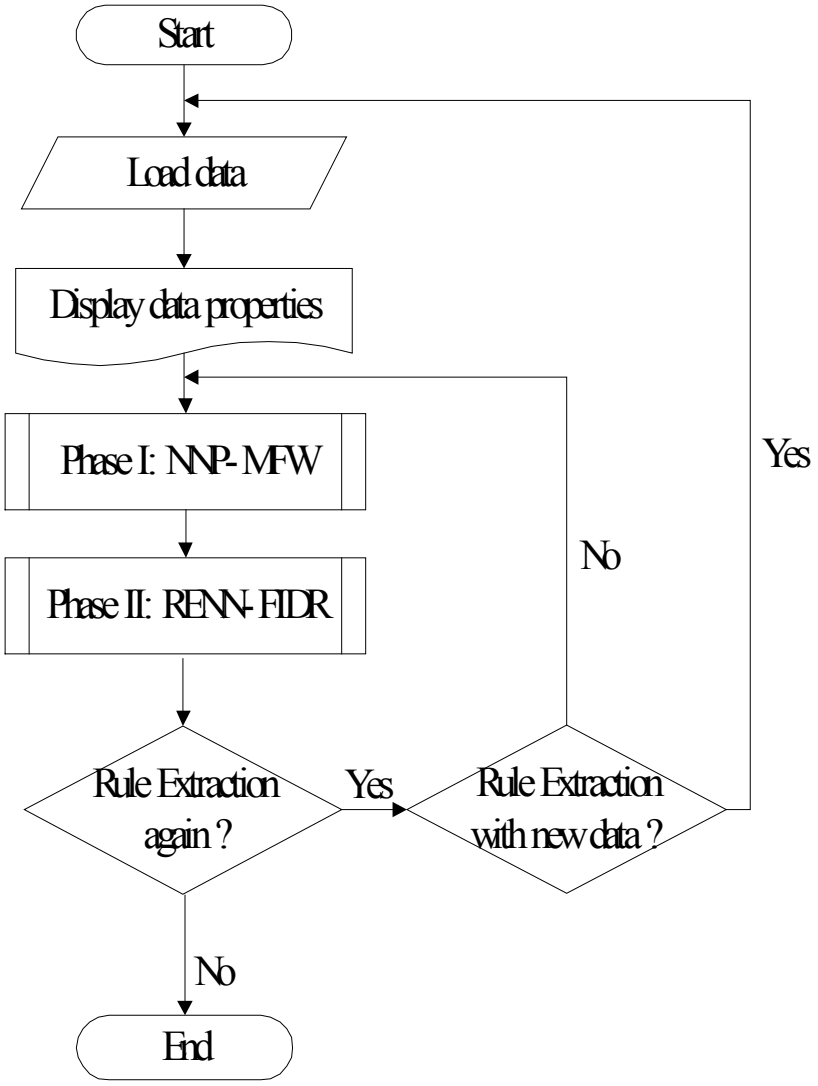
บทที่ 5

โปรแกรมการสกัดกฎภาษาธรรมชาติโดยใช้โครงข่ายประสาทเทียม

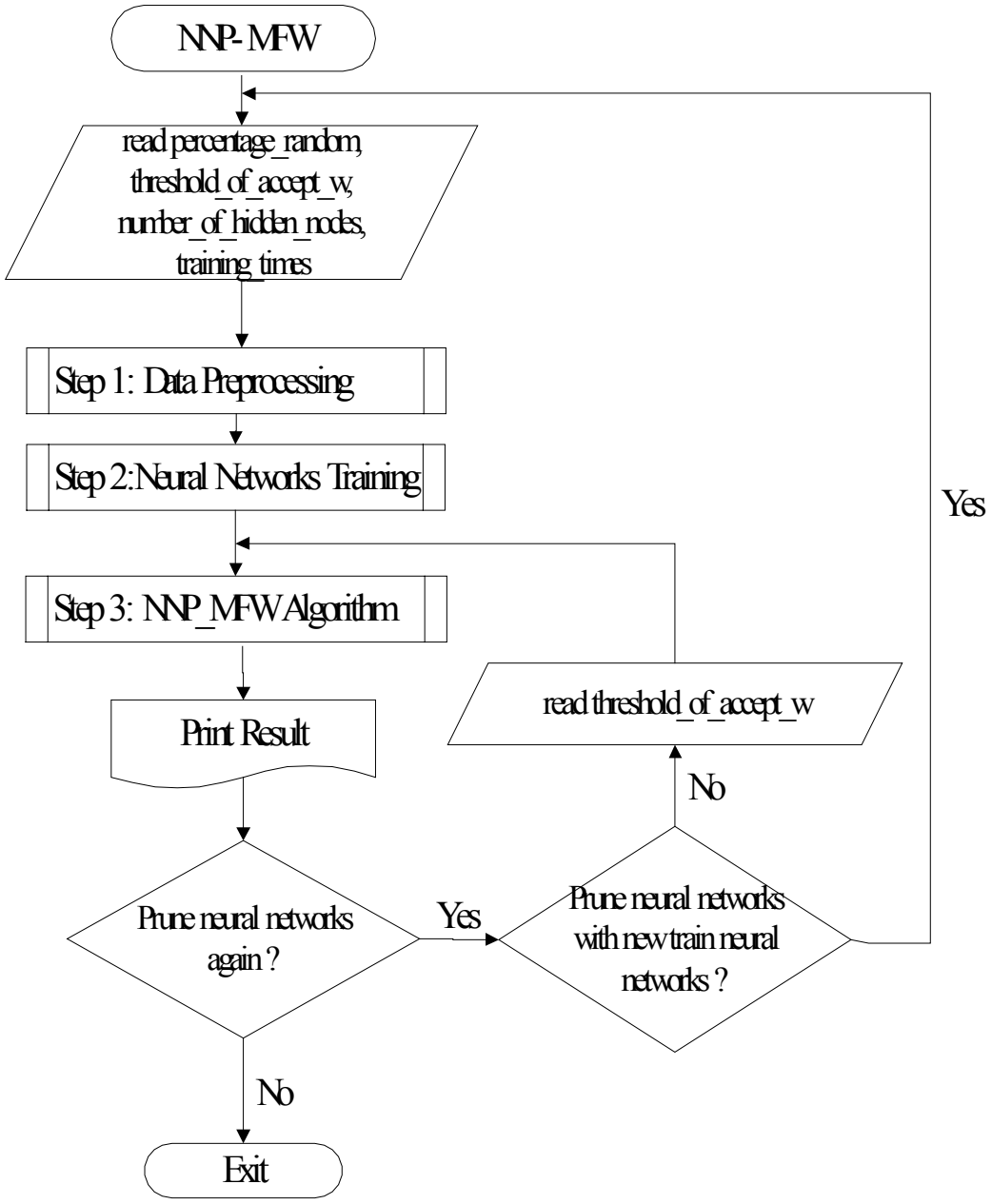
เพื่อให้ผู้ใช้สามารถใช้โปรแกรมการสกัดกฎภาษาธรรมชาติโดยใช้โครงข่ายประสาทเทียมได้ง่าย จึงได้ออกแบบส่วนการติดต่อกับผู้ใช้ให้อยู่ในรูปแบบที่เป็นมิตรและง่ายต่อการใช้งานด้วย Graphic User Interface ในการทำงานของโปรแกรมจะอธิบายด้วยผังการทำงานของโปรแกรม นอกจากนี้ยังแสดงตัวอย่างผลการทำงานของโปรแกรมด้วย

