A First Glance to the Quality Assessment of Dental Photostimulable Phosphor Plates with Deep Learning

Ariana Bermudez*, Saul Calderon-Ramirez†, Trevor Thang‡, Pascal Tyrrell‡, Armaghan Moemeni§, Shengxiang Yang‡, Jordina Torrents-Barrena¶

*Pattern Recognition and Machine Learning Group
School of Computing, Costa Rica Institute of Technology, Costa Rica
†Faculty of Dentistry, University of Toronto, Canada
‡Department of Medical Imaging, University of Toronto, Canada
§Centre for Computational Intelligence (CCI), De Montfort University, United Kingdom
¶University of Nottingham, United Kingdom

Department of Mathematics and Computer Engineering, Universitat Rovira i Virgili, Spain

Abstract—Photostimulable Phosphor Plates are commonly used in digital X-ray imaging for dentistry. During its usage, these plates get damaged, influencing the diagnosis performance and confidence of the dentistry professional. We propose a deep learning based classifier to discard or extend the usage of photostimulable phosphor (PSP) plates based on their physical damage. The system automatically assesses, for the first time in the literature, when dentists should discard their plates. To validate our methodology, an in-house dataset is built on 25 PSP artifact masks (Carestream, CS 7600) digitally superimposed over 100 Complementary Metal-oxide-semiconductor (CMOS) periapical images (Carestream, RVG 6200) with known radiologic interpretations. From these 2500 images, unique subsets of 100 images were evaluated by 25 dentists to find periapical inflammatory lesions on the tooth. Doctors’ opinion on whether the plates should be discarded or not was also collected. State-of-the-art deep convolutional networks were tested using five-fold cross validation, yielding classification accuracies from 87% to almost 89%. Specifically, InceptionV3 and Resnet50 obtained the best performances with statistical significance. Qualitative heat-maps showed that such models can identify and employ artifacts to decide on whether to discard the PSP plate or not. This work intends to be the base line for future works related to the automatic PSP plate assessment.

Index Terms—Dentistry, Photostimulable Phosphor Plates, Deep Learning, Quality Assessment.

I. INTRODUCTION

X-ray imaging, also known as radiography, is used by medical experts to explore internal body parts. This technique was developed and published by Wilhelm Rontgen in December 28th, 1895 [16]. Later, Otto Walkhoff used this method to obtain a radiography of the teeth [2]. Before X-ray machines were developed, dentists based their diagnosis on patients’ symptoms or by directly making incisions into the patients’ soft tissues. Radiographies are employed to detect decay between teeth and changes in bone density caused by gum disease. They are also useful in determining the proper fit of a crown and the marginal integrity of fillings [3]. Radiographic examinations are recommended periodically to determine the state of the teeth, bones and gums. Afterwards, additional imaging is prescribed based on individuals’ specific needs [1].

Digital radiography is becoming increasingly popular in dentistry due to its low radiation. It does not require film development, yields faster results and is easier to store [31]. There are two types of digital radiography: direct and indirect digital image receptors [4]. An example of the former is the CMOS sensor, in which a CMOS device converts the amount of X-ray energy into a digital signal using photosensitive pixels and on-chip circuitry [19]. PSP plates are indirect digital receptors made of plastic coated with a barium fluorohalide polymer that temporarily stores the X-ray energy onto the plate. With special scanners, this energy is read out and transformed into a digital image [17]. However, these plates are minimally protected, thus suffering physical damage caused by bends, peels and bites. This damage precludes the effective diagnosis of the teeth through radiologic images.

The dental community currently lacks up-to-date guidelines to replace damaged PSP plates (see Figure 1) [20] [35]. The few attempts of defining PSP plate usage guidelines used an objective image quality approach, ignoring the dentistry professional diagnosis performance [35]. Plates are discarded based on each dentist’s subjective opinion of whether the plate is going to interfere with the diagnosis. Dentists have varying degrees of knowledge and experience using PSP plates, thus considering different criteria to discontinue its usage. Subjective PSP plate disposal may increase the number of early discarded plates, or the continued usage of damaged PSP plates with diagnosis hindering artifacts. Using damaged PSP plates might require acquiring new radiologic data to ensure a faithful analysis of the region of interest, exposing patients to additional irradiation.
Given the urgent demand of a standardized procedure to discard or keep plates, we propose, for the first time to our knowledge, a deep learning classification method to estimate whether a dental PSP plate should be replaced due to diagnostically-hindering artifacts. Section II addresses previous related work. Afterwards, section III details the designed approach and the dataset used. Experiments and results are then described in Section IV. Finally, conclusions and future work are discussed in Section V.

