CHAPTER III

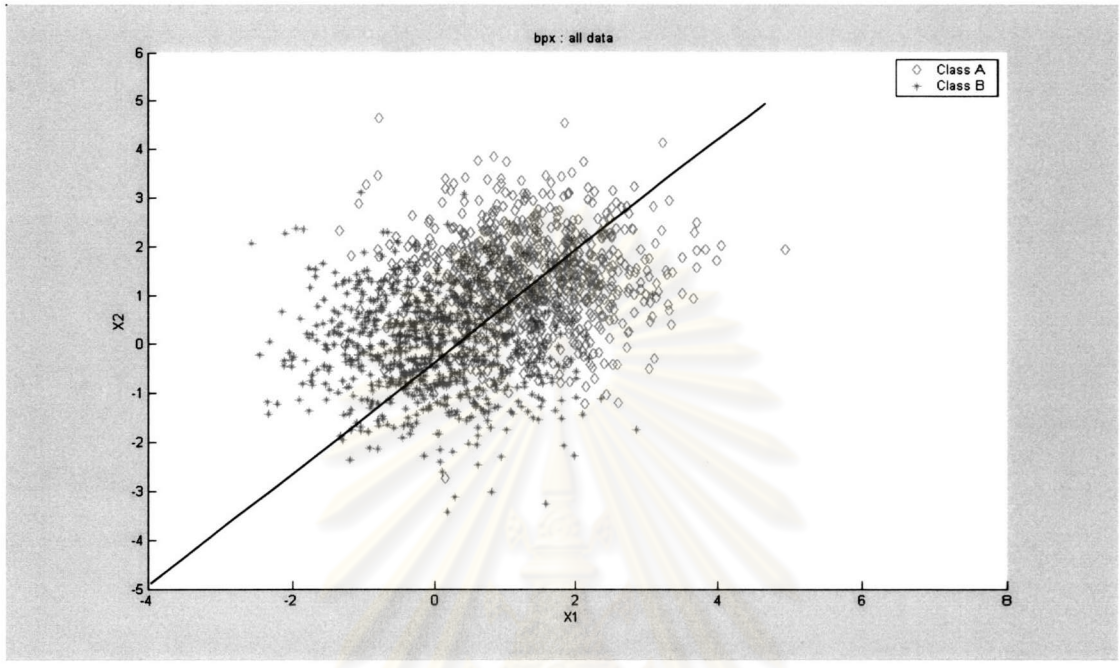
EXPERIMENTAL RESULTS

We tested the rule extraction approach by using the generated ambiguous data set and the real world classification problems such as the Glass data set [27], Iris data set [28] and Wisconsin breast cancer data set [29].

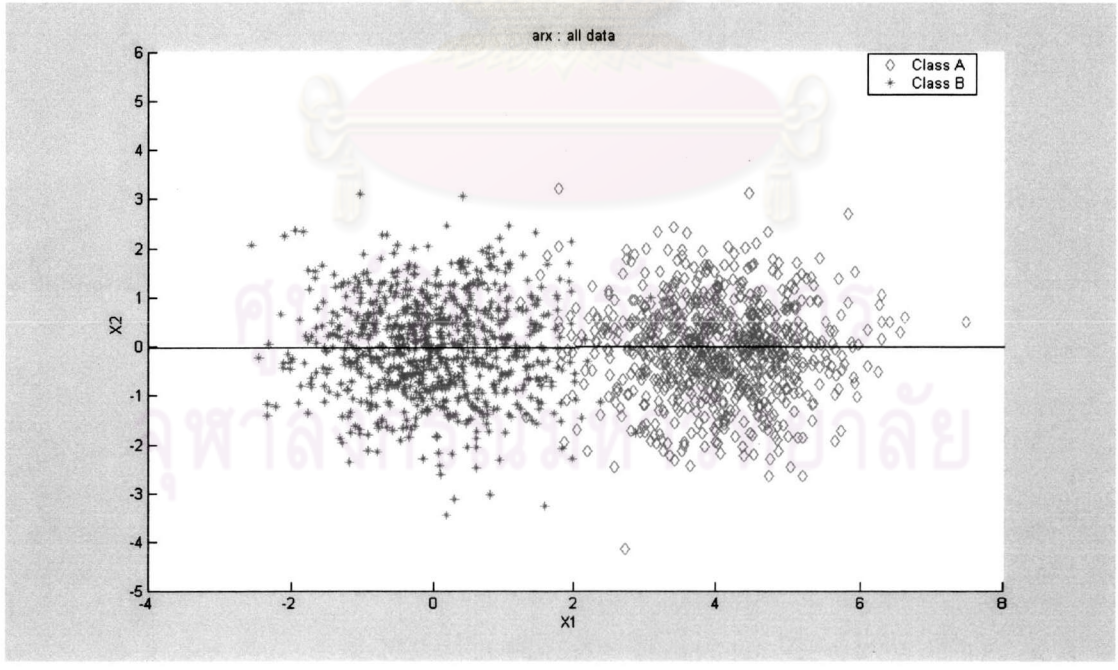
3.1 Generated Ambiguous Data Set.

We generate data set of two classes (class A and class B) with two input features (x_1 and x_2). There are 1,500 points for both class A and class B. The data are generated in two orientations: (i) on the x-axis orientation and (ii) on the diagonal orientation. There are three different degrees of overlapping data between two classes at 2σ, $2\sqrt{2}\sigma$, and 4σ distances apart, where σ "sigma" is the standard deviation (SD) of the data. The six data sets for the experiment are (1) bpx, (2) apx, (3) bqx, (4) aqx, (5) brx, and (6) arx. The representation of the data set names are as follows. The first character in the data set names indicates the orientation of the two classes, where "*a*" indicates *x-axis orientation* and "*b*" indicates *diagonal orientation*. The second character of the data set names indicates the SD of the distances between two classes of overlapping data, where "*p*" indicates 2σ ambiguous SD, "*q*" indicates $2\sqrt{2}\sigma$ ambiguous SD, and "*r*" indicates 4σ ambiguous SD. The examples of the generated ambiguous data sets are given in Figure 3. Figure 3.1(a) shows the most complex data set with 2σ ambiguous SD of overlapping data on the diagonal orientation from the data set "(1) bpx" while Figure 3.1(b) shows the less complex data set with 4σ ambiguous SD of overlapping data on the x-axis

orientation from the data set "(6) arx".



(a)



(b)

Figure 3.1 The generated ambiguous data set. (a) Data set with bpx format has the most complex data set. (b) Data set with arx format has the less complex data set.

In the first step, a neural network was trained by using the Stuttgart Neural Network Simulator (SNNS) software. The data set was divided into two sets. The training set contained 750 input vectors and the test set contained 750 input vectors selected randomly. The network has two input units (x_1 and x_2), ten hidden nodes, and one output node (2:10:1). The value of the output node equals to 1 when the training input is in class A and 0 when it is in class B. In the REAP algorithm, the parameter setting for the maximum number of accepted ambiguous activation value ψ is 5%. The threshold value for the feature extraction ω is 50%. In the RECF algorithm, the threshold value for the accepting the certainty factor value ξ is 50%.

