UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

Odontological lesion detection using deep learning

Daniel Sebastián Porras Dávila

Ingeniería en Ciencias de la Computación

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniero en Ciencias de la Computación

Quito, 19 de diciembre de 2021

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

HOJA DE CALIFICACIÓN
DE TRABAJO DE FIN DE CARRERA

Odontological lesion detection using deep learning

Daniel Sebastián Porras Dávila

Nombre del profesor, Título académico **Noel Pérez, Ph. D**

Quito, 19 de diciembre de 2021

© DERECHOS DE AUTOR

Nombres y apellidos:	Daniel Sebastián Porras Dávila
Código:	00200948
Cédula de identidad:	171910789-6
Lugar y fecha:	Quito, 19 de diciembre de 2021

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en http://bit.ly/COPETheses.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on http://bit.ly/COPETheses.

RESUMEN

En el mundo odontológico, la detección temprana de diferentes lesiones bucales es de vital importancia para realizar un tratamiento certero y preciso con el fin de cuidar la salud del paciente. Sin embargo, no siempre es posible que los dentistas hagan un diagnóstico precoz debido a la escasez de síntomas o la falta de experiencia. En este contexto, este trabajo tiene como objetivo ayudar a los dentistas a detectar lesiones de forma temprana mediante el uso de redes neuronales convolucionales. El conjunto de datos de entrada para el entrenamiento de la red está compuesto por imágenes de rayos X orales con dos clases, las que presentan caries y las que no. Para el paso de preprocesamiento, cada imagen fue editada con CLAHE para mejorar el entrenamiento. Luego, cada región de interés (ROI) se recortó de la imagen original y se guardó como una nueva. Para estandarizar el tamaño de la capa de entrada de la CNN, se agregó algo de ruido gaussiano como relleno para las ROI más pequeñas, y para las ROI más grandes se utilizó un simple cambio de tamaño. Si alguna imagen presentaba algún otro tipo de lesión, se descartaba automáticamente.

Finalmente, el conjunto de datos obtenido se utilizó para entrenar 3 arquitecturas CNN diferentes. Cada modelo se entrenó variando el número de épocas y el tamaño del lote, y medimos métricas como AUC, exactitud, pérdida, precisión y recuperación. El mejor modelo fue el primero, obteniendo un valor AUC de 0.998 y un valor de pérdida de 0.04.

Palabras clave: redes neuronales convolucionales, caries, clahe, región de interés, clasificación binaria, detección de objetos.

ABSTRACT

In the dental world, the early detection of different oral lesions is of vital importance to carry out an accurate and precise treatment in order to take care of the patient's health. However, it is not always possible for dentists to make an early diagnosis due to scarce symptoms or lack of experience. In this context, this work aims to help dentists detect lesions early by using convolutional neural networks. The input dataset for training the network is composed by oral x-ray images with two classes, those which present cavities and those which not. For the preprocessing step, each image was edited with CLAHE to improve training. Then, every region of interest (ROI) was cropped from the original image and saved as a new one. To standardize the size for the input layer of the CNN, some Gaussian Noise was added as padding tom the smaller ROIs, and for the larger ROIs a simple resize was used. If any image had any other type of lesion, it was automatically discarded.

Finally, the obtained dataset was used to train 3 different CNN architectures. Each model was trained varying the number of epochs and the batch size, and we measured metrics such as AUC, accuracy, loss, precision and recall. The best model was the first one, obtaining a AUC value of 0.998 and a loss value of 0.04.

Key words: convolutional neural networks, cavities, clahe, region of interest, binary classification, object detection.

TABLE OF CONTENTS

INDEX OF TABLES

INDEX OF FIGURES

INTRODUCTION

Within Medicine, the branch of Dentistry has undergone lots of changes in recent years due to the technological innovation in tools that help dentists to perform a more precise work. However, one of the main problems for dentists remains the early detection of lesions, especially caries on occlusal surfaces (Ohnishi, Y. et al, 2011). This surface has grooves and spaces that are difficult to access for cleaning, which is why it is here where most cavities occur. Caries can be defined as a demineralization or destruction of dental tissue (enamel and dentine) produced by the decomposition of food debris that generate harmful acids in the dental plaque, which can lead to pain, infections and eventually, tooth loss (World Health Organization, 2020) (Liebe-Harkort, C., Astvaldsdóttir, A., Tranæus, S., 2009). Among the main causes for the formation of cavities we find poor cleaning, excessive consumption of sugary drinks, desserts or carbohydrates; and the little generation of saliva in the mouth (Datta, Soma., Chaki, Nabendu., Modak, Biswajit., 2020).

Caries treatment turns out to be so important that, in 2014, a study published in Chapter 6 of the book "Intelligent Data Analysis: From Data Gathering to Data Comprehension" was carried out, which shows a global map of caries in children aged 12 years or less according to the DMFT (Decayed, Missing and Filled Teeth) metric (Young, D. A., Nový, B. B., Zeller, G. G., et al, 2015). The map perfectly illustrates the number of children affected by cavities around the world, leading to treatments ranging from tooth repair to tooth extraction and filling (see Figure 1). The most striking thing about the map is that, if we analyze the South American region, there is not a single country where the DMFT level is classified as “very low” (Datta, Soma., Chaki, Nabendu., Modak, Biswajit., 2020). In fact, countries like Ecuador or Bolivia are classified as “high”, which means that most children under 12 years old suffer from cavities that require urgent treatment.

