

DEEP LEARNING APPROACH  
ON SYMPTOM QUESTIONNAIRE AND ABDOMINAL RADIOGRAPHY  
FOR DIAGNOSIS OF DYSSYNERGIC DEFECATION



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 2021

Copyright of Chulalongkorn University

เทคนิคการเรียนรู้เชิงลึกบนข้อมูลแบบสอบถามอาการและภาพเอกซเรย์ช่องท้อง  
เพื่อวินิจฉัยภาวะกล้ามเนื้อควบคุมการถ่ายอุจจาระทำงานไม่ประสานกัน



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต  
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์  
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย  
ปีการศึกษา 2564  
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Thesis Title	DEEP LEARNING APPROACH ON SYMPTOM QUESTIONNAIRE AND ABDOMINAL RADIOGRAPHY FOR DIAGNOSIS OF DYSSYNERGIC DEFECATION
By	Miss Sornsiri Poovongsaroj
Field of Study	Computer Engineering
Thesis Advisor	Assistant Professor PEERAPON VATEEKUL, Ph.D.
Thesis Co Advisor	Assistant Professor Tanisa Patcharatrakul, 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  
(Associate Professor SUPOT TEACHAVORASINSKUN, D.Eng.)

THESIS COMMITTEE

----- Chairman  
(Professor BOONSERM KIJSIRIKUL, Ph.D.)

----- Thesis Advisor  
(Assistant Professor PEERAPON VATEEKUL, Ph.D.)

----- Thesis Co-Advisor  
(Assistant Professor Tanisa Patcharatrakul, M.D.)

----- Examiner  
(PAKKAPON RATTANACHAISIT, M.D.)

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

สรณสิริ พวงศาโรจน์ : เทคนิคการเรียนรู้เชิงลึกบนข้อมูลแบบสอบถามอาการและภาพเอกซเรย์ช่องท้องเพื่อวินิจฉัยภาวะกล้ามเนื้อควบคุมการถ่ายอุจจาระทำงานไม่ประสานกัน. ( DEEP LEARNING APPROACH ON SYMPTOM QUESTIONNAIRE AND ABDOMINAL RADIOGRAPHY FOR DIAGNOSIS OF DYSSYNERGIC DEFECATION) อ.ที่ปรึกษาหลัก : ผศ. ดร.พีรพล เวทีกุล, อ.ที่ปรึกษาร่วม : ผศ.(พิเศษ) พญ.ฐนิสา พัชรตระกูล

ภาวะกล้ามเนื้อควบคุมการถ่ายอุจจาระทำงานไม่ประสานกันเป็นหนึ่งในสาเหตุที่มักพบบ่อยของภาวะท้องผูกเรื้อรัง ผู้ป่วยที่มีภาวะนี้ไม่สามารถขับถ่ายได้เนื่องจากมีปัญหาในการควบคุมกล้ามเนื้อบริเวณอุ้งเชิงกราน ซึ่งทำงานประสานกันเพื่อทำหน้าที่ในการขับถ่ายอุจจาระ การวินิจฉัยภาวะนี้ผู้ป่วยจำเป็นต้องเข้ารับการตรวจในโรงพยาบาลตติยภูมิ ทางผู้จัดทำจึงต้องการพัฒนาโมเดลการเรียนรู้เชิงลึกสำหรับคัดกรองผู้ป่วยเบื้องต้นจากสถานพยาบาลระดับปฐมภูมิและทุติยภูมิเพื่อส่งต่อมารับการวินิจฉัยในโรงพยาบาลที่มีความพร้อมต่อไป โดยใช้ข้อมูลทางการแพทย์ที่สามารถจัดทำได้ง่ายๆ เช่น แบบสอบถามอาการ และภาพเอกซเรย์ช่องท้อง เป็นต้น การพัฒนาโมเดลการเรียนรู้เชิงลึกถูกแบ่งออกเป็น 3 ส่วนด้วยกัน ส่วนแรก การพัฒนาโมเดลจากอัลกอริทึมการเรียนรู้ต้นไม้และโมเดลการเรียนรู้เชิงลึก โดยใช้ข้อมูลจากแบบสอบถามอาการเพียงอย่างเดียว นอกจากนี้ทางผู้จัดทำยังทำการเลือกคุณลักษณะ โดยความรู้ที่ได้มาจากผู้เชี่ยวชาญ และโดยวิธีการดั้งเดิม เพื่อค้นหาชุดข้อมูลที่เหมาะสมที่สุด ส่วนที่สอง การพัฒนาโมเดลจากโมเดลจำแนกรูปภาพที่เป็น state-of-the-art โดยใช้ข้อมูลจากภาพเอกซเรย์ช่องท้องอย่างเดียว และมีการใช้เทคนิคในการเพิ่มข้อมูลรูปภาพในการประมวลผลก่อนด้วย ส่วนที่สาม ทางผู้จัดทำได้นำเสนอโมเดลแบบบูรณาการ ซึ่งใช้ทั้งข้อมูลจากแบบสอบถามอาการและภาพเอกซเรย์ช่องท้อง คุณลักษณะที่ได้รับคัดเลือกจากแบบสอบถามอาการจะถูกนำมาต่อกันกับคุณลักษณะภาพที่สกัดออกมาจากภาพเอกซเรย์ช่องท้องที่ขึ้นต่อกัน (concatenate layer) โดยแนวทางดังกล่าวมีต้นแบบมาจากการวินิจฉัยของแพทย์ผู้เชี่ยวชาญในการปฏิบัติงานจริง นอกจากนี้ทางผู้จัดทำยังนำเสนอวิธีการประมวลผลก่อนและประมวลผลหลังที่เหมาะสมกับชุดข้อมูลขนาดเล็กเพื่อเพิ่มประสิทธิภาพและความแม่นยำของตัวโมเดล จากผลการทดลองแสดงให้เห็นว่าโมเดลแบบบูรณาการที่ทางผู้จัดทำเสนอเอาชนะโมเดลฐานด้วยค่าความแม่นยำ 66.01%

สาขาวิชา วิศวกรรมคอมพิวเตอร์

ปีการศึกษา 2564

ลายมือชื่อนิสิต .....

ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

# # 6370290621 : MAJOR COMPUTER ENGINEERING

KEYWORD: medical diagnosis, artificial intelligence, deep learning, dyssynergic defecation, symptom questionnaire, abdominal radiography

Sornsiri Poovongsaraj : DEEP LEARNING APPROACH ON SYMPTOM QUESTIONNAIRE AND ABDOMINAL RADIOGRAPHY FOR DIAGNOSIS OF DYSSYNERGIC DEFECATION.

Advisor: Asst. Prof. PEERAPON VATEEKUL, Ph.D. Co-advisor: Asst. Prof. Tanisa Patcharatrakul, M.D.

Dyssynergic defecation is one of the most common causes of chronic constipation. It is a behavioral problem in which the pelvic floor muscles are unable to coordinate with the surrounding muscles and nerves to evacuate stool. Patients are required to undergo specialized tests only available at tertiary healthcare centers for diagnosis. The aim of this thesis is to develop deep learning-based models to prescreen potential patients from primary and secondary healthcare centers for further diagnostic tests by using easily obtainable data such as symptom questionnaire and abdominal radiography. First, we developed a model which uses symptom questionnaire as an input from tree-based machine learning algorithms and deep learning model. Feature selection based on expert knowledge and based on traditional method were performed to find the best set of input features. Second, we developed a model which uses abdominal radiography as an input from the state-of-the-art image classification models. Several image augmentation techniques were applied as data preprocessing. Third, we proposed an integrated model which uses both symptom questionnaire and abdominal radiography as inputs. The selected input features from symptom questionnaire were combined with image features extracted from the abdominal radiography using a concatenate layer. This approach was meant to imitate how human experts diagnose in real life. We also proposed data preprocessing and postprocessing suitable for small dataset to improve the model accuracy and efficiency. The results show that our proposed integrated model outperforms the baseline models with an accuracy of 66.01%.

Field of Study: Computer Engineering

Student's Signature .....

Academic Year: 2021

Advisor's Signature .....

Co-advisor's Signature .....

## ACKNOWLEDGEMENTS

The completion of this thesis could not have been possible without the help, support, and guidance of many people.

I would like to thank my advisor, Assoc. Prof. Dr. Peerapon Vateekul, for all his help and guidance that he has given me over the past two years.

I would also like to thank Asst. Prof. Dr. Tanisa Patcharatrakul and Dr. Pakkapon Rattanachaisit for providing experimental data, as well as guidance and feedback throughout this research.

I would like to express my gratitude to Prof. Dr. Boonserm Kijirikul and Dr. Thanapat Kangkachit for their time and valuable insights.

I would like to extend my sincere thanks to my lab seniors from DataMind for their support and cooperation.

Finally, I would like to thank my parents for their continuous support and understanding when undertaking my research and writing this thesis. Without their support, I do not think that I could overcome the difficulties during these years.

Sornsiri Poovongsaroj

## TABLE OF CONTENTS

	Page
ABSTRACT (THAI) .....	iii
ABSTRACT (ENGLISH) .....	iv
ACKNOWLEDGEMENTS .....	v
TABLE OF CONTENTS .....	vi
LIST OF TABLES .....	ix
LIST OF FIGURES.....	x
CHAPTER I INTRODUCTION.....	1
1.1 Background and Rationale.....	1
1.2 Aims and Objectives.....	2
1.3 Scope of Work .....	3
1.4 Contribution .....	3
1.5 Research Method.....	3
1.6 Research Plan .....	4
1.7 Research Ethics.....	5
1.8 Publication.....	5
CHAPTER II LITERATURE REVIEW .....	6
2.1 Tabular Data.....	6
2.2 X-Ray Images.....	7
2.3 Related Work .....	9
CHAPTER III RESEARCH DESIGN AND METHODOLOGY.....	10
3.1 Data Preprocessing .....	10

3.1.1 Symptom Questionnaire.....	10
3.1.2 Abdominal Radiography.....	11
3.2 Model Development.....	13
3.2.1 Symptom Model.....	13
3.2.2 Image Model.....	13
3.2.3 Integrated Model.....	14
3.3 Data Postprocessing .....	15
CHAPTER IV EXPERIMENTAL SETUP.....	17
4.1 Dataset.....	17
4.1.1 Symptom Questionnaire.....	17
4.1.2 Abdominal Radiography.....	19
4.1.3 Combined Dataset .....	19
4.2 Hyperparameters Tuning.....	20
4.2.1 Symptom Model.....	20
4.2.2 Image Model.....	21
4.2.3 Integrated Model.....	21
4.3 Evaluation Metrics.....	22
CHAPTER V RESULTS .....	23
5.1 Symptom Model .....	23
5.2 Image Model .....	25
5.3 Integrated Model .....	26
CHAPTER VI CONCLUSION.....	30
6.1 Conclusion .....	30
6.2 Future Work.....	30



