Multimodal Oral Lesion Classification

BY

Pun Dissorn 64011587

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE BIOMEDICAL ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY
LADKRABANG
ACADEMIC YEAR 2023

SCHOOL OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

PROJECT CERTIFICATE

Project Title	Multimodal Oral Lesion Classification
Student Name	Mr. Pun Dissorn Student ID. 64011587
Degree	Bachelor of Engineering in Biomedical Engineering
Project Advisor	Signed: ____________________
	(Asst. Prof. Dr. TreesukonTreebupachatsakul)
Project Co-Advisor	Signed: ____________________
	(Asst.Prof. Dr. Suvit Poomrittigul)
Project Co-Advisor	Signed: ____________________
	(Assoc.Prof. Dr. Siribang-on Piboonniyom Khovidhunkit)
Committee	Signed: ____________________
	(Prof. Dr. Chuchart Pintavirooj)
Committee	Signed: ____________________
	(Asst. Prof. Dr. Kasama Srirussamee)
Committee	Signed: ____________________
	(Assoc. Prof. Wibool Piyawattanametha)
Head of Department	Signed: ____________________
	(Assoc. Prof. Dr. Sarinporn Visitsattapongse)

Project Title	Multimodal Oral Lesion Classification
Student Name	Mr. Pun Dissorn
Degree	Bachelor of Engineering in Biomedical Engineering
Project Advisor	Asst. Prof. Dr. Treesukon Treebupachatsakul
Project Co-Advisor	Asst.Prof. Dr. Suvit Poomrittigul
Project Co-Advisor	Assoc.Prof. Dr. Siribang-on Piboonniyom Khovidhunkit
Academic Years	2023

ABSTRACT

This research aims to improve the classification of oral lesions using different types of deep learning processes. Some oral lesions are a pre-cancerous state that could develop into oral cancer, early detection of pre-cancerous lesions is needed before they progress to cancer. Some oral lesions have a white color, known as white lesions. The classification between pre-cancerous and non-pre-cancerous of white lesions can be challenging due to their similarities. Early detection of pre-cancerous lesions is needed before they progress to cancer. The pre-cancerous of white lesions include oral lichen planus and leukoplakia. This research proposes two approaches for oral lesion classification. The first part uses pre-processing to segment interferences, utilizing the U-net model with different backbones such as EfficientNet B3, SE-ResNet 18, ResNet 18, MobileNet V2, Inception V3, ResNeXt 50, DenseNet 121, VGG 19, and SE-ResNeXt 50. The pre-processing model's performance is tested using DenseNet 121. The second part introduces a Visual Question Answering (VQA) model that implements medical history to train the model. The dataset for classification contains five classes: leukoplakia, lichen planus, candidiasis, other white, and ulcer. The results of EfficientUNet and DenseUNet achieved the highest validation and Intersection over Union (IoU) scores at 92% and 98%, respectively. The proposed approach segmented interferences from the images, demonstrating the success of these models in the pre-processing phase. Subsequently, a CNN model using DenseNet 121 was used for

classification, achieving accuracy a training of 99.1%, with validation and test reaching 86.1% and 75.5%, respectively. In the VQA part, implementing medical history resulted in an improvement in test accuracies, reaching 85.9%.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Asst. Prof. Dr. Treesukon Treebupachatsakul and my co-advisors Asst. Prof. Dr. Suvit Poomrittigul and Assoc. Prof. Dr. Siribang-on Piboonniyom Khovidhunkit, who have consistently supported and provided valuable guidance throughout the completion of this work. Their suggestions and guidelines have been instrumental in ensuring the success of each stage of the project. Additionally, I extend my thanks to my family and friends for their support throughout this entire journey.

Pun Dissorn

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

Symbols/Abbreviations	Terms
VQA	Visual Question Answer
CNNs	Convolutional Neural Networks
IoU	Intersection over Union
RNN	Recurrent Neural Networks
LSTM	Long Short-Term Memory

CHAPTER 1
INTRODUCTION

1.1 Oral Lesion

White oral lesions are categorized into two types: pre-cancerous and non-pre-cancerous. The pre-cancerous is the stage before the lesion develop to cancer. The detection of oral pre-cancerous stage can reduce the risk of cancer development of patient. Oral cancer is often associated with risky behaviors such as tobacco smoking, alcohol consumption, and betel nut chewing. Early detection of precancerous white lesions such as leukoplakia and lichen planus are crucial to facilitate timely treatment and prevent further progression [1]. Distinguishing between pre-cancerous types presents a challenge due to their appearance and coloration are not various. Non-pre-cancerous lesions, including candidiasis, ulcers, and linear alba. Traditional diagnostic methods for oral cancer, involving clinical pathology and biopsy procedures, which are not only time-consuming but may also cause discomfort. Moreover, the diagnosis of pre-cancerous requires the expert who has an experience and expertis in oral cancer.

1.2 CNN, U-net and VQA

Medical diagnostics have been improved due to the application of deep learning models, particularly Convolutional Neural Networks (CNNs) [2]. These models have demonstrated their capability to identify and categorize various medical conditions, including distinct types of oral lesions [3]. CNNs operate by dividing an image into smaller parts, extracting essential features from each element, and merging these features then make accurate predictions regarding the image [4]. Nevertheless, despite this improvement, the results achieved in terms of detection and classification are still not good for medical professionals [2].

Therefore, this project proposes to applied deep learning technology to classify the multi-class of oral white lesion images. In the first part, we use the U-Net model for preprocessing before classification and in the second part, our research has transitioned to the Visual Question Answer (VQA) model coupled with patient medical history data for classification purposes. This approach gives the unique advantages of both multimodal analysis and deep learning to get higher performance of oral lesion classification model and improving the traditional, time-consuming clinical pathology approach. This approach aims to reduce the workload of specialists in oral lesions, given the limited number of available specialists in this field.

1.3 Objective

1.Our research aims to study and develop the classification models by integrating segmentation techniques as a pre-processing.

2.Develop the model-based patient-specific medical history incorporating image dataset.

1.4 Scope of the study

Utilizing a deep learning model for the classification of oral lesions, there were 5 classes of leukoplakia, lichen planus, candidiasis, other white lesions, and ulcers. One approach involves segmenting teeth interference out of the images. And another approach construct the models by using both the image data and medical history. The image dataset in this research were collected from the patient of Faculty of Dentist Mahidol University. All images used in this research was acquired the approval of Ethical Committee of the Faculty of Dentistry/Faculty of Pharmacy, Mahidal University COA.NO.MU-DT/PY-IRB 2021/DT098).

1.5 Report Outline

Chapter 2 reviews of theories that are relevant to or utilized in this research

Chapter 3 describes the design method of each part of the research

Chapter 4 demonstrates the results and discussion

Chapter 5 conclusion of the results

CHAPTER 2
REVIEW OF THEORY RELATED

2.1 Normal oral mucosa

In the mouth, the oral mucosa is a crucial structure comprised of several layers essential for oral health and function as shown in Figure 1. The connective tissue, a vital component, lends it a pink hue and houses a vascular network in the bottom layer. Above the connective tissue is the basement membrane, a line in the middle part covered by epithelium, and the outermost layer is the keratin layer that has a white color as shown in Figure 2 [5].

Color: pink

Surface : smooth

Consitency:soft

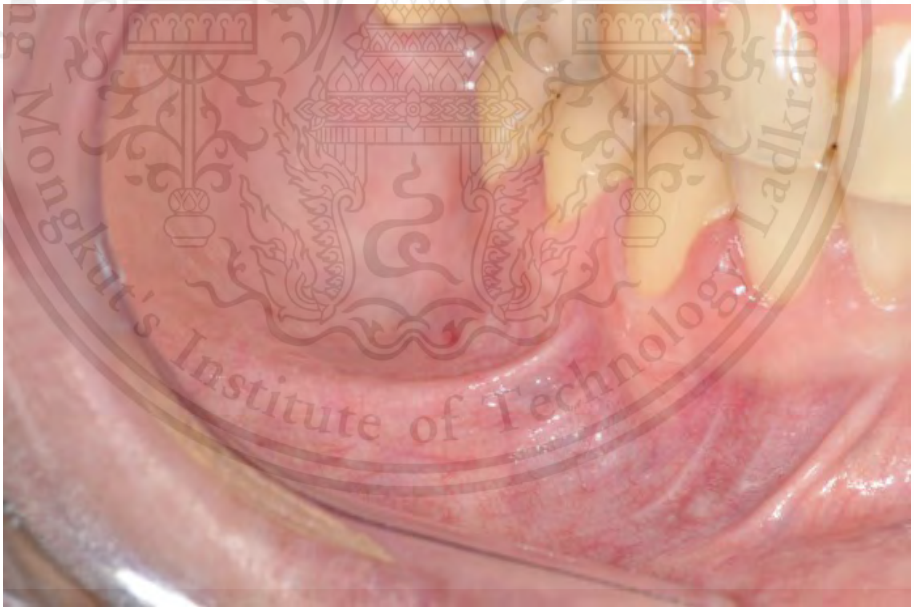

Figure 1 Normal oral mucosa[6]

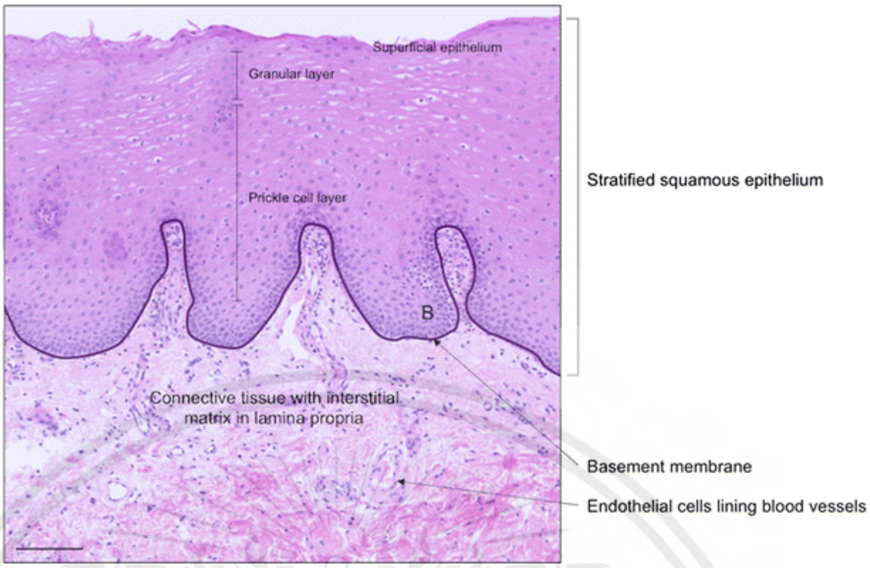

Figure 2 The images of oral mucosa [7]

2.2 White lesion

White lesion is the abnormal part in the oral cavity that have 4 different types [8]

- High keratin production (hyperkeratosis)
- High epithelial cells (acanthosis)
- Death of surface epithelium (necrosis)
- Fungal colonies (candidiasis)

In clinical diagnosis, white lesion can be catagorize in to 2 classes depend on the clinival diagnostic test

- Can be scraped off
- Cannot be scraped off

In case of can be scraped off, the type of lesion can be

- Candidiasis
- Food debris
- Detached keratin
- Necrotic tissues

In case of can not be scraped off, the type of lesion can be

- Linear albar

- Friction keratosis
- Lichen planus
- Leukoplakia

2.2.1 Detached keratin

This is one of the white lesions that can be scraped off due to high keratin production. This is not a pre-cancerous class. The detachment of keratin which shown in Figure 3 can be occur by following causes this condition.

