

# การจำแนกพันธุ์ไม้ป่าชายเลนเขตร้อนด้วยอัลกอริธึมป่าสุ่ม และข้อมูลภาพถ่ายจากดาวเทียมรายละเอียดสูง

## Tropical Mangrove Species Classification Using Random Forest Algorithm and Very High-Resolution Satellite Imagery

กฤษัญญาน อินทรัตน์ และ สุชาวดี ศิลปรัตน์

Kritchayan Intarat and Suchawadee Sillaparatt

สาขาวิชาภูมิศาสตร์ คณะศิลปศาสตร์ มหาวิทยาลัยธรรมศาสตร์

Geography Sector, Faculty of Liberal Arts, Thammasat University

Received : 5 March 2019

Revised : 23 April 2019

Accepted : 28 April 2019

### บทคัดย่อ

การผสมของค่าสเปคตรัมในจุดภาพของภาพถ่ายจากดาวเทียมเป็นปัญหาหลักที่พบในกระบวนการจำแนกเชิงจุดภาพโดยเฉพาะการจำแนกในระดับของสายพันธุ์พืช การศึกษานี้มีวัตถุประสงค์เพื่อประเมินประสิทธิภาพในการใช้อัลกอริธึมการจำแนกแบบเชิงจุดภาพป่าสุ่มเพื่อปรับปรุงความถูกต้องในการจำแนกพันธุ์ไม้ป่าชายเลนเขตร้อนบริเวณพื้นที่อนุรักษ์ป่าชายเลนอำเภอปากพนัง จังหวัดนครศรีธรรมราช ในการดำเนินงานใช้ภาพถ่ายจากดาวเทียมรายละเอียดสูง Quickbird ที่ได้รับการปรับแก้เชิงรังสี และเชิงเรขาคณิต ร่วมกับข้อมูลที่ได้จากการสำรวจภาคสนาม ข้อมูลจะถูกแบ่งออกเป็นสองชุด ได้แก่ ชุดที่ใช้สำหรับการทำการสอน และชุดที่ใช้สำหรับการทดสอบ กระบวนการสอนจะทำการปรับค่าพารามิเตอร์ในแบบจำลองที่มีความเหมาะสมสำหรับการจำแนกด้วย RF ได้แก่ ค่าความลึกสูงสุดของต้นไม้ (Maximum total of tree depth) ค่าจำนวนตัวอย่างต่ำสุดของแต่ละโหนด (Minimum number of sample node) และจำนวนของต้นไม้สูงสุดในป่า (Maximum number of tree in forest) ผลที่ได้จะถูกนำไปเปรียบเทียบประสิทธิภาพการจำแนกกับวิธีการจำแนกเชิงจุดภาพแบบความน่าจะเป็นสูงสุด โดยใช้ค่าความถูกต้องโดยรวมของการจำแนก ค่าสถิติแคปปา และค่าสถิติการทดสอบ Z เป็นตัวชี้วัด ผลการจำแนกพบว่า การใช้อัลกอริธึมป่าสุ่มให้ประสิทธิภาพในการจำแนกที่สูงกว่าวิธีการจำแนกเชิงจุดภาพแบบความน่าจะเป็นสูงสุด โดยมีค่าความถูกต้องโดยรวมที่ร้อยละ 78.00 และค่าสถิติแคปปาที่ 0.72 ในขณะที่ผลการจำแนกด้วยความน่าจะเป็นสูงสุดให้ค่าความถูกต้องโดยรวมที่ร้อยละ 56.00 และค่าสถิติแคปปาที่ 0.44 ผลจากการทดสอบค่าสถิติ Z ( $Z = 3.68$ ) ช่วยยืนยันถึงความแตกต่างระหว่างทั้งสองวิธีการจำแนกอย่างมีนัยสำคัญที่ค่าความเชื่อมั่นร้อยละ 95.00

**คำสำคัญ :** ป่าสุ่ม, ภาพถ่ายจากดาวเทียมรายละเอียดสูง, การจำแนก, พันธุ์ไม้ป่าชายเลนเขตร้อน

\*Corresponding author. E-mail : intaratt@tu.ac.th

## Abstract

The spectral mixing was a challenge that principally found in the pixel-based classification method, particularly in the species level. The objective of this study was to evaluate the effectiveness of the random forest (RF) algorithm in order to improve the accuracy of the tropical mangrove species classification in Pak Phanang mangrove conservation, Nakhon Si Thammarat Province. The study utilized the very high-resolution, the Quickbird image, which was pre-calibrated using radiometric and geometric correction to incorporate with the field observation data. The process divided the input data into the training and the validation sets. The training process adjusted the input parameters for instances, the tree depth, the number of sample node, and the number of trees to acquire the best RF classification model. The validation compared the classified result with the conventional pixel-based maximum likelihood classification (MLC). The overall accuracy (OA), the kappa statistic, and the Z-statistic were indications of the RF classification evaluation. The result revealed that the RF algorithm achieved higher efficiency with the overall accuracy of 78.00% and 0.72 for the kappa statistic. Meanwhile, for MLC, the OA and the kappa statistic presented 56.00% and 0.44, respectively. The Z statistic ( $Z = 3.68$ ) result also significantly confirmed the difference between RF and MLC at the 95% confidence level.

