บทความวิจัย

# การศึกษาเปรียบเทียบระหว่างโครงข่ายประสาทเทียมแบบคอนโวลูชันและวิชันทรานส์ฟอร์เมอร์สำหรับการจำแนกโรคใบอ้อยในสภาพแวดล้อมจริง

จักรพันธ์ ศรีวงษา* ทินกร คุณาสิทธิ์ และ ศุภชัย ใบปอด
สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏมหาสารคาม

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 08 1670 7863 อีเมล: Jackaphan@rmu.ac.th
DOI: 10.14416/j.kmutnb.2026.11.004
รับเมื่อ 24 กุมภาพันธ์ 2569 แก้ไขเมื่อ 7 พฤษภาคม 2569 ตอบรับเมื่อ 6 กรกฎาคม 2569 เผยแพร่ออนไลน์ 20 พฤศจิกายน 2569
© 2026 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

**บทคัดย่อ**

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบสมรรถนะของโมเดลการเรียนรู้เชิงลึก 2 กลุ่มสถาปัตยกรรม ได้แก่ โครงข่ายประสาทเทียมแบบคอนโวลูชัน (ResNet-50 และ DenseNet-121) และวิชันทรานส์ฟอร์เมอร์ (ViT-Base และ Swin-Tiny) สำหรับการจำแนกโรคใบอ้อย 4 ประเภท ได้แก่ ใบปกติ โรคเหี่ยวแดง โรคใบจุดวงแหวน และโรคราสนิม ในสภาพแวดล้อมจริง ชุดข้อมูลที่ใช้ประกอบด้วยภาพถ่ายใบอ้อยความละเอียดสูงจำนวน 4,000 ภาพ แบ่งเป็น 4 คลาส ๆ ละ 1,000 ภาพ โดยแบ่งสัดส่วนสำหรับฝึกสอน ตรวจสอบ และทดสอบในอัตรา 80:10:10 โมเดลทั้งหมดถูกฝึกสอนจากค่าเริ่มต้นบน หน่วยประมวลผลกราฟิก (GPU) NVIDIA Tesla T4 โดยใช้ PyTorch 1.12 และกำหนดพารามิเตอร์การฝึกสอน (Hyperparameters) ร่วมกัน ได้แก่ ขนาดกลุ่มข้อมูล (Batch Size) เท่ากับ 16 จำนวนรอบการฝึกสอน (Epochs) เท่ากับ 15 และตัวปรับแต่งน้ำหนักแบบอดัม (Adam Optimizer) ที่อัตราการเรียนรู้ (Learning Rate) 0.001 ผลการทดลองพบว่า DenseNet-121 มีสมรรถนะสูงสุดในทุกตัวชี้วัด ด้วยความแม่นยำรวม (Accuracy) ร้อยละ 99.25 ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าเอฟวัน (F1-Score) เท่ากับ 0.9926 0.9925 และ 0.9925 ตามลำดับ และใช้เวลาฝึกสอนน้อยที่สุดเพียง 10 นาที 56 วินาที รองลงมาคือ ResNet-50 (ร้อยละ 99.00) และ Swin-Tiny (ร้อยละ 98.50) ในขณะที่ ViT-Base ให้ความแม่นยำต่ำสุดที่ร้อยละ 97.00 ผลการวิจัยชี้ให้เห็นว่าสถาปัตยกรรมโครงข่ายประสาทเทียมแบบคอนโวลูชัน (CNN) มีความเหมาะสมกับชุดข้อมูลขนาดจำกัดมากกว่าสถาปัตยกรรมวิชันทรานส์ฟอร์เมอร์ (Transformer) เนื่องจากมีความเอนเอียงเชิงอุปนัย (Inductive Bias) ที่เหมาะสมกับการสกัดคุณลักษณะเฉพาะของรอยโรคขนาดเล็ก องค์ความรู้ที่ได้สามารถนำไปเป็นแนวทางในการพัฒนาระบบวินิจฉัยโรคพืชอัตโนมัติบนอุปกรณ์พกพาเพื่อส่งเสริมเกษตรอัจฉริยะต่อไป

**คำสำคัญ:** การจำแนกโรคใบอ้อย การเรียนรู้เชิงลึก เกษตรอัจฉริยะ โครงข่ายประสาทเทียมแบบคอนโวลูชัน วิชันทรานส์ฟอร์เมอร์

การอ้างอิงบทความ: จักรพันธ์ ศรีวงษา, ทินกร คุณาสิทธิ์ และ ศุภชัย ใบปอด, "การศึกษาเปรียบเทียบระหว่างโครงข่ายประสาทเทียมแบบคอนโวลูชันและวิชันทรานส์ฟอร์เมอร์สำหรับการจำแนกโรคใบอ้อยในสภาพแวดล้อมจริง," *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 37, ฉบับที่ 1, หน้า 1–12, ม.ค.–มี.ค. 2570. เลขที่บทความ 271-088385, doi: 10.14416/j.kmutnb.2026.11.004.

Research Article

# A Comparative Study of Convolutional Neural Networks and Vision Transformers for Sugarcane Leaf Disease Classification in Real-World Environments

Jackaphan Sriwongsa*, Tinnakorn Kunasit and Supachai Baipod

Department of Computer Science, Faculty of Science and Technology, Rajabhat Mahasarakham University, Maha Sarakham, Thailand

* Corresponding Author, Tel. 08 1670 7863, E-mail: Jackaphan@rmu.ac.th
DOI: 10.14416/j.kmutnb.2026.11.004
Received 24 February 2026; Revised 7 May 2026; Accepted 6 July 2026; Published online: 20 November 2026
© 2026 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

**Abstract**

This study aimed to investigate and compare the performance of two deep learning architecture groups, namely Convolutional Neural Networks (ResNet-50 and DenseNet-121) and Vision Transformers (ViT-Base and Swin-Tiny), for classifying four types of sugarcane leaf conditions, namely healthy leaves, red wilt, ring spot, and rust, under real-world field conditions. The dataset comprised 4,000 high-resolution sugarcane leaf images, balanced across four classes with 1,000 images per class, and was partitioned into training, validation, and testing sets at a ratio of 80:10:10. All models were trained from scratch on an NVIDIA Tesla T4 GPU using PyTorch 1.12 under identical hyperparameter settings, including a batch size of 16, 15 epochs, and the Adam optimizer with a learning rate of 0.001. Experimental results demonstrated that DenseNet-121 achieved the highest performance across all evaluation metrics, with an overall accuracy of 99.25%, and Precision, Recall, and F1-Score values of 0.9926, 0.9925, and 0.9925, respectively, while requiring the shortest training time of only 10 minutes and 56 seconds. ResNet-50 and Swin-Tiny followed with accuracies of 99.00% and 98.50%, respectively, whereas ViT-Base achieved the lowest accuracy of 97.00%. These findings indicate that CNN architectures are more suitable than Vision Transformer architectures for limited datasets because their inherent inductive bias is better suited for extracting discriminative features from small disease lesions. The findings of this study provide a practical foundation for developing automated plant disease diagnosis systems on mobile devices to support smart agriculture.

