

การเปรียบเทียบอัลกอริทึมการเรียนรู้ของเครื่องสำหรับการจำแนกการใช้ประโยชน์ที่ดิน และสิ่งปกคลุมดินจากภาพถ่ายดาวเทียม Sentinel-2 จังหวัดจันทบุรี

A Comparison of Machine Learning Algorithms for Land Use and Land Cover

Classification from Sentinel-2 Satellite Imagery in Chanthaburi Province

คุณากร หอมขาว^{1*}, ชูเดช โลศิริ², สลิลลา เขี่ยมอิทธิพล³, วิญญู ไครมกระโทก⁴, กษิดิศ พร้อมเพราะ¹ และ วชิรธร จันทร์ชมภู¹

Kunakorn Homkhaow^{1*}, Chudech Losiri², Salila Iamittipon³, Winyoo Kromkratoke⁴,

Kasidid Promproh¹ and Wachirathorn Janchomphu¹

¹สาขาวิชาวิทยาศาสตร์ประยุกต์ คณะเทคโนโลยีสังคม มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก วิทยาเขตจันทบุรี ประเทศไทย

²ภาควิชาภูมิศาสตร์ คณะสังคมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประเทศไทย

³กลุ่มวิเคราะห์สภาพการใช้ที่ดิน กองนโยบายและแผนการใช้ที่ดิน กรมพัฒนาที่ดิน ประเทศไทย

⁴กองวิจัยเศรษฐกิจยาง ฝ่ายเศรษฐกิจยาง การยางแห่งประเทศไทย ประเทศไทย

¹Department of Applied Science, Faculty of Social Technology,

Rajamangala University of Technology Tawan-ok, Chanthaburi Campus, Thailand

²Department of Geography, Faculty of Social Sciences, Srinakharinwirot University, Thailand

³Land Use Analysis Group, Land Use Policy and Planning Division, Land Development Department, Thailand

⁴Rubber Economics Research Division, Rubber Economics Department, Rubber Authority of Thailand, Thailand

Received: 8 April 2026, Received in revised form: 25 May 2026, Accepted: 28 May 2026

Available online: 8 June 2026

บทคัดย่อ

วัตถุประสงค์และที่มา: จังหวัดจันทบุรีเป็นพื้นที่ที่มีความหลากหลายของการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน ครอบคลุมทั้งพื้นที่เกษตรกรรม ไม้ผล ป่าไม้ พื้นที่ชุมชน และพื้นที่ชายฝั่งทะเล ซึ่งมีการเปลี่ยนแปลงอย่างต่อเนื่องจากการขยายตัวของเกษตรกรรม การพัฒนาเมือง และการเปลี่ยนแปลงสภาพแวดล้อม ส่งผลให้มีการมีข้อมูลการใช้ที่ดินที่ถูกต้อง ทันสมัย และมีความละเอียดสูงเป็นสิ่งสำคัญต่อการวางแผนและการจัดการทรัพยากรธรรมชาติ อีกทั้งในปัจจุบัน เทคโนโลยีภูมิสารสนเทศโดยเฉพาะภาพถ่ายดาวเทียม Sentinel-2 ร่วมกับเทคนิคการเรียนรู้ของเครื่อง ได้รับความนิยมในการจำแนกการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน เนื่องจากสามารถจัดการข้อมูลขนาดใหญ่และซับซ้อนได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม อัลกอริทึมแต่ละประเภทมีข้อดีและข้อจำกัดแตกต่างกัน ดังนั้นการเปรียบเทียบประสิทธิภาพของอัลกอริทึมจึงมีความจำเป็นเพื่อเลือกใช้วิธีที่เหมาะสมที่สุด การศึกษานี้จึงมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่อง 4 อัลกอริทึม ได้แก่ ป่าสุ่ม (Random Forest: RF) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) ต้นไม้ตัดสินใจ (Decision Tree: DT) และเอนโทรปีสูงสุด (Maximum Entropy: MaxEnt) สำหรับการจำแนกการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดินจำนวน 7 ประเภทในจังหวัดจันทบุรี โดยมุ่งเน้นการประเมินความถูกต้องและความเหมาะสมของแต่ละวิธีในการประยุกต์ใช้กับข้อมูลภาพถ่ายดาวเทียม Sentinel-2

วิธีดำเนินการวิจัย: ใช้ข้อมูลภาพถ่ายจากดาวเทียม Sentinel-2 ระดับ 2A ซึ่งผ่านกระบวนการแก้ไขชั้นบรรยากาศแล้ว บันทึกภาพเมื่อวันที่ 12 มกราคม พ.ศ.2569 ซึ่งมีปริมาณเมฆน้อยกว่าร้อยละ 5 เพื่อให้ได้ข้อมูลที่มีคุณภาพสูง และลดผลกระทบจากสภาพอากาศ ข้อมูลทั้งหมดถูกนำมาประมวลผลบนแพลตฟอร์ม Google Earth Engine (GEE) ประกอบไปด้วย แบนด์ที่ใช้ในการวิเคราะห์ ได้แก่ แบนด์ 2 จัดเก็บช่วงคลื่นสีน้ำเงิน (Blue) แบนด์ 3 จัดเก็บช่วงคลื่นสีเขียว (Green) แบนด์ 4 จัดเก็บช่วงคลื่นสีแดง (Red) แบนด์ 8 จัดเก็บช่วงคลื่นอินฟราเรดใกล้ (NIR) และแบนด์ 11 จัดเก็บช่วงคลื่นอินฟราเรดคลื่นสั้น (SWIR) ร่วมกับดัชนีเชิงสเปกตรัม 4 ดัชนี ได้แก่ NDVI (Normalized Difference Vegetation Index), NDBI (Normalized Difference Built-up Index), MNDWI (Modified Normalized Difference Water Index) และ BSI (Bare Soil Index) เพื่อเพิ่มศักยภาพในการแยกแยะลักษณะการใช้ประโยชน์ที่ดินที่มีความคล้ายคลึงกันทางสเปกตรัม ดำเนินการสุ่มตัวอย่างแบบแบ่งชั้นภูมิ โดยใช้ตัวอย่างฝึกสอน 1,000 ตัวอย่าง และตัวอย่างทดสอบ 400 ตัวอย่างจาก 7 ประเภทการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน อัลกอริทึมทุกตัวได้รับการปรับค่าพารามิเตอร์ (Hyperparameter Tuning) ผ่าน Grid Search และ 5-fold Cross-Validation บนแพลตฟอร์ม Google Earth Engine เพื่อให้ได้ประสิทธิภาพสูงสุดก่อนการจำแนก และผลลัพธ์ที่ได้ถูกประเมินด้วยตัวชี้วัดทางสถิติหลายรูปแบบ ได้แก่ ค่าความถูกต้องโดยรวม (Overall Accuracy: OA) ค่าสถิติแคปปา (Kappa Coefficient) ค่า F1-score รายคลาส และการทดสอบความแตกต่างทางสถิติด้วย Z-test ที่ระดับความเชื่อมั่นร้อยละ 95 เพื่อประเมินความแตกต่างของประสิทธิภาพระหว่างอัลกอริทึมอย่างมีนัยสำคัญ

ผลการวิจัย: การจำแนกการใช้ประโยชน์ที่ดินพบว่า อัลกอริทึมป่าสุ่ม (RF) ให้ผลลัพธ์ที่ดีที่สุด โดยมีค่าความถูกต้องโดยรวม (OA) สูงถึงร้อยละ 90.25 และค่าแคปปาเท่ากับ 0.88 แสดงถึงความสอดคล้องของผลการจำแนกกับข้อมูลจริงในระดับสูง รองลงมาคือวิธีต้นไม้ตัดสินใจ (DT) ซึ่งให้ค่า OA ร้อยละ 86.25 และค่าแคปปาเท่ากับ 0.83 ตามด้วยซัพพอร์ตเวกเตอร์แมชชีน (SVM) ที่มีค่า OA ร้อยละ 84.25 และค่าแคปปาเท่ากับ 0.81 ในขณะที่วิธีเอนโทรปีสูงสุด (MaxEnt) ให้ผลลัพธ์ต่ำสุด โดยมีค่า OA ร้อยละ 69.50 และค่าแคปปาเท่ากับ 0.63 เมื่อพิจารณาในระดับรายคลาส พบว่าคลาสพื้นที่น้ำมีค่าความแม่นยำสูงที่สุด โดยมีค่า F1-score สูงถึง 0.99 ในวิธีป่าสุ่ม เนื่องจากมีลักษณะสเปกตรัมที่แตกต่างจากคลาสอื่นอย่างชัดเจน ในทางตรงกันข้าม คลาสพื้นที่ไม่ผลมีค่าความแม่นยำต่ำที่สุดในทุกอัลกอริทึม เนื่องจากมีลักษณะสเปกตรัมใกล้เคียงกับพื้นที่ป่าไม้หรือไม้ยืนต้น ทำให้เกิดความสับสนในการจำแนก ในส่วนของผลการทดสอบทางสถิติด้วย Z-test ยืนยันว่าอัลกอริทึมป่าสุ่มมีประสิทธิภาพแตกต่างจากอัลกอริทึมอื่นอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่นร้อยละ 95 แสดงให้เห็นถึงประสิทธิภาพของวิธีนี้ในการจัดการกับข้อมูลที่มีความซับซ้อนและไม่เป็นเชิงเส้น

สรุปผลการวิจัย: อัลกอริทึมป่าสุ่ม (RF) มีประสิทธิภาพสูงสุดในการจำแนกการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดินจำนวน 7 ประเภทในจังหวัดจันทบุรีจากข้อมูลภาพถ่ายดาวเทียม Sentinel-2 เมื่อเปรียบเทียบกับอัลกอริทึมอื่น โดยเฉพาะในกรณีที่มีข้อมูลมีความซับซ้อนและมีความแปรปรวนสูง นอกจากนี้ การประยุกต์ใช้แพลตฟอร์ม Google Earth Engine ยังช่วยเพิ่มประสิทธิภาพในการประมวลผลข้อมูลขนาดใหญ่ ลดข้อจำกัดด้านทรัพยากรคอมพิวเตอร์ และสามารถดำเนินการวิเคราะห์ได้อย่างรวดเร็วและมีประสิทธิภาพ องค์ความรู้ที่ได้จากการวิจัยนี้สามารถนำไปประยุกต์ใช้ในการวางแผนการใช้ที่ดิน การบริหาร

จัดการทรัพยากรธรรมชาติ การติดตามการเปลี่ยนแปลงการใช้ที่ดิน รวมถึงการสนับสนุนการตัดสินใจเชิงนโยบายในระดับ
จังหวัดและระดับภูมิภาคได้อย่างมีประสิทธิภาพ

คำสำคัญ: อัลกอริทึมการเรียนรู้ของเครื่อง; การใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน; ภาพถ่ายจากดาวเทียม;
เซนติเนล-2; ภูเก็ต เอิร์ธ เอนจิน

Abstract

Background and Objectives: Chanthaburi province is characterized by highly diverse land use and land cover (LULC), encompassing agricultural lands, orchards, forests, urban areas, and coastal zones. These areas undergo continuous modifications due to agricultural expansion, urbanization, and environmental transformations. Consequently, possessing accurate, up-to-date, and high-resolution LULC data is crucial for effective natural resource planning and management. Currently, geoinformatics technology, particularly Sentinel-2 satellite imagery combined with machine learning techniques, has gained significant popularity for LULC classification due to its efficiency in managing large and complex datasets. However, each algorithm has distinct advantages and limitations; therefore, evaluating their performance is necessary to determine the most optimal approach. This study aimed to compare the performance of four machine learning algorithms: Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), and Maximum Entropy (MaxEnt), for classifying seven LULC categories in Chanthaburi province. The primary focus was to evaluate the accuracy and suitability of each method when applied to satellite imagery.