REFERENCES.....	31
APPENDICES .....	34
APPENDIX A.....	35
APPENDIX B.....	36
VITA .....	39



## LIST OF TABLES

	Page
Table 1 Gantt chart for research timeline .....	4
Table 2 Diagnostic criteria for DD.....	17
Table 3 Symptom questionnaire .....	18
Table 4 Optimal parameters for random forest during second-round performance (feature selection) from GridSearchCV .....	21
Table 5 First-round performance (129 features).....	24
Table 6 Second-round performance (feature selection) .....	24
Table 7 Best symptom model with decision threshold adjustment on train set .....	24
Table 8 Best symptom model performance in each iteration of 5-fold cross-validation .....	25
Table 9 SOTA image classification model performance .....	25
Table 10 EfficientNet model performance .....	26
Table 11 Datasets with different setups using EfficientNet-B3.....	26
Table 12 Best image model with decision threshold adjustment on validate set .....	26
Table 13 Integrated model performance (from Figure 10).....	27
Table 14 Best integrated model with decision threshold adjustment on validate set	28
Table 15 Model performance comparison on each dataset .....	28

## LIST OF FIGURES

	Page
Figure 1 TabNet's sparse feature selection (from Fig. 1 [16]).....	7
Figure 2 EfficientNet Compound Scaling.....	8
Figure 3 Structure of MBConv and Fused-MBConv (from Fig. 2 [23]).....	8
Figure 4 Grad-CAM on KUB radiograph: (a) original image and (b) Grad-CAM.....	9
Figure 5 Research design and methodology workflow.....	10
Figure 6 Illustration of k-fold cross-validation.....	11
Figure 7 Data preprocessing for abdominal radiograph: (a) original image, (b) original image with CLAHE, (c) auto crop, (d) manual crop, (e) manual crop with manual contrast and (f) auto crop with auto contrast.....	12
Figure 8 Examples of augmented abdominal radiographs.....	13
Figure 9 Image model architecture.....	14
Figure 10 Multiple-input single-output architecture: (a) concatenating input features from the symptom questionnaire directly with image features and (b) adding some dense layer to the input features from the symptom questionnaire before concatenating with the image features.....	15
Figure 11 Grad-CAM: (a) patients with NTC (b) patients with DD.....	16
Figure 12 Abdominal radiographs.....	19
Figure 13 Class distribution in each dataset.....	20
Figure 14 Structure of the confusion matrix.....	22
Figure 15 Feature importance bar chart: (a) input features based on expert knowledge and (b) input features based on traditional method.....	29

# CHAPTER I

## INTRODUCTION

### 1.1 Background and Rationale

People usually experience constipation from time to time. It is one of the most common gastrointestinal disorders which affects people of all ages. Most cases of constipation are mild and can be treated with simple changes in diet and lifestyle. Constipation is considered to be chronic if the condition persists for several weeks or longer. Although chronic constipation rarely has any life-threatening effects, it largely interferes ability to function and patient's daily life [1]. More than one half of chronic constipation is caused by dyssynergic defecation (DD). It is a health condition with difficulty passing stool due to an inability to coordinate abdominal, rectoanal, and pelvic floor muscles which support the bowel movement [2]. Patients with DD often reported feelings such as excessive straining and incomplete evacuation which has negative impact on their quality of life [3, 4].

Diagnostic criteria for DD require specialized testing, namely anorectal manometry (ARM), and balloon expulsion test (BET), and defecography. Chronic constipation symptoms are considered along with 2 out of these 3 tests [5]. However, these tests are not widely available and clinicians with a high level of expertise are required. A thorough history or stool diary are also used to predict the presence of DD in patients. Therefore, we would like to propose a more practical approach to diagnose DD accurately in an accessible manner.

Previous studies have emphasized the possibility of using tabular data such as a detailed history or stool diary to diagnose DD [2-6]. One of the studies suggested an approach to determine the predictability of DD in patients with chronic constipation using a standardized self-reported symptom questionnaire. They claimed that questions regarding need to strain, straining duration, urge to defecate, and incomplete evacuation are useful for predicting the presence of DD [6]. Medical images can also be used in the diagnosis as shown in prior study which adopted the use of computerized tomography (CT) images to identify rectal evacuation disorders

in constipated patients by measuring the rectal gas volume. The results show that the rectal gas volume in patients with constipation are greater than those with normal transit constipation (NTC) [7]. Although CT scan is very effective for diagnosis, it is relatively expensive and may not be available in rural areas. Instead of CT image, we decided to use plain radiographs which are more affordable in our experiment.

In the past few years, deep learning has exhibited impressive performances in medicine and healthcare with its advanced computing power and the increasing availability of big data [8]. There have been numerous studies to improve the efficiency and accuracy of medical diagnosis using machine learning and deep learning [9, 10]. The ability to analyze clinical data at an exceptional speed and provide diagnostic suggestions comparable with medical experts greatly help clinicians to narrow down the diagnostic range and develop a more precise and complete diagnosis.

This thesis proposes deep learning approaches to diagnose DD by using symptom questionnaire and abdominal radiography of the kidney, ureter, and bladder (KUB) view. We attempt to develop deep learning-based models to help in pre-screening eligible patients from primary and secondary healthcare centers for further diagnostic tests. Several preprocessing and postprocessing techniques are applied to overcome small data limitations and maximize the model's performance.

## 1.2 Aims and Objectives

We aimed to propose deep learning-based models for diagnosis by using symptom questionnaire and abdominal radiography (KUB view).

The objectives of this research are as follows:

- To propose deep learning algorithms for data and image analysis.
- To overcome limited size of healthcare datasets by applying suitable preprocessing and postprocessing techniques.
- To examine the possibility of pre-screening patients with DD by using symptom questionnaire and abdominal radiography (KUB view).

### 1.3 Scope of Work

- Develop deep learning-based models using symptom questionnaire and abdominal radiography (KUB view).
- Establish appropriate data preprocessing and postprocessing methods for limited data problems.
- Compare 3 different datasets: (1) symptom questionnaire, (2) abdominal radiography, and (3) symptom questionnaire and abdominal radiography.

### 1.4 Contribution

- Deep learning-based models for DD diagnosis will be proposed.
- Suitable techniques for handling with small datasets will be proposed.
- Allow primary and secondary healthcare centers to pre-screen potential patients to be referred to tertiary healthcare centers.

### 1.5 Research Method

- 1) Studying background and related work of the selected research topic.
- 2) Analyzing and understanding the data received from the Center of Excellence in Neurogastroenterology and Motility.
- 3) Learning essential deep learning tools, TensorFlow and Keras.
- 4) Designing and implementing deep learning-based models on symptom questionnaire and abdominal radiography.
- 5) Evaluating and improving the model performance using feedbacks from medical experts.
- 6) Comparing the results of each dataset.
- 7) Presenting and publishing at conference proceedings.
- 8) Writing thesis.

1.6 Research Plan

Task	2563		2564				2565	
	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2
Choose a research topic								
Study background and related work								
Data preparation and analysis								
Model design and implementation								
Phase 1: apply model on symptom questionnaire								
Phase 2: apply model on abdominal radiography								
Phase 3: apply model on symptom questionnaire and abdominal radiography								
Model evaluation and refinement								
Thesis proposal writing								
Thesis proposal submission								
Thesis writing								
Thesis submission								

Table 1 Gantt chart for research timeline

## 1.7 Research Ethics

This research was approved by the Institutional Review Board of Faculty of Medicine, Chulalongkorn University (IRB number 016/64).

## 1.8 Publication

- Poovongsaroj S., Rattanachaisit P., Patcharatrakul T., Gonlachanvit S., and Vateekul P., AI-ASSISTED DIAGNOSIS OF DYSSYNERGIC DEFECATION USING DEEP LEARNING APPROACH ON ABDOMINAL RADIOGRAPHY AND SYMPTOM QUESTIONNAIRE. Paper presented at International Joint Conference on Computer Science and Software Engineering 2022, Bangkok, Thailand, 22–25 June 2022.
- Rattanachaisit P., Poovongsaroj S., Patcharatrakul T., Gonlachanvit S., and Vateekul P., DIAGNOSIS OF DYSSYNERGIC DEFECATION (DD) USING MACHINE LEARNING APPROACH ON SYMPTOMS QUESTIONNAIRE. Paper presented at Digestive Disease Week 2022, San Diego, California, United States, 21–24 May 2022.
- Rattanachaisit P., Poovongsaroj S., Patcharatrakul T., Gonlachanvit S., and Vateekul P., MEDICAL IMAGING WITH DEEP LEARNING APPROACH FOR DIAGNOSIS OF DYSSYNERGIC DEFECATION (DD). Paper presented at Digestive Disease Week 2022, San Diego, California, United States, 21–24 May 2022.
- Rattanachaisit P., Poovongsaroj S., Patcharatrakul T., Gonlachanvit S., and Vateekul P., Sa412 ABDOMINAL RADIOGRAPHY WITH ARTIFICIAL INTELLIGENCE FOR DIAGNOSIS OF DYSSYNERGIC DEFECATION (DD). *Gastroenterology*, 2021. 160: p. S-498.



## CHAPTER II

### LITERATURE REVIEW

#### 2.1 Tabular Data

Tabular data refers to data displayed in rows and columns. Symptom questionnaire is an example of tabular data which contains a set of questions useful for collecting information on relevant symptoms. Previous study has applied the usage of self-reported pain severity and frequency validate questionnaire to classify constipated patients with a high or low symptom burden [11].

Despite the fact that deep learning has been growing rapidly during the past few years, traditional machine learning is still a good option when dealing with tabular data [12]. One of the most popular and successful machine learning algorithms are the tree-based models. They are famous for their simplicity with high accuracy, stability, and ease of interpretation. Another study has recognized tree-based methods as powerful data analytic tools for clinical decision-making using complex data structures [13]. The standard tree-based model performance can also be improved by combining several decision trees to reduce variance. Among ensembling methods, random forest is constructed in randomly chosen subspaces using random subsets of data [14]. XGBoost is one of the most recent ensemble tree-based models which utilizes a gradient descent algorithm. It applies the principle of boosting weak learner to minimize the loss function [15].

Although interpretable is one of the key strengths of tree-based models, a recent study proposed a high-performance and interpretable deep tabular data learning model. The authors claimed that their model, TabNet, outperforms other traditional machine learning algorithms while providing interpretable feature attributions and insights into its global behavior. It employs a single deep learning architecture which uses an instance-wise feature selection to focus on features that are important for the example at hand (Figure 1). This is accomplished through the use of sequential attention performed at each decision step [16].