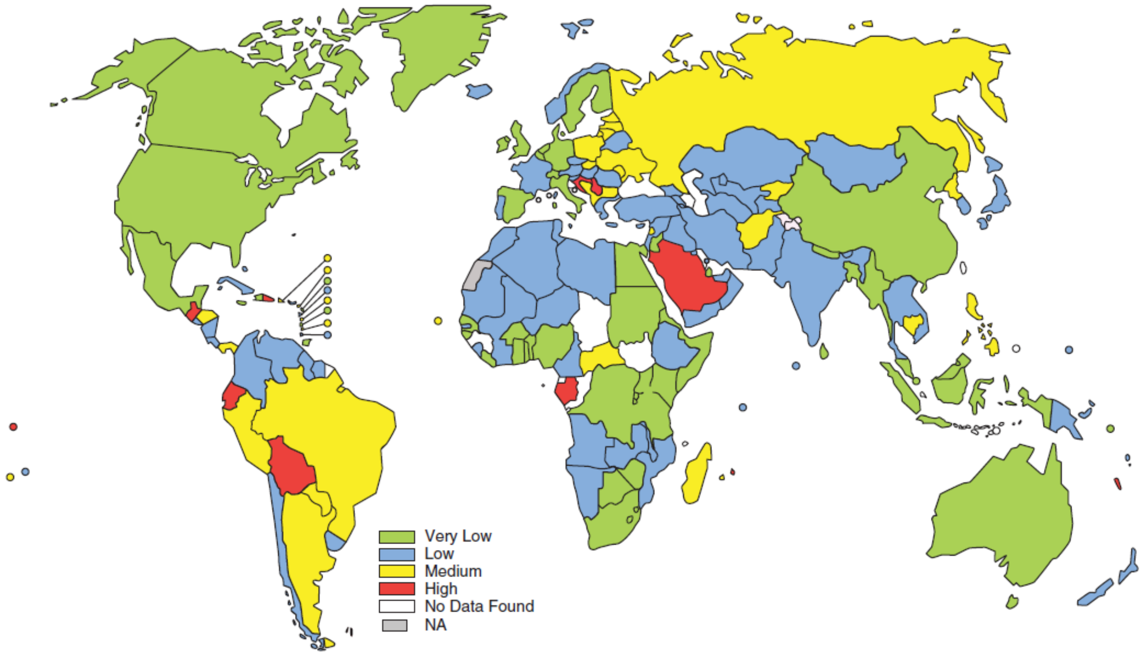

Figure 1: Global overview of caries involvement in 12-year-old children according to the DMFT classification.

This is due not only to the fact that detection is not easy, but in most South American countries, dentists do not have the appropriate technology for early diagnoses. Therefore, those who do not have complete equipment, generally perform the detection by "naked eye", a technique that introduces a lot of human error and a high probability of missing injuries in their initial phase. Nowadays, there are some techniques for caries detection, among which are visual inspection, x-rays, use of optical tools, etc (Oliveira, J., Proença, H, 2011). For example, taking radiographs periodically helps dentists a lot in terms of diagnosis and proper treatment (Birdal, Ramiz., Gumus, Ergun., Sertbaş, A., Sinem, I., 2015) (Shahbazian, M., Vandewoude, C., Wyatt, J. et al, 2013). However, radiographs by themselves are not enough, a mistake in which many dentists fall into. In fact, the British Dental Journal maintains that only 3% of healthcare professionals combine more than one technique, so there is still a lot of human interpretation in the detection process (Slimani, Amel., Terrer, Elodie. et al, 2020).

For all the techniques mentioned above, specificity and sensitivity can be determined. Sensitivity is defined as the capability to detect lesions when they are present, while sensibility is the capability to determine a healthy surface when it really is. For example, visual detection often has a range of 14-24% in sensitivity and 70-99% in sensibility (Martignon, S., Uribe, S., Pulido, A., Cortés, A. et al, 2013). In order to reduce human error and improve values of specificity and sensitivity, slightly more automated techniques have been introduced. One of them was "Digital Substraction" in 2008 (Oprea, S., Marinescu, C., Lita, I., Jurianu, M. et al, 2008). It consisted of feeding a software with X-ray images, which went through processing to isolate the tooth, convert it into black and white (binary), fill the image and then subtract it from the original to isolate cavities. Although it is true this is a useful process, it is not used much today since more details are needed such as the type of injury and the degree of progress it presents (Igarashi, Y., Yoshida, S., Kanazawa, E., 2017).

Another technique is the ICDAS II (International Caries Detection and Assessment System), which is a system agreed upon in Maryland in 2005 that classifies cavities into 7 stages based on histological activity (Braun, A., Guiraud, L., Frankenberger, R., 2016). The aim of this was to detect caries on patients in the initial clinical examination. Thus, dentists analyze the tooth, compare it with the ICDAS II matrix and can determine the degree of injury that the patient presents. In this way, dentist could prevent the development of a more serious lesion in the future and improve the oral health of people. This technique has shown results in a range of 70-85% for sensitivity and 80-90% for specificity in detection of caries in occlusal surfaces and lesions located in the external half of the enamel (Carvalho, R., Letieri, A., Vieira, I., Santos, T. et al, 2018). However, this tool alone is not yet the ideal way for detection because it still requires a lot of visual inspection by the dentist, who can miss some very small lesions that are not easily perceptible to the human eye.

The main objective of this research is to provide dentists a more specialized tool that can help them to detect cavities lesions early on x-ray images. To achieve this, we used convolutional neural networks, which is the most recommended technique to extract features such as contours or edge points (Diaa, M., Hany, Ammar, 2006) due to the combination of convolutional and pooling layers on its architecture (Yamashita, R., Nishio, M., Do, R. et al, 2018) (Schwendicke, F., Golla, T., Dreher, M., Krois, J., 2019). The input images were preprocessed and then were used to feed 3 different DCNN models. The aim was to maximize the AUC metric and minimize the loss function to achieve results with, at least, 90\% level of confidence.

