



การออกแบบและพัฒนาการจำแนกรังไหมดีและรังไหมเสียด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชันและการเรียนรู้แบบถ่ายโอน

A Design and Development of Good and Bad Cocoon Classification Models using Convolutional Neural Network and Transfer Learning

รัชชานนท์ สิมสวัสดิ์ (Ratchanon Simsawat)* และอรอุมา พรำโม่ต (On-Uma Pramote)*

Received: April 9, 2024
Revised: November 3, 2024
Accepted: November 13, 2024

* ผู้นิพนธ์ประสานงาน: ออรอุมา พรำโม่ต (On-Uma Pramote) อีเมล: onbee@psru.ac.th

DOI:10.14416/j.it.2026.v1.008

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อออกแบบและพัฒนาการจำแนกรังไหมดีและรังไหมเสียด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชันและการเรียนรู้แบบถ่ายโอน ประกอบด้วย 3 ขั้นตอน ได้แก่ 1) การจัดเก็บรวบรวมข้อมูล 2) การเตรียมข้อมูล และ 3) การพัฒนาแบบจำลองจำแนกเสียง โดยข้อมูลที่ใช้เป็นไฟล์เสียงนามสกุล WAV บันทึกเสียงโดยใช้ไมโครโฟนและโปรแกรมคอมพิวเตอร์ซึ่งให้ผู้เชี่ยวชาญเขย่าจากรังไหมดีจำนวน 500 ไฟล์ และรังไหมเสียจำนวน 500 ไฟล์ รวมทั้งหมด 1,000 ไฟล์ ข้อมูลถูกแปลงจากข้อมูลเสียงเป็นข้อมูลภาพด้วยวิธีสเปกตรัมเมลได้ 1,000 ภาพ และตัดภาพด้วยวิธีการ 2 วิธี ได้แก่ 1) การหาคึ่งกลางภาพ และ 2) การปรับขนาดภาพ ใช้แบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชันและแบบจำลองการเรียนรู้แบบถ่ายโอน 2 วิธี ได้แก่ MobileNetV2 และ NASNetMobile ในการพัฒนาแบบจำลองโดยข้อมูลภาพถูกแบ่งเป็น 2 ชุดในอัตราส่วน 80 ต่อ 20 ใช้เป็นข้อมูลสำหรับฝึกสอนแบบจำลอง 800 ภาพ และข้อมูลทดสอบแบบจำลอง 200 ภาพ ผลการวัดประสิทธิภาพแบบจำลองพบว่าวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชันให้ค่าความถูกต้องร้อยละ 100 ด้วยภาพที่ได้หลังการตัดภาพจากทั้ง 2 วิธี ยังคงคุณลักษณะเด่นอยู่บริเวณกลางภาพและภาพที่ใช้ในการฝึกสอนเป็นภาพของสเปกตรัมเมลรังไหมดีและรังไหมเสียทำให้ลดการรบกวนและอคติจากสัญญาณภาพแบบอื่นในแบบจำลอง ส่วนการเรียนรู้การถ่ายโอน วิธี MobileNetV2 ให้ค่าความถูกต้องร้อยละ 95 และ วิธี NASNetMobile ให้ค่าความถูกต้องร้อยละ 94.5

คำสำคัญ: การจำแนกเสียง โครงข่ายประสาทเทียมแบบคอนโวลูชัน การเรียนรู้แบบถ่ายโอน รังไหม

Abstract

This research aims to design and development of good and bad cocoon classification models using convolutional neural network and transfer learning. There are 3 steps: 1) data acquisition, 2) data preparation, and 3) model development. The data used was a WAV audio file, recorded using a microphone and computer program, which had experts shake out 500 good cocoons and 500 bad cocoons, for a total of 1000 files. The data was converted from audio data into images using Mel spectrogram technique, The data was converted from audio data into images using the Mel Spectrum technique into 1000 images and crop the image using 2 methods, including 1) finding the center of the image and 2) resizing the image. Using a convolutional neural network. and transfer learning in modeling. The data is divided into two parts, 800 images of training data and 200 images of test data. The results show that the convolutional neural network model has an accuracy of 100 percent, Since the images obtained after cropping the images from both methods, there are still prominent features in the center of the image and the images used for training are images of the spectrum of good and bad cocoons, reducing noise or bias in classification model. As for the MobileNetV2 transfer learning model has an accuracy of 95 percent, and the NASNetMobile has an accuracy of 94.5 percent.

* หลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏพิบูลสงคราม

* Computer Science Program, Faculty of Science and Technology, Pibulsongkram Rajabhat University.

Keywords: Audio classification, Convolutional Neural Network, Transfer Learning, Cocoon.

1. บทนำ

ไหม (Silkworm) เป็นแมลงมีประโยชน์ทางเศรษฐกิจของประเทศ รังไหมเป็นเส้นใยธรรมชาติใช้ทำเสื้อผ้าเครื่องนุ่งห่ม เคหะสิ่งทอต่าง ๆ สารสกัดจากรังไหมสามารถใช้เป็นองค์ประกอบในเครื่องสำอางและอาหารเพื่อสุขภาพ ใช้ประโยชน์ทางการแพทย์และการศึกษาวิจัยทางวิทยาศาสตร์ ดักแต่ไหมมีคุณค่าทางโภชนาการสูงอุดมด้วยโปรตีนและไขมันประเภทไม่อิ่มตัว ใช้เป็นอาหารทั้งของคนและสัตว์เศรษฐกิจที่สำคัญคือ ฝ้ายไหมไทยมีชื่อเสียงไปทั่วโลกในด้านความสวยงามเป็นเอกลักษณ์เป็นสิ่งทอหลักที่นารายได้เข้าประเทศ การเลี้ยงไหมทำให้เกษตรกรมีอาชีพสร้างรายได้สำหรับดำรงชีวิตไม่ละถิ่นฐานไปทำงานที่อื่นมีโอกาสสร้างครอบครัวและชุมชนที่เข้มแข็งมั่นคง พึ่งพาตนเองได้ นอกจากนี้ไหมยังช่วยสร้างรายได้ให้กับผู้ประกอบการภาคเอกชนที่เกี่ยวข้องในอุตสาหกรรมสิ่งทอและการสร้างผลิตภัณฑ์จากไหม [1]

ประเทศไทยมีกรมหม่อนไหมเป็นหน่วยงานที่ดูแลงานด้านหม่อนไหมของประเทศ เพื่อให้งานด้านหม่อนไหมของไทยมีความยั่งยืน และมีศูนย์ศิลปาชีพบางไทรเป็นแหล่งจำหน่ายผลิตภัณฑ์จากไหม นอกจากนี้ยังมีบริษัทเอกชนที่ทำธุรกิจด้านไหม เช่น บริษัท จุลไหมไทย [2] การได้มาซึ่งผลิตภัณฑ์ไหมที่ดีมีปัจจัยสำคัญคือ รังไหม โดยรังไหมที่ดีหรือมีคุณภาพต้องเป็นรังไหมที่มีหนอนไหมเพราะหนอนไหมเป็นตัวผลิตเส้นใย ดังนั้นระหว่างการผลิตรังไหมต้องดูแลหนอนไหมไม่ให้ตายก่อนขยารัง และก่อนขยารังต้องตรวจสอบก่อนเพื่อขยารังเฉพาะรังไหมที่มีตัวหนอน ซึ่งการตรวจสอบในปัจจุบันเกษตรกรใช้วิธีเขยารังเพื่อฟังเสียงเป็นความสามารถที่ต้องใช้เวลาในการฝึกฝนและฝึกสอน ส่งผลให้มีเกษตรกรน้อยรายที่เชี่ยวชาญในทักษะด้านนี้ ซึ่งการผลิตรังไหมในแต่ละครั้งมีจำนวนไม่ต่ำกว่า 50,000 ตัว ทำให้ต้องใช้เวลาในการตรวจสอบนาน