5.1 ผังการทำงานของโปรแกรม

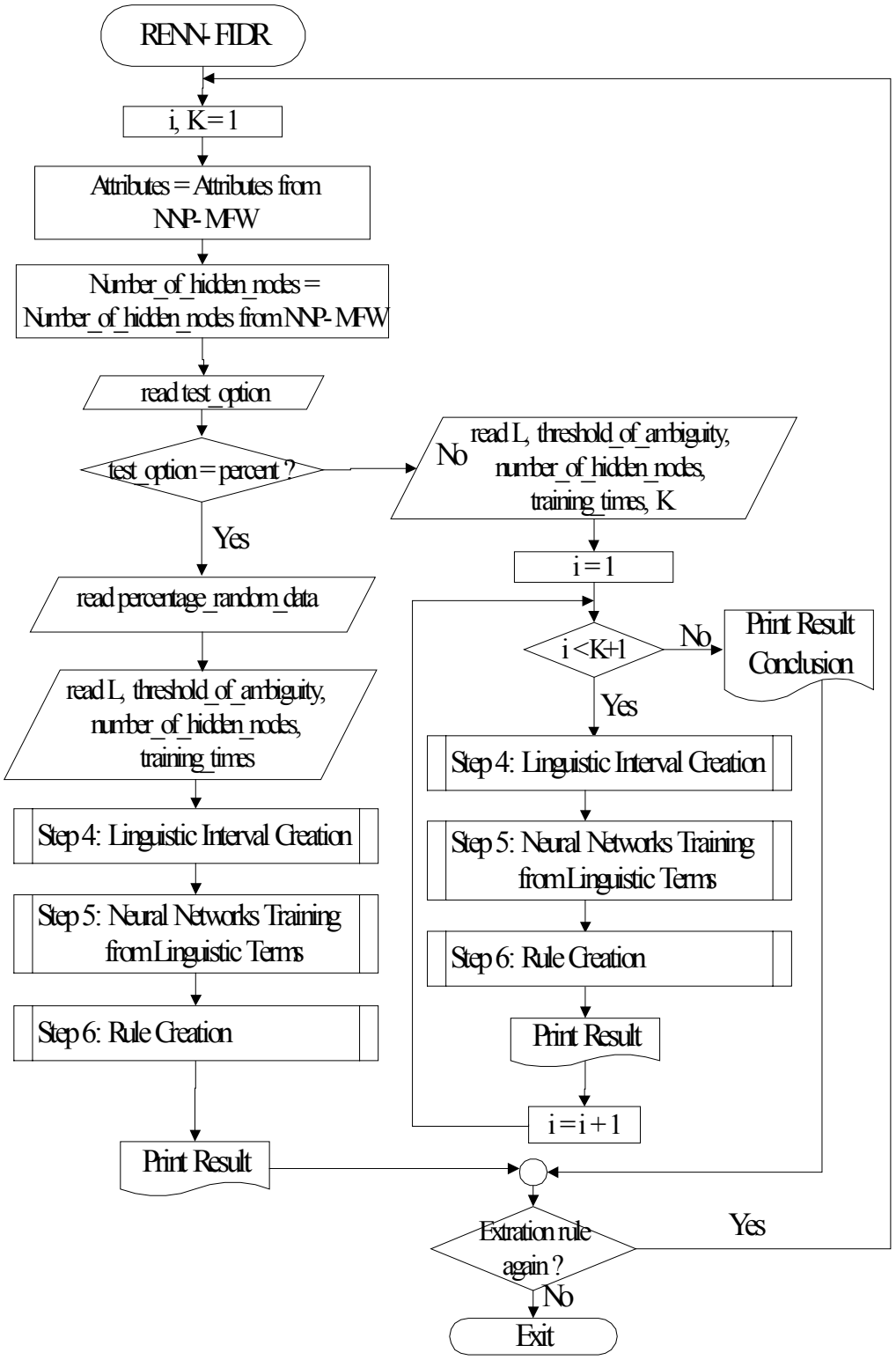
ในทำงานของโปรแกรมการสกัดกฎภาษาธรรมชาติโดยใช้โครงข่ายประสาทเทียมสามารถอธิบายด้วยผังการทำงานของโปรแกรมดังภาพประกอบ 5.1 การทำงานของขั้นตอนหลักที่ I การตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) สามารถอธิบายด้วยผังการทำงานดังภาพประกอบ 5.2 และการทำงานของขั้นตอนหลักที่ II