

บทที่ 3

ผลการทดลองและบทวิจารณ์

บทนี้จะนำเสนอผลลัพธ์ที่ได้จากการทดลองตามแบบจำลองการกรองข้อมูลสำหรับโครงข่ายประสาทเทียม (Data Filtering for Neural Network Forecasting: DFNNF) ใช้ข้อมูลอากาศจากแหล่งข้อมูลประเทศไทย (ข้อมูลจากสถานีอากาศ กรมอุตุนิยมวิทยา ประเทศไทย) ข้อมูลอากาศจากแหล่งข้อมูลต่างประเทศ (กลุ่มฐานข้อมูลอากาศเมืองริโอเดอจาเนโร และเมืองเซาเปาโล ประเทศบราซิล) ข้อมูลทางการแพทย์จากประเทศไทย (เก็บตัวอย่างมาจากโรงพยาบาลในประเทศไทย) และข้อมูลทางการแพทย์ที่เป็นมาตรฐานจาก University of California at Irvine (UCI) ผลการทดลองแสดงได้ตามลำดับการทำงานของแบบจำลอง DFNNF ดังนี้

3.1 ชุดข้อมูลอากาศประเทศไทยจังหวัดชลบุรี

ข้อมูลจัดเก็บโดยกรมอุตุนิยมวิทยาประเทศไทย จำนวน 17,520 เรคอร์ด (6 ปี) เก็บข้อมูลเป็นราย 3 ชั่วโมง คือ เวลา 01.00น. 04.00น. 07.00น. 10.00น. 13.00น. 16.00น. 19.00น. และ 22.00น. ของแต่ละวัน ประกอบด้วยข้อมูลเข้า 7 ตัวแปรคือตัวแปร x_1 ถึง x_7 และข้อมูลผลลัพธ์ 1 ตัวแปรคือตัวแปร y รายละเอียดของข้อมูลแสดงได้ดังตารางที่ 3.1

ตารางที่ 3.1 ตารางแสดงตัวแปรข้อมูลจากสถานีอากาศ กรมอุตุนิยมวิทยาประเทศไทย
ที่มา : ข้อมูลจากกรมอุตุนิยมวิทยา ประเทศไทย

ตัวแปร	ความหมาย	หน่วย
x_1	ปริมาณเมฆในท้องฟ้า	(1-10) ส่วน
x_2	อุณหภูมิจุดน้ำค้าง	องศาเซลเซียส
x_3	ความกดอากาศ	มิลลิบาร์
x_4	ความชื้นสัมพัทธ์	เปอร์เซ็นต์
x_5	อุณหภูมิจุดน้ำค้าง	องศาเซลเซียส
x_6	ความเร็วลม	นอต
x_7	ทิศทางลม	องศา
y	ปรากฏการณ์ฝนตก	มิลลิเมตร

3.1.1 การเตรียมข้อมูล

ตัวอย่างข้อมูลดิบของชุดข้อมูลประเทศไทยแสดงได้ดังตารางที่ 3.2 โดยประกอบด้วยค่าข้อมูลสูญหาย 3 ค่า คือ ค่าข้อมูลของตัวแปร x_2 : อุณหภูมิจุดน้ำค้างที่เวลา 04.00น. ค่าข้อมูลของตัว

ฝ่ายหอสมุด
ศูนย์หนังสือ อรรถกวีวิบูลย์

แปร x_1 : ความชื้นสัมพัทธ์ที่เวลา 16.00น. และค่าข้อมูลของตัวแปร x_5 : อุณหภูมิที่เวลา 04.00น. เมื่อผ่านขั้นตอนการแทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยของข้อมูลรอบข้าง นั่นคือ ค่าข้อมูลสูญหายเท่ากับ (ค่าก่อนหน้าค่าสูญหาย+ค่าหลังค่าสูญหาย/2) ตัวอย่างข้อมูลที่ผ่านการแทนค่าข้อมูลสูญหายแสดงได้ดังตารางที่ 3.3

ตารางที่ 3.2 ตารางแสดงตัวอย่างข้อมูลดิบของจังหวัดชลบุรี ประเทศไทย ราย 3 ชั่วโมง

เวลา	x_1 เมฆ	x_2 อุณหภูมิจุด น้ำค้าง	x_3 ความกด อากาศ	x_4 ความชื้น สัมพัทธ์	x_5 อุณหภูมิ	x_6 ความเร็ว ลม	x_7 ทิศทางลม	y ปริมาณ ฝน
01.00	10	24.9	1008.98	89	26.9	0	0	1
04.00	10		1008.55	96		0	0	3
07.00	3	25.8	1010.81	86	28.3	3	270	0
10.00	2	24.3	1009.91	76	28.9	3	250	0
13.00	3	23.8	1009.46	71	29.5	2	240	0
16.00	3	22.4	1009.26		30.0	0	0	0
19.00	0	22.0	1010.36	61	30.3	0	0	0
22.00	3	25.0	1011.00	78	29.2	0	0	0

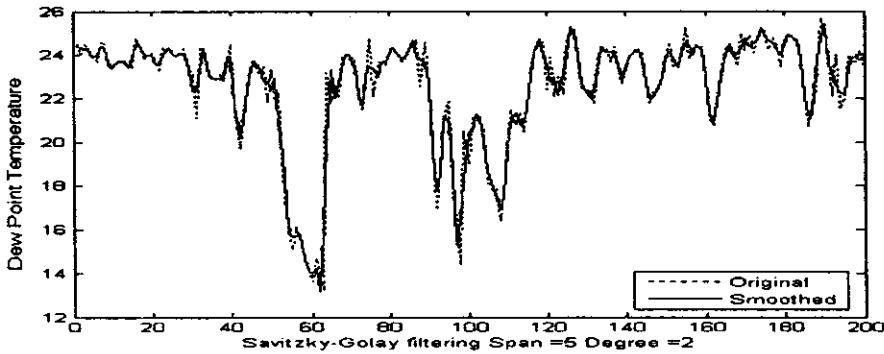
ตารางที่ 3.3 ตารางแสดงตัวอย่างข้อมูลที่ผ่านการแทนค่าข้อมูลสูญหาย

เวลา	x_1 เมฆ	x_2 อุณหภูมิจุด น้ำค้าง	x_3 ความกด อากาศ	x_4 ความชื้น สัมพัทธ์	x_5 อุณหภูมิ	x_6 ความเร็ว ลม	x_7 ทิศทาง ลม	y 0=ฝนไม่ตก 1=ฝนตก
01.00	10	24.9	1008.98	89	26.9	0	0	1
04.00	10		1008.55	96		0	0	1
07.00	3	25.8	1010.81	86	28.3	3	270	0
10.00	2	24.3	1009.91	76	28.9	3	250	0
13.00	3	23.8	1009.46	71	29.5	2	240	0
16.00	3	22.4	1009.26		30.0	0	0	0
19.00	0	22.0	1010.36	61	30.3	0	0	0
22.00	3	25.0	1011	78	29.2	0	0	0

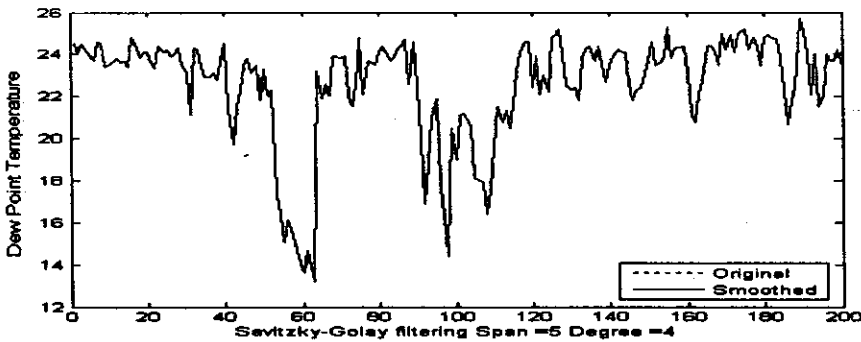
ในกรณีที่ เป็นข้อมูลอากาศจะทำนายผลลัพธ์ที่เวลาถัดไป กล่าวคือ จะกำหนดค่าข้อมูลเข้าของโครงข่ายประสาทเทียมที่เวลาปัจจุบัน (t) และกำหนดค่าข้อมูลออกของโครงข่ายประสาทเทียมที่เวลาถัดไป ($t+1$) เช่น จากตารางที่ 3.3 กำหนดให้เวลา 04.00 น. เป็นเวลาปัจจุบัน (t) ค่าข้อมูลเข้าของโครงข่ายประสาทเทียม คือ ค่าของทุกแอคทริบิวที่เวลา 04.00 น. และ ค่าข้อมูลออกของโครงข่ายประสาทเทียม คือ ค่าของแอคทริบิวผลลัพธ์ที่เวลาถัดไป 07.00 น. ($t+1$) เป็นต้น

3.1.2 การกรองข้อมูลและแบ่งช่วงข้อมูล

3.1.2.1 การกรองข้อมูล ในขั้นตอนนี้จะเป็นการทดสอบเทคนิคการกรองข้อมูลแต่ละแบบว่าเทคนิคการกรองข้อมูลใดให้ค่าความถูกต้องของการทำนายมากที่สุด ในขั้นตอนแรกจะทำกรทดสอบเทคนิคการกรองข้อมูลแบบ Savitzky-Golay Filtering เนื่องจากเทคนิคการกรองข้อมูลนี้การกำหนดพารามิเตอร์สองค่า คือ ขนาดหน้าต่าง และ ระดับขั้นพหุนาม ซึ่งต้องทดลองหาค่าระดับขั้นพหุนามที่เหมาะสม โดยมีข้อกำหนดว่าขนาดหน้าต่างต้องมีค่ามากกว่าระดับขั้นพหุนามเสมอ



