Received: 24 มี.ค. 2568 **Revised:** 21 เม.ย. 2568 **Accepted:** 23 เม.ย. 2568

การคาดการณ์พื้นที่น้ำท่วมบริเวณพื้นที่ลุ่มน้ำชุมชนเมืองจังหวัดเชียงใหม่ โดยใช้การสำรวจระยะไกลและการเรียนรู้การถดถอย

Flood Area Prediction in Urban Watershed of Chiang Mai Province Using Remote Sensing and Regression Learning

เพชรรัช ปะระไทย[1], นฤพร เต็งไตรรัตน์[1*] และ Wai Lok Woo[2]

[1]สาขาวิชาวิศวกรรมซอฟต์แวร์ คณะบริหารธุรกิจ มหาวิทยาลัยพายัพ เชียงใหม่ ประเทศไทย

[2]Department of Computer and Information Sciences, Northumbrai University, Newcastle Upon Tyne, United Kingdom

Phetcharat Parathai[1], Naruephorn Tengtrairat[1*] and Wai Lok Woo [2]

[1]School of Software Engineering, Faculty of Business Administration, Payap University, Chiang Mai, Thailand.

[2]Department of Computer and Information Sciences, Northumbrai University, Newcastle Upon Tyne, United Kingdom

[*]Corresponding author: naruephorn_t@payap.ac.th

Abstract

The current state of global warming has led to more frequent and severe flooding. Accurate flood warnings in advance can help mitigate damage to lives and property. This research presents a modeling approach for predicting flood-prone areas using satellite imagery data extraction. The study focuses on urban areas in Chiang Mai Province, covering a total area of 902.025 square meters. Six key factors influencing flood prediction are considered: water balance, evapotranspiration, land cover, rainfall, NDVI, and flood extent. These factors are derived from satellite images captured by Sentinel, MODIS, and Landsat satellites via Google Earth Engine. Satellite imagery provides extensive and continuous spatial coverage, while radar satellite (SAR) data remains operational even in cloudy weather or at night, reducing the cost and time required for field surveys. The flood prediction models were developed using seven regression methods: LR, SVR, Decision Tree, Random Forest, Gradient Boosting, XGBoost, and LightGBM. These methods were selected for their high efficiency and

computational speed, making them suitable for large datasets without requiring high processing power. The Random Forest method demonstrated the highest accuracy in flood area prediction, with lower RMSE and higher R-squared values compared to XGBoost, LightGBM, Decision Tree, Gradient Boosting, SVR, and LR, with average improvements of 3%, 12%, 19%, 31%, 51%, and 65%, respectively. This research holds significant potential for practical application, particularly through the development of an early warning system for flood hazards at the local level. Such a system would be especially beneficial in urban areas with high population density, enhancing the efficiency and timeliness of emergency response planning and evacuation efforts by relevant agencies in a systematic and coordinated manner.

Keywords: *remote sensing; flood prediction; regression; machine learning; satellite imagery*

บทคัดย่อ

ภาวะโลกร้อนปัจจุบันทำให้เกิดอุทกภัยรุนแรงและมีความถี่ในการเกิดสูงขึ้น หากสามารถแจ้งเตือนพื้นที่น้ำท่วมล่วงหน้าได้อย่างตรงจุด จะช่วยลดความเสียหายแก่ชีวิตและทรัพย์สินของประชาชนได้ งานวิจัยนี้นำเสนอ การสร้างแบบจำลองเพื่อคาดการณ์พื้นที่น้ำท่วมโดยใช้การสกัดข้อมูลจากภาพจากดาวเทียม บริเวณพื้นที่ชุมชนเมือง จ. เชียงใหม่ ครอบคลุมพื้นที่ทั้งหมด 902.025 ตารางเมตร ปัจจัยที่เกี่ยวข้องกับการคาดการณ์พื้นที่น้ำท่วม ประกอบด้วย 6 ปัจจัย ได้แก่ สมดุลน้ำ การคายระเหยของน้ำ พื้นผิวปกคลุมดิน ปริมาณน้ำฝน NDVI และ พื้นที่น้ำท่วมจากภาพถ่ายดาวเทียม Sentinel ดาวเทียม MODIS และ ดาวเทียม Landsat ผ่าน Google Earth Engine ภาพถ่ายดาวเทียมครอบคลุมพื้นที่กว้างและต่อเนื่อง ข้อมูลจากดาวเทียมเรดาร์ (SAR) สามารถใช้งานได้แม้ในสภาพอากาศที่มีเมฆมากหรือเวลากลางคืน ช่วยลดต้นทุนและเวลาในการสำรวจ สำหรับแบบจำลองได้ถูกสร้างโดยใช้วิธีทาง Regression จำนวน 7 วิธี ได้แก่ LR, SVR, Decision Tree, Random Forest, Gradient Boosting, XGBoost และ LightGBM เนื่องจากมีประสิทธิภาพสูงและคำนวณได้เร็วสามารถใช้กับชุดข้อมูลขนาดใหญ่ได้โดยไม่ต้องการพลังประมวลผลสูง จากการศึกษาพบว่า วิธี Random Forest สามารถคาดการณ์พื้นที่น้ำท่วมได้ถูกต้องสูงสุด มีค่า RMSE และ R-Square สูงกว่าวิธี XGBoost LightGBM Decision Tree Gradient Boosting SVR และ LR คิดเป็นค่าเฉลี่ยร้อยละ 3 ร้อยละ 12 ร้อยละ 19 ร้อยละ 31 ร้อยละ 51 และ ร้อยละ 65 ตามลำดับ งานวิจัยนี้สามารถพัฒนาต่อยอดให้เกิดประโยชน์ โดยการนำแบบจำลองที่พัฒนาขึ้นไปประยุกต์ใช้เป็นระบบแจ้งเตือนภัย

น้ำท่วมล่วงหน้าในระดับพื้นที่ โดยเฉพาะในเขตเมืองซึ่งมีความหนาแน่นของประชากรสูง ซึ่งจะช่วยเสริมประสิทธิภาพในการวางแผนรับมือและดำเนินการอพยพประชาชนของหน่วยงานที่เกี่ยวข้องได้อย่างทันท่วงทีและเป็นระบบ

คำสำคัญ: *การสำรวจระยะไกล; การคาดการณ์น้ำท่วม; การถดถอย; การเรียนรู้ของเครื่อง; ภาพถ่ายดาวเทียม*

1. บทนำ

การเกิดน้ำท่วมเป็นปัญหาสำคัญที่ส่งผลกระทบต่อชีวิตและทรัพย์สินของประชาชน รวมถึงเศรษฐกิจและสิ่งแวดล้อมของประเทศ การพัฒนาแบบจำลองเพื่อคาดการณ์พื้นที่น้ำท่วมที่มีประสิทธิภาพจึงเป็นสิ่งจำเป็น เพื่อช่วยในการวางแผนและจัดการทรัพยากรอย่างเหมาะสม จังหวัดเชียงใหม่เป็นพื้นที่ที่มีความเสี่ยงต่อการเกิดน้ำท่วม เนื่องจากมีลักษณะภูมิประเทศเป็นที่ราบลุ่มแม่น้ำปิงและมีแม่น้ำสาขาหลายสายไหลผ่าน ประกอบกับหากในช่วงฤดูฝนจะมีปริมาณน้ำฝนสะสมจำนวนมาก อาจทำให้เกิดปัญหาน้ำท่วมหนักในหลายพื้นที่ของจังหวัด ดังที่เกิดขึ้นในปี พ.ศ. 2565 และพ.ศ. 2567 ในพื้นที่เขตเทศบาลนครเชียงใหม่ พื้นที่ริมฝั่งแม่น้ำปิง พื้นที่เกษตรกรรมในเขตอำเภอต่างๆ พื้นที่เศรษฐกิจสำคัญ ข้อมูลจากรายงานสถานการณ์สาธารณภัย กรมป้องกันและบรรเทาสาธารณภัย ในวันที่ 10 กันยายน พ.ศ.2567 พบว่าจังหวัดเชียงใหม่มีพื้นที่ประสบภัยน้ำท่วม 2 อำเภอ 5 ตำบล 22 หมู่บ้าน และมีครัวเรือนได้รับผลกระทบ 2,928 ครัวเรือน และมีผู้เสียชีวิตจากเหตุดินถล่ม 6 ราย และผู้บาดเจ็บ 3 ราย(กรมป้องกันและบรรเทาสาธารณภัย,2567) ข้อมูลจากดาวเทียมเป็นทรัพยากรที่มีคุณค่าอย่างยิ่งในการเสริมสร้างความแม่นยำ และประสิทธิภาพของการคาดการณ์น้ำท่วม เนื่องจากสามารถให้ข้อมูลที่มีความละเอียดเชิงพื้นที่สูง ครอบคลุมพื้นที่ขนาดใหญ่ และสามารถติดตามการเปลี่ยนแปลงของสภาพแวดล้อมได้อย่างต่อเนื่องทั้งในมิติของเวลาและพื้นที่ โดยเฉพาะข้อมูลจากดาวเทียมแบบหลายช่วงคลื่น (multispectral) เช่น Sentinel-2 และ Landsat ซึ่งสามารถนำมาใช้ในการวิเคราะห์ตัวชี้วัดที่สำคัญ เช่น ดัชนีพืชพรรณ (NDVI) ความชื้นในดิน ลักษณะการใช้ที่ดิน และระดับน้ำท่วมได้อย่างมีประสิทธิภาพ (Pham et al., 2021) ในขณะเดียวกัน ข้อมูลจากดาวเทียมเรดาร์ (Synthetic Aperture Radar: SAR) เช่น Sentinel-1 ยังมีศักยภาพในการตรวจจับผิวน้ำได้แม้ในสภาพอากาศที่มีเมฆมากหรือในช่วงเวลากลางคืน ซึ่งช่วยลดข้อจำกัดของการสำรวจภาคสนามและเพิ่มความแม่นยำในการระบุขอบเขตของพื้นที่ที่ได้รับผลกระทบจากน้ำท่วม (Nhangumbe et al., 2023) เมื่อข้อมูลดังกล่าวถูกรวมเข้ากับเทคนิคด้านปัญญาประดิษฐ์ เช่น Random Forest, XGBoost หรือแบบจำลอง Deep Learning ข้อมูลจากดาวเทียมจึงสามารถพัฒนาเป็นองค์ประกอบ

หลักของระบบคาดการณ์น้ำท่วมที่มีความแม่นยำสูง ทั้งในระดับท้องถิ่นและระดับภูมิภาค (Sanderson et al., 2023)

