

บทที่ 5

สรุปและอภิปรายผล

การวิจัยนี้เป็นการพัฒนาดัชนีวัดความสำคัญของตัวแปรและเกณฑ์จุดตัดในการคัดเลือกตัวแปรบนพื้นฐานวิธีกำลังสองน้อยที่สุดบางส่วน สำหรับการจำแนกประเภทข้อมูล โดยวิธีการคัดเลือกตัวแปรแบบใหม่ที่ได้จากดัชนีวัดความสำคัญของตัวแปร และเกณฑ์จุดตัดที่พัฒนาขึ้นนี้ได้แก่ วิธี VIP-BCT และวิธี VIIC-BCT วิธีดังกล่าวได้นำไปเปรียบเทียบกับวิธีการคัดเลือกตัวแปรแบบดั้งเดิม คือ วิธี VIP-1 ซึ่งเป็นการคัดเลือกตัวแปรที่ใช้ดัชนี VIP ที่มีเกณฑ์จุดตัด 1.0 ภายใต้สถานการณ์จำลอง 180 สถานการณ์ ($4 \times 3 \times 5 \times 3$) คือเปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูล 4 ระดับ (1%, 3%, 5% และ 10%) ความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้อง 3 ระดับ (1, 3 และ 5 หน่วย) แบบแผนความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้อง 5 แบบแผน (ไม่สัมพันธ์กัน (Σ_1) สัมพันธ์กันที่ระดับสัมพันธ์สหสัมพันธ์คงที่ 0.5 (Σ_2) และ 0.9 (Σ_3) สัมพันธ์กันที่ระดับสัมพันธ์สหสัมพันธ์ไม่คงที่ซึ่งกำหนดสหสัมพันธ์ระหว่างตัวแปรทำนายที่ i กับ j เท่ากับ $\alpha^{|i-j|}$ เมื่อ α เท่ากับ 0.5 (Σ_4) และ 0.9 (Σ_5)) และขนาดตัวอย่าง 3 ขนาด (40, 70 และ 100 ตัวอย่าง) และได้ นำวิธีการคัดเลือกตัวแปรไปคัดเลือกยีนที่เกี่ยวข้องกับการจำแนกประเภทเนื้อเยื่อมะเร็งลำไส้ใหญ่ และเนื้อเยื่อปกติ เพื่อจำแนกประเภทเนื้อเยื่อมะเร็งลำไส้ใหญ่ และเนื้อเยื่อปกติ จากผลการศึกษาสามารถสรุปผลได้ดังนี้

สรุปผลการวิจัย

1. ดัชนีวัดความสำคัญของตัวแปรทำนายที่พัฒนาขึ้นโดยอาศัยดัชนี VIP (Variable Influence on the Projection) ซึ่งมีการพัฒนาสูตรเพื่อให้เหมาะสมกับกรณีข้อมูลที่ตัวแปรตามเป็นตัวแปรเชิงคุณภาพ ด้วยการใช้ค่า ACR_o สำหรับวัดความสอดคล้องขององค์ประกอบ α แทนการใช้ค่า $R^2(Y, t_o)$ ในดัชนี VIP และเรียกดัชนีที่พัฒนาขึ้นว่า VIIC (Variable Importance Index for Classification) โดยดัชนี VIIC ของตัวแปรทำนายที่ j คำนวณดังนี้

$$VIIC_j = \sum_{\alpha=1}^h ACR_{\alpha} \cdot \left(w_{\alpha j} / \|w_{\alpha}\| \right)^2 \quad (5.1)$$

