

บทที่ 4

ผลการทดลองและวิจารณ์

วิทยานิพนธ์นี้ได้ทดลองกับฐานข้อมูลทางการแพทย์ที่เป็นมาตรฐานจาก University of California at Irvine [30] เป็นฐานข้อมูลระดับสากลได้แก่ฐานข้อมูลโรคมะเร็งเต้านม (Wisconsin Breast Cancer Database) ฐานข้อมูลโรคหัวใจจากคลินิกรักษาโรคหัวใจ (Heart Disease from Cleveland Clinic Database) ฐานข้อมูลโรคหัวใจจากโปรเจกต์สแตตส์อียุโรป (Heart Disease from European Statlog Project Database) ฐานข้อมูลโรคเบาหวาน (Pima Indians Diabetes Database) และฐานข้อมูลทางการแพทย์ที่เก็บตัวอย่างจากโรงพยาบาลในประเทศไทยได้แก่ ฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจ (Electrocardiogram Database of Heart Disease Patients) [39]

4.1 ฐานข้อมูลโรคมะเร็งเต้านม

ฐานข้อมูลโรคมะเร็งเต้านม (Wisconsin breast cancer database) ประกอบด้วยข้อมูล 2 กลุ่มคือ กลุ่มที่ไม่เป็นโรคมะเร็งเต้านม (Benign) มี 458 คน และกลุ่มที่เป็นโรคมะเร็งเต้านม (Malignant) มี 241 คน รวมทั้งทั้งหมด 699 คน ตัวแปรข้อมูลเข้ามีทั้งหมด 9 ตัวแปร x_1 คือ Clump thickness x_2 คือ Uniformity of cell size x_3 คือ Uniformity of cell shape x_4 คือ Marginal adhesion x_5 คือ Single epithelial cell size x_6 คือ Bare nuclei x_7 คือ Bland chromatin x_8 คือ Normal nucleoli และ x_9 คือ Mitoses ดังตารางที่ 4.1

4.1.1 การตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW)

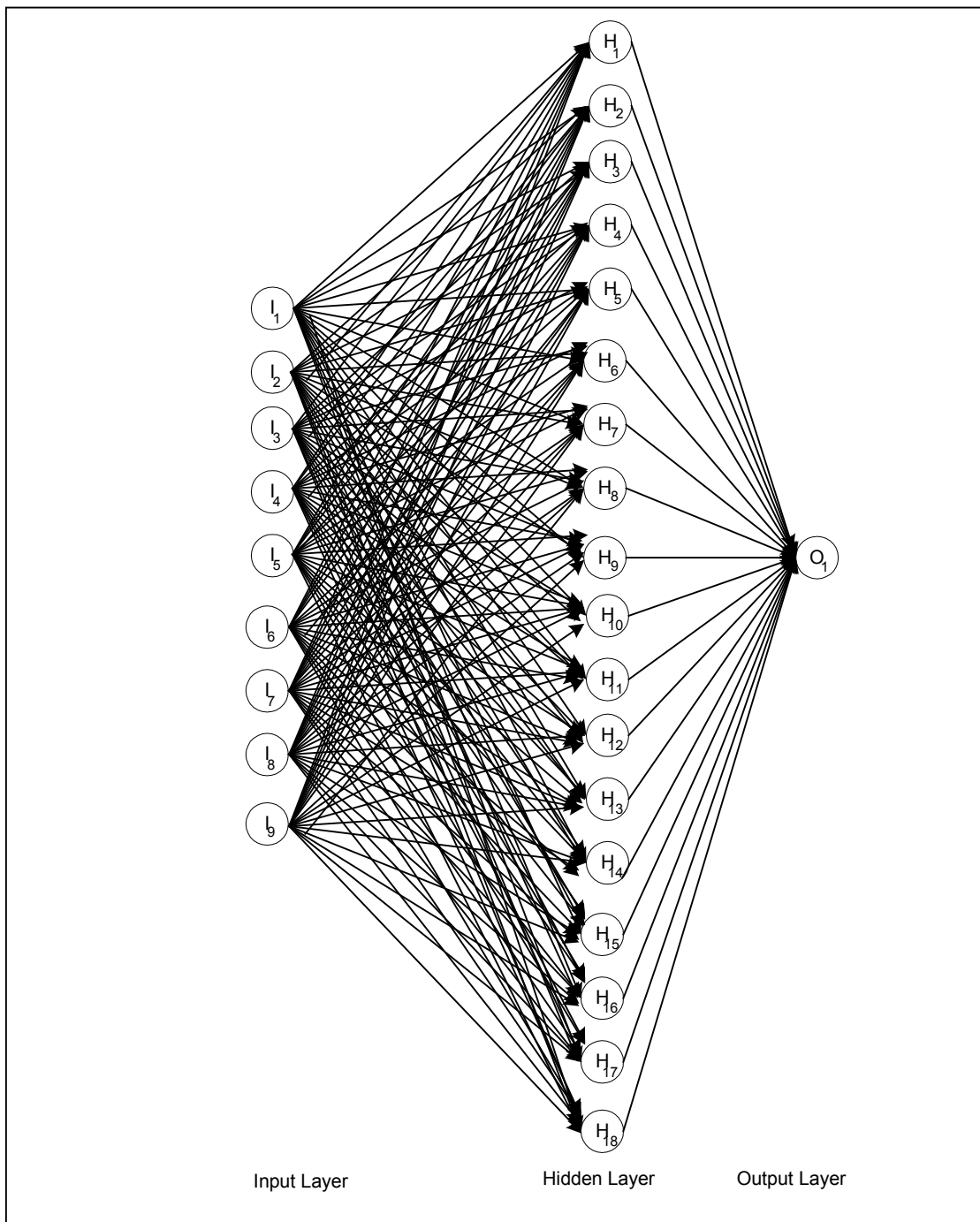
ในการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุดมีขั้นตอนในการทำงาน 3 ขั้นตอนมีรายละเอียดดังนี้

ขั้นตอนที่ 1 การเตรียมข้อมูล เนื่องจากในฐานข้อมูลโรคมะเร็งเต้านมมีแถว ข้อมูลที่มีค่าของตัวแปรข้อมูลเข้าที่ 7 หายไป (Missing Values) 16 แถว ดังนั้นจึงแทนค่าตัวแปรข้อมูลเข้าที่หายไปด้วยค่าเฉลี่ยของกลุ่ม (Class Mean) ซึ่งมีค่าเท่ากับ 3 ในการทดลองมีการเลือกข้อมูลเพื่อใช้ในการสอน (Training set) 90 % และใช้ในการทดสอบ (Testing set) 10 %

ตารางที่ 4.1 รายละเอียดของฐานข้อมูลโรคมะเร็งเต้านม

ตัวแปร	ชื่อตัวแปร	ชนิด	ค่าที่เป็นไปได้
x_1	clump thickness	Nominal Scale	1-10
x_2	uniformity of cell size	Nominal Scale	1-10
x_3	uniformity of cell shape	Nominal Scale	1-10
x_4	marginal adhesion	Nominal Scale	1-10
x_5	single epithelial cell size	Nominal Scale	1-10
x_6	bare nuclei	Nominal Scale	1-10
x_7	bland chromatin	Nominal Scale	1-10
x_8	normal nucleoli	Nominal Scale	1-10
x_9	mitoses	Nominal Scale	1-10
c	class	Nominal Scale	1 = Malignant 2 = Benign

ขั้นตอนที่ 2 การสอนโครงข่ายประสาทเทียมด้วยการเรียนรู้แบบแพร่ย้อนกลับ กำหนดให้มีโหนดข้อมูลเข้า 9 โหนด โหนดซ่อน 18 โหนด และโหนดผลลัพธ์ 1 โหนด ดังนั้น โครงสร้างของโครงข่ายประสาทเทียมคือ 9:18:1 ดังภาพประกอบ 4.1 โดยจะทำการสอนโครงข่ายประสาทเทียมจนกระทั่งครบจำนวนรอบที่ตั้งไว้ 200 รอบ (Epochs) หรือได้ค่าความผิดพลาด (MSE) น้อยกว่าค่าที่ยอมรับได้ซึ่งมีค่าเท่ากับ 0 ในขั้นตอนนี้จะได้ค่าน้ำหนักของโครงข่ายประสาทเทียม ตัวอย่างค่าน้ำหนักระหว่างชั้นซ่อนและชั้นผลลัพธ์แสดง ดังตารางที่ 4.2 ตัวอย่างค่าน้ำหนักระหว่างชั้นข้อมูลเข้าและชั้นซ่อนแสดงดังตารางที่ 4.3 โดยที่ ค่าน้ำหนักที่ได้จะถูกนำไปใช้ในขั้นตอนที่ 3 ขั้นตอนวิธีการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด



ภาพประกอบ 4.1 โครงสร้างของโครงข่ายประสาทเทียมก่อนการตัดโหนด
ของฐานข้อมูลโรคมะเร็งเต้านม

ตารางที่ 4.2 คำนวณน้ำหนักระหว่างชั้นซ่อนและชั้นผลลัพธ์ของฐานข้อมูลโรคมะเร็งเต้านม

H	O₁
H₁	15.4
H₂	-0.4
H₃	6.5
H₄	19.6
H₅	13.6
H₆	-0.1
H₇	-27.8
H₈	0.8
H₉	17.7
H₁₀	9.4
H₁₁	-26.9
H₁₂	-13.0
H₁₃	-13.8
H₁₄	-23.4
H₁₅	22.2
H₁₆	37.7
H₁₇	-17.7
H₁₈	8.9

ตารางที่ 4.3 คำนวณน้ำหนักระหว่างชั้นข้อมูลเข้าและชั้นซ่อนของฐานข้อมูลโรคมะเร็งเต้านม

I	H₁	H₂	H₃	H₄	H₅	H₆	H₇	H₈	H₉	H₁₀	H₁₁	H₁₂	H₁₃	H₁₄	H₁₅	H₁₆	H₁₇	H₁₈
I₁	2.9	-2.0	-0.5	1.2	0.3	-1.3	-1.2	-1.1	-2.7	-1.5	3.3	0.4	0.1	0.3	-0.3	4.0	0.3	1.2
I₂	-4.4	-3.2	0.0	0.7	0.8	-1.4	0.9	-0.6	-1.3	0.3	2.4	0.7	0.3	-0.6	2.2	6.2	1.8	3.9
I₃	2.1	1.4	0.0	-0.2	-0.2	-1.0	2.7	-0.2	-2.8	1.2	-2.8	-1.4	1.1	1.5	2.7	5.3	-3.0	3.7
I₄	-3.3	-2.0	-0.5	0.6	-5.5	-0.7	-1.7	-0.6	0.2	-2.3	0.0	-0.2	-1.2	-2.6	2.5	-6.3	4.3	0.3
I₅	-3.4	-3.0	0.8	-1.0	0.5	1.4	-3.5	-1.2	-1.1	-3.8	1.0	-1.4	-2.7	-0.2	-2.5	-3.3	2.8	-4.6

I_6	1.8	-0.3	-0.6	-0.8	-0.5	-1.0	-0.5	-0.7	6.2	1.6	-1.1	-0.6	-3.0	0.3	-2.1	0.3	-7.0	-5.7
I_7	-2.6	0.0	1.6	2.9	2.2	-1.1	0.0	-0.7	-0.2	5.2	-1.3	2.4	2.4	-0.3	0.5	-1.6	1.0	2.7
I_8	2.7	1.4	0.1	0.2	-2.6	1.0	-2.4	-0.2	0.6	0.2	-1.9	-1.7	1.7	2.3	0.4	-2.7	-0.3	-1.4
I_9	-2.1	-1.2	-0.6	0.0	6.1	-3.5	-1.1	-0.8	-1.7	-1.8	-1.2	-2.9	-0.2	-0.9	-1.3	-2.4	-1.8	0.3

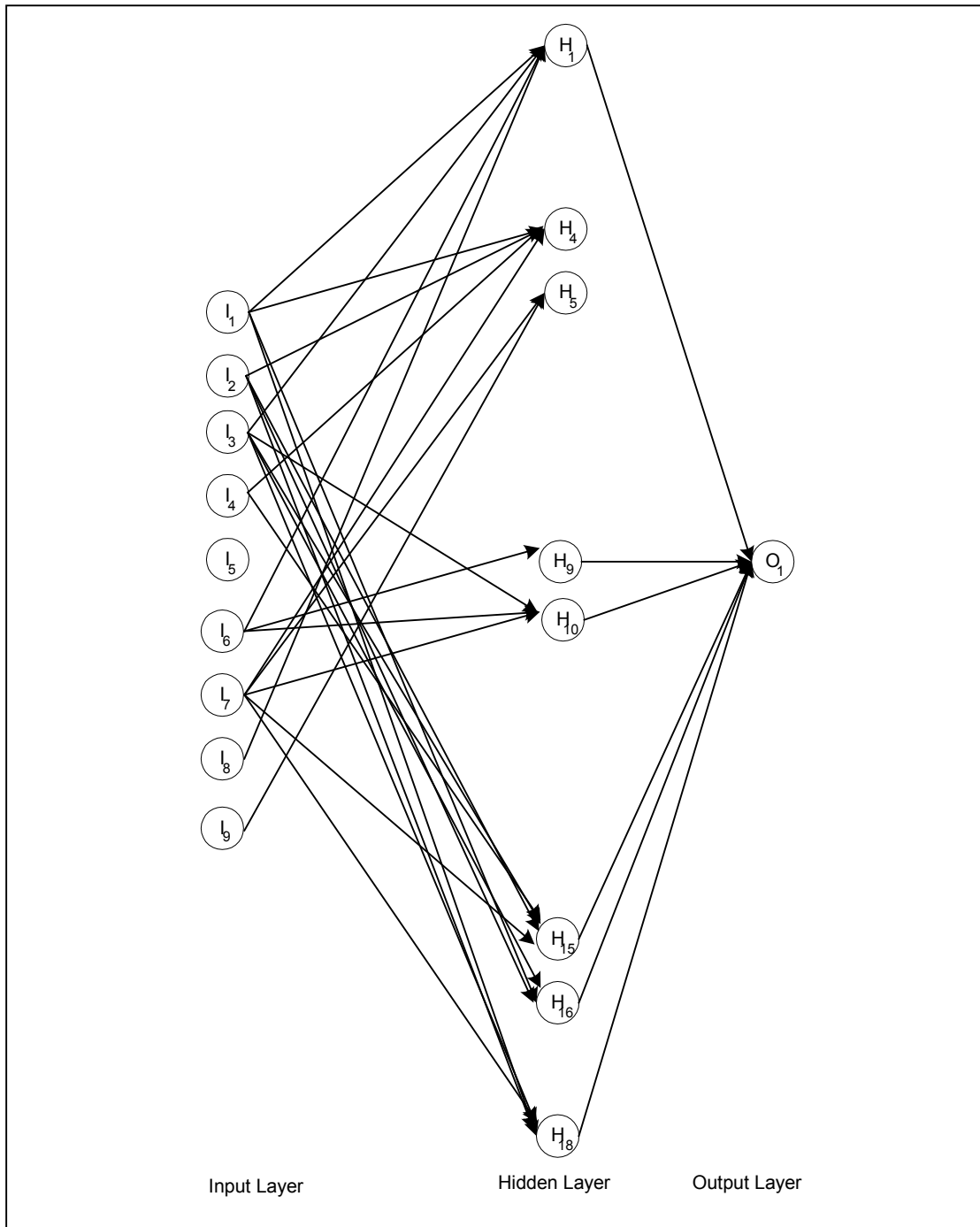
ขั้นตอนที่ 3 ขั้นตอนวิธีการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด ตัวอย่างกำหนดให้ค่าร้อยละของน้ำหนักที่ยอมรับได้ $\square = 20$ จากตารางที่ 4.2 ค่าน้ำหนักที่มีค่ามากที่สุดของแต่ละโหนดผลลัพธ์ซึ่งในที่นี้มีโหนดเดียวคือ O_1 (\max_w) = 37.7 ดังนั้นสามารถคำนวณหาค่าน้ำหนักที่ยอมรับได้ (accept_w) = $(37.7 \times 20) / 100$ หรือ 7.5 และยอมรับโหนดซ่อน H_1 H_4 H_5 H_9 H_{10} H_{15} H_{16} H_{18} รวมทั้งหมด 7 โหนดเนื่องจากมีค่าน้ำหนักมากกว่าค่าน้ำหนักที่ยอมรับได้ดังตารางที่ 4.4 จากตารางที่ 4.3 สามารถหาค่าน้ำหนักที่มีค่ามากที่สุดของแต่ละโหนดซ่อนที่ยอมรับ (\max_w') และคำนวณหาค่าน้ำหนักที่ยอมรับได้ (accept_w') ดังตารางที่ 4.5 ตัวอย่างเช่น โหนดซ่อนที่ 1 (H_1) มีค่าน้ำหนักที่มีค่ามากที่สุด (\max_w') = 2.9 มีค่าน้ำหนักที่ยอมรับได้ (accept_w') = 0.6 และโหนดข้อมูลเข้า I_3 I_6 I_8 มีค่าน้ำหนักมากกว่าค่าที่ยอมรับได้ดังนั้นจึงยอมรับโหนดข้อมูลเข้านั้น เมื่อพิจารณาทั้ง 7 โหนดซ่อนแล้วพบว่าจะยอมรับโหนดข้อมูลเข้า I_1 I_2 I_3 I_4 I_6 I_7 I_8 I_9 นั่นคือโหนดข้อมูลเข้า I_5 ถูกตัดออกจากโครงข่ายประสาทเทียม เพราะฉะนั้นโครงสร้างของโครงข่ายประสาทเทียมที่ผ่านการตัดโหนดคือ 8:8:1 ดังภาพประกอบ 4.2

ตารางที่ 4.4 การพิจารณายอมรับโหนดก่อนของฐานข้อมูลโรคมะเร็งเต้านม

	H	O₁	ยอมรับโหนดก่อน H
✓	H₁	15.4	
×	H₂	-0.4	
×	H₃	6.5	
✓	H₄	19.6	
✓	H₅	13.6	✓
×	H₆	-0.1	
×	H₇	-27.8	
×	H₈	0.8	
✓	H₉	17.7	
✓	H₁₀	9.4	
×	H₁₁	-26.9	
×	H₁₂	-13.0	
×	H₁₃	-13.8	
×	H₁₄	-23.4	
✓	H₁₅	22.2	
✓	H₁₆	37.7	
×	H₁₇	-17.7	
✓	H₁₈	8.9	
	max_w	37.7	
	accept_w	7.5	

ตารางที่ 4.5 การพิจารณายอมรับโหนดข้อมูลเข้าของฐานข้อมูลโรคมะเร็งเต้านม

I	H ₁	H ₄	H ₅	H ₉	H ₁₀	H ₁₅	H ₁₆	H ₁₈	ยอมรับโหนดข้อมูลเข้า I
✓ I ₁	2.9	1.2	0.3	-2.7	-1.5	-0.3	4.0	1.2	
✓ I ₂	-4.4	0.7	0.8	-1.3	0.3	2.2	6.2	3.9	
✓ I ₃	2.1	-0.2	-0.2	-2.8	1.2	2.7	5.3	3.7	
✓ I ₄	-3.3	0.6	-5.5	0.2	-2.3	2.5	-6.3	0.3	
✗ I ₅	-3.4	-1.0	0.5	-1.1	-3.8	-2.5	-3.3	-4.6	
✓ I ₆	1.8	-0.8	-0.5	6.2	1.6	-2.1	0.3	-5.7	
✓ I ₇	-2.6	2.9	2.2	-0.2	5.2	0.5	-1.6	2.7	
✓ I ₈	2.7	0.2	-2.6	0.6	0.2	0.4	-2.7	-1.4	
✓ I ₉	-2.1	0.0	6.1	-1.7	-1.8	-1.3	-2.4	0.3	
max_w'	2.9	3	6.1	6.2	5.2	2.7	6.2	3.9	
accept_w'	0.6	0.6	1.2	1.2	1.0	0.5	1.2	0.8	



ภาพประกอบ 4.2 โครงสร้างของโครงข่ายประสาทเทียมหลังการตัดโหนด
ของฐานข้อมูลโรคมะเร็งเต้านม

4.1.2 การสกัดกฏภาษาธรรมชาติจากโครงข่ายประสาทเทียมโดยการแทนค่าความถี่ของช่วงข้อมูล (RENN-FIDR)

ในการสกัดกฏภาษาธรรมชาติจากโครงข่ายประสาทเทียมโดยการแทนค่าความถี่ของช่วงข้อมูลจะพิจารณาเฉพาะโหนดซ่อน และโหนดข้อมูลเข้าที่เสนอแนะจากขั้นตอนหลักที่ I การตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) ในการทดลองจะเลือกกลุ่มข้อมูลแบบความเที่ยงตรง 10 กลุ่ม (10-Fold Cross Validation) เพื่อให้ข้อมูลทุกตัวได้ถูกนำมาเป็นข้อมูลชุดทดสอบ (Testing Set) ซึ่งมีขั้นตอนในการทำงาน 3 ขั้นตอนมีรายละเอียดดังนี้

ขั้นตอนที่ 4 การสร้างรูปแบบภาษาธรรมชาติจากความถี่ของช่วงข้อมูล กำหนดให้จำนวนช่วงภาษาธรรมชาติ $L = 3$ $L = 5$ และ $L = 7$ ตามลำดับ ตัวอย่างช่วงภาษาธรรมชาติที่แบ่งได้เมื่อ $\square = 20$ $\phi = 50$ $L = 3$ แสดงดังตารางที่ 4.6 เมื่อ $\square = 20$ $\phi = 50$ $L = 5$ แสดงดังตารางที่ 4.7 และเมื่อ $\square = 20$ $\phi = 50$ $L = 7$ แสดงดังตารางที่ 4.8

ตารางที่ 4.6 ช่วงของข้อมูลเมื่อ $\square = 20$ $\phi = 50$ และ $L = 3$ ของฐานข้อมูลโรคมะเร็งเต้านม

x	small	medium	large
x_1	[1,6)	[6,9)	[9,10]
x_2	[1,4)	[4,10)	[10,10]
x_3	[1,4)	[4,9)	[9,10]
x_4	[1,3)	[3,9)	[9,10]
x_6	[1,9)	[9,10)	[10,10]
x_7	[1,4)	[4,8)	[8,10]
x_8	[1,3)	[3,10)	[10,10]
x_9	[1,7)	[7,8)	[8,10]

ตารางที่ 4.7 ช่วงของข้อมูลเมื่อ $\square = 20$ $\phi = 50$ และ $L = 5$ ของฐานข้อมูลโรคมะเร็งเต้านม

x	very small	small	medium	large	very large
x_1	[1,5)	[5,6)	[6,8)	[8,9)	[9,10]
x_2	[1,4)	[4,5)	[5,6)	[6,10)	[10,10]
x_3	[1,3)	[3,4)	[4,6)	[6,9)	[9,10]

x_4	[1,2)	[2,3)	[3,7)	[7,9)	[9,10]
x_6	[1,6)	[6,7)	[7,9)	[9,10)	[10,10]
x_7	[1,4)	[4,6)	[6,7)	[7,8)	[8,10]
x_8	[1,3)	[3,6)	[6,8)	[8,10)	[10,10]
x_9	[1,5)	[5,6)	[6,7)	[7,8)	[8,10]

ตารางที่ 4.8 ช่วงของข้อมูลเมื่อ $\square = 20$ $\phi = 50$ และ $L = 7$ ของฐานข้อมูลโรคมะเร็งเต้านม

x	tiny	very small	small	medium	large	very large	huge
x_1	[1,3)	[3,5)	[5,6)	[6,7)	[7,8)	[8,9)	[9,10]
x_2	[1,3)	[3,4)	[4,5)	[5,6)	[6,8)	[8,10)	[10,10]
x_3	[1,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,9)	[9,10]
x_4	[1,2)	[2,3)	[3,4)	[4,5)	[5,7)	[7,9)	[9,10]
x_6	[1,4)	[4,5)	[5,6)	[6,7)	[7,9)	[9,10)	[10,10]
x_7	[1,2)	[2,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,10]
x_8	[1,3)	[3,4)	[4,6)	[6,7)	[7,8)	[8,10)	[10,10]
x_9	[1,2)	[2,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,10]

ขั้นตอนที่ 5 การสอนโครงข่ายประสาทเทียมด้วยรูปแบบภาษาธรรมชาติ โดยมีการเรียนรู้แบบแพร่ย้อนกลับ โหนดข้อมูลเข้าและโหนดซ่อนจะกำหนดจากการเสนอแนะจาก ขั้นตอนการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) ตัวอย่างเช่นถ้า โหนดข้อมูลเข้า 8 โหนด และโหนดซ่อน 8 โหนด กรณีที่ $L = 7$ จะได้โครงสร้างของโครงข่ายประสาทเทียมเป็น $(8 \times 7):8:1$ หรือ $56:8:1$ โดยจะทำการสอนโครงข่ายประสาทเทียมจนกระทั่งครบจำนวนรอบที่ตั้งไว้ 200 รอบ (Epochs) หรือได้ค่าความผิดพลาด (MSE) น้อยกว่าค่าที่ยอมรับได้ซึ่งมีค่าเท่ากับ 0 ตัวอย่างค่าน้ำหนักระหว่างชั้นซ่อนและชั้นผลลัพธ์แสดงดังตารางที่ 4.9 ตัวอย่างค่าน้ำหนักระหว่างชั้นข้อมูลเข้าและชั้นซ่อนแสดงดังตารางที่ 4.10 ซึ่งน้ำหนักที่ได้จากการสอนโครงข่ายประสาทเทียมถูกนำไปใช้ในขั้นตอนที่ 6 ขั้นตอนวิธีการสร้างกฎโดยพิจารณา ค่าน้ำหนักที่เป็นบวก

