

Real-Time Image Classification for Malignant Biliary Strictures on Cholangioscopy
Images Based on Deep Learning Approach



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering in Computer Engineering

Department of Computer Engineering

FACULTY OF ENGINEERING

Chulalongkorn University

Academic Year 2022

Copyright of Chulalongkorn University

การจำแนกรูปภาพของภาวะท่อน้ำดีอุดตันที่สงสัยมะเร็งแบบทันทีผ่านภาพจากการส่องกล้องภายใน
ท่อน้ำด้วยวิธีการเรียนรู้เชิงลึก



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Thesis Title	Real-Time Image Classification for Malignant Biliary Strictures on Cholangioscopy Images Based on Deep Learning Approach
By	Mr. Passakron Phuangthongkham
Field of Study	Computer Engineering
Thesis Advisor	Associate Professor PEERAPON VATEEKUL, Ph.D.
Thesis Co Advisor	Associate Professor PHONTHEP ANGSUWATCHARAKON, M.D.

Accepted by the FACULTY OF ENGINEERING, Chulalongkorn University in
Partial Fulfillment of the Requirement for the Master of Engineering

..... Dean of the FACULTY OF
ENGINEERING
(Professor SUPOT TEACHAVORASINSKUN, D.Eng.)

THESIS COMMITTEE

..... Chairman
(Professor BOONSERM KJSIRIKUL, Ph.D.)

..... Thesis Advisor
(Associate Professor PEERAPON VATEEKUL, Ph.D.)

..... Thesis Co-Advisor
(Associate Professor PHONTHEP ANGSUWATCHARAKON,
M.D.)

..... Examiner
(PUNNARAI SIRICHAROEN, Ph.D.)

..... External Examiner
(Thanapat Kangkachit, Ph.D.)

ภาสกร พ่วงทองคำ : การจำแนกรูปภาพของภาวะท่อน้ำดีอุดตันที่สงสัยมะเร็งแบบทันทีผ่านภาพจากการส่องกล้องภายในท่อน้ำดีด้วยวิธีการเรียนรู้เชิงลึก. (Real-Time Image Classification for Malignant Biliary Strictures on Cholangioscopy Images Based on Deep Learning Approach) อ.ที่ปรึกษาหลัก : รศ. ดร.พีรพล เวทีกุล, อ.ที่ปรึกษาร่วม : รศ. นพ.พรเทพ อังศุวัชรการ

การระบุสาเหตุของการตีบตันของท่อน้ำดีว่าเป็นมะเร็งหรือไม่นั้นเป็นเรื่องยาก ซึ่งในปัจจุบันการส่องกล้องตรวจท่อน้ำดีแบบใช้ digital single-operator cholangioscopy ช่วยให้แพทย์ส่องกล้องตรวจท่อน้ำดีได้แม่นยำยิ่งขึ้นจึงสามารถตรวจพบรอยโรคในท่อน้ำดีได้โดยตรงด้วยตาของตนเอง อย่างไรก็ตามยังคงมี การวินิจฉัยที่ไม่สอดคล้องกันของแพทย์ส่องกล้อง ดังนั้นการตรวจชิ้นเนื้อจึงถือเป็นมาตรฐานทองคำในการวินิจฉัยโรคนี ซึ่งหากมีความผิดพลาดในการนำชิ้นเนื้อออกจากท่อน้ำดีอาจทำให้ต้องทำการตัดชิ้นเนื้อใหม่อีกครั้ง ในวิทยานิพนธ์นี้เราได้เสนอเครือข่ายประสาทเทียมที่ออกแบบมาโดยเฉพาะสำหรับการจำแนกการตีบตันของท่อน้ำดีแบบทันที เราได้ทำการพัฒนาแบบจำลองของเราให้สามารถจำแนกรอยโรคออกมาได้ว่าเป็นมะเร็งหรือไม่ใช่มะเร็ง อีกทั้งแบบจำลองของเรายังสามารถบอกจุดของรอยโรคเพื่อที่จะสามารถตัดชิ้นเนื้อออกมาตรวจสอบได้โดยไม่ต้องใช้ข้อมูลที่บอกตำแหน่งในรูปภาพแต่ใช้แค่ประเภทของรูปภาพเท่านั้น เรายังคิดค้น guide wire augmentation ขึ้นมาเพื่อลดปัญหาของแบบจำลองที่ไปสงสัยรูปภาพที่มีอุปกรณ์และบังคับให้มองหาเนื้อเยื่อที่เป็นรอยโรคมมากขึ้น อีกทั้งเราได้นำแบบจำลองที่ได้ไปวัดผลต่อในรูปแบบของวีดีโอและออกแบบวิธีการใช้แบบจำลองเพื่อเพิ่มประสิทธิภาพในวีดีโอส่องกล้องจริง ในการทดลองของเรา เราจะวัดผลด้วยข้อมูล 3 ชุดโดยแบ่งข้อมูลตามคนไข้ เราได้รับข้อมูลจาก ศูนย์ส่องกล้องโรงพยาบาลจุฬาลงกรณ์โดยมีข้อมูลผู้ป่วยทั้งหมด 104 คน ได้รูปภาพมาทั้งหมด 885 รูป โดยประสิทธิภาพของแบบจำลองสามารถทำได้ที่ 0.8577 และ 0.8395 ในรูปแบบของ sensitivity และ F1 ตามลำดับและสามารถทำความเร็วได้ที่ 83 เฟรมต่อวินาที

สาขาวิชา วิศวกรรมคอมพิวเตอร์

ปีการศึกษา 2565

ลายมือชื่อนิสิต

ลายมือชื่อ อ.ที่ปรึกษาหลัก

ลายมือชื่อ อ.ที่ปรึกษาร่วม

6470419421 : MAJOR COMPUTER ENGINEERING

KEYWORD: Deep learning, Indeterminate biliary strictures, Real-time image classification, Explainable deep learning

Passakron Phuangthongkham : Real-Time Image Classification for Malignant Biliary Strictures on Cholangioscopy Images Based on Deep Learning Approach. Advisor: Assoc. Prof. PEERAPON VATEEKUL, Ph.D. Co-advisor: Assoc. Prof. PHONTHEP ANGSUWATCHARAKON, M.D.

It is challenging to determine if the cause of bile duct strictures is benign or malignant. Currently, endoscopists may more precisely inspect the bile duct thanks to computerized single-operator cholangioscopy. As a result, lesions in the bile duct can be seen with the naked eye. However, endoscopists continue to diagnose patients differently. Consequently, a biopsy is typically regarded as the gold standard. The necessity to repeat operations results from a biopsy sample mistake that results in a false-negative cancer diagnosis. In this study, we suggest a convolutional neural network developed particularly for real-time malignant biliary stricture classification. Our approach, which relies purely on an image-level label rather than annotation position, can produce output for both categorization and showing sections of tissue. An augmentation known as "guide-wire augmentation" makes the model focus on tissues rather than equipment, like a guide wire. Our model for still images has been updated to use video inference. All models in our experiment are performed on three patient-based bootstraps. The collection includes 885 images and 104 patient records from King Chulalongkorn Memorial Hospital. The model's sensitivity and F1 performance for still images are 0.8577 and 0.8395, respectively. With a speed of 83 frames per second, the model can be used for real-time inference.

Field of Study: Computer Engineering

Student's Signature

Academic Year: 2022

Advisor's Signature

Co-advisor's Signature

ACKNOWLEDGEMENTS

The research grant funds have been provided by the 72nd Anniversary of His Majesty King Bhumibol Adulyadej Scholarship and the 90th Anniversary Chulalongkorn University Fund (Ratchadapiseksomphot Endowment Fund), Assoc. Prof. Dr. Peerapon Vateekul was an outstanding supervisor, and I'd want to thank him for all of his hard work and encouragement. His advice and comments were extremely beneficial in overcoming any difficulties I had. Additionally, I am grateful to the Centre of Excellence in Gastrointestinal Oncology at Chulalongkorn University and the National Research Council of Thailand for funding this study. I'd want to extend my gratitude to the skilled endoscopists, particularly Assoc. Prof. Dr. Phonthep Angsuwatcharakon, Dr. Santi Kulpatcharapong, and Rungsun Rerknimitr, for the information they've provided, and the time and effort they've invested. Their advice and insightful observations on cholangioscopy helped us accomplish great research. My thank is given to Mr. Phanukorn Sunthornwetchapong for his contribution to deploying the AI model, his original source code and comment is very useful and practical use. Thank Mr. Pasit Jakkrawankul for his suggestion and constructive comments on crucial parts of our proposed method. Last but not least, my thank is given to my family and my girlfriend for several terms of support during my studies which include mental support and financial support, etc. My thank is also given to my friend, Mr. Passin Pornvoraphat and Mr. Mam Sothornin, etc., for helping and encouraging me throughout my studies. Finally, thank Data Mind Laboratory for facilitating the location and high-end GPU to complete this thesis.

Passakron Phuangthongkham

TABLE OF CONTENTS

	Page
.....	iii
ABSTRACT (THAI).....	iii
.....	iv
ABSTRACT (ENGLISH).....	iv
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS.....	vi
LIST OF TABLES.....	ix
LIST OF FIGURES.....	x
CHAPTER I INTRODUCTION.....	1
1.1 Aims and Objectives.....	2
1.2 The Scope of Work.....	3
1.3 Research Funding.....	3
1.4 Publication.....	3
CHAPTER II BACKGROUND.....	4
2.1 Image classification.....	4
2.1.1 Convolution layer.....	5
2.1.2 Pooling.....	5
2.1.3 Activation function.....	6
2.1.4 Fully connected.....	6
2.2 Supervise learning.....	7
2.2.1 Loss function.....	7

2.2.2 Optimizer	7
2.3 Evaluation Metrics	8
2.4 Cholangioscopy.....	9
CHAPTER III RELATED WORKS	11
3.1 Model network	11
3.1.1 Resnet (2016).....	11
3.1.2 Xception (2017).....	12
3.1.3 EfficientNet (2019).....	14
3.1.4 PYLON (2022).....	15
3.2 Medical image classification.....	16
3.3 Explainable deep learning.....	17
3.4 Real time medical image classification.....	19
3.5 Indeterminate biliary stricture image classification	20
CHAPTER IV CONCEPT AND RESEARCH METHODOLOGY.....	22
4.1 Data preparation.....	22
4.1.1 Still image dataset.....	22
4.1.2 Video dataset.....	23
4.2 Data augmentation	24
4.2.1 Cut image augmentation.....	24
4.2.3 Jigsaw augmentation.....	25
4.3 Model improvement.....	26
4.4 Video inference for biliary stricture	28
4.5 Model deployment.....	29
4.6 Model evaluations	30

4.6.1 Still image evaluation	30
4.6.2 Video evaluation	30
CHAPTER V EXPERIMENTS AND RESULTS	31
5.1 Comparing model result	31
5.2 Ablation study for modifying pylon	32
5.3 Comparing Effect of guide wire augmentation	33
5.4 Comparing effect of jigsaw augmentation	35
5.5 Video classification result	35
5.5.1 heatmap result	35
5.5.2 video classification result	36
5.6 Comparison of experts results	36
5.7 Heatmap output from our model	38
5.8 Error analysis	40
5.8.1 Image analysis	40
5.7.2 Videos analysis	43
5.8 Model deployments	49
5.8.1 Heatmap overlay	50
5.8.2 Contour overlay	51
CHAPTER VI CONCLUSION	52
REFERENCES	54
VITA	59

LIST OF TABLES

	Page
Table 1. The performance comparison between our model and other model on our testing set, Boldface refers to the winner	32
Table 2. The effect of modify PYLON from original on our test dataset, effnet refers to EfficientNet, boldface refers to the winner.....	33
Table 3. Effect of guide wire cut in augmentation on our model, boldface refers to winner	33
Table 4. Effect of Jigsaw augmentation on our model, boldface refers to winner.....	35
Table 5. Comparative of video classification on testing data, boldface refers to winner	36
Table 6. Comparison between our model and two expert endoscopists on the testing of still images and videos. Boldface refers to the winner.....	37

LIST OF FIGURES

	Page
Figure 1. an overview of the deep convolutional neural network architecture [14]....	4
Figure 2. Convolutional operation on widthxheightx3 input and 3x3x3 filter [15].....	5
Figure 3. GAP and Max pooling from the 4x4 feature with pooling 2x2 [16].....	6
Figure 4. overview of the instrument used for cholangioscopy [18].....	9
Figure 5. Example of X-ray image from ERCP [18].....	10
Figure 6. Example of SpyGlass® Cholangioscopy image that has RGB image and higher resolution.....	10
Figure 7. building a Residual block [19].....	12
Figure 8. the modified depthwise separable convolution with n=3 [24]	13
Figure 9. An overall of Architecture of Xception model [22].....	14
Figure 10. (a) baseline network, (b)-(d) are width scaling, depth scaling and resolution scaling respectively, (e) Compound scaling [25]	15
Figure 11. PYLON's architecture [27].....	16
Figure 12. The procedure of generating class activation map (CAM).....	18
Figure 13. (a) the example of malignant label images (b) the example of benign label images	23
Figure 14. Dataset preparation for training model in still image and evaluating model per fold.....	23
Figure 15. typical augmentation (a) normal image, (b) rotation , (c) horizontal flip, (d) translation, (e) auto contrast.....	24
Figure 16. (a) normal image (b) cut guide wire and paste in for augmentation.....	25
Figure 17. Jigsaw augmentation example. (a) original image (b) 2 x 2 jigsaw ratio (c) 4 x 4 jigsaw ratio (d) 5 x 8 jigsaw ratio (e) 10 x 16 jigsaw ratio.	26

Figure 18. Architecture of our model compared to the original PYLON: (A) Our model's architecture was enhanced from PYLON in 3 parts: (1) update the backbone from ResNet50 to be EfficientNetB3, (2) add the prediction head in the encoder, and (3) maintain the decoder to generate heatmaps and modify the prediction head here as auxiliary head. (B) The original PYLON's architecture.	27
Figure 19. The operation of the video classification algorithm.	29
Figure 20. effect of augmentation on CAM, (a) normal images	34
Figure 21. The result heatmap from model (a) normal image (b) second output from model which is 64x64 resolution image (c) Mapping 64x64 resolution image to heatmaps for visualization.....	36
Figure 22. Comment about heatmap from the endoscopist.....	38
Figure 23. Comment about heatmap from the endoscopist (cont).	39
Figure 24. Grad-Cam from True positive.....	40
Figure 25. Grad-Cam from True negative.....	41
Figure 26. Grad-Cam from False positive.....	42
Figure 27. Grad-Cam from false negative.	43
Figure 28. The true positive plotting between frame and malignant score.	45
Figure 29. The true negative plotting between frame and malignant score.	46
Figure 30. The false positive plotting between frame and malignant score.	47
Figure 31. The false negative plotting between frame and malignant score.	48
Figure 32. Overview of UI design for deployment.....	49
Figure 33. A heatmap overlay is shown on the left of the screen.	50
Figure 34. Heatmap overlay on main screen of the experiment.	50
Figure 35. A contour overlay is shown on the left side of the screen.	51
Figure 36. Contour overlay on main screen of the experiment.....	51



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

CHAPTER I

INTRODUCTION

Bile duct cancer, also known as cholangiocarcinoma (CCA), is an uncommon cancer that is rarely found in the global population but is prevalent in Thailand [1]. Moreover, the mortality rate for CCA is relatively high, with a five-year survival rate of 5% [2].

Malignant biliary stricture diagnosis is still an essential topic in cancer diagnosis since the differentiation between cancer and non-cancer has various aspects [3]. It has been reported that 25% of patients with biliary stricture that perform biopsies was not found malignancy [4].

The technologies utilized in diagnostics Whether a radiological examination or a laboratory examination has been consistently invented and developed [5]. Digital single-operator (DSOC) is a diagnostic technology that allows endoscopists to see the picture of the bile duct directly, which improves the distinction between benign and malignant lesions [6]. DSOC defines malignant features as dilated and tortuous vessels, irregular papillary projection, infiltrative lesion, ulceration, and polypoid nodule or mass [7]. However, low levels of interobserver agreement (IOA) and uneven performance have been found among experienced endoscopists categorizing cholangioscopy images of biliary strictures [8, 9].

Deep learning with convolutional neural networks (CNN) is a powerful method in machine learning for analyzing data patterns. Currently, CNN is used for a variety of tasks in the real world, as well as medical terms. They develop CNN to analyze the disease and determine whether it is abnormal or not [10]. CNN is also used in endoscopy to find cancer patterns or to assist endoscopists in classifying when performing endoscopy [11].

