การศึกษาการจำแนกข้อมูลภาพด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน กรณีศึกษา พื้นที่เพาะปลูกข้าว อำเภอจุน จังหวัดพะเยา

A Study on Support Vector Machine Image Classification: A Case Study on Rice Plantation Area in Chun District, Phayao Province

อานนท์ เบียงแล[1] และ สวรินทร์ ฤกษ์อยู่สุข[2*]

Anon Bianglae[1] and Sawarin Lerk-u-suke[2*]

[1] บริษัท ไอบิตซ์ จำกัด อำเภอเมือง จังหวัดขอนแก่น 65000

[2] หลักสูตรภูมิสารสนเทศศาสตร์ คณะเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยพะเยา อำเภอเมือง จังหวัดพะเยา 56000

[*]Corresponding author: sawarin.le@up.ac.th

Received: November 21, 2020 Revised: December 18, 2020 Accepted: December 25, 2020

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ เพื่อศึกษาประสิทธิภาพของเทคนิคการจำแนกข้อมูลภาพด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine – SVM) สำหรับการจำแนกพื้นที่ปลูกข้าวในเขตอำเภอจุน จังหวัดพะเยา ด้วยข้อมูลภาพดาวเทียมแลนแซต 8 ร่วมกับชุดข้อมูลจุดตัวอย่างและจุดตรวจสอบสำหรับการประเมินความถูกต้องนั้นพัฒนาจากชุดข้อมูลการใช้ประโยชน์ที่ดินของกรมพัฒนาที่ดินและการสำรวจภาคสนาม จำนวนจุดตัวอย่างประกอบด้วย 6 ระดับ ได้แก่ 25, 50, 75, 100, 125 และ 150 จุด และจุดตรวจสอบจำนวนทั้งสิ้น 406 จุด การประเมินความถูกต้องของการจำแนกจะอาศัยการพิจารณาค่าความถูกต้องโดยรวมของการจำแนก (Overall Accuracy) และค่าสัมประสิทธิแคปปา (Kappa Coefficient) โดยนำผลที่ได้จากการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนเปรียบเทียบกับการจำแนกข้อมูลภาพด้วยวิธีความน่าจะเป็นสูงสุด (Maximum Likelihood Classifier – MLC) ที่ได้รับความนิยมและมีการประยุกต์ใช้งานอย่างแพร่หลายในการจำแนกข้อมูลภาพ ผลจากการศึกษาพบว่า การจำแนกข้อมูลโดยอาศัยวิธีซัพพอร์ตเวกเตอร์แมชชีนโดยอาศัยจำนวนจุดตัวอย่างที่แตกต่างกัน ผลการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องเท่ากับ 89.66%, 89.90%, 90.89, 91.63%, 93.60% และ 94.33% ตามลำดับ ส่วนวิธีการจำแนกข้อมูลภาพด้วยวิธีความน่าจะเป็นสูงสุดให้ค่าความถูกต้องโดยรวมของการจำแนกเท่ากับ 83.25%, 83.50%, 86.21%, 87.19%, 85.96% และ 86.95% ตามลำดับ ค่าสัมประสิทธิ์แคปปาที่ได้จากวิธีซัพพอร์ตเวกเตอร์แมชชีนเท่ากับ 0.79, 0.80, 0.82, 0.83, 0.87, และ 0.89 ส่วนการจำแนกข้อมูลภาพด้วยวิธีความน่าจะเป็นสูงสุดให้ค่าสัมประสิทธิ์แคปปาเท่ากับ 0.67, 0.67, 0.72, 0.74, 0.72, และ 0.74 ตามลำดับ จากผลการจำแนกดังกล่าวพบว่า วิธีซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องโดยรวมของการจำแนกและค่าสัมประสิทธิ์แคปปาสูงกว่าวิธีความน่าจะเป็นสูงสุด เมื่อแยกระดับจำนวนจำนวนจุดตัวอย่างที่เลือกใช้งาน นอกจากนี้ หากลดจำนวนจุดตัวอย่างลงก็ยังให้ค่าความถูกต้องโดยรวมและค่าสัมประสิทธิ์แคปปาที่สูงกว่าอย่างชัดเจน ทำให้สรุปได้ว่าวิธีซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพมากกว่าวิธีความน่าจะเป็นสูงสุด

คำสำคัญ: ซัพพอร์ตเวกเตอร์แมชชีน / ความน่าจะเป็นสูงสุด / การจำแนกข้อมูลภาพ / พื้นที่ปลูกข้าว

Abstract

This research aims to study the performance of Support Vector Machine (SVM) using Landsat 8 satellite imagery to identify the paddy field in Chun District, Phayao Province. Training point and the checkpoint for classification accuracy assessment were developed from the Land Use Data Set from the Land Development Department and field surveys. Training point consists of 6 levels including 25, 50, 75, 100, 125 and 150 points. Supervised classification using SVM and Maximum Likelihood Classification (MLC) was compared by considering the statistical values; overall accuracy and Kappa coefficient. The overall accuracy values from SVM are reported at a high level of 89.66%, 89.90%, 90.89, 91.63%, 93.60%, and 94.33% respectively. And The overall accuracy values are reported at 83.25%, 83.50%, 86.21%, 87.19%, 85.96% and 86.95% respectively when MLC are applied. The Kappa coefficient values from SVM are also reported at a high level of 0.79, 0.80, 0.82, 0.83, 0.87, and 0.89, respectively. And the kappa are reported at 0.67, 0.67, 0.72, 0.74, 0.72, and 0.74 respectively when MLC are in use. According the results, it was found that SVM is more effective than MLC to identify the paddy field of the study area. It is hope that the method proposed in the study can be applied to other similar image classification.

Keyword: Support Vector Machine / Maximum Likelihood / Image classification / Paddy field

1. บทนำ

ข้าวเป็นอาหารหลักของไทยตั้งแต่อดีตและยังเป็นอาหารหลักของประชากรโลกมากกว่าสามพันล้านคน และข้าวยังเป็นพืชเศรษฐกิจอันดับหนึ่งของประเทศไทยที่ทำรายได้ให้กับประเทศอย่างมหาศาล ข้าวหอมมะลิไทย (Thai Hom Mali) ยังขึ้นชื่อว่าเป็นข้าวที่ดีที่สุดในโลกจากการประกวดแข่งขันข้าวที่ดีที่สุดในโลก (The World's Best Rice Contest) ในปีค.ศ. 2016 และ 2017 หรือเมื่อปีพ.ศ. 2559 และ 2560 (The Rice Trader, 2020) บอกได้ถึงคุณภาพที่ดีของคุณภาพข้าวไทย ดังนั้นการบริหารจัดการพื้นที่เพาะปลูกข้าวนั้นจึงมีความสำคัญและจำเป็นอย่างยิ่ง