ตารางที่ 4.9 ค่าน้ำหนักระหว่างชั้นซ่อนและชั้นผลลัพธ์ของฐานข้อมูลโรคมะเร็งเต้านม

H	O₁
H₁	15.3
H₂	-27.6
H₃	6.8
H₄	7.4
H₅	1.6
H₆	-26.6
H₇	-10.6
H₈	20.9

ตารางที่ 4.10 คำนวณน้ำหนักระหว่างชั้นข้อมูลเข้าและชั้นซ่อนของฐานข้อมูลโรคมะเร็งเต้านม

I	H ₁	H ₂	H ₃	H ₄	H ₅	H ₆	H ₇	H ₈
I ₁₁	0.7	0.1	0.2	0.3	-1.5	1.3	1.3	-3.0
I ₁₂	0.7	-1.3	-2.0	-2.4	0.8	3.0	-3.4	-0.7
I ₁₃	1.1	0.5	0.6	0.3	-1.2	-1.1	4.0	-2.2
I ₁₄	0.1	2.4	0.7	0.5	1.1	-1.4	1.9	2.1
I ₁₅	-0.6	-1.9	0.3	0.0	1.1	-1.9	-1.2	2.5
I ₁₆	0.9	-0.1	0.3	-1.7	0.4	-2.4	0.2	-0.7
I ₁₇	-0.4	0.2	1.9	1.5	0.0	-0.6	-4.0	4.4
I ₂₁	-0.2	1.9	2.5	1.7	-1.0	5.7	0.0	-1.7
I ₂₂	-0.2	-0.5	0.0	0.5	-0.2	-0.8	0.0	-1.4
I ₂₃	-1.2	-1.0	0.9	-4.4	1.3	-1.4	-1.8	-3.5
I ₂₄	-0.9	-1.3	1.2	1.5	-1.1	-0.1	-0.9	1.5
I ₂₅	-1.4	1.4	0.5	1.1	0.5	-0.2	0.4	-2.5
I ₂₆	1.6	0.7	-0.6	0.7	1.1	0.1	-1.2	2.4
I ₂₇	1.5	0.4	0.0	-1.1	-1.1	-2.1	-0.2	4.9
I ₃₁	-2.8	-0.3	1.0	0.8	0.0	0.3	0.4	-0.6
I ₃₂	1.0	0.1	-0.9	1.4	-0.8	0.8	-1.4	3.8
I ₃₃	-2.9	2.0	0.0	0.6	-0.4	-1.0	0.4	-2.0
I ₃₄	1.5	-1.7	0.3	0.7	-1.0	2.0	-1.4	2.4
I ₃₅	1.2	-2.1	-1.6	-2.0	-0.1	0.0	1.3	-0.9
I ₃₆	0.9	1.0	-0.8	-1.2	-0.5	-1.0	-0.5	0.8
I ₃₇	-0.1	0.5	0.1	-0.2	-0.5	-3.9	-2.4	0.8
I ₄₁	0.6	-1.4	-1.4	-1.3	1.1	0.9	2.9	-0.7
I ₄₂	1.6	3.5	-0.4	0.4	-1.0	1.2	-0.4	2.7
I ₄₃	-1.2	-1.0	-0.4	1.4	-0.5	-0.1	-1.2	1.8
I ₄₄	1.8	-0.2	1.4	0.0	0.6	-0.4	0.7	-0.8
I ₄₅	1.0	-0.9	0.3	-1.1	0.7	0.6	-0.3	-1.0
I ₄₆	0.2	0.4	1.9	-0.9	0.3	-1.4	-0.8	1.0

\mathbf{I}_{47}	-1.6	-0.2	0.6	-0.5	-0.4	-0.2	0.7	0.3
-------------------	------	------	-----	------	------	------	-----	-----

ตารางที่ 4.10 (ต่อ)

I	H ₁	H ₂	H ₃	H ₄	H ₅	H ₆	H ₇	H ₈
I ₆₁	-2.4	-0.2	-1.1	1.5	0.7	2.4	-1.7	-2.1
I ₆₂	-2.9	-0.5	3.0	0.3	-1.0	0.1	1.1	2.0
I ₆₃	2.7	3.4	-0.6	0.1	0.5	-4.5	0.9	1.2
I ₆₄	2.1	-3.3	-0.4	-1.1	-1.3	4.1	1.0	-1.4
I ₆₅	0.5	-0.7	-1.9	0.9	-0.5	4.2	-0.7	-0.7
I ₆₆	0.2	0.6	0.0	0.9	-0.9	-3.0	-2.2	0.3
I ₆₇	1.6	1.1	-0.3	1.8	1.2	-3.5	-0.6	3.6
I ₇₁	-0.5	-2.4	0.6	-1.8	-0.3	5.5	-0.2	-1.2
I ₇₂	1.5	3.7	-1.6	-0.5	0.8	-5.3	-2.5	-1.8
I ₇₃	-2.5	0.1	2.2	0.8	1.5	-2.7	1.5	1.5
I ₇₄	2.8	0.5	-0.1	1.0	0.9	5.6	-0.2	-0.2
I ₇₅	-1.6	-1.0	1.4	0.0	-0.8	-1.0	0.2	2.4
I ₇₆	-4.5	0.2	2.9	1.6	0.9	-0.3	0.6	3.7
I ₇₇	1.6	-0.6	0.9	1.8	0.5	-1.9	-1.3	3.3
I ₈₁	0.4	0.9	1.1	1.9	0.3	-0.4	1.9	-0.7
I ₈₂	0.0	0.0	0.3	0.9	-1.1	0.7	-0.1	0.5
I ₈₃	-1.0	-1.9	-0.2	-3.0	0.1	2.8	0.0	-2.0
I ₈₄	0.1	-0.6	0.8	0.4	0.9	1.8	-0.9	3.9
I ₈₅	0.1	-0.9	-0.1	-0.3	0.8	-1.2	0.3	-4.4
I ₈₆	-0.2	-1.5	0.2	-0.2	-0.3	-0.5	-2.1	1.5
I ₈₇	1.8	1.6	1.3	1.8	-1.0	-1.8	-1.4	1.3
I ₉₁	-0.9	-0.6	-1.1	-2.4	0.6	-1.3	2.7	-2.8
I ₉₂	2.9	0.7	-2.0	1.8	-0.7	-0.6	-0.6	3.3
I ₉₃	-1.5	0.6	2.7	-1.0	-0.4	-0.3	-2.8	3.4
I ₉₄	-2.0	-2.2	0.3	1.6	0.8	2.6	-1.8	1.2
I ₉₅	-2.1	2.2	0.9	-0.3	-0.2	-1.7	1.5	-1.9
I ₉₆	1.0	0.7	0.2	0.1	0.2	2.3	2.6	-1.2

I_{97}	1.9	0.0	0.6	0.1	-0.3	-1.5	0.1	2.0
----------	-----	-----	-----	-----	------	------	-----	-----

ขั้นตอนที่ 6 ขั้นตอนวิธีการสร้างกฎโดยพิจารณาค่าน้ำหนักที่เป็นบวก จากตารางที่ 4.9 โหนดชอนที่มีค่าเส้นเชื่อมน้ำหนักเป็นบวกคือโหนดชอน H_1, H_3, H_4, H_5, H_8 และจากตารางที่ 4.10 เมื่อพิจารณาเส้นเชื่อมน้ำหนักระหว่างโหนดข้อมูลเข้าและโหนดชอนทั้ง 5 โหนดนี้จะพบว่ามีเส้นเชื่อมน้ำหนักที่มีค่าเป็นบวก 157 เส้นเชื่อม เมื่อนำมาสร้างกฎจะได้จำนวนกฎทั้งหมด $(150 \times (150+1))/2 = 12,403$ กฎ มีกฎที่มีค่าความถูกต้องในช่วง (IC) มากกว่าหรือเท่ากับ 0.5 อยู่จำนวน 217 กฎ ซึ่งใน 217 กฎนี้มีกฎที่มีค่าความถูกต้องนอกช่วง (OC) มากกว่าหรือเท่ากับ 0.5 อยู่จำนวน 18 กฎ โดยมีกฎที่ไม่มีตัวเชื่อม (and หรือ or) ของเงื่อนไข 12 กฎ ดังนั้นจึงนำกฎทั้ง 12 กฎนี้มาเชื่อมด้วยเงื่อนไข or ทำให้รวมกันได้เป็น 1 กฎและมีจำนวนตัวเชื่อมของเงื่อนไข 11 ตัวเชื่อม เพราะฉะนั้นจำนวนกฎที่สกัดได้คือ 7 กฎและจำนวนตัวเชื่อมของเงื่อนไขที่สกัดได้คือ 18 ตัวเชื่อม ดังภาพประกอบ 4.3 ซึ่งมีค่าความถูกต้อง 98.3 % (ทำนายถูกต้อง 687 คน และทำนายผิด 12 คน)

R_1 : If $x_6 = \text{huge}$ or $x_1 = \text{huge}$ or $x_2 = \text{huge}$ or $x_3 = \text{huge}$ or $x_8 = \text{huge}$ or

$x_7 = \text{huge}$ or $x_4 = \text{large}$ or $x_2 = \text{medium}$ or $x_1 = \text{very large}$ or $x_9 = \text{small}$ or

$x_6 = \text{large}$ or $x_6 = \text{medium}$ then Malignant

R_2 : If $x_4 = \text{huge}$ and $x_3 = \text{small}$ then Malignant

R_3 : If $x_3 = \text{very small}$ and $x_7 = \text{large}$ then Malignant

R_4 : If $x_4 = \text{medium}$ and $x_2 = \text{very small}$ then Malignant

R_5 : If $x_9 = \text{very small}$ and $x_3 = \text{very small}$ then Malignant

R_6 : If $x_6 = \text{very large}$ and $x_8 = \text{medium}$ then Malignant

R_7 : If $x_8 = \text{small}$ and $x_6 = \text{small}$ and $x_7 = \text{small}$ then Malignant

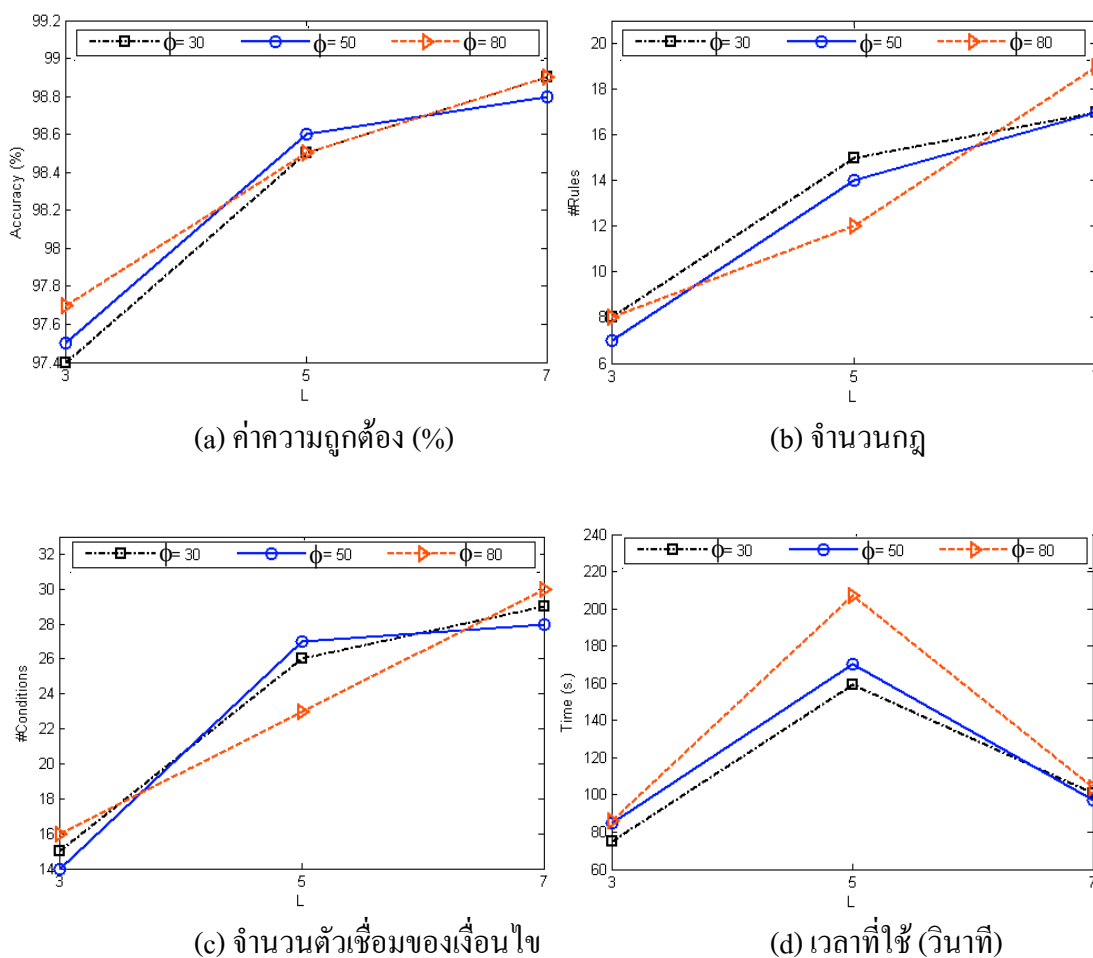
else Benign

ภาพประกอบ 4.3 ตัวอย่างกฎของฐานข้อมูลโรคมะเร็งเต้านมเมื่อ $\square = 20$ $\phi = 50$ และ $L = 7$

4.1.3 เปรียบเทียบผลการทดลองและวิจารณ์ผลการทดลอง

เมื่อกำหนดให้ค่าร้อยละของน้ำหนักรับได้ $\square = 0$ (ไม่มีการตัดโหนดก่อนสกัดกฎภาษาธรรมชาติ) $\square = 20$ และ $\square = 40$ จะได้ผลการทดลองดังนี้

1. ค่าร้อยละที่อมรับได้ $\square = 0$ (ไม่มีการตัดโหนดก่อนสกัดกฎภาษาธรรมชาติ) โครงสร้างโครงข่ายประสาทเทียมหลังผ่านการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) คือ 9:18:1 เมื่อกำหนดให้ค่าความกำวมที่ยอมรับได้ $\phi = 30$ $\phi = 50$ และ $\phi = 80$ กำหนดจำนวนรูปแบบภาษาธรรมชาติ $L = 3$ $L = 5$ และ $L = 7$ จะได้ผลการทดลองดังภาพประกอบ 4.4



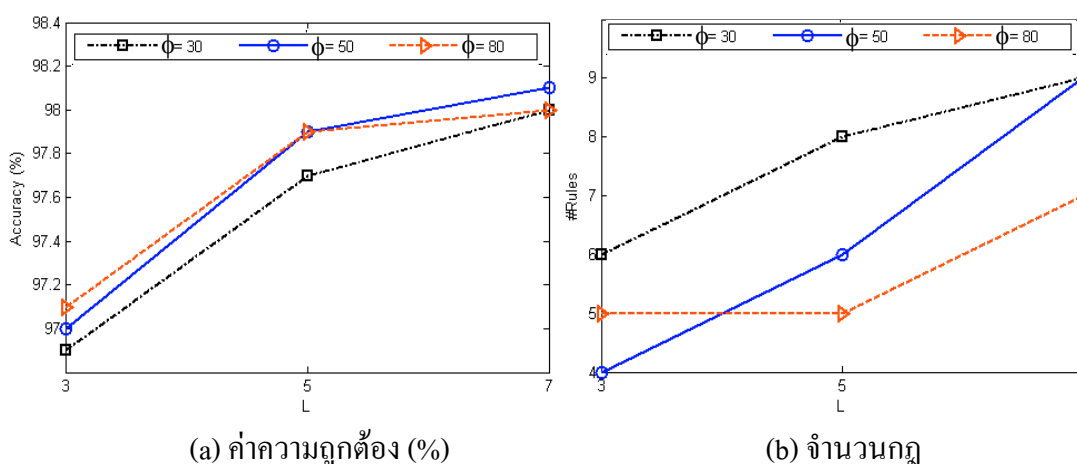
ภาพประกอบ 4.4 กราฟของผลการทดลองการสกัดกฎภาษาธรรมชาติด้วยวิธีการ RENN-P-FIDR ของฐานข้อมูลโรคมะเร็งเต้านม เมื่อกำหนดค่าร้อยละที่ยอมรับได้ $\square = 0$

จากภาพประกอบ 4.4 (a) พบว่าค่าความถูกต้องจะขึ้นอยู่กับจำนวนรูปแบบภาษาธรรมชาติ L กล่าวคือเมื่อกำหนดช่วง L ให้มีค่าจำนวนช่วงมากขึ้น (เช่น $L = 7$) หรือมี

การแบ่งช่วงละเอียดขึ้น จะได้ค่าความถูกต้องที่สูงกว่าค่า L ที่มีจำนวนช่วงน้อยกว่า (เช่น $L = 3$) หรือมีการแบ่งช่วงที่กว้างกว่าในการแทนค่าข้อมูล แต่อย่างไรก็ตามจำนวนกฎและจำนวนตัวเชื่อมของเงื่อนไขเพิ่มขึ้นตามค่า L ที่เพิ่มขึ้นด้วย ดังภาพประกอบ 4.4(b) และภาพประกอบ 4.4(c) นอกจากนี้จากภาพประกอบ 4.4(d) พบว่าเมื่อกำหนดให้ $L = 7$ จะใช้เวลาในการสกัดกฎภาษาธรรมชาติน้อยกว่าเมื่อกำหนดให้ $L = 5$ ทั้งนี้เนื่องจากโครงข่ายประสาทเทียมมีการเรียนรู้ได้เร็วกว่า

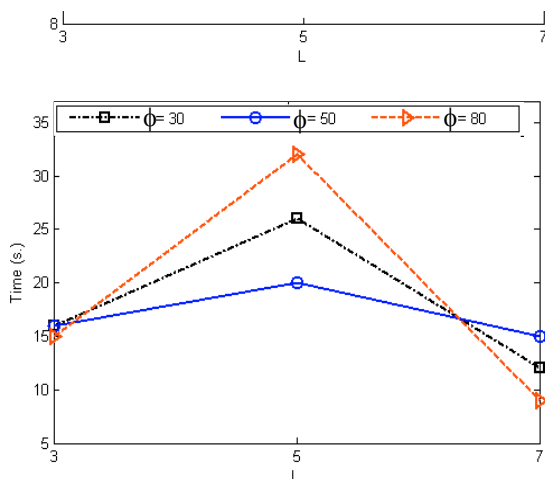
จากผลการทดลองพบว่าถึงแม้ว่ากำหนดให้ค่าความกำกวมที่ยอมรับได้ ϕ มีค่าแตกต่างกันดังนี้ $\phi = 30$ $\phi = 50$ $\phi = 80$ แต่คงยังได้ค่าความถูกต้อง จำนวนกฎ จำนวน ตัวเชื่อมของเงื่อนไข และเวลาที่ใช้ในการสกัดกฎใกล้เคียงกันเนื่องจากช่วงภาษาธรรมชาติที่แบ่งได้มีค่าของช่วงใกล้เคียงกันมาก ซึ่งเกิดมาจากการรวมช่วง นอกจากนี้ค่าความถูกต้อง จำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข และเวลาที่ใช้ในการสกัดกฎ ยังขึ้นอยู่กับขั้นตอนการสอนโครงข่าย ประสาทเทียม ซึ่งส่งผลมายังขั้นตอนการสร้างกฎอีกด้วย

2. ค่าร้อยละที่ยอมรับได้ $\square = 20$ โครงสร้างโครงข่ายประสาทเทียมหลังผ่านการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) คือ 8:8:1 โดยที่ตัวแปรข้อมูลเข้าที่ 5 ถูกตัดออกจากโครงข่ายประสาทเทียม เมื่อกำหนดให้ค่าความกำกวมที่ยอมรับได้ $\phi = 30$ $\phi = 50$ และ $\phi = 80$ กำหนดจำนวนรูปแบบภาษาธรรมชาติ $L = 3$ $L = 5$ และ $L = 7$ จะได้ผลการทดลองดังภาพประกอบ 4.5 ซึ่งมีแนวโน้มเหมือนกับผลการทดลองเมื่อกำหนดให้ $\square = 0$ กล่าวคือ ถ้ากำหนดให้รูปแบบภาษาธรรมชาติ L เพิ่มขึ้นจะทำให้ได้กฎที่มีค่าความถูกต้องของข้อมูล จำนวนกฎ และจำนวนตัวเชื่อมของเงื่อนไขเพิ่มขึ้นดังภาพประกอบ 4.5(a) 4.5(b) และ 4.5(c)

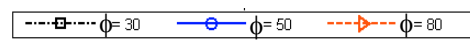


(a) ค่าความถูกต้อง (%)

(b) จำนวนกฎ



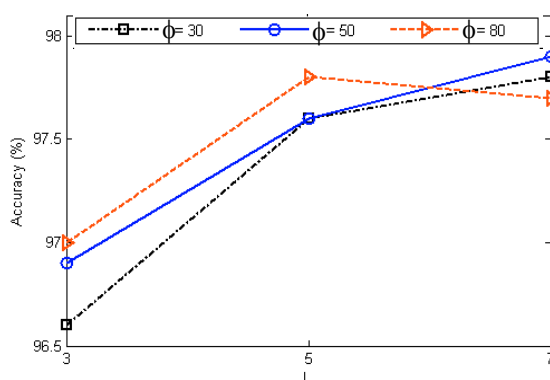
(c) จำนวนตัวเชื่อมของเงื่อนไข



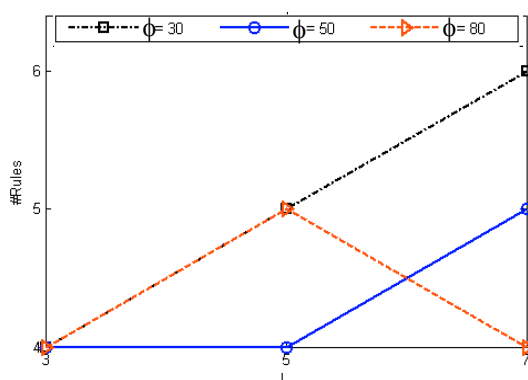
(d) เวลาที่ใช้ (วินาที)

ภาพประกอบ 4.5 กราฟของผลการทดลองการสกัดกฎภาษาธรรมชาติด้วยวิธีการ RENN-P-FIDR ของฐานข้อมูลโรคมะเร็งเต้านม เมื่อค่าร้อยละที่ยอมรับได้ $\square = 20$

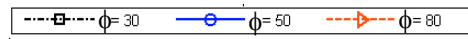
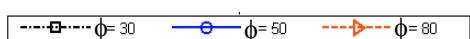
3. ค่าร้อยละที่ยอมรับได้ $\square = 40$ โครงสร้างโครงข่ายประสาทเทียมหลังผ่านการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่านำหนักสูงสุด (NNP-MFW) คือ 7:5:1 โดยที่ ตัวแปรข้อมูลเข้าที่ 5 และตัวแปรข้อมูลเข้าที่ 9 ถูกตัดออกจากโครงข่ายประสาทเทียม เมื่อกำหนดให้ค่าความกำกวมที่ยอมรับได้ $\phi = 30$ $\phi = 50$ และ $\phi = 80$ กำหนดจำนวนรูปแบบภาษาธรรมชาติ $L = 3$ $L = 5$ และ $L = 7$ จะได้ผลการทดลองดังภาพประกอบ 4.6 ซึ่งมีแนวโน้มเหมือนกับผลการทดลองเมื่อกำหนดให้ $\square = 0$ และ $\square = 20$ กล่าวคือ ถ้ากำหนดให้รูปแบบภาษาธรรมชาติ L เพิ่มขึ้นจะทำให้ได้กฎที่มีค่าความถูกต้องของข้อมูล จำนวนกฎ และจำนวนตัวเชื่อมของเงื่อนไขเพิ่มขึ้นดังภาพประกอบ 4.6(a) 4.6(b) และ 4.6(c)

