

บทที่ 2

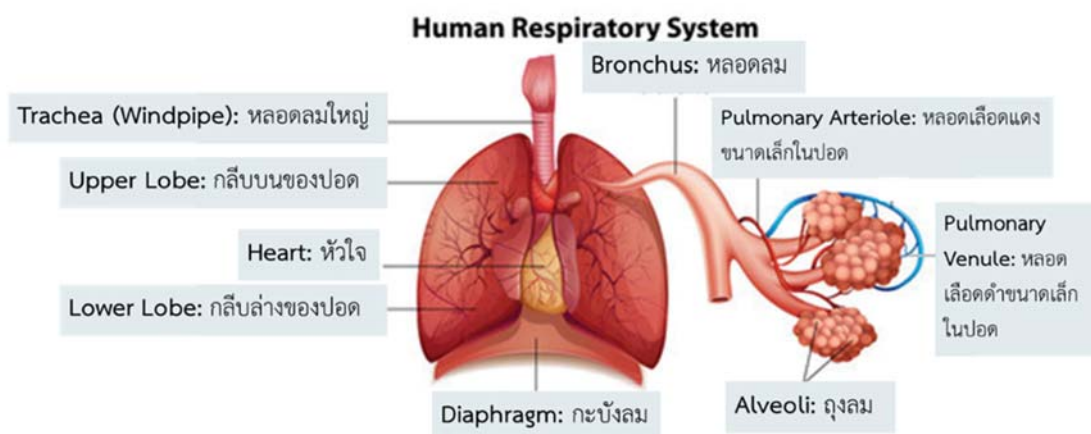
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

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 (เสียงแตก กรอบ / เสียงป๊อป)	- เป็นเสียงสั้น ๆ คล้ายเสียงแตกหรือเสียงป๊อปเป็นช่วง ๆ - มักได้ยินชัดในช่วงหายใจเข้า - เกิดจากของเหลวในทางเดินหายใจหรือการเปิดของถุงลม/หลอดลมที่แฟบ - เสียงอาจมีตั้งแต่เบาและแหลม ไปจนถึงดังและคล้ายเสียงน้ำกล้ว	- 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 (เสียงครี๊ดคราดต่ำ)	- เป็นเสียงต่ำ คล้ายเสียงกรน - มักได้ยินชัดในช่วงหายใจออก - เกิดจากเสมหะหรือสารคัดหลั่งในทางเดินหายใจขนาดใหญ่ - มักลดลงหรือหายไปหลังไอ	- ทางเดินหายใจอุดกั้น - หลอดลมอักเสบ (Bronchitis) - ปอดอักเสบ
Diminished (เสียงปอดเบา)	- 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) - คroup (Croup)

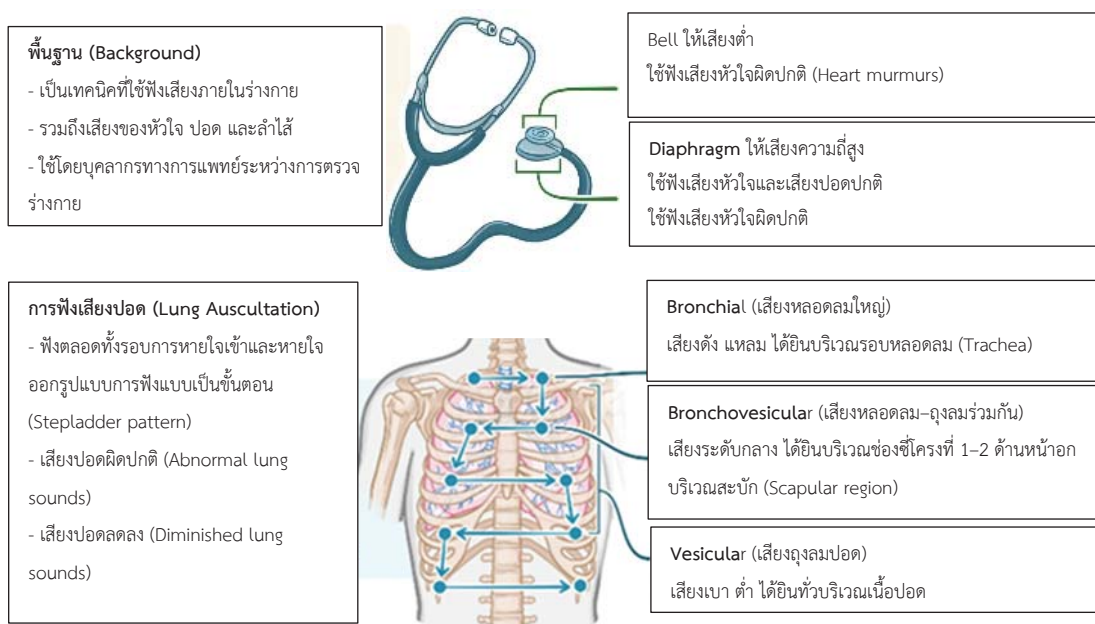
ตารางที่ 2.1 แนวคิดเชิงกลไกของการเกิดเสียงปอด* (ต่อ)

เสียงปอด	ลักษณะสำคัญ	โรคที่เกี่ยวข้อง
Pleural Rub (เสียงเสียดสี ของเยื่อหุ้ม ปอด)	- 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