ในปัจจุบัน เทคโนโลยีการเรียนรู้ของเครื่อง (Machine Learning) ได้รับความสนใจอย่างมากในการนำมาประยุกต์ใช้คาดการณ์และจำลองสถานการณ์น้ำท่วม เนื่องจากความสามารถในการวิเคราะห์ข้อมูลจำนวนมากและซับซ้อนได้อย่างมีประสิทธิภาพ โดยเฉพาะการใช้เทคนิคการฝังตัว (Embeddings) เพื่อแปลงข้อมูลที่มีมิติสูงให้อยู่ในรูปแบบเวกเตอร์ที่มีมิติต่ำลง ทำให้สามารถวิเคราะห์และประมวลผลข้อมูลได้รวดเร็วและแม่นยำยิ่งขึ้น การสร้างแบบจำลองด้วยวิธีการเรียนรู้การถดถอย (Regression Learning) ได้ถูกนำมาใช้อย่างกว้างขวางเพื่อประมาณการการเกิดน้ำท่วมในพื้นที่เป้าหมาย จากบทความ (Manikandan et al., 2024) ได้ใช้วิธี Regression Learning พิจารณาลักษณะของน้ำฝนที่เฉพาะเจาะจงในแต่ละพื้นที่ เพื่อตัดสินความน่าจะเป็นของเหตุการณ์น้ำท่วม โดยใช้ข้อมูลน้ำฝนระดับเขตจากประเทศอินเดีย ซึ่งเป็นฐานข้อมูลในการฝึกฝนและพัฒนาโมเดลการทำนาย เทคนิคการเรียนรู้ของเครื่อง (ML) หลายประเภท เช่น XGBoost และ K-Nearest Neighbors (KNN) ถูกใช้ในการฝึกฝนอัลกอริธึมทำนายการเกิดน้ำท่วม (Hakim et al., 2024) พบว่าการทำงานที่ผสานการสังเกตการณ์จากดาวเทียมหลายดวงและการจำลองทางอุทกพลศาสตร์ เพื่อให้การพยากรณ์น้ำท่วมขนาดใหญ่เป็นไปอย่างรวดเร็ว แม่นยำ โดยไม่ขึ้นอยู่กับความละเอียดของข้อมูล นอกจากนี้ การสร้างแผนที่น้ำท่วม ด้วยการพัฒนาอัลกอริธึมการเรียนรู้เชิงลึก จากภาพเรดาร์สังเคราะห์และภาพออปติคัลมัลติสเปกตรัมโดยใช้แค่แถบสีแดง สีเขียว และสีน้ำเงิน (RGB) จากภาพ Sentinel-2 ซึ่งแสดงให้เห็นว่า ภาพเต็ม 13 แบนด์เป็นตัวเลือกที่ได้เปรียบกว่าสำหรับการฝึกโมเดลการเรียนรู้เชิงลึก (Sanderson et al., 2023)

การคาดการณ์น้ำท่วมในปัจจุบันยังคงมีความท้าทายหลายประการ อาทิ การรวบรวมข้อมูลที่ครอบคลุมและมีคุณภาพสูง การพัฒนาแบบจำลองที่สามารถทำงานได้ในสภาพแวดล้อมที่เปลี่ยนแปลงตลอดเวลาเป็นสิ่งที่ท้าทาย ซึ่งอาจส่งผลต่อความแม่นยำและประสิทธิภาพของแบบจำลองคาดการณ์ การวิจัยนี้มุ่งเน้นการพัฒนาแบบจำลองคาดการณ์พื้นที่น้ำท่วม ด้วยการสกัดข้อมูลจากภาพถ่ายดาวเทียมเพื่อสร้างชุดข้อมูลพื้นที่น้ำท่วม และใช้เทคนิคการเรียนรู้แบบ Regression Learning 7 วิธี เพื่อหาวิธีที่มีความแม่นยำในการคาดการณ์พื้นที่น้ำท่วมสูงสุด

โครงสร้างของบทความนี้ประกอบด้วยส่วนหลัก 5 ส่วน ได้แก่ ส่วนที่ (1) บทนำ อธิบายถึงที่มา ความสำคัญ งานวิจัยที่มีความเกี่ยวข้องกับ ส่วนที่ (2) การสกัดและรวบรวมข้อมูลน้ำท่วม ส่วนที่ (3) อธิบายหลักการของแบบจำลองการเรียนรู้ของเครื่องจักรสำหรับคาดการณ์พื้นที่น้ำท่วม ในส่วนที่ (4) นำเสนอการทดลองและเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่องจักร และส่วนสุดท้าย ส่วนที่ (5) สรุปผลงานวิจัย

2. การสกัดและรวบรวมข้อมูลน้ำท่วม

เทคโนโลยีการสำรวจระยะไกล (Remote Sensing) โดยเฉพาะอย่างยิ่งการใช้ภาพถ่ายดาวเทียม จึงเป็นเครื่องมือสำคัญในการรวบรวมข้อมูลเกี่ยวกับน้ำท่วม เนื่องจาก Remote Sensing สามารถให้ข้อมูลที่ครอบคลุมในพื้นที่กว้างและอัปเดตได้อย่างต่อเนื่อง กระบวนการสกัดข้อมูลน้ำท่วมจากภาพถ่ายดาวเทียมประกอบด้วย 2 ขั้นตอนหลัก คือ 1) การเลือกประเภทของดาวเทียมที่เหมาะสม และ 2) การประมวลผลภาพเพื่อสกัดข้อมูลพื้นที่น้ำท่วมและปัจจัยที่เกี่ยวข้องเพื่อสร้างชุดข้อมูลพื้นที่น้ำท่วม

ดาวเทียมที่ใช้ในการวิเคราะห์น้ำท่วมสามารถแบ่งออกเป็น 2 ประเภทหลัก ได้แก่ ดาวเทียมสำรวจโลกในช่วงแสงที่ตามองเห็น (Optical Satellite) และดาวเทียมเรดาร์ตรวจจับระยะไกล (Synthetic Aperture Radar: SAR) ในงานวิจัยนี้ได้ใช้ข้อมูลจากดาวเทียมต่อไปนี้

ดาวเทียม Sentinel (European Space Agency, n.d.). โดยใช้ข้อมูล "Harmonized" หรือ “S2_Harmonized” ข้อมูลนี้เป็นข้อมูลระดับ 2A (Level-2A) เป็นข้อมูลที่ได้รับการปรับปรุงให้มีความสอดคล้องกันมากขึ้นระหว่างข้อมูลจากดาวเทียม Sentinel-2A และ Sentinel-2B มีความละเอียดสูง (สูงสุดที่ 10 เมตร) ทำให้สามารถนำข้อมูลจากทั้งสองดาวเทียมมาวิเคราะห์ร่วมกันได้อย่างมีประสิทธิภาพ เช่น การคำนวณดัชนีการเจริญเติบโตของพืช (Normalized Difference Vegetation Index: NDVI)

ดาวเทียม MODIS (NASA, n.d.) มีหน้าที่ในการตรวจวัดรังสีที่สะท้อนและแผ่ออกมาจากพื้นผิวโลกและชั้นบรรยากาศ ชุดข้อมูล MOD16A2 ให้ข้อมูลเกี่ยวกับการระเหยของน้ำ (Evapotranspiration) ซึ่งเป็นกระบวนการที่น้ำระเหยจากพื้นดินและพืชสู่ชั้นบรรยากาศ ความละเอียดเชิงพื้นที่พิกเซลในภาพครอบคลุมพื้นที่ 500 x 500 เมตรบนพื้นผิวโลก

ชุดข้อมูล WWF/HydroSHEDS/03VFDEM (Lehner, Verdin, & Jarvis, 2008) เป็นชุดข้อมูลแผนที่ที่จัดทำโดยโครงการ HydroSHEDS (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales) ชุดข้อมูลนี้ให้ข้อมูลภูมิประเทศในรูปแบบของโมเดลระดับความสูงดิจิทัล (Digital Elevation Model: DEM) ที่ถูกเติมเต็มช่องว่าง (void-filled) เพื่อความสมบูรณ์ของข้อมูล

ชุดข้อมูล JRC Global Surface Water Mapping Layers จัดทำโดยศูนย์วิจัยร่วมแห่งสหภาพยุโรป (Joint Research Centre: JRC) เป็นแผนที่แสดงตำแหน่งและการกระจายตัวของน้ำผิวดินทั่วโลกตั้งแต่ปี พ.ศ. 2527 ถึง พ.ศ. 2564 โดยใช้ภาพถ่ายจากดาวเทียม Landsat 5, 7 และ 8 (Pekel et al., 2016)

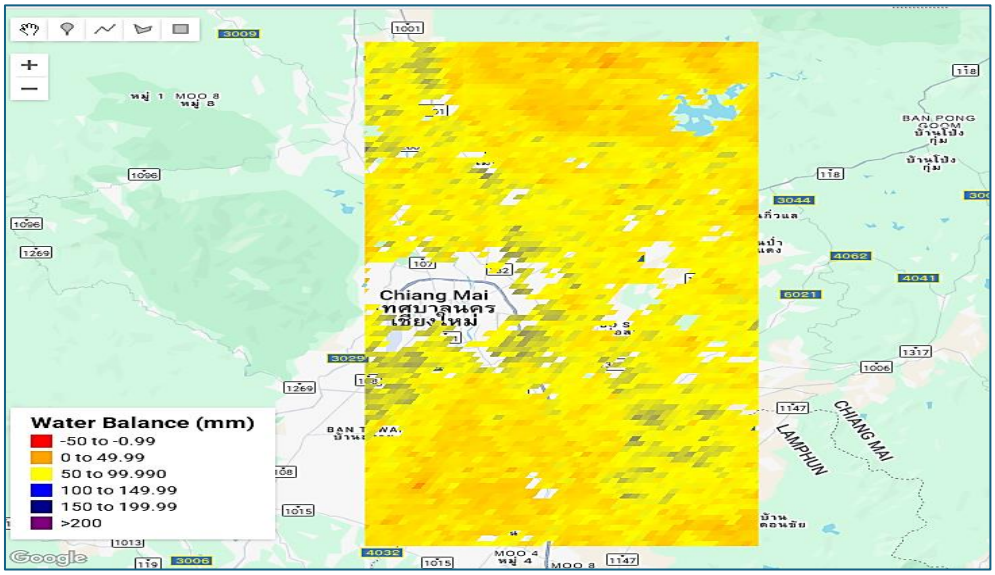
ชุดข้อมูล CHIRPS Pentad (Climate Hazards Center InfraRed Precipitation with Station data) เป็นชุดข้อมูลปริมาณน้ำฝนที่พัฒนาโดย Climate Hazards Center แห่งมหาวิทยาลัยแคลิฟอร์เนีย ซานตาบาร์บารา (UCSB) ภาพถ่ายดาวเทียมอินฟราเรดความละเอียดเชิงพื้นที่ประมาณ 5 เมตร ความถี่การรวมข้อมูลปริมาณน้ำฝนเป็นช่วง 5 วัน (Funk et al., 2015)

การประมวลผลภาพเพื่อสกัดพื้นที่น้ำท่วมและปัจจัยที่เกี่ยวข้อง กระบวนการสกัดและรวบรวมข้อมูลน้ำท่วมเป็นขั้นตอนสำคัญในการสร้างแบบจำลองการคาดการณ์พื้นที่น้ำท่วม ข้อมูลที่จำเป็นสำหรับการวิเคราะห์ประกอบด้วย 6 ข้อมูล ได้แก่