Methodology: This study utilized atmospherically corrected Level-2A Sentinel-2 satellite imagery acquired on January 12, 2026. The selected imagery had less than 5% cloud cover to ensure high data quality and minimize weather-related impacts. All data were processed on the Google Earth Engine (GEE) platform. The spectral bands incorporated into the analysis included Band 2 (Blue), Band 3 (Green), Band 4 (Red), Band 8 (Near Infrared: NIR), and Band 11 (Shortwave Infrared: SWIR). Furthermore, four spectral indices—Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), Modified Normalized Difference Water Index (MNDWI), and Bare Soil Index (BSI)—were integrated to enhance the discrimination capability for LULC types sharing similar spectral signatures. Stratified random sampling was conducted, utilizing 1,000 training samples and 400 testing samples across the seven LULC categories. All algorithms underwent hyperparameter tuning via Grid Search and 5-fold Cross-Validation on the GEE platform to optimize their performance prior to classification. The final results were evaluated using various statistical metrics, including Overall Accuracy (OA), Kappa Coefficient, class-specific

F1-scores, and a Z-test at a 95% confidence level to assess statistically significant differences in performance among the models.

Main Results: The LULC classification revealed that the Random Forest (RF) algorithm yielded the most superior results, achieving a high Overall Accuracy (OA) of 90.25% and a Kappa coefficient of 0.88, which indicated a robust agreement between the classified outputs and the ground truth data. This was followed by the Decision Tree (DT) method, which achieved an OA of 86.25% and a Kappa coefficient of 0.83, and the Support Vector Machine (SVM) with an OA of 84.25% and a Kappa coefficient of 0.81. Conversely, the Maximum Entropy (MaxEnt) method produced the lowest performance, yielding an OA of 69.50% and a Kappa coefficient of 0.63. At the individual class level, the water body class exhibited the highest accuracy, reaching an F1-score of 0.99 under the RF method, owing to its distinct spectral characteristics. In contrast, the orchard class demonstrated the lowest accuracy across all algorithms, as its spectral signature closely resembled that of forest or perennial tree areas, resulting in classification confusion. The statistical Z-test confirmed that the performance of the RF algorithm was significantly different from the other algorithms at the 95% confidence level, demonstrating the efficacy of this method in handling complex and non-linear data.

Conclusions: The Random Forest (RF) algorithm is the most effective method for classifying the seven LULC types in Chanthaburi province using Sentinel-2 satellite imagery when compared to the other evaluated algorithms, particularly in scenarios involving highly complex and variable data. Additionally, the application of the Google Earth Engine platform significantly enhances the processing efficiency of large-scale spatial data, mitigates computational resource constraints, and facilitates rapid analysis. The knowledge derived from this research can be effectively applied to land use planning, natural resource management, LULC change monitoring, and supporting policy-level decision-making at both provincial and regional scales.

Keywords: Machine Learning Algorithms; Land Use and Land Cover; Satellite Imagery; Sentinel-2; Google Earth Engine

*Corresponding author. E-mail : kunakorn_ho@rmutto.ac.th

Introduction

การเพิ่มขึ้นของประชากรและการขยายตัวทางเศรษฐกิจส่งผลโดยตรงต่อการเปลี่ยนแปลงลักษณะการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน (Land Use and Land Cover: LULC) เช่น การขยายตัวของพื้นที่อยู่อาศัย พื้นที่เมือง และพื้นที่เกษตรกรรม ข้อมูล LULC จึงมีความจำเป็นอย่างยิ่งสำหรับหน่วยงานที่เกี่ยวข้อง ไม่ว่าจะเป็นหน่วยงานภาครัฐ องค์กรปกครอง

ส่วนท้องถิ่น และภาคประชาชน เพื่อนำไปใช้ในการบริหารจัดการและกำหนดนโยบายด้านการพัฒนาพื้นที่อย่างมีประสิทธิภาพ (Mazzia *et al.*, 2020) จังหวัดจันทบุรีเป็นจังหวัดหนึ่งในภาคตะวันออกของประเทศไทยที่มีความโดดเด่นด้านการเกษตร โดยเฉพาะการปลูกไม้ผลเขตร้อน เช่น ทุเรียน มังคุด และเงาะ อีกทั้งยังมีพื้นที่ป่าไม้และประเภทการใช้ที่ดินที่หลากหลาย ด้วยบทบาทการเป็นแหล่งเกษตรกรรมที่สำคัญของประเทศ ประกอบกับการพัฒนาโครงสร้างพื้นฐานที่เกิดขึ้นอย่างต่อเนื่อง ส่งผลให้เกิดความสนใจในการลงทุนและการพัฒนาพื้นที่ในจังหวัดจันทบุรีเพิ่มขึ้นอย่างมาก นอกจากนี้ กิจกรรมทางการเกษตรเชิงพาณิชย์ยังเป็นอีกปัจจัยสำคัญที่ก่อให้เกิดการเปลี่ยนแปลงการใช้ที่ดินภายในจังหวัดอย่างต่อเนื่อง ในการจำแนกการใช้ประโยชน์ที่ดิน ข้อมูลภาพถ่ายจากดาวเทียมได้กลายมาเป็นแหล่งข้อมูลที่มีความสำคัญและมีประสิทธิภาพสูง เนื่องจากสามารถครอบคลุมพื้นที่ขนาดใหญ่ และช่วยลดระยะเวลาและค่าใช้จ่ายในการดำเนินงานเมื่อเปรียบเทียบกับวิธีการสำรวจภาคสนาม (Intarat & Sillaparat, 2019) ปัจจุบันมีข้อมูลภาพถ่ายจากดาวเทียมที่เปิดให้ใช้งานโดยไม่มีค่าใช้จ่าย เช่น Landsat ซึ่งมีความละเอียดเชิงพื้นที่ 30 เมตร และ Sentinel-2 ซึ่งมีความละเอียดเชิงพื้นที่สูงสุดถึง 10 เมตร โดยข้อมูลดังกล่าวเป็นข้อมูลแบบหลายช่วงคลื่น (Multispectral) ที่ครอบคลุมตั้งแต่ช่วงคลื่นแสงที่ตามองเห็นไปจนถึงช่วงคลื่นอินฟราเรดใกล้ และช่วงคลื่นอินฟราเรดคลื่นสั้น ทำให้ผู้ใช้งานสามารถนำข้อมูลไปประยุกต์ใช้ได้อย่างหลากหลาย ทั้งในการจำแนกข้อมูลภายในภาพ และการติดตามการเปลี่ยนแปลงของพื้นที่ในแต่ละช่วงเวลา (Piedelobo *et al.*, 2019; Abdi, 2020; Mazzia *et al.*, 2020) ข้อมูลภาพถ่ายจากดาวเทียมมักถูกนำมาใช้ร่วมกับวิธีการจำแนก ซึ่งสามารถแบ่งออกได้เป็น 2 ประเภทหลัก ได้แก่ การจำแนกแบบไม่กำกับดูแล (Unsupervised classification) และการจำแนกแบบกำกับดูแล (Supervised classification) โดยวิธีการจำแนกแบบกำกับดูแลจำเป็นต้องใช้ข้อมูลสอน (training data) เพื่อให้ตัวจำแนก (classifier) สามารถเรียนรู้และนำไปเปรียบเทียบกับค่าของจุดภาพในแต่ละชั้นข้อมูล เพื่อใช้ในการทำนายผลลัพธ์ของการจำแนก (Richards, 2013) สำหรับการจำแนกภาพถ่ายจากดาวเทียมที่มีความละเอียดปานกลาง นิยมใช้วิธีการจำแนกแบบกำกับดูแล ร่วมกับการจำแนกเชิงจุดภาพ (pixel-based classification) ซึ่งมีวิธีการหลากหลาย เช่น เอนโทรปีสูงสุด (Maximum Entropy: MaxEnt) ความน่าจะเป็นสูงสุด (Maximum Likelihood: MLC) พาราเรลไปป์ (Parallelepiped: PC) และระยะทางใกล้ที่สุด (Minimum Distance: MDC) (Aggarwal *et al.*, 2016) โดยเอนโทรปีสูงสุด (Maximum Entropy: MaxEnt) ถือเป็นวิธีการจำแนกข้อมูลเชิงสถิติที่ได้รับความนิยมอย่างแพร่หลายในงานด้านการสำรวจระยะไกลและการสร้างแบบจำลองเชิงพื้นที่ โดยทั่วไปให้ความถูกต้องสูงเมื่อเปรียบเทียบกับวิธีการจำแนกหลายประเภท (Richards & Jia, 2006) หลักการของ MaxEnt เป็นการประมาณการแจกแจงความน่าจะเป็นภายใต้เงื่อนไขที่มีค่าเอนโทรปีสูงสุด โดยอาศัยข้อมูลเงื่อนไขจากตัวแปรนำเข้า อย่างไรก็ตาม วิธีดังกล่าวยังมีข้อจำกัดด้านประสิทธิภาพการจำแนก เนื่องจากมีความไวต่อสัญญาณรบกวน (noise) คุณภาพของข้อมูลฝึกสอน ความไม่สมดุลของข้อมูลตัวอย่าง และการเลือกตัวแปรนำเข้า ซึ่งอาจส่งผลต่อความสามารถในการจำแนกของแบบจำลอง (Rodriguez-Galiano *et al.*, 2012) ด้วยเหตุนี้ จึงได้มีการพัฒนาอัลกอริทึมการเรียนรู้ของเครื่องที่มีความยืดหยุ่นและสามารถจัดการกับข้อมูลที่มีความซับซ้อนได้ดียิ่งขึ้น ในปัจจุบัน อัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning: ML) ได้เข้ามามีบทบาทสำคัญในงานด้านการประมวลผลภาพมากยิ่งขึ้น โดยอัลกอริทึมเหล่านี้สามารถเรียนรู้รูปแบบและความสัมพันธ์ของข้อมูลที่มีความซับซ้อนและไม่เป็นเชิงเส้นได้อย่างมีประสิทธิภาพ อัลกอริทึม ML ได้ถูกพัฒนา