การสกัดกฎภาษาธรรมชาติจากโครงข่ายประสาทเทียมโดยการแทนค่าความถี่ของช่วงข้อมูล (RENN-FIDR) สามารถอธิบายด้วยผังการทำงานดังภาพประกอบ 5.3



ภาพประกอบ 5.1 ผังการทำงานของโปรแกรมการสกัดกฎภาษาธรรมชาติโดยใช้โครงข่ายประสาทเทียม



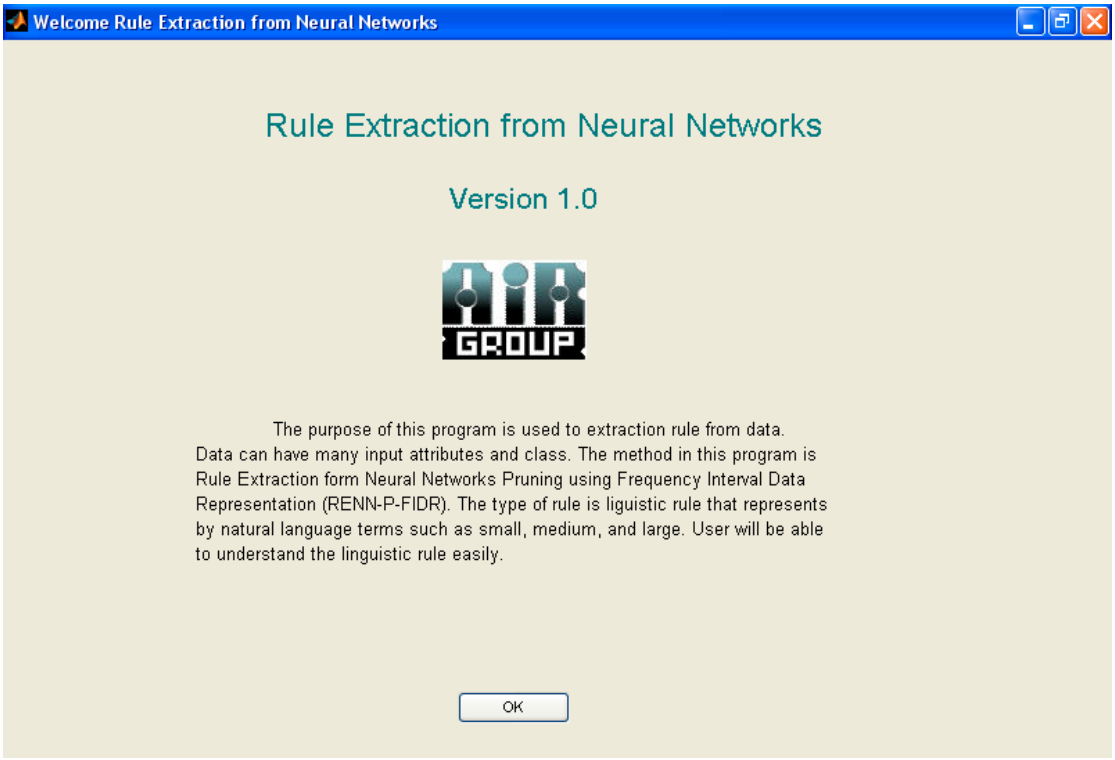
ภาพประกอบ 5.2 ผังการทำงานของขั้นตอนหลักที่ I การตัดโหนดโครงข่ายประสาทเทียม
โดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW)