II. RELATED WORK

To the best of our knowledge, no work on using machine and deep learning methods for automatic qualitative assessment of PSP plates has been reported yet. Previous techniques used subjective classification schemes to categorize PSP plates based on their damage [20]. Particularly, authors in [17] studied the presence, frequency, and causes of artifacts in intraoral images obtained using PSP plates. Recently, a dataset composed with PSP plates assessed has been developed in [35], where the interference of artifacts in the PSP plates in diagnosis performance of periapical inflammatory disease is evaluated. The study concludes that the PSP plate artifacts interfere mostly in the confidence of the dentistry professional. The dataset developed in [35] is used in this work.

Nevertheless, machine and deep learning approaches have been successfully used in medicine and dentistry. Kavitha et al. [22] employed support vector machines to diagnose osteoporosis from dental panoramic radiographs. Lately, Prajapati et al. [28] developed a VGG16-based Convolutional Neural Networks (CNN) to identify dental caries, periapical infection and periodontitis. Another deep network [21] was recently trained on optical coherence tomography images to classify human oral tissues and detect dental caries.

The PSP plate quality assessment can be generalized as a visual quality inspection of materials task, a problem widely tackled in literature [26]. In Sun et al. [36], authors argued that the prevalence of deep learning solutions have increased considerably in metal quality assessment since 2014 due to their high accuracy and flexible data-driven feature extraction. Therefore, automatic visual inspection techniques based on traditional statistical or filtering approaches have declined in the past few years. A recent example of material visual quality assessment using deep learning can be found in [9], where the authors implemented a neural network system for visual fabric (material) inspection. Specifically, the accurate and fast region-based CNN proposed by [29] was applied to damage detection.

III. MATERIALS AND METHODS

In this work, ResNet [18] and Inception [33] architectures were tested to classify PSP plates. Furthermore, exact damage localization was performed via class activation maps [30].

A. Dataset

Blank images of PSP plates (Carestream, CS 7600) were acquired from the oral radiology clinic at the Faculty of Dentistry (University of Toronto, Canada). A sample is shown in Figure 1. PSP plates were wiped with a 0.6% w/v sodium hypochlorite solution, covered in a plastic hygiene sheathes, and subsequently exposed using a typical posterior periapical setting (Belmont Phot-X IIs, 70 kV, 6.0 mA, 0.22 seconds). The protective sheathes were removed and the plates were scanned and digitalized. These images were exported from the workstation to a portable network graphics format. More concretely, we selected 25 PSP plates: 15 and 5 plates presented severe and intermediate damage, respectively, while 4 plates are new, and the last one is a blank mask image.

A total of 531 dental images were obtained from the Faculty of Dentistry’s picture archiving and communication system (PACS) in tagged image file format (TIFF), and acquired using a CMOS digital sensor (Carestream, RVG 6200). A sample image is depicted in Figure 2). CMOS images are artifact free, as no physical plates are used for its acquisition. Therefore, we used them to semi-artificially produce the PSP plate image samples with different degrees of damage, avoiding any additional patient radiation exposure.

In short, 100 CMOS cases were combined with 25 PSP plate images to augment the data up to 2500 images. The remaining 431 images superimposed with a blank plate correspond to non-disposable samples (see Figure 3). The dataset generation pipeline is depicted in Figure 4. Pixel addition was performed to create these 2500 samples, where each pixel was weighted and summed with the according pixels’ intensities, producing a new image. More details about the pixel weighting sum procedure can be found in [35]. Hence, our dataset is finally composed of 2931 images, where 1320 and 1164 images stand for both no PSP plate disposal and dispose PSP plate classes, respectively.