1. Water Balance สมดุลน้ำเป็นการคำนวณปริมาณน้ำเข้าและออกในพื้นที่เฉพาะในช่วงเวลาที่กำหนด สมการสำหรับคำนวณค่า Water Balance มีรูปแบบพื้นฐาน (Oki & Kanae, 2006) ดังสมการที่ (1)

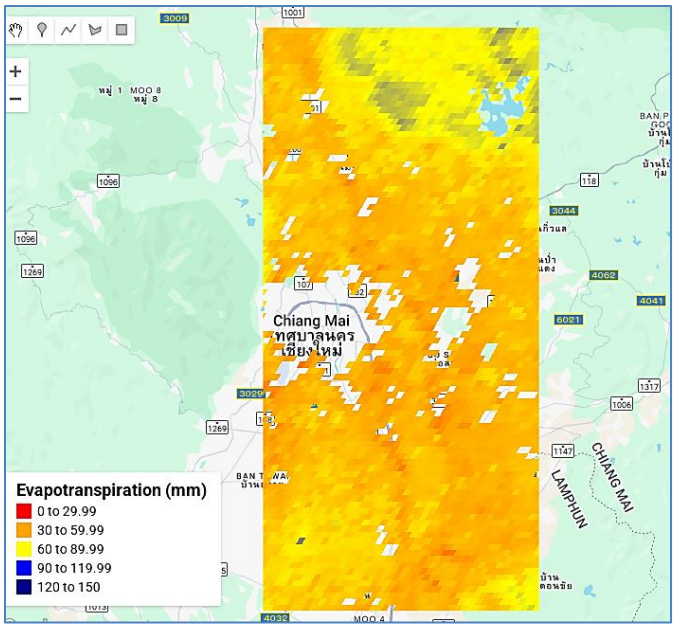
$$\Delta S = P - ET - R - G \quad (1)$$

โดยที่ ตัวแปร ΔS คือ การเปลี่ยนแปลงปริมาณน้ำที่เก็บสะสม (change in water storage) โดยทั่วไปใช้ มิลลิเมตร (mm) ต่อหน่วยเวลา ตัวแปร P คือ ปริมาณน้ำฝน (precipitation) ตัวแปร ET คือ การระเหยและการคายน้ำ (evapotranspiration) ตัวแปร R คือ การไหลบ่าของน้ำ (runoff) และตัวแปร G คือ การไหลของน้ำใต้ดิน (groundwater flow) ตัวอย่างการสกัดค่า Water Balance จากภาพถ่ายดาวเทียม แสดงในภาพที่ 1



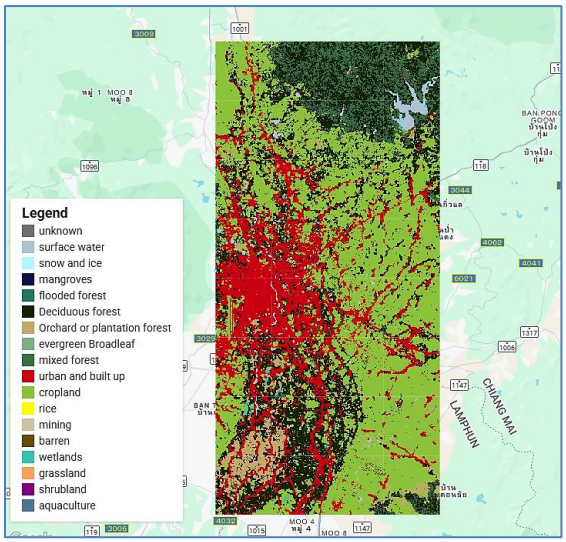
ภาพที่ 1 ตัวอย่างภาพค่าเฉลี่ย Water Balance ของพื้นที่ศึกษา จังหวัดเชียงใหม่ ช่วงปี พ.ศ.2562 ถึง พ.ศ. 2565 จากข้อมูลภาพถ่ายดาวเทียมของดาวเทียม MODIS และ CHIRPS Pentad

2. Evapotranspiration (ET) การคายระเหยของน้ำ ปริมาณน้ำทั้งหมดที่ถูกถ่ายโอนจากพื้นผิวโลกไปยังชั้นบรรยากาศในรูปของไอน้ำ กระบวนการนี้ผสมผสานกลไกที่แตกต่างกันสองประการเข้าด้วยกัน (Yang et al., 2021) การระเหย และ การคายระเหยเป็นกระบวนการที่พืชดูดซับน้ำผ่านรากแล้วปล่อยออกสู่ชั้นบรรยากาศในรูปไอน้ำผ่านรูพรุนเล็กๆ ที่เรียกว่าปากใบ ซึ่งส่วนใหญ่อยู่ที่ใบ ET เป็นองค์ประกอบสำคัญของวัฏจักรของน้ำ ซึ่งมีอิทธิพลต่อการก่อตัวของเมฆ ความชื้น และรูปแบบการตกตะกอน ดังแสดงตัวอย่างในภาพที่ 2



ภาพที่ 2 ตัวอย่างภาพค่าเฉลี่ย Evapotranspiration ของพื้นที่ศึกษา จังหวัดเชียงใหม่ ช่วงปี พ.ศ. 2562 ถึง พ.ศ. 2565 จากข้อมูลภาพถ่ายดาวเทียมของดาวเทียม MODIS และ CHIRPS Pentad

3. ข้อมูล Landcover หรือข้อมูลพื้นผิวปกคลุมดิน มีบทบาทสำคัญในการศึกษาน้ำท่วม เนื่องจากแต่ละประเภทของที่ดินมีผลต่อการไหลของน้ำ การซึมลงสู่ดิน และการกักเก็บน้ำแตกต่างกัน พื้นผิวปกคลุมดิน ของงานวิจัยนี้แบ่งออกเป็น 18 ประเภท ดังแสดงตัวอย่างในภาพที่ 3



ภาพที่ 3 ข้อมูล Landcover ของพื้นที่ศึกษา จังหวัดเชียงใหม่ ปี พ.ศ. 2565 จากข้อมูลภาพถ่ายดาวเทียมของดาวเทียม MODIS และ CHIRPS Pentad

4. Rainfall หรือ ปริมาณน้ำฝน หมายถึง ปริมาณน้ำที่ตกลงมาสู่พื้นดินในรูปของฝนภายในระยะเวลาที่กำหนด โดยวัดเป็น มิลลิเมตร (mm) น้ำที่ปกคลุมพื้นที่ 1 ตารางเมตร มีความลึก 1 มิลลิเมตร หรือ นิ้ว (inches) ซึ่งเป็นค่าที่สำคัญในการศึกษาวัฏจักรของน้ำ อุทกวิทยา และการคาดการณ์น้ำท่วม โดยสกัดข้อมูลจาก ชุดข้อมูล CHIRPS Pentad

5. NDVI (Normalized Difference Vegetation Index) หรือ ดัชนีความแตกต่างเชิงปกติของพืชพรรณ เป็นวิธีการวิเคราะห์ภาพถ่ายดาวเทียมที่ใช้ในการตรวจสอบและวัดระดับความเขียวหรือความหนาแน่นของพืชพรรณในพื้นที่ต่าง ๆ หลักการของ NDVI อ้างอิงจากลักษณะทางกายภาพของพืช คือ พืชที่มีสุขภาพดีจะ ดูดกลืนแสงสีแดง (Red) เพื่อใช้ในการสังเคราะห์แสง ในขณะเดียวกันพืชจะ สะท้อนแสงในช่วงอินฟราเรดใกล้ (Near-Infrared หรือ NIR) ออกมาในปริมาณมาก เพราะพืชไม่ต้องการแสงชนิดนี้ การวัดความแตกต่างระหว่างอินฟราเรดใกล้ เช่น พืชพรรณชนิดใดที่สะท้อนแสงสีแดงได้อย่างชัดเจน NDVI (Normalized Difference Vegetation Index) จะวัดพืชพรรณ สูตรคำนวณ NDVI ดังสมการที่ (2)

$$NDVI = \frac{(NIR-RED)}{(NIR+RED)} \quad (2)$$

โดยที่ ตัวแปร NIR (Near-Infrared Reflectance) คือแสงที่สะท้อนในสเปกตรัมอินฟราเรดใกล้ ปริมาณการสะท้อนแสงในช่วงคลื่น อินฟราเรดใกล้ (Near Infrared, ~760–900 nm) และตัวแปร

RED (Red Reflectance) คือแสงที่สะท้อนจากช่วงสีแดงของสเปกตรัม ปริมาณการสะท้อนแสงในช่วงคลื่น แสงสีแดง (Red, ~600–700 nm) โดยสกัดข้อมูลจากดาวเทียม Sentinel ค่าที่ได้จาก NDVI จะอยู่ในช่วง -1 ถึง +1 โดย ค่าที่ใกล้ +1 หมายถึง พืชมีความเขียวชอุ่ม หนาแน่น และสุขภาพดี ค่าที่ใกล้ 0 หมายถึง พื้นที่ว่าง หรือดินเปล่า ค่าที่ติดลบ (เช่น -0.2) มักแสดงถึงน้ำ หรือวัตถุที่ไม่ใช่พืช

6. Flooded Area หรือ "พื้นที่น้ำท่วม" หมายถึง บริเวณที่ถูกปกคลุมด้วยน้ำในช่วงระยะเวลาหนึ่ง ซึ่งปกติแล้วไม่ได้อยู่ใต้น้ำ พื้นที่น้ำท่วมสามารถเกิดขึ้นได้จากหลายสาเหตุ และมีลักษณะที่แตกต่างกันไป การตรวจจับพื้นที่น้ำท่วมด้วยการใช้ Remote Sensing และ ภาพถ่ายดาวเทียม เป็นวิธีสำคัญในการตรวจสอบและวิเคราะห์พื้นที่น้ำท่วม โดยใช้การวิเคราะห์ข้อมูล ระดับความสูงของพื้นดิน ปริมาณน้ำผิวดิน ร่วมกับการใช้เรดาร์ (SAR - Synthetic Aperture Radar) (Singh et al., 2021) การเปรียบเทียบภาพก่อนและหลังน้ำท่วม (Change Detection Analysis) มีกระบวนการในการวิเคราะห์พื้นที่น้ำท่วม (UN-SPIDER, n.d.) ดังนี้