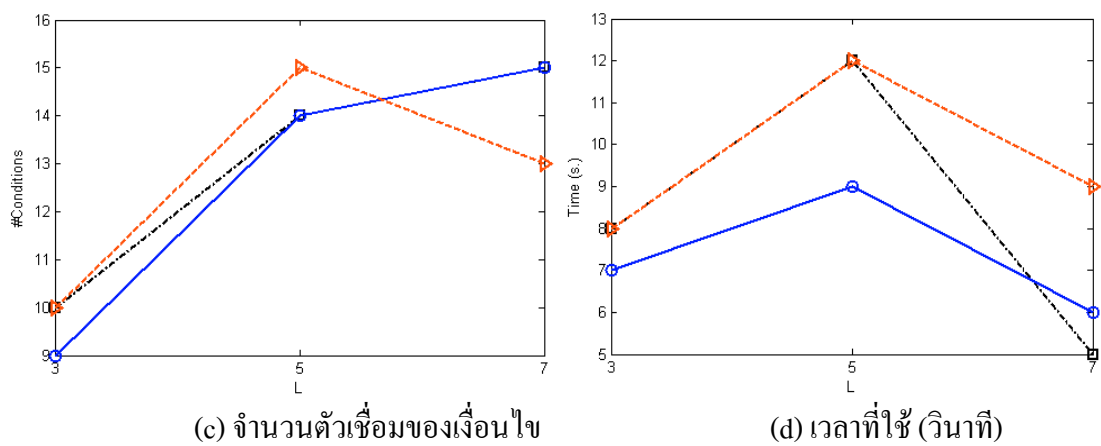


(a) ค่าความถูกต้อง (%)



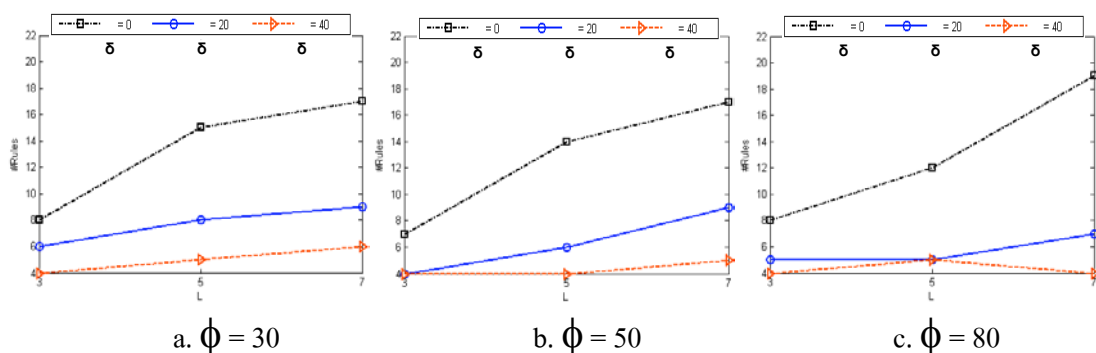
(b) จำนวนกฎ



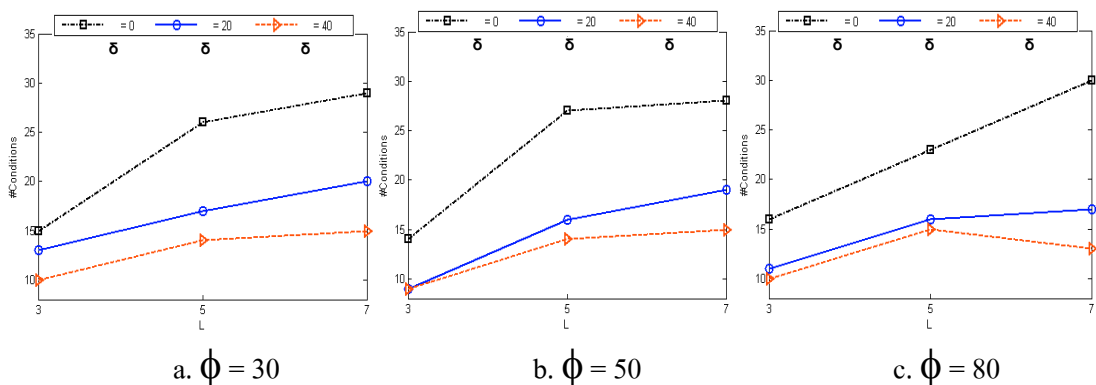


ภาพประกอบ 4.6 กราฟของผลการทดลองการสกัดคุณภาษาธรรมชาติด้วยวิธีการ RENN-P-FIDR ของฐานข้อมูลโรคมะเร็งเต้านม เมื่อค่าร้อยละที่ยอมรับได้ $\square = 40$

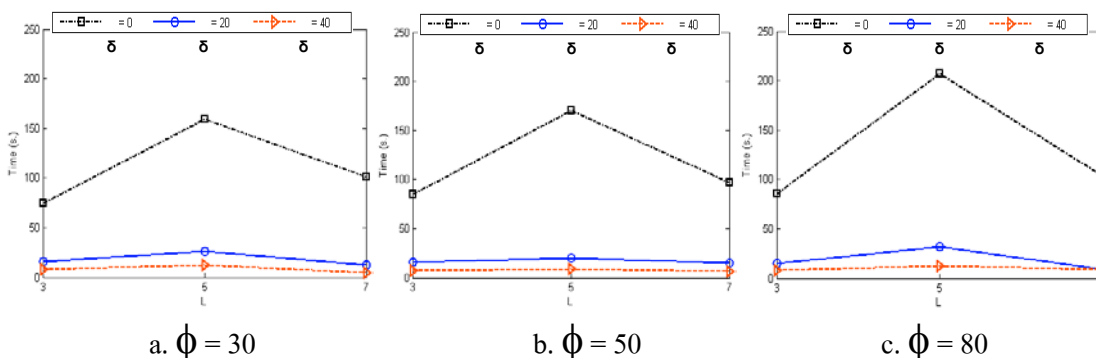
สามารถเปรียบเทียบผลการทดลองจากการกำหนดค่าร้อยละที่ยอมรับได้ $\square = 0$ $\square = 20$ และ $\square = 40$ ของจำนวนกฎแสดงได้ดังภาพประกอบ 4.7 ของจำนวนตัวเชื่อมของ เส้นใยแสดงได้ดังภาพประกอบ 4.8 ของเวลาที่ใช้แสดงได้ดังภาพประกอบ 4.9 และของค่าความถูกต้องแสดงได้ดังภาพประกอบ 4.10 ซึ่งพบว่าเมื่อกำหนดให้ค่าร้อยละที่ยอมรับได้ \square เพิ่มขึ้นจะทำให้ได้จำนวนกฎ และจำนวนตัวเชื่อมของเส้นใยลดลง เวลาที่ใช้ในการสกัดกั้นลดลงมาก แต่ค่าความถูกต้องลดลงเพียงเล็กน้อย



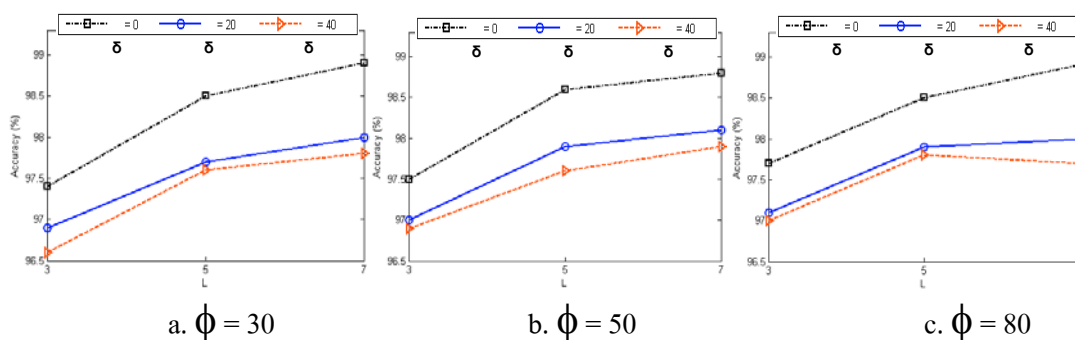
ภาพประกอบ 4.7 จำนวนกฎของฐานข้อมูลโรคมะเร็งเต้านม



ภาพประกอบ 4.8 จำนวนตัวเชื่อมของเส้นใยของฐานข้อมูลโรคมะเร็งเต้านม



ภาพประกอบ 4.9 เวลาที่ใช้ในการสกัดกั้น (วินาที) ของฐานข้อมูลโรคมะเร็งเต้านม



ภาพประกอบ 4.10 ค่าความถูกต้อง (%) ของฐานข้อมูลโรคมะเร็งเต้านม

ตารางที่ 4.11 แสดงผลการทดลองการสกัดกฎภาษาธรรมชาติ RENN-P-FIDR ของฐานข้อมูลโรคมะเร็งเต้านมทั้งหมด ตัวแปรที่มีผลต่อการสกัดกฎภาษาธรรมชาติ RENN-P-FIDR มี 3 ตัวแปรคือ ค่าร้อยละของน้ำหนักที่ยอมรับได้ \square จำนวนรูปแบบภาษาธรรมชาติ L แลค่าความกำกวมที่ยอมรับได้ ϕ ซึ่งแต่ละตัวแปรแนวโน้มผลการทดลองดังนี้

1. ค่าร้อยละของน้ำหนักที่ยอมรับได้ \square ถ้ากำหนดให้ \square มีค่ามากจะตัดโหนดโครงข่ายประสาทเทียมได้มาก กฎที่สกัดได้มีจำนวนกฎและจำนวนตัวเชื่อมของเงื่อนไขลดลง เวลาที่ใช้ในการสกัดกฎลดลงมาก แต่อย่างไรก็ตามค่าความถูกต้องอาจลดลงไปด้วย ทั้งนี้เราสามารถกำหนดได้ว่าจะยอมรับค่าความถูกต้องลดลงได้จนถึงระดับใด ซึ่งจากผลการทดลองควรกำหนดค่า \square ให้อยู่ในช่วง 0 ถึง 40

2. จำนวนรูปแบบภาษาธรรมชาติ L ถ้ากำหนดให้ L มีค่าจำนวนช่วงมากขึ้น (เช่น $L = 7$) หรือมีการแบ่งช่วงละเอียดขึ้น จะได้ค่าความถูกต้องที่สูงกว่าค่า L ที่มีจำนวนช่วงน้อยกว่า (เช่น $L = 3$) หรือมีการแบ่งช่วงที่กว้างกว่าในการแทนค่าข้อมูล แต่อย่างไรก็ตามจำนวนกฎและจำนวนตัวเชื่อมของเงื่อนไขเพิ่มขึ้นตามค่า L ที่เพิ่มขึ้นด้วย

3. ค่าความกำกวมที่ยอมรับได้ ϕ ถ้ากำหนดให้ $\phi = 30$ หมายถึงจะแบ่งช่วงข้อมูลโดยยอมให้ค่าความกำกวมของช่วงไม่เกินร้อยละ 30 จากผลการทดลองแม้ว่ากำหนดให้ ค่าความกำกวมที่ยอมรับได้ ϕ มีค่าแตกต่างกันคือ $\phi = 30$ $\phi = 50$ $\phi = 80$ พบว่าได้ ค่าความถูกต้อง จำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข และเวลาที่ใช้ในการสกัดกฎใกล้เคียงกันเนื่องจากช่วงภาษาธรรมชาติที่แบ่งได้มีค่าของช่วงใกล้เคียงกัน

ตารางที่ 4.12 แสดงผลการทดลองเปรียบเทียบการสกัดกฎภาษาธรรมชาติ RENN-P-FIDR กับวิธีการอื่นๆ เมื่อกำหนดให้ $L=7$ และ $\phi = 50$ พบว่าวิธีการสกัดกฎภาษาธรรมชาติ RENN-P-FIDR ได้ผลการทดลองที่ดีกว่า NEFCLASS [3] ที่ได้กฎภาษาธรรมชาติ (Linguistic Rule)

เช่นกัน และการสกัดกฎภาษาธรรมชาติ RENN-P-FIDR ได้ผลการทดลองที่ดีกว่า DIMLP [1] และ J48 [24] ที่ใช้กฎทั่วไป (Crisp Rule) ในการแทนค่ากฎ

ตารางที่ 4.11 ผลการทดลองการสกัดกฎภาษาธรรมชาติ RENN-P-FIDR
ของฐานข้อมูลโรคมะเร็งเต้านม

δ Method		L	ϕ	Total Accuracy (%)	#Rules	#Con.	Time (s.)
RENN-P-FIDR	0	3	30	97.4	8	15	75
RENN-P-FIDR	0	5	30	98.5	15	26	159
RENN-P-FIDR	0	7	30	98.9	17	29	101
RENN-P-FIDR	0	3	50	97.5	7	14	85
RENN-P-FIDR	0	5	50	98.6	14	27	170
RENN-P-FIDR	0	7	50	98.8	17	28	97
RENN-P-FIDR	0	3	80	97.7	8	16	86
RENN-P-FIDR	0	5	80	98.5	12	23	207
RENN-P-FIDR	0	7	80	98.9	19	30	104
RENN-P-FIDR	20	3	30	96.9	6	13	16
RENN-P-FIDR	20	5	30	97.7	8	17	26
RENN-P-FIDR	20	7	30	98.0	9	20	12
RENN-P-FIDR	20	3	50	97.0	4	9	16
RENN-P-FIDR	20	5	50	97.9	6	16	20
RENN-P-FIDR	20	7	50	98.1	9	19	15
RENN-P-FIDR	20	3	80	97.1	5	11	15
RENN-P-FIDR	20	5	80	97.9	5	16	32
RENN-P-FIDR	20	7	80	98.0	7	17	9
RENN-P-FIDR	40	3	30	96.6	4	10	8
RENN-P-FIDR	40	5	30	97.6	5	14	12
RENN-P-FIDR	40	7	30	97.8	6	15	5
RENN-P-FIDR	40	3	50	96.9	4	9	7
RENN-P-FIDR	40	5	50	97.6	4	14	9

RENN-P-FIDR	40	7	50	97.9	5	15	6
RENN-P-FIDR	40	3	80	97.0	4	10	8
RENN-P-FIDR	40	5	80	97.8	5	15	12
RENN-P-FIDR	40	7	80	97.7	4	13	9

ตารางที่ 4.12 ผลการทดลองเมื่อเปรียบเทียบการสกัดกฎด้วย RENN-P-FIDR กับขั้นตอนวิธีอื่นๆ ของฐานข้อมูลโรคมะเร็งเต้านม

Method	Type	#Rules	#Con.	Total Accuracy (%)
RENN-P-FIDR ($\square = 0, L=7, \phi =50$)	Linguistic Rule	17	28	98.8
RENN-P-FIDR ($\square = 20, L=7, \phi =50$)	Linguistic Rule	9	19	98.1
RENN-P-FIDR ($\square = 40, L=7, \phi =50$)	Linguistic Rule	5	15	97.9
C-MLP2LN [3]	Crisp Rule	-	-	99.0
DIMLP [1]	Crisp Rule	-	-	97.0
NEFCLASS [3]	Linguistic Rule	-	-	96.5
J48 [24]	Crisp Rule	28	29	94.4

4.2 ฐานข้อมูลโรคหัวใจจากคลินิกคลิฟแลนด์

ฐานข้อมูลโรคหัวใจจากคลินิกคลิฟแลนด์ (Heart Disease from Cleveland Clinic Database) ประกอบด้วยข้อมูล 2 กลุ่มคือ กลุ่มผู้ที่ไม่เป็นโรคหัวใจ (Absence) มี 164 คน และกลุ่มผู้ที่เป็นโรคหัวใจ (Presence) มี 139 คน รวมทั้งหมด 303 คน ตัวแปรข้อมูลเข้ามีทั้งหมด 13 ตัวแปร ดังตารางที่ 4.13

4.2.1 การตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW)

ในการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุดมีขั้นตอนในการทำงาน 3 ขั้นตอนมีรายละเอียดดังนี้

ขั้นตอนที่ 1 การเตรียมข้อมูล เนื่องจากในโรคหัวใจจากคลินิกคลิฟแลนด์มีแถวข้อมูลที่มีค่าของตัวแปรข้อมูลเข้าที่ 12 หายไป (Missing Values) 4 แถว และแถวข้อมูลที่มีค่าของตัวแปรข้อมูลเข้าที่ 13 หายไป (Missing Values) 2 แถว ดังนั้นจึงแทนค่าตัวแปรข้อมูลเข้าที่หายไปด้วยค่าเฉลี่ยของกลุ่ม (Class Mean) ซึ่งสำหรับตัวแปรข้อมูลเข้าที่ 12 มีค่าเฉลี่ยเท่ากับ 0 สำหรับตัวแปรข้อมูลเข้าที่ 13 ในกลุ่มที่ไม่เป็นโรคหัวใจมีค่าเฉลี่ยเท่ากับ 3 และกลุ่มที่เป็นโรคหัวใจมีค่าเฉลี่ยเท่ากับ 7 ในการทดลองมีการเลือกชุดข้อมูลเพื่อใช้ในการสอน (Training set) 90 % และใช้ในการทดสอบ (Testing set) 10 %

ตารางที่ 4.13 รายละเอียดของฐานข้อมูลโรคหัวใจจากคลินิกคลิฟแลนด์

ตัวแปร	ชื่อตัวแปร	ชนิด	ค่าที่เป็นไปได้
x ₁	age in years	Integer	29-76
x ₂	gender	Nominal Scale	1 = male 2 = female
x ₃	chest pain type	Nominal Scale	1 = typical angina 2 = atypical angina 3 = non-anginal pain 4 = asymptomatic
x ₄	resting blood pressure (in mm Hg)	Integer	94-200
x ₅	serum cholestorol in mg/dl	Integer	126-417
x ₆	fasting blood sugar > 120 mg/dl	Nominal Scale	1= true 0 = false
x ₇	resting electrocardiographic results	Nominal Scale	0 = normal 1= having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) 2 = showing probable or definite left ventricular hypertrophy by Estes' criteri
x ₈	maximum heart rate achieved	Integer	71-202
x ₉	exercise induced angina	Nominal Scale	1 = yes 0 = no
x ₁₀	ST depression induced by exercise relative to rest	Real	0-6.2
x ₁₁	the slope of the peak exercise ST segment	Nominal Scale	1 = upsloping 2 = flat 3 = downsloping

x_{12}	number of major vessels (0-3) colored by flourosopy	Nominal Scale	0,1,2,3
x_{13}	thal	Nominal Scale	3 = normal 6 = fixed defect 7 = reversable defect
c	class	Nominal Scale	0 = Absence 1 = Presence

ขั้นตอนที่ 2 การสอนโครงข่ายประสาทเทียมด้วยการเรียนรู้แบบแพร่ย้อนกลับ กำหนดให้มีโหนดข้อมูลเข้า 13 โหนด โหนดซ่อน 13 โหนด และโหนดผลลัพธ์ 1 โหนด ดังนั้นโครงสร้างของโครงข่ายประสาทเทียมคือ 13:13:1 โดยจะทำการสอนโครงข่ายประสาทเทียมจนกระทั่งครบจำนวนรอบที่ตั้งไว้ 200 รอบ (Epochs) หรือได้ค่าความผิดพลาด (MSE) น้อยกว่าค่าที่ยอมรับได้ซึ่งมีค่าเท่ากับ 0 ในขั้นตอนนี้จะได้ค่าน้ำหนักของโครงข่ายประสาทเทียมเพื่อนำไปตัดโหนด

ขั้นตอนที่ 3 ขั้นตอนวิธีการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด ตัวอย่างของโครงสร้างโครงข่ายประสาทเทียมที่ผ่านการตัดโหนดเมื่อกำหนดให้ค่าร้อยละของน้ำหนักที่ยอมรับได้ $\square = 20$ คือ 11:7:1 กล่าวคือเหลือโหนดข้อมูลเข้า 11 โหนด โหนดซ่อน 7 โหนด และโหนดผลลัพธ์ 1 โหนด โดยมีโหนดข้อมูลเข้าที่ 1 และ 8 ถูกตัดออกไป

4.2.2 การสกัดคุณลักษณะจากโครงข่ายประสาทเทียมโดยการแทนค่าความถี่ของช่วงข้อมูล (RENN-FIDR)

ในการสกัดคุณลักษณะจากโครงข่ายประสาทเทียมโดยการแทนค่าความถี่ของช่วงข้อมูลจะพิจารณาเฉพาะโหนดซ่อนและโหนดข้อมูลเข้าที่เสนอแนะจากขั้นตอนหลักที่ I การตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) ในการทดลองจะเลือกสุ่มข้อมูลแบบความเที่ยงตรง 10 กลุ่ม (10-Fold Cross Validation) เพื่อให้ข้อมูลทุกตัวได้ถูกนำมาเป็นข้อมูลชุดทดสอบ (Testing Set) ซึ่งมีขั้นตอนในการทำงาน 3 ขั้นตอนมีรายละเอียดดังนี้

ขั้นตอนที่ 4 การสร้างรูปแบบภาษาธรรมชาติจากความถี่ของช่วงข้อมูล ตัวอย่างช่วงภาษาธรรมชาติที่แบ่งได้ เมื่อ $\square = 20$ $\phi = 50$ และ $L = 7$ แสดงดังตารางที่ 4.14 ซึ่งมี ตัวแปรข้อมูลบางตัวแปรที่สามารถแบ่งช่วงภาษาธรรมชาติได้จำนวนน้อยกว่าค่า L ที่กำหนดไว้เช่นตัวแปร

ข้อมูลเข้าที่ 2 สามารถแบ่งช่วงได้ 2 ช่วงเท่านั้น เนื่องจากตัวแปรข้อมูลเข้าที่ 2 มีค่าเป็น 0 หรือ 1 เท่านั้น

ตารางที่ 4.14 ช่วงของข้อมูลเมื่อ $\square = 20$ $\phi = 50$ และ $L = 7$ ของฐานข้อมูลโรคหัวใจจากคลินิก
ลิฟแลนด์

x	tiny	very small	small	medium	large	very large	huge
x_2	[0,1)	[1,1]	-	-	-	-	-
x_3	[1,2)	[2,3)	[3,4)	[4,4]	-	-	-
x_4	[94,100)	[100,108)	[108,110)	[110,112)	[112,114)	[114,115)	[115,200]
x_5	[126,341)	[341,342)	[342,353)	[353,354)	[354,407)	[407,417)	[417,564]
x_6	[0,1)	[1,1]	-	-	-	-	-
x_7	[0,1)	[1,2)	[2,2]	-	-	-	-
x_9	[0,1)	[1,1]	-	-	-	-	-
x_{10}	[0,0.1)	[0.1,0.5)	[0.5,0.6)	[0.6,0.7)	[0.7,3.6)	[3.6,4.4)	[4.4,6.2]
x_{11}	[1,2)	[2,3)	[3,3]	-	-	-	-
x_{12}	[0,1)	[1,2)	[2,3)	[3,3]	-	-	-
x_{13}	[3,6)	[6,7)	[7,7]	-	-	-	-

ขั้นตอนที่ 5 การสอนโครงข่ายประสาทเทียมด้วยรูปแบบภาษาธรรมชาติ โดยมีการเรียนรู้แบบแพร่ย้อนกลับ โหนดข้อมูลเข้า และโหนดซ่อนจะกำหนดจากการเสนอแนะจาก ขั้นตอนการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) โดยจะทำการสอนโครงข่ายประสาทเทียมจนกระทั่งครบจำนวนรอบที่ตั้งไว้ 200 รอบ (Epochs) หรือได้ ค่าความผิดพลาด (MSE) น้อยกว่าค่าที่ยอมรับได้ซึ่งมีค่าเท่ากับ 0 ตัวอย่างเช่น ถ้าโหนดข้อมูลเข้า 11 โหนด และโหนดซ่อน 7 โหนด กรณีที่ $L = 7$ จะได้โครงสร้างของโครงข่ายประสาทเทียมเป็น 44:7:1 โดยมีโหนดข้อมูลเข้า 44 โหนดเพราะช่วงภาษาธรรมชาติที่แบ่งได้มี 44 ช่วงดังตารางที่ 4.14 น้ำหนักที่ได้จากการสอนโครงข่ายประสาทเทียมถูกนำไปใช้ในขั้นตอนที่ 6 ขั้นตอนวิธีการสร้างกฎโดยพิจารณาค่าน้ำหนักที่เป็นบวก

ขั้นตอนที่ 6 ขั้นตอนวิธีการสร้างกฎโดยพิจารณาค่าน้ำหนักที่เป็นบวก ตัวอย่างกฎในกรณีที่มีค่า $\square = 20$ $\phi = 50$ ที่ $L = 7$ แสดงได้ดังภาพประกอบ 4.11 ซึ่งจำนวนกฎที่สกัดได้คือ 6 กฎและจำนวนตัวเชื่อม (and หรือ or) ของเงื่อนไขที่สกัดได้คือ 7 ตัวเชื่อม ซึ่งมี ค่าความถูกต้อง 85.8 % (ทำนายถูกต้อง 260 คน และทำนายผิด 43 คน)