Cholangioscopy is one of the tasks in endoscopy that focuses on the bile duct. In 2021, CNN is applied to finding malignancy in cholangioscopy image through spyglass DSOC [12]. As a remarkable result, they achieved 95% overall accuracy. Anyway, these works are based solely on still images and may not be applicable in real-world scenarios. Recently studies [13] show how to deal with real-world scenarios for assisting diagnosis by using real case videos for evaluation with a moving average of predicted malignancy in 900 frames and achieve impressive values of 0.933 and 0.906 in terms of sensitivity and accuracy. However, an endoscopist may perform a biopsy to prove the malignancy of the bile duct. In the case of biopsies, the result will be more precise if the model provides the location of the malignancy in real-time.

The purpose is to enhance the performance of a model that is used to aid experts during cholangioscopy more practical and effective by classifying malignant and benign cholangioscopy images on our dataset. Following are the contributions: (1) Enhancement the classification model that provide heatmap more accurate in our specific real dataset (2) Ablation tests for guide wire paste in augmentation demonstrate that the model can deal with lesions more generally and precisely. (3) propose the algorithm to apply the model for real-world scenarios with more efficiency.

1.1 Aims and Objectives

To propose a deep learning model that provides real-time classification and heat maps to assist endoscopists during cholangioscopy.

1.2 The Scope of Work

1. Evaluate the proposed deep learning network addition to the following
 - a. Experiment on our private dataset of Biliary strictures from the Center of Excellence for Innovation and Endoscopy in Gastrointestinal Oncology, Chulalongkorn University, Thailand.
 - b. Cholangioscopy images and videos from our dataset were acquired by experienced endoscopists.
2. The proposed network can classify malignancy biliary stricture from biliary stricture.
3. The inference speed of real-time classification is more than 25 fps.

1.3 Research Funding

This research project was funded by the National Research Council of Thailand (NRCT; N42A640330), Chulalongkorn University (CU-GRS-64), and Chulalongkorn University (CU-GRS-62-02-30-01) and supported by the Center of Excellence in Gastrointestinal Oncology, Chulalongkorn University annual grant. It was also funded by the University Technology Center (UTC) at Chulalongkorn University. Additionally, The research grant funds have been provided by the 72nd Anniversary of His Majesty King Bhumibol Adulyadej Scholarship and the 90th Anniversary Chulalongkorn University Fund (Ratchadapiseksomphot Endowment Fund).

1.4 Publication

- P. Phuangthongkham, P. Angsuwatcharakon, S. Kulpatcharapong, P. Vateekul and R. Rerknimitr, "Real-Time Identification of Malignant Biliary Strictures on Cholangioscopy Images Using Explainable Convolutional Neural Networks With Heatmaps," in *IEEE Access*, vol. 11, pp. 49943-49956, 2023, doi: 10.1109/ACCESS.2023.3276642.
- *IEEE Access*, Q1
- Impact Factor = 3.476

CHAPTER II

BACKGROUND

The background knowledge for the thesis is covered in this chapter. It is consisting of Image classification, supervise learning, Data Augmentation, Evaluation Matrix, and Cholangioscopy.

2.1 Image classification

Image classification is the process of labeling or categorizing an entire image. Images should only include a single class. The model receives input as images, then extracts crucial features from the images and shows the output as a class. In deep learning, there are many important modules for Image classification

Convolutional Neural Networks (CNNs), layering makes CNNs strong. CNNs simultaneously process red, green, and blue image components using a three-dimensional neural network. This requires fewer artificial neurons to analyze a picture than feed forward neural networks.

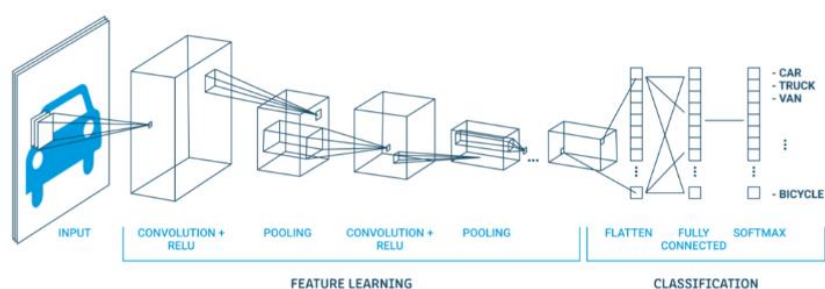


Figure 1. an overview of the deep convolutional neural network architecture [14]

Typically, a convolutional network's design has four types of layers: convolution, pooling, activation, and fully connected.

2.1.1 Convolution layer

Convolution layer it is comprised of a set of convolutional filters also known as kernels. The filter convolved feature matrices as input to produce a features map as output. This kernel is a weight for the model and will be change after optimizing. After training, this filter will be the data pattern filter. The mathematical in this layer start with each filter uses a different channel input value to multiply the weights. The sum of all the inputs gives a different value for each filter position. This operation shows in Figure 2.

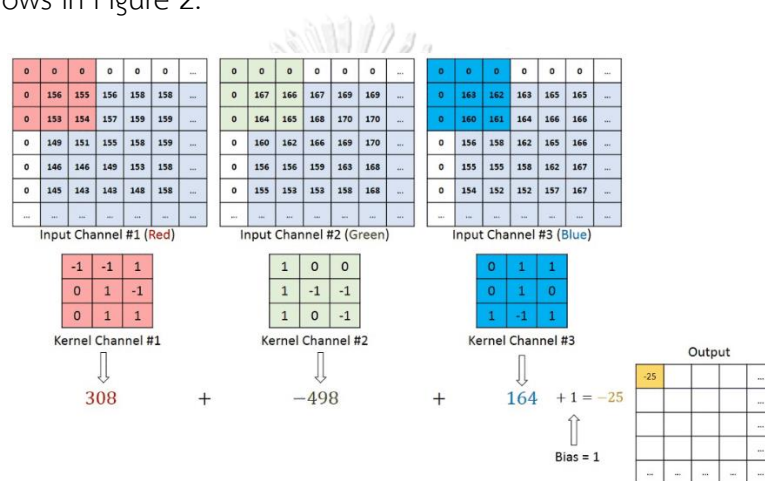


Figure 2. Convolutional operation on widthxheightx3 input and 3x3x3 filter [15]

2.1.2 Pooling

The pooling layers progressively shrink the image size, retaining just the most vital details. These important features are determined by the method of pooling. Max, min, and GAP pooling are the most common types of pooling, see figure 3 show the example of pooling.

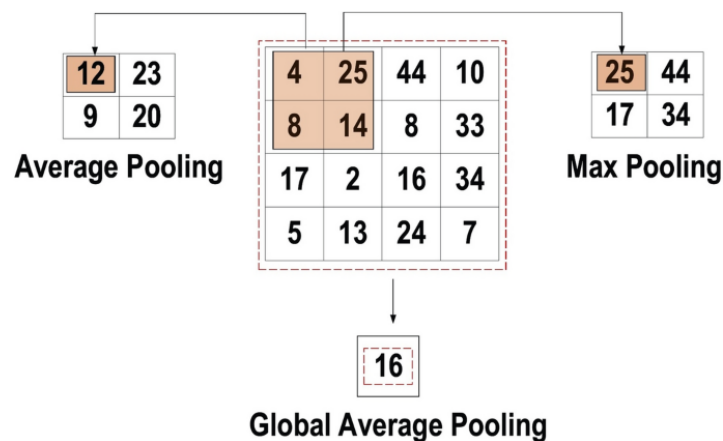


Figure 3. GAP and Max pooling from the 4x4 feature with pooling 2x2 [16]

2.1.3 Activation function

Activation function is the function that use for change value to non-linear value, since nonlinear function having the capacity to discriminate, which is a crucial feature, the most frequently applied to the model are sigmoid function in Sigmoid and ReLU. Sigmoid can produce the output between 0-1 when input is the real number. The equation can be expressed equation follows:

$$f(x)_{\text{sigm}} = \frac{1}{1+e^{-x}} \quad (1)$$

ReLU is the most often employed function within the CNN. It transforms the input values to positive values and lower computing cost when compute back propagation. The equation can be expressed equation follows:

$$f(x)_{\text{ReLU}} = \max(0, x) \quad (2)$$

2.1.4 Fully connected

This layer is frequently added near the end of any CNN architecture. Each neuron in this layer is connected to every neuron in the previous layer. When we compute this layer, we refer to the fundamental method of traditional multilayer perceptron neural networks. The input of these layer is refined from previous layer

such as convolutional layer, pooling and activation which is flattened. The output of the last FC layer also referred CNNs output.

2.2 Supervise learning

Supervised learning is part of machine learning and artificial intelligence. Model is learned by input and label output pairs. The process of learning is diverse and varies based on the type of model. For this work, we use deep learning for image classification. The training procedure entails calculating loss from the model's output using a loss function. Losses will be calculated to adjust the model's weight in order to minimize the loss itself. To do so, we'll need an optimizer to compute gradient descent and adjust weight during the model's training process.

2.2.1 Loss function

The loss function is stand for loss from the output of the model. In this work, use loss function for categorized two class. Therefore, binary, and categorical cross entropy is applied for the main loss:

$$L = - \sum_{n \in N} \sum_{c \in C} W_c \log \frac{\exp(x_{n,c})}{\sum_{i \in C} \exp(x_{n,i})} y_{n,c} \quad (3)$$

Where $y_{n,c}$ is a pair of labels, $x_{n,c}$ is the model output, W_c is the weight of all class, and N is a sample from minibatch

2.2.2 Optimizer

Optimizer is used for finding the minimum of the loss. If the loss is minimum, we said that is the best weight of the model. To do that, we start with compute loss from the output, then compute the gradient. In mathematical, the gradient has a

direction that points from the lower point to the higher point. For finding minimum, we must change to the opposite direction by add minus to the equation. Thus, this kind of action is gradient descent, we use gradient descent for update weight whole model with the chain rule, the equation of update rule is explained as follows:

$$\omega_{t+1} = \omega_t - \alpha \frac{\partial L}{\partial \omega_t} \quad (4)$$

Where ω_t denote to weight, α is learning rate, $\frac{\partial L}{\partial \omega_t}$ is gradient of the loss.

2.3 Evaluation Metrics

Due to model must categories biliary strictures in to two classes. If model output and label are Malignant, the result is true positive (TP). In other hand, if those output is Benign, the result is true negative (TN). False positive (FP) is defined when output is malignant, but label is benign. Likewise, False negative is defined when output is benign, but label is malignant, the evaluation metrics of this work are accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and F1-score. The evaluation metrics are described below:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$recall_{Malignant}(sensitivity) = \frac{TP}{TP+FN} \quad (6)$$

$$recall_{Benign}(specificity) = \frac{TN}{TN+FP} \quad (7)$$

$$Precision_{Malignant}(PPV) = \frac{TP}{TP+FP} \quad (8)$$

$$Precision_{Benign}(NPV) = \frac{TN}{TN+FN} \quad (9)$$

$$F1 = \frac{2 \times \text{recall} \times \text{Precision}}{\text{recall} + \text{precision}} \quad (10)$$

2.4 Cholangioscopy

Cholangioscopy is an endoscopic method that does not involve cutting into the body. It is used to look at the bile ducts visually and treat them at the same time [17]. There is two types are represented below

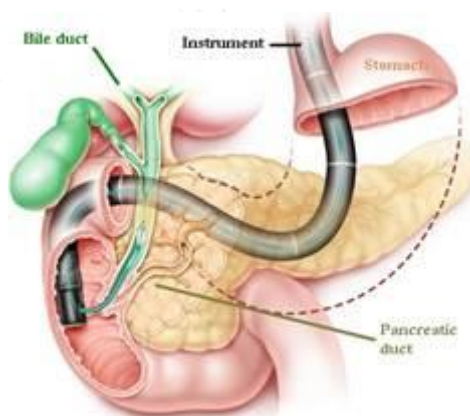


Figure 4. overview of the instrument used for cholangioscopy [18]

Endoscopic Retrograde Cholangio Pancreatography (ERCP) is a radiographic examination of the bile ducts (small drainage tubes), gallbladder, and/or pancreatic duct performed in real-time. ERCP assists your gastroenterologist in diagnosing and treating numerous biliary illnesses, such as bile duct obstruction owing to stones or cancer, or pancreatic disorders, such as pancreatitis or bile duct cancer. However, ERCP only provide x-ray images show in figure 5. With this limitation, it hard obtains biopsy and to find out whether biliary stricture is cancer.

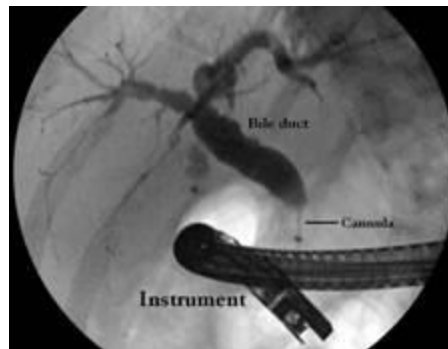


Figure 5. Example of X-ray image from ERCP [18]

Spyglass® cholangioscopy is the attachment for the basic ERCP. With this tool, endoscopists can directly observe the bile ducts and acquire a more accurate biopsy, the picture is shown in figure 6. With higher resolution and RGB images. In this work, we use this image as a dataset for proposed model.



Figure 6. Example of SpyGlass® Cholangioscopy image that has RGB image and higher resolution

CHAPTER III

RELATED WORKS

This chapter describes the relevant deep neural networks for the thesis, medical image classification, cut image data augmentation, Real time medical image classification and recently work that relate to the Indeterminate biliary stricture image classification.

3.1 Model network

Over the past few years, deep neural networks are consistently developed. Model is more accuracy and efficiency, also practical for using in real-world problem solving. This section aims to provide information on the network used for this thesis. The detail is represented following.

3.1.1 Resnet (2016)

In 2016, Model named ResNet [19] is proposed. This model affects the underlying structure of several models by proposed the residual learning framework show in figure.7. Typically, neural networks layers will feed forward layer to layer, but residual block not only feed forward but also directly feed the input skip to next layer. Due to skip connection, the vanishing gradient from gradient descent is no longer much affected since back propagation is calculated through the input of the layer.

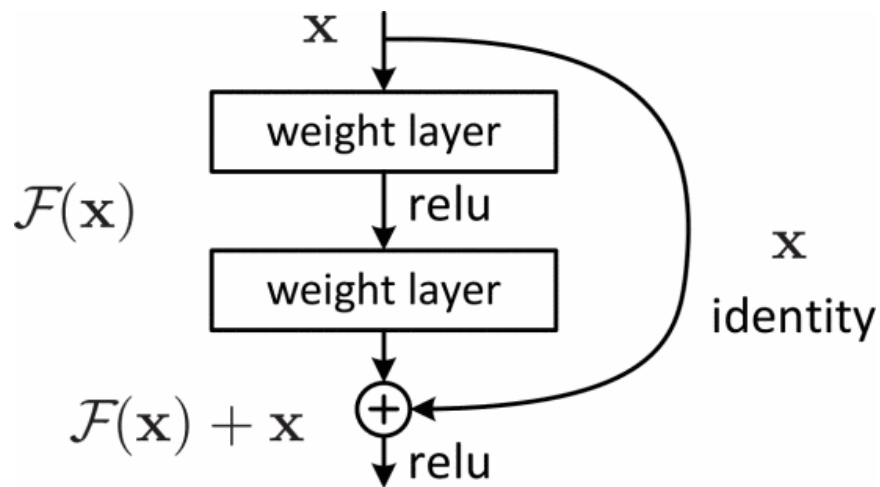


Figure 7. building a Residual block [19]

When compare with the more complexity neural networks such as VGG [20], the result from the ImageNet dataset [21] confirmed the model with residual block is better. Testing on ILSVRC 2015 classification competition, ResNet [19] has 3.57% top-5 error, while VGG [20] has 7.32% top-5 error.

3.1.2 Xception (2017)

Xception [22] is represented as advance version of Inception [23], by modified depthwise separable convolution module. It achieves high performance from ImageNet dataset [21] by 0.790 in terms of accuracy when VGG-16 [20] and ResNet-152 [19] achieve 0.715 and 0.770 respectively.

Firstly, depthwise separable convolution module is consist of depthwise convolution and Pointwise convolution respectively. Depthwise convolution the channel-wise $n \times n$ spatial convolution. For example, if it has 10 channels, the module will have 10 $n \times n$ spatial convolution. Pointwise convolution is 1×1 convolution to make the dimension we need.

