automatic multi-class segmentation of 3D surfaces has proven to be accurate and effective, as well as easy to integrate into existing processing pipelines; there is no sub-sampling of the surface as it is required by the competing approaches such as MeshSegNet (open-source code available at https://github.com/Tai-Hsien/MeshSegNet) and PointConv (open-source code available at https://github.com/DylanWusee/pointconv). A great advantage of this method is the ability to predict the universal IDs of the crowns in the upper and lower jaws. The results reported by the competing approaches focused on upper models only or use a classification model to identify upper/lower jaws. Our approach is fully automated and labels both upper and lower crowns with a single model. The model learns to identify features specific to each jaw. We are aware that this model can still be improved and we can safely expect better results for segmentation of wisdom teeth as our sample size increases.

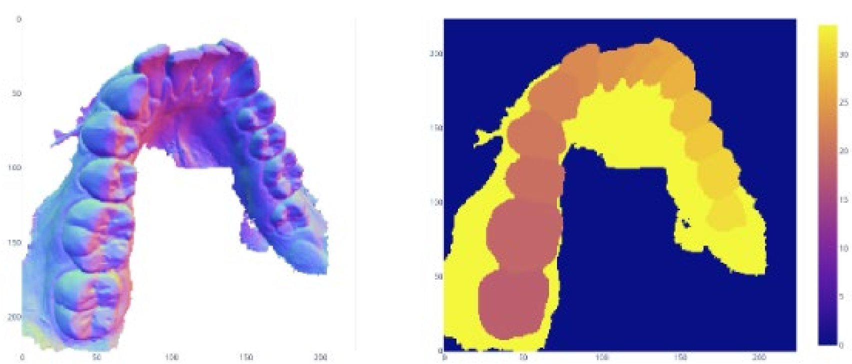

Figure 3. Rendered 2D view of the IOS. Left: surface normal encoded in RGB components, (additional surface properties may be rendered or extracted from the surface using the face-ID maps). Right: segmentation of the dental crowns rendered with a color map.

ALI - IOS (Automatic Landmark Identification in IOS)

Proper placement of the dental crowns is crucial to treatment planning, tooth movement, fabrication of dental restorations, monitoring and maintaining periodontal health, and attaining stable treatment outcomes and occlusal function. IOS are 3D surface models of the upper and lower dentition used to accurately evaluate the clinical crown position without exposing the patient to radiation. Given that, intraoral scanning and digitization of tooth geometries is a fundamental step in dental digital workflow, the accuracy of measurements in IOS is a must. Experts need to segment each tooth and annotate the corresponding anatomical landmarks to plan and assess crowns' position and/or movement for restorative, orthodontics movements, and/or implant dentistry. Manually performing these tasks is time-consuming and prone to inconsistency. There is a clinical need to develop fully automatic methods instead of manual operation. Facing this challenge, we developed a new algorithm for Automatic Landmark Identification on IntraOral Scans (ALLIOS), which combines image processing, image segmentation, and machine learning approaches to automatically and accurately identify commonly used landmarks on IntraOral Scans. More than 150 digital dental models were pre-processed by clinicians to manually annotate the landmarks on each dental crown.

The AI was then trained using the pre-labeled data. A virtual camera moved around each crown of the dental model taking pictures at various locations. The network (U-Net) was then trained to estimate the landmark position in each image. The final predicted landmark position was an average of the location on each picture (Figure 4). Our results showed an average distance error of 0.29 ± 0.23 mm between the prediction and the clinician's landmarks.

