



การจำแนกการใช้ประโยชน์ที่ดินในจังหวัดนครนายก ด้วยอัลกอริทึมการเรียนรู้เครื่องและภาพถ่ายจากดาวเทียม Sentinel-2

Land Use Classification in Nakhon Nayok Province

Using Machine Learning Algorithms and Sentinel-2 Image

กฤษฎาณ อินทร์ตัน^{*}

Kritchayan Intarat^{*}

คณะศิลปศาสตร์ สาขาวิชาภูมิศาสตร์ มหาวิทยาลัยธรรมศาสตร์

Faculty of Liberal Arts, Department of Geography, Thammasat University

Received : 14 October 2021

Revised : 8 December 2021

Accepted : 17 December 2021

บทคัดย่อ

ลักษณะความเป็นพลวัตของการใช้ประโยชน์ที่ดินมีส่วนเกี่ยวข้องกับสภาพเศรษฐกิจและสังคมในท้องถิ่นที่เปลี่ยนแปลงไป จังหวัดนครนายกมีลักษณะการใช้ที่ดินหลักเป็นพื้นที่ปลูกพืชไร่และสถานที่ท่องเที่ยวทางธรรมชาติซึ่งมีการเปลี่ยนแปลงอย่างรวดเร็ว การมีข้อมูลการใช้ประโยชน์ที่ดินที่ทันสมัยเป็นสิ่งสำคัญต่อการบริหารจัดการภายในจังหวัด ที่ผ่านมามีภาพถ่ายจากดาวเทียมที่ให้บริการแบบไม่มีค่าใช้จ่ายนิยมถูกนำมาใช้สำหรับช่วยในกระบวนการจำแนกร่วมกับตัวจำแนกประเภทต่าง ๆ นอกจากนี้การเรียนรู้เครื่องได้ถูกนำมาใช้กับงานด้านการจำแนกการใช้ประโยชน์ที่ดินเพื่อเพิ่มประสิทธิภาพของวิธีการจำแนก อย่างไรก็ตามในการจำแนกการใช้ประโยชน์ที่ดินจำเป็นต้องดำเนินการทดสอบกระบวนการจำแนกเพื่อหาตัวจำแนกที่ให้ผลลัพธ์การจำแนกได้มีความถูกต้องมากที่สุด งานวิจัยนี้ได้ทดสอบประสิทธิภาพของตัวจำแนกการเรียนรู้เครื่องจำนวน 4 วิธีการ ได้แก่ ข่ายประสาทเทียม (artificial neural network) ต้นไม้ตัดสินใจ (decision tree) ป่าสุ่ม (random forest) และซัพพอร์ทเวกเตอร์แมชชีน (support vector machine) รวมถึงวิธีการจำแนกแบบมาตรฐานแบบความน่าจะเป็นสูงสุด (maximum likelihood) ในกระบวนการจะทดสอบเพื่อหาไฮเพอร์พารามิเตอร์ของตัวจำแนกที่เหมาะสม ผลที่ได้จะถูกนำมาพิจารณาเพื่อเลือกวิธีที่ให้ค่าความถูกต้องการจำแนกสูงสุดด้วยค่าความถูกต้องโดยรวม F-1score ค่าสถิติแคปปา (kappa coefficient) และค่าสถิติ Z-test ผลการจำแนกพบว่าตัวจำแนกป่าสุ่มมีประสิทธิภาพในการจำแนกการใช้ประโยชน์ที่ดินมากที่สุดโดยมีค่าความถูกต้องโดยรวมที่ร้อยละ 92.00 ตามด้วยต้นไม้ตัดสินใจ ข่ายประสาทเทียม ซัพพอร์ทเวกเตอร์แมชชีน และความน่าจะเป็นสูงสุด (ร้อยละ 84.00 69.00 65.00 และ 63.00 ตามลำดับ) ผลการทดสอบ Z-test ช่วยยืนยันถึงความแตกต่างของตัวจำแนกป่าสุ่มกับตัวจำแนกอื่นอย่างมีนัยสำคัญที่ระดับความเชื่อมั่นร้อยละ 95

คำสำคัญ : การใช้ประโยชน์ที่ดิน ; การจำแนก ; การเรียนรู้เครื่อง ; ภาพถ่ายจากดาวเทียม ; เซนติเนล-2



Abstract

The dynamic of land use can be recognized as the transformation of local socio-economic. Nakhon Nayok Province exhibits the field crop and environmental destination as significant sites, and it further reveals an active transition of local land use. Therefore, acquiring updated land-use information is significant to local management and administration. Recently, satellite imagery has been freely available and incorporated with the classification process using any classifiers. In addition, machine learning classifiers have been introduced to increase the potential of classification. However, applying these classifiers demands experiments to identify the fittest algorithm. This article examined the potential of four machine learning algorithms: analytical neural network (ANN), decision tree (DT), random forest (RF), and support vector machine (SVM), including the standard classifier (maximum likelihood: MLC). Hyperparameters of each machine learning algorithm have been inquired properly. Classification results were then selected to identify the most desirable classifier using overall accuracy, F1-score, kappa coefficient, and Z-statistics. According to the result, the random forest has been reported as the best classifier, which gains 92.00% of overall accuracy—followed by a decision tree, analytical neural network, support vector machine, and maximum likelihood classifier (84.00, 69.00, 65.00, and 63.00 of overall accuracy, respectively). The Z-statistics also confirmed that the random forest had a significant difference compared with other classifiers at a 95% confidence level.

Keywords : land use ; classification ; machine learning ; satellite image ; Sentinel-2

บทนำ

การเพิ่มขึ้นของประชากรส่งผลต่อการเพิ่มขึ้นของอุปสงค์ที่จำเป็นในด้านต่าง ๆ และการเปลี่ยนแปลงลักษณะการใช้ประโยชน์ที่ดิน เช่น ที่อยู่อาศัย พื้นที่เมือง และการเกษตรกรรม ข้อมูลดังกล่าวมีความจำเป็นต่อหน่วยงานไม่ว่าจะเป็นรัฐบาล หน่วยงานภาครัฐส่วนท้องถิ่น และประชาชนทุกภาคส่วนสำหรับใช้บริหารจัดการด้านนโยบายในการพัฒนาพื้นที่ (Mazzia *et al.*, 2020) จังหวัดนครนายกเป็นจังหวัดที่ตั้งอยู่ใกล้กรุงเทพมหานคร แต่ลักษณะของพื้นที่และการใช้ประโยชน์ที่ดินมีความหลากหลายเนื่องจากอีกด้านหนึ่งของจังหวัดมีขอบเขตติดกับอุทยานแห่งชาติเขาใหญ่ ด้วยการเป็นแหล่งท่องเที่ยวที่อยู่ไม่ไกลจากกรุงเทพมหานคร เดินทางสะดวก และมีสภาพแวดล้อมของเมืองที่ยังมีความเป็นธรรมชาติ ทำให้ผู้คนสนใจเข้ามาซื้อที่ดินในจังหวัดนครนายกเป็นจำนวนมาก นอกจากนี้กิจกรรมด้านการท่องเที่ยวเป็นอีกปัจจัยหนึ่งที่ก่อให้เกิดการเปลี่ยนแปลงพื้นที่ภายในจังหวัดอยู่ตลอดเวลา