MATERIALS AND METHODS

A. Database

The database used for this research was obtained from the private dental clinic of Universidad San Francisco de Quito USFQ. It is important to mention that all images were used with the authorization of the patients and the clinic. It had in total 928 x-ray images which were not annotated on its original version. By using an online free software named Roboflow, and with the help of experts in odontological lesions, we were able to draw bounding boxes manually that identified caries on each image (see Figure 2). Some images had a wide variety of lesions, not just caries, so in those cases we ignored other lesions and focused only in all the caries that the image presented. Also, there were some images on the database that were not clear enough to identify any kind of lesion, so they were discarded automatically. Finally, we decided to export all the annotations in a xml format since it was easier to read that kind of file, identify the coordinates of the bounding boxes and filter them into our two classes.

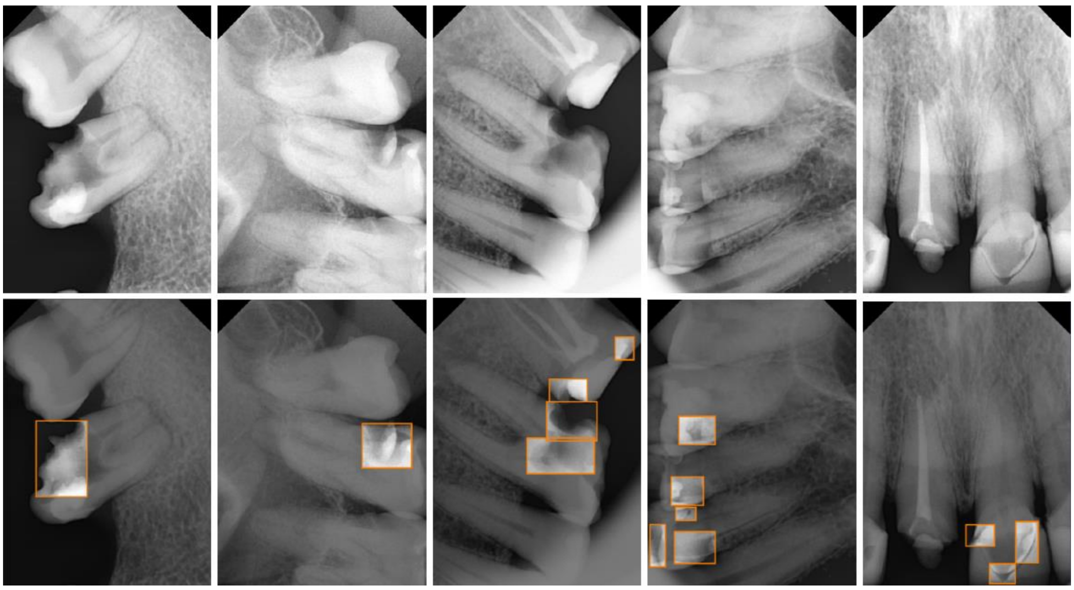

Figure 2: Some samples of images in database: original images provided by the clinic (first row) and images annotated by odontological experts identifying caries (second row) after using Roboflow for drawing the bounding boxes.

B. Deep learning models

Deep learning is a machine learning technique used to give computers the capability to learn from data or patterns in order to understand the real world (Li, X., Zhang, G., Li, K., Zheng, W, 2016). Models developed for deep learning are gaining more and more popularity since they have lots of possible applications. Moreover, they are optimized to work with big data, which is a very important advantage due to the huge amount of data that is generated daily all over the world. Contrary to traditional machine learning algorithms, which work with smaller datasets. There are lots of models in deep learning optimized for different tasks, such as the traditional DNN, deep autoencoders, RNN, CNN, etc (Navamani, T., 2019).

The deep convolutional neural networks are one of the bests kinds of neural networks for computer vision (Lee, J., Kim, D., Jeong, S., Choi, S., 2018). The traditional architecture of CNN consists of, at least, a convolutional and max pooling layers. Starting from this, lots of configurations could be created, transforming a traditional CNN into deep by adding an input layer, several hidden layers and an output layer. This helps the network to extract more features to learn better. The very first successful deep CNN is Alexnet, designed by Alex Krizhevsky, and known as the winner of the ImageNet Large Scale Visual Recognition Challenge in 2012 (Alake, R., 2020), where it was proved for the first time that a deep CNN was used for image classification (Krizhevsky, A., Sutskever, I., Hinton, H., 2012). Alexnet´s architecture is composed of 5 convolutional layers, 3 max pooling layers, 2 normalization layers, 3 fully connected layers and softmax as the activation function on the last layer for categorical classification (Wei, J., 2019) (Great Learning Team, 2020). Using these details as a prototype, a new architecture was created to fit better into our initial problem.

C. Proposed method

Although the traditional architecture of Alexnet is a very good model, it has details that does not match the aim of this research. Taking as starting point the skeleton and principals of Alexnet, we created 3 different deep CNN models. In terms of architecture, they are very similar, however, they were designed to have different number of layers, different kernel sizes, etc. All the models were programmed using Python version 3.7.12 and some deep learning libraries such as Tensorflow and Keras version 2.7 (Keras Team, 2021).

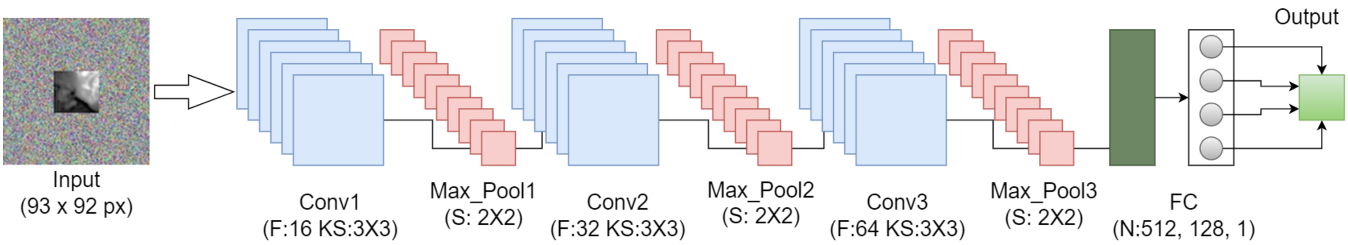

Figure 3: Model 1 of the proposed method: F - number of filters; KS - Conv. Kernel Size; S - Max Pool Kernel Size; FC - Fully connected; N - number of neurons.