R_1 : If x_{12} = medium or x_{10} = huge then Presence

R_2 : If x_{13} = small and x_{11} = very small then Presence

R_3 : If x_3 = medium and x_9 = very small and x_{10} = large then Presence

R_4 : If x_{12} = very small and x_3 = medium then Presence

R_5 : If x_3 = medium and x_{12} = small then Presence

R_6 : If x_{12} = small and x_{10} = large then Presence

else Absence

ภาพประกอบ 4.11 ตัวอย่างกฎของฐานข้อมูลโรคหัวใจจากคลินิกคลิฟแลนด์เมื่อ

$$\square = 20 \quad \phi = 50 \quad \text{และ} \quad L = 7$$

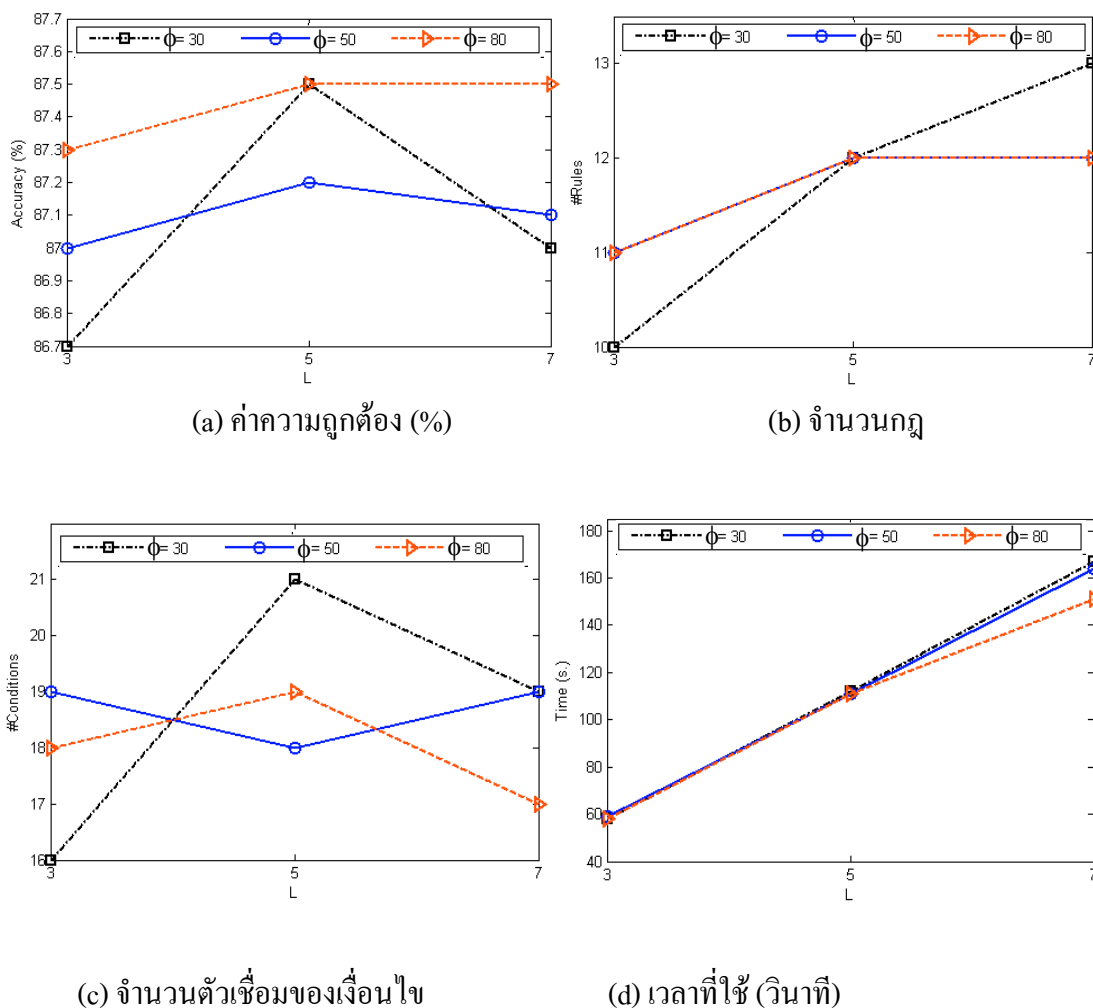
4.2.3 เปรียบเทียบผลการทดลองและวิจารณ์ผลการทดลอง

เมื่อกำหนดให้ค่าร้อยละของน้ำหนักที่ยอมรับได้ $\square = 0$ $\square = 20$ และ $\square = 40$ จะได้ผลการทดลองดังนี้

1. ค่าร้อยละที่ยอมรับได้ $\square = 0$ (ไม่มีการตัดโหนดก่อนสกัดกฎหมายธรรมชาติ) โครงสร้างโครงข่ายประสาทเทียมหลังผ่านการจัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) คือ 13:13:1 เมื่อกำหนดให้ค่าความกำกวมที่ยอมรับได้ $\phi = 30$ $\phi = 50$ และ $\phi = 80$ กำหนดจำนวนรูปแบบภาษาธรรมชาติ $L = 3$ $L = 5$ และ $L = 7$ จะได้ผลการทดลองดังภาพประกอบ 4.12

จากผลการทดลองพบว่าถึงแม้กำหนดให้จำนวนรูปแบบภาษาธรรมชาติ L มีค่าต่างกันดังนี้ $L = 7$ $L = 5$ และ $L = 3$ แต่ยังคงได้กฎที่มีค่าความถูกต้อง จำนวนกฎ และจำนวนตัวเชื่อมของเงื่อนไขใกล้เคียงกัน ดังภาพประกอบ 4.12(a) 4.12(b) และ 4.12(c) ตามลำดับ ทั้งนี้เนื่องจากช่วงภาษาธรรมชาติที่แบ่งได้มีค่าของช่วงใกล้เคียงกัน ทั้งนี้เพราะมีตัวแปรข้อมูลเข้า 8 ตัวแปร จาก 13 ตัวแปรที่มีลักษณะเป็น Nominal Scale (เกลื่อนามกำหนด) ที่สามารถแบ่งได้ช่วงภาษาธรรมชาติมากที่สุด 3 หรือ 4 ช่วงเท่านั้น ดังนั้นเมื่อกำหนดให้ $L = 5$ หรือ $L = 7$ ก็ได้ช่วงภาษาธรรมชาติ 3 หรือ 4 ช่วงเท่านั้น เพราะฉะนั้นจะเห็นได้ว่าเมื่อกำหนดให้ $L = 3$ $L = 5$ หรือ $L = 7$ จะได้ช่วงภาษาธรรมชาติใกล้เคียงกันมาก จากผลการทดลองยังพบอีกว่าถึงแม้ว่ากำหนดให้ค่าความกำกวมที่ยอมรับได้ ϕ มีค่าแตกต่างกันดังนี้ $\phi = 30$ $\phi = 50$ $\phi = 80$ แต่ยังคงยังได้กฎที่มีค่าความถูกต้องใกล้เคียงกัน เนื่องจากช่วงภาษาธรรมชาติที่แบ่งได้มีค่าของช่วงใกล้เคียงกัน นอกจากนี้ค่าความถูกต้องของกฎยัง

ขึ้นอยู่กับขั้นตอนการสอนโครงข่ายประสาทเทียม และขั้นตอนการสร้างกฎอีกด้วย จากภาพประกอบ 4.12(d) พบว่าเมื่อกำหนดให้จำนวนช่วงภาษาธรรมชาติสูงขึ้นจะใช้เวลาในการสกัดกฎนานขึ้น

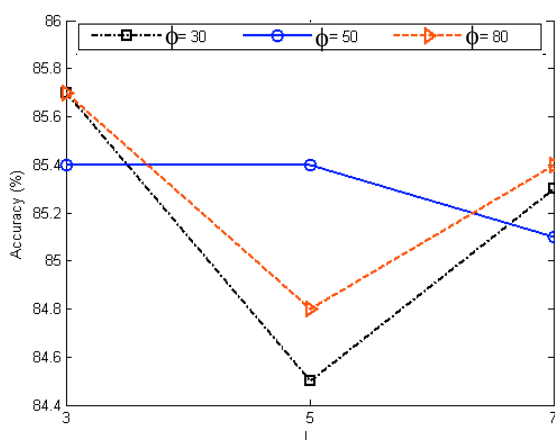


ภาพประกอบ 4.12 กราฟของผลการทดลองการสกัดกฎภาษาธรรมชาติด้วยวิธีการ RENN-P-FIDR ของฐานข้อมูลโรคหัวใจจากคลินิกคลิฟแลนด์ เมื่อค่าร้อยละที่ยอมรับได้

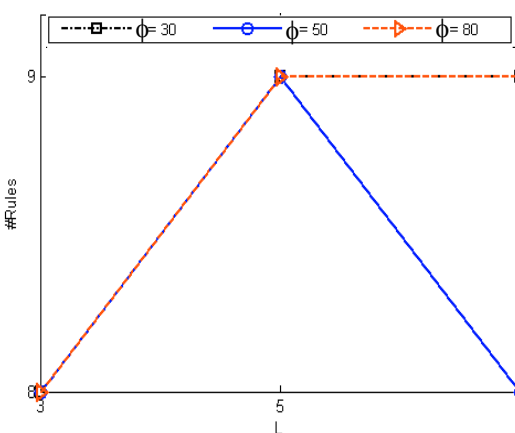
$$\square = 0$$

2. ค่าร้อยละที่ยอมรับได้ $\square = 20$ โครงสร้างโครงข่ายประสาทเทียมหลังผ่านการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) คือ 11:7:1 โดยที่ตัวแปรข้อมูลเข้าที่ 1 และ 8 ถูกตัดออกจากโครงข่ายประสาทเทียม เมื่อกำหนดให้ ค่าความกำกวมที่ยอมรับได้ $\phi = 30$ $\phi = 50$ และ $\phi = 80$ กำหนดจำนวนรูปแบบภาษาธรรมชาติ $L = 3$ $L = 5$ และ $L = 7$ จะได้ผลการทดลองดังภาพประกอบ 4.13 ซึ่งมีแนวโน้มเหมือนกับผลการทดลองเมื่อกำหนดให้ $\square = 0$ กล่าวคือถึงแม้กำหนดให้จำนวนรูปแบบภาษาธรรมชาติ L มีค่าต่างกันดังนี้ $L = 7$ $L = 5$ และ $L = 3$ แต่ยังคงได้กฎที่มีค่าความถูกต้อง จำนวนกฎ และจำนวนตัวเชื่อมของเงื่อนไขใกล้เคียงกัน

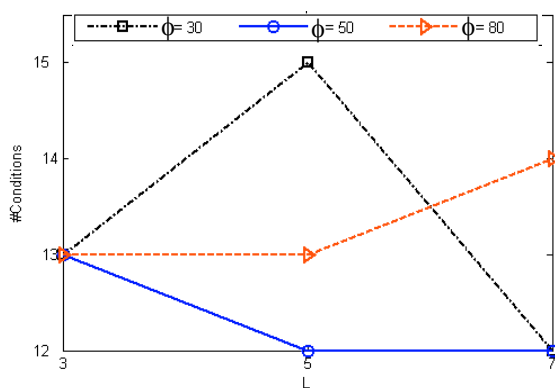
ดังภาพประกอบ 4.13(a) 4.13(b) และ 4.13(c) ตามลำดับ ทั้งนี้เนื่องจากช่วงภาษาธรรมชาติที่แบ่งได้มีค่าของช่วงใกล้เคียงกัน จากผลการทดลองยังพบอีกว่าถึงแม้ว่ากำหนดให้ค่าความกำกวมที่ยอมรับได้ ϕ มีค่าแตกต่างกันดังนี้ $\phi = 30$ $\phi = 50$ $\phi = 80$ แต่คงยังได้กฎที่มีค่าความถูกต้องใกล้เคียงกันเนื่องจากช่วงภาษาธรรมชาติที่แบ่งได้มีค่าของช่วงใกล้เคียงกัน นอกจากนี้จากภาพประกอบ 4.13(d) พบว่าเมื่อกำหนดให้จำนวนช่วงภาษาธรรมชาติสูงขึ้น จะใช้เวลาในการสกัดกฎภาษาธรรมชาติมากขึ้น



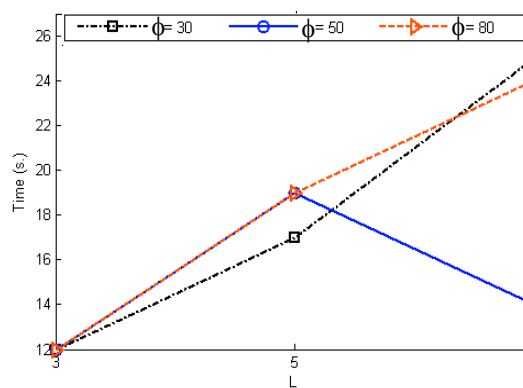
(a) ค่าความถูกต้อง (%)



(b) จำนวนกฎ



(c) จำนวนตัวเชื่อมของเงื่อนไข

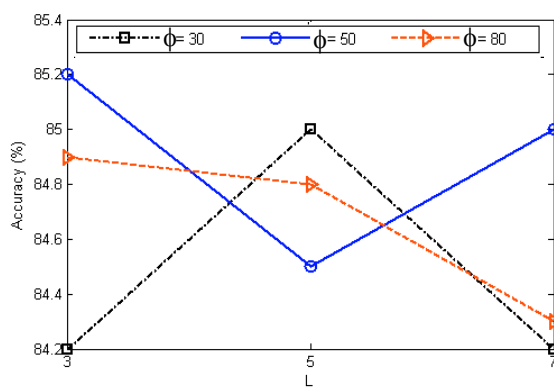


(d) เวลาที่ใช้ (วินาที)

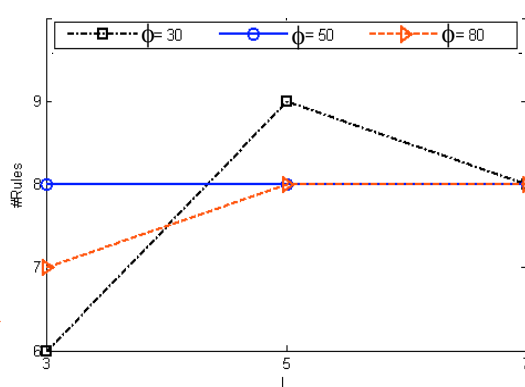
ภาพประกอบ 4.13 กราฟของผลการทดลองการสกัดกฎภาษาธรรมชาติด้วยวิธีการ RENN-P-FIDR ของฐานข้อมูลโรคหัวใจจากคลินิกฟแลนค์ เมื่อค่าร้อยละที่ยอมรับได้

$$\square = 20$$

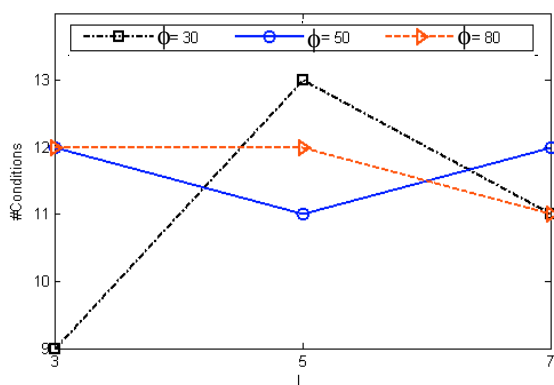
3. ค่าร้อยละที่ยอมรับได้ $\square = 40$ โครงสร้างโครงข่ายประสาทเทียมหลังผ่านการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) คือ 10:5:1 โดยที่ ตัวแปรข้อมูลเข้าที่ 1 5 และ 8 ถูกตัดออกจากโครงข่ายประสาทเทียม เมื่อกำหนดให้ค่า ความกำกวมที่ยอมรับได้ $\phi = 30$ $\phi = 50$ และ $\phi = 80$ กำหนดจำนวนรูปแบบภาษาธรรมชาติ $L = 3$ $L = 5$ และ $L = 7$ จะได้ผลการทดลองดังภาพประกอบ 4.14 ซึ่งมีแนวโน้มเหมือนกับผลการทดลองเมื่อกำหนดให้ $\square = 0$ และ $\square = 20$ กล่าวคือถึงแม้กำหนดให้จำนวนรูปแบบภาษาธรรมชาติ L มีค่าต่างกันดังนี้ $L = 7$ $L = 5$ และ $L = 3$ แต่ยังคงได้กฎที่มีค่าความถูกต้อง จำนวนกฎ และจำนวนตัวเชื่อมของเงื่อนไขใกล้เคียงกัน ดังภาพประกอบ 4.14(a) 4.14(b) และ 4.14(c) ตามลำดับ จากภาพประกอบ 4.14(d) พบว่าเมื่อกำหนดให้จำนวนช่วงภาษาธรรมชาติสูงขึ้น จะใช้เวลาในการสกัดกฎภาษาธรรมชาตินานขึ้น



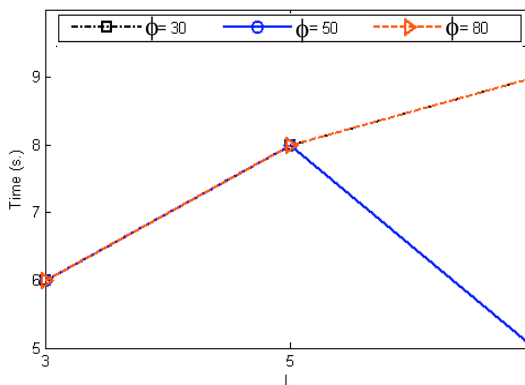
(a) ค่าความถูกต้อง (%)



(b) จำนวนกฎ



(c) จำนวนตัวเชื่อมของเงื่อนไข

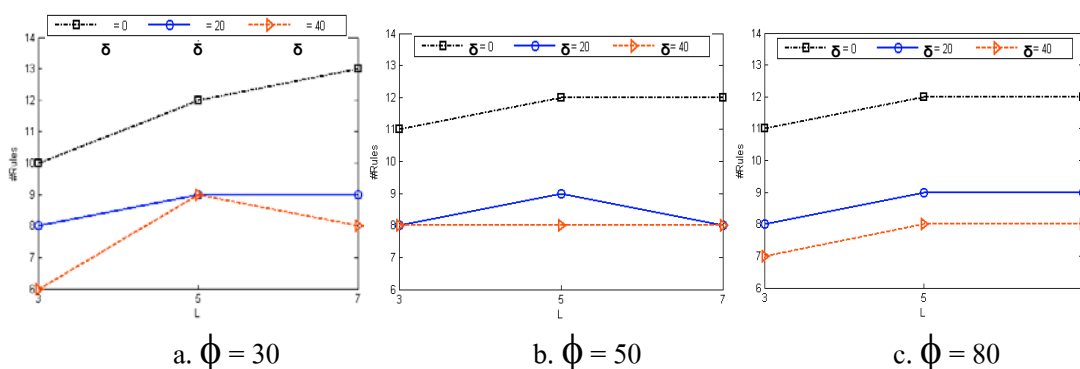


(d) เวลาที่ใช้ (วินาที)

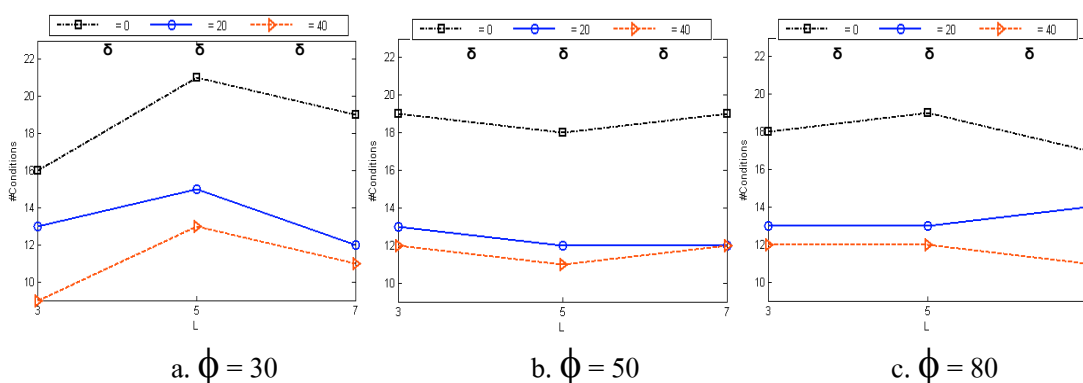
ภาพประกอบ 4.14 กราฟของผลการทดลองการสกัดกฎภาษาธรรมชาติด้วยวิธีการ RENN-P-FIDR ของฐานข้อมูลโรคหัวใจจากคลินิกคลิฟแลนด์ เมื่อค่าร้อยละที่ยอมรับได้

$$\square = 40$$

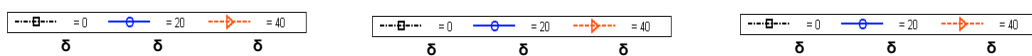
สามารถเปรียบเทียบผลการทดลองจากการกำหนดค่าร้อยละที่ยอมรับได้ $\square = 0$ $\square = 20$ และ $\square = 40$ ของจำนวนกฎแสดงได้ดังภาพประกอบ 4.15 ของจำนวนตัวเชื่อมของเงื่อนไขแสดงได้ดังภาพประกอบ 4.16 ของเวลาที่ใช้แสดงได้ดังภาพประกอบ 4.17 และของค่าความถูกต้องแสดงได้ดังภาพประกอบ 4.18 ซึ่งพบว่าเมื่อกำหนดให้ค่าร้อยละที่ยอมรับได้ \square เพิ่มขึ้นจะทำให้ได้จำนวนกฎ และจำนวนตัวเชื่อมของเงื่อนไขเวลาที่ใช้ในการสกัดกฎลดลงมากแต่ค่าความถูกต้องลดลงเพียงเล็กน้อย

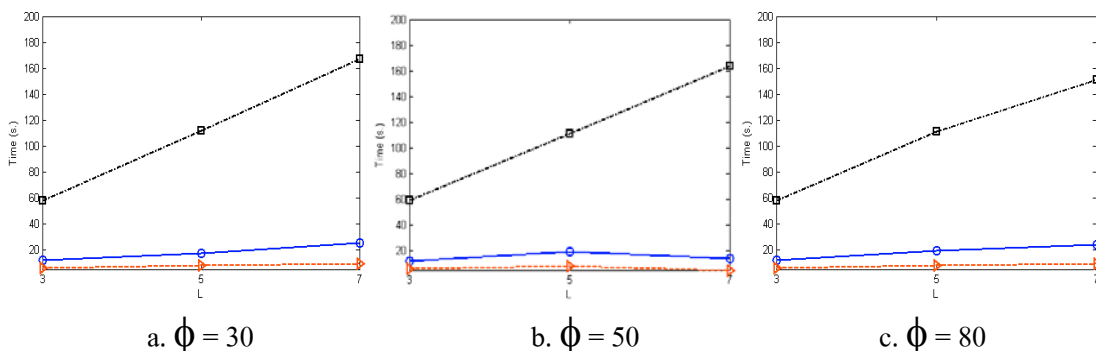


ภาพประกอบ 4.15 จำนวนกฎของฐานข้อมูลโรคหัวใจจากคลินิกคลิฟแลนด์

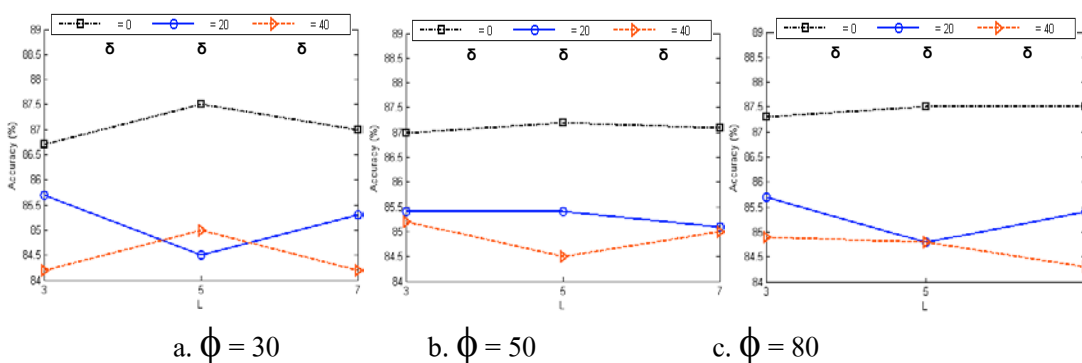


ภาพประกอบ 4.16 จำนวนตัวเชื่อมของเงื่อนไขของฐานข้อมูลโรคหัวใจจากคลินิกคลิฟแลนด์





ภาพประกอบ 4.17 เวลาที่ใช้ในการสกัดกฎ (วินาที) ของฐานข้อมูลโรคหัวใจจากคลินิกคลิฟแลนด์



ภาพประกอบ 4.18 ค่าความถูกต้อง (%) ของฐานข้อมูลโรคหัวใจจากคลินิกคลิฟแลนด์

ตารางที่ 4.15 แสดงผลการทดลองการสกัดกฎภาษาธรรมชาติ RENN-P-FIDR ของฐานข้อมูลโรคหัวใจจากคลินิกคลิฟแลนด์ทั้งหมด ตัวแปรที่มีผลต่อการสกัดกฎภาษาธรรมชาติ RENN-P-FIDR มี 3 ตัวแปรคือ ค่าร้อยละของน้ำหนักที่ยอมรับได้ \square จำนวนรูปแบบภาษาธรรมชาติ L และค่าความกำกวมที่ยอมรับได้ ϕ ซึ่งแต่ละตัวแปรมีแนวโน้มผลการทดลอง ดังนี้

