

การตรวจจับสิ่งแปลกปลอมบนขวดแก้วด้วยโครงข่ายประสาทเชิงลึก

ANOMALY DETECTION ON GLASS BOTTLE USING DEEP NEURAL  
NETWORK



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาปรัชญาดุษฎีบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์  
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2566

KMITL-2023-SC-D-002-058

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# ANOMALY DETECTION ON GLASS BOTTLE USING DEEP NEURAL NETWORK



A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE  
DEGREE OF DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE  
DEPARTMENT OF COMPUTER SCIENCE SCHOOL OF SCIENCE  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2023

KMITL-2023-SC-D-002-058

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2023

SCHOOL OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การตรวจจับสิ่งแปลกปลอมบนขวดแก้วด้วยโครงข่ายประสาทเชิงลึก
ชื่อนักศึกษา	นิพัทธ์ คล้ายโพธิ์
รหัสประจำตัว	58605020
ปริญญา	ปรัชญาดุษฎีบัณฑิต (วิทยาการคอมพิวเตอร์)
ภาควิชา	วิทยาการคอมพิวเตอร์
พ.ศ.	2566
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผู้ช่วยศาสตราจารย์ ดร.อนันตพร ทรราชคุณด้าย

### บทคัดย่อ

กระบวนการตรวจจับสิ่งแปลกปลอมบนขวดแก้วเป็นส่วนสำคัญของกระบวนการควบคุมคุณภาพในอุตสาหกรรมการผลิตขวดแก้ว ในการผลิตขวดแก้วต้องทำการตรวจสอบคุณภาพของขวดก่อนส่งให้ลูกค้า ในปัจจุบันได้มีการประยุกต์ใช้เทคโนโลยีแมชชีนวิชันเพื่อทำการวิเคราะห์หาสิ่งแปลกปลอมบนขวดแก้วโดยอัตโนมัติ ในงานวิจัยนี้เสนอวิธีการตรวจจับสิ่งแปลกปลอมสำหรับงานตรวจสอบคุณภาพบนขวดแก้วสองวิธีการ 1) วิธีการตรวจจับสิ่งแปลกปลอมแบบผสมโดยใช้ CNNLSTM และการจำแนกประเภทตามตัวอย่าง (instance-based) ที่เป็นวิธีที่มีความแม่นยำสูงในการตรวจจับสิ่งแปลกปลอม 2) วิธีการตรวจจับสิ่งแปลกปลอมโดยใช้โครงข่ายประสาทเชิงลึกแบบ Autoencoder และ Threshold ของข้อผิดพลาดที่เรียกว่า DDAET ข้อดีของวิธีการนี้คือสามารถตรวจสอบได้รวดเร็วโดยใช้สถาปัตยกรรมของโครงข่าย Autoencoder ซึ่งออกแบบให้มีขนาดเล็กและใช้ชุดข้อมูลภาพขวดดีในการฝึกสอนเท่านั้น ประสิทธิภาพของวิธีการที่นำเสนอทั้งสองวิธีการถูกนำไปเปรียบเทียบกับวิธีอื่นๆ หลายวิธีคือ ResNet50 VGG16 MobileNetV3 ADA Edge detection และ Image threshold ในงานวิจัยนี้ใช้ชุดข้อมูลในการทดลองสามชุดได้แก่ ชุดข้อมูลภาพบริเวณลำตัวของขวด ชุดข้อมูลภาพคอขวด และชุดข้อมูลภาพขวดหมุน ผลการทดลองแสดงให้เห็นว่าวิธีการ CNNLSTM ที่นำเสนอให้ความแม่นยำที่สูงสำหรับชุดข้อมูลภาพขวดหมุน ซึ่งให้ค่าความถูกต้องเท่ากับ 97.66% DDAET เป็นวิธีการที่มีความแม่นยำและมีความรวดเร็วในการตรวจจับสิ่งแปลกปลอมมากที่สุดสำหรับชุดข้อมูลข้อมูลภาพบริเวณลำตัวของขวด และชุดข้อมูลภาพคอขวด ซึ่งให้ค่าความถูกต้องเท่ากับ 80.57% สำหรับชุดข้อมูลข้อมูลภาพบริเวณลำตัวของขวด และ 92.43% สำหรับชุดข้อมูลภาพคอขวด

**คำสำคัญ :** การตรวจจับสิ่งแปลกปลอม การเรียนรู้ของเครื่อง โครงข่ายประสาทเชิงลึก แมชชีนวิชัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

<b>Title</b>	Anomaly Detection on Glass Bottle Using Deep Neural Network
<b>Student Name</b>	Niphath Claypo
<b>Student ID</b>	58605020
<b>Degree</b>	Doctor of Philosophy (Computer Science)
<b>Department</b>	Computer Science
<b>Year</b>	2023
<b>Thesis Advisor</b>	Asst. Prof. Dr. Anantaporn Hanskunatai

### Abstract

The process of detecting anomalies on glass bottles is an essential part of quality control in the glass bottle manufacturing industry. Glass bottles must undergo quality inspection before being delivered to customers. Currently, machine vision technology is being applied for automatic analysis and identification of anomalies on glass bottles. This research proposes two methods for detecting anomalies in glass bottle quality inspection: 1) A hybrid method that combines CNN-LSTM with instance-based classification, achieving high accuracy in detecting counterfeit anomalies. 2) An anomaly detection method using the Autoencoder deep neural network, and a threshold is called DDAET. DDAET offers the advantage of fast detection using a small-sized Autoencoder architecture, trained solely on a good bottle image dataset. The effectiveness of both proposed methods is compared with several other techniques, including ResNet50, VGG16, MobileNetV3, ADA Edge detection, and Image threshold. This research conducted experiments on three datasets: images of the bottle body region, neck images, and rotated bottle images. The experimental results show that the CNNLSTM method presented achieves high accuracy for the rotated bottle image dataset, with an accuracy rate of 97.66%. DDAET is the most accurate and fastest method for detecting counterfeit anomalies on the bottle body and neck datasets, with accuracy rates of 80.57% and 92.43%, respectively, for each dataset.

**Keywords:** Anomaly Detection, Machine Learning, Deep Neural Networks, Machine Vision

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## กิตติกรรมประกาศ

วิทยานิพนธ์นี้มีอาจสำเร็จลุล่วงไปได้ด้วยดีหากไม่ได้รับคำแนะนำ คำชี้แจง ความรู้และความเอาใจใส่จาก ผู้ช่วยศาสตราจารย์ ดร.อนันตพร ทรราชคุณชัย ผู้เป็นอาจารย์ที่ปรึกษา ซึ่งท่านได้สละเวลาให้กับข้าพเจ้าอย่างเต็มที่ จึงใคร่ขอขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณรองศาสตราจารย์ ดร.อนุชิต จิตพัฒนกุล ผู้ช่วยศาสตราจารย์ ดร.ศรัณย์ อินทโกสุม ผู้ช่วยศาสตราจารย์ ดร.นวลสวาท หิรัญสกุลวงศ์ และผู้ช่วยศาสตราจารย์ ดร.ปัทมา เจริญพร ที่กรุณาเป็นคณะกรรมการสอบหัวข้อวิทยานิพนธ์ในครั้งนี้

ขอขอบคุณบุคลากร คณะวิทยาศาสตร์สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุกๆท่านที่ให้ความช่วยเหลือและคำปรึกษาตลอดระยะเวลาการดำเนินงานวิจัยจนสำเร็จลุล่วงไปได้ด้วยดี

ขอกราบขอบพระคุณบิดา มารดา สมาชิกในครอบครัว ตลอดจนญาติพี่น้อง และเพื่อนๆทุกคน ที่สนับสนุนและเป็นกำลังใจในระหว่างการศึกษาเป็นอย่างดี ขอขอบคุณทุกๆ ท่านที่ไม่ได้กล่าวมา ณ ที่นี้ที่มีส่วนร่วมในการทำวิทยานิพนธ์ในครั้งนี้

นายนิพัทธ์ คล้ายโพธิ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	จ
สารบัญภาพ	ฉ
<b>บทที่ 1 บทนำ</b>	<b>1</b>
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย	2
1.3 ขอบเขตของงานวิจัย	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
<b>บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง</b>	<b>4</b>
2.1 ทฤษฎีที่เกี่ยวข้อง	4
2.1.1 การเรียนรู้เชิงลึก (Deep learning)	4
2.1.2 Convolutional Neural Network	4
2.1.3 AlexNet	6
2.1.4 VGG16	7
2.1.5 ResNet50	8
2.1.6 MobileNet	9
2.1.7 Long short-term memory (LSTM)	9
2.1.8 โครงข่ายประสาทแบบ Autoencoder	10
2.1.9 การขยายข้อมูล (Data augmentation)	11
2.2 งานวิจัยที่เกี่ยวข้อง	12
<b>บทที่ 3 วิธีการดำเนินงานวิจัย</b>	<b>16</b>
3.1 ชุดข้อมูลภาพสิ่งแปลกปลอม	16
3.2 วิธีการตรวจจับสิ่งผิดปกติแบบผสมโดยใช้ CNNLSTM และการจำแนกตามตัวอย่าง	18
3.2.1 สถาปัตยกรรมของโครงข่ายประสาท CNNLSTM	18
3.2.2 วิธีการเรียนรู้ข้อมูลสิ่งแปลกปลอม	19
3.2.3 วิธีการตรวจจับสิ่งแปลกปลอมด้วยวิธีการจำแนกประเภทตามตัวอย่าง	23

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3	วิธีการตรวจจับสิ่งแปลกปลอมโดยใช้โครงข่ายประสาทแบบ Autoencoder และ Threshold ของข้อผิดพลาด	25
3.3.1	สถาปัตยกรรมของ Autoencoder	25
3.3.2	สร้างภาพพื้นหลังปกติและค่า Threshold ของข้อผิดพลาด	26
3.3.3	การตรวจจับสิ่งแปลกปลอมด้วยภาพพื้นหลังปกติและค่า Threshold ของข้อผิดพลาด	29
<b>บทที่ 4</b>	<b>ผลการวิจัยและการอภิปรายผล</b>	<b>33</b>
4.1	ชุดข้อมูลภาพสิ่งแปลกปลอมที่ใช้ในการฝึกสอนและทดสอบโมเดล	33
4.2	การตั้งค่าการทดลอง	33
4.3	การประเมินผลการทดลอง	35
4.4	ผลการทดลอง	38
<b>บทที่ 5</b>	<b>สรุปผลการวิจัยและข้อเสนอแนะ</b>	<b>49</b>
5.1	สรุปผลการวิจัย	49
5.2	ข้อเสนอแนะ	49
	เอกสารอ้างอิง	51
	ภาคผนวก	53
	ภาคผนวก ก ผลงานที่ได้รับการตีพิมพ์	54
	ภาคผนวก ข ข้อมูลวารสารที่ได้รับการตีพิมพ์	75
	ประวัติผู้เขียน	83

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญตาราง

ตารางที่	หน้า
ตารางที่ 1.1 สรุปงานวิจัยที่เกี่ยวข้อง	15
ตารางที่ 4.1 จำนวนภาพที่ใช้ในการทดลองแยกประเภทคำหยาบแต่ละประเภทในชุดข้อมูลบริเวณลำตัวคอ และขวิดหมุ่น	33
ตารางที่ 4.2 ผลการทดลองเปรียบเทียบของชุดข้อมูลบริเวณลำตัวของขวิดด้วยค่าความถูกต้อง ค่า Recall ความแม่นยำ และ F1	40
ตารางที่ 4.3 ผลการทดลองเปรียบเทียบของชุดข้อมูลบริเวณคอขวิดด้วยค่าความถูกต้อง ค่า Recall ความแม่นยำ และ F1	42
ตารางที่ 4.4 ผลการทดลองเปรียบเทียบประสิทธิภาพของชุดข้อมูลภาพขวิดหมุ่นด้วยค่าความถูกต้อง ค่า Recall ความแม่นยำ และ F1	44
ตารางที่ 4.5 เวลาในการฝึกสอนและเวลาในการตรวจจับหาสิ่งแปลกปลอมของชุดข้อมูลภาพบริเวณลำตัว ชุดข้อมูลคอขวิด และชุดข้อมูลภาพขวิดหมุ่น	46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญรูป

รูปที่	หน้า
รูปที่ 2.1 Convolutional Neural Network	6
รูปที่ 2.2 AlexNet	7
รูปที่ 2.3 VGG16	8
รูปที่ 2.4 ResNet50	8
รูปที่ 2.5 MobileNet	9
รูปที่ 2.6 CNNLSTM	9
รูปที่ 2.7 Autoencoder	10
รูปที่ 2.8 ตัวอย่างภาพการขยายข้อมูล	11
รูปที่ 3.1 เครื่องตรวจสอบ (ก) เครื่องตรวจสอบขวดแก้วแบบตรง (ข) เครื่องตรวจสอบขวดแก้วแบบ ขวดหมุน	17
รูปที่ 3.2 ภาพถ่ายผนังด้านข้างขวดบริเวณลำตัวและคอขวด	18
รูปที่ 3.3 ตัวอย่างสิ่งแปลกปลอม (ก) หิน (ข) ฉีกขาด (ค) ฟองอากาศ	18
รูปที่ 3.4 CNNLSTM	19
รูปที่ 3.5 ขั้นตอนการฝึกสอนโมเดล CNNLSTM	20
รูปที่ 3.6 ขั้นตอนวิธีการฝึกสอนเพื่อรู้จำคุณลักษณะของสิ่งแปลกปลอม	22
รูปที่ 3.7 ขั้นตอนวิธีการตรวจจับสิ่งแปลกปลอม	23
รูปที่ 3.8 วิธีการจำแนกประเภทจำแนกประเภทตามตัวอย่าง	24
รูปที่ 3.9 กราฟแสดงความซับซ้อนของวิธีการจำแนกประเภทจำแนกประเภทตามตัวอย่าง	25
รูปที่ 3.10 Autoencoder	26
รูปที่ 3.11 ภาพที่สร้างขึ้นใหม่ (ก) ภาพส่วนลำตัว (ข) ภาพส่วนคอ	27
รูปที่ 3.12 ภาพพื้นหลังปกติ	27
รูปที่ 3.13 กราฟแสดงความซับซ้อนของกระบวนการสร้างภาพพื้นหลังปกติและค่า Threshold ของ ข้อผิดพลาด	29
รูปที่ 3.14 วิธีการตรวจจับสิ่งแปลกปลอมด้วยภาพพื้นหลังและค่า Threshold ของข้อผิดพลาด	29
รูปที่ 3.15 วิธีการตรวจจับสิ่งแปลกปลอม	30
รูปที่ 3.16 กราฟแสดงความซับซ้อนของขั้นตอนการตรวจจับสิ่งแปลกปลอมด้วยภาพพื้นหลังปกติ และค่า Threshold ของข้อผิดพลาด	32
รูปที่ 4.1 ตัวอย่างผลการจำแนกประเภทของ DDAET (ก) พื้นที่ที่มีสิ่งแปลกปลอมคอขวด (ข) ผลการ ตรวจจับสิ่งแปลกปลอมคอขวด (ค) พื้นที่ที่มีสิ่งแปลกปลอมบริเวณลำตัว (ง) ผลการตรวจจับสิ่ง	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แปลงปลอมบริเวณลำตัวของขวด (จ) พื้นที่ที่มีสิ่งแปลงปลอมขวดหมุน (ฉ) ผลการตรวจจับสิ่งแปลงปลอมขวดหมุน	37
รูปที่ 4.2 ผลการทดลองเปรียบเทียบความถูกต้อง CNNLSTM ที่มีค่า K ต่างกัน	38
รูปที่ 4.3 Learning Curve ประสิทธิภาพการเรียนรู้ของ CNNLSTM	39



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

การตรวจจับสิ่งแปลกปลอมเป็นงานที่สำคัญถูกนำมาใช้ในหลาย ๆ ด้าน เช่น ด้านความปลอดภัย การผลิต ความปลอดภัยทางไซเบอร์ และด้านการเกษตร เป็นต้น วิธีการตรวจจับสิ่งแปลกปลอมแบบดั้งเดิมอาศัยความเชี่ยวชาญของมนุษย์และแบบจำลองทางสถิติ ซึ่งอาจไม่สามารถจับรูปแบบสิ่งแปลกปลอมที่ซับซ้อนในข้อมูลที่มีมิติสูงเช่นข้อมูลภาพได้ ในช่วงไม่กี่ปีที่ผ่านมา วิธีการตรวจจับสิ่งแปลกปลอมโดยใช้โครงข่ายประสาทเทียมเชิงลึกได้ประสิทธิภาพที่ดีในการสร้างเครื่องมือในการตรวจจับสิ่งแปลกปลอมโดยเฉพาะอย่างยิ่งในข้อมูลรูปภาพและวิดีโอ

ในอุตสาหกรรมการผลิตบรรจุภัณฑ์แก้วมีการตรวจจับสิ่งแปลกปลอมบนขวดแก้วก่อนการส่งออกไปยังลูกค้า การตรวจจับตำหนิบนขวดแก้วจึงมีความสำคัญอย่างมากในงานควบคุมคุณภาพในอุตสาหกรรม การตรวจจับตำหนิบนขวดแก้วช่วยในการรักษาคุณภาพผลิตภัณฑ์ที่มีความสมบูรณ์และปราศจากข้อบกพร่อง นั้นหมายความว่าผู้บริโภคจะได้รับสินค้าที่มีคุณภาพดีและปลอดภัยตามคาดหวัง ช่วยสร้างความเชื่อถือจากลูกค้า โดยทำให้ลูกค้ามั่นใจว่าสินค้าที่ได้รับจากบริษัทมีคุณภาพสูงและถูกต้องตามมาตรฐานที่กำหนด และช่วยให้บริษัทสามารถประกอบกิจการตามมาตรฐานและข้อกำหนดของอุตสาหกรรมได้อย่างถูกต้อง ซึ่งเป็นสิ่งสำคัญในการรักษาความเชื่อถือจากผู้อื่นในอุตสาหกรรมนี้ และยังช่วยลดการสูญเสียทรัพยากรที่เกิดจากการผลิตสินค้าที่มีข้อบกพร่อง โดยป้องกันการส่งออกสินค้าที่มีตำหนิหรือเสียหาย ซึ่งอาจเป็นการประหยัดเวลาและค่าใช้จ่ายในการผลิตในระยะยาว

วิธีการแบบดั้งเดิมของการตรวจสอบขวดแก้วมีการใช้เครื่องตรวจสอบคุณภาพบรรจุภัณฑ์แก้ว (Inspection Machine) และใช้มนุษย์ที่มีความเชี่ยวชาญในการตรวจจับสิ่งแปลกปลอมเพื่อใช้ในการตรวจสอบคุณภาพของขวดแก้วบนไลน์การผลิต Inspection Machine จะทำการตรวจจับสิ่งผิดปกติบนผนังด้านข้างขวด ก้น คอ และปากขวด โดยใช้ข้อมูลภาพขวดที่ทำการถ่ายจากกล้องอุตสาหกรรม สิ่งผิดปกติที่เกิดขึ้นบนขวดแก้วได้แก่ รอยแตก ฟองอากาศ ขวดสกปรก และหินดำ เป็นต้น วิธีการตรวจจับสิ่งแปลกปลอมจะนำข้อมูลภาพถ่ายของแก้วมาผ่านวิธีการประมวลผลภาพ (Image processing) เช่น Image threshold และ edge detection เป็นต้นเพื่อตรวจจับสิ่งแปลกปลอมที่อยู่บนของขวดแก้ว วิธีการ image processing สามารถตรวจจับได้อย่างรวดเร็วกว่าการใช้มนุษย์และช่วยเพิ่มประสิทธิภาพในการตรวจจับตำหนิและเพิ่มคุณภาพของขวดแก้วในขบวนการผลิตของอุตสาหกรรมการผลิตบรรจุภัณฑ์แก้วได้

ปัญหาของการตรวจจับสิ่งแปลกปลอมบนขวดแก้วโดยการตรวจสอบด้วยมนุษย์คือมาจากการตรวจสอบที่ไม่สามารถตรวจจับสิ่งแปลกปลอมได้อย่างครบถ้วนทุกด้านของขวดแก้ว และเกิดจาก

การเหนื่อยล้าในการทำงาน ทำให้เกิดมีสิ่งแปลกปลอมหลุดรอดไปจากขั้นตอนการตรวจสอบคุณภาพ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และเกิดขึ้นอยู่บ่อยครั้ง การตรวจจับสิ่งแปลกปลอมโดย inspection machine แบบดั้งเดิมไม่สามารถตรวจจับสิ่งแปลกปลอมของขวดแก้วได้อย่างถูกต้องแม่นยำเท่าที่ควร มีความละเอียดในการตรวจสอบน้อย ตรวจสอบได้ดีเฉพาะตำหนิที่มีขนาดใหญ่และมีสีเข้มเด่นชัดแตกต่างจากผนังขวดอย่างชัดเจน ทำให้เกิดปัญหาขวดมีตำหนิส่งผ่านไปถึงลูกค้าเกิดผลเสียเป็นอย่างมากในด้านการประกันคุณภาพและความน่าเชื่อถือของกระบวนการผลิตขวดแก้ว

วิทยานิพนธ์นี้เสนอวิธีการตรวจจับสิ่งแปลกปลอมสำหรับการตรวจจับสิ่งแปลกปลอมบนขวดแก้วสองวิธีการประกอบด้วย วิธีการที่หนึ่งการตรวจจับสิ่งแปลกปลอมแบบผสมโดยใช้โครงข่ายประสาทเชิงลึกที่เรียกว่า Convolutional Neural Networks (CNNs) เพื่อใช้ในการสกัดคุณลักษณะเชิงลึกจากข้อมูลภาพถ่าย และวิธีการจำแนกประเภทตามตัวอย่าง วิธีการที่สองในการตรวจจับสิ่งแปลกปลอมโดยใช้โครงข่ายประสาทเชิงลึกที่เรียกว่า autoencoder และวิธีการตรวจจับสิ่งแปลกปลอมด้วยค่าความผิดปกติ (Outlier detection) ถูกพัฒนาขึ้นเพื่อแก้ไขปัญหาเวลาที่ใช้ในการฝึกสอนที่ยาวนานและการใช้ทรัพยากรของเครื่องในวิธีการแรก วิธีการนี้ออกแบบมาเพื่อให้สามารถทำงานได้อย่างรวดเร็วโดยใช้จำนวนตัวอย่างข้อมูลสำหรับการฝึกสอนโครงข่ายประสาทที่น้อยลงและใช้ภาพสำหรับฝึกสอนเฉพาะภาพขวดดีเพื่อให้ได้ผลลัพธ์ที่มีประสิทธิภาพ วิธีการตรวจจับสิ่งแปลกปลอมที่นำเสนอเหล่านี้มีการประเมินประสิทธิภาพด้วยชุดข้อมูลสิ่งแปลกปลอมบนขวดแก้วทั้งหมดสามชุดได้แก่ ชุดข้อมูลภาพบริเวณลำตัวขวด ชุดข้อมูลภาพคอขวด และชุดข้อมูลภาพขวดหมุน

## 1.2 วัตถุประสงค์ของงานวิจัย

- 1) เพื่อพัฒนาอัลกอริทึมสำหรับการตรวจจับสิ่งแปลกปลอมโดยใช้โครงข่ายประสาทเชิงลึกในการช่วยตรวจจับสิ่งแปลกปลอมบนขวดแก้ว
- 2) เพื่อเปรียบเทียบประสิทธิภาพของวิธีการที่เสนอกับวิธีการตรวจจับสิ่งแปลกปลอมที่ใช้โครงข่ายประสาทเชิงลึกแบบ Convolutional Neural Network Autoencoder และวิธีการทาง Image processing

### 1.3 ขอบเขตของงานวิจัย

- 1) ศึกษาวิธีการตรวจจับสิ่งแปลกปลอมแบบผสมโดยใช้โครงข่ายประสาทเชิงลึก และวิธีการจำแนกประเภทตามตัวอย่างเพื่อพัฒนาอัลกอริทึมสำหรับการตรวจจับสิ่งแปลกปลอมบนผนังขวดแก้ว
- 2) ศึกษาวิธีการตรวจจับสิ่งแปลกปลอมบนขวดแก้วโดยใช้ Autoencoder และ Threshold ของข้อผิดพลาด
- 3) ศึกษาการตรวจจับสิ่งแปลกปลอมจากภาพถ่ายเครื่องตรวจสอบแบบตรงและแบบขวดหมุนที่มีชุดข้อมูลภาพสำหรับฝึกสอนเพื่อสร้างโมเดล และชุดข้อมูลภาพทดสอบสำหรับทดสอบประสิทธิภาพของโมเดล ประกอบด้วยภาพที่มีตำหนิจำนวน 688 ภาพ และภาพไม่มีตำหนิจำนวน 573 ภาพ

### 1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ได้อัลกอริทึมที่สามารถตรวจจับสิ่งแปลกปลอมบนขวดแก้วได้อย่างถูกต้องแม่นยำ
- 2) ได้อัลกอริทึมการตรวจจับสิ่งแปลกปลอมบนขวดแก้วที่มีความรวดเร็วในกระบวนการฝึกสอน
- 3) ได้โมเดลในการตรวจจับสิ่งแปลกปลอมบนขวดแก้วที่มีโครงสร้างขนาดเล็กใช้ทรัพยากรของเครื่องน้อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในการดำเนินการวิจัยครั้งนี้ได้ทำการศึกษาค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้องเกี่ยวกับการตรวจจับสิ่งแปลกปลอมบนขวดแก้วเพื่อใช้เป็นแนวทางในการวิจัยดังนี้

#### 2.1 ทฤษฎีที่เกี่ยวข้อง

##### 2.1.1 การเรียนรู้เชิงลึก (Deep learning)

การเรียนรู้เชิงลึกเป็นส่วนย่อยในของการเรียนรู้ของเครื่อง (Machine learning) ซึ่งโดยพื้นฐานแล้วเป็นโครงข่ายประสาทเทียมที่มีเลเยอร์ตั้งแต่สามเลเยอร์ขึ้นไปเป็นโครงข่ายประสาทเชิงลึก โครงข่ายประสาทเหล่านี้พยายามจำลองพฤติกรรมของสมองมนุษย์ ซึ่งทำให้สามารถเรียนรู้ข้อมูลจำนวนมากได้

##### 2.1.2 Convolutional Neural Network

Convolutional Neural Networks (CNN) เป็นอัลกอริทึมการเรียนรู้โครงข่ายเชิงลึกที่นิยมนำมาใช้กับข้อมูลภาพ (Lecun, 1998) มักใช้สำหรับการจำแนกประเภทและงานการมองเห็นด้วยคอมพิวเตอร์ (computer vision) CNN ประกอบด้วยชั้นต่อไปนี้

ชั้นนำเข้า (Input Layer): เป็นชั้นแรกของ CNN ซึ่งเป็นชั้นที่รับข้อมูลเข้าสู่โมเดล โดยทั่วไปจะเป็นภาพหรือข้อมูลที่ต้องการให้เครือข่ายทำนายหรือจำแนก แต่ Input Layer สามารถรับข้อมูลที่มีรูปแบบอื่นๆ ได้ตามที่ต้องการ ในกรณีที่ Input Layer เป็นภาพ ส่วนใหญ่จะเป็นภาพที่มีความกว้าง (width) และความสูง (height) และสามารถมีช่องสี (channel) เพื่อแสดงสีของภาพได้ ตัวอย่างเช่น ภาพสีที่มีความกว้าง 256 พิกเซล ความสูง 256 พิกเซล และมีช่องสี RGB 3 ช่อง ซึ่งหมายความว่าข้อมูลสีแบบแยกส่วนในภาพ แต่ละพิกเซลของภาพจะถูกนำเข้าไปใน Input Layer แยกตามพิกเซลและช่องสี เพื่อให้เครือข่ายประสาทเทียมสามารถเรียนรู้คุณลักษณะและรูปร่างของภาพได้

ชั้นคอนโวลูชัน (Convolutional Layer): เป็นส่วนหนึ่งที่สำคัญใน CNN เป็นชั้นที่ใช้ในการคัดสรรคุณลักษณะ (feature extraction) จากภาพ โดยใช้กระบวนการคอนโวลูชัน (convolution) เพื่อสกัดคุณลักษณะที่สำคัญออกมาจากภาพต้นฉบับ หลังจากนั้นเกิดการลดขนาดของภาพ (downsampling) ทำให้การคำนวณที่เกิดขึ้นลดลง และช่วยลดปริมาณข้อมูลที่ต้องนำมาใช้ในการฝึกฝนของโมเดลเครือข่ายประสาทเทียม สำหรับ Convolutional Layer ใน CNN มักประกอบด้วยข้อมูล (feature map) ที่สกัดคุณลักษณะจากภาพต้นฉบับ และกรอง (filter) หรือเมทริกซ์คอนโวลูชัน (convolutional kernel) ที่ใช้ในกระบวนการคอนโวลูชัน กระบวนการคอนโวลูชันคือการทำดอทโปรดักของเมทริกซ์คอนโวลูชันกับข้อมูลในพื้นที่ในภาพต้นฉบับ ทำให้ได้ข้อมูลใหม่ที่เป็นผลลัพธ์จากกระบวนการนี้ ตัวอย่างเช่น กรณีของภาพ 2 มิติ เมทริกซ์คอนโวลูชันคือตัวกรองขนาดเล็กที่อยู่ในรูปของเมทริกซ์และถูกเลื่อนบนพื้นที่ข้อมูลของภาพ การใช้ Convolutional Layer ใน CNN ทำให้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

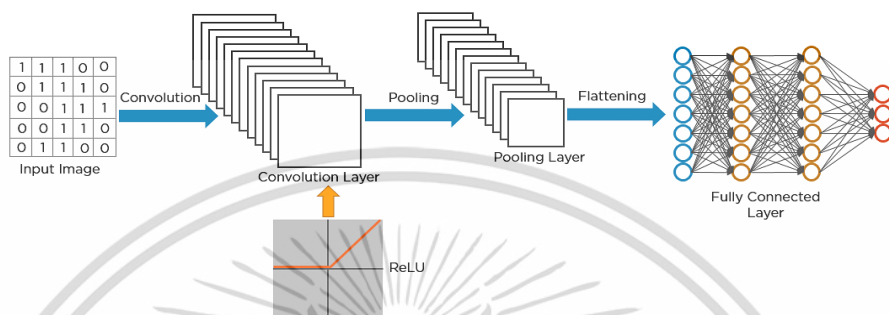
เครือข่ายสามารถตรวจจับลักษณะที่โดดเด่นในภาพได้อย่างมีประสิทธิภาพ และช่วยให้เครือข่ายสามารถจำแนกวัตถุในภาพได้ดีขึ้น

**ชั้นการกระตุ้น (Activation Layer):** เป็นชั้นที่สำคัญในเครือข่ายประสาทที่ใช้ในการเปลี่ยนค่าข้อมูลที่ผ่านมาจากชั้นก่อนหน้าเพื่อส่งต่อไปให้กับชั้นถัดไปของโมเดล ฟังก์ชันการกระตุ้น (Activation Function) เป็นฟังก์ชันที่ใช้ในการคำนวณค่าใหม่ของข้อมูล โดยมีหน้าที่กำหนดเงื่อนไขในการเปิดใช้งานหรือปิดใช้งานช่องของโครงข่าย ฟังก์ชันการกระตุ้นที่ใช้กันอย่างแพร่หลายคือ Rectified Linear Unit (ReLU) ซึ่งกำหนดค่าน้อยกว่าศูนย์เป็นศูนย์และค่าอื่นให้คงที่ซึ่งช่วยให้กรอบค่าที่มีความเปลี่ยนแปลงต่ำและลดปัญหาการคำนวณได้

**ชั้นพูลลิ่ง (Pooling Layer):** ชั้นที่ใช้ในการลดขนาดของภาพหรือข้อมูลที่ผ่านมาจากชั้นก่อนหน้าเพื่อลดข้อจำกัดต่าง ๆ และลดปริมาณข้อมูลที่ต้องนำมาใช้ในการฝึกฝนของโมเดลเครือข่ายประสาท ชั้นนี้เป็นอีกหนึ่งองค์ประกอบที่ช่วยเสริมสร้างความสามารถในการสกัดคุณลักษณะของโครงข่าย การทำงานของ Pooling Layer คือการลดขนาดของข้อมูลจากชั้นก่อนหน้ามาสร้างพื้นที่รวมข้อมูล (pooling region) แล้วทำการสกัดข้อมูลที่สำคัญจากพื้นที่รวมนั้น ซึ่งการสกัดข้อมูลนี้สามารถทำได้หลายรูปแบบ โดยที่ Max Pooling และ Average Pooling เป็นวิธีที่ใช้กันมากที่สุด  
**Max Pooling:** ในขั้นตอนนี้จะเลือกค่าสูงสุดในพื้นที่รวมข้อมูล เช่น หากเรากำหนดให้พื้นที่รวมข้อมูลมีขนาด 2x2 พิกเซล จะเลือกค่าที่มีค่าสูงสุดจากพื้นที่นี้เป็นค่าเดียวที่สำคัญ  
**Average Pooling:** ในขั้นตอนนี้จะคำนวณค่าเฉลี่ยของข้อมูลในพื้นที่รวมข้อมูล เช่น หากเรากำหนดให้พื้นที่รวมข้อมูลมีขนาด 2x2 พิกเซล จะคำนวณค่าเฉลี่ยของพื้นที่นี้เป็นค่าเดียวที่สำคัญ

**ชั้นเชื่อมต่อ (Fully Connected Layer):** เป็นชั้นที่เชื่อมต่อค่าคุณลักษณะที่สกัดมาจากชั้นก่อนหน้ากับชั้นส่งออก (Output Layer) เพื่อทำนายหรือจำแนกประเภทข้อมูล ชั้นนี้เป็นส่วนที่ทำให้โมเดลเครือข่ายประสาทเทียมเป็นโครงสร้างของ Multi-Layer Perceptron (MLP) ที่มีความสามารถในการจำแนกข้อมูลที่ซับซ้อนและคัดแยกคุณลักษณะของข้อมูล เมื่อข้อมูลผ่านชั้นคอนโวลูชัน (Convolutional Layer) หรือชั้นพูลลิ่ง (Pooling Layer) แล้ว จะได้คุณลักษณะที่สกัดมาจากข้อมูลนั้นเป็นข้อมูลแบบเมตริกซ์ (Matrix) หรือเวกเตอร์ (Vector) ใน Fully Connected Layer นี้จะนำคุณลักษณะที่ได้มาจากชั้นก่อนหน้าทั้งหมด และนำมาเชื่อมต่อกันเป็นเลเยอร์เดียว ทำให้ข้อมูลนี้เป็นลักษณะของเวกเตอร์ที่มีขนาดความยาวเป็นตัวแปร ซึ่งตัวแปรนี้จะเป็นข้อมูลที่เป็นภาพรวมของคุณลักษณะที่สกัดมาจากภาพหรือข้อมูลต้นฉบับ เมื่อมีการเชื่อมต่อกันของชั้นคอนโวลูชันหรือชั้นพูลลิ่งกับ Fully Connected Layer จะสามารถนำข้อมูลนี้ไปยังชั้นส่งออก (Output Layer) ที่เป็นชั้นที่ใช้ในการจำแนกหรือทำนายข้อมูล ในชั้นส่งออกมักจะใช้ฟังก์ชันการกระตุ้น (Activation Function) เพื่อจำแนกประเภทข้อมูล และให้ผลลัพธ์ที่เป็นความน่าจะเป็นในการจำแนกหรือทำนายข้อมูลของเครือข่าย

ชั้นส่งออก (Output Layer): เป็นชั้นที่ใช้ในการคำนวณผลลัพธ์หรือการทำนายข้อมูล ชั้นนี้มีบทบาทสำคัญในการตัดสินใจส่งออกคำตอบหรือผลลัพธ์จากเครือข่ายประสาทเทียม โดยมักจะใช้ฟังก์ชันการกระตุ้น (Activation Function) เพื่อทำนายความน่าจะเป็นในผลลัพธ์หรือส่งออกค่าข้อมูลที่เป็นผลลัพธ์ของโมเดล ในหนึ่งเครือข่ายประสาทเทียม อาจมีชั้นส่งออกมากกว่าหนึ่งชั้น ขึ้นอยู่กับความต้องการในการจำแนกประเภท ดังรูปที่ 2.1



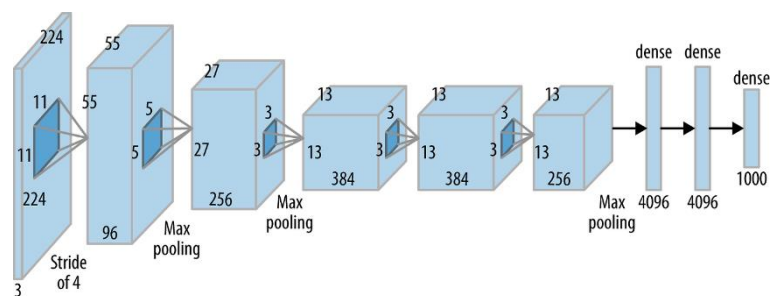
รูปที่ 2.1 Convolutional Neural Network

(ที่มา : <https://www.simplilearn.com/tutorials/deep-learning-tutorial/convolutional-neural-network>)

### 2.1.3 AlexNet

AlexNet เป็นชื่อของสถาปัตยกรรมโครงข่ายประสาทเทียม (CNN) ซึ่งออกแบบโดย Alex Krizhevsky ร่วมกับ Ilya Sutskever และ Geoffrey AlexNet เป็นเครือข่าย convolution แรกที่ใช้ GPU เพื่อเพิ่มประสิทธิภาพการทำงานของโครงข่ายประสาท (Krizhevsky, 2012)

สถาปัตยกรรมของ AlexNet ประกอบด้วย ชั้น convolution จำนวน 5 ชั้น ชั้น max-pooling จำนวน 3 ชั้น ชั้น Normalization จำนวน 2 ชั้น ชั้น fully connected จำนวน 2 ชั้น และชั้น Softmax จำนวน 1 ชั้น ชั้น Convolution แต่ละชั้นประกอบด้วย Convolutional Filter และ ReLU ฟังก์ชันกระตุ้นแบบไม่เชิงเส้น ชั้น pooling ใช้เป็นแบบ max-pooling ขนาดอินพุตส่วนใหญ่เป็น  $224 \times 224 \times 3$  แต่เนื่องจากมีช่องว่างภายในซึ่งเกิดขึ้น (padding) จึงกลายเป็น  $227 \times 227 \times 3$  AlexNet โดยรวมมี 60 ล้านพารามิเตอร์ ดังรูปที่ 2.2