To illustrate better our proposed method, Figure 3 shows the architecture created for model 1, which is very similar to the other two models just varying the number of layers and filters. In this figure, we show that the input ROI image (size of 93x92 px) is fed to the first convolutional layer, which is the one in charge of creating a feature map representation based on the kernel size 3x3 established and the convolutional filters. Once this is created, the map serves as input for the max pooling layer with kernel size 2x2, whose job is to reduce irrevelant features from it to obtain a more specialized map in terms of features. The debugged map then enters the second convolutional layer and the process repeats. Finally, the fully connected section of the network is composed by four layers. The first one is the flatten layer, which reduces bidimensional features into a single vector of characteristics. The reamining three are dense layers which help on the classification task. It is important to mention that the activation function of the last layer, in our case, has to be sigmoid (Hentschke, H., 2019) (Brownlee, J., 2021). Since its output ranging is from 0 to 1, it is the best one for binary classification (Karakaya, M., 2020). Also, four dropout layers were introduced, each with ratio 0.1, in order to reduce the probability of overfitting.

Model 2 is very similar to the previous architecture, with the slightly difference that we do not have the same number of convolutional and max pooling layers. Also, the position of each layer in the network varies. The kernel sizes for all the convolutional layer was 5x5, and for the max pooling was 3x3. The input ROIs are fed to a convolutional layer with 32 filters. Then, it passes through a max pooling layer. Later, two consecutive convolutional layers appear, each with 64 filters, and finally the last max pooling layer. The fully connected section of the network remains with 4 layers composed by the flatten layer and three dense layers with 256, 256 and 1 neurons respectively. This network did not had dropout layers.

Model 3 was created to be larger than the other two in order to demonstrate the effects of training the dataset used in this research in a deeper network with more filters. This model presented four convolutional layers. This layers had 64, 128, 256 and 128 filters respectively. Moreover, the kernel sizes were 7x7, 5x5 and 3x3 for the remaining two layers. After each convolutional layer, a max pooling layer was included. In total, there were four of this layers, each with kernel size of 2x2. Once again, the fully connected section remained the same with four layers composed by the flatten layer and the remaining three were dense layers with 512, 128 and 1 neurons respectively. For this model, five dropout layers were included, each one with ratio 0.1.

D. Experimental setup

The experimental setup was divided into 4 stages.

1) *Data preprocessing:* as explained before, the original database contained 928 x-ray images from some patients of the odontological clinic at USFQ. The first step consisted on making annotations of the different lesions present on each image of the database using Roboflow. Once the annotations were finished, the next step was to filter and separate the images that present caries from the images that did not present caries. The algorithm was optimized so that, while being filtered, each image was passed through a histogram equalization or CLAHE, as shown in Figure 4. This is a tool that helps to improve the contrasts of the image, and the result was saved on different directories depending on the presence or not of caries on the image.

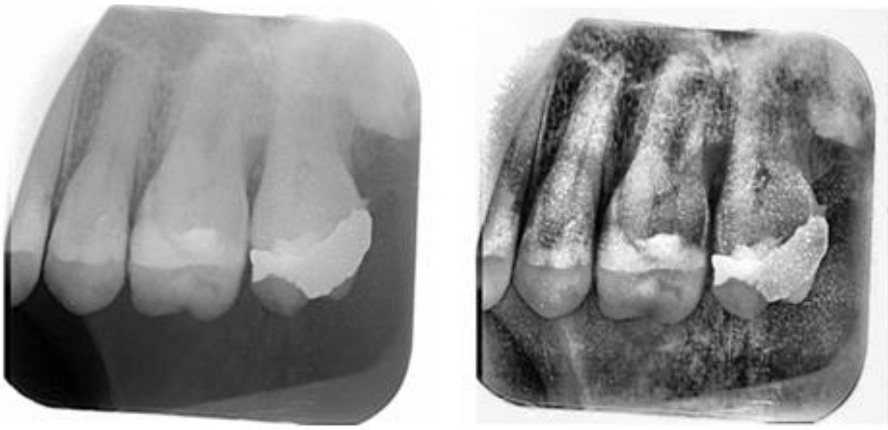

Figure 4: Histogram equalization preprocessing step: example of an original image of the database (left) and the same image with improved contrasts after passing through CLAHE (right).

At the end of this part of the process, we ended with 230 caries images and 148 non caries images. The remaining images were discarded due to poor quality or the lesion size was to small to be considered important for training. Since the final image repository did not have too many samples, we did data augmentation to increase the number of trainable samples. For this, we used 4 transformations, rotation of 90 degrees to the right, rotation of 90 degrees to the left, vertical flip and horizontal flip. Finally, in order to have a more specialized training in the network, we decided to crop regions of interest (portions of images with presence of caries and absence of caries) of each image so the network could learn better the differences of each class. For this, we obtained the biggest and the smallest ROI in terms of size, and calculated the average size, which was 93x92 px. ROIs that were bigger to the average were downscaled, while ROIs that were smaller where centered and some Gaussian noise was added as padding to achieve the desired size. At the end, we obtained an unbalanced dataset of 4545 images, 1515 ROIs samples with presence of caries and 3030 ROIs samples with absence of caries.