1. ค่าร้อยละของน้ำหนักที่ยอมรับได้ \square ถ้ากำหนดให้ \square มีค่ามากจะตัดโหนดโครงข่ายประสาทเทียมได้มาก กฎที่สกัดได้มีจำนวนกฎและจำนวนตัวเชื่อมของเงื่อนไขลดลง เวลาที่ใช้ในการสกัดกฎลดลงมาก แต่อย่างไรก็ตามค่าความถูกต้องอาจลดลงตามไปด้วย ทั้งนี้เราสามารถกำหนดได้ว่า จะยอมรับค่าความถูกต้องลดลงได้จนถึงระดับใด ซึ่งจากผลการทดลองควรกำหนดค่า \square ให้อยู่ในช่วง 0 ถึง 40

2. จำนวนรูปแบบภาษาธรรมชาติ L ถ้ากำหนดให้ L มีค่าจำนวนช่วงมากขึ้น (เช่น $L = 7$) หรือมีการแบ่งช่วงละเอียดขึ้น จะได้ค่าความถูกต้อง จำนวนกฎและจำนวนตัวเชื่อม

ของเงื่อนไขที่ใกล้เคียงกับค่า L ที่มีจำนวนช่วงน้อยกว่า (เช่น $L = 3$) ทั้งนี้จะใช้เวลาในการสกัดคุณ
 นานกว่า

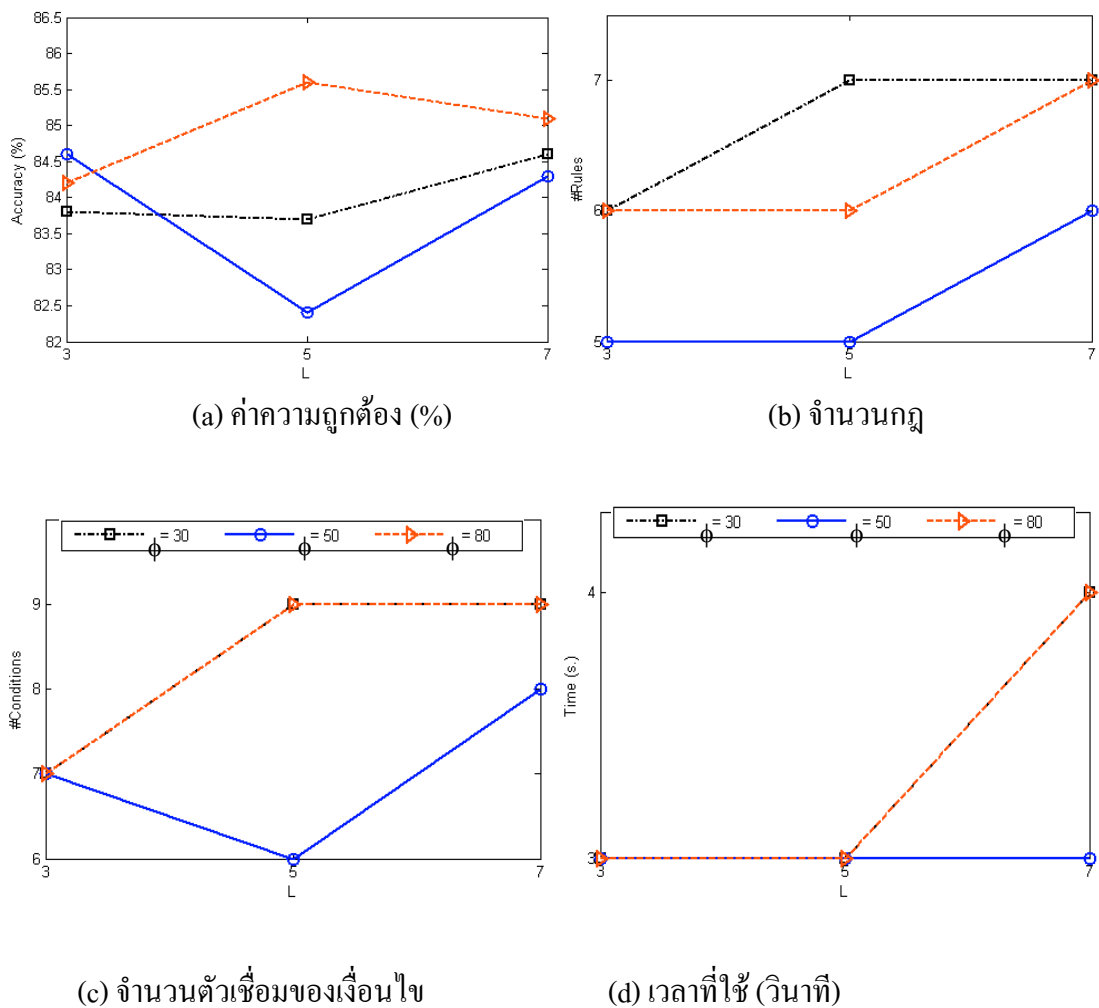
3. ค่าความกำกวมที่ยอมรับได้ ϕ ถ้ากำหนดให้ $\phi = 30$ หมายถึงจะแบ่งช่วง
 ข้อมูลโดยยอมให้ค่าความกำกวมของช่วงไม่เกินร้อยละ 30 จากผลการทดลองแม้ว่ากำหนดให้ ค่า
 ความกำกวมที่ยอมรับได้ ϕ มีค่าแตกต่างกันคือ $\phi = 30$ $\phi = 50$ $\phi = 80$ พบว่าได้ ค่าความถูก
 ต้อง จำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข และเวลาที่ใช้ในการสกัดกฎใกล้เคียงกันเนื่องจากช่วง
 ภาษาธรรมชาติที่แบ่งได้มีค่าของช่วงใกล้เคียงกัน

ตารางที่ 4.16 แสดงผลการทดลองเปรียบเทียบการสกัดกฎภาษาธรรมชาติ RENN-
 P-FIDR กับวิธีการอื่นๆ เมื่อกำหนดให้ $\square = 0$, $L=7$ และ $\phi = 50$ พบว่าวิธีการสกัดกฎภาษาธรรม
 ชาติ RENN-P-FIDR ได้ผลการทดลองที่ดีกว่า C-MLP2LN [3] DIMLP [1] และ J48 [24] เมื่อ
 กำหนดให้ $\square = 20$ หรือ $\square = 20$ $L=7$ และ $\phi = 50$ พบว่าวิธีการสกัดกฎภาษา ธรรมชาติ RENN-P-
 FIDR ได้ผลการทดลองที่ดีกว่า DIMLP [1] และ J48 [24]

ตารางที่ 4.15 ผลการทดลองการสกัดกฎด้วย RENN-P-FIDR ของฐานข้อมูลโรคหัวใจ
 จากคลินิกคลิฟแลนด์

δ	Method		L	ϕ	Total Accuracy (%)	#Rules	#Con.	Time (s.)
	RENN-P-FIDR	0	3	30	86.7	10	16	58
	RENN-P-FIDR	0	5	30	87.5	12	21	112
	RENN-P-FIDR	0	7	30	87.0	13	19	167
	RENN-P-FIDR	0	3	50	87.0	11	19	59
	RENN-P-FIDR	0	5	50	87.2	12	18	111
	RENN-P-FIDR	0	7	50	87.1	12	19	164
	RENN-P-FIDR	0	3	80	87.3	11	18	58
	RENN-P-FIDR	0	5	80	87.5	12	19	111
	RENN-P-FIDR	0	7	80	87.5	12	17	151
	RENN-P-FIDR	20	3	30	85.7	8	13	12
	RENN-P-FIDR	20	5	30	84.5	9	15	17
	RENN-P-FIDR	20	7	30	85.3	9	12	25
	RENN-P-FIDR	20	3	50	85.4	8	13	12

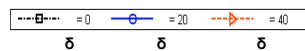
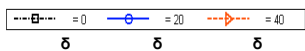
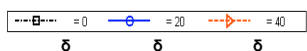
RENN-P-FIDR	20	5	50	85.4	9	12	19
RENN-P-FIDR	20	7	50	85.1	8	12	14
RENN-P-FIDR	20	3	80	85.7	8	13	12
RENN-P-FIDR	20	5	80	84.8	9	13	19
RENN-P-FIDR	20	7	80	85.4	9	14	24
RENN-P-FIDR	40	3	30	84.2	6	9	6
RENN-P-FIDR	40	5	30	85.0	9	13	8
RENN-P-FIDR	40	7	30	84.2	8	11	9
RENN-P-FIDR	40	3	50	85.2	8	12	6
RENN-P-FIDR	40	5	50	84.5	8	11	8
RENN-P-FIDR	40	7	50	85.0	8	12	5
RENN-P-FIDR	40	3	80	84.9	7	12	6
RENN-P-FIDR	40	5	80	84.8	8	12	8
RENN-P-FIDR	40	7	80	84.3	8	11	9

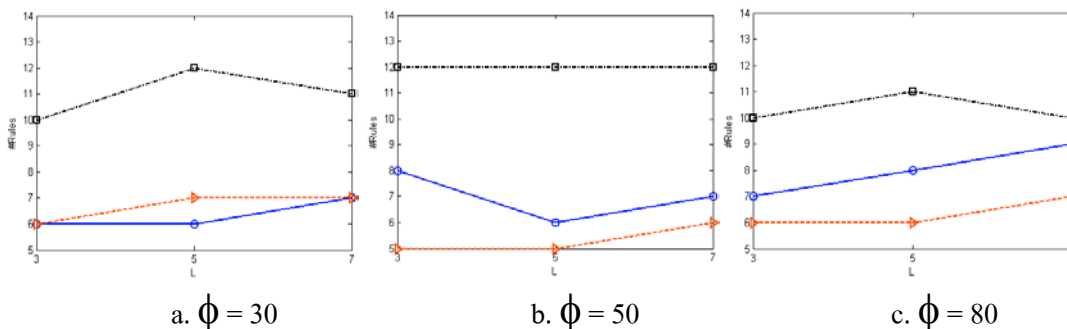


ภาพประกอบ 4.22 กราฟของผลการทดลองการสกัดกฎภาษาธรรมชาติด้วยวิธีการ RENN-P-FIDR ของฐานข้อมูลโรคหัวใจจากโปรเจกต์สแตตส์ออคยุโรป เมื่อค่าร้อยละที่ยอมรับได้

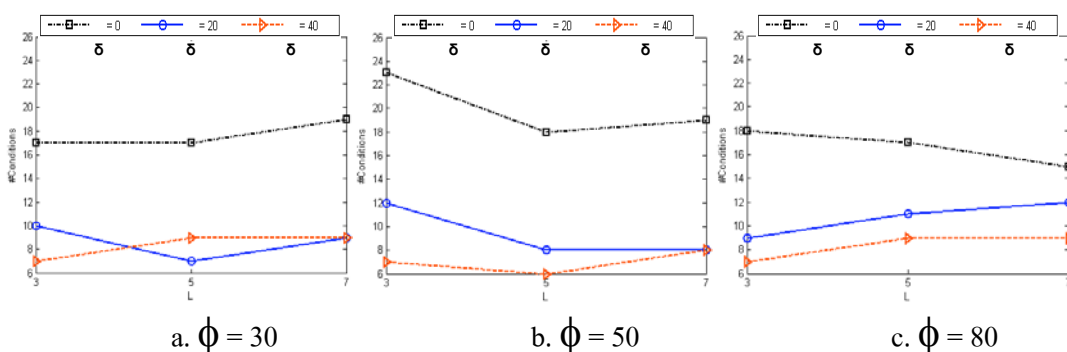
$\square = 40$

สามารถเปรียบเทียบผลการทดลองจากการกำหนดค่าร้อยละที่ยอมรับได้ $\square = 0$ $\square = 20$ และ $\square = 40$ ของจำนวนกฎแสดงได้ดังภาพประกอบ 4.23 ของจำนวนตัวเชื่อมของเงื่อนไขแสดงได้ดังภาพประกอบ 4.24 ของเวลาที่ใช้แสดงได้ดังภาพประกอบ 4.25 และของ ค่าความถูกต้องแสดงได้ดังภาพประกอบ 4.26 ซึ่งพบว่าเมื่อกำหนดให้ค่าร้อยละที่ยอมรับได้ \square เพิ่มขึ้นจะทำให้ได้จำนวนกฎ และจำนวนตัวเชื่อมของเงื่อนไขลดลงเวลาที่ใช้ในการสกัดกฎลดลงแต่ค่าความถูกต้องลดลงเพียงเล็กน้อย

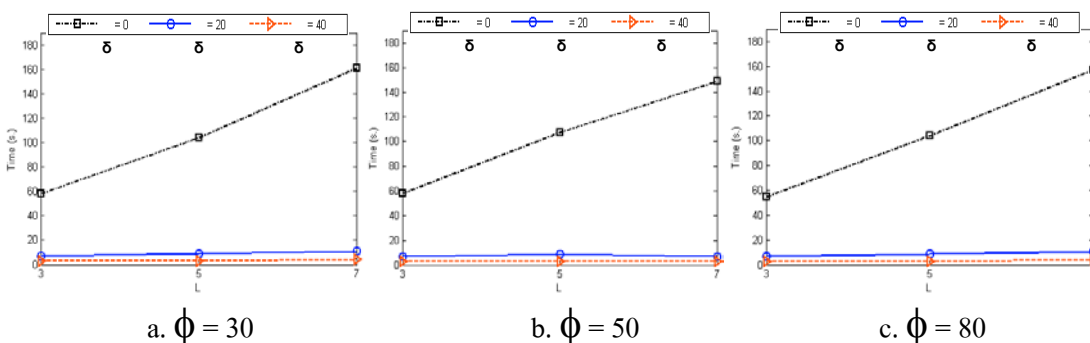




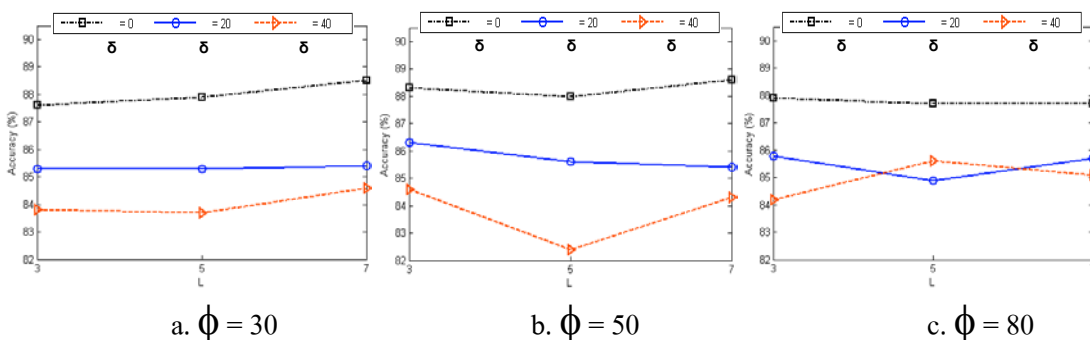
ภาพประกอบ 4.23 จำนวนกฎของฐานข้อมูลโรคหัวใจจากโปรเจกต์สแตตส์ออคยูโรป



ภาพประกอบ 4.24 จำนวนตัวเชื่อมของเงื่อนไขของฐานข้อมูลโรคหัวใจจากโปรเจกต์สแตตส์ออคยูโรป



ภาพประกอบ 4.25 เวลาที่ใช้ในการสกัดกฎ (วินาที) ของฐานข้อมูลโรคหัวใจจากโปรเจกต์สแตตส์ออคยูโรป



ภาพประกอบ 4.26 ค่าความถูกต้อง (%) ของฐานข้อมูลโรคหัวใจจากโปรเจกต์สแตตส์ออคยุโรป

ตารางที่ 4.19 แสดงผลการทดลองการสกัดกัญญาธรรมชาติ RENN-P-FIDR ของฐานข้อมูลโรคหัวใจจากคลินิกคลิฟแลนด์ทั้งหมด ตัวแปรที่มีผลต่อการสกัดกัญญาธรรมชาติ RENN-P-FIDR มี 3 ตัวแปรคือ ค่าร้อยละของน้ำหนักที่ยอมรับได้ \square จำนวนรูปแบบภาษาธรรมชาติ L และค่าความกำกวมที่ยอมรับได้ ϕ ซึ่งแต่ละตัวแปรมีแนวโน้มผลการทดลอง ดังนี้

1. ค่าร้อยละของน้ำหนักที่ยอมรับได้ \square ถ้ากำหนดให้ \square มีค่ามากจะตัดโหนดโครงข่ายประสาทเทียมได้มาก กฎที่สกัดได้มีจำนวนกฎและจำนวนตัวเชื่อมของเงื่อนไขลดลง เวลาที่ใช้ในการสกัดกัญญาลดลงมาก แต่อย่างไรก็ตามค่าความถูกต้องอาจลดลงตามไปด้วย ทั้งนี้เราสามารถกำหนดได้ว่า จะยอมรับค่าความถูกต้องลดลงได้จนถึงระดับใด ซึ่งจากผลการทดลองควรกำหนดค่า \square ให้อยู่ในช่วง 0 ถึง 40

2. จำนวนรูปแบบภาษาธรรมชาติ L ถ้ากำหนดให้ L มีค่าจำนวนช่วงมากขึ้น (เช่น $L = 7$) หรือมีการแบ่งช่วงละเอียดขึ้น จะได้ค่าความถูกต้อง จำนวนกฎและจำนวนตัวเชื่อมของเงื่อนไขที่ใกล้เคียงกับค่า L ที่มีจำนวนช่วงน้อยกว่า (เช่น $L = 3$) ทั้งนี้จะใช้เวลาในการสกัดกัญญานานกว่า

3. ค่าความกำกวมที่ยอมรับได้ ϕ ถ้ากำหนดให้ $\phi = 30$ หมายถึงจะแบ่งช่วงข้อมูลโดยยอมให้ค่าความกำกวมของช่วงไม่เกินร้อยละ 30 จากผลการทดลองแม้ว่ากำหนดให้ ค่าความกำกวมที่ยอมรับได้ ϕ มีค่าแตกต่างกันคือ $\phi = 30$ $\phi = 50$ $\phi = 80$ พบว่าได้ ค่าความถูกต้อง จำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข และเวลาที่ใช้ในการสกัดกัญญาใกล้เคียงกันเนื่องจากช่วงภาษาธรรมชาติที่แบ่งได้มีค่าของช่วงใกล้เคียงกัน

ตารางที่ 4.20 แสดงผลการทดลองเปรียบเทียบการสกัดกัญญาธรรมชาติ RENN-P-FIDR กับวิธีการอื่นๆ เมื่อกำหนดให้ $L=7$ และ $\phi = 50$ พบว่าวิธีการสกัดกัญญาธรรมชาติ RENN-P-FIDR ได้ผลการทดลองที่ดีกว่า C-MLP2LN [3] DIMLP [1] และ J48 [24]

ตารางที่ 4.19 ผลการทดลองการสกัดกฎด้วย RENN-P-FIDR ของฐานข้อมูลโรคหัวใจ
จากโปรเจกต์สแตตส์ออคยูโรป

δ	Method		L	ϕ	Total Accuracy (%)	#Rules	#Con.	Time (s.)
	RENN-P-FIDR	0	3	30	87.6	10	17	58
	RENN-P-FIDR	0	5	30	87.9	12	17	104
	RENN-P-FIDR	0	7	30	88.5	11	19	161
	RENN-P-FIDR	0	3	50	88.3	12	23	58
	RENN-P-FIDR	0	5	50	88.0	12	18	107
	RENN-P-FIDR	0	7	50	88.6	12	19	149
	RENN-P-FIDR	0	3	80	87.9	10	18	55
	RENN-P-FIDR	0	5	80	87.7	11	17	104
	RENN-P-FIDR	0	7	80	87.7	10	15	157
	RENN-P-FIDR	20	3	30	85.3	6	10	7
	RENN-P-FIDR	20	5	30	85.3	6	7	9
	RENN-P-FIDR	20	7	30	85.4	7	9	11
	RENN-P-FIDR	20	3	50	86.3	8	12	7
	RENN-P-FIDR	20	5	50	85.6	6	8	9
	RENN-P-FIDR	20	7	50	85.4	7	8	7
	RENN-P-FIDR	20	3	80	85.8	7	9	7
	RENN-P-FIDR	20	5	80	84.9	8	11	9
	RENN-P-FIDR	20	7	80	85.7	9	12	11
	RENN-P-FIDR	40	3	30	83.8	6	7	3
	RENN-P-FIDR	40	5	30	83.7	7	9	3
	RENN-P-FIDR	40	7	30	84.6	7	9	4
	RENN-P-FIDR	40	3	50	84.6	5	7	3
	RENN-P-FIDR	40	5	50	82.4	5	6	3
	RENN-P-FIDR	40	7	50	84.3	6	8	3
	RENN-P-FIDR	40	3	80	84.2	6	7	3
	RENN-P-FIDR	40	5	80	85.6	6	9	3

RENN-P-FIDR	40	7	80	85.1	7	9	4
-------------	----	---	----	------	---	---	---

ตารางที่ 4.20 ผลการทดลองเมื่อเปรียบเทียบการสกัดกฎด้วย RENN-P-FIDR กับขั้นตอนวิธีอื่นๆ ของฐานข้อมูลโรคหัวใจจากโปรเจกต์สแตตส์ออคยูโรป

Method	Type	#Rules	#Con.	Total Accuracy (%)
RENN-P-FIDR ($\alpha = 0, L=7, \phi =50$)	Linguistic Rule	12	19	88.6
RENN-P-FIDR ($\alpha = 20, L=7, \phi =50$)	Linguistic Rule	7	8	85.4
RENN-P-FIDR ($\alpha = 40, L=7, \phi =50$)	Linguistic Rule	6	8	84.3
C-MLP2LN [3]	Crisp Rule	-	-	-
DIMLP [1]	Crisp Rule	-	-	84.0
J48 [24]	Crisp Rule	18	75	77.0

4.4 ฐานข้อมูลโรคเบาหวาน

ฐานข้อมูลโรคเบาหวาน (Pima Indians Diabetes Database) ประกอบด้วยข้อมูล 2 กลุ่มคือ กลุ่มผู้ที่ไม่เป็นโรคเบาหวาน (Negative) มี 500 คน และกลุ่มผู้ที่เป็นโรคเบาหวาน (Positive) มี 268 คน รวมทั้งหมด 768 คน ตัวแปรข้อมูลเข้ามีทั้งหมด 8 ตัวแปร ดังตารางที่ 4.21

ตารางที่ 4.21 รายละเอียดของฐานข้อมูลโรคเบาหวาน

ตัวแปร	ชื่อตัวแปร	ชนิด	ค่าที่เป็นไปได้
x_1	number of times pregnant	Integer	0-17
x_2	plasma glucose concentration a 2 hours in an oral glucose tolerance test	Integer	0-199
x_3	diastolic blood pressure (mm Hg)	Integer	0-122
x_4	triceps skin fold thickness (mm)	Integer	0-99
x_5	2-hour serum insulin (mu U/ml)	Integer	0-846
x_6	body mass index (weight in kg/ (height in m) ²)	Real	0-67.1
x_7	diabetes pedigree function	Real	0.078-2.42
x_8	age (years)	Integer	21-81
c	class	Nominal Scale	0 = Negative

			1 = Positive
--	--	--	--------------

4.4.1 การตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW)

ในการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุดมีขั้นตอนในการทำงาน 3 ขั้นตอนมีรายละเอียดดังนี้

ขั้นตอนที่ 1 การเตรียมข้อมูล เพื่อใช้ในการทดลองมีการเลือกกลุ่มข้อมูลเพื่อใช้ในการสอน (Training set) 90 % และใช้ในการทดสอบ (Testing set) 10 %

ขั้นตอนที่ 2 การสอนโครงข่ายประสาทเทียมด้วยการเรียนรู้แบบแพร่ย้อนกลับ กำหนดให้มีโหนดข้อมูลเข้า 8 โหนด โหนดซ่อน 16 โหนด และโหนดผลลัพธ์ 1 โหนด ดังนั้นโครงสร้างของโครงข่ายประสาทเทียมคือ 8:16:1 โดยจะทำการสอนโครงข่ายประสาทเทียมจนกระทั่งครบจำนวนรอบที่ตั้งไว้ 200 รอบ (Epochs) หรือได้ค่าความผิดพลาด (MSE) น้อยกว่าค่าที่ยอมรับได้ซึ่งมีค่าเท่ากับ 0 ในขั้นตอนนี้จะได้ค่าน้ำหนักของโครงข่ายประสาทเทียมเพื่อนำไปตัดโหนดในขั้นตอนที่ 3 ขั้นตอนวิธีการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด

ขั้นตอนที่ 3 ขั้นตอนวิธีการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด ตัวอย่างของโครงสร้างโครงข่ายประสาทเทียมที่ผ่านการตัดโหนดเมื่อกำหนดให้ค่าร้อยละของน้ำหนักที่ยอมรับได้ $\square = 20$ คือ 7:8:1 กล่าวคือเหลือโหนดข้อมูลเข้า 7 โหนด โหนดซ่อน 8 โหนด และโหนดผลลัพธ์ 1 โหนด โดยมีโหนดข้อมูลเข้าที่ 5 ถูกตัดออกไป