To modify depthwise separable convolution module for Xception model such that it is not significantly different from the original module, Swap the point

wise in front of the depthwise and pull intermediate activation out from the module
Figure 8 represents the modified depthwise separable convolution.

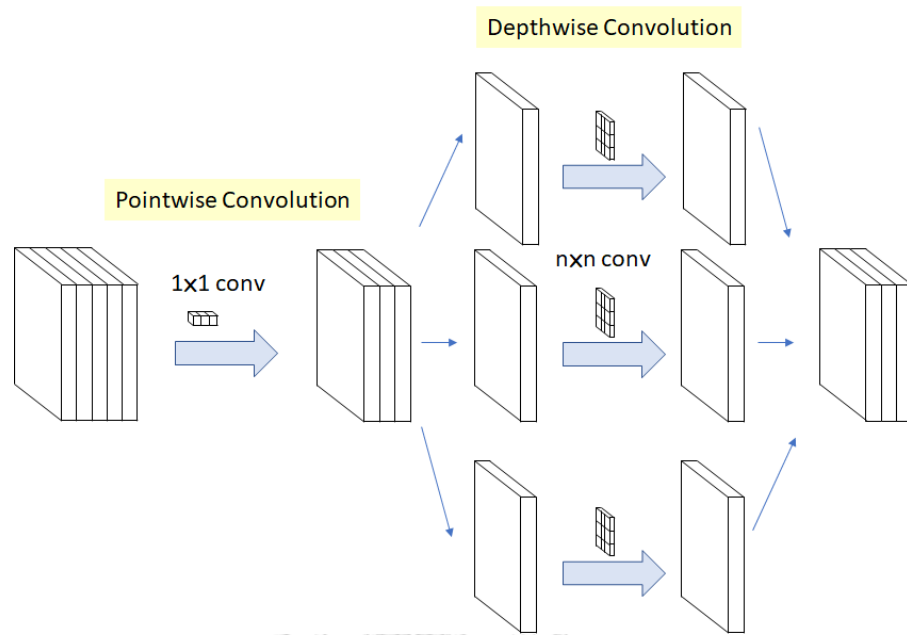


Figure 8. the modified depthwise separable convolution with $n=3$ [24]

The overall architecture in Xception model [22] are describe in Figure 9. The architecture is divided in 3-part, Entry flow, Middle flow and Exit flow. The modified depthwise separable convolution is illustrate as SeparableConv.

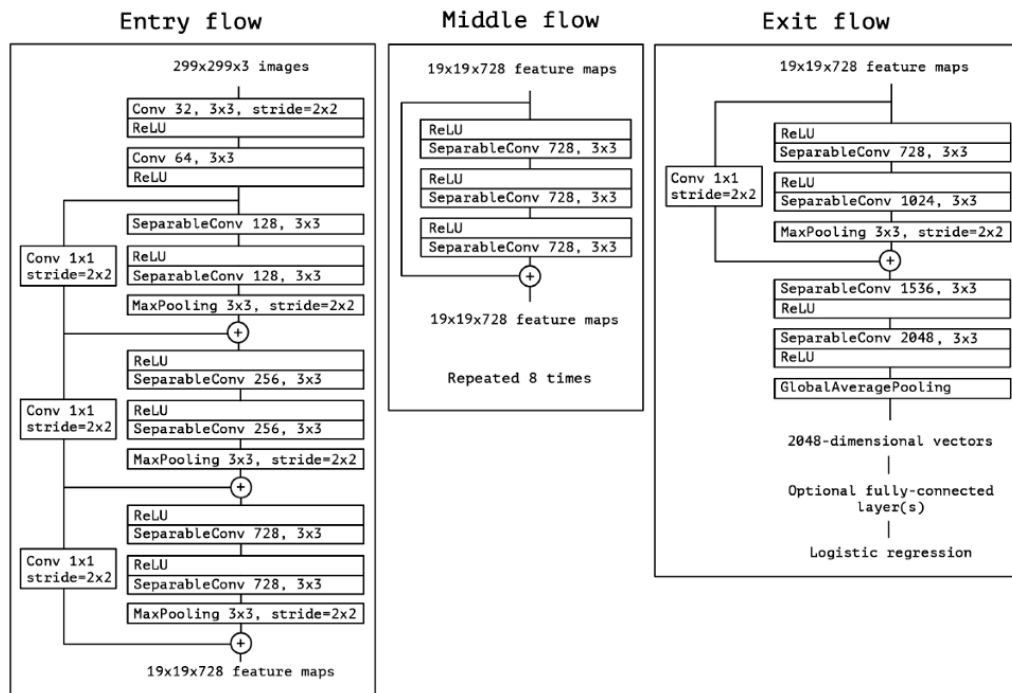


Figure 9. An overall of Architecture of Xception model [22]

3.1.3 EfficientNet (2019)

In the past, for improve model efficiency, model must scale up something such as layers, resolution of image and width(channel) such as ResNet [19] that can scale up model from ResNet18 to ResNet200. Likewise, accuracy of the model increases from scaling up layers. However, this phenomenon needs manual adjustment and considerable time, leading frequently in little or no performance enhancement. EfficientNet [25] is a result from Compound model scaling method show in figure 10, that scale CNN model with width, depth, and image resolution.

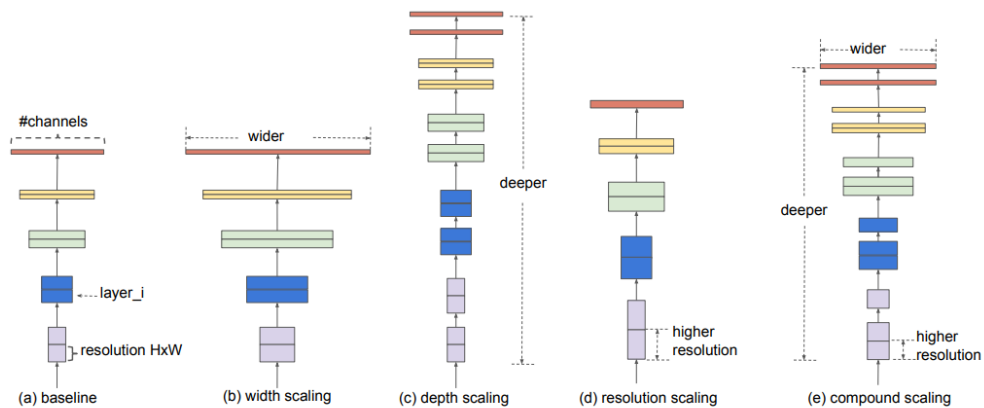


Figure 10. (a) baseline network, (b)-(d) are width scaling, depth scaling and resolution scaling respectively, (e) Compound scaling [25]

The compound scaling approach is based on the concept of balancing width, depth, and resolution through constant ratio scaling. In addition, the compound scaling improved the model efficiency and accuracy of earlier CNN models such as MobileNet [26] and ResNet [19] by around 1.4% and 0.7% ImageNet accuracy, respectively, compared to other random scaling techniques. The architecture of the EfficientNet employs a mobile inverted bottleneck and scaled up to create a EfficientNetB0 – B7, EfficientNetB7 is the biggest model from family and achieve impressive results is 84.4% top-1 accuracy and 97.3% top-5 accuracy on ImageNet.

3.1.4 PYLON (2022)

Pyramid localization Network (PYLON) [27], this model aims to improve resolution of Heatmaps by CAM method, in fact, PYLON does not require expert annotation of label position and may instead be trained with solely image-level labels. This functionality is particularly crucial for domains where expert annotation is frequently unavailable or expensive. For the output, PYLON has two outputs, the first one is classification output model, the second one is heatmaps that process through upsampling module (UP) and pyramid attention (PA), These two modules allow

PYLON compute Heatmap with CAM method in high resolution. The model architecture is described in figure 11.

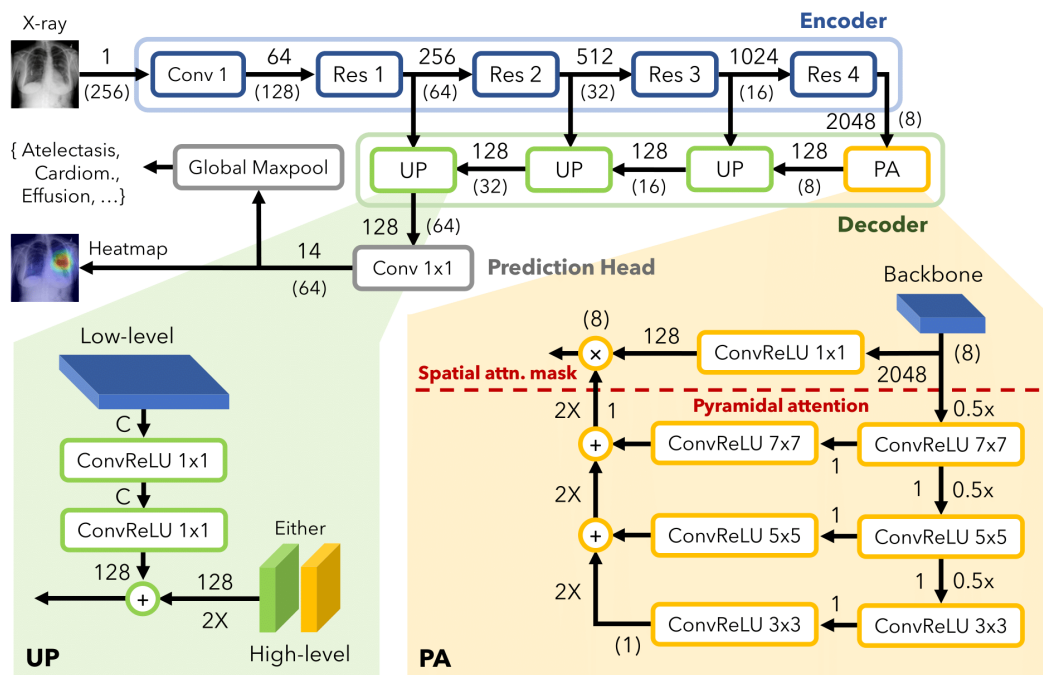


Figure 11. PYLON's architecture [27]

Encoder is a ResNet backbone used for encoding to embedded vectors and then feeding to PA, this module consists of ConvReLU layer with several filter size, ConvReLU is convolutional, ReLU activation and Batch normalization. PA is focus on find the crucial features and send to the UP module. Up module is used for increase image size when image size is 64 x 64 then feed to the conv 1x1 for shrink channel to class label.

3.2 Medical image classification

Deep learning image classification has been used in a variety of tasks over the last decade. Medical tasks are some of the most successful in the deep learning era [10]. In gastrointestinal disease, diagnosis by GI endoscopy also contributes

consistently [11]. In 2018 Chen PJ et al [28], proposed using a deep neural network to classify two types of polyps with a size less than 5 mm. They had 1476 images for neoplastic polyps and 681 images for hyperplastic polyps, which they benchmarked with an endoscopist. They achieve high performance that is 96.3, 78.1, 89.6, and 91.5 in terms of sensitivity, specificity, PPV, and NPV respectively, which is more than endoscopist's NPV ranging from 73.9% to 94% from six person.

In that year, Jun-Yan He et al. [29] propose work related to hookworm detection through wireless capsule endoscopy (WCE) image, they use two networks of CNN to classify whether patient is infected by hookworm. The first one is built for edge extraction to refine the second CNN feature. The second one is based model for hookworm classification. They have 4,828 hookworm images and 436,796 images for non-hookworm which is pretty imbalanced, however, their method does not miss any infected patients. Their method reaches 0.895 in terms of ROC-AUC while compare with GoogLeNet [23] and AlexNet [30] have 0.883 and 0.769 in terms of ROC-AUC respectively.

In 2020, Poundel et al. [31], presented a method for classifying colorectal disease on their own dataset with five classes. By modify original ResNet with adding dilated rated for convolution and add DropBlock to the model, their data was collected 364 images for adenocarcinoma, 775 images for adenoma, 563 images for Crohb's disease, 773 images for ulcerative colitis and 770 images for normal. Their model achieves an impressive F1-score of 0.93, while other methods achieves F1-score ranging from 0.87 to 0.91

3.3 Explainable deep learning

CNN is a combination of layer, we don't know what is going within a plenty of mathematical value inside that layers and what that layers stand for, it is truly black box. In 2016, B. Zhou et al. proposed a method for understanding the "black box" of image classification called "Class activation maps" (CAM) [32]. This method modifies

CNN architectures for image classification by using global average pooling (GAP), the procedure is map back the predicted class to previous convolution layer and sum all weight in layer to produce a class activation map. Describe in figure 12.

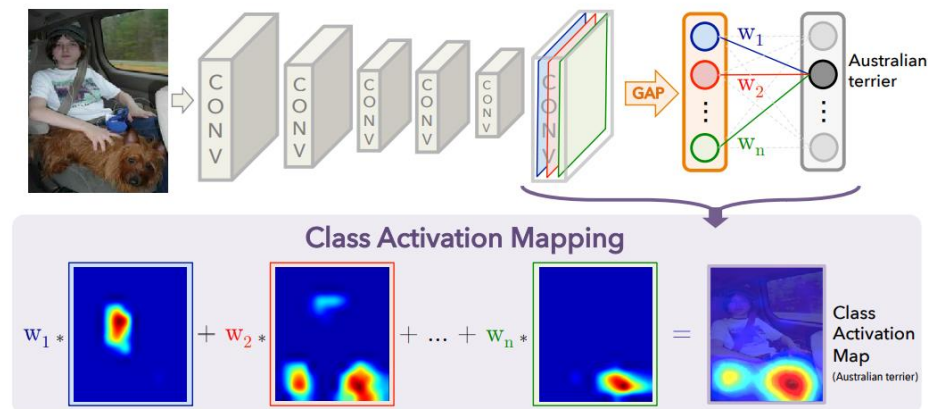


Figure 12. The procedure of generating class activation map (CAM)

The resolution of cam depends on size of image in that layer which always low resolution. However, some tasks need more resolution, such as medical tasks for aid expert in classification tasks.

In 2022, Preechakul et al. [27] proposed the Pyramid Localization Network (PYLON), which generates more resolution through an upsampling module and a pyramid attention module. PYLON does not need to map output and sum weight like traditional CAM. They proposed the PA module for finding localization of the image, and then used the upsampling module to generate more resolution of the CAM and give output in term of classification result and heat-map, they perform PYLON in NIH's Chest X-ray14 [33], that Dataset is good for evaluating accuracy because it has more than 100,00 images, which 1,000 of them were annotated with bounding boxes of disease location. For model evaluation, they compare their model with CAM [32], Grad-CAM [34], Grad-CAM++ [35], XGrad-CAM [36], Li et al.'s method [33], and FPN [37]. Their method achieves 0.65 in terms of weight average

point localization accuracy, which is more than other methods that range from 0.06 to 0.61 in a 512x512 input image.

3.4 Real time medical image classification

Some tasks in a real-world situation, especially in medical terms, required real-time assistance, for more precise and accurate diagnosis. Endoscopy is one of them. In 2018, Michael et al. [38] developed a model based on the Inception [23] model for classifying diminutive adenomas from hyperplastic polyps and provided an algorithm for use in real-time situations by video evaluation, this algorithm adds credibility value that was calculated frame by frame in form of exponential smoothing. For training model, they comprised video to frame by 223 videos, validating 40 videos, and testing 125 videos. The model accuracy was 94%, sensitivity was 98%, and NPV was 97% in 106 polyps video testing, which is high confidence in prediction calculated by credibility. However, this model had poor value in term of frame per second (FPS) which is 20 FPS,

In 2022, Yi Lu et al. [39] proposed a model for classifying the histopathology of colorectal polyps, using ResNeSt [40] as a model for image classification. In the dataset, they divided it into 5 classes: hyperplastic or inflammatory polyps, adenomatous polyps, intramucosal cancer, deep submucosal invasive cancer, and normal mucosa. Split the dataset with 5-fold cross validation, which is 7,032 images for polyps and 3,541 images for normal mucous, and leave 116 consecutive polyp videos for test performance of model. They refine image features by adding an edge channel to the input, which is computed by edge extraction, the overall accuracy of the model is 93%. For video testing, their model achieves 84.62%, 86.27%, and 85.34% in terms of sensitivity, specificity, and accuracy, respectively. However, this work does not provide any algorithms for more practical real-case scenarios in video testing, which may add more accuracy and be more practical.

