

การประยุกต์ใช้การเรียนรู้แบบถ่ายโอนเชิงลึกในการจำแนกประเภทเสียงทางเดินหายใจในเด็ก

Deep Transfer Learning for Pediatric Respiratory Sound Classification

สพล ศรีหล่มสัก (Sapon Srilomsak)* ปรัชญา เปรมปราณีรัชต์ (Pradya Prempraneerach)^{1**}

สุภาวดี ทับกล่ำ (Supawadee Tubglam)*** คุณากร อนุกิตติพานิชย์ (Khunakon Anuwatpanich)^{****}

(Received: January 18, 2024; Revised: March 27, 2024; Accepted: April 2, 2024)

บทคัดย่อ

การติดเชื้อในระบบทางเดินหายใจส่วนล่างในเด็ก (LRTIs) เป็นปัญหาสุขภาพที่ก่อให้เกิดการเจ็บป่วยและเสียชีวิตในเด็กทั่วโลก สาเหตุของการติดเชื้ออาจเกิดจากเชื้อแบคทีเรียหรือไวรัสหลายชนิด ส่งผลให้เกิดลักษณะเสียงหายใจที่แตกต่างกันออกไป การวินิจฉัย LRTIs ในเด็กทารกมีความซับซ้อนเนื่องจากปอดและหัวใจของเด็กอยู่ในตำแหน่งที่ใกล้กัน การศึกษานี้มุ่งพัฒนาระบบปัญญาประดิษฐ์ (AI) สำหรับการจำแนกเสียงหายใจของเด็กทารกโดยใช้เทคนิค Deep Transfer Learning (DTL) ด้วยโมเดล DTL จำนวน 5 ชนิด คือ VGG-16, VGG-19, EfficientNet B0, EfficientNet B7 และ MobileNet ที่นำมาประยุกต์ใช้จำแนกภาพเสียงการหายใจในโดเมนความถี่และเวลา จำนวน 5 ประเภท คือ เสียงการหายใจแบบ Normal, Crackle, Rhonchi, Stridor และ Wheezing ที่วัดเสียงได้จากตุ๊กตาเด็กเสมือนจริง ด้วยหูฟังแพทย์แบบดิจิทัล ผลการวิจัยพบว่าโมเดล MobileNet สามารถจำแนกเสียงการหายใจได้อย่างแม่นยำที่สุด โดยมีความแม่นยำ (accuracy) สูงกว่า 80% เมื่อเทียบกับโมเดล DTL อีก 4 ชนิด ฉะนั้นโมเดล MobileNet มีศักยภาพในการนำมาประยุกต์ใช้สำหรับช่วยบุคลากรทางการแพทย์ ในการวินิจฉัยโรค LRTIs ในเด็กทารก และ ช่วยให้ผู้ป่วยเด็กได้รับการรักษาที่เหมาะสมและทันเวลาที่

ABSTRACT

Lower respiratory tract infections (LRTIs) in infants are a major global health problem that can lead to illness and death. The cause of infections can be from a variety of bacteria or viruses, resulting in different breathing sounds. Diagnosis of LRTIs in infants is complex because lungs and heart are close together. This study emphasized on a development of Artificial Intelligence (AI) systems for classifying infant breathing sounds using Deep Transfer Learning (DTL) techniques. Five DTL models: VGG-16, VGG-19, EfficientNet B0, EfficientNet B7, and MobileNet were applied to classify respiratory-sound images in both frequency and time domains for 5 different types of respiratory sounds: Normal, Crackle, Rhonchi, Stridor and Wheezing. MobileNet model achieves the highest accuracy of more than 80%, comparing to the other 4 DTL models. Thus, MobileNet model has a strong potential to be used for assisting medical personnel to accurate diagnosis of LRTIs in infants and enabling appropriate and timely treatment.

คำสำคัญ: การเรียนรู้แบบถ่ายโอนเชิงลึก จำแนกเสียงหายใจเด็ก วิเคราะห์ภาพสเปกโตรแกรม

Keywords: Deep transfer learning, Classifying pediatric respiratory sound, Spectrogram image analysis

¹Corresponding author: ppradya@engr.tu.ac.th

*นักศึกษา หลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมเครื่องกล คณะวิศวกรรมศาสตร์ มหาวิทยาลัยธรรมศาสตร์

Students, Master of Engineering Program, Department of Mechanical Engineering, Faculty of Engineering, Thammasat University

**ผู้ช่วยศาสตราจารย์ ภาควิชาวิศวกรรมเครื่องกล คณะวิศวกรรมศาสตร์ มหาวิทยาลัยธรรมศาสตร์

Assistant Professor, Department of Mechanical Engineering, Faculty of Engineering, Thammasat University

***อาจารย์ ภาควิชาการพยาบาลกุมารเวชศาสตร์ คณะพยาบาลศาสตร์ มหาวิทยาลัยธรรมศาสตร์

Lecturer, Department of Pediatric Nursing, Faculty of Nursing, Thammasat University

****อาจารย์ คณะครุศาสตร์อุตสาหกรรม มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

Lecturer, Faculty of Industrial Education, Rajamangala University of Technology Thanyaburi

บทนำ

การติดเชื้อในระบบทางเดินหายใจส่วนล่าง (LRTIs) เป็นกลุ่มการติดเชื้อที่เกิดจากจุลินทรีย์ชนิดต่างๆ เช่น แบคทีเรีย ไวรัส และเชื้อรา ซึ่งการติดเชื้อก่อให้เกิดโรคหลายประเภท เช่น โรคปอดบวม หลอดลมอักเสบเฉียบพลัน ไข้หวัดใหญ่ โรคไอกรน โรคซิสติกไฟโบรซิส ในปัจจุบันมีการพบการติดเชื้อไวรัสร่วมกับการติดเชื้อแบคทีเรีย ดังการศึกษาของ Tchatchouang, S [1] ซึ่งส่งผลกระทบต่อการทำงานของ Mucociliary clearance mechanism (MSM) หรือกระบวนการของร่างกายในการขับอนุภาคหรือเชื้อโรครายในระบบทางเดินหายใจ ด้วยการกลืนลงกระเพาะอาหารหรือการไอ ทำให้มีประสิทธิภาพลดลง [2-3] และเมื่อการติดเชื้อนี้เกิดขึ้นในเด็กวัยแรกเกิดไปจนถึงวัยเด็กตอนต้น อาจส่งผลให้เกิดการบีบตัวของทางเดินหายใจและการอักเสบในสภาวะความดันสูงสะสมระหว่างการไอ ซึ่งนำไปสู่การไออย่างต่อเนื่องจะส่งผลเสียต่อคุณภาพการนอนหลับ ความทุกข์ และความเหนื่อยล้าของเด็ก และยังสามารถส่งผลให้อาการ LRTIs แย่ลงได้ [4]

การติดเชื้อในระบบทางเดินหายใจส่วนล่าง (LRTIs) เป็นโรคที่พบบ่อยและก่อให้เกิดการเสียชีวิตทั่วโลก โดยศูนย์ควบคุมและป้องกันโรค (CDC) ประเมินการว่า เพียงแค่เฉพาะในสหรัฐอเมริกา LRTIs ก็เป็นสาเหตุที่ทำให้ผู้ป่วยต้องเข้าโรงพยาบาลมากกว่า 1.6 ล้านรายและเสียชีวิตมากกว่า 29,000 รายต่อปี [5] รวมถึงยังเป็นหนึ่งในโรคที่พบบ่อยที่สุดในทารกและเด็กซึ่งมีอัตราการเสียชีวิตที่สูง [6-7] โดยจากข้อมูลทางสถิติพบว่า ผู้ป่วยที่อายุน้อยกว่า 5 ปี เสียชีวิตด้วยโรคปอดอักเสบมากกว่า 2 ล้านคน และนอกจากนี้ ยังพบผู้ป่วยที่เป็น LRTIs มากกว่า 13.9 ล้านคนต่อปี และในจำนวนนั้นเป็นผู้ป่วยเด็กอายุต่ำกว่า 5 ปี และเสียชีวิตเป็นจำนวนกว่า 13.1% ของทั้งหมด นอกจากนี้ LRTIs แบบเฉียบพลันยังเป็นสาเหตุหลักอันดับที่ 4 ของภาวะการเจ็บป่วยและการเสียชีวิตของทารกทั่วโลกอีกด้วย [8-9] โดยทั่วไปการรักษา LRTIs นั้นมักจะใช้ยาต้านไวรัสและแบคทีเรียชนิดต่างๆ และยาสเตียรอยด์ รวมไปถึงการดูแลต่างๆ เช่น การพักผ่อนและการดื่มน้ำให้เพียงพอ แต่ในกรณีที่เป็นการติดเชื้อชนิดรุนแรง หรือเกิดภาวะแทรกซ้อนทำให้มีอาการรุนแรงมากขึ้น อาจจำเป็นต้องเข้ารับการรักษาในโรงพยาบาลเพื่อรับการรักษาอย่างเข้มข้น เช่น การให้ออกซิเจน การใส่ท่อช่วยหายใจ การเจาะคอ การเจาะระบายน้ำในเยื่อหุ้มปอด การทำกายภาพบำบัดทรวงอก รวมไปถึงการให้ยาปฏิชีวนะ [10]