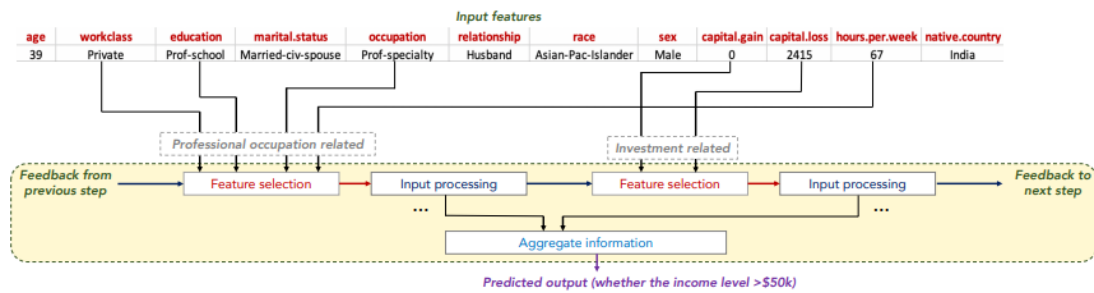


Figure 1 TabNet's sparse feature selection (from Fig. 1 [16])

## 2.2 X-Ray Images

Unlike tabular data, deep learning has been playing a prominent role in the field of image recognition. Regarding the coronavirus disease (COVID-19) pandemic, there have been numerous studies in deep learning on chest X-ray images [17]. Many studies attempted to develop a COVID-19 automated detection by applying the state-of-the-art (SOTA) deep learning models on chest X-ray images. [18-20]. A recent study suggests a fusion technique of combining clinical data and chest X-ray images to improve diagnostic accuracy [21]. Several SOTA image classification models were applied in this thesis and their results were compared.

In this thesis, we implemented several SOTA image classification models to find the most suitable model for the task. The first model is EfficientNet, a family of deep artificial neural networks, which achieves the competitive performance with the use of compound scaling at all dimensions of depth, width, and resolution (Figure 2). The EfficientNet model comprises of 8 models from B0 to B7 with each subsequent model number referring to variants with more parameters and higher accuracy [22]. After a few years, the authors released a new family of convolutional networks which was developed from EfficientNet known as EfficientNetV2. They mentioned using a combination of training-aware neural architecture search (NAS) and scaling to speed up the training time and reduce the number of parameters. They also replaced MBConv in early stages with Fused-MBConv as shown in Figure 3. An improved method of progressive learning by adding stronger regularization as the image size increases was also proposed in the research [23]. Visual Geometry Group (VGG), Residual networks (ResNet), and Dense Convolutional Network (DenseNet) achieve SOTA performance in image recognition and were used in numerous studies [24-26].

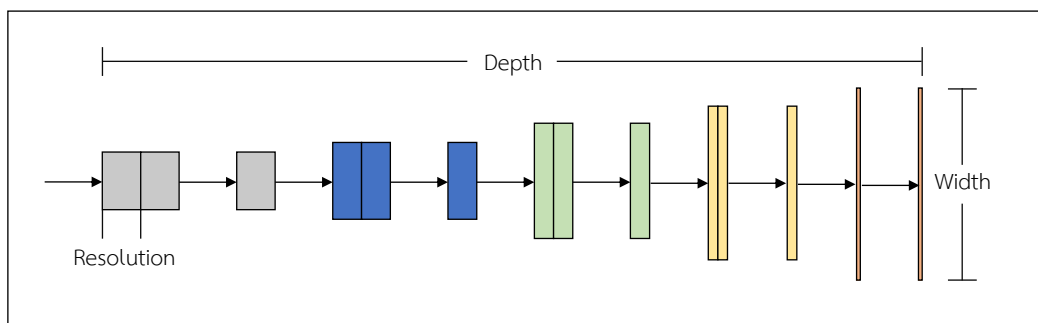


Figure 2 EfficientNet Compound Scaling

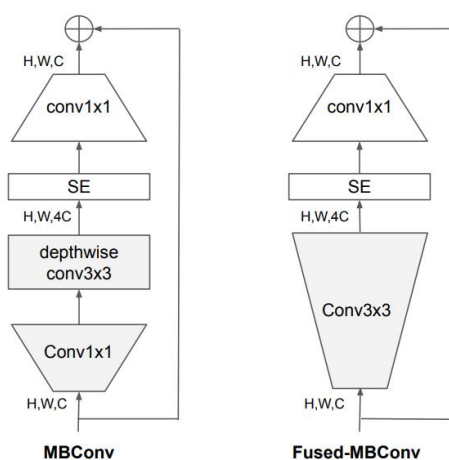


Figure 3 Structure of MBConv and Fused-MBConv (from Fig. 2 [23])

One of the key challenges for deep learning in healthcare is the lack of data due to patient privacy and data protection laws. Some authors have driven the further development of meta-transfer learning approach to transfer knowledge from big data and reduce the search space in data with small sample sizes. The research successfully overcome data scarcity and data heterogeneity using few-shot learning algorithms integrated with meta-learning [27]. Prior research suggested a new method to enhance the microstructure images by using Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm. It is a variant of adaptive histogram equalization [28] in which the contrast amplification is limited to reduce the noise amplification. Instead of the entire image, CLAHE operates on small regions in the image called tiles. The neighboring tiles are combined using bilinear interpolation to remove the artificial boundaries which results in improving the contrast of the image [29]. In

addition, one of the studies proposed a technique called Gradient-weighted Class Activation Mapping (Grad-CAM) which has been widely used in recent deep learning-based models to produce visual explanations for their decisions as shown in Figure 4. Hence, this technique allows humans to understand how these models make classification predictions [30].

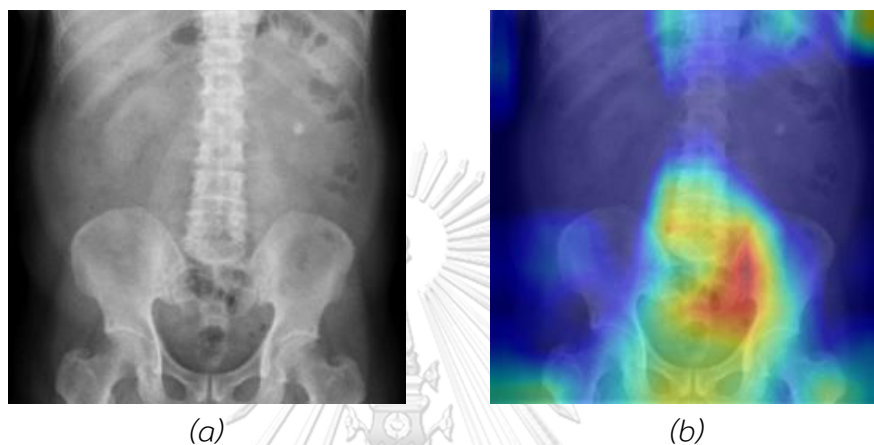


Figure 4 Grad-CAM on KUB radiograph: (a) original image and (b) Grad-CAM

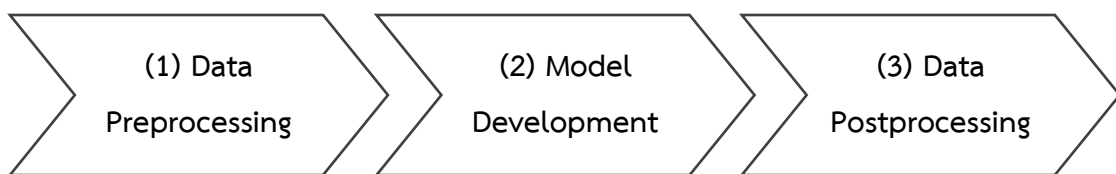
### 2.3 Related Work

Previous research can only be considered a first step towards a more profound understanding of AI applications in DD diagnosis. The author proposed a deep learning algorithm to evaluate DD using a newly developed technology called high-definition anorectal manometry (3D-HDAM). Spatial-temporal data extracted from the 3D-HDAM studies at a tertiary healthcare center was used as input data [31]. Although their outcomes were outstanding with comparable diagnostic accuracy, there is a limited access for the 3D-HDAM technology which makes it difficult to obtain the data in small hospitals.

## CHAPTER III

### RESEARCH DESIGN AND METHODOLOGY

This chapter discusses in detail the methodological procedures used to conduct the research including data preprocessing, model development, and data postprocessing (Figure 5).



*Figure 5 Research design and methodology workflow*

### 3.1 Data Preprocessing

#### 3.1.1 Symptom Questionnaire

For tabular data, data preparation and cleaning are very important. Data cleaning is performed to ensure that the data is correct, usable, and consistent by removing noise and filling missing values. Bar charts and histograms are used to visualize the distribution of the answers for each question in the symptom questionnaire. Before we start cleaning the data, questions related to ground truth or the specialized tests are excluded. The missing values are replaced by mode value for categorical data (ex. gender, sense of incomplete evacuation, presence of congenital disease) and mean value for numerical data (ex. age, weight, height). According to medical experts, symptom frequency and severity are closely related to each other. By multiplying the two features, we can create a more appropriate feature for training the model (ex. a mild symptom that frequently occurs can be considered more serious than a severe symptom that rarely occurs).

We also applied feature selection to find the best possible subset of input features for the symptom model and the integrated model. Feature selection based on expert knowledge and traditional method were compared in the experiment. In addition, k-fold cross-validation is also applied to handle the small dataset problem.

The dataset is divided into  $k$  non-overlapping folds. Therefore, each fold is used as a test set once. In this research, 5-fold cross-validation is applied to the dataset ( $k = 5$ ). One of the folds will be used as a test set while the rest are used to train the model. The process is repeated  $k$  times until every fold is used as a test set and the mean performance of the model is reported (Figure 6). K-fold cross-validation significantly reduces bias and allows the model to become more generalized.

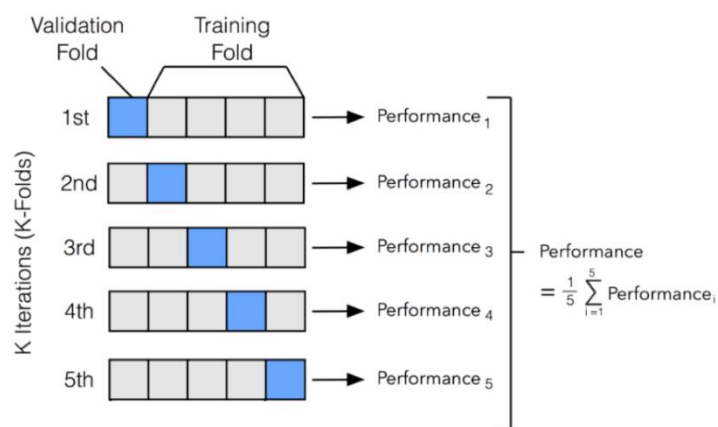


Figure 6 Illustration of  $k$ -fold cross-validation

### 3.1.2 Abdominal Radiography

The received Digital Imaging and Communications in Medicine (DICOM) files are extracted to Portable Network Graphics (PNG) images. The images are modified to reproduce exactly how they were viewed during the examination. As a result, the images with a high-resolution of 2300 x 2900 pixels.

We have tried several approaches of cropping and contrasting the images both manually and by using python code (auto) to compare the results (Figure 7). The images are cropped to focus on the pelvic cavity and contrasted to be similar to when medical experts interpret them. Due to the data limitations, a lot of image augmentation techniques such as rotation, shear, zoom, horizontal flip, and vertical flip are also used to expand the amount of training data (Figure 8).

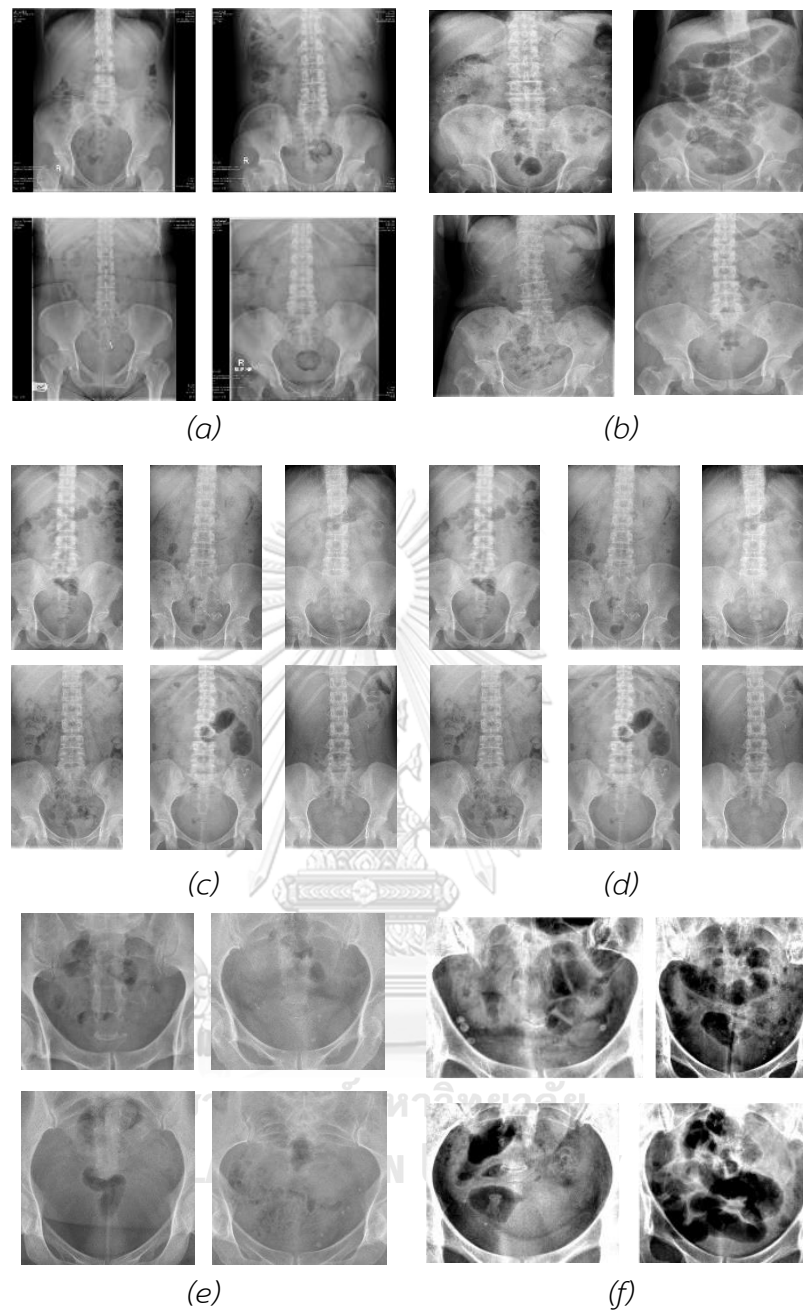


Figure 7 Data preprocessing for abdominal radiograph: (a) original image, (b) original image with CLAHE, (c) auto crop, (d) manual crop, (e) manual crop with manual contrast and (f) auto crop with auto contrast

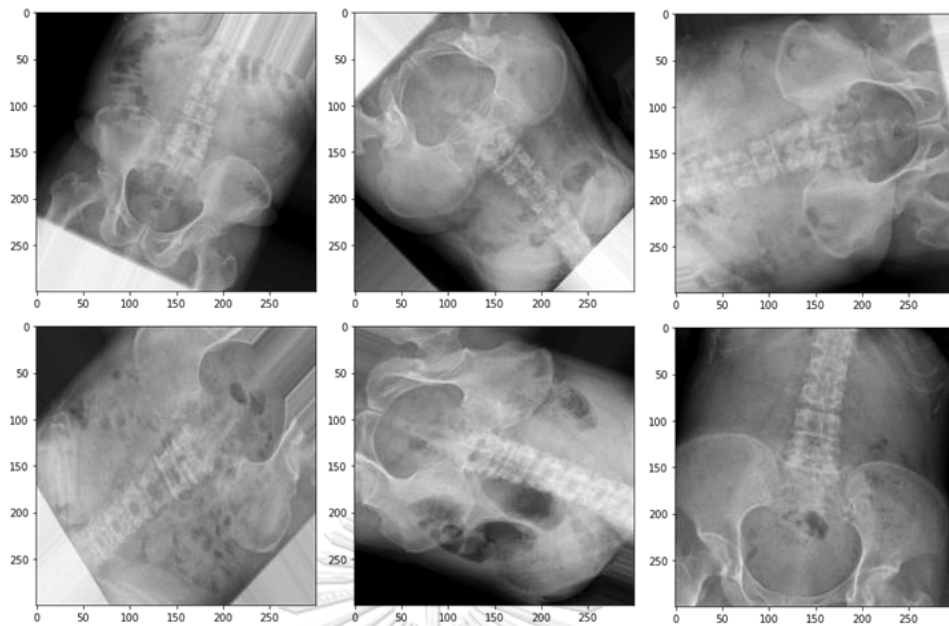


Figure 8 Examples of augmented abdominal radiographs

## 3.2 Model Development

### 3.2.1 Symptom Model

Some popular tree-based machine learning algorithms are applied on the symptom questionnaire since they are easy to use and provide high accuracy with stability. One of the biggest advantages of tree-based models is that they can be visualized and interpreted by humans unlike the black box deep learning models. Aside from decision tree, random forest, and XGBoost, we also applied the recent interpretable deep tabular data learning model, TabNet, in this research.

### 3.2.2 Image Model

We have applied several SOTA image classification models, VGG, DenseNet, ResNet, EfficientNet, and EfficientNetV2, on the abdominal radiographs to find the model which provides the best performance. The best model is used as a backbone and some layers are added to accommodate the task. Dropout layers are also added between the dense layers to reduce overfitting. The final output layer is added as the last layer of the model (Figure 9).



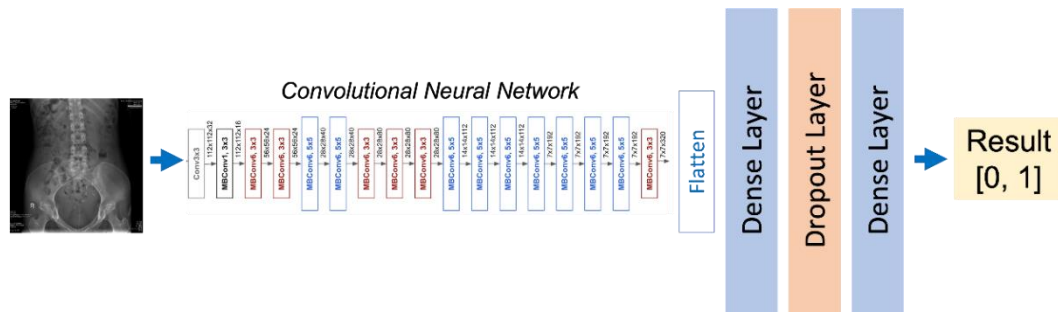


Figure 9 Image model architecture

### 3.2.3 Integrated Model

The integrated model is based on a multiple-input single-output structure with symptom questionnaire and abdominal radiograph as inputs. We decided to combine both data together since they are used by medical experts during the diagnostic process to determine DD.

The architecture of the integrated model is shown in Figure 10. The model consists of 2 input layers: (1) image features and (2) input features from the symptom questionnaire. The image features are extracted using a SOTA image classification model as backbone and some dense layers are also added to reduce the number of parameters to be similar to the input features from the symptom questionnaire. The next layer is the concatenate layer which combines the image features with the input features from the symptom questionnaire. A few dense and dropout layers are also added after the concatenate layer. The final output layer is added as the last layer of the model. The difference between Figure 10(a) and 10(b) is that we added some dense layers to filter the input features from the symptom questionnaire before combining with the image features.

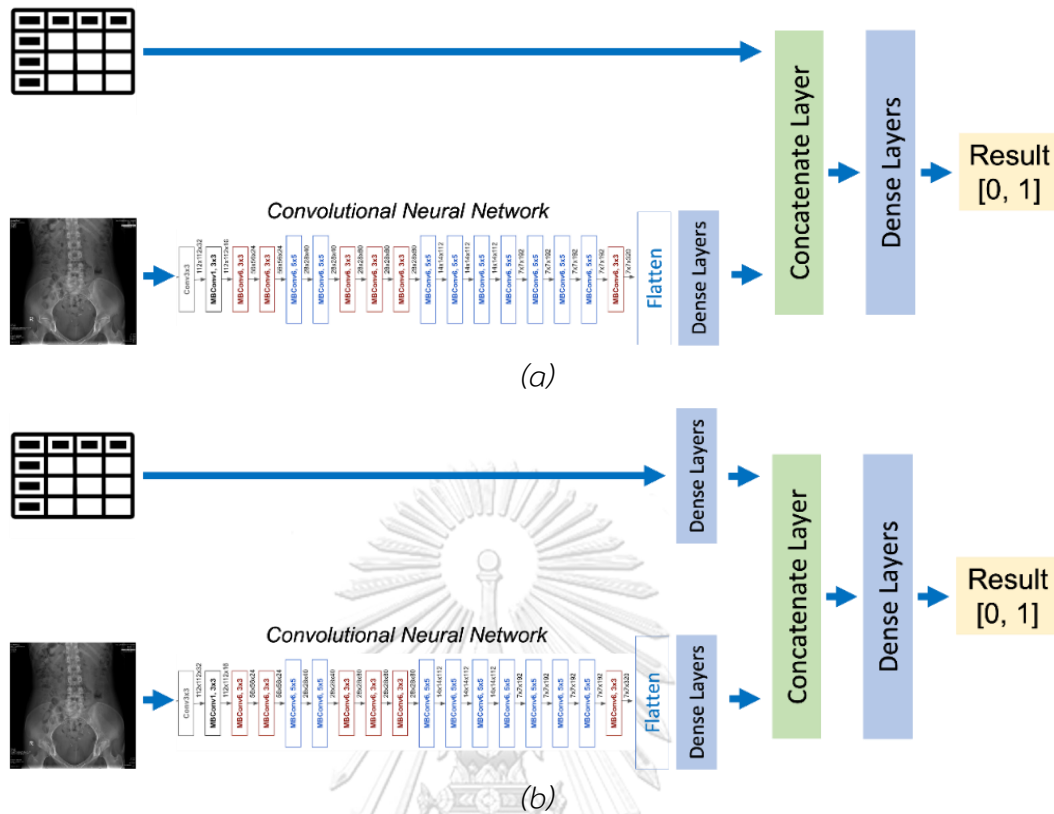


Figure 10 Multiple-input single-output architecture: (a) concatenating input features from the symptom questionnaire directly with image features and (b) adding some dense layer to the input features from the symptom questionnaire before concatenating with the image features

### 3.3 Data Postprocessing

For symptom questionnaire, the decision threshold is adjusted to improve the sensitivity of the tree-based models. Hence, the decision threshold is adjusted from 0.30 to 0.70 and the optimal value which provides the highest sensitivity and adequate specificity is selected. Feature importance calculation and decision tree diagram are also used for error analysis. Evaluating the feature importance allows us to understand which features have a lot of impact to the model than the other. Meanwhile, the decision tree diagram is plotted to demonstrate how the model approach to the conclusions.