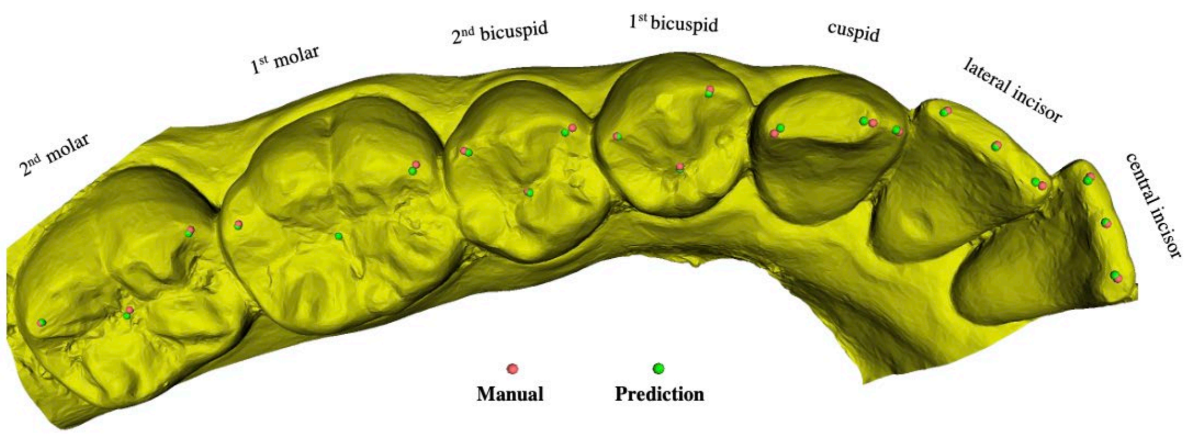

Figure 4. Comparison between manual landmarks and predicted landmarks on different teeth. The red spheres represent the clinician's landmarks (manual) and the green spheres were predicted by ALLIOS.

CONCLUSIONS

Machine learning tools such as ALI-CBCT and FiboSeg allow fast segmentation with high accuracy and robustness on CBCT images and IOS with heterogeneous acquisition parameters. We imagine that in future projects, other structures could be added to the segmentation and a larger dataset will allow better generalization of the application for our tool to work with other types of scans such as magnetic resonance imaging.

The landmarks detected by ALI-CBCT and ALI-IOS will be used to initialize two registration methods, voxel or surface-based registration, and registration of IOS and CBCT scans. They may also increase the efficiency and accuracy of quantitative assessment in clinical practice without the need for human interactions/annotations.

Given the high robustness and time performance, the developed tools are being implemented in an open-source web-based clinical decision support system, the Data Storage for Computation and Integration (DSCI) [17], and in user-friendly 3D Slicer modules (Figure 5). The computer-aided diagnostic tools will aid in therapy planning, and provide a step towards the implementation of dentistry decision support systems, as machine learning techniques are becoming important to automatically analyze dental images.

Machine learning tools are still in their early implementation in orthodontics, and current research on ML raises important questions regarding interpretability and dataset sample reliability. Therefore, better collaboration between orthodontic experts and ML engineers is urged to achieve a positive symbiosis between AI and clinician experts.