การวินิจฉัยโรค LRTIs ในปัจจุบันจะอาศัยการฟังเสียงหายใจจากปอดของผู้ป่วยด้วยหูฟังแพทย์ หรือ Stethoscope ซึ่งผลการวินิจฉัยอาจไม่แม่นยำเสมอไป เนื่องจากจำนวนบุคลากรทางการแพทย์ผู้เชี่ยวชาญที่อาจไม่เพียงพอ ประกอบกับเชื้อก่อโรคบางชนิดอาจตรวจพบในคนที่ไม่มีสุขภาพดีโดยไม่มีอาการใดๆ หรืออาจพบในคนที่มีอาการเล็กน้อยซึ่งอาการแสดงมีความใกล้เคียงกับ LRTIs แต่ไม่ได้ติดเชื้อไวรัสหรือแบคทีเรียใดๆ ทำให้ในบางกรณีแพทย์อาจจ่ายยาปฏิชีวนะให้ผู้ป่วยโดยไม่จำเป็น ดังเช่นในงานวิจัยของ Kevin L. Schwartz [11] พบว่าการจ่ายยาปฏิชีวนะโดยไม่จำเป็นแก่ผู้ป่วยเด็กอายุต่ำกว่า 2 ปี เป็นอัตรา 17.6% ของผู้ป่วยที่รับยาปฏิชีวนะทั้งหมด ซึ่งอาจส่งผลให้เกิดความล่าช้าในการรักษาอย่างถูกต้องและจัดการดูแลอย่างทั่วถึง ทำให้อาจเกิดอันตรายต่อผู้ป่วย โดยเฉพาะในกรณีที่ผู้ป่วยเป็นเด็กเล็กที่มีอายุน้อยและมีภูมิคุ้มกันโรคอ่อนแอ ซึ่งมีความเสี่ยงต่อภาวะแทรกซ้อนจาก LRTIs อย่างมาก

ในปัจจุบันเทคนิคการเรียนรู้เชิงลึก (Deep Learning) ได้ถูกพัฒนาอย่างต่อเนื่องและได้มีการนำไปประยุกต์ใช้งานทางวิศวกรรมหลากหลายสาขา เช่น การประมวลผลภาษาธรรมชาติ (natural language processing) การจดจำภาพ (visual recognition) และ การวิเคราะห์ภาพทางการแพทย์ (medical image analysis) [12-13] เป็นต้น โดยทั่วไปแล้วการประเมินประสิทธิภาพของการเรียนรู้เชิงลึกจะต้องแลกเปลี่ยนความแม่นยำในการเรียนรู้กับความเร็วในการประมวลผลการเรียนรู้ อย่างไรก็ตาม ซึ่งจะขึ้นอยู่กับขนาดและโครงสร้างของโครงข่ายประสาทเทียม (Neural Network) ที่นำมาประยุกต์ใช้ ฉะนั้นในงานวิจัยนี้จะประยุกต์ใช้และเปรียบเทียบการเรียนรู้เชิงลึกหลายเทคนิค เพื่อใช้ในการจำแนกเสียงการหายใจทั้งปกติและผิดปกติในหลายหลากแบบที่เกิดจากโรค LRTIs

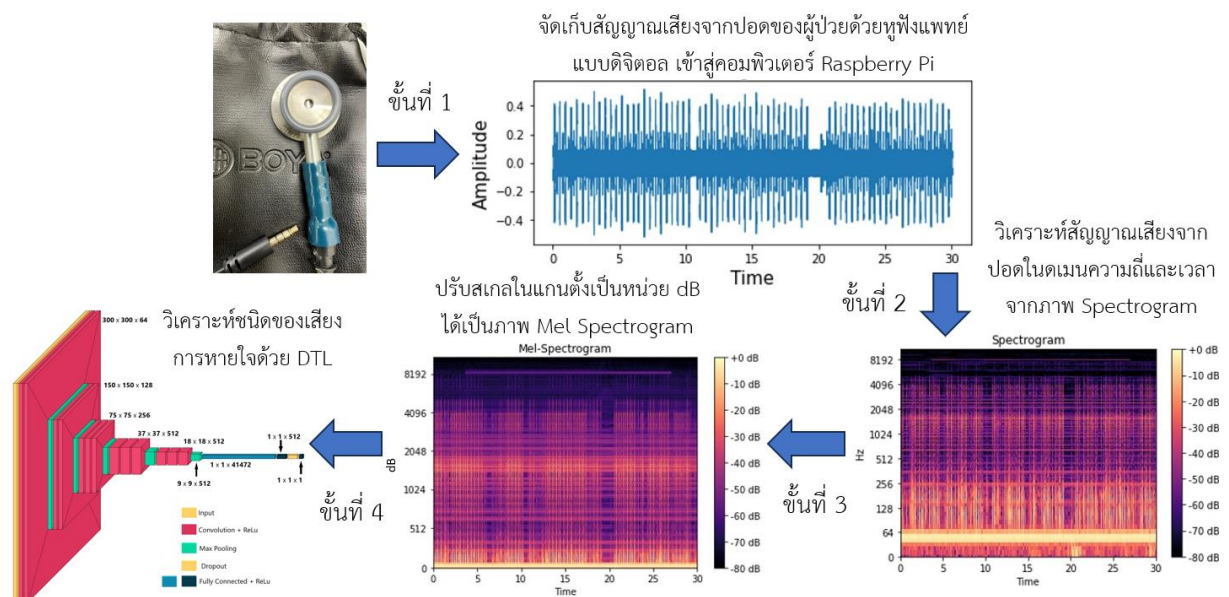
วัตถุประสงค์ของงานวิจัย

เพื่อแก้ไขปัญหาดังกล่าวข้างต้น ผู้วิจัยจึงได้พัฒนานวัตกรรมใหม่ในรูปแบบของการทำนายและจำแนกเสียงหายใจปกติและเสียงหายใจที่ผิดปกติในแบบต่างๆ ของผู้ป่วยเด็กที่อาจเป็นโรค LRTIs ด้วยเทคนิคการเรียนรู้แบบถ่ายโอนเชิงลึก (Deep Transfer Learning หรือ DTL) ที่พัฒนาจาก Convolution Neural Network (CNN) architecture [13]

โดยจุดเด่นของงานวิจัยนี้ คือการใช้ DTL ในการช่วยบุคลากรทางการแพทย์จำแนกเสียงหายใจของผู้ป่วยเด็กที่มีความผิดปกติในส่วนต่างๆ ของระบบทางเดินหายใจ ซึ่งจะช่วยลดความผิดพลาดในการวินิจฉัยของแพทย์และพยาบาลที่ต้องอาศัยประสบการณ์และความเชี่ยวชาญในการฟังเสียงการหายใจของผู้ป่วย ที่มีคุณลักษณะที่ใกล้เคียงกันมากในแต่ละโรค นอกจากนั้นเสียงรบกวนจากสิ่งแวดล้อมก็อาจจะทำให้การวินิจฉัยการฟังเสียงหายใจของผู้ป่วยผิดพลาดได้เช่นกัน [14] มากไปกว่านั้นแล้วนวัตกรรมนี้สามารถช่วยลดระยะเวลาในการวินิจฉัย ส่งผลให้ผู้ป่วยได้รับการวินิจฉัยและรักษาที่ถูกต้องอย่างทันท่วงที ซึ่งจะช่วยลดความเสี่ยงต่อการเกิดภาวะแทรกซ้อนและ ช่วยลดอัตราการเสียชีวิตของผู้ป่วย