และนำมาใช้เป็นตัวจำแนกร่วมกับข้อมูลภาพถ่ายจากดาวเทียม เพื่อปรับปรุงความถูกต้องของผลการจำแนกให้สูงขึ้นกว่าวิธีการจำแนกแบบดั้งเดิม งานวิจัยที่ผ่านมาได้มีการนำอัลกอริทึม ML ไปประยุกต์ใช้ในหลากหลายด้าน เช่น การจำแนกการใช้ประโยชน์ที่ดิน (Abdi, 2020; Rodriguez-Galiano *et al.*, 2012) การจำแนกพืชพรรณ (Macintyre *et al.*, 2020) และการจำแนกพันธุ์ไม้ป่าชายเลน (Intarat & Sillaparat, 2019) ซึ่งผลการศึกษาพบว่าอัลกอริทึม ML สามารถให้ผลลัพธ์ที่มีความถูกต้องสูงขึ้นอย่างมีนัยสำคัญ

อย่างไรก็ตาม ยังปรากฏช่องว่างทางการวิจัย (research gap) ที่สำคัญในบริบทของจังหวัดจันทบุรี กล่าวคือ การจำแนกพื้นที่ไม้ผล (orchards) พื้นที่ไม้ยืนต้น (perennial vegetation) และพื้นที่ป่าไม้ (forests) ออกจากกัน ด้วยข้อมูลภาพถ่ายดาวเทียมแบบหลายช่วงคลื่น (multispectral imagery) ยังคงเป็นข้อจำกัดสำคัญในการศึกษาด้านการสำรวจระยะไกล เนื่องจากเป็นพื้นที่ที่มีลักษณะการตอบสนองเชิงสเปกตรัมที่ใกล้เคียงกัน โดยเฉพาะค่าดัชนีพืชพรรณและการสะท้อนพลังงานในช่วงคลื่นอินฟราเรดใกล้ (Near Infrared: NIR) ซึ่งส่งผลให้การแยกแยะประเภทพื้นที่เกิดความสับสนได้ง่าย โดยเฉพาะในพื้นที่จังหวัดจันทบุรี ซึ่งเป็นแหล่งเพาะปลูกไม้ผลเศรษฐกิจสำคัญของประเทศไทย เช่น ทุเรียน มังคุด และเงาะ ที่มีลักษณะโครงสร้างเรือนยอดและความหนาแน่นของพืชพรรณใกล้เคียงกับพื้นที่ป่าไม้ นอกจากนี้ ยังพบว่าการศึกษาเชิงเปรียบเทียบประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่องสำหรับการจำแนกพื้นที่ไม้ผลในบริบทดังกล่าวยังมีอยู่อย่างจำกัด ดังนั้น การศึกษานี้จึงมีความสำคัญในการประเมินว่าอัลกอริทึม ML ชนิดใดมีศักยภาพสูงสุดในการจำแนกประเภทการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดินภายใต้บริบทเฉพาะของพื้นที่ศึกษาได้อย่างมีประสิทธิภาพ

สำหรับอัลกอริทึม ML ที่นำมาเปรียบเทียบในการศึกษานี้ ประกอบด้วย 4 วิธี ที่ถูกนำมาใช้อย่างกว้างขวาง ได้แก่ อัลกอริทึมป่าสุ่ม (Random Forest: RF) ซึ่งเป็นวิธีแบบ ensemble learning ที่สร้างต้นไม้ตัดสินใจจำนวนมากและรวมผลลัพธ์ด้วยการโหวต โดยมีความสามารถในการลดปัญหา overfitting และให้ความถูกต้องสูงในการจำแนกการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน (LULC) จากข้อมูลภาพถ่ายดาวเทียม (Breiman, 2001; Rodriguez-Galiano *et al.*, 2012; Belgiu & Drăguț, 2016) อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) ซึ่งเป็นอัลกอริทึมที่ทำการสร้างเส้นแบ่ง (Hyperplane) ที่เหมาะสมที่สุดเพื่อแบ่งแยกข้อมูล และมีประสิทธิภาพสูงในการจัดการข้อมูลที่มีมิติสูง (Arora & Watanachaturaporn, 2004; Abdi, 2020) อัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree: DT) เป็นแบบจำลองที่จำแนกข้อมูลโดยการแบ่งข้อมูลแบบต่อเนื่องตามเงื่อนไขของตัวแปรแต่ละตัวในรูปแบบต้นไม้ (Breiman *et al.*, 1984) ซึ่งเป็นวิธีที่สามารถตีความผลลัพธ์ได้ง่าย และเป็นพื้นฐานสำคัญของอัลกอริทึมแบบ ensemble หลายประเภท (Richards, 2013) และอัลกอริทึมเอนโทรปีสูงสุด (Maximum Entropy: MaxEnt) ซึ่งเป็นวิธีเชิงสถิติที่อาศัยหลักการ Maximum Entropy ในการประมาณการแจกแจงความน่าจะเป็นของข้อมูลภายใต้เงื่อนไขที่ได้จากข้อมูลฝึกสอน เพื่อใช้ในการจำแนกจุดภาพออกเป็นประเภทต่าง ๆ (Richards & Jia, 2006) การเลือกอัลกอริทึมทั้ง 4 วิธีนี้ช่วยให้การศึกษาคอบคลุมแนวทางการจำแนกข้อมูลที่หลากหลาย ทั้งแบบ ensemble-based (RF) แบบ kernel-based (SVM) แบบ rule-based (DT) และแบบ statistical-based (MaxEnt) จึงเหมาะสมสำหรับการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเชิงพื้นที่อย่างรอบด้าน นอกจากนี้แบบคัสเปกตรัมของดาวเทียม Sentinel-2 แล้ว การศึกษานี้ยังนำดัชนีเชิงสเปกตรัมจำนวน 4 ชนิดมาใช้เป็นตัวแปรนำเข้าเพิ่มเติม ได้แก่ NDVI

(Normalized Difference Vegetation Index) ซึ่งใช้บ่งชี้ความหนาแน่นและความสมบูรณ์ของพืชพรรณ รวมถึงช่วยแยกพื้นที่พืชพรรณออกจากพื้นที่ที่ไม่มีพืชปกคลุม (Tucker, 1979) NDBI (Normalized Difference Built-up Index) ซึ่งใช้ในการตรวจจับพื้นที่ชุมชนและสิ่งปลูกสร้าง โดยมีการตอบสนองสูงต่อพื้นผิวเมือง เช่น คอนกรีตและหลังคาอาคาร (Zha *et al.*, 2003) MNDWI (Modified Normalized Difference Water Index) ซึ่งใช้ในการตรวจสอบพื้นที่น้ำ โดยอาศัยช่วงคลื่น SWIR เพื่อลดความสับสนกับพื้นที่ชุมชน และมีประสิทธิภาพสูงกว่า NDWI แบบดั้งเดิม (Xu, 2006) และ BSI (Bare Soil Index) ใช้สำหรับเน้นลักษณะของพื้นที่ดินเปล่า พื้นที่เปิดโล่ง และพื้นที่ที่มีพืชปกคลุมน้อย รวมถึงช่วยแยกแยะพื้นที่ดังกล่าวออกจากพื้นที่พืชพรรณ แหล่งน้ำ และพื้นที่สิ่งปลูกสร้างบางประเภท โดยประยุกต์ใช้ข้อมูลจากช่วงคลื่น SWIR, Red, NIR และ Blue ร่วมกัน (Rikimaru *et al.*, 2002) การนำดัชนีเชิงสเปกตรัมเหล่านี้มาใช้ร่วมกับแบนด์สเปกตรัมช่วยเพิ่มความหลากหลายของข้อมูลนำเข้า และเพิ่มความสามารถในการแยกแยะประเภทการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน (LULC) ที่มีลักษณะการตอบสนองทางสเปกตรัมใกล้เคียงกันได้ดียิ่งขึ้น

ดังนั้น การศึกษานี้จึงมีวัตถุประสงค์เพื่อทดสอบและเปรียบเทียบประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่องในการจำแนกการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดินในจังหวัดจันทบุรี โดยเลือกใช้อัลกอริทึมจำนวน 4 วิธี ได้แก่ ป่าสุ่ม (Random Forest: RF) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) ต้นไม้ตัดสินใจ (Decision Tree: DT) และเอนโทรปีสูงสุด (Maximum Entropy: MaxEnt) ร่วมกับข้อมูลภาพถ่ายจากดาวเทียม Sentinel-2 ภายใต้แนวทางการจำแนกเชิงจุดภาพ (pixel-based classification) อัลกอริทึมทั้งหมดได้รับการปรับค่าไฮเพอร์พารามิเตอร์เพื่อเพิ่มประสิทธิภาพของแบบจำลอง โดยใช้ชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ (training and testing datasets) และผลลัพธ์ที่ได้จะถูกนำไปประเมินความถูกต้อง เพื่อระบุวิธีการจำแนกที่เหมาะสมและมีประสิทธิภาพสูงสำหรับการจำแนกการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดินในพื้นที่ศึกษา

Methodology

1. พื้นที่ศึกษา

พื้นที่ศึกษาครอบคลุมพื้นที่ทั้งหมดของจังหวัดจันทบุรี โดยมีเนื้อที่ประมาณ 3,961,250 ไร่ หรือ 6,338 ตารางกิโลเมตร จังหวัดจันทบุรีเป็นจังหวัดทางชายฝั่งทะเลภาคตะวันออกของประเทศไทย สภาพภูมิประเทศประกอบไปด้วยป่าไม้ภูเขา ที่ราบสูง ที่ราบลุ่มน้ำ และที่ราบชายฝั่งทะเล ในส่วนของพื้นที่ป่าไม้มีประมาณ 3 ใน 10 ของพื้นที่ทั้งจังหวัด ประชากรส่วนใหญ่ของจังหวัดจันทบุรีอาศัยอยู่ทางตอนใต้ของจังหวัด (DOPA, 2023; NSO, 2023)

ทิศเหนือ	ติดต่อกับจังหวัดฉะเชิงเทรา และจังหวัดสระแก้ว
ทิศใต้	ติดต่อกับทะเลอ่าวไทย
ทิศตะวันออก	ติดต่อกับจังหวัดตราด และราชอาณาจักรกัมพูชา
ทิศตะวันตก	ติดต่อกับจังหวัดระยอง และจังหวัดชลบุรี