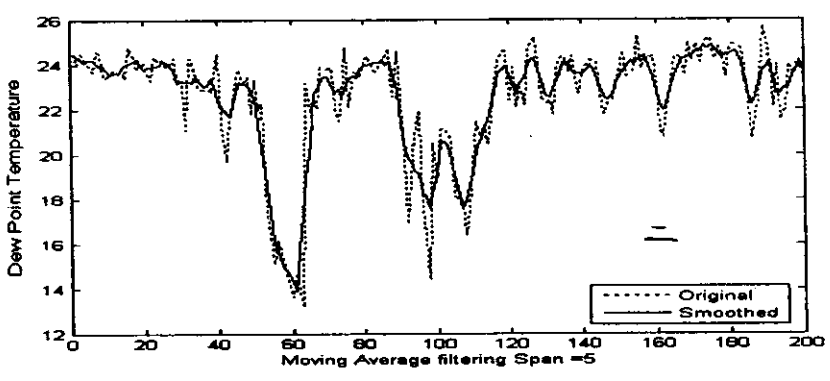
ภาพประกอบที่ 3.1 แสดงตัวอย่างการกรองข้อมูลอากาศจังหวัดชลบุรีด้วยเทคนิค Savitzky-Golay ที่ขนาดหน้าต่าง เท่ากับ 5 และระดับขั้นพหุนามเท่ากับ 2



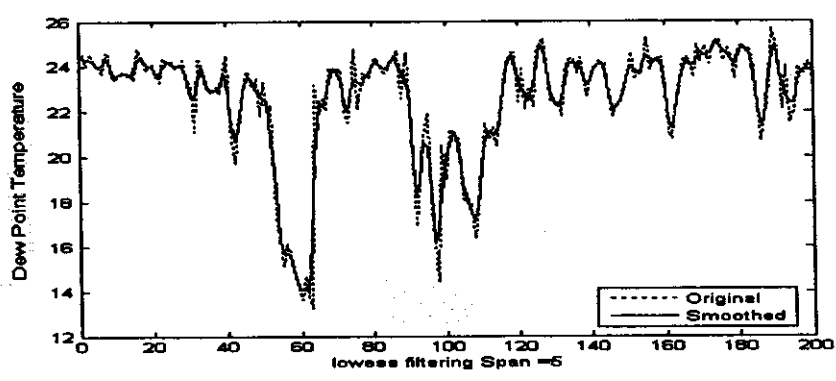
ภาพประกอบที่ 3.2 แสดงตัวอย่างการกรองข้อมูลอากาศจังหวัดชลบุรีด้วยเทคนิค Savitzky-Golay ที่ขนาดหน้าต่าง เท่ากับ 5 และระดับขั้นพหุนามเท่ากับ 4

จากภาพประกอบที่ 3.1 และ 3.2 จะพบว่าเมื่อระดับขั้นพหุนามเพิ่มขึ้นจะทำให้ค่าข้อมูลที่กรองแล้ว (Smoothed) มีความใกล้เคียงกับค่าข้อมูลเดิมมาก เนื่องจากเทคนิคการกรองข้อมูลแบบ Savitzky-Golay นั้นใช้หลักการของสมการถดถอยแบบกำลังสองน้อยที่สุด เมื่อระดับขั้นพหุนามเพิ่มขึ้นจะทำให้มีค่าสัมประสิทธิ์เพิ่มขึ้นตามจำนวนของระดับขั้นพหุนาม และทำให้ค่าข้อมูลที่กรองแล้วมีค่าใกล้เคียงกับค่าข้อมูลเดิมมากเกินไป (Over Fitting) จนเกือบจะไม่มีกรกรองข้อมูลเกิดขึ้น และใช้เวลาในการกรองนานกว่าระดับขั้นพหุนามที่มีค่าน้อย ดังนั้นในงานวิจัยนี้จึงใช้ระดับ

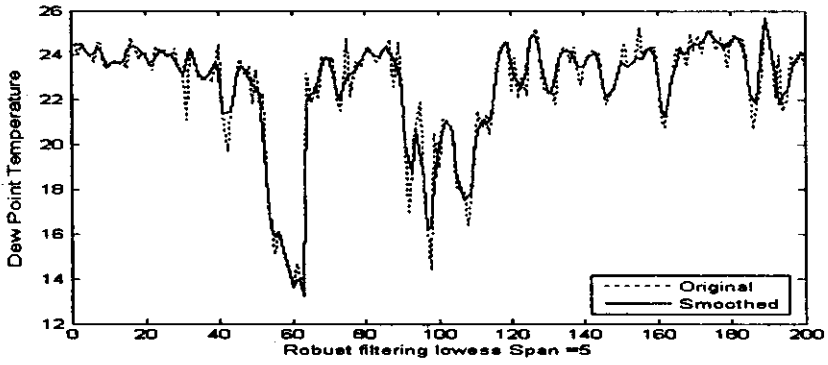
ชั้นของพหุนามเท่ากับ 2 จากนั้นจะทำการทดสอบเทคนิคการกรองข้อมูลแต่ละแบบ ภาพประกอบที่ 3.3 แสดงผลการกรองข้อมูลด้วยเทคนิคการกรองข้อมูลแบบ Moving Average Filtering ภาพประกอบที่ 3.4 แสดงผลการกรองข้อมูลด้วยเทคนิคการกรองข้อมูล Local Regression Filtering แบบ lowess ภาพประกอบที่ 3.5 แสดงผลการกรองข้อมูลด้วยเทคนิคการกรองข้อมูล Local Regression Filtering แบบ rlowess ภาพประกอบที่ 3.6 แสดงผลการกรองข้อมูลด้วยเทคนิคการกรองข้อมูล Local Regression Filtering แบบ loess ภาพประกอบที่ 3.7 แสดงผลการกรองข้อมูลด้วยเทคนิคการกรองข้อมูล Local Regression Filtering แบบ rloess ภาพประกอบที่ 3.8 แสดงผลการกรองข้อมูลด้วยเทคนิคการกรองข้อมูลแบบ Savitzky-Golay Filtering ที่ระดับชั้นพหุนามเท่ากับสอง และ ภาพประกอบที่ 3.9 แสดงผลการกรองข้อมูลด้วยเทคนิคการกรองข้อมูลแบบ Hamming Window โดยทุกเทคนิคการกรองข้อมูลจะมีขนาดหน้าต่างเท่ากับ 5



ภาพประกอบที่ 3.3 แสดงผลการกรองข้อมูลอากาศจังหวัดชลบุรีด้วยเทคนิคการกรองข้อมูลแบบ Moving Average Filtering

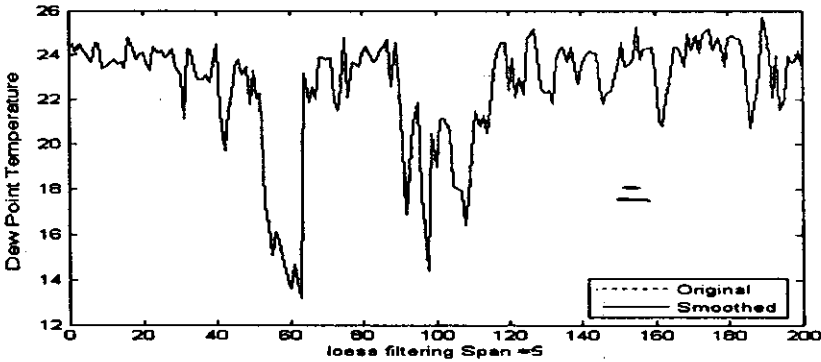


ภาพประกอบที่ 3.4 แสดงผลการกรองข้อมูลอากาศจังหวัดชลบุรีด้วยเทคนิคการกรองข้อมูล Local Regression Filtering แบบ lowess



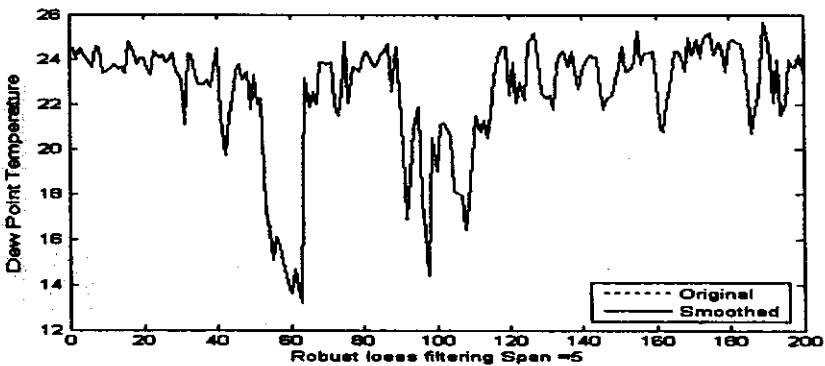
ภาพประกอบที่ 3.5 แสดงผลการกรองข้อมูลอากาศจังหวัดชลบุรีด้วยเทคนิคการกรองข้อมูล

Local Regression Filtering แบบ lowess



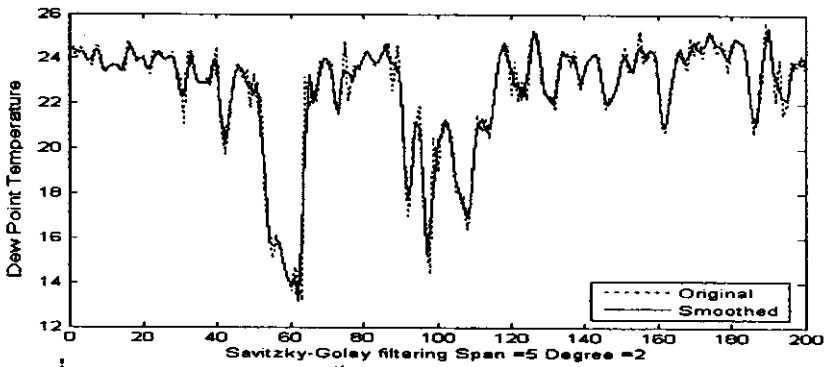
ภาพประกอบที่ 3.6 แสดงผลการกรองข้อมูลอากาศจังหวัดชลบุรีด้วยเทคนิคการกรองข้อมูล

