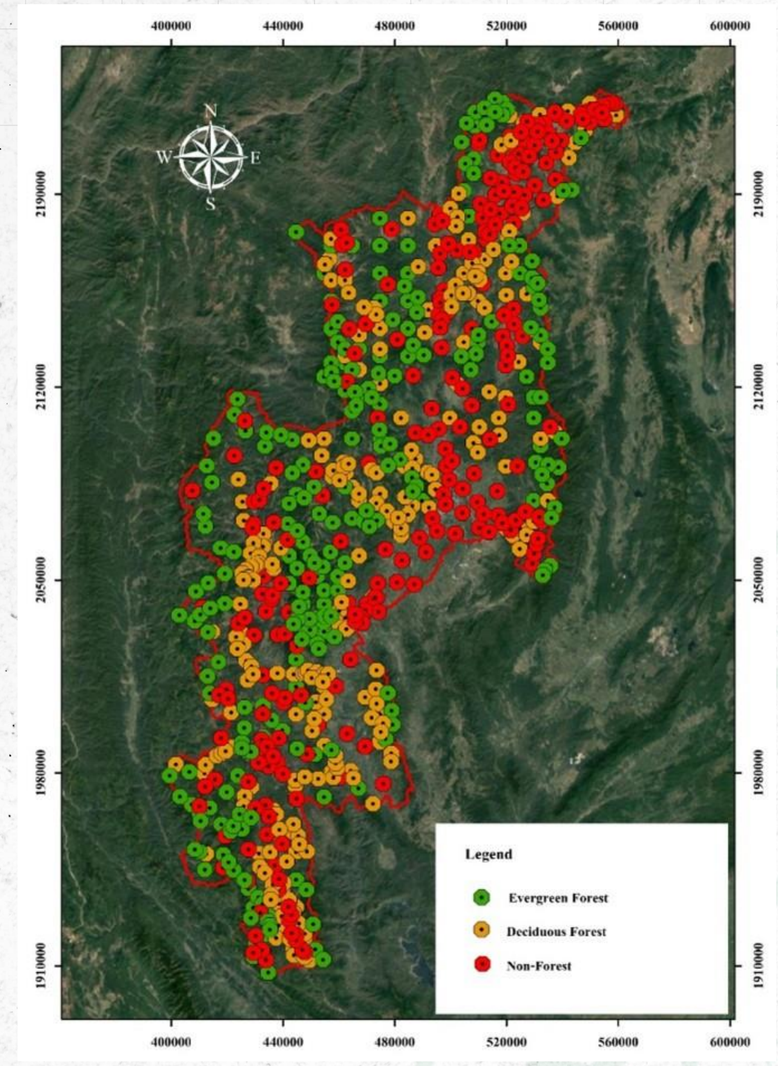
การประยุกต์อัลกอริทึมป่าสุ่มและข้อมูลดาวเทียมหลายช่วงเวลาในการจำแนกชนิดป่า