For abdominal radiographs, we also adjusted the decision threshold ranging from 0.30 to 0.70 to find the optimal value with the highest sensitivity and adequate specificity. As for model interpretation, Grad-CAM is applied to produce visual explanations which can also be used for error analysis. By applying Grad-CAM, we can see which part of the image the model is focusing and whether it is related to our region of interest or not (Figure 11).

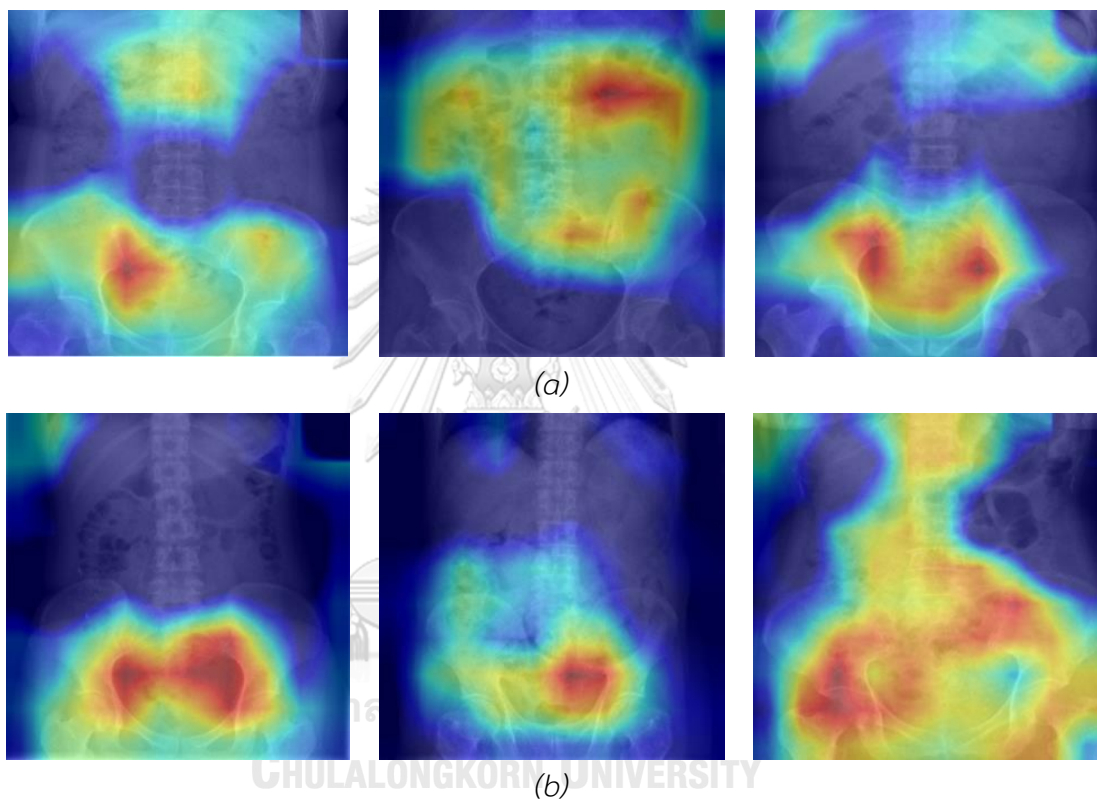


Figure 11 Grad-CAM: (a) patients with NTC (b) patients with DD

## CHAPTER IV

### EXPERIMENTAL SETUP

This chapter describes the datasets along with hyperparameters used in each model. Evaluation metrics used to evaluate the performance of the models are defined at the end of the chapter.

#### 4.1 Dataset

The data were collected from chronic constipated patients who visited the Center of Excellence in Neurogastroenterology and Motility, Faculty of Medicine, Chulalongkorn University during B.E. 2554 – 2563 and undergone ARM, BET, and colonic transit (CTT) tests. The samples were carefully selected according to the following criteria in Table 2.

*Table 2 Diagnostic criteria for DD*

Test \ Outcome	Patients with DD	Patients with Normal Transit Constipation (NTC)	
ARM	✓	✗	✓
BET	✓	✗	✗
CTT	✗	✗	✗

In sum, there are 440 consecutive patients who fulfilled the above criteria. Among the 440 patients, there are 223 patients who were diagnosed as having DD and 217 patients having NTC.

##### 4.1.1 Symptom Questionnaire

Patients are required to complete a validated questionnaire reflecting their gastrointestinal symptoms for the past 3 months when they first visited. Pain severity was graded into mild, moderate, and severe based on how much it affects the patient's day-to-day life on a scale of 0-6. There are a total of 129 questions regarding patient demographics, gastrointestinal symptoms including severity and

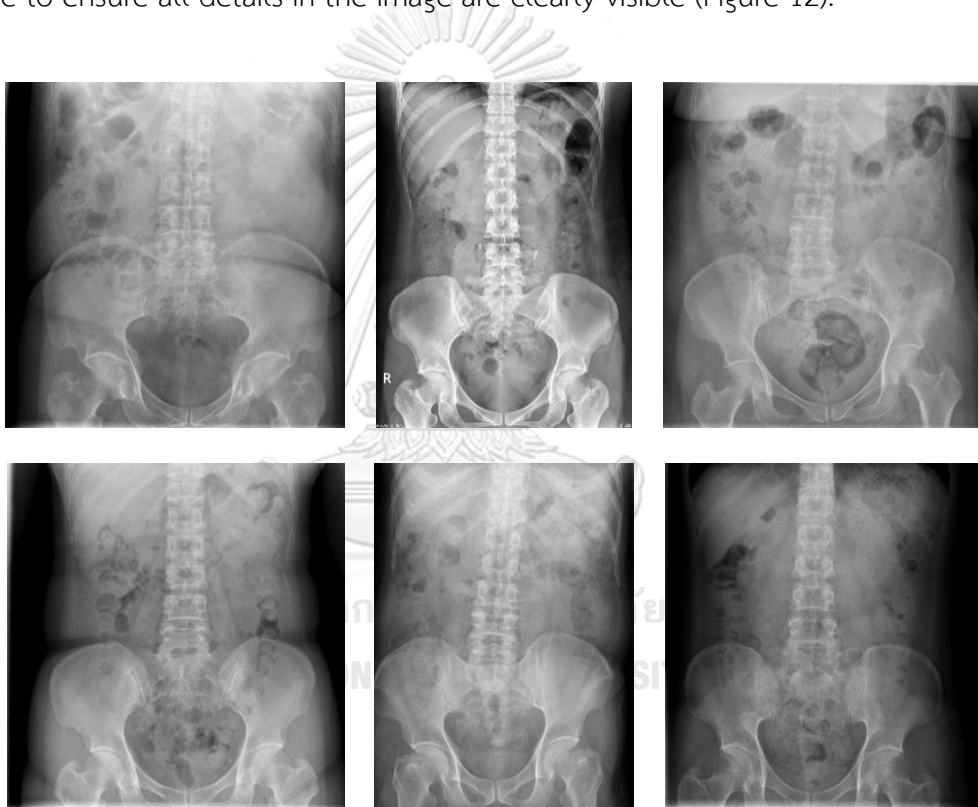
duration, bowel movement details, and a self-reported measure of symptom burden related to constipation. Baseline characteristics and constipation-related symptoms that were asked in the symptom questionnaire are shown in Table 3. In total, there are 440 samples (223 DD and 217 NTC) since each patient only have one set of data. The dataset was divided into a ratio of 80:20 and 5-fold cross-validation was performed during the training process to prevent overfitting.

*Table 3 Symptom questionnaire*

<b>Baseline Characteristics</b>	
Age	Underlying Disease
Gender	Surgical History
Body Mass Index (BMI)	Delivery History
<b>Gastrointestinal Symptoms</b>	
Abdominal Discomfort	Early Satiety and Postprandial Discomfort
Abdominal Pain	Epigastric Burning
Stool Characteristics	Defecation Problems
Fecal Incontinence	Dysphagia
Globus	Acid Regurgitation
Belching	Food Regurgitation
Heartburn	Chest Pain
Nausea and Vomiting	Loss of Appetite
<b>Extra Gastrointestinal Symptoms</b>	
Night Choking	Insomnia
Chronic Cough	Hoarseness of Voice
Sore Throat	Weight Loss

#### 4.1.2 Abdominal Radiography

Abdominal radiographs of KUB view were taken when the patients underwent the CTT test. These radiographs are taken on the fifth day after the patients ingested radio-opaque markers. Since the radiographs of their first and third day are usually also taken for comparison, one patient may have up to 3 images available. In total, there are 764 images (376 DD and 388 NTC). The dataset was divided into train, validate, and test sets in a ratio of 60:20:20. In addition, the abdominal radiographs (KUB view) used in this research were in DICOM format which retains a high-resolution image to ensure all details in the image are clearly visible (Figure 12).



*Figure 12 Abdominal radiographs*

#### 4.1.3 Combined Dataset

Combined dataset consists of both patient's symptom questionnaire and their abdominal radiographs. Since we paired the abdominal radiographs with the symptom questionnaire, the samples in the dataset are equal to the total number of images. In total, there are 764 samples (376 DD and 388 NTC). The dataset was divided into train, validate, and test sets in a ratio of 60:20:20.

Class distribution of each dataset is summarized and shown in Figure 13. The symptom questionnaire has the smallest sample size of 440 samples since each patient only completed the questionnaire once. On the other hand, there are 764 abdominal radiographs available due to several images taken during the CTT test.