ขั้นตอนการดำเนินการวิจัย

การสร้างโมเดลการเรียนรู้เชิงลึกด้วยเทคนิค CNN เพื่อใช้ในการวิเคราะห์เสียงการหายใจจากปอดของผู้ป่วยโดยใช้ภาพ Spectrogram ขั้นที่หนึ่งเริ่มจากการพัฒนาหูฟังแพทย์แบบดิจิทัล (Digital stethoscope) ในการจัดเก็บสัญญาณเสียงจากปอดของผู้ป่วยเข้าสู่เครื่องคอมพิวเตอร์ขนาดเล็ก Raspberry Pi อธิบายในหัวข้อที่ 3.1 ขั้นที่สองในหัวข้อที่ 3.2 และ 3.3 จะนำสัญญาณเสียงนั้นมาวิเคราะห์หาค่าประกอบความถี่ในช่วงเวลาสั้นๆ เพื่อให้ได้เป็นภาพสเปกโตรแกรม (Spectrogram) ของเสียงการหายใจ จากนั้นขั้นที่สามจึงนำภาพสเปกโตรแกรม (Spectrogram) นั้นมาปรับสเกลในแกนตั้งเป็นหน่วยเดซิเบลหรือที่เรียกว่า Mel Spectrogram แล้วในขั้นที่สี่ได้ทำการวิเคราะห์ชนิดของเสียงการหายใจ โดยโครงข่ายประสาทเทียมแบบการเรียนรู้เชิงลึกด้วยเทคนิค DTL เป็นเทคนิคที่ให้ความแม่นยำในการพยากรณ์ได้อย่างรวดเร็ว ดังอธิบายในหัวข้อที่ 3.4 และ 3.5 ขั้นตอนทั้งหมดของงานวิจัยนี้ในการวิเคราะห์และจำแนกเสียงการหายใจด้วยเทคนิค DTL ได้แสดงในแผนภาพขั้นตอนการเชื่อมโยงการดำเนินการ ดังในภาพที่ 1



ภาพที่ 1 แผนภาพแสดงขั้นตอนการดำเนินงานวิจัย ในการจำแนกเสียงการหายใจจากภาพด้วยเทคนิคการเรียนรู้เชิงลึก

1. การเตรียมการและการจัดเก็บข้อมูล

สำหรับการจัดเก็บข้อมูลเสียงการหายใจในการทดลองที่ใช้ในการทดสอบขั้นถัดไป ผู้วิจัยได้จัดทำหูฟังแพทย์แบบดิจิตอล (Digital Stethoscope) ดังแสดงในภาพที่ 2 (ก) เพื่อเก็บเสียงการหายใจของตุ๊กตาเด็กทารกเสมือนจริง เข้าสู่คอมพิวเตอร์ Raspberry Pi4 ขนาดเล็ก โดยขั้นตอนการสร้างหูฟังแพทย์แบบดิจิตอลจะทำการตัดต่อรับเสียงของ Stethoscope ยี่ห้อ Littmann รุ่น Classic II Infant ที่ตำแหน่งใกล้เคียงกับแผ่นรับเสียงให้มากที่สุด แล้วต่อไมโครโฟนยี่ห้อ Boya รุ่น BY-LM10 เข้ากับท่อรับเสียง จากนั้นต่อสายจากไมโครโฟนเข้าไปยังคอมพิวเตอร์ Raspberry Pi4 โดยหูฟังแพทย์ Littmann รุ่นที่นำมาใช้นี้มีน้ำหนักโดยรวมประมาณ 90 กรัม และแผ่นรับเสียงมีขนาดที่เล็กเป็นพิเศษ โดยด้าน Bell มีขนาดเส้นผ่านศูนย์กลางเพียง 1.9 เซนติเมตร และด้าน Diaphragm มีขนาดเส้นผ่านศูนย์กลางเพียง 2.7 เซนติเมตร

ส่วนการวัดเสียงหายใจทั้งแบบปกติและแบบไม่ปกติจะเก็บข้อมูลจากตุ๊กตาเด็กทารกเสมือนจริง CAE Luna Simulator เป็นเครื่องมือสำหรับฝึกอบรมการดูแลทารกแรกเกิดที่มีความเที่ยงตรงสูง สามารถใช้ฝึกอบรมบุคลากรทางการแพทย์ได้อย่างสมจริง คุณสมบัติหลักประการหนึ่งของตุ๊กตาเด็ก Luna คือความสามารถในการสร้างเสียงที่สมจริง โดยตุ๊กตาเด็ก Luna ใช้ลำโพงและเซ็นเซอร์ร่วมกันเพื่อสร้างเสียงในช่วงความถี่ที่หลากหลาย ทำให้สามารถสร้างเสียงทางสรีรวิทยาที่หลากหลายได้ รวมถึงเสียงหัวใจ ปอด ลำไส้ และหลอดเลือด ตลอดจนเสียงหายใจที่ผิดปกติของเด็กทารก เช่น เสียงกรอบแกรบ (Crackle) เสียงหวีด (Wheezing) เสียงอืด (Rhonchi) และ เสียงอืด (Stridor) เป็นต้น เสียงการหายใจแบบปกติ (normal respiratory sound) นั้นจะมีความถี่อยู่ในช่วง 50-2,500 Hz ส่วนเสียงกรอบแกรบ (crackle sound) จะเป็นเสียงดังแบบไม่ต่อเนื่องที่เกิดขึ้นเป็นช่วงที่น้อยกว่า 20 ms ที่ประกอบไปด้วยความถี่หลักในช่วงระหว่าง 350 ถึง 650 Hz [15] ที่เชื่อมโยงกับการติดเชื้ในปอด (หรือโรคปอดบวม) และการผิดปกติจากการหายใจ (pulmonary fibrosis หรือการล้มเหลวของหัวใจ) ส่วนเสียงหวีด (wheezing sound) ที่เกิดขึ้นเป็นหน่วงๆ ที่มีคาบระหว่าง 50 และ 250 ms [16] และ เสียงอืด (rhonchi sound) ที่มีความถี่หลักประมาณ 150 Hz ซึ่งมีคาบมากกว่า 80 ms [17] และ เสียงอืด (stridor sound) ที่มีความถี่หลักประมาณ 400 Hz และมีความถี่สูงที่สุดมากกว่า 2000 Hz และมีคาบมากกว่า 250 ms [17-20]



(ก)

(ข)

ภาพที่ 2 (ก) การพัฒนาหูฟังแพทย์แบบดิจิตอล (Digital Stethoscope) สำหรับเก็บข้อมูลสัญญาณเสียงหายใจของผู้ป่วย
(ข) การเก็บเสียงหายใจจากปอดของตุ๊กตาเด็กเสมือนจริง CAE Luna

2. ชุดข้อมูลเสียงการหายใจสำหรับการทดลอง

ในการสอนโมเดลการเรียนรู้เชิงลึก ได้ทำการเก็บข้อมูลเสียงหายใจของตุ๊กตาเด็กทารก CAE Luna เป็นจำนวน 45 เสียง ซึ่งประกอบด้วยเสียงหายใจแบบปกติ 5 เสียง และเสียงหายใจแบบผิดปกติ 40 เสียง โดยแบ่งออกเป็นเสียงหายใจที่ผิดปกติออกเป็นแบบ Crackle, Rhonchi, Stridor, Wheezing อย่างละ 5 เสียง และเสียงหายใจแบบผิดปกติที่มีเสียงรบกวนจากเสียงหัวใจเต้นแทรกอยู่ อย่างละ 5 เสียง ดังแสดงในตารางที่ 1