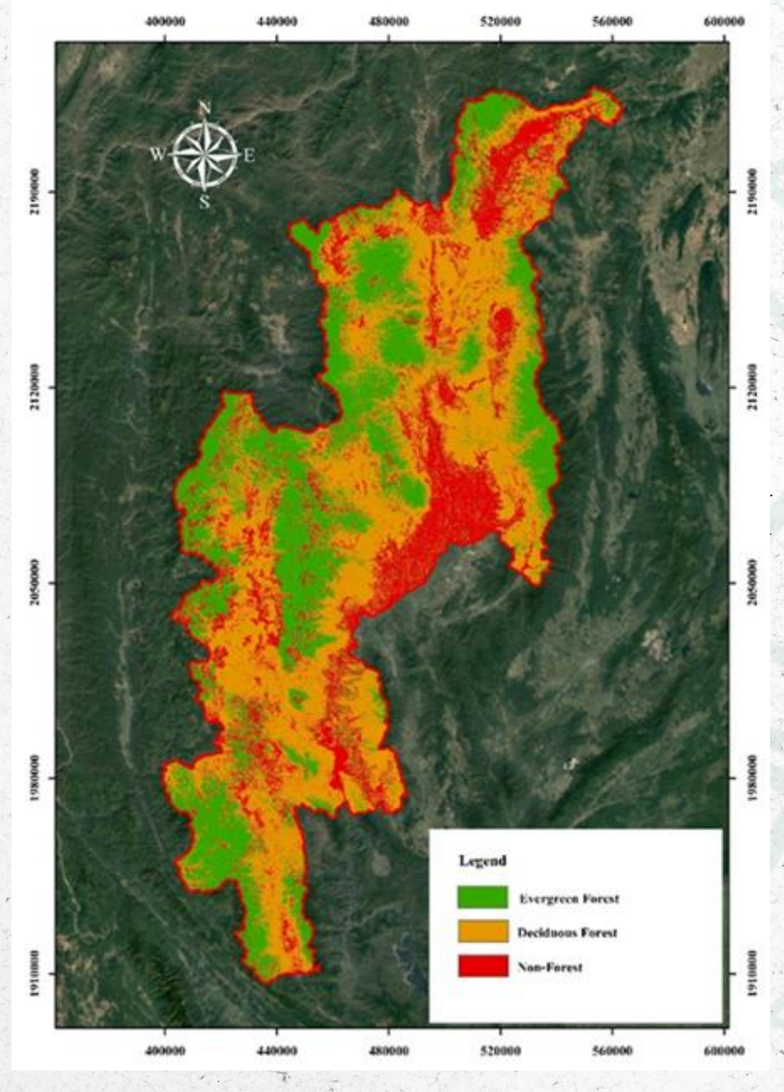
Application of Random Forest Algorithm and Multi-Temporal Satellite Data for Forest Types Classification

วิชญ์ภาส สังพาลี[1], อัฐวีร์ พิมพา[2] และวีระภาส คุณรัตนสิริ[3]

[1] สาขาวิชาเกษตรศาสตร์ คณะผลิตกรรมการเกษตร มหาวิทยาลัยแม่โจ้ เชียงใหม่
[2] สาขาวิชาการจัดการป่าไม้ มหาวิทยาลัยแม่โจ้-แพร่ เฉลิมพระเกียรติ แพร่
[3] ภาควิชาการจัดการป่าไม้ คณะวนศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ กรุงเทพฯ



การกระจายจุดอ้างอิง 750 จุดสำหรับการทดสอบแบบจำลอง ซึ่งสร้างขึ้นโดยใช้การสุ่มตัวอย่างแบบแบ่งชั้นทั่วในพื้นที่จังหวัดเชียงใหม่



แผนที่แสดงการจำแนกประเภทป่าจากแบบจำลองของจังหวัดเชียงใหม่ ปี 2567

ดาวโหลดเพิ่มเติมในบทความฉบับเต็มที่: อัฐวีร์ พิมพา, สุธีระ เหิมฮึก วิชญ์ภาส สังพาลี และวีระภาส คุณรัตนสิริ. 2569. การประยุกต์อัลกอริทึมป่าสุ่มและข้อมูลดาวเทียมหลายช่วงเวลาในการจำแนกชนิดป่า จังหวัดเชียงใหม่. วารสารวิจัยนิเวศวิทยาป่าไม้เมืองไทย, 10(1), 1-22. https://doi.org/10.34044/tferj.2026.10.1.6595