- Herbal whitening tooth paste
- The ingredients of tooth paste including mint, cinnamon, clove oil, sage, eucalyptus oil

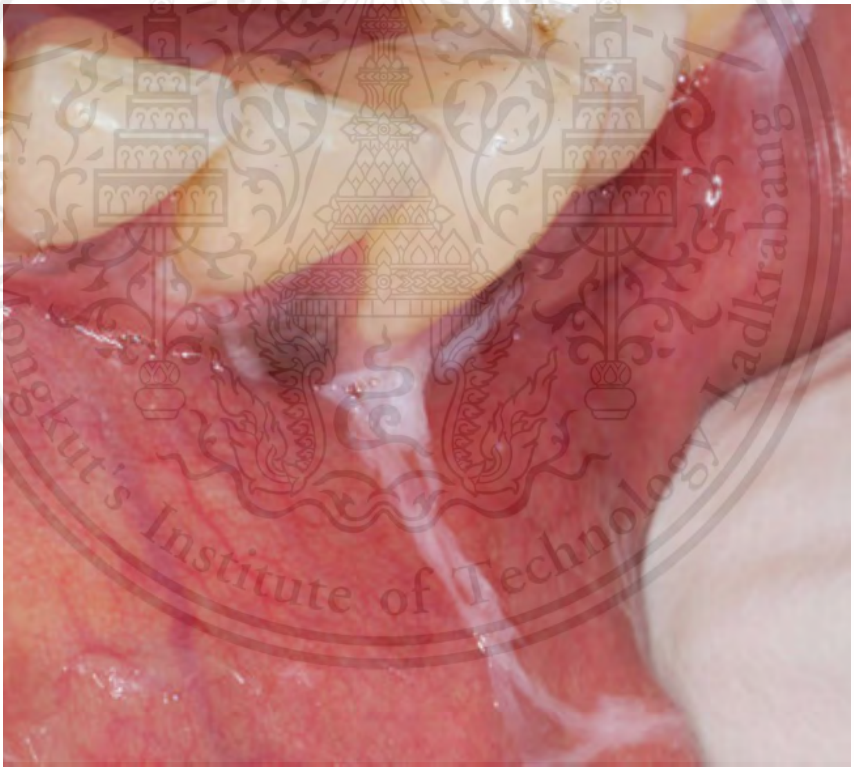

Figure 3 Detached keratin [9]

2.2.2 Candidiasis

Within the category of scrapeable lesions, candidiasis which shown in Figure 4 is the oral lesion that is infected by 2 types of fungi

- Superficial candidiasis
- Systemic candidiasis

This oral lesion doesn't have a fatal effect and it is non-pre-cancerous. Candidiasis can appear anywhere in the oral cavity and is commonly found in patients with a weakened immune system or using steroids for an extended period. It might be asymptomatic or cause dryness in the mouth [10-12].

Treatment for Candidiasis

- Elimination of predisposing factor
- Antifungal drugs e.g. Nystatin oral suspension, Micronazole gel

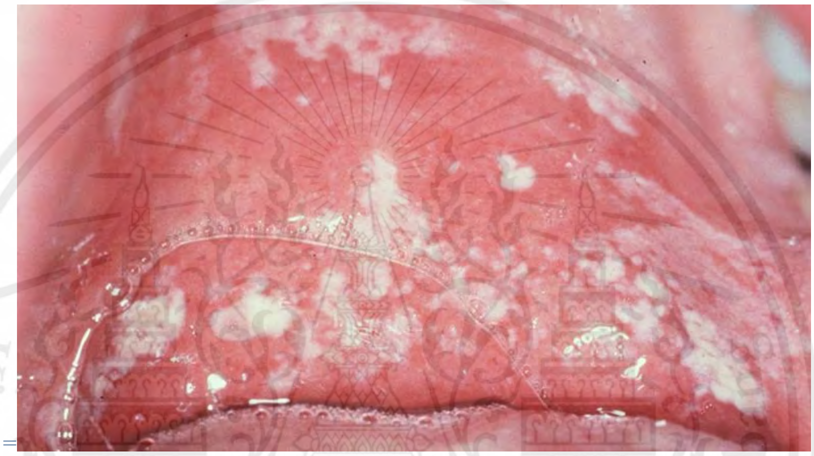

Figure 4 Candidiasis [13]

2.2.3 Necrotic epithelium

This oral lesion is classified as having a high number of epithelial cells (acanthosis). It's often found next to a tooth pain. Clinically, it appears as a white plaque that can be rubbed off, revealing sloughed tissue. It can also cause a burning sensation and pain [14].

Treatment for necrotic epithelium

- Self recovery
- Solcoseryl dental adhesive paste

2.2.4 Linear alba

This is one of the white lesions that can not be scraped off. Linear alba is a line-like area usually found in the cheek which shown in Figure 5, caused by the biting of teeth [15].

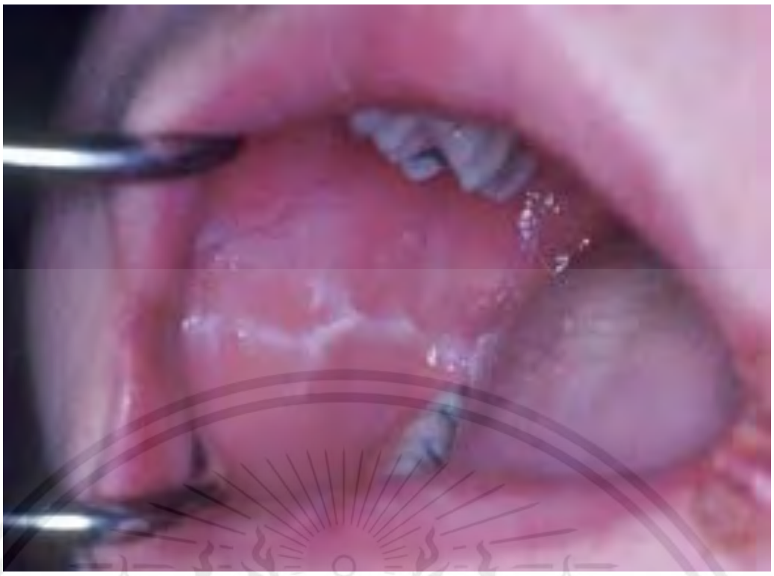

Figure 5 Linear alba [16]

2.2.5 Frictional Keratosis

This is a lesion that has a thicker layer of keratin which shown in Figure 6, caused by friction on the lip or cheek. Clinical features include a white lesion adjacent to an irritating factor, sometimes appearing with an ulcer [17,18].

Treatment of Friction Keratosis

- Remove irritation factor
- Observe and biopsy if necessary

Figure 6 Frictional Keratosis [19]

2.2.6 Lichen planus

Lichen planus which shown in Figure 7 is oral pre-cancerous. This lesion is the destruction of basal keratinocytes by cytotoxic CD8 T cells, but the specific cause of Lichen planus is unknown. Clinical features include white and red areas, along with white striations (Wickham's striae), categorizing it as a pre-cancerous class [20-22].

Treatment of Lichen planus

- Using steroid

Figure 7 Lichen planus [23]

2.2.7 Leukoplakia

Leukoplakia which shown in Figure 8 is a member of the non-scrapeable lesion category and is considered a potentially malignant disorder, with a premalignant nature that carries a 10% to 20% chance of developing into cancer. It belongs to the pre-cancerous class and is among the most common pre-cancerous lesions, characterized by the appearance of a white patch without symptoms [24-27].

Clinical feature

- White plaque that can't be scraped off
- Usually unilateral
- Asymyomic
- Homogenous
- Non-homogenous

Figure 8 Leukoplakia [28]

The pre-cancerous white lesions include leukoplakia and lichen planus, which should be under the supervision of dentist and require the treatment to prevent progression to oral cancer.

2.3 Ulcer

Ulcers are a type of lesion resembling white lesions but are not categorized as abnormal in oral lesions. They occur due to complete loss of the epithelial layer, often surrounded by a red color which shown in Figure 9.

Cause of the ulcer

- Bite the tongue or lip
- Lack of sleep
- Lack of vitamin B12
- Dehydrate

Figure 9 Ulcer [29]

2.4 Medical History

The medical history in this research is the data collected from dentists during patient visits. There's a checklist for patients to note symptoms, causes, and oral lesion diagnoses. We believe this data could significantly assist our research implementation.

- Burning sensation when have spicy food
- Smoking
- Alcohol consumption
- Betel nut chewing
- Can be scraped off
- Constant discomfort
- No symptoms

2.5 Convolutional Neural Networks(CNNs)

Convolutional Neural Networks (CNNs) are artificial neural networks that mimic human vision to comprehend images. They utilize a layered architecture comprising convolution, pooling, and fully connected layers to extract features from pictures. The convolution process dissects the image into smaller sections and employs filters to extract specific features [30].

2.5.1 Filter

It's a 2D table with different sizes (kernels). We'll use these filters to multiply with the input image but only in small sections as shown in Figure 10. The filter helps exclude unnecessary parts of the input image. It slides continuously, multiplying and summing to generate a single value for each position across the entire image. This method ensures that only essential features of the image are obtained. In a CNN model, numerous filter layers or convolution layers can be included. This process also helps in reducing the size of the input image.

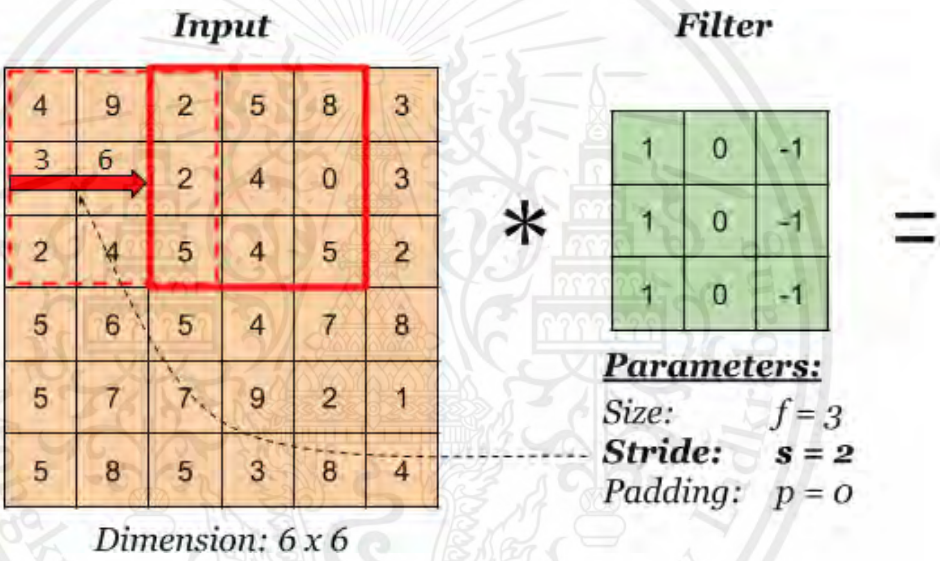

Figure 10 Filter kernel [31]

2.5.2 Pooling layer

The pooling layer is used to reduce the spatial size of the convolutional feature and decrease the amount of processed data. Similar to the convolutional layer, pooling also requires a window.