3.5 Indeterminate biliary stricture image classification

Indeterminate biliary stricture is one of the new challenging tasks for the deep learning approach. Saraiva MM et al. developed CNN-based on DSOC images for detecting malignancy in biliary strictures in 2021 [12], obtaining 9,695 malignant images and 2,160 benign images from 85 patients. For the deep learning approach, they employed Xception [22] for diagnostics and evaluated model by splitting the dataset for 5-fold cross-validation. Their model reaches high performance, with an overall accuracy of 94.9%, a sensitivity of 94.7%, a specificity of 92.1%, and an AUC of 0.988. However, this study focuses only on still images, which may not aid the endoscopist in a real-time situation.

In 2022, Marya et al. [13] propose a deep learning model and algorithm for diagnosis in real-world scenarios. They benchmarked their model with two traditional techniques, brush cytology and forceps biopsy sampling, they employed ResNet50V2 [41] as a model for classification, and collected cholangioscopy images from 2012 to 2021, which totaled 2,388,439 still images with 154 patients. In the dataset, expert endoscopists are used to categorize and annotate images. This data is classified into five categories: high-quality malignant, high-quality benign, high-quality suspicious, and low-quality uninformative. They pick 14,381 images from the database for training and 5,348 images for testing, that image is from 132 patients and leave 22 patients for video testing. The result is impressive, which the model had high quality malignant AUROCs of 0.941. They use the moving average in video testing to predict whether a video is malignant or not based on the moving average result of high quality malignant. The overall accuracy of video testing is 0.906 in a 900 frame-average, which is much higher than brush cytology and forceps biopsy sampling, which are 0.625 and 0.609, respectively. However, this model provides a woefully inadequate real-time classification result, and endoscopists must sometimes perform a biopsy while performing cholangioscopy. This work is not an option for an AIDS endoscopist to perform a biopsy. For this limitation, our contribution is focused

on not only classification assistance but also providing real-time heatmaps of malignant in order to perform biopsies more precisely.



CHAPTER IV

CONCEPT AND RESEARCH METHODOLOGY

This chapter will illustrate about experiment setup and how this model work for real-world scenarios, this chapter are represented by 6 part which contain Data preparation, Data augmentation, Enhancement model, video inference for real case problem, model evaluation, and model deployment.

4.1 Data preparation

This private cholangioscopy image dataset was collected from 2014 to 2021 from the Center of Excellence for Innovation and Endoscopy in Gastrointestinal Oncology, Chulalongkorn University, Thailand. This data contains 104 patients, from patient get 885 still image data which is labeled by the diagnosis result and second screening by two expert endoscopist. In addition, video data also collected from the same source as image which contains 5 patients only for testing model by video. This work is split data by patient based in 3-fold without same patient is testing set.

4.1.1 Still image dataset

As mentioned above, dataset is divided in to 3-fold by patient based for prevent leakage data since still image from patient is not equal, it is ranging from 1 to 28 image, the label image was categorized in two class, 447 images for malignant and 438 images for benign, the class label image are shown in figure 13. For training model, we divided still image data by 70: 15: 15 per fold by patient based which contains 72 patients for training, 16 patients for validating, and 16 patients for testing. Prior to training, we also increased the number of sample training images by randomly duplicating images from patients who had fewer than five images to five images.

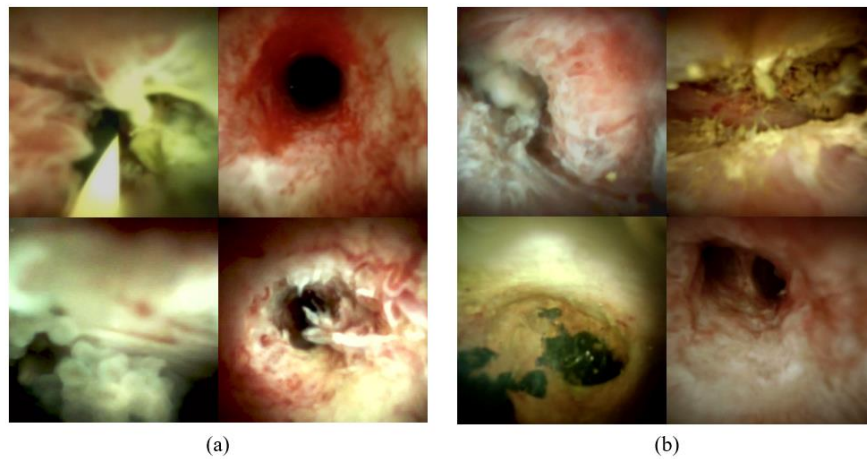


Figure 13. (a) the example of malignant label images (b) the example of benign label images

4.1.2 Video dataset

Some patients in still image data also have video data, we left rest of video data for solely testing. For testing video, we use test video from patient who divided to be testing in still image for preventing leakage data. Additionally, we have 5 patients with only have video data, we assign those 5 patients to testing video for get result in all 3-fold evaluations. Finally, we illustrate this data preparation in Figure 14

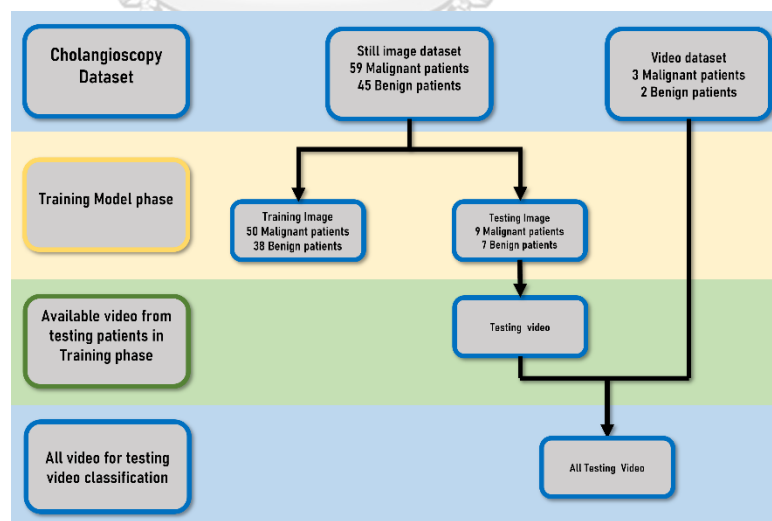


Figure 14. Dataset preparation for training model in still image and evaluating model per fold

4.2 Data augmentation

For make model more generalize, the most frequently used is data augmentation. The work of augmentation is generating more image with other perspective with computer vision method for training model, there are many methods for augmentation. In this work we applied typical method that always applies in deep learning is rotation, horizontal flip, affine transformation, and auto contrast which is shown in figure 15.

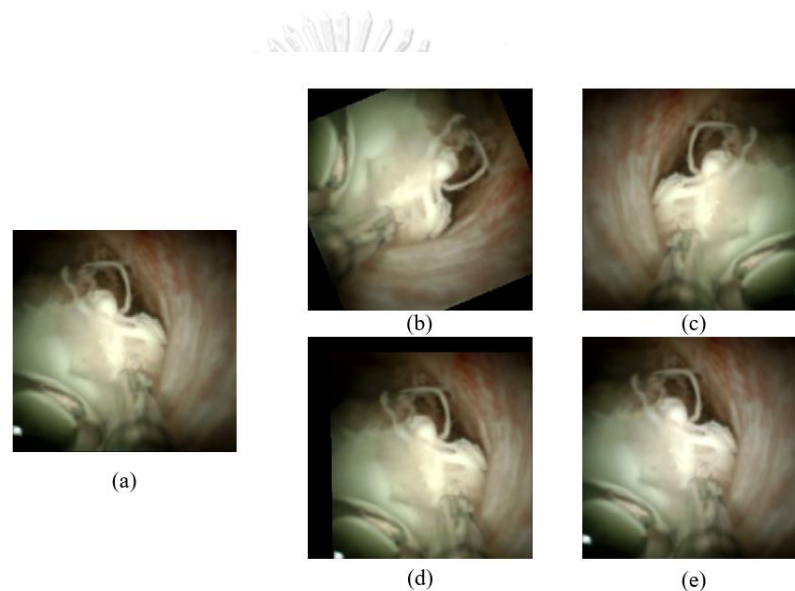


Figure 15. typical augmentation (a) normal image, (b) rotation , (c) horizontal flip, (d) translation, (e) auto contrast

4.2.1 Cut image augmentation

There is also work for hard augmentation with cut image and generate more generalization model. In 2019, Yun, Sangdoon, et al. [44], proposed cut mix augmentation. This augmentation is employed by cut image from some label image to another label image and weight that two class together. For our work, we can not employ this cut mix practically because some area of biliary stricture image cannot represent their class and sometime endoscopist must perform guide wire or other tools during cholangioscopy. In that case, we proposed specific augmentation

technique for biliary stricture image classification task, first we cut image from labeled image with applied guide wire and prepare for augmentation, second, we randomly paste in those cut image to the training image in vertical and make sure that guide wire not bigger than half of height or width of image the augmentation illustrate in figure 16.

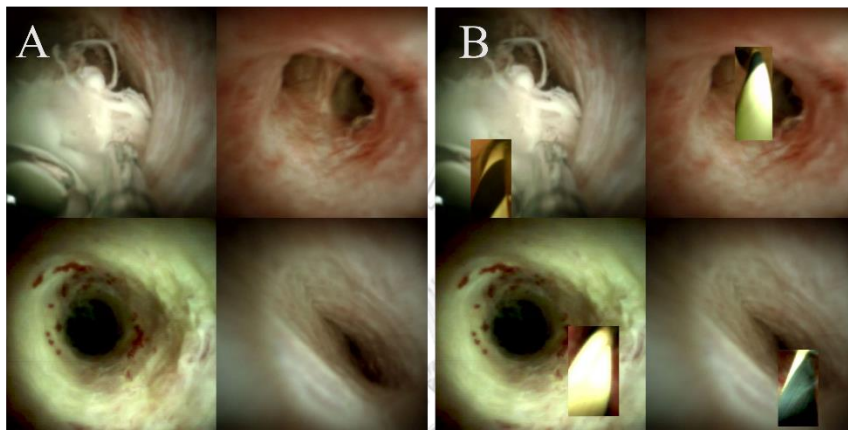


Figure 16. (a) normal image (b) cut guide wire and paste in for augmentation

4.2.3 Jigsaw augmentation

Jigsaw augmentation [42] was invented to destroy image structure due to the fact that sometimes image structure is the reason that a model has a location bias. The jigsaw technique splits an image into identically sized rectangular pieces before shuffling and assembling them to its original sizes, see figure 17. The procedure enables the model to establish a direct connection during training.

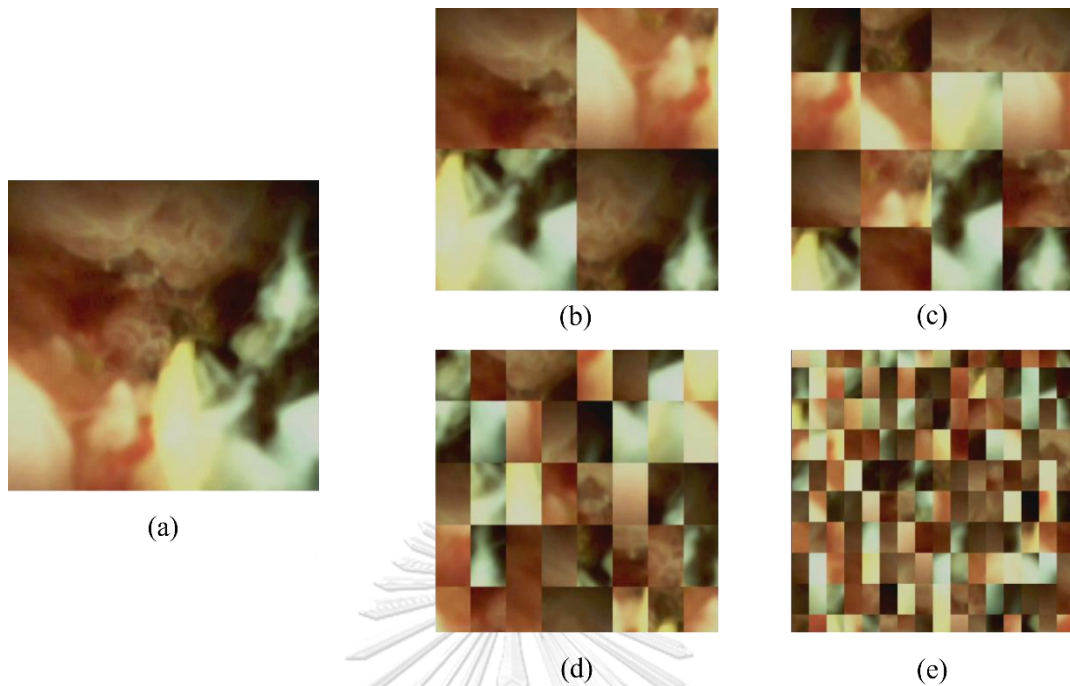


Figure 17. Jigsaw augmentation example. (a) original image (b) 2 x 2 jigsaw ratio (c) 4 x 4 jigsaw ratio (d) 5 x 8 jigsaw ratio (e) 10 x 16 jigsaw ratio.

4.3 Model improvement

PYLON is very good for this task since model provide not only classification but also heatmaps that necessary for help endoscopist perform biopsy more precisely. However, this model lack of image classification efficiency, thus, we propose our model that based on PYLON that will raise more efficiency but still provide heatmaps correctly, the model architecture illustrates in figure 18.

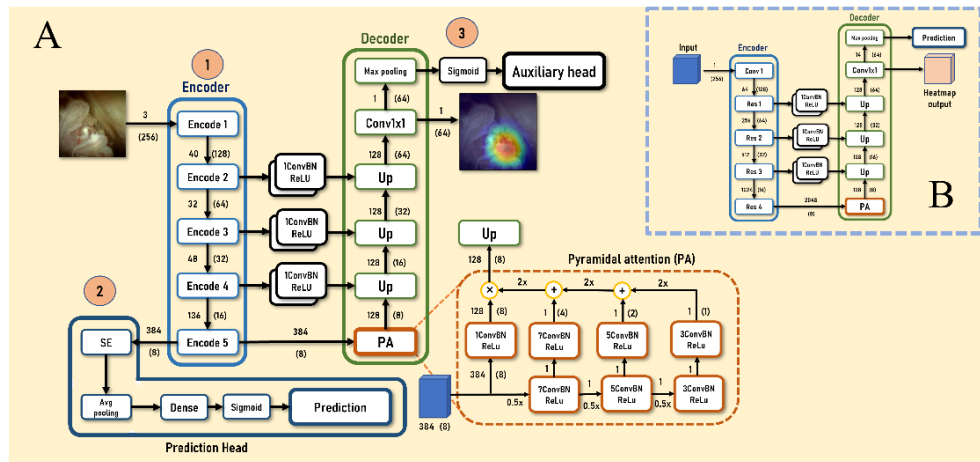


Figure 18. Architecture of our model compared to the original PYLON: (A) Our model's architecture was enhanced from PYLON in 3 parts: (1) update the backbone from ResNet50 to be EfficientNetB3, (2) add the prediction head in the encoder, and (3) maintain the decoder to generate heatmaps and modify the prediction head here as auxiliary head. (B) The original PYLON's architecture.

According to figure 18, the essential module is contained three parts. Firstly, encoder module is use for encoding image to embedded vector, we use EfficientNetB3 as backbone. While compute encoder, model also send feature from encoder 2, encoder 3 and encoder 4 to decoder module and change channel by 2 stacks of 1ConvBNReLU (this module contains 1x1 conv, Batch normalization, and ReLU activation respectively). After encoding, model will separate to two paths. Prediction Head and Decoder, prediction head is stand for classification task which contain squeeze and executed block [42] that allow model made more capacity with channel attention, after refining features with SE block, model will send that feature to traditional classification head consist of Avg pooling, Dense, Sigmoid, and prediction with 0.5 threshold. Decoder block is use for generating malignant heatmaps of images. The two most important modules are pyramidal attention (PA) and UP module. PA is used for refining features more precisely and find crucial features, PA is built by 1ConvBNReLU, 7ConvBNReLU, 5ConvBNReLU, and 3

ConvBNReLU following the figure 18. Inside the PA, the features are reduced channel from 384 to 1 channel by max pooling and reduce resolution by convolution which show in figure 18 by 0.5x, and also interpolate resolution to combine with another features are shown in figure 18 by 2x, after PA module, model will send feature to UP module, in UP module, feature resolution will be up scaling by interpolate features from previous block and combine with encoder feature that up channel by 1convBNReLU. Lastly, the second output of model has 64x64 resolution, which can be mapping to heatmap. To train this model, we use two binary losses for optimizing model the first one is on prediction head for image classification and the second one is auxiliary head for generate output heatmaps, we combine losses following:

$$Loss = Loss_{prediction} + Loss_{Auxiliary} \quad (11)$$

where $Loss_{prediction}$ denote to BCE from prediction head and $Loss_{Auxiliary}$ is BCE from auxiliary head.

