



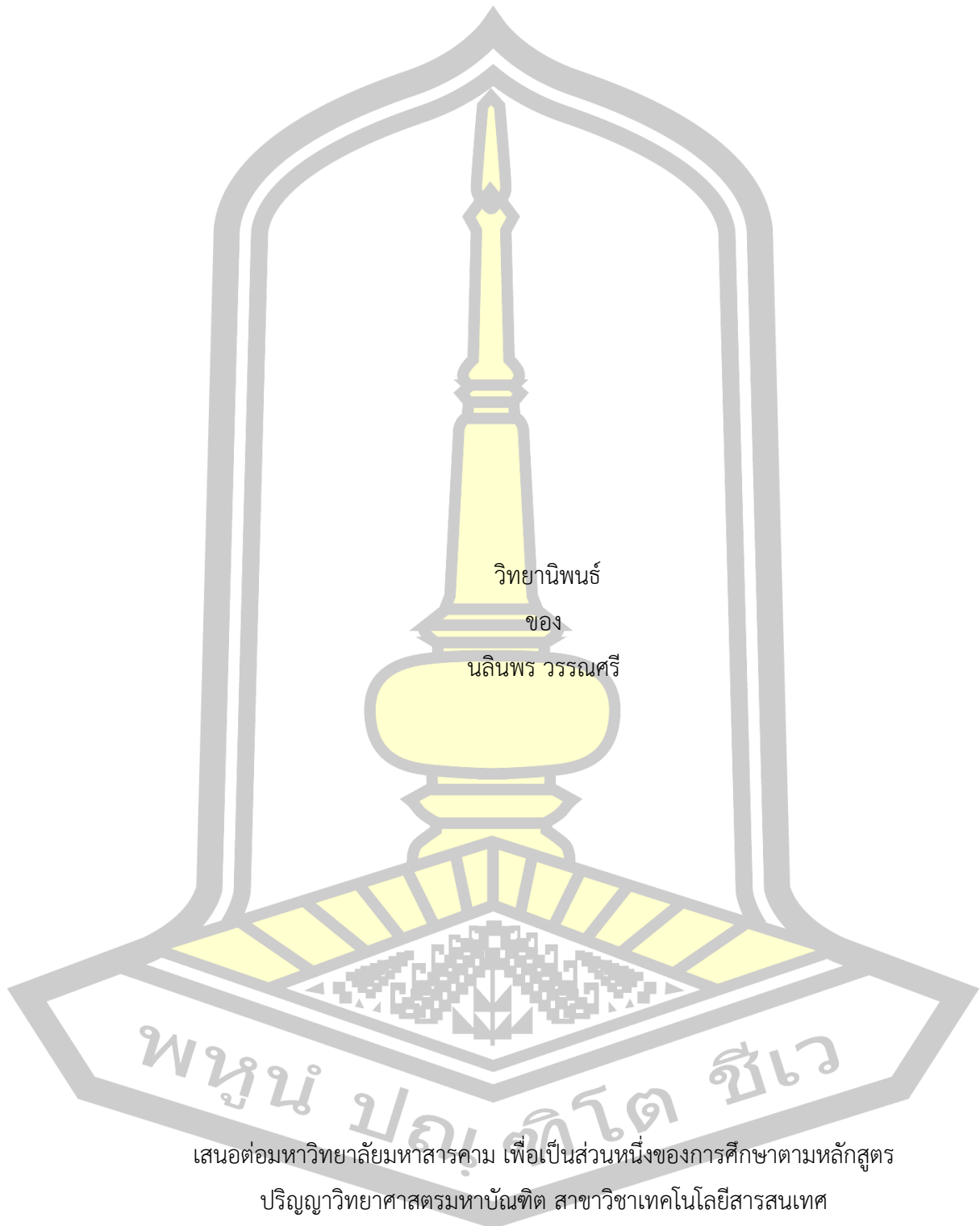
การจำแนกสิ่งปกคลุมดินจากข้อมูลภาพถ่ายดาวเทียมเซนติเนล -2 ด้วยเทคนิคการเรียนรู้ของเครื่อง

วิทยานิพนธ์  
ของ  
นลินพร วรรณศรี

เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ  
พฤษภาคม 2568

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

การจำแนกสิ่งปกคลุมดินจากข้อมูลภาพถ่ายดาวเทียมเซนติเนล -2 ด้วยเทคนิคการเรียนรู้ของเครื่อง

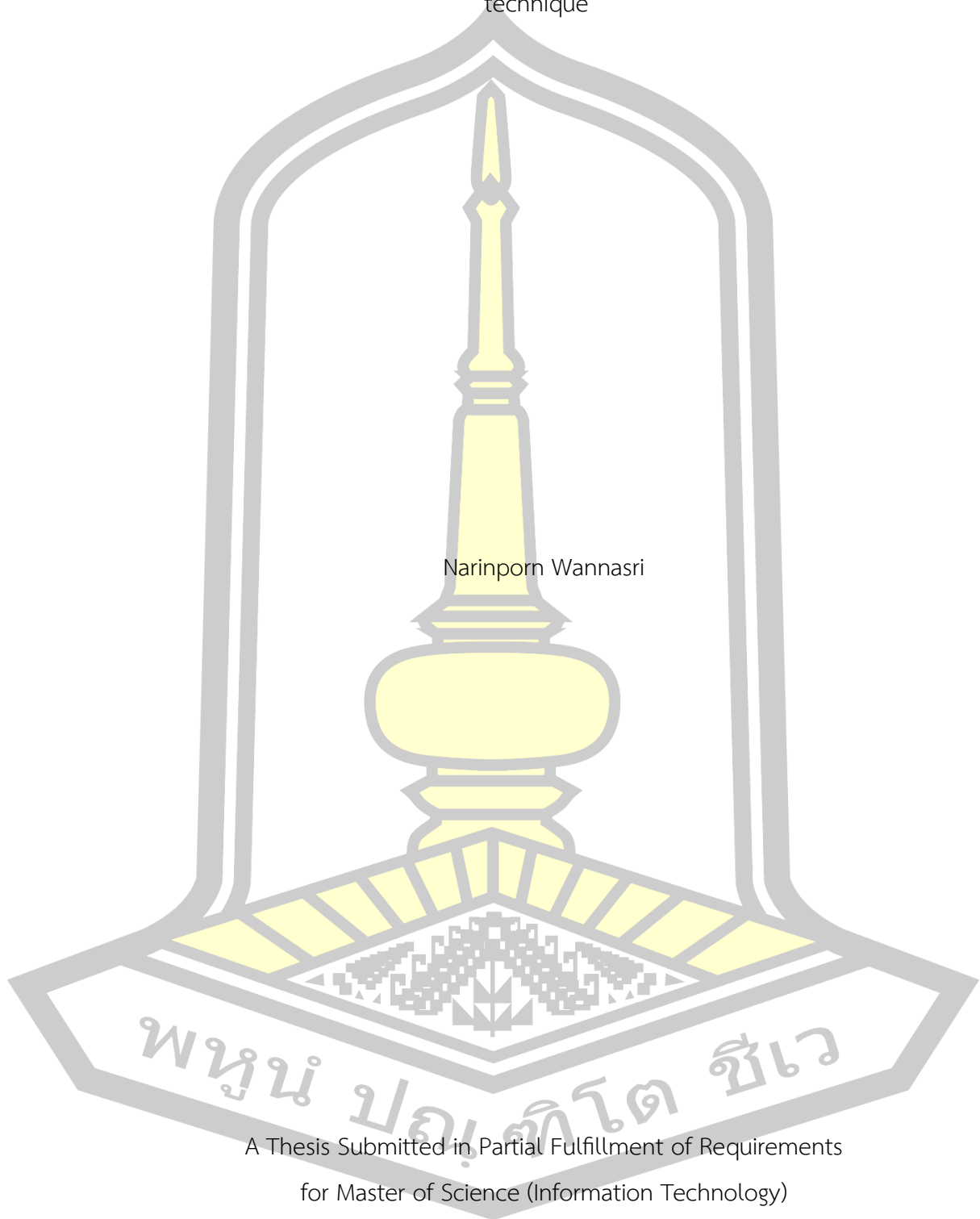


เสนอต่อมหาวิทยาลัยมหาสารคาม เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

พฤษภาคม 2568

ลิขสิทธิ์เป็นของมหาวิทยาลัยมหาสารคาม

Land Cover Classification from Sentinel -2 Satellite Imagery Using Machine Learning  
technique



Narinporn Wannasri

A Thesis Submitted in Partial Fulfillment of Requirements  
for Master of Science (Information Technology)

May 2025

Copyright of Maharakham University



คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาวิทยานิพนธ์ของนางสาวนลินพร วรรณศรี  
แล้วเห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต  
สาขาวิชาเทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

คณะกรรมการสอบวิทยานิพนธ์

ประธานกรรมการ

(ผศ. ดร. ธาดา จันทะคุณ )

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผศ. ดร. สาธิต แสงประดิษฐ์ )

กรรมการ

(ผศ. ดร. จารีย์ ทองคำ )

กรรมการผู้ทรงคุณวุฒิภายนอก

(ผศ. ดร. นฤเศรษฐ์ ประเสริฐศรี )

มหาวิทยาลัยอนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญา วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ของมหาวิทยาลัยมหาสารคาม

(รศ. ดร. จันทิมา พลพินิจ )

คณบดีคณะวิทยาการสารสนเทศ

(ผศ. ดร. พลเดช เขาวรัตน์ )

คณบดีบัณฑิตวิทยาลัย

ชื่อเรื่อง	การจำแนกสิ่งปกคลุมดินจากข้อมูลภาพถ่ายดาวเทียมเซนติเนล -2 ด้วยเทคนิคการเรียนรู้ของเครื่อง		
ผู้วิจัย	นลินพร วรรณศรี		
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. สานิต แสงประดิษฐ์		
ปริญญา	วิทยาศาสตรมหาบัณฑิต	สาขาวิชา	เทคโนโลยีสารสนเทศ
มหาวิทยาลัย	มหาวิทยาลัยมหาสารคาม	ปีที่พิมพ์	2568

### บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อจำแนกประเภทสิ่งปกคลุมดินในอำเภอเมืองร้อยเอ็ด โดยใช้ข้อมูลภาพถ่ายดาวเทียม Sentinel-2 และประเมินประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่อง 4 วิธี ได้แก่ โครงข่ายประสาทเทียม, ต้นไม้ตัดสินใจ, ป่าสุ่ม และซัพพอร์ตเวกเตอร์แมชชีน. พื้นที่ศึกษาถูกจำแนกออกเป็น 8 ประเภท คือ พื้นที่เกษตรกรรม, เมืองและย่านการค้า, พื้นที่สวนราชการ, พื้นที่ว่างเปล่า, พื้นที่อยู่อาศัย, ป่าไม้, แหล่งน้ำตามธรรมชาติ และพื้นที่สวนสาธารณะ. กระบวนการวิจัยประกอบด้วยเตรียมข้อมูล การสร้างแบบจำลอง การประเมินความถูกต้อง และการวิเคราะห์ผล.

ผลการศึกษาพบว่า เทคนิคโครงข่ายประสาทเทียม ให้ค่าความแม่นยำโดยรวมสูงสุดที่ 83% และค่าสัมประสิทธิ์แคปปา 0.79 ซึ่งสูงกว่าเทคนิคอื่น ๆ. ในขณะที่เทคนิคต้นไม้ตัดสินใจ มีความแม่นยำต่ำสุดที่ 70%. การจำแนกพื้นที่สวนสาธารณะและพื้นที่สวนราชการมีความแม่นยำต่ำเนื่องจากลักษณะพื้นที่ที่มีความซับซ้อนและค่าการสะท้อนแสงที่คล้ายคลึงกัน ทำให้เกิดความสับสนในการจำแนก. การศึกษานี้ชี้ให้เห็นถึงศักยภาพของเทคนิคการเรียนรู้ของเครื่องในการประมวลผลข้อมูลสำรวจระยะไกล ซึ่งเป็นประโยชน์ต่อการจัดการทรัพยากรและการวางแผนการใช้ที่ดิน.

คำสำคัญ : การเรียนรู้ของเครื่อง, การจำแนกสิ่งปกคลุมดิน, โครงข่ายประสาทเทียม

พูน ปณ ทิโต ชีเว

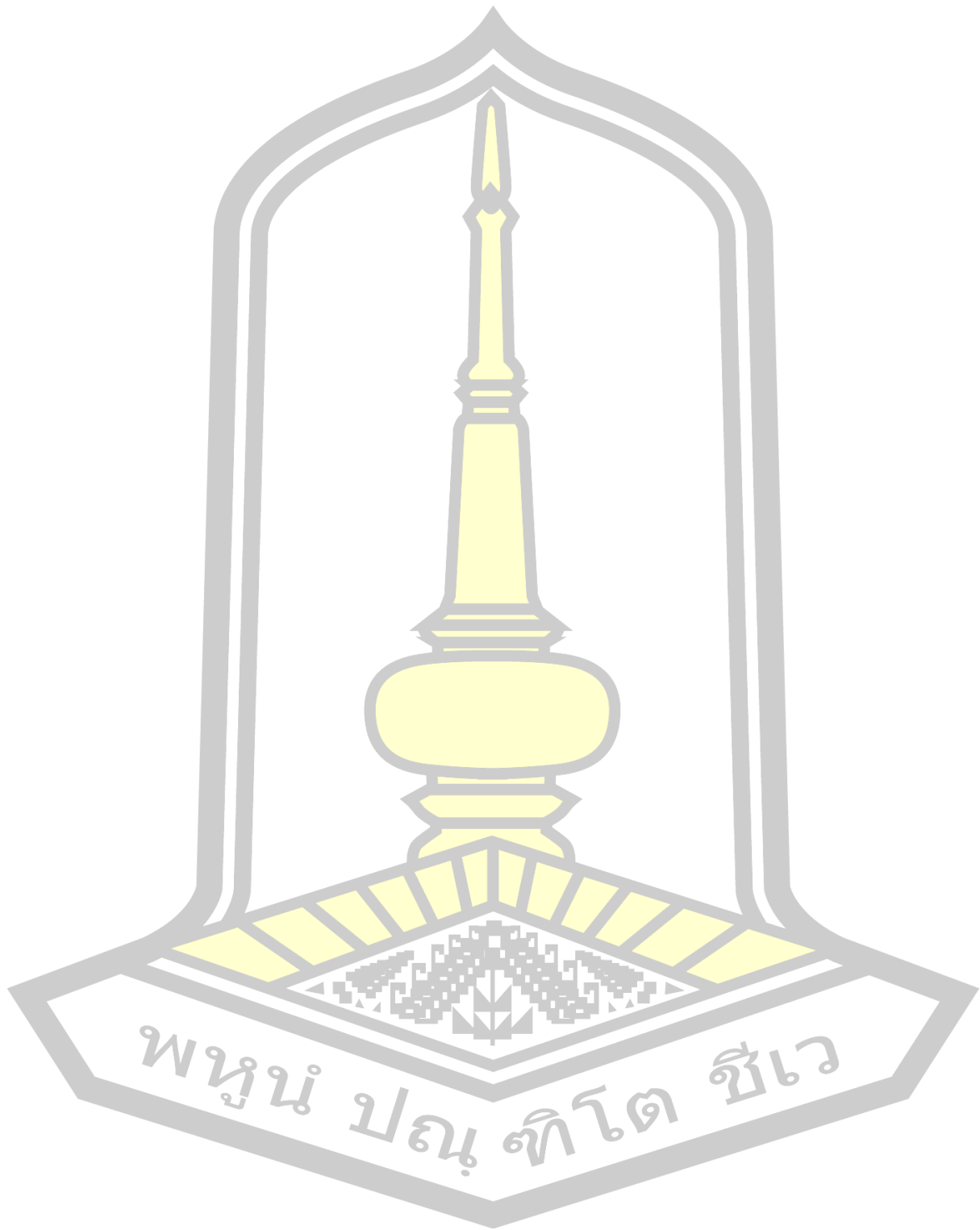
<b>TITLE</b>	Land Cover Classification from Sentinel -2 Satellite Imagery Using Machine Learning technique		
<b>AUTHOR</b>	Narinporn Wannasri		
<b>ADVISORS</b>	Assistant Professor Satith Saengpreedit , Ph.D.		
<b>DEGREE</b>	Master of Science	<b>MAJOR</b>	Information Technology
<b>UNIVERSITY</b>	Maharakham University	<b>YEAR</b>	2025

### ABSTRACT

This study aimed to classify land cover types in Mueang Roi Et District using Sentinel-2 satellite imagery and evaluate the performance of four machine learning techniques: Artificial Neural Network, Decision Tree, Random Forest, and Support Vector Machine. The study area was classified into eight categories: agricultural land, urban and commercial areas, government areas, vacant land, residential areas, forests, natural water sources, and public parks. The research methodology involved data preprocessing, model development, accuracy assessment, and result analysis.

The findings indicated that the Artificial Neural Network technique achieved the highest overall accuracy 83% and Kappa coefficient 0.79, outperforming the other methods. Conversely, the Decision Tree technique yielded the lowest accuracy at 70%. Classification accuracy for public parks and government areas was notably lower, attributed to the complex land use characteristics and similar spectral reflectance values within these areas, leading to classification confusion. This research underscores the potential of machine learning techniques in processing remote sensing data, thereby supporting effective resource management and land-use planning

Keyword : Remote sensing, Machine learning, Sentinel-2



## กิตติกรรมประกาศ

กราบขอบพระคุณ ผศ.ดร.สาธิต แสงประดิษฐ์ อาจารย์ที่ปรึกษาอย่างสูง ที่ได้กรุณาให้คำแนะนำ ปรึกษา และตรวจแก้ไขวิทยานิพนธ์ฉบับนี้มาโดยตลอด คำแนะนำอันเป็นประโยชน์ของอาจารย์ ทำให้ดิฉันสามารถทำวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบคุณ กรมพัฒนาที่ดิน ที่ได้กรุณาให้ข้อมูลเกี่ยวกับสิ่งปกคลุมดินในอำเภอเมืองร้อยเอ็ด และ องค์การอวกาศยุโรป ที่เปิดเผยข้อมูลภาพถ่ายดาวเทียม Sentinel-2 ให้สามารถนำมาใช้ในการวิจัยได้

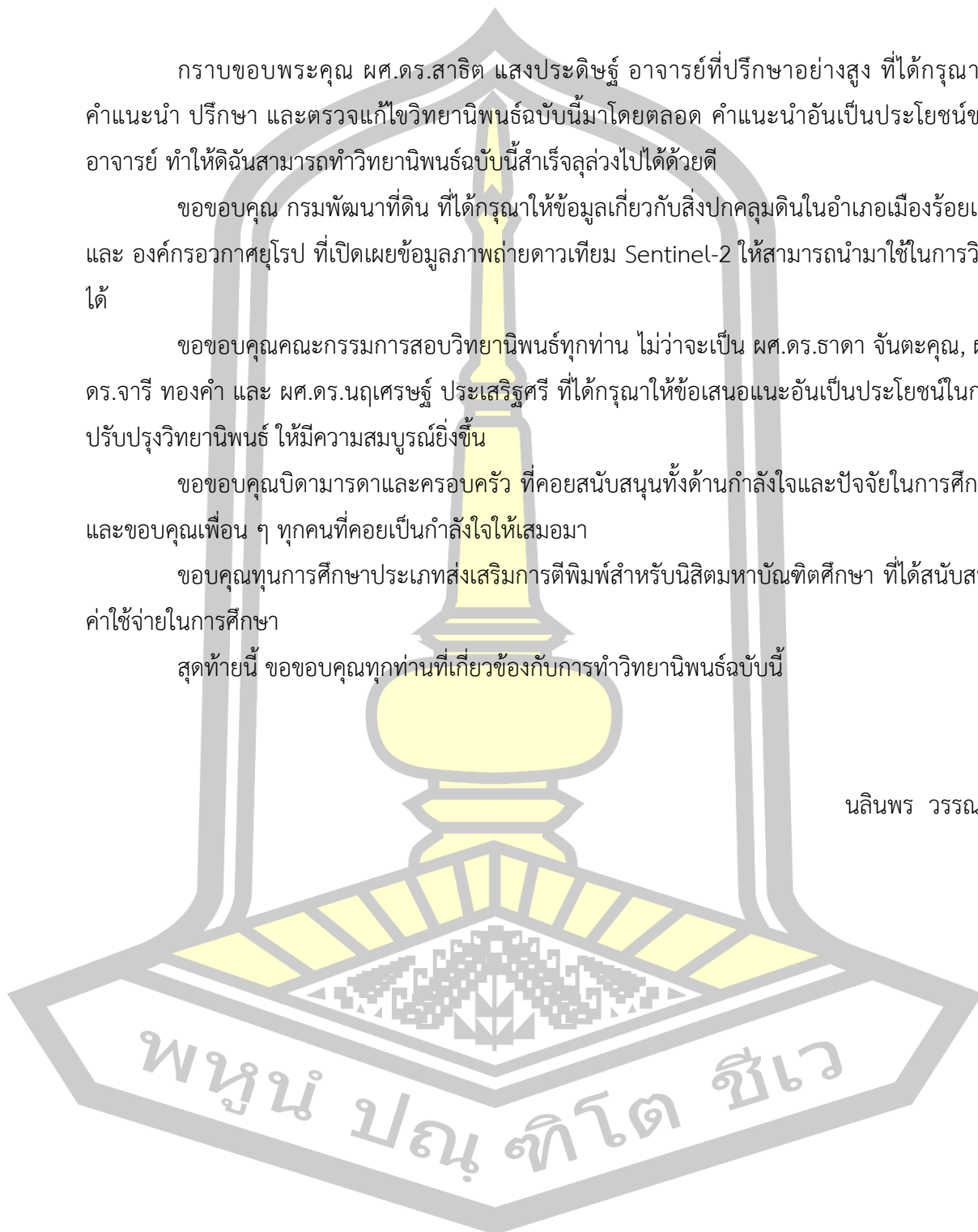
ขอขอบคุณคณะกรรมการสอบวิทยานิพนธ์ทุกท่าน ไม่ว่าจะเป็น ผศ.ดร.ธาดา จันตะคุณ, ผศ.ดร.จารี ทองคำ และ ผศ.ดร.นฤเศรษฐ์ ประเสริฐศรี ที่ได้กรุณาให้ข้อเสนอแนะอันเป็นประโยชน์ในการปรับปรุงวิทยานิพนธ์ ให้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบคุณบิดามารดาและครอบครัว ที่คอยสนับสนุนทั้งด้านกำลังใจและปัจจัยในการศึกษา และขอบคุณเพื่อน ๆ ทุกคนที่คอยเป็นกำลังใจให้เสมอมา

ขอบคุณทุนการศึกษาประเภทส่งเสริมการตีพิมพ์สำหรับนิสิตมหาบัณฑิตศึกษา ที่ได้สนับสนุนค่าใช้จ่ายในการศึกษา

สุดท้ายนี้ ขอขอบคุณทุกท่านที่เกี่ยวข้องกับการทำวิทยานิพนธ์ฉบับนี้

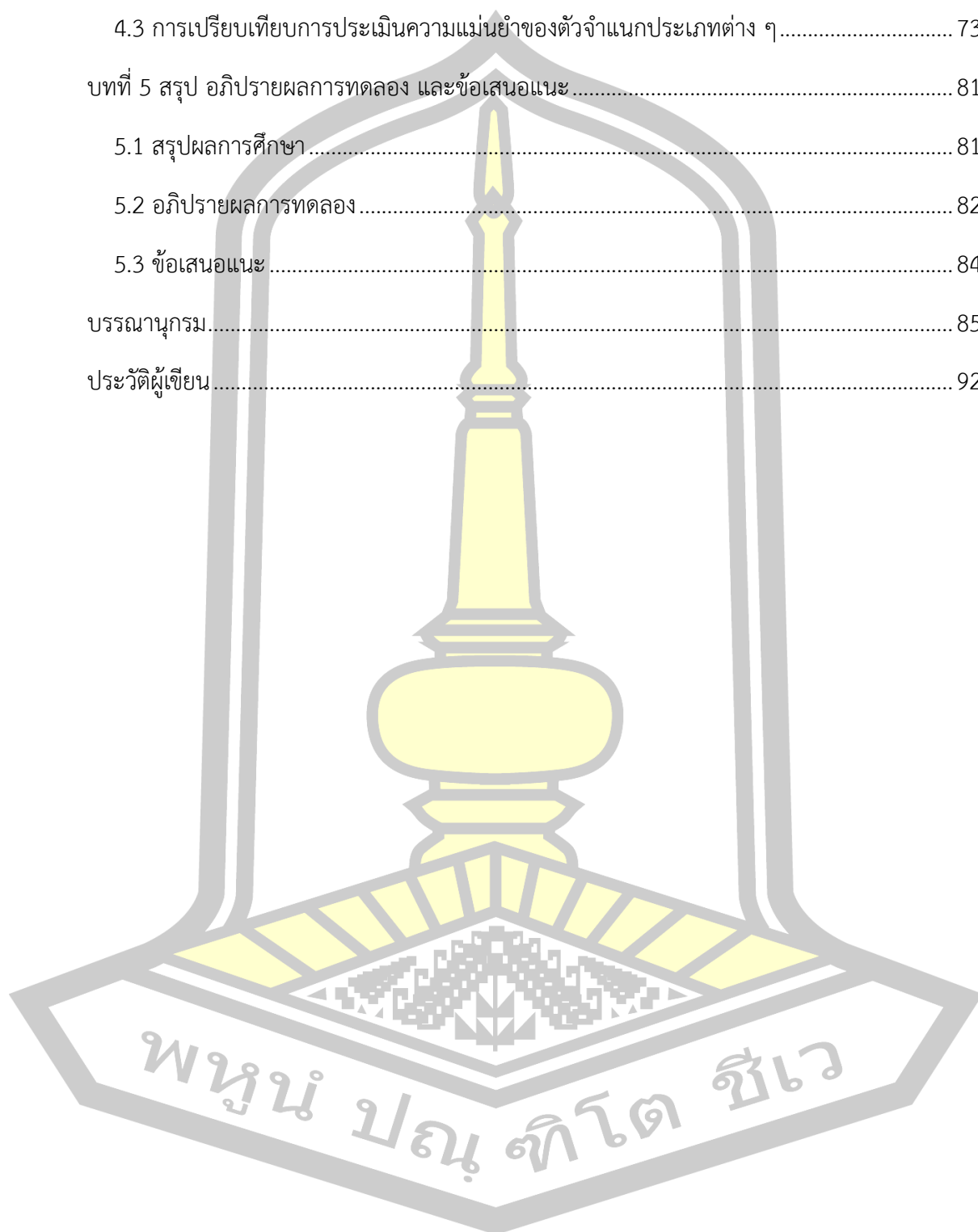
นลินพร วรรณศรี



## สารบัญ

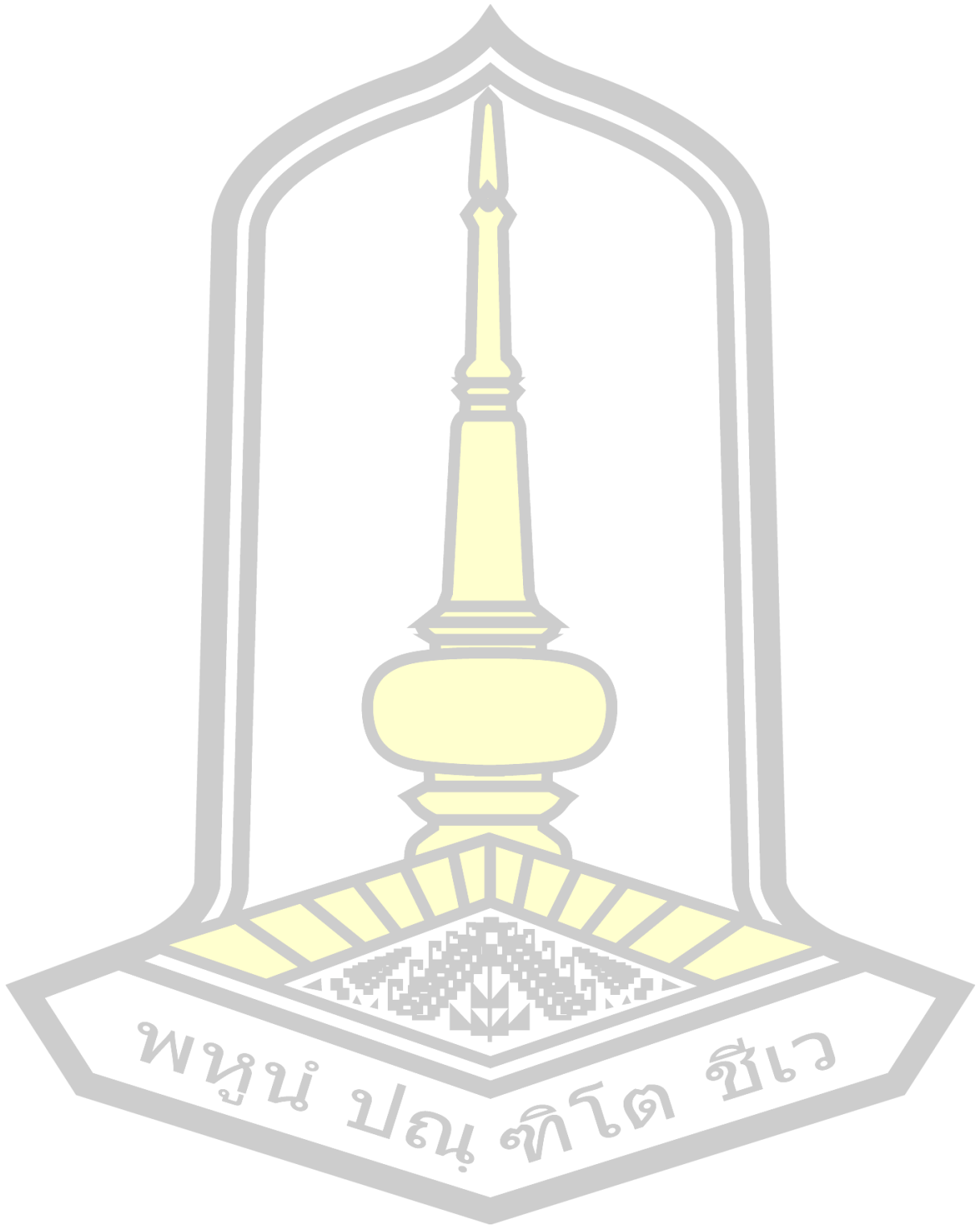
	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ช
สารบัญ.....	ช
สารบัญตาราง.....	ญ
สารบัญรูปภาพ.....	ฎ
บทที่ 1 บทนำ.....	1
1.1 หลักการและเหตุผล.....	1
1.2 วัตถุประสงค์.....	2
1.3 ผลที่คาดว่าจะได้รับ.....	2
1.4 นิยามศัพท์เฉพาะ.....	3
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	4
2.1 ทฤษฎีที่เกี่ยวข้อง.....	4
2.2 งานวิจัยที่เกี่ยวข้อง.....	24
บทที่ 3 ขั้นตอนการดำเนินงาน.....	38
3.1 ข้อมูลที่ใช้ศึกษา.....	38
3.2 การเตรียมข้อมูล.....	39
3.3 การสร้างแบบจำลอง (Model Building).....	40
3.4 การตรวจสอบความถูกต้อง (Accuracy Assessment).....	43
บทที่ 4 ผลการศึกษา.....	46
4.1 ผลการวิเคราะห์ข้อมูลภาพการจำแนกสิ่งปกคลุมดิน.....	46

4.2 ประสิทธิภาพการจำแนกสิ่งปกคลุมดิน .....	67
4.3 การเปรียบเทียบการประเมินความแม่นยำของตัวจำแนกประเภทต่าง ๆ .....	73
บทที่ 5 สรุป อภิปรายผลการทดลอง และข้อเสนอแนะ .....	81
5.1 สรุปผลการศึกษา .....	81
5.2 อภิปรายผลการทดลอง .....	82
5.3 ข้อเสนอแนะ .....	84
บรรณานุกรม .....	85
ประวัติผู้เขียน .....	92



## สารบัญตาราง

	หน้า
ตารางที่ 2.1 ข้อมูลดาวเทียม SENTINEL- 2.....	7
ตารางที่ 2.2 การจำแนกการใช้ประโยชน์ที่ดินของกรมพัฒนาที่ดิน ปี พ.ศ.2555 ระดับที่ 1 และ 2... 8	8
ตารางที่ 2.3 ระดับความสอดคล้องของสถิติแคปตามแนวทางของ Landis และ Koch.....	23
ตารางที่ 4.1 Overall Accuracy และ Overall Kappa Statistics การจำแนก.....	49
ตารางที่ 4.2 ความถูกต้องและความคลาดเคลื่อนของผู้ใช้และผู้ผลิตของเทคนิคต้นไม้ตัดสีนใจ.....	51
ตารางที่ 4.3 แสดงสัดส่วนพื้นที่ที่ได้จากการจำแนกสิ่งปกคลุมดินเทคนิคต้นไม้ตัดสีนใจ.....	52
ตารางที่ 4.4 Overall Accuracy และ Overall Kappa Statistics การจำแนกเทคนิคป่าสุ่ม.....	54
ตารางที่ 4.5 ความถูกต้องและความคลาดเคลื่อนของผู้ใช้และผู้ผลิตของเทคนิคป่าสุ่ม.....	55
ตารางที่ 4.6 แสดงสัดส่วนพื้นที่ที่ได้จากการจำแนกสิ่งปกคลุมดินเทคนิคป่าสุ่ม.....	56
ตารางที่ 4.7 Overall Accuracy และ Overall Kappa Statistics การจำแนกเทคนิคซัพพอร์ต เวกเตอร์แมชชีน.....	59
ตารางที่ 4.8 ความถูกต้องและความคลาดเคลื่อนของผู้ใช้และผู้ผลิตของเทคนิค.....	61
ตารางที่ 4.9 แสดงสัดส่วนพื้นที่ที่ได้จากการจำแนกสิ่งปกคลุมดินเทคนิคซัพพอร์ต.....	62
ตารางที่ 4.10 Overall Accuracy และ Overall Kappa Statistics การจำแนกเทคนิคโครงข่าย ประสาทเทียม.....	64
ตารางที่ 4.11 ความถูกต้องและความคลาดเคลื่อนของผู้ใช้และผู้ผลิตของเทคนิคโครงข่าย.....	65
ตารางที่ 4.12 แสดงพื้นที่ที่ได้จากการจำแนกสิ่งปกคลุมดินเทคนิคโครงข่ายประสาทเทียม.....	66
ตารางที่ 4.13 แสดงค่าความถูกต้องของการจำแนกสิ่งปกคลุมดินที่มีค่าสูงสุดและต่ำสุด.....	68
ตารางที่ 4.14 ค่าเฉลี่ยของผลลัพธ์จาก k fold.....	70
ตารางที่ 4.15 อธิบายผลลัพธ์.....	71
ตารางที่ 4.16 เปรียบเทียบค่า Training Accuracy และ Validation Accuracy.....	72
ตารางที่ 4.17 การเปรียบเทียบการประเมินความแม่นยำของเทคนิคต่าง ๆ.....	74



## สารบัญรูปภาพ

	หน้า
ภาพประกอบที่ 2.1 ขั้นตอนการประมวลผลข้อมูล .....	5
ภาพประกอบที่ 2.2 ตารางประเมินความถูกต้อง .....	13
ภาพประกอบที่ 2.3 แบบโมเดลของโครงข่ายประสาทเทียม .....	16
ภาพประกอบที่ 2.4 Characteristics of Random Forest .....	17
ภาพประกอบที่ 2.5 แบบโมเดลต้นไม้ตัดสินใจ .....	18
ภาพประกอบที่ 2.6 หลักการทำงานของ k-Fold Cross-Validation .....	21
ภาพประกอบที่ 3.1 ดาวนโหนดภาพถ่ายจากดาวเทียม Sentinel -2 .....	39
ภาพประกอบที่ 3.2 แสดงพื้นที่ศึกษา .....	39
ภาพประกอบที่ 4.1 ผลการจำแนกสิ่งปกคลุมดินเทคนิค ต้นไม้ตัดสินใจ .....	48
ภาพประกอบที่ 4.2 ผลการจำแนกสิ่งปกคลุมดินเทคนิคป่าสุ่ม .....	53
ภาพประกอบที่ 4.3 ผลการจำแนกสิ่งปกคลุมดินเทคนิคซัพพอร์ตเวกเตอร์แมชชีน .....	58
ภาพประกอบที่ 4.4 ผลการจำแนกสิ่งปกคลุมดินเทคนิคโครงข่ายประสาทเทียม .....	63
ภาพประกอบที่ 4.5 ประสิทธิภาพการจำแนกความถูกต้องสูงสุดของพื้นที่แหล่งน้ำแต่ละเทคนิค .....	68
ภาพประกอบที่ 4.6 ประสิทธิภาพการจำแนกความถูกต้องสูงสุดของพื้นที่ป่าไม้แต่ละเทคนิค ..	69
ภาพประกอบที่ 4.7 ประสิทธิภาพการจำแนกความถูกต้องสูงสุดของพื้นที่เกษตรกรรมแต่ละเทคนิค	69
ภาพประกอบที่ 4.8 กราฟการเปรียบเทียบประสิทธิภาพของเทคนิค .....	73
ภาพประกอบที่ 4.9 กราฟเปรียบเทียบความแม่นยำรวม (Overall Accuracy) และค่าสัมประสิทธิ์ แคปปา (Overall Kappa Statistics) ของแต่ละเทคนิคการจำแนกสิ่งปกคลุมดิน .....	76
ภาพประกอบที่ 4.10 กราฟเปรียบเทียบค่าประสิทธิภาพของแต่ละเทคนิค .....	77
ภาพประกอบที่ 4.11 กราฟนี้แสดงผลลัพธ์ความแม่นยำ (Accuracy) เทคนิคโครงข่าย .....	79

# บทที่ 1

## บทนำ

### 1.1 หลักการและเหตุผล

แผนที่การใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน (Land Use and Land Cover - LULC) เป็นข้อมูลพื้นฐานที่มีความสำคัญอย่างยิ่งต่อการวางแผนและบริหารจัดการทรัพยากรที่ดินอย่างยั่งยืนและมีประสิทธิภาพ การมีข้อมูล LULC ที่ถูกต้อง แม่นยำ และทันสมัย ช่วยให้เกิดความเข้าใจในสถานการณ์ปัจจุบันของการใช้ที่ดิน และสามารถนำไปใช้ประกอบการตัดสินใจในการวางแผนพัฒนาในอนาคตได้อย่างเหมาะสม [1] ในอดีต การจัดทำแผนที่ LULC มักใช้วิธีการสำรวจภาคพื้นดิน ซึ่งแม้จะให้ข้อมูลที่ละเอียด แต่ก็มีข้อจำกัดด้านเวลา ค่าใช้จ่าย และกำลังคนที่ต้องใช้ในการสำรวจพื้นที่ขนาดใหญ่ [2]

ด้วยความก้าวหน้าของเทคโนโลยีการสำรวจข้อมูลระยะไกล (Remote Sensing - RS) ทำให้การจัดทำแผนที่ LULC มีประสิทธิภาพมากขึ้น ข้อมูลภาพถ่ายดาวเทียมสามารถครอบคลุมพื้นที่กว้างขึ้นที่ข้อมูลได้หลายช่วงเวลา และประหยัดค่าใช้จ่ายกว่าการสำรวจภาคพื้นดิน [3] ดาวเทียมสมัยใหม่ เช่น Sentinel-2 ให้ข้อมูลที่มีความละเอียดเชิงพื้นที่และเชิงสเปกตรัมสูงขึ้น [4], [5] ซึ่งเอื้อต่อการวิเคราะห์และจำแนกประเภท LULC ได้ละเอียดยิ่งขึ้น

อย่างไรก็ตาม การวิเคราะห์ข้อมูลภาพถ่ายดาวเทียมที่มีปริมาณมหาศาลและมีความซับซ้อนสูง ยังคงเป็นความท้าทาย เทคนิคการเรียนรู้ของเครื่อง (Machine Learning - ML) ได้เข้ามามีบทบาทสำคัญในการประมวลผลและจำแนกข้อมูล LULC จากภาพถ่ายดาวเทียมได้อย่างรวดเร็วและแม่นยำ [6] เทคนิค ML เช่น โครงข่ายประสาทเทียม (Artificial Neural Network - ANN), ต้นไม้ตัดสินใจ (Decision Tree - DT), ป่าสุ่ม (Random Forest - RF), และซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine - SVM) ได้รับการนำมาประยุกต์ใช้อย่างแพร่หลายในการจำแนก LULC

ถึงแม้ว่าจะมีงานวิจัยจำนวนมากที่ใช้เทคนิค ML ในการจำแนก LULC แต่ประสิทธิภาพของแต่ละเทคนิคอาจแตกต่างกันไป ขึ้นอยู่กับปัจจัยหลายอย่าง เช่น ลักษณะของข้อมูลดาวเทียมที่ใช้, ความซับซ้อนของพื้นที่ศึกษา, การเลือกใช้พารามิเตอร์ของแบบจำลอง, และคุณภาพของข้อมูลฝึกสอน [7] ดังนั้น การศึกษาเปรียบเทียบประสิทธิภาพของเทคนิค ML ต่าง ๆ ในการจำแนก LULC สำหรับพื้นที่และชุดข้อมูลที่เฉพาะเจาะจง จึงยังคงมีความสำคัญ เพื่อให้ทราบว่าเทคนิคใดมีความเหมาะสมและสามารถสร้างแผนที่ LULC ที่มีความแม่นยำสูงสุด

งานวิจัยนี้มุ่งเน้นการประเมินและเปรียบเทียบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ที่ได้รับความนิยม ได้แก่ โครงข่ายประสาทเทียม (Artificial Neural Network), ต้นไม้ตัดสินใจ (Decision Tree), ป่าสุ่ม (Random Forest) และซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine) สำหรับการจำแนกประเภทสิ่งปกคลุมดินในพื้นที่อำเภอเมืองร้อยเอ็ด จังหวัดร้อยเอ็ด ซึ่งเป็นพื้นที่ที่มีการขยายตัวของชุมชนและการเปลี่ยนแปลงการใช้ประโยชน์ที่ดินอย่างต่อเนื่อง โดยใช้ข้อมูลจากภาพถ่ายดาวเทียม Sentinel-2 ที่มีความละเอียดเชิงพื้นที่เหมาะสมสำหรับการวิเคราะห์ระดับพื้นที่เมืองและกึ่งเมืองกึ่งชนบท ทั้งนี้ วัตถุประสงค์หลักของการวิจัย คือ (1) เพื่อพัฒนาวิธีการจำแนกสิ่งปกคลุมดินจากภาพถ่ายดาวเทียม Sentinel-2 โดยใช้เทคนิค ANN, DT, RF และ SVM และ (2) เพื่อวิเคราะห์และเปรียบเทียบความแม่นยำและความเชื่อถือได้ของผลการจำแนกที่ได้จากแต่ละเทคนิค โดยประเมินผ่านค่าความถูกต้องโดยรวม (Overall Accuracy) และค่าสถิติแคปปา (Kappa Statistic) ผลการศึกษาจะช่วยชี้แนวทางในการเลือกใช้เทคนิคที่เหมาะสมสำหรับการจัดทำข้อมูลการใช้ที่ดินและสิ่งปกคลุมดิน (LULC) ที่มีความถูกต้องและสามารถประยุกต์ใช้ในการวางแผนพัฒนาและจัดการทรัพยากรในระดับท้องถิ่นอย่างมีประสิทธิภาพ

## 1.2 วัตถุประสงค์

1. เพื่อพัฒนาวิธีจำแนกสิ่งปกคลุมดินในพื้นที่ศึกษาจากภาพถ่ายดาวเทียม Sentinel-2 ด้วยเทคนิค โครงข่ายประสาทเทียม (Artificial Neural Network), ต้นไม้ตัดสินใจ (Decision Tree), ป่าสุ่ม (Random Forest) และซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine)
2. เพื่อวิเคราะห์ความแม่นยำและความเชื่อถือได้ของผลการจำแนกสิ่งปกคลุมดินจากเทคนิคต่าง ๆ โดยใช้ค่าความถูกต้อง (Accuracy) และสถิติแคปปา (Kappa Statistics) เป็นตัวชี้วัด

## 1.3 ผลที่คาดว่าจะได้รับ

1. งานวิจัยนี้จะช่วยเติมเต็มช่องว่างทางวิชาการในด้านการใช้เทคโนโลยี Remote Sensing และ Machine Learning ในการจำแนกสิ่งปกคลุมดิน ซึ่งยังขาดการศึกษาอย่างละเอียดในบางบริบทหรือพื้นที่เฉพาะ
2. ทราบถึงผลจากการทดสอบด้วยประสิทธิภาพของตัวจำแนกการเรียนรู้เครื่อง (Machine Learning) จำนวน 4 เทคนิค จากภาพถ่ายดาวเทียม Sentinel-2