โดยที่ ACR_{α} แทน Adjusted Count R^2 ขององค์ประกอบ t_{α}

$w_{\alpha j}$ แทน น้ำหนักของตัวแปรทำนายที่ j ในองค์ประกอบ t_{α}

$\|w_{\alpha}\|$ แทน ขนาดของเวกเตอร์น้ำหนักขององค์ประกอบ t_{α}

สำหรับการพัฒนาเกณฑ์จุดตัดนั้นได้อาศัยแนวทางการตรวจหาค่าผิดปกติของดัชนีวัดความสำคัญของตัวแปรบวกรด้วยวิธีแผนภาพกล่อง หรือ Boxplot และเรียกเกณฑ์จุดตัดนี้ว่า เกณฑ์จุดตัด Boxplot (Boxplot Cutoff Threshold) หรือเกณฑ์จุดตัด BCT ซึ่งคำนวณดังนี้

$$BCT = Q_3 + 1.5 \times IQR \quad (5.2)$$

โดยที่ BCT แทน เกณฑ์จุดตัด BCT ซึ่งใช้สัญลักษณ์ BCT_{VIIC} และ BCT_{VIP} สำหรับเกณฑ์จุดตัดของ VIIC และ VIP ตามลำดับ

Q_1 และ Q_3 แทน ควอร์ไทล์ที่ 1 และ 3 ของค่าดัชนีของตัวแปรบวกร ตามลำดับ

IQR แทน พิสัยระหว่างควอร์ไทล์ (Interquartile Range: $IQR = Q_3 - Q_1$)

จากดัชนีวัดความสำคัญของตัวแปร และเกณฑ์จุดตัดสำหรับการคัดเลือกตัวแปรที่พัฒนาขึ้น นำมาใช้ในการคัดเลือกตัวแปรที่ประกอบด้วยดัชนีวัดความสำคัญของตัวแปรทำนาย และเกณฑ์จุดตัด ได้ 2 วิธี ดังนี้

1) วิธี VIP-BCT เป็นการคัดเลือกตัวแปรโดยใช้ดัชนี VIP และใช้เกณฑ์จุดตัด BCT_{VIP}

2) วิธี VIIC-BCT เป็นการคัดเลือกตัวแปรโดยใช้ดัชนี VIIC และใช้เกณฑ์จุดตัด BCT_{VIIC}

2. ผลการศึกษาการคัดเลือกตัวแปรด้วยวิธีที่พัฒนาขึ้นซึ่งได้แก่วิธี VIP-BCT และวิธี VIIC-BCT โดยเปรียบเทียบความแม่นยำในการจำแนกประเภทตัวแปรกับวิธีการคัดเลือกตัวแปรแบบดั้งเดิม คือวิธี VIP-1 สรุปได้ดังนี้

2.1 วิธี VIP-BCT มีค่าเฉลี่ย BAC สูงกว่าวิธี VIP-1 ซึ่งเป็นไปตามสมมติฐานการวิจัย ในสถานการณ์โดยส่วนใหญ่ของ $Prel=1, 3, 5$

2.2 วิธี VIIC-BCT มีค่าเฉลี่ย BAC สูงกว่าวิธี VIP-1 ซึ่งเป็นไปตามสมมติฐานการวิจัย ในสถานการณ์โดยส่วนใหญ่ของ $Prel=1, 3, 5$ และเมื่อ $Mdif=1$

นอกจากนี้เมื่อเปรียบเทียบวิธีการคัดเลือกตัวแปรที่พัฒนาขึ้นทั้งสองวิธีปรากฏว่า วิธี VIP-BCT มีค่าเฉลี่ย BAC สูงกว่าวิธี VIIC-BCT ในสถานการณ์ส่วนใหญ่ ยกเว้นในสถานการณ์ที่ $Mdif=1$ และ $n=40$ ที่วิธี VIIC-BCT มีค่าเฉลี่ย BAC สูงกว่าวิธี VIP-BCT

3. การศึกษาความสามารถของยีนที่ได้จากการคัดเลือกตัวแปรในการจำแนกประเภทเนื้อเยื่อมะเร็งลำไส้ใหญ่ และเนื้อเยื่อปกติ ซึ่งข้อมูลประกอบด้วยเนื้อเยื่อจำนวน 62 ชิ้นเนื้อ และมียีนจำนวน 2,000 ยีน ได้ผลสรุปดังนี้

3.1 การคัดเลือกตัวแปรด้วยวิธี VIP-BCT คัดเลือกได้ยีนจำนวนน้อยที่สุด โดยมีจำนวน 190 ยีน รองลงมาคือวิธี VIIC-BCT คัดเลือกยีนได้ 452 ยีน และวิธี VIP-1 คัดเลือกยีนได้จำนวน 683 ยีน

