2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Combining CNN with Pathological Information for the Detection of Transmissive Lesions of Jawbones from CBCT Images

Zimo Huang[1], Tian Xia[1], Jinman Kim[1], Lefei Zhang[2] and Bo Li[3*]

Abstract— **Cone-Beam Computed Tomography (CBCT) imaging modality is used to acquire 3D volumetric image of the human body. CBCT plays a vital role in diagnosing dental diseases, especially cyst or tumour-like lesions. Current computer-aided detection and diagnostic systems have demonstrated diagnostic value in a range of diseases, however, the capability of such a deep learning method on transmissive lesions has not been investigated. In this study, we propose an automatic method for the detection of transmissive lesions of jawbones using CBCT images. We integrated a pre-trained DenseNet with pathological information to reduce the intra-class variation within a patient's images in the 3D volume (stack) that may affect the performance of the model. Our proposed method separates each CBCT stacks into seven intervals based on their disease manifestation. To evaluate the performance of our method, we created a new dataset containing 353 patients' CBCT data. A patient-wise image division strategy was employed to split the training and test sets. The overall lesion detection accuracy of 80.49% was achieved, outperforming the baseline DenseNet result of 77.18%. The result demonstrates the feasibility of our method for detecting transmissive lesions in CBCT images.**

Clinical relevance — **The proposed strategy aims at providing automatic detection of the transmissive lesions of jawbones with the use of CBCT images that can reduce the workload of clinical radiologists, improve their diagnostic efficiency, and meet the preliminary requirement for the diagnosis of this kind of disease when there is a lack of radiologists.**

I. Introduction

Cone-Beam Computed Tomography (CBCT) is an emerging medical imaging modality that uses divergent, cone forming X-ray CT. CBCT image is acquired in three-dimensional space and constituent 2D slices can be used for multiple plane reconstruction and display a range of morphological features. CBCT images are routinely used in dental applications, e.g., for disease assessment and facilitate drafting pre-treatment plans, as the imaging modality started to be adopted by dentists in recent years [1]. CBCT has several advantages over other imaging modality in dental practices, it is an easily available clinical imaging modality that can offer an efficient non-invasive scanning in combination with higher spatial resolution while reducing the radiation dose and scan costs comparing to X-ray and conventional CTs [2].

*Corresponding author
[1]School of Computer Science, The University of Sydney, ({zhua2125, txia6085}@uni.sydney.edu.au, jinman.kim@sydney.edu.au)
[2]Institute of Artificial Intelligence and School of Computer Science, Wuhan University, Wuhan, China, (zhanglefei@whu.edu.cn)
[3]School of Stomatology, Wuhan University, Wuhan, China, (libocn@whu.edu.cn)

The adoption of CBCT is improving dental disease diagnosis, such as diagnosing the transmissive lesions of jawbones, also known as the space-occupying lesions of jawbones. Its clinical manifestation involves progressive growth and size of the lesions that can lead to expansion of the jaws, facial deformity, and continuous absorption that keep thinning the bone plate and leading to pathological fractures. Such disease can also induce decreasing of the pulp vitality of the affected teeth and abnormal tooth germ development in the affected area, which not only seriously affects the health of the patients' jaw and teeth, but also brings great impact on the aesthetics, function and psychology of patients and affects the prognosis [3]. The lesions are usually located inside the jawbones of the patient, which presents challenges in the diagnosis due to the surrounding bones. The establishment of accurate diagnosing of such diseases is of vital importance for patients' treatment and recovery. CBCT has been selected as the 'gold standard' to diagnose transmissive lesions because it could provide high accuracy for differential diagnosis between cysts and granulomas with excellent inter-rater and good to excellent intra-rater reliabilities [4].