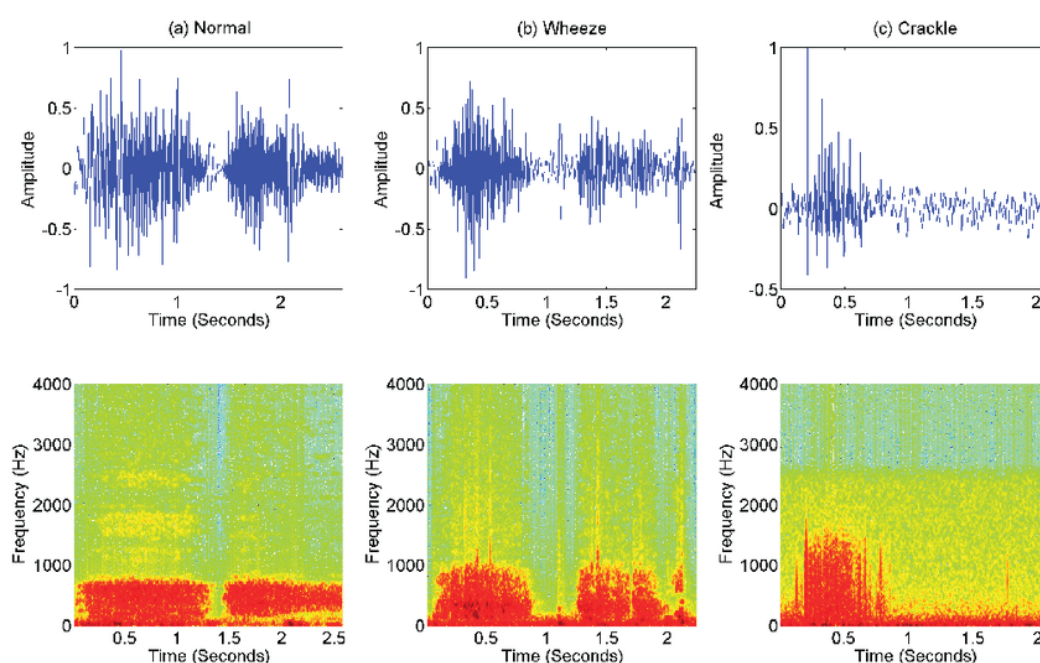
รูปที่ 2.2 ส่วนประกอบของหูฟังแพทย์ และตำแหน่งมาตรฐานสำหรับการตรวจฟังเสียง

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

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

2.1.3 ประเภทของเสียงปอดและลักษณะเชิงความถี่ของสัญญาณ

เสียงปอดสามารถจำแนกเป็นเสียงปอดภาวะปกติ (Normal lung sounds) และเสียงปอดภาวะผิดปกติ (Abnormal lung sounds) โดยเสียงภาวะผิดปกติที่พบบ่อย ได้แก่ เสียงแตกกรอบ เสียงหวีด เสียงครีตคราดต่ำ และเสียงหวีดแหลมดัง ซึ่งสัมพันธ์กับพยาธิสภาพของระบบทางเดินหายใจและมีลักษณะเชิงเวลาและเชิงความถี่เฉพาะตัว (Sarkar, M., Madabhavi, I., Niranjana, 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) ซึ่งส่งผลกระทบต่ออากาศ ทำให้ทางเดินหายใจตีบหรืออุดตัน (2) โรคของเนื้อเยื่อปอด (Lung tissue diseases) ซึ่งส่งผลกระทบต่อโครงสร้างเนื้อเยื่อปอด ทำให้ปอดขยายได้ไม่เต็มที่ และ (3) โรคการไหลเวียนของปอด (Lung circulation diseases) ซึ่งส่งผลกระทบต่อหลอดเลือดในปอดและสัมพันธ์กับภาระการทำงานของหัวใจ (โรงพยาบาลบำรุงราษฎร์, 2567) ในงานวิจัยนี้มุ่งศึกษาสัญญาณเสียงปอดเพื่อสนับสนุนการจำแนกโรคที่เกี่ยวข้องกับระบบทางเดินหายใจซึ่งพบได้บ่อย และสอดคล้องกับแนวทางการวิจัยด้านปัญญาประดิษฐ์ที่กำหนดจำแนกโรคที่สำคัญ เช่น โรคปอดอักเสบ โรคหอบหืด และโรคปอดอุดกั้นเรื้อรัง เป็นต้น

2.1.4.1 โรคปอดอักเสบ

เรียกอีกชื่อว่า โรคปอดบวม เป็นการอักเสบของเนื้อปอดที่พบได้บ่อย โดยเฉพาะในเด็กเล็ก ผู้สูงอายุ และผู้ที่มีภาวะภูมิคุ้มกันต่ำ ซึ่งอาจรุนแรงจนก่อให้เกิดภาวะแทรกซ้อนและการเสียชีวิตได้ สาเหตุของโรคแบ่งได้เป็น 2 กลุ่ม ได้แก่ ปอดอักเสบจากการติดเชื้อ ซึ่งเป็นชนิดที่พบบ่อยที่สุด โดยเชื้อไวรัส เชื้อแบคทีเรีย หรือเชื้อรา สามารถเข้าสู่ปอดและทำให้เกิดการอักเสบของถุงลมและเนื้อเยื่อโดยรอบ และปอดอักเสบจากการไม่ติดเชื้อ เช่น การสูดสารระคายเคือง (ฝุ่น คาร์บอน 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

เป็นภาวะอักเสบเรื้อรังของเยื่อหลอดลมร่วมกับความไวของหลอดลมต่อสิ่งกระตุ้น ทำให้เกิดอาการหายใจลำบาก เหนื่อยหอบ และสมรรถภาพปอดลดลง โดยโรคนี้รักษาไม่หายขาดแต่สามารถควบคุมไม่ให้อาการกำเริบได้ สาเหตุสำคัญประกอบด้วยปัจจัยสิ่งแวดล้อม (สารก่อภูมิแพ้ การติดเชื้อทางเดินหายใจ อากาศเย็น กรดไหลย้อน ความเครียด หรือยาบางชนิด) และปัจจัยทางพันธุกรรม อาการที่พบบ่อย ได้แก่ ไอมีเสมหะ แน่นหน้าอก หายใจไม่อิ่ม หายใจมีเสียงหวีด และเหนื่อยหอบ 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 เนื่อปอด 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

พัฒนาเพื่อรองรับการศึกษาและพัฒนาอัลกอริทึมทางการเรียนรู้ของเครื่อง และปัญญาประดิษฐ์ สำหรับการตรวจจับและจำแนกโรคปอด