**Keywords**: Sugarcane Leaf Disease Classification, Deep Learning, Smart Agriculture, Convolutional Neural Networks, Vision Transformer

Please cite this article as: J. Sriwongsa, T. Kunasit and S. Baipod, "Prediction of feature importance for youtube premium cancellation using artificial neural network and ensemble learning algorithms," *The Journal of KMUTNB*, vol. 37, no. 1, pp. 1–12, Jan.–Mar. 2027 (in Thai), Art. no. 271-088385, doi: 10.14416/j.kmutnb.2026.11.004.

## 1. บทนำ

อ้อย (Sugarcane) เป็นพืชเศรษฐกิจที่มีความสำคัญอย่างยิ่งต่ออุตสาหกรรมการผลิตน้ำตาล และพลังงานทดแทน อย่างไรก็ตามปัญหาหลักที่ส่งผลกระทบโดยตรงต่อปริมาณและคุณภาพของผลผลิตคือ การระบาดของโรคพืชทางใบ ได้แก่ โรคเหี่ยวแดง (Red Rot) โรคใบจุดวงแหวน (Ring Spot) และโรคราสนิม (Rust) [1] ในอดีตการประเมินและวินิจฉัยโรคเหล่านี้ต้องอาศัยนักวิชาการเกษตรหรือผู้เชี่ยวชาญในการลงพื้นที่สำรวจ ซึ่งเป็นกระบวนการที่ใช้เวลานาน สิ้นเปลืองแรงงาน และมีโอกาสเกิดความผิดพลาดจากดุลยพินิจของมนุษย์ (Human Error) โดยเฉพาะอย่างยิ่งการสำรวจในสภาพแวดล้อมจริง (Real-World Environments) ที่มีความผันแปรของแสง เงา และความซับซ้อนของพื้นหลังแปลงปลูก ซึ่งเป็นอุปสรรคสำคัญต่อการวินิจฉัยโรคอย่างแม่นยำ

ในปัจจุบันเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence; AI) และในช่วงไม่กี่ปีที่ผ่านมาเทคโนโลยีการเรียนรู้เชิงลึก (Deep Learning) ได้รับความนิยมและประสบความสำเร็จอย่างสูงในการนำมาประยุกต์ใช้ เพื่อตรวจจับและจำแนกโรคพืชหลากหลายชนิด เช่น ข้าวโพด ข้าว และมะเขือเทศ [2]–[4] อย่างไรก็ตามโครงข่ายประสาทเทียมแบบคอนโวลูชัน (Convolutional Neural Networks; CNNs) ถูกนำมาใช้อย่างแพร่หลาย และประสบความสำเร็จในการสกัดจุดเด่น (Feature Extraction) ของรอยโรคพืช ดังที่ Upadhye และคณะ [5] ได้นำเสนอการใช้เทคนิค CNN ในการตรวจจับโรคอ้อย ซึ่งสามารถเพิ่มความรวดเร็วและลดภาระงานได้อย่างเป็นรูปธรรม อย่างไรก็ตามนอกเหนือจากกลุ่ม CNN แล้ว สถาปัตยกรรมวิชันทรานส์ฟอร์เมอร์ (Vision Transformers; ViTs) ได้ถูกพัฒนาขึ้นและสร้างความก้าวหน้าอย่างมากในงานด้านคอมพิวเตอร์วิทัศน์ เนื่องจากมีกลไกความสนใจในตนเอง (Self-Attention Mechanism) ที่ช่วยให้โมเดลสามารถวิเคราะห์และเชื่อมโยงบริบทภาพรวม (Global Context) ได้ดีกว่า CNN ซึ่งเน้นเฉพาะจุดเด่นบริเวณย่อย (Local Features) การศึกษาของ Özden [6] ได้เปรียบเทียบ Vision Transformer และ CNN ในการวิเคราะห์ภาพโรคพืชหลายชนิดพบว่า Vision Transformer มีจุดเด่นในการวิเคราะห์บริบทภาพรวม แต่มีข้อจำกัดด้านความต้องการข้อมูลจำนวนมากและทรัพยากรการประมวลผลสูง ในทำนองเดียวกัน Demir และคณะ [7] ได้ศึกษาการประยุกต์ใช้ปัญญาประดิษฐ์ในการวินิจฉัยโรคพืชโดยเปรียบเทียบระหว่าง Vision Transformer และ CNN พบว่า CNN มีความได้เปรียบในด้านความเสถียรบนชุดข้อมูลขนาดจำกัด ในขณะที่ Vision Transformer มีศักยภาพสูงกว่าเมื่อมีข้อมูลเพียงพอ ทั้งสองงานวิจัยชี้ให้เห็นถึงความจำเป็นในการศึกษาเปรียบเทียบเชิงระบบบนชุดข้อมูลโรคพืชเฉพาะชนิดในสภาพแวดล้อมจริง ซึ่งเป็นสิ่งที่งานวิจัยนี้มุ่งดำเนินการ

จากการทบทวนงานวิจัยที่ผ่านมาพบว่ามีช่องว่างสำคัญ 3 ประการ ประการแรกคือ งานวิจัยด้านการจำแนกโรคอ้อยส่วนใหญ่มุ่งเน้นการใช้สถาปัตยกรรม CNN กลุ่มใดกลุ่มหนึ่งเพียงกลุ่มเดียว [5], [8] ยังขาดการเปรียบเทียบเชิงระบบระหว่าง CNN และ Vision Transformer บนชุดข้อมูลโรคใบอ้อยโดยตรง ประการที่สองคือ งานที่ผ่านมาหลายชิ้นทดสอบโมเดลในสภาพแวดล้อมที่ถูกควบคุม กล่าวคือ การรวบรวมภาพในพื้นที่ปิดที่มีแสงสม่ำเสมอ พื้นหลังสีเดียวและระยะถ่ายภาพคงที่ ซึ่งไม่สะท้อนความซับซ้อนของแปลงปลูกจริง [4] ประการที่สามคือ ยังขาดการศึกษาที่ประเมินสถาปัตยกรรมทั้งสอง