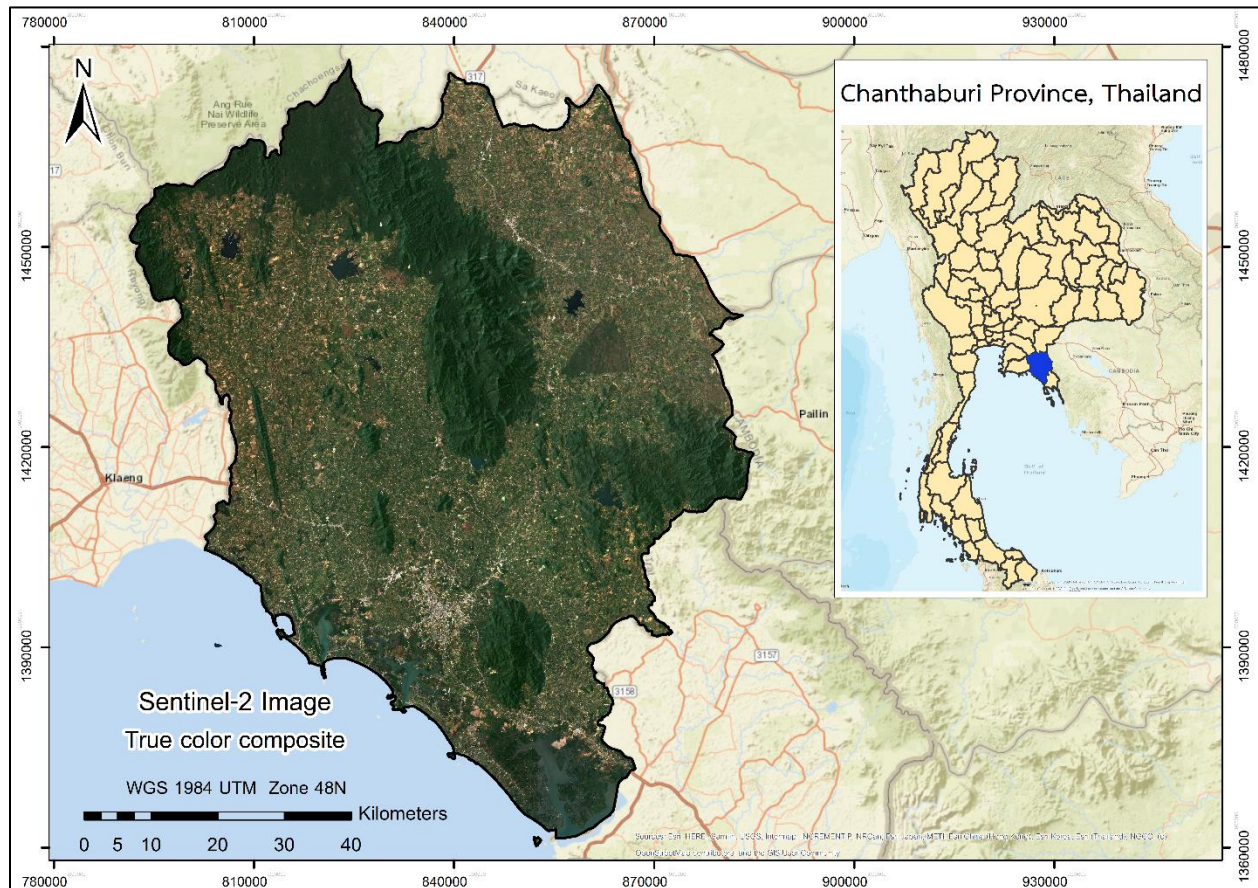


Figure 1 Study area; Chanthaburi Province

2. ข้อมูลที่ใช้ในการศึกษา

2.1 ภาพถ่ายจากดาวเทียม Sentinel-2 MSI (MultiSpectral Instrument) ระดับประมวลผล Level-2A (Bottom-of-Atmosphere Reflectance) ที่ผ่านการปรับแก้เชิงชั้นบรรยากาศ บันทึกภาพเมื่อวันที่ 12 มกราคม พ.ศ.2569 ที่ความละเอียด 20 เมตร ระบบพิกัดอ้างอิง WGS 1984 UTM Zone 48N โดยเรียกใช้ข้อมูลผ่าน Google Earth Engine ด้วย Collection ID: COPERNICUS/S2_SR_HARMONIZED ประกอบไปด้วยแบนด์ที่ใช้ในการวิเคราะห์ ได้แก่ แบนด์ 2 จัดเก็บช่วงคลื่นสีน้ำเงิน (Blue) แบนด์ 3 จัดเก็บช่วงคลื่นสีเขียว (Green) แบนด์ 4 จัดเก็บช่วงคลื่นสีแดง (Red) แบนด์ 8 จัดเก็บช่วงคลื่นอินฟราเรดใกล้ (NIR) และแบนด์ 11 จัดเก็บช่วงคลื่นอินฟราเรดคลื่นสั้น (SWIR) ซึ่งครอบคลุมช่วงคลื่นหลักที่ใช้กันอย่างแพร่หลายในการศึกษา LULC ระดับจังหวัด

2.2 ข้อมูลดัชนีเชิงสเปกตรัม (Spectral Indices) ที่ใช้ในการศึกษา โดยผู้วิจัยได้ทำการวิเคราะห์ดัชนีเชิงสเปกตรัมเพิ่มเติม เพื่อใช้เป็นคุณลักษณะของตัวแปรอัลกอริทึมในการจำแนกการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน โดยมีรายละเอียดดังแสดงใน Table 1

Table 1 Spectral indices used in this study

Index	Equation	Source
Normalized Difference Vegetation Index (NDVI)	$(\text{NIR}-\text{RED})/(\text{NIR}+\text{RED})$	(Tucker, 1979)
Normalized Difference Built-up Index (NDBI)	$(\text{SWIR1}-\text{NIR})/(\text{SWIR1}+\text{NIR})$	(Zha <i>et al.</i> , 2003)
Modified Normalized Difference Water Index (MNDWI)	$(\text{GREEN}-\text{SWIR1})/(\text{GREEN}+\text{SWIR1})$	(Xu, 2006)
Bare Soil Index (BSI)	$((\text{SWIR1}+\text{RED})-(\text{NIR}+\text{BLUE}))/(\text{SWIR1}+\text{RED}+\text{NIR}+\text{BLUE})$	(Rikimaru <i>et al.</i> , 2002)

*Note: BLUE = Reflectance of blue band, GREEN = Reflectance of green band, RED = Reflectance of red band, NIR = Reflectance of near infrared band, SWIR1 = Reflectance of short-wave infrared 1 band

2.3 ข้อมูลการใช้ประโยชน์ที่ดินของจังหวัดจันทบุรี พ.ศ.2568 ในรูปแบบเวกเตอร์ จากกรมพัฒนาที่ดิน

2.4 ข้อมูลขอบเขตจังหวัดจันทบุรี ในรูปแบบเวกเตอร์ จากกรมพัฒนาที่ดิน

2.5 ข้อมูลอ้างอิงภาคสนามถูกนำมาใช้ในการจัดเตรียมชุดข้อมูลสำหรับการฝึกสอน และการทดสอบแบบจำลอง โดยแบ่งข้อมูลออกเป็นสัดส่วน 70:30 สำหรับชุดข้อมูลฝึกสอน (training dataset) และชุดข้อมูลทดสอบ (testing dataset) ตามลำดับ ด้วยวิธีการสุ่มตัวอย่างแบบแบ่งชั้น (stratified random splitting) เพื่อรักษาสัดส่วนของข้อมูลในแต่ละประเภทการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดินให้สอดคล้องกันทั้งสองชุดข้อมูล ประเภทการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดินที่ใช้ในการศึกษาประกอบด้วย 7 ประเภท ได้แก่ พื้นที่ชุมชนและสิ่งปลูกสร้าง (Urban and Built-up Area: Ur) พื้นที่น้ำ (Water Body: Wa) พื้นที่ป่าไม้และไม้ละเมาะ (Forest and Shrubland: Fr) พื้นที่ไม้ยืนต้น (Perennial Vegetation: Pe) พื้นที่ไม้ผล (Orchards: Or) พื้นที่เกษตรกรรมอื่น ๆ (Other Agricultural Land: Ag) และพื้นที่ดินเปล่า (Bare Land: Br) โดยกำหนดจำนวนตัวอย่างให้มีความใกล้เคียงกันในแต่ละประเภท เพื่อลดผลกระทบจากความไม่สมดุลของข้อมูล (class imbalance) ชุดข้อมูลประกอบด้วยตัวอย่างฝึกสอนจำนวน 1,000 ตัวอย่าง และตัวอย่างทดสอบจำนวน 400 ตัวอย่าง นอกจากนี้เพื่อจัดการปัญหาความสัมพันธ์เชิงพื้นที่ (spatial autocorrelation) ที่อาจเกิดขึ้นจากการที่จุดตัวอย่างฝึกสอนและทดสอบอยู่ใกล้กันมากเกินไป หรืออยู่ภายในแปลงการใช้ประโยชน์ที่ดินเดียวกัน ซึ่งอาจส่งผลให้เกิดการเรียนรู้เกินพอดี (overfitting) และทำให้ค่าความถูกต้องสูงเกินจริง ได้มีการกำหนดระยะห่างขั้นต่ำระหว่างจุดตัวอย่าง (minimum separation distance) โดยจุดตัวอย่างภายในประเภทเดียวกันต้องมีระยะห่างไม่น้อยกว่า 60 เมตร (3 pixels สำหรับข้อมูล Sentinel-2 ความละเอียดเชิงพื้นที่ 20 เมตร) ซึ่งเป็นระยะที่มากกว่าความละเอียดเชิงพื้นที่ของข้อมูลหลายเท่า เพื่อลดอิทธิพลของ spatial dependency ระหว่างจุดตัวอย่างที่อยู่ใกล้เคียงกัน และเพิ่มความเป็นอิสระเชิงพื้นที่ของข้อมูลฝึกสอนและข้อมูลทดสอบ ทั้งนี้ ข้อมูลแบนด์ทั้งหมด

ถูกปรับให้อยู่ในความละเอียดเชิงพื้นที่ 20 เมตร ก่อนนำไปใช้ในการวิเคราะห์ เพื่อให้ข้อมูลมีความสอดคล้องกันในทุกตัวแปร นำเข้าในการจำแนกข้อมูลเชิงพื้นที่

3. การประมวลผลข้อมูลและการฝึกสอนอัลกอริทึม

3.1 การฝึกสอนอัลกอริทึม

โดยใช้ชุดข้อมูลฝึกสอนชุดเดียวกันสำหรับการฝึกสอนในแต่ละอัลกอริทึม โดยการประมวลผลทั้งหมดดำเนินการบน Google Earth Engine (GEE) ซึ่งเป็นแพลตฟอร์มคลาวด์สำหรับการวิเคราะห์ข้อมูลภูมิสารสนเทศขนาดใหญ่ เข้าถึงคลังข้อมูล Sentinel-2 Level-2A (COPERNICUS/S2_SR_HARMONIZED) โดยใช้ Classifier API อัลกอริทึมป่าสุ่ม (ee.Classifier.smileRandomForest) อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (ee.Classifier.libsvm) อัลกอริทึมต้นไม้ตัดสินใจ (ee.Classifier.smileCart) และอัลกอริทึมเอนโทรปีสูงสุด (ee.Classifier.maximumEntropy)

3.2 การหาพารามิเตอร์ที่เหมาะสม (Hyperparameter Optimization)