4.4.2 การสกัดคุณลักษณะจากโครงข่ายประสาทเทียมโดยการแทนค่าความถี่ของช่วงข้อมูล (RENN-FIDR)

ในการสกัดคุณลักษณะจากโครงข่ายประสาทเทียมโดยการแทนค่าความถี่ของช่วงข้อมูลจะพิจารณาเฉพาะโหนดซ่อน และโหนดข้อมูลเข้าที่เสนอแนะจากขั้นตอนหลักที่ I การตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) ในการทดลองจะเลือกกลุ่มข้อมูลแบบความถี่ตรง 10 กลุ่ม เพื่อให้ข้อมูลทุกตัวได้ถูกนำมาเป็นข้อมูลชุดทดสอบ ซึ่งมีขั้นตอนในการทำงาน 3 ขั้นตอนมีรายละเอียดดังนี้

ขั้นตอนที่ 4 การสร้างรูปแบบภาษาธรรมชาติจากความถี่ของช่วงข้อมูล ตัวอย่างช่วงภาษาธรรมชาติที่แบ่งได้ เมื่อ $\square = 20$ $\phi = 50$ และ $L = 7$ แสดงดังตารางที่ 4.22

ตารางที่ 4.22 ช่วงของข้อมูลเมื่อ $\square = 20$ $\phi = 50$ และ $L = 7$ ของฐานข้อมูลโรคเบาหวาน

x	tiny	very small	small	medium	large	very large	huge
x_1	[0,3)	[1,7)	[7,10)	[10,11)	[11,12)	[12,14)	[14,17]

x_2	[0,175)	[175,179)	[179,183)	[183,194)	[194,197)	[197,198)	[198,199]
x_3	[0,55)	[55,56)	[56,98)	[98,100)	[100,102)	[102,114)	[114,114]
x_4	[0,50)	[50,51)	[51,52)	[52,56)	[56,60)	[60,63)	[63,99]
x_6	[0,45.4)	[45.4,46.2)	[46.2,48.3)	[48.3,52.9)	[52.9,57.3)	[57.3,59.4)	[59.4,67.1]
x_7	[0.08,0.09)	[0.09,1.28)	[1.28,1.4)	[1.4,2.14)	[2.14,2.33)	[2.33,2.42)	[2.42,2.42]
x_8	[21,36)	[36,60)	[60,67)	[67,68)	[68,70)	[70,72)	[72,81]

ขั้นตอนที่ 5 การสอนโครงข่ายประสาทเทียมด้วยรูปแบบภาษารธรรมชาติ โดยมี การเรียนรู้แบบแพร่ย้อนกลับ โดยจะทำการสอนโครงข่ายประสาทเทียมจนกระทั่งครบจำนวนรอบ ที่ตั้งไว้ 200 รอบ (Epochs) หรือได้ค่าความผิดพลาด (MSE) น้อยกว่าค่าที่ยอมรับได้ซึ่งมีค่าเท่ากับ 0 โหนดข้อมูลเข้า และโหนดซ่อนจะกำหนดจากการเสนอแนะจาก ขั้นตอนการตัดโหนดโครงข่าย ประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) ตัวอย่างเช่นถ้าโหนดข้อมูลเข้า 7 โหนด และ โหนดซ่อน 8 โหนด กรณีที่ $L = 7$ จะได้โครงสร้างของโครงข่ายประสาทเทียมเป็น $(7 \times 7):8:1$ หรือ $49:8:1$ น้ำหนักที่ได้จากการสอนโครงข่ายประสาทเทียมในขั้นตอนที่ 3 ขั้นตอนวิธีการสร้างกฎโดย พิจารณาค่าน้ำหนักที่เป็นบวก

ขั้นตอนที่ 6 ขั้นตอนวิธีการสร้างกฎโดยพิจารณาค่าน้ำหนักที่เป็นบวก ตัวอย่างกฎ ในกรณีที่มีค่า $\square = 20$ $\phi = 50$ ที่ $L = 7$ แสดงได้ดังภาพประกอบ 4.27 ซึ่งจำนวนกฎที่สกัดได้คือ 2 กฎและจำนวนตัวเชื่อม (and หรือ or) ของเงื่อนไขที่สกัดได้คือ 17 ตัวเชื่อม ซึ่งมี ค่าความถูกต้อง 74.2 % (ทำนายถูกต้อง 570 คน และทำนายผิด 198 คน)

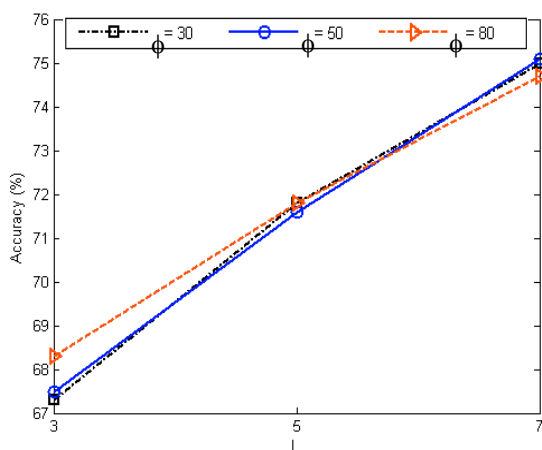
R_1 : If $x_2 = \text{medium}$ or $x_2 = \text{small}$ or $x_7 = \text{small}$ or $x_6 = \text{medium}$ or $x_6 = \text{very small}$ or $x_1 = \text{large}$ or $x_3 = \text{very large}$ or $x_2 = \text{large}$ or $x_2 = \text{very small}$ or $x_1 = \text{huge}$ or $x_6 = \text{large}$ or $x_2 = \text{very large}$ or $x_7 = \text{large}$ or $x_2 = \text{huge}$ or $x_8 = \text{very large}$ or $x_3 = \text{huge}$ then Positive R_2 : If $x_1 = \text{tiny}$ and $x_8 = \text{small}$ and $x_3 = \text{small}$ then Positive else Negative

ภาพประกอบ 4.27 ตัวอย่างกฎฐานข้อมูลโรคเบาหวานเมื่อ $\square = 20$ $\phi = 50$ และ $L = 7$

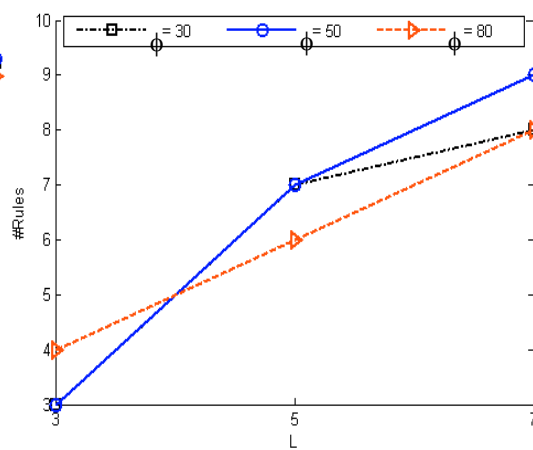
4.4.3 เปรียบเทียบผลการทดลองและวิจารณ์ผลการทดลอง

เมื่อกำหนดให้ค่าร้อยละของน้ำหนักที่ยอมรับได้ $\phi = 0$ $\phi = 20$ และ $\phi = 40$ จะได้ผลการทดลองดังนี้

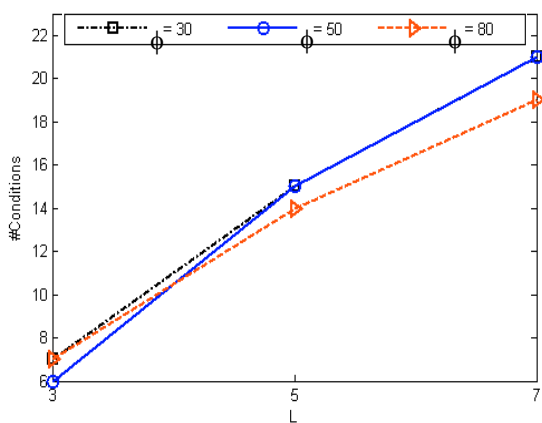
1. ค่าร้อยละที่ยอมรับได้ $\phi = 0$ (ไม่มีการตัดโหนดก่อนสกัดคุณภาษา ธรรมชาติ) โครงสร้างโครงข่ายประสาทเทียมหลังผ่านการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ น้ำหนักสูงสุด (NNP-MFW) คือ 8:16:1 เมื่อกำหนดให้ค่าความกำวมที่ยอมรับได้ $\phi = 30$ $\phi = 50$ และ $\phi = 80$ กำหนดจำนวนรูปแบบภาษาธรรมชาติ $L = 3$ $L = 5$ และ $L = 7$ จะได้ผลการทดลองดัง ภาพประกอบ 4.28



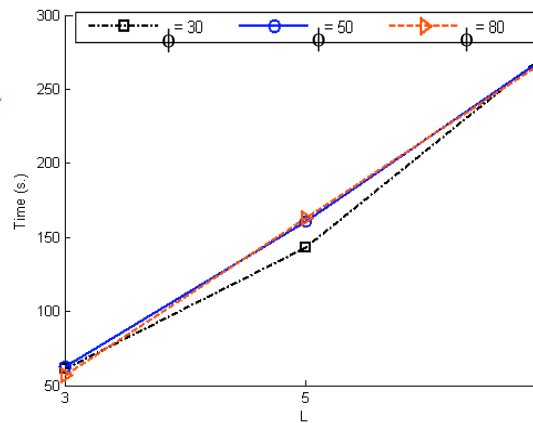
(a) ค่าความถูกต้อง (%)



(b) จำนวนกฎ



(c) จำนวนตัวเชื่อมของเงื่อนไข

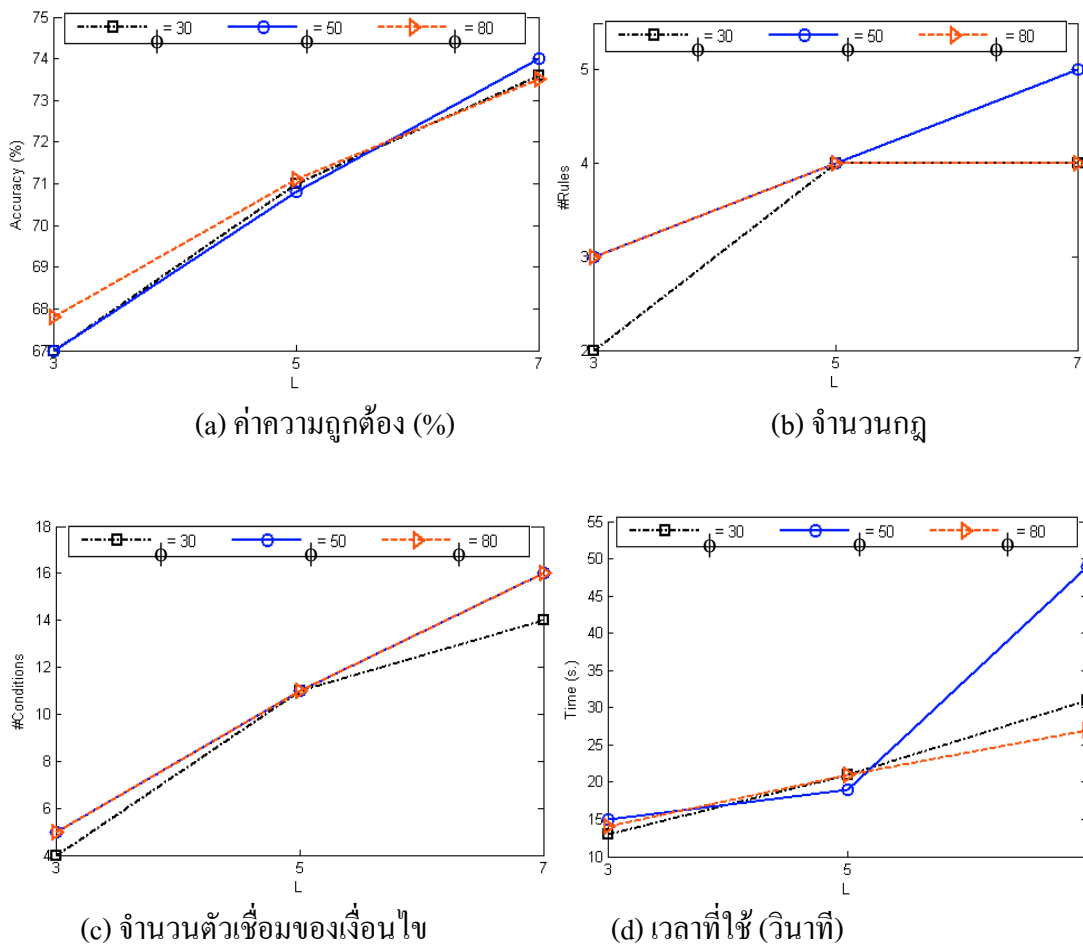


(d) เวลาที่ใช้ (วินาที)

ภาพประกอบ 4.28 กราฟของผลการทดลองการสกัดกัญญาธรรมชาติด้วยวิธีการ RENN-P-FIDR ของฐานข้อมูลโรคเบาหวานเมื่อค่าร้อยละที่ยอมรับได้ $\square = 0$

จากภาพประกอบ 4.28(a) พบว่าค่าความถูกต้องจะขึ้นอยู่กับข้อกำหนดจำนวนรูปแบบภาษาธรรมชาติ L กล่าวคือเมื่อกำหนดช่วง L ให้มีค่าจำนวนช่วงมากขึ้น (เช่น $L = 7$) หรือมีการแบ่งช่วงละเอียดขึ้น จะได้ค่าความถูกต้องที่สูงกว่าค่า L ที่มีจำนวนช่วงน้อยกว่า (เช่น $L = 5$ และ $L = 3$) หรือมีการแบ่งช่วงที่กว้างกว่าในการแทนค่าข้อมูล แต่อย่างไรก็ตามจำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข เวลาในการสกัดกัญญาธรรมชาติเพิ่มขึ้นตามค่า L ที่เพิ่มขึ้นด้วย ดังภาพประกอบ 4.28(b) ภาพประกอบ 4.28(c) และภาพประกอบ 4.28(d) นอกจากนี้จากผลการทดลองพบว่าถึงแม้ว่ากำหนดให้ค่าความกำกวมที่ยอมรับได้ ϕ มีค่าแตกต่างกันดังนี้ $\phi = 30$ $\phi = 50$ $\phi = 80$ แต่คงยังได้ค่าความถูกต้อง จำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข และเวลาที่ใช้ในการสกัดกัญญา ใกล้เคียงกันเนื่องจากช่วงภาษาธรรมชาติที่แบ่งได้มีค่าของช่วงใกล้เคียงกัน ทั้งนี้เนื่องจากช่วงที่แบ่งได้มีจำนวนไม่เท่ากับช่วงภาษาธรรมชาติที่ต้องการ ทำให้ต้องมีการรวมช่วงในกรณีที่มีจำนวนช่วงที่แบ่งได้มากกว่าจำนวนช่วงภาษาธรรมชาติที่ต้องการ หรือแบ่งครึ่งช่วงในกรณีที่มีจำนวนช่วงที่แบ่งได้น้อยกว่าจำนวนช่วงภาษาธรรมชาติที่ต้องการ นอกจากนี้ ค่าความถูกต้องของกฎยังขึ้นอยู่กับขั้นตอนการสอนโครงข่ายประสาทเทียม และขั้นตอนการสร้างกฎอีกด้วย

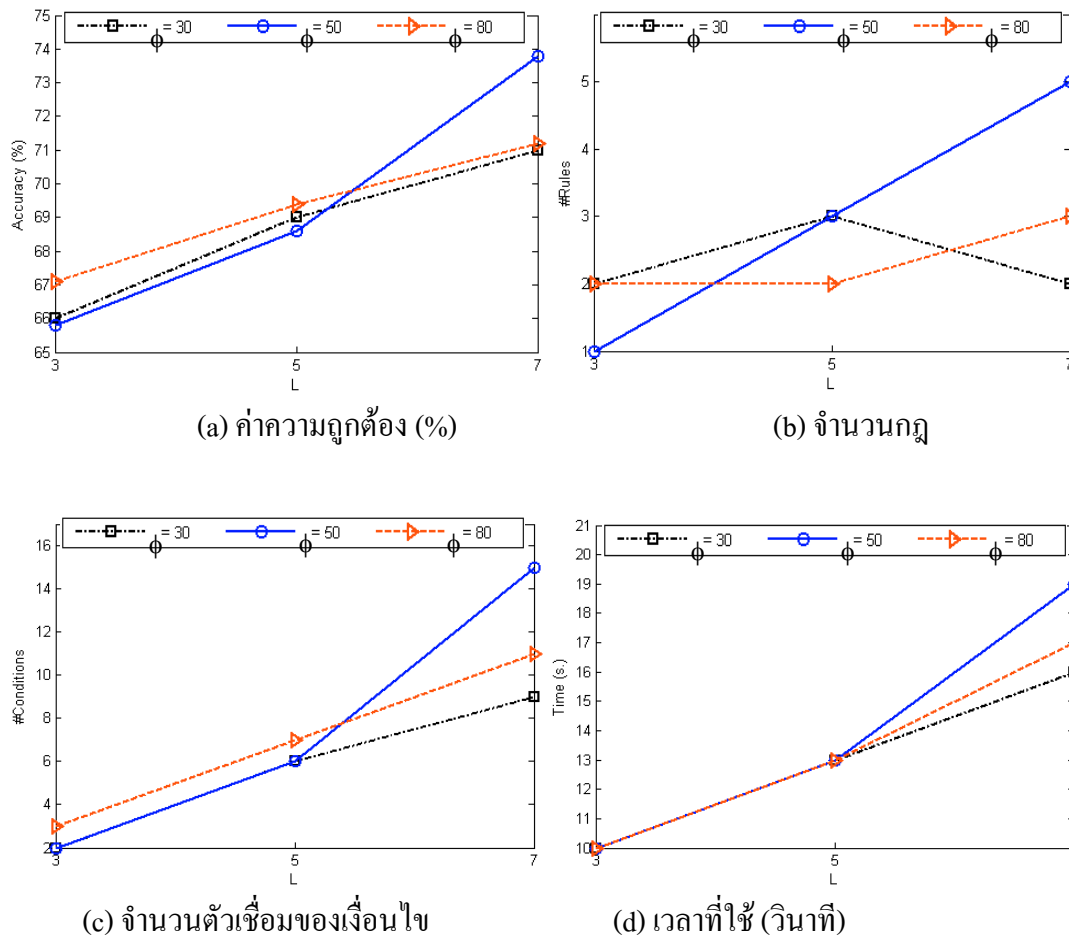
2. ค่าร้อยละที่ยอมรับได้ $\square = 20$ โครงสร้างโครงข่ายประสาทเทียมหลังผ่านการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) คือ 7:8:1 โดยที่ตัวแปรข้อมูลเข้าที่ 5 ถูกตัดออกจากโครงข่ายประสาทเทียม เมื่อกำหนดให้ค่าความกำกวมที่ยอมรับได้ $\phi = 30$ $\phi = 50$ และ $\phi = 80$ กำหนดจำนวนรูปแบบภาษาธรรมชาติ $L = 3$ $L = 5$ และ $L = 7$ จะได้ผลการทดลองดังภาพประกอบ 4.29 ซึ่งมีแนวโน้มเหมือนกับผลการทดลองเมื่อกำหนดให้ $\square = 0$ กล่าวคือ จากภาพประกอบ 4.29(a) พบว่าค่าความถูกต้องจะขึ้นอยู่กับข้อกำหนดจำนวนรูปแบบภาษาธรรมชาติ L กล่าวคือเมื่อกำหนดช่วง L ให้มีค่าจำนวนช่วงมากขึ้น จะได้ ค่าความถูกต้องที่สูงกว่าค่า L ที่มีจำนวนช่วงน้อยกว่า แต่อย่างไรก็ตามจำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข เวลาในการสกัดกัญญาธรรมชาติเพิ่มขึ้นตามค่า L ที่เพิ่มขึ้นด้วย ดังภาพประกอบ 4.29(b) ภาพประกอบ 4.29(c) และภาพประกอบ 4.29(d) นอกจากนี้จากผลการทดลองพบว่าถึงแม้ว่ากำหนดให้ค่าความกำกวมที่ยอมรับได้ ϕ มีค่าแตกต่างกันดังนี้ $\phi = 30$ $\phi = 50$ $\phi = 80$ แต่คงยังได้ค่าความถูกต้อง จำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข และเวลาที่ใช้ในการสกัดกัญญา ใกล้เคียงกันเนื่องจากช่วงภาษาธรรมชาติที่แบ่งได้มีค่าของช่วงใกล้เคียงกัน



ภาพประกอบ 4.29 กราฟของผลการทดลองการสกัดกฎภาษาธรรมชาติด้วยวิธีการ RENN-P-FIDR ของฐานข้อมูลโรคเบาหวานเมื่อค่าร้อยละที่ยอมรับได้ $\square = 20$

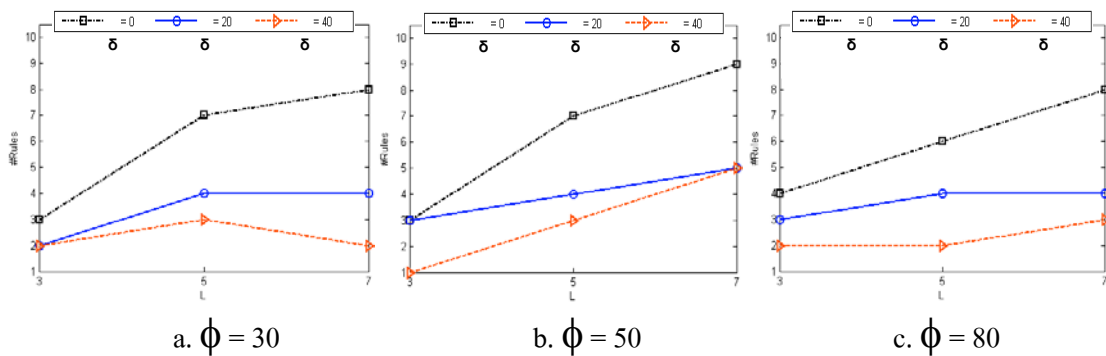
3. ค่าร้อยละที่ยอมรับได้ $\square = 40$ โครงสร้างโครงข่ายประสาทเทียมหลังผ่านการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) คือ 7:7:1 โดยที่ตัวแปรข้อมูลเข้าที่ 5 และตัวแปรข้อมูลเข้าที่ 4 ถูกตัดออกจากโครงข่ายประสาทเทียม เมื่อกำหนดให้ค่าความกำกวมที่ยอมรับได้ $\phi = 30$ $\phi = 50$ และ $\phi = 80$ กำหนดจำนวนรูปแบบภาษาธรรมชาติ $L = 3$ $L = 5$ และ $L = 7$ จะได้ผลการทดลองดังภาพประกอบ 4.30 ซึ่งมี แนวโน้มผลการทดลองเหมือนกับเมื่อกำหนดให้ $\square = 0$ และ $\square = 40$ กล่าวคือจากภาพประกอบ 4.30(a) พบว่าค่าความถูกต้องจะขึ้นอยู่กับกำหนดจำนวนรูปแบบภาษาธรรมชาติ L กล่าวคือเมื่อกำหนดช่วง L ให้มีค่าจำนวนช่วงมากขึ้น จะได้ค่าความถูกต้องที่สูงกว่าค่า L ที่มีจำนวนช่วงน้อยกว่า แต่อย่างไรก็ตามจำนวนกฎ

จำนวนตัวเชื่อมของเงื่อนไข เวลาในการสกัดกฎภาษาธรรมชาติเพิ่มขึ้นตามค่า L ที่เพิ่มขึ้นด้วย ดังภาพประกอบ 4.30(b) ภาพประกอบ 4.30(c) และภาพประกอบ 4.30(d)

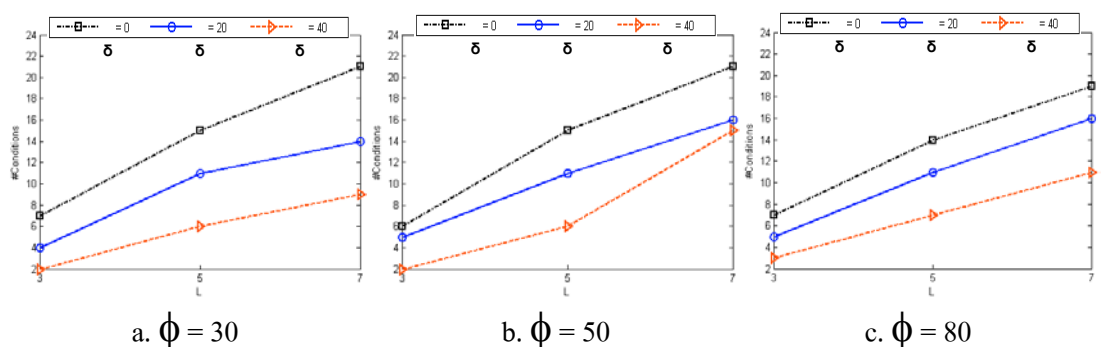


ภาพประกอบ 4.30 กราฟของผลการทดลองการสกัดกฎภาษาธรรมชาติด้วยวิธีการ RENN-P-FIDR ของฐานข้อมูลโรคเบาหวานเมื่อกำหนดค่าร้อยละที่ยอมรับได้ $\square = 40$