3.2 เมื่อจำแนกประเภทเนื้อเยื่อมะเร็งลำไส้ใหญ่ และเนื้อเยื่อปกติด้วยวิธี PLS-DA ความสามารถของยีนที่คัดเลือกได้จากวิธีการคัดเลือกตัวแปรแต่ละวิธี ปรากฏว่า ค่าเฉลี่ย BAC ของการจำแนกประเภทการเป็นมะเร็งลำไส้ใหญ่เมื่อคัดเลือกตัวแปรด้วยวิธี VIP-1 วิธี VIP-BCT และวิธี VIIC-BCT เท่ากับ 82.02% 82.71% และ 83.21% ตามลำดับ ซึ่งมีค่ามากกว่า 80% อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ทั้งสามวิธี ซึ่งเป็นไปตามสมมติฐานการวิจัย

อภิปรายผล

การวิจัยนี้ได้สิ่งที่ค้นพบซึ่งนำมาอภิปรายได้ดังนี้

1. ดัชนี VIIC ได้พัฒนาขึ้นเพื่อให้เหมาะสมกับงานการจำแนกประเภท ซึ่งอาศัยแนวทางจากดัชนี VIP โดยมีการปรับค่าด้วยการใช้ Adjusted Count R^2 มาเป็นค่าวัดความสอดคล้องขององค์ประกอบ t_o ในการทำนายตัวแปรตามซึ่งเป็นตัวแปรเชิงคุณภาพ แทนการใช้ R^2 ซึ่งเหมาะสมกับการวัดความสอดคล้องกรณีที่ตัวแปรตามเป็นตัวแปรเชิงปริมาณ การใช้ Adjusted Count R^2 มาเป็นค่าวัดความสอดคล้องขององค์ประกอบ t_o มีความเหมาะสมกว่าการใช้ R^2 เนื่องจากการจำแนกประเภทนั้นตัวแปรตามเป็นตัวแปรเชิงคุณภาพ นอกจากนี้เกณฑ์จุดตัดที่กำหนดเป็นค่าคงที่นั้นแม้จะง่ายกับการใช้งาน แต่ไม่เหมาะสมกับทุกลักษณะของโครงสร้างข้อมูล การใช้เกณฑ์จุดตัดควรแตกต่างกันในแต่ละข้อมูล เนื่องจากแต่ละข้อมูลมีลักษณะที่แตกต่างกัน เกณฑ์จุดตัดที่เหมาะสมจึงควรแตกต่างกันในแต่ละข้อมูล ซึ่งสอดคล้องกับแนวความคิดของ Chong and Jun (2005)

2. ในสถานการณ์ส่วนใหญ่ วิธี VIP-BCT มีความแม่นยำในการคัดเลือกตัวแปรมากกว่าวิธี VIP-1 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ซึ่งเกณฑ์จุดตัดที่ได้จากวิธี VIP-BCT มีค่าเฉลี่ยมากกว่า 1.0 ในทุกสถานการณ์ โดยเกณฑ์จุดตัดมีค่ามากในกรณี Prel, Mdif และ n มีค่าน้อย และเกณฑ์จุดตัดมีค่าเข้าใกล้ 1.0 เมื่อ Prel, Mdif และ n มีค่ามาก ซึ่งในพารามิเตอร์ Prel นั้นสอดคล้องกับการหาค่าเกณฑ์จุดตัดที่เหมาะสมในแต่ละสถานการณ์โดย Chong and Jun (2005) ที่เมื่อ Prel มีค่าน้อย เกณฑ์จุดตัดที่เหมาะสมควรมีค่ามากกว่าเมื่อ Prel มีค่ามาก ซึ่งในงานดังกล่าวศึกษาในพารามิเตอร์ Prel=10 เป็นระดับต่ำสุด (ในขณะที่งานวิจัยนี้ศึกษาพารามิเตอร์ Prel=10 เป็นระดับสูงสุด) โดยในระดับดังกล่าวเกณฑ์จุดตัดที่เหมาะสมโดยส่วนใหญ่มีค่าเกิน 1.0 ดังนั้นเมื่อ Prel มีค่าน้อย การคัดเลือกตัวแปรด้วยวิธี VIP-1 ซึ่งใช้เกณฑ์จุดตัดค่าคงที่ 1.0 จึงมีค่าเกณฑ์จุดตัดต่ำกว่าเกณฑ์จุดตัดที่เหมาะสม ซึ่งมีแนวโน้มที่จะคัดเลือกตัวแปรจำนวนที่มากเกินไปจำนวนตัวแปรที่เกี่ยวข้อง การใช้เกณฑ์จุดตัด BCT สำหรับวิธี VIIC-BCT ได้ผลใกล้เคียงกับวิธี VIP-BCT ในด้านของสถานการณ์ที่มีความแม่นยำกว่าวิธี VIP-1 ซึ่งเป็นผลอันเนื่องมาจากการกำหนดเกณฑ์จุดตัดที่ไม่ได้เป็นค่าคงที่