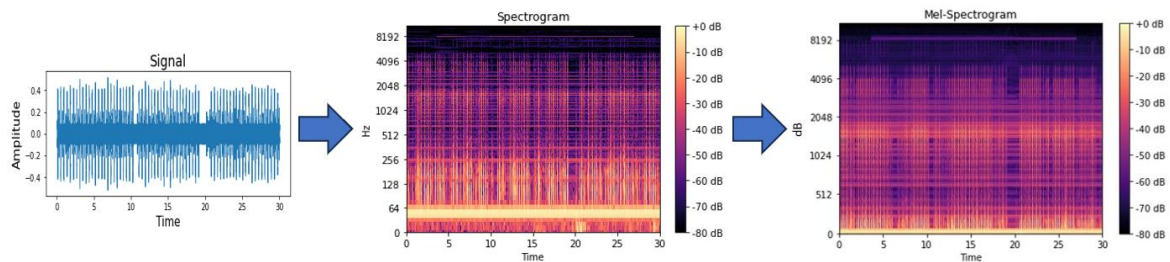
ตารางที่ 1 ชุดข้อมูลเสียงหายใจที่วัดด้วยหูฟังแพทย์แบบดิจิทัล จากตุ๊กตาเด็กทารก CAE Luna

ชุดข้อมูล (ชนิดของเสียงหายใจ)	เสียงหายใจ ปกติ	เสียงหายใจ ผิดปกติแบบ Crackle	เสียงหายใจ ผิดปกติแบบ Rhonchi	เสียงหายใจ ผิดปกติแบบ Stridor	เสียงหายใจ ผิดปกติแบบ Wheezing
เสียงหายใจที่ไม่มีเสียงการ เต้นหัวใจรบกวน	5	5	5	5	5
เสียงหายใจที่มีเสียงการ เต้นหัวใจรบกวน	-	5	5	5	5

3. ชุดข้อมูลภาพ Spectrogram ของเสียงการหายใจ

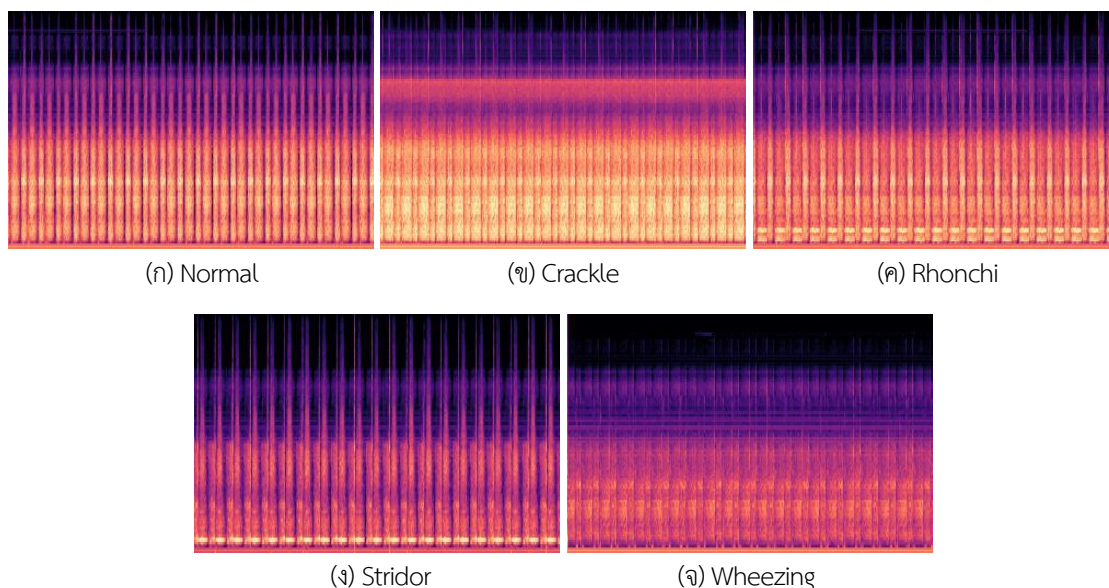
สัญญาณเสียงการหายใจที่วัดได้นั้น ประกอบไปด้วยคลื่นเสียงหลายความถี่ จึงต้องนำมาผ่านกระบวนการ แปลงฟูเรียร์แบบเร็ว (FFT) ซึ่งเป็นอัลกอริทึมที่ใช้อย่างแพร่หลายในการคำนวณการแปลงสัญญาณเสียงให้อยู่ในโดเมนความถี่ เพื่อประมวลผลสัญญาณ (Signal Processing) และจะช่วยให้สามารถแยกสัญญาณเสียงออกเป็นแอมพลิจูดและมุมเฟสที่แต่ละความถี่ แล้วนำมาคำนวณเป็นผลลัพธ์กำลังของสัญญาณ ที่เรียกว่าสเปกตรัม (Spectrum) จะทำให้สามารถวิเคราะห์และจำแนกเสียงแต่ละประเภทตามแอมพลิจูดที่เกิดขึ้นที่ความถี่ต่างกันออกไปของสัญญาณที่วัดได้

ในการสร้างภาพสเปกโตรแกรมแบบสองมิติ นั้นสร้างได้โดยการแปลงแบบฟูเรียร์ในช่วงเวลาสั้น (short-time Fourier transform) ที่ใช้การแปลงฟูเรียร์แบบเร็ว (FFT) ที่คำนวณกำลังของสัญญาณหรือแอมพลิจูดของสัญญาณในช่วงสั้นที่ซ้อนเหลื่อมกัน (Overlapping Window) แล้วนำมาเรียงต่อกันในแนวตั้ง ที่แสดงด้วยภาพโดยมีแกนนอนเป็นเวลาช่วงสั้นและแกนตั้งเป็นความถี่ ทำให้ได้เป็นภาพ Spectrogram ในระนาบสองมิติและเพื่อให้ได้สัญญาณเสียงที่เน้นในย่านคลื่นเสียงต่ำ ซึ่งเป็นคลื่นเสียงของมนุษย์ และลดความสำคัญของเสียงอื่นๆ ที่ไม่ใช่ของเสียงมนุษย์ให้เป็นเสียงสัญญาณรบกวน (หรือ Noise) จะใช้การปรับสเกลแบบ Mel หรือการปรับสเกลแกนตั้งให้เป็นหน่วยเดซิเบลทำให้ได้ภาพ Mel Spectrogram ดังแสดงในภาพที่ 3



ภาพที่ 3 ขั้นตอนการแปลงสัญญาณเสียงให้เป็นภาพ Mel Spectrogram

และ ภาพที่ 4 แสดงตัวอย่างของภาพ Mel Spectrogram ของเสียงการหายใจที่ไม่มีเสียงการเต้นหัวใจรบกวนซึ่งวัดได้จากตุ๊กตาเสมือนจริง CAE Luna ที่ได้อธิบายในตารางที่ 1 ก่อนนำภาพ Mel Spectrogram ไปใช้สอนแอมโมเดล DTL ได้ทำการตัดขอบภาพสีขาวและแกนตั้งและแกนนอนของภาพออกไป



ภาพที่ 4 ภาพ Mel Spectrogram ของสัญญาณเสียงหายใจแบบต่างๆ ที่ไม่มีเสียงการเต้นหัวใจมารบกวน ซึ่งวัดได้จากตุ๊กตาเสมือนจริง CAE Luna