สำหรับการจำแนกการใช้ประโยชน์ที่ดิน ภาพถ่ายจากดาวเทียมกลายเป็นข้อมูลที่สำคัญและมีประสิทธิภาพ นอกจากนี้ยังช่วยลดเวลาและค่าใช้จ่ายในการดำเนินงาน (Intarat & Sillaparat, 2019) ปัจจุบันมีข้อมูลภาพถ่ายจากดาวเทียมที่ให้บริการโดยไม่มีค่าใช้จ่าย เช่น Landsat (รายละเอียดจุดภาพ 30 เมตร) และ Sentinel-2 (รายละเอียดจุดภาพที่ดีที่สุด 10 เมตร) ซึ่งเป็นภาพถ่ายจากดาวเทียมรายละเอียดปานกลางและมีข้อมูลแบบ multispectral ที่มีช่วงคลื่นให้เลือกใช้ตั้งแต่ช่วงคลื่นตามมองเห็นไปจนถึงช่วงคลื่นอินฟราเรดความร้อนทำให้ผู้ใช้สามารถดาวน์โหลดมาประยุกต์ใช้งานได้หลากหลายไม่ว่าจะเป็นด้านการจำแนกข้อมูลภายในภาพหรือการติดตามการเปลี่ยนแปลงของพื้นที่ในแต่ละช่วงเวลา (Abdi, 2020; Mazzia *et al.*, 2020; Piedadlobo *et al.*, 2019) ข้อมูลภาพถ่ายจากดาวเทียมจะถูกนำมาใช้งานร่วมกับวิธีการจำแนกซึ่งมีอยู่สองประเภท ได้แก่ วิธีการจำแนกแบบไม่กำกับดูแล (unsupervised classification method) และวิธีการจำแนกแบบกำกับดูแล (supervised classification method) โดยวิธีหลังจำเป็นต้องสร้างข้อมูลสอน (training data) เพื่อให้ตัวจำแนก (classifier) สามารถนำไปเปรียบเทียบกับค่าที่อยู่ในจุดภาพของแต่ละชั้นข้อมูลและทำนายผลลัพธ์ออกมา (Richards, 2013) ในการจำแนกภาพถ่ายจากดาวเทียมรายละเอียดปานกลางนิยมใช้วิธีการจำแนกแบบกำกับดูแลร่วมกับวิธีการจำแนกเชิงจุดภาพ (pixel-based classification) ซึ่งมีหลายวิธีการ เช่น วิธีความน่าจะเป็นสูงสุด (maximum likelihood: MLC) พาราเรลไปป์ (parallelepiped: PC) และวิธีระยะทางใกล้ที่สุด (minimum distance: MDC) (Aggarwal, *et al.*, 2016) โดยมีวิธีความน่าจะเป็นสูงสุดเป็นวิธีจำแนกมาตรฐานที่นิยมใช้และให้ความถูกต้องสูงกว่าอีก 2 วิธีการ (de Oliveira Duarte, *et al.*, 2018) อย่างไรก็ตามวิธีความน่าจะเป็นสูงสุดยังพบปัญหาในด้านของความถูกต้องของผลลัพธ์ที่ได้จากการจำแนกเนื่องมาจากสัญญาณรบกวน (noise) ที่เกิดขึ้นและตัวอย่างที่นำมาฝึกสอนต้องมีการกระจายแบบปกติ (Rodriguez-Galiano *et al.*, 2012) ซึ่งในบางครั้งไม่สามารถจัดการให้ข้อมูลฝึกสอนดังกล่าวมีลักษณะการกระจายแบบปกติตามข้อกำหนดได้ จึงได้มีการพัฒนาอัลกอริทึมที่ใช้เป็นตัวจำแนกให้มีประสิทธิภาพที่สูงขึ้นและสามารถแก้ไขปัญหาดังกล่าว

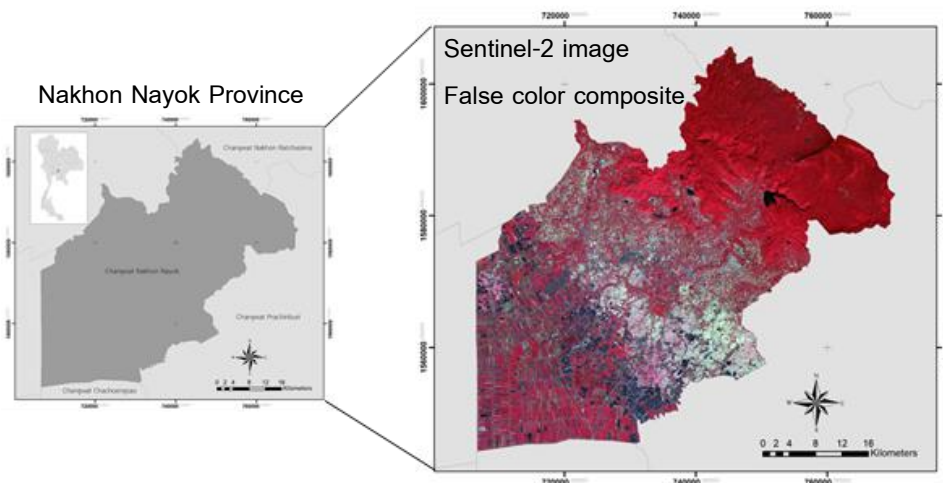
วิธีการจำแนกการใช้ประโยชน์ที่ดินได้รับการพัฒนาให้มีความถูกต้องในการจำแนกสูงมากขึ้น ปัจจุบันอัลกอริทึมการเรียนรู้เครื่อง (machine learning; ML) ได้เข้ามามีบทบาทในการทำงานด้านการประมวลผลภาพมากยิ่งขึ้น อัลกอริทึมต่าง ๆ ได้ถูกพัฒนาเป็นวิธีจำแนกสำหรับใช้งานร่วมกับภาพถ่ายจากดาวเทียมและช่วยปรับปรุงประสิทธิภาพในการจำแนกให้เพิ่มขึ้นสูงกว่าวิธีจำแนกมาตรฐานแบบความน่าจะเป็นสูงสุด ได้มีการทดสอบการจำแนกข้อมูลภาพถ่ายจากดาวเทียมด้วยวิธีการ

จำแนกที่มาจากอัลกอริธึม ML ในงานด้านต่าง ๆ เช่น การจำแนกการใช้ประโยชน์ที่ดิน (Abdi, 2020; Majnoui-Toutakhane, 2020; Rodriguez-Galiano, *et al.*, 2012) การจำแนกพืชพรรณ (Macintyre, *et al.*, 2020) และการจำแนกพันธุ์ไม้ป่าชายเลน (Intarat & Sillapat, 2019; Jhonnerie, 2015) ซึ่งวิธีการจำแนกด้วยอัลกอริธึม ML ให้ผลลัพธ์ที่มีความถูกต้องสูงขึ้นไปอย่างมีนัยสำคัญ

ดังนั้น เพื่อเป็นการทดสอบถึงประสิทธิภาพในการจำแนกการใช้ประโยชน์ที่ดินในจังหวัดนครนายกของอัลกอริธึม ML กระบวนการจำแนกในงานนี้ได้ใช้อัลกอริธึม ML จำนวน 4 ชนิด ได้แก่ ข่ายประสาทเทียม (artificial neural network: ANN) ต้นไม้การตัดสินใจ (decision tree: DT) ป่าสุ่ม (random forest: RF) และซัพพอร์ตเวกเตอร์แมชชีน (support vector machine: SVM) มาใช้เป็นตัวจำแนกร่วมกับภาพถ่ายดาวเทียม Sentinel-2 ด้วยวิธีการจำแนกเชิงจุดภาพ เพื่อเปรียบเทียบหาอัลกอริธึมที่ให้ค่าความถูกต้องในการจำแนกการใช้ประโยชน์ที่ดิน อัลกอริธึมทุกตัวจะถูกทดสอบปรับค่าไฮเพอร์พารามิเตอร์ เพื่อให้มีประสิทธิภาพสำหรับการจำแนกที่ดีที่สุดด้วยข้อมูลการสอนและทดสอบ (train and test datasets) ผลลัพธ์ที่ได้จะถูกนำมาทดสอบความถูกต้องเพื่อแสดงถึงวิธีการจำแนกการใช้ประโยชน์ที่ดินในจังหวัดนครนายกที่ให้ประสิทธิภาพที่ดีที่สุด

วิธีดำเนินการวิจัย

ในงานวิจัยนี้ได้ใช้จังหวัดนครนายกเป็นพื้นที่ศึกษามีพิกัด UTM 47N ระหว่าง 706540 E ถึง 770275 E และ 1544228 N ถึง 1605162 N มีเนื้อที่ 2,122 ตารางกิโลเมตร มีพื้นที่ติดจังหวัดโดยรอบ ได้แก่ ทิศเหนือติดจังหวัดนครราชสีมา และสระบุรี ทิศตะวันออกติดจังหวัดปราจีนบุรี ทิศใต้ติดจังหวัดฉะเชิงเทรา และทิศตะวันตกติดจังหวัดปทุมธานี (ภาพที่ 1) ลักษณะการใช้ประโยชน์ที่ดินหลักเป็นพื้นที่ทำนาและสวนผลไม้ ในปัจจุบันมีการเปลี่ยนแปลงของพื้นที่จากพื้นที่นาและสวนเป็นที่อยู่อาศัยและกิจกรรมด้านการท่องเที่ยวเพิ่มมากขึ้นเนื่องจากจังหวัดนครนายกเป็นจังหวัดที่อยู่ใกล้กรุงเทพฯ และมีแหล่งท่องเที่ยวสำหรับพักผ่อนหย่อนใจ ส่งผลให้ผู้คนนิยมเดินทางมาเที่ยวที่จังหวัดนครนายกในช่วงวันหยุด



ภาพที่ 1 ขอบเขตพื้นที่ศึกษาจังหวัดนครนายก (พื้นที่สี่เหลี่ยม) แสดงภาพถ่ายจากดาวเทียม Sentinel-2 บันทึกภาพเมื่อวันที่ 13 ธันวาคม 2563 ผสมสีเท็จ (false color composite)

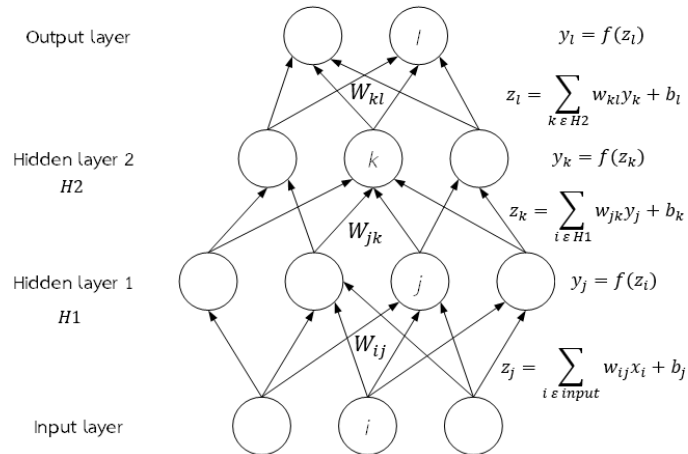


ข้อมูลที่ใช้ในงานวิจัยประกอบด้วยข้อมูลภาพถ่ายจากดาวเทียม Sentinel-2 และข้อมูลที่ได้จากการสำรวจภาคสนาม ข้อมูลภาพถ่ายจากดาวเทียม Sentinel-2 ใช้ภาพระดับ 1C จำนวน 2 ข้อมูลภาพ ได้แก่ T47PQR และ T47PQS ซึ่งเป็นผลิตภัณฑ์ที่ผ่านการจัดทำภาพถ่ายออร์โธสตีและได้รับการปรับแก้ค่าเชิงเรขาคณิตและเชิงรังสีขนาด 12 บิต ค่าที่อยู่ในแต่ละจุดภาพได้รับการปรับให้เป็นค่าการสะท้อนแสง (reflectance) ในระดับใต้ชั้นบรรยากาศ (bottom-of-atmosphere: BOA) ขนาด 16 บิต (Int16) บันทึกภาพเมื่อวันที่ 13 ธันวาคม 2562 เวลา 10.41 น. โดยมีระบบพิกัดอ้างอิงแบบ World Geodetic System 1984 (WGS1984) และระบบพิกัดแบบ Universal Transverse Mercator (UTM) Zone 47 North (ภาพที่ 1) ความละเอียดเชิงพื้นที่ของภาพถ่ายมีขนาด 10 × 10 ตารางเมตร เลือกช่วงคลื่นจำนวนทั้งหมด 4 ช่วงคลื่น ได้แก่ B2 (blue) B3 (green) B4 (red) และ B8 (NIR) สร้างข้อมูลดัชนีพืชพรรณผลต่างนอร์มอลไลซ์ (normalized different vegetation index; NDVI) โดยใช้ช่วงคลื่น B4 และ B8 ดังสมการที่ 1

$$NDVI = \frac{NIR-red}{NIR+red} \quad (1)$$

เมื่อ *NIR* คือช่วงคลื่นอินฟราเรดใกล้และ *red* คือช่วงคลื่นสีแดงของภาพถ่ายจากดาวเทียม Sentinel-2 จากนั้นนำข้อมูลที่ได้ไปรวมกับช่วงคลื่นรวมทั้งหมด 5 ช่วงคลื่นด้วยกระบวนการ layer stack ก่อนนำเข้ากระบวนการวิเคราะห์ ดำเนินการสุ่มข้อมูลตัวอย่างฝึกสอนและทดสอบความถูกต้องของการจำแนกการใช้ประโยชน์ที่ดินจำนวน 1,000 และ 400 ตัวอย่างตามลำดับด้วยวิธีการเลือกตัวอย่างแบบแบ่งชั้นภูมิ (stratified random sampling) โดยใช้แผนที่การใช้ประโยชน์ที่ดินจากกรมพัฒนาที่ดินปี พ.ศ. 2561 ตามลักษณะการใช้ประโยชน์ที่ดินในพื้นที่จำนวน 7 ประเภท ได้แก่ พื้นที่ว่างเปล่า (bareland: Br) สิ่งปลูกสร้าง (buildup: Bu) ไร่ (fieldcrop: Fe) พื้นที่ป่าไม้ (forest: Fr) พื้นที่สวน (orchard: Or) พื้นที่ไม้ยืนต้น (perennial: Pe) และน้ำ (water: Wa) สำหรับการปรับค่าไฮเพอร์พารามิเตอร์ของแต่ละอัลกอริทึม ตัวอย่างฝึกสอนจำนวน 1,000 ตัวอย่างจะถูกแบ่งสัดส่วน 70:30 สำหรับสอนและทดสอบประสิทธิภาพของอัลกอริทึมที่ใช้จำแนก ข้อมูลทั้งหมดได้รับการสำรวจในเดือนพฤศจิกายน 2563 และเดือนกุมภาพันธ์ 2564

ตัวจำแนกที่ใช้จะเป็นอัลกอริทึม ML ทั้งหมด 4 ชนิด ได้แก่ ANN, DT, RF, และ SVM โดยมี MLC เป็นพื้นฐานสำหรับเปรียบเทียบ (baseline) โดยแต่ละวิธีต้องมีการทดสอบไฮเพอร์พารามิเตอร์ (ตารางที่ 1) ANN ได้รับความนิยมในการนำมาใช้ในงานด้านการจำแนกภาพถ่ายจากดาวเทียม (Ge *et al.*, 2020; Majnoui-Toutakhane, 2020; Saboori *et al.*, 2019; Silva *et al.*, 2020) โดยเป็นแบบจำลองทางคณิตศาสตร์ที่จำลองลักษณะการทำงานของประสาทมนุษย์มาใช้สำหรับจดจำและวิเคราะห์ข้อมูลนำเข้าเพื่อทำนายผลลัพธ์ แบบจำลองจะมีการจัดเก็บค่าน้ำหนักภายในโครงข่ายและปรับปรุงให้มีความเหมาะสมเมื่อมีการเรียนรู้เพิ่มเติมคล้ายกับความสามารถของประสาทที่จดจำสิ่งต่าง ๆ ที่ได้เรียนรู้เพิ่มหรือเรียนรู้บ่อยจนสามารถนำไปใช้ในการแก้ปัญหาได้ ตัวโครงข่ายจะแบ่งการเชื่อมต่อกออกเป็นชั้น ประกอบด้วย ชั้นข้อมูลนำเข้า (input layer) ชั้นข้อมูลแอบแฝง (hidden layer) และชั้นข้อมูลส่งออก (output layer) ในการทำงานกับข้อมูลที่มีความซับซ้อน โครงข่ายจะมีจำนวนหลายชั้น (multi-layer perceptron หรือ deep neural network) ดังภาพที่ 2

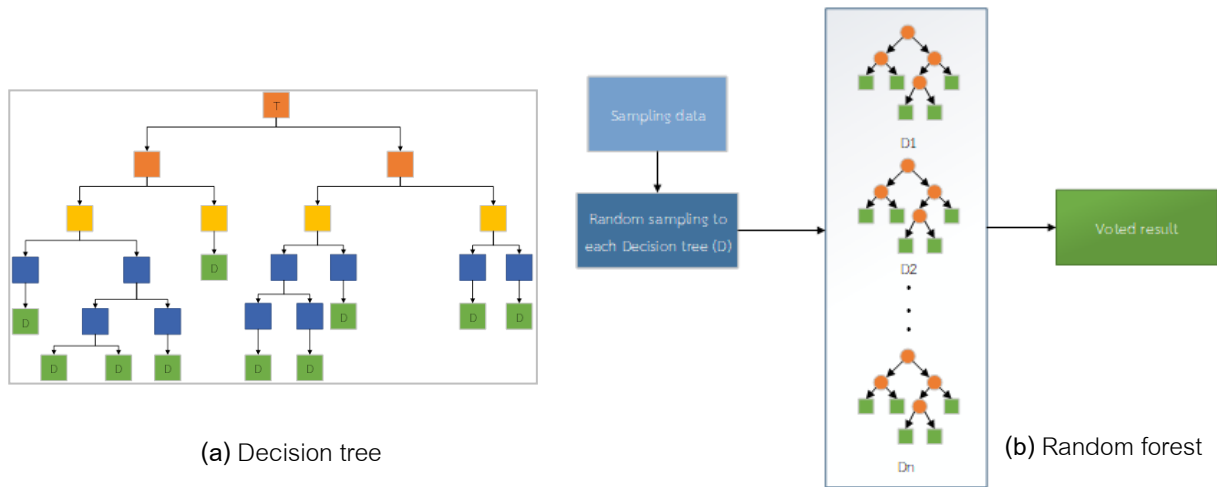


ภาพที่ 2 ข่ายประสาทเทียมแบบ feed forward มีชั้นแอบแฝงจำนวน 2 ชั้น (H1 และ H2)
 ดัดแปลงจาก Shrestha & Mahmood (2019)