Local Regression Filtering แบบ loess

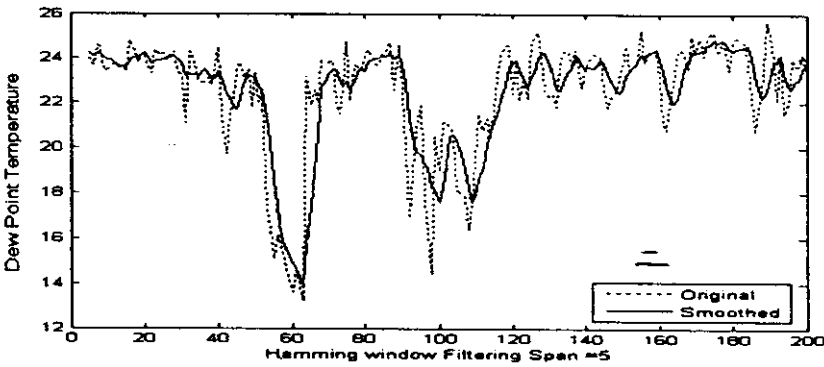


ภาพประกอบที่ 3.7 แสดงผลการกรองข้อมูลอากาศจังหวัดชลบุรีด้วยเทคนิคการกรองข้อมูล

Local Regression Filtering แบบ rloess



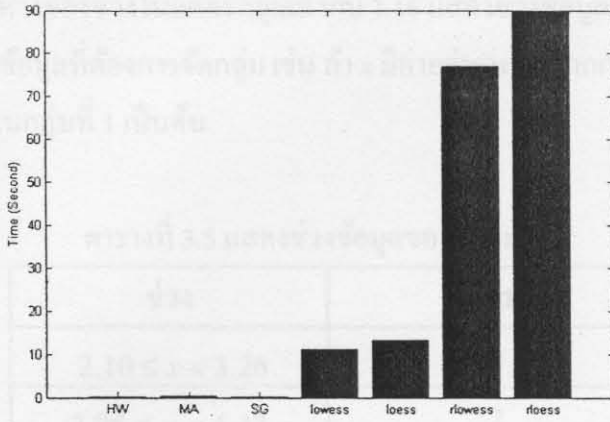
ภาพประกอบที่ 3.8 แสดงผลการกรองข้อมูลอากาศจังหวัดชลบุรีด้วยเทคนิคการกรองข้อมูลแบบ Savitzky-Golay Filtering ที่ระดับขั้นพหุนามเท่ากับสอง



ภาพประกอบที่ 3.9 แสดงผลการกรองข้อมูลอากาศจังหวัดชลบุรีด้วยเทคนิคการกรองข้อมูลแบบ Hamming Window

3.1.2.2 เปรียบเทียบเวลาในการกรองข้อมูล จากผลการทดลองการคำนวณเวลาที่ใช้ในการกรองข้อมูลของทุกเทคนิคพบว่าสามารถแบ่งกลุ่มเวลาที่ใช้ออกเป็นสามกลุ่ม เทคนิคการกรองข้อมูลที่ใช้เวลามาก คือ เทคนิคการกรองข้อมูล Local Regression Filtering แบบ rlowess และแบบ floess เทคนิคการกรองข้อมูลที่ใช้เวลาปานกลาง คือ เทคนิคการกรองข้อมูล Local Regression Filtering แบบ lowess และแบบ loess เทคนิคการกรองข้อมูลที่ใช้น้อย คือ เทคนิคการกรองข้อมูลแบบ Hamming Window Filtering เทคนิค Moving Average Filtering และ เทคนิค Savitzky-Golay แสดงได้ดังภาพประกอบที่ 3.10

ซึ่งในทุกการทดลองในงานวิจัยนี้พบว่าเวลาที่แต่ละเทคนิคการกรองข้อมูลใช้จะได้ผลลัพธ์เช่นเดียวกับภาพประกอบที่ 3.10



ภาพประกอบที่ 3.10 แสดงการเปรียบเทียบเวลาที่ใช้ในการกรองข้อมูล
ของเทคนิคการกรองข้อมูลแต่ละแบบ

3.1.2.3 การแบ่งช่วงข้อมูล เมื่อทำการกรองข้อมูลแล้ว ในกรณีที่เราต้องการแบ่งข้อมูลออกเป็นกลุ่มโดยแบ่งช่วงค่าเท่ากันเพื่อให้โครงข่ายประสาทเทียมเรียนรู้ได้ดีขึ้น ตารางที่ 3.4 แสดงตัวอย่างข้อมูลความเร็วลมของชุดข้อมูลอากาศจังหวัดชลบุรี

ตารางที่ 3.4 แสดงตัวอย่างข้อมูลความเร็วลมของชุดข้อมูลอากาศจังหวัดชลบุรี

เวลา	01.00	04.00	07.00	10.00	13.00	16.00	19.00	22.00
ความเร็วลม	6.4	6.9	3.3	2.1	4.2	6.2	7	7.9

จากตารางที่ 4.4 สามารถแบ่งกลุ่มโดยคำนวณค่าระยะห่างของช่วงในแต่ละกลุ่มได้จากสมการที่ (2.1) ในบทที่ 2 คือ

$$r = \frac{\max - \min}{q}$$

โดยที่ ตัวแปร r คือระยะห่างของช่วงในแต่ละกลุ่ม ตัวแปร q คือจำนวนกลุ่มที่ต้องการ ตัวแปร \max คือค่าข้อมูลสูงสุด และตัวแปร \min คือค่าข้อมูลต่ำสุด

ในงานวิจัยนี้จะแบ่งกลุ่มออกเป็น 5 กลุ่ม นั่นคือ q มีค่าเท่ากับ 5 \max คือค่ามากที่สุดของข้อมูลและ \min คือ ค่าน้อยสุดของข้อมูล เมื่อแทนค่าลงในสมการจะได้ ดังนี้

$$r = \frac{7.9 - 2.1}{5} = 1.16$$

นั่นคือระยะห่างของช่วงในแต่ละกลุ่มเท่ากับ 1.16 แสดงช่วงข้อมูลของแต่ละกลุ่มในตารางที่ 3.5 โดยที่ x คือ ค่าข้อมูลที่ต้องการจัดกลุ่ม เช่น ถ้า x มีค่าอยู่ระหว่างมากกว่าหรือเท่ากับ 2.10 และน้อยกว่า 3.26 จะอยู่ในกลุ่มที่ 1 เป็นต้น

ตารางที่ 3.5 แสดงช่วงข้อมูลของแต่ละกลุ่ม

ช่วง	กลุ่ม
$2.10 \leq x < 3.26$	1
$3.26 \leq x < 4.42$	2
$4.42 \leq x < 5.58$	3
$5.58 \leq x < 6.74$	4
$6.74 \leq x \leq 7.90$	5

เมื่อได้ช่วงของแต่ละกลุ่มมาแล้ว จากนั้นจะทำการกำหนดกลุ่มให้กับข้อมูลแต่ละค่า โดยพิจารณาตามช่วงค่า แสดงข้อมูลที่จัดกลุ่มแล้วตามตารางที่ 3.6

ตารางที่ 3.6 แสดงค่าข้อมูลความเร็วลมของชุดข้อมูลอากาศจังหวัดชลบุรีที่จัดกลุ่มแล้ว

เวลา	01.00	04.00	07.00	10.00	13.00	16.00	19.00	22.00
ความเร็วลม	4	5	2	1	2	4	5	5

3.1.2.4 เปลี่ยนรูปข้อมูลให้อยู่ในช่วง $[0,1]$ เหตุที่ต้องมีการแปลงข้อมูลให้อยู่ในช่วง $[0,1]$ เนื่องจากต้องเปลี่ยนให้อยู่ในรูปแบบที่โครงข่ายประสาทเทียมสามารถนำไปประมวลผลได้ เช่น ข้อมูลจากตารางที่ 4.4 แปลงข้อมูลให้อยู่ในช่วง $[0,1]$ โดยใช้สมการที่ (1.2)

$$\text{ค่าข้อมูลใหม่} = \frac{(\text{ค่าข้อมูลเดิม} - \text{ค่าต่ำสุดของช่วง})}{(\text{ค่าสูงสุดช่วง} - \text{ค่าต่ำสุดช่วง})}$$

เมื่อแทนค่าข้อมูลตัวที่ 1 จากตารางที่ 3.4 ลงในสมการ โดยค่าสูงสุดช่วงเท่ากับ 7.9 และ ค่าต่ำสุดช่วงเท่ากับ 2.1 จะได้ดังนี้

$$\text{ค่าข้อมูลใหม่} = \frac{(6.4 - 2.1)}{(7.9 - 2.1)} = 0.74$$

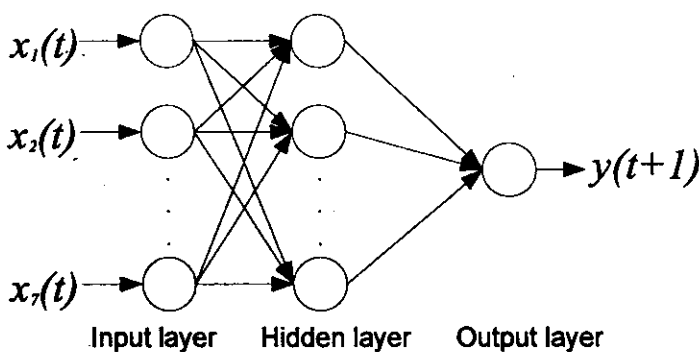
ตารางที่ 3.7 แสดงตัวอย่างค่าข้อมูลความเร็วลมที่ผ่านการเปลี่ยนรูปให้อยู่ในช่วง [0,1] แล้ว จะเห็นได้ว่าค่าน้อยสุดจะถูกแทนค่าเป็น 0 และค่ามากที่สุดจะถูกแทนค่าเป็น 1

ตารางที่ 3.7 แสดงตัวอย่างค่าข้อมูลความเร็วลมของชุดข้อมูลอากาศจังหวัดชลบุรีที่ผ่านการเปลี่ยนรูปให้อยู่ในช่วง [0, 1] แล้ว

เวลา	01.00	04.00	07.00	10.00	13.00	16.00	19.00	22.00
ความเร็วลม	0.74	0.83	0.21	0.00	0.36	0.71	0.84	1.00