ภาพประกอบ 5.3 ผังการทำงานของการทำงานของขั้นตอนหลักที่ II การสกัดกฎภาษาธรรมชาติจากโครงข่ายประสาทเทียมโดยการแทนค่าความถี่ของช่วงข้อมูล (RENN-FIDR)

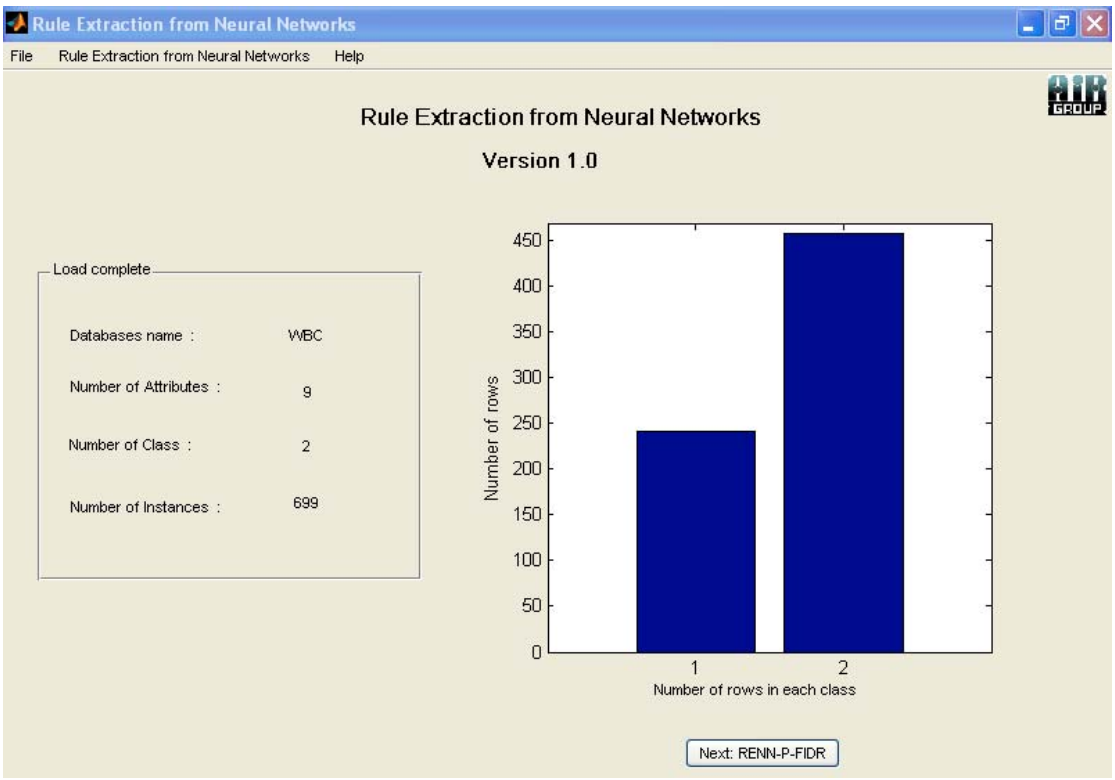
5.2 ผลการทำงานของโปรแกรม

เมื่อเปิดโปรแกรมการสกัดกฎภาษาธรรมชาติโดยใช้โครงข่ายประสาทเทียมจะปรากฏหน้าจอที่อธิบายข้อมูลเบื้องต้นเกี่ยวกับโปรแกรม ดังภาพประกอบ 5.4