1. การลดสัญญาณรบกวนแบบ Speckle ในภาพถ่ายเรดาร์ (SAR) ที่ใช้ในการวิเคราะห์น้ำท่วม Speckle เป็นสัญญาณรบกวนแบบ multiplicative ที่เกิดขึ้นในภาพถ่ายเรดาร์ (SAR) สมการ Refined Lee Speckle Filter ใช้สำหรับกำจัดสัญญานรบกวน สามารถแสดงได้ดังสมการที่ (3)

$$I_{filtered} = I_{mean} + W \cdot (I_{center} - I_{mean}) \quad (3)$$

โดยที่ $I_{filtered}$ คือค่าพิกเซลที่ถูกกรอง ณ ตำแหน่ง (x, y) I_{mean} คือค่าเฉลี่ยของพิกเซลในหน้าต่างกรอง ณ ตำแหน่ง (x, y) I_{center} คือ ค่าเฉลี่ยของพิกเซลในบริเวณใกล้เคียง ณ ตำแหน่ง (x, y) และ W คือค่าน้ำหนัก ณ ตำแหน่ง (x, y) ซึ่งคำนวณจากค่าสัมประสิทธิ์ความแปรปรวน (Coefficient of Variation) ซึ่งคำนวณจาก $W = \frac{C_{local}^2}{C_{local}^2 + C_{noise}^2}$ โดยที่ C_{local}^2 คือความแปรปรวนของพิกเซลในหน้าต่างกรอง และ C_{noise}^2 คือค่าความแปรปรวนของ noise ซึ่งโดยทั่วไปประมาณจากค่าคงที่ของภาพเรดาร์

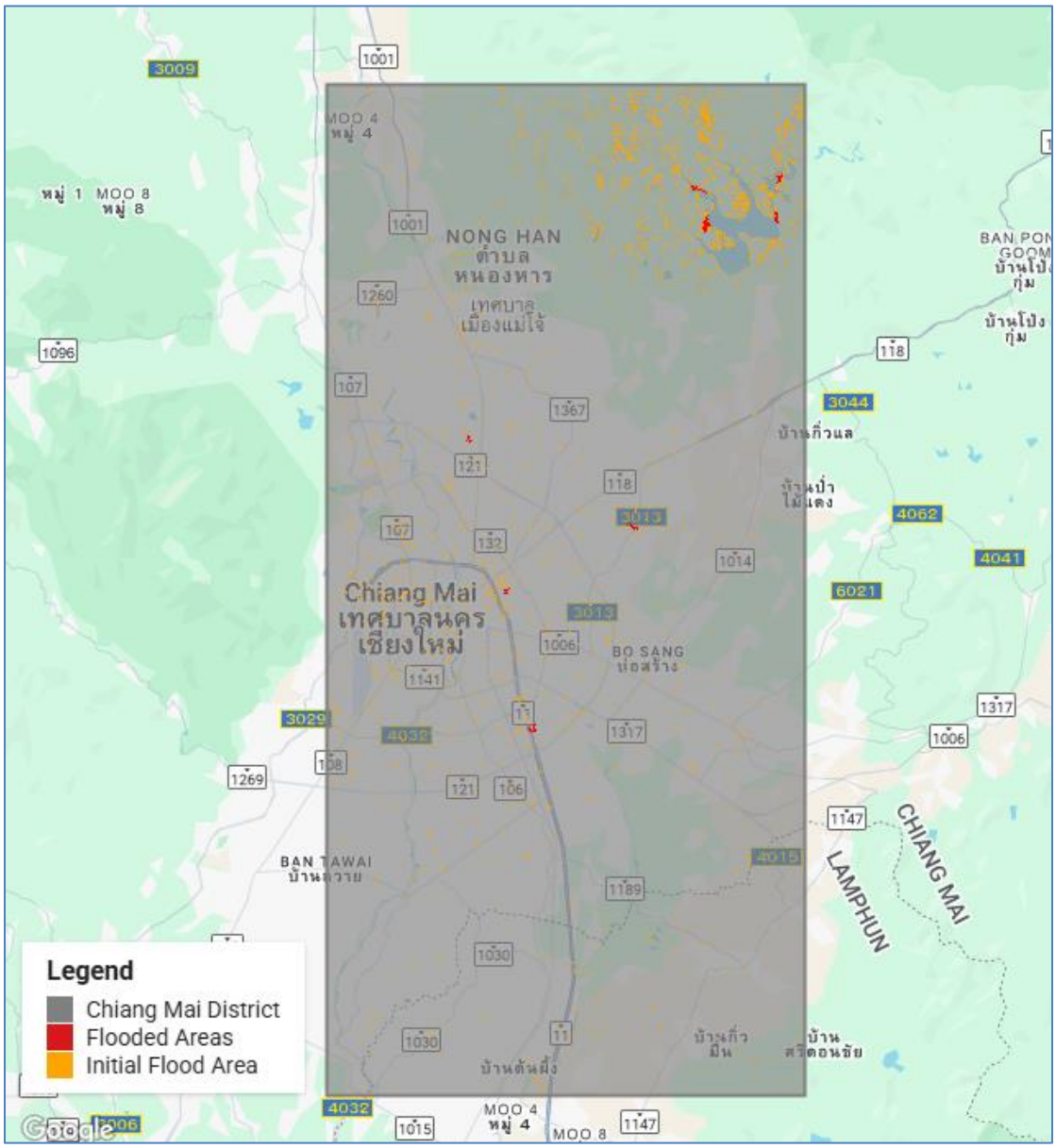
2. การกรองข้อมูลและการประมวลผลพารามิเตอร์ที่กำหนดไว้ล่วงหน้า จากคลังข้อมูล Sentinel-1 GRD ทั้งหมดที่เรียกว่า ImageCollection ใน Google Earth Engine จะถูกกรองตามโหมดเครื่องมือ โพลาไรเซชัน ทิศทางผ่าน ความละเอียดเชิงพื้นที่

และถูกตัดให้เหลือตามขอบเขตของพื้นที่ที่สนใจ จากนั้น ImageCollection ที่กรองแล้วจะถูกย่อให้เหลือช่วงเวลาที่เลือก (ก่อนและหลังเหตุการณ์น้ำท่วม)

3. การตรวจจับการเปลี่ยนแปลง ด้วยการแบ่งรูปภาพด้วยวิธีโมเสก (Image Tiling หรือ Image Mosaicking) เป็นกระบวนการที่ใช้ในการจัดการกับภาพถ่ายที่มีขนาดใหญ่ (เช่น ภาพจากดาวเทียม หรือภาพแผนที่) โดยแบ่งออกเป็นชิ้นเล็ก ๆ ที่มีขนาดเท่ากัน เรียกว่า ไทล์ (Tile) เพื่อให้สามารถประมวลผลได้ง่ายขึ้น รวดเร็วขึ้น หรือเพื่อใช้ในงานวิเคราะห์เชิงพื้นที่ หลังน้ำท่วมจะถูกแบ่งด้วยโมเสกก่อนน้ำท่วม ส่งผลให้ได้เลเยอร์แรสเตอร์ที่แสดงระดับการเปลี่ยนแปลงต่อพิกเซล ค่าสูง (พิกเซลสว่าง) แสดงถึงการเปลี่ยนแปลงสูง ค่าต่ำ (พิกเซลมืด) แสดงถึงการเปลี่ยนแปลงเพียงเล็กน้อย ค่าติดลบ หมายถึงพื้นที่ที่มีระดับน้ำลดลง เกณฑ์ที่กำหนดไว้ล่วงหน้าที่ 1.25 จะถูกนำไปใช้ โดยกำหนด 1 ให้กับค่าทั้งหมดที่มากกว่า 1.25 และ 0 ให้กับค่าทั้งหมดที่น้อยกว่า 1.25 เลเยอร์แรสเตอร์ไบนารีที่สร้างขึ้นโดยกระบวนการนี้จะแสดงขอบเขตน้ำท่วมที่อาจเกิดขึ้น

4. ปรับปรุงเลเยอร์ระดับน้ำท่วม มีการใช้ชุดข้อมูลเพิ่มเติมหลายชุดเพื่อปรับค่าภายในเลเยอร์ระดับน้ำท่วม โดยใช้แบบจำลองระดับความสูงแบบดิจิทัล (WWF HydroSHEDS) ซึ่งอิงตามข้อมูล SRTM และมีความละเอียดเชิงพื้นที่ประมาณ 92.6 เมตร นอกจากนี้ ยังมีการประเมินการเชื่อมต่อของพิกเซลน้ำท่วมเพื่อขจัดพิกเซลที่เชื่อมต่อกับพิกเซลพื้นที่เพื่อนบ้าน 8 ทิศทางรอบพิกเซลเป้าหมาย การดำเนินการนี้จะช่วยลดสัญญาณรบกวนของผลิตภัณฑ์ระดับน้ำท่วม

5. การคำนวณพื้นที่น้ำท่วม ในการคำนวณพื้นที่น้ำท่วม จะต้องสร้างเลเยอร์แรสเตอร์ใหม่โดยคำนวณพื้นที่เป็นตารางเมตรสำหรับแต่ละพิกเซล เมื่อรวมพิกเซลทั้งหมดแล้ว จะได้ข้อมูลพื้นที่น้ำท่วม ดังแสดงตัวอย่างในภาพที่ 4



ภาพที่ 4 ข้อมูล Flooded Area ของพื้นที่ศึกษา จังหวัดเชียงใหม่ ปี พ.ศ. 2565 จากข้อมูลภาพถ่ายดาวเทียมของ • ชุดข้อมูล JRC Global Surface Water Mapping Layers

3. แบบจำลองการเรียนรู้ของเครื่องจักรสำหรับคาดการณ์พื้นที่น้ำท่วม

การออกแบบชุดข้อมูลพื้นที่น้ำท่วม เพื่อใช้สำหรับการสร้างแบบจำลอง สำหรับการคาดการณ์พื้นที่น้ำท่วม ซึ่งได้กำหนดพื้นที่ที่ทำการศึกษาเป็นพื้นที่เขตเมืองของ จังหวัดเชียงใหม่ มีตำแหน่ง เลขละติจูด (Latitude) และเลขลองจิจูด (Longitude) ที่ 18° 36' 0" N, 98° 57' 0" E และ

19° 0' 0" N, 99° 9' 0" E ตามลำดับ เพื่อให้การวิเคราะห์ข้อมูลมีความละเอียดและสามารถระบุลักษณะเฉพาะของพื้นที่ได้อย่างแม่นยำ จึงจัดเป็นโครงข่ายแบบกริดขนาด 8 แถว (แนวนอน) และ 4 หลัก (แนวตั้ง) รวมหน่วยย่อยทั้งหมด 32 เซลล์ ซึ่งมีวัตถุประสงค์หลักเพื่อเพิ่มความสามารถในการประมวลผลข้อมูลเชิงพื้นที่ (spatial analysis) และลดความคลาดเคลื่อนของแบบจำลอง การแบ่งพื้นที่ในลักษณะนี้ยังช่วยให้สามารถตรวจสอบความแม่นยำเชิงพื้นที่ (spatial accuracy) ของผลลัพธ์จากแบบจำลองในแต่ละเซลล์ได้โดยตรง การแบ่งพื้นที่ออกเป็นเซลล์ย่อยช่วยให้สามารถจับคู่ข้อมูลจากภาพถ่ายดาวเทียมกับตำแหน่งทางภูมิศาสตร์ที่ชัดเจน และสามารถนำไปใช้ในการฝึกแบบจำลอง Machine Learning ได้อย่างมีประสิทธิภาพมากยิ่งขึ้น โดยเฉพาะอย่างยิ่งในงานคาดการณ์พื้นที่น้ำท่วมที่ต้องการข้อมูลที่มีความละเอียดเชิงพื้นที่ระดับท้องถิ่น การจัดเก็บข้อมูลในรูปแบบตาราง (tabular data) โดยแต่ละแถว (row) แทนข้อมูลของหนึ่งเซลล์ในหนึ่งเดือน และแต่ละคอลัมน์ (column) แทนตัวแปรต้นและตัวแปรตาม ซึ่งตัวแปรต้น ประกอบด้วย Water balance Evapotranspiration Landcover Rainfall และ NDVI ตัวแปรตาม คือ พื้นที่น้ำท่วม ช่วงเวลาเก็บข้อมูลรายเดือน ตั้งแต่ปี 2019 ถึง 2022 (48 เดือน) ขนาดของ Dataset มีจำนวนข้อมูลทั้งหมด 32 เซลล์ × 48 เดือน = 1,536 แถว