3.1.3 การสอนโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมที่ใช้ในงานวิจัยชิ้นนี้ใช้โครงข่ายประสาทเทียมแบบ Multilayer Perceptron ใช้ Activation Function แบบ Sigmoid Function โดยกำหนดให้ จำนวนโหนดในชั้นข้อมูลเข้า และชั้นซ่อนมีจำนวนเท่ากับจำนวนตัวแปรเข้าของชุดข้อมูล และจำนวนโหนดในชั้นผลลัพธ์มีหนึ่งโหนด ชุดข้อมูลอากาศจังหวัดชลบุรีมีตัวแปรเข้า 7 ตัวแปร ดังนั้นสถาปัตยกรรมของโครงข่ายประสาทเทียมเป็น 7:7:1 มีจำนวนโหนดในชั้นข้อมูลเข้า และชั้นซ่อน 7 โหนด และจำนวนโหนดในชั้นผลลัพธ์ 1 โหนด แสดงได้ดังภาพประกอบที่ 3.11 จากนั้นจะทำการสอนโครงข่ายประสาทเทียมด้วยข้อมูลชุดสอนที่ผ่านการกรองข้อมูลแล้ว



ภาพประกอบที่ 3.11 สถาปัตยกรรมโครงข่ายประสาทเทียมที่ใช้สำหรับข้อมูลอากาศจังหวัดชลบุรี

3.1.4 การพยากรณ์ผลลัพธ์

ในขั้นตอนนี้จะทำการพยากรณ์ผลลัพธ์โดยใช้ข้อมูลชุดทดสอบที่ผ่านการกรองข้อมูลแล้ว ตารางที่ 3.8 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลเพียงอย่างเดียวที่ขนาดหน้าต่างแตกต่างกัน ผลการทดลองพบว่าเทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องมากที่สุด คือ เทคนิค loess ให้ค่าความถูกต้อง 95.6% ที่ขนาดหน้าต่างเท่ากับ 3 ตารางที่ 3.9 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลและช่วงข้อมูลที่ขนาดหน้าต่างแตกต่างกัน ผลการทดลองพบว่าเทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องมากที่สุด คือ เทคนิค loess และ rlowess ให้ค่าความถูกต้อง 94.5% ที่

ขนาดหน้าต่างเท่ากับ 3 จะเห็นได้ว่าการกรองข้อมูลเพียงอย่างเดียวจะให้ค่าความถูกต้องมากกว่าการกรองข้อมูลและแบ่งช่วงข้อมูล

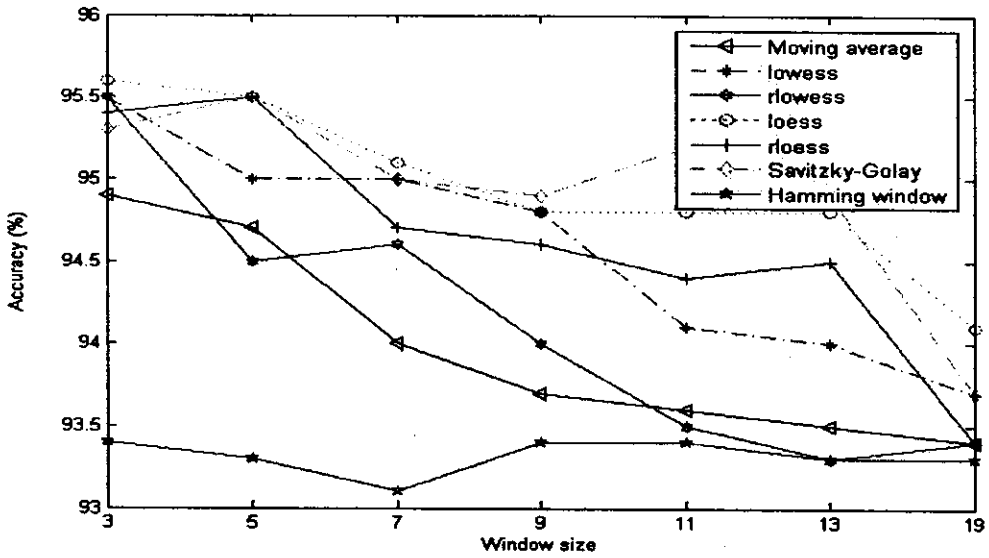
ตารางที่ 3.8 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลเพียงอย่างเดียวที่ขนาดหน้าต่างแตกต่างกันของข้อมูลอากาศจังหวัดชลบุรี

เทคนิค	ขนาดหน้าต่าง						
	3	5	7	9	11	13	19
Moving Average	94.9	94.7	94.0	93.7	93.6	93.5	93.4
lowess	95.5	95.0	95.0	94.8	94.1	94.0	93.7
rloess	95.5	94.5	94.6	94.0	93.5	93.3	93.4
loess		95.5	95.1	94.8	94.8	94.8	94.1
rloess	95.4	95.5	94.7	94.6	94.4	94.5	93.4
Savitzky-Golay	95.3	95.5	95.0	94.9	95.2	94.9	93.7
Hamming Window	93.4	93.3	93.1	93.4	93.4	93.3	93.3

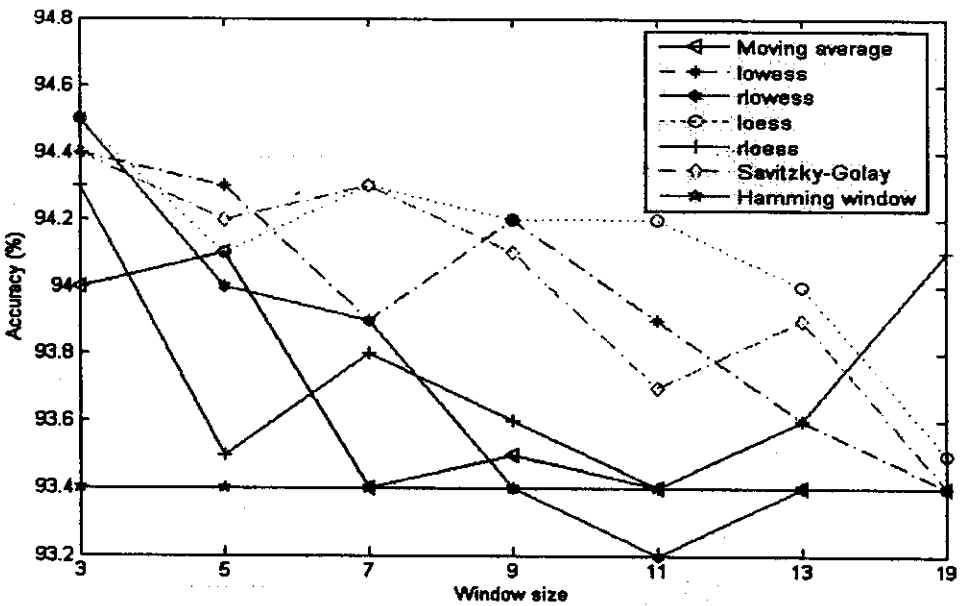
ตารางที่ 3.9 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลและแบ่งช่วงข้อมูลที่ขนาดหน้าต่างแตกต่างกันของข้อมูลอากาศจังหวัดชลบุรี

เทคนิค	ขนาดหน้าต่าง						
	3	5	7	9	11	13	19
Moving Average	94.0	94.1	93.4	93.5	93.4	93.4	93.4
lowess	94.4	94.3	93.9	94.2	93.9	93.6	93.4
rloess		94.0	93.9	93.4	93.2	93.4	93.4
loess		94.1	94.3	94.2	94.2	94.0	93.5
rloess	94.3	93.5	93.8	93.6	93.4	93.6	94.1
Savitzky-Golay	94.4	94.2	94.3	94.1	93.7	93.9	93.4
Hamming Window	93.4	93.4	93.4	93.4	93.4	93.4	93.4

ภาพประกอบที่ 3.12 แสดงแนวโน้มของค่าความถูกต้องเมื่อขนาดหน้าต่างเพิ่มขึ้นของการทดลองการกรองข้อมูลเพียงอย่างเดียว พบว่าเมื่อขนาดหน้าต่างเพิ่มขึ้นค่าความถูกต้องมีแนวโน้มลดลง ภาพประกอบที่ 3.13 แสดงแนวโน้มของค่าความถูกต้องเมื่อขนาดหน้าต่างเพิ่มขึ้นของการทดลองการกรองข้อมูลและแบ่งช่วงข้อมูล พบว่าเมื่อขนาดหน้าต่างเพิ่มขึ้นค่าความถูกต้องมีแนวโน้มลดลงเช่นเดียวกัน



ภาพประกอบที่ 3.12 แสดงแนวโน้มของค่าความถูกต้องเมื่อขนาดหน้าต่างเพิ่มขึ้นของการทดลองการกรองข้อมูลเพียงอย่างเดียวของข้อมูลอากาศจังหวัดชลบุรี



ภาพประกอบที่ 3.13 แสดงแนวโน้มของค่าความถูกต้องเมื่อขนาดหน้าต่างเพิ่มขึ้นของการทดลองการกรองข้อมูลและแบ่งช่วงข้อมูลของข้อมูลอากาศจังหวัดชลบุรี

3.2 ชุดข้อมูลพายุเมืองริโอเดอจาเนโร ประเทศบราซิล

จำนวน 1,794 เรคอร์ด ประกอบด้วยข้อมูลเข้า 10 ตัวแปรคือตัวแปร x_1 ถึง x_{10} และข้อมูลผลลัพธ์ 1 ตัวแปรคือตัวแปร y รายละเอียดของข้อมูลแสดงได้ดังตารางที่ 3.10

ตารางที่ 3.10 ตารางแสดงตัวแปรข้อมูลพายุเมืองริโอเดอจาเนโร ประเทศบราซิล

ตัวแปร	ความหมาย	หน่วย
x_1	อุณหภูมิสูงสุด	ฟาเรนไฮต์
x_2	อุณหภูมิต่ำสุด	ฟาเรนไฮต์
x_3	อุณหภูมิจุดน้ำค้างสูงสุด	ฟาเรนไฮต์
x_4	อุณหภูมิจุดน้ำค้างต่ำสุด	ฟาเรนไฮต์
x_5	ความชื้นสัมพัทธ์สูงสุด	เปอร์เซ็นต์
x_6	ความชื้นสัมพัทธ์ต่ำสุด	เปอร์เซ็นต์
x_7	ความกดอากาศสูงสุด	นิ้ว
x_8	ความกดอากาศต่ำสุด	นิ้ว
x_9	ความเร็วลม	ไมล์ต่อชั่วโมง
x_{10}	ปริมาณเมฆ	0 ถึง 10 ส่วนในท้องฟ้า
y	ปรากฏการณ์การเกิดพายุ	เกิด/ไม่เกิดพายุ