Recent advances in deep convolutional neural networks (CNNs) have achieved state-of-the-art performances on multiple computer vision applications, especially on image processing. CNNs learns and quantify sophisticated imaging representations by leveraging large volumes of image data [5]. In the study by Ker et al. [6], the researchers suggested that using multiple machine learning algorithms including CNN is applicable and important in key research areas of medical image analysis, especially on Computer-Aided Detection (CAD) tasks. Published studies have applied several well-known CNN architectures, such as GoogLeNet [7], UNet [8] and Overfeat [9] while achieving state-of-the-art performances in the detection tasks on medical data.

The employment of CNNs in medical image detection tasks of abnormalities is of great significance for the improvement of diagnostic accuracy and efficiency for the clinic. Despite CNNs have been employed for the diagnosis of dental diseases including cephalometric landmark detection [10] and caries detection [11], their role and diagnostic values have yet been thoroughly investigated in the emerging CBCT images.

In this paper, we propose an automatic method for detecting transmissive lesions of jawbones using CBCT images. Specifically, we integrated a pre-trained DenseNet with pathological information which was employed to reduce the intra-class variation within each sample that may affect the performance of the model. Our key contributions are:

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

- We quantified imaging representations of the transmissive lesions using a pre-trained DenseNet that was fine-tuned with a curated dataset. This is the first time for a CNN architecture to be applied to the transmissive lesions in the jaw shown on CBCT images.
- We propose a novel technique to integrate pathology information to the training by separating a single CBCT image into multiple intervals of slices. Such a method was determined to resolve the difficulty of applying CNNs on raw CBCT images. It could also reduce the computational cost while enhancing the overall performance.
- We built a CBCT image dataset of the transmissive lesions of jawbones – the largest dataset evaluation for this disease.
- We conducted experiments on the detection task and evaluated the performance of the model, the result suggested that the detecting system can serve as a computer-aided assistant for the dentists with the clinical diagnosing procedures.

II. Materials and Methods

A. Data Collection and Preprocessing

The DICOM-formatted CBCT images from a total of 353 patients were collected under ethical approval from the Stomatological Hospital of Wuhan University, China. Those patients were retrospectively selected with inclusion criteria as follows: 1) The pathological diagnosis is clear and the image data are complete, 2) Images with excessive pseudopacities (including motion pseudopacities and metal prosthetic pseudopacities), poor image quality or beyond the field of view (lesions are not completely displayed) are excluded, 3) The lesion should have a single pathological type (multiple lesions should also have the same pathological type), images containing multiple lesions with the coexistence of multiple pathological types were excluded. 4) All cases should be confirmed by histopathology examinations after the surgery. An anonymization process was taken to remove all personally identifiable information (PII) for each patient after the inclusion.

There were multiple subtypes of the transmissive lesions of jawbones, we manually selected four types of sub diseases to be allocated including the ameloblastoma, periapical cyst, dentigerous cyst and odontogenic keratocyst. The curated dataset is comprised of 89 ameloblastoma patients, 83 periapical cyst patients, 51 dentigerous cyst patients and 130 odontogenic keratocyst patients were collected and stored in the database. The detailed information of the dataset is shown in Table I.

TABLE I
The Generated CBCT Image Dataset

Sub-disease Category	Patients #	Class I Slices	Class II Slices
Ameloblastoma	89	31488	13073
Periapical Cyst	83	35173	5336
Dentigerous Cyst	51	19739	4339
Odontogenic Keratocyst	130	45293	19692

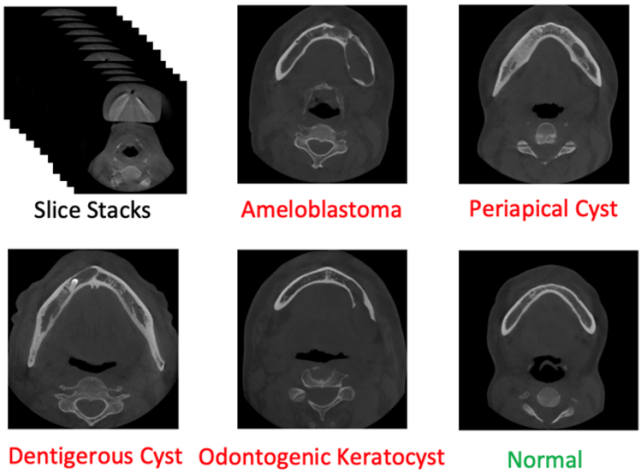

Fig. 1. Illustration of CBCT. The Slice Stacks showed the nature of CBCT that each CBCT data contains multiple 2D slices, forming a stack. The example slices with red names were all classified as disease slices during the labelling process, including 4 types of sub-diseases of the transmissive lesions of jawbones. The bottom right slice shows an example of normal slices that contain no lesion.