**Keywords :** random forest, very high resolution satellite imagery, classification, tropical mangrove species

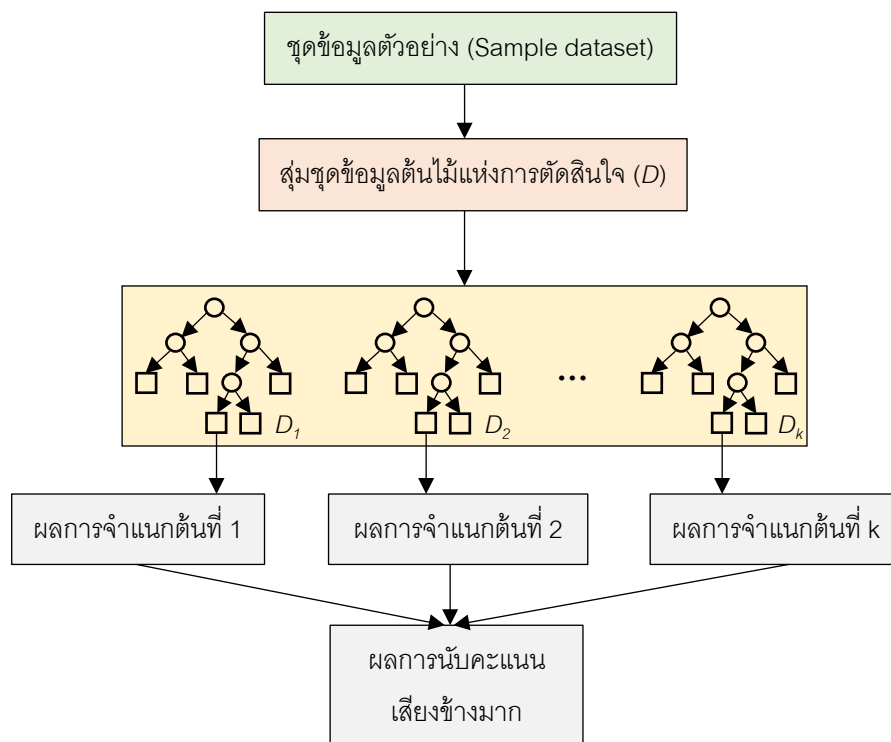
## บทนำ

ป่าชายเลนเป็นระบบนิเวศที่พบได้ทั่วไปในบริเวณระหว่างละติจูดที่ 30° เหนือ และใต้ โดยเฉพาะในพื้นที่เขตร้อน จะพบป่าชายเลนขึ้นอยู่ทั่วไปตามชายฝั่งบริเวณปากแม่น้ำ ป่าชายเลนมีความสำคัญในการเชื่อมโยงพื้นดินและพื้นน้ำเข้าด้วยกัน (Hamdan *et al.*, 2013; Heumann, 2011) เป็นแหล่งรวมของสังคมพืชและสัตว์หลากหลายชนิด นอกจากนี้ ยังให้ประโยชน์ทางด้านสิ่งแวดล้อมและเศรษฐกิจ (Lee *et al.*, 2014; Sasmito *et al.*, 2016) ส่งผลให้พื้นที่ป่าชายเลนในบางพื้นที่ได้รับการเปลี่ยนแปลงเป็นพื้นที่เกษตร บ่อเลี้ยงสัตว์น้ำ พื้นที่อยู่อาศัย และอื่น ๆ (Jhonnerie *et al.*, 2015) ป่าชายเลนบริเวณแหลมตะลุมพุก อำเภอปากพนัง จังหวัดนครศรีธรรมราช (พื้นที่ศึกษา) ซึ่งเป็นป่าชายเลนอนุรักษ์ที่อยู่ภายใต้การดูแลของสถานีพัฒนาทรัพยากรป่าชายเลนที่ 15 นครศรีธรรมราช กรมทรัพยากรทางทะเลและชายฝั่งได้รับผลกระทบจากปัญหาดังกล่าว จึงได้มีการบูรณาการเพื่อให้ชุมชนที่เข้ามาตั้งถิ่นฐานในพื้นที่ป่าชายเลนสามารถใช้ประโยชน์จากพื้นที่ร่วมกันได้อย่างยั่งยืนและไม่กระทบกับระบบนิเวศป่าชายเลน ในการจัดการและพัฒนาพื้นที่ป่าชายเลน แผนที่ป่าชายเลนที่ทันสมัยและมีความถูกต้องเป็นข้อมูลที่จำเป็นในการนำมาช่วยตัดสินใจเชิงนโยบาย โดยเฉพาะข้อมูลเกี่ยวกับพันธุ์ไม้ป่าชายเลนนั้น ยังเป็นข้อมูลที่พบเห็นได้น้อยและไม่เป็นที่นิยมจัดทำ เนื่องจากในการจัดทำแผนที่ดังกล่าว ต้องทำการสำรวจพื้นที่ป่าชายเลนอย่างละเอียด ส่งผลต่อการสิ้นเปลืองงบประมาณและเวลาที่ใช้ในการทำงาน (Kuenzer *et al.*, 2011)

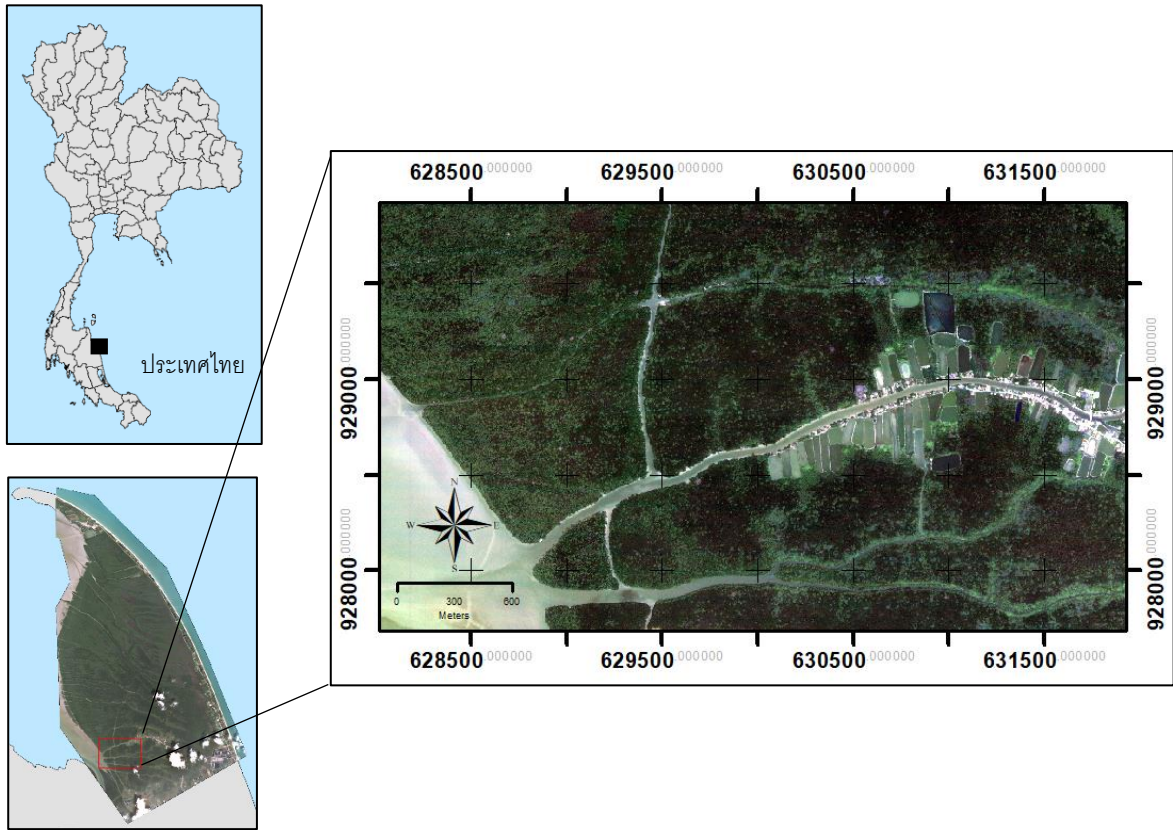
รีโมทเซนซิงได้ถูกนำมาใช้เป็นเครื่องมือเพื่อช่วยในการจัดทำข้อมูลพันธุ์ไม้ป่าชายเลน โดยนิยมใช้ร่วมกับวิธีการจำแนกแบบไร้พารามิเตอร์ (Non-parametric classification) ประเภทต่าง ๆ เช่น การจำแนกแบบความน่าจะเป็นไปได้สูงสุด ต้นไม้แห่งการตัดสินใจ และวิธีการจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีน (Chan & Paelinckx, 2008; Jhonnerie *et al.*, 2015; Watanakij & Vaiphasa, 2016) การจำแนกแบบป่าสุ่ม (Random forest: RF) เป็นวิธีการจำแนกที่จัดอยู่ใน