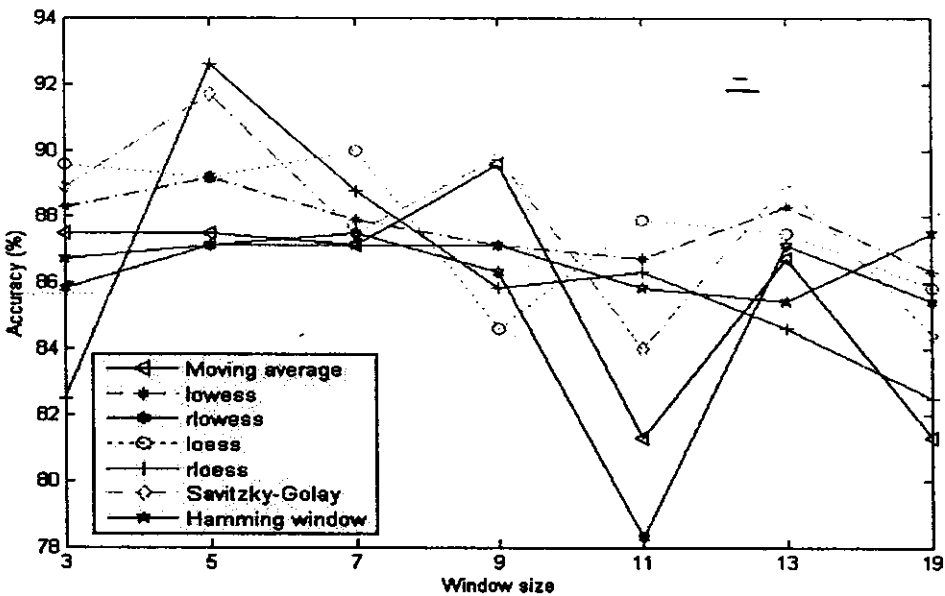
3.2.1 การเตรียมข้อมูล ข้อมูลทั้งหมด 1,794 เรคอร์ด แบ่งข้อมูลออกเป็นชุดสอน 1,457 และแบ่งเป็นชุดทดสอบ 337 เรคอร์ด แทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยข้อมูลรอบข้าง กำหนดค่าข้อมูลเข้าของโครงข่ายประสาทเทียมที่เวลาปัจจุบัน (t) และกำหนดค่าข้อมูลออกของโครงข่ายประสาทเทียมที่เวลาถัดไป ($t+1$)

3.2.2 การกรองข้อมูลและแบ่งช่วงข้อมูล ทำการกรองข้อมูลทั้งชุดสอนและชุดทดสอบโดยใช้เทคนิคการกรองทั้ง 4 แบบ จากผลการทดลองการคำนวณเวลาที่ใช้ในการกรองข้อมูลของทุกเทคนิคพบว่าสามารถแบ่งกลุ่มเวลาที่ใช้ออกเป็นสามกลุ่ม เทคนิคการกรองข้อมูลที่ใช้เวลานาน คือ เทคนิคการกรองข้อมูล Local Regression Filtering แบบ rlowess และแบบ rloess เทคนิคการกรองข้อมูลที่ใช้เวลายานกลาง คือ เทคนิคการกรองข้อมูล Local Regression Filtering แบบ lowess และแบบ loess เทคนิคการกรองข้อมูลที่ใช้เวลาน้อย คือ เทคนิคการกรองข้อมูลแบบ Hamming Window Filtering เทคนิค Moving Average Filtering และ เทคนิค Savitzky-Golay

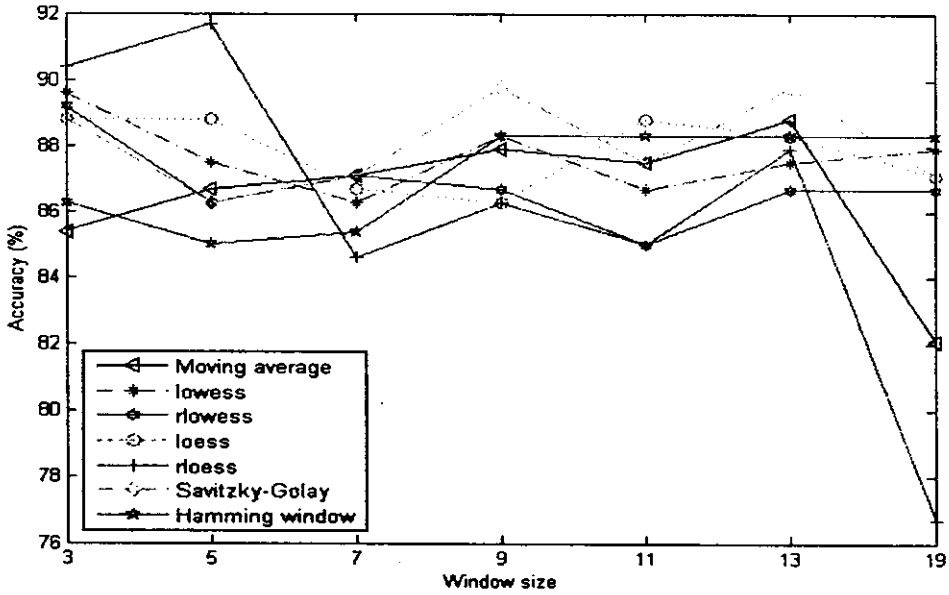
ในกรณีที่ต้องการทดลองการกรองข้อมูลและแบ่งช่วงข้อมูล จะทำการแบ่งช่วงข้อมูล และจากนั้นทำการแปลงข้อมูลให้อยู่ในช่วง $[0, 1]$

3.2.3 การสอนโครงข่ายประสาทเทียม ชุดข้อมูลพายุเมืองริโอเดอจาเนโรมีตัวแปรเข้า 10 ตัวแปร ดังนั้นสถาปัตยกรรมของโครงข่ายประสาทเทียมเป็น 10:10:1 มีจำนวนโหนดในชั้นข้อมูลเข้า และชั้นซ่อน 10 โหนด และจำนวนโหนดในชั้นผลลัพธ์ 1 โหนด จากนั้นจะทำการสอนโครงข่ายประสาทเทียมด้วยข้อมูลชุดสอนที่ผ่านการกรองข้อมูลแล้ว

3.2.4 การพยากรณ์ผลลัพธ์ ทำการพยากรณ์ผลลัพธ์โดยใช้ข้อมูลชุดทดสอบที่ผ่านการกรองข้อมูลแล้ว ภาพประกอบที่ 3.14 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลเพียงอย่างเดียวที่ขนาดหน้าต่างแตกต่างกัน ผลการทดลองพบว่าเทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องมากที่สุดคือ เทคนิค rloess ให้ค่าความถูกต้อง 92.6% ที่ขนาดหน้าต่างเท่ากับ 5 ภาพประกอบที่ 3.15 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลและช่วงข้อมูลที่ขนาดหน้าต่างแตกต่างกัน ผลการทดลองพบว่าเทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องมากที่สุดคือ เทคนิค rloess ให้ค่าความถูกต้อง 91.7% ที่ขนาดหน้าต่างเท่ากับ 5 จะเห็นได้ว่าการกรองข้อมูลเพียงอย่างเดียวจะให้ค่าความถูกต้องมากกว่าการกรองข้อมูลและแบ่งช่วงข้อมูล



ภาพประกอบที่ 3.14 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลเพียงอย่างเดียวที่ขนาดหน้าต่างแตกต่างกัน ข้อมูลพายุเมืองริโอเดอจาเนโร



ภาพประกอบที่ 3.15 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลและแบ่งช่วงข้อมูลที่ขนาดหน้าต่างแตกต่างกัน ข้อมูลพายุเมืองริโอเดอจาเนโร

3.3 ชุดข้อมูลพายุเมืองเซาเปาโล ประเทศบราซิล

จำนวน 1,420 เรคอร์ด ประกอบด้วยข้อมูลเข้า 10 ตัวแปรคือตัวแปร x_1 ถึง x_{10} และข้อมูลผลลัพธ์ 1 ตัวแปรคือตัวแปร y รายละเอียดของข้อมูลแสดงได้ดังตารางที่ 3.11

ตารางที่ 3.11 ตารางแสดงตัวแปรข้อมูลพายุเมืองเซาเปาโล ประเทศบราซิล

ตัวแปร	ความหมาย	หน่วย
x_1	อุณหภูมิสูงสุด	ฟาเรนไฮต์
x_2	อุณหภูมิต่ำสุด	ฟาเรนไฮต์
x_3	อุณหภูมิจุดน้ำค้างสูงสุด	ฟาเรนไฮต์
x_4	อุณหภูมิจุดน้ำค้างต่ำสุด	ฟาเรนไฮต์
x_5	ความชื้นสัมพัทธ์สูงสุด	เปอร์เซ็นต์
x_6	ความชื้นสัมพัทธ์ต่ำสุด	เปอร์เซ็นต์
x_7	ความกดอากาศสูงสุด	นิ้ว
x_8	ความกดอากาศต่ำสุด	นิ้ว
x_9	ความเร็วลม	ไมล์ต่อชั่วโมง
x_{10}	ปริมาณเมฆ	0 ถึง 10 ส่วนในท้องฟ้า
y	ปรากฏการณ์การเกิดพายุ	เกิด/ไม่เกิดพายุ

