Image Processing Techniques Applied to Oral Medicine

Maira B. H. Moran[1], **Aura Conci**[1]

[1]Instituto de Computação – Universidade Federal Fluminense (UFF)
24210-310 – Niterói – RJ – Brazil

{mhernandez,aconci}@ic.uff.br

Abstract. *This thesis' main objective is to investigate the applicability of image processing techniques in the oral imaging field, through a selection of oral issues use cases that can be mapped into different computational tasks. In order to investigate the validity of this approach, a series of methods are proposed employing different IP techniques. This work presents a compilation of these different solutions as an unique toolset to be employed in clinical scenarios, assisting oral medicine experts in several tasks with decision-supporting algorithms for diagnosis of oral issues, instead of only providing visualization interfaces for image exams as current commercial solutions.*

Resumo. *O objetivo principal desta tese é investigar a aplicabilidade de técnicas de processamento de imagem no campo de imagens orais, por meio de uma seleção de casos de uso de problemas orais que podem ser mapeados em diferentes tarefas computacionais. Para investigar a validade desta abordagem, um conjunto de novos métodos é proposto empregando diferentes técnicas de IP. Este trabalho apresenta um conjunto de ferramentas para serem empregadas em cenários clínicos, auxiliando especialistas em medicina oral em diversas tarefas, através de algoritmos de suporte à decisão para o diagnóstico de problemas orais, ao invés de de apenas fornecer interfaces de visualização para exames de imagem como as soluções comerciais atuais.*

1. Problem characterization

Digital image examinations are currently a prevalent technology in medical and dental practice being employed in the diagnosis of several oral diseases. Medical imaging in general presents an excellent potential for image processing (IP) solutions since examination images is an essential part of a wide range of medical areas [Schwendicke et al. 2019]. The integration of digital imaging with computer-assisted analysis has revolutionized the oral health field. Image evaluation constitutes an important step in the diagnosis of several oral issues, including oral diseases being applied in cariology, endodontics, periodontology, orthodontics, and forensic dentistry [Schwendicke et al. 2019]. Several IP solutions for oral imaging have emerged in the last few years, aiming to automate the identification and evaluation of oral diseases, potentially reducing possible errors resulting from experts' subjectivity, showing promising results for this problem [Abdalla-Aslan et al. 2020]. By harnessing the power of computer algorithms, clinicians can now detect subtle abnormalities, assess bone density, and track changes over time with greater precision than ever before. Computer methods have the potential to improve the accuracy and consistency of diagnosis serving as supporting tools for experts [Abdalla-Aslan et al. 2020]. It also lowers costs by eliminating routine tasks and modernizing the diagnosis process [Schwendicke et al. 2020]. Computational algorithms can analyze vast amounts of imaging data quickly and accurately, assisting clinicians in making more informed decisions. As technology continues to advance, IP-powered solutions are likely to play an increasingly important role in improving patient outcomes and reducing healthcare costs.

2. Research motivation

There are several challenges to overcome before computational solutions can be widely implemented in oral clinical practice, such as data privacy, algorithm transparency, lack of a regulatory approval ensuring the safe and ethical use of computational methods in healthcare, the lack of available data, the subjectivity of oral diseases' diagnosis, etc. The limiting issues for the adoption IP solutions in this field are aggravated by social and economic inequality. This creates a disparity in the access to advanced diagnosis tools between affluent and lower-income regions, exacerbating existing healthcare inequalities. Even considering that resource-poor regions were initially neglected in the recent growth of health computational solution, especially AI-based solutions, it can be observed that there are several recent factors that encourage the development of such solutions in neglected places. This scenario reinforces the importance of implementing solutions that can be accessed by a wide range of the population, including in public health units.

3. Objectives

This thesis' main objective was to investigate the applicability of image processing techniques in the oral/dental imaging field, through a selection of oral issues use cases that can be mapped to different computational tasks demanding the application of diverse IP techniques. For that, the following specific objectives were also considered: 1 - Selection of use cases in the Dentistry area; 2 - Design and implementation solutions for the considered use cases, selecting IP and AI techniques; 3 - Definition of metrics for evaluating the proposed solutions; 4 - Collection, curation and annotation of data (oral image examinations); 5 - Performance of experiments to evaluate the proposed solutions including quantitative and/or qualitative analysis.