การวิเคราะห์นี้ใช้จังหวัดเชียงใหม่เป็นตัวแทนการศึกษา ซึ่งมีความสำคัญเชิงยุทธศาสตร์ในฐานะพื้นที่ป่าต้นน้ำลำธารหลักของระบบนิเวศลุ่มน้ำปิง และเป็นพื้นที่ที่มีความหลากหลายทางชีวภาพสูงที่สุดแห่งหนึ่งของประเทศไทย อย่างไรก็ตาม ในปัจจุบันพื้นที่ดังกล่าวกำลังเผชิญกับวิกฤตการณ์ด้านสิ่งแวดล้อมที่รุนแรงและท้าทาย โดยเฉพาะปัญหาการลดลงของพื้นที่ป่าอย่างรวดเร็วและการเกิดไฟป่าซ้ำซากในช่วงฤดูแล้ง ส่งผลกระทบต่อเนื่องทำให้เกิดปัญหามลพิษทางอากาศ ข้ามพรมแดนและฝุ่นละอองขนาดเล็ก (PM 2.5) ที่เป็นอันตรายต่อสุขภาพของประชาชนและเศรษฐกิจในระดับภูมิภาค ความรุนแรงและพฤติกรรมการลุกลามของไฟป่านั้นมีความสัมพันธ์โดยตรงกับประเภทของเชื้อเพลิงและชนิดป่า โดยเฉพาะพื้นที่ป่าผลัดใบ (Deciduous forests) เช่น ป่าเต็งรังและป่าเบญจพรรณ ซึ่งมีการทิ้งใบและสะสมเชื้อเพลิงชีวมวลแห้งปริมาณมากในช่วงหน้าแล้ง ย่อมมีความเสี่ยงต่อการเกิดไฟและลุกลามรุนแรงกว่าป่าไม่ผลัดใบ (Evergreen forests) ที่มีความชื้นสูงกว่า ดังนั้นการมีฐานข้อมูลแผนที่จำแนกประเภทป่าไม้ที่มีความถูกต้องแม่นยำ ความละเอียดสูง และเป็นปัจจุบัน จึงเป็นสิ่งจำเป็นเร่งด่วนในการวางแผนจัดการเชื้อเพลิง การกำหนดเขตพื้นที่เสี่ยงไฟป่า และการฟื้นฟูทรัพยากรธรรมชาติ แต่ทว่า การจำแนกประเภทป่าไม้ในพื้นที่ที่มีสภาพภูมิประเทศเป็นเทือกเขาสูงสลับซับซ้อนอย่างจังหวัดเชียงใหม่ นับเป็นความท้าทายสำคัญของการสำรวจระยะไกล เนื่องจากปัจจัยทางกายภาพเรื่องเงาภูเขาและความคล้ายคลึงกันของค่าการสะท้อนแสงทางสเปกตรัมของพืชพรรณต่างชนิดในบางช่วงเวลา ทำให้วิธีการดั้งเดิมที่ใช้ภาพถ่ายดาวเทียมเพียงช่วงเวลาเดียวไม่สามารถแยกแยะป่าผลัดใบและป่าไม่ผลัดใบได้อย่างมีประสิทธิภาพ เพื่อแก้ไขข้อจำกัดดังกล่าว การบูรณาการข้อมูลภาพถ่ายดาวเทียมหลายช่วงเวลา (Multi-temporal imagery) ที่สามารถติดตามการเปลี่ยนแปลงทางชีพลักษณ์ (Phenology) ของพืชพรรณในรอบปี ร่วมกับเทคโนโลยีการเรียนรู้ของเครื่อง (Machine learning) บนแพลตฟอร์มประมวลผลข้อมูลขนาดใหญ่อย่าง Google Earth Engine (GEE) จึงเป็นแนวทางที่ทรงประสิทธิภาพ งานวิจัยนี้มีวัตถุประสงค์หลักเพื่อ 1) จัดทำแผนที่จำแนกประเภทป่าไม้ของจังหวัดเชียงใหม่ ประจำปี พ.ศ. 2567 โดยประยุกต์ใช้อัลกอริทึมป่าสุ่ม (Random forest) และ 2) วิเคราะห์ลำดับความสำคัญของตัวแปร (Feature importance) ทั้งด้านดัชนีเชิงสเปกตรัมและปัจจัยสภาพภูมิประเทศ เพื่อระบุปัจจัยทางนิเวศวิทยาที่มีอิทธิพลสูงสุดต่อความถูกต้องของการจำแนก