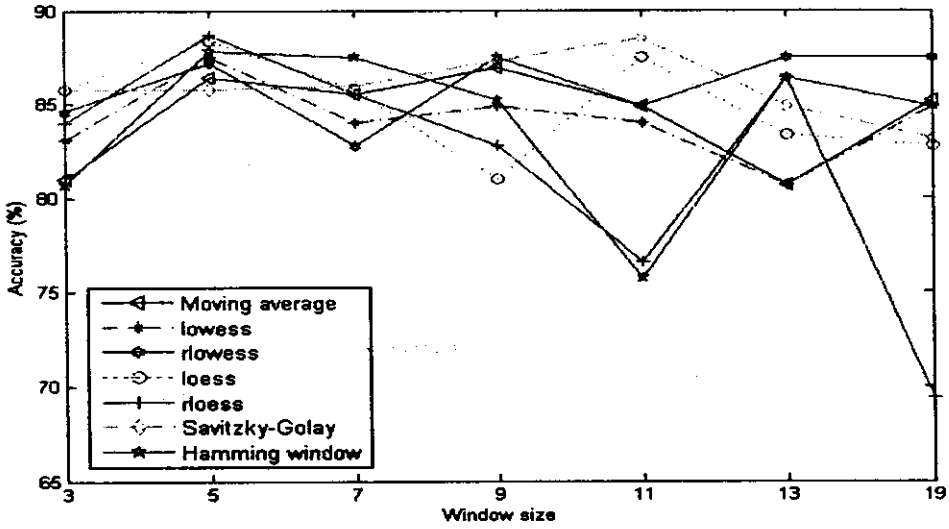
3.3.1 **การเตรียมข้อมูล** ข้อมูลทั้งหมด 1,420 เรคอร์ด แบ่งข้อมูลออกเป็นชุดสอน 1,083 และแบ่งเป็นชุดทดสอบ 337 เรคอร์ด แทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยข้อมูลรอบข้าง กำหนดค่าข้อมูลเข้าของโครงข่ายประสาทเทียมที่เวลาปัจจุบัน (t) และกำหนดค่าข้อมูลออกของโครงข่ายประสาทเทียมที่เวลาถัดไป ($t+1$)

3.2.2 **การกรองข้อมูลและแบ่งช่วงข้อมูล** ทำการกรองข้อมูลทั้งชุดสอนและชุดทดสอบ โดยใช้เทคนิคการกรองทั้ง 4 แบบ จากผลการทดลองการคำนวณเวลาที่ใช้ในการกรองข้อมูลของทุกเทคนิคพบว่าสามารถแบ่งกลุ่มเวลาที่ใช้ออกเป็นสามกลุ่ม เทคนิคการกรองข้อมูลที่ใช้เวลามากคือ เทคนิคการกรองข้อมูล Local Regression Filtering แบบ rlowess และแบบ rloess เทคนิคการกรองข้อมูลที่ใช้เวลาปานกลาง คือ เทคนิคการกรองข้อมูล Local Regression Filtering แบบ lowess และแบบ loess เทคนิคการกรองข้อมูลที่ใช้เวลาน้อย คือ เทคนิคการกรองข้อมูลแบบ Hamming Window Filtering เทคนิค Moving Average Filtering และเทคนิค Savitzky-Golay เช่นเดียวกับชุดข้อมูลก่อนหน้า

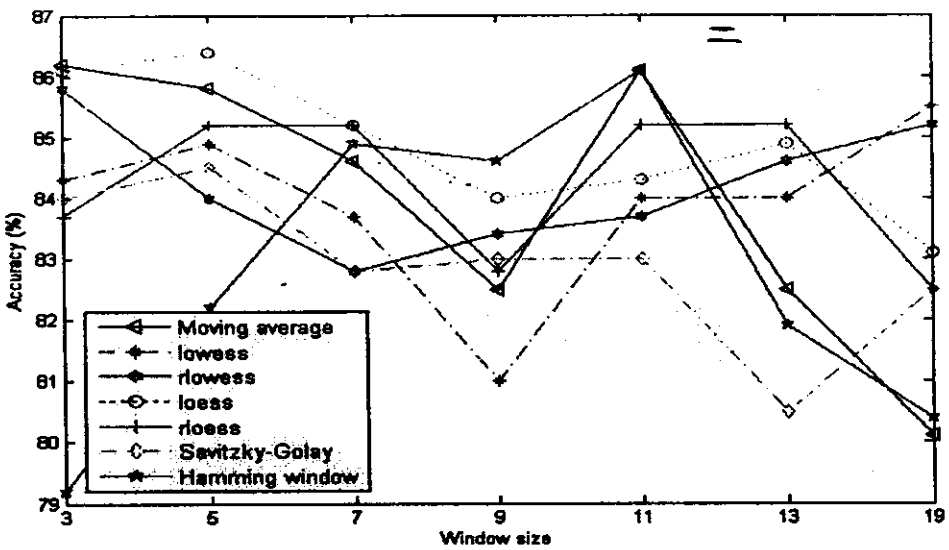
ในกรณีที่ต้องการทดลองการกรองข้อมูลและแบ่งช่วงข้อมูล จะทำการแบ่งช่วงข้อมูล และจากนั้นทำการแปลงข้อมูลให้อยู่ในช่วง $[0, 1]$

3.2.3 **การสอนโครงข่ายประสาทเทียม** ชุดข้อมูลพายุเมืองเซาเปาโลมีตัวแปรเข้า 10 ตัวแปร ดังนั้นสถาปัตยกรรมของโครงข่ายประสาทเทียมเป็น 10:10:1 มีจำนวนโหนดในชั้นข้อมูลเข้า และชั้นซ่อน 10 โหนด และจำนวนโหนดในชั้นผลลัพธ์ 1 โหนด จากนั้นจะทำการสอนโครงข่ายประสาทเทียมด้วยข้อมูลชุดสอนที่ผ่านการกรองข้อมูลแล้ว

3.2.4 **การพยากรณ์ผลลัพธ์** ทำการพยากรณ์ผลลัพธ์โดยใช้ข้อมูลชุดทดสอบที่ผ่านการกรองข้อมูลแล้ว ภาพประกอบที่ 3.16 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลเพียงอย่างเดียวที่ขนาดหน้าต่างแตกต่างกัน ผลการทดลองพบว่าเทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องมากที่สุดคือ เทคนิค rloess ให้ค่าความถูกต้อง 88.7% ที่ขนาดหน้าต่างเท่ากับ 5 ภาพประกอบที่ 3.17 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลและช่วงข้อมูลที่ขนาดหน้าต่างแตกต่างกัน ผลการทดลองพบว่าเทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องมากที่สุดคือ เทคนิค loess ให้ค่าความถูกต้อง 86.4% ที่ขนาดหน้าต่างเท่ากับ 5



ภาพประกอบที่ 3.16 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลเพียงอย่างเดียว ที่ขนาดหน้าต่างแตกต่างกัน ข้อมูลพายุเมืองเซาเปาโล



ภาพประกอบที่ 3.17 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลและแบ่งช่วงข้อมูลที่ขนาดหน้าต่างแตกต่างกัน ข้อมูลพายุเมืองเซาเปาโล

3.4 ชุดข้อมูลโรคเบาหวาน (The pima-indians-diabetes)

จำนวน 768 เรคอร์ด ประกอบด้วยข้อมูลเข้า 8 ตัวแปรคือตัวแปร x_1 ถึง x_8 และข้อมูลผลลัพธ์ 1 ตัวแปรคือตัวแปร y รายละเอียดของข้อมูลแสดงได้ดังตารางที่ 3.12

ตารางที่ 3.12 ตารางแสดงตัวแปรข้อมูล โรคเบาหวาน

ตัวแปร	ความหมาย	หน่วย
x_1	Number of pregnant times	-
x_2	Plasma glucose concentration with 2 hours in an oral glucose tolerance test	-
x_3	Diastolic blood pressure	mm Hg
x_4	Triceps of skin fold thickness	mm
x_5	2-hour serum insulin	mu U/ml
x_6	Body mass index	weight in kg/(height in m ²)
x_7	Diabetes pedigree function	-
x_8	Age	years
y	Data of diabetes	— Yes/No

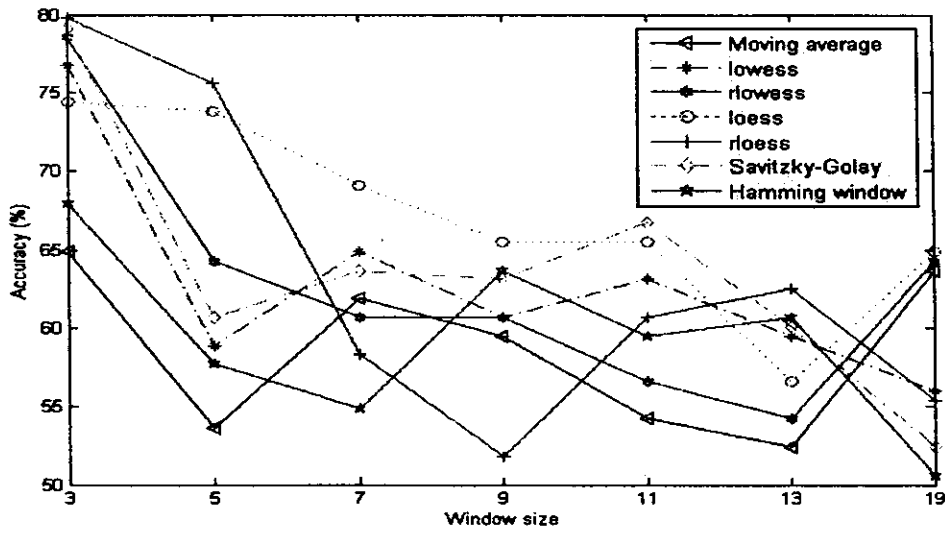
3.4.1 การเตรียมข้อมูล ข้อมูลทั้งหมด 768 เรคอร์ด แบ่งข้อมูลออกเป็นชุดสอน 600 และแบ่งเป็นชุดทดสอบ 168 เรคอร์ด แทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยข้อมูลรอบข้าง

3.4.2 การกรองข้อมูลและแบ่งช่วงข้อมูล กรองข้อมูลทั้งชุดสอนและชุดทดสอบ โดยใช้เทคนิคการกรองทั้ง 4 แบบ ในกรณีที่ต้องการทดลองการกรองข้อมูลและแบ่งช่วงข้อมูล จะทำการแบ่งช่วงข้อมูล และจากนั้นทำการแปลงข้อมูลให้อยู่ในช่วง [0, 1]