การเตรียมข้อมูล ประกอบด้วย 4 ขั้นตอน ได้แก่ ขั้นตอนแรก การตรวจสอบและกำหนดค่าให้กับข้อมูลที่หาย ด้วยการเติมค่าเฉลี่ย ขั้นตอนที่ 2 แปลงข้อมูลตัวอักษรเป็นรหัสตัวเลขด้วยวิธี One-Hot Encoding ขั้นตอนที่ 3 การ normalize ข้อมูล ปรับสเกลของตัวแปรต้นให้อยู่ในช่วงเดียวกัน (เช่น 0 ถึง 1) เพื่อให้โมเดลเรียนรู้ได้ดีขึ้น และสุดท้าย แบ่งข้อมูลเป็นชุดฝึก (training set) และชุดทดสอบ (test set) เช่น 80% สำหรับ ใช้เทคนิค K-Fold Cross-Validation (5-Fold) เพื่อฝึกและเลือกแบบจำลอง และ กำหนด 20% สำหรับชุดทดสอบ (testing)

การเลือกโมเดล การศึกษาวิธีการจำลองตามแบบจำลองทั้ง 7 วิธีสำหรับการพยากรณ์พื้นที่น้ำท่วม ได้แก่ Linear Regression (LR), Support Vector Regression (SVR), Decision Tree Regression (DTR), Random Forest Regression (RFR), Gradient Boosting Regression (GBR), XGBoost Regression (XGBR) และ LightGBM Regression (LGBMR)

4. การวิเคราะห์ผลการทดลอง

ในส่วนนี้จะศึกษาประสิทธิภาพของแบบจำลองแต่ละวิธีสำหรับคาดการณ์พื้นที่น้ำท่วมแต่ละวิธี โดยการทดลองแบ่งออกเป็น 2 ส่วน ส่วนแรก แสดงการเลือกพารามิเตอร์ที่เหมาะสมที่สุดของการเรียนรู้ของวิธี Random Forest XGBoost Regressor และ LightGBM Regressor ส่วนที่ 2 นำเสนอการเปรียบเทียบประสิทธิภาพในการคาดการณ์พื้นที่น้ำท่วมของวิธีการเรียนรู้ 7 วิธี การประเมิน

โมเดล ใช้ Root Mean Squared Error (RMSE) และ R-squared (R^2) สำหรับ RMSE เป็นค่าที่ใช้วัดความแตกต่างระหว่างค่าที่แบบจำลองทำนายกับค่าจริง สำหรับค่า R-squared บอกถึงความสัมพันธ์ของตัวแปรในโมเดล เป็นค่าที่ใช้วัดสัดส่วนของความแปรปรวน (variance) ในตัวแปรเป้าหมาย

พื้นที่ศึกษาครบคลุมพื้นที่เมืองของจังหวัดเชียงใหม่ ซึ่งเป็นศูนย์กลางเศรษฐกิจและวัฒนธรรมของภาคเหนือ และบางส่วนของจังหวัดลำพูนที่ติดกับจ.เชียงใหม่ มีแม่น้ำปิงไหลผ่านพื้นที่ ภูมิประเทศมีลักษณะผสมผสานระหว่างภูมิประเทศที่เป็นภูเขาและหุบเขา พื้นที่เกษตรกรรม และพื้นที่เมือง มีตำแหน่ง ละติจูด: 18.6°N ถึง 19.0°N ลองจิจูด: 98.95°E ถึง 99.15°E ขนาดความกว้าง 21.1 กม. (ทิศตะวันออก-ตะวันตก) ความยาว 42.75 กม. (ทิศเหนือ-ใต้) คิดเป็นพื้นที่ 902.025 ตร.กม. พื้นที่ทั้งหมดถูกแบ่งออกเป็นเซลล์ย่อย 8 แถว (แนวนอน) และ 4 หลัก (แนวตั้ง) รวมทั้งหมด 32 เซลล์ แต่ละเซลล์มีขนาด 5.275 กม.X 5.34 กม. คิดเป็น 28.17 ตร.กม.

4.1 การเลือกพารามิเตอร์ที่เหมาะสมที่สุดของ Random Forest XGBoost Regressor และ LightGBM Regressor

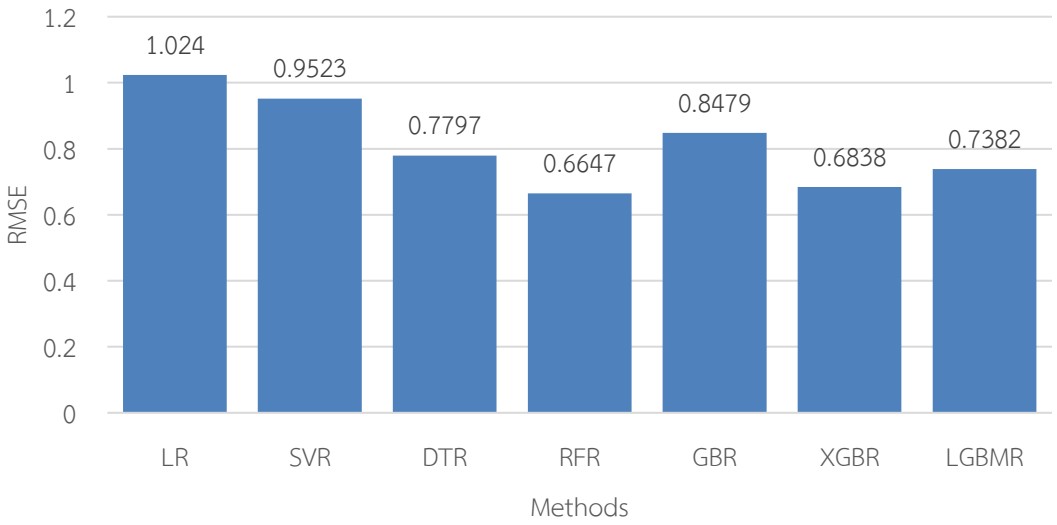
การทดลองเพื่อระบุพารามิเตอร์ที่เหมาะสมที่สุด สำหรับแบบจำลองการเรียนรู้ของเครื่องแต่ละวิธี ได้แก่ Random Forest XGBoost และ LightGBM กำหนดชุดตัวแปรสำหรับทดสอบดังนี้

1) วิธี Random Forest กำหนด จำนวนต้นไม้: [50, 100, 200] ความลึกของต้นไม้ [None, 10, 20] จำนวนตัวอย่างขั้นต่ำในการแบ่งโหนด [2, 5, 10] และ จำนวนตัวอย่างขั้นต่ำในโหนดใบไม้ [1, 2, 4]
2) วิธี XGBoost กำหนด จำนวนต้นไม้: [50, 100, 200], ความลึกของต้นไม้: [3, 4, 5], อัตราการเรียนรู้: [0.01, 0.1, 0.2], gamma: [0, 0.1, 0.2] สัดส่วนของข้อมูลที่จะใช้ในการฝึกฝนแต่ละต้นไม้: [0.7, 0.8, 0.9] การสุ่มตัวอย่างคอลัมน์ต่อหนึ่งต้นไม้: [0.7, 0.8, 0.9] พารามิเตอร์ของ L1 regularization (ค่า α): [0, 0.1, 1] พารามิเตอร์ของ L2 regularization (ค่า λ): [0, 0.1, 1]
3) วิธี LightGBM กำหนดตัวแปรเช่นเดียวกับวิธี XGBoost โดยเพิ่ม จำนวน Leaf Node สูงสุดในแต่ละต้นไม้: [20, 30, 40] และจำนวนข้อมูลขั้นต่ำที่จำเป็นในการสร้าง Leaf Node: [20, 30]

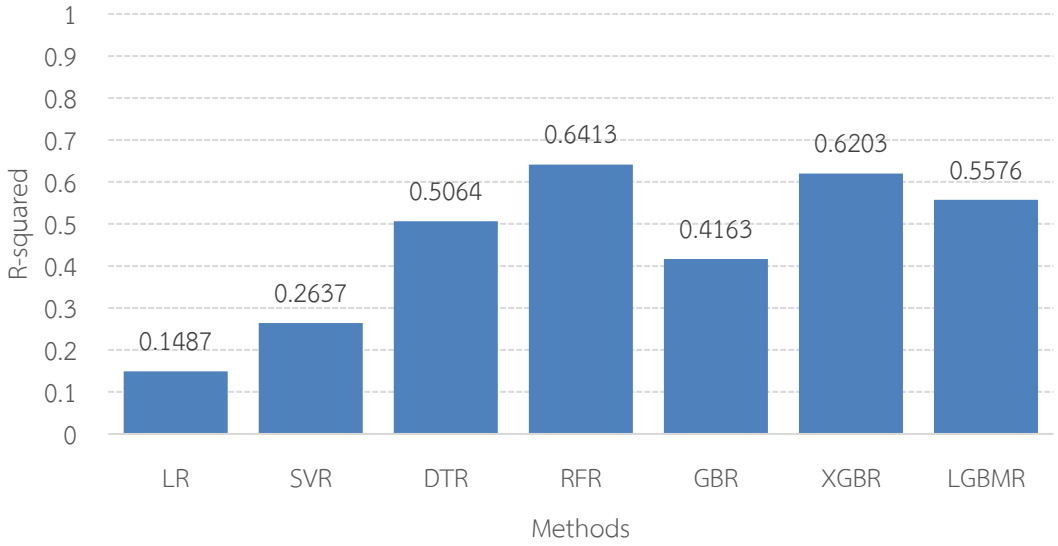
ผลการคาดการณ์พื้นที่น้ำท่วมของแบบจำลองจากการทดสอบชุดของตัวแปรต่างๆ ให้ผลความถูกต้องสูงสุดเมื่อกำหนดค่าตัวแปรของแต่ละวิธีดังต่อไปนี้ วิธี Random Forest กำหนดให้ความลึกของต้นไม้ = 20 จำนวนตัวอย่างขั้นต่ำในแต่ละใบ = 2 จำนวนตัวอย่างขั้นต่ำในการแบ่งโหนด = 2 จำนวนต้นไม้ = 100 กำหนดค่าพารามิเตอร์ของวิธี XGBoost ดังต่อไปนี้ จำนวนต้นไม้ = 200 ความลึกสูงสุด = 5 สัดส่วนคุณลักษณะที่ใช้ต่อต้นไม้ = 0.7 แกมมา = 0 อัตราการเรียนรู้ = 0.1 ค่าลดความซับซ้อน L1 = 1 ค่าลดความซับซ้อน L2 = 0.1 สัดส่วนของตัวอย่างที่จะใช้ในการสร้างแต่ละต้นไม้ = 0.9 สำหรับวิธี LightGBM จากผลการทดลองหาค่าตัวแปรที่เหมาะสมด้วยวิธี GridSearch กำหนดค่าพารามิเตอร์ต่างๆ ดังนี้ จำนวนต้นไม้ = 200 ความลึกสูงสุด = 5 สัดส่วนคุณลักษณะ = 0.7 อัตราการเรียนรู้ = 0.2 จำนวนตัวอย่างขั้นต่ำที่จำเป็นต้องมีในโหนดลูก = 20 จำนวนใบสูงสุดในแต่ละต้นไม้ ค่าลดความซับซ้อน L1 = 0.1 ค่าลดความซับซ้อน L2 = 1 สัดส่วนของตัวอย่างที่จะใช้ในการสร้างแต่ละต้นไม้ = 0.7 การกำหนดพารามิเตอร์ข้างต้นได้ถูกกำหนดให้กับทั้ง 3 วิธีที่กล่าวมาเพื่อใช้สำหรับคาดการณ์พื้นที่น้ำท่วม