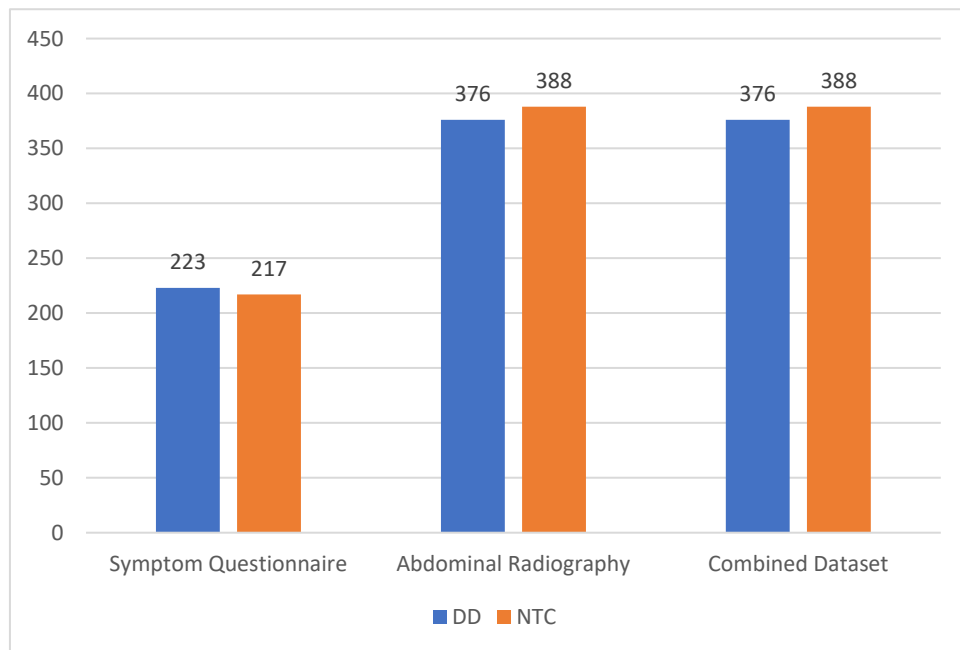


Figure 13 Class distribution in each dataset

## 4.2 Hyperparameters Tuning

### 4.2.1 Symptom Model

To find optimal parameters for each tree-based model and TabNet, GridSearchCV method was used with random state set as 0. During the first-round performance which applies all 129 features, the best parameters for decision tree were *criterion='entropy'*, *max\_depth=12*, and *min\_samples\_leaf=7*. The optimal parameters for random forest were *n\_estimators=300*, *criterion='entropy'*, *max\_depth=2*, and *min\_samples\_leaf=3*. The optimal parameters for XGBoost were *n\_estimator=300*, *max\_depth=16*, and *learning\_rate=0.003*. The optimal parameters for TabNet were *n\_steps=6* and *gamma=1.4*.

During the second-round performance, feature selection was performed and several subsets of input features were created. We applied these subsets on random

forest using GridSearchCV to find the optimal parameters for each dataset. The parameters for each subset of input features are listed in Table 4.

*Table 4 Optimal parameters for random forest during second-round performance (feature selection) from GridSearchCV*

	<i>n_estimators</i>	<i>criterion</i>	<i>max_depth</i>	<i>min_samples_leaf</i>
<b><i>Based on expert knowledge</i></b>				
15 features	250	gini	7	5
68 features	100	gini	10	4
<b><i>Based on traditional method</i></b>				
10 features	500	entropy	2	5
15 features	150	entropy	2	5
20 features	100	gini	3	5
25 features	400	gini	4	10
30 features	100	entropy	3	4

#### 4.2.2 Image Model

A dense layer of 256 with rectified linear activation function is applied between two dropout layers with a dropout rate of 0.4. Adam optimization with learning rate 0.0001, sigmoid function, and binary cross entropy loss were chosen for training the model. Total epochs trained were 200 epochs using early stopping with patience of 50 epochs and batch size of 32.

#### 4.2.3 Integrated Model

The integrated model is trained from scratch using the set of input features obtained from feature selection in the symptom model. The image features are extracted using the same SOTA image classification model as the best image model with two dense layers of 1024 and 64, respectively. For figure 10(b), a dense layer with the same size as the number of input features from the symptom questionnaire is applied to extract the features before concatenating with the image features. After the concatenate layer, two dense layers of 1024 and 256, respectively, are applied. Two dropout layers using a dropout rate of 0.4 is applied after each dense layer to prevent overfitting. Adam optimization using learning rate 0.0001, sigmoid function,



and binary cross entropy are used. Total epochs trained were 500 epochs using early stopping with patience of 80 epochs and batch size of 32.

### 4.3 Evaluation Metrics

For classification tasks, accuracy, F1-score, sensitivity and specificity are common metrics for model evaluation. Positive predictive value (PPV) and negative predictive value (NPV) are also considered in this research since they are crucial metrics in medical field. The mathematic equations of the evaluation metrics are defined using True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Confusion matrix was also constructed to reflect how the model is disorganized and confused while making predictions (Figure 14).

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \quad (4.3.1)$$

$$\text{Sensitivity/Recall} = \frac{TP}{TP + FN} \quad (4.3.2)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4.3.3)$$

$$\text{PPV/Precision} = \frac{TP}{TP + FP} \quad (4.3.4)$$

$$\text{NPV} = \frac{TN}{FN + TN} \quad (4.3.5)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.3.6)$$

	Predicted Negative (0)	Predicted Positive (1)
Actually Negative (0)	TN	FP
Actually Positive (1)	FN	TP

Figure 14 Structure of the confusion matrix

## CHAPTER V

### RESULTS

Chapter 5 shows the results of each dataset: (1) symptom questionnaire, (2) abdominal radiography, and (3) symptom questionnaire and abdominal radiography. Thus, the chapter is divided into 3 parts according to the model that was used for each dataset: (1) symptom model, (2) image model, and (3) integrated model.

#### 5.1 Symptom Model

In the first-round, we trained the tree-based models using all 129 features as input features. The results in Table 5 show that random forest achieved the highest accuracy of 55.68%. Therefore, we decided to apply feature selection using random forest. The feature importance scores were calculated and the top 10, 15, 20, 25, and 30 most important features were created as new subsets of input features. We applied the new subsets to random forest algorithm and compared the results to find the optimal subset of input features (Table 5).

According to Table 5 and Table 6, the best model for the symptom model is the random forest and the best subset of input features consist of 15 features selected based on traditional method. For data postprocessing, we adjusted the decision threshold of the best model varying from 0.30 to 0.70 to find the optimal value. A brief result of this is shown in Table 7 and the full report can be found in Appendix B. Since random forest with default decision threshold of 0.50 on train set provides the best performance, we will apply this value to the test set. As a result, random forest with default decision threshold of 0.50 achieves a sensitivity of 56.95%, specificity of 62.67%, f1-score of 59.76%, and accuracy of 59.77%. In addition, the model performance in each iteration of the 5-fold cross-validation is also shown in Table 8.

Table 5 First-round performance (129 features)

	Sensitivity	Specificity	F1	Accuracy
Decision Tree	56.50	50.69	53.57	53.64
<b>Random Forest</b>	<b>53.81</b>	<b>57.60</b>	<b>55.68</b>	<b>55.68</b>
XGBoost	51.12	55.76	53.40	53.41
TabNet	52.18	57.71	53.85	54.77

Table 6 Second-round performance (feature selection)

	Sensitivity	Specificity	F1	Accuracy
<i>Based on expert knowledge</i>				
15 features	55.16	57.60	56.36	56.36
68 features	54.26	56.68	55.45	55.45
<i>Based on traditional method</i>				
10 features	52.47	64.52	58.30	58.41
<b>15 features</b>	<b>56.95</b>	<b>62.67</b>	<b>59.76</b>	<b>59.77</b>
20 features	55.16	60.37	57.71	57.73
25 features	53.36	62.21	57.67	57.73
30 features	56.05	61.29	58.62	58.64

Table 7 Best symptom model with decision threshold adjustment on train set

	Sensitivity	Specificity	F1	Accuracy
0.46	87.33	27.42	53.12	57.78
0.47	81.95	38.82	58.33	60.68
0.48	75.89	48.15	61.16	62.22
0.49	70.29	58.40	64.17	64.43
<b>0.50</b>	<b>63.78</b>	<b>68.19</b>	<b>65.90</b>	<b>65.97</b>
0.51	56.38	75.68	65.57	65.91
0.52	48.87	80.87	63.72	64.66
0.53	41.13	85.25	61.04	62.90
0.54	33.85	89.63	58.24	61.36
0.55	28.59	91.82	55.46	59.77

*Table 8 Best symptom model performance in each iteration of 5-fold cross-validation*

	Sensitivity	Specificity	F1	Accuracy
1-fold	51.11	55.81	53.40	53.41
2-fold	60.00	72.09	65.84	65.91
3-fold	57.78	79.07	67.92	68.18
4-fold	61.36	50.00	55.54	55.68
5-fold	54.55	56.82	55.68	55.68

## 5.2 Image Model

Table 9 and Table 10 show the results of SOTA image classification models and the results of EfficientNet-B0 to B3 performance on plain abdominal radiograph (original image), respectively. We also attempted several setups for the dataset such as cropping and contrasting the images to find the suitable customization for the dataset. The model with the best performance is applied on 6 datasets with different augmentation techniques to find the most suitable dataset for the task. The results from Table 11 show that applying EfficientNet-B3 on plain abdominal radiograph provides the best performance. Therefore, we adjusted the decision threshold of EfficientNet-B3 on plain abdominal radiograph varying from 0.30 to 0.70 on the validate set to find the optimal value. A brief result can be found in Table 12, while the full report can be found in Appendix B.

As a result, the decision threshold of 0.54 provides the best performance on validate set. After applying this value to EfficientNet-B3 on the test set using plain abdominal radiograph, we achieved a sensitivity of 61.04%, specificity of 56.58%, f1-score of 58.80%, and accuracy of 58.82%.