รัฐบาลไทยได้ส่งเสริมการใช้เทคโนโลยีหรือนวัตกรรมให้กับเกษตรกรทั้งสตาร์ทอัพสายเกษตรหรือส่งเสริมให้เกษตรกรเป็นเกษตรกรอัจฉริยะใช้ข้อมูล และปัญญาประดิษฐ์ในการทำเกษตรเพิ่มขึ้น และพยายามผลักดันให้เกษตรกรนำมาใช้ได้จริง [3] เทคโนโลยีปัญญาประดิษฐ์ในปัจจุบัน

เน้นการใช้ภาพถ่ายเป็นข้อมูลนำเข้าโดยวิธีที่นิยม ได้แก่ วิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน และการเรียนรู้แบบถ่ายโอน ซึ่งมีงานวิจัยที่ใช้ปัญญาประดิษฐ์ทั้ง 2 แบบในงานด้านเกษตรกรรมที่หลากหลาย ได้แก่ การจำแนกโรคทางใบ [4] - [6] การจำแนกสายพันธุ์ [7] การจำแนกระดับความสุก [8] - [9]

จากปัญหาและเทคโนโลยีข้างต้น ผู้วิจัยจึงเสนอการออกแบบและพัฒนาการจำแนกรังไหมดี และรังไหมเสียโดยการวิเคราะห์ภาพสเปกโทรแกรมของเสียงด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน และการเรียนรู้แบบถ่ายโอนเพื่อช่วยจำแนกรังไหมที่ตัวหนอนยังมีชีวิต กับรังไหมที่ตัวหนอนตายแล้ว ลดการใช้ผู้เชี่ยวชาญเฉพาะทางช่วยในการจำแนกมีความถูกต้องรวดเร็ว ลดภาระค่าใช้จ่ายของเกษตรกร

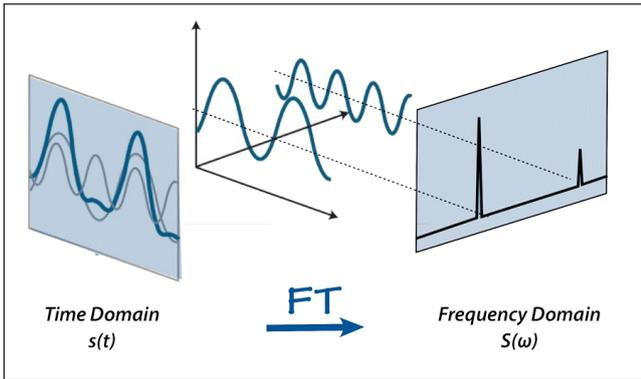
2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 เสียงและสเปกตรัมเมล (Sound and Mel Spectrogram)

เสียงเป็นพลังงานที่เกิดจากการอัดและการขยายโมเลกุลอากาศสลับกัน ซึ่งเป็นความเปลี่ยนแปลงตามการเคลื่อนที่ของโมเลกุลอากาศในรูปแบบคลื่น ทำให้มีคุณลักษณะเฉพาะของเสียงที่สามารถระบุได้ เช่น ความยาวช่วงคลื่น แอมพลิจูด ความเร็ว ความเข้มเสียง [10] ส่งผลให้เสียงมีความแตกต่างกัน เช่น เสียงสูง เสียงต่ำ เสียงดัง เสียงเบา สัญญาณเสียงอยู่ในโดเมนเวลา ในงานวิจัยนี้จะแปลงสัญญาณจากโดเมนเวลาเป็นโดเมนความถี่ ด้วยวิธีการแปลงฟูรีเย (Fourier Transform) ส่งผลให้ได้ความถี่และแอมพลิจูดของความถี่ที่ถูกแสดงผลในแบบสเปกตรัม (Spectrum) [11] ซึ่งในภาพที่ 1 สัญญาณเสียงในโดเมนเวลาจะแสดงเป็นคลื่นไซน์แต่ละคลื่นมีความสูงแตกต่างกัน เพราะในคลื่นเสียงสามารถมีหลายความยาวช่วงคลื่นหลายแอมพลิจูดได้ (เสียงที่สนใจ เสียงรบกวน) เมื่ออยู่ในโดเมนความถี่คลื่นเสียงจะถูกแยกออกจากกันให้สามารถเลือกความถี่ที่สนใจได้

สเปกตรัมหรือแผนภาพของสเปกตรัม นิยมใช้กับการแสดงผลของแถบคลื่นความถี่ของเสียง คลื่นแม่เหล็กไฟฟ้าหรือคลื่นของแสง เพราะแสดงความถี่ของสัญญาณตามเวลา ทำให้เข้าใจได้ง่าย โดยแนวนอนเป็นแกนเวลา แนวตั้งเป็นความถี่ที่มีหน่วยเป็นเฮิรตซ์ (Hz) [12] ในงานวิจัยนี้ได้นำสเปกตรัมแปลงเป็นสเกลเมล ซึ่งเป็นสเกลที่ให้ความสำคัญช่วงความถี่ต่ำ

มากกว่าช่วงความถี่สูง จึงเก็บรายละเอียดของสัญญาณเสียง ช่วงความถี่ต่ำได้มากกว่า [13]



ภาพที่ 1 การแปลงสัญญาณเวลาเป็นสัญญาณความถี่

2.2 ไบรราลี ลิปโรซ่า (Librosa)

ไลบรารี ลิปโรซ่า เป็นโอเพ่นซอร์สไลบรารีสำหรับวิเคราะห์การประมวลผลเสียงและเพลง มีฟังก์ชันใช้ในการทำงาน ได้แก่ อ่านไฟล์เสียงในรูปแบบต่าง ๆ คำนวณสเปกโตรแกรม สเปกโตรแกรมเมล และโครมาแกรม ดึงข้อมูลคุณลักษณะจากข้อมูลเสียง เช่น ค่าสัมประสิทธิ์เซปสตรัลความถี่เมล เซนทรอยด์สเปกตรัม อัตราการข้ามจุดศูนย์ มียูทิลิตี้สำหรับการสร้างภาพข้อมูลเสียง เป็นต้น [14]

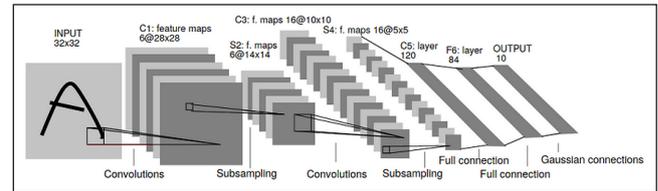
2.3 โครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network: CNN)

โครงข่ายประสาทเทียมแบบคอนโวลูชันเป็นโครงข่ายประสาทเทียมหนึ่งในกลุ่ม Bio-inspired มีหลักการการทำงานคือ มองภาพเป็นพื้นที่ย่อย ๆ และพิจารณาแยกคุณลักษณะ (Feature) พื้นที่ย่อยนั้น เช่น ลายเส้น การตัดกันของสี โดยใช้คอนโวลูชันเชิงพื้นที่ (Spatial Convolution) เพื่อดึงคุณลักษณะออกมาและนำมาผสมกัน โครงสร้างของโครงข่ายประสาทเทียม (Neural Network) ประกอบด้วย 3 ชั้น ดังนี้ 1) ชั้นนำเข้า (Input Layer) 2) ชั้นซ่อน (Hidden Layer) โดยในชั้นนี้แยกย่อยได้อีกคือ ชั้นคอนโวลูชัน (Convolutional Layer) ชั้นพูลลิ่ง (Pooling Layer) ชั้นเชื่อมโยงแบบสมบูรณ์ (Fully-Connected Layer) และ 3) ชั้นแสดงผลลัพธ์ (Output Layer) หรือเรียกว่า Multi-Layer Perceptron (MLP) [15] ดังแสดงในภาพที่ 2

2.4 การเรียนรู้การถ่ายโอน (Transfer Learning: TL)

การเรียนรู้แบบถ่ายโอนเป็นการใช้แบบจำลองที่ถูกสร้างฝึกและปรับแต่งไว้แล้ว มาปรับใช้ใหม่สำหรับงานใหม่