วิธีการ: ดำเนินการบนแพลตฟอร์ม Google Earth Engine (GEE) เพื่อประมวลผลข้อมูลเชิงพื้นที่ขนาดใหญ่ โดยใช้ข้อมูลหลักจากภาพถ่ายดาวเทียม Sentinel-2 ระดับ 2A (Surface reflectance) ที่ผ่านการแก้ไขค่าการกระเจิงของบรรยากาศแล้ว ครอบคลุมพื้นที่ศึกษา 22,436 ตารางกิโลเมตร กระบวนการเตรียมข้อมูลแบ่งออกเป็นสองส่วนสำคัญ ส่วนแรกคือการสร้างภาพสังเคราะห์แบบค่ามัธยฐาน (Median composite) ในช่วงฤดูแล้ง (1 ธันวาคม พ.ศ. 2566 – 31 มีนาคม พ.ศ. 2567) โดยคัดเลือกเฉพาะภาพที่มีเมฆปกคลุมน้อยกว่าร้อยละ 60 เพื่อใช้เป็นข้อมูลฐานที่ปราศจากเมฆสำหรับวิเคราะห์ความสัมพันธ์ของตัวแปร ส่วนที่สองคือการสร้างชุดข้อมูลอนุกรมเวลา (Time-series stack) ครอบคลุมภาพตัวแทนรายเดือนฤดูแล้ง (4 ช่วงเวลา) และฤดูฝน (1 ช่วงเวลา) เพื่อดึงลักษณะเด่นทางชีพลักษณ์ของการผลัดใบ นอกจากนี้ได้ผนวกข้อมูลสภาพภูมิประเทศ ได้แก่ ระดับความสูง (Elevation) ความลาดชัน (Slope) และทิศด้านลาด (Aspect) จากแบบจำลองความสูงเชิงเลข (SRTM DEM) โดยปรับความละเอียดเชิงพื้นที่ให้เป็น 20 เมตร โดยกำหนดตัวแปรตั้งต้นจำนวน 26 ตัวแปร ครอบคลุมดัชนีพืชพรรณ ดัชนีน้ำและดิน ดัชนีเฉพาะทางป่าไม้ และคลื่นสเปกตรัมดั้งเดิม กระบวนการคัดเลือกตัวแปร (Feature selection) ทำใน 2 ขั้นตอน คือ การคัดกรองความซ้ำซ้อนด้วยสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (ตัดตัวแปรที่มีค่า $r > 0.90$ และ $r < -0.90$) และการจัดลำดับความสำคัญตามค่าการลดลงของความไม่บริสุทธิ์จีนี (Mean decrease in gini impurity) ด้วยอัลกอริทึมป่าสุ่ม แบบจำลองกำหนดชั้นเป้าหมาย 3 ประเภท ได้แก่ ป่าผลัดใบ ป่าไม่ผลัดใบ และพื้นที่ไม่ใช่ป่า โดยใช้จุดตัวอย่างอ้างอิงมาตรฐาน 750 จุด ที่ได้จากการสุ่มแบบแบ่งชั้นภูมิ (Stratified random sampling) แบ่งเป็นชุดข้อมูลฝึกสอนร้อยละ 80 และชุดข้อมูลทดสอบร้อยละ 20 โดยกำหนดค่าพารามิเตอร์จำนวนต้นไม้ ในแบบจำลองเท่ากับ 500 ต้น