4.2 การเปรียบเทียบประสิทธิภาพในการคาดการณ์พื้นที่น้ำท่วมของวิธีการเรียนรู้ 7 วิธี

การแสดงประสิทธิภาพการคาดการณ์พื้นที่น้ำท่วม ได้กำหนดค่าพารามิเตอร์ของวิธีดังกล่าวตามผลการทดลองในหัวข้อที่ 2 ในส่วนนี้จะแสดงการเปรียบเทียบความถูกต้องในการคาดการณ์พื้นที่น้ำท่วมของวิธีในกลุ่ม Regression ได้แก่ LR, SVR, DTR, RF, GBR, XGBR และ LGBMR โดยวัดค่าความถูกต้องด้วยวิธี RMSE และ R-Square ผลลัพธ์แสดงไว้ในภาพที่ 7 และ ภาพที่ 8 ตามลำดับ

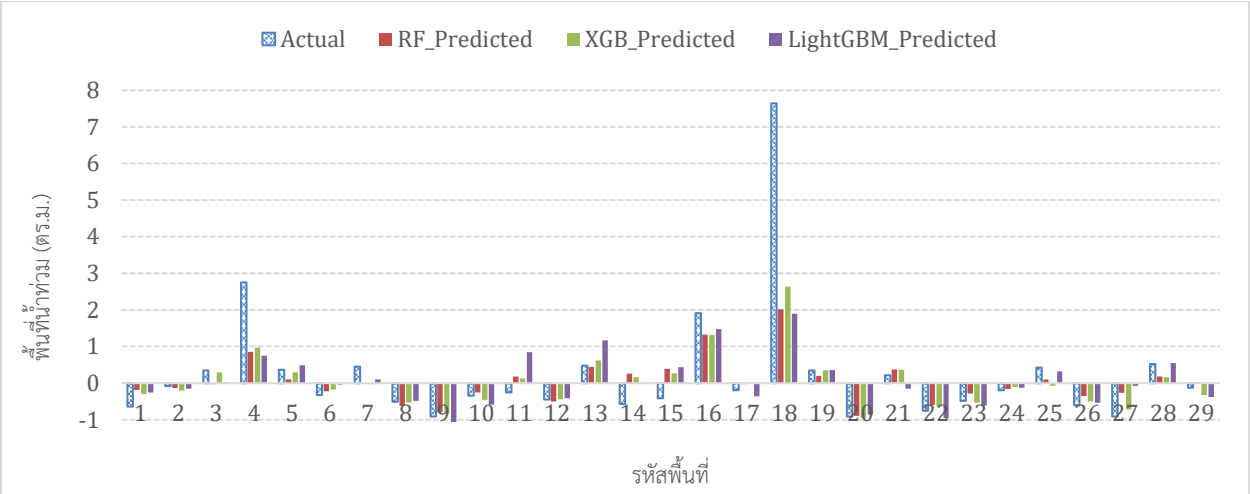


ภาพที่ 7 ประสิทธิภาพการคาดการณ์พื้นที่น้ำท่วมของแบบจำลองวิธีต่างๆ แสดงด้วยค่า RMSE

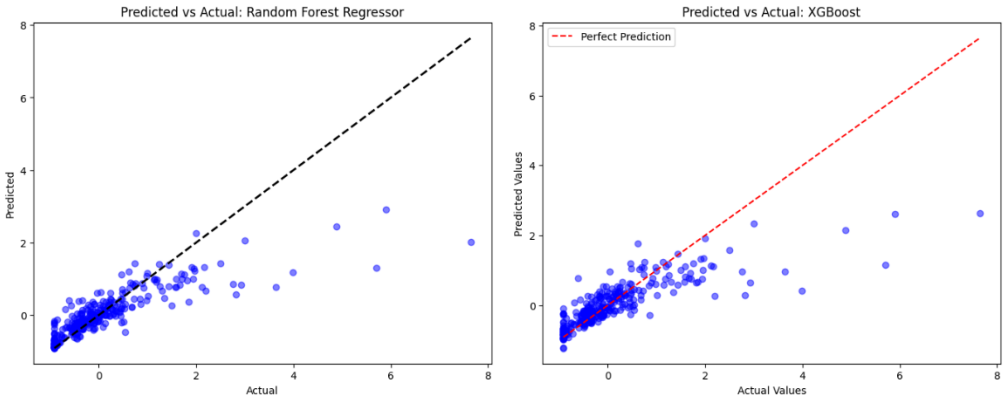


ภาพที่ 8 ประสิทธิภาพการคาดการณ์พื้นที่น้ำท่วมของแบบจำลองวิธีต่างๆ แสดงด้วยค่า R-Square

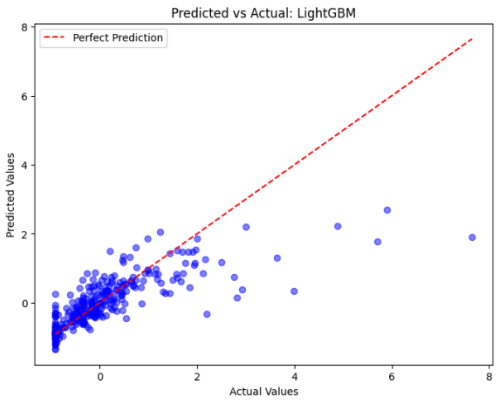
จากค่า RMSE และค่า R-Square ในภาพที่ 7 และ ภาพที่ 8 พบว่าวิธี Random Forest วิธี XGBoost และ LightGBM ให้คะแนนที่ดีที่สุด 3 อันดับแรก วิธี Random Forest มีค่า RMSE และค่า R-Square ดีกว่าวิธี XGBoost และ LightGBM คิดเป็นค่าเฉลี่ยร้อยละ 3 และ ร้อยละ 12 ตามลำดับ ตัวอย่างผลการคาดการณ์พื้นที่น้ำท่วมของทั้ง 3 วิธีเปรียบเทียบกับค่าพื้นที่น้ำท่วมจริง แสดงในภาพที่ 9 โดยทั่วไปพื่นที่น้ำท่วมมีขนาดน้อยกว่า 2 ตร.ม. ซึ่งทั้ง 3 วิธีสามารถประมาณพื้นที่ได้ใกล้เคียงกับค่าจริง ยกเว้นพื้นที่รหัส 4 และรหัส 18 มีพื้นที่น้ำท่วมบริเวณกว้าง ทั้ง 3 วิธีคาดการณ์พื้นที่น้ำท่วมแตกต่างจากค่าจริงไปมาก จากชุดข้อมูลทดสอบจำนวน 308 พื้นที่ ได้แสดงการกระจายผลการประมาณการพื้นที่น้ำท่วมของทั้ง 3 วิธี เปรียบเทียบกับค่าจริง ดังแสดงในภาพที่ 10



ภาพที่ 9 เปรียบเทียบตัวอย่างผลการคาดการณ์พื้นที่น้ำท่วมจำนวน 30 ตัวอย่างของวิธี Random Forest XGBoost และ LightGBM