#### 1.4 นิยามศัพท์เฉพาะ

1. การจำแนกสิ่งปกคลุมดิน (Land Cover Classification) คือ กระบวนการที่นำเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) มาประยุกต์ใช้ในการแปลความหมายจากข้อมูลภาพถ่ายดาวเทียม เพื่อจำแนกประเภทของสิ่งปกคลุมดินในพื้นที่ศึกษาตามลักษณะการสะท้อนค่าพลาคิ (Spectral Signature) ที่แตกต่างกันในแต่ละประเภท เช่น พื้นที่เกษตรกรรม พื้นที่เมือง แหล่งน้ำ ป่าไม้ เป็นต้น โดยในงานวิจัยนี้ได้เลือกใช้เทคนิค 4 รูปแบบ ได้แก่ โครงข่ายประสาทเทียม (Artificial Neural Network ANN), ต้นไม้ตัดสินใจ (Decision Tree), ป่าสุ่ม (Random Forest) และซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine) เพื่อทำการจำแนกและสร้างผลลัพธ์ในรูปแบบแผนที่การจำแนกสิ่งปกคลุมดิน (Land Cover Classification Map) ซึ่งสามารถนำไปใช้ในการวางแผนและบริหารจัดการการใช้ที่ดินในพื้นที่ได้อย่างมีประสิทธิภาพ สิ่งปกคลุมดินพื้นที่รอยเอ็ด (Land Cover in Roi Et Province) หมายถึง ประเภทของสิ่งปกคลุมดินที่พบในพื้นที่จังหวัดร้อยเอ็ด ประเทศไทย ซึ่งโดยทั่วไปประกอบด้วย พื้นที่เกษตรกรรม (โดยเฉพาะนาข้าว), พื้นที่ชุมชนและสิ่งปลูกสร้าง, แหล่งน้ำ (แม่น้ำ, อ่างเก็บน้ำ), และอาจมีพื้นที่ป่าไม้หรือทุ่งหญ้าอยู่บ้าง การทราบประเภทและการกระจายตัวของสิ่งปกคลุมดินในพื้นที่มีความสำคัญต่อการวางแผนและการจัดการทรัพยากร
2. ภาพถ่ายดาวเทียม (Satellite Imagery) คือ ภาพของพื้นผิวโลกที่บันทึกโดยเซนเซอร์หรือกล้องที่ติดตั้งอยู่บนดาวเทียมที่โคจรรอบโลก ภาพเหล่านี้บันทึกการสะท้อนหรือการแผ่รังสีพลังงานแม่เหล็กไฟฟ้าจากพื้นผิวโลกในช่วงความยาวคลื่นต่างๆ กัน ข้อมูลจากภาพถ่ายดาวเทียมถูกนำมาใช้ประโยชน์หลากหลายด้าน เช่น การทำแผนที่ การติดตามการเปลี่ยนแปลงสิ่งแวดล้อม การเกษตร และการจัดการภัยพิบัติ
3. เทคนิคการเรียนรู้ของเครื่อง (Machine Learning Techniques) เป็นสาขาหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence - AI) ที่มุ่งเน้นการพัฒนาอัลกอริทึมที่ช่วยให้ระบบคอมพิวเตอร์สามารถ "เรียนรู้" จากข้อมูลได้ด้วยตนเอง โดยไม่ต้องมีการเขียนโปรแกรมสั่งงานโดยตรงในทุกขั้นตอน ในบริบทของการจำแนกสิ่งปกคลุมดิน เทคนิคการเรียนรู้ของเครื่อง เช่น Support Vector Machine (SVM), Random Forest, Neural Networks จะถูกนำมาใช้เพื่อฝึกสอนโมเดลให้สามารถจำแนกประเภทสิ่งปกคลุมดินจากข้อมูลภาพถ่ายดาวเทียม (เช่น ข้อมูลจาก Sentinel-2) ได้อย่างอัตโนมัติและแม่นยำ

## บทที่ 2

### เอกสารและงานวิจัยที่เกี่ยวข้อง

จากการทบทวนทฤษฎีและงานวิจัยที่เกี่ยวข้องเพื่อเป็นแนวคิดในการศึกษาวิจัยการจำแนกสิ่งปกคลุมดินจากข้อมูลภาพถ่ายดาวเทียมเซนติเนล -2 ด้วยเทคนิคการเรียนรู้ของเครื่อง มีแนวความคิดและงานวิจัยที่เกี่ยวข้อง เพื่อเป็นแนวทางในการนำมาประยุกต์ใช้ในการศึกษา ประกอบไปด้วย การใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน เทคโนโลยีสำรวจระยะไกล ข้อมูลภาพถ่ายจากดาวเทียม การประมวลผลข้อมูลภาพถ่ายจากดาวเทียม การเรียนรู้ของเครื่อง ข้อมูลอำเภอเมืองร้อยเอ็ด และงานวิจัยที่เกี่ยวข้อง

#### 2.1 ทฤษฎีที่เกี่ยวข้อง

##### 2.1.1 อำเภอเมืองร้อยเอ็ด

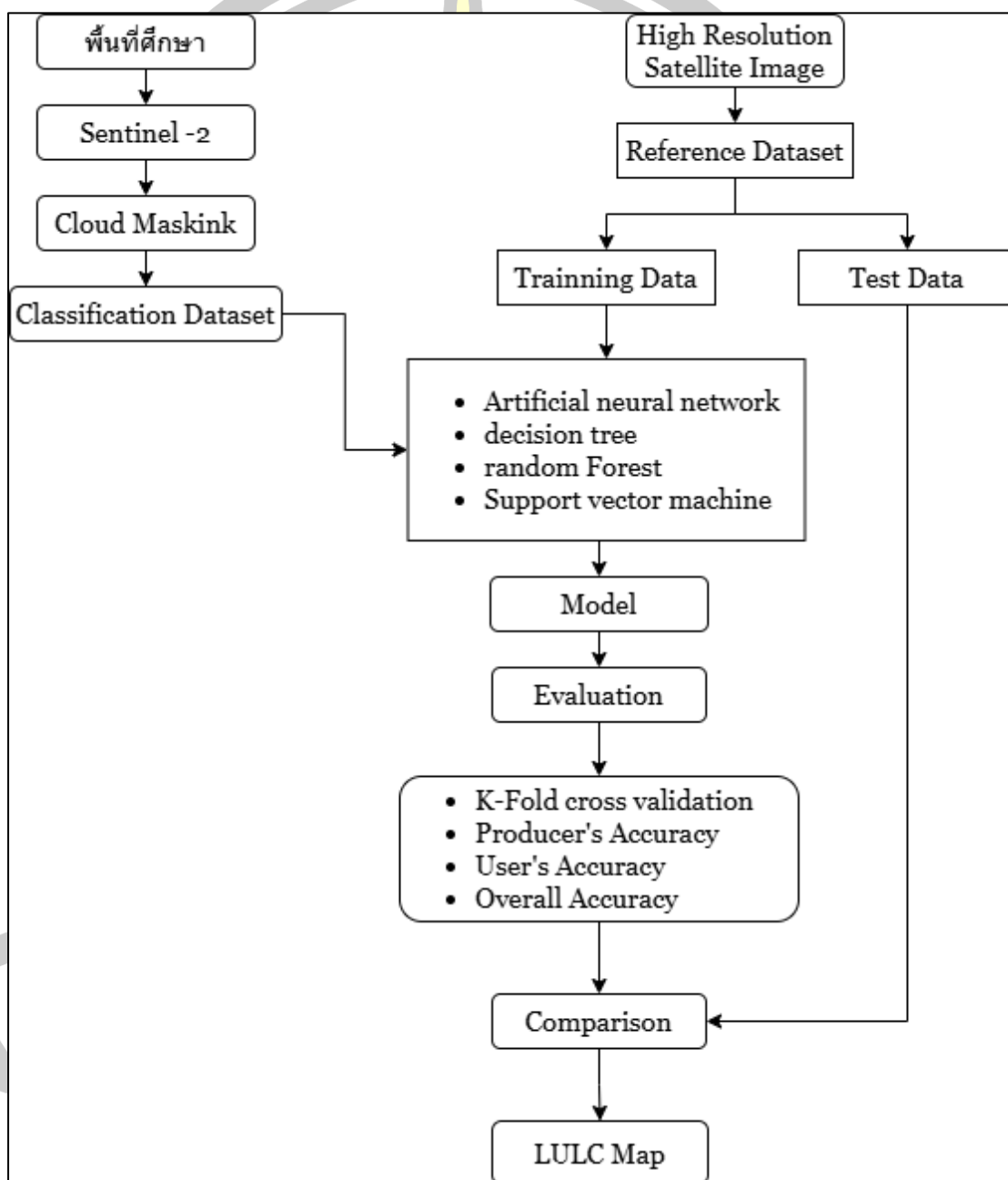
อำเภอเมืองร้อยเอ็ดในอนาคตที่จะมีการขยายตัวขยายตัว เนื่องจากปัจจัยต่าง ๆ ความต้องการที่อยู่อาศัยจะเพิ่มขึ้นเมื่อเมืองร้อยเอ็ดเติบโต ส่งผลให้เกิดการพัฒนาโครงการบ้านจัดสรร คอนโดมิเนียม และพื้นที่ชุมชนใหม่ ๆ โดยเฉพาะในเขตชานเมืองและบริเวณรอบตัวเมืองที่ยังมีพื้นที่ว่าง และมีการเติบโตของประชากรที่มีแนวโน้มเพิ่มขึ้นในอนาคต การย้ายถิ่นฐานจากชนบทสู่เมืองเพื่อเข้าถึงที่อยู่อาศัย การศึกษา ขนส่งสาธารณะ อนาคตอำเภอเมืองร้อยเอ็ดจะมีการสร้างสถานีรถไฟพื้นที่รอบสถานีรถไฟจะเป็นจุดที่มีการพัฒนาอย่างรวดเร็ว เมื่อระบบรถไฟเข้ามา พื้นที่รอบสถานีจะกลายเป็นศูนย์กลางทางการค้าและบริการ เช่น การก่อสร้างศูนย์การค้า โรงแรม และอาคารสำนักงาน ซึ่งจะส่งผลให้เกิดการขยายตัวของเมืองในพื้นที่อย่างรวดเร็ว เทศบาลเมืองร้อยเอ็ดเป็นส่วนหนึ่งของจังหวัดร้อยเอ็ด โดยถือเป็นศูนย์กลางด้านการปกครองและเศรษฐกิจของจังหวัด พื้นที่ทั้งหมด 493.6 ตร.กม ประชากรทั้งหมด 157,085 คน (ข้อมูล ณ ปี พ.ศ. 2567) ความหนาแน่น 318.24 คน/ตร.กม

ที่ตั้งและอาณาเขตอำเภอเมืองร้อยเอ็ดตั้งอยู่ทางทิศตะวันตกของจังหวัด มีอาณาเขตติดต่อกับเขตการปกครองข้างเคียงดังต่อไปนี้

- |             |  |
|-------------|--|
| ทิศเหนือ    | ติดกับ อำเภอเมืองมหาสารคาม (จังหวัดมหาสารคาม) และอำเภอจังหาร |
| ทิศตะวันออก | ติดกับ อำเภอธวัชบุรี   |
| ทิศใต้      | ติดกับ อำเภออาจสามารถ และอำเภอจตุรพักตรพิมาน                 |
| ทิศตะวันตก  | ติดกับ อำเภอศรีสมเด็จ  |

### 2.1.2 ขั้นตอนการประมวลผลข้อมูล

กระบวนการวิจัยทั้งหมด ตั้งแต่การกำหนดพื้นที่ศึกษา การเตรียมข้อมูล การสร้างแบบจำลอง จนถึงการประเมินผลและการจัดทำแผนที่ แสดงให้เห็นเป็นแผนภาพขั้นตอนการดำเนินงานดังภาพประกอบที่ 2.1



ภาพประกอบที่ 2.1 ขั้นตอนการประมวลผลข้อมูล

### 2.1.3 เทคโนโลยีการสำรวจข้อมูลระยะไกล (Remote sensing)

การรับรู้จากระยะไกล เป็นวิทยาศาสตร์และศิลปะของการได้มาซึ่งข้อมูลเกี่ยวกับวัตถุพื้นผิว และปรากฏการณ์บนพื้นโลก จากเครื่องรับรู้ (Sensor) โดยปราศจากการเข้าไปสัมผัสวัตถุเป้าหมาย ทั้งนี้อาศัยพลังงานของคลื่นแม่เหล็กไฟฟ้า (Electromagnetic energy) เป็นสื่อในการได้มาของข้อมูลซึ่งมีคุณสมบัติ 3 ประการ คือ ลักษณะการสะท้อนช่วงคลื่นแม่เหล็กไฟฟ้า (Spectral characteristics) ลักษณะเชิงพื้นที่ของวัตถุบนพื้นผิวโลก (Spatial characteristics) และลักษณะการเปลี่ยนแปลงของวัตถุตามช่วงเวลา (Temporal characteristics)

กระบวนการสำรวจข้อมูลระยะไกลแบ่งออกเป็น 2 กระบวนการใหญ่ คือ การบันทึกข้อมูล (Data Acquisition) และการวิเคราะห์ประมวลผลข้อมูล (Data Analysis) การบันทึกข้อมูลเป็น กระบวนการบันทึก พลังงานที่สะท้อนหรือส่งผ่านของสิ่งปกคลุมดินโดยเครื่องมือบันทึกบนยานสำรวจ แล้วส่งข้อมูลทั้งในรูปแบบภาพถ่ายไปยังภาคพื้นดิน เพื่อผ่านกรรมวิธีการผลิตเป็นข้อมูลทั้งในรูปแบบ ภาพถ่าย (Photographic) และข้อมูลเชิงตัวเลข (Digital Data) การวิเคราะห์ประมวลผลข้อมูล ประกอบด้วยการ วิเคราะห์ด้วยสายตา และวิเคราะห์ด้วยเครื่องคอมพิวเตอร์ ก่อนที่ผู้ใช้จะนำมาใช้ประโยชน์โดยเครื่องบันทึกข้อมูลที่ใช้พลังงานคลื่นแม่เหล็กไฟฟ้าตามธรรมชาติ Passive remote sensing เป็นระบบที่ใช้กันกว้างขวางตั้งแต่เริ่มแรกจนถึงปัจจุบัน โดยมีแหล่งพลังงานที่เกิดตามธรรมชาติ คือ ดวงอาทิตย์เป็นแหล่งกำเนิดพลังงาน ระบบนี้จะรับและบันทึก ข้อมูลได้ส่วนใหญ่ในเวลากลางวัน และมีข้อจำกัดด้านภาวะอากาศ ไม่สามารถรับข้อมูลได้ในฤดูฝน หรือเมื่อมีเมฆ หมอกฝน Active remote sensing เป็นระบบที่แหล่งพลังงานเกิดจากการสร้างขึ้นในตัวของเครื่องมือสำรวจ เช่น ช่วงคลื่นไมโครเวฟที่สร้างในระบบเรดาร์ แล้วส่งพลังงานนั้นไปยังพื้นที่เป้าหมาย ระบบนี้สามารถทำการรับและบันทึกข้อมูล ได้โดยไม่มีข้อจำกัดด้านเวลา หรือ ด้านสภาวะภูมิอากาศ คือสามารถรับส่งสัญญาณได้ทั้งกลางวันและกลางคืน อีกทั้งยังสามารถทะลุผ่านกลุ่มเมฆ หมอก ฝนได้ในทุกฤดูกาล.[8]

#### 2.1.3.1 ดาวเทียม Sentinel-2A

ดาวเทียม Sentinel-2 ถูกพัฒนาขึ้นภายใต้โครงการ Global Monitoring for Environment and Security (GMES) ซึ่งเป็นความร่วมมือระหว่าง คณะกรรมาธิการยุโรป (European Commission) และ องค์การอวกาศยุโรป (European Space Agency) โดยมีวัตถุประสงค์เพื่อสร้าง ชีตความสามารถของสหภาพยุโรปในการจัดหาและใช้ประโยชน์สารสนเทศที่เกี่ยวข้องกับสิ่งแวดล้อม และความมั่นคง ดาวเทียม Sentinel-2 เป็นดาวเทียมวงโคจรกว้าง (Wide-

swath) ถูกสร้างขึ้นเพื่อ บันทึกภาพพื้นผิวโลกต่อเนื่องจากดาวเทียม Landsat และ SPOT ดาวเทียม Sentinel-2 ถูกส่งขึ้นสู่วงโคจรเมื่อปี พ.ศ. 2556 ประกอบด้วย ดาวเทียม S2A และ S2B ปฏิบัติงานภายในวงโคจรเดียวกัน แต่ตำแหน่งของดาวเทียมต่างกัน 180 องศา ที่ระดับความสูง 786 กิโลเมตร ความกว้างแนว บันทึกภาพ 290 กิโลเมตร รอบเวลาในการโคจรกลับมาบันทึกภาพที่เดิม ณ ตำแหน่งเส้นศูนย์สูตร ทุก ๆ 5 วัน Sentinel ถูกพัฒนาขึ้นเพื่อรองรับ application ต่าง ๆ เช่น การเกษตร การสำรวจทรัพยากร การจัดการภัยพิบัติ การจัดการทรัพยากรน้ำ และอื่น ๆ [9] คุณลักษณะจำเพาะของช่วงคลื่นต่างๆ ของดาวเทียม Sentinel-2 ที่ใช้ในการศึกษานี้ แสดงไว้ดังตารางที่ 2.1

**ตารางที่ 2.1** ข้อมูลดาวเทียม SENTINEL- 2

Description	Central Wavelength	Resolution
1. Coastal aerosol	443 nm	60 m
2. Blue	490 nm	10 m
3. Green	560 nm	
4. Red	665 nm	
5. Vegetation red edge	705 nm	20 m
6. Vegetation red edge	740 nm	
7. Vegetation red edge	783 nm	
8. NIR	842 nm	10 m
9. Narrow NIR	865 nm	20 m
10. Water vapour	940 nm	60 m
11. Short Wave Infrared (SWIR)	1375 nm	20 m
12. Short Wave Infrared (SWIR)	1610 nm	
13. Short Wave Infrared (SWIR)	2190 nm	

( ที่มา: ESA, <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial>)

### 2.1.3.2 การใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน

การจำแนกประเภทการใช้ประโยชน์ที่ดินในการศึกษานี้อ้างอิงตามระบบของกรมพัฒนาที่ดิน พ.ศ. 2555 [10] ซึ่งแบ่งการใช้ที่ดินออกเป็นระดับต่างๆ ดังรายละเอียดการจำแนกในระดับที่ 1 และระดับที่ 2 ที่แสดงดังตารางที่ 2.2

**ตารางที่ 2.2** การจำแนกการใช้ประโยชน์ที่ดินของกรมพัฒนาที่ดิน ปี พ.ศ.2555 ระดับที่ 1 และ 2

Level I	Code	Level II	Code
1. พื้นที่ชุมชนและสิ่งปลูกสร้าง (Urban and built-up land)	U	เมืองและย่านการค้า (Urban and Commercial area)	U1
		ที่อยู่อาศัย (Residential area)	U2
		สถานที่ราชการและสถาบัน (Governmental and Institutional land)	U3
		สถานีคมนาคม การสื่อสาร และสาธารณูปโภค (Transportation, Communications and Utilities)	U4
		ย่านอุตสาหกรรม (Industrial land)	U5
		สิ่งปลูกสร้างอื่น ๆ (Other Built-up land)	U6
		สนามกอล์ฟ (Golf Course)	U7
2. พื้นที่เกษตรกรรม (Agricultural Land)	A	พื้นที่นา (Paddy field)	A1
		พืชไร่ (Field crop)	A2
		ไม้ยืนต้น (Perennial)	A3
		ไม้ผล (Orchard)	A4
		พืชสวน (Horticulture)	A5
		ไร่หมุนเวียน (Swidden cultivation)	A6

ตารางที่ 2.2 การจำแนกการใช้ประโยชน์ที่ดินของกรมพัฒนาที่ดิน ปี พ.ศ.2555 ระดับที่ 1 และ 2 (ต่อ)

Level I	Code	Level II	Code
2. พื้นที่เกษตรกรรม (Agricultural Land)	A	ทุ่งหญ้าเลี้ยงสัตว์และโรงเรือน (Pasture and Farm house)	A7
		พืชน้ำ (Aquatic plant)	A8
		สถานที่เพาะเลี้ยงสัตว์น้ำ (Aquacultural land)	A9
		เกษตรผสมผสาน (Integrated farm)	A0
3. พื้นที่ป่าไม้ (Forest Land)	F	ป่าไม่ผลัดใบ (Evergreen forest)	F1
		ป่าผลัดใบ (Deciduous forest)	F2
		ป่าชายเลน (Mangrove forest)	F3
		ป่าพรุ (Swamp forest)	F4
		ป่าปลูก (Forest plantation)	F5
		วนเกษตร (Agro – forestry)	F6
		ป่าชายหาด (Beach forest)	F7
4. พื้นที่แหล่งน้ำ (Water body)	W	พื้นที่แหล่งน้ำ (Natural water body)	W1
		แหล่งน้ำที่มนุษย์สร้างขึ้น (Artificial water body)	W2
5. พื้นที่เบ็ดเตล็ด (Miscellaneous land)	M	ทุ่งหญ้าและไม้ละเมาะ (Rangeland and Scrub)	M1
		พื้นที่ลุ่มชื้นแฉะและพื้นที่น้ำขัง (Marsh and Swamp)	M2
		เหมืองและบ่อขุด (Mine and pit)	M3
		พื้นที่เบ็ดเตล็ดอื่น ๆ (Other Miscellaneous land)	M4
		นาเกลือ (Salt flat)	M5
		หาดทราย (Beach)	M6
		ที่ทิ้งขยะ (Garbage dump)	M7

## 2.1.4 ขั้นตอนการจำแนกสิ่งปกคลุมดิน

### 2.1.4.1 กระบวนการก่อนการประมวลผลภาพ (pre-processing)

กระบวนการก่อนการประมวลผลภาพ (Pre-processing) การปรับแก้ภาพมีวัตถุประสงค์เพื่อปรับแก้ความคลาดเคลื่อนของข้อมูล (Data error) สัญญาณที่ รบกวน (Noise) และความบิดเบี้ยวเชิงเรขาคณิตที่เกิดขึ้นในระหว่างกระบวนการถ่ายภาพ การบันทึกข้อมูล สัญญาณ การสะท้อนคลื่นแม่เหล็กไฟฟ้า การส่งสัญญาณ และการโคจรของดาวเทียม มีกระบวนการปรับแก้ 2 กระบวนการที่ต้องกระทำ ดังนี้

### 2.1.4.2 การตรวจแก้คลื่นรังสี (Radiometric correction)

ก่อนที่จะส่งข้อมูลจากระยะไกลให้กับผู้ใช้งาน ข้อมูลเหล่านี้จะต้องผ่านการตรวจแก้คลื่นรังสีจากสถานีรับสัญญาณดาวเทียมภาคพื้นดินมาแล้วระดับหนึ่ง แต่บางครั้งยังปรากฏข้อบกพร่องทางเชิงคลื่น เนื่องจากสาเหตุหลายประการ เช่น การรบกวนจากชั้นบรรยากาศ หรือจากความบกพร่องของเครื่องรับสัญญาณ ทำให้เกิดความไม่ชัดเจนพรมัว มีลายเส้นปะปน (Strip/Noise) ปรากฏบนข้อมูลจากดาวเทียม ปัญหาอีกด้านหนึ่งที่ต้องมีการตรวจแก้คลื่นรังสี คือ เมื่อต้องการใช้ข้อมูลหลายช่วงเวลาเพื่อการศึกษาการเปลี่ยนแปลงของปรากฏการณ์ใด ปรากฏการณ์หนึ่งต้องทำการปรับแก้ค่ามุมยกของดวงอาทิตย์ (Sun elevation correction) ที่เปลี่ยนแปลงไปตามแต่ละช่วงเวลาและแต่ละฤดูกาล การแก้ไขปัญหาดังกล่าวทำได้โดยการตรวจแก้คลื่นรังสี พร้อมทั้งต้องมีรายละเอียดตัวแปร (Parameters) เกี่ยวกับการรับสัญญาณมุมที่ดาวเทียม กระทำกับแสงดวงอาทิตย์ (Solar illumination angles) ค่ารังสีตกกระทบ (Irradiance) การกระจายแสงในเส้นทางผ่าน (Path radiance) ค่าการสะท้อนของวัตถุเป้าหมาย (Reflectance of target) ค่าการส่งผ่านของบรรยากาศ เป็นต้น และข้อมูลสภาวะอากาศในขณะที่ทำการบันทึกข้อมูล การปรับแก้มีกรรมวิธีในการคำนวณที่ซับซ้อนมาก โดยต้องใช้ซอฟต์แวร์ที่มีโปรแกรมเฉพาะสำหรับการตรวจแก้คลื่นรังสี ซึ่งโดยทั่วไปการแก้ไขข้อบกพร่องเชิงคลื่นจะต้องทำ ได้แก่

- 1) การชดเชยค่าการสะท้อนที่บิดเบือนของสภาวะอากาศ (Haze compensation) เกิดขึ้นจากการกระจัดกระจายแสงในบรรยากาศ จึงทำให้เกิดการสลัวของแสง (Haze) ทำให้ภาพมีลักษณะไม่ชัดเจน ไม่คมชัดการแก้ไขทำได้โดยการลดผลการกระจัดกระจายแสงในบรรยากาศให้น้อยที่สุด โดยการเปรียบเทียบค่าความสว่างทั่วไปกับค่าความสว่างต่ำที่สุด

(Zero reflectance) ซึ่งส่วนใหญ่จะเป็นวัตถุที่ดูดกลืนพลังงานมาก เช่น น้ำใสจะมีการดูดกลืนสูงมากในช่วงคลื่นอินฟราเรด

2) การเปลี่ยนค่าความสว่างเป็นค่าการแผ่รังสีสมบูรณ์ (Conversion of digital numbers to absolute radiance value) เป็นการตรวจแก้คลื่นรังสีอีกวิธีหนึ่ง โดยการแปลงค่าความสว่างเป็นค่าการแผ่รังสี คำนวณจากข้อมูลค่าการแผ่รังสีสูงสุด และการแผ่รังสีต่ำสุดในแต่ละช่วงคลื่น

#### 2.1.4.3 การตรวจแก้เชิงเรขาคณิต (Geometric correction)

ก่อนนำข้อมูลจากดาวเทียมไปใช้ประโยชน์มีความจำเป็นต้องปรับแก้เชิงเรขาคณิต เพราะพิกัดตำแหน่งของวัตถุต่าง ๆ มีความคลาดเคลื่อนไปจากความเป็นจริง เนื่องจากข้อบกพร่องทางเครื่องรับสัญญาณและรูปลักษณ์ของวัตถุ การตรวจแก้เชิงเรขาคณิตมีความจำเป็นมากยิ่งขึ้น เมื่อต้องการนำข้อมูลจากระยะไกลไปใช้งานร่วมกับข้อมูลเชิงแผนที่อื่น ๆ เพื่อให้สามารถซ้อนทับกันได้ หรือหากต้องการศึกษาปรากฏการณ์อย่างหนึ่งในหลายช่วงเวลา การเปรียบเทียบข้อมูลแต่ละช่วงเวลาต้องมีระบบพิกัดเดียวกัน จึงจะสามารถซ้อนข้อมูลแต่ละเวลาลงกันได้สนิท จึงจำเป็นต้องทำการปรับแก้เชิงเรขาคณิต

#### 2.1.4.4 การเน้นข้อมูลภาพ (Image enhancement)

กระบวนการปรับเปลี่ยนค่าของจุดภาพหรือค่าระดับสีเทา เพื่อเพิ่มรายละเอียด ความชัดเจนของ ข้อมูลภาพ หรือการเพิ่มระดับความแตกต่างระหว่างวัตถุ ทำให้สังเกตเห็นขอบเขตของวัตถุต่างชนิดกันได้ชัดเจนมากขึ้น หรือสามารถที่จะเน้นความคมชัดเฉพาะในส่วนที่ต้องการศึกษาช่วยให้การตีความประเภทวัตถุง่ายขึ้น แล้วนำผลลัพธ์ที่ได้ไปทำการแปลตีความ โดยทำการศึกษาแปลตีความด้วยสายตา (Visual interpretation) เพื่อกำหนดประเภทข้อมูลก่อนที่นำไปใช้เพื่อการจำแนกประเภทข้อมูลเทคนิคต่าง ๆ ที่จะใช้ในการเน้นข้อมูลภาพนั้นอยู่กับ - ค่าจุดภาพ ซึ่งประกอบด้วยข้อมูลจากหลายช่วงคลื่นของข้อมูลจากดาวเทียมที่ออกแบบมาเฉพาะ ผู้ใช้จะต้องทำความเข้าใจรูปแบบของปฏิสัมพันธ์ระหว่างพลังงานในแต่ละช่วงคลื่นกับวัตถุบนพื้นโลก - วัตถุประสงค์ของการเน้นข้อมูล คือ ต้องเปลี่ยนแปลงข้อมูลเพื่อให้เห็นรายละเอียดต่าง ๆ ในเรื่องที่ต้องการศึกษาได้ดียิ่งขึ้น - ผลที่คาดว่าจะได้รับการเน้นข้อมูลภาพ - พื้นฐานของผู้วิเคราะห์ซึ่งจะต้องมีประสบการณ์ในการวิเคราะห์และเทคนิคการเน้นข้อมูลระดับ ความแตกต่างระหว่างวัตถุในข้อมูลภาพ ซึ่งวัดได้จากสัดส่วนของระดับความแตกต่างของความชัดเจน (Contrast ratio) โดยนำค่าความสว่างสูงสุดมาหารด้วยค่าการสะท้อนต่ำที่สุด

#### 2.1.4.5 การเน้นข้อมูลแบบเชิงพื้นที่ (Spatial enhancement)

การเน้นข้อมูลโดยพิจารณาค่าของจุดภาพรอบข้างแตกต่างกับการเน้นข้อมูลช่วงคลื่นซึ่งพิจารณาค่าของจุดภาพเดี่ยว ๆ เท่านั้น โดยทั่วไปการเน้นข้อมูลแบบนี้จะพิจารณาจากความถี่ของกลุ่มข้อมูลข้างเคียง หรือความแตกต่างของค่าจุดภาพสูงสุดและต่ำสุดของกลุ่มข้อมูล ภาพรอบข้าง หมายถึงการเปลี่ยนแปลงในค่าของจุดภาพต่อหนึ่งหน่วยระยะทางสำหรับบางส่วนของภาพที่กำหนด กระบวนการเฉลี่ยค่าของกลุ่มขนาดเล็กของจุดภาพจำนวนหนึ่งผ่านเข้าไปในภาพหนึ่ง เพื่อเปลี่ยนแปลงความถี่ของ กลุ่มจุดภาพนั้น ๆ เรียกว่า การกรองแบบวนรอบ (Convolution filtering) กระบวนการนี้ใช้ส่วนกลางแบบวนรอบ (Convolution kernel) ซึ่งเป็นเมทริกซ์ของตัวเลขที่ใช้ในการเปลี่ยนค่าของแต่ละจุดภาพด้วยค่าของจุดภาพรอบข้าง ในรูปแบบเฉพาะจำนวนของตัวเลขในเมทริกซ์จะเป็นตัวกำหนดน้ำหนักในการปรับเปลี่ยนค่าของจุดภาพเฉพาะที่ค่าตัวเลขนี้มักจะเรียกว่า ค่าสัมประสิทธิ์ เพราะใช้รูปแบบและวิธีทางสมการคณิตศาสตร์ ตัวอย่างการเน้นข้อมูลแบบเชิงพื้นที่

#### 2.1.4.6 กรองภาพ (Image filtering)

การปรับเปลี่ยนค่าของจุดภาพที่มีจำนวนน้อยเข้ากับจุดภาพที่มีจำนวนมากหรือกำจัดจุดภาพที่มีส่วนน้อยออกจากภาพ เพื่อความเข้าใจเกี่ยวกับการปรับเปลี่ยนค่าของจุดภาพที่ไม่ต้องการแบบวนรอบจะใช้เมทริกซ์ซ้อนทับบนแฟ้มข้อมูลจุดภาพ จุดภาพที่ถูกเปลี่ยนค่าจะอยู่บนหน้าต่าง ตัวกรองภาพที่ใช้กับข้อมูลจากดาวเทียม เป็นกรอบตารางกริดรูปสี่เหลี่ยมจัตุรัส (Matrix) โดยมีจำนวนจุดภาพทั้งในแนวดิ่งและ แนวนอนเป็นเลขคี่เสมอ เช่น 3x3 5x5 และ 7x7 เป็นต้น ทั้งนี้เพื่อให้เกิดความสมดุล (Symmetry) กับจุดภาพที่อยู่ตรงกลางกรอบของตัวกรอง การแปลงค่าใหม่ของจุดภาพโดยวิธีการกรองภาพ ทำได้โดยการเคลื่อนกรอบตัวกรองไปตาม ข้อมูลภาพจากซ้ายไปขวา ทีละแถวจนหมดทั้งภาพ ในระหว่างที่ตัวกรองเคลื่อนที่ไปจะทำการคำนวณค่าใหม่ให้แก่จุด ภาพที่อยู่ตรงกลางกรอบของตัวกรอง จากลักษณะการเคลื่อนที่ดังกล่าว ตัวกรองภาพจึงมีชื่อเรียกว่า ส่วนกลาง (Kernel) การใช้ตัวกรองภาพสามารถประยุกต์ใช้ได้โดยตรงกับข้อมูลเชิงพื้นที่ที่อยู่ในรูปข้อมูลเชิงเลข เช่น ข้อมูล จากดาวเทียม และเรียกว่า วิธีการแบบวนรอบ ซึ่งหมายถึง กระบวนการทางคณิตศาสตร์ที่คำนวณค่าข้อมูลเชิงเลข หรือ ค่าการสะท้อนเชิงคลื่น (Radiometric) ของแต่ละจุดภาพกึ่งกลางใหม่ โดยใช้ค่าจากทุกจุดภาพที่ตกอยู่ในกรอบของตัวกรองในขณะที่ทำการกวาดผ่านข้อมูลภาพมาเป็นค่าพื้นฐานเริ่มต้น แล้วแปลงค่าเหล่านั้นด้วยวิธีการคณิตศาสตร์ หลายรูปแบบ เช่น การคูณ การหาร การหาค่าเฉลี่ยจากค่ามัชฌิมาเลขคณิต (Mean) มัชฌิมาน (Median) ฐานนิยม (Mode) หรือการหาค่า

ความเบี่ยงเบนในกรอบตัวกรอง (Local standard deviation) ฯลฯ และบางครั้งอาจมีการให้ค่าน้ำหนัก (Weighting) ร่วมกันด้วย [10]

#### 2.1.4.7 การตรวจสอบภาคสนาม (Ground Check)

การตรวจสอบแก้ไขผล การแปลภาพให้สมบูรณ์ยิ่งขึ้น ในการสำรวจข้อมูลภาคพื้นดินควรคำนึงถึง ช่วงเวลาในการสำรวจ ควรเป็นช่วงเวลาที่ใกล้เคียงกับการบันทึกข้อมูลภาพถ่ายดาวเทียม ดังนั้นการวางแผนก่อนการสำรวจจึงจำเป็นและขึ้นอยู่กับเทคนิค การสุ่มตัวอย่างที่เหมาะสมด้วย การกำหนดจุดตัวอย่าง (Sample Point) ควรให้กระจายทั่วพื้นที่ศึกษาเพื่อจะได้เป็นตัวแทนของทุกกลุ่มประเภทข้อมูล ซึ่งการสุ่มตัวอย่างอาจทำได้ทั้งวิธีการสุ่มตัวอย่างแบบจุด หรือวิธีการสุ่มตัวอย่างแบบพื้นที่หลังจากนั้นข้อมูลที่ได้จากการสำรวจกลุ่ม ตัวอย่างจะนำมาวิเคราะห์เปรียบเทียบผลโดยอาศัยหลักความน่าจะเป็นกับอัตราส่วนความผิดพลาดที่ยอมรับได้ [11] การประเมินความถูกต้องของการจำแนก จะทำการเปรียบเทียบผลลัพธ์กับข้อมูลอ้างอิงภาคพื้นดิน และสรุปผลในรูปแบบตารางประเมินความถูกต้อง (Confusion Matrix) ซึ่งมีโครงสร้างดังแสดงดังภาพประกอบที่ 2.2

		ข้อมูลทดสอบอ้างอิงทางภาคพื้นดิน (Ground reference test information)				
ผลการจำแนกจาก การรับรู้จาก ระยะไกล (Remote sensing classification)	ประเภท (Class)	1	2	3	K	ผลรวมของแถว (Row total)
	1	$x_{1,1}$	$x_{1,2}$	$x_{1,3}$	$x_{1,k}$	$x_{1+}$
	2	$x_{2,1}$	$x_{2,2}$	$x_{2,3}$	$x_{2,k}$	$x_{2+}$
	3	$x_{3,1}$	$x_{3,2}$	$x_{3,3}$	$x_{3,k}$	$x_{3+}$
	K	$x_{k,1}$	$x_{k,2}$	$x_{k,3}$	$x_{k,k}$	$x_{k+}$
	ผลรวมของคอลัมน์ (Column total)	$x_{+1}$	$x_{+2}$	$x_{+3}$	$x_{+k}$	N

ภาพประกอบที่ 2.2 ตารางประเมินความถูกต้อง  
(ที่มา: <https://files.oaiusercontent.com>)

#### 2.1.5 Machine Learning

การทำให้คอมพิวเตอร์ สามารถเรียนรู้สิ่งต่าง ๆ และพัฒนาการทำงานให้ดีขึ้นได้ด้วยตัวเองจากข้อมูลและสภาพแวดล้อมที่ได้รับจากการเรียนรู้ของระบบ โดยไม่ต้องมีมนุษย์คอยกำกับหรือเขียนโปรแกรมเพิ่มเติม และไม่ว่าในอนาคตมันจะมีข้อมูลรูปแบบใหม่ ๆ ที่เกิดขึ้นมามนุษย์ก็ไม่

จำเป็นที่จะต้องไปนั่งเขียนโปรแกรมใหม่ เพราะคอมพิวเตอร์สามารถตีความและตอบสนองได้ด้วยตัวเอง แน่แน่นอนว่าธุรกิจหรืออุตสาหกรรมไหนนำเทคโนโลยีนี้ไปปรับใช้ได้อย่างถูกต้องวิธี จะทำให้ได้เปรียบในเชิงการแข่งขันของธุรกิจอย่างมาก เพราะสามารถลดเวลาการทำงานในการวิเคราะห์ข้อมูลต่าง ๆ และลดต้นทุนแรงงานที่เกิดขึ้นได้มากเลยทีเดียว

### 2.1.5.1 หลักการทำงานของ Machine Learning

การเรียนรู้ของ Machine Learning นั้นจะมีหลักการคล้าย ๆ กับของมนุษย์ที่จำเป็นต้องเรียนรู้จากประสบการณ์ โดยให้นักภาพถึงการสอนเด็กคนหนึ่งให้แยกความแตกต่างระหว่าง ดินสอและปากกา เราจำเป็นต้องสอนเด็กคนนั้นก่อน ว่าดินสอเป็นอย่างไร และปากกาเป็นอย่างไร เพื่อให้เด็กเรียนรู้และแยกแยะความแตกต่างระหว่างของสองสิ่งได้ Machine Learning เองก็ทำงานในลักษณะเดียวกัน ด้วยการป้อนชุดข้อมูลพื้นฐานและชุดคำสั่งต่าง ๆ เพื่อให้คอมพิวเตอร์ “เรียนรู้” และจำแนกแยกแยะวัตถุต่าง ๆ รวมถึงบุคคล สิ่งของ เป็นต้น และเพื่อให้ได้ผลลัพธ์ที่แม่นยำมากขึ้น โปรแกรมเมอร์จึงมีการป้อนชุดข้อมูลใหม่ ๆ และต้องฝึกฝนระบบอย่างสม่ำเสมอ เพื่อให้เกิดวิวัฒนาการของระบบ ทำให้ระบบคิดได้ด้วยตัวเองในเวลาต่อมา การทำงานของระบบ Machine Learning มี 3 รูปแบบ โดยหลักการของ Machine Learning จะแบ่งออกเป็น 3 ประเภทตามรูปแบบการเรียนรู้ด้วยกัน นั่นก็คือ

1. Supervised Learning หรือการเรียนรู้แบบมีผู้สอนเป็นการทำให้คอมพิวเตอร์สามารถหาคำตอบของปัญหาได้ด้วยตัวเอง หลังจากเรียนรู้จากชุดข้อมูลตัวอย่างไปแล้ว ระยะเวลาหนึ่ง ยกตัวอย่างเวลาเราป้อนข้อมูลให้กับคอมพิวเตอร์ (Input) เช่น รูปปากกา เบื้องต้นคอมพิวเตอร์จะยังไม่รู้ว่ารูปที่เราป้อนเข้าไป คือ รูปปากกา เราจึงต้องสอนให้คอมพิวเตอร์รู้จักเพื่อนำไปวิเคราะห์ (Feature Extraction) ว่า ปากกาจะมีปลายด้ามเป็นปุ่ม และใช้หมึกในการเขียน เป็นต้น จากนั้นคอมพิวเตอร์ก็นำข้อมูลดังกล่าวไปประมวล/จัดหมวดหมู่ (Classification) เพื่อให้หลังจากนี้มันสามารถแยกออกได้ว่าอะไรคือปากกาอะไรไม่ใช่ปากกา

2. วิเคราะห์ข้อมูลการใช้ประโยชน์ที่ดินทดสอบด้วยประสิทธิภาพของตัวจำแนกการเรียนรู้เครื่อง (Machine learning) จำนวน 4 วิธีการ ได้แก่

- 2.1 โครงข่ายประสาทเทียม (Artificial Neural Networks หรือ ANNs) เป็นแบบจำลองทางคณิตศาสตร์ที่สร้างขึ้นเพื่อเลียนแบบการทำงานของโครงข่ายประสาทในสมองของมนุษย์ โดยมีจุดประสงค์หลักคือการทำให้คอมพิวเตอร์สามารถเรียนรู้และแก้ไขปัญหาที่ซับซ้อนได้ โดยไม่ต้องมีการเขียนโปรแกรมคำสั่งอย่างชัดเจน หลักการทำงานพื้นฐานโครงข่ายประสาทเทียม

ประกอบด้วยหน่วยประมวลผลขนาดเล็กจำนวนมากที่เรียกว่า "นิวรอน" (neurons) ซึ่งเชื่อมต่อกันเป็นชั้น ๆ แต่ละนิวรอนจะรับข้อมูลเข้า ประมวลผล และส่งผลลัพธ์ไปยังนิวรอนในชั้นถัดไป การประมวลผลข้อมูลในแต่ละนิวรอนจะใช้ฟังก์ชันทางคณิตศาสตร์ที่เรียกว่า "ฟังก์ชันกระตุ้น" (activation function) ซึ่งทำหน้าที่กำหนดว่านิวรอนจะส่งสัญญาณออกไปหรือไม่

โครงข่ายประสาทเทียมจะเรียนรู้โดยการปรับค่าน้ำหนัก (weights) ของการเชื่อมต่อระหว่างนิวรอน ซึ่งเป็นการปรับปรุงความแม่นยำในการทำนายหรือจำแนกข้อมูล โดยกระบวนการเรียนรู้จะเกิดขึ้นจากการป้อนข้อมูลจำนวนมากให้กับโครงข่าย และปรับค่าน้ำหนักให้เหมาะสมกับข้อมูลเหล่านั้น

องค์ประกอบหลักชั้นนำเข้า (Input layer) รับข้อมูลเข้าสู่โครงข่าย เช่น ภาพ เสียง หรือข้อความ ชั้นซ่อน (Hidden layer) เป็นชั้นที่อยู่ระหว่างชั้นนำเข้าและชั้นส่งออก ทำหน้าที่ประมวลผลข้อมูลและดึงลักษณะเด่นของข้อมูลออกมา ชั้นส่งออก (Output layer) ส่งผลลัพธ์ของการประมวลผล เช่น การจำแนกประเภทของภาพ หรือการแปลภาษา

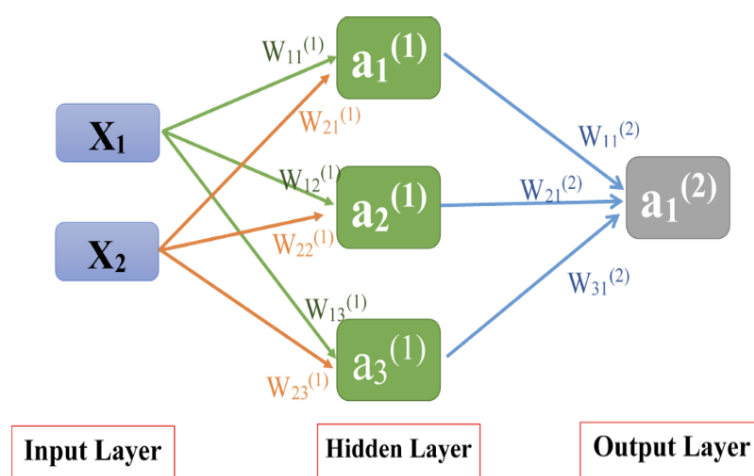
ประเภทของโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมแบบป้อนไปข้างหน้า (Feedforward neural networks) ข้อมูลจะไหลจากชั้นนำเข้าไปยังชั้นส่งออกโดยไม่มีการไหลย้อนกลับ เหมาะสำหรับงานจำแนกประเภทและการถดถอย

โครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent neural networks) มีการเชื่อมต่อระหว่างนิวรอนที่ทำให้ข้อมูลสามารถไหลวนกลับได้ เหมาะสำหรับข้อมูลที่มีลำดับเวลา เช่น การรู้จำเสียงและการแปลภาษา

โครงข่ายประสาทเทียมแบบ convolutional (Convolutional Neural Networks หรือ CNNs) เหมาะสำหรับข้อมูลภาพและวิดีโอ โดยใช้เทคนิคการ convolutional เพื่อดึงลักษณะเด่นของภาพ

โครงสร้างพื้นฐานของโครงข่ายประสาทเทียม (ANN) ประกอบด้วยชั้นนำเข้า (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้นส่งออก (Output Layer) ซึ่งมีการเชื่อมต่อและประมวลผลข้อมูลตั้งแต่ภาพแบบจำลองถึงภาพประกอบที่ 2.3



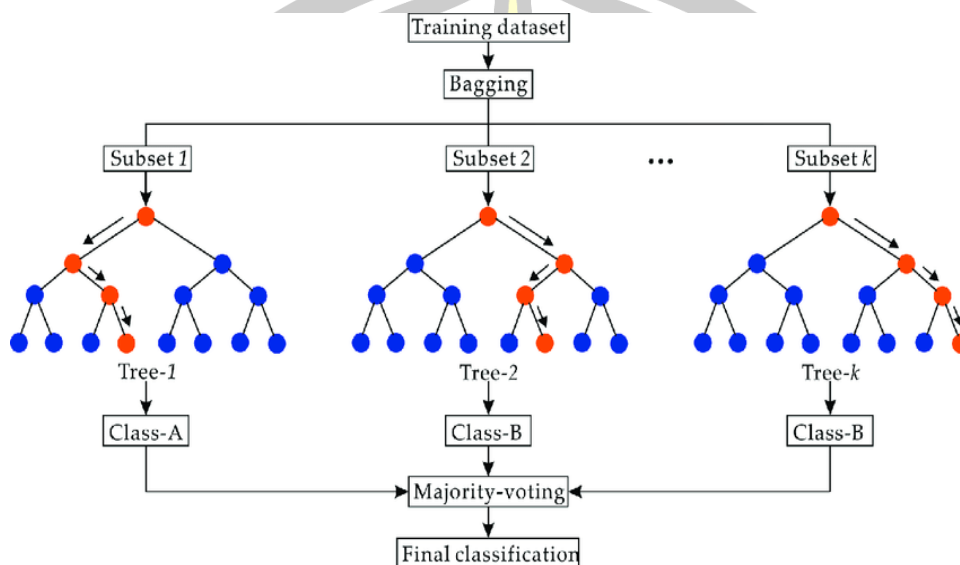
ภาพประกอบที่ 2.3 แบบโมเดลของโครงข่ายประสาทเทียม

( ที่มา: <https://medium.com/machine-learning-researcher/artificial-neural-network-ann-4481fa33d85a>)

โหนดอินพุตรับข้อมูลในรูปแบบที่แสดงเป็นตัวเลขได้ ข้อมูลจะแสดงเป็นค่าการเปิดใช้งาน โดยแต่ละโหนดจะรับตัวเลขข้อมูลนี้จะถูกส่งไปยังเครือข่ายขึ้นอยู่กับความแรงของการเชื่อมต่อ (น้ำหนัก) น้ำหนักเหล่านี้ใช้ในการคำนวณผลรวมถ่วงน้ำหนักของอินพุตและการใช้ฟังก์ชันการเปิดใช้งาน ในเลเยอร์ที่ซ่อนอยู่เพื่อสร้างข้อมูลในเลเยอร์เอาต์พุตของเลเยอร์เหล่านั้น มีการใช้แบบจำลอง ANN หลายแบบในการจำแนกประเภทการใช้ที่ดิน เช่น เครือข่ายฮอปฟิลด์ การแข่งขันที่จัดระเบียบตนเอง ฟังก์ชันพื้นฐานแบบรัศมีการรับรู้หลายชั้น และการรับรู้หลายชั้นพร้อมการแพร่กระจายย้อนกลับ บทความนี้ใช้การแพร่กระจายย้อนกลับมาตรฐานสำหรับตัวจำแนกประเภทคลาสการใช้ที่ดิน การแพร่กระจายย้อนกลับของข้อผิดพลาดใช้สำหรับการย้อนกลับผ่านเครือข่ายและการปรับน้ำหนักโดยใช้วิธีการเรียกซ้ำ [12]

2.2 การสุ่มพื้นที่ป่าไม้ (Random Forest) เทคนิคที่ทำการสุ่มเลือกคุณสมบัติออกมาจากชุดข้อมูลหลาย ๆ ชุด จากนั้น นำเอาชุดของคุณสมบัติเหล่านั้นมาสร้างแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจหลาย ๆ ต้น โดยเทคนิค การสุ่มพื้นที่ป่าไม้ถูกนำเสนอครั้งแรกในปี ค.ศ. 1995 โดย Tin Kam ซึ่งต่อมาเทคนิคนี้ถูกต่อยอดโดย Leo Breiman ลักษณะของต้นไม้ที่อยู่ภายในป่าของเทคนิคการสุ่มพื้นที่ป่าไม้จะถูกควบคุมด้วย 3 ปัจจัยคือ 1. ต้นไม้แต่ละต้นจะถูกสอน (Train) โดยการใช้เซตย่อยจากข้อมูลตัวอย่าง 2. เมื่อต้นไม้โตขึ้น จะสามารถค้นหาโหนด (Node) แต่ละโหนดที่อยู่ในกิ่งที่ดีที่สุดของต้นไม้โดยใช้การสุ่ม เลือกคุณสมบัติจาก  $N$  คุณสมบัติ 3 ต้นไม้แต่ละต้นจะไม่มีการตัดออก แต่จะปล่อยให้ต้นไม้โตขึ้นไปเรื่อย ๆ จนได้ผลลัพธ์ที่ดีที่สุดหลัง จากการสร้างป่าแล้วทำการให้

คะแนน (Vote) โดยต้นไม้ภายในป่า เทคนิคป่าสุ่ม (Random Forest) ทำงานโดยสร้างต้นไม้ตัดสินใจหลายต้น (Tree-1, Tree-2, ..., Tree-k) จากข้อมูลย่อยที่สุ่มมา และใช้การลงคะแนนเสียงข้างมาก (Majority Voting) ในการจำแนกขั้นสุดท้าย ดังแผนภาพหลักการทำงานดังภาพประกอบที่ 2.4



ภาพประกอบที่ 2.4 Characteristics of Random Forest

(ที่มา: [https://www.researchgate.net/figure/Flowchart-of-classification-using-RF-algorithm\\_fig3\\_332298301](https://www.researchgate.net/figure/Flowchart-of-classification-using-RF-algorithm_fig3_332298301))

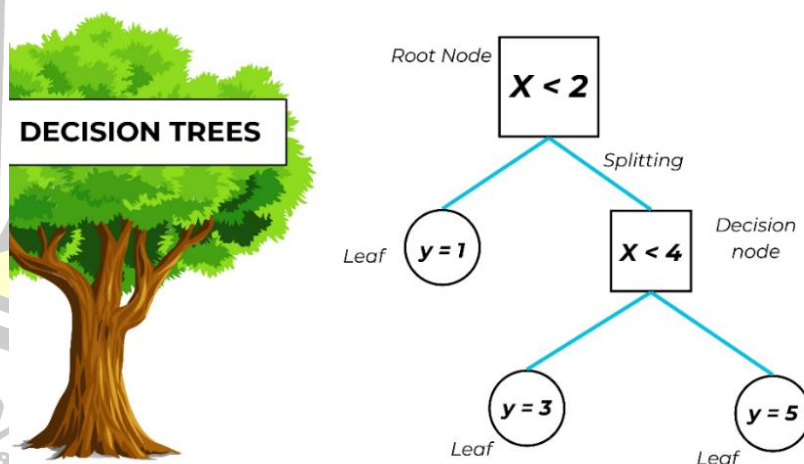
2.3 ซัพพอร์ตเวกเตอร์แมชชีน (Support vector machine) เป็นเทคนิคที่สามารถนำมาช่วยแก้ปัญหาการจำแนกข้อมูล ใช้ในการวิเคราะห์ข้อมูลและจำแนกข้อมูล โดยอาศัยหลักการของการหาสมมาตรของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้โดยเน้นไปยังเส้นแบ่งแยกแยะกลุ่มข้อมูลได้ดีที่สุด SVM ได้ถูกพัฒนาในกรอบแนวคิดของทฤษฎีการเรียนรู้เชิงสถิติและถูกนำไปประยุกต์ใช้ในหลาย ๆ ด้าน ไม่ว่าจะเป็นการประยุกต์ใช้ SVM เพื่อสนับสนุนการตัดสินใจทางการแพทย์, การทำนายข้อมูลที่เป็นอนุกรมเวลา (Time series), การระบุตัวตนด้วยใบหน้า เป็นต้น [13]

2.4 ต้นไม้ตัดสินใจ (Decision Tree) เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์ระหว่างตัวแปรต่าง ๆ และนำไปสู่การตัดสินใจ โดยแสดงผลในรูปแบบโครงสร้างต้นไม้ ทำให้เข้าใจง่ายและสามารถนำไปประยุกต์ใช้ได้หลากหลายด้าน ซึ่งโครงสร้างของต้นไม้ตัดสินใจ จะมี โหนดราก (Root Node) จุดเริ่มต้นของต้นไม้ แสดงถึงตัวแปรหลักที่ใช้ในการตัดสินใจ โหนดภายใน (Internal Node) แสดงถึงตัวแปรที่ใช้ในการแบ่งข้อมูลออกเป็นกลุ่มย่อย

กิ่ง (Branch) แสดงถึงผลลัพธ์ที่เป็นไปได้ของตัวแปรในโหนดภายใน โหนดใบ (Leaf Node) จุดสิ้นสุดของกิ่ง แสดงถึงผลลัพธ์สุดท้ายหรือการตัดสินใจ ประเภทของต้นไม้ตัดสินใจ 1.แบบจำแนก (Classification Tree) ใช้สำหรับจำแนกข้อมูลออกเป็นกลุ่มต่าง ๆ เช่น การจำแนกประเภทของลูกค้า หรือการวินิจฉัยโรค 2.ต้นไม้ตัดสินใจแบบถดถอย (Regression Tree) ใช้สำหรับทำนายค่าตัวแปรต่อเนื่อง เช่นการทำนายยอดขาย หรือราคาบ้าน หลักการทำงานของต้นไม้ตัดสินใจ

1. การเลือกตัวแปร เลือกตัวแปรที่สามารถแบ่งข้อมูลออกเป็นกลุ่มย่อยได้ดีที่สุด โดยพิจารณาจากค่าความบริสุทธิ์ของข้อมูล (impurity) หรือค่าความแปรปรวน (variance)
2. การแบ่งข้อมูล แบ่งข้อมูลออกเป็นกลุ่มย่อยตามค่าของตัวแปรที่เลือก
3. การสร้างกิ่ง สร้างกิ่งของต้นไม้เพื่อแสดงผลลัพธ์ที่เป็นไปได้ของตัวแปร
4. การสร้างโหนดภายใน ทำซ้ำขั้นตอนที่ 1-3 สำหรับแต่ละกลุ่มย่อย จนกว่าจะได้ผลลัพธ์สุดท้าย
5. การสร้างโหนดใบ กำหนดผลลัพธ์สุดท้ายหรือการตัดสินใจให้กับแต่ละโหนดใบ

ต้นไม้ตัดสินใจ (Decision Tree) มีโครงสร้างประกอบด้วยโหนดราก (Root Node) โหนดตัดสินใจ (Decision Node) และโหนดใบ (Leaf Node) ซึ่งแสดงเงื่อนไขการแบ่งข้อมูลและการตัดสินใจ ดังตัวอย่างโครงสร้างดังภาพประกอบที่ 2.5



ภาพประกอบที่ 2.5 แบบโมเดลต้นไม้ตัดสินใจ

(ที่มา :<https://www.insidealgorithms.com>)

3. Unsupervised Learning หรือ การเรียนรู้โดยไม่มีผู้สอน เป็นการเรียนรู้ที่ให้เครื่องจักรนั้นสามารถเรียนรู้ได้ด้วยตนเอง โดยไม่ต้องมีค่าเป้าหมายของแต่ละข้อมูล ซึ่งวิธีการคือ

มนุษย์จะเป็นผู้ใส่ข้อมูลต่าง ๆ และกำหนดสิ่งที่ต้องการจากข้อมูลเหล่านั้น ทำให้เครื่องจักรวิเคราะห์จากการจำแนกและสร้างแบบแผนจากข้อมูลที่ได้รับมา เรียกได้ว่าตรงกันข้ามกับรูปแบบแรกเลย ตัวอย่างเช่น การที่เราป้อนข้อมูล (Input) รูปปากกาเข้าไป แต่ไม่ได้บอกว่ารูปที่ป้อนเข้าไปเป็นรูปปากกา เมื่อคอมพิวเตอร์นำไปวิเคราะห์ (Feature Extraction) ก็ยังสามารถวิเคราะห์ได้ว่ารูปที่ใส่เข้าไปมีลักษณะยังไง แต่คราวนี้มันไม่สามารถเอาไปประมวล/จัดหมวดหมู่ (Classification) ได้แล้วมันจะใช้วิธีการแบ่งกลุ่มแทน (Clustering) ซึ่งคอมพิวเตอร์ก็อาจเอารูปปากกาไปจัดกลุ่มกับปากกาไฮไลท์หรือเครื่องเขียนอื่น ๆ ที่มีปลายด้ามเป็นปุ่ม และใช้หมึกในการเขียนเหมือนกัน เป็นต้น

4. Reinforcement Learning หรือ การเรียนรู้แบบเสริมกำลัง เป็นวิธีการเรียนรู้แบบหนึ่งที่ใช้การเรียนรู้เกิดมาจากการปฏิสัมพันธ์ (interaction) ระหว่างผู้เรียนรู้ (agent) กับสิ่งแวดล้อม (environment) ที่มีการเรียนรู้สิ่งต่าง ๆ จาก Agent ภายใต้การเลือกกระทำสิ่งต่าง ๆ ให้ได้ผลลัพธ์ ที่มากที่สุดผ่านการลองผิดลองถูกภายใต้สถานการณ์หรือระบบจำลองที่พัฒนาระบบการตัดสินใจให้ดีขึ้นเรื่อย ๆ หรือพูดให้ง่ายขึ้น คือการที่เรากำหนดเงื่อนไขบางอย่างให้กับคอมพิวเตอร์ แล้วทำให้คอมพิวเตอร์บรรลุหรือทำตามเงื่อนไขนั้นได้ผ่านการลองผิดลองถูก โดยผู้พัฒนาอาจตั้งเป้าหมาย Feedback Loop และเงื่อนไขในการได้รับรางวัล ยกตัวอย่างเช่น Alpha Go เงื่อนไขของการเล่นเกมหมากล้อมให้ชนะคือ ใช้หมากของตนล้อมพื้นที่บนกระดานให้ครอบคลุมดินแดนมากกว่าคู่ต่อสู้ ที่นี้ Alpha Go ก็จะเรียนรู้ว่าหากคู่ต่อสู้เดินหมากนี้ ตัวมันเองจะเดินหมากไหนเพื่อให้บรรลุเงื่อนไขที่กำหนดไว้ให้ นั่นคือการยึดพื้นที่บนกระดานให้ได้มากที่สุด

#### 2.1.6 การวัดประสิทธิภาพของแบบจำลอง

2.1.6.1 การวิเคราะห์หลังการจำแนก (Post-processing) ประมวลผลเพื่อจำแนกวัตถุแต่ละชนิด จะนำผลลัพธ์ที่ได้มาวิเคราะห์และแก้ไขให้มีความถูกต้องมากขึ้น เช่น บริเวณขอบของพื้นที่เนื่องจากข้อจำกัดของความละเอียดคุณภาพของข้อมูลภาพถ่ายดาวเทียมทำให้บางจุดภาพมีข้อมูลของวัตถุมากกว่า 1 ชนิด เช่น มีข้อมูลจากพื้นที่นา 50% และมีพื้นที่ดิน 50% เป็นต้น ส่งผลให้การแปลผิดพลาดเพราะจุดภาพดังกล่าวเป็นนาข้าวแค่ 50% จึงไม่สามารถบอกได้ว่าจุดดังกล่าวเป็นพื้นที่นาหรือไม่ใช่พื้นที่นา [14]

2.1.6.2 K-Fold Cross-Validation เป็นเทคนิคสำคัญในการประเมินประสิทธิภาพของเทคนิค Machine Learning โดยหลักการคือการแบ่งข้อมูลทั้งหมดออกเป็นส่วน ๆ (folds) จำนวน k ส่วน จากนั้นจะทำการวนรอบฝึกและทดสอบเทคนิค k ครั้ง โดยในแต่ละรอบจะใช้ข้อมูล k-1 ส่วนในการฝึก และใช้ส่วนที่เหลือ 1 ส่วนในการทดสอบ ทำเช่นนี้เพื่อให้เทคนิคได้เรียนรู้จากข้อมูล

หลากหลายรูปแบบ และเพื่อให้ได้ค่าประเมินประสิทธิภาพที่น่าเชื่อถือมากขึ้น เมื่อวนรอบครบทุกครั้ง จะนำผลลัพธ์จากการประเมินในแต่ละรอบมาหาค่าเฉลี่ย เพื่อให้ได้ค่าประเมินประสิทธิภาพของเทคนิคโดยรวม เทคนิคนี้ช่วยลดปัญหา Overfitting และทำให้เทคนิคมีความสามารถในการทั่วไป (Generalization) ที่ดีขึ้น

2.1.6.3 กระบวนการฝึกและทดสอบเทคนิค สำหรับแต่ละ fold จะมีการทำงาน k รอบ โดยในแต่ละรอบ เทคนิคจะถูกฝึกด้วยข้อมูล k-1 fold (training set) และทำการทดสอบด้วยข้อมูล 1 fold ที่เหลือ (validation set) กระบวนการนี้จะทำซ้ำทั้งหมด k ครั้ง โดยแต่ละ fold จะถูกใช้เป็นชุดทดสอบหนึ่งครั้ง ตัวอย่างเช่น ถ้า  $k = 5$  จะมีการแบ่งข้อมูลเป็น 5 fold และทำการฝึก 5 ครั้ง โดยแต่ละ fold จะถูกใช้เป็นชุดทดสอบหนึ่งครั้งและใช้ส่วนที่เหลือสำหรับฝึกเทคนิค

2.1.6.4 การคำนวณค่าเฉลี่ยของผลการทดสอบ ผลลัพธ์ของการทดสอบในแต่ละ fold จะถูกบันทึกไว้ เช่น ค่า accuracy, precision, recall, F1 score หรือค่าความผิดพลาด (error) เช่น Mean Squared Error (MSE) หลังจากทดสอบครบทุก fold จะทำการเฉลี่ยค่าผลลัพธ์เหล่านี้ เพื่อให้ได้ค่าประสิทธิภาพของเทคนิคที่แสดงถึงความสามารถในการทำนายบนข้อมูลที่ไม่เคยเห็นมาก่อนการคำนวณค่าเฉลี่ยนี้จะช่วยลดความเบี่ยงเบนในการประเมินประสิทธิภาพของเทคนิคและทำให้ประเมินได้แม่นยำขึ้น

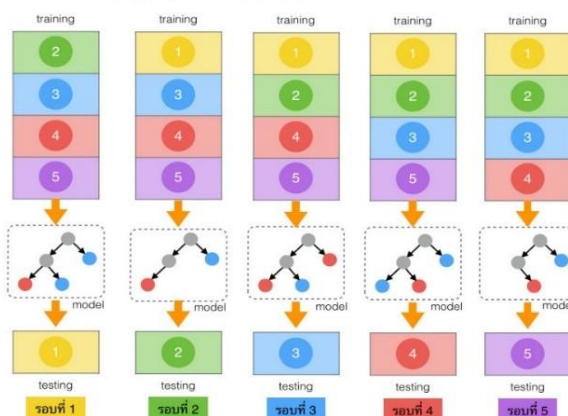
2.1.6.5 การประเมินผลรวมของเทคนิคค่าเฉลี่ยของผลลัพธ์จาก k fold จะเป็นค่าที่ใช้ในการประเมินประสิทธิภาพของเทคนิค ค่าเหล่านี้ช่วยให้เข้าใจว่าเทคนิคทำงานได้ดีแค่ไหนในการทำนายข้อมูลใหม่ ๆ (generalization) นอกจากนี้ยังสามารถดูค่าความแปรปรวนของผลลัพธ์ (variance) เพื่อวิเคราะห์ความเสถียรของเทคนิค

การประเมินความเสถียรของแบบจำลองสามารถทำได้ด้วยวิธี K-Fold Cross Validation ซึ่งเป็นการแบ่งข้อมูลออกเป็น k ส่วน และสลับกันใช้ส่วนหนึ่งเป็นข้อมูลทดสอบ (Testing) และส่วนที่เหลือเป็นข้อมูลฝึกสอน (Training) ดังตัวอย่างการทำงานแบบ 5-Fold ดังภาพประกอบที่ 2.6

## Cross-validation



## • ตัวอย่าง 5-fold cross-validation



ภาพประกอบที่ 2.6 หลักการทำงานของ k-Fold Cross-Validation

(ที่มา: [www.dataminingtrend.com](http://www.dataminingtrend.com))

2.1.6.6 วิธีการประเมินความถูกต้อง การประเมินความถูกต้องของการจำแนกข้อมูลสำรวจระยะไกลนั้นเป็นการพิจารณาข้อมูล ผลการจำแนกชั้นข้อมูล ณ ตำแหน่งที่ตั้งหน่วยตัวอย่าง (จุดสำรวจ) เปรียบเทียบความสอดคล้องกันกับความเป็นจริงที่พบในภาคสนามที่ถือว่าเป็นข้อมูลอ้างอิง แล้วทำการแจกแจงให้อยู่ในรูปของตารางข้อมูล ที่เรียกว่า Error Matrix หรือ Confusion Matrix หรือ Contingency Table โดยสามารถใช้วิเคราะห์ หาค่าแสดงความถูกต้องได้หลายลักษณะ ดังนี้

2.1.6.7 ความถูกต้องของผู้ผลิต (Producer's Accuracy) ประสิทธิภาพของการชั้นข้อมูลของตัวจำแนก (Classifier) เป็นการวัดว่าในชั้นข้อมูลหนึ่ง ๆ ตัว จำแนกสามารถจัดชั้นได้ดีมากน้อยเพียงใดเมื่อพิจารณาความผิดพลาดที่เกิดจากการละข้อมูลไว้ (Omission Error) อีกลักษณะหนึ่งคือ จากข้อมูลอ้างอิงของชั้นข้อมูลหนึ่ง ๆ ที่ใช้ตรวจสอบมีการจำแนกข้อมูลถูกต้องเท่าไรในความผิดพลาดของการจำแนกที่เกิดขึ้นจากการที่ไม่ได้จำแนกข้อมูลชั้นนั้นตามที่มีอยู่จริง เป็นประโยชน์ต่อการวินิจฉัยว่าตัวจำแนกทำงานได้ดีเพียงใดในการ จำแนกข้อมูลประเภทหนึ่ง ๆ อย่างไรก็ตามคำอธิบายเชิงหลักการดังกล่าวมานี้อาจดูเข้าใจยาก

ค่าความถูกต้องของผู้ผลิต (Producer's Accuracy - PA) ซึ่งเป็นสัดส่วนของจำนวน 픽เซลที่จำแนกได้ถูกต้องในแต่ละคลาสเมื่อเทียบกับค่าจริง สามารถคำนวณได้ดังสมการที่ 2.1

$$PA = \frac{\text{True Positive}}{\text{Column Total}} \times 100 \quad (2.1)$$

2.1.6.8 ความถูกต้องของผู้ใช้ (User's Accuracy) การจำแนกสามารถบ่งบอกความน่าเชื่อถือในการนำข้อมูลในแต่ละชั้นข้อมูลไปใช้ เป็นการวัดว่าผล การจำแนกชั้นข้อมูลหนึ่ง ๆ ถูกต้องมากน้อยเพียงใด โดยเป็นความผิดพลาดที่เกิดจากการจำแนกข้อมูล ประเภทนั้นมากกว่าความเป็นจริงในลักษณะที่เป็นการรวมพื้นที่ประเภทอื่นเข้ามา (Commission Errors) เป็นประโยชน์ต่อการวินิจฉัยการจำแนกชั้นข้อมูลใดที่ได้ผลดีเป็นที่เชื่อถือได้ ถ้าอธิบายตาม แนวทางของตัวอย่างข้างต้น ก็จะได้ว่าในการแปลภาพเพื่อจำแนกชนิดป่าหนึ่งเมื่อนำไปประเมินในพื้นที่จริงมีความถูกต้องมากน้อยเพียงใด โดยเอาจุดที่ถูกต้องตรงกันหารด้วยจำนวนจุดที่แปลภาพว่าเป็นชนิดป่านั้นนำไปสำรวจในภาคสนามทั้งหมดคูณด้วยร้อยเพื่อคิดเป็นเปอร์เซ็นต์ ซึ่งจะสะท้อนผลการแปลภาพได้ถูกต้องแค่ไหน ส่วนที่ผิดก็คือการแปลที่เกินความเป็นจริงเพราะเป็นการนำจุดภาพที่ตามจริงแล้วเป็นชนิดป่าอื่นเข้ามาด้วย [15]

ขณะที่ค่าความถูกต้องของผู้ใช้ (User's Accuracy - UA) ซึ่งบ่งบอกความน่าเชื่อถือของผลการจำแนกในแต่ละคลาสบนแผนที่ คำนวณได้ ดังสมการที่ 2.2

$$UA = \frac{\text{True Positive}}{\text{Row Total}} \times 100 \quad (2.2)$$

2.1.6.9 ค่าความถูกต้องโดยรวม (Overall Accuracy - OA) ซึ่งเป็นการวัดความแม่นยำของการจำแนกทั้งหมดโดยไม่แยกประเภท คำนวณโดยนำผลรวมของค่าบนเส้นทแยงมุมใน Confusion Matrix หารด้วยจำนวนตัวอย่างทั้งหมด ดังสมการที่ 2.3

- นำค่าบนเส้นทแยงมุมมาบวกกัน (True Positives ของแต่ละคลาส)
- หารด้วยจำนวนตัวอย่างทั้งหมด

$$OA = \frac{\sum \text{Diagonal Elements}}{\text{Total Samples}} \times 100 \quad (2.3)$$

2.1.6.10 ค่าสถิติแคปปา (Kappa Statistics) การแสดงความถูกต้องโดยรวมเช่นกัน แต่มีการพิจารณาลักษณะของความผิดพลาดที่เกิดขึ้นใน Error Matrix ด้วยการใช้สถิติแคปปาเป็นการประเมินความถูกต้องที่ถือว่าการพิจารณาทั้งความสอดคล้องกันระหว่างข้อมูลตรวจสอบกับข้อมูลอ้างอิงที่เป็นไปตามโอกาส (Chance Agreement) และที่เป็นไปตามจริง (Actual Agreement) วิเคราะห์เพื่อแสดงควบคู่ไปกับค่าความถูกต้องโดยรวม โดยหากค่าทั้งสองมีค่าที่ใกล้เคียงกันไปในทิศทางสูงก็ถือว่าผลการจำแนกข้อมูลนั้นดีกล่าวคือ การที่ความถูกต้องโดยรวมมีค่า (%) ที่สูง ไม่ได้

หมายความว่าผลการจำแนกครั้งนั้นดีหากค่าสถิติแคปามีค่าที่ต่ำ เพราะนั่นแสดงว่าความถูกต้องที่สูงเกินไปโดยบังเอิญ หากพิจารณาในตารางแจกแจงความผิดพลาดก็จะพบว่ามีค่าที่ไม่อยู่ในแนวทแยงบนซ้าย - ล่างขวากระจายอยู่มาก แสดงว่าการจำแนกมีลักษณะผันผวนอยู่ [15]

ผลที่ได้จากการจำแนกจะถูกนำไปตรวจสอบความถูกต้องโดยหาค่าคลเคลื่อนประกอบ 2 ประเภท ได้แก่ Error of omission และ Error of commission ส่วนความเที่ยงตรงของการจำแนกมี 2 ประเภท ได้แก่ ความถูกต้อง Producer 's Accuracy และ User's Accuracy ซึ่งจะนำมาตรวจสอบความถูกต้องเที่ยงตรงของการ จำแนกโดยรวม (Overall accuracy) และค่าสัมประสิทธิ์ Kappa จากนั้นเปรียบเทียบผลการจำแนก [16]

เพื่อประเมินความสอดคล้องของการจำแนกโดยพิจารณาถึงความถูกต้องที่อาจเกิดขึ้นโดยบังเอิญ จะใช้ค่าสัมประสิทธิ์แคป (Kappa Coefficient - K) ซึ่งคำนวณได้จากสมการที่ 2.4 เป็นตัวชี้วัดที่ใช้ในการประเมินความแม่นยำของการจำแนกประเภท โดยเปรียบเทียบกับค่าความถูกต้องที่เกิดขึ้นแบบสุ่ม ซึ่งช่วยให้เข้าใจได้ว่าเทคนิคทำงานดี โดยไม่ถูกลำเอียงจากความน่าจะเป็นที่จำแนกถูกต้องแบบสุ่ม [17] ตารางที่ 2.3 แสดงเกณฑ์การแปลความหมายค่าสถิติ Kappa ตามแนวทางของ Landis และ Koch [18] เพื่อประเมินระดับความสอดคล้องของการจำแนก

$$K = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_i + Xx_{+1})}{N^2 - \sum_{i=1}^r (x_{i.} Xx_{+1})} \quad (2.4)$$

ตารางที่ 2.3 ระดับความสอดคล้องของสถิติแคปตามแนวทางของ Landis และ Koch

ค่าสถิติ Kappa	ระดับความสอดคล้องระหว่างผู้ประเมิน
0.81-1.00	ความสอดคล้องดีมาก (Almost Perfect)
0.61-0.80	ความสอดคล้องดี (Substantial)
0.41-0.60	ความสอดคล้องปานกลาง (Moderate)
0.21-0.40	ความสอดคล้องพอใช้ (Fair)
0.00-0.20	ความสอดคล้องเล็กน้อย (Slight)
น้อยกว่า 0.00	ไม่มีความสอดคล้อง (Poor)

จากการคำนวณค่าสัมประสิทธิ์แคปสามารถอธิบายได้ว่า สัดส่วนของค่าความผิดพลาดที่ได้รับจากการจำแนกประเภทของวัตถุเปรียบเทียบกับ ค่าความผิดพลาดที่ได้รับจากการ

จำแนกจากการสุ่มตัวอย่างที่สมบูรณ์แล้ว ค่าของสัมประสิทธิ์จะมีค่าอยู่ในช่วง 0-1 ถ้าจากการคำนวณได้ผลเท่ากับ 0.75 แสดงว่าการจำแนกจากการทำงานนี้มีค่าเท่ากับ 75% ของค่าความผิดพลาดที่ได้รับจากการทำงานในการกำหนดกลุ่มตัวอย่างหรือข้อมูลที่ใช้ในการอ้างอิง [19]

## 2.2 งานวิจัยที่เกี่ยวข้อง

Intarat และ Sillaparat [20] ศึกษาการจำแนกพันธุ์ไม้ป่าชายเลนเขตร้อนด้วยเทคนิคป่าสุ่ม และข้อมูลภาพถ่ายจากดาวเทียมรายละเอียดสูง ผลที่ได้จะถูกนำไปเปรียบเทียบประสิทธิภาพการจำแนกกับวิธีการจำแนกเชิงจุดภาพแบบความน่าจะเป็นสูงสุด โดยใช้ค่าความถูกต้องโดยรวมของการจำแนกค่าสถิติแคปปา และค่าสถิติการทดสอบ Z เป็นตัวชี้วัด ผลการจำแนกพบว่า การใช้เทคนิคป่าสุ่มให้ประสิทธิภาพในการจำแนกที่สูงกว่าวิธีการจำแนกเชิงจุดภาพแบบความน่าจะเป็นสูงสุด โดยมีค่าความถูกต้องโดยรวมที่ร้อยละ 78.00 และค่าสถิติ แคปปาที่ 0.72 ในขณะที่ผลการจำแนกด้วยความน่าจะเป็นสูงสุดให้ค่าความถูกต้องโดยรวมที่ร้อยละ 56.00 และค่าสถิติ แคปปาที่ 0.44 ผลจากการทดสอบค่าสถิติ Z ( $Z = 3.68$ ) ช่วยยืนยันถึงความแตกต่างระหว่างทั้งสองวิธีการจำแนก อย่างมีนัยสำคัญที่ค่าความเชื่อมั่นร้อยละ 95.00

Siri-on, Diloksumpun และ Khunrattanasiri [21] ศึกษาเทคนิคที่เหมาะสมในการจำแนกการใช้ประโยชน์ที่ดินด้วยเทคนิคการจำแนกเชิงจุดภาพ (pixel-based classification) และเทคนิคการจำแนกเชิงวัตถุ (object-based classification) เปรียบเทียบกับการจำแนกด้วยสายตา ผลการศึกษาพบว่า เทคนิคการจำแนกเชิงจุดภาพ ร้อยละความถูกต้องของการจำแนกโดยรวม เท่ากับ 39.4 ค่าสัมประสิทธิ์ Kappa เท่ากับ 0.27 และ เทคนิคการจำแนกเชิงวัตถุ ร้อยละความถูกต้องของการจำแนกโดยรวม เท่ากับ 34.2 ค่าสัมประสิทธิ์ Kappa เท่ากับ 0.22 ซึ่งเทคนิคการจำแนกเชิงจุดภาพให้ค่าที่สูงกว่าเล็กน้อยแต่ อยู่ในเกณฑ์การพิจารณาระดับความสอดคล้องของสัมประสิทธิ์ Kappa ระดับพอใช้ ทั้งสองเทคนิค แสดงว่าภาพถ่าย เชิงเลขจากอากาศยานไร้คนขับที่มีเพียงแถบช่วงคลื่น RGB ไม่เหมาะกับการจำแนกโดยใช้เครื่องคอมพิวเตอร์ช่วย มีความจำเป็นต้องศึกษาเทคนิคเพิ่มเติม

Intarat [22] ทำการศึกษาการจำแนกการใช้ประโยชน์ที่ดินของจังหวัดนครนายกด้วยเทคนิค การเรียนรู้เครื่อง(Machine Learning) ร่วมกับภาพถ่ายจากดาวเทียม Sentinel-2 ผลการศึกษาพบว่า จำแนกป่าสุ่มมีประสิทธิภาพในการจำแนกการใช้ประโยชน์ที่ดินมากที่สุดโดยมีค่าความถูกต้องโดยรวมที่ร้อยละ 92.00 ตามด้วยต้นไม้ตัดสินใจ ข่ายประสาทเทียม ซัพพอร์ตเวกเตอร์แมชชีน และความน่าจะเป็นสูงสุด (ร้อยละ 84.00 69.00 65.00 และ 63.00 ตามลำดับ) ผลการ

ทดสอบ Z-test ช่วยยืนยันถึงความแตกต่างของตัวจำแนกป่าสุ่มกับตัวจำแนกอื่นอย่างมีนัยสำคัญที่ระดับความเชื่อมั่นร้อยละ 95

Fang และคณะ [23] การทำแผนที่ที่มีความสำคัญอย่างยิ่งต่อการจัดการการผลิตทางการเกษตรและความมั่นคงทางอาหาร และเทคนิคการเรียนรู้ของเครื่องจักร เช่น SVM ,RF และ CART บนพื้นฐานของแพลตฟอร์ม GEE ใช้ภาพถ่ายหลายสเปกตรัมความละเอียด Sentinel-2 10 m. และรวมเข้ากับเทคนิค SVM ,RF และ CART เพื่อระบุและแผนที่ข้าวสาลีฤดูหนาวในพื้นที่ขนาดใหญ่ ผลการศึกษาพบว่า เมื่อเปรียบเทียบกับเทคนิค RF และ CART แล้ว SVM บรรลุที่มีประสิทธิภาพมากที่สุดในการระบุข้าวสาลีฤดูหนาว แม้ว่าเทคนิค SVM จะไวต่อพารามิเตอร์ แต่ก็ได้รับ ค่าความแม่นยำสูงสุดและค่าคงเหลือโดยการปรับไฮเปอร์พารามิเตอร์ ดังนั้นเราขอแนะนำว่า SVM จะเป็นเทคนิคที่เหมาะสมและมีประสิทธิภาพ

Aziz และคณะ [24] ได้นำเทคนิคขั้นสูงที่ใช้เทคนิค Machine Learning และ Deep Learning มาใช้ในการจำแนกประเภทพื้นที่ป่าไม้ในเขตเมืองแอบบอตทาบัต ประเทศปากีสถาน เนื่องจากเป็นประเทศที่มีพื้นที่ป่าไม้น้อย โดยมีพื้นที่ป่าไม้ปกคลุมน้อยกว่า 6% ของพื้นที่ทั้งหมด เพื่อแก้ไขปัญหาเทคนิค Artificial Neural Network (ANN) และ Random Forest จึงถูกนำมาใช้ ซึ่งให้ผลลัพธ์ที่ทันสมัยที่สุดในแง่ของความแม่นยำโดยรวม และค่าสัมประสิทธิ์ Kappa เทคนิค ANN แสดงให้เห็นถึงประสิทธิภาพที่โดดเด่น โดยให้ความแม่นยำที่ 97.75% และค่าสัมประสิทธิ์ Kappa ที่ 0.965 อย่างไรก็ตาม เพื่อแก้ไขปัญหาที่มีอยู่ใน ANN จึงได้นำเทคนิค Random Forest มาใช้ วิธีนี้ทำให้ได้ความแม่นยำที่ 96.98% และค่าสัมประสิทธิ์ Kappa ที่ 0.954 โดยเฉพาะอย่างยิ่งเมื่อใช้ความลึกสูงสุด 20 สำหรับภาพซ้อนของ Temporal Layer การนำเทคนิค ANN ไปใช้กับพื้นที่ทั้งหมด 166,103 เฮกตาร์ของเมืองแอบบอตทาบัต แสดงให้เห็นพื้นที่ป่าไม้ 51,613 เฮกตาร์ คิดเป็น 31.07% ของพื้นที่ทั้งหมด ในขณะที่เดียวกับการใช้เทคนิค Random Forest สามารถระบุพื้นที่ป่าไม้ทั้งหมดได้ 51,774 เฮกตาร์ คิดเป็น 31.17% ของพื้นที่เขตเมืองแอบบอตทาบัต เพื่อเพิ่มความแม่นยำในการจำแนกพื้นที่ป่าไม้ ขอแนะนำให้รวมภาพถ่ายดาวเทียมแบบไฮเปอร์สเปกตรัม นอกจากนี้ สำหรับการปรับปรุงในอนาคต จะมีการสำรวจเทคนิคการเรียนรู้เชิงลึก เช่น Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM) และ Gated Recurrent Units (GRU) เพื่อดูศักยภาพในการเพิ่มความแม่นยำในการจำแนก

Saini และ Singh [25] การทำแผนที่การใช้ที่ดินและการตรวจจัดการปกคลุมหิมะในภูมิภาคหิมาลัยโดยใช้การเรียนรู้ของเครื่องจักรและมัลติสเปกตรัมภาพถ่ายดาวเทียม Sentinel-2 มี

วัตถุประสงค์เพื่อทำแผนที่การปกคลุมดินเพื่อการใช้ที่ดิน (LULC) ตรวจสอบการปกคลุมหิมะในเขตเทือกเขาหิมาลัย Chamoli รัฐUttarakhand ประเทศอินเดีย โดยใช้ภาพถ่าย Sentinel-2 แบบวันเดียว โดยใช้วิธีการเรียนรู้ของเครื่องจักร สามวิธี (Random Forest (RF), Support Vector Machine (SVM) และ k-nearest Neighbor (KNN)) เพื่อจำแนกภาพถ่ายดาวเทียม Sentinel-2 ซึ่งรวมถึงแถบแสงอินฟราเรดใกล้(NIR) และแสงที่มองเห็นได้ ผลลัพธ์แสดงให้เห็นถึงการทำแผนที่ชั้น LULC ผลการศึกษานี้ประสบความสำเร็จในการจำแนกประเภทโดยรวมที่ 91.01% ,89.67% และ 87.88% สำหรับ SVM, RF และ KNN ตามลำดับ SVM รายงานความแม่นยำสูงสุดและเพิ่มขึ้น +1.31% และ+3.11% เมื่อเทียบกับ RF และ KNN ตามลำดับ ชั้นหิมะ ป่าไม้และทราย มีความแม่นยำสูงกว่าในขณะที่สิ่งก่อสร้างและแหล่งน้ำ มีความแม่นยำเฉพาะชั้นต่ำกว่า โดยการจำแนก 3 เทคนิค การวิเคราะห์เผยให้เห็นว่า การจำแนกทั้งหมดทำแผนที่ปกคลุมหิมะ ได้อย่างมีประสิทธิภาพ โดยมีคะแนน F1 97.97%, 97.72% และ 96.80% โดย SVM, RF และ KNN ตามลำดับ ผลลัพธ์บ่งชี้ว่าการประมาณหิมะ พื้นที่ครอบคลุม 2758 ตร.กม. 2,744 กม. โดย SVM, RF และ KNN ตามลำดับ ผลการศึกษานี้แสดงให้เห็นว่าการใช้ข้อมูลดาวเทียมมัลติสเปกตรัม Sentinel-2 เป็นทางเลือกที่เหมาะสมสำหรับการทำแผนที่ปกคลุมหิมะ 2,742 ตร.กม

Buathongkuea และคณะ [11] ประยุกต์ใช้ข้อมูลดาวเทียม SENTINEL -2 กับ LANDSAT 8 เพื่อจำแนกการใช้ประโยชน์ที่ดินระดับ 2 พื้นที่คาบสมุทรสทิงพระ จังหวัดสงขลา ด้วยเทคโนโลยีการสำรวจข้อมูลระยะไกล โดยการเปรียบเทียบความถูกต้องการจำแนกการใช้ประโยชน์ที่ดินแบบไม่กำกับดูแล (Unsupervised Classification) ผลการศึกษาพบว่า ดาวเทียม SENTINEL -2 ให้ค่าความ ถูกต้องสูงกว่าดาวเทียม LANDSAT 8 ในช่วงฤดูร้อน ค่าความถูกต้องเป็น ร้อยละ 44.94 และร้อยละ 38.38 ตามลำดับ จากนั้นได้ศึกษาการใช้ประโยชน์ที่ดินระดับ 2 โดยการจำแนกแบบกำกับดูแล (Supervised Classification) โดยใช้ข้อมูลดาวเทียม SENTINEL -2 จากการวิเคราะห์ ค่าความน่าเชื่อถือข้อมูลเทียบกับข้อมูลจุดตรวจภาคสนาม มีค่าระหว่างร้อยละ 86.18 - 94.62 การใช้ประโยชน์ที่ดินที่ไม่มีข้อผิดพลาดจากการจำแนก พบ 6 ประเภท คือ พื้นที่ชุมชนและสิ่งปลูกสร้าง ไม้ยืนต้น ไร่ หมุนเวียน ป่าไม้ผลัดใบ ป่าชายหาด และหาดทราย ตามลำดับ ส่วน ประเภทที่ไม่เหมาะสมในการจำแนกการใช้ประโยชน์ที่ดิน คือ พืชสวน เนื่องจากมีค่าความถูกต้องน้อยที่สุด คิดเป็นร้อยละ 25 การจำแนกการใช้ประโยชน์ที่ดินในพื้นที่ศึกษา พื้นที่เกษตรกรรมมีเนื้อที่การใช้ประโยชน์ที่ดินมากที่สุด และรองลงมาเป็นพื้นที่ชุมชนและสิ่งปลูกสร้างพื้นที่แหล่งน้ำ พื้นที่ป่าไม้ และพื้นที่เบ็ดเตล็ด ตามลำดับ ซึ่งมีความสอดคล้องกับข้อมูลการใช้ประโยชน์ที่ดินของกรมพัฒนาที่ดิน