กลุ่ม Machine learning ชนิดหนึ่ง que เริ่มเป็นที่นิยมในการนำมาใช้กับการจำแนกข้อมูลริโมทเซนซิงในงานประเภทต่าง ๆ เช่น งานด้านการจำแนกการใช้ประโยชน์ที่ดิน สิ่งปกคลุมดิน (Dubeau *et al.*, 2017; Zhu *et al.*, 2017) พันธุ์ไม้ป่าชายเลน (Wang *et al.*, 2018) รวมไปถึงการติดตามพื้นที่ที่ถูกเผาไหม้ (Ramo & Chuvieco, 2017) RF เป็นการรวมต้นไม้ตัดสินใจ (Decision tree) จำนวนหลายต้นไว้ด้วยกันให้มีลักษณะคล้ายป่า และทำการลงคะแนนเพื่อเลือกค่าฐานนิยม (Mode) ของผลลัพธ์จากต้นไม้ทั้งหมด (ภาพที่ 1) ส่งผลให้คำตอบมีความถูกต้องและความแม่นยำมากกว่าการใช้ผลลัพธ์จากต้นไม้เพียงต้นเดียว (Chan & Paelinckx, 2008; Luo *et al.*, 2016) ข้อได้เปรียบของ RF คือ ในการทำงานไม่จำเป็นต้องทราบฟังก์ชันการแจกแจง (Distribution function) ทำให้การทำงานมีความยืดหยุ่นสูง และผลลัพธ์ที่ได้จากการจำแนกมีความถูกต้องสูง (Rodriguez-Galiano *et al.*, 2012) มากไปกว่านั้น RF ยังสามารถกำหนดความสำคัญให้กับตัวแปร รวมถึงสามารถทำนายค่าที่เป็นข้อมูลสูญหายได้ อย่างไรก็ตาม งานวิจัยที่นำ RF มาใช้จำแนกพันธุ์ไม้ป่าชายเลนยังมีจำนวนไม่มาก (Jhonnerie *et al.*, 2015)

การศึกษานี้มีวัตถุประสงค์เพื่อประเมินประสิทธิภาพของการนำวิธีการจำแนกแบบ RF มาใช้ร่วมกับข้อมูลภาพถ่ายจากดาวเทียมแบบรายละเอียดสูง Quickbird ในการจำแนกพันธุ์ไม้ป่าชายเลนเชิงจุดภาพ (Pixel based classification) บริเวณแหลมตะลุมพุก อำเภอปากพนัง จังหวัดนครศรีธรรมราช โดยทำการทดสอบพารามิเตอร์ภายในอัลกอริทึม RF เพื่อให้ผลลัพธ์ที่ได้มีความถูกต้องสูงที่สุด จากนั้นทำการตรวจสอบความถูกต้องและเปรียบเทียบผลลัพธ์ที่ได้จากการจำแนกกับวิธีการจำแนกแบบความน่าจะเป็นสูงสุด (Maximum likelihood classification, MLC) ซึ่งเป็นวิธีการจำแนกมาตรฐานสำหรับงานริโมทเซนซิง



ภาพที่ 1 ขั้นตอนการทำงานของวิธีการจำแนกด้วยป่าสุ่ม ดัดแปลงจาก Luo *et al.* (2016)



ภาพที่ 2 พื้นที่ศึกษาบริเวณพื้นที่อนุรักษ์ป่าชายเลนแหลมตะดุมพุก อำเภอปากพนัง จังหวัดนครศรีธรรมราช

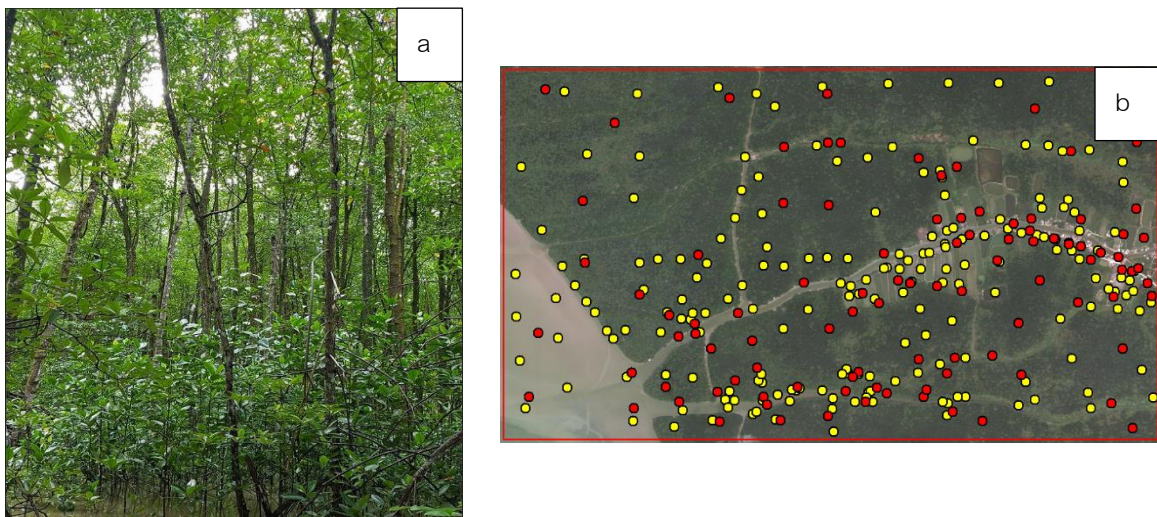
### วิธีดำเนินการวิจัย

พื้นที่ศึกษาตั้งอยู่ในเขตพื้นที่อนุรักษ์ป่าชายเลนบริเวณแหลมตะดุมพุก อำเภอปากพนัง จังหวัดนครศรีธรรมราช โดยการศึกษาครั้งนี้ได้ทำการเลือกแปลงทดสอบบริเวณทางตอนใต้ของพื้นที่อนุรักษ์ป่าชายเลนขนาด 8.5 ตารางกิโลเมตร ตำแหน่งพิกัด UTM โซน 47N พิกัด X 628000 - 632000 และพิกัด Y 927700 - 930000 ซึ่งเป็นพื้นที่ที่ได้รับการสำรวจอย่างทั่วถึง (ภาพที่ 2) และมีการปรับข้อมูลให้เป็นปัจจุบัน ในพื้นที่ศึกษาพบพันธุ์ไม้ป่าชายเลนที่โดดเด่นอยู่ 3 สายพันธุ์ ได้แก่ โกงกางใบเล็ก (*Rhizophora apiculata* Blume, Ra) โกงกางใบใหญ่ (*Rhizophora mucronata* Poir, Rm) และ แสมขาว (*Avicennia alba* Blume, Aa) โดยมีต้นโกงกางใบเล็กเป็นพันธุ์ไม้ที่มีจำนวนมากที่สุด แทรกด้วยโกงกางใบใหญ่ที่ฝักจะลอยน้ำมาติดกับต้นโกงกางใบเล็ก และเจริญเติบโต (ภาพที่ 3a) ในส่วนของต้นแสมขาวจะพบเจริญเติบโตอยู่ตามริมคลองทั้งสองฝั่ง นอกจากนี้ยังพบพื้นที่ที่พันธุ์ไม้ป่าชายเลนเจริญเติบโตในพื้นที่เดียวกันกับบ่อเลี้ยงกุ้ง และชุมชนของชาวบ้านที่ตั้งอยู่ในพื้นที่ศึกษา