2) *Training and test sets:* we used the 10 stratified cross validation technique to build training and testing sets. This method is very good to build this kind of sets for two important reasons, among others. First of all, it helps to ensure the sample ratio for all folds. Second of all, this technique introduces some interesting variations because in every fold, both training and test sets changes, which makes our results for each model trustworthy (Muralidhar, K., 2021).

   In addition, we created an extra separated dataset manually containing 60 sample images. This dataset was used to validate the training results obtained with the mean of the folds and to create the ROC and Precision-Recall curves (by making predictions with the selected as best model) as part of the final analysis of this experiment.

3) *Model configuration:* besides the variety of architectures explained in previous sections, we used some hyperparameters as constants for training and testing all models. In terms of number of epochs used, we set the iteration to 200 epochs, with increment of 50 units. It was not worth it to increase over 200 iterations because overfitting also increased and the metrics did not improve. Moreover, datasets were trained with two different batches sizes, 16 and 32 units for each model. The optimizer used to compile the model was Adam, which showed the best result comparing to RMSprop or SGD

(stochastic gradient descendent). Moreover, as the loss function we used binary crossentropy instead of categorical crossentropy since we only have 2 target classes. Finally, in models 1 and 3, some dropout layers were added with a rate of 0.1 to reduce overfitting. This means that, in each training epoch, some neurons are randomly disabled so the remaining activated neurons are obliged to learn more by their own.

4) *Assessment metrics:* several metrics were used to monitor the training and later the predictions made by the models on the testing set. During training, for each epoch, we monitored accuracy, loss, precision, recall and AUC. With this, we were able to draw some important plots to measure the performance for each model and select the best one according to the greatest AUC and the lowest loss metrics. The combination of this two metrics helped us to achieve a level of confidence greater than 90%.

RESULTS AND DISCUSSION

A. Performance evaluation

Table I shows the results obtained for the three models according to the variation of epochs and batch sizes. For DCNN1, the AUC metric was always above 0.95, which is an outstanding result for a binary classification problem like the one held in this research. Moreover, analyzing the loss metric, the highest value on this first architecture was 0.09, which means that with this model we achieved a level of confidence above 90% in every iteration and with the two values of batch sizes proposed. In addition, the model never presented overfitting in any of its variants. Considering the higher AUC, there were two models that achieved 0.998. However, if we add the loss metric to the consideration and search the lower of the two models, the best one was achieved with 32 as batch size and 150 epochs. The mentioned model obtained an AUC of 0.998 and a loss of 0.04 (see the bold line in Table I).

On the other hand, DCNN2 was not as stable as DCNN1 on every iteration. If we analyze just the results obtained with batch size 16, it is easy to see that with 100 epochs or less, the model obtained a very good AUC above 0.90. However, the loss function slightly stays around 0.10, which is a good result, but in the medical world every hundredth counts since that could be the difference between a patient living or dying. Moreover, beyond 150 epochs, the model losses its generalization power and falls into overfitting. This is shown by the precision and recall metrics for this number of iterations since their measure is 0. Additionally, loss increases over 0.5 while AUC could not even equalize that value. This means that the model, in this point, had a level of confidence lower than 50%, which is not acceptable. Instead, with a batch size of 32, this model works very well on each iteration, obtaining once again outstanding results above 95%. The best model was achieved with this batch size and 200 epochs, obtaining an AUC of 0.998 and a loss of 0.05 (see the bold line in Table I).

Table 1: Performance results of deep learning models

Model	Conv. Layer (f)	Kernel Size	FC Layer (n)	Batch Size	Epochs (u)	AUC (u)	Loss B.C	ACC (%)	PRE (u)	REC (u)
DCNN1	(16,32,64)	(3x3)	(512,128,1)	16	50	0.996	0.05	96	0.98	0.92
				32	50	0.993	0.09	95	0.95	0.92
				16	100	0.995	0.07	96	0.98	0.92
				32	100	0.996	0.06	97	0.97	0.94
				16	150	0.997	0.05	97	0.99	0.93
				32	**150**	**0.998**	**0.04**	**97**	**0.99**	**0.94**
				16	200	0.998	0.05	97	0.99	0.93
				32	200	0.997	0.05	97	0.98	0.94
DCNN2	(32,64,64)	(5x5)	(256,256,1)	16	50	0.993	0.09	95	0.96	0.92
				32	50	0.991	0.11	94	0.93	0.92
				16	100	0.994	0.11	96	0.97	0.91
				32	100	0.998	0.06	97	0.97	0.93
				16	150	0.490	0.65	66	0.00	0.00
				32	150	0.996	0.06	97	0.96	0.94
				16	200	0.490	0.65	66	0.00	0.00
				32	**200**	**0.998**	**0.05**	**97**	**0.99**	**0.94**
DCNN3	(64,128,256, 128)	(7x7) (5x5) (3x3)	(512,128,1)	16	50	0.491	0.67	63	0.47	0.01
				32	50	0.723	0.61	67	0.71	0.67
				16	100	0.493	0.65	63	1.00	0.01
				32	100	0.851	0.46	76	0.82	0.73

16	150	0.478	0.66	63	1.00	0.01
32	150	0.889	0.40	80	0.85	0.77
16	200	0.488	0.66	63	1.00	0.01
32	**200**	**0.942**	**0.29**	**86**	**0.89**	**0.84**

Conv.- convolutional; f- number of filters per layer; n- number of neurons per layer; FC- fully; connected; u- units; B.C- binary crossentropy; AUC, ACC, PRE, REC, Loss - mean of metrics AUC, accuracy, precision, recall and loss over ten folds.