รูปที่ 2.2 AlexNet

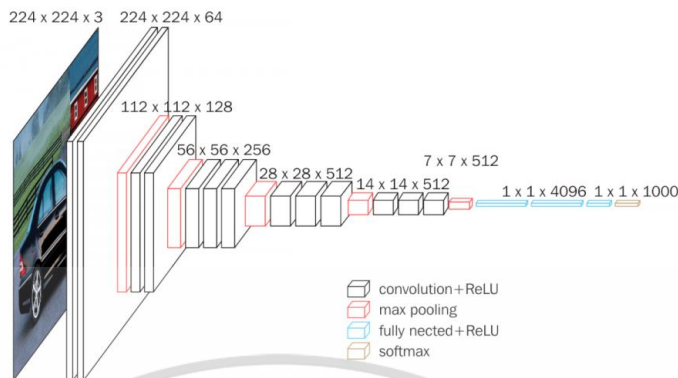
(ที่มา : <https://towardsdatascience.com/the-w3h-of-alexnet-vggnet-resnet-and-inception-7bbaaecccc96>)

### 2.1.4 VGG16

ชื่อเต็มของ VGG คือ Visual Geometry Group ซึ่งเป็นวิธีการของ Department of Science and Engineering of Oxford University โดย (Simonyan and Zisserman, 2014) ได้เปิดตัวชุดโมเดลโครงข่ายแบบหมุนวนที่ขึ้นต้นด้วย VGG ซึ่งสามารถนำไปใช้กับการจดจำใบหน้าและการจำแนกรูปภาพ ตั้งแต่ VGG16 ถึง VGG19 จุดประสงค์ดั้งเดิมของการวิจัยของ VGG เกี่ยวกับความลึกของโครงข่ายคอนโวลูชันคือเพื่อทำความเข้าใจว่าความลึกของโครงข่ายคอนโวลูชันมีผลต่อความแม่นยำในการจำแนกประเภทและการจดจำภาพขนาดใหญ่อย่างไร เพื่อเพิ่มจำนวนเลเยอร์เครือข่ายให้ลึกขึ้นและเพื่อหลีกเลี่ยงพารามิเตอร์มากเกินไป จึงใช้คอร์เนลขนาดเล็ก 3x3 ในทุกเลเยอร์

โครงสร้างของ VGG16 มีดังนี้ Input เป็นภาพ RGB 244 x 244 x 3 ซึ่งส่งต่อไปที่เลเยอร์ Convolutional ชั้นถัดไปเรื่อย ๆ ทำการส่งต่อจำนวน 2 ถึง 3 ครั้งจะได้ 64 ถึง 512 Feature Map และ Filter ที่มีขนาด 3x3 และใช้วิธีการขยับ (Stride) ทีละ 1 ครั้ง ชั้นตอน Convolutional จะทำการสกัดคุณลักษณะด้วย Max Pooling ขนาด Filter 3 x 3 และใช้วิธีการขยับ (Stride) ทีละ 2 ครั้ง ซึ่งจะทำให้ภาพมีขนาดที่เล็กลง เล็กลงเรื่อยๆ ในครั้งสุดท้ายจะเหลือเพียง Feature Map จากนั้นภาพที่เหลือแผ่ออกเป็นแนวตั้ง (Flatter Layer) ได้จำนวน 25,088 และจะส่งไปชั้น Fully Connected ดังรูปที่ 2.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



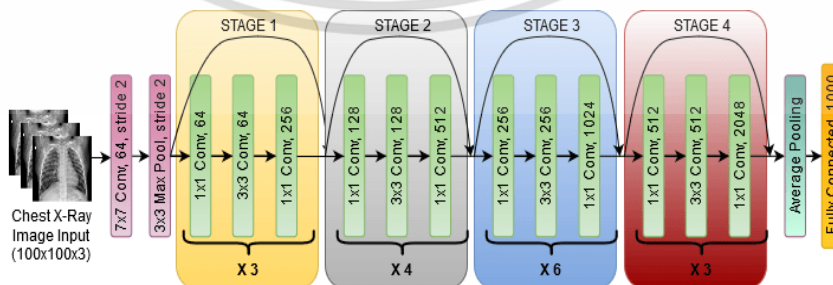
รูปที่ 2.3 VGG16

(ที่มา: [www.kaggle.com/blurredmachine/vggnet-16-architecture-a-complete-guide](http://www.kaggle.com/blurredmachine/vggnet-16-architecture-a-complete-guide))

2.1.5 ResNet50

ResNet ย่อมาจาก Deep Residual Network ได้มีการนำเสนอโครงข่ายประสาทนี้ในงานวิจัย Deep residual learning for image recognition มีวัตถุประสงค์ในการแก้ปัญหา vanishing gradient ของโครงข่ายประสาทแบบเก่าที่เกิดขึ้นกับโครงข่ายที่มีความลึกค่อนข้างมาก (He, 2016)

โครงสร้างของ ResNet50 มีจำนวนชั้นของของโครงข่ายประสาทจำนวน 152 ชั้น โดยใช้เทคนิคการออกแบบ module ที่มีลักษณะทางลัดลงใน network โครงข่าย ResNet50 นี้ประกอบด้วย 4 block โดยมีจำนวนพารามิเตอร์ที่ใช้สำหรับฝึกสอนโครงข่ายประสาททั้งหมดตามชั้นที่เราใช้เรียกชื่อ คือ ResNet50 จะหมายถึงจำนวน 50 เลเยอร์ ซึ่งจะอธิบายขนาดว่า [3, 4, 6, 3] ซึ่งคือ  $(3 + 4 + 6+3) \times 3 = 48$  ชั้น + 2 ชั้น = 50 ซึ่ง ResNet50 ดังรูปที่ 2.4



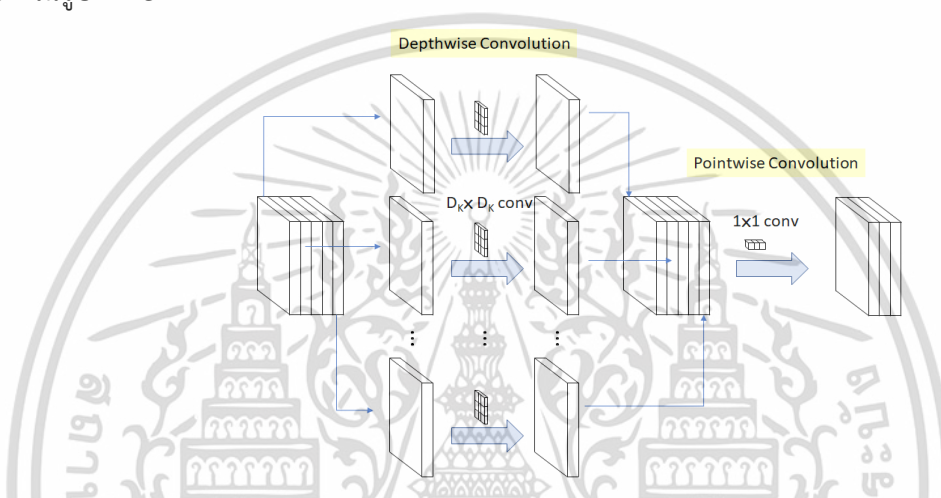
รูปที่ 2.4 ResNet50

(ที่มา: <https://open-instruction.com/dl-algorithms/overview-of-residual-neural-network-resnet>)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 2.1.6 MobileNet

MobileNet เป็นโครงข่ายประสาทเทียมแบบคอนโวลูชันรูปแบบหนึ่งที่ถูกพัฒนาขึ้นโดย Andrew G. Howard และในปีพ.ศ.2560 โดยรูปแบบของ MobileNet ถูกพัฒนาให้เป็นโมเดลโครงข่ายประสาทเชิงลึกที่มีขนาดเล็ก ที่ทำงานได้อย่างรวดเร็ว Latency ต่ำ ใช้ทรัพยากรของเครื่องในการประมวลผลไม่มาก ถูกออกแบบมาสำหรับงานที่มีทรัพยากรจำกัด MobileNet สามารถใช้งานได้ทั้ง Classification Detection Embedding และ Segmentation เหมือนกับโมเดลที่เป็นที่นิยมอื่น ๆ (Sandler et al., 2018 ; Howard et al., 2019) สถาปัตยกรรมของเครือข่าย MobileNet แสดงในรูปที่ 2.5

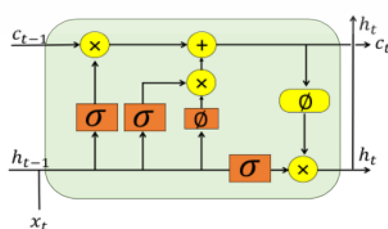


รูปที่ 2.5 MobileNet

(ที่มา: <https://towardsdatascience.com/review-mobilenetv1-depthwise-separable-convolution-light-weight-model-a382df364b69>)

### 2.1.7 Long short-term memory (LSTM)

Long short-term memory (LSTM) เป็นโครงข่ายประเภท Long short-term memory (LSTM) เป็นโครงข่ายประเภท Recurrent Neural Network อัลกอริทึม LSTM ได้รับการออกแบบมาเพื่อหลีกเลี่ยงปัญหาการพังทลายในระยะยาว (Hochreiter and Schmidhuber, 1997) รายละเอียดหน่วย LSTM แสดงในรูปที่ 2.6



รูปที่ 2.6 CNNLSTM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

LSTM ประกอบด้วยหน่วยความจำหลายหน่วย เซลล์ แต่ละเซลล์หน่วยความจำมีสามองค์ประกอบ: เขียน (write) อ่าน (read) และลืม (forget) ที่ขึ้นตอนเวลา  $t$  หน่วย forget gate ได้รับการปรับปรุงดังนี้

$$f_t = \sigma_g(W_f \mathbf{x}_t + U_f h_t + b_f), \quad (2.1)$$

$$i_t = \sigma_g(W_i \mathbf{x}_t + U_i h_t + b_i), \quad (2.2)$$

$$o_t = \sigma_g(W_o \mathbf{x}_t + U_o h_t + b_o), \quad (2.3)$$

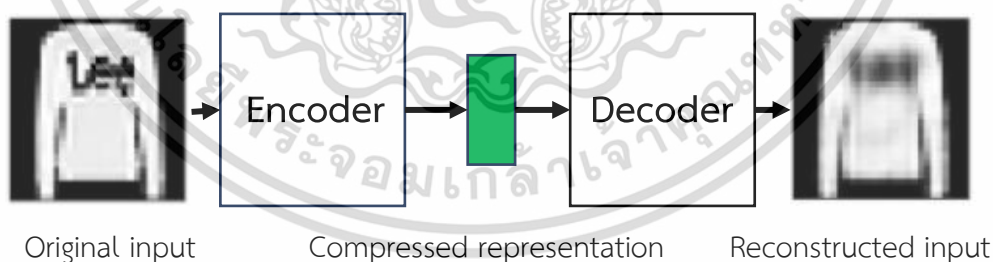
$$c_t = f_t \odot c_{t-1} + i_t \odot \sigma_c(W_c \mathbf{x}_t + b_c), \quad (2.4)$$

$$h_t = o_t \odot \sigma_h(c_t), \quad (2.5)$$

โดยที่  $\mathbf{x}_t$  เป็นเวกเตอร์ข้อมูลอินพุต  $f_t$  เป็น forget gate  $i_t$  เป็นเกตอินพุต/อัปเดต  $o_t$  เป็นประตูทางออก  $c_t$  เป็นเวกเตอร์สถานะของเซลล์ ตัวดำเนินการ  $\sigma$  เป็นฟังก์ชันการเปิดใช้งานซิกมอยด์  $\odot$  คือ Hadamard product (element-wise product) และ  $h_t$  เป็นเวกเตอร์เอาต์พุตของหน่วย LSTM

### 2.1.8 โครงข่ายประสาทแบบ Autoencoder

โครงข่ายประสาทแบบ Autoencoder เป็นแบบจำลองโครงข่ายประสาทที่มีวัตถุประสงค์หลักในการเรียนรู้การเข้ารหัส (encoding) และการถอดรหัส (decoding) ข้อมูลภาพเพื่อใช้ในการเรียนรู้ลักษณะหรือโครงสร้างของข้อมูล โดยส่วน encoder ทำหน้าที่แปลงข้อมูลเข้าเป็นคุณลักษณะที่เป็นตัวแทนข้อมูล (representation) ที่มีมิติต่ำลง ซึ่งส่วนนี้จะเป็นการลดขนาดข้อมูลนำเข้าให้ลดลง โดยส่วน decoder จะถอดรหัสคุณลักษณะที่เป็นตัวแทนข้อมูลเพื่อให้ข้อมูลกลับมาใกล้เคียงกับข้อมูลเข้าเดิม ตัวอย่าง autoencoder แสดงในรูปที่ 2.7



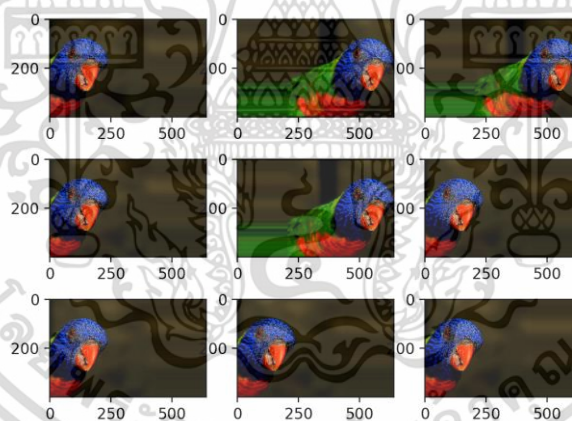
รูปที่ 2.7 Autoencoder

ในกระบวนการเรียนรู้ของ Autoencoder โครงข่ายประสาททำการเรียนรู้และถอดรหัสข้อมูลตัวอย่างใหม่มีความคล้ายคลึงกับข้อมูลเข้าเดิมให้มากที่สุด ซึ่งจะเกิดขึ้นได้โดยที่โครงข่ายมีการเปรียบเทียบคู่ระหว่างข้อมูลเข้าและข้อมูลออกเพื่อคำนวณค่าความสูญเสีย (loss) และใช้วิธีการปรับแต่งพารามิเตอร์ในโครงข่ายประสาทเพื่อลดค่าความสูญเสียให้น้อยที่สุด ในกระบวนการเรียนรู้นี้ เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Autoencoder สามารถเรียนรู้รูปแบบหรือลักษณะสำคัญของข้อมูลโดยอัตโนมัติโดยไม่ต้องการคลาส คำตอบของข้อมูลเป้าหมาย Autoencoder มีการประยุกต์ใช้ในหลายงาน เช่นในการลดมิติข้อมูล (dimensionality reduction) หรือสกัดลักษณะ (feature extraction) ข้อมูล เพื่อใช้ในงานแบบ จำแนก (classification) หรือจัดกลุ่ม (clustering) ข้อมูล โดย Autoencoder มีความสามารถในการ จับคู่ลักษณะที่สำคัญและจัดระเบียบข้อมูลเข้าในเนื้อที่สกัดลักษณะมากขึ้น นอกจากนี้ Autoencoder ยังเป็นพื้นฐานสำคัญในการสร้างโมเดลประสาทที่ซับซ้อนมากขึ้น เช่น Variational Autoencoder (VAE) และ Generative Adversarial Networks (GAN) ซึ่งมีการประยุกต์ใช้ในการสร้างข้อมูลที่ เหมือนจริง

### 2.1.9 การขยายข้อมูล (Data augmentation)

การขยายข้อมูล (Data augmentation) หรือ การสร้างข้อมูลภาพใหม่ให้มีจำนวนเพิ่มขึ้น โดยทำการดัดแปลงภาพเดิมที่มีด้วยวิธีการต่างๆ เช่น บิด ตัด หมุน เปลี่ยนสีทำภาพให้มืดหรือสว่าง ขึ้น หรือใส่สัญญาณรบกวน (Noise) ลงไปในภาพ จะทำให้ได้รูปภาพแบบต่าง ๆ ไม่จำกัด (Shorten and Khoshgoftaar, 2019) ตัวอย่างภาพการขยายข้อมูลแสดงในรูปที่ 2.8



รูปที่ 2.8 ตัวอย่างภาพการขยายข้อมูล

(ที่มา: <https://machinelearningmastery.com/how-to-configure-image-data-augmentation-when-training-deep-learning-neural-networks>)

วิธีการขยายข้อมูลเกิดขึ้นมาเนื่องจากประสิทธิภาพความแม่นยำของโมเดลโครงข่ายประสาท Deep Learning นั้นขึ้นกับปริมาณข้อมูลฝึกสอนเป็นปัจจัยสำคัญ การเพิ่มข้อมูลได้ใช้เป็นเทคนิคที่สำคัญมากสำหรับ Machine Learning โดยเฉพาะคอมพิวเตอร์วิทัศน์ (Computer Vision) การทำ Augmentation นั้นภาพที่สร้างขึ้นใหม่จะต้องไม่เหมือนข้อมูลภาพเดิมจากต้นฉบับมากเกินไป เพราะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โมเดล deep learning นั้นอาจจะไปเรียนรู้สิ่งที่ไม่สำคัญบนภาพและอาจจะจำสิ่งที่ไม่จำเป็นไป ซึ่งจะทำให้การดำเนินงานนั้นมีความแม่นยำน้อยลง

## 2.2 งานวิจัยที่เกี่ยวข้อง

Ma (2002) มีการนำเสนอระบบตรวจจับตำหนิประกอบด้วยกล้องแปดตัวทำการประมวลผลบนคอมพิวเตอร์สองเครื่อง และชุดควบคุมหนึ่งชุด กล้องวิดีโอติดตั้งอยู่ข้างสายการผลิตและจับภาพขวดขณะเคลื่อนผ่านกล้องบนสายการผลิต จากนั้นภาพถ่ายขวดจะถูกส่งไปยังคอมพิวเตอร์สองเครื่อง ซึ่งจะถูกรประมวลผลโดยอัลกอริทึมการมองเห็นของคอมพิวเตอร์ (Computer vision) อัลกอริทึมการมองเห็นของคอมพิวเตอร์จะระบุข้อบกพร่องในภาพและส่งสัญญาณไปยังชุดควบคุม จากนั้นหน่วยควบคุมจะนำขวดที่ตรวจพบตำหนิออกจากสายการผลิต ระบบได้รับการทดสอบกับขวดแก้วหลายประเภทที่มีข้อบกพร่องประเภทต่างๆ ผลปรากฏว่าระบบสามารถตรวจจับข้อบกพร่องได้แม่นยำกว่า 90% ระบบยังสามารถตรวจจับข้อบกพร่องขนาดต่างๆ รวมถึงข้อบกพร่องขนาดเล็กที่มองไม่เห็นด้วยตาเปล่า ระบบที่อธิบายในบทความนี้เป็น การปรับปรุงที่สำคัญกว่าวิธีการตรวจจับข้อบกพร่องในขวดแก้วแบบเดิมๆ ระบบไม่สัมผัส ซึ่งหมายความว่าไม่ทำให้ขวดเสียหาย ระบบยังมีความแม่นยำมาก ระบบนี้มีศักยภาพในการปรับปรุงคุณภาพของขวดแก้วและเพื่อลดต้นทุนการผลิต ระบบนี้ยังใช้เพื่อปรับปรุงความปลอดภัยของขวดแก้วได้ด้วยการตรวจจับข้อบกพร่องที่อาจทำให้ขวดแตกได้

Li (2017) เสนอวิธีการตรวจจับตำหนิบนขวดแก้วโดยใช้ Machine vision วิธีแรกจะประมวลผลภาพของขวดเพื่อขจัดสัญญาณรบกวน จากนั้นจึงใช้อัลกอริทึม thresholding เพื่อแบ่งภาพออกเป็นสองส่วน ได้แก่ ขวดและพื้นหลัง จากนั้นระบุปากขวดโดยการวิเคราะห์พื้นที่รูรับแสงและความกว้างของพื้นที่เชื่อมต่อของปากขวด ขั้นตอนสุดท้าย การจับคู่ของพิกเซลโดเมนที่เชื่อมต่อจะใช้เพื่อตัดสินว่าปากขวดมีคุณสมบัติเหมาะสมหรือไม่ และตรวจจับตำแหน่งของตำหนิที่ตรวจพบ ผลการทดลองการตรวจจับขวดแสดงให้เห็นว่าวิธีนี้สามารถระบุตำแหน่งของตำหนิได้อย่างแม่นยำว่าปากขวดมีข้อบกพร่องอยู่บริเวณไหน และตรวจจับตำหนิของขวดได้ ความแม่นยำสูงในการตรวจจับสิ่งผิดปกติบนขวดแก้ว

Fu (2019) เสนอวิธีการที่ใช้ vision system สำหรับการตรวจจับหาตำหนิบนขวดแก้ว วิธีการนี้ ใช้การให้แสงพื้นหลัง การกรองค่ามัธยฐานของภาพ การปรับปรุง และการตรวจจับขอบเพื่อวิเคราะห์ลักษณะของโดเมนที่เชื่อมต่อของตำแหน่งของตำหนิ วิธีการนี้สามารถตรวจจับข้อบกพร่องต่างๆ เช่น รอยแตก ขอบขวด ขวดสกปรก จุดต่างดำ และอื่นๆ วิธีนี้ใช้งานง่ายและใช้กันอย่างแพร่หลาย

Zhou (2020) เสนอกรอบการตรวจจับตำหนิบนพื้นผิวสำหรับการตรวจสอบกันขวด รวมถึงวิธีการโลคัลไลเซชัน ROI ที่แม่นยำโดยใช้การตรวจจับวงกลมซูเปอร์พิกเซลอัตราเอนโทรปี (ERSCD)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และการแบ่งพื้นที่การวัดออกเป็นสองส่วน ในส่วนแรกได้รับการวิเคราะห์ด้วยวิธีการแบ่งส่วนซูเปอร์พิกเซลแบบกระจายแบบแอนไอโซโทรปิกแบบกระจายความถี่ (FTADSP) ในขณะที่ส่วนสองได้รับการวิเคราะห์ด้วยการกรองหลายระดับการแปลงเวฟเล็ต (WTMF) เพื่อลดอิทธิพลของพื้นผิวและปรับปรุงความทนทาน วิธีการ WTMF บรรลุผลลัพธ์ที่คาดหวังสำหรับการตรวจจับข้อบกพร่องที่กันขวดอย่างมีประสิทธิภาพ

Zhou (2020) ได้เสนอกรอบการทำงานสำหรับการตรวจสอบด้วยภาพของกันขวดแก้วอัตโนมัติโดยใช้การตรวจจับความเด่น (saliency detection) และการจับคู่เทมเพลต (template matching) กรอบการทำงานที่เสนอประกอบด้วยสามขั้นตอนหลัก: 1) การได้มาของภาพ: ขั้นตอนแรกนี้คือการได้ภาพของกันขวด รับภาพโดยใช้กล้องที่ติดตั้งอยู่เหนือสายพานลำเลียงขวด 2) การตรวจจับส่วนเด่น: ขั้นตอนที่สองนี้คือการตรวจจับส่วนเด่นในภาพ พื้นที่เด่นคือพื้นที่ของภาพที่ดึงดูดสายตาหรือมีความสำคัญ การตรวจจับความสนใจใช้เพื่อระบุบริเวณของภาพที่น่าจะมีข้อบกพร่อง 3) การจับคู่เทมเพลต: ขั้นตอนที่สามคือการจับคู่เทมเพลตกับพื้นที่ที่โดดเด่น เทมเพลตเป็นภาพขนาดเล็กที่แสดงถึงข้อบกพร่อง การจับคู่เทมเพลตใช้เพื่อยืนยันการมีอยู่ของข้อบกพร่องในพื้นที่สำคัญ วิธีการนี้สามารถตรวจสอบข้อบกพร่องของกันขวดแก้วได้อย่างมีประสิทธิภาพ กรอบการทำงานมีความแม่นยำและมีประสิทธิภาพ และสามารถปรับปรุงคุณภาพของขวดแก้วได้

Lu (2021) เสนอวิธีการสำหรับตรวจจับข้อบกพร่องบริเวณปากขวดแก้ว วิธีการนี้ใช้วิธีการแบ่งส่วนพื้นที่เพื่อระบุปากขวด จากนั้นใช้ชุดเทคนิคการประมวลผลภาพเพื่อตรวจจับข้อบกพร่อง วิธีการทำงานโดยแบ่งภาพขวดแก้วออกเป็นสองส่วนก่อน คือ ส่วนปากขวดและส่วนอื่นๆ ของขวด จากนั้นจึงระบุปากขวดเป็นพื้นที่ที่มีพื้นที่สูงสุด เมื่อระบุปากขวดได้แล้ว ชุดของเทคนิคการประมวลผลภาพจะถูกนำไปใช้เพื่อตรวจจับข้อบกพร่อง เทคนิคการประมวลผลภาพประกอบด้วย การตรวจจับขอบ (edge detection) เทคนิคนี้ใช้เพื่อระบุขอบปากขวด Thresholding เทคนิคนี้ใช้เพื่อแปลงภาพเป็นภาพไบนารี โดยที่พิกเซลเป็นสีดำหรือขาว Morphology เทคนิคนี้ใช้เพื่อระบุและลบสัญญาณรบกวนออกจากภาพ เมื่อมีการระบุข้อบกพร่องแล้ว จะจำแนกข้อบกพร่องเหล่านั้นออกเป็น 1 ใน 3 ประเภท ได้แก่ รอยร้าว รอยบิ่น และรอยขีดข่วน วิธีการนี้สามารถตรวจสอบข้อบกพร่องได้ด้วยความแม่นยำสูง สามารถตรวจจับข้อบกพร่องได้ด้วยความแม่นยำถึง 98% วิธีการนี้ยังทนทานต่อสัญญาณรบกวนและการเปลี่ยนแปลงของสภาพแสงอีกด้วย

He (2020) ได้เสนอวิธีการใหม่ในการตรวจจับข้อบกพร่องในผลิตภัณฑ์อุตสาหกรรม วิธีนี้ใช้โครงข่ายประสาทเทียมแบบถดถอยเชิงลึก (deep regression neural network) เพื่อทำนายความรุนแรงของข้อบกพร่องที่แต่ละพิกเซลในภาพ วิธีการทำงานโดยแบ่งภาพออกเป็นตารางพิกเซลก่อน จากนั้นโครงข่ายประสาทเทียมแบบถดถอยเชิงลึกจะถูกใช้เพื่อทำนายความรุนแรงของข้อบกพร่องในแต่ละพิกเซล ความรุนแรงของข้อบกพร่องจะแสดงด้วยตัวเลขระหว่าง 0 ถึง 1 โดยที่ 0 หมายถึงไม่มีข้อบกพร่อง และ 1 หมายถึงข้อบกพร่องร้ายแรง ความรุนแรงของข้อบกพร่อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ที่คาดการณ์ไว้จะถูกนำมาใช้เพื่อระบุตำแหน่งและความรุนแรงของข้อบกพร่องในภาพ วิธีการนี้สามารถตรวจสอบข้อบกพร่องได้ด้วยความแม่นยำสูง ในการทดลองที่ดำเนินการโดยนักวิจัย วิธีการดังกล่าวสามารถตรวจจับข้อบกพร่องได้ด้วยความแม่นยำถึง 98% วิธีการนี้ยังทนทานต่อสัญญาณรบกวนและการเปลี่ยนแปลงของสภาพแสงอีกด้วย

Liang (2022) เสนอวิธีการข้อบกพร่องโดยใช้โครงข่ายประสาทเทียมเชิงลึกแบบใหม่สำหรับการตรวจจับข้อบกพร่องของวัตถุขนาดเล็กในกระเบื้องแม่เหล็ก กรอบงานที่เสนอเรียกว่า ELCNN ประกอบด้วยสามส่วน: โมดูลสกัดคุณลักษณะ (feature extraction) โมดูลปรับปรุงคุณลักษณะระดับต้น (feature enhancement) และโมดูลฟิวชันคุณลักษณะหลายสเกล (multi-scale feature fusion) โมดูลสกัดคุณลักษณะจะสกัดคุณลักษณะจากภาพอินพุต โมดูลปรับปรุงคุณลักษณะแบบต้นจะปรับปรุงคุณลักษณะที่สกัดออกมา และโมดูลฟิวชันคุณลักษณะหลายสเกลจะรวมคุณลักษณะขั้นสูงเข้ากับคุณลักษณะจากโมดูลการสกัดคุณลักษณะ วิธีการนี้ได้รับการประเมินบนชุดข้อมูลขนาดใหญ่ชุดใหม่ที่เรียกว่า MT-SOD และผลที่ได้แสดงให้เห็นว่ามีประสิทธิภาพและมีความแม่นยำ

Koodtalang (2019) เสนอวิธีการในการตรวจสอบกันขวดแก้วเพื่อหาจุดบกพร่อง วิธีการนี้ใช้การผสมผสานระหว่างการประมวลผลภาพ (image processing) และเทคนิคการเรียนรู้เชิงลึก (deep learning) เพื่อระบุข้อบกพร่องที่กันขวด วิธีการทำงานโดยจับภาพกันขวดก่อน ภาพจะถูกประมวลผลโดยใช้เทคนิคการประมวลผลภาพที่หลากหลายเพื่อขจัดสัญญาณรบกวนและเพิ่มความคมชัด ภาพที่ประมวลผลแล้วจะถูกส่งเข้าสู่โมเดลการเรียนรู้เชิงลึกที่ได้รับการฝึกอบรมเพื่อระบุข้อบกพร่องในกันขวดแก้ว โมเดลการเรียนรู้เชิงลึกแสดงความเป็นไปได้ที่กันขวดมีข้อบกพร่อง วิธีการนี้ประเมินจากชุดข้อมูล 1,000 ภาพของกันขวดแก้ว วิธีการนี้สามารถระบุข้อบกพร่องด้วยความแม่นยำ 98% และสามารถระบุข้อบกพร่องประเภทต่างๆ ได้ เช่น รอยแตก ชิป และรอยขีดข่วน

ตารางที่ 2.1 สรุปรายละเอียดงานวิจัยที่เกี่ยวข้องในการตรวจจับสิ่งแปลกปลอม ตารางแสดงรายละเอียดของแต่ละงานวิจัยประกอบด้วย ชื่องานวิจัย ปีที่ตีพิมพ์ ข้อมูลสิ่งแปลกปลอมที่นำมาใช้ในการทดลอง และสรุปผลการทดลองของอัลกอริทึมการตรวจจับสิ่งแปลกปลอม

## ตารางที่ 2.1 สรุปงานวิจัยที่เกี่ยวข้อง

ลำดับ	ชื่องานวิจัย	ปีที่ตีพิมพ์	อัลกอริทึมที่ใช้	ชุดข้อมูล	ผลการทดลอง
1	A glass bottle defect detection system without touching	2002	Image Thresholding	ภาพถ่ายบริเวณปากและขอบปากขวด	90%
2	The method for glass bottle defects detecting based on machine vision	2017	Image Thresholding	ภาพถ่ายบริเวณคอขวด	99.3%
3	Medicine Glass Bottle Defect Detection Based on Machine Vision	2019	การกรองข้อมูลภาพ (Image filtering) การปรับปรุงภาพ (Image enhancement) และการหาขอบภาพ (Edge Detection)	ขวดแก้วสำหรับบรรจุยา	91.6%
4	A Surface Defect Detection Framework for Glass Bottle Bottom Using Visual Attention Model and Wavelet Transform	2020	วิธีการแบ่งส่วนซูเปอร์พิกเซลแบบกระจายแบบแอนไอโซโทรปิกแบบกระจายความถี่ (FTADSP) และการกรองหลายระดับการแปลงเวฟเลต (WTMF)	ภาพถ่ายบริเวณก้นขวด	93.24%
5	Automated Visual Inspection of Glass Bottle Bottom With Saliency Detection and Template Matching	2019	การตรวจจับความเด่น (saliency detection) และการจับคู่เทมเพลต (template matching)	ภาพถ่ายบริเวณก้นขวด	88.83%
6	A defect detection method for glass bottle mouth based on area segmentation	2021	การตรวจจับขอบ (edge detection) และ Image thresholding	ภาพถ่ายบริเวณปากขวด	98%
7	Deep Regression Neural Network for Industrial Surface Defect Detection	2020	Deep Regression Neural Network	ชุดข้อมูลภาพถ่ายตำหนิบริเวณพื้นผิว DAGM2007 dataset	98%
8	ELCNN: A Deep Neural Network for Small Object Defect Detection of Magnetic Tile	2022	Feature enhancement and loop-shaped fusion convolutional neural network (ELCNN)	ภาพถ่ายตำหนิบนแผ่นกระเบื้องแม่เหล็ก	97.61%
9	Glass Bottle Bottom Inspection Based on Image Processing and Deep Learning	2019	การประมวลผลภาพ (image processing) และเทคนิคการเรียนรู้เชิงลึก (deep learning)	ภาพถ่ายบริเวณก้นขวด	98%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

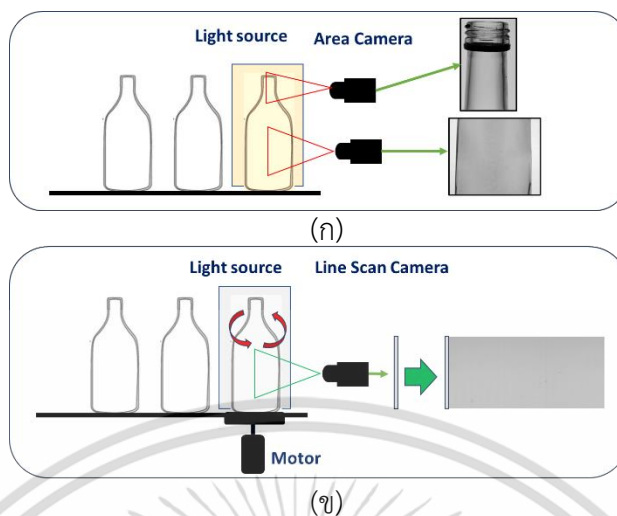
## บทที่ 3

### วิธีการดำเนินงานวิจัย

ในบทนี้อธิบายวิธีการสร้างชุดข้อมูลภาพสิ่งแปลกปลอมและนำเสนอวิธีการสำหรับตรวจจับสิ่งแปลกปลอมโดยใช้โครงข่ายประสาทเชิงลึกประกอบด้วยวิธีการตรวจจับสิ่งแปลกปลอมที่นำเสนอสองวิธีการคือ 1) วิธีการตรวจจับสิ่งแปลกปลอมแบบผสมโดยใช้ CNNLSTM และการจำแนกประเภทตามตัวอย่าง (instance-based) 2) วิธีการตรวจจับสิ่งแปลกปลอมโดยใช้โครงข่ายประสาทเชิงลึก Autoencoder และ Threshold error โดยมีรายละเอียดดังต่อไปนี้

#### 3.1 ชุดข้อมูลภาพสิ่งแปลกปลอม

ในกระบวนการผลิตขวดแก้วมีเครื่องตรวจสอบคุณภาพขวดแก้วอยู่สองประเภทหลักคือเครื่องตรวจสอบขวดแก้วแบบตรง (straight glass bottle inspection machine) และเครื่องตรวจสอบขวดแก้วแบบขวดหมุน (rotary glass bottle inspection machine) รูปที่ 3.1 แสดงตัวอย่างของเครื่องสำหรับตรวจสอบคุณภาพขวดแก้ว โดยเครื่องตรวจสอบขวดแก้วแบบตรงเป็นระบบอัตโนมัติที่ใช้เทคโนโลยีที่ทันสมัย เช่น วิธีการการคอมพิวเตอร์วิทัศน์ (Computer vision) และการประมวลผลภาพ (Image processing) เพื่อตรวจจับและระบุสิ่งแปลกปลอมในขวดแก้ว โดยจะบันทึกรูปภาพของขวดจากมุมต่างๆ หลายมุมรอบขวด และวิเคราะห์ภาพเหล่านั้นโดยทันทีหลังจากถ่ายภาพเสร็จ เครื่องตรวจสอบนี้ถูกนำมาใช้เพื่อทำให้มั่นใจในคุณภาพสินค้าและป้องกันปัญหาในระหว่างการผลิตและการขนส่ง ภาพของเครื่องตรวจสอบขวดแก้วแบบตรงแสดงในรูปที่ 3.1 (ก) ในขณะที่เครื่องตรวจสอบขวดแก้วแบบหมุนเป็นระบบอัตโนมัติที่ออกแบบมาเฉพาะสำหรับการตรวจสอบขวดแก้วที่มีทรงกลมโดยใช้กลไกการหมุนขวด โดยการหมุนขวดเป็นการตรวจสอบพื้นผิวทั้งหมดเพื่อให้ได้ระดับความแม่นยำสูงในการตรวจจับข้อผิดพลาด เนื่องจากขวดมีการหมุนทำให้ต้องใช้กล้องถ่ายภาพแบบ line scan โดยที่กล้อง line scan ใช้ในการบันทึกภาพของวัตถุหรือพื้นผิวในรูปแบบของเส้นพิกเซลเดียว กล้องสแกนแบบเส้นจะเคลื่อนที่ในแนวขนานกับวัตถุหรือพื้นผิวที่ต้องการบันทึกภาพ โดยส่วนของวัตถุที่เคลื่อนที่จะถูกสแกนบนเซนเซอร์ของกล้อง กล้องจะบันทึกภาพเป็นแถวเดียวโดยที่แต่ละพิกเซลจะเป็นข้อมูลจากตำแหน่งเวลาที่แตกต่างกัน ซึ่งจะช่วยให้ได้ภาพที่ยาวและแบ่งแยกส่วนของวัตถุหรือสิ่งของได้อย่างละเอียด ภาพจะถูกบันทึกเป็นค่าสีหรือค่าความเข้มของพิกเซลในแต่ละจุด แต่ละเส้นพิกเซลเดียวที่ถ่ายได้จะถูกนำรวมให้กลายเป็นภาพที่มีขนาดกว้างยาวตามที่กำหนด เครื่องตรวจสอบขวดแก้วแบบหมุนแสดงในรูปที่ 3.1 (ข)



รูปที่ 3.1 เครื่องตรวจสอบ (ก) เครื่องตรวจสอบขวดแก้วแบบตรง (ข) เครื่องตรวจสอบขวดแก้วแบบขวดหมุน

เทคนิคการใช้แสงในการถ่ายภาพของเครื่องตรวจสอบขวดจะใช้แฟลชโพลาไรซ์ใน light source สำหรับการถ่ายภาพขวดเป็นหนึ่งในเทคนิคที่นิยมใช้ในการสร้างแสงและเงาที่เหมาะสมเพื่อเพิ่มความชัดเจนและความโดดเด่นของภาพขวดในกระบวนการถ่ายภาพหรือการตรวจสอบ แฟลชโพลาไรซ์ถูกออกแบบมาเพื่อทำหน้าที่เป็นตัวกรองแสงที่ทำให้แสงที่ออกจากแหล่งแสงเป็นแสงที่มีความเนียนและแนวเดียวกัน การใช้แฟลชโพลาไรซ์ใน light source สามารถสร้างแสงที่มีความสม่ำเสมอในการถ่ายภาพขวดได้ ทำให้เพิ่มความชัดเจนและความโดดเด่นของภาพขวดในการตรวจสอบและปรับปรุงคุณภาพของผลิตภัณฑ์ เทคนิคนี้ใช้กับเครื่องตรวจสอบขวดทั้งสองประเภท