การเรียนรู้การถ่ายโอนมี 2 ประเภท ได้แก่ การเรียนรู้การถ่ายโอนแบบ Transductive และการเรียนรู้การถ่ายโอนแบบ Inductive ในงานวิจัยนี้ใช้การเรียนรู้การถ่ายโอนแบบ Inductive เช่น VGG, Xception, MobileNet ซึ่งได้รับการฝึกล่วงหน้าบนชุดข้อมูลภาพขนาดใหญ่จากนั้นจะถูกปรับแต่งเพื่อพัฒนาการจำแนกวัตถุ [16]



ภาพที่ 2 โครงสร้างของโครงข่ายประสาทเทียม

2.5 งานวิจัยที่เกี่ยวข้อง

ณัฐนิชา ขอบใจ และคณะ [17] นำเสนอเรื่องการจำแนกเสียงของแมลง 5 ชนิด ได้แก่ แมลงกระชอน ต๊กแตน จิ้งหรีด จิ้งโกร่ง แมลงดานา ซึ่งใช้การสกัดคุณลักษณะข้อมูลโดยข้อมูลนำเข้าได้มาจากวิดีโอเสียงของแมลง ใน YouTube 14 ไฟล์ แต่ละไฟล์ถูกตัดเป็น ไฟล์ย่อยมีความยาว 5 วินาที ได้ไฟล์ทั้งสิ้น 2,500 ไฟล์ แบ่งเป็นข้อมูลตั้งนี้ Train Set ร้อยละ 64 Validation Set ร้อยละ 16 และ Test Set ร้อยละ 20 มีการเปรียบเทียบวิธี 2 วิธี ได้แก่ แบบจำลองโครงข่ายประสาท (Neural Networks) กับแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Network) ซึ่งผลการทดลองโครงข่ายประสาทเทียมแบบคอนโวลูชันให้ค่าความถูกต้องร้อยละ 99.80 ส่วนโครงข่ายประสาทให้ค่าความถูกต้องร้อยละ 85.20

อภิสรา ภิรมย์โต และคณะ [18] นำเสนอเรื่อง การพัฒนาแบบจำลองจำแนกเสียงนกด้วยวิธีการเรียนรู้เชิงลึก โดยมีเสียงนก 20 ชนิด ในอุทยานแห่งชาติเขาใหญ่ จำนวน 1,613 ไฟล์ ได้มาจากเว็บไซต์ xenocanto และนำมาตัดให้สั้นเหลือไฟล์ละ 8 วินาที ได้ไฟล์ทั้งสิ้น 4,800 ไฟล์ แล้วนำมาสกัดคุณลักษณะและใช้วิธีโครงข่ายประสาทเทียมคอนโวลูชันในการสร้างแบบจำลองจำแนก ซึ่งแบ่งข้อมูลสำหรับ Train Set ร้อยละ 60 Validation Set ร้อยละ 20 และ Test Set ร้อยละ 20 ได้ค่าความถูกต้องของแบบจำลองร้อยละ 90.62

ธีรภัทร กาญจนาคม และคณะ [19] นำเสนอเรื่อง การจำแนกเสียงของสุนัขด้วยอุปกรณ์ประมวลผลปัญญาประดิษฐ์

บนขอบเครือข่าย เพื่อให้เข้าใจความหมายที่สุนัขสื่อสารได้ไฟล์จากการอัดเสียงสุนัขอายุ 12 เดือนขึ้นไป เป็นสุนัขโตเต็มวัย และเสียงต้องมีความดังประมาณ 44.10 กิโลเฮิร์ต ไฟล์เสียงมีความยาว 5 - 7 วินาที โดยแบ่งกลุ่มเสียงเป็น 5 กลุ่ม ได้แก่ เสียงของสุนัขเมื่อได้รับความเจ็บปวด เสียงเห่าของสุนัข เสียงทอนของสุนัข เสียงขู่ของสุนัข และเสียงอื่น ๆ ใช้วิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน ได้ค่าความถูกต้องของ Train Data ร้อยละ 94 และ Test Data ร้อยละ 87

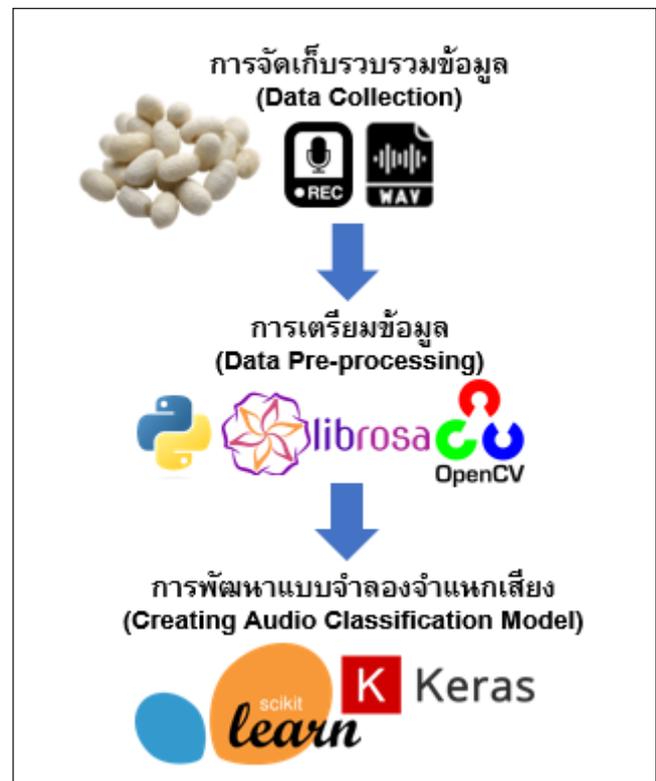
ภูวตล เขียงจันทร์ และจรัสศรี รุ่งรัตนอุบล [20] นำเสนอเรื่องโปรแกรมรู้จำเสียงสัตว์ 9 ชนิด ได้แก่ แมว หมู วัว ช้าง ไก่ แกะ ลิง กบ และสุนัข ข้อมูลเสียงแต่ละชนิดมี 30 ไฟล์ รวม 270 ไฟล์ อัดและพัฒนาแบบจำลองด้วยทีชเอเบิล แมทซอิน แบ่งเป็นข้อมูล Train จำนวน 189 ไฟล์ และ ข้อมูล Test จำนวน 81 ไฟล์ และนำไปประยุกต์ใช้บนแอนดรอยด์ แอปพลิเคชันที่พัฒนาด้วยโปรแกรม Flutter โดยแบบจำลองให้ค่าความถูกต้องเมื่อทดสอบผ่านคอมพิวเตอร์ ร้อยละ 71.6 และให้ค่าความถูกต้องในชุดข้อมูลเดียวกันที่ทดสอบผ่านโทรศัพท์แอนดรอยด์ ร้อยละ 37

ชั้นนี้ กวินเสกสรรค์ และคณะ [21] นำเสนอเรื่องการจำแนกเสียงอ่านร้อยกรองไทย โดยใช้โครงข่ายประสาทเทียม โดยแยกการอ่านทำนองเสนาะ และการขับเสภา ใช้การสกัดคุณลักษณะด้วยวิธี Neural Network ได้ข้อมูลจากวิดีโอ การอ่านทำนองเสนาะ และการขับเสภาใน YouTube แบ่งเป็นการอ่านทำนองเสนาะ 7 ไฟล์ การขับเสภา 5 ไฟล์ และนำมาตัดเป็นไฟล์ขนาด 10 วินาที ได้ทั้งหมด 271 ไฟล์ ซึ่งในการแบ่งข้อมูลแบ่งเป็นข้อมูลชุดฝึก ร้อยละ 80 และข้อมูลชุดทดสอบ ร้อยละ 20 ได้ค่าความถูกต้องร้อยละ 100

