

XÂY DỰNG HỆ DỮ LIỆU HỖ TRỢ MÁY HỌC ĐỂ CHẨN ĐOÁN SÂNG RĂNG SÂU RĂNG GIAI ĐOẠN SỚM

Mai Thị Giang Thanh^{1,✉}, Võ Trương Như Ngọc², Ngô Văn Toàn²
Lương Minh Hằng², Trần Ngọc Phương Thảo³

¹Trường Cao đẳng Y tế Hà Đông

²Trường Đại học Y Hà Nội

³Trường Đại học Văn Lang, Hồ Chí Minh

Nghiên cứu nhằm xây dựng hệ dữ liệu hỗ trợ máy học để chẩn đoán sàng lọc sâu răng giai đoạn sớm ở trẻ 3 - 6 tuổi trên một số cơ sở thuộc địa bàn thành phố Hà Nội. Kết quả nghiên cứu cho thấy: Số lượng ảnh tổn thương sâu răng giai đoạn sớm trong kho dữ liệu của nghiên cứu này là 478 ảnh. Có sự đa dạng về số lượng và vị trí tổn thương sâu răng giai đoạn sớm ở cả 5 góc chụp ảnh trong miệng như ảnh toàn cảnh, ảnh bên phải, ảnh bên trái, ảnh hàm trên, ảnh hàm dưới lần lượt là: 505 răng, 362 răng, 363 răng, 50 răng và 90 răng. Kho dữ liệu được xây dựng có sự phân bố của các tổn thương sâu răng giai đoạn sớm tập trung nhiều ở ảnh vùng mặt ngoài của răng với 994 tổn thương, mặt nhai là 65 tổn thương và mặt bên chỉ là 14 tổn thương.

Từ khóa: Kho dữ liệu, sâu răng giai đoạn sớm, trẻ 3-6 tuổi.

I. ĐẶT VẤN ĐỀ

Ngày nay với sự phát triển mạnh mẽ của công nghệ 4.0 trên mọi lĩnh vực trong đời sống và xã hội, các ứng dụng công nghệ giúp con người tiết kiệm thời gian, công sức và đạt được hiệu quả cao hơn trong công việc. Trong nha khoa với mong muốn hỗ trợ các bác sĩ và bệnh nhân trong chẩn đoán bệnh các nhà nghiên cứu đã sử dụng công cụ học máy để hỗ trợ chẩn đoán tự động các bệnh lý răng miệng dựa trên phim x-quang như nghiên cứu của Srivastava,¹ Vinayahalingam,²... Tại Việt Nam, tác giả Võ Trương Như Ngọc và cộng sự (2020)³ cũng đóng góp nghiên cứu về phương pháp chẩn đoán tự động biến chứng răng khôn dựa trên phim x-quang.

Sâu răng giai đoạn sớm là giai đoạn tổn thương không có lỗ sâu. Các tổn thương được đặc trưng bởi sự thay đổi về màu sắc, độ bóng và cấu trúc bề mặt men răng. Điều trị sâu răng giai đoạn sớm bằng phương pháp tái khoáng hóa thì cấu trúc men răng có thể phục hồi hoàn toàn.⁴ Theo nghiên cứu Pitts 2013⁵ sâu răng giai đoạn sớm là giai đoạn đầu tiên phát hiện thấy sự thay đổi bề mặt men răng như đổi màu với đốm trắng hoặc nâu và chưa có biểu hiện phá hủy bề mặt men hay vùng tối ở lớp ngà răng. Để chẩn đoán sâu răng giai đoạn sớm có một số phương pháp thường được sử dụng trong lâm sàng như: khám trực tiếp, X- quang, phương pháp laser huỳnh quang, định lượng ánh sáng huỳnh quang.⁶ Năm 2012, tác giả Boye⁷ đã chứng minh hiệu quả chẩn đoán sâu răng qua ảnh chụp bằng iPhone. Tác giả Kohara (2018)⁸ sử dụng hình ảnh smartphone để chẩn đoán các giai đoạn sâu răng ở mặt nhai trên trẻ 3 - 6 tuổi. Tại Việt Nam, năm 2018 tác giả Lê Thị