เทคโนโลยีทางด้านการสำรวจระยะไกลเป็นแนวทางหนึ่งที่สามารถนำมาช่วยในด้านการบริหารจัดการพื้นที่เพาะปลูกเพื่อการเกษตรได้อย่างมีประสิทธิภาพ เช่น ตรวจวัดการเปลี่ยนแปลงของพื้นที่เพาะปลูก การวางแผนการใช้ที่ดิน การหาพื้นที่เหมาะสมเพื่อส่งเสริมการเพาะปลูกหรือการคาดการผลผลิตล่วงหน้าจากขนาดพื้นที่เพาะปลูก วิธีการจำแนกข้อมูลภาพเพื่อใช้ในการค้นหาและจำแนกพื้นที่เพาะปลูกพืชในปัจจุบันนั้น มีวิธีการที่นิยมใช้การอย่างแพร่หลาย ได้แก่ การจำแนกข้อมูลภาพแบบระยะห่างต่ำสุด (Minimum distance to means – MD) การจำแนกข้อมูลภาพแบบสี่เหลี่ยมคู่ขนาน (Parallelepiped classification or Box classifier – PP) และการจำแนกข้อมูลภาพแบบความน่าจะเป็นสูงที่สุด (Maximum Likelihood Classification – MLC) ซึ่งการจำแนกข้อมูลภาพแบบความน่าจะเป็นสูงที่สุด เป็นวิธีที่นิยมใช้สำหรับการจำแนกข้อมูลภาพอย่างกว้างขวาง (Lillesand & Kiefer, 2000) (สุวิทย์ อ๋องสมหวัง, 2559) (Jensen, 2005) เพราะให้ค่าความถูกต้องของการจำแนกข้อมูลภาพมากที่สุดเมื่อเปรียบเทียบกับวิธีอื่น ๆ ข้างต้น (สำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ, 2552) อย่างไรก็ดี วิธีการจำแนกข้อมูลภาพแบบความน่าจะเป็นสูงที่สุดยังมีข้อจำกัดหลายประการ เช่น จำเป็นต้องใช้พื้นที่ตัวอย่าง (Training area) จำนวนมากเพื่อให้เกิดความถูกต้องในการจำแนกข้อมูลภาพและการกระจายตัวของข้อมูลต้องอยู่บนสมมติฐานของการกระจายตัวแบบปกติ ในขณะวิธีการจำแนกข้อมูลสมัยใหม่อย่างการจำแนกแบบซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine – SVM) ที่ให้ค่าความถูกต้องของการจำแนกข้อมูลภาพมีความถูกต้องแม่นยำสูงเช่นเดียวกับการจำแนกภาพแบบความ

น่าจะเป็นสูงสุด (MLC) (Mountrakis, Im & Ogole, 2011) โดยต้องการจำนวนพื้นที่ตัวอย่างไม่มากนักซึ่งจะช่วยประหยัดเวลาในการทำงาน ลดทรัพยากรในการประมวลผลข้อมูล และไม่จำเป็นต้องอาศัยพื้นฐานของการแจกแจงแบบปกติของข้อมูล (Miller, 2017) ดังนั้นผู้วิจัยจึงมีความสนใจที่จะศึกษาการจำแนกภาพแบบซัพพอร์ตเวกเตอร์แมชชีน (SVM) สำหรับการจำแนกข้อมูลภาพเพื่อสกัดข้อมูลพื้นที่เพาะปลูกข้าว

การดำเนินงานวิจัยครั้งนี้เป็นการศึกษาวิธีการจำแนกข้อมูลภาพด้วยวิธีซัพพอร์ตเว็กเตอร์แมชชีนในการจำแนกพื้นที่ปลูกข้าว และเปรียบเทียบความถูกต้องในการจำแนกพื้นที่ปลูกข้าวระหว่างวิธีซัพพอร์ตเว็กเตอร์แมชชีนกับวิธีการจำแนกข้อมูลภาพแบบความน่าจะเป็นสูงสุดบนชุดข้อมูลภาพดาวเทียมแลนแซต 8 (LANDSAT 8 : OLI) ในพื้นที่อำเภอจุน จังหวัดพะเยา

2. วัตถุประสงค์

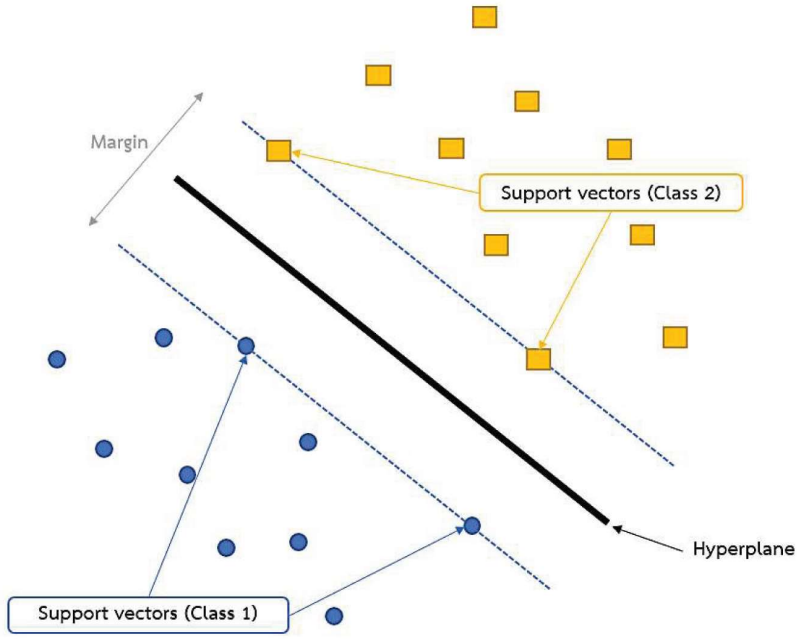
งานวิจัยครั้งนี้มุ่งเน้นศึกษาประสิทธิภาพของเทคนิคการจำแนกข้อมูลภาพด้วยวิธีซัพพอร์ตเว็กเตอร์แมชชีน (SVM) โดยการเปรียบเทียบกับวิธีการจำแนกข้อมูลภาพแบบความน่าจะเป็นสูงสุด (MLC) เพื่อสกัดข้อมูลพื้นที่เพาะปลูกข้าวจากข้อมูลภาพดาวเทียมแลนด์แซต 8 การเปรียบเทียบนั้นจะทำการประเมินความถูกต้องผลการจำแนกข้อมูลภาพ ที่มีการปรับใช้จำนวนจุดตัวอย่างที่แตกต่างกันในการจำแนกข้อมูลภาพด้วยเทคนิคดังกล่าว ผลการศึกษาจะเป็นแนวทางที่สำคัญในการประยุกต์ใช้งานเทคโนโลยีการรับรู้จากระยะไกลด้านการเกษตรและการจำแนกข้อมูลภาพสำหรับงานประยุกต์ที่เกี่ยวข้อง

