การวิเคราะห์ปัจจัยกำหนดความเสี่ยงของโรคเลปโตสไปโรซิสในประเทศไทยโดยใช้การเรียนรู้ของเครื่อง

*Pikkanet Suttirat, Sudarat Chadsuthi, **Charin Modchang**, and Joacim Rocklöv*

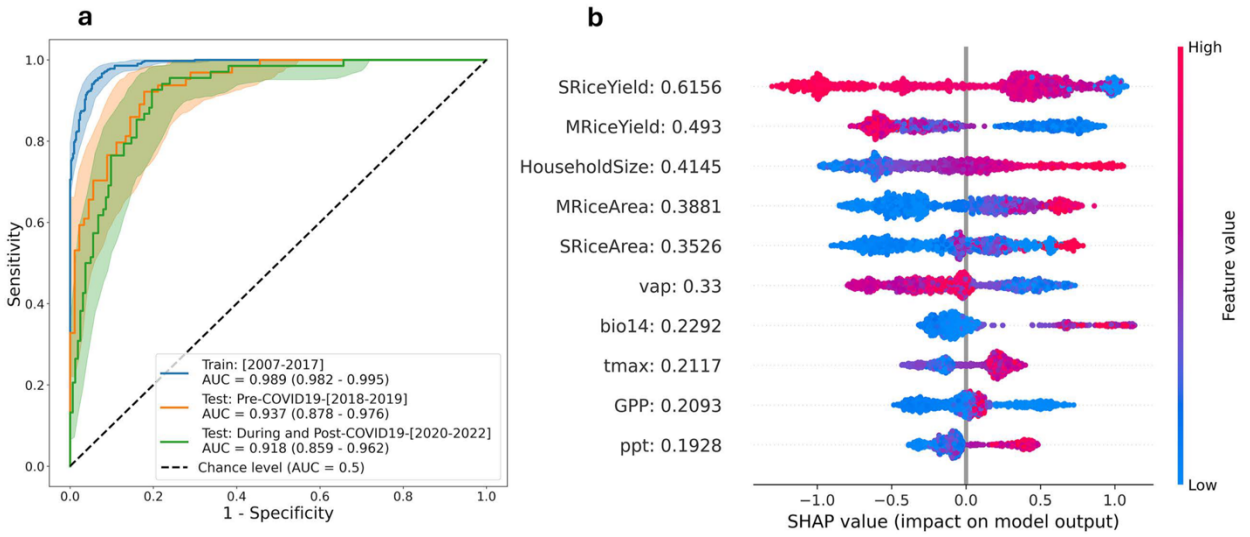
หลักการและเป้าหมาย: โรคเลปโตสไปโรซิสเป็นโรคติดเชื้อแบคทีเรียที่ร้ายแรงและเป็นภัยคุกคามสำคัญต่อสาธารณสุขของประเทศไทย โดยมีรายงานผู้ป่วยมากกว่า 51,000 รายระหว่างปี 2550-2565 เชื้อแบคทีเรียนี้เจริญเติบโตได้ดีในน้ำและดินที่ปนเปื้อน โดยเฉพาะในพื้นที่ทำนาข้าว และคนติดเชื้อจากการสัมผัสกับสิ่งแวดล้อมที่ปนเปื้อนเหล่านี้ เนื่องจากโรคนี้เกิดจากปฏิสัมพันธ์ที่ซับซ้อนระหว่างสภาพอากาศ วิธีการทำการเกษตร และสภาพเศรษฐกิจสังคม นักวิจัยจึงมุ่งพัฒนาแบบจำลองการเรียนรู้ของเครื่องที่สามารถทำนายว่าจังหวัดใดในประเทศไทยมีความเสี่ยงสูงต่อการระบาดของโรคเลปโตสไปโรซิส โดยวิเคราะห์ปัจจัยหลายอย่างเหล่านี้ร่วมกัน

สรุป: ทีมวิจัยสร้างแบบจำลองปัญญาประดิษฐ์โดยใช้เทคนิคที่เรียกว่า XGBoost เพื่อวิเคราะห์ข้อมูลเฝ้าระวังโรค 16 ปี ร่วมกับข้อมูลเกี่ยวกับปริมาณฝน อุณหภูมิ กิจกรรมการทำนาข้าว ขนาดครัวเรือน และตัวชี้วัดทางเศรษฐกิจในจังหวัดต่าง ๆ ของประเทศไทย พวกเขาฝึกแบบจำลองด้วยข้อมูลปี 2550-2560 เพื่อแยกแยะระหว่างจังหวัดที่มีความเสี่ยงสูงกับต่ำ จากนั้นทดสอบความแม่นยำด้วยข้อมูลปี 2561-2565 เพื่อทำความเข้าใจว่าปัจจัยใดสำคัญที่สุดในการทำนาย พวกเขาใช้วิธีการวิเคราะห์ที่เรียกว่า SHAP ซึ่งแสดงให้เห็นว่าแต่ละปัจจัยมีอิทธิพลต่อการตัดสินใจของแบบจำลองอย่างไร