Tác giả liên hệ: Mai Thị Giang Thanh,

Trường Cao đẳng Y tế Hà Đông

Email: maithigiangthanh@gmail.com

Ngày nhận: 20/10/2020

Ngày được chấp nhận: 28/12/2020

Thùy Linh⁹ đã sử dụng iPhone 6 để chụp ảnh tổn thương sâu răng trên trẻ 8 tuổi cho thấy với tổn thương sâu răng giai đoạn sớm có độ nhạy của phương pháp chụp ảnh là 72,62% và độ đặc hiệu là 97,88%. Trong nghiên cứu này tác giả cũng đưa ra kiến nghị các bác sĩ răng hàm mặt có thể sử dụng ảnh chụp bằng smartphone để chẩn đoán sâu răng từ xa vì thao tác đơn giản, hiệu quả, độ nhạy của chẩn đoán đạt tiêu chuẩn.

Qua khảo sát các nghiên cứu trên thế giới có thể thấy rằng, để chẩn đoán tự động với sự hỗ trợ của công cụ học máy thì bước đầu phải nghiên cứu và xây dựng kho dữ liệu đào tạo học máy. Hiện nay, các nghiên cứu quan tâm nhiều đến vấn đề xây dựng kho dữ liệu cho học máy. Một trong số đó là xây dựng kho dữ liệu cho bệnh sâu răng để hỗ trợ chẩn đoán tự động trên phim x-quang.¹ Tuy nhiên chưa có nhiều nghiên cứu quan tâm đến vấn đề xây dựng kho dữ liệu sâu răng giai đoạn sớm qua ảnh chụp tại Việt Nam. Tiếp theo dòng nghiên cứu về sâu răng giai đoạn sớm tại Việt Nam, mục tiêu của nghiên cứu này là xây dựng kho dữ liệu sâu răng giai đoạn sớm để đào tạo học máy nhằm hỗ trợ trong quá trình chẩn đoán của bác sĩ.

II. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP

1. Đối tượng

Ảnh chụp điện thoại iPhone 7 của tổn thương sâu răng giai đoạn sớm trên trẻ 3 - 6 tuổi.

Tiêu chuẩn lựa chọn: Ảnh được bác sĩ chẩn đoán có tổn thương sâu răng giai đoạn sớm trên khám lâm sàng và ảnh thấy rõ các chi tiết tổn thương sâu răng giai đoạn sớm khi bác sĩ khám trên ảnh: thay đổi màu sắc như đốm trắng, đốm vàng nâu trên bề mặt men răng, bề mặt men răng không có lỗ sâu. Tổn thương nằm ở khoảng giữa của ảnh.

Tiêu chuẩn loại trừ: Ảnh của tổn thương sâu răng giai đoạn giữa và giai đoạn muộn: có đổi

màu ngà răng, lỗ sâu rõ ràng lộ ngà răng.

2. Phương pháp

Thời gian nghiên cứu: Từ 07/2019 đến tháng 05/2020.

Địa điểm nghiên cứu: Nghiên cứu được thực hiện tại Trung tâm Khám chữa bệnh kỹ thuật cao nhà A7 - Viện đào tạo Răng Hàm Mặt - Trường Đại học Y Hà Nội, Khoa Răng Hàm Mặt - Bệnh viện Việt Nam CuBa và một số địa điểm khác ở Hà Nội: Trường mầm non Hàng Đào và trường mầm non Mậu Lương - Hà Đông - Hà Nội.