2.5.3 Fully Connected Layer

This is the layer of a neural network that connects all input layer to activation to next layer In the context of oral lesion classification as shown in Figure 11. The images of oral lesions serve as the initial input data, representing features within these images. This data is then connected to a hidden layer, which functions through a series of mathematical transformations. The hidden layer, being intermediate, remains unseen and comprises nodes that utilize weights to establish relationships between input and output. The configuration of hidden layers depends on the specific model used. Finally, the output layer yields the predictive results, particularly in tasks like classifying oral lesions.

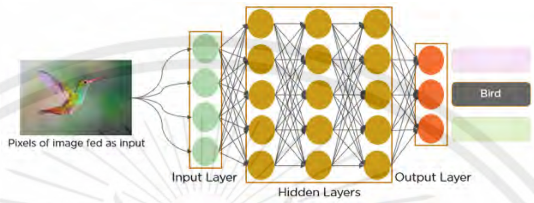

Figure 11 Layer in CNNs [32]

2.5.4 U-net

The U-net model is a neural network architecture designed for image segmentation. It consists of two main parts: the encoding path and the decoding path [33].

The encoding path, also known as the contracting path, is responsible for extracting features from the images by reducing spatial dimensions in layers. This path incorporates skip connections to preserve the positions of the feature maps in different layers and scale for segmentation as shown in Figure 12.

The decoding path, also known as the expansive path, involves up-sampling and convolutional layers used to enhance spatial resolution and concatenate with the skip connection from the encoding path to achieve precise segmentation.

This model to train it require the label of the ground to of the interesting area on the image for the model to learn. And this model can apply to medical image segmentation. The common evaluation for U-net is IoU (Intersection over Union).

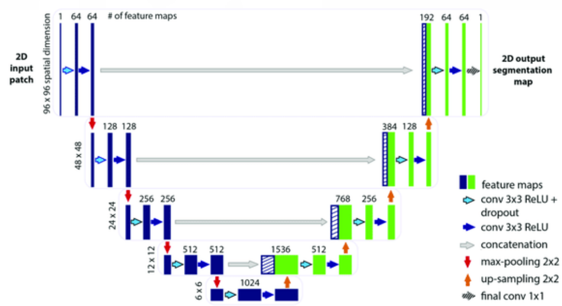

Figure 12 U-net architecture [34]

2.6 Densenet

The DenseNet model is a deep learining architecture that use dense connectivity to each neural network layers. Dense layer is the dense connectivity where every layer receives input from all previous layers as shown in Figure 13. Dense blocks are the sub-units of the densenet each blocks is use for reduce spatial dimension between each dense block. Densenet has many type the difference of each is the number of Dense layer such as Densenet121, Densenet 201and Densenet 169 [4].

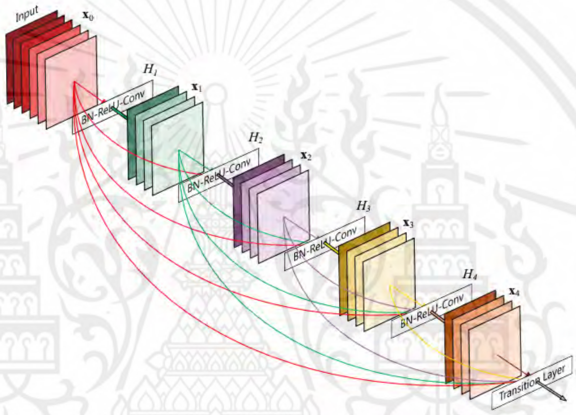

Figure 13 Dense Layer [4]

2.7 LSTM (long short-term memory)

Long Short-Term Memory (LSTM) as shown on Figure 14, a type of recurrent neural network, that designed for capture long-range of sequential data. This model is the improvement of RNN model that can mitigate the vanishing gradient that happen when working with long sequences. The cell structure contain previous cell, previous hidden state and input data for the left part. Right part is to send the information of process to next part contain new cell state and new hidden state. The application usually work with the frequency data, natural language processing or time series [35].

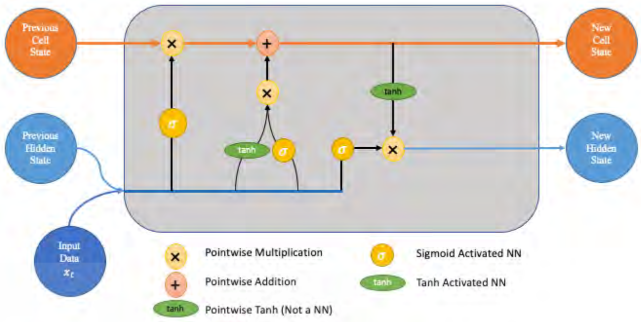

Figure 14 LSTM architecture [36]

2.8 VQA model

Visual Question Answering (VQA) model as shown in Figure 15, this model designed for answer question about the image. VQA model contain with 2 parts combining image analysis and the natural language processing components together to answer the question [37].

Image Encoding, VQA is the CNNs to extract the feature from the picture and we can adjust the model that we want to use by ourselves such as DenseNet121or ResNet.

Question Encoding, this part uses the RNN, LSTM or Transformer model to convert and extract the feature from the text base analysis.

Fusion part, after encode both image and text it will fusion together by create a joint embedding. Next to make a answer prediction to have fully connected layer to output the possible answers.

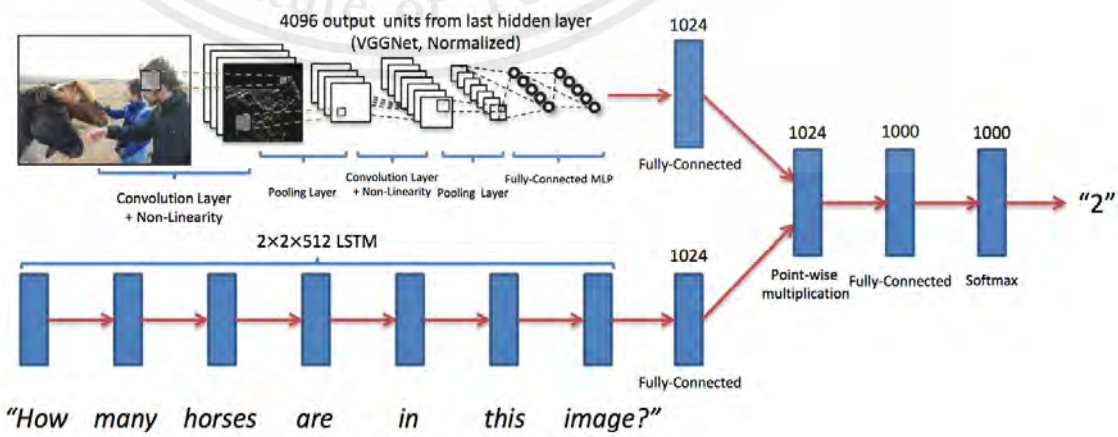

Figure 15 VQA model architecture [38]

2.9 Evaluation of the model performance

To evaluation our model performance we will use 4 metric to understand the capabilities of the model performance.

- Accuracy: use to measure the correction of prediction in total dataset, it calculated the ratio of correctly predicted data to total data
- Dice loss: is the metric that commonly use in image segmentation, it evaluates the similarity between two data and calculated using the dice coefficient. If the dice loss is high, it mean that overlop of the predicted and ground truth mask is high.
- IoU (Intersection over Union): use for overlap between the predicted and ground truth bounding mask, it calculated ratio of intersection area to the union area of prediction and ground truth. Higher IoU mean better prediction.
- Precision: measures the accuracy of positive prediction from model. It calculated the ratio of true positive predictions to the sum of true positive and false positive predictions.

2.10 Colab notebook

We explored Collaboratory (Colab) notebooks, which is a Google Colab an online platform for Jupyter notebooks. In this research, we used all the coding part with the python in Jupyter notebooks. The Feature of the colab is the free access to GPU and TPU resources and easy to access to data in the Google Drive. We provided the interface of the coding, text cells, and the output cells. Collaborating feature to coding with other people in real-time. Colab notebooks come with the libraries such as TensorFlow and PyTorch.

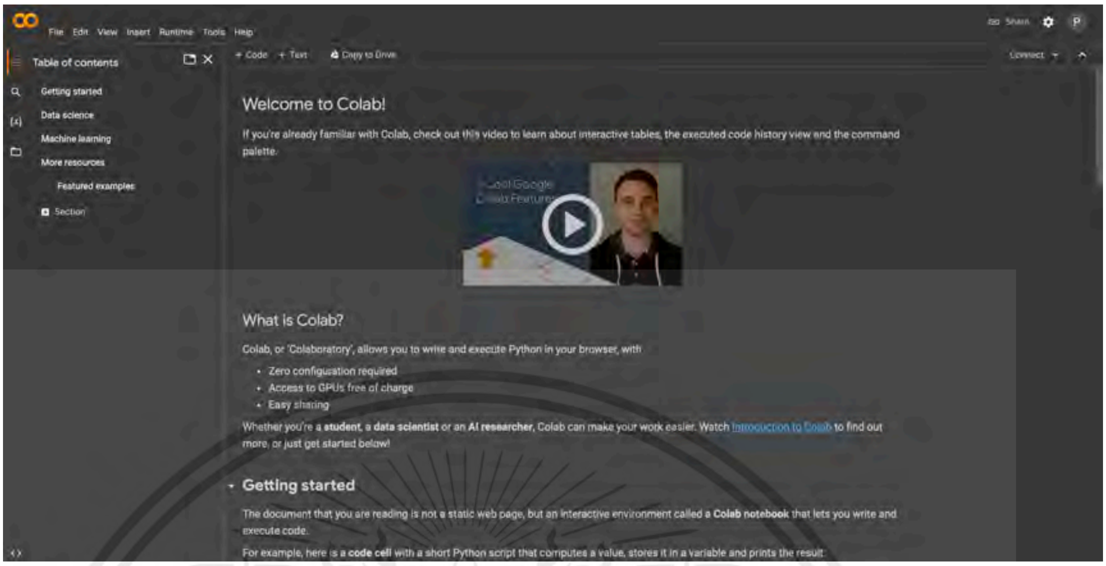

Figure 16 Web page of Colab notebook

2.11 Chapter Summary

Oral lesion has many classes, some are the pre-cancerous class that need an urgent medical treatments such as leukoplakia and lichen planus. The classification these white is a challenging task and require specialist. We propose to construct the models of oral lesion classification by Convolution Neural Networks (CNNs) that have been used to classify and identify many types of medical problem. CNNs can break down an image into smaller parts and extracting features from each component, and then combine to make a prediction. However, the result of the classification did not achieve same level as medical professional so we aim to improve the classification model by using the pre-processsing segmentation approach and including the medical history data with image dataset.

Pre-processing of segmenation implements the technique of removing interference cause by teeth has been accomplished. Segmentation part we used U-net which is a neural network architecture that segmenting the region of interest in images and remove the part that we don’t need.

Using medical history data with image dataset to improve classification, is the propose that we consulted with dentist that we hope that it can cooperate to help improving the model performance. This part we implemented the VQA model which is the multimodal model that can combine text base and images base classification by extract the feature of each data and conduct out the answer.