ข้อมูลที่ใช้ในการวิจัยประกอบด้วยข้อมูลรีโมทเซนซิง และข้อมูลจากการสำรวจภาคสนาม ข้อมูลด้านรีโมทเซนซิงได้ใช้ภาพถ่ายจากดาวเทียม Quickbird ซึ่งได้ทำการบันทึกข้อมูลเมื่อวันที่ 13 ตุลาคม 2552 จำนวน 4 ช่วงคลื่น ประกอบด้วย น้ำเงิน เขียว แดง และอินฟราเรดใกล้ ข้อมูลภาพถ่ายได้ถูกทำการปรับแก้เชิงคลื่น (Radiometric correction) ในระดับ 2A พร้อมได้ทำการปรับแก้ข้อมูลชั้นบรรยากาศ (Atmospheric correction) พร้อมแปลงค่าการสะท้อนด้วยแบบจำลอง MODTRAN-4 ร่วมกับ FLAASH จากนั้นทำการปรับแก้เชิงเรขาคณิต (Geometric correction) โดยใช้ข้อมูลภาพถ่ายออร์โธรีคจากสำนักงานเศรษฐกิจการเกษตรเป็นข้อมูลอ้างอิง ทำการประมาณค่า (Resampling) ด้วยวิธี Nearest

neighbor และกำหนดค่าคลาดเคลื่อนให้ไม่เกิน 1 จุดภาพ ภาพแต่ละช่วงคลื่นที่ได้รับการปรับแก้แล้วถูกนำมารวมชั้นข้อมูล (Layer stacking) เพื่อนำเข้าสู่การวิเคราะห์จำแนกในขั้นต่อไป

ข้อมูลภาคสนามได้ทำการบันทึกช่วงเดือนกุมภาพันธ์ และมีนาคม 2554 โดย Koedsin & Vaiphasa (2013) ซึ่งทำการวางแปลงขนาด 30×30 ตารางเมตร โดยได้เลือกข้อมูลในแปลงทดสอบมาจำนวน 300 แปลง ในแต่ละแปลงทำการบันทึกพันธุ์ไม้ที่มีขนาดสูงเกิน 2.5 เมตร จากนั้น ข้อมูลพันธุ์ไม้ในแต่ละแปลงทั้งหมดจะถูกนำมาจัดกลุ่มอีกครั้งโดยอ้างอิงจากพันธุ์ไม้ที่โดดเด่นจำนวน 5 สายพันธุ์ที่พบในพื้นที่ป่าชายเลนอำเภอปากพนัง ข้อมูลได้รับการสำรวจ และปรับปรุงตรวจสอบข้อมูลให้เป็นปัจจุบันครั้งที่ 2 ในปี พ.ศ. 2558 โดย Watanakij & Vaiphasa (2016) จากนั้น ได้ลงสำรวจเพื่อเก็บข้อมูลเพิ่มเติม และตรวจสอบความถูกต้องของข้อมูลเป็นครั้งที่ 3 ในเดือนธันวาคม 2560 จากข้อมูลที่ได้รับการปรับปรุงนั้น แสดงให้เห็นว่าพื้นที่ป่าชายเลนเขตร้อนในพื้นที่ศึกษามีสภาพเช่นเดียวกับช่วงเวลาที่ทำการบินถ่ายภาพถ่ายจากดาวเทียม เนื่องจากบริเวณดังกล่าว เป็นพื้นที่อนุรักษ์ที่ปราศจากการบุกรุกของชาวบ้านในพื้นที่ และจากการศึกษาของ Koedsin & Vaiphasa (2013); Watanakij & Vaiphasa (2016) พบว่า ในช่วงเวลาตั้งแต่ทำการบันทึกภาพถ่ายจากดาวเทียมจนถึงการสำรวจเก็บข้อมูลภาคสนาม ป่าชายเลนเขตร้อนในพื้นที่ศึกษาไม่มีการเปลี่ยนแปลงอย่างชัดเจน ข้อมูลแปลงทดสอบถูกนำมาแบ่งเป็นสองชุด ได้แก่ ชุดสำหรับการสอนจำนวน 200 แปลง และชุดสำหรับการทดสอบจำนวน 100 แปลง (ภาพที่ 3b) ด้วยวิธีการสุ่มแบบเลือกลำดับชั้นภูมิ (Stratified random sampling) ในการศึกษานี้ได้แบ่งคลาสที่ใช้ในการจำแนกออกเป็น 6 คลาส ได้แก่ ต้นโกงกางใบเล็ก (Ra) ต้นโกงกางใบใหญ่ (Rm) ต้นแสมขาว (Aa) พันธุ์ไม้ผสม (Mx) พื้นที่เมือง (Ur) และพื้นที่น้ำ (Wa)



ภาพที่ 3 a) ต้นไม้ป่าชายเลนที่ขึ้นอย่างหนาแน่น และปะปนกัน (ต้นโกงกางใบเล็ก และโกงกางใบใหญ่)

b) ข้อมูลชุดการสอน (จุดสีเหลือง) และการทดสอบ (จุดสีแดง) สำหรับการจำแนกด้วย RF และ MLC

การจำแนกด้วยวิธีการ RF เป็นการนำต้นไม้ตัดสินใจ (Decision tree/ CART) ซึ่งเสนอโดย Breiman *et al.* (1984) มาวิเคราะห์ร่วมกับการจำแนกแบบไรพารามิเตอร์ โดยต้นไม้ตัดสินใจมีลักษณะเป็นการวิเคราะห์แบบลำดับชั้น เริ่มจากรากของต้นไม้ (Root node) ไปยังแต่ละส่วนของกิ่ง และโหนดภายใน (Inner node) จนไปถึงสิ้นสุดที่ส่วนของใบ (Leaf node) เพื่อให้ได้ตัวแทนของคลาส เมื่อนำต้นไม้ตัดสินใจมาวิเคราะห์ร่วมกันเป็นกลุ่มเปรียบเสมือนป่าของต้นไม้

ตัดสินใจ ผลลัพธ์สุดท้ายจะถูกทำการเลือกจากค่าฐานนิยม (ผลลัพธ์ที่มีจำนวนมากที่สุด) ดังแสดงในสมการที่ 1 ดังนั้น คลาสที่ได้จากต้นไม้ตัดสินใจจำนวนมากที่สุดจะเป็นคำตอบของการวิเคราะห์ในแต่ละครั้ง

$$H_r(X) = \arg \max_j \sum_{i=1}^k (I(h(X, \theta_i) = j)) \quad (1)$$

เมื่อ  $H(X)$  แทนค่าแบบจำลองการจำแนกรวม ( $I(h(X, \theta_i))$ ) เป็นผลลัพธ์ที่ได้จากการจำแนกด้วยต้นไม้ตัดสินใจที่  $i$  ของค่านำเข้า  $X$

อย่างไรก็ตาม ในการจำแนกด้วย RF จำเป็นต้องทำการปรับพารามิเตอร์ให้มีความเหมาะสมต่อการจำแนก (Jhonnerie et al., 2015) ในการศึกษานี้ได้ทำการทดสอบปรับค่าพารามิเตอร์ที่มีผลต่อการจำแนกจำนวน 3 ค่า (ตารางที่ 1) ได้แก่ ค่าความลึกสูงสุดของต้นไม้ (Maximum total of tree depth) ค่าจำนวนตัวอย่างต่ำสุดของแต่ละโหนด (Minimum number of sample node) และ จำนวนของต้นไม้สูงสุดในป่า (Maximum number of tree in forest) ด้วยวิธีการ Out of bag (OOB) ซึ่งสามารถใช้แทนการนำข้อมูลชุดการสอนมาทำการทดสอบความถูกต้องของการทำนายค่าการจำแนก และช่วยลดค่า Overfitting จากการทดสอบ (Breiman, 1996) ค่าพารามิเตอร์จะถูกนำไปใช้ในการทดสอบ อัลกอริทึม RF โดยค่าที่ให้ความถูกต้องโดยรวม (Overall accuracy, OA) สูงที่สุดจะถูกนำมาใช้เป็นพารามิเตอร์ของ RF ในการจำแนกพันธุ์ไม้ป่าชายเลนเขตร้อน

