

การเปรียบเทียบการรู้จำลายมือเขียนตัวอักษรไทยในการจำแนกกลุ่ม
ระหว่างวิธีเคเน็ยเรสเนเบอร์กับวิธีเคอร์เนล

นายรัชชัย กิติวงษ์ประทีป

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2556

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์นี้ทั้งหมดปีการศึกษา 2556 ที่เก็บไว้ในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the Graduate School.

COMPARISON OF THAI ALPHABET HANDWRITING RECOGNITION
IN CLASSIFICATION USING
K-NEAREST NEIGHBOUR METHODS AND KERNEL METHODS

Mr. Thawatchai Kitiwongprateep

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2013

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การเปรียบเทียบการรู้จำลายมือเขียนตัวอักษรไทยใน การจำแนกกลุ่มระหว่างวิธีเคเนียร์เรสเนเบอร์กับ วิธีเคอร์เนล
โดย	นายธวัชชัย กิตติวงษ์ประทีป
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	อาจารย์ ดร.อัครินทร์ ไพบุลย์พานิช

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัย
ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาโทบริหารธุรกิจ

..... คณบดีคณะพาณิชยศาสตร์และการบัญชี
(รองศาสตราจารย์ ดร.พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.เสกสรร เกียรติสุขไพบุลย์)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(อาจารย์ ดร.อัครินทร์ ไพบุลย์พานิช)

..... กรรมการ
(อาจารย์ ดร.อนันตณัฐ กัณฑ์บุญรัตน์)

..... กรรมการภายนอกมหาวิทยาลัย
(อาจารย์ ดร.บุญลืทธิ ยี่มวาสนา)

ธวัชชัย กิตติวงษ์ประทีป: การเปรียบเทียบการรู้จำลายมือเขียนตัวอักษรไทยในการจำแนกกลุ่มระหว่างวิธีเคเน็ยเรสเนเบอร์กับวิธีเคอร์เนล. (COMPARISION OF THAI ALPHABET HANDWRITING RECOGNITION IN CLASSIFICATION USING K-NEAREST NEIGHBOUR METHODS AND KERNEL METHODS) อ. ที่ปรึกษาวิทยานิพนธ์หลัก: อ. ดร. อัครินทร์ ไพบุลย์พานิช, 157 หน้า.

การวิจัยในครั้งนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบการรู้จำลายมือเขียนตัวอักษรไทยในการจำแนกกลุ่มระหว่างวิธีเคเน็ยเรสเนเบอร์กับวิธีเคอร์เนล โดยการใช้การแปลงฟูเรียร์มาพิจารณาคุณลักษณะของตัวอักษร ซึ่งทำการทดลองกับข้อมูลตัวอักษรจำนวน 2 ชุด คือ ข้อมูลลายมือเขียนตัวอักษรไทยมาจากการเขียน 1 คน และข้อมูลลายมือเขียนตัวอักษรไทยมาจากการเขียน 10 คน จำนวนสัมประสิทธิ์ฟูเรียร์ที่ใช้ในการศึกษา คือ 200, 500 และ 1,000 และข้อมูลลายมือเขียนตัวอักษรไทยที่ใช้ในการศึกษา คือ ตัวอักษรละ 90 ตัว แบ่งเป็นตัวอย่างเรียนรู้ 60 ตัวและตัวอย่างทดสอบ 30 ตัว ซึ่งศึกษากรณีตัวอย่างเรียนรู้ขนาด 10, 20, 30, 40, 50 และ 60 และใช้ค่าอัตราความผิดพลาดของการจำแนกกลุ่ม (error rate of misclassification) เป็นเกณฑ์ในการเปรียบเทียบ

จากผลการศึกษา เมื่อพิจารณาข้อมูลลายมือเขียนตัวอักษรไทยมาจากการเขียน 1 คน พบว่า วิธีเคเน็ยเรสเนเบอร์ โดยวัดระยะห่างระยะทางแบบ Euclidean กับ Cosine ที่ $k = 2$ จำนวนขนาดตัวอย่างเรียนรู้ 60 ที่ค่าสัมประสิทธิ์ฟูเรียร์ 500 ให้ประสิทธิภาพในการจำแนกกลุ่มดีกว่ากรณีอื่นๆ โดยให้ผลลัพธ์ค่าอัตราความผิดพลาดของการจำแนกกลุ่ม คือ 3.89% และเมื่อพิจารณาข้อมูลลายมือเขียนตัวอักษรไทยมาจากการเขียน 10 คน พบว่า วิธีเคเน็ยเรสเนเบอร์ โดยวัดระยะห่างระยะทางแบบ Euclidean ที่ $k = 5$ จำนวนขนาดตัวอย่างเรียนรู้ 60 ที่ค่าสัมประสิทธิ์ฟูเรียร์ 500 ให้ประสิทธิภาพในการจำแนกกลุ่มดีกว่ากรณีอื่นๆ โดยให้ผลลัพธ์ค่าอัตราความผิดพลาดของการจำแนกกลุ่ม คือ 15.28%

ภาควิชา.....สถิติ.....ลายมือเขียนนิติต

สาขาวิชา.....สถิติ.....ลายมือ ชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก

ปีการศึกษา.....2556.....

5481581326: MAJOR STATISTICS

KEYWORDS: Handwriting recognition/ K-nearest neighbour method/ Kernel method

THAWATCHAI KITIWONGPRATEEP: COMPARISON OF THAI ALPHABET
HANDWRITING RECOGNITION IN CLASSIFICATION USING K-NEAREST
NEIGHBOUR METHODS AND KERNEL METHODS. ADVISOR: AKARIN
PHAIBULPANICH, Ph.D., 157 pp.

The objective of this research is to compare recognition two nonparametric classification methods: K-nearest neighbour and Kernel with normal kernel, LSCV and LKCV when applied to Thai alphabet handwriting data, when collected from both one person and ten persons. To turn images into numerical data, fourier transform was used with fourier coefficients: 200, 500 and 1,000. Each handwritten letter were collected 90 times, from which the Test data set of 60 samples were randomly selected. The Training data set consists of the remaining samples with the size ranging from 10, 20, 30, 40, 50 and 60. The error rate of misclassification was used as the criterion of comparison.

Result of study shows that, for the Thai alphabet handwriting data from one person, K-nearest neighbour methods, using euclidean distance and cosine distance with 60 training sets at $k = 2$ and 500 coefficeints, has misclassification rate of 3.89% which is than other cases and error rate of misclassification. For the Thai alphabet handwriting data from ten persons, K-nearest neighbour methods using euclidean distance with 60 training sets at $k = 5$ and 500 coefficeints has better classification rate than other cases and error rate of misclassification is 15.28%.

Department: Statistics Student's Signature.....

Field of Study: Statistics Advisor's Signature.....

Academic Year: 2013

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สามารถสำเร็จลุล่วงไปได้ด้วยดี ด้วยความช่วยเหลือและเอาใจใส่อย่างดีของ อาจารย์ ดร. อัครินทร์ ไพบูลย์พานิช อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้วิจัยขอกราบขอบพระคุณท่านอาจารย์เป็นอย่างสูงที่กรุณาให้คำปรึกษา อบรมสั่งสอน และให้ข้อคิดเห็นต่างๆ ตลอดจนให้ความช่วยเหลือ คำแนะนำเพื่อปรับปรุงแก้ไข จนกระทั่งวิทยานิพนธ์เสร็จสมบูรณ์

ผู้วิจัยขอกราบขอบพระคุณรองศาสตราจารย์ ดร. เสกสรรค์ เกียรติสุขไพบูลย์ ประธานกรรมการสอบวิทยานิพนธ์ อาจารย์ ดร. อนันตฉัตร กันต์ธัญญรัตน์ และอาจารย์ ดร. บุญสิทธิ์ ยิ้มวาสนา กรรมการสอบวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ ตรวจสอบ และแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น

สุดท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณครอบครัวที่ช่วยส่งเสริม สนับสนุนและให้กำลังใจเสมอมาจนสำเร็จการศึกษา รวมทั้งขอขอบคุณเพื่อนๆ ทุกคนที่คอยให้กำลังใจผู้วิจัยมาโดยตลอด

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ณ
สารบัญภาพ.....	ด
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ในการศึกษา.....	3
1.3 สมมติฐานของการศึกษา.....	4
1.4 คำจำกัดความที่ใช้ในการศึกษา.....	4
1.5 ขอบเขตของการศึกษา.....	4
1.6 วิธีดำเนินการศึกษา.....	6
1.7 ประโยชน์ที่คาดว่าจะได้รับ.....	7
บทที่ 2 ทฤษฎีและตัวสถิติที่เกี่ยวข้อง.....	8
2.1 ทฤษฎีพื้นฐาน.....	8
2.2 วิธีการจำแนกที่ใช้ในการศึกษา.....	17
2.2.1 วิธี Kernel.....	18
2.2.2 วิธี K-nearest neighbour.....	21
2.3 เกณฑ์ที่ใช้ในการพิจารณาประสิทธิภาพในการจำแนก.....	22
บทที่ 3 วิธีดำเนินการศึกษา.....	24
3.1 ขอบเขตของการศึกษา.....	24
3.2 แหล่งที่มาของข้อมูล.....	25
3.3 ขั้นตอนวิธีการรู้จำ.....	26

	หน้า
3.4 ขั้นตอนในการดำเนินการศึกษา.....	28
3.5 ขั้นตอนการทำงานของโปรแกรม.....	29
บทที่ 4 ผลการวิเคราะห์ข้อมูล.....	31
4.1 ผลการศึกษาจากลายมือเขียนตัวอักษรไทย 1 คน.....	32
4.2 ผลการศึกษาจากลายมือเขียนตัวอักษรไทย 10 คน.....	84
บทที่ 5 สรุปผลการศึกษาและข้อเสนอแนะ.....	136
5.1 สรุปผลการศึกษา.....	137
5.2 ด้านการศึกษาวิจัย.....	147
5.3 ข้อเสนอแนะ.....	147
รายการอ้างอิง.....	148
บรรณานุกรม.....	149
ภาคผนวก.....	150
ประวัติผู้เขียนวิทยานิพนธ์.....	157

สารบัญตาราง

ตารางที่	หน้า
3.1 แสดงตัวอย่างค่าสัมประสิทธิ์ 10 ค่าที่ได้จากการแปลงฟูเรียร์ของตัวอักษร “ข”.....	27
4.1 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดย วัดระยะห่างระยะทางแบบ Euclidean.....	33
4.2 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดย วัดระยะห่างระยะทางแบบ Euclidean.....	34
4.3 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดย วัดระยะห่างระยะทางแบบ Euclidean.....	35
4.4 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดย วัดระยะห่างระยะทางแบบ Cityblock.....	36
4.5 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดย วัดระยะห่างระยะทางแบบ Cityblock.....	37
4.6 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดย วัดระยะห่างระยะทางแบบ Cityblock.....	38
4.7 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดย วัดระยะห่างระยะทางแบบ Correlation.....	39
4.8 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดย วัดระยะห่างระยะทางแบบ Correlation.....	40

ตารางที่	หน้า
4.9 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดย วัดระยะห่างระยะทางแบบ Correlation.....	41
4.10 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดย วัดระยะห่างระยะทางแบบ Cosine.....	42
4.11 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดย วัดระยะห่างระยะทางแบบ Cosine.....	43
4.12 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดย วัดระยะห่างระยะทางแบบ Cosine.....	44
4.13 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี Normal Kernel.....	45
4.14 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี LSCV.....	46
4.15 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี LKCV.....	47
4.16 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดย วัดระยะห่างระยะทางแบบ Euclidean.....	50
4.17 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดย วัดระยะห่างระยะทางแบบ Euclidean.....	51
4.18 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดย วัดระยะห่างระยะทางแบบ Euclidean.....	52

ตารางที่	หน้า
4.19 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดย วัดระยะห่างระยะทางแบบ Cityblock.....	53
4.20 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดย วัดระยะห่างระยะทางแบบ Cityblock.....	54
4.21 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดย วัดระยะห่างระยะทางแบบ Cityblock.....	55
4.22 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดย วัดระยะห่างระยะทางแบบ Correlation.....	56
4.23 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดย วัดระยะห่างระยะทางแบบ Correlation.....	57
4.24 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดย วัดระยะห่างระยะทางแบบ Correlation.....	58
4.25 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดย วัดระยะห่างระยะทางแบบ Cosine.....	59
4.26 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดย วัดระยะห่างระยะทางแบบ Cosine.....	60
4.27 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดย วัดระยะห่างระยะทางแบบ Cosine.....	61

ตารางที่	หน้า
4.28 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี Normal Kernel.....	62
4.29 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี LSCV.....	63
4.30 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี LKCV.....	64
4.31 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดย วัดระยะห่างระยะทางแบบ Euclidean.....	67
4.32 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดย วัดระยะห่างระยะทางแบบ Euclidean.....	68
4.33 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดย วัดระยะห่างระยะทางแบบ Euclidean.....	69
4.34 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดย วัดระยะห่างระยะทางแบบ Cityblock.....	70
4.35 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดย วัดระยะห่างระยะทางแบบ Cityblock.....	71
4.36 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดย วัดระยะห่างระยะทางแบบ Cityblock.....	72
4.37 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดย วัดระยะห่างระยะทางแบบ Correlation.....	73

ตารางที่	หน้า
4.38 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดย วัดระยะห่างระยะทางแบบ Correlation.....	74
4.39 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดย วัดระยะห่างระยะทางแบบ Correlation.....	75
4.40 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดย วัดระยะห่างระยะทางแบบ Cosine.....	76
4.41 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดย วัดระยะห่างระยะทางแบบ Cosine.....	77
4.42 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดย วัดระยะห่างระยะทางแบบ Cosine.....	78
4.43 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี Normal Kernel.....	79
4.44 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี LSCV.....	80
4.45 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี LKCV.....	81
4.46 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Euclidean.....	85
4.47 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Euclidean.....	86

ตารางที่	หน้า
4.48 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดยวัดระยะห่างระยะทางแบบ Euclidean.....	87
4.49 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดยวัดระยะห่างระยะทางแบบ Cityblock.....	88
4.50 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดยวัดระยะห่างระยะทางแบบ Cityblock.....	89
4.51 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดยวัดระยะห่างระยะทางแบบ Cityblock.....	90
4.52 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดยวัดระยะห่างระยะทางแบบ Correlation.....	91
4.53 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดยวัดระยะห่างระยะทางแบบ Correlation.....	92
4.54 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดยวัดระยะห่างระยะทางแบบ Correlation.....	93
4.55 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดยวัดระยะห่างระยะทางแบบ Cosine.....	94
4.56 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดยวัดระยะห่างระยะทางแบบ Cosine.....	95

ตารางที่	หน้า
4.57 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดยวัดระยะห่างระยะทางแบบ Cosine.....	96
4.58 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี Normal Kernel.....	97
4.59 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี LSCV.....	98
4.60 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี LKCV.....	99
4.61 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดยวัดระยะห่างระยะทางแบบ Euclidean.....	102
4.62 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดยวัดระยะห่างระยะทางแบบ Euclidean.....	103
4.63 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดยวัดระยะห่างระยะทางแบบ Euclidean.....	104
4.64 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดยวัดระยะห่างระยะทางแบบ Cityblock.....	105
4.65 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดยวัดระยะห่างระยะทางแบบ Cityblock.....	106
4.66 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดยวัดระยะห่างระยะทางแบบ Cityblock.....	107

ตารางที่	หน้า
4.67 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดยวัดระยะห่างระยะทางแบบ Correlation.....	108
4.68 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดยวัดระยะห่างระยะทางแบบ Correlation.....	109
4.69 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดยวัดระยะห่างระยะทางแบบ Correlation.....	110
4.70 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดยวัดระยะห่างระยะทางแบบ Cosine.....	111
4.71 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดยวัดระยะห่างระยะทางแบบ Cosine.....	112
4.72 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดยวัดระยะห่างระยะทางแบบ Cosine.....	113
4.73 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี Normal Kernel.....	114
4.74 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี LSCV.....	115
4.75 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี LKCV.....	116
4.76 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดยวัดระยะห่างระยะทางแบบ Euclidean.....	119

ตารางที่	หน้า
4.77 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดยวัดระยะห่างระยะทางแบบ Euclidean.....	120
4.78 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดยวัดระยะห่างระยะทางแบบ Euclidean.....	121
4.79 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดยวัดระยะห่างระยะทางแบบ Cityblock.....	122
4.80 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดยวัดระยะห่างระยะทางแบบ Cityblock.....	123
4.81 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดยวัดระยะห่างระยะทางแบบ Cityblock.....	124
4.82 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดยวัดระยะห่างระยะทางแบบ Correlation.....	125
4.83 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดยวัดระยะห่างระยะทางแบบ Correlation.....	126
4.84 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดยวัดระยะห่างระยะทางแบบ Correlation.....	127
4.85 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดยวัดระยะห่างระยะทางแบบ Cosine.....	128

ตารางที่	หน้า
4.86 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k=3$ โดยวัดระยะห่างระยะทางแบบ Cosine.....	129
4.87 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k=5$ โดยวัดระยะห่างระยะทางแบบ Cosine.....	130
4.88 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี Normal Kernel.....	131
4.89 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี LSCV.....	132
4.90 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย จาก 10 คนสำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี LKCV.....	133
5.1 แสดงการเปรียบเทียบค่าอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียน ตัวอักษรไทย ด้วยวิธี K-nearest neighbour และวิธี Kernel จาก 1 คน.....	138
5.2 แสดงการเปรียบเทียบค่าอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียน ตัวอักษรไทย ด้วยวิธี K-nearest neighbour และวิธี Kernel จาก 10 คน.....	141

สารบัญภาพ

ภาพที่	หน้า
3.1 ข้อมูลต้นแบบของลายมือเขียนตัวอักษรไทย.....	26
3.2 ตัวอย่างการหาขอบของรูปภาพแต่ละตัวอักษร เมื่อผ่านกระบวนการปรับปรุง คุณภาพ.....	27
5.1 แสดงการเปรียบเทียบค่าอัตราความผิดพลาดน้อยที่สุดในแต่ละวิธีการ สำหรับ ตัวอย่างเรียนรู้จำนวนต่างๆ และจำนวนสัมประสิทธิ์ 500 กรณีลายมือเขียนมา จาก 1 คน.....	144
5.2 แสดงการเปรียบเทียบค่าอัตราความผิดพลาดน้อยที่สุดในแต่ละวิธีการ สำหรับ ตัวอย่างเรียนรู้จำนวนต่างๆ และจำนวนสัมประสิทธิ์ 500 กรณีลายมือเขียนมา จาก 10 คน.....	145

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การรู้จำ (Recognition) เป็นศาสตร์หนึ่งที่มีความสำคัญในสาขาวิชาปัญญาประดิษฐ์ (Artificial Intelligence) ซึ่งการรู้จำได้ถูกนำไปใช้ประโยชน์ในด้านต่างๆ มากมาย เช่น การรู้จำภาพ ใบหน้า การรู้จำรูปภาพ การรู้จำตัวอักษร เป็นต้น วิธีการรู้จำรูปแบบมีการศึกษาและวิจัยในสาขาวิชาคอมพิวเตอร์และวิศวกรรมศาสตร์ รวมทั้งสาขาสถิติก็มีวิธีที่สามารถใช้ในการรู้จำรูปแบบ คือ การนำมาวิเคราะห์จำแนกประเภท (discriminant analysis) ซึ่งเป็นการวิเคราะห์เพื่อหากฎการตัดสินใจในการจำแนกกลุ่มของสิ่งของต่างๆ การรู้จำรูปแบบเชิงสถิติจึงนับว่ามีความเกี่ยวข้องกันอย่างใกล้ชิดกับทฤษฎีการตัดสินใจเชิงสถิติด้วย

การรู้จำตัวอักษรเป็นการรู้จำรูปแบบหนึ่งที่มีผู้สนใจทำการวิจัยมานานพอสมควร โดยเริ่มจากการวิจัยการรู้จำตัวพิมพ์ทั้งตัวเลขและตัวอักษรในภาษาต่างๆ หลายภาษา ได้แก่ ภาษาอังกฤษ จีน ญี่ปุ่น เป็นต้น ต่อมาเมื่อผู้วิจัยมาศึกษาการรู้จำรูปแบบลายมือเขียน เนื่องจากได้พบปัญหาว่าลายมือเขียนของแต่ละบุคคลมีความแตกต่างกัน ความผิดพลาดในการอ่านลายมือเขียนนั้นจึงเกิดขึ้นได้มาก จึงมีวัตถุประสงค์ที่จะสร้างเครื่องมือหรือระบบคอมพิวเตอร์ที่สามารถรับข้อมูลเข้าจากการเขียนด้วยลายมือของผู้ใช้ เพื่อสร้างความสะดวกและความถูกต้องในการรู้จำให้ผู้ใช้มากยิ่งขึ้น

การรู้จำลายมือเขียนตัวอักษรประกอบด้วย 3 ขั้นตอน (กนกศรี, 2540) ขั้นตอนที่หนึ่งเป็นขั้นตอนการเก็บข้อมูลโดยใช้เครื่องมือต่างๆ เช่น สแกนเนอร์ เพื่อรับข้อมูลในลักษณะที่เป็นภาพ และปากกาอิเล็กทรอนิกส์ เครื่องอ่านพิกัด (digitizer) เพื่อรับข้อมูลในลักษณะที่เป็นจุด (pixel) ขั้นตอนที่สองเป็นขั้นตอนการแปลงข้อมูลที่วัดค่าได้ให้อยู่ในรูปแบบที่สามารถนำไปคำนวณต่อไป ส่วนขั้นตอนสุดท้ายคือขั้นตอนการสร้างกฎการตัดสินใจสำหรับการแนกกลุ่มที่ได้จากขั้นตอนที่สอง โดยใช้วิธีการต่างๆ ที่มีอยู่หลายวิธี ซึ่งลายมือเขียนตัวอักษรของแต่ละบุคคลนั้นมีลักษณะที่

แตกต่างกันอย่างแน่นอน ไม่ว่าจะ เป็นขนาดของตัวอักษรที่เขียนไม่เท่ากันหรือการเขียนเอียงของตัวอักษร จึงทำให้เกิดปัญหาของการเขียนตัวอักษรบางตัวนั้นไปคล้ายคลึงกับตัวอักษรอื่น ทำให้การรู้จำของตัวอักษรมีความผิดพลาดเกิดขึ้น ได้มีผู้วิจัยได้นำการรู้จำรูปแบบเชิงสถิติมาแก้ไขปัญหาดังกล่าวด้วยวิธีต่างๆ เช่น วิธีเพื่อนบ้านใกล้เคียงที่สุด k ค่า ซึ่งวิธีนี้เป็นวิธีตัดสินใจว่ากลุ่มใดจะสามารถแทนเงื่อนไขใดได้บ้าง โดยการตรวจสอบค่า k ของเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด ซึ่งมีข้อดีในการจำแนกกลุ่มได้ง่ายและรวดเร็ว แต่ถ้าข้อมูลมีจำนวนมากจะทำให้ประสิทธิภาพในการคำนวณมีความผิดพลาดเกิดขึ้น ส่วนวิธีเคอร์เนล เป็นวิธีที่จะใช้ฟังก์ชันความหนาแน่นความน่าจะเป็นมาคำนวณแล้วตัดสินใจการจำแนกกลุ่มด้วยความน่าจะเป็นภายหลัง ซึ่งมีข้อดีในการจำแนกกลุ่มได้ง่าย แต่จะมีปัญหากับข้อมูลที่มี outlier นอกจากนี้ได้มีการใช้วิธีการจัดกลุ่มแบบฟัซซีหรือการจำแนกกลุ่มด้วยวิธีซัพพอตเวกเตอร์แมชชีนมาใช้ ดังนั้นผู้วิจัยจึงสนใจที่จะศึกษาการรู้จำของลายมือเขียนตัวอักษรเพื่อให้เกิดปัญหาที่เกิดขึ้นในการรู้จำนั้นมีความผิดพลาดเกิดขึ้นน้อยที่สุด จึงได้ทำการศึกษางานวิจัยที่เกี่ยวข้องกับการรู้จำลายมือเขียนตัวอักษรต่างๆ ด้วยวิธีต่างๆ โดยมีงานวิจัยที่ศึกษา ดังนี้

กนกศรี ศรีนนภากร (2540) ได้ทำการศึกษาการรู้จำลายมือเขียนตัวอักษรภาษาไทยโดยวิธีการจำแนกกลุ่มด้วยวิธีการจำแนกกลุ่ม 3 วิธี คือ วิธีการวิเคราะห์จำแนกประเภทเชิงเส้น วิธีเพื่อนบ้านใกล้เคียงที่สุด k ค่าโดย k กำหนดเป็น 3, 5 และ 8 และวิธีการวิเคราะห์จำแนกประเภทแบบเคอร์เนลที่ใช้ฟังก์ชันปกติเป็นฟังก์ชันความหนาแน่นเคอร์เนล การเก็บข้อมูลลายมือเขียนตัวอักษรเก็บโดยใช้ปากกาอิเล็กทรอนิกส์เป็นค่าพิกัดของจุดต่างๆของตัวอักษรบนระนาบ (X, Y) จากผลการศึกษาพบว่า วิธีการวิเคราะห์จำแนกประเภทเคอร์เนลเป็นวิธีที่ให้ค่าอัตราความผิดพลาดต่ำที่สุด