กลุ่มภายใต้เงื่อนไขที่เท่าเทียมกันบนชุดข้อมูลภาพถ่ายจากสภาพแวดล้อมจริงในบริบทเกษตรกรรมไทย [6], [7] ดังนั้นความใหม่ (Novelty) ของงานวิจัยนี้จึงอยู่ที่การเป็นงานแรกที่เปรียบเทียบ CNN และ Vision Transformer 4 สถาปัตยกรรมภายใต้เงื่อนไขที่เท่าเทียมกันอย่างเป็นระบบบนชุดข้อมูลใบอ้อยที่รวบรวมจากแปลงปลูกจริงในประเทศไทย เพื่อให้ได้คำตอบที่ชัดเจนสำหรับการนำไปใช้จริงในบริบทเกษตรอัจฉริยะ ทั้งนี้ เพื่อเป็นแนวทางสำคัญในการนำไปประยุกต์และพัฒนาต่อยอดในการพัฒนาแอปพลิเคชันบนสมาร์ทโฟนสำหรับเกษตรกร การบูรณาการกับระบบโดรนเพื่อสำรวจแปลงปลูกในวงกว้าง และการเชื่อมโยงกับแพลตฟอร์มเกษตรอัจฉริยะ (Smart Farming) ที่ช่วยให้เกษตรกรได้รับการแจ้งเตือนและคำแนะนำการจัดการโรคได้อย่างทันท่วงที ซึ่งจะช่วยลดการสูญเสียผลผลิต และลดต้นทุนการใช้สารเคมีได้อย่างมีประสิทธิภาพ

## 2. วิธีการทดลอง

การศึกษานี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการเรียนรู้เชิงลึกต่างสถาปัตยกรรมในการจำแนกประเภทโรคใบอ้อย โดยมีรายละเอียดของขั้นตอนการดำเนินงานวิจัยดังต่อไปนี้

### 2.1 การเตรียมและจัดการชุดข้อมูล

ชุดข้อมูลภาพถ่ายใบอ้อยที่ใช้ในการศึกษานี้นำมาจากฐานข้อมูลที่รวบรวมจากสภาพแวดล้อมการปลูกจริงตามที่ได้นำเสนอไว้ในการศึกษาของ [8] ซึ่งรวบรวมจาก 4 จังหวัด ได้แก่ จังหวัดมหาสารคาม ขอนแก่น กาฬสินธุ์ และร้อยเอ็ด ชุดข้อมูลจัดเก็บในช่วงเดือนกุมภาพันธ์ พ.ศ. 2567 ถึง เมษายน พ.ศ. 2567 ถ่ายภาพด้วยกล้องโทรศัพท์มือถือชนิดเดียวกันทุกภาพ ความละเอียด 12MP (Mega Pixels) ประกอบด้วยภาพถ่ายใบอ้อยจำนวนทั้งสิ้น 4,000 ภาพ แบ่งออกเป็น 4 กลุ่ม (Classes) กลุ่มละ 1,000 ภาพ ได้แก่ 1) ใบอ้อยปกติ (Healthy) 2) โรคเหี่ยวแดง (Red Rot/Red Stripe) 3) โรคใบจุดวงแหวน และ 4) โรคราสนิม

ก่อนนำข้อมูลเข้าสู่กระบวนการฝึกสอนภาพทั้งหมดถูกนำมาปรับขนาดมาตรฐานให้มีความละเอียด 224 × 224 พิกเซล เนื่องจากสถาปัตยกรรมโมเดลทั้ง 4 ที่ทดสอบในงานวิจัยนี้ล้วนออกแบบมาให้รับอินพุต (Input) ขนาดดังกล่าวนี้ ทำให้การเปรียบเทียบระหว่างสถาปัตยกรรมมีความเป็นธรรมแม้การลดขนาดภาพอาจทำให้รายละเอียดบางส่วนสูญเสียไปแต่คุณลักษณะสำคัญของรอยโรค เช่น สี พื้นผิว และรูปร่าง ยังคงปรากฏอยู่ในระดับที่เพียงพอสำหรับการจำแนกประเภท จากนั้นชุดข้อมูลถูกแบ่งออก เป็น 3 ส่วนหลักอย่างเป็นระบบ (Systematic Data Split) ในอัตราส่วน 80:10:10 ได้แก่ ชุดข้อมูลสำหรับฝึกสอน (Training Set) จำนวน 3,200 ภาพ ชุดข้อมูลสำหรับตรวจสอบความถูกต้องระหว่างฝึก (Validation Set) จำนวน 400 ภาพ และชุดข้อมูลสำหรับทดสอบประสิทธิภาพ (Testing Set) จำนวน 400 ภาพ นอกจากนี้งานวิจัยนี้ได้ประยุกต์ใช้เทคนิคการเพิ่มปริมาณข้อมูล (Data Augmentation) [9] เฉพาะในชุดข้อมูลฝึกสอนเพื่อเพิ่มความหลากหลายของลักษณะภาพ และลดปัญหาการเรียนรู้แบบท่องจำ (Overfitting) โดยกำหนดพารามิเตอร์ดังนี้ การหมุนภาพแบบสุ่ม (Random Rotation) การพลิกภาพตามแนวนอน (Horizontal Flip) และการปรับความสว่าง (Brightness Jitter)

### 2.2 สถาปัตยกรรมโมเดลการเรียนรู้เชิงลึก

งานวิจัยนี้มุ่งเน้นการศึกษาเปรียบเทียบระหว่างโมเดล

2 กลุ่มสถาปัตยกรรมหลัก ที่ได้รับความนิยมในงานคอมพิวเตอร์วิทัศน์ ได้แก่ 1) กลุ่มโครงข่ายประสาทเทียมแบบคอนโวลูชัน ประกอบด้วยโมเดล ResNet-50 [10] ซึ่งมีโครงสร้างการเชื่อมต่อข้ามชั้น (Skip Connection) ช่วยแก้ปัญหาความชันสูญหายในโมเดลที่มีความลึก และโมเดล DenseNet-121 [11] ซึ่งมีจุดเด่นในการเชื่อมต่อข้อมูลแบบหนาแน่น (Dense Block) ทำให้ทุกชั้นย่อยสามารถนำคุณลักษณะ (Feature Maps) จากชั้นก่อนหน้ามาใช้ซ้ำได้อย่างมีประสิทธิภาพ 2) กลุ่มวิชันทรานส์ฟอร์เมอร์ ประกอบด้วยโมเดล ViT-Base [12] ซึ่งประยุกต์ใช้กลไกความสนใจในตนเอง (Self-Attention) โดยการแบ่งภาพออกเป็นแพตช์ (Patches) เพื่อเรียนรู้ความสัมพันธ์ของบริบทภาพในมุมกว้าง และโมเดล Swin-Tiny [13] ซึ่งเป็นสถาปัตยกรรมทรานส์ฟอร์เมอร์แบบลำดับชั้น (Hierarchical Transformer) ที่คำนวณความสนใจผ่านหน้าต่างแบบเลื่อน (Shifted Windows) ช่วยลดภาระการคำนวณและเพิ่มความแม่นยำในการวิเคราะห์ภาพที่มีความละเอียดหลากหลาย