สิ่งแปลกปลอมของขวดแก้วส่วนใหญ่มักพบมากที่ส่วนผนังด้านข้างของขวด ซึ่งเป็นส่วนที่มีพื้นที่ของขวดมากที่สุดอยู่บริเวณนี้ เป้าหมายของชุดข้อมูลนี้คือการนำภาพบริเวณผนังด้านข้างของขวดมาทำการวิเคราะห์ตรวจสอบเพื่อตรวจจับสิ่งแปลกปลอมที่เกิดขึ้น ชุดข้อมูลนี้จะเน้นไปที่บริเวณคอขวด ลำตัวขวด และภาพขวดหมุน ซึ่งเป็นส่วนที่สำคัญที่จำเป็นต้องตรวจจับสิ่งแปลกปลอมในกระบวนการผลิต

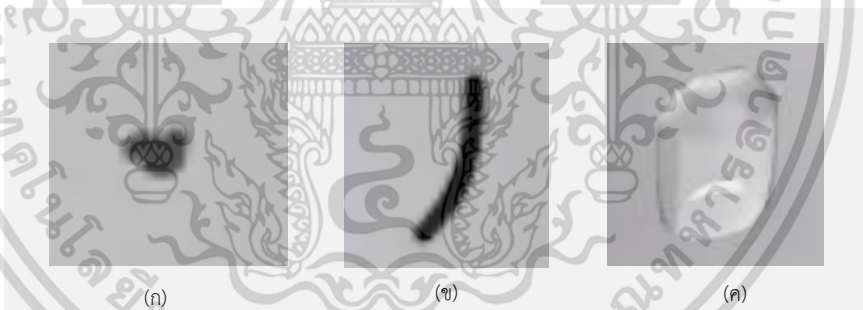
ในการเก็บรวบรวมภาพขวดของเครื่องตรวจสอบแบบตรงจะใช้กล้องสองตัวเพื่อถ่ายภาพขวด กล้องตัวแรกโฟกัสที่บริเวณคอและกล้องตัวที่สองโฟกัสที่บริเวณลำตัวของขวด โดยที่กล้องแต่ละตัวจะทำการถ่ายภาพเมื่อมีขวดไหลผ่านหน้ากล้อง โดยที่กล้องแต่ละตัวจะถ่ายภาพหนึ่งขวดต่อหนึ่งภาพตามสัญญาณสั่งถ่ายที่ได้รับมาจากระบบควบคุม

รูปที่ 3.2 แสดงภาพตัวอย่างภาพถ่ายผนังด้านข้างขวดบริเวณลำตัวและคอขวดด้วยกล้องถ่ายภาพประเภท area scan เมื่อขวดแก้วในสายการผลิตผ่านหน้ากล้องของเครื่องตรวจสอบ



รูปที่ 3.2 ภาพถ่ายผนังด้านข้างขวดบริเวณลำตัวและคอขวด

ส่วนการเก็บรวบรวมภาพขวดของเครื่องตรวจสอบแบบขวดหมุนจะโฟกัสภาพเน้นไปที่บริเวณลำตัวขวด เมื่อกล้องทำการถ่ายภาพขวดแล้วภาพที่ถ่ายได้จะถูกส่งไปยังโปรแกรมสำหรับตรวจจับสิ่งแปลกปลอม หากขวดมีสิ่งแปลกปลอมอยู่ที่ผนังด้านข้างขวด โปรแกรมตรวจจับจะส่งสัญญาณ reject ไปยังระบบควบคุมเพื่อกำจัดขวดที่มีตำหนินี้ออกไป สิ่งแปลกปลอมที่ตรวจพบตัวอย่างเช่น เศษหินที่ไม่สามารถหลอมได้ รอยฉีก รอยแตก และตุ่มฟองอากาศที่ผนังด้านข้างของขวด รูปที่ 3.3 แสดงภาพขวดที่พบข้อบกพร่องในชุดข้อมูลที่เสนอ รูปที่ 3.3 (ก) แสดงรอยตำหนิของหิน (ข) รอยฉีกขาด และ (ค) รอยตำหนิฟองอากาศที่ผนังด้านข้างขวด



รูปที่ 3.3 ตัวอย่างสิ่งแปลกปลอม (ก) หิน (ข) ฉีกขาด (ค) ฟองอากาศ

### 3.2 วิธีการตรวจจับสิ่งผิดปกติแบบผสมโดยใช้ CNNLSTM และการจำแนกตามตัวอย่าง

ส่วนนี้จะอธิบายถึงสถาปัตยกรรมของโครงข่ายประสาทเชิงลึกที่เรียกว่า CNNLSTM และอธิบายวิธีการตรวจจับสิ่งแปลกปลอมแบบผสม

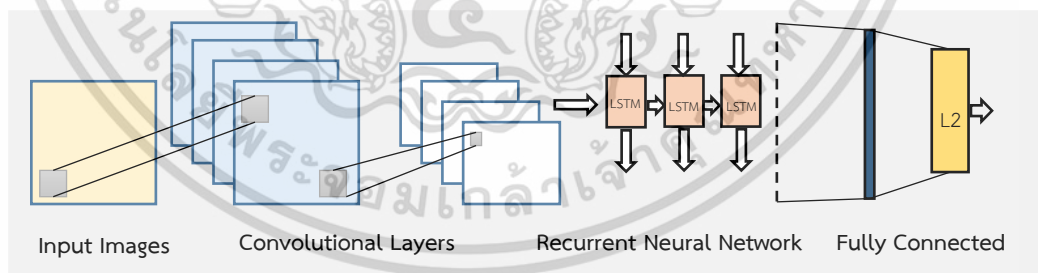
#### 3.2.1 สถาปัตยกรรมของโครงข่ายประสาท CNNLSTM

ในงานวิจัยนี้โครงข่ายประสาทเชิงลึก CNNLSTM ถูกนำมาใช้สำหรับการสกัดคุณลักษณะเด่นจากข้อมูลภาพ คุณลักษณะที่ผ่านการสกัดโดย CNNLSTM นำมาใช้เป็นข้อมูลตัวอย่างใหม่สำหรับการเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ฝึกสอนการเรียนรู้แบบตามตัวอย่าง (instance-based) เพื่อให้สามารถรู้จำรูปแบบของสิ่งแปลกปลอมที่เกิดขึ้นในข้อมูลภาพ

โครงสร้างชั้นของ CNNLSTM สำหรับการสกัดคุณลักษณะจากภาพได้รับแรงบันดาลใจมาจากโครงข่ายประสาทเทียมเชิงลึกที่เรียกว่า AlexNet ซึ่งได้ทำการปรับปรุงโครงสร้างชั้นของ AlexNet โดยการเพิ่มโครงข่ายประสาทเทียมแบบ LSTM ย่อมาจาก Long Short-Term Memory ซึ่งเป็นชนิดของ Recurrent Neural Network (RNN) เข้าไปในชั้นคอนโวลูชัน ที่ถูกออกแบบมาเพื่อแก้ไขปัญหา vanishing gradient problem LSTM ประกอบด้วยหลายเซลล์หน่วยความจำ แต่ละเซลล์หน่วยความจำมีสามองค์ประกอบ: เขียน (write) อ่าน (read) และลบ (forget)

ชั้นคอนโวลูชันใน CNNLSTM มีการขยายจากชั้นของ AlexNet ดังต่อไปนี้ ในชั้นคอนโวลูชันแรกและชั้นที่สองให้มีจำนวน kernel คอนโวลูชันเป็น 256 kernel และในชั้นคอนโวลูชันชั้นที่ห้า กำหนดให้มีจำนวน kernel ของคอนโวลูชันเท่ากับ  $512 \times 5$  kernel วิธีการ Batch-Normalization (BN) ถูกนำมาใช้กับผลลัพธ์ของชั้นคอนโวลูชัน โดยที่ผลลัพธ์ของการทำ BN เป็นชุดของจำนวนเต็มที่ส่งไปยังชั้นถัดไป ในชั้นคอนโวลูชันชั้นสุดท้ายได้มีการเพิ่ม LSTM เข้าไป LSTM นำเข้าข้อมูลจากการทำ BN มาใช้ ผลลัพธ์ของ LSTM เท่ากับ 512 ผลลัพธ์จากจำนวนหน่วยเอาต์พุตของ LSTM ชั้น fully connected ประกอบด้วย hidden neurons จำนวน 1,000 โหนด ฟังก์ชันกระตุ้น ของชั้น convolutional ทั้งหมดเป็นแบบ Rectified Linear Unit (ReLU) ชั้น output ประกอบด้วย neuron จำนวนหนึ่งโหนดและใช้ Sigmoid เป็นฟังก์ชันกระตุ้น โมเดล CNNLSTM ในการศึกษาถูกฝึกฝนด้วยวิธีการ Stochastic Gradient Descent (SGD) (Zhang, 2004) ภาพรวมของโครงสร้างโครงข่ายประสาท CNNLSTM แสดงในรูปที่ 3.4



รูปที่ 3.4 CNNLSTM

### 3.2.2 วิธีการเรียนรู้ข้อมูลสิ่งแปลกปลอม

ขั้นตอนวิธีการเรียนรู้ข้อมูลสิ่งแปลกปลอมและสกัดคุณลักษณะจากข้อมูลสิ่งแปลกปลอมด้วยโครงข่ายประสาท CNNLSTM แสดงได้ดังรูปที่ 3.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

---

**อัลกอริทึม 1:** วิธีการฝึกสอนและสร้างชุดโมเดล CNNLSTM
 

---

**Input:**  $\mathbf{S}$ : A training set,  $K$ : The number of the subdatasets,  $\theta$ : A performance threshold,  $\mathbf{m}_{k-1}$ : The pre-trained weights

**Output:**  $E$ : A set of CNNLSTM models,  $C$ : A set of representation datasets

1. Randomly split the training dataset into  $K$  sub datasets:  $\mathbf{S} \rightarrow \{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \dots, \mathbf{s}_K\}$ .
  2. For  $k = 1, 2, \dots, K$  do:
  3.  $AUC = 0$
  4. While  $AUC > \theta$  do:
    5. Train CNNLSTM from  $\mathbf{m}_{k-1}$  with  $\mathbf{s}_k$  to obtain the  $\mathbf{m}_k$  model.
    6. Extract deep features  $\mathbf{F}_k$  from  $\mathbf{s}_k$  by  $\mathbf{m}_k$ .
    7. Normalize every feature vectors  $\mathbf{f}_{ik}$  in  $\mathbf{F}_k$  with an  $L2$ -norm technique.
    8. For  $i$  in  $y \in \{0,1\}$  do:
      9. Compute the centroid of a class  $\mathbf{c}_{ik}$  from  $\mathbf{F}_{k \in N_{ik}}$ .
    10. End For
    11. Calculate the Euclidean distances from centroid  $\mathbf{c}_{ik}$  to every feature of all  $\mathbf{F}_k$  instances.
    12. Predicted label  $h_j$  of  $\mathbf{s}_k$  by using  $\mathbf{c}_{ik}$  and  $\mathbf{F}_k$ .
    13. Compute  $AUC$  value from predicted labels  $h_j$  with true label  $y$ .
  14. End While
  15.  $\mathbf{m}_k \in E$  and  $\mathbf{c}_{ik} \in C$ .
  16. End For
- 

**รูปที่ 3.5** ขั้นตอนการฝึกสอนโมเดล CNNLSTM

ในการฝึกสอนและสร้างโมเดล CNNLSTM จะทำการแบ่งชุดข้อมูลฝึกสอนออกเป็นชุดข้อมูลย่อย ด้วยการเลือกชุดข้อมูลตัวอย่างมาจากข้อมูลในชุดข้อมูลฝึกสอน  $\mathbf{S}$  ด้วยวิธีการสุ่มเลือกและนำข้อมูลที่สุ่มได้เหล่านั้นมาสร้างชุดข้อมูลย่อยหลาย ๆ ชุด โดยชุดข้อมูลย่อยให้แทนด้วย  $\mathbf{s}_k$  ประกอบด้วยเวกเตอร์ข้อมูล  $\mathbf{x}$  และคลาสคำตอบ  $y$  โดยมีข้อมูลทั้งหมดในชุดข้อมูลเท่ากับ  $N$  ข้อมูล:  $y: \mathbf{s}_k = \{(\mathbf{x}_{1k}, y_{1k}), \dots, (\mathbf{x}_{jk}, y_{jk}), \dots, (\mathbf{x}_{Nk}, y_{Nk})\}$  (บรรทัดที่ 1) โดยเป้าหมายของการฝึกสอนด้วยชุดข้อมูลย่อยเหล่านี้คือการสร้างความหลากหลายของโมเดล CNNLSTM และสร้างตัวแทนของข้อมูลของแต่ละคลาสคำตอบ (class representation) สำหรับใช้ในการจำแนกประเภทข้อมูล

ขั้นตอนการฝึกสอนกำหนดให้ชุดข้อมูลย่อย  $\mathbf{s}_k$  เป็นชุดข้อมูลฝึกสอนสำหรับ CNNLSTM โมเดลที่  $\mathbf{m}_k$  (บรรทัดที่ 5) โดยที่  $k$  คือลำดับปัจจุบัน โมเดล CNNLSTM ที่ได้รับการฝึกสอนในรอบชุดข้อมูลปัจจุบันจะนำค่าน้ำหนักของโมเดลที่ผ่านการฝึกสอนในรอบก่อนหน้า  $\mathbf{m}_{k-1}$  มาใช้เป็นค่าน้ำหนักเริ่มต้นสำหรับการฝึกสอนในรอบที่ปัจจุบัน ในขั้นตอนนี้ทำการฝึกสอนโมเดล  $\mathbf{m}_k$  จำนวน 5 epoch ด้วยชุดข้อมูลย่อย โมเดลที่ได้จากการฝึกสอนจะถูกนำไปใช้ในการสกัดคุณลักษณะจากข้อมูลภาพ

ขั้นตอนต่อมาทำการสร้างเซตของเวกเตอร์คุณลักษณะเด่น  $\mathbf{F}_k = \{\mathbf{f}_{1k}, \mathbf{f}_{2k}, \dots, \mathbf{f}_{Nk}\}$  จาก การสกัดคุณลักษณะด้วยโมเดล  $\mathbf{m}_k$  (บรรทัดที่ 6) ซึ่งคุณลักษณะ  $\mathbf{f}_{jk}$  ถูกสกัดมาจากข้อมูล  $\mathbf{x}_{jk}$  โดยที่  $\mathbf{f}_{jk} = \mathbf{m}_k(\mathbf{x}_{jk})$

ทุกเวกเตอร์คุณลักษณะใน  $\mathbf{F}_k$  ถูกนำมาทำการปรับปรุงให้เป็นรูปแบบ L2-norm (บรรทัดที่ 7) โดยกำหนดให้  $\mathbf{f}_{jk} = L2(\mathbf{f}_{jk})$  ค่า norm ของ  $\mathbf{f}_{jk}$  คำนวณได้ดังสมการที่ 3.1

$$|\mathbf{f}_{jk}| = \sqrt{\sum_{l=1}^{n^k} |v_{lk}|^2}, \quad (3.1)$$

โดยที่  $v_{lk}$  เป็นคุณลักษณะที่ได้จากการสกัดคุณลักษณะในภาพเป็นสมาชิกอยู่ใน  $\mathbf{f}_{jk}$  และ  $n^k$  คือขนาดของเวกเตอร์ของคุณลักษณะ  $\mathbf{f}_{jk}$  เพื่อให้วิธีการที่เสนอนี้สามารถรู้จำรูปแบบของของสิ่งแปลกปลอมที่เกิดขึ้นในภาพได้ จึงได้มีการทำค่านวนหาจุดศูนย์กลางของข้อมูลคลาสคำตอบด้วย  $\mathbf{c}_{ik}$  ในขั้นตอนนี้ จุดศูนย์กลางของคลาสคำตอบ  $\mathbf{c}_{ik}$  คือค่าเฉลี่ยของเวกเตอร์ของคุณลักษณะใน  $\mathbf{F}_k$  เฉพาะที่เป็นสมาชิกของคลาสคำตอบที่ลำดับที่  $i$  วิธีการคำนวณหาจุดศูนย์กลางของคลาสคำตอบ  $\mathbf{c}_{ik}$  แสดงได้ดังสมการที่ 3.2

$$\mathbf{c}_{ik} = \frac{1}{N_{ik}} \sum_{\mathbf{f}_{jk} \in N_{ik}} \mathbf{f}_{jk}, \quad (3.2)$$

โดยที่  $N_{ik}$  เป็นจำนวนของเวกเตอร์คุณลักษณะทั้งหมดของคลาสคำตอบที่  $i$  ที่เป็นสมาชิกอยู่ในชุดข้อมูลย่อย  $s_k$  (บรรทัดที่ 9) ทำการหาค่าความคล้ายคลึงของทุกคลาสคำตอบจากการหาค่าระหว่างเวกเตอร์คุณลักษณะ  $\mathbf{f}_{jk}$  และจุดศูนย์กลางของคลาสคำตอบ  $\mathbf{c}_{ik}$  วิธีการคำนวณค่าความคล้ายคลึงถูกคำนวณในบรรทัดที่ 11 ทำการคำนวณหาค่าความคล้ายกับทุกจุดศูนย์กลางของคลาสคำตอบโดยการคำนวณหาค่าความคล้ายจะใช้วิธีการหาระยะทางแบบ Euclidean และทำนายคลาสคำตอบ  $h_j$  จากการเลือกคลาสคำตอบที่มีระยะทางน้อยที่สุดซึ่งเป็นคลาสคำตอบที่มีความคล้ายคลึงกับข้อมูลที่สกัดได้นั้นมากที่สุดและจะทำนายคลาสคำตอบนั้นออกมา (บรรทัดที่ 12) ค่าความคล้ายคลึงและคลาสคำตอบที่ทำนายออกมา  $h_j$  ถูกนิยามดังสมการที่ 3.3

$$h_j = \arg \min_i \|\mathbf{f}_{jk} - \mathbf{c}_{ik}\|_2, \text{ for } j = 1 : Nk, \quad (3.3)$$

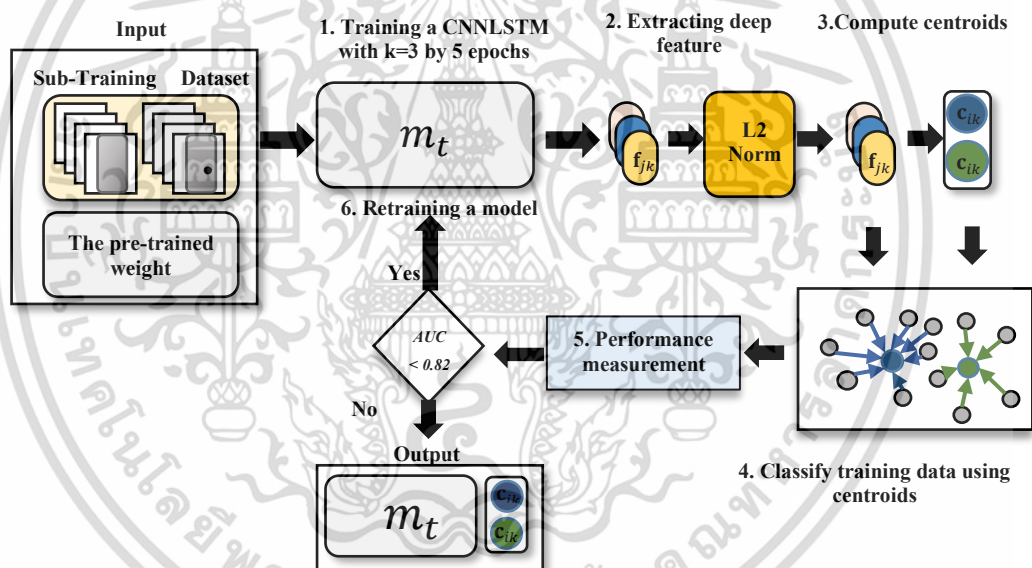
โดยที่  $Nk$  เป็นจำนวนของสมาชิกทั้งหมดในชุดข้อมูลย่อย  $s_k$

ขั้นตอนต่อมาของการฝึกสอนจะทำการประเมินประสิทธิภาพของโมเดล  $m_k$  โดยใช้เกณฑ์การวัดประสิทธิภาพแบบ AUC (Area Under the Curve) ประสิทธิภาพของโมเดลจะถูกวัดด้วยวิธีการนี้ในทุก ๆ 5 รอบของการฝึกสอนโมเดล หากค่า AUC ของโมเดลจากการทำนายคลาสคำตอบของชุดฝึกสอนมีค่าน้อยกว่าค่า  $\theta$  ที่กำหนดโดยผู้ใช้ โมเดล  $m_k$  จะต้องกลับไปทำการฝึกสอนโมเดลใหม่อีก 5 epoch (ย้อนกลับไปบรรทัดที่ 5) และคุณลักษณะ  $\mathbf{F}_k$  และ จุดศูนย์กลางของคลาสคำตอบ  $\mathbf{c}_{ik}$  จะถูกคำนวณใหม่ไปด้วย ในทางกลับกันหากค่า AUC ในรอบการฝึกสอนนั้นมากกว่าค่า  $\theta$  การฝึกสอนในรอบปัจจุบันจะหยุดลงและโมเดล  $m_k$  จะเป็นสมาชิกในชุดโมเดล  $E$  โดยที่โมเดล  $m_k \in E$  และ ศูนย์กลางของคลาสคำตอบ  $\mathbf{c}_{ik} \in C$  (บรรทัดที่ 15)

กระบวนการฝึกสอนที่นำเสนอจะช่วยให้การสกัดคุณลักษณะเด่นจากภาพที่มีคลาสคำตอบเดียวกันกับศูนย์กลางของคลาสคำตอบนั้นมีความคล้ายกันมากที่สุดและมีระยะห่างระหว่าง

ศูนย์กลางที่ไม่ได้เป็นคลาสคำตอบเดียวกันกับภาพมากที่สุด ทำให้สามารถตรวจจับสิ่งแปลกปลอมได้อย่างถูกต้อง

ค่าน้ำหนักของโครงข่ายประสาท CNNLSTM ที่ผ่านการฝึกสอนในครั้งปัจจุบันที่  $k$  จะถูกนำมาใช้เป็นค่าน้ำหนักเริ่มต้นของโครงข่ายประสาทในการฝึกสอนครั้งถัดไป  $m_k \rightarrow m_{k+1}$  และทำการฝึกสอนด้วยชุดข้อมูลย่อยที่  $S_{k+1}$  วิธีการนี้ช่วยให้เวลาที่ใช้ในการฝึกสอนโครงข่ายประสาทในแต่ละครั้งมีเวลาที่ลดลงและปรับปรุงความแม่นยำในตรวจจับสิ่งแปลกปลอมจากการฝึกสอนครั้งก่อนหน้า และเป็นวิธีการที่สะดวกต่อการใช้น้ำหนักที่ถูกฝึกจากโมเดลก่อนหน้านี้เพื่อมาปรับปรุงในรอบของการฝึกสอนรอบ ถัดๆไป วิธีการฝึกสอนเพื่อสร้างโมเดลที่น่าเสนอนี้สามารถเห็นได้ว่าโมเดลที่เป็นสมาชิกในชุดโมเดล  $E$  เป็นชุดของโมเดลที่มีขอบเขตการตัดสินใจที่แตกต่างกันแต่มีชั้นโครงสร้างของโครงข่ายประสาทที่เหมือนกัน และชุดของจุดศูนย์กลางของคลาสคำตอบที่เป็นสมาชิกทั้งหมดใน  $C$  เป็นตัวแทนของรูปแบบของข้อมูลสิ่งแปลกปลอมที่ทำการฝึกสอนในแต่ละครั้ง ภาพรวมของวิธีฝึกสอนเพื่อรู้จำคุณลักษณะของสิ่งแปลกปลอมได้แสดงไว้ในภาพที่ 3.6



รูปที่ 3.6 ขั้นตอนวิธีการฝึกสอนเพื่อรู้จำคุณลักษณะของสิ่งแปลกปลอม

ความซับซ้อนของขั้นตอนการฝึกสอนโมเดล CNNLSTM นั้นมีการวนซ้ำรูป for เพื่อสร้างชุดของโมเดล CNNLSTM จำนวน  $k$  รอบ ฟังก์ชันมีความซับซ้อนของขั้นตอนนี้คือ  $O(n)$  ในการวนรอบแต่ละครั้งจะลดจำนวนการทำซ้ำไปที่ละหนึ่ง แต่ในรอบของ  $k$  จะมีการวนซ้ำเพื่อฝึกสอนโมเดล CNNLSTM ดังนั้น ฟังก์ชันความซับซ้อนของ CNNLSTM คือ  $O(\log_{ker}(n))$  โดยที่  $ker$  คือ ขนาดของ kernel และ  $n$  จำนวนของข้อมูล และมีการคำนวณค่า AUC ซึ่งอาจมีผลต่อประสิทธิภาพของโมเดล จำนวนรอบของลูป While ขึ้นอยู่กับค่า AUC เมื่อมีค่ามากกว่าเกณฑ์  $\theta$  ที่กำหนด ฟังก์ชันมีความซับซ้อนของขั้นตอนนี้คือ  $O(m)$  โดยที่  $m$  แทนจำนวนรอบการทำซ้ำที่มากที่สุดของลูป While

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้นโปรแกรมนี้จะมี ฟังก์ชันมีความซับซ้อนอยู่ในช่วงของ  $O(nm \times \log_{ker}(n))$  นั่นเอง อาจมีความแตกต่างของเวลาการทำงานของโปรแกรมในแต่ละครั้งขึ้นอยู่กับขนาดของชุดของโมเดลและค่า AUC ที่เกิดขึ้นในแต่ละรอบของการทำซ้ำ

### 3.2.3 วิธีการตรวจจับสิ่งแปลกปลอมด้วยวิธีการจำแนกประเภทตามตัวอย่าง

วิธีการตรวจจับสิ่งแปลกปลอมโดยใช้ CNNLSTM จะกล่าวถึงในส่วนนี้ เป็นเทคนิคที่ช่วยในการป้องกันอคติในการทำนายคลาสคำตอบ มีการคำนวณค่าน้ำหนักระยะทางสำหรับทุกคลาสคำตอบ ขั้นตอนวิธีการของการจำแนกประเภทแสดงในรูปที่ 3.7

#### อัลกอริทึม 2: ขั้นตอนวิธีการตรวจจับสิ่งแปลกปลอมด้วยวิธีการจำแนกประเภทตามตัวอย่าง

Input:

$E$ : CNNLSTM models,

$C$ : A set of centroid class data,

$X$ : An unseen data

Output:  $h_f = \operatorname{argmax}_i \sum_i \sum_k \log \left( \frac{1}{ds_{ik}} \right)$

1. For  $k = 1, 2, \dots, K$  do:
2.     Extract features  $\mathbf{f}_k = m_k(\mathbf{x})$ .
3.     Normalize the feature vectors  $\mathbf{f}_k$  by L2-norm.
4.     For  $i = 1: y$  do:
5.         Calculate the distance  $d_{ik}$  between centroid  $\mathbf{c}_{ik}$  to features  $\mathbf{f}_k$ .
6.         Compute the distance weight  $ds_{ik}$ .
7.     End For
8. End For

รูปที่ 3.7 ขั้นตอนวิธีการตรวจจับสิ่งแปลกปลอม

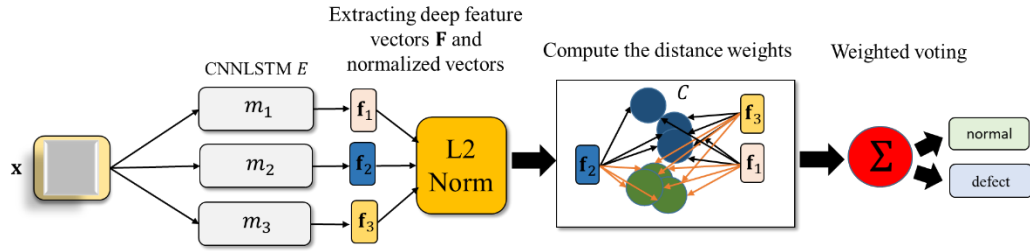
จากรูปที่ 3.8 อินพุตของอัลกอริทึมนี้ประกอบด้วยชุดของโมเดล  $E$  และเซตของ  $C$  ส่วนผลลัพธ์ที่ได้จากการทำงานนี้คือคลาสคำตอบประกอบด้วย 0 คือขจัดและ 1 คือขจัดที่มีสิ่งแปลกปลอมคำนวณได้ดังสมการ 3.4

$$h_f = \operatorname{argmax}_i \sum_i \sum_k \log \left( \frac{1}{ds_{ik}} \right), \quad (3.4)$$

$h_f$  คือคลาสคำตอบที่ทำนาย  $i$  คือดัชนีคลาสคำตอบ  $k$  คือจำนวนโมเดลใน  $E$  และ  $ds_{ik}$  คือค่าน้ำหนักระยะทางที่คำนวณโดยสมการที่ 3.6 จากสมการนี้ค่าน้ำหนักระยะทาง  $ds_{ik}$  ที่มีค่าสูงจะให้ค่าน้ำหนักในการโหวตคลาสคำตอบที่มีค่าต่ำ และในทางกลับกันค่าน้ำหนักระยะทางที่มีค่าต่ำจะมีค่าน้ำหนักในการโหวตคลาสคำตอบที่สูง

รูปที่ 3.8 แสดงขั้นตอนวิธีการจำแนกประเภทตามตัวอย่าง ซึ่งในขั้นตอนแรกทำการสกัด

คุณลักษณะจากข้อมูลภาพที่ต้องการตรวจจับสิ่งแปลกปลอม:  $\mathbf{f}_k = m_k(\mathbf{x})$  หลังจากนั้นคุณลักษณะเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.8 วิธีการจำแนกประเภทจำแนกประเภทตามตัวอย่าง

$f_k$  จะถูกทำให้เป็นมาตรฐานโดยใช้เทคนิค L2-norm:  $f_k = L2(f_k)$  โดยที่  $f_k$  ถูกทำให้เป็นมาตรฐานในปริภูมิเวกเตอร์เดียวกันของจุดศูนย์กลาง  $c_{ik}$  เพื่อใช้ในการคำนวณค่าระยะความเหมือน  $\alpha_{ik}$  ระหว่างคุณลักษณะ  $f_k$  และจุดศูนย์กลางของคลาส  $c_{ik}$  ค่าระยะความเหมือนคำนวณโดยสมการ 3.5

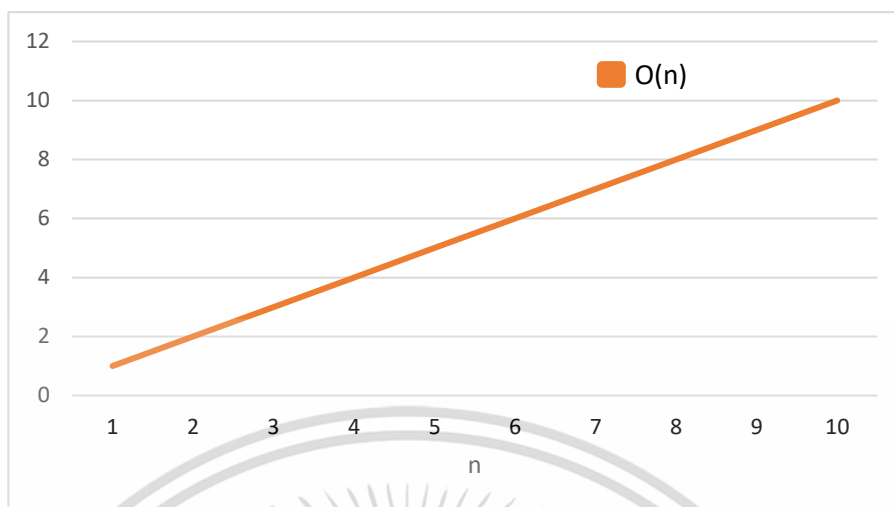
$$\alpha_{ik} = \|f_k - c_{ik}\|_2, \text{ for } k = 1:K, \quad (3.5)$$

โดยที่  $\alpha_{ik}$  คือค่าระยะทางความคล้ายของคลาส  $ith$  เวกเตอร์คุณลักษณะที่สกัดได้จากแต่ละโมเดล ถูกนำมาคำนวณระยะทางความคล้ายของทุกจุดศูนย์กลางของคลาสคำตอบใน  $C$  คำนวณสำหรับการโหวตคลาสคำตอบจะคำนวณดังสมการ 3.6

$$ds_{ik} = \frac{\alpha_{ik}}{\sum_i \alpha_{ik}} \quad (3.6)$$

น้ำหนักสำหรับโหวตคำตอบ  $ds_{ik}$  คือระยะทางความคล้ายเฉลี่ยระหว่างเวกเตอร์คุณลักษณะและศูนย์กลางของคลาสคำตอบ หากข้อมูลที่ต้องการตรวจจับสิ่งแปลกปลอม  $x$  เป็นข้อมูลที่มีสิ่งแปลกปลอม คุณลักษณะที่สกัดได้จากภาพด้วย CNNLSTM จะคล้ายกับจุดศูนย์กลางของคลาสสิ่งแปลกปลอม และน้ำหนักระยะทางที่คำนวณได้โดยสมการ 3.6 จะมีค่าที่สูง ในทางกลับกันค่าน้ำหนักระยะทางจะมีค่าที่ต่ำ

ขั้นตอนวิธีการตรวจจับสิ่งแปลกปลอมด้วยวิธีการจำแนกประเภทตามตัวอย่างนั้น อัลกอริทึมนี้มีความซับซ้อน ฟังก์ชันมีความซับซ้อนคือ  $O(n)$  เนื่องจากมีการวนซ้ำรูป for เป็นจำนวน  $k$  รอบการทำซ้ำเท่ากับจำนวนชุดของโมเดล โดยที่อัลกอริทึมนี้จะมีการทำซ้ำด้านในรูป for อีกครั้ง การทำซ้ำเท่ากับจำนวนของคลาสคำตอบเสมอ ดังนั้น ฟังก์ชันมีความซับซ้อนคือ  $O(1)$  ขั้นตอนนี้ทำเพื่อหาค่าระยะทางระหว่างจุดศูนย์กลางของคลาสคำตอบและคุณลักษณะที่สกัดโดย CNNLSTM จึงไม่มีผลต่อการเติบโตของโปรแกรม ดังนั้นอัตราการเติบโตของฟังก์ชันหรือความซับซ้อนที่มากที่สุดของวิธีการตรวจจับสิ่งแปลกปลอมด้วยวิธีการจำแนกประเภทตามตัวอย่างนั้นคือ  $O(n)$  โดยกราฟจะเป็นเส้นตรงหรือ linear graph ความซับซ้อนของอัลกอริทึมการตรวจจับสิ่งแปลกปลอมจะคงที่ตลอดเวลาเมื่อมีการเปลี่ยนแปลงขนาดของชุดของโมเดล  $k$  ดังแสดงในรูปที่ 3.9 จากกราฟแกนอนแสดงค่าที่เพิ่มขึ้นของจำนวนโมเดล CNNLSTM และแกนตั้งแสดงค่าความซับซ้อนของเวลา



รูปที่ 3.9 กราฟแสดงความซับซ้อนของวิธีการจำแนกประเภทจำแนกประเภทตามตัวอย่าง

### 3.3 วิธีการตรวจจับสิ่งแปลกปลอมโดยใช้โครงข่ายประสาทแบบ Autoencoder และ Threshold ของข้อผิดพลาด

วิธีการตรวจจับสิ่งแปลกปลอมบนขวดแก้วโดยใช้โครงข่ายประสาทแบบ Autoencoder และ Threshold ของข้อผิดพลาดในงานวิจัยนี้เรียกว่า DDAET (Defect detection for glass bottle using autoencoder and threshold error) แนวคิดหลักของวิธีการ DDAET สำหรับตรวจจับสิ่งแปลกปลอมวิธีการนี้คือ 1) แก้ปัญหาความซับซ้อนในการฝึกสอนโมเดลและลดการใช้ข้อมูลภาพตัวอย่างจำนวนมากในการฝึกสอนของโครงข่ายประสาทเชิงลึก 2) สร้างวิธีการตรวจจับสิ่งแปลกปลอมที่ใช้เฉพาะชุดข้อมูลภาพดีในการฝึกสอน 3) การลดเวลาในการตรวจจับสิ่งแปลกปลอม

#### 3.3.1 สถาปัตยกรรมของ Autoencoder

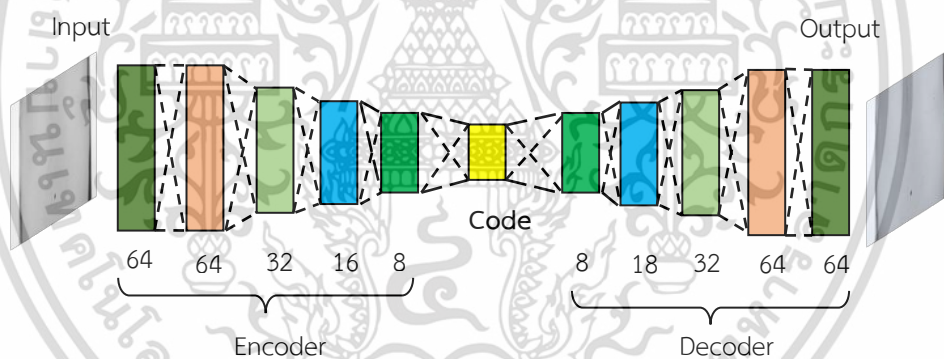
สำหรับวิธีการ DDAET ได้ทำการออกแบบโครงสร้างของโครงข่ายประสาทแบบ Autoencoder ให้มีขนาดเล็กและใช้สำหรับการเรียนรู้เฉพาะชุดข้อมูลภาพขวดดีเท่านั้น การออกแบบโครงสร้างชั้นของโครงข่ายประสาท Autoencoder ที่มีขนาดเล็กเพื่อให้การฝึกสอนโมเดลมีความรวดเร็วและใช้ทรัพยากรของเครื่องน้อย ใช้จำนวนของภาพที่นำมาฝึกสอนโมเดลไม่มาก มีจำนวนพารามิเตอร์ที่ต้องทำการปรับในขั้นตอนการฝึกสอนน้อย ตรวจจับข้อบกพร่องบนขวดแก้วได้รวดเร็ว และเป็นมิตรกับผู้ใช้งานจริง