Alshari และคณะ [26] ซึ่งนำเสนอการจำแนกประเภทการใช้ที่ดินและสิ่งปกคลุมดิน (LULC) โดยใช้เทคนิคปัญญาประดิษฐ์แบบผสมผสาน Artificial Neural Network (ANN) ร่วมกับ Random Forest (RF) ในรูปแบบโมเดล ANN\_RF โดยการใช้ข้อมูลจาก Sentinel-2A และ Landsat-8 พร้อมทั้งรวมข้อมูลจากแบบจำลองความสูง (DEM) การทดลองแสดงให้เห็นว่าโมเดล ANN\_RF ให้ค่า Overall Accuracy สูงกว่า ANN เดี่ยวอย่างชัดเจน เช่น Sentinel-2A ให้ผลจาก 61.69% เพิ่มเป็น 82.52% และ Landsat-8 จาก 62.07% เป็น 80.00% ซึ่งสะท้อนถึงศักยภาพของการรวม ANN กับ RF ในการลด Overfitting และเพิ่มประสิทธิภาพในการ generalize โมเดลนี้ผ่านการทดสอบด้วย Confusion Matrix และการประเมินค่า Kappa Coefficient ซึ่งชี้ให้เห็นว่าความแม่นยำในการจำแนกประเภท LULC เพิ่มขึ้นอย่างมีนัยสำคัญ โดยเฉพาะในบริบทของภูมิภาคหลากหลาย เช่น เมือง Sana'a ประเทศเยเมน การเลือกใช้ ANN เพื่อเรียนรู้ลักษณะเชิงลึกของข้อมูลร่วมกับ RF ที่จัดการข้อมูลเชิงมิติสูงได้ดี ทำให้ ANN\_RF กลายเป็นทางเลือกที่มีประสิทธิภาพสำหรับงานจำแนกประเภทจากข้อมูลภาพถ่ายดาวเทียม การศึกษานี้สนับสนุนการใช้ Hybrid AI Model เพื่อเพิ่ม Validation Accuracy ของโมเดล และสามารถนำมาเป็นแนวทางเสริมในการเลือกโมเดลที่เหมาะสมในบริบทของงานวิจัยฉบับนี้ได้เป็นอย่างดี

Aroonsri และ Sangpradid [12] ศึกษาการจำแนกพื้นที่ฟาร์มกุ่มจากภาพถ่ายดาวเทียม Sentinel-2 โดยใช้เทคนิค Artificial Neural Networks (ANN) ซึ่งเน้นถึงความสำคัญของการเปลี่ยนแปลงการใช้ประโยชน์ที่ดินอันเนื่องมาจากการขยายตัวของอุตสาหกรรมการเลี้ยงกุ่มในประเทศไทย งานวิจัยนี้มีวัตถุประสงค์สองประการ ได้แก่ (1) การจำแนกพื้นที่ฟาร์มกุ่มด้วย ANN และ (2) การตรวจสอบการเปลี่ยนแปลงการใช้ที่ดินในช่วงปี 2015, 2018 และ 2020 โดยใช้เทคนิค image differencing ซึ่งผลการศึกษาพบว่า ANN สามารถจำแนกพื้นที่ฟาร์มกุ่มได้ครอบคลุมถึง 80% ของพื้นที่ศึกษา และแสดงให้เห็นแนวโน้มการเพิ่มขึ้นของพื้นที่ฟาร์มกุ่มอย่างต่อเนื่อง สอดคล้องกับความต้องการในตลาดโลกที่สูงขึ้นสำหรับผลิตภัณฑ์กุ่ม ผลลัพธ์ดังกล่าวแสดงให้เห็นว่า ANN สามารถประมวลผลลักษณะ spectral ที่ซับซ้อนและเปลี่ยนแปลงได้ในระยะเวลาหลายปี จึงเป็นเครื่องมือสำคัญในการติดตามการใช้ประโยชน์ที่ดินในเชิงนโยบายและธุรกิจ

Pacheco และคณะ [27] ได้วิเคราะห์การเปลี่ยนแปลงการใช้ที่ดินในพื้นที่เหมืองทองคำในรัฐ Pernambuco ประเทศบราซิล โดยใช้ข้อมูลภาพถ่ายจากดาวเทียม PlanetScope และเทคนิค k-Nearest Neighbors (kNN) พร้อมการประเมินด้วย NDVI และ NDWI พบว่าโมเดลสามารถให้ค่า Overall Accuracy มากกว่า 99% และ Kappa เท่ากับ 0.98 ซึ่งสะท้อนความน่าเชื่อถือในการ

ประเมินผลการจำแนกในสภาพพื้นที่ซับซ้อน โดยเน้นว่าการวิเคราะห์หลายช่วงเวลาและการใช้ค่าดัชนีร่วมกันมีบทบาทอย่างมากในการเพิ่มความแม่นยำของโมเดลจำแนกภาพถ่ายดาวเทียม ผลการประเมินประสิทธิภาพจาก  $K=10$  CV ผลลัพธ์จาก 10-fold CV เผยให้เห็นว่าค่า 'k' ที่เหมาะสมที่สุดแตกต่างกันระหว่างปี โดย  $k=5$  ให้ความแม่นยำสูงสุด (ประมาณ 0.9899) สำหรับข้อมูลปี 2018 ในขณะที่  $k=7$  เหมาะสมที่สุดสำหรับข้อมูลปี 2023 (ประมาณ 0.9900) ความแม่นยำโดยรวมที่ได้จาก 10-fold CV นั้นสูงอย่างสม่ำเสมอ (ประมาณ 99%) ซึ่งบ่งชี้ถึงประสิทธิภาพและความน่าเชื่อถือที่ยอดเยี่ยมของโมเดล kNN สำหรับการจำแนก LULC ทั้งสี่ประเภทโดยใช้ข้อมูล PlanetScope ในพื้นที่ศึกษา นี้ นอกจากนี้ ช่วงความเชื่อมั่น (Confidence Interval) ที่แคบซึ่งได้จากผลลัพธ์ CV ยังสนับสนุนความสอดคล้องของการจำแนกอีกด้วย

Sharma และคณะ [28] งานวิจัยนี้ใช้ข้อมูลจาก Sentinel-2 และแพลตฟอร์ม Google Earth Engine เพื่อจำแนกประเภทสิ่งปกคลุมดินในเขต Jammu ประเทศอินเดีย โดยเปรียบเทียบประสิทธิภาพของอัลกอริธึม Machine Learning หลายแบบ ได้แก่ Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting Trees (GTB) และ Classification and Regression Trees (CART) . ผลการศึกษาแสดงให้เห็นว่า RF มีประสิทธิภาพสูงสุด โดยมีค่า Overall Accuracy เท่ากับ 99.36% และ Kappa Coefficient เท่ากับ 99.11% รองลงมาคือ SVM ที่มีค่า Overall Accuracy เท่ากับ 90.48% และ Kappa Coefficient เท่ากับ 86.70%. ผลลัพธ์เหล่านี้เน้นย้ำถึงความสามารถของ RF ในการจำแนกประเภทสิ่งปกคลุมดินอย่างแม่นยำ

Sim และคณะ [29] งานวิจัยนี้ประเมินผลกระทบของการปรับเปลี่ยนเงื่อนไขการฝึกอบรมของโมเดล Deep Learning ต่อความแม่นยำในการจำแนกประเภทสิ่งปกคลุมดิน โดยใช้ข้อมูลจาก Sentinel-2 และข้อมูลเสริม เช่น ข้อมูลพื้นผิวและข้อมูลเชิงเนื้อผ้า (textural information) . โมเดล U-Net ที่ถูกปรับปรุงด้วยการรวมข้อมูล spectral, textural และ terrain แสดงให้เห็นถึงความแม่นยำที่สูงขึ้น โดยมีค่า Overall Accuracy เท่ากับ 90.3% และ Kappa Coefficient เท่ากับ 0.78. การใช้ฟังก์ชันการสูญเสียแบบผสมและการปรับอัตราการเรียนรู้แบบไดนามิกช่วยลดปัญหา overfitting และเพิ่มความแม่นยำในการจำแนกประเภทสิ่งปกคลุมดิน

Hamdia และคณะ [30] งานนี้ตั้งเป้าพัฒนาแผนที่ LULC รายละเอียดระดับ Level II ของเขตทะเลทรายชอตรูมกลางตุนิเซียซึ่งพื้นที่มีพื้นผิวสูงสะท้อนและลักษณะสเปกตรัมเป็นเนื้อทรายหยาบ ผู้วิจัยดึงภาพ Sentinel-2 L2A 32 ชิน (ปี 2021) ทำ atmospheric-correction แล้วสร้างดัชนี NDBI, BSI, NDVI รวมถึงเมตริก texture (GLCM entropy, variance) ใน scale  $9 \times 9$  pixel

เพื่อขยายข้อมูลเชิงพื้นที่ชุด training-validation 1 320 จุดเก็บด้วย stratified random sampling ระบุ 7 คลาส (bare soil, irrigated cropland, oasis-urban mixed ฯลฯ) ใช้ Random Forest (ntree = 600, mtry = sqrt(p)) และประมวลผลบน Google Earth Engine API ผล Overall Accuracy (OA) 89 %, **K** 0.86 ซึ่งดีกว่า Sentinel-2 band-only  $\approx 8\%$  และ Landsat-8  $\approx 13\%$  Variable importance plot แสดงว่า SWIR-2 (B12) และ B11-B8A texture variance ลด error rate มากสุด Analysis of confusion matrix ชี้ปัญหาหลักคือความสับสน oasis-urban vs. built-up mix แต่ recall ยังสูง 0.81 ข้อจำกัดคือขาดข้อมูล phenology ต่างฤดูกาลและสนามโปรไฟล์ความชื้น ผลงานให้แนวทางแก่หน่วยงานกระทรวงสิ่งแวดล้อมในการจัดทำฐานข้อมูล LULC พื้นที่แห้งแล้งเพื่อวางแผนโครงข่ายชลประทานขนาดเล็ก

Qiu และ Fang [31] งานนี้สืบค้นผลกระทบของขนาด  $k$  ใน  $k$ -fold CV ต่อตัวชี้วัด OA / **K** สำหรับ RF classification winter - wheat (China) ใช้ Sentinel-2 NDVI time-series 12 ช่วง (Jan-Jun) สกัดจาก 80 scene L2A แปลงเป็น feature vector 48 ค่า (mean, std, derivative) RF (ntree = 500, mtry = sqrt(p)) ใช้ training samples 1200 จุดทดลอง  $k = 5, 7, 10, 12, 15$  กราฟ variance-OA แสดงแนวโน้ม inverse relation ;  $k < 7$  variance  $\geq \pm 2\%$ ,  $k > 12$  ค่า OA stable แต่เวลา CV เพิ่ม  $> 40\%$  สรุป  $k = 10$  เป็น trade-off และแนะนำใช้ในงานเกษตรพื้นที่ใหญ่  $> 10000$  h เพื่อลด selection-bias<sup>2</sup>.

Elmotawakkil และคณะ [32] Advanced machine learning models for robust prediction of water quality index and classification เพื่อเปรียบเทียบประสิทธิภาพของ อัลกอริทึม ML ขั้นสูงหลายตัว (ANN, SVM, DT, RF, XGBoost, LSTM) อย่างครอบคลุมและเป็นระบบ สำหรับการทำนายดัชนีคุณภาพน้ำ (Water Quality Index - WQI) และการจำแนกประเภทคุณภาพน้ำ (Water Quality Classification - WQC) โดยมีเป้าหมายเพื่อเติมเต็มช่องว่างในงานวิจัย เดิมด้วยการใช้ชุดข้อมูลที่ทันสมัยและตัวชี้วัดประสิทธิภาพที่หลากหลาย และเพื่อเป็นแนวทางในการเลือกแบบจำลองสำหรับการจัดการคุณภาพน้ำ ใช้ข้อมูลอนุกรมเวลาของคุณภาพน้ำ (เช่น อุณหภูมิ, ออกซิเจนละลายน้ำ (DO), pH) จากสถานีตรวจวัดหลายแห่ง มีการประมวลผลข้อมูลเบื้องต้น เช่น การเติมค่าที่หายไปและการทำมาตรฐานข้อมูล ใช้อัลกอริทึม ANN, RF, DT, SVM, XGBoost สำหรับการจำแนกประเภท (WQC) และ LSTM, XGBoost, DT, RF สำหรับการถดถอย (WQI) มีการปรับจูนไฮเปอร์พารามิเตอร์อย่างเข้มงวดด้วย Grid Search และ 3-fold Cross-Validation (CV) ประเมินผลด้วย Accuracy, Precision, Recall, F1-score (สำหรับ Classification) และ  $R^2$ ,

RMSE, MSE (สำหรับ Regression) วิธีการที่เข้มงวดนี้สะท้อนถึงความสำคัญของระเบียบวิธีวิจัยที่นำเชื่อถือ ผลการวิจัย XGBoost มีประสิทธิภาพสูงสุดในการจำแนกประเภท WQC (Accuracy > 99.9%) ในขณะที่ LSTM เหนือกว่าในการทำนาย WQI ( $R^2=0.9999$ , RMSE/MSE ต่ำสุด) การวิเคราะห์ความสำคัญของคุณลักษณะ (Feature Importance) ชี้ว่าออกซิเจนละลายน้ำ (DO) เป็นปัจจัยที่สำคัญที่สุดอย่างสม่ำเสมอ ผลลัพธ์นี้แสดงให้เห็นแนวโน้มที่อัลกอริทึมขั้นสูงหรือแบบ Ensemble มักมีประสิทธิภาพเหนือกว่าอัลกอริทึมพื้นฐานในงานที่ซับซ้อน ข้อสรุป XGBoost และ LSTM มีประสิทธิภาพเหนือกว่าอัลกอริทึม ML อื่นๆ ที่ทดสอบอย่างมีนัยสำคัญในการทำนายคุณภาพน้ำ แสดงให้เห็นถึงความแม่นยำและความทนทานที่เหมาะสมสำหรับการใช้งานจริง 1 อย่างไรก็ตามงานวิจัยยอมรับข้อจำกัด เช่น โอกาสเกิด Overfitting ในบางแบบจำลอง และความจำเป็นในการวิจัยเพิ่มเติมเกี่ยวกับเทคนิคการประมวลผลข้อมูลล่วงหน้าและสถาปัตยกรรม Deep Learning ที่ซับซ้อนขึ้น 1 สิ่งนี้ตอกย้ำแนวคิดที่ว่าแม้แต่อัลกอริทึมที่ดีที่สุดก็มีข้อจำกัดและบริบทที่มีความสำคัญ

Boateng และคณะ [33] ทบทวนและเปรียบเทียบตัวจำแนกประเภทแบบนอนพาราเมตริก 4 ตัว (KNN, SVM, RF, NN) โดยพิจารณาจากคุณลักษณะต่างๆ เช่น ความทนทาน ความไวต่อการเปลี่ยนแปลง ความเสถียร การจัดการข้อมูลขนาดใหญ่ ความไวต่อสัญญาณรบกวน เวลาในการปรับจูน และความแม่นยำ เพื่อระบุจุดแข็ง/จุดอ่อน วิธีการทบทวนวรรณกรรมจากบทความวิจัย 68 เรื่อง และตำราที่ใช้อัลกอริทึมเหล่านี้ วิเคราะห์เปรียบเทียบตามคุณลักษณะที่กำหนดไว้ ผลการวิจัยพบว่า RF มีความไวต่อการเปลี่ยนแปลงเล็กน้อยในชุดข้อมูลฝึกสอน บางครั้งไม่เสถียรและมีแนวโน้มที่จะ Overfit KNN ทำงานช้าเมื่อข้อมูลมีขนาดใหญ่ และการกำหนดค่า K ที่เหมาะสมทำได้ยาก SVM และ RF ถูกระบุว่าไม่ไวต่อสัญญาณรบกวนหรือการฝึกมากเกินไป (Overtraining) เหมาะสำหรับข้อมูลที่ไม่สมดุล เวลาในการจำแนกประเภทของ NN/KNN เพิ่มขึ้นตามขนาดข้อมูลมากกว่า SVM/RF NN มีประสิทธิภาพสูงแต่ซับซ้อนและใช้เวลาปรับจูนนาน สอดคล้องกับข้อสังเกตเกี่ยวกับความซับซ้อนของ ANN SVM/RF มักถูกแนะนำว่าใช้งานง่ายกว่า เร็วกว่า และแม่นยำ ทุกอัลกอริทึมมีข้อดีข้อเสีย SVM และ RF มักถูกแนะนำเนื่องจากใช้งานง่าย รวดเร็ว และแม่นยำ โดยเฉพาะอย่างยิ่งความทนทานต่อสัญญาณรบกวนและการฝึกมากเกินไป ประสิทธิภาพแตกต่างกันไปตามชุดข้อมูล ไม่มีวิธีใดที่ดีที่สุดสำหรับทุกปัญหา ตอกย้ำว่าประสิทธิภาพขึ้นอยู่กับบริบท

Huang และคณะ [34] ประเมินผลกระทบของขนาดตัวอย่างฝึกสอนและกลยุทธ์การปรับสมดุลข้อมูล (balanced/imbalanced) ต่อความแม่นยำของผลลัพธ์.วิธีการวิจัยใช้ภาพถ่ายดาวเทียม Sentinel-2A ที่ผ่านการประมวลผลแล้วของพื้นที่สามเหลี่ยมปากแม่น้ำแดง ประเทศ

เวียดนาม ซึ่งครอบคลุม 6 ประเภทการใช้ประโยชน์ที่ดิน/สิ่งปกคลุมดิน.สร้างชุดข้อมูลฝึกสอน 14 ชุดที่แตกต่างกัน โดยมีขนาดตัวอย่างและกลยุทธ์การปรับสมดุลข้อมูล (balanced/imbalanced) ที่หลากหลาย ฝึกอัลกอริทึม RF, kNN และ SVM ด้วยพารามิเตอร์ที่ปรับให้เหมาะสมที่สุดสำหรับแต่ละชุดข้อมูล ประเมินประสิทธิภาพโดยใช้ค่าความแม่นยำโดยรวม (Overall Accuracy - OA) เป็นหลัก ผลการวิจัยตัวจำแนกประเภททั้งสามแบบให้ค่า OA สูง อยู่ในช่วง 90% ถึง 95%.โดยทั่วไป SVM ให้ค่า OA สูงสุดและมีความไวต่อขนาดตัวอย่างฝึกสอนน้อยที่สุด ตามมาด้วย RF และ kNN ตามลำดับ เมื่อมีข้อมูลฝึกสอนเพียงพอ (มากกว่า 750 พิกเซลต่อคลาส หรือคิดเป็นประมาณ 0.25% ของพื้นที่ศึกษาทั้งหมด) ทั้งสามอัลกอริทึมแสดงประสิทธิภาพที่ใกล้เคียงกันและสูง ( $OA > 93.85\%$ ) โดยไม่คำนึงถึงความสมดุลของข้อมูล(balanced/imbalanced) สำหรับชุดข้อมูลขนาดใหญ่ที่สมดุล SVM มีความแม่นยำสูงสุด (95.29%) ตามด้วย RF (94.59%) และ kNN (94.10%) อัลกอริทึมทั้ง RF, kNN และ SVM มีประสิทธิภาพสูงสำหรับการจำแนกประเภทการใช้ประโยชน์ที่ดิน/สิ่งปกคลุมดินด้วยภาพถ่ายดาวเทียม Sentinel-2. SVM มีประสิทธิภาพเหนือกว่าเล็กน้อยและมีความไวต่อกว้างกว่าต่อขนาดตัวอย่างฝึกสอนเมื่อเทียบกับ RF และ kNN งานวิจัยแนะนำว่าขนาดตัวอย่างฝึกสอนควรคิดเป็นประมาณ 0.25% ของพื้นที่ศึกษาทั้งหมดเพื่อให้ได้ผลลัพธ์ที่ดี

Nascimento และคณะ [35] เปรียบเทียบอัลกอริทึม ML (kNN, SVM, RF) สำหรับการตรวจจับ Software Aging โดยเฉพาะการทำนายการใช้ RAM จนหมดใน SQL Server เพื่อระบุอัลกอริทึมที่แม่นยำที่สุด ประยุกต์ใช้ในวิศวกรรมซอฟต์แวร์/ความน่าเชื่อถือของระบบ ใช้ชุดข้อมูลเกี่ยวกับการใช้ RAM ของ SQL Server ใช้อัลกอริทึม kNN, SVM และ RF ประเมินประสิทธิภาพโดยใช้ MAE, RMSE และ R2 เน้นตัวชี้วัดการถดถอยสำหรับการทำนายเวลาจนกว่าทรัพยากรจะหมด บทคัดย่อ ระบุว่ามีการเปรียบเทียบ แต่ไม่ได้ระบุว่าอัลกอริทึมใดดีที่สุด จำเป็นต้องดูเนื้อหาฉบับเต็มเพื่อผลลัพธ์เฉพาะ อย่างไรก็ตาม อธิบายอัลกอริทึม kNN ใช้ค่าเฉลี่ยของเพื่อนบ้านสำหรับการถดถอย, SVM ใช้ฟังก์ชันการสูญเสีย (Loss functions) และ Kernel เฉพาะ, RF ใช้ค่าเฉลี่ยการทำนายจากต้นไม้ตัดสินใจหลายต้น ข้อสรุป งานวิจัยระบุอัลกอริทึมที่แม่นยำที่สุดในบรรดา kNN, SVM และ RF สำหรับการทำนายการใช้ RAM จนหมด โดยพิจารณาจากตัวชี้วัด MAE, RMSE, R2

Patel และคณะ ประเมิน DT, SVM, RF และ ANN สำหรับจำแนก LULC โดยใช้ RS และ GIS ANN แม่นยำที่สุดในพื้นที่ซับซ้อน RF และ SVM ดีในพื้นที่ขอบเขตชัดเจน DT ให้ผลต่ำที่สุด งานวิจัยนี้ชี้ว่าโมเดลต้องเลือกตามลักษณะข้อมูล

Kasahun และ Legesse [36] เมือง Dilla ประเทศเอธิโอเปีย งานวิจัยนี้เปรียบเทียบโดยตรงระหว่าง ANN, RF และ SVM สำหรับการหาแผนที่ LULC ในเขตเมืองโดยใช้ภาพความ

ละเอียดสูง ซึ่งตอบสนองความต้องการข้อมูลที่ถูกต้องการสำหรับการวางผังเมืองในบริบทของประเทศกำลังพัฒนาวัตถุประสงค์เพื่อประเมินและระบุตัวจำแนก ML ที่เหมาะสมที่สุด (ANN, SVM, RF) สำหรับการจำแนก LULC และการตรวจจับการเปลี่ยนแปลงในเมือง Dilla ประเทศเอธิโอเปีย เพื่อสนับสนุนหน่วยงานวางแผนอัลกอริทึมที่เปรียบเทียบ Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF) ภาพถ่ายดาวเทียมความละเอียดสูงจาก Google Earth (ปี 2006, 2014, 2022) พื้นที่ศึกษาเมือง Dilla เขต Gedeo ทางตอนใต้ของประเทศเอธิโอเปีย ผลการศึกษาเชิงเปรียบเทียบที่สำคัญ RF มีประสิทธิภาพเหนือกว่า SVM และ ANN อย่างมีนัยสำคัญในทุกตัวชี้วัด ความแม่นยำโดยรวม (OA) (RF 0.97, SVM 0.96, ANN 0.89), ค่าสัมประสิทธิ์ Kappa (RF 0.98, SVM 0.97, ANN 0.81), ความแม่นยำของผู้ผลิต (Producer's Accuracy) (RF 0.99, SVM 0.98, ANN 0.94), ความแม่นยำของผู้ใช้ (User's Accuracy) (RF 0.97, SVM 0.97, ANN 0.88) RF แสดงให้เห็นถึงประสิทธิภาพเชิงการคำนวณที่เหนือกว่า (เวลาฝึก 6.33 วินาที เทียบกับ 15 วินาทีสำหรับ SVM และ 30 วินาทีสำหรับ ANN) การประเมินคุณภาพทางสายตา (Visual assessment) ยืนยันว่า RF สร้างแผนที่ที่มีความแม่นยำสูงกว่าและขอบเขตระหว่างประเภท LULC ชัดเจนกว่า ผลการตรวจจับการเปลี่ยนแปลงพื้นที่สิ่งปลูกสร้างของ RF สอดคล้องกับข้อมูลบันทึกอย่างเป็นทางการ ซึ่งแตกต่างจาก ANN และ SVM งานวิจัยนี้เน้นย้ำถึงความแข็งแกร่งและประสิทธิภาพของ RF โดยเฉพาะอย่างยิ่งในสภาพแวดล้อมเมืองที่ซับซ้อนซึ่งใช้ข้อมูลความละเอียดสูง แสดงให้เห็นถึงความเหมาะสมสำหรับการทำแผนที่เชิงปฏิบัติการที่ต้องการทั้งความแม่นยำและความเร็ว ความเหนือกว่าอย่างสม่ำเสมอของ RF ในหลายๆ ด้าน (ความแม่นยำ, Kappa, ความแม่นยำของผู้ผลิต/ผู้ใช้, ความเร็ว, คุณภาพทางสายตา และความสอดคล้องกับการเปลี่ยนแปลงจริง) บ่งชี้ถึงความสามารถในการจัดการข้อมูลและการจัดการการจำแนกที่เฉพาะเจาะจงในงานวิจัยนี้ได้ดีกว่า ANN และ SVM

Bhungeni, Ramjatan, และ Gebreslasie [37] กลุ่มน้ำ uMngeni ประเทศแอฟริกาใต้ งานวิจัยนี้ทำการทำแผนที่ LULC ในพื้นที่ลุ่มน้ำสำคัญโดยใช้ข้อมูล Landsat 8 และเปรียบเทียบอัลกอริทึม MLA แบบ non-parametric ที่ใช้ pixel-based จำนวน 4 ตัว รวมถึง Naïve Bayes ซึ่งไม่ค่อยมีการเปรียบเทียบมากนัก วัตถุประสงค์ เพื่อประเมินประสิทธิภาพของ Naïve Bayes (NB), Random Forests (RF), Support Vector Machines (SVM), และ Artificial Neural Networks (ANN) สำหรับการทำแผนที่ LULC ในลุ่มน้ำ uMngeni ประเทศแอฟริกาใต้ โดยใช้ภาพ Landsat 8 OLI เพื่อเป็นข้อมูลสำหรับการจัดการทรัพยากรน้ำ อัลกอริทึมที่เปรียบเทียบ Naïve Bayes (NB),

Random Forests (RF), Support Vector Machines (SVM), Artificial Neural Networks (ANN) ข้อมูลดาวเทียม Landsat 8 Operational Land Imager (L8-OLI) (ความละเอียด 30 เมตร) พื้นที่ศึกษาลุ่มน้ำ uMngeni จังหวัด KwaZulu-Natal ประเทศแอฟริกาใต้ ผลการศึกษาเชิงเปรียบเทียบที่สำคัญ RF ให้ความแม่นยำสูงสุด (OA=97.02%, Kappa=0.96) ซึ่งสูงกว่า SVM (OA=89.74%, Kappa=0.88) และ ANN (OA=87%, Kappa=0.86) อย่างมีนัยสำคัญ NB มีประสิทธิภาพต่ำ (OA=68.64%, Kappa=0.58) โดยมีปัญหาในการแยกแยะประเภท LULC และเกิด overfitting RF แสดงให้เห็นถึงความแข็งแกร่ง โดยเฉพาะอย่างยิ่งสำหรับการทำแผนที่ประเภท LULC ที่มีพื้นที่ขนาดเล็ก (พื้นที่โล่ง, เขื่อน, พื้นที่ชุ่มน้ำ) SVM และ ANN แสดงจุดอ่อนเฉพาะ SVM มีข้อผิดพลาดในการละเว้น (omission error) สำหรับพิกเซลเขื่อน/พื้นที่ชุ่มน้ำ; ANN เกิด overfitting ในประเภทเขื่อน/พื้นที่ชุ่มน้ำ และมีปัญหาในการจำแนกแหล่งน้ำขนาดเล็ก ผลการศึกษานี้ตอกย้ำความเหนือกว่าโดยทั่วไปของ RF แต่ยืนยันถึงความสำคัญของการประเมินประสิทธิภาพในประเภท LULC ที่เป็นส่วนน้อย (minority classes) ซึ่ง RF แสดงความแข็งแกร่งเป็นพิเศษเมื่อเทียบกับ SVM และ ANN ในกรณีนี้ ประสิทธิภาพที่ต่ำของ NB ชี้ให้เห็นถึงข้อจำกัดสำหรับงาน LULC ที่ซับซ้อนซึ่งมีคุณลักษณะ (features) ที่อาจมีความสัมพันธ์กัน ช่องว่างความแม่นยำ ที่มา: ระหว่าง RF กับอัลกอริทึมอื่นๆ (SVM, ANN, NB) เน้นย้ำถึงประสิทธิภาพของ RF ความสำเร็จเฉพาะกับประเภท LULC ส่วนน้อยชี้ให้เห็นถึงความสามารถในการจัดการชุดข้อมูลที่ไม่สมดุลหรือจับลักษณะสเปกตรัมที่ละเอียดอ่อนได้ดีกว่าอัลกอริทึมอื่นที่ทดสอบในการศึกษานี้ซึ่งใช้ข้อมูล Landsat ความล้มเหลวของ NB น่าจะเกิดจากสมมติฐานหลักเกี่ยวกับความเป็นอิสระของคุณลักษณะ ซึ่งไม่ค่อยเป็นจริงในข้อมูลสเปกตรัม

Ouma และคณะ [38] เมือง Gaborone ประเทศบอตสวานา งานวิจัยนี้เปรียบเทียบอัลกอริทึม ML สี่ตัว (รวมถึง GTB และ MLP-ANN ควบคู่ไปกับ RF และ SVM) ในช่วงเวลายาวนาน (1984-2020) โดยใช้ข้อมูล Landsat ในพื้นที่ที่มีการขยายตัวของเมือง และมีการสำรวจการผสมผสานผลลัพธ์หลังการจำแนก (post-classification fusion) ซึ่งเป็นแนวทางที่ไม่เหมือนใคร วัตถุประสงค์ เพื่อนำไปใช้และประเมินความแม่นยำของ Gradient Tree Boosting (GTB), Random Forest (RF), Support Vector Machine (SVM), และ Multilayer Perceptron Neural Networks (MLP-ANN) สำหรับการทำแผนที่ LULC ในเขตเมืองในช่วง 35 ปี ในพื้นที่วางแผน Greater Gaborone (GGPA) ประเทศบอตสวานา และเพื่อปรับปรุงความแม่นยำผ่านการผสมผสานผลลัพธ์หลังการจำแนก อัลกอริทึมที่เปรียบเทียบ Gradient Tree Boosting (GTB), Random

Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron Neural Networks (MLP-ANN) ข้อมูลดาวเทียม ชุดข้อมูล Landsat หลายเซ็นเซอร์ (L4-MSS, L5-TM, L7-ETM+, L8-OLI) ตั้งแต่ปี 1984-2020 (ความละเอียด 30 เมตร, MSS ถูก resample) พื้นที่ศึกษา พื้นที่วางแผน Greater Gaborone (GGPA) ประเทศบอตสวานา ผลการศึกษาเชิงเปรียบเทียบที่สำคัญ จากค่า OA เฉลี่ยตลอดช่วงเวลาศึกษา RF มีประสิทธิภาพดีที่สุด (92.8%) ตามด้วย MLP-ANN (91.2%), SVM (90.9%), และ GTB (87.8%) ประสิทธิภาพแตกต่างกันไปตามประเภท LULC MLP-ANN เหมาะสำหรับพื้นที่สิ่งปลูกสร้าง/แหล่งน้ำ, RF เหมาะสำหรับพื้นที่ไม้พุ่ม/ป่าไม้ แนวทางการผสมผสานหลังการจำแนกแบบ FEI-FEO ที่แปลกใหม่ช่วยปรับปรุงความแม่นยำโดยรวมของการจำแนก LULC ได้มากกว่า 2% โดยการรวมจุดแข็งของตัวจำแนกต่างๆ แสดงให้เห็นถึงความสามารถในการประยุกต์ใช้ตัวจำแนก ML ในระยะยาว และยืนยันความแข็งแกร่งโดยทั่วไปของ RF ตลอดหลายทศวรรษของข้อมูล Landsat ที่สำคัญคือ แสดงให้เห็นว่าอัลกอริทึมที่แตกต่างกันอาจมีความสามารถโดดเด่นในประเภท LULC ที่แตกต่างกัน ซึ่งชี้ให้เห็นว่าแนวทางแบบผสมผสาน (hybrid) หรือการหลอมรวม (fusion) สามารถให้ผลลัพธ์ที่เหนือกว่าการพึ่งพาตัวจำแนก "ที่ดีที่สุด" เพียงตัวเดียว การจัดอันดับที่สอดคล้องกันตลอดช่วงเวลาที่ยาวนานช่วยเพิ่มน้ำหนักให้กับความน่าเชื่อถือโดยทั่วไปของ RF ความแตกต่างของประสิทธิภาพเฉพาะประเภท LULC เน้นย้ำว่าความแม่นยำโดยรวมไม่ได้บอกเล่าเรื่องราวทั้งหมด อัลกอริทึมอาจดีที่สุดโดยรวมแต่อาจไม่เหมาะสมที่สุดสำหรับประเภท LULC ที่สนใจ โดยเฉพาะ ความสำเร็จของวิธีการหลอมรวมพิสูจน์โดยตรงว่าการใช้ประโยชน์จากจุดแข็งของหลายอัลกอริทึมสามารถเอาชนะจุดอ่อนของแต่ละอัลกอริทึมได้ นำไปสู่แผนที่ขั้นสุดท้ายที่แม่นยำยิ่งขึ้น นี้แสดงถึงเส้นทางที่เป็นไปได้ นอกเหนือจากการเลือกผู้ชนะเพียงรายเดียว

Talukdar และคณะ [39] บทความปริทัศน์ บทความนี้เป็นการทบทวนวรรณกรรมที่มีคุณค่า โดยสังเคราะห์ผลการศึกษางานวิจัยจำนวนมากที่เปรียบเทียบตัวจำแนก ML สำหรับการจำแนก LULC โดยใช้การสังเกตการณ์จากดาวเทียม วัตถุประสงค์ เพื่อทบทวนและตรวจสอบความแม่นยำของอัลกอริทึม ML ต่าง ๆ (RF, SVM, ANN, Fuzzy ARTMAP, SAM และ MD) สำหรับการจำแนก LULC จากการศึกษาสังเกตการณ์ด้วยดาวเทียม โดยมีเป้าหมายเพื่อระบุตัวจำแนกที่ดีที่สุดโดยรวม อัลกอริทึมที่เปรียบเทียบ (ในวรรณกรรมที่ทบทวน) ส่วนใหญ่เป็น Random Forest (RF), Support Vector Machine (SVM) และ Artificial Neural Network (ANN) แต่ยังคงกล่าวถึง Decision Tree (DT), Fuzzy ARTMAP, Spectral Angle Mapper (SAM), Mahalanobis Distance (MD), K-Nearest Neighbor (KNN), Classification and Regression Trees (CART) และ Gradient Tree

Boost (GTB) ข้อมูลดาวเทียม (ที่กล่าวถึงในวรรณกรรมที่ทบทวน) ส่วนใหญ่เป็น Landsat, Sentinel-2 และอื่นๆ โดยนัยยะ พื้นที่ศึกษา ไม่เฉพาะเจาะจงพื้นที่เดียว ทบทวนงานวิจัยจากบริบทต่าง ๆ ทั่วโลก รวมถึงตัวอย่างการประยุกต์ใช้ Landsat 8 สำหรับแม่น้ำคงคา ประเทศอินเดีย ผลการศึกษาเชิงเปรียบเทียบที่สำคัญ (สังเคราะห์จากวรรณกรรม) RF และ ANN มักถูกอ้างถึงว่าเป็นตัวจำแนก LULC ที่ดีที่สุดบ่อยครั้ง SVM และ RF มักถูกพิจารณาว่าเป็นเทคนิคชั้นนำโดยรวม RF โดยทั่วไปแสดงความแม่นยำสูงสุด SVM ยังแสดงความแม่นยำสูง ซึ่งมักจะเทียบเท่ากับ RF อัลกอริทึมเช่น SAM และ MD (ซึ่งเป็นแบบดั้งเดิมหรือเรียบง่ายกว่า) มีแนวโน้มที่จะให้ความแม่นยำต่ำกว่า ประสิทธิภาพขึ้นอยู่กับลักษณะของเซ็นเซอร์ คุณสมบัติของข้อมูล (ความละเอียด) เครื่องมือประมวลผล และสภาพพื้นที่ศึกษาอย่างมาก บทความนี้ยืนยันฉันทามติทั่วไปจากงานวิจัยแต่ละชิ้น RF และ SVM มักเป็นตัวจำแนก ML ที่แม่นยำและน่าเชื่อถือที่สุดสำหรับการทำแผนที่ LULC ซึ่งมักมีประสิทธิภาพเหนือกว่า ANN และวิธีการแบบดั้งเดิม อย่างไรก็ตาม มันตอกย้ำอย่างยิ่งถึงข้อสังเกตที่สำคัญว่าประสิทธิภาพขึ้นอยู่กับบริบท และไม่มีอัลกอริทึมใดที่ดีที่สุดในทุกสถานการณ์ การรวบรวมผลการศึกษาจากหลายงานวิจัย ทำให้การทบทวนนี้เป็นการยืนยันแนวโน้มที่สังเกตได้จากกรณีศึกษาเดียวในระดับที่สูงขึ้น การกล่าวถึง RF และ SVM อย่างสม่ำเสมอว่าเป็นผู้มีประสิทธิภาพสูงสุดในงานวิจัยต่างๆ ช่วยเสริมความแข็งแกร่งให้กับข้อค้นพบนี้ ในขณะเดียวกัน การระบุอย่างชัดเจนเกี่ยวกับความแปรผันของประสิทธิภาพที่ขึ้นอยู่กับปัจจัยต่าง ๆ เน้นย้ำถึงข้อควรระวังที่สำคัญในการประกาศผู้ชนะสากล และเน้นย้ำถึงความจำเป็นในการเลือกและทดสอบอย่างรอบคอบในสถานการณ์การใช้งานเฉพาะ

Ghayour และคณะ [40] การเปรียบเทียบ Sentinel-2 กับ Landsat 8 งานวิจัยนี้เปรียบเทียบประสิทธิภาพของหลายอัลกอริทึม (SVM, ANN, MLC, MD, MH) โดยตรงโดยใช้ทั้งข้อมูล Sentinel-2 และ Landsat 8 สำหรับพื้นที่เดียวกัน (ไม่ได้ระบุ) ทำให้สามารถประเมินผลกระทบทั้งจากอัลกอริทึมและแหล่งข้อมูลได้ นอกจากนี้ยังตรวจสอบการปรับจูนพารามิเตอร์ของ SVM วัตถุประสงค์เพื่อประเมินและเปรียบเทียบประสิทธิภาพของอัลกอริทึม SVM, ANN, MLC, MD และ MH สำหรับการทำให้แผนที่ LULC โดยใช้ข้อมูลทั้งจาก Sentinel-2 และ Landsat 8, ตรวจสอบผลกระทบของพารามิเตอร์ penalty ของ SVM และระบุการผสมผสานที่ดีที่สุด อัลกอริทึมที่เปรียบเทียบ Support Vector Machine (SVM), Artificial Neural Network (ANN), Maximum Likelihood Classification (MLC), Minimum Distance (MD), Mahalanobis (MH) ข้อมูลดาวเทียม Sentinel-2 และ Landsat 8 พื้นที่ศึกษา ไม่ได้ระบุ เน้นการเปรียบเทียบวิธีการ ผล

การศึกษาเชิงเปรียบเทียบที่สำคัญ SVM ให้ค่า OA สูงสุด (เฉลี่ยรวม 94%, 95.82% สำหรับ S2, 94.78% สำหรับ L8) ด้วยการปรับจูนที่เหมาะสม วิธีการแบบพารามेटริก (MD, MLC) ให้ความแม่นยำต่ำกว่า โดยเฉพาะ MLC กับ Landsat 8 (74.68%) ประสิทธิภาพของ ANN ไรต์ต่อจำนวนชั้นซ่อน (hidden layers); การเพิ่มชั้นลดความแม่นยำลงอย่างมาก OA ที่ดีที่สุดของ ANN 88.46% (S2), 91.31% (L8) ประสิทธิภาพของ SVM ขึ้นอยู่กับฟังก์ชันเคอร์เนลและพารามิเตอร์ penalty อย่างมาก; เคอร์เนลเชิงเส้น (linear kernel) ที่มีค่า penalty เฉพาะ (150 สำหรับ S2, 200 สำหรับ L8) ให้ผลดีที่สุด ข้อมูล Sentinel-2 ให้ความแม่นยำโดยรวมสูงกว่าเล็กน้อยเมื่อเทียบกับ Landsat 8 สำหรับอัลกอริทึมที่มีประสิทธิภาพสูงสุด (SVM) แต่ผลนี้ไม่สอดคล้องกันในทุกอัลกอริทึม (ANN, MD, MH บางครั้งดีกว่าด้วย L8) งานวิจัยนี้เน้นย้ำอย่างยิ่งถึงบทบาทสำคัญของการปรับค่าพารามิเตอร์ให้เหมาะสม (hyperparameter optimization) โดยเฉพาะสำหรับ SVM และ ANN แสดงให้เห็นว่า SVM ที่ปรับจูนอย่างดีสามารถทำงานได้ดีกว่าวิธีอื่น ๆ รวมถึง ANN นอกจากนี้ยังให้หลักฐานเชิงประจักษ์ว่าแม้ Sentinel-2 อาจมีข้อได้เปรียบเล็กน้อยเหนือ Landsat 8 (น่าจะเนื่องมาจากความละเอียด/แถบคลื่น) แต่การเลือกอัลกอริทึมและการปรับจูนอาจมีอิทธิพลมากกว่าแหล่งข้อมูลเพียงอย่างเดียว การวิเคราะห์โดยละเอียดเกี่ยวกับเคอร์เนล/ค่า penalty ของ SVM และชั้นซ่อนของ ANN เชื่อมโยงการเลือกพารามิเตอร์กับการเปลี่ยนแปลงความแม่นยำอย่างมีนัยสำคัญโดยตรง สิ่งนี้แสดงให้เห็นว่าการใช้อัลกอริทึมแบบ "สำเร็จรูป" นั้นไม่เพียงพอ ข้อค้นพบที่ว่า SVM มีประสิทธิภาพเหนือกว่า ANN ชัดแย้งกับงานวิจัยอื่น ๆ บางชิ้น (เช่น บางครั้งจัดอันดับ ANN สูงกว่า) แต่สอดคล้องกับงานวิจัยอื่น ๆ ซึ่งตอกย้ำประเด็นเรื่องความแปรปรวน การเปรียบเทียบที่ละเอียดอ่อนของ S2 และ L8 แสดงให้เห็นว่าการเลือกเซ็นเซอร์ไม่ใช่การตัดสินใจแบบ "ดีกว่า/แยกว่า" อย่างง่ายๆ ปฏิสัมพันธ์กับอัลกอริทึมมีความสำคัญ

Khunrattanasiri และคณะ [41] เขตพัฒนาพิเศษภาคตะวันออก (EEC) ประเทศไทย งานวิจัยนี้เป็นการเปรียบเทียบที่เพิ่งดำเนินการในบริบทเฉพาะของประเทศไทย (เขตพัฒนาพิเศษภาคตะวันออก - EEC) โดยเปรียบเทียบการแปลตีความด้วยสายตา (visual interpretation) กับ RF และตัวจำแนกแบบดั้งเดิม Maximum Likelihood โดยใช้ข้อมูล Landsat 9 ที่ใหม่กว่า วัตถุประสงค์เพื่อเปรียบเทียบผลการจำแนกการใช้ประโยชน์ที่ดินจากข้อมูลดาวเทียม Landsat 9 OLI-2 โดยใช้วิธีการแปลตีความด้วยสายตา, การจำแนกแบบกำกับดูแล (supervised classification) ด้วยวิธี Maximum Likelihood (MLC) และเทคนิคการเรียนรู้ของเครื่อง Random Forest (RF) ในพื้นที่ EEC ประเทศไทย วิธีการที่เปรียบเทียบ การแปลตีความด้วยสายตา (Visual Interpretation),