ตารางที่ 1 พารามิเตอร์ และค่าที่ใช้ในการปรับพารามิเตอร์ เพื่อให้ได้ค่าที่เหมาะสมสำหรับการจำแนก

พารามิเตอร์	ค่าที่ใช้ในการปรับพารามิเตอร์
ค่าความลึกของต้นไม้สูงสุด	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1,000
ค่าจำนวนตัวอย่างต่ำสุดของแต่ละโหนด	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1,000
จำนวนของต้นไม้สูงสุดในป่า	10, 20, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1,000

ผลลัพธ์ที่ได้จากการจำแนก จะถูกทำการวัดประสิทธิภาพด้วยเมตริกซ์ความคลาดเคลื่อน (Confusion matrix) เพื่อตรวจสอบความถูกต้องของผลจำแนก (Congalton & Green, 2009) และใช้ค่าสถิติโคเฮนส์ แคปป์ (Cohen's kappa) ดังสมการที่ 2 ตรวจสอบความสอดคล้องในการจำแนก (Anthony & Joanne, 2005) จากนั้น นำผลลัพธ์ที่ได้จากวิธี MLC (วิธีการจำแนกเชิงจุดภาพ) มาเปรียบเทียบกับข้อมูลที่ใช้ในการสอนจะเป็นข้อมูลเดียวกันกับที่ใช้ใน RF เพื่อลดความลำเอียงในการจำแนกจุด

$$K = \frac{P(O) - P(E)}{1 - P(E)} \quad (2)$$

ให้  $P(O)$  เป็นค่าความสอดคล้องของข้อมูลทำการจำแนก และ  $P(E)$  แทนค่าความสอดคล้องของข้อมูลทำการอ้างอิง ค่าโคเฮนส์แคปป์ สามารถแบ่งออกได้เป็นสามช่วง (Congalton & Green, 2009) ได้แก่ 1) มากกว่า 0.8 หมายถึงมีความสอดคล้องกันมาก 2) ระหว่าง 0.4 ถึง 0.8 หมายถึง มีความสอดคล้องกันปานกลาง และ 3) น้อยกว่า 0.4 มีความสอดคล้องกันต่ำ

สุดท้ายผลจากการจำแนกที่ทำการเปรียบเทียบกัน จะถูกทดสอบด้วยค่าสถิติการทดสอบ Z ดังสมการที่ 3 (Congalton & Green, 2009) เพื่อพิสูจน์ว่าผลการจำแนกที่ได้จากวิธีการที่ต่างกันนั้น มีความแตกต่างกันอย่างมีนัยสำคัญหรือไม่

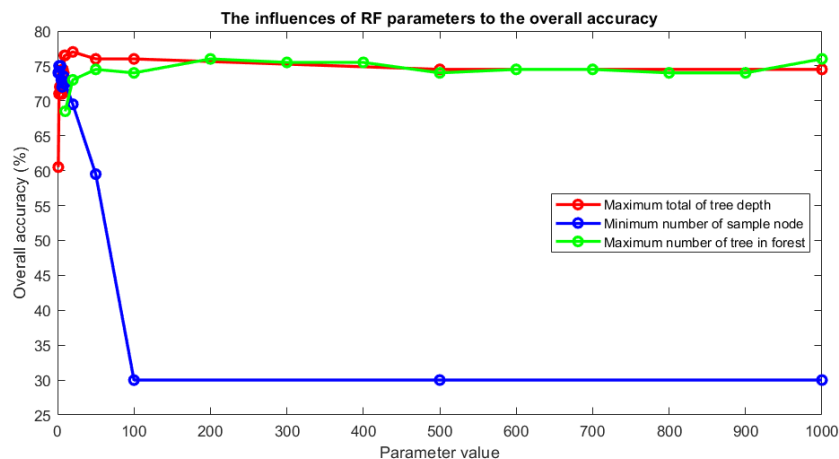
$$Z = \frac{|\hat{K}_1 - \hat{K}_2|}{\sqrt{\hat{v}ar(\hat{K}_1) - \hat{v}ar(\hat{K}_2)}} \quad (3)$$

เมื่อ  $\hat{K}_1, \hat{K}_2$  เป็นค่าประมาณแคปปาของเมตริกซ์ค่าคลาดเคลื่อนที่ 1 และ 2 และ  $\hat{v}ar(\hat{K}_1) - \hat{v}ar(\hat{K}_2)$  ค่าประมาณความแปรปรวนของแคปปา จากสมการที่ 2-30 ค่าสถิติ Z เป็นค่ามาตรฐาน และมีการแจกแจงแบบปกติ กำหนดให้สมมติฐานหลัก  $H_0: (K_1 - K_2) = 0$  (ค่าไม่แตกต่าง) และสมมติฐานรอง  $H_1: (K_1 - K_2) \neq 0$  (ค่ามีความแตกต่าง) โดยที่  $H_0$  จะถูกปฏิเสธเมื่อ  $Z \geq Z_{\alpha/2}$  เมื่อ  $\alpha/2$  คือค่าระดับความเชื่อมั่นของการทดสอบค่าสถิติ Z แบบสองทาง และให้องศาอิสระมีค่าอนันต์

การเตรียมข้อมูลสารสนเทศภูมิศาสตร์ การวิเคราะห์ภาพถ่ายจากดาวเทียม และการแสดงผลใช้ซอฟต์แวร์รหัสเปิด QGIS เวอร์ชัน 2.18 การทดสอบพารามิเตอร์ และการจำแนกพื้นที่ป่าชายเลนเขตร้อนใช้ซอฟต์แวร์รหัสเปิด Python เวอร์ชัน 3.6 ทำการติดตั้ง Library ที่เกี่ยวข้อง ได้แก่ Scikitlearn Gdal และ Numpy การวิเคราะห์ค่าสถิติ ใช้ซอฟต์แวร์รหัสเปิด Octave GUI

