การตรวจจับโรคปอดโดยใช้สัญญาณเสียงจากปอดผ่านโครงข่ายประสาทเทียมลึกแบบหลายงาน

นางสาวขวัญจิต ออกเวหา

มหาวิทยาลัยเทคโนโลยีสุรนารี

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาปรัชญาดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมโทรคมนาคมและคอมพิวเตอร์
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2568

PULMONARY DISEASE DETECTION USING LUNG SOUND SIGNALS WITH MULTI-TASK DEEP NEURAL NETWORKS

KHWANJIT ORKWEHA

มหาวิทยาลัยเทคโนโลยีสุรนารี

A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in Telecommunication
and Computer Engineering
Suranaree University of Technology
Academic Year 2025

การตรวจจับโรคปอดโดยใช้สัญญาณเสียงจากปอดผ่านโครงข่ายประสาทเทียมลึกแบบหลายงาน

มหาวิทยาลัยเทคโนโลยีสุรนารี อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

คณะกรรมการสอบวิทยานิพนธ์

.....

(รองศาสตราจารย์ ดร. จักรี ศรีนนท์ฉัตร)
ประธานกรรมการ

.....

(ศาสตราจารย์ ดร. พีระพงษ์ อุฑารสกุล)
กรรมการ (อาจารย์ที่ปรึกษาวิทยานิพนธ์)

.....

(รองศาสตราจารย์ ดร. คมเดช ภาพัฒนบุรี)
กรรมการ (อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม)

.....

(ผู้ช่วยศาสตราจารย์ ดร. เศรษฐวิทย์ ภูฉายา)
กรรมการ

.....

(ศาสตราจารย์ ดร. มนต์ทิพย์ภา อุฑารสกุล)
กรรมการ

.....

(รองศาสตราจารย์ ดร. นุชนาฏ สันทาลุนัย)
กรรมการ

.....

(รองศาสตราจารย์ ดร.ยุพาพร รักสกุลพิวัฒน์)
รักษาการแทนรองอธิการบดีฝ่ายวิชาการและประกันคุณภาพ

.....

(รองศาสตราจารย์ ดร.พรศิริ จงกล)
คณบดีสำนักวิชาวิศวกรรมศาสตร์

มหาวิทยาลัยเทคโนโลยีสุรนารี

ขวัญจิต ออกเวหา:การตรวจจับโรคปอดโดยใช้สัญญาณเสียงจากปอดผ่านโครงข่ายประสาทเทียมลึกแบบหลายงาน (PULMONARY DISEASE DETECTION USING LUNG SOUND SIGNALS WITH MULTI-TASK DEEP NEURAL NETWORKS)
อาจารย์ที่ปรึกษา: ศาสตราจารย์ ดร.พีระพงษ์ อุฑารสกุล, 96 หน้า.

คำสำคัญ : การวินิจฉัยด้วยสัญญาณเสียง ออโตเอนโคเดอร์แบบหลายงาน การเรียนรู้เชิงลึก

วิทยานิพนธ์ฉบับนี้นำเสนอวิธีการสำหรับการตรวจจับและจำแนกโรคระบบทางเดินหายใจจากสัญญาณเสียงปอด (Lung sounds) โดยพัฒนาแบบจำลองออโตเอนโคเดอร์แบบหลายงาน (Multi-Task Autoencoder) เพื่อเรียนรู้ตัวแทนเชิงแฝงที่คงสารสนเทศสำคัญของสัญญาณ พร้อมเพิ่มความสามารถในการจำแนกแบบหลายกลุ่ม และเสนอแนวทางไฮบริดด้วยการผสานแบบจำลองออโตเอนโคเดอร์แบบหลายงานกับซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เพื่อปรับปรุงประสิทธิภาพในการจำแนกประเภท

การประเมินประสิทธิภาพของวิธีการที่นำเสนอโดยใช้เทคนิคการตรวจสอบความถูกต้องแบบไขว้ 5 ส่วน (5-fold cross-validation) บนชุดข้อมูลเสียงปอดที่เปิดเผยสาธารณะจาก King Abdullah University Hospital (KAUH) สำหรับการจำแนกประเภทแบบ 4 กลุ่ม ได้แก่ ภาวะปกติ โรคปอดอักเสบ โรคหอบหืด และโรคปอดอุดกั้นเรื้อรัง พบว่าแบบจำลองออโตเอนโคเดอร์แบบหลายงานให้ค่าความถูกต้องรวม 89.47 เปอร์เซ็นต์ และค่าคะแนน F1 89.96 เปอร์เซ็นต์ ขณะที่แนวทางไฮบริดกับซัพพอร์ตเวกเตอร์แมชชีน ค่าความถูกต้องรวมสูงสุด 91.49 เปอร์เซ็นต์ และค่าคะแนน F1 92.13 เปอร์เซ็นต์ ผลลัพธ์สะท้อนว่าการผสานตัวจำแนกซัพพอร์ตเวกเตอร์แมชชีนช่วยเพิ่มความชัดเจนและความเสถียรของขอบเขตการตัดสินใจบนคุณลักษณะเชิงแฝง การจำแนกโดยรวมมีความถูกต้องและความสมดุลสูงขึ้นอย่างมีนัยสำคัญ และมีศักยภาพต่อการนำไปประยุกต์ใช้งานในอนาคต

สาขาวิชาวิศวกรรมโทรคมนาคม
ปีการศึกษา 2568

ลายมือชื่อนักศึกษา ขวัญจิต ออกเวหา
ลายมือชื่ออาจารย์ที่ปรึกษา .....
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

KHWANJIT ORKWEHA: PULMONARY DISEASE DETECTION USING LUNG SOUND SIGNALS WITH MULTI-TASK DEEP NEURAL NETWORKS.
THESIS ADVISOR: PROF. PEERAPONG UTHANSAKUL, Ph.D., 96 PP.

Keyword: AUDIO DIAGNOSTICS MULTI-TASK AUTOENCODER DEEP LEARNING

This thesis proposes an approach for detecting and classifying respiratory diseases from lung sound signals. The proposed method develops a Multi-Task Autoencoder (MTAE) to learn latent representations that preserve essential signal information while enhancing multi-class classification capability. In addition, a hybrid approach is introduced by integrating the MTAE with a Support Vector Machine (SVM) to improve the classification performance

An evaluation was conducted using 5-fold cross-validation on the King Abdullah University Hospital (KAUH) lung sound dataset for four-class classification: normal, pneumonia, asthma, and chronic obstructive pulmonary disease. The MTAE achieved an overall accuracy of 89.47% and F1-score of 89.96%, whereas the hybrid MTAE–SVM achieved the highest overall accuracy of 91.49% and F1-score of 92.13%. The results indicate that incorporating an SVM classifier enhances the clarity and stability of the decision boundary in the latent feature space, leading to significantly improved overall accuracy and more balanced classification performance, and demonstrating strong potential for future practical applications.

มหาวิทยาลัยเทคโนโลยีสุรนารี

School of Telecommunication Engineering
Academic Year 2025

Student's Signature ขวัญจิต อ่อกเมฆา
Advisor's Signature .....
Co-advisor's Signature .....

สารบัญ

หน้า

บทคัดย่อ (ภาษาไทย) ..... ก

บทคัดย่อ (ภาษาอังกฤษ) ..... ข

กิตติกรรมประกาศ ..... ค

สารบัญ ..... ง

สารบัญตาราง ..... ช

สารบัญรูป ..... ซ

บทที่

1 บทนำ 1

1.1 ความเป็นมาและความสำคัญ 1

1.2 วัตถุประสงค์ของงานวิจัย 3

1.3 ขอบเขตของงานวิจัย 3

1.4 จุดเด่นของงานวิจัยวิทยานิพนธ์ 4

1.5 ประโยชน์ที่คาดว่าจะได้รับ 4

1.6 โครงสร้างและการจัดลำดับเนื้อหาวิทยานิพนธ์ 4

2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 6

2.1 พื้นฐานสรีรวิทยาและสัญญาณเสียงปอด 6

2.1.1 สรีรวิทยาของระบบทางเดินหายใจ 6

2.1.2 การส่งผ่าน การตรวจจับ และการฟังเสียงปอด 9

2.1.3 ประเภทของเสียงปอดและลักษณะเชิงความถี่ของสัญญาณ 10

2.1.4 โรคที่เกี่ยวข้องกับระบบทางเดินหายใจและความสัมพันธ์กับเสียงปอด 11

2.1.5 แหล่งที่มาของข้อมูล 12

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญ (ต่อ)

หน้า

2.2 การสกัดคุณลักษณะ....13
2.2.1 การเตรียมสัญญาณเพื่อรองรับการสกัดคุณลักษณะ....14
2.2.2 สัมประสิทธิ์เซปสตรัมความถี่เมล....15
2.2.3 คุณลักษณะเชิงพลวัตที่สะท้อนการเปลี่ยนแปลงเชิงเวลา....16
2.2.4 การจัดรูปคุณลักษณะเพื่อใช้งานกับแบบจำลอง....17
2.3 วิธีการที่มีอยู่เดิม....18
2.3.1 โครงข่ายประสาทเทียมแบบคอนโวลูชัน....18
2.4 วิธีการที่นำเสนอ....22
2.4.1 ออโตเอนโคเดอร์แบบหลายงาน....23
2.4.2 แบบจำลองไฮบริด....27
2.5 ปริทัศน์วรรณกรรม....31

3 วิธีการดำเนินการวิจัย....36

3.1 ชุดข้อมูลและการเตรียมสัญญาณเสียงปอด....36
3.1.1 ชุดข้อมูลเสียงปอด....36
3.1.2 การเตรียมสัญญาณและการแบ่งช่วงสัญญาณ....38
3.1.3 การจัดการข้อมูลไม่สมดุล....39
3.2 การออกแบบวิธีการที่นำเสนอ....41
3.2.1 ออโตเอนโคเดอร์แบบหลายงาน....41
3.2.2 ออโตเอนโคเดอร์แบบหลายงานร่วมกับซัพพอร์ตเวกเตอร์แมชชีน....44
3.3 ทรัพยากรที่ใช้ในงานวิจัยวิทยานิพนธ์....47
3.4 การประเมินประสิทธิภาพ....48
3.4.1 เมตริกซ์สับสนและตัวชี้วัดการประเมินผลรายกลุ่ม....48
3.4.2 การวิเคราะห์เชิงภาพด้วยเทคนิค t-SNE....51

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญ (ต่อ)

หน้า

4 ผลการวิจัยและอภิปรายผล 53

4.1 ผลการดำเนินงานวิจัยในส่วนของการหาความยาวของเสียง 53

4.2 ผลการดำเนินงานวิจัยในส่วนของการเลือกอัตราการเรียนรู้ 55

4.3 ผลการดำเนินงานวิจัยในส่วนการกำหนดค่าการเข้ารหัสและการถอดรหัส 56

4.4 ผลการดำเนินงานวิจัยในส่วนของจำนวนชั้นเชื่อมโยงสมบูรณ์ 57

4.5 การเปรียบเทียบแบบจำลอง 58

4.6 ผลการดำเนินงานในส่วนการเรียนรู้หลานงายร่วมกับซัพพอร์ตเวกเตอร์แมชชีน 61

4.7 ผลการเปรียบเทียบกับแนวทางที่มีอยู่เดิมตามงานวิจัยอ้างอิง 64

5 สรุปและข้อเสนอแนะ 67

5.1 สรุปผลการวิจัย 67

5.2 อภิปรายผลการวิจัย 68

5.3 ข้อจำกัดของงานวิจัย 69

5.4 ข้อเสนอแนะและแนวทางการพัฒนาในอนาคต 69

รายการอ้างอิง 70

ภาคผนวก

ภาคผนวก ก บทความวิชาการที่ได้รับการตีพิมพ์เผยแพร่ในระหว่างการศึกษา 76

ประวัติผู้เขียน 96

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญตาราง

ตารางที่ **หน้า**

2.1 แนวคิดเชิงกลไกของการเกิดเสียงปอด..... 7

3.1 ข้อมูลประชากรของกลุ่มตัวอย่างจากฐานข้อมูล KAUH..... 37

3.2 พารามิเตอร์การสกัดคุณลักษณะด้วยสัมประสิทธิ์เซปสตรัมความถี่เมล..... 42

3.3 สถาปัตยกรรมของแบบจำลองออโตเอนโคเดอร์แบบหลายงาน..... 43

3.4 พารามิเตอร์สำคัญที่ใช้ในการฝึกแบบจำลอง..... 47

4.1 ผลของความยาวเฟรมของแบบจำลองออโตเอนโคเดอร์แบบหลายงาน..... 54

4.2 ผลการทดลองของอัตราการเรียนรู้..... 55

4.3 ประสิทธิภาพของการกำหนดชั้นการเข้ารหัสและการถอดรหัส..... 57

4.4 ประสิทธิภาพของชั้นเชื่อมโยงสมบูรณ์..... 58

4.5 ผลการทดลองแบบจำลอง..... 59

4.6 ผลการประเมินประสิทธิภาพของแบบจำลองในการจำแนกโรคปอด..... 63

4.7 การเปรียบเทียบผลการจำแนกโรคปอดรายกลุ่มโรค..... 63

4.8 การเปรียบเทียบผลการจำแนก 4 กลุ่มกับแนวทางที่มีอยู่เดิม..... 65

มหาวิทยาลัยเทคโนโลยีสุรนารี

สารบัญรูป

รูปที่		หน้า
2.1	โครงสร้างของระบบทางเดินหายใจ	7
2.2	ส่วนประกอบของหูฟังแพทย์ และตำแหน่งมาตรฐานสำหรับการตรวจฟังเสียง	9
2.3	ตัวอย่างเสียงปอดภาวะปกติและเสียงปอดภาวะผิดปกติ	10
2.4	การเลื่อนเคอร์เนลบนอินพุตและการสร้างแผนที่คุณลักษณะ	19
2.5	สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบคอนโวลูชันสองชั้น	22
2.6	ออโตเอนโคเดอร์	24
2.7	การเรียนรู้แบบหลายงาน	25
3.1	การแบ่งข้อมูลแบบการตรวจสอบไขว้แบบห้าส่วน	40
3.2	แบบจำลองออโตเอนโคเดอร์แบบหลายงานร่วมกับซัพพอร์ตเวกเตอร์แมชชีน	46
3.3	เมตริกซ์สับสนแบบจำแนก 4 กลุ่ม	48
3.4	ตัวอย่างการแสดงภาพด้วยเทคนิค t-SNE	51
4.1	การกระจายภาพ t-SNE ของคุณลักษณะแบบต่าง ๆ	60
4.2	เมตริกซ์สับสนของแบบจำลอง	62

มหาวิทยาลัยเทคโนโลยีสุรนารี

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี อันเนื่องมาจากความช่วยเหลือและการสนับสนุนเป็นอย่างดียิ่งจากหลายฝ่าย ผู้เขียนขอกราบขอบพระคุณทุกท่านไว้ ณ โอกาสนี้

ขอขอบคุณมหาวิทยาลัยเทคโนโลยีสุรนารีเป็นอย่างสูง สำหรับการสนับสนุนทุนการศึกษาและทุนวิจัย ซึ่งเอื้ออำนวยต่อการดำเนินงานวิจัยฉบับนี้จนสำเร็จลุล่วงด้วยดี ทั้งนี้ ขอขอบคุณมหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก สำหรับการอนุญาตให้ลาเพื่อศึกษาต่อเต็มเวลา ซึ่งมีส่วนสำคัญต่อการสนับสนุนให้การศึกษาวิจัยครั้งนี้สามารถดำเนินไปได้อย่างต่อเนื่องและมีประสิทธิภาพ

ขอขอบพระคุณอาจารย์ที่ปรึกษาวิทยานิพนธ์ ได้แก่ ศาสตราจารย์ ดร.พีระพงษ์ อุฑารสกุล และรองศาสตราจารย์ ดร.คมเดช ภาพัฒนบุรี อย่างสูง สำหรับการให้คำแนะนำ การกำกับดูแลอย่างใกล้ชิด ตลอดจนข้อคิดเห็นเชิงวิชาการอันเป็นประโยชน์ ซึ่งมีส่วนสำคัญอย่างยิ่งต่อการพัฒนาคุณภาพและความสมบูรณ์ของงานวิจัยฉบับนี้

ขอขอบพระคุณ รองศาสตราจารย์ ดร.จักรี ศรีนนท์ฉัตร ศาสตราจารย์ ดร.มนต์ทิพย์ภา อุฑารสกุล ผู้ช่วยศาสตราจารย์ ดร.เศรษฐวิทย์ ภูฉายา และรองศาสตราจารย์ ดร.นุชนาฏ สันทาลุนัย ในฐานะคณะกรรมการสอบวิทยานิพนธ์ ที่ได้กรุณาให้ข้อเสนอแนะและคำชี้แนะอันทรงคุณค่า ซึ่งช่วยยกระดับคุณภาพและความถูกต้องทางวิชาการของงานวิจัยฉบับนี้

นอกจากนี้ ขอขอบคุณคณาจารย์และบุคลากรทุกท่านของสำนักวิชาวิศวกรรมโทรคมนาคม รวมถึงเพื่อนร่วมงานทุกคน สำหรับความอนุเคราะห์ ความช่วยเหลือ และการสนับสนุนด้วยดีตลอดระยะเวลาการดำเนินงานวิจัย

ท้ายที่สุดนี้ ขอขอบคุณโครงการทุนการศึกษา บิดา มารดา และอาจารย์ที่ปรึกษา สำหรับการสนับสนุน กำลังใจ และความเอื้ออาทรอย่างดียิ่งตลอดระยะเวลาการศึกษา ณ มหาวิทยาลัยเทคโนโลยีสุรนารี ซึ่งเป็นแรงผลักดันสำคัญให้การศึกษาครั้งนี้สำเร็จสมบูรณ์

ขวัญจิต ออกเวหา

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญ

เสียงปอด (Lung sounds) เป็นสัญญาณชีวการแพทย์ที่เกิดจากการไหลของอากาศภายในระบบทางเดินหายใจ ซึ่งสามารถสะท้อนถึงสภาวะปกติและความผิดปกติของปอดได้ เสียงปอดมีบทบาทสำคัญในการวินิจฉัยโรคระบบทางเดินหายใจ เช่น โรคปอดอักเสบ (Pneumonia) โรคหอบหืด (Asthma) และโรคปอดอุดกั้นเรื้อรัง (Chronic Obstructive Pulmonary Disease: COPD) ซึ่งเป็นโรคที่ส่งผลกระทบต่อคุณภาพชีวิตและมีอัตราการเกิดโรคเพิ่มขึ้น การตรวจวินิจฉัยโรคเหล่านี้ตั้งแต่ระยะเริ่มต้นจึงมีความสำคัญต่อการรักษาและลดอัตราการเสียชีวิตของผู้ป่วย (Kazem, Z. and Zahra, M., 2006; Gibson, G.J., Loddenkemper, R., Lundbäck, B. and Sibille, Y., 2013)

การประเมินเสียงปอดด้วยวิธีการฟังผ่านหูฟังแพทย์ (Auscultation) อาศัยประสบการณ์และความเชี่ยวชาญของแพทย์ ซึ่งมีข้อจำกัดด้านความเป็นอัตวิสัย ความไม่สม่ำเสมอในการตีความ และความยากในการจำแนกเสียงที่มีภาวะผิดปกติที่มีลักษณะใกล้เคียงกัน นอกจากนี้ เสียงปอดบางประเภทอาจปรากฏในช่วงความถี่ที่อยู่นอกเหนือการรับรู้ของหูมนุษย์ ส่งผลให้เกิดความคลาดเคลื่อนในการวินิจฉัยได้ (Pasterkamp, H., Kraman, S. S., and Wodicka, G. R., 1997) ปัญหาดังกล่าวเป็นอุปสรรคต่อความถูกต้องและความน่าเชื่อถือของการตรวจวินิจฉัยในทางคลินิก

มีการนำเทคนิคการประมวลผลสัญญาณและการเรียนรู้ของเครื่อง (Machine Learning: ML) มาใช้ในการพัฒนาระบบช่วยวินิจฉัยโรคจากเสียงปอด โครงสร้างของระบบส่วนใหญ่ประกอบด้วยสองขั้นตอนหลัก ได้แก่ การสกัดคุณลักษณะ (Feature extraction) และการจำแนกประเภท (Classification) โดยที่ค่าสัมประสิทธิ์เซปสตรัมความถี่เมล (Mel Frequency Cepstral Coefficients: MFCCs) เป็นหนึ่งในคุณลักษณะที่ได้รับความนิยมในการนำมาใช้ เนื่องจากสามารถแทนลักษณะสเปกตรัมของสัญญาณเสียงได้อย่างมีประสิทธิภาพ และมีความทนทานต่อสัญญาณรบกวน (Sengupta, N., Sahidullah, M. and Saha, G., 2016; Bahoura, M. and Pelletier, C., 2004; Jung, S.Y., Liao, C.H., Wu, Y.S., Yuan, S.M. and Sun, C.T., 2021) วิธีการจำแนกที่ใช้กันอย่างแพร่หลาย ได้แก่ ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) แบบจำลอง

ผสมแบบเกาส์เซียน (Gaussian Mixture Model: GMM), วิธีเพื่อนบ้านใกล้ที่สุดแบบเค (k-Nearest Neighbors: KNN) และโครงข่ายประสาทเทียมแบบต่าง ๆ (Azmy, M. M., 2015; Prajapati, S. K., Choudhary, T. S. and Mishra, S., 2023)

วิธีการเรียนรู้เชิงลึก (Deep Learning: DL) โดยโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) จะช่วยเพิ่มความแม่นยำในการจำแนกโรคจากเสียงปอดได้อย่างมีนัยสำคัญ โดย Bardou, D., Zhang, K., and Ahmad, S. M. (2018) ได้นำเสนอการจำแนกเสียงปอดด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชัน และรายงานผลการจำแนกที่มีประสิทธิภาพสูงขึ้นเมื่อเทียบกับวิธีการเรียนรู้ของเครื่อง แนวทางการผสานคุณลักษณะเชิงลึกจากโครงข่ายที่ผ่านการฝึกได้ถูกนำมาใช้เพื่อเพิ่มความสามารถในการเรียนรู้รูปแบบเชิงเวลาในงานจำแนกสัญญาณเสียง งานด้านเสียงปอดใช้โครงข่ายคอนโวลูชันโดยตรงกับตัวแทนเวลาและความถี่ (Shi, L., Du, K., Zhang, C., Ma, H., and Yan, W., 2019) งานวิจัยมุ่งเน้นเฉพาะงานการจำแนก โดยไม่ได้ให้ความสำคัญกับการรักษาข้อมูลเชิงโครงสร้างของสัญญาณเสียงดั้งเดิม ซึ่งอาจทำให้รายละเอียดบางส่วนที่มีนัยสำคัญทางคลินิกสูญหายไปในกระบวนการสกัดคุณลักษณะ นอกจากนี้แนวทางการเรียนรู้แบบงานเดียว ยังมีข้อจำกัดในการเรียนรู้คุณลักษณะที่ครอบคลุม และส่งผลต่อความสามารถในการอธิบายข้อมูลใหม่หรือข้อมูลทั่วไป เมื่อมีการเปลี่ยนแปลงของสภาวะแวดล้อมการบันทึกเสียง ผู้ป่วย หรืออุปกรณ์ที่ใช้เก็บข้อมูล

เพื่อลดข้อจำกัดดังกล่าว แนวคิดของออโตเอนโคเดอร์แบบหลายงาน (Multi-Task Autoencoder: MTAE) จึงได้รับความสนใจมากขึ้นในงานด้านสัญญาณชีวภาพ เนื่องจากเป็นแบบจำลองที่สามารถเรียนรู้งานการสร้างสัญญาณกลับคืน (Reconstruction) และงานการจำแนกประเภทไปพร้อมกัน ผ่านตัวแทนเชิงแฝงร่วม (Shared latent representation) ส่งผลให้แบบจำลองสามารถคงไว้ซึ่งสารสนเทศสำคัญของสัญญาณต้นฉบับ ขณะเดียวกันก็เพิ่มประสิทธิภาพในการจำแนกประเภทของข้อมูลได้โดย Ditthapron, A., Banluesombatkul, N., Ketrat, S., Chuangsuwanich, E. and Wilaiprasitporn, T. (2019) ได้นำเสนอการใช้ออโตเอนโคเดอร์แบบหลายงาน เพื่อสกัดคุณลักษณะร่วมสำหรับการจำแนกสัญญาณการตรวจคลื่นไฟฟ้าสมอง (Electroencephalography) ขณะที่ Autthasan, P., Duan, R., Zhang, Z., Li, H., Guan, C., and Wilaiprasitporn, T. (2021) เสนอแบบจำลองสำหรับการจำแนกสัญญาณการตรวจคลื่นไฟฟ้าสมอง ในงานการจินตนาการการเคลื่อนไหว (Motor imagery) ซึ่งแสดงให้เห็นประสิทธิภาพของการเรียนรู้แบบหลายงานในบริบทที่ต้องการความสามารถแบบไม่ขึ้นกับผู้ใช้ (Subject-independent) นอกจากนี้ Bashir-Gonbadi, F. and Khotanlou, H. (2021) ยังได้ประยุกต์แนวคิดออโตเอนโคเดอร์เชิงคอนโวลูชันแบบหลายงาน

(Multi-task convolutional autoencoder) ในงานจำแนกเนื้องอกสมอง ซึ่งสะท้อนว่าแนวทางการเรียนรู้ร่วมกันระหว่างการสร้างคืน และการจำแนกสามารถช่วยยกระดับคุณภาพของตัวแทนเชิงลึกได้อย่างมีนัยสำคัญ จากการทบทวนวรรณกรรมยังพบว่ามีงานวิจัยจำนวนไม่มากที่นำแนวคิดออโตเอนโคเดอร์แบบหลายงานมาประยุกต์ใช้โดยตรงกับการจำแนกโรคปอดจากเสียงปอด ซึ่งยังเป็นช่องว่างของงานวิจัยที่ควรได้รับการศึกษาเพิ่มเติม

ดังนั้น งานวิจัยนี้จึงมุ่งเสนอกรอบวิธีการใหม่สำหรับการตรวจจับโรคปอดจากสัญญาณเสียงปอด โดยประยุกต์ใช้แบบจำลองออโตเอนโคเดอร์แบบหลายงานร่วมกับซัพพอร์ตเวกเตอร์แมชชีน เพื่อเพิ่มความแม่นยำในการจำแนกโรค พร้อมทั้งรักษาคุณลักษณะเชิงโครงสร้างของสัญญาณเสียงไว้ ซึ่งจะช่วยเพิ่มความน่าเชื่อถือของระบบช่วยวินิจฉัยโรค และสามารถต่อยอดไปสู่การใช้งานจริงในระบบสุขภาพอัจฉริยะและการแพทย์ทางไกล

1.2 วัตถุประสงค์ของงานวิจัย

1.2.1 เพื่อวิเคราะห์สัญญาณเสียงปอด ใช้ในการคาดการณ์โรคที่เกี่ยวข้องกับระบบทางเดินหายใจ

1.2.2 ประยุกต์ใช้การสกัดคุณลักษณะของเสียงปอด เพื่อเป็นข้อมูลนำเข้าให้อัลกอริทึมการเรียนรู้เชิงลึก

1.2.3 เพื่อพัฒนาอัลกอริทึมการเรียนรู้เชิงลึกในการคาดการณ์โรคที่เกี่ยวข้องกับระบบทางเดินหายใจ

1.3 ขอบเขตของงานวิจัย

1.3.1 นำเสนอการประยุกต์ใช้การสกัดคุณลักษณะของเสียงปอดด้วยเทคนิคออโตเอนโคเดอร์แบบหลายงาน

1.3.2 ศึกษาและนำเสนอแนวทางการจำแนกโรคจากเสียงปอดด้วย วิธีการไฮบริด (Hybrid Method) โดยผสานการสกัดคุณลักษณะเชิงลึกด้วยเทคนิคออโตเอนโคเดอร์แบบหลายงานร่วมกับตัวจำแนกการเรียนรู้ของเครื่องแบบดั้งเดิม คือ ซัพพอร์ตเวกเตอร์แมชชีน เพื่อเพิ่มประสิทธิภาพในการแยกแยะกลุ่มโรค

1.3.3 แบบจำลองที่พัฒนาขึ้นสามารถตรวจสอบประสิทธิภาพ ในการคาดการณ์โรคที่เกี่ยวข้องกับระบบทางเดินหายใจ จำแนกเป็น 4 กลุ่มโรค ได้แก่ ภาวะปกติ (Normal) โรคหอบหืด

มหาวิทยาลัยเทคโนโลยีสุรนารี

(Asthma) โรคปอดอักเสบ (Pneumonia) และโรคปอดอุดกั้นเรื้อรัง (Chronic Obstructive Pulmonary Disease: COPD) โดยใช้ชุดข้อมูลเสียงปอดจากแหล่งข้อมูลที่เป็นมาตรฐานและเปิดเผยต่อสาธารณะ แบบจำลองออโตเอนโคเดอร์แบบหลายงานสามารถเพิ่มประสิทธิภาพในการตรวจจับโรคปอดจากสัญญาณเสียงปอดได้หรือไม่

1.4 จุดเด่นของงานวิจัย

1.4.1 พัฒนาแบบจำลองออโตเอนโคเดอร์แบบหลายงาน เพื่อเรียนรู้การแทนลักษณะของสัญญาณเสียงปอดที่มีประสิทธิภาพ

1.4.2 นำเสนอแนวทางแบบผสม โดยผสานแบบจำลองออโตเอนโคเดอร์แบบหลายงานและซัพพอร์ตเวกเตอร์แมชชีน เพื่อยกระดับสมรรถนะในการจำแนกข้อมูล

1.4.3 การประเมินผลแบบไม่ขึ้นกับตัวบุคคลโดยใช้การแบ่งข้อมูลออกเป็นห้าส่วน (5-fold) โดยกำหนดให้ข้อมูลของบุคคลเดียวกันไม่ปรากฏทั้งในชุดฝึกและชุดทดสอบ เพื่อป้องกันการรั่วไหลของข้อมูล และทำให้ผลการประเมินสะท้อนความสามารถของแบบจำลองในการประยุกต์ใช้

1.4.4 อธิบายแบบจำลองด้วยการวิเคราะห์เชิงคุณภาพด้วยเทคนิค t-Distributed Stochastic Neighbor Embedding (t-SNE) และเมตริกซ์สับสน (Confusion matrix)

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้วิธีการใหม่สำหรับการตรวจจับโรคปอดจากสัญญาณเสียงปอดที่ผสานแนวคิดการเรียนรู้แบบหลายงานเข้ากับการเรียนรู้ของเครื่อง

1.5.2 ได้แบบจำลองที่สามารถรักษาสารสนเทศเชิงโครงสร้างของสัญญาณเสียงปอดไว้ควบคู่กับการเพิ่มความถูกต้องในการจำแนกโรค

1.5.3 ได้แนวทางที่สามารถนำไปประยุกต์ใช้เป็นระบบช่วยวินิจฉัยโรคปอดแบบอัตโนมัติเพื่อช่วยสนับสนุนการตัดสินใจของบุคลากรทางการแพทย์

1.6 โครงสร้างและการจัดลำดับเนื้อหาวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้แบ่งเนื้อหาออกเป็น 5 บท เพื่ออธิบายที่มา แนวคิด ระเบียบวิธีการวิจัย ผลการทดลอง และข้อสรุปของการพัฒนาแบบจำลองการเรียนรู้เชิงลึกสำหรับการคาดการณ์โรคระบบทางเดินหายใจจากสัญญาณเสียงปอด

1.6.1 บทที่ 1 บทนำ นำเสนอความเป็นมาและความสำคัญของปัญหา วัตถุประสงค์ ขอบเขตการวิจัย และคุณูปการของงาน โดยกำหนดโจทย์การจำแนก 4 กลุ่มโรค ได้แก่ ภาวะปกติ, โรคหอบหืด, โรคปอดอักเสบ และโรคปอดอุดกั้นเรื้อรัง บนชุดข้อมูลมาตรฐานแบบเปิดเผย

1.6.2 บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง ทบทวนวรรณกรรมด้านการวิเคราะห์เสียงปอด วิธีการประมวลผลสัญญาณและการสกัดคุณลักษณะ รวมถึงแนวทางการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึก

1.6.3 บทที่ 3 วิธีการดำเนินการวิจัย ข้อมูลและการเตรียมสัญญาณ การสกัดคุณลักษณะด้วยสัมประสิทธิ์เซปสตรัมความถี่เมล การพัฒนาแบบจำลองออโตเอนโคเดอร์เชิงลึกแบบหลายงานและแบบไฮบริดร่วมกับซัพพอร์ตเวกเตอร์แมชชีน รวมถึงการประเมินประสิทธิภาพ

1.6.4 บทที่ 4 ผลการวิจัยและอภิปรายผล ผลการดำเนินงานวิจัยของพารามิเตอร์ที่สำคัญ การเปรียบเทียบแบบจำลองที่นำเสนอกับแบบจำลองอ้างอิงและงานวิจัยที่มีอยู่เดิม และการวิเคราะห์เชิงภาพด้วยเทคนิค t-distributed Stochastic Neighbor Embedding (t-SNE) และเมตริกซ์สับสน (Confusion matrix)

1.6.5 บทที่ 5 สรุปและข้อเสนอแนะ สรุปผลการวิจัยตามวัตถุประสงค์และคุณูปการของวิทยานิพนธ์ อภิปรายข้อจำกัดของงานวิจัย และเสนอแนวทางการพัฒนาต่อยอดในอนาคต เช่น การตรวจสอบความถูกต้องกับข้อมูลภายนอก และการเพิ่มความทนทานต่อความแปรปรวนของสัญญาณเสียงปอด

มหาวิทยาลัยเทคโนโลยีสุรนารี

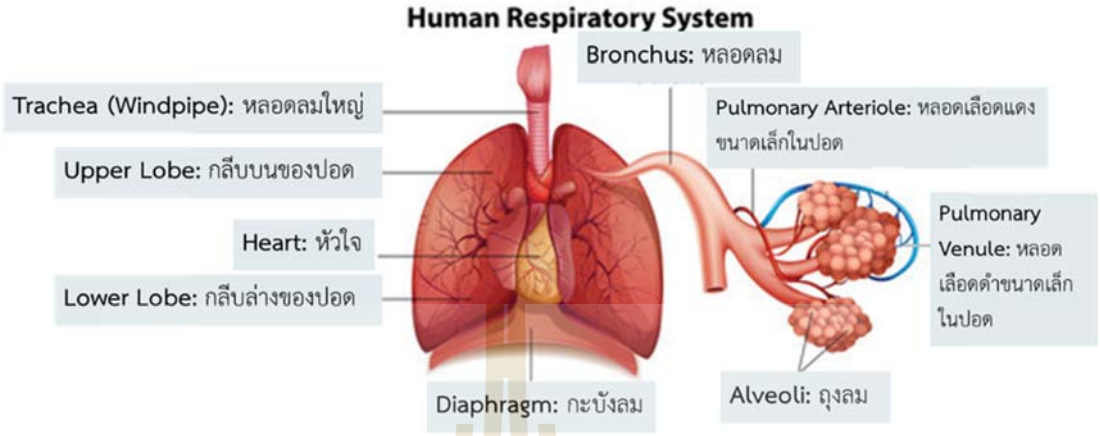
บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

นำเสนอการทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์สัญญาณเสียงทางการแพทย์ โดยเฉพาะอย่างยิ่งงานด้านการวิเคราะห์เสียงปอดเพื่อการตรวจจับและจำแนกโรคระบบทางเดินหายใจ ซึ่งเป็นรากฐานสำคัญในการพัฒนาแนวคิด และวิธีการของงานวิจัยวิทยานิพนธ์นี้ การอธิบายหลักการพื้นฐานของสัญญาณเสียงปอดและคุณลักษณะเชิงสเปกตรัมที่ใช้ในการแทนข้อมูลเสียง การทบทวนแนวคิดและทฤษฎีของแบบจำลองการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึกที่นำมาประยุกต์ใช้กับข้อมูลเสียงทางการแพทย์ จากนั้นการอธิบายแนวคิดของออโตเอนโคเดอร์ การเรียนรู้แบบหลายงาน และตัวจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีน ซึ่งเป็นส่วนสำคัญของวิธีการที่นำเสนอ และงานวิจัยที่เกี่ยวข้องเพื่อชี้ให้เห็นถึงข้อจำกัดและช่องว่างงานวิจัย นำไปสู่การกำหนดแนวทางและวิธีการของงานวิจัยวิทยานิพนธ์นี้

2.1 พื้นฐานสรีรวิทยาและสัญญาณเสียงปอด

2.1.1 สรีรวิทยาของระบบทางเดินหายใจ

โครงสร้างของระบบทางเดินหายใจของมนุษย์ (Human respiratory system) โดยอากาศจะเข้าสู่ร่างกายผ่านหลอดลมใหญ่ (Trachea) แล้วแยกไปยังหลอดลม (Bronchus) ภายในปอดซึ่งแบ่งออกเป็นกลีบบนและกลีบล่าง ถัดไปหลอดลมจะแตกแขนงเป็นท่อขนาดเล็กจนถึงถุงลมปอด (Alveoli) ซึ่งเป็นบริเวณสำคัญสำหรับการแลกเปลี่ยนก๊าซออกซิเจนและคาร์บอนไดออกไซด์กับหลอดเลือดฝอย โดยมีหลอดเลือดแดงและหลอดเลือดดำทำหน้าที่ลำเลียงเลือด รูปที่ 2.1 แสดงตำแหน่งของหัวใจที่อยู่ระหว่างปอดทั้งสองข้าง และกระบังลม (Diaphragm) ซึ่งเป็นกล้ามเนื้อสำคัญที่ช่วยควบคุมการหายใจเข้าและออกของมนุษย์



รูปที่ 2.1 โครงสร้างของระบบทางเดินหายใจ

ที่มา: โรงพยาบาลบำรุงราษฎร์, 2567

การไหลของอากาศระหว่างการหายใจเข้าออกสามารถก่อให้เกิดเสียงปอดจากการไหลแบบปั่นป่วน (Turbulent flow) การสั่นของผนังหลอดลม และการเปิดปิดของทางเดินหายใจขนาดเล็กในระดับหลอดลมฝอยและถุงลมปอด ซึ่งสามารถอธิบายเชิงแนวคิดได้ ดังแสดงในตารางที่ 2.1 เมื่อเกิดความผิดปกติของโครงสร้างหรือสภาวะทางพยาธิวิทยา เช่น การอักเสบ การตีบแคบ หรือการมีสารคัดหลั่ง รูปแบบของเสียงปอดจะเปลี่ยนแปลงไปอย่างมีนัยสำคัญ (Pasterkamp et al., 1997; Loudon, R. G., and Murphy, R. L. H., 1984)

ตารางที่ 2.1 แนวคิดเชิงกลไกของการเกิดเสียงปอด*

เสียงปอด	ลักษณะสำคัญ	โรคที่เกี่ยวข้อง
Crackles (เสียงแตกกรอบ / เสียงป๊อป)	- เป็นเสียงสั้น ๆ คล้ายเสียงแตกหรือเสียงป๊อปเป็นช่วง ๆ - มักได้ยินชัดในช่วงหายใจเข้า - เกิดจากของเหลวในทางเดินหายใจหรือการเปิดของถุงลม/หลอดลมที่แฟบ - เสียงอาจมีตั้งแต่เบาและแหลม ไปจนถึงดังและคล้ายเสียงน้ำกลั้ว	- ปอดอักเสบ - ภาวะหัวใจล้มเหลว (Heart failure) - ภาวะน้ำท่วมปอด (Pulmonary edema) - ภาวะถุงลมแฟบ (Atelectasis) - โรคปอดเนื้อเยื่อคั่น (Interstitial lung disease) - การกำเริบของโรคปอดอุดกั้นเรื้อรัง (COPD exacerbations)

ตารางที่ 2.1 แนวคิดเชิงกลไกของการเกิดเสียงปอด* (ต่อ)

เสียงปอด	ลักษณะสำคัญ	โรคที่เกี่ยวข้อง
Wheezes (เสียงหวีด)	- เป็นเสียงต่อเนื่องคล้ายเสียงดนตรี เกิดจากทางเดินหายใจแคบลง - มักได้ยินชัดในช่วงหายใจออก - มักเกิดจากหลอดลมขนาดเล็กหรือหลอดลมฝอย	- โรคหอบหืด - โรคปอดอุดกั้นเรื้อรัง - ภาวะหัวใจล้มเหลว
Rhonchi (เสียงครืดคราดต่ำ)	- เป็นเสียงต่ำ คล้ายเสียงกรน - มักได้ยินชัดในช่วงหายใจออก - เกิดจากเสมหะหรือสารคัดหลั่งในทางเดินหายใจขนาดใหญ่ - มักลดลงหรือหายไปหลังไอ	- ทางเดินหายใจอุดกั้น - หลอดลมอักเสบ (Bronchitis) - ปอดอักเสบ
Diminished (เสียงปอดเบาลง)	- เสียงปอดเบากว่าปกติ - เกิดจากการไหลเวียนของอากาศลดลงหรือทางเดินหายใจอุดกั้น - อาจเกิดจากน้ำหรืออากาศในช่องเยื่อหุ้มปอด - บ่งชี้ว่าการระบายอากาศลดลงหรือการขยายตัวของปอดจำกัด	- โรคปอดอุดกั้นเรื้อรัง - โรคหอบหืด - ภาวะลมรั่วในช่องเยื่อหุ้มปอด (Pneumothorax) - น้ำในช่องเยื่อหุ้มปอด (Pleural effusion) - โรคอ้วนหรือผนังทรวงอกหนาตัว - ภาวะถุงลมแฟบรุนแรง (Severe atelectasis)
Stridor (เสียงหวีดแหลมดัง)	- เป็นเสียงแหลมและดังมาก - มักได้ยินชัดขณะฟังบริเวณหลอดลม (trachea) - บ่งชี้การตีบแคบของทางเดินหายใจส่วนบน - ต้องได้รับการประเมินอย่างเร่งด่วน	- การสำลักสิ่งแปลกปลอม - ฝาปิดกล่องเสียงอักเสบ (Epiglottitis) - ครูป (Croup)