จากการทบทวนวรรณกรรมข้างต้นพบว่างานวิจัยจำแนกเสียงนำมาใช้จำแนกได้ทั้งเสียงคน สัตว์ หรือแมลง โดยการเก็บข้อมูลจากเว็บไซต์หรือบันทึกเสียงเอง ซึ่งใช้วิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน และแบ่งชุดข้อมูลสำหรับการสร้างแบบจำลองจำแนกในอัตราส่วนข้อมูลฝึกสอนร้อยละ 80 และข้อมูลทดสอบร้อยละ 20 ดังนั้นผู้วิจัยจึงนำการทบทวนข้างต้นมาบูรณาการในงานวิจัยเสียงรังใหม่ที่เก็บข้อมูลเอง โดยใช้วิธีสำหรับการสร้างแบบจำลองและการแบ่งชุดข้อมูลเดียวกับวรรณกรรมที่ได้ทบทวน

3. วิธีดำเนินการวิจัย

การออกแบบและพัฒนาระบบการจำแนกรังใหม่ดี และรังใหม่เสียด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน และการเรียนรู้แบบถ่ายโอนมีกรอบแนวคิดซึ่งประกอบไปด้วย 3 ส่วน ดังนี้ ส่วนแรก การจัดเก็บรวบรวมข้อมูล (Data Collection) เป็นการเก็บบันทึกข้อมูลเสียงของรังใหม่ทั้งรังดีและรังเสียให้อยู่ในรูปแบบดิจิทัลที่สามารถนำไปประมวลผลในคอมพิวเตอร์ได้ ส่วนที่สอง การเตรียมข้อมูล (Data Pre-processing) เป็นการทำความสะอาดข้อมูล เช่น ลดเสียงรบกวน จัดการสัญญาณเสียงให้เหมาะสมกับการสร้างแบบจำลอง ส่วนที่สาม การพัฒนาแบบจำลอง (Creating Audio Classification Model) เป็นการพัฒนาแบบจำลองโดยแบ่งเป็น 2 ขั้นตอน ได้แก่ การฝึกสอนแบบจำลอง (Training) และการทดสอบแบบจำลอง (Testing) ดังแสดงในภาพที่ 3



ภาพที่ 3 กรอบแนวคิดของการจำแนกรังใหม่

การออกแบบและพัฒนาระบบการจำแนกรังใหม่ดี และรังใหม่เสียด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชันและการเรียนรู้แบบถ่ายโอนมีรายละเอียดดังนี้

3.1 การจัดเก็บรวบรวมข้อมูล

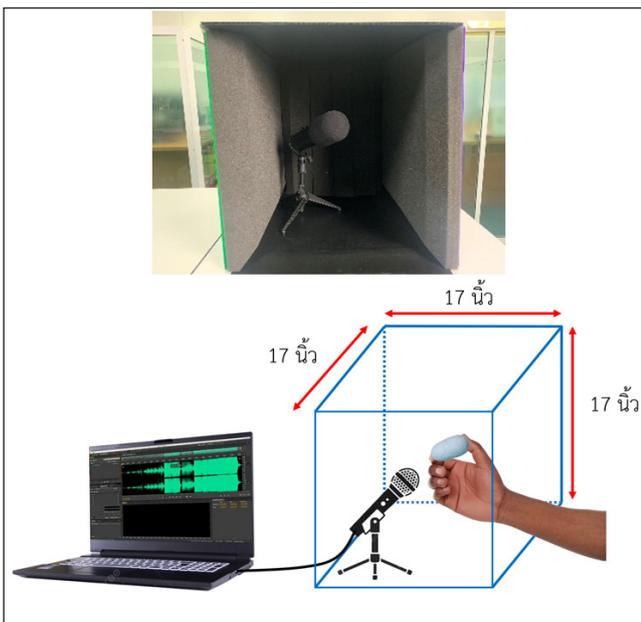
3.1.1 ข้อมูลที่ใช้ในงานวิจัย (Dataset) รังใหม่ที่ใช้ในงานวิจัยนี้เป็นพันธุ์ใหม่แม่พันธุ์สายพันธุ์จีน แบ่งออกเป็น 2 แบบ รังดี (รังใหม่ที่ตัวหนอนมีชีวิต) จำนวน 500 รัง

และรังเสีย (รังไหมที่ตัวหนอนตาย) จำนวน 500 รัง ดังแสดงในภาพที่ 4 ใช้ไฟล์เสียง 1 รัง 1 ไฟล์ ดังนั้นไฟล์เสียงทั้งหมดที่ใช้จำนวน 1,000 ไฟล์ แยกเป็นเสียงรังดี 500 ไฟล์ และเสียงรังเสีย 500 ไฟล์



ภาพที่ 4 ลักษณะของรังไหมดีและรังไหมเสีย

3.1.2 สภาพแวดล้อมที่ใช้ในการเก็บบันทึกข้อมูลเสียง ผู้วิจัยสร้างกล่องด้วยฟิวเจอร์บอร์ดมีขนาดความกว้าง ความยาว และความสูง 17 นิ้ว บุผนังด้านในด้วยแผ่นฟองน้ำซับเสียง (Acoustic Foam) เพื่อป้องกันการเสียงรบกวนจากภายนอก และเสียงสะท้อนภายใน รับข้อมูลเสียงด้วยไมโครโฟนคอนเดนเซอร์ยี่ห้อแฟนเทค (Fantech Microphone condenser) ดังแสดงในภาพที่ 5



ภาพที่ 5 สภาพแวดล้อมที่ใช้ในการบันทึกเสียง

การเขย่ารังไหมให้ผู้เชี่ยวชาญเป็นคนเขย่า และกำหนดผลลัพธ์ (Label) ของรังไหม โดยการเขย่าผู้เชี่ยวชาญจะยื่นมือที่ถือรังไหมเข้าไปในกล่องที่สร้างขึ้น การเขย่า 1 ครั้งคือการเคลื่อนที่ของตัวไหมไป - กลับ จากปลายรังด้านหนึ่งสู่อีกด้านหนึ่งนับเป็น 1 รอบ การวางรังไหมสำหรับการเขย่าจะวางรังในแนวนอนขนานกับพื้น ไฟล์เสียงที่ใช้ในการประมวลผลเป็นไฟล์นามสกุลเวฟ (WAVE)

3.2 การเตรียมข้อมูล

3.2.1 การตัดข้อมูลเสียง หลังการบันทึกเสียงได้ข้อมูลเป็นไฟล์ WAVE โดยแต่ละไฟล์มีเวลาในการบันทึกไม่เท่ากัน ดังนั้นต้องทำการเลือกช่วงข้อมูลเสียงโดยพิจารณาข้อมูลในมิติของเวลา คือ วินาที (Second) ข้อมูลเสียงรังไหมจะอยู่ช่วงกลางของไฟล์ เวลาที่ใช้บันทึกเสียงสั้นที่สุดเท่ากับ 1.09 วินาที และเวลาที่ใช้บันทึกเสียงยาวที่สุดเท่ากับ 3.83 วินาที ผู้วิจัยจึงพิจารณาเลือกความยาว 1 วินาที ในการทำการทดลอง ดังนั้นการเลือกข้อมูลจะพิจารณาหาที่กึ่งกลางเวลาในการบันทึกเสียงแต่ละไฟล์ และ ± 0.5 วินาที ดังชัดโค๊ดในภาพที่ 6 ส่วนในภาพที่ 7 เป็นตัวอย่างข้อมูลเสียงก่อนใช้ขั้นตอนวิธีการตัด แนวตั้งแสดงความดังของเสียง (เดซิเบล) แนวนอนแสดงระยะเวลาที่ใช้บันทึก ภาพที่ 8 เป็นข้อมูลเสียงหลังจากขั้นตอนวิธีการตัดแล้วจะเห็นได้ว่าข้อมูลที่ส่วนหัวและท้ายของไฟล์ซึ่งไม่มีเสียงของรังไหมถูกตัดออกไป