Finally, DCNN3 had a wide variety of interesting results. Taking into account the results of batch size 16, the model did a very poor performance. Within the four epochs number proposed, none of them obtained an AUC greater than 0.50, and the loss metric in every case was above 0.60, which translates into a level of confidence of 40% or lower, depicting more mistakes than hits. In addition, it presents overfitting in all epochs. Despite this, results with batch size 32 improved considerably in terms of AUC, increasing its value through epochs. Just with 200 epochs, the model achieved more than 0.90 in AUC. However, the loss function never decreased to 0.10 or lower, indicating that this model did not achieve the level of confidence wanted on this research. As mentioned, the best model was obtained with 200 epochs and 32 as batch size. This model got an AUC of 0.942 and a loss of 0.29 (see the bold line in Table I).

Figure 5 shows the ROC Curve, Precision vs Recall and Loss vs Epochs for the best model. It is important to mention that the first two curves (ROC and Precision-Recall) were created using the separated dataset explained in previous sections, not the mean of the folds, to validate if the results obtained during training were well done. The tendency of the curves demonstrate that, increasing the number of samples for predictions, the value obtained will be increasing until reaching a little less than 1, which is aligned to the results presented on Table 1 (mean of training). In terms of the loss function, this curve was built with the mean of the folds during training. The curve shows that, in each iteration, the loss value decreases and tries to reach the default learning rate of Adam optimizer, which es 0.001. This is an ideal behaviour for a deep learning model. The highest number in the y-scale of the mean loss plot is less than 4, which means a level of confidence greater than 96%. Moreover, the Precision-Recall and the ROC curves both are 90% or higher, proving DCNN1 as the best model.

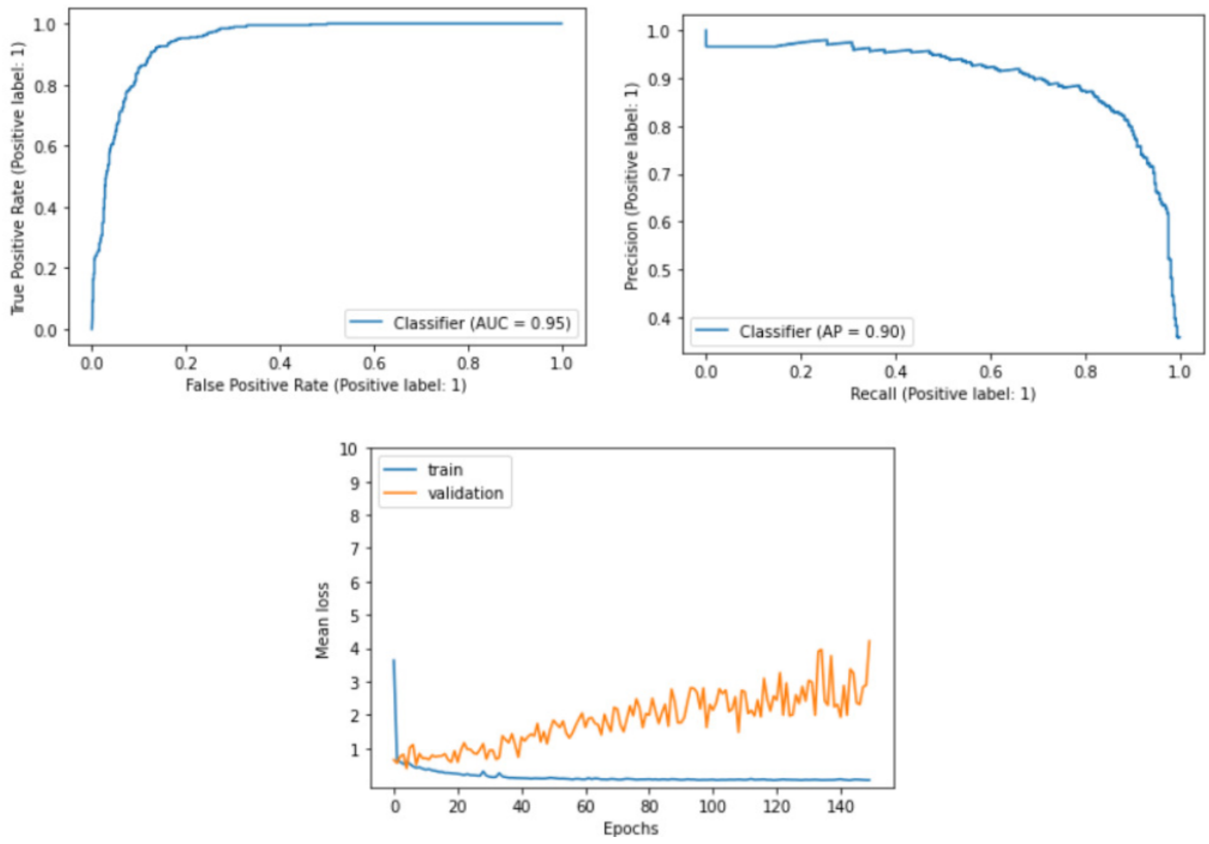

Figure 5: Performance of the best proposed deep learning model (DCNN1) on predictions testing based on the higher AUC (left), precision-recall (right) and the lower mean of the loss function (center) over ten folds.

CONCLUSIONS AND FUTURE WORK

In this research, we explored three different deep CNN models with a wide variety of epochs and hyperparameters. At the end, model 3 was the one with worst performance, taking into account that this architecture was deeper and had more filters in the convolutional layers than the others. Although the best alternative of this model achieved a good AUC, the loss metric was also very high, over 10%, which is something very risky in the medical world, where everything has to be as precise and exact as possible. Moreover, with this model we proved that deeper neural networks have higher probability of falling in overfitting if the dataset is not big enough.