เป็นขั้นตอนที่มีความสำคัญในการสร้างแบบจำลองการเรียนรู้ของเครื่อง เนื่องจากพารามิเตอร์ที่ไม่เหมาะสมอาจทำให้แบบจำลองเกิดปัญหาความซับซ้อนต่ำเกินไป (underfitting) หรือความซับซ้อนสูงเกินไป (overfitting) ซึ่งส่งผลโดยตรงต่อความแม่นยำในการจำแนกข้อมูล โดยใช้ค่า Accuracy เชillyจาก Cross-Validation เป็นเกณฑ์ในการคัดเลือก โดยกระบวนการปรับค่าไฮเพอร์พารามิเตอร์บนแพลตฟอร์ม GEE ดำเนินการด้วยวิธี Grid Search ซึ่งเป็นการทดสอบทุกชุดค่าพารามิเตอร์ที่กำหนดไว้ในตาราง (Table 2) อย่างครบถ้วน และใช้ 5-Fold Cross-Validation บนชุดข้อมูลฝึกสอน เพื่อประเมินประสิทธิภาพเฉลี่ยของแต่ละชุดพารามิเตอร์และลดความเสี่ยงจาก overfitting ก่อนนำค่าที่ดีที่สุดไปใช้ในการจำแนกจริง โดยมีรายละเอียดดังแสดงใน Table 2

Table 2 Hyperparameters Tested for Each Algorithm

Algorithms	Testing hyperparameters
Random Forest (RF)	numberOfTrees: 50, 100, 200, 300 minLeafPopulation: 1, 5, 10 bagFraction: 0.5, 0.6, 0.7
Support Vector Machine (SVM)	kernelType: RBF gamma: 0.1, 0.5, 1.0 cost: 1, 10, 100
Decision Tree (DT)	maxNodes: 10, 20, 50, 100 minLeafPopulation: 1, 5, 10
Maximum Entropy (MaxEnt)	Default parameters (GEE implementation)

3.3 การตรวจสอบความถูกต้อง

ผลลัพธ์การจำแนกทั้งหมดจะถูกนำมาวัดประสิทธิภาพด้วยตารางเมตริกซ์ความสับสน (Confusion Matrix) คำนวณค่า True Positive (TP), False Positive (FP), True Negative (TN) และ False Negative (FN) จากชุดข้อมูลทดสอบ

จากนั้นคำนวณค่าความถูกต้องโดยรวม (OA), ค่า Precision, ค่า Recall, ค่า F1-score, ค่าสัมประสิทธิ์แคปปา (Cohen's Kappa: K) และทดสอบค่า Z-test ที่ระดับความเชื่อมั่นร้อยละ 95 (Viera & Garrett, 2005)

Results

1. ผลของไฮเพอร์พารามิเตอร์ที่เหมาะสม

การทดสอบไฮเพอร์พารามิเตอร์ของอัลกอริทึมป่าสุ่ม พบว่า ชุดที่ให้ค่าความถูกต้องสูงสุด (ร้อยละ 90.25) ได้แก่ จำนวนต้นไม้ (numberOfTrees) = 200 จำนวนตัวอย่างใบไม้ต่ำสุด (minLeafPopulation) = 1 และสัดส่วนข้อมูลสุ่ม (bagFraction) = 0.5 สำหรับอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ค่า cost = 10 และ gamma = 0.5 (kernelType = RBF) ให้ประสิทธิภาพสูงสุดที่ร้อยละ 80.25 สำหรับอัลกอริทึมต้นไม้ตัดสินใจ ค่าจำนวนโหนดสูงสุด (maxNodes) = 50 และจำนวนตัวอย่างใบไม้ต่ำสุด (minLeafPopulation) = 1 ให้ประสิทธิภาพสูงสุดที่ร้อยละ 84.00 ดังแสดงรายละเอียดใน Table 3

Table 3 Optimal hyperparameters for each algorithm

Algorithms	Testing hyperparameters
Random Forest (RF)	numberOfTrees: 200 minLeafPopulation: 1 bagFraction: 0.5
Support Vector Machine (SVM)	kernelType: RBF gamma: 0.5 cost: 10
Decision Tree (DT)	maxNodes: 50 minLeafPopulation: 1
Maximum Entropy (MaxEnt)	Default parameters (GEE implementation)

2. ผลการตรวจสอบความถูกต้อง

ผลการจำแนกด้วยวิธีทั้ง 4 วิธี ได้รับการประเมินความถูกต้องโดยใช้ชุดข้อมูลทดสอบชุดเดียวกัน พบว่าอัลกอริทึมป่าสุ่ม (RF) ให้ค่าความถูกต้องโดยรวม (OA) สูงสุด ที่ร้อยละ 90.25 และค่า Kappa เท่ากับ 0.88 รองลงมาคือต้นไม้ตัดสินใจ (DT) ให้ค่าความถูกต้องโดยรวม (OA) ร้อยละ 86.25 และ Kappa เท่ากับ 0.83 ซัพพอร์ตเวกเตอร์แมชชีน (SVM) ให้ค่าความถูกต้องโดยรวม (OA) ร้อยละ 84.25 และ Kappa เท่ากับ 0.81 และเอนโทรปีสูงสุด (MaxEnt) ให้ค่าความถูกต้องโดยรวม (OA) ร้อยละ 69.50 และ Kappa เท่ากับ 0.63 ตามลำดับ

เมื่อนำผลการจำแนกด้วยวิธีป่าสุ่มมาทดสอบความแตกต่างทางสถิติด้วยวิธี Z-test โดยกำหนดสมมติฐานหลัก (H_0) คือ ผลลัพธ์จากการจำแนกด้วยวิธีป่าสุ่มไม่แตกต่างจากผลลัพธ์ที่ได้จากการจำแนกด้วยวิธีอื่น และสมมติฐานรอง (H_1) คือ ผลลัพธ์ที่ได้จากการจำแนกด้วยวิธีป่าสุ่มมีความแตกต่างจากผลลัพธ์ที่ได้จากการจำแนกด้วยวิธีอื่น (Intarat & Sillaparat, 2019) ผลการทดสอบแสดงให้เห็นว่า วิธีป่าสุ่มให้ผลลัพธ์แตกต่างจากวิธีอื่นอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่นร้อยละ 95 (ค่าวิกฤติ ± 1.96) โดยมีค่า Z-test เท่ากับ 7.22, 2.49 และ 2.43 เมื่อเปรียบเทียบกับวิธีเอนโทรปีสูงสุด วิธีซัพพอร์ต

เวกเตอร์แมชชีน และวิธีต้นไม้ตัดสินใจ ตามลำดับ ทั้งนี้ ค่าทดสอบ Z-test ในทุกกรณีมีค่าสูงกว่าค่าวิกฤติที่กำหนด จึงสามารถปฏิเสธสมมติฐานหลัก (H_0) ได้อย่างมีนัยสำคัญทางสถิติ ดังแสดงรายละเอียดใน Table 4

Table 4 Comparison of OA, Kappa and Z-test results

Algorithms	OA (%)	Kappa (K)	Z-test	Test result
Random Forest (RF)	90.25	0.88	-	-
Support Vector Machine (SVM)	84.25	0.81	2.49	Reject H_0
Decision Tree (DT)	86.25	0.83	2.43	Reject H_0
Maximum Entropy (MaxEnt)	69.50	0.63	7.22	Reject H_0

เมื่อพิจารณาค่า F1-score จำแนกตามประเภทการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน (LULC) ของแต่ละอัลกอริทึม จาก Table 5 พบว่า ป่าสุ่มมีประสิทธิภาพในการจำแนกสูงที่สุดในทุกประเภท โดยมีค่า F1-score สูงกว่าซัพพอร์ตเวกเตอร์แมชชีน ต้นไม้ตัดสินใจ และเอนโทรปีสูงสุด และเมื่อพิจารณาค่า F1-score รายประเภทการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน จาก Table 5 พบว่า พื้นที่แหล่งน้ำ (Water body) มีค่าความแม่นยำสูงที่สุดในทุกอัลกอริทึม โดยที่อัลกอริทึมป่าสุ่ม ซัพพอร์ตเวกเตอร์แมชชีน และต้นไม้ตัดสินใจให้ค่า F1-score เท่ากับ 0.99 เท่ากัน ขณะที่อัลกอริทึมเอนโทรปีสูงสุดมีค่า 0.90 ทั้งนี้ เนื่องจากพื้นที่น้ำมีลักษณะสเปกตรัมที่แตกต่างจากประเภทอื่นอย่างชัดเจน โดยเฉพาะค่า MNDWI ที่สูง รองลงมาคือพื้นที่ป่าไม้และไม้ละเมาะ (Forest and shrubland) ซึ่งอัลกอริทึมป่าสุ่มให้ค่า F1-score สูงสุดที่ 0.92 ต้นไม้ตัดสินใจ 0.88 ซัพพอร์ตเวกเตอร์แมชชีน 0.86 และเอนโทรปีสูงสุด 0.73 ในทางตรงกันข้าม พื้นที่ไม้ยืนต้น (Perennial vegetation) มีค่า F1-score ต่ำที่สุดในทุกอัลกอริทึม โดยอัลกอริทึมป่าสุ่ม มีค่าเท่ากับ 0.87 ต้นไม้ตัดสินใจ 0.82 ซัพพอร์ตเวกเตอร์แมชชีน 0.79 และเอนโทรปีสูงสุด 0.61 ซึ่งเป็นผลจากลักษณะสเปกตรัมที่ใกล้เคียงกับพื้นที่ไม้ผล และพื้นที่ป่าไม้และไม้ละเมาะ ทำให้เกิดความสับสนในการจำแนกระหว่างประเภทดังกล่าว สำหรับพื้นที่ไม้ผล (Orchards) อัลกอริทึมป่าสุ่มให้ค่า F1-score เท่ากับ 0.90 ต้นไม้ตัดสินใจ 0.88 ซัพพอร์ตเวกเตอร์แมชชีน 0.85 และเอนโทรปีสูงสุด 0.74 ส่วนพื้นที่ชุมชนและสิ่งปลูกสร้าง พื้นที่เกษตรกรรมอื่น ๆ และพื้นที่ดินเปล่า ให้ค่า F1-score อยู่ในระดับปานกลางถึงดี โดยอัลกอริทึมป่าสุ่ม มีค่าเท่ากับ 0.86, 0.88 และ 0.89 ตามลำดับ ขณะที่เอนโทรปีสูงสุด ให้ค่าต่ำที่สุดในหลายประเภท โดยเฉพาะพื้นที่ชุมชนและสิ่งปลูกสร้าง และพื้นที่ไม้ยืนต้น ที่มีค่าเพียง 0.61 เท่ากัน รองลงมาคือพื้นที่เกษตรกรรมอื่น ๆ ที่มีค่า 0.58 ซึ่งสะท้อนข้อจำกัดของอัลกอริทึมเอนโทรปีสูงสุดในการจำแนกประเภทที่มีความหลากหลายและทับซ้อนทางสเปกตรัมสูง ดังแสดงรายละเอียดใน Table 5 Table 6 และ Figure 3

Table 5 Comparison of F1-score by land use/land cover class

Class	Test sample	F1-score			
		RF	SVM	DT	MaxEnt
Urban and built-up area	34	0.86	0.76	0.78	0.61
Water body	44	0.99	0.99	0.99	0.90
Forest and shrubland	82	0.92	0.86	0.88	0.73
Perennial vegetation	68	0.87	0.79	0.83	0.61
Orchards	98	0.90	0.85	0.88	0.74
Other agricultural land	22	0.88	0.81	0.81	0.58
Bare land	52	0.89	0.87	0.84	0.63