4. Contributions

The contributions of this thesis can be summarized as: 1 - the creation of new original datasets composed of oral image examinations; 2 - new original methods for each use-case; 3 - free tool with a web graphic interface; 4 - scientific dissemination (publications in conferences, journals and book chapters).

5. Obtained results

In order to develop novel IP solutions for oral imaging, and achieve the proposed objectives, five different studies were performed, each one focusing on different medical/dental imaging tasks, referred here as use cases. These use case were selected during the initial evaluation of the clinical needs and demands present in a Brazilian public health unit (Oral Radiology Center in Policlinica Piquet Carneiro, part of the Universidade do Estado do Rio de Janeiro's health complex). The selection of use cases to be assessed in different studies to compose this thesis work is also one of its specific objectives, as described in the

Tabela 1. Use cases and respective information.

Use case/problems	Data	Algorithms	Metrics
Mandibular canal segmentation	Computed tomographs	IP techniques (Hough transform, mathematical morphology, equalization, thresholding)	IoU
Image quality improvement: super-resolution for periapical images	Intraoral radiographs (periapical)	IP techniques (interpolations) and AI models (CNN and GAN)	MSE, PNSR, SSIM, MOS
Periapical radiographs classification considering periodontal bone destruction	Intraoral radiographs (periapical)	AI models (CNNs: ResNet and Inception) and IP techniques (equalization)	Confusion matrices, PPV (precision), sensitivity (recall), specificity, NPV, AUC-ROC curve, AUC-PR curve
Approximal caries classification in bitewing radiographs	Intraoral radiographs (bitewing)	AI models (CNNs: ResNet and Inception) and IP techniques (mathematical morphology, equalization, thresholding)	Confusion matrices, PPV (precision), sensitivity (recall), specificity, NPV, AUC-ROC curve
Segmentation of tissue layers in ultrasound	Ultrasound image	AI models (YOLO and UNet)	Dice coefficient

Tabela 2. Average (and standard deviation) of MSE, PSNR, SSIM, and MOS for each method.

Method	MSE	PSNR	SSIM	MOS
Nearest interpolation	101.50 (± 46.44)	28.84 (± 3.02)	0.92 (± 0.03)	1.20 (± 0.45)
Bilinear interpolation	61.21 (± 29.37)	31.19 (± 3.41)	0.94 (± 0.03)	1.40 (± 0.55)
Bicubic interpolation	52.69 (± 25.57)	31.90 (± 3.55)	0.94 (± 0.02)	1.80 (± 0.45)
Lanczos interpolation	48.03 (± 23.68)	32.36 (± 3.67)	0.95 (± 0.02)	2.40 (± 0.55)
SRCNN [Dong et al. 2015]	33.27 (± 13.79)	33.55 (± 2.79)	0.98 (± 0.01)	2.60 (± 0.55)
SRGAN [Ledig et al. 2017]	16.99 (± 7.50)	36.40 (± 2.47)	**1.00 (± 0.00)**	**3.60 (± 0.55)**
SRGAN – TL – Chest X-ray	14.79 (± 6.46)	36.99 (± 2.42)	**1.00 (± 0.00)**	3.40 (± 0.89)
SRGAN – TL – Flowers	**13.19 (± 6.31)**	**37.70 (±2.96)**	**1.00 (± 0.00)**	3.40 (± 0.55)

Introduction chapter. These studies vary regarding the employed data and the techniques that compose their solutions, as summarized in Table 1. The definition of the techniques that compose the proposed solutions also consisted of an specific objective of this thesis work, as well as the selection of evaluation metrics to validate them.

The evaluation of the first method's (mandibular canal segmentation) performance considered the similarity between the canal volumes defined by the specialist and the canal volumes obtained by the method through their IoU. The IoU values obtained between the volumes defined by the proposed method and the corresponding ground truth volumes were, in average 0.790 (std.: 0.098). For comparison purposes, the IoU values of the volumes obtained by the 3D-UNet method, which is currently the state of the art for this problem were on average 0.755 (std.: 0.081).

Regarding the super-resolution problem, the methods performance were quantitatively compared (Table 2). This study used the Wilcoxon test to verify if the superiority denoted by the mean values in Table 2 is statistically valid. The confidence interval of 99% was considered. The H_A hypothesis is valid for all tests for all metrics, except for the SSIM comparing the two SRGAN trained with transfer learning.