ต้นไม้ตัดสินใจเป็นการสร้างแบบจำลองทางคณิตศาสตร์รูปแบบหนึ่งเพื่อหาผลลัพธ์ที่ดีที่สุดตามเงื่อนไขที่กำหนด สามารถวิเคราะห์ได้ทั้งแบบการจำแนกและแบบจำลองการถดถอย (Shrestha & Mahmood, 2019) ลักษณะโครงสร้างเป็นแบบลำดับชั้นคล้ายต้นไม้กลับหัว ต้นไม้ตัดสินใจจะเรียนรู้ข้อมูลสอน (training dataset) เพื่อจัดกลุ่มข้อมูลตามเงื่อนไข หลังจากเสร็จสิ้นกระบวนการเรียนรู้ ต้นไม้ตัดสินใจสามารถนำมาใช้จำแนกชุดข้อมูลในพื้นที่ศึกษาเดียวกันที่ไม่รู้จัก (unknown dataset) ได้ โครงสร้างของต้นไม้ตัดสินใจประกอบด้วย 1) โหนด (node) แสดงคุณสมบัติของแต่ละเงื่อนไขเป็นจุดแบ่ง (threshold) สำหรับแยกข้อมูลไปตามทิศทาง 2) กิ่ง (branch) เป็นคุณสมบัติของโหนดแต่ละโหนดที่แบ่งแยกออกมา จำนวนของกิ่งจะมีจำนวนเท่ากับคุณสมบัติของโหนด และ 3) ใบ (leaf) เป็นกลุ่มของผลลัพธ์ในการแยกแยะกลุ่มข้อมูลต้นไม้ตัดสินใจจะไม่แบ่งข้อมูลต่อหลังจากข้อมูลดังกล่าวถูกจัดกลุ่มให้อยู่ที่ใบ (ภาพที่ 3a)

ได้มีการพัฒนาต้นไม้ตัดสินใจให้มีประสิทธิภาพมากขึ้นโดย Breiman *et al.* (1984) และเรียกอัลกอริธึมดังกล่าวว่า ป่าสุ่มซึ่งนิยมนำมาประยุกต์ร่วมกับข้อมูลวิมิตเซนซิงเพื่อช่วยในการจำแนกการจำแนกเฉพาะเรื่อง เช่น งานด้านการจำแนกการใช้ประโยชน์ที่ดิน สิ่งปกคลุมดิน การจำแนกพันธุ์ไม้ป่าชายเลน และการหาพื้นที่ถูกเผาไหม้ (Dubeau *et al.*, 2017; Intarat & Sillaparat, 2019; Jhonnerie *et al.*, 2015; Ramo & Chuvieco, 2017; Wang *et al.*, 2018; Zhang *et al.*, 2017; Zhu *et al.*, 2017) มีข้อได้เปรียบตรงที่มีความยืดหยุ่นในการทำงานสูง ข้อมูลที่นำเข้าสู่การประมวลผลไม่จำเป็นต้องทราบฟังก์ชันการแจกแจง (distribution function) (Rodriguez-Galiano *et al.*, 2012) การทำงานของป่าสุ่มจะนำต้นไม้ตัดสินใจมารวมกัน มีลักษณะคล้ายป่า (ภาพที่ 3b) ต้นไม้ตัดสินใจแต่ละต้นจะจัดกลุ่มและจำแนกข้อมูลซึ่งป่าสุ่มจะนำผลลัพธ์ทั้งหมดที่ได้จากต้นไม้ตัดสินใจแต่ละต้นมาตรวจสอบหาค่าฐานนิยม (mode) เพื่อหาผลลัพธ์ที่มีจำนวนมากที่สุดให้เป็นผลลัพธ์สุดท้าย

ส่งผลให้ผลลัพธ์ดังกล่าวมีความถูกต้องและแม่นยำเนื่องจากได้ถูกกลั่นกรองมาจากต้นไม้ตัดสินใจจำนวนหลายต้น (Intarat & Sillapat, 2019)



ภาพที่ 3 (a) โครงสร้างการทำงานของต้นไม้ตัดสินใจและ (b) ลักษณะการประมวลผลของป่าสุ่ม ผลลัพธ์สุดท้าย จะได้จากการโหวตผลลัพธ์จากต้นไม้ตัดสินใจ (D_n) ที่มีจำนวนมากที่สุด

วิธีจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีนถูกนำเสนอเพื่อแก้ปัญหาการจำแนกข้อมูลแบบทวินาม (binary classification) ข้อมูลดังกล่าวต้องสามารถถูกแบ่งออกเป็นสองคลาสด้วยเส้นแบ่ง (hyperplane) ได้ ในการฝึกสอน ตัวข้อมูลนำเข้าจะมีจำนวนทั้งสิ้น k ตัวอย่างแทนค่าด้วย $(x_1, y_1), \dots, (x_k, y_k)$ เมื่อ $x_i \in R^N$ เป็นค่าเวกเตอร์ของข้อมูลขนาด N มิติที่ข้อมูลนำเข้าเป็นสมาชิกอยู่และคลาสของข้อมูลที่มีจำนวนสองค่าแทนด้วย $y_i \in \{-1, +1\}$ (Watanachaturaporn *et al.*, 2008) เส้นแบ่งสร้างขึ้นจากฟังก์ชันการตัดสินใจเชิงเส้น (linear decision function) ในงานด้านการจำแนกภาพถ่ายจากดาวเทียม ข้อมูลที่นำมาจำแนกส่วนใหญ่จะเป็นข้อมูลตัวแปรแบบพหุ (multivariable) ซึ่งจำเป็นต้องใช้ฟังก์ชันแบบพหุเพื่อจำแนกข้อมูลและนำเคอร์เนลประเภทไม่เชิงเส้น (non-linear kernel) เข้ามาร่วมในกระบวนการ (Aurélien, 2017) ซึ่งประเภทของเคอร์เนลที่นิยมใช้ คือ เกาเซียน ราบีเบฟ (gaussian radial basis function kernel) ในการจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีนจำเป็นต้องคำนวณเพิ่มเติมจากแบบทวิโดยใช้วิธีการจำแนกแบบจับคู่ (pairwise classification) เพื่อสร้างแบบจำลองการจำแนกแบบไบนารี (binary classifier) ในทุกข้อมูลฝึกสอนและนำไปจำแนกชุดเวกเตอร์ของข้อมูลทดสอบเพื่อโหวตหาคลาสที่จำแนกได้ถูกต้องมากที่สุด (Arora & Watanachaturaporn, 2004) โดยจะสร้างตัวจำแนกซัพพอร์ตเวกเตอร์แมชชีนของแต่ละคู่ข้อมูลได้ตัวจำแนกทั้งหมด $\frac{1}{2}M(M - 1)$ ในกรณีที่มีคลาสทั้งหมด M คลาส

สำหรับการจำแนกด้วยเคอร์เนลเกาเซียนราบีเบฟ (gaussian radial basis function kernel) ต้องปรับค่าพารามิเตอร์ในแบบจำลองให้มีความเหมาะสม (grid search) พารามิเตอร์ที่ต้องพิจารณา ได้แก่ ค่าพารามิเตอร์ปรับเคอร์



เนล (γ) และค่าพารามิเตอร์ขอบเขตแบ่งจำแนก (C) โดยพารามิเตอร์ทั้งสองจะมีช่วงค่า $C \in \{2^{-5}, 2^{-3}, \dots, 2^{13}, 2^{15}\}; \gamma \in \{2^{-15}, 2^{-13}, \dots, 2^1, 2^8\}$ หากให้ค่าพารามิเตอร์มากหรือต่ำเกินไปอาจส่งผลให้แบบจำลองเกิด overfit ได้ (Meyer et al., 2003)

ตารางที่ 1 ไฮเพอร์พารามิเตอร์ที่ใช้ทดสอบอัลกอริทึม ML และวิธีการจำแนกแบบความน่าจะเป็นสูงสุด (MLC)

Algorithms	Testing hyperparameters
Artificial neural network (ANN)	hidden layer number, max iteration, learning rate, momentum
Decision tree (DT)	max number of tree depth, min number of sample node
Random forest (RF)	max number of tree depth, min number of sample node, max number of tree
Support vector machine (SVM)	kernel parameter (γ), error control (C)
Maximum likelihood (MLC)	threshold probability

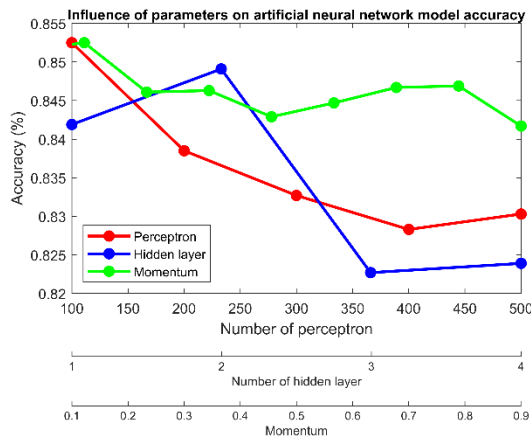
ผลลัพธ์การจำแนกทั้งหมดจะถูกนำมาวัดประสิทธิภาพโดยเริ่มจากคำนวณค่า true positive (TP) false positive (FP) true negative (TN) และ false negative (FN) จากตารางเมตริกซ์ความสับสน (confusion matrix) ข้อมูลที่ใช้วัดมาจากข้อมูลตัวอย่างสำหรับทดสอบ (testing data) ต่อมา ค่า TP TN FP และ FN ใช้คำนวณค่าความถูกต้อง (overall accuracy; OA) ค่าความแม่นยำเฉพาะในส่วนที่แบบจำลองทำนายถูก (precision หรือ positive predictive value) ค่าความแม่นยำของการทำนายในส่วนของเหตุการณ์ที่เกิดขึ้นจริง (recall หรือ sensitivity) และค่า F1-score ในส่วนของการเปรียบเทียบประสิทธิภาพการจำแนกของแต่ละวิธีจะใช้ค่าสัมประสิทธิ์โคเฮน แคปปา (Cohen's kappa; K) ซึ่งเป็นค่าสถิติที่ใช้สำหรับทดสอบโดยจะตรวจสอบความสอดคล้อง (consistency) ของข้อมูลที่ได้จากการจำแนกและข้อมูลอ้างอิง (Viera & Garrett, 2005) และทดสอบคะแนนมาตรฐาน (Z-test) ที่ระดับความเชื่อมั่นร้อยละ 95 (Intarat & Sillaparat, 2019) เพื่อเป็นการยืนยันถึงประสิทธิภาพของวิธีการจำแนกแต่ละวิธี