Cỡ mẫu: Cỡ mẫu nghiên cứu sử dụng công thức để đánh giá độ nhạy và độ đặc hiệu của một phương pháp chẩn đoán bệnh:

$$n_{se} = \left(\frac{Z_{\alpha}^2 \times p_{se}(1 - p_{se})}{\omega^2} \right) / p_{dis}$$

Trong đó:

n_{se} : cỡ mẫu nghiên cứu cho độ nhạy,

Z_{α}^2 là hằng số của phân phối chuẩn (chọn $\alpha = 0,05$ thì Z_{α} bằng 1,96),

p_{se} là xác suất dương tính thật (độ nhạy) (ước tính là 80%), ((độ nhạy trong nghiên cứu Boye 2012 là 81,3%)⁷

ω là sai số của hai xác suất dương tính thật (độ nhạy) và âm tính thật (độ đặc hiệu), lấy $\omega = 0,05$.

p_{dis} là tỷ lệ hiện mắc bệnh trong quần thể.

Lấy $p_{dis} = 0,645$ (theo nghiên cứu của Lưu Văn Tường,¹⁰ với $N = 1203$).

$$n_{se} = \left(\frac{1,96^2 \times 0,8(1 - 0,8)}{0,05^2} \right) / 0,645 = 381$$

Lấy cỡ mẫu nghiên cứu là 400 trường hợp. Kho dữ liệu sẽ có tối thiểu 400 ảnh có tổn thương sâu răng giai đoạn sớm với độ phân giải tối thiểu Full HD 1920 x 1080 pixel.

Thực tế trong nghiên cứu này chúng tôi đã xây dựng được kho dữ liệu với 478 ảnh có tổn thương sâu răng giai đoạn sớm với độ phân giải 4032 x 3024 pixel. Ảnh được chụp bằng điện thoại iPhone 7. 478 ảnh này đã được bác

sĩ chẩn đoán qua khám lâm sàng có răng tổn thương sâu răng giai đoạn sớm và bác sĩ khám trên ảnh cũng kết luận có răng tổn thương sâu răng giai đoạn sớm.

Lập phiếu thu thập thông tin

Phiếu thu thập thông tin được lập theo mẫu thiết kế sẵn gồm các phần sau: Hành chính, bệnh sử, tiền sử, khám lâm sàng, chụp ảnh

Nội dung, chỉ số nghiên cứu:

Đặc điểm ảnh: Góc toàn cảnh, Góc bên trái, Góc bên phải, Góc hàm trên, Góc hàm dưới.

Đặc điểm vị trí tổn thương sâu răng giai đoạn sớm trên răng: Mặt ngoài, mặt trong, mặt gần, mặt xa, mặt nhai .

Quy trình xây dựng hệ dữ liệu cho học máy

Quá trình thu thập dữ liệu:

Quy trình xây dựng hệ dữ liệu sâu răng giai đoạn sớm được thiết lập dựa theo nghiên cứu của tác giả Roh và cộng sự.¹¹ Quy trình này gồm 6 bước như sau:

Bước 1: Khảo sát tỷ lệ mắc bệnh

Theo nghiên cứu của Lưu Văn Tường¹⁰ trên 1203 trẻ 3 tuổi tại 4 trường Mầm non- Thành phố Hà Nội: Tỷ lệ sâu răng sữa giai đoạn sớm là 64,5%.

Bước 2: Phân tích và đánh giá

Yêu cầu tối thiểu về chất lượng hình ảnh hỗ trợ chẩn đoán qua ảnh:

Trong nghiên cứu của K. T. Ashique và các cộng sự¹² đã thực hiện việc so sánh chất lượng ảnh chụp sử dụng các loại điện thoại thông minh so sánh với máy ảnh kỹ thuật số thông thường. Để xác định chất lượng ảnh tối thiểu phục vụ chẩn đoán bệnh răng miệng thì ảnh chất lượng phù hợp là những ảnh được xác nhận tiêu chuẩn ảnh tối thiểu Full HD 1920 x 1080 pixel.

Phương thức truyền dữ liệu hình ảnh giữ được chất lượng ảnh gốc:

Tuân thủ theo thông tư số: 53/2014/TT- BYT của Bộ Y Tế, DICOM được sử dụng trong hệ

thống hỗ trợ chẩn đoán và hội chẩn từ xa như giao thức trao đổi, lưu trữ, xử lý, thu hoặc nhận, in ấn và chia sẻ dữ liệu hình ảnh số giữa các thiết bị.

Trong hệ thống hỗ trợ chẩn đoán từ xa trong nha khoa, các ảnh được thu thập bằng điện thoại để phục vụ quá trình chẩn đoán sẽ được chuyển về định dạng JPEG 2000, sau đó chuyển sang định dạng DICOM (dcm) để gửi nhận và lưu trữ trong server dicom. Giao thức này đảm bảo chất lượng ảnh được giữ nguyên như ảnh gốc trên server.

Bước 3: Thu thập dữ liệu

Trong quá trình thu thập dữ liệu, nghiên cứu này đã sử dụng thiết bị di động iPhone 7 có độ phân giải 12 megapixel và chỉ sử dụng những phần mềm chụp ảnh trên thiết bị di động giữ nguyên yếu tố thực tại của ảnh và không có yếu tố chỉnh sửa ảnh tự động. Ảnh chụp được lưu dưới định dạng JPEG 2000.

Mỗi đối tượng được chụp 5 ảnh tương ứng với 5 góc chụp ảnh trong miệng: Góc toàn cảnh, góc bên trái, góc bên phải, góc hàm trên, góc hàm dưới. Tư thế chụp ảnh tuân theo hướng dẫn chụp ảnh trong miệng bằng smartphone có sử dụng băng miệng.¹³

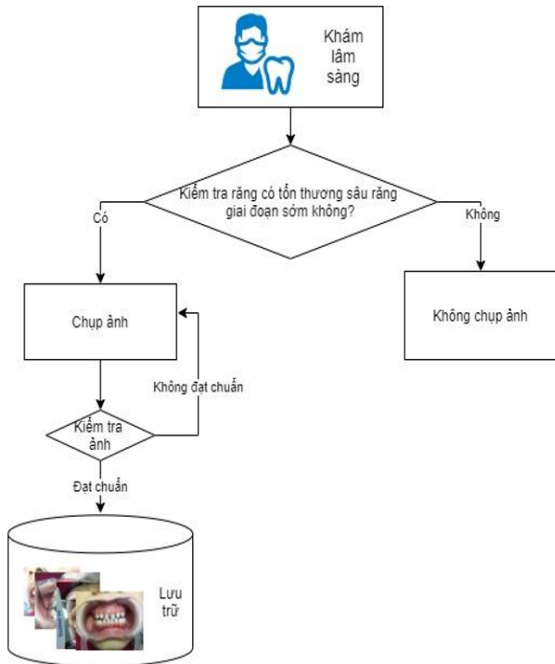
Bước 4: Chuẩn hóa dữ liệu

Sàng lọc 750 ảnh thu được 478 ảnh có tổn thương sâu răng giai đoạn sớm trên ảnh. Trong đó có 272 ảnh không đáp ứng được các tiêu chí ảnh kho dữ liệu học máy. Loại bỏ các dữ liệu không hợp lệ bao gồm: ảnh mờ, ảnh nhiễu, ảnh không rõ tổn thương sâu răng giai đoạn sớm.

Bước 5: Mã hóa dữ liệu

Các file dữ liệu hình ảnh tổn thương của đối tượng nghiên cứu được đặt tên theo mã đối tượng và số thứ tự ảnh: <mã đối tượng>_<số thứ tự>.

Sơ đồ quy trình thu thập dữ liệu:



Dữ liệu hình ảnh của các đối tượng sẽ được tạo thành hai bản sao. Bản sao thứ nhất là tệp ảnh gốc các bức ảnh được giữ nguyên, bản sao còn lại các bác sĩ sẽ phân loại và phân vùng vùng răng bị tổn thương trên từng bức ảnh chụp hàm răng mỗi đối tượng nghiên cứu.

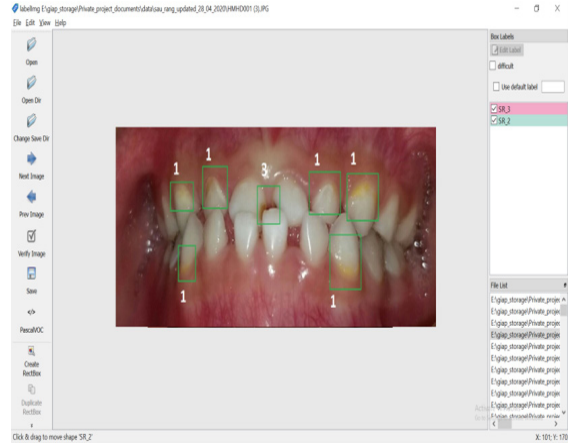
Quá trình xử lý dữ liệu cho học máy

Quá trình xử lý dữ liệu cho học máy gồm các bước chính sau:¹¹

Bước 1: Kiểm tra và chuẩn hóa dữ liệu sâu răng giai đoạn sớm đã được thu thập cho học máy. Loại bỏ những ảnh không đạt chuẩn trong học máy bao gồm: ảnh nhiễu, ảnh mờ, ảnh không rõ tổn thương, ảnh không đúng định dạng, kích thước ảnh không đạt chuẩn, ...

Bước 2: Gán nhãn tổn thương cho học máy.

Những ảnh sâu răng giai đoạn sớm đã đạt chuẩn được các bác sĩ sử dụng phần mềm gán nhãn tổn thương Labeling để gán nhãn tổn thương trên ảnh cho học máy. Kết quả thu được sau khi sử dụng phần mềm gán nhãn bao gồm: Ảnh gốc và file XML (đây là định dạng mã hóa và lưu trữ khoanh vùng tổn thương trên ảnh)



Hình 2. Phần mềm gán nhãn tổn thương trên ảnh

Bước 3: Tổ chức cấu trúc và lưu dữ liệu đã gán nhãn cho học máy.

Dữ liệu đã gán nhãn được tổ chức theo định dạng PASCAL_VOC (đây là định dạng yêu cầu cho đầu vào của phương pháp học máy). Kho dữ liệu được lưu trữ trong một thư mục gồm 3 thư mục con:

- JPEGImages: Thư mục chứa ảnh gốc của ảnh tổn thương sâu răng giai đoạn sớm
- Annotations: Chứa toàn bộ các file XML, được mã hóa và lưu trữ cho vùng gán nhãn tổn thương trên ảnh.
- Imagesets: Gồm 2 file “train.txt” và “val.txt”.

Đây là hai file phân chia dữ liệu để thực hiện quá trình huấn luyện và kiểm tra cho phương pháp học máy.

3. Xử lý số liệu

Số liệu được nhập và quản lý bằng phần mềm Microsoft Excel 2010 cho thông tin hồ sơ bệnh nhân. Số liệu được xử lý bằng lập trình SPSS 20.0.

4. Đạo đức nghiên cứu

Lựa chọn đối tượng nghiên cứu đúng với tiêu chuẩn đề ra. Các quy trình khám lấy số liệu đảm bảo vô khuẩn, không ảnh hưởng tới sức khỏe của đối tượng nghiên cứu. Các thông tin

thu được phải đảm bảo bí mật và chỉ dùng cho mục đích nghiên cứu. Bổ sung bảng thông tin nghiên cứu và có bản cam kết tham gia nghiên cứu. Đề tài đã được chấp thuận bởi hội đồng đạo đức của Trường đại học Y Hà Nội theo quyết định số NCS21 /ĐHYHN- HĐĐĐ ngày 14 tháng 2 năm 2019 và đề tài tuân thủ các quy trình, quy định đã được ban hành.

III. KẾT QUẢ

Đặc điểm trên ảnh của kho dữ liệu sâu răng giai đoạn sớm

Bảng 1. Tỷ lệ ảnh có tổn thương sâu răng giai đoạn sớm

Tỷ lệ \ Góc chụp	Góc chụp					
	Toàn cảnh	Bên trái	Bên phải	Hàm dưới	Hàm trên	Tổng
N	151	125	135	26	41	478
%	31,6%	26,2%	28,2%	5,4%	8,6%	100%

Kho dữ liệu có hình ảnh tổn thương sâu răng giai đoạn sớm trên ảnh toàn cảnh là cao nhất với 151 ảnh (31,6%) và thấp nhất trên ảnh hàm dưới với 26 ảnh (5,4%).

Bảng 2. Tỷ lệ răng có tổn thương sâu răng giai đoạn sớm

Tỷ lệ \ Góc chụp	Góc chụp					
	Toàn cảnh	Bên trái	Bên phải	Hàm dưới	Hàm trên	Tổng
N	505	362	363	50	90	1370
%	36,9%	26,4%	26,5%	3,6%	6,6%	100%

Tổng số răng tổn thương sâu răng giai đoạn sớm là 1370 tổn thương. Trong đó ảnh chụp mặt ngoài của cung răng: ảnh toàn cảnh, bên trái, bên phải có số tổn thương cao hơn ảnh chụp mặt nhai hàm trên và hàm dưới.

Bảng 3. Tỷ lệ mặt răng tổn thương của góc chụp mặt ngoài cung răng

Mặt răng \ Góc chụp	Góc chụp			Tổng	
	Toàn cảnh	Bên trái	Bên phải	N	%
Mặt ngoài	323	330	341	994	80,8%
Mặt gần	129	16	11	156	12,7%
Mặt xa	53	16	11	80	6,5%
Tổng	505	362	363	1230	100%

Tổn thương sâu răng giai đoạn sớm ở mặt ngoài của răng là cao hơn mặt xa và mặt gần với 994 tổn thương

Bảng 4. Tỷ lệ mặt răng tổn thương của góc chụp mặt nhai cung răng

Mặt răng \ Góc chụp	Hàm dưới	Hàm trên	Tổng	
			N	%
Mặt nhai	31	34	65	46,4
Mặt gần	5	3	8	5,7
Mặt xa	4	2	6	4,3
Mặt trong	10	51	61	43,6
Tổng	50	90	140	100

Tổn thương sâu răng giai đoạn sớm ở mặt nhai của răng là cao hơn mặt gần, mặt xa, mặt trong với 65 tổn thương.

IV. BÀN LUẬN

Kết quả nghiên cứu của chúng tôi cho thấy, hệ dữ liệu của nghiên cứu này được xây dựng với 478 ảnh tổn thương sâu răng giai đoạn sớm. Theo nghiên cứu của tác giả Berdouses và cộng sự¹⁴ năm 2015 để phát hiện và phân loại sâu răng trên mặt nhai bằng công cụ học máy đã tiến hành nghiên cứu và đánh giá kết quả học máy chỉ trên 103 ảnh. Do đó cho thấy, kho dữ liệu của nghiên cứu đang thực sự chiếm ưu thế so với kho dữ liệu của nhóm tác giả Berdouses. Năm 2017 tác giả Rana¹⁵ dùng học máy như một công cụ để phân đoạn tự động các bệnh về lợi từ các hình ảnh trong miệng cũng chỉ xây dựng cơ sở dữ liệu với 258 ảnh phục vụ cho kho dữ liệu. Khi nghiên cứu tổng quan về học máy trong lĩnh vực nha khoa tác giả Hwang¹⁶ từ năm 2016 đến tháng 12 năm 2018 quy mô trung bình của kho dữ liệu đào tạo cho học máy có xu hướng ngày càng tăng. Điều này được giải thích vì khi cơ sở dữ liệu để đào tạo học máy càng nhiều thì sự chính xác trong chuẩn đoán càng cao, do đó cỡ mẫu trong kho dữ liệu của chúng tôi phù hợp với xu thế phát triển lĩnh vực trí tuệ nhân tạo. Bên cạnh đó trong kho dữ liệu của nghiên cứu có một sự đa dạng về vị trí của tổn thương theo cả 5 góc chụp trong miệng như góc toàn cảnh,

bên phải, bên trái, hàm trên và hàm dưới. Sự đa dạng này cũng thể hiện trên cả 5 mặt của răng như mặt ngoài, mặt trong, mặt gần, mặt xa và mặt nhai. Đi cùng với sự đa dạng về vị trí sẽ là sự đa dạng về hình ảnh của tổn thương sâu răng giai đoạn sớm, yếu tố này là yếu tố quan trọng quyết định chất lượng của học máy. Tuy nhiên, kho dữ liệu này không tránh được một nhược điểm là sự phân bố của các tổn thương sâu răng giai đoạn sớm sẽ tập trung nhiều ở ảnh chụp vùng mặt ngoài của răng trên cung hàm với 994 tổn thương, mặt nhai là 65 tổn thương và mặt bên chỉ 14 tổn thương. Nhưng kết quả tổn thương trên ảnh hoàn toàn phản ánh đúng thực tế lâm sàng khi các bác sĩ khám răng trực tiếp trên bệnh nhân thì tổn thương sâu răng giai đoạn sớm ở mặt nhai thường được phát hiện nhiều hơn. Như vậy, kho dữ liệu trong nghiên cứu này hoàn toàn phù hợp với xu thế phát triển công nghệ và đáp ứng được các yêu cầu về một kho dữ liệu hỗ trợ máy học để chẩn đoán sâu răng giai đoạn sớm.

V. KẾT LUẬN

Qua nghiên cứu xây dựng hệ dữ liệu hỗ trợ máy học để chẩn đoán sàng lọc sâu răng giai đoạn sớm trên trẻ 3 - 6 tuổi tại Hà Nội đã xây dựng được kho dữ liệu với 478 ảnh có tổn thương sâu răng giai đoạn sớm. Nghiên cứu

cũng đã đưa ra quy trình xây dựng hệ dữ liệu hỗ trợ máy học cần thực hiện đầy đủ 6 bước: Khảo sát tỷ lệ mắc bệnh, phân tích và đánh giá, thu thập dữ liệu, chuẩn hóa dữ liệu, mã hóa dữ liệu, lưu trữ dữ liệu. Kho dữ liệu sâu răng giai đoạn sớm của nghiên cứu này đáp ứng được yêu cầu về số lượng, tính đa dạng của tổn thương và chất lượng của hình ảnh. Tuy nhiên cần tiếp tục thu thập bộ dữ liệu phong phú hơn để huấn luyện cho mô hình học máy.

Lời cảm ơn

Chúng tôi xin chân thành cảm ơn sự hỗ trợ, giúp đỡ của Viện đào tạo Răng Hàm Mặt, bệnh viện hữu nghị Việt Nam CuBa, hai trường mầm non Hàng Đào và Mậu Lương, quận Hà Đông, thành phố Hà Nội.

TÀI LIỆU THAM KHẢO

1. Srivastava MM, Kumar P, Pradhan L, Varadarajan S . Detection of tooth caries in bitewing radiographs using deep learning. *arXiv preprint arXiv*: 2017; 1711. 07312.
2. Vinayahalingam S, Xi T, Bergé S, et al. Automated detection of third molars and mandibular nerve by deep learning. *Scientific reports*. 2019; 9(1), 1 - 7.
3. Ngoc VTN, Agwu AC, Son LH, et al. The Combination of Adaptive Convolutional Neural Network and Bag of Visual Words in Automatic Diagnosis of Third Molar Complications on Dental X-Ray Images. *Diagnostics*. 2020; 10(4), 209.
4. Young DA, Nový BB, Zeller GG, et al. The American Dental Association caries classification system for clinical practice: a report of the American Dental Association Council on Scientific Affairs. *The Journal of the American Dental Association*. 2015; 146(2), 79 - 86.
5. Pitts NB, Ekstrand KR & ICDAS Foundation. International Caries Detection and Assessment System (ICDAS) and its International Caries Classification and Management System (ICCMS)—methods for staging of the caries process and enabling dentists to manage caries. *Community dentistry and oral epidemiology*. 2013; 41(1), e41 - e52.
6. Gomez J, Tellez M, Pretty IA, et al. Non-cavitated carious lesions detection methods: a systematic review. *Community Dentistry and Oral Epidemiology*. 2013; 41(1), 55 - 66.
7. Boye U, Walsh T, Pretty IA, Tickle M. Comparison of photographic and visual assessment of occlusal caries with histology as the reference standard. *BMC Oral Health*. 2012; 12(1), 10.
8. Kohara EK, Abdala CG, Novaes TF, et al. Is it feasible to use smartphone images to perform telediagnosis of different stages of occlusal caries lesions? *PloS one*. 2018; 13(9), e0202116.
9. Nguyễn Thị Thùy Linh. Thực trạng bệnh sâu răng của học sinh 8 tuổi tại trường tiểu học Kim Liên, Đống Đa, Hà Nội qua khám lâm sàng và chụp ảnh bằng smartphone. *Đại Học Y Hà Nội*, 2018: p. 44.
10. Lưu Văn Tường., Thực trạng bệnh sâu răng sữa sớm và một số yếu tố liên quan với sâu răng với học sinh mầm non 03 tuổi tại Hà Nội. *Tạp chí Y dược học Việt Nam*. 2019; 480 (1 & 2): p. 94-99.
11. Roh Y, Heo G, Whang SE. A survey on data collection for machine learning: a big data-ai integration perspective. *IEEE Transactions on Knowledge and Data Engineering*. 2019
12. Ashique KT, Kaliyadan F. Clinical photography for trichology practice: Tips and tricks. *International journal of trichology*. 2011; 3(1), 7.
13. Võ Trương Như Ngọc. Chụp ảnh nha khoa bằng điện thoại di động thông minh in Ghi

hình răng hàm mặt. 2018; p. 135 - 143.

14. Berdouses ED, Koutsouri GD, Tripoliti EE, et al. A computer-aided automated methodology for the detection and classification of occlusal caries from photographic color images. *Computers in biology and medicine*. 2015; 62, 119 - 135.

15. Rana, A., Yauney, G., Wong, L. C.,

Gupta, O., Muftu, A., & Shah, P. Automated segmentation of gingival diseases from oral images. *IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)*. 2017; (pp. 144-147).

16. Hwang JJ, Jung YH, Cho BH, et al. An overview of deep learning in the field of dentistry. *Imaging science in dentistry*. 2019; 49(1), 1 - 7.

Summary

BUILDING A DATA SYSTEM TO SUPPORT DEEP LEARNING FOR DIAGNOSIS OF EARLY- STAGE DENTAL CARIES

The purpose of this study is to build a data system to support deep learning for screening diagnosis of early - stage dental caries of 3 to 6 – years - old on various facilities in Hanoi. The results of the study illustrate that: The number of images with early - stage caries in the database of this study is 478. There is a variety in the number and locations of early - stage cavities in all 5 angles of oral imaging; panorama, right side, left side, upper jaw, lower jaw account for 505 teeth, 362 teeth, 363 teeth, 50 teeth and 90 teeth, respectively. In conclusion, the constructed database shows that the distribution of early caries lesions concentrated on the outer surface of the tooth with 994 lesions; there were 65 lesions on the occlusal surface and only 14 on the lateral surface.

Keywords: Database, early-stages dental caries, 3 to 6 year-old children.