CHAPTER 3
METHODOLOGY

3.1 Ethics approval

This research was approved by the Ethical Committee of the Faculty of Dentistry/Faculty of Pharmacy, Mahidol University COA.NO.MU-DT/PY-IRB 2021/DT098. This research was in full compliance with International Guidelines for Human Research Protection including the Helsinki Declaration, the Belmont Report, CIOMS Guideline, and the International Conference on Harmonization in Good Clinical Practice. The image dataset have six classes including Leukoplakia, Lichen planus, Candidiasis, Other white, Ulcer and Normal mucosa. All image was collected from 2022-2023.

3.2 Dataset preparation

3.2.1 Dataset for pre-processing by segmentation

In this research , we utilized an oral lesion dataset that provided by an oral medicine specialist from the Faculty of Dentistry, Mahidol University. The dataset has six classes: leukoplakia (93 images), ulcer (93 images), normal mucosa (61 images), lichen planus (61 images), other white lesions (52 images), and pseudomembranous candidiasis (40 images) as shown in Table 1. To enhance model predictions, we created true masks and saved them as .png files through the RectLabel program.

For the segmentaion of the teeth which are the interference in the images.We can combined each classes and selected the dataset that inclued teeth to create model of segmentation. We divided the data into 295 training images, 80 validation images, and 25 test images.

Table 1 The number of images dataset for segmentation of the teeth interference

Oral lesion classes	Number of images
Normal mucosa	61
Ulcer	93
Lichen planus	61
Leukoplakia	93
Pseudomembranous Candidiasis	40
Other white lesions	52

3.2.2 Dataset segmented by U-net for classification

After segmented teeth out by U-net as shown in Figure 18, dataset included of five classes: leukoplakia (450 images), pseudomembranous candidiasis (238 images), lichen planus (273 images), other white lesions (283 images), and ulcer (335 images), then we separate the Test set for 40 images and the other split in 80:20 for training set and validation set as shown in Table 2.

Table 2 Dataset that used for classification after segmented by U-net

Oral lesion classes	Training set (images)	Validation set (images)	Test set (images)
Other white lesions	195	48	40
Lichen planus	187	46	40
Leukoplakia	328	82	40
Pseudomembranous Candidiasis	159	39	40
Ulcer	236	59	40

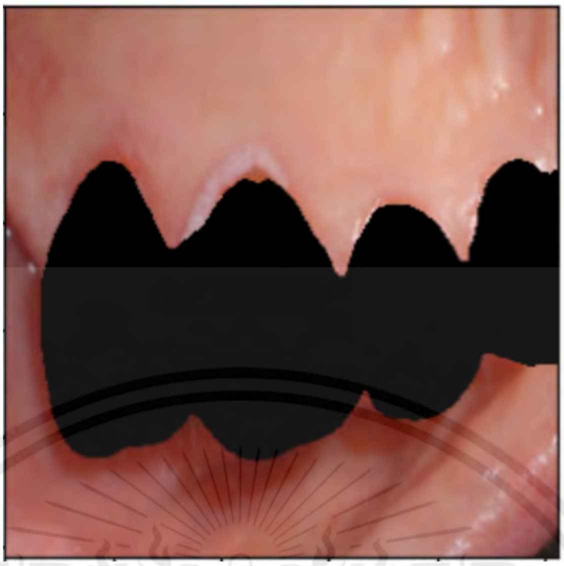

Figure 17 The example of the output of segmented teeth out that used for further classification

3.2.3 Dataset for VQA model

1. Image datasets for VQA model contain dataset consisted of five classes: ulcer (236 images), lichen planus (187 images), leukoplakia (328 images), pseudomembranous candidiasis (159 images), and other white lesions (195 images) and we splite for 80:20 for training and test as shown in Table 3.

Table 3 Data that use for images classes in VQA model

Oral lesion classes	Training set (images)	Test set (images)
Other white lesions	154	41
Lichen planus	148	39
Leukoplakia	260	68
Pseudomembranous Candidiasis	129	30
Ulcer	193	43

2. Dataset of Medical History the medical history are the cause of symptoms and diagnosis data of the oral lesion.

- Other white 100% no symtoms
- Lichenplanus 97% burning sensation when have spicy food, 3% no symptoms
- Leukoplakia 32% smoking 14%alcohol consumption 8%betel nut chewing 46% no symptoms
- Candidiasis 100% can be scraped off
- Ulcer 87%constant discomfort 13% no symptoms

To train the VQA model tokenized the medical history. To reduce the complicated full sentence. We defined can be scraped off as 'b', burning sensation when have spicy food as 'c', alcohol consumption as 'd', smoking as 'e', betel nut chewing as 'f', no symptoms as 'g' and constant discomfort as 'h' as shown in Table 4. The raw data that we collected from the dentist for image by image and list the medical history as shown in Figure 19. Lastly we path the data to our model by used image path, question which is the tokenized and the answer as a class of the output as shown in Figure 20.

Table 4 Medical History that tokenization for VQA model

Medical history	Tokenization
Can be scraped off	b
Burning sensation when have spicy food	c
Alcohol consumption	d
Smoking	e
Betel nut chewing	f
No symptoms	g
Constant discomfort	h

A	B	C	D	E	F	G	H	I
-v	เช็ดออกได้	กินเผ็ดแสบ	ดื่มเหล้า	สูบบุหรี่	เคี้ยวหมาก	ไม่มีอาการ	เจ็บหรือแสบตลอดเวลา	
16062011035_resized.jpg				1				
20140911_170015_resized.jpg						1		
20140911_170018_resized.jpg						1		
CIMG0122_resized.jpg				1				
CIMG0124_resized.jpg				1				
CIMG0890_resized.jpg				1				
CIMG1638-6-05-09_resized.jpg						1		
CIMG4142_resized.jpg						1		
CIMG4145_resized.jpg						1		
CIMG4932_resized.jpg						1		
CIMG5519_resized.jpg						1		
CIMG5520_resized.jpg						1		
CIMG5524_resized.jpg						1		

Figure 18 Raw data of medical history that we correct from dentist

```
    {
        "image_path": "/content/drive/MyDrive/Oral lesion 5 classes final train model resize imbalance/train/A_Leukoplakia/le18_File_003_resized.jpg",
        "question": "de",
        "answer": "A_Leukoplakia"
    },
```

Figure 19 Coding path in VQA model for leukoplakia image and medical history of alcohol consumption and smoking in question section

3.3 Segmentation

Segmenting oral lesions is challenging because of their dynamic shapes and undefined margins. We collaborated with a dentist to gain insights into different oral lesion types and shapes, which significantly enhanced our segmentation approach to address this challenge. Due to limited data and a wide distribution of lesions, achieving a perfect segmentation for the lesions to be used as true masks was challenging. Our model was tested on a set of oral lesion images from the leukoplakia class. However, the results, as shown in Figure 21, were unsatisfactory, showing an Intersection over Union (IoU) value of 0.43. Hence, our approach has shifted to focusing on segmenting interferences within the images, primarily targeting teeth, the most common interference. Because of their static shape and white color, teeth are easier to identify. We applied a mask to segment teeth, inverted it, and multiplied it to remove teeth from the images. We performed this preprocessing step to improve the accuracy of our classification model. We employed the RectLabel program to generate true masks for the model to predict, outputting them as PNG files, as shown in Figure 22. We divided the data into 295 training images, 80 validation images, and 25 test images after segmenting teeth.

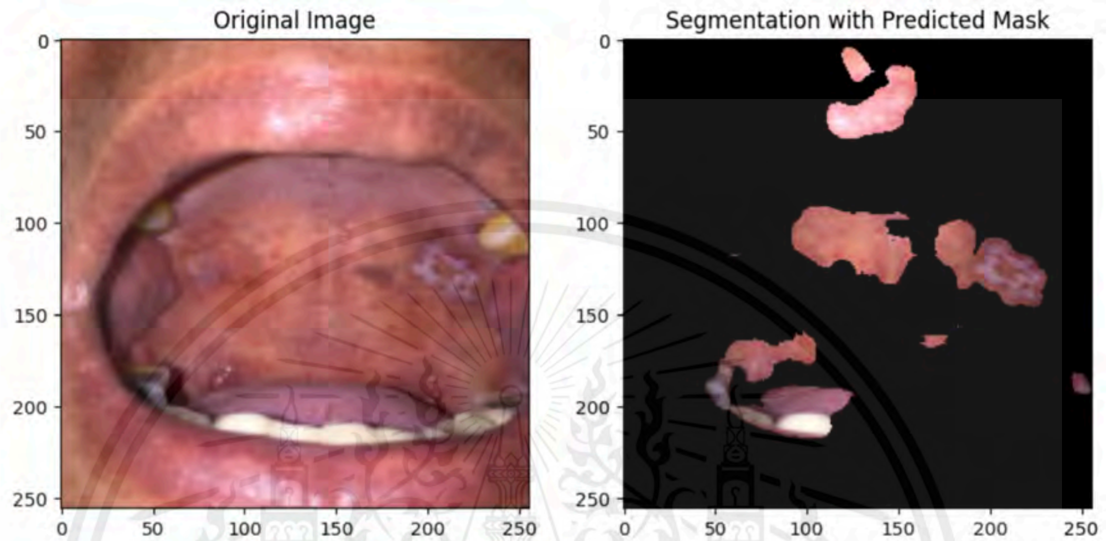

Figure 20 The output generated by the U-Net model for segmenting oral lesions

Figure 21 The mask that generated using the Rectlabel program

3.4 Implementation Tools

This research used Keras version 2.5 and Tensorflow 2.2.1. The model architecture utilized was sourced from segmentation_model [23], based on the U-net backbone architecture. The research utilized Python version 3.9.0 and conducted the model training on Google Colab. The GPUs accessible on Google Colab included Nvidia K80 or T4, boasting 12 GB of GPU memory and a GPU memory clock of 0.82 GHz, offering a performance of 4.1 TFLOPS. Alongside, Google Colab supplied two

CPU cores, 12 GB of accessible RAM, and 358 GB of disk space, with 5 GB for each session.

3.5 Model Architecture for Segmentation

The segmentation models utilized in this research were all based on the U-net backbone architecture, as shown in Figure 23 to produce the dataset for subsequent classification. The U-net backbone employs a standard convolution size of 3x3, whereas various other CNN backbones including ResNet 18, SE-ResNet 18, ResNeXt 50, Inception V3, MobileNet V2, SE-ResNeXt 50, DenseNet 121, EfficientNet B3, and VGG 19 were employed for feature extraction, as shown in Table 5. To measure the models' efficacy, we assessed dice loss, accuracy, and Intersection over Union (IoU) as shown in equations (1) to (3). We configured the Adam optimizer with a learning rate of 0.0001.