### 2.3 การตั้งค่าสภาพแวดล้อมการทดลองและการประเมินผล

การพัฒนาโมเดลและการทดลองทั้งหมดดำเนินการบนเครื่องที่มีหน่วยประมวลผลกราฟิก NVIDIA Tesla T4 (หน่วยความจำ 16 GB) หน่วยความจำหลัก 32 GB ระบบปฏิบัติการ Ubuntu 20.04 LTS โดยใช้ Python 3.8 และเฟรมเวิร์ก PyTorch 1.12 ร่วมกับไลบรารี Timm (PyTorch Image Models) สำหรับโหลดสถาปัตยกรรมโมเดล ResNet-50, DenseNet-121, ViT-Base และ Swin-Tiny ที่นำมาเปรียบเทียบในการศึกษานี้ สำหรับพารามิเตอร์การฝึกสอนงานวิจัยนี้กำหนดค่าขนาดกลุ่มข้อมูลเท่ากับ 16 และอัตราการเรียนรู้ เท่ากับ 0.001 โดยกำหนดจากการทดลองเบื้องต้น (Preliminary Experiment) และจำนวนรอบการฝึกสอนที่ 15 รอบ ซึ่งเป็นค่าที่กำหนดขึ้นจากการสังเกตว่าโมเดลทุกสถาปัตยกรรมเข้าสู่ภาวะลู่เข้า (Convergence) โดยค่าความสูญเสียบนชุดตรวจสอบ (Validation Loss) ไม่ลดลงอย่างมีนัยสำคัญหลังรอบการฝึกสอน ที่ 12 โดยพิจารณาจากการที่ค่า Validation Loss ลดลงน้อยกว่า 0.001 ต่อรอบการฝึกสอน ในช่วงรอบการฝึกสอนที่ 12–15 สำหรับโมเดลทุกสถาปัตยกรรมบ่งชี้ว่าโมเดลเข้าสู่ภาวะลู่เข้าแล้ว และโมเดลทั้งหมดถูกตั้งค่าโดยไม่ใช้น้ำหนักที่ฝึกสอนล่วงหน้า (Train from Scratch) [14] เพื่อให้โครงข่ายเรียนรู้คุณลักษณะเฉพาะของโรคใบอ้อยจากชุดข้อมูลนี้โดยตรง โดยไม่มีอคติจากคุณลักษณะของ ImageNet ซึ่งแตกต่างอย่างมีนัยสำคัญจากภาพใบอ้อยในสภาพแวดล้อมแปลงปลูก แนวคิดนี้สอดคล้องกับที่ Dosovitskiy และคณะ [12] ได้อธิบายถึงข้อจำกัดของ ViT เมื่อฝึกสอนจากค่าเริ่มต้นบนชุดข้อมูลขนาดจำกัด เพื่อให้โครงข่ายเรียนรู้คุณลักษณะเฉพาะของโรคใบอ้อยอย่างเต็มที่ งานวิจัยนี้เลือกใช้ฟังก์ชันความสูญเสียแบบ Cross-Entropy Loss ทำงานร่วมกับตัวปรับแต่งน้ำหนัก (Optimizer) ชนิด Adam [15] โดยมีอัตราการเรียนรู้ (Learning Rate) เริ่มต้นที่ 0.001

## 3. ผลการทดลองและอภิปรายผลการทดลอง

การศึกษานี้มุ่งเน้นการประเมินและเปรียบเทียบประสิทธิภาพของโมเดลการเรียนรู้เชิงลึก 2 กลุ่มสถาปัตยกรรม ได้แก่ โครงข่ายประสาทเทียมแบบคอนโวลูชัน (ResNet-50 และ DenseNet-121) และวิชันทรานส์ฟอร์เมอร์ (ViT-Base และ Swin-Tiny) ในการจำแนกโรคใบอ้อยภายใต้การแบ่งชุดข้อมูลจำนวน 4,000 ภาพ ในสัดส่วน 80:10:10 การประเมินผลแบ่งออกเป็น

การวิเคราะห์พฤติกรรมการเรียนรู้ระหว่างการฝึกสอนและการทดสอบประสิทธิภาพขั้นสุดท้ายกับชุดข้อมูลทดสอบจำนวน 400 ภาพ

### 3.1 พฤติกรรมการเรียนรู้และระยะเวลาการฝึกสอนโมเดล

ในกระบวนการฝึกสอนโมเดลใช้จำนวนรอบการฝึกสอน 15 รอบ งานวิจัยนี้ได้ใช้ชุดข้อมูลสำหรับตรวจสอบความถูกต้องระหว่างฝึก จำนวน 400 ภาพ เพื่อเฝ้าระวังและป้องกันปัญหาการเรียนรู้แบบท่องจำ จากการบันทึกผลการทดลองพบว่าโมเดลแต่ละสถาปัตยกรรมใช้ระยะเวลาในการฝึกสอนและให้ค่าความแม่นยำสูงสุด (Validation Accuracy) ที่แตกต่างกัน โดยสถาปัตยกรรม DenseNet-121 ใช้เวลาประมวลผลน้อยที่สุดเพียง 10 นาที 56 วินาที พร้อมให้ความแม่นยำสูงสุดที่ร้อยละ 98.50 รองลงมาคือ Swin-Tiny ใช้เวลา 13 นาที 42 วินาที (ความแม่นยำร้อยละ 98.25) และ ResNet-50 ใช้เวลา 30 นาที 57 วินาที (ความแม่นยำร้อยละ 98.25) ในขณะที่ ViT-Base ใช้เวลาประมวลผลนานที่สุดถึง 35 นาที 47 วินาที โดยให้ความแม่นยำที่ร้อยละ 97.75