มหาวิทยาลัยเทคโนโลยีสุรนารี

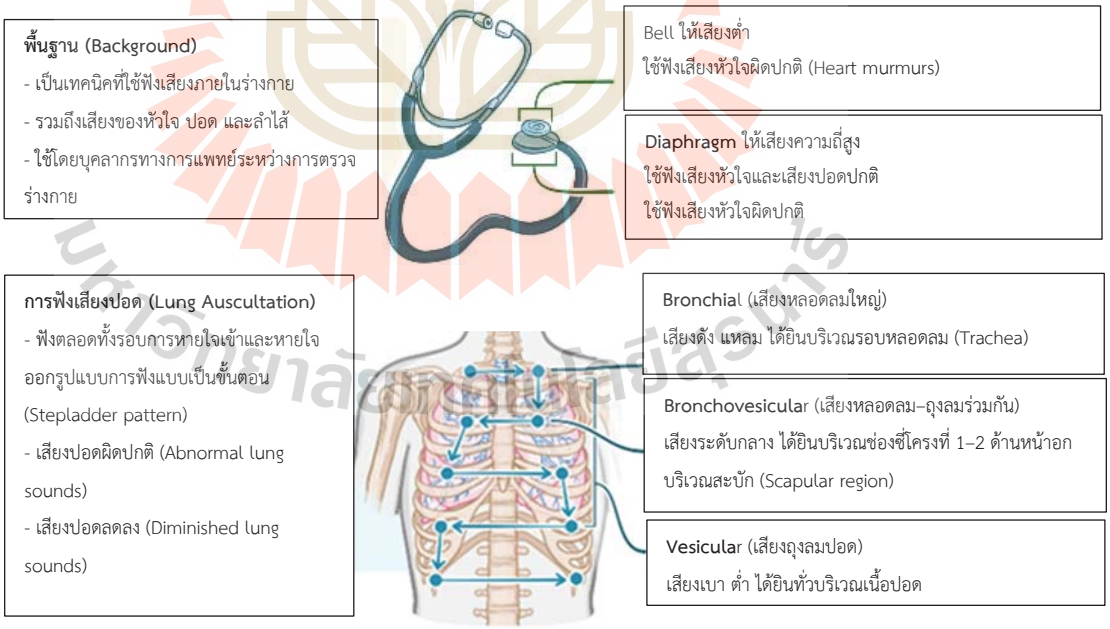
ตารางที่ 2.1 แนวคิดเชิงกลไกของการเกิดเสียงปอด* (ต่อ)

เสียงปอด	ลักษณะสำคัญ	โรคที่เกี่ยวข้อง
Pleural Rub (เสียงเสียดสีของเยื่อหุ้มปอด)	- เป็นเสียงเสียดสีจากเยื่อหุ้มปอดที่อักเสบ - ได้ยินทั้งขณะหายใจเข้า และบางครั้งหายใจออก - ได้ยินชัดบริเวณด้านล่างหรือด้านข้างของทรวงอก	- เยื่อหุ้มปอดอักเสบ (Pleurisy) - เยื่อหุ้มหัวใจอักเสบ (Pericarditis) - น้ำในช่องเยื่อหุ้มหัวใจ

ที่มา: *ดัดแปลงจาก Pasterkamp et al. (1997) และ Loudon and Murphy (1984)

2.1.2 การส่งผ่าน การตรวจจับ และการฟังเสียงปอด

เสียงปอดที่เกิดขึ้นภายในปอดจะถูกส่งผ่านเนื้อปอด เยื่อหุ้มปอด และผนังทรวงอก ก่อนถูกตรวจจับด้วยหูฟังแพทย์ (Stethoscope) หรือหูฟังแพทย์แบบดิจิทัล (Digital stethoscope) สัญญาณที่ตรวจวัดได้จึงสะท้อนทั้งแหล่งกำเนิดเสียงและผลของการลดทอนหรือการกรองจากโครงสร้างทรวงอก ดังรูปที่ 2.2



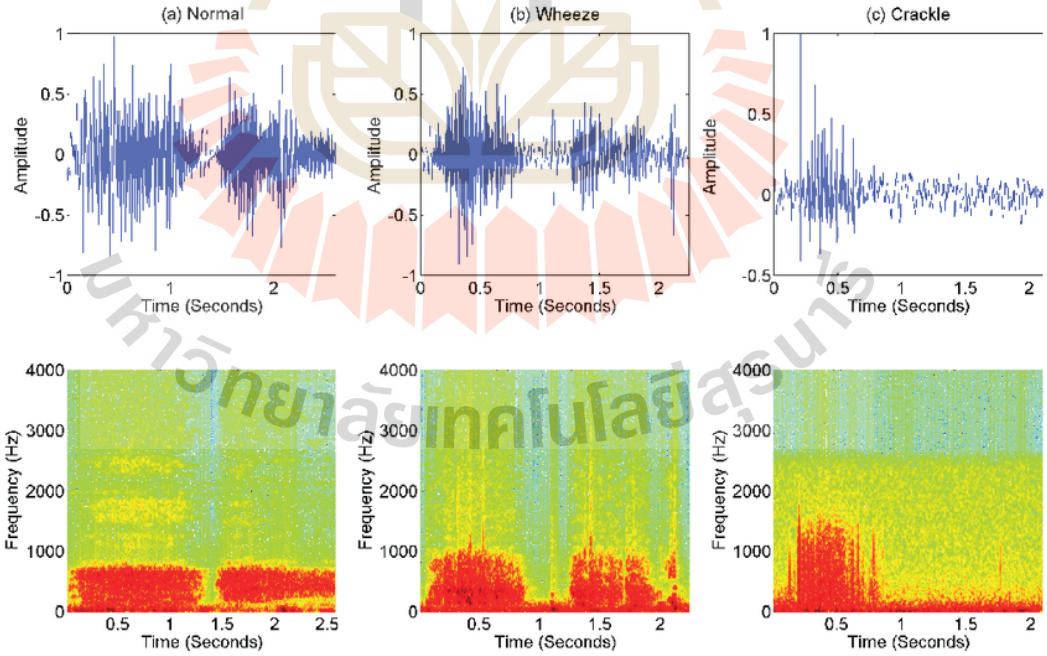
รูปที่ 2.2 ส่วนประกอบของหูฟังแพทย์ และตำแหน่งมาตรฐานสำหรับการตรวจฟังเสียง

ที่มา: ดัดแปลงจาก Bohadana, A., Izbicki, G., and Kraman, S. S. (2014)

การกำหนดตำแหน่งการวางหูฟังให้ครอบคลุมบริเวณทรวงอกที่สำคัญช่วยเพิ่มความสม่ำเสมอของการตรวจวัด เป็นข้อกำหนดสำคัญในการวิเคราะห์เชิงวิศวกรรม (Rangayyan, R. M., 2015) นอกจากนั้นในทางปฏิบัติการประเมินเสียงปอดด้วยการตรวจฟังเสียง ยังมีข้อจำกัดสำคัญ ได้แก่ ความเป็นอัตวิสัยของผู้ตรวจ ประสบการณ์และความเชี่ยวชาญของแต่ละบุคคล ความเสี่ยงในการพลาดเบาะแสเชิงวินิจฉัย เป็นต้น โดยเฉพาะเมื่อสัญญาณมีความเข้มต่ำหรือมีสัญญาณรบกวน ส่งผลให้เกิดการพัฒนาแนวทางการวินิจฉัยโดยใช้คอมพิวเตอร์ช่วย (Computer-aided diagnosis) เพื่อเพิ่มความเป็นวัตถุวิสัยและความแม่นยำของการประเมินเสียง

2.1.3 ประเภทของเสียงปอดและลักษณะเชิงความถี่ของสัญญาณ

เสียงปอดสามารถจำแนกเป็นเสียงปอดภาวะปกติ (Normal lung sounds) และเสียงปอดภาวะผิดปกติ (Abnormal lung sounds) โดยเสียงภาวะผิดปกติที่พบบ่อย ได้แก่ เสียงแตกกรอบ เสียงหวีด เสียงครืดคราดต่ำ และเสียงหวีดแหลมดัง ซึ่งสัมพันธ์กับพยาธิสภาพของระบบทางเดินหายใจและมีลักษณะเชิงเวลาและเชิงความถี่เฉพาะตัว (Sarkar, M., Madabhavi, I., Niranjan, N., and Dogra, M., 2015) เพื่อความชัดเจนสามารถแสดงตัวอย่างสัญญาณในโดเมนเวลาและโดเมนเวลาและความถี่ ดังรูปที่ 2.3



รูปที่ 2.3 ตัวอย่างเสียงปอดภาวะปกติและเสียงปอดภาวะผิดปกติ

ที่มา: Bahoura, M. (2009)

เสียงปอดภาวะปกติ มีองค์ประกอบความถี่หลักอยู่ในช่วงประมาณ 100 ถึง 600 เฮิรตซ์ ขณะที่เสียงปอดภาวะผิดปกติบางชนิดมีองค์ประกอบความถี่สูงถึงระดับ 1 ถึง 2 กิโลเฮิรตซ์ ซึ่งเป็นพื้นฐานสำคัญต่อการกำหนดขั้นตอนการกรองความถี่และการเลือกคุณลักษณะสัญญาณเพื่อจำแนกประเภทเสียง (Kompis, M., Pasterkamp, H., and Wodicka, G. R., 2001; Bahoura, 2009) เชื่อมโยงจากลักษณะเสียงไปสู่การคาดการณ์โรค จำเป็นต้องพิจารณาว่าสัญญาณเสียงปอดเป็นสัญญาณไม่คงที่ตามเวลา และปนเปื้อนสัญญาณรบกวนจากเสียงหัวใจ การเคลื่อนไหวของร่างกาย และเสียงสิ่งแวดล้อม ทำให้การวิเคราะห์เชิงเวลาและความถี่ รวมถึงการออกแบบคุณลักษณะมีบทบาทสำคัญต่อความสามารถในการแยกความแตกต่างของโรค (Rangayyan, 2015)

2.1.4 โรคที่เกี่ยวข้องกับระบบทางเดินหายใจและความสัมพันธ์กับเสียงปอด

โรคที่เกี่ยวข้องกับระบบทางเดินหายใจ หมายถึงปัญหาใด ๆ ในปอดที่ทำให้ปอดทำงานผิดปกติ โดยสามารถจัดกลุ่มเชิงแนวคิดได้ 3 ประเภท ได้แก่ (1) โรคทางเดินหายใจ (Airway diseases) ซึ่งส่งผลต่อท่ออากาศ ทำให้ทางเดินหายใจตีบหรืออุดกั้น (2) โรคของเนื้อเยื่อปอด (Lung tissue diseases) ซึ่งส่งผลต่อโครงสร้างเนื้อเยื่อปอด ทำให้ปอดขยายได้ไม่เต็มที่ และ (3) โรคการไหลเวียนของปอด (Lung circulation diseases) ซึ่งส่งผลต่อหลอดเลือดในปอดและสัมพันธ์กับภาระการทำงานของหัวใจ (โรงพยาบาลบำรุงราษฎร์, 2567) ในงานวิจัยนี้มุ่งศึกษาสัญญาณเสียงปอดเพื่อสนับสนุนการจำแนกโรคที่เกี่ยวข้องกับระบบทางเดินหายใจซึ่งพบได้บ่อย และสอดคล้องกับแนวทางงานวิจัยด้านปัญญาประดิษฐ์ที่กำหนดจำแนกโรคที่สำคัญ เช่น โรคปอดอักเสบ โรคหอบหืด และโรคปอดอุดกั้นเรื้อรัง เป็นต้น

2.1.4.1 โรคปอดอักเสบ

เรียกอีกชื่อว่า โรคปอดบวม เป็นการอักเสบของเนื้อปอดที่พบได้บ่อย โดยเฉพาะในเด็กเล็ก ผู้สูงอายุ และผู้ที่มีภาวะภูมิต้านทานต่ำ ซึ่งอาจรุนแรงจนก่อให้เกิดภาวะแทรกซ้อนและการเสียชีวิตได้ สาเหตุของโรคแบ่งได้เป็น 2 กลุ่ม ได้แก่ ปอดอักเสบจากการติดเชื้อ ซึ่งเป็นชนิดที่พบบ่อยที่สุด โดยเชื้อไวรัส เชื้อแบคทีเรีย หรือเชื้อรา สามารถเข้าสู่ปอดและทำให้เกิดการอักเสบของถุงลมและเนื้อเยื่อโดยรอบ และปอดอักเสบจากการไม่ติดเชื้อ เช่น การสูดสารระคายเคือง (ฝุ่น ควัน สารเคมีระเหย) หรือผลข้างเคียงจากยาบางชนิด อาการสำคัญ ได้แก่ ไอมีเสมหะ เจ็บหน้าอกขณะหายใจหรือไอ มีไข้ หนาวสั่น อ่อนเพลีย และอาจมีคลื่นไส้อาเจียนหรือท้องเสียร่วม (โรงพยาบาลบำรุงราษฎร์, 2567) ในเชิงการฟังเสียงปอด ปอดอักเสบสัมพันธ์กับการอักเสบและสารคัดหลั่งในถุงลมหรือเนื้อปอด ส่งผลให้พบเสียงผิดปกติกลุ่มเสียงแตกกรอบ และเกิด

การเปลี่ยนแปลงของสัญญาณในโดเมนเวลาและความถี่ ซึ่งเป็นเหตุผลที่งานวิจัยจำนวนมากใช้ข้อมูลเสียงปอดเพื่อช่วยคัดกรองโรคปอดอักเสบ (Bohadana et al., 2014)

2.1.4.2 โรคหอบหืด

เป็นภาวะอักเสบเรื้อรังของเยื่อบุหลอดลมร่วมกับความไวของหลอดลมต่อสิ่งกระตุ้น ทำให้เกิดอาการหายใจลำบาก เหนื่อยหอบ และสมรรถภาพปอดลดลง โดยโรคนี้รักษาไม่หายขาดแต่สามารถควบคุมไม่ให้อาการกำเริบได้ สาเหตุสำคัญประกอบด้วยปัจจัยสิ่งแวดล้อม (สารก่อภูมิแพ้ การติดเชื้อทางเดินหายใจ อากาศเย็น กรดไหลย้อน ความเครียด หรือยาบางชนิด) และปัจจัยทางพันธุกรรม อาการที่พบบ่อย ได้แก่ ไอมีเสมหะ แน่นหน้าอก หายใจไม่อิ่ม หายใจมีเสียงหวีดและเหนื่อยหอบ โดยมักเป็น ๆ หาย ๆ แต่จะแสดงอาการช่วงกลางคืนหรือเช้ามืด (โรงพยาบาลบำรุงราษฎร์, 2567) ในเชิงสัญญาณเสียง โรคหอบหืดสัมพันธ์กับการตีบแคบของหลอดลมและการเพิ่มความต้านทานของทางเดินหายใจ จึงมักพบเสียง wheezes ที่มีลักษณะต่อเนื่องและเด่นในโดเมนความถี่ ซึ่งเป็นลักษณะสำคัญต่อการจำแนกด้วยวิธีเชิงเวลาและความถี่ (Sarkar et al., 2015; Bahoura, 2009)

2.1.4.3 โรคปอดอุดกั้นเรื้อรัง

เป็นกลุ่มโรคปอดอักเสบเรื้อรังที่พบบ่อยและเป็นสาเหตุการเสียชีวิตอันดับต้น ของประชากรโลก โดยลักษณะสำคัญ คือ หลอดลม เนื้อปอด และหลอดเลือดปอดเกิดการอักเสบเสียหายจากการได้รับแก๊สหรือสารระคายเคืองเป็นเวลานาน ส่งผลให้หลอดลมค่อย ๆ ตีบแคบหรือถูกอุดกั้นและไม่สามารถฟื้นคืนสู่สภาพปกติได้ โรคหลักในกลุ่มนี้ ได้แก่ ถุงลมโป่งพอง และหลอดลมอักเสบเรื้อรัง รวมถึงหลอดลมพอง ปัจจัยเสี่ยงสำคัญได้แก่การสูบบุหรี่ พันธุกรรม และมลพิษทางอากาศหรือการสัมผัสสารเคมีต่อเนื่อง อาการมักเริ่มจากไอเรื้อรัง มีเสมหะมาก เหนื่อยหอบ หายใจลำบาก แน่นหน้าอก หายใจมีเสียงหวีด และติดเชื้อทางเดินหายใจบ่อย (โรงพยาบาลบำรุงราษฎร์, 2567) ในมุมมองการจำแนกจากเสียงปอด โรคปอดอุดกั้นเรื้อรังอาจมีรูปแบบเสียงบางส่วนคล้ายโรคหอบหืด จึงมีความท้าทายในการจำแนกโรคปอดอุดกั้นเรื้อรังกับโรคหอบหืด เมื่อใช้แบบจำลองจำแนกประเภทแบบหลายกลุ่มโรค

2.1.5 แหล่งที่มาของข้อมูล

ในการวิจัยนี้ ผู้วิจัยได้เลือกใช้ชุดข้อมูลเสียงปอดจากงานวิจัยเรื่อง *“A dataset of lung sounds recorded from the chest wall using an electronic stethoscope”* ซึ่งเป็นชุดข้อมูลที่จัดทำขึ้นโดยใช้หูฟังแพทย์อิเล็กทรอนิกส์ในการบันทึกเสียงจากผู้ป่วยจริง ชุดข้อมูลดังกล่าวถูก

พัฒนาเพื่อรองรับการศึกษาและพัฒนาอัลกอริทึมทางด้านการเรียนรู้ของเครื่อง และปัญญาประดิษฐ์สำหรับการตรวจจับและจำแนกโรคปอด

ข้อมูลถูกเก็บรวบรวมจากผู้เข้าร่วมจำนวนทั้งสิ้น 112 ราย ประกอบด้วยผู้ที่มีสุขภาพปกติ และผู้ป่วยที่มีความผิดปกติของระบบทางเดินหายใจ โดยครอบคลุมโรคปอดที่สำคัญ ได้แก่ โรคหอบหืด โรคปอดอักเสบ โรคปอดอุดกั้นเรื้อรัง ภาวะหัวใจล้มเหลว พังผืดในปอด หลอดลมอักเสบ และภาวะน้ำในเยื่อหุ้มปอด

2.1.5.1 วิธีการและขั้นตอนการจัดเก็บข้อมูลรคปอดอักเสบ

การเก็บข้อมูลเสียงปอดดำเนินการโดยใช้หูฟังแพทย์อิเล็กทรอนิกส์รุ่น Littmann Electronic Stethoscope Model 3200 ซึ่งสามารถบันทึกสัญญาณเสียงและส่งข้อมูลไปยังคอมพิวเตอร์ผ่านระบบ Bluetooth ขั้นตอนการเก็บข้อมูลสามารถสรุปได้ดังนี้

1. ทำการวางหูฟังแพทย์บนตำแหน่งต่าง ๆ ของผนังทรวงอก (Chest wall) โดยตำแหน่งดังกล่าวถูกกำหนดโดยแพทย์ผู้เชี่ยวชาญ

2. บันทึกเสียงการหายใจของผู้ป่วยในแต่ละตำแหน่ง โดยระยะเวลาการบันทึกอยู่ระหว่าง 5 ถึง 30 วินาที ซึ่งเพียงพอสำหรับการครอบคลุมอย่างน้อยหนึ่งรอบของการหายใจ

3. สำหรับการบันทึกแต่ละครั้ง มีการบันทึกซ้ำในรูปแบบตัวกรองความถี่ (Filter) ที่แตกต่างกันจำนวน 3 แบบ ได้แก่ Bell, Diaphragm และ Extended เพื่อเน้นช่วงความถี่ของเสียงที่แตกต่างกัน

4. ข้อมูลเสียงที่ได้จะถูกส่งไปยังคอมพิวเตอร์และจัดเก็บในรูปแบบไฟล์เสียงชนิด .wav ผ่านซอฟต์แวร์ Littmann StethAssist

ทั้งนี้ การเก็บข้อมูลมุ่งเน้นเฉพาะเสียงปอด โดยที่ไม่มีวัตถุประสงค์ในการบันทึกเสียงหัวใจ

2.2 การสกัดคุณลักษณะ

จากหัวข้อ 2.1 เสียงปอดเป็นสัญญาณชีวการแพทย์ที่สะท้อนทั้งสรีรวิทยาและพยาธิสภาพของระบบทางเดินหายใจ โดยมีลักษณะไม่คงที่ตามเวลา และปนเปื้อนสัญญาณรบกวนจากเสียงหัวใจ การเคลื่อนไหวของร่างกาย และสิ่งแวดล้อม ทำให้การจำแนกโรคจากสัญญาณต้นฉบับโดยตรงมีข้อจำกัด กระบวนการสกัดคุณลักษณะจึงมีบทบาทสำคัญในการแปลงสัญญาณดิบให้เป็นตัวแทนเชิงตัวเลขที่คงสารสนเทศสำคัญของโรค และเอื้อต่อการเรียนรู้ของแบบจำลอง

มหาวิทยาลัยเทคโนโลยีสุรนารี

การสกัดคุณลักษณะ มีขั้นตอนการเตรียมสัญญาณ เพื่อจำกัดย่านความถี่ที่สนใจและลดสัญญาณรบกวน การแบ่งช่วงสัญญาณ (Segmentation) เพื่อรองรับการวิเคราะห์แบบช่วงเวลา (Short-time analysis) และการคำนวณคุณลักษณะ เช่น สัมประสิทธิ์เซปสตรัมความถี่เมล และอนุพันธ์ตามเวลาเพื่อเน้นความแตกต่างของคุณลักษณะเสียงที่มีความสัมพันธ์กับโรค

2.2.1 การเตรียมสัญญาณเพื่อรองรับการสกัดคุณลักษณะ

เสียงปอดที่ทำการบันทึกมีสัญญาณรบกวน จึงใช้ตัวกรองแถบผ่าน (Bandpass filter) เพื่อคงองค์ประกอบความถี่ที่เกี่ยวข้องและลดผลกระทบจากสัญญาณรบกวนก่อนสกัดคุณลักษณะ งานวิจัยนี้ใช้ตัวกรองแถบผ่านแบบบัตเตอร์เวิร์ธ (Butterworth bandpass filter) และกำหนดฟังก์ชันถ่ายโอนในโดเมนความถี่ โดยกำหนดลำดับฟิลเตอร์ N_{BW} = 12 และกำหนดย่านความถี่ที่สนใจ 120 ถึง 1800 เฮิรตซ์ (Pasterkamp et al., 1997; Rangayyan, 2015) สมการที่ (2.1) แสดงการตอบสนองเชิงขนาดของตัวกรองแถบผ่านแบบบัตเตอร์เวิร์ธแบบความถี่ต่ำผ่านในรูปทั่วไป โดยในการทดลองจริงใช้การออกแบบเป็น bandpass ด้วยความถี่ตัดต่ำ f_L และความถี่ตัดสูง f_H ตามกระบวนการออกแบบตัวกรองมาตรฐาน

$$H(f) = \frac{1}{\sqrt{1 + \left(\frac{f}{f_c}\right)^{2N_{BW}}}} \tag{2.1}$$

เมื่อ

$H(f)$ คือ ฟังก์ชันการตอบสนองเชิงความถี่ของตัวกรอง

f คือ ความถี่ของสัญญาณ (เฮิรตซ์)

f_c คือ ความถี่ตัดของตัวกรอง (เฮิรตซ์)

N_{BW} คือ อันดับของตัวกรองบัตเตอร์เวิร์ธ

ทั้งนี้ ตัวกรองแถบผ่านแบบบัตเตอร์เวิร์ธ มีคุณสมบัติการตอบสนองในย่านผ่านที่เรียบ จึงช่วยลดความผิดเพี้ยนของพลังงานในย่านความถี่ที่สนใจและทำให้คุณลักษณะเชิงสเปกตรัมมีเสถียรภาพมากขึ้น (Oppenheim, A. V., and Schafer, R. W., 2010; Rangayyan, 2015)

แม้สัญญาณเสียงปอดจะไม่คงที่ แต่สามารถประมาณให้คงที่ได้ในช่วงเวลาสั้น จึงทำให้การวิเคราะห์แบบช่วงเวลาและการคำนวณสเปกตรัมมีความหมาย (Rangayyan, 2015; Bahoura

and Pelletier, 2003; Sengupta et al., 2015) นอกจากนี้ การคำนวณสัมประสิทธิ์เซปสตรัมความถี่เมลยังต้องแบ่งสัญญาณเป็นเฟรมสั้นระดับมิลลิวินาที เพื่อให้ได้ความสมดุลที่เหมาะสมระหว่างความละเอียดเชิงเวลาและเชิงความถี่ในบริบทเสียงปอด (Li, L., Xu, W., Hong, Q., Tong, F., and Wu, J., 2016; Bahoura, 2009)

2.2.2 สัมประสิทธิ์เซปสตรัมความถี่เมล

สัมประสิทธิ์เซปสตรัมความถี่เมลเป็นคุณลักษณะเชิงสเปกตรัมที่สรุปรูปทรงของสเปกตรัมในช่วงเวลาสั้นให้อยู่ในรูปค่าสัมประสิทธิ์จำนวนน้อย กระบวนการประกอบด้วยการคำนวณสเปกตรัมกำลังผ่านการแปลงฟูเรียร์ช่วงเวลาสั้น (Short-Time Fourier Transform: STFT) การรวมพลังงานผ่านแถบกรองความถี่เมล (Mel-filter bank) ทำการแปลงลอการิทึม (Logarithm) และการแปลงด้วยการแปลงโคไซน์ไม่ต่อเนื่อง (Discrete Cosine Transform: DCT) เพื่อให้ได้สัมประสิทธิ์เซปสตรัม (Rangayyan, 2015; Bahoura and Pelletier, 2003; Sengupta et al., 2015; Li et al., 2016) สมการพื้นฐานของสัมประสิทธิ์เซปสตรัมความถี่เมล ดังสมการที่ (2.2)

$$MFCC(c) = \sum_{m=0}^{M-1} \log(E(m)) \cos\left(\frac{\pi c \left(m + \frac{1}{2} \right)}{M} \right) \tag{2.2}$$

เมื่อ

$MFCC(c)$ คือ ค่าสัมประสิทธิ์เซปสตรัมความถี่เมลลำดับที่ c

c คือ ดัชนีของสัมประสิทธิ์เซปสตรัม โดยทั่วไปเลือกเฉพาะลำดับต้น ๆ ที่มีสารสนเทศสำคัญ $c = 0, 1, \ldots, N_c - 1$

N_c คือ จำนวนสัมประสิทธิ์เซปสตรัมความถี่เมลที่เลือกใช้งาน

$E_{(m)}$ คือ ค่าพลังงานของสัญญาณเสียงในแถบความถี่เมลลำดับที่ m หลังจากผ่านชุดตัวกรองเมล

m คือ ดัชนีของตัวกรองเมล $m = 0, 1, \ldots, M - 1$

M คือ จำนวนตัวกรองเมลทั้งหมด

$\log(\cdot)$ คือ ฟังก์ชันลอการิทึม ซึ่งใช้เพื่อจำลองการรับรู้ความดังของเสียง

$\cos(\cdot)$ คือ ฟังก์ชันโคไซน์ ซึ่งเป็นส่วนหนึ่งของการแปลงโคไซน์แบบไม่ต่อเนื่อง

มหาวิทยาลัยเทคโนโลยีสุรนารี

สัมประสิทธิ์เซปสตรัมความถี่เมล มีความเหมาะสมสำหรับเสียงปอด เนื่องจากสามารถจับความแตกต่างของสเปกตรัมระหว่างกลุ่ม เช่น เสียงหวีด ที่มีพลังงานเด่นและต่อเนื่องในย่านความถี่บางช่วง หรือเสียงกรอบแตกที่เกิดเป็นช่วงสั้นและมีการเปลี่ยนแปลงฉับพลัน

2.2.3 คุณลักษณะเชิงพลวัตที่สะท้อนการเปลี่ยนแปลงเชิงเวลา

เสียงปอดภาวะผิดปกติมักมีลักษณะเปลี่ยนแปลงตามเวลา เช่น เสียงหวีดที่ปรากฏเด่นในช่วงหายใจออก หรือเสียงกรอบแตกที่เกิดเป็นเหตุการณ์สั้น ๆ ดังนั้น การเพิ่มคุณลักษณะเชิงพลวัตจึงช่วยให้แบบจำลองรับรู้การเปลี่ยนแปลงของสเปกตรัมตามเวลาได้ดียิ่งขึ้น โดยทั่วไปค่าอนุพันธ์อันดับหนึ่ง (Delta coefficient) ของสัมประสิทธิ์เซปสตรัมความถี่เมล สามารถคำนวณเป็นอนุพันธ์เชิงเวลาประมาณด้วยผลต่างถ่วงน้ำหนักดังสมการที่ (2.3)

$$\Delta c_t = \frac{\sum_{n=1}^{T} n\left(c_{t+n} - c_{t-n}\right)}{2\sum_{n=1}^{T} n^2} \tag{2.3}$$

เมื่อ

Δc_t คือ ค่าอนุพันธ์อันดับหนึ่งของสัมประสิทธิ์เซปสตรัมความถี่เมล ณ เฟรมเวลา t

c_t คือ ค่าสัมประสิทธิ์เซปสตรัมความถี่เมล ณ เฟรมเวลา t

T คือ จำนวนเฟรมด้านข้างที่ใช้ในการประมาณอนุพันธ์เชิงเวลา

n คือ ดัชนีผลรวมสำหรับการคำนวณเดลตา $(n = 1, \ldots, T)$

t คือ ดัชนีของเฟรมเวลา

สมการที่ (2.4) อนุพันธ์อันดับสอง (Delta–Delta coefficient) ของสัมประสิทธิ์เซปสตรัมความถี่เมล

$$\Delta^2 c_t = \Delta(\Delta c_t) \tag{2.4}$$

เมื่อ

$\Delta^2 c_t$ คือ ค่าอนุพันธ์อันดับสองของสัมประสิทธิ์เซปสตรัมความถี่เมล ณ เฟรมเวลา t

มหาวิทยาลัยเทคโนโลยีสุรนารี

2.2.4 การจัดรูปคุณลักษณะเพื่อใช้งานกับแบบจำลอง

เมื่อได้ชุดคุณลักษณะแล้วขั้นตอนถัดไป คือ การจัดรูปแบบให้เหมาะกับตัวจำแนก โดยตัวจำแนกแบบดั้งเดิม จะรับอินพุตเป็นเวกเตอร์คุณลักษณะ ขณะที่ตัวจำแนกแบบการเรียนรู้เชิงลึก เช่น โครงข่ายประสาทเทียมแบบคอนโวลูชัน สามารถรับอินพุตเป็นเวกเตอร์หรือภาพคุณลักษณะจากตัวแทนเวลา-ความถี่ หรือเมทริกซ์สัมประสิทธิ์เซปสตรัมความถี่เมล

หลักการทำงานของโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN) หลังการสกัดคุณลักษณะเชิงสเปกตรัม ประกอบด้วยชั้นคอนโวลูชันสำหรับเรียนรู้ตัวกรองและสร้างแผนที่คุณลักษณะจากข้อมูล ชั้นฟังก์ชันกระตุ้นเพื่อเพิ่มความไม่เชิงเส้น และชั้นการรวมค่า (Pooling) เพื่อลดมิติของข้อมูลและเพิ่มความทนทานต่อความแปรปรวนของสัญญาณ ในขั้นตอนสุดท้าย ชั้นเชื่อมโยงสมบูรณ์ (Fully connected layer) ทำหน้าที่สรุปสารสนเทศและตัดสินใจในการจำแนก ซึ่งช่วยให้การเรียนรู้รูปแบบของเสียงปอดจากตัวแทนสัมประสิทธิ์เซปสตรัมความถี่เมลในโดเมนเวลาและความถี่มีประสิทธิภาพมากขึ้น

แนวทางงานวิจัย เพื่อให้การสกัดคุณลักษณะสอดคล้องกับการจำแนกโรคระบบทางเดินหายใจแบบหลายกลุ่ม แนวทางสำคัญประกอบด้วย (1) การกรองสัญญาณด้วย ตัวกรองแถบผ่านแบบบัตเตอร์เวิร์ธ ย่าน 120 ถึง 1800 เฮิรตซ์ อันดับ 12 ตามสมการที่ (2.1) (2) การแบ่งสัญญาณระดับช่วงสัญญาณ 4 วินาที แบบไม่ซ้อนทับ (3) การคำนวณสัมประสิทธิ์เซปสตรัมความถี่เมล ตามสมการที่ (2.2) โดยควบคุมพารามิเตอร์ ใช้หน้าต่างแฮมมิง (Hamming window) และการทับซ้อน 50 เปอร์เซ็นต์ และ (4) การเพิ่มอนุพันธ์อันดับหนึ่งหรืออนุพันธ์อันดับสอง ตามสมการที่ (2.3) และ (2.4) เพื่อสะท้อนพลวัตของสัญญาณ ซึ่งเป็นองค์ประกอบที่เกี่ยวข้องกับสมรรถนะการจำแนก กล่าวโดยสรุป การสกัดคุณลักษณะเป็นขั้นตอนสำคัญที่ทำให้สัญญาณเสียงปอดซึ่งไม่คงที่ตามเวลาและมีสัญญาณรบกวนสูง สามารถถูกแทนด้วยชุดคุณลักษณะที่เหมาะต่อการเรียนรู้ของปัญญาประดิษฐ์ โดยสัมประสิทธิ์เซปสตรัมความถี่เมลและคุณลักษณะอนุพันธ์ตามเวลา มีบทบาทสำคัญในการจับคุณลักษณะเสียงที่สัมพันธ์กับโรค เช่น โรคปอดอักเสบ โรคหอบหืด และโรคปอดอุดกั้นเรื้อรัง ชี้ให้เห็นพารามิเตอร์และขั้นตอนสำคัญที่ส่งผลต่อประสิทธิภาพการจำแนก (Rangayyan, 2015; Bahoura, 2009; Sengupta et al., 2016)

2.3 วิธีการที่มีอยู่เดิม

2.3.1 โครงข่ายประสาทเทียมแบบคอนโวลูชัน

ในบริบทวิธีการที่มีอยู่เดิม งานจำแนกสัญญาณเสียงจำนวนมากสามารถจัดกลุ่มได้เป็น 2 กลุ่ม ได้แก่ แนวทางการเรียนรู้ของเครื่องแบบดั้งเดิมที่ใช้คุณลักษณะร่วมกับตัวจำแนก และแนวทางการเรียนรู้เชิงลึกที่ให้แบบจำลองเรียนรู้รูปแบบจากข้อมูลโดยตรง โดยเฉพาะโครงข่ายประสาทเทียมแบบคอนโวลูชัน ซึ่งเป็นสถาปัตยกรรมที่โดดเด่นสำหรับข้อมูลเชิงสองมิติ และสามารถประยุกต์กับสัญญาณเสียงได้เมื่อแปลงสัญญาณเสียงให้อยู่ในรูปภาพคุณลักษณะ (Feature image) เช่น สเปกโตรแกรมเมลแบบลอการิทึม หรือแผนภาพสัมประสิทธิ์เซปสตรัมความถี่เมล (O'Shea, K., and Nash, R., 2015; Piczak, K. J., 2015; Hershey, S., Chaudhuri, S., Ellis, D. P. W., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., Slaney, M., and Wilson, K., 2017)

ข้อมูลเสียงโดยธรรมชาติเป็นสัญญาณอนุกรมเวลา เมื่อถูกแปลงให้อยู่ในรูปตัวแทนเวลาและความถี่ เช่น สเปกโตรแกรม หรือเมลสเปกโตรแกรมแบบลอการิทึม สัญญาณเสียงจะมีโครงสร้างเชิงสองมิติในแกนเวลาและความถี่ ซึ่งสอดคล้องกับรูปแบบอินพุตของโครงข่ายประสาทเทียมแบบคอนโวลูชัน ทำให้แบบจำลองสามารถเรียนรู้รูปแบบที่ปรากฏบนระนาบเวลาและความถี่ได้อย่างมีประสิทธิภาพ แนวคิดดังกล่าวเป็นการผสานที่สำคัญระหว่างการสกัดคุณลักษณะและการประยุกต์ใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน โดยงานวิจัยด้านการจำแนกเสียงได้แสดงให้เห็นอย่างต่อเนื่องว่าวิธีการดังกล่าวสามารถให้ผลลัพธ์ที่ดีในงานจำแนกเสียงทั้งในบริบททั่วไปและงานด้านเสียงปอด (Piczak, 2015; Hershey et al., 2017; Perna, D. and Tagarelli, A., 2019; Zhang, Z., Li, H., Zhang, R., Wang, Y., and Guan, C., 2021)

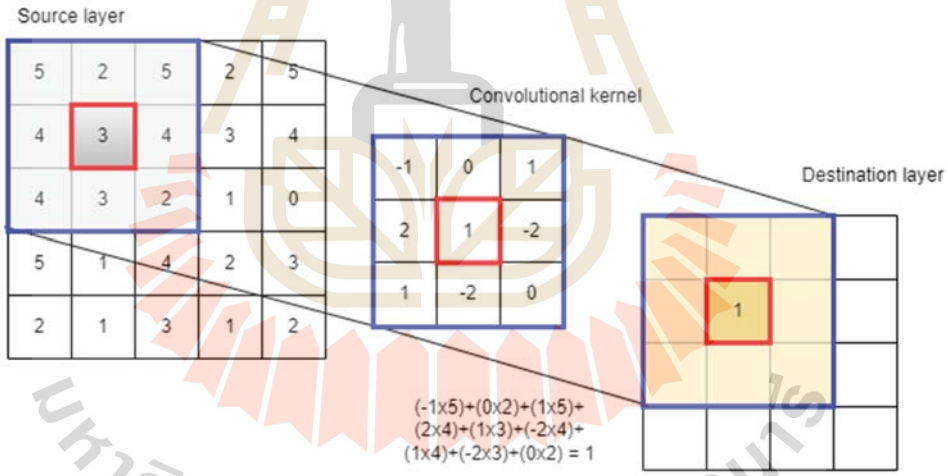
2.3.1.1 หลักการของโครงข่ายประสาทเทียมแบบคอนโวลูชัน

โครงข่ายประสาทเทียมแบบคอนโวลูชัน ประกอบด้วยชั้นคอนโวลูชัน (Convolutional layer) ซึ่งใช้ตัวกรองหรือเคอร์เนล (Kernel) เลื่อนผ่านอินพุตเพื่อคำนวณผลคูณจุด (Dot product) และสร้างแผนที่คุณลักษณะ แนวคิดสำคัญ คือ การเชื่อมต่อเฉพาะที่ (Local connectivity) และการใช้ค่าน้ำหนักร่วมกัน (Weight sharing) ทำให้แบบจำลองสามารถเรียนรู้รูปแบบที่เกิดซ้ำในบริเวณต่าง ๆ ของอินพุตได้อย่างมีประสิทธิภาพ (O'Shea and Nash, 2015) เมื่อกำหนดให้อินพุตเป็นเมทริกซ์สองมิติ และกำหนดตัวกรอง ดังสมการที่ (2.5)

$$Y_{(i.j)} = \left(\sum_{r=0}^{k_{h-1}} \sum_{s=0}^{k_{w-1}} X_{(i+r,j+s)} W_{(r,s)} + b_{nn} \right) \quad (2.5)$$

เมื่อ

$X_{(i,j)}$	คือ ค่าข้อมูลนำเข้า ณ ตำแหน่ง (i, j) ซึ่งอาจอยู่ในรูปของเมทริกซ์คุณลักษณะ
$W(r,s)$	คือ ค่าน้ำหนักของเคอร์เนล ณ ตำแหน่ง (r,s) โดยมีขนาด $k_h \times k_w$
b_{nn}	คือ ค่าไบแอสของโหนดประสาทในชั้นคอนโวลูชัน
$Y_{(i,j)}$	คือ ค่าผลลัพธ์บนแผนที่คุณลักษณะ ที่ตำแหน่ง (i,j)
i, j	คือ ดัชนีตำแหน่งของผลลัพธ์บนแผนที่คุณลักษณะ
r,s	คือ ดัชนีตำแหน่งภายในเคอร์เนล
kh และ kw	คือ ความสูงและความกว้างของเคอร์เนลตามลำดับ



รูปที่ 2.4 การเลื่อนเคอร์เนลบนอินพุตและการสร้างแผนที่คุณลักษณะ

การกำหนดพารามิเตอร์ของชั้นคอนโวลูชันมีผลต่อความละเอียดเชิงพื้นที่ของการเรียนรู้และขนาดของแผนที่คุณลักษณะ โดยเฉพาะค่าการเลื่อน (Stride) และการเติมขอบ (Padding) ซึ่งเป็นองค์ประกอบพื้นฐานในการออกแบบโครงข่ายประสาทเทียมแบบคอนโวลูชัน