In contrast, models 1 and 2 achieved very similar results on its best performance. In fact, the only difference is a hundredth in the loss metric, since model 1 obtained 0.04 and model 2 obtained 0.05, which are outstanding. However, DCNN1 used 50 less epochs to get those results, which in terms of computational cost and time spent, is a very important detail to consider. In addition, DCNN1 showed off as a more robust and solid model since it never fell on overfitting, something which a DCNN2 with small batch size could not avoid. Due to all the explanations above, we can conclude that model 1 was the best one among the models tested in this research.

Figure 6 shows some examples of original x-ray images, ground truth and predictions made by DCNN1. The figure demonstrate that the model has a very high accuracy measure when it comes to predict the location of caries in an image that it had not seen before. From the examples protrayed, just one of them predicted wrong a caries, while the other three examples were successfully detected according to the ground truth. The level of confidence in the detection of each caries is maintained over the 95%, which supports the results obtained during the training phase.

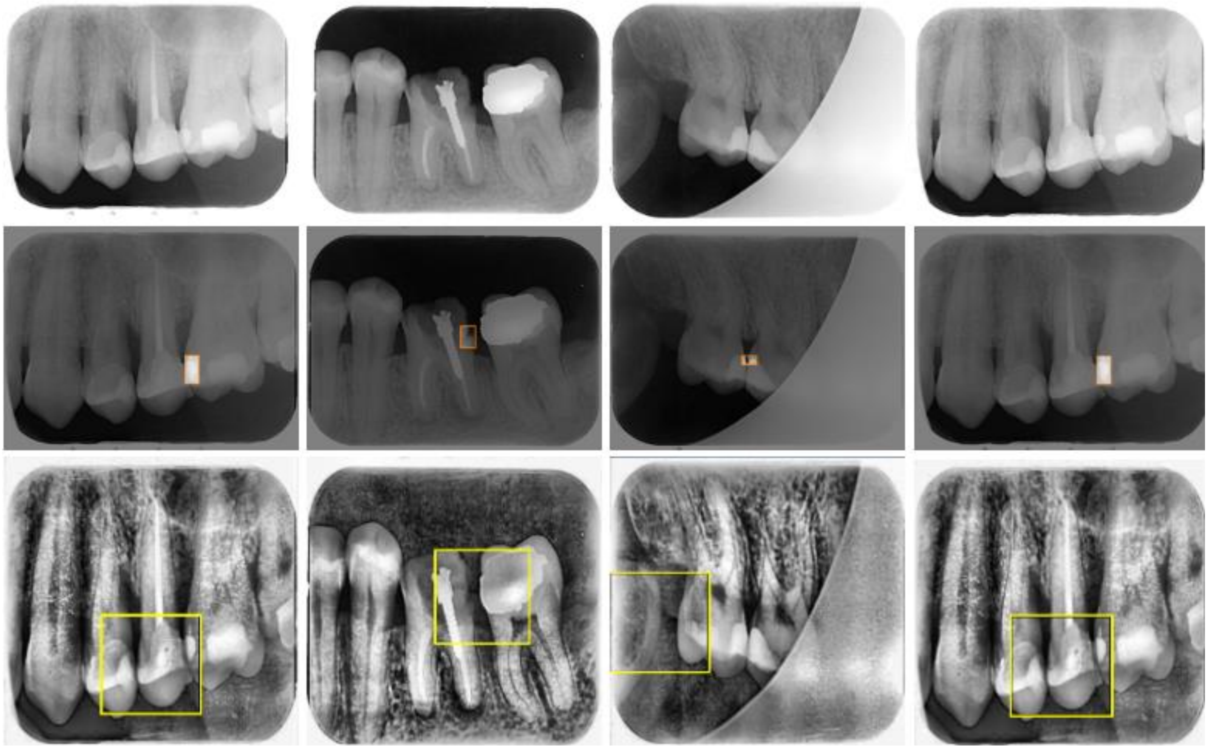

Figure 6: Some examples of the selected best model outputs: original images provided by the clinic (first row), ground truth images annotated by odontological experts identifying caries (second row) and caries detection on images predicted by DCNN1 (third row).

As future work, we plan to increase the samples in the dataset and train the model not only with ROIs of caries, but also with other types of lesions. With this, the problem will stop being only cataloged as binary classification. Instead, it will be a categorical classification to explore if the DCNN1 has the same results in another type of classification.

ACKNOWLEDGEMENT

Authors thank to Anahi Orellana and Johanna Monar for helping in the annotations of caries in the original image database, and to the Applied Signal Processing and Machine Learning Research Group of USFQ for providing the computing infrastructure (NVidia DGX workstation) to implement and execute the developed source code. Publication of this article was funded by the Academic Articles Publication Fund of Universidad San Francisco de Quito USFQ.

REFERENCES

Alake, R. (2020). *Implementing alexnet CNN architecture using tensorflow 2.0+ and keras*. Medium. Retrieved September 20, 2021, from https://towardsdatascience.com/implementing-alexnet-cnn-architecture-using-tensorflow-2-0-and-keras-2113e090ad98.

Birdal, Ramiz., Gumus, Ergun., Sertbaş, A., Sinem, I. (2015). *Automated lesion detection in panoramic dental radiographs*. Oral Radiology. 32. 10.1007/s11282-015-0222-8.

Braun, A., Guiraud, L., Frankenberger, R. (2016). *Histological validation of ICDAS II and radiological assessment of occlusal carious lesions in permanent teeth. Odontology*. 2017 Jan;105(1):46-53. doi: 10.1007/s10266-016-0245-6. Epub 2016 Apr 18. PMID: 27090647.

Brownlee, J. (2021). *How to choose an activation function for deep learning*. Machine Learning Mastery. Retrieved December 12, 2021, from https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/.