Random Forest (RF), Maximum Likelihood Classifier (MLC) ข้อมูลดาวเทียม Landsat 9 OLI-2 (ความละเอียด 30 เมตร) พื้นที่ศึกษา เขตพัฒนาพิเศษภาคตะวันออก (EEC) ประเทศไทย ผลการศึกษาเชิงเปรียบเทียบที่สำคัญ การแปลตีความด้วยสายตาให้ความแม่นยำโดยรวมสูงสุด (97.30%) ในบรรดาวิธีการอัตโนมัติ RF (OA=67.57%) มีประสิทธิภาพเหนือกว่า MLC (OA=64.86%) อย่างมีนัยสำคัญ เมื่อเปรียบเทียบกับเกณฑ์มาตรฐานการแปลตีความด้วยสายตา RF จำแนกพื้นที่ได้ถูกต้องมากกว่า (63.26% ของพื้นที่ EEC) ในทุกประเภทการใช้ประโยชน์ที่ดินเมื่อเทียบกับ MLC แม้ว่าการแปลตีความด้วยสายตาโดยผู้เชี่ยวชาญจะยังคงมีความแม่นยำสูง (แต่ใช้เวลานานและอาจขึ้นอยู่กับดุลยพินิจ) งานวิจัยนี้ยืนยันในบริบทของประเทศไทยว่า RF ให้ผลลัพธ์ที่ดีกว่าวิธี MLC แบบดั้งเดิมอย่างมากสำหรับการจำแนกอัตโนมัติโดยใช้ข้อมูล Landsat ล่าสุด ค่า OA ที่ค่อนข้างต่ำของ RF เมื่อเทียบกับงานวิจัยอื่น ๆ อาจบ่งชี้ถึงความท้าทายเฉพาะในภูมิภาคของ EEC, ลักษณะข้อมูล Landsat 9, หรือรายละเอียดการดำเนินงาน (เช่น ข้อมูลฝึกสอน) การเปรียบเทียบโดยตรงแสดงให้เห็นข้อได้เปรียบที่ชัดเจนของ RF เหนือ MLC ซึ่งสอดคล้องกับแนวโน้มที่กว้างขึ้นที่ระบุไว้ในวรรณกรรม ความแม่นยำสูงของการแปลตีความด้วยสายตาทำหน้าที่เป็นเกณฑ์มาตรฐานที่มีประโยชน์ แม้ว่าอาจไม่สามารถปฏิบัติได้จริงในระดับใหญ่ ความคลาดเคลื่อนระหว่างความแม่นยำของ RF ที่นี้กับในงานวิจัยอื่น ๆ (เช่น >90% ใน ) เน้นย้ำถึงปัจจัยที่ขึ้นอยู่กับบริบท ซึ่งชี้ให้เห็นว่าแม้แต่อัลกอริทึมที่แข็งแกร่งอย่าง RF ก็ต้องการการประยุกต์ใช้อย่างระมัดระวังและอาจเผชิญกับความท้าทายในสภาพแวดล้อมบางอย่างหรือกับชุดข้อมูล/ขั้นตอนการฝึกสอนที่เฉพาะเจาะจง



## บทที่ 3

### ขั้นตอนการดำเนินงาน

บทนี้อธิบายถึงขั้นตอนและวิธีการที่ใช้ในการดำเนินงาน วิจัยเรื่องการจำแนกสิ่งปกคลุมดินจากข้อมูลภาพถ่ายดาวเทียม Sentinel -2 ด้วยเทคนิคการเรียนรู้ของเครื่อง โดยครอบคลุมตั้งแต่การกำหนดขอบเขตพื้นที่ศึกษา การรวบรวมข้อมูล การเตรียมข้อมูล การสร้างแบบจำลองการเรียนรู้ของเครื่อง และการประเมินความถูกต้องของผลการจำแนก รายละเอียดของการดำเนินงานมีดังต่อไปนี้

#### 3.1 ข้อมูลที่ใช้ศึกษา

การวิจัยครั้งนี้ใช้ข้อมูลหลัก 2 ประเภท ดังนี้

1. **ข้อมูลภาพถ่ายดาวเทียม** ภาพถ่ายจากดาวเทียม Sentinel-2 ซึ่งบันทึกภาพพื้นที่ศึกษาครอบคลุมอำเภอเมืองร้อยเอ็ด จังหวัดร้อยเอ็ด โดยเลือกภาพถ่ายที่ปราศจากเมฆปกคลุมมากที่สุด ณ วันที่ 28 ธันวาคม พ.ศ. 2567 ข้อมูล Sentinel-2 มีความละเอียดเชิงพื้นที่ตั้งแต่ 10 เมตร และมีหลายช่วงคลื่น (Multispectral) ซึ่งเหมาะสำหรับการวิเคราะห์และจำแนกสิ่งปกคลุมดินประเภทต่าง ๆ
2. **ข้อมูลอ้างอิงภาคพื้นดิน**
  - ข้อมูลภาพความละเอียดสูงจาก Google Earth เพื่อใช้ในการกำหนดพื้นที่ตัวอย่างสำหรับการฝึกสอน (Training Area) และการตรวจสอบความถูกต้องเบื้องต้น
  - ข้อมูลจากการสำรวจภาคสนาม (Field Survey) ในบางพื้นที่ตัวอย่าง เพื่อยืนยันประเภทสิ่งปกคลุมดินที่แท้จริง และใช้เป็นข้อมูลอ้างอิง (Ground Truth) สำหรับการประเมินความแม่นยำขั้นสุดท้าย

#### เครื่องมือและซอฟต์แวร์

- โปรแกรมระบบสารสนเทศภูมิศาสตร์ (GIS Software) เช่น QGIS หรือ ArcGIS สำหรับการจัดการข้อมูลเชิงพื้นที่ การประมวลผลภาพถ่ายดาวเทียม และการสร้างแผนที่
- สภาพแวดล้อมการเขียนโปรแกรมภาษา Python พร้อมไลบรารีสำหรับการเรียนรู้ของเครื่อง เช่น Scikit-learn, TensorFlow หรือ Keras สำหรับการสร้างและประเมินแบบจำลอง

## 3.2 การเตรียมข้อมูล

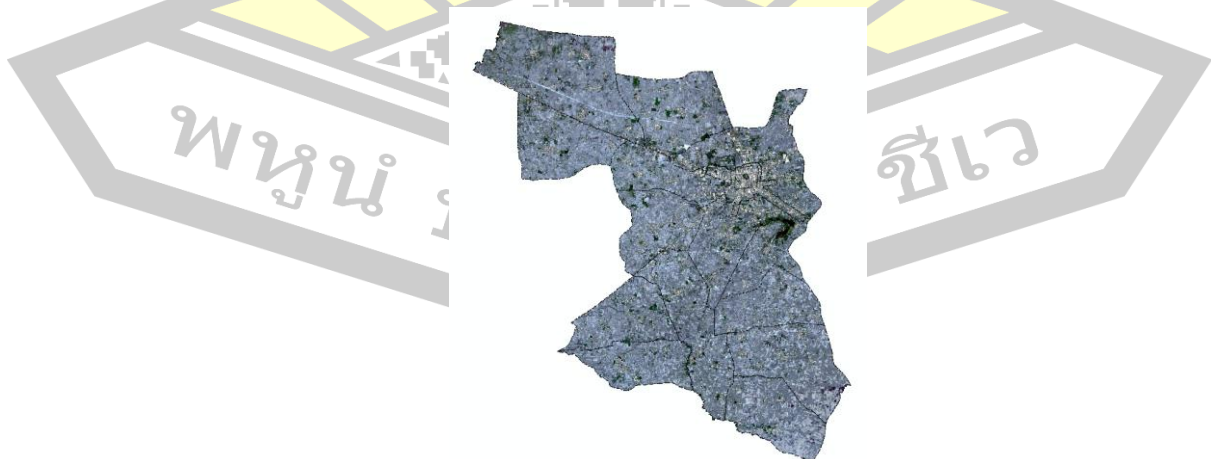
3.2.1 ดาวเทียมถ่ายภาพจากดาวเทียม Sentinel -2 สืบค้นได้จากเว็บไซต์ <https://browser.dataspace.copernicus.eu/> ซึ่งได้เลือกข้อมูลดาวเทียมที่ปราศจากเมฆมากที่สุดของบริเวณที่ศึกษา โดยเลือกข้อมูลภาพ วันที่ 28 ธันวาคม พ.ศ. 2567 สำหรับหน้าเว็บไซต์ได้แสดงออกมาดังภาพประกอบที่ 3.1



ภาพประกอบที่ 3.1 ดาวเทียมถ่ายภาพจากดาวเทียม Sentinel -2

3.2.2 การปรับแก้เชิงเรขาคณิต (Geometric Correction) ตรวจสอบและปรับแก้ความคลาดเคลื่อนทางเรขาคณิตของภาพถ่ายดาวเทียม (หากจำเป็น) เพื่อให้ตำแหน่งพิกัดบนภาพมีความถูกต้องตรงกับตำแหน่งจริงบนพื้นผิวโลก และสามารถซ้อนทับกับข้อมูลเชิงพื้นที่อื่น ๆ ได้อย่างแม่นยำ

3.2.3 การตัดขอบเขตพื้นที่ศึกษา (Subset Image) ตัดข้อมูลภาพถ่ายดาวเทียมให้เหลือเฉพาะขอบเขตการปกครองของอำเภอเมืองร้อยเอ็ด จังหวัดร้อยเอ็ด ซึ่งเป็นพื้นที่เป้าหมายของการศึกษา แสดงดังภาพประกอบที่ 3.2



ภาพประกอบที่ 3.2 แสดงพื้นที่ศึกษา

3.2.4 สร้างเวกเตอร์ชั้นใหม่ (Layer) เพื่อใช้กำหนดพื้นที่ฝึกสอน (Training Areas) ในการสร้างและประเมินแบบจำลอง ได้มีการรวบรวมจุดตัวอย่างสำหรับสิ่งปกคลุมดินแต่ละประเภท ทั้งจากข้อมูลภาพถ่ายดาวเทียม Sentinel-2 เพื่อใช้ในการฝึกสอน (Training) และจาก Google Maps/การสำรวจภาคสนาม เพื่อใช้ในการตรวจสอบความถูกต้อง (Validation) โดยมีจำนวนจุดตัวอย่างดังตารางที่ 3.1

**ตารางที่ 3.1** กลุ่มตัวอย่างที่ใช้ในการฝึกสอนและตรวจสอบความถูกต้อง

ประเภทข้อมูล สิ่งปกคลุมดิน	จำนวนจุดตัวอย่าง Sentinel-2 (Training)	จำนวนจุดตัวอย่าง Google Map/ ภาคสนาม (Validation)
พื้นที่เกษตรกรรม (A)	30	44
เมืองและย่านการค้า	30	60
พื้นที่ส่วนราชการ	30	43
พื้นที่ว่างเปล่า	30	43
พื้นที่อยู่อาศัย	30	44
พื้นที่ป่าไม้	30	44
พื้นที่แหล่งน้ำ	30	40
พื้นที่สวนสาธารณะ	30	20
รวม	240	350

**หมายเหตุ** การกำหนดจำนวนจุดตัวอย่างสำหรับการตรวจสอบความถูกต้อง (Validation) ในแต่ละประเภทสิ่งปกคลุมดินนี้ พิจารณาจากความหลากหลายเชิงพื้นที่และความสำคัญของประเภทข้อมูลนั้น ๆ เช่น เมืองและย่านการค้ามีจำนวนจุดมากกว่า (60 จุด) เนื่องจากเป็นพื้นที่ที่มีการใช้ประโยชน์หลากหลายและเปลี่ยนแปลงเร็ว ขณะที่พื้นที่สวนสาธารณะมีจำนวนจุดน้อยกว่า (20 จุด) เพราะมีขนาดและการกระจายที่จำกัด การกระจายจุดลักษณะนี้ช่วยให้ได้ข้อมูลที่ครอบคลุมและสอดคล้องกับลักษณะจริงของพื้นที่ศึกษามากยิ่งขึ้น

### 3.3 การสร้างแบบจำลอง (Model Building)

การวิจัยนี้ใช้เทคนิคการเรียนรู้ของเครื่องแบบกำกับดูแล (Supervised Learning) จำนวน 4 เทคนิค เพื่อจำแนกสิ่งปกคลุมดินจากข้อมูลภาพถ่ายดาวเทียม Sentinel-2 โดยมีรายละเอียดของแต่ละแบบจำลองดังนี้

**3.3.1 ต้นไม้ตัดสินใจ (Decision Tree - DT)** สร้างแบบจำลองที่ใช้โครงสร้างการตัดสินใจแบบลำดับขั้นเพื่อจำแนกข้อมูลตามคุณลักษณะเชิงสเปกตรัม

- เกณฑ์การแบ่งโหนด (Criterion) Gini Index

— ความลึกสูงสุด (Max Depth) กำหนดค่าเป็น 10 เพื่อควบคุมความซับซ้อนและป้องกัน Overfitting

3.3.2 ป่าสุ่ม (Random Forest - RF) เป็นเทคนิค Ensemble ที่สร้างต้นไม้ตัดสินใจหลาย ๆ ต้น ( $n_{\text{estimators}}$ ) จากชุดข้อมูลย่อยแบบสุ่ม และรวมผลการทำนายเพื่อเพิ่มความแม่นยำและลดความแปรปรวน

- จำนวนต้นไม้ ( $n_{\text{estimators}}$ ) 100 ต้น
- ความลึกสูงสุด (Max Depth) กำหนดค่าเป็น 20

3.3.3 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine - SVM) เทคนิคที่พยายามหาเส้นแบ่ง (Hyperplane) ที่ดีที่สุดเพื่อแยกข้อมูลระหว่างคลาสต่าง ๆ โดยให้มีระยะห่าง (Margin) กว้างที่สุด

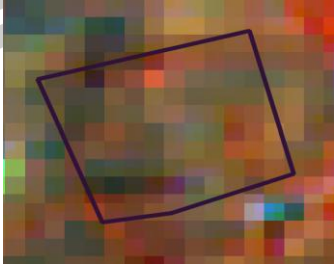

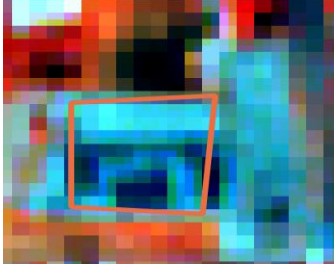



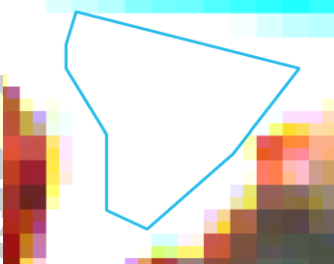

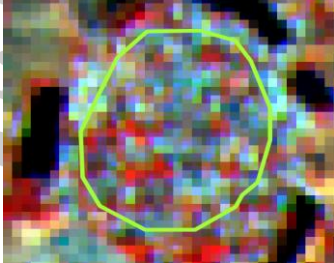

- เคอร์เนล (Kernel) Radial Basis Function (RBF) เพื่อจัดการกับความสัมพันธ์ที่ไม่เป็นเชิงเส้น
- พารามิเตอร์ C 10 (ควบคุม Trade-off ระหว่าง Margin กับ Classification Error)
- พารามิเตอร์ Gamma 0.1 (ควบคุมอิทธิพลของจุดข้อมูลแต่ละจุด)

3.3.4 โครงข่ายประสาทเทียม (Artificial Neural Network - ANN) แบบจำลองที่เรียนแบบโครงข่ายประสาทในสมองมนุษย์ ประกอบด้วยชั้น Input, Hidden Layers, และ Output Layer เพื่อเรียนรู้รูปแบบที่ซับซ้อนในข้อมูล

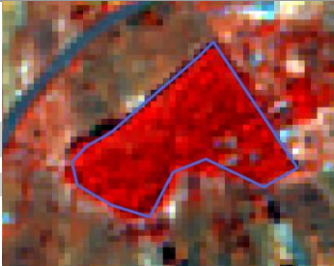





- โครงสร้าง 3 ชั้นซ่อน (Hidden Layers) โดยมีจำนวนโหนด 64, 32, และ 16 ตามลำดับ
- ฟังก์ชันกระตุ้น (Activation Function) ReLU (Rectified Linear Unit) สำหรับ Hidden Layers
- Optimizer Adam
- อัตราการเรียนรู้ (Learning Rate) 0.001

เพื่อให้เห็นภาพลักษณะของข้อมูลตัวอย่างที่ใช้ ตารางที่ 3.2 แสดงตัวอย่างภาพของสิ่งปกคลุมดินแต่ละประเภทที่ปรากฏบนภาพถ่ายดาวเทียม Sentinel-2 และภาพจาก Google Maps ที่ใช้เป็นข้อมูลอ้างอิง

ตารางที่ 3.2 กลุ่มตัวอย่าง

ประเภทข้อมูลสิ่งปกคลุมดิน	Sentinel -2	Google Map
พื้นที่เกษตรกรรม (A)		
เมืองและย่านการค้า		
พื้นที่ส่วนราชการ		
พื้นที่ว่างเปล่า		
พื้นที่อยู่อาศัย		

ตารางที่ 3.2 กลุ่มตัวอย่าง (ต่อ)

พื้นที่ป่าไม้		
พื้นที่แหล่งน้ำ		
พื้นที่สวนสาธารณะ		

### 3.4 การตรวจสอบความถูกต้อง (Accuracy Assessment)

หลังจากสร้างแบบจำลองและได้ผลการจำแนกสิ่งปกคลุมดินแล้ว จะทำการประเมินประสิทธิภาพและความแม่นยำของแต่ละเทคนิค โดยเปรียบเทียบผลลัพธ์กับข้อมูลอ้างอิงภาคพื้นดิน (Ground Truth) ที่ได้จากการสุ่มตัวอย่างจาก Google Earth และการสำรวจภาคสนาม (ซึ่งเป็นชุดข้อมูลที่แยกต่างหากจากข้อมูลที่ใช้ฝึกสอน) โดยใช้วิธีการและตัวชี้วัดดังต่อไปนี้

1. การสร้างตาราง Confusion Matrix เป็นตารางที่แสดงจำนวนจุดตัวอย่างที่ถูกจำแนกได้อย่างถูกต้องและผิดพลาดสำหรับแต่ละประเภทสิ่งปกคลุมดิน เปรียบเทียบระหว่างผลการจำแนกจากเทคนิคกับข้อมูลอ้างอิง

2. การคำนวณค่าความถูกต้องรายประเภท ความถูกต้องของผู้ผลิต (Producer's Accuracy - PA) สัดส่วนของจุดอ้างอิงภาคพื้นดินสำหรับประเภทหนึ่ง ๆ ที่ถูกจำแนกได้อย่างถูกต้อง บ่งชี้ว่าเทคนิคจำแนกพื้นที่ประเภทนั้นได้ครบถ้วนเพียงใด (คำนวณจากคอลัมน์ใน Confusion Matrix)

ค่าความถูกต้องของผู้ผลิต (Producer's Accuracy - PA) คำนวณจากสัดส่วนของ True Positive เทียบกับผลรวมของคอลัมน์ใน Confusion Matrix ดังสมการที่ 3.1

$$PA = \frac{\text{True Positive}}{\text{Column Total}} \times 100 \quad (3.1)$$

ค่าความถูกต้องของผู้ใช้ (User's Accuracy - UA) คำนวณจากสัดส่วนของ True Positive เทียบกับผลรวมของแถวใน Confusion Matrix ดังสมการที่ 3.2

$$UA = \frac{\text{True Positive}}{\text{Row Total}} \times 100 \quad (3.2)$$

ค่าความถูกต้องโดยรวม (Overall Accuracy - OA) คำนวณจากผลรวมของค่าบนเส้นทแยงมุมหารด้วยจำนวนตัวอย่างทั้งหมด ดังสมการที่ 3.3

- นำค่าบนเส้นทแยงมุมมาบวกกัน (True Positives ของแต่ละคลาส)
- หารด้วยจำนวนตัวอย่างทั้งหมด

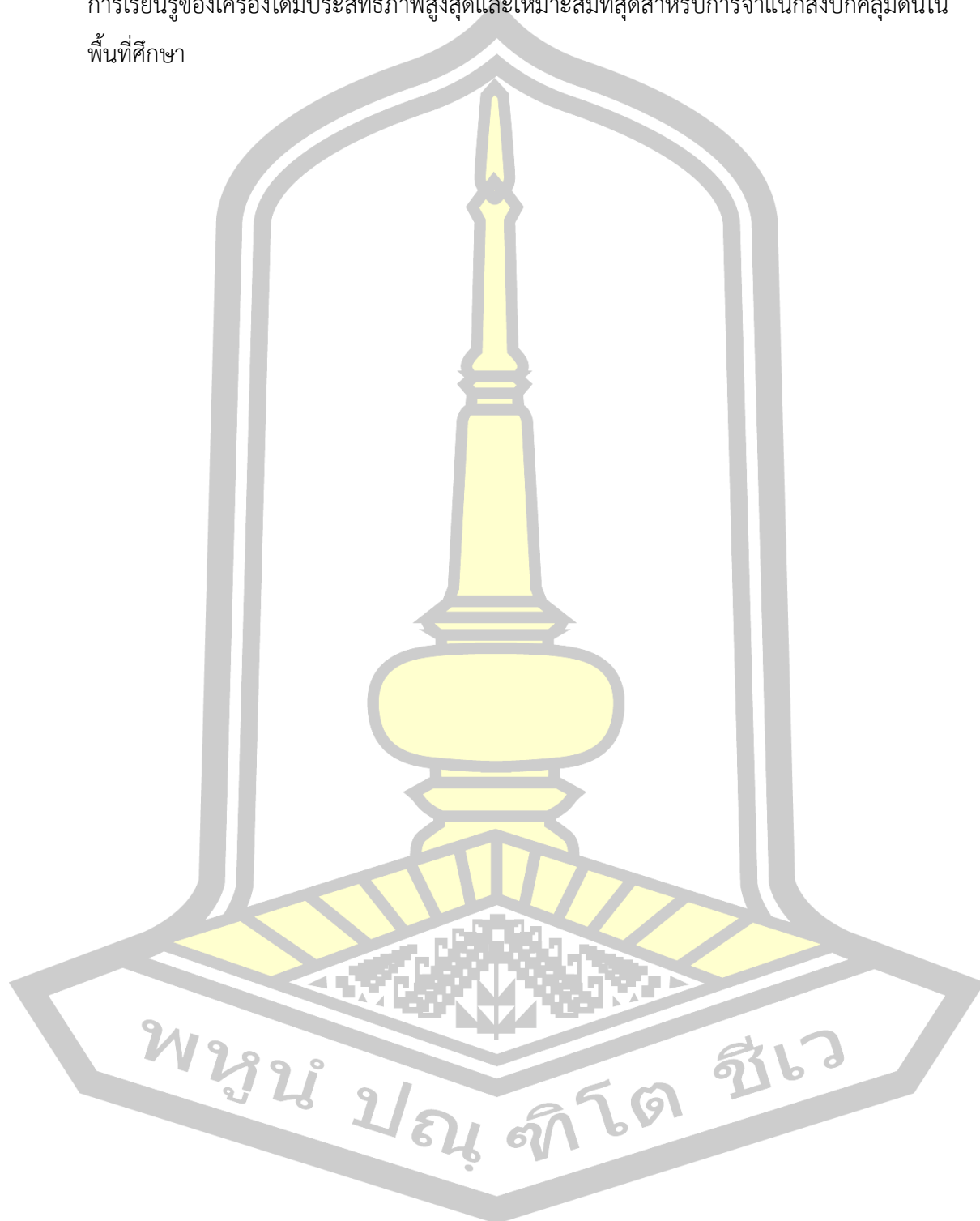
$$OA = \frac{\sum \text{Diagonal Elements}}{\text{Total Samples}} \times 100 \quad (3.3)$$

ค่าสัมประสิทธิ์แคปปา (Kappa Coefficient - K) ซึ่งวัดความสอดคล้องที่แท้จริง คำนวณได้จากสมการที่ 3.4 [17]

$$K = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_i + Xx_{+1})}{N^2 - \sum_{i=1}^r (x_{ii} Xx_{+1})} \quad (3.4)$$

การประเมินความเสถียรด้วย K-Fold Cross Validation ใช้วิธี K-Fold Cross Validation โดยตั้งค่า k=10 เพื่อประเมินประสิทธิภาพและความเสถียรของเทคนิคแต่ละชนิดอย่างน่าเชื่อถือยิ่งขึ้น โดยจะคำนวณค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของตัวชี้วัดต่าง ๆ จากการทดสอบทั้ง 10 รอบ ได้แก่ Accuracy, Kappa, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Precision, Recall, และ F1-score เพื่อเปรียบเทียบความสามารถในการ Generalize ของแต่ละเทคนิค

ผลการประเมินความถูกต้องทั้งหมดจะถูกนำมาวิเคราะห์และเปรียบเทียบ เพื่อสรุปว่าเทคนิคการเรียนรู้ของเครื่องใดมีประสิทธิภาพสูงสุดและเหมาะสมที่สุดสำหรับการจำแนกสิ่งปกคลุมดินในพื้นที่ศึกษา



## บทที่ 4

### ผลการศึกษา

บทนี้นำเสนอผลการวิเคราะห์ข้อมูลที่ได้จากการดำเนินงานวิจัยตามระเบียบวิธีที่กล่าวไว้ในบทที่ 3 โดยมีวัตถุประสงค์เพื่อจำแนกสิ่งปกคลุมดินในพื้นที่อำเภอเมืองร้อยเอ็ด จังหวัดร้อยเอ็ด และประเมินประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่อง 4 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ, เทคนิคป่าสุ่ม, เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และเทคนิคโครงข่ายประสาทเทียม การวิเคราะห์ใช้ข้อมูลภาพถ่ายดาวเทียม Sentinel-2 ซึ่งบันทึกเมื่อวันที่ 28 ธันวาคม พ.ศ. 2567 ผลการศึกษาที่นำเสนอในบทนี้ประกอบด้วย ผลการจำแนกสิ่งปกคลุมดินด้วยเทคนิคแต่ละวิธี การประเมินประสิทธิภาพเชิงลึกและการเปรียบเทียบความแม่นยำระหว่างเทคนิค เพื่อให้เห็นถึงความเหมาะสมและข้อจำกัดของแต่ละวิธีในการประยุกต์ใช้กับข้อมูลภาพถ่ายดาวเทียมในบริบทของพื้นที่ศึกษา

#### 4.1 ผลการวิเคราะห์ข้อมูลภาพการจำแนกสิ่งปกคลุมดิน

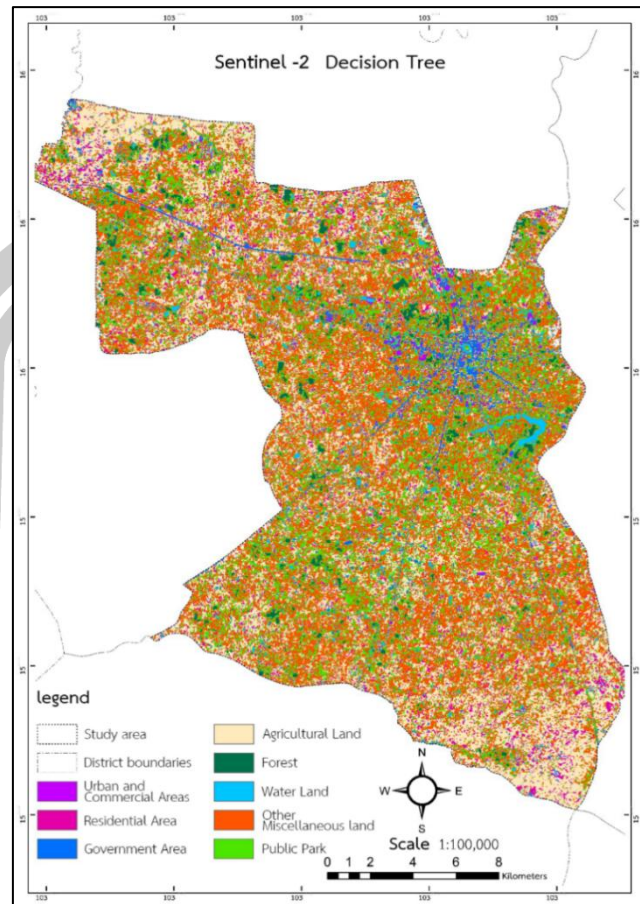
ส่วนนี้เป็นการนำเสนอผลลัพธ์ที่ได้จากการนำเทคนิคการเรียนรู้ของเครื่องทั้ง 4 วิธี มาใช้ในการจำแนกประเภทสิ่งปกคลุมดิน 8 ประเภท ได้แก่ พื้นที่เกษตรกรรม, เมืองและย่านการค้า, พื้นที่สวนราชการ, พื้นที่ว่างเปล่า, พื้นที่อยู่อาศัย, พื้นที่ป่าไม้, พื้นที่แหล่งน้ำ และพื้นที่สวนสาธารณะ ผลการวิเคราะห์จะแสดงทั้งในเชิงคุณภาพ ลักษณะแผนที่ผลการจำแนก ดังภาพประกอบที่ 4.1 ดังภาพประกอบที่ 4.2 ดังภาพประกอบที่ 4.3 และดังภาพประกอบที่ 4.4.เชิงปริมาณ ค่าความแม่นยำและสัดส่วนพื้นที่ ดังตารางที่ 4.1 ถึง ตารางที่ 4.12

##### 4.1.1 เทคนิคต้นไม้ตัดสินใจ (Decision Tree - DT)

ผลการจำแนกสิ่งปกคลุมดินด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree - DT) ซึ่งแสดงการกระจายตัวของแต่ละประเภทในพื้นที่ศึกษา ดังภาพประกอบที่ 4.1 แสดงแผนที่ผลการจำแนกสิ่งปกคลุมดิน 8 ประเภทในพื้นที่ศึกษา โดยใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree - DT) ซึ่งมีความถูกต้องโดยรวม (Overall Accuracy OA) เท่ากับ 70% และค่าสัมประสิทธิ์ Kappa เท่ากับ 0.63 โดยกำหนดค่า Max Depth ของต้นไม้ตัดสินใจเป็น 10 เพื่อควบคุมความซับซ้อนของแบบจำลองและป้องกันปัญหา Overfitting จากการวิเคราะห์ด้วยสายตา (Visual Interpretation) จากผลการจำแนกพบว่า การจำแนกพื้นที่แหล่งน้ำ (สีฟ้า) บริเวณเส้นขอบของคูเมืองเก่าในเขตเมืองร้อยเอ็ด ด้วยเทคนิคการจำแนกแบบต้นไม้ตัดสินใจ ไม่สามารถจำแนกเป็นแหล่งน้ำได้ถูกต้อง แต่ถูกจำแนกเป็นที่พื้นที่สวนสาธารณะ (สีเขียวอ่อน) และพื้นที่ราชการ (สีน้ำเงิน) ปะปนอยู่ร่วมกัน

โดยเฉพาะบริเวณโดยรอบพื้นที่ในเมือง ซึ่งปรากฏการจำแนกประเภทอื่น ๆ แทรกซ้อนอยู่ พื้นที่ราชการ (สีน้ำเงิน) มีการกระจายตัวตามแนวถนนสายหลักและสายรอง ในขณะที่พื้นที่เมืองและย่านการค้า (สีม่วง) มีลักษณะการตั้งเรียงตัวของอาคารบ้านเรือนตามแนวถนนสายหลัก ซึ่งแสดงถึงลักษณะพื้นที่เชิงพาณิชย์เป็นส่วนใหญ่ พื้นที่ราชการ ได้แก่ โรงเรียนและศูนย์ราชการต่าง ๆ ส่วนใหญ่จะตั้งอยู่บริเวณโดยรอบศูนย์กลางของเมือง อย่างไรก็ตาม ผลการจำแนกพบว่าการปะปนระหว่างพื้นที่ราชการกับพื้นที่เมืองและย่านการค้า (สีม่วง) หมู่บ้าน (สีชมพู) พื้นที่เกษตรกรรม (สีครีม) พื้นที่อื่น ๆ (สีส้ม) และพื้นที่สาธารณะ (สีเขียวอ่อน) ซึ่งตั้งอยู่ทั้งในใจกลางเมืองและบริเวณชานเมือง บริเวณชานเมืองพบว่าการจำแนกพื้นที่เกษตร พื้นที่ป่าไม้ และพื้นที่หมู่บ้านต่าง ๆ ที่รวมตัวกันเป็นกลุ่มย่อย โดยมีการปะปนของพื้นที่เมืองและย่านการค้า (สีม่วง) พื้นที่อื่น ๆ (สีส้ม) และพื้นที่สาธารณะ (สีเขียวอ่อน) อย่างไรก็ตาม การจำแนกพื้นที่ป่าไม้ (สีเขียวเข้ม) และพื้นที่เกษตรกรรม (สีครีม) บริเวณรอบนอกเมืองยังสามารถแปลผลได้ แม้จะพบว่าการปะปนของพื้นที่อื่น ๆ และพื้นที่สาธารณะในบางส่วน โดยเฉพาะบริเวณพื้นที่ป่าไม้ที่มีการปะปนกับพื้นที่เกษตรกรรมและพื้นที่สวนสาธารณะเป็นจำนวนมากทั้งในเมืองและโดยรอบเมือง ความซับซ้อนในการปะปนของประเภทสิ่งปกคลุมดินในพื้นที่ศึกษาดังกล่าวส่งผลให้ค่าความถูกต้องของการจำแนกภาพถ่ายดาวเทียมมีผลการจำแนกที่ต่ำ ดังภาพประกอบที่ 4.1





ภาพประกอบที่ 4.1 ผลการจำแนกสิ่งปกคลุมดินเทคนิค  
ต้นไม้ตัดสินใจ

ดังตารางที่ 4.1 แสดงค่า Overall Accuracy และ Kappa Statistic พร้อมทั้ง Confusion Matrix ที่ได้จากการจำแนกด้วยเทคนิคต้นไม้ตัดสินใจ (DT) ค่าความถูกต้องโดยรวม (Overall Accuracy) ที่ 0.70 และค่าสัมประสิทธิ์แคปปา (Kappa Coefficient) เท่ากับ 0.63 ค่า Kappa ที่ 0.63 นี้ จัดอยู่ในระดับ "ดี" (Substantial Agreement) ตามเกณฑ์การแปลผลของ Landis and Koch (1977) ซึ่งหมายความว่าความสอดคล้องระหว่างผลการจำแนกกับข้อมูลอ้างอิงนั้นดีกว่าการสุ่มอย่างมีนัยสำคัญ แต่ยังไม่ถึงระดับดีมาก

ตารางที่ 4.1 Overall Accuracy และ Overall Kappa Statistics การจำแนก  
เทคนิคต้นไม้ตัดสินใจ

ประเภทการจำแนก	จุดสำรวจภาคสนาม								
	เมืองและย่านการค้า	พื้นที่อยู่อาศัย	พื้นที่ส่วนราชการ	พื้นที่เกษตรกรรม	พื้นที่ป่าไม้	พื้นที่แหล่งน้ำ	พื้นที่ว่างเปล่า	พื้นที่สวนสาธารณะ	รวม
เมืองและย่านการค้า	7	2	0	1	0	1	2	0	13
พื้นที่อยู่อาศัย	1	39	0	8	0	0	0	0	48
พื้นที่ส่วนราชการ	4	8	9	0	0	0	0	0	21
พื้นที่เกษตรกรรม	0	0	0	78	2	2	5	0	87
พื้นที่ป่าไม้	0	0	0	2	60	0	0	0	62
พื้นที่แหล่งน้ำ	0	0	0	0	2	31	0	0	33
พื้นที่ว่างเปล่า	0	5	2	20	6	1	14	0	48
พื้นที่สวนสาธารณะ	0	5	0	9	10	2	6	6	38
รวม	12	59	11	118	80	37	27	6	350
ค่าความถูกต้องโดยรวม (Overall Accuracy) = 0.70									
ค่าสัมประสิทธิ์ Kappa = 0.63									

ผลลัพธ์จากตาราง confusion matrix ซึ่งแสดงให้เห็นถึงจำนวนของตัวอย่างในแต่ละประเภทพื้นที่ที่ถูกจำแนกได้อย่างถูกต้องและผิดพลาดนั้น พบว่า ประเภทพื้นที่ที่มีความแม่นยำในการจำแนกสูง ได้แก่ พื้นที่เกษตรกรรม พื้นที่ป่าไม้ และแหล่งน้ำ โดยพื้นที่เกษตรกรรมสามารถจำแนกได้อย่างถูกต้อง 39 ตัวอย่าง ขณะที่พื้นที่ป่าไม้ถูกจำแนกได้อย่างถูกต้อง 60 จุด และพื้นที่แหล่งน้ำสามารถจำแนกได้อย่างถูกต้อง 14 จุด แสดงให้เห็นว่าพื้นที่ดังกล่าวมีลักษณะเฉพาะที่โดดเด่นทางด้านลักษณะทางสเปกตรัมและโครงสร้างเชิงพื้นที่ ซึ่งส่งผลให้แบบจำลองสามารถเรียนรู้และแยกแยะได้อย่างมีประสิทธิภาพ ในทางกลับกัน ยังพบข้อผิดพลาดในการจำแนกประเภทพื้นที่บางประเภทในระดับที่ควรพิจารณา เช่น พื้นที่ว่างเปล่า พื้นที่สวนสาธารณะ พื้นที่เมืองและย่านการค้า รวมถึงพื้นที่ส่วนราชการ ซึ่งมีอัตราการจำแนกผิดพลาดสูง โดยพื้นที่สวนสาธารณะจำแนกได้อย่างถูกต้องเพียง 6 จุด และพื้นที่เมืองและย่านการค้าถูกจำแนกได้ 7 จุด ข้อมูลดังกล่าวสะท้อนให้เห็นถึงความคลาดเคลื่อนของแบบจำลองในบริเวณที่มีลักษณะทางกายภาพหรือโครงสร้างพื้นฐานที่คล้ายคลึงกัน เช่น อาคาร ถนน

พื้นที่โล่ง และสิ่งปลูกสร้าง ซึ่งอาจทำให้แบบจำลองไม่สามารถแยกแยะประเภทพื้นที่เหล่านี้ได้อย่างชัดเจน ส่งผลให้เกิดการสับสนระหว่างหมวดหมู่ในการจำแนก

นอกจากนี้ยังพบว่า พื้นที่บางประเภท เช่น พื้นที่สวนราชการและพื้นที่ว่างเปล่า มีจำนวนตัวอย่างที่ถูกจำแนกผิดไปยังหมวดหมู่อื่นในระดับที่สูง ซึ่งอาจมีสาเหตุมาจากลักษณะสเปกตรัมของภาพที่ใกล้เคียงกัน หรืออาจเนื่องจากขอบเขตของพื้นที่ในภาพถ่ายดาวเทียมมีการผสมผสานระหว่างองค์ประกอบของพื้นที่หลายประเภท เช่น พื้นที่ราชการที่ประกอบด้วยอาคาร ต้นไม้ ถนน และลานโล่ง รวมถึงการที่แบบจำลองอาจไม่ได้รับข้อมูลเพียงพอหรือมีความหลากหลายน้อยในกระบวนการเรียนรู้ ผลการจำแนกสามารถสรุปได้ว่า เทคนิคต้นไม้ตัดสินใจ มีศักยภาพในการจำแนกพื้นที่บางประเภทที่มีลักษณะเฉพาะชัดเจนได้ดี อย่างไรก็ตาม ในประเภทพื้นที่ที่มีความคล้ายคลึงกันทางกายภาพหรือโครงสร้างยังคงมีข้อจำกัดที่จำเป็นต้องได้รับการปรับปรุง เพื่อเพิ่มความแม่นยำของการจำแนก โดยอาจพิจารณาการใช้ข้อมูลเสริมเพิ่มเติม เช่น ดัชนีจากภาพถ่ายดาวเทียม (เช่น NDVI, NDBI) ข้อมูลความสูงภูมิประเทศ หรือข้อมูลภาพในช่วงคลื่นอื่น ๆ นอกจากนี้ การนำเทคนิคการเรียนรู้เชิงลึก หรือการรวมหลายเทคนิคเข้าด้วยกัน (ensemble methods) อาจเป็นแนวทางหนึ่งที่สามารถช่วยเพิ่มประสิทธิภาพในการจำแนกและลดข้อผิดพลาดของการจำแนกประเภทพื้นที่จากข้อมูลภาพถ่ายดาวเทียมในอนาคตได้อย่างมีประสิทธิภาพยิ่งขึ้น

เพื่อวิเคราะห์ประสิทธิภาพการจำแนกในแต่ละประเภทสิ่งปกคลุมดินของเทคนิค DT จึงได้คำนวณค่าความถูกต้องของผู้ผลิต (Producer's Accuracy) และความถูกต้องของผู้ใช้ (User's Accuracy) รวมถึงค่าความคลาดเคลื่อน ดังแสดงในตารางที่ 4.2 ให้ข้อมูลเชิงลึกเกี่ยวกับประสิทธิภาพของโมเดลในการจำแนกแต่ละคลาสโดยเปรียบเทียบผลการจำแนกกับข้อมูลอ้างอิงภาคสนามหรือข้อมูลความละเอียดสูง ค่านี้บ่งชี้ว่าโมเดลสามารถตรวจจับพื้นที่ประเภทนั้น ๆ ที่มีอยู่จริงได้ดีเพียงใด พบว่า พื้นที่สวนสาธารณะ มี PA สูงสุด (100%) ตามด้วย พื้นที่แหล่งน้ำ (83.78%) และ พื้นที่สวนราชการ (81.82%) แสดงว่าโมเดลตรวจพบพื้นที่เหล่านี้ส่วนใหญ่ได้ถูกต้อง อย่างไรก็ตาม เมืองและย่านการค้า และ พื้นที่ว่างเปล่า มี PA ค่อนข้างต่ำ (58.33% และ 51.85% ตามลำดับ) หมายความว่าพื้นที่จริงของสองคลาสนี้จำนวนมากที่โมเดลจำแนกผิดไปเป็นคลาสอื่น (Error of Omission สูง)

ตารางที่ 4.2 ความถูกต้องและความคลาดเคลื่อนของผู้ใช้และผู้ผลิตของเทคนิคต้นไม้ตัดสินใจ

ประเภท	Producer's Accuracy	Error omission	Users Accuracy	Error Commission
เมืองและย่านการค้า	58.33	41.67	53.85	46.15
พื้นที่อยู่อาศัย	66.10	33.90	81.25	18.75
พื้นที่ส่วนราชการ	81.82	18.18	42.86	57.14
พื้นที่เกษตรกรรม	66.10	33.90	89.66	10.34
พื้นที่ป่าไม้	75.00	25.00	96.77	3.23
พื้นที่แหล่งน้ำ	83.78	16.22	93.94	6.06
พื้นที่ว่างเปล่า	51.85	48.15	29.17	70.83
พื้นที่สวนสาธารณะ	100.00	0.00	15.79	84.21

User's Accuracy (UA) คำนึงชี้ว่าพื้นที่ที่ถูกจำแนกเป็นประเภทนั้น ๆ ในแผนที่ผลลัพธ์มีความน่าเชื่อถือเพียงใด (ความถูกต้อง) พบว่า พื้นที่ป่าไม้ และ พื้นที่แหล่งน้ำ มี UA สูงมาก (96.77% และ 93.94% ตามลำดับ) แสดงว่าแผนที่ในส่วนนี้มีความน่าเชื่อถือสูงมาก ตามมาด้วย พื้นที่เกษตรกรรม (89.66%) และ พื้นที่อยู่อาศัย (81.25%) ซึ่งถือว่ามีความน่าเชื่อถือดี ในทางตรงกันข้าม พื้นที่สวนสาธารณะ และ พื้นที่ว่างเปล่า มี UA ต่ำมาก (15.79% และ 29.17% ตามลำดับ) บ่งชี้ว่าแผนที่ในส่วนนี้มีความน่าเชื่อถือต่ำอย่างยิ่ง มีการจำแนกพื้นที่อื่นผิดมาเป็นสองคลาสนี้จำนวนมาก (Error of Commission สูงถึง 84.21% และ 70.83% ตามลำดับ) พื้นที่ส่วนราชการ และ เมืองและย่านการค้า ก็มี UA ไม่สูงนัก (42.86% และ 53.85%) แสดงว่ามีความน่าเชื่อถือปานกลางถึงต่ำ

พหุ ประถมศึกษา

ผลการจำแนกด้วยเทคนิค DT ทำให้ได้สัดส่วนพื้นที่ของสิ่งปกคลุมดินแต่ละประเภทในพื้นที่ศึกษา คิดเป็นตารางกิโลเมตร ไร่ และร้อยละ ดังรายละเอียดในตารางที่ 4.3