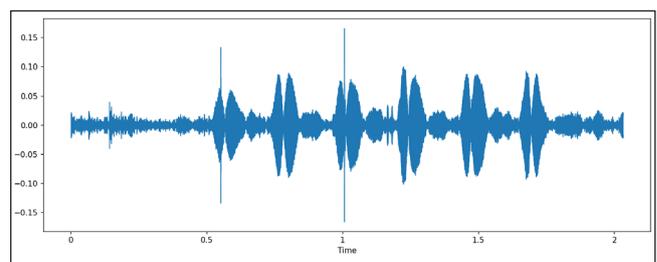
Algorithm: Finding the middle sound

Input: Sound File WAV

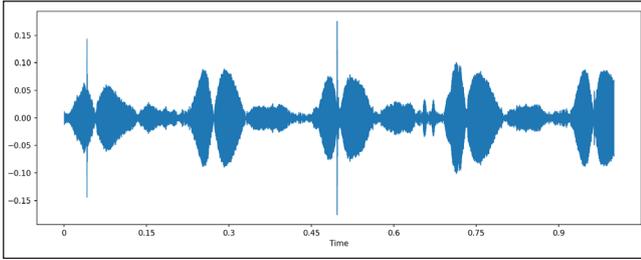
Output: Sound File WAV

- 1: Read data sound
- 2: $len_sound = \text{round}(\text{Data_sound.frame}/\text{Data_sound.samplerate}, 2)$
- 3: **if** $len_sound > 0.5$ **then**
- 4: $mid_time = \text{round}(len_sound/2), 2)$
- 5: $start_time = mid_time - 0.5$
- 6: $end_time = mid_time + 0.5$
- 7: $data = \text{int}((end_time - start_time) * \text{framerate})$
- 8: $new_sound = \text{int}(len(data)/\text{samplewidth})$
- 9: Write new_sound
- 10: **end**

ภาพที่ 6 ขั้นตอนวิธีการตัดไฟล์เสียง

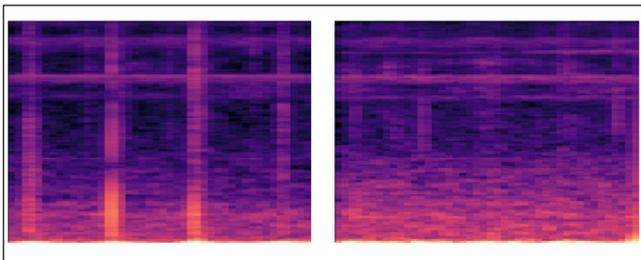


ภาพที่ 7 ข้อมูลเสียงก่อนใช้ขั้นตอนวิธีการตัด



ภาพที่ 8 ข้อมูลเสียงหลังใช้ขั้นตอนวิธีการตัด

3.2.2 การแปลงข้อมูลเสียงเป็นภาพไลบรารี ลิปโรซ่า จะอ่านข้อมูลไฟล์เสียงนามสกุลเวฟ (WAV) และแปลงให้เป็นอาร์เรย์ด้วยนัมพาย (Numpy) เพื่อคำนวณ สเปกโตรแกรมเมลสำหรับการวิเคราะห์เนื้อหาความถี่ของสัญญาณเสียงในช่วงเวลาต่าง ๆ และสร้างภาพจากฟังก์ชันการพล็อตของไลบรารี ลิปโรซ่า เพื่อสร้างภาพข้อมูลเสียง ดังแสดงในภาพที่ 9 และใช้ไลบรารีโอเพนซีวี (OpenCV) เพื่อบันทึกข้อมูลภาพลงในคอมพิวเตอร์สำหรับใช้สร้างแบบจำลองการจำแนกต่อไป



ภาพที่ 9 ภาพข้อมูลเสียงจากรังใหม่ดีและรังใหม่เสีย

3.3 การพัฒนาแบบจำลอง

ในงานวิจัยนี้พัฒนาแบบจำลองการจำแนกเสียงด้วยวิธีการ 2 วิธี ได้แก่ วิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน และวิธีการเรียนรู้การถ่ายโอน

3.3.1 การพัฒนาแบบจำลองการจำแนกเสียงด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน จะแบ่งข้อมูลภาพออกเป็น 2 ส่วน ได้แก่ ส่วนฝึกสอนแบบจำลอง (Training) และส่วนทดสอบแบบจำลอง (Testing) โดยข้อมูลจะถูกแบ่งเป็นส่วนฝึกสอน ร้อยละ 80 หรือจำนวน 800 ไฟล์ และส่วนทดสอบ ร้อยละ 20 หรือจำนวน 200 ไฟล์ ใช้วิธี 10 - Fold Cross Validation ในการแบ่งข้อมูลเข้าแบบจำลอง กำหนดพารามิเตอร์ดังนี้ ค่า Epochs เท่ากับ 20 ค่า Learning เท่ากับ 0.01 ค่า Batch size เท่ากับ 32 โดยมีโครงสร้างของโครงข่ายประสาทเทียมแบบคอนโวลูชันดังภาพที่ 10

Model: "sequential"		
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 254, 254, 16)	448
max_pooling2d (MaxPooling2D)	(None, 127, 127, 16)	0
conv2d_1 (Conv2D)	(None, 125, 125, 32)	4640
max_pooling2d_1 (MaxPooling2D)	(None, 62, 62, 32)	0
dropout (Dropout)	(None, 62, 62, 32)	0
flatten (Flatten)	(None, 123008)	0
dropout_1 (Dropout)	(None, 123008)	0
dense (Dense)	(None, 120)	14761080
dense_1 (Dense)	(None, 2)	242
=====		
Total params: 14,766,410		
Trainable params: 14,766,410		
Non-trainable params: 0		

ภาพที่ 10 โครงข่ายประสาทเทียมแบบคอนโวลูชันที่ใช้ในการทดลอง

3.3.2 การพัฒนาแบบจำลองการจำแนกเสียงด้วยการเรียนรู้แบบถ่ายโอน (Transfer Learning) เป็นวิธีที่ช่วยลดเวลาการเทรนแบบจำลอง Deep Learning ด้วยการนำบางส่วนของแบบจำลองที่ฝึกสอนเรียบร้อยแล้วกับงานที่ใกล้เคียงกัน มาใช้เป็นส่วนหนึ่งของแบบจำลองใหม่ ซึ่งในงานวิจัยนี้เลือก MobileNetV2 และ NASNetMobile มาใช้ในการทดสอบเพราะมีจำนวนพารามิเตอร์น้อยส่งผลให้ใช้เวลาในการประมวลผลเร็วกว่าแบบจำลองอื่น และมีค่าความถูกต้องอันดับ 1 (Top-1 Accuracy) มากกว่าร้อยละ 70 ค่าความถูกต้อง 5 อันดับแรก (Top-5 Accuracy) มากกว่าร้อยละ 90 โดยทั้งสองวิธีมีรายละเอียดดังตารางที่ 1

ตารางที่ 1 แสดงข้อมูลแบบจำลองการเรียนรู้การถ่ายโอนที่ใช้ในงานวิจัย

Model	Top-1 Accuracy	Top-5 Accuracy	Parameters
MobileNet V2	71.30 %	90.10%	3.54 M
NASNet Mobile	74.40 %	91.90 %	5.33 M

3.3.3 การวัดประสิทธิภาพ (Classifier Evaluation)