ศิริรัตน์ ทินตะนัย ดร. อโนทัย ตริวานิช และ ชัชวาลย์ เรืองประพันธ์ (2544) ได้ทำการศึกษาการรู้จำลายมือเขียนตัวเลขอารบิกด้วยวิธีการจำแนกกลุ่ม 2 วิธี คือ วิธีจำแนกแบบใช้พารามิเตอร์ ได้แก่ วิธีการวิเคราะห์จำแนกประเภทเชิงเส้น และวิธีจำแนกแบบไม่ใช้พารามิเตอร์ ได้แก่ วิธีเคอร์เนล โดยการแบ่งพื้นที่ย่อยของตัวเลขในขนาดพื้นที่ย่อยต่างๆ จากผลการศึกษาพบว่า วิธีจำแนกแบบไม่ใช้พารามิเตอร์เป็นวิธีที่ให้ค่าอัตราความผิดพลาดต่ำที่สุด

กานต์ เสารยะวิเศษ (2545) ได้ทำการศึกษาคำรู้จำลายมือเขียนตัวเลขไทยโดยใช้การจัดกลุ่มแบบฟัซซี่ โดยการหาขอบภาพของตัวอักษรมาทำการแปลงฟูเรียร์ ซึ่งต้องการจำแนกลายมือเขียนจาก 1 คนและลายมือเขียนจากหลายคน จากผลการศึกษาพบว่า การจำแนกลายมือเขียนจาก 1 คนให้ค่าอัตราความผิดพลาดต่ำที่สุด

จากงานวิจัยที่กล่าวมาข้างต้นและได้ทำการศึกษาข้อมูลต่างๆ จึงทำให้ผู้วิจัยสนใจที่จะศึกษาคำรู้จำลายมือเขียนตัวอักษร โดยจะศึกษาลายมือเขียนตัวอักษรไทย เนื่องจากตัวอักษรไทยเป็นภาษาประจำชาติและมีความเป็นเอกลักษณ์ แต่ตัวอักษรไทยบางตัวนั้นก็ยังมีลักษณะที่คล้ายคลึงกัน ทำให้เกิดปัญหาในการจำแนกกลุ่มอาจมีความผิดพลาดเกิดขึ้นได้ง่าย ผู้วิจัยจึงได้หาวิธีที่จะทำให้การจำแนกกลุ่มลายมือเขียนตัวอักษรไทยมีความผิดพลาดน้อยที่สุด ผู้วิจัยจึงสนใจที่จะใช้วิธีการจำแนกกลุ่มด้วยวิธีจำแนกกลุ่มแบบไม่ใช้พารามิเตอร์ 2 วิธี คือ วิธีเคอร์เนล (Kernel method) และวิธี K-nearest neighbour เนื่องจาก 2 วิธีนี้ให้ความผิดพลาดในการจำแนกกลุ่มที่ต่ำ และจะศึกษา 2 วิธีนี้ในรูปแบบต่างๆเพิ่มขึ้น คือ วิธีเคอร์เนล จะศึกษาเคอร์เนล 3 แบบ คือ Normal kernel, Least square cross-validation (LSCV) และ Likelihood crossvalidation (LKCV) และวิธี K-nearest neighbour โดยจะวัดระยะห่างระยะทาง 4 แบบ คือ Euclidean, Cityblock, Correlation และ Cosine เพื่อให้เกิดความผิดพลาดในการจำแนกกลุ่มที่ได้น้อยลง โดยการหาขอบภาพของตัวอักษรมาทำการแปลงฟูเรียร์ เนื่องจาก การแปลงฟูเรียร์นั้นมีความคงทนต่อการเลื่อนจุด การหมุน และการย่อหรือขยายของตัวอักษร เพื่อให้เห็นถึงประสิทธิภาพข้อดีและข้อเสียของคำรู้จำลายมือเขียนตัวอักษรไทยในการจำแนกกลุ่ม และเพื่อเป็นอีกทางเลือกหนึ่งให้ผู้สนใจสามารถเอาไปประยุกต์กับชีวิตประจำวันได้

1.2 วัตถุประสงค์ในการศึกษา

ศึกษาเปรียบเทียบประสิทธิภาพในการวิเคราะห์ลายมือเขียนตัวอักษรไทยและคำรู้จำด้วยวิธีจำแนกกลุ่มแบบไม่ใช้พารามิเตอร์ระหว่างวิธีเคอร์เนลกับวิธี K-nearest neighbour ในแต่ละสถานการณ์ที่กำหนด โดยใช้ค่าอัตราความผิดพลาดของการจำแนกเป็นเกณฑ์ในการเปรียบเทียบ

1.3 สมมติฐานของการศึกษา

ข้อมูลลายมือเขียนตัวอักษรไทยในลักษณะข้อมูลที่แตกต่างกัน คือ ตัวอักษรลายมือเขียนไทย ค่าสัมประสิทธิ์ฟูเรียร์ และขนาดตัวอย่างเรียนรู้เมื่อวิเคราะห์การจำแนกกลุ่มด้วย 2 วิธี คือ วิธี K-nearest neighbour และวิธี Kernel มีผลทำให้เกิดอัตราความผิดพลาดในการจำแนกกลุ่มที่แตกต่างกัน

1.4 คำจำกัดความที่ใช้ในการศึกษา

การรู้จำ หมายถึง การนำชุดข้อมูลจำหนึ่งป้อนให้ระบบเพื่อให้ระบบทำการจำข้อมูลนั้น เรียกว่า ขั้นตอนการเรียนรู้ และการนำข้อมูลอีกชุดหนึ่งมาป้อนให้ระบบทำการพิจารณาว่าข้อมูลที่อ่านเข้ามาเปรียบเทียบกับข้อมูลใดในข้อมูลชุดแรก เรียกว่า ขั้นตอนการทดสอบ

ประสิทธิภาพ หมายถึง ความสามารถในการจำแนกกลุ่มข้อมูลของการจำแนกทั้งสองวิธี โดยเปรียบเทียบจากอัตราความผิดพลาดในการจำแนกกลุ่มวิธีหนึ่งต่ออัตราความผิดพลาดในการจำแนกกลุ่มอีกวิธีหนึ่ง

1.5 ขอบเขตการศึกษา

ในการวิจัยครั้งนี้จะทำการศึกษากายใต้ขอบเขตดังนี้

1. เก็บบันทึกข้อมูลลายมือเขียนตัวอักษรไทย คือ ข, ชม, ซ, ซ, ฉ, ฉ, ก, ก, ก, ก, ค, ค, บ และ ม โดยแต่ละตัวอักษรจะเขียนจำนวน 90 ครั้งจำนวน 2 ชุด คือ ชุดที่ 1 จากลายมือเขียนตัวอักษรไทย 1 คน และชุดที่ 2 จากลายมือเขียนตัวอักษรไทย 10 คนในตารางขนาด 1×1 นิ้วต่อ 1 ตาราง และทำการสแกนโดยเครื่องสแกน แล้วเก็บผลลัพธ์เป็นแฟ้มข้อมูลรูปภาพ 1 ตัวอักษรต่อ 1 แฟ้มข้อมูล
2. นำตัวอักษรไทยมาทำการแปลงฟูเรียร์โดยกำหนดค่าสัมประสิทธิ์ที่ได้คือ 200, 500 และ 1,000

3. ทำการสุ่มตัวอักษรแต่ละตัวอักษรไทยแบบไม่ใส่คืนจำนวน 30 ตัวจากตัวอักษร 90 ตัวของแต่ละตัวอักษรไทย เพื่อเป็นตัวอย่างทดสอบ และตัวอักษรที่เหลือแต่ละตัวอักษรไทยจำนวน 60 ตัวจะเก็บเป็นตัวอย่างเรียนรู้

4. จากตัวอย่างเรียนรู้ตัวอักษรที่ศึกษา ซึ่งแต่ละตัวอักษรไทยที่ศึกษามี 60 ตัว จะทำการสุ่มจำนวน 10, 20, 30, 40, 50 และ 60 เพื่อเป็นตัวอย่างเรียนรู้

5. นำตัวอย่างเรียนรู้และตัวอย่างทดสอบมาทำการจำแนกกลุ่ม 2 วิธี คือ

5.1 วิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทาง 4 แบบ

คือ

5.1.1 ระยะทางแบบ Euclidean

5.1.2 ระยะทางแบบ Cityblock

5.1.3 ระยะทางแบบ Correlation

5.1.4 ระยะทางแบบ Cosine

5.2 วิธี Kernel 3 แบบ คือ

5.2.1 Normal kernel

5.2.2 Least square cross-validation (LSCV)

5.2.3 Likelihood cross-validation (LKCV)

6. คำนวณหาค่าอัตราความผิดพลาดในการจำแนก

7. ในการศึกษาครั้งนี้ทำการทดลองซ้ำสำหรับตัวอย่างเรียนรู้ที่แตกต่างกัน 3 ครั้ง

1.6 วิธีดำเนินการศึกษา

1. ศึกษาค้นคว้าเอกสารต่างๆทั้งทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการจำแนกกลุ่ม
 - วิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระหว่างระยะทาง 4 แบบ คือ
 1. ระยะทางแบบ Euclidean
 2. ระยะทางแบบ Cityblock
 3. ระยะทางแบบ Correlation
 4. ระยะทางแบบ Cosine
 - วิธี Kernel 3 แบบ คือ
 1. Normal kernel
 2. Least square cross-validation (LSCV)
 3. Likelihood cross-validation (LKCV)
 - ค่าอัตราความผิดพลาดในการจำแนกกลุ่ม
2. เก็บข้อมูลลายมือเขียนตัวอักษรไทยตามขนาดตัวอย่างที่กำหนด
3. นำข้อมูลลายมือเขียนตัวอักษรไทยมาประมวลผล ซึ่งแบ่งได้เป็น 2 ขั้นตอน คือ
 - ขั้นตอนที่ 1 Pre-Processing เป็นกระบวนการปรับปรุงคุณภาพและหาขอบของภาพ
 - ขั้นตอนที่ 2 Feature Extraction เป็นกระบวนการค้นหาคุณลักษณะพิเศษที่แฝงอยู่ในรูปภาพตัวอักษร โดยใช้ตัวบรรยายลักษณะฟูเรียร์ (Fourier Descriptors)
4. ทำการสุ่มขนาดตัวอย่างแบบไม่ใส่คืนเพื่อเป็นตัวอย่างทดสอบและตัวอย่างเรียนรู้ตามที่กำหนด
5. คำนวณหาค่าเฉลี่ยอัตราความผิดพลาดในการจำแนกกลุ่มด้วยวิธี K-nearest neighbour และวิธี Kernel
6. วิเคราะห์และสรุปผลการเปรียบเทียบวิธีการที่ใช้ในการวิจัย

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1. ทราบประสิทธิภาพในการจำแนกกลุ่มภายใต้สถานการณ์ที่ศึกษา
2. เพื่อเป็นแนวทางในการพัฒนาวิธีการรู้จำลายมือเขียนตัวอักษร เพื่อให้มีการประยุกต์ใช้กับภาษาไทยได้อย่างเต็มรูปแบบ และให้มีการวิจัยในระดับสูงเพื่อให้ได้การรู้จำที่มีประสิทธิภาพสูงขึ้นต่อไป
3. เพื่อเป็นแนวทางในการศึกษาเปรียบเทียบเกณฑ์ทางสถิติอื่น ๆ ที่เกี่ยวข้องต่อไป

บทที่ 2

ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

2.1 ทฤษฎีพื้นฐาน

2.1.1 การรู้จำรูปแบบ (Pattern Recognition)

มนุษย์มีความสามารถในการรู้จำได้จากการเห็นภาพ การได้ยินเสียง การสัมผัส การได้กลิ่นและการรู้รส ซึ่งทำให้มนุษย์สามารถแยกแยะวัตถุต่างๆ ที่อยู่รอบตัวที่มีรูปแบบต่างกันอย่างชัดเจนได้ โดยทั่วไปแล้วการรู้จำรูปแบบสามารถแบ่งออกเป็น 3 ประเภทคือ (สมยศ, 2540)

2.1.1.1 การรู้จำรูปแบบของมนุษย์ สามารถแบ่งเป็น 2 ประเภทคือ

1. การรู้จำในสิ่งที่เป็นรูปธรรม (The Recognition of Concrete Items)

การรู้จำประเภทนี้ยังสามารถจำแนกย่อยได้อีกสองกลุ่มคือ การรู้จำในสิ่งที่มีรูปแบบอยู่อย่างถาวร เช่น ตัวอักษร และวัตถุต่างๆ เป็นต้น และการรู้จำในสิ่งที่มีรูปแบบอยู่ชั่วคราว เช่น คลื่น เสียง อนุกรมเวลา เป็นต้น

2. การรู้จำในสิ่งที่เป็นนามธรรม (The Recognition of Abstract Items)

เป็นการรู้จำในสิ่งที่ไม่รูปร่าง หรือการรู้จำในสิ่งที่ได้จากความคิด (Conceptual Recognition)

2.1.1.2 การรู้จำแบบจำแนกตามวิธีการรู้จำ สามารถแบ่งเป็น 4 วิธีคือ

1. วิธีการเทียบคู่รูปแบบ

วิธีนี้เป็นวิธีการแรกๆ ที่ใช้ในการรู้จำตัวอักษร หลักการโดยทั่วไปคือ จะต้องมีต้นแบบที่สร้างขึ้นมาสำหรับอ่านตัวอักษรปกติจะเป็นข้อมูล 2 มิติ โดยมีการกำหนดตำแหน่งที่สำคัญที่สามารถใช้แยกแยะความแตกต่างระหว่างตัวอักษรแต่ละตัว การทำงานของระบบจะนำรูปภาพที่ต้องการไปเปรียบเทียบกับต้นแบบเพื่อวัดความคล้ายคลึงของภาพกับตัวแบบ จากนั้นก็ระบุว่าเป็นรหัสของตัวอักษรอะไร วิธีการนี้ค่อนข้างจะอ่อนไหวต่อข้อมูลแทรกซ้อน ขนาดของภาพ และการเอียงของตัวอักษร

2. วิธีการทางสถิติ

วิธีนี้เป็นวิธีการที่ใช้ค่าความน่าจะเป็นหรือฟังก์ชันความน่าจะเป็นมาใช้ในการตัดสินใจรูปภาพอินพุตที่ได้มาจากขั้นตอนการสกัดลักษณะสำคัญ และจะส่งเข้าไปในส่วนการรู้จำเฉพาะของแต่ละตัวอักษร ซึ่งได้ผลลัพธ์ออกมาเป็นค่าความน่าจะเป็นที่อินพุตเป็นตัวอักษรใดเมื่ออินพุตได้ผ่านส่วนการรู้จำได้ครบทุกตัวแล้วก็นำเอาผลลัพธ์ที่ได้ทั้งหมดมาเปรียบเทียบกับว่าค่าความน่าจะเป็นของตัวอักษรตัวใดมากที่สุด ผลลัพธ์จะแสดงออกมาเป็นตัวอักษรนั้น

3. วิธีการวิเคราะห์ทางโครงสร้าง

เป็นการนำภาพตัวอักษรแต่ละตัวมาวิเคราะห์โครงสร้าง โดยถือว่าตัวอักษรทุกตัวประกอบด้วยองค์ประกอบพื้นฐานที่ได้มาจากการสกัดลักษณะสำคัญที่ส่งมาให้กับขั้นตอนการรู้จำแบบการวิเคราะห์ทางโครงสร้างนี้มักจะใช้ชื่อหรือค่าที่บอกว่าลักษณะโครงสร้างสำคัญนั้นเป็นอะไร เช่น เส้นตรง วงกลม เป็นต้น แทนที่จะเป็นค่าจำนวนจริง วิธีการนี้มีข้อดีตรงที่มีความยืดหยุ่นต่อความหลากหลายของตัวอักษรค่อนข้างมาก แต่อย่างไรก็ตามอัตราความถูกต้องของวิธีนี้ขึ้นอยู่กับ การสร้างกฎ และการวิเคราะห์กฎที่มีประสิทธิภาพซึ่งเป็นส่วนสำคัญที่สุดของวิธีการนี้

4. วิธีทางโครงข่ายประสาทเทียม

วิธีนี้เป็นเทคนิคที่พยายามเลียนแบบการทำงานของมนุษย์ ที่มีโครงข่ายเชื่อมต่อกันของหน่วยความจำย่อยจำนวนมากที่สะสมความรู้เอาไว้ ความรู้เหล่านี้จะได้รับการฝึกสอนไว้ก่อน โครงข่ายประสาทเทียมจะเรียนรู้ถึงรูปแบบตัวอักษรที่หลากหลายของตัวอักษรเหล่านั้น เพื่อว่าเวลาทำงานจริงจะได้มีความสามารถพอที่จะรับมือกับภาพตัวอักษรในหลายๆรูปแบบ สิ่งที่สอนให้กับโครงข่ายประสาทเทียมไม่จำเป็นต้องเป็นรูปของตัวอักษรอย่างที่เรเห็นก็ได้ ข้อมูลที่ส่งให้มักจะผ่านขั้นตอนการสกัดลักษณะสำคัญ และกระบวนการประมวลผลเบื้องต้นอื่นๆ ก่อนเสมอ

2.1.1.3 ระบบของการรู้จำรูปแบบโดยคอมพิวเตอร์ ประกอบด้วย

1. ส่วนของการนำเข้า ทำหน้าที่ในการแปลงรูปแบบที่ต้องการวิเคราะห์ให้เป็นสัญญาณอิเล็กทรอนิกส์ อุปกรณ์นำข้อมูลเข้าที่นิยมใช้ ได้แก่ กล้องวิดีโอ (Video Camera) เครื่องอ่านพิกัดภาพ (Image Digitizers) เครื่องสแกน เป็นต้น

2. ส่วนการประมวลผลขั้นต้น เป็นการกระทำกับสัญญาณตามเงื่อนไขที่เพิ่มเติมขึ้นและอาจรวมถึงหน้าที่ในการขยายสัญญาณสเปเชียลฟิลเตอร์ริง (Spatial Filtering) การวิเคราะห์สเปกตรัม และการแปลงสัญญาณอนาล็อกไปเป็นสัญญาณดิจิทัล เป็นต้น

3. ส่วนของการค้นหาตัวบรรยายลักษณะ หรือเรียกว่าส่วนของการแสดงให้เห็นความแตกต่าง (Discriminators) ทำหน้าที่จับคู่ต้นแบบ (Template Matching) เป็นต้น

4. ส่วนของการเลือกผลลัพธ์ (Response Selector) หรือส่วนของการรู้จำรูปแบบ เป็นขั้นตอนการเปรียบเทียบคุณลักษณะพิเศษของสิ่งที่ทำการรู้จำ เพื่อเลือกรูปแบบเก็บไว้ในกลุ่มที่เหมาะสม ซึ่งเป็นรูปแบบที่เหมือนกับรูปแบบที่ต้องการรู้จำมากที่สุด

5. ส่วนของการแสดงผลลัพธ์ เป็นกระบวนการแสดงผลของสิ่งที่ได้ผ่านกระบวนการรู้จำ ระบบการแสดงผลลัพธ์อาจใช้ ตัวกำเนิดเสียง รูปภาพ หรือจอภาพเป็นสื่อในการแสดงผล

2.1.2 การรู้จำตัวอักษร

การรู้จำตัวอักษรสามารถที่จะแบ่งแยกย่อยออกเป็นประเภทที่แตกต่างกันเป็นจำนวนมาก และที่มีความสำคัญเป็นอย่างมากก็คือการแยกคุณลักษณะพิเศษของตัวอักษรที่ต้องการนำไปรู้จำ การรู้จำตัวอักษรสามารถแยกออกเป็น 2 ประเภท คือ (Vuokko Vuori, 1999)

2.1.2.1 การรู้จำตัวอักษรแบบมีการประมวลผลในขณะเขียน (On-line Recognition)

การรู้จำตัวอักษรแบบมีการประมวลผลในขณะเขียน ข้อความต้นฉบับ ก็คือข้อมูลเข้า และจะทำการรู้จำในเวลาขณะนั้น (Real Time) อุปกรณ์นำข้อมูลเข้าที่ใช้กันมากได้แก่ Tablet ซึ่งสามารถที่จะเก็บข้อมูลในตำแหน่งที่ปากกาเคลื่อนไปบนแผ่นดิจิทัลไจเซอร์ ซึ่งตำแหน่ง หรือพิกัดที่ใช้ปากกาเขียนข้อมูลจะถูกส่งเป็นค่าพิกัดของจุดไปยังคอมพิวเตอร์โดยตรง เพื่อทำการรู้จำในช่วงเวลานั้น

โปรแกรมหลักของการรู้จำตัวอักษรลายมือเขียนแบบมีการประมวลผลในขณะเขียนมีข้อได้เปรียบก็คือ ระบบการรู้จำตัวอักษรลายมือเขียนสามารถที่จะจำลองเพื่อใช้แทนเมาส์และคีย์บอร์ด ดังนั้นระบบจึงจำเป็นอย่างมากที่จะใช้ Tablet สำหรับเขียน เพื่อส่งข้อมูลไปทำการรู้จำและผลลัพธ์ที่ได้จากการรู้จำจะถูกส่งกลับไปยังโปรแกรมหลักเพื่อแสดงผลต่อไป

2.1.2.2 การรู้จำตัวอักษรแบบไม่มีการประมวลผลในขณะเขียน (Off-line

Recognition)

การรู้จำตัวอักษรลายมือเขียนถือว่าเป็นเรื่องที่ยุ่งยาก ซึ่งจะขึ้นอยู่กับตัวโปรแกรม (Application) ซึ่งโปรแกรมนี้เป็นระบบซึ่งทำการรู้จำตัวอักษรต้นฉบับเดิมที่เขียนเป็นลายลักษณ์อักษรลงบนกระดาษจากนั้นแปลงจากแผ่นกระดาษให้อยู่ในลักษณะของรูปภาพสองมิติ ซึ่งเป็นรูปภาพแบบบิตแม็พ ซึ่งเรียกกันว่าการประมวลผลรูปภาพแบบดิจิทัล ซึ่งการการรู้จำประเภทนี้ เรียกว่าการรู้จำตัวอักษรแบบไม่มีการประมวลผลในขณะเขียน ซึ่งระบบจะยังไม่ทำการรู้จำตัวอักษรในเวลานั้น แต่จะทำการรู้จำเมื่อทำการเขียนตัวอักษรเสร็จสิ้น การรู้จำตัวอักษรแบบไม่มีการประมวลผลในขณะเขียนนั้นเป็นส่วนประกอบหนึ่งของ Optical Character Recognition (OCR)

ตัวอักษรที่นำมาประยุกต์ใช้กับการรู้จำตัวอักษรแบ่งเป็น 2 ประเภทคือ (สมยศม ,2540)

1) ตัวอักษรตัวพิมพ์ (Printed Character)

รูปแบบของตัวอักษรชนิดนี้จะได้มาจากการพิมพ์ เช่น จากเครื่องพิมพ์ดีด หรืออุปกรณ์การพิมพ์อื่นๆ โดยมีรูปแบบตัวอักษรที่แน่นอนตามชนิดของแต่ละอุปกรณ์การพิมพ์ การพิมพ์เหล่านี้ ได้แก่ สิ่งพิมพ์ต่างๆ เช่น หนังสือ เป็นต้น

การรู้จำตัวอักษรตัวพิมพ์สามารถแบ่งได้เป็น 2 ประเภทคือ

ก) การรู้จำตัวอักษรตัวพิมพ์รูปแบบเดียว (Recognition of Single-Font Characters)

การรู้จำตัวอักษรประเภทนี้มีจุดมุ่งหมาย เพื่อให้เครื่องคอมพิวเตอร์สามารถอ่าน และรู้จำตัวอักษรตัวพิมพ์ได้เฉพาะรูปแบบของตัวอักษรที่เก็บไว้เพียงรูปแบบเดียวเท่านั้น

ข) การรู้จำตัวอักษรตัวพิมพ์หลายรูปแบบ (Recognition of Multiple-Font Characters)

การรู้จำตัวอักษรตัวพิมพ์ประเภทนี้มีจุดมุ่งหมายเพื่อให้เครื่องคอมพิวเตอร์สามารถอ่าน และรู้จำตัวอักษรตัวพิมพ์ได้มากกว่าหนึ่งรูปแบบของตัวอักษร

2) ตัวอักษรลายมือเขียน (Handwritten Character)

รูปแบบตัวอักษรลายมือเขียนได้มาจากการเขียนทั่วไปซึ่งไม่สามารถกำหนดรูปแบบที่แน่นอนได้ หากแต่จะขึ้นอยู่กับพฤติกรรมการเขียนของแต่ละบุคคล สามารถแบ่งออกเป็น 2 ประเภท คือ

ก) ลายมือเขียนแบบตัวโดด (Isolated Handprint Character)

ตัวอักษรของกลุ่มนี้เป็นลายมือเขียนที่มักจะถูกกำหนดให้เขียนในกรอบที่จัดไว้ โดยเขียนครั้งละตัวแยกจากกัน การรู้จำลายมือเป็นเรื่องที่ยุ้งยากเนื่องจากตัวอักษรที่ได้จากการเขียนของแต่ละบุคคลมีลักษณะแตกต่างกันมากกว่าตัวอักษรที่ได้จากเครื่องพิมพ์

ข) ลายมือแบบเขียนต่อเนื่อง (Scrip Recognition)

เป็นการรู้จำที่ยากที่สุดในกลุ่มของการรู้จำทั้งหมด การรู้จำแบบนี้บางครั้งเรียกรู้จำแบบลายมือเขียนอิสระ (Freestyle Handwritten Recognition) เพราะตัวอักษรที่จะต้องรู้จำเป็นตัวอักษรที่เป็นลายมือเขียนโดยไม่มีข้อกำหนดใดๆ ผู้เขียนสามารถเขียนได้ตามธรรมชาติอย่างต่อเนื่อง ดังนั้นตัวอักษรที่ได้จากมีเส้นที่ลากเชื่อมตัวอักษรหลายๆ ตัวให้ติดกันประกอบกับความแตกต่างอย่างมากของลายมือแต่ละคน