*Table 9 SOTA image classification model performance*

	Sensitivity	Specificity	F1	Accuracy
VGG19	24.36	77.33	46.88	50.33
DenseNet121	21.79	69.33	42.07	45.10
ResNet50	52.56	54.67	53.59	53.59
<b>EfficientNet-B3</b>	<b>68.83</b>	<b>52.63</b>	<b>60.50</b>	<b>60.78</b>
EfficientNetV2-s	11.54	88.00	40.80	49.02

Table 10 EfficientNet model performance

	Sensitivity	Specificity	F1	Accuracy
EfficientNet-B0	46.15	48.00	47.06	47.06
EfficientNet-B1	74.36	33.33	52.02	54.25
EfficientNet-B2	35.90	52.00	43.50	43.79
<b>EfficientNet-B3</b>	<b>68.83</b>	<b>52.63</b>	<b>60.50</b>	<b>60.78</b>

Table 11 Datasets with different setups using EfficientNet-B3

	Sensitivity	Specificity	F1	Accuracy
<b>Original image</b>	<b>68.83</b>	<b>52.63</b>	<b>60.50</b>	<b>60.78</b>
Original image with CLAHE	67.95	25.33	44.31	47.06
Auto crop	57.69	48.00	52.78	52.94
Auto crop with auto contrast	81.82	14.47	41.62	48.37
Manual crop	53.95	56.58	55.26	55.26
Manual crop with manual contrast	63.16	42.11	52.10	52.63

Table 12 Best image model with decision threshold adjustment on validate set

	Sensitivity	Specificity	F1	Accuracy
0.46	91.03	16.00	46.26	54.25
0.47	88.46	21.33	49.50	55.56
0.48	84.62	26.67	51.86	56.21
0.49	80.77	30.67	53.00	56.21
0.50	74.36	36.00	53.65	55.56
0.51	73.08	44.00	57.78	58.82
0.52	67.95	52.00	59.79	60.13
0.53	62.82	56.00	59.39	59.48
<b>0.54</b>	<b>60.26</b>	<b>62.67</b>	<b>61.44</b>	<b>61.44</b>
0.55	56.41	66.67	61.38	61.44

### 5.3 Integrated Model

According to prior experiments, EfficientNet-B3 was chosen as backbone to extract the feature vectors from the X-ray image before concatenating with the selected input features from symptom questionnaire. Table 13 demonstrates the performance of our proposed integrated models (see Figure 10) on 4 different sets of input features from the symptom questionnaire, i.e., all 129 features, 15 and 68

features selected based on expert knowledge, and 15 features selected based on traditional method.

From Table 14, the integrated model from Figure 10(a) which uses 15 input features selected based on expert knowledge achieves the highest accuracy. Therefore, we adjusted the decision threshold of the model from 0.30 to 0.70 on the validate set to find the optimal decision threshold. Hence, a brief result is shown in Table 14 and the full report can be found in Appendix B. As a result, the decision threshold of 0.49 provides the best performance on validate set. After applying this value to the test set, we achieved a sensitivity of 74.36%, specificity of 57.33%, f1-score of 65.68%, and accuracy of 66.01%.

Table 13 Integrated model performance (from Figure 10)

	Sensitivity	Specificity	F1	Accuracy
<b>Figure 10(a)</b>				
<i>Based on expert knowledge</i>				
<b>15 features</b>	<b>71.19</b>	<b>62.67</b>	<b>67.21</b>	<b>67.32</b>
68 features	62.82	58.67	60.74	60.78
<i>Based on traditional method</i>				
15 features	66.67	42.67	54.12	54.90
129 features	61.54	56.00	58.76	58.82
<b>Figure 10(b)</b>				
<i>Based on expert knowledge</i>				
15 features	21.79	62.67	39.50	41.83
68 features	96.00	5.13	37.29	49.67
<i>Based on traditional method</i>				
15 features	96.15	6.67	39.66	52.29
129 features	46.15	77.33	60.62	61.44

*Table 14 Best integrated model with decision threshold adjustment on validate set*

	Sensitivity	Specificity	F1	Accuracy
0.46	74.36	28.00	48.63	51.63
0.47	69.23	37.33	52.21	53.59
0.48	65.38	41.33	52.79	53.59
<b>0.49</b>	<b>65.38</b>	<b>44.00</b>	<b>54.27</b>	<b>54.90</b>
0.50	62.82	45.33	53.80	54.25
0.51	55.13	52.00	53.56	53.59
0.52	52.56	53.33	52.94	52.94
0.53	48.72	53.33	50.97	50.98
0.54	46.15	57.33	51.53	51.63
0.55	43.59	57.33	50.15	50.33

In summary, the results of the best model from each dataset (i.e., symptom questionnaire, abdominal radiography, and combined dataset) are demonstrated in Table 15. For symptom model, applying random forest on 15 input features selected based on traditional method provides an accuracy of 59.77%. For image model, applying EfficientNet-B3 on plain abdominal radiograph provides an accuracy of 58.82%. For integrated model, applying integrated model without additional dense layers on 15 input features selected based on expert knowledge and using EfficientNet-B3 as backbone provides the highest accuracy of 66.01% which outperforms other baseline models.

*Table 15 Model performance comparison on each dataset*

	Sensitivity	Specificity	F1	Accuracy
Symptom Model	56.95	62.67	59.76	59.77
Image Model	61.04	56.58	58.80	58.82
<b>Integrated Model</b>	<b>74.36</b>	<b>57.33</b>	<b>65.68</b>	<b>66.01</b>

Although the symptom model which utilizes 15 input features selected based on traditional method outperforms the symptom model which utilizes 15 input features selected based on expert knowledge, the integrated model achieves better performance than other baseline models while using 15 input features selected based on expert knowledge with EfficientNet-B3 as backbone. Therefore, we

attempted to investigate the difference between datasets selected based on expert knowledge and traditional method by plotting feature importance bar charts for each dataset. According to Figure 15, stool frequency, duration of abdominal bloating, and duration of abdominal distention were chosen by both expert and model. We can see that medical experts tended to choose features related to gastrointestinal symptoms, while the features selected based on traditional method also include baseline characteristics and extra gastrointestinal symptoms, especially numerical data. Thus, future research should focus on the relationship between each characteristic or symptom and the target variable to improve the model performance.

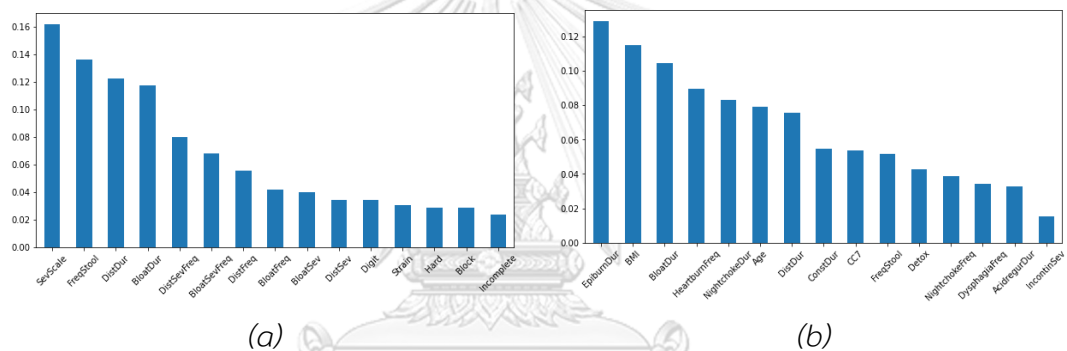


Figure 15 Feature importance bar chart: (a) input features based on expert knowledge and (b) input features based on traditional method



## CHAPTER VI

### CONCLUSION

#### 6.1 Conclusion

To our knowledge, this is the first report of DD diagnosis using deep learning approach on easily obtainable clinical data such as symptom questionnaire and abdominal radiography. We proposed deep learning-based models using 3 different datasets: (1) symptom questionnaire, (2) abdominal radiography, and (3) symptom questionnaire and abdominal radiography. We also presented several preprocessing and postprocessing techniques for both tabular data and image data to overcome the problem of small size data. The results show that our proposed integrated model which uses both type of data outperforms the baseline models which uses only one type of data. This is very much the key component in future attempts to improve the accuracy and efficiency of deep learning-based models for DD diagnosis.

In the end, our proposed deep learning-based models which utilize clinical data obtained from primary and secondary healthcare centers are able to support clinicians prescreen potential patients with promising accuracy. Furthermore, we provided several alternatives by using only symptom questionnaire as an input, using only abdominal radiograph as an input, or using both symptom questionnaire and abdominal radiograph as inputs. Therefore, primary and secondary healthcare centers shall be able to prescreen patients with possibility of having DD to be transferred to tertiary healthcare centers for specialized tests and treatment.

#### 6.2 Future Work

Future research should examine the features in symptom questionnaire and the appropriate image augmentation techniques for the abdominal radiographs. Data preprocessing and postprocessing should integrate expert knowledge to maximize the quality of inputs. Feature selection should also be investigated. By enhancing the data quality, we can improve the performance of the deep learning-based models.

## REFERENCES

1. Basilisco, G. and M. Coletta, *Chronic constipation: A critical review*. Digestive and Liver Disease, 2013. **45**(11): p. 886-893.
2. Rao, S.S.C. and T. Patcharatrakul, *Diagnosis and Treatment of Dyssynergic Defecation*. Journal of neurogastroenterology and motility, 2016. **22**(3): p. 423-435.
3. Rao, S.S.C., et al., *Dyssynergic Defecation: Demographics, Symptoms, Stool Patterns, and Quality of Life*. Journal of Clinical Gastroenterology, 2004. **38**(8).
4. Rao, S.S.C., *Dyssynergic defecation and biofeedback therapy*. Gastroenterology clinics of North America, 2008. **37**(3): p. 569-viii.
5. Tanner, S., et al., *Prevalence and Clinical Characteristics of Dyssynergic Defecation and Slow Transit Constipation in Patients with Chronic Constipation*. Journal of clinical medicine, 2021. **10**(9): p. 2027.
6. Parker, C.H., et al., *The Use of Standardized Questions in Identifying Patients with Dyssynergic Defecation*. Journal of the Canadian Association of Gastroenterology, 2018. **1**(2): p. 60-66.
7. Park, S.-Y., et al., *Rectal Gas Volume Measured by Computerized Tomography Identifies Evacuation Disorders in Patients With Constipation*. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association, 2017. **15**(4): p. 543-552.e4.
8. Ahuja, A.S., *The impact of artificial intelligence in medicine on the future role of the physician*. PeerJ, 2019. **7**: p. e7702-e7702.
9. Obermeyer, Z. and E.J. Emanuel, *Predicting the Future - Big Data, Machine Learning, and Clinical Medicine*. N Engl J Med, 2016. **375**(13): p. 1216-9.
10. Bohr, A. and K. Memarzadeh, *The rise of artificial intelligence in healthcare applications*. Artificial Intelligence in Healthcare, 2020: p. 25-60.
11. Ruffle, J., et al., *PWE-097 Machine learning can accurately classify chronic constipation patients by symptom burden using pain measures alone*. Gut, 2019. **68**(Suppl 2): p. A218.