3. วิธีการศึกษา

งานวิจัยครั้งนี้ เป็นการศึกษาประสิทธิภาพของเทคนิคการจำแนกข้อมูลภาพด้วยวิธีซัพพอร์ตเว็กเตอร์แมชชีน (SVM) ซึ่งเป็นวิธีการหนึ่งของการเรียนรู้ของเครื่องจักร (Machine Learning – ML) แบบมีผู้สอน (Supervised learning) (Goodfellow, Bengio & Courville, 2016) โดยการเปรียบเทียบกับวิธีความน่าจะเป็นสูงสุด (MLC) ที่เป็นวิธีการจำแนกข้อมูลภาพที่ได้รับความนิยมอย่างยาวนานแพร่หลาย (Jensen, 2005) ทั้งสองวิธีนี้ต่างก็เป็นเทคนิคที่นิยมใช้ในการจำแนกข้อมูลและอยู่ในกลุ่มของการจำแนกข้อมูลแบบผู้สอน (Supervised Classification) โดยการจำแนกข้อมูลภาพทั้งสองวิธีรวมถึงการประเมินความต้องของการจำแนกข้อมูลจะดำเนินการโดยอาศัยโปรแกรมด้านการประมวลผลภาพเชิงเลข (Digital Image Processing – DIP) ที่มีความน่าเชื่อถือและได้รับความนิยมอย่างแพร่หลายในงานการรับรู้จากระยะไกล (Remote Sensing – RS)

การจำแนกข้อมูลภาพด้วยวิธีซัพพอร์ตเว็กเตอร์แมชชีน (SVM) นั้นอยู่บนพื้นฐานของวิธีทางสถิติแบบไม่อาศัยพารามิเตอร์ (Non–parametric statistics) ที่ไม่ทราบคุณสมบัติหรือลักษณะของประชากร (Population) หรือที่เรียกว่าพารามิเตอร์ (Parameter) จึงไม่ตั้งอยู่บนสมมุติฐานของการกระจายตัวของประชากรหรือข้อมูลที่นำมาทำการจำแนก การจำแนกข้อมูลทางสถิติ (Statistical data classification) นั้นจะเป็นการแบ่ง (Division) หรือการคัดแยกข้อมูลจากกลุ่มตัวอย่างประชากรให้เป็นหมวดหมู่ที่มีความหมายหรือมีคุณลักษณะที่เหมือนกัน (บางครั้งเรียกว่า กลุ่มประชากรย่อย หรือ Sub–populations) เพื่อใช้ในการวิเคราะห์ข้อมูลต่อไป (Miller, 2017) โดยทั่วไปแล้วข้อมูลประชากรตัวอย่างที่นำมาจัดกลุ่มหรือคัดแยกข้อมูลนั้น ผู้คัดแยกจะไม่ทราบความแตกต่างของแต่ละกลุ่ม (Class/Category) อย่างชัดเจน การคัดแยกจะอาศัยการพิจารณาความเหมือนกัน (Similarity) ของคุณลักษณะบางประการของประชากรหรือข้อมูล โดยจะทำการจัดกลุ่มข้อมูลที่มีคล้ายคลึงกันให้อยู่กลุ่มเดียวกัน และคัดแยกข้อมูลที่แตกต่างกันแยกออกไปอยู่กันคนละกลุ่ม วิธีซัพพอร์ตเว็กเตอร์แมชชีน (SVM) จะประยุกต์หลักการจำแนกประเภทข้อมูลโดยอาศัยการจดจำข้อมูลที่อยู่

บริเวณขอบของแต่ละคลาส ที่เรียกว่า เวกเตอร์ซัพพอร์ต (Vector support) ในการช่วยจดจำและจำแนกข้อมูล วิธีการนี้มีประสิทธิภาพในการจำแนกข้อมูลได้ดีและสามารถรองรับข้อมูลที่มีปริภูมิ (Space) ได้ทั้งแบบเชิงเส้น (Linear) และแบบไม่เป็นเชิงเส้น (Non–linear) การแบ่งคลาสหรือการจำแนกข้อมูล (Classification) นั้นจะอาศัยระนาบเกิน (Hyperplane) ในการแบ่งแยกข้อมูล โดยระนาบเกินนี้จะพิจารณาจากระยะขอบ (Margin) ของข้อมูลแต่ละกลุ่มในการตัดสินใจกำหนดเลือกระนาบเกิน เวกเตอร์ข้อมูลที่อยู่บริเวณขอบของแต่ละคลาสจะเรียกว่า เวกเตอร์ซัพพอร์ต (Support vector) ลักษณะของเวกเตอร์ซัพพอร์ต (ปริญญา สงวนสัตย์, 2562) ระยะขอบและระนาบเกินสามารถแสดงได้ดังรูปที่ 1 ดังนี้



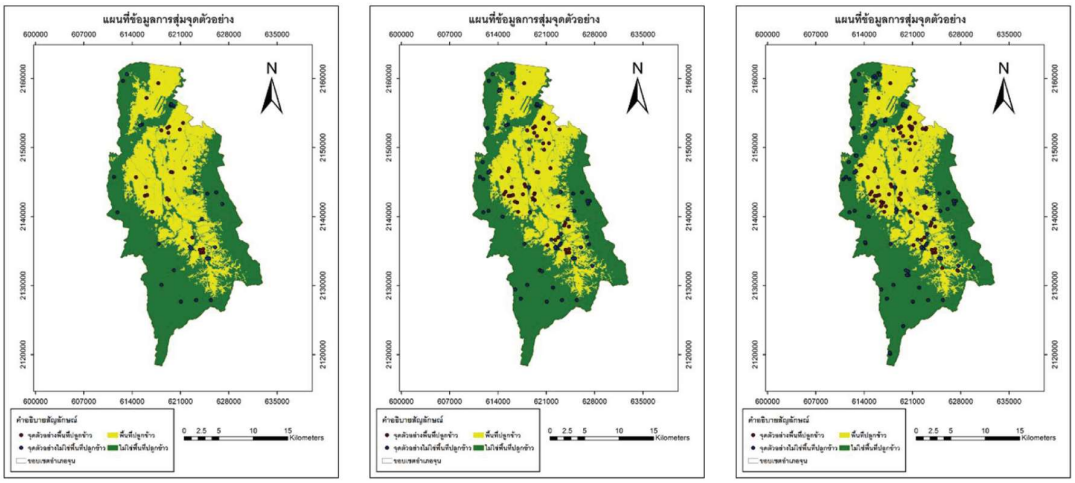
รูปที่ 1 เวกเตอร์ซัพพอร์ต ระยะขอบและระนาบเกินของวิธีซัพพอร์ตเว็กเตอร์แมชชีน (SVM)
ปรับปรุงจาก Artificial intelligence with Machine learning – AI สร้างได้ด้วยแมชชีนเลิร์นนิง (Python edition), ปริญญา สงวนสัตย์ (2562)