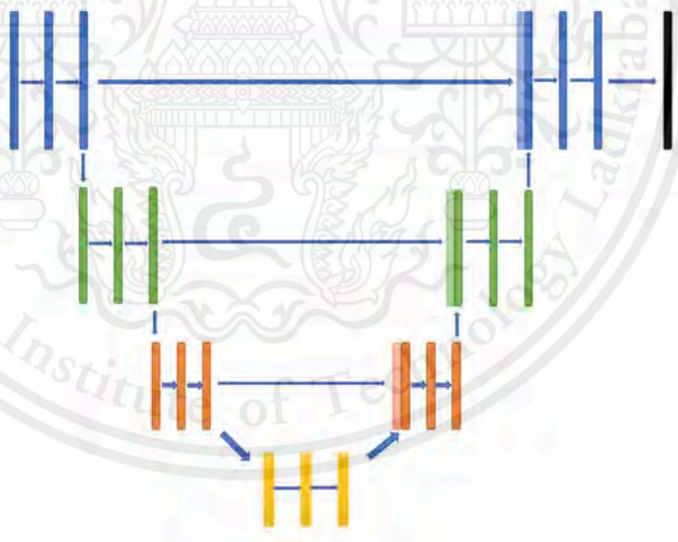

Figure 22 The U-net Architecture

Table 5 list of U-net backbone

U-net backbone
Original U-net
ResNeXt 50
VGG 19
EfficientNet B3
SE-ResNet 18
ResNet 18
DenseNet 121
SE-ResNeXt 50
Mobilenet V2
Inception V3

$$\text{Dice loss} = \frac{2*TP}{2*TP+FP+FN} \quad (1)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$\text{IoU} = \frac{TP}{TP+FP+FN} \quad (3)$$

3.6 Model Architecture for Classification

Previously, we executed teeth segmentation as a pre-processing step for further classification. This segmentation step generated a new dataset that segmented teeth out. The classification models in this research are built upon the DenseNet 121 architecture [4], utilizing a transfer learning approach. DenseNet 121 employs direct connections between layers, enabling efficient information flow, as shown in Figures 24 and 25. Transfer learning involves initial the weights of a pre-trained model from a different task to initialize our model's weights. Our classification model was trained 100 epochs of training, using a batch size of 32. To evaluate the classification performance, we employed loss, accuracy, and precision metrics represented by equations (2), (4), and (5) respectively. Additionally, we analyzed the confusion matrix and learning curve to gain deeper insights into the model's efficacy. The Adam optimizer was configured with a learning rate of 0.0001.

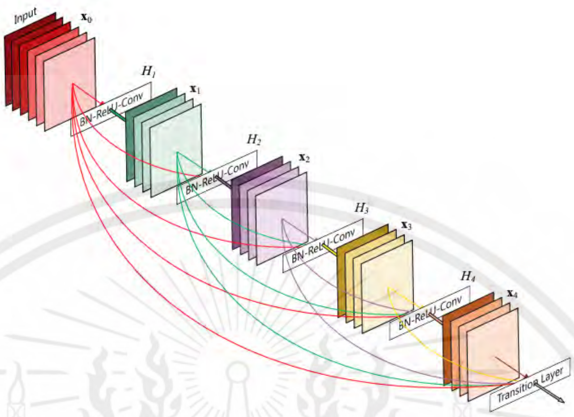

Figure 23 A block with 5 layers in DenseNet [4]

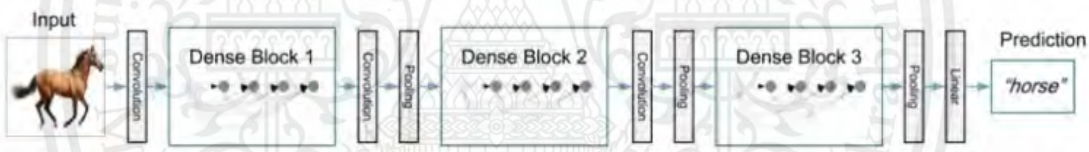

Figure 24 A DenseNet with 3 dense blocks

$$\text{loss} = -\sum(y\_true \times \log(y\_prediction)) \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

3.7 Model Architecture for VQA

VQA model contains 2 sub-models, one from image feature extraction and another one is text feature extraction, which combines 2 features to make a prediction as shown in Figure 26. Image feature extraction can change to any CNNs model, in our study we chose DenseNet 201 pre-trained model to extract image feature. Text feature extraction we used LSTM model that can detect the pattern or the frequency of text and could adjust the hidden layer. To evaluate the classification performance, we employed loss, accuracy, and precision metrics represented by equations (2), (4), and (5) respectively. Additionally, we analyzed the confusion matrix and learning curve to gain deeper insights into the model's efficacy. The Adam optimizer was configured with a learning rate of 0.0001.

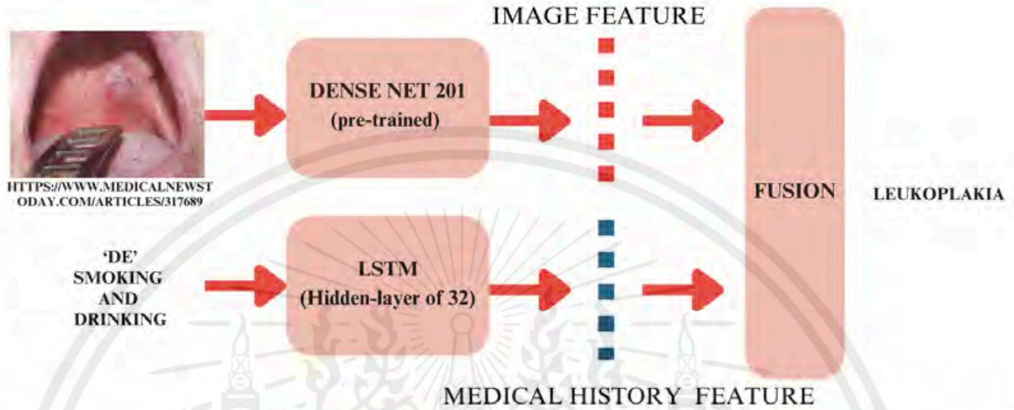

Figure 25 the VQA model architecture

3.8 Summary

This chapter introduce the dataset preparation for each model. In this research contains with 2 main methods of classification. The first one is the segmentation part that aim to segment the interference of the images which is the teeth by difference backbone of the U-net model. The segmented the teeth images were use as a new dataset for classification model of DenseNet 121. In the second part of this research we applied the medical data by using VQA model to train images and text s simultaneously. The evaluation of model performance for segmented, we used accuracy, dice loss and IoU. And for the classification, we used accuracy, loss and precision.

CHAPTER 4
EXPERIMENTAL RESULT AND DISCUSSION

4.1 Introduction

In this chapter showed the result of the each model start with the segmentation model that we used many type of backbone U-net models to evaluate the performance of the segmentation. Next part of the result would be utilize the best model to create a new dataset of segmented teeth out and used the DenseNet 121 architecture to construct the classification model. Lastly would be the result of the VQA model performance which use both image and medical history.

4.2 Result and Discussion for Segmentation U-net Model

For further classification of oral white lesions, we successfully preprocessed the image dataset by eliminating interference. Because of the limited data and the variability in lesion distribution, achieving a high accuracy segmentation of the lesions to create true masks was challenging. Hence, we chose to segment interferences in the images, with teeth being the most common interference owing to their shape and white color. This approach enabled us to tackle the segmentation challenge more effectively despite the limitations of the dataset. We attained high accuracy and IoU scores by removing teeth from the images. The outcomes of teeth segmentation using the U-net backbone with various additional backbones are shown in Table 6. The conventional U-net models we employed attained an IoU of 74.9% on the training set and 81.7% on the validation set. The resulting mask generated by the constructed U-net is shown in Figure 27. In the case of DenseUNet, the IoU for both training and validation stood at 92.4%. Figure 28 shows the output mask produced by DenseUNet. Significantly, EfficientUNet achieved the highest IoU, scoring 93.4% in training and 92.9% in validation. The EfficientUNet output mask is shown in Figure 29. Following this, Figure 30 shows the inverted mask, effectively eliminating teeth from the original image processed by EfficientUNet. Figure 31 shows the learning curve for EfficientUNet.

Table 6. The performance of segmentation models

U-net backbone	Metrics			
	Training IoU	**Validation IoU**	**Validation Accuracy**	**Validation Dice loss**
U-net Original	0.749	0.817	0.881	0.318
EfficientNet B3	0.934	0.929	0.986	-0.101
VGG 19	0.925	0.924	0.985	-0.098
SE-ResNet 18	0.911	0.917	0.984	-0.088
ResNeXt 50	0.927	0.921	0.985	-0.091
SE-ResNeXt 50	0.939	0.932	0.972	-0.104
Mobilenet V2	0.918	0.921	0.985	0.095
DenseNet 121	0.924	0.924	0.985	-0.096
ResNet 18	0.876	0.897	0.981	-0.068
Inception V3	0.919	0.917	0.984	-0.086