เมื่อพิจารณาเฉพาะวิธีการที่ใช้เกณฑ์จุดตัด BCT ได้แก่วิธี VIP-BCT และ VIIC-BCT ซึ่งมีความแตกต่างกันในส่วนของดัชนีที่ใช้วัดความสำคัญของตัวแปร วิธี VIIC-BCT นั้นให้ผลดีกว่าวิธี

VIP-BCT ในสถานการณ์ที่ยากแก่การจำแนกประเภทข้อมูล คือในสถานการณ์ที่ $Mdif$ และ n มีค่าน้อย ($Mdif=1$ และ $n=40$) สาเหตุอาจเนื่องมาจากการสร้างเกณฑ์จุดตัด BCT อาศัยเกณฑ์การตรวจสอบค่าผิดปกติของ $VIIC_{noise}$ หรือ VIP_{noise} สำหรับวิธี VIP-BCT และวิธี VIIC-BCT ตามลำดับ ซึ่ง $VIIC_{noise}$ มีความแม่นยำมากกว่า และหางของการแจกแจงยาวกว่าของ VIP_{noise} ทำให้มีโอกาสเกิดค่าผิดปกติมากกว่า และเกณฑ์จุดตัดที่ได้มีค่าต่ำ ซึ่งในงานการจำแนกประเภทข้อมูลที่มีลักษณะยากต่อการจำแนกนั้น ค่า VIIC และ VIP ของตัวแปรที่เกี่ยวข้องและตัวแปรที่ไม่เกี่ยวข้องจะแตกต่างกันอย่างไม่ชัดเจน การใช้เกณฑ์จุดตัดต่ำจะทำให้ตัวแปรที่เกี่ยวข้องมีโอกาสได้รับเลือกมากกว่าเมื่อใช้เกณฑ์จุดตัดที่มีค่ามาก ซึ่งส่งผลให้ค่า Sen มีค่ามาก ในขณะที่ Spe ได้รับการควบคุมค่าเมื่อใช้เกณฑ์จุดตัด BCT เนื่องจากการใช้เกณฑ์จุดตัด BCT สามารถควบคุมโอกาสที่จะผิดพลาดในการจำแนกตัวแปรที่ไม่เกี่ยวข้องว่าเป็นตัวแปรที่เกี่ยวข้อง ซึ่งสังเกตว่าค่า Spe ของวิธีที่ใช้เกณฑ์จุดตัด BCT จะมีค่ามากและเปลี่ยนแปลงไม่มากในแต่ละสถานการณ์ ดังนั้นจึงนำไปสู่ค่า BAC ที่มีค่ามากเนื่องจาก BAC เป็นค่าเฉลี่ยระหว่าง Sen กับ Spe เมื่อเทียบกับวิธี VIP-BCT ซึ่งมีความแม่นยำกว่าทำให้เกณฑ์จุดตัดที่ได้มีค่ามาก ทำให้ตัวแปรที่เกี่ยวข้องมีโอกาสได้รับเลือกน้อย ส่งผลให้ค่า Sen มีค่าน้อยนำไปสู่ค่า BAC ที่มีค่าน้อย ดังนั้นในสถานการณ์ที่ยากแก่การจำแนกประเภทข้อมูล วิธี VIIC-BCT นั้นให้ผลดีกว่าวิธี VIP-BCT