ค่าการเลื่อน หมายถึง ระยะก้าวของการเลื่อนเคอร์เนล หากค่าการเลื่อนมีค่ามากขึ้น เคอร์เนลจะเลื่อนทีละหลายตำแหน่ง ส่งผลให้จำนวนตำแหน่งที่คำนวณลดลง ขนาดแผนที่

คุณลักษณะลดลง และระดับการซ้อนทับระหว่างหน้าต่างคอนโวลูชันลดลง สมการที่ (2.6) ความสัมพันธ์ของขนาดผลลัพธ์เมื่อไม่มีการเติมขอบ

$$Output\ size = \frac{I_{size} - F_s}{S_{stride}} + 1 \tag{2.6}$$

เมื่อ

I_{size} คือ ขนาดของข้อมูลนำเข้า (Input size) ในมิติที่พิจารณา

F_s คือ ขนาดของตัวกรองหรือเคอร์เนล

S_{stride} คือ ค่าการเลื่อนของตัวกรอง

$Output\ size$ คือ ขนาดของผลลัพธ์ที่ได้จากชั้นคอนโวลูชันหรือชั้นการรวมค่า

การเติมขอบ เป็นเทคนิคการเติมค่ารอบขอบอินพุต นิยมใช้การเติมศูนย์ (Zero-padding) เพื่อช่วยรักษาข้อมูลบริเวณขอบและควบคุมขนาดแผนที่คุณลักษณะไม่ให้เล็กลงอย่างรวดเร็วเมื่อความลึกของโครงข่ายเพิ่มขึ้น

ฟังก์ชันกระตุ้น (Active function) หลังชั้นโครงข่ายประสาทเทียมแบบคอนโวลูชัน ใช้ฟังก์ชันกระตุ้นแบบไม่เชิงเส้นเพื่อเพิ่มความสามารถในการเรียนรู้รูปแบบที่ซับซ้อน ฟังก์ชันเรกติไฟต์ลิเนียร์ยูนิต (Rectified Linear Unit: ReLU) เป็นฟังก์ชันกระตุ้นที่ช่วยลดปัญหาการหายไปของกราเดียนต์และเพิ่มความเร็วในการฝึกของโครงข่าย เป็นตัวเลือกที่นิยมเนื่องจากช่วยให้การเรียนรู้มีเสถียรภาพและลดปัญหาการไล่ระดับ ในบางกรณี (Nair, V., and Hinton, G. E., 2010)

$$\mathrm{Re}LU\left(x_{\mathrm{Re}LU}\right) = \max\left(0, x_{\mathrm{Re}LU}\right) \tag{2.7}$$

เมื่อ

$x_{\mathrm{Re}LU}$ คือ ข้อมูลนำเข้าของฟังก์ชันเรกติไฟต์ลิเนียร์ยูนิต

$\mathrm{Re}LU(\cdot)$ คือ ฟังก์ชันเรกติไฟต์ลิเนียร์ยูนิต

$\max\left(0, x_{\mathrm{Re}LU}\right)$ คือ ฟังก์ชันที่ให้ผลลัพธ์เป็นศูนย์เมื่ออินพุตมีค่าน้อยกว่าหรือเท่ากับศูนย์ และให้ผลลัพธ์เท่ากับอินพุตเมื่ออินพุตมีค่ามากกว่าศูนย์

มหาวิทยาลัยเทคโนโลยีสุรนารี

ชั้นการรวมค่า ใช้เพื่อลดมิติของข้อมูล ลดความไวต่อการเลื่อนตำแหน่ง และช่วยควบคุมความซับซ้อนการคำนวณของแบบจำลอง ในงานเสียงปอดการลดมิติและการเน้นสารสนเทศสำคัญมีความสำคัญ เนื่องจากอินพุตเวลาและความถี่มีขนาดใหญ่และมีสัญญาณรบกวน ในกรณีการจำแนกหลายกลุ่มนิยมใช้ฟังก์ชันซอฟต์แมกซ์ (Softmax) คือ ฟังก์ชันกระตุ้นที่ใช้ในชั้นเอาต์พุตของแบบจำลองจำแนก เพื่อแปลงเวกเตอร์ค่าคะแนนเชิงเส้น (logits) ให้เป็นการแจกแจงความน่าจะเป็นของแต่ละกลุ่ม โดยผลรวมของความน่าจะเป็นทั้งหมดมีค่าเท่ากับ 1 ดังสมการที่ (2.8)

$$p_k = \frac{e^{a_k}}{\sum_{j_{soft}=1}^{K_{soft}} e^{aj_{soft}}} \tag{2.8}$$

เมื่อ

a_k คือ ค่าลอจิกของกลุ่มที่ k ซึ่งเป็นผลลัพธ์จากชั้นเชื่อมโยงสมบูรณ์

k คือ ดัชนีของกลุ่มเป้าหมาย โดย $k = 1, 2, \ldots, K_{\text{soft}}$

j_{soft} คือ ดัชนีสำหรับการรวม โดย $j = 1, 2, \ldots, K_{\text{soft}}$

K_{soft} คือ จำนวนกลุ่มทั้งหมดของปัญหาการจำแนก

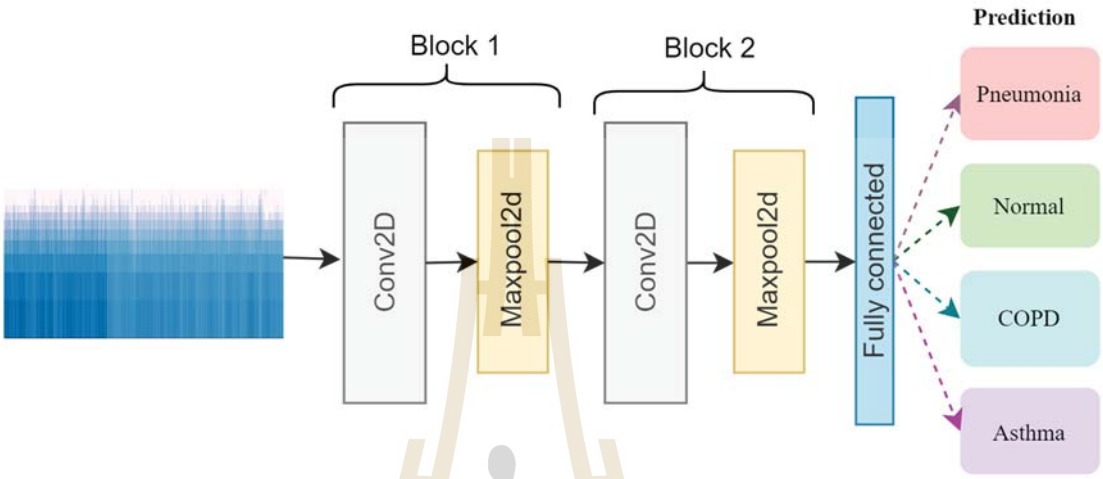
P_k คือ ความน่าจะเป็นที่แบบจำลองทำนายว่าข้อมูลอยู่ในกลุ่มที่ k

2.3.1.2 สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบคอนโวลูชัน

โครงข่ายประสาทเทียมเชิงลึก นำมาประยุกต์ใช้ในการตรวจคัดแยกโรคจากสัญญาณชีวภาพ ซึ่งมีความสามารถในการเรียนรู้รูปแบบเชิงพื้นที่และเชิงเวลาได้อย่างมีประสิทธิภาพ งานวิเคราะห์เสียงปอด Aykanat, M., Kılıç, Ö., Kurt, B., and Saryal, S. (2017) ได้เสนอการใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันสองชั้น (Convolutional Neural Network with Two Layers: CNN-2L) เป็นแบบจำลองพื้นฐานที่มีโครงสร้างไม่ซับซ้อน แต่มีประสิทธิภาพในการจำแนก

สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบคอนโวลูชันสองชั้น ประกอบด้วยชั้นคอนโวลูชันจำนวนสองบล็อก โดยแต่ละบล็อกประกอบด้วยชั้นคอนโวลูชัน (Conv2D) สำหรับสกัดคุณลักษณะเชิงพื้นที่จากข้อมูลนำเข้า และชั้นแม็กซ์พูลลิง (MaxPool2d) เพื่อทำการลดมิติของข้อมูลและเพิ่มความทนทานต่อการเลื่อนตำแหน่งของคุณลักษณะ หลังจากผ่าน

กระบวนการดังกล่าวข้อมูลจะถูกแปลงเป็นเวกเตอร์และส่งต่อไปยังชั้นเชื่อมโยงสมบูรณ์ เพื่อทำการจำแนกประเภทของเสียงปอดในขั้นสุดท้าย



รูปที่ 2.5 สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบคอนโวลูชันสองชั้น

ขั้นตอนการประมวลผลข้อมูลอินพุตของแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชันสองชั้น คือ คุณลักษณะสัมประสิทธิ์เซปสตรัมความถี่เมลถูกใช้เป็นตัวแทนเชิงเวลาและความถี่ของสัญญาณเสียง การใช้สัมประสิทธิ์เซปสตรัมความถี่เมลร่วมกับโครงข่ายประสาทเทียมแบบคอนโวลูชัน ช่วยให้แบบจำลองสามารถเรียนรู้รูปแบบของพลังงานและโครงสร้างเชิงสเปกตรัมที่เกี่ยวข้องกับลักษณะเสียงปอดภาวะปกติและภาวะผิดปกติ เพื่อให้การเปรียบเทียบมีสอดคล้องกับแนวทางงานวิจัยก่อนหน้า งานวิจัยนี้จึงนำโครงข่ายประสาทเทียมแบบคอนโวลูชันสองชั้นมาใช้เป็นแบบจำลองอ้างอิง (Baseline model) สำหรับการจำแนกโรคจากคุณลักษณะเสียงปอด เพื่อประเมินประสิทธิภาพของการเรียนรู้เชิงลึกเมื่อเทียบกับวิธีการที่อาศัยการสกัดคุณลักษณะ และตัวจำแนกแบบดั้งเดิม (Tripathy, R. K., Das, D., Ghose, A., and Ghosh, A., 2022)

2.4 วิธีการที่นำเสนอ

โครงข่ายประสาทเทียมแบบคอนโวลูชันจะมีประสิทธิภาพสูงในการจำแนกสัญญาณเสียงจากสเปกโตรแกรม แต่การฝึกแบบการจำแนกแบบงานเดียว (Single-task classification) ซึ่งมุ่งเน้นเฉพาะการเพิ่มความถูกต้องของการจำแนก อาจทำให้กระบวนการเรียนรู้อาศัยฟังก์ชันความสูญเสียที่สะท้อนเพียงเป้าหมายการแยกกลุ่มเพียงด้านเดียว ส่งผลให้แบบจำลองมีแนวโน้มเรียนรู้รูปแบบเชิง

สถิติที่เฉพาะเจาะจงกับชุดข้อมูลฝึกมากกว่าการเรียนรู้โครงสร้างเชิงกายภาพหรือสรีรวิทยาของสัญญาณเสียงปอด

การนำเสนอแนวคิดออโตเอนโคเดอร์แบบหลายงาน ซึ่งควบคุมให้ตัวเข้ารหัสเรียนรู้ตัวแทนเชิงแฝงที่รองรับทั้งการรักษาโครงสร้างสัญญาณผ่านการสร้างกลับคืน และการจำแนกโรคพร้อมกัน ทำให้การเรียนรู้มีความสมดุลและทำหน้าที่เป็นเทคนิคการควบคุมความซับซ้อนของแบบจำลอง ส่งผลให้ได้คุณลักษณะเชิงลึกที่มีเสถียรภาพ เหมาะกับข้อมูลเสียงปอดที่มีความซับซ้อนและสัญญาณรบกวน

2.4.1 ออโตเอนโคเดอร์แบบหลายงาน

ออโตเอนโคเดอร์ (Autoencoder: AE) เป็นโครงข่ายประสาทเทียมแบบไม่ต้องมีป้ายกำกับ หรือการเรียนรู้โดยไม่มีผู้สอน (Unsupervised learning) ที่มีวัตถุประสงค์หลักเพื่อเรียนรู้ตัวแทนเชิงแฝง (Latent representation) ของข้อมูล โดยให้แบบจำลองสามารถสร้างข้อมูลนำเข้าให้กลับคืนได้ใกล้เคียงที่สุดผ่านกระบวนการเข้ารหัส (Encoder) และถอดรหัส (Decoder) แนวคิดสำคัญ คือ การบีบอัดสารสนเทศจากข้อมูลมิติสูงให้เหลือแกนกลางที่สำคัญ แล้วใช้แกนกลางนั้นสร้างข้อมูลเดิมกลับคืน ซึ่งช่วยให้แบบจำลองเรียนรู้โครงสร้างภายในของข้อมูลได้

โดยทั่วไปออโตเอนโคเดอร์ประกอบด้วยฟังก์ชัน 2 ส่วน ได้แก่

1. ตัวเข้ารหัส ทำหน้าที่แปลงข้อมูลอินพุต (Input data) ไปเป็นตัวแทนเชิงแฝง
2. ตัวถอดรหัส ทำหน้าที่สร้างข้อมูลประมาณค่าจากตัวแทนเชิงแฝง

$$z = \phi(v) \tag{2.9}$$

$$\hat{v} = \psi(z) \tag{2.10}$$

เมื่อ

v คือ ข้อมูลอินพุต

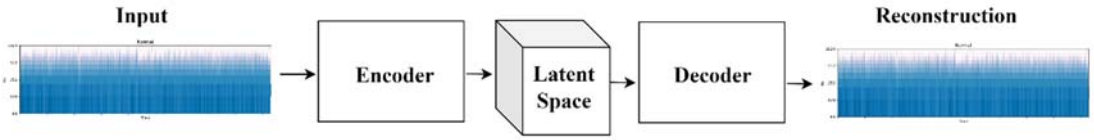
z คือ ตัวแทนเชิงแฝง

$\hat{v}$ คือ ข้อมูลที่ถูกสร้างขึ้นใหม่จากตัวแทนเชิงแฝง

$\phi(\cdot)$ คือ ฟังก์ชันของตัวเข้ารหัส (Encoder function) ซึ่งเรียนรู้การแปลงข้อมูลนำเข้าสู่ปริภูมิเชิงแฝง

$\psi(\cdot)$ คือ ฟังก์ชันของตัวถอดรหัส (Decoder function) ซึ่งเรียนรู้การสร้างข้อมูลประมาณค่ากลับจากตัวแทนเชิงแฝง

มหาวิทยาลัยเทคโนโลยีสุรนารี



รูปที่ 2.6 ออโตเอนโคเดอร์

ฟังก์ชันการสูญเสีย (Loss function) เป็นตัวชี้วัดเชิงปริมาณที่ใช้ประเมินระดับความผิดพลาดของแบบจำลอง โดยวัดความแตกต่างระหว่างค่าที่แบบจำลองทำนายได้กับค่าจริงของข้อมูล ฟังก์ชันดังกล่าวทำหน้าที่เป็นเกณฑ์ ในการปรับปรุงพารามิเตอร์ของแบบจำลองผ่านกระบวนการเรียนรู้แบบไล่ระดับ (Gradient-based optimization) เพื่อให้ค่าการสูญเสียลดลงอย่างต่อเนื่องและแบบจำลองสามารถอธิบายข้อมูลได้ดีขึ้น

ออโตเอนโคเดอร์วัตถุประสงค์หลักของการฝึก คือ การสร้างข้อมูลอินพุต ให้กลับคืนมาในรูปของข้อมูลที่ถูกสร้างขึ้นใหม่จากตัวแทนเชิงแฝง ให้มีความใกล้เคียงกับข้อมูลเดิมมากที่สุด ดังนั้น ฟังก์ชันการสูญเสียที่ใช้จึงมุ่งวัดความคลาดเคลื่อนระหว่างข้อมูลนำเข้าและข้อมูลที่สร้างกลับคืน โดยหนึ่งในรูปแบบที่นิยมใช้และได้รับการยอมรับอย่างแพร่หลาย คือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error: MSE) เพราะเป็นฟังก์ชันที่มีความต่อเนื่องและสามารถหาค่าอนุพันธ์ได้ง่าย ทำให้เหมาะสมกับการฝึกโครงข่ายประสาทเทียมด้วยวิธีการแพร่กระจายย้อนกลับ (Backpropagation) ให้โทษกับความคลาดเคลื่อนขนาดใหญ่สูงกว่าความคลาดเคลื่อนขนาดเล็ก ส่งผลให้แบบจำลองถูกบังคับให้เรียนรู้ตัวแทนเชิงแฝงที่รักษาโครงสร้างสำคัญของสัญญาณได้ดี มีความเหมาะสมกับข้อมูลเชิงปริมาณแบบต่อเนื่อง

การให้ค่าการสูญเสียจากการสร้างกลับคืนมีค่าน้อย ทำให้แบบจำลองไม่สามารถมุ่งเรียนรู้เฉพาะคุณลักษณะที่ช่วยการจำแนกเพียงอย่างเดียว แต่จำเป็นต้องเก็บรักษาสารสนเทศเชิงโครงสร้างของสัญญาณเสียงปอดไว้ในตัวแทนเชิงแฝง ส่งผลให้ได้คุณลักษณะเชิงลึกที่มีความเสถียรและเหมาะสมต่อการนำไปใช้ร่วมกับงานอื่น เช่น การจำแนกโรคหรือการเรียนรู้หลายงานในขั้นตอนถัดไป

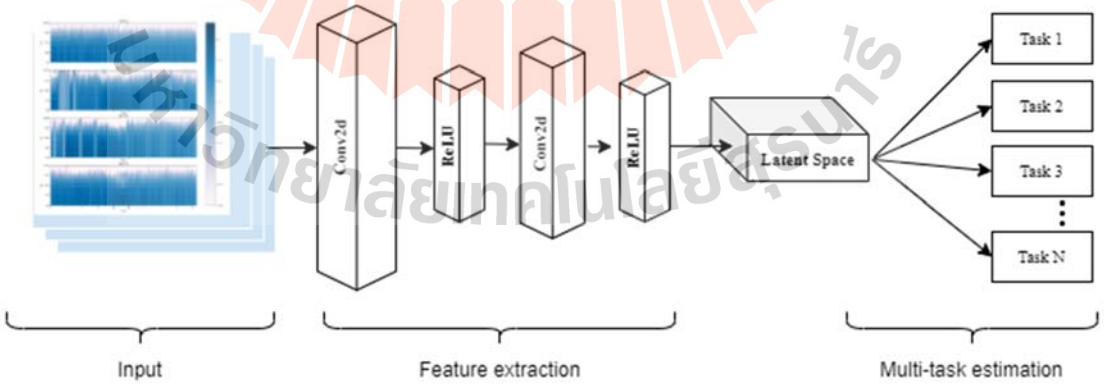
สมการฟังก์ชันการสูญเสียจากการสร้างข้อมูลกลับคืน (Reconstruction loss) ดังสมการที่ (2.11)

$$L_{recon} = \left(\frac{1}{n_{recon}} \right) \sum_{j_{recon}=1}^{n_{recon}} \left\| x_{j_{recon}} - \hat{x}_{j_{recon}} \right\|^2 \tag{2.11}$$

เมื่อ

L_{recon}	คือ ค่าฟังก์ชันการสูญเสียจากการสร้างข้อมูลกลับคืน
n_{recon}	คือ จำนวนตัวอย่างข้อมูลทั้งหมด
$x_{j_{recon}}$	คือ เวกเตอร์ข้อมูลของตัวอย่างลำดับที่ j
$\hat{x}_{j_{recon}}$	คือ เวกเตอร์ข้อมูลที่สร้างกลับคืนของข้อมูลตัวอย่างลำดับที่ j
$\|\cdot\|$	คือ ค่านอร์มแบบยุคลิด (Euclidean norm)

การเรียนรู้แบบหลายงาน แนวทางที่ให้แบบจำลองเรียนรู้มากกว่าหนึ่งวัตถุประสงค์พร้อมกัน โดยใช้ตัวแทนร่วม (Shared representation) เพื่อให้เกิดการถ่ายโอนสารสนเทศระหว่างงานช่วยลดโอกาสการเรียนรู้เกิน และเพิ่มความสามารถในการประยุกต์กับข้อมูลใหม่ ในบริบทของสัญญาณชีวภาพ การใช้การเรียนรู้หลายงานมีความเหมาะสมเพราะข้อมูลสัญญาณรบกวนสูงและความแปรปรวนระหว่างบุคคล



รูปที่ 2.7 การเรียนรู้แบบหลายงาน

ออโตเอนโคเดอร์แบบหลายงานเป็นการขยายแนวคิดออโตเอนโคเดอร์ให้สามารถทำอย่างน้อยสองงานพร้อมกัน ได้แก่ งานการสร้างกลับคืนของข้อมูล และงานการจำแนกประเภท โครงสร้างหลักประกอบด้วย ตัวเข้ารหัสร่วม และมีหัวงาน (Task heads) แยกสำหรับตัวถอดรหัสและตัวจำแนก (Classifier head) ตัวเข้ารหัสทำหน้าที่สร้างตัวแทนร่วม z ซึ่งถูกใช้ทั้งในการสร้าง $\hat{v}$ และการทำนายกลุ่ม $\hat{u}$ สมการการทำนายกลุ่มตัวจำแนก ดังสมการที่ (2.12)

$$\hat{u} = h(z) \tag{2.12}$$

เมื่อ

u คือ ป้ายกำกับจริงของข้อมูล

$\hat{u}$ คือ เวกเตอร์ความน่าจะเป็นของแต่ละกลุ่มที่ได้จากการทำนาย

$h(\cdot)$ คือ ฟังก์ชันตัวจำแนก

สำหรับงานจำแนกหลายกลุ่ม ใช้ฟังก์ชันการสูญเสียแบบครอสเอนโทรปี (Cross entropy loss) เป็นตัววัดความเหมาะสมของการทำนาย

$$L_{class} = -\sum_{k_{class}=1}^{K_{class}} y_{k_{class}} \log\left(\hat{y}_{k_{class}}\right) \tag{2.13}$$

เมื่อ

L_{class} คือ ค่าฟังก์ชันการสูญเสียสำหรับงานจำแนกประเภท

K_{class} คือ จำนวนกลุ่มทั้งหมดของการจำแนก

k_{class} คือ ดัชนีของกลุ่มการจำแนก โดย $k_{class} = 1, 2, \ldots, K_{class}$

$y_{k_{class}}$ คือ ค่าป้ายกำกับจริงของกลุ่มที่ k_{class}

$\hat{y}_{k_{class}}$ คือ ค่าความน่าจะเป็นที่แบบจำลองทำนายสำหรับกลุ่มที่ k_{class}

แนวคิดฟังก์ชันการสูญเสียของออโตเอนโคเดอร์แบบหลายงาน คือ การรวมค่าความสูญเสียของทั้งสองงานเข้าด้วยกัน เพื่อให้การเรียนรู้ของตัวเข้ารหัสถูกกำกับด้วยเป้าหมายคู่ (Dual objectives) สมการรวมค่าการสุญเสีย (Total loss) ดังสมการที่ (2.14)

มหาวิทยาลัยเทคโนโลยีสุรนารี

$$L_{total} = L_{recon} + L_{class} \quad (2.14)$$

เมื่อ

L_{total} คือ ฟังก์ชันการสูญเสียรวมของแบบจำลองออโตเอนโคเดอร์แบบหลายงาน

L_{recon} คือ ฟังก์ชันการสูญเสียจากการสร้างข้อมูลกลับคืน

L_{class} คือ ฟังก์ชันการสูญเสียสำหรับการจำแนก

การรวม L_{recon} ทำให้แบบจำลองรักษาโครงสร้างของสัญญาณและลดการสูญเสียสารสนเทศ ขณะที่ L_{class} ทำให้ตัวแทนเชิงแฝงมีความสามารถการจำแนกได้ดีเมื่อฝึกพร้อมกัน ตัวแทนเชิงแฝงจึงมีแนวโน้มเป็นคุณลักษณะเชิงลึกที่เสถียรมากกว่าการฝึกเพื่อจำแนกอย่างเดียว ซึ่งสอดคล้องกับเป้าหมายของงานตรวจจับโรคจากเสียงปอดที่มีสัญญาณรบกวนและความแปรปรวนสูง

2.4.2 แบบจำลองไฮบริด

ในบริบทของการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึก การออกแบบระบบที่มีบทบาทให้แต่ละองค์ประกอบทำหน้าที่ในสิ่งที่ถนัดและเชื่อมต่อกันเป็นกระบวนการเดียว เพื่อสร้างสมดุลระหว่างคุณภาพของตัวแทนข้อมูล (Data representation) และความคมชัดของขอบเขตการตัดสินใจ (Decision boundary) โดยแบบจำลองเชิงลึกสามารถเรียนรู้ตัวแทนเชิงแฝงที่ช่วยลดความซับซ้อน และสัญญาณรบกวนของข้อมูล ขณะที่ตัวจำแนกแบบดั้งเดิม โดยเฉพาะตัวจำแนกแบบเพิ่มระยะขอบ (Margin-based classifier) สามารถกำหนดขอบเขตการตัดสินใจที่มีเสถียรภาพและมีความสามารถในการทั่วไปได้ดี ข้อได้เปรียบของแนวทางแบบไฮบริดในกรณีที่ชุดข้อมูลสำหรับการฝึกมีจำนวนจำกัด เนื่องจากการเรียนรู้คุณลักษณะเชิงลึกช่วยแปลงข้อมูลดิบที่มีมิติสูงให้เป็นตัวแทนที่กระชับและมีความหมาย ส่งผลให้ตัวจำแนกสามารถทำงานได้อย่างมีประสิทธิภาพ โดยไม่ต้องเรียนรู้พารามิเตอร์จำนวนมาก จึงช่วยลดความเสี่ยงของการเรียนรู้เกิน เพิ่มความเสถียรของการตัดสินใจ และเอื้อให้ระบบมีความทนทานต่อสัญญาณรบกวนภายใต้ข้อจำกัดด้านข้อมูล (Bengio, Y., Courville, A., and Vincent, P., 2013; Goodfellow, I., Bengio, Y., and Courville, A., 2016; LeCun, Y., Bengio, Y., and Hinton, G., 2015)

สำหรับงานวิจัยนี้ แนวคิดไฮบริดถูกนำมาใช้ในรูปแบบตัวแทนเชิงลึก ผสมผสานกับตัวจำแนกแบบระยะขอบ โดยให้ออโตเอนโคเดอร์แบบหลายงานทำหน้าที่เรียนรู้ตัวแทนเชิงแฝงที่รองรับทั้งงานการสร้างกลับคืนและงานการจำแนกผ่านการเรียนรู้หลายงาน จากนั้นนำตัวแทนไปใช้กับซัพพอร์ตเวกเตอร์แมชชีนเป็นตัวจำแนกขั้นสุดท้ายแทนชั้นเชื่อมโยงสมบูรณ์ หลักการสำคัญ คือ

เมื่อคุณลักษณะที่ได้จากออโตเอนโคเดอร์แบบหลายงานมีการแยกกลุ่มเชิงโครงสร้างได้ดีขึ้น ซัพพอร์ตเวกเตอร์แมชชีนซึ่งอาศัยแนวคิดการเพิ่มระยะขอบสูงสุดจะช่วยค้นหาไฮเปอร์เพลนที่เหมาะสมและเพิ่มความคมชัดของการตัดสินใจ โดยเฉพาะในกรณีคลาสที่มีลักษณะใกล้เคียงกันซึ่งพบได้บ่อยในข้อมูลเสียงปอดและชีวสัญญาณทางการแพทย์ (Caruana, R., 1997; Cortes, C., and Vapnik, V., 1995)

ดังนั้น แนวทางไฮบริดในงานวิจัยนี้จึงเป็นการออกแบบเชิงระบบที่มีเหตุผลรองรับอย่างชัดเจน โดยออโตเอนโคเดอร์แบบหลายงานช่วยยกระดับคุณภาพและความเสถียรของตัวแทนคุณลักษณะ ขณะที่ซัพพอร์ตเวกเตอร์แมชชีนช่วยยกระดับกลไกการตัดสินใจให้มีความคมชัดและความสามารถในการทั่วไปสูงขึ้น ส่งผลให้กรอบแนวคิดการทำงานร่วมกันระหว่างออโตเอนโคเดอร์แบบหลายงานกับซัพพอร์ตเวกเตอร์แมชชีน มีศักยภาพในการเพิ่มทั้งความแม่นยำและความสามารถในการประยุกต์ใช้งานจริงภายใต้ข้อจำกัดของข้อมูลเสียงทางการแพทย์

ซัพพอร์ตเวกเตอร์แมชชีน เป็นวิธีการเรียนรู้ของเครื่องแบบมีผู้สอน ที่มีจุดเด่นด้านการจำแนกประเภท โดยอาศัยหลักการเพิ่มระยะขอบระหว่างกลุ่มข้อมูลเพื่อสร้างขอบเขตการตัดสินใจที่มีเสถียรภาพ คุณสมบัตินี้ทำให้ซัพพอร์ตเวกเตอร์แมชชีนถูกนำมาใช้อย่างแพร่หลายในแนวทางแบบจำลองไฮบริดร่วมกับแบบจำลองเชิงลึก เพราะสามารถทำงานได้ดีบนชุดคุณลักษณะที่ผ่านการสกัดหรือเรียนรู้แล้ว แม้ในกรณีที่จำนวนข้อมูลฝึกมีจำกัดหรือข้อมูลมีมิติสูง

ในกรณีข้อมูลเชิงเส้น (Linearly separable) จะหาสมการไฮเปอร์เพลน (Hyperplane) ที่เหมาะสมที่สุดเพื่อแยกข้อมูลสองกลุ่ม แนวคิดทางคณิตศาสตร์และการเรียนรู้ของเครื่องที่ หมายถึง พื้นผิวหรือขอบเขตที่ใช้แบ่งข้อมูลออกเป็นกลุ่ม ในปริภูมิหลายมิติ

$$w_{h_{yp}}^{Transpose} x_{h_{yp}} + b_{h_{yp}} = 0 \tag{2.15}$$

เมื่อ

$w_{h_{yp}}$ คือ เวกเตอร์ค่าน้ำหนักของไฮเปอร์เพลน

$x_{h_{yp}}$ คือ เวกเตอร์คุณลักษณะของข้อมูล

$b_{h_{yp}}$ คือ ค่าคงที่

แนวคิดของระยะขอบมากที่สุด โดยต้องการทำให้ขนาดของเวกเตอร์น้ำหนักเล็กที่สุดภายใต้เงื่อนไขการแยกกลุ่มที่ถูกต้อง ดังสมการที่ (2.16)

มหาวิทยาลัยเทคโนโลยีสุรนารี

$$\min\left(\frac{1}{2}\right)\left\|w_{h_{yp}}\right\|^2 s.t. y_q\left(w_{h_{yp}}^{Transpose}x_q + b_{h_{yp}}\right) \geq 1 \quad (2.16)$$

เมื่อ

$w_{h_{yp}}$ คือ เวกเตอร์ค่าน้ำหนักของไฮเปอร์เพลน

y_q คือ ป้ายกำกับจริงของตัวอย่างลำดับที่ q โดยมีค่าอยู่ในเซต $\{-1,+1\}$ สำหรับกรณีการจำแนกสองกลุ่ม

x_q คือ เวกเตอร์คุณลักษณะของตัวอย่างลำดับที่ q

$b_{h_{yp}}$ คือ ค่าคงที่

$\left\|w_{h_{yp}}\right\|$ คือ ค่านอร์มแบบยูคลิดของเวกเตอร์น้ำหนัก

เงื่อนไข $y_q\left(w_{h_{yp}}^{Transpose}x_q + b_{h_{yp}}\right) \geq 1$ ใช้กำหนดให้ข้อมูลทุกตัวอย่างถูกจำแนกอย่างถูกต้อง และอยู่บนหรือภายนอกเส้นระยะขอบ

ในทางปฏิบัติ ข้อมูลไม่สามารถแยกได้สมบูรณ์จึงเพิ่มตัวแปรผ่อนผัน (Slack variables) เพื่อให้แบบจำลองยืดหยุ่นต่อจุดข้อมูลที่ทับซ้อนกัน และควบคุมด้วยพารามิเตอร์ c_{svm} สมการซัพพอร์ตเวกเตอร์แมชชีนแบบซอฟต์มาร์จิน (Soft-margin SVM)

$$\min\left(\frac{1}{2}\right)\left\|w_{h_{yp}}\right\|^2 + c_{svm}\sum_{sv=1}^{n_{svm}} y_{sv}\left(w_{h_{yp}}^{Transpose}x_{sv} + b_{h_{yp}}\right) \geq 1-\xi_{sv}, \xi_{sv} \geq 0 \quad (2.17)$$

เมื่อ

$w_{h_{yp}}$ คือ เวกเตอร์ค่าน้ำหนักของไฮเปอร์เพลน

y_{sv} คือ ป้ายกำกับจริงของตัวอย่างลำดับที่ sv โดยมีค่าอยู่ในเซต $\{-1,+1\}$ สำหรับกรณีการจำแนกสองกลุ่ม

x_{sv} คือ เวกเตอร์คุณลักษณะของตัวอย่างลำดับที่ sv

$b_{h_{yp}}$ คือ ค่าคงที่

n_{svm} คือ จำนวนตัวอย่างข้อมูลทั้งหมด

ξ_{sv} คือ ค่าตัวแปรผ่อนผันของตัวอย่างลำดับที่ i ใช้วัดระดับการละเมิดเงื่อนไขระยะขอบ

มหาวิทยาลัยเทคโนโลยีสุรนารี

c_{svm} คือ พารามิเตอร์ควบคุมการแลกเปลี่ยนระหว่างความกว้างของระยะขอบและความผิดพลาดจากการจำแนก

$\left\| w_{h_{yp}} \right\|^2$ คือ กำลังสองของค่านอร์มแบบยูคลิดของเวกเตอร์น้ำหนัก

เคอร์เนลและซัพพอร์ตเวกเตอร์แมชชีนสำหรับข้อมูลไม่เชิงเส้น เพื่อแปลงข้อมูลจากปริภูมิดั้งเดิมไปยังปริภูมิที่แยกได้ดีขึ้นโดยไม่ต้องคำนวณการแปลงอย่างชัดเจน ฟังก์ชันตัดสินใจของซัพพอร์ตเวกเตอร์แมชชีน ในรูปแบบทวิภาค ดังสมการที่ (2.18)

$$f(x_{pre}) = sign\left(\sum_{d=1}^{n_{data}} \alpha_d \gamma_d Q(x_d, x_{pre}) + b_{model} \right) \tag{2.18}$$

เมื่อ

$f(x_{pre})$ คือ ฟังก์ชันตัดสินใจสำหรับการจำแนกกลุ่มของข้อมูลตัวอย่าง x_{pre}

α_d คือ ค่าสัมประสิทธิ์จากการแก้ปัญหาในรูปทวิภาค

γ_d คือ ป้ายกำกับจริงของตัวอย่างลำดับที่ d โดยมีค่าอยู่ในเซต $\{-1, +1\}$

x_d คือ เวกเตอร์คุณลักษณะของตัวอย่างลำดับที่ d ในชุดข้อมูลฝึก

x_{pre} คือ เวกเตอร์คุณลักษณะของข้อมูลตัวอย่างที่ต้องการทำนาย

b_{model} คือ ค่าคงที่ของแบบจำลอง

n_{data} คือ จำนวนตัวอย่างข้อมูลทั้งหมด

$Q(x_d, x_{pre})$ คือ ฟังก์ชันเคอร์เนลที่ใช้วัดความคล้ายคลึงระหว่างตัวอย่าง x_d และ x_{pre}

$sign(\cdot)$ คือ ฟังก์ชันกำหนดกลุ่มผลลัพธ์จากเครื่องหมายของค่าที่คำนวณได้

เคอร์เนลที่ใช้ คือ เคอร์เนลฟังก์ชันเรเดียล (Radial Basis Function: RBF kernel) ซึ่งมีความเหมาะสมกับข้อมูลสัญญาณชีวภาพที่มีความไม่เชิงเส้นสูง

$$Q(x_d, x_{pre}) = \exp\left(-\gamma \left\| x_d - x_{pre} \right\|^2 \right) \tag{2.19}$$

เมื่อ

γ คือ พารามิเตอร์ที่ควบคุมความกว้างของเคอร์เนล

มหาวิทยาลัยเทคโนโลยีสุรนารี

จากข้อจำกัดของแบบจำลองเชิงลึกที่มุ่งเน้นการจำแนกเพียงอย่างเดียว และศักยภาพของการเรียนรู้หลายงานผ่านออโตเอนโค้ดหลายงาน งานวิจัยนี้จึงเสนอแนวทางซึ่งผสานข้อดีของการเรียนรู้ตัวแทนเชิงแฝงและการตัดสินใจแบบมาร์จินเข้าด้วยกัน โดยมีแนวคิดหลัก ได้แก่ ใช้ออโตเอนโคเดอร์แบบหลายงานเพื่อเรียนรู้ตัวแทนเชิงแฝง ที่รักษาโครงสร้างของข้อมูลเสียงปอดผ่านงานการสร้างกลับคืน พร้อมกับเสริมความสามารถในการจำแนกผ่านงานจำแนกประเภท และใช้ซัพพอร์ตเวกเตอร์แมชชีนเป็นตัวจำแนกขั้นสุดท้ายแทนชั้นเชื่อมโยงสมบูรณ์ของโครงข่าย เพื่อเพิ่มเสถียรภาพของการจำแนก โดยเฉพาะในกรณีที่ข้อมูลมีจำนวนจำกัดและมีลักษณะไม่เชิงเส้น ตัวแทนเชิงแฝงที่ได้จากออโตเอนโคเดอร์แบบหลายงานสามารถมองเป็นคุณลักษณะเชิงลึกที่เรียนรู้โดยตรงจากข้อมูลเสียงปอด จากนั้นจึงนำไปใช้เป็นอินพุตของซัพพอร์ตเวกเตอร์แมชชีนเพื่อกำหนดขอบเขตการตัดสินใจในปริภูมิของตัวแทนเชิงแฝง ดังแสดงในสมการที่ (2.20)

$$f(z) = sign\left(\sum_{d=1}^{n_{data}} \alpha_d \gamma_d Q(x_d, x_{pre}) + b_{\mathrm{model}} \right) \tag{2.20}$$

โดยในสมการที่ (2.20) ตัวจำแนกซัพพอร์ตเวกเตอร์แมชชีน ทำงานบนเวกเตอร์ตัวแทนเชิงแฝง แทนการทำงานบนข้อมูลอินพุตโดยตรง ซึ่งช่วยให้การตัดสินใจอาศัยคุณลักษณะที่มีมิติและโครงสร้างเหมาะสมต่อการหาขอบเขตการตัดสินใจแบบระยะห่างสูงสุด

2.5 ปริทัศน์วรรณกรรม

จากการทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องด้านการตรวจจับโรคระบบทางเดินหายใจจากเสียงปอด พบว่างานวิจัยในสาขานี้มีการศึกษาอย่างต่อเนื่อง ทั้งในด้านการประมวลผลสัญญาณ การสกัดคุณลักษณะ และแบบจำลองการเรียนรู้ของเครื่อง โดยสามารถจัดกลุ่มแนวทางที่สำคัญตามลักษณะวิธีการ ได้แก่ แนวทางที่อาศัยการสกัดคุณลักษณะร่วมกับการเรียนรู้ของเครื่องแบบดั้งเดิม ซึ่งมุ่งออกแบบตัวแทนสัญญาณ เช่น คุณลักษณะเชิงเวลาและความถี่ และเชปสตรัม แล้วใช้ตัวจำแนกในการแยกกลุ่มเสียง การเรียนรู้เชิงลึกแบบต้นทางถึงปลายทาง (End-to-end deep learning) ที่ให้แบบจำลองเรียนรู้รูปแบบเสียงจากตัวแทนสัญญาณ เช่น สเปกโตรแกรม หรือจากสัญญาณดิบโดยตรง เพื่อเพิ่มศักยภาพในการจำแนก และแนวทางที่เริ่มผสานการเรียนรู้เชิงตัวแทน (Representation learning) หรือการเรียนรู้แบบหลายงาน เพื่อรองรับข้อจำกัดของข้อมูลติดป้ายกำกับที่มีจำนวนจำกัด และเพิ่มความทนทานของแบบจำลองต่อความแปรปรวนของข้อมูล แนวทางเหล่านี้ช่วยยกระดับ

สมรรถนะของระบบอย่างมีนัยสำคัญ งานวิจัยบางส่วนยังพบข้อจำกัดเกี่ยวกับความไม่สมดุลของข้อมูล ในหัวข้อต่อไปจึงนำเสนอการปริทัศน์วรรณกรรมโดยจัดเรียงตามกลุ่มแนวทางดังกล่าว เพื่อให้เห็นจุดเด่น ข้อจำกัด และประเด็นที่นำไปสู่การกำหนดแนวคิดของงานวิจัยนี้