a) วิธี Random Forest b) วิธี XGBoost



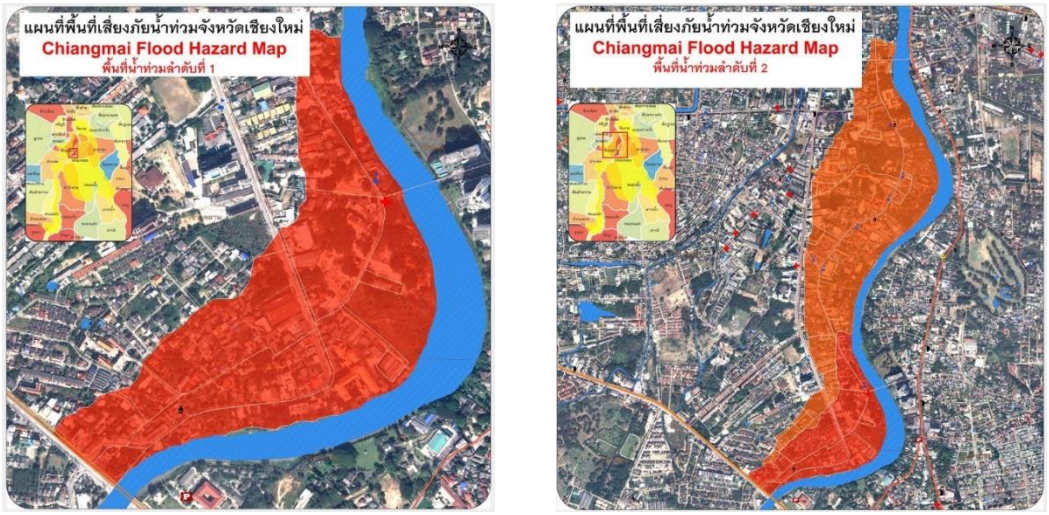
c) วิธี LightGBM

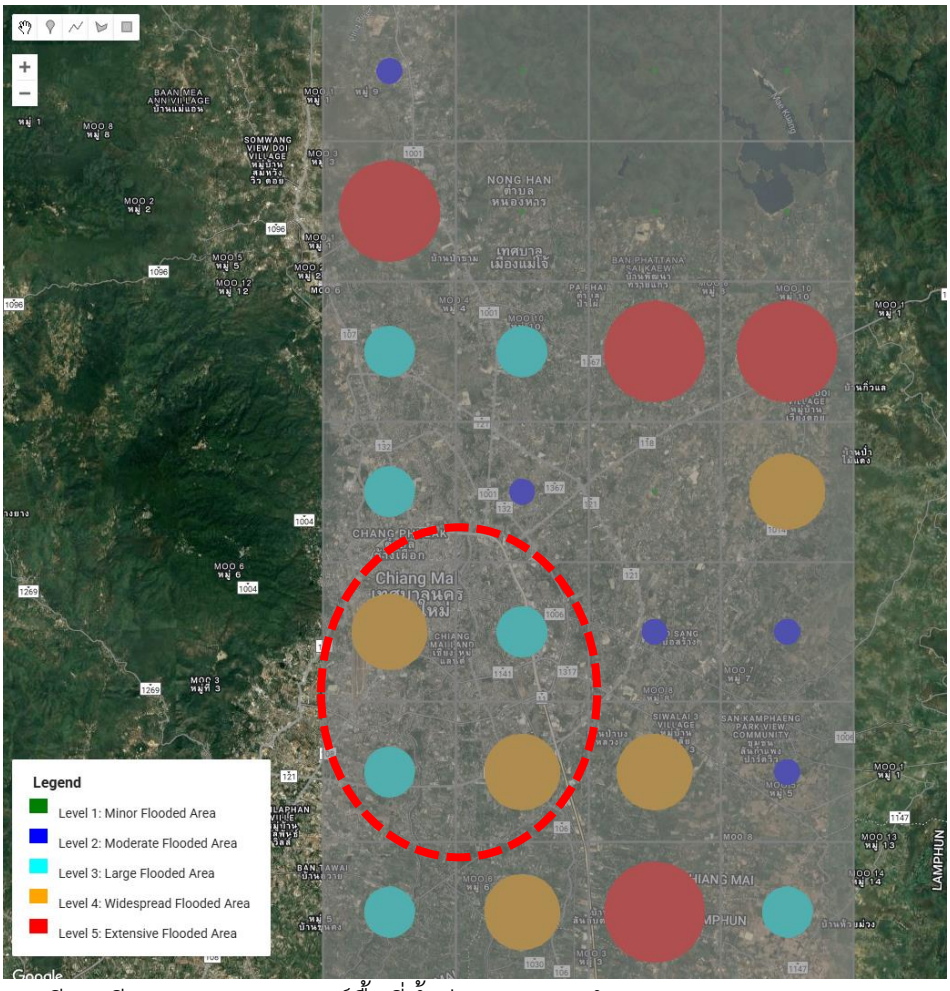
ภาพที่ 10 การเปรียบเทียบการคาดการณ์พื้นที่น้ำท่วมและข้อมูลจริงของวิธี a) Random Forest b) XGBoost และ c) วิธี LightGBM

จากภาพที่ 10 แสดงแกน X (แนวนอน) "Actual" แสดงถึงค่าจริงของตัวแปรเป้าหมาย แกน Y (แนวตั้ง) "Predicted" แสดงถึงค่าที่โมเดลทำนาย เส้นทแยงมุม: เส้นสีดำประ (dashed line) เป็นเส้นอ้างอิงที่แสดงถึงค่าที่ทำนายเท่ากับค่าจริง (y = x) จะเห็นได้ว่าจุดข้อมูลส่วนใหญ่กระจุกตัวอยู่ใกล้เส้นทแยงมุม ซึ่งแสดงว่าโมเดล Random Forest Regressor มีความแม่นยำในการทำนายที่ดี การที่จุดข้อมูลกระจุกตัวใกล้เส้นทแยงมุมมากกว่าอีก 2 วิธี วิธี Random Forest และ วิธี XGBoost มีความสามารถในการจัดการกับข้อมูลที่มีความสัมพันธ์ไม่เชิงเส้น (Non-linear Relationships) แต่วิธี Random Forest สามารถคาดการณ์พื้นที่น้ำท่วมนี้ได้ดีกว่าวิธี XGBoost เนื่องจากชุดข้อมูลพื้นที่น้ำท่วมเกี่ยวกับปัจจัยของการเกิดน้ำท่วม เช่น อากาศ ฝน และปัจจัยทางภูมิศาสตร์ เป็นต้น ซึ่งทำให้อาจมีความแปรปรวนของข้อมูล ทำให้ชุดข้อมูลมีความซับซ้อนและกระจายตัวสูง วิธี Random Forest เป็นโมเดลที่ใช้การรวมหลาย ๆ ต้นไม้ตัดสิน (Ensemble Learning) โดยการสร้างต้นไม้ตัดสินหลาย ๆ ต้นจากข้อมูลที่สุ่ม (Bootstrap Sampling) การเลือกแบบสุ่มในแต่ละต้นไม้ ช่วยให้โมเดลสามารถหลีกเลี่ยงการเรียนรู้จาก ข้อมูลที่ไม่สำคัญหรือข้อมูล noise ซึ่งจะช่วยลดความเสี่ยงจากการ overfitting และมีความสามารถในการปรับตัวที่ดีต่อข้อมูลที่มีความไม่สมบูรณ์หรือข้อมูล noise มีความทนทานต่อ outliers คือข้อมูลที่มีความผิดปกติ ในทางตรงกันข้าม XGBoost อาจทำให้เกิด overfitting ได้ง่ายกว่า โดยเฉพาะในกรณีที่มีการปรับค่าพารามิเตอร์มากเกินไป เช่น ค่า อัตราการเรียนรู้ที่ต่ำหรือ ความลึกของต้นไม้ หรือ จำนวนข้อมูลขั้นต่ำที่โหนด ซึ่งทำให้โมเดลสามารถเรียนรู้จากข้อมูล noise และข้อมูลที่ผิดปกติได้มากขึ้น XGBoost เป็นวิธีที่มีความแม่นยำสูง เนื่องจากใช้ Gradient Boosting ซึ่งสร้างโมเดลใหม่โดยการปรับปรุงข้อผิดพลาดจากโมเดลที่ผ่านมาทีละน้อย แต่การที่มันสร้างโมเดลทีละต้น (sequential learning) อาจทำให้โมเดลมีความเสี่ยงที่จะไม่สามารถจัดการกับความแปรปรวนในข้อมูลได้ดีเท่าที่ควร งานวิจัยนี้ให้ผลสอดคล้องกับการศึกษาของ งานวิจัยนี้ (Islam et al., 2025) ได้ให้ภาพรวมที่ครอบคลุมของการใช้วิธีการต่างๆ ในการทำแผนที่ความเสี่ยงต่ออุทกภัยในเขตเมือง, และความสำคัญของการบูรณาการเทคโนโลยีการสำรวจระยะไกล อาทิ ข้อมูลดาวเทียมที่มีความละเอียดสูง เช่น Landsat, Sentinel และแบบจำลองความสูงดิจิทัล (DEM) มีความสำคัญอย่างยิ่งต่อการทำแผนที่น้ำท่วมที่มีประสิทธิภาพ การใช้เทคนิคการเรียนรู้ของเครื่อง ได้แก่ Support Vector Machines, Random Forests และ Artificial Neural Networks เป็นต้น และการสร้างแบบจำลองทางอุทกวิทยาโดยใช้ปัจจัยสำคัญของน้ำท่วมในเมืองจำนวน 35 ปัจจัย อาทิ ได้รับอิทธิพลจากการใช้ที่ดิน รูปแบบของฝน ความใกล้ชิดกับแหล่งน้ำ และคุณสมบัติของดิน ผล

การศึกษาพบว่า แบบจำลอง Random Forest ถือเป็นหนึ่งในแบบจำลองการเรียนรู้ของเครื่องจักรที่ดีที่สุดสำหรับการคาดการณ์ความเสี่ยงต่อน้ำท่วมในเมือง โดยมีความแม่นยำประมาณ 85% แบบจำลองนี้สามารถจัดการความสัมพันธ์ที่ซับซ้อนและไม่เชิงเส้นระหว่างการเกิดน้ำท่วมและปัจจัยกำหนดเงื่อนไขได้อย่างมีประสิทธิภาพ

ข้อมูลน้ำท่วมจังหวัดเชียงใหม่ในปี พ.ศ. 2565 ข้อมูลจาก สำนักงานชลประทานที่ 1 กรมชลประทาน จังหวัดเชียงใหม่ รายงานสถานการณ์น้ำในแม่น้ำปิง ช่วงวันที่ 3 ถึง 5 ตุลาคม พ.ศ. 2565 สถานการณ์น้ำท่วมเขตเทศบาลนครเชียงใหม่ (กรุงเทพธุรกิจ, 2565) อาทิ บริเวณชุมชนเชียงใหม่แลนด์ และชุมชนเวียงทอง ตลาดไนท์บาซาร์ ถนนช้างคลาน ต.ป่าแดด ต.ช้างคลาน ต.หนองหอย เชิงสะพานนวรัตน์ และย่านถนนเจริญประเทศ โดยมีพื้นที่น้ำท่วมหนักบริเวณโซนเสี่ยงภัยที่ 1 และ 2 จากทั้งหมด 7 โซนตามแผนที่เสี่ยงภัยน้ำท่วมจังหวัดเชียงใหม่ (ศูนย์วิชาการสนับสนุนการบริหารจัดการน้ำ มหาวิทยาลัยเชียงใหม่, 2567) จากผลการคาดการณ์พื้นที่น้ำท่วมจากแบบจำลอง RFR เดือนตุลาคม พ.ศ.2565 ของงานวิจัยนี้ ได้นำเสนอเปรียบเทียบกับข้อมูลพื้นที่น้ำท่วมจริงตามแหล่งข้อมูลข้างต้น และแผนที่เสี่ยงภัยน้ำท่วมจังหวัดเชียงใหม่ โซนเสี่ยงภัยที่ 1 และ 2 ดังแสดงในภาพที่ 11 พบว่าแบบจำลอง RFR สามารถคาดการณ์พื้นที่น้ำท่วมได้สอดคล้องกับเหตุการณ์จริง โดยระดับ (Levels) แสดงพื้นที่ที่จะเกิดน้ำท่วม แบ่งเป็น 5 ระดับประกอบด้วย พื้นที่ขนาดเล็ก ไปจนถึงน้ำท่วมเป็นวงกว้าง จากระดับ 1 จนถึงระดับ 5 ตามลำดับ





ภาพที่ 11 เปรียบเทียบผลการคาดการณ์พื้นที่น้ำท่วมจากแบบจำลอง Random Forest Regression ในเดือนตุลาคม พ.ศ.2565 กับบริเวณจังหวัดเชียงใหม่