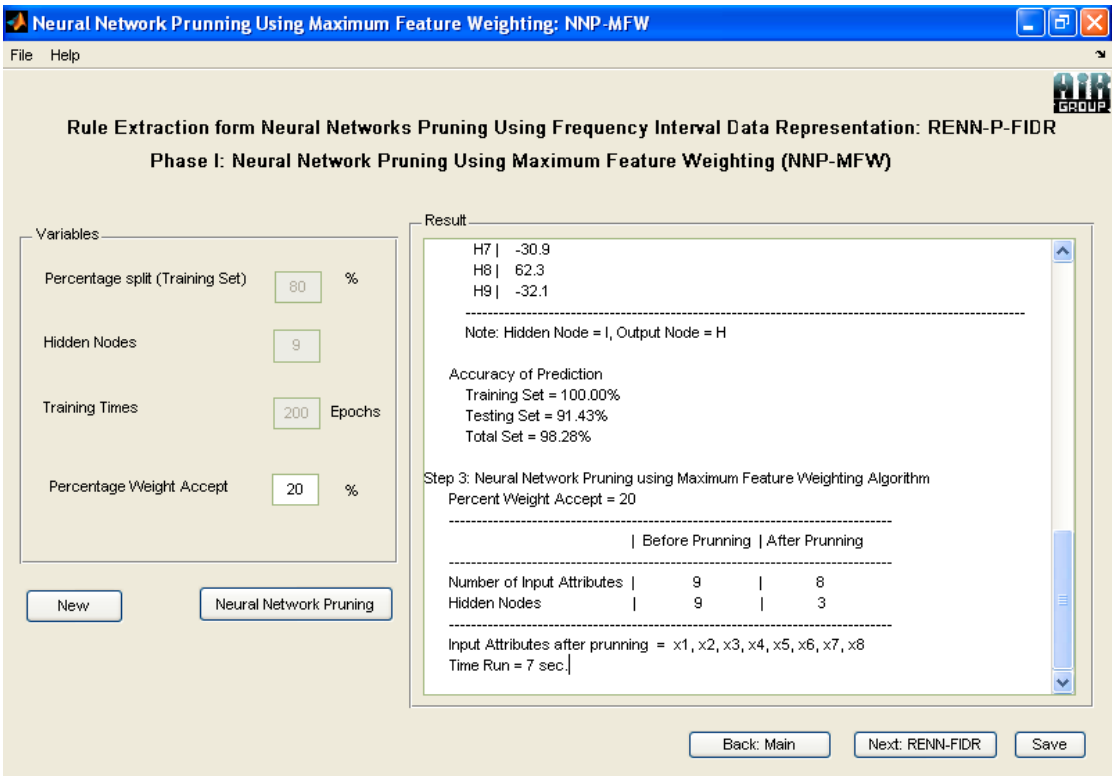
ภาพประกอบ 5.4 โปรแกรมการสกัดกฎภาษาธรรมชาติโดยใช้โครงข่ายประสาทเทียม

เริ่มต้นสกัดกฎภาษาธรรมชาติโดยใช้โครงข่ายประสาทเทียมได้โดยคลิกปุ่ม OK และโหลดไฟล์ข้อมูลเข้าสู่โปรแกรม โปรแกรมจะแสดงรายละเอียดของข้อมูล ตัวอย่างเมื่อโหลดไฟล์ฐานข้อมูลโรคมะเร็งเต้านมดังภาพประกอบ 5.5