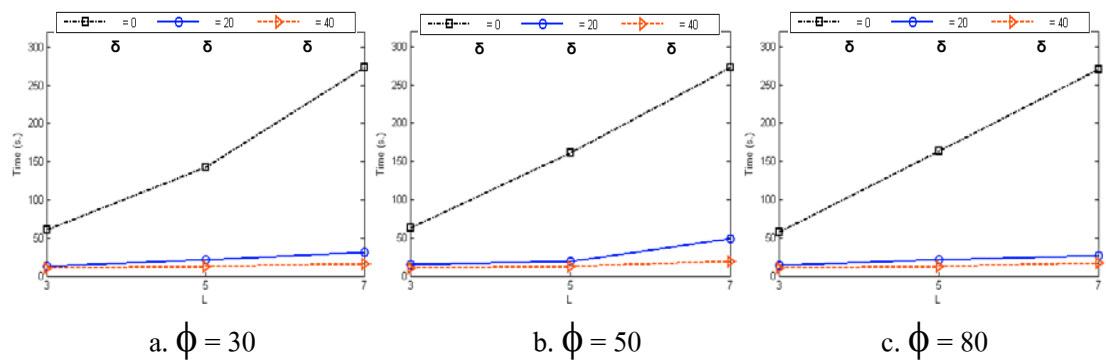
สามารถเปรียบเทียบผลการทดลองจากการกำหนดค่าร้อยละที่ยอมรับได้ $\square = 0$ $\square = 20$ และ $\square = 40$ ของจำนวนกฎแสดงได้ดังภาพประกอบ 4.31 ของจำนวนตัวเชื่อมของเงื่อนไขแสดงได้ดังภาพประกอบ 4.32 ของเวลาที่ใช้แสดงได้ดังภาพประกอบ 4.33 และของ ค่าความถูกต้องแสดงได้ดังภาพประกอบ 4.34 ซึ่งพบว่าเมื่อกำหนดให้ค่าร้อยละที่ยอมรับได้ \square เพิ่มขึ้นจะทำให้ได้จำนวนกฎ และจำนวนตัวเชื่อมของเงื่อนไขลดลง เวลาที่ใช้ในการสกัดกฎลดลงมากแต่ค่าความถูกต้องลดลงเพียงเล็กน้อย



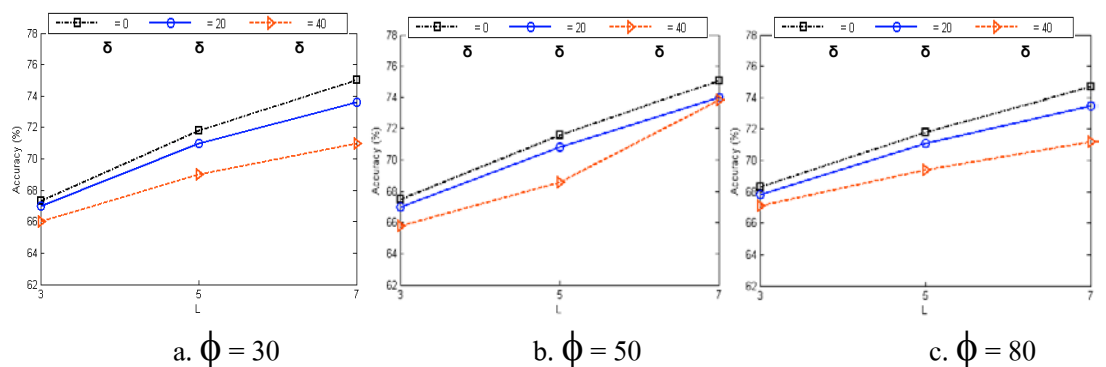
ภาพประกอบ 4.31 จำนวนกฎของฐานข้อมูลโรคเบาหวาน



ภาพประกอบ 4.32 จำนวนตัวเชื่อมของเงื่อนไขของฐานข้อมูลโรคเบาหวาน



ภาพประกอบ 4.33 เวลาที่ใช้ในการสกัดกฎ (วินาที) ของฐานข้อมูลโรคเบาหวาน



ภาพประกอบ 4.34 ค่าความถูกต้อง (%) ของฐานข้อมูลโรคเบาหวาน

ตารางที่ 4.23 แสดงผลการทดลองการสกัดกัญญาธรรมชาติ RENN-P-FIDR ของฐานข้อมูลโรคเบาหวานทั้งหมด ตัวแปรที่มีผลต่อการสกัดกัญญาธรรมชาติ RENN-P-FIDR มี 3 ตัวแปรคือ ค่าร้อยละของน้ำหนักที่ยอมรับได้ จำนวนรูปแบบภาษา ธรรมชาติ L และ ค่าความกำกวมที่ยอมรับได้ ϕ ซึ่งแต่ละตัวแปรแนวโน้มผลการทดลองดังนี้

1. ค่าร้อยละของน้ำหนักที่ยอมรับได้ ถ้ากำหนดให้ มีค่ามากจะตัด โหนดโครงข่ายประสาทเทียมได้มาก กฎที่สกัดได้มีจำนวนกฎและจำนวนตัวเชื่อมของเงื่อนไขลดลง เวลาที่ใช้ในการสกัดกฎลดลงมาก แต่อย่างไรก็ตามค่าความถูกต้องอาจลดลงไปด้วย ทั้งนี้เราสามารถกำหนดได้ว่าจะยอมรับค่าความถูกต้องลดลงได้จนถึงระดับใด ซึ่งจากผลการทดลองควร กำหนดค่า ให้อยู่ในช่วง 0 ถึง 40

2. จำนวนรูปแบบภาษาธรรมชาติ L ถ้ากำหนดให้ L มีค่าจำนวนช่วงมากขึ้น (เช่น $L = 7$) หรือมีการแบ่งช่วงละเอียดขึ้น จะได้ค่าความถูกต้องที่สูงกว่าค่า L ที่มีจำนวนช่วงน้อยกว่า (เช่น $L = 3$) หรือมีการแบ่งช่วงที่กว้างกว่าในการแทนค่าข้อมูล แต่อย่างไรก็ตามจำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข และเวลาเพิ่มขึ้นตามค่า L ที่เพิ่มขึ้นด้วย

3. ค่าความกำกวมที่ยอมรับได้ ϕ ถ้ากำหนดให้ $\phi = 30$ หมายถึงจะแบ่งช่วง ข้อมูลโดยยอมให้ค่าความกำกวมของช่วงไม่เกินร้อยละ 30 จากผลการทดลองแม้ว่ากำหนดให้ ค่าความกำกวมที่ยอมรับได้ ϕ มีค่าแตกต่างกันคือ $\phi = 30$ $\phi = 50$ $\phi = 80$ พบว่าได้ ค่าความถูกต้อง จำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข และเวลาที่ใช้ในการสกัดกฎใกล้เคียงกันเนื่องจากช่วง ภาษาธรรมชาติที่แบ่งได้มีค่าของช่วงใกล้เคียงกัน

ตารางที่ 4.24 แสดงผลการทดลองเปรียบเทียบการสกัดกัญญาธรรมชาติ RENN-P-FIDR กับวิธีการอื่นๆ เมื่อกำหนดให้ = 0 $L=7$ และ $\phi = 50$ พบว่าวิธีการสกัดกัญญาธรรมชาติ RENN-P-FIDR ได้ผลการทดลองที่ดีกว่า J48 [24]

ตารางที่ 4.23 ผลการทดลองการสกัดกฎด้วย RENN-P-FIDR ของฐานข้อมูลโรคเบาหวาน

δ Method		L	ϕ	Total Accuracy (%)	#Rules	#Con.	Time (s.)
RENN-P-FIDR	0	3	30	67.3	3	7	61
RENN-P-FIDR	0	5	30	71.8	7	15	143
RENN-P-FIDR	0	7	30	75.0	8	21	274
RENN-P-FIDR	0	3	50	67.5	3	6	63
RENN-P-FIDR	0	5	50	71.6	7	15	161
RENN-P-FIDR	0	7	50	75.1	9	21	272
RENN-P-FIDR	0	3	80	68.3	4	7	57
RENN-P-FIDR	0	5	80	71.8	6	14	163
RENN-P-FIDR	0	7	80	74.7	8	19	270
RENN-P-FIDR	20	3	30	67.0	2	4	13
RENN-P-FIDR	20	5	30	71.0	4	11	21
RENN-P-FIDR	20	7	30	73.6	4	14	31
RENN-P-FIDR	20	3	50	67.0	3	5	15
RENN-P-FIDR	20	5	50	70.8	4	11	19
RENN-P-FIDR	20	7	50	74.0	5	16	49
RENN-P-FIDR	20	3	80	67.8	3	5	14
RENN-P-FIDR	20	5	80	71.1	4	11	21
RENN-P-FIDR	20	7	80	73.5	4	16	27
RENN-P-FIDR	40	3	30	66.0	2	2	10
RENN-P-FIDR	40	5	30	69.0	3	6	13
RENN-P-FIDR	40	7	30	71.0	2	9	16
RENN-P-FIDR	40	3	50	65.8	1	2	10
RENN-P-FIDR	40	5	50	68.6	3	6	13
RENN-P-FIDR	40	7	50	73.8	5	15	19
RENN-P-FIDR	40	3	80	67.1	2	3	10
RENN-P-FIDR	40	5	80	69.4	2	7	13
RENN-P-FIDR	40	7	80	71.2	3	11	17

ตารางที่ 4.24 ผลการทดลองเมื่อเปรียบเทียบการสกัดกฎด้วย RENN-FIDR กับขั้นตอนวิธีอื่นๆ
ของฐานข้อมูลโรคเบาหวาน

Method	Type	#Rules	#Con.	Total Accuracy (%)
RENN-P-FIDR ($\alpha = 0, L=7, \phi =50$)	Linguistic Rule	9	21	75.1
RENN-P-FIDR ($\alpha = 20, L=7, \phi =50$)	Linguistic Rule	5	16	74.0
RENN-P-FIDR ($\alpha = 40, L=7, \phi =50$)	Linguistic Rule	5	15	73.8
C-MLP2LN [3]	Crisp Rule	-	-	77.7
DIMLP [1]	Crisp Rule	-	-	76.4
J48 [24]	Crisp Rule	20	91	71.5

4.5 ฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจ

ฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจจากโรงพยาบาลในประเทศไทย (Electrocardiogram Database of Heart Disease Patients) [39] ซึ่งเก็บรวบรวมมาจาก โรงพยาบาล 4 โรงพยาบาลคือ โรงพยาบาลสงขลานครินทร์ โรงพยาบาลราชวิถี โรงพยาบาลพญาไท 2 และโรงพยาบาลจุฬา โดยฐานข้อมูลนี้ใช้ในการศึกษาโครงการเฉพาะเรื่องซึ่งเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจจากโรงพยาบาลในประเทศไทยมีการจำแนกผู้ป่วยเป็น 2 ภาวะได้แก่ภาวะปกติ (Normality) มี 325 คน และภาวะผิดปกติ (Abnormality) มี 545 คน รวมทั้งหมด 870 คน ตัวแปรข้อมูลเข้ามีทั้งหมด 5 ตัวแปร x_1 คือคลื่นปกติ P Wave หัวตั้งเสมอ และมี QRS Complex ตามมา (P Wave) x_2 คือ Normal P-R Interval วัดจุดเริ่ม P ถึงจุดเริ่ม QRS เป็นวินาที (P-R Interval) x_3 คือคลื่นปกติวัดได้ 0.1 วินาที หรือน้อยกว่า และตามหลัง P Wave (QRS Complex) x_4 คือจังหวะ: สม่่าเสมอแทนด้วย 1 ไม่สม่่าเสมอแทนด้วย 0 (Rhythm) และ x_5 คืออัตรา: คนปกติอยู่ระหว่าง 60-100 ครั้งต่อนาที (Rate) ดังตารางที่ 4.25

ตารางที่ 4.25 รายละเอียดของฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจ

ตัวแปร	ชื่อตัวแปร	ชนิด	ค่าที่เป็นไปได้
x_1	P wave	Nominal Scale	0,1
x_2	P-R interval	Real	0.103-0.272
x_3	QRS complex	Real	0.04-0.92
x_4	rhythm	Nominal Scale	0,1
x_5	rate	Integer	20-166
c	class	Nominal Scale	1 = Normality 2 = Abnormality

4.5.1 การตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW)

ในการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุดมีขั้นตอนในการทำงาน 3 ขั้นตอนมีรายละเอียดดังนี้

ขั้นตอนที่ 1 การเตรียมข้อมูล ในการทดลองมีการเลือกกลุ่มข้อมูลเพื่อใช้ในการสอน (Training set) 90 % และใช้ในการทดสอบ (Testing set) 10 %

ขั้นตอนที่ 2 การสอนโครงข่ายประสาทเทียมด้วยการเรียนรู้แบบแพร่ย้อนกลับ กำหนดให้มีโหนดข้อมูลเข้า 5 โหนด โหนดซ่อน 10 โหนด และโหนดผลลัพธ์ 1 โหนด ดังนั้นโครงสร้างของโครงข่ายประสาทเทียมคือ 5:10:1 โดยจะทำการสอนโครงข่ายประสาทเทียมจนกระทั่งครบจำนวนรอบที่ตั้งไว้ 200 รอบ (Epochs) หรือได้ค่าความผิดพลาด (MSE) น้อยกว่าค่าที่ยอมรับได้ซึ่งมีค่าเท่ากับ 0 ในขั้นตอนนี้จะได้ค่าน้ำหนักของโครงข่ายประสาทเทียมเพื่อนำไปตัดโหนดในขั้นตอนที่ 3 ขั้นตอนวิธีการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด

ขั้นตอนที่ 3 ขั้นตอนวิธีการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด ตัวอย่างของโครงสร้างโครงข่ายประสาทเทียมที่ผ่านการตัดโหนดเมื่อกำหนดให้ค่าร้อยละของน้ำหนักที่ยอมรับได้ $\square = 20$ คือ 5:6:1 กล่าวคือเหลือโหนดข้อมูลเข้า 5 โหนด โหนดซ่อน 6 โหนด และโหนดผลลัพธ์ 1 โหนด

4.5.2 การสกัดคุณลักษณะชาติจากโครงข่ายประสาทเทียมโดยการแทนค่าความถี่ของช่วงข้อมูล (RENN-FIDR)

ในการสกัดคุณลักษณะชาติจากโครงข่ายประสาทเทียมโดยการแทนค่าความถี่ของช่วงข้อมูลจะพิจารณาเฉพาะโหนดซ่อน และโหนดข้อมูลเข้าที่เสนอแนะจากขั้นตอนหลักที่ I

การตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) ในการทดลองจะเลือก ลุ่มข้อมูลแบบความเที่ยงตรง 10 กลุ่ม (10-Fold Cross Validation) เพื่อให้ข้อมูลทุกตัวได้ถูกนำมา เป็นข้อมูลชุดทดสอบ (Testing Set) ซึ่งมีขั้นตอนในการทำงาน 3 ขั้นตอนมีรายละเอียดดังนี้

ขั้นตอนที่ 4 การสร้างรูปแบบภาษาธรรมชาติจากความถี่ของช่วงข้อมูล ตัวอย่าง ช่วงภาษาธรรมชาติที่แบ่งได้ จากการทดลองใน เมื่อ $\square = 20$ $\phi = 50$ และ $L = 7$ แสดงดัง ตาราง ที่ 4.26

ตารางที่ 4.26 ช่วงของข้อมูลเมื่อ $\square = 20$ $\phi = 50$ และ $L = 7$ ของฐานข้อมูลภาวะของผู้ป่วย โรคหัวใจ

x	tiny	very small	small	medium	large	very large	huge
x_1	[0,1)	[1,1]	-	-	-	-	-
x_2	[0.103,0.1 2)	[0.12,0.12 2)	[0.122,0.12 6)	[0.126,0.15 1)	[0.151,0.15 3)	[0.153,0.15 7)	[0.157,0.2 72]
x_3	[0.04,0.05 4)	[0.054,0.0 6)	[0.06,0.064)	[0.064,0.07)	[0.07,0.071)	[0.071,0.07 4)	[0.074,0.9 2]
x_4	[0,1)	[1,1]	-	-	-	-	-
x_5	[32,65)	[65,91)	[91,94)	[94,95)	[95,97)	[97,100)	[100,166]

ขั้นตอนที่ 5 การสอนโครงข่ายประสาทเทียมด้วยรูปแบบภาษาธรรมชาติ โดยมีการเรียนรู้แบบแพร่ย้อนกลับ โดยจะทำการสอนโครงข่ายประสาทเทียมจนกระทั่งครบจำนวนรอบ ที่ตั้งไว้ 200 รอบ (Epochs) หรือได้ค่าความผิดพลาด (MSE) น้อยกว่าค่าที่ยอมรับได้ซึ่งมีค่า เท่ากับ 0 โหนดข้อมูลเข้าและโหนดซ่อนจะกำหนดจากการเสนอแนะจากขั้นตอนการตัดโหนด โครงข่าย ประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) ตัวอย่างเช่นถ้าโหนดข้อมูลเข้า 5 โหนด และ โหนดซ่อน 6 โหนด กรณีที่ $L = 7$ จะได้โครงสร้างของโครงข่ายประสาทเทียมเป็น 25:6:1 โดยมี โหนดข้อมูลเข้า 25 โหนดเพราะช่วงภาษาธรรมชาติที่แบ่งได้มี 25 ช่วงดังตารางที่ 4.26 น้ำหนักที่ได้ จากการสอนโครงข่ายประสาทเทียมถูกนำไปใช้ในขั้นตอนที่ 6 ขั้นตอนวิธีการสร้างกฎโดยพิจารณา ค่าน้ำหนักที่เป็นบวก

ขั้นตอนที่ 6 ขั้นตอนวิธีการสร้างกฎโดยพิจารณาค่าน้ำหนักที่เป็นบวก ตัวอย่างกฎ ในกรณีที่มีค่า $\square = 20$ $\phi = 50$ ที่ $L = 7$ แสดงได้ดังภาพประกอบ 4.35 ซึ่งจำนวนกฎที่สกัดได้คือ 2

กฎและจำนวนตัวเชื่อม (and หรือ or) ของเงื่อนไขที่สกัดได้คือ 2 ตัวเชื่อม ซึ่งมี ค่าความถูกต้อง 89.3 % (ทำนายถูกต้อง 777 คน และทำนายผิด 93 คน)

R_1 : If $x_5 = \text{huge}$ or $x_4 = \text{tiny}$ then Abnormality
 R_2 : If $x_5 = \text{tiny}$ and $x_3 = \text{huge}$ then Abnormality
 else Normality

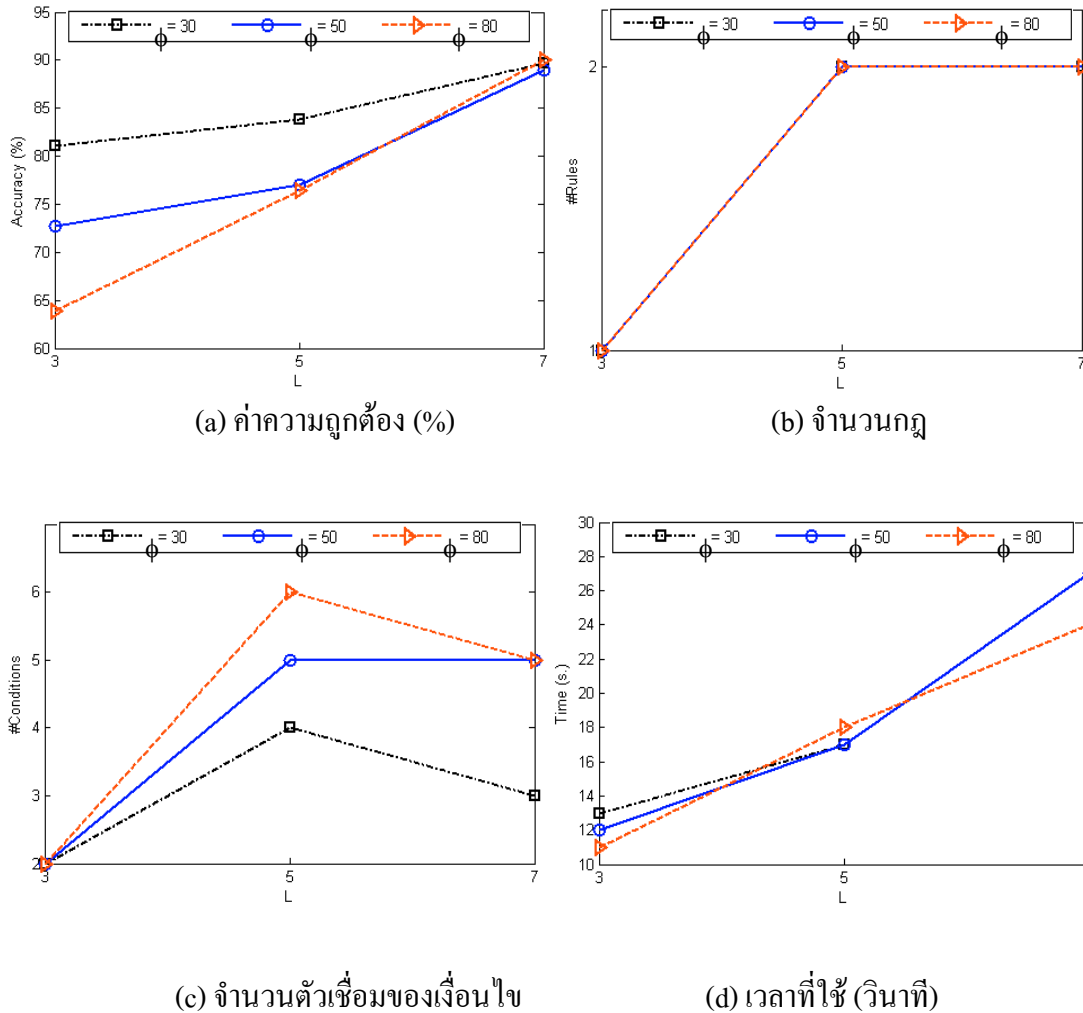
ภาพประกอบ 4.35 ตัวอย่างกฎของฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจเมื่อ

$$\square = 20 \quad \phi = 50 \quad \text{และ} \quad L = 7$$

4.5.3 เปรียบเทียบผลการทดลองและวิจารณ์ผลการทดลอง

เมื่อกำหนดให้ค่าร้อยละของน้ำหนักที่ยอมรับได้ $\square = 0$ $\square = 20$ และ $\square = 40$ จะได้ผลการทดลองดังนี้

1. ค่าร้อยละที่ยอมรับได้ $\square = 0$ (ไม่มีการตัดโหนดก่อนสกัดกฎหมาย ธรรมชาติ) โครงสร้างโครงข่ายประสาทเทียมหลังผ่านการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) คือ 5:10:1 เมื่อกำหนดให้ค่าความกำกวมที่ยอมรับได้ $\phi = 30$ $\phi = 50$ และ $\phi = 80$ กำหนดจำนวนรูปแบบภาษาธรรมชาติ $L = 3$ $L = 5$ และ $L = 7$ จะได้ผลการทดลองดังภาพประกอบ 4.36 จากภาพประกอบ 4.36(a) พบว่าค่าความถูกต้องจะขึ้นอยู่กับข้อกำหนดจำนวนรูปแบบภาษาธรรมชาติ L กล่าวคือเมื่อกำหนดช่วง L ให้มีค่าจำนวนช่วงมากขึ้น (เช่น $L = 7$) หรือมีการแบ่งช่วงละเอียดขึ้น จะได้ค่าความถูกต้องที่สูงกว่าค่า L ที่มีจำนวนช่วงน้อยกว่า (เช่น $L = 3$) หรือมีการแบ่งช่วงที่กว้างกว่าในการแทนค่าข้อมูล แต่อย่างไรก็ตามเวลาที่ใช้ในการสกัดกฎก็มากขึ้นด้วยดังภาพประกอบ 4.36(b) นอกจากนี้จะพบว่าเมื่อกำหนดให้ L มีค่ามากขึ้นจะทำให้มีจำนวนกฎและจำนวนตัวเชื่อมใกล้เคียงกัน ดังภาพประกอบ 3.36(c) และภาพประกอบ 3.36(d)