Bahoura and Pelletier (2003) ศึกษาปัญหาการตรวจจับเสียงหวีดจากเสียงหายใจ ซึ่งเป็นเสียงภาวะผิดปกติที่มีลักษณะไม่คงที่และถูกรบกวนด้วยสัญญาณรบกวนสูง งานวิจัยเสนอการใช้คุณลักษณะเชิงเซปสตรัม สัมประสิทธิ์เซปสตรัมความถี่เมลร่วมกับวิธีจำแนกแบบการควอนไทซ์เชิงเวกเตอร์ (Vector Quantization: VQ) เพื่อแยกเสียงหวีดออกจากเสียงภาวะปกติ โดยทำการเปรียบเทียบกับคุณลักษณะประเภทอื่น เช่น การแปลงฟูเรียร์แบบเร็ว (Fast Fourier Transform: FFT) และการเข้ารหัสเชิงพยากรณ์เชิงเส้น (Linear Predictive Coding: LPC) จากผลการทดลองพบว่าวิธีที่ใช้สัมประสิทธิ์เซปสตรัมความถี่เมลให้ประสิทธิภาพสูงกว่า โดยมีค่าความถูกต้องมากกว่า 90 เปอร์เซ็นต์ เมื่อเทียบกับวิธีอื่น งานวิจัยนี้ชี้ให้เห็นถึงความเหมาะสมของคุณลักษณะเชิงสเปกตรัมสำหรับการวิเคราะห์เสียงปอด แม้ว่ายังจำกัดอยู่ที่การจำแนกแบบสองกลุ่มและใช้การออกแบบคุณลักษณะโดยบุคคล

Sengupta et al. (2015) มุ่งศึกษาปัญหาความไม่เสถียรของการจำแนกเสียงปอดซึ่งเกิดจากสัญญาณรบกวนและความแปรปรวนของเสียง งานวิจัยทำการเปรียบเทียบคุณลักษณะเชิงเซปสตรัมหลายชนิด ได้แก่ สัมประสิทธิ์เซปสตรัมความถี่เมล, ค่าสัมประสิทธิ์เซปสตรัมความถี่เชิงเส้น (Linear Frequency Cepstral Coefficients: LFCC), ค่าสัมประสิทธิ์เซปสตรัมจากการพยากรณ์เชิงเส้น (Linear Predictive Cepstral Coefficients: LPCC) และค่าสัมประสิทธิ์เซปสตรัมจากการพยากรณ์เชิงเส้นเชิงการรับรู้ (Perceptual Linear Prediction Cepstral Coefficients: PLPCC) และใช้โครงข่ายประสาทเทียมแบบฟีดฟอร์เวิร์ด (Feedforward Neural Network: FNN) เป็นตัวจำแนก ข้อมูลทดลองประกอบด้วยเสียงปอดประเภทต่าง ๆ เช่น เสียงภาวะปกติ เสียงหวีด และเสียงแตกกรอบ ผลการทดลองสัมประสิทธิ์เซปสตรัมความถี่เมลให้ผลการจำแนกดีที่สุด โดยมีค่าความถูกต้องมากกว่า 92 เปอร์เซ็นต์ และมีความไวต่อเสียงภาวะผิดปกติที่มีประสิทธิภาพ แสดงให้เห็นบทบาทสำคัญของการเลือกคุณลักษณะที่เหมาะสมต่อประสิทธิภาพของการจำแนกเสียงปอด

Li et al. (2016) เสนอแนวทางการจำแนกเสียงปอดโดยอัตโนมัติภายใต้สภาพแวดล้อมจริงซึ่งมีลักษณะเชิงเวลาและสัญญาณรบกวนสูง งานวิจัยนำเสนอวิธีการแบบผสมระหว่าง โครงข่ายประสาทเทียมเชิงลึก (Deep Neural Network: DNN) และแบบจำลองมาร์คอฟฮิดเดน (Hidden Markov Model: HMM) โดยใช้สัมประสิทธิ์เซปสตรัมความถี่เมลเป็นอินพุต และใช้โครงข่ายประสาทเทียมเชิงลึกเพื่อประมาณค่าความน่าจะเป็นของสถานะในแบบจำลองมาร์คอฟซ่อนฮิดเดน ผลการ

ทดลองพบว่าวิธีนี้ให้ประสิทธิภาพสูงกว่าวิธีแบบจำลองมาร์คอฟซ์ฮิดเดนที่ใช้การผสมแบบเกาส์เซียนอย่างมีนัยสำคัญ โดยให้ความถูกต้องสูงขึ้นและลดอัตราความผิดพลาดลง งานนี้แสดงให้เห็นศักยภาพของการผสานการเรียนรู้เชิงลึกเข้ากับแบบจำลองเชิงลำดับเวลาในงานวิเคราะห์เสียงปอด

Chambres, G., Hanna, P., and Desainte-Catherine, M. (2018) ตั้งข้อสังเกตว่าการตัดสินใจจากเสียงปอดในระดับรอบการหายใจ (Cycle-level) อาจไม่สะท้อนสถานะของผู้ป่วยได้อย่างเหมาะสม งานวิจัยจึงเสนอการตรวจจับผู้ป่วยโรคทางเดินหายใจด้วยการตัดสินใจในระดับผู้ป่วย (Patient-level decision) โดยรวมผลจากหลายรอบการหายใจของผู้ป่วยรายเดียวกัน ข้อมูลทดลองใช้ฐานข้อมูล ICBHI ผลการทดลองแสดงว่าการตัดสินใจในระดับผู้ป่วยช่วยเพิ่มค่าความไว (Sensitivity) และความจำเพาะ (Specificity) อย่างมีนัยสำคัญ งานนี้เน้นย้ำความสำคัญของการสรุปผลในระดับผู้ป่วยเพื่อให้สอดคล้องกับการใช้งานจริงทางคลินิก

Perna and Tagarelli (2019) ศึกษาการตรวจจับโรคทางเดินหายใจจากเสียงโดยใช้การเรียนรู้เชิงลึก เพื่อแก้ข้อจำกัดของวิธีการแบบดั้งเดิมที่พึ่งพาการออกแบบคุณลักษณะด้วยมือ งานวิจัยเสนอการใช้โครงข่ายประสาทเทียมเชิงคอนโวลูชันแบบการเรียนรู้เชิงลึกแบบต้นทางถึงปลายทาง โดยใช้ตัวแทนสัญญาณในโดเมนเวลาและความถี่เป็นอินพุต และทำการประเมินบนฐานข้อมูลเสียงปอดมาตรฐาน ผลการทดลอง พบว่าแนวทางเรียนรู้เชิงลึกให้ประสิทธิภาพที่เทียบเท่าหรือดีกว่าวิธีการเรียนรู้ของเครื่องแบบดั้งเดิม งานนี้แสดงถึงศักยภาพของโครงข่ายประสาทเทียมเชิงคอนโวลูชันสำหรับการวิเคราะห์เสียงปอด

Zhang et al. (2021) กล่าวถึงความท้าทายของการจำแนกเสียงปอดซึ่งเสียงภาวะผิดปกติจะเกิดขึ้นในช่วงเวลาสั้นและมีความแตกต่างระหว่างกลุ่มไม่ชัดเจน งานวิจัยเสนอการใช้โครงข่ายประสาทเทียมแบบคอนโวลูชันร่วมกับกลไกความสนใจโดยใช้เมลสเปกโตรแกรมเป็นอินพุต เพื่อเน้นพื้นที่เชิงเวลาและความถี่ที่สำคัญ ผลการทดลองแสดงแบบจำลองที่ใช้ให้ประสิทธิภาพการจำแนกที่ดีกว่าโครงข่ายประสาทเทียมเชิงคอนโวลูชันแบบดั้งเดิม โดยเฉพาะในการจำแนกเสียงภาวะผิดปกติ แสดงให้เห็นบทบาทของสถาปัตยกรรมในการเพิ่มความสามารถของแบบจำลองการเรียนรู้เชิงลึก

Engin, M., Engin, A. Z., Senol, Y., Memis, S., Yıldırım, A., and Kafalı, H. (2022) เสนอระบบอัตโนมัติสำหรับการจำแนกเสียงปอดที่รวมขั้นตอนการตรวจจับรอบการหายใจและการจำแนกเสียงเข้าด้วยกัน ระบบใช้การสกัดคุณลักษณะจากโดเมนเวลาและความถี่ เช่น สัมประสิทธิ์เซปสตรัมความถี่เมล และการเข้ารหัสเชิงพยากรณ์เชิงเส้น แล้วทำการจำแนกเสียงปอดหลายกลุ่ม ได้แก่ เสียงภาวะปกติ เสียงครืดคราดต่ำ และเสียงแตกกรอบ หลายประเภท จากผลการทดลองระบบที่เสนอให้ค่าความถูกต้องสูงสุด 90.63 เปอร์เซ็นต์ และมีค่าพื้นที่ใต้โค้ง (Area Under the Curve: AUC) เฉลี่ย

0.94 แสดงให้เห็นความสำคัญของการรวมกระบวนการประมวลผลสัญญาณเข้ากับการจำแนกเสียงปอดแบบอัตโนมัติ

Xia, Y., Han, J., and Mascolo, C. (2022) นำเสนอบทความปริทัศน์เกี่ยวกับการตรวจคัดกรองโรคทางเดินหายใจจากสัญญาณเสียง โดยครอบคลุมฐานข้อมูล วิธีการประมวลผลสัญญาณ และเทคนิคการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึกที่ใช้ในงานที่ผ่านมา ชี้ให้เห็นว่างานจำนวนมากรายงานค่าความถูกต้องในระดับสูง แต่ยังคงประสบปัญหาความไม่สมดุลของข้อมูล ความสามารถในการอธิบายผลลัพธ์ และการนำไปใช้งานจริง งานนี้ทำหน้าที่เป็นภาพรวมที่สะท้อนความท้าทายและทิศทางการวิจัยในอนาคตของการวิเคราะห์เสียงปอด

Soni, N., Chowdhury, A., Bhatia, M., and Chakrabarti, I. (2022) ศึกษาปัญหาการขาดแคลนข้อมูลที่มีการติดป้ายกำกับในงานวิเคราะห์เสียงปอด โดยเสนอแนวทางการเรียนรู้แบบไม่ต้องใช้ป้ายกำกับด้วยเทคนิคการเรียนรู้แบบเปรียบต่าง (Contrastive learning) เพื่อเรียนรู้ตัวแทนข้อมูลของสัญญาณเสียง วิธีการใช้ข้อมูลทางคลินิกเพื่อกำหนดคู่ข้อมูลเชิงบวกและเชิงลบ ผลการทดลองพบว่าแนวทางการเรียนรู้แบบกำกับตนเอง (Self-supervised learning) ที่นำเสนอสามารถปรับปรุงประสิทธิภาพของการจำแนกในขั้นงานปลายน้ำ (Downstream task) ได้อย่างมีนัยสำคัญ โดยเฉพาะในกรณีที่มีข้อมูลติดป้ายกำกับเพียงส่วนน้อย แสดงให้เห็นศักยภาพของการเรียนรู้ตัวแทนข้อมูลสำหรับการวิเคราะห์เสียงปอดภายใต้ข้อจำกัดของข้อมูล

Hassan, U., Singhal, A., and Chaudhary, P. (2024) นำเสนอ EasyNet ซึ่งเป็นโครงข่ายประสาทเทียมเชิงลึกที่มีโครงสร้างเรียบง่ายสำหรับการจำแนกเสียงปอด โดยออกแบบให้เรียนรู้จากสัญญาณเสียงปอดแบบดิบโดยตรงโดยไม่ต้องแปลงเป็นสเปกโตรแกรมหรือสกัดคุณลักษณะด้วยมือ การทดลองครอบคลุมการจำแนกแบบสองกลุ่ม แบบหลายกลุ่ม และแบบหลายป้ายกำกับ บนฐานข้อมูลเสียงปอดหลายชุด ผลการทดลองแสดงให้เห็นว่าแบบจำลองที่นำเสนอให้ค่าความถูกต้อง ความไว และความจำเพาะในระดับสูง งานนี้ชี้ให้เห็นว่าแบบจำลองที่มีโครงสร้างไม่ซับซ้อนสามารถให้ประสิทธิภาพที่แข่งขันได้ในการวิเคราะห์เสียงปอด

จากการปริทัศน์วรรณกรรมพบว่า งานวิจัยจำนวนมากสามารถพัฒนาแบบจำลองเพื่อวิเคราะห์เสียงปอดได้อย่างมีประสิทธิภาพภายใต้เงื่อนไขการทดลองของตนเอง แต่เมื่อพิจารณาเชิงลึกยังพบประเด็นที่น่าสนใจหลายด้าน ได้แก่ ความแปรปรวนของสัญญาณจากแหล่งบันทึกและสภาวะแวดล้อมที่แตกต่างกัน ความไม่สมดุลของข้อมูลซึ่งส่งผลต่อความไวของการตรวจจับกลุ่มผิดปกติ ตลอดจนความจำเป็นในการกำหนดขั้นตอนประเมินผลให้สะท้อนการใช้งานจริง โดยเฉพาะการออกแบบกระบวนการเตรียมข้อมูล การกำหนดพารามิเตอร์ เช่น เฟรม ช่วงสัญญาณ รอบการหายใจ

เป็นต้น การเลือกตัวชี้วัดที่เหมาะสมเพื่อประเมินความสามารถของแบบจำลอง นอกจากนี้ แนวทางดังกล่าวยังแสดงให้เห็นความสำคัญของการเรียนรู้ตัวแทนสัญญาณที่ดีและการออกแบบสถาปัตยกรรมที่สามารถเน้นสารสนเทศสำคัญของเสียงภาวะผิดปกติได้ ซึ่งเป็นปัจจัยที่มีผลต่อประสิทธิภาพและความน่าเชื่อถือ

บทที่ 3
วิธีการดำเนินการวิจัย

นำเสนอรายละเอียดระเบียบวิธีวิจัยและขั้นตอนการทดลองของงานวิจัยอย่างเป็นระบบ อธิบายกระบวนการออกแบบและการพัฒนาแบบจำลองที่นำเสนอ รวมถึงวิธีการประเมินประสิทธิภาพของแบบจำลองในการตรวจจับโรคปอดโดยใช้สัญญาณเสียงจากปอดผ่านโครงข่ายประสาทเทียมลึกแบบหลายงาน

3.1 ชุดข้อมูลและการเตรียมสัญญาณเสียงปอด

3.1.1 ชุดข้อมูลเสียงปอด

งานวิจัยนี้ใช้ชุดข้อมูลเสียงปอดจากฐานข้อมูลที่เผยแพร่สาธารณะ (Fraiwan, M., Fraiwan, L., Khassawneh, B., and Ibnian, A. A., 2021) ฐานข้อมูลที่ใช้สำหรับงานจำแนกเสียงปอด ซึ่งเป็นชุดข้อมูลที่มีการเก็บข้อมูลจากโรงพยาบาล King Abdullah University Hospital (KAUH) อย่างไรก็ตาม งานวิจัยนี้เลือกใช้ข้อมูล KAUH ในการทดลองเพื่อประเมินประสิทธิภาพของวิธีการที่นำเสนอ เพราะมีความเหมาะสมต่อการทดลองแบบการตรวจสอบไขว้แบบห้าส่วน โดยแยกรายบุคคล (Subject-independent fivefold cross-validation) เพราะในแต่ละกลุ่มโรคมีจำนวนอาสาสมัครมากกว่า 4 ราย (ปกติ, โรคหอบหืด, โรคปอดอักเสบ, โรคปอดอุดกั้นเรื้อรัง) จึงเหมาะสมต่อการแบ่งส่วนข้อมูลโดยหลีกเลี่ยงการปะปนของข้อมูลบุคคลเดียวกันระหว่างชุดฝึกและชุดทดสอบ

การทดลองดำเนินการบนชุดข้อมูลสาธารณะ โดยมีการจัดกลุ่มผู้เข้าร่วมตามสภาวะสุขภาพ 4 กลุ่ม ได้แก่ กลุ่มภาวะปกติ 35 ราย โรคหอบหืด 32 ราย โรคปอดอักเสบ 5 ราย และโรคปอดอุดกั้นเรื้อรัง 9 ราย สัญญาณเสียงปอดถูกบันทึกที่ความถี่การสุ่มตัวอย่าง 4 กิโลเฮิรตซ์ รายละเอียดเชิงประชากรศาสตร์ (ช่วงอายุ และสัดส่วนเพศ) รวมทั้งจำนวนสัญญาณหลังการแบ่งช่วงสัญญาณ ดังตารางที่ 3.1

ตารางที่ 3.1 ข้อมูลประชากรของกลุ่มตัวอย่างจากฐานข้อมูล KAUH

กลุ่มตัวอย่าง	จำนวนอาสาสมัคร (คน)	ช่วงอายุ (ปี)	เพศ (หญิง, ชาย)	จำนวนสัญญาณเสียงแบ่งส่วน (*จำนวนสัญญาณ)
ภาวะปกติ	35	18–81	หญิง 11, ชาย 24	415 (105)
โรคหอบหืด	32	12–72	หญิง 17, ชาย 15	357 (96)
โรคปอดอักเสบ	5	36–70	หญิง 2, ชาย 3	66 (15)
โรคปอดอุดกั้นเรื้อรัง	9	42–76	หญิง 1, ชาย 8	102 (27)

หมายเหตุ *ตัวเลขในวงเล็บแสดงจำนวนสัญญาณเสียงปอดต้นฉบับ (Original lung sounds)

การเก็บข้อมูลดำเนินการที่ King Abdullah University Hospital ประเทศจอร์แดน โดยใช้หูฟังแพทย์อิเล็กทรอนิกส์ 3M™ Littmann® Electronic Stethoscope model 3200 เป็นอุปกรณ์หลักในการบันทึกเสียงปอด อุปกรณ์ดังกล่าวสามารถบันทึกเสียงจากร่างกายผู้ป่วยและส่งข้อมูลไปยังคอมพิวเตอร์ผ่าน Bluetooth adaptor จากนั้นผู้วิจัยใช้ซอฟต์แวร์ 3M™ Littmann® StethAssist Visualization Software ในการดึงข้อมูลเสียงจากหน่วยความจำของหูฟังแพทย์และส่งออกไฟล์เสียงในรูปแบบ .wav

ในการบันทึกเสียง แพทย์ผู้เชี่ยวชาญเป็นผู้วางหูฟังแพทย์บนตำแหน่งต่าง ๆ ของผนังทรวงอก โดยตำแหน่งที่ใช้บันทึกเสียงไม่ได้ถูกกำหนดให้เหมือนกันทุกคนแบบตายตัว แต่ขึ้นอยู่กับการประเมินและการตรวจวินิจฉัยของแพทย์ในแต่ละกรณี ตำแหน่งการเก็บเสียงถูกระบุเป็นเขตของทรวงอก ได้แก่ ด้านหน้าและด้านหลังของทรวงอก ด้านซ้ายและด้านขวา รวมถึงบริเวณส่วนบน ส่วนกลาง และส่วนล่างของทรวงอก การระบุตำแหน่งดังกล่าวมีความสำคัญ เนื่องจากเสียงปอดที่บันทึกได้อาจแตกต่างกันตามตำแหน่งของปอดและลักษณะของพยาธิสภาพในผู้ป่วยแต่ละราย

เสียงที่บันทึกจากผู้เข้าร่วมแต่ละรายมีความยาวประมาณ 5–30 วินาที ซึ่งเป็นระยะเวลาที่เพียงพอสำหรับครอบคลุมอย่างน้อยหนึ่งรอบการหายใจ อย่างไรก็ตาม บทความต้นฉบับระบุว่าไม่ได้มีการกำหนดระยะเวลาขั้นต่ำในการบันทึกเสียงอย่างเคร่งครัด เนื่องจากการบันทึกขึ้นอยู่กับการตรวจของแพทย์และข้อจำกัดของอุปกรณ์ โดยระยะเวลาสูงสุดของไฟล์เสียงถูกจำกัดจากความสามารถของหูฟังแพทย์อิเล็กทรอนิกส์ที่ใช้ในการศึกษา

ไฟล์เสียงในชุดข้อมูลถูกส่งออกเป็นไฟล์ .wav ภายใต้ตัวกรองความถี่ 3 รูปแบบ ได้แก่ Bell mode, Diaphragm mode และ Extended mode โดยแต่ละตัวกรองมีคุณสมบัติใน

การเน้นช่วงความถี่ของเสียงแตกต่างกัน Bell mode เน้นเสียงความถี่ต่ำ Diaphragm mode เน้นเสียงในช่วงความถี่ที่เหมาะกับการฟังเสียงปอดมากขึ้น และ Extended mode ครอบคลุมช่วงความถี่ที่กว้างและเน้นเสียงในช่วงกลาง การมีไฟล์เสียงจากตัวกรองทั้งสามรูปแบบช่วยให้สามารถนำข้อมูลไปวิเคราะห์เปรียบเทียบผลของช่วงความถี่ต่อประสิทธิภาพของแบบจำลอง หรือเลือกใช้ตัวกรองที่เหมาะสมกับวัตถุประสงค์ของการวิเคราะห์ได้

ชุดข้อมูลยังประกอบด้วยไฟล์ data annotation.xlsx ซึ่งระบุข้อมูลประกอบของแต่ละตัวอย่าง ได้แก่ อายุ เพศ การวินิจฉัยโรค ชนิดของเสียงปอด และตำแหน่งที่บันทึกเสียงบนทรวงอก ข้อมูลเหล่านี้ถูกใช้เป็นข้อมูลกำกับในการวิเคราะห์และสร้างแบบจำลองจำแนกประเภทเสียงปอดหรือโรคระบบทางเดินหายใจ ทั้งนี้ ข้อมูลที่เผยแพร่ไม่มีข้อมูลที่สามารถระบุตัวตนของผู้เข้าร่วมได้โดยตรง และการเก็บข้อมูลได้รับการอนุมัติจากคณะกรรมการจริยธรรมการวิจัยของ King Abdullah University Hospital และ Jordan University of Science and Technology

3.1.2 การเตรียมสัญญาณและการแบ่งช่วงสัญญาณ

เพื่อเตรียมข้อมูลสำหรับการเรียนรู้ของแบบจำลอง สัญญาณเสียงปอดในแต่ละไฟล์ถูกแบ่งออกเป็นช่วงสัญญาณด้วยหน้าต่างเวลาแบบไม่ทับซ้อน (Non-overlapping window) โดยกำหนดความยาวหน้าต่างเท่ากับ 4 วินาที ค่าดังกล่าวได้มาจากการทดลองของงานวิจัยนี้ ซึ่งมีความสอดคล้องกับแนวทางในงานวิจัยก่อนหน้า ที่ใช้การแบ่งสัญญาณระดับหลายวินาที เพื่อครอบคลุมลักษณะเชิงคลินิกของเสียงปอด (Abdulhameed, M. F., Al-Zubaidi, A. S., Mohammed, M. A., and Al-Hayani, A. A., 2020; Rocha, B. M., Filos, D., Mendes, L., Vogiatzis, I., Perantoni, E.,Kaimakamis, E., Natsiavas, P., Jaume-i-Capó, A., and Paiva, R. P., 2020)

ผลจากการแบ่งช่วงสัญญาณด้วยหน้าต่างความยาว 4 วินาที การเลือกความยาวหน้าต่าง 4 วินาที เป็นค่าประนีประนอมระหว่างการครอบคลุมลักษณะเชิงคลินิกของเสียงปอดในระดับรอบการหายใจ และการควบคุมจำนวนเฟรมให้เหมาะสมต่อการฝึกแบบจำลอง (Chambres et al., 2018; Rangayyan, 2015) ทำให้ได้จำนวนเฟรมแตกต่างกันในแต่ละกลุ่มโรค ดังที่สรุปไว้ในตารางที่ 3.1 โดยกลุ่มภาวะปกติมีจำนวนเฟรมมากที่สุด จำนวน 415 เฟรม ข้อมูลที่ใช้ในการศึกษาได้มาจากแฟ้มข้อมูลสัญญาณเสียงปอดต้นฉบับจำนวน 105 แฟ้ม รองลงมา คือ กลุ่มโรคหอบหืด จำนวน 357 เฟรม จาก 96 แฟ้ม กลุ่มโรคปอดอุดกั้นเรื้อรัง จำนวน 102 เฟรม จาก 27 แฟ้ม และกลุ่มโรคปอดอักเสบ จำนวน 66 เฟรม จาก 15 แฟ้ม ตามลำดับ ความแตกต่างของจำนวนเฟรม

ระหว่างกลุ่มโรคแสดงถึงลักษณะความไม่สมดุลของข้อมูล ซึ่งเป็นปัจจัยสำคัญที่ต้องนำมาพิจารณาในการออกแบบการฝึก และการประเมินประสิทธิภาพของแบบจำลอง

3.1.3 การจัดการข้อมูลไม่สมดุล

ชุดข้อมูลเสียงปอดที่ใช้ในงานวิจัยนี้มีลักษณะเป็นข้อมูลไม่สมดุล (Imbalanced dataset) เนื่องจากจำนวนข้อมูลในแต่ละกลุ่มโรคมีความแตกต่างกัน ทั้งในระดับจำนวนอาสาสมัคร จำนวนแฟ้มข้อมูลเสียง และจำนวนช่วงสัญญาณที่ได้หลังจากการแบ่งข้อมูล โดยงานวิจัยนี้พิจารณาการจำแนกโรคปอดออกเป็น 4 กลุ่ม ได้แก่ ภาวะปกติ โรคหอบหืด โรคปอดอักเสบ และโรคปอดอุดกั้นเรื้อรัง ซึ่งจากฐานข้อมูล KAUH พบว่าจำนวนอาสาสมัครของแต่ละกลุ่มไม่เท่ากัน โดยกลุ่มภาวะปกติและโรคหอบหืดมีจำนวนตัวอย่างมากกว่ากลุ่มโรคปอดอักเสบและโรคปอดอุดกั้นเรื้อรังอย่างชัดเจน ฐานข้อมูลต้นฉบับประกอบด้วยอาสาสมัครกลุ่มปกติ 35 ราย โรคหอบหืด 32 ราย โรคปอดอักเสบ 5 ราย และโรคปอดอุดกั้นเรื้อรัง 9 ราย ซึ่งสะท้อนให้เห็นว่าชุดข้อมูลมีการกระจายตัวของจำนวนตัวอย่างไม่เท่ากันตั้งแต่ระดับแหล่งข้อมูลต้นฉบับ

ภายหลังจากการเตรียมสัญญาณและการแบ่งช่วงสัญญาณเสียง ข้อมูลที่ใช้ในการศึกษายังคงมีจำนวนตัวอย่างในแต่ละกลุ่มแตกต่างกัน โดยกลุ่มโรคหอบหืดมีจำนวน 357 เฟรม จาก 96 แฟ้ม กลุ่มโรคปอดอุดกั้นเรื้อรังมีจำนวน 102 เฟรม จาก 27 แฟ้ม และกลุ่มโรคปอดอักเสบมีจำนวน 66 เฟรม จาก 15 แฟ้ม ความแตกต่างของจำนวนเฟรมระหว่างกลุ่มโรคดังกล่าวถือเป็นปัจจัยสำคัญที่อาจส่งผลต่อกระบวนการเรียนรู้ของแบบจำลอง เนื่องจากแบบจำลองอาจมีแนวโน้มเรียนรู้ลักษณะของกลุ่มที่มีจำนวนข้อมูลมากกว่าได้ดีกว่ากลุ่มที่มีจำนวนข้อมูลน้อยกว่า

ในงานวิจัยนี้ไม่ได้ใช้วิธีการปรับสมดุลข้อมูลโดยตรง เช่น การเพิ่มจำนวนข้อมูลด้วยการสุ่มซ้ำ (Oversampling) การลดจำนวนข้อมูลจากกลุ่มที่มีมากกว่า (Undersampling) หรือการสร้างข้อมูลสังเคราะห์ เนื่องจากข้อมูลเสียงปอดเป็นข้อมูลชีวการแพทย์ที่มีลักษณะเฉพาะของแต่ละบุคคลและแต่ละโรค การเพิ่มหรือลดจำนวนข้อมูลโดยตรงอาจทำให้รูปแบบของสัญญาณผิดไปจากลักษณะข้อมูลจริง หรืออาจทำให้เกิดอคติจากข้อมูลที่ถูกสร้างขึ้นซ้ำ ดังนั้น งานวิจัยนี้จึงเลือกจัดการผลกระทบของข้อมูลไม่สมดุลผ่านการออกแบบกระบวนการทดลอง การแบ่งข้อมูล และการเลือกตัวชี้วัดประเมินผลให้เหมาะสมกับลักษณะของชุดข้อมูล

แนวทางแรก คือ การเลือกใช้ข้อมูลจากฐานข้อมูล KAUH เฉพาะกลุ่มโรคที่มีจำนวนอาสาสมัครเพียงพอต่อการประเมินผลแบบการตรวจสอบไขว้ห้าส่วนโดยแยกตามบุคคล หรือ Subject-independent 5-fold cross-validation ได้แก่ ภาวะปกติ โรคหอบหืด โรคปอดอักเสบ และโรคปอดอุดกั้นเรื้อรัง โดยข้อมูลของอาสาสมัครรายเดียวกันจะถูกจัดให้อยู่ในส่วนเดียวกันเสมอ

ไม่ให้เฟรมเสียงจากบุคคลเดียวกันปรากฏอยู่ทั้งในชุดฝึกและชุดทดสอบ วิธีดังกล่าวช่วยลดปัญหาการรั่วไหลของข้อมูล และทำให้ผลการประเมินสะท้อนความสามารถของแบบจำลองในการจำแนกข้อมูลจากบุคคลใหม่ได้เหมาะสมมากขึ้น

งานวิจัยนี้ใช้การแบ่งข้อมูลแบบการตรวจสอบไขว้แบบห้าส่วน โดยแยกตามบุคคลเพื่อหลีกเลี่ยงการรั่วไหลของข้อมูลระหว่างชุดฝึกและชุดทดสอบ โดยกำหนดให้ข้อมูลเสียงปอดทั้งหมดของอาสาสมัครรายเดียวกันต้องอยู่ในส่วนเดียวกันเสมอ ไม่ว่าจะเป็นเฟรมเสียงหรือช่วงสัญญาณย่อยใด ๆ แนวทางดังกล่าวมีความสำคัญอย่างยิ่งในงานวิเคราะห์เสียงทางการแพทย์ ช่วยให้การประเมินผลแสดงความสามารถของแบบจำลองในการทำให้เป็นสากลต่อผู้ป่วยรายใหม่ได้อย่าง การแบ่งข้อมูลดำเนินการโดยจัดอาสาสมัครออกเป็น 5 ส่วน และหมุนเวียนให้แต่ละส่วนทำหน้าที่เป็นชุดทดสอบ 1 ครั้ง ในขณะที่อีก 4 ส่วน ใช้เป็นชุดฝึกและชุดตรวจสอบในแต่ละรอบการทดลอง การออกแบบในลักษณะนี้ช่วยลดอคติจากการเลือกชุดทดสอบ และเพิ่มความน่าเชื่อถือของผล

Subjects ⟶ [Fold 1] [Fold 2] [Fold 3] [Fold 4] [Fold 5]

Iteration 1: Test = Fold 1, Train = Fold 2–5

Iteration 2: Test = Fold 2, Train = Fold 1,3–5

...

Iteration 5: Test = Fold 5, Train = Fold 1-4

รูปที่ 3.1 การแบ่งข้อมูลแบบการตรวจสอบไขว้แบบห้าส่วน

แนวทางที่สอง คือ การใช้แบบจำลองออโตเอนโคเดอร์แบบหลายงาน (Multi-Task Autoencoder: MTAE) เพื่อเรียนรู้ตัวแทนเชิงแฝงของสัญญาณเสียงปอด โดยแบบจำลองเรียนรู้งานการสร้างข้อมูลกลับคืนร่วมกับงานการจำแนกประเภท ทำให้ตัวแทนข้อมูลที่ได้ยังคงรักษาสารสนเทศสำคัญของสัญญาณเสียงต้นฉบับไว้ ขณะเดียวกันก็เพิ่มความสามารถในการแยกแยะกลุ่มโรค นอกจากนี้ งานวิจัยยังนำตัวแทนเชิงแฝงที่ได้จาก MTAE ไปใช้ร่วมกับซัพพอร์ตเวกเตอร์แมชชีน เพื่อช่วยกำหนดขอบเขตการตัดสินใจระหว่างกลุ่มโรคให้ชัดเจนยิ่งขึ้น ซึ่งสอดคล้องกับแนวทางหลักของวิทยานิพนธ์ที่มุ่งพัฒนาแบบจำลองแบบผสมระหว่างการเรียนรู้เชิงลึกและการเรียนรู้ของเครื่อง

แนวทางที่สาม คือ การประเมินประสิทธิภาพของแบบจำลองโดยไม่พิจารณาเฉพาะค่าความถูกต้องรวม (Accuracy) เพียงอย่างเดียว เนื่องจากในกรณีที่ข้อมูลแต่ละกลุ่มมีจำนวนไม่เท่ากัน ค่าความถูกต้องอาจได้รับอิทธิพลจากกลุ่มข้อมูลที่มีจำนวนมากกว่า และอาจไม่สะท้อนความสามารถของแบบจำลองต่อกลุ่มโรคที่มีข้อมูลน้อยได้อย่างชัดเจน ดังนั้น งานวิจัยนี้จึงรายงานตัวชี้วัดหลายด้าน ได้แก่ ความแม่นยำ (Precision) ความไวหรือรีคอลล์ (Sensitivity/Recall) และค่าคะแนน F1 (F1-score) รวมถึงการวิเคราะห์เมตริกซ์สับสน เพื่อใช้พิจารณาความสามารถของแบบจำลองในแต่ละกลุ่มโรคอย่างครอบคลุม โดยเฉพาะค่าคะแนน F1 ซึ่งสะท้อนความสมดุลระหว่างความแม่นยำ และความไวหรือรีคอลล์ ได้ดีกว่าการใช้ความถูกต้องเพียงค่าเดียวในกรณีที่ข้อมูลมีความไม่สมดุล

ดังนั้น การจัดการข้อมูลไม่สมดุลในงานวิจัยนี้เป็นการจัดการในเชิงระเบียบวิธีการทดลองมากกว่าการปรับจำนวนข้อมูลโดยตรง กล่าวคือ งานวิจัยนี้ควบคุมผลกระทบของข้อมูลไม่สมดุลผ่านการเลือกชุดข้อมูลที่เหมาะสม การแบ่งข้อมูลแบบแยกตามบุคคล การใช้แบบจำลองที่สามารถเรียนรู้ตัวแทนเชิงแฝงของสัญญาณเสียงปอดได้อย่างมีประสิทธิภาพ และการใช้ตัวชี้วัดประเมินผลที่ครอบคลุมหลายมิติ แนวทางดังกล่าวช่วยให้การประเมินประสิทธิภาพของแบบจำลองมีความน่าเชื่อถือมากขึ้น และลดความเอนเอียงของผลลัพธ์ที่อาจเกิดจากจำนวนข้อมูลที่แตกต่างกันระหว่างกลุ่มโรค

3.2 การออกแบบวิธีการที่นำเสนอ

3.2.1 ออโตเอนโคเดอร์แบบหลายงาน

เมื่อผ่านขั้นตอนการเตรียมสัญญาณ สัญญาณเจะถูกนำเข้าสู่กระบวนการสกัดคุณลักษณะ คือ สัมประสิทธิ์เซปสตรัมความถี่เมล ซึ่งเป็นคุณลักษณะที่ออกแบบมาเพื่ออธิบายลักษณะสเปกตรัมระยะสั้น (Short-term spectrum) ของสัญญาณเสียง และแปลงสเกลความถี่ให้สอดคล้องกับการรับรู้ของการได้ยินของมนุษย์ การสร้างคุณลักษณะประกอบด้วย (1) การคำนวณสเปกตรัมกำลังด้วยฟูริเยร์ (Power spectrum calculation using Fourier transform) (2) การทำแผนที่ความถี่เข้าสู่สเกลเมลผ่านชุดตัวกรองเมล (Mel filter bank) (3) การแปลงลอการิทึมเพื่อลดช่วงไดนามิกของพลังงาน และ (4) การแปลงโคไซน์แบบไม่ต่อเนื่องเพื่อให้ได้ค่าสัมประสิทธิ์เซปสตรัมที่มีการรวมตัวของข้อมูลและลดมิติได้อย่างมีประสิทธิภาพ

กำหนดให้การแทนข้อมูลด้วยสัมประสิทธิ์เซปสตรัมความถี่เมล เป็นคุณลักษณะหลักที่ใช้ป้อนเข้าสู่แบบจำลอง มีการกำหนดพารามิเตอร์ ได้แก่ ความยาวเฟรม (Frame length) ทำการ

ทดสอบหลายระดับ การใช้หน้าต่างแฮมมิง (Hamming window) การซ้อนทับของเฟรม 50 เปอร์เซ็นต์ และการคำนวณค่าสัมประสิทธิ์เซปสตรัมความถี่เมลหลัก รวมถึงการเพิ่มอนุพันธ์อันดับหนึ่งและอันดับสองเพื่อแสดงการเปลี่ยนแปลงเชิงเวลา

ตารางที่ 3.2 พารามิเตอร์การสกัดคุณลักษณะด้วยสัมประสิทธิ์เซปสตรัมความถี่เมล

พารามิเตอร์	**ค่าที่กำหนด**
ความยาวช่วงสัญญาณที่ใช้แบ่ง (Segment duration)	4 วินาที (ไม่ซ้อนทับ)
ความถี่สุ่มตัวอย่าง (Sampling frequency)	4 กิโลเฮิรตซ์
ความยาวเฟรม	20, 40, 128, 256, 512, 1000 มิลลิวินาที
หน้าต่างสัญญาณ	หน้าต่างแฮมมิง
การซ้อนทับเฟรม	50% overlap
จำนวนสัมประสิทธิ์เซปสตรัมความถี่เมล	13 มิติ

แนวคิดการเรียนรู้หลายงานในลักษณะการแชร์ตัวแทนเชิงแฝง สอดคล้องกับงานวิจัยดั้งเดิมด้านการเรียนรู้แบบหลายงาน ซึ่งชี้ให้เห็นว่าการเรียนรู้หลายงานสามารถทำหน้าที่เป็นกลไกการควบคุมความซับซ้อนของแบบจำลอง ช่วยลดการเรียนรู้เกิน และเพิ่มความสามารถในการทำให้แบบจำลองสามารถใช้งานได้กับข้อมูลที่ไม่เคยพบมาก่อน (Caruana, 1997)

ในบริบทของข้อมูลเสียงปอด ซึ่งมีความแปรปรวนสูงจากปัจจัยผู้ป่วย สภาพแวดล้อม และวิธีการบันทึก การใช้การเรียนรู้หลายงานจึงมีบทบาทสำคัญในการเพิ่มคุณภาพของตัวแทนเชิงแฝง งานการสร้างข้อมูลกลับคืนช่วยรักษาคุณลักษณะสัญญาณพื้นฐาน ขณะที่งานการจำแนกช่วยเน้นสารสนเทศเชิงคลินิกที่เกี่ยวข้องกับโรค ส่งผลให้ตัวแทนเชิงแฝงที่ได้มีความสมดุลและเหมาะสมต่อการนำไปใช้ในขั้นตอนการจำแนก การออกแบบสถาปัตยกรรมออโตเอนโคเดอร์แบบหลายงานมีพื้นฐานทั้งจากแนวคิดระเบียบวิธีและจากทฤษฎีการเรียนรู้หลายงานที่ได้รับการยอมรับ ช่วยสนับสนุนการใช้แบบจำลองกับการจำแนกเสียงปอดที่มีข้อมูลจำกัดและมีความแปรปรวนสูง

ตารางที่ 3.3 สถาปัตยกรรมของแบบจำลองออโตเอนโคเดอร์แบบหลายงาน

Module	Layer	Kernel	Layer configuration	Activation
Autoencoder (Encoder1) Autoencoder (Encoder2) Autoencoder (Encoder3)	Conv2d 1 Conv2d 2 Conv2d 3	3x3 3x3 3x3	in_channels=1, out_channels=16 in_channels=16, out_channels=8 in_channels=8, out_channels=2	ReLU ReLU ReLU
Autoencoder (Decoder1) Autoencoder (Decoder2) Autoencoder (Decoder3)	ConvTranspose2d 1 ConvTranspose2d 2 ConvTranspose2d 3	3x3 3x3 3x3	in_channels=2, out_channels=8 in_channels=8, out_channels=16 in_channels=16, out_channels=1	ReLU ReLU ReLU and Sigmoid
Classification (Block1) Classification (Block2) Classification (Block3)	FC (Linear 1) FC (Linear 2) FC (Linear 3) Classifier (Linear 4)	- - - -	in_features=[Flatten latent size], out_features=128 in_features=128, out_features=128 in_features=128, out_features=128 in_features=128, out_features=4	ReLU ReLU ReLU Softmax

ในตารางที่ 3.3 ออกแบบให้มีการแบ่งปันตัวแทนเชิงแฝงร่วมกันระหว่างงานการสร้างข้อมูลกลับคืน และงานการจำแนก การออกแบบในลักษณะนี้ช่วยให้ตัวแทนเชิงแฝงรักษาทั้งสารสนเทศเชิงโครงสร้างและสารสนเทศที่เกี่ยวข้องกับการจำแนกโรคไว้พร้อมกัน