โดยทั่วไปแล้วในการเรียนรู้ด้วยโมเดลปัญญาประดิษฐ์แบบ CNN นั้น จำเป็นต้องมีข้อมูลจำนวนมาก ในการสอนโมเดล CNN เพื่อให้เรียนรู้ข้อมูลได้อย่างถูกต้องและมีความแม่นยำ จึงควรจะมีการเพิ่มจำนวนข้อมูลให้มากขึ้น โดยเฉพาะอย่างยิ่งเมื่อมีข้อมูลภาพ Mel Spectrogram ของเสียงการหายใจที่จะใช้สอนโมเดล DTL น้อยเกินไป จนอาจทำให้การเรียนรู้เกิดความผิดพลาดได้ จึงจำเป็นที่จะต้องนำภาพ Mel Spectrogram ที่ได้มาผ่านกระบวนการ Data Augmentation โดยจะมีวิธีการในการแปลงรูปภาพให้เพิ่มจำนวนมากขึ้น เช่น การหมุน การซูม การเบลอภาพ การเลื่อนรูปภาพไปทางซ้ายหรือขวา การสลับซ้าย-ขวา ของภาพ เป็นต้น ในงานวิจัยนี้ได้เพิ่มจำนวนภาพ Mel Spectrogram โดยการสลับภาพซ้าย-ขวา และเลื่อนภาพซ้าย-ขวา ด้วยเทคนิค Data Augmentation เพื่อให้คงคุณสมบัติของย่านความถี่ โดยผู้วิจัยได้แบ่งข้อมูลภาพ Mel Spectrogram ทั้งหมดออกเป็น 2 ชุด โดยชุดที่ 1 คือชุดข้อมูลสำหรับการทดสอบโดยใช้ภาพเสียงหายใจที่วัดได้จากการทดลอง โดยแบ่งเป็นภาพเสียงหายใจ 5 ประเภท ประเภทละ 1 ภาพ รวมทั้งหมด 5 ภาพ และ ชุดข้อมูลที่ 2 สำหรับใช้สอนแก่เทคนิค DTL เป็นภาพเสียงหายใจที่วัดได้ที่เหลืออีก จำนวน 40 ภาพ โดยจะนำภาพในชุดที่ 2 ที่ผ่านการทำ Data Augmentation เพื่อเพิ่มจำนวนของข้อมูลภาพให้เป็น 320 ภาพ ที่ใช้สอนและทดสอบแก่เทคนิค DTL โดยจะแบ่งชุดข้อมูลภาพเสียงหายใจชุดข้อมูลที่ 2 ออกเป็น 80% (หรือ 258 ภาพ) และ 20% (หรือ 62 ภาพ) สำหรับการสอน (train) และทดสอบ (test หรือ validate) ความถูกต้องของเทคนิค DTL ตามลำดับ

4. โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolution Neural Network)

โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolution Neural Network หรือ CNN) เป็นเทคนิคพื้นฐานที่สำคัญของการเรียนรู้แบบเชิงลึก (Deep Learning) ที่ถูกนำมาใช้ในการแก้ปัญหการจัดหมวดหมู่ (Classification) และ จำแนกภาพที่มีประสิทธิภาพใกล้เคียงกับมนุษย์ โดยจะจำลองการมองเห็นของมนุษย์ที่มองภาพเป็นพื้นที่ย่อยๆ โดยใช้ตัวกรอง (Filter) หรือ เคอร์เนล (kernel) ที่จะใช้ในการคำนวณและดึงคุณลักษณะเฉพาะของวัตถุออกมาจากภาพ โดยปกติแล้วตัวกรองอันหนึ่งจะดึงคุณลักษณะเฉพาะออกมาได้เพียงหนึ่งอย่าง เช่น เส้นขอบ จุดต่าง สี ขนาด หรือรูปร่าง เป็นต้น แล้วนำข้อมูลคุณลักษณะเด่นมาผสมกัน เพื่อจำแนกว่าวัตถุที่อยู่ในรูปภาพนั้นมีความคล้ายกับอะไร ด้วยความน่าจะเป็นมากน้อยเพียงใด

โดยทั่วไปโครงข่ายประสาทเทียมแบบคอนโวลูชัน จะประกอบไปด้วยชั้น (Layer) ต่างๆ คือ 1) ชั้นคอนโวลูชัน (Convolution Layer) จะใช้ตัวกรองที่เป็นเมทริกซ์สองมิติที่มีขนาดตามที่ต้องการพิจารณา ในการคำนวณหาลักษณะเฉพาะต่างๆ ของรูปภาพนั้น โดยเมทริกซ์ที่กำหนดจะถูกนำมาคอนโวลูต์กับรูปภาพแบบสองมิติ ทำให้ได้ผลลัพธ์เป็น feature map ที่แสดงถึงคุณลักษณะเด่นต่างๆ ของวัตถุในภาพ 2) ชั้นการรวม (Pooling Layer) เป็นเทคนิคในการย่อเมทริกซ์ลงด้วยตัวกรองเพื่อดึงเอาส่วนที่สำคัญที่สุดของภาพ ที่ได้จากชั้นคอนโวลูชันออกมาและช่วยในการลดความแปรปรวนและความซับซ้อนในการคำนวณ โดยชั้นการรวมที่ใช้กันอย่างแพร่หลาย มี 2 ประเภท คือ Max Pooling ที่จะเลือกข้อมูลภายในตัวกรองที่มีค่ามากที่สุด และ Average Pooling ที่จะหาค่าเฉลี่ยของข้อมูลภายในตัวกรอง และ 3) ชั้นเชื่อมโยงแบบสมบูรณ์ (Fully-connected Layer) ซึ่งเป็นขั้นตอนจำแนกข้อมูลจากกระบวนการต่างๆ ก่อนหน้าผ่านชั้นของ neuron ที่เชื่อมโยงกันเป็นโครงข่ายประสาทเทียมที่ใช้จำแนกหมวดหมู่ของข้อมูล โดยการเรียนรู้แบบ back-propagation ที่ต้องทำการสอนซ้ำหลายๆ ครั้ง

5. การเรียนรู้แบบถ่ายโอนเชิงลึก (Deep Transfer Learning)

การเรียนรู้แบบถ่ายโอนเชิงลึก นั้นพัฒนาจากโมเดลโครงข่ายประสาทเทียมแบบคอนโวลูชันแบบเชิงลึก ที่ได้พัฒนาโครงสร้างและได้รับการฝึกสอนมาก่อนด้วยชุดข้อมูลขนาดใหญ่ ให้สามารถนำไปประยุกต์ใช้ในการเรียนรู้ชุดข้อมูลใหม่ที่ขนาดเล็กกว่าและมีความคล้ายกับชุดข้อมูลเดิมที่ได้ใช้ในการฝึกสอนแล้ว [21] การฝึกสอนการเรียนรู้แบบถ่ายโอนเชิงลึกนั้นสามารถให้ผลการเรียนรู้ที่แม่นยำได้ดียิ่งขึ้น เมื่อใช้ร่วมกับกระบวนการที่เรียกว่า fine tuning เพื่อเพิ่มประสิทธิภาพการเรียนรู้ [22] ซึ่งจะช่วยให้อัตราการเรียนรู้ของคุณลักษณะเฉพาะของชุดข้อมูลใหม่ที่อาจไม่มีอยู่ในชุดข้อมูลดั้งเดิมของการฝึกสอนโมเดลมาก่อน [23-24] โดยผู้วิจัยได้เลือกใช้โมเดลการเรียนรู้แบบถ่ายโอนเชิงลึก 5 โมเดล ในงานวิจัยนี้ คือ 1) EfficientNet B0, 2) EfficientNet B7, 3) VGG-16, 4) VGG-19 และ 5) MobileNet ที่เขียนด้วยภาษา Python

โมเดล VGG-16 เป็นโมเดลโครงข่ายประสาทเทียมที่ถูกสร้างขึ้นโดยมีพื้นฐานจาก CNN โดยมีชั้นคอนโวลูชันแบบสองมิติ จำนวน 16 ชั้นซ้อนกัน ซึ่งถูกนำมาใช้กันอย่างแพร่หลายในการจำแนกภาพ ต่อมาโมเดล VGG-19 ได้ถูกพัฒนาต่อยอดจากโมเดล VGG-16 โดยได้เพิ่มจำนวนชั้นคอนโวลูชันแบบสองมิติ เป็นจำนวน 19 ชั้น [25] โดยทั้งสองโมเดลจะตามด้วยชั้นเชื่อมโยงแบบสมบูรณ์ จำนวน 3 ชั้นเหมือนกัน การเรียนรู้แบบถ่ายโอนด้วยโมเดล VGG-16 และ VGG-19 มีการเรียนรู้จากชุดข้อมูล ImageNet Dataset ที่มีขนาดใหญ่ [12]