3. เมื่อจำแนกประเภทเนื้อเยื่อมะเร็งลำไส้ใหญ่ และเนื้อเยื่อปกติด้วยวิธี PLS-DA ถึงแม้ว่าความสามารถของยีนที่คัดเลือกได้จากวิธีการคัดเลือกตัวแปรแต่ละวิธีไม่แตกต่างกันมากนัก แต่จำนวนยีนที่ใช้ในการจำแนกประเภทเมื่อใช้วิธี VIP-BCT มีจำนวนน้อยที่สุด รองลงมาคือวิธี VIIC-BCT และวิธี VIP-1 โดยมีจำนวนเท่ากับ 190 ยีน 452 ยีน และ 683 ยีน ตามลำดับ ซึ่งการใช้ยีนจำนวนน้อยกว่าทำให้สามารถลดค่าใช้จ่าย และเวลาที่ใช้ในการรวบรวมข้อมูลได้ ในขณะที่มีประสิทธิภาพไม่แตกต่างกัน

การเปรียบเทียบวิธีการคัดเลือกตัวแปรเมื่อใช้ข้อมูลจำลองนั้น วัดประสิทธิภาพด้วยความแม่นยำในการจำแนกประเภทตัวแปรว่าเป็นตัวแปรที่เกี่ยวข้อง หรือตัวแปรไม่เกี่ยวข้อง ซึ่งเป็นการวัดประสิทธิภาพของการคัดเลือกตัวแปร ในขณะที่การศึกษาความสามารถของยีนที่คัดเลือกได้จากวิธีการคัดเลือกตัวแปรแต่ละวิธีนั้น เป็นการคัดเลือกยีนจากวิธีต่าง ๆ และนำยีนที่ได้ไปใช้ในการสร้างตัวแบบจำแนกประเภทด้วยวิธี PLS-DA และวัดประสิทธิภาพด้วยความแม่นยำในการจำแนกประเภทเนื้อเยื่อว่าเป็นเนื้อเยื่อมะเร็งลำไส้ใหญ่ หรือเนื้อเยื่อปกติ ซึ่งเป็นการวัดประสิทธิภาพการทำนายประสิทธิภาพของการคัดเลือกตัวแปรอาจไม่มีความสัมพันธ์กันอย่างมากกับประสิทธิภาพการทำนาย (Chong & Jun, 2005) จึงมีโอกาสที่ผลการเปรียบเทียบวิธีการคัดเลือกตัวแปรจากข้อมูลจำลองให้ผลที่ต่างจากข้อมูลที่ใช้ในการศึกษาความสามารถของยีน

ยีนที่ได้รับการคัดเลือกจากวิธีการคัดเลือกตัวแปรทั้ง 3 ก่อนข้างสอดคล้องกับงานวิจัยอื่น (Ben-Dor et al., 2000; Chen & Nguyen, n.d.; Fajarewicz & Wiench, 2003; Wang, Zhu, & Zou, 2008) ซึ่งได้รวมยีนที่มีความสัมพันธ์กับการเป็นมะเร็งลำไส้ใหญ่ ทำให้สามารถนำไปใช้ในการทำนายการเป็นมะเร็งลำไส้ใหญ่ได้แม่นยำโดยทำนายถูกต้องมากกว่า 80% ยีนที่อยู่ในอันดับต้นของทั้ง 3 วิธีการคัดเลือกตัวแปร ได้แก่

Human Cysteine-Rich Protein (CRP) gene (หมายเลขยีน 765: M76378) โปรตีนกลุ่มย่อยที่อยู่ในตระกูลของ CRP (Metallothionein) ที่มีความเกี่ยวข้องกับการเกิดมะเร็งลำไส้ใหญ่ (Arriaga et al., 2012)