**ตารางที่ 4.3** แสดงสัดส่วนพื้นที่ที่ได้จากการจำแนกสิ่งปกคลุมดินเทคนิคต้นไม้ตัดสินใจ

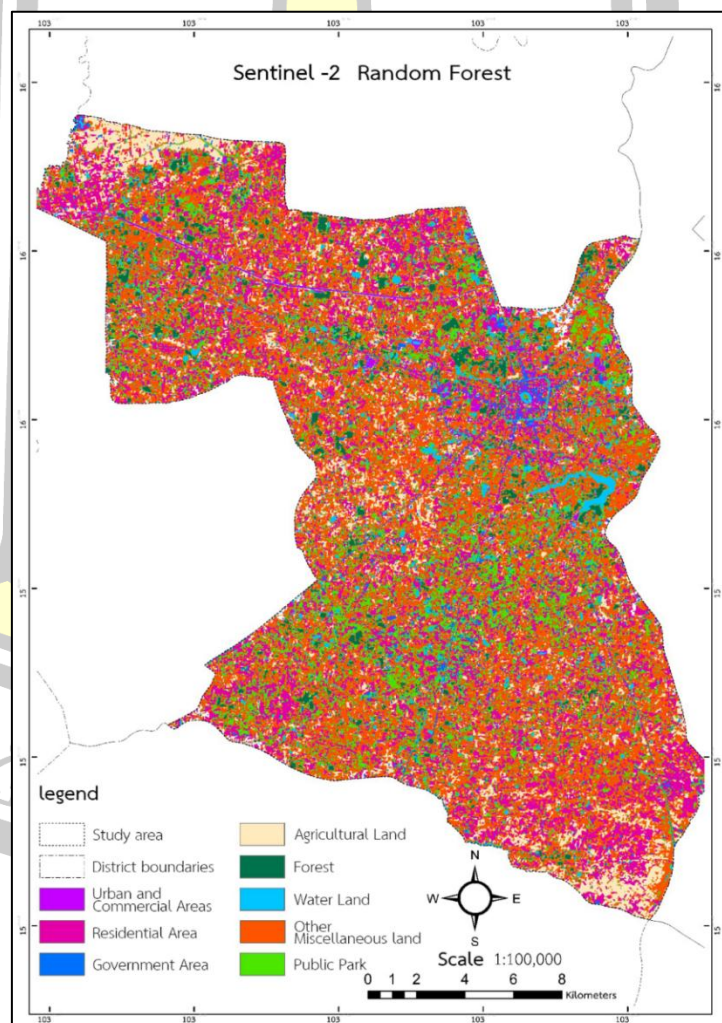
ประเภท	ตร.กม.	ตร.ม.	ไร่	ร้อยละ
พื้นที่เกษตรกรรม (A)	2	1847026	1154	0.37
เมืองและย่านการค้า	31	30625140	19141	6.20
พื้นที่ส่วนราชการ	27	26553816	16596	5.38
พื้นที่ว่างเปล่า	113	112733025	70458	22.84
พื้นที่อยู่อาศัย	8	8449664	5281	1.71
พื้นที่ป่าไม้	20	19911519	12445	4.03
พื้นที่แหล่งน้ำ	161	160936732	100585	32.60
พื้นที่สวนสาธารณะ	133	132543079	82839	26.85
รวม	493.6000	493600000	308500	100

จากตารางที่ 4.3 พบว่าพื้นที่ส่วนใหญ่ถูกจำแนกเป็นพื้นที่แหล่งน้ำ (32.60%) และพื้นที่สวนสาธารณะ (26.85%) ตามด้วยพื้นที่ว่างเปล่า (22.84%) สัดส่วนพื้นที่สวนสาธารณะและพื้นที่ว่างเปล่าที่สูงมากที่สุดสอดคล้องโดยตรงกับค่า User's Accuracy ที่ต่ำของสองคลาสนี้ ซึ่งเป็นการยืนยันเชิงปริมาณถึงแนวโน้มการจำแนกพื้นที่ประเภทอื่นผิดพลาดมาเป็นสองคลาสนี้มากเกินไป (Over-classification) ของเทคนิค DT

#### 4.1.2 เทคนิคป่าสุ่ม (Random Forest - RF)

ผลการจำแนกสิ่งปกคลุมดินด้วยเทคนิคป่าสุ่ม (Random Forest - RF) ซึ่งแสดงการกระจายตัวของแต่ละประเภทในพื้นที่ศึกษา แสดงดังแผนที่ดังภาพประกอบที่ 4.2 โดยมีค่าความถูกต้องโดยรวม (Overall Accuracy OA) เท่ากับ 74% และค่าสัมประสิทธิ์ Kappa เท่ากับ 0.68 ดังตารางที่ 4.4 ใช้จำนวนต้นไม้ในป่า (Number of Trees) เท่ากับ 100 ต้น และกำหนดค่า Max Depth เป็น 20 เพื่อเพิ่มประสิทธิภาพของแบบจำลองในการเรียนรู้ลักษณะข้อมูล จากการวิเคราะห์ด้วยสายตา (Visual Interpretation) พบว่า การจำแนกพื้นที่แหล่งน้ำ (สีฟ้า) ซึ่งเป็นบริเวณคูคลองเก่าที่มีลักษณะเป็นคลองรูปทรงสี่เหลี่ยมโดยรอบพื้นที่สาธารณะ สามารถจำแนกได้อย่างแม่นยำและชัดเจน แม้ยังคงมีการปะปนเล็กน้อยกับพื้นที่ราชการ (สีน้ำเงิน) และพื้นที่สวนสาธารณะ (สีเขียวอ่อน) บริเวณใจกลางเมืองพบการกระจุกตัวของพื้นที่ราชการ (สีน้ำเงิน) ที่ตั้งอยู่ในจุดกึ่งกลางของเมือง

ร่วมกับพื้นที่เมืองและย่านการค้า (สีม่วง) รวมถึงพื้นที่หมู่บ้าน (สีชมพู) ที่มีการกระจายตัวไปตามแนวถนนทั้งสายหลักและสายรอง การกระจายนี้ขยายเข้าสู่พื้นที่โดยรอบเมืองและชานเมือง ซึ่งส่วนใหญ่เป็นพื้นที่เกษตรกรรม (สีเขียว) จากการวิเคราะห์ทางสายตาพบเพิ่มเติมว่าพื้นที่อื่น ๆ พื้นที่สวนสาธารณะ และพื้นที่หมู่บ้าน มีการปะปนในจำนวนมาก ทั้งในพื้นที่เกษตรกรรมและตามเส้นทางถนนในหมู่บ้าน อย่างไรก็ตาม แบบจำลองสามารถจำแนกพื้นที่ป่าไม้ (สีเขียวเข้ม) ได้อย่างมีประสิทธิภาพ แม้บางพื้นที่จะมีการปะปนกับพื้นที่สวนสาธารณะ (สีเขียวอ่อน) และพื้นที่หมู่บ้าน (สีชมพู) ผลลัพธ์ดังกล่าวแสดงให้เห็นว่า แบบจำลอง RF มีความสามารถในการเรียนรู้และแยกแยะลักษณะเฉพาะของพื้นที่ได้ดี โดยสามารถจำแนกพื้นที่ที่มีลักษณะแตกต่างกันอย่างชัดเจนจากประเภทพื้นที่อื่น ๆ ได้อย่างแม่นยำมากยิ่งขึ้น ผลการประเมินความแม่นยำของเทคนิคป่าสุ่ม สำหรับเทคนิคป่าสุ่ม (RF) ผลการประเมินความแม่นยำโดยรวมและค่าสถิติแคปปา สรุปลได้จาก Confusion Matrix ดังตารางที่ 4.4



ภาพประกอบที่ 4.2 ผลการจำแนกสิ่งปกคลุมดินเทคนิคป่าสุ่ม

ตารางที่ 4.4 Overall Accuracy และ Overall Kappa Statistics การจำแนกเทคนิคป่าสุ่ม

ประเภทการจำแนก	จุดสำรวจภาคสนาม								
	เมืองและย่านการค้า	พื้นที่อยู่อาศัย	พื้นที่ส่วนราชการ	พื้นที่เกษตรกรรม	พื้นที่ป่าไม้	พื้นที่แหล่งน้ำ	พื้นที่ว่างเปล่า	พื้นที่สวนสาธารณะ	รวม
เมืองและย่านการค้า	8	4	1	0	1	3		0	17
พื้นที่อยู่อาศัย	1	47	3	5	0	0	0	0	56
พื้นที่ส่วนราชการ	3	4	5	1	0	0	2	0	15
พื้นที่เกษตรกรรม	0	2	2	93	2	0	9	1	109
พื้นที่ป่าไม้	0	0	0	2	63	0	0	0	65
พื้นที่แหล่งน้ำ	0	0	0	2	2	30	0	0	34
พื้นที่ว่างเปล่า	0	2	0	5	3	1	11	2	24
พื้นที่สวนสาธารณะ	0	3	0	10	6	3	5	3	30
รวม	12	62	11	118	77	37	27	6	350
ค่าความถูกต้องโดยรวม (Overall Accuracy) = 0.74									
ค่าสัมประสิทธิ์ Kappa = 0.68									

ผลลัพธ์ดังตารางที่ 4.4 confusion matrix แสดงให้เห็นถึงความแม่นยำในการจำแนกที่แตกต่างกันไปในแต่ละประเภทพื้นที่ที่สามารถจำแนกได้อย่างแม่นยำมากที่สุด คือ พื้นที่เกษตรกรรม พื้นที่พื้นที่ป่าไม้ และแหล่งน้ำ โดยพื้นที่เกษตรกรรมสามารถจำแนกได้ถูกต้องจำนวน 93 จุด ขณะที่พื้นที่พื้นที่ป่าไม้มีการจำแนกถูกต้องถึง 63 จุด และพื้นที่แหล่งน้ำจัดอยู่ในกลุ่มที่มีความถูกต้องสูงเช่นกัน โดยถูกจำแนกได้อย่างถูกต้องจำนวน 30 จุด ผลลัพธ์ดังกล่าวแสดงให้เห็นว่าแบบจำลองสามารถเรียนรู้และแยกแยะลักษณะเฉพาะของพื้นที่เหล่านี้ได้ดี เนื่องจากแต่ละประเภทมีลักษณะทางสเปกตรัมที่ค่อนข้างชัดเจนและมีความแตกต่างจากประเภทพื้นที่อื่น ๆ อย่างมีนัยสำคัญ ในทางตรงกันข้าม ประเภทพื้นที่ที่มีความคลาดเคลื่อนในการจำแนกในระดับที่ควรได้รับการพิจารณา ได้แก่ พื้นที่เมืองและย่านการค้า พื้นที่ว่างเปล่า และพื้นที่สวนสาธารณะ โดยพื้นที่เมืองและย่านการค้าถูกจำแนกผิดเป็นประเภทอื่นจำนวนหนึ่ง และพื้นที่ส่วนราชการมีการจำแนกถูกต้องเพียง 5 ตัวอย่าง ซึ่งอาจมีสาเหตุมาจากลักษณะของพื้นที่ที่คล้ายคลึงกับพื้นที่เกษตรกรรมหรือพื้นที่ชุ่มน้ำ เช่นเดียวกันกับพื้นที่สวนสาธารณะที่มีความสับสนในการจำแนกสูง อาจเนื่องมาจากองค์ประกอบภายในพื้นที่ที่

หลากหลาย เช่น ต้นไม้ ทางเดิน อาคาร หรือแหล่งน้ำขนาดเล็ก ทำให้แบบจำลองไม่สามารถแยกแยะได้ชัดเจนเพียงพอ นอกจากนี้ยังพบว่าพื้นที่ที่มีโครงสร้างคล้ายคลึงกัน เช่น พื้นที่ส่วนราชการ พื้นที่เมืองและย่านการค้า มักจะเกิดความสับสนระหว่างกันในการจำแนก โดยเฉพาะในกรณีที่มีการพัฒนาโครงสร้างพื้นฐานร่วมกัน หรือมีลักษณะทางกายภาพที่คล้ายคลึง เช่น พื้นผิวแข็ง ถนน และอาคาร ซึ่งสะท้อนถึงความท้าทายในการจำแนกประเภทพื้นที่ในเขตเมืองที่มีความหลากหลายทางการใช้ประโยชน์และโครงสร้างที่ทับซ้อนกัน ประสิทธิภาพการจำแนกรายประเภทของเทคนิค RF ซึ่งวัดจากค่า Producer's Accuracy และ User's Accuracy รวมถึงค่าความคลาดเคลื่อน แสดงดังตารางที่ 4.5

**ตารางที่ 4.5** ความถูกต้องและความคลาดเคลื่อนของผู้ใช้และผู้ผลิตของเทคนิคป่าสุ่ม

ประเภท	Producers Accuracy	Error Omission	Users Accuracy	Error commission
เมืองและย่านการค้า	67	33	47	53
พื้นที่อยู่อาศัย	76	24	84	16
พื้นที่ส่วนราชการ	45	55	33	67
พื้นที่เกษตรกรรม	79	21	85	15
พื้นที่ป่าไม้	82	18	97	3
พื้นที่แหล่งน้ำ	81	19	88	12
พื้นที่ว่างเปล่า	41	59	46	54
พื้นที่สวนสาธารณะ	50	50	10	90

User's Accuracy (UA) RF แสดงการปรับปรุง UA ในหลายคลาส เช่น พื้นที่อยู่อาศัย (84% vs 81.25% ของ DT) และ พื้นที่ว่างเปล่า (46% vs 29.17% ของ DT) ซึ่งแสดงว่าผลการจำแนกในแผนที่สำหรับคลาสเหล่านี้มีความน่าเชื่อถือมากขึ้น อย่างไรก็ตาม ปัญหาความแม่นยำต่ำยังคงปรากฏชัดเจนอย่างยิ่งใน พื้นที่ส่วนราชการ (UA เพียง 33%) และโดยเฉพาะ พื้นที่สวนสาธารณะ ซึ่งมี UA ต่ำมากเพียง 10% และมี Error of Commission สูงถึง 90% แสดงว่า RF ยังคงมีปัญหาอย่างรุนแรงในการแยกแยะสองคลาสนี้ออกจากคลาสนี้ อื่น ๆ

Producer's Accuracy (PA) ค่า PA สำหรับ พื้นที่ส่วนราชการ และ พื้นที่ว่างเปล่า ลดลงอย่างมากเมื่อเทียบกับ DT (45% vs 81.82% และ 41% vs 51.85% ตามลำดับ) บ่งชี้ว่า RF อาจจะทำนายพื้นที่จริงของสองคลาสนี้ผิดไปเป็นคลาสนี้มากขึ้น (Error of Omission สูงขึ้น) ในขณะที่ PA

ของคลาสอื่น ๆ ส่วนใหญ่ใกล้เคียงเดิมหรือดีขึ้นเล็กน้อย เทคนิค RF ให้ผลการจำแนกสัดส่วนพื้นที่ของสิ่งปกคลุมดินแต่ละประเภทแตกต่างจากเทคนิค DT ดังตารางที่ 4.6

**ตารางที่ 4.6** แสดงสัดส่วนพื้นที่ที่ได้จากการจำแนกสิ่งปกคลุมดินเทคนิคป่าสุ่ม

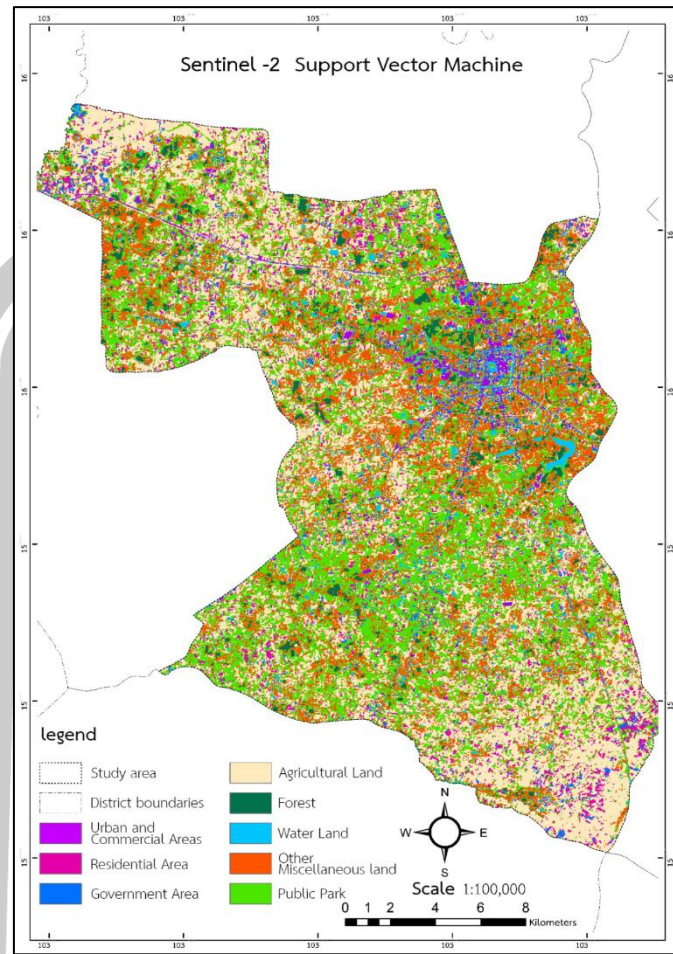
ประเภท	ตร.กม.	ตร.ม.	ไร่	ร้อยละ
พื้นที่เกษตรกรรม (A)	4	3,891,000	2,432	0.79
เมืองและย่านการค้า	113	113,077,899	70,674	22.91
พื้นที่ส่วนราชการ	13	12,927,036	8,079	2.62
พื้นที่ว่างเปล่า	51	50,806,931	31,754	10.29
พื้นที่อยู่อาศัย	16	15,974,438	9,984	3.24
พื้นที่ป่าไม้	19	19,342,616	12,089	3.92
พื้นที่แหล่งน้ำ	200	199,554,619	124,722	40.43
พื้นที่สวนสาธารณะ	78	78,025,461	48,766	15.81
รวม	493.6000	493600000	308500	100

เทคนิค RF จำแนกพื้นที่แหล่งน้ำเป็นสัดส่วนใหญ่ที่สุด (40.43%) ตามด้วยเมืองและย่านการค้า (22.91%) และสวนสาธารณะ (15.81%) สัดส่วนพื้นที่สวนสาธารณะลดลงจาก DT แต่ยังคงถือว่าสูงเมื่อพิจารณาจากค่า User's Accuracy ที่ต่ำมากเพียง 10% ซึ่งยังคงบ่งชี้ถึงปัญหาการจำแนกเกิน (Over-classification) อย่างมีนัยสำคัญสำหรับคลาสนี้



#### 4.1.3 เทคนิคซ์พอร์ตเวกเตอร์แมชชีน (Support Vector Machine - SVM)

แผนที่ผลลัพธ์การจำแนกสิ่งปกคลุมดินด้วยเทคนิคซ์พอร์ตเวกเตอร์แมชชีน (SVM) ดังภาพประกอบที่ 4.3 แสดงซึ่งให้ค่าความถูกต้องโดยรวม (Overall Accuracy OA) เท่ากับ 78% และค่าสัมประสิทธิ์แคปปา (Kappa Coefficient) เท่ากับ 0.73 ดังแสดงในตารางที่ 4. โดยในการจำแนกครั้งนี้ใช้เคอร์เนลแบบเรเดียลเบสฟังก์ชัน (Radial Basis Function RBF) เพื่อรองรับลักษณะข้อมูลที่ไม่เป็นเชิงเส้น โดยกำหนดค่าพารามิเตอร์ C เท่ากับ 10 และค่าพารามิเตอร์ Gamma เท่ากับ 0.1 จากการวิเคราะห์ด้วยสายตา (Visual Interpretation) พบว่า การจำแนกพื้นที่แหล่งน้ำ (สีฟ้า) ซึ่งเป็นพื้นที่คูคลองเก่าที่มีลักษณะเป็นคลองรูปทรงสี่เหลี่ยมล้อมรอบพื้นที่สาธารณะ ยังไม่สามารถจำแนกได้อย่างแม่นยำมากนัก โดยมีการปะปนของพื้นที่สวนสาธารณะ (สีเขียวอ่อน) พื้นที่เมืองและย่านการค้า (สีม่วง) และพื้นที่ราชการ (สีน้ำเงิน) ในระดับหนึ่ง อย่างไรก็ตาม ลักษณะของแหล่งน้ำที่มีรูปร่างสี่เหลี่ยมยังสามารถสังเกตเห็นได้ชัดเจน พื้นที่เมืองและย่านการค้า (สีม่วง) มีการกระจายตัวไปตามแนวถนนสายหลัก และสามารถจำแนกได้อย่างแม่นยำ เนื่องจากเขตอำเภอเมืองร้อยเอ็ดมีลักษณะ เมืองแบบกระจุกตัว โดยมีอาคารพาณิชย์เป็นองค์ประกอบหลัก พื้นที่ราชการ (สีน้ำเงิน) พบว่ามีการกระจายตัวตามแนวถนนสายหลัก และขยายตัวออกไปยังรอบนอกของพื้นที่เมืองทั้งภายในและนอกเขตคูคลองเมืองเก่า ในส่วนของพื้นที่สวนสาธารณะ (สีเขียวอ่อน) และพื้นที่ว่างเปล่า พบว่ามีการกระจายตัวเป็นวงกว้างตามพื้นที่ต่าง ๆ มีทั้งในพื้นที่เกษตรกรรมและชุมชนรอบนอกเมือง พื้นที่ชุมชนโดยรอบเขตชานเมือง ลักษณะการตั้งถิ่นฐานแบบกระจุกตัวเป็นวงกลม หรือขยายไปตามแนวถนนขนบสำหรับพื้นที่ป่าไม้ การจำแนกสามารถทำได้อย่างชัดเจน โดยมีขอบเขตของพื้นที่ป่าที่แน่นอน แม้จะมีการปะปนจากพื้นที่โดยรอบบ้างเล็กน้อย พื้นที่ที่จำแนกได้ถูกต้องมากที่สุด คือพื้นที่เกษตรกรรม ซึ่งครอบคลุมพื้นที่ประมาณ 60% ของพื้นที่ทั้งหมดในเขตอำเภอเมืองร้อยเอ็ด โดยเฉพาะในบริเวณพื้นที่ชานเมือง ซึ่งแสดงให้เห็นถึงศักยภาพของข้อมูลและเทคนิคที่ใช้ในการจำแนกประเภทสิ่งปกคลุมดินในพื้นที่ศึกษา ผลการประเมินความแม่นยำโดยรวมและค่าสถิติแคปปาของเทคนิคซ์พอร์ตเวกเตอร์แมชชีน (SVM) สรุปลได้จาก Confusion Matrix ดังตารางที่ 4.7



ภาพประกอบที่ 4.3 ผลการจำแนกสิ่งปกคลุมดินเทคนิคซ์ฟอว์ทเวกเตอร์แมชชีน

ค่าความถูกต้องโดยรวม (Overall Accuracy) ที่ 0.78 และค่าสัมประสิทธิ์แคปปา (Kappa Coefficient) เท่ากับ 0.73 ซึ่งสูงกว่า DT และ RF และยังคงจัดอยู่ในระดับ "ดี" (Substantial Agreement) แสดงถึงประสิทธิภาพโดยรวมที่ดีขึ้นอีกระดับดังตารางที่ 4.7 Overall Accuracy และ Overall Kappa Statistics การจำแนกเทคนิคซ์ฟอว์ทเวกเตอร์แมชชีน

พหุ ประถม โท ชีวะ

ตารางที่ 4.7 Overall Accuracy และ Overall Kappa Statistics การจำแนกเทคนิคซ์ฟอर्थ  
เวกเตอร์แมชชีน

ประเภทการจำแนก	จุดสำรวจภาคสนาม								
	เมืองและย่านการค้า	พื้นที่อยู่อาศัย	พื้นที่สวนราชการ	พื้นที่เกษตรกรรม	พื้นที่ป่าไม้	พื้นที่แหล่งน้ำ	พื้นที่ว่างเปล่า	พื้นที่สวนสาธารณะ	รวม
เมืองและย่านการค้า	10	7	2	1	2	1	0	0	23
พื้นที่อยู่อาศัย	0	43	0	5	1	0	0	0	49
พื้นที่สวนราชการ	0	3	4	3	0	0	0	0	10
พื้นที่เกษตรกรรม	0	0	0	93	0	0	4	0	97
พื้นที่ป่าไม้	0	1	0	2	66	0	0	0	69
พื้นที่แหล่งน้ำ	1	0	0	1	0	32	0	0	34
พื้นที่ว่างเปล่า	0	0	1	7	0	0	16	0	24
พื้นที่สวนสาธารณะ	1	8	2	5	8	4	7	9	44
รวม	12	62	9	117	77	37	27	9	350
ค่าความถูกต้องโดยรวม (Overall Accuracy) = 0.78									
ค่าสัมประสิทธิ์ Kappa = 0.73									

ผลลัพธ์ดังตารางที่ 4.7 confusion matrix ซึ่งแสดงรายละเอียดของจำนวนจุดสำรวจภาคสนามที่ถูกจำแนกได้อย่างถูกต้องและผิดพลาดในแต่ละประเภทพื้นที่ พบว่าพื้นที่ที่จำแนกได้อย่างแม่นยำที่สุดคือพื้นที่เกษตรกรรม โดยสามารถจำแนกได้ถูกต้อง 93 จุด แสดงให้เห็นถึงลักษณะเชิงสเปกตรัมที่ชัดเจนและมีความแตกต่างจากประเภทอื่น ส่งผลให้แบบจำลองสามารถเรียนรู้และแยกแยะได้ดี รองลงมาคือพื้นที่ป่าไม้ซึ่งจำแนกถูกต้อง 66 จุด และพื้นที่อยู่อาศัยซึ่งจำแนกถูกต้อง 43 จุด ข้อมูลดังกล่าวสะท้อนถึงความสามารถของแบบจำลองในการรับรู้ลักษณะเฉพาะของประเภทพื้นที่ที่มีความแตกต่างกันอย่างชัดเจน โดยเฉพาะประเภทพื้นที่ที่มีลักษณะทางธรรมชาติหรือการใช้ประโยชน์ที่เฉพาะเจาะจง พื้นที่แหล่งน้ำ พบว่ามีการจำแนกถูกต้อง 32 จุด ซึ่งถือว่ามี ความแม่นยำในระดับสูงเช่นกัน โดยลักษณะของน้ำในภาพถ่ายดาวเทียมมักมีความแตกต่างอย่างชัดเจนจากพื้นที่อื่น ๆ โดยเฉพาะในแถบคลื่นแม่เหล็กไฟฟ้าช่วงอินฟราเรดใกล้ ซึ่งสามารถแยกน้ำออกจากพืชพรรณและดินได้อย่างมีประสิทธิภาพ ยังพบว่าการจำแนกประเภทพื้นที่บางประเภทมีความ

คลาดเคลื่อนสูง เช่น พื้นที่สวนสาธารณะ ซึ่งจำแนกถูกต้องเพียง 9 จุด และมีการจำแนกผิดไปยังประเภทพื้นที่อื่น ๆ จำนวนมาก เช่น พื้นที่อยู่อาศัย พื้นที่เกษตรกรรม และพื้นที่ป่าไม้ ความคลาดเคลื่อนดังกล่าวอาจเกิดจากลักษณะเฉพาะของสวนสาธารณะที่มีองค์ประกอบหลากหลาย เช่น พื้นที่สีเขียว ทางเดิน สิ่งก่อสร้าง และแหล่งน้ำขนาดเล็ก ทำให้มีลักษณะคล้ายคลึงกับประเภทพื้นที่อื่น โดยเฉพาะในระดับภาพถ่ายจากดาวเทียมที่ไม่สามารถแยกรายละเอียดภายในพื้นที่ได้อย่างชัดเจน พื้นที่ว่างเปล่า พบว่ามีการจำแนกถูกต้อง 16 จุด ขณะที่พื้นที่เมืองและย่านการค้า มีการจำแนกถูกต้อง 10 จุด และพื้นที่ส่วนราชการจำแนกได้ถูกต้อง 4 จุด ซึ่งสะท้อนให้เห็นว่าพื้นที่ที่มีโครงสร้างทางกายภาพคล้ายคลึงกัน หรือมีการใช้งานร่วมกัน อาจทำให้เกิดการสับสนในการจำแนกได้ โดยเฉพาะอย่างยิ่งในบริเวณเขตเมืองที่มีการพัฒนาและใช้ประโยชน์ที่ดินอย่างหลากหลายและต่อเนื่อง สรุปการจำแนกประเภทพื้นที่ที่ใช้เทคนิค SVM จากภาพถ่ายดาวเทียม Sentinel-2 มีความแม่นยำอยู่ในระดับที่ดี โดยเฉพาะในพื้นที่ที่มีลักษณะเฉพาะที่ชัดเจน เช่น พื้นที่เกษตรกรรม พื้นที่ป่าไม้ และพื้นที่แหล่งน้ำ อย่างไรก็ตาม ในบางประเภทพื้นที่ที่มีลักษณะผสมผสานและคล้ายคลึงกัน เช่น พื้นที่สวนสาธารณะ พื้นที่เมือง และพื้นที่ราชการ ยังพบข้อผิดพลาดในการจำแนกในระดับหนึ่ง ซึ่งควรได้รับการปรับปรุงในอนาคต โดยอาจพิจารณาใช้ข้อมูลเสริมเพิ่มเติม เช่น ดัชนีพืชพรรณ (NDVI), ดัชนีความชื้น, ข้อมูลภาพถ่ายหลายช่วงเวลา หรือการปรับปรุงคุณภาพของข้อมูลการฝึกสอนแบบจำลอง เพื่อเพิ่มความแม่นยำในการจำแนก และลดข้อผิดพลาดที่อาจส่งผลกระทบต่อ การวิเคราะห์และวางแผนการใช้ที่ดินในระดับพื้นที่ การวิเคราะห์ความถูกต้องรายประเภทของเทคนิค SVM โดยพิจารณาค่า Producer's Accuracy, User's Accuracy และค่าความคลาดเคลื่อน แสดงไว้ในตารางที่ 4.8



**ตารางที่ 4.8** ความถูกต้องและความคลาดเคลื่อนของผู้ใช้และผู้ผลิตของเทคนิค  
ซัพพอร์ตเวกเตอร์แมชชีน

ประเภท	Producers Accuracy	Error Omission	Users Accuracy	Error commission
เมืองและย่านการค้า	83	16.67	43.48	57
พื้นที่อยู่อาศัย	69	30.65	87.76	12
พื้นที่ส่วนราชการ	44	55.56	40.00	60
พื้นที่เกษตรกรรม	79	20.51	95.88	4
พื้นที่ป่าไม้	86	14.29	95.65	4
พื้นที่แหล่งน้ำ	86	13.51	94.12	6
พื้นที่ว่างเปล่า	59	40.74	66.67	33
พื้นที่สวนสาธารณะ	100	0.00	20.45	80

User's Accuracy (UA) SVM แสดงประสิทธิภาพที่โดดเด่นอย่างมากในแง่ของ UA สำหรับพื้นที่เกษตรกรรม (95.88%), ป่าไม้ (95.65%), และแหล่งน้ำ (94.12%) ซึ่งสูงกว่าสองเทคนิคแรกอย่างชัดเจน บ่งชี้ความน่าเชื่อถือที่สูงขึ้นอย่างมีนัยสำคัญของแผนที่อยู่ในคลาสเหล่านี้ นอกจากนี้ UA ของพื้นที่ว่างเปล่า ก็ดีขึ้นอย่างมาก (66.67% vs 46% ของ RF) และ พื้นที่อยู่อาศัย ก็ยังคงสูง (87.76%) อย่างไรก็ตาม SVM ยังคงมีปัญหาอย่างมากในการจำแนก พื้นที่ส่วนราชการ (UA 40.00%) และ พื้นที่สวนสาธารณะ (UA เพียง 20.45%) ซึ่งมีค่า Error of Commission สูงมาก (60% และ 80% ตามลำดับ) แสดงว่า SVM ยังไม่สามารถแยกแยะสองคลาสนี้จากคลาสอื่น ๆ ได้ดีนัก

Producer's Accuracy (PA) ค่า PA ของ SVM มีความน่าสนใจที่ พื้นที่สวนสาธารณะ กลับมาเป็น 100% อีกครั้ง (เหมือน DT) แต่ PA ของ พื้นที่ส่วนราชการ ยังคงต่ำ (44%) และ PA ของ พื้นที่อยู่อาศัย ลดลง (69% vs 76% ของ RF)

พูน ปณ ทิโต ชเว

ขนาดและสัดส่วนพื้นที่ร้อยละของสิ่งปกคลุมดินที่จำแนกโดยเทคนิค SVM แสดงไว้ดังตารางที่ 4.9

**ตารางที่ 4.9** แสดงสัดส่วนพื้นที่ที่ได้จากการจำแนกสิ่งปกคลุมดินเทคนิคซัพพอร์ต

เวกเตอร์แมชชีน

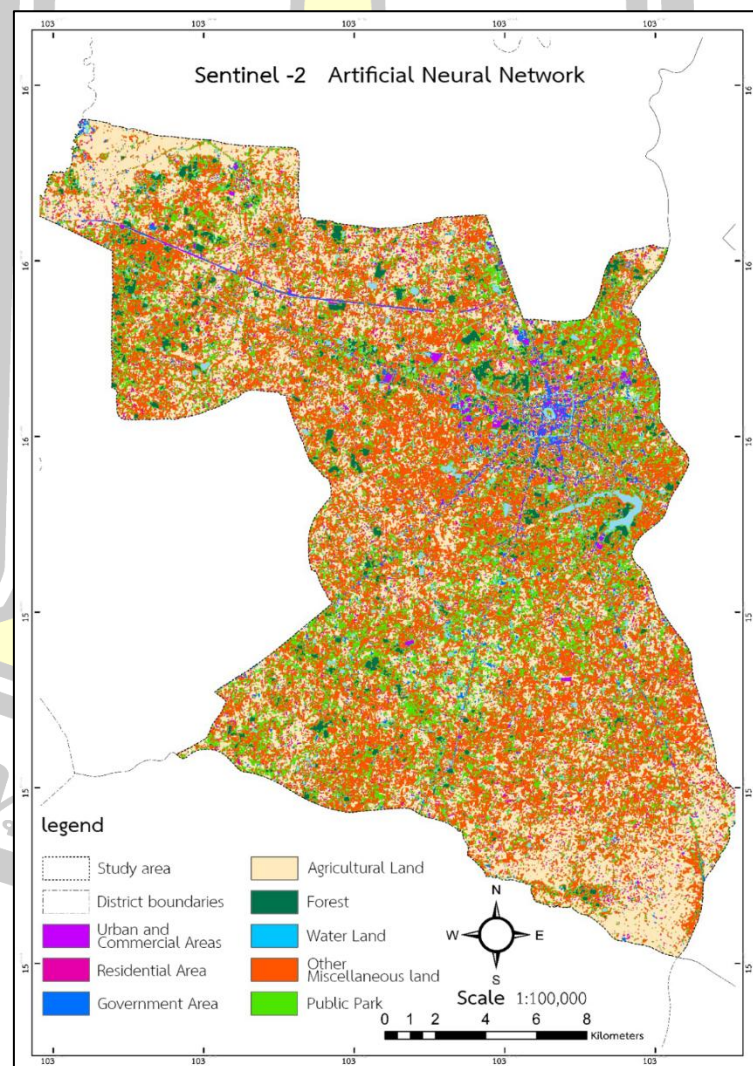
ประเภท	ตร.กม.	ตร.ม.	ไร่	ร้อยละ
พื้นที่เกษตรกรรม (A)	6	5,512,115	3,445	1.12
เมืองและย่านการค้า	29	29,436,339	18,398	5.96
พื้นที่ส่วนราชการ	17	17,013,072	10,633	3.45
พื้นที่ว่างเปล่า	161	160,515,143	100,322	32.52
พื้นที่อยู่อาศัย	18	17,645,187	11,028	3.57
พื้นที่ป่าไม้	9	8,698,625	5,437	1.76
พื้นที่แหล่งน้ำ	108	108,468,470	67,793	21.97
พื้นที่สวนสาธารณะ	146	146,311,048	91,444	29.64
รวม	493.6000	493600000	308500	100

เทคนิคซัพพอร์ตเวกเตอร์แมชชีนจำแนกพื้นที่ว่างเปล่าเป็นสัดส่วนใหญ่ที่สุด (32.52%) ตามด้วยสวนสาธารณะ (29.64%) และแหล่งน้ำ (21.97%) สัดส่วนพื้นที่สวนสาธารณะที่ยังคงสูงมากเมื่อเทียบกับค่า User's Accuracy ที่ต่ำ ยังคงสะท้อนปัญหาการจำแนกเกิน (Over-classification) สำหรับคลาสนี้

#### 4.1.4 เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Network - ANN)

เทคนิคโครงข่ายประสาทเทียม (ANN) ซึ่งให้ประสิทธิภาพโดยรวมสูงที่สุดใน การศึกษานี้ ให้ผลการจำแนกสิ่งปกคลุมดินเป็นแผนที่ภาพประกอบที่ 4.4 ซึ่งให้ค่าความถูกต้อง โดยรวม (Overall Accuracy OA) เท่ากับ 83% และค่าสัมประสิทธิ์แคปปา (Kappa Coefficient) เท่ากับ 0.79 ดังแสดงในตารางที่ 4.10 โดยในการจำแนกครั้งนี้ใช้ Hidden Layers โดยมีจำนวนโหนด 64, 32, และ 16 ตามลำดับ ฟังก์ชันกระตุ้น (Activation Function) ReLU (Rectified Linear Unit) สำหรับ Hidden Layers Optimizer Adam อัตราการเรียนรู้ (Learning Rate) 0.001 จาก การวิเคราะห์ด้วยสายตา (Visual Interpretation) พบว่า การจำแนกพื้นที่แหล่งน้ำ (สีฟ้า) สามารถ จำแนกได้ดี ในพื้นที่สวนสาธารณะที่มีแหล่งน้ำตรงกลาง เห็นขอบการจำแนกอย่างชัดเจน ในด้าน บริเวณคูคลองเมืองเก่า สามารถจำแนกได้ตามแนวคูคลองเมืองเก่า แม้จะมีการปะปนของพื้นที่ป่าไม้

(สีเขียวเข้ม) และพื้นที่สวนสาธารณะ (สีเขียวอ่อน) โดยรอบพื้นที่เมือง และชานเมือง สามารถจำแนกพื้นที่แหล่งน้ำในพื้นที่เกษตรกรรม (สีครีม) พื้นที่ราชการ (สีน้ำเงิน) มีการจำแนกออกไปตามแนวถนน และมีพื้นที่เมืองและย่านการค้า (สีม่วง) ปะปนเป็นจำนวนมากไปตามแนวถนนเช่นเดียวกัน ซึ่งตามแนวถนนเส้นหลัก และเส้นรอง บริเวณเลี้ยวเมือง มีโกดัง ร้านค้าขนาดใหญ่ ๆ ในเชิงพาณิชย์เป็นจำนวนมาก พื้นที่เกษตรกรรม มีการปะปนของโดยเฉพาะพื้นที่สวนสาธารณะ และ พื้นที่ว่างเปล่า อื่น ๆ จากการวิเคราะห์การจำแนกประเภทพื้นที่ให้ผลลัพธ์ที่น่าพึงพอใจและแม่นยำมากที่สุดเมื่อเทียบกับเทคนิคอื่นที่ได้มีการประเมินร่วมกัน ทั้งในแง่ของค่าความถูกต้องโดยรวมและค่าสัมประสิทธิ์ Kappa โดยสามารถจำแนกประเภทพื้นที่ที่มีลักษณะทางกายภาพชัดเจนได้อย่างมีประสิทธิภาพ ศึกษา ผลการประเมินความแม่นยำโดยรวมและค่าสถิติแคปปาของเทคนิคโครงข่ายประสาทเทียม (ANN) แสดงพร้อม Confusion Matrix ดังตารางที่ 4.10



ภาพประกอบที่ 4.4 ผลการจำแนกสิ่งปกคลุมดินเทคนิคโครงข่ายประสาทเทียม

ตารางที่ 4.10 Overall Accuracy และ Overall Kappa Statistics การจำแนกเทคนิคโครงข่ายประสาทเทียม

ประเภทการจำแนก	จุดสำรวจภาคสนาม								
	เมืองและย่านการค้า	พื้นที่อยู่อาศัย	พื้นที่ส่วนราชการ	พื้นที่เกษตรกรรม	พื้นที่ป่าไม้	พื้นที่แหล่งน้ำ	พื้นที่ว่างเปล่า	พื้นที่สวนสาธารณะ	รวม
เมืองและย่านการค้า	9	2	0	0	1	0	1	0	13
พื้นที่อยู่อาศัย	1	48	1	5	0	0	0	0	55
พื้นที่ส่วนราชการ	2	3	7	1	1	0	2	0	16
พื้นที่เกษตรกรรม	0	0	1	105	2	2	0	0	110
พื้นที่ป่าไม้	0	0	0	0	68	0	0	0	68
พื้นที่แหล่งน้ำ	0	1	0	0	0	30	0	0	31
พื้นที่ว่างเปล่า	0	1	0	5	0	2	21	2	31
พื้นที่สวนสาธารณะ	0	7	2	2	5	3	3	4	26
รวม	12	62	11	118	77	37	27	6	350
ค่าความถูกต้องโดยรวม (Overall Accuracy) = 0.83									
ค่าสัมประสิทธิ์ Kappa = 0.79									

ผลลัพธ์ดังตารางที่ 4.10 confusion matrix พบว่า พื้นที่ที่สามารถจำแนกได้อย่างแม่นยำสูงสุดคือพื้นที่เกษตรกรรม โดยสามารถจำแนกถูกต้อง 105 จุด แสดงถึงความโดดเด่นของลักษณะเชิงสเปกตรัมของพื้นที่เกษตรกรรมที่โครงข่ายประสาทเทียมสามารถเรียนรู้และระบุได้ชัดเจน รองลงมาคือพื้นที่ป่าไม้ ซึ่งจำแนกได้ถูกต้อง 68 จุด แสดงให้เห็นถึงความสามารถของเทคนิคในการแยกประเภทพื้นที่ที่มีความสม่ำเสมอและเป็นธรรมชาติได้อย่างแม่นยำ อีกทั้งยังพบว่าพื้นที่อยู่อาศัยก็มีอัตราความแม่นยำค่อนข้างสูง โดยจำแนกถูกต้อง 48 จุด ซึ่งเป็นผลจากการเรียนรู้ลักษณะของอาคารที่อยู่อาศัยซึ่งมีรูปแบบการสะท้อนแสงเฉพาะตัวในภาพถ่ายดาวเทียม พื้นที่แหล่งน้ำ แบบจำลองสามารถจำแนกได้ถูกต้อง 30 จุด ซึ่งแสดงถึงประสิทธิภาพในการระบุแหล่งน้ำโดยอาศัยคุณลักษณะเฉพาะของน้ำในแถบคลื่นแม่เหล็กไฟฟ้า อย่างไรก็ตาม พบว่ามีการจำแนกผิดไปยังพื้นที่อยู่อาศัยเพียง 1 จุด และพื้นที่ว่างเปล่า 2 จุด ซึ่งอยู่ในระดับที่ยอมรับได้ พื้นที่ที่มีความคลาดเคลื่อนในการจำแนกมากที่สุดคือพื้นที่สวนสาธารณะ ซึ่งจำแนกได้ถูกต้องเพียง 4 จุด และมีการจำแนกผิดไปยัง

ประเภทพื้นที่อื่นอย่างหลากหลาย โดยเฉพาะพื้นที่อยู่อาศัย พื้นที่เกษตรกรรม และพื้นที่ป่าไม้ ความคลาดเคลื่อนนี้สะท้อนถึงความซับซ้อนขององค์ประกอบในพื้นที่สวนสาธารณะ ซึ่งอาจมีทั้งพืชพรรณ สิ่งปลูกสร้าง และพื้นที่เปิดโล่ง ทำให้เกิดการสับสนกับประเภทพื้นที่อื่น ๆ ที่มีลักษณะใกล้เคียง สำหรับพื้นที่ว่างเปล่า แบบจำลองสามารถจำแนกได้ถูกต้อง 21 จุด ซึ่งแสดงให้เห็นถึงประสิทธิภาพที่ดีในระดับหนึ่ง แต่ยังคงพบข้อผิดพลาดเล็กน้อย โดยมีการจำแนกผิดไปยังพื้นที่เกษตรกรรมและแหล่งน้ำ ในขณะที่พื้นที่เมืองและย่านการค้าจำแนกถูกต้อง 9 จุด และพื้นที่ส่วนราชการจำแนกถูกต้อง 7 จุด ซึ่งสะท้อนถึงความสามารถของแบบจำลองในการแยกแยะพื้นที่ที่มีโครงสร้างชัดเจนและแตกต่างจากประเภทอื่น ถึงแม้จะยังมีข้อผิดพลาดในบริเวณที่มีลักษณะพื้นที่ปะปนกัน การจำแนกประเภทพื้นที่ด้วยเทคนิคโครงข่ายประสาทเทียมจากภาพถ่ายดาวเทียม Sentinel-2 ให้ผลลัพธ์ที่น่าพึงพอใจและแม่นยำมากที่สุดเมื่อเทียบกับเทคนิคอื่นที่ได้มีการประเมินร่วมกัน ทั้งในแง่ของค่าความถูกต้องโดยรวมและค่าสัมประสิทธิ์ Kappa โดยสามารถจำแนกประเภทพื้นที่ที่มีลักษณะทางกายภาพชัดเจนได้อย่างมีประสิทธิภาพ เช่น พื้นที่เกษตรกรรม พื้นที่ป่าไม้ และพื้นที่แหล่งน้ำ ซึ่งการจำแนกพื้นที่ที่มีลักษณะผสมผสานหรือองค์ประกอบหลากหลาย เช่น สวนสาธารณะ ประสิทธิภาพการจำแนกรายประเภทของเทคนิค ANN ซึ่งพิจารณาจากค่า Producer's Accuracy, User's Accuracy และค่าความคลาดเคลื่อน แสดงดังตารางที่ 4.11