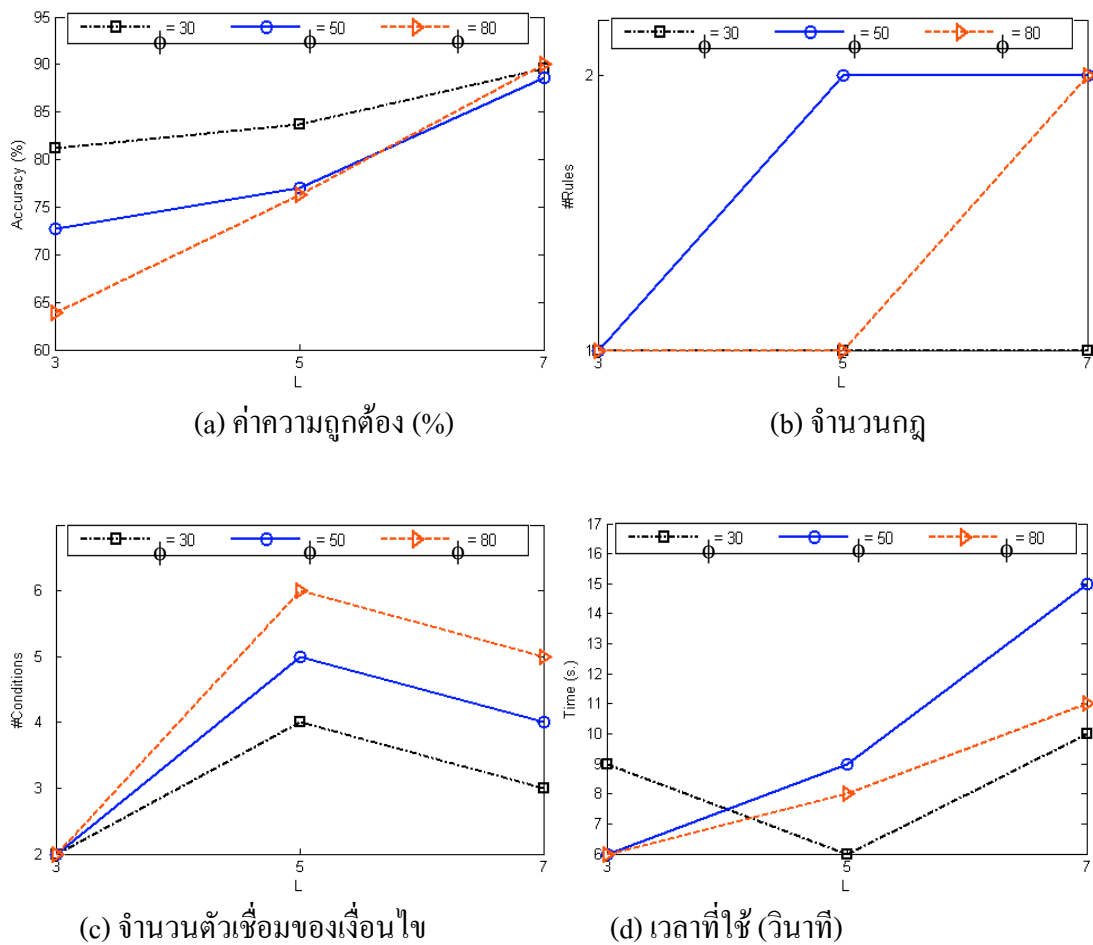


ภาพประกอบ 4.36 กราฟของผลการทดลองการสกัดคุณลักษณะด้วยวิธีการ RENN-P-FIDR ของฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจ เมื่อค่าร้อยละน้ำหนักที่ยอมรับได้

$$\square = 0$$

2. ค่าร้อยละที่ยอมรับได้ $\square = 20$ โครงสร้างโครงข่ายประสาทเทียมหลังผ่านการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) คือ 5:6:1 เมื่อกำหนดให้ค่าความกำกวมที่ยอมรับได้ $\phi = 30$ $\phi = 50$ และ $\phi = 80$ กำหนดจำนวนรูปแบบภาษาธรรมชาติ $L = 3$ $L = 5$ และ $L = 7$ จะได้ผลการทดลองดังภาพประกอบ 4.37 ซึ่งมีแนวโน้มเหมือนกับผลการทดลองเมื่อกำหนดให้ $\square = 0$ กล่าวคือจากภาพประกอบ 4.37(a) พบว่า ค่าความถูกต้องจะขึ้นอยู่กับ

การกำหนดจำนวนรูปแบบภาษาธรรมชาติ L กล่าวคือเมื่อกำหนดช่วง L ให้มีค่าจำนวนช่วงมากขึ้น จะได้ค่าความถูกต้องที่สูงกว่าค่า L ที่มีจำนวนช่วงน้อยกว่า แต่อย่างไรก็ตามเวลาที่ใช้ในการสกัดกฎก็มากขึ้นด้วยดังภาพประกอบ 4.37(d) นอกจากนี้จะพบว่าเมื่อกำหนดให้ L มีค่ามากขึ้นก็ยังคงได้จำนวนกฎและจำนวนตัวเชื่อมใกล้เคียงกัน ดังภาพประกอบ 4.37(b) และภาพประกอบ 4.37(c)

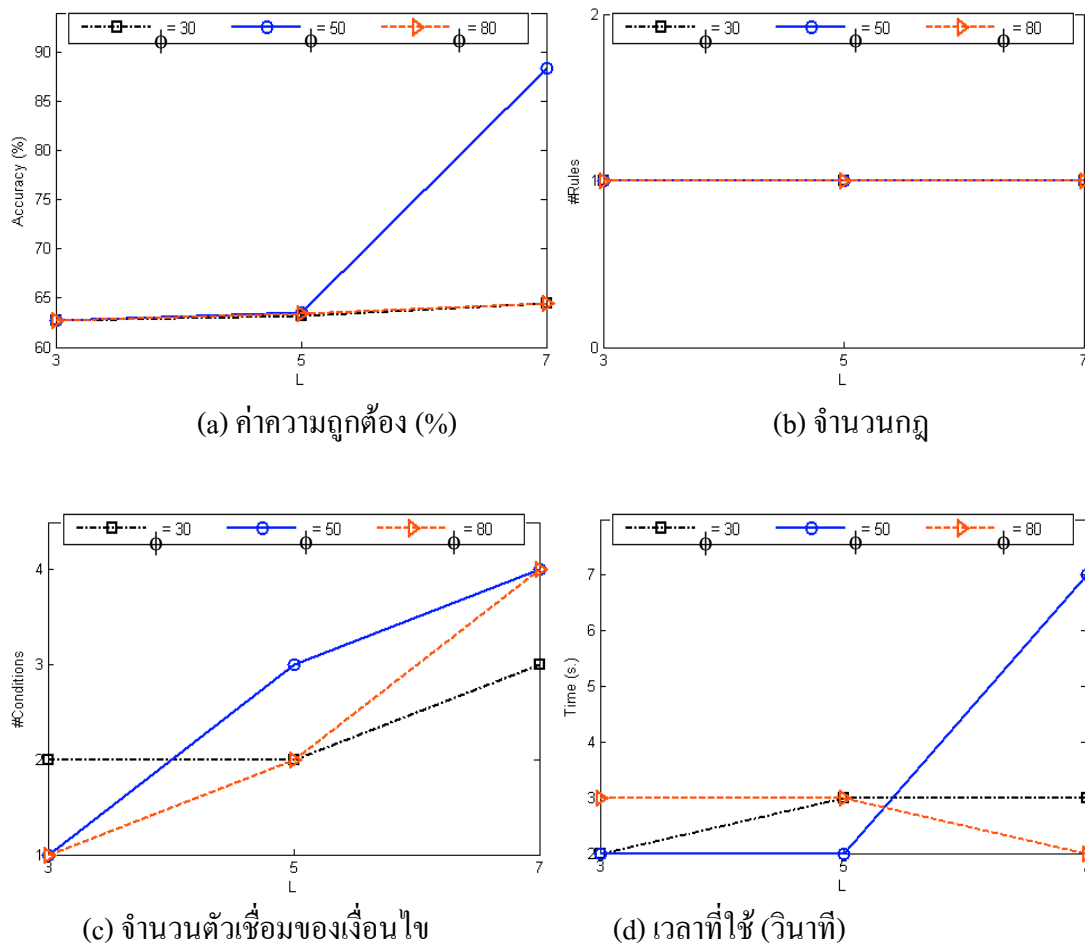


ภาพประกอบ 4.37 กราฟของผลการทดลองการสกัดกฎภาษาธรรมชาติด้วยวิธีการ RENN-P-FIDR ของฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจ เมื่อค่าร้อยละน้ำหนักที่ยอมรับได้

□ = 20

3. ค่าร้อยละที่ยอมรับได้ □ = 40 โครงสร้างโครงข่ายประสาทเทียมหลังผ่านการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) คือ 5:3:1 เมื่อกำหนดให้ค่าความกำกวมที่ยอมรับได้ $\phi = 30$ $\phi = 50$ และ $\phi = 80$ กำหนดจำนวนรูปแบบภาษาธรรมชาติ $L = 3$ $L = 5$ และ $L = 7$ จะได้ผลการทดลองดังภาพประกอบ 4.38 ซึ่งมี แนวโน้มเหมือนกับผลการ

ทดลองเมื่อกำหนดให้ $\phi = 0$ กล่าวคือจากภาพประกอบ 4.38(a) พบว่าค่าความถูกต้องจะขึ้นอยู่กับ
กำหนดจำนวนรูปแบบภาษาธรรมชาติ L กล่าวคือเมื่อกำหนดช่วง L ให้มีค่าจำนวนช่วงมากขึ้นจะ
ได้ค่าความถูกต้องที่สูงกว่าค่า L ที่มีจำนวนช่วงน้อยกว่า นอกจากนี้จะพบว่าเมื่อกำหนดให้ L มีค่า
มากขึ้นก็ยังคงได้จำนวนกฎและจำนวนตัวเชื่อมใกล้เคียงกัน ดังภาพประกอบ 3.38(b) และภาพ
ประกอบ 48(c)

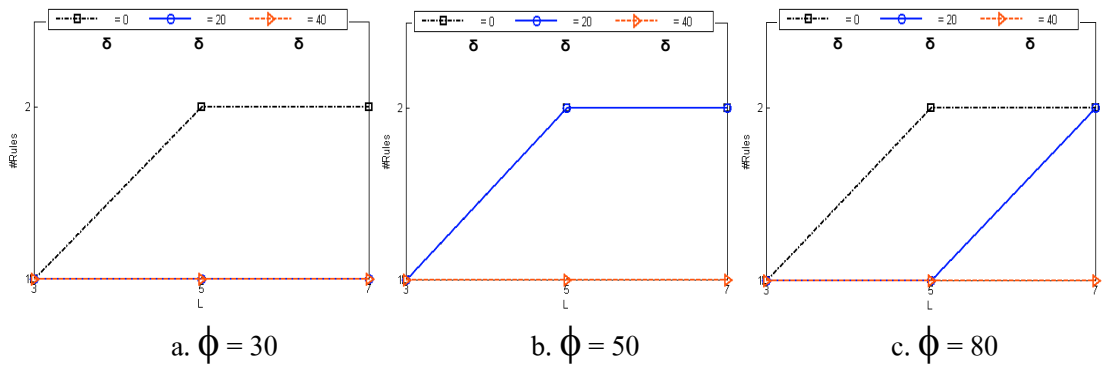


ภาพประกอบ 4.38 กราฟของผลการทดลองการสกัดกฎภาษาธรรมชาติด้วยวิธีการ RENN-P-FIDR
ของฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจ เมื่อกำร้อยละน้ำหนักที่ยอมรับได้

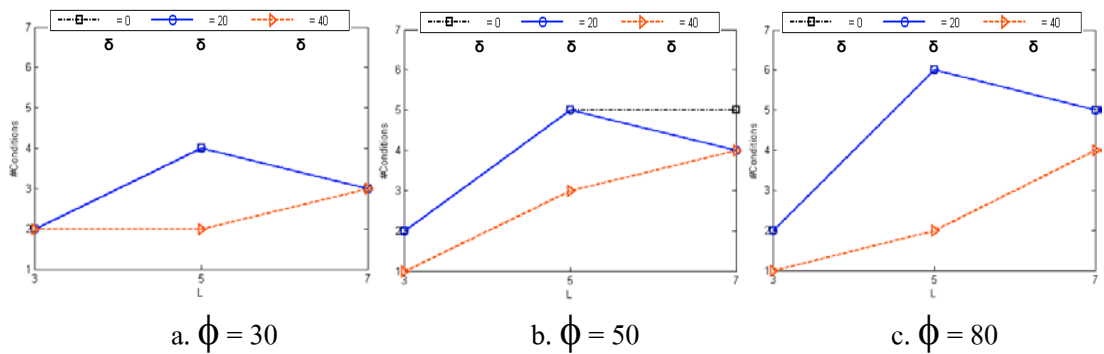
$$\square = 40$$

สามารถเปรียบเทียบผลการทดลองจากการกำหนดค่าร้อยละที่ยอมรับได้ $\square = 0$
 $\square = 20$ และ $\square = 40$ พบว่าเมื่อกำหนดให้ค่าร้อยละที่ยอมรับได้ \square เพิ่มขึ้นจะยังคงทำให้ได้

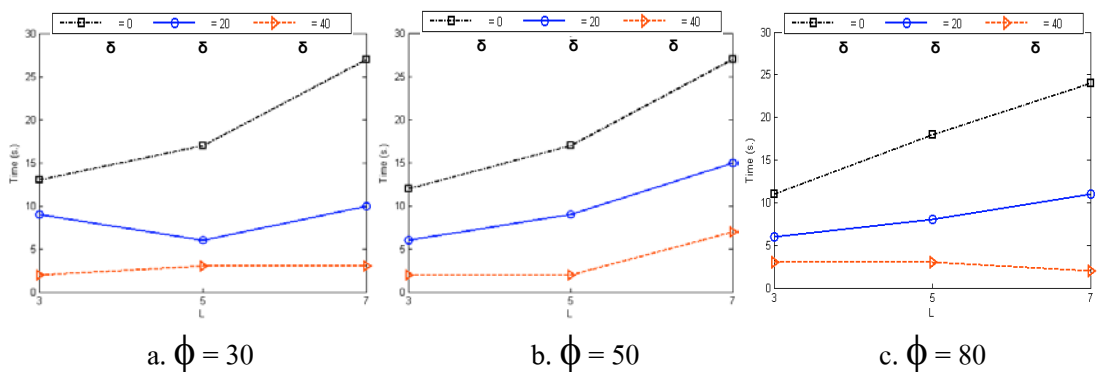
จำนวนกฎ และจำนวนตัวเชื่อมของเงื่อนไขใกล้เคียงกัน ดังภาพประกอบ 4.39 และภาพประกอบ 4.40 เวลาที่ใช้ในการสกัดกฎลดลงดังภาพประกอบ 4.41 แต่อย่างไรก็ตามค่าความถูกต้องก็ ลดลงด้วยดังภาพประกอบ 4.42



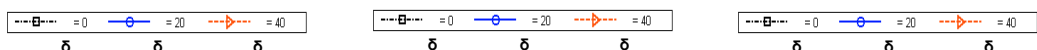
ภาพประกอบ 4.39 จำนวนกฎของฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจ

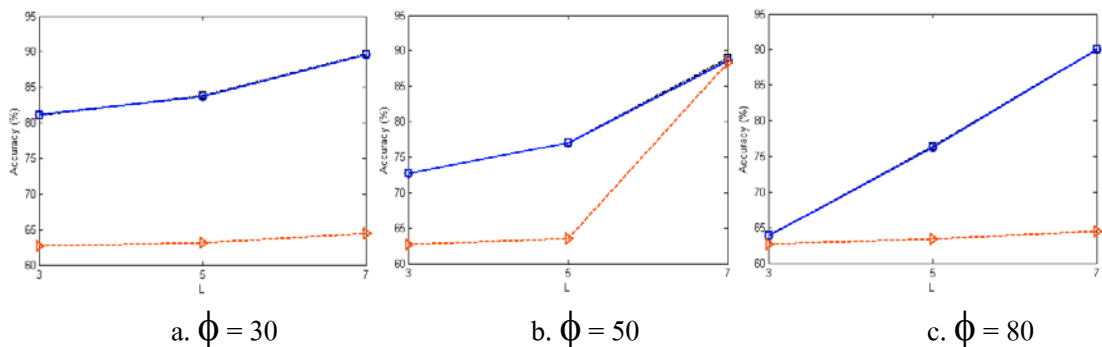


ภาพประกอบ 4.40 จำนวนตัวเชื่อมของเงื่อนไขของฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจ



ภาพประกอบ 4.41 เวลาที่ใช้ในการสกัดกฎ (วินาที) ของฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจ





ภาพประกอบ 4.42 ค่าความถูกต้อง (%) ของฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจ

ตารางที่ 4.27 แสดงผลการทดลองการสกัดกฎภาษาธรรมชาติ RENN-P-FIDR ของฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจทั้งหมด ตัวแปรที่มีผลต่อการสกัดกฎภาษาธรรมชาติ RENN-P-FIDR มี 3 ตัวแปรคือ ค่าร้อยละของน้ำหนักที่ยอมรับได้ \square จำนวนรูปแบบภาษาธรรมชาติ L และค่าความกำกวมที่ยอมรับได้ ϕ ซึ่งแต่ละตัวแปรแนวโน้มผลการทดลองดังนี้

1. ค่าร้อยละของน้ำหนักที่ยอมรับได้ \square ถ้ากำหนดให้ \square มีค่ามากจะตัดโหนดโครงข่ายประสาทเทียมได้มาก กฎที่สกัดได้มีจำนวนกฎและจำนวนตัวเชื่อมของเงื่อนไขลดลง เวลาที่ใช้ในการสกัดกฎลดลงมาก แต่อย่างไรก็ตามค่าความถูกต้องอาจลดลงตามไปด้วย ทั้งนี้เราสามารถกำหนดได้ว่าจะยอมรับค่าความถูกต้องลดลงได้จนถึงระดับใด ซึ่งจากผลการทดลองควรกำหนดค่า \square ให้อยู่ในช่วง 0 ถึง 40

2. จำนวนรูปแบบภาษาธรรมชาติ L ถ้ากำหนดให้ L มีค่าจำนวนช่วงมากขึ้น (เช่น $L = 7$) หรือมีการแบ่งช่วงละเอียดขึ้น จะได้ค่าความถูกต้องที่สูงกว่าค่า L ที่มีจำนวนช่วงน้อยกว่า (เช่น $L = 3$) หรือมีการแบ่งช่วงที่กว้างกว่าในการแทนค่าข้อมูล แต่อย่างไรก็ตามจำนวนกฎและจำนวนตัวเชื่อมของเงื่อนไขยังคงใกล้เคียงกัน

3. ค่าความกำกวมที่ยอมรับได้ ϕ ถ้ากำหนดให้ $\phi = 30$ หมายถึงจะแบ่งช่วงข้อมูลโดยยอมให้ค่าความกำกวมของช่วงไม่เกินร้อยละ 30 จากผลการทดลองแม้ว่ากำหนดให้ ค่าความกำกวมที่ยอมรับได้ ϕ มีค่าแตกต่างกันคือ $\phi = 30$ $\phi = 50$ $\phi = 80$ พบว่าได้ ค่าความถูกต้อง จำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข และเวลาที่ใช้ในการสกัดกฎใกล้เคียงกันเนื่องจากช่วงภาษาธรรมชาติที่แบ่งได้มีค่าของช่วงใกล้เคียงกัน

ตารางที่ 4.27 ผลการทดลองการสกัดกฎด้วย RENN-P-FIDR ของฐานข้อมูลภาวะของผู้ป่วยโรคหัวใจ

δ	Method	L	ϕ	Total Accuracy (%)	#Rules	#Con.	Time (s.)	
	RENN-P-FIDR	0	3	30	81.1	1	2	13

RENN-P-FIDR	0	5	30	83.8	2	4	17
RENN-P-FIDR	0	7	30	89.7	2	3	27
RENN-P-FIDR	0	3	50	72.7	1	2	12
RENN-P-FIDR	0	5	50	77.7	2	5	17
RENN-P-FIDR	0	7	50	89.0	2	5	27
RENN-P-FIDR	0	3	80	63.9	2	2	11
RENN-P-FIDR	0	5	80	76.4	1	6	18
RENN-P-FIDR	0	7	80	90.0	2	5	24
RENN-P-FIDR	20	3	30	81.2	1	2	9
RENN-P-FIDR	20	5	30	83.7	1	4	6
RENN-P-FIDR	20	7	30	89.6	1	3	10

ตารางที่ 4.27 (ต่อ)

δ Method		L	ϕ	Total Accuracy (%)	#Rules	#Con.	Time (s.)
RENN-P-FIDR	20	3	50	72.7	1	2	6
RENN-P-FIDR	20	5	50	77.0	2	5	9
RENN-P-FIDR	20	7	50	88.6	2	4	15
RENN-P-FIDR	20	3	80	63.9	1	2	6
RENN-P-FIDR	20	5	80	76.3	1	6	8
RENN-P-FIDR	20	7	80	90.0	2	5	11
RENN-P-FIDR	40	3	30	62.7	1	2	2
RENN-P-FIDR	40	5	30	63.2	1	2	3
RENN-P-FIDR	40	7	30	64.5	1	3	3
RENN-P-FIDR	40	3	50	62.7	1	1	2
RENN-P-FIDR	40	5	50	63.5	1	3	2
RENN-P-FIDR	40	7	50	88.4	1	4	7
RENN-P-FIDR	40	3	80	62.7	1	1	3
RENN-P-FIDR	40	5	80	63.4	1	2	3
RENN-P-FIDR	40	7	80	64.5	1	4	3

จากผลการทดลองทั้ง 5 ฐานข้อมูลที่กล่าวมาจะเห็นได้ว่ากฎภาษาธรรมชาติที่สกัดได้แพทย์สามารถเข้าใจได้เพราะกฎแทนด้วยภาษาธรรมชาติ เช่น เล็ก กลาง และใหญ่ เป็นต้น ตัวแปรที่มีผลต่อการสกัดกฎภาษาธรรมชาติ RENN-P-FIDR มี 3 ตัวแปรคือ ค่าร้อยละของน้ำหนักที่ยอมรับได้ \square จำนวนรูปแบบภาษาธรรมชาติ L และค่าความกำกวมที่ยอมรับได้ ϕ ซึ่งแต่ละตัวแปรมีแนวโน้มผลการทดลอง ดังนี้

1. ค่าร้อยละของน้ำหนักที่ยอมรับได้ \square ถ้ามีการตัดโหนดโครงข่ายประสาทเทียมก่อนการสกัดกฎ ($\square > 0$) กฎที่ได้จะมีจำนวนกฎ และจำนวนตัวเชื่อมของเงื่อนไขน้อยกว่าไม่มีการตัดโหนดโครงข่ายประสาทเทียมก่อนการสกัดกฎ ($\square = 0$) เวลาที่ใช้ในการสกัดกฎเมื่อมีการตัดโหนดโครงข่ายประสาทเทียมน้อยกว่ามากด้วย แต่อย่างไรก็ตามค่าความถูกต้องอาจน้อยกว่าไม่มีการตัดโหนดโครงข่ายประสาทเทียม ทั้งนี้เราสามารถกำหนดได้ว่าจะยอมรับค่าความถูกต้องน้อยกว่าได้จนถึงระดับใด ซึ่งจากผลการทดลองควรกำหนดให้ \square อยู่ในช่วง 0 ถึง 40 โดยที่ถ้ากำหนดให้ \square มีค่าน้อยจะตัดโหนดโครงข่ายประสาทเทียมได้น้อย ในถ้ามกลับกันถ้ากำหนดให้ \square มีค่ามากจะตัดโหนดโครงข่ายประสาทเทียมได้มาก

2. จำนวนรูปแบบภาษาธรรมชาติ L ถ้ากำหนดให้ L มีค่าจำนวนช่วงมากขึ้น (เช่น $L = 7$) หรือมีการแบ่งช่วงละเอียดขึ้น จะได้ค่าความถูกต้องที่สูงกว่าค่า L ที่มีจำนวนช่วงน้อยกว่า (เช่น $L = 3$) หรือมีการแบ่งช่วงที่กว้างกว่าในการแทนค่าข้อมูล แต่อย่างไรก็ตามจำนวนกฎ และจำนวนตัวเชื่อมของเงื่อนไข และเวลาเพิ่มขึ้นตามค่า L ที่เพิ่มขึ้นด้วย ซึ่งขึ้นอยู่กับ ผู้ใช้จะเลือก

3. ค่าความกำกวมที่ยอมรับได้ ϕ ถ้ากำหนดให้ $\phi = 30$ หมายถึงจะแบ่งช่วงข้อมูลโดยยอมให้ค่าความกำกวมของช่วงไม่เกินร้อยละ 30 จากผลการทดลองแม้ว่ากำหนดให้ ค่าความกำกวมที่ยอมรับได้ ϕ มีค่าแตกต่างกันคือ $\phi = 30$ $\phi = 50$ $\phi = 80$ พบว่าได้ค่าความ-ถูกต้อง จำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข และเวลาที่ใช้ในการสกัดกฎใกล้เคียงกันเนื่องจากช่วงภาษาธรรมชาติที่แบ่งได้มีค่าของช่วงใกล้เคียงกัน ซึ่งได้ผลซึ่งเกิดมาจากการรวมช่วง นอกจากนี้ค่าความถูกต้อง จำนวนกฎ จำนวนตัวเชื่อมของเงื่อนไข และเวลาที่ใช้ในการสกัดกฎ ยังขึ้นอยู่กับขั้นตอนการสอนโครงข่ายประสาทเทียม ซึ่งส่งผลมายังขั้นตอนการสร้างกฎอีกด้วย