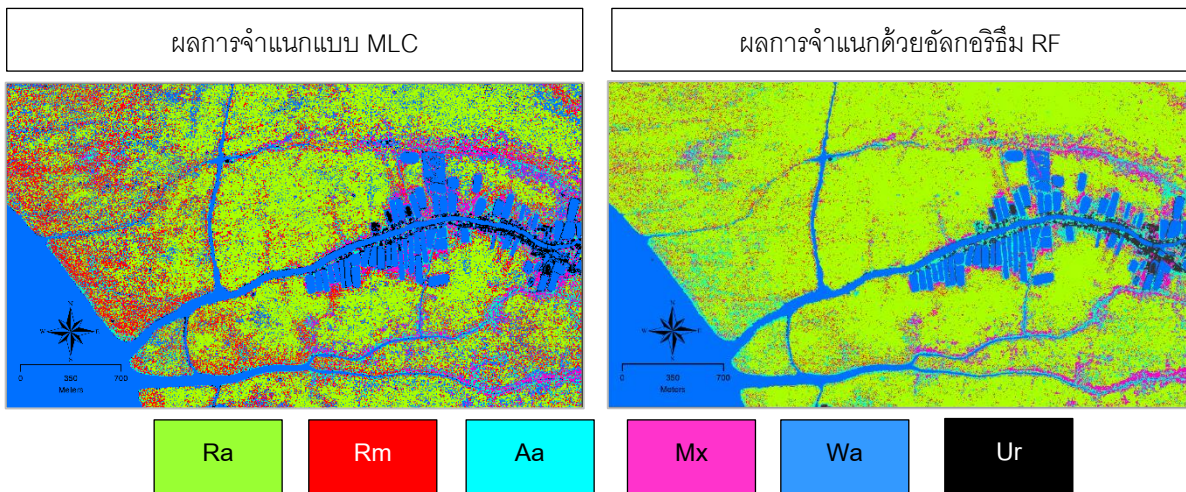
## ผลการวิจัย

การทดสอบพารามิเตอร์สำหรับการจำแนกด้วยอัลกอริทึม RF เริ่มจากการใช้ค่าความลึกสูงสุดของต้นไม้เริ่มต้นที่ 1 พบว่า OA มีค่าร้อยละ 60.50 เมื่อทำการเพิ่มค่าพารามิเตอร์ความลึกของต้นไม้ขึ้นไป พบว่า ค่า OA เพิ่มขึ้นตาม และเริ่มมีความเสถียรที่ค่าความลึกของต้นไม้เท่ากับ 7 มีค่า OA เฉลี่ยร้อยละ 75.50 โดยค่าความลึกสูงสุดของต้นไม้ที่ 20 ให้ค่า OA ร้อยละ 77.00 ค่าจำนวนตัวอย่างต่ำสุดของแต่ละโหนดให้ค่า OA สูงที่สุดที่ค่า 2 โดยมีความถูกต้องที่ร้อยละ 75.00 จากนั้น ทำการทดสอบค่าจำนวนต้นไม้สูงสุดในปี พบว่า จำนวนต้นไม้ที่ 200 และ 1,000 ให้ค่า OA สูงที่สุดที่ร้อยละ 76.00 เมื่อนำค่าพารามิเตอร์ทั้งสามมาทำการทดสอบความเหมาะสม พบว่า ชุดพารามิเตอร์ค่าความลึกสูงสุดของต้นไม้ ค่าจำนวนตัวอย่างต่ำสุดของแต่ละโหนด และค่าจำนวนต้นไม้สูงสุดในปี ที่ 20 2 และ 200 (ภาพที่ 4) ให้ค่าความเหมาะสมที่ร้อยละ 75.50

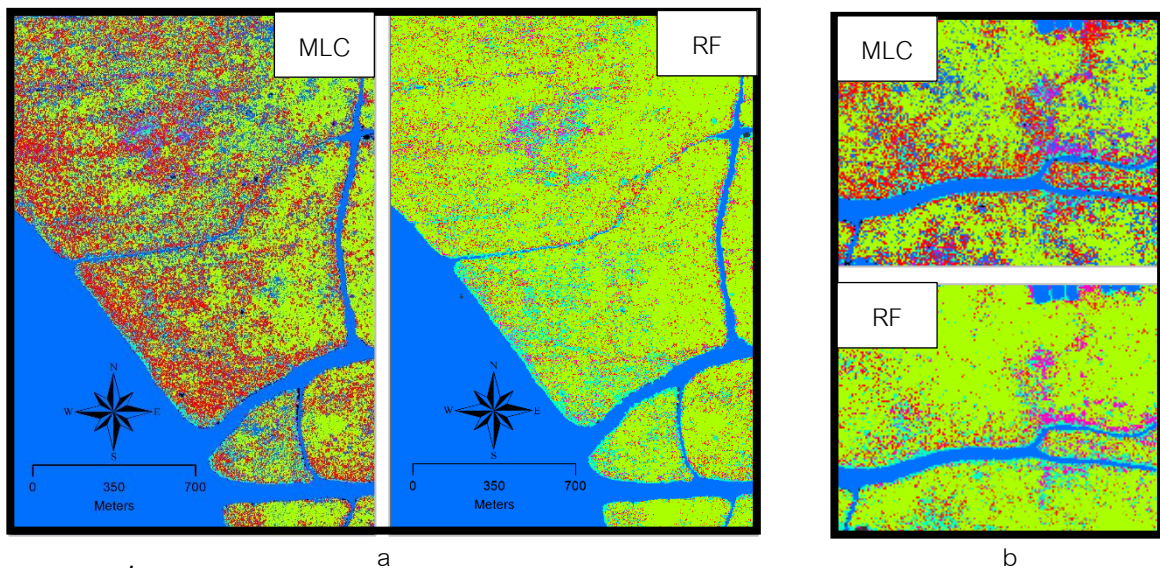


ภาพที่ 4 ค่าความเหมาะสมของพารามิเตอร์ที่ส่งผลต่อความถูกต้องโดยรวมของการจำแนกด้วยอัลกอริทึม RF

ผลการจำแนกด้วย RF และ MLC ได้แสดงในภาพที่ 4 ในการตรวจสอบความถูกต้องของการจำแนก ได้ใช้ชุดข้อมูลสำหรับการทดสอบชุดเดียวกัน ผลจากการจำแนกด้วย RF ให้ค่า OA ร้อยละ 78.00 (ตารางที่ 2) ค่าสถิติโคเฮนส์ แคปปาที่ 0.72 (ตารางที่ 4) สูงกว่าผลการจำแนกด้วย MLC ที่ให้ค่า OA ร้อยละ 56.00 (ตารางที่ 3) ค่าสถิติโคเฮนส์ แคปปาที่ 0.44 (ตารางที่ 4) จากค่าสถิติโคเฮนส์ แคปปา แสดงให้เห็นว่า ความถูกต้องจากการจำแนกข้อมูลจากภาพถ่ายจากดาวเทียม และความถูกต้องจากการตีความหมายข้อมูลภาคพื้นดินจากวิธีการจำแนก RF ให้ค่าความสอดคล้องที่ดีกว่าวิธีการจำแนกแบบ MLC (Anthony & Joanne, 2005) เมื่อทำการทดสอบด้วยค่าสถิติการทดสอบ Z (ตารางที่ 4) พบว่า ผลที่ได้จากการจำแนกทั้งสองวิธีมีความแตกต่างกันอย่างมีนัยสำคัญ (Congalton & Green, 2009) โดยมีค่า 3.68 (มากกว่า 1.96) ที่ระดับความเชื่อมั่น 0.05 เมื่อพิจารณาจากภาพที่ 5a สามารถเห็นได้ชัดเจนว่า ต้นโกงกางใบใหญ่ที่ปรากฏให้เห็นเป็นจำนวนมากในผลจากการจำแนกด้วย MLC ถูกแทนที่ด้วยพื้นที่ของต้นโกงกางใบเล็กในการจำแนกด้วย RF ซึ่งให้ผลตรงกับผลการสำรวจภาคสนามที่พื้นที่ในบริเวณดังกล่าว นอกจากนี้ ตามแนวคลอง (ภาพที่ 5b) ที่การจำแนกด้วย MLC ให้ค่าเป็นต้นโกงกางใบใหญ่นั้น การจำแนกด้วย RF ให้ผลลัพธ์เป็นต้นแสมขาว ซึ่งตรงกับผลที่ได้จากการสำรวจภาคสนามเช่นกัน