นอกจากนี้ เมื่อพิจารณาผลการประเมินค่าน้ำหนักความสำคัญของตัวแปรเชิงสเปกตรัม (Feature Importance) ด้วยวิธีป่าสุ่ม (Random Forest: RF) เพื่อวิเคราะห์เชิงลึกถึงปัจจัยสเปกตรัมหลักที่มีอิทธิพลต่อผลการจำแนกประเภทการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน (LULC) ในพื้นที่จังหวัดจันทบุรี พบว่า ตัวแปรที่นำมาใช้ทั้งแบนด์สเปกตรัมและดัชนีพืชพรรณ มีอิทธิพลต่อการจำแนกแตกต่างกัน โดยตัวแปรที่มีความสำคัญต่อการตัดสินใจของแบบจำลองมากที่สุด 3 อันดับแรก ได้แก่ NDVI ร้อยละ 22.50 MNDWI ร้อยละ 18.20 และช่วงคลื่นอินฟราเรดคลื่นสั้น หรือ แบนด์ 11 ร้อยละ 15.40 ตามลำดับ ในขณะที่ช่วงคลื่นสีน้ำเงิน หรือ แบนด์ 2 ร้อยละ 1.70 เป็นตัวแปรที่มีอิทธิพลต่อแบบจำลองน้อยที่สุด ผลลัพธ์นี้แสดงให้เห็นว่าการนำดัชนีเชิงสเปกตรัมเข้ามาบูรณาการร่วมกับข้อมูลช่วงคลื่นพื้นฐานของ Sentinel-2 มีส่วนสำคัญอย่างยิ่งในการเพิ่มความแม่นยำในการแยกแยะประเภทการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดินในพื้นที่ศึกษา ดังแสดงรายละเอียดใน Figure 2

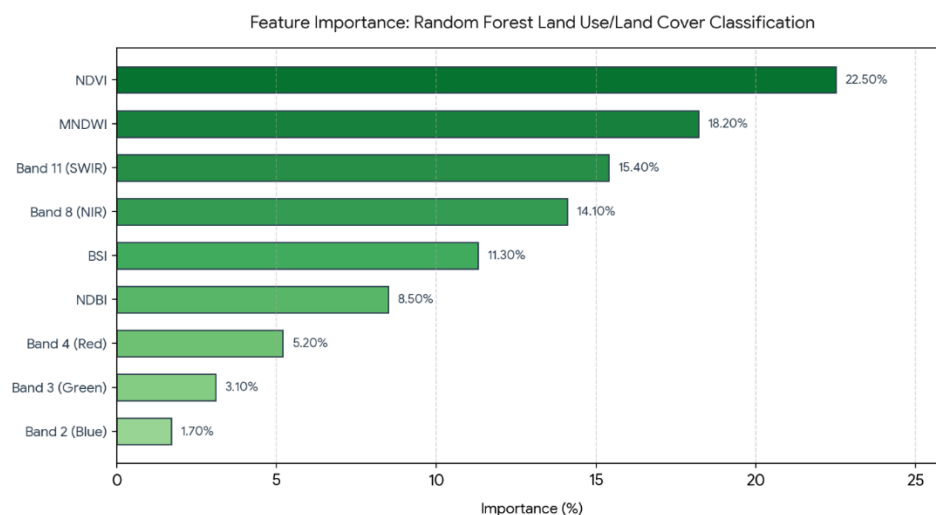
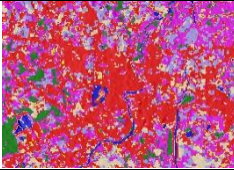
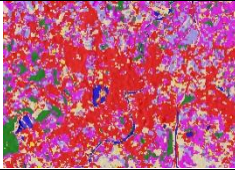
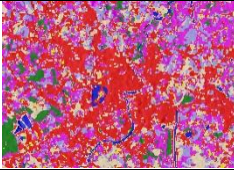
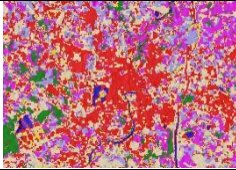
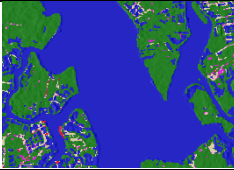
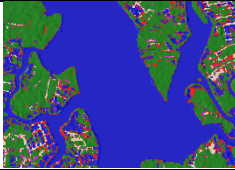
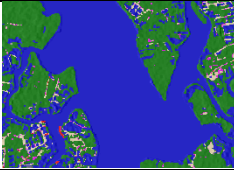
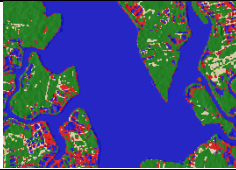
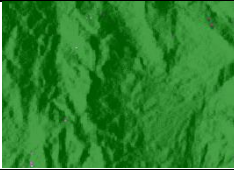
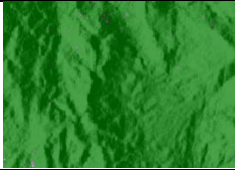
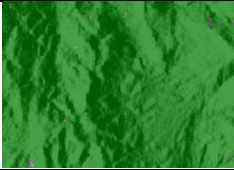
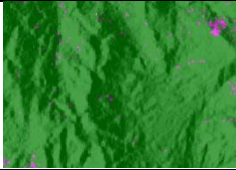
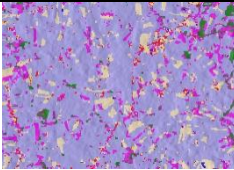
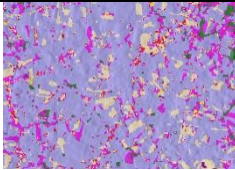
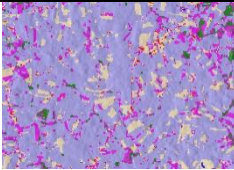
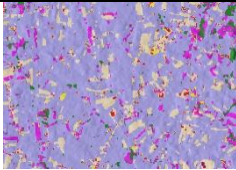
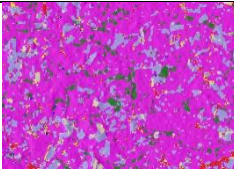
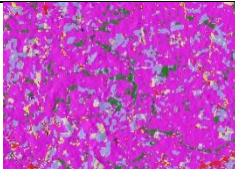
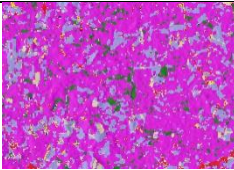
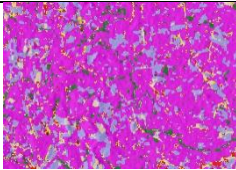
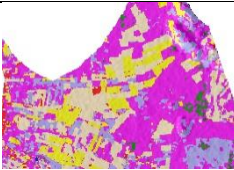
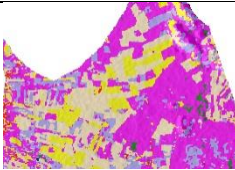
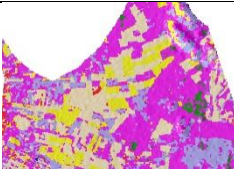
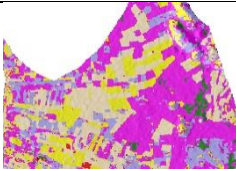
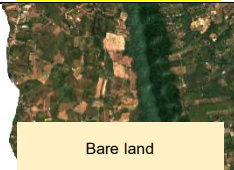
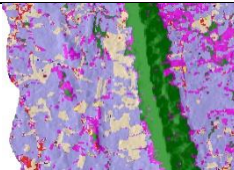
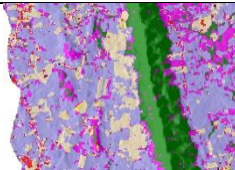
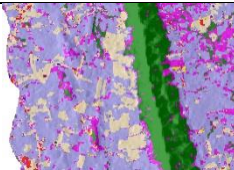
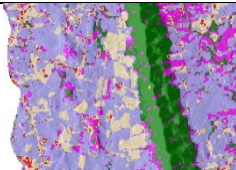


Figure 2 Feature importance scores derived from the Random Forest classifier for land use and land cover classification in Chanthaburi Province

Table 6 Comparison of land use and land cover classification results among different algorithms

Sentinel-2	RF	SVM	DT	MaxEnt
				
				
				
				
				
				
				

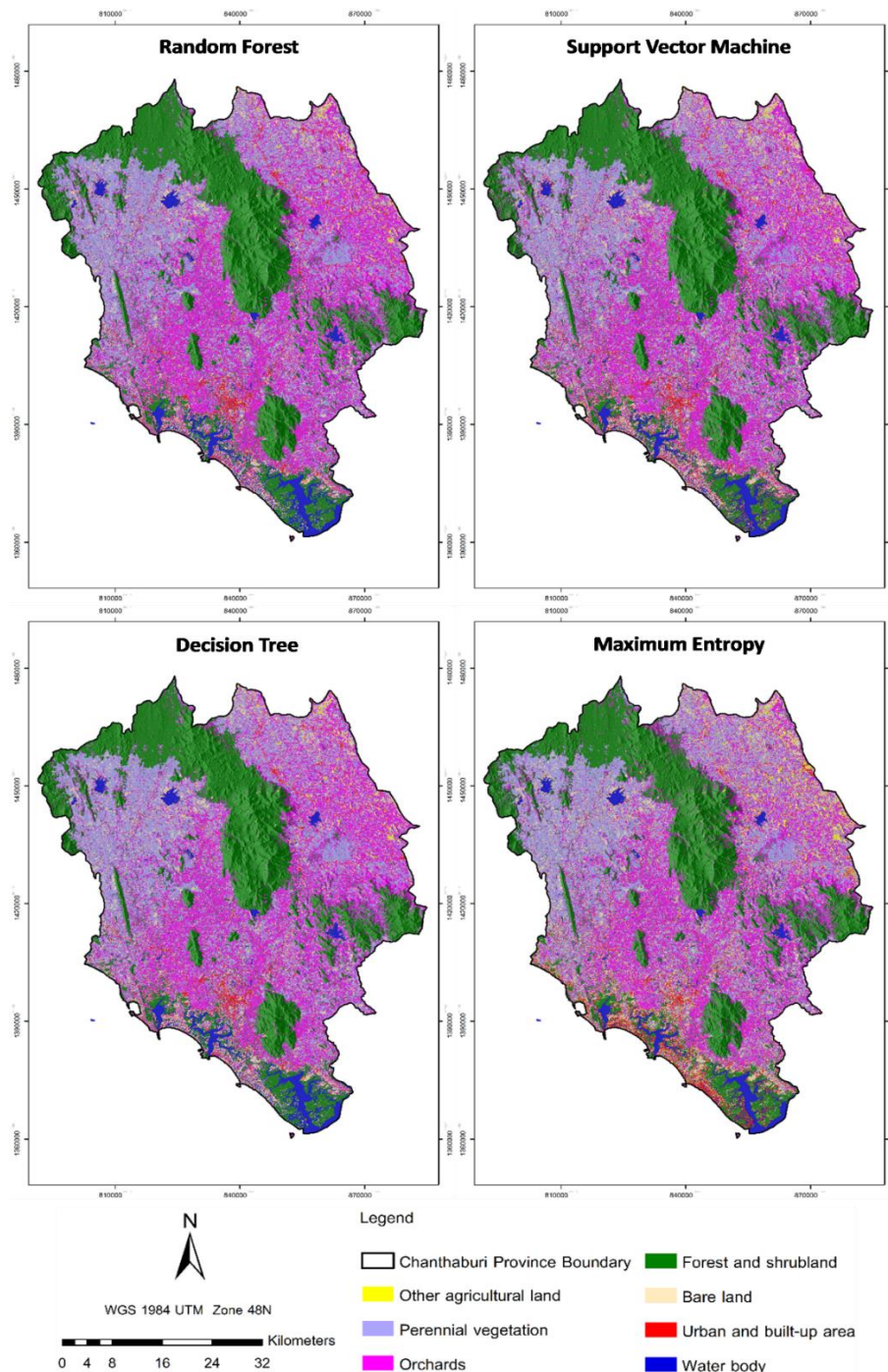


Figure 3 Classification results of land use and land cover for each algorithm

Table 7 Confusion Matrix and PA/UA of Random Forest (RF)