มหาวิทยาลัยเทคโนโลยีสุรนารี

3.2.2 ออโตเอนโคเดอร์แบบหลายงานร่วมกับซัพพอร์ตเวกเตอร์แมชชีน

หลังจากได้ตัวแทนเชิงแฝงที่เรียนรู้จากแบบจำลองออโตเอนโค้ดแบบหลายงานในหัวข้อก่อน ตัวแทนดังกล่าวถูกนำมาใช้เป็นอินพุตของกระบวนการจำแนกโรค โดยงานวิจัยนี้เลือกใช้ซัพพอร์ตเวกเตอร์แมชชีน เป็นตัวจำแนกหลักในขั้นตอนสุดท้ายแทนการใช้ชั้นฟังก์ชันซอฟต์แมกซ์ของโครงข่ายประสาทเทียม ทั้งนี้ การเลือกใช้ซัพพอร์ตเวกเตอร์แมชชีนมีเหตุผลเชิงระเบียบวิธีที่สอดคล้องกับลักษณะของข้อมูลเสียงปอด ซึ่งมีมิติของคุณลักษณะที่มีแนวโน้มสูง แต่มีจำนวนตัวอย่างจำกัด และมีความแปรปรวนสูง

หลักการของซัพพอร์ตเวกเตอร์แมชชีน เป็นตัวจำแนกแบบอาศัยระยะขอบ (Margin) ที่มุ่งหาขอบเขตการตัดสินใจที่มีระยะขอบมากที่สุดระหว่างกลุ่มต่าง ๆ ซึ่งช่วยลดความเสี่ยงของการเรียนรู้เกิน และเพิ่มความสามารถในการทำให้แบบจำลองทั่วไปได้ดีกับข้อมูลที่ไม่เคยพบมาก่อน โดยเฉพาะในกรณีที่ข้อมูลมีการกระจายไม่สมดุลหรือมีการซ้อนทับกันของกลุ่มบางส่วน จากลักษณะดังกล่าวทำให้ซัพพอร์ตเวกเตอร์แมชชีนมาใช้อย่างแพร่หลายในงานจำแนกสัญญาณชีวการแพทย์ รวมถึงงานวิเคราะห์เสียงปอดและเสียงทางเดินหายใจ

ในงานวิจัยขั้นตอนการจำแนกขั้นสุดท้ายถูกออกแบบให้ใช้ซัพพอร์ตเวกเตอร์แมชชีนร่วมกับตัวแทนเชิงแฝงหรือคุณลักษณะเชิงลึกที่ได้จากแบบจำลองออโตเอนโคเดอร์แบบหลายงาน เพื่อเพิ่มความเสถียรของขอบเขตการตัดสินใจ โดยเฉพาะเมื่อข้อมูลมีความแปรปรวนสูงและมีจำนวนตัวอย่างจำกัด ซึ่งเป็นลักษณะสำคัญของข้อมูลเสียงปอดในบริบททางคลินิก กระบวนการทำงานโดยรวมดำเนินการในรูปแบบ 2 ขั้นตอน คือ (1) ฝึกออโตเอนโคเดอร์แบบหลายงาน ให้เรียนรู้คุณลักษณะที่มีความหมายจากข้อมูลเสียงปอด และ (2) ใช้ออโตเอนโคเดอร์แบบหลายงานที่ฝึกแล้วเป็นตัวสกัดคุณลักษณะ ก่อนนำคุณลักษณะดังกล่าวเข้าสู่ซัพพอร์ตเวกเตอร์แมชชีนเพื่อทำการจำแนกโรคต่อไป

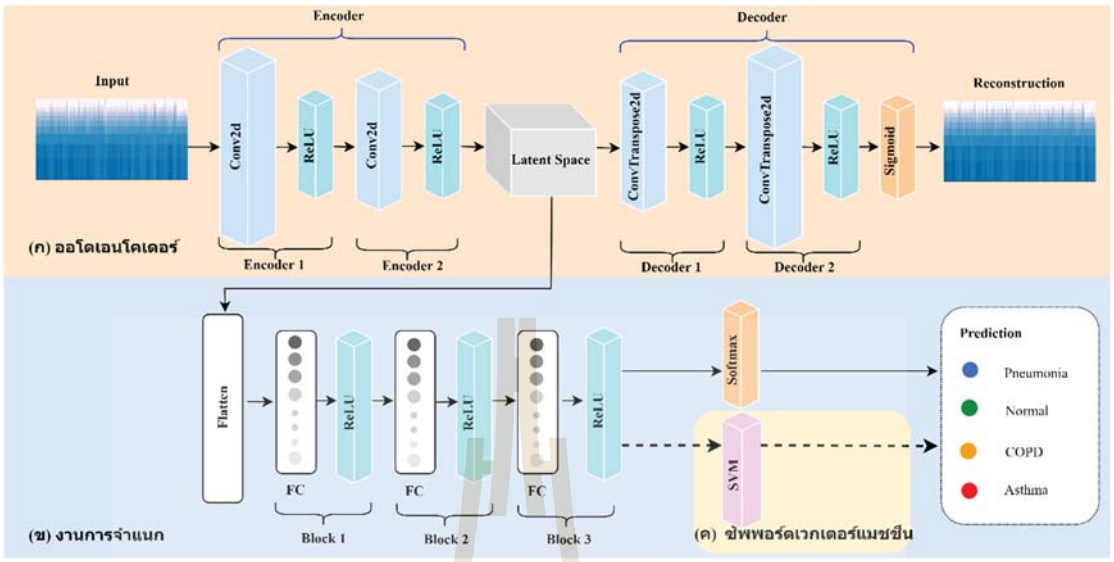
เมื่อทำการฝึกออโตเอนโคเดอร์แบบหลายงาน นำคุณลักษณะจากชั้นเชื่อมโยงสมบูรณ์ชั้นสุดท้ายของออโตเอนโคเดอร์แบบหลายงาน มาใช้เป็นข้อมูลนำเข้าสำหรับซัพพอร์ตเวกเตอร์แมชชีนแทนการใช้ชั้นจำแนกแบบเดิมของโครงข่ายประสาทเทียม จากนั้นในช่วงทดสอบจะใช้ออโตเอนโคเดอร์แบบหลายงานที่ฝึกแล้วสกัดคุณลักษณะจากข้อมูลทดสอบ และให้ซัพพอร์ตเวกเตอร์แมชชีนทำนายกลุ่มโรคจากคุณลักษณะดังกล่าว แนวทางนี้มีนัยเชิงระเบียบวิธีที่สำคัญ คือ แยกบทบาทระหว่างการเรียนรู้ตัวแทน และการสร้างขอบเขตการตัดสินใจ ทำให้การประเมินแสดงความสามารถในการทำให้เป็นสากลของคุณลักษณะที่เรียนรู้ได้ชัดเจน

การตั้งค่าพารามิเตอร์ของซัพพอร์ตเวกเตอร์แมชชีน การจำแนกโดยใช้เคอร์เนลฟังก์ชันเรเดียล ซึ่งพิจารณาพารามิเตอร์สำคัญ 2 ตัว ได้แก่ พารามิเตอร์ควบคุมการแลกเปลี่ยนระหว่างความกว้างของระยะขอบและความผิดพลาดจากการจำแนก และพารามิเตอร์ที่ควบคุมความกว้างของเคอร์เนล คือ พารามิเตอร์แกมมา การเลือกค่าพารามิเตอร์อาศัยการวิเคราะห์บนชุดข้อมูลตรวจสอบ (Validation) ของแบบจำลองที่ฝึกจากชุดฝึก จากทฤษฎีของซัพพอร์ตเวกเตอร์แมชชีนและงานศึกษาพื้นฐานด้านการเรียนรู้ด้วยเคอร์เนล พบว่าค่าพารามิเตอร์ที่ควบคุมความกว้างของเคอร์เนลที่มีค่าสูงจะทำให้ขอบเขตการตัดสินใจมีความซับซ้อนและมีความไวต่อข้อมูลรบกวน ขณะที่ค่าขนาดเล็กจะส่งผลให้ขอบเขตการตัดสินใจมีความเรียบและช่วยเพิ่มความสามารถในการใช้งานทั่วไปของแบบจำลอง โดยเฉพาะในกรณีที่ข้อมูลมีมิติสูงหรือจำนวนตัวอย่างมีจำกัด (Schölkopf, B., and Smola, A. J., 2002; Hastie, T., Tibshirani, R., and Friedman, J., 2009) แนวคิดดังกล่าวสอดคล้องกับหลักการของซัพพอร์ตเวกเตอร์แมชชีน ที่มุ่งเน้นการเพิ่มระยะขอบสูงสุด (maximum margin) เพื่อควบคุมความซับซ้อนของแบบจำลองและลดความเสี่ยงของการเรียนรู้เกิน

นอกจากนี้ งานวิจัยที่ประยุกต์ใช้ซัพพอร์ตเวกเตอร์แมชชีนร่วมกับคุณลักษณะเชิงลึกหรือคุณลักษณะที่ผ่านการเรียนรู้แล้ว ระบุว่าเมื่อคุณลักษณะมีการแยกกลุ่มเชิงโครงสร้างที่ดีอยู่แล้ว การใช้ค่าพารามิเตอร์แกมมา ในช่วงต่ำสามารถช่วยให้ซัพพอร์ตเวกเตอร์แมชชีนทำหน้าที่เป็นกลไกการตัดสินใจที่มีเสถียรภาพ โดยไม่จำเป็นต้องปรับจูนพารามิเตอร์อย่างละเอียด งานของ Niu, X., and Suen, C. Y. (2012) แสดงให้เห็นว่าการใช้ซัพพอร์ตเวกเตอร์แมชชีนบนคุณลักษณะที่สกัดจากโครงข่ายเชิงลึกสามารถให้ผลการจำแนกที่มีความเสถียรโดยอาศัยการกำหนดค่าพารามิเตอร์ในช่วงมาตรฐาน ขณะที่ Huang, G., Song, S., Gupta, J. N. D., and Wu, C. (2014) ชี้ให้เห็นว่าตัวจำแนกแบบมาร์จิน มีประสิทธิภาพสูงเมื่อทำงานบนตัวแทนข้อมูลที่มีโครงสร้างชัดเจนจากกระบวนการเรียนรู้ล่วงหน้า

สถาปัตยกรรมและลำดับขั้นตอนการทำงานของแบบจำลองที่นำเสนอ ซึ่งถูกออกแบบขึ้นเพื่อแก้ไขข้อจำกัดของการเรียนรู้แบบงานเดียวในการวิเคราะห์เสียงปอด โดยผสานแนวคิดการเรียนรู้หลายงานเข้ากับตัวจำแนกแบบระยะขอบ เพื่อยกระดับทั้งความแม่นยำและความเสถียรของการจำแนกโรคระบบทางเดินหายใจ ลำดับขั้นตอนตั้งแต่การรับสัญญาณ การสกัดคุณลักษณะ การเรียนรู้ตัวแทนเชิงแฝง ไปจนถึงการจำแนกโรค ดังรูปที่ 3.2 แบบจำลองออโตเอนโคเดอร์แบบหลายงานร่วมกับซัพพอร์ตเวกเตอร์แมชชีน ซึ่งประกอบด้วย 3 ส่วนหลัก ได้แก่

มหาวิทยาลัยเทคโนโลยีสุรนารี



รูปที่ 3.2 แบบจำลองออโตเอนเดอร์แบบหลายงานร่วมกับซัพพอร์ตเวกเตอร์แมชชีน

ในส่วนที่ (ก) ออโตเอนโคเดอร์ ทำหน้าที่เรียนรู้ตัวแทนเชิงแฝงจากข้อมูลอินพุต โดยส่วนตัวเข้ารหัส ประกอบด้วยบล็อกคอนโวลูชันสองชั้น (Encoder 1 และ Encoder 2) ซึ่งใช้เลเยอร์ Convolution 2D ร่วมกับฟังก์ชันกระตุ้นฟังก์ชันเรกติไฟต์ลิเนียร์ยูนิต เพื่อสกัดคุณลักษณะเชิงลึกและลดมิติของข้อมูล จากนั้นข้อมูลจะถูกแปลงให้อยู่ในรูปของปริภูมิเชิงแฝง ซึ่งเป็นตัวแทนเชิงแฝงที่สรุปสารสนเทศสำคัญของสัญญาณเสียงปอด ส่วนตัวถอดรหัสประกอบด้วยเลเยอร์คอนโวลูชันทรานสโพสแบบสองมิติ(Convolution Transpose 2D) และฟังก์ชันกระตุ้นฟังก์ชันเรกติไฟต์ลิเนียร์ยูนิต เพื่อสร้างสัญญาณกลับคืน ให้ใกล้เคียงกับอินพุตเดิม โดยมีวัตถุประสงค์เพื่อบังคับให้ตัวแทนเชิงแฝงรักษาโครงสร้างของสัญญาณต้นฉบับไว้

ในส่วนที่ (ข) งานการจำแนก ตัวแทนเชิงแฝงจากตัวแทนเชิงแฝง จะถูกนำมาทำการปรับรูปข้อมูลเป็นเวกเตอร์หนึ่งมิติ (Flatten) และป้อนเข้าสู่ชั้นเชื่อมโยงสมบูรณ์ จำนวนหลายบล็อก โดยแต่ละบล็อกใช้ฟังก์ชันกระตุ้นฟังก์ชันเรกติไฟต์ลิเนียร์ยูนิต เพื่อเพิ่มความไม่เชิงเส้นในการเรียนรู้คุณลักษณะ จากนั้นมีการใช้ฟังก์ชันซอฟต์แม็กซ์ เพื่อให้ได้ค่าความน่าจะเป็นของแต่ละคลาสในกรณีการจำแนกหลายกลุ่ม ซึ่งสะท้อนผลการทำนายเบื้องต้นของโครงข่าย

ในส่วนที่ (ค) ส่วนตัวจำแนกขั้นสุดท้ายด้วยซัพพอร์ตเวกเตอร์แมชชี**น** ตัวแทนเชิงแฝงหรือคุณลักษณะเชิงลึกที่ได้จากส่วนการจำแนกจะถูกส่งต่อไปยังตัวจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีน ซึ่งทำหน้าที่เป็นตัวจำแนกขั้นสุดท้าย โดยใช้หลักการระยะขอบสูงสุด เพื่อค้นหาไฮเปอร์เพลน

ที่เหมาะสมในการแยกกลุ่มข้อมูล ส่งผลให้การตัดสินใจมีความคมชัดและเสถียรมากขึ้น โดยเฉพาะในกรณีที่กลุ่มโรคมีลักษณะใกล้เคียงกัน ผลสุดท้าย คือ การจำแนกโรคระบบทางเดินหายใจเป็น 4 กลุ่ม

โดยสรุป สถาปัตยกรรมดังกล่าวสะท้อนแนวคิดการผสมผสานระหว่างการเรียนรู้หลายงานเพื่อยกระดับคุณภาพของตัวแทนเชิงแฝง และการใช้ตัวจำแนกซัพพอร์ตเวกเตอร์แมชชี**น**เพื่อเพิ่มความเสถียรของการตัดสินใจ

3.3 ทรัพยากรที่ใช้ในการวิจัยวิทยานิพนธ์

เครื่องมือที่ใช้ในการพัฒนางานวิจัยนี้ ใช้หน่วยประมวลผลกลาง : AMD Ryzen Threadripper 3970x หน่วยประมวลผลกราฟฟิก : NVIDIA GeForce RTX 3090, หน่วยความจำหลัก : 128 GB หน่วยความจำสำรอง SSD/ HDD : 2TB/6TB ระบบปฏิบัติการ : Ubuntu 20.04 ภาษาที่ใช้ในการพัฒนา : Python 3 และเฟรมเวิร์ก : PyTorch ในการฝึกแบบจำลอง มีการกำหนดค่าพารามิเตอร์การเรียนรู้ที่สำคัญ เช่น จำนวนรอบการฝึก ขนาดกลุ่มข้อมูลย่อย ตัวปรับค่าพารามิเตอร์ และฟังก์ชันกระตุ้น เพื่อให้ผลการทดลองมีความเสถียรและสามารถทำซ้ำได้ รายละเอียดของพารามิเตอร์ที่ใช้ในการฝึกแบบจำลอง ดังแสดงในตารางที่ 3.4

ตารางที่ 3.4 พารามิเตอร์สำคัญที่ใช้ในการฝึกแบบจำลอง

พารามิเตอร์	**ค่าที่กำหนด**
ขนาดแบตช์ (Batch size)	32
จำนวนยุคการฝึก (Epochs)	350
ตัวปรับค่าเหมาะสม (Optimizer)	อดัม (Adam)
ค่าความสูญเสีย (Loss)	ครอสเอนโทรปี
ฟังก์ชันกระตุ้นของชั้นเอาต์พุต (Output activation)	ซอฟต์แมกซ์
ฟังก์ชันกระตุ้นของชั้นซ่อน (Hidden activation)	ฟังก์ชันเรกติไฟต์ลิเนียร์ยูนิต

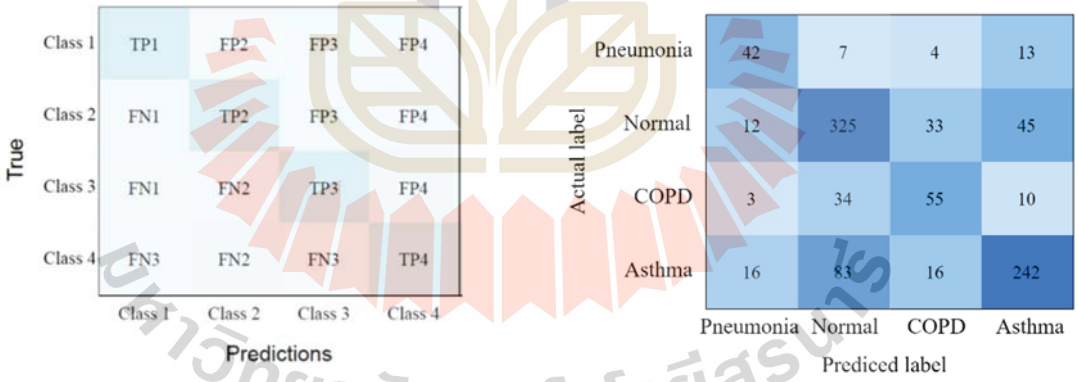
ตารางที่ 3.4 แสดงพารามิเตอร์สำคัญที่ใช้ในการฝึกโครงข่าย ซึ่งทำหน้าที่เรียนรู้ตัวแทนเชิงแฝงของสัญญาณเสียงปอดสำหรับการนำไปใช้ในขั้นตอนการจำแนก โดยกำหนดขนาดแบตช์ 32 เพื่อรักษาสมดุลระหว่างความเสถียรของการเรียนรู้และประสิทธิภาพการคำนวณ จำนวนยุคการฝึก 350 รอบ เพื่อให้สามารถเรียนรู้โครงสร้างเชิงแฝงของข้อมูลได้อย่างเพียงพอ

ในขั้นตอนการฝึกออโตเอนโคเดอร์ ฟังก์ชันความสูญเสียแบบครอสเอนโทรปี ถูกใช้สำหรับงานการจำแนกที่ผสานอยู่ในกรอบการเรียนรู้หลายงาน ขณะที่ฟังก์ชันเรกติไฟต์ลิเนียร์ยูนิตถูกใช้เพื่อเพิ่มความไม่เชิงเส้นของแบบจำลอง ส่วนฟังก์ชันซอฟต์แมกซ์ ถูกใช้ในชั้นเอาต์พุตของงานการจำแนกภายในออโตเอนโคเดอร์เท่านั้น โดยผลลัพธ์จากชั้นเชิงแฝงที่ได้หลังการฝึกจะถูกนำไปใช้เป็นเวกเตอร์คุณลักษณะอินพุตสำหรับซัพพอร์ตเวกเตอร์แมชชีน ซึ่งทำหน้าที่เป็นตัวจำแนกขั้นสุดท้ายของงานแยกจากกระบวนการฝึกโครงข่ายออโตเอนโคเดอร์

3.4 การประเมินประสิทธิภาพ

3.4.1 เมตริกซ์สับสนและตัวชี้วัดการประเมินผลรายกลุ่ม

เมตริกซ์สับสนเป็นเครื่องมือมาตรฐานที่ใช้ในการประเมินประสิทธิภาพของแบบจำลองสำหรับปัญหาการจำแนกประเภท โดยเฉพาะในกรณีการจำแนกหลายกลุ่ม (Multi-class classification) เมตริกซ์ดังกล่าวแสดงความสัมพันธ์ระหว่างป้ายกำกับจริง (True labels) และผลการทำนายของแบบจำลอง (Predicted labels) ในรูปแบบตาราง ซึ่งช่วยให้สามารถวิเคราะห์รูปแบบความถูกต้องและความสับสนระหว่างกลุ่มได้อย่างเป็นระบบ



(ก) พื้นฐานสำหรับการจำแนก 4 กลุ่ม (ข) การจำแนกเสียงปอดออกเป็น 4 กลุ่ม

รูปที่ 3.3 เมตริกซ์สับสนแบบจำแนก 4 กลุ่ม

รูปที่ 3.3 (ก) แสดงโครงสร้างพื้นฐานของเมตริกซ์สับสนสำหรับการจำแนก 4 กลุ่ม โดยค่าในแนวทแยงมุม (Diagonal elements) แทนจำนวนตัวอย่างที่แบบจำลองจำแนกได้ถูกต้องในแต่ละกลุ่ม ซึ่งเรียกว่าผลจริงบวก (True Positive: TP) ของกลุ่มนั้น ซึ่งเป็นข้อมูลจริงที่มีผลเฉลยเป็น

ค่าจริง และแบบจำลองทำนายผลว่าอยู่ในเป็นกลุ่มนั้น ซึ่งถือว่าแบบจำลองทำนายถูกต้องตรงกับข้อมูลจริง ขณะที่ค่านอกแนวทแยงมุมแสดงกรณีการจำแนกผิดพลาด ซึ่งสามารถแบ่งออกเป็นค่าผลบวกเทียม (False Positive: FP) ข้อมูลจริงที่มีผลเฉลยเป็นกลุ่มอื่น แต่แบบจำลองทำนายผลว่าอยู่ในกลุ่มค่าจริงบวก ซึ่งถือว่าแบบจำลองทำนายผิดกลุ่มข้อมูล และผลลบเทียม (False Negative: FN) ข้อมูลจริงที่มีผลเฉลยเป็นคลาสบวก แต่แบบจำลองทำนายผลว่าอยู่ในกลุ่มอื่น ซึ่งถือว่าแบบจำลองทำนายผิดจากข้อมูลจริง ทั้งนี้ การกระจายของค่าเหล่านี้สะท้อนระดับความสับสนระหว่างกลุ่มต่าง ๆ อย่างชัดเจน สำหรับงานวิจัยนี้ จำแนกเสียงปอดออกเป็น 4 กลุ่ม ได้แก่ ภาวะปกติ โรคปอดอักเสบ โรคปอดอุดกั้นเรื้อรัง และโรคหอบหืด โดยรูปที่ 3.3 (ข) แสดงตัวอย่างเมตริกซ์สับสนที่ได้จากผลการทดลอง จากเมตริกซ์สับสนสามารถคำนวณตัวชี้วัดการประเมินผลได้ดังนี้

ความแม่นยำ สำหรับรายกลุ่ม ดังสมการที่ (3.1)

$$\text{Precision}_{id} = \frac{TP_{id}}{TP_{id} + FP_{id}} \tag{3.1}$$

เมื่อ

Precision_{id} คือ ค่าความแม่นยำของแบบจำลองสำหรับกลุ่มที่ id

FP_{id} คือ จำนวนตัวอย่างที่แบบจำลองทำนายว่าเป็นกลุ่มที่ id แต่ในความเป็นจริงเป็นกลุ่มอื่น

ความไว สำหรับรายกลุ่ม ดังสมการที่ (3.2)

$$\text{Recall}_{id} = \frac{TP_{id}}{TP_{id} + FN_{id}} \tag{3.2}$$

เมื่อ

Recall_{id} คือ ค่าความไวของแบบจำลองสำหรับกลุ่มที่ id

FN_{id} คือ จำนวนตัวอย่างของกลุ่มที่ id ที่แบบจำลองไม่สามารถตรวจพบได้และถูกจำแนกเป็นกลุ่มอื่น

มหาวิทยาลัยเทคโนโลยีสุรนารี

ค่าคะแนน F1 สำหรับรายกลุ่ม ดังสมการที่ (3.3)

$$\mathrm{F1}_{id} = 2 * \frac{\mathrm{Precision}_{id} + \mathrm{Recall}_{id}}{\mathrm{Precision}_{id} \times \mathrm{Recall}_{id}} \tag{3.3}$$

เมื่อ

$\mathrm{F1}_{id}$ คือ ค่าคะแนน F1 ของแบบจำลองสำหรับกลุ่มที่ id

ค่าคะแนน F1 เป็นตัวชี้วัดเชิงสมดุลระหว่างความแม่นยำและความไว ซึ่งเหมาะสมสำหรับการประเมินประสิทธิภาพในกรณีที่จำนวนตัวอย่างของแต่ละกลุ่มไม่สมดุล

ความถูกต้องโดยรวม (Overall Accuracy: OA) ดังสมการที่ (3.4)

$$OA = \sum_{C_{id=1}}^{C_{all}} \frac{TP_{id}}{N_{all}} \tag{3.4}$$

เมื่อ

C_{id} คือ ดัชนีของกลุ่ม

C_{all} คือ จำนวนกลุ่มทั้งหมด

TP_{id} คือ จำนวนตัวอย่างของกลุ่มที่ id ที่แบบจำลองจำแนกได้ถูกต้อง

N_{all} คือ จำนวนตัวอย่างทั้งหมด

id คือ ดัชนีของกลุ่มหรือคลาสที่ใช้ในการจำแนก โดย $id = 1, 2, \ldots, C_{all}$

ในงานวิจัยนี้มีการรายงานค่าตัวชี้วัดรายกลุ่ม เพื่อสะท้อนสมรรถนะของแบบจำลองในแต่ละโรคจากเมตริกซ์สับสน โดยสำหรับกลุ่มที่ id ได้คำนวณความไว (Recall_{id}) ซึ่งนิยามเป็นอัตราส่วนระหว่างจำนวนตัวอย่างที่ถูกจำแนกถูกต้องของกลุ่มนั้นต่อจำนวนตัวอย่างทั้งหมดของกลุ่ม ดังแสดงในสมการที่ (3.3) และในงานวิจัยนี้กำหนดสัญลักษณ์การรายงานรายโรคเป็น $\mathrm{ACC}_{\mathrm{Normal}}$, $\mathrm{ACC}_{Pnuemonia}$, ACC_{Asthma} และ ACC_{COPD} สำหรับกลุ่มภาวะปกติ กลุ่มโรคปอดอักเสบ กลุ่มโรคปอดอุดกั้นเรื้อรัง และโรคหอบหืด ตามลำดับ เพื่อให้การนำเสนอผลลัพธ์ในแต่ละกลุ่มมีความชัดเจน

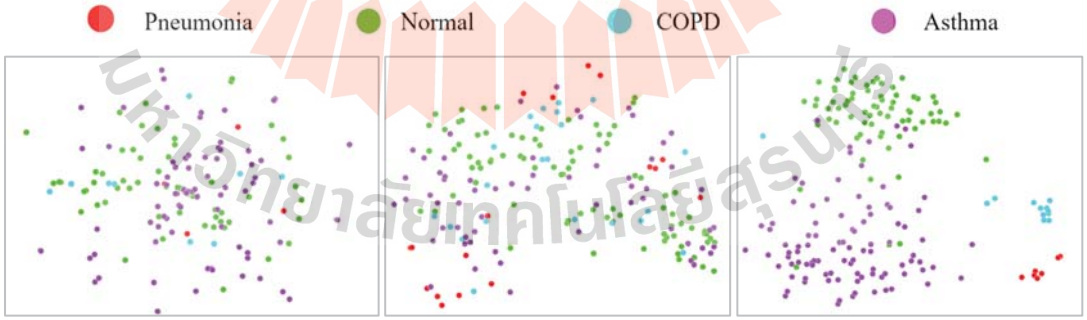
เพื่อให้การประเมินผลมีความครอบคลุมมากยิ่งขึ้น งานวิจัยนี้รายงานทั้งค่าตัวชี้วัดรายกลุ่มและค่าเฉลี่ยรวม โดยให้ความสำคัญกับค่าคะแนน F1 เนื่องจากเป็นตัวชี้วัดที่สะท้อนสมดุล

ระหว่างความแม่นยำ และความไวได้ดีกว่าค่าความถูกต้องเพียงอย่างเดียว โดยเฉพาะในกรณีที่จำนวนตัวอย่างของแต่ละกลุ่มไม่สมดุล นอกจากนี้ เมตริกซ์สับสนยังช่วยสนับสนุนการตีความเชิงคลินิก โดยทำให้เห็นรูปแบบความสับสนระหว่างกลุ่มโรค และเป็นข้อมูลสำคัญสำหรับการปรับปรุงแบบจำลอง (Powers, 2011; Sokolova and Lapalme, 2009)

3.4.2 การวิเคราะห์เชิงภาพด้วยเทคนิค t-SNE

นอกเหนือจากการประเมินประสิทธิภาพของแบบจำลองด้วยตัวชี้วัดเชิงปริมาณและเมตริกซ์สับสนแล้ว งานวิจัยนี้ยังใช้การวิเคราะห์เชิงภาพด้วยเทคนิค t-SNE เพื่อแสดงลักษณะการกระจายตัวของข้อมูลและคุณลักษณะเชิงลึกหลังผ่านกระบวนการสกัดคุณลักษณะหรือการเรียนรู้ตัวแทนเชิงแฝง เทคนิค t-SNE เป็นวิธีการลดมิติข้อมูล ซึ่งสามารถแปลงข้อมูลจากปริภูมิมิติสูงลงสู่ปริภูมิสองมิติได้ โดยยังคงโครงสร้างของข้อมูลไว้ ทำให้สามารถสังเกตแนวโน้มการรวมกลุ่มและการแยกกลุ่มของข้อมูลได้อย่างชัดเจน (van der Maaten, L., and Hinton, G. (2008)

ในงานวิจัยนี้ t-SNE ถูกนำมาใช้เพื่อสนับสนุนการอธิบายความสามารถของแบบจำลองในการจำแนกประเภทของโรค โดยตัวอย่างผลการแสดงภาพ แสดงไว้ในรูปที่ 3.3 ซึ่งเป็นผลลัพธ์จากตัวอย่าง 1 ส่วน ของกระบวนการตรวจสอบความถูกต้องแบบไขว้ห้าส่วน ทำการเปรียบเทียบการกระจายตัวของข้อมูลจาก 3 กรณี ได้แก่ ข้อมูลต้นฉบับ (Raw features) คุณลักษณะที่ได้จากแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชันซึ่งใช้เป็นแบบจำลองอ้างอิง และตัวแทนเชิงแฝงที่ได้จากแบบจำลองออโตเอนโคเดอร์แบบหลายงาน



(ก) ต้นฉบับ (ข) แบบจำลองอ้างอิง (ค) แบบจำลอง MTAE

รูปที่ 3.4 ตัวอย่างการแสดงภาพด้วยเทคนิค t-SNE

จากรูปที่ 3.4 สามารถตีความสัญลักษณ์กลุ่ม ได้ดังนี้ สีแดง หมายถึง โรคปอดอักเสบ (Pneumonia), สีเขียว หมายถึง ปกติ (Normal), สีฟ้า หมายถึง โรคปอดอุดกั้นเรื้อรัง (COPD) และสีบานเย็น หมายถึง โรคหอบหืด (Asthma) ทั้งนี้ ผลการแสดงภาพด้วยเทคนิค t-SNE ถูกใช้เพื่อสนับสนุนข้อสังเกตเชิงคุณภาพควบคู่กับผลการประเมินเชิงปริมาณ และไม่ได้ถูกนำมาใช้เป็นเกณฑ์หลักในการตัดสินประสิทธิภาพของแบบจำลอง

สังเกตได้ว่า การกระจายตัวของข้อมูลในกรณีอินพุตต้นฉบับมีการซ้อนทับกันของกลุ่มค่อนข้างมาก แสดงให้เห็นถึงความยากในการแยกประเภทของโรคโดยอาศัยคุณลักษณะดั้งเดิมเพียงอย่างเดียว ขณะที่กรณีของแบบจำลองอ้างอิง เริ่มแสดงแนวโน้มการรวมกลุ่มของข้อมูลในบางคลาส แต่ยังคงมีความสับสนระหว่างกลุ่มที่มีลักษณะใกล้เคียงกัน ในทางตรงกันข้าม ตัวแทนเชิงแฝงที่ได้จากแบบจำลองออโตเอนโคเดอร์แบบหลายงาน แสดงให้เห็นถึงการแยกกลุ่มของข้อมูลในแต่ละโรคได้ชัดเจนมากขึ้น ซึ่งสะท้อนถึงความสามารถของการเรียนรู้หลายงานในการสร้างตัวแทนข้อมูลที่มีโครงสร้างและความจำแนกได้ดีกว่า

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 4
ผลการวิจัยและอภิปรายผล

ในบทนี้นำเสนอผลจากการดำเนินงานวิจัย ทั้งผลการกำหนดพารามิเตอร์ การสกัดคุณลักษณะ การปรับแต่งพารามิเตอร์การฝึกของแบบจำลอง การออกแบบสถาปัตยกรรมของแบบจำลองการเรียนรู้หลายงานแบบออโตเอนโคเดอร์แบบหลายงาน และการประยุกต์ใช้ตัวจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีน เพื่อเพิ่มความสามารถในการจำแนกของโรคที่เกี่ยวข้องกับระบบทางเดินหายใจจากเสียงปอด ตลอดจนการเปรียบเทียบประสิทธิภาพกับแบบจำลองอ้างอิง

4.1 ผลการดำเนินงานวิจัยในส่วนของการหาความยาวของเสียง

การกำหนดความยาวเฟรม สำหรับการสกัดคุณลักษณะของสัมประสิทธิ์เซปสตรัมความถี่เมลเป็นปัจจัยสำคัญต่อคุณภาพของคุณลักษณะเชิงเวลาและความถี่ ที่ป้อนเข้าแบบจำลองโดยเฉพาะสำหรับสัญญาณเสียงปอดซึ่งมีลักษณะไม่คงที่และมีองค์ประกอบความถี่ที่ซับซ้อน ความยาวเฟรมที่สั้นเกินไปอาจทำให้สัญญาณมีความแปรปรวนสูงและเกิดความไวต่อสัญญาณรบกวน ขณะที่ความยาวเฟรมที่ยาวเกินไปจะลดความสามารถในการเก็บรายละเอียดการเปลี่ยนแปลงเชิงเวลา ส่งผลให้ความสามารถในการจำแนกลดลง

เป็นสิ่งที่สำคัญที่จะระบุความยาวของเฟรมให้เหมาะสมของสัมประสิทธิ์เซปสตรัมความถี่เมลโดยทำการตั้งค่าเบื้องต้น ประกอบด้วยตัวเข้ารหัส 3 ชั้น และตัวถอดรหัส 3 ชั้น และชั้นเชื่อมโยงสมบูรณ์ จำนวน 3 ชั้น อัตราการเรียนรู้ (Learning rate: LR) ถูกตั้งค่าเริ่มต้นเป็น 0.00001 ผลการทดลองแสดงให้เห็นผลกระทบของความยาวของเฟรมสัมประสิทธิ์เซปสตรัมความถี่เมลที่แตกต่างกันว่ามีผลต่อประสิทธิภาพของแบบจำลอง ดังตารางที่ 4.1 การศึกษาของสัมประสิทธิ์เซปสตรัมความถี่เมลกับระยะเวลาของเฟรม 20 มิลลิวินาที ในแบบจำลองแสดงให้เห็นถึงประสิทธิภาพสูงสุดในการจำแนกประเภทของโรคโดยใช้สัญญาณเสียงปอด ระยะเฟรมจึงเป็นสิ่งสำคัญในการศึกษาของสัญญาณเสียงปอด เพราะช่วยในการเก็บรวบรวมข้อมูลที่ถูกต้องของคุณสมบัติเสียง รายละเอียดที่จำเป็นสำหรับการแยกเสียงภาวะปกติกับเสียงที่มีความภาวะผิดปกติ ในทางตรงกันข้ามความยาว

เฟรมที่ยาวกว่ามีแนวโน้มที่จะทำให้คุณลักษณะเฉพาะลดลง ส่งผลให้ความถูกต้องลดลง การทดลองนี้สอดคล้องกับการศึกษาของงานวิจัยที่มีอยู่ก่อนหน้าซึ่งเน้นความสำคัญระยะเฟรมในการศึกษาของสัญญาณเสียงปอด

ตารางที่ 4.1 ผลของความยาวเฟรมของแบบจำลองออโตเอนโคเดอร์แบบหลายงาน

ความยาวเฟรม (มิลลิวินาที)	**ความถูกต้อง (เปอร์เซ็นต์)**	**ความแม่นยำ (เปอร์เซ็นต์)**	**ความไว (เปอร์เซ็นต์)**	**ค่าคะแนน F1 (เปอร์เซ็นต์)**
20	71.28 ± 3.30	74.40 ± 9.21	52.60 ± 6.11	54.60 ± 5.13
40	61.17 ± 5.10	47.60 ± 16.58	41.80 ± 5.50	40.20 ± 6.76
128	56.60 ± 1.71	36.80 ± 11.34	35.40 ± 2.30	32.60 ± 2.70
256	55.75 ± 4.62	33.00 ± 12.02	33.80 ± 2.28	31.20 ± 2.95
512	50.74 ± 3.21	25.40 ± 2.61	30.80 ± 1.48	27.40 ± 1.52
1000	51.38 ± 4.71	25.40 ± 2.51	31.40 ± 2.51	27.80 ± 2.49

หมายเหตุ ± ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: SD)

ผลการทดลองในตารางที่ 4.1 ความยาวเฟรม 20 มิลลิวินาที ให้ประสิทธิภาพสูงสุด เมื่อเทียบกับความยาวเฟรมอื่น ๆ โดยมีค่าความถูกต้องเท่ากับ 71.28 ± 3.30 เปอร์เซ็นต์ และค่าความแม่นยำเท่ากับ 74.40 ± 9.21 เปอร์เซ็นต์ ซึ่งสูงที่สุดในชุดการทดลองนี้ ขณะเดียวกันค่าความไวเท่ากับ 52.60 ± 6.11 เปอร์เซ็นต์ และค่าคะแนน F1 เท่ากับ 54.60 ± 5.13 เปอร์เซ็นต์ แสดงให้เห็นว่าความยาวเฟรม 20 มิลลิวินาที สามารถรักษาสมดุลระหว่างการตรวจจับตัวอย่างภาวะผิดปกติ และความแม่นยำของการทำนายได้ดีที่สุดภายใต้เงื่อนไขการทดลอง

เมื่อเพิ่มความยาวเฟรมเป็น 40 มิลลิวินาที พบว่าประสิทธิภาพลดลงอย่างมีนัยสำคัญ โดยความถูกต้องลดลงเหลือ 61.17 ± 5.10 เปอร์เซ็นต์ และค่าคะแนน F1 ลดลงเหลือ 40.20 ± 6.76 เปอร์เซ็นต์ และเมื่อใช้ความยาวเฟรมที่ยาวมากขึ้น (128 ถึง 1000 มิลลิวินาที) ค่าตัวชี้วัดลดลงต่อเนื่อง เช่น ที่ 512 มิลลิวินาที ค่าความถูกต้องลดลงเหลือ 50.74 ± 3.21 เปอร์เซ็นต์ และค่าคะแนน F1 ลดลงเหลือ 27.40 ± 1.52 เปอร์เซ็นต์ แสดงให้เห็นว่าเฟรมที่ยาวทำให้คุณลักษณะจำเพาะของเสียงปอดถูกเฉลี่ยหรือเจือจาง ส่งผลให้แบบจำลองจำแนกกลุ่มได้ยากขึ้น ดังนั้นเพื่อให้ได้คุณลักษณะ

สัมประสิทธิ์เซปสตรัมความถี่เมลที่เหมาะสมสำหรับการวินิจฉัยโรคระบบทางเดินหายใจจากเสียงปอด งานวิจัยนี้จึงเลือกใช้ความยาวเฟรม 20 มิลลิวินาที เป็นพารามิเตอร์มาตรฐานสำหรับการทดลอง

4.2 ผลการดำเนินงานวิจัยในส่วนของการเลือกอัตราการเรียนรู้

อัตราการเรียนรู้ เป็นพารามิเตอร์หลักที่ควบคุมขนาดการปรับปรุงค่าน้ำหนักในระหว่างการฝึก หากอัตราการเรียนรู้สูงเกินไปอาจทำให้การปรับค่าน้ำหนักข้ามจุดเหมาะสม และไม่สามารถเข้าหาคำตอบที่เสถียรได้ ขณะที่อัตราการเรียนรู้ต่ำเกินไปจะทำให้การฝึกช้าและอาจติดอยู่ในจุดที่ต่ำกว่าค่าที่เหมาะสมที่สุด (Sub-optimal) การศึกษาการใช้อัตราการเรียนรู้หลายค่า และผลลัพธ์ที่ได้นำเสนอในตารางที่ 4.2

ตารางที่ 4.2 ผลการทดลองของอัตราการเรียนรู้

อัตราการเรียนรู้	ความถูกต้อง (เปอร์เซ็นต์)	ความแม่นยำ (เปอร์เซ็นต์)	ความไว (เปอร์เซ็นต์)	ค่าคะแนน F1 (เปอร์เซ็นต์)
0.01	44.15 ± 3.63	11.00 ± 0.71	25.00 ± 0.00	15.40 ± 0.89
0.001	75.96 ± 10.33	68.60 ± 20.50	67.20 ± 13.94	67.00 ± 17.59
0.0001	81.70 ± 3.44	80.20 ± 5.93	77.20 ± 3.56	78.00 ± 4.30
0.00001	71.28 ± 3.03	74.40 ± 9.21	55.60 ± 5.50	59.20 ± 7.33
0.000001	46.17 ± 5.30	18.40 ± 7.40	27.60 ± 3.21	21.40 ± 6.11