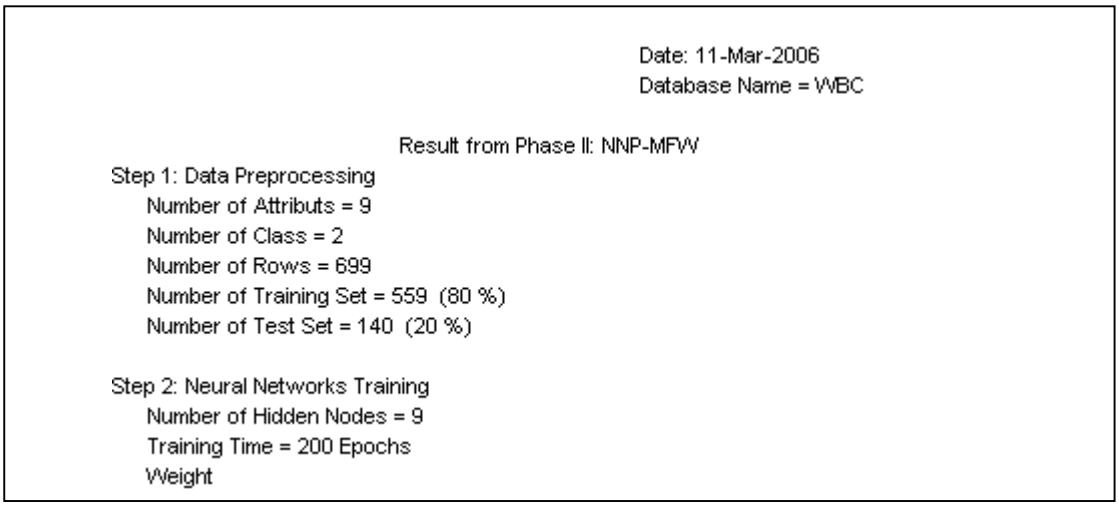


ภาพประกอบ 5.5 ตัวอย่างการแสดงรายละเอียดของฐานข้อมูลโรคมะเร็งเต้านม

คลิกเลือกเมนู Rule Extraction> Rule Extraction form Neural Networks Pruning using Frequency Interval Data Representation: RENN-P-FIDR หรือคลิกปุ่ม Next: RENN-P-FIDR เพื่อสกัดกฎภาษาธรรมชาติโดยใช้โครงข่ายประสาทเทียมที่ผ่านการตัดโหนดโดยใช้การแทนค่าความถี่ของช่วงข้อมูล ซึ่งจะเข้าสู่หน้าจอขั้นตอนหลักที่ I การตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) ตัวอย่างผลลัพธ์เมื่อกำหนดให้ร้อยละของข้อมูลชุดสอน = 80 ร้อยละของน้ำหนักที่ยอมรับได้ = 20 จำนวนโหนดซ่อน = 9 โหนด และจำนวนรอบในการเรียนรู้ = 200 รอบ แสดงดังภาพประกอบ 5.6 ซึ่งมีรายละเอียดดังภาพประกอบ 5.7



ภาพประกอบ 5.6 หน้าจอผลลัพธ์ของขั้นตอนหลักที่ I (NNP-MFW)



ภาพประกอบ 5.7 ผลลัพธ์ของขั้นตอนหลักที่ I (NNP-MFW) ในแต่ละขั้นตอน

Input Layer to Hidden Layer

I/H	H1	H2	H3	H4	H5	H6	H7	H8	H9
I1	-0.1	-0.5	-0.4	1.5	-5.6	-4.5	2.1	2.4	1.0
I2	-3.2	-2.0	2.8	-2.5	8.4	-0.2	1.9	-0.3	-1.3
I3	1.3	0.4	-0.6	-0.1	4.3	-2.3	0.2	1.8	-1.3
I4	1.5	-1.7	-2.0	-2.0	4.3	-3.3	-0.1	0.9	-2.1
I5	-0.1	-3.9	-0.5	-3.1	-5.7	-2.0	-4.2	4.0	-0.8
I6	2.7	3.3	-2.5	2.9	-0.3	4.9	2.0	3.9	4.3
I7	-2.1	4.8	-0.8	2.7	1.1	-1.8	-4.4	-6.0	-0.8
I8	0.9	-1.1	-0.8	2.2	-1.0	0.8	-4.6	-2.7	1.2
I9	-4.3	5.3	3.6	0.5	-4.8	5.7	2.5	0.4	3.1

Note: Input Node = I, Hidden Node = H

Hidden Layer to Output Layer

H/O	O1
H1	-34.0
H2	3.5
H3	-2.7
H4	61.7
H5	34.2
H6	-32.7
H7	-30.9
H8	62.3
H9	-32.1

Note: Hidden Node = I, Output Node = H

Accuracy of Prediction
Training Set = 100.00%
Testing Set = 91.43%
Total Set = 98.28%

Step 3: Neural Network Pruning using Maximum Feature Weighting Algorithm
Percent Weight Accept = 20