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

2.1.5.1 วิธีการและขั้นตอนการเก็บข้อมูลโรคปอดอักเสบ

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

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 เฮิร์ตซ์ (Pasterkamp et al., 1997; Rangayyan, 2015) สมการที่ (2.1) แสดงการตอบสนองเชิงขนาดของตัวกรองแถบผ่านแบบบัตเตอร์เวิร์ธแบบความถี่ต่ำผ่านในรูปทั่วไป โดยในการทดลองจริงใช้การออกแบบเป็น bandpass ด้วยความถี่ตัดต่ำ f_L และความถี่ตัดสูง f_H ตามกระบวนการออกแบบตัวกรองมาตรฐาน

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

เมื่อ

$H(f)$	คือ ฟังก์ชันการตอบสนองเชิงความถี่ของตัวกรอง
f	คือ ความถี่ของสัญญาณ (เฮิร์ตซ์)
f_c	คือ ความถี่ตัดของตัวกรอง (เฮิร์ตซ์)
N_{BW}	คือ อันดับของตัวกรองบัตเตอร์เวิร์ธ

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

แม้สัญญาณเสียงปอดจะไม่คงที่ แต่สามารถประมาณให้คงที่ได้ในช่วงเวลาสั้น จึงทำให้การวิเคราะห์แบบช่วงเวลาและการคำนวณสเปกตรัมมีความหมาย (Rangayyan, 2015; Bahoura

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

2.2.2 สัมประสิทธิ์เซปสตรีมความถี่เมล

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

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

เมื่อ

$MFCC(c)$	คือ ค่าสัมประสิทธิ์เซปสตรีมความถี่เมลลำดับที่ c
c	คือ ดัชนีของสัมประสิทธิ์เซปสตรีม โดยทั่วไปเลือกเฉพาะลำดับต้น ๆ ที่มีสารสนเทศสำคัญ $c = 0, 1, \dots, N_c - 1$
N_c	คือ จำนวนสัมประสิทธิ์เซปสตรีมความถี่เมลที่เลือกใช้งาน
$E_{(m)}$	คือ ค่าพลังงานของสัญญาณเสียงในแถบความถี่เมลลำดับที่ m หลังจากผ่านชุดตัวกรองเมล
m	คือ ดัชนีของตัวกรองเมล $m = 0, 1, \dots, M - 1$
M	คือ จำนวนตัวกรองเมลทั้งหมด
$\log(\cdot)$	คือ ฟังก์ชันลอการิทึม ซึ่งใช้เพื่อจำลองการรับรู้ความดังของเสียง
$\cos(\cdot)$	คือ ฟังก์ชันโคไซน์ ซึ่งเป็นส่วนหนึ่งของการแปลงโคไซน์แบบไม่ต่อเนื่อง

สัมประสิทธิ์เซปสตรีมความถี่เมล มีความเหมาะสมสำหรับเสียงปอด เนื่องจากสามารถจับความแตกต่างของสเปกตรัมระหว่างกลุ่ม เช่น เสียงหวีด ที่มีพลังงานเด่นและต่อเนื่องในย่านความถี่บางช่วง หรือเสียงกรอบแตกที่เกิดขึ้นช่วงสั้นและมีการเปลี่ยนแปลงฉับพลัน

2.2.3 คุณลักษณะเชิงพลวัตที่สะท้อนการเปลี่ยนแปลงเชิงเวลา

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

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

เมื่อ

Δc_t	คือ ค่าอนุพันธ์อันดับหนึ่งของสัมประสิทธิ์เซปสตรีมความถี่เมล ณ เฟรมเวลา t
c_t	คือ ค่าสัมประสิทธิ์เซปสตรีมความถี่เมล ณ เฟรมเวลา t
T	คือ จำนวนเฟรมด้านข้างที่ใช้ในการประมาณอนุพันธ์เชิงเวลา
n	คือ ดัชนีผลรวมสำหรับการคำนวณเดลตา ($n = 1, \dots, T$)
t	คือ ดัชนีของเฟรมเวลา

สมการที่ (2.4) อนุพันธ์อันดับสอง (Delta-Delta coefficient) ของสัมประสิทธิ์เซปสตรีมความถี่เมล

$$\Delta^2 c_t = \Delta(\Delta c_t) \quad (2.4)$$

เมื่อ

$\Delta^2 c_t$	คือ ค่าอนุพันธ์อันดับสองของสัมประสิทธิ์เซปสตรีมความถี่เมล ณ เฟรมเวลา t
----------------	--

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 ซึ่งช่วยให้การเรียนรู้รูปแบบของเสียงปลอดภัยจากตัวแทนสัมประสิทธิ์เซปสตรัมความถี่เมลในโดเมนเวลาและความถี่มีประสิทธิภาพมากขึ้น

แนวทางการวิจัย เพื่อให้การสกัดคุณลักษณะสอดคล้องกับการจำแนกโรกระบบทางเดินหายใจแบบหลายกลุ่ม 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 การสกัดคุณลักษณะเป็นขั้นตอนสำคัญที่ทำให้สัญญาณเสียงปลอดภัยไม่คงที่ตามเวลาและมีสัญญาณรบกวนสูง สามารถถูกแทนด้วยชุดคุณลักษณะที่เหมาะสมต่อการเรียนรู้ของปัญญาประดิษฐ์ โดยสัมประสิทธิ์เซปสตรัมความถี่เมลและคุณลักษณะอนุพันธ์ตามเวลา มีบทบาทสำคัญในการจับคุณลักษณะเสียงที่สัมพันธ์กับโรค เช่น โรคปอดอักเสบ โรคหอบหืด และโรคปอดอุดกั้นเรื้อรัง ซึ่งให้เห็นพารามิเตอร์และขั้นตอนสำคัญที่ส่งผลต่อประสิทธิภาพการจำแนก (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 สเปกโตรแกรมแบบลอการิทึม 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 แนวคิดดังกล่าวเป็นการผสมผสานที่สำคัญระหว่างการสกัดคุณลักษณะและการประยุกต์ใช้โครงข่ายประสาทเทียมแบบคอนโวลูชัน 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)