LULC	Or	Fr	Pe	Br	Wa	Ur	Ag	Total	UA (%)
Or	86	5	7	0	0	0	0	98	87.76
Fr	2	76	4	0	0	0	0	82	92.68
Pe	5	2	61	0	0	0	0	68	89.71
Br	0	0	0	45	0	5	2	52	86.54
Wa	0	1	0	0	43	0	0	44	97.73
Ur	0	0	0	3	0	31	0	34	91.18
Ag	0	0	0	1	0	2	19	22	86.36
Total	93	84	72	49	43	38	21	400	
PA (%)	92.47	90.48	84.72	91.84	100.0	81.58	90.48		90.25%

Table 8 Confusion Matrix and PA/UA of Decision Tree (DT)

LULC	Or	Fr	Pe	Br	Wa	Ur	Ag	Total	UA (%)
Or	83	4	9	0	0	2	0	98	84.69
Fr	3	72	6	0	0	0	0	82	87.80
Pe	4	5	58	0	0	0	0	68	85.29
Br	0	0	0	43	0	5	3	52	82.69
Wa	0	1	0	0	43	0	0	44	97.73
Ur	0	0	0	5	0	29	0	34	85.29
Ag	0	0	0	2	0	3	17	22	77.27
Total	91	82	74	50	43	40	20	400	
PA (%)	92.31	87.80	78.38	86.00	100.0	72.50	85.00		86.25%

Table 9 Confusion Matrix and PA/UA of Support Vector Machine (SVM)

LULC	Or	Fr	Pe	Br	Wa	Ur	Ag	Total	UA (%)
Or	80	6	10	0	0	2	0	98	81.63
Fr	3	72	7	0	0	0	0	82	87.80
Pe	7	6	55	0	0	0	0	68	80.88
Br	0	0	0	42	0	7	3	52	80.77
Wa	0	1	0	0	43	0	0	44	97.73
Ur	0	0	0	6	0	28	0	34	82.35
Ag	0	0	0	2	0	3	17	22	72.27
Total	90	85	72	44	43	40	20	400	
PA (%)	88.89	84.71	76.39	95.45	100.00	70.00	85.00		84.25%

Table 10 Confusion Matrix and PA/UA of Maximum Entropy (MaxEnt)

LULC	Or	Fr	Pe	Br	Wa	Ur	Ag	Total	UA (%)
Or	71	8	17	0	0	0	2	98	72.45
Fr	8	62	12	0	0	0	0	82	75.61
Pe	15	10	43	0	0	0	0	68	63.24
Br	0	0	2	31	0	16	3	52	59.62
Wa	0	8	0	0	36	0	0	44	81.82
Ur	0	0	0	10	0	24	0	34	70.59
Ag	0	0	0	6	0	5	11	22	50.00
Total	94	88	74	47	36	45	16	400	
PA (%)	75.53	70.45	58.11	65.96	100.0	53.33	68.75		69.50%

Discussion

การทดสอบประสิทธิภาพการจำแนกโดยใช้อัลกอริทึมจำนวน 4 วิธี ได้แก่ ป่าสุ่ม (Random Forest: RF) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) ต้นไม้ตัดสินใจ (Decision Tree: DT) และเอนโทรปีสูงสุด (Maximum Entropy: MaxEnt) พบว่า อัลกอริทึมการเรียนรู้ของเครื่อง ได้แก่ RF, SVM และ DT ให้ค่าความถูกต้องในการจำแนกสูงกว่า MaxEnt อย่างชัดเจน สะท้อนให้เห็นถึงศักยภาพของอัลกอริทึมการเรียนรู้ของเครื่องในการประยุกต์ใช้ร่วมกับข้อมูลภาพถ่ายดาวเทียมสำหรับการจำแนกการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน (Intarat & Sillaparat, 2019; Abdi, 2020)

โดยเฉพาะอย่างยิ่ง วิธีป่าสุ่มให้ค่าความถูกต้องโดยรวม (Overall Accuracy: OA) สูงที่สุด เท่ากับร้อยละ 90.25 เนื่องจากเป็นวิธีที่พัฒนาต่อยอดจากต้นไม้ตัดสินใจ โดยใช้การรวมผลลัพธ์จากต้นไม้หลายต้น (ensemble learning) และตัดสินใจผลด้วยวิธีการโหวตแบบค่าฐานนิยม (majority voting) ซึ่งช่วยลดความคลาดเคลื่อน และเพิ่มความเสถียรของแบบจำลอง นอกจากนี้ วิธีป่าสุ่มยังสามารถประเมินความสำคัญของตัวแปร (variable importance) และจัดการกับข้อมูลสูญหายได้อย่างมีประสิทธิภาพ (Breiman, 2001) อีกทั้งผลการทดสอบทางสถิติด้วย Z-test ยังยืนยันว่าประสิทธิภาพของวิธีป่าสุ่มสูงกว่าตัวจำแนกแบบการเรียนรู้ของเครื่องอื่นอย่างมีนัยสำคัญทางสถิติ (Intarat, 2022)

ความท้าทายสำคัญในการจำแนกข้อมูลในการศึกษานี้ คือความสับสนระหว่างพื้นที่ไม่ผลและพื้นที่ไม่ยืนต้น ซึ่งเป็นประเภทการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดินที่มีจำนวนตัวอย่างมากในพื้นที่ศึกษา ทั้งนี้เนื่องจากทั้งสองประเภทมีลักษณะการตอบสนองเชิงสเปกตรัมที่ใกล้เคียงกัน โดยเฉพาะในช่วงคลื่นอินฟราเรดใกล้ (Near Infrared: NIR) การเพิ่มดัชนีเชิงสเปกตรัม เช่น NDVI, MNDWI และ BSI เข้ามาเป็นตัวแปรนำเข้าเพิ่มเติมสามารถช่วยลดความสับสนดังกล่าวได้ในระดับหนึ่ง โดยเฉพาะเมื่อใช้ร่วมกับอัลกอริทึมป่าสุ่ม (Random Forest: RF) ซึ่งสามารถจัดการกับความสัมพันธ์เชิงซับซ้อนและไม่เป็นเชิงเส้นของข้อมูลได้อย่างมีประสิทธิภาพ เมื่อเปรียบเทียบกับอัลกอริทึมเอนโทรปีสูงสุด (Maximum Entropy: MaxEnt) ซึ่งเป็นวิธีการจำแนกเชิงสถิติ พบว่า อัลกอริทึมการเรียนรู้ของเครื่อง ได้แก่ RF, SVM และ DT ให้ค่าความถูกต้องสูงกว่าอย่างต่อเนื่อง

เนื่องจากสามารถจัดการกับข้อมูลที่มีการกระจายตัวซับซ้อนและมีลักษณะไม่เป็นเชิงเส้นได้ดีกว่า จึงเหมาะสมกับการจำแนกข้อมูลการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดินที่มีลักษณะสเปกตรัมใกล้เคียงกัน (Rodriguez-Galiano *et al.*, 2012; Intarat, 2022)

อย่างไรก็ตาม การศึกษานี้ยังมีข้อจำกัดจากการใช้ข้อมูลเชิงแสง (optical data) เพียงอย่างเดียว ซึ่งยังมีข้อจำกัดในการจำแนกพื้นที่ที่มีความหลากหลายเชิงสเปกตรัมสูง (spectral heterogeneity) และพื้นที่ที่มีลักษณะการปกคลุมผสมภายในจุดภาพ (mixed pixels) โดยเฉพาะสวนผลไม้ที่มีการปลูกพืชหลายชนิดร่วมกันในพื้นที่เดียว เช่น ทุเรียน เงาะ และมังคุด ซึ่งพืชแต่ละชนิดมีลักษณะทางกายภาพของใบและโครงสร้างเรือนยอดที่แตกต่างกันอย่างชัดเจน เช่น ใบของมังคุดที่มีความหนาและมันวาว ลักษณะทรงพุ่มของเงาะที่ค่อนข้างทึบ และการจัดเรียงตัวของใบทุเรียนที่ทำให้เกิดความขรุขระของผิวด้านบนเรือนยอด (canopy roughness) ความแตกต่างทางกายภาพเหล่านี้ทำให้เกิดเงาตกกระทบบลับซับซ้อนภายในทรงพุ่ม ส่งผลให้การกระเจิงของแสงภายในโครงสร้างใบ (spongy mesophyll) ไม่สม่ำเสมอ ทำให้ค่าการสะท้อนในช่วงคลื่นอินฟราเรดใกล้ (Near Infrared: NIR) มีความแปรปรวนสูง นอกจากนี้ ความแตกต่างของปริมาณน้ำในใบ (leaf water content) ของพืชแต่ละชนิดยังส่งผลให้ระดับการดูดกลืนพลังงานในช่วงคลื่นอินฟราเรดคลื่นสั้น (Shortwave Infrared: SWIR) ภายในพื้นที่สวนผลไม้มีความแตกต่างกันสูงตามไปด้วย เมื่อเปรียบเทียบกับพื้นที่ป่าดิบแล้งในจังหวัดจันทบุรี ซึ่งมีลักษณะโครงสร้างเรือนยอดแบบหลายชั้น (multi-layered canopy) ที่ปกคลุมต่อเนื่องและมีมวลชีวภาพสูง ทำให้ป่าดิบแล้งมีการกระเจิงของคลื่น NIR ที่ค่อนข้างสูงและคงที่ และมีการดูดกลืนคลื่น SWIR ที่สม่ำเสมอจากความชื้นในป่าที่หนาแน่นกว่า อย่างไรก็ตาม แม้กลไกการสะท้อนระดับทรงพุ่มจะต่างกัน แต่เมื่อวิเคราะห์ในระดับจุดภาพ (pixel) การผสมผสานของค่าการสะท้อนที่แปรปรวนในสวนผลไม้แบบผสม กลับส่งผลให้ค่าเฉลี่ยของทั้งสองกลุ่มมีความทับซ้อนกัน เนื่องจากทั้งสองประเภทมีความหนาแน่นของเรือนยอดและค่าดัชนีพืชพรรณใกล้เคียงกัน ซึ่งสอดคล้องกับผลการวิเคราะห์ความสำคัญของตัวแปร (Feature Importance) ใน Figure 3 ที่แสดงให้เห็นว่าแบบจำลองจำเป็นต้องพึ่งพาข้อมูลการสะท้อนและการดูดกลืนพลังงานในช่วงคลื่นเสริม ได้แก่ แบนด์ 11 (SWIR) และแบนด์ 8 (NIR) อย่างมากในการเพิ่มความสามารถในการแยกแยะและลดความสับสนของกลุ่มคลาสพืชพรรณที่มีลักษณะเชิงสเปกตรัมทับซ้อนกันเหล่านี้