4.4 Video inference for biliary stricture

In real-world scenarios, we proposed algorithms for using model predicted patients, whether malignant or not, this algorithm is represented in figure 19. When endoscopist diagnosis biliary stricture while using spyglass DSOC cholangioscopy, they must see several perspectives for diagnosis the patient. Same as this algorithm, we use model for predict frame of video and average predicted frame, in 100 frames if malignant more than half, we will predict that patient is malignant. But if malignant is not exceed 50 frames, we predict that patient is benign. In real case, the endoscopist may stop to focus the lesion on bile duct, since this condition, predicted frame will increase unnecessary and it causes wrong prediction. For solving this problem, we add variance calculate to the algorithm, before feeding frame to model, we calculate the variance of that frame to make sure the image is changed, if new frame is more

or less than 5% variance from last predicted frame, we will feed that frame to the model and redo the process of prediction.

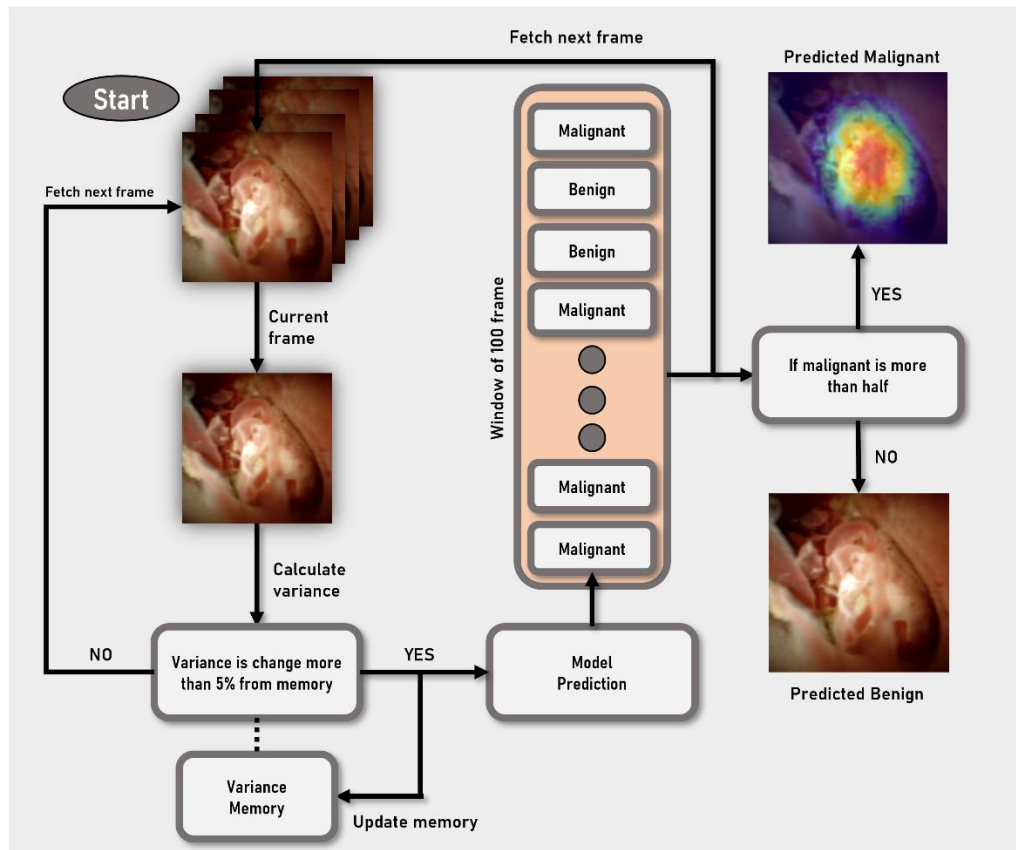


Figure 19. The operation of the video classification algorithm.

4.5 Model deployment

Model are optimized to open neural network exchange (ONNX) format for evaluation and deployment in the Center of Excellence in Gastrointestinal Oncology. Model in ONNX format is faster than typical PyTorch model formats, and it more practical using in real world scenarios.

4.6 Model evaluations

For model evaluation, we convert the model to ONNX format and inference on the TensorRT backend, whose computation is RTX3090. The computational complexity of all models was calculated in terms of multiply-adds operators (MAdds) and model parameters.

4.6.1 Still image evaluation

we use classification metric that mention in 2.3, evaluation only still image in test set in 3-fold and average them to final result, we also evaluate speed of the model in frame per second by start from receive uint8 image to predict class of that image.

4.6.2 Video evaluation

Same as above, we use classification metric that mention in 2.3, but instead of evaluating still image, we use video from the patient that mention in 4.1.2, evaluating per patient in classification metric, and evaluate speed test by frame per second from uint8 through our algorithm and end with prediction result.

CHAPTER V

EXPERIMENTS AND RESULTS

In this chapter, preliminary experiments about biliary stricture image classification and video classification from chapter 4 are explained

5.1 Comparing model result

For the best performance in prediction, we conduct experiments by training an image classification model with several different models. According to their work [21], EfficientNet is pretty fast and has high accuracy when compared with others. In addition, biliary stricture image classification studies [41, 42] have used Xception [18] and ResNet50v2 [43] in their work, lastly, we modified PYLON [23] to make our own model. For all that reason, we compare the performance of the models, which are EfficientNetB2, EfficientNetB3, Xception, ResNet50v2, PYLON, and our model. We separate the type of model by output, the first one provides only classification results, and the last one provides both classification results and heatmaps that will be used for the video classification experiment. All models were trained under the same condition, which is augmented by basic augmentation and cut image augmentation that are mentioned in 4.2.1. However, the input resolution was chosen based on their work. We use ImageNet as pre-train of all models and optimize model by AdamW as optimizer with 0.0007 for learning rate and 64 batch size for training, we train 150 epoch and choose model from best validation on F1-score in validation set, we pick model from that best validation and testing in test set. The result is shown in Table 1. our model achieves the highest in terms of sensitivity, NPV, F1-score, and accuracy, which are 0.8577, 0.8443, 0.8395, and 0.8415, respectively, For trading of FPS, our model can provide heatmaps in real time while FPS is 84.1.

Table 1. The performance comparison between our model and other model on our testing set, Boldface refers to the winner

Model	Sensitivity	Specificity	PPV	NPV	F1	Accuracy	FPS
ResNet50v2 [41]	0.7517	0.8660	0.8577	0.7661	0.8067	0.8084	188.6
Xception [22]	0.7807	0.8141	0.8223	0.7758	0.7954	0.7966	156.7
EfficientNetB2 [25]	0.7904	0.8560	0.8575	0.7908	0.8222	0.8233	181.8
EfficientNetB3 [25]	0.7879	0.8573	0.8624	0.7888	0.8220	0.8236	172.4
PYLON [27] (heatmaps)	0.7908	0.7658	0.7924	0.7949	0.7800	0.7842	86.4
Our model (heatmaps)	0.8577	0.8188	0.8418	0.8443	0.8395	0.8415	84.1

5.2 Ablation study for modifying pylon

As we mentioned in 4.3, we modified PYLON to our model by change original encoder backbone, adding prediction head, and add auxiliary loss for training CAM, before we found the best architecture, we also test other backbone, which is EfficientNetB2, EfficientNetB3, ResNet50 after change backbone we also test with effect of prediction head by original pylon and add prediction head to model the result are illustrate in Table 2.

Table 2. The effect of modify PYLON from original on our test dataset, effnet refers to EfficientNet, boldface refers to the winner.

Method	Sensitivity	Specificity	PPV	NPV	F1	Accuracy	FPS
PYLON (resnet50) (original PYLON)	0.8165	0.7594	0.7913	0.7888	0.7878	0.7900	86.4
PYLON (resnet50) + modify head	0.8333	0.7885	0.8188	0.8140	0.8117	0.8142	86.2
PYLON (EffnetB2) + modify head	0.7605	0.8433	0.8424	0.7626	0.8001	0.8012	84.5
PYLON (EffnetB3) + modify head	0.8582	0.7930	0.8281	0.8459	0.8276	0.8313	84.2
PYLON (EffnetB3) + modify head + SE (our model)	0.8577	0.8188	0.8418	0.8443	0.8395	0.8415	84.1
PYLON (EffnetB4) + modify head + SE	0.7525	0.8812	0.8864	0.7655	0.8149	0.8173	67.5
PYLON (EffnetB5) + modify head + SE	0.7668	0.8688	0.8728	0.7730	0.8157	0.8173	59.1
PYLON (EffnetB6) + modify head + SE	0.7395	0.8637	0.8567	0.7599	0.8021	0.8021	51.5
PYLON (EffnetB7) + modify head + SE	0.7264	0.8648	0.8635	0.7553	0.7949	0.7972	39.2

5.3 Comparing Effect of guide wire augmentation

From experiments, we know that our model is pretty good at predicting when compared with others under the same conditions. In this section, we show that our guide wire augmentation is one of the reasons that our model reaches 0.8395 in terms of F1-score, the effect of the augmentation result is shown in Table 3.

Table 3. Effect of guide wire cut in augmentation on our model, boldface refers to winner

Method	Sensitivity	Specificity	PPV	NPV	F1	Accuracy
Our model with guide wire augmentation	0.7858	0.8328	0.8415	0.7875	0.8086	0.8106
Our model without guide wire augmentation	0.8577	0.8188	0.8418	0.8443	0.8395	0.8415

Moreover, we also present effect of augmentation in terms of explanation by CAM result to show our where our model is considering from figure 20. Without augmentation, model is looking tools as benign class which is not lesion from biliary stricture. With augmentation, the model changes attention to the lesion, which is correctly predicted as malignant.

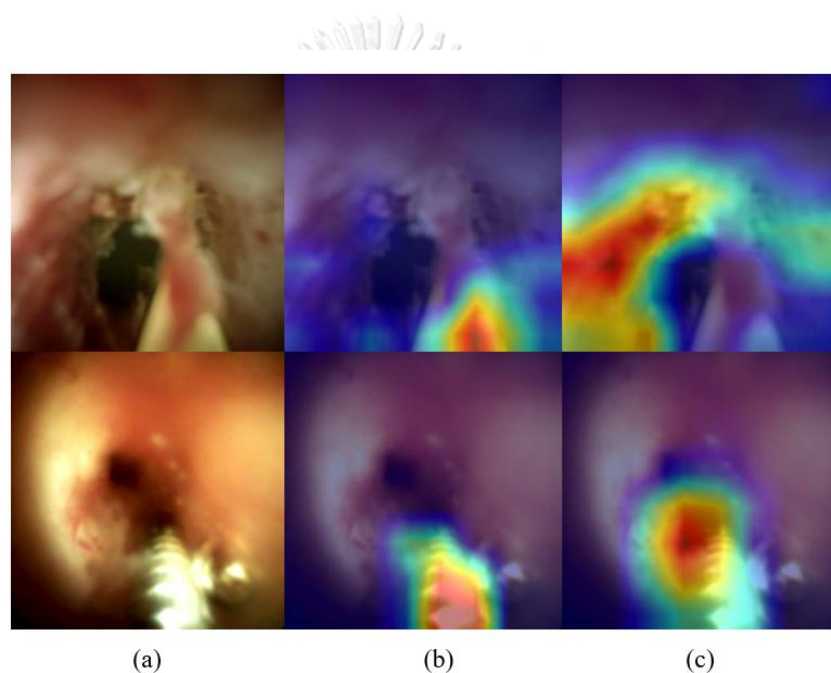


Figure 20. effect of augmentation on CAM, (a) normal images
(b) without augmentation (c) with augmentation

5.4 Comparing effect of jigsaw augmentation

The jigsaw augmentation will destroy the image structure that causes location bias in the model. Unfortunately, our data has only an image level label, so the model will be confused with the destroyed image. In Table 4, the results show that without applying jigsaw augmentation, the model has a better classification result.

Table 4. Effect of Jigsaw augmentation on our model, boldface refers to winner.

Method	Sensitivity	Specificity	PPV	NPV	F1	Accuracy
Our model with jigsaw 2x2	0.8236	0.8216	0.8387	0.8058	0.8217	0.8228
Our model with jigsaw 4x4	0.8380	0.8113	0.8349	0.8251	0.8262	0.8288
Our model	0.8577	0.8188	0.8418	0.8443	0.8395	0.8415

5.5 Video classification result

5.5.1 heatmap result

This model provide real-time heatmaps for assisting endoscopist classify malignant from biliary stricture and perform biopsy, heatmaps of malignant is shown in figure 21.

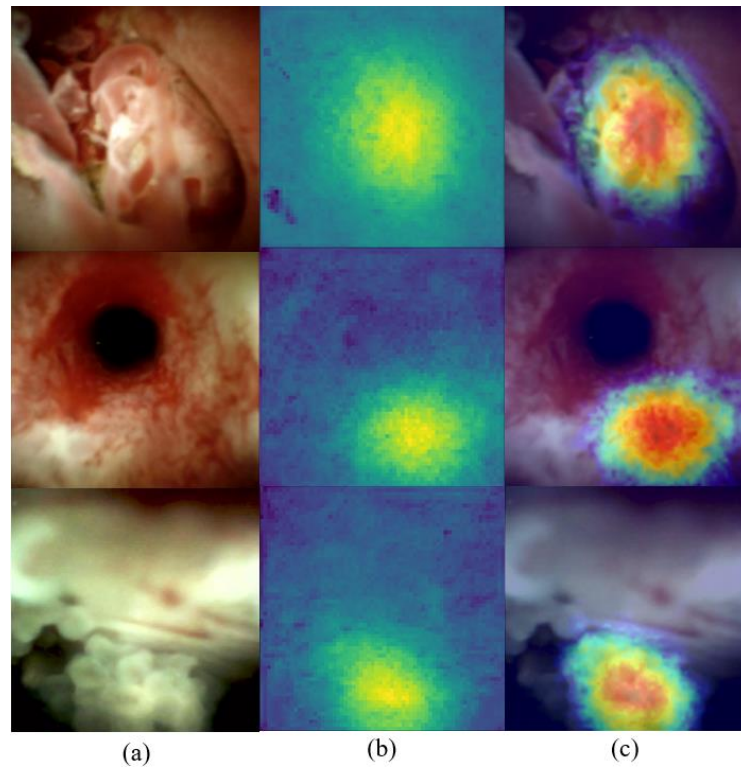


Figure 21. The result heatmap from model (a) normal image (b) second output from model which is 64x64 resolution image (c) Mapping 64x64 resolution image to heatmaps for visualization

5.5.2 video classification result

For evaluation of video classification algorithms, the testing data from Section 4.1.2 was used, and we compared our algorithms with a typical moving average. We show the result in Table 5. Our algorithms achieve high performance of 0.9024, 0.9394, 0.9333, 0.9154, 0.9193, and 0.9197 in terms of sensitivity, specificity, PPV, NPV, F1-score, and accuracy, respectively.

5.6 Comparison of experts results

Table 5. Comparative of video classification on testing data, boldface refers to winner

Model	Sensitivity	Specificity	PPV	NPV	F1	Accuracy	FPS
Moving average	0.8956	0.7765	0.7748	0.8998	0.8263	0.8297	83.0
Our algorithm	0.9024	0.9394	0.9333	0.9154	0.9193	0.9197	84.0

In this section, we represent the model's efficiency by comparing it with two expert endoscopists. Still image datasets and video datasets were prepared for the two experts. We carried out the experiment in the same manner as determined by the model. In Table 6, the prediction results from the two experts are illustrated. Surprisingly, with the still images dataset, our model is seen to provide more robustness than humans, achieving 0.8577, 0.8443, 0.8395, and 0.8415 in terms of sensitivity, NPV, F1, and accuracy, respectively. In addition to the video dataset, our model demonstrated an impressive performance more than the experts, achieving 0.9024, 0.9394, 0.9333, 0.9154, 0.9193, and 0.9197 in terms of sensitivity, specificity, PPV, NPV, F1, and accuracy, respectively.

Table 6. Comparison between our model and two expert endoscopists on the testing of still images and videos. Boldface refers to the winner.