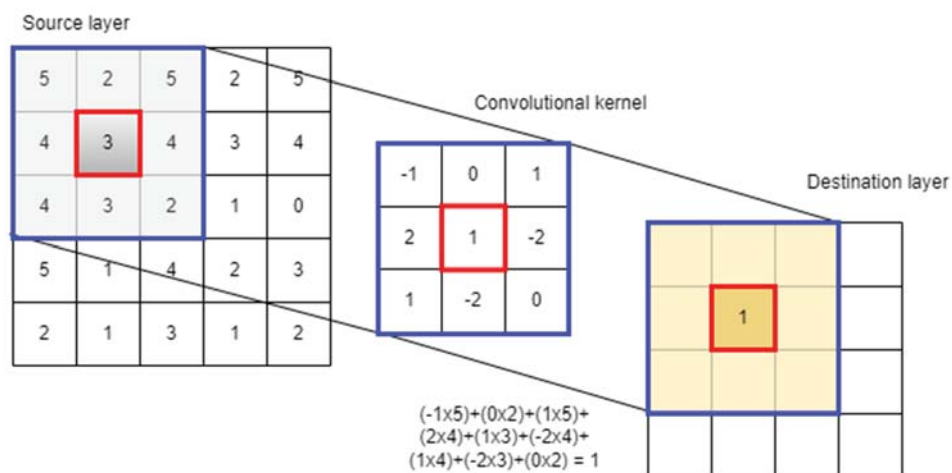
2.3.1.1 หลักการของโครงข่ายประสาทเทียมแบบคอนโวลูชัน

โครงข่ายประสาทเทียมแบบคอนโวลูชัน ประกอบด้วยชั้นคอนโวลูชัน (Convolutional layer) ซึ่งใช้ตัวกรองหรือเคอร์เนล (Kernel) เลื่อนผ่านอินพุตเพื่อคำนวณผลคูณจุด (Dot product) และสร้างแผนที่คุณลักษณะ แนวคิดสำคัญ คือ การเชื่อมต่อเฉพาะที่ (Local connectivity) และการใช้ค่าน้ำหนักร่วมกัน (Weight sharing) ทำให้แบบจำลองสามารถเรียนรู้รูปแบบที่เกิดขึ้นในบริเวณต่าง 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 ค่าน้ำหนักของคอร์เนล 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 ดัชนีตำแหน่งภายในคอร์เนล
k_h และ k_w	คือ ความสูงและความกว้างของคอร์เนลตามลำดับ



รูปที่ 2.4 การเลื่อนคอร์เนลบนอินพุตและการสร้างแผนที่คุณลักษณะ

การกำหนดพารามิเตอร์ของชั้นคอนโวลูชันมีผลต่อความละเอียดเชิงพื้นที่ของการเรียนรู้และขนาดของแผนที่คุณลักษณะ โดยเฉพาะค่าการเลื่อน (Stride) และการเติมขอบ (Padding) ซึ่งเป็นองค์ประกอบพื้นฐานในการออกแบบโครงข่ายประสาทเทียมแบบคอนโวลูชัน

ค่าการเลื่อน หมายถึง ระยะก้าวของการเลื่อนคอร์เนล หากค่าการเลื่อนมีค่ามากขึ้น คอร์เนลจะเลื่อนที่หลายตำแหน่ง ส่งผลให้จำนวนตำแหน่งที่คำนวณลดลง ขนาดแผนที่

คุณลักษณะลดลง และระดับการซ้อนทับระหว่างหน้าต่างคอนโวลูชันลดลง สมการที่ (2.6) ความสัมพันธ์ของขนาดผลลัพธ์เมื่อไม่มีการเติมขอบ

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

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 ฟังก์ชันเรกติไฟต์ไลเนียร์ยูนิต (Rectified Linear Unit: ReLU) เป็นฟังก์ชันกระตุ้นที่ช่วยลดปัญหาการหายไปของกราเดียนต์และเพิ่มความเร็วในการฝึกของโครงข่าย เป็นตัวเลือกที่นิยมเนื่องจากช่วยให้การเรียนรู้มีเสถียรภาพและลดปัญหาการไล่ระดับ ในบางกรณี (Nair, V., and Hinton, G. E., 2010)

$$ReLU(x_{ReLU}) = \max(0, x_{ReLU}) \quad (2.7)$$

เมื่อ

x_{ReLU}	คือ ข้อมูลนำเข้าของฟังก์ชันเรกติไฟต์ไลเนียร์ยูนิต
$ReLU(\cdot)$	คือ ฟังก์ชันเรกติไฟต์ไลเนียร์ยูนิต
$\max(0, x_{ReLU})$	คือ ฟังก์ชันที่ให้ผลลัพธ์เป็นศูนย์เมื่ออินพุตมีค่าน้อยกว่าหรือเท่ากับศูนย์ และให้ผลลัพธ์เท่ากับอินพุตเมื่ออินพุตมีค่ามากกว่าศูนย์

ชั้นการรวมค่า ใช้เพื่อลดมิติของข้อมูล ลดความไวต่อการเลื่อนตำแหน่ง และช่วยควบคุมความซับซ้อนการคำนวณของแบบจำลอง ในงานเสียงปอดการลดมิติและการเน้นสารสนเทศสำคัญมีความสำคัญ เนื่องจากอินพุตเวลาและควมถี่มีขนาดใหญ่และมีสัญญาณรบกวน ในกรณีการจำแนกหลายกลุ่มนิยมใช้ฟังก์ชันซอฟต์แวร์แมกซ์ (Softmax) คือ ฟังก์ชันกระตุ้นที่ใช้ในชั้นเอาต์พุตของแบบจำลองจำแนก เพื่อแปลงเวกเตอร์ค่าคะแนนเชิงเส้น (Logits) ให้เป็นการแจกแจงความน่าจะเป็นของแต่ละกลุ่ม โดยผลรวมของความน่าจะเป็นทั้งหมดมีค่าเท่ากับ 1 ดังสมการที่ (2.8)

$$p_k = \frac{e^{a_k}}{\sum_{j_{\text{soft}}=1}^{K_{\text{soft}}} e^{a_{j_{\text{soft}}}}} \quad (2.8)$$