2.1.3 การประมวลผลภาพ (Image Processing) (Castleman, 1995)

2.1.3.1 การปรับปรุงคุณภาพของภาพ (Image Enhancement)

วัตถุประสงค์หลักของการปรับปรุงคุณภาพของภาพ คือเพื่อให้ได้ภาพที่มีลักษณะเหมาะสมกับความต้องการใช้งาน ซึ่งขึ้นอยู่กับลักษณะเฉพาะที่เราต้องการ เช่นการปรับปรุงรายละเอียดของภาพที่ปรากฏ กำจัดลดลายที่ไม่ต้องการ ปรับปรุงความคมชัด (Contrast) เป็นต้น ดังนั้นปริมาณวิฤต (Criteria Quality) ซึ่งใช้เป็นหลักในการประเมินคุณภาพของภาพที่ได้จากการประมวลผลด้วยวิธีการนี้จึงค่อนข้างเป็นเชิงนามธรรม (Subjective)

2.1.3.2 การเรียกคืนภาพ (Image Restoration)

ขณะบันทึกภาพอาจมีปัญหาเกิดขึ้น เช่นการสั่นสะเทือนของอุปกรณ์บันทึกภาพ ความผิดพลาดเนื่องมาจากการปรับโฟกัส จนทำให้ได้ภาพที่เบลอ เพื่อที่จะให้ได้ภาพซึ่งมีคุณภาพดีเช่นปกติที่ควรจะเป็น เราจำเป็นต้องใช้กระบวนการซึ่งเรียกว่า การเรียกคืน (คุณภาพของ) ภาพเข้ามาประมวลผลภาพ กระบวนการนี้สามารถใช้ปรับปรุงคุณภาพของภาพได้โดยการลดหรือกำจัดผลที่เกิดจากความคาดเคลื่อน (Degradation) ข้างต้น โดยสร้างแบบจำลองทางคณิตศาสตร์ตามชนิดของความคาดเคลื่อนที่เกิดขึ้น แล้วใช้กระบวนการผวนกลับ (InverseProcess) เพื่อที่จะเรียกคืน (Recover) ภาพต้นฉบับ

2.1.3.3 การบีบอัดและขยายกลับคืนภาพ (Image Compression and Decompression)

เพื่อลดปัญหาขนาดของหน่วยความจำเพื่อใช้จัดเก็บภาพเพิ่มขึ้นอย่างรวดเร็วเมื่อภาพและจำนวนบิตต่อจุดภาพเพิ่มขึ้นจึงจำเป็นต้องใช้วิธีการบีบอัด (Compression) เข้าช่วยในการจัดเก็บและส่งผ่านข้อมูลภาพ วิธีการนี้มีหลักการสำคัญคือ ลดจำนวนข้อมูลที่ซ้ำซ้อน (Redundancy) ที่มีอยู่ในภาพที่กำลังพิจารณา ดังนั้นผลที่ได้ก็คือข้อมูลภาพที่มีขนาดเล็กลง ทำให้ประหยัดเนื้อที่ในการจัดเก็บและสามารถส่งไปทางสายสัญญาณสื่อสารได้รวดเร็ว เมื่อข้อมูลภาพเดินทางถึงปลายทางหรือ เมื่อต้องการใช้ข้อมูลนี้โดยอ่านจากอุปกรณ์จัดเก็บก็จะใช้วิธีการขยายกลับคืน(Decompression) เพื่อสร้างภาพต้นฉบับขึ้นมาใหม่

2.1.3.4 การตัดแบ่งภาพ (Image Segmentation)

จุดประสงค์หลักของการตัดแบ่งภาพ คือ เพื่อหาอาณาเขตของบริเวณต่างๆ ในภาพซึ่งแต่ละบริเวณอาจสื่อความหมายที่แตกต่างกันออกไป เช่น บริเวณที่เป็นพื้นหลัง บริเวณที่มีค่าความเข้มที่เท่ากันซึ่งมีความเป็นไปได้สูงที่จะเป็นบริเวณที่สมนัยกับวัตถุเชิงกายภาพวัตถุเดียวกัน เป็นต้น เพื่อให้ระบบคอมพิวเตอร์สามารถเข้าใจความหมายของข่าวสารที่แฝงอยู่ในภาพและใช้ประโยชน์ในระดับสูงขึ้น โดยทั่วไปการตัดแบ่งภาพ แบ่งออกเป็น 2 กลุ่ม คือการตัดแบ่งโดยอาศัยความสม่ำเสมอของลักษณะ ของบริเวณแต่ละบริเวณ (Similarity- based Segmentation) การตัดแบ่งโดยใช้ความไม่ต่อเนื่องของลักษณะของแต่ละบริเวณ (Discontinuity-based Segmentation) โดยพิจารณาตำแหน่งของจุดเส้นขอบ ซึ่งเกิดจากการเชื่อมกันของจุดภาพหลายๆ จุดภาพเข้าด้วยกัน

2.1.3.5 การบรรยายและแทนความหมายของภาพ (Image Description and Representation)

โดยทั่วไปภาพแต่ละภาพจะมีคุณลักษณะ หรือความหมาย (Meaning) แฝงไว้ ผู้ที่เห็นภาพแต่ละคนอาจแปลความหมายต่างกันออกไปซึ่งส่วนหนึ่งขึ้นอยู่กับประสบการณ์ของคนผู้นั้น ดังนั้น จึงควรมีการกำหนดข้อตกลงเกี่ยวกับลักษณะวิธีการนำเสนอ และการแทนความหมายของภาพในรูปของตัวบรรยายลักษณะ เพื่อที่จะได้สื่อสารระหว่างกันได้อย่างถูกต้องตรงกันและมีประสิทธิภาพยิ่งขึ้น

ตัวบรรยายลักษณะที่สำคัญของภาพสามารถค้นหาได้จาก

1. ขอบภาพ (Edge)
2. ขอบเขต (Boundary)
3. บริเวณ (Region)
4. รูปทรง (Shape)
5. ลวดลาย (Texture)
6. สี (Color)

หลังจากที่ตัวบรรยายลักษณะเหล่านี้ถูกเฟ้นออกมาจากภาพ จะถูกใช้เพื่อนำเสนอความหมายของภาพ หนึ่งภาพหรือบริเวณใดๆ ของภาพเดียวกันสามารถถูกอธิบาย โดยใช้ตัวบรรยายลักษณะที่ต่างกัน บริเวณเดียวกันอาจถูกแทนในรูปของตัวบรรยายลักษณะที่เฟ้นมาจากขอบเขตหรือจากลวดลาย

2.1.3.6 การแปลความหมายและรู้จำภาพ (Image Interpretation and Recognition)

หลังจากที่บริเวณต่างๆ ในภาพถูกนำเสนอโดยใช้คุณลักษณะแบบต่างๆ แล้วงานลำดับถัดไปคือการแปลความหมายและรู้จำว่าบริเวณนั้นๆ คือวัตถุใดๆ ซึ่งจะนำไปสู่การสรุปความหมายของภาพ และท้ายสุดก็คือการตอบสนองที่เหมาะสมของหน่วยประมวลผลต่อสิ่งเร้า เช่นการจัดแบ่งกลุ่ม

การประมวลผลภาพนั้นจะรวมถึงกระบวนการหาขอบของวัตถุในภาพ ปรับปรุงคุณภาพขอบของวัตถุในภาพ และกระบวนการเฟ้นหาคุณลักษณะพิเศษที่แฝงอยู่ในรูปภาพ ตัวอักษร โดยใช้ตัวบรรยายลักษณะฟูเรียร์ ฉะนั้นจำเป็นต้องมีความรู้เกี่ยวกับการหาขอบของวัตถุในภาพ และตัวบรรยายลักษณะฟูเรียร์

2.1.4 การหาขอบของวัตถุในภาพ

การหาขอบของวัตถุในภาพนับเป็นขั้นตอนที่สำคัญขั้นตอนหนึ่งของกระบวนการประมวลผลภาพ การหาขอบของวัตถุในภาพนั้นมีเงื่อนไขที่สำคัญ 2 ประการคือ ภาพนั้นจะต้องเป็นรูปที่ต่อเนื่องกัน และภาพที่จะนำมาหาขอบนั้นจะต้องเป็นภาพที่อยู่ในโทนขาว-ดำเท่านั้น อาจใช้เทคนิคการขยายภาพออกจากขอบ 1 จุดภาพ จากนั้นนำไปหาผลต่างขอบของภาพเดิม หรือการหาโดยใช้รหัสสลับสี เคลื่อนที่ไปรอบๆ ขอบภาพนั้นจนกระทั่งครบรอบการเคลื่อนที่ ซึ่งผลของการกระทำดังกล่าวจะได้ขอบของภาพตามต้องการ

2.1.4.1 การแยกภาพวัตถุออกจากพื้นหลัง (Binarization)

การแยกวัตถุออกจากพื้นหลังนั้นคือการเปลี่ยนรูปภาพที่อยู่ใน Gray Level ให้เป็นรูปภาพที่อยู่ใน Black and White Level รูปภาพถูกแบ่งออกเป็นสองส่วน ส่วนแรกจะเรียกว่าวัตถุ ส่วนที่สองเรียกว่าพื้นหลัง โดยทั้งวัตถุและพื้นหลังจะอยู่บนภาพเดียวกัน ทั้งนี้วัตถุจะมีขนาดเล็กกว่าพื้นหลังวิธีการนี้เรียกว่าการแยกภาพวัตถุออกจากพื้นหลัง วิธีการแยกจะพิจารณาแต่ละจุดภาพว่ามีค่าความเข้มแสงน้อยกว่า หรือมากกว่าค่ามาตรฐานที่กำหนดค่านี้เรียกว่า ธรี่สไฮด์ดิ้ง แวลู (Thresholding Value) แสดงดังสมการที่ 2-1

$$g(x, y) = \begin{cases} 0 & \text{if } g(x, y) < T \\ 1 & \text{if } g(x, y) \geq T \end{cases} \quad (2-1)$$

ข้อมูลภาพของลายมือเขียนภาษาไทย เมื่อนำมาผ่านเครื่องกวาดตรวจแบบ Gray Scale 256 ระดับ (มีค่าตั้งแต่ 0-255) จะถูกจัดเก็บเป็นแฟ้มข้อมูลภาพโดย 1 จุดภาพจะมีค่าได้เพียง 2 ค่าเท่านั้นคือ

0 หมายถึง เนื้อภาพ (Object)

1 หมายถึง พื้นหลัง (Background)

2.1.4.2 การแยกจุดภาพบริเวณขอบนอกออกจากภาพวัตถุ

หลังจากที่สามารถแยกวัตถุออกจากพื้นหลังได้แล้ว ภาพที่ได้จะเหลือเพียงวัตถุที่มีค่าความเข้มแสงของจุดภาพทั้งหมดเป็น 1 บนพื้นหลังที่มีค่าความเข้มแสงของจุดภาพเป็น 0 ทั้งหมดวิธีการที่จะหาขอบของภาพคือการแยกจุดภาพที่อยู่บริเวณขอบนอกของวัตถุออกจาก

คุณภาพของวัตถุทั้งหมด วิธีการในการแยกคุณภาพสามารถทำได้หลายวิธี เช่นวิธีการของรหัสลูกโซ่ วิธีการของมอโฟโลยี (Morphology) เป็นต้น

2.1.4.3 ตัวบรรยายลักษณะฟูรีเยร์ (Fourier Descriptors) (นิรันดร์, 2538)

สัญญาณ (Signal) หรือ ฟังก์ชัน (Function) ต่าง ๆ ที่เกิดขึ้นในปัญหาทางกายภาพ (Physical Problem) และทางวิศวกรรม (Engineering Problem) ที่พบเห็นอยู่เสมอ มักจะเป็นฟังก์ชันคาบ (Period Function) ซึ่งฟังก์ชันเหล่านี้สามารถแทนให้อยู่ในเทอม หรือพจน์ของฟังก์ชันคาบของสัญญาณไซน์และโคไซน์ โดยทั่วไปแล้วสัญญาณทั้งสองเป็นสัญญาณพื้นฐานสำหรับพิจารณาสัญญาณอื่นๆ โดยการแทนแต่ละเทอมของอนุกรม (Series) ด้วยฟังก์ชันไซน์กับโคไซน์ ซึ่งเกิดขึ้นจากความคิดของนักฟิสิกส์ และนักคณิตศาสตร์ชาว ฝรั่งเศสชื่อ Joseph Fourier ดังนั้นเพื่อเป็นเกียรติแก่เขาจึงเรียกฟังก์ชัน ดังกล่าวนี้ว่า อนุกรมฟูรีเยร์ (Fourier Series) ซึ่งอนุกรมฟูรีเยร์เสมือนเป็นเครื่องมือที่นำมาใช้แก้ปัญหาต่างๆ เช่น สมการเชิงอนุพันธ์สามัญ (Ordinary Differential Equation) และสมการเชิงอนุพันธ์ย่อย (Partial Differential Equation)

1. การแปลงฟูรีเยร์ (Fourier Transform)

จุดมุ่งหมายอันหนึ่งในการวิเคราะห์สัญญาณ คือ เพื่อที่จะสามารถแทนสัญญาณใดๆ ให้อยู่ในเทอมของฟังก์ชันคาบ (Periodic Function) และนอกจากนั้นการวิเคราะห์ความถี่ของระบบเชิงเส้น (Linear System) นั้นเราต้องการแทนฟังก์ชันอินพุต หรือฟังก์ชันขับ (Driving Function) ทุกๆ ฟังก์ชันให้อยู่ในรูปผลบวกแบบดิสครีต (Discrete) หรือแบบต่อเนื่อง (Continuous) ของฟังก์ชันคาบ ซึ่งเป้าหมายนี้ ถ้าเป็นฟังก์ชันคาบ $f(t)$ ที่เกิดขึ้นตลอดช่วงทั้งหมด $(-\infty < t < \infty)$ ซึ่งสามารถแทนให้อยู่ในรูปผลรวมแบบดิสครีตของฟังก์ชันคาบ และถ้าฟังก์ชันไม่เป็นคาบ (Nonperiodic Function) เราก็สามารถแทนให้อยู่ในเทอมของฟังก์ชันคาบ ในช่วงที่จำกัด $(t_0 < t < t_0 + T)$ ทั้งนี้สำหรับจุดมุ่งหมายของการวิเคราะห์ความถี่นั้น เราต้องการที่จะแทนฟังก์ชันอินพุตทุก ๆ ฟังก์ชันในเทอมของฟังก์ชันคาบที่เกิดขึ้นตลอดช่วงทั้งหมด ไม่ใช่ในช่วงที่จำกัด

คุณสมบัติของการแปลงฟูรีเยร์เสมือนเป็นเครื่องมืออันหนึ่ง ซึ่งนำมาใช้สำหรับแสดง หรือแทนฟังก์ชันให้อยู่ในเทอมของความถี่ต่างๆ ถ้าฟังก์ชัน $f(t)$ สามารถหาแปลงฟูรีเยร์ $F(\omega)$ ได้ แล้วในทางกลับกันถ้าทราบ $F(\omega)$ ก็จะสามารถจะแปลงกลับให้เป็น $f(t)$ ได้ซึ่งสามารถเขียนแทนได้ว่า

$$f(t) \leftrightarrow F(\omega) \quad (2-2)$$

การแปลงฟูรีเยร์ให้ค่าของตัวบรรยายลักษณะฟูรีเยร์ที่มีความคงทนต่อการเลื่อนจุด การหมุนและการย่อหรือขยาย

2. ลักษณะของฟูรีเยร์ (Fourier Features)

ลักษณะของฟูรีเยร์ใช้สำหรับพรรณนารูปร่างลักษณะ ขอบเขตของวัตถุในภาพ กำหนดให้ (x_k, y_k) โดยที่ $k = 0, 1, \dots, N-1$ คือพิกัดจำนวน N ค่าของขอบเขตพื้นที่ใด ๆ ซึ่งแต่ละพิกัด (x_k, y_k) เป็นจำนวนเชิงซ้อน (Complex Number) นั่นคือ

$$u_k = x_k + iy_k \quad \text{โดยที่ } i \text{ คือ } \sqrt{-1} \quad (2-3)$$

ดังนั้นจะได้มาซึ่งดีสครีตฟูรีเยร์ทรานส์ฟอร์ม (DFT : Discrete Fourier Transform)

$$f_l = \sum_{k=0}^{N-1} u_k \exp\left(-j \frac{2\pi}{N} lk\right) \quad l = 0, 1, \dots, N-1 \quad (2-4)$$

ผลของการแปลงฟูรีเยร์ ดังแสดงในสมการที่ (2-4) จะได้เวกเตอร์ของค่าสัมประสิทธิ์ (Coefficient) 1 ชุด เวกเตอร์เป็นเวกเตอร์แบบ 1 มิติ ขนาด $1 \times n$ โดยที่สามารถกำหนดขนาดของ n ได้

2.2 วิธีการจำแนกที่ใช้ในการศึกษา

วิธีการจำแนกแบบไม่ใช้พารามิเตอร์ (nonparametric discriminant) โดยทั่วไปจะใช้กับข้อมูลที่มีการแจกแจงไม่เป็นแบบปกติพหุ หรือเมื่อไม่ทราบการแจกแจงของข้อมูล การจำแนกแบบไม่ใช้พารามิเตอร์บางวิธีจะสร้างกฎการจำแนกจากฟังก์ชันความหนาแน่นความน่าจะเป็นของแต่ละกลุ่มที่ประมาณด้วยวิธีต่างๆ เช่น วิธีเคอร์เนล (kernel) วิธีเพื่อนบ้านใกล้เคียงที่สุด k ค่า (K-Nearest neighbour, K-NN) รูปโครงสร้างต้นไม้ (tree-structured classification) เป็นต้น

1. วิธีเคอร์เนล (kernel method)

เป็นวิธีหนึ่งที่ใช้การประมาณค่าฟังก์ชันความหนาแน่นความน่าจะเป็นแบบไม่ใช้พารามิเตอร์ของประชากรแต่ละกลุ่มเพื่อนำไปสร้างกฎการจำแนก วิธีการคือกำหนดฟังก์ชันความหนาแน่น $K(\cdot)$ ที่เรียกว่า ฟังก์ชันความหนาแน่นเคอร์เนล (kernel density) ให้มีศูนย์กลางอยู่ที่แต่ละค่าสังเกต x_i และประมาณค่าฟังก์ชันความหนาแน่นความน่าจะเป็นที่ x ใดๆ ด้วยค่าเฉลี่ยความหนาแน่นความน่าจะเป็นของค่าสังเกต x_i ทุกจุด ค่าสังเกต x_i ที่อยู่ใกล้ x จะมีน้ำหนักและส่งผลมากถึงฟังก์ชันความหนาแน่นความน่าจะเป็นของ x

ระยะห่างระหว่างค่าสังเกตวัดโดยใช้ระยะทาง Mahalanobis หรือระยะทาง Euclidean สำหรับระยะทาง Mahalanobis อาจคำนวณจากเมทริกซ์ความแปรปรวนร่วมรวม (pooled covariance matrix) หรือ เมทริกซ์ความแปรปรวนร่วมภายในแต่ละกลุ่ม (within-group covariance matrix) นั่นคือ ระยะทางกำลังสองระหว่างเวกเตอร์ของค่าสังเกตสองเวกเตอร์ x และ y ใดๆในกลุ่ม g คือ

$$d_g^2(x, y) = (x - y)'C^{-1}(x - y)$$

เมื่อ C แทน เมทริกซ์ความแปรปรวนร่วมรวม หรือ เมทริกซ์ความแปรปรวนร่วมภายในแต่ละกลุ่ม

ระยะทาง Euclidean คือ

$$d_g^2(x, y) = (x - y)'(x - y)$$

การประมาณฟังก์ชันความหนาแน่นเคอร์เนล (Hand, 1981, 1982; Silverman, 1986) จะได้ว่า

$$\hat{f}_g(x) = \frac{1}{n_g h_g^p} \sum_{i=1}^{n_g} K_N[(x - X_i)/h_g], \quad h_g > 0$$

ให้	X_i	แทนค่าสังเกตของเวกเตอร์ที่มีมิติ p
	h_g	แทนค่าราบเรียบที่เรียกว่า bandwidth ของกลุ่ม g
	n_g	แทนค่าขนาดตัวอย่างค่าสังเกตของเวกเตอร์ของกลุ่ม g
	p	แทนค่าขนาดของมิติ

ในการใช้วิธีเคอร์เนลเพื่อประมาณฟังก์ชันความหนาแน่นความน่าจะเป็นของค่าสังเกต x ในประชากรกลุ่มที่ g จะต้องหาค่า h ของขอบเขตค่าสังเกตที่อยู่รอบ x ซึ่งค่า h เรียกว่า ค่าราบเรียบ (bandwidth) ซึ่งเป็นตัวกำหนดระดับของความไม่สม่ำเสมอในการประมาณฟังก์ชันความหนาแน่น

เมื่อ	ค่าราบเรียบมีค่าเข้าใกล้ 0	ลักษณะโครงสร้างของข้อมูลจะมีลักษณะกว้าง ไปมาทำให้ไม่แสดงถึงลักษณะของข้อมูล แท้จริง
	ค่าราบเรียบมีค่าเล็ก	ลักษณะโครงสร้างของข้อมูลจะปรับเรียบ
	ค่าราบเรียบมีค่าใหญ่	ลักษณะโครงสร้างของข้อมูลจะปรับเรียบมาก ขึ้น
	ค่าราบเรียบมีค่าเข้าใกล้อนันต์	ลักษณะโครงสร้างของข้อมูลจะแบนราบ

ฟังก์ชันความหนาแน่นความน่าจะเป็นเคอร์เนลที่เลือกใช้ในงานวิจัย (Epanechnikov, 1969) คือ

- เคอร์เนลปกติ (Normal kernel)

ฟังก์ชันความหนาแน่นความน่าจะเป็นเคอร์เนลของกลุ่ม g คือ

$$K_N[(x - X_i) / h_g] = (2\pi)^{-p/2} |C_g|^{-1/2} \exp(-0.5(h_g^{-2} d_g^2))$$

เมื่อ $d_g^2 = (x - X_i)' C_g^{-1} (x - X_i), \quad i = 1, 2, \dots, n_g$

โดย
$$h_g = \left(\frac{4 / (2m + 1)}{n_g} \right)^{1/(m+4)}$$

ให้ X_i แทนค่าสังเกตของเวกเตอร์

h_g แทนค่าราบเรียบที่เรียกว่า bandwidth ของกลุ่ม g

n_g แทนค่าขนาดตัวอย่างค่าสังเกตของเวกเตอร์ของกลุ่ม g

m แทนค่าจำนวนตัวแปรค่าสังเกตของเวกเตอร์

วิธี Least square cross-validation (LSCV) และวิธี Likelihood cross-validation (LKCV) สองวิธีนี้มีหลักการในการหาค่าราบเรียบ (h) ที่เหมือนกัน แต่แตกต่างกันที่ฟังก์ชันของแต่ละวิธี โดยการหาค่าราบเรียบที่เหมาะสมของสองวิธีนี้ จะทำให้ฟังก์ชันของแต่ละวิธีมีค่าน้อยที่สุด

- Least square cross-validation (LSCV) (Silverman, 1986, Worton, 1995)

$$LSCV(h_g) = \frac{1}{n_g^2 h_g^p} \sum_i \sum_j K^*[(X_i - X_j) / h_g] + \frac{2}{n_g h_g^p} K(0), \quad h_g > 0$$

$$\text{เมื่อ} \quad K^*[(X_i - X_j) / h_g] = K^*(z) = K^{(2)}(z) - 2K(z),$$

$$K(z) = (2\pi)^{-p/2} \exp(-z'z / 2), \quad K(0) = (2\pi)^{-p/2}$$

$$\text{และ} \quad K^{(2)}(z) = \left(\frac{1}{2\sqrt{\pi}} \right)^p \exp(-z'z / 4)$$

ให้ X แทนค่าสังเกตของเวกเตอร์ที่มีมิติ p

h_g แทนค่าราบเรียบที่เรียกว่า bandwidth ของกลุ่ม g

n_g แทนค่าขนาดตัวอย่างค่าสังเกตของเวกเตอร์ของกลุ่ม g

- Likelihood cross-validation (LKCV) (Silverman, 1986)

$$LKCV(h_g) = \frac{1}{n_g} \sum_{i=1} \log \left[\frac{1}{(n_g - 1)h_g^p} \sum_{j=1} K \left(\frac{X_i - X_j}{h_g} \right) \right], \quad h_g > 0$$

ให้ X แทนค่าสังเกตของเวกเตอร์ที่มีมิติ p

h_g แทนค่าราบเรียบที่เรียกว่า bandwidth ของกลุ่ม g

n_g แทนค่าขนาดตัวอย่างค่าสังเกตของเวกเตอร์ของกลุ่ม g

p แทนค่าขนาดของมิติ

จากฟังก์ชันความหนาแน่นความน่าจะเป็นของแต่ละกลุ่มที่ประมาณขึ้นจะนำมาวิเคราะห์ เพื่อหาการตัดสินใจในการจำแนกกลุ่ม ซึ่งจะคำนวณหาความน่าจะเป็นภายหลัง (posterior probability) ที่ค่าสังเกต x จะเป็นสมาชิกกลุ่มที่ g ได้จาก

$$p(g|x) = \frac{p_g f_g(x)}{\sum_{g=1}^n p_g f_g(x)} \quad \text{เมื่อ } p_g \text{ เป็นค่าความจะเป็นก่อนหน้า (prior probability)}$$