ตารางคอนฟิวชัน เมทริกซ์ (Confusion Matrix) เป็นตารางที่ใช้ประเมินประสิทธิภาพของแบบจำลองด้วยการประเมินค่าจริง (Actual Value) เทียบกับค่าผลลัพธ์ที่ได้จากแบบจำลอง (Predicted Value) โดยสรุปอยู่ในรูปแบบของตารางเมทริกซ์ซึ่งประกอบด้วยค่า 4 ค่า แบ่งเป็น 2 กลุ่ม คือ 1) กลุ่มที่สนใจ (Positive) โดยแยกพิจารณาได้เป็นกรณีย่อย ดังนี้ True Positive (TP) คือ ผลลัพธ์ไปในทางเดียวกัน ค่าจริง และแบบจำลองให้ผลลัพธ์เป็นสิ่งที่สนใจ True Negative (TN) ผลลัพธ์ไปในทางเดียวกัน ค่าจริง และแบบจำลองให้ผลลัพธ์เป็นสิ่งที่ไม่สนใจ 2) กลุ่มที่ไม่สนใจ (Negative) มีกรณีย่อย ดังนี้ False Positive (FP) ผลลัพธ์ไปตรงข้ามกัน ค่าจริงเป็นสิ่งที่ไม่สนใจ แต่แบบจำลองให้ผลลัพธ์เป็นสิ่งที่สนใจ False Negative (FN) ผลลัพธ์ตรงข้ามกัน ค่าจริงเป็นสิ่งที่สนใจ แต่แบบจำลองให้ผลลัพธ์เป็นสิ่งที่ไม่สนใจ

ความแม่นยำในการประเมินในสิ่งที่สนใจ (Precision) พิจารณาเฉพาะที่เป็น True Positives (TP) ความสามารถในการตรวจจับสิ่งที่สนใจ (Recall) ความถูกต้อง (Accuracy) ใช้บอกความถึงความสามารถของแบบจำลองโดยรวมให้ผลลัพธ์ไปในทางเดียวกัน ทั้งในกรณี Positive และ Negative ค่าเฉลี่ยแบบฮาร์โมนิกในการวัดประสิทธิภาพแบบจำลอง (F1-Score) เป็นค่าความสมดุลระหว่างความแม่นยำในการประเมินในสิ่งที่สนใจ และความแม่นยำในการประเมินในสิ่งที่สนใจ [22]

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Accuracy} = \frac{TP + TN}{(TP + FP + FN + TN)} \quad (3)$$

$$\text{F1 - Score} = 2 \times \frac{TP}{TP + FP} \quad (4)$$

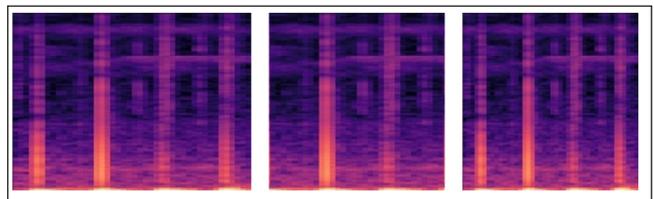
4. ผลการดำเนินงาน

การออกแบบและพัฒนาการจำแนกรั้วใหม่ดี และรั้วใหม่เสียด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชัน และการเรียนรู้แบบถ่ายโอน แบ่งการทดลองออกเป็น 3 ส่วน ได้แก่ การทดสอบภาพจากเสียงเพื่อใช้ในการวิจัย การทดสอบภาพกับแบบจำลองการเรียนรู้แบบถ่ายโอน และการทดสอบ

กับแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชัน

4.1 ผลการแปลงภาพจากข้อมูลเสียง

การวิจัยนี้ใช้คอมพิวเตอร์โน้ตบุ๊กที่มีหน่วยประมวลผลกลางรหัส Intel Core i7 Gen11 ความเร็ว 2.30 กิกะเฮิร์ต หน่วยความจำหลัก 48 กิกะไบต์ หน่วยความจำสำรอง 1 เทระไบต์ หน่วยประมวลผลกราฟิก Nvidia GeForce RTX 3070 ใช้ซอฟต์แวร์ภาษาไพธอน ไบรารีไลเซนส์ไฟล์ (Soundfile) ทำการตัดเสียงให้มีความยาวในมิติของเวลาเท่ากัน คือ 1 วินาที จากนั้นแปลงเสียงเป็นภาพด้วยไลบรารีลิบโรซ่า และไลบรารีโอเพนซีวี ภาพที่ได้มีขนาดความกว้าง 610 พิกเซล สูง 450 พิกเซล ความละเอียด 100 ดีพีไอ ระดับความลึก 24 บิต จากนั้นทำการปรับขนาดภาพให้ความกว้างและความสูงมีขนาดเท่ากัน โดยใช้ค่า 450 พิกเซล เพื่อการทำงานที่เหมาะสมกับเคเนล (Kernel) ของแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชัน และแบบจำลองการเรียนรู้การถ่ายโอน การลดความกว้างของภาพใช้ 2 วิธี ได้แก่ 1) วิธีกึ่งกลางภาพ (Center) เป็นหาที่กึ่งกลางของความกว้างแล้วขยายพื้นที่ทางซ้าย - ขวาออกจากกึ่งกลางเท่า ๆ 2) วิธีปรับขนาด (Resize) เป็นการใช้ฟังก์ชัน Resize ของไลบรารีโอเพนซีวี เพื่อลดพื้นที่ความกว้างลง ดังภาพที่ 11



(ก) (ข) (ค)

ภาพที่ 11 การปรับขนาดภาพ

จากภาพที่ 11 (ก) เป็นภาพที่ได้จากการแปลงข้อมูลเสียงเป็นภาพ 11 (ข) เป็นภาพที่ใช้วิธี Center และภาพ 11 (ค) เป็นภาพที่ใช้วิธี Resize ซึ่งภาพที่ได้จาก 2 วิธี จะถูกนำไปทดสอบกับแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชัน และแบบจำลองการเรียนรู้การถ่ายโอน

4.2 ผลการทดลองโดยใช้แบบจำลองการเรียนรู้การถ่ายโอน

การทดลองนี้ใช้ Google Computer Engine และไพธอน 3 โดยมีฮาร์ดแวร์ ดังนี้ Intel(R) Xeon(R) CPU @ 2.20 กิกะเฮิร์ต จำนวน 2 หน่วย หน่วยความจำหลัก 12.7 กิกะไบต์ หน่วยความจำสำรอง 225.8 กิกะไบต์ ซอฟต์แวร์ประกอบด้วย

ไลบรารีไซเคิลรีน ไลบรารีนิ้มพาย ไลบรารีแพนดาส และไลบรารีซีบอน กำหนดค่า Epoch เท่ากับ 20 ค่า Learning เท่ากับ 0.01 ค่า Batch Size เท่ากับ 32 แสดงผลดังตารางที่ 2

ตารางที่ 2 แสดงประสิทธิภาพของแบบจำลองการเรียนรู้แบบถ่ายโอน

Classifier Evaluation	MobileNetV2		NASNetMobile	
	Center	Resize	Center	Resize
Precision	95.00	94.00	95.00	91.00
Recall	95.00	94.50	94.50	91.00
Accuracy	95.00	94.00	94.50	91.00
F1-Score	95.00	94.50	94.50	91.00

จากตารางที่ 2 พบว่า แบบจำลองการเรียนรู้แบบถ่ายโอนวิธี MobileNet ให้ค่าให้ค่าความแม่นยำ ค่าความสามารถในการตรวจจับสิ่งที่สนใจ ค่าความถูกต้อง และค่าเฉลี่ยแบบฮาร์โมนิกที่สูงและใกล้เคียงกันกว่าแบบจำลองการเรียนรู้แบบถ่ายโอนวิธี NASNetMobile และภาพข้อมูลรังไหมที่ใช้การลดความกว้างของภาพด้วยวิธี Center นำไปใช้ในการฝึกสอนโมเดลช่วยทำให้ประสิทธิภาพในการจำแนกของโมเดลสูงกว่าการลดความกว้างของภาพด้วยวิธี Resize

4.3 ผลการทดลองโดยใช้แบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชัน

การทดลองนี้ใช้ฮาร์ดแวร์ และซอฟต์แวร์ชุดเดียวกับการทดลองโดยการเรียนรู้แบบถ่ายโอน โดยให้ผลการทดลองดังตารางที่ 3