Dataset	Classifier	Sen.	Spec.	PPV	NPV	F1	Acc.
Still images	Expert 1	0.7900	0.8422	0.8646	0.7834	0.8143	0.8149
	Expert 2	0.6932	0.8473	0.8375	0.7139	0.7663	0.7677
	Our model	0.8577	0.8188	0.8418	0.8443	0.8395	0.8415
Videos	Expert 1	0.7542	0.5000	0.5365	0.7250	0.6080	0.6106
	Expert 2	0.7913	0.9280	0.8843	0.8515	0.8604	0.8664
	Our model	0.9024	0.9394	0.9333	0.9154	0.9193	0.9197

As observed in Table 6, results reveal that humans are confused by benign since the sensitivity of the model exceeds the sensitivity of the two experts on both still images and videos. Therefore, the model plays an important role in assisting endoscopists while performing cholangioscopy.

5.7 Heatmap output from our model

This section, we present the heatmap output from the model. To confirm the heatmap output which is good enough to use in the real-world scenarios, the only way to confirm is biopsy which is the gold standard for this task. However, we cannot perform biopsy to assure the result due to clinical limitations. Thus, we will confirm the result in compromise way, we show the output of the heatmap then get the comment from the expert in 3 ways: strongly agree, agree, and disagree in figure 22, and figure 23.




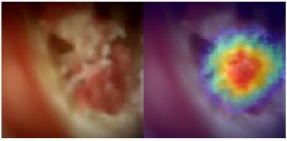
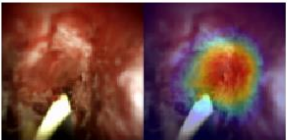
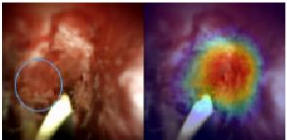
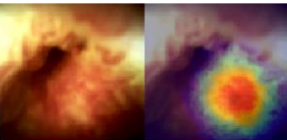
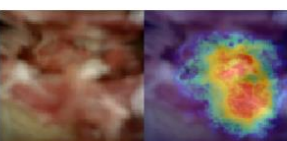
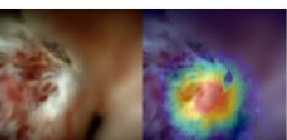
Image	Comment from endoscopist			The heatmap should also be this area
	Strongly agree	Agree	Disagree	
	✓			
		✓		
	✓			
	✓			
	✓			

Figure 22. Comment about heatmap from the endoscopist.

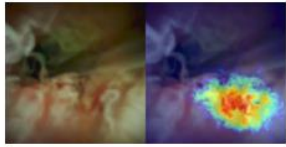
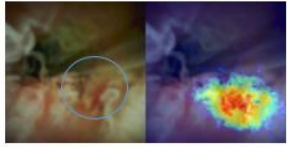
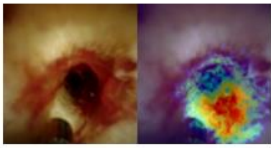
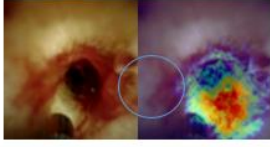
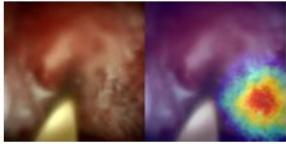
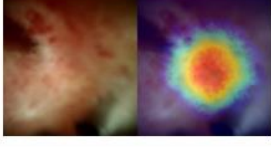
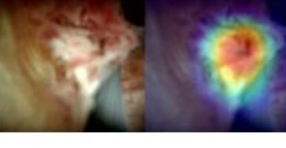
Image	Comment from endoscopist			The heatmap should also be this area
	Strongly agree	Agree	Disagree	
		✓		
		✓		
	✓			
	✓			
	✓			

Figure 23. Comment about heatmap from the endoscopist (cont).

5.8 Error analysis

In this section, we investigate the best model by looking at the image and video with human knowledge. An error from the model was presented.

5.8.1 Image analysis

We first investigate the testing image from bootstrap 1, and the results will be discussed in terms of true positive, true negative, false positive, and false negative. The first one is a true positive. From figure 24, the heat-map result, also known as the suspicious area, shows the lesion of the biliary stricture.

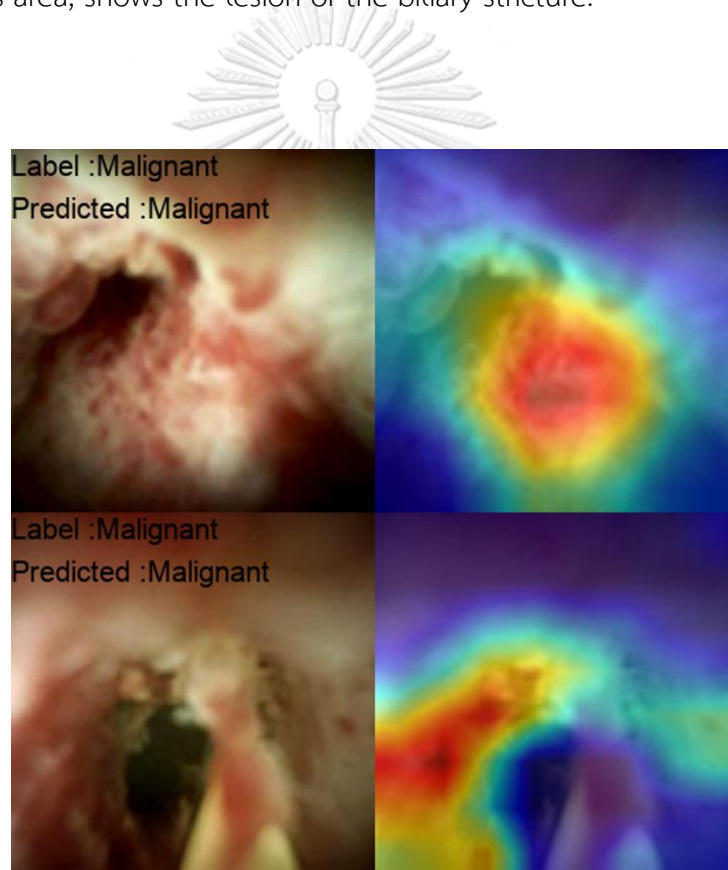


Figure 24. Grad-Cam from True positive.

In the second one, we focus on the true negative image. Figure 25: Due to the fact that the classifier is a binary classifier, the heat map of the image will show only suspicious areas. A true negative image will not find anything if it does not have a suspicious area. Sometimes the heatmap will show, but the confidence is not enough to recognize it as malignant.

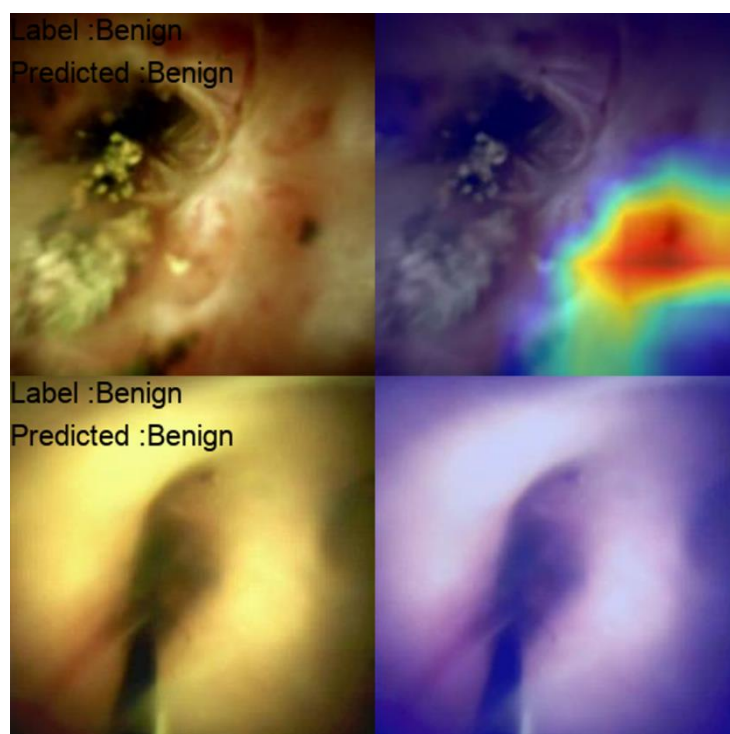


Figure 25. Grad-Cam from True negative.

In the third one, we investigate the false positive. Figure 26 shows the suspicious areas in the benign class that result in false positives. Some lesions are not the cause of malignancy. However, the image can confuse the model with the lesion, which seems malignant, and predict that lesion with high confidence.

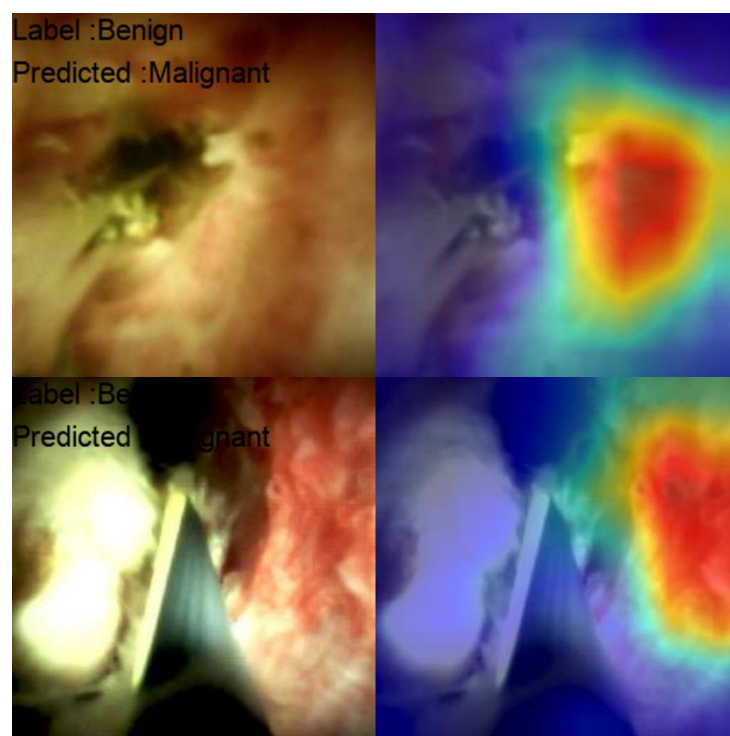


Figure 26. Grad-Cam from False positive.

The last one is a false negative. Figure 27 shows the heat-map with the Grad-Cam method. The model accurately looked at the malignant areas but was not confident enough to predict the image as malignant. Due to this reason, the images are considered benign, which causes false negatives.

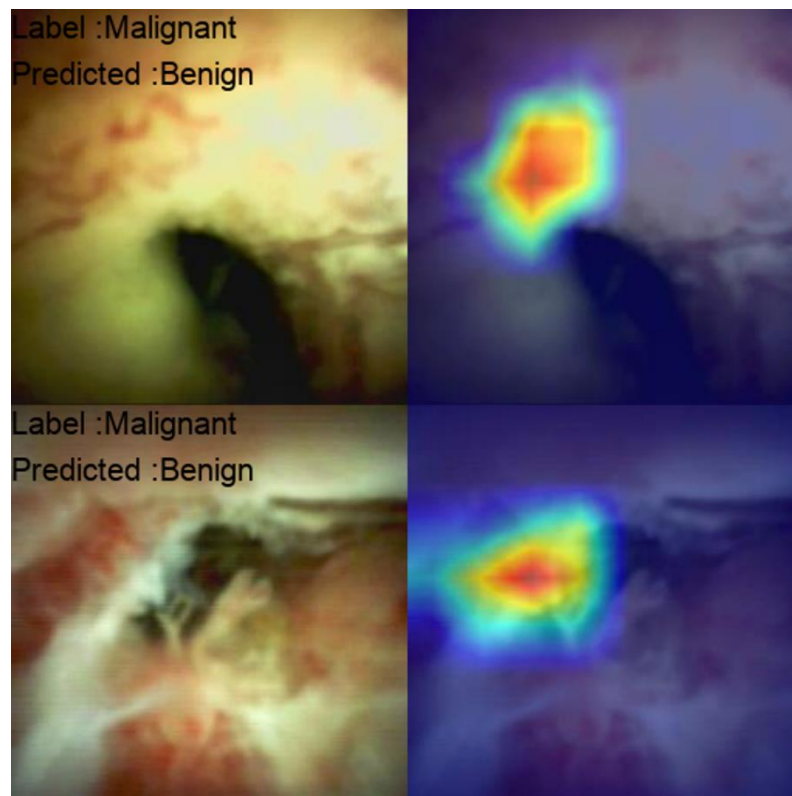


Figure 27. Grad-Cam from false negative.

5.7.2 Videos analysis

We also investigated the testing videos. The results will be discussed in terms of true positive, true negative, false positive, and false negative. In this section, we will show two graphs: the image predicted from the video frame by frame and the working of our algorithm on the videos.

We start at investigate true positive video. In figure 28 the result are show the prediction frame by frame and our algorithm. The video from the real-case come with the noisy images, when the image show the noisy the model will predict to be benign because there are no suspicious area. From the algorithm, we set threshold with 0.5 from malignant score. We can observed in the last part of the video the model focuses on the malignant area and score rise up more than 0.5 then we predict this video as malignant.

The second we show the true negative case in figure 29. The model mostly predict the frame as benign cause there is no suspicious areas and some of the frame come with the noisy image. The prediction of video is benign cause the malignant score does not exceed 0.5.

The third we present the false positive video prediction. From figure 30, the prediction start with malignant that cause the moving average high in the first place. Then the model is confused by the lesion in the bile duct with continuous predict as malignant until the malignant score exceed 0.5.

The last one we present the false negative video prediction. In figure 31, the very first part of the video come with the noisy until the model found the malignancy areas. However due to the length of the video, the malignant score from frame prediction almost touch 0.5 threshold in the last part of the video lead to the result with benign prediction.

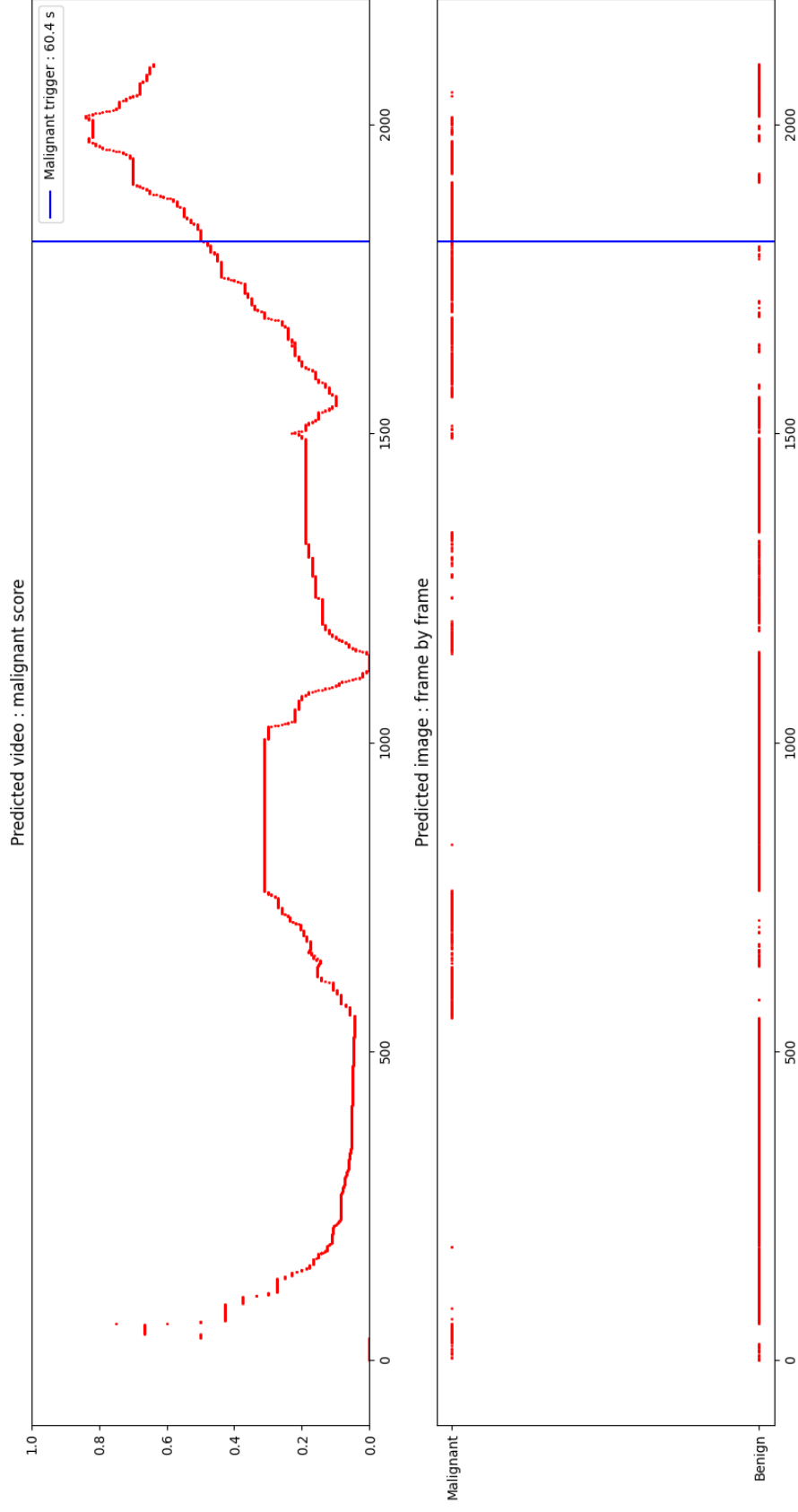


Figure 28. The true positive plotting between frame and malignant score.