โครงข่ายประสาทแบบ Autoencoder ประกอบด้วยสองส่วนหลักคือ encoder และ decoder รายละเอียดของส่วน encoder และ decoder ในโครงสร้างโมเดล autoencoder ที่ได้ทำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การออกแบบใหม่อธิบายดังต่อไปนี้ 1) ชั้นนำเข้า (Input layer) ให้มีขนาด 224x224 พิกเซลเท่ากับขนาดของภาพที่ใช้ฝึกสอน 2) ส่วน encoder ประกอบด้วยชั้น convolutional จำนวน 5 ชั้นแต่ละชั้นประกอบด้วยคอร์เนลจำนวน 64 64 32 16 และ 8 คอร์เนลตามลำดับ 3) ส่วน decoder ประกอบด้วยชั้น convolutional ทั้งหมด 5 ชั้นประกอบด้วยคอร์เนลจำนวน 64 64 32 16 และ 8 คอร์เนลตามลำดับ ขนาดคอร์เนลของทุกชั้น convolutional ถูกตั้งค่าให้มีขนาดเท่ากับ 2x2 และทำการ stride เป็น 1 ครั้งและ padding เป็นแบบ same อีกด้วย

Autoencoder ที่ออกแบบนี้ใช้ฟังก์ชัน activation แบบ Rectified Linear Unit (ReLU) ในทุกชั้น convolutional และฟังก์ชัน activation ของชั้นเอาต์พุตสุดท้ายนั้นใช้ฟังก์ชัน activation แบบ Sigmoid และใช้ Adam optimizer ในการปรับค่าพารามิเตอร์ต่างๆด้วยอัตราการเรียนรู้กำหนดให้เท่ากับ 0.001 และ Batch size เท่ากับ 5 ภาพ และโครงข่ายประสาทนี้มีจำนวนพารามิเตอร์ที่ต้องปรับทั้งหมดจำนวนเท่ากับ 57,299 พารามิเตอร์ ผลลัพธ์ที่ได้จาก autoencoder คือการสร้างรูปภาพที่ผ่านการ encoder และ decoder จากภาพต้นฉบับที่ป้อนเข้าไปยังโครงข่ายประสาท โครงสร้างของ autoencoder ที่นำเสนอแสดงในรูปที่ 3.10



รูปที่ 3.10 Autoencoder

### 3.3.2 สร้างภาพพื้นหลังปกติและค่า Threshold ของข้อผิดพลาด

วิธีการฝึกสอนโมเดล autoencoder ของ DDAET ทำการฝึกสอนด้วยชุดข้อมูลภาพที่ไม่มีสิ่งแปลกปลอม (normal images) เพื่อใช้ตรวจจับสิ่งแปลกปลอมเป็นแนวคิดหลักของวิธีการฝึกโมเดล autoencoder ที่นำเสนอในงานวิจัยนี้ ชุดข้อมูลภาพสำหรับฝึกสอนโครงข่ายประสาทใช้ภาพขวดที่ไม่มีข้อบกพร่องมาใช้ในการฝึกสอนโมเดลและใช้วิธีการขยายข้อมูล (Data augmentation) เพื่อเพิ่มภาพขวดที่ไม่มีข้อบกพร่องให้มีจำนวนเพิ่มขึ้นทำให้ประสิทธิภาพของโมเดลมีมากขึ้นในขั้นตอนการฝึกสอน ในการตรวจจับสิ่งแปลกปลอมของขวดในไลน์การผลิตจริง การตรวจจับสิ่งแปลกปลอมของขวดจะต้องใช้ความเชี่ยวชาญของผู้ตรวจจับเพื่อระบุข้อบกพร่องของขวดในไลน์การผลิตขวดแต่ละชนิด ขวดที่มีสิ่งแปลกปลอมเป็นขวดที่น้อยกว่าขวดปกติที่ไม่มีสิ่งแปลกปลอมอย่างมาก ด้วยเหตุนี้ การเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

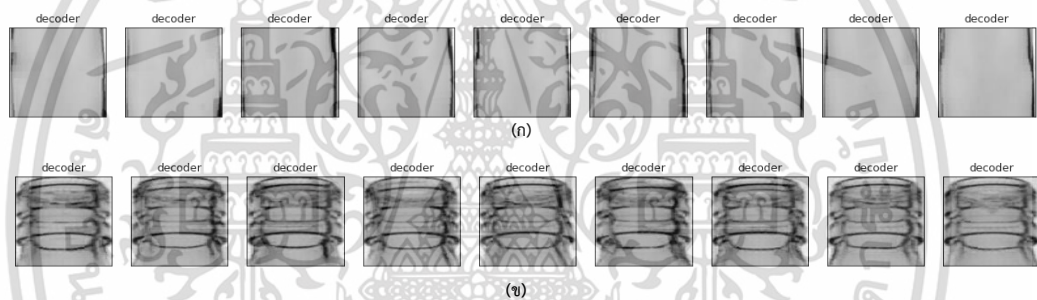
เก็บรวบรวมภาพขดที่ผิดปกติหรือมีสิ่งแปลกปลอม (anomaly images) เพื่อนำมาใช้ฝึกสอนโมเดล autoencoder จึงเป็นงานที่ทำได้อย่างง่ายสำหรับการรวบรวมภาพของขดที่มีสิ่งแปลกปลอมให้ได้จำนวนมาก และได้ชนิดของสิ่งแปลกปลอมที่ครบถ้วนสำหรับนำมาฝึกสอนโมเดล autoencoder

ขั้นตอนการสร้างภาพพื้นหลังปกติและค่า threshold ของข้อผิดพลาด สำหรับการตรวจจับสิ่งแปลกปลอมบนขดแก้วที่ได้นำเสนอในงานวิจัยนี้ได้อธิบายดังต่อไปนี้

1) ทำการฝึกสอนโครงข่ายประสาท autoencoder เพื่อให้โมเดลรู้จักภาพของขดที่ไม่มีสิ่งแปลกปลอม โมเดล autoencoder สามารถลบข้อมูลรบกวนออกจากภาพได้ ภาพที่สร้างขึ้นจาก autoencoder หลังจากทำการเข้ารหัสและถอดรหัสจากภาพเดิมถูกกำหนดโดยสมการที่ 3.7

$$\mathbf{y} = \varphi(\mathbf{x}), \quad (3.7)$$

กำหนดให้  $\varphi$  คือโมเดล autoencoder  $\mathbf{x}$  คือภาพขดปกติที่ไม่มีสิ่งแปลกปลอมที่เป็นอินพุตของโมเดล และ  $\mathbf{y}$  คือภาพที่ถูกสร้างขึ้นจากการเข้ารหัสและถอดรหัสกลับจากภาพเดิม รูปที่ 3.11 (ก) แสดงภาพที่ถูกสร้างใหม่บริเวณส่วนลำตัวของขด และ (ข) แสดงภาพที่ถูกสร้างใหม่ของส่วนคอขดแก้ว

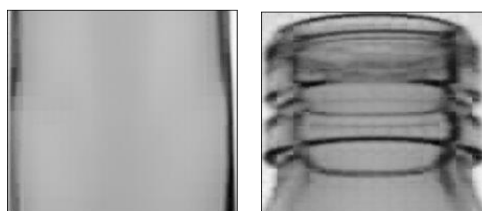


รูปที่ 3.11 ภาพที่สร้างขึ้นใหม่ (ก) ภาพส่วนลำตัว (ข) ภาพส่วนคอ

2) สร้างภาพพื้นหลังปกติโดยการหาค่าเฉลี่ยของภาพที่ถูกเข้ารหัสของชุดข้อมูลภาพขดปกติทั้งหมดที่ได้ทำการฝึกสอนโมเดล ให้ภาพพื้นหลังปกติแทนด้วย  $\mathbf{B}$  ซึ่งได้นิยามโดยสมการ 3.8

$$\mathbf{B}_{ij} = \frac{1}{N} \sum_{n=1}^N y_{ij}^n, \quad (3.8)$$

โดยที่  $N$  คือจำนวนรวมของภาพขดแก้วปกติในชุดข้อมูลฝึกสอน  $i$  เป็นจำนวนแถวของภาพ  $j$  เป็นจำนวนคอลัมน์ของภาพ ภาพตัวอย่างพื้นหลังปกติของลำตัวขดแก้วและบริเวณคอของขดแก้วแสดงดังรูปที่ 3.12



รูปที่ 3.12 ภาพพื้นหลังปกติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) คำนวณเมทริกซ์ของค่าเกณฑ์การตัดสินใจให้แทนด้วย  $T$  ค่าเกณฑ์การตัดสินใจคำนวณมาจากค่าเฉลี่ยของค่า threshold ของข้อผิดพลาดรวมกับค่าเบี่ยงเบนที่คำนวณหาค่าความผิดพลาดมาจากภาพขจัดปกติที่เป็นข้อมูลการฝึกสอนและภาพที่สร้างขึ้นจากโมเดล autoencoder ด้วยสมการ 3.9

$$T_{ij} = e_{ij} + s_{ij}, \quad (3.9)$$

โดยที่  $T_{ij}$  คือค่า threshold ของข้อผิดพลาด  $e_{ij}$  คือค่าเฉลี่ยความผิดพลาด (MSE) ของคุณลักษณะระหว่างภาพข้อมูลนำเข้า  $\mathbf{x}$  และภาพผลลัพธ์การเข้ารหัส  $\mathbf{y}$  ในแถวเดียวกันที่  $i$  และคอลัมน์เดียวกันที่  $j$  และ  $s_{ij}$  คือส่วนเบี่ยงเบนมาตรฐานของค่าความผิดพลาด ค่าเฉลี่ยความผิดพลาด  $e_{ij}$  คำนวณตามสมการ 3.10

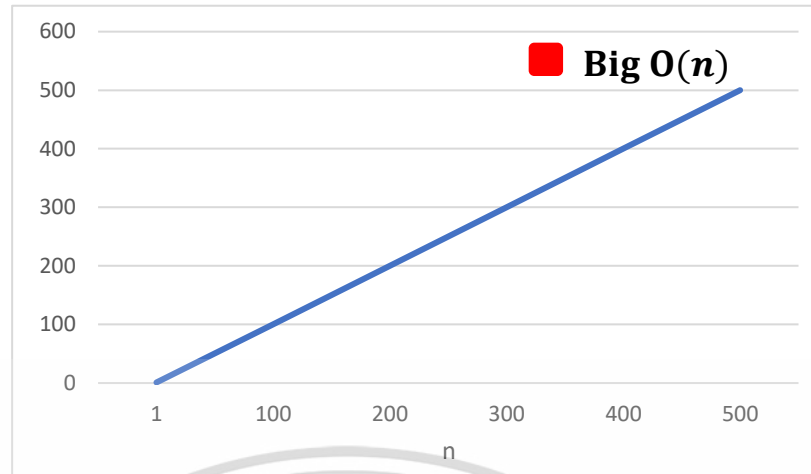
$$e_{ij} = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_{ij}^n - \mathbf{y}_{ij}^n)^2. \quad (3.10)$$

ค่าส่วนเบี่ยงเบนมาตรฐานของค่าความผิดพลาด  $S$  จากชุดข้อมูลฝึกสอนของขจัดปกติคำนวณโดยสมการ 3.11

$$s_{ij} = \sqrt{\frac{1}{N} \sum_{n=1}^N (e_{ij} - \overline{e_{ij}})^2}. \quad (3.11)$$

ขนาดของเมทริกซ์ของค่าเกณฑ์การตัดสินใจ  $T$  มีขนาดเท่ากับขนาดของภาพอินพุต (224 แถว 224 คอลัมน์) ตรวจสอบสิ่งแปลกปลอมบนขจัดปกติโดยใช้ค่าเกณฑ์ความผิดพลาดเมื่อเทียบระหว่างภาพที่ต้องการตรวจสอบกับภาพพื้นหลังปกติ

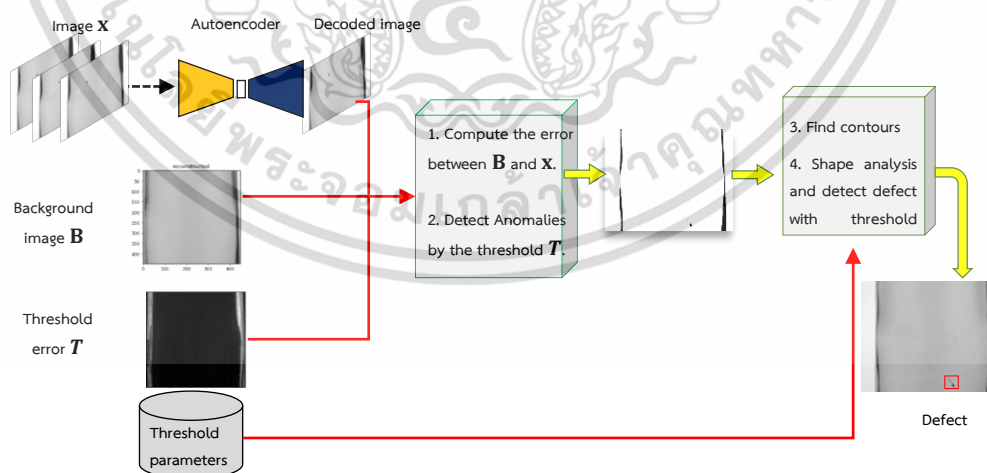
ในกระบวนการสร้างภาพพื้นหลังปกติและค่า Threshold ของข้อผิดพลาดของวิธีการ DDAET ทั้งหมดนี้มีขั้นตอนการทำงานแต่ละขั้นตอนไม่มีการทำซ้ำในรูปดังนั้นจะสามารถประมาณค่าฟังก์ชันความซับซ้อนของแต่ละขั้นตอนได้ดังนี้ ขั้นตอนฝึกสอนโครงข่ายประสาท autoencoder มีค่าฟังก์ชันมีความซับซ้อนของขั้นตอนนี้คือ  $O(n)$  โดยที่  $n$  เป็นจำนวน epoch ของการฝึกสอนโมเดล คำนวณหาค่าความซับซ้อนของกระบวนการทั้งหมดจะเป็นค่าที่สูงสุดของขั้นตอนที่ใช้เวลามากที่สุดซึ่งในกรณีนี้คือ  $O(n)$  โดยกราฟจะเป็นเส้นตรงหรือ linear graph ซึ่งหมายความว่าเมื่อขนาดของ epoch เพิ่มขึ้นอย่างเป็นเส้นตรง (เพิ่ม  $n$ ) ความซับซ้อนของอัลกอริทึมจะเพิ่มขึ้นเสมอแสดงในรูปที่ 3.13 จากกราฟแกนอนแสดงค่าที่เพิ่มขึ้นของจำนวน epoch ของการฝึกสอนโมเดล และแกนตั้งแสดงค่าความซับซ้อนของเวลา



รูปที่ 3.13 กราฟแสดงความซับซ้อนของกระบวนการสร้างภาพพื้นหลังปกติและค่า Threshold ของข้อผิดพลาด

### 3.3.3 การตรวจจับสิ่งแปลกปลอมด้วยภาพพื้นหลังปกติและค่า Threshold ของข้อผิดพลาด

ในส่วนนี้อธิบายวิธีการตรวจจับสิ่งแปลกปลอมของ DDAET ด้วยภาพพื้นหลังปกติและ Threshold ของข้อผิดพลาด ตรวจจับสิ่งแปลกปลอมที่เกิดขึ้นบนขวดได้ด้วยการหาค่าความผิดพลาดจากภาพขวดที่ถูกเข้ารหัสจากโครงข่ายประสาท autoencoder เมื่อเทียบกับภาพพื้นหลังปกติ หากค่าความผิดพลาดมีมากกว่าเกณฑ์ค่าความผิดพลาดจากชุดข้อมูลฝึกสอนขวดปกติ ขั้นตอนการตรวจจับสิ่งแปลกปลอมบนขวดแก้วโดยรวมแสดงในรูปที่ 3.14



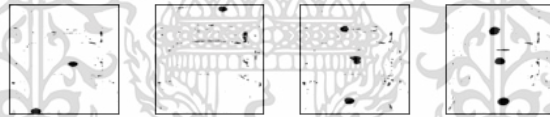
รูปที่ 3.14 วิธีการตรวจจับสิ่งแปลกปลอมด้วยภาพพื้นหลังและค่า Threshold ของข้อผิดพลาด

จากรูปที่ 3.14 มีขั้นตอนการทำงาน ดังนี้

1) คำนวณค่าความผิดพลาดของแต่ละพิกเซลระหว่างภาพพื้นหลังปกติ  $\mathbf{B}$  (ซึ่งเป็นค่าเฉลี่ยของภาพที่ถอดรหัสของชุดข้อมูลภาพขจัดปกติตามที่กำหนดในสมการที่ 3.8) และภาพขจัดที่ผ่านการเข้ารหัสแล้ว  $\mathbf{x} = \varphi(\mathbf{x})$  (ซึ่งเป็นภาพที่ต้องการตรวจจับสิ่งแปลกปลอม) กำหนดให้ค่าความผิดพลาดของพิกเซลเป็น  $\rho_{ij}$  ทำการคำนวณหาค่าความผิดพลาดตามสมการที่ 3.12.

$$\rho_{ij} = \sqrt{(\mathbf{B}_{ij} - \mathbf{x}_{ij})^2}. \quad (3.12)$$

2) การตรวจจับสิ่งแปลกปลอมด้วยเกณฑ์การตัดสินใจ  $T$  เพื่อแปลงภาพขจัดแก้วที่ต้องการตรวจจับสิ่งแปลกปลอม  $\mathbf{x}$  ให้เป็นภาพที่มีพิกเซลแบบไบนารี จากการคำนวณหาค่าความผิดพลาด  $\rho_{ij}$  ระหว่างพิกเซลกับพิกเซลของภาพพื้นหลังปกติจากชุดข้อมูลฝึกสอนขจัดปกติ ถ้าค่าความผิดพลาด  $\rho_{ij}$  มากกว่าค่าเกณฑ์ความผิดพลาด  $T_{ij}$ :  $\rho_{ij} > T_{ij}$  กำหนดให้พิกเซลที่  $\mathbf{x}_{ij}$  มีค่าเท่ากับ 0 ถ้าค่าความผิดพลาด  $\rho_{ij}$  น้อยกว่าค่าเกณฑ์ความผิดพลาด  $T_{ij}$ :  $\rho_{ij} < T_{ij}$  กำหนดให้พิกเซล  $\mathbf{x}_{ij}$  มีค่าเท่ากับ 1 จากนั้นทำการแปลงค่าพิกเซลของภาพ  $\mathbf{x}$  ให้มีค่าอยู่ในช่วงตั้งแต่ 0 ถึง 255 โดย  $\mathbf{x} = \mathbf{x} \times 255$  วิธีการแปลงพิกเซลนี้สามารถที่จะทำให้เห็นพิกเซลที่เป็นบริเวณส่วนของสิ่งแปลกปลอมจะเป็นพิกเซลสีดำ ที่เกิดขึ้นในภาพจากค่าความผิดพลาดที่เกินเกณฑ์ ดังรูปที่ 3.15



รูปที่ 3.15 วิธีการตรวจจับสิ่งแปลกปลอม

รูปที่ 3.15 แสดงการใช้ autoencoder สำหรับการเข้ารหัสและถอดรหัสภาพ ทำให้สามารถเน้นขอบกรอบในภาพที่ถอดรหัส ทำให้แยกแยะความแตกต่างจากพื้นหลังได้ และช่วยให้ตรวจจับได้ง่ายขึ้นเมื่อเทียบกับภาพต้นฉบับที่ยังไม่ได้ประมวลผล

3) คั่นหารูปร่าง (contour) ของบริเวณพิกเซลสีดำที่ปรากฏในภาพที่ได้ทำการแปลงด้วยขั้นตอนที่สอง

4) DDAET จะจำแนกประเภทสิ่งแปลกปลอมตามลักษณะของ contours รูปทรงของสิ่งแปลกปลอมสามารถอธิบายได้ด้วยรูปร่างและเส้นโค้งของสิ่งแปลกปลอม contours เป็นเครื่องมือที่มีประโยชน์สำหรับการวิเคราะห์รูปทรงในงานการตรวจจับและจดจำวัตถุ (object detection) สำหรับการจำแนกประเภทขอบกรอบใช้ปัจจัย 4 อย่างได้แก่ ขนาด (size) ความหนา (thickness) ความกะทัดรัด (compactness) และการกรอก (filling) เพื่อนำมาสร้างกฎการจำแนกประเภทสิ่งแปลกปลอมสามข้อดังต่อไปนี้:

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. if  $size \geq 15$  and  $compactness \leq 2$  then:

the defect is stone;

2. if  $compactness > 2$  and  $thickness \geq 10$  then:

the defect is a tear;

3. if  $compactness \leq 2$  and  $filling < 45$  then:

the defect is a blister.

1) ค่า  $size$  คือขนาดของวัตถุที่สกัดได้เป็นค่าจำนวน pixel คำนวณได้ดังสมการ 3.13

$$Size = width \times height \quad (3.13)$$

โดยที่  $width$  คือความกว้างของวัตถุ และ  $height$  เป็นความสูงของวัตถุ

2) ค่า  $compactness$  คือค่าความกะทัดรัดของวัตถุเป็นค่าที่นำมาใช้ในการจำแนกประเภทวัตถุว่ามีรูปทรงเป็นทางยาวหรือเป็นรูปทรงกลม คำนวณได้จากสมการ 3.14

$$Compactness = \frac{Thickness}{Length} \quad (3.14)$$

โดยที่  $Thickness$  คือความหนาของวัตถุ และ  $Length$  เป็นความยาวของวัตถุ

3) ค่า  $Filling$  เป็นค่าที่บ่งบอกว่าวัตถุนั้นมีค่าความทึบของเส้นขอบมากเท่าใด เป็นการหาเปอร์เซ็นต์ระหว่าง pixel ขาว และ pixel ดำในวัตถุ ค่า  $filling$  นำมาใช้ในการตรวจจับฟองอากาศ คำนวณได้ดังสมการ 3.15

$$Filling = \frac{size \times 100}{Thickness \times length} \quad (3.15)$$

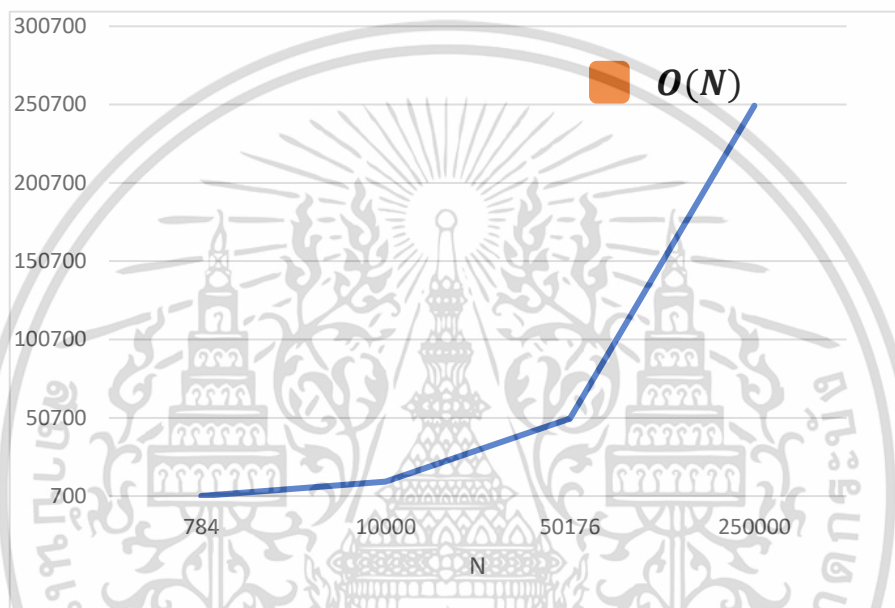
ในกระบวนการจำแนกประเภทสิ่งแปลกปลอมนี้ ผู้ใช้งานเป็นผู้กำหนดค่าให้กับพารามิเตอร์ทั้งสี่ พารามิเตอร์ข้างต้น ซึ่งจะใช้ในการกำหนดเกณฑ์สำหรับการจำแนกประเภทของสิ่งแปลกปลอม วิธีการจำแนกประเภทนี้มักใช้ใน machine vision เพื่อระบุข้อบกพร่องของขวดแก้วในสายการผลิต

ขั้นตอนการตรวจจับสิ่งแปลกปลอมด้วยภาพพื้นหลังปกติและค่า Threshold ของข้อผิดพลาดของวิธีการ DDAET มีการคำนวณค่าความผิดพลาดของแต่ละพิกเซลระหว่างภาพพื้นหลังปกติ  $B$  และภาพขวดที่ผ่านการเข้ารหัสด้วย autoencoder ขั้นตอนนี้มีการวนลูปคำนวณค่าความผิดพลาดของแต่ละพิกเซล ดังนั้น ฟังก์ชันความซับซ้อนของขั้นตอนนี้คือ  $O(N)$  โดยที่  $N$  เป็นจำนวนพิกเซลในภาพพื้นหลังปกติ การตรวจจับสิ่งแปลกปลอมด้วยเกณฑ์การตัดสินใจ  $T$  จากการคำนวณหา

ค่าความผิดพลาด  $p_{ij}$  ของทุกพิกเซลระหว่างพิกเซลกับพิกเซลของภาพพื้นหลังปกติจากชุดข้อมูล เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ฝึกสอนขวดปกติ ขั้นตอนนี้มีการวนลูปเพื่อคำนวณค่าความผิดพลาดของทุกพิกเซล ดังนั้น ฟังก์ชันความซับซ้อนของขั้นตอนนี้คือ  $O(N)$  โดยกราฟจะเป็นเส้นตรงหรือ linear graph ซึ่งหมายความว่าเมื่อเพิ่มจำนวนพิกเซลในภาพพื้นหลังปกติอย่างเป็นเส้นตรง (เพิ่ม  $N$ ) ความซับซ้อนของอัลกอริทึมจะเพิ่มขึ้นเช่นกันเสมอแสดงในรูปที่ 3.16 จากกราฟแกนนอนแสดงค่าที่เพิ่มขึ้นของจำนวนพิกเซลของภาพพื้นหลังปกติ และแกนตั้งแสดงค่าความซับซ้อนของเวลา ช่วงของแกนตั้งและแกนนอนไม่ได้ใช้สัดส่วนเดียวกันกราฟจึงไม่เป็นเส้นตรง



รูปที่ 3.16 กราฟแสดงความซับซ้อนของขั้นตอนการตรวจจับสิ่งแปลกปลอมด้วยภาพพื้นหลังปกติและค่า Threshold ของข้อผิดพลาด

## บทที่ 4

### ผลการวิจัยและการอภิปรายผลการวิจัย

ในบทนี้จะนำเสนอผลการวิจัยและอภิปรายผลการวิจัยที่ได้จากการทดลองของอัลกอริทึมการตรวจจับสิ่งแปลกปลอมบนขวดแก้ว ที่ได้พัฒนาขึ้นทั้งสองวิธีคือวิธี CNNLSTM และ DDAET

#### 4.1 ชุดข้อมูลภาพสิ่งแปลกปลอมที่ใช้ในการฝึกสอนและทดสอบโมเดล

การทดลองในงานวิจัยนี้ใช้ชุดข้อมูลภาพขวดแก้วทั้งหมด 3 ชุด ได้แก่ ชุดข้อมูลภาพขวดบริเวณลำตัว (body dataset) ชุดข้อมูลภาพขวดบริเวณคอ (neck dataset) และ ชุดข้อมูลภาพขวดหมุน (rotating dataset) ชุดข้อมูลเหล่านี้ประกอบด้วยชนิดของตำหนิ 3 ชนิดหลัก ได้แก่ ตำหนิหิน (stone) ตำหนิรอยแตก (tear) และ ตำหนิรูพรุน (blister)

ชุดข้อมูลภาพขวดบริเวณลำตัวประกอบด้วยภาพขวดทั้งหมด 571 ภาพ ซึ่งประกอบด้วย 368 ภาพที่มีตำหนิและ 203 ภาพที่ไม่มีตำหนิ ชุดข้อมูลภาพขวดบริเวณคอประกอบด้วยภาพขวดจำนวน 570 ภาพ ซึ่งประกอบด้วย 250 ภาพที่มีตำหนิและ 320 ภาพที่ไม่มีตำหนิ ชุดข้อมูลภาพขวดหมุนประกอบด้วยภาพขวดจำนวน 120 ภาพ ซึ่งประกอบด้วย 70 ภาพที่มีตำหนิและ 50 ภาพที่ไม่มีตำหนิ ดังตารางที่ 4.1

ตารางที่ 4.1 จำนวนภาพที่ใช้ในการทดลองแยกประเภทตำหนิแต่ละประเภทในชุดข้อมูลบริเวณลำตัวคอ และขวดหมุน

ตำหนิ	ลำตัว	คอ	ขวดหมุน
หิน	207	151	30
ทางยาว	58	37	17
ฟองอากาศ	102	62	23
ขวดดี	203	320	50
รวม	571	570	120

#### 4.2 การตั้งค่าการทดลอง

การทดลองนี้ทำการวัดประสิทธิภาพโดยการทำ 5-fold cross validation เป็นเทคนิคที่ใช้ในการประเมินประสิทธิภาพของโมเดลของการเรียนรู้เครื่อง (machine learning models) โดยจะทำการแบ่งข้อมูลออกเป็น 5 กลุ่มเท่าๆ กันหรือที่เรียกว่า fold โดยแต่ละ fold จะถูกใช้เป็นชุดทดสอบ (testing set) ในครั้งหนึ่งของการทดลองและใช้ fold ที่เหลือเป็นข้อมูลฝึกสอนโมเดล (training set) ขั้นตอนการทำ 5-fold cross validation ดังนี้

1. แบ่งข้อมูลทั้งหมดออกเป็น 5 ส่วนย่อยโดยมีขนาดเท่าๆ กัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. เลือก fold แรกเป็นชุดข้อมูลทดสอบ และ fold ที่เหลือเป็นชุดข้อมูลฝึกสอน
3. ฝึกโมเดลบนชุดข้อมูลฝึกสอนและทดสอบโมเดลบนชุดข้อมูลทดสอบ
4. บันทึกค่าประสิทธิภาพของโมเดล
5. ทำขั้นตอนที่ 2-4 สำหรับ fold อื่นๆ ที่เหลือ
6. คำนวณค่าเฉลี่ยของประสิทธิภาพที่ได้จากการทดลองของทุก fold เพื่อประเมินประสิทธิภาพของโมเดล

การใช้งาน 5-fold cross validation นี้จะช่วยทำให้ได้รับประกันว่าการประเมินประสิทธิภาพของโมเดลนั้นมีความถูกต้องและมีความเที่ยงตรง เนื่องจากวิธีการนี้ใช้ข้อมูลทั้งหมดในการทดสอบและการฝึกสอนโมเดล ทำให้การประเมินประสิทธิภาพมีความเทียบเท่าและแท้จริงกับสภาพการใช้งานจริง

เพื่อประเมินประสิทธิภาพของวิธีการ CNNLSTM และ DDAET ที่ได้นำเสนอสำหรับการตรวจจับตำหนิบนขวดแก้ว ในการทดลองนี้ วิธีการ CNNLSTM นั้นใช้ transfer learning อัลกอริทึมเพื่อนำมาใช้ในการทดลองเปรียบเทียบประสิทธิภาพ โดยที่ เป็นเทคนิคสำคัญในการใช้งานโครงข่ายประสาทเชิงลึก Transfer learning เป็นกระบวนการที่นำโครงข่ายประสาทเชิงลึกที่ได้ถูกฝึกสอนด้วยข้อมูลมหาศาล ที่เรียกว่า pre-trained model มาใช้ในงานที่มีข้อมูลน้อยหรือในงานที่คล้ายคลึงกัน โดยใช้ pre-trained model ที่มีความสามารถในการจำแนกภาพที่ถูกฝึกสอนด้วยข้อมูลมหาศาล โมเดลสามารถนำมาใช้งานในงานอื่นๆ โดยทำการปรับแต่งหรือจัดการเล็กน้อยเพื่อให้เหมาะสมกับงานที่ต้องการ โมเดลเหล่านี้เมื่อนำมาใช้ต้องทำการ fine-tuning เป็นกระบวนการที่จะปรับแต่งหรือปรับค่าของ pre-trained model ที่ถูกนำมาใช้งานใหม่ เพื่อให้มีประสิทธิภาพและความแม่นยำสูง การปรับค่าสามารถทำได้โดยการเปลี่ยนแปลงหรือปรับค่าในชั้นบนสุดของโครงข่าย ซึ่งเป็นชั้นที่รับรู้ลักษณะเฉพาะของงานที่ต้องการ เช่น การปรับแต่งค่าของชั้น fully-connected layer หรือชั้น classifier โดยจะใช้เทคนิค backpropagation เพื่อปรับค่าและปรับปรุงน้ำหนักของชั้นต่างๆ

วิธีการตรวจจับสิ่งแปลกปลอมแบบดั้งเดิมที่ใช้เทคนิคการประมวลผลภาพที่นำมาใช้ในการทดลองเปรียบเทียบผลการทดลองของวิธีการที่นำเสนอคือวิธี image thresholding และ edge detection และวิธีการตรวจจับสิ่งแปลกปลอมที่ใช้โครงข่ายประสาทแบบลึกที่นำมาใช้ในการทดลองได้แก่ วิธี Anomaly Detection with Autoencoder (ADA) ResNet50 AlexNet VGG16 และ MobileNetV3

วิธีการ image thresholding เป็นเทคนิคที่ใช้กันอย่างแพร่หลายในการประมวลผลภาพโดยใช้ OpenCV ซึ่งเป็นการกำหนดค่าพิกเซลเป็น 0 หรือค่าสูงสุดตามค่าเกณฑ์ที่กำหนด (Rafae and Richard, 2007) การตรวจจับเส้นขอบ (edge detection) เป็นเทคนิคการประมวลผลภาพดิจิทัลที่ใช้ในการระบุและสกัดเส้นขอบในภาพ วิธีนี้ทำการสกัดเส้นขอบโดยการตรวจจับการเปลี่ยนแปลงความเข้มหรือสีระหว่างพิกเซลที่อยู่ติดกัน (Singh and Kaur) โมเดล ResNet50 เป็นโครงสร้างของ

เครือข่ายประสาทเชิงลึกที่มีการนำเสนอในปี 2016 โมเดล ResNet-50 ประกอบด้วยชั้นหน้าที่มีความลึก 50 ชั้นและใช้บล็อกที่เหลืออยู่ ซึ่งประกอบด้วยชั้นที่มีการคำนวณด้วยแบบลดรูปและการเชื่อมต่อทางลัด โมเดลนี้ได้รับการยอมรับอย่างกว้างขวางในการรู้จำภาพในงานต่างๆ โมเดล AlexNet เป็นโครงสร้างของเครือข่ายประสาทเชิงลึกที่ประกอบด้วย 8 ชั้น ซึ่งประกอบด้วย 5 ชั้นคอนโวลูชันและ 3 ชั้น fully connected AlexNet ใช้ ReLU เป็นฟังก์ชันกระตุ้นและการปรับปรุงการประสานกันโดยการใช้ dropout regularization โมเดล VGG16 เป็นโมเดลที่มีประสิทธิภาพที่ยอดเยี่ยมในงานทาง computer vision แบบต่างๆ โมเดล MobileNetV3 เป็นโมเดลที่มีความสมดุลระหว่างขนาดของโมเดลและความแม่นยำ ทำให้เหมาะสำหรับการใช้งานอย่างมีประสิทธิภาพบนอุปกรณ์ที่มีทรัพยากรการคำนวณจำกัด นอกเหนือจากโมเดลที่กล่าวถึงข้างต้นยังได้ทำการทดลองเปรียบเทียบกับวิธีการตรวจจับสิ่งแปลกปลอมที่ใช้เครือข่ายประสาทแบบ autoencoder ที่เรียกว่า ADA วิธีการนี้เป็นวิธีการที่มีการ encode และ decode ภาพ และใช้ภาพที่ผ่านการ decode มาคำนวณค่า threshold error ในกระบวนการตรวจจับสิ่งแปลกปลอม ADA จะเปรียบเทียบภาพที่ผ่านการ decode กับค่า threshold error หากมีค่า error เกินค่า threshold error แสดงว่าภาพนั้นเป็นภาพที่มีสิ่งแปลกปลอม

สำหรับการฝึกสอนวิธีการที่นำเสนอและโมเดลที่ใช้เทคนิค deep learning จำนวนรอบการฝึกสอน (epochs) ถูกกำหนดเป็น 500 สำหรับโมเดล VGG16, ResNet50, MobileNetV3 ADA และ DDAET ค่าเกณฑ์ความผิดพลาด  $\theta$  ของข้อมูลนำเข้า (input error threshold) สำหรับวิธีการ CNNLSTM ตั้งค่าไว้เป็น 0.2 และจำนวนโมเดล k ของ CNNLSTM ให้มีค่าเท่ากับ 3 ภาพการฝึกของเครือข่ายประสาท autoencoder ของ DDAET ใช้ภาพขจัดติในการฝึกสอน 10 ภาพ และทำ augment เพิ่มอีก 10 ภาพรวมเป็น 20 ภาพเพื่อทำการฝึกสอนสำหรับชุดข้อมูลทั้งสามชุด จำนวนภาพขจัดติที่นำมาใช้ในการสร้างภาพพื้นหลังปกตินั้นไม่จำเป็นต้องใช้ภาพในการสร้างภาพพื้นหลังเนื่องจาก ภาพขจัดเป็นภาพที่มีสภาพแวดล้อมที่ไม่ซับซ้อน ภาพขจัดปกตินั้นทั้งหมดจะถูกถ่ายในสภาพแวดล้อมเดียวกันมุมในการถ่ายเดียวกันอีกทั้งขจัด อีกทั้งยังมีรูปทรงเหมือนกันทุกขจัดอีกด้วย

การทดลองตรวจจับสิ่งแปลกปลอมบนขจัดแก้วสำหรับวิธีการ CNNLSTM DDAET และวิธีการที่นำมาใช้ในการทดลองเปรียบเทียบประสิทธิภาพทั้งหมดได้ทำการทดลองบนคอมพิวเตอร์ส่วนบุคคลที่มีซีพียูรุ่น AMD Ryzen 7 5000 และการ์ดจอ NVIDIA GTX 1650 เพื่อให้การคำนวณและการวิเคราะห์ผลลัพธ์เป็นไปอย่างมีประสิทธิภาพ

### 4.3 การประเมินผลการทดลอง

ในการประเมินประสิทธิภาพของวิธีการตรวจจับสิ่งแปลกปลอมบนขจัดแก้วที่นำเสนอ ประเมินโดยใช้เมตริกต่างๆ ได้แก่ Accuracy Recall Precision และค่า F1-score โดยที่ Accuracy Recall Precision และค่า F1-score คำนวณได้ดังสมการที่ 4.1- 4.4 ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (4.1)$$