โมเดล EfficientNet เป็นโมเดลโครงข่ายประสาทเทียมที่พัฒนาโดย Google AI โดยมีพื้นฐานจาก CNN ที่ได้รับการออกแบบมาเพื่อเป็นโมเดลโครงข่ายประสาทเทียมขั้นสูงสำหรับงานจำแนกรูปภาพ ซึ่งมุ่งเน้นทั้งความแม่นยำและประสิทธิภาพ โดยลดขนาดของโมเดลและจำนวนการคำนวณ แต่ยังรักษาคุณภาพของโมเดลไว้ โดยอ้างอิงจากการวัด FLOPS (floating-point operations เป็นหน่วยวัดประสิทธิภาพในการทำงานของโมเดล) โครงสร้างของ EfficientNets ได้รับการออกแบบโดยใช้เทคนิค Neural Architecture Search ซึ่งเป็นการค้นหาโครงสร้างเครือข่ายประสาทเทียมโดยอัตโนมัติ ทำให้ EfficientNet เหมาะสมทั้งในด้านประสิทธิภาพและความแม่นยำ โดย EfficientNet มีการเรียนรู้จาก ImageNet และใช้ประโยชน์จากหลักการต่าง ๆ เช่น การปรับขนาดฟีเจอร์ (feature scaling) และการปรับขนาดอัตราการเรียนรู้ (Learning rate scaling) เพื่อปรับปรุงประสิทธิภาพของโมเดล ซึ่งในงานวิจัยนี้จะประยุกต์ใช้ทั้งโมเดล EfficientNet B0 และ EfficientNet B7 โดย EfficientNet B0 เป็นโมเดลขนาดเล็กที่มีประสิทธิภาพเพียงพอสำหรับงานจำแนกภาพหลายประเภทมีชั้นการเรียนรู้ 86 ชั้น ในขณะที่ EfficientNet B7 เป็นโมเดลขนาดใหญ่ที่มีประสิทธิภาพสูงสำหรับงานจำแนกภาพที่มีความซับซ้อนมีชั้นการเรียนรู้ 600 ชั้น [26]

โมเดล MobileNet เป็นโมเดลโครงข่ายประสาทเทียม ที่ได้รับการออกแบบสำหรับการประมวลผลภาพบนอุปกรณ์พกพาและอุปกรณ์ฝังตัว มาโดยมีขนาดเล็กและสามารถประมวลผลได้รวดเร็วทำให้สามารถใช้งานในอุปกรณ์พกพาได้ MobileNet ใช้ Convolution แบบแยกความลึก (Depth-wise Separable Convolutions) เพื่อสร้างโครงข่ายประสาทเทียมแบบคอนโวลูชันขนาดเล็ก โดยผู้พัฒนาได้ทดลองเปรียบเทียบกับโมเดลชนิดอื่นๆ และพบว่า MobileNet มีประสิทธิภาพที่เหนือกว่าและ

รวดเร็วกว่าโมเดลชนิดอื่นๆ อย่างมาก ในการจัดประเภทรูปภาพบนชุดข้อมูล ImageNet รวมถึงมีประสิทธิภาพในการจำแนกข้อมูลอื่นๆ อีกมากมายเช่น การตรวจจับวัตถุ การวิเคราะห์ใบหน้า และการระบุตำแหน่ง เป็นต้น [27]

ตารางที่ 2 ความแตกต่างของการเรียนรู้แบบถ่ายโอนเชิงลึกด้วยโมเดล MobileNet VGG และ EfficientNet

คุณสมบัติ	MobileNet	VGG-16 & -19	EfficientNet B0 & B7
ชั้น Convolution	28	16/19	86/600
จุดเด่น	ประสิทธิภาพ ความเร็ว	ความแม่นยำ	ความแม่นยำ ประสิทธิภาพ
พารามิเตอร์ ($\times 10^6$)	45.69	35.95/41.26	69.58/195.17
FLOPS	0.24	13.8/35.5	0.37/685.4

ผลการทดสอบและการวิเคราะห์ผลลัพธ์

1. ผลลัพธ์การทดลองใช้ Deep Transfer Learning

ในการจำแนกภาพด้วยโครงข่ายประสาทเทียมแบบคอนโวลูชันกำหนดให้มีการสอนโครงข่ายประสาทเทียมในการเรียนรู้ลักษณะเฉพาะของภาพและเปรียบเทียบประสิทธิภาพความแม่นยำในการจำแนกภาพ Mel Spectrogram จำนวน 50 รอบ และเปรียบเทียบประสิทธิภาพความแม่นยำในการจำแนกภาพ Mel Spectrogram ระหว่างโมเดล DTL ทั้ง 5 ชนิด ซึ่งใช้เทคนิคการหาค่าเหมาะสมที่สุด (optimizer) ในกระบวนการเรียนรู้แบบ backpropagation เป็น Stochastic Gradient Descent (SGD) และ ใช้ค่า hyperparameters เป็นค่า default ในแต่ละโมเดล โดยได้ทำการสอน (train) และทดสอบ (test หรือ validate) โมเดลทั้ง 5 ชนิด ซึ่งค่าถ่วงน้ำหนักเฉพาะในชั้นเชื่อมต่อแบบสมบูรณ์ (Fully-connected Layer) จะได้รับการสอน

ตารางที่ 3 แสดงผลลัพธ์ความแม่นยำในการสอนโมเดล EfficientNet B0, EfficientNet B7, VGG-16 VGG-19 และ MobileNet ตามลำดับ โดยแสดงให้เห็นว่า ประสิทธิภาพของโมเดลในการทดสอบและประสิทธิภาพการสอนของโมเดล EfficientNet B0, EfficientNet B7, VGG-16 และ VGG-19 มีค่าที่ต่ำ ซึ่งสาเหตุที่มีประสิทธิภาพที่ต่ำ คาดว่าจะเกิดจากปัจจัยหลายประการเช่น ความซับซ้อนของโมเดล และ ค่าถ่วงน้ำหนัก (weight) ในชั้นคอนโวลูชันนั้นไม่ได้มีการสอนใหม่ และ hyperparameters ต้องมีการ fine tuning เป็นต้น

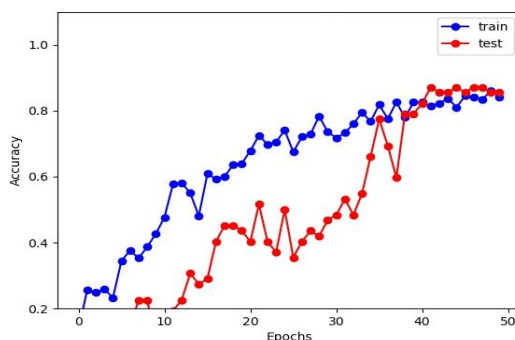
ตารางที่ 3 ตารางแสดงผลเปรียบเทียบประสิทธิภาพความแม่นยำในการจำแนกภาพ Mel Spectrogram ของโมเดล DTL จาก การสอน (train) และทดสอบ (validate) โมเดลทั้ง 5 ชนิด

	EfficientNet B0	EfficientNet B7	VGG-16	VGG-19	MobileNet
ความแม่นยำในการสอน	25.19%	25.19%	26.36%	27.13%	86.05%
ความแม่นยำในการทดสอบ	22.58%	22.58%	22.58%	22.58%	87.10%

จากกราฟที่ 5 แสดงถึงผลลัพธ์การสอนโมเดล MobileNet โดยกราฟจะแสดงถึงประสิทธิภาพความแม่นยำของโมเดลในการสอน (train) และในการทดสอบ (test) โดยแกนนอนแสดงถึงจำนวนรอบการสอน (epoch) แกนโมเดลและแกนตั้งแสดงถึงประสิทธิภาพความแม่นยำของโมเดล (accuracy) จากกราฟแสดงให้เห็นว่า ประสิทธิภาพของโมเดลในการทดสอบ

และการสอนจะเพิ่มขึ้นเรื่อย ๆ ตามจำนวนรอบการสอนที่เพิ่มขึ้น ในช่วงแรกๆ ประสิทธิภาพของโมเดลในการทดสอบจะเพิ่มขึ้นอย่างรวดเร็ว จากนั้นจะเพิ่มขึ้นอย่างช้า ๆ จนกระทั่งถึงจุดอิ่มตัว โดยเส้นสีน้ำเงินแสดงถึงประสิทธิภาพของโมเดลในการสอน มีค่าสูงสุดที่ 86.05% และเส้นสีแดงแสดงถึงประสิทธิภาพของโมเดลในการทดสอบกับชุดข้อมูลทดสอบ มีค่าสูงสุดอยู่ที่ 87.1% ซึ่งพบว่าค่าความแม่นยำในการจำแนกภาพจากชุดข้อมูลทดสอบมีค่าสูงใกล้เคียงกับค่าที่ได้จากชุดข้อมูลสอน นอกจากนี้ จะเห็นได้ว่าประสิทธิภาพของโมเดลในการทดสอบจะสูงกว่าประสิทธิภาพของโมเดลในการสอน ซึ่งแสดงให้เห็นว่าโมเดลถูกปรับให้เหมาะสมกับชุดข้อมูลการสอบแล้ว