จากภาพที่ 11 แสดงให้เห็นว่า แบบจำลอง RFR สามารถคาดการณ์พื้นที่น้ำท่วมในบริเวณ เขตเทศบาลนครเชียงใหม่ ถนนช้างคลาน ต.ป่าแดด ต.ช้างคลาน ต.หนองหอย เชิงสะพานนวรัตน์ และย่านถนนเจริญประเทศ และสามารถแสดงผลการคาดการณ์พื้นที่บริเวณน้ำท่วมเฉพาะของแต่ละพื้นที่ย่อยจำนวน 32 พื้นที่ ซึ่งให้ข้อมูลที่เฉพาะของแต่ละพื้นที่ได้ละเอียดกว่าเมื่อเปรียบเทียบกับแผนที่เสี่ยงภัยน้ำท่วมจังหวัดเชียงใหม่

5. สรุปผล

บทความวิจัยนี้ได้นำเสนอ การสร้างแบบจำลองการคาดการณ์พื้นที่น้ำท่วมจากข้อมูลภาพถ่ายดาวเทียม พื้นที่ศึกษาครอบคลุมพื้นที่เมืองของจังหวัดเชียงใหม่และบางส่วนของจังหวัดลำพูน คิดเป็นพื้นที่ 902.025 ตร.ม. โดยสกัดปัจจัยที่เกี่ยวข้อง 6 ปัจจัย ประกอบด้วย สมดุลน้ำ การคายระเหยของน้ำ พื้นผิวปกคลุมดิน ปริมาณน้ำฝน NDVI และ พื้นที่น้ำท่วม จากภาพถ่าย ดาวเทียม Sentinel ดาวเทียม MODIS และ ดาวเทียม Landsat ผ่าน Google Earth Engine จากนั้นทำการฝึกฝนแบบจำลองด้วยวิธีทาง Machine Learning ในกลุ่ม Regression 7 วิธี ได้แก่ LR, SVR, Decision Tree, Random Forest, Gradient Boosting, XGBoost และ LightGBM เพื่อหาแบบจำลองที่มีประสิทธิภาพสูงสุด พบว่า วิธี Random Forest เป็นวิธีที่มีประสิทธิภาพสูงสุด มีค่า RMSE และ R-Square ที่ 0.6647 ตร.ม. และ 0.6413 ตามลำดับ เนื่องจากคุณลักษณะเด่นของวิธี Random Forest ทำให้สามารถรับมือกับข้อมูลที่มีการกระจายตัวแปรปรวนได้ดีกว่าวิธีอื่นๆ

การใช้ภาพถ่ายดาวเทียมร่วมกับ Regression Learning เพื่อคาดการณ์พื้นที่น้ำท่วมมีจุดเด่นหลายประการ ได้แก่ 1) การครอบคลุมพื้นที่กว้างและต่อเนื่อง สามารถตรวจสอบข้อมูลในพื้นที่ที่เข้าถึงได้ยาก 2) ข้อมูลจากดาวเทียมเรดาร์ (SAR) สามารถใช้งานได้แม้ในสภาพอากาศที่มีเมฆมากหรือเวลากลางคืน ซึ่งเป็นข้อได้เปรียบสำคัญในช่วงที่เกิดน้ำท่วม 3) การลดต้นทุนและเวลาในการสำรวจ การใช้ภาพถ่ายดาวเทียมลดความจำเป็นในการส่งทีมสำรวจภาคสนาม ทำให้ประหยัดงบประมาณและลดความเสี่ยงในการเก็บข้อมูล ทำให้ช่วยลดเวลาในการรวบรวมข้อมูล 4) การสร้างแบบจำลองด้วยวิธี Regression นั้นเป็นวิธีที่เข้าใจง่ายและพัฒนาได้ง่าย มีประสิทธิภาพสูงและคำนวณได้เร็วสามารถใช้กับชุดข้อมูลขนาดใหญ่ได้โดยไม่ต้องการพลังประมวลผลสูง การใช้ภาพถ่ายดาวเทียมมีข้อจำกัด ประการแรก คือ ด้านช่วงเวลา ดาวเทียมแต่ละดวงมีวงโคจรและช่วงเวลาในการเก็บข้อมูลที่แตกต่างกัน ทำให้ไม่สามารถเก็บข้อมูลปัจจุบันได้ตลอดเวลา ประการที่ 2 ด้านข้อมูลเชิงลึก ภาพถ่ายดาวเทียมสามารถให้ข้อมูลเกี่ยวกับพื้นที่ที่ถูกน้ำท่วมได้ แต่ไม่สามารถให้ข้อมูลเชิงลึกเกี่ยวกับความลึกของน้ำ ความเร็วของกระแสน้ำ โดยตรงจากภาพได้

แนวทางการพัฒนาต่อยอดในอนาคต มุ่งเน้นไปที่การปรับปรุงประสิทธิภาพ โดยการใช้ Stacking Regression การเพิ่มความสามารถของแบบจำลอง รวมกับวิธี Deep Learning อาทิ LSTM เพื่อให้ได้แบบจำลองที่มีประสิทธิภาพยิ่งขึ้น นอกจากนี้ แนวทางสำหรับการพัฒนาต่อยอดสำหรับพัฒนาแบบจำลองคาดการณ์ความลึกของน้ำท่วมโดยใช้ข้อมูลจากพื้นที่น้ำท่วมที่คาดการณ์ได้เป็นหนึ่งในข้อมูลนำเข้า

เอกสารอ้างอิง

กรมป้องกันและบรรเทาสาธารณภัย. (2567). **รายงานสถานการณ์สาธารณภัย**. ค้นเมื่อ [14 เมษายน 2568], ค้นจาก https://backofficeminisite.disaster.go.th/apiv1/apps/minisite_directing/194/content/8728/download?filename=26c236095f0e9fc4a4e0af7edf6fa9f0.pdf

กรุงเทพธุรกิจ. (2565, 4 ตุลาคม). **อัปเดต "น้ำท่วมเชียงใหม่" ล่าสุด! เช็กจุดน้ำท่วมขัง-ปิดการจราจร**. ค้นเมื่อ [16 เมษายน 2568], ค้นจาก https://www.bangkokbiznews.com/news/news-update/1030426

ศูนย์วิชาการสนับสนุนการบริหารจัดการน้ำ มหาวิทยาลัยเชียงใหม่. (2567). **แผนที่เสี่ยงภัยน้ำท่วม (Flood Hazard Map)**. ค้นเมื่อ [16 เมษายน 2568], ค้นจาก https://watercenter.scmc.cmu.ac.th/cmflood/floodmap

Chitwatkulsiri, D., & Miyamoto, H. (2023). Real-Time Urban Flood Forecasting Systems for Southeast Asia—A Review of Present Modelling and Its Future Prospects. **Water**, 15(1), 178. https://doi.org/10.3390/w15010178

European Space Agency. (n.d.). **Sentinel-2 Level-2A products**. ค้นเมื่อ [19 พฤศจิกายน 2567], ค้นจาก https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... & Husak, G. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. **Scientific data**, *2*(1), 1-13.

Hakim DK, Gernowo R, and Nirwansyah AW (2024). Flood prediction with time series data mining: Systematic review. **Natural Hazards Research**, 4(2): 194-220. https://doi.org/10.1016/j.nhres.2023.10.001

Islam, T., Zeleke, E. B., Afroz, M., & Melesse, A. M. (2025). A Systematic Review of Urban Flood Susceptibility Mapping: Remote Sensing, Machine Learning, and Other Modeling **Approaches. Remote Sensing**, 17(3), 524. https://doi.org/10.3390/rs17030524

Lehner, B., Verdin, K., & Jarvis, A. (2008). **HydroSHEDS**. ค้นเมื่อ [19 พฤศจิกายน 2567], ค้นจาก https://www.hydrosheds.org/

Manikandan, P., Vivek, S. T. S., Thejaswi, M., Tejaswini, B., Manoj, S., & Reddy, P. T. (2024). Flood Risk Assessment System using Logistic Regression. **In 2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA)** (pp. 1145-1149). IEEE. https://doi.org/10.1109/ICECA63461.2024.10801008

NASA. (n.d.). **Moderate Resolution Imaging Spectroradiometer (MODIS)**. ค้นเมื่อ [19 พฤศจิกายน 2567], ค้นจาก https://lifehacker.ru/veb-sajt-nasa/.

Nhangumbe, M., Nascetti, A., & Ban, Y. (2023). Multi-Temporal Sentinel-1 SAR and Sentinel-2 MSI Data for Flood Mapping and Damage Assessment in Mozambique. ISPRS International **Journal of Geo-Information**, 12(2), 53. https://doi.org/10.3390/ijgi12020053

Tengtrairat, N., Woo, W. L., Parathai, P., Sundaranaga, C., Na Ayutthaya, T. K., & Rinchumphu, D. (2022). Impact of extreme class-imbalance on landslide-risk prediction and mitigation using two-stage deep neural network. **In 2022 14th International Conference on Signal Processing Systems (ICSPS)** (pp. 712–719). IEEE. https://doi.org/10.1109/ICSPS58776.2022.00131

Oki, T., & Kanae, S. (2006). Global hydrological cycles and world water resources. **Science**, 313(5790), 1068-1072.

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. **Nature**, 540(7633), 418–422. https://doi.org/10.1038/nature20584

Pham, H. V., Brakenridge, G. R., Anderson, E., & Kettner, A. J. (2021). Satellite-based flood detection and monitoring using multi-temporal Sentinel-1 SAR imagery. **Remote Sensing**, 13(3), 453. https://doi.org/10.3390/rs13030453

Sanderson, J., Tengtrairat, N., Woo, W. L., Mao, H., & Al-Nima, R. R. (2023). XFIMNet: an Explainable deep learning architecture for versatile flood inundation mapping with synthetic aperture radar and multi-spectral optical images. **International Journal of Remote Sensing**, 44(24), 7755–7789. https://doi.org/10.1080/01431161.2023.2288945

Sanderson, J., Mao, H., Tengtrairat, N., Al-Nima, R. R. O., & Woo, W. L. (2024). Explainable deep semantic segmentation for flood inundation mapping with class activation mapping techniques. **Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART)** (Vol. 3, pp. 1028–1035).

Singh, G., & Rawat, K. S. (2024). Mapping flooded areas utilizing Google Earth Engine and open SAR data: A comprehensive approach for disaster response. **Discov Geosci**, 2(5). https://doi.org/10.1007/s44288-024-00006-4

UN-SPIDER. (n.d.). **Step 10: Area calculation of flood extent**. จาก [19 พฤษจิกายน 2567] https://www.un-spider.org/advisory-support/recommended-practices/recommended-practice-google-earth-engine-flood-mapping/step-by-step#Step%2010:%20Area%20calculation%20of%20flood%20extent

Yang, X., Wang, N., Liang, Q., Chen, A., & Wu, Y. (2021). Impacts of Human Activities on the Variations in Terrestrial Water Storage of the Aral Sea Basin. **Remote Sensing**, 13(15), 2923. https://doi.org/10.3390/rs13152923.