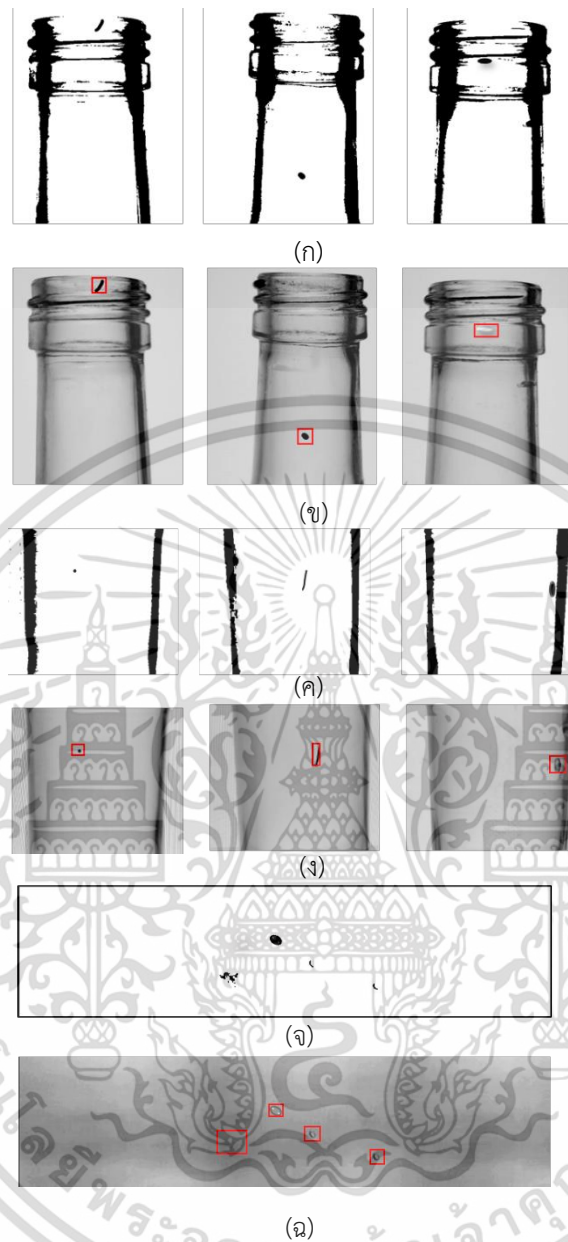
$$\text{Recall} = \frac{TP}{TP + FN} \times 100 \quad (4.2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100 \quad (4.3)$$

$$F_1 - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \quad (4.4)$$

Accuracy เป็นค่าที่ใช้ในการวัดประสิทธิภาพของอัลกอริทึมการจำแนกประเภทข้อมูลทั้งหมด ค่า Accuracy คำนวณได้จากอัตราส่วนระหว่างจำนวนของตัวอย่างที่ถูกทำนายถูกต้อง (ตรงกับคลาสดำเนิน) กับจำนวนของตัวอย่างทั้งหมดในชุดข้อมูลที่ใช้ทดสอบหรือทดลอง เป็นการวัดประสิทธิภาพหรือความถูกต้องของโมเดลในการทำนายหรือจำแนกประเภทของข้อมูล Recall เป็นค่าที่ใช้วัดประสิทธิภาพของอัลกอริทึมการจำแนกประเภทในการตรวจจับภาพโดยเฉพาะตัวอย่างภาพขจัดที่มีตำหนิ (Positive Class) โดยการคำนวณจำนวนภาพที่อัลกอริทึมจำแนกประเภทถูกต้องว่าเป็นภาพขจัดที่มีตำหนิ (TP) หารด้วยจำนวนทั้งหมดของตัวอย่างภาพขจัดที่มีตำหนิจริงทั้งหมด (TP + FN) ซึ่งแสดงถึงความสามารถของระบบในการจับสิ่งแปลกปลอมอย่างมีประสิทธิภาพ ค่า Precision คือการวัดความแม่นยำของการทำนายในการจำแนกประเภทข้อมูลในคลาสนี้ที่เป็นขจัดที่มีตำหนิ (Positive Class) โดยค่าความเที่ยงตรงจะวัดด้วยสัดส่วนของจำนวนภาพขจัดที่ถูกต้องว่าเป็นภาพขจัดที่มีตำหนิ (TP) หารด้วยผลรวมของจำนวนภาพขจัดที่โมเดลจำแนกประเภทได้ถูกต้องว่าเป็นขจัดมีตำหนิ (TP) และจำนวนภาพที่โมเดลทำนายว่าเป็นภาพขจัดมีตำหนิ (Positive Class) แต่ในความเป็นจริงแล้วเป็นภาพขจัดที่ไม่มีตำหนิ (FP) ค่าความเที่ยงตรงจะเป็นตัวชี้วัดให้ทราบถึงความแม่นยำของอัลกอริทึมในการตรวจจับและจำแนกประเภทข้อมูลในคลาสนี้ที่เป็นภาพขจัดมีตำหนิว่ามีความแม่นยำเพียงใด F1-score ค่าที่แสดงประสิทธิภาพการจำแนกประเภทของอัลกอริทึม โดยการนำค่าความเที่ยงตรง และค่า Recall มาคำนวณหาค่าเฉลี่ย หรือเรียกว่า Harmonic Mean ซึ่ง Model ที่มีค่า F1-score สูงจะถือว่าโมเดลนั้นมีประสิทธิภาพในการตรวจจับสิ่งแปลกปลอมได้ดี ค่า F1-score ในการตรวจจับสิ่งแปลกปลอมของแต่ละวิธีการสามารถใช้เป็นตัวชี้วัดเพื่อเปรียบเทียบประสิทธิภาพระหว่างวิธีการ และมีการพิจารณาค่าความระลึกและค่า Precision ร่วมกันเพื่อให้ได้ภาพรวมของประสิทธิภาพการทำงานของวิธีการทั้งหมดในชุดข้อมูลนั้นๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.1 ตัวอย่างผลการจำแนกประเภทของ DDAET (ก) พื้นที่ที่มีสิ่งแปลกปลอมคอขวด (ข) ผลการตรวจจับสิ่งแปลกปลอมคอขวด (ค) พื้นที่ที่มีสิ่งแปลกปลอมบริเวณลำตัว (ง) ผลการตรวจจับสิ่งแปลกปลอมบริเวณลำตัวของขวด (จ) พื้นที่ที่มีสิ่งแปลกปลอมขวดหมุน (ฉ) ผลการตรวจจับสิ่งแปลกปลอมขวดหมุน

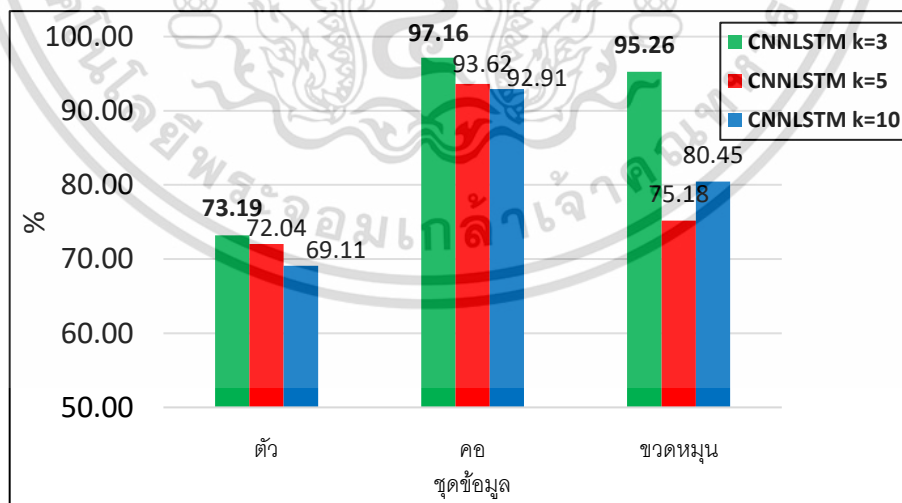
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.4 ผลการทดลอง

งานวิจัยนี้ผลการทดลองจะถูกแบ่งออกเป็นสองส่วนย่อยคือ ผลการประเมินประสิทธิภาพการตรวจจับสิ่งแปลกปลอมบนขวดแก้วด้วยวิธีการ CNNLSTM และ DDAET โดยเปรียบเทียบกับวิธีการอื่นๆ ที่กล่าวในหัวข้อ 4.2 และเปรียบเทียบเวลาที่ใช้ในการฝึกสอนโมเดล

รูปที่ 4.1 แสดงตัวอย่างผลการตรวจจับสิ่งแปลกปลอมของวิธีการ DDAET สำหรับชุดข้อมูลภาพของบริเวณลำตัวของขวด คอขวด และภาพขวดหมุน รูปที่ 4.1 (ก) (ค) และ(จ) พิกเซลสีดำแสดงพื้นที่เน้นให้เห็นตำหนิที่ตรวจพบของลำตัว คอ และขวดหมุนตามลำดับ พิกเซลสีดำเหล่านี้แสดงถึงข้อผิดพลาดที่มีค่าเกินค่าเกณฑ์ที่ระบุไว้ พื้นที่สีดำเหล่านี้จะถูกนำไปจำแนกประเภทว่าเป็นชอบหรือไม่ ถ้าไม่ใช่ชอบจะถูกจำแนกประเภทว่าเป็นตำหนิในชนิดต่างๆ ตามค่าพารามิเตอร์ที่ผู้ใช้กำหนด รูปที่ 4.1 (ข) (ง) และ(ฉ) นำเสนอตัวอย่างของผลลัพธ์การตรวจจับตำหนิสำหรับภาพของส่วนลำตัว คอ และขวดหมุนตามลำดับ ผลลัพธ์เหล่านี้ถูกจำแนกตามพารามิเตอร์เกณฑ์ที่ใช้สำหรับการตรวจจับข้อบกพร่อง

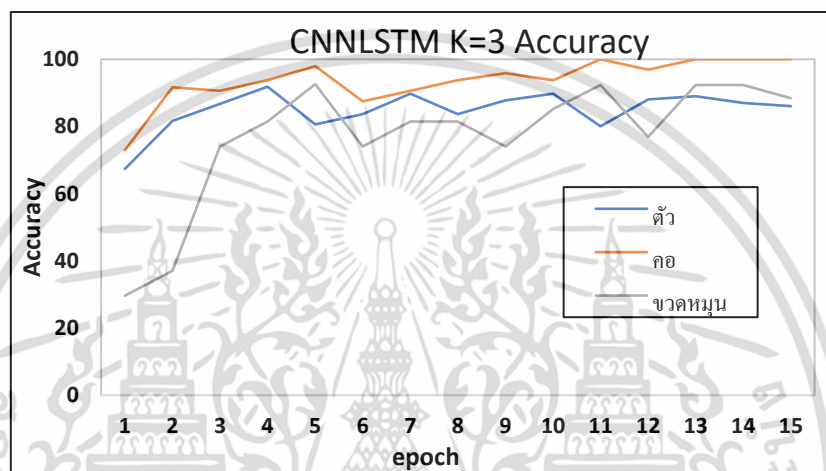
รูปที่ 4.2 แสดงการเปรียบเทียบค่าความถูกต้องของวิธี CNNLSTM สำหรับการตรวจจับสิ่งแปลกปลอมชุดข้อมูลลำตัวขวด คอ และขวดหมุน วิธี CNNLSTM ที่มีการกำหนดค่า  $K$  เท่ากับ 3 มีความถูกต้องในการตรวจจับสิ่งแปลกปลอม สำหรับส่วนลำตัวขวด คอ และ ขวดหมุน มากกว่า CNNLSTM ที่มีการกำหนดค่า  $K$  เป็น 5 และ 10 ทุกชุดข้อมูล จากรูปแสดงให้เห็นว่าวิธีการ CNNLSTM ที่กำหนดให้มีค่า  $K$  เพิ่มขึ้นไม่ทำให้วิธีการนี้มีประสิทธิภาพการจำแนกประเภทภาพขวดที่มีสิ่งแปลกปลอมเพิ่มมากขึ้นตามไปด้วย ในงานวิจัยนี้จึงกำหนดค่า  $K$  สำหรับวิธีการ CNNLSTM มีค่าเท่ากับ 3 เพื่อใช้ในการทดลอง



รูปที่ 4.2 ผลการทดลองเปรียบเทียบความถูกต้อง CNNLSTM ที่มีค่า K ต่างกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4.3 แสดง Learning Curve ความเปลี่ยนแปลงของประสิทธิภาพการเรียนรู้ของโมเดล CNNLSTM ที่มีค่า  $K = 3$  ตามจำนวนรอบการฝึกสอน (epochs) ในแต่ละรอบฝึกสอน โดยประสิทธิภาพการเรียนรู้จะวัดโดยค่า Accuracy ที่ได้จากการตรวจหาสิ่งแปลกปลอมของข้อมูลทดสอบทั้งสามชุดข้อมูล รูปนี้แสดงการเปลี่ยนแปลงของประสิทธิภาพการเรียนรู้ของโมเดลเมื่อทำการฝึกสอนในช่วงระยะเวลาการฝึกสอน ตั้งแต่เริ่มต้นทำการฝึกสอนไปจนถึงจุดที่ประสิทธิภาพการเรียนรู้ค่อนข้างคงที่หรือไม่เปลี่ยนแปลงมากขึ้น



รูปที่ 4.3 Learning Curve ประสิทธิภาพการเรียนรู้ของ CNNLSTM

ตารางที่ 4.2 ถึง 4.4 แสดงผลการทดลองเปรียบเทียบการตรวจจับสิ่งแปลกปลอมของชุดข้อมูลบริเวณลำตัวของขวด ชุดข้อมูลคอขวด และชุดข้อมูลขวดหมุนตามลำดับ ตารางแสดงค่า Accuracy Recall Precision และ F1-score ของอัลกอริทึมต่างๆที่ใช้ในการทดลองนี้ ค่าเหล่านี้แสดงให้เห็นถึงประสิทธิภาพของแต่ละอัลกอริทึมในการตรวจจับสิ่งแปลกปลอมบนขวดแก้ว แต่ละคอลัมน์มีการคำนวณค่าส่วนเบี่ยงเบนมาตรฐานแสดงหลังสัญลักษณ์  $\pm$  และคำนวณหาค่า t-Test ที่มีค่าความเชื่อมั่นเท่ากับ 95% ( $\alpha = 0.05$ ) โดยที่สัญลักษณ์ \* แสดงว่าอัลกอริทึม DDAET มีความถูกต้องในการตรวจจับสิ่งแปลกปลอมสูงกว่าวิธีการอื่นๆอย่างมีนัยสำคัญ สัญลักษณ์ # แสดงอัลกอริทึม CNNLSTM มีความถูกต้องในการตรวจจับสิ่งแปลกปลอมสูงกว่าวิธีการอื่นๆอย่างมีนัยสำคัญ และสัญลักษณ์ + แสดงวิธีการที่มีค่าความถูกต้องในการตรวจจับสิ่งแปลกปลอมมากที่สุดนั้นมีค่าสูงกว่าวิธีการ CNNLSTM และ DDAET อย่างมีนัยสำคัญ

ตารางที่ 4.2 ผลการทดลองเปรียบเทียบของชุดข้อมูลบริเวณลำตัวของขวดด้วยค่า Accuracy Recall Precision และ F1-score

วิธีการ	Accuracy	Recall	Precision	F1-score
DDAET	80.57±0.90	71.94±0.45 <sup>+</sup>	74.78±1.08 <sup>+</sup>	73.49±1.1 <sup>+</sup>
CNNLSTM	74.61±2.14 <sup>*+</sup>	82.06±1.73	74.08±2.22 <sup>+</sup>	74.55±2.2 <sup>+</sup>
ResNet50	79.84±1.39	71.91±1.48 <sup>#</sup>	<b>88.83±2.39</b>	74.26±1.0
AlexNet	75.28±3.23 <sup>*</sup>	73.95±3.27 <sup>#</sup>	76.29±2.495	71.88±2.7
VGG16	<b>81.48±2.04</b>	<b>84.81±3.65</b>	78.92±2.75	<b>78.32±1.8</b>
MobileNetV3	71.29±1.86 <sup>*#</sup>	58.19±2.03 <sup>*#</sup>	58.31±2.94 <sup>*#</sup>	58.35±2.2 <sup>*#</sup>
ADA	75.20±3.11 <sup>*</sup>	52.54±2.88 <sup>*#</sup>	57.18±2.59 <sup>*#</sup>	55.24±2.8 <sup>*#</sup>
Image Threshold	63.52±2.41 <sup>*#</sup>	51.87±2.27 <sup>*#</sup>	54.44±2.37 <sup>*#</sup>	51.00±0.0 <sup>*#</sup>
Edge detection	71.14±4.80 <sup>*</sup>	78.87±4.82	72.74±4.76	71.80±10.1

ตาราง 4.2 แสดงผลการทดลองเปรียบเทียบการตรวจจับสิ่งแปลกปลอมบริเวณลำตัวของขวด ตารางนี้แสดงค่า Accuracy Recall Precision และ F1-score ของอัลกอริทึมต่างๆที่ได้ทำการตรวจจับสิ่งแปลกปลอมบนชุดข้อมูลลำตัวขวด ค่า Accuracy ของแต่ละอัลกอริทึมในตารางนี้เป็น การบอกถึงประสิทธิภาพของอัลกอริทึมดังกล่าวในการตรวจจับสิ่งแปลกปลอมบนขวดดังแสดงใน คอลัมน์ Accuracy จากตาราง VGG16 เป็นอัลกอริทึมที่มีค่า Accuracy จากการตรวจจับสิ่งแปลกปลอมสูงที่สุดในชุดข้อมูลนี้เมื่อเทียบกับวิธีอื่นๆ DDAET เป็นอัลกอริทึมที่มีความถูกต้องในการ ตรวจจับที่สูงใกล้เคียงกับ VGG16 และสูงกว่าวิธีการตรวจจับสิ่งแปลกปลอมส่วนใหญ่ ResNet50 ยัง เป็นอีกหนึ่งอัลกอริทึมที่มีค่า Accuracy ในการตรวจจับที่สูงเมื่อเทียบกับอัลกอริทึมที่มีค่า Accuracy สูงที่สุด AlexNet ADA และ CNNLSTM มีความถูกต้องปานกลางในการตรวจจับสิ่งแปลกปลอมของ ชุดข้อมูลนี้ DDAET มีค่า Accuracy ที่สูงกว่าทั้งสามวิธีการนี้อย่างมีนัยสำคัญ MobileNetV3 เป็น อัลกอริทึมที่ใช้โครงข่ายประสาทเชิงลึกที่มีค่า Accuracy น้อยที่สุด และน้อยกว่าอัลกอริทึมที่นำเสนอ อย่างมีนัยสำคัญทั้งสองอัลกอริทึม Edge detection เป็นอัลกอริทึมที่มีค่า Accuracy น้อยเมื่อเทียบ กับ DDAET แล้วอัลกอริทึมนี้มีค่า Accuracy ที่น้อยกว่าอย่างมีนัยสำคัญ Image Threshold เป็น อัลกอริทึมที่มีความถูกต้องต่ำกว่าอัลกอริทึมอื่น ๆ ในการทดลองของชุดข้อมูลนี้ ทำให้อัลกอริทึมนี้มี ค่า Accuracy ที่น้อยกว่าวิธีการที่นำเสนออย่างมีนัยสำคัญทั้งสองวิธีการอีกด้วย

คอลัมน์ที่สามแสดงการเปรียบเทียบค่า Recall โดยที่ VGG16 มีค่า Recall สูงที่สุดสำหรับ การตรวจจับสิ่งแปลกปลอมในชุดข้อมูลนี้ และยิ่งสูงกว่า DDAET อย่างมีนัยสำคัญ CNNLSTM เป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อัลกอริทึมที่มีค่า Recall ที่สูงเมื่อเทียบกับวิธีการอื่นๆ แต่น้อยกว่า VGG16 เล็กน้อยเมื่อเปรียบเทียบกันแล้วไม่ได้น้อยกว่าอย่างมีนัยสำคัญ Edge detection เป็นอัลกอริทึมประมวลผลภาพที่มีค่า Recall ที่สูงเมื่อเทียบกับอัลกอริทึมอื่นๆแต่มีค่าความแปรปรวนสูงตามไปด้วย AlexNet CNNLSTM-DDAET และ ResNet50 เป็นอัลกอริทึมที่ใช้โครงข่ายประสาทเชิงลึกที่มีค่า Recall ปานกลางตามลำดับเมื่อเทียบกับอัลกอริทึมที่มีค่าสูงที่สุด และ AlexNet และ ResNet50 มีค่า Recall น้อยกว่า CNNLSTM ที่นำเสนออย่างมีนัยสำคัญ MobileNetV3 ADA และ Image threshold เป็นอัลกอริทึมที่มีค่า Recall น้อยที่สุดตามลำดับ และทั้งสามวิธีการมีค่าน้อยกว่าอัลกอริทึมที่นำเสนออย่างมีนัยสำคัญทั้งสองอัลกอริทึม

ค่า Precision ของแต่ละอัลกอริทึมจากผลการทดลองภาพถ่ายบริเวณลำตัวขวดแสดงในคอลัมน์ที่สี่ จากการทดลอง Resnet50 เป็นอัลกอริทึมที่มีค่า Precision มากที่สุดในชุดข้อมูลนี้ และยังมีค่า Precision ที่มากกว่า DDAET และ CNNLSTM อย่างมีนัยสำคัญ VGG16 AlexNet DDAET และ CNNLSTM เป็นอัลกอริทึมที่มีค่า Precision ที่สูงตามลำดับ Edge detection เป็นอัลกอริทึมการประมวลผลภาพที่มีค่า Precision สูงเมื่อเทียบกับ MobileNetV3 และ ADA ในขณะที่ MobileNetV3 ADA และ Image threshold เป็นอัลกอริทึมที่มีค่า Precision น้อยที่สุดการทดลองนี้ ทั้งสามอัลกอริทึมมีค่า Precision น้อยกว่าอัลกอริทึมที่นำเสนออย่างมีนัยสำคัญทั้งสองอัลกอริทึม

การเปรียบเทียบค่า F1-score จากผลการทดลองของทุกอัลกอริทึมในการทดลองภาพถ่ายบริเวณลำตัวขวดแสดงในคอลัมน์ที่ห้า จากการทดลอง VGG16 ยังคงเป็นอัลกอริทึมที่มีค่า F1-score สูงที่สุด และมากกว่าวิธีการที่นำเสนออย่างมีนัยสำคัญ CNNLSTM ResNet50 DDAET AlexNet และ Edge detection มีค่า F1-score สูงตามลำดับ MobileNetV3 และ ADA เป็นอัลกอริทึมที่มีค่า F1-score น้อย และมี Image Threshold เป็นอัลกอริทึมที่มีค่า F1-score น้อยที่สุดในชุดข้อมูลภาพถ่ายบริเวณลำตัวของขวด ทั้งสามวิธีการนี้มีค่า F1-score น้อยกว่าอัลกอริทึมที่นำเสนอทั้งสองอันกอริทึมอย่างมีนัยสำคัญ

จากผลการทดลองสามารถสรุปได้ว่า VGG16 DDAET และ ResNet50 เหมาะสมและมีประสิทธิภาพในการตรวจจับสิ่งแปลกปลอมเมื่อพิจารณาจากทุกค่าที่ใช้ในการเปรียบเทียบผลการทดลอง ในขณะที่ ADA MobileNetV3 และ Edge detection อัลกอริทึมมีประสิทธิภาพในการตรวจจับสิ่งแปลกปลอมที่ค่อนข้างต่ำ Image Threshold อัลกอริทึมมีประสิทธิภาพต่ำที่สุดในการตรวจจับสิ่งแปลกปลอมบริเวณลำตัวของขวด

ตารางที่ 4.3 ผลการทดลองเปรียบเทียบของชุดข้อมูลบริเวณคอขวดด้วยค่า Accuracy Recall Precision และ F1-score

วิธีการ	Accuracy	Recall	Precision	F1-score
DDAET	92.43±1.58	94.04±2.45	91.86±2.45 <sup>+</sup>	92.26±1.96
CNNLSTM	89.82±0.96 <sup>*</sup>	85.86±0.86 <sup>*</sup>	93.74±1.12	82.41±2.04 <sup>*</sup>
ResNet50	80.00±1.87 <sup>*#</sup>	73.28±4.78 <sup>*#</sup>	88.88±1.92 <sup>*#</sup>	74.10±1.10 <sup>*#</sup>
AlexNet	92.32±1.40	89.59±0.95 <sup>*</sup>	95.00±1.82	91.76±1.66
VGG16	84.84±2.26 <sup>*#</sup>	79.72±0.88 <sup>*#</sup>	90.34±0.66 <sup>#</sup>	81.96±1.08 <sup>*</sup>
MobileNetV3	72.44±0.94 <sup>*#</sup>	63.05±1.22 <sup>*#</sup>	80.88±1.45 <sup>*#</sup>	64.13±1.37 <sup>*#</sup>
ADA	76.42±1.28 <sup>*#</sup>	71.53±1.59 <sup>*#</sup>	75.06±2.30 <sup>*#</sup>	72.65±1.32 <sup>*#</sup>
Image Threshold	66.22±1.90 <sup>*#</sup>	52.56±2.22 <sup>*#</sup>	53.72±1.69 <sup>*#</sup>	52.29±1.52 <sup>*#</sup>
Edge detection	72.24±2.19 <sup>*#</sup>	65.57±2.12 <sup>*#</sup>	71.36±2.21 <sup>*#</sup>	66.90±2.08 <sup>*#</sup>

ผลการทดลองเปรียบเทียบค่า Accuracy Recall Precision และ F1-score ของการตรวจจับสิ่งแปลกปลอมบริเวณคอขวดแสดงดังแสดงในตาราง 4.3 คอลัมน์ที่สองแสดงค่า Accuracy ของการตรวจจับสิ่งแปลกปลอมของแต่ละอัลกอริทึม โดยที่ DDAET เป็นอัลกอริทึมมีค่า Accuracy มากสุดในชุดข้อมูลนี้และมีส่วนเบี่ยงเบนมาตรฐานที่ต่ำ วิธี AlexNet มีค่า Accuracy ที่สูงใกล้เคียงกันและยังสูงกว่าอัลกอริทึมส่วนใหญ่ของวิธีการตรวจจับสิ่งแปลกปลอมอื่นๆ CNNLSTM มีผลการทดลองค่า Accuracy ที่สูงและมีความแปรปรวนที่น้อยเมื่อเทียบกับวิธีการอื่นๆ แต่น้อยกว่า DDAET อย่างมีนัยสำคัญ VGG16 และ ResNet50 มีค่า Accuracy ปานกลางเมื่อเทียบกับอัลกอริทึมที่มีความถูกต้องสูงสุด และที่ค่า Accuracy น้อยกว่าอัลกอริทึมที่นำเสนออย่างมีนัยสำคัญทั้งสอง อัลกอริทึม ADA MobileNetV3 และ Edge detection อัลกอริทึมเหล่านี้มีค่า Accuracy ในการตรวจจับสิ่งแปลกปลอมที่ต่ำในการทดลองชุดข้อมูลนี้ และต่ำกว่าอัลกอริทึมที่นำเสนออย่างมีนัยสำคัญ โดยที่ Image Threshold เป็นอัลกอริทึมที่มีประสิทธิภาพต่ำสุดและมีความแปรปรวนที่สูงเมื่อเทียบกับวิธีการอื่นๆ

สำหรับค่า Recall ของการทดลองในชุดข้อมูลนี้แสดงในคอลัมน์ที่สาม DDAET อัลกอริทึมมีค่า Recall ที่มากที่สุดในการตรวจจับสิ่งแปลกปลอมในชุดข้อมูลนี้ และมี AlexNet มีค่า Recall ที่สูงที่สุดใกล้เคียงกันเมื่อเทียบกับอัลกอริทึมอื่นๆ CNNLSTM มีค่า Recall ที่สูงเมื่อเทียบกับอัลกอริทึมอื่นๆ แต่น้อย DDAET อย่างมีนัยสำคัญ VGG16 ResNet50 และ ADA มีค่า Recall อยู่ในระดับปานกลางสำหรับชุดข้อมูลนี้ และน้อยกว่าวิธีการที่นำเสนอทั้งสองวิธีการอย่างมีนัยสำคัญ Edge

detection และ MobileNetV3 มีค่า Recall ที่ต่ำ โดยที่ Image Threshold มีค่า Recall ต่ำที่สุด อัลกอริทึมเหล่านี้ยังคงมีค่า Recall ที่ต่ำกว่าวิธีการที่นำเสนออย่างมีนัยสำคัญอีกด้วย

คอลัมน์ที่สี่แสดงการเปรียบเทียบค่า Precision ของแต่ละอัลกอริทึมในการทดลองชุด ข้อมูลภาพคอขวด โดยมี AlexNet เป็นอัลกอริทึมที่มีค่า Precision สูงที่สุด และมีค่า Precision ที่สูงกว่า DDAET ที่นำเสนออย่างมีนัยสำคัญ CNNLSTM DDAET VGG16 และ ResNet50 อัลกอริทึมเหล่านี้มีค่า Precision ที่สูงเมื่อเทียบกับอัลกอริทึมอื่นๆ โดยที่ CNNLSTM มีค่าน้อยกว่า AlexNet เล็กน้อยไม่ต่ำกว่าอย่างมีนัยสำคัญ ขณะที่ VGG16 มีค่า Precision ที่สูงแต่ก็ยังน้อยกว่า CNNLSTM อย่างมีนัยสำคัญด้วย ResNet50 ก็มีค่า Precision ที่น้อยกว่าอัลกอริทึมที่นำเสนออย่างมีนัยสำคัญทั้งสองอัลกอริทึม MobileNetV3 ADA และ Edge detection มีค่า Recall อยู่ในระดับปานกลาง ทั้งสามวิธีการนี้มีค่า Precision ที่น้อยกว่าวิธีการที่นำเสนออย่างมีนัยสำคัญทั้งสองวิธีการ โดยที่วิธี Image Threshold ยังคงมีค่า Precision น้อยที่สุด และน้อยกว่าอย่างมีนัยสำคัญ

การเปรียบเทียบค่า F1-score ของทุกอัลกอริทึมในการทดลองนี้แสดงในคอลัมน์ที่ห้า จากการทดลอง DDAET เป็นอัลกอริทึมที่มีค่า F1-score สูงที่สุดในการตรวจจับสิ่งแปลกปลอม AlexNet มีค่า F1-score สูงมากกว่าอัลกอริทึมอื่นๆ CNNLSTM และ VGG16 มีค่า F1-score ที่สูง โดยที่ทั้งสองวิธีการนี้มีค่า F1-score น้อยกว่า DDAET อย่างมีนัยสำคัญ ResNet50 และ ADA มีค่า F1-score อยู่ในระดับปานกลาง และน้อยกว่าอัลกอริทึมที่นำเสนออย่างมีนัยสำคัญทั้งสองอัลกอริทึม Edge detection และ MobileNetV3 เป็นอัลกอริทึมที่มีค่า F1-score น้อย และ Image Threshold มีค่า F1-score น้อยที่สุดในชุดข้อมูลภาพถ่ายบริเวณคอของขวด ทำให้วิธีการเหล่านี้มีค่า F1-score น้อยกว่าอัลกอริทึมที่นำเสนออย่างมีนัยสำคัญ

จากตารางเปรียบเทียบผลการทดลองแสดงให้เห็นว่า DDAET เป็นอัลกอริทึมที่เหมาะสมที่สุดในการตรวจจับสิ่งแปลกปลอมบริเวณคอขวด โดยพิจารณาจากค่า Accuracy Recall และ F1-score สูงที่สุดในชุดข้อมูล AlexNet เหมาะสมและมีประสิทธิภาพในการตรวจจับสิ่งแปลกปลอม อัลกอริทึมนี้มีค่า Precision มากที่สุดแสดงถึงผลลัพธ์ที่มีความแม่นยำในการตรวจจับสิ่งแปลกปลอม ในขณะที่ Image Threshold, Edge detection, และ MobileNetV3 ไม่เหมาะสมและมีประสิทธิภาพต่ำในการตรวจจับสิ่งแปลกปลอมในชุดข้อมูลนี้ อัลกอริทึมมีประสิทธิภาพต่ำที่สุดสำหรับการตรวจจับสิ่งแปลกปลอมบริเวณคอขวด

**ตารางที่ 4.4** ผลการทดลองเปรียบเทียบประสิทธิภาพของชุดข้อมูลภาพขวดหมึกด้วยค่า Accuracy Recall Precision และ F1-score

วิธีการ	Accuracy	Recall	Precision	F1-score
DDAET	95.43±1.12 <sup>#</sup>	93.84±2.19 <sup>#</sup>	94.22±1.00 <sup>#</sup>	92.62±1.33 <sup>#</sup>
CNNLSTM	<b>97.66±1.21</b>	<b>97.52±0.45</b>	<b>97.17±0.60</b>	<b>96.60±0.64</b>
ResNet50	60.42±0.80 <sup>*#</sup>	55.44±0.95 <sup>*#</sup>	59.10±0.46 <sup>*#</sup>	52.82±1.09 <sup>*#</sup>
AlexNet	60.44±1.07 <sup>*#</sup>	52.61±0.41 <sup>*#</sup>	79.47±0.39 <sup>*#</sup>	42.83±0.39 <sup>*#</sup>
VGG16	95.63±0.53 <sup>#</sup>	95.57±1.41 <sup>*#</sup>	94.74±1.00 <sup>#</sup>	95.60±0.85 <sup>#</sup>
MobileNetV3	70.82±1.52 <sup>*#</sup>	62.206±1.67 <sup>*#</sup>	79.32±1.11 <sup>*#</sup>	52.85±1.75 <sup>*#</sup>
ADA	61.19±0.92 <sup>*#</sup>	58.836±0.79 <sup>*#</sup>	55.30±1.17 <sup>*#</sup>	56.30±0.20 <sup>*#</sup>
Image Threshold	65.70±2.09 <sup>*#</sup>	77.612±1.44 <sup>*#</sup>	69.68±0.62 <sup>*#</sup>	63.44±0.61 <sup>*#</sup>
Edge detection	63.222±2.05 <sup>*#</sup>	75.622±1.59 <sup>*#</sup>	68.31±1.02 <sup>*#</sup>	60.84±1.07 <sup>*#</sup>

ผลการทดลองเปรียบเทียบประสิทธิภาพการตรวจจับสิ่งแปลกปลอมด้วยค่า Accuracy Recall Precision และ F1-score ของอัลกอริทึมต่างๆโดยใช้ชุดข้อมูลภาพขวดหมึกแสดงในตารางที่ 4.4 ผลการทดลองเปรียบเทียบประสิทธิภาพด้วยค่า Accuracy ของการตรวจจับสิ่งแปลกปลอมของชุดข้อมูลภาพขวดหมึกแสดงให้เห็นถึงความสามารถในการตรวจจับสิ่งแปลกปลอมของแต่ละอัลกอริทึม จากตารางผลการทดลองเปรียบเทียบค่า Accuracy พบว่า CNNLSTM มีค่า Accuracy สูงที่สุด แสดงให้เห็นถึงความถูกต้องในการตรวจจับสิ่งแปลกปลอมในชุดข้อมูลภาพขวดหมึก และยังมีค่า Accuracy มากกว่าอัลกอริทึมอื่นๆที่ใช้ในการเปรียบเทียบผลการทดลองอย่างมีนัยสำคัญทุกอัลกอริทึม VGG16 และ DDAET มีค่า Accuracy ที่สูงเมื่อเทียบกับอัลกอริทึมอื่นๆอย่างมีนัยสำคัญ แต่น้อยกว่า CNNLSTM ที่เป็นอัลกอริทึมที่มีค่า Accuracy มากที่สุดอย่างมีนัยสำคัญเช่นกัน MobileNetV3 มีค่าความถูกต้องอยู่ในระดับปานกลาง และมีค่าความถูกต้องน้อยกว่าอัลกอริทึมที่นำเสนออย่างมีนัยสำคัญทั้งสองอัลกอริทึม Image Threshold และ Edge detection เป็นอัลกอริทึมการประมวลผลภาพที่มีค่า Accuracy ต่ำและมีค่าความแปรปรวนของผลการทดลองที่สูง แต่ก็ยังมีค่า Accuracy ที่สูงกว่า ADA AlexNet และ ResNet50 ที่เป็นอัลกอริทึมที่ใช้โครงข่ายประสาทเชิงลึก และยังต่ำกว่าวิธีการที่นำเสนออย่างมีนัยสำคัญอีกด้วย

คอลัมน์ที่สามแสดงค่า Recall ของแต่ละวิธีการ ผลการทดลองแสดง CNNLSTM มีค่า Recall สูงที่สุด เมื่อเทียบกับอัลกอริทึมอื่นๆแล้ว CNNLSTM มีค่า Accuracy ที่มากกว่า อย่างมีนัยสำคัญทุกอัลกอริทึมในการทดลองนี้ VGG16 และ DDAET เป็นอัลกอริทึมเหล่านี้ค่า Recall ที่สูงเมื่อเทียบกับอัลกอริทึมอื่นๆ วิธี Image threshold และ Edge detection เป็นวิธีการที่มีค่า Recall

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปานกลางและมากกว่าอัลกอริทึมที่ใช้โครงข่ายประสาทแบบลึกบางวิธีการ แต่ยังมีค่า Recall ที่น้อยกว่าอัลกอริทึมที่นำเสนออย่างมีนัยสำคัญ MobileNetV3 ADA ResNet50 อัลกอริทึมเหล่านี้มีค่า Recall น้อยเมื่อเทียบกับอัลกอริทึมอื่นๆ โดยมี AlexNet เป็นอัลกอริทึมที่มีค่า Recall น้อยที่สุดในชุดข้อมูลนี้

คอลัมน์ที่สี่แสดงค่า Precision ของทุกอัลกอริทึมการตรวจจับสิ่งแปลกปลอม CNNLSTM อัลกอริทึมที่มีค่า Precision สูงที่สุดทำให้ CNNLSTM เป็นอัลกอริทึมที่มีค่า Precision มากกว่าอัลกอริทึมอื่นๆ อย่างมีนัยสำคัญทุกอัลกอริทึมในการทดลองนี้ VGG16 และ DDAET มีค่า Precision สูงกว่าอัลกอริทึมอื่นๆ AlexNet และ MobileNetV3 มีค่า Precision อยู่ในระดับปานกลาง และน้อยกว่าวิธีการทำนำเสนออย่างมีนัยสำคัญ โดยมี Image Threshold และ Edge detection มีค่า Precision ที่ต่ำ ResNet50 และ ADA มีค่า Precision ในการตรวจจับสิ่งแปลกปลอมต่ำสุดในชุดข้อมูลนี้