การเตรียมข้อมูลสำหรับการวิเคราะห์และการแสดงผลข้อมูลใช้ซอฟต์แวร์ ArcGIS Desktop เวอร์ชัน 10.8.1 (E-Contract No.: E21-0076) การทดสอบไฮเพอร์พารามิเตอร์การจำแนกการใช้ประโยชน์ที่ดิน และการทดสอบความถูกต้องใช้ซอฟต์แวร์สเปค Python เวอร์ชัน 3.9 ติดตั้ง library ได้แก่ Scikit-learn, Gdal, NumPy, Pandas และ Seaborn

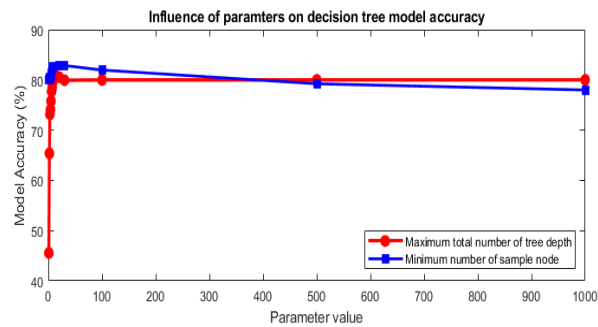
ผลการวิจัย

การทดสอบไฮเพอร์พารามิเตอร์ของแต่ละวิธีการจำแนกเพื่อหาค่าที่ทำให้การจำแนกมีประสิทธิภาพสูงที่สุด แบ่งการทดสอบออกตามแต่ละอัลกอริทึม ผลการทดสอบค่าไฮเพอร์พารามิเตอร์ของข่ายประสาทเทียม (ภาพที่ 4a) พบว่าชุดไฮเพอร์พารามิเตอร์ที่ให้ผลลัพธ์ดีที่สุด (ร้อยละ 85.25) ได้แก่ จำนวนชั้นแอบแฝง = 2 (จำนวนเพอร์เซปตรอนในแต่ละชั้น = 100) จำนวนรอบสูงสุด = 10,000 อัตราการเรียนรู้ = Adaptive โมเมนตัม = 0.2 ฟังก์ชันกระตุ้น (activation function) = rectified linear unit และฟังก์ชันต้นทุน (cost function or solver) = ADAM

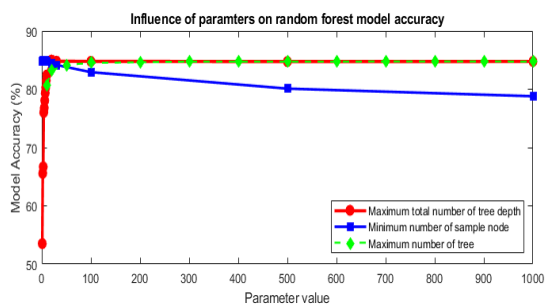
ไฮเพอร์พารามิเตอร์ที่เหมาะสมสำหรับต้นไม้ตัดสินใจพิจารณาจากค่าความลึกของต้นไม้และจำนวนตัวอย่างของใบไม้ในโนดเพื่อให้ค่าเอนโทรปีน้อยที่สุดดังแสดงในภาพที่ 4b ค่าที่เหมาะสมสำหรับการจำแนกการใช้ประโยชน์ที่ดิน ได้แก่ ค่าความลึกของต้นไม้ที่ 10 และค่าจำนวนตัวอย่างของใบไม้ต่ำสุดในโนดที่ 20 โดยประสิทธิภาพของต้นไม้ตัดสินใจอยู่ที่ร้อยละ 81.46 การทดสอบไฮเพอร์พารามิเตอร์ที่เหมาะสมสำหรับการจำแนกการใช้ประโยชน์ที่ดินด้วยอัลกอริธึมป่าสุ่มในพื้นที่ศึกษา (ภาพที่ 4c) พบว่าค่าความลึกของต้นไม้ที่ 20 ค่าจำนวนตัวอย่างของใบไม้ต่ำสุดในโนดที่ 10 และค่าจำนวนต้นไม้ที่ 600 ให้ค่าประสิทธิภาพของแบบจำลองได้สูงที่สุดที่ร้อยละ 85.00 สำหรับไฮเพอร์พารามิเตอร์ของซัพพอร์ตเวกเตอร์แมชชีน (ภาพที่ 4d) พบว่าค่าไฮเพอร์พารามิเตอร์สำหรับเคอร์เนลเกาเซียนอาร์บีเอฟ คือ $C = 30,000$ และ $\gamma = 30$ เป็น Grid search ที่ให้ประสิทธิภาพสูงที่สุดที่ร้อยละ 79.07



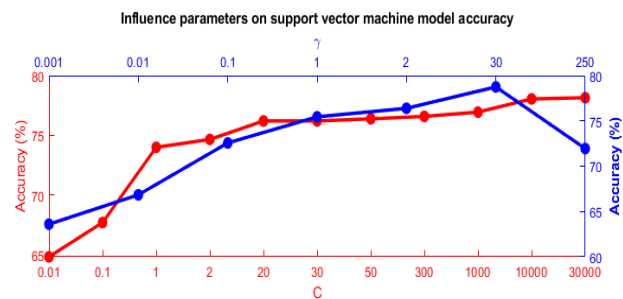
(a)



(b)



(c)



(d)

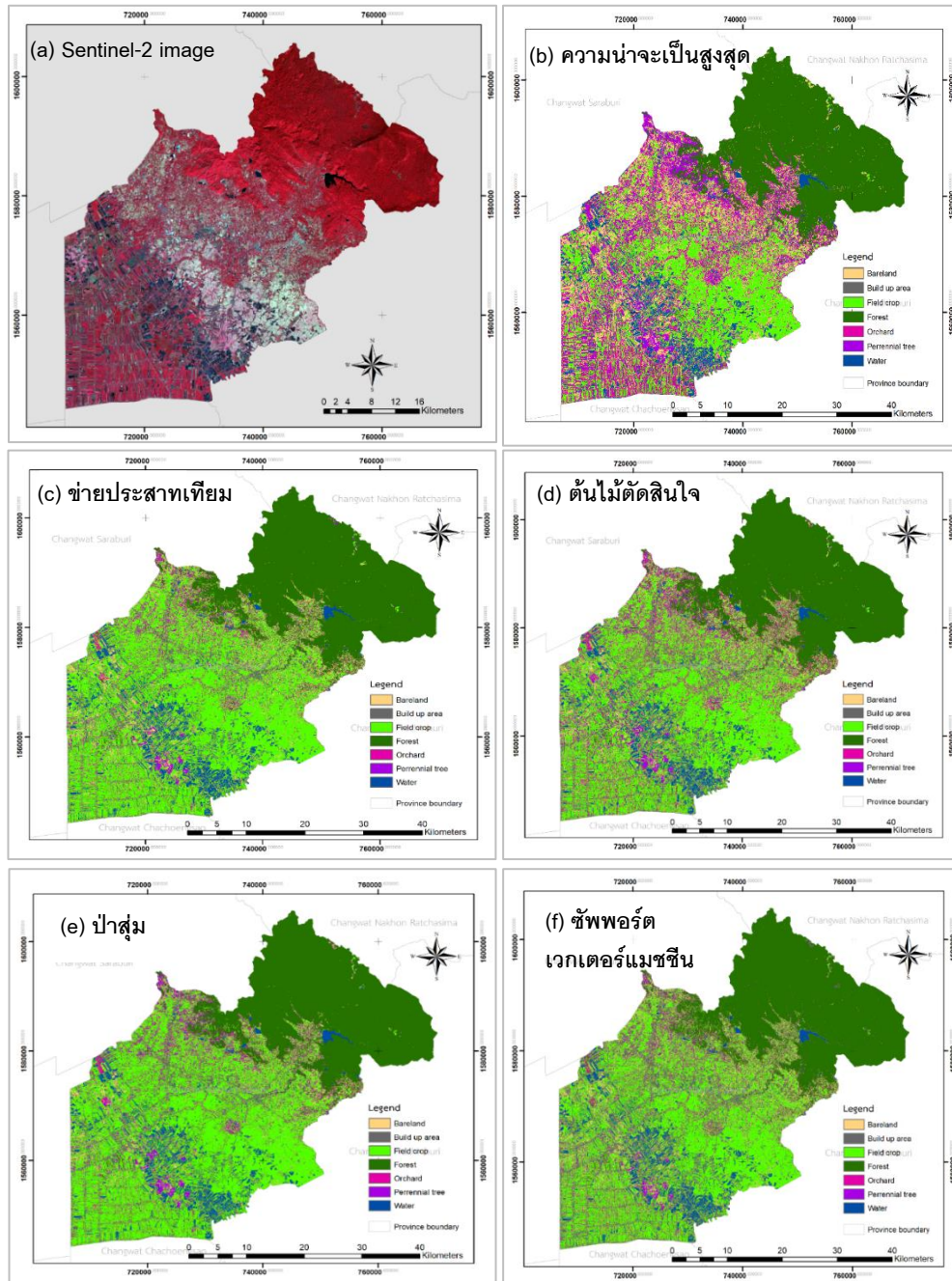
ภาพที่ 4 ผลการทดสอบไฮเพอร์พารามิเตอร์ของ (a) ข่ายประสาทเทียม (b) ต้นไม้ตัดสินใจ (c) ป่าสุ่ม และ (d) ซัพพอร์ตเวกเตอร์แมชชีน