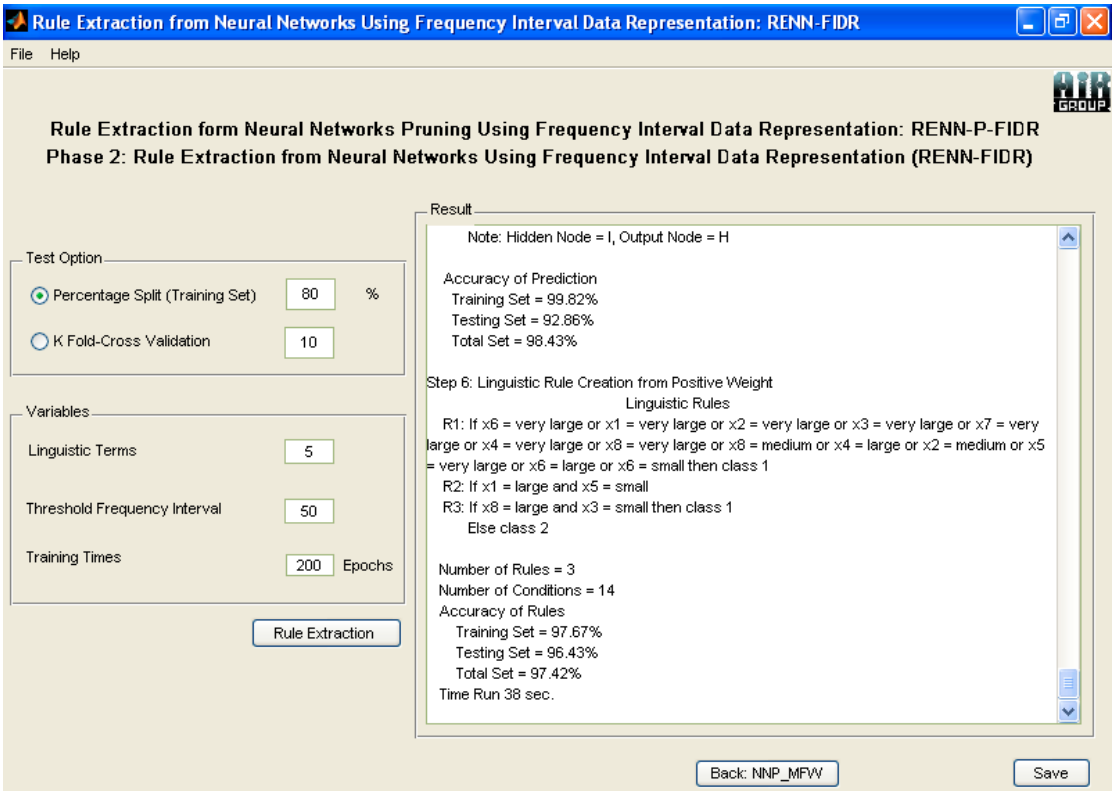
	Before Prunning	After Prunning
Number of Input Attributes	9	8
Hidden Nodes	9	3

Input Attributes after prunning = x1, x2, x3, x4, x5, x6, x7, x8
Time Run = 7 sec.

ภาพประกอบ 5.7 (ต่อ)

เมื่อคลิกปุ่ม Next: RENN-FIDR จะปรากฏหน้าจอขั้นตอนหลักที่ II การสกัดกฎภาษาธรรมชาติจากโครงข่ายประสาทเทียมโดยการแทนค่าความถี่ของช่วงข้อมูล (RENN-

FIDR) ตัวอย่างผลลัพธ์เมื่อเลือกสุ่มข้อมูลแบบร้อยละ โดยกำหนดให้ร้อยละของข้อมูล ชุดสอน = 80 เปอร์เซ็น จำนวนรูปแบบภาษาธรรมชาติ = 5 และจำนวนรอบในการเรียนรู้ = 200 รอบ แสดงดังภาพประกอบ 5.8 ซึ่งมีรายละเอียดดังภาพประกอบ 5.9



ภาพประกอบ 5.8 หน้าจอผลลัพธ์ในขั้นตอนหลักที่ II (RENN-FIDR)

Date: 11-Mar-2006
Database Name = WBC
Percent Weight Accept = 20

Result from Phase II: RENN-FIDR

Test Option = Percentage
Number of Training Set = 559 (80 %)
Number of Testing Set = 140 (20 %)
Number of Total Set = 699 (100 %)

Step Linguistic Intervals Creation from Frequency Algorithm
Linguistic Terms = 5
Threshold Frequency Interval = 50

Linguistic Intervals

x / ling. terms	very small,	small,	medium,	large,	very large
x1	[1.0,5.0),	[5.0,6.0),	[6.0,7.0),	[7.0,9.0),	[9.0,10.0]
x2	[1.0,4.0),	[4.0,5.0),	[5.0,6.0),	[6.0,10.0),	[10.0,10.0]
x3	[1.0,3.0),	[3.0,4.0),	[4.0,6.0),	[6.0,9.0),	[9.0,10.0]
x4	[1.0,3.0),	[3.0,4.0),	[4.0,7.0),	[7.0,10.0),	[10.0,10.0]
x5	[1.0,3.0),	[3.0,4.0),	[4.0,6.0),	[6.0,10.0),	[10.0,10.0]
x6	[1.0,6.0),	[6.0,7.0),	[7.0,9.0),	[9.0,10.0),	[10.0,10.0]
x7	[1.0,4.0),	[4.0,5.0),	[5.0,6.0),	[6.0,8.0),	[8.0,10.0]
x8	[1.0,3.0),	[3.0,4.0),	[4.0,6.0),	[6.0,10.0),	[10.0,10.0]