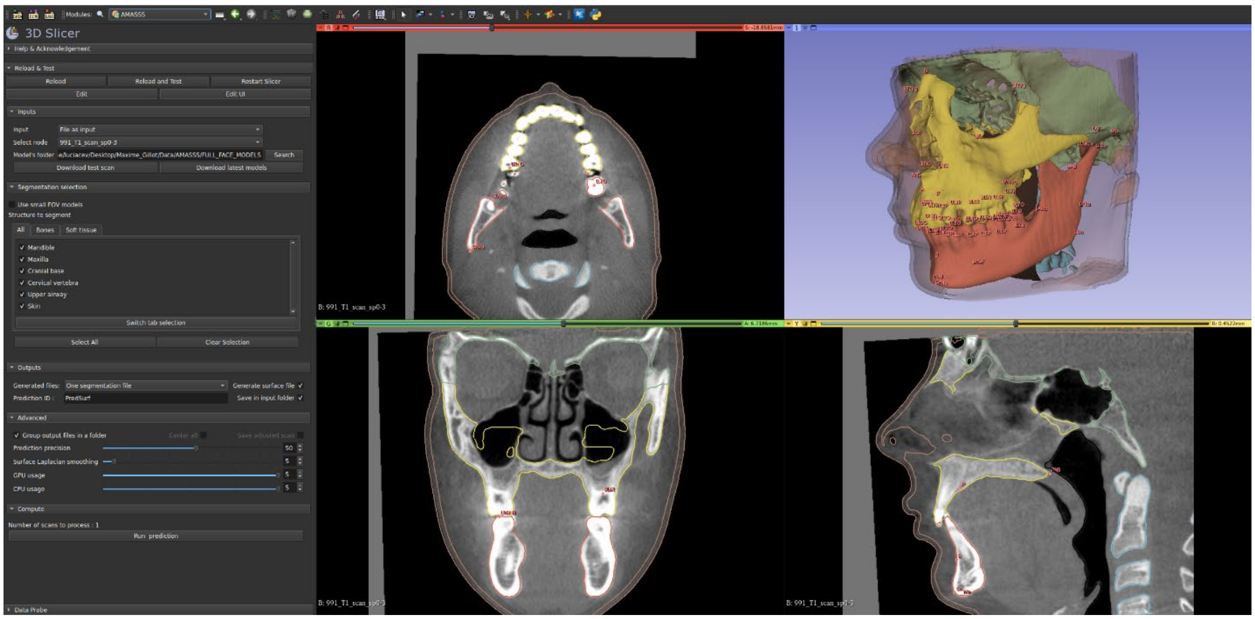

Figure 5. 3D Slicer AMASSS module on the left panel, with a visualization of ALI-CBCT/AMASSS result on the right.

ACKNOWLEDGEMENTS

Supported by NIDCR R01 024450, AA0F Dewel Memorial Biomedical Research award, by Research Enhancement Award Activity 141 from the University of the Pacific, Arthur A. Dugoni School of Dentistry and the American Association of Orthodontists Foundation (AAOF).

REFERENCES

1. Chaudhry, S., Dhawan, S. (2020). AI-Based Recommendation System for Social Networking. In: Pant, M., Sharma, T., Verma, O., Singla, R., Sikander, A. (eds) Soft Computing: Theories and Applications. Advances in Intelligent Systems and Computing, vol 1053. Springer, Singapore. https://doi.org/10.1007/978-981-15-0751-9_58
2. Hoyer, WD, Kroschke, M, Schmitt, B, Kraume, K, & Shankar, V. Transforming the customer experience through new technologies. Journal of Interactive Marketing 2020, 51(1), 57–71.
3. Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: a review. Sensors (Basel) 2018; 18(8):2674.

4. Bachute MR, and Subhedar JM. Autonomous driving architectures: insights of machine learning and deep learning algorithms. Machine Learning With Applications 2021; 6100164.
5. Jumper J et al. AlphaFold 2. London, England and Seoul, South Korea: Fourteenth Critical Assessment of Techniques for Protein Structure Prediction; 2020.
6. Heo L, Feig M. Modeling of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) proteins by machine learning and physics-based refinement. bioRxiv [Preprint] 2020; Mar 28:2020.03.25.008904.
7. Yushkevich PA, Yang G, Gerig G: ITK-SNAP: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. Annu Int Conf IEEE Eng Med Biol Soc 2016, 2016:3342-3345.
8. Pieper S, Halle M, Kikinis R. 3D Slicer. Arlington, Virginia: 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821); 2004; 1:632-635.
9. Yegnanarayana B. Artificial neural networks. India: Prentice-Hall of India Pvt. Ltd; 2009.
10. Dale R. GPT-3: What's it good for? Natural Language Engineering 2021; 27,113-118.
11. Zhang J. Basic neural units of the brain: neurons, synapses and action potential. arXiv [Preprint] 2019;1906.01703.
12. Ayodele TO. Types of machine learning algorithms. In: New Advances in Machine Learning. London, England: IntechOpen; 2010.
13. Song I, Kim H, Barom Jeon P. Deep learning for real-time robust facial expression recognition on a smartphone. Las Vegas, Nevada: 2014 IEEE International Conference on Consumer Electronics (ICCE). IEEE 2014, 564-567.
14. Monamo P, Marivate V, Twala B, "Unsupervised learning for robust Bitcoin fraud detection," 2016 Information Security for South Africa (ISSA), Johannesburg, South Africa, 2016, 129-134.
15. Chen, J. (2016). The Evolution of Computing: AlphaGo. Computing in Science & Engineering. 18(4): 4-7.
16. Boston Dynamics, accessed Feb 03 2023, https://www.youtube.com/watch?v=fn3KWM1kuAw.
17. Brosset S, Dumont M, Cevidanes L, Soroushmehr R, Bianchi J, Gurgel M, Deleat-Besson R, Le C, Ruellas A, Yatabe M et al. Web infrastructure for data management, storage and computation. Proc SPIE Int Soc Opt Eng 2021 Feb;11600:116001N.