แสดงผลการจำแนกด้วยวิธีความน่าจะเป็นสูงสุด ข่ายประสาทเทียม ต้นไม้ตัดสินใจ ป่าสุ่ม และซัพพอร์ตเวกเตอร์แมชชีน ในภาพที่ 5 ดำเนินการตรวจสอบ OA และ K ของแต่ละการจำแนกด้วยชุดข้อมูลทดสอบชุดเดียวกันทั้งหมดได้ OA ร้อยละ 63.00 69.00 84.00 92.00 และ 65.00 และค่า K เท่ากับ 0.56 0.63 0.80 0.90 และ 0.57 ตามลำดับ (ภาพที่ 6) เมื่อพิจารณาจาก OA และ K พบว่าผลการจำแนกด้วยป่าสุ่มให้ความถูกต้องสูงที่สุด (ร้อยละ 92.00) และมีความสอดคล้องของความถูกต้องจากการจำแนกข้อมูล (K = 0.90) และความถูกต้องจากการตีความหมายข้อมูลภาคพื้นดินสูงกว่าการจำแนกวิธีอื่น (Viera & Garrett, 2005) เมื่อนำผลการจำแนกด้วยป่าสุ่มมาทดสอบเปรียบเทียบกับผลการจำแนกอื่นด้วยการทดสอบ Z-test มีสมมติฐานหลัก (H_0) คือผลลัพธ์จากการจำแนกด้วยป่าสุ่มไม่แตกต่างจากผลลัพธ์ที่ได้จากวิธีอื่นและมีสมมติฐานรอง (H_1) คือผลลัพธ์ที่ได้จากการจำแนกด้วยป่าสุ่มมีความแตกต่างจากผลลัพธ์ที่ได้จากวิธีอื่น (Intarat & Sillaparat, 2019) ผลการทดสอบพบว่าวิธีการจำแนกด้วยป่าสุ่มให้ผลลัพธ์ที่แตกต่างอย่างมีนัยสำคัญที่ระดับความเชื่อมั่นร้อยละ 95 (ค่าวิกฤตที่ 1.96 ทั้งสองทาง) โดยมีค่า Z-test ที่ 10.09 8.32 3.46 และ 9.89 เมื่อทดสอบกับวิธีความน่าจะเป็นสูงสุด ข่ายประสาทเทียม ต้นไม้ตัดสินใจ และซัพพอร์ตเวกเตอร์แมชชีนตามลำดับซึ่งอยู่นอกช่วงค่าวิกฤตส่งผลให้ต้องปฏิเสธสมมติฐานหลักและยอมรับสมมติฐานรอง

ตารางที่ 3 ตารางเปรียบเทียบค่า F1-score ของคลาสที่ได้รับการจำแนกในแต่ละวิธีการจำแนก

Class	Test sample	F1-score				
		MLC	ANN	DT	RF	SVM
Br	55	0.51	0.63	0.82	0.94	0.60
Bu	30	0.84	0.82	0.89	0.92	0.72
Fe	105	0.66	0.76	0.85	0.91	0.72
Fr	31	0.74	0.67	0.81	0.93	0.64
Or	73	0.46	0.48	0.79	0.91	0.42
Pe	73	0.65	0.69	0.87	0.95	0.66
Wa	33	0.77	0.81	0.84	0.84	0.75

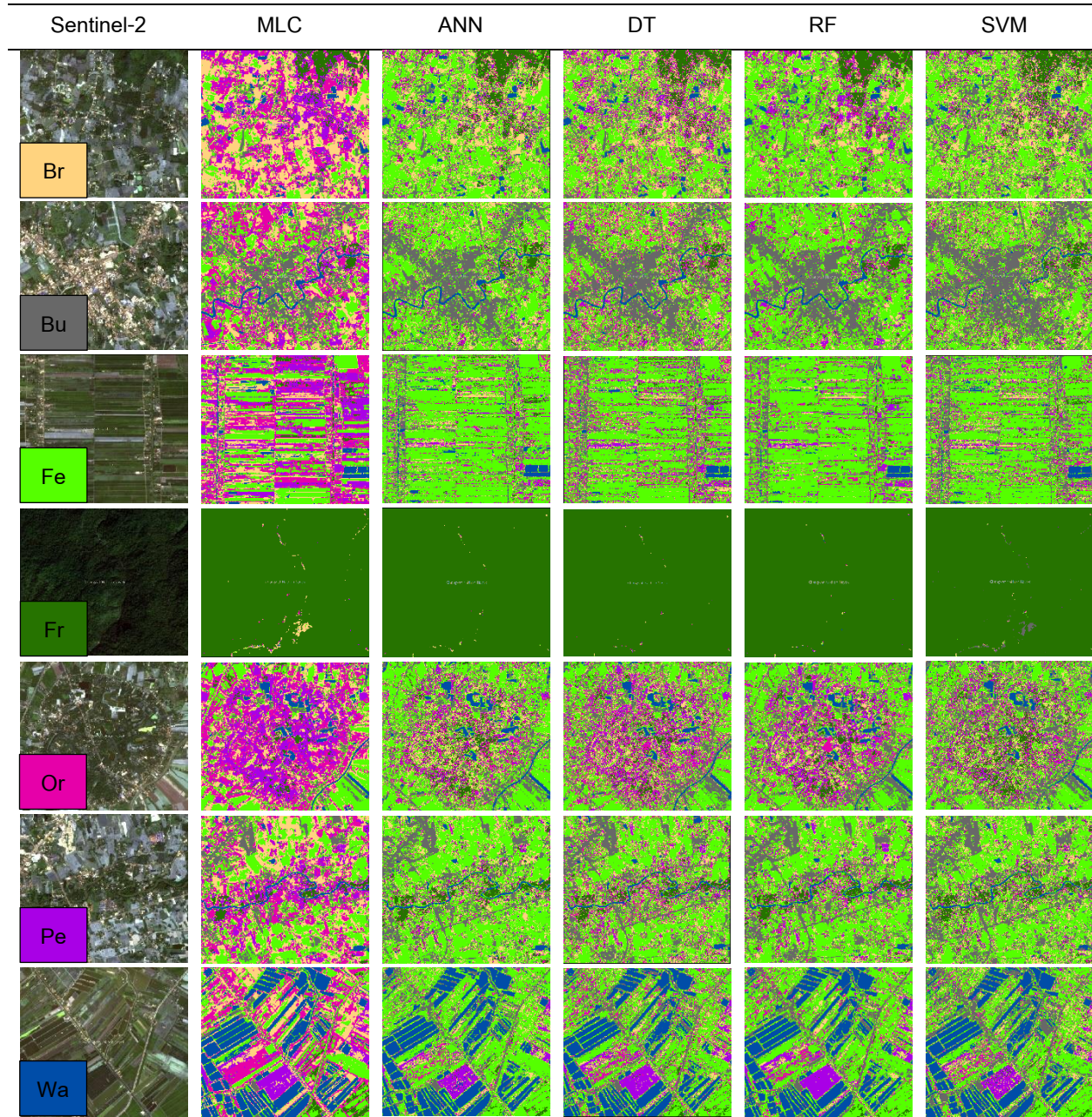


ภาพที่ 5 ผลการจำแนกการใช้ประโยชน์ที่ดิน (a) ภาพถ่ายจากดาวเทียม Sentinel-2 (b) ความน่าจะเป็นสูงสุด (c) ข่ายประสาทเทียม (d) ต้นไม้ตัดสินใจ (e) ป่าสุ่ม และ (f) ซัพพอร์ตเวกเตอร์แมชชีน