The result from RECF algorithm is shown in Table 3.1. The names of all six data sets are (1) bpx, (2) apx, (3) bqx, (4) aqx, (5) brx, and (6) arx while the accuracy results for the SNNS are 84.1%, 87.7%, 91.7%, 93.4%, 97.3%, and 99%, respectively. The accuracy results of all six data sets from the test set for the RECF algorithm and RENL algorithm with parameter setting ψ = 5% and ω = 50% are 63.7%, 82.7%, 81.2%, 92.8%, 97.2%, and 98.3%, respectively. The accuracy results of all six data sets from the test set for the RECF algorithm and RENL algorithm with parameter setting ψ = 1% and ω = 40% are 56.9%, 67.7%, 72.4%, 88.0%, 92.5%, and 98.0%, respectively. A number of rules with parameter setting ψ = 5% and ω = 50% for of all six data sets are 5, 4, 5, 3, 5, and 3, respectively. A number of rules with parameter setting ψ = 1% and ω = 40% of all six data sets are 5, 3, 5, 3, 5, and 3, respectively. For the diagonal orientation, there are five rules for all three standard deviations. But for the x-axis orientation, there are three to four rules for 2σ ambiguous standard deviation and three rules for $2\sqrt{2}\sigma$ and 4σ ambiguous standard deviation.

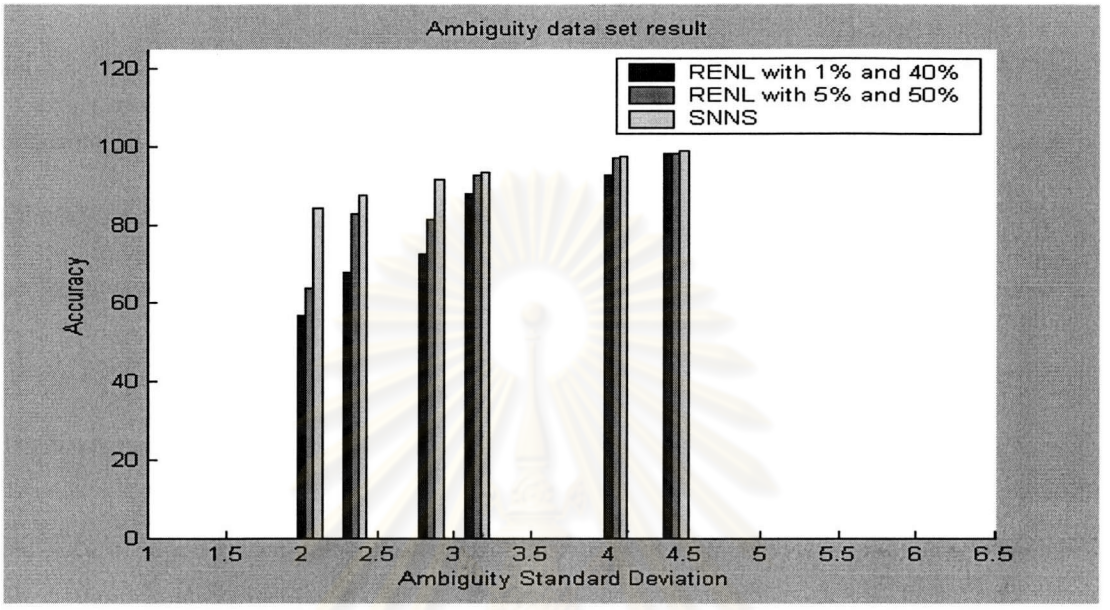
Table 3.1 The experimental results of generated ambiguous data set.

Data Set Name	Ambiguous Orientation	Ambiguous Standard Derivation	SNNS (%)	RECF& RENL (%) $\psi = 5\%$ $\omega = 50\%$	RECF& RENL (%) $\psi = 1\%$ $\omega = 40\%$	Number of Rules $\psi = 5\%$ $\omega = 50\%$	Number of Rules $\psi = 1\%$ $\omega = 40\%$
(1) bpx	diagonal	2σ	84.1	63.7	56.9	5	5
(2) apx	x-axis	2σ	87.7	82.7	67.7	4	3
(3) bqx	diagonal	$2\sqrt{2}\sigma$	91.7	81.2	72.4	5	5
(4) aqx	x-axis	$2\sqrt{2}\sigma$	93.4	92.8	88.0	3	3
(5) brx	diagonal	4σ	97.4	97.2	92.5	5	5
(6) arx	x-axis	4σ	99.0	98.3	98.0	3	3

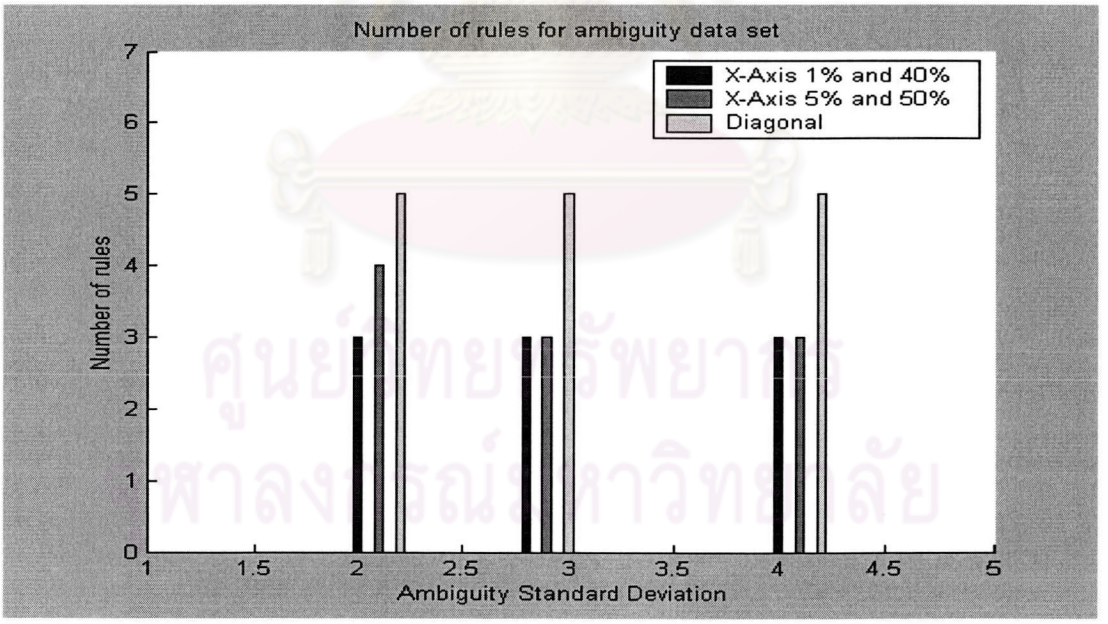
The comparison of the accuracy among RENL algorithm with parameter setting $\psi = 5\%$ and $\omega = 50\%$, $\psi = 1\%$ and $\omega = 40\%$, and SNNS is shown in Figure 3.2(a). The more ambiguous standard deviation (or less complex data set) is, the higher the classification accuracy for both SNNS and RENL algorithm will be. The comparison of the number rules for x-axis orientation with parameter setting $\psi = 5\%$ and $\omega = 50\%$, $\psi = 1\%$ and $\omega = 40\%$, and the diagonal orientation is shown in Figure 3.2(b). The number of rules for diagonal orientation is higher because the diagonal orientation is more complex than the x-axis orientation. The more ambiguous standard deviation (or less complex data set) is, the smaller numbers of rules for x-axis orientation will be.

The CF rule with parameter setting $\psi = 5\%$ and $\omega = 50\%$ from all six data sets is shown in Table 3.2. From REAP algorithm, the feature extraction eliminates the input feature x_2 from the data set with x-axis orientation. So, the rules from the x-axis orientation consist of input feature x_1 only. Meanwhile, the rules from the diagonal

orientation consist of both input features x_1 and x_2. The CF values that represent the certainty factor of the data are indicated on each rule.



(a)



(b)

Figure 3.2 The experimental results of generated ambiguous data set.
(a) The comparison of the accuracy. (b) The comparison of number of rules.

Table 3.2 The CF rules of generated ambiguous data set with parameter setting $\psi = 5\%$ and $\omega = 50\%$.

Data Set Name	Rule No	CF Rules		
(1) bpx	$R11_{cf}$	**If** $x_1 \geq 2$	**Then** class A	$CF_A = 95$
[diagonal, 2σ]	$R12_{cf}$	**If** $x_2 \geq 2$	**Then** class A	$CF_A = 96$
	$R13_{cf}$	**If** $x_1 \leq -1.1$	**Then** class B	$CF_B = 97$
	$R14_{cf}$	**If** $-1 \leq x_1 \leq 1.9$ and $-1 \leq x_2 \leq 1.9$	**Then** class B	$CF_B = 55$
	$R15_{cf}$	**If** $x_2 \leq -1.1$	**Then** class B	$CF_B = 97$
(2) apx	$R21_{cf}$	**If** $x_1 \geq 1.6$	**Then** class A	$CF_A = 92$
[x-axis, 2σ]	$R22_{cf}$	**If** $0.7 \leq x_1 \leq 1.5$	**Then** class A	$CF_A = 58$
	$R23_{cf}$	**If** $-0.4 \leq x_1 \leq 0.5$	**Then** class B	$CF_B = 86$
	$R24_{cf}$	**If** $x_1 \leq -0.5$	**Then** class B	$CF_B = 98$
(3) bqx	$R31_{cf}$	**If** $x_1 \geq 1.5$	**Then** class A	$CF_A = 91$
[diagonal, $2\sqrt{2}\sigma$]	$R32_{cf}$	**If** $x_2 \geq 2$	**Then** class A	$CF_A = 98$
	$R33_{cf}$	**If** $x_1 \leq -0.5$	**Then** class B	$CF_B = 98$
	$R34_{cf}$	**If** $-0.4 \leq x_1 \leq 1.4$ and $-0.6 \leq x_2 \leq 1.9$	**Then** class B	$CF_B = 76$
	$R35_{cf}$	**If** $x_2 \leq -0.7$	**Then** class B	$CF_B = 100$
(4) aqx	$R41_{cf}$	**If** $x_1 \geq 1.8$	**Then** class A	$CF_A = 97$
[x-axis, $2\sqrt{2}\sigma$]	$R42_{cf}$	**If** $x_1 \leq 0.4$	**Then** class B	$CF_B = 99$
	$R43_{cf}$	**If** $0.5 \leq x_1 \leq 1.7$	**Then** class B	$CF_B = 72$
(5) brx	$R51_{cf}$	**If** $x_1 \geq 2$	**Then** class A	$CF_A = 98$
[diagonal, 4σ]	$R52_{cf}$	**If** $x_2 \geq 2.5$	**Then** class A	$CF_A = 99$
	$R53_{cf}$	**If** $x_1 \leq 0.7$	**Then** class B	$CF_B = 99$
	$R54_{cf}$	**If** $0.7 \leq x_1 \leq 1.9$ and $0.4 \leq x_2 \leq 2.4$	**Then** class B	$CF_B = 63$
	$R55_{cf}$	**If** $x_2 \leq 0.3$	**Then** class B	$CF_B = 99$
(6) arx	$R61_{cf}$	**If** $x_1 \geq 2.1$	**Then** class A	$CF_A = 99$
[x-axis, 4σ]	$R62_{cf}$	**If** $1.6 \leq x_1 \leq 2$	**Then** class B	$CF_B = 72$
	$R63_{cf}$	**If** $x_1 \leq 1.5$	**Then** class B	$CF_B = 99$

For examples, the rules that have CF value greater than 90 (high certainty to be in the given class) are rule numbers R15, R21, R24, R31, R32, R33, R35, R41, R42, R51, R52, R53, R55, R61, and R63. While the rules that have the CF value less than 60 (low certainty to be in the given class) are rule numbers R14 and R22. The larger CF values are, the smaller ambiguous region between the data set will be. Thus, the CF value can

help user gain the knowledge of how certain the answer should be in a given class. The CF rule base with parameter setting $\psi = 1\%$ and $\omega = 40\%$ from all six data sets is shown in Table 3.3.

Table 3.3 The CF rules of generated ambiguous data set with parameter setting $\psi = 1\%$ and $\omega = 40\%$.

Data Set Name	Rule No	CF Rules		
(1) bpx	$R11_{cf}$	**If** $x_1 \geq 3.1$	**Then** class A	CF_A=100
[diagonal, 2σ]	$R12_{cf}$	**If** $x_2 \geq 3.2$	**Then** class A	CF_A=100
	$R13_{cf}$	**If** $-1.5 \leq x_1 \leq 3$ and $-1.4 \leq x_2 \leq 3.1$	**Then** class A	$CF_B = 53$
	$R14_{cf}$	**If** $x_1 \leq -1.6$	**Then** class B	CF_B=100
	$R15_{cf}$	**If** $x_2 \leq -1.5$	**Then** class B	CF_B=100
(2) apx	$R21_{cf}$	**If** $x_1 \geq 2.4$	**Then** class A	$CF_A = 99$
[x-axis, 2σ]	$R22_{cf}$	**If** $x_1 \leq -0.7$	**Then** class B	CF_B=100
	$R23_{cf}$	**If** $-0.6 \leq x_1 \leq 2.3$	**Then** class B	$CF_B = 53$
(3) bqx	$R31_{cf}$	**If** $x_1 \geq 2.3$	**Then** class A	$CF_A = 99$
[diagonal, $2\sqrt{2}\sigma$]	$R32_{cf}$	**If** $x_2 \geq 2.5$	**Then** class A	$CF_A = 99$
	$R33_{cf}$	**If** $x_1 \leq -0.7$	**Then** class B	CF_B=100
	$R34_{cf}$	**If** $-0.6 \leq x_1 \leq 2.2$ and $- \leq x_2 \leq 2.4$	**Then** class B	$CF_B = 64$
	$R35_{cf}$	**If** $x_2 \leq -1.1$	**Then** class B	CF_B=100
(4) aqx	$R41_{cf}$	**If** $x_1 \geq 2.2$	**Then** class A	$CF_A = 99$
[x-axis, $2\sqrt{2}\sigma$]	$R42_{cf}$	**If** $x_1 \leq -0.1$	**Then** class B	CF_B=100
	$R43_{cf}$	**If** $0 \leq x_1 \leq 2.1$	**Then** class B	$CF_B = 67$
(5) brx	$R51_{cf}$	**If** $x_1 \geq 2.3$	**Then** class A	$CF_A = 99$
[diagonal, 4σ]	$R52_{cf}$	**If** $x_2 \geq 3$	**Then** class A	$CF_A = 99$
	$R53_{cf}$	**If** $x_1 \leq 0$	**Then** class B	CF_B=100
	$R54_{cf}$	**If** $0 \leq x_1 \leq 2.2$ and $0 \leq x_2 \leq 2.9$	**Then** class B	$CF_B = 64$
	$R55_{cf}$	**If** $x_2 \leq -0.1$	**Then** class B	CF_B =100
(6) arx	$R61_{cf}$	**If** $x_1 \geq 2.3$	**Then** class A	$CF_A = 99$
[x-axis, 4σ]	$R62_{cf}$	**If** $1.1 \leq x_1 \leq 2.2$	**Then** class B	$CF_B = 78$
	$R63_{cf}$	**If** $x_1 \leq 1$	**Then** class B	CF_B =100