ผลการศึกษา: ผลการคัดเลือกตัวแปรได้ชุดข้อมูลที่เหมาะสมและปราศจากความซ้ำซ้อนจำนวน 12 ตัวแปร ผลการวิเคราะห์พบว่าปัจจัยที่มีอิทธิพลสูงสุดต่อความถูกต้องของแบบจำลองคือ ข้อมูลสภาพภูมิประเทศ โดยระดับความสูง (Elevation) มีค่าความสำคัญทิ้งห่างเป็นอันดับหนึ่ง (629.27) รองลงมาคือความลาดชัน (492.16) สำหรับดัชนีเชิงสเปกตรัมพบว่า ดัชนีเงา (SI) (215.45) และดัชนีความต่างพืชพรรณสีเขียว (GNDVI) (184.90) มีความสำคัญสูงสุด เนื่องจากสามารถสะท้อนความซับซ้อนของโครงสร้างเรือนยอดและพฤติกรรมการทิ้งใบในหน้าแล้งได้อย่างเด็ดขาด แบบจำลอง Random Forest ที่พัฒนาขึ้นแสดงประสิทธิภาพที่ยอดเยี่ยม โดยให้ค่าความถูกต้องแบบ Out-of-Bag (OOB) สูงถึงร้อยละ 90.30 และเมื่อประเมินกับชุดข้อมูลทดสอบพบว่ามีความถูกต้องโดยรวม (Overall accuracy) สูงถึงร้อยละ 95.92 พร้อมค่าสัมประสิทธิ์แคปปา (Kappa coefficient) 0.94 ผลการจำแนกรายประเภทพบว่า พื้นที่ไม่ใช่ป่า มีความถูกต้องสูงสุด (ความถูกต้องของผู้ผลิต 97.92% ความถูกต้องของผู้ใช้ 100%) รองลงมาคือ ป่าไม่ผลัดใบ (ความถูกต้องของผู้ผลิต 94.12% ความถูกต้องของผู้ใช้ 96.00%) และป่าผลัดใบ (ความถูกต้องของผู้ผลิต 95.83% ความถูกต้องของผู้ใช้ 92.00%) แผนที่ผลลัพธ์เชิงพื้นที่ระบุว่าจังหวัดเชียงใหม่มีพื้นที่ป่าผลัดใบประมาณร้อยละ 48.70 (10,779.68 ตร.กม.) กระจายตัวหนาแน่นในพื้นที่ราบเชิงเขาและระดับความสูงปานกลาง ในขณะที่ป่าไม่ผลัดใบครอบคลุมพื้นที่ร้อยละ 32.37 (7,164.34 ตร.กม.) ซึ่งสอดคล้องกับหลักนิเวศวิทยาป่าไม้บนพื้นที่สูง อย่างไรก็ตาม การตรวจสอบความถูกต้องเชิงพื้นที่ยังคงพบข้อจำกัดบางประการ โดยเฉพาะบริเวณรอยต่อทางนิเวศวิทยา (Ecotone) ที่ระดับความสูง 800 – 1,150 เมตร ซึ่งสังคมพืชมีความซับซ้อน และการจำแนกผิดพลาดในพื้นที่เกษตรกรรมประเภทไม้ผลยืนต้นอายุยาว เช่น สวนลำไยและสวนส้ม ที่มีลักษณะเรือนยอดสีเขียวตลอดปีคล้ายคลึงกับป่าธรรมชาติ