B. Image labeling

Our 2500 generated images were equally distributed to 25 dentists (100 images each) to evaluate them. Thus, one label per observation (image) was obtained. The minimum clinical requirements to participate in this study were: 1)
being a licensed member of the Ontario dental legislative body (Canada), and 2) being a dentist with at least 1 year of experience, specifically in detecting periapical pathologies from radiologic images. Oral and maxillofacial specialists were excluded from the study. All images were assessed using the same procedure and display configuration. We utilized a LED Dell 24 inches monitor: P2417H, 1920 × 1080, 60 Hz, brightness and contrast of 75%, with the Windows Photo Viewer. Experts labeled each image by thinking whether they would discard the plate or not.

C. Proposed computational method

We implemented three state-of-the-art deep CNN architectures (i.e., ResNet18, ResNet50 and InceptionV3) to positively or negatively assess PSP plates. The selected models were sampled to represent low (ResNet18), mid (ResNet50) and high (InceptionV3) performing architectures, according to [6].

These models were trained with the aforementioned superimposed images (see Section III-A). Additionally, pretrained weights obtained from the ImageNet Large Scale Visual Recognition Challenge [11] were loaded and tuned for binary PSP plate classification. Our input samples were standarized and resized to 299×299 pixels for its usage in InceptionV3 and to 224×224 to feed the ResNet architectures.

ResNet [18] is a very deep CNN that attenuates the problem of vanishing and exploding gradients by using skip connections. Also, the ResNet architecture implements as a top model a global average pooling step, lowering the total number of parameters to estimate. The tested ResNet18 consists of 18 layers with 11 million of parameters, while ResNet50 is composed of 50 layers and 25 million of parameters.

The Inception CNN architecture (also known as GoogleNet) [33] adds consecutive inception modules with small kernel convolutions to drastically reduce the number of parameters. The InceptionV3 architecture [32] consists of nearly 25 million of parameters to estimate during training.

To visualize the damages detected in the PSP plates, we employed the Gradient Class Activation Maps (Grad-CAM) developed by Selvaraju et al. [41]. Grad-CAM performs a forward pass and evaluates the gradients per class. These gradients are subsequently used to generate an image with the activated neurons in the target layer.

IV. EXPERIMENTS AND RESULTS

We performed a 5-fold cross validation. A total of 732 images (20%) were used for model evaluation, while the remaining 2199 images (80%) were used for training. The accuracy, sensitivity and specificity were measured for all tested CNN models to robustly analyse their success rate.

As for the hyper-parameters of the tested architectures, a cross entropy loss function optimized with stochastic gradient descent with a momentum of 0.9, and a learning rate of \( \alpha = 0.001 \) was used, and a batch size of 4 and a maximum of 10 epochs was used during training, as model convergence proved to be fast in our empirical tests.

Table I summarizes the aforementioned performance measures, along with a statistical description of the results for \( n = 20 \) experiments, \( k = 5 \) folds and \( i = 100 \) observations, thus yielding \( d = 100 - 1 = 99 \) degrees of freedom. To reproduce our results and statistical analysis, the CSV files with the yielded values can be found in our Github account https://github.com/saul1917/PSPclassification.git.

<table>
<thead>
<tr>
<th>Architecture</th>
<th>( \bar{\text{F}}_{\text{acc}} )</th>
<th>( s_{\text{acc}} )</th>
<th>( \bar{\text{F}}_{\text{sen}} )</th>
<th>( s_{\text{sen}} )</th>
<th>( \bar{\text{F}}_{\text{spe}} )</th>
<th>( s_{\text{spe}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>InceptionV3</td>
<td>88.40</td>
<td>1.58</td>
<td>86.82</td>
<td>3.17</td>
<td>87.13</td>
<td>3.03</td>
</tr>
<tr>
<td>Resnet50</td>
<td>88.08</td>
<td>1.61</td>
<td>88.56</td>
<td>2.79</td>
<td>85.56</td>
<td>3.35</td>
</tr>
<tr>
<td>Resnet18</td>
<td>87.39</td>
<td>1.55</td>
<td>88.79</td>
<td>2.99</td>
<td>83.57</td>
<td>2.91</td>
</tr>
</tbody>
</table>

According to Table I, InceptionV3 seems to slightly outperform ResNet in terms of accuracy and specificity, confirming its sample mean higher performance yielded in [6]. However, a statistical analysis is needed to validate our findings. First,
we employed a Kolmogorov-Smirnov test to confirm that the distribution of our results is normal in terms of accuracy, sensitivity and specificity for all $i = 100$ observations. Figures 7, 8 and 9 shows the valid box-plot for normally distributed data, which graphically depicts the small differences of the obtained accuracy, specificity and sensitivity, respectively, for the three models.