All the patients were patients with the disease and has one CBCT data which was composed of volumes of slices. The annotation process was manually taken under the guidance of radiologists to label each slice as one of the two classes: Class I: normal slices, and Class II: disease slices. An example of CBCT slices and the sub-diseases were illustrated in Fig 1.

B. CNN Architecture

DenseNet was selected for our experiment because of its generalisability for different tasks and its capability to comprehensively quantify imaging representations within the data. The unique network architecture of DenseNet helps to make it easier to capture all the necessary information as dense connections concatenate the output of the previous layer with the future layer [12]. The framework is designed to ensure the maximum information flow between layers in the network as all layers are connected directly with each other. To preserve the feed-forward nature, each layer obtains additional inputs from all preceding layers and passes on its feature maps to all subsequent layers. Hence, the l^{th} layer has l inputs, consisting of the feature-maps of all preceding convolutional blocks. Its feature maps were passed on to all the $L-1$ subsequent layers. This introduces $\frac{L(L+1)}{2}$ connections in an L-layer network, instead of just L in traditional architectures.

For each function H_l produces k feature-maps, it follows that the l^{th} layer has $k_0 + k \times (l-1)$ input feature-maps, where k_0 is the number of channels in the input layer. The hyperparameter k was referred to as the growth rate of the network, it provides the important advantage as DenseNet can have very narrow layers by adjusting k. DenseNet has demonstrated its performance in the medical field: CheXNet

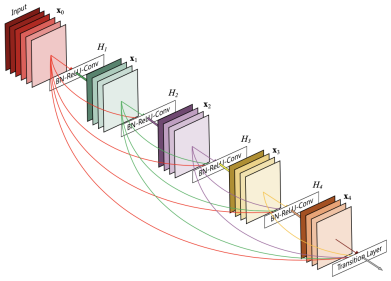

Fig. 2. Dense Block Architecture: Each layer takes all preceding feature-maps as input [12].

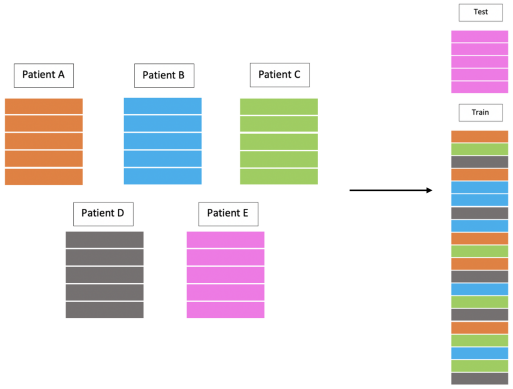

Fig. 3. Patient-wise image division method: An illustration of the division method we utilized during the splitting of training and testing sets. In this example, there are five patients, each uniquely colour-coded. One patient (Patient E) is assigned to the testing set. The four remaining patients are allocated to the training set. The validation set is split out using the same method and then the slices in the training set will be shuffled.

is a modified DenseNet that achieved an average AUROC of 84.14% on 14 pathologies in the ChestX-ray14 dataset, which had a margin of > 0.05 over the previous state of the art results [13].

III. Experiment

A. Dataset

Each CBCT image of the transmissive lesions within our dataset consisted of approximately 500 2D image slices. However, there were only around an average of 100 slices that were labelled as disease slices; the remaining slices before and after the lesion were all labelled as normal. As the CBCT scaned through the entire head section of the patient, the slices located in the bottom jaw area showing the teeth can have the same label as the slices located in the top of the head containing the brain, which increased the intra-class variation (the slices with the same class are having huge differences in their imaging features) within the normal class due to the different biological feature of brain and jaw.