ภาพที่ 4 ผลการจำแนกด้วยอัลกอริธึม RF และ MLC



ภาพที่ 5 a) การจำแนกด้วย RF ให้ค่าคลาสดต้นโกงกางใบเล็กได้ถูกต้องมากกว่าการจำแนกด้วย MLC

b) ผลการจำแนกต้นแสมขาวตามแนวคลอง ได้รับการปรับปรุงด้วยวิธีการ RF ให้มีความถูกต้องมากขึ้น



**ตารางที่ 2** เมตริกซ์ความคลาดเคลื่อนของการจำแนกด้วย RF

		ข้อมูลจากการสำรวจภาคสนาม						ผลรวม	ความถูกต้อง ผู้ใช้ (ร้อยละ)	ความถูกต้อง ผู้ผลิต (ร้อยละ)
		Ra	Rm	Aa	Mx	Wa	Ur			
ผลการจำแนกแบบ RF	Ra	27	2	1	0	0	0	30	90.00	72.97
	Rm	4	5	0	1	0	0	10	50.00	71.43
	Aa	1	0	9	2	0	0	12	75.00	64.29
	Mx	5	0	3	10	0	0	18	55.56	76.92
	Wa	0	0	1	0	12	2	15	80.00	100.00
	Ur	0	0	0	0	0	15	15	100.00	88.24
	ผลรวม		37	7	14	13	12	17	100	

**ตารางที่ 3** เมตริกซ์ความคลาดเคลื่อนของการจำแนกด้วย MLC

		ข้อมูลจากการสำรวจภาคสนาม						ผลรวม	ความถูกต้อง ผู้ใช้ (ร้อยละ)	ความถูกต้อง ผู้ผลิต (ร้อยละ)
		Ra	Rm	Aa	Mx	Wa	Ur			
ผลการจำแนกแบบ MLC	Ra	21	3	2	0	4	0	30	70.00	75.00
	Rm	2	5	1	2	0	0	10	50.00	33.34
	Aa	1	2	8	1	0	0	12	66.67	53.34
	Mx	4	5	4	3	4	0	18	16.67	50.00
	Wa	0	0	0	0	12	3	15	80.00	41.38
	Ur	0	0	0	0	9	6	15	40.00	66.67
	ผลรวม		28	15	15	6	29	9	100	

**ตารางที่ 4** ผลการทดสอบความถูกต้องโดยรวม ค่าความสอดคล้อง และเปรียบเทียบประสิทธิภาพของวิธีการจำแนก

วิธีการจำแนก	ความถูกต้องโดยรวม (ร้อยละ)	ค่าสถิติโคเฮนส์ แคปปา	ค่าสถิติการทดสอบ Z (ระดับความเชื่อมั่น ร้อยละ 95)
RF	78.00	0.72	3.68
MLC	56.00	0.44	

**วิจารณ์ผลการวิจัย**

ผลการจำแนกระดับพันธุ์ไม้ป่าชายเลนในงานวิจัยนี้ ให้ค่าความถูกต้องของการจำแนกโดยรวมที่ร้อยละ 78.00 ซึ่งมีค่าต่ำกว่าค่าความถูกต้องของผลการจำแนกโดยรวมในระดับการใช้ที่ดิน และพื้นที่ปกคลุมดิน (Land use and land cover: LULC) ที่ให้ความถูกต้องโดยรวมสูงกว่าร้อยละ 80.00 ขึ้นไป (Jhonnerie *et al.*, 2015; Luo *et al.*, 2016) เนื่องจากการผสมค่าช่วงคลื่น (Spectral mixing) ของคลาสระดับสายพันธุ์มีความสับสนสูงกว่าช่วงคลื่นการจำแนกในระดับการใช้ที่ดิน และพื้นที่ปกคลุมดิน (Jhonnerie *et al.*, 2015) อย่างไรก็ตาม จากการตรวจสอบภาคสนาม พบว่า RF ช่วยแก้ปัญหาด้านการจำแนกที่ผิดพลาด (Misclassification) ที่พบในการจำแนกด้วย MLC (Jhonnerie *et al.*, 2015;

Koedsin & Vaiphasa, 2013) โดยเฉพาะต้นโกงกางใบเล็ก และต้นโกงกางใบใหญ่ ที่มีค่าการสะท้อนแสงแทบจะเป็นค่าเดียวกัน มีเพียงคลาสพื้นที่น้ำเท่านั้น ที่มีความแตกต่างกับคลาสนอื่น ๆ ชัดเจน ผลที่ได้จากงานวิจัยนี้ ช่วยยืนยันถึงประสิทธิภาพของการจำแนกเชิงจุดภาพด้วย RF ที่ให้ผลการจำแนกดีกว่าวิธี MLC

การใช้ข้อมูลภาพถ่ายจากดาวเทียมที่มีจำนวนช่วงคลื่นหลายช่วงคลื่น หรือมีความละเอียดสูงสามารถช่วยแก้ปัญหาการผสมค่าช่วงคลื่นได้ ในการศึกษาของ Koedsin & Vaiphasa (2013) ได้นำภาพถ่ายจากดาวเทียมแบบไฮเพอร์สเปกตรัล ซึ่งมีช่วงคลื่นที่แคบจำนวนหลายช่วงคลื่นมาใช้ในการจำแนกพันธุ์ไม้ป่าชายเลน โดยผลลัพธ์ที่ได้จากการจำแนกมีความถูกต้องโดยรวมของการจำแนกสูงถึงร้อยละ 92.00 อย่างไรก็ตาม ด้วยความละเอียดของจุดภาพที่มีขนาด  $30 \times 30$  เมตร ส่งผลให้การจำแนกต้นไม้ป่าชายเลนในบางสายพันธุ์ เช่น ต้นโกงกางใบเล็ก และต้นโกงกางใบใหญ่ มีความถูกต้องต่ำ ในส่วนของภาพถ่ายดาวเทียมรายละเอียดสูง มีข้อได้เปรียบในเรื่องความละเอียดของจุดภาพ สามารถประยุกต์ร่วมกับวิธีการจำแนกเชิงวัตถุได้เป็นอย่างดี (Watanakij & Vaiphasa, 2016) อย่างไรก็ตาม เมื่อนำภาพถ่ายจากดาวเทียมรายละเอียดสูงมาทำการจำแนกแบบเชิงจุดภาพ ผลการจำแนกยังให้ค่าความถูกต้องโดยรวมต่ำ เป็นผลมาจากการผสมค่าช่วงคลื่นในแต่ละจุดภาพที่มีค่าการสะท้อนที่ใกล้เคียงกันมาก