การจำแนกเวกเตอร์ของค่าสังเกต x จะพิจารณาจากค่าความน่าจะเป็นภายหลัง ดังนี้

$$\text{จัด } x \text{ ให้อยู่ในกลุ่มที่ } g \text{ ถ้า } p(g|x) = \max_l p(l|x)$$

2. วิธี K-nearest neighbour

K-Nearest neighbour (K-NN) คือ วิธีการจัดแบ่งกลุ่มวิธีหนึ่ง ซึ่งเทคนิคนี้จะตัดสินใจว่า กลุ่มใดที่จะแทนเงื่อนไขหรือกรณีใหม่ๆ ได้บ้าง โดยการตรวจสอบจำนวน K ("K" ใน K-nearest neighbour) ของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด

วิธี K-NN เป็นวิธีการวัดค่าระยะห่างระหว่างแต่ละคุณลักษณะของข้อมูลของสองเวกเตอร์ x และ y ใดๆในกลุ่ม g ซึ่งสามารถคำนวณค่าระยะทางได้จาก

2.1 ระยะทาง Euclidean

$$d_g^2(x, y) = (x - y)'(x - y)$$

2.2 ระยะทาง Cityblock

$$d_g(x, y) = \sum_{i=1}^n |x_i - y_i|$$

2.3 ระยะทาง Correlation

$$d_g(x, y) = 1 - \frac{(x - \bar{x})(y - \bar{y})'}{\sqrt{(x - \bar{x})(x - \bar{x})'} \sqrt{(y - \bar{y})(y - \bar{y})'}}$$

$$\text{เมื่อ } \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{และ} \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

2.4 ระยะทาง Cosine

$$d_g(x, y) = \left(1 - \frac{xy'}{\sqrt{(xx')(yy')}} \right)$$

- ข้อดีของ K-NN เป็นวิธีง่ายในการจำแนกกลุ่ม แต่ก็มีข้อเสีย คือ
 - 1. ใช้ระยะเวลาในการคำนวณนาน เพราะต้องคำนวณระยะทางระหว่างค่าสังเกตที่ต้องการจำแนกกลุ่มกับทุกค่าข้อมูลที่มีอยู่
 - 2. ถ้าคุณลักษณะของข้อมูลมีจำนวนมากจะเกิดปัญหาในการคำนวณค่า

2.3 เกณฑ์ที่ใช้ในการพิจารณาประสิทธิภาพในการจำแนก

อัตราความผิดพลาดของการจำแนก (Error rate of misclassification)

ความผิดพลาดของการรู้จำรูปแบบเกิดขึ้นเมื่อรูปแบบหนึ่งถูกจำแนกให้เข้าอยู่ในกลุ่มที่ไม่ตรงกับรูปแบบทางกายภาพที่แท้จริงของกลุ่มนั้น กฎการจำแนกกลุ่มจะดีหรือไม่จึงจะประเมินจากประสิทธิภาพในการจำแนกค่าสังเกตโดยพิจารณาจากอัตราความผิดพลาดในการจำแนกกลุ่มหรือความน่าจะเป็นในการจำแนกกลุ่มผิดของกฎนั้น การประมาณค่าอัตราความผิดพลาดมีหลายแบบที่นิยมใช้คือ อัตราความผิดพลาดที่เห็นชัด (apparent error rate, APER) สามารถคำนวณได้จาก

$$\text{APER} = \frac{\sum_{i=1}^g n_{im}}{\sum_{i=1}^g n_i}$$

เมื่อ n_{im} แทนจำนวนตัวอักษรทั้งหมดที่ถูกจำแนกผิดเป็นกลุ่มอื่น
 n_i แทนจำนวนตัวอักษรทั้งหมดในการทดสอบ

APER เป็นค่าประมาณอัตราความผิดพลาดที่คำนวณได้ง่าย แต่เป็นค่าประมาณที่เอนเอียง เนื่องจากเป็นอัตราความผิดพลาดที่คำนวณจากชุดข้อมูลที่นำมาสร้างกฎ ค่าประมาณอัตราความผิดพลาดจึงต่ำกว่าอัตราความผิดพลาดที่แท้จริง ค่าประมาณอัตราความผิดพลาดที่ดีกว่าคือ อัตราความผิดพลาดในตัวอย่างทดสอบ กล่าวคือ แบ่งตัวอย่างค่าสังเกตที่รวบรวมมาออกเป็น 2 ส่วน ส่วนหนึ่งเป็นตัวอย่างการเรียนรู้ซึ่งเป็นชุดข้อมูลที่ใช้ในการเรียนรู้ และสร้างกฎการจำแนกกลุ่ม อีกส่วนหนึ่งเป็นตัวอย่างทดสอบที่จะนำมาใช้ในการประเมินประสิทธิภาพของกฎการจำแนกกลุ่มที่สร้างจากตัวอย่างการเรียนรู้ โดยคำนวณอัตราการจำแนกกลุ่มผิดพลาดที่เกิดขึ้นในตัวอย่างทดสอบเมื่อใช้กฎการจำแนกที่สร้างขึ้นนั้น

บทที่ 3

วิธีการดำเนินการศึกษา

การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบการรู้จำลายมือเขียนตัวอักษรไทยในการจำแนกกลุ่มด้วยวิธี K-nearest neighbour กับวิธี Kernel โดยในการเปรียบเทียบจากการศึกษานี้ จะพิจารณาจากตัวบรรยายลักษณะฟูเรียร์จากการแปลงฟูเรียร์ เนื่องจากตัวบรรยายลักษณะฟูเรียร์จะได้ค่าสัมประสิทธิ์ฟูเรียร์ที่มีความคงทนต่อการเลื้อจุด การหมุน และการย่อหรือการขยาย และใช้ค่าอัตราความผิดพลาดในการจำแนกกลุ่มเป็นเกณฑ์ในการเปรียบเทียบ เพื่อศึกษาว่าวิธีการใดมีความผิดพลาดในการจำแนกกลุ่มน้อยที่สุด

การศึกษานี้ทำการเก็บข้อมูลและพิจารณาข้อมูลโดยใช้โปรแกรม Matlab เวอร์ชัน 7.12.0.635 (R2011a) ในการทำการศึกษากายใต้ขอบเขตการณศึกษาต่อไปนี้

3.1 ขอบเขตของการศึกษา

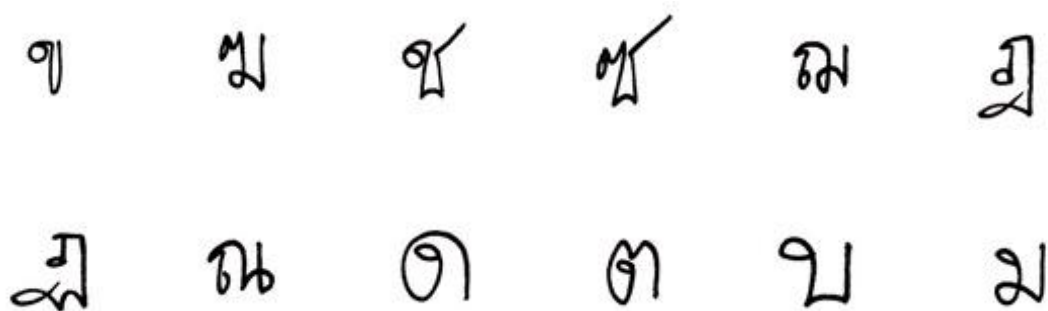
ในการวิจัยครั้งนี้จะทำการศึกษากายใต้ขอบเขตดังนี้

1. เก็บบันทึกข้อมูลลายมือเขียนตัวอักษรไทย คือ ข, ชม, ซ, ซ, ฉ, ฎ, ฏ, ฌ, ด, ต, บ และ ม โดยแต่ละตัวอักษรจะเขียนจำนวน 90 ครั้งจำนวน 2 ชุด คือ ชุดที่ 1 จากลายมือเขียนตัวอักษรไทย 1 คน และชุดที่ 2 จากลายมือเขียนตัวอักษรไทย 10 คนในตารางขนาด 1×1 นิ้วต่อ 1 ตาราง และทำการสแกนโดยเครื่องสแกน แล้วเก็บผลลัพธ์เป็นแฟ้มข้อมูลรูปภาพ 1 ตัวอักษรต่อ 1 แฟ้มข้อมูล
2. นำตัวอักษรไทยมาทำการแปลงฟูเรียร์โดยกำหนดค่าสัมประสิทธิ์ที่ได้คือ 200, 500 และ 1,000
3. ทำการสุ่มตัวอักษรแต่ละตัวอักษรไทยแบบไม่ใส่คืนจำนวน 30 ตัวจากตัวอักษร 90 ตัวของแต่ละตัวอักษรไทย เพื่อเป็นตัวอย่างทดสอบ และตัวอักษรที่เหลือแต่ละตัวอักษรไทยจำนวน 60 ตัวจะเก็บเป็นตัวอย่างเรียนรู้

4. จากตัวอย่างเรียนรู้ตัวอักษรที่ศึกษา ซึ่งแต่ละตัวอักษรไทยที่ศึกษามี 60 ตัว จะทำการสุ่มจำนวน 10, 20, 30, 40, 50 และ 60 เพื่อเป็นตัวอย่างเรียนรู้
5. นำตัวอย่างเรียนรู้และตัวอย่างทดสอบมาทำการจำแนกกลุ่ม 2 วิธี คือ
 - 5.1 วิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทาง 4 แบบ คือ
 - 5.1.1 ระยะทางแบบ Euclidean
 - 5.1.2 ระยะทางแบบ Cityblock
 - 5.1.3 ระยะทางแบบ Correlation
 - 5.1.4 ระยะทางแบบ Cosine
 - 5.2 วิธี Kernel 3 แบบ คือ
 - 5.2.1 Normal kernel
 - 5.2.2 Least square cross-validation (LSCV)
 - 5.2.3 Likelihood cross-validation (LKCV)
6. คำนวณหาค่าอัตราความผิดพลาดในการจำแนก
7. ในการศึกษาครั้งนี้ทำการทดลองซ้ำสำหรับตัวอย่างเรียนรู้ที่แตกต่างกัน 3 ครั้ง

3.2 แหล่งที่มาของข้อมูล

รูปภาพตัวอักษร คือข้อมูลที่นำมาใช้ทดสอบการรู้จำลายมือเขียนตัวอักษรไทย ได้มาจากลายมือเขียนซึ่งเขียนลงบนกระดาษในตารางขนาด 1×1 นิ้วต่อ 1 ตาราง และผ่านเครื่องสแกนที่มีความละเอียดแบบ Gray Scale 256 นำมาเก็บไว้เป็นข้อมูลต้นแบบ โดยเก็บแต่ละตัวอักษรลงแฟ้มข้อมูลโดย 1 ตัวอักษรต่อ 1 แฟ้มข้อมูล ข้อมูลในรูปแบบ .jpg โดยตัวอักษรที่ใช้สำหรับการทดสอบประกอบด้วย ข, ส, ช, ซ, ฉ, ฎ, ฏ, ณ, ด, ต, บ และ ม โดยตัวอักษรที่ใช้สำหรับการทดสอบนั้นมีขนาด 64×64 pixel



ภาพที่ 3.1 ข้อมูลต้นแบบของลายมือเขียนตัวอักษรไทย

3.3 ขั้นตอนวิธีการรู้จำ

3.3.1 ขั้นตอน Pre-processing

เป็นกระบวนการประมวลผลภาพเพื่อปรับปรุงคุณภาพของรูปภาพตัวอักษรก่อนที่จะนำรูปภาพมาทำการเรียนรู้

3.3.1.1 การปรับปรุงคุณภาพของรูปภาพตัวอักษร

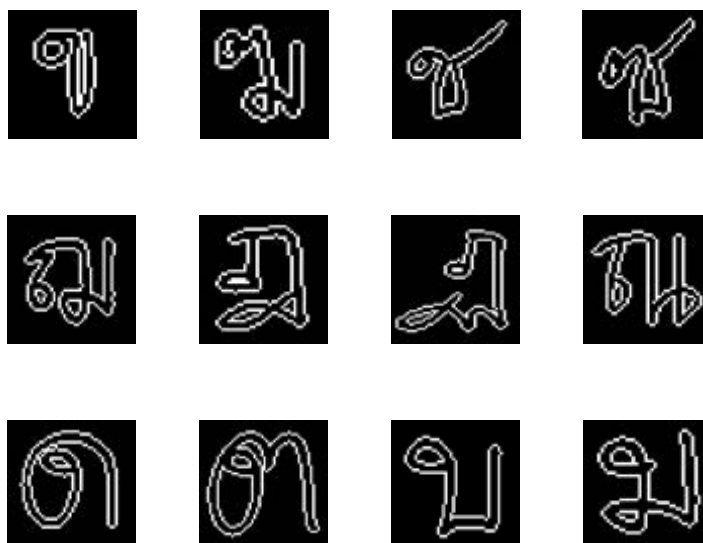
นำรูปภาพของลายมือเขียนตัวอักษรไทยผ่านเครื่องสแกน จากนั้นทำการแปลงรูปภาพตัวอักษรที่ได้จากการสแกนจากโหมด Gray Scale 256 ให้เป็นภาพโหมดขาว-ดำ มีค่าเป็น 0 กับ 1

โดย 0 หมายถึง เนื้อภาพ (Object)

1 หมายถึง พื้นหลัง (Background)

3.3.1.2 การหาขอบของภาพ

หลังจากที่ทำการปรับปรุงคุณภาพของรูปภาพตัวอักษรได้ภาพในโหมดขาว-ดำ จากนั้นจะทำการหาขอบของภาพ ได้ดังภาพที่ 3.2



ภาพที่ 3.2 ตัวอย่างการหาขอบของรูปภาพแต่ละตัวอักษร เมื่อผ่านกระบวนการปรับปรุงคุณภาพ

3.3.2 ขั้นตอน Feature Extraction

เป็นกระบวนการค้นหาคุณลักษณะพิเศษที่แฝงอยู่ในรูปภาพของแต่ละตัวอักษร โดยใช้ตัวบรรยายลักษณะฟูเรียร์ ซึ่งการแปลงฟูเรียร์จะส่งให้ค่าของตัวบรรยายลักษณะฟูเรียร์มีความคงทนต่อการเลื่อนจุด การหมุน และการย่อหรือขยาย

$$f_l = \sum_{k=0}^{N-1} u_k \exp\left(-j \frac{2\pi}{N} lk\right) \quad l = 0, 1, \dots, N-1 \quad (3-1)$$

จากสมการที่ (3-1) ทำให้ได้มาซึ่งตัวบรรยายลักษณะฟูเรียร์ ซึ่งเป็นผลของการแปลงฟูเรียร์จะได้เวกเตอร์ของค่าสัมประสิทธิ์ (Coefficient) 1 ชุด เวกเตอร์เป็นเวกเตอร์แบบ 1 มิติ ขนาด $1 \times n$ โดยที่สามารถกำหนดขนาดของ n ได้

ตารางที่ 3.1 แสดงตัวอย่างค่าสัมประสิทธิ์ 10 ค่าที่ได้จากการแปลงฟูเรียร์ของตัวอักษร “ข”

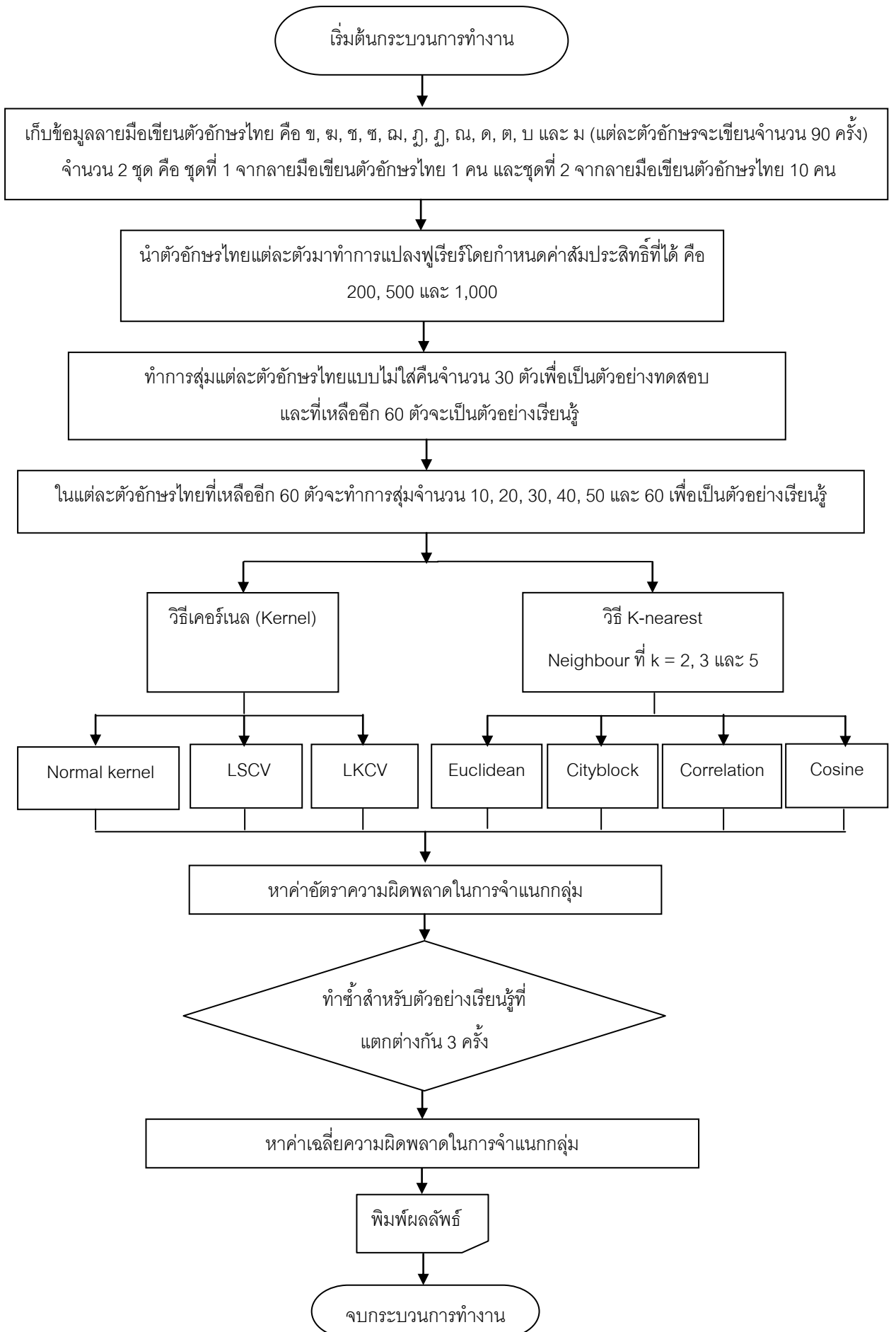
0.4173	0.1429	0.1356	0.0753	0.0279	0.0429	0.0553	0.0310	0.0299	0.0417
--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

3.4 ขั้นตอนในการดำเนินการศึกษา

1. ศึกษาค้นคว้าเอกสารต่างๆทั้งทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการจำแนกกลุ่ม
 - วิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระหว่างระยะทาง 4 แบบ คือ
 1. ระยะทางแบบ Euclidean
 2. ระยะทางแบบ Cityblock
 3. ระยะทางแบบ Correlation
 4. ระยะทางแบบ Cosine
 - วิธี Kernel 3 แบบ คือ
 1. Normal kernel
 2. Least square cross-validation (LSCV)
 3. Likelihood cross-validation (LKCV)
 - ค่าอัตราความผิดพลาดในการจำแนกกลุ่ม
2. เก็บข้อมูลลายมือเขียนตัวอักษรไทยตามขนาดตัวอย่างที่กำหนด
3. นำข้อมูลลายมือเขียนตัวอักษรไทยมาประมวลผล ซึ่งแบ่งได้เป็น 2 ขั้นตอน คือ
 - ขั้นตอนที่ 1 Pre-Processing เป็นกระบวนการปรับปรุงคุณภาพและหาขอบของภาพ
 - ขั้นตอนที่ 2 Feature Extraction เป็นกระบวนการค้นหาคุณลักษณะพิเศษที่แฝงอยู่ในรูปภาพตัวอักษร โดยใช้ตัวบรรยายลักษณะฟูเรียร์ (Fourier Descriptors)
4. ทำการสุ่มขนาดตัวอย่างแบบไม่ใส่คืนเพื่อเป็นตัวอย่างทดสอบและตัวอย่างเรียนรู้ตามที่กำหนด
5. คำนวณหาค่าเฉลี่ยอัตราความผิดพลาดในการจำแนกกลุ่มด้วยวิธี K-nearest neighbour และวิธี Kernel
6. วิเคราะห์และสรุปผลการเปรียบเทียบวิธีการที่ใช้ในการวิจัย

3.5 ขั้นตอนการทำงานของโปรแกรม

โปรแกรมที่ใช้ในการศึกษาการวิจัยในครั้งนี้เขียนด้วยโปรแกรม Matlab เวอร์ชัน 7.12.0.635 (R2011a) ซึ่งในการศึกษาครั้งนี้ทำการทดลองซ้ำสำหรับตัวอย่างเรียนรู้ที่แตกต่างกัน 3 ครั้ง สามารถแสดงขั้นตอนการทำงานของโปรแกรมได้ดังนี้



บทที่ 4

ผลการวิเคราะห์ข้อมูล

จากการศึกษาเปรียบเทียบประสิทธิภาพในการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย คือ ข, ชม, ช, ซ, ฉ, ฎ, ฏ, ณ, ด, ต, บ และ ม โดยทำการศึกษาลายมือเขียนตัวอักษรไทยตัวละ 90 ตัวอักษรจำนวน 2 ชุด คือ ชุดที่ 1 จากลายมือเขียนตัวอักษรไทย 1 คน และชุดที่ 2 จากลายมือเขียนตัวอักษรไทย 10 คน ซึ่งในการพิจารณาลายมือเขียนตัวอักษรไทย จะทำการพิจารณาลักษณะตัวอักษรโดยใช้การแปลงฟูเรียร์ (Fourier transform) ซึ่งจะได้ค่าสัมประสิทธิ์ (coefficient) จำนวนค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ที่ใช้ในการศึกษา ได้แก่ 200, 500 และ 1,000 โดยแต่ละตัวอักษรจะแบ่งข้อมูลเป็น 2 กลุ่ม คือ ตัวอย่างเรียนรู้นี้ขนาด 60 ตัวอักษร และตัวอย่างทดสอบขนาด 30 ตัวอักษร โดยตัวอย่างเรียนรู้นี้แบ่งย่อยออกเป็น 6 กลุ่ม คือ 10, 20, 30, 40, 50 และ 60 โดยการศึกษาครั้งนี้จะใช้วิธีในการจำแนกกลุ่ม 2 วิธี คือ วิธีเพื่อนบ้านใกล้เคียงที่สุด k ค่า (K-nearest neighbour method) และ วิธีเคอร์เนล (Kernel method) ซึ่งวิธี K-nearest neighbour จะทำการพิจารณาค่า k 3 ค่า คือ 2, 3, และ 5 และพิจารณาระยะห่างระยะทาง 4 แบบ คือ Euclidean, Cityblock, Correlation และ Cosine ส่วนวิธี Kernel จะทำการพิจารณา 3 แบบ คือ Normal kernel, Least squares cross-validation (LSCV) และ Likelihood cross-validation (LKCV) จากการศึกษาครั้งนี้ เกณฑ์ที่ใช้ในการพิจารณาเปรียบเทียบ คือ ค่าอัตราความผิดพลาดของการจำแนก (Error rate of misclassification)

4.1 ผลการศึกษาจากลายมือเขียนตัวอักษรไทย 1 คน

ตารางที่ 4.1 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	2	1	5	2	1	2	3	3	5	2	4	2	2	3	2	2	2	2
ฃ	0	2	2	2	2	2	2	2	1	2	1	2	2	2	2	2	2	2
ช	2	6	6	4	1	2	9	1	3	6	1	2	4	1	2	2	2	2
ฅ	12	8	13	10	10	10	7	10	8	9	9	9	8	9	9	9	9	9
ฉ	13	8	4	6	8	8	6	7	4	8	7	5	5	5	7	7	7	7
ฐ	1	12	2	2	2	2	1	1	1	0	0	0	1	0	0	0	0	0
ฎ	0	0	2	0	0	1	5	0	1	3	1	2	2	1	1	1	1	1
ณ	9	11	10	9	8	8	13	7	9	9	8	9	9	8	7	7	7	7
ด	9	1	5	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1
ต	9	7	13	7	5	5	6	5	8	2	5	2	8	5	5	5	5	5
บ	1	1	3	2	1	1	4	1	2	1	2	2	2	2	2	2	2	2
ม	4	2	3	4	3	3	4	4	3	2	3	2	2	3	2	2	2	2
Total	62	59	68	49	42	45	61	42	46	46	42	38	46	40	40	40	40	40
%	17.22	16.39	18.89	13.61	11.67	12.50	16.94	11.67	12.78	12.78	11.67	10.56	12.78	11.11	11.11	11.11	11.11	11.11
Mean (%)	17.50			12.59			13.80			11.67			11.67			11.11		

ตารางที่ 4.2 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	2	2	5	2	2	2	3	2	5	2	3	2	2	2	2	1	1	1
ฃ	2	1	1	2	1	1	3	0	2	1	1	1	1	1	1	1	1	1
ช	2	5	7	4	3	3	8	1	5	5	3	4	5	2	3	2	2	2
ฅ	16	8	14	11	7	7	10	8	9	8	7	8	7	7	7	7	7	7
ฉ	11	10	15	9	9	9	3	8	6	4	6	3	2	2	3	3	3	3
ฐ	6	14	2	4	7	7	1	1	1	0	0	0	1	0	0	0	0	0
ฎ	0	0	2	1	0	0	1	0	0	1	0	1	1	1	1	1	1	1
ณ	9	13	11	11	11	11	11	10	10	7	10	7	11	10	9	10	10	10
ด	9	2	3	1	1	1	0	1	0	3	1	1	1	1	1	1	1	1
ต	10	12	12	13	12	12	4	11	7	1	6	2	5	6	5	5	5	5
บ	1	1	2	1	1	1	2	1	2	0	1	1	2	1	0	1	1	1
ม	6	3	3	4	3	3	6	4	3	4	3	4	4	3	3	3	3	3
Total	74	71	77	63	57	57	52	47	50	36	41	34	42	36	35	35	35	35
%	20.56	19.72	21.38	17.50	15.83	15.83	14.44	13.05	13.89	10.00	11.38	9.44	11.67	10.00	9.72	9.72	9.72	9.72
Mean (%)	20.55			16.39			13.79			10.27			10.46			9.72		

ตารางที่ 4.3 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	2	1	5	3	0	0	4	2	5	3	3	2	4	2	3	3	3	3
ฃ	2	3	4	1	1	1	3	1	1	1	1	0	1	1	1	1	1	1
ช	6	5	6	6	3	3	5	3	7	4	6	5	6	5	6	6	6	6
ฅ	17	8	15	8	8	8	10	9	9	10	8	11	9	8	7	8	8	8
ฉ	8	8	14	7	6	6	4	5	7	4	7	4	4	3	4	5	5	5
ง	7	13	2	5	8	8	1	4	1	1	1	1	1	1	1	1	1	1
จ	0	0	2	1	0	0	3	0	0	0	0	0	1	0	0	0	0	0
ฉ	12	13	12	10	11	11	16	9	12	8	11	8	11	9	10	10	10	10
ด	9	1	3	1	2	2	1	2	1	2	1	2	1	1	1	1	1	1
ต	10	14	10	11	12	12	6	11	6	2	5	3	6	5	5	5	5	5
บ	1	2	1	1	1	1	2	1	2	2	1	2	2	1	1	1	1	1
ม	3	4	4	4	3	3	7	4	4	3	4	4	2	4	3	2	2	2
Total	77	72	78	58	55	55	62	51	55	40	48	42	48	40	42	43	43	43
%	21.39	20.00	21.67	16.11	15.28	15.28	17.22	14.17	15.28	11.11	13.33	11.67	13.33	11.11	11.67	11.94	11.94	11.94
Mean (%)	21.02			15.56			15.56			12.04			12.04			11.94		

ตารางที่ 4.4 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	1	4	7	4	2	2	2	1	4	2	1	2	2	2	2	2	2	2
ฃ	1	5	2	4	3	3	2	3	1	1	1	1	1	1	1	1	1	1
ช	0	4	3	3	1	1	11	1	3	6	2	3	3	2	4	2	2	2
ฅ	14	7	8	8	9	9	8	8	7	8	7	9	8	7	7	8	8	8
ฉ	10	10	13	7	8	8	8	7	7	9	7	8	7	7	8	8	8	8
ฐ	3	17	5	8	7	7	2	7	5	2	5	2	3	4	4	3	3	3
ฎ	1	0	3	1	0	0	5	0	0	4	0	2	2	1	2	1	1	1
ณ	15	14	10	10	11	11	14	11	9	11	9	10	9	8	8	9	9	9
ด	12	3	3	0	0	0	3	0	0	4	0	3	1	1	1	1	1	1
ต	8	10	10	9	8	8	10	7	9	6	7	5	9	7	8	7	7	7
บ	2	2	3	3	3	3	5	4	3	3	4	3	2	3	3	3	3	3
ม	6	4	5	3	4	4	5	1	2	4	1	4	3	2	3	2	2	2
Total	73	80	72	60	56	56	75	50	50	60	44	52	50	45	51	47	47	47
%	20.28	22.22	20.00	16.67	15.56	15.56	20.83	13.89	13.89	16.67	12.22	14.44	13.89	12.50	14.17	13.06	13.06	13.06
Mean (%)	20.83			15.93			16.20			14.44			13.52			13.06		

ตารางที่ 4.5 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	3	7	1	0	0	5	1	6	1	1	1	2	1	1	0	0	0
ฃ	0	3	0	2	2	2	2	1	0	0	0	0	0	0	1	0	0	0
ช	1	4	5	6	1	1	8	1	6	5	3	4	7	3	4	3	3	3
ฌ	18	3	13	8	10	10	8	9	6	9	7	9	7	7	6	7	7	7
ฎ	11	11	14	9	9	9	5	7	7	5	7	6	5	5	5	5	5	5
ฎ	7	17	3	9	10	10	2	6	4	1	3	2	2	3	2	2	2	2
ฏ	1	0	3	1	0	0	3	0	1	1	0	1	2	1	1	1	1	1
ณ	14	13	10	10	14	14	16	11	10	11	9	9	10	9	8	9	9	9
ด	11	2	3	0	1	1	2	1	0	5	1	3	0	1	1	1	1	1
ต	8	12	10	9	10	10	8	9	7	3	8	1	7	8	9	8	8	8
บ	2	2	3	3	1	1	4	1	3	1	2	3	3	2	1	2	2	2
ม	5	4	5	4	5	5	6	5	4	6	4	3	3	3	4	3	3	3
Total	78	74	76	62	63	63	69	52	54	48	45	42	48	43	43	41	41	41
%	21.67	20.56	21.11	17.22	17.50	17.50	19.17	14.44	15.00	13.33	12.50	11.67	13.33	11.94	11.94	11.39	11.39	11.39
Mean (%)	21.11			17.41			16.20			12.50			12.40			11.39		

ตารางที่ 4.6 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	2	6	2	1	1	2	1	3	0	0	0	3	0	0	0	0	0
ฃ	3	5	0	1	1	1	2	0	1	0	0	0	1	0	1	0	0	0
ช	2	5	4	2	4	4	6	2	6	4	6	4	5	5	4	5	5	5
ฅ	16	5	14	9	6	6	8	8	9	10	9	9	7	8	9	8	8	8
ฉ	8	11	16	11	8	8	3	9	7	6	7	5	4	4	6	5	5	5
ฐ	9	16	3	6	11	11	1	6	4	1	3	1	1	2	2	1	1	1
ฎ	1	0	2	1	0	0	4	1	1	1	0	1	1	1	1	1	1	1
ณ	15	12	13	9	14	14	19	12	10	13	11	9	12	10	11	11	11	11
ด	9	2	3	1	1	1	1	1	0	2	1	2	0	1	1	1	1	1
ต	9	11	8	9	9	9	7	8	7	3	7	6	6	6	6	6	6	6
บ	1	1	2	3	1	1	3	1	3	1	2	1	3	2	1	2	2	2
ม	5	3	4	3	3	3	6	4	4	5	4	5	4	4	5	5	5	5
Total	78	73	75	57	59	59	62	53	55	46	50	43	47	43	47	45	45	45
%	21.67	20.28	20.83	15.83	16.39	16.39	17.22	14.72	15.27	12.78	13.89	11.94	13.06	11.94	13.06	12.50	12.50	12.50
Mean (%)	20.93			16.20			15.74			12.87			12.69			12.50		

ตารางที่ 4.7 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	5	1	3	1	1	1	6	2	1	1	4	1	2	2	1	1	1
ฃ	3	1	3	3	5	5	1	1	4	1	1	1	0	0	0	0	0	0
ช	1	5	9	6	5	5	11	7	4	10	6	6	8	9	6	8	8	8
ฅ	16	13	6	9	6	6	7	8	7	7	9	7	8	6	7	7	7	7
ฉ	12	13	7	5	8	8	2	4	6	4	3	6	3	5	4	4	4	4
ฐ	3	4	9	5	3	3	3	4	3	3	4	4	3	3	4	3	3	3
ฎ	0	3	0	0	0	0	2	0	0	1	0	0	1	0	0	0	0	0
ณ	11	8	9	10	9	9	15	11	8	8	7	9	11	9	10	9	9	9
ด	14	1	6	2	3	3	1	2	2	3	1	2	2	3	2	2	2	2
ต	7	9	5	5	4	4	14	5	4	0	0	4	4	3	3	3	3	3
บ	3	4	3	2	3	3	7	2	3	3	3	2	2	2	2	2	2	2
ม	0	1	4	4	5	5	8	4	4	4	4	4	3	3	4	3	3	3
Total	80	67	62	54	52	52	62	54	47	45	39	49	46	45	44	42	42	42
%	22.22	18.61	17.22	15.00	14.44	14.44	17.22	15.00	13.06	12.50	10.83	13.61	12.78	12.50	12.22	11.67	11.67	11.67
Mean (%)	19.35			14.63			15.09			12.31			12.50			11.67		

ตารางที่ 4.8 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	1	4	2	0	0	4	1	5	2	3	2	2	2	2	1	1	1
ฃ	2	3	2	3	3	3	0	2	0	0	0	0	0	0	0	0	0	0
ช	2	9	4	6	5	5	8	3	8	6	7	6	7	6	6	6	6	6
ฅ	20	6	15	9	8	8	8	9	8	10	8	11	7	8	8	8	8	8
ฉ	8	7	16	5	7	7	3	8	5	6	5	4	3	3	4	4	4	4
ง	5	12	4	4	4	4	3	3	4	3	4	4	3	4	3	3	3	3
จ	0	0	4	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ฉ	12	12	5	7	8	8	12	8	11	10	9	12	10	9	8	9	9	9
ด	13	3	4	3	1	1	1	1	10	3	1	2	1	1	1	1	1	1
ต	7	8	9	7	8	8	2	7	4	1	4	1	4	4	4	4	4	4
บ	4	2	4	3	3	3	8	2	3	3	3	3	3	3	3	3	3	3
ม	11	3	2	3	5	5	7	3	3	6	3	3	3	4	4	3	3	3
Total	84	66	73	53	52	52	57	47	61	50	47	48	43	44	43	42	42	42
%	23.33	18.33	20.28	14.72	14.44	14.44	15.83	13.06	16.94	13.89	13.06	13.33	11.94	12.22	11.94	11.67	11.67	11.67
Mean (%)	20.65			14.53			15.28			13.43			12.03			11.67		

ตารางที่ 4.9 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	0	5	2	0	0	4	1	5	1	4	2	3	2	1	1	1	1
ฃ	3	2	1	2	2	2	0	2	0	0	0	0	0	0	0	0	0	0
ช	2	7	2	4	3	3	8	2	9	7	7	6	6	7	7	6	6	6
ฅ	19	6	16	10	10	10	10	11	8	11	9	12	8	8	8	8	8	8
ฉ	9	8	19	7	9	9	4	9	5	5	6	4	1	3	4	3	3	3
ฐ	8	13	4	5	6	6	3	4	3	3	2	1	3	3	3	3	3	3
ฎ	0	0	3	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0
ณ	14	7	6	6	12	12	16	7	11	11	8	11	12	10	10	9	9	9
ด	13	4	3	2	4	4	1	2	0	4	1	3	1	1	1	1	1	1
ต	6	8	8	8	8	8	3	7	5	0	4	1	5	4	4	4	4	4
บ	4	3	5	3	4	4	9	3	3	4	3	4	3	4	3	3	3	3
ม	7	5	4	3	3	3	9	3	3	5	3	2	3	3	3	2	2	2
Total	85	63	76	53	61	61	70	51	52	51	47	46	45	45	44	40	40	40
%	23.61	17.50	21.11	14.72	16.94	16.94	19.44	14.17	14.44	14.17	13.06	12.78	12.50	12.50	12.22	11.11	11.11	11.11
Mean (%)	20.74			16.20			16.02			13.34			12.41			11.11		

ตารางที่ 4.10 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	2	4	3	1	1	3	2	6	2	4	2	2	3	3	2	2	2
ฃ	1	2	2	3	2	2	2	3	1	2	1	2	1	1	1	1	1	1
ช	2	6	5	4	2	2	11	2	5	7	2	3	6	2	4	3	3	3
ฅ	14	6	11	9	7	7	6	7	8	7	7	8	8	7	6	7	7	7
ฉ	12	6	13	5	6	6	2	5	3	3	4	3	3	3	3	3	3	3
ฎ	1	10	3	4	2	2	1	2	1	0	0	0	1	0	0	0	0	0
ฏ	0	0	2	0	0	0	3	0	1	2	0	1	1	0	0	0	0	0
ณ	9	11	9	11	10	10	10	9	9	6	7	7	9	8	6	7	7	7
ด	13	2	1	1	1	1	1	1	1	3	1	2	1	1	1	1	1	1
ต	7	6	11	5	4	4	5	4	6	1	4	1	6	4	4	4	4	4
บ	1	2	3	2	1	1	5	2	2	2	2	1	2	2	2	2	2	2
ม	5	3	2	3	3	3	5	3	3	2	3	1	1	2	1	1	1	1
Total	65	56	66	50	39	39	54	40	46	37	35	31	41	37	31	31	31	31
%	18.06	15.56	18.33	13.89	10.83	10.83	15.00	11.11	12.78	10.28	9.72	8.61	11.39	10.28	8.61	8.61	8.61	8.61
Mean (%)	17.32			11.85			12.96			9.54			10.09			8.61		

ตารางที่ 4.11 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	2	4	2	0	0	3	1	5	2	3	2	2	3	2	1	1	1
ฃ	2	2	1	2	1	1	1	0	2	1	1	0	1	1	1	1	1	1
ช	1	7	6	8	4	4	9	4	9	7	7	6	7	7	7	6	6	6
ฅ	20	7	14	10	6	6	9	8	8	9	7	9	7	7	7	7	7	7
ฉ	9	5	16	5	7	7	1	7	5	2	4	3	1	4	2	1	1	1
ฐ	5	13	3	5	4	4	1	2	3	1	2	1	1	2	1	1	1	1
ฎ	0	0	3	1	0	0	1	0	0	1	0	1	1	0	1	1	1	1
ณ	10	13	8	9	12	12	13	10	8	9	8	9	8	8	8	9	9	9
ด	11	2	3	1	1	1	1	1	0	3	1	2	1	1	1	1	1	1
ต	7	10	11	8	9	9	3	8	4	1	4	1	4	4	4	4	4	4
บ	1	1	2	3	1	1	6	1	3	2	2	2	3	2	1	2	2	2
ม	6	3	2	3	2	2	6	3	3	3	3	0	2	3	1	2	2	2
Total	72	65	73	57	47	47	54	45	50	41	42	36	38	42	36	36	36	36
%	20.00	18.06	20.28	15.83	13.06	13.06	15.00	12.50	13.89	11.39	11.67	10.00	10.56	11.67	10.00	10.00	10.00	10.00
Mean (%)	19.45			13.98			13.80			11.02			10.74			10.00		

ตารางที่ 4.12 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	1	6	1	0	0	4	2	5	2	3	2	4	2	2	2	2	2
ฃ	3	3	5	2	1	1	2	1	1	0	1	0	1	1	1	1	1	1
ช	4	9	4	5	3	3	7	3	8	6	7	4	7	7	7	7	7	7
ฌ	19	6	16	0	10	10	10	10	9	10	8	11	9	8	8	8	8	8
ฎ	9	8	18	7	6	6	2	8	5	2	4	1	1	0	2	1	1	1
ฎ	7	13	3	5	7	7	1	3	2	1	1	1	1	1	1	1	1	1
ฏ	0	0	3	1	0	0	3	0	0	0	0	0	1	0	0	0	0	0
ณ	14	11	6	8	12	12	15	9	11	8	8	9	11	9	10	10	10	10
ด	11	1	2	2	2	2	1	2	0	2	1	1	1	1	1	1	1	1
ต	8	11	7	10	9	9	4	8	6	0	4	1	6	4	4	4	4	4
ป	1	2	2	2	1	1	3	2	3	3	2	1	3	2	3	3	3	3
ม	3	4	4	2	3	3	6	2	2	2	2	1	1	1	2	1	1	1
Total	79	69	76	55	54	54	58	50	52	36	41	32	46	36	41	39	39	39
%	21.94	19.17	21.11	15.28	15.00	15.00	16.11	13.89	14.44	10.00	11.39	8.89	12.78	10.00	11.39	10.83	10.83	10.83
Mean (%)	20.74			15.09			14.81			10.09			11.39			10.83		

ตารางที่ 4.13 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี Normal kernel

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
ฃ	5	5	5	4	4	4	6	5	5	5	5	4	4	4	4	4	4	4
ช	2	2	4	3	2	2	4	2	2	4	2	2	4	2	3	4	4	4
ฅ	19	12	13	14	15	13	11	15	12	13	14	12	11	13	11	13	13	13
ฉ	4	10	10	17	8	8	4	12	9	3	8	8	3	7	4	3	3	3
ฐ	9	15	9	10	12	12	1	9	6	1	7	5	3	4	3	3	3	3
ฎ	1	0	1	0	0	0	5	0	0	1	0	0	0	0	0	0	0	0
ณ	17	7	10	2	13	13	24	4	9	24	5	6	17	7	12	15	15	15
ด	7	3	7	2	2	2	14	2	7	12	2	7	10	3	10	9	9	9
ต	12	11	12	12	12	11	13	10	10	12	11	11	5	10	9	10	10	10
บ	1	1	3	1	1	2	3	1	2	2	1	1	2	2	1	1	1	1
ม	11	10	9	8	8	8	24	9	8	20	8	10	17	10	12	15	15	15
Total	88	77	84	74	78	76	110	70	71	97	64	67	77	63	70	78	78	78
%	24.44	21.39	23.33	20.56	21.67	21.11	30.56	19.44	19.72	26.94	17.78	18.61	21.39	17.50	19.44	21.67	21.67	21.67
Mean (%)	23.05			21.11			23.24			21.11			19.44			21.67		

ตารางที่ 4.14 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี LSCV

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	2	3	2	1	2	2	3	3	3	3	3	3	2	2	2	2	2	2
ฃ	5	5	5	4	5	4	5	5	5	5	3	4	3	3	5	3	3	3
ช	3	4	4	2	2	3	4	3	4	4	4	2	4	4	3	4	4	4
ฅ	18	16	16	16	15	16	12	12	13	12	10	12	12	14	15	15	15	15
ฉ	3	3	3	2	3	3	3	2	2	3	3	3	3	3	3	3	3	3
ฐ	11	12	12	12	11	12	11	10	11	6	8	5	2	2	2	3	3	3
ฎ	0	1	0	0	0	0	2	1	1	0	1	0	0	0	1	0	0	0
ณ	20	19	19	19	18	19	17	17	16	15	17	17	17	16	17	16	16	16
ด	6	5	5	2	5	4	5	4	5	5	5	4	10	9	9	9	9	9
ต	12	12	11	12	12	12	10	10	11	12	11	12	4	4	4	8	8	8
บ	1	2	2	1	2	1	2	2	2	2	2	2	2	2	1	1	1	1
ม	9	10	9	8	8	9	10	9	10	10	9	9	10	10	9	9	9	9
Total	90	92	87	79	83	85	84	78	83	77	76	73	69	69	71	73	73	73
%	25.00	25.56	24.17	21.94	23.06	23.61	23.33	21.67	23.06	21.39	21.11	20.28	19.17	19.17	19.72	20.28	20.28	20.28
Mean (%)	24.91			22.87			22.69			20.93			19.35			20.28		

ตารางที่ 4.15 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี LKCV

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	3	5	5	5	4	4	5	4	5	4	5	5	5	4	4	3	3	3
ฃ	3	3	3	4	3	3	3	3	3	3	3	3	3	2	3	2	2	2
ช	4	5	4	3	3	3	5	5	4	5	4	4	4	5	4	4	4	4
ฅ	18	16	15	12	12	11	8	11	10	10	10	11	10	10	11	11	11	11
ฉ	6	7	6	5	6	5	3	4	4	4	5	4	4	4	3	3	3	3
ง	8	9	8	8	6	7	1	1	2	1	2	2	2	2	2	2	2	2
จ	1	2	3	0	2	3	4	5	4	2	4	2	0	0	1	0	0	0
ฉ	13	12	13	13	14	11	12	12	11	10	12	11	11	11	12	12	12	12
ด	9	9	9	9	8	9	9	9	9	9	9	8	4	5	3	5	5	5
ต	12	11	12	11	11	11	11	10	12	5	6	6	6	7	6	8	8	8
บ	1	2	2	1	1	3	3	2	2	2	4	3	2	2	2	2	2	2
ม	5	5	6	4	5	4	5	6	5	9	6	8	6	6	4	5	5	5
Total	83	86	86	75	75	74	69	72	71	64	70	67	57	58	55	57	57	57
%	23.06	23.89	23.89	20.83	20.83	20.56	19.17	20.00	19.72	17.78	19.44	18.61	15.83	16.11	15.28	15.83	15.83	15.83
Mean (%)	23.61			20.74			19.63			18.61			15.74			15.83		

จากตารางที่ 4.1-4.3 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Euclidean พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะเพิ่มขึ้น สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 3$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 9.72%

จากตารางที่ 4.4-4.6 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Cityblock พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 3$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 11.39%

จากตารางที่ 4.7-4.9 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Correlation พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะเพิ่มขึ้น สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 5$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 11.11%

จากตารางที่ 4.10-4.12 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Cosine พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะเพิ่มขึ้น สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 2$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 8.61%

จากตารางที่ 4.13-4.15 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 200 ค่าด้วยวิธี LKCV พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 50 ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 15.74%

จากการศึกษาการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 200 ค่า สรุปได้ว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง ซึ่งวิธีที่ให้ประสิทธิภาพในการจำแนกกลุ่มดีที่สุด คือ วิธี K-nearest neighbour ที่ $k = 2$ โดยวัดระยะทางแบบ Cosine ที่จำนวนตัวอย่างเรียนรู้ขนาด 60 ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 8.61%

ตารางที่ 4.16 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	2	3	6	2	1	2	2	3	4	2	3	2	0	2	2	0	0	0
ฃ	0	3	0	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0
ช	3	2	2	0	2	2	4	1	1	3	1	2	1	1	2	1	1	1
ฅ	11	7	9	9	6	8	3	7	9	4	7	8	5	7	3	5	5	5
ฉ	8	6	6	5	4	4	2	3	3	3	3	2	1	2	2	2	2	2
ฎ	3	6	3	3	1	1	0	2	0	0	0	0	0	0	0	0	0	0
ฏ	1	0	2	0	1	1	6	1	1	2	1	1	1	1	1	1	1	1
ณ	4	7	5	5	2	2	5	2	3	4	1	3	2	1	1	1	1	1
ด	4	0	1	0	0	0	2	0	0	2	0	1	0	0	1	0	0	0
ต	7	7	11	6	4	4	4	4	4	1	2	1	4	2	2	2	2	2
บ	1	2	1	2	1	1	1	1	2	1	1	1	2	1	1	1	1	1
ม	1	1	4	2	1	2	1	2	2	1	2	1	1	1	1	1	1	1
Total	45	44	50	35	23	27	31	26	29	24	21	22	17	18	17	14	14	14
%	12.50	12.22	13.89	9.72	6.39	7.50	8.61	7.22	8.06	6.67	5.83	6.11	4.72	5.00	4.72	3.89	3.89	3.89
Mean (%)	12.87			7.87			7.96			6.20			4.81			3.89		

ตารางที่ 4.17 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	2	2	4	3	2	2	2	3	4	2	3	2	2	2	2	2	2	2
ฃ	0	2	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
ช	3	1	3	0	2	2	4	1	2	3	1	2	1	1	3	2	2	2
ฌ	13	8	8	11	9	9	4	9	7	4	7	7	5	6	3	4	4	4
ฉ	5	6	6	5	6	6	3	4	2	3	3	2	2	1	2	2	2	2
ฎ	5	12	3	5	4	4	0	4	0	0	0	0	0	0	0	0	0	0
ฏ	0	0	2	0	0	0	1	0	0	1	0	1	1	1	1	1	1	1
ณ	3	5	4	5	3	3	7	5	1	2	3	5	3	5	2	3	3	3
ด	5	0	1	1	0	0	2	0	1	2	0	1	1	1	2	1	1	1
ต	9	9	11	10	10	10	4	8	5	0	3	1	3	3	2	2	2	2
บ	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
ม	2	4	2	1	1	1	2	1	0	1	0	1	0	0	0	0	0	0
Total	48	50	45	42	38	38	30	36	24	19	21	23	19	21	18	18	18	18
%	13.33	13.89	12.50	11.67	10.56	10.56	8.33	10.00	6.67	5.28	5.83	6.39	5.28	5.83	5.00	5.00	5.00	5.00
Mean (%)	13.24			10.93			8.33			5.83			5.37			5.00		

ตารางที่ 4.18 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	2	0	5	4	3	3	2	4	4	2	4	2	3	3	2	3	3	3
ฃ	2	3	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0
ช	5	1	1	0	3	3	4	0	3	3	3	1	4	2	1	1	1	1
ฌ	15	9	10	9	8	8	6	8	7	8	7	5	6	6	5	6	6	6
ฉ	5	6	7	3	4	4	2	3	2	3	2	3	1	1	3	2	2	2
ฎ	8	13	2	4	5	5	0	5	0	0	0	0	0	0	0	0	0	0
ฏ	0	0	2	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
ถ	4	8	6	4	2	2	13	3	4	4	2	3	4	3	3	4	4	4
ด	5	1	1	1	2	2	1	1	1	2	1	1	1	1	1	1	1	1
ต	8	13	7	12	12	12	1	10	4	1	3	3	4	3	3	3	3	3
บ	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
ม	2	3	3	2	2	2	1	2	1	2	1	1	0	1	1	0	0	0
Total	57	57	46	41	44	44	34	37	27	26	24	23	24	21	20	21	21	21
%	15.83	15.83	12.78	11.39	12.22	12.22	9.44	10.28	7.50	7.22	6.67	6.39	6.67	5.83	5.56	5.83	5.83	5.83
Mean (%)	14.81			11.94			9.07			6.76			6.02			5.83		

ตารางที่ 4.19 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	4	6	3	0	0	2	2	3	1	2	2	2	2	1	1	1	1
ฃ	0	4	2	2	0	0	3	0	0	1	0	1	1	1	1	1	1	1
ช	2	1	2	1	1	1	4	1	3	2	1	1	3	1	2	2	2	2
ฅ	12	6	7	9	5	5	3	6	7	3	6	8	5	7	3	5	5	5
ฉ	8	7	8	4	6	6	3	4	4	4	3	3	2	3	4	3	3	3
ฐ	4	12	5	8	6	6	1	8	6	1	6	1	2	4	2	2	2	2
ฎ	2	2	7	4	0	0	5	2	2	1	1	1	3	1	1	1	1	1
ณ	8	8	9	6	4	4	10	5	7	6	6	8	5	6	4	4	4	4
ด	5	1	2	1	1	1	2	1	1	2	1	2	1	1	1	1	1	1
ต	7	8	7	7	8	8	6	6	8	5	6	6	7	6	7	5	5	5
บ	1	2	1	2	1	1	1	1	2	0	1	1	2	1	1	1	1	1
ม	2	2	3	2	1	1	2	2	2	1	2	1	1	1	1	1	1	1
Total	51	57	59	49	33	33	42	38	45	27	35	35	34	34	28	27	27	27
%	14.17	15.83	16.39	13.61	9.17	9.17	11.67	10.56	12.50	7.50	9.72	9.72	9.44	9.44	7.78	7.50	7.50	7.50
Mean (%)	15.46			10.65			11.58			8.98			8.89			7.50		

ตารางที่ 4.20 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	3	6	2	0	0	1	0	3	1	1	1	3	1	1	1	1	1
ฃ	10	2	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ช	2	1	2	0	1	1	4	0	2	1	0	0	1	0	1	0	0	0
ฌ	12	7	9	7	6	6	6	6	5	6	4	6	6	4	4	5	5	5
ฎ	6	9	11	6	6	6	3	4	4	3	3	1	1	2	4	1	1	1
ฎ	7	15	6	9	9	9	1	7	4	1	4	3	1	4	1	1	1	1
ฏ	0	1	4	1	0	0	6	0	1	2	0	2	2	2	2	2	2	2
ณ	7	9	8	7	7	7	9	6	7	7	5	5	6	5	6	6	6	6
ด	6	1	2	1	2	2	1	2	1	2	2	2	1	2	2	2	2	2
ต	6	10	9	9	12	12	7	11	11	5	8	5	9	8	9	7	7	7
บ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ม	2	3	2	2	1	1	1	2	2	2	2	0	0	0	0	0	0	0
Total	59	62	61	47	45	45	40	39	41	31	30	26	31	29	31	26	26	26
%	16.39	17.22	16.94	13.06	12.50	12.50	11.11	10.83	11.39	8.61	8.33	7.22	8.61	8.06	8.61	7.22	7.22	7.22
Mean (%)	16.85			12.69			11.11			8.05			8.43			7.22		

ตารางที่ 4.21 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	0	5	4	0	0	1	1	3	0	0	0	2	1	1	1	1	1
ฃ	0	1	1	2	0	0	1	1	0	0	0	0	0	0	0	0	0	0
ช	3	3	4	0	3	3	4	1	3	3	2	2	3	1	2	2	2	2
ฅ	14	7	9	7	7	7	3	6	5	4	5	5	7	6	6	7	7	7
ฉ	7	9	7	5	6	6	4	3	4	4	3	2	2	2	5	2	2	2
ฐ	7	13	5	8	9	9	1	6	2	1	2	2	1	2	1	1	1	1
ฎ	0	1	3	1	0	0	4	1	2	2	0	2	3	2	2	2	2	2
ณ	6	10	6	7	6	6	13	7	6	5	5	6	3	5	6	4	4	4
ด	5	1	2	2	3	3	2	2	2	3	2	2	2	2	2	2	2	2
ต	5	11	9	10	11	11	6	9	10	5	7	5	9	8	9	7	7	7
บ	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ม	2	4	2	2	3	3	3	2	2	4	2	1	2	1	3	2	2	2
Total	50	59	54	49	49	49	43	40	40	32	29	28	35	31	38	31	31	31
%	13.89	16.39	15.00	13.61	13.61	13.61	11.94	11.11	11.11	8.89	8.06	7.78	9.72	8.61	10.56	8.61	8.61	8.61
Mean (%)	15.09			13.61			11.39			8.24			9.63			8.61		

ตารางที่ 4.22 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	1	3	2	2	0	0	1	4	2	1	1	3	0	1	1	0	0	0
ฃ	1	1	3	3	1	1	1	0	1	1	1	0	1	1	1	1	1	1
ช	3	0	7	4	4	4	6	5	3	6	5	4	5	4	4	4	4	4
ฌ	13	11	3	5	5	5	5	4	5	5	8	4	4	3	4	4	4	4
ฎ	7	13	7	5	5	5	3	4	3	4	2	3	2	4	2	2	2	2
ฎ	3	3	8	4	3	3	2	1	4	1	2	2	1	1	2	1	1	1
ฏ	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ณ	5	4	5	4	2	2	7	3	3	2	3	2	3	1	3	2	2	2
ด	9	1	3	1	3	3	2	1	1	2	1	1	3	2	1	2	2	2
ต	5	7	2	3	1	1	2	3	1	1	1	1	2	1	1	1	1	1
บ	1	4	2	2	1	1	3	3	1	1	1	1	3	1	1	1	1	1
ม	3	1	1	2	2	2	2	2	2	1	1	1	0	0	1	0	0	0
Total	53	50	44	36	27	27	34	30	26	25	26	22	24	19	21	18	18	18
%	14.72	13.89	12.22	10.00	7.50	7.50	9.44	8.33	7.22	6.94	7.22	6.11	6.67	5.28	5.83	5.00	5.00	5.00
Mean (%)	13.61			8.33			8.33			6.76			5.93			5.00		

ตารางที่ 4.23 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	1	1	3	2	0	0	1	0	4	1	2	1	1	2	2	1	1	1
ฃ	1	3	1	3	1	1	0	1	0	0	0	0	0	0	0	0	0	0
ช	2	9	0	2	4	4	5	2	5	4	5	4	4	5	5	4	4	4
ฅ	16	7	10	8	7	7	5	8	9	6	7	9	7	7	4	6	6	6
ฉ	7	8	13	9	5	5	3	4	3	3	3	2	2	2	3	2	2	2
ฐ	5	9	3	5	8	8	1	4	2	1	3	1	1	2	1	1	1	1
ฎ	0	1	4	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0
ณ	6	4	2	4	4	4	8	2	4	3	1	2	4	1	2	2	2	2
ด	11	2	2	2	2	2	2	1	1	2	1	2	2	1	2	2	2	2
ต	4	6	3	5	2	2	4	2	2	0	2	1	3	2	1	2	2	2
บ	2	2	3	2	1	1	3	1	2	1	1	1	2	1	1	1	1	1
ม	4	3	2	2	2	2	2	1	2	1	1	1	0	1	1	0	0	0
Total	59	55	46	45	36	36	36	26	34	22	26	24	26	24	22	21	21	21
%	16.39	15.28	12.78	12.50	10.00	10.00	10.00	7.22	9.44	6.11	7.22	6.67	7.22	6.67	6.11	5.83	5.83	5.83
Mean (%)	14.82			10.83			8.89			6.67			6.67			5.83		

ตารางที่ 4.24 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	1	1	2	1	0	0	0	1	2	1	2	1	1	1	1	1	1	1
ฃ	3	4	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0
ช	4	6	0	4	5	5	5	3	6	5	5	5	4	4	4	4	4	4
ฌ	20	6	11	7	8	8	6	8	7	8	6	8	6	6	6	5	5	5
ฉ	5	9	12	9	5	5	4	5	3	4	3	4	3	3	3	4	4	4
ฎ	10	12	3	5	8	8	2	5	1	1	1	1	1	1	1	1	1	1
ฏ	0	0	2	1	0	0	1	0	1	0	0	0	1	0	1	0	0	0
ณ	6	3	3	4	4	4	10	3	4	2	2	2	4	2	4	3	3	3
ด	12	4	1	2	6	6	2	3	2	3	2	3	2	2	2	2	2	2
ต	4	7	8	5	3	3	2	2	1	1	0	1	2	1	2	1	1	1
บ	2	3	4	2	1	1	3	2	2	1	2	1	2	2	2	2	2	2
ม	3	5	2	2	2	2	2	2	2	1	2	1	1	1	1	1	1	1
Total	70	60	49	43	43	43	37	35	31	27	25	24	27	23	22	24	24	24
%	19.44	16.67	13.61	11.94	11.94	11.94	10.28	9.72	8.61	7.50	6.94	6.67	7.50	6.39	6.11	6.67	6.67	6.67
Mean (%)	16.57			11.94			9.54			7.04			6.67			6.67		

ตารางที่ 4.25 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเขียนรู้																	
	10			20			30			40			50			60		
ข	1	2	4	2	0	0	2	2	4	2	3	2	0	2	2	0	0	0
ฃ	0	3	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0	0
ช	3	8	3	7	3	3	5	3	4	4	2	2	2	1	2	1	1	1
ฌ	13	4	10	8	5	5	3	6	7	4	6	8	4	6	3	5	5	5
ฉ	8	6	11	6	4	4	2	3	3	3	3	2	1	2	2	2	2	2
ฎ	2	7	3	3	2	2	0	3	0	0	1	0	0	0	0	0	0	0
ฏ	2	1	2	1	1	1	2	1	1	1	0	1	1	1	1	1	1	1
ณ	3	4	5	3	1	1	7	1	3	3	2	4	2	3	1	1	1	1
ด	7	0	1	0	1	1	2	1	0	2	2	1	0	1	2	1	1	1
ต	5	5	10	5	2	2	4	2	5	1	1	1	3	2	2	2	2	2
ป	1	2	2	2	1	1	3	1	2	1	1	1	2	1	1	1	1	1
ม	3	1	2	2	2	2	1	2	1	1	1	1	0	0	0	0	0	0
Total	48	43	41	40	22	22	32	25	30	23	22	24	15	19	16	14	14	14
%	13.33	11.94	11.38	11.11	6.11	6.11	8.89	6.94	8.33	6.39	6.11	6.67	4.17	5.28	4.44	3.89	3.89	3.89
Mean (%)	12.22			7.78			8.05			6.39			4.63			3.89		

ตารางที่ 4.26 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	1	2	4	2	1	1	3	2	4	2	2	2	2	2	2	1	1	1
ฃ	0	3	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
ช	3	5	1	1	2	2	5	0	5	3	4	3	3	4	4	3	3	3
ฅ	15	6	9	9	7	7	4	8	8	6	8	8	5	7	4	4	4	4
ฉ	4	6	10	7	5	5	2	5	3	2	4	2	1	1	1	1	1	1
ฐ	7	10	3	5	7	7	0	4	0	1	1	1	0	1	1	1	1	1
ฎ	0	0	2	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1
ณ	5	6	3	4	3	3	7	3	4	3	2	3	5	3	2	3	3	3
ด	7	0	1	0	1	1	2	1	1	2	2	1	1	1	2	1	1	1
ต	6	8	7	10	7	7	4	6	4	0	2	1	4	2	1	2	2	2
บ	1	2	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1
ม	2	3	3	1	1	1	2	2	1	1	1	1	0	1	1	0	0	0
Total	51	51	44	42	36	36	32	33	33	22	27	24	23	24	20	18	18	18
%	14.17	14.17	12.22	11.67	10.00	10.00	7.78	9.17	9.17	6.11	7.50	6.67	6.39	6.67	5.56	5.00	5.00	5.00
Mean (%)	13.52			10.56			8.71			6.76			6.21			5.00		

ตารางที่ 4.27 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	1	1	4	2	2	2	2	2	4	2	4	2	3	2	2	2	2	2
ฃ	2	3	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
ช	5	5	0	5	5	5	6	2	6	4	3	4	5	3	3	3	3	3
ฌ	19	6	12	8	8	8	3	8	7	6	6	9	7	6	5	7	7	7
ฉ	4	7	9	6	3	3	4	4	3	4	3	2	2	1	3	2	2	2
ฎ	9	14	2	6	6	6	0	5	1	0	0	0	0	0	0	0	0	0
ฏ	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ณ	4	4	3	5	3	3	14	4	4	3	2	3	4	3	3	3	3	3
ด	10	2	1	0	6	6	2	1	0	3	1	1	0	1	2	1	1	1
ต	5	8	11	10	10	10	2	7	3	1	1	1	3	1	2	1	1	1
บ	1	2	2	2	1	1	3	1	2	1	1	1	0	1	1	1	1	1
ม	2	1	2	1	2	2	1	1	1	1	1	1	0	1	1	0	0	0
Total	62	53	48	47	47	47	37	35	31	25	22	24	24	19	22	20	20	20
%	17.22	14.72	13.33	13.06	13.06	13.06	10.28	9.72	8.61	6.94	6.11	6.67	6.67	5.27	6.11	5.56	5.56	5.56
Mean (%)	15.09			13.06			9.54			6.57			6.02			5.56		

ตารางที่ 4.28 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี Normal kernel

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
ฃ	2	4	3	4	5	3	5	3	3	3	2	3	2	2	2	2	2	2
ช	2	4	4	0	3	3	4	4	4	4	4	3	3	3	3	3	3	3
ฅ	19	14	13	11	13	14	9	12	12	14	11	14	10	12	10	11	11	11
ฉ	2	3	8	10	3	8	3	2	3	2	2	3	2	3	2	2	2	2
ฐ	8	5	5	9	4	5	0	2	1	0	2	2	0	1	2	2	2	2
ฎ	1	3	1	0	1	1	4	0	0	2	0	0	0	0	0	0	0	0
ณ	10	10	10	2	8	10	20	12	8	20	13	12	8	9	9	7	7	7
ด	5	6	7	2	5	4	9	6	6	10	4	5	2	2	3	2	2	2
ต	12	10	9	9	9	8	6	7	11	4	4	6	3	3	6	5	5	5
บ	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	1
ม	9	8	11	2	6	10	16	9	14	11	10	9	7	7	8	8	8	8
Total	73	70	74	52	60	69	79	60	65	72	55	59	40	45	47	45	45	45
%	20.28	19.44	20.56	14.44	16.67	19.17	21.94	16.67	18.06	20.00	15.28	16.39	11.11	12.50	13.06	12.50	12.50	12.50
Mean (%)	20.09			16.76			18.89			17.22			12.22			12.50		

ตารางที่ 4.29 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี LSCV

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
ฃ	2	4	3	4	4	4	4	2	4	2	3	2	3	3	2	2	2	2
ช	0	2	1	0	1	1	4	3	3	4	2	3	3	2	2	2	2	2
ฌ	20	19	18	14	14	15	15	16	15	17	14	15	11	10	13	17	17	17
ฎ	2	2	2	1	2	3	3	4	3	2	2	2	2	2	2	2	2	2
ฎ	11	10	11	10	8	9	4	3	3	0	3	2	0	1	2	1	1	1
ฏ	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
ณ	18	18	17	16	16	17	17	16	17	18	15	14	13	11	12	10	10	10
ด	5	6	5	2	4	3	5	5	4	10	7	9	2	2	3	2	2	2
ต	15	14	14	11	9	11	8	8	9	4	6	4	3	4	3	4	4	4
บ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ม	12	11	12	8	7	8	5	6	5	6	5	5	5	5	5	8	8	8
Total	88	91	85	70	69	75	69	67	67	68	61	60	46	44	48	52	52	52
%	24.44	25.28	23.61	19.44	19.17	20.83	19.17	18.61	18.61	18.89	16.94	16.67	12.78	12.22	13.33	14.44	14.44	14.44
Mean (%)	24.44			19.81			18.79			17.50			12.78			14.44		

ตารางที่ 4.30 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี LKCV

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	3	5	5	5	4	4	4	4	4	4	4	4	4	4	4	4	4	4
ฃ	3	2	3	4	3	4	3	4	4	2	3	3	3	3	3	2	2	2
ช	4	4	3	2	4	4	5	4	5	3	4	3	4	3	3	4	4	4
ฅ	17	16	15	12	12	11	9	9	10	13	11	10	9	10	9	9	9	9
ฉ	4	4	5	8	7	8	4	3	2	3	4	2	2	2	2	2	2	2
ง	7	6	7	7	5	6	0	1	1	0	1	0	0	1	0	1	1	1
จ	1	3	2	0	4	3	4	3	2	1	0	0	0	0	0	0	0	0
ฉ	9	10	11	2	10	9	15	12	13	13	9	11	7	8	8	7	7	7
ด	5	4	5	1	2	1	5	5	5	5	5	5	2	2	3	2	2	2
ต	11	11	10	11	11	11	4	3	4	3	4	3	3	4	3	5	5	5
บ	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ม	13	11	11	2	7	5	10	9	8	8	7	8	4	4	3	4	4	4
Total	79	78	79	55	70	67	64	58	59	56	53	50	39	42	39	41	41	41
%	21.94	21.67	21.94	15.28	19.44	18.61	17.78	16.11	16.39	15.56	14.72	13.89	10.83	11.67	10.83	11.39	11.39	11.39
Mean (%)	21.85			17.78			16.76			14.72			11.11			11.39		

จากตารางที่ 4.16-4.18 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Euclidean พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะเพิ่มขึ้น สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 2$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 3.89%

จากตารางที่ 4.19-4.21 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Cityblock พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 3$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 7.22%

จากตารางที่ 4.22-4.24 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Correlation พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะเพิ่มขึ้น สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 2$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 5.00%

จากตารางที่ 4.25-4.27 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระหว่างแบบ Cosine พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะเพิ่มขึ้น สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 2$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 3.89%

จากตารางที่ 4.28-4.30 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 500 ค่าด้วยวิธี LKCV พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 50 ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 11.11%

จากการศึกษาการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 500 ค่า สรุปได้ว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง ซึ่งวิธีที่ให้ประสิทธิภาพในการจำแนกกลุ่มดีที่สุด คือ วิธี K-nearest neighbour ที่ $k = 2$ โดยวัดระยะห่างระหว่างแบบ Euclidean กับ Cosine ที่จำนวนตัวอย่างเรียนรู้ขนาด 60 ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 3.89%

ตารางที่ 4.31 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	2	2	6	2	1	3	0	3	4	0	3	1	0	0	0	0	0	0
ฃ	0	1	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ช	1	1	4	1	0	0	5	0	1	2	1	2	1	1	1	1	1	1
ฅ	6	8	4	4	8	7	9	6	9	7	5	4	4	5	4	5	5	5
ฉ	12	5	4	2	5	5	3	4	3	3	4	3	2	2	3	2	2	2
ง	4	7	9	7	3	3	0	6	0	0	1	1	0	0	0	0	0	0
จ	0	0	2	0	0	0	6	0	1	3	0	0	1	0	1	0	0	0
ฉ	6	6	9	5	4	4	8	4	3	4	3	3	4	3	3	3	3	3
ด	3	0	0	0	0	1	2	0	0	2	0	1	0	0	1	0	0	0
ต	11	9	14	8	6	6	6	6	4	9	6	10	7	5	5	5	5	5
บ	1	1	2	1	1	1	0	1	2	0	1	1	1	1	1	1	1	1
ม	3	1	6	5	2	2	3	6	2	3	5	5	4	5	3	5	5	5
Total	49	41	62	37	30	32	42	36	29	33	29	31	24	22	22	22	22	22
%	13.61	11.39	17.22	10.28	8.33	8.89	11.67	10.00	8.06	9.17	8.06	8.61	6.67	6.11	6.11	6.11	6.11	6.11
Mean (%)	14.07			9.17			9.91			8.61			6.30			6.11		

ตารางที่ 4.32 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	2	3	6	3	1	1	2	2	4	1	2	1	0	1	1	0	0	0
ฃ	0	2	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ช	5	2	8	1	2	2	5	1	2	5	0	4	3	0	3	1	1	1
ฌ	12	8	5	6	7	7	7	5	6	5	4	5	3	4	3	3	3	3
ฉ	12	6	5	3	4	4	4	2	3	4	4	4	3	4	4	4	4	4
ฎ	4	14	4	6	6	6	0	4	0	0	0	1	0	1	0	0	0	0
ฏ	1	0	2	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
ถ	4	7	5	7	4	4	14	4	4	6	3	4	6	4	4	4	4	4
ด	3	0	1	0	0	0	2	0	0	2	0	1	0	0	1	0	0	0
ต	9	17	14	15	13	13	7	13	13	8	12	11	11	12	12	12	12	12
บ	1	0	1	1	1	1	0	1	1	0	1	0	1	1	0	1	1	1
ม	2	4	4	3	3	3	3	2	3	4	4	2	2	3	3	3	3	3
Total	55	63	56	47	41	41	45	34	36	36	30	33	29	30	31	28	28	28
%	15.28	17.50	15.56	13.06	11.39	11.39	12.50	9.44	10.00	10.00	8.33	9.17	8.06	8.33	8.61	7.78	7.78	7.78
Mean (%)	16.11			11.95			10.65			9.17			8.33			7.78		

ตารางที่ 4.33 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	1	2	6	3	0	0	2	3	4	1	3	1	1	1	1	1	1	1
ฃ	2	3	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	0
ช	5	1	4	3	4	4	5	3	4	5	3	4	4	2	4	3	3	3
ฌ	13	7	5	6	7	7	5	6	5	5	6	7	4	4	5	3	3	3
ฎ	9	5	5	2	5	5	4	3	4	3	3	3	3	2	3	3	3	3
ฎ	8	14	3	6	7	7	0	4	0	0	0	0	0	0	0	0	0	0
ฏ	1	0	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
ณ	5	8	8	4	6	6	15	7	6	4	3	5	7	2	5	5	5	5
ด	3	1	0	0	0	0	1	0	0	2	0	1	0	0	2	0	0	0
ต	8	17	14	17	15	15	6	13	15	8	12	9	13	12	12	12	12	12
บ	1	0	1	1	1	1	0	1	1	0	1	1	1	1	0	1	1	1
ม	2	4	3	3	3	3	4	2	3	3	3	5	2	2	3	3	3	3
Total	58	62	51	47	49	49	44	43	42	31	34	36	35	26	35	31	31	31
%	16.11	17.22	14.17	13.06	13.61	13.61	12.22	11.94	11.67	8.61	9.44	10.00	9.72	7.22	9.72	8.61	8.61	8.61
Mean (%)	15.83			13.43			11.94			9.35			8.89			8.61		

ตารางที่ 4.34 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	2	8	3	0	0	1	1	2	0	1	1	0	1	0	0	0	0
ฃ	1	4	3	2	2	2	1	2	1	1	1	1	1	1	1	1	1	1
ช	1	2	6	2	1	1	7	1	2	5	2	4	3	2	3	3	3	3
ฅ	4	7	3	4	4	4	6	3	3	4	3	3	4	3	4	4	4	4
ฉ	9	6	9	4	5	5	5	3	5	5	4	4	5	4	6	4	4	4
ฐ	6	10	8	7	2	2	1	4	2	0	3	2	2	2	0	2	2	2
ฎ	1	2	5	2	2	2	7	2	4	3	2	2	6	3	3	3	3	3
ณ	6	9	9	7	9	9	10	7	7	8	7	8	7	7	6	7	7	7
ด	3	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1
ต	9	13	13	12	12	12	8	10	11	10	11	11	11	11	13	11	11	11
บ	1	2	1	1	1	1	0	1	1	0	1	1	1	1	0	1	1	1
ม	4	4	6	6	4	4	5	6	4	5	5	5	4	5	4	5	5	5
Total	45	62	72	51	43	43	52	41	43	43	41	43	45	41	41	42	42	42
%	12.50	17.22	20.00	14.17	11.94	11.94	14.44	11.39	11.94	11.94	11.39	11.94	12.50	11.39	11.39	11.67	11.67	11.67
Mean (%)	16.57			12.68			12.59			11.76			11.76			11.67		

ตารางที่ 4.35 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	2	8	3	0	0	2	0	1	1	0	0	1	0	1	1	1	1
ฃ	1	4	3	2	2	2	0	2	1	0	0	0	1	1	1	1	1	1
ช	5	3	6	2	2	2	5	1	3	6	1	5	4	1	2	1	1	1
ฅ	8	5	3	5	4	4	5	4	4	4	3	3	4	3	3	3	3	3
ฉ	9	6	8	4	5	5	5	3	4	4	2	3	5	3	4	3	3	3
ฐ	7	11	7	6	6	6	1	6	2	0	2	0	0	0	0	0	0	0
ฎ	1	0	6	2	0	0	5	1	2	2	3	2	4	2	2	2	2	2
ณ	3	10	6	7	4	4	12	6	5	7	7	6	7	7	7	7	7	7
ด	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1
ต	8	15	16	14	14	14	9	13	13	7	14	10	12	14	11	13	13	13
บ	0	1	1	1	0	0	0	1	1	0	1	0	1	1	0	1	1	1
ม	4	4	2	1	3	3	5	2	3	5	4	3	3	4	4	4	4	4
Total	48	62	67	48	41	41	51	40	40	37	38	33	43	37	36	37	37	37
%	13.33	17.22	18.61	13.33	11.39	11.39	14.17	11.11	11.11	10.28	10.56	9.17	11.94	10.29	10.00	10.29	10.29	10.29
Mean (%)	16.39			12.04			12.13			10.00			10.74			10.29		