เมื่อ

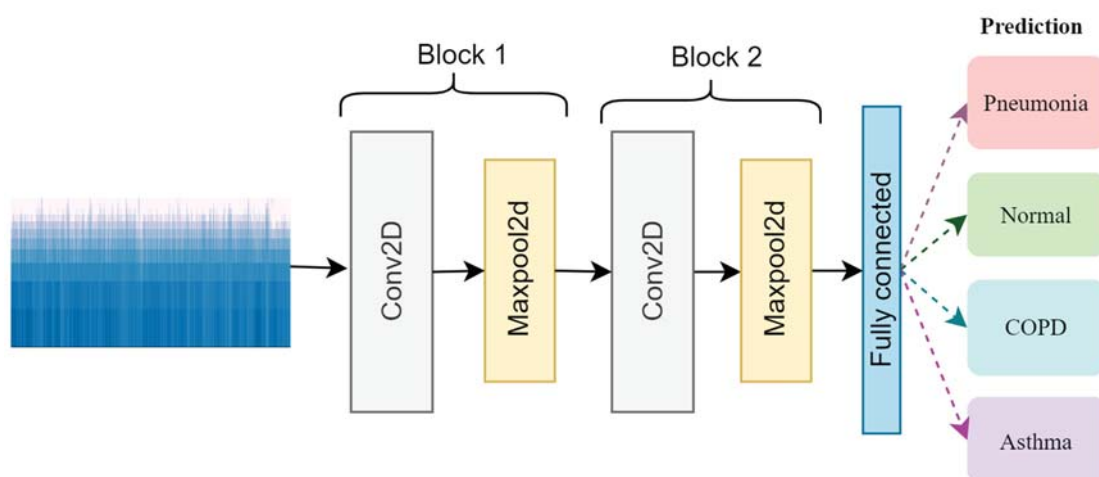
a_k	คือ ค่าลอจิกของกลุ่มที่ k ซึ่งเป็นผลลัพธ์จากชั้นเชื่อมโยงสมบูรณ์
k	คือ ดัชนีของกลุ่มเป้าหมาย โดย $k = 1, 2, \dots, K_{\text{soft}}$
j_{soft}	คือ ดัชนีสำหรับการรวม โดย $j = 1, 2, \dots, K_{\text{soft}}$
K_{soft}	คือ จำนวนกลุ่มทั้งหมดของปัญหาการจำแนก
P_k	คือ ความน่าจะเป็นที่แบบจำลองทำนายว่าข้อมูลอยู่ในกลุ่มที่ 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 และชั้นแมกซ์พูลลิง (MaxPool2d) เพื่อทำการลดมิติของข้อมูลและเพิ่มความทนทานต่อการเลื่อนตำแหน่งของคุณลักษณะ หลังจากผ่าน

กระบวนการดังกล่าวข้อมูลจะถูกแปลงเป็นเวกเตอร์และส่งต่อไปยังชั้นเชื่อมโยงสมบูรณ์ เพื่อทำการจำแนกประเภทของเสียงปอดในขั้นสุดท้าย



รูปที่ 2.5 สถาปัตยกรรมของโครงข่ายประสาทเทียมแบบคอนโวลูชันสองชั้น

ขั้นตอนการประมวลผลข้อมูลอินพุตของแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชันสองชั้น คือ คุณลักษณะสัมประสิทธิ์เซปสตรัมความถี่เมลถูกใช้เป็นตัวแทนเชิงเวลาและความถี่ของสัญญาณเสียง การใช้สัมประสิทธิ์เซปสตรัมความถี่เมลร่วมกับโครงข่ายประสาทเทียมแบบคอนโวลูชัน ช่วยให้แบบจำลองสามารถเรียนรู้รูปแบบของพลังงานและโครงสร้างเชิงสเปกตรัมที่เกี่ยวข้องกับลักษณะเสียงปอดภาวะปกติและภาวะผิดปกติ เพื่อให้การเปรียบเทียบมีสอดคล้องกับแนวทางการวิจัยก่อนหน้านี้ 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

สถิติที่เฉพาะเจาะจงกับชุดข้อมูลฝึกมากกว่าการเรียนรู้โครงสร้างเชิงกายภาพหรือสรีรวิทยาของสัญญาณเสียงพูด

การนำเสนอแนวคิดออโตเอนโคเดอร์แบบหลายงาน ซึ่งควบคุมให้ตัวเข้ารหัสเรียนรู้ตัวแทนเชิงแฝงที่รองรับทั้งการรักษาโครงสร้างสัญญาณผ่านการสร้างกลับคืน และการจำแนกโรคพร้อมกัน ทำให้การเรียนรู้มีความสมดุลและทำหน้าที่เป็นเทคนิคการควบคุมความซับซ้อนของแบบจำลอง ส่งผลให้ได้คุณลักษณะเชิงลึกที่มีเสถียรภาพ เหมาะกับข้อมูลเสียงพูดที่มีความซับซ้อนและสัญญาณรบกวน

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

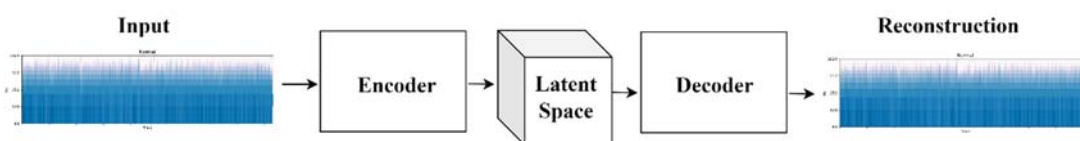
1. ตัวเข้ารหัส ทำหน้าที่แปลงข้อมูลอินพุต (Input data) ไปเป็นตัวแทนเชิงแฝง
2. ตัวถอดรหัส ทำหน้าที่สร้างข้อมูลประมาณค่าจากตัวแทนเชิงแฝง