พื้นที่ศึกษาในครั้งนี้ คือ อำเภอจุน จังหวัดพะเยา มีขนาดพื้นที่ประมาณ 570 ตารางกิโลเมตร โดยอำเภอจุนเป็นแหล่งเพาะปลูกข้าวหอมมะลิอินทรีย์ที่มีชื่อเสียงของจังหวัด ตามสภาพภูมิประเทศแล้วอำเภอจุนจัดอยู่ในเขตลุ่มน้ำโขงเช่นเดียวกับอำเภอเมืองพะเยา อำเภอแม่ใจ อำเภอดอกคำใต้และบางส่วนของอำเภอปง พื้นที่เพาะปลูกข้าวส่วนมากจะเน้นการเพาะปลูกข้าวหอมมะลิแบบอินทรีย์ที่ให้ผลตอบแทนสูงกว่าข้าวชนิดอื่น นอกจากนี้เกษตรกรของอำเภอจุนยังได้รับการรับรองระบบจากโครงการพัฒนาคุณภาพมาตรฐานสินค้าเกษตร (Good Agricultural Practices – GAP) จำนวนมากที่สุดของจังหวัดพะเยา (สำนักงานปลัดกระทรวงเกษตรและสหกรณ์, 2563)

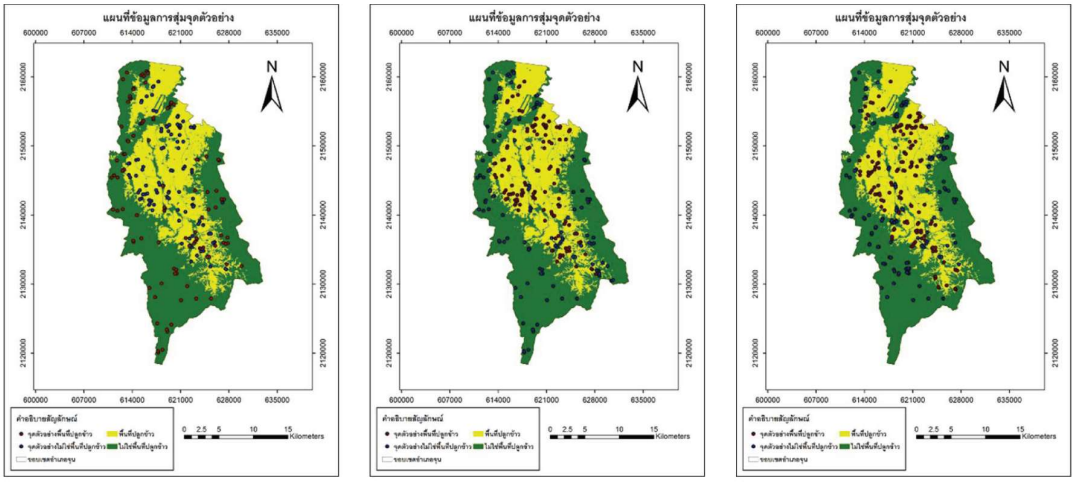
ข้อมูลภาพที่ใช้ในการศึกษาครั้งนี้ คือ ข้อมูลภาพถ่ายจากดาวเทียมแลนด์แซท 8 (LANDSAT 8) ที่บันทึกภาพในช่วงปีพ.ศ. 2557 (ค.ศ. 2014) จากเครื่องรับรู้ (Sensor) เครื่องสร้างภาพพื้นดินดำเนินการ (Operational Land Imager – OLI) ที่ได้รับการพัฒนาร่วมกันระหว่างองค์การบริหารการบินและอวกาศแห่งชาติ (National Aeronautics

and Space Administration – NASA) และสำนักงานสำรวจธรณีวิทยาแห่งสหรัฐอเมริกา (United States Geological Survey– USGS) โดยการดาวน์โหลดข้อมูลภาพจาก https://earthexplorer.usgs.gov ข้อมูลที่ใช้เป็นข้อมูลที่มีการปรับแก้ความเคลื่อนเชิงเรขาคณิต (Geometric correction) และความคลาดเคลื่อนเชิงรังสี (Radiometric correction) จากระบบข้อมูลที่ให้บริการ

จุดตัวอย่าง (Training Point – TP) ที่ใช้ในการสอนให้คอมพิวเตอร์เรียนรู้สำหรับการจำแนกข้อมูลภาพนั้นจะทำการคัดเลือกแบบสุ่ม (Random) มาจากชุดข้อมูลการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดินของจังหวัดพะเยา ปี 2558 จากกรมพัฒนาที่ดิน ซึ่งเป็นช่วงเวลาใกล้เคียงกับชุดข้อมูลภาพถ่ายดาวเทียมและการลงพื้นที่สำรวจของกรมพัฒนาที่ดินในการปรับปรุงและจำแนกพื้นที่การใช้ประโยชน์ที่ดินชุดดังกล่าว สำหรับจุดตัวอย่างของพื้นที่เพาะปลูกข้าวนี้ ผู้วิจัยได้ทำการแบ่งจำนวนจุดตัวอย่างสำหรับการจำแนกข้อมูลออกเป็น 6 ระดับ ได้แก่ จุดตัวอย่างพื้นที่เพาะปลูกข้าว จำนวน 25, 50, 75, 100, 125 และ 150 จุด ตามลำดับ ในส่วนของจุดตรวจสอบ (Check Point – CP) ที่ใช้สำหรับการประเมินความถูกต้องของการจำแนกข้อมูลภาพจากทั้งสองวิธีนั้นจะอาศัยชุดข้อมูลการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดินของจังหวัดพะเยา ปี 2558 ของกรมพัฒนาที่ดิน ร่วมกับการสำรวจภาคสนามเช่นเดียวกับการคัดเลือกจุดตัวอย่าง (TP) จำนวนจุดตรวจสอบ (CP) ที่ใช้ในการศึกษาครั้งนี้ ผู้วิจัยได้ทำการคัดเลือกมาจำนวนทั้งสิ้น 406 จุด ซึ่งมากกว่าความต้องการขั้นต่ำของการประเมินความถูกต้องจากการคำนวณขนาดตัวอย่างบนพื้นฐานทฤษฎีความน่าจะเป็นของการแจกแจงแบบทวินาม (Binomial distribution probability theory) ที่ต้องการความถูกต้องโดยรวมของการประเมินความถูกต้องการจำแนกข้อมูลระดับที่ดีกว่าร้อยละ 85 และค่าความผิดพลาดที่ยอมให้เกิดขึ้นได้ร้อยละ 5 (204 จุด) โดยจุดตรวจสอบจะอยู่บนพื้นที่เพาะปลูกข้าวและไม่ใช่พื้นที่เพาะปลูกข้าวจำนวนพื้นที่ละ 203 จุด ตำแหน่งของจุดตัวอย่าง (TP) และจุดตรวจสอบ (CP) สามารถแสดงได้ดังรูปที่ 3 และ 4 ดังนี้