The NL rules with parameter setting $\psi = 5\%$ and $\omega = 50\%$ are shown in Table 3.4.

Table 3.4 The NL rules of generated ambiguous data set with parameter setting $\psi = 5\%$ and $\omega = 50\%$.

Data Set Name	Rule No	NL Rules		
(1) bpx [diagonal, 2σ]	$R11_{nl}$	**If** x_1 is large	**Then** class A	$CF_A = 95$
	$R12_{nl}$	**If** x_2 is large	**Then** class A	$CF_A = 96$
	$R13_{nl}$	**If** x_1 is small	**Then** class B	$CF_B = 97$
	$R14_{nl}$	**If** x_1 is medium and x_2 is medium	**Then** class B	$CF_B = 55$
	$R15_{nl}$	**If** x_2 is small	**Then** class B	$CF_B = 97$
(2) apx [x-axis, 2σ]	$R21_{nl}$	**If** x_1 is very large	**Then** class A	$CF_A = 95$
	$R22_{nl}$	**If** x_1 is large	**Then** class A	$CF_A = 58$
	$R23_{nl}$	**If** x_1 is medium	**Then** class B	$CF_B = 86$
	$R24_{nl}$	**If** x_1 is small	**Then** class B	$CF_B = 98$
(3) bqx [diagonal, $2\sqrt{2}\sigma$]	$R31_{nl}$	**If** x_1 is large	**Then** class A	$CF_A = 91$
	$R32_{nl}$	**If** x_2 is large	**Then** class A	$CF_A = 98$
	$R33_{nl}$	**If** x_1 is small	**Then** class B	$CF_B = 98$
	$R34_{nl}$	**If** x_1 is medium and x_2 is medium	**Then** class B	$CF_B = 76$
	$R35_{nl}$	**If** x_2 is small	**Then** class B	$CF_B = 100$
(4) aqx [x-axis, $2\sqrt{2}\sigma$]	$R41_{nl}$	**If** x_1 is large	**Then** class A	$CF_A = 97$
	$R42_{nl}$	**If** x_1 is small	**Then** class B	$CF_B = 99$
	$R43_{nl}$	**If** x_1 is medium	**Then** class B	$CF_B = 72$
(5) brx [diagonal, 4σ]	$R51_{nl}$	**If** x_1 is large	**Then** class A	$CF_A = 98$
	$R52_{nl}$	**If** x_2 is large	**Then** class A	$CF_A = 99$
	$R53_{nl}$	**If** x_1 is small	**Then** class B	$CF_B = 99$
	$R54_{nl}$	**If** x_1 is medium and x_2 is medium	**Then** class B	$CF_B = 63$
	$R55_{nl}$	**If** x_2 is small	**Then** class B	$CF_B = 99$
(6) arx [x-axis, 4σ]	$R61_{nl}$	**If** x_1 is large	**Then** class A	$CF_A = 99$
	$R62_{nl}$	**If** x_1 is medium	**Then** class B	$CF_B = 72$
	$R63_{nl}$	**If** x_1 is small	**Then** class B	$CF_B = 99$

For the diagonal orientation data sets ((1) bpx, (3) bqx, and (5) brx), there are three intervals to represent the input features x_1 and x_2 with the NL terms as "small", "medium", and "large". For data set (1) bpx, the ranges of NL terms for x_1 are [-10,-0.9], [-1,1.9], and [2,10] and for x_2 are [-10,-0.9], [-1,1.9], and [2,10], respectively.

For data set (3) bqx, the ranges of NL terms for x_1 are [-10,-0.3], [-0.4,1.4], and [1.5,10] and for x_2 are [-10,-0.5], [-0.6,1.9], and [2,10], respectively. For data set (5) brx, the ranges of NL terms for x_1 are [-10,0.6], [0.7,1.9], and [2, 10] and for x_2 are [-10,0.3], [0.4,2.4], and [2.5,10], respectively. On the other hand for the x-axis orientation data sets, since we use only x_1 to represent the NL rules while the data set (2) apx uses four intervals as "small", "medium", "large", and "very large" where the ranges of NL terms for x_1 are [-10,-0.3], [-0.4,0.5], [0.6,1.6], and [1.7,10], respectively. The data set (4) aqx and (6) arx use three intervals as "small", "medium", and "large". For data set (4) aqx, the ranges of NL terms for x_1 are [-10,0.4], [0.5,1.7], and [1.8,10] while for data set (6) arx, the ranges of NL terms for x_1 are [-10,1.5], [1.6,2], and [2.1,10], respectively.

3.2 Glass Data Set

The data set concerns the classification of mixture of different minerals for making various glass products e.g. building window, vehicle window, container, tableware, or headlamp. The glass data set has 214 entities with six classes [27]. The six class types are as follows:

class 1: "*building window with float processed*",
class 2: "*building window with non float processed*",
class 3: "*vehicle window with float processed*",
class 4: "*container*",
class 5: "*tableware*", and
class 6: "*headlamp*".

The data consist of nine input features of real values which are:

RI:	*Refractive index* (x_1) with range	[1.151115,1.153393],
Na:	*Sodium* (x_2) with range	[10.73,17.38],
Mg:	*Magnesium* (x_3) with range	[0.33,4.49],

Al:	*Aluminum* (x_4) with range	[0.29,3.5],
Si:	*Silicon* (x_5) with range	[69.81,75.41],
K:	*Potassium* (x_6) with range	[0.02,6.21],
Ca:	*Calcium* (x_7) with range	[5.43,16.19],
Ba:	*Barium* (x_8) with range	[0.06,3.15], and
Fe:	*Iron* (x_9) with range	[0.01,0.51].