Regarding the periapical radiographs classification, the generated models were evaluated with the test set. The results are in the confusion matrices presented in Table 3, where it is possible to see that the proportion of examples correctly classified (test accuracy) was 0.740 for the ResNet model and 0.817 for the Inception model. Table 4

Tabela 3. Confusion matrices for ResNet, Inception and SVM.

		n = 104	Predicted	
ResNet			Healthy	Bone loss
	Act.	Healthy	**38**	14
		Bone loss	13	39
Inception		n = 104	Predicted	
			Healthy	Bone loss
	Act.	Healthy	37	15
		Bone loss	4	**48**
SVM		n = 104	Predicted	
			Healthy	Bone loss
	Act.	Healthy	14	38
		Bone loss	8	44

Tabela 4. Results considering the test set.

Metric	ResNet	Inception	SVM
PPV (precision)	0.736	**0.762**	0.537
Sensitivity (recall)	0.750	**0.923**	0.846
Specificity	**0.731**	0.711	0.241
NPV	0.745	**0.902**	0.636
AUC-ROC curve	**0.864**	0.860	0.510
AUC-PR curve	**0.868**	0.847	0.564

shows the values achieved for the other metrics.

For the approximal caries classification in bitewing radiographs problem, The overall results and the specific results for each class are summarized in the confusion matrices (Tables 6 and 7). Finally, in order to evaluate the segmentation of tissue layers in ultrasound imaging, the Dice coefficient values were obtained considering the test dataset (Table 8).

6. Discussion

The solutions presented in this work compose a toolset that unifies them into an unique web graphic interface. The presentation of such solutions in this form is a novelty since there is no other current solution (commercial or free) that consolidates different IP and AI solutions for the considered oral issues. Also, each solution alone constitutes a novel method, either a completely new and original IP procedure or a established AI algorithm not previously proposed in the literature for this application field.

Regarding the mandibular canal segmentation, we highlight that our method eliminates the need for a large dataset to achieve good results. This advantage is significant, since in contexts such as the one discussed here, data samples, especially annotated data, are scarce [Moran et al. 2020].

Considering the image quality problem of super resolution for periapical images, the MOS metric analysis indicates that the difference observed by the expert in the perceptual quality of the SRGAN models (with and without transfer learning) is not statistically significant, which leads to a reflexion of the degree of details that can be perceived by humans, and how much of the improvement achieved is really relevant. Another important phenomena shown by the results is that the use of datasets of a similar domain in the pre-training step is not necessary for increasing the model's final performance. The results suggest that the super-resolution DL algorithms provide high quality images significantly improving the ability of visual analysis of the images.

Tabela 5. Performance of each CNN model considering the test set.

CNN	Learning Rate	Class	Precision	Recall	Specificity	NPV	AUC–ROC
Inception	0.001	Normal	**0.818**	0.600	0.933	0.823	0.643
		Incipient	**0.722**	**0.866**	0.833	**0.926**	**0.861**
		Advanced	0.687	0.733	0.833	**0.862**	0.810
	0.01	Normal	0.371	0.866	0.266	0.800	0.600
		Incipient	0.333	0.200	0.800	0.666	0.670
		Advanced	**1.000**	0.667	**1.000**	0.682	0.560
	0.1	Normal	0.000	0.000	**1.000**	0.667	0.500
		Incipient	0.000	0.000	**1.000**	0.667	0.500
		Advanced	0.333	**1.000**	0.000	0.000	0.500
ResNet	0.001	Normal	0.416	**1.000**	0.300	**1.000**	**0.807**
		Incipient	0.600	0.200	0.933	0.700	0.747
		Advanced	**1.000**	0.267	1.000	0.731	0.730
	0.01	Normal	0.379	0.733	0.400	0.750	0.612
		Incipient	0.333	0.200	0.800	0.667	0.609
		Advanced	0.714	0.333	0.933	0.737	**0.819**
	0.1	Normal	0.382	0.867	0.300	0.818	0.688
		Incipient	0.500	0.267	0.867	0.703	0.789
		Advanced	**1.000**	0.200	1.000	0.714	0.779

Tabela 6. Confusion matrices of each Inception model.