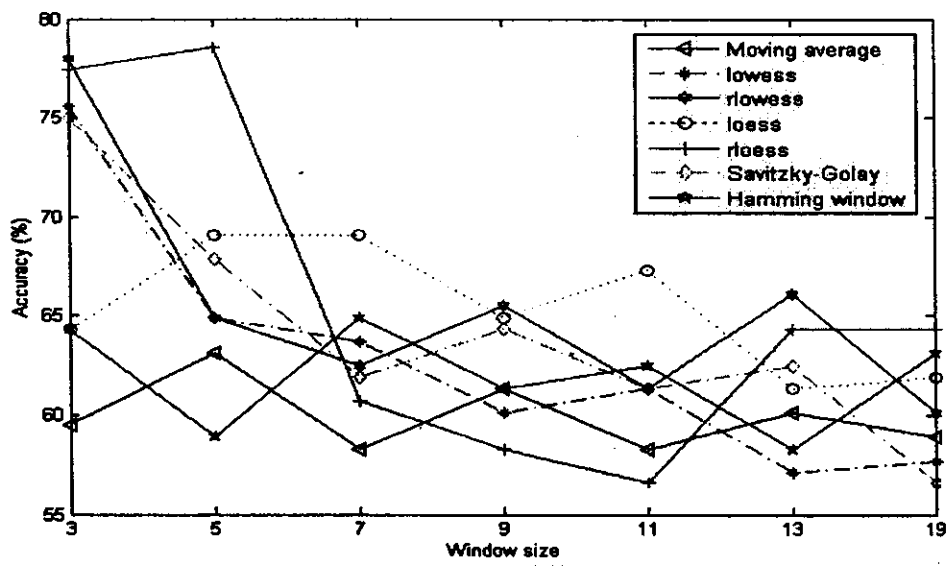
3.4.3 การสอนโครงข่ายประสาทเทียม ชุดข้อมูล โรคเบาหวานมีตัวแปรเข้า 8 ตัวแปร ดังนั้นสถาปัตยกรรมของโครงข่ายประสาทเทียมเป็น 8:8:1 มีจำนวนโหนด ในชั้นข้อมูลเข้า และชั้นซ่อน 8 โหนด และจำนวนโหนดในชั้นผลลัพธ์ 1 โหนด จากนั้นจะทำการสอนโครงข่ายประสาทเทียมด้วยข้อมูลชุดสอนที่ผ่านการกรองข้อมูลแล้ว

3.4.4 การพยากรณ์ผลลัพธ์ ทำการพยากรณ์ผลลัพธ์โดยใช้ข้อมูลชุดทดสอบที่ผ่านการกรองข้อมูลแล้ว ภาพประกอบที่ 3.18 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลเพียงอย่างเดียวที่ขนาดหน้าต่างแตกต่างกัน ผลการทดลองพบว่าเทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องมากที่สุดคือ เทคนิค floess ให้ค่าความถูกต้อง 79.8% ที่ขนาดหน้าต่างเท่ากับ 3 ภาพประกอบที่ 3.19 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลและช่วงข้อมูลที่ขนาดหน้าต่างแตกต่างกัน ผลการทดลอง

พบว่าเทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องมากที่สุด คือ เทคนิค rloess ให้ค่าความถูกต้อง 78.6% ที่ขนาดหน้าต่างเท่ากับ 5



ภาพประกอบที่ 3.18 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลเพียงอย่างเดียว ที่ขนาดหน้าต่างแตกต่างกัน ข้อมูลโรคเบาหวาน



ภาพประกอบที่ 3.19 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลและแบ่งช่วงข้อมูลที่ขนาดหน้าต่างแตกต่างกัน ข้อมูลโรคเบาหวาน

3.5 ชุดข้อมูลโรคหัวใจ จากโรงพยาบาลในประเทศไทย

เป็นข้อมูลจากโรงพยาบาล 5 แห่งในประเทศไทย ประกอบด้วย โรงพยาบาลสงขลานครินทร์ โรงพยาบาลราชวิถี โรงพยาบาลพญาไท2 และ โรงพยาบาลจุฬา ซึ่งเป็นข้อมูลจากการศึกษาภาวะที่ผิดปกติของ Sino-Atrial Node (SA Node) โดยแปลงข้อมูลจากกราฟคลื่นไฟฟ้าหัวใจ และจำแนกประเภทของความผิดปกติเป็น 4 ประเภท ได้แก่ประเภทที่ 1 ถึงประเภทที่ 4 ในการทดลองจะทำการทดลองที่ละประเภทความผิดปกติ เช่น ในการทดลองทำนายความผิดปกติประเภทที่ 1 เรคอร์ดผลลัพธ์ที่มีความผิดปกติประเภทที่ 1 จะถูกจัดให้อยู่ใน Class 1 และเรคอร์ดผลลัพธ์อื่นๆจะถูกจัดอยู่ใน Class 0 เป็นต้น ในการทดลองนี้จะทำการทดลองข้อมูลโรคหัวใจประเภทที่ 4 ประกอบด้วยข้อมูลจำนวน 870 เรคอร์ด มีข้อมูลเข้า 5 ตัวแปรคือตัวแปร x_1 ถึง x_5 และข้อมูลผลลัพธ์ 1 ตัวแปรคือตัวแปร y เป็นหรือไม่เป็นโรคหัวใจ รายละเอียดของข้อมูลแสดงได้ดังตารางที่ 3.13

ตารางที่ 3.13 ตารางแสดงตัวแปรข้อมูลโรคหัวใจ

ตัวแปร	ความหมาย
x_1	P Wave
x_2	P-R Interval
x_3	QRS Complex
x_4	Rhythm
x_5	Rate
y	โรคหัวใจ (Yes/No)

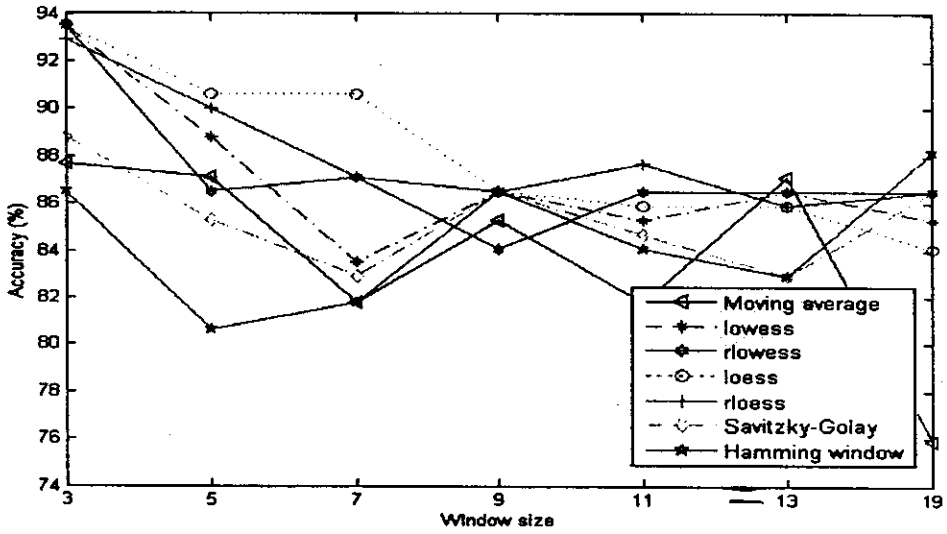
3.5.1 การเตรียมข้อมูล ข้อมูลทั้งหมด 870 เรคอร์ด แบ่งข้อมูลออกเป็นชุดสอน 600 และแบ่งเป็นชุดทดสอบ 170 เรคอร์ด แทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยข้อมูลรอบข้าง

3.5.2 การกรองข้อมูลและแบ่งช่วงข้อมูล กรองข้อมูลทั้งชุดสอนและชุดทดสอบ โดยใช้เทคนิคการกรองทั้ง 4 แบบ ในกรณีที่ต้องการทดลองการกรองข้อมูลและแบ่งช่วงข้อมูล จะทำการแบ่งช่วงข้อมูล และจากนั้นทำการแปลงข้อมูลให้อยู่ในช่วง $[0, 1]$

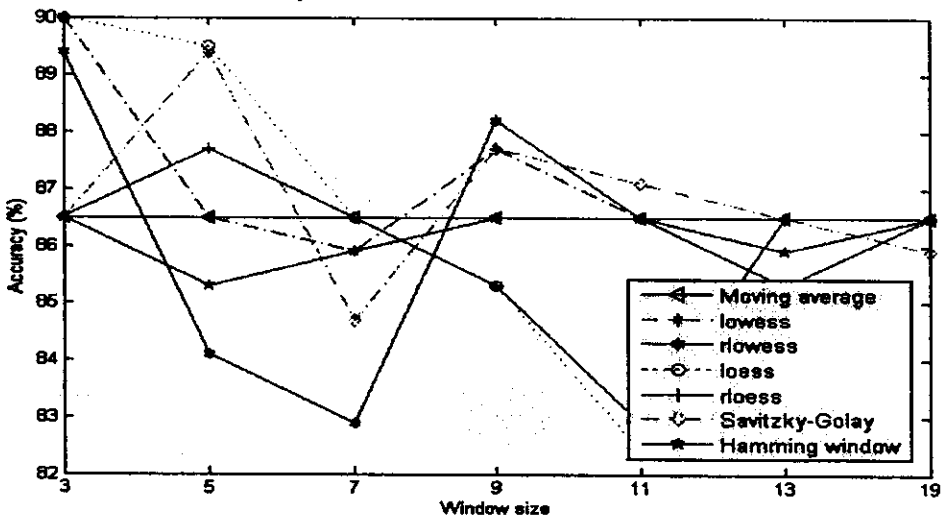
3.5.3 การสอนโครงข่ายประสาทเทียม ชุดข้อมูลโรคหัวใจมีตัวแปรเข้า 5 ตัวแปร ดังนั้นสถาปัตยกรรมของโครงข่ายประสาทเทียมเป็น 5:5:1 มีจำนวนโหนด ในชั้นข้อมูลเข้า และชั้นซ่อน 5 โหนด และจำนวนโหนดในชั้นผลลัพธ์ 1 โหนด จากนั้นจะทำการสอนโครงข่ายประสาทเทียมด้วยข้อมูลชุดสอนที่ผ่านการกรองข้อมูลแล้ว