ภาพที่ 6 OA ของวิธีการจำแนกแต่ละวิธี (a) และการทดสอบความถูกต้องด้วยเมตริกซ์ความสับสนของ (b) ความน่าจะเป็นสูงสุด (c) ข่ายประสาทเทียม (d) ต้นไม้ตัดสินใจ (e) ป่าสุ่ม และ (f) ซัพพอร์ตเวกเตอร์แมชชีนด้วยค่าความแม่นยำเฉพาะในส่วนที่แบบจำลองทำนายถูก (precision)

ตารางที่ 2 เปรียบเทียบผลการจำแนกการใช้ประโยชน์ที่ดินของแต่ละวิธีการจำแนก





เมื่อพิจารณาค่าความถูกต้องการจำแนกรายคลาสจากผลการจำแนกแต่ละวิธี (ตารางที่ 2) และ F1-score (ตารางที่ 3) ปรากฏว่าวิธีการจำแนกแบบป่าสุ่มเป็นวิธีที่ให้ความถูกต้องของการจำแนกรายคลาสสูงที่สุด โดยในคลาส Br Bu Fe Fr Or และ Pe ที่ให้ค่า F1-score มากกว่า 0.90 ขึ้นไป มีเพียงคลาส Wa ที่ให้ค่า F1-score 0.84 ซึ่งต่ำกว่าคลาสอื่น แต่เมื่อเปรียบเทียบกับวิธีจำแนกอื่นแล้ว คลาส Wa ที่จำแนกด้วยวิธีการป่าสุ่มยังมีผลการจำแนกสูงกว่าวิธีการจำแนกอื่นทั้งหมด อย่างไรก็ตาม ในคลาส Bu และ Fr ผลการจำแนกด้วยวิธีการความน่าจะเป็นสูงสุดให้ค่าความถูกต้องดีกว่าวิธีข่ายประสาทเทียมและวิธีซัพพอร์ตเวกเตอร์แมชชีน และคลาส Wa ที่วิธีความน่าจะเป็นสูงสุดให้ค่าความถูกต้องการจำแนกสูงกว่าวิธีซัพพอร์ตเวกเตอร์แมชชีน

วิจารณ์ผลการวิจัย

การทดสอบประสิทธิภาพการจำแนก ML จำนวน 4 วิธี การ ได้แก่ ข่ายประสาทเทียม ต้นไม้ตัดสินใจ ป่าสุ่ม และซัพพอร์ตเวกเตอร์แมชชีนพร้อมด้วยวิธีความน่าจะเป็นสูงสุดซึ่งเป็นตัวจำแนกมาตรฐานสำหรับวิธีการจำแนกเชิงคุณภาพ (Rodriguez-Galiano *et al.*, 2012) วิธีการจำแนกแบบ ML ให้ความถูกต้องการจำแนกที่สูงกว่าวิธีความน่าจะเป็นสูงสุดโดยพิจารณาจาก OA (ภาพที่ 6a) แสดงให้เห็นถึงประสิทธิภาพของการจำแนกแบบ ML ในการประยุกต์ร่วมกับภาพถ่ายจากดาวเทียมสำหรับการจำแนกการใช้ประโยชน์ที่ดิน (Intarat & Sillapat, 2019; Abdi, 2020; Macintyre *et al.*, 2020) ข้อมูลตัวอย่างที่ใช้ได้ดำเนินการสุ่มตัวอย่างแบบเชิงลำดับขั้นที่ให้จำนวนตัวอย่างตามขนาดพื้นที่ของคลาสการใช้ประโยชน์ที่ดินที่กำหนด คลาสที่มีพื้นที่มากจะมีจำนวนตัวอย่างมากกว่าคลาสน้อยเช่นเดียวกับงานของ Intarat & Sillapat (2019) และ Vasilakos *et al.* (2020) แต่ในงานวิจัยกลับพบปัญหาความอคติ (bias) ที่เกิดขึ้นจากการเก็บตัวอย่างจำนวนไม่เท่ากัน (Abdi, 2020) และอาจส่งผลต่อความถูกต้องในการจำแนก

ภาพถ่ายจากดาวเทียม Sentinel-2 ที่ใช้ในการจำแนกมีรายละเอียดเชิงจุดภาพ 10 เมตรแบบหลายช่วงคลื่น (VNIR) ซึ่งให้ความละเอียดเชิงรังสี 16 บิตช่วยให้สามารถพิจารณาค่าการสะท้อนของคลาสน้ำที่จำแนกได้อย่างมีประสิทธิภาพ (Piedelobo *et al.*, 2019) ในงานนี้เลือกใช้ภาพถ่ายจากดาวเทียม Sentinel-2 ในช่วงเดือนธันวาคมซึ่งเป็นช่วงที่มีเมฆปกคลุมต่ำ อย่างไรก็ตามการใช้ข้อมูลภาพเพียงหนึ่งช่วงเวลาส่งผลให้เกิดความสับสนในคลาส Fe และ Br เนื่องจากลักษณะสิ่งปกคลุมดินในพื้นที่มีการเปลี่ยนแปลงไปตามช่วงเวลา (เช่น พื้นที่พืชไร่ที่มีช่วงปลูกกับช่วงเก็บเกี่ยวและพื้นที่ว่างเปล่าที่มีหญ้าปกคลุมกับพื้นที่ว่างเปล่าที่เป็นดิน) การแก้ปัญหาจำเป็นต้องเพิ่มคลาสการจำแนกตามลักษณะของพื้นที่และรวมคลาสน้ำหลังเสร็จสิ้นการจำแนกซึ่งตัวจำแนก ML บางตัว เช่นวิธีป่าสุ่มและต้นไม้ตัดสินใจสามารถแก้ปัญหาในจุดนี้ได้ (พิจารณาจากค่า F1-score ในตารางที่ 3) แต่ในการจำแนกด้วยวิธีความน่าจะเป็นสูงสุดยังมีค่าคลาดเคลื่อนเกิดขึ้นในคลาสดังกล่าวซึ่งเป็นผลมาจากการผสมค่าช่วงคลื่น (spectral mixing) ที่มีความใกล้เคียงกันมากในแต่ละคลาส (Intarat & Sillapat, 2019) นอกจากนี้การใช้ภาพถ่ายจากดาวเทียมหลายช่วงเวลาสามารถช่วยพิจารณาลักษณะของสิ่งปกคลุมดินที่เปลี่ยนแปลงไปและเพิ่มความถูกต้องให้แก่ผลการจำแนก (Mazzia *et al.*, 2020; Macintyre *et al.*, 2020; Vasilakos *et al.*, 2020)

วิธีจำแนกการใช้ประโยชน์ที่ดินด้วยตัวจำแนก ML เป็นลักษณะการจำแนกแบบไร้พารามิเตอร์จึงไม่จำเป็นต้องทราบฟังก์ชันการแจกแจง (distribution function) ของตัวอย่างข้อมูล (Rodriguez-Galiano *et al.*, 2012) ผลจากการจำแนกแสดงให้เห็นว่าวิธีป่าสุ่มให้ค่า OA (ภาพที่ 6a) และ F1-score (ตารางที่ 3) ในแต่ละคลาสได้ดีที่สุดสำหรับงานนี้เนื่องจากวิธีป่าสุ่มเป็นการพัฒนาจากวิธีต้นไม้ตัดสินใจโดยการรวมต้นไม้ตัดสินใจเอาไว้ด้วยกันจนมีลักษณะคล้ายป่า (Breiman *et al.*, 1984) จากนั้นจะหาค่าผลลัพธ์จากต้นไม้ตัดสินใจทุกต้นจากค่าฐานนิยม (mode) ซึ่งผลลัพธ์ที่หานั้นได้ผ่านการกลั่นกรองจากต้นไม้ตัดสินใจมาก่อนแล้ว วิธีป่าสุ่มยังสามารถให้ความสำคัญกับตัวแปรและทำนายค่าที่เป็นข้อมูลสูญหายได้ (Jhonnerie *et al.*, 2015; Intarat & Sillaparat, 2019) นอกจากนี้ผลการทดสอบ Z-test ยังช่วยยืนยันถึงประสิทธิภาพของวิธีป่าสุ่มที่สูงกว่าตัวจำแนก ML อื่นอย่างมีนัยสำคัญ