Human Desmin Gene (หมายเลขยีน 249: M63391) การแสดงออกที่สูงขึ้นของ Desmin มีความสัมพันธ์กับความรุนแรงและการแบ่งแยกของมะเร็งลำไส้ใหญ่และทำให้อัตราการอยู่รอดของผู้ป่วยมะเร็งลำไส้ใหญ่ลดลง (Ma et al., 2009) และในปัจจุบันมีหลักฐานที่แสดงให้เห็นถึงความเกี่ยวข้องทางคลินิกของผู้ป่วยมะเร็งลำไส้เป็นจำนวนมาก

Uroguanylin Precursor (หมายเลขยีน 377: Z50753) การรักษาด้วย Uroguanylin มีผลอย่างมีนัยสำคัญในเชิงบวกต่อการลดลงของระยะก่อนเป็นมะเร็งลำไส้ใหญ่ (Shailubhai et al., 2000)

Myosin Heavy Chain, Non-Muscle (หมายเลขยีน 493:R87126) บทบาทที่สำคัญของ Myosin คือเป็นองค์ประกอบที่แสดงออกมาในเซลล์ที่ไม่ใช่กล้ามเนื้อ (Non-muscle cell) รวมทั้งเป็นส่วนหนึ่งของเครือข่ายสารโครงร่างของเซลล์ (Cytoskeletal Network) (Yam, Chan, & Hsiao, 2001) มีหลักฐานที่แสดงให้เห็นถึงบทบาทของ Myosin ในการสร้างสารไซโตไคน์ (Cytokinesis) (De Lozanne & Spudich, 1987) ภายในเซลล์ซึ่งเกี่ยวข้องกับการเจริญขึ้นมาใหม่ (Proliferation) ของเซลล์มะเร็ง (Rovner, Murphy, & Owens, 1986; Simons & Rosenberg, 1992)

Myosin Regulatory Light Chain 2 (หมายเลขยีน 1423: J02854) เป็นเป้าหมายภายในเซลล์ของกลุ่มโปรตีนที่ฝังอยู่เนื้อผิวเปลือกเซลล์ (integrins) ที่มีผลต่อการเคลื่อนไหวของเซลล์ ซึ่งมีอำนาจจำแนกในการเป็นมะเร็ง (Keely, Parise, & Juliano, 1998) การเปลี่ยนแปลงระดับของ Myosin Regulatory Light Chain 2 สอดคล้องกับบทบาทในการป้องกันโรคมะเร็งลำไส้ใหญ่ (Breikers et al., 2006)

ข้อเสนอแนะ

ข้อเสนอแนะในการนำไปใช้

1. วิธี VIIC-BCT เป็นวิธีที่มีความแม่นยำในการจำแนกประเภทตัวแปรมากที่สุดอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 เมื่อใช้เกณฑ์การเปรียบเทียบค่าเฉลี่ย BAC ในสถานการณ์ที่ยากแก่

การจำแนกประเภทข้อมูล กล่าวคือความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้องมีค่าน้อย และขนาดตัวอย่างมีค่าน้อย ($Mdif=1$ และ $n=40$)

2. วิธี VIP-BCT เป็นวิธีที่มีความแม่นยำในการจำแนกประเภทตัวแปรมากที่สุดอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 เมื่อใช้เกณฑ์การเปรียบเทียบค่าเฉลี่ย BAC ในสถานการณ์ที่เปอร์เซ็นต์ตัวแปรที่เกี่ยวข้องกับการจำแนกประเภทข้อมูลเท่ากับ 1, 3 และ 5 ยกเว้นสถานการณ์ที่ยากแก่การจำแนกประเภทข้อมูล

3. การที่จะทราบได้ว่าตัวแปรใดที่เกี่ยวข้องกับการจำแนกประเภทข้อมูลนั้น เป็นไปได้ยากในทางปฏิบัติ ดังนั้นการพิจารณาการใช้วิธีการคัดเลือกตัวแปรจากขนาดตัวอย่างโดยพิจารณาครั้งนี้ในกรณีที่ขนาดตัวอย่างไม่เล็กและไม่ใหญ่จนเกินไป ($70 \leq n \leq 100$) การคัดเลือกตัวแปรด้วยวิธี VIP-BCT จะให้ค่าเฉลี่ย BAC มากที่สุดในสถานการณ์ส่วนใหญ่ และเมื่อขนาดตัวอย่างเล็ก ($n=40$) ถึงแม้วิธี VIIC-BCT จะให้ผลดีที่สุดเฉพาะเมื่อความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มข้อมูลของตัวแปรที่เกี่ยวข้องเท่ากับ 1 แต่เมื่อพิจารณาในทุกสถานการณ์เมื่อขนาดตัวอย่างเท่ากับ 40 พบว่า วิธี VIIC-BCT มีค่าเฉลี่ย BAC ที่ไม่ผันผวนไปในแต่ละสถานการณ์เมื่อเปรียบเทียบกับวิธีอื่น โดยมีค่าเฉลี่ย BAC อยู่ในช่วง $0.9132 - 0.9741$ ในขณะที่ VIP-BCT มีค่าเฉลี่ย BAC อยู่ในช่วง $0.8498 - 0.9930$ และ VIP-1 มีค่าเฉลี่ย BAC อยู่ในช่วง $0.8374 - 0.9818$ ดังนั้นจึงควรใช้วิธี VIIC-BCT เมื่อตัวแปรมีขนาดเล็ก

4. การใช้ดัชนีวัดความสำคัญของตัวแปร และเกณฑ์จุดตัดในการคัดเลือกตัวแปรที่นำเสนอนี้ไม่ได้มุ่งเน้นหาเซตย่อยที่เล็กที่สุดที่มีประสิทธิภาพสูงสุด เนื่องจากการทำในลักษณะดังกล่าวใช้เวลาในการประมวลผลมากโดยเฉพาะอย่างยิ่งเมื่อมีตัวแปรเป็นจำนวนมาก ดังนั้นการคัดเลือกตัวแปรที่นำเสนอสามารถใช้เป็นการคัดเลือกตัวแปรในเบื้องต้นเพื่อลดจำนวนตัวแปรลง ซึ่งจะทำให้สามารถนำไปใช้กับวิธีการจำแนกประเภทบางวิธีที่มีข้อกำหนดในด้านจำนวนตัวแปร และเพื่อลดเวลาในการประมวลผลได้

ข้อเสนอแนะในการศึกษาต่อไป

1. การศึกษาที่ผู้วิจัยได้ศึกษากับกรณีที่ข้อมูลจำแนกเป็น 2 กลุ่ม (Binary Class) เท่านั้น แต่ดัชนี VIIC และเกณฑ์ BCT สามารถนำไปใช้กับกรณีที่ข้อมูลมีมากกว่า 2 กลุ่มได้ โดยกำหนดจำนวนตัวแปรตามเท่ากับจำนวนกลุ่ม จึงอาจศึกษาในกรณีดังกล่าวได้ด้วย
2. การศึกษาที่ผู้วิจัยใช้เกณฑ์จุดตัด BCT สำหรับดัชนี VIIC และดัชนี VIP ในการศึกษาครั้งนี้ต่อไปอาจนำเกณฑ์จุดตัด BCT ไปใช้กับดัชนีอื่น ๆ ซึ่งดัชนีโดยส่วนใหญ่ไม่ได้กำหนดเกณฑ์จุดตัดที่ชัดเจนในการคัดเลือกตัวแปร
3. การศึกษาความสามารถของยีนที่ได้จากการคัดเลือกตัวแปรแต่ละวิธี ในการจำแนกประเภทเนื้อเยื่อมะเร็งลำไส้ใหญ่ และเนื้อเยื่อปกติ ผู้วิจัยใช้วิธีการจำแนกประเภท PLS-DA ในการ

สร้างตัวแบบจำแนกประเภท จึงควรได้มีการศึกษาต่อไปว่าวิธีการจำแนกประเภทวิธีอื่นให้ผลลัพธ์ในลักษณะเดียวกับ PLS-DA หรือไม่ เนื่องจากอาจมีอันตรกิริยา (Interaction) ระหว่างวิธีการคัดเลือกตัวแปร และวิธีการจำแนกประเภทข้อมูล ซึ่งอาจนำไปสู่ผลลัพธ์ที่แตกต่างกัน

มหาวิทยาลัยบูรพา
Burapha University