หมายเหตุ ± ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: SD)

ผลการทดลองในตารางที่ 4.2 แสดงให้เห็นว่า อัตราการเรียนรู้ เท่ากับ 0.0001 ให้ผลลัพธ์ดีที่สุด โดยมีค่าความถูกต้องเท่ากับ 81.70 ± 3.44 เปอร์เซ็นต์ และค่าความแม่นยำ ค่าความไว และค่าคะแนน F1 เท่ากับ 80.20 ± 5.93 เปอร์เซ็นต์, 77.20 ± 3.56 เปอร์เซ็นต์ และ 78.00 ± 4.30 เปอร์เซ็นต์ ตามลำดับ ซึ่งสะท้อนการเรียนรู้ที่มีเสถียรภาพและให้ผลสมดุลมากที่สุดในทุกตัวชี้วัด เมื่อใช้อัตราการเรียนรู้ เท่ากับ 0.001 แม้ค่าความถูกต้องอยู่ที่ 75.96 ± 10.33 เปอร์เซ็นต์ และค่าคะแนน F1 อยู่ที่ 67.00 ± 17.59 เปอร์เซ็นต์ แต่มีส่วนเบี่ยงเบนมาตรฐานสูงมาก แสดงถึงความไม่เสถียรระหว่างพับกับจำนวนยุคการฝึก ในขณะที่อัตราการเรียนรู้ เท่ากับ 0.01 ให้ผลลัพธ์ต่ำสุดอย่างชัดเจน (ความถูกต้อง 44.15 ± 3.63 เปอร์เซ็นต์, ค่าคะแนน F1 15.40 ± 0.89 เปอร์เซ็นต์) สอดคล้องกับ

ภาวะการปรับค่าเกินจุดเหมาะสม (Over-shooting) และการฝึกที่ไม่เข้าหาอย่างเหมาะสม ส่วนอัตราการเรียนรู้ที่มีค่าต่ำมาก เช่น 0.000001 ทำให้ค่าความถูกต้อง ลดลงเหลือ 46.17 ± 5.30 เปอร์เซ็นต์ และค่าคะแนน F1 21.40 ± 6.11 เปอร์เซ็นต์ สะท้อนว่าแบบจำลองอัปเดตน้ำหนักช้าเกินไปจนไม่สามารถเรียนรู้รูปแบบจำแนกได้อย่างมีประสิทธิภาพ จากผลการวิเคราะห์นี้จึงเลือกใช้ อัตราการเรียนรู้ เท่ากับ 0.0001 สำหรับการฝึกแบบจำลองในการทดลองขั้นตอนต่อไป

4.3 ผลการดำเนินงานวิจัยในส่วนการกำหนดค่าการเข้ารหัสและการถอดรหัส

ผลการดำเนินงานวิจัยในส่วนของการกำหนดสถาปัตยกรรมการเข้ารหัสและการถอดรหัสของแบบจำลองออโตเอนโคเดอร์แบบหลายงาน แสดงให้เห็นว่า จำนวนชั้นของตัวเข้ารหัส และตัวถอดรหัส ส่งผลโดยตรงต่อความสามารถของแบบจำลองในการเรียนรู้ตัวแทนเชิงแฝง และความสามารถในการทำให้คุณลักษณะสามารถใช้ทั่วไปได้กับข้อมูลที่ไม่เคยปรากฏในชุดฝึก หากสถาปัตยกรรมมีความตื้นเกินไป แบบจำลองอาจไม่สามารถจับคุณลักษณะสำคัญของสัญญาณได้อย่างครบถ้วน ในทางกลับกัน หากสถาปัตยกรรมมีความลึกมากเกินไป อาจนำไปสู่ปัญหาการเรียนรู้เกินและลดประสิทธิภาพในการใช้งานทั่วไปของแบบจำลอง

นอกจากนี้ ประสิทธิภาพของแบบจำลองออโตเอนโคเดอร์แบบหลายงานยังได้รับอิทธิพลอย่างมีนัยสำคัญจากการออกแบบสถาปัตยกรรมในส่วนของหัวงานการจำแนก (Classification head) โดยเฉพาะจำนวนชั้นของชั้นเชื่อมโยงสมบูรณ์ ทำหน้าที่แปลงตัวแทนเชิงแฝงไปสู่การตัดสินใจเชิงกลุ่ม การกำหนดจำนวนชั้น และโครงสร้างของส่วนนี้อย่างเหมาะสมมีความสำคัญต่อการรักษาสมดุลระหว่างความสามารถในการแยกกลุ่ม และความเสถียรของแบบจำลอง

เพื่อศึกษาผลกระทบของสถาปัตยกรรมต่อประสิทธิภาพ ทำการทดลองโดยปรับเปลี่ยนจำนวนชั้นและการจัดเรียงของบล็อกการเข้ารหัสและการถอดรหัส รวมถึงจำนวนชั้นของชั้นเชื่อมโยงสมบูรณ์ในส่วนการจำแนก จากนั้นจึงประเมินผลที่ได้โดยอาศัยค่าความถูกต้องของการจำแนกเป็นตัวชี้วัดหลัก ผลการทดลองที่สรุปไว้ในตารางที่ 4.3 แสดงให้เห็นแนวโน้มที่ชัดเจนของผลกระทบของการออกแบบแบบจำลอง

ตารางที่ 4.3 ประสิทธิภาพของการกำหนดชั้นการเข้ารหัสและการถอดรหัส

เข้ารหัส-ถอดรหัส	ความถูกต้อง (เปอร์เซ็นต์)	ความแม่นยำ (เปอร์เซ็นต์)	ความไว (เปอร์เซ็นต์)	ค่าคะแนน F1 (เปอร์เซ็นต์)
1-1	88.51 ± 7.10	91.60 ± 5.13	89.00 ± 5.52	89.80 ± 5.72
2-2	89.47 ± 4.21	92.05 ± 3.15	88.72 ± 6.63	89.96 ± 5.07
3-3	81.70 ± 3.44	80.20 ± 5.93	77.20 ± 3.56	78.00 ± 4.30
4-4	72.87 ± 10.21	69.60 ± 14.84	63.80 ± 12.03	64.60 ± 12.28
5-5	71.17 ± 9.28	66.60 ± 9.66	65.60 ± 13.58	65.40 ± 11.63

หมายเหตุ ± ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: SD)

ในตารางที่ 4.3 แสดงว่าแบบจำลองที่มีโครงสร้างแบบสมมาตร 2-2 ให้สมดุลดีที่สุด โดยมีค่าความถูกต้อง เท่ากับ 89.47 ± 4.21 เปอร์เซ็นต์ และค่าความแม่นยำ ค่าความไว และค่าคะแนน F1 เท่ากับ 92.05 ± 3.15 เปอร์เซ็นต์, 88.72 ± 6.63 เปอร์เซ็นต์ และ 89.96 ± 5.07 เปอร์เซ็นต์ ตามลำดับ ซึ่งเป็นค่าที่สูงที่สุดหรือใกล้เคียงค่าสูงสุดในทุกตัวชี้วัด โครงสร้าง 1-1 ให้ค่าความถูกต้องสูงใกล้เคียง (88.51 ± 7.10 เปอร์เซ็นต์) แต่มีส่วนเบี่ยงเบนมาตรฐานสูงกว่า และค่าเฉลี่ยโดยรวมน้อยกว่า 2-2 แสดงให้เห็นว่า 1-1 ยังไม่เพียงพอสำหรับการเรียนรู้คุณลักษณะที่ซับซ้อนของเสียงปอดในทุกกลุ่ม เมื่อเพิ่มจำนวนชั้นความลึกเป็น 3-3, 4-4 และ 5-5 พบว่าค่าความถูกต้องลดลงอย่างต่อเนื่องจาก 81.70 เปอร์เซ็นต์ เป็น 72.87 เปอร์เซ็นต์ และ 71.17 เปอร์เซ็นต์ ตามลำดับ พร้อมกับค่าคะแนน F1 ลดลงจาก 78.00 เปอร์เซ็นต์ เป็น 64.60 เปอร์เซ็นต์ และ 65.40 เปอร์เซ็นต์ ตามลำดับ แสดงให้เห็นว่า การเพิ่มความซับซ้อนของแบบจำลองไม่ได้ทำให้ดีขึ้นเสมอไป และมีแนวโน้มเข้าสู่ภาวะการเรียนรู้เกิน หรือเพิ่มภาระการเรียนรู้จนทำให้แบบจำลองไม่เสถียร โครงสร้างการเข้ารหัสและการถอดรหัส แบบสมมาตร 2-2 เป็นสถาปัตยกรรมที่เหมาะสม

4.4 ผลการดำเนินงานวิจัยในส่วนของจำนวนชั้นเชื่อมโยงสมบูรณ์

ชั้นเชื่อมโยงสมบูรณ์ ทำหน้าที่แปลงคุณลักษณะเชิงแฝงจากพื้นที่เชิงแฝง (Latent space) ไปสู่การตัดสินใจจำแนกกลุ่ม จำนวนชั้นเชื่อมโยงสมบูรณ์มากเกินไปอาจเพิ่มพารามิเตอร์จนเกิดการเรียนรู้ที่มากเกิน ขณะที่น้อยเกินไปอาจไม่เพียงพอในการสร้างขอบเขตการจำแนกที่ซับซ้อน

ผลการทดลองในตารางที่ 4.4 แสดงว่า ชั้นเชื่อมโยงสมบูรณ์ จำนวน 3 ชั้น ให้ผลดีที่สุด โดยมีค่าความถูกต้อง เท่ากับ 89.47 ± 4.21 เปอร์เซ็นต์ และค่าคะแนน F1 เท่ากับ 89.96 ± 5.07 เปอร์เซ็นต์ ซึ่งเป็นค่าที่สูงที่สุดในชุดนี้ แสดงให้เห็นว่า 3 ชั้น สามารถเพิ่มความสามารถในการสร้างขอบเขตการจำแนกโดยไม่ทำให้แบบจำลองซับซ้อนเกินจำเป็น เมื่อใช้ชั้นเชื่อมโยงสมบูรณ์ จำนวน 2 ชั้น ค่าความถูกต้องอยู่ที่ 87.55 ± 1.50 เปอร์เซ็นต์ ซึ่งน้อยกว่า 3 ชั้น และเมื่อใช้ชั้นเชื่อมโยงสมบูรณ์ จำนวน 1 ชั้น ค่าความถูกต้องลดลงเป็น 85.84 ± 5.79 เปอร์เซ็นต์ สะท้อนว่าโครงสร้างที่ตื้นเกินไปจำกัดความสามารถในการแยกกลุ่ม ในทางตรงกันข้ามเมื่อเพิ่มชั้นเป็น 4 และ 5 ชั้น ค่าความถูกต้องลดลงเป็น 85.70 ± 4.86 เปอร์เซ็นต์ และ 86.38 ± 2.15 เปอร์เซ็นต์ ตามลำดับ โดยค่าคะแนน F1 อยู่ที่ 83.80 เปอร์เซ็นต์ ทั้งสองกรณี แสดงแนวโน้มว่าการเพิ่มชั้นมากเกินไปไม่ก่อให้เกิดประโยชน์เพิ่ม และอาจทำให้แบบจำลองเริ่มเรียนรู้มากเกินไป หรือมีความไม่เสถียร ดังนั้นงานวิจัยจึงเลือกชั้นเชื่อมโยงสมบูรณ์ จำนวน 3 ชั้น เป็นค่าที่เหมาะสมสำหรับแบบจำลอง

ตารางที่ 4.4 ประสิทธิภาพของชั้นเชื่อมโยงสมบูรณ์

ชั้นเชื่อมโยงสมบูรณ์	ความถูกต้อง (เปอร์เซ็นต์)	ความแม่นยำ (เปอร์เซ็นต์)	ความไว (เปอร์เซ็นต์)	ค่าคะแนน F1 (เปอร์เซ็นต์)
1	85.84 ± 5.79	91.20 ± 2.77	84.60 ± 7.50	86.80 ± 5.36
2	87.55 ± 1.50	89.40 ± 2.88	85.80 ± 7.43	86.80 ± 4.44
3	89.47 ± 4.21	92.05 ± 3.15	88.72 ± 6.63	89.96 ± 5.07
4	85.70 ± 4.86	86.80 ± 5.12	84.20 ± 7.89	83.80 ± 9.18
5	86.38 ± 2.15	85.80 ± 3.63	82.20 ± 1.79	83.80 ± 1.48

หมายเหตุ ± ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: SD)

4.5 การเปรียบเทียบแบบจำลอง

ผลการทดลองเปรียบเทียบประสิทธิภาพการจำแนกแบบ 4 กลุ่ม ระหว่างแบบจำลองอ้างอิงโครงข่ายประสาทเทียมแบบคอนโวลูชันสองชั้น กับแบบจำลองออโตเอนโคเดอร์แบบหลายงานซึ่งเป็นวิธีที่นำเสนอ พบว่าแบบจำลองที่นำเสนอให้ผลที่ดีกว่าอย่างชัดเจนในตัวชี้วัด ได้แก่ ความถูกต้อง ความแม่นยำ ความไว และค่าคะแนน F1 ดังตารางที่ 4.5

ตารางที่ 4.5 ผลการทดลองแบบจำลอง

แบบจำลอง	ความถูกต้อง (เปอร์เซ็นต์)	ความแม่นยำ (เปอร์เซ็นต์)	ความไว (เปอร์เซ็นต์)	ค่าคะแนน F1 (เปอร์เซ็นต์)
CNN-2L (Baseline)	70.64 ± 8.66	70.60 ± 10.43	65.40 ± 8.44	65.20 ± 8.70
MTAE	89.47 ± 4.21	92.05 ± 3.15	88.72 ± 6.63	89.96 ± 5.07

หมายเหตุ ± ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: SD)

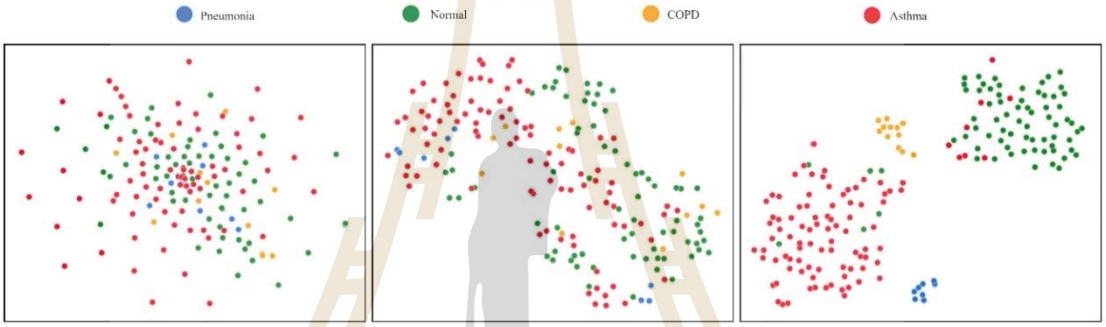
โดยโครงข่ายประสาทเทียมแบบคอนโวลูชันสองชั้น ให้ค่าความถูกต้อง เท่ากับ 70.64 ± 8.66 เปอร์เซ็นต์ และมีค่าความแม่นยำ ค่าความไว และค่าคะแนน F1 เท่ากับ 70.60 ± 10.43 เปอร์เซ็นต์, 65.40 ± 8.44 เปอร์เซ็นต์ และ 65.20 ± 8.70 เปอร์เซ็นต์ ตามลำดับ แสดงว่าแบบจำลองอ้างอิงยังมีข้อจำกัดด้านความสามารถในการจำแนกกลุ่ม โดยเฉพาะเมื่อพิจารณาค่าความไว และค่าคะแนน F1 ที่อยู่ในระดับปานกลาง ซึ่งบ่งชี้ถึงความผิดพลาดในการจำแนกและความสมดุลของการจำแนกที่ยังไม่สูง ในทางตรงกันข้าม แบบจำลองที่นำเสนอ ให้ค่าความถูกต้องสูงถึง 89.47 ± 4.21 เปอร์เซ็นต์ และมีค่าความแม่นยำ ค่าความไว และค่าคะแนน F1 เท่ากับ 92.05 ± 3.15 เปอร์เซ็นต์ 88.72 ± 6.63 เปอร์เซ็นต์ และ 89.96 ± 5.07 เปอร์เซ็นต์ ตามลำดับ แสดงให้เห็นว่าแบบจำลองที่นำเสนอ สามารถเพิ่มทั้งความแม่นยำในการทำนาย และความไวในการตรวจพบกลุ่มได้พร้อมกัน ส่งผลให้ค่าคะแนน F1 สูงขึ้นอย่างมีนัยสำคัญเมื่อเทียบกับแบบจำลองอ้างอิง

นอกจากนี้ค่าส่วนเบี่ยงเบนมาตรฐานของแบบจำลองที่นำเสนอ มีแนวโน้มต่ำกว่าแบบจำลองอ้างอิง ในหลายตัวชี้วัด โดยเฉพาะค่าความถูกต้อง และค่าความแม่นยำ ซึ่งแสดงว่ามีความเสถียรในการทำนายมากกว่าเมื่อประเมินแบบการตรวจสอบไขว้ (Cross-validation) สนับสนุนว่าแนวทางการเรียนรู้ตัวแทนเชิงแฝงแบบหลายงานของแบบจำลองออโตเอนโคเดอร์แบบหลายงาน ช่วยยกระดับประสิทธิภาพการจำแนกโรคระบบทางเดินหายใจจากเสียงปอดได้สูงกว่าแบบจำลองอ้างอิง ทั้งค่าความถูกต้องและความสมดุลของตัวชี้วัดรายกลุ่ม

หลังจากสรุปค่าพารามิเตอร์ที่เหมาะสมกับงานวิจัย กำหนดการทดลองหลักของแบบจำลองออโตเอนโคเดอร์แบบหลายงาน โดยใช้อัตราการเรียนรู้ เท่ากับ 0.0001 โครงสร้างเข้ารหัสและถอดรหัส 2-2 และชั้นเชื่อมโยงสมบูรณ์ จำนวน 3 ชั้น วิเคราะห์ความสามารถในการจำแนกกลุ่มของ

คุณลักษณะด้วย t-SNE เพื่อเปรียบเทียบ (ก) สัญญาณเสียงต้นฉบับ (ข) คุณลักษณะจาก CNN-2L และ (ค) คุณลักษณะจาก MTAE

แสดงการลดมิติของข้อมูลด้วยเทคนิค t-SNE เปรียบเทียบการกระจายตัวของข้อมูล 4 กลุ่ม โดยใช้ตัวอย่างจาก 1 ใน 5 ส่วน การวิเคราะห์ภาพดังกล่าวสามารถตีความได้ในเชิงคุณภาพของตัวแทนคุณลักษณะ กล่าวคือ หากตัวแทนคุณลักษณะสามารถจำแนกกลุ่มได้ดี จะปรากฏเป็นกลุ่มที่มีความหนาแน่นภายในกลุ่มสูง และมีระยะห่างระหว่างกลุ่มมากขึ้น ขณะที่การทับซ้อนกันของกลุ่มสะท้อนความยากในการสร้างขอบเขตจำแนก



(ก) สัญญาณต้นฉบับ (ข) คุณลักษณะของ CNN-2L (ค) คุณลักษณะของ MTAE

รูปที่ 4.1 การกระจายภาพ t-SNE ของคุณลักษณะแบบต่าง ๆ

สัญญาณต้นฉบับ การกระจายตัวของข้อมูลสัญญาณต้นฉบับมีลักษณะปะปนและทับซ้อนกันสูงระหว่างกลุ่ม ซึ่งแสดงว่าในสัญญาณเสียงปอดต้นฉบับมีความแปรปรวนมากจากหลายปัจจัย เช่น ระดับความดัง ความยาวรอบการหายใจ สัญญาณรบกวน และความแตกต่างระหว่างบุคคล ส่งผลให้รูปแบบเชิงโครงสร้างที่จำแนกโรคแต่ละชนิดยังไม่เด่นชัด การที่กลุ่มยังจำแนกไม่ออกในภาพนี้จึงเป็นเหตุผลให้เห็นว่าจำเป็นต้องมีขั้นตอนการสกัดคุณลักษณะหรือการเรียนรู้ตัวแทนที่เหมาะสมก่อนการจำแนก

คุณลักษณะจาก CNN-2L เมื่อพิจารณาคุณลักษณะที่สกัดโดยแบบจำลองอ้างอิง พบว่าเกิดการรวมกลุ่มของข้อมูลมากขึ้นเมื่อเทียบกับสัญญาณต้นฉบับ มีบางกลุ่มที่เริ่มจับตัวเป็นกลุ่ม และมีทิศทางการจำแนกจากกลุ่มอื่นดีขึ้น มีพื้นที่รอยต่อ (Boundary overlap) ซึ่งตีความได้ว่าแบบจำลองอ้างอิงสามารถเรียนรู้คุณลักษณะพื้นฐานของสเปกตรัมได้บางส่วน แต่ยังไม่สามารถดึงคุณลักษณะ

จำเพาะที่จำแนกกลุ่มที่คล้ายกันได้อย่างชัดเจน ผลภาพนี้สอดคล้องกับผลตัวเลขในตารางที่ 4.5 ที่แบบจำลองอ้างอิงให้ค่าความถูกต้องรวมต่ำ ซึ่งบ่งชี้ว่าคุณลักษณะที่ได้จากแบบจำลองอ้างอิงยังมีความสามารถในการจำแนกจำกัดเมื่อเจอกลุ่มข้อมูลที่มีลักษณะเสียงใกล้เคียงกัน

คุณลักษณะของ MTAE ที่สกัดจากแบบจำลองออโตเอนโคเดอร์แบบหลายงาน พบว่ากลุ่มของข้อมูลมีความชัดเจนขึ้นอย่างเห็นได้ชัด โดยแนวโน้มสำคัญ คือ ความกระชับภายในกลุ่มเพิ่มขึ้น (Intra-class compactness) และความแยกจากกันระหว่างกลุ่มมากขึ้น (Inter-class separation) เมื่อเทียบกับรูปที่ 4.1 (ข) ซึ่งแสดงว่าแบบจำลองสามารถเรียนรู้ตัวแทนที่จัดระเบียบเชิงโครงสร้างให้กลุ่มเดียวกันอยู่ใกล้กันมากขึ้น และผลักกลุ่มต่างกันให้ออกห่างกันมากขึ้น

แบบจำลองออโตเอนโคเดอร์แบบหลายงาน มีลักษณะการเรียนรู้ที่ควบคุมให้แบบจำลองสร้างตัวแทนเชิงแฝงที่ดีทั้งต่อการสร้างใหม่ และต่อการจำแนกพร้อมกัน ทำให้ตัวแทนที่เรียนรู้มีความทั่วไปและทนทานต่อความแปรปรวนของสัญญาณมากกว่าแบบจำลองที่เน้นจำแนกเพียงอย่างเดียว ผลภาพจากเทคนิค t-SNE จึงสนับสนุนข้อค้นพบว่าแบบจำลองออโตเอนโคเดอร์แบบหลายงานให้คุณลักษณะที่เหมาะสมต่อการจำแนกโรคทางเดินหายใจจากเสียงปอดได้ดีกว่า

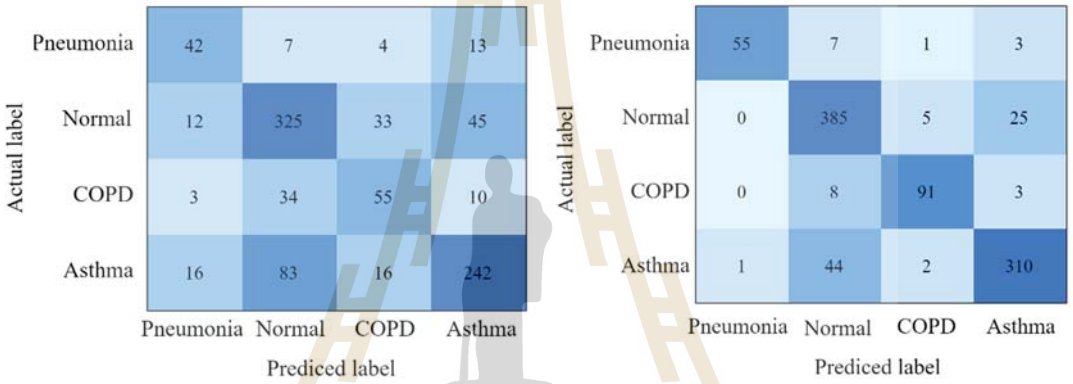
โดยสรุปผลภาพเทคนิค t-SNE ชี้ให้เห็นว่า แบบจำลองออโตเอนโคเดอร์แบบหลายงานสามารถสร้างตัวแทนคุณลักษณะที่มีการรวมกลุ่มภายในกลุ่มมากขึ้นและเพิ่มการแยกจากกันระหว่างกลุ่มได้ดีกว่าแบบจำลองอ้างอิง ส่งผลให้การจำแนกโรคจากเสียงปอดทำได้แม่นยำขึ้น ทั้งนี้ผลเชิงภาพดังกล่าวมีความสอดคล้องกับผลเชิงตัวเลขจากตารางที่ 4.5 และรูปที่ 4.2 เมตริกซ์สับสนที่แสดงการลดความสับสนระหว่างกลุ่มโรค

4.6 ผลการดำเนินงานในส่วนการเรียนรู้หลายงานร่วมกับซัพพอร์ตเวกเตอร์แมชชีน

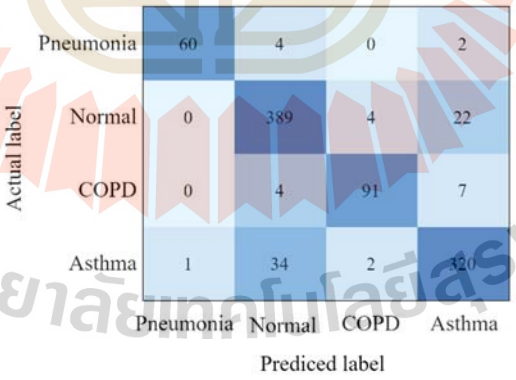
แบบจำลองออโตเอนโคเดอร์แบบหลายงานร่วมกับซัพพอร์ตเวกเตอร์แมชชีนทำหน้าที่เป็นตัวจำแนกขั้นสุดท้ายบนตัวแทนเชิงแฝงที่ได้จากการเรียนรู้หลายงาน พารามิเตอร์สำคัญของซัพพอร์ตเวกเตอร์แมชชีนที่มีผลต่อรูปแบบของขอบเขตการตัดสินใจ คือ พารามิเตอร์แกมมา ซึ่งควบคุมความกว้างของฟังก์ชันเคอร์เนลแบบเรเดียล และส่งผลโดยตรงต่อความซับซ้อนของไฮเปอร์เพลนที่ใช้ในการจำแนกข้อมูล

ด้วยเหตุผลเชิงทฤษฎีและหลักฐานจากงานวิจัยก่อนหน้า งานวิจัยนี้จึงเลือกกำหนดค่าพารามิเตอร์แกมมา เท่ากับ 0.1 สำหรับซัพพอร์ตเวกเตอร์แมชชีน ตามแนวทางที่ได้รับการยอมรับใน

วรรณกรรม เพื่อหลีกเลี่ยงการเพิ่มความซับซ้อนของกระบวนการทดลองและเพื่อให้การประเมินประสิทธิภาพของวิธีการแบบจำลองออโตเอนโคเดอร์แบบหลายงานร่วมกับซัพพอร์ตเวกเตอร์แมชชีนมุ่งเน้นไปที่คุณภาพของตัวแทนเชิงแฝงที่ได้จากการเรียนรู้หลายงานเป็นหลัก เมื่อใช้ค่าพารามิเตอร์แกมมาค่าดังกล่าว ซัพพอร์ตเวกเตอร์แมชชีนสามารถกำหนดขอบเขตการตัดสินใจที่มีความคมชัด โดยอาศัยการจำแนกกลุ่มโครงสร้างของตัวแทนเชิงแฝงจากออโตเอนโคเดอร์แบบหลายงาน ดังที่แสดงในรูปที่ 4.2



(ก) แบบจำลอง CNN-2L (ข) แบบจำลอง MTAE



(ค) แบบจำลอง MTAE-SVM

รูปที่ 4.2 เมตริกซ์สับสนของแบบจำลอง

ตารางที่ 4.6 ผลการประเมินประสิทธิภาพของแบบจำลองในการจำแนกโรคปอด

แบบจำลอง	**ความถูกต้อง (เปอร์เซ็นต์)**	**ความแม่นยำ (เปอร์เซ็นต์)**	**ความไว (เปอร์เซ็นต์)**	**ค่าคะแนน F1 (เปอร์เซ็นต์)**
CNN-2L	70.64 ± 8.66	70.60 ± 10.43	65.40 ± 8.44	65.20 ± 8.70
MTAE	89.47 ± 4.21	92.05 ± 3.15	88.72 ± 6.63	89.96 ± 5.07
MTAE-SVM	91.49 ± 3.03	93.45 ± 2.32	91.22 ± 4.17	92.13 ± 3.14

หมายเหตุ ± ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: SD)

ตารางที่ 4.6 แสดงผลการประเมินประสิทธิภาพของแบบจำลอง ซึ่งรายงานเป็นค่าเฉลี่ยพร้อมส่วนเบี่ยงเบนมาตรฐานจากการทดสอบแบบการตรวจสอบไขว้แบบห้าส่วน ผลการทดลองพบว่าแบบจำลองออโตเอนโคเดอร์แบบหลายงานร่วมกับซัพพอร์ตเวกเตอร์แมชชีนให้ค่าประสิทธิภาพสูงสุดในทุกตัวชี้วัด โดยมีค่าความถูกต้อง 91.49 ± 3.03 เปอร์เซ็นต์ และมีค่าความแปรปรวนอยู่ในระดับต่ำ สะท้อนถึงความถูกต้องและความเสถียรของแบบจำลองในการจำแนกข้อมูลหลายกลุ่มโรค

ผลดังกล่าวแสดงให้เห็นว่าการผสานตัวแทนเชิงแฝงจากโครงสร้างแบบจำลองออโตเอนโคเดอร์แบบหลายงานเข้ากับตัวจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีนช่วยเพิ่มความสามารถในการจำแนกข้อมูลระหว่างกลุ่มได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม การประเมินผลในภาพรวมยังไม่สามารถแสดงสมรรถนะของแบบจำลองในแต่ละกลุ่มโรคได้อย่างครบถ้วน ดังนั้น ตารางที่ 4.7 จึงนำเสนอผลการเปรียบเทียบประสิทธิภาพในระดับรายกลุ่มโรค เพื่อให้เห็นพฤติกรรมของแบบจำลอง

ตารางที่ 4.7 การเปรียบเทียบผลการจำแนกโรคปอดรายกลุ่มโรค

วิธีการสกัดคุณลักษณะและตัวจำแนก	ACC_{Normal} **(เปอร์เซ็นต์)**	$ACC_{Pnuemonia}$ **(เปอร์เซ็นต์)**	ACC_{Asthma} **(เปอร์เซ็นต์)**	ACC_{COPD} **(เปอร์เซ็นต์)**	OA **(เปอร์เซ็นต์)**
MFCC and CNN-2L	78.50	62.50	67.40	52.44	70.64
MFCC and MTAE	92.77	83.33	86.83	89.22	89.47
MFCC and MTAE-SVM	93.73	90.91	89.64	89.22	91.49

มหาวิทยาลัยเทคโนโลยีสุรนารี

จากตารางที่ 4.7 แสดงผลการเปรียบเทียบประสิทธิภาพของวิธีการสกัดคุณลักษณะ และตัวจำแนกในการจำแนกโรคปอดแบบแยกรายกลุ่ม โดยพิจารณาค่าความถูกต้องในการจำแนกของแต่ละกลุ่มโรค ได้แก่ กลุ่มภาวะปกติ (ACC_{Normal}) กลุ่มโรคปอดอักเสบ ($ACC_{Pnuemonia}$) กลุ่มโรคหอบหืด (ACC_{Asthma}) และกลุ่มโรคปอดอุดกั้นเรื้อรัง (ACC_{COPD}) รวมถึงค่าความถูกต้องรวม (Overall Accuracy: OA) จากผลการทดลองพบว่าวิธีแบบจำลองออโตเอนโคเดอร์แบบหลายงานเข้ากับตัวจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องรวมสูงสุดที่ 91.49 เปอร์เซ็นต์ และให้ผลการจำแนกที่ดีในทุกกลุ่มโรคเมื่อเทียบกับวิธีอื่น

เมื่อพิจารณาในระดับรายกลุ่มโรค พบว่าแบบจำลองออโตเอนโคเดอร์แบบหลายงานเข้ากับตัวจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีนสามารถเพิ่มความถูกต้องในการจำแนกกลุ่มปอดอักเสบและกลุ่มหอบหืดได้อย่างชัดเจน โดยค่าความถูกต้องของกลุ่มโรคปอดอักเสบเพิ่มจาก 83.33 เปอร์เซ็นต์เป็น 90.91 เปอร์เซ็นต์ และกลุ่มโรคหอบหืดเพิ่มจาก 86.83 เปอร์เซ็นต์เป็น 89.64 เปอร์เซ็นต์ ขณะที่ค่าความถูกต้องรวมเพิ่มขึ้นจากแบบจำลองออโตเอนโคเดอร์แบบหลายงาน 89.47 เปอร์เซ็นต์ เป็น 91.49 เปอร์เซ็นต์ ผลลัพธ์ดังกล่าวสะท้อนให้เห็นว่าการใช้ซัพพอร์ตเวกเตอร์แมชชีนช่วยปรับขอบเขตการตัดสินใจหมาะสมยิ่งขึ้นบนคุณลักษณะที่เรียนรู้จากแบบจำลองออโตเอนโคเดอร์แบบหลายงาน ส่งผลให้แบบจำลองสามารถแยกแยะโรคปอดแต่ละประเภทได้อย่างมีประสิทธิภาพมากขึ้น โดยเฉพาะในกลุ่มโรคที่มีลักษณะสัญญาณใกล้เคียงกัน

4.7 ผลการเปรียบเทียบกับแนวทางที่มีอยู่เดิมตามงานวิจัยอ้างอิง

หัวข้อนี้นำเสนอการเปรียบเทียบวิธีที่นำเสนอทั้งแบบสัมประสิทธิ์เซปสตรัมความถี่เมลร่วมกับแบบจำลองออโตเอนโคเดอร์แบบหลายงาน และสัมประสิทธิ์เซปสตรัมความถี่เมลร่วมกับแบบจำลองออโตเอนโคเดอร์แบบหลายงานผสานตัวจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีนกับแนวทางที่มีอยู่เดิม ประเมินการจำแนก 4 กลุ่ม แสดงค่าความถูกต้องรายกลุ่มและความถูกต้องรวม เพื่อให้เห็นความสามารถในการแข่งขันของวิธีที่นำเสนอเมื่อเทียบกับงานวิจัยก่อนหน้า

ในตารางที่ 4.8 แนวทางที่นำเสนอให้ค่าความถูกต้องรวมสูงกว่าแนวทางที่มีอยู่เดิม แบบจำลองออโตเอนโคเดอร์แบบหลายงานที่มีกระบวนการสกัดคุณลักษณะสัมประสิทธิ์เซปสตรัมความถี่เมลมีค่าความถูกต้องรวม เท่ากับ 89.47 เปอร์เซ็นต์ และแบบจำลองออโตเอนโคเดอร์แบบหลายงานร่วมกับตัวจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีนที่มีกระบวนการสกัดคุณลักษณะสัมประสิทธิ์เซปสตรัมความถี่เมลมีค่าความถูกต้องรวม เท่ากับ 91.49 เปอร์เซ็นต์ ซึ่งสูงกว่า โครงข่าย

ประสาทเทียมแบบคอนโวลูชันสองชั้นที่มีกระบวนการสกัดคุณลักษณะสัมประสิทธิ์เซปสตรัมความถี่เมลมีค่าความถูกต้องรวม (Aykanat et al.,2017) เท่ากับ 70.64 เปอร์เซ็นต์ และสูงกว่าการแปลงเวฟเลตเชิงประจักษ์ (Empirical Wavelet Transform: EWT) ร่วมกับการประมวลผลด้วยชุดตัวกรอง (Filter Bank-based Processing: FBP) และการจำแนกด้วยไลต์กราเดียนต์บูสติ้งแมชชีน (Light Gradient Boosting Machine: LGBM) (Tripathy et al., 2022) ที่มีค่าความถูกต้องรวมเท่ากับ 77.13 เปอร์เซ็นต์ สะท้อนว่าแบบจำลองออโตเอนโคเดอร์แบบหลายงานสามารถสกัดคุณลักษณะที่เหมาะสมต่อการจำแนกโรคได้ดีกว่าวิธีแบบดั้งเดิม

ตารางที่ 4.8 การเปรียบเทียบผลการจำแนก 4 กลุ่มกับแนวทางที่มีอยู่เดิม

วิธีสกัดคุณลักษณะและตัวจำแนก	ACC_{Normal} **(เปอร์เซ็นต์)**	$ACC_{Pnuemonia}$ **(เปอร์เซ็นต์)**	ACC_{Asthma} **(เปอร์เซ็นต์)**	ACC_{COPD} **(เปอร์เซ็นต์)**	OA **(เปอร์เซ็นต์)**
MFCC+CNN-2L	78.50	62.50	67.40	52.44	70.64
EWT-FBP + LGBM	–	–	–	–	77.13
MFCC + MTAE	92.77	83.33	86.83	89.22	89.47
MFCC + MTAE-SVM	93.73	90.91	89.64	89.22	91.49

เมื่อเปรียบเทียบระหว่างแบบจำลองออโตเอนโคเดอร์แบบหลายงานกับแบบจำลองออโตเอนโคเดอร์แบบหลายงานร่วมกับตัวจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีนพบว่า ค่าความถูกต้องรวมเพิ่มจาก 89.47 เปอร์เซ็นต์ เป็น 91.49 เปอร์เซ็นต์ (เพิ่มขึ้น 2.02 เปอร์เซ็นต์) ซึ่งสอดคล้องกับคำอธิบายในหัวข้อก่อนหน้าว่าการใช้ตัวจำแนกซัพพอร์ตเวกเตอร์แมชชีน ช่วยเพิ่มความไวและความสามารถในการตัดสินใจบนคุณลักษณะเชิงแฝง โดยการเพิ่มระยะห่างระหว่างคุณลักษณะกับไฮเปอร์เพลนของซัพพอร์ตเวกเตอร์แมชชีน และให้ผลดีขึ้นกว่าการใช้ตัวจำแนกของแบบจำลองออโตเอนโคเดอร์แบบหลายงานเพียงอย่างเดียว (Pathonsuwan, W., Phapatanaburi, K., Buayai, P., Jumphoo, T., Anchuen, P., Uthansakul, M., and Uthansakul, P., 2022)

ในเชิงรายกลุ่มจุดที่เพิ่มเด่นที่สุด คือ กลุ่มโรคปอดอักเสบ เพิ่มจาก 83.33 เปอร์เซ็นต์ (แบบจำลองออโตเอนโคเดอร์แบบหลายงาน) เป็น 90.91 เปอร์เซ็นต์ (แบบจำลองออโตเอนโคเดอร์

แบบหลายงานร่วมกับตัวจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีน) หรือเพิ่ม 7.58 เปอร์เซ็นต์ แสดงให้เห็นว่าตัวจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีน มีบทบาทสำคัญต่อการลดความสับสนในกลุ่มที่ทับซ้อนกับกลุ่มอื่น ขณะที่กลุ่มโรคปอดอุดกั้นเรื้อรัง มีค่าคงที่ 89.22 เปอร์เซ็นต์ ทั้งสองวิธี แสดงว่าความยากในการจำแนกบางกลุ่มอาจยังขึ้นกับลักษณะสัญญาณของโรคนั้น ๆ และควรพิจารณาร่วมกับเมตริกซ์สับสนเพื่ออธิบายคู่กลุ่มที่สับสน

มหาวิทยาลัยเทคโนโลยีสุรนารี

บทที่ 5
สรุปและข้อเสนอแนะ

นำเสนอการสรุปผลการวิจัย และข้อเสนอแนะสำหรับการพัฒนางานในอนาคต ซึ่งแสดงให้เห็นว่าแนวทางที่นำเสนอสามารถตรวจจับโรคปอดจากสัญญาณเสียงปอดได้อย่างมีประสิทธิภาพ ผลการทดลองดังกล่าวสนับสนุนแนวคิดการเรียนรู้หลายงานผ่านโครงสร้างออโตเอนโคเดอร์แบบหลายงานที่ช่วยคงไว้ซึ่งสารสนเทศเชิงโครงสร้างของสัญญาณ และสร้างตัวแทนเชิงแฝงที่มีประสิทธิภาพในการจำแนกกลุ่ม ขณะที่แนวทางไฮบริดร่วมกับซัพพอร์ตเวกเตอร์แมชชีน ช่วยเสริมความเหมาะสมของขอบเขตการตัดสินใจและเพิ่มความแข็งแรงของแบบจำลอง ภายใต้สภาวะข้อมูลที่มีความไม่สมดุลและมีการซ้อนทับของลักษณะสัญญาณระหว่างกลุ่มโรค