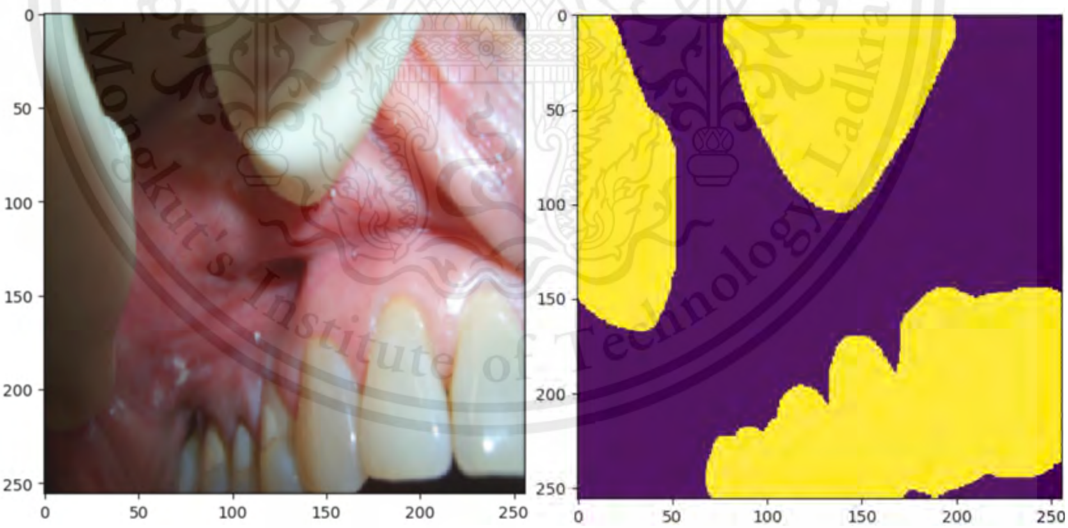

Figure 26 The U-net's output mask

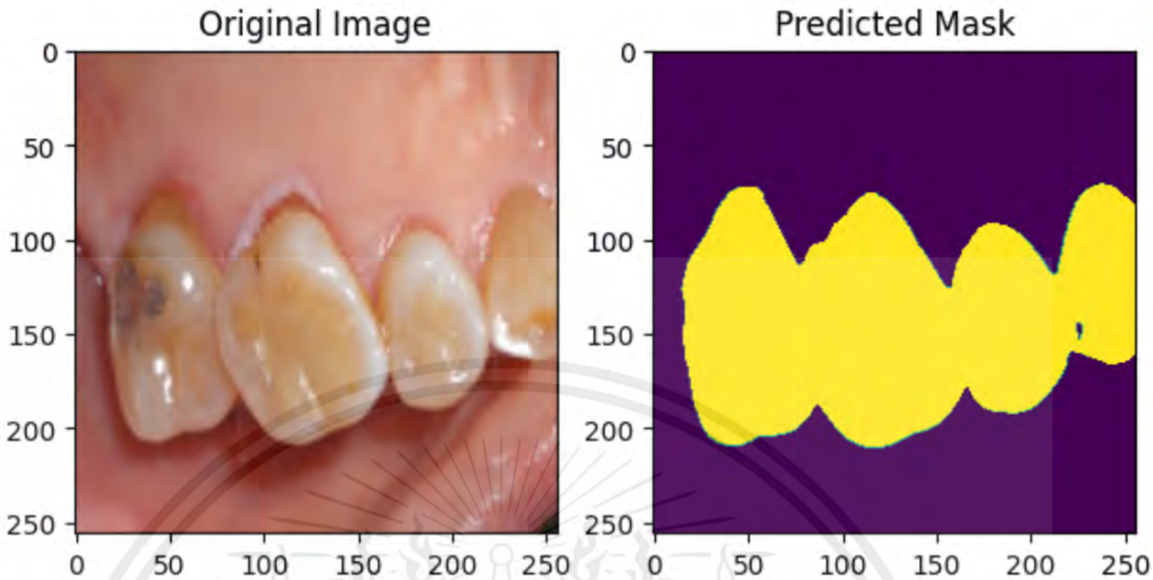

Figure 27 The DenseUNet’s output mask

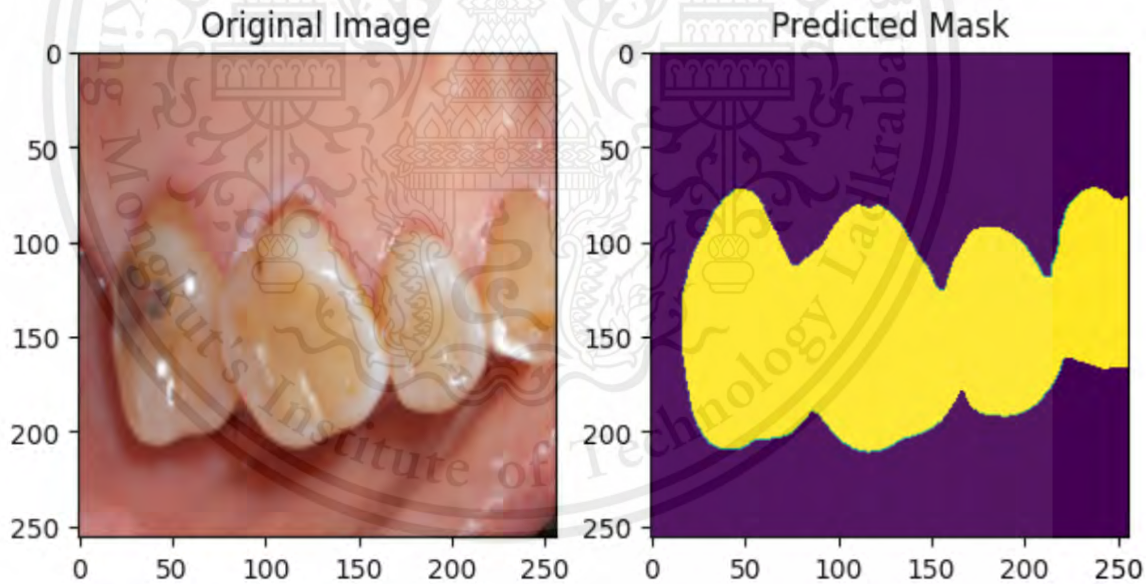

Figure 28 The EffcientUNet’s output mask

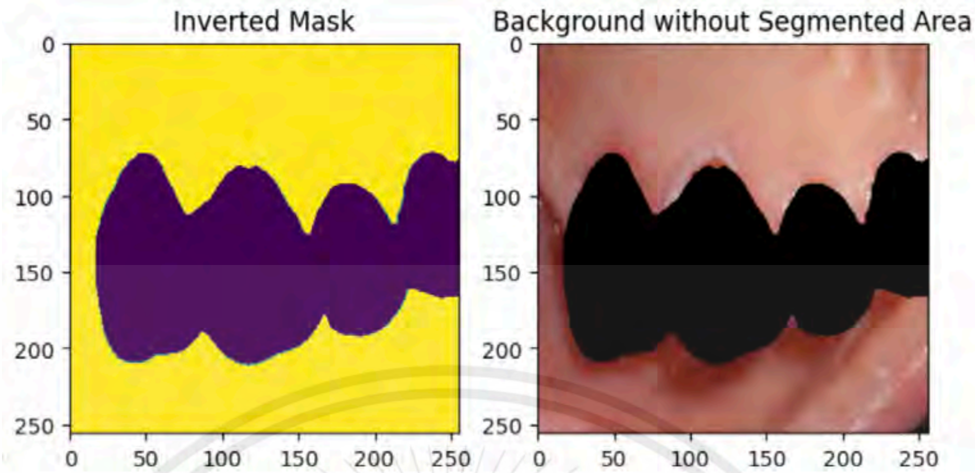

Figure 29 EfficientUNet's inverted mask and removed teeth from the original image

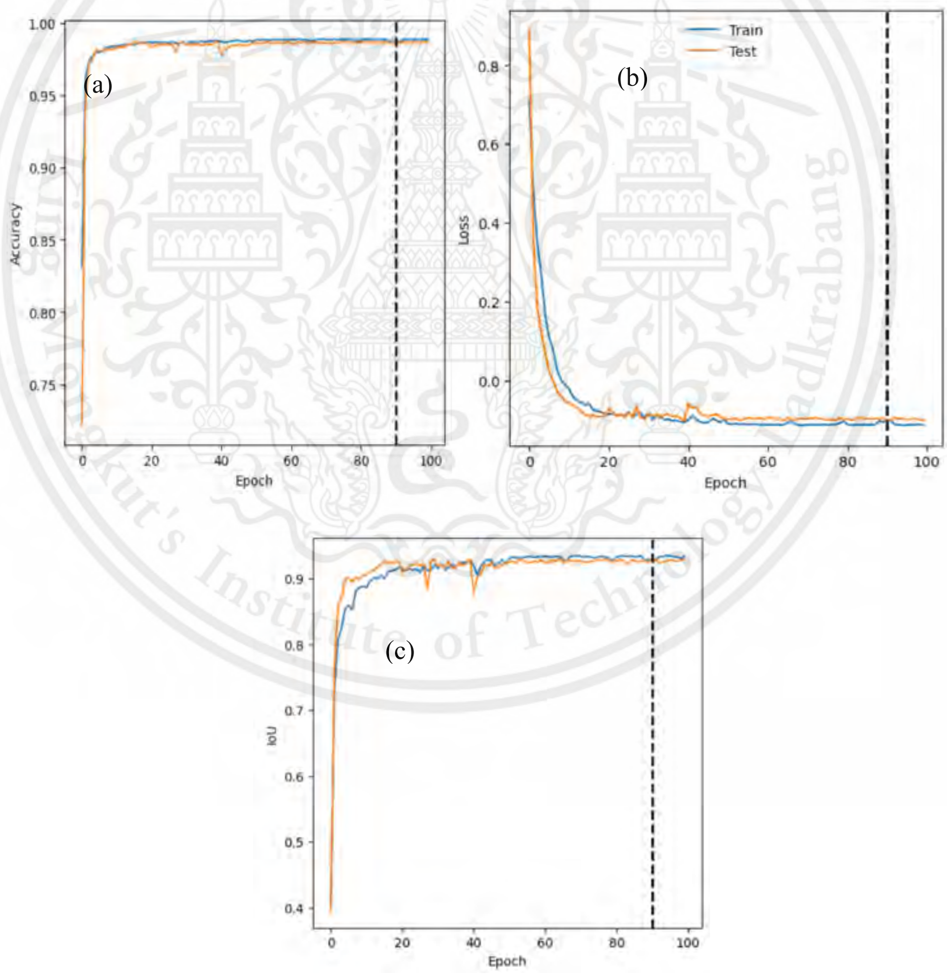

Figure 30 The accuracies for training and validation (a), the loss for training and validation (b), and the IoU for training and validation (c) of EfficientUNet

Our research's findings indicate that different model architectures significantly affect tooth segmentation accuracy. Among the implemented models, the original U-net model has the lowest IoU scores and the highest loss dice validation across all models. This implies that the U-net model may not be the optimal choice for teeth segmentation. Conversely, both DenseUNet and EfficientUNet models showed the best performance among of all backbones, giving high IoU scores on both the training and validation sets. Considering the learning curve, EfficientUNet reached its peak performance and showed stable learning after 20 epochs. However, there wasn't a significant difference in performance when evaluated on the test set among these models.

4.3 Result and Discussion for classification of segmented teeth out dataset

After using the EfficientUNet segmentation model to create a new dataset, which excludes teeth from the original images, we found it suitable for training the DenseNet 121-based classification model. The attained loss values during training, validation, and testing were 0.033, 0.54, and 1.054, respectively. These results are shown in Table 7. The accuracy achieved at 99.1% for training, 86.1% for validation, and 75.5% for testing. Precision scores were also acquired, registering 99.2% for training, 85% for validation, and 76.1% for testing. The learning curve of the DenseNet 121 model is shown in Figure 32 and the confusion matrix is shown in Figure 33. Based on these results, the model's performance was not satisfactory in comparison to our previous research [39]. Moreover, the test set accuracy didn't notably higher that achieved using the original image dataset. However, various CNN architectures like GoogleNet, Inception, and others have yet to be explored. Thus, exploring alternative classification architecture networks is a promising approach to yield more effective models.

Table 7. The classificant performance of DenseNet 121 model

DenseNet 121	Accuracy	Precision	Loss
training	0.991	0.992	0.033
validation	0.861	0.85	0.54
test	0.755	0.761	1.054