Learning rate	Actual and predicted cases per class				
0.001			PREDICTED		
			Normal	Incipient	Advanced
	TRUE	Normal	60%(9)	13%(2)	27%(4)
		Incipient	7%(1)	**86%(13)**	7%(1)
		Advanced	7%(1)	20%(3)	73%(11)
0.01			PREDICTED		
			Normal	Incipient	Advanced
	TRUE	Normal	**87%(13)**	13%(2)	0%(0)
		Incipient	80%(12)	20%(3)	0%(0)
		Advanced	67%(10)	27%(4)	7%(1)
0.1			PREDICTED		
			Normal	Incipient	Advanced
	TRUE	Normal	0%(0)	0%(0)	100%(15)
		Incipient	0%(0)	0%(0)	100%(15)
		Advanced	0%(0)	0%(0)	**100%(15)**

Tabela 7. Confusion matrices of each Resnet model.

Learning rate		Actual and predicted cases per class			
0.001			PREDICTED		
			Normal	Incipient	Advanced
		Normal	**100%(15)**	0%(0)	0%(0)
	TRUE	Incipient	80%(12)	20%(3)	0%(0)
		Advanced	60%(9)	13%(2)	27%(4)
0.01			PREDICTED		
			Normal	Incipient	Advanced
		Normal	73%(11)	13%(2)	13%(2)
	TRUE	Incipient	80%(12)	20%(3)	0%(0)
		Advanced	40%(6)	27%(4)	**33%(5)**
0.1			PREDICTED		
			Normal	Incipient	Advanced
		Normal	87%(13)	13%(2)	0%(0)
	TRUE	Incipient	73%(11)	**27%(4)**	0%(0)
		Advanced	67%(10)	13%(2)	20%(3)

Tabela 8. Dice coefficient values achieved for the different segmentation approaches considered in the experiment

Algorithm	Overall	Epidermis	Dermis	Hipodermis	Muscle components	Cortical bone
UNet	0.873	0.873	0.868	0.899	**0.864**	0.859
YOLO	**0.888**	**0.878**	**0.910**	**0.917**	0.863	**0.871**

Regarding the classification models, for the periodontal bone loss classification task, the ResNet's misclassifications are almost equally distributed on the considered classes, in constrast, Inception's incorrect classifications are mostly related to healthy regions misclassified as periodontal bone destruction cases. Moreover, SVM's bad results reinforces that using DL leads to a substantial improvement in the results for this task.

For the caries classification task, the best model present a very promising result, but there is a huge impairment in all other models' performance for the three different classes. It is possible to observe that none of the classes were unanimously favored by all models, suggesting that this problem is not focused on one class only, but affect the entire dataset. An increase in the whole dataset size would probably lead to better results.

Finally, regarding the tissue segmentation use case, the overall results can be considered promising. However, it can be observed that the algorithms still presented a significant difficult in handling with some specific samples. A more larger dataset would allow a better assessment of these phenomena, as well as promoting a more robust solution.

Analyzing the results of experiments for all use cases' solutions, some insights can be obtained. The first one are the promising results that suggest the potential of these solutions to be applied in the oral imaging field. Another point to be highlighted is the positive combination of traditional IP and DL techniques. The experiments suggested that these techniques can be somehow complimentary in the sense that some of their limitations can be minimized by combining them. For example, in the canal segmentation use case, the use of a DL algorithm would demand a huge amount of data that were not available for this work.

Referências

Abdalla-Aslan, R., Yeshua, T., Kabla, D., Leichter, I., and Nadler, C. (2020). An artificial intelligence system using machine-learning for automatic detection and classification of dental restorations in panoramic radiography. *Oral surgery, oral medicine, oral pathology and oral radiology*, 130(5):593–602.

Dong, C., Loy, C. C., He, K., and Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690.

Moran, M. B., Faria, M. D., Giraldi, G. A., Bastos, L. F., and Conci, A. (2020). Using super-resolution generative adversarial network models and transfer learning to obtain high resolution digital periapical radiographs. *Computers in biology and medicine*, 129:104139.

Schwendicke, F., Golla, T., Dreher, M., and Krois, J. (2019). Convolutional neural networks for dental image diagnostics: A scoping review. *Journal of Dentistry*, 91:103226.

Schwendicke, F., Samek, W., and Krois, J. (2020). Artificial intelligence in dentistry: Chances and challenges. *Journal of Dental Research*, page 0022034520915714.