จากข้อมูลข้างต้นแสดงให้เห็นว่า DenseNet-121 มีความโดดเด่นสูงสุดทั้งในด้านความแม่นยำและการใช้ทรัพยากรเวลาอย่างมีประสิทธิภาพ ในขณะที่ ViT-Base ใช้เวลาในการประมวลผลนานที่สุดแต่กลับให้ความแม่นยำต่ำที่สุด ซึ่งสอดคล้องกับข้อจำกัดของสถาปัตยกรรมกลุ่มวิชันทรานส์ฟอร์เมอร์ดั้งเดิมเมื่อเผชิญกับชุดข้อมูลขนาดจำกัด

### 3.2 ประสิทธิภาพบนชุดข้อมูลทดสอบ

เพื่อประเมินความสามารถในการนำไปประยุกต์ใช้งานจริง โมเดลทั้ง 4 สถาปัตยกรรมถูกนำมาทดสอบกับชุดข้อมูลสำหรับทดสอบประสิทธิภาพ จำนวน 400 ภาพ ซึ่งเป็นชุดข้อมูลที่โมเดลไม่เคยเห็นมาก่อนในกระบวนการฝึกสอน การประเมินผลอาศัยตัวชี้วัดประสิทธิภาพระดับมหภาค (Macro-Average) ได้แก่ ความแม่นยำรวม ความแม่น ความระลึก และค่าเอฟวัน โดยคำนวณจากสมการที่ (1)–(4) ดังต่อไปนี้ [16]

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precison = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

**ตารางที่ 1** เปรียบเทียบประสิทธิภาพของโมเดลในการจำแนกโรคใบอ้อยบนชุดข้อมูลทดสอบ

| โมเดล | ค่าความแม่นยำรวม (%) | ค่าความแม่นยำ | ค่าความระลึก | ค่าเอฟวัน |
|---|---|---|---|---|
| ResNet-50 | 99.00 | 0.9902 | 0.9900 | 0.9900 |
| DenseNet-121 | 99.25 | 0.9926 | 0.9925 | 0.9925 |
| ViT-Base | 97.00 | 0.9706 | 0.9700 | 0.9700 |
| Swin-Tiny | 98.50 | 0.9850 | 0.9850 | 0.9850 |

โดยที่ TP (True Positive) หมายถึง จำนวนภาพที่โมเดลทำนายว่ามีโรคและผลจริงมีโรคนั้นตรงกัน TN (True Negative) หมายถึง จำนวนภาพที่โมเดลทำนายว่าปกติและผลจริงเป็นปกติ FP (False Positive) หมายถึง จำนวนภาพที่โมเดลทำนายว่ามีโรคแต่ผลจริงเป็นภาพ

ปกติ และ FN (False Negative) หมายถึง จำนวนภาพที่โมเดลทำนายว่าปกติแต่ผลจริงเป็นภาพที่มีโรค ซึ่งผลการเปรียบเทียบประสิทธิภาพของโมเดลทั้ง 4 สถาปัตยกรรมบนชุดข้อมูลสำหรับทดสอบประสิทธิภาพได้ถูกรวบรวมและแสดงไว้ในตารางที่ 1

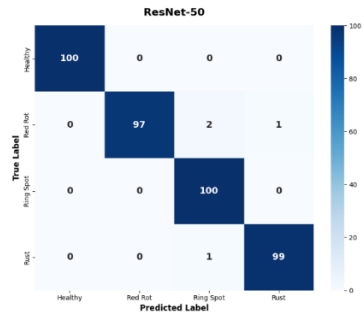
จากตารางที่ 1 ผลการประเมินสมรรถนะของโมเดลบนชุดข้อมูลทดสอบจำนวน 400 ภาพ พบว่าโมเดลในกลุ่มสถาปัตยกรรมโครงข่ายประสาทเทียมแบบคอนโวลูชันให้ผลลัพธ์โดยรวมที่สูงกว่ากลุ่มวิชันทรานส์ฟอร์เมอร์อย่างชัดเจน โดยสถาปัตยกรรม DenseNet-121 มีสมรรถนะดีที่สุดในทุกตัวชี้วัด ซึ่งให้ค่าคือ ความแม่นยำรวม สูงสุดถึงร้อยละ 99.25 พร้อมทั้งมีค่าความแม่นยำ ค่าความระลึก และค่าเอฟวัน สูงถึง 0.9926 0.9925 และ 0.9925 ตามลำดับ รองลงมาคือ สถาปัตยกรรม ResNet-50 ที่ให้ความแม่นยำร้อยละ 99.00 ในขณะเดียวกัน สำหรับกลุ่มวิชันทรานส์ฟอร์เมอร์พบว่าโมเดล Swin-Tiny สามารถทำความแม่นยำได้ที่ร้อยละ 98.50 ซึ่งสูงกว่าโมเดล ViT-Base ดั้งเดิมที่ให้ผลการประเมินต่ำที่สุดในกลุ่มการทดลองนี้ด้วยความแม่นยำร้อยละ 97.00

**ตารางที่ 2** เปรียบเทียบสมรรถนะของโมเดลในงานวิจัยนี้กับงานที่ผ่านมา

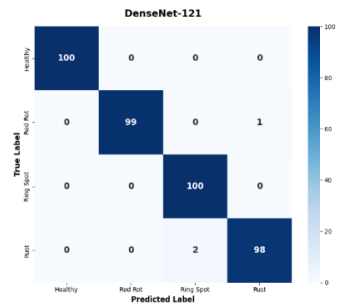
| งานวิจัย | โมเดล | ชุดข้อมูล | ค่าความแม่นยำรวม (%) |
|---|---|---|---|
| Upadhye และคณะ [5] | CNN | โรคอ้อย (ไม่ระบุจำนวน) | 98.69 |
| Baipod และคณะ [8] | CNN | ใบอ้อย 4 คลาส | 97.28 |
| Ali และคณะ [17] | ViT | โรคพืช 55 คลาส 83,603 ภาพ | 89.30 |
| งานวิจัยนี้ - DenseNet-121 | CNN | ใบอ้อย 4 คลาส 4,000 ภาพ | 99.25 |

เพื่อวางบริบทเชิงเปรียบเทียบของผลลัพธ์ที่ได้ตารางที่ 2 แสดงการเปรียบเทียบสมรรถนะของโมเดลที่มีประสิทธิภาพสูงสุดในงานวิจัยนี้ ได้แก่ DenseNet-121 กับงานวิจัยที่เกี่ยวข้องในด้านการจำแนกโรคใบอ้อย และโรคพืชในกลุ่มใกล้เคียง