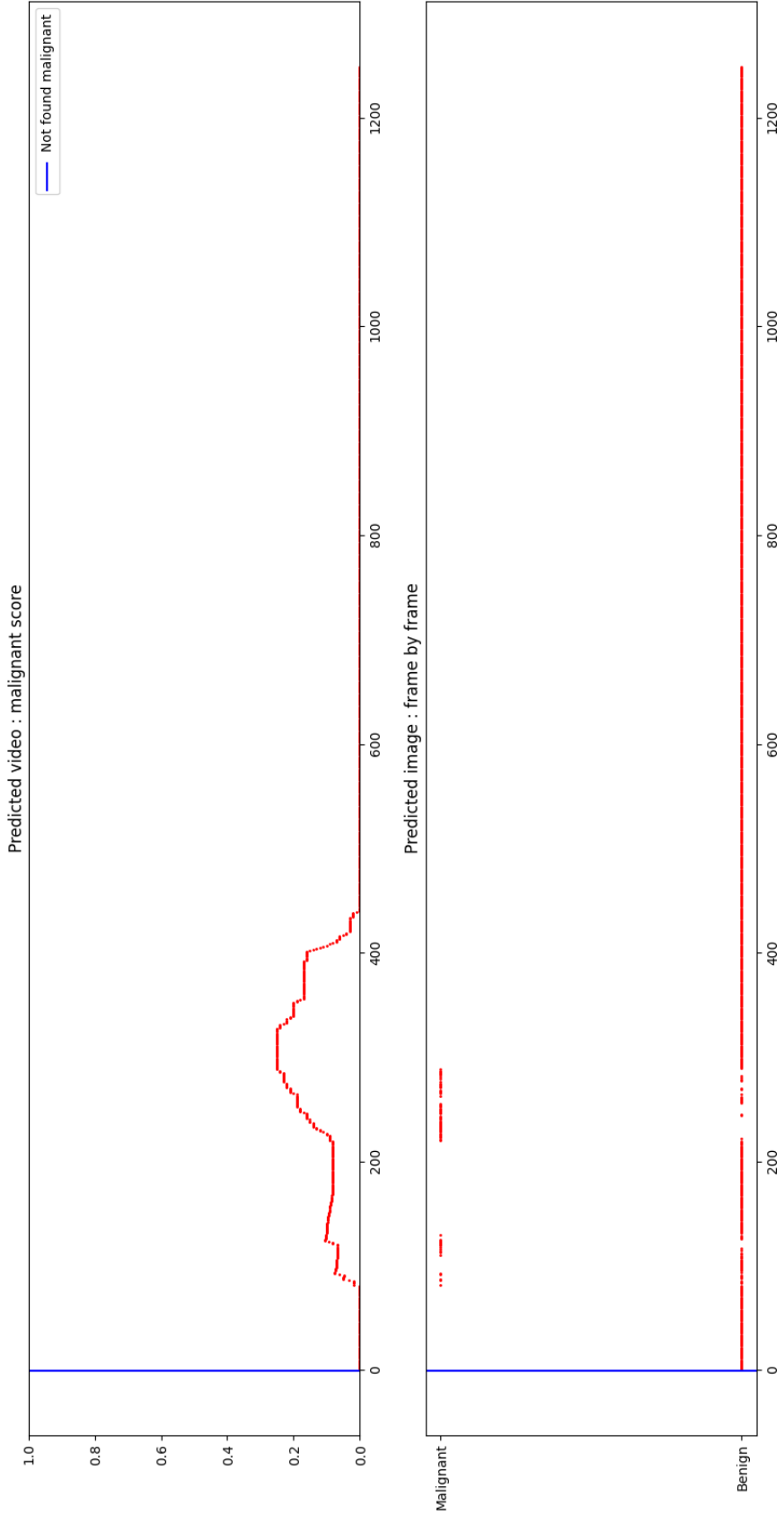


Figure 29. The true negative plotting between frame and malignant score.

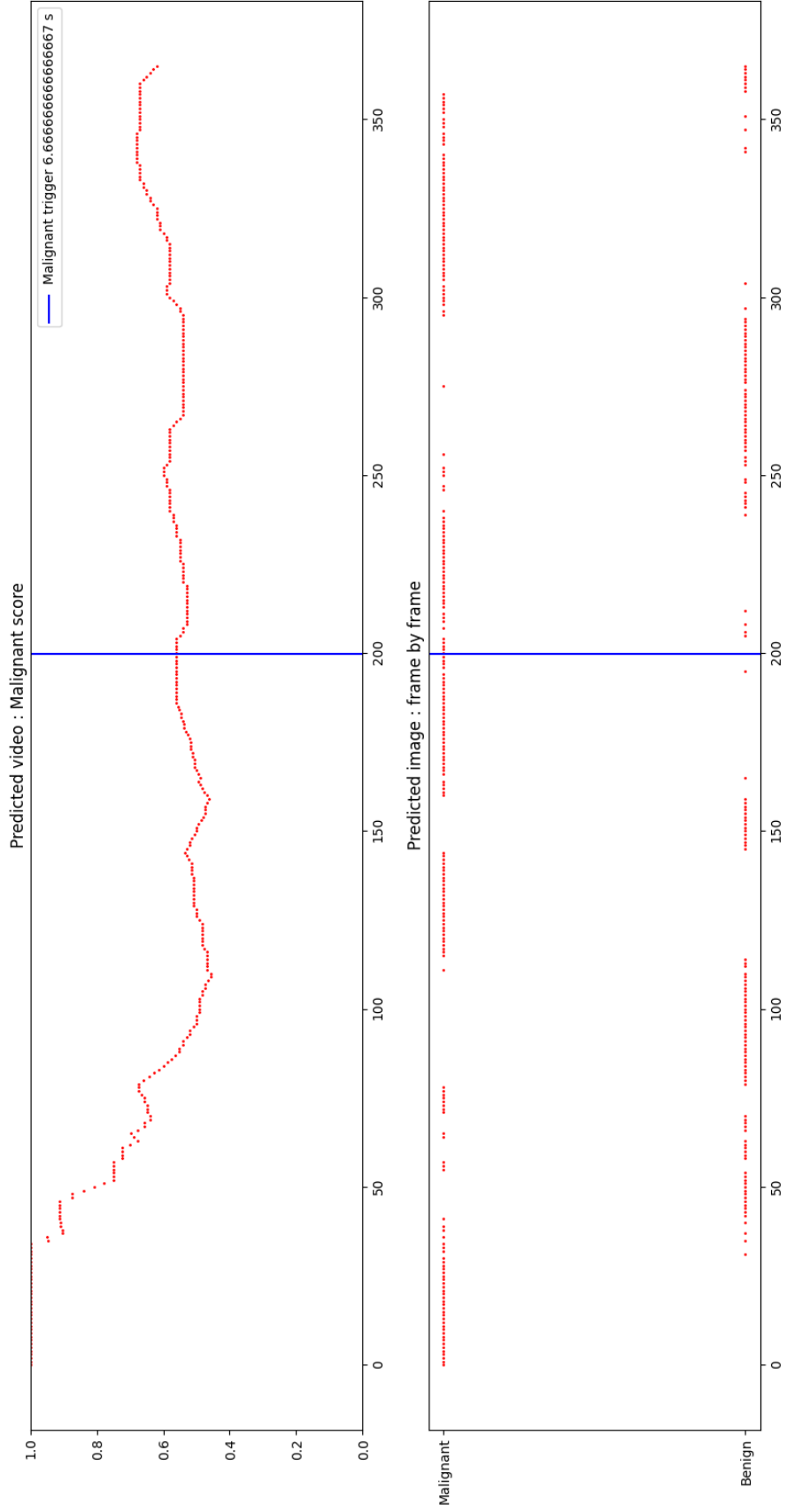


Figure 30. The false positive plotting between frame and malignant score.

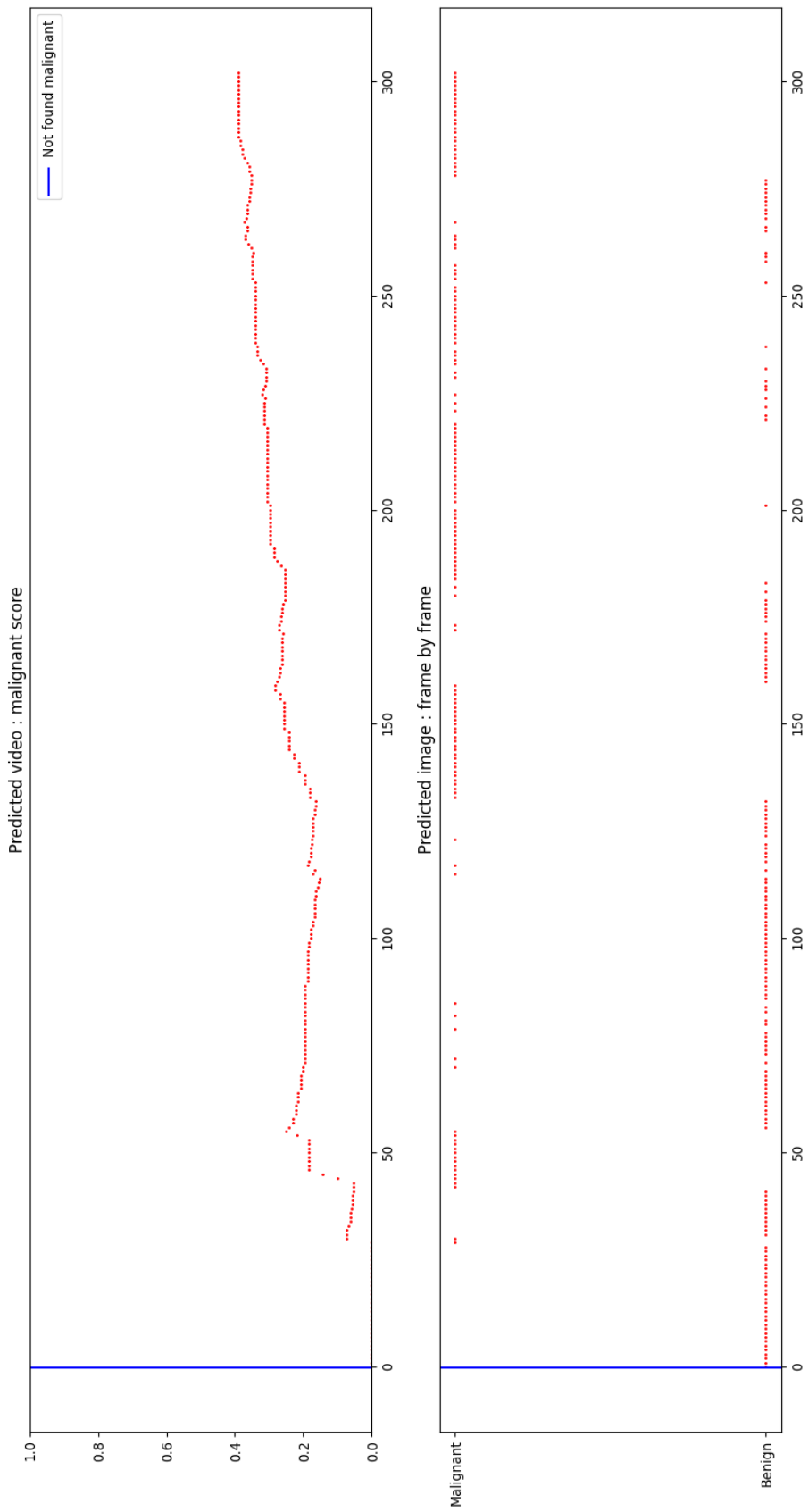


Figure 31. The false negative plotting between frame and malignant score.

5.8 Model deployments

In this section, we present the program that is used in real-world scenarios with our model deployments. The crucial features in this program consist of predictions, heatmaps, contours, and pictures in pictures.

We first present an overview of the program in figure 32. The program's UI contains several features. Firstly, malignant point thresholds can adjust the prediction threshold of moving averages. Secondly, heatmap areas on the main picture can be toggled on or OFF. Thirdly, draw style, which can toggle the draw output style from heatmap to contour, Fourthly, the malignant score is calculated from the moving average through 100 frames. Lastly, the prediction shown in the output of the model is that if malignant scores exceed the threshold, the prediction will be malignant. The calculation of the malignant score and prediction can be reset if you push the reset button. The one more important thing is the pictures that show on the left side of the screen, which always draw the heatmap to show the user's result without interrupting the user's experiment. In addition, All UI can be toggled to OFF, which means the experiment is not using AI assistance.

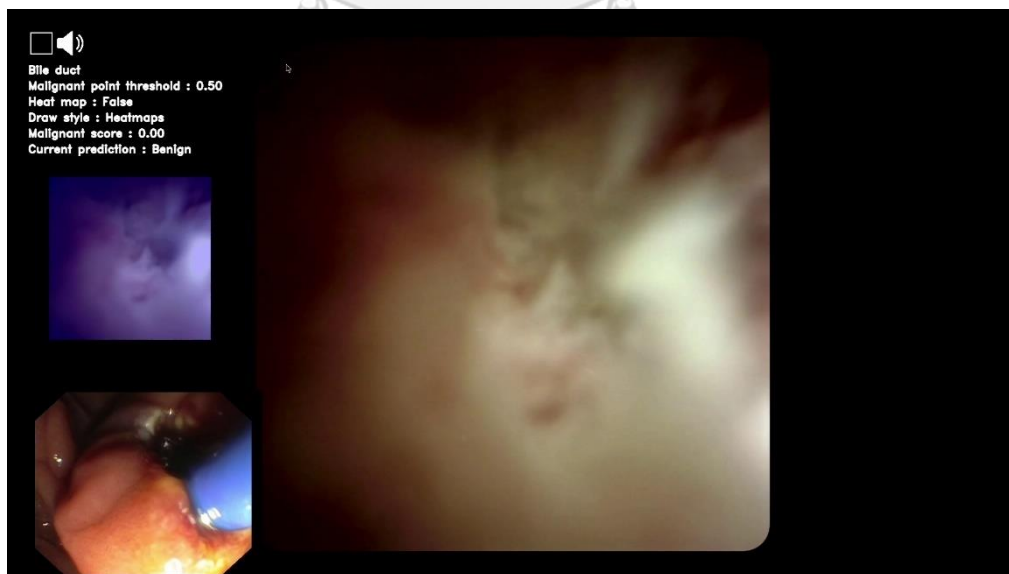


Figure 32. Overview of UI design for deployment.

5.8.1 Heatmap overlay

We present the heatmap overlay that will overlay on the suspicious area of the image. In figure 33, the heatmap is overlaid on the left side of the screen to show the user that an area is suspiciously malignant.