12. Schwartz-Ziv, R. and A. Armon, *Tabular data: Deep learning is not all you need*. Information Fusion, 2022. **81**: p. 84-90.
13. Banerjee, M., et al., *Tree-Based Analysis*. Circulation. Cardiovascular quality and outcomes, 2019. **12**(5): p. e004879-e004879.
14. Tin Kam, H., *The random subspace method for constructing decision forests*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998. **20**(8): p. 832-844.
15. Chen, T. and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, Association for Computing Machinery: San Francisco, California, USA. p. 785–794.
16. Arik, S.Ö. and T. Pfister, *TabNet: Attentive Interpretable Tabular Learning*. Proceedings of the AAAI Conference on Artificial Intelligence, 2021. **35**(8): p. 6679-6687.
17. Çallı, E., et al., *Deep learning for chest X-ray analysis: A survey*. Medical Image Analysis, 2021. **72**: p. 102125.
18. Diallo, P.A.K.K. and Y. Ju. *Accurate Detection of COVID-19 Using K-EfficientNet Deep Learning Image Classifier and K-COVID Chest X-Ray Images Dataset*. in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*. 2020.
19. Chetoui, M., et al., *Explainable COVID-19 Detection on Chest X-rays Using an End-to-End Deep Convolutional Neural Network Architecture*. Big Data and Cognitive Computing, 2021. **5**(4).
20. El Asnaoui, K. and Y. Chawki, *Using X-ray images and deep learning for automated detection of coronavirus disease*. Journal of Biomolecular Structure and Dynamics, 2021. **39**(10): p. 3615-3626.
21. Khan, I.U., et al., *Using a Deep Learning Model to Explore the Impact of Clinical Data on COVID-19 Diagnosis Using Chest X-ray*. Sensors (Basel), 2022. **22**(2).
22. Tan, M. and Q. Le, *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*, in *Proceedings of the 36th International Conference on Machine Learning*, C. Kamalika and S. Ruslan, Editors. 2019, PMLR: Proceedings

- of Machine Learning Research. p. 6105--6114.
23. Tan, M. and Q. Le, *EfficientNetV2: Smaller Models and Faster Training*, in *Proceedings of the 38th International Conference on Machine Learning*, M. Marina and Z. Tong, Editors. 2021, PMLR: Proceedings of Machine Learning Research. p. 10096--10106.
  24. He, K., et al. *Deep Residual Learning for Image Recognition*. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.
  25. Simonyan, K. and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv 1409.1556, 2014.
  26. Huang, G., et al. *Densely Connected Convolutional Networks*. in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.
  27. Park, Y., A.-C. Hauschild, and D. Heider, *Transfer learning compensates limited data, batch effects and technological heterogeneity in single-cell sequencing*. *NAR genomics and bioinformatics*, 2021. **3**(4): p. lqab104-lqab104.
  28. Pizer, S.M., et al., *Adaptive histogram equalization and its variations*. *Computer Vision, Graphics, and Image Processing*, 1987. **39**(3): p. 355-368.
  29. Mishra, A., *Contrast Limited Adaptive Histogram Equalization (CLAHE) Approach for Enhancement of the Microstructures of Friction Stir Welded Joints*. arXiv preprint arXiv:2109.00886, 2021.
  30. Selvaraju, R.R., et al. *Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization*. in *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017.
  31. Levy, J.J., et al., *Video-Based Deep Learning to Detect Dyssynergic Defecation with 3D High-Definition Anorectal Manometry*. *bioRxiv*, 2021: p. 2021.12.11.472233.



APPENDICES

จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

**APPENDIX A**  
**LIST OF ABBREVIATIONS**

Abbreviation	Description
AcidregurDur	Duration of acid regurgitation
BloatDur	Duration of abdominal bloating
BloatFreq	Abdominal bloating frequency
BloatSev	Abdominal bloating severity
BloatSevFreq	Combination of abdominal bloating severity and frequency
Block	Sense of blockage
CC7	Chief complaint of abdominal distension
ConstDur	Duration of constipation
Digit	Using digital maneuvers
DistDur	Duration of abdominal distension
DistFreq	Abdominal distension frequency
DistSev	Abdominal distension severity
DistSevFreq	Combination of abdominal distension severity and frequency
DysphagiaFreq	Dysphagia frequency
EpibumDur	Duration of epigastric burning
FreqStool	Stool frequency/week
Hard	Hard stool
HeartburnFreq	Heartburn frequency
Incomplete	Incomplete evacuation
IncontinSev	Fecal incontinence severity
NightchokeFreq	Duration of night choking
SevScale	Disturbing severity scale
Strain	Excessive straining

**APPENDIX B**  
**FULL REPORT OF DECISION THRESHOLD ADJUSTMENT**

*Table B. 1 Full report of best symptom model with decision threshold adjustment*

	Sensitivity	Specificity	F1	Accuracy
0.30	100.00	0	33.63	50.68
0.31	100.00	0	33.63	50.68
0.32	100.00	0	33.63	50.68
0.33	100.00	0	33.63	50.68
0.34	100.00	0	33.63	50.68
0.35	100.00	0.12	33.76	50.74
0.36	100.00	0.12	33.76	50.74
0.37	100.00	0.12	33.76	50.74
0.38	100.00	0.12	33.76	50.74
0.39	100.00	0.46	34.14	50.91
0.40	99.89	1.39	35.11	51.31
0.41	99.78	2.42	36.19	51.76
0.42	99.33	5.53	39.23	53.07
0.43	98.65	8.30	41.72	54.09
0.44	97.08	12.33	44.98	55.28
0.45	92.60	18.78	48.80	56.19
0.46	87.33	27.42	53.12	57.78
0.47	81.95	38.82	58.33	60.68
0.48	75.89	48.15	61.16	62.22
0.49	70.29	58.40	64.17	64.43
<b>0.50</b>	<b>63.78</b>	<b>68.19</b>	<b>65.90</b>	<b>65.97</b>
0.51	56.38	75.68	65.57	65.91
0.52	48.87	80.87	63.72	64.66
0.53	41.13	85.25	61.04	62.90
0.54	33.85	89.63	58.24	61.36
0.55	28.59	91.82	55.46	59.77
0.56	25.00	94.24	53.78	59.15
0.57	22.64	95.51	52.46	58.58
0.58	18.94	96.66	49.89	57.27
0.59	16.70	97.46	48.33	56.53
0.60	14.35	98.04	46.52	55.63
0.61	12.89	98.73	45.47	55.23
0.62	11.55	99.19	44.41	54.77
0.63	10.20	99.31	43.22	54.15
0.64	8.86	99.54	41.99	53.58
0.65	7.51	99.65	40.73	52.95
0.66	6.50	99.77	39.78	52.50
0.67	5.72	99.89	39.02	52.16
0.68	4.71	99.89	37.98	51.65
0.69	4.37	99.89	37.65	51.48
0.70	3.70	100.00	36.98	51.19

Table B. 2 Full report of best image model with decision threshold adjustment

	Sensitivity	Specificity	F1	Accuracy
0.30	98.72	0	33.48	50.33
0.31	98.72	1.33	34.92	50.98
0.32	98.72	1.33	34.92	50.98
0.33	98.72	1.33	34.92	50.98
0.34	98.72	1.33	34.92	50.98
0.35	98.72	1.33	34.92	50.98
0.36	98.72	2.67	36.34	51.63
0.37	98.72	2.67	36.34	51.63
0.38	98.72	2.67	36.34	51.63
0.39	98.72	2.67	36.34	51.63
0.40	98.72	2.67	36.34	51.63
0.41	98.72	4.00	37.72	52.29
0.42	98.72	6.67	40.40	53.59
0.43	96.15	6.67	39.66	52.29
0.44	93.59	8.00	40.16	51.63
0.45	92.31	10.67	42.17	52.29
0.46	91.03	16.00	46.26	54.25
0.47	88.46	21.33	49.50	55.56
0.48	84.62	26.67	51.86	56.21
0.49	80.77	30.67	53.00	56.21
0.50	74.36	36.00	53.65	55.56
0.51	73.08	44.00	57.78	58.82
0.52	67.95	52.00	59.79	60.13
0.53	62.82	56.00	59.39	59.48
<b>0.54</b>	<b>60.26</b>	<b>62.67</b>	<b>61.44</b>	<b>61.44</b>
0.55	56.41	66.67	61.38	61.44
0.56	47.44	72.00	58.97	59.48
0.57	33.33	74.67	51.74	53.59
0.58	24.36	51.00	49.41	53.59
0.59	15.38	88.00	44.01	50.98
0.60	14.10	92.00	44.28	52.29
0.61	10.26	93.33	41.35	50.98
0.62	6.41	96.00	38.54	50.33
0.63	6.41	98.67	39.29	51.63
0.64	5.13	95.67	38.00	50.98
0.65	1.28	98.67	33.99	49.02
0.66	1.28	98.67	33.99	49.02
0.67	1.28	98.67	33.99	49.02
0.68	1.28	98.67	33.99	49.02
0.69	1.28	98.67	33.99	49.02
0.70	0	98.67	32.60	48.37



Table B. 3 Full report of best integrated model with decision threshold adjustment

	Sensitivity	Specificity	F1	Accuracy
0.30	98.72	4.00	37.72	52.29
0.31	98.72	5.33	39.07	52.94
0.32	98.72	6.67	40.40	53.59
0.33	96.15	6.67	39.66	52.29
0.34	92.31	6.67	38.54	50.33
0.35	91.03	6.67	38.17	49.67
0.36	91.03	6.67	38.17	49.67
0.37	91.03	6.67	38.17	49.67
0.38	91.03	6.67	38.17	49.67
0.39	88.46	6.67	37.42	48.37
0.40	88.46	6.67	37.42	48.37
0.41	88.46	8.00	38.61	49.02
0.42	83.33	9.33	38.17	47.06
0.43	80.77	16.00	42.65	49.02
0.44	79.49	16.00	42.19	48.37
0.45	78.21	25.33	48.40	52.29
0.46	74.36	28.00	48.63	51.63
0.47	69.23	37.33	52.21	53.59
0.48	65.38	41.33	52.79	53.59
<b>0.49</b>	<b>65.38</b>	<b>44.00</b>	<b>54.27</b>	<b>54.90</b>
0.50	62.82	45.33	53.80	54.25
0.51	55.13	52.00	53.56	53.59
0.52	52.56	53.33	52.94	52.94
0.53	48.72	53.33	50.97	50.98
0.54	46.15	57.33	51.53	51.63
0.55	43.59	57.33	50.15	50.33
0.56	42.31	57.33	49.46	49.67
0.57	41.03	66.67	52.94	53.59
0.58	33.33	68.00	48.96	50.33
0.59	30.77	69.33	47.93	49.67
0.60	29.49	70.67	47.66	49.67
0.61	28.21	70.67	46.84	49.02
0.62	26.92	74.67	47.58	50.33
0.63	24.36	77.33	46.88	50.33
0.64	20.51	78.67	44.65	49.02
0.65	20.51	78.67	44.65	49.02
0.66	19.23	81.33	44.67	49.67
0.67	19.23	81.33	44.67	49.67
0.68	19.23	82.67	45.15	50.33
0.69	19.23	82.67	45.15	50.33
0.70	19.23	82.67	45.15	50.33

## VITA

NAME	Sornsiri Poovongsaroj
DATE OF BIRTH	9 December 1999
PLACE OF BIRTH	Bangkok, Thailand
INSTITUTIONS ATTENDED	Civil Aviation Training Center
HOME ADDRESS	616/18 Techawanit Road Bang Sue, Bangkok Thailand 10800



จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**