Step 5: Neural Networks Training with Linguistic Terms
Number of Hidden Nodes = 3
Training Times = 200 Epochs
Weight

Input Layer to Hidden Layer

I/H	H1	H2	H3
I1_1	-2.4	5.9	4.1
I1_2	-4.1	0.8	-0.0
I1_3	3.4	1.3	-4.6
I1_4	1.2	-2.1	2.6
I1_5	2.7	-9.1	-1.2
I2_1	0.8	-5.9	-2.6
I2_2	2.1	6.9	1.2
I2_3	0.5	-3.9	-0.9
I2_4	0.1	-0.3	-1.3
I2_5	-0.5	-1.9	2.3
I3_1	-2.2	0.7	-2.7
I3_2	1.6	0.5	0.9
I3_3	5.1	-3.0	-4.6
I3_4	-0.8	-0.0	4.8
I3_5	-0.6	-2.5	2.2
I4_1	-1.3	-1.5	-5.7
I4_2	-1.5	2.4	-0.0
I4_3	2.6	0.8	-0.3
I4_4	0.1	-1.0	3.0
I4_5	0.1	-3.3	5.5

ภาพประกอบ 5.9 ผลลัพธ์ในขั้นตอนหลักที่ II (RENN-FIDR) ในแต่ละขั้นตอน

```
I5_1 |   -2.9   4.1   4.5
I5_2 |   1.0   0.4   -4.5
I5_3 |   -2.4   3.3   0.6
I5_4 |   3.1   1.1   1.3
I5_5 |   0.7   -7.7   0.8
I6_1 |   -2.4   3.6   -1.6
I6_2 |   2.4   -2.1   1.2
I6_3 |   0.8   -0.9   -4.6
I6_4 |   1.5   -2.3   5.4
I6_5 |   0.7   -1.3   2.5
I7_1 |   0.2   3.7   1.4
I7_2 |   4.6   -0.6   -2.1
I7_3 |   -0.6   -1.2   -3.4
I7_4 |   0.1   2.2   0.9
I7_5 |   -0.0   -7.7   0.4
I8_1 |   -2.6   1.1   -1.9
I8_2 |   2.5   -2.8   -3.3
I8_3 |   2.0   5.5   4.0
I8_4 |   1.8   -2.5   -0.4
I8_5 |   0.6   -2.8   -1.1
-------------------------------------------------------------------------------------------------
      Note: Input Node = I, Hidden Node = H

                              Hidden Layer to Output Layer
-------------------------------------------------------------------------------------------------
H/O   |   O1
-------------------------------------------------------------------------------------------------
H1    |   7.2
H2    |   -28.8
H3    |   29.6
-------------------------------------------------------------------------------------------------
      Note: Hidden Node = I, Output Node = H

  Accuracy of Prediction
    Training Set = 99.82%
    Testing Set = 92.86%
    Total Set = 98.43%

Step 6: Linguistic Rule Creation from Positive Weight
                              Linguistic Rules
  R1: If x6 = very large or x1 = very large or x2 = very large or x3 = very large or x7 = very
large or x4 = very large or x8 = very large or x8 = medium or x4 = large or x2 = medium or x5
= very large or x6 = large or x6 = small then class 1
  R2: If x1 = large and x5 = small
  R3: If x8 = large and x3 = small then class 1
      Else class 2

  Number of Rules = 3
  Number of Conditions = 14
  Accuracy of Rules
    Training Set = 97.67%
    Testing Set = 96.43%
    Total Set = 97.42%
  Time Run 38 sec.
```

ภาพประกอบ 5.9 (ต่อ)

ผู้ใช้สามารถใช้งานโปรแกรมการสกัดกฎภาษาธรรมชาติโดยใช้โครงข่ายประสาทเทียมได้ง่าย เนื่องจากมีส่วนการติดต่อกับผู้ใช้เป็นแบบ Graphic User Interface ผู้ใช้สามารถกำหนดตัวแปรต่างๆ ได้เอง ในการทำงานของโปรแกรมมี 3 หน้าจอหลักคือ หน้าจอสำหรับโหลดข้อมูล หน้าจอสำหรับการตัดโหนดโครงข่ายประสาทเทียมโดยใช้ค่าน้ำหนักสูงสุด (NNP-MFW) และหน้าจอสำหรับการสกัดกฎภาษาธรรมชาติจากโครงข่ายประสาทเทียมโดย การแทนค่าความถี่ของช่วงข้อมูล (RENN-FIDR) โดยมีการแสดงผลลัพธ์ของการทำงานในแต่ละ ขั้นตอนอย่างชัดเจนและเข้าใจได้ง่าย โดยเฉพาะอย่างยิ่งกฎภาษาธรรมชาติที่สกัดได้