$$z = \phi(v) \quad (2.9)$$

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

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



รูปที่ 2.6 ออโตเอนโคเดอร์

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

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

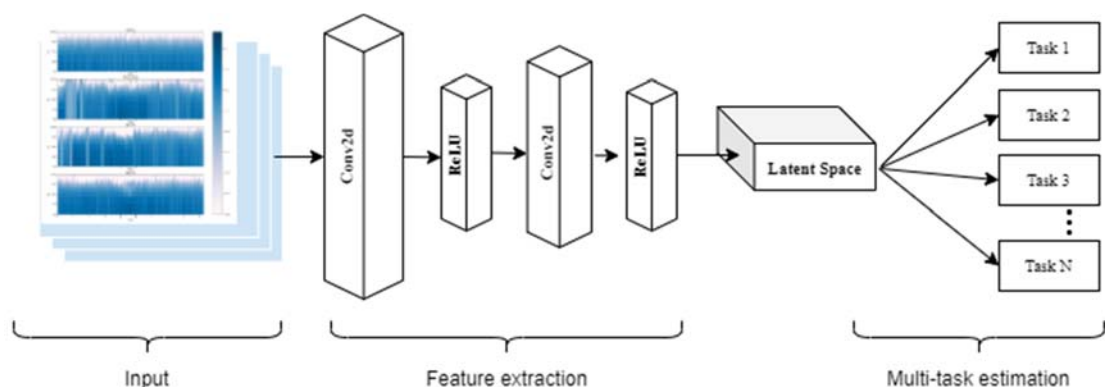
สมการฟังก์ชันการสูญเสียจากการสร้างข้อมูลกลับคืน (Reconstruction loss) ดังสมการที่ (2.11)

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

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 ตัวเข้ารหัสรวม และมีหัวงาน (Task heads) แยกสำหรับตัวถอดรหัส และตัวจำแนก (Classifier head) ตัวเข้ารหัสทำหน้าที่สร้างตัวแทนร่วม z ซึ่งถูกใช้ทั้งในการสร้าง $\hat{v}$ และการทำนายกลุ่ม $\hat{u}$ สมการการทำนายกลุ่มตัวจำแนก ดังสมการที่ (2.12)

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

เมื่อ

u	คือ ป้ายกำกับจริงของข้อมูล
$\hat{u}$	คือ เวกเตอร์ความน่าจะเป็นของแต่ละกลุ่มที่ได้จากการทำนาย
$h(\cdot)$	คือ ฟังก์ชันตัวจำแนก

สำหรับงานจำแนกหลายกลุ่ม ใช้ฟังก์ชันการสูญเสียแบบครอสเอนโทรปี (Cross entropy loss) เป็นตัววัดความเหมาะสมของการทำนาย

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

เมื่อ

L_{class}	คือ ค่าฟังก์ชันการสูญเสียสำหรับงานจำแนกประเภท
K_{class}	คือ จำนวนกลุ่มทั้งหมดของการจำแนก
k_{class}	คือ ดัชนีของกลุ่มการจำแนก โดย $k_{class} = 1, 2, \dots, K_{class}$
$y_{k_{class}}$	คือ ค่าป้ายกำกับจริงของกลุ่มที่ k_{class}
$\hat{y}_{k_{class}}$	คือ ค่าความน่าจะเป็นที่แบบจำลองทำนายสำหรับกลุ่มที่ k_{class}

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

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

เมื่อ

L_{total}	คือ ฟังก์ชันการสูญเสียรวมของแบบจำลองออโตเอนโคเดอร์แบบหลายงาน
L_{recon}	คือ ฟังก์ชันการสูญเสียจากการสร้างข้อมูลกลับคืน
L_{class}	คือ ฟังก์ชันการสูญเสียสำหรับการจำแนก

การรวม L_{recon} ทำให้แบบจำลองรักษาโครงสร้างของสัญญาณและลดการสูญเสียสารสนเทศ ขณะที่ L_{class} ทำให้ตัวแทนเชิงแฝงมีความสามารถในการจำแนกได้ดีเมื่อฝึกพร้อมกัน 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

ในกรณี ข้อมูลเชิงเส้น (Linearly separable) จะหาสมการไฮเปอร์เพลน (Hyperplane) ที่เหมาะสมที่สุดเพื่อแยกข้อมูลสองกลุ่ม แนวคิดทางคณิตศาสตร์และการเรียนรู้ของเครื่องที่ หมายถึง พื้นผิวหรือขอบเขตที่ใช้แบ่งข้อมูลออกเป็นกลุ่ม ในปริภูมิหลายมิติ

$$w_{hyp}^{Transpose} x_{hyp} + b_{hyp} = 0 \quad (2.15)$$

เมื่อ

w_{hyp}	คือ เวกเตอร์ค่าน้ำหนักของไฮเปอร์เพลน
x_{hyp}	คือ เวกเตอร์คุณลักษณะของข้อมูล
b_{hyp}	คือ ค่าคงที่

แนวคิดของระยะขอบมากที่สุด โดยต้องการทำให้ขนาดของเวกเตอร์น้ำหนักเล็กที่สุด ภายใต้เงื่อนไขการแยกกลุ่มที่ถูกต้อง ดังสมการที่ (2.16)

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

เมื่อ

w_{hyp}	คือ เวกเตอร์ค่าน้ำหนักของไฮเปอร์เพลน
y_q	คือ ป้ายกำกับจริงของตัวอย่างลำดับที่ q โดยมีค่าอยู่ในเซต $\{-1, +1\}$ สำหรับกรณีการจำแนกสองกลุ่ม
x_q	คือ เวกเตอร์คุณลักษณะของตัวอย่างลำดับที่ q
b_{hyp}	คือ ค่าคงที่
$\ w_{hyp}\ $	คือ ค่านอร์มแบบยุคลิดของเวกเตอร์น้ำหนัก

เงื่อนไข $y_q (w_{hyp}^{Transpose} x_q + b_{hyp}) \geq 1$ ใช้กำหนดให้ข้อมูลทุกตัวอย่างถูกจำแนกอย่างถูกต้อง และอยู่บนหรือภายนอกเส้นระยะขอบ