We introduced a novel technique that employs the pathology information to eliminate the intra-class variation within each sample that may affect the performance of the model. As the largest amount of slices for a single patient was 521 within all collected samples, we divided the 500+ slices into multiple volume sections with a fixed number of slices and trained them separately. Hence, within each section, slices with the same labels share a bunch of much similar biological features. We determined that every 75 slices will be considered as a slice section because the disease images are concentrated in the slice range of 100 to 200, with the use of 75 slices, we can establish image sets with similar disease vs. normal class ratio. As a result, image volumes were partitioned into 7 sections. However, the distribution of disease-containing slices in section 6 and 7 was lower than 1%, hence these two sections were excluded. With reduced image slices, the training computation efficiency is increased because smaller input volume will allow the network to focus more on the subtle differences among images.

The training, validating and testing sets were split using the patient-wise image division method as shown in Fig 3. This method enables the model to become robust, consistent and more generalizable as it treats each patient rather than each slice as a training sample by splitting all the slices belonging to the same patient into one of the three image sets (training, validation or testing) respectively. A 5-fold cross-validation method was also adopted for the average accuracy calculation that is more reflective of the performance of the model on detecting the disease as well as showing the robustness with the use of different pairs of training and testing sets.

B. Implementation Details

We utilized the pre-trained torchvision Densenet-161 model, all the slices were resampled into a fixed size of 224×224. During the training process, the parameters on dense blocks 3 and 4 and the corresponding transition blocks were frozen. Another adjustment made to the model was the fully connected layer, it was modified to output the required predictions as there were two classes in total. The model was trained end-to-end with the Adam optimization method and the learning rate was set to 1×10^{-4} so as not to overshoot the optimal solution. Moreover, the learning rate schedule has been applied, whereby the learning rate is decayed by 5% for each training epoch. All the experiments were carried out using the NVIDIA GeForce RTX 2080 Ti GPU with 11 GB memories. The batch size was set to 32 for the training and validating processes, the testing process was set to have one sample per batch so as it will return the real prediction for that particular slice.

IV. Result and Discussion

We applied the same model architecture to all five slice sections, the accuracy and AUC score for each section have been obtained and shown in Table II. The overall percentage within the table was calculated as:

$$OP = \frac{Correctly\ Predicted\ Sample\ Number}{Total\ Number\ of\ Testing\ Samples} \quad (1)$$

where OP stands for overall percentage. Because of the different number of testing samples for each section, calculating the average of all 5 accuracies cannot reflect the true ability of the model.

As shown in Table II, the averaged accuracy across all 5 sections was 80.49%, suggesting that the application of CNN is capable of detecting transmissive lesions displayed on CBCT images. However, the worst corresponding AUC score was 0.5122, which was caused by numerous false positive predictions, as the samples that have the actual class as disease were misclassified as normal. We attributed this finding to the exaggerated proportion of disease slices (6%) within section 5 that made the model heavily skewed to normal samples.

The result was also higher than the baseline experiment that was trained without the image partition method (77.18%). The result with the proposed method was not only having a higher average accuracy but also having a smaller number of false-positive predictions than the baseline, demonstrating the importance of the proposed image partition based on the integration of pathological information.

We attribute the variance between the training accuracies across different sections to the differences in anatomical features. Images belonging to section 1 contain the hyoid bone and body of the mandible. The posterior area within the image is the ridged cone which is not in the dental scope. From the images in section two, mandibular alveolar bone, mandibular teeth, and posterior mandible can be observed. Mandibular ascending ramus and maxillary teeth are the main component of the images within section three. In section 4, maxilla and maxillary sinus can be seen as well as some nasal tissue and organs such as nasal septum and turbinate. The last section contains the images from the zygomatic to the zygomatic arch. With the difference in the presenting organs and tissue and their various complexity, the features to be learnt are different. Differences between image proportions were also considered as a factor that contributed to the changes in the training accuracies.