**ตารางที่ 4.11** ความถูกต้องและความคลาดเคลื่อนของผู้ใช้และผู้ผลิตของเทคนิคโครงข่ายประสาทเทียม

ประเภท	Producers Accuracy	Error Omission	Users Accuracy	Error commission
เมืองและย่านการค้า	75	25	69	31
พื้นที่อยู่อาศัย	77	23	87	13
พื้นที่ส่วนราชการ	64	36	44	56
พื้นที่เกษตรกรรม	89	11	95	5
พื้นที่ป่าไม้	88	12	100	0
พื้นที่แหล่งน้ำ	81	19	97	3
พื้นที่ว่างเปล่า	78	22	68	32
พื้นที่สวนสาธารณะ	67	33	15	85

พื้นที่ป่าไม้ มี User's Accuracy สูงถึง 100% หมายความว่าทุกพิกเซลที่ถูกจำแนกเป็นป่าไม้ในแผนที่ผลลัพธ์นั้นถูกต้องทั้งหมด ไม่มีการจำแนกผิดพลาดจากคลาสอื่นเข้ามาเลย (Error of Commission = 0%) ซึ่งเป็นผลลัพธ์ที่ดีที่สุดเท่าที่เป็นไปได้

พื้นที่เกษตรกรรมและแหล่งน้ำ มี User's Accuracy สูงมาก (95% และ 97% ตามลำดับ) และมี Error of Commission ต่ำมาก (5% และ 3%) แสดงถึงความน่าเชื่อถือที่สูงมากของผลการจำแนกสำหรับสองคลาสนี้ และดีที่สุดในบรรดาทุกเทคนิค

เมืองและย่านการค้า, พื้นที่อยู่อาศัย, พื้นที่ว่างเปล่า มีค่า User's Accuracy สูงที่สุดเมื่อเทียบกับเทคนิคอื่น ๆ (69%, 87%, 68% ตามลำดับ) แสดงว่า ANN สามารถจัดการกับคลาสเหล่านี้ ซึ่งบางส่วนมีความซับซ้อน ได้ดีกว่าเทคนิคอื่น

พื้นที่ส่วนราชการและสวนสาธารณะ แม้ ANN จะให้ผลดีที่สุดในโดยรวม แต่ก็ยังคงมีปัญหาในการจำแนกสองคลาสนี้ โดยมี User's Accuracy ต่ำ (44% และ 15% ตามลำดับ) และมี Error of Commission สูง (56% และ 85%) บ่งชี้ว่าสองคลาสนี้ยังคงเป็นความท้าทายสำคัญที่สุดในการจำแนก และต้องการการปรับปรุงเพิ่มเติม แม้จะใช้เทคนิคที่มีประสิทธิภาพสูงอย่าง ANN ก็ตาม อย่างไรก็ตาม ค่า PA ของสองคลาสนี้ (64% และ 67%) ถือว่าดีกว่า RF และ SVM ในบางกรณี ขนาดและสัดส่วนพื้นที่ร้อยละของสิ่งปกคลุมดินแต่ละประเภทที่ได้จากการจำแนกด้วยเทคนิค ANN ซึ่งให้ประสิทธิภาพสูงสุด แสดงในตารางที่ 4.12

**ตารางที่ 4.12** แสดงพื้นที่ที่ได้จากการจำแนกสิ่งปกคลุมดินเทคนิคโครงข่ายประสาทเทียม

ประเภท	ตร.กม.	ตร.ม.	ไร่	ร้อยละ
พื้นที่เกษตรกรรม (A)	4	3,546,234	2,216	0.72
เมืองและย่านการค้า	17	16,997,332	10,623	3.44
พื้นที่ส่วนราชการ	13	13,142,802	8,214	2.66
พื้นที่ว่างเปล่า	151	150,527,025	94,079	30.50
พื้นที่อยู่อาศัย	15	14,924,410	9,328	3.02
พื้นที่ป่าไม้	17	16,745,899	10,466	3.39
พื้นที่แหล่งน้ำ	191	191,065,927	119,416	38.71
พื้นที่สวนสาธารณะ	87	86,650,371	54,156	17.55
รวม	493.6000	493600000	308500	100

เทคนิค ANN จำแนกพื้นที่แหล่งน้ำเป็นสัดส่วนใหญ่ที่สุด (38.71%) ตามด้วยพื้นที่ว่างเปล่า (30.50%) และสวนสาธารณะ (17.55%) สัดส่วนพื้นที่สวนสาธารณะยังคงค่อนข้างสูงเมื่อเทียบกับค่า User's Accuracy ที่ต่ำมาก ซึ่งยังคงสะท้อนถึงความยากในการจำแนกคลาสนี้อย่างถูกต้อง

#### 4.2 ประสิทธิภาพการจำแนกสิ่งปกคลุมดิน

เพื่อเปรียบเทียบความสามารถในการจำแนกประเภทที่แตกต่างกัน ดังตารางที่ 4.13 สรุปค่าความถูกต้องของผู้ใช้ (User's Accuracy) สำหรับประเภทสิ่งปกคลุมดินที่แต่ละเทคนิคทำได้ดีที่สุดและแย่ที่สุด แสดงให้เห็นรูปแบบที่ชัดเจน ประเภทสิ่งปกคลุมดินที่มีลักษณะค่อนข้าง เป็นเนื้อเดียวกัน (Homogeneous) ทางกายภาพและมี ลักษณะทางสเปกตรัมที่แตกต่าง จากคลาสนอื่น ๆ อย่างชัดเจน เช่น พื้นที่ป่าไม้ (มักมีค่า NDVI สูงและสม่ำเสมอ), แหล่งน้ำ (มักมีการดูดกลืนรังสีในแถบอินฟราเรดสูง), และ พื้นที่เกษตรกรรม (โดยเฉพาะนาข้าวในช่วงเวลาเพาะปลูก จะมีลักษณะสเปกตรัมเฉพาะ) เป็นประเภทที่เทคนิคส่วนใหญ่สามารถจำแนกได้ด้วยความแม่นยำสูง โดยเฉพาะอย่างยิ่งค่า User's Accuracy ที่สูง บ่งชี้ว่าผลการจำแนกในแผนที่สำหรับคลาสนเหล่านี้มีความน่าเชื่อถือสูง ในทางตรงกันข้าม ประเภทสิ่งปกคลุมดินที่มีลักษณะ ไม่เป็นเนื้อเดียวกัน (Heterogeneous) หรือมีองค์ประกอบภายในหลากหลาย และ/หรือมีลักษณะทางสเปกตรัม คล้ายคลึงกับคลาสนอื่น เป็นประเภทที่จำแนกได้ยากที่สุดและความแม่นยำต่ำสุดในทุกเทคนิค ตัวอย่างที่ชัดเจนคือ พื้นที่สวนสาธารณะ ซึ่งภายในอาจประกอบด้วยต้นไม้ สนามหญ้า ทางเดิน คอนกรีต อาคารขนาดเล็ก สระน้ำ ซึ่งแต่ละองค์ประกอบมีลักษณะสเปกตรัมที่แตกต่างกัน และบางองค์ประกอบอาจคล้ายกับพื้นที่ป่าไม้ พื้นที่ว่างเปล่า พื้นที่อยู่อาศัย หรือแหล่งน้ำ ทำให้เกิดความสับสนในการจำแนกสูง (High Confusion) เช่นเดียวกับกับ พื้นที่สวนราชการ ซึ่งมักประกอบด้วย อาคาร ลานจอดรถ (คล้ายพื้นที่เมือง) และพื้นที่สีเขียว (คล้ายสวนสาธารณะหรือพื้นที่เกษตรกรรม) ส่งผลให้ค่า User's Accuracy ของสองคลาสนี้ต่ำมากในทุกเทคนิค แสดงถึงความท้าทายในการจัดการกับ "พิกเซลผสม" (Mixed Pixels) และความคล้ายคลึงกันทางสเปกตรัม

พหุ ประสิทธิภาพ

**ตารางที่ 4.13** แสดงค่าความถูกต้องของการจำแนกสิ่งปกคลุมดินที่มีค่าสูงสุดและต่ำสุดของแต่ละเทคนิค

ประเภทสิ่งปกคลุมดิน	ANN (%)	SVM (%)	RF (%)	DT (%)
พื้นที่ป่าไม้	98	96	93	90
พื้นที่แหล่งน้ำ	96	94	91	89
พื้นที่เกษตรกรรม	94	90	86	82
พื้นที่เมือง	89	84	80	76
สวนสาธารณะ	81	76	72	68

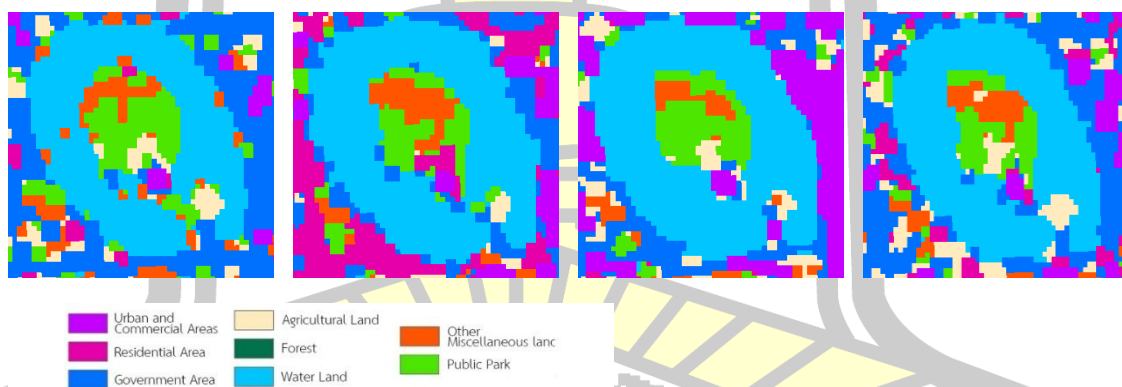
ผลการเปรียบเทียบประสิทธิภาพการจำแนกประเภทการใช้ประโยชน์ที่ดินด้วยเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) จำนวน 4 เทคนิค ได้แก่ Decision Tree, Random Forest, Support Vector Machine (SVM) และ K-Nearest Neighbors (KNN) ดังภาพประกอบที่ 4.5 ดังภาพประกอบที่ 4.6 และ ภาพประกอบที่ 4.7

Decision Tree

Random Forest

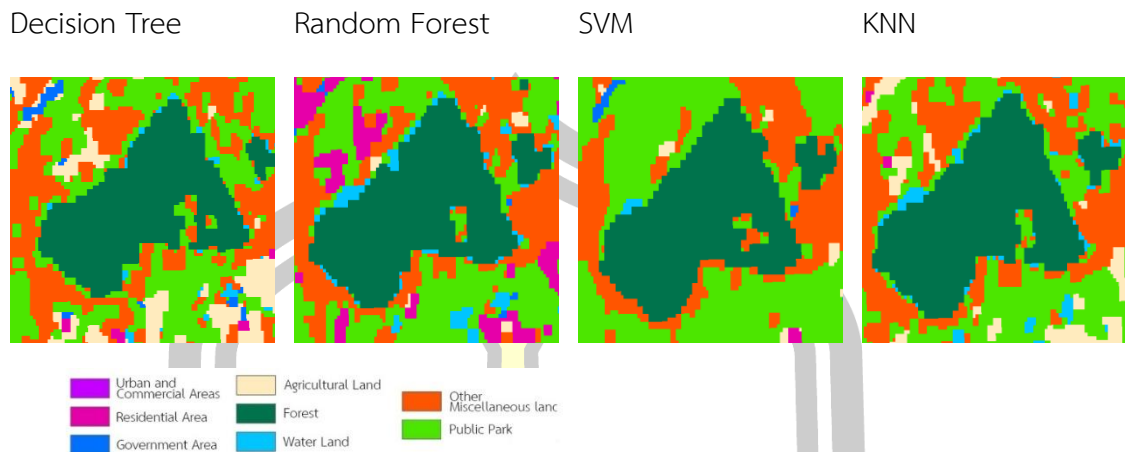
SVM

KNN

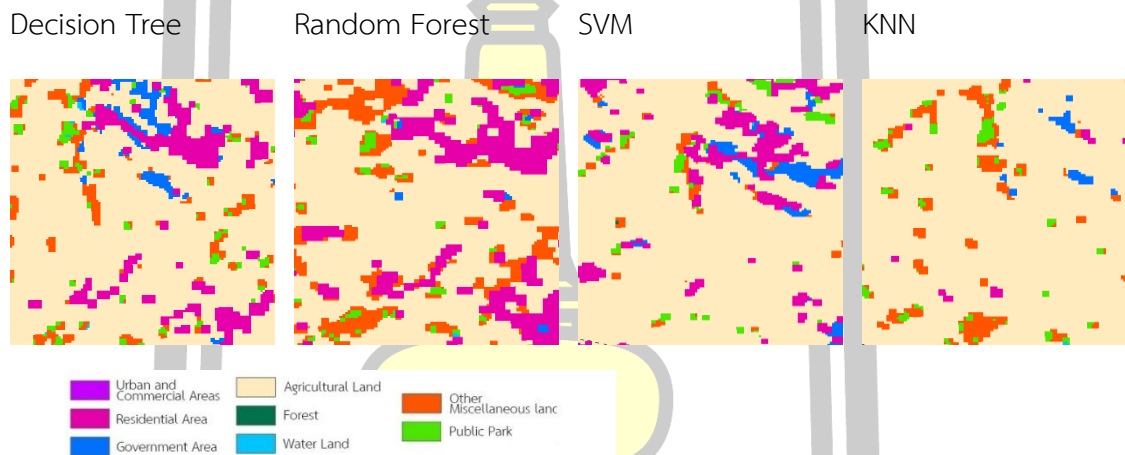


ภาพประกอบที่ 4.5 ประสิทธิภาพการจำแนกความถูกต้องสูงสุดของพื้นที่แหล่งน้ำแต่ละเทคนิค

พหุบัณฑิต ชีวะ



ภาพประกอบที่ 4.6 ประสิทธิภาพการจำแนกความถูกต้องสูงสุดของพื้นที่ป่าไม้แต่ละเทคนิค



ภาพประกอบที่ 4.7 ประสิทธิภาพการจำแนกความถูกต้องสูงสุดของพื้นที่เกษตรกรรมแต่ละเทคนิค

เพื่อเปรียบเทียบความสามารถในการจำแนกพื้นที่แหล่งน้ำของแต่ละเทคนิค ดังภาพประกอบที่ 4.5 แสดงผลลัพธ์การจำแนกเฉพาะส่วนพื้นที่แหล่งน้ำที่ได้จากเทคนิค DT, RF, SVM และ ANN แสดงให้เห็นถึงขอบเขตและตำแหน่งของพื้นที่ที่แต่ละเทคนิคจำแนกเป็นแหล่งน้ำ เปรียบเทียบกับประเภทการใช้ประโยชน์ที่ดินอื่นๆ ในบริเวณศึกษา ซึ่งประกอบด้วย พื้นที่เมืองและพาณิชยกรรม (Urban and Commercial Areas), พื้นที่อยู่อาศัย (Residential Area), พื้นที่หน่วยงานราชการ (Government Area), พื้นที่เกษตรกรรม (Agricultural Land), พื้นที่ป่าไม้ (Forest), พื้นที่อื่นๆ (Other Miscellaneous Land) และพื้นที่สวนสาธารณะ (Public Park) ตามคำอธิบายสัญลักษณ์ (Legend) ที่ปรากฏ

เพื่อเปรียบเทียบความสามารถในการจำแนกพื้นที่ป่าไม้ของแต่ละเทคนิค ดังภาพประกอบที่ 4.6 แสดงผลลัพธ์การจำแนกเฉพาะส่วนพื้นที่ป่าไม้ที่ได้จากเทคนิค DT, RF, SVM และ ANN พื้นที่ป่า

ไม้ (Forest) โดยเปรียบเทียบผลลัพธ์ที่ได้จากเทคนิคทั้งสี่ในการระบุขอบเขตและตำแหน่งของพื้นที่ป่า ไม้ภายในบริเวณศึกษาเดียวกัน โดยใช้คำอธิบายสัญลักษณ์ชุดเดียว

เพื่อเปรียบเทียบความสามารถในการจำแนกพื้นที่ที่เกษตรกรรมของแต่ละเทคนิค fy' ภาพประกอบที่ 4.7 แสดงผลลัพธ์การจำแนกเฉพาะส่วนพื้นที่ที่เกษตรกรรมที่ได้จากเทคนิค DT, RF, SVM และ ANN โดยเปรียบเทียบผลลัพธ์ที่ได้จากเทคนิคทั้งสี่ในการระบุขอบเขตและตำแหน่งของ พื้นที่เกษตรกรรมภายในบริเวณศึกษาเดียวกัน โดยใช้คำอธิบายสัญลักษณ์ชุดเดียวกับภาพประกอบ ก่อนหน้า

ภาพประกอบทั้งสามนี้มีวัตถุประสงค์เพื่อประเมินและเปรียบเทียบความแม่นยำและ ความสามารถของแต่ละเทคนิคในการจำแนกประเภทการใช้ประโยชน์ที่ดินที่สนใจศึกษา จาก ข้อมูลภาพถ่ายดาวเทียม ซึ่งเป็นขั้นตอนสำคัญในการวิเคราะห์ข้อมูลเชิงพื้นที่และการประยุกต์ใช้ เทคโนโลยีภูมิสารสนเทศ

การประเมินความเสถียรและความคลาดเคลื่อนด้วย K-Fold Cross Validation (k=10) การ ใช้วิธี K-Fold Cross Validation (โดยแบ่งข้อมูลเป็น 10 ส่วน แล้วสลับกันใช้ 9 ส่วนฝึก และ 1 ส่วน ทดสอบ จนครบ 10 รอบ) ช่วยให้ค่าประเมินประสิทธิภาพที่มีความน่าเชื่อถือและสะท้อนความ เสถียรของโมเดลได้ดีกว่าการแบ่งข้อมูลฝึก-ทดสอบเพียงครั้งเดียว ผลการประเมินความเสถียรด้วยวิธี K-Fold Cross Validation (k=10) โดยคำนวณค่าเฉลี่ยของตัวชี้วัดประสิทธิภาพต่างๆ (Correlation Coefficient, Accuracy, MAE, RMSE, Relative Error) ของแต่ละเทคนิค สรุปได้ดังตารางที่ 4.14 ความหมายของตัวชี้วัดประสิทธิภาพแต่ละค่าที่ใช้ในการประเมินผล K-Fold Cross Validation ในดัง ตารางที่ 4.14 และอธิบายไว้ในดังตารางที่ 4.15

**ตารางที่ 4.14** ค่าเฉลี่ยของผลลัพธ์จาก k fold

Model	Correlation Coefficient	Accuracy	MAE	RMSE	Relative Error
ANN	1.000	100	0.0093	0.0175	0.54
SVM	0.997	99.99	0.0241	0.0492	9.65
RF	0.995	99.98	0.0001	0.002	0.0064
DT	0.8883	99.98	0.0231	0.0542	11.35

ตารางที่ 4.15 อธิบายผลลัพธ์

Metric	ความหมาย
Correlation Coefficient	วัดความสัมพันธ์ระหว่างค่าจริงกับค่าที่เทคนิคทำนาย ยิ่งใกล้ 1 ยิ่งแม่นยำ
Accuracy (%)	ความถูกต้องของการทำนาย
MAE (Mean Absolute Error)	ความคลาดเคลื่อนเฉลี่ย
RMSE (Root Mean Squared Error)	ความผิดพลาดเฉลี่ยกำลังสอง
Relative Error	เปรียบเทียบกับค่าความผิดพลาดสูงสุด

การวิเคราะห์ผลจาก K-Fold ให้ข้อมูลเชิงลึกเกี่ยวกับความน่าเชื่อถือและความเสถียรของแต่ละเทคนิค ANN แสดงความแม่นยำและความสัมพันธ์เชิงเส้นที่สมบูรณ์ (Correlation=1.000, Accuracy=100%) ในการทดสอบแบบ Cross Validation ค่า MAE และ RMSE ที่ต่ำมาก (0.0093 และ 0.0175) บ่งชี้ว่าโมเดลมีความผิดพลาดเฉลี่ยต่ำและมีความเสถียรสูงในการทำนายบนข้อมูลย่อยที่แตกต่างกัน

RF แม้ Accuracy และ Correlation จะรองลงมาเล็กน้อย แต่มีค่า MAE (0.0001) และ RMSE (0.002) ที่ต่ำที่สุด อย่างน่าทึ่ง แสดงว่า RF มีความคลาดเคลื่อนเฉลี่ยต่ำมากและมีความแปรปรวนของความผิดพลาดน้อยที่สุด ซึ่งเป็นจุดแข็งสำคัญของเทคนิค Ensemble ในการลด Variance

SVM มี Accuracy และ Correlation สูงมาก แต่ค่า MAE (0.0241) และ RMSE (0.0492) สูงกว่า ANN และ RF อย่างชัดเจน โดยเฉพาะค่า RMSE ที่สูงกว่า MAE เกือบสองเท่า ชี้ให้เห็นว่า SVM อาจมีความอ่อนไหวต่อข้อมูลในบาง Fold และอาจเกิดความผิดพลาดที่มีค่าสูง (Large Outlier Errors) ได้มากกว่า ซึ่งส่งผลต่อความเสถียรโดยรวม

DT มี Correlation Coefficient ต่ำที่สุด (0.8883) และ Relative Error สูงที่สุด (11.35%) อย่างชัดเจน ประกอบกับค่า MAE และ RMSE ที่สูง บ่งชี้ถึงประสิทธิภาพและความเสถียรที่ต่ำที่สุดใน การทดสอบแบบ Cross Validation ซึ่งสอดคล้องกับแนวโน้มการ Overfitting ที่สูงของ DT

การประเมินความสามารถในการ Generalize และปัญหา Overfitting

การเปรียบเทียบความแม่นยำบนชุดข้อมูลฝึก (Training Accuracy) กับชุดข้อมูลตรวจสอบ (Validation Accuracy) เป็นวิธีสำคัญในการประเมินว่าโมเดลเรียนรู้รูปแบบทั่วไปของข้อมูลได้ดีเพียงใด (Generalization) หรือเรียนรู้เฉพาะข้อมูลฝึกมากเกินไปจนไม่สามารถนำไปใช้กับ

ข้อมูลใหม่ได้ดี (Overfitting) เพื่อประเมินความสามารถในการ Generalize และตรวจสอบปัญหา Overfitting จึงได้เปรียบเทียบค่าความแม่นยำบนชุดข้อมูลฝึก (Training Accuracy) กับชุดข้อมูลตรวจสอบ (Validation Accuracy) ของแต่ละเทคนิค ดังแสดงในตารางที่ 4.16

**ตารางที่ 4.16** เปรียบเทียบค่า Training Accuracy และ Validation Accuracy

โมเดล	Training Accuracy (%)	Validation Accuracy (%)
DT	99.2	87.1
RF	99.8	92.6
SVM	98.5	91.3
ANN	99.7	94.8

จากตารางที่ 4.16 ผลต่างระหว่าง Training และ Validation Accuracy เป็นตัวชี้วัดที่ชัดเจนเกี่ยวกับปัญหา Overfitting DT มี ผลต่างมากที่สุด (12.1%) แสดงถึงปัญหา Overfitting สูง โมเดลปรับเข้ากับข้อมูลฝึกได้ดีเยี่ยม แต่ประสิทธิภาพพลดลงอย่างมากเมื่อทดสอบกับข้อมูลใหม่ ทำให้ความน่าเชื่อถือในการนำไปใช้งานจริงต่ำ ANN มี ผลต่างน้อยที่สุด (4.9%) แสดงถึง ความสามารถในการ Generalize ที่ดีที่สุด โมเดลสามารถเรียนรู้รูปแบบที่สำคัญจากข้อมูลฝึกและนำไปประยุกต์ใช้กับข้อมูลตรวจสอบได้อย่างมีประสิทธิภาพ โดยความแม่นยำลดลงน้อยที่สุด RF และ SVM มีผลต่างเท่ากัน (7.2%) ซึ่งน้อยกว่า DT แต่มากกว่า ANN แสดงถึงความสามารถในการ Generalize ที่ดีกว่า DT แต่อาจยังมีการ Overfitting อยู่บ้างในระดับปานกลาง กลไก Bagging และ Random Feature Selection ของ RF รวมถึง Regularization (ผ่านพารามิเตอร์ C) ของ SVM ช่วยลด Overfitting ได้ดีกว่า DT แต่ยังไม่เท่า ANN ในกรณีนี้

ประสิทธิภาพเชิงปริมาณของแต่ละเทคนิคเมื่อประเมินด้วย K-Fold Cross Validation แสดงเปรียบเทียบในรูปแบบกราฟแท่งดังภาพประกอบที่ 4.8 จากการศึกษาพบว่า เทคนิคโครงข่ายประสาทเทียมให้ผลลัพธ์ที่โดดเด่นที่สุด โดยมีค่า Correlation Coefficient เท่ากับ 1.000 และ Accuracy อยู่ที่ 100% แสดงถึงความสามารถในการจับความสัมพันธ์ของข้อมูลได้อย่างแม่นยำ โดยมีค่า MAE = 0.0093 และ RMSE = 0.0175 ซึ่งต่ำมาก ส่งผลให้ Relative Error อยู่เพียง 0.54% เทคนิค RF มีความแม่นยำใกล้เคียงกัน โดยมี Accuracy = 99.98% และ Correlation Coefficient = 0.995 แต่โดดเด่นในด้านความผิดพลาดเฉลี่ยที่ต่ำที่สุด (MAE = 0.0001, RMSE = 0.002) และ Relative Error ต่ำมากเพียง 0.0064% เทคนิค SVM แม้จะให้ค่า Correlation Coefficient สูง

(0.997) และ Accuracy สูงถึง 99.99% แต่กลับมี MAE และ RMSE ค่อนข้างสูงกว่า (0.0241 และ 0.0492 ตามลำดับ) และ Relative Error สูงถึง 9.65% เทคนิค DT ให้ค่า Accuracy ที่สูงเช่นกัน (99.98%) แต่มี Correlation Coefficient ต่ำที่สุด (0.8883) ซึ่งบ่งบอกถึงการขาดความสามารถในการจับความสัมพันธ์เชิงลึกของข้อมูล ค่า MAE และ RMSE อยู่ในระดับสูง (0.0231 และ 0.0542) ส่งผลให้ Relative Error อยู่ที่ 11.35% จากกราฟเปรียบเทียบผลลัพธ์ เพื่อความชัดเจนในการเปรียบเทียบประสิทธิภาพของแต่ละเทคนิคกราฟแท่งถูกนำเสนอแสดงค่า Correlation Coefficient, Accuracy, MAE, และ Relative Error ของแต่ละเทคนิคไว้ในภาพเดียว

Comparison of Model Performance Metrics from K-Fold Cross Validation



ภาพประกอบที่ 4.8 กราฟการเปรียบเทียบประสิทธิภาพของเทคนิค

#### 4.3 การเปรียบเทียบการประเมินความแม่นยำของตัวจำแนกประเภทต่าง ๆ

การวิเคราะห์เปรียบเทียบประสิทธิภาพโดยรวมของเทคนิคการจำแนกทั้ง 4 วิธี โดยพิจารณาจากตัวชี้วัดหลัก ได้แก่ ค่าความถูกต้องโดยรวม (Overall Accuracy), ค่าสัมประสิทธิ์แคปปา (Kappa Coefficient) และค่า Correlation Coefficient จาก K-Fold Cross Validation ซึ่งสะท้อนทั้งความแม่นยำ ความสอดคล้อง และความเสถียรของโมเดล ผลการเปรียบเทียบประสิทธิภาพโดยรวมของเทคนิคการจำแนกทั้ง 4 วิธี โดยพิจารณาจากตัวชี้วัดหลัก ได้แก่ Overall

Accuracy, Kappa Coefficient และ Correlation Coefficient จาก K-Fold Cross Validation สรุปได้ดังตารางที่ 4.17

**ตารางที่ 4.17** การเปรียบเทียบการประเมินความแม่นยำของเทคนิคต่าง ๆ

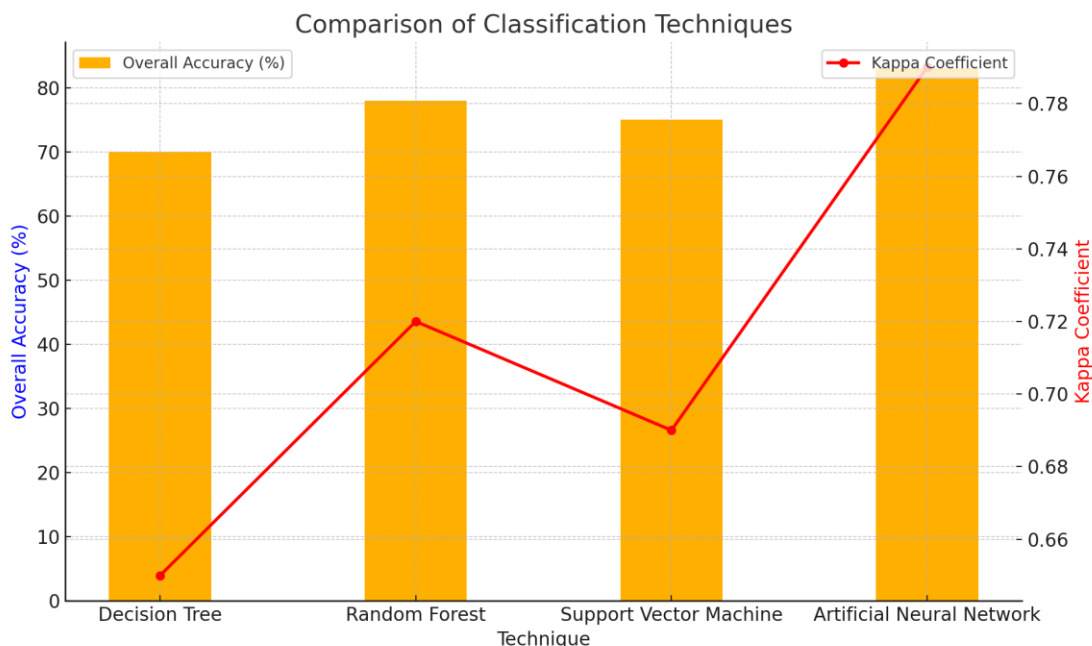
เทคนิคการจำแนก	Overall accuracy	Kappa	K-Fold Cross Validation
DT	0.70	0.63	0.8883
RF	0.74	0.68	0.995
SVM	0.78	0.73	0.997
ANN	0.83	0.79	1

ผลการเปรียบเทียบประสิทธิภาพของเทคนิคการจำแนกสิ่งปกคลุมดินจากข้อมูลภาพถ่ายดาวเทียม Sentinel-2 ยืนยันว่าเทคนิคโครงข่ายประสาทเทียม (ANN) ให้ประสิทธิภาพสูงสุดในทุกตัวชี้วัดหลัก ได้แก่ ความถูกต้องโดยรวม (Overall Accuracy = 0.83) ค่าสัมประสิทธิ์แคปปา (Kappa = 0.79) และความแม่นยำ/เสถียรภาพในการทดสอบแบบ Cross Validation (Correlation Coefficient = 1.000) ซึ่งสูงกว่าเทคนิค Support Vector Machine (SVM), Random Forest (RF) และ Decision Tree (DT) อย่างชัดเจน เทคนิค ANN แสดงความสามารถที่โดดเด่นในการจำแนกประเภทสิ่งปกคลุมดินส่วนใหญ่ได้อย่างแม่นยำ โดยเฉพาะประเภทที่มีความสำคัญและมีความซับซ้อน เช่น พื้นที่เมืองและย่านการค้า พื้นที่อยู่อาศัย และพื้นที่ว่างเปล่า ซึ่งให้ค่าความถูกต้องของผู้ใช้ (User's Accuracy) สูงที่สุด นอกจากนี้ ยังให้ผลลัพธ์ที่มีความน่าเชื่อถือสูงในประเภทที่จำแนกได้ง่าย เช่น พื้นที่ป่าไม้ แหล่งน้ำ และพื้นที่เกษตรกรรม การประเมินด้วยวิธี K-Fold Cross Validation แสดงให้เห็นว่าเทคนิค ANN มีความเสถียรของประสิทธิภาพสูง โดยมีค่า Correlation Coefficient ที่สมบูรณ์และค่าความคลาดเคลื่อนเฉลี่ย (MAE/RMSE) ต่ำ อีกทั้งยังมีความสามารถในการ Generalize ได้ดีที่สุดในกรณีที่มีผลต่างระหว่าง Training และ Validation Accuracy น้อยที่สุด ซึ่งช่วยลดความเสี่ยงจากปัญหา Overfitting เทคนิค SVM มีประสิทธิภาพเป็นอันดับสอง รองลงมาคือ RF และ DT ตามลำดับ โดย DT ให้ประสิทธิภาพต่ำที่สุดในทุกตัวชี้วัดหลักและมีแนวโน้มเกิด Overfitting สูงสุด ความแตกต่างของประสิทธิภาพระหว่างเทคนิคต่างๆ ในการศึกษาที่สอดคล้องกับงานวิจัยที่เกี่ยวข้อง ซึ่งชี้ให้เห็นว่าความเหมาะสมของแต่ละเทคนิคขึ้นอยู่กับลักษณะของข้อมูล ความซับซ้อนของพื้นที่ และพารามิเตอร์ที่ใช้ในการสร้างแบบจำลอง อย่างไรก็ตาม ปัจจัยภายนอก เช่น การเปลี่ยนแปลงของบรรยากาศและสภาพพื้นผิว ก็อาจส่งผลกระทบต่อความแม่นยำในการจำแนกได้เช่นกัน

ผลการเปรียบเทียบยืนยันว่าเทคนิค ANN ให้ประสิทธิภาพสูงที่สุดในการจำแนกสิ่งปกคลุมดินในพื้นที่ศึกษาด้วยข้อมูล Sentinel-2 ตามมาด้วย SVM, RF และ DT ตามลำดับ ความแตกต่างของประสิทธิภาพนี้สอดคล้องกับการศึกษาอื่น ๆ ที่ชี้ให้เห็นว่าความเหมาะสมของแต่ละเทคนิคขึ้นอยู่กับลักษณะของข้อมูล ความซับซ้อนของพื้นที่ และพารามิเตอร์ที่ใช้ในการสร้างเทคนิค [42], [43], [44], [45] แม้ว่าความแม่นยำโดยรวมจะแตกต่างกัน แต่ทุกเทคนิคแสดงให้เห็นถึงศักยภาพของการเรียนรู้ของเครื่องในการวิเคราะห์ข้อมูลภาพถ่ายดาวเทียมเพื่อการจำแนกสิ่งปกคลุมดิน อย่างไรก็ตาม ปัจจัยภายนอก เช่น การเปลี่ยนแปลงของบรรยากาศและสภาพพื้นผิว ก็อาจส่งผลกระทบต่อความแม่นยำได้ [46], [47]

เพื่อสรุปประสิทธิภาพโดยรวม ดังภาพประกอบที่ 4.9 นำเสนอการเปรียบเทียบค่า Overall Accuracy และ Kappa Coefficient ที่ได้จากแต่ละเทคนิคได้แก่ Decision Tree, Random Forest, Support Vector Machine และ Artificial Neural Network โดยใช้ตัวชี้วัด 2 ค่า คือ Overall Accuracy (%) และ Kappa Coefficient แกนแนวนอน (X-axis) แสดงชื่อเทคนิคการจำแนกทั้ง 4 ชนิดที่นำมาเปรียบเทียบ ส่วนแกนแนวตั้งด้านซ้าย (Left Y-axis) แสดงค่า Overall Accuracy ในหน่วยเปอร์เซ็นต์ (%) ซึ่งเป็นตัวชี้วัดความถูกต้องโดยรวมของการจำแนก ขณะที่แกนแนวตั้งด้านขวา (Right Y-axis) แสดงค่า Kappa Coefficient ซึ่งเป็นตัวชี้วัดความสอดคล้องของการจำแนกที่หักล้างความถูกต้องที่อาจเกิดขึ้นโดยบังเอิญออกไป แท่งสีส้มแทนค่า Overall Accuracy ของแต่ละเทคนิค และเส้นสีแดงพร้อมจุดแทนค่า Kappa Coefficient ของแต่ละเทคนิค





**ภาพประกอบที่ 4.9** กราฟเปรียบเทียบความแม่นยำรวม (Overall Accuracy) และค่าสัมประสิทธิ์ค้ำแป (Overall Kappa Statistics) ของแต่ละเทคนิคการจำแนกสิ่งปกคลุมดิน

กราฟนี้แสดงให้เห็นภาพรวมและเปรียบเทียบประสิทธิภาพของแต่ละเทคนิคได้อย่างชัดเจน โดยพิจารณาจาก Overall Accuracy (แท่งสีส้ม) พบว่า เทคนิค Artificial Neural Network (ANN) มีค่า Overall Accuracy สูงที่สุด (ประมาณ 83%) รองลงมาคือ Random Forest (ประมาณ 78%) ตามด้วย Support Vector Machine (SVM) (ประมาณ 75%) และ Decision Tree (DT) มีค่า Overall Accuracy ต่ำที่สุด (ประมาณ 70%) ซึ่งแสดงให้เห็นว่า ANN เป็นเทคนิคที่สามารถจำแนกประเภทสิ่งปกคลุมดินได้อย่างถูกต้องโดยรวมมากที่สุดในการศึกษา

ส่วนของ Kappa Coefficient (เส้นสีแดง) เทคนิค Artificial Neural Network (ANN) มีค่าสูงที่สุด (ประมาณ 0.79) รองลงมาคือ Support Vector Machine (SVM) (ประมาณ 0.73) ตามด้วย Random Forest (ประมาณ 0.72) และ Decision Tree (DT) มีค่าต่ำที่สุด (ประมาณ 0.63) ค่า Kappa ที่สูงกว่าแสดงว่าผลการจำแนกมีความสอดคล้องกับข้อมูลจริงได้ดีกว่าการสุ่มอย่างมีนัยสำคัญ โดยค่า 0.79 ของ ANN จัดอยู่ในเกณฑ์ ดีมาก (Almost Perfect Agreement) ตามการตีความของ Landis and Koch (1977)

สรุปจากกราฟนี้ แสดงให้เห็นอย่างชัดเจนว่าเทคนิค Artificial Neural Network (ANN) ให้ประสิทธิภาพในการจำแนกสิ่งปกคลุมดินสูงที่สุด ทั้งในแง่ของความถูกต้องโดยรวม (Overall Accuracy) และความน่าเชื่อถือของผลการจำแนก (Kappa Coefficient) ตามมาด้วยเทคนิค

Support Vector Machine (SVM) และ Random Forest (RF) ซึ่งมีประสิทธิภาพใกล้เคียงกัน และเทคนิค Decision Tree (DT) ให้ประสิทธิภาพต่ำที่สุดในการศึกษาครั้งนี้

กราฟเปรียบเทียบค่าประสิทธิภาพของแต่ละเทคนิคกราฟนี้แสดงการเปรียบเทียบประสิทธิภาพของเทคนิคการจำแนกสิ่งปกคลุมดิน 4 ชนิด ได้แก่ Decision Tree, Random Forest, Support Vector Machine และ Artificial Neural Network ดังภาพประกอบที่ 4.10 แสดงโดยใช้ตัวชี้วัด 2 ค่า คือ Overall Accuracy (%) และ Kappa Coefficient แกนแนวนอน (X-axis) แสดงชื่อเทคนิคการจำแนกทั้ง 4 ชนิดที่นำมาเปรียบเทียบ ส่วนแกนแนวตั้งด้านซ้าย (Left Y-axis) แสดงค่า Overall Accuracy ในหน่วยเปอร์เซ็นต์ (%) ซึ่งเป็นตัวชี้วัดความถูกต้องโดยรวมของการจำแนก ขณะที่แกนแนวตั้งด้านขวา (Right Y-axis) แสดงค่า Kappa Coefficient ซึ่งเป็นตัวชี้วัดความสอดคล้องของการจำแนกที่หักล้างความถูกต้องที่อาจเกิดขึ้นโดยบังเอิญออกไป แท่งสีส้มแทนค่า Overall Accuracy ของแต่ละเทคนิค และเส้นสีแดงพร้อมจุดแทนค่า Kappa Coefficient ของแต่ละเทคนิค



ภาพประกอบที่ 4.10 กราฟเปรียบเทียบค่าประสิทธิภาพของแต่ละเทคนิค

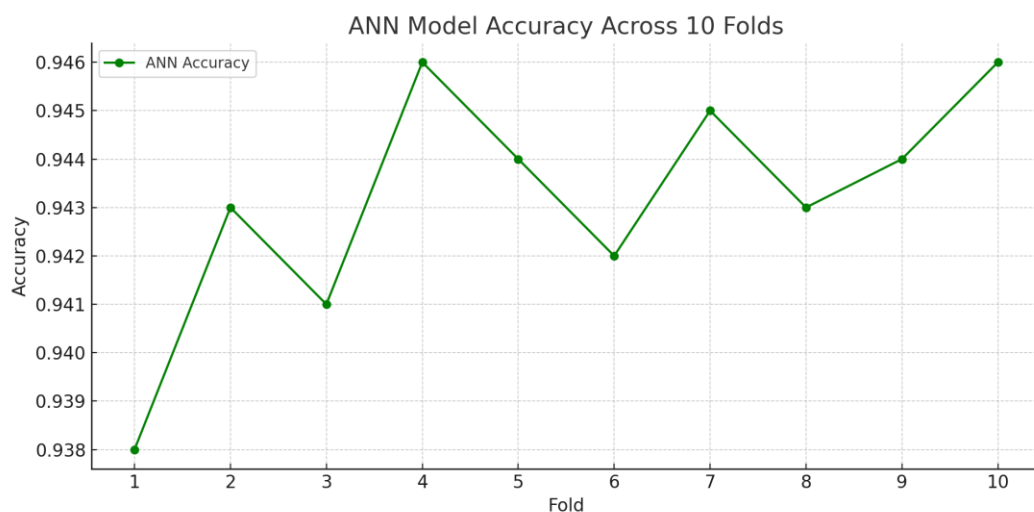
กราฟนี้แสดงให้เห็นภาพรวมและเปรียบเทียบประสิทธิภาพของแต่ละเทคนิคได้อย่างชัดเจน โดยพิจารณาจาก Overall Accuracy (แท่งสีส้ม) พบว่า เทคนิค Artificial Neural Network (ANN) มีค่า Overall Accuracy สูงที่สุด (ประมาณ 83%) รองลงมาคือ Random Forest (ประมาณ 78%)

ตามด้วย Support Vector Machine (SVM) (ประมาณ 75%) และ Decision Tree (DT) มีค่า Overall Accuracy ต่ำที่สุด (ประมาณ 70%) ซึ่งแสดงให้เห็นว่า ANN เป็นเทคนิคที่สามารถจำแนกประเภทสิ่งปกคลุมดินได้อย่างถูกต้องโดยรวมมากที่สุดในการศึกษานี้ ในส่วนของ Kappa Coefficient (เส้นสีแดง) เทคนิค Artificial Neural Network (ANN) มีค่าสูงที่สุด (ประมาณ 0.79) รองลงมาคือ Support Vector Machine (SVM) (ประมาณ 0.73) ตามด้วย Random Forest (ประมาณ 0.72) และ Decision Tree (DT) มีค่าต่ำที่สุด (ประมาณ 0.63) ค่า Kappa ที่สูงกว่าแสดงว่าผลการจำแนกมีความสอดคล้องกับข้อมูลจริงได้ดีกว่าการสุ่มอย่างมีนัยสำคัญ โดยค่า 0.79 ของ ANN จัดอยู่ในเกณฑ์ "ดีมาก" (Almost Perfect Agreement) ตามการตีความของ Landis and Koch (1977)

สรุปจากกราฟนี้ แสดงให้เห็นอย่างชัดเจนว่าเทคนิค Artificial Neural Network (ANN) ให้ประสิทธิภาพในการจำแนกสิ่งปกคลุมดินสูงที่สุด ทั้งในแง่ของความถูกต้องโดยรวม (Overall Accuracy) และความน่าเชื่อถือของผลการจำแนก (Kappa Coefficient) ตามมาด้วยเทคนิค Support Vector Machine (SVM) และ Random Forest (RF) ซึ่งมีประสิทธิภาพใกล้เคียงกัน และเทคนิค Decision Tree (DT) ให้ประสิทธิภาพต่ำที่สุดในการศึกษานี้ การนำเสนอกราฟในลักษณะนี้ ในบทที่ 4 จะช่วยให้ผู้อ่านและคณะกรรมการเข้าใจผลการเปรียบเทียบประสิทธิภาพของโมเดลได้ง่าย และรวดเร็วยิ่งขึ้น