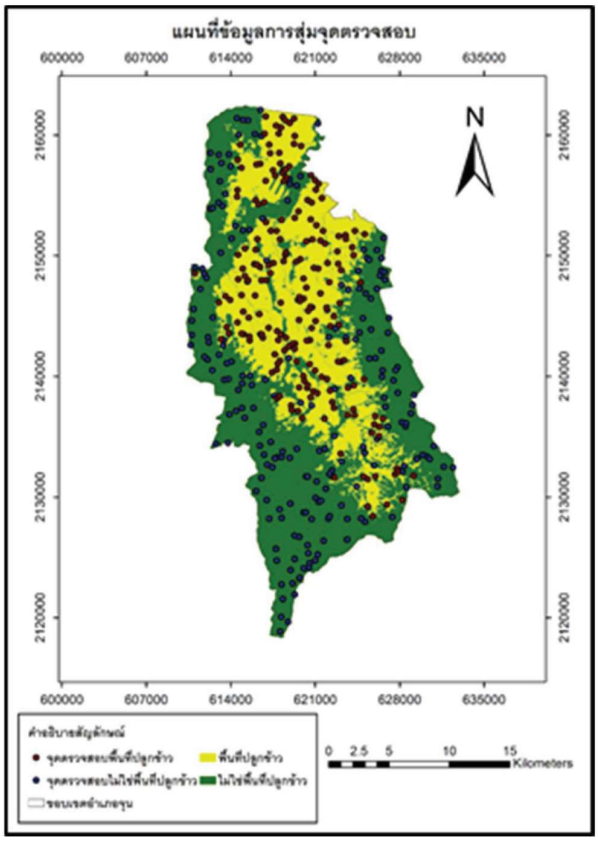


(ก) (ข) (ค)



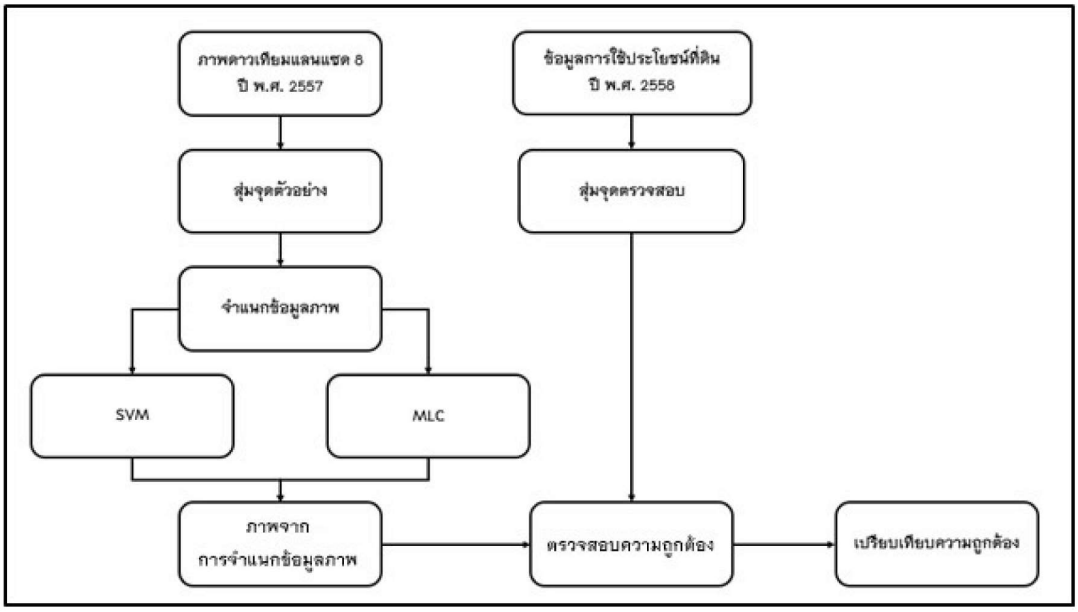
(ง) (จ) (ฉ)

รูปที่ 2 แสดงจุดตัวอย่าง (TP) จำนวน (ก) 25 จุด (ข) 50 จุด (ค) 75 จุด (ง) 100 จุด (จ) 125 จุดและ (ฉ) 150 จุด



รูปที่ 3 แสดงจุดตรวจสอบ (CP) ที่ใช้สำหรับการประเมินความถูกต้องของการจำแนกข้อมูลภาพ

การประเมินความถูกต้องของการจำแนก (Classification Accuracy Assessment) นั้นจะอาศัยวิธีการทางสถิติโดยอาศัยตารางเมทริกซ์ความคลาดเคลื่อน (Error matrix) โดยอาศัยการคำนวณทางสถิติในการประเมินความแม่นยำหรือความถูกต้องของการจำแนกจากการพิจารณาค่าความถูกต้องโดยรวม (Overall Accuracy – OA) และค่าสัมประสิทธิ์แคปปา (Kappa Coefficient – Kappa) ในการตรวจสอบความถูกต้องของการจำแนกข้อมูลภาพที่ได้จากวิธีซัพพอร์ตเว็กเตอร์แมชชีน (SVM) และวิธีความน่าจะเป็นสูงสุด (MLC) ผลลัพธ์ที่ได้จะสะท้อนถึงความถูกต้องและประสิทธิภาพของวิธีการจำแนกข้อมูลภาพของทั้งสองวิธีการ ขั้นตอนดำเนินงานวิจัย สามารถสรุปได้ดังรูปที่ 4 ดังนี้