TABLE II
PERFORMANCE OF THE MODEL ON DIFFERENT SECTIONS

Section	Accuracy	AUC	Disease Slices %
1	0.9201	0.8100	18%
2	0.7366	0.7505	56%
3	0.6407	0.6437	50%
4	0.7586	0.6332	29%
5	0.9680	0.5122	6%
Overall	0.8049		

V. CONCLUSION

We propose an automatic method for detecting transmissive lesions of jawbones using CBCT images. A pre-trained DenseNet was integrated with pathological information where the pathology information was employed to eliminate the intra-class variation. Our results show that the CBCT image data-set, comprising of different types of transmissive lesions of the jawbone, can be effectively classified into disease and normal with an overall accuracy of 80.49%.

For future research, we will explore 3D CNN models to better learn the sequential information within CBCT data. Our future work will take into consideration of the growing CBCT datasets. We will further optimise our methods to be more generalisable for different datasets and to different disease types.

REFERENCES

[1] J. Zhang, M. Liu, L. Wang, S. Chen, P. Yuan, J. Li, S. G. Shen, Z. Tang, K. C. Chen, J. J. Xia and D. Shen, “Context-guided fully convolutional net-works for joint craniomaxillofacial bone segmentation and landmark digitization”. In: Medical Image Anal. vol. 60, 2020.

[2] Y. Huang, J. V. Dessel, M. Depypere, M. EzEldeen, A. A. Iliescu, E. D. Santos, I. Lambrichts, X. Liang and R. Jacobs, “Validating cone-beam computed tomography for per-implant bone morphometric analysis”. In: Bone Research. vol. 2, 2014.

[3] D. Baumhoer and S. Höller, “Cystic lesions of the jaws”. In: Der Pathologe. vol. 39, pp. 71–84, 2018.

[4] J. Guo, J. Simon, P. Sedghizadeh, O. Soliman, T. Chapman and R. Enciso, “Evaluation of the reliability and accuracy of using cone-beam computed tomography for diagnosing periapical cysts from granulomas”. In: Journal of Endodontics. vol. 39, pp. 1485–1490, 2013.

[5] R. Yamashita, M. Nishio, R. K. G. Do and K. Togashi, "Convolutional neural networks: an overview and application in radiology”. In: Insights Imaging. vol. 9, pp. 611–629, 2018.

[6] J. Ker, L. Wang, J. Rao and T. C. C. Lim, “Deep Learning Applications in Medical Image Analysis”. In: IEEE Access. vol. 6, pp. 9375–9389, 2018.

[7] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks”. In: Nature. vol. 542.7639, pp. 115–118, 2017.

[8] O. Ronneberger, P. Fischer and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: MICCAI. vol. 9351, pp. 234–241, 2015.

[9] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus and Y. LeCun, “OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks”. In: ICLR.2014.

[10] M. Zeng, Z. Yan, S. Liu, Y. Zhou and L. Qiu, “Cascaded convolutional networks for automatic cephalometric landmark detection”. In: Medical Image Anal. vol. 68, 2021.

[11] F. Casalegno, T. Newton, R. Daher, M. Abdelaziz, A. Lodi-Rizzini, F. Schürmann, I. Krejci and H. Markram, “Caries Detection with Near-Infrared Transillumination Using Deep Learning”. In: Journal of Dental Research. vol. 98, pp. 1227–1233, 2019.

[12] G. Huang, Z. Liu, L. Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks". In: CVPR, pp. 2261-2269, 2017

[13] P. Rajpurkar, J. Irvin, K. Zhu, B.Yang, H. Mehta, T. Duan, D. Y. Ding, A. Bagul, C. Langlotz, K. S. Shpanskaya, M. P. Lungren and A.Y. Ng, “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning”. In: CoRRabs. vol. 1711.05225, 2017.