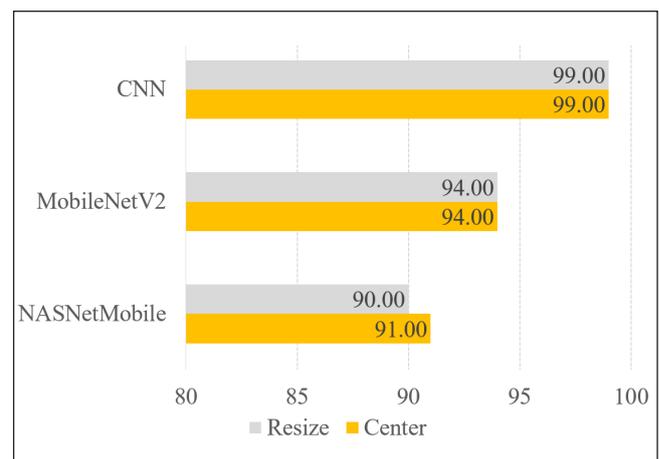
ตารางที่ 3 แสดงประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชัน

Classifier Evaluation	CNN	
	Center	Resize
Precision	100.00	100.00
Recall	100.00	100.00
Accuracy	100.00	100.00
F1-Score	100.00	100.00

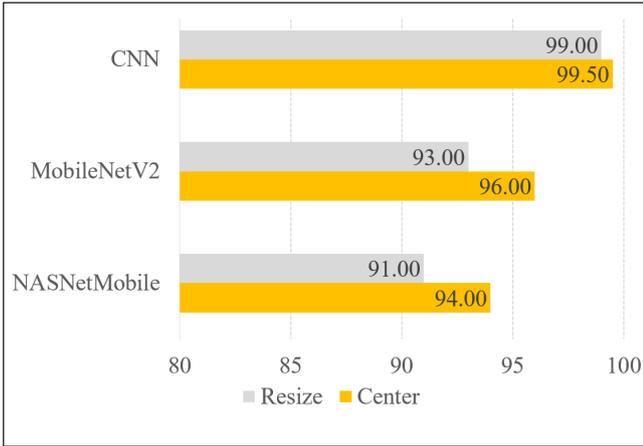
จากตารางที่ 3 ค่าประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชันทั้งค่าความแม่นยำ ค่าความสามารถในการตรวจจับสิ่งที่สนใจ ค่าความถูกต้อง และค่าเฉลี่ยแบบฮาร์โมนิก ให้ค่าสูงสุดร้อยละ 100 และใช้ได้ดีกับภาพที่ได้จากการลดความกว้างของภาพ ทั้งวิธี Center และวิธี Resize เนื่องจากลักษณะเด่นของภาพอยู่บริเวณกลางภาพ เมื่อใช้การลดความกว้างของภาพด้วยวิธี Center ทำให้คุณลักษณะเด่นยังคงอยู่ และการลดความกว้างของภาพด้วยวิธี Resize เป็นการลดพื้นที่ความกว้างลงแต่ยังคงคุณลักษณะเด่นไว้ได้ ประกอบกับค่าพารามิเตอร์และจำนวนชั้นเลเยอร์ของวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชันถูกกำหนดเพื่อการจำแนกภาพสเปกโตแกรมข้อมูลเสียงรังไหมดีและรังไหมเสียโดยปราศจากรูปอื่นปะปนส่งผลให้ประสิทธิภาพของแบบจำลองทั้ง 4 ค่าจึงมีค่าสูงสุด

4.4 ผลประสิทธิภาพของการลดความกว้างของภาพที่แตกต่างตามจำนวนครั้งที่ใช้ข้อมูล

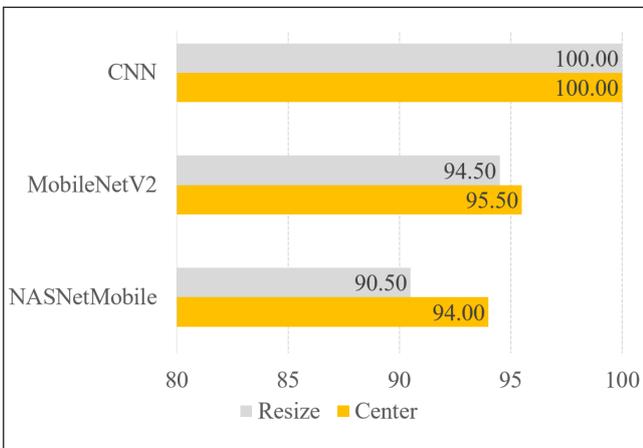
ประสิทธิภาพของแบบจำลองแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชัน แบบจำลองการเรียนรู้แบบถ่ายโอนวิธี MobileNet และวิธี NASNetMobile โดยใช้การลดความกว้างของภาพ ด้วยวิธี Center และวิธี Resize จะพิจารณาค่าความถูกต้องในจำนวนครั้งที่ใช้ข้อมูลแสดงดังภาพที่ 12 - 15



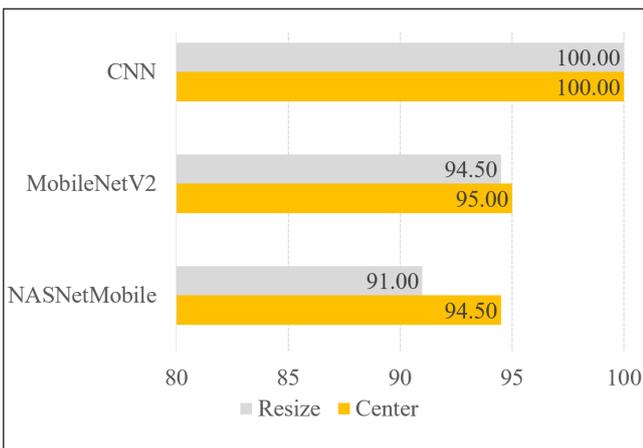
ภาพที่ 12 จำนวนครั้งที่ใช้ข้อมูล 5 Epochs



ภาพที่ 13 จำนวนครั้งที่ใช้ข้อมูล 10 Epochs



ภาพที่ 14 จำนวนครั้งที่ใช้ข้อมูล 15 Epochs



ภาพที่ 15 จำนวนครั้งที่ใช้ข้อมูล 20 Epochs

จากภาพที่ 12 - 15 จำนวนครั้งที่ใช้ข้อมูล 5, 10, 15, และ 20 Epochs พบว่าการปรับขนาดภาพด้วยวิธี Center ส่งผลให้ค่าความถูกต้องในแบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชัน แบบจำลองการเรียนรู้แบบถ่ายโอนวิธี

MobileNet และวิธี NASNetMobile สูงกว่า หรือเท่ากับการลดความกว้างของภาพด้วยวิธี Resize แบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชันที่ใช้การลดความกว้างของภาพด้วย Center ให้ค่าความถูกต้องสูงกว่าแบบจำลองการเรียนรู้แบบถ่ายโอนวิธี MobileNet และวิธี NASNetMobile ในทุก Epochs ซึ่งใน Epochs 15 แบบจำลองโครงข่ายประสาทเทียมแบบคอนโวลูชันที่ใช้การลดความกว้างของภาพด้วย Center ให้ค่าความถูกต้องร้อยละ 100 ในขณะที่แบบจำลองการเรียนรู้แบบถ่ายโอนวิธี MobileNet ให้ค่าความถูกต้อง ร้อยละ 95.00 และวิธี NASNetMobile ให้ค่าความถูกต้องร้อยละ 94.50

5. สรุป

งานวิจัยนี้นำเสนอการออกแบบและพัฒนาการจำแนกรูปภาพใหม่ดี และรังไหมเสียด้วยวิธีโครงข่ายประสาทเทียมแบบคอนโวลูชันและการเรียนรู้แบบถ่ายโอน โดยมีข้อมูลภาพจำนวน 1,000 ภาพ จาก 2 คลาส ๆ ละ 500 ภาพ แบ่งภาพเป็น 2 ชุด เพื่อฝึกสอนและทดสอบแบบจำลอง โดยใช้อัตราส่วน 80 ต่อ 20 การเตรียมข้อมูลจากเสียงเป็นภาพใช้ 2 วิธี ได้แก่ วิธีกึ่งกลางภาพ (Center) และวิธีปรับขนาด (Resize) ซึ่งใช้ภาษาไพธอนไลบรารีลิลิปโรซ่า ไลบรารีโอเพนซีวี ไลบรารีเคอร์ส และไลบรารีไซคิเคิลรินในการพัฒนางานวิจัย