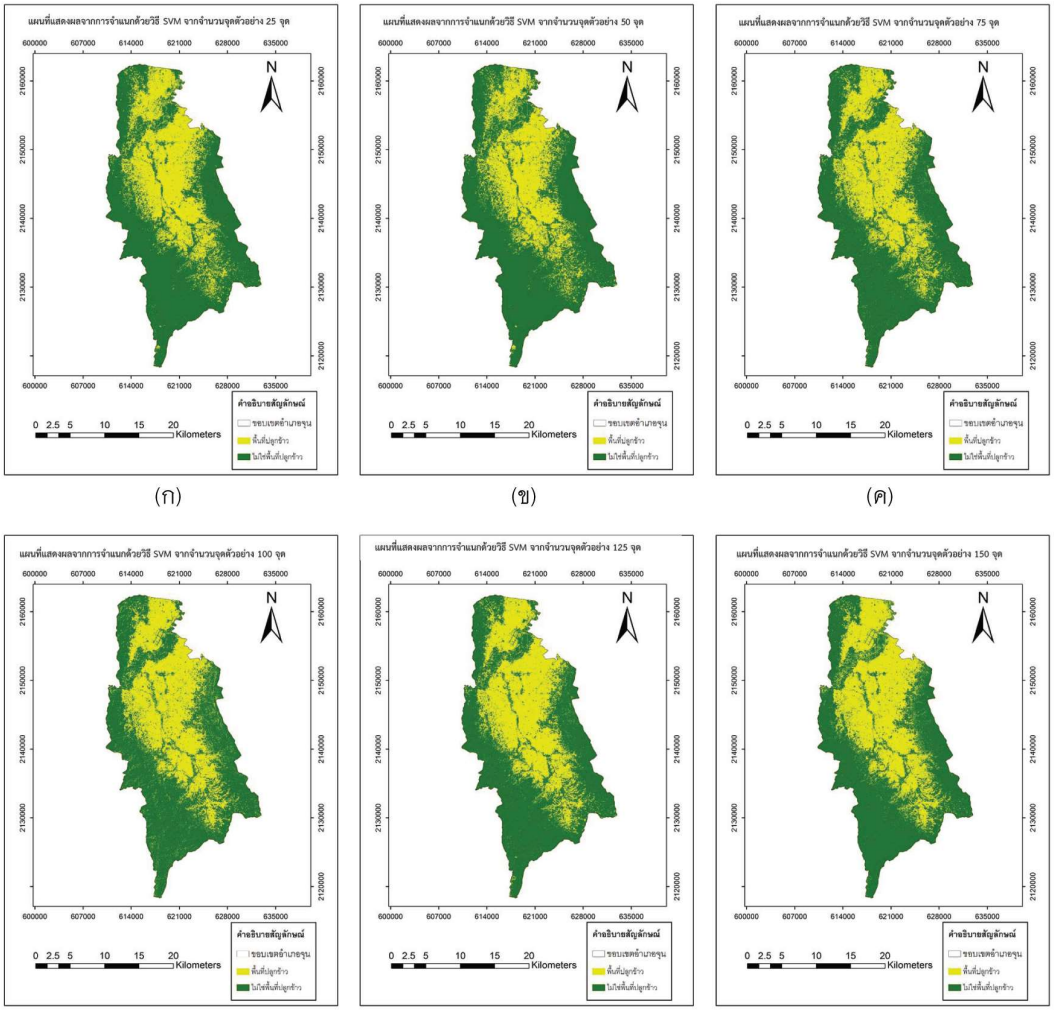
รูปที่ 4 ขั้นตอนการดำเนินงานวิจัย

4. ผลการศึกษา

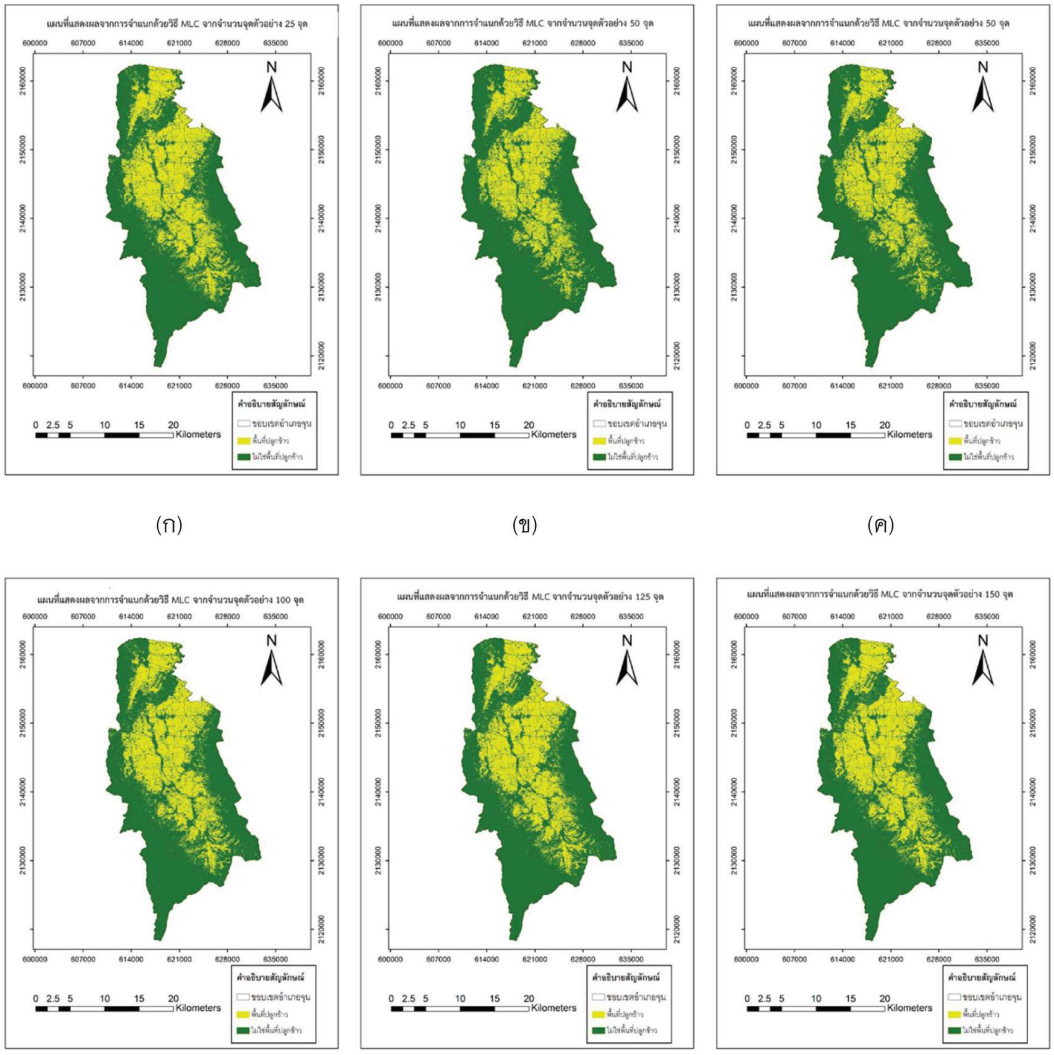
การศึกษาครั้งนี้เน้นการศึกษาเปรียบเทียบประสิทธิภาพของเทคนิควิธีการจำแนกข้อมูลแบบดั้งเดิม (การจำแนกข้อมูลภาพด้วยวิธีความน่าจะเป็นสูงสุด (MLC)) ที่ได้รับความนิยมมาอย่างยาวนานและการจำแนกข้อมูลภาพด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM) ที่อยู่ในหมวดการเรียนรู้ของเครื่องจักร (Machine Learning – ML) โดยการแบ่งข้อมูลที่ต้องการให้เครื่องจักรจำแนกเป็น 2 ประเภท คือ พื้นที่เพาะปลูกข้าวและไม่ใช่พื้นที่เพาะปลูกข้าว การตรวจสอบความถูกต้องของการจำแนกและการประเมินประสิทธิภาพของทั้งสองวิธีจะอาศัยจุดตัวอย่าง (TP) ที่แตกต่างกันเพื่อศึกษาถึงความถูกต้องของการจำแนกที่ได้รับ ในงานด้านการจำแนกข้อมูลภาพจากการรับรู้จากระยะไกลนั้น ต้นทุนและค่าใช้จ่ายในภาคสนามถือได้ว่าเป็นต้นทุนที่มีความสำคัญอย่างมาก เนื่องจากมีค่าใช้จ่ายที่สูงและใช้ระยะเวลานานในการสำรวจข้อมูล รวมถึงกำลังคนหรือจำนวนผู้ปฏิบัติงาน การลดจำนวนความต้องจุดตัวอย่างสำหรับการจำแนกข้อมูลจะช่วยให้ผู้ที่สนใจและผู้ที่ทำงานด้านการรังวัดสำรวจและการรับรู้จากระยะไกลสามารถทำงานให้บรรลุเป้าหมายได้โดยอาศัยทรัพยากรและต้นทุนที่ต่ำกว่า แต่ยังทำให้ได้ผลลัพธ์ที่มีความถูกต้องแม่นยำอยู่ในเกณฑ์มาตรฐานงานที่ยอมรับได้ รายละเอียดของการจำแนกข้อมูลภาพด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM) นั้นผู้ใช้งานสามารถเลือกชนิดของเคอร์แนลได้ 4 รูปแบบประกอบด้วย รูปแบบเชิงเส้น (Linear), โพลิโนเมียล (Polynomial), อาร์บีเอฟ (RBF) และ ซิกมอยด์ (Sigmoid) โดยผลการประเมินความถูกต้องจากการทดลองพบว่า