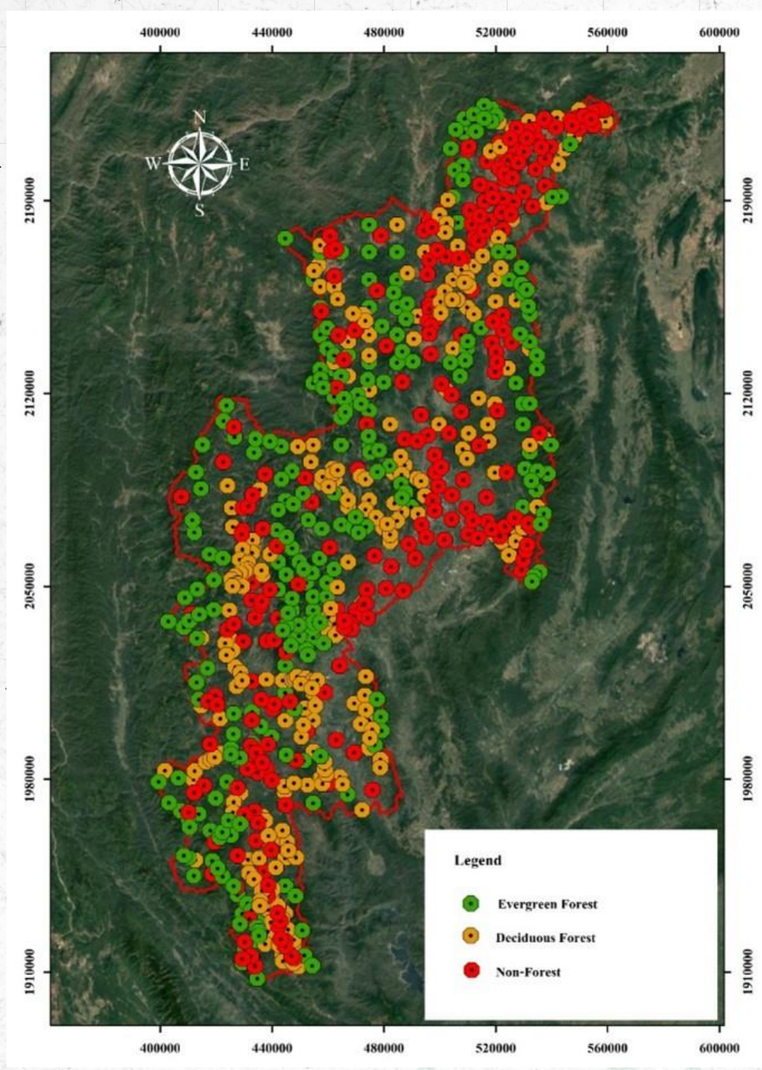
การศึกษาครั้งนี้พิสูจน์ให้เห็นว่า การประยุกต์ใช้อัลกอริทึมป่าสุ่มร่วมกับข้อมูลภาพถ่ายดาวเทียม Sentinel-2 แบบอนุกรมเวลาบนแพลตฟอร์มคลาวด์คอมพิวติ้ง เป็นเครื่องมือที่มีประสิทธิภาพสูงและมีความทนทาน (Robustness) ในระดับแนวหน้าสำหรับการจัดทำแผนที่ป่าไม้ในภูมิประเทศที่สลับซับซ้อน ข้อค้นพบเชิงประจักษ์ยืนยันว่า ระดับความสูงทำหน้าที่เสมือนเส้นแบ่งเขตทางนิเวศวิทยาที่สำคัญที่สุดในการจำแนกประเภทป่าไม้ในภาคเหนือของไทย แผนที่ผลลัพธ์ที่ได้จากงานวิจัยนี้มีความแม่นยำสูงและเหมาะสมอย่างยิ่งต่อการนำไปประยุกต์ใช้เป็นฐานข้อมูลเชิงพื้นที่ เพื่อกำหนดขอบเขตพื้นที่เสี่ยงและบริหารจัดการไฟป่าระดับพื้นที่ได้อย่างตรงจุด เพื่อการพัฒนาแบบจำลองในอนาคต ผู้วิจัยเสนอแนะให้มีการปรับปรุงการจำแนกชั้นข้อมูลพื้นที่ไม่ใช่ป่าให้มีรายละเอียดมากขึ้น โดยแยกพื้นที่สวนผลไม้ออกเป็นประเภทย่อย และพิจารณาบูรณาการข้อมูลจากเซนเซอร์เรดาร์ (Sentinel-1 SAR) หรือการวิเคราะห์พื้นผิว (Texture analysis) เพื่อยกระดับความสามารถในการแยกแยะโครงสร้างทางกายภาพของพืชพรรณให้สมบูรณ์ยิ่งขึ้น

053-875-603 งานวิจัยและบริการวิชาการ คณะผลิตกรรมการเกษตร มหาวิทยาลัยแม่โจ้
ap@mju.ac.th ห้องเอื้องเงินหลวง อาคารเฉลิมพระเกียรติสมเด็จพระศรีนครินทร์ มหาวิทยาลัยแม่โจ้

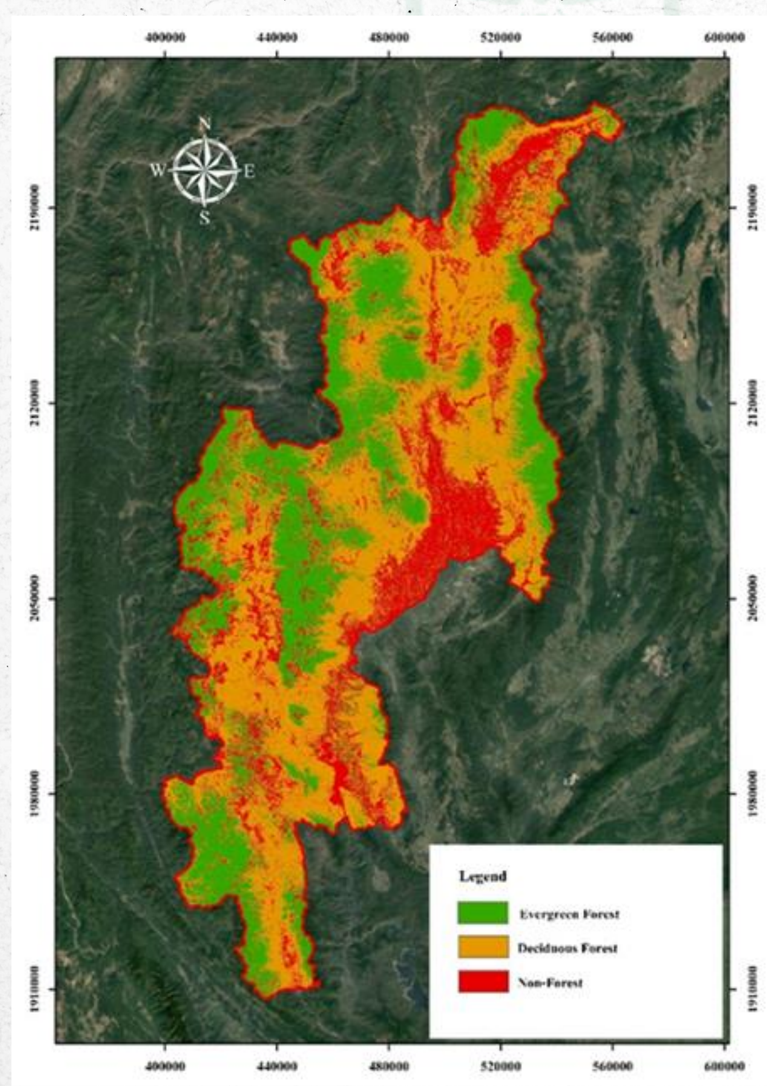
Application of Random Forest Algorithm and Multi-Temporal Satellite Data for Forest Types Classification