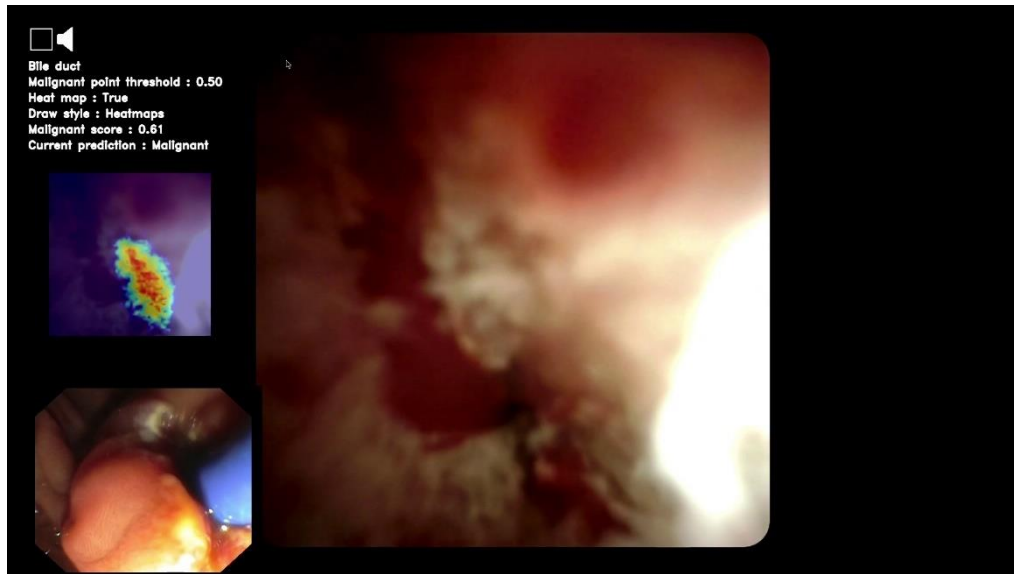


Figure 33. A heatmap overlay is shown on the left of the screen.

Additionally, the heatmap overlay can be shown on the main screen of the experiment, The result is shown in figure 34.

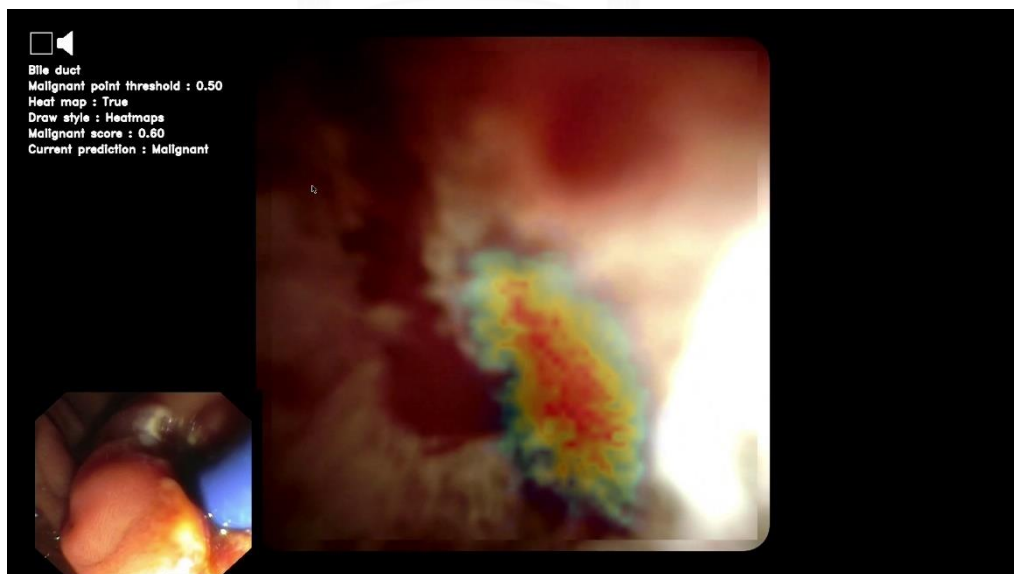


Figure 34. Heatmap overlay on main screen of the experiment.

5.8.2 Contour overlay

As the same as heatmap the contour are used as overlay to the image, we employ the heatmap which is the output of the model to contour by thresholding. Figure 35 shows the contour overlay on the left side of the screen same as heatmap.

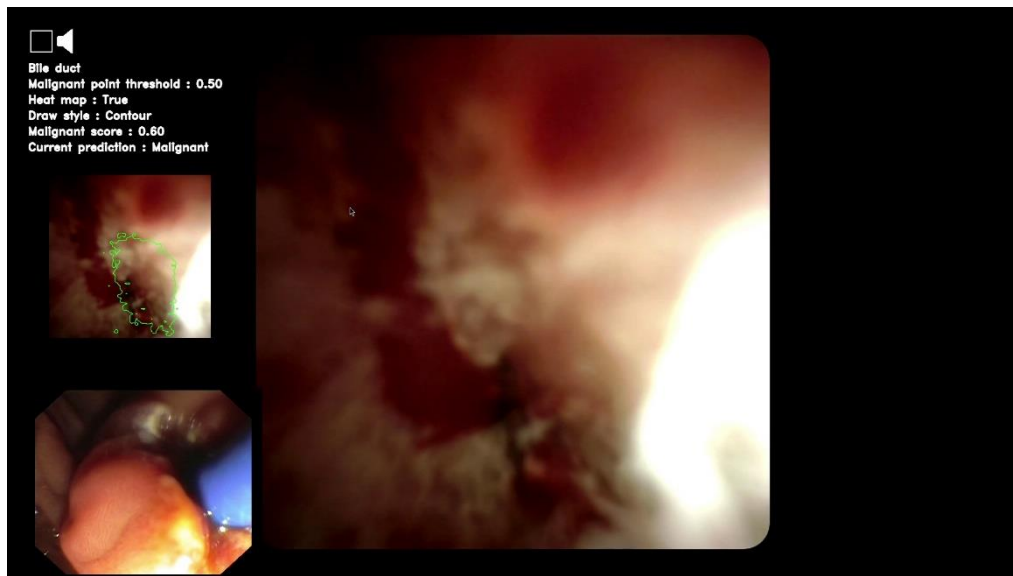


Figure 35. A contour overlay is shown on the left side of the screen.

A contour overlay can also shown on the main screen as same as heatmap overlay. The result of contour is demonstrated on figure 36.

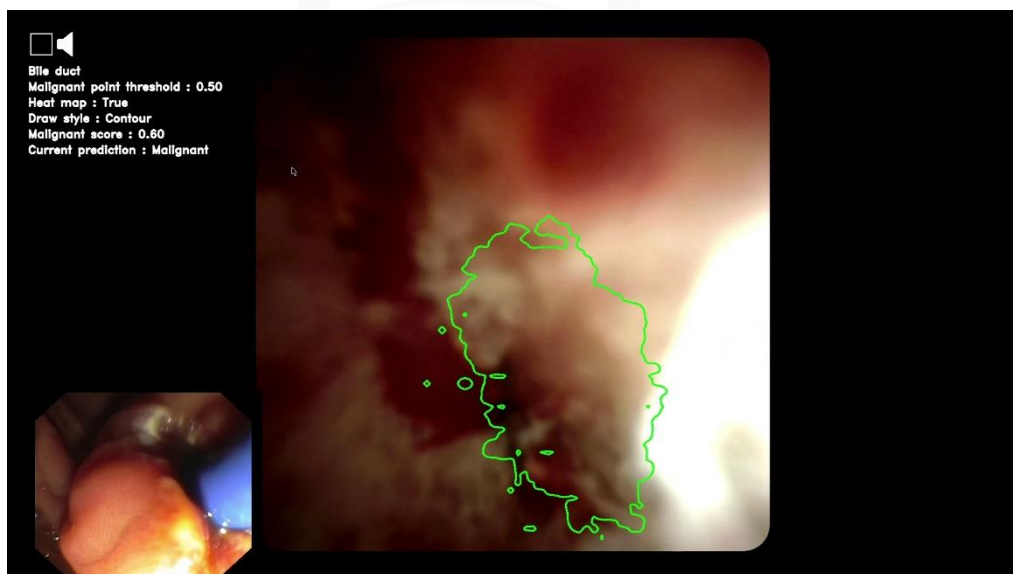


Figure 36. Contour overlay on main screen of the experiment.

CHAPTER VI

CONCLUSION

In the course of this thesis, a novel kind of deep learning model was developed. This model is able to categorize biliary strictures and give heatmaps to aid in the performance of biopsies in real time. The information for the dataset came from actual patients who were treated at King Chulalongkorn Memorial Hospital. Because of the outcomes of our experiment, we were able to conclude that our model improved its classification performance by adding a prediction head and an auxiliary loss to the original PYLON. Despite these additions, the model was still able to identify potentially malicious areas. The generalizability of the model was increased because of the addition of guide wires, which significantly enhanced classification performance. In addition, we provided an approach for applying the model in a real-world situation that included a particular job for cholangioscopy, a moving average of the expected frame, and a variance difference for eliminating needless frames.

Endoscopists who do biopsies will greatly benefit from this work. Two constraints must be discussed in this section. First, the model must be able to operate on the TensorRT engine and convert to the ONNX format. This method can greatly increase the speed of model inference, especially for convolutional networks like EfficientNet. Some trendy models, such as Transformer, cannot completely profit from this procedure since their ONNX-formatted model only sees a slight boost in inference speed. This makes several modern models inappropriate for use in our real-time inference situation. Second, there is no publicly accessible data set on the indeterminate biliary strictures task; this may lead to a result with less of another aspect.

It can be improved further in three areas in the future: First, we could apply a strategy to overcome the limited training data. The diffusion models can create synthesis data, which increases the quantity of training data in our private dataset, lowering the potential for overfitting and making the model more generic. Second, after the data is larger and more sufficient, the p-value from the statistical significance test may be correctly provided in the future. Third, we plan to place our model on medical-grade hardware to enable endoscopists to perform cholangioscopy and biopsies. In the clinical trial, we will confirm our model's performance in real-world scenarios.



REFERENCES

1. Sripa, B. and C. Pairojkul, *Cholangiocarcinoma: lessons from Thailand*. Current Opinion in Gastroenterology, 2008. **24**(3).
2. Mosconi, S., et al., *Cholangiocarcinoma*. Critical Reviews in Oncology/Hematology, 2009. **69**(3): p. 259-270.
3. Draganov, P., et al., *Diagnostic accuracy of conventional and cholangioscopy-guided sampling of indeterminate biliary lesions at the time of ERCP: A prospective, long-term follow-up study*. Gastrointestinal endoscopy, 2012. **75**: p. 347-53.
4. Clayton, R., et al., *Incidence of benign pathology in patients undergoing hepatic resection for suspected malignancy*. The surgeon : journal of the Royal Colleges of Surgeons of Edinburgh and Ireland, 2003. **1**: p. 32-8.
5. Victor, D., et al., *Current endoscopic approach to indeterminate biliary strictures*. World journal of gastroenterology : WJG, 2012. **18**: p. 6197-205.
6. Arvanitakis, M., *Digital single-operator cholangioscopy-guided biopsy for indeterminate biliary strictures: Seeing is believing?* Gastrointestinal Endoscopy, 2020. **91**: p. 1114-1116.
7. Bowlus, C.L., K.A. Olson, and M.E. Gershwin, *Erratum: Evaluation of indeterminate biliary strictures*. Nature Reviews Gastroenterology & Hepatology, 2017. **14**(12): p. 749-749.
8. Stassen, P., et al., *Diagnostic accuracy and interobserver agreement of digital single-operator cholangioscopy for indeterminate biliary strictures*. Gastrointestinal endoscopy, 2021. **94**.
9. Behary, J., M. Keegan, and P. Craig, *The inter-observer agreement of optical features used to differentiate benign from neoplastic biliary lesions assessed at balloon-assisted cholangioscopy*. Journal of Gastroenterology and Hepatology, 2018. **34**.
10. Egger, J., et al., *Medical deep learning—A systematic meta-review*. Computer Methods and Programs in Biomedicine, 2022. **221**: p. 106874.

11. Du, W., et al., *Review on the Applications of Deep Learning in the Analysis of Gastrointestinal Endoscopy Images*. IEEE Access, 2019. **7**: p. 142053-142069.
12. Mascarenhas, M., et al., *Artificial intelligence for automatic diagnosis of biliary stricture malignancy status in single-operator cholangioscopy: a pilot study*. Gastrointestinal Endoscopy, 2021. **95**.
13. Marya, N., et al., *Identification of patients with malignant biliary strictures using a cholangioscopy-based deep learning artificial intelligence (with video)*. Gastrointestinal Endoscopy, 2022.
14. *Understanding Deep Convolutional Neural Networks*.
15. Saha, S. *A Comprehensive Guide to Convolutional Neural Networks — the ELI5 Way*. 16 Nov. 2022; Available from: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>.
16. Alzubaidi, L., et al., *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*. Journal of Big Data, 2021. **8**.
17. Zhou, Y., et al., *Three modalities on common bile duct exploration*. Z Gastroenterol, 2017. **55**(9): p. 856-860.
18. *Digestive Diseases & Nutrition | USF Health*.
19. He, K., et al. *Deep Residual Learning for Image Recognition*. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.
20. Simonyan, K. and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2014.
21. Russakovsky, O., et al., *ImageNet Large Scale Visual Recognition Challenge*. International Journal of Computer Vision, 2014. **115**.
22. Kaiser, L., A. Gomez, and F. Chollet, *Depthwise Separable Convolutions for Neural Machine Translation*. 2017.
23. Szegedy, C., et al. *Going deeper with convolutions*. in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015.
24. Tsang, S.-H. *Review: Xception — With Depthwise Separable Convolution, Better Than Inception-v3 (Image Classification)*. Medium, . 20 Mar. 2019; Available from: <https://towardsdatascience.com/review-xception-with-depthwise->

[separable-convolution-better-than-inception-v3-image-dc967dd42568](#).

25. Tan, M. and Q. Le. *Efficientnet: Rethinking model scaling for convolutional neural networks*. in *International conference on machine learning*. 2019. PMLR.
26. Howard, A.G., et al., *Mobilenets: Efficient convolutional neural networks for mobile vision applications*. arXiv preprint arXiv:1704.04861, 2017.
27. Preechakul, K., et al., *Improved image classification explainability with high-accuracy heatmaps*. *iScience*, 2022. **25**(3): p. 103933.
28. Chen, P.-J., et al., *Accurate classification of diminutive colorectal polyps using computer-aided analysis*. *Gastroenterology*, 2018. **154**(3): p. 568-575.
29. He, J.Y., et al., *Hookworm Detection in Wireless Capsule Endoscopy Images With Deep Learning*. *IEEE Transactions on Image Processing*, 2018. **27**(5): p. 2379-2392.
30. Krizhevsky, A., I. Sutskever, and G.E. Hinton, *Imagenet classification with deep convolutional neural networks*. *Communications of the ACM*, 2017. **60**(6): p. 84-90.
31. Poudel, S., et al., *Colorectal Disease Classification Using Efficiently Scaled Dilation in Convolutional Neural Network*. *IEEE Access*, 2020. **8**: p. 99227-99238.
32. Zhou, B., et al. *Learning deep features for discriminative localization*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
33. Li, Z., et al. *Thoracic Disease Identification and Localization with Limited Supervision*. in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018.
34. Selvaraju, R.R., et al. *Grad-cam: Visual explanations from deep networks via gradient-based localization*. in *Proceedings of the IEEE international conference on computer vision*. 2017.
35. Chattopadhyay, A., et al. *Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks*. in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2018.
36. Fu, R., et al., *Axiom-based grad-cam: Towards accurate visualization and explanation of cnns*. arXiv preprint arXiv:2008.02312, 2020.
37. Kirillov, A., et al. *Panoptic Feature Pyramid Networks*. in *2019 IEEE/CVF*

- Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019.
38. Byrne, M., et al., *Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model*. *Gut*, 2017. **68**: p. gutjnl-2017.
39. Lu, Y., et al., *Real-Time Artificial Intelligence-Based Histologic Classifications of Colorectal Polyps Using Narrow-Band Imaging*. *Frontiers in Oncology*, 2022. **12**: p. 879239.
40. Zhang, H., et al. *ResNeSt: Split-Attention Networks*. in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2022.
41. He, K., et al., *Identity Mappings in Deep Residual Networks*. Vol. 9908. 2016. 630-645.
42. Hu, J., L. Shen, and G. Sun. *Squeeze-and-Excitation Networks*. in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018.





จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

VITA

NAME Passakron Phuangthongkham

DATE OF BIRTH 03 Sep 1998

PLACE OF BIRTH Rayong, Thailand

INSTITUTIONS ATTENDED Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University

HOME ADDRESS 79/56 Kheharomklao 64 Rd. Klongsongtonnun Sub-District Ladkrabang District Bangkok 10520

PUBLICATION P. Phuangthongkham, P. Angsuwatcharakon, S. Kulpatcharapong, P. Vateekul and R. Rerknimitr, "Real-Time Identification of Malignant Biliary Strictures on Cholangioscopy Images Using Explainable Convolutional Neural Networks With Heatmaps," in IEEE Access, vol. 11, pp. 49943-49956, 2023, doi: 10.1109/ACCESS.2023.3276642.