ผลที่ได้: แบบจำลองสามารถทำนายความเสี่ยงของโรคเลปโตสไปโรซิสได้อย่างแม่นยำสูง (93.7% ก่อนการระบาดของโควิด-19 และ 95.1% หลังการระบาด) รูปแบบการผลิตข้าวกลายเป็นตัวทำนายที่สำคัญที่สุดของความเสี่ยงโรค โดยพื้นที่ที่ใช้วิธีการทำนาแบบดั้งเดิมที่ใช้เครื่องจักรน้อยแสดงความเสี่ยงที่สูงกว่า ขนาดครัวเรือนและปัจจัยภูมิอากาศบางอย่างเช่นความดันไอน้ำก็มีบทบาทสำคัญเช่นกัน ที่น่าสนใจคือ ประสิทธิภาพของแบบจำลองลดลงในช่วงการระบาดของโควิด-19 (2563-2564) ซึ่งน่าจะเกิดจากการเฝ้าระวังโรคที่หยุดชะงักและการรายงานที่ต่ำกว่าความเป็นจริงเมื่อระบบสาธารณสุขมีภาระหนัก งานวิจัยนี้ให้กรอบการทำงานที่มีคุณค่าสำหรับการพัฒนาระบบเตือนภัยล่วงหน้าที่สามารถช่วยหน่วยงานสาธารณสุขไทยมุ่งเป้าความพยายามในการป้องกันโรคไปยังชุมชนที่เปราะบางที่สุด

ทุนวิจัยและกิตติกรรมประกาศ: The Rockefeller Foundation under the Strengthening the Early Warning and Outbreak Detection Systems through Nation-wide Event-based and Syndromic Surveillance (STONES) project (Grant No. 2021PPI005).

เป้าหมาย SDGs ที่เกี่ยวข้อง: 3. การมีสุขภาพและความเป็นอยู่ที่ดี



สรุปภาพ: ตัวชี้วัดประสิทธิภาพของแบบจำลอง XGBoost และการวิเคราะห์ความสำคัญของปัจจัยด้วย SHAP **(a)** กราฟ ROC (Receiver Operating Characteristic) แสดงความสามารถในการแยกแยะของแบบจำลองในชุดข้อมูลต่าง ๆ พร้อมค่าพื้นที่ใต้กราฟ (AUC) ผลลัพธ์ได้มาจากพารามิเตอร์ที่เหมาะสมที่สุดซึ่งระบุผ่านการตรวจสอบไขว้ 5 ส่วน เส้นแนวทแยงแสดงการจำแนกแบบสุ่ม (AUC = 0.5) ในขณะที่ AUC = 1.0 บ่งชี้การทำนายที่สมบูรณ์แบบ พื้นที่แรเงาแสดงช่วงความเชื่อมั่น 95% (95% CIs) จากการทำ bootstrap 1,000 ครั้ง ค่า 95% CIs ของ AUC แสดงในวงเล็บ **(b)** การวิเคราะห์ SHAP (SHapley Additive exPlanations) ของปัจจัยทำนาย 10 อันดับแรก จัดอันดับตามค่าเฉลี่ยสัมบูรณ์ของ SHAP (ปี 2550-2565) ข้อมูลแต่ละจุดแสดงเป็นจุด โดยความเข้มของสีบ่งบอกขนาดของปัจจัย (สีแดง = สูง, สีน้ำเงิน = ต่ำ) ค่า SHAP บนแกน x แสดงการมีส่วนร่วมของแต่ละปัจจัยต่อความน่าจะเป็นที่ทำนาย โดยมีจุดกึ่งกลางที่ log odds = 0 สอดคล้องกับความน่าจะเป็นที่ทำนายเท่ากับ 0.5

งานวิจัยที่เกี่ยวข้อง:

Suttirat P, Chadsuthi S, **Modchang C**, and Rocklöv J. Unraveling the drivers of leptospirosis risk in Thailand using machine learning. *PLOS Neglected Tropical Diseases*. https://doi.org/10.1371/journal.pntd.0013618