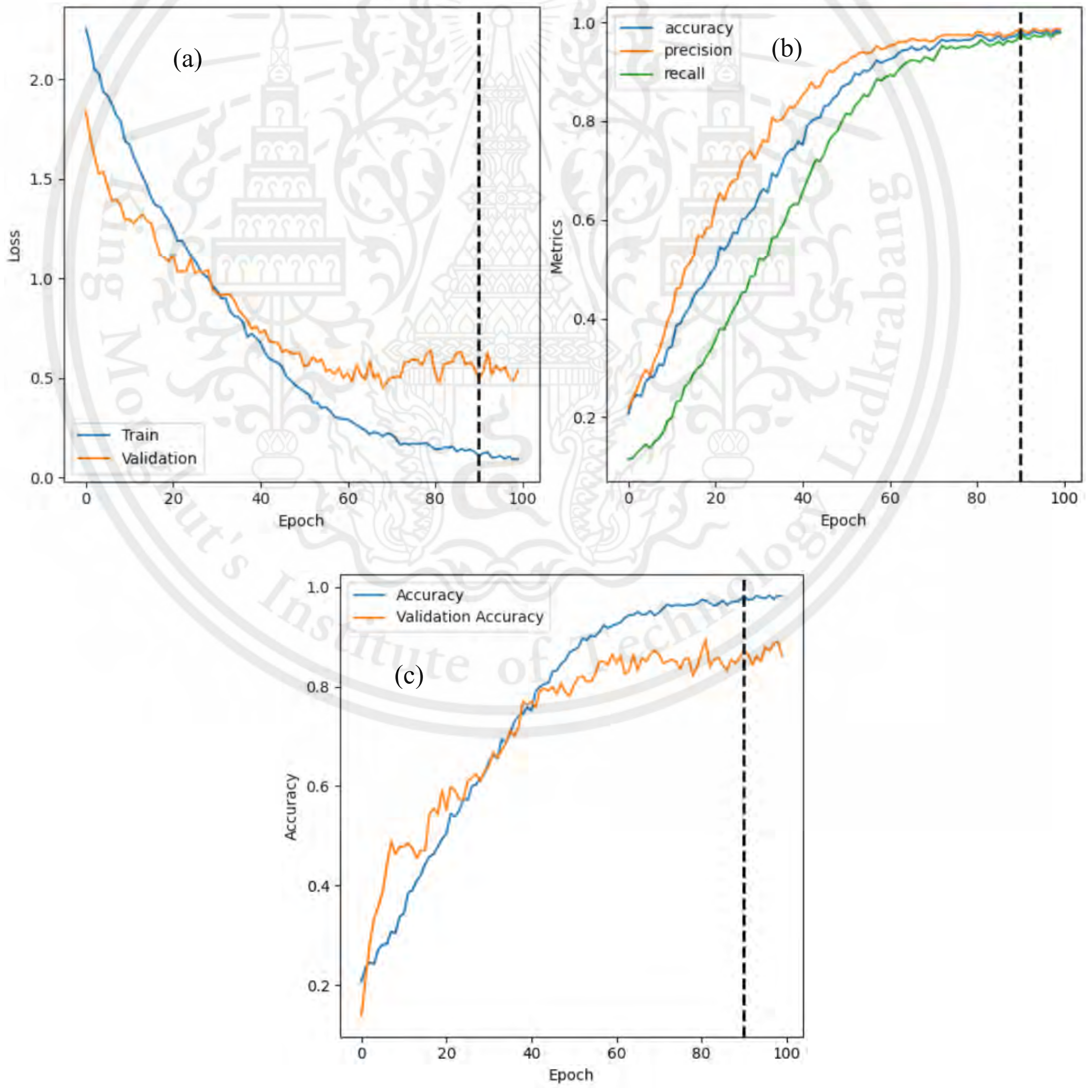

Figure 31 The loss for training and validation (a), accuracy, precision and recall of training (b), and the accuracy for training and validation (c) of DenseNet 121

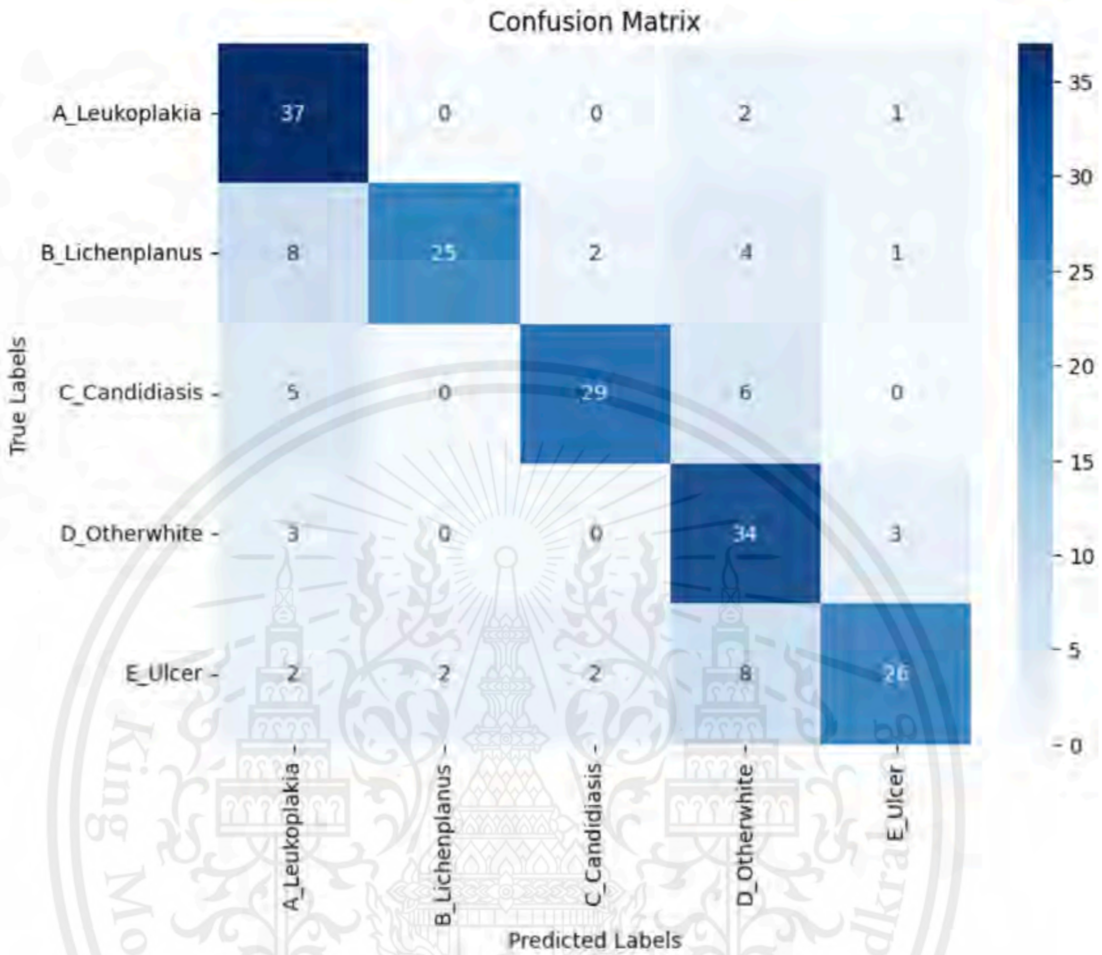

Figure 32 The Confusion Matrix representing the test dataset for the trained DenseNet 121 model

4.4 Result and Discussion for VQA model

The result from the VQA model which used DenseNet201 and LSTM as a feature extraction. In the first attempt we used 256 of hidden layers the result was very high. We used hypothesis of the weight adjustment of the images feature and text feature to make the model attened on the image feature more than text feature. The result of this hypothesis was not satisfied. Lastly, we reduce the hidden layer of the LSTM part which is the medical history. Reducing hidden layers of LSTM could improve the performance of VQA model that the model was not overfit to the medical history. VQA training achieved an accuracy of 0.9921, loss of 0.0256 and precision of 1.000. For the test achieved accuracy of 0.8597, loss of 0.3332 and precision of 0.9762 as shown in Table 8.

Table 8 The performance of the VQA model

VQA	Accuracy	Precision	Loss
training	0.9921	1.000	0.0256
test	0.8597	0.9762	0.3332

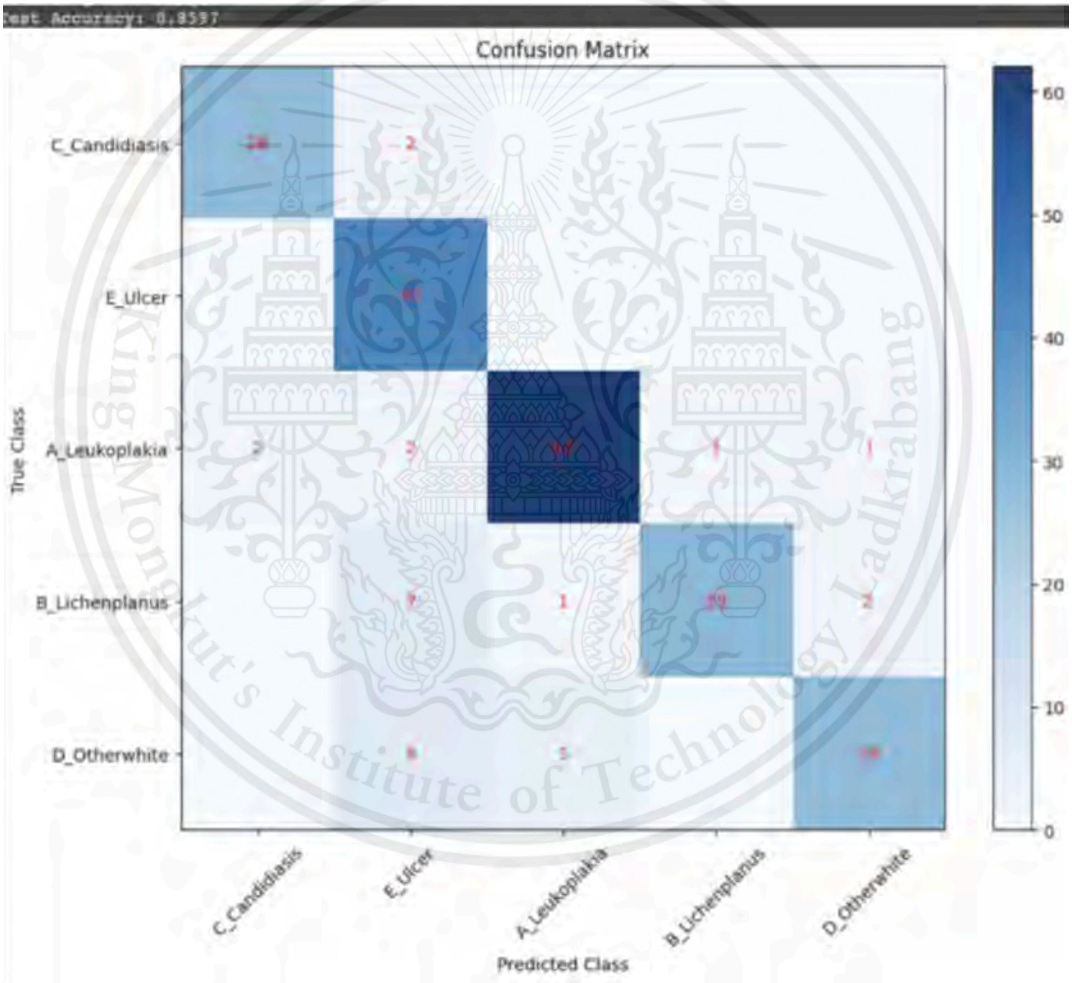

Figure 33 The confusion matrix of the VQA model prediction

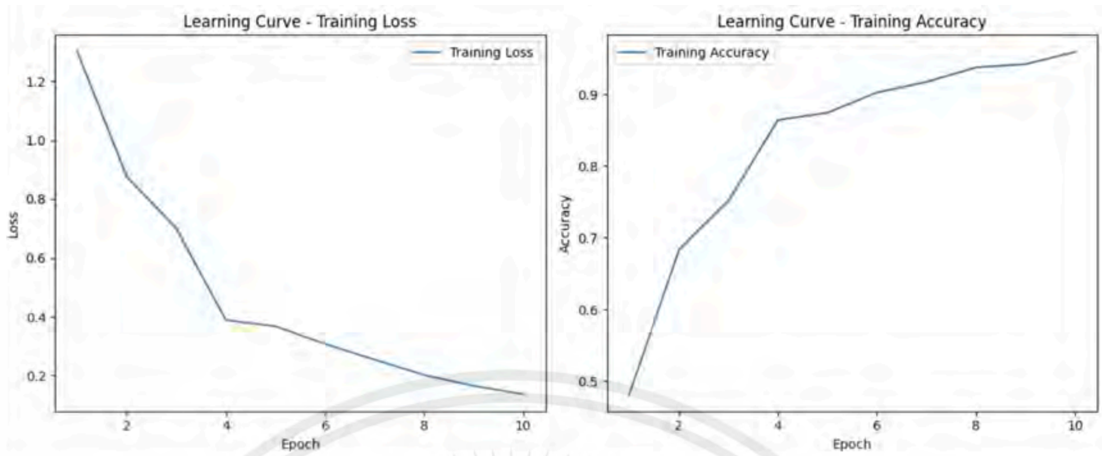

Figure 34 The learning curve of the VQA model

4.5 Compared all classification result

Our first result of pre-processed interference teth out by segmentation was not much improve compare to out previous research with raw image [39] in both loss and accuracy. From the result of the VQA model we could see that the improvement of the model performance was significant in both loss and accuracy. The results of each proposed techniques were compared as shown in Table 9.