คอลัมน์ที่ห้าแสดงการเปรียบเทียบค่า F1-score จากผลการทดลอง CNNLSTM ยังคงเป็นอัลกอริทึมที่มีค่า F1 ที่มากที่สุดในชุดข้อมูลนี้ และมีค่า F1-score มากกว่าอัลกอริทึมอื่นๆ อย่างมีนัยสำคัญทุกอัลกอริทึม VGG16 และ DDAET มีค่า F1 ค่อนข้างสูงเมื่อเทียบกับอัลกอริทึมอื่นๆ Image Threshold และ Edge detection มีค่า F1 อยู่ในระดับปานกลาง และน้อยกว่าอัลกอริทึมที่นำเสนออย่างมีนัยสำคัญ ADA MobileNetV3 และ ResNet50 มีค่า F1 ต่ำ โดยมี AlexNet เป็นอัลกอริทึมที่มีค่า F1 ต่ำที่สุดในชุดข้อมูลนี้

จากตารางแสดงผลการทดลองแสดงให้เห็นว่าวิธี CNNLSTM เป็นอัลกอริทึมที่มีประสิทธิภาพที่ดีและเหมาะสมสำหรับการตรวจจับสิ่งแปลกปลอมในข้อมูลภาพขดหมุ่น มีค่า Accuracy Recall Precision และ F1-score ที่สูงที่สุดในชุดข้อมูลนี้ VGG16 และ DDAET เป็นอัลกอริทึมที่เหมาะสมและมีประสิทธิภาพสูงในการตรวจจับสิ่งแปลกปลอม ในขณะที่ Image Threshold, Edge detection, ADA, MobileNetV3, และ ResNet50 มีประสิทธิภาพต่ำ ส่วน AlexNet มีประสิทธิภาพต่ำที่สุดสำหรับการตรวจจับสิ่งแปลกปลอมสำหรับภาพขดหมุ่น

**ตารางที่ 4.5** เวลาในการฝึกสอนและเวลาในการตรวจจับสิ่งแปลกปลอมของชุดข้อมูลภาพบริเวณลำตัว ชุดข้อมูลคอขวด และชุดข้อมูลภาพขวดหมุน

Methods	ชุดข้อมูลบริเวณลำตัว		ชุดข้อมูลคอขวด		ชุดข้อมูลภาพขวดหมุน	
	เวลาการฝึกสอน (วินาที)	เวลาการตรวจจับ (วินาที)	เวลาการฝึกสอน (วินาที)	เวลาการตรวจจับ (วินาที)	เวลาการฝึกสอน (วินาที)	เวลาการตรวจจับ (วินาที)
DDAET	<b>61.8</b>	0.004	61.8	0.004	61.8	0.004
CNNLSTM	103.59	0.12	<b>59.22</b>	0.065	<b>59.22</b>	0.065
ResNet50	1,500	0.5	1,500	0.5	1,500	0.5
AlexNet	690	0.3	690	0.3	690	0.3
VGG16	1,800	0.5	1,800	0.5	1,800	0.5
MobileNetV3	780	0.3	780	0.3	780	0.3
ADA	180	0.2	180	0.2	180	0.2
Image Threshold	0	0.002	0	0.002	0	0.002
Edge detection	0	<b>0.001</b>	0	<b>0.001</b>	0	<b>0.001</b>

ตารางที่ 4.5 แสดงเวลาในการฝึกสอนโมเดลและเวลาในการตรวจจับสิ่งแปลกปลอมของชุดข้อมูลภาพบริเวณลำตัว ชุดข้อมูลคอขวด และชุดข้อมูลภาพขวดหมุน วิธีการ DDAET ใช้เวลาในการฝึกสอนน้อยที่สุดสำหรับชุดข้อมูลบริเวณลำตัวขวดเมื่อเทียบกับอัลกอริทึมอื่นๆ และเวลาในการตรวจจับสิ่งแปลกปลอมน้อยกว่าวิธีการอื่นๆที่ใช้โครงข่ายประสาทในการตรวจจับสิ่งแปลกปลอม และใช้เวลาในการตรวจจับเท่ากันทั้งสามชุดข้อมูล วิธีการ CNNLSTM ใช้เวลาในการฝึกสอนเพื่อสร้างชุดของโมเดลโครงข่ายประสาทน้อยที่สุดในชุดข้อมูลคอขวด และชุดข้อมูลภาพขวดหมุน สอดคล้องกับผลการทดลองที่มีค่าความถูกต้องที่สูงของสองชุดข้อมูลนี้ เนื่องจากในแต่ละรอบการฝึกสอนโมเดลของ CNNLSTM มีการประเมินประสิทธิภาพทุกๆ 5 epoch หาก CNNLSTM มีค่าความถูกต้องมากกว่าค่าเกณฑ์ที่กำหนดจะถือว่าทำการฝึกสอนเสร็จสมบูรณ์ CNNLSTM สามารถเรียนรู้ทั้งสองชุดข้อมูลนี้ได้ดีทำให้มีรอบการฝึกสอนที่น้อย ส่งผลให้อัลกอริทึมนี้ใช้เวลาในการฝึกสอนโมเดลน้อยกว่าอัลกอริทึมอื่นๆ และยังใช้เวลาในการตรวจจับสิ่งแปลกปลอมน้อยเมื่อเทียบกับ ResNet50 AlexNet VGG16 MobileNetV3 และ ADA โดยที่วิธี VGG16 และ ResNet50 ใช้เวลาในการฝึกสอนมากกว่าวิธีการอื่นๆอย่างเห็นได้ชัดเนื่องจากสองวิธีการนี้เป็นโครงข่ายประสาทที่มีขนาดใหญ่ที่ซับซ้อนและพารามิเตอร์จำนวนมากที่ต้องทำการ fine-tuning จำนวนมาก ในทางกลับกัน จากตารางนี้แสดงให้เห็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เห็นว่าวิธีการฝึกสอนของ CNNLSTM ที่นำเสนอในงานวิจัยนี้สามารถช่วยลดเวลาในการฝึกสอนโมเดลลงอย่างเห็นได้ชัดเมื่อเทียบเวลาการฝึกสอนกับ ResNet50 VGG16 AlexNet และ MobileNetV3 อย่างไรก็ตามการออกแบบโมเดล autoencoder ที่มีโครงสร้างขนาดเล็กและใช้ชุดข้อมูลฝึกสอนจำนวนน้อยของวิธี DDAET ส่งผลให้ใช้เวลาฝึกสอนโครงข่ายประสาทเชิงลึกน้อยที่สุดทั้งสามชุดข้อมูล Image Threshold และ Edge detection ไม่มีเวลาในการฝึกสอนเนื่องจากเป็นวิธีการประมวลผลภาพที่ไม่ต้องการการฝึกสอนและใช้เวลาในการตรวจจับสิ่งแปลกปลอมน้อยและคงที่ การเลือกใช้วิธีการในการตรวจจับสิ่งแปลกปลอมควรพิจารณาค่า Precision และค่า F1 รวมถึงเวลาในการตรวจจับสิ่งแปลกปลอมเพื่อให้ได้ผลลัพธ์ที่เหมาะสมตามความต้องการของแต่ละกรณีการตรวจจับสิ่งแปลกปลอม

ผลการทดลองชุดข้อมูลภาพบริเวณลำตัวขวด VGG16 เป็นอัลกอริทึมที่มีความถูกต้องในการตรวจจับสิ่งแปลกปลอมมากที่สุด โดยมี DDAET อัลกอริทึมให้ความถูกต้องที่สูงและมีค่าความแปรปรวนของผลการทดลองน้อยกว่า VGG16 แสดงให้เห็นถึงประสิทธิภาพของ DDAET ที่ได้นำเสนอและเมื่อพิจารณาถึงเวลาในการฝึกสอนโมเดลเพื่อสร้างโครงข่ายประสาท autoencoder และการใช้ชุดข้อมูลภาพสำหรับฝึกสอนที่น้อยกว่าอัลกอริทึมอื่นๆแล้ว ทำให้ DDAET มีความเหมาะสมที่จะนำไปใช้ในการตรวจจับสิ่งแปลกปลอมบริเวณลำตัวขวดมากที่สุดอีกหนึ่งอัลกอริทึม

ผลการทดลองชุดข้อมูลภาพคอขวดแสดงให้เห็นถึงประสิทธิภาพของโมเดล DDAET ที่ได้นำเสนอก็มีความถูกต้องที่สูง ใช้เวลาในการฝึกสอนโมเดลน้อย และการใช้ชุดข้อมูลภาพสำหรับฝึกสอนที่น้อยกว่าวิธีการอื่นๆที่ใช้โครงข่ายประสาทแบบลึกเป็นอย่างมาก ทำให้ DDAET มีความเหมาะสมที่สุดที่จะนำไปใช้ในการตรวจจับสิ่งแปลกปลอมในชุดข้อมูลนี้ เนื่องจากชุดข้อมูลภาพคอขวดได้ทำการถ่ายภาพที่โฟกัสไปที่คอขวดโดยตรงทำให้ภาพที่ถ่ายมาได้นั้นสามารถถ่ายภาพสิ่งแปลกปลอมได้คมชัดกว่าภาพบริเวณลำตัวเมื่อมีขนาดพิกเซลที่เท่ากัน ทำให้อัลกอริทึมที่นำเสนอทั้งสองอัลกอริทึมมีความถูกต้องมากกว่าภาพบริเวณลำตัว

ผลการทดลองชุดข้อมูลภาพขวดหมุน CNNLSTM แสดงให้เห็นถึงประสิทธิภาพดีที่สุดในการตรวจจับสิ่งแปลกปลอมของภาพขวดหมุนเมื่อพิจารณาจากค่าความถูกต้องที่มากที่สุด CNNLSTM ยังใช้โครงข่ายประสาทเชิงลึกที่ได้ออกแบบทำให้ใช้เวลาในการฝึกสอนโมเดลน้อย วิธี DDAET ที่ได้นำเสนอก็มีความถูกต้องที่สูงกว่าอัลกอริทึมอื่นๆอย่างมีนัยสำคัญ และยังใช้เวลาในการฝึกสอนโมเดลและการใช้ชุดข้อมูลภาพสำหรับฝึกสอนที่น้อย จึงเป็นอีกหนึ่งอัลกอริทึมที่มีความเหมาะสมที่จะนำไปใช้ในการตรวจจับสิ่งแปลกปลอมในชุดข้อมูลนี้

ผลการทดลองแสดงให้เห็นถึงประสิทธิภาพของโมเดล CNNLSTM ที่นำเสนอสำหรับการตรวจจับข้อบกพร่องบนชุดข้อมูลภาพขวดหมุน โมเดล CNNLSTM ที่นำเสนอมีความแม่นยำและมีประสิทธิภาพดีตามเกณฑ์การประเมินที่ใช้ ซึ่งเวลาในการฝึกสอนเพื่อสร้างชุดของโครงข่ายประสาทเชิงลึกที่น้อยเมื่อเทียบกับวิธีการที่ใช้โครงข่ายประสาทวิธีการอื่นๆ ผลการทดลองเหล่านี้สามารถใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เป็นแนวทางในการเลือกโมเดลที่เหมาะสมโดยพิจารณาความแม่นยำและเวลาในการประมวลผลเพื่อตอบสนองความต้องการในการใช้งานในประเด็นที่กำหนดไว้ ผลการทดลองของ DDAET ที่ใช้ภาพพื้นหลังปกติและค่า threshold ของความผิดพลาดในการตรวจจับสิ่งแปลกปลอมในภาพพบว่าประสิทธิภาพและรวดเร็วในการระบุข้อบกพร่อง ความแม่นยำในการตรวจจับมีค่าสูง การตั้งค่าพารามิเตอร์จากผู้ใช้ของวิธี DDAET ที่แม่นยำเป็นสิ่งสำคัญที่ช่วยให้วิธีการนี้สามารถจำแนกประเภทสิ่งแปลกปลอมได้อย่างถูกต้อง วิธีการที่ใช้โครงข่ายประสาทเชิงลึกบางวิธีมีความแม่นยำที่สูงกว่าวิธีการที่นำเสนอในบางกรณีเนื่องจากวิธีการเหล่านี้ได้ทำการฝึกสอนด้วยข้อมูลภาพจำนวนมาก และภาพต้องมีความหลากหลายของข้อมูลตัวอย่างจึงมีความสามารถในการรู้จำลักษณะต่าง ๆ ที่เกี่ยวข้องกับข้อบกพร่องขวดได้ดี ในขณะที่วิธีการ Image threshold และ Edge detection 1 ในการทำงานน้อยที่สุด โดยไม่จำเป็นต้องใช้การฝึกสอนโมเดล



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 5

### สรุปผลการวิจัยและข้อเสนอแนะ

#### 5.1 สรุปผลการวิจัย

งานวิจัยนี้นำเสนอวิธีการตรวจจับสิ่งแปลกปลอมบนขดแก้วโดยใช้โครงข่ายประสาทเชิงลึกสองวิธีการประกอบด้วย CNNLSTM และ DDAET 1) CNNLSTM เป็นวิธีการตรวจจับสิ่งแปลกปลอมบนขดแก้วที่ใช้การผสมผสานของโครงข่ายประสาทเชิงลึกระหว่าง CNN และ LSTM เพื่อนำมาสกัดคุณลักษณะเชิงลึกจากภาพขดแก้ว และใช้อัลกอริทึมการจำแนกประเภทตามตัวอย่าง (instance-based classification) มาใช้ในการตรวจจับสิ่งแปลกปลอมบนขดแก้ว 2) DDAET เป็นวิธีการที่ออกแบบมาให้มีการฝึกสอนโมเดลโครงข่ายประสาทสามเชิงลึกได้อย่างรวดเร็ว ใช้เฉพาะชุดข้อมูลภาพขดแก้วที่ดีในการฝึกสอนทำให้งานต่อผู้ใช้งาน DDAET สร้างโครงข่ายประสาทเชิงลึกแบบออโตเอนโคเดอร์ (autoencoder) สำหรับการสร้างภาพพื้นหลังปกติ และใช้อัลกอริทึมตรวจจับข้อผิดพลาดในข้อมูล (outlier detection algorithms) โดยใช้ค่า Threshold ของข้อผิดพลาดที่ได้จากความผิดพลาดในการเข้ารหัสและถอดรหัสภาพขดแก้วปกติของโมเดล autoencoder มาทำการสกัดหาสิ่งแปลกปลอมบนขดแก้ว

งานวิจัยนี้ทำการทดลองเปรียบเทียบประสิทธิภาพของอัลกอริทึมการตรวจจับสิ่งแปลกปลอมที่นำเสนอกับชุดข้อมูลภาพขดแก้วสามชุดข้อมูลคือชุดข้อมูลภาพบริเวณลำตัวของขด ขดข้อมูลภาพคอขด และชุดข้อมูลภาพขดหมุน ผลการตรวจจับสิ่งแปลกปลอมของวิธีที่ CNNLSTM แสดงให้เห็นว่าวิธีนี้มีประสิทธิภาพในการตรวจจับข้อบกพร่องที่ดีสำหรับชุดข้อมูลภาพขดหมุนเมื่อเทียบกับวิธีการตรวจจับข้อบกพร่องอื่นๆ นอกจากนี้ CNNLSTM ใช้เวลาฝึกสอนน้อยกว่าวิธีการที่ใช้โครงข่ายประสาทเทียมวิธีการอื่นๆ แต่ก็ยังใช้เวลาในการฝึกสอนที่มากกว่า DDAET ผลการทดลองทำให้เห็นว่าวิธีการ CNNLSTM เหมาะสำหรับการตรวจจับข้อบกพร่องในขดแก้วในกระบวนการผลิตที่มีประสิทธิภาพ ผลการทดลองของวิธีการ DDAET แสดงให้เห็นว่าวิธีตรวจจับสิ่งแปลกปลอมที่นำเสนอมีประสิทธิภาพที่ดีในการตรวจจับตำหนิสำหรับชุดข้อมูลภาพบริเวณลำตัว และคอขดแก้ว อัลกอริทึมนี้ยังใช้เวลาฝึกสอนโครงข่ายประสาทเชิงลึกน้อยที่สุดและใช้ข้อมูลขดแก้วที่ดีที่สุดในการฝึกสอนโมเดล autoencoder ทำให้เหมาะสำหรับการตรวจจับสิ่งแปลกปลอมบนขดแก้วในกระบวนการผลิต

#### 5.2 ข้อเสนอแนะ

วิธีการ CNNLSTM สามารถพัฒนาต่อยอดการตรวจจับสิ่งแปลกปลอมไปยังขดประเภทต่างๆ และตรวจจับสิ่งแปลกปลอมประเภทอื่นเพิ่มมากขึ้น ความยืดหยุ่นของ CNNLSTM ช่วยให้สามารถเปลี่ยนแปลงจำนวนโมเดลและออกแบบขั้นตอนการตรวจจับที่กำหนดเองเพื่อปรับให้เข้ากับชุดข้อมูลเฉพาะทิศทางในอนาคตเหล่านี้จะช่วยเพิ่มขีดความสามารถและการประยุกต์ใช้วิธีการ

CNNLSTM ในการตรวจจับสิ่งแปลกปลอมและนำไปสู่ความก้าวหน้าของระบบควบคุมคุณภาพในอุตสาหกรรมต่างๆ

วิธีการ DDAET สามารถนำไปใช้กับขนาดประเภทต่างๆ และยังสามารถขยายการตรวจจับสิ่งแปลกปลอมประเภทต่างๆให้มากขึ้น สถาปัตยกรรมของ autoencoder สามารถเปลี่ยนแปลงได้และออกแบบให้เหมาะสมกับชุดข้อมูลเฉพาะ ค่า Threshold ของข้อผิดพลาดสามารถปรับค่าขึ้นหรือลดลงให้เหมาะสมกับการตรวจจับสิ่งแปลกปลอมประเภทอื่นๆ วิธีนี้ยังสามารถขยายไปสู่การเรียนรู้แบบ semi-supervised learning เพื่อตรวจจับข้อบกพร่องของผลิตภัณฑ์อื่นๆ ได้มากขึ้น



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## เอกสารอ้างอิง

- Fu, L., Zhang, S., Gong, Y., and Huang, Q. (2019). "Medicine Glass Bottle Defect Detection Based on Machine Vision." In **Chinese Control and Decision Conference (CCDC)** (pp. 5681-5685).
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In **Proceedings of the IEEE conference on computer vision and pattern recognition** (pp. 770-778).
- He, Z., and Liu, Q. (2020). "Deep Regression Neural Network for Industrial Surface Defect Detection." *IEEE Access*, 8, 35583-35591.
- Hochreiter, S., and Schmidhuber, J. (1997). "Long short-term memory." *Neural computation*, 9(8), 1735-1780.
- Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., and Adam, H. (2019). "Searching for mobilenetv3." In **Proceedings of the IEEE international conference on computer vision** (pp. 1314-1324).
- Howard, A., Sandler, M., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018). "MobileNetV2: Inverted residuals and linear bottlenecks." In **Proceedings of the IEEE conference on computer vision and pattern recognition** (pp. 4510-4520).
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "ImageNet classification with deep convolutional neural networks." In *Advances in neural information processing systems* (pp. 1097-1105).
- Koodtalang, W., Sangsuwan, T., and Sukanna, S. (2019). "Glass Bottle Bottom Inspection Based on Image Processing and Deep Learning." In **Research, Invention, and Innovation Congress (RI2C)**, Bangkok, Thailand, pp. 1-5.
- Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). "Gradient-based learning applied to document recognition." **Proceedings of the IEEE**, 86(11), 2278-2324.
- Li, F., Hang, Z., Yu, G., Wei, G., and Xinyu, C. (2017). "The method for glass bottle defects detecting based on machine vision." In **29th Chinese Control and Decision Conference (CCDC)** (pp. 7618-7621).
- Liang, W., and Sun, Y. (2022). "ELCNN: A Deep Neural Network for Small Object Defect Detection of Magnetic Tile." *IEEE Transactions on Instrumentation and Measurement*, 71, 1-10.

Lu, J., and Zhang, X. (2021). "A defect detection method for glass bottle mouth based on machine vision." *IEEE Transactions on Instrumentation and Measurement*, 70, 1-10.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- on area segmentation.” In *Proceedings of SPIE 12076, 2021 International Conference on Image, Video Processing, and Artificial Intelligence*, Shanghai.
- Ma, H. M., Su, G. D., Wang, J. Y., and Ni, Z. (2002). “A glass bottle defect detection system without touching.” In *Proceedings of the International Conference on Machine Learning and Cybernetics* (Vol. 2, pp. 628-632).
- Rafae, C. G. and Richard, E. W. (2007) “Digital Image Processing (3th ed.)” **Pearson International Edition.**
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018). “MobileNetV2: Inverted residuals and linear bottlenecks.” In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).
- Shorten, C., and Khoshgoftaar, T. M. (2019). “Data augmentation in machine learning: A survey.” *Big data*, 7(3), 1-26.
- Simonyan, K., and Zisserman, A. (2014). “Very deep convolutional networks for large-scale image recognition.” *arXiv preprint arXiv:1409.1556*.
- Singh, H., and Kaur, N. (2018). “A review of edge detection techniques for image segmentation.” *Journal of Computational and Theoretical Nanoscience*, vol.15, no.9, pp.4131-414.
- Zhou, X., Wang, Y., Xiao, C., Zhu, Q., Lu, X., Zhang, H., Ge, J., and Zhao, H. (2019). “Automated Visual Inspection of Glass Bottle Bottom With Saliency Detection and Template Matching.” *IEEE Transactions on Instrumentation and Measurement*, 68(11), 4253-4267.
- Zhou, X., Wang, Y., Zhu, Q., Mao, J., Xiao, C., Lu, X., and Zhang, H. (2020). “A Surface Defect Detection Framework for Glass Bottle Bottom Using Visual Attention Model and Wavelet Transform.” *IEEE Transactions on Industrial Informatics*, 16(4), 2189-2201.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก  
ผลงานที่ได้รับการตีพิมพ์



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## FACE SPOOFING DETECTION BASED ON DEEP FEATURE EXTRACTION AND INSTANCE-BASED CLASSIFICATION

NIPHAT CLAYPO<sup>1,\*</sup>, SAICHON JAIYEN<sup>2</sup> AND ANANTAPORN HANSKUNATAI<sup>1,\*</sup>

<sup>1</sup>Data Science and Computational Intelligence Lab  
Department of Computer Science  
School of Science

King Mongkut's Institute of Technology Ladkrabang  
No. 1, Chalong Krung 1, Chalong Krung Road, Lat Krabang Sub-District, Lat Krabang District  
Bangkok 10520, Thailand

\*Corresponding authors: {58605020; anantaporn.ha}@kmitl.ac.th

<sup>2</sup>School of Information Technology  
King Mongkut's University of Technology Thonburi  
No. 126, Pracha Uthit Road, Bang Mot Sub-District, Thung Khru District  
Bangkok 10140, Thailand  
saichon.jai@kmutt.ac.th

Received May 2022; accepted July 2022

**ABSTRACT.** Face recognition is an important task in smart home security for detecting a face or monitoring a person in a live video and verifying the identity of an authentic user. However, there have been spoofing face methods that can trick a face recognition algorithm into wrongly verifying the identity of the person. In this paper, we propose a new hybrid framework for spoofing face detection based on Convolutional Neural Network and Long Short-Term Memory (CNNLSTM) and instance-based learning algorithm. In addition, a new dataset called FSA-CCTV is proposed, which contains face images from CCTV video clips with many types of spoofing attacks. The performance of our method was compared to several other anti-spoofing methods: CNN and RI-LBP, SLRNN, HSV+YCbCr, ResNet50, YCbCr+SVM and YCbCr+KNN. The experimental results show that our method yielded 93.2% of Accuracy, 96.8% of Recall, 94% of Precision, 94.8% of  $F_1$ -score and 0.93 of AUC on the FSA CCTV dataset. From the experimental results we can conclude that the proposed algorithm outperforms other approaches and yielded the most stable classification accuracy on the proposed dataset.

**Keywords:** Convolutional neural network, Face spoofing attack detection, Feature extraction, Smart security

**1. Introduction.** Smart home technology provides home automation systems for homeowners and facilities that provide security, comfort, convenience, and energy efficiency [1]. In terms of security, residential smart security cameras can monitor, in real time, the happenings around the house. Face recognition is a feature of smart security [2]. It is a necessary method for identifying and verifying faces as well as authenticating users. Face recognition algorithms can automatically detect a face from a real-time video stream and search a face database to find a match [3]. A face recognition system must be able to detect spoofs. A facial liveness detection algorithm, which can decide whether a recognized face is real or spoofed, is an important subroutine in a face recognition system. Its main goals are to stop fraud and ensure an authentic face before the recognized face is matched to a face in the database for identification.

Types of face spoofing attacks are such as printed photo, mask, 3D mask, smartphone, tablet, and ID card. The widespread of Internet and smartphones makes it easy to find

DOI: 10.24507/iceel.17.02.235

images for spoofing faces. A high-performance anti-spoofing method can stop an attack outright and inform the user of the detected anomaly.

Spoofing detection methods have been proposed to solve face spoofing problems by identifying face liveness. The feature extraction method is the main type of data preprocessing for face spoofing detection methods. Texture-based methods are used to analyze color-texture information in face images by extracting low-level feature descriptions from different color spaces. Examples of these methods are Local Binary Patterns (LBP) [4-6], local similar patterns (LSP) [7], and YCbCr color [8]. These methods are traditional face spoofing detection approaches. Recently, Deep Learning (DL) is one of widely used methods for extracting features in an image. Convolutional Neural Network (CNN), a type of DL, is especially good at extracting features in images. Several studies have applied CNNs to detecting face spoofing. Similarly, many researchers have used CNNs to extract features that suit facial liveness detection. The review of these methods will show in Related Works section.

This paper presents a method for eliminating face spoofing problem. The main contributions from our paper are as follows:

- We create a new dataset called FSA-CCTV. This dataset is presented of diverse attacks type for developing the home security system. The FSA-CCTV is a collected image from video frames in long-distance shooting video clips of CCTV cameras. It consists of several different lighting conditions.
- We propose a new hybrid algorithm for classifying spoofed faces. A new structure of the CNN and LSTM is proposed and called CNNLSTM. The proposed method creates a set of CNNLSTM models and uses the models to extract features from images. Finally, instance-based algorithm is used to classify spoofed faces.

The rest of this paper is organized as follows: Section 2 reviews anti-spoofing face detection articles in the literature; Section 3 suggests the proposed dataset; Section 4 describes the proposed method; Section 5 explains the evaluation methods, the experimental setup, and the experimental outcomes; and finally, Section 6 suggests directions for future works and concludes the paper.

**2. Related Works.** In this section, we review existing face spoofing detection approaches mainly focused on features extraction with the texture-based and CNN methods. The related methods are divided into five groups as follows.

**2.1. Texture based methods.** In [9], Yang et al. proposed a person-specific face anti-spoofing method based on HOG and MsLBP feature extraction methods. Li et al. [10] proposed a method that converted images into YCbCr color space and then extracted features by LBP method. In the same vein, Fourati et al. [11] used Image Quality Assessment (IQA) and motion cues of face image to identify face spoofing.

**2.2. CNN based methods.** CNN is a specialization method for detected patterns in image. Chen et al. [12] proposed a Two-Stream Convolutional Neural Network (TSCNN). The TSCNN uses a Multi-Scale Retinex (MSR) space to solve illumination problem. Muhammad and Melo [13] proposed an SLRNN for spoofed face classification. The SLRNN method combined CNN and adding Long Short-Term Memory (LSTM) together. Rehman et al. [14] presented a face liveness detection method that incorporated a disparity layer in the CNN to learn dynamic disparity maps. Wirianto and Mauritsius [15] presented an Indonesia Labelled Face in the Wild (ILFW) dataset. The researchers suggested a DCNN network architecture called ResNet100 for face recognition.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**2.3. Texture-based and CNN based methods.** The texture-based feature and CNN based features were proposed by Chen et al. [16]. The research introduced Face Anti-spoofing Region-based Convolutional Neural Network (FARCNN), an instance of FARCNN that is based on improved faster Region-based Convolutional Neural Network (R-CNN) framework for face extraction and extraction features. They extended the faster R-CNN method and Retinex-based LBP to cover face anti-spoofing tasks with diverse illumination conditions. It classifies spoofing faces by using an SVM classifier.

**2.4. CNN based and machine learning methods.** Shao et al. [17] proposed a feature learning model for 3D-mask face anti-spoofing. A VGG was used to extract features and deep dynamic textures. Li et al. [18] presented a technique for extracting features by CNN, and detected spoofed faces by SVM. George and Marcel [19] presented a Multi-Channel Convolutional Neural Network (MCCNN). The new loss function was presented in that study.

**2.5. Texture based and CNN based and machine learning methods.** Khammari [20] presented a new method that extracted features by using LBP and WLD. The output features were encoded by CNN. The output from CNN was inputted to SVM to identify live or spoofed face. Chen et al. [21] extended the face algorithm from Li et al. [10]. That extension was based on RI-LBP and CNN and used SVM to identify face spoofing.

Texture-based methods extract features from low-level feature descriptions from different color spaces, and these lose some features, which makes the accuracy disappear. CNN-based methods can extract deep challenging features and extract from multi-view of an image by convolution layers, providing high performance for spoofing face detection.

**3. The FSA-CCTV Dataset.** The new dataset called FSA-CCTV is proposed. We fetched video streams from the IP cameras, following the Real Time Streaming Protocol (RTSP). Faces are detected by using a Haar Cascades technique [22]. Haar Cascades method is fast to detect and is an effective detection method. Then, we resized images to  $224 \times 224$ , and used data augmentation to increase the number of images. The types of face spoofing attacks of this dataset are fake images and images of a person wearing a mask from smartphone, iPad, office card, and ID card under multiple lighting conditions. Sample images from our proposed dataset are shown in Figure 1. The first row in the figure shows spoofed faces. The first two images in the second row are images of spoofed faces taken in IR mode, and the last two pictures are genuine faces. The dataset consists of 2,045 spoofed face images and 1,015 genuine face images. This dataset uses the data augmentation technique for increasing number of data, reduces overfitting and decreases variance of model. The spoofed faces are expanding to 3,724 images and the genuine faces increase to 2,396 images.



FIGURE 1. Sample images from the FSA-CCTV dataset

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**4. Face Spoofing Detection Method.** This section explains the proposed anti-spoofing approach.

**4.1. Network architecture of CNNLSTM.** CNNLSTM is used for feature extraction in this study. The outputs of CNNLSTM are the new instances for instance-based learning to recognize spoofing faces. The architecture of CNNLSTM for extracting features from images is inspired by AlexNet [23]. We improve the neural network model of AlexNet by adding LSTM algorithm to its convolutional layer. The LSTM is designed to avoid a long-term dependency problem [13]. It consists of several memory cells. Each memory cell has three elements: write, read, and forget (delete). The convolutional layers in the CNNLSTM are extended from those of AlexNet as explained as follows. In the first and second convolutional layers, we set the number of convolutional kernels to 256 kernels, and in the fifth convolutional layer was set to  $512 \times 5$  kernels. In the final representation layer, LSTM, a Recurrent Neural Network (RNN) was used to receive inputs from prior convolutional layers. A Batch-Normalization (BN) method was applied to the outputs of convolutional layers before they were inputted into the LSTM. The output of BN is tuple of integer fed into the LSTM layer. A set of 512 outputs from LSTM output units was fed into the next layer. The fully connected layers contain 1,000 hidden neurons. The activation function of all the convolutional layers is a Rectified Linear Unit (ReLU). The output layer contains one output neuron with a sigmoid activation function. The CNNLSTM model in this study was trained with a Stochastic Gradient Descent (SGD) method.

**4.2. The proposed feature learning network.** The overall process of creating feature learning networks of the proposed algorithm is illustrated in Algorithm 1.

---

**Algorithm 1:** Pseudo code of the feature learning network method.

---

**Input:**  $S$ : A training set,  $K$ : The number of the subdatasets,  $\theta$ : A performance threshold,  $m_{k-1}$ : The pre-trained weights

**Output:**  $E$ : A set of CNNLSTM models,  $C$ : A set of representation datasets

- 1: Randomly split the training dataset into  $K$  sub datasets:  $S \rightarrow \{s_1, s_2, s_3, \dots, s_K\}$ .
- 2: **For**  $k = 1, 2, \dots, K$  **do**:
- 3:      $AUC = 0$
- 4:     **While**  $AUC < \theta$  **do**:
- 5:         Train CNNLSTM from  $m_{k-1}$  with  $s_k$  to obtain the  $m_k$  model.
- 6:         Extract deep features  $F_k$  from  $s_k$  by  $m_k$ .
- 7:         Normalize every feature vectors  $f_k$  in  $F_k$  with an  $L2$ -norm technique.
- 8:         **For**  $i$  in  $y \in \{0, 1\}$  **do**:
- 9:             Compute the centroid of a class  $c_{ik}$  from  $F_{k \in N_{ik}}$ .
- 10:         **End For**
- 11:         Calculate the Euclidean distances from centroid  $c_{ik}$  to every feature of all  $F_k$  instances.
- 12:         Predicted label  $h_j$  of  $s_k$  by using  $c_{ik}$  and  $F_k$ .
- 13:         Compute  $AUC$  value from predicted labels  $h_j$  with true label  $y$ .
- 14:     **End While**
- 15:      $m_k \in E$  and  $c_{ik} \in C$ .
- 16: **End For**

---

To construct CNNLSTM models, our method randomly selects instances of data in the  $S$  and put them into several subdatasets. The subdataset  $s_k$  consists of  $N$  input data vectors  $x$  and label  $y$ :  $s_k = \{(x_{1k}, y_{1k}), \dots, (x_{jk}, y_{jk}), \dots, (x_{Nk}, y_{Nk})\}$  (Line1). These subdatasets aim to create diversity CNNLSTM models and representation of class data.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The subdataset  $s_k$  is the training set for a CNNLSTM,  $m_k$  (Line5). The current model gets the weights from previous model  $m_{k-1}$ .

A set of extracted feature vectors  $F_k = \{f_{1k}, f_{2k}, \dots, f_{Nk}\}$  is extracted from model  $m_k$  (Line6). The features  $f_{jk}$  are extracted from a data  $x_{jk}$ :  $f_{jk} = m_k(x_{jk})$ .

All feature vectors in  $F_k$  are normalized by using an  $L2$ -norm technique (Line7), and let  $f_{jk} = L2(f_{jk})$ . The norm of  $f_{jk}$  is calculated as follows.

$$|f_{jk}| = \sqrt{\sum_{l=1}^{n^k} |v_{lk}|^2}, \quad (1)$$

where  $v_{lk}$  is an extracted value in  $f_{jk}$ ,  $n^k$  is size of  $f_{jk}$ . For the proposed method to be able to recognize patterns of genuine and spoofing faces, a centroid of a class,  $c_{ik}$  is computed in this step. The centroid  $c_{ik}$  is an average of every feature vector in  $F_k$  from instances of class  $i$ th. The centroids  $c_{ik}$  is the representation of training data. The centroid  $c_{ik}$  is calculated as follows:

$$c_{ik} = \frac{1}{N_{ik}} \sum_{f_{jk} \in N_{ik}} f_{jk}, \quad (2)$$

where  $N_{ik}$  is the number of all data of class  $i$  in  $s_k$  (Line9). The similarity distances of every class between  $f_{jk}$  and  $c_{ik}$  were computed in Line11. The distance is a Euclidean distance. The prediction class  $h_j$  is selected by the minimum distance of a class (Line12). Compute the similarity value and predicted label  $h_j$  is defined as

$$h_j = \arg \min_j \|f_{jk} - c_{ik}\|_2, \text{ for } j = 1 : Nk, \quad (3)$$

where  $Nk$  is the size of  $s_k$ . The performance of  $m_k$  measures in every five training epochs, in terms of the Area Under the Curve (AUC). If AUC value is less than a threshold  $\theta$ , the  $m_k$  must be retrained for another 5 epochs (go to Line5),  $F_k$  and  $c_{ik}$  are recomputed, otherwise the training would be stopped and the model  $m_k$  is a member of  $E$ ;  $m_k \in E$ ; and  $c_{ik} \in C$  (Line15). Our training process minimizes the distance between the extracted features and a centroid of a class which have the same class label and maximizes the distance between the centroid of genuine and the centroid of spoof faces.

For the next subdataset  $s_{k+1}$ , we use pre-trained weights from the previous iteration  $k$  to learn the current data  $m_{k+1} \rightarrow m_k$ . To reduce training time and improve classification accuracy, it was convenient to use a pre-trained weight from a previous model. The models in  $E$  are different from a decision boundary model of the same architecture. The centroids

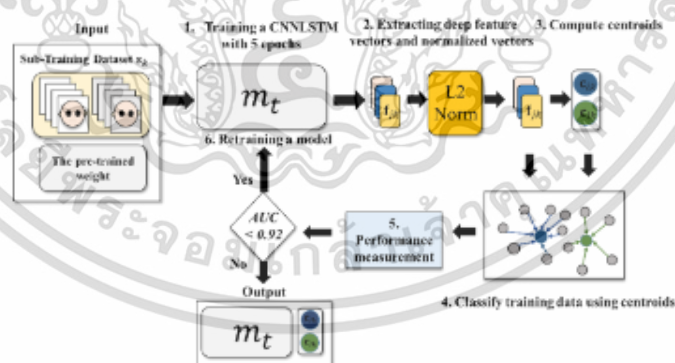


FIGURE 2. Overall process of feature learning networks for face spoofing detection

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

in  $C$  are the representation of the patterns of spoofing attack. The overview of feature learning method is illustrated in Figure 2.

**4.3. Classification of face spoofing.** The proposed instance-based classification method is discussed in this section. This method is a technique for preventing bias in class prediction. The distance weights are computed for every class. The pseudo code of the classification steps is shown in Algorithm 2.

---

**Algorithm 2:** Pseudo code of spoof face classification.

---

**Input:**  $E$ : CNNLSTM models,  $C$ : A set of centroid class data,  $x$ : An unseen data

**Output:**  $h_{final} = \arg \max_i \sum_i \sum_k \log \left( \frac{1}{ds_{ik}} \right)$

1: For  $k = 1, 2, \dots, K$  do:

2: Extract features  $f_k = m_k(x)$ .

3: Normalize the feature vectors  $f_k$  by  $L2$ -norm.

4: For  $i = 1 : y$  do:

5: Calculate the distance  $d_{ik}$  between centroid  $c_{ik}$  to features  $f_k$ .

6: Compute the distance weight  $ds_{ik}$ .