สำหรับความท้าทายในการแยกแยะพื้นที่ไม้ผลออกจากพื้นที่เกษตรกรรมอื่น ๆ นั้น ยังสามารถพัฒนาแนวทางเพิ่มเติมได้หลายวิธี ได้แก่ (1) การใช้ข้อมูลหลายช่วงเวลา (multi-temporal analysis) โดยนำภาพถ่ายดาวเทียมหลายช่วงเวลาในรอบปีมาวิเคราะห์ร่วมกัน เพื่อศึกษารูปแบบการเปลี่ยนแปลงเชิงฤดูกาลของพืชแต่ละประเภท เนื่องจากไม้ผลในจังหวัดจันทบุรีมีวงจรการเจริญเติบโตและการเก็บเกี่ยวแตกต่างจากพืชเกษตรประเภทอื่น เช่น ข้าวหรือพืชไร่ ส่งผลให้เกิดรูปแบบการเปลี่ยนแปลงเชิงสเปกตรัมเฉพาะตัว ซึ่งมีแนวโน้มช่วยเพิ่มความแม่นยำในการจำแนกข้อมูลได้มากยิ่งขึ้น (Mazzia *et al.*, 2020; Macintyre *et al.*, 2020) (2) การใช้ข้อมูลสารสนเทศภูมิศาสตร์เพิ่มเติม เช่น ข้อมูลแปลงเกษตร (farm parcel data) หรือข้อมูลทะเบียนสวนผลไม้ของหน่วยงานภาครัฐ เพื่อใช้เป็นข้อมูลสนับสนุนในการจำแนก (3) การเพิ่มแบนด์ Red-Edge ของ Sentinel-2 (Band 5, 6 และ 7) ซึ่งมีความไวต่อปริมาณคลอโรฟิลล์และโครงสร้างใบพืช ช่วยเพิ่มความสามารถในการแยกแยะชนิดพืชที่มีค่าดัชนีพืชพรรณใกล้เคียงกัน และ (4) การใช้คุณลักษณะเชิงพื้นผิว (texture features) จากการวิเคราะห์

Grey-Level Co-occurrence Matrix (GLCM) เป็นตัวแปรนำเข้าเพิ่มเติม เนื่องจากพื้นที่ไม้ผลมักมีรูปแบบการปลูกเป็นแถวที่สม่ำเสมอ ในขณะที่พื้นที่เกษตรกรรมอื่น ๆ มักมีลักษณะพื้นผิวที่ต่างกัน

Conclusions

งานวิจัยนี้แสดงให้เห็นถึงศักยภาพของอัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning: ML) สำหรับการจำแนกการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดินในจังหวัดจันทบุรี โดยอาศัยข้อมูลภาพถ่ายจากดาวเทียม Sentinel-2 บนแพลตฟอร์ม Google Earth Engine (GEE) ร่วมกับข้อมูลตัวอย่างที่ได้จากการสำรวจภาคสนาม ผลการศึกษาพบว่า อัลกอริทึมการเรียนรู้ของเครื่อง ได้แก่ ป่าสุ่ม (Random Forest: RF) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) และต้นไม้ตัดสินใจ (Decision Tree: DT) ให้ค่าความถูกต้องในการจำแนกสูงกว่าวิธีเอนโทรปีสูงสุด (Maximum Entropy: MaxEnt) อย่างชัดเจน โดยเฉพาะอัลกอริทึมป่าสุ่ม ซึ่งให้ค่าความถูกต้องโดยรวม (Overall Accuracy: OA) สูงที่สุด และมีความแตกต่างอย่างมีนัยสำคัญทางสถิติเมื่อเปรียบเทียบกับอัลกอริทึมอื่นที่ระดับความเชื่อมั่นร้อยละ 95 ผลการศึกษายังสะท้อนให้เห็นว่าการประยุกต์ใช้ตัวแปรนำเข้าประเภทดัชนีเชิงสเปกตรัม (เช่น NDVI, MNDWI และ BSI) ร่วมกับช่วงคลื่นอินฟราเรด (เช่น SWIR และ NIR) มีบทบาทสำคัญอย่างยิ่งในการเพิ่มประสิทธิภาพและความแม่นยำของการจำแนก ซึ่งสอดคล้องกับผลการวิเคราะห์ความสำคัญของตัวแปร (Feature Importance) ที่ยืนยันว่าตัวแปรเหล่านี้ช่วยลดความสับสนในการจำแนกพื้นที่ได้อย่างมีประสิทธิภาพ โดยเฉพาะในพื้นที่ที่มีความซับซ้อนของเรือนยอดพืชพรรณและมีลักษณะการตอบสนองเชิงสเปกตรัมใกล้เคียงกัน

สำหรับข้อเสนอแนะในการศึกษาครั้งต่อไป ควรพิจารณาการผสมผสานข้อมูลเชิงโครงสร้าง (data fusion) ระหว่างข้อมูล Sentinel-2 ร่วมกับข้อมูลเรดาร์ Sentinel-1 (SAR) หรือข้อมูล LiDAR เพื่อเพิ่มความสามารถในการแยกแยะความแตกต่างของโครงสร้างเรือนยอดระหว่างพื้นที่ไม้ผล พื้นที่ไม้ยืนต้น และพื้นที่ป่าไม้และไม้ละเมาะ นอกจากนี้ การเพิ่มช่วงคลื่น Red-Edge ของ Sentinel-2 (Band 5, 6 และ 7) เป็นตัวแปรนำเข้าเพิ่มเติม ยังมีศักยภาพในการเพิ่มประสิทธิภาพการจำแนก เนื่องจากช่วงคลื่นดังกล่าวมีความไวต่อปริมาณคลอโรฟิลล์และโครงสร้างใบพืช ซึ่งสามารถช่วยแยกแยะชนิดพืชพรรณที่มีลักษณะเชิงสเปกตรัมใกล้เคียงกันได้ดียิ่งขึ้น รวมถึงการทดสอบในพื้นที่ศึกษาอื่นเพื่อพิสูจน์ความสามารถในการนำไปใช้ทั่วไป (Generalizability) ของแบบจำลอง

Acknowledgments

งานวิจัยนี้ขอขอบพระคุณกลุ่มวิเคราะห์สภาพการใช้ที่ดิน กรมพัฒนาที่ดิน ที่ให้การสนับสนุนข้อมูลขอบเขตจังหวัดจันทบุรี และข้อมูลการใช้ประโยชน์ที่ดิน เพื่อใช้เป็นข้อมูลอ้างอิงในการวิจัย ตลอดจนผู้มีส่วนร่วมในการสำรวจภาคสนามและตรวจสอบข้อมูลภาคพื้นดิน ซึ่งมีผลสำคัญต่อความสำเร็จของงานวิจัย นอกจากนี้ ขอขอบคุณแพลตฟอร์ม Google Earth Engine และโครงการ Copernicus สำหรับการให้บริการข้อมูลภาพถ่ายดาวเทียม Sentinel-2 ที่ใช้ในการศึกษานี้

References

- Abdi, A. M. (2020). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience & Remote Sensing*, 57(1), 1–20.
<https://doi.org/10.1080/15481603.2019.1650447>
- Aggarwal, N., Srivastava, M., & Dutta, M. (2016). Comparative analysis of pixel-based and object-based classification of high resolution remote sensing images – A review. *International Journal of Engineering Trends and Technology*, 38(1), 5–11. <https://doi.org/10.14445/22315381/IJETT-V38P202>
- Arora, M.K., & Watanachaturaporn, P. (2004). Support Vector Machines for Classification of Multi- and Hyperspectral Data. In: *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-05605-9_11
- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31.
<https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Breiman, L. (2001). *Random forests*. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC Press.
- Department of Provincial Administration (DOPA). (2023). *Population statistics and administrative area data*. Bangkok, Thailand: Ministry of Interior.
- Intarat, K., & Sillaparat, S. (2019). Tropical mangrove species classification using random forest algorithm and very high-resolution satellite imagery. *Burapha Science Journal*, 24(2), 742–753. (in Thai)
- Intarat, K. (2022). Land use classification in Nakhon Nayok Province using machine learning algorithms and Sentinel-2 imagery. *Burapha Science Journal*, 27(2), 1153–1171. (in Thai)
- Macintyre, P., van Niekerk, A., & Mucina, L. (2020). Efficacy of multi-season Sentinel-2 imagery for compositional vegetation classification. *International Journal of Applied Earth Observation and Geoinformation*, 85, 101980. <https://doi.org/10.1016/j.jag.2019.101980>

- Mazzia, V., Khaliq, A., & Chiaberge, M. (2020). Improvement in land cover and crop classification based on temporal features learning from Sentinel-2 data using recurrent-convolutional neural network. *Applied Sciences*, 10(1), 238. <https://doi.org/10.3390/app10010238>
- National Statistical Office (NSO). (2023). *Statistical data of Thailand: Population and housing statistics*. Bangkok, Thailand: Ministry of Digital Economy and Society.
- Piedelobo, L., Hernandez-Lopez, D., Ballesteros, R., Chakhar, A., Del Pozo, S., Gonzalez-Aguilera, D., & Moreno, M. A. (2019). Scalable pixel-based crop classification combining Sentinel-2 and Landsat-8 data time series. *Agricultural Systems*, 171, 36–50. <https://doi.org/10.1016/j.agsy.2019.01.005>
- Richards, J.A. , & Jia, X. (2006) Remote Sensing Digital Image Analysis: An Introduction. *Springer*, Berlin, 207.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>
- Richards, J. A. (2013). Remote sensing digital image analysis: An introduction (5th ed.). *Springer*.
<https://doi.org/10.1007/978-3-642-30062-2>
- Rikimaru, A., Roy, P. S., & Miyatake, S. (2002). Tropical forest cover density mapping. *Tropical Ecology*, 43(1), 39–47.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, 37(5), 360–363.
- Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27(14), 3025–3033.
<https://doi.org/10.1080/01431160600589179>

Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24(3), 583–594.
doi.org/10.1080/01431160304987