In the first step of rule extraction process, the neural network was trained by SNNS software. The data set is divided into two sets: training set and test set. The training set contained 107 input vectors selected randomly while the test set contained the other 107 input vectors. The network structure has nine input nodes, four hidden nodes, and one output node (9:4:1). For the training process, each range is normalized between [0,1]. The target of all data in the considered class is set to 1 while the target of all data not in the considered class is set to 0. There are no missing data. There are six network structures to be trained: class 1 to class 6. In the REAP algorithm, the parameter setting for the maximum number of accepted ambiguous activation values ψ is 5%. The threshold value for the feature extraction ω is 50%. In the RECF algorithm, the threshold value for the accepting the certainty factor value ξ is 50%.

From the experimental result, the network structure has *99%* classification accuracy for SNNS from the training data sets of all six classes. The accuracy results from all data of class 1 to class 6 for REAP and RECF algorithm are *78%, 76%, 83%, 95%, 92%,* and *94%,* respectively. The average accuracy for all classes is *86%*. The accuracy results from all data of class 1 to class 6 for RENL algorithm are *77%, 75%, 82%, 94%, 92%,* and *94%,* respectively. The average accuracy for all classes is *86%* as shown in Table 3.5.

Table 3.5 The experimental results of glass data set .

Glass type	SNNS	REAP & RECF Algorithm	RENL Algorithm
Class 1	99%	78%	77%
Class 2	99%	76%	75%
Class 3	100%	83%	82%
Class 4	100%	95%	94%
Class 5	100%	92%	92%
Class 6	100%	94%	94%

The classification rates of some classes, e.g. class 1 and class 2, in the experiment are rather low compared to the results from the neural network. It is due to an obvious reason. The data of all classes are highly overlapped in some dimensions. This distribution is not good for interval rule extraction which each separating hyperplane must be orthogonal to its corresponding principal dimensional axis. In this situation, the neural network can give a better result since each separating hyperplane is not necessary orthogonal to the principal dimensional axis. However, the accuracy of interval rule extraction can be increased by adding more rules to further refine the classification of some highly overlapped data space. But this approach can create an "overfit" situation which reduces the generalization of the rules. The example of ambiguity glass data set for class 1 of Mg (x_3) and Al (x_4) is shown in Figure 3.3. Note that Mg (x_3) and Al (x_4) are not eliminated in the feature extraction when we set the threshold value for the feature extraction ω to 50%.

There are six NL rules. For class 1 to class 6 as shown in Figure 3.4. Four NL terms as "small", "medium", "large", and "very large" are used in this experiment. The

NL rules provide the concept of how input features x_2, x_3, x_4, x_6, x_7, x_8 and x_9 effect the result of the rules defining class 1 to class 6. Note that x_1 and x_5 can be eliminated. For example, in $R6_{nl}$, if Na is very large and Mg is small or medium, then the glass type is headlamp (class 6) with 94% certainty. The ranges for x_2 to x_9 are shown as follows:

Na (x_2) range: [10.73,12.53], [12.54,13.39], [13.40,14.06], [14.07,17.38],
Mg (x_3) range: [0.33,2.58], [2.59,3.45], [3.46,3.99], [4.00,4.49],
Al (x_4) range: [0.29,1.09], [1.10,1.41], [1.42,2.47], [2.48,3.50],
K (x_6) range: [0.02,0.64], [0.65,3.12], [3.13,4.66], [4.67,6.21],
Ca (x_7) range: [5.43,6.83], [6.84,8.43], [8.44,9.2], [9.21,16.19],
Ba (x_8) range: [0.06,0.37], [0.38,1.61], [1.62,2.38], [2.39,3.15], and
Fe (x_9) range: [0.01,0.06], [0.07,0.26], [0.27,0.39], [0.40,0.51].

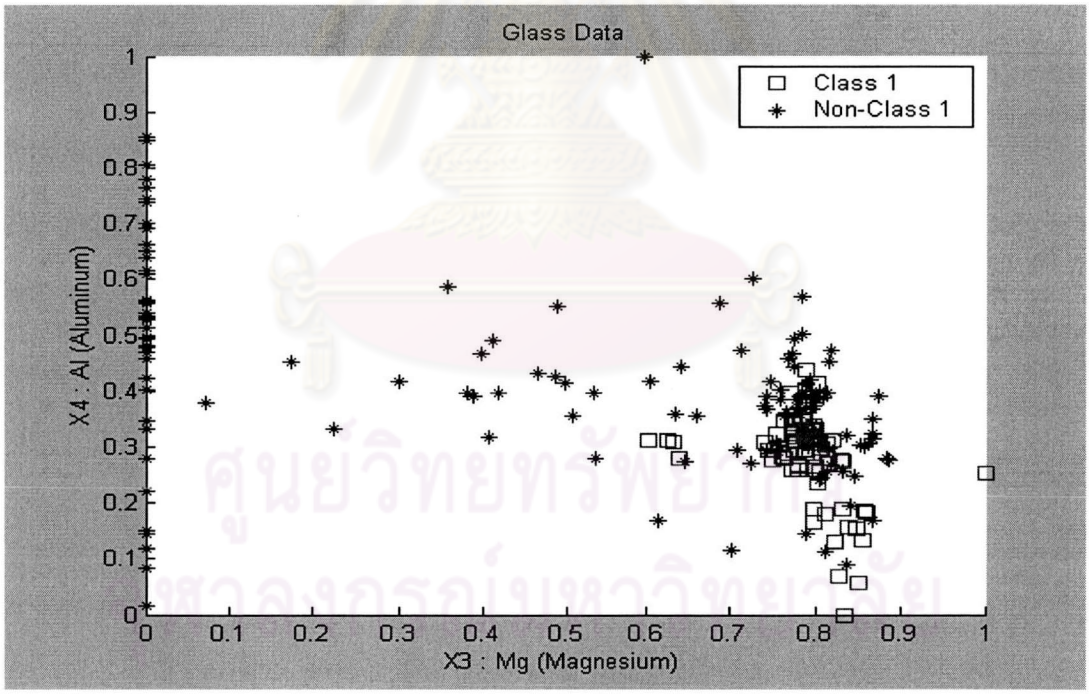


Figure 3.3 The ambiguity glass data set.