รูปแบบที่เหมาะสมในการจำแนก คือ เคอร์แนลแบบโพลีโนเมียล (Polynomial) และค่าพารามิเตอร์ที่เหมาะสมในการจำแนกคือ ค่าลำดับ (degree) เท่ากับ 5 ทำให้ผู้วิจัยเลือกใช้การกำหนดค่าพารามิเตอร์ดังกล่าวเป็นตัวแทนการจำแนกข้อมูลภาพด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM) ซึ่งเคอร์แนลแบบโพลีโนเมียล (Polynomial) นี้เป็นเคอร์แนลที่ให้ค่าความถูกต้องของการจำแนกในระดับสูงเช่นเดียวกับงานวิจัยของอูสตูเนอร์และคณะ (Üstüner, Sanli & Dixon, 2015)

ในส่วนของจุดตัวอย่าง (TP) นั้นผู้วิจัยได้ทำการแบ่งออกเป็น 6 ระดับที่มีจำนวนจุดตัวอย่างเท่ากับ 25, 50, 75, 100, 125 และ 150 จุด ตามลำดับ และอาศัยจุดตรวจสอบ (CP) เพื่อใช้ในการประเมินความถูกต้องของการจำแนกจำนวนทั้งสิ้น 406 จุด ผลการจำแนกข้อมูลภาพด้วยวิธีพพอร์ตเวกเตอร์แมชชีน (SVM) และวิธีความน่าจะเป็นสูงสุด (MLC) แสดงดังรูปที่ 5 และรูปที่ 6 กราฟเปรียบเทียบค่าความถูกต้องโดยรวม (OA) และค่าสัมประสิทธิ์แคปปา (Kappa) แสดงดังรูปที่ 7 และรูปที่ 8 และตารางประเมินความถูกต้องของการจำแนกเพื่อใช้ประกอบการศึกษาประสทธิภาพของแต่ละวิธีสามารถแสดงดังตารางที่ 1 ดังนี้



รูปที่ 5 ผลการจำแนกข้อมูลภาพด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM) โดยอาศัยจุดตัวอย่างจำนวน (ก) 25 จุด (ข) 50 จุด (ค) 75 จุด (ง) 100 จุด (จ) 125 จุดและ (ฉ) 150 จุด



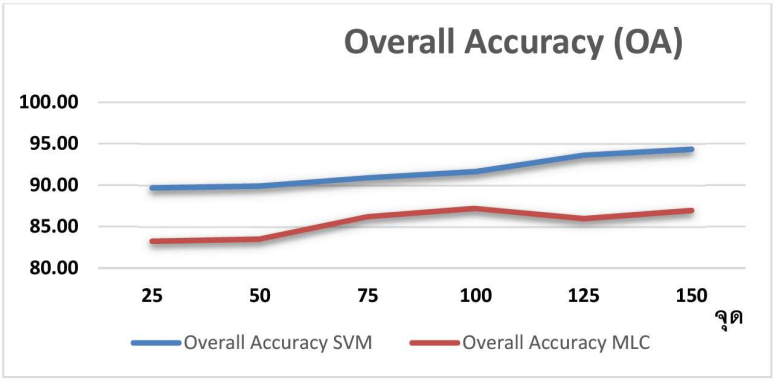
(ก) (ข) (ค)

(ง) (จ) (ฉ)

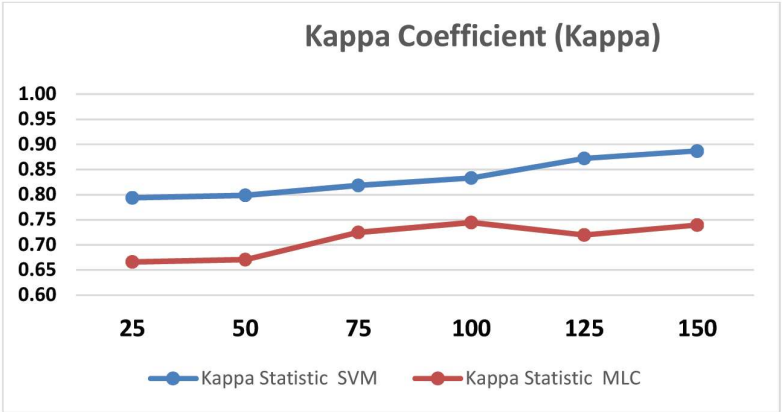
รูปที่ 6 ผลการจำแนกข้อมูลภาพด้วยวิธีความน่าจะเป็นสูงสุด (MLC) โดยอาศัยจุดตัวอย่างจำนวน (ก) 25 จุด (ข) 50 จุด (ค) 75 จุด (ง) 100 จุด (จ) 125 จุดและ (ฉ) 150 จุด

ตารางที่ 1 ผลการประเมินความถูกต้องของการจำแนกข้อมูลภาพด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM) กับวิธีความน่าจะเป็นสูงสุด (MLC)

จำนวนจุดตัวอย่าง	วิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM)		วิธีความน่าจะเป็นสูงสุด (MLC)	
	ค่าความถูกต้องโดยรวม (OA)	ค่าสัมประสิทธิ์แคปปา (Kappa)	ค่าความถูกต้องโดยรวม (OA)	ค่าสัมประสิทธิ์แคปปา (Kappa)
25	89.66	0.79	83.25	0.67
50	89.90	0.80	83.50	0.67
75	90.89	0.82	86.21	0.72
100	91.63	0.83	87.19	0.74
125	93.60	0.87	85.96	0.72
150	94.33	0.89	86.95	0.74