7: End For

8: End For

---

The sets of models  $E$  and centroids  $C$  are the inputs of Algorithm 2. The final output is calculated as follows:

$$h_{final} = \arg \max_i \sum_i \sum_k \log \left( \frac{1}{ds_{ik}} \right), \quad (4)$$

where  $h_{final}$  is the predicted class,  $i$  is class index,  $k$  is the number of models and  $ds_{ik}$  is the distance weights calculated by Equation (6). According to this equation, a high distance weight  $ds_{ik}$  would produce a low voting value, and vice versa. The predicted class  $h_{final}$  is selected with a maximum voting value of the individual class. The proposed classification method is illustrated in Figure 3.

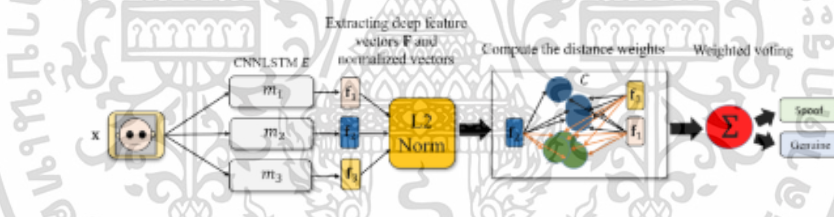


FIGURE 3. The overall process of face spoofing classification

The first step extracts feature from unseen data:  $f_k = m_k(x)$  (Line2). The similarity distance  $\alpha_{ik}$  between  $f_k$  and  $c_{ik}$  is calculated by Equation (5). Before  $\alpha_{ik}$  is computed, the  $f_k$  is normalized by  $L2$ -norm technique:  $f_k = L2(f_k)$  (Line3). The  $f_k$  is normalized into the same vector space of  $c_{ik}$ . The distances values are expressed as follows:

$$\alpha_{ik} = \|f_k - c_{ik}\|_2, \text{ for } k = 1 : K, \quad (5)$$

where  $\alpha_{ik}$  is a distance value of the  $i$ th class (Line5). Each feature vector is computed distance of every centroid of a class data in  $C$ . The voting weights are computed as follows:

$$ds_{ik} = \frac{\alpha_{ik}}{\sum_i \alpha_{ik}}. \quad (6)$$

The weights are the average distance between the feature vectors and centroids of spoofed and authentic faces (Line6). The main concept of distance measure is inspired by a face

identification method [24], applied to spoofing detection. If an unseen data  $\mathbf{x}$  is an authentic face, then the features extracted from CNNLSTM would be similar to the centroids of an authentic face, and the distance weight computed by Equation (6) would be high. Otherwise, the distance weight would be low.

**5. Experiment and Results.** This section describes and discusses the experiment and experimental results on our proposed face-spoofing dataset. The results of each method on our dataset were shown of the performance of spoofed face detection for smart home security.

**5.1. Experimental setup.** The FSA-CCTV dataset has 6,120 images, including 3,724 positive class images, and 2,396 negative class images. Let the genuine faces be negative class and spoofed faces be positive class. The experiment was done with five-fold cross-validation technique. The results were averaged over five testing folds. The input training threshold,  $\theta$ , was set to 0.92 and  $K$ , was set to 3, 5, and 10. To evaluate our method, the results were compared to those achieved by several widely known face anti-spoofing algorithms.

**5.2. Experimental results.** This research conducted the experiment only on the FSA-CCTV dataset because of limitations in the personal data. The performances of the proposed method and several state-of-the-art methods were evaluated. Performance values are reported as an average of five replicates from five runs.

Table 1 shows the classification accuracy on each testing fold, the average accuracy of five folds, and the Standard Deviation ( $SD$ ) value on the FSA-CCTV dataset. The result shows that the average classification accuracies of the proposed methods with all  $K$  sizes are better than those of the other compared methods in identifying spoofed faces. The accuracy results of our methods are still high, with a low variance in several testing folds (shown by  $SD$  value behind  $\pm$  in the average column). For the accuracy of the first and third testing folds, the accuracies of the proposed methods are higher than those of the other compared methods. The fifth testing fold of our methods with  $K = 5$  and  $K = 10$  is equal to CNN and RI-LBP, and HSV+YCbCr with 93% of accuracy. Considering the accuracy of the second and the fourth folds, SLRNN method has the highest accuracy in this testing fold. YCbCr+SVM method has the lowest classification performance in spoofed faces detection than those of other methods. On the other hand, the accuracy of CNN and RI-LBP, SLRNN, HSV+YCbCr, and ResNet50 are dropped on some testing folds. These methods have high variance of accuracy, especially SLRNN is very sensitive to a spoofing attack in this dataset. We used a  $t$ -test with alpha equal to 0.01 to evaluate the differences between the accuracy of the proposed methods and other methods in this

TABLE 1. The performance on the FSA-CCTV dataset in terms of Accuracy

Methods	Fold (%)					Average (%)
	1	2	3	4	5	
CNN and RI-LBP	93	92	89	89	93	91.2 $\pm$ 2* $\circ\star$
SLRNN	91	96	84	95	84	90 $\pm$ 5.7* $\circ\star$
HSV+YCbCr	91	88	88	89	93	89.8 $\pm$ 2.1* $\circ\star$
ResNet50	93	91	87	91	84	89.2 $\pm$ 3.6* $\circ\star$
YCbCr+SVM	85	87	88	88	88	87.2 $\pm$ 1.3* $\circ\star$
CNNLSTM	89	93	89	91	91	90.6 $\pm$ 1.6* $\circ\star$
YCbCr+KNN	92	92	91	91	92	91.6 $\pm$ 0.5* $\circ\star$
The proposed method $K = 3$	93	93	95	93	91	93 $\pm$ 1.4
The proposed method $K = 5$	93	93	94	93	93	93.2 $\pm$ 0.4
The proposed method $K = 10$	94	92	94	93	93	93.2 $\pm$ 0.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

experiment (\*, o, †, indicate the proposed method with  $K = 3$ ,  $K = 5$ ,  $K = 10$  is significantly better than the corresponding algorithm, respectively). The results show that our method with every  $K$  size is significantly better than the other methods by comparing with the average accuracy.

Table 2 shows the average values of *Recall*, *Precision*, and  $F_1$ -score of the proposed method and the other compared methods on the FSA-CCTV dataset. The proposed method of all  $K$  sizes has higher average *Recall* value than other methods. The proposed method with  $K = 3$  has the highest value of *Recall*. While YCbCr+SVM provides the lowest *Recall* with 88%. In addition, our method of all  $K$  sizes is significantly better than the other algorithms. The third column contains *Precision* values of spoofed faces. The obtained *Precision* values indicate an acceptable classification accuracy for the model. It can be seen in the column that the proposed method is inferior to SLRNN, and CNN and IR-LBP. HSV+YCbCr is the worst of *Precision* value. The  $F_1$ -score of all methods is shown in the fourth column. Our models with  $K = 3$  and  $K = 5$  are significantly better than all other methods except CNN and RI-LBP in terms of  $F_1$ -score. While the proposed method with  $K = 10$  is significantly better than YCbCr+SVM and equal to CNN and RI-LBP method. YCbCr+SVM provides the lowest value of  $F_1$ -score.

Figure 4 shows the *ROC* curves of all compared algorithms on the FSA-CCTV dataset. The highest point of the three *ROC* curves of the proposed method with different  $K$  is at

TABLE 2. The performance on the FSA-CCTV dataset in term of *Recall*, *Precision*, and  $F_1$ -score

Methods	Recall (%)	Precision (%)	$F_1$ -score (%)
CNN and RI-LBP	93.8* o†	95.2	94.6
SLRNN	90.4* o†	96	92.8*o
HSV+YCbCr	92.6* o†	91.6o	92.2*o
ResNet50	91.2* o†	93.6	92.2*o
YCbCr+SVM	88* o†	93	90.4* o†
CNNLSTM	92* o†	92o	92*o
YCbCr+KNN	91.6* o†	94	93*o
Proposed method $K = 3$	96.8	92	94.8
Proposed method $K = 5$	96	94	94.8
Proposed method $K = 10$	95.8	94	94.6

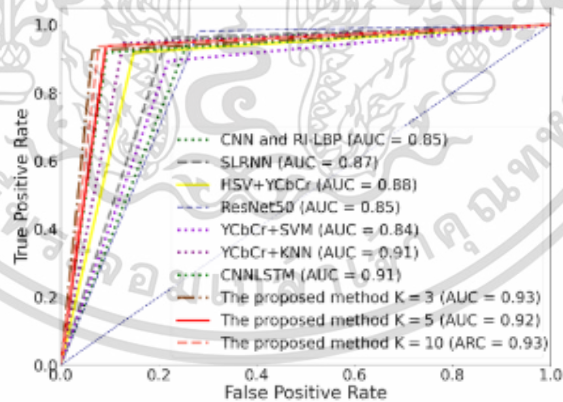


FIGURE 4. ROC curve on the FSA-CCTV dataset

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

the upper left corner. The proposed method of all  $K$  sizes outperforms all other methods – providing the higher value of  $AUC$  at 0.93, 0.92, and 0.93, respectively. YCbCr+SVM has the lowest  $AUC$  than other methods. The classification results from the proposed method are indicated that our method is efficient in spoofed face detection and provides high accuracy in identifying authentic faces. While the other comparison methods are high variance and have low  $ROC$  value meaning that those approaches are unbalanced in terms of classification performance.

The experimental results indicate that the CNN-based methods have higher precision than the color texture-based methods. The merging of the features extracted from CNN with the instance-based classification method of our algorithm can enhance the performance of the detection. However, when classifying with high lighting cases, the proposed method and those of other methods misclassified these images because these images have few gradients and depth shapes. Conversely, these have high performance on IR images.

**6. Conclusions.** This paper proposes the FSA-CCTV dataset and a new hybrid approach for face spoofing attack problem based on deep neural network and instance-based algorithms. The classification results of the proposed method on the FSA-CCTV dataset are compared with other widely known anti-spoofing methods. In the proposed method, a new architecture of deep neural network, called CNNLSTM, is used to extract features from facial images. CNNLSTM is trained on the proposed face-spoofing dataset. A new distance weight voting procedure is used to predict the final class. The performance of the proposed method is evaluated with 5-fold cross-validation technique. Experimental results show that the proposed method outperforms in classification performance than the other anti-spoofing classification methods. The proposed method is suitable for smart security on CCTV.

For future work, we plan to apply the proposed algorithm to different types of anti-spoofing tasks. The number of models generated by CNNLSTM can be varied, and the structure of CNNLSTM can be designed suitable for a particular dataset. The number of extracted features can be reduced, and other normalization methods can be applied to relieving span resources. This method can also be extended to an incremental learning that learns incremental data.

#### REFERENCES

- [1] M. R. Alam, M. B. I. Reaz and M. A. M. Ali, A review of smart homes-past, present, and future, *IEEE Trans. Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol.42, no.6, pp.1190-1203, 2012.
- [2] M. R. Chandra, B. V. Kumar and B. S. Babu, IoT enabled home with smart security, *International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, Chennai, India, 2017.
- [3] B. Ríos-Sánchez, D. Costa-da-Silva, N. Martín-Yuste and C. Sánchez-Avila, Deep learning for facial recognition on single sample per person scenarios with varied capturing conditions, *Applied Sciences*, vol.9, no.24, 5474, 2019.
- [4] D. He and L. Wang, Texture unit, texture spectrum, and texture analysis, *IEEE Trans. Geoscience and Remote Sensing*, vol.28, pp.509-512, 1990.
- [5] G. Zhao and M. Pietikäinen, Dynamic texture recognition using local binary patterns with an application to facial expressions, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.29, no.6, pp.915-928, 2007.
- [6] M. Heikkilä and M. Pietikäinen, A texture-based method for modeling the background and detecting moving objects, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.28, pp.657-662, 2006.
- [7] H. R. Pourreza, M. Masoudifar and M. Manafzade, LSP: Local similarity pattern, a new approach for rotation invariant noisy texture analysis, *2011 18th IEEE International Conference on Image Processing*, Belgium, 2011.
- [8] C. Poynton, *Digital Video and HDTV*, Morgan Kaufmann, 2003.
- [9] J. Yang, Z. Lei, D. Yi and S. Z. Li, Person-specific face antispoofing with subject domain adaptation, *IEEE Trans. Information Forensics and Security*, vol.10, no.4, pp.797-809, 2015.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [10] L. Li, P. L. Correia and A. Hadid, Face recognition under spoofing attacks: Countermeasures and research directions, *IET Biometrics*, vol.7, no.1, pp.3-14, 2018.
- [11] E. Fourati, W. Elloumi and A. Chetouani, Anti-spoofing in face recognition-based biometric authentication using image quality assessment, *Multimedia Tools and Applications*, pp.865-889, 2019.
- [12] H. Chen, G. Hu, Z. Lei, Y. Chen, N. M. Robertson and S. Z. Li, Attention-based two-stream convolutional networks for face spoofing detection, *IEEE Trans. Information Forensics and Security*, vol.15, pp.578-593, 2020.
- [13] U. Muhammad and W. Melo, Face anti-spoofing via sample learning based Recurrent Neural Network (RNN), *The British Machine Vision Conference (BMVC)*, Cardiff, UK, 2019.
- [14] Y. A. U. Rehman, L.-M. Po and M. Liu, SLNet: Stereo face liveness detection via dynamic disparity maps and convolutional neural network, *Expert Systems with Applications*, vol.142, 2020.
- [15] Wirianto and T. Mauritsius, The development of face recognition model in Indonesia pandemic context based on DCNN and Arcface loss function, *International Journal of Innovative Computing, Information and Control*, vol.17, no.5, pp.1513-1530, 2021.
- [16] H. Chen, Y. Chen, X. Tian and R. Jiang, A cascade face spoofing detector based on face, *IEEE Access*, vol.7, pp.170116-170133, 2019.
- [17] R. Shao, X. Lan and P. C. Yuen, Joint discriminative learning of deep dynamic textures for 3D mask face anti-spoofing, *IEEE Trans. Information Forensics and Security*, vol.14, no.4, pp.923-938, 2018.
- [18] H. Li, P. He, S. Wang, A. Rocha, X. Jiang and A. C. Kot, Learning generalized deep feature representation for face anti-spoofing, *IEEE Trans. Information Forensics and Security*, vol.13, no.10, pp.2639-2652, 2018.
- [19] A. George and S. Marcel, Learning one class representations for face presentation attack detection using multi-channel convolutional neural networks, *IEEE Trans. Information Forensics and Security*, vol.16, pp.361-375, 2021.
- [20] M. Khammar, Robust face anti-spoofing using CNN with LBP and WLD, *IET Image Processing*, vol.13, no.11, pp.1880-1884, 2019.
- [21] F. M. Chen, C. Wen, K. Xie, F. Q. Wen, G. Q. Sheng and X. G. Tang, Face liveness detection: Fusing colour texture feature and deep feature, *IET Biometrics*, vol.8, no.6, pp.369-377, 2019.
- [22] A. B. Shetty, Bhoomika, Deeksha, J. Rebeiro and Ramyashree, Facial recognition using Haar cascade and LBP classifiers, *Global Transitions Proceedings*, vol.2, pp.330-335, 2021.
- [23] A. Krizhevsky, I. Sutskever and G. E. Hinton, Advances in neural information processing systems, *Advances in Neural Information Processing Systems (NIPS2012)*, vol.25, 2012.
- [24] F. Schroff, D. Kalenichenko and J. Philbin, FaceNet: A unified embedding for face recognition and clustering, *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.815-823, 2015.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# Inspection System for Glass Bottle Defect Classification Based on Deep Neural Network

Niphath Claypo<sup>1</sup>, Saichon Jaiyen<sup>2</sup> and Anantaporn Hanskunatai<sup>3</sup>

Automation, Innovation, Intelligence, and Data Science Research Unit  
Department of Computer Science  
School of Science

King Mongkut's Institute of Technology Ladkrabang  
Chalong Krung 1, Chalong Krung Road, Lat Krabang Sub-District, Lat Krabang District  
Bangkok 10520, Thailand<sup>1,3</sup>

58605020@kmitl.ac.th; anantaporn.ha@kmitl.ac.th<sup>1,3</sup>  
King Mongkut's University of Technology Thonburi  
Pracha Uthit Road, Bang Mot Sub-District, Thung Khru District  
Bangkok 10140, Thailand<sup>2</sup>  
saichon.jai@kmutt.ac.th<sup>2</sup>

**Abstract**—The problem of defects in glass bottles is a significant issue in glass bottle manufacturing. There are various types of defects that can occur, including cracks, scratches, and blisters. Detecting these defects is crucial for ensuring the quality of glass bottle production. The inspection system must be able to accurately detect and automatically determine that the defects in a bottle affect its appearance and functionality. Defective bottles must be identified and removed from the production line to maintain product quality. This paper proposed glass bottle defect classification using Convolutional Neural Network with Long Short-Term Memory (CNNLSTM) and instant base classification. CNNLSTM is used for feature extraction to create a representation of the class data. The instant base classification predicts anomalies based on the similarity of representations of class data. The convolutional layer of the CNNLSTM method incorporates a transfer learning algorithm, using pre-trained models such as ResNet50, AlexNet, MobileNetV3, and VGG16. In this experiment, the results were compared with ResNet50, AlexNet, MobileNetV3, VGG16, ADA, Image threshold, and Edge detection methods. The experimental results demonstrate the effectiveness of the proposed method, achieving high classification accuracies of 77% on the body dataset, 95% on the neck dataset, and an impressive 98% on the rotating dataset.

**Keywords**—Convolution neural network; Glass bottle; Defect detection; Long short-term memory; Inspection machine

## I. INTRODUCTION

Defect detection plays a vital role in glass bottle production, safeguarding product quality, consumer safety, brand reputation, and value. By investing in robust flaw detection systems and processes, manufacturers can maintain impeccable standards, safe, and visually appealing glass bottles to their customers [1,2]. Common types of defects that can occur in glass bottles include: 1) Stones are foreign stone grains embedded in the glass, degrading the quality of the bottle. 2) Tears are deformed breaks or fractures on the surface of the bottle. 3) Blisters are raised or swollen areas caused by uneven cooling during manufacturing, affecting both strength and

appearance. 4) Cracks are breaks in the surface that compromise the structural integrity of bottle.

The inspection machine uses various technologies to detect defects, including cameras, lasers, and sensors. The inspection machine uses cameras to capture images of the products, and the software analyzes these images to detect defects. When defects are detected, the software sends a signal to the control system to remove the faulty product from the production line.

As for the problem with traditional inspection machines, they cannot inspect defects in complex areas and require the use of imaging techniques to visualize the defects. Because these machines utilize image processing techniques and require various parameter settings to be adjusted by users for defect detection. It results in the inability to completely remove defective bottles from the production line, undermining confidence in the manufacturing process. The detection method on inspection machine for detecting the defects must be accurate, precise, capable of recognizing defect well, and fast in the learning process of defect patterns.

This research proposes a new method for detecting defects in glass bottles using a deep neural network for extracting deep features and instance-based classification. The method uses CNN combined with LSTM to recognize distinctive features of defects and extract those features. The training process is designed to create a set of CNNLSTM models with fewer training iterations, helping to reduce the training time. The instance-based classification is used to classify the defect in the images. It reduces parameter settings by users for defect detection, resulting in a reduction in user workload.

The rest of this paper is organized as follows: Section II presents a background study of glass bottle defect detection techniques. Section III explains the glass bottle defect dataset and describes the proposed glass bottle defect classification method. Section IV explains the evaluation methods, the experimental setup, and the experimental outcomes; and finally.

Section V suggests directions for future works and concludes the paper.

## II. RELATED WORKS

The existing defect detection techniques mainly focused on the defect in glass bottles. The related methods are as follows. Latina et al. [3] presented defect detection method for detecting defects in glass bottles for the purpose of reusing them and highlights the limitations of the manual inspection in micro, small, and medium enterprises (MSMEs). The research introduced a cost-effective deep learning-based method using the SSD MobileNetV2 model to detect various defects in glass bottles. The method used transfer learning and data augmentation techniques to achieve high accuracy, with up to 98.07% overall system accuracy. Gong et al. [4] presented a machine vision system designed for the automatic online inspection of defects in transparent labels on curved glass bottles. The system used an area-array camera and a custom-made blue dome illumination device to capture high-quality still images by minimizing reflection. To address the challenge of distorted curved geometry, a deformable template matching method was employed for precise defect location. The method also included an adaptive threshold selection strategy that effectively detected small scratches by using global and local threshold values along with a Gaussian fitting algorithm. Additionally, techniques such as skeleton extraction and distance transformation were applied to detect the complete edge contour of Chinese characters with a special font, considering the golden edge printing error. Field tests demonstrated an impressive detection accuracy of 99.5% at a speed of 60 bottles per minute, covering over 60,000 bottles. Vitis et al. [5] proposes an algorithm that achieves high detection accuracy while significantly reducing processing time. By using adaptive thresholding and analyzing luminous intensity variations, the algorithm effectively detects blob and airline defects while mitigating the impact of tube curvature, rotation, and vibration. Comparative evaluations demonstrate an 86% reduction in processing time, a 268% increase in throughput, and improved detection accuracy compared to existing methods. The algorithm also incorporates Region of Interest reduction techniques and a tuning procedure for parameter adjustment during production batch changes. The performance of the algorithm was assessed in a real environment, and it successfully identified misclassified tubes, suggesting its practical applicability. Zhang et al. [6] suggested a machine learning-based acoustic defect detection (LearningADD) system to replace manual inspection. The system used an improved Hilbert-Huang transform (HHT) to extract features from acoustic signals and a Shuffled Frog Leaping Algorithm (SFLA) to select features. Five deployment strategies were compared and optimized to improve real-time performance. The LearningADD system was validated using data from a real-life beverage factory. The F-measure of the system reached 98.48%. The proposed deployment strategies were verified using experiments on private cloud platforms. The Distributed Heavy Edge deployment strategy outperformed other strategies, with a defect detection time of less than 2.061 seconds for 99% of bottles. Zhou et al. [7] proposed a surface defect detection framework, which consisted of three main components. Firstly, a novel

localization method called entropy rate superpixel circle detection (ERSCD) was introduced. It combined least-squares circle detection, entropy rate superpixel (ERS), and an improved randomized circle detection to accurately identify the region of interest (ROI) on the bottle bottom. The ROI was then divided into two measurement regions: the central panel region and the annular texture region. For defect detection in the central panel region, a method named frequency-tuned anisotropic diffusion super-pixel segmentation (FTADSP) was proposed. It integrated frequency-tuned salient region detection (FT), anisotropic diffusion, and improved superpixel segmentation to accurately detect defect regions and boundaries. For defect detection in the annular texture region, a strategy called wavelet transform multiscale filtering (WTMF) was proposed. It employed wavelet transform and a multiscale filtering algorithm to reduce texture influence and enhance robustness to localization errors. Zhou et al. [8] presented a new apparatus for real-time bottle bottom inspection. The apparatus used a combination of Hough circle detection and size prior to locating the bottom of the bottle. The region of interest was then divided into three measurement regions: central panel region, annular panel region, and annular texture region. A saliency detection method was used to find defective areas in the central panel region. A multiscale filtering method is used to search for defects in the annular panel region. Template matching was combined with multiscale filtering to detect defects in the annular texture region.

The problem with traditional methods of detecting anomalies on glass bottles using image processing is the difficulty in detecting anomalies, especially when there are color and shape variations on the bottles. Inconsistent lighting conditions also pose challenges in detecting anomalies. The disadvantage of using deep neural networks for detecting counterfeit on glass bottles is the complexity involved in building and training the network. Deep neural network-based methods require a large amount of data and longer training time. Additionally, fine-tuning and selecting appropriate parameters for deep networks can be challenging.

## III. METHODOLOGY

### A. Glass Bottle Dataset

There are two main categories of glass bottle inspection machines: straight glass bottle inspection machines and rotary glass bottle inspection machines. Fig. 1 shows an example of an inspection machine for glass bottles. The straight glass bottle inspection machine is an automated system that utilizes advanced technologies, such as computer vision and image processing, to detect and identify defects in glass bottles. It captures images of the bottles from different angles and analyzes them in real-time to ensure product quality and prevent issues during production and transportation. The straight glass bottle inspection machine is depicted in Fig. 1(a). On the other hand, the rotating glass bottle inspection machine is a specialized automated system designed specifically for inspecting glass bottles using a rotating mechanism. By rotating the bottles, it enables a comprehensive examination of the entire surface, ensuring a high level of accuracy in defect detection. The rotating glass bottle inspection machine is shown in Fig. 1(b).

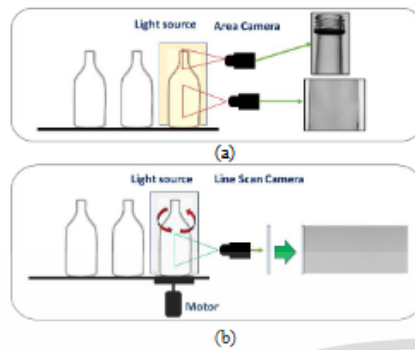


Fig. 1. Inspection machine (a) Straight glass bottle inspection machine (b) Rotating glass bottle inspection machine

The bottle inspection machine utilizes a vision camera-based inspection method for fully automated visual inspection of glass bottles. The process is as follows. Set the bottles in the designated area on the conveyor belt of the bottle inspection machine. The bottles are properly aligned and spaced to ensure accurate and consistent imaging. Provide proper lighting conditions for capturing clear and well-illuminated bottle images. Set up the camera parameters such as focus, exposure, and white balance to optimize image quality and clarity. Initiate the image capture process using the bottle inspection machine to trigger the camera to capture images of the bottles, as they pass through the inspection area. The machine utilizes two cameras to capture images of the bottle; one camera is focused on the neck, while the other is focused on the body of the bottle. Fig. 2 illustrates the composition of the side wall of glass bottles. (a) displays images from the neck dataset, (b) provides

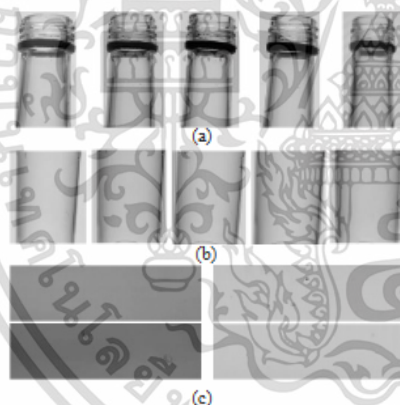


Fig. 2. Glass bottle dataset (a) Neck dataset (b) Body dataset (c) Rotating bottle dataset

an example of the body dataset, and (c) showcases sample bottle images from the rotating dataset. The body dataset for glass bottles comprises 571 images, with 368 images showing defects and 203 images without defects. Similarly, the neck dataset consists of 570 images, including 250 images with defects and 320 images without defects. As for the rotating dataset, it comprises 120 images, with 70 images displaying defects and 50 images representing undamaged bottles. Each image in the body and neck datasets is standardized to a size of 900x800 pixels (width x height), ensuring consistency across the dataset. On the other hand, the images in the rotating dataset are resized to a dimension of 1024x800 pixels (width x height). The captured images are transmitted to the detection system for analysis. If any defects, such as stones, tears, or blisters, are identified on the side wall of a bottle, the detection system generates a signal to reject the faulty bottle. This process ensures that only bottles without defects continue downstream in the production line. Fig. 3 (a) is shown tear defect, (b) is a blister, and (c) is stone defect on slid wall of bottle. These defects can appear throughout the bottle.



Fig. 3. Defect images. (a) Tear, (b) Blister, (c) Stone

### B. The proposed method

This section describes the proposed defect detection approach, which is divided into three parts: 1) The network architecture of CNNLSTM for extracting deep features in images, 2) the training process for creating a set of CNNLSTM models and computing the representation of class data, 3) the classification approach that suggests using the extracted deep features from unseen bottle image and applying the distance weight method to classify defect or normal glass bottle images.

#### 1) Network architecture of CNNLSTM

CNNLSTM is a convolutional neural network (CNN) combined [9] with LSTM (Long Short-Term Memory) [10] to extract deep features in images. CNNs are effective at extracting spatial features from images, while LSTMs are capable of capturing temporal dependencies in sequential data. The fusion of these two architectures enables CNNLSTM to extract deep features from images. The convolutional layers choose pre-trained deep-learning models available in the Keras library, including VGG16, ResNet50, MobileNetV3, and AlexNet. Using a pre-trained model can reduce the amount of time and resources needed to train a model. It already learned to recognize a variety of features and improve the accuracy of the model. Added LSTM to its convolutional layer. The detail of LSTM unit is shown in Fig. 4.

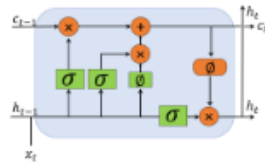


Fig 4. LSTM

LSTM comprises multiple memory cells, each consisting of three essential elements: write, read, and forget (delete). During each time step ( $t$ ), the forget gate unit is updated according to the following process,

$$f_t = \sigma_z(W_f x_t + U_f h_t + b_f), \quad (1)$$

$$i_t = \sigma_z(W_i x_t + U_i h_t + b_i), \quad (2)$$

$$o_t = \sigma_z(W_o x_t + U_o h_t + b_o), \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \sigma_z(W_c x_t + b_c), \quad (4)$$

$$h_t = o_t \odot \sigma_h(c_t), \quad (5)$$

where  $x_t$  is the input data,  $f_t$  is the forget gate,  $i_t$  is the input/update gate,  $o_t$  is the output gate,  $c_t$  is the cell state vector,  $\sigma$  is the sigmoid activation function,  $\odot$  is the Hadamard product (element-wise product), and  $h_t$  is the output of the LSTM unit. In the final representation layer, an LSTM, which is a Recurrent Neural Network (RNN) suitable for sequential data, receives inputs from previous convolutional layers. The LSTM produces a set of 512 outputs, which are then passed to the next layer. The outputs of the convolutional layers are subjected to Batch-Normalization (BN) before they are fed into the LSTM layer. The fully connected layers contain 1,000 hidden neurons. The activation function of all the convolutional layers (with a max pooling size of  $2 \times 2$ ) is a Rectified Linear Unit (ReLU) [11]. Deep features are extracted from these layers. The output layer contains one output neuron with a sigmoidal activation function. The size of the input images is set as  $224 \times 224 \times 3$  pixels. The CNNLSTM is trained using a Stochastic Gradient Descent (SGD) method [12]. The learning rate is set to 0.01, and the batch size is 8 images. The CNNLSTM neural network architecture is shown in Fig. 5.

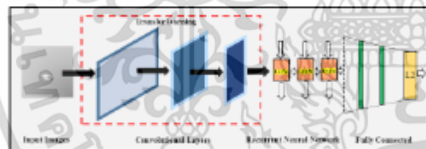


Fig 5. CNNLSTM

**Algorithm 1:** Pseudo code of the feature learning network method.

**Input:**

$S$ : A training bottle dataset:  $\{(x_1, y_1), \dots, (x_n, y_n)\}$

$K$ : The number of the sub datasets

$\theta$ : A performance threshold value

$m_{k-1}$ : The pre-trained model

**Output:**

$E$ : A set of CNNLSTM models

$C$ : A set of representation of class data

- 1: Randomly split the training dataset into  $K$  sub datasets:  $S \rightarrow \{s_1, s_2, \dots, s_K\}$ .
- 2: **For**  $k = 1, 2, \dots, K$  **do**:
- 3:  $MSE = 0$  and  $m_k = m_{k-1}$
- 4: **While**  $MSE < \theta$  **do**:
- 5: Train model  $m_k$  model with  $s_k$
- 6: Extract features  $F_k$  in images by  $m_k$ .
- 7: Normalize all features in  $F_k$  with an L2-norm technique.
- 8: **For**  $i$  in  $y \in \{0,1\}$  **do**:
- 9: Compute the representation of each class data  $c_{ik}$  from  $F_k$ .
- 10: **End For**
- 11: Evaluate the distances from centroid  $c_{ik}$  to every feature of all  $F_k$  instances.
- 12: Predicted a bottle is normal or abnormal ( $h$ ) is based on the minimum distance between  $c_{ik}$  and  $F_{k \in \{n,ik\}}$ .
- 13: Compute means squared error MSE from  $h$  with true label  $y$ .
- 14: **End While**
- 15:  $m_k \in E$  and  $c_{ik} \in C$ .
- 16: **End For**

## 2) The proposed learning process

Our training process aims to minimize the similarity distance between the extracted deep feature vectors and the centroid of the corresponding class with the same label. Simultaneously, it maximizes the similarity distance between the centroid of good bottles and the centroid of defective bottles.

The overall learning process of creating a set of CNNLSTM and representing class data is proposed in this section. The process of creating a set of CNNLSTM models and generating representations of class data is illustrated in Algorithm 1. The inputs of the learning process include a training dataset  $S$ , a specified number of sub-training sets  $K$ , a threshold value  $\theta$ , and the pre-trained weight  $m_{k-1}$ . The parameter  $K$  is a parameter used to determine the number of sub-training sets and the number of CNNLSTM models, which needs to be appropriately adjusted. The value of  $K$  affects the number of samples in each sub-training set used for training CNNLSTM and the diversity of CNNLSTM models.  $\theta$  is a parameter used to define the initial performance of CNNLSTM in each training iteration. The pre-trained weight  $m_{k-1}$  helps to reduce training time and resource requirements, improve performance, and enable the model to perform well in limited data conditions.

To construct a set of CNNLSTM models, our method involves randomly selecting images in the  $S$  training dataset and organizing them into several partitions or subsets, i.e.,  $S \rightarrow \{s_1, s_2, \dots, s_k, \dots, s_K\}$ . The purpose of creating these sub-datasets is to foster diversity in CNNLSTM models and generate representation vectors for both normal bottles and bottles with defects.

The current  $s_k$  is the training bottle images for a CNNLSTM model,  $m_k$ . Let the set models is ensemble  $E$ . The convolutional layer of the current model uses pre-train weight of previous round. The output of training with five epochs is  $m_k$  (Step 5).

In Step 6, deep features vectors  $F_k$  are extracted from the images in  $s_k$ . This process involves removing the output layer from the CNNLSTM and obtaining the output features through the fully connected layers. Let the deep feature vector  $f_{jk}$  be an instance extracted from an image in  $F_k : f_{jk} = m_k(x_{jk})$ . The feature vector consists of 1,000 components, which corresponds to the number of hidden neurons:  $f_{jk} = \{v_1, v_2, \dots, v_{n_k}\}, f_{jk} \in F_k$ .

Each feature vector is normalized using an L2-norm technique (Step 7). This normalization technique is commonly used to scale and standardize vectors in machine learning and data analysis. The L2-norm, also known as the Euclidean norm or the L2-norm, is a mathematical measure of the length or magnitude of a vector. In the context of feature normalization, the L2-norm technique calculates the square root of the sum of the squares of each component in the vector, resulting in a normalized vector with a magnitude of 1. Here, we provide an explanation of the concept of L2-norm. Consider a deep feature vector as

$$F_k = [f_{1k}, f_{2k}, f_{3k}, \dots, f_{nk}] \tag{6}$$

where

$$|f_{jk}| = \sqrt{\sum_{i=1}^{n^k} |v_{ik}|^2} \tag{7}$$

where  $v_{ik}$  represents an extracted feature within  $f_{jk}$ , and  $n^k$  denotes the total number of extracted features  $v$  in  $f_{jk}$ . L2 serves as the final output layer of CNNLSTM, producing the extracted features as the output.

The representation of the class data  $c_{ik}$  is computed as the average of all feature vectors in  $F_k$ , which have been extracted from instances of the  $i$ -th class sample data (Step 9). In this step, we compute the centroids for the two classes: the centroid of the good bottles (class 0) and the centroid of the bottles with defects (class 1):  $y = \{0, 1\}$ . These class centroids enable the recognition of patterns on the side walls of the bottles. The centroid vector  $c_{ik}$  is calculated as follows:

$$c_{ik} = \frac{1}{N_{ik}} \sum_{f_{jk} \in N_{ik}} f_{jk} \tag{8}$$

where  $c_{ik}$  constrain the average values  $a : c_{ik} = \{a_1, a_2, \dots, a_n\}$ ; and  $N_{ik}$  is the number of all data instances of class  $i$  in  $s_k$ . The centroids are representation of class data.

Predicted a bottle is normal class or abnormal class ( $h$ ) is based on the similarity distance between  $c_{ik}$  and  $F_{k \in N_{ik}}$  (Step 12). The similarity distances between each class are compared using the Euclidean distance to predict the category of the bottle. Predict normal or abnormal bottles by minimizing the similarity distance value defined as

$$h_j = \arg \min_i \|f_{jk} - c_{ik}\|_2, \text{ for } j = 1 : N_k, \tag{9}$$

where  $N_k$  is the number of instances in  $s_k$ .

To assess the extraction and classification performance of a trained CNNLSTM model  $m_k$ , the Mean Squared Error (denoted as MSE) is measured. If the model error (MSE) exceeds an error threshold value ( $\theta$ ), the deep neural model will undergo an additional training process of 5 epochs (proceed to Step 5). If the condition is met, the deep feature vectors  $F_k$  and the centroid  $c_{ik}$  are recomputed. Otherwise, the training process is stopped, and the model is obtained. The training iteration  $k$  is completed based on the performance condition. The threshold value  $\theta$  is used to measure the performance and reduce the number of training iterations in each model generation. It serves as a criterion for selecting a set of models that require fewer training iterations compared to the training

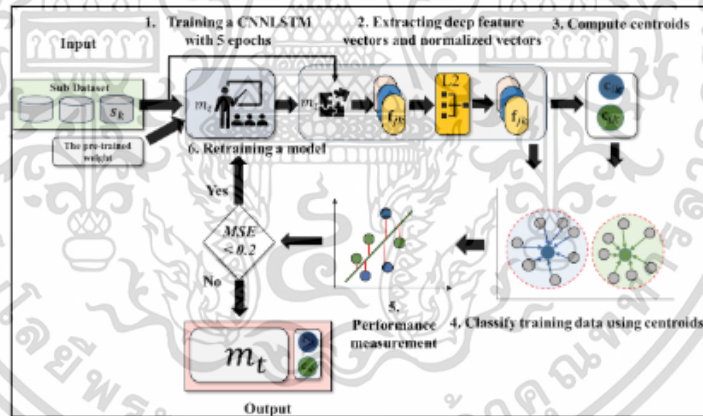


Fig. 6. Over all process of feature learning networks for glass bottle

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

process of a normal deep neural network. At this point, the model  $m_k$  is added to the ensemble  $E$  ( $m_k \in E$ ), and the centroid  $c_{ik}$  is included in the set  $c_{ik} \in C$  (Step 15). For the next iteration, we used pre-trained CNNLSTM weights from the previous iteration  $k$  to learn the current data  $m_{k+1} \rightarrow m_k$ . This method eliminates the need to reset the weights each time and only requires fine-tuning on the new set of bottle images. The overview of feature learning method is illustrated in Fig. 6.