$R1_{cf}$:	**If**	(0.28 ≤ Na ≤ 0.4) and (0.73 ≤ Mg ≤ 0.81) and (0.25 ≤ Al ≤ 0.35)	
	Then	Building window with float processed	CF_{bwf} = 77
$R2_{cf}$:	**If**	(0.14 ≤ Ca ≤ 0.27)	
	Then	Building window with non float processed	CF_{bwnf} = 75
$R3_{cf}$:	**If**	(0.75 ≤ Mg ≤ 0.88) and (0.08 ≤ K ≤ 0.1) and (0.28 ≤ Ca ≤ 0.35)	
	Then	Vehicles with float processed	CF_{vf} = 82
$R4_{cf}$:	**If**	(0 ≤ Mg ≤ 0.42) and (0.42 ≤ Al ≤ 0.68) and (Ba = 0) and (Fe = 0)	
	Then	Container	CF_{ct} = 94
$R5_{cf}$:	**If**	(0 ≤ Mg ≤ 0.54) and (Ba = 0) and (Fe = 0)	
	Then	Tableware	CF_{tw} = 92
$R6_{cf}$:	If	(0.5 ≤ Na ≤ 1) and (0 ≤ Mg ≤ 0.72)	
	Then	Headlamp	CF_{hl} = 94

(a)

$R1_{nl}$:	**If**	(Na is medium) and (Mg is large) and (Al is medium)	
	Then	Building window with float processed	CF_{bwfp} = 77
$R1_{nl}$:	**If**	(Ca is medium)	
	Then	Building window with non float processed	CF_{bwnfp} = 75
$R1_{nl}$:	**If**	(Mg is large) and (K is small) and (Ca is large)	
	Then	Vehicles with float processed	CF_{vfp} = 82
$R1_{nl}$:	**If**	(Mg is small) and (Al is large) and (Ba is small) and (Fe is small)	
	Then	Container	CF_{ctn} = 94
$R1_{nl}$:	**If**	(Mg is small) and (Ba is small) and (Fe is small)	
	Then	Tableware	CF_{tbw} = 92
$R1_{nl}$:	**If**	(Na is very large) and (Mg is less than or equal to medium)	
	Then	Headlamp	CF_{hl} = 94

(b)

Figure 3.4 The glass rules. (a) CF rules. (b) NL rules.

3.3 Iris Data Set

The well-known Iris data set has been used in the rule extraction process. The Iris data set consists of three classes of flowers with 50 patterns each. The Setosa class is linearly separable while the Versicolor and Virginica classes are not. The data consist of

real values and have a feature dimensionality of four. The four features are sepal length (x_1), sepal width (x_2), petal length (x_3), and petal width (x_4) [28].

In the first step of rule extraction process, the neural network was trained by SNNS software. The data set is divided into two sets: training set and test set. The training set contains 120 input vectors (40 input vectors from each class) selected randomly and the test set contains 30 input vectors (10 input vectors from each class). The network structure has four input units, four hidden nodes, and one output node (4:4:1). There are three network structures to be trained (the Setosa network structure, the Versicolor network structure, and the Virginica network structure). In the training process, the Setosa network structure will set the data in the class of Setosa to have the output value equals to 1 while the data not in the class of Setosa will have the output value equals to 0. This training process is also applied to other two network structures. In the REAP algorithm, the parameter setting for the maximum number of accepted ambiguous activation value ψ is 5%. The threshold value for the feature extraction ω is 50%. In the RECF algorithm, the threshold value for the accepting the certainty factor value ξ is 50%.

The experimental results with the Iris data set using the REAP, RECF, and RENL algorithm are shown in Table 3.6. The results of Setosa network structure show 100% accuracy on all of the data set for all SNNS, REAP, RECF, and RENL algorithms. The Versicolor network structure shows 100% accuracy for SNNS, 94% accuracy for all REAP algorithm, RECF algorithm, and RENL algorithm on all of the data set with 3 data out of 50 are classified into the wrong group. The Virginica network structure shows 99% for SNNS and 88% accuracy for REAP algorithm on all of the data set with 18 data out of 150 are classified into the wrong group, and 100% accuracy for RECF

algorithm and RENL algorithm on all of the data set. The averages of classification accuracy on all of the data set with three network structures are 99.7% for SNNS, 94% for REAP, and 98% for and RECF algorithm and RENL algorithm.

Table 3.6 The experimental results of Iris data set.

Iris Class	SNNS	REAP Algorithm	RECF & RENL Algorithm
Setosa	100%	100%	100%
Versicolor	100%	94.0%	94.0%
Virginica	99.0%	88.0%	100%
Overall Performance	99.7%	94.0%	98.0%

The crisp rule from REAP algorithm, the CF rule base from RECF algorithm, and the NL rule base from RENL algorithm are shown in Figure 3.5(a), 3.5(b), and 3.5(c), respectively. From REAP algorithm with feature extraction, we can eliminate sepal length (x_1) and sepal width (x_2) and the rule base results use only petal length (x_3) and petal width (x_4). There are four rules for REAP rule and seven rules for both the RECF rule and the NL rule. The natural language terms in NL rule are "small", "medium", "large", and "very large". The ranges of natural language terms for x_3 are [1,2.9], [3,4.9], [5,5.1], and [5.2,6.9] and for x_4 are [0.1,0.9], [1,1.6], [1.7,1.8], and [1.9,2.5]. All of the conditional parts of the rules are projected on the input vector values, so we can understand the relationship between the conditions of the rules and the given class. For examples, in Figure 3.5(a), the crisp rule $R1_{cr}$ is "If ($1.0 \leq$ petal length $\leq$ 1.9) Then Setosa". In Figure 3.5(b), the CF rule $R1_{cf}$ provides the CF value equal to 100. So, we can be certain that there is only Setosa class in the given ranges. In Figure

3.5(c) the NL rule $R1_{nl}$ is "If petal length is small Then Setosa" with CF equals to 100.

$R1_{cr}$:	**If** (1.0 ≤ petal length ≤ 1.9)	**Then**	Setosa
$R2_{cr}$:	**If** (0.1 ≤ petal width ≤ 0.6)	**Then**	Setosa
$R3_{cr}$:	**If** (3.0 ≤ petal length ≤ 5.1) and (1.0 ≤ petal width ≤ 1.8)	**Then**	Versicolor
$R4_{cr}$:	**If** (4.5 ≤ petal length ≤ 6.9) and (1.4 ≤ petal width ≤ 2.5)	**Then**	Virginica

(a)

$R1_{cf}$:	**If** (1.0 ≤ petal length ≤ 2.9)	**Then**	Setosa	CF_{st} = 100
$R2_{cf}$:	**If** (0.1 ≤ petal width ≤ 0.9)	**Then**	Setosa	CF_{st} = 100
$R3_{cf}$:	**If** (3.0 ≤ petal length ≤ 4.9) and (1.0 ≤ petal width ≤ 1.6)	**Then**	Versicolor	CF_{vs} = 100
$R4_{cf}$:	**If** (5.2 ≤ petal length ≤ 6.9)	**Then**	Virginica	CF_{vg} = 100
$R5_{cf}$:	**If** (1.9 ≤ petal width ≤ 2.5)	**Then**	Virginica	CF_{vg} = 100
$R6_{cf}$:	**If** (5.0 ≤ petal length ≤ 5.1) and (1.0 ≤ petal width ≤ 1.8)	**Then**	Virginica	CF_{vg} = 67
$R7_{cf}$:	**If** (3.0 ≤ petal length ≤ 5.1) and (1.7 ≤ petal width ≤ 1.8)	**Then**	Virginica	CF_{vg} = 67

(b)

$R1_{nl}$:	**If** (petal length is small)	**Then**	Setosa	CF_{st} = 100
$R2_{nl}$:	**If** (petal width is small)	**Then**	Setosa	CF_{st} = 100
$R3_{nl}$:	**If** (petal length is medium) and (petal width is medium)	**Then**	Versicolor	CF_{vs} = 100
$R4_{nl}$:	**If** (petal length is very large)	**Then**	Virginica	CF_{vg} = 100
$R5_{nl}$:	**If** (petal width is very large)	**Then**	Virginica	CF_{vg} = 100
$R6_{nl}$:	**If** (petal length is large) and (petal width ≥ medium)	**Then**	Virginica	CF_{vg} = 67
$R7_{nl}$:	**If** (petal length ≥ medium) and (petal width is large)	**Then**	Virginica	CF_{vg} = 67

(c)

Figure 3.5 The Iris rules. (a) Crisp rules. (b) CF rules. (c) NL rules.