การประยุกต์วิธีการจำแนกเชิงวัตถุร่วมกับ RF ร่วมกับภาพถ่ายจากดาวเทียมรายละเอียดสูง สามารถปรับปรุงผลการจำแนกพันธุ์ไม้ป่าชายเลนเขตร้อนให้มีความถูกต้องโดยรวมเพิ่มสูงขึ้น เช่นการศึกษาของ Wang *et al.* (2018) ที่สามารถเพิ่มความถูกต้องโดยรวมของการจำแนกจากร้อยละ 79.00 เป็นร้อยละ 82.00 ดังนั้น การประยุกต์วิธีการจำแนกเชิงวัตถุ ร่วมกับการนำช่วงคลื่นแบบ Panchromatic มาผสมรวมกับช่วงคลื่นที่ตามองเห็นของภาพถ่ายจากดาวเทียมรายละเอียดสูง จะสามารถช่วยให้ผลลัพธ์จากการจำแนกมีความถูกต้องมากยิ่งขึ้น และเป็นเรื่องที่น่าสนใจสำหรับการศึกษาด้านการจำแนกพันธุ์ไม้ป่าชายเลนในครั้งต่อไป

### สรุปผลการวิจัย

การศึกษานี้แสดงให้เห็นถึงประสิทธิภาพของการนำ RF มาใช้เพื่อจำแนกพันธุ์ไม้ป่าชายเลนเขตร้อนจากข้อมูลภาพถ่ายจากดาวเทียมรายละเอียดสูง Quickbird ร่วมกับข้อมูลที่ได้จากการสำรวจภาคสนาม ผลที่ได้จากการจำแนกพบว่า การใช้วิธีการ RF ในการจำแนกพันธุ์ไม้ป่าชายเลนเขตร้อนให้มีความถูกต้องโดยรวมสูงกว่า และสามารถช่วยแยกแยะพันธุ์ไม้ที่มีความสับสนเชิงคลื่นออกจากกันได้ดีกว่าเมื่อเปรียบเทียบกับวิธีการจำแนกแบบ MLC อย่างมีนัยสำคัญ อย่างไรก็ตาม หากมีการนำไปประยุกต์ร่วมกับวิธีการจำแนกเชิงวัตถุ และเพิ่มความละเอียดของจุดภาพถ่ายจากดาวเทียมด้วยช่วงคลื่น Panchromatic ความถูกต้องของผลการจำแนกพันธุ์ไม้ป่าชายเลนเขตร้อนมีแนวโน้มที่จะได้รับการปรับปรุงให้มีประสิทธิภาพมากขึ้นกว่าเดิม

### กิตติกรรมประกาศ

ขอขอบคุณ ดร.ชัยโชค ไวภาษา ดร.วีรพงษ์ เกิดสิน และ ดร.ณกร วัฒนกิจ ที่ได้อนุเคราะห์ข้อมูลบางส่วนในการศึกษา ขอขอบคุณสถานีพัฒนาทรัพยากรป่าชายเลนที่ 15 และ 44 นครศรีธรรมราช กรมทรัพยากรทางทะเลและชายฝั่ง ที่ให้ความอนุเคราะห์บุคคลากร และเรือสำหรับการสำรวจเก็บข้อมูลภาคสนาม

### เอกสารอ้างอิง

Anthony, J. V., & Joanne, M. G. (2005). Understanding Interobserver Agreement: The Kappa Statistics. *Family Medicine*, 37(5), 3.

- Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *Annals of Statistics*, 24(6), 2350-2383.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees*: Taylor & Francis.
- Chan, J. C. W., & Paelinckx, D. (2008). Evaluation of random forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing Environment*, 112, 2999-3011.
- Congalton, R. G., & Green, K. (2009). *Assessing the accuracy of remotely sensed data. [electronic resource] : principles and practices*: Boca Raton : CRC Press/Taylor & Francis, c2009. 2nd ed.
- Dubeau, P., King, D. J., Unbushe, D. G., & Rebelo, L. M. (2017). Mapping the Dabus Wetlands, Ethiopia, Using Random Forest Classification of Landsat, PALSAR and Topographic Data. *Remote Sensing*, 9(10). doi:10.3390/rs9101056
- Hamdan, O., Khairunnisa, M. R., Ammar, A. A., Hasmadi, I. M., & Aziz, H. K. (2013). MANGROVE CARBON STOCK ASSESSMENT BY OPTICAL SATELLITE IMAGERY. *Penilaian stok karbon hutan paya laut menggunakan imej satelit optik.*, 25(4), 554-565.
- Heumann, B. W. (2011). Satellite remote sensing of mangrove forests: Recent advances and future opportunities. *Progress in Physical Geography*, 35(1), 87-108. doi:10.1177/0309133310385371
- Jhonnerie, R., Siregar, V. P., Nababan, B., Prasetyo, L. B., & Wouthuyzen, S. (2015). Random Forest Classification for Mangrove Land Cover Mapping Using Landsat 5 TM and Alos Palsar Imageries. *Procedia Environmental Sciences*, 24, 215-221. doi:https://doi.org/10.1016/j.proenv.2015.03.028
- Koedsin, W., & Vaiphasa, C. (2013). Discrimination of Tropical Mangroves at the Species Level with EO-1 Hyperion Data. *Remote Sensing*, 5(7), 3562-3582. doi:10.3390/rs5073562
- Kuenzer, C., Bluemel, A., Gebhardt, S., Quoc, T. V., & Dech, S. (2011). Remote Sensing of Mangrove Ecosystems: A Review. *Remote Sensing*, 3(5), 878-928. doi:10.3390/rs3050878
- Lee, S. Y., Primavera, J. H., Dahdouh-Guebas, F., McKee, K., Bosire, J. O., Cannicci, S., Record, S. (2014). Ecological role and services of tropical mangrove ecosystems: a reassessment. *Global Ecology and Biogeography*, 23(7), 726-743. doi:10.1111/geb.12155
- Luo, Y. M., Huang, D. T., Liu, P. Z., & Feng, H. M. (2016). An novel random forests and its application to the classification of mangroves remote sensing image. *Multimedia Tools and Applications*, 75(16), 9707-9722. doi:10.1007/s11042-015-2906-9
- Ramo, R., & Chuvieco, E. (2017). Developing a Random Forest Algorithm for MODIS Global Burned Area Classification. *Remote Sensing*, 9(11). doi:10.3390/rs9111193
- 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. doi:10.1016/j.isprsjprs.2011.11.002

- Sasmito, S. D., Murdiyarso, D., Friess, D. A., & Kurnianto, S. (2016). Can mangroves keep pace with contemporary sea level rise? A global data review. *Wetlands Ecology and Management*, 24(2), 263-278. doi:10.1007/s11273-015-9466-7
- Wang, D. Z., Wan, B., Qiu, P. H., Su, Y. J., Guo, Q. H., & Wu, X. C. (2018). Artificial Mangrove Species Mapping Using Pleiades-1: An Evaluation of Pixel-Based and Object-Based Classifications with Selected Machine Learning Algorithms. *Remote Sensing*, 10(2). doi:10.3390/rs10020294
- Watanakij, N., & Vaiphasa, C. (2016). Improving the accuracy of mangrove species discrimination using object based and high spatial resolution imagery: A case study in Pak Phanang, Thailand. *International Journal of Geoinformatics*, 12(3), 41-49.
- Zhu, Y. H., Liu, K., Liu, L., Myint, S. W., Wang, S. G., Liu, H. X., & He, Z. (2017). Exploring the Potential of WorldView-2 Red-Edge Band-Based Vegetation Indices for Estimation of Mangrove Leaf Area Index with Machine Learning Algorithms. *Remote Sensing*, 9(10). doi:10.3390/rs9101060