**Algorithm 2:** Pseudocode for classifying bottle defects.

**Input:**

$E$ : A set of CNNLSTM models  
 $C$ : A set of centroids of class data  
 $x$ : A bottle images

**Output:**

$$h_i = \operatorname{argmin}_i \sum_k sd_{ik}$$

- 1: **For**  $k = 1, 2, \dots, K$  **do**:
- 2:   Extract deep features  $f_k = m_k(x)$
- 3:   Normalize the deep feature vectors  $f_k$  by L2-norm
- 4:   **For**  $i = 1:y$  **do**:
- 5:     Calculate the distance  $d_{ik}$  between centroid  $c_{ik}$  to features  $f_k$
- 6:   **End For**
- 7:   Compute the distance weight  $sd_{ik}$
- 8: **End For**

### 3) The classifying process

This section presents a detailed discussion of the proposed similarity voting method. This classification method focuses on utilizing distinctive features of defects for bottle defect detection. The similarity distance is calculated between the representation features of normal bottles and abnormal bottles. The classification steps of our method are outlined in Algorithm 2.

The input for Algorithm 2 consists of a set of CNNLSTM models  $E$ , the centroid of class data  $C$ , and the bottle image  $x$ . The models in set  $E$  differ from the decision boundary models of the same architecture. Extract features in bottle image using each model in  $E$  and combine them to predict the final output. Deep feature vectors are extracted from the bottle image:  $f_k = m_k(x)$  (Line 2). Then, the deep feature vector  $f_k$  is normalized using the L2-norm technique:  $f_k = L2(f_k)$  (Step 3). The deep feature vector is normalized to ensure it resides in the same vector space as  $c_{ik}$ , enabling the calculation of similarity

values. The similarity distance  $d_{ik}$  between feature and centroid is expressed as follows:

$$\alpha_{ik} = \|f_k - c_{ik}\|_2, \text{ for } k = 1:K, \quad (10)$$

where  $K$  is the number of ensemble model. Each centroid calculates the similarity distance with the features of the bottle image. The voting weight of each model is computed as follows:

$$sd_{ik} = \frac{\alpha_{ik}}{\sum_i \alpha_{ik}}, \quad (11)$$

where  $sd_{ik}$  is the voting weight of class  $i$  at model  $k$ . The voting weight is determined by calculating the average distance between the centroid and the features of both normal and abnormal bottles (Step 6).

The final output of this method is prediction class that compute as follows:

$$h_i = \operatorname{argmin}_i \sum_k sd_{ik}, \quad (12)$$

where  $h_i$  is the class or category that the method predicts the input image belongs to base on its defect patterns and relationships from the training bottle images. The proposed classification method is illustrated in Fig. 7.

## IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we present a comprehensive set of experiments conducted to assess the performance of the CNNLSTM method.

### A. Experiment setup

These experiments used three glass bottle image datasets, namely the body dataset, the neck dataset, and the rotating dataset. These datasets consist of three primary types of defects: stone, tear, and blister. The body dataset of glass bottles comprises a total of 571 images. For the training set, 292 images, consisting of 161 images with defects and 131 images without defects were used to train models. The remaining 279 images constituted the testing set, with 207 images containing defects and 72 images without defects. The neck dataset encompasses 570 images of glass bottles. The training set include 288 images, with 150 images exhibiting defects and 138 images without defects. The testing set includes 282 images, with 100 images containing defects and 182 images without defects. The rotating dataset encompasses 120 images of glass bottles. The training set include 80 images, with 47 images exhibiting defects and 33 images without defects. The testing set includes 40 images, with 23 images containing defects and 17 images without defects. The distribution of each defect type

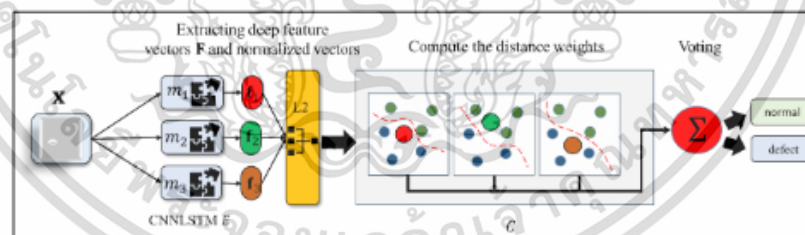


Fig. 7. The overall process of glass bottle classification

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

in both the training and testing datasets for body, neck and rotating bottles is presented in Table 1.

TABLE 1. NUMBER OF TRAINING AND TESTING IMAGES FOR EACH DEFECT TYPE IN THE BODY, NECK AND ROTATING DATASETS

Defect type	Body		Neck		Rotating	
	Train	Test	Train	Test	Train	Test
Stone	81	127	80	71	20	10
Tear	30	28	20	17	12	5
Blisters	50	52	50	12	15	8
Normal	131	72	138	182	33	17
Total	292	279	288	282	80	40

To assess the effectiveness of our proposed method on the defect dataset, various image processing techniques, including image thresholding and edge detection, along with machine learning models such as Anomaly Detection with Autoencoder (ADA), ResNet50, AlexNet, VGG16, and MobileNetV3 were used to construct models for comparison. Image thresholding, a commonly used technique in image processing using OpenCV, involves setting pixel values to either 0 or a maximum value based on a predefined threshold [13]. Edge detection is a digital image processing technique used to identify and extract edges in an image, representing abrupt changes in intensity or color between adjacent pixels [14]. ResNet50 is a deep convolutional neural network architecture introduced in 2016 [15]. ResNet-50 consists of 50 layers and employs residual blocks, each containing convolutional layers and shortcut connections. It has achieved state-of-the-art performance in various image recognition tasks. AlexNet is a deep convolutional neural network architecture proposed in 2012 [16]. AlexNet comprises eight layers, including five convolutional layers and three fully connected layers. It introduced several key innovations, such as using ReLU as an activation function and applying dropout regularization. VGG16 developed by the Visual Geometry Group (VGG) at the University of Oxford in 2014, is a widely recognized deep convolutional neural network architecture primarily used for image classification [17]. It consists of 13 convolutional layers followed by 3 fully connected layers, and it has demonstrated outstanding performance on numerous computer vision tasks. MobileNetV3, introduced by Google in 2019, is specifically designed for mobile and embedded devices that have limited computational resources [18]. This architecture aims to strike a balance between model size and accuracy, making it suitable for efficient deployment on devices with constrained hardware capabilities. In addition to these models, we also employed the Anomaly Detection with Autoencoder (ADA) technique. ADA detects anomalies by evaluating the reconstruction loss, comparing it to a predefined threshold. If the reconstruction loss surpasses the threshold, the input is classified as an anomaly.

For training our method and deep learning-based models, we used a training set comprising 292 images for the body dataset and 288 images for the neck dataset. The training epochs are set at 500 for VGG16, ResNet50, MobileNetV3,

and ADA. The input error threshold,  $\theta$ , of the proposed method is set to 0.2, and the number of ensemble models,  $K$ , is set to 3. All the experiments, including the proposed defect detection method and the comparison methods, were conducted on a personal computer equipped with an AMD Ryzen 7 5000 series processor and an NVIDIA GTX 1650 GPU, allowing for efficient computation and analysis of the results.

### B. Experimental Result

Four evaluation metrics are adopted for the analysis: Accuracy, Recall, Precision, and F1-score. These metrics are widely used in image defect detection and provide comprehensive insights into the performance of the methods. They are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (16)$$

$TP$  refers to the number of defect bottles (positive) that are correctly predicted as positive by the model.  $FN$  represents the number of defect bottles (positive) that are incorrectly predicted as bottle without defects (negative) by the model.  $TN$  indicates the number of the bottles without defect (negative) that are correctly predicted as negative by the model. Lastly,  $FP$  signifies the number of the bottles without defects (negative) that are mistakenly predicted as positive by the model.

Fig. 8 presents visualizations of defect classification results of CNNLSTM for the body, neck, and rotation of bottle datasets.

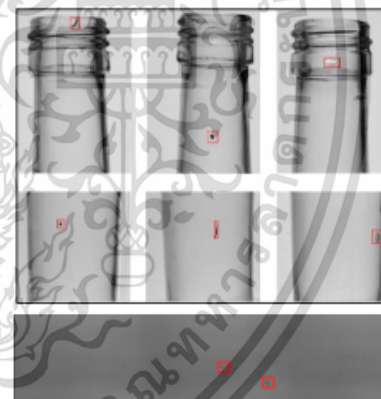


Fig. 8. Examples of classification results of the CNNLSTM

TABLE II. THE PERFORMANCE ON THE BODY IMAGES, NECK IMAGES AND ROTATING IMAGES DATASETS IN TERM OF ACCURACY

Methods	Body dataset	Neck dataset	Rotating dataset
	Accuracy (%)	Accuracy (%)	Accuracy (%)
CNNLSTM-ResNet50	71	95	62
CNNLSTM-AlexNet	75	89	97
CNNLSTM-MobileNetV3	64	92	98
CNNLSTM-VGG16	77	80	90
ResNet50	79	79	60
AlexNet	74	92	60
VGG16	80	85	95
MobileNetV3	70	72	70
ADA	74	75	60
Image threshold	62	65	65
Edge detection	70	72	63

The classification accuracy of various methods on the defect datasets is presented in Table II. The results for the body dataset demonstrate that the standalone VGG16 method outperformed all other methods, achieving an accuracy of 80%. ResNet50 is ranked second in terms of accuracy. Among the CNNLSTM models, CNNLSTM-VGG16 achieved the highest accuracy compared to the other CNNLSTM variants. Conversely, the image threshold method exhibited the lowest accuracy on this dataset.

The neck dataset results reveal that CNNLSTM-ResNet50 attained the highest accuracy of 95%, closely followed by CNNLSTM-AlexNet, CNNLSTM-MobileNetV3, CNNLSTM-VGG16, and AlexNet. Among the standalone methods, AlexNet exhibited the highest accuracy on this dataset. Conversely, the image threshold method yielded the lowest accuracy for the neck dataset.

Upon comparing the results of the rotating dataset images, it is obvious that CNNLSTM-MobileNetV3 achieved the highest accuracy. CNNLSTM-AlexNet and the standalone VGG16 also demonstrated good performance with high accuracies. Conversely, ADA exhibited the lowest accuracy percentages on this dataset.

Overall, the results indicate that different methods perform differently on the three datasets. The standalone VGG16 achieved the highest accuracy on the body dataset. CNNLSTM generally achieved the highest accuracy on both the neck and rotating datasets.

In Table III, the performance measures, including Recall, Precision, and F1-score, are presented for the body dataset. VGG16 demonstrates the highest level of correctness, evident from its superior metrics across Recall, Precision, and F1-score. Following closely in terms of correctness, CNNLSTM-AlexNet achieves relatively high Precision and F1-score. On the other hand, Image Threshold exhibits the lowest level of correctness among the methods listed.

TABLE III. THE PERFORMANCE ON THE BODY IMAGES DATASETS IN TERMS OF RECALL, PRECISION, AND F1-SCORE

Methods	Recall (%)	Precision (%)	F1-score (%)
CNNLSTM-ResNet50	80	73	70
CNNLSTM-AlexNet	82	74	74
CNNLSTM-MobileNetV3	71	66	63
CNNLSTM-VGG16	83	75	75
ResNet50	71	75	73
AlexNet	73	74	70
VGG16	86	78	77
MobileNetV3	57	57	57
ADA	51	55	53
Image Threshold	50	53	51
Edge detection	78	72	67

VGG16 outperforms the CNNLSTM method in terms of accuracy on the body image dataset. The VGG16 is advantaged by being pre-trained on a large and diverse collection of images. Its capacity to identify various features associated with bottle defects, coupled with the requirement for a comprehensive and diverse training set, contributes to its superior performance.

TABLE IV. THE PERFORMANCE ON THE NECK IMAGES DATASETS IN TERMS OF RECALL, PRECISION, AND F1-SCORE

Methods	Recall (%)	Precision (%)	F1-score (%)
CNNLSTM-ResNet50	95	92	96
CNNLSTM-AlexNet	89	85	92
CNNLSTM-MobileNetV3	92	90	93
CNNLSTM-VGG16	80	77	78
ResNet50	79	71	88
AlexNet	92	89	94
VGG16	85	79	90
MobileNetV3	72	62	80
ADA	75	70	75
Image Threshold	65	52	53
Edge detection	72	65	70

Table IV presents the Recall, Precision, and F1-score metrics of the proposed method and the compared methods for neck image datasets. Among the CNNLSTM methods, CNNLSTM-ResNet50 demonstrates the highest recall, precision, and F1-score, indicating its strong performance. CNNLSTM-MobileNetV3 follows closely behind with good recall, precision, and F1-score. Among the individual models, AlexNet demonstrates good performance, achieving high recall, precision, and F1-score. MobileNetV3 exhibits moderate performance with decent recall and F1-score, but its precision is relatively lower. ADA demonstrates moderate performance across all metrics. On the other hand, Image Threshold and Edge detection methods exhibit lower performance, with lower recall, precision, and F1-score.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

TABLE V. THE PERFORMANCE ON THE ROTATING IMAGES DATASETS IN TERMS OF RECALL, PRECISION, AND F1-SCORE

Methods	Recall (%)	Precision (%)	F1-score (%)
CNNLSTM-ResNet50	66	71	61
CNNLSTM-AlexNet	97	96	96
CNNLSTM-MobileNetV3	97	98	98
CNNLSTM-VGG16	91	90	89
ResNet50	55	58	52
AlexNet	52	79	42
VGG16	95	94	94
MobileNetV3	62	79	52
ADA	58	55	56
Image Threshold	77	69	63

Table V presents the Recall, Precision, and F1-score metrics of the proposed method and the compared methods for rotating image datasets. Among the CNNLSTM methods, CNNLSTM-MobileNetV3 and CNNLSTM-AlexNet demonstrate outstanding performance, achieving high recall, precision, and F1-score. CNNLSTM-VGG16 also performs well, with consistently high metrics. However, CNNLSTM-ResNet50 exhibits comparatively lower performance across the metrics. When considering individual models, VGG16 stands out with consistently high scores for recall, precision, and F1-score. ResNet50, AlexNet, and MobileNetV3 display varying levels of performance, with some metrics being higher or lower than others. In terms of the additional methods, Image Threshold exhibits relatively high recall but lower precision and F1-score. ADA demonstrates quite low performance across the metrics.

TABLE VI. THE TRAINING TIME AND DETECTION TIME ON THE BODY IMAGES AND NECK IMAGES DATASETS

Methods	Body dataset		Neck dataset		Rotating dataset	
	Training time (second)	Detection time (second)	Training time (second)	Detection time (second)	Training time (second)	Detection time (second)
CNNLSTM-ResNet50	154.29	0.11	148.55	0.054	65.4	0.054
CNNLSTM-AlexNet	161.59	0.12	59.23	0.005	67.12	0.009
CNNLSTM-MobileNetV3	141.93	0.12	80.69	0.001	60.28	0.002
CNNLSTM-VGG16	304.41	0.074	304.44	0.025	161.44	0.025
ResNet50	1,500	0.5	1,500	0.5	1,500	0.5
AlexNet	300	0.3	300	0.3	300	0.3
VGG16	1,800	0.5	1,800	0.5	1,800	0.5
MobileNetV3	700	0.3	700	0.3	700	0.3
ADA	100	0.2	100	0.2	100	0.2
Image Threshold	-	0.002	-	0.002	-	0.002
Edge detection	-	0.001	-	0.001	-	0.001

Table VI displays the training time and detection time of all detection methods. Among the CNNLSTM methods, CNNLSTM-VGG16 has the longest training time on all three datasets, followed by CNNLSTM-ResNet50. However, CNNLSTM-ResNet50 has the shortest detection time across all datasets, indicating faster inference speed. CNNLSTM-AlexNet and CNNLSTM-MobileNetV3 also demonstrate relatively shorter training and detection times. For the individual models, ResNet50, AlexNet, VGG16, and MobileNetV3 have consistent training and detection times across all three datasets. Among the additional methods, ADA has relatively short training and detection times, while Image Threshold and Edge detection methods have negligible training times but slightly longer detection times compared to the CNNLSTM and individual models.

Overall, the experimental results demonstrate the effectiveness of the proposed CNNLSTM models for defect detection on all three datasets. The proposed models achieved competitive accuracy and performed well in terms of evaluation metrics. The training and detection times varied among the models, with CNNLSTM-AlexNet demonstrating the shortest training time. These findings can guide the selection of appropriate models based on the trade-off between accuracy and computational time requirements in practical applications.

## V. CONCLUSIONS

This research presents an approach for detecting defects in glass bottles using a combination of deep convolutional neural networks with long short-term memory (CNNLSTM) and instance-based classification algorithms. The CNNLSTM combines two types of deep neural networks to extract features and create a representation of class data, which is then used for defect detection. The proposed method is compared with other well-known defect detection methods. Experimental results demonstrate that the proposed method outperforms other defect detection methods in terms of detection performance. Additionally, the proposed method requires less training time, making it suitable for efficient glass bottle defect detection in production lines.

In future work, we intend to extend the application of the proposed algorithm to different types of bottles and explore additional defect types. The flexibility of the CNNLSTM allows for varying the number of models and designing customized detection layers to adapt to specific datasets. Furthermore, there is potential to extend this method to semi-supervised learning, enabling the detection of defects in other product categories. These future directions will further enhance the capabilities and applicability of our approach in defect detection and contribute to the advancement of quality control systems in various industries.

## REFERENCES

- [1] H. Xie, F. Lu, G. Ouyang, X. Sheng and Z. Zhao, "A rapid inspection method for encapsulating quality of PET bottles based on machine vision," *IEEE International Conference on Computer and Communications (ICCC)*, Chengdu, China, pp. 2025-2028, 2017.
- [2] L. Fu, S. Zhang, Y. Gong and Q. Huang, "Medicine Glass Bottle Defect Detection Based on Machine Vision," *Chinese Control and Decision Conference (CCDC)*, Nanchang, China, pp. 5681-5685, 2019.

- [3] M. A. E. Latina, J. Van Russel R. Dela Cruz and F. D. Delos Santos, "Empty Glass Bottle Defect Detection Based on Deep Learning with CNN Using SSD MobileNetV2 Model," *IEEE 14th International Conference on Humanoid Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, Boracay Island, Philippines, pp. 1-6, 2022.
- [4] W. Gong, K. Zhang, C. Yang, M. Yi and J. Wu, "Adaptive Visual Inspection Method for Transparent Label Defect Detection of Curved Glass Bottle," *International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, Chongqing, China, pp. 90-95, 2020.
- [5] G. A. De Vitis, A. Di Tecco, P. Foglia, and C. A. Prete, "Fast Blob and Air Line Defects Detection for High Speed Glass Tube Production Lines," *Journal of Imaging*, vol. 7, no. 11, p. 223, Oct. 2021.
- [6] T. Zhang, B. Ding, X. Zhao, G. Liu, Z. Pang, "LearningADD: Machine learning based acoustic defect detection in factory automation," *Journal of Manufacturing Systems*, vol. 60, pp. 48-58, 2021.
- [7] X. Zhou, Y. Wang, Q. Zhu, J. Mao, C. Xiao, X. Lu and H. Zhang, "A Surface Defect Detection Framework for Glass Bottle Bottom Using Visual Attention Model and Wavelet Transform," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2189-2201, 2020.
- [8] X. Zhou, Y. Wang, Q. Zhu, J. Mao, C. Xiao, X. Lu, and H. Zhang, "Automated Visual Inspection of Glass Bottle Bottom With Saliency Detection and Template Matching," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 11, pp. 4253-4267, 2019.
- [9] S. Montaha, S. Azam, A. K. M. R. H. Rafid, M. Z. Hasni, A. Karim and A. Islam, "TimeDistributed-CNN-LSTM: A Hybrid Approach Combining CNN and LSTM to Classify Brain Tumor on 3D MRI Scans Performing Ablation Study," in *IEEE Access*, vol. 10, pp. 60039-60059, 2022.
- [10] S. Xiang and B. Tang, "CSLM: Convertible Short-Term and Long-Term Memory in Differential Neural Computers," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 9, pp. 4026-4038, 2021.
- [11] V. Nair, and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 807-814, 2010.
- [12] L. Bottou, "Large-Scale Machine Learning with Stochastic Gradient Descent," In: Lechevallier, Y., Saporta, G. (eds) *Proceedings of COMPSTAT'2010*. Physica-Verlag HD, 2010.
- [13] R. C. Gonzalez and R. E. Woods, *Digital Image Processing* (3th ed.), Pearson International Edition, 2008.
- [14] H. Singh, and N. Kaur, "A review of edge detection techniques for image segmentation," *Journal of Computational and Theoretical Nanoscience*, vol. 15, no. 9, pp.4131-414, 2018.
- [15] K. He, X. Zhang, S. Ran and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778, 2016.
- [16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM* 60, vol. 6, pp. 84-90, 2017.
- [17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Proc. Int. Conf. Learn. Represent.*, San Diego, CA, USA, pp. 1409-1556, 2014.
- [18] A. Howard, M. Sandler, B. Chen, W. Wang, L. C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, F. Pang, H. Adam and Q. Le, "Searching for MobileNetV3," *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), pp. 1314-1324, 2019.

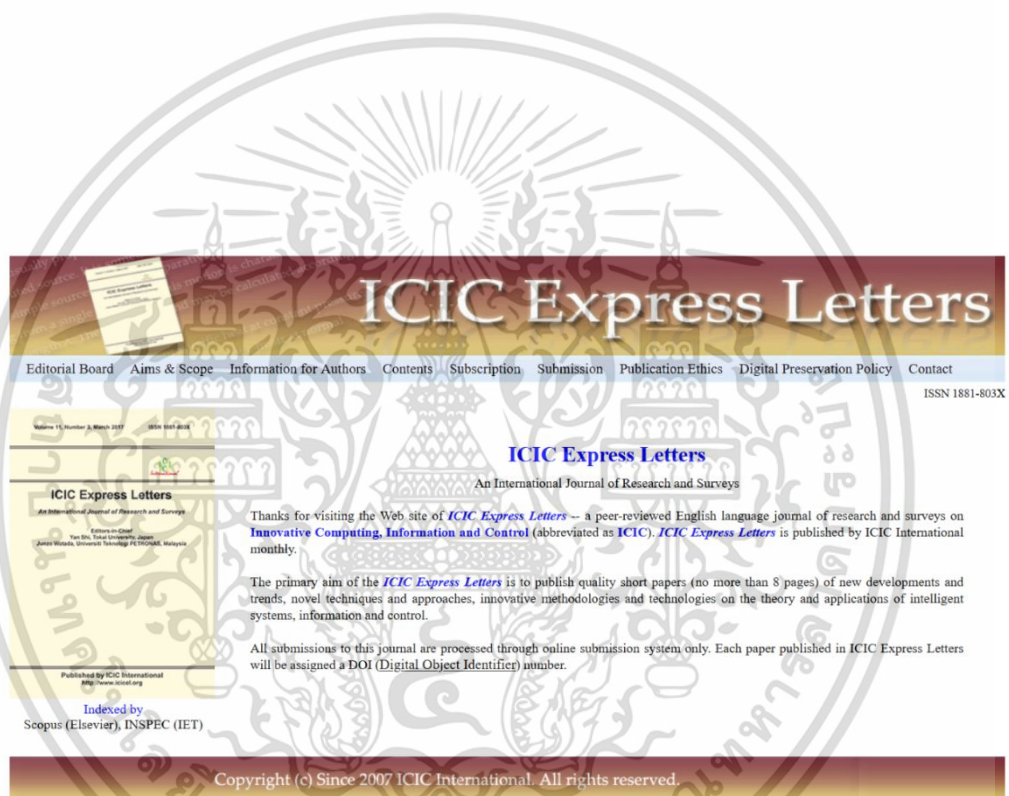


เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข  
ข้อมูลวารสารที่ได้รับการตีพิมพ์



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



**ICIC Express Letters**

Editorial Board Aims & Scope Information for Authors Contents Subscription Submission Publication Ethics Digital Preservation Policy Contact

ISSN 1881-803X

Volume 11, Number 3, March 2017 ISSN 1881-803X

**ICIC Express Letters**  
An International Journal of Research and Surveys

Thanks for visiting the Web site of *ICIC Express Letters* – a peer-reviewed English language journal of research and surveys on **Innovative Computing, Information and Control** (abbreviated as **ICIC**). *ICIC Express Letters* is published by ICIC International monthly.

The primary aim of the *ICIC Express Letters* is to publish quality short papers (no more than 8 pages) of new developments and trends, novel techniques and approaches, innovative methodologies and technologies on the theory and applications of intelligent systems, information and control.

All submissions to this journal are processed through online submission system only. Each paper published in *ICIC Express Letters* will be assigned a DOI (Digital Object Identifier) number.

Published by ICIC International  
http://www.icicet.org

Indexed by  
Scopus (Elsevier), INSPEC (IET)

Copyright (c) Since 2007 ICIC International. All rights reserved.  
辽ICP备12003320号-1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



Scopus Preview

## Source details

[Feedback >](#) [Compare sources >](#)

## ICIC Express Letters

Scopus coverage years: from 2009 to Present

Publisher: ICIC Express Letters Office

ISSN: 1881-803X

Subject area: [Computer Science: General Computer Science](#) [Engineering: Control and Systems Engineering](#)

Source type: Journal

[View all documents >](#)
[Set document alert](#)
[Save to source list](#)

CiteScore 2022

1.4

SJR 2022

0.261

SNIP 2022

0.329

[CiteScore](#) [CiteScore rank & trend](#) [Scopus content coverage](#)

## Improved CiteScore methodology

CiteScore 2022 counts the citations received in 2019-2022 to articles, reviews, conference papers, book chapters and data papers published in 2019-2022, and divides this by the number of publications published in 2019-2022. [Learn more >](#)

CiteScore 2022

1.4

864 Citations 2019 - 2022

599 Documents 2019 - 2022

Calculated on 05 May, 2023

CiteScoreTracker 2023

1.2

652 Citations to date

522 Documents to date

Last updated on 05 July, 2023 • Updated monthly

CiteScore rank 2022

Category Rank Percentile

Computer Science  
 General Computer Science #167/233 28th

Engineering  
 Control and Systems Engineering #219/286 23rd

[View CiteScore methodology >](#) [CiteScore FAQ >](#) [Add CiteScore to your site >](#)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SJR Scimago Journal & Country Rank

Enter Journal Title, ISSN or Publisher Name

Home Journal Rankings Country Rankings Viz Tools Help About Us

### ICIC Express Letters

**COUNTRY**  
Japan  
Universities and research institutions in Japan  
Media Ranking in Japan

**SUBJECT AREA AND CATEGORY**  
Computer Science  
Computer Science (miscellaneous)  
Engineering  
Control and Systems Engineering

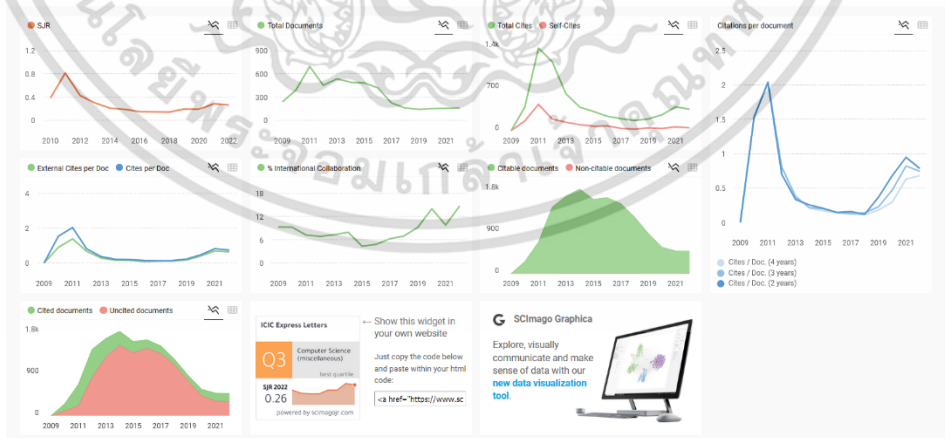
**PUBLISHER**  
ICIC Express Letters Office

**H-INDEX**  
23

**SCOPE**

The ICIC Express Letters is a peer reviewed English language journal of research and surveys on Innovative Computing, Information and Control (abbreviated as ICIC), and is published by ICIC International monthly. The primary aim of the ICIC Express Letters is to publish quality short papers (no more than 8 pages) of new developments and trends, novel techniques and approaches, innovative methodologies and technologies on the theory and applications of intelligent systems, information and control. The scope of the ICIC Express Letters is in the broad areas of intelligent systems, information and control, includes but not limited to: -Intelligent Systems and Innovative Computing: Artificial intelligence and expert systems, approximate reasoning, fuzzy systems and rough sets, fuzzy optimization and decision making, evolutionary computation, intelligent information systems, image and speech signal processing, human-machine learning, neural networks, knowledge discovery, data mining, clustering and data analysis, decision support systems, intelligent fault detection and identification, knowledge-based systems, hybrid intelligent systems, soft computing, management science. -Information and Control: Information theory and applications, optimal, adaptive, predictive, robust, fuzzy and reliable control and filtering on linear and nonlinear systems, stochastic systems, hybrid systems, supervisory systems, guidance and navigation systems, real-time systems, discrete event systems, signal and image processing, system identification, optimization techniques, robustness analysis, learning systems, fault detection, robotics and motion control, manufacturing systems, power systems, networked control systems, environmental and bio-engineering, complex and large scale systems, bioinformatics, control education, transportation systems.

Join the conversation about this journal



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



The Science and Information Organization

Q

---

HOME
ABOUT US
JOURNALS
CONFERENCES
CONTACT US

---

Home
Call for Papers
Indexing
Submit your Paper
Guidelines
Fees
Current Issue
Archives
Editors
Reviewers
Subscribe

---

## International Journal of Advanced Computer Science and Applications (IJACSA)

Computer Science Journals - Open Access - Scopus Indexed

---

IJACSA is a scholarly computer science journal representing the best in research. Its mission is to provide an outlet for quality research to be publicised and published to a global audience. The journal aims to publish papers selected through rigorous double-blind peer review to ensure originality, timeliness, relevance, and readability.

In sync with the Journal's vision "to be a respected publication that publishes peer reviewed research articles, as well as review and survey papers contributed by international community of Authors", we have drawn reviewers and editors from institutions and universities across the globe. A double blind peer review process is conducted to ensure that we retain high standards. At IJACSA, we stand strong because we know that global challenges make way for new innovations, new ways and new talent.

International Journal of Advanced Computer Science and Applications publishes research, review and survey papers which offer a significant contribution to the computer science literature, and which are of interest to a wide audience. Coverage extends to all main-stream branches of computer science and related applications.

Open Access Statement: This is an open access journal which means that all content is freely available without charge to the user or his/her Institution. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles, or use them for any other lawful purpose, without asking prior permission from the publisher or the author. All past issues are available in the [Archives](#).

\*\* Indexing: The papers of this journal are submitted to Web of Science, Scopus, Inspec, Ei Compendex, Semantic Scholar, EBSCOhost, Microsoft Academic Search and Index Copernicus. The Journal is Scientific level 1 in the Norwegian Register and included in the Federation of Finnish Learned Societies/Publication Forum.\*\*

CALL FOR PAPERS
SUBMIT YOUR PAPER
CURRENT ISSUE



### Upcoming Conferences



**Future of Information and Communication Conference (FICC) 2024**  
4-5 April 2024 Berlin, Germany



**Computing Conference 2024**  
11-12 July 2024 London, United Kingdom



**Intell Sys 2023**  
7-8 September 2023 Amsterdam, The Netherlands



**Future Technologies Conference (FTC) 2023**  
2-3 November 2023 San Francisco, United States

### Bibliographic Information

- ✓ ISSN : 2156-5570 (Online)
- ✓ ISSN : 2156-107X (Print)
- ✓ DOI : [10.14569/issn.2156-5570](https://doi.org/10.14569/issn.2156-5570)
- ✓ Publication Frequency : Monthly

**Impact Factor**  
✓ 0.9

**CiteScore**  
✓ 2.1 (2022) Q3

**Dr Kohei Arai**  
Editor-in-Chief





Affiliation : Saga University, Japan.

Domains of Research : Computer Science, Human-Computer Interaction, Networking, Information Retrievals, Optimization Theory, Modelling and Simulation, Satellite Remote Sensing, Computer Vision, Decision Making, Methodology

[editor@ijacsa.org](mailto:editor@ijacsa.org)


BACK TO TOP

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้


HOME | ABOUT US | **JOURNALS** | CONFERENCES | CONTACT US

Home | Call for Papers | Indexing | Submit your Paper | Guidelines | Fees | Current Issue | Archives | Editors | Reviewers | [Subscribe](#)

## Indexing

To be recognized as an authoritative, high-quality source of information, a journal must be widely available. Authors publish with us for the heightened visibility, research activity, and industry credibility.

### IJACSA is covered by following databases and archives:

- Scopus (Q3 | Cite Score 2.1)
- Web of Science (Clarivate | JIF 0.9)
- Inspec (IET)
- Ei Compendex
- Norwegian Register for Scientific Journals, Series and Publishers (NSD)
- Julkaisuforum Publication Forum (Federation of Finnish Learned Societies)
- Genamics JournalSeek
- Semantic Scholar
- EBSCOhost
- Ex Libris
- Microsoft Academic Search
- WorldCat
- CNKI

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## Source details

[Feedback >](#) [Compare sources >](#)

## International Journal of Advanced Computer Science and Applications

Scopus coverage years: from 2017 to Present

Publisher: Science and Information Organization

ISSN: 2158-107X E-ISSN: 2156-5570

Subject area: [Computer Science: General Computer Science](#)

Source type: Journal

[View all documents >](#)
[Set document alert](#)
[Save to source list](#) [Source Homepage](#)
CiteScore 2022  
2.1SJIR 2022  
0.258SNIP 2022  
0.512
[CiteScore](#) [CiteScore rank & trend](#) [Scopus content coverage](#)

## Improved CiteScore methodology

CiteScore 2022 counts the citations received in 2019-2022 to articles, reviews, conference papers, book chapters and data papers published in 2019-2022, and divides this by the number of publications published in 2019-2022. [Learn more >](#)

CiteScore 2022

$$2.1 = \frac{9,537 \text{ Citations 2019 - 2022}}{4,553 \text{ Documents 2019 - 2022}}$$

Calculated on 05 May, 2023

CiteScoreTracker 2023

$$1.8 = \frac{7,294 \text{ Citations to date}}{4,104 \text{ Documents to date}}$$

Last updated on 08 July, 2023 • Updated monthly

CiteScore rank 2022

Category

Rank - Percentile

Computer Science  
General Computer Science

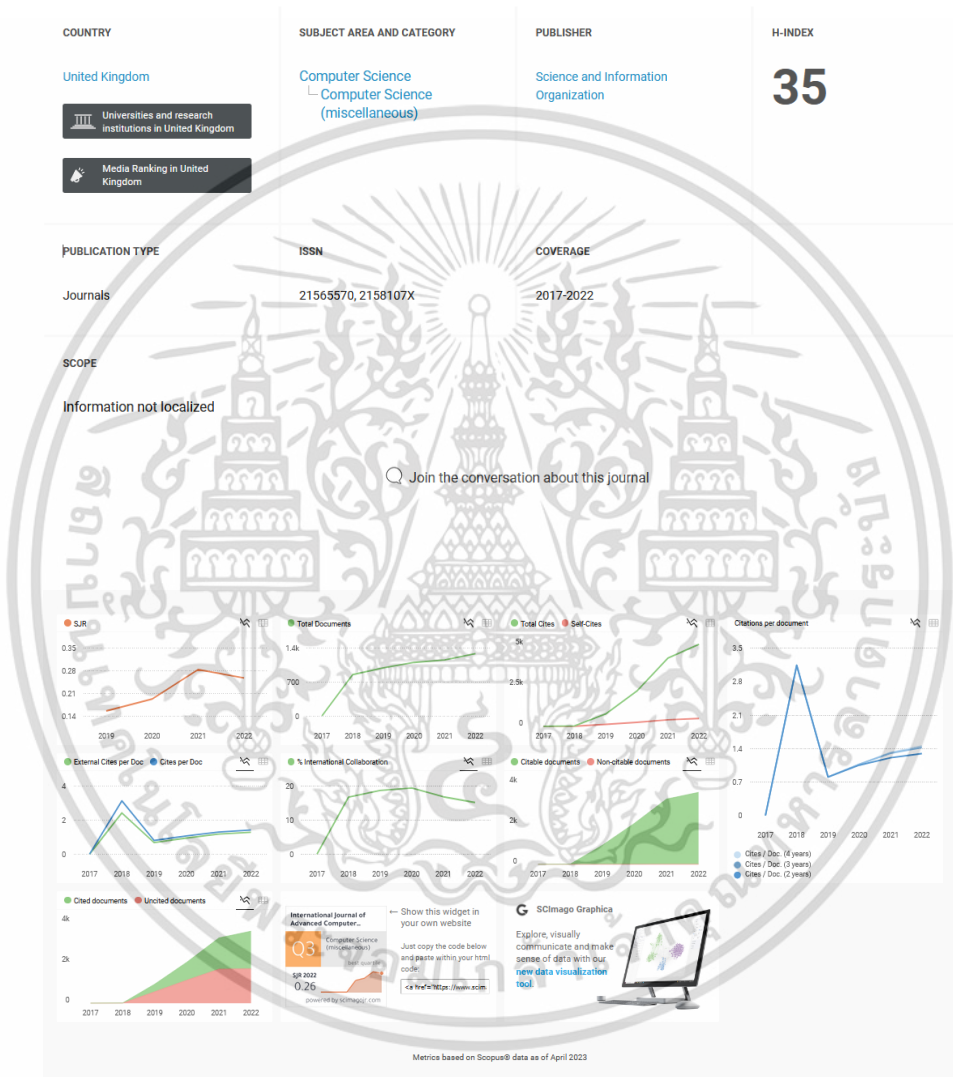
#129/233 — 44th

[View CiteScore methodology >](#) [CiteScore FAQ >](#) [Add CiteScore to your site >](#)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



## International Journal of Advanced Computer Science and Applications



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ประวัติผู้เขียน

ชื่อ นายนิพัทธ์ คล้ายโพธิ์  
 วัน เดือน ปีเกิด 12 กันยายน 2532  
 ที่อยู่ปัจจุบัน 37 หมู่ 6 ตำบลบ้านกร่าง อำเภอศรีประจันต์ จังหวัดสุพรรณบุรี 72140  
 ประวัติการศึกษา พ.ศ.2554 บริหารธุรกิจบัณฑิต (บธ.บ.)

สาขาวิชาการระบบสารสนเทศ (คอมพิวเตอร์ธุรกิจ) คณะบริหารธุรกิจ  
 มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี เขตตลิ่งชัน 3.12  
 พ.ศ.2557 วิทยาศาสตรมหาบัณฑิต (วท.ม.)

สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์  
 สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
 เขตตลิ่งชัน 3.18



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้