งานวิจัยใช้การวัดประสิทธิภาพด้วยตารางคอนฟิวชันแมทริกซ์ ค่าความแม่นยำ ค่าความสามารถในการตรวจจับสิ่งที่สนใจ ค่าความถูกต้อง และค่าเฉลี่ยแบบฮาร์โมนิก ซึ่งพบว่าแบบจำลองที่ออกแบบและพัฒนาสามารถใช้จำแนกเสียงได้ โดยแบบจำลองที่ใช้วิธีโครงข่ายประสาทเทียมแบบคอนโวลูชันให้ค่าความแม่นยำ ค่าความสามารถในการตรวจจับสิ่งที่สนใจ ค่าความถูกต้อง และค่าเฉลี่ยแบบฮาร์โมนิก ร้อยละ 100 ซึ่งสามารถใช้ภาพนำเข้าโดยการแปลงเสียงเป็นภาพ และการลดความกว้างของภาพได้ทั้งวิธี Center หรือวิธี Resize เพราะคุณลักษณะเด่นของข้อมูลอยู่กลางไฟล์ ส่งผลให้ค่าความถูกต้องในการจำแนกสูง และสามารถใช้ประสิทธิภาพของแบบจำลองได้ตั้งแต่ Epochs ที่ 15 ส่วนการใช้การถ่ายโอนการเรียนรู้แบบจำลอง MobileNetV2 ให้ค่าประสิทธิภาพสูงกว่า NASNetMobile และควรใช้การลดความกว้างของภาพด้วยวิธี Center ในการเตรียมภาพนำเข้าแบบจำลอง

แนวทางวิจัยในอนาคตสามารถทำได้โดยการใช้อุปกรณ์บันทึกเสียงที่หลากหลาย เพิ่มประเภทไฟล์เสียง เพิ่มสัญญาณรบกวน

เพื่อให้ได้ข้อมูลที่ตรงกับสภาพความเป็นจริง (เสียงที่เกิดขึ้นตามธรรมชาติ) ใช้วิธีการแปลงภาพแบบอื่น เช่น สัมประสิทธิ์เซปสตรัมบนสเกลเมล (Mel frequency cepstrum coefficient, MFCC) หรือการสกัดคุณลักษณะเด่นอื่น หรือใช้แบบจำลองอื่น ๆ

6. กิตติกรรมประกาศ

ขอขอบคุณหลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏพิบูลสงคราม ที่ให้ความอนุเคราะห์สถานที่ และอุปกรณ์ในการทำงานวิจัย และคุณสมบูรณ์ สิมสวัสดิ์ เกษตรกรผู้เลี้ยงไหม ตำบลเขาค้อ อำเภอเขาค้อ จังหวัดเพชรบูรณ์

7. เอกสารอ้างอิง

- [1] Kasertsart University, *Silk production technology and value added of silk products*. Available Online at https://www3.rdi.ku.ac.th/exhibition/53/group06/silk_center/silk.html, accessed on 9 April 2024.
- [2] Chul Thai Silk, *Until it becomes Thai silk*. Available Online at <https://www.chulthaisilk.com/aboutus>, accessed on 9 April 2024.
- [3] Thairut Online, *Update on agricultural innovation 2024 survival and opportunities with AI*. Available Online at <https://www.thairath.co.th/scoop/interview/2754702>, accessed on 9 April 2024.
- [4] T. Worasit, et al. "Rice bacterial blight and blast diseases recognition using deep learning techniques." *Khon Kaen Agriculture Journal*, Vol. 50, No. 1, pp. 216 - 228, January - February, 2021.
- [5] C. Suppatoomsin and A. Srikaew. "Imagery Grape Leaf Disease Diagnosis Based on a GA-SASOM Algorithm." *The Journal of Industrial Technology*, Vol. 14, No. 3, pp. 44 - 61, September - December, 2018.
- [6] N. Chinpanthana. "Image Classification of Tomato Leaf Diseases using Convolutional Neural Network." *Journal of Information Science and Technology*, Vol. 13, No. 2, pp. 40 - 49, July - December, 2023.
- [7] S. Thipsrirach, S. Kunhareang, and S. Suwanlee. "The Use of Deep Learning Technique in the Classification of Pradu Hang Dam Thai Native Chicken Images." *Rajabhat Agriculture Journal*, Vol. 22, No. 1, pp. 50 - 59, January - June, 2023.
- [8] O.-U. Pramote and J. Undub. "An Analysis of Banana Ripeness Level by Image using RGB Color Model and Supervised Learning." *Journal of science and technology RMUTSB*, Vol. 5, pp. 34 - 43, August, 2021.
- [9] P. Saechen and T. Siriborvornratanakul. "Oil Palm Fresh Fruit Bunch Ripeness Classification by Deep Learning." *KMUTT Research & Development Journal*, Vol. 46, No. 1, pp. 81 - 104, January - March, 2023.
- [10] Wikipedia, *Sound*. Available Online at <https://en.wikipedia.org/wiki/Sound>, accessed on 9 April 2024.
- [11] Kaggle, *Audio Processing using Librosa for Beginners*. Available Online at <https://www.kaggle.com/code/jaseemck/audio-processing-using-librosa-for-beginners>, accessed on 9 April 2024.
- [12] N. Sritasuwan, *Development of Computer Program for Study Bird Vocal Communication*. Available Online at https://archive.lib.cmu.ac.th/full/res/2538/tressct490508_38_full.pdf, accessed on 9 April 2024.
- [13] P. Doungmala, *Mel frequency cepstrum coefficient: MFCC*. Available Online at <https://dr-pathasu-doung.medium.com/mel-frequency-cepstrum-coefficient-mfcc-b735a349c312>, accessed on 9 April 2024.
- [14] M. Anthony, *Audio analysis with librosa*. Available Online at <https://blog.neurotech.africa/audio-analysis-with-librosa/>, accessed on 9 April 2024.
- [15] K. Zoumana, *An Introduction to Convolutional Neural Networks (CNNs)*. Available Online at <https://www.datacamp.com/tutorial/introduction-to-convolutional-neural-networks-cnns>, accessed on 9 April 2024.
- [16] Datascientest, *Transfer Learning: What is it*. Available Online at <https://datascientest.com/en/transfer-learning-what-is-it>, accessed on 9 April 2024.
- [17] N. Khobjai, et al. "Sound Classification of five insect species." *The 10th Asia Undergraduate Conference on Computing (AUCC2022)*, Thailand, pp. 512 - 520, 2022.



- [18] A. Piromto, et al. "Developing a Model for Classifying Birds' Voices by Deep Learning Technique." *The 10th Asia Undergraduate Conference on Computing (AUCC2022)*, Thailand, pp. 498 - 505, 2022.
- [19] T. Kanchanakom, et al. "Classification of Dog Vocalization on an Edge-AI Device." *The 9th Asia Undergraduate Conference on Computing (AUCC2021)*, Thailand, pp. 464 - 472, 2021.
- [20] P. Chiangchan and J. Rungrattanaubol. "Animal voice recognition application." *The 10th Asia Undergraduate Conference on Computing (AUCC2022)*, Thailand, pp. 902 - 909, 2022.
- [21] S. Kawinseksan, et al. "Thai Poetry Reading Sound Classification using Neural Network." *The 9th Asia Undergraduate Conference on Computing (AUCC2021)*, Thailand, pp. 799 - 804, 2021.
- [22] Wikipedia, *Confusion matrix*. Available Online at https://en.wikipedia.org/wiki/Confusion_matrix, accessed on 9 April 2024.
-