Given the normal distribution of our dataset, an ANOVA test is also performed [15] for the accuracy, sensitivity and specificity of the three models. Table IV shows that, with a $p < 0.05$, InceptionV3 is better in terms of accuracy than Resnet18 with statistical significance. However, there is not a statistical meaningful difference of InceptionV3 when compared to Resnet50. Specificity wise, InceptionV3 outperforms both models with statistical significance. Particularly, it outperforms Resnet18 by a large margin.

<table>
<thead>
<tr>
<th>Resnet50</th>
<th>Resnet18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.1409 (ND)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>5.4737e-05 (-1.74)</td>
</tr>
<tr>
<td>Specificity</td>
<td>6.1710e-04 (+1.57)</td>
</tr>
</tbody>
</table>

TABLE II

INCEPTION3 ANOVA p VALUE OF RESNET50 AND RESNET18. ITS IMPROVEMENT (+) OR DETERIORATION (-) IN PERFORMANCE, AND NO DIFFERENCE (ND) IS REPORTED WITH 95% OF CONFIDENCE ($p < 0.05$).

Finally, Table IV depicts that Resnet50 outperforms its Resnet18 in terms of accuracy with statistical significance by a modest margin, given its statistically equal sensitivity and larger specificity boost.

<table>
<thead>
<tr>
<th>Resnet18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
</tr>
<tr>
<td>Sensitivity</td>
</tr>
<tr>
<td>Specificity</td>
</tr>
</tbody>
</table>

TABLE III

RESNET50 ANOVA p VALUE OF RESNET18. ITS IMPROVEMENT (+) OR DETERIORATION (-) IN PERFORMANCE, AND NO DIFFERENCE (ND) IS REPORTED WITH 95% OF CONFIDENCE ($p < 0.05$).

As for a qualitative performance assessment, we generated heatmaps of the on the input images [30] to map exactly which pixels influenced the classification process. The qualitative results obtained for the tested images are fully consistent. When the selected class is dispose PSP plate, most of the pixels belong to artifacts of the plate (see Figure 5). Otherwise, when the selected class is no PSP plate disposal, the activation map is frequently focused on image regions with no artifacts (see Figure 6).

V. DISCUSSION AND CONCLUSIONS

This work proposes the first fully automatic system for PSP plate quality assessment. Our semi-artificial training dataset can mimic PSP plate images with teeth without exposing patients to high levels of radiation. The presented model can precisely and efficiently assist the reuse or disposal of PSP plates, which reduces the variability of opinions among dentists and favors the clinical consensus.

All tested models yielded accuracies around 86% and 88% after performing a 5-fold cross validation. A Kolmogorov-Smirnov test was conducted to confirm a Gaussian distribution of the results. An ANOVA test was also performed to statistically measure result confidence [5], concluding that InceptionV3 outperforms statistically better than the tested ResNet models, except for ResNet50, accuracy wise. We stress the importance of analysing statistically the obtained results in order to reach rigorous and accurate conclusions, as also seen in [12], [5]. Unfortunately, both machine and deep learning communities often skip these statistical tests.

Additionally, we employed the Grad-CAM technique to
might serve as a baseline for future automated PSP plate quality assessment.

**REFERENCES**


Another improvement for the implemented computer aided assessment is the usage of uncertainty techniques as the ones found in [34], [7], [8], [23]. We plan to evaluate the feasibility of the current implementation to determine if the yielded accuracy is practically enough, and also to evaluate the impact of uncertainty estimation in the automated assessment.

Finally, object-based CNN architectures [10] will be explored deeply to enhance damage localization in PSP plates. The dataset aims to be publicly available soon, as this work