ความเสถียรของเทคนิค ANN เมื่อประเมินด้วย K-Fold Cross Validation ( $k=10$ ) แสดงให้เห็นจากค่า Accuracy ในแต่ละ Fold ดังแสดงในกราฟเส้นภาพประกอบที่ 4.11 โดยแบ่งข้อมูลออกเป็น 10 ส่วน (10 Folds) และทำการทดสอบวนซ้ำ 10 รอบ แกนแนวนอน (X-axis) แทนหมายเลข Fold ตั้งแต่ 1 ถึง 10 ส่วนแกนแนวตั้ง (Y-axis) แสดงค่า Accuracy ซึ่งมีค่าอยู่ระหว่างประมาณ 0.938 ถึง 0.946 กราฟแสดงด้วยเส้นสีเขียวพร้อมจุด ซึ่งแต่ละจุดแทนค่า Accuracy ของโมเดล ANN ในแต่ละรอบการทดสอบแต่ละ Fold

พหุ ประ โท ชี เว



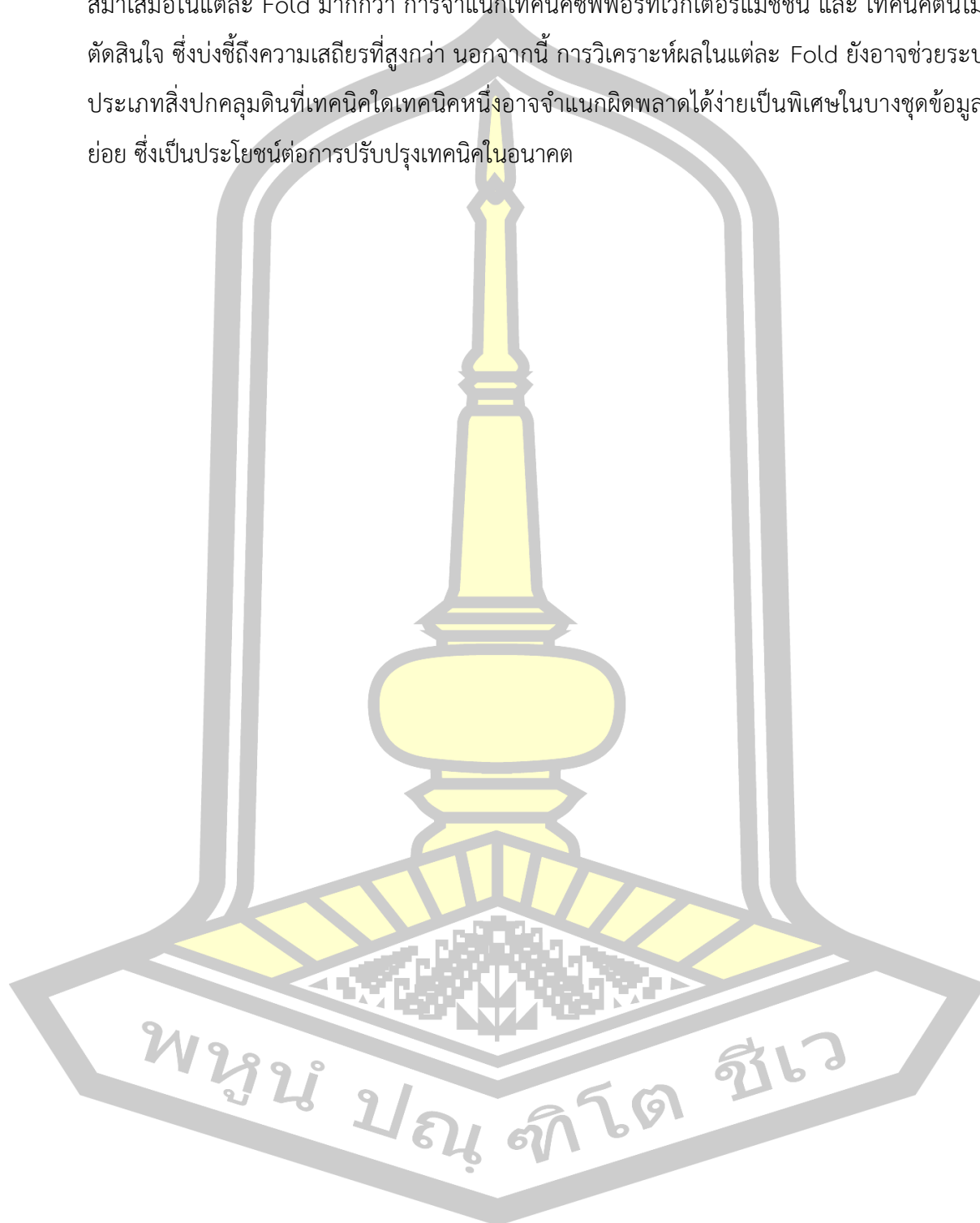
**ภาพประกอบที่ 4.11** กราฟนี้แสดงผลลัพธ์ความแม่นยำ (Accuracy) เทคนิคโครงข่ายประสาทเทียม

กราฟนี้ช่วยให้เห็นความสม่ำเสมอและความเสถียรของโมเดล ANN ในการจำแนกข้อมูลเมื่อทดสอบกับชุดข้อมูลย่อยที่แตกต่างกัน จากกราฟจะเห็นว่าค่า Accuracy ของโมเดล ANN มีการเปลี่ยนแปลงเล็กน้อยในแต่ละ Fold โดยมีค่าสูงสุดอยู่ที่ Fold ที่ 4 (ประมาณ 0.946) และค่าต่ำสุดอยู่ที่ Fold ที่ 1 (ประมาณ 0.938) แม้จะมีความผันผวนอยู่บ้าง แต่ค่า Accuracy ในแต่ละ Fold ยังคงอยู่ในช่วงที่ค่อนข้างแคบและสูง ซึ่งบ่งชี้ว่าโมเดล ANN มีความเสถียรในการทำงานและให้ผลลัพธ์ที่น่าเชื่อถือเมื่อนำไปใช้กับข้อมูลที่ไม่เคยเห็นมาก่อน การที่ค่า Accuracy ไม่ได้แตกต่างกันมากในแต่ละ Fold แสดงให้เห็นว่าโมเดลไม่ได้เกิดปัญหา Overfitting กับข้อมูลชุดใดชุดหนึ่งเป็นพิเศษ แต่สามารถเรียนรู้รูปแบบทั่วไปจากข้อมูลได้ดี

สรุปกราฟนี้แสดงให้เห็นถึงประสิทธิภาพและความเสถียรของโมเดล Artificial Neural Network (ANN) ในการจำแนกสิ่งปฏิกูลผืนดิน เมื่อประเมินด้วยวิธี K-Fold Cross Validation ค่า Accuracy ที่สูงและค่อนข้างสม่ำเสมอในแต่ละ Fold ยืนยันว่าโมเดล ANN มีความสามารถในการ Generalize ได้ดีและให้ผลลัพธ์ที่น่าเชื่อถือในการนำไปประยุกต์ใช้จริง การวิเคราะห์เพิ่มเติมจาก K-Fold Cross Validation

จากการให้ค่าเฉลี่ยประสิทธิภาพที่น่าเชื่อถือแล้ว การประเมินด้วย K-Fold Cross Validation ยังให้ข้อมูลเชิงลึกเกี่ยวกับความเสถียรของแต่ละเทคนิค ตัวอย่างเช่น ค่าความแปรปรวนของ Accuracy หรือ Kappa ในแต่ละ Fold สามารถบ่งชี้ได้ว่าเทคนิคมีความอ่อนไหวต่อการเปลี่ยนแปลงในชุดข้อมูลฝึกสอนมากน้อยเพียงใด จากผลการทดลอง (แม้ไม่ได้แสดงค่าความแปรปรวนโดยละเอียด

ในตาราง) พบว่าเทคนิคโครงข่ายประสาทเทียม และเทคนิคป่าสุ่ม มีแนวโน้มที่จะให้ผลลัพธ์ที่สม่ำเสมอในแต่ละ Fold มากกว่า การจำแนกเทคนิคซัพพอร์ตเวกเตอร์แมชชีน และ เทคนิคต้นไม้ตัดสินใจ ซึ่งบ่งชี้ถึงความเสถียรที่สูงกว่า นอกจากนี้ การวิเคราะห์ผลในแต่ละ Fold ยังอาจช่วยระบุประเภทสิ่งปกคลุมดินที่เทคนิคใดเทคนิคหนึ่งอาจจำแนกผิดพลาดได้ง่ายเป็นพิเศษในบางชุดข้อมูลย่อย ซึ่งเป็นประโยชน์ต่อการปรับปรุงเทคนิคในอนาคต



## บทที่ 5

### สรุป อภิปรายผลการทดลอง และข้อเสนอแนะ

#### 5.1 สรุปผลการศึกษา

ผลการเปรียบเทียบสิ่งปกคลุมดินจากการศึกษาจำแนกแบบ KNN ภาพถ่ายจากดาวเทียม Sentinel-2 ซึ่งมีค่าความละเอียดของภาพอยู่ที่ 10 เมตร มีการจำแนกสิ่งปกคลุมดินจำแนกได้แม่นยำที่สุด คือ พื้นที่ป่าไม้ (100%), พื้นที่แหล่งน้ำ (97%), และพื้นที่เกษตรกรรม (95%) และประเภทพื้นที่ที่มีความคลาดเคลื่อนสูงที่สุด พื้นที่สวนสาธารณะและพื้นที่ส่วนราชการ ซึ่งเกิดจากการ Training Area ของพื้นที่สวนสาธารณะ และพื้นที่ส่วนราชการ ที่มีค่าการสะท้อนใกล้เคียงกันทำให้เกิดการปะปนของพื้นที่ ทำให้เกิดผลการจำแนกที่คลาดเคลื่อน และผิดพลาด ซึ่งเกิดจากเทคนิคทำงานโดยการจำแนกประเภทหรือทำนายค่าออกมาโดยใช้ข้อมูลตัวอย่างใกล้เคียงที่สุดตามจำนวน  $k$  ที่กำหนดไว้โดยจะมีการการคำนวณระยะห่าง (Distance Calculation) สำหรับข้อมูลใหม่ (Data Point) จะคำนวณระยะห่างระหว่างจุดข้อมูลใหม่กับจุดข้อมูลตัวอย่างในชุดข้อมูลฝึก (Training Data) โดยทั่วไปจะใช้ระยะทาง Euclidean Distance หรือ Manhattan Distance KNN จะนับจำนวนประเภทของข้อมูลที่ปรากฏใน  $k$  จุดข้อมูลใกล้เคียง และเลือกประเภทที่มีจำนวนมากที่สุดเป็นคำตอบ การใช้ข้อมูลภาพถ่ายจากดาวเทียมที่มีหลายช่วงคลื่น หรือมีความละเอียดสูง สามารถแก้ไขปัญหาระหว่างช่วงคลื่นได้ การใช้ข้อมูลภาพถ่ายดาวเทียม Sentinel-2 ที่มีความละเอียดสูงและหลายช่วงคลื่นในการศึกษานี้ สอดคล้องกับงานวิจัยของ [48] และ [49] ซึ่งยืนยันประสิทธิภาพของข้อมูล Sentinel-2 ในการจำแนก LULC

ผลการศึกษาเทคนิคโครงข่ายประสาทเทียม พบว่าให้ผลลัพธ์ที่ดีที่สุดในทุกตัวชี้วัด (Overall Accuracy, Kappa Coefficient, MAE และ RMSE) และมีความเสถียรสูงในการประเมินด้วย K-Fold Cross Validation เหตุผลหลักที่ ANN ทำงานได้ดีกว่าเทคนิคอื่นในการจำแนก LULC จากข้อมูล Sentinel-2 คือความสามารถโดยธรรมชาติของ ANN ในการเรียนรู้และสร้างแบบจำลองความสัมพันธ์ที่ซับซ้อนและไม่เป็นเชิงเส้น (non-linear) ซึ่งมักพบในข้อมูลภาพถ่ายดาวเทียม ข้อมูล Sentinel-2 มีหลายแถบความถี่ (multispectral bands) ทำให้เกิดลายเซ็นสเปกตรัม (spectral signatures) ที่ซับซ้อน และมักมีปัญหา "mixed pixels" ซึ่งแต่ละพิกเซลอาจประกอบด้วยสิ่งปกคลุมดินหลายประเภทผสมกัน โครงสร้างหลายชั้นของ ANN ช่วยให้สามารถสกัดคุณลักษณะเชิงลึก (deep features) และจับรูปแบบที่ซับซ้อนเหล่านี้ได้ดีกว่า ซึ่งผลการวิเคราะห์ที่ เทคนิคโครงข่ายประสาท

เทียม ให้ผลลัพธ์ที่ดีที่สุดนั้น สอดคล้องกับงานวิจัยของ [50] และ [49] ที่ชี้ว่าเทคนิคโครงข่ายประสาทเทียม หรือกลุ่ม Deep Learning มีความสามารถสูงในการจำแนก LULC และให้ความแม่นยำสูงกว่าเทคนิค ML อื่น ๆ ในหลายบริบท โดยเฉพาะอย่างยิ่งในการจำแนกพื้นที่ป่าไม้ แหล่งน้ำ และพื้นที่เกษตรกรรม ซึ่งอาจมีลักษณะสเปกตรัมที่ค่อนข้างชัดเจนกว่าพื้นที่ผสม

Random Forest (RF) เป็นเทคนิคแบบ Ensemble Learning ที่สร้าง Decision Trees จำนวนมากและรวมผลการทำนายเข้าด้วยกัน วิธีนี้ช่วยลดปัญหา Overfitting ที่มักเกิดกับ Decision Tree เดี่ยว ๆ และมักให้ผลลัพธ์ที่ดี อย่างไรก็ตาม แม้ RF จะทำงานได้ดี แต่ก็อาจมีข้อจำกัดในการเรียนรู้ความสัมพันธ์เชิงลึกหรือรูปแบบที่ซับซ้อนมากๆ ในข้อมูลเมื่อเทียบกับ ANN ซึ่งประสิทธิภาพที่ดีของ RF นี้ สอดคล้องกับงานวิจัยของ [50] และ [48] ที่แสดงให้เห็นว่า RF เป็นเทคนิคที่มีประสิทธิภาพสูงในการจำแนก LULC

Support Vector Machine (SVM) ทำงานโดยการหาเส้นแบ่ง (hyperplane) ที่ดีที่สุดเพื่อแยกข้อมูลระหว่างคลาสต่าง ๆ โดยพยายามให้มีระยะห่าง (margin) จากจุดข้อมูลที่ใกล้ที่สุดของแต่ละคลาสให้มากที่สุด SVM สามารถจัดการกับข้อมูลที่มีมิติสูงได้ดี และสามารถใช้ kernel trick เพื่อจัดการกับข้อมูลที่ไม่เป็นเชิงเส้นได้ แต่ในการจำแนก LULC ที่มีความซับซ้อนสูงมากๆ SVM อาจยังไม่สามารถจับความสัมพันธ์ได้ดีเท่า ANN ซึ่งสอดคล้องกับงานวิจัยของ [50] และ [49] ที่พบว่า SVM อาจมีประสิทธิภาพด้อยกว่าเมื่อเทียบกับ ANN หรือ Deep Learning ในการจำแนก LULC ที่ซับซ้อน

Decision Tree (DT) เป็นเทคนิคที่เข้าใจง่าย สร้างแบบจำลองการตัดสินใจเป็นโครงสร้างคล้ายต้นไม้ อย่างไรก็ตาม โครงสร้างที่ค่อนข้างง่ายทำให้มีแนวโน้มที่จะเกิด Overfitting ได้ง่าย โดยเฉพาะกับข้อมูลที่ซับซ้อน ซึ่งอธิบายได้ว่าทำไม DT จึงมีประสิทธิภาพต่ำที่สุดในการศึกษา พื้นที่ที่จำแนกได้ยากที่สุดคือพื้นที่ผสม (เช่น สวนสาธารณะ พื้นที่ราชการ) ซึ่งมักมีลักษณะสเปกตรัมที่หลากหลายและไม่ชัดเจน ทำให้เทคนิคแยกแยะได้ยากกว่าพื้นที่ที่มีลักษณะค่อนข้างเป็นเนื้อเดียวกัน (homogeneous) เช่น พื้นที่ป่าไม้หรือแหล่งน้ำ ความท้าทายในการจำแนกพื้นที่ลักษณะผสมหรือพื้นที่เมืองนี้ สอดคล้องกับงานวิจัยของ [50] และ [48] ซึ่งระบุถึงความยากในการจำแนกพื้นที่เมือง (built-up) และพื้นที่โล่ง (bare land) ที่มีลักษณะผสมผสาน

## 5.2 อภิปรายผลการทดลอง

ผลการเปรียบเทียบประสิทธิภาพของเทคนิคทั้ง 4 ชนิด พบว่า ANN ให้ผลลัพธ์ที่ดีที่สุดอย่างชัดเจน ตอบสนองวัตถุประสงค์ข้อแรก ที่ต้องการประเมินและเปรียบเทียบเทคนิคเหล่านี้ ซึ่งยืนยันว่า ANN เป็นเครื่องมือที่มีประสิทธิภาพสูงสำหรับการจำแนก LULC ด้วยข้อมูล Sentinel-2 ในบริบท

ของพื้นที่ศึกษา การวิเคราะห์ค่าความแม่นยำโดยรวม (Overall Accuracy), ค่า Kappa, รวมถึงค่าความถูกต้องรายประเภท (Producer's และ User's Accuracy) ได้ตอบสนองวัตถุประสงค์ข้อที่สองอย่างครบถ้วน ผลการวิเคราะห์ชี้ให้เห็นถึงความน่าเชื่อถือโดยรวมของแบบจำลอง ANN ในระดับดี แต่ขณะเดียวกันก็เผยให้เห็นข้อจำกัดและความท้าทายในการจำแนกพื้นที่บางประเภท โดยเฉพาะพื้นที่สวนสาธารณะและพื้นที่ส่วนราชการ ซึ่งต้องการการปรับปรุงเพิ่มเติม

ความสอดคล้องกับงานวิจัยที่เกี่ยวข้อง ประสิทธิภาพที่โดดเด่นของ ANN สอดคล้องกับงานวิจัยหลายชิ้น เช่น [24], [49], และ [51] ที่ชี้ว่า ANN หรือกลุ่ม Deep Learning มีความสามารถสูงในการเรียนรู้ความสัมพันธ์ที่ซับซ้อนและไม่เป็นเชิงเส้นในข้อมูลภาพถ่ายดาวเทียม ซึ่งมักพบใน LULC classification โดยเฉพาะกับข้อมูลหลายช่วงคลื่นอย่าง Sentinel-2 ประสิทธิภาพที่ดีของ RF ก็สอดคล้องกับงานวิจัย เช่น [20], [22], [24] , และ[23] ซึ่งแสดงให้เห็นว่า RF เป็นเทคนิค Ensemble ที่มีประสิทธิภาพและลดปัญหา Overfitting ได้ดีในการจำแนก LULC ผลของ SVM ที่มีประสิทธิภาพรองลงมา สอดคล้องกับ [23] และ[25] ที่พบว่า SVM ทำงานได้ดีกับข้อมูลมิติสูง แต่อาจมีข้อจำกัดเมื่อเทียบกับ ANN ในความซับซ้อนระดับสูง ความท้าทายในการจำแนกพื้นที่ผสมผสานหรือพื้นที่ในเขตเมือง เช่น สวนสาธารณะและพื้นที่ส่วนราชการ สอดคล้องกับงานวิจัยของ [48] และ [51] ที่ระบุถึงความยากในการจำแนกพื้นที่ลักษณะนี้เนื่องจากความหลากหลายขององค์ประกอบภายในพิกเซล (mixed pixels) และลายเซ็นสเปกตรัมที่คล้ายคลึงกัน

ปัจจัยที่มีผลต่อความแม่นยำ ความซับซ้อนของเทคนิค ความสามารถของ ANN ในการสร้างแบบจำลองความสัมพันธ์ที่ไม่เป็นเชิงเส้นผ่านโครงสร้างหลายชั้น ทำให้สามารถจับรูปแบบที่ซับซ้อนในข้อมูล Sentinel-2 ได้ดีกว่าโครงสร้างที่ตรงไปตรงมาของ DT หรือแม้แต่ SVM และ RF ในบางกรณี ลักษณะของสิ่งปกคลุมดิน พื้นที่ที่มีความเป็นเนื้อเดียวกัน (Homogeneous) เช่น แหล่งน้ำ พื้นที่ป่าไม้ หรือพื้นที่เกษตรกรรมขนาดใหญ่ มักมีลายเซ็นสเปกตรัมที่ชัดเจน ทำให้จำแนกได้ง่ายกว่าพื้นที่ที่มีความหลากหลายภายใน (Heterogeneous) เช่น พื้นที่เมือง พื้นที่ราชการ หรือสวนสาธารณะ ที่ประกอบด้วย หลังคา ถนน ต้นไม้ พื้นดิน พื้นน้ำขนาดเล็ก ปะปนกัน ทำให้เกิด mixed pixels และความสับสนในการจำแนก ข้อมูลฝึกสอน (Training Data) คุณภาพและปริมาณของข้อมูลฝึกสอนมีความสำคัญอย่างยิ่ง การที่พื้นที่สวนสาธารณะและพื้นที่ส่วนราชการมีจำนวนตัวอย่างน้อยกว่า หรือมีความหลากหลายไม่ครอบคลุม อาจเป็นสาเหตุหนึ่งที่ทำให้เทคนิคจำแนกพื้นที่เหล่านี้ได้ไม่แม่นยำเท่าที่ควร

### 5.3 ข้อเสนอแนะ

จากผลการศึกษานี้ ใช้ ANN แสดงประสิทธิภาพที่เหนือกว่า โดยเฉพาะในการจัดการกับความซับซ้อนของข้อมูล LULC จาก Sentinel-2 จึงแนะนำให้ใช้เป็นเทคนิคหลัก โดยเฉพาะในพื้นที่ที่มีความซับซ้อนหรือมีลักษณะผสมผสานกันมาก พัฒนาความแม่นยำเพิ่มเติม Deep Learning เทคนิค Deep Learning เช่น Convolutional Neural Networks (CNN) ถูกออกแบบมาเพื่อประมวลผลข้อมูลเชิงพื้นที่ (spatial data) อย่างภาพได้ดีเยี่ยม การใช้ CNN อาจช่วยให้สกัดคุณลักษณะจากภาพถ่ายดาวเทียมได้ดียิ่งขึ้น Ensemble Learning การผสมผสานเทคนิคหลาย ๆ แบบเข้าด้วยกัน (เช่น การรวมผลการทำนายจาก ANN, RF, และ SVM) อาจช่วยเพิ่มความแม่นยำและความเสถียรโดยรวมได้ ข้อเสนอแนะในการพัฒนาความแม่นยำเพิ่มเติมด้วย Deep Learning หรือ Ensemble Learning สอดคล้องกับงานวิจัยของ [50] ที่นำเสนอแบบจำลองไฮบริด ANN-RF และ [49] ที่แสดงให้เห็นประสิทธิภาพของ Deep Learning เพิ่มข้อมูลฝึกสอน การมีข้อมูลฝึกสอน (Training Data) ที่มีคุณภาพและปริมาณเพียงพอเป็นสิ่งสำคัญสำหรับเทคนิค Machine Learning ทุกประเภท โดยเฉพาะสำหรับคลาสที่จำแนกได้ยาก (เช่น พื้นที่ผสม) การเพิ่มข้อมูลตัวอย่างสำหรับคลาสเหล่านี้ จะช่วยให้เทคนิคเรียนรู้ลักษณะเฉพาะของคลาสนั้น ๆ ได้ดีขึ้น และลดความผิดพลาดในการจำแนกขยายความแนวทางการวิจัยในอนาคต

#### แนวทางการวิจัยในอนาคตต่อยอดจากผลลัพธ์

การวิเคราะห์ภาพถ่ายดาวเทียมจากหลายช่วงเวลา (เช่น Sentinel-2 มีรอบการถ่ายภาพซ้ำที่สั้น ) ช่วยให้สามารถติดตามและทำความเข้าใจพลวัตการเปลี่ยนแปลงของสิ่งปกคลุมดินได้ดีขึ้น ซึ่ง LSTM เป็นเครื่องมือที่เหมาะสมสำหรับงานนี้ วิเคราะห์การเปลี่ยนแปลงเชิงพื้นที่ - เวลา การรวมการวิเคราะห์ทั้งในมิติพื้นที่และเวลา จะช่วยให้เข้าใจรูปแบบและแนวโน้มการเปลี่ยนแปลงการใช้ประโยชน์ที่ดินได้อย่างครอบคลุม ซึ่งข้อมูลเชิงลึกนี้มีความสำคัญอย่างยิ่งต่อการวางแผนพัฒนาเมือง การจัดการทรัพยากรธรรมชาติ และการรักษาสิ่งแวดล้อมอย่างยั่งยืน

ผลการศึกษาสามารถนำไปประยุกต์ใช้ในงานด้าน การจัดการทรัพยากรที่ดิน การวางผังเมือง และการบริหารจัดการสิ่งแวดล้อม ในเขตเมืองร้อยเอ็ดได้อย่างมีประสิทธิภาพมากขึ้น

## บรรณานุกรม

- [1] Dipanwita Dutta, Atiqur Rahman, S. K. Paul, and a. A. Kundu, "Changing pattern of urban landscape and its effect on land surface temperature in and around Delhi," 2019, doi: <https://doi.org/10.1007/s10661-019-7645-3>.
- [2] Philip Kibet Langat, Lalit Kumar, Richard Koech, and a. M. K. Ghosh, "Monitoring of land use/land-cover dynamics using remote sensing: a case of Tana River Basin, Kenya," *Geocarto International*, vol. 36, no. 13, 2021, doi: <https://doi.org/10.1080/10106049.2019.1655798>.
- [3] J. r. Hoffmann, "The future of satellite remote sensing in hydrogeology," *Hydrogeol*, vol. 13, pp. 247-250, 2005, doi: 10.1007/s10040-004-0409-2.
- [4] Yuei-An Liou, An-Ming Wu, and a. H.-Y. Lin, *FORMOSAT-2 Quick Imaging*. 2015.
- [5] Elizabeth A. Wentz, David Nelson, Atiqur Rahman, William L. Stefanov, and a. S. S. Roy, "Expert system classification of urban land use/cover for Delhi, India," ed, 2008.
- [6] Ce Zhang *et al.*, "Joint Deep Learning for land cover and land use classification," *Remote Sensing of Environment*, vol. 221, pp. 173-187, 2019, doi: <https://doi.org/10.1016/j.rse.2018.11.014>.
- [7] Swapan Talukdar *et al.*, "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review," 2020, doi: <https://doi.org/10.3390/rs12071135>.
- [8] สำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ, ตำราเทคโนโลยีอวกาศและภูมิสารสนเทศศาสตร์ *Space Technology and Geo-Informatics*. 2552.
- [9] ส. สาขาสระบุรี, เทคนิคการใช้ภาพถ่ายดาวเทียมในการติดตามการเปลี่ยนแปลงของสภาพพื้นที่ป่าและตรวจติดตามไฟป่า. 2021.
- [10] GISTDA, มาตรฐานข้อกำหนดข้อมูลภูมิสารสนเทศพื้นฐาน (FGDS) ชั้นข้อมูลการใช้ประโยชน์ที่ดิน. 2013.
- [11] B. T. C. Buathongkuea, A. Chaiyarat, P. Rattanathavorn, and N. Kaewthong, "Application of SENTINEL-2 and LANDSAT 8 to classifying second-level land use classification in the SathingPra peninsular, Songkhla province," presented at the

- The 27th National Convention on Civil Engineering Chiang Rai, THAILAND 2565.  
[Online]. Available:  
<https://conference.thaince.org/index.php/ncce27/article/view/1738/869>.
- [12] Ilada AROONSRI and a. S. SANGPRADID, "ARTIFICIAL NEURAL NETWORKS FOR THE CLASSIFICATION OF SHRIMP FARM FROM SATELLITE IMAGERY," presented at the Geographia Technica, Vol. 16, Issue 2, 2021, 2021.
- [13] M. P. Theodoros Evgeniou, "Regularized Multi-Task Learning," 2001.
- [14] อ. เลิศอร่ามแสง, "การประยุกต์ใช้การวิเคราะห์ข้อมูลภาพเชิงวัตถุร่วมกับออนไลน์เพื่อวิเคราะห์การใช้ที่ดินในจังหวัดพิษณุโลก," ปรชญาตษฎีบัณฑิต, มหาวิทยาลัยนเรศวร, 2565.  
[Online]. Available: สืบค้นจาก  
<http://nuir.lib.nu.ac.th/dspace/handle/123456789/5808>
- [15] ก. ชูชีพ, "การประเมินความถูกต้องในการสำรวจระยะไกล," presented at the Remote Sensing Technical Note No. 3, Faculty of Forestry, Kasetsart University, 2561, 03.  
[Online]. Available: สืบค้นจาก <https://forest-admin.forest.ku.ac.th/304xxx/?q=system/files/book/3%282018%29%20Accuracy%20Assessment.pdf>.
- [16] H.L. Harris, A.S. Cupp, and a. R. N. Funston, "Utilization of soybeans or corn milling by-products in beef heifer development diets," *Proceedings, Western Section, American Society of Animal Science*, vol. 56, 2005, pp. 317-321, 2008.  
[Online]. Available:  
[https://www.researchgate.net/publication/5905685\\_Utilization\\_of\\_soybeans\\_or\\_corn\\_milling\\_by-products\\_in\\_beef\\_heifer\\_development\\_diets](https://www.researchgate.net/publication/5905685_Utilization_of_soybeans_or_corn_milling_by-products_in_beef_heifer_development_diets).
- [17] P. SERRA, X. PONS, and a. D. SAURI, "Post-classification change detection with data from different sensors: some accuracy considerations," *International Journal of Remote Sensing*, vol. 24, pp. 3311-3340, 2003, doi: 10.1080/0143116021000021189.
- [18] J. Richard Landis and Gary G. Koch, "An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers," vol. 33, no. 2, pp. 363-374, 1977, doi: <https://doi.org/10.2307/2529786>.
- [19] Russel G. Congalton and a. K. Green, *Assessing the Accuracy of Remotely Sensed Data*, 2018.

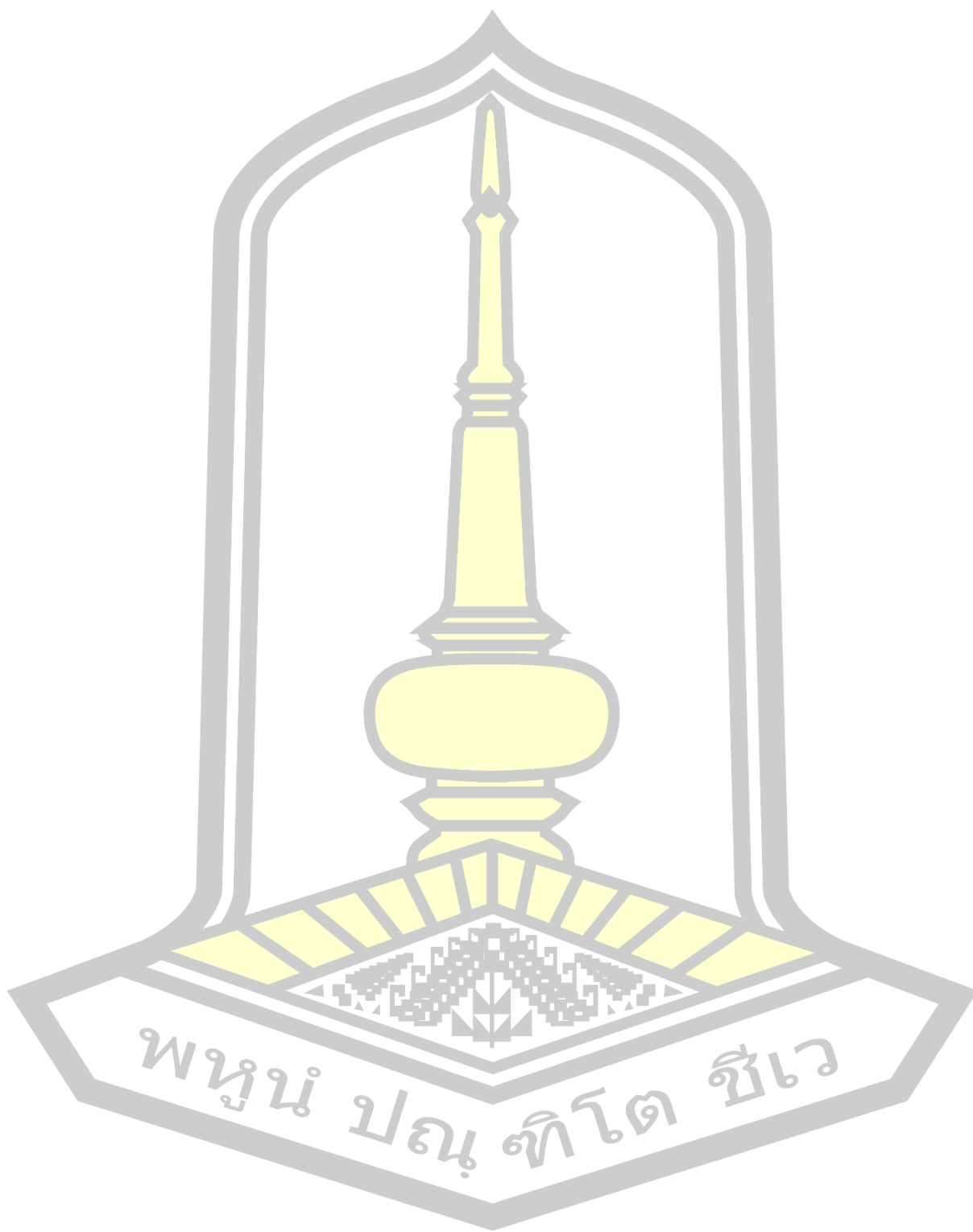
- [20] K. Intarat, & Sillaparatt, S., "การจำแนกพื้นที่ป่าชายเลนเขตร้อนด้วยอัลกอริธึมป่าสุ่ม และ ข้อมูลภาพถ่ายจากดาวเทียมรายละเอียดสูง " presented at the วารสารวิทยาศาสตร์บูรพา ปีที่ 24 (ฉบับที่ 2), พฤษภาคม – สิงหาคม พ.ศ. 2562 2562.
- [21] K. Siri-on, Diloksumpun, P., & Khunrattanasiri, W. , "เทคนิคการจำแนกการใช้ประโยชน์ ที่ดินด้วยภาพถ่ายเชิงเลขจากอากาศยานไร้คนขับ," 2564.
- [22] K. Intarat, "การจำแนกการใช้ประโยชน์ที่ดินในจังหวัดนครนายก ด้วยอัลกอริธึมการเรียนรู้ เครื่องและภาพถ่ายจากดาวเทียม Sentinel-2," วารสารวิทยาศาสตร์บูรพา ปีที่ 27 (ฉบับที่ 2) พฤษภาคม – สิงหาคม พ.ศ. 2565 2565.
- [23] Peng Fang *et al.*, "The Classification Performance and Mechanism of Machine Learning Algorithms in Winter Wheat Mapping Using Sentinel-2 10 m Resolution Imagery," *Appl. Sci.*10(15), 5075, 2020. [Online]. Available: สืบค้นจาก <https://www.mdpi.com/2076-3417/10/15/5075>.
- [24] Gouhar Aziz, Nasru Minallah, Aamir Saeed, Jaroslav Frnda, and a. W. Khan, "Remote sensing based forest cover classification using machine learning," 2024.
- [25] Rashmi Saini and S. Singh, "Land use land cover mapping and snow cover detection in Himalayan region using machine learning and multispectral Sentinel-2 satellite imagery," presented at the *Int. j. inf. tecnol.*, 2024. [Online]. Available: <https://doi.org/10.1007/s41870-023-01673-1>.
- [26] Eman A. Alshari, Mohammed B. Abdulkareem, and a. B. W. Gawali, "Classification of land use/land cover using artificial intelligence (ANN-RF)," vol. 5-2022, 2023, doi: <https://doi.org/10.3389/frai.2022.964279>.
- [27] Admilson da Penha Pacheco, João Alexandre Silva do Nascimento, Antonio Miguel Ruiz-Armenteros, and A. U. J. d. S. Junior, "Land Cover Transformations in Mining-Influenced Areas Using PlanetScope Imagery, Spectral Indices, and Machine Learning: A Case Study in the Hinterlands de Pernambuco, Brazil," *Land*, vol. 14, no. 2, p. 325, 30 November 2025, doi: <https://doi.org/10.3390/land14020325>.
- [28] Sheeba Khan, Ashutosh Bhardwaj, and a. M. Sakthivel, "Accuracy Assessment of Land Use Land Cover Classification Using Machine Learning Classifiers in Google Earth Engine; A Case Study of Jammu District," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information*

- Sciences*, vol. XLVIII, no. 4, 2024, doi: <https://doi.org/10.5194/isprs-archives-XLVIII-4-2024-263-2024>.
- [29] Woo-Dam Sim, Jong-Su Yim, and and Jung-Soo Lee, "Assessing Land Cover Classification Accuracy: Variations in Dataset Combinations and Deep Learning Models," *Remote Sensing Applications in Land Use, Land-Use Change and Forestry (LULUCF)*, vol. 16, no. 14, 2024, doi: <https://doi.org/10.3390/rs16142623>.
- [30] M. Hamdia, N. Bannari, and F. H. Chokmani, "Sentinel-2 and random-forest fusion for detailed land-use/land-cover mapping in arid environments," *Remote Sensing*, vol. 15, no. 7, 2023.
- [31] J. Qiu and P. Fang, "Evaluating K-fold sizes on land-cover classification accuracy: A case of winter wheat, China," *Applied Sciences*, vol. 12, no. 2, 2022.
- [32] A. Elmotawakkil, N. Enneya, S. K. Bhagat, M. M. Ouda, and and V. Kumar, ""Advanced machine learning models for robust prediction of water quality index and classification,"" *J. Hydroinform*, vol. 27, no. 2, pp. 299-319, 2025, doi: [10.2166/hydro.2025.290](https://doi.org/10.2166/hydro.2025.290).
- [33] E. Y. Boateng, J. Otoo, and and D. A. Abaye, "Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review," vol. 8, no. 4, pp. 341-357, 2022, doi: [10.4236/jdaip.2020.84020](https://doi.org/10.4236/jdaip.2020.84020).
- [34] Guannan Huang, Y. C. Yani Guo, and and Zhengwei Nie, "Application of Machine Learning in Material Synthesis and Property Prediction," *Materials 2023*, vol. 16, no. 17, pp. 159-164, 2023, doi: [10.1145/3697090.3699798](https://doi.org/10.1145/3697090.3699798).
- [35] M. G. Nascimento, R. J. Moura, F. Machida, and and E. Andrade, "Comparison of Machine Learning Algorithms for Detecting Software Aging in SQL Server," in *Proc. 13th Lat.-Am. Symp. Dependable Secure Comput. (LADC '24)*, pp. 159-164, 2024, doi: <https://doi.org/10.1145/3697090.3699798>.
- [36] M. Kasahun and and A. Legesse, "Machine learning for urban land use/ cover mapping: Comparison of artificial neural network, random forest and support vector machine, a case study of Dilla town Heliyon," vol. 10, no. 20, p. e39146, 2024, doi: <https://doi.org/10.1016/j.heliyon.2024.e39146>.
- [37] O. Bhungeni, A. Ramjatan, and and M. Gebreslasie, "Evaluating Machine-Learning

- Algorithms for Mapping LULC of the uMngeni Catchment Area, KwaZulu-Natal," *Remote Sensing*, vol. 16, no. 12, p. 2219, 2024, doi: <https://doi.org/10.3390/rs16122219>.
- [38] Y. O. Ouma, A. Keitsile, B. Nkwae, P. Odirile, D. Moalafhi, and and J. Qi, "Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach," *European Journal of Remote Sensing*, vol. 56, no. 1, 2023, doi: <https://doi.org/10.1080/22797254.2023.2173659>.
- [39] S. Talukdar, P. Singha, S. Mahato, Shahfahad, Y.-A. L. S. Pal, and and A. Rahman, "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review," *Remote Sensing*, vol. 12, no. 7, p. 1135, 2020, doi: <https://doi.org/10.3390/rs12071135>.
- [40] L. Ghayour, Neshat, *et al.*, "Performance Evaluation of Sentinel-2 and Landsat 8 OLI Data for Land Cover/Use Classification Using a Comparison between Machine Learning Algorithms," *Remote Sensing*, vol. 13, no. 7, 2021, doi: <https://doi.org/10.3390/rs13071349>.
- [41] ว. คุณรัตน์ศิริ, จ. ศรีคงรักษ์, ส. สุขสะอาด, and ว. ดำรงสัจจศิริ, "การเปรียบเทียบวิธีการจำแนกการใช้ประโยชน์ที่ดิน จากข้อมูลภาพถ่ายดาวเทียม Landsat 9 บริเวณเขตพัฒนาพิเศษภาคตะวันออก ประเทศไทย," *Thai Journal of Forestry*, vol. 43, no. 2, pp. 51-61, 2024. [Online]. Available: Available: <https://li01.tci-thaijo.org/index.php/tjf/article/view/262376>.
- [42] E. Maxwell, Timothy A. Warner, and a. F. Fang, "Implementation of machine-learning classification in remotesensing: an applied review," *INTERNATIONAL JOURNAL OF REMOTE SENSING*, vol. 39, pp. 2784-2817, 2018, doi: <https://doi.org/10.1080/01431161.2018.1433343>.
- [43] Sophia S. Rwanga and a. J. M. Ndambuki, "Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS," *International Journal of Geosciences*, vol. 8, pp. 611-622, 2017, doi: <https://www.scirp.org/journal/paperinformation?paperid=75926>.
- [44] Kamrul Islam, Mohammed Jashimuddin, Biswajit Nath , and a. T. K. Nath, "Land use classification and change detection by using multi-temporal remotely,"

2018.

- [45] Stefan Leyk, Johannes H. Uhl, Deborah Balk, and a. B. Jones, "Assessing the accuracy of multi-temporal built-up land layers across rural-urban trajectories in the United States," vol. 204, pp. 898-917, 2018, doi: <https://doi.org/10.1016/j.rse.2017.08.035>.
- [46] Xianju Li, Weitao Chen, Xinwen Cheng, and a. L. Wang, "A Comparison of Machine Learning Algorithms for Mapping of Complex Surface-Mined and Agricultural Landscapes Using ZiYuan-3 Stereo Satellite Imagery," vol. 8, no. 6, 2016, doi: <https://doi.org/10.3390/rs8060514>.
- [47] Ramita Manandhar, Inakwu O. A. Odeh, and a. T. Ancev, "Improving the Accuracy of Land Use and Land Cover Classification of Landsat Data Using Post-Classification Enhancement," vol. 1, no. 3, pp. 330-344, 2009, doi: <https://doi.org/10.3390/rs1030330>.
- [48] Loubelo Madiela Bill Donatien, Bouka Biona Clobite, Missamou Lemvo, and a. M. Midel, "Comparing Sentinel-2 and Landsat 9 for land use and land cover mapping assessment in the north of Congo Republic: a case study in Sangha region," *INTERNATIONAL JOURNAL OF REMOTE SENSING*, vol. 45, no. 22, pp. 8015-8036, 2024, doi: <https://doi.org/10.1080/01431161.2024.2394238>.
- [49] Benyamin Hosseiny, Abdulhakim M. Abdic, and a. S. Jamalnia, "Urban land use and land cover classification with interpretable," *Remote Sensing Applications: Society and Environment* vol. 28, 2022, doi: <https://doi.org/10.1016/j.rsase.2022.100843>.
- [50] M. Sharafat, "Comparison of accuracy and reliability of random forest, support vector machine, artificial neural network and maximum likelihood method in land use/cover classification of urban setting," vol. 14, pp. 1-16, January 2024 2024, doi: <https://doi.org/10.1016/j.envc.2023.100800>



พหุบัณฑิตยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

## ประวัติผู้เขียน

ชื่อ	นลินพร วรรณศรี
วันเกิด	31 ตุลาคม 2541
สถานที่เกิด	โรงพยาบาลสรรพสิทธิประสงค์
สถานที่อยู่ปัจจุบัน	279 หมู่ 5 ตำบลขุขันธ์ อำเภอตระการพืชผล จังหวัดอุบลราชธานี
ประวัติการศึกษา	พ.ศ. 2560 จบการศึกษาโรงเรียนมัธยมศึกษาการพืชผล พ.ศ. 2564 จบการศึกษาระดับปริญญาตรี สาขาภูมิสารสนเทศ คณะ วิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม พ.ศ. 2568 ระดับปริญญาโท สาขาเทคโนโลยีสารสนเทศ คณะวิทยาการ สารสนเทศ มหาวิทยาลัยมหาสารคาม

พูนัน ปณฺ ทิโต ชีเว