ตารางที่ 4.36 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	2	7	2	0	0	1	0	2	0	0	0	0	0	0	0	0	0
ฃ	2	4	2	2	2	2	0	2	0	0	0	0	0	0	0	0	0	0
ช	6	3	5	2	3	3	4	1	5	5	2	5	5	3	4	3	3	3
ฅ	8	5	4	5	5	5	5	3	2	4	1	3	6	1	3	3	3	3
ฉ	8	7	7	6	7	7	4	4	3	4	3	2	3	3	4	4	4	4
ฐ	10	15	5	9	12	12	0	6	1	0	2	0	0	0	0	0	0	0
ฎ	0	0	3	2	0	0	7	1	1	4	1	2	2	2	4	2	2	2
ณ	4	9	5	6	6	6	12	4	5	5	5	7	7	6	6	7	7	7
ด	3	0	1	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1
ต	9	17	14	15	13	13	11	13	13	7	12	8	10	10	11	9	9	9
บ	0	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0
ม	3	6	3	1	4	4	5	3	3	5	4	3	3	4	3	4	4	4
Total	53	68	57	52	54	54	52	39	37	36	31	31	38	30	36	33	33	33
%	14.72	18.89	15.83	14.44	15.00	15.00	14.44	10.83	10.28	10.00	8.61	8.61	10.56	8.33	10.00	8.33	8.33	8.33
Mean (%)	16.48			14.81			11.85			9.07			9.63			8.33		

ตารางที่ 4.37 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	4	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ฃ	2	2	3	2	3	3	0	0	2	1	1	1	0	1	1	1	1	1
ช	4	10	5	5	3	3	9	7	3	7	8	6	6	5	6	5	5	5
ฅ	6	5	6	4	6	6	6	2	4	4	3	2	3	2	2	3	3	3
ฉ	10	8	5	3	6	6	5	6	5	5	4	4	6	5	4	5	5	5
ฐ	2	4	8	5	3	3	1	3	5	0	1	3	0	0	1	0	0	0
ฎ	1	5	0	3	0	0	3	2	1	1	1	1	2	1	1	1	1	1
ณ	5	4	8	7	6	6	9	6	6	4	4	4	6	4	4	4	4	4
ด	5	1	1	1	1	1	2	1	1	3	2	1	1	1	1	1	1	1
ต	7	13	4	4	3	3	3	4	3	5	9	3	3	3	4	3	3	3
บ	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ม	4	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	3
Total	48	61	45	39	35	35	42	35	34	34	37	39	31	25	38	27	27	27
%	13.33	16.94	12.50	10.83	9.72	9.72	11.67	9.72	9.44	9.44	10.28	10.83	8.61	6.94	10.56	7.50	7.50	7.50
Mean (%)	14.26			10.09			10.28			10.18			8.70			7.50		

ตารางที่ 4.38 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระหว่างแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	0	2	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
ฃ	1	2	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	0
ช	5	5	9	8	4	4	9	4	8	8	6	6	7	5	6	7	7	7
ฅ	12	4	5	6	6	6	5	6	4	6	4	5	4	4	4	3	3	3
ฉ	10	6	8	4	5	5	4	5	4	4	3	4	4	4	4	4	4	4
ง	9	9	5	6	5	5	0	5	3	0	2	2	0	2	0	0	0	0
จ	1	0	4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ฉ	5	7	3	5	4	4	13	4	4	4	3	4	4	3	3	3	3	3
ด	7	2	1	1	2	2	2	2	1	2	1	2	2	2	2	2	2	2
ต	5	8	15	12	7	7	7	7	11	4	7	6	9	7	5	7	7	7
บ	1	2	2	1	2	2	3	2	1	1	1	1	1	1	1	1	1	1
ม	5	4	3	2	4	4	5	4	2	4	2	1	2	2	3	2	2	2
Total	61	49	58	48	40	40	49	40	39	33	29	31	33	30	28	29	29	29
%	16.94	13.61	16.11	13.33	11.11	11.11	13.61	11.11	10.83	9.17	8.06	8.61	9.17	8.33	7.78	8.06	8.06	8.06
Mean (%)	15.55			11.85			11.85			8.61			8.43			8.06		

ตารางที่ 4.39 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	0	1	3	2	0	0	1	0	2	1	1	1	0	0	0	0	0	0
ฃ	2	4	2	2	1	1	1	2	1	0	0	0	0	0	0	0	0	0
ช	5	4	8	8	5	5	8	6	9	7	7	8	9	6	6	7	7	7
ฅ	11	6	6	5	7	7	4	4	3	4	4	6	5	4	3	4	4	4
ฉ	9	6	8	4	7	7	5	4	4	4	3	3	5	3	4	4	4	4
ฎ	7	9	4	8	8	8	0	5	2	0	1	1	1	1	1	0	0	0
ฏ	1	0	4	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ณ	4	8	4	7	4	4	15	3	7	4	4	4	6	4	4	4	4	4
ด	8	2	1	2	4	4	2	1	2	3	1	2	2	2	2	2	2	2
ต	6	9	14	10	7	7	6	7	10	4	6	7	7	6	5	6	6	6
บ	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1
ม	5	5	1	2	3	3	5	3	2	5	3	3	2	3	4	2	2	2
Total	59	55	56	52	47	47	51	36	43	33	31	36	38	30	30	30	30	30
%	16.39	15.28	15.56	14.44	13.06	13.06	14.17	10.00	11.94	9.17	8.61	10.00	10.56	8.33	8.33	8.33	8.33	8.33
Mean (%)	15.74			13.52			12.04			9.26			9.07			8.33		

ตารางที่ 4.40 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	1	1	5	1	0	0	0	0	4	0	1	0	0	0	0	0	0	0
ฃ	1	1	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	0
ช	2	3	7	5	2	2	6	2	2	5	2	5	2	2	2	2	2	2
ฅ	5	7	6	5	6	6	7	4	4	5	3	4	4	4	2	4	4	4
ฉ	12	4	4	2	5	5	3	4	4	4	4	4	3	3	4	3	3	3
ฎ	3	4	5	4	1	1	0	3	1	0	1	1	0	0	0	0	0	0
ฏ	1	0	3	0	0	0	4	0	0	2	0	0	1	0	2	0	0	0
ณ	6	5	5	4	5	5	9	5	5	5	4	5	5	4	4	4	4	4
ด	3	1	0	0	1	1	2	0	0	3	0	1	0	0	1	0	0	0
ต	9	3	12	6	2	2	4	3	5	5	3	8	5	3	3	3	3	3
บ	1	1	2	1	1	1	0	1	1	0	1	1	1	1	0	1	1	1
ม	3	2	5	4	2	2	3	4	4	3	4	4	4	4	3	4	4	4
Total	47	32	55	34	26	26	38	27	30	32	23	33	25	21	21	21	21	21
%	13.06	8.89	15.28	9.44	7.22	7.22	10.56	7.50	8.33	8.89	6.39	9.17	6.94	5.83	5.83	5.83	5.83	5.83
Mean (%)	12.41			7.96			8.80			8.15			6.20			5.83		

ตารางที่ 4.41 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	1	2	5	1	0	0	2	0	3	1	1	1	0	0	0	0	0	0
ฃ	0	1	1	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0
ช	6	4	8	4	2	2	5	1	3	5	2	2	3	2	3	2	2	2
ฅ	13	6	5	6	6	6	5	4	3	5	4	5	3	4	3	3	3	3
ฉ	10	5	5	3	4	4	4	2	3	3	3	3	2	2	2	2	2	2
ฐ	8	11	4	4	4	4	0	3	0	0	0	1	0	1	0	0	0	0
ฎ	1	0	2	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0
ณ	5	7	3	5	4	4	14	4	3	5	3	5	5	3	4	4	4	4
ด	4	1	0	0	0	0	2	0	0	2	0	1	0	0	2	0	0	0
ต	8	11	12	12	12	12	6	11	11	8	0	9	9	9	9	9	9	9
บ	1	1	1	1	1	1	0	1	1	0	1	0	1	1	0	1	1	1
ม	2	2	4	3	3	3	5	1	3	4	2	2	3	2	3	2	2	2
Total	59	51	50	41	36	36	45	28	30	34	26	29	26	24	26	23	23	23
%	16.39	14.17	13.89	11.39	10.00	10.00	12.50	7.78	8.33	9.44	7.22	8.06	7.22	6.67	7.22	6.39	6.39	6.39
Mean (%)	14.82			10.46			9.53			8.24			7.04			6.39		

ตารางที่ 4.42 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	1	1	5	1	0	0	1	1	3	1	1	1	0	0	0	0	0	0
ฃ	3	1	1	2	1	1	0	1	1	0	0	0	0	0	0	0	0	0
ช	5	3	6	3	4	4	5	4	7	4	3	5	4	1	4	2	2	2
ฅ	12	5	5	6	7	7	5	6	4	6	3	5	4	4	5	3	3	3
ฉ	8	3	7	3	4	4	3	2	4	3	4	4	3	3	3	3	3	3
ง	6	13	3	6	9	9	0	4	1	0	1	0	0	0	0	0	0	0
จ	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ฉ	4	8	3	5	6	6	17	5	4	5	5	3	6	4	4	4	4	4
ด	4	0	1	0	1	1	2	0	0	3	0	1	1	0	2	0	0	0
ต	7	14	13	12	12	12	6	11	12	6	10	8	10	11	10	11	11	11
บ	1	1	1	1	1	1	0	1	1	0	1	1	1	1	0	1	1	1
ม	3	4	3	3	3	3	5	3	3	4	3	3	2	3	4	2	2	2
Total	54	53	49	42	48	48	45	38	40	32	31	31	31	27	32	26	26	26
%	15.00	14.72	13.61	11.67	13.33	13.33	12.50	10.56	11.11	8.89	8.61	8.61	8.61	7.50	8.89	7.22	7.22	7.22
Mean (%)	14.44			12.78			11.39			8.70			8.33			7.22		

ตารางที่ 4.43 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี Normal kernel

ตัวอักษร	จำนวนตัวอย่างเขียนรู้																	
	10			20			30			40			50			60		
ข	2	4	3	5	2	2	4	3	4	4	2	2	3	2	3	3	3	3
ฃ	3	3	4	5	4	5	5	5	5	3	4	3	3	3	3	3	3	3
ช	4	4	4	3	4	3	4	3	3	4	3	3	4	4	2	3	3	3
ฅ	12	12	11	8	10	8	10	10	10	11	8	12	10	10	8	9	9	9
ฉ	4	4	7	10	4	8	2	4	3	3	3	6	3	3	4	3	3	3
ง	8	4	6	10	5	3	0	3	2	0	4	2	0	1	4	4	4	4
จ	0	0	0	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0
ฉ	6	9	8	1	10	11	22	9	9	18	10	5	11	11	10	10	10	10
ด	5	5	5	0	5	4	3	4	6	2	4	3	2	2	2	2	2	2
ต	12	11	13	15	11	10	13	12	12	12	11	13	11	10	13	13	13	13
บ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ม	9	8	8	3	6	10	12	10	10	13	9	11	8	7	10	8	8	8
Total	66	65	70	61	63	66	77	64	65	71	59	61	56	55	60	59	59	59
%	18.33	18.05	19.44	16.94	17.50	18.33	21.39	17.78	18.05	19.72	16.39	16.94	15.56	15.28	16.67	16.39	16.39	16.39
Mean (%)	18.61			17.59			19.07			17.68			15.84			16.39		

ตารางที่ 4.44 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี LSCV

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	1	3	3	3	3	3	4	3	4	3	3	3	3	3	3	3	3	3
ฃ	1	3	5	5	4	4	6	3	5	3	4	4	4	3	4	3	3	3
ช	0	3	1	1	2	3	4	4	4	4	4	3	4	4	4	4	4	4
ฅ	20	19	18	15	16	15	15	16	14	15	13	14	12	11	12	12	12	12
ฉ	5	6	6	10	8	7	3	4	3	3	3	3	2	2	2	3	3	3
ง	16	16	16	15	16	16	15	15	15	14	13	14	13	10	11	6	6	6
จ	0	1	2	0	2	1	0	1	1	2	2	2	2	2	2	2	2	2
ฉ	8	8	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ด	2	2	2	0	3	2	3	2	3	3	3	2	2	2	2	2	2	2
ต	15	15	16	17	15	16	16	15	16	15	13	15	13	14	13	14	14	14
บ	1	1	1	1	1	1	1	1	1	1	1	1	22	1	1	1	1	1
ม	15	16	15	7	9	8	10	9	9	8	10	9	8	10	8	11	11	11
Total	84	93	93	74	79	76	77	73	75	71	69	70	65	62	62	61	61	61
%	23.33	25.83	25.83	20.56	21.94	21.11	21.39	20.28	20.83	19.72	19.17	19.44	18.06	17.22	17.22	16.94	16.94	16.94
Mean (%)	25.00			21.20			20.83			19.44			17.50			16.94		

ตารางที่ 4.45 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี LKCV

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	2	5	3	5	4	4	4	4	5	4	5	4	5	5	4	5	5	5
ฃ	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
ช	5	5	5	3	5	4	5	5	4	4	4	4	4	4	4	4	4	4
ฅ	12	12	11	7	10	9	10	11	10	12	10	11	9	10	9	9	9	9
ฉ	6	5	6	5	4	5	3	4	3	3	3	4	3	3	3	3	3	3
ง	9	10	8	7	4	6	1	1	1	0	0	0	0	0	0	0	0	0
จ	0	0	1	0	1	1	2	1	1	0	0	0	0	0	0	0	0	0
ฉ	14	15	15	15	15	15	15	14	15	15	13	12	10	11	9	10	10	10
ด	3	4	4	0	3	2	3	3	2	2	2	2	2	2	2	2	2	2
ต	10	13	12	15	13	13	12	10	12	9	10	10	11	10	11	12	12	12
บ	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ม	10	9	11	4	7	8	12	10	9	10	9	10	7	7	8	8	8	8
Total	76	83	81	66	71	72	72	68	67	64	61	62	56	57	55	58	58	58
%	21.11	23.06	22.50	18.33	19.72	20.00	20.00	18.89	18.61	17.78	16.94	17.22	15.56	15.83	15.28	16.11	16.11	16.11
Mean (%)	22.22			19.35			19.17			17.31			15.56			16.11		

จากตารางที่ 4.31-4.33 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Euclidean พบว่าเมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะเพิ่มขึ้น สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 2$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 6.11%

จากตารางที่ 4.34-4.36 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Cityblock พบว่าเมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 5$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 8.33%

จากตารางที่ 4.37-4.39 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Correlation พบว่าเมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะเพิ่มขึ้น สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 2$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 7.50%

จากตารางที่ 4.40-4.42 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Cosine พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะเพิ่มขึ้น สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 2$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 5.83%

จากตารางที่ 4.43-4.45 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 1,000 ค่าด้วยวิธี LKCV พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 50 ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 15.56%

จากการศึกษาการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 1,000 ค่า สรุปได้ว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง ซึ่งวิธีที่ให้ประสิทธิภาพในการจำแนกกลุ่มดีที่สุด คือ วิธี K-nearest neighbour ที่ $k = 2$ โดยวัดระยะห่างระยะทางแบบ Cosine ที่จำนวนตัวอย่างเรียนรู้ขนาด 60 ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 5.83%

4.2 ผลการศึกษามาจากลายมือเขียนตัวอักษรไทย 10 คน

ตารางที่ 4.46 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	6	5	5	4	4	4	4	3	3	3	4	7	4	5	8	5	5	5
ฃ	13	13	14	13	10	11	12	11	11	12	11	11	11	11	11	11	11	11
ช	6	6	10	6	4	6	5	3	7	5	3	6	4	3	5	4	4	4
ฅ	10	12	5	6	4	3	4	3	2	2	1	3	1	1	1	1	1	1
ฉ	14	14	17	14	16	16	15	14	12	14	13	11	13	12	13	12	12	12
ง	13	14	11	13	14	9	13	11	9	10	12	11	12	14	12	14	14	14
จ	7	8	10	9	9	11	9	9	9	9	7	9	8	7	8	8	8	8
ฉ	8	10	14	9	10	10	9	10	8	9	10	8	9	9	8	9	9	9
ด	8	10	14	9	3	15	10	4	11	10	11	12	9	11	12	9	9	9
ต	8	7	12	5	8	6	4	5	4	3	4	5	3	4	5	4	4	4
บ	7	10	8	8	10	7	8	9	8	8	10	9	9	10	9	9	9	9
ม	18	17	13	14	13	10	10	9	8	8	8	9	8	7	9	7	7	7
Total	118	126	133	110	105	108	103	101	92	93	94	101	91	94	101	93	93	93
%	32.78	35.00	36.94	30.56	29.17	30.00	28.61	28.06	25.56	25.83	26.11	28.06	25.28	26.11	28.06	25.83	25.83	25.83
Mean (%)	34.91			29.91			27.41			26.67			26.48			25.83		

ตารางที่ 4.47 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	6	3	5	3	3	4	3	4	5	3	4	6	5	4	8	5	5	5
ฃ	14	12	14	13	11	10	13	10	10	11	11	10	11	12	11	12	12	12
ช	4	8	8	4	7	4	3	3	5	13	3	6	3	3	5	3	3	3
ฅ	12	13	8	7	6	4	6	4	2	2	3	2	1	2	1	1	1	1
ฉ	12	15	17	11	14	16	13	13	12	12	13	11	13	11	11	11	11	11
ฐ	9	15	19	11	13	8	12	10	9	9	13	9	11	12	10	11	11	11
ฎ	8	10	10	6	8	8	8	9	8	9	5	7	6	7	7	7	7	7
ณ	9	10	15	6	9	10	9	9	9	10	8	9	9	8	7	7	7	7
ด	8	10	15	9	15	17	10	14	11	10	10	13	9	12	12	11	11	11
ต	9	8	11	6	7	6	4	4	5	3	2	6	3	3	6	4	4	4
บ	6	9	8	6	7	5	6	5	4	4	6	5	6	6	6	6	6	6
ม	19	19	18	13	14	10	12	9	10	10	8	10	9	7	10	9	9	9
Total	116	132	148	95	114	102	99	94	90	96	86	94	86	87	94	86	86	86
%	32.22	36.67	41.11	26.39	31.67	28.33	27.50	26.11	25.00	26.67	23.89	26.11	23.89	24.17	26.11	23.89	23.89	23.89
Mean (%)	36.67			28.80			26.20			25.56			24.72			23.89		

ตารางที่ 4.48 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	6	6	3	3	2	5	2	4	4	2	4	6	4	4	7	4	4	4
ฃ	15	14	15	12	10	11	13	10	10	11	10	9	10	10	8	10	10	10
ช	3	7	8	4	7	6	4	4	3	3	4	6	3	4	4	3	3	3
ฅ	11	12	8	8	7	6	6	6	5	5	3	4	1	3	1	1	1	1
ฉ	10	14	12	11	11	14	11	12	11	13	12	11	13	12	13	14	14	14
ฐ	11	16	19	10	16	12	10	9	8	9	9	11	9	10	10	9	9	9
ฎ	9	11	7	6	7	9	7	6	8	7	6	8	6	6	5	6	6	6
ณ	11	8	16	8	9	9	9	8	7	9	9	8	10	8	8	9	9	9
ด	9	13	13	12	12	17	10	14	14	12	11	14	10	11	12	11	11	11
ต	17	8	10	6	7	6	5	3	3	4	4	5	3	4	5	4	4	4
บ	17	7	10	5	6	6	6	5	4	5	6	2	6	3	3	6	6	6
ม	21	19	17	15	17	12	13	13	8	10	11	8	10	11	8	10	10	10
Total	120	135	138	100	111	113	96	94	85	90	89	92	85	86	84	87	87	87
%	33.33	37.50	38.33	27.78	30.83	31.39	26.67	26.11	23.61	25.00	24.72	25.56	23.61	23.89	23.33	24.17	24.17	24.17
Mean (%)	36.39			30.00			25.46			25.09			23.61			24.17		

ตารางที่ 4.49 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	8	5	7	5	5	5	4	4	14	4	4	5	4	6	5	5	5	5
ฃ	19	16	16	16	13	13	15	11	13	12	11	13	12	11	12	12	12	12
ช	6	5	9	5	4	6	6	4	6	6	3	5	4	3	3	4	4	4
ฅ	10	11	4	7	3	4	2	4	2	2	4	4	3	3	2	2	2	2
ฉ	14	15	20	12	19	17	15	16	14	14	16	13	15	16	15	15	15	15
ฐ	14	17	13	5	14	11	15	13	11	13	13	12	12	14	13	13	13	13
ฎ	11	11	11	10	10	12	10	11	10	10	7	11	8	9	9	9	9	9
ณ	12	12	14	9	12	10	11	13	9	11	12	7	11	10	7	9	9	9
ด	10	14	18	9	15	18	11	16	14	12	12	15	12	14	15	14	14	14
ต	9	10	14	8	8	11	7	9	9	7	8	10	6	7	8	6	6	6
บ	11	12	9	11	12	11	12	13	12	13	12	12	12	11	11	11	11	11
ม	18	20	17	17	14	12	14	10	12	11	9	13	10	10	12	10	10	10
Total	142	148	152	124	129	130	122	124	116	115	111	120	109	114	112	110	110	110
%	39.44	41.11	42.22	34.44	35.83	36.11	33.89	34.44	32.22	31.94	30.83	33.33	30.28	31.67	31.11	30.56	30.56	30.56
Mean (%)	40.92			35.46			33.52			32.03			31.02			30.56		

ตารางที่ 4.50 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	6	5	5	5	5	3	4	5	4	4	3	5	4	6	4	4	4	4
ฃ	18	16	17	16	6	12	15	11	12	11	11	12	11	10	11	10	10	10
ช	5	6	11	4	5	6	4	4	7	4	5	6	4	5	5	4	4	4
ฅ	9	8	7	7	4	4	4	4	2	2	3	3	3	12	2	2	2	2
ฉ	9	16	18	10	17	15	16	14	14	13	14	13	13	14	14	13	13	13
ฐ	13	18	15	14	13	10	13	11	11	11	13	13	14	14	12	14	14	14
ฎ	12	12	10	9	9	8	9	9	9	9	8	11	8	10	10	10	10	10
ณ	11	10	14	9	13	10	11	13	9	11	13	6	11	11	7	9	9	9
ด	9	12	19	6	14	17	11	16	14	13	12	15	12	14	15	14	14	14
ต	8	9	14	9	8	9	7	7	8	7	7	8	6	7	8	7	7	7
บ	9	11	8	9	12	8	9	9	6	7	8	6	8	9	8	8	8	8
ม	20	21	10	19	14	15	15	10	14	11	10	16	11	11	12	11	11	11
Total	129	144	157	117	130	117	118	113	110	103	107	114	105	113	108	106	106	106
%	35.83	40.00	43.61	32.50	36.11	32.50	32.78	31.39	30.56	28.61	29.72	31.67	29.17	31.39	30.00	29.44	29.44	29.44
Mean (%)	39.81			33.70			31.58			30.00			30.19			29.44		

ตารางที่ 4.51 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	5	3	4	3	3	3	4	3	5	3	4	3	2	5	3	2	2	2
ฃ	18	16	18	15	13	10	15	10	11	10	11	13	10	9	11	9	9	9
ช	4	7	11	2	5	7	3	5	5	4	5	6	4	5	6	4	4	4
ฌ	11	11	10	9	5	6	6	4	5	3	3	6	2	2	2	2	2	2
ฉ	11	16	14	8	19	14	13	18	11	15	14	13	13	14	12	13	13	13
ฎ	13	18	17	13	16	10	12	10	11	10	10	13	11	11	12	11	11	11
ฏ	12	13	9	8	8	11	8	9	7	8	8	10	8	10	10	10	10	10
ณ	12	12	14	9	11	12	11	11	8	10	11	6	8	9	8	9	9	9
ด	10	16	16	8	12	15	9	15	11	11	11	14	12	13	14	13	13	13
ต	6	10	11	9	9	6	7	8	6	7	8	7	7	8	7	7	7	7
บ	10	9	10	10	8	7	10	7	7	8	8	7	9	6	6	7	7	7
ม	20	19	17	18	18	12	15	14	15	12	10	15	9	13	11	10	10	10
Total	132	150	151	112	127	113	113	114	102	101	103	113	95	105	102	97	97	97
%	36.67	41.67	41.94	31.11	35.28	31.39	31.39	31.67	28.33	28.06	28.61	31.39	26.39	29.17	28.33	26.94	26.94	26.94
Mean (%)	40.09			32.59			30.46			29.35			27.96			26.94		

ตารางที่ 4.52 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	5	5	3	2	4	2	2	3	2	2	3	4	2	4	4	3	3	3
ฃ	14	14	7	15	12	7	13	9	9	11	9	10	10	9	10	10	10	10
ช	7	4	11	5	5	7	4	4	8	4	4	7	4	4	5	4	4	4
ซ	10	12	6	8	4	4	4	3	3	2	2	5	2	2	3	2	2	2
ฅ	14	13	16	15	13	16	14	12	13	13	12	12	13	12	12	13	13	13
ฉ	11	18	17	13	17	12	14	14	11	12	12	12	12	14	13	14	14	14
ฐ	7	8	9	8	9	11	8	9	9	8	7	9	7	7	7	7	7	7
ณ	11	8	12	8	9	14	7	11	10	8	10	9	8	9	9	8	8	8
ด	8	14	17	11	17	17	13	16	13	14	9	13	9	10	10	9	9	9
ต	8	5	9	3	4	5	2	3	4	2	2	6	2	4	6	4	4	4
บ	15	15	16	16	15	15	17	14	14	15	14	12	15	11	11	12	12	12
ม	19	19	19	17	17	17	16	15	16	14	14	14	12	13	13	11	11	11
Total	129	135	142	121	126	127	114	113	112	105	98	113	96	99	103	97	97	97
%	35.83	37.50	39.44	33.61	35.00	35.28	31.67	31.39	31.11	29.17	27.22	31.39	26.67	27.50	28.61	26.94	26.94	26.94
Mean (%)	37.59			34.63			31.39			29.26			27.59			26.94		