3.5.4 การพยากรณ์ผลลัพธ์ ทำการพยากรณ์ผลลัพธ์โดยใช้ข้อมูลชุดทดสอบที่ผ่านการกรองข้อมูลแล้ว ภาพประกอบที่ 3.20 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลเพียงอย่างเดียวที่

ขนาดหน้าต่างแตกต่างกัน ผลการทดลองพบว่าเทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องมากที่สุด คือ เทคนิค rlowess, lowess และ loess ให้ค่าความถูกต้อง 93.5% ที่ขนาดหน้าต่างเท่ากับ 3 ภาพประกอบที่ 3.21 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลและช่วงข้อมูลที่ขนาดหน้าต่างแตกต่างกัน ผลการทดลองพบว่าเทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องมากที่สุด คือ เทคนิค loess และ lowess ให้ค่าความถูกต้อง 90.0% ที่ขนาดหน้าต่างเท่ากับ 3



ภาพประกอบที่ 3.20 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลเพียงอย่างเดียว ที่ขนาดหน้าต่างแตกต่างกัน ข้อมูลโรคหัวใจ



ภาพประกอบที่ 3.21 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลและแบ่งช่วงข้อมูลที่ขนาดหน้าต่างแตกต่างกัน ข้อมูลโรคหัวใจ

3.6 ชุดข้อมูลโรคมะเร็งเต้านม (Wisconsin breast cancer data set)

จำนวน 699 เรคอร์ด ประกอบด้วยข้อมูลเข้า 9 ตัวแปรคือตัวแปร x_1 ถึง x_9 และข้อมูลผลลัพธ์ 1 ตัวแปรคือตัวแปร y รายละเอียดของข้อมูลแสดงได้ดังตารางที่ 3.14

ตารางที่ 3.14 ตารางแสดงตัวแปรข้อมูล โรคมะเร็งเต้านม

ตัวแปร	ความหมาย
x_1	Clump thickness
x_2	Uniformity of cell size
x_3	Uniformity of cell shape
x_4	Marginal adhesion
x_5	Single epithelial cell size
x_6	Bare nuclei
x_7	Bland chromatin
x_8	Normal nucleoli
x_9	Mitoses
y	Data of breast cancer (Yes/No) —

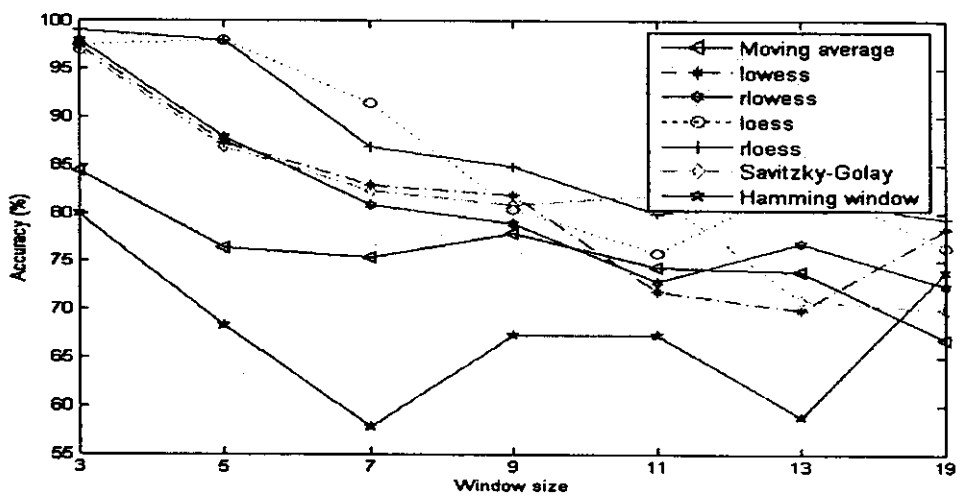
3.6.1 การเตรียมข้อมูล ข้อมูลทั้งหมด 699 เรคอร์ด แบ่งข้อมูลออกเป็นชุดสอน 500 และแบ่งเป็นชุดทดสอบ 199 เรคอร์ด แทนค่าข้อมูลสูญหายด้วยค่าเฉลี่ยข้อมูลรอบข้าง

3.6.2 การกรองข้อมูลและแบ่งช่วงข้อมูล ทำการกรองข้อมูลทั้งชุดสอนและชุดทดสอบ โดยใช้เทคนิคการกรองทั้ง 4 แบบ ในกรณีที่ต้องการทดลองการกรองข้อมูลและแบ่งช่วงข้อมูล จะทำการแบ่งช่วงข้อมูล และจากนั้นทำการแปลงข้อมูลให้อยู่ในช่วง $[0, 1]$

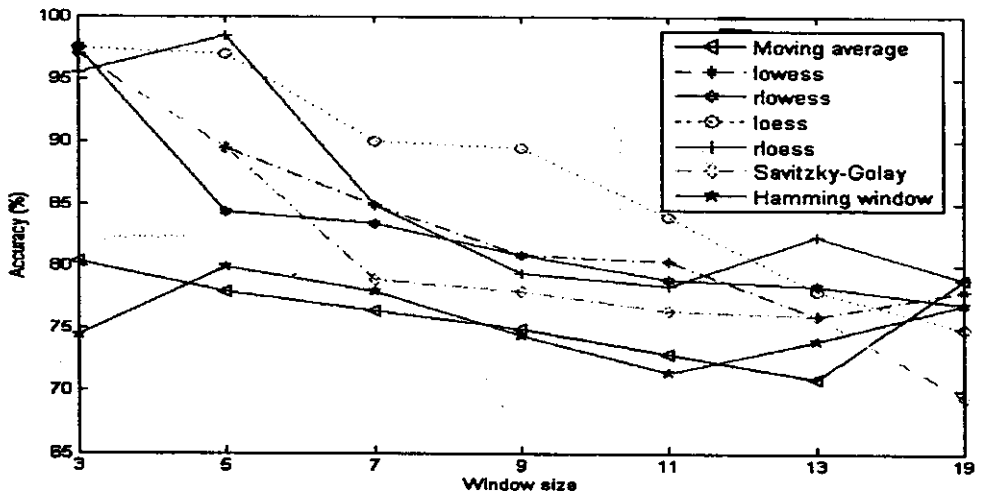
3.6.3 การสอนโครงข่ายประสาทเทียม ชุดข้อมูล โรคมะเร็งเต้านมมีตัวแปรเข้า 9 ตัวแปร ดังนั้นสถาปัตยกรรมของโครงข่ายประสาทเทียมเป็น 9:9:1 มีจำนวนโหนด ในชั้นข้อมูลเข้า และชั้นซ่อน 9 โหนด และจำนวนโหนดในชั้นผลลัพธ์ 1 โหนด จากนั้นจะทำการสอน โครงข่ายประสาทเทียมด้วยข้อมูลชุดสอนที่ผ่านการกรองข้อมูลแล้ว

3.6.4 การพยากรณ์ผลลัพธ์ ทำการพยากรณ์ผลลัพธ์ โดยใช้ข้อมูลชุดทดสอบที่ผ่านการกรองข้อมูลแล้ว ภาพประกอบที่ 3.22 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลเพียงอย่างเดียวที่ขนาดหน้าต่างแตกต่างกัน ผลการทดลองพบว่าเทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องมากที่สุดคือ เทคนิค floess ให้ค่าความถูกต้อง 99.0% ที่ขนาดหน้าต่างเท่ากับ 3 ภาพประกอบที่ 3.23 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลและช่วงข้อมูลที่ขนาดหน้าต่างแตกต่างกัน ผลการทดลอง

พบว่าเทคนิคการกรองข้อมูลที่ให้ค่าความถูกต้องมากที่สุด คือ เทคนิค rloess ให้ค่าความถูกต้อง 98.5% ที่ขนาดหน้าต่างเท่ากับ 5



ภาพประกอบที่ 3.22 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลเพียงอย่างเดียว ที่ขนาดหน้าต่างแตกต่างกัน ข้อมูล โรคมะเร็งเต้านม



ภาพประกอบที่ 3.23 แสดงค่าความถูกต้องของการทดลองกรองข้อมูลและแบ่งช่วงข้อมูลที่ขนาดหน้าต่างแตกต่างกัน ข้อมูล โรคมะเร็งเต้านม

ตารางที่ 3.15 แสดงการเปรียบเทียบค่าความถูกต้องระหว่างการกรองข้อมูลอย่างเดียว การกรองข้อมูลและแบ่งช่วงข้อมูล และ ไม่กรองข้อมูล จากชุดข้อมูล 6 ชุดข้อมูล โดยใช้เทคนิคการกรองข้อมูลแบบ rloess ที่ขนาดหน้าต่างเท่ากับ 3 (สำหรับข้อมูลอายุใช้ขนาดหน้าต่างเท่ากับ 5) พบว่าการกรองข้อมูลเพียงอย่างเดียวให้ค่าความถูกต้องมากกว่าการกรองข้อมูลและแบ่งช่วงข้อมูล และการไม่กรองข้อมูล

ตารางที่ 3.15 ตารางแสดงการเปรียบเทียบค่าความถูกต้อง โดยใช้เทคนิคการกรองข้อมูลแบบ rloess ที่ขนาดหน้าต่างเท่ากับ 3 (สำหรับข้อมูลพายุใช้ขนาดหน้าต่างเท่ากับ 5)

Data set	Filtering	Non-Filtering	Filtering and dividing
Pima Indian Diabetes	79.8	76.2	77.4
Sao Paulo	88.7	86.9	85.2
Rio de Janeiro	92.6	89.6	91.7
Heart disease type1	96.5	95.3	81.2
Heart disease type2	98.8	98.8	97.7
Heart disease type3	95.9	95.9	81.8
Heart disease type4	92.9	90.0	86.5
Chuburi Province	95.4	95.4	94.3
WBCD	99.0	99.0	95.5

=