5.1 สรุปผลการวิจัย

งานวิจัยนี้นำเสนอกรอบวิธีการใหม่สำหรับการตรวจจับและจำแนกโรคระบบทางเดินหายใจจากสัญญาณเสียงปอด โดยมุ่งเพิ่มประสิทธิภาพการแทนข้อมูลของสัญญาณที่มีความแปรปรวนสูงผ่านการเรียนรู้ตัวแทนเชิงแฝงด้วยแบบจำลองออโตเอนโคเดอร์แบบหลายงาน และยกระดับความชัดเจนของขอบเขตการตัดสินใจด้วยแนวทางแบบไฮบริดที่ผสานออโตเอนโคเดอร์แบบหลายงานเข้ากับซัพพอร์ตเวกเตอร์แมชชีน เพื่อให้เหมาะต่อการประยุกต์ใช้ในระบบช่วยคัดกรองและการแพทย์ทางไกล

การประเมินประสิทธิภาพของวิธีการที่นำเสนอโดยใช้เทคนิคการตรวจสอบความถูกต้องแบบไขว้ 5 ส่วน บนชุดข้อมูลเสียงปอดจาก KAUH สำหรับการจำแนก 4 กลุ่ม ได้แก่ ภาวะปกติ โรคปอดอักเสบ โรคหอบหืด และโรคปอดอุดกั้นเรื้อรัง พบว่าแบบจำลองออโตเอนโคเดอร์แบบหลายงานให้ค่าความถูกต้องรวมเท่ากับ 89.47 ± 4.21 เปอร์เซ็นต์ และค่าคะแนน F1 เท่ากับ 89.96 ± 5.07 เปอร์เซ็นต์ ขณะที่แนวทางไฮบริดออโตเอนโคเดอร์แบบหลายงานร่วมกับตัวจำแนกซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องรวมสูงสุด 91.49 ± 3.03 เปอร์เซ็นต์ และค่าคะแนน F1 เท่ากับ 92.13 ± 3.14 เปอร์เซ็นต์ แสดงว่าการผสานซัพพอร์ตเวกเตอร์แมชชีนช่วยเพิ่มความคมชัดและความเสถียรของขอบเขตการตัดสินใจบนคุณลักษณะเชิงแฝง ส่งผลให้การจำแนกโดยรวมมีความถูกต้องและความ

สมดุลดีขึ้นอย่างมีนัยสำคัญโดยสรุป วิธีการที่นำเสนอสามารถเรียนรู้คุณลักษณะเชิงลึกจากเสียงปอดได้อย่างมีประสิทธิภาพ และการต่อยอดเป็นวิธีการแบบไฮบริดด้วยออโตเอนโคเดอร์แบบหลายงานร่วมกับตัวจำแนกซัพพอร์ตเวกเตอร์แมชชีนช่วยยกระดับสมรรถนะให้สูงขึ้นทั้งในมิติความถูกต้องรวมและความสมดุลของผลการจำแนก จึงมีศักยภาพต่อการนำไปประยุกต์ใช้งานจริงเพื่อสนับสนุนการตรวจวินิจฉัยโรคระบบทางเดินหายใจ

5.2 อภิปรายผลการวิจัย

ผลการทดลองบ่งชี้ว่าแนวทางที่นำเสนอให้สมรรถนะเหนือกว่าแบบจำลองอ้างอิงอย่างเป็นระบบ ดังนี้ คือ คุณภาพของตัวแทนเชิงแฝงที่เกิดจากการเรียนรู้หลายงาน และประสิทธิภาพของกลไกการตัดสินใจเมื่อเสริมด้วยตัวจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีน ทั้งสองส่วนทำงานเสริมกันในลักษณะเรียนรู้ตัวแทนให้เหมาะสม และกำหนดขอบเขตการจำแนกให้ชัดเจน

การเรียนรู้หลายงานในโครงสร้างออโตเอนโคเดอร์ทำหน้าที่เป็นข้อจำกัดเชิงโครงสร้างที่กำกับให้ตัวเข้ารหัสสร้างคุณลักษณะเชิงแฝงที่รองรับหลายวัตถุประสงค์พร้อมกัน ส่งผลให้ตัวแทนดังกล่าวมีแนวโน้มลดองค์ประกอบที่ไม่สัมพันธ์กับการจำแนก (เช่น สัญญาณรบกวนหรือความแตกต่างเฉพาะบุคคล) และคงไว้ซึ่งสารสนเทศที่แสดงความแตกต่างทางพยาธิสภาพ จึงช่วยให้พื้นที่คุณลักษณะมีความกระชับ และจำแนกกลุ่มได้ดีขึ้น ลดความเสี่ยงของการเรียนรู้แบบยึดติดข้อมูลฝึกและเอื้อต่อการใช้งานกับข้อมูลใหม่

การผสานซัพพอร์ตเวกเตอร์แมชชีนกับคุณลักษณะเชิงแฝงช่วยเพิ่มความเข้มแข็งของขอบเขตการตัดสินใจผ่านหลักการเพิ่มระยะห่างระหว่างกลุ่ม ซึ่งมีความเหมาะสมกับข้อมูลเสียงปอดที่มีความแปรผันสูง และอาจเกิดการทับซ้อนของรูปแบบเสียงระหว่างกลุ่มโรคบางส่วน การแยกขั้นตอนการเรียนรู้ตัวแทนออกจากขั้นตอนการจำแนกจึงช่วยลดความคลุมเครือในการตัดสินใจ ทำให้ผลการจำแนกมีความเสถียรขึ้นเมื่อเผชิญตัวอย่างที่ยากหรืออยู่ใกล้ขอบเขตระหว่างกลุ่ม

การประยุกต์ใช้ ตัวชี้วัดสำคัญไม่ได้จำกัดอยู่ที่ความถูกต้องรวม แต่ยังรวมถึงความสามารถในการตรวจพบกลุ่มโรคและความสมดุลของการจำแนก เนื่องจากการพลาดตรวจพบผู้ป่วยอาจส่งผลกระทบสูงในบริบทการคัดกรองหรือการตรวจวินิจฉัย ดังนั้นการที่แบบจำลองสามารถยกระดับสมรรถนะโดยรวมควบคู่กับแนวโน้มที่ดีของความไว และค่าคะแนน F1 จึงบ่งชี้ว่าแนวทางที่นำเสนอมีความสามารถในการลดข้อผิดพลาดในตรวจพบ พร้อมกับรักษาสมดุลระหว่างการตรวจพบและการทำนายผิดได้เหมาะสมมากขึ้น

นอกจากนี้ การประเมินผลด้วยการตรวจสอบไขว้แบบห้าส่วน ช่วยให้การตีความผลการทดลองมีความใกล้เคียงกับสถานการณ์ใช้งานจริง กล่าวคือ แบบจำลองต้องทำงานกับผู้ป่วยรายใหม่ที่ไม่เคยพบมาก่อน จึงสนับสนุนว่าประสิทธิภาพที่เพิ่มขึ้นมีนัยต่อความสามารถในการทำให้เป็นสากลมากกว่าการจดจำรูปแบบเฉพาะบุคคล

5.3 ข้อจำกัดของงานวิจัย

แม้ว่าวิธีที่นำเสนอจะแสดงสมรรถนะที่ดีภายใต้เงื่อนไขการทดลองของงานวิจัยนี้ แต่การตีความผลลัพธ์และการนำไปประยุกต์ใช้ควรพิจารณาภายใต้ขอบเขตของการศึกษา เพื่อให้สอดคล้องกับหลักความน่าเชื่อถือเชิงวิชาการ

5.3.1 ด้านชุดข้อมูล ผลการทดลองแสดงให้เห็นถึงประสิทธิภาพของแบบจำลองภายใต้ชุดข้อมูลและบริบทการบันทึกที่ใช้ในการศึกษา ควรทดสอบกับข้อมูลจากแหล่งอื่นเพิ่มเติมเพื่อลดความเสี่ยงของการเรียนรู้แบบยึดติดข้อมูล และเอื้อต่อการใช้งานกับข้อมูลใหม่

5.3.2 การแทนสัญญาณด้วยการใช้สัมประสิทธิ์เซปสตรัมความถี่เมลเป็นตัวแทนข้อมูลหลัก อาจไม่ครอบคลุมในทุกมิติ อาจจะพัฒนาต่อยอดด้วยการใช้ตัวแทนข้อมูลรูปแบบอื่น ๆ

5.3.3 ด้านสภาพแวดล้อมการประเมิน การประเมินผลในสภาพแวดล้อมแบบควบคุม ทำให้ผลลัพธ์ แตกต่างจากการใช้งานจริง จึงควรศึกษาความทนทานของระบบในบริบทภาคสนามต่อไป

5.4 ข้อเสนอแนะและแนวทางการพัฒนาในอนาคต

เพื่อยกระดับความน่าเชื่อถือและเพิ่มศักยภาพการประยุกต์ใช้ของแนวทางที่นำเสนองานวิจัยในอนาคตควรพิจารณาประเด็นต่อไปนี้

5.4.1 ทดสอบแบบจำลองกับชุดข้อมูลจากแหล่งอื่นที่มีความหลากหลายมากขึ้น ทั้งด้านผู้ป่วย อุปกรณ์ และสภาพแวดล้อมการบันทึก เพื่อประเมินความคงทนของสมรรถนะและลดอิทธิพลจากบริบทของชุดข้อมูลเดิม

5.4.2 ศึกษาและเปรียบเทียบตัวแทนข้อมูลอื่นเพิ่มเติม หรือผสานหลายคุณลักษณะร่วมกัน เพื่อเพิ่มความครอบคลุมของสารสนเทศที่เกี่ยวข้องกับรูปแบบเสียงปอด และตรวจสอบผลต่อประสิทธิภาพอย่างเป็นระบบ

5.4.3 ศึกษาความทนทานต่อเสียงรบกวน ความแปรผันจากการวางอุปกรณ์ และข้อกำหนดเชิงระบบ เพื่อรองรับการพัฒนาไปสู่การใช้งานแบบเรียลไทม์หรือภาคสนาม

รายการอ้างอิง

โรงพยาบาลบำรุงราษฎร์. (2567). ภาวะการเจ็บป่วยและการรักษา. สืบค้นเมื่อ 2 กุมภาพันธ์ 2567, จาก https://www.bumrungrat.com/th/conditions-treatments

Abdulhameed, M. F., Al-Zubaidi, A. S., Mohammed, M. A., and Al-Hayani, A. A. (2020). A deep learning approach for respiratory sound classification. *Biomedical Signal Processing and Control*, *61*, 102035. https://doi.org/10.1016/j.bspc.2020.102035

Autthasan, P., Duan, R., Zhang, Z., Li, H., Guan, C., and Wilaiprasitporn, T. (2021). MIN2Net: End-to-end multi-task learning for subject-independent motor imagery EEG classification. *IEEE Transactions on Biomedical Engineering*, *68*(7), 2106–2116. https://doi.org/10.1109/TBME.2020.3045381

Aykanat, M., Kılıç, Ö., Kurt, B., and Saryal, S. (2017). Classification of lung sounds using convolutional neural networks. *EURASIP Journal on Image and Video Processing*, *2017*, 1–9. https://doi.org/10.1186/s13640-017-0213-4

Azmy, M. M. (2015). Classification of lung sounds based on linear prediction cepstral coefficients and support vector machine. In *Proceedings of the IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)* (pp. 1–5). IEEE. https://doi.org/10.1109/AEECT.2015.7360527

Bahoura, M. (2009). Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. *Computers in Biology and Medicine*, *39*(9), 824–843. https://doi.org/10.1016/j.compbiomed.2009.06.011

Bahoura, M., and Pelletier, C. (2004). Respiratory sounds classification using cepstral analysis and Gaussian mixture models. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)* (pp. 2661–2664). IEEE.

Bardou, D., Zhang, K., and Ahmad, S. M. (2018). Lung sounds classification using convolutional neural networks. *Artificial Intelligence in Medicine*, *88*, 58–69. https://doi.org/10.1016/j.artmed.2018.04.008

Bashir-Gonbadi, F., and Khotanlou, H. (2021). Multi-task convolutional autoencoder for brain tumor classification. *Multimedia Tools and Applications*, *80*, 21919–21939. https://doi.org/10.1007/s11042-020-10345-0

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798–1828. https://doi.org/10.1109/TPAMI.2013.50

Bohadana, A., Izbicki, G., and Kraman, S. S. (2014). Fundamentals of lung auscultation. *The New England Journal of Medicine*, *370*(8), 744–751. https://doi.org/10.1056/NEJMra1302901

Caruana, R. (1997). Multitask learning. *Machine Learning*, *28*(1), 41–75.

Chambres, G., Hanna, P., and Desainte-Catherine, M. (2018). Automatic detection of patient with respiratory diseases using lung sound analysis. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 181–185). IEEE. https://doi.org/10.1109/ICASSP.2018.8461404

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297. https://doi.org/10.1007/BF00994018

Ditthapron, A., Banluesombatkul, N., Ketrat, S., Chuangsuwanich, E., and Wilaiprasitporn, T. (2019). Universal joint feature extraction for P300 EEG classification using multi-task autoencoders. *IEEE Access*, *7*, 68415–68428. https://doi.org/10.1109/ACCESS.2019.2917982

Engin, M., Engin, A. Z., Şenol, Y., Memiş, S., Yıldırım, A., and Kafalı, H. (2022). Classification of lung sounds using machine learning methods. *Biomedical Signal Processing and Control*, *71*, 103101. https://doi.org/10.1016/j.bspc.2021.103101

Fraiwan, M., Fraiwan, L., Khassawneh, B., and Ibnian, A. A. (2021). A dataset of lung sounds recorded from the chest wall using an electronic stethoscope. *Data in Brief*, *35*, 106913. https://doi.org/10.1016/j.dib.2021.106913

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT Press.

Hassan, U., Singhal, A., and Chaudhary, P. (2024). Lung disease detection using EasyNet. *Biomedical Signal Processing and Control*, *91*, 105944. https://doi.org/10.1016/j.bspc.2024.105944

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Hershey, S., Chaudhuri, S., Ellis, D. P. W., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., Slaney, M., and Wilson, K. (2017). CNN architectures for large-scale audio classification. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 131–135). IEEE. https://doi.org/10.1109/ICASSP.2017.7952132

Huang, G., Song, S., Gupta, J. N. D., and Wu, C. (2014). Semi-supervised and unsupervised extreme learning machines. *IEEE Transactions on Cybernetics*, 44(12), 2405–2417. https://doi.org/10.1109/TCYB.2014.2307349

Jung, S.-Y., Liao, C.-H., Wu, Y.-S., Yuan, S.-M., and Sun, C.-T. (2021). Efficiently classifying lung sounds through depthwise separable CNN models with fused STFT and MFCC features. *Diagnostics*, *11*(4), Article 732. https://doi.org/10.3390/diagnostics11040732

Kompis, M., Pasterkamp, H., and Wodicka, G. R. (2001). Acoustic imaging of the human chest. *Chest*, *120*(4), 1309–1321. https://doi.org/10.1378/chest.120.4.1309

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

มหาวิทยาลัยเทคโนโลยีสุรนารี

Li, L., Xu, W., Hong, Q., Tong, F., and Wu, J. (2016). Classification between normal and adventitious lung sounds using deep neural network. In *Proceedings of the International Symposium on Chinese Spoken Language Processing (ISCSLP)* (pp. 1–5). IEEE.

Loudon, R. G., and Murphy, R. L. H. (1984). Lung sounds. *The American Review of Respiratory Disease*, *130*(4), 663–673. https://doi.org/10.1164/arrd.1984.130.4.663

Nair, V., and Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (pp. 807–814). Omnipress.

Niu, X., and Suen, C. Y. (2012). A novel hybrid CNN–SVM classifier for recognizing handwritten digits. *Pattern Recognition*, 45(4), 1318–1325. https://doi.org/10.1016/j.patcog.2011.09.021

O'Shea, K., and Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

Oppenheim, A. V., and Schafer, R. W. (2010). Discrete-time signal processing (3rd ed.). Pearson.

Pasterkamp, H., Kraman, S. S., and Wodicka, G. R. (1997). Respiratory sounds: Advances beyond the stethoscope. *American Journal of Respiratory and Critical Care Medicine*, *156*(3), 974–987. https://doi.org/10.1164/ajrccm.156.3.9701135

Pathonsuwan, W., Phapatanaburi, K., Buayai, P., Jumphoo, T., Anchuen, P., Uthansakul, M., and Uthansakul, P. (2022). RS-MSConvNet: A novel end-to-end pathological voice detection model. *IEEE Access*, *10*, 120450–120461. https://doi.org/10.1109/ACCESS.2022.3220171

Perna, D., and Tagarelli, A. (2019). Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks. In *Proceedings of the IEEE International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 50–55). IEEE. https://doi.org/10.1109/CBMS.2019.00020

Piczak, K. J. (2015). Environmental sound classification with convolutional neural networks. In *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1–6). IEEE. https://doi.org/10.1109/MLSP.2015.7324337

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, *2*(1), 37–63.

Rangayyan, R. M. (2015). *Biomedical signal analysis: A case-study approach* (2nd ed.). Wiley–IEEE Press.

Rocha, B. M., Filos, D., Mendes, L., Vogiatzis, I., Perantoni, E., Kaimakamis, E., Natsiavas, P., Jaume-i-Capó, A., and Paiva, R. P. (2020). An open access database for the evaluation of respiratory sound classification algorithms. *Physiological Measurement*, *41*(1), 015001. https://doi.org/10.1088/1361-6579/ab63f6

Sarkar, M., Madabhavi, I., Niranjan, N., and Dogra, M. (2015). Auscultation of the respiratory system. *Annals of Thoracic Medicine*, *10*(3), 158–168. https://doi.org/10.4103/1817-1737.160831

Sengupta, N., Sahidullah, M., and Saha, G. (2016). Lung sound classification using cepstral-based statistical features. *Computers in Biology and Medicine*, *75*, 118–129. https://doi.org/10.1016/j.compbiomed.2016.06.013

Schölkopf, B., and Smola, A. J. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press.

Shi, L., Du, K., Zhang, C., Ma, H., and Yan, W. (2019). Lung sound recognition using convolutional neural networks. *International Journal of Distributed Sensor Networks*, *15*(3), 1–9. https://doi.org/10.1177/1550147719834465

Sokolova, M., and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Soni, N., Chowdhury, A., Bhatia, M., and Chakrabarti, I. (2022). Self-supervised contrastive learning for respiratory sound classification. *Biomedical Signal Processing and Control*, *73*, 103431. https://doi.org/10.1016/j.bspc.2021.103431

Tripathy, R. K., Das, D., Ghose, A., and Ghosh, A. (2022). Automated detection of pulmonary diseases from lung sound signals using fixed-boundary-based empirical wavelet transform. *IEEE Sensors Letters*, *6*, 1–4. https://doi.org/10.1109/LSENS.2022.3147487

van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605.

Xia, Y., Han, J., and Mascolo, C. (2022). A survey on respiratory sound analysis for disease screening. *IEEE Reviews in Biomedical Engineering*, *15*, 220–236. https://doi.org/10.1109/RBME.2021.3117566

Zhang, Z., Li, H., Zhang, R., Wang, Y., and Guan, C. (2021). Attention-based convolutional neural network for lung sound classification. *Biomedical Signal Processing and Control*, *69*, 102917. https://doi.org/10.1016/j.bspc.2021.102917

ภาคผนวก ก

บทความวิชาการที่ได้รับการตีพิมพ์เผยแพร่ในระหว่างการศึกษา

มหาวิทยาลัยเทคโนโลยีสุรนารี

รายชื่อบทความวิชาการที่ได้รับการตีพิมพ์เผยแพร่ระหว่างการศึกษา

Orkweha, K., Phapatanaburi, K., Pathonsuwan, W., Jumphoo, T., Rattanasak, A., Anchuen, P., Pinthurat, W., Uthansakul, M., and Uthansakul, P. (2024). A framework for detecting pulmonary diseases from lung sound signals using a hybrid multi-task autoencoder–SVM model. Symmetry, 16(11), 1413. https://doi.org/10.3390/sym16111413

Duangmanee, P., Phapatanaburi, K., Pathonsuwan, W., Jumphoo, T., Rattanasak, A., Orkweha, K., Anchuen, P., Uthansakul, M., and Uthansakul, P. (2025). Triplet multi-kernel CNN for detection of pulmonary diseases from lung sound signals. IEEE Access, 13, 50137–50151. https://doi.org/10.1109/ACCESS.2025.3552108

มหาวิทยาลัยเทคโนโลยีสุรนารี

symmetry

MDPI

Article

A Framework for Detecting Pulmonary Diseases from Lung Sound Signals Using a Hybrid Multi-Task Autoencoder-SVM Model

Khwanjit Orkweha [1], Khomdet Phapatanaburi [2,*], Wongsathon Pathonsuwan [3], Talit Jumphoo [3], Atcharawan Rattanasak [1], Patikorn Anchuen [4], Watcharakorn Pinthurat [5], Monthippa Uthansakul [1] and Peerapong Uthansakul [1,*]

1 School of Telecommunication Engineering, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand; d6500597@g.sut.ac.th (K.O.); d6500658@g.sut.ac.th (A.R.); mtp@sut.ac.th (M.U.)
2 Department of Telecommunication Engineering, Faculty of Engineering and Technology, Rajamangala University of Technology Isan (RMUTI), Nakhon Ratchasima 30000, Thailand
3 Institute of Research and Development, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand; pathonsuwan.wst@sut.ac.th (W.P.); jumphoo@sut.ac.th (T.J.)
4 Navaminda Kasatriyadhiraj Royal Air Force Academy, Saraburi 18180, Thailand; patikorn_an@rtaf.mi.th
5 Department of Electrical Engineering, Rajamangala University of Technology Tawan-Ok, Chanthaburi 22210, Thailand; watcharakorn_pi@rmutto.ac.th
* Correspondence: khomdet.ph@rmuti.ac.th (K.P.); uthansakul@sut.ac.th (P.U.)

Abstract: Research focuses on the efficacy of Multi-Task Autoencoder (MTAE) models in signal classification due to their ability to handle many tasks while improving feature extraction. However, researchers have not thoroughly investigated the study of lung sounds (LSs) for pulmonary disease detection. This paper introduces a new framework that utilizes an MTAE model to detect lung diseases based on LS signals. The model integrates an autoencoder and a supervised classifier, simultaneously optimizing both classification accuracy and signal reconstruction. Furthermore, we propose a hybrid approach that combines an MTAE and a Support Vector Machine (MTAE-SVM) to enhance performance. We evaluated our model using LS signals from a publicly available database from King Abdullah University Hospital. The model attained an accuracy of 89.47% for four classes (normal, pneumonia, asthma, and chronic obstructive pulmonary disease) and 90.22% for three classes (normal, pneumonia, and asthma cases). Using the MTAE-SVM, the accuracy was further improved to 91.49% for four classes and 93.08% for three classes, respectively. The results indicate that the MTAE and MTAE-SVM have a considerable potential for detecting pulmonary diseases from lung sound signals. This could aid in the creation of more user-friendly and effective diagnostic tools.

Keywords: pulmonary disease detection; multi-task autoencoder; MFCC; audio diagnostics; health informatics

check for updates

Citation: Orkweha, K.; Phapatanaburi, K.; Pathonsuwan, W.; Jumphoo, T.; Rattanasak, A.; Anchuen, P.; Pinthurat, W.; Uthansakul, M.; Uthansakul, P. A Framework for Detecting Pulmonary Diseases from Lung Sound Signals Using a Hybrid Multi-Task Autoencoder-SVM Model. *Symmetry* **2024**, *16*, 1413. https://doi.org/10.3390/sym16111413

Academic Editor: Tomohiro Inagaki

Received: 12 September 2024
Revised: 17 October 2024
Accepted: 22 October 2024
Published: 23 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Breathing-generated lung sounds (LSs) carry valuable health information [1] crucial for diagnosing lung illnesses [2]. Given the highly transmissible and potentially severe nature of many diseases, prompt and accurate identification is paramount. However, this task is challenging as LSs often occur at frequencies beyond human hearing, increasing the risk of missing important diagnostic cues. Furthermore, the assessment of LSs, known as auscultation, by healthcare practitioners is inherently subjective and may lead to diagnostic errors [3]. To tackle these challenges [4], there is a growing trend towards adopting computer-aided diagnosis methods. These technologies offer a more objective and reliable approach to diagnosing respiratory disorders, enhancing the accuracy and efficiency of lung disease diagnostics. In this paper, we focus on a framework for detecting pulmonary diseases (PDs) from LS signals, leveraging pattern recognition techniques in the field

of computer science. This approach falls within the broader domain of symmetry in computational models.

In recent years, significant advancements have been made in detecting PDs using LS signals and conventional machine learning techniques. The typical structure of these LS detection systems comprises two main phases: initial feature extraction and subsequent classification. During feature extraction, these systems meticulously analyze breathing sounds, employing various classification models. Previous studies have demonstrated a preference for wavelet-based features [5–7], Mel Frequency Cepstral Coefficients (MFCCs) [8–11], Linear Predictive Cepstral Coefficients (LPCCs) [12], bispectrum [8], and Empirical Wavelet Transform with Fixed Boundary Points (EWT-FBP) [13]. Particularly, MFCCs have gained widespread adoption due to their exceptional ability to capture crucial spectral characteristics of LSs, essential for distinguishing between different pulmonary conditions. They provide a compact yet comprehensive representation of audio signals, facilitating efficient analysis. Their small size and robustness in noisy environments make them ideal for clinical settings where LS recordings may contain background noise. In the classification phase, various models including Support Vector Machines (SVMs) [5,12,14], Gaussian Mixture Models (GMMs) [15,16], k-Nearest Neighbors (KNN) [17], Linear Discriminant Analysis (LDA) [17], Decision Trees (DTs) [18], Multilayer Perceptron (MLP) [13], Random Forest (RF) [13,19], Extreme Gradient Boosting (XGBoost) [13], and Light Gradient Boosting Machine (LGBM) [13,16,17] have been utilized. The choice of classifier significantly impacts the system's effectiveness in diagnosing PDs from LSs, underscoring the importance of proficient feature extraction design. Therefore, the development of these systems necessitates expertise in feature extraction methodologies to ensure accurate and reliable detection of pulmonary conditions.

In addition to the classifiers previously discussed, the utilization of Deep Neural Network (DNN) methods in detecting PDs has gained traction due to their growing effectiveness [20,21]. A noteworthy development in this domain is the application of Convolutional Neural Networks (CNNs) for LS analysis, elucidated in [22]. This approach commences with the extraction of MFCC features, which are then inputted into a two-layer CNN (CNN-2L) for classification. The CNN-2L has demonstrated superior performance compared with traditional SVM-based techniques, particularly in distinguishing normal LSs from those indicative of asthma or pneumonia. This success underscores the potential of combining CNNs with MFCCs for more precise classification. The CNN-2L serves as a benchmark model in this study for PDs detection from LS signals. Furthermore, there has been notable progress in merging DNNs with Hidden Markov Models (DNN-HMMs), as detailed in [23]. This fusion harnesses the predictive capabilities of DNNs and the proficiency of HMMs in recognizing temporal patterns, significantly enhancing classification performance. Another significant advancement is the integration of the VGGish network with the bidirectional gated recurrent unit neural network (BiGRU), as proposed in [24]. This hybrid model utilizes Mel-spectrogram features, capitalizing on the VGGish network's advanced feature extraction and the BiGRU's capability to process sequential data. The VGGish-BiGRU model exemplifies the benefits of amalgamating these advanced neural network architectures, leading to enhanced accuracy and reliability in LS analysis for disease identification, thereby paving the way for future advancements.

Deep learning algorithms have demonstrated remarkable potential in accurately identifying diseases from LS signals. However, these systems often focus on single tasks, potentially overlooking critical information. To address this issue, researchers have developed Multi-Task Autoencoders (MTAEs), aiming to simultaneously optimize both the reconstruction fidelity of the input signal and the accuracy of disease classification. This dual-focus approach ensures not only precise disease detection but also the preservation of detailed information from the original LSs, essential for complex medical diagnostics. The versatility of MTAEs is further evidenced by their successful deployment in various domains, demonstrating their robustness and adaptability in handling intricate data-driven challenges. A notable example is the Event-Related Potential Encoder Network

(ERPENet) [25], which integrates CNNs with Long Short-Term Memory (LSTM) within an autoencoder architecture to efficiently compress Electroencephalography (EEG) signals and extract critical P300 features, essential for tasks like the P300 speller. Similarly, a novel end-to-end multi-task learning method, referred to as MIN2Net, was introduced in [26], which represents an advanced MTAE model for independent motor imagery EEG classification, outperforming standard benchmarks in brain–machine interfaces on the SMR-BCI and OpenBMI datasets. Another significant advancement is the multi-task convolutional autoencoder-based neural network for brain tumor classification, as reported in [27]. Applied to the Cheng dataset, this method achieves remarkable accuracy, categorizing brain tumors into meningioma, glioma, and pituitary with a 99.30% success rate. These instances collectively underscore the sophisticated capabilities and adaptability of MTAEs in managing complex, data-centric tasks. However, their application in PD detection using LS signals remains uncharted territory, offering opportunities for significant advancements in medical diagnostics.

In this paper, we introduce an innovative framework utilizing an MTAE, specifically tailored for comprehensive analysis of LS signals. This novel framework integrates an autoencoder and a supervised classifier in a synergistic combination. The autoencoder, comprising an encoder and a decoder, is adept at creating a latent vector that encapsulates the fundamental characteristics of LSs. Simultaneously, the supervised classifier interprets this latent vector to extract clinically significant information. This model meticulously designs its training process to simultaneously optimize two crucial factors: the accuracy of disease classification and the fidelity of signal reconstruction. This dual-focus approach is pivotal for enhancing the model's capability to accurately detect various lung diseases while preserving the integrity of the original LS signals. We were inspired by [28,29], which combined a CNN/Raw Speech-Based Multi-Scale Convolutional Neural Network (RS-MSConvNet) for trainable feature representation with an SVM as the classifier. This model was more accurate than CNN/RS-MSConvNet alone, which was mostly due to its ability to use both the CNN's feature extraction strengths and the SVM's classification strengths. In particular, the distance between the CNN/RS-MSConvNet features that were automatically extracted and the trained SVM's hyperplane made it more specific, which improved its ability to detect. In this paper, we propose a hybrid MTAE-SVM model that replaces the final fully connected layers of the MTAE after full training, resulting in further improvements in classification accuracy.

In this paper, we introduce an innovative framework utilizing an MTAE, specifically tailored for comprehensive analysis of LS signals. This novel framework integrates an autoencoder and a supervised classifier in a synergistic combination. We chose the MTAE model for its distinctive capacity to simultaneously improve classification accuracy and signal reconstruction quality, which is essential for the effective analysis of complex patterns in LS data. Our model addresses the shortcomings of current methods that concentrate only on classification, which could miss essential diagnostic information comprised in the LS signals. The autoencoder, comprising an encoder and a decoder, is adept at creating a latent vector that encapsulates the fundamental characteristics of LSs. Simultaneously, the supervised classifier interprets this latent vector to extract clinically significant information. This model meticulously designs its training process to simultaneously optimize two crucial factors: the accuracy of disease classification and the fidelity of signal reconstruction. This dual-focus approach is pivotal for enhancing the model's capability to accurately detect various lung diseases while preserving the integrity of the original LS signals. Our experimental results demonstrate that the proposed MTAE model achieves greater accuracy in PD detection from LS waves, notably outperforming current approaches, including the benchmark CNN-2L model. This increase in performance demonstrates the efficacy of our method and its potential to advance medical diagnostics.

The salient contributions of this paper can be summarized as:

1. This study introduces the inaugural design of a novel framework utilizing MTAE models tailored specifically for the detection of PDs from LS signals. To the best of

our knowledge, this is the first attempt to introduce a framework that integrates both regression and classification methods for the detection of PDs using LS signals.

2. We investigate our model on the dataset provided by [30] from King Abdullah University Hospital (KAUH). Our framework demonstrates remarkable accuracy, achieving 89.47% reliability in distinguishing between normal, pneumonia, asthma, and chronic obstructive pulmonary disease (COPD) using fivefold cross-validation on the KAUH dataset. Additionally, it attains a 90.22% accuracy in classifying normal, pneumonia, and asthma cases.
3. The MTAE model is further refined by incorporating an SVM method, replacing the final fully connected layer to classify the automatically extracted features from the fully trained MTAE. This modified version, referred to as the hybrid MTAE-SVM model architecture, delivers enhanced accuracy compared with the MTAE classifier alone.

The subsequent sections of the paper are structured as follows. Related work is presented and discussed in Section 2. The proposed framework is explained in Section 3, while the experimental setup and the obtained results are deliberated in Sections 4 and 5, respectively. Finally, Section 6 concludes the paper and offers possible directions for future investigation.

2. Related Works

This section offers a concise overview of commonly employed CNN methodologies for detecting PDs using LS signals, which will serve as our baseline for comparison with our proposed approach. The emphasis lies on utilizing MFCCs for feature extraction, followed by CNNs for classification.

2.1. MFCC

The MFCC is pivotal in audio signal processing, especially in capturing the short-term power spectrum of a sound. Obtaining the MFCC involves a sequence of successive steps. Initially, the process utilizes the Fourier Transform to generate the power spectrum. Then, the spectra are mapped onto the mel scale, which reflects the perceived frequency of a tone. This is followed by logarithmic scaling and ultimately culminates in the application of the inverse Fourier Transform. The mathematical formulation of the MFCC is presented as,

$$MFCC(i) = \sum_{k=0}^{N-1} \log(P(k)) \cdot \cos\left((2k+1)\frac{\pi i}{2N}\right), \quad \text{for } i = 1, 2, \ldots, M, \tag{1}$$

where $P(k)$ denotes the power spectrum, N represents the count of mel filters, and M indicates the number of cepstral coefficients.

Following the approach outlined in [31], we resized the initial MFCC images to a more practical 28×28 pixel grayscale format. This resizing was intentionally performed to reduce computational burden while preserving data integrity.

2.2. CNN-Based Classifier

CNNs play a pivotal role in classifying PDs using LS signals. Following feature extraction using the MFCC, the processed data are fed into a CNN, renowned for its exceptional proficiency in detecting patterns and conducting classification tasks. Specifically designed to scrutinize the precise features of the sound signals represented in MFCC images, the CNN architecture comprises convolutional layers, pooling layers, and fully connected (FC) layers arranged hierarchically. Each layer of the CNN contributes to the detection process by extracting and enhancing features from the MFCC data, ultimately leading to robust and precise classification of LSs.

The CNN methodology commences with convolutional layers, which utilize filters on the input MFCC images to produce feature maps, highlighting significant characteristics in the sound signals. This approach is pivotal for discerning the intricate patterns

in LSs indicative of various lung diseases. Subsequently, max-pooling layers reduce the dimensionality of the feature maps, thereby enhancing the computational efficiency of the network and mitigating the risk of overfitting. The final layers of the CNN are densely connected, enabling advanced reasoning processes based on the extracted features. These layers amalgamate information from the feature maps to ultimately classify the LSs, distinguishing between healthy and diseased conditions. The CNN architecture configuration can be observed in Figure 1.

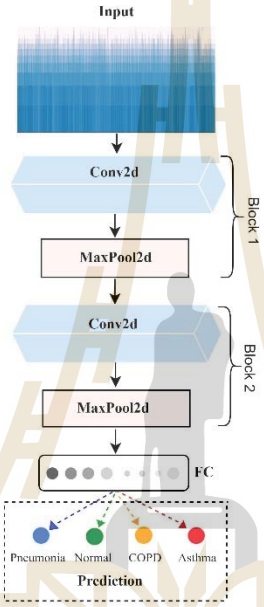


Figure 1. The configuration of CNN architecture.

3. Proposed Framework

This section presents our innovative approach for detecting PDs from LS signals using the proposed MTAE model and the MTAE-SVM hybrid architecture. The methodology is structured into three key phases: data preparation and feature extraction, the design and training of the MTAE model, and the integration of the MTAE-SVM hybrid model architecture.

3.1. Data Preparation and Feature Extraction

The dataset comprises audio recordings demonstrating various types of PDs, with each recording indexed by its associated file path and disease label. The preprocessing step involves applying a Butterworth bandpass filter to the audio data. The transfer function of the filter is given by,

$$H(f) = \frac{1}{\sqrt{1 + \left(\frac{f}{f_c}\right)^{2n}}}, \tag{2}$$

where n represents the filter order (set to 12), f is the frequency, and f_c is the cut-off frequencies, ranging between 120 Hz and 1800 Hz. Following this, MFCCs, as elaborated in the preceding section, are extracted from these filtered signals.

3.2. MTAE Model Architecture

This section presents a novel MTAE model meticulously crafted as a sophisticated neural network consisting of two primary components: an autoencoder for precise signal reconstruction and a classifier for accurate disease identification. These elements work in tandem to analyze LS signals, offering vital insights for advanced medical diagnostics. The structure of the proposed MTAE framework is depicted in Figure 2.

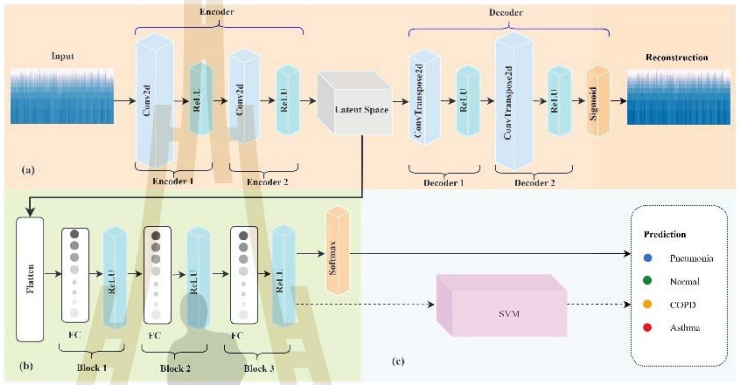


Figure 2. The architecture of our proposed framework, illustrating: (**a**) the autoencoder for unsupervised feature learning, (**b**) the MTAE-based classifier for multi-task representation learning, and (**c**) the MTAE-SVM hybrid classifier, which integrates MTAE features with an SVM for final classification.

3.2.1. Autoencoder

The autoencoder, an integral component of our strategic framework, has been meticulously crafted to transform intricate LS data into a condensed, lower-dimensional latent vector, facilitating swift feature extraction. This transformation process is essential for capturing inherent patterns within the data, crucial for precise disease detection. The autoencoder systematically restores the data to their original high dimensionality, thereby preserving the quality of the signals.

Encoder

The encoder plays a pivotal role in constraining the input data. It adeptly captures and encodes the most crucial characteristics into a hidden space representation, a critical step for effective pattern recognition. This is achieved through a series of convolutional layers, each activated by a Rectified Linear Unit (ReLU) function. The ReLU function, denoted as $\mathrm{ReLU}(x) = \max(0, x)$, introduces non-linearity, enabling the model to better analyze intricate and non-linear relationships observed in LS data. By exclusively preserving positive values, this function significantly enhances the model's learning process and addresses challenges like the vanishing gradient problem, commonly encountered in DNN training.

Decoder

The decoder functions as the counterpart to the encoder, aiming to accurately reconstruct the original input from its compressed latent representation. This task is achieved by employing transposed convolutional layers, each accompanied by ReLU activations. These layers progressively enhance the resolution of the compressed, data-dense latent vector to align with the original input, maintaining a high number of dimensions. The symmetrical arrangement of the encoder and decoder ensures the retention of vital data

characteristics, enabling the model to obtain a concise yet comprehensive representation of the input. This capability proves particularly advantageous for unsupervised learning tasks, such as anomaly identification and noise reduction in LS analysis, where grasping the underlying structure of the data is paramount.

3.2.2. MTAE-Based Classifier

The classifier is intricately linked to the latent vector of the autoencoder and has been meticulously crafted for the specific task of disease classification. Comprising multiple linear layers, the system is adeptly structured to interpret the encoder's output, thus transforming it into precise disease detection results. The architecture of the classifier is meticulously designed to enhance its capability to discern subtle variations in LS patterns, which serve as indicators of different PDs. The output of the classifier plays a crucial role in computing the cross-entropy loss, a fundamental metric for training and refining the model. With its emphasis on precision and accuracy in disease detection, the classifier emerges as an indispensable component in the diagnostic process.

3.2.3. Training Procedure for MTAE

Throughout the training phase, our model is meticulously engineered to achieve two primary objectives: precise disease classification and effective signal reconstruction. Through the utilization of a dual-target approach, the model's ability to generalize to unfamiliar data and uphold accuracy in disease categorization is enhanced.

The focus of the training revolves around optimizing a composite loss function, which is defined as,

$$L = L_{recon} + L_{class}, \tag{3}$$

where L_{recon} represents the reconstruction loss, quantified by the Mean Squared Error (MSE), and L_{class} is the classification loss, determined using the cross-entropy (CE) loss. These components are crucial in steering the model's learning process:

- **MSE for reconstruction:** this loss function evaluates the average squared discrepancies between the reconstructed signals and the original inputs. It is formulated as,

$$L_{recon} = \frac{1}{S}\sum_{i=1}^{S}(Y_i - \hat{Y}_i)^2, \tag{4}$$

  where S is the total number of samples, Y_i is the actual signal value, and $\hat{Y}_i$ the reconstructed version. The MSE is vital for ensuring the model accurately reproduces the input signals, a key aspect for effective diagnosis.
- **CE for classification:** this loss assesses the model's accuracy in disease classification and is formulated as,

$$L_{class} = -\sum_{c=1}^{C} y_{o,c}\log(p_{o,c}), \tag{5}$$

  where C is the number of disease categories, y is a binary indicator (1 if the class label c is correct for observation o, 0 otherwise), and p is the predicted probability of observation o belonging to class c. This loss refines the model's capacity to distinguish between diseases.