โดยสาเหตุที่โมเดล MobileNet มีความแม่นยำสูงกว่าโมเดลชนิดอื่นๆ นั้นมาจากการใช้ชั้นคอนโวลูชันแบบ Depth-wise Separable Convolution ซึ่งมีพารามิเตอร์น้อยกว่าชั้นคอนโวลูชันแบบปกติ ในการลดพารามิเตอร์นี้ทำให้โมเดลมีประสิทธิภาพในการใช้หน่วยความจำและลดความเสี่ยงในการเกิดการเรียนรู้ผิดพลาด [28] โดยเฉพาะในกรณีที่มีข้อมูลการฝึกที่จำกัด รวมทั้งการใช้เทคนิค inverted residuals ในการเรียงลำดับการทำงานของโมเดลโครงข่ายประสาทเทียมแบบคอนโวลูชัน ซึ่งจะทำให้โมเดลมีประสิทธิภาพที่ดีกว่าโมเดลโดยทั่วไปที่ใช้การเรียงลำดับแบบปกติ (residual blocks) [29]



ภาพที่ 5 ผลลัพธ์ความแม่นยำในการสอน (train) และทดสอบ (test) ด้วยโมเดล MobileNet ซึ่งเป็นเทคนิค DTL

2. การตรวจสอบประเมินประสิทธิภาพของโมเดล

การวัดประสิทธิภาพของโมเดล MobileNet เป็นสิ่งจำเป็นสำหรับการประเมินความสามารถของโมเดล โดยในงานวิจัยนี้ จะทำการประเมินความแม่นยำในการจำแนกเสียงการหายใจแบบปกติและแบบผิดปกติจากภาพ Mel Spectrogram หลังการสอบโมเดล โดยผู้วิจัยได้พิจารณาจาก confusion matrix ของภาพเสียงหายใจในชุดที่ 2 จำนวน 62 ภาพ ที่ใช้ในการตรวจสอบ เพื่อแสดงให้เห็นถึงประสิทธิภาพการสอนหรือการเรียนรู้ของโมเดล MobileNet ดังในภาพที่ 6

Confusion Matrix Model Comparison

	Crackle	Normal	Rhonchi	Stridor	Wheezing
Crackle	14	0	0	0	0
Normal	0	6	0	0	0
Rhonchi	3	0	11	0	0
Stridor	0	0	6	8	0
Wheezing	0	0	0	0	14

ภาพที่ 6 Confusion matrix ที่บ่งบอกถึงประสิทธิภาพในการจำแนกภาพเสียงการหายใจทั้ง 5 ประเภท ด้วยโมเดล MobileNet ช่วยการประเมินความแม่นยำในการจำแนกของเสียงหายใจแต่ละชนิดได้อย่างถูกต้องหรือสับสนกับเสียงการหายใจแบบใดได้ด้วย

นอกจากนี้ผู้วิจัยได้ทำการทดสอบโมเดล MobileNet กับเสียงการหายใจที่วัดได้จากหูฟังแพทย์แบบดิจิทัล (Digital stethoscope) ที่ได้พัฒนาขึ้น เพื่อทดสอบการใช้งานกับเสียงหายใจที่วัดได้ในชุดข้อมูลชุดที่ 1 สำหรับเสียงทั้ง 5 ชนิด เพื่อทดสอบความสามารถในการจำแนกชนิดของเสียงหายใจ จากการทดสอบพบว่ามีความแม่นยำในการทำนาย ชนิดของเสียงหายใจที่ได้จากการทดสอบได้ผลลัพธ์ดังตารางที่ 4 ซึ่งแสดงถึงความถูกต้องของโมเดลในการทำนายชนิดของเสียงหายใจ

ตารางที่ 4 ผลลัพธ์ความน่าจะเป็นในการทดสอบใช้โมเดล MobileNet ในการจำแนกเสียงหายใจจากการทดลอง

เสียงที่ใช้ทดสอบ	ความน่าจะเป็นจากการทำนาย				
	Crackle	Rhonchi	Stridor	Wheezing	Normal
Crackle	48.45%	14.24%	8.03%	21.00%	7.40%
Rhonchi	1.96%	70.46%	11.12%	2.64%	13.81%
Stridor	0.81%	6.85%	79.55%	1.43%	11.34%
Wheezing	26.58%	13.72%	11.35%	41.24%	7.08%
Normal	0.04%	0.79%	1.09%	0.74%	97.34%

สรุปและขอเสนอแนะในการพัฒนาต่อไป

เป้าหมายหลักในงานวิจัยนี้ต้องการใช้เทคนิคการเรียนรู้แบบเชิงลึกมาช่วยในการจำแนกชนิดของเสียงหายใจของเด็กทั้งแบบปกติและเสียงหายใจที่ผิดปกติอีก 4 ประเภท ซึ่งจะช่วยให้การวินิจฉัยสาเหตุของโรค LRTIs ให้แก่บุคลากรทางการแพทย์ที่มีความแม่นยำมากยิ่งขึ้น และ ช่วยลดความผิดพลาดรวมถึงระยะเวลาในการวินิจฉัยลงได้ ส่งผลให้ผู้ป่วยได้รับการรักษาที่ถูกต้องและทันเวลาที่ ซึ่งจะช่วยลดความเสี่ยงต่อการเกิดภาวะแทรกซ้อนได้

ผู้วิจัยได้ทำการพัฒนาหูฟังแพทย์แบบดิจิทัลสำหรับการเก็บข้อมูลเสียงการหายใจจากปอดของตุ๊กตาเด็กทารกเสมือนจริง CAE Luna และแปลงสัญญาณเสียงให้เป็นภาพ Mel Spectrogram ผ่านกระบวนการแปลงฟูเรียร์ในช่วงเวลาสั้น แล้วนำภาพที่ได้มาใช้สอนโมเดลการเรียนรู้เชิงลึกแบบถ่ายโอน 5 เทคนิค จากผลการทดสอบพบว่า โมเดล EfficientNet B0, EfficientNet B7, VGG-16 และ VGG-19 ให้ความแม่นยำในการเรียนรู้และในการทำนายที่ต่ำ สำหรับภาพของเสียงการหายใจที่ปกติและไม่ปกติต่างๆ แต่โมเดล MobileNet มีความแม่นยำในการทำนายชนิดของเสียงการหายใจโดยมีความแม่นยำสูงสุด 87.1% รวมถึงเมื่อทดสอบกับเสียงการหายใจที่วัดจริงด้วยหูฟังแพทย์แบบดิจิทัลจากตุ๊กตาเด็กทารก ก็ยังสามารถทำนายชนิดของเสียงหายใจได้อย่างแม่นยำ

แม้ว่าโมเดลการเรียนรู้เชิงลึกจะสามารถจำแนกชนิดเสียงการหายใจที่วัดได้จากหูฟังแพทย์แบบดิจิทัลที่สร้างขึ้นได้อย่างถูกต้องในผลการวิจัยได้ตามจุดประสงค์ อย่างไรก็ตาม งานวิจัยนี้ยังสามารถพัฒนาได้อีกในอนาคต เนื่องจากชุดข้อมูลเสียงการหายใจที่ใช้สำหรับการศึกษาในงานวิจัยนี้จัดเก็บมาจากตุ๊กตาเด็กเสมือนจริงแทนการวัดเสียงการหายใจจากผู้ป่วยเด็กจริง เนื่องจากมีข้อจำกัดด้านเวลาการทำวิจัยและขั้นตอนในการขออนุญาตจากผู้ปกครองหรือผู้ดูแลเด็ก ทำให้โมเดลการเรียนรู้เชิงลึกที่สร้างขึ้นมานั้นอยู่เพียงแค่อันดับการทดลองในสภาพแวดล้อมจำลอง ฉะนั้นควรต้องมีการพัฒนาต่อยอดเพื่อให้สามารถนำเทคนิคในงานวิจัยนี้ไปใช้ในงานช่วยเหลือบุคลากรทางการแพทย์ในสภาพแวดล้อมจริงในโรงพยาบาลได้อย่างมีประสิทธิภาพ ในลำดับต่อไป

กิตติกรรมประกาศ

ขอขอบคุณสำนักงานคณะกรรมการวิจัยแห่งชาติ (วช.) สำหรับทุนสนับสนุนโครงการวิจัยเรื่อง “การพัฒนาผลิตภัณฑ์และประสิทธิภาพของ Automatic Pediatric Posture-Gyro: Intelligence for Sick Babies” ปี 2564

เอกสารอ้างอิง

1. Tchatchouang S, Nzouankeu A, Kenmoe S, Ngando L, Penlap V, Fonkoua M-C, et al. Bacterial Aetiologies of Lower Respiratory Tract Infections among Adults in Yaoundé, Cameroon. *BioMed Research International*. 2019; 2019: 4834396; DOI 10.1155/2019/4834396.
2. Matloobi A, Buday T, Brozmanova M, Konarska M, Poljacek I, Martvon L, et al. The effect of stimulation and unloading of baroreceptors on cough in experimental conditions. *Respiratory Physiology & Neurobiology*. 2022; 303: 103921; DOI: 10.1016/j.resp.2022.103921.
3. Musher DM, Abers MS, Bartlett JG. Evolving Understanding of the Causes of Pneumonia in Adults, With Special Attention to the Role of Pneumococcus. *Clinical Infectious Diseases*. 2017; 65(10): 1736-1744.
4. Shields MD, Bush A, Everard ML, McKenzie S, Primhak R. BTS guidelines: Recommendations for the assessment and management of cough in children. *Thorax*. 2008 Apr; 63 Suppl 3: iii1-iii15; DOI: 10.1136/thx.2007.077370.
5. Jain S, Self WH, Wunderink RG, Fakhran S, Balk R, Bramley AM, et al. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. *N Engl J Med*. 2015; 373 (5): 415-427; DOI: 10.1056/NEJMoal500245.
6. Lanata CF, Black RE. Acute Lower Respiratory Infections. In: Semba RD, Bloem MW, Piot P, editors. *Nutrition and Health in Developing Countries*. Totowa, NJ: Humana Press; 2008; 179-214; DOI: 10.1007/978-1-59745-464-3_7.
7. Morrow BM. Airway clearance therapy in acute paediatric respiratory illness: A state-of-the-art review. *S Afr J Physiother*. 2019; 75(1): 1295; DOI: 10.4102/sajp.v75i1.1295.
8. Mamouris P, Henrard S, Molenberghs G, Verhaegen J, Lin G, Vaes B. Pneumococcal vaccination prevented severe LRTIs in adults: a causal inference framework applied in registry data. *J Clin Epidemiol*. 2022; 143: 118-127; DOI: 10.1016/j.jclinepi.2021.12.008.
9. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. 2020; 396(10258): 1204-1222; DOI: 10.1016/S0140-6736(20)30925-9.
10. Suwanjutha S, Chantarojsiri T, Prapphal N, Suntornlohanakul S, Laohapand C, Wongpaitoon N. Acute Lower Respiratory Infections. In: Deerojanawong J, Boonjindasap V, editors. *Guideline for the management of Acute Respiratory Infections in Children*. Nonthaburi: Beyond Enterprise Company LIM; 2019. 72-155.
11. Schwartz KL, Langford BJ, Daneman N, Chen B, Brown KA, McIsaac W, et al. Unnecessary antibiotic prescribing in a Canadian primary care setting: a descriptive analysis using routinely collected electronic medical record data. *CMAJ Open*. 2020; 8(2): E360-E369; DOI: 10.9778/cmajo.20190175.
12. Kumar S, Kumar H. Classification of COVID-19 X-ray images using transfer learning with visual geometrical groups and novel sequential convolutional neural networks. *MethodsX*. 2023; 11: 102295; DOI: 10.1016/j.mex.2023.102295.

13. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 3rd International Conference on Learning Representations (ICLR 2015), Computational and Biological Learning Society. 2014; pp. 1–14; San Diego, CA, USA; DOI: 10.48550/arXiv.1409.1556.
14. Porter P, Brisbane J, Tan J, Bear N, Choveaux J, Della P, et al. Diagnostic Errors Are Common in Acute Pediatric Respiratory Disease: A Prospective, Single-Blinded Multicenter Diagnostic Accuracy Study in Australian Emergency Departments. *Front Pediatr.* 2021; 9: 736018; DOI: 10.3389/fped.2021.736018.
15. Serbes G, Sakar CO, Kahya Y, Aydin N. Pulmonary crackle detection using time-frequency analysis. *Digital Signal Processing.* 2013; 23(3): 1012-1021; DOI: 10.1016/j.dsp.2012.12.009.
16. McGee S. Auscultation of the Lungs. In: McGee S, editor. *Evidence-Based Physical Diagnosis* (Fourth Edition). Philadelphia: Elsevier; 2018; p: 261-274. e4.
17. Zulfiqar R, Majeed F, Irfan R, Rauf HT, Benkhelifa E, Belkacem AN. Abnormal Respiratory Sounds Classification Using Deep CNN Through Artificial Noise Addition. *Front Med (Lausanne).* 2021; 8:714811; DOI: 10.3389/fmed.2021.714811.
18. Reichert S, Gass R, Brandt C, Andrès E. Analysis of respiratory sounds: state of the art. *Clin Med Circ Respir Pulm Med.* 2008; 2: 45-58; DOI: 10.4137/ccrpm.s530.
19. Andrès E, Gass R, Charloux A, Brandt C, Hentzler A. Respiratory sound analysis in the era of evidence-based medicine and the world of medicine 2.0. *J Med Life.* 2018; 11(2): 89-106.
20. Kim Y, Hyon Y, Jung SS, Lee S, Yoo G, Chung C, et al. Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. *Scientific reports.* 2021;11(1):1-11; DOI: 10.1038/s41598-021-96724-7.
21. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? *Advances in neural information processing systems (NIPS).* 2014; 27: 3320-3328; DOI: 10.48550/arXiv.1411.1792.
22. Huh M, Agrawal P, Efros AA. What makes ImageNet good for transfer learning? *arXiv: 1608.08614v2.* 2016; DOI: 10.48550/arXiv.1608.08614.
23. Faiza MS, Moftah RAM, Bozed KA. Applying Convolutional Neural Network and Transfer Learning to Mel-Spectrograms for COVID-19 Diagnosis via Cough Sounds. *Int J Sci Res Publ.* 2023; 13(06): 32-37; DOI: 10.29322/USRP.13.06.2023.
24. Ali AA, Al-Marzouqi H. Melanoma detection using regular convolutional neural networks. 2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA), United Arab Emirates, 2017: 1-5; DOI: 10.1109/ICECTA.2017.8252041.
25. Bansal M, Kumar M, Sachdeva M, Mittal A. Transfer learning for image classification using VGG19: Caltech-101 image data set. *J Ambient Intell and Humaniz Comput.* 2023;14(4): 3609–3620; DOI: 10.1007/s12652-021-03488-z.
26. Hoang VT, Jo KH. Practical Analysis on Architecture of EfficientNet. 2021 14th International Conference on Human System Interaction (HSI); 2021 July 8-10; pp. 1-4; Gdansk, Poland: Institute of Electrical and Electronics Engineers; 2021; DOI: 10.1109/HSI52170.2021.9538782.



27. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv: 1704.04861v1. 2017; DOI: 10.48550/arXiv.1704.04861.
28. Srivastava H, Sarawadekar K. A Depthwise Separable Convolution Architecture for CNN Accelerator. 2020 IEEE Applied Signal Processing Conference (ASPCON); 2020 Oct 7-9; pp. 1-5; Kolkata, India: Institute of Electrical and Electronics Engineers; 2020; DOI: 10.1109/ASPCON49795.2020.9276672.
29. Li H, Wang J, Xiong N, Zhang Y, Vasilakos AV, Luo X. A Siamese Inverted Residuals Network Image Steganalysis Scheme based on Deep Learning. ACM Trans Multimedia Comput Commun Appl. 2023;19(6):214. DOI: 10.1145/3579166.