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

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

เมื่อ

w_{hyp}	คือ เวกเตอร์ค่าน้ำหนักของไฮเปอร์เพลน
y_{sv}	คือ ป้ายกำกับจริงของตัวอย่างลำดับที่ sv โดยมีค่าอยู่ในเซต $\{-1, +1\}$ สำหรับกรณีการจำแนกสองกลุ่ม
x_{sv}	คือ เวกเตอร์คุณลักษณะของตัวอย่างลำดับที่ sv
b_{hyp}	คือ ค่าคงที่
n_{svm}	คือ จำนวนตัวอย่างข้อมูลทั้งหมด
ξ_{sv}	คือ ค่าตัวแปรผ่อนผันของตัวอย่างลำดับที่ i ใช้วัดระดับการละเมิดเงื่อนไขระยะขอบ

C_{svm} คือ พารามิเตอร์ควบคุมการแลกเปลี่ยนระหว่างความกว้างของระยะขอบและความผิดพลาดจากการจำแนก

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

เคอร์เนลและซัพพอร์ตเวกเตอร์แมชชีนสำหรับข้อมูลไม่เชิงเส้น เพื่อแปลงข้อมูลจากปริภูมิดั้งเดิมไปยังปริภูมิที่แยกได้ดีขึ้นโดยไม่ต้องคำนวณการแปลงอย่างชัดเจน ฟังก์ชันตัดสินใจของซัพพอร์ตเวกเตอร์แมชชีน ในรูปแบบทวิภาค ดังสมการที่ (2.18)

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

เมื่อ

$f(x_{pre})$ คือ ฟังก์ชันตัดสินใจสำหรับการจำแนกกลุ่มของข้อมูลตัวอย่าง x_{pre}
 α_d คือ ค่าสัมประสิทธิ์จากการแก้ปัญหาในรูปทวิภาค
 γ_d คือ บ้ายกำกับจริงของตัวอย่างลำดับที่ 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}
 $\text{sign}(\cdot)$ คือ ฟังก์ชันกำหนดกลุ่มผลลัพธ์จากเครื่องหมายของค่าที่คำนวณได้

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

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

เมื่อ

γ คือ พารามิเตอร์ที่ควบคุมความกว้างของเคอร์เนล

จากข้อจำกัดของแบบจำลองเชิงลึกที่มุ่งเน้นการจำแนกเพียงอย่างเดียว และ ศักยภาพของการเรียนรู้หลายงานผ่านออโตเอนโค้ดหลายงาน งานวิจัยนี้จึงเสนอแนวทางซึ่งผสาน ข้อดีของการเรียนรู้ตัวแทนเชิงแฝงและการตัดสินใจแบบมาร์จินเข้าด้วยกัน โดยมีแนวคิดหลัก ได้แก่ ใช้ออโตเอนโคเดอร์แบบหลายงานเพื่อเรียนรู้ตัวแทนเชิงแฝง ที่รักษาโครงสร้างของข้อมูลเสียงปอด ผ่านงานการสร้างกลับคืน พร้อมกับเสริมความสามารถในการจำแนกผ่านงานจำแนกประเภท และใช้ ซัพพอร์ตเวกเตอร์แมชชีนเป็นตัวจำแนกขั้นสุดท้ายแทนชั้นเชื่อมโยงสมบูรณ์ของโครงข่าย เพื่อเพิ่ม เสถียรภาพของการจำแนก โดยเฉพาะในกรณีที่ข้อมูลมีจำนวนจำกัดและมีลักษณะไม่เชิงเส้น ตัวแทน เชิงแฝงที่ได้จากออโตเอนโคเดอร์แบบหลายงานสามารถมองเป็นคุณลักษณะเชิงลึกที่เรียนรู้โดยตรง จากข้อมูลเสียงปอด จากนั้นจึงนำไปใช้เป็นอินพุตของซัพพอร์ตเวกเตอร์แมชชีนเพื่อกำหนดขอบเขต การตัดสินใจในปริภูมิของตัวแทนเชิงแฝง ดังแสดงในสมการที่ (2.20)

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

โดยในสมการที่ (2.20) ตัวจำแนกซัพพอร์ตเวกเตอร์แมชชีน ทำงานบนเวกเตอร์ ตัวแทนเชิงแฝง แทนการทำงานบนข้อมูลอินพุตโดยตรง ซึ่งช่วยให้การตัดสินใจอาศัยคุณลักษณะที่มี มิติและโครงสร้างเหมาะสมต่อการหาขอบเขตการตัดสินใจแบบระยะห่างสูงสุด

2.5 ปรัชญาวรรณกรรม

จากการทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องด้านการตรวจจับโรคระบบทางเดินหายใจ จากเสียงปอด พบว่างานวิจัยในสาขานี้มีการศึกษาอย่างต่อเนื่อง ทั้งในด้านการประมวลผลสัญญาณ การสกัดคุณลักษณะ และแบบจำลองการเรียนรู้ของเครื่อง โดยสามารถจัดกลุ่มแนวทางที่สำคัญตาม ลักษณะวิธีการ ได้แก่ แนวทางที่อาศัยการสกัดคุณลักษณะร่วมกับการเรียนรู้ของเครื่องแบบดั้งเดิม ซึ่ง มุ่งออกแบบตัวแทนสัญญาณ เช่น คุณลักษณะเชิงเวลาและความถี่ และเซปสตรีม แล้วใช้ตัวจำแนกใน การแยกกลุ่มเสียง การเรียนรู้เชิงลึกแบบต้นทางถึงปลายทาง (End-to-end deep learning) ที่ให้ แบบจำลองเรียนรู้รูปแบบเสียงจากตัวแทนสัญญาณ เช่น สเปกโตรแกรม หรือจากสัญญาณดิบโดยตรง เพื่อเพิ่มศักยภาพในการจำแนก และแนวทางที่เริ่มผสานการเรียนรู้เชิงตัวแทน (Representation learning) หรือการเรียนรู้แบบหลายงาน เพื่อบรรจุข้อจำกัดของข้อมูลติดป้ายกำกับที่มีจำนวนจำกัด และเพิ่มความทนทานของแบบจำลองต่อความแปรปรวนของข้อมูล แนวทางเหล่านี้ช่วยยกระดับ