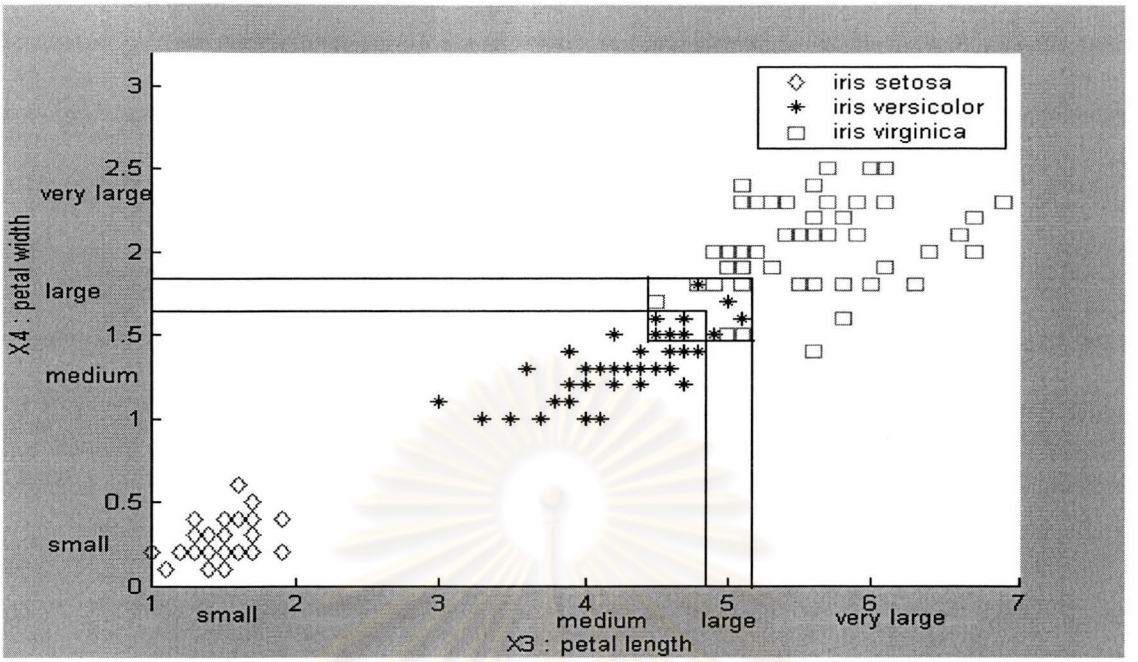
There are many rule extraction algorithms that try to extract the rules from a neural network. If the rule is in the mathematical equations then the user will have difficulty to understand the relationship between the input features. For the comprehensible rules, the rule should not be in the form of mathematical equations. The performance comparison for Iris data set problem of our approach and other rule extraction techniques such as Fuzzy Neural Network using Particle Swarm Learning

Algorithm [13], Evolving Fuzzy Neural Networks [14], Clustering Genetic Algorithm Rules [19], C4.5 [21], Modified RX Algorithm [20], and Novelty Detector Modeling with MLP [30] are shown in Table 3.7.

Table 3.7 Performance comparison for Iris data set.

Rule Extraction Algorithm	Rule in the form of Math Equations	Accuracy
RECF & RENL Algorithm (our approach)	No	98.0%
REAP Algorithm (our approach)	No	94.0%
Rule Extraction from Radial Basis function [28]	No	40.0%
Novelty Detector Modelling with MLP [30]	No	94.0%
Evolving Fuzzy Neural Networks [14]	No	95.3%
Fuzzy Neural Network using Particle Swarm Learning Algorithm [13]	No	97.0%
C4.5 [21]	No	97.3%
FULL-RE [12]	No	97.3%
Clustering Genetic Algorithm Rules [19]	Yes	97.0%
Modified RX Algorithm [20]	Yes	97.3 %

The REAP rule receives 94% accuracy. Because the crisp rule from REAP algorithm has the ambiguous input vector intervals. The ambiguous input vector intervals for x_3 is [4.5, 5.1] and for x_4 is [1.4, 1.8] as illustrated in the box area of Figure 3.6(a). The examples of the ambiguous input vector data points with 26 data points (8 data points from Versicolor class and 18 data points from Virginica class) are shown in Figure 3.6(b).



(a)

x_1	x_2	x_3	x_4	Class	x_1	x_2	x_3	x_4	Class
6.2	2.2	4.5	1.5	versicolor	6.3	2.8	5.1	1.5	virginica
6.5	2.8	4.6	1.5	versicolor	6.0	2.2	5.0	1.5	virginica
6.7	3.0	5.0	1.7	versicolor	4.9	2.5	4.5	1.7	virginica
5.9	3.2	4.8	1.8	versicolor	6.0	3.0	4.8	1.8	virginica

(b)

Figure 3.6 Examples of Iris data set. (a) The graph displays vague input vector intervals on x_3 and x_4. (b) Examples of Iris vague input vector intervals data points.

From the experimental result, the RENL rules are comprehensible with 98% accuracy. The example of the natural language terms in Rules from Evolving Fuzzy Neural Network is shown in Figure 3.7(a). He *et al.* [13] uses a particle swarm optimization algorithm to extract rules from a fuzzy neural network which has a complex network structure and receives the result at 97% accuracy using two groups of low and high. The examples of incomprehensible rules in the form of mathematical equations is

shown in Figure 3.7(b) clustering genetic algorithm rules [19] and Figure 3.7(c) modified RX algorithm [20].