รูปที่ 7 กราฟเปรียบเทียบค่าความถูกต้องโดยรวม (OA) ระหว่างวิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM) และวิธีความน่าจะเป็นสูงสุด (MLC)



รูปที่ 8 กราฟเปรียบเทียบค่าสัมประสิทธิ์แคปปา (Kappa) ระหว่างวิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM) และวิธีความน่าจะเป็นสูงสุด (MLC)

5. สรุปผลการศึกษา

การจำแนกพื้นที่ปลูกข้าวด้วยวิธีซัพพอร์ตเว็กเตอร์แมชชีน (SVM) ให้ค่าความถูกต้องโดยรวม (SVM–OA) ของการจำแนก ในทุกจุดตัวอย่าง (25, 50, 75, 100, 125 และ 150 จุด) เท่ากับร้อยละ 89.66, 89.90, 90.89, 91.63, 93.60 และร้อยละ 94.33 ตามลำดับ ส่วนวิธีการจำแนกข้อมูลภาพแบบความน่าจะเป็นสูงสุดให้ค่าความถูกต้องโดยรวมของการจำแนก (MLC–OA) เท่ากับร้อยละ 83.25, 83.50, 86.21, 87.19, 85.96 และร้อยละ 86.95 ตามลำดับ และค่าสัมประสิทธ์แคปปา (Kappa) จากการจำแนกข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM–Kappa) เท่ากับ 0.79, 0.80, 0.82, 0.83, 0.87, และ 0.89 ตามลำดับ ส่วนค่าสัมประสิทธ์แคปปา (Kappa) จากการจำแนกข้อมูลแบบความน่าจะเป็นสูงสุด (MLC–Kappa) ให้ค่าเท่ากับ 0.67, 0.67, 0.72, 0.74, 0.72, และ 0.74 ตามลำดับ

จากผลการจำแนกพื้นที่ปลูกข้าวในงานวิจัยครั้งนี้พบว่า เทคนิคการจำแนกข้อมูลภาพด้วยวิธีซัพพอร์ตเว็กเตอร์แมชชีน (SVM) ให้ค่าความถูกต้องโดยรวมของการจำแนกข้อมูล (OA) และค่าสัมประสิทธิ์แคปปา (Kappa) สูงกว่าวิธีการจำแนกข้อมูลภาพวิธีความน่าจะเป็นสูงสุด (MLC) ในทุกระดับจำนวนจุดตัวอย่างที่ใช้งาน แสดงให้เห็นถึงประสิทธิภาพของวิธีซัพพอร์ตเว็กเตอร์แมชชีน (SVM) ที่สูงกว่าวิธีความน่าจะเป็นสูงสุด (MLC) อย่างชัดเจนทั้งในด้านความถูกต้องของการจำแนกข้อมูลและความต้องการจำนวนจุดตัวอย่างที่น้อยกว่า ซึ่งสอดคล้องกับงานวิจัยของมอนดอลและคณะ (Mondal et al., 2012) โรกนีและคณะ (Rokni et al., 2014) และ ทรัน วัน นินห์และชาติชาย ไวยสุระสิงห์ (2560) และจุดเด่นที่สำคัญของวิธีซัพพอร์ตเว็กเตอร์แมชชีน (SVM) คือ ความต้องการจุดตัวอย่างจำนวนน้อยแต่ก็ยังให้ค่าความถูกต้องโดยรวม (OA) ในระดับสูง ที่สอดคล้องกับงานวิจัยของอูสตูเนอร์และคณะ (Üstüner et al., 2015) ทำให้สามารถสรุปได้ว่า วิธีซัพพอร์ตเว็กเตอร์แมชชีน (SVM) เป็นวิธีการจำแนกข้อมูลภาพที่มีประสิทธิภาพสูงกว่าวิธีความน่าจะเป็นสูงสุด (MLC) ที่ได้รับความนิยมอย่างมากในอดีตอย่างชัดเจน

6. เอกสารอ้างอิง

[1] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[2] Jensen, J. R. (2005). Introductory Digital Image Processing: A Remote Sensing Perspective. Prentice Hall.

[3] Lillesand, T. M. & Kiefer, R. W. (2000). Remote sensing and image interpretation. John Wiley & Sons.

[4] Miller, J. D. (2017). Statistics for data science. Packt.

[5] Mondal, A., Kudu, S., Chandniha, D. K. & Shukla, R. (2012). Comparison of support vector machine and maximum likelihood classification technique using satellite imagery. International Journal of Remote Sensing and GIS. 1. 116–123.

[6] Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing. 66. 247–259.

[7] Rokni, B., Bin Ahmad, B. & Zabihi, H. (2014). Comparison of two Classification methods (MLC and SVM) to extract land use and land cover in Johor Malaysia. IOP Conference Series: Earth and Environmental Science. 20. 012052. 10.1088/1755–1315/20/1/012052.

[8] The Rice trader. (2020, Dec 5). World's best rice contest. https://thericetrader.com/conference–worlds–best–rice/

[9] Üstüner, M., Sanli, B. F., & Dixon, B. (2015). Application of Support Vector Machines for Landuse Classification Using High–Resolution RapidEye Images: A Sensitivity Analysis. European Journal of Remote Sensing. 48. 403 – 422.

[10] ทรัน วัน นินห์ และ ชาติชาย ไวยสุระสิงห์. (2560).การศึกษาเปรียบเทียบวิธีแบบความน่าจะเป็นได้สูงที่สุดกับวิธีซัพพอร์ตเวกเตอร์แมชชีนในการวิเคราะห์ข้อมูลภาพดาวเทียมแลนแซตเพื่อประเมินการเปลี่ยนแปลงการใช้ที่ดินเมืองเก่า. มหาวิทยาลัยขอนแก่น

[11] ปริญญา สงวนสัตย์. (2562). Artificial intelligence with Machine learning – AI สร้างได้ด้วยแมชชีนเลิร์นนิง (Python edition). นนทบุรี: ไอดีซี พรีเมียร์.

[12] สำนักงานปลัดกระทรวงเกษตรและสหกรณ์. (ม.ป.ป). ข้อมูลพื้นฐานจังหวัดพะเยา. สืบค้นเมื่อ 1 พฤษจิกายน 2563, จาก https://www.opsmoac.go.th/phayao–dwl–files–401291791096

[13] สำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ (องค์การมหาชน). (2552). ตำราเทคโนโลยีอวกาศและภูมิสารสนเทศศาสตร์ (พิมพ์ครั้งที่ 1). กรุงเทพฯ : สำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ (องค์การมหาชน).

[14] สุวิทย์ อ๋องสมหวัง. (2559). ระบบการรับรู้จากระยะไกลและการวิเคราะห์ข้อมูลภาพเชิงเลข. นครราชสีมา: สาขาวิชาการรับรู้จากระยะไกล สำนักวิชาวิทยาศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี.