Sameh Talaat, Ahmed Kaboudan, Wael Talaat, Budi Kusnoto, Flavio Sanchez, Mohammed H. Elnagar, Ahmed Ghoneima, Christoph Bourauel

Improving the accuracy of publicly available search engines in recognizing and classifying dental visual assets using convolutional neural networks

Abstract

Aim: To assess the accuracy of DigiBrain4, Inc. (DB4) Dental Classifier and DB4 Smart Search Engine* in recognizing, categorizing, and classifying dental visual assets as compared with Google Search Engine, one of the largest publicly available search engines and the largest data repository.

Materials and methods: Dental visual assets were collected and labeled according to type, category, class, and modifiers. These dental visual assets contained radiographs and clinical images of patients' teeth and occlusion from different angles of view. A modified SqueezeNet architecture was implemented using the TensorFlow r1.10 framework. The model was trained using two NVIDIA Volta graphics processing units (GPUs). A program was built to search Google Images, using Chrome driver (Google web driver) and submit the returned images to DB4 Dental Classifier and DB4 Smart Search Engine. The categorical accuracy of the DB4 Dental Classifier and DB4 Smart Search Engine in recognizing, categorizing, and classifying dental visual assets was then compared with that of Google Search Engine.

Results: The categorical accuracy achieved using the DB4 Smart Search Engine for searching dental visual assets was 0.93, whereas that achieved using Google Search Engine was 0.32.

Conclusion: The current DB4 Dental Classifier and DB4 Smart Search Engine application and add-on have proved to be accurate in recognizing, categorizing, and classifying dental visual assets. The search engine was able to label images and reject non-relevant results.

Keywords: *dental visual assets, artificial intelligence, dental radiographs, dental clinical images, dental classifier, smart search engine, machine learning, deep learning, convolutional neural network*

Introduction

Machine learning (ML) is an exciting branch of artificial intelligence (AI). Deep learning (DL) is a subfield of ML that powers the most human-like AI. DL models have the capability to learn and master tasks such as the extraction of meaningful patterns, which is a characteristic of human intelligence. DL applications in dentistry are geared toward developing programs capable of detection of pathologies, diagnosis of diseases, automatic identification of cephalometric landmarks, and segmentation of teeth and other structures.

Different programming approaches have been explored, ranging from traditional computer vision algorithms to recent ML techniques, and acceptable results have been achieved[1-4]. However, no known attempts have been made to build a real dental cognitive system capable of recognizing and interpreting the content of dental radiographs, clinical images, and other visual assets. The need for such a system is crucial, as using most of the currently available generic search engines to search for specialized data such as dental visual assets might yield inaccurate results, leading to a waste of time and effort. An ML approach that provides a model designed to learn and improve by experience through exposure to large datasets would be beneficial to overcome this problem[5]. Therefore, the present authors developed a model – the DB4 Dental Classifier and DB4 Smart Search Engine filter – to achieve the objective of accurately recognizing, interpreting, and classifying dental visual assets. The scope of the model includes grouping dental visual assets such as radiographs, clinical images, images of dental equipment and instruments, and others, into types, and categorizing each type relative to its nature. The model also classifies items pertaining to each type and category. If a viewing standard for a class exists, the model autocorrects the orientation of relevant detected items according to standards. The model also recognizes modifiers such as upper, lower, right, and left, and autogenerates accurate metadata relative to each asset. The model integrates with web search engines and can be used as a standalone in-house search engine. It is expected to signifi-

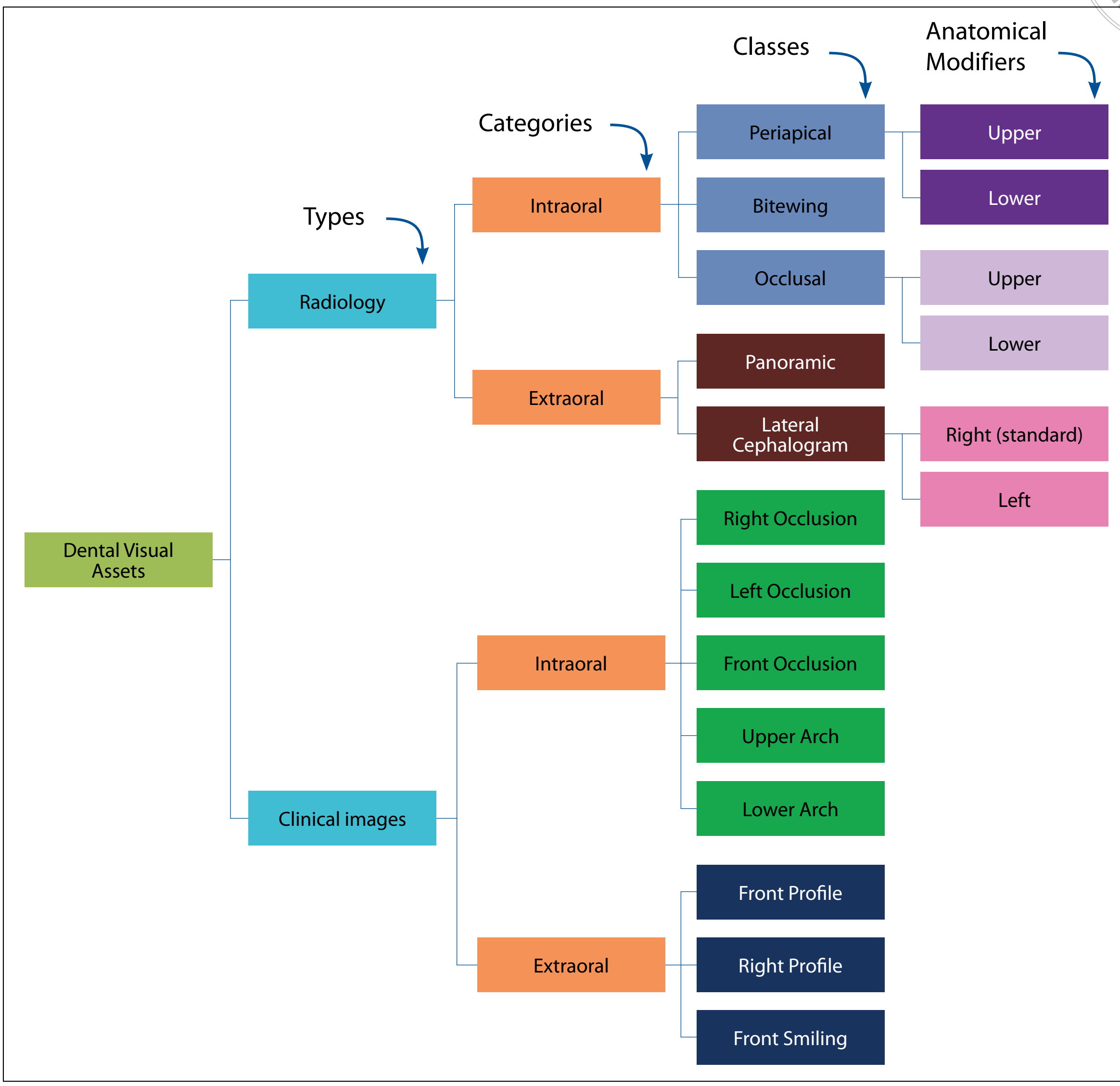


Fig 1 Labelling of deidentified assets according to type, category, class, and anatomical modifiers.

cantly enhance clinical practice in the field of dentistry and in the medical field in general by saving the time and effort required for image interpretation and diagnostic decision making. This will be achieved because the model will yield the most relevant results upon searching for dental visual assets through boosting the accuracy of search engines. In addition, the categorization of big data using a smart search engine will enrich medical research and allow better control over big data in medical and research institutions.

The present study aimed to assess the accuracy of DB4 Dental Classifier and DB4 Smart Search Engine in recognizing, categorizing, and classifying dental visual assets as compared with Google Search Engine.

Materials and methods

Datasets

The present research met the requirements of the Western Institutional Review Board (WIRB; WA, USA) for a waiver of consent under 45 CFR 46 116 (f; IRB approval 20193360). Dental visual assets were collected from the Foundation for Modern Bioprogressive Orthodontics, 650 West Colfax Ave, Denver, Colorado 80204, USA, and labeled according to standard terminology. The deidentified assets were labeled according to type, category, class, and anatomical modifiers (Fig 1). Labeling was done manually by a licensed orthodontist with 17 years of experience. These dental visual assets contained radiographs (periapical, bitewing, occlusal, panoramic, lateral cephalograms) and clinical images of patients' teeth and occlusion from different angles of view. The final dataset contained 3000 dental images for each class, except for occlusal radiographs that had a sample size of only 300. No identifying information such as name, age, gender, or time of exposure was used for the dataset.

Generalization of datasets

To this repository, 3000 miscellaneous image samples were added from Common Objects in Context (COCO)[6], which is a large-scale object detection, segmentation, and captioning dataset. This addition to the repository ensured that the model engine would be capable of differentiating generic images from our highly specialized dataset.

Partitioning of datasets

The fully generalized dataset was shuffled randomly and then divided into three datasets:

1. Training dataset (65%): This dataset is used for learning and fitting the network parameters (weights).
2. Validation dataset (15%): This dataset is used to tune the architecture and validate the training.
3. Test dataset (20%): This dataset is independent of the training dataset but follows the same probability distribution. It is used to assess the performance of the fully trained model.

Convolutional Neural Networks (CNNs) model

Authors: please add the city.

The model was based on a modified implementation of SqueezeNet[7], implemented using the TensorFlow r1.10 framework (Google, city, CA, USA), which offered the necessary depth with considerably less computational load, but with the highest accuracy, as required by our model search engine. A compression repeating-block design was implemented to reduce the total number of trainable parameters from approximately 60 million to 729,165, resulting in the optimization of the computational and time requirements for training. Three color channels were used to handle both color and grayscale images equally. Following the preliminary convolution and pooling layers, the later layers were mostly composed of a repeating structure that started with an X number of channels of 1 × 1 convolution filters, where the variable X ranged from 16 to 128, followed by two parallel convolutions, the first was 4X channels and 1 × 1 filters, and the second was 4X channels and 3 × 3 filters. The results of both parallel convolutions were then concatenated in an 8X channel structure. The structure composition depth started at 16, 64, 128, and ended at 64, 256, 512 depth channels. Maximum pooling function was implemented all over the network, except for the penultimate layer, where average pooling was implemented. Original images had a width and height within the range of 1000 to 2000 pixels. All images were converted to three-channel JPG formats.

The neural network was trained for 300 epochs. The Root Mean Square Propagation Optimizer was utilized in order to adapt the learning rate for every parameter, where the start learning rate was 0.0001. Weight was initialized using the Glorot uniform initializer, which extracted samples from a uniform distribution within [-limit, limit], where limit is the square root of (6/M + L), where M is the number of input units in the weight tensor and L is the number of output units in the weight tensor. The model was trained using two NVIDIA Volta graphics processing units (GPUs; NVIDIA, CA, USA).

Categorical accuracy was measured to assess the mean accuracy rate across all predictions and was calculated as the proportion of correctly classified classes achieved. In addition, categorical cross entropy (log loss function) and top-k categorical accuracy were used to measure the performance of the model at the end of the training process. Categorical cross-entropy was used to measure the total entropy between the probability distributions, and was calculated as the sum of separate loss for each class label per observation, whereas top-k categorical accuracy indicated success when the target class was within the top-k predictions provided, where k was set at 5.

Google Search Engine web interface

A program was built to search Google Images, using Chrome driver (Google web driver), then fetch and submit the returned images to DB4 Dental Classifier and DB4 Smart Search Engine add-on in order to filter and display only the dental image classes and modifiers that were searched for. The DB4 Dental Classifier and DB4 Smart Search Engine were then tested on the completely foreign dataset that was randomly returned from the Google search. The categorical accuracy of the DB4 Dental Classifier and DB4 Smart Search Engine was assessed in terms of recognizing, categorizing, and classifying these foreign datasets randomly returned from the Google search. The categorical accuracy measured the mean accuracy rate across all predictions and was calculated as the proportion of correctly classified classes achieved.

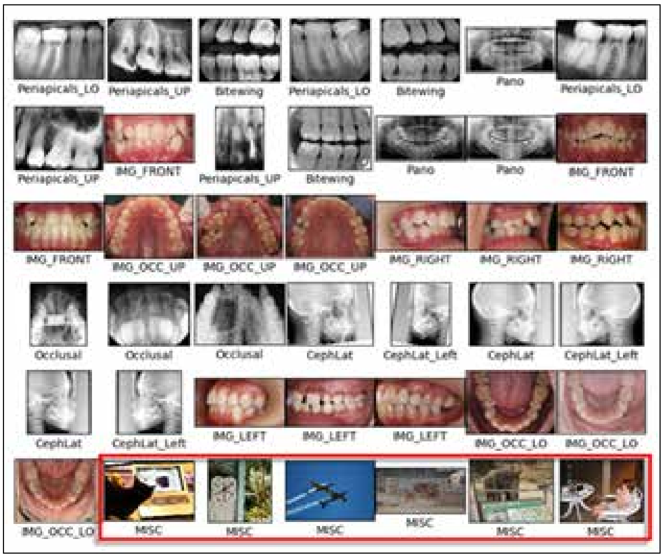


Fig 2 Results achieved from the DB4 Smart Search Engine when searching dental visual assets. The search engine displays the recognized class under each image. The red rectangle shows the wrong image type, classified as 'Misc.'

Results

The scores obtained at the end of the training process were: loss 0.0120; categorical accuracy 0.9961; top-k categorical accuracy 1.0000; validation loss 0.1282; validation categorical accuracy 0.9749; and validation top-k categorical accuracy 0.9985.

The mean categorical accuracy achieved using the DB4 Smart Search Engine for searching dental visual assets (clinical images and radiographs) on a foreign dataset was 0.93 (Fig 2), whereas the mean categorical accuracy achieved using Google Search Engine for searching the same assets was 0.3 (Fig 3). Categorical accuracy was calculated as the number of correct results achieved divided by the number of all results. An orthodontist interpreted the results on two separate occasions with a 1-week interval between them. This orthodontist had 35 years of experience and was a different orthodontist from the one who did the labeling. Using double quotes to enforce Google Search Engine to match the full search phrase resulted in a zero to 20% match (0 to 0.2 categorical accuracy). The results of categorical accuracy, categorical cross entropy, and top-k categorical accuracy showed the optimal performance of DB4 Smart Search Engine in recognizing and classifying dental visual assets and illustrated the positive effect of training on the model (Fig 4). The model was able to eliminate inaccurate search results that did not pertain to the search phrase (Fig 5).

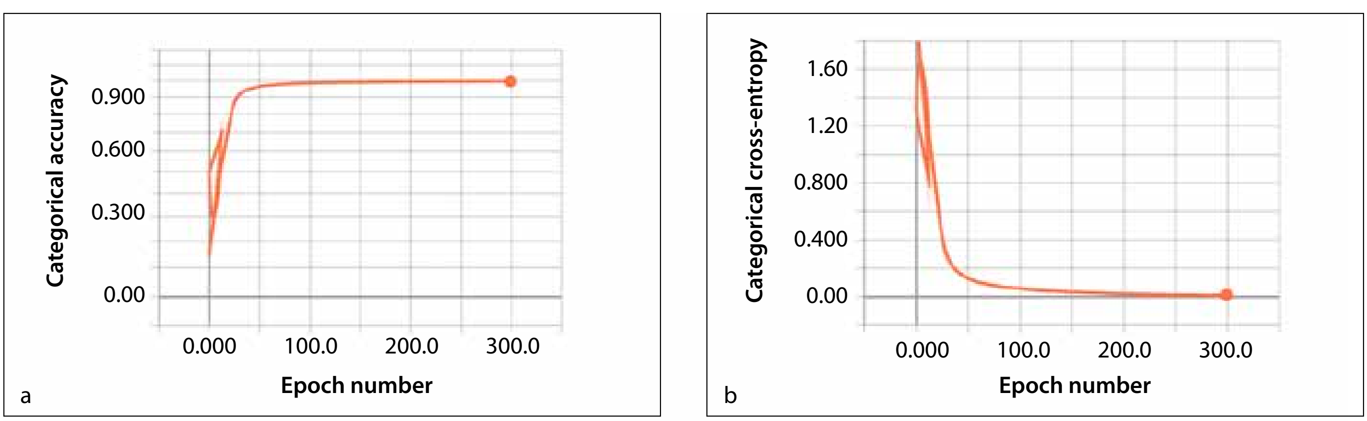
Discussion

The use of DL in the medical field is currently directed toward improving work efficiency, saving time, and achieving more accurate results using computer-aided technologies. In the last two decades, computer-aided technologies have developed significantly and are currently being used to assist physicians and radiologists in diagnosing several diseases in different medical fields[9]. DL is a subdomain of AI, which allows a computer to be trained to recognize patterns using large datasets. In the present study, the authors trained a CNNs model to recognize, categorize, and classify dental visual assets.

It has been suggested that configuring an accurate model requires training using heterogeneous data collected from different centers and at different time points. The dataset should also contain miscellaneous data from outside the area of focus to allow for generalizability[10]. To achieve acceptable performance, the authors of the present study designed a network with a depth of more than 40 layers. The implementation of CNNs of such depth might result in approximately 60 million trainable parameters, leading to an increase in the computational and time requirements for training. To overcome this risk, the authors implemented a compression repeating block inspired by the SqueezeNet model[7], which resulted in a reduction of the total number of trainable parameters to 729,165.

CNNs have been used in different applications in medicine such as in the detection of colitis on computed tomography[11], and the detection of brain lesions on magnetic resonance images[12]. However, the applications in dentistry are still limited. The CNNs model used in the present study has the advantage of improved accuracy and fewer computational requirements compared with other CNNs. It proved to be highly accurate and is expected to solve the shortcomings related to searching specialized data using currently available search engines. This is particularly clear with generic search engines such as Google, which is reliable when it comes to generic data but fails systematically with specialized data. Using double quotations to search the exact phrase in Google usually results in zero to 20% matches, whereas unquoted free word search – regardless of the search terms – results in a mixture of the required assets, but mostly unrequired and completely unrelated assets. In addition, the results fail to filter anatomical modifiers, like distinguishing upper from lower, or right from left, or any combination of these terms, and fail to filter the results by type, category or class. This can be explained by the fact that the Google Search Engine is not

Fig 3 Screen capture showing the results achieved using Google Search Engine to search for upper periapical radiograph. The red rectangles show the wrong image type.



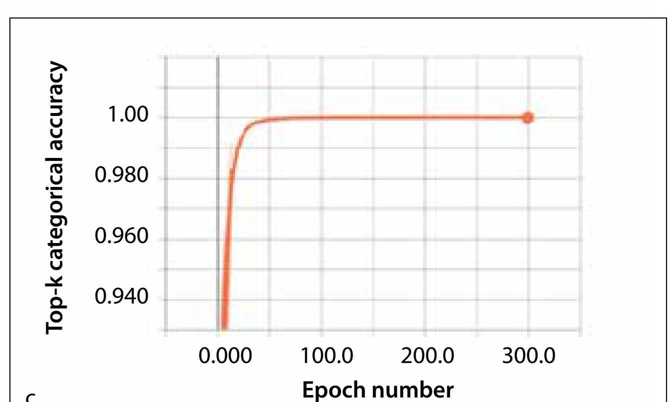


Fig 4 The results of categorical accuracy (a), categorical cross entropy (b), and top-k categorical accuracy (c) carried out on the test dataset.

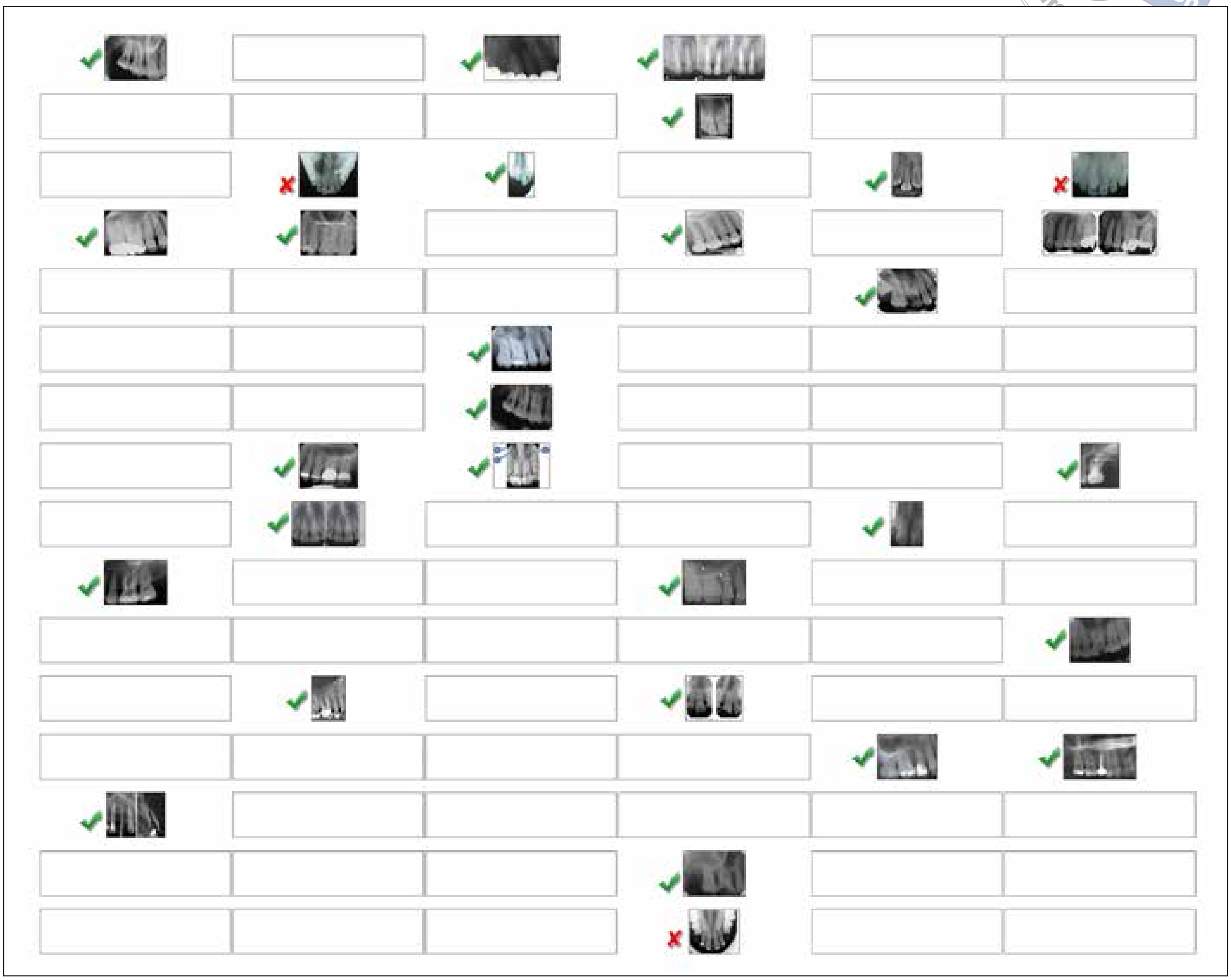


Fig 5 The precision of the DB4 Smart Search Engine compared with Google Search Engine, using the same search phrase "upper periapical radiographs". Empty boxes show that the images were filtered out (removed) by the search engine as they do not pertain to the given search phrase. ✓ indicates true positive results; ✗ indicates false positive results.

trained to recognize and differentiate between the fine-grained classes of dental visual assets.

The results of this study confirm the progressive improvement of the present model's performance as a result of training, which is a step on the way to developing a full dental cognitive system that can reliably render classified dental visual assets, as needed. Numerous benefits of such a system are foreseen for different applications such as: automatic selection of the correct images to be uploaded to patient databases; searching for specific pathologies in large medical images archiving and communication systems (PACSs); the possibility of automated evaluation of patient data with insurance companies to facilitate decision making; testing the accuracy of the labeled images that have been stored in different databases; and sorting big data accumulated at universities and insurance companies with better granularity.

The accuracy of the presented model would have increased from 0.93 to 0.96 if occlusal images were excluded because the training dataset had a limited number of occlusal radiographs. Accordingly, the main limitation of the current model was the recognition of occlusal radiographs due to the inadequacy of the sample size of the images in the training phase.

The first testing phase in the development of the present model was done on the dataset described above (see Methods section above). The second testing phase was done on completely foreign datasets, randomly returned from Google Search Engine. These datasets were completely new for our model and were generated from different statistical spaces. No new training was involved in the second testing phase, and the accuracy was measured as the probability of true positive hits.

Further improvement to the current model output can be achieved by using bigger and more diverse training datasets through progressively introducing more types, classes, and categories such as dental instruments and dental equipment. CNNs models can be gradually enhanced and tuned, further training can be done, and the techniques used can be improved[13,14]. The results obtained in previous training models can be used for the fine-tuning of these models and the training of new models. DB4 Smart Search Engine is an efficient tool for big data mining, as it allows the investigators to use publically available big data and enables data mining for dental images in unlabeled big datasets. Furthermore, it can help in data preparation for developing more sophisticated dental image analysis algorithms.

Further investigations and improvements are expected to translate the current research attempts for use in clinical practice in the near future. The refinement of the current models will allow the performance of the system to either match or exceed the experienced practitioners' levels[15]. The present authors believe that this will boost the performance of practitioners in clinical practice, but that it will never replace them.

Conclusion

The current DB4 Dental Classifier and DB4 Smart Search Engine have proved to be accurate. The models fulfilled all the required functionalities to recognize, categorize, and classify dental visual assets as well as label them. The DB4 Smart Search Engine was able to recognize and reject non-relevant images. It can work as a filter booster add-on for generic web search engines. It can also act as a native application for searching organizations' locally stored imaging data, and autogenerating precise dental metadata associated with those images. It is also an efficient tool for big data mining of dental visual assets. The present authors intend to continue developing it so that it recognizes more dental assets such as equipment, devices, instruments, and other objects usually encountered in the dental field.

Acknowledgment

Dr Sameh Talaat's position at the University of Bonn was supported by Quintessenz Verlag, Germany.

References

1. Wang S, Li H, Li J, Zhang Y, Zou B. Automatic analysis of lateral cephalograms based on multiresolution decision tree regression voting. J Healthc Eng 2018;2018:1797502.
2. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2:230–243.
3. Fazal MI, Patel ME, Tye J, Gupta Y. The past, present and future role of artificial intelligence in imaging. Eur J Radiol 2018;105:246–250.
4. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.
5. Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. AJR Am J Roentgenol 2017;208:754–760.
6. Lin TY, Maire M, Belongie S, et al. Microsoft COCO: common objects in context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds). ECCV 2014. LNCS. Cham: Springer, 2014:740–755.
7. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <10.5MB model size. arXiv preprint arXiv 20161602.07360:1–13.
8. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. Ann Surg 2018;268:70–76.
9. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imaging Graph 2007;31:198–211.
10. Son LH, Tuan TM, Fujita H, et al. Dental diagnosis from X-ray images: an expert system based on fuzzy computing. Biomed Signal Process Control 2018;39:64–73.
11. Liu J, Wang D, Lu L, et al. Detection and diagnosis of colitis on computed tomography using deep convolutional neural networks. Med Phys 2017;44:4630–4642.
12. Hsieh YZ, Luo YC, Pan C, Su MC, Chen CJ, Hsieh KL. Cerebral small vessel disease biomarkers detection on MRI-sensor-based image and deep learning. Sensors (Basel) 2019;19:2573.
13. Tuzoff DV, Tuzova LN, Bornstein MM, et al. Tooth detection and numbering in panoramic radiographs using convolutional neural networks. Dentomaxillofac Radiol 2019;48:20180051.
14. Kohli MD, Summers RM, Geis JR. Medical image data and datasets in the era of machine learning - Whitepaper from the 2016 C-MIMI meeting dataset session. J Digit Imaging 2017;30:392–399.
15. Ren S, He K, Girshick R, Sun J. Faster R–CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 2017;39:1137–1149.

Sameh Talaat

Sameh Talaat, BDS, MSc, PhD
Department of Orthodontics, Faculty of Dental Medicine, Future University in Egypt; Oral Technology, University of Bonn, Bonn, Germany

Ahmed Kaboudan, BDS, MSc Orthodontics, PhD Computer Science
CRO DigiBrain4 Inc, Chicago, Illinois, USA

Wael Talaat, BDS, MSc, PhD
Department of Oral and Craniofacial Health Sciences, College of Dental Medicine, University of Sharjah, UAE; Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, Suez Canal University, Egypt

Budi Kusnoto, DDS, MS
Department of Orthodontics, College of Dentistry, University of Illinois at Chicago, Chicago, Illinois, USA

Flavio Sanchez, BDS, MS
Department of Orthodontics, College of Dentistry, University of Illinois at Chicago, Chicago, Illinois, USA

Mohammed H. Elnagar, DDS, MSc, PhD
Department of Orthodontics, College of Dentistry, University of Illinois at Chicago, Chicago, Illinois, USA; Department of Orthodontics, Faculty of Dentistry, Tanta University, Egypt

Ahmed Ghoneima, BDS, MSc, MSD, PhD
Department of Orthodontics, Hamdan Bin Mohammed College of Dental Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, UAE

Christoph Bourauel, Prof Dr rer nat
Oral Technology, University of Bonn, Bonn, Germany

Address Wael M. Talaat, BDS, MSc, PhD, College of Dental Medicine , University Dental Hospital Sharjah, Room F125, P.O. Box: 27272, UAE; Tel: +971 65057605; Email: wtaha@sharjah.ac.ae

Authors: past positions can not be listed

Authors: Positions and roles are generally not listed in IJCD anymore