Table 9 Comparison of the classification result of test set

Techniques	Model architectures	Result (test set)	
		Loss	Accuracy
Our previous research of classification of raw data without pre-processing [39]	DenseNet 121	0.755	0.7755
Pre-processed interference teeth out by segmentation (our proposed model)	DenseNet 121	1.054	0.755
Classification by incorporating medical history and image datasets (our proposed model)	VQA	0.3332	0.8597

4.6 Summary

The segmentation model's results showed the best performance with both DenseUNet and EfficientUNet models, achieving significantly high IoU scores on both training and validation sets among the adapted backbones. Considering the learning curve, EfficientUNet reached its peak performance and showed stable learning after 20 epochs. However, the performance assessed by the test set did not show significant differences among these models. Then, we utilized the EfficientUNet model to segment the teeth from the images and generate a new dataset. When we used new dataset for classification the result was not improve, but the result of the VQA model which use medical history and image improved.

CHAPTER 5
CONCLUSION

The aim of this project was to improve the classification of oral lesion. Segmenting lesions remains challenging due to limited data and undefined margins. Removing definable objects like teeth via segmentation is an attractive approach to explore the classification performance of deep learning models. This research successfully conducted the interference teeth segmentation using a combination of U-net backbone, DenseNet 121, and EfficientNet B3. The pre-processed images were utilized to train the DenseNet 121 CNN model for oral white lesion classification. The classification model's performance indices for validation accuracy and precision were promising, reaching 0.861 and 0.85, respectively. The performance achieved using the interference segmentation technique doesn't significantly outperform the original image dataset without interference segmentation. Comparatively, focusing on the DenseNet 121 model, the assumption that the teeth's white color is a main interference with classification might not hold. Next for the VQA model part that we implemented the medical history to help classification we could see the performance improved. We could conclude that the medical history can improve model classification. Finally from this research that we implement both segmented and medical history to improved classification performance, we could propose that medical history can improve the classification more from our result.

5.1 Suggestion

1. For the segmentation part if we have more data of the oral lesion, we could try the YOLO V8 model which require more computational unit.

2. If the limitation data still remains we could try other model that we don’t use yet, such as DenseNet201, GoogLeNet.

3. Try to use unsupervised learning model to train the both medical data and image classification.

REFERENCES

[1] Neville, B. W., Damm, D. D., Allen, C. M., and Chi, A. C. published "Oral and Maxillofacial Pathology" in its 4th edition with Elsevier in 2019.

[2] Park, S. Y., Kim, J. Y., Lee, S. K., and Lee, J. H. authored "A deep learning-based approach for the detection and classification of oral lesions" in the Journal of Clinical Medicine, Volume 9, Issue 12, on page 3852, in 2020.

[3] Li, H., Li, T., Li, L., and Xia, L. published the article "A Deep Learning Model for Leukoplakia and Lichen Planus Recognition in Oral Mucosa Images" in the Journal of Healthcare Engineering in 2021.

[4] Huang, G., Liu, Z., Pleiss, G., Van Der Maaten, L., & Weinberger, K. (2019). Convolutional Networks with Dense Connectivity. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[5] Brizuela, M., & Winters, R. (2023). Histology, Oral Mucosa.

[6] Benites, Bernar, et al. "Late recurrence of Burkitt's lymphoma in the jaw: numb chin syndrome as the only symptom." Autopsy Case Reports 11 (2021): 10.4322/acr.2020.218.

[7] Wardill, Hannah, et al. "Tight junction defects are seen in the buccal mucosa of patients receiving standard dose chemotherapy for cancer." Supportive Care in Cancer: Official Journal of the Multinational Association of Supportive Care in Cancer 24 (2016): 10.1007/s00520-015-2964-6.

[8] Mortazavi, H., Safi, Y., Baharvand, M., Jafari, S., Anbari, F., & Rahmani, S. (2019). Oral White Lesions: An Updated Clinical Diagnostic Decision Tree. Dental Journal, 7(1), 1-15. DOI: 10.3390/dj7010015. PMID: 30736423. PMCID: PMC6473409.

[9] Müller, Susan. "Frictional Keratosis, Contact Keratosis and Smokeless Tobacco Keratosis: Features of Reactive White Lesions of the Oral Mucosa." Head and Neck Pathology 13 (2019): 10.1007/s12105-018-0986-3.

[10] Benito-Cruz B, Aranda-Romo S, López-Esqueda FJ, de la Rosa-García E, Rosas-Hernández R, Sánchez-Vargas LO. Oral Candida isolates and fluconazole susceptibility patterns in older Mexican women. Arch Gerontol Geriatr. 2016 Jul-Aug;65:204-10.

[11] Mahajan B, Bagul N, Desai R, Reddy M, Mahajan A, Shete A, Risbud A, Mane A. Pseudomembranous Type of Oral Candidiasis is Associated with Decreased Salivary Flow Rate and Secretory Immunoglobulin A Levels. Mycopathologia. 2015 Aug;180(1-2):75-80.

[12] Akpan A, Morgan R. Oral candidiasis. Postgrad Med J. 2002 Aug;78(922):455-9.

[13] Edminister, John R., and William D. James. "Mucosal Candidiasis Clinical Presentation." Medscape. Medscape, [URL: https://emedicine.medscape.com/article/1075227-clinical?form=fpf].

[14] Brown HR, Hardisty JF. 1990. Oral cavity, esophagus and stomach. In: Pathology of the Fischer Rat (Boorman GA, Montgomery CA, MacKenzie WF, eds). Academic Press, San Diego, CA, 9-30.

[15] Kansky AA, Didanovic V, Dovsak T, Brzak BL, Pelivan I, Terlevic D Radiol Oncol 2018 Sep 11;52(3):263-266. doi: 10.2478/raon-2018-0031. PMID:30210036

[16] Local Plymouth, MN Dentist. "What is Linea Alba?" Local Plymouth, MN Dentist, 17 November 2015, [URL: https://www.life-smiles.net/dr-geislers-blog/white-line-cheeks].

[17] Chi AC, Lambert PR, 3rd, Pan Y, Li R, Vo DT, Edwards E, Gangarosa P, Neville BW. Is alveolar ridge keratosis a true leukoplakia? A clinicopathologic comparison of 2,153 lesions. J Am Dent Assoc. 2007;138(5):641–651. doi: 10.14219/jada.archive.2007.0236.

[18] Haisley-Royster CA, Allingham RR, Klintworth GK, Prose NS. Hereditary benign intraepithelial dyskeratosis: report of two cases with prominent oral lesions. J Am Acad Dermatol. 2001;45(4):634–636. doi: 10.1067/mjd.2001.116336.

[19] Jose Luis Tapia, DDS; Chief Editor: Jeff Burgess, DDS, MSD. "Oral Frictional Hyperkeratosis." Updated: Feb 19, 2021, [URL: https://emedicine.medscape.com/article/1076089-overview].

[20] Scully C, Beyli M, Ferreiro MC, Ficarra G, Gill Y, Griffiths M, et al. Update on oral lichen planus:etiopathogenesis and management. Crit Rev Oral Biol Med. 1998;9:86–122.

[21] Regezi JA, Dekker NP, MacPhail LA, Lozada-Nur F, McCalmont TH. Vascular adhesion molecules in oral lichen planus. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 1996;81:682–90.

[22] Zhou XJ, Sugerman PB, Savage NW, Walsh LJ. Matrix metalloproteinases and their inhibitors in oral lichen planus. J Cutan Pathol. 2001;28:72–82.

[23] N. Burkhart and the AAOM Web Writing Group. "Oral Lichen Planus." Updated: September 2013, [URL: https://www.aaom.com/oral-lichen-planus].

[24] Hsue SS, Wang WC, Chen CH, et al. Malignant transformation in 1458 patients with potentially malignant oral mucosal disorders: a follow-up study based in a Taiwanese hospital. J Oral Pathol Med. 2007;36:25–9.

[25] Sugar L, Banoczy J. Follow-up Studies in Oral Leukoplakia. Bull. Org. mond. Sante. 1969;41:289–293.

[26] Liu W, Wang YF, Zhou HW, et al. Malignant transformation of oral leukoplakia: a retrospective cohort study of 218 Chinese patients. BMC Cancer. 2010;10:685–685.

[27] Silverman S, Gorsky M, Lozada F. Oral leukoplakia and malignant transformation. A follow up study of 257 patients. Cancer. 1984;53:563–568.

[28] Dr. Aparna Pandya. "Oral Leukoplakia." Published on August 30, 2022, [URL: https://dentistchannel.online/oral-leukoplakia/article?for=dentist].

[29] Dr. Mahima Krongold. "What Causes Mouth Ulcers?" Published on April 28, 2023, [URL: https://familydentalclinic.com.au/news/what-causes-mouth-ulcers/].

[30] Li, Z., Yang, W., Peng, S., & Liu, F. (2020). A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. arXiv preprint arXiv:2004.02806.

[31] Uzzal Podder. "How to include a custom filter in a Keras based CNN?" Asked on August 20, 2018, Modified on August 27, 2019, [URL: https://stackoverflow.com/questions/51930312/how-to-include-a-custom-filter-in-a-keras-based-cnn].

[32] Manav Mandal. "Introduction to Convolutional Neural Networks (CNN)." Updated on November 16th, 2023. [URL: https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/].

[33] Iakubovskii, P. (2019). Segmentation Models. GitHub repository. Retrieved from \url{https://github.com/qubvel/segmentation_models}

[34] Livne, Michelle et al. "A U-Net Deep Learning Framework for High Performance Vessel Segmentation in Patients With Cerebrovascular Disease." Frontiers in Neuroscience, vol. 13, 2019, doi: 10.3389/fnins.2019.00097.

[35] Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks. arXiv preprint arXiv:1909.09586.

[36] Dolphin, Rian. "LSTM Networks | A Detailed Explanation: A Comprehensive Introduction to LSTMs." Towards Data Science, Oct 22, 2020.

[37] Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., & Zhang, L. (2018). Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[38] Anantzoid. "VQA-Keras-Visual-Question-Answering." GitHub, 2023, github.com/anantzoid/VQA-Keras-Visual-Question-Answering.

[39] K. Phosri, T. Treebupachatsakul, W. Chomkwah, T. Tanpatanan, B. Thanathornwong, S. P. Khovidhunkit, and S. Poomrittigul, "The Comparison of Deep Learning Model Efficiency for Classification of Oral White Lesions," 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), Phuket, Thailand, 2022, pp. 1-4