Witchaphart Sungpalee[1], Atthavee Phimpa[2] and Weeraphart Khunrattanasiri[3]

[1] Program in Agriculture, Faculty of Agricultural Production, Maejo University, Chiang Mai Province
[2] Program in Forest Management, Maejo University Phrae Campus, Phrae Province
[3] Department of Forest Management, Faculty of Forestry, Kasetsart University, Bangkok



Distribution of 750 reference points for model training and testing, generated using stratified random sampling across Chiang Mai Province



Forest types classification map of Chiang Mai province in 2024

Full Paper: Atthavee Phimpa, Sutheera Hermhuk, Witchaphart Sungpalee and Weeraphart Khunrattanasiri. 2026. Application of Random ForestAlgorithm and Multi-Temporal Satellite Data for Forest TypesClassification in Chiang Mai Province. Thai Forest Ecological Research Journal, 10(1), 1-22. https://doi.org/10.34044/tferj.2026.10.1.6595

This analysis uses Chiang Mai province as a representative study; the city is strategically important as a major watershed area for the Ping River basin ecosystem and boasts one of the highest biodiversity levels in Thailand. However, this area is currently facing a severe and challenging environmental crisis, particularly the rapid decline in forest area and recurring wildfires during the dry season. These fires have ongoing consequences, leading to transboundary air pollution and PM 2.5 levels that are hazardous to public health and the regional economy. The severity and spread of wildfires are directly related to fuel type and forest types. Deciduous forests, such as dry dipterocarp forest and mixed deciduous forest, which ecologically shed leaves and accumulate large amounts of dry biomass fuel during the dry season, are more susceptible to fire and more severe fire spread than evergreen forests, which have higher humidity levels. Therefore, an accurate, high-resolution, and up-to-date forest classification map database is urgently needed for fuel management planning, wildfire risk zone identification, and natural resource restoration. However, classifying forest types in the complex mountainous terrain of Chiang Mai Province presents a significant challenge for remote sensing due to physical factors such as mountain shadows and the similarity of spectral reflectance values among different vegetation species at certain times. Traditional methods using single-temporal satellite imagery cannot effectively distinguish between deciduous and evergreen forests. To overcome this limitation, integrating multi-temporal satellite imagery capable of tracking vegetation phenology changes throughout the year, along with machine learning technology on a large-scale data processing platform like Google Earth Engine (GEE), is a powerful approach. This research therefore aims to: 1) create a forest classification map of Chiang Mai Province for the year 2024 using a random forest (RF) algorithm, and 2) analyze the feature importance of both spectral indices and topographic factors to identify the ecological factors that have the greatest influence on classification accuracy.