By merging these two loss functions, the model is effectively guided towards accurate disease categorization while simultaneously proficiently reconstructing input signals. Implementing this well-balanced strategy is pivotal in crafting a robust model highly suitable for practical diagnostic applications. This joint optimization implies that the common latent representation includes features that are both distinctive for classification and informative for reconstruction, as suggested in [32,33]. Consequently, the model augments its capacity to identify small deviations in LS signals, thereby enhancing classification accuracy while

preserving high-quality signal reconstruction, which is crucial for the development of dependable diagnostic tools.

3.3. MTAE-SVM Model Architecture

The SVM-based approach has proven effective in both classification and regression tasks [34], making it widely applicable across various domains [35–37], including image classification [38] and audio classification [39]. In addition, the SVM-based method is one popular method for detecting pulmonary diseases due to its robust handling of complex, high-dimensional feature spaces, which are common in medical signal processing. Previous studies [28,29] motivated us by demonstrating the efficacy of integrating a CNN or an RS-MSConvNet for feature extraction alongside an SVM classifier. The hybrid models demonstrated superior accuracy compared with utilizing a CNN or RS-MSConvNet independently, primarily due to their ability to utilize the advantages of the CNN's feature extraction alongside the SVM's classification strengths. The features extracted automatically by the CNN/RS-MSConvNet, when fed into the SVM, leveraged the SVM's ability to identify optimal separating hyperplanes within the feature space. This improvement increased the model's precision and overall detection capabilities. This paper creates the use of an SVM-based method to classify the features derived from the fully trained MTAE at the final FC layer. To implement the SVM within the hybrid model, we utilize the Radial Basis Function (RBF) kernel. The penalty parameter, C, and the optimal kernel parameter, γ, are determined through validation data analysis on the hybrid model, which is trained using the training data.

The implementation of the MTAE-SVM model, as depicted in Figure 2, involves several key steps. First, during the training phase, the training set samples are input into the MTAE model. Once fully trained, the MTAE classifier automatically extracts the relevant features for each input. We then replace the FC layers with an SVM-based classifier, which learns from the extracted feature vectors produced by the MTAE. During testing, the fully trained MTAE processes the samples from the test set to generate feature representations. We then feed these extracted features into the trained SVM to detect PDs.

4. Experimental Setup

In this section, the datasets used to verify the performance our framework, feature extraction, and training setup, as well as performance criteria, are systematically presented.

4.1. Database

Two widely used databases for LS classification are the International Conference on Biomedical Health Informatics (ICBHI) and KAUH databases. In this paper, we employed the KAUH database to assess the effectiveness of our proposed methods. The primary reason for selecting this database was its suitability for subject-independent classification experiments based on fivefold cross-validation, as it includes more than four subjects per disease category (normal, asthma, pneumonia, COPD). The details of the KAUH database are shown in Table 1. The objective of this study was to assess the effectiveness of our proposed approach in identifying PDs from LS signals. The evaluation was conducted using a publicly available dataset, as cited in [30]. The dataset includes participants categorized into various health conditions: 35 classified as normal, 32 as asthma, 5 as pneumonia, and 9 as COPD. The LS signal was recorded at a sampling frequency of 4 kHz.

Table 1. The demographic information of the subjects based on the KAUH database.

Pulmonary Disease	No. of Subjects	Age Range	Gender (Female, Male)	Segmented Waveform No. (Original LS)
Normal	35	18–81	11 female, 24 male	415 (105)
Asthma	32	12–72	17 female, 15 male	357 (96)
Pneumonia	5	36–70	2 female, 3 male	66 (15)
COPD	9	42–76	1 female, 8 male	102 (27)

A crucial preprocessing stage in our methodology involved segmenting every LS signal into individual frames. Segmentation was accomplished by employing a non-overlapping window approach. In this study, we opted for window durations of 4 s due to their substantial impact on the overall detection performance of our method. Through extensive testing, the suitable window length was determined to ensure it captured essential clinical characteristics in the LS signals while optimizing the efficiency of the algorithm. The quantities of frames with a duration of 4 s are given in Table 1.

4.2. MFCC Configuration

The MFCC was utilized as the primary feature for inputting data into our model. The coefficients were calculated using frame lengths of 20, 40, 128, 256, 512, and 1000 ms. Each frame underwent processing with a Hamming window and had a 50% overlap. This approach resulted in a 13-dimensional MFCC. To investigate temporal changes in the audio frames, we computed both the delta and double-delta coefficients of the MFCC. Consistent with [40], we utilized a segment duration of 4 s in our investigation. Following the recommendation in [31], we converted the 39-dimensional coefficients of the CNN-2L-based classifier into a grayscale format with dimensions of 28 × 28 pixels. In contrast, our proposed model utilized the original 39-dimensional coefficients directly as input.

4.3. Network Training

The evaluation models were constructed using the PyTorch v1.10.1 framework and trained on an NVIDIA RTX3090 GPU with 24 GB of RAM. The training parameters can be found in Table 2.

Table 2. Model parameters of the MTAE.

Parameters	MTAE
Batch Size	32
Epochs	350
Optimizer	ADAM
Loss	Cross-entropy
Output Activation	Softmax
Hidden Layer Activation	ReLU

4.4. Performance Criteria

When assessing the classifier's performance, we employed commonly used metrics outlined in [41], described as precision (Pr), recall (Re), F1-score ($F1$), and accuracy (Acc), which can be defined as,

$$Pr = \frac{TP}{TP + FP}, \tag{6}$$

$$Re = \frac{TP}{TP + FN}, \tag{7}$$

$$F1 = 2 \times \frac{Pr \times Re}{Pr + Re}, \tag{8}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \tag{9}$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative values, respectively.

5. Results and Discussion

This section showcases the results achieved by our proposed framework, alongside a comparative analysis of its performance with similar studies.

มหาวิทยาลัยเทคโนโลยีสุรนารี

5.1. Results Based on MTAE Classification

This section unveils the outcomes of our proposed MTAE module. The subsequent subsections detail the findings obtained from adjusting the parameter configurations.

- The duration of the frame in the MFCC significantly influences the classification accuracy in deep learning models. Hence, identifying the optimal frame length for the MFCC is crucial to enhance the effectiveness of our proposed MTAE. This study employed an MTAE comprising three encoder and three decoder blocks for regression, with an additional three layers for classification. The architectural and configuration details of this MTAE are outlined in Table 3. A Learning Rate (LR) of 0.00001 was selected. The experimental results, illustrating the impact of various MFCC frame lengths on the performance of our MTAE, are presented in Table 4. As depicted in Table 4, the MFCC with a frame duration of 20 ms in our MTAE exhibited the highest efficacy in detecting PDs using LS signals. This selected frame length proved to be optimal as it strikes a balance between capturing spectral details and preserving temporal resolution, closely resembling the typical duration of a vowel sound in speech. This duration holds pivotal importance in LS studies, enabling precise collection of detailed acoustic characteristics necessary for distinguishing between normal and diseased sounds. Conversely, longer frame lengths tend to oversimplify these distinct characteristics, leading to a reduction in classification accuracy. This finding aligns with prior research presented in [28], underscoring its significance in LS studies for PD diagnosis. Consequently, we consistently adopted a frame length of 20 ms in the MFCC as the input parameter for our MTAE in all subsequent tests.
- Choosing the LR significantly influences the effectiveness of a deep learning model. An excessively high LR can cause the model to overshoot optimal weights, whereas an excessively low LR may lead to slow convergence or becoming stuck in suboptimal solutions. This study experimented with various LR values, and the results highlighted in Table 5 underscore the substantial impact of this hyperparameter. Remarkably, an LR of 0.0001 (LR = 0.0001) achieved the highest accuracy of 81.70%. This LR effectively adjusted the weights while preventing overshooting. Additionally, employing an LR of 0.001 (LR = 0.001) resulted in a commendable accuracy of 75.96%, demonstrating its suitability for the task at hand. However, setting the LR too high (0.01) or too low (0.000001) led to diminished model performance, underscoring the importance of careful LR adjustment for optimal training and convergence. Upon analyzing the results, we opted to utilize an LR of 0.0001 as the default setting for our MTAE in all subsequent studies, given its proven effectiveness in achieving optimal performance.
- The efficacy of our MTAE model is notably influenced by its architecture, particularly the configuration of encoder and decoder blocks during the regression phase, along with the number of FC layers in the classification component. The careful selection of these architectural elements plays a pivotal role in maximizing its effectiveness. We systematically conducted experiments by varying both the combinations of encoder–decoder blocks and the number of FC layers to assess their impact on accuracy. The outcomes of these adjustments, outlined in Tables 6 and 7, provide valuable insights into the optimal structural arrangement for our model. The results presented in Table 6 regarding encoder–decoder combinations reveal a significant relationship between model accuracy and architectural complexity. The configuration of 2-2 exhibits the highest accuracy rate of 89.47%, indicating that a well-balanced level of complexity is most effective for this model. This implies that having two layers in both the encoder and decoder achieves an optimal balance between capturing essential characteristics and preventing overfitting. In contrast, as the model's complexity increases with 3-3, 4-4, and 5-5 arrangements, there is a noticeable decline in accuracy. This decline is likely attributed to overfitting, where the model becomes overly attuned to the training data, thus compromising its ability to generalize. Remarkably, even the simplest 1-1 configuration still achieves a reasonably high accuracy of 88.51%. This suggests that

minimal configurations can be beneficial, although they are not as favorable as the 2-2 configuration. Based on the insights provided in Table 7, the highest accuracy is attained when utilizing three FC layers, achieving an accuracy of 89.47%. The second-highest accuracy, at 87.55%, is achieved with a configuration of two FC layers. This trend indicates that increasing the number of FC layers up to a certain threshold can enhance the model's data classification capabilities. However, the introduction of additional layers beyond this optimal threshold does not necessarily lead to performance improvement, as evidenced by the slight declines in accuracy observed for configurations with four and five FC layers (86.70% and 86.38%, respectively). This observation suggests the potential for overfitting or a limitation in learning capacity with additional layers. The model's performance is most compromised when employing only one FC layer, resulting in an accuracy of 85.84%. This underscores the necessity of multiple layers to achieve effective performance.

- To assess the effectiveness of multi-task learning, we adapted our model by removing the decoder component and relied solely on entropy loss for classification tasks. This modification led to a decrease in accuracy, with performance dropping to 88.40%. This outcome supports the hypothesis that a multi-task learning approach, utilizing a shared latent vector, is more efficient in detecting PDs from LS signals compared with a single-task learning technique that employs a non-shared latent vector. Such a technique aligns with similar methodologies outlined in the literature [27], suggesting that shared representations in multi-task learning may yield stronger and more adaptable models, particularly in complex tasks such as medical diagnosis.
- Following thorough parameter optimization, we settled on an LR of 0.0001 and finalized the design of our MTAE system. This architecture, featuring a 2-2 encoder–decoder arrangement and three FC layers, corresponds to the design depicted in Figure 2.

Furthermore, we evaluated the effectiveness of characteristics extracted from the trained MTAE in discerning various types of PDs in LSs. To accomplish this, we employed t-distributed Stochastic Neighbor Embedding (t-SNE) [42] for dimensionality reduction and comparison of raw sound signals, features extracted from the trained CNN-2L model, and features extracted from the learned MTAE model. We conducted an analysis of the distributions of four classes using t-SNE, selecting one out of five folds for each analysis. Figure 3 illustrates the t-SNE-based feature distributions obtained by comparing raw sound signals, features from the trained CNN-2L model, and features from the trained MTAE model. Specifically, Figure 3a showcases a noticeable convergence of class distributions when using raw sound signals, leading to limited distinction among diseases. In contrast, Figure 3b illustrates improved differentiation across classes using the CNN-2L model, albeit with some remaining overlap. Figure 3c demonstrates that features obtained via MTAE exhibit clearer boundaries and shorter gaps between different classes, indicating enhanced potential for effectively detecting PDs from LS signals.

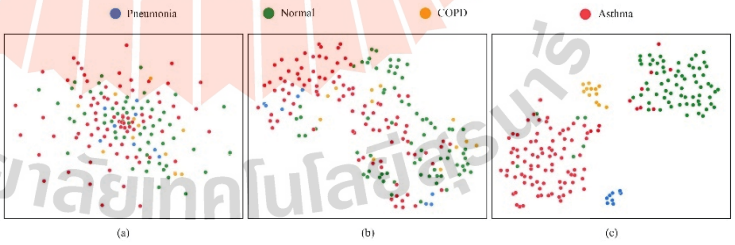


Figure 3. t-SNE visualization of feature distributions: (**a**) raw LS signal features, (**b**) features extracted from the baseline CNN model, and (**c**) features learned by the MTAE.

Lastly, to explore the relationship between each disease category, we present the confusion matrices. Figure 4 showcases the confusion matrices generated by our proposed framework. The confusion between COPD and asthma primarily stems from the similarities in their clinical presentations. However, our approach notably improves the categorization of pneumonia compared with the CNN-2L model, as evidenced by the increased true positive rate for pneumonia in matrix (b). The suggested MTAE exhibits a high overall accuracy in predicting the 'normal' class, while also mitigating the occurrence of false positives compared with the baseline. This suggests that the MTAE more effectively captures the unique characteristics of each disease, although it still encounters challenges in differentiating between diseases with similar symptoms.

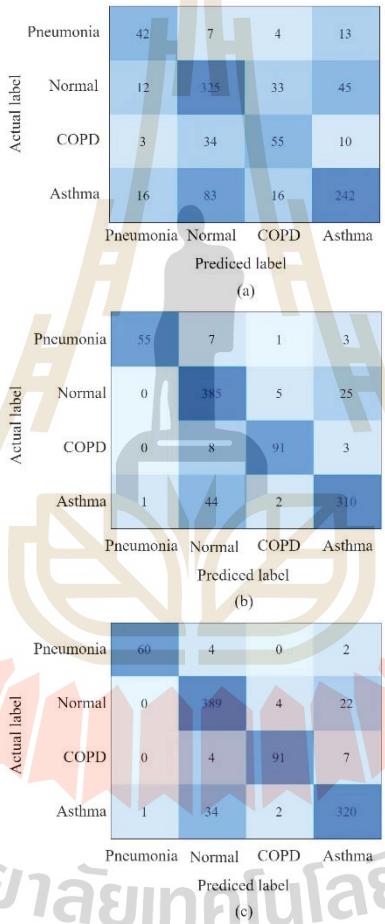


Figure 4. Confusion matrices for (**a**) CNN-based baseline, (**b**) MTAE, and (**c**) MTAE-SVM. The color intensity represents the proportion of correctly or incorrectly classified instances, with darker shades indicating higher values.

มหาวิทยาลัยเทคโนโลยีสุรนารี

Table 3. Configuration of the MTAE designed for investigating the impact of varying frame lengths in MFCC analysis.

Module	Layer	Kernel	Layer Configuration	Activation
Auto. (Encoder1)	Conv2d 1	3 × 3	in_channels = 1, out_channels = 16	ReLU
Auto. (Encoder2)	Conv2d 2	3 × 3	in_channels = 16, out_channels = 8	ReLU
Auto. (Encoder3)	Conv2d 3	3 × 3	in_channels = 8, out_channels = 2	ReLU
Auto. (Decoder1)	ConvTranspose2d 1	3 × 3	in_channels = 2, out_channels = 8	ReLU
Auto. (Decoder2)	ConvTranspose2d 2	3 × 3	in_channels = 8, out_channels = 16	ReLU
Auto. (Decoder3)	ConvTranspose2d 3	3 × 3	in_channels = 16, out_channels = 1	ReLU and Sigmoid
Class. (Block1)	FC (Linear 1)	-	in_features = [flatten latent size], out_features = 128	ReLU
Class. (Block2)	FC (Linear 2)	-	in_features = 128, out_features = 128	ReLU
Class. (Block3)	FC (Linear 3)	-	in_features = 128, out_features = 128	ReLU
	Classifier (Linear 4)	-	in_features = 128, out_features = 4	Softmax

Auto.: autoencoder, Class.: classification.

Table 4. Effect of MFCC frame length on classification performance in the MTAE model.

Frame Length	Acc (%)	Pr (%)	Re (%)	F1 (%)
20 ms	71.28 ± 3.30	74.40 ± 9.21	52.60 ± 6.11	54.60 ± 5.13
40 ms	61.17 ± 5.10	47.60 ± 16.58	41.80 ± 5.50	40.20 ± 6.76
128 ms	56.60 ± 1.71	36.80 ± 11.34	35.40 ± 2.30	32.60 ± 2.70
256 ms	55.75 ± 4.62	33.00 ± 12.02	33.80 ± 2.28	31.20 ± 2.95
512 ms	50.74 ± 3.21	25.40 ± 2.61	30.80 ± 1.48	27.40 ± 1.52
1 s	51.38 ± 4.71	25.40 ± 2.51	31.40 ± 2.51	27.80 ± 2.49

Table 5. LR on MTAE-based classification.

LR	Acc (%)	Pr (%)	Re (%)	F1 (%)
0.01	44.15 ± 3.63	11.00 ± 0.71	25.00 ± 0.00	15.40 ± 0.89
0.001	75.96 ± 10.33	68.60 ± 20.50	67.20 ± 13.94	67.00 ± 17.59
0.0001	81.70 ± 3.44	80.20 ± 5.93	77.20 ± 3.56	78.00 ± 4.30
0.00001	71.28 ± 3.30	74.40 ± 9.21	55.60 ± 5.50	59.20 ± 7.33
0.000001	46.17 ± 5.30	18.40 ± 7.40	27.60 ± 3.21	21.40 ± 6.11

Table 6. Performance of different encoder–decoder configurations.

Encoder–Decoder	Acc (%)	Pr (%)	Re (%)	F1 (%)
1-1	88.51 ± 7.10	91.60 ± 5.13	89.00 ± 5.52	89.80 ± 5.72
2-2	89.47 ± 4.21	92.05 ± 3.15	88.72 ± 6.63	89.96 ± 5.07
3-3	81.70 ± 3.44	80.20 ± 5.93	77.20 ± 3.56	78.00 ± 4.30
4-4	72.87 ± 10.21	69.60 ± 14.84	63.80 ± 12.03	64.60 ± 12.28
5-5	71.17 ± 9.28	66.60 ± 9.66	65.60 ± 13.58	65.40 ± 11.63

Table 7. FC layer performance.

FC Layers	Acc (%)	Pr (%)	Re (%)	F1 (%)
1	85.84 ± 5.79	91.20 ± 2.77	84.60 ± 7.50	86.80 ± 5.36
2	87.55 ± 1.50	89.40 ± 2.88	85.80 ± 7.43	86.80 ± 4.44
3	89.47 ± 4.21	92.05 ± 3.15	88.72 ± 6.63	89.96 ± 5.07
4	86.70 ± 4.86	86.80 ± 5.12	84.20 ± 7.89	83.80 ± 9.18
5	86.38 ± 2.15	85.80 ± 3.63	82.20 ± 1.79	83.80 ± 1.48

5.2. *Results Based on MTAE-SVM Classification*

This subsection presents the results of the MTAE-SVM model. Since the γ value significantly affects SVM detection performance, identifying the optimal γ is essential.

Testing γ values from 0.1 to 30 in 0.1 increments, 0.1 emerged as optimal, offering the highest accuracy on the validation set of the trained hybrid model. The fully trained MTAE model with $\gamma = 0.1$ was then used on the test set because the distance between the MTAE-extracted features and the SVM hyperplane gave the best rate of correct classification. Figure 5 presents a comparison between the performance of the MTAE-SVM model and the MTAE model.

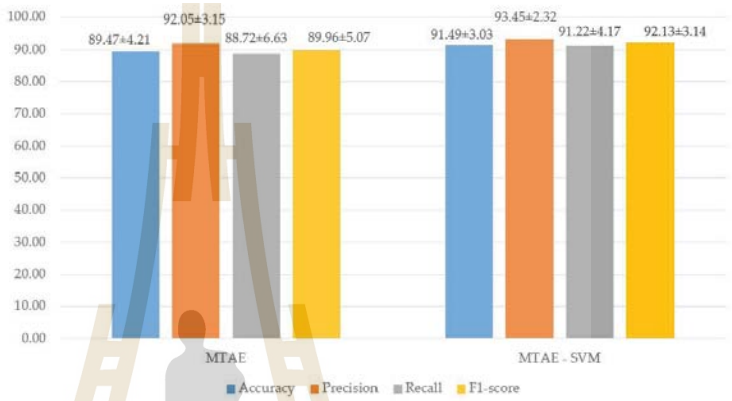


Figure 5. Comparison performance of the proposed MTAE classifier and MTAE-SVM classifier.

As shown in Figure 5, the hybrid model achieved enhanced accuracy, which can be attributed to its superior decision-making process. The increased distance between the automatically extracted features from the MTAE and the hyperplane of the trained SVM resulted in higher specificity compared with the MTAE classifier alone, as summarized in [28]. This directly contributed to improved detection accuracy, as evidenced by the greater detail in the MTAE-SVM confusion matrix compared with the MTAE and CNN methods, as illustrated in Figure 4a–c.

5.3. Comparison with Existing Approaches

In this subsection, we present a comprehensive comparison between our proposed feature extraction and classification technique and a range of established solutions. The evaluation is conducted by comparing the classification accuracy across four disease classes: normal (Acc_N), pneumonia (Acc_P), asthma (Acc_A), and COPD (Acc_C), as well as the Overall Accuracy (OA) of the system. The comparison of various feature extraction methods and classifiers for both the four-class and three-class scenarios is provided in Table 8 for the four-class case and in Table 9 for the three-class case.

Based on the information provided in Table 8, the proposed MTAE and MTAE-SVM achieve an OA of 89.47% and 91.49%, respectively, significantly outperforming the CNN 2L and EWT-FBP with LGBM, which attain OAs of 70.64% and 77.13%, respectively. These results highlight the superior capability of the MTAE and MTAE-SVM models in effectively capturing essential features for disease classification, leading to improved performance across multiple classes. The approach we propose demonstrates superior performance compared with alternative methods by utilizing the strengths of the MTAE's effective feature extraction alongside the strong classification abilities of the SVM. The MTAE model boosts feature learning through the simultaneous optimization of signal reconstruction and classification tasks, leading to the development of more distinct features. The incorporation of the SVM enhances classification performance by adeptly managing the representative features from the MTAE and optimizing the margin between classes, resulting in improved accuracy.

Table 8. Comparison with some systems for four classes.

Feature Extraction Method and Classifier	Acc_N (%)	Acc_P (%)	Acc_A (%)	Acc_C (%)	OA (%)
MFCC and CNN-2L, as in [31]	78.50	62.50	67.40	52.44	70.64
EWT-FBP and LGBM, as in [13]	-	-	-	-	77.13
MFCC and MTAE (ours)	92.77	83.33	86.83	89.22	89.47
MFCC and MTAE-SVM (ours)	93.73	90.91	89.64	89.22	91.49

Table 9. Comparison of feature extraction methods and classifiers for three classes.

Feature Extraction Method and Classifier	Acc_N (%)	Acc_P (%)	Acc_A (%)	OA (%)
MFCC and CNN-2L, as in [31]	85.30	40.01	64.79	72.31
MFCC and DNN-HMM, as in [24]	86.30	84.62	75.93	82.73
EWT-FBP and LGBM, as in [13]	77.58	100.00	76.72	84.76
MFCC and SVM, as in [24]	88.61	81.84	77.16	82.64
MFCC and KNN, as in [24]	76.67	72.22	66.67	72.05
MFCC and RF, as in [24]	80.28	79.27	59.26	73.96
Mel spectrogram and DNN, as in [24]	74.44	78.85	78.40	77.34
Mel spectrogram and LSTM, as in [24]	92.50	83.33	80.25	85.33
Mel spectrogram and GRU, as in [24]	92.22	81.41	78.70	83.94
Mel spectrogram & BiLSTM, as in [24]	94.44	84.62	79.63	86.28
Mel spectrogram & BiGRU, as in [24]	91.94	86.75	83.33	87.41
MFCC and MTAE (ours)	93.01	89.39	87.11	90.22
MFCC and MTAE-SVM (ours)	93.73	96.97	91.60	93.08

In the three-class classification task, as presented in Table 9, our proposed MTAE and MTAE-SVM systems achieved OAs of 90.22% and 93.08%, respectively, surpassing several well-known methods. The strong performance of the MTAE can be attributed to its multi-task learning approach, which effectively captures key MFCC features. Additionally, the hybrid methodology employed in the MTAE-SVM enhances performance, particularly in scenarios with limited training data, by fully leveraging the strengths of the MTAE. The impressive results of our models originate from their capacity to effectively learn complex patterns in LS signals and classify them with precision, a process further improved by the working together integration of the MTAE and SVM. This leads to enhanced precision in identifying PDs when compared with current methods.

6. Conclusions

In this paper, we introduced a novel MTAE model tailored specifically for the detection of PDs through the analysis of LS signals. The model employed a unique training approach that simultaneously enhances both classification accuracy and reconstruction quality. This model proposal has the potential to improve the efficiency and flexibility of audio-based medical diagnostics. The experimental findings demonstrate that our MTAE model outperforms CNN-2L and other existing systems in OA, achieving 89.47% accuracy in four-class disease classification and 90.22% accuracy in three-class classification. Furthermore, the MTAE-SVM could enhance the accuracy further. The accuracy could be improved to 91.49 % and 93.08 %, respectively. The remarkable outcome underscores the MTAE model's ability to capture the intricate and distinctive characteristics necessary for precise categorization of PDs. By fully using the strengths of the MTAE model, the hybrid approach of the MTAE-SVM also improves performance, especially when there are not many training data.

However, our study has certain limitations that need to be acknowledged:

- The model was evaluated using a specific dataset from KAUH, which may limit the generalizability of the results to other populations or recording conditions.
- The performance of the model relies on the availability of large numbers of labeled data for training, which can be challenging to obtain in medical settings.

- The model has not yet been implemented or validated in clinical settings, which is necessary to demonstrate its practical effectiveness in diagnosing PDs.

In future research direction, we plan to collaborate with healthcare institutions to test and refine the model in real-world scenarios, which will help demonstrate its potential in effectively diagnosing PDs. Additionally, we will explore the effectiveness of integrating phase feature extraction [43,44] and multi-scale CNNs [28] to improve our proposed MTAE methods. Furthermore, we will utilize score aggregation [45] to complement our suggested MTAE model with other classifiers.

Author Contributions: Conceptualization, K.O., K.P., W.P. (Watcharakorn Pinthurat), M.U. and P.U.; methodology, K.O.; Simulation, K.O.; validation, K.O., K.P., W.P. (Wongsathon Pathonsuwan), T.J., A.R., P.A., W.P. (Watcharakorn Pinthurat), M.U. and P.U.; visualization, K.O. and K.P.; formal analysis, K.O., K.P. and W.P. (Watcharakorn Pinthurat); writing—original draft, K.O.; writing—review and editing, K.O., K.P., W.P. (Wongsathon Pathonsuwan), T.J., A.R., P.A., W.P. (Watcharakorn Pinthurat), M.U. and P.U.; supervision, K.P., M.U. and P.U. All authors have read and agreed to the published version of the manuscript.

Funding: This research project is supported by Science Research and Innovation Fund. Agreement No. FF67/P1-044.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study were collected from King Abdullah University Hospital and are openly available in [Mendeley Data] at [https://doi.org/10.17632/jwyy9np4gv.3 accessed on 9 September 2024], reference number [30].

Acknowledgments: This research project is supported by Suranaree University of Technology (SUT), Thailand.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kazem, Z.; Zahra, M. Fundamentals of respiratory sounds and analysis. *Morgan Claypool Publ. Ser. Synth. Lect. Biomed. Eng.* **2006**, *8*, 19–21.
2. Gibson, G.J.; Loddenkemper, R.; Lundbäck, B.; Sibille, Y. Respiratory health and disease in Europe: The new European Lung White Book. *Eur. Respir. J.* **2013**, *42*, 559–563. [CrossRef] [PubMed]
3. Shi, L.; Zhang, Y.; Zhang, J. Lung sound recognition method based on wavelet feature enhancement and time-frequency synchronous modeling. *IEEE J. Biomed. Health Inform.* **2022**, *27*, 308–318. [CrossRef]
4. Shuvo, S.B.; Ali, S.N.; Swapnil, S.I.; Hasan, T.; Bhuiyan, M.I.H. A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWT-based hybrid scalogram. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 2595–2603. [CrossRef]
5. Abbasi, S.; Derakhshanfar, R.; Abbasi, A.; Sarbaz, Y. Classification of normal and abnormal lung sounds using neural network and support vector machines. In Proceedings of the 21st Iranian Conference on Electrical Engineering (ICEE), Mashhad, Iran, 14–16 May 2013; pp. 1–4.
6. Uysal, S.; Uysal, H.; Bolat, B.; Yıldırım, T. Classification of normal and abnormal lung sounds using wavelet coefficients. In Proceedings of the Signal Processing and Communications Applications Conference (SIU), Trabzon, Turkey, 23–25 April 2014; pp. 2138–2141.
7. Liu, Y.; Zhang, C.-M.; Zhao, Y.H.; Dong, L. The feature extraction and classification of lung sounds based on wavelet packet multiscale analysis. *Chin. J. -Comput.-Chin. Ed.* **2006**, *29*, 769.
8. Sengupta, N.; Sahidullah, M.; Saha, G. Lung sound classification using cepstral-based statistical features. *Comput. Biol. Med.* **2016**, *75*, 118–129. [CrossRef]
9. Jakovljević, N.; Lončar-Turukalo, T. Hidden Markov model-based respiratory sound classification. In Proceedings of the Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017, Thessaloniki, Greece, 18–21 November 2017; pp. 39–43.
10. Bahoura, M.; Pelletier, C. New parameters for respiratory sound classification. In Proceedings of the CCECE 2003-Canadian Conference on Electrical and Computer Engineering. Toward a Caring and Humane Technology, Montreal, QC, Canada, 4–7 May 2003; Volume 3, pp. 1457–1460.
11. Jung, S.-Y.; Liao, C.-H.; Wu, Y.-S.; Yuan, S.-M.; Sun, C.-T. Efficiently classifying lung sounds through depthwise separable CNN models with fused STFT and MFCC features. *Diagnostics* **2021**, *11*, 732. [CrossRef]

12. Azmy, M.M. Classification of lung sounds based on linear prediction cepstral coefficients and support vector machine. In Proceedings of the IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), Amman, Jordan, 3–5 November 2015; pp. 1–5.
13. Tripathy, R.K.; Das, D.; Ghose, A.; Ghosh, A. Automated detection of pulmonary diseases from lung sound signals using fixed-boundary-based empirical wavelet transform. *IEEE Sens. Lett.* **2022**, *6*, 1–4. [CrossRef]
14. Prajapati, S.K.; Choudhary, T.S.; Mishra, S. Early Detection of Lung Disease Using Multi-Class Classifiers. In Proceedings of the 2023 IEEE 4th Annual Flagship India Council International Subsections Conference (INDISCON), Mysore, India, 5–7 August 2023; pp. 1–6.
15. Reynolds, D.A.; Rose, R.C. Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Trans. Audio Speech Lang. Process.* **1995**, *3*, 72–83. [CrossRef]
16. Haider, A.; Ashraf, M.D.; Azhar, M.U.; Maruf, S.O.; Naqvi, M.; Khawaja, S.G.; Akram, M.U. Separation and classification of crackles and bronchial breath sounds from normal breath sounds using Gaussian mixture model. In Proceedings of the Neural Information Processing: 21st International Conference, Kuching, Malaysia, 3–6 November 2014; pp. 495–502.
17. Fraiwan, L.; Hassanin, O.; Fraiwan, M.; Khassawneh, B.; Ibnian, A.M.; Alkhodari, M. Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers. *Biocybern. Biomed. Eng.* **2021**, *41*, 1–14. [CrossRef]
18. Chambres, G.; Hanna, P.; Desainte-Catherine, M. Automatic detection of patients with respiratory diseases using lung sound analysis. In Proceedings of the International Conference on Content-Based Multimedia Indexing (CBMI), La Rochelle, France, 4–6 September 2018; pp. 1–6.
19. Jaber, M.M.; Abd, S.K.; Shakeel, P.M.; Burhanuddin, M.A.; Mohammed, M.A.; Yussof, S. A telemedicine tool framework for lung sounds classification using ensemble classifier algorithms. *Measurement* **2020**, *162*, 107883. [CrossRef]
20. García-Ordás, M.T.; Benítez-Andrades, J.A.; García-Rodríguez, I.; Benavides, C.; Alaiz-Moretón, H. Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalanced data. *Sensors* **2020**, *20*, 1214. [CrossRef] [PubMed]
21. Wanasinghe, T.; Bandara, S.; Madusanka, S.; Meedeniya, D.; Bandara, M.; de la Torre Díez, I. Lung Sound Classification With Multi-Feature Integration Utilizing Lightweight CNN Model. *IEEE Access* **2024**, *12*, 21262–21276. [CrossRef]
22. Bardou, D.; Zhang, K.; Ahmad, S.M. Lung sounds classification using convolutional neural networks. *Artif. Intell. Med.* **2018**, *88*, 58–69. [CrossRef]
23. Li, L.; Xu, W.; Hong, Q.; Tong, F.; Wu, J. Classification between normal and adventitious lung sounds using deep neural network. In Proceedings of the International Symposium on Chinese Spoken Language Processing (ISCSLP), Tianjin, China, 17–20 October 2016; pp. 1–5.
24. Shi, L.; Du, K.; Zhang, C.; Ma, H.; Yan, W. Lung sound recognition algorithm based on VGGish-BIGRU. *IEEE Access* **2019**, *7*, 139438–139449. [CrossRef]
25. Ditthapron, A.; Banluesombatkul, N.; Ketrat, S.; Chuangsuwanich, E.; Wilaiprasitporn, T. Universal joint feature extraction for P300 EEG classification using multi-task autoencoder. *IEEE Access* **2019**, *7*, 68415–68428. [CrossRef]
26. Autthasan, P.; Chaisaen, R.; Sudhawiyangkul, T.; Rangpong, P.; Kiatthaveephong, S.; Dilokthanakul, N.; Bhakdisongkhram, G.; Phan, H.; Guan, C.; Wilaiprasitporn, T. MIN2Net: End-to-end multi-task learning for subject-independent motor imagery EEG classification. *IEEE Trans. Biomed. Eng.* **2021**, *69*, 2105–2118. [CrossRef]
27. Bashir-Gonbadi, F.; Khotanlou, H. Brain tumor classification using deep convolutional autoencoder-based neural network: Multi-task approach. *Multimed. Tools Appl.* **2021**, *80*, 19909–19929. [CrossRef]
28. Pathonsuwan, W.; Phapatanaburi, K.; Buayai, P.; Jumphoo, T.; Anchuen, P.; Uthansakul, M.; Uthansakul, P. RS-MSConvNet: A novel end-to-end pathological voice detection model. *IEEE Access* **2022**, *10*, 120450–120461. [CrossRef]
29. Niu, X.-X.; Suen, C.Y. A novel hybrid CNN-SVM classifier for recognizing handwritten digits. *Pattern Recognit.* **2012**, *45*, 1318–1325. [CrossRef]
30. Fraiwan, M.; Fraiwan, L.; Khassawneh, B.; Ibnian, A. A dataset of lung sounds recorded from the chest wall using an electronic stethoscope. *Data Br.* **2021**, *35*, 106913. [CrossRef] [PubMed]
31. Aykanat, M.; Kılıç, Ö.; Kurt, B.; Saryal, S. Classification of lung sounds using convolutional neural networks. *Eurasip J. Image Video Process.* **2017**, *2017*, 1–9. [CrossRef]
32. Nukitram, J.; Saengmolee, W.; Chaisaen, R.; Autthasan, P.; Sengnon, N.; Wungsintaweekul, J.; Cheaha, D.; Kumarnsit, E.; Sudhawiyangkul, T.; Wilaiprasitporn, T. ANet: Autoencoder-Based Local Field Potential Feature Extractor for Evaluating an Antidepressant Effect in Mice After Administering Kratom Leaf Extracts. *IEEE Trans. Biomed. Circuits Syst.* **2023**, *17*, 67–76. [CrossRef]
33. Das, B.B.; Ram, S.K.; Babu, K.S.; Mohapatra, R.K.; Mohanty, S.P. Person Identification Using Autoencoder-CNN Approach with Multitask-Based EEG Biometric. *Multimed. Tools Appl.* **2024**, *83*, 83205–83225.
34. Li, Y.; Gong, S.; Liddell, H. Support vector regression and classification based multi-view face detection and recognition. In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 28–30 March 2000; pp. 300–305.
35. Toledo-Pérez, D.C.; Rodríguez-Reséndiz, J.; Gómez-Loenzo, R.A.; Jauregui-Correa, J.C. Support vector machine-based EMG signal classification techniques: A review. *Appl. Sci.* **2019**, *9*, 4402. [CrossRef]

36. Ismail, A.; Abdlerazek, S.; El-Henawy, I.M. Development of smart healthcare system based on speech recognition using support vector machine and dynamic time warping. *Sustainability* **2020**, *12*, 2403. [CrossRef]
37. Khairandish, M.O.; Sharma, M.; Jain, V.; Chatterjee, J.M.; Jhanjhi, N.Z. A hybrid CNN-SVM threshold segmentation approach for tumor detection and classification of MRI brain images. *Irbm* **2022**, *43*, 290–299. [CrossRef]
38. Liu, G.; Wang, L.; Liu, D.; Fei, L.; Yang, J. Hyperspectral image classification based on non-parallel support vector machine. *Remote Sens.* **2022**, *14*, 2447. [CrossRef]
39. Lin, C.-C.; Chen, S.-H.; Truong, T.-K.; Chang, Y. Audio classification and categorization based on wavelets and support vector machine. *IEEE Trans. Speech Audio Process.* **2005**, *13*, 644–651.
40. Sengupta, N.; Sahidullah, M.; Saha, G. Optimization of cepstral features for robust lung sound classification. In Proceedings of the IEEE India Conference (INDICON), New Delhi, India, 17–20 December 2015; pp. 1–6.
41. Thuwajit, P.; Rangpong, P.; Sawangjai, P.; Autthasan, P.; Chaisaen, R.; Banluesombatkul, N.; Boonchit, P.; Tatsaringkansakul, N.; Sudhawiyangkul, T.; Wilaiprasitporn, T. EEGWaveNet: Multiscale CNN-based spatiotemporal feature extraction for EEG seizure detection. *IEEE Trans. Ind. Inform.* **2021**, *18*, 5547–5557. [CrossRef]
42. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
43. Phapatanaburi, K.; Pathonsuwan, W.; Wang, L.; Anchuen, P.; Jumphoo, T.; Buayai, P.; Uthansakul, M.; Uthansakul, P. Whispered speech detection using glottal flow-based features. *Symmetry* **2022**, *14*, 777. [CrossRef]
44. Phapatanaburi, K.; Wang, L.; Liu, M.; Nakagawa, S.; Jumphoo, M.T.; Uthansakul, M. Significance of relative phase features for shouted and normal speech classification. *EURASIP J. Audio Speech Music Process.* **2024**, *1*, 2. [CrossRef]
45. Phapatanaburi, K.; Buayai, P.; Kupimai, M.; Yodrot, T. Linear prediction residual-based constant-Q cepstral coefficients for replay attack detection. In Proceedings of the International Electrical Engineering Congress (iEECON), Chiang Mai, Thailand, 4–6 March 2020; pp. 1–4.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

มหาวิทยาลัยเทคโนโลยีสุรนารี

ประวัติผู้เขียน

นางสาวขวัญจิต ออกเวหา เกิดที่จังหวัดราชบุรี ประเทศไทย ในปี พ.ศ. 2523 สำเร็จการศึกษาระดับปริญญาตรี สาขาวิศวกรรมไฟฟ้า (แขนงวิศวกรรมโทรคมนาคม) ในปี พ.ศ. 2546 จากสถาบันเทคโนโลยีราชมงคล และระดับปริญญาโท สาขาวิศวกรรมไฟฟ้า พ.ศ. 2556 จากมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี ทำงานเป็นพนักงานในสถาบันอุดมศึกษา สายวิชาการ ตำแหน่ง ผู้ช่วยศาสตราจารย์ คณะวิศวกรรมศาสตร์บูรณาการและเทคโนโลยี ณ มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก วิทยาเขตจันทบุรี จังหวัดจันทบุรี ปัจจุบันกำลังศึกษาระดับปริญญาเอก สาขาวิศวกรรมโทรคมนาคมและคอมพิวเตอร์ ณ มหาวิทยาลัยเทคโนโลยีสุรนารี จังหวัดนครราชสีมา ประเทศไทย งานวิจัยที่สนใจด้านปัญญาประดิษฐ์และการประมวลผลสัญญาณ การเรียนรู้ของเครื่อง การเรียนรู้เชิงลึก และสิ่งประดิษฐ์นวัตกรรมที่เกี่ยวข้องด้านการเกษตร

มหาวิทยาลัยเทคโนโลยีสุรนารี