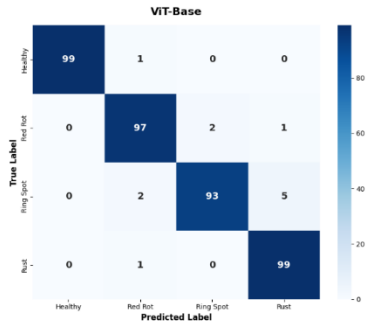
นอกจากผลการเปรียบเทียบประสิทธิภาพของโมเดลในตารางที่ 1 แล้ว งานวิจัยนี้ได้นำเสนอเมทริกซ์ความสับสน (Confusion Matrix) ของโมเดลทั้ง 4 สถาปัตยกรรม ดังแสดงในรูปที่ 1 เพื่อวิเคราะห์พฤติกรรมการทำนาย และประเมินความสามารถในการจำแนกแบบรายคลาส (Class-wise Classification) ซึ่งเป็นวิธีการที่ช่วยให้สามารถระบุจุดอ่อนของโมเดลในการแยกแยะลักษณะเด่นของรอยโรคบนใบอ้อยแต่ละชนิดได้ [18] ดังรูปที่ 1–4 ตามลำดับ



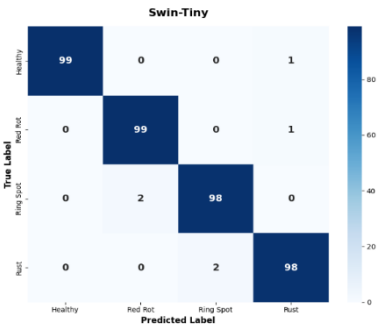

**รูปที่ 1** Confusion Matrix ในการจำแนกโรคใบอ้อยด้วยชุดข้อมูล Testing set ของโมเดล ResNet-50




**รูปที่ 2** Confusion Matrix ในการจำแนกโรคใบอ้อยด้วยชุดข้อมูล Testing set ของโมเดล DenseNet-121




**รูปที่ 3** Confusion Matrix ในการจำแนกโรคใบอ้อยด้วยชุดข้อมูล Testing set ของโมเดล ViT-Base




**รูปที่ 4** Confusion Matrix ในการจำแนกโรคใบอ้อยด้วยชุดข้อมูล Testing set ของโมเดล Swin-Tiny

จากรูปที่ 1–4 เมื่อพิจารณาค่าความแม่นยำและความระลึกแยกรายคลาสพบว่า DenseNet-121 มีสมรรถนะที่โดดเด่นและสม่ำเสมอที่สุด โดยได้ค่าความแม่นยำเท่ากับ 1.0000 ในคลาส Healthy และ Red Rot และมีค่าไม่ต่ำกว่า 0.9800 ในทุกคลาส ในทางตรงข้าม ViT-Base แสดงจุดอ่อนที่ชัดเจนที่สุดในคลาส Ring Spot ซึ่งได้ค่าความระลึกต่ำสุดที่ 0.9300 และค่าความแม่นยำของคลาส Rust ที่ 0.9429 สะท้อนให้เห็นถึงความสับสนระหว่างสองคลาสนี้ เนื่องจากทั้งสองโรคปรากฏเป็นจุดรอยโรคขนาดเล็กที่มีลักษณะสีและรูปร่างใกล้เคียงกัน ซึ่งต้องอาศัยการสกัดคุณลักษณะเฉพาะจุดที่เป็นจุดแข็งของสถาปัตยกรรมโครงข่ายประสาทเทียมแบบคอนโวลูชันมากกว่าสถาปัตยกรรมวิชันทรานส์ฟอร์เมอร์ที่เน้นการวิเคราะห์บริบทภาพรวม

### 3.3 การวิเคราะห์เชิงลึกและอภิปรายผล

จากผลการประเมินในตารางที่ 1 และเมทริกซ์ความสับสนในรูปที่ 1–4 สามารถวิเคราะห์อภิปรายผลเชิงลึกทางสถาปัตยกรรมได้ดังนี้

3.3.1 โครงข่าย DenseNet-121 แสดงให้เห็นถึงประสิทธิภาพสูงสุดในทุกมิติ ทั้งเวลาการฝึกสอนที่สั้นที่สุด และอัตราการทำนายผิดพลาดที่ต่ำที่สุดในทุกคลาส ดังจะเห็นได้จากรูปที่ 2 ที่โมเดลมีการทำนายผิดพลาดเพียง 3 ภาพ จากชุดทดสอบทั้งหมด 400 ภาพ ปัจจัยสำคัญมาจากโครงสร้างการส่งต่อข้อมูลแบบหนาแน่น (Dense Connectivity) [11] ที่ช่วยให้เครือข่ายสามารถนำคุณลักษณะ (Features) จากชั้นก่อนหน้ามาใช้ซ้ำได้อย่างมีประสิทธิภาพลดปัญหาความชันสูญหาย (Vanishing Gradient) [19] ทำให้เครือข่ายเรียนรู้ลักษณะของรอยโรคที่มีขนาดเล็ก และซับซ้อนบนใบอ้อยได้อย่างรวดเร็วและแม่นยำ

3.3.2 ความท้าทายของสถาปัตยกรรมวิชันทรานส์ฟอร์เมอร์บนข้อมูลแม้ ViT-Base จะมีกลไกการให้ความสนใจในตนเองที่ทรงพลัง แต่เมื่อฝึกสอนด้วยชุดข้อมูลเพียง 4,000 ภาพ โมเดลแสดงสมรรถนะต่ำกว่ากลุ่มสถาปัตยกรรมโครงข่ายประสาทเทียมแบบคอนโวลูชันอย่างมีนัยสำคัญ ดังที่ปรากฏในรูปที่ 3 ซึ่งพบความสับสนในการแยกแยะระหว่างโรคใบจุดวงแหวนและโรคราสนิมมากที่สุด โดยมีการทำนายผิดพลาดรวม 12 ภาพ อันเป็นผลจากความคล้ายคลึงของลักษณะสีและรูปร่างรอยโรคระหว่างสองคลาสดังกล่าว ทั้งนี้ สอดคล้องกับ Dosovitskiy และคณะ [12] ที่ระบุว่าสถาปัตยกรรมวิชัน