Carvalho, R., Letieri, A., Vieira, I., Santos, T. et al. (2018). *Accuracy of visual and image-based ICDAS criteria compared with a micro-CT gold standard for caries detection on occlusal surfaces*. Brazilian oral research, 32, e60. https://doi.org/10.1590/1807-3107bor-2018.vol32.0060

Datta, Soma., Chaki, Nabendu., Modak, Biswajit. (2020). *A Systematic Review on the Evolution of Dental Caries Detection Methods and Its Significance in Data Analysis Perspective*. 10.1002/9781119544487.ch6.

Diaa, M., Hany, Ammar. (2006). *A neural network system for matching dental radiographs*. Pattern Recognition, Volume 40, Issue 1, January 2007, Pages 65-79.

Great Learning Team. (2020). *AlexNet: The first CNN to Win Image Net: What is AlexNet?* AlexNet: The First CNN to win Image Net. Retrieved November 12, 2021, from https://www.mygreatlearning.com/blog/alexnet-the-first-cnn-to-win-image-net/.

Hentschke, H. (2019). *Sigmoid activation and binary crossentropy-a less than perfect match?* Medium. Retrieved December 12, 2021, from https://towardsdatascience.com/sigmoid-activation-and-binary-crossentropy-a-less-than-perfect-match-b801e130e31.

Igarashi, Y., Yoshida, S., Kanazawa, E. (2017). *The prevalence and morphological types of non-carious cervical lesions (NCCL) in a contemporary sample of people*. Odontology 105, 443–452. https://doi.org/10.1007/s10266-017-0300-y

Karakaya, M. (2020). *How to solve binary classification problems in deep learning with Tensorflow & Keras?* Medium. Retrieved December 12, 2021, from https://medium.com/deep-learning-with-keras/which-activation-loss-functions-part-a-e16f5ad6d82a.

Keras Team. (2021). *Keras Documentation: Model training apis*. Keras. Retrieved September 1, 2021, from https://keras.io/api/models/model_training_apis/.

Krizhevsky, A., Sutskever, I., Hinton, H. (2012). *Imagenet classification with deep convolutional neural networks*. [Online]. Available: https://proceedings.neurips.cc/paper/2012/file/ c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

Lee, J., Kim, D., Jeong, S., Choi, S. (2018). *Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm*. Journal of dentistry, 77, 106–111. https://doi.org/10.1016/j.jdent.2018.07.015

Li, X., Zhang, G., Li, K., Zheng, W. (2016). *Deep learning and its parallelization*. Big Data, 95–118. https://doi.org/10.1016/b978-0-12-805394-2.00004-0

Liebe-Harkort, C., Astvaldsdóttir, A., Tranæus, S,. (2009). *Quantification of Dental Caries by Osteologists and Odontologists - A Validity and Reliability Study*. International Journal of Osteoarchaeology. 20. 525 - 539. 10.1002/oa.1079.

Martignon, S., Uribe, S., Pulido, A., Cortés, A. et al (2013). *Comparación entre el examen radiográfico y el visual-táctil para detectar y valorar caries dental interproximal*. Dialnet. Retrieved November 12, 2021, from https://dialnet.unirioja.es/servlet/articulo?codigo=5026007.

Muralidhar, K. (2021, February 22). *What is stratified cross-validation in machine learning?* Medium. Retrieved December 12, 2021, from https://towardsdatascience.com/what-is-stratified-cross-validation-in-machine-learning-8844f3e7ae8e.

Navamani, T. (2019). *Efficient Deep Learning Approaches for Health Informatics*. Deep Learning and Parallel Computing Environment for Bioengineering Systems, 123–137. https://doi.org/10.1016/b978-0-12-816718-2.00014-2

Ohnishi, Y. et al. (2011). *Early detection of oral precancerous and early cancerous lesions by fluorescence visualization*. Journal of Osaka Dental University, 45(2), 215-219.

Oliveira, J., Proença, H. (2011). *Caries detection in panoramic dental X-ray images*. In Computational Vision and Medical Image Processing (pp. 175-190). Springer, Dordrecht.

Oprea, S., Marinescu, C., Lita, I., Jurianu, M. et al. (2008). *Image processing techniques used for dental x-ray image analysis*. 31st International Spring Seminar on Electronics Technology, 2008, pp. 125-129, doi: 10.1109/ISSE.2008.5276424.

Schwendicke, F., Golla, T., Dreher, M., Krois, J. (2019*). Convolutional neural networks for dental image diagnostics: A scoping review*. Journal of dentistry, 91, 103226. https://doi.org/10.1016/j.jdent.2019.103226

Shahbazian, M., Vandewoude, C., Wyatt, J. et al. (2013). *Comparative assessment of periapical radiography and CBCT imaging for radiodiagnostics in the posterior maxilla*. Odontology 103, 97–104. https://doi.org/10.1007/s10266-013-0144-z

Slimani, Amel., Terrer, Elodie. et al. (2020). *Carious lesion detection technologies: factual clinical approaches.* British Dental Journal. 229. 432-442. 10.1038/s41415-020-2116-3.

Wei, J. (2019). *Alexnet: The architecture that challenged CNNS*. Medium. Retrieved December 12, 2021, from https://towardsdatascience.com/alexnet-the-architecture-that-challenged-cnns-e406d5297951.

World Health Organization. (2020, March 25). *Oral Health*. World Health Organization. Retrieved November 12, 2021, from https://www.who.int/news-room/fact-sheets/detail/oral-health.

Yamashita, R., Nishio, M., Do, R. et al. (2018). *Convolutional neural networks: an overview and application in radiology*. Insights Imaging 9, 611–629. https://doi.org/10.1007/s13244-018-0639-9

Young, D. A., Nový, B. B., Zeller, G. G. et al. (2015). *The American Dental Association Caries Classification System for Clinical Practice*. The Journal of the American Dental Association, 146(2), 79–86. https://doi.org/10.1016/j.adaj.2014.11.018