R1:	**If**	high2 and low3 and low4	**Then**	Setosa
R2:	**If**	high2 and high3 and low4	**Then**	Versicolor
R3:	**If**	high2 and high3 and high4	**Then**	Virginica

(a)

R1:	**If**	$(3.6 \le a_1 \le 4.2$ and $85.8 \le a_2 \le 95.3)$	**Then**	Setosa
R2:	**If**	$(2.3 \le a_1 \le 3.5$ and $69.7 \le a_2 \le 85.2)$	**Then**	Setosa
R3:	**If**	$(3.2 \le a_1 \le 4.5$ and $82.9 \le a_2 \le 103.5)$	**Then**	Setosa
R4:	**If**	$(4.0 \le a_1 \le 5.4$ and $95.2 \le a_2 \le 113.1)$	**Then**	Setosa
R5:	**If**	$(a_1 = 5.5$ and $a_2 = 111.6)$	**Then**	Setosa
	where	$a_1 = 0.22A_s - 2.16A_p + 0.84$	$a_2 = 3.12A_s - 11.61A_p + 41.99$	
		A_s is sepal area	A_p is petal area	

(b)

R1:	**If**	$(A_s - 3.98A_p \ge 2.343)$ and $(11.21 \le A_s - 5.56A_p \le 21.87$ or $A_p - 0.18A_s = 1.47)$	**Then**	Setosa
	where	A_s is sepal area and A_p is petal area		

(c)

Figure 3.7 Examples of incomprehensible rules. (a) Rules from Evolving Fuzzy Neural Network [13]. (b) Clustering genetic algorithm rules [19].(c) Modified RX algorithm [20].

3.4 Wisconsin Breast Cancer Data Set

The Wisconsin breast cancer data set has 699 entities drawn from the classes Benign (458 samples) and Malignant (241). Each entry is a 9-dimensional vector of which each element has value between 1 and 10. The nine features are clump thickness (x_1), uniformity of cell size (x_2), uniformity of cell shape (x_3), marginal adhesion (x_4), single epithelial cell size (x_5), bare nuclei (x_6), bland chromatin (x_7), normal nucleoli (x_8), and mitoses (x_9). The missing value of bare nuclei (x_6) is set to 3 which is equal to the average number.

In the first step of rule extraction process, the neural network was trained by SNNS software. The data set is divided into two sets: training set and test set. The training set contained 350 randomly selected input vectors, and the test set contained 349 input vectors. The network structure has nine input units, nine hidden nodes, and one output node (9:9:1). In the training process of the network structure, the data in the class of Malignant is set to have the output value equals to 1 while the data in the class of Benign are set to 0. In REAP algorithm, the parameter setting for the maximum number of accepted ambiguous activation value ψ is 1%. The threshold value for the feature extraction ω is 50%. In the RECF algorithm, the threshold value for the accepting the certainty factor value ξ is 50%.

Table 3.8 The experimental results of Wisconsin breast cancer data set.

Malignant Class	SNNS	REAP & RECF Algorithm	RENL Algorithm NL terms = 4	RENL Algorithm NL terms = 3
Overall Performance	96%	94%	94%	95%

SNNS has 96% classification accuracy for the training data sets. REAP algorithm, RECF algorithm, and RENL algorithm with four NL terms achieve 94% accuracy, while RENL algorithm with three NL terms achieves 95% accuracy on all of the data set as shown in Table 3.8.

$R1_{cf}$:	**If** clump thickness is {7,9,10}	**Then**	Malignant	CF_{ma} = 99.7
$R2_{cf}$:	**If** uniformity of cell size is {5,9,10}	**Then**	Malignant	CF_{ma} = 99.7
$R3_{cf}$:	**If** uniformity of cell shape is {9,10}	**Then**	Malignant	CF_{ma} = 100
$R4_{cf}$:	**If** marginal adhesion is {7,8,10}	**Then**	Malignant	CF_{ma} = 99.7
$R5_{cf}$:	**If** single epithelial cell size is {9}	**Then**	Malignant	CF_{ma} = 100
$R6_{cf}$:	**If** bare nuclei is {6,9}	**Then**	Malignant	CF_{ma} = 100
$R7_{cf}$:	**If** bland chromatin is {6,8,9,10}	**Then**	Malignant	CF_{ma} = 99.7
$R8_{cf}$:	**If** normal nucleoli is {10}	**Then**	Malignant	CF_{ma} = 100
$R9_{cf}$:	**If** mitoses is {4,6,8,9,10}	**Then**	Malignant	CF_{ma} = 99.7

(a)

$R1_{nl}$:	**If** clump thickness is greater than or equal to large	**Then**	Malignant	CF_{ma} = 95
$R2_{nl}$:	**If** uniformity of cell size is greater than or equal to medium	**Then**	Malignant	CF_{ma} = 97
$R3_{nl}$:	**If** uniformity of cell shape is very large	**Then**	Malignant	CF_{ma} = 100
$R4_{nl}$:	**If** marginal adhesion is greater than or equal to large	**Then**	Malignant	CF_{ma} = 98
$R5_{nl}$:	**If** single epithelial cell size is very large	**Then**	Malignant	CF_{ma} = 95
$R6_{nl}$:	**If** bare nuclei is greater than or equal to medium	**Then**	Malignant	CF_{ma} = 90
$R7_{nl}$:	**If** bland chromatin is greater than or equal to medium	**Then**	Malignant	CF_{ma} = 87
$R8_{nl}$:	**If** normal nucleoli is very large	**Then**	Malignant	CF_{ma} = 97
$R9_{nl}$:	**If** mitoses is greater than or equal to medium	**Then**	Malignant	CF_{ma} = 100

(b)

$R1_{nl}$:	**If** clump thickness is greater than or equal to medium	**Then**	Malignant	CF_{ma} = 95
$R2_{nl}$:	**If** uniformity of cell size is greater than or equal to medium	**Then**	Malignant	CF_{ma} = 96
$R3_{nl}$:	**If** uniformity of cell shape is large	**Then**	Malignant	CF_{ma} = 100
$R4_{nl}$:	**If** marginal adhesion is greater than or equal to medium	**Then**	Malignant	CF_{ma} = 98
$R5_{nl}$:	**If** single epithelial cell size is greater than or equal to medium	**Then**	Malignant	CF_{ma} = 88
$R6_{nl}$:	**If** bare nuclei is greater than or equal to medium	**Then**	Malignant	CF_{ma} = 94
$R7_{nl}$:	**If** bland chromatin is greater than or equal to medium	**Then**	Malignant	CF_{ma} = 88
$R8_{nl}$:	**If** normal nucleoli is large	**Then**	Malignant	CF_{ma} = 97
$R9_{nl}$:	**If** mitoses is greater than or equal to medium	**Then**	Malignant	CF_{ma} = 100

(c)

Figure 3.8 The Wisconsin breast cancer rules. (a) CF rules.
(b) NL rules with four NL terms. (c) NL rules with three NL terms.

The CF rule has 9 rules for Malignant class as shown in Figure 3.8(a). The NL rule provides the concept of how all of the input features x_1 to x_9 effect the result of the rules to have a high risk of breast cancer (Malignant). Each rule is represented with the input features x_1 to x_9. The NL rule with four NL terms has 9 rules for Malignant class as shown in Figure 3.8(b). There are four natural language terms as "small", "medium", "large", and "very large" where the ranges are [1, 4], [5, 6], [7, 8], and [9, 10], respectively. The NL rule with three NL terms has 9 rules for Malignant class as shown in Figure 3.8(c). There are three natural language terms as "small", "medium", and "large" where the ranges are [1, 6], [7, 8], and [9, 10], respectively. For example, in rule $R3_{nl}$ of Figure 3.8(c), if the uniformity of cell shape is large, then the patient has a high risk to get the breast cancer with 100% certainty.