ทรานส์ฟอร์เมอร์ขาดความเอนเอียงเชิงอุปนัย (Inductive Bias) ทำให้ต้องการชุดข้อมูลขนาดใหญ่เพื่อการเรียนรู้ที่มีประสิทธิภาพและสอดคล้องกับ Ali และคณะ [17] ที่พบว่า ViT ให้ความแม่นยำเพียงร้อยละ 89.30 แม้ฝึกสอนบนชุดข้อมูลโรคพืชกว่า 83,000 ภาพ สะท้อนให้เห็นข้อจำกัดเชิงสถาปัตยกรรมของ ViT ในการประยุกต์ใช้กับชุดข้อมูลขนาดจำกัดในงานวิเคราะห์ภาพทางการเกษตร

3.3.3 การเพิ่มประสิทธิภาพด้วย Swin Transformer เป็นที่น่าสังเกตว่า Swin-Tiny ซึ่งประยุกต์ใช้กลไกการคำนวณความสนใจภายในหน้าต่างย่อยแบบเลื่อน (Shifted Window Attention) [13] แสดงให้เห็นถึงความสามารถในการแก้ปัญหาข้อจำกัดของ ViT-Base ได้อย่างมีนัยสำคัญ ดังที่ปรากฏในรูปที่ 5 ซึ่งมีอัตราการทำนายผิดพลาดลดลงอย่างชัดเจนเมื่อเทียบกับรูปที่ 3 ทั้งนี้ เนื่องจากโครงสร้างเชิงลำดับชั้น (Hierarchical Structure) ของ Swin-Tiny ที่ผสานคุณสมบัติเชิงพื้นที่คล้ายสถาปัตยกรรมโครงข่ายประสาทเทียมแบบคอนโวลูชันเข้ากับกลไก Attention ส่งผลให้สามารถสกัดคุณลักษณะรอยโรครายคลาสได้ดีกว่า ViT-Base อย่างมีนัยสำคัญ

เมื่อพิจารณาในมิติความคุ้มค่าเชิงทรัพยากร DenseNet-121 มีความโดดเด่นสูงสุดในเชิงสมดุลระหว่างความแม่นยำและทรัพยากรที่ใช้ โดยมีจำนวนพารามิเตอร์เพียง 8.0 ล้านพารามิเตอร์ [11] น้อยกว่า ResNet-50 (25.6 ล้านพารามิเตอร์) [10] Swin-Tiny (28.3 ล้านพารามิเตอร์) [13] และ ViT-Base (86.6 ล้านพารามิเตอร์) [12] อย่างมีนัยสำคัญ ในขณะที่ยังคงให้ความแม่นยำสูงสุดจึงเป็นตัวเลือกที่เหมาะสมที่สุดสำหรับการนำไปประยุกต์ใช้บนอุปกรณ์ที่มีทรัพยากรจำกัด เช่น สมาร์ทโฟนหรืออุปกรณ์ Edge Computing ในระบบเกษตรอัจฉริยะ

**4. สรุปผลการทดลอง**

งานวิจัยนี้ได้นำเสนอการศึกษาเชิงเปรียบเทียบสมรรถนะของโมเดลการเรียนรู้เชิงลึก 2 กลุ่มสถาปัตยกรรม ได้แก่ กลุ่มโครงข่ายประสาทเทียมแบบคอนโวลูชัน (ResNet-50 และ DenseNet-121) และกลุ่มวิชันทรานส์ฟอร์เมอร์ (ViT-Base และ Swin-Tiny) สำหรับการจำแนกโรคใบอ้อย 4 ประเภท ได้แก่ ใบปกติ โรคเหี่ยวแดง โรคใบจุดวงแหวน และโรคราสนิม จากชุดข้อมูลภาพถ่ายในสภาพแวดล้อมจริงจำนวน 4,000 ภาพ

ผลการทดลองแสดงให้เห็นอย่างชัดเจนว่าสถาปัตยกรรมโครงข่ายประสาทเทียมแบบคอนโวลูชันมีความเหมาะสมสูงสุดสำหรับการวิเคราะห์ภาพทางการเกษตรที่มีข้อจำกัดด้านขนาดชุดข้อมูล โดย DenseNet-121 เป็นสถาปัตยกรรมที่มีสมรรถนะสูงสุดในการศึกษานี้ด้วยค่าความแม่นยำรวมบนชุดข้อมูลทดสอบร้อยละ 99.25 เมื่อเปรียบเทียบกับงานวิจัยที่ผ่านมาดังแสดงในตารางที่ 2 พบว่า DenseNet-121 ให้ความแม่นยำสูงกว่าร้อยละ 95.60 ของงานวิจัยก่อนหน้าที่ใช้สถาปัตยกรรมใกล้เคียงกันถึงร้อยละ 3.65 และยังให้ความแม่นยำสูงกว่างานวิจัยอื่นที่ใช้ CNN บนชุดข้อมูลลักษณะเดียวกันอีกด้วยอันเป็นผลจากสถาปัตยกรรม Dense Connectivity ที่สามารถสกัดคุณลักษณะรอยโรคขนาดเล็กได้อย่างละเอียด ประกอบกับค่าความแม่นยำ ค่าความระลึก และค่าเอฟวัน ที่ 0.9926 0.9925 และ 0.9925 ตามลำดับ สะท้อนให้เห็นถึงความสม่ำเสมอของสมรรถนะในทุกคลาสโรค นอกจากนี้ DenseNet-121 ยังมีความคุ้มค่าเชิงทรัพยากรสูงสุด โดยใช้ระยะเวลาฝึกสอนเพียง 10 นาที 56 วินาที และมีจำนวนพารามิเตอร์น้อยที่สุดในบรรดาโมเดลทั้งหมดที่ทดสอบ ในขณะที่ ViT-Base ยังคงมีข้อจำกัดเมื่อฝึกสอนด้วยชุดข้อมูลขนาดจำกัด อย่างไรก็ตาม Swin-Tiny แสดงให้เห็นว่าการปรับ

ปรุงกลไก Attention ด้วยโครงสร้าง Shifted Window สามารถยกระดับสมรรถนะของกลุ่มสถาปัตยกรรมวิชันทรานส์ฟอร์เมอร์ให้ใกล้เคียงกับสถาปัตยกรรมโครงข่ายประสาทเทียมแบบคอนโวลูชันได้อย่างมีนัยสำคัญ