สมรรถนะของระบบอย่างมีนัยสำคัญ งานวิจัยบางส่วนยังพบข้อจำกัดเกี่ยวกับความไม่สมดุลของข้อมูล ในหัวข้อต่อไปจึงนำเสนอการปฏิสน์วรรณกรรมโดยจัดเรียงตามกลุ่มแนวทางดังกล่าว 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 แม้ว่ายังจำกัดอยู่กับการจำแนกแบบสองกลุ่มและใช้การออกแบบคุณลักษณะโดยบุคคล

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) เป็นตัวจำแนกข้อมูลทดลองประกอบด้วยเสียงปอดประเภทต่าง ๆ เช่น เสียงภาวะปกติ เสียงหวีด และเสียงแตกกรอบ ผลการทดลองสัมประสิทธิ์เซปสตรัมความถี่เมลให้ผลการจำแนกที่ดีที่สุด โดยมีค่าความถูกต้องมากกว่า 92 เปอร์เซ็นต์ และมีความไวต่อเสียงภาวะผิดปกติที่มีประสิทธิภาพ แสดงให้เห็นบทบาทสำคัญของการเลือกคุณลักษณะที่เหมาะสมต่อประสิทธิภาพของการจำแนกเสียงปอด

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

ทดลองพบว่าวิธีนี้ให้ประสิทธิภาพสูงกว่าวิธีแบบจำลองมาร์คอฟซิดเดนที่ใช้การผสมแบบเกาส์เซียนอย่างมีนัยสำคัญ โดยให้ความถูกต้องสูงขึ้นและลดอัตราความผิดพลาดลง งานนี้แสดงให้เห็นศักยภาพของการผสมการเรียนรู้เชิงลึกเข้ากับแบบจำลองเชิงลำดับเวลาในงานวิเคราะห์เสียงปอด

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 โดยเฉพาะในการจำแนกเสียงภาวะผิดปกติแสดงให้เห็นบทบาทของสถาปัตยกรรมในการเพิ่มความสามารถของแบบจำลองการเรียนรู้เชิงลึก

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

0.94 แสดงให้เห็นความสำคัญของการรวมกระบวนการประมวลผลสัญญาณเข้ากับการจำแนกเสียงพูดแบบอัตโนมัติ

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

Soni, N., Chowdhury, A., Bhatia, M., and Chakrabarti, I. (2022) ศึกษาปัญหาการขาดแคลนข้อมูลที่มีการติดป้ายกำกับในงานวิเคราะห์เสียงพูด โดยเสนอแนวทางการเรียนรู้แบบไม่ต้องใช้ป้ายกำกับด้วยเทคนิคการเรียนรู้แบบเปรียบเทียบ (Contrastive learning) เพื่อเรียนรู้ตัวแทนข้อมูลของสัญญาณเสียง วิธีการใช้ข้อมูลทางคลินิกเพื่อกำหนดคู่ข้อมูลเชิงบวกและเชิงลบ ผลการทดลองพบว่าแนวทางการเรียนรู้แบบกำกับตนเอง (Self-supervised learning) ที่นำเสนอสามารถปรับปรุงประสิทธิภาพของการจำแนกในขั้นตอนปลายน้ำ (Downstream task) ได้อย่างมีนัยสำคัญ โดยเฉพาะในกรณีที่มีข้อมูลติดป้ายกำกับเพียงส่วนน้อย แสดงให้เห็นศักยภาพของการเรียนรู้ตัวแทนข้อมูลสำหรับการวิเคราะห์เสียงพูดภายใต้ข้อจำกัดของข้อมูล

Hassan, U., Singhal, A., and Chaudhary, P. (2024) นำเสนอ EasyNet ซึ่งเป็นโครงข่ายประสาทเทียมเชิงลึกที่มีโครงสร้างเรียบง่ายสำหรับการจำแนกเสียงพูด โดยออกแบบให้เรียนรู้จากสัญญาณเสียงพูดแบบดิบโดยตรงโดยไม่ต้องแปลงเป็นสเปกโตรแกรมหรือสกัดคุณลักษณะด้วยมือ การทดลองครอบคลุมการจำแนกแบบสองกลุ่ม แบบหลายกลุ่ม และแบบหลายป้ายกำกับ บนฐานข้อมูลเสียงพูดหลายชุด ผลการทดลองแสดงให้เห็นว่าแบบจำลองที่นำเสนอให้ค่าความถูกต้อง ความไว และความจำเพาะในระดับสูง งานนี้ชี้ให้เห็นว่าแบบจำลองที่มีโครงสร้างไม่ซับซ้อนสามารถให้ประสิทธิภาพที่แข่งขันได้ในการวิเคราะห์เสียงพูด

จากการปริทัศน์วรรณกรรมพบว่า งานวิจัยจำนวนมากสามารถพัฒนาแบบจำลองเพื่อวิเคราะห์เสียงพูดได้อย่างมีประสิทธิภาพภายใต้เงื่อนไขการทดลองของตนเอง แต่เมื่อพิจารณาเชิงลึกยังพบประเด็นที่น่าสนใจหลายด้าน ได้แก่ ความแปรปรวนของสัญญาณจากแหล่งบันทึกและสภาวะแวดล้อมที่แตกต่างกัน ความไม่สมดุลของข้อมูลซึ่งส่งผลกระทบต่อความไวของการตรวจจับกลุ่มผิดปกติ ตลอดจนความจำเป็นในการกำหนดขั้นตอนประเมินผลให้สะท้อนการใช้งานจริง โดยเฉพาะการออกแบบกระบวนการเตรียมข้อมูล การกำหนดพารามิเตอร์ เช่น เฟรม ช่วงสัญญาณ รอบการหายใจ

เป็นต้น การเลือกตัวชี้วัดที่เหมาะสมเพื่อประเมินความสามารถของแบบจำลอง นอกจากนี้ แนวทางดังกล่าวยังแสดงให้เห็นความสำคัญของการเรียนรู้ตัวแทนสัญญาณที่ดีและการออกแบบสถาปัตยกรรมที่สามารถเน้นสารสนเทศสำคัญของเสียงภาวะผิดปกติได้ ซึ่งเป็นปัจจัยที่มีผลต่อประสิทธิภาพและความน่าเชื่อถือ