Method: This study was conducted on the Google Earth Engine (GEE) platform to process large-scale geospatial data. The primary dataset comprised Sentinel-2 Level 2A (Surface Reflectance) imagery, which had been atmospherically corrected, covering the entire 22,436 square kilometers. The data preparation process was divided into two main parts. The first involved creating a median composite image during the dry season (December 1, 2023 – March 31, 2024), selecting only images with less than 60% cloud cover to serve as a cloud-free baseline for analyzing the relationships between variables. The second component involved constructing a time-series stack comprising representative monthly images from four dry-season periods and one wet-season period. This multi-temporal approach was specifically designed to capture the distinct phenological signatures of leaf shedding and greening. Furthermore, topographic data, including elevation, slope, and aspect, were derived from the Shuttle Radar Topography Mission Digital Elevation Model (SRTM DEM) and resampled to a 20-meter spatial resolution. The researchers initially calculated a comprehensive set of 26 predictor variables, encompassing vegetation indices, water and soil indices, forest-specific indices, and original spectral bands. To optimize model performance, a two-step Feature Selection process was implemented. Initially, Pearson's correlation coefficient analysis was used to eliminate highly redundant variables (excluding those with $r > 0.90$ or $r < -0.90$). Subsequently, the remaining variables were ranked based on the Mean Decrease in Gini Impurity metric using the Random Forest algorithm. The classification model targeted three distinct classes: deciduous forest, evergreen forest, and non-forest. Reference data consisted of 750 standard ground-truth points, collected via stratified random sampling to ensure spatial independence. These points were randomly partitioned into an 80% training set (600 points) and a 20% testing set (150 points). The RF classifier was parameterized with 500 decision trees (ntrees) to maximize stability.

Results: The feature selection process successfully refined the dataset to an optimal subset of 12 non-redundant variables. The analysis revealed that topographic features were the most influential factors governing the model's predictive capability. Specifically, elevation was dominance the ranking with the highest importance score (629.27), followed by slope (492.16). Among the spectral predictors, the shadow index (SI) (215.45) and the Green Normalized Difference Vegetation Index (GNDVI) (184.90) proved to be the most critical variables, as they effectively captured the complexities of canopy structures and mitigated topographic shadow effects during the dry season. The developed RF model demonstrated exceptional performance, achieving a high Out-of-Bag (OOB) accuracy of 90.30%. When evaluated against the independent testing set, the model yielded an Overall Accuracy of 95.92% and a Kappa coefficient of 0.94. Class-specific performance analysis indicated that the non-forest class achieved the highest accuracy (Producer's Accuracy 97.92%, User's Accuracy 100%), followed by evergreen forest (Producer's Accuracy 94.12%, User's Accuracy 96.00%), and deciduous forest (Producer's Accuracy 95.83%, User's Accuracy 92.00%). The final spatial map revealed that deciduous forests cover approximately 48.70% (10,779.68 km2) of the province, predominantly distributed in foothills and mid-elevation zones, while evergreen forests account for 32.37% (7,164.34 km2), densely dominating the higher mountain ranges, which perfectly aligns with highland forest ecology principles. However, spatial verification identified specific limitations. Minor misclassifications were observed within ecological transition zones (ecotones) at elevations between 800 and 1,150 meters due to highly mixed forest structures. Additionally, spectral confusion occurred in agricultural areas containing perennial, long-living fruit orchards (e.g., longan and orange orchards), which maintain permanent green canopies that closely resemble the spectral signatures of natural evergreen forests.

This study demonstrates that applying a RF algorithm with multi-temporal Sentinel-2 imagery on a cloud computing platform provides a highly robust, cutting-edge tool for mapping complex forest ecosystems in mountainous regions. Empirical findings confirm that elevation serves as the most important ecological boundary for forest classification in northern Thailand. The resulting high-precision classification map serves as a vital spatial database for accurately delineating wildfire risk zones and supporting targeted natural resource management. For further enhance model accuracy in future research, it is highly recommended to refine the non-forest class by explicitly separating perennial fruit orchards into distinct sub-categories. Furthermore, incorporating Synthetic Aperture Radar (Sentinel-1 SAR) data or employing advanced texture analysis should be strongly considered to improve the differentiation of physical vegetation structures across complex landscapes.

053-875-603
Faculty of agricultural production, Maejo University
ap@mju.ac.th
The Princess Mother Building, Maejo University