ผลการวิจัยยืนยันศักยภาพของ DenseNet-121 สำหรับการใช้งานจริงในสภาพแวดล้อมที่มีความผันแปร และเป็นพื้นฐานสำหรับการพัฒนาต่อยอดสู่ระบบวินิจฉัยและแจ้งเตือนโรคใบอ้อยแบบเรียลไทม์ โดยบูรณาการกับเทคโนโลยีอินเทอร์เน็ตของสรรพสิ่ง (Internet of Things; IoT) ระบบคลาวด์ และแอปพลิเคชันบนสมาร์ทโฟน เพื่อส่งเสริมการเฝ้าระวังโรคพืชเชิงรุกและระบบเกษตรอัจฉริยะ (Smart Agriculture) ต่อไป

### 4.1 ข้อจำกัดและทิศทางการวิจัยในอนาคต

งานวิจัยนี้มีข้อจำกัดด้านขนาดชุดข้อมูลจำนวนคลาสโรคและการฝึกสอนจากค่าเริ่มต้น โดยไม่ประยุกต์ใช้การเรียนรู้เชิงถ่ายโอน (Transfer Learning) ซึ่งอาจจำกัดศักยภาพของสถาปัตยกรรมกลุ่มวิชันทรานส์ฟอร์เมอร์ดังที่ปรากฏในผลการทดลองของ ViT-Base ดังนั้นงานวิจัยในอนาคตจึงมีแผนขยายขอบเขตชุดข้อมูล และจำนวนคลาสโรค ศึกษาผลของการเรียนรู้เชิงถ่ายโอนจาก ImageNet ประเมินโมเดลขนาดเบา (Lightweight Model) เช่น MobileNetV3 และ EfficientNet-Lite สำหรับอุปกรณ์ปลายทาง (Edge Device) และพัฒนาต้นแบบระบบวินิจฉัยโรคแบบเรียลไทม์โดยบูรณาการ DenseNet-121 กับแอปพลิเคชันบนสมาร์ทโฟน และระบบเทคโนโลยีอินเทอร์เน็ตของสรรพสิ่ง เพื่อรองรับการใช้งานในระบบเกษตรอัจฉริยะต่อไป

### เอกสารอ้างอิง

[1] J. Mauya, R. Amin, M. I. Hossain, S. Ruhi, and M. S. Reza, “Improved multiscale attention based deep learning approach for automated sugarcane leaf disease detection using BSRI data,” *Scientific Reports*, vol. 15, no. 1, Dec. 2025, Art. no. 45474, doi: 10.1038/s41598-025-28947-x.

[2] M. H. Saleem, J. Potgieter, and K. M. Arif, “Plant disease detection and classification by deep learning,” *Plants (Basel, Switzerland)*, vol. 8, no. 11, Oct. 2019, Art. no. 468, doi: 10.3390/plants8110468.

[3] L. Chen, Y. Yuan, and H. He, “Image recognition for crop diseases using a novel multi-attention module,” *International Journal of Agricultural and Biological Engineering*, vol. 18, no. 1, pp. 238–244, Mar. 2025, doi: 10.25165/j.ijabe.20251801.8136.

[4] I. Pacal, I. Kunduracioğlu, M. H. Alma, M. Deveci, S. Kadry, J. Nedoma, V. Slaný, and R. Martinek, “A systematic review of deep learning techniques for plant diseases,” *Artificial Intelligence Review*, vol. 57, no. 11, Sep. 2024, Art. no. 304, doi: 10.1007/s10462-024-10944-7.

[5] S. A. Upadhye, M. R. Dhanvijay, and S. M. Patil, “Sugarcane disease detection using CNN-deep learning method: An Indian perspective,” *Universal Journal of Agricultural Research*, vol. 11, no. 1,

pp. 80–97, Feb. 2023, doi: 10.13189/ujar.2023.110108.

[6] C. Özden, " Vision transformer and CNN-based models for image analysis of plant diseases: An approach for agricultural decision support systems," *ISPEC Journal of Agricultural Sciences*, vol. 10, no. 1, pp. 109–121, Jan. 2026, doi: 10.5281/zenodo.18228928.

[7] A. Demir, F. Sarı, M. Arzu, and M. Kaya, " Artificial intelligence-aided diagnosis in agriculture: Plant disease classification with vision transformer and CNN models," *NATURENGS*, vol. 6, no. 2, pp. 22–31, Dec. 2025, doi: 10.46572/naturengs.1832476.

[8] S. Baipod, C. Savithi, and J. Sriwongsa, " Applying convolutional neural networks for sugarcane leaf disease image classification and integration with LINE chatbot to support smart agriculture," *Journal of Computer and Creative Technology*, vol. 3, no. 2, pp. 317–332, Dec. 2025 (in Thai), doi: 10.14456/jcct.2025.24.

[9] S. D. Daphal and S. M. Koli, " Enhanced deep learning technique for sugarcane leaf disease classification and mobile application integration," *Heliyon*, vol. 10, no. 8, Apr. 2024, Art. no. e29438, doi: 10.1016/j.heliyon.2024.e29438.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[11] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, " Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA*, 2017, pp. 4700–4708, doi: 10.1109/CVPR.2017.243.

[12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, " An image is worth 16x16 words: Transformers for image recognition at scale," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021, doi: 10.48550/arXiv.2010.11929.

[13] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, " Swin transformer: Hierarchical vision transformer using shifted windows," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada*, 2021, pp. 10012–10022, doi: 10.1109/ICCV48922. 2021.00986.

[14] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of*

*Big Data*, vol. 3, no. 1, May 2016, Art. no. 9, doi: 10.1186/s40537-016-0043-6.

[15] D. P. Kingma and J. Ba, " Adam: A method for stochastic optimization, " *International Conference on Learning Representation*, May 2015, doi: 10.48550/arXiv.1412.6980.

[16] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm. 2009.03.002.

[17] M. Ali, M. Salma, M. El Haji, and B. Jamal, " Plant disease detection using vision transformers," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 15, no. 2, pp. 2334–2344, Apr. 2025, doi: 10.11591/ijece.v15i2.pp2334-2344.

[18] D. A. Padilla, G. V. Magwili, A. L. A. Marohom, C. M. G. Co, J. C. C. Gaño, and J. M. U. Tuazon, " Portable yellow spot disease identifier on sugarcane leaf via image processing using support vector machine," in *2019 5th International Conference on Control, Automation and Robotics (ICCAR) 2019*, pp. 901–905, doi: 10.1109/ICCAR.2019.8813495.

[19] S. Hochreiter, " The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 2, pp. 107–116, Apr. 1998, doi: 10.1142/S0218488598000094.