ตัวจำแนก ML มีการเรียนรู้เพื่อปรับแบบจำลองที่ใช้สำหรับจำแนก แต่ละตัวจำแนกจำเป็นต้องมีการทดสอบไฮเพอร์พารามิเตอร์เพื่อหาค่าที่มีความเหมาะสมกับชุดข้อมูลสอนและให้ค่าทดสอบการจำแนกของแบบจำลองมีความถูกต้องหรือมีประสิทธิภาพสูงที่สุดซึ่งเป็นข้อที่ได้เปรียบเมื่อเทียบกับวิธีความน่าจะเป็นสูงสุดซึ่งเป็นการจำแนกแบบอาศัยพารามิเตอร์ที่ต้องทราบการแจกแจงของข้อมูลตัวอย่าง (Jhonnerie *et al.*, 2015; Intarat & Sillaparat, 2019; Abdi, 2020) อย่างไรก็ตามค่า F1-score ของคลาส Bu Fr และ Wa ในตารางที่ 3 แสดงให้เห็นว่าวิธีความน่าจะเป็นสูงสุดมีค่าสูงกว่าวิธีชายประสาทเทียมและซัพพอร์ตเวกเตอร์แมชชีนเนื่องจากคลาสดังกล่าวมีการเปลี่ยนแปลงน้อยและฮิสโตแกรมของค่าการสะท้อนแสงของคลาสดังกล่าวมีลักษณะเฉพาะ (unique) ใกล้เคียงกันในแต่ละจุดภาพ เมื่อพิจารณาค่าความน่าจะเป็นของคลาสจากค่าความแปรปรวนที่ต่างกันไม่มากและสหสัมพันธ์ของข้อมูลที่มีความสัมพันธ์กันส่งผลให้คลาสดังกล่าวถูกจัดกลุ่มให้อยู่กลุ่มเดียวกัน (Lennon, 2002; Lillesand *et al.*, 2015) ค่า F1-score ที่แสดงจึงมีค่าสูง

การเพิ่มข้อมูลที่ให้สารสนเทศที่สำคัญกับการจำแนกสามารถช่วยเพิ่มให้ผลลัพธ์มีความถูกต้องสูงขึ้นได้ (Mazzia, 2020) การจำแนกการใช้ประโยชน์ที่ดินในงานนี้ใช้ชั้นข้อมูลภาพถ่ายจากดาวเทียมจำนวน 5 ชั้นข้อมูล หากมีการเพิ่มข้อมูลอื่นที่มีนัยสำคัญต่อการจำแนก เช่น การใช้ดัชนีพืชพรรณอื่นมาช่วยในการจำแนกรวมทั้งการใช้ภาพถ่ายจากดาวเทียมประเภทอื่นที่มีรายละเอียดสูงกว่า (Intarat & Sillaparat, 2019; Sun *et al.*, 2019; Imran *et al.*, 2020) เป็นการเพิ่มชั้นข้อมูลที่ช่วยให้การจำแนกมีประสิทธิภาพเพิ่มมากขึ้น

สรุปผลการวิจัย

งานวิจัยนี้แสดงถึงประสิทธิภาพของวิธีการจำแนกแบบ ML ที่ใช้สำหรับจำแนกการใช้ประโยชน์ที่ดินในจังหวัดนครนายกจากข้อมูลภาพถ่ายจากดาวเทียม Sentinel-2 ร่วมกับข้อมูลตัวอย่างจากการสุ่มและการสำรวจภาคสนาม ผลการจำแนกพบว่าวิธีการแบบ ML มีประสิทธิภาพสูงกว่าวิธีการจำแนกมาตรฐานแบบความน่าจะเป็นสูงสุดโดยมีป่าสุ่มเป็นวิธีจำแนกที่ให้ค่า OA สูงที่สุดเมื่อเปรียบเทียบกับวิธีชายประสาทเทียม ต้นไม้ตัดสินใจ และซัพพอร์ตเวกเตอร์แมชชีนที่ระดับความเชื่อมั่นร้อยละ 95 หากมีการเพิ่มข้อมูลสารสนเทศที่มีนัยสำคัญต่อการจำแนกหรือเพิ่มรายละเอียดเชิงพื้นที่ของภาพถ่ายจากดาวเทียมมีแนวโน้มที่จะช่วยเพิ่มประสิทธิภาพการจำแนกให้สูงขึ้นกว่าเดิม



กิตติกรรมประกาศ

งานวิจัยนี้ได้รับการสนับสนุนจากกองทุนวิจัยคณะศิลปศาสตร์ มหาวิทยาลัยธรรมศาสตร์ ประจำปีงบประมาณ 2564

เอกสารอ้างอิง

- 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.
doi: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. *Int. J. Eng. Trends Technol*, 38(1), 5-11.
- 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* (pp. 237-255): Springer.
- Aurélien, G. (2017). Hands-on machine learning with scikit-learn & tensorflow. *Geron Aurelien*.
- Congalton, R. G., & Green, K. (2019). *Assessing the accuracy of remotely sensed data: principles and practices*: CRC press.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*: CRC press.
- 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), 1056.
- Ge, G., Shi, Z., Zhu, Y., Yang, X., & Hao, Y. (2020). Land use/cover classification in an arid desert-oasis mosaic landscape of China using remote sensed imagery: Performance assessment of four machine learning algorithms. *Global Ecology and Conservation*, 22, e00971. doi:https://doi.org/10.1016/j.gecco.2020.e00971



- Imran, A. B., Khan, K., Ali, N., Ahmad, N., Ali, A., & Shah, K. (2020). Narrow band based and broadband derived vegetation indices using Sentinel-2 Imagery to estimate vegetation biomass. *Global Journal of Environmental Science and Management-Gjesm*, 6(1), 97-108. doi:10.22034/gjesm.2020.01.08
- 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)
- 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
- Lennon, R. (2002). Remote sensing digital image analysis: An introduction. *United States: Esa/Esrin*.
- Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). Remote sensing and image interpretation: *John Wiley & Sons*.
- Majnouni-Toutakhane, A. (2020). Modeling the Land Use Change Process on the South Coast of the Caspian Sea Using Logistic Regression and Artificial Neural Network. *Journal of Environmental Accounting and Management*, 8(2), 111-123. doi:10.5890/jeam.2020.06.001
- 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 (R-CNN). *Applied Sciences-Basel*, 10(1). doi:10.3390/app10010238
- 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. doi:10.1016/j.jag.2019.101980
- Meyer, D., Leisch, F., & Hornik, K. (2003). The support vector machine under test. *Neurocomputing*, 55(1-2), 169-186.



de Oliveira Duarte, D. C., Zanetti, J., Junior, J. G., & das Graças Medeiros, N. (2018). Comparison of supervised classification methods of maximum likelihood, minimum distance, parallelepiped and neural network in images of unmanned air vehicle (UAV) in Viçosa-MG. *Revista Brasileira de Cartografia*, 70(2), 437-452.

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: Case study of the Duero river basin. *Agricultural Systems*, 171, 36-50.
doi:10.1016/j.agry.2019.01.005

Ramo, R., & Chuvieco, E. (2017). Developing a random forest algorithm for MODIS global burned area classification. *Remote Sensing*, 9(11), 1193.

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.

Saboori, M., Torahi, A. A., & Bakhtyari, H. R. R. (2019). Combining multi-scale textural features from the panchromatic bands of high spatial resolution images with ANN and MLC classification algorithms to extract urban land uses. *International Journal of Remote Sensing*, 40(22), 8608-8634.
doi:10.1080/01431161.2019.1620371

Shrestha, A., & Mahmood, A. (2019). Review of Deep Learning Algorithms and Architectures. *Ieee Access*, 7, 53040-53065. doi:10.1109/access.2019.2912200

Silva, L. P. e., Xavier, A. P. C., da Silva, R. M., & Santos, C. A. G. (2020). Modeling land cover change based on an artificial neural network for a semiarid river basin in northeastern Brazil. *Global Ecology and Conservation*, 21, e00811. doi:https://doi.org/10.1016/j.gecco.2019.e00811

Sun, C. L., Bian, Y., Zhou, T., & Pan, J. L. (2019). Using of Multi-Source and Multi-Temporal Remote Sensing Data Improves Crop-Type Mapping in the Subtropical Agriculture Region. *Sensors*, 19(10).
doi:10.3390/s19102401



- Vasilakos, C., Kavroudakis, D., & Georganta, A. (2020). Machine Learning Classification Ensemble of Multitemporal Sentinel-2 Images: The Case of a Mixed Mediterranean Ecosystem. *Remote Sensing*, 12(12). doi:10.3390/rs12122005
- Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam med*, 37(5), 360-363.
- Wang, D., Wan, B., Qiu, P., Su, Y., Guo, Q., & Wu, X. (2018). Artificial mangrove species mapping using pléiades-1: An evaluation of pixel-based and object-based classifications with selected machine learning algorithms. *Remote Sensing*, 10(2), 294.
- Watanachaturaporn, P., Arora, M. K., & Varshney, P. K. (2008). Multisource classification using support vector machines: An empirical comparison with decision tree and neural network classifiers. *Photogrammetric Engineering and Remote Sensing*, 74(2), 239-246. doi:10.14358/pers.74.2.239
- Zhang, K. X., Wu, X. L., Niu, R. Q., Yang, K., & Zhao, L. R. (2017). The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China. *Environmental Earth Sciences*, 76(11). doi:10.1007/s12665-017-6731-5
- Zhu, Y., Liu, K., Liu, L., Myint, S. W., Wang, S., Liu, H., & 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), 1060.