ตารางที่ 4.53 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	3	4	3	3	2	2	2	3	3	3	4	3	3	4	3	3	3	3
ฃ	15	13	9	14	10	9	10	8	10	10	9	10	9	9	10	10	10	10
ช	6	8	10	7	7	6	6	4	6	6	4	6	6	4	6	6	6	6
ฌ	14	14	9	8	7	3	7	4	5	3	4	6	3	3	5	3	3	3
ฉ	13	11	17	12	15	16	12	14	12	12	13	10	13	11	11	11	11	11
ฎ	9	17	20	15	15	14	15	13	11	11	11	12	12	14	12	15	15	15
ฏ	8	9	9	8	7	10	6	7	8	6	6	7	6	7	8	7	7	7
ณ	12	10	13	8	8	11	9	9	9	7	9	8	7	8	8	7	7	7
ด	10	12	5	10	16	16	12	15	12	13	12	13	9	12	10	10	10	10
ต	8	5	8	4	3	5	4	3	5	3	3	6	3	4	5	4	4	4
บ	16	16	14	17	14	16	16	14	15	15	12	12	13	11	11	12	12	12
ม	20	20	18	19	20	17	17	18	17	14	16	16	15	15	16	14	14	14
Total	134	139	145	125	124	125	116	112	113	103	103	109	99	102	105	102	102	102
%	37.22	38.61	40.28	34.72	34.44	34.72	32.22	31.11	31.39	28.61	28.61	30.28	27.50	28.33	29.17	28.33	28.33	28.33
Mean (%)	38.70			34.63			31.57			29.17			28.33			28.33		

ตารางที่ 4.54 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	7	4	3	3	3	3	3	2	5	3	2	5	4	3	4	4	4	4
ฃ	17	12	10	15	10	8	10	8	8	9	7	8	7	8	9	8	8	8
ช	7	7	11	4	8	10	6	4	5	6	4	7	6	4	6	6	6	6
ฅ	12	15	9	13	6	8	8	5	8	5	5	11	3	6	8	5	5	5
ฉ	11	8	9	12	14	17	12	13	10	11	12	10	11	15	11	13	13	13
ฐ	10	16	20	11	18	18	13	15	13	13	12	14	9	13	11	11	11	11
ฎ	9	9	11	8	6	9	7	8	8	7	5	6	5	7	7	7	7	7
ณ	14	9	14	7	9	10	8	8	9	8	8	8	8	6	8	7	7	7
ด	9	15	16	8	14	16	10	17	14	13	11	15	10	11	11	10	10	10
ต	6	4	8	3	3	3	4	3	5	4	2	7	3	4	5	14	14	14
บ	15	14	15	16	16	15	16	15	16	15	13	13	13	12	11	12	12	12
ม	20	23	20	20	20	19	18	18	15	17	12	15	16	18	18	15	15	15
Total	137	136	146	120	127	136	115	116	116	111	99	119	95	107	109	102	102	102
%	38.06	37.78	40.56	33.33	35.28	37.78	31.94	32.22	32.22	30.83	27.50	33.06	26.39	29.72	30.28	28.33	28.33	28.33
Mean (%)	38.80			35.46			32.13			30.46			28.80			28.33		

ตารางที่ 4.55 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k= 2 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	4	4	4	3	3	3	3	3	3	3	4	6	4	5	7	5	5	5
ฃ	15	13	9	13	10	6	12	9	8	10	9	9	10	9	11	10	10	10
ช	6	4	12	4	4	6	4	3	7	4	3	6	4	3	5	4	4	4
ฅ	12	11	5	7	4	3	4	3	2	2	1	3	1	1	1	1	1	1
ฉ	13	13	18	13	15	18	12	14	11	11	14	10	11	14	12	11	11	11
ฐ	10	16	17	12	16	11	12	12	10	10	12	11	12	14	12	14	14	14
ฎ	6	8	8	8	8	10	9	9	10	10	7	10	8	7	8	8	8	8
ณ	9	9	13	7	8	11	8	8	10	8	8	8	8	8	8	7	7	7
ด	7	12	14	9	14	15	10	15	11	12	10	12	9	10	11	9	9	9
ต	8	7	10	3	5	5	2	3	4	2	2	5	2	3	4	3	3	3
บ	7	12	11	11	13	10	12	11	8	10	12	9	11	11	10	10	10	10
ม	17	18	15	15	14	12	12	10	12	10	11	13	11	10	12	10	10	10
Total	114	127	136	105	114	110	100	100	96	92	93	102	91	95	101	92	92	92
%	31.67	35.27	37.78	29.17	31.67	30.56	27.78	27.78	26.67	25.56	25.83	29.33	25.38	26.39	28.06	25.56	25.56	25.56
Mean (%)	34.91			30.47			27.41			26.91			26.61			25.56		

ตารางที่ 4.56 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	5	4	3	4	2	3	3	2	4	3	2	4	3	2	5	3	3	3
ฃ	15	12	9	12	11	6	11	8	8	9	9	7	9	9	10	9	9	9
ช	4	6	9	3	7	4	4	3	6	3	4	6	3	4	5	3	3	3
ฌ	14	12	7	7	6	5	7	3	4	3	3	4	2	2	2	2	2	2
ฉ	10	9	18	10	12	17	10	12	11	9	12	10	11	10	11	9	9	9
ฎ	7	17	20	12	16	10	12	11	9	11	12	12	11	13	12	12	12	12
ฏ	8	10	8	7	6	7	7	7	7	8	5	7	5	6	6	6	6	6
ณ	10	9	14	8	7	9	9	8	11	9	8	9	8	8	8	8	8	8
ด	7	12	13	7	15	14	10	14	10	10	10	11	11	12	11	11	11	11
ต	7	8	10	4	5	3	4	3	5	3	2	6	3	4	5	4	4	4
บ	6	10	9	9	12	8	10	11	7	9	12	7	10	9	8	8	8	8
ม	18	19	16	16	15	11	13	12	11	10	11	12	11	11	11	10	10	10
Total	111	128	136	99	114	97	100	94	93	87	90	95	87	90	94	85	85	85
%	30.83	35.56	37.78	27.50	31.67	26.94	27.78	26.11	25.83	24.17	25.00	26.39	24.17	25.00	26.11	23.61	23.61	23.61
Mean (%)	34.72			28.70			26.57			25.19			25.09			23.61		

ตารางที่ 4.57 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	7	3	5	4	3	4	3	3	4	3	3	4	2	4	4	2	2	2
ฃ	17	14	11	13	10	7	12	9	9	10	8	9	7	8	8	7	7	7
ช	4	7	12	4	5	7	3	4	3	3	4	6	3	4	4	4	4	4
ฅ	10	12	8	10	5	6	7	6	7	6	3	8	2	2	4	2	2	2
ฉ	8	9	9	11	12	13	12	13	11	13	4	10	14	13	12	13	13	13
ฐ	9	15	18	9	16	15	11	13	12	9	11	13	9	13	10	9	9	9
ฎ	7	11	7	9	7	11	7	7	8	7	5	7	5	7	6	6	6	6
ณ	12	7	16	6	10	9	9	7	8	9	7	8	8	8	6	6	6	6
ด	11	13	13	11	13	16	12	13	12	12	10	13	10	11	12	11	11	11
ต	6	7	10	3	6	5	4	3	4	4	2	6	3	3	5	4	4	4
บ	7	10	10	9	11	8	8	8	6	8	7	7	8	8	9	8	8	8
ม	20	20	17	16	17	12	14	11	13	12	11	9	10	12	11	10	10	10
Total	118	128	136	105	115	113	102	97	97	96	85	100	81	93	91	82	82	82
%	32.78	35.56	37.78	29.17	31.94	31.39	28.33	26.94	26.94	26.67	23.61	27.78	22.50	25.83	25.28	22.78	22.78	22.78
Mean (%)	35.37			30.83			27.40			26.02			24.54			22.78		

ตารางที่ 4.58 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี Normal kernel

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	6	4	5	4	5	4	4	4	4	4	4	4	4	4	4	4	4	4
ฃ	20	21	20	19	19	19	17	17	17	15	17	17	18	18	16	17	17	17
ช	5	7	8	9	6	7	5	5	5	4	4	4	3	3	3	3	3	3
ฅ	15	15	15	14	14	14	13	13	13	14	13	14	13	13	13	14	14	14
ฉ	10	9	10	8	7	9	6	8	8	8	8	8	7	7	7	7	7	7
ฐ	16	15	16	15	15	15	15	15	15	14	15	15	17	16	16	16	16	16
ฎ	15	13	14	11	12	11	10	10	11	12	12	11	11	10	10	9	9	9
ณ	14	12	14	10	10	12	11	11	11	10	10	10	10	10	10	10	10	10
ด	18	18	15	17	13	13	10	11	10	9	10	9	8	8	7	5	5	5
ต	4	4	6	7	7	7	5	5	6	6	6	6	7	7	7	3	3	3
บ	0	3	4	3	3	4	3	3	3	1	2	1	1	2	1	2	2	2
ม	20	20	20	19	20	19	19	18	17	17	17	17	16	16	16	17	17	17
Total	143	141	147	136	131	135	118	120	120	114	118	116	115	114	110	107	107	107
%	39.72	39.17	40.83	37.78	36.39	37.50	32.78	33.33	33.33	31.67	32.78	32.22	31.94	31.67	30.56	29.72	29.72	29.72
Mean (%)	39.91			37.22			33.15			32.22			31.39			29.72		

ตารางที่ 4.59 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี LSCV

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	4	5	4	4	3	4	3	3	3	4	4	4	4	4	4	4	4	4
ฃ	19	18	20	17	16	17	16	17	16	17	17	17	17	16	16	15	15	15
ช	2	3	2	3	4	3	4	4	2	2	3	2	4	2	3	1	1	1
ฌ	18	18	18	15	16	15	18	17	18	17	17	17	16	16	16	16	16	16
ฎ	8	10	8	12	13	13	10	10	10	13	15	14	13	15	14	15	15	15
ฎ	17	16	17	18	17	18	18	17	16	16	16	15	15	15	16	17	17	17
ฏ	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
ณ	11	11	10	11	11	11	11	11	10	9	9	9	9	9	9	9	9	9
ด	12	11	11	11	10	10	12	10	11	9	8	9	9	10	9	12	12	12
ต	9	9	10	10	10	9	7	6	9	9	5	7	5	4	3	4	4	4
บ	2	2	2	1	1	1	1	0	1	1	1	0	0	1	1	1	1	1
ม	20	19	20	20	19	19	20	18	19	19	18	19	18	17	18	17	17	17
Total	134	134	134	134	132	132	132	125	127	128	125	122	122	121	121	123	123	123
%	37.22	37.22	37.22	37.22	36.67	36.67	36.67	34.72	35.27	35.56	34.72	33.89	33.89	33.61	33.61	34.17	34.17	34.17
Mean (%)	37.22			36.85			35.55			34.72			33.70			34.17		

ตารางที่ 4.60 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 200 ค่าด้วยวิธี LKCV

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	6	5	5	4	4	4	5	4	5	5	4	4	4	4	4	4	4	4
ฃ	20	20	19	20	19	20	20	18	19	20	18	18	18	18	19	18	18	18
ช	3	5	4	6	3	4	3	3	3	2	2	2	2	2	2	2	2	2
ฅ	12	12	12	13	15	15	18	15	16	15	15	15	16	15	15	15	15	15
ฉ	10	9	9	6	5	6	4	6	6	10	10	9	10	9	10	9	9	9
ง	11	11	10	18	18	17	20	17	18	15	14	15	15	13	14	12	12	12
จ	17	14	15	8	9	9	9	9	9	11	10	11	10	9	11	9	9	9
ฉ	13	12	13	10	11	10	11	10	11	9	10	9	10	9	9	9	9	9
ด	20	18	17	15	13	14	6	8	7	8	8	7	5	6	6	6	6	6
ต	3	5	4	6	6	5	6	7	7	8	5	7	8	7	6	5	5	5
บ	0	2	1	2	3	2	3	3	3	1	2	1	2	1	1	1	1	1
ม	20	20	20	20	20	20	20	19	19	19	19	19	18	18	18	18	18	18
Total	135	133	129	128	126	126	124	119	123	123	117	117	118	111	115	108	108	108
%	37.50	36.94	35.83	35.56	35.00	35.00	34.44	33.06	34.17	34.17	32.50	32.50	32.78	30.83	31.94	30.00	30.00	30.00
Mean (%)	36.76			35.19			33.89			33.06			31.85			30.00		

จากตารางที่ 4.46-4.48 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Euclidean พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 50 ที่ $k = 5$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 23.61%

จากตารางที่ 4.49-4.51 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Cityblock พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 5$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 26.94%

จากตารางที่ 4.52-4.54 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Correlation พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะเพิ่มขึ้น สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 2$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 26.94%

จากตารางที่ 4.55-4.57 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 200 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Cosine พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 5$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 22.78%

จากตารางที่ 4.58-4.60 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 200 ค่าด้วยวิธี Normal kernel พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 29.72%

จากการศึกษาการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 200 ค่า สรุปได้ว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง ซึ่งวิธีที่ให้ประสิทธิภาพในการจำแนกกลุ่มดีที่สุด คือ วิธี K-nearest neighbour ที่ $k = 5$ โดยวัดระยะทางแบบ Cosine ที่จำนวนตัวอย่างเรียนรู้ขนาด 60 ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 22.78%

ตารางที่ 4.61 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	4	2	4	2	2	4	2	3	2	2	4	3	3	4	3	4	4	4
ฃ	10	16	15	12	15	10	10	11	7	9	10	7	8	11	6	8	8	8
ช	7	5	8	4	4	3	3	2	5	3	3	4	3	4	2	3	3	3
ฅ	13	6	6	5	4	3	5	2	3	2	3	4	2	4	3	3	3	3
ฉ	12	10	17	8	9	16	7	9	11	7	9	12	7	9	12	8	8	8
ฐ	11	17	9	16	13	7	11	11	8	10	12	7	11	12	9	11	11	11
ฎ	12	8	17	9	6	7	8	7	7	8	5	8	7	6	7	7	7	7
ณ	9	7	8	8	5	2	6	3	6	5	4	3	6	4	4	5	5	5
ด	9	10	13	7	9	11	10	10	9	9	9	11	8	8	10	8	8	8
ต	7	6	6	5	3	4	1	3	3	2	5	3	3	5	4	3	3	3
บ	5	11	10	4	10	7	5	7	4	3	7	6	4	8	6	6	6	6
ม	16	11	12	10	10	10	10	7	8	6	6	6	6	5	7	5	5	5
Total	115	109	115	90	90	84	78	75	73	66	77	74	68	80	73	71	71	71
%	31.94	30.28	31.94	25.00	25.00	23.33	21.67	20.83	20.28	18.33	21.39	20.56	18.89	22.22	20.28	19.72	19.72	19.72
Mean (%)	31.39			24.44			20.93			20.09			20.46			19.72		

ตารางที่ 4.62 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	4	2	4	2	2	3	2	2	2	2	3	4	2	4	6	3	3	3
ฃ	10	15	13	10	14	10	10	11	7	10	11	7	10	11	6	9	9	9
ช	11	7	10	1	6	6	2	4	3	2	4	2	2	4	2	2	2	2
ซ	9	10	7	6	5	4	7	2	4	3	3	4	2	3	2	2	2	2
ฅ	10	8	13	5	9	13	9	10	10	9	9	11	9	9	11	9	9	9
ฉ	9	17	13	13	10	5	11	10	7	11	9	9	11	8	9	12	12	12
ฐ	12	6	9	9	6	9	8	6	10	8	4	10	8	6	8	9	9	9
ณ	7	8	11	8	7	5	6	4	2	4	4	3	4	5	2	5	5	5
ด	9	11	13	8	8	12	8	10	12	10	9	11	9	9	10	9	9	9
ต	6	5	6	4	3	4	2	2	5	2	3	4	2	2	5	2	2	2
บ	5	7	9	5	9	6	5	6	3	3	7	2	5	6	4	4	4	4
ม	19	12	13	10	9	10	9	8	7	6	6	6	7	6	6	6	6	6
Total	111	108	121	81	88	87	79	75	72	70	72	73	71	73	71	72	72	72
%	30.83	30.00	33.61	22.50	24.44	24.17	21.94	20.83	20.00	19.44	20.00	20.28	19.72	20.28	19.72	20.00	20.00	20.00
Mean (%)	31.33			23.70			20.92			19.91			19.91			20.00		

ตารางที่ 4.63 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	6	2	4	2	3	4	2	2	3	2	3	3	2	4	4	3	3	3
ฃ	16	17	13	11	14	8	11	11	9	9	9	8	9	9	8	10	10	10
ช	8	9	12	3	8	7	5	4	5	2	3	5	2	2	3	2	2	2
ฅ	11	11	5	4	6	6	6	2	5	3	2	5	2	2	4	3	3	3
ฉ	10	11	13	7	7	11	4	7	9	8	9	10	8	8	11	9	9	9
ฐ	9	18	17	9	15	8	8	7	6	7	7	6	6	7	5	6	6	6
ฎ	13	6	12	7	6	10	8	7	11	8	5	10	7	5	9	7	7	7
ณ	9	7	12	7	8	1	6	4	3	3	4	1	3	3	1	2	2	2
ด	9	12	14	8	11	12	8	13	10	8	8	10	5	8	6	5	5	5
ต	16	9	7	2	4	3	1	4	2	2	4	1	2	2	3	1	1	1
บ	3	7	9	5	8	4	4	5	3	3	5	4	4	3	4	3	3	3
ม	18	13	18	14	10	9	9	7	7	6	6	5	5	5	5	4	4	4
Total	118	122	136	79	100	83	72	73	73	61	65	68	55	58	63	55	55	55
%	32.78	33.89	37.78	21.94	27.78	23.06	20.00	20.28	20.28	16.94	18.06	18.89	15.28	16.11	17.50	15.28	15.28	15.28
Mean (%)	34.82			24.26			20.19			17.96			16.30			15.28		

ตารางที่ 4.64 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	7	4	3	4	4	2	4	4	4	4	4	4	4	4	3	5	5	5
ฃ	13	16	17	15	16	14	13	14	9	9	13	12	9	13	11	11	11	11
ช	4	8	7	3	5	5	2	6	2	3	4	4	2	5	2	3	3	3
ฌ	10	8	9	5	4	3	5	2	3	2	2	3	1	3	2	2	2	2
ฉ	11	11	17	7	13	14	11	13	10	10	12	10	9	11	9	9	9	9
ฎ	13	18	13	13	14	10	12	13	10	12	13	10	12	13	11	12	12	12
ฏ	14	11	7	11	7	7	8	9	8	8	8	8	7	9	7	7	7	7
ถ	11	13	11	11	8	10	7	7	8	7	7	6	7	6	6	6	6	6
ด	11	11	17	8	11	13	10	10	12	10	8	11	8	7	8	7	7	7
ต	9	8	9	8	6	9	6	5	7	5	6	7	5	7	7	5	5	5
บ	7	12	8	9	11	12	8	12	9	10	11	6	9	10	8	8	8	8
ม	18	16	17	12	14	10	11	10	9	9	7	10	9	8	8	9	9	9
Total	128	136	135	106	113	109	97	105	91	89	95	91	82	96	82	84	84	84
%	35.56	37.78	37.50	29.44	31.39	30.28	26.94	29.17	25.28	24.72	26.39	25.28	22.78	26.67	22.78	23.33	23.33	23.33
Mean (%)	36.95			30.37			27.13			25.46			24.08			23.33		

ตารางที่ 4.65 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	7	3	3	4	3	2	3	3	4	3	3	5	3	3	3	3	3	3
ฃ	15	16	17	16	13	13	14	12	9	10	12	11	10	13	11	12	12	12
ช	7	8	9	2	8	7	3	6	4	4	6	3	4	6	3	4	4	4
ฅ	10	9	11	5	6	3	5	3	3	3	2	4	1	2	2	2	2	2
ฉ	11	9	16	5	13	15	10	14	13	10	13	13	9	12	11	9	9	9
ฐ	13	18	13	10	14	10	10	12	10	10	11	10	10	12	10	10	10	10
ฎ	14	9	9	9	6	9	7	8	10	8	8	9	8	9	8	9	9	9
ณ	12	10	10	10	8	7	8	6	6	6	7	6	7	7	7	7	7	7
ด	12	13	16	8	11	12	9	9	10	8	8	10	8	7	7	6	6	6
ต	9	7	9	8	4	7	5	6	4	5	5	5	4	7	6	5	5	5
บ	5	13	8	8	10	10	6	9	8	6	9	4	6	8	5	7	7	7
ม	17	16	16	10	14	10	13	10	9	9	8	12	8	9	10	9	9	9
Total	132	131	137	95	110	105	93	98	90	82	92	92	78	95	83	83	83	83
%	36.67	36.39	38.06	26.39	30.56	29.17	25.83	27.22	25.00	22.78	25.56	25.56	21.67	26.39	23.06	23.06	23.06	23.06
Mean (%)	37.04			28.71			26.02			24.63			23.71			23.06		

ตารางที่ 4.66 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	4	2	4	2	3	3	1	3	3	2	2	3	2	3	3	3	3	3
ฃ	14	19	16	15	15	11	12	13	8	8	11	17	8	9	8	8	8	8
ช	6	11	9	2	6	7	4	11	4	5	7	4	2	6	2	2	2	2
ฅ	10	8	11	7	7	7	5	6	6	3	3	4	2	3	4	3	3	3
ฉ	13	9	14	7	8	12	5	7	9	10	12	11	8	12	9	8	8	8
ฐ	13	18	14	13	12	10	9	17	12	11	10	11	10	11	8	12	12	12
ฎ	15	12	11	11	6	9	8	10	11	10	8	10	9	8	9	9	9	9
ณ	12	10	14	5	8	7	6	8	7	6	8	5	7	7	6	7	7	7
ด	11	12	12	8	11	14	11	14	11	8	8	10	7	8	5	5	5	5
ต	7	6	8	4	4	9	1	5	7	4	4	5	5	6	4	4	4	4
บ	5	11	8	8	11	8	8	9	9	11	9	5	8	7	5	6	6	6
ม	19	12	12	10	12	11	12	12	11	10	8	12	7	11	9	8	8	8
Total	129	130	133	92	103	108	82	115	108	88	90	97	75	91	72	75	75	75
%	35.83	36.11	36.94	25.56	28.61	30.00	22.78	31.94	30.00	24.44	25.00	26.94	20.83	25.28	20.00	20.83	20.83	20.83
Mean (%)	36.14			28.06			28.24			25.46			22.04			20.83		

ตารางที่ 4.67 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	3	3	4	0	3	4	2	3	3	3	2	4	2	3	3	3	3	3
ฃ	10	9	11	8	10	6	8	4	6	5	5	7	6	5	7	6	6	6
ช	10	8	10	4	5	7	3	4	9	3	4	8	2	4	4	2	2	2
ฅ	12	9	9	7	5	3	5	2	3	3	2	6	3	3	4	4	4	4
ฉ	14	10	13	7	8	12	7	9	7	6	9	7	6	7	7	6	6	6
ฐ	17	19	16	16	16	13	16	14	13	14	14	13	14	13	12	13	13	13
ฎ	16	10	9	10	8	9	8	8	9	8	7	8	7	6	7	6	6	6
ณ	8	9	8	7	8	9	6	9	5	7	8	5	7	8	6	7	7	7
ด	8	16	14	10	14	13	12	13	13	13	8	11	8	8	8	8	8	8
ต	17	1	2	2	1	2	0	1	4	0	2	5	2	2	6	2	2	2
บ	12	13	17	11	11	13	9	12	10	10	10	6	10	7	7	7	7	7
ม	19	16	15	18	13	12	14	10	12	11	9	13	10	9	12	10	10	10
Total	136	123	128	100	102	103	90	89	94	83	80	93	77	75	83	74	74	74
%	37.78	34.17	35.56	27.78	28.33	28.61	25.00	24.72	26.11	23.06	22.22	25.83	21.39	20.83	23.06	20.56	20.56	20.56
Mean (%)	35.84			28.24			25.28			23.70			21.76			20.56		

ตารางที่ 4.68 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	2	1	3	0	1	3	1	2	2	2	2	3	2	2	4	2	2	2
ฃ	12	9	9	9	9	6	8	4	6	5	5	7	6	5	7	6	6	6
ช	10	11	12	4	8	7	2	5	5	4	4	3	3	4	4	3	3	3
ฅ	13	10	10	6	5	5	5	2	6	3	4	6	3	4	6	4	4	4
ฉ	12	9	13	7	9	12	9	11	7	8	13	7	7	12	7	7	7	7
ฐ	16	18	18	16	14	12	15	12	12	12	13	4	13	14	12	14	14	14
ฎ	15	11	10	11	10	1	9	9	11	8	9	10	8	7	8	6	6	6
ณ	8	9	11	7	8	10	5	8	5	5	7	5	4	7	5	4	4	4
ด	9	18	15	1	16	14	10	13	11	11	8	10	7	8	7	7	7	7
ต	5	1	4	1	0	2	1	0	5	1	1	5	2	2	6	3	3	3
บ	17	13	12	12	12	15	10	13	14	12	11	7	12	7	6	8	8	8
ม	20	18	20	18	15	14	17	10	12	12	12	14	14	10	14	12	12	12
Total	139	128	137	102	107	111	92	89	96	83	89	81	81	82	86	76	76	76
%	38.61	35.56	38.06	28.33	29.72	30.83	25.56	24.72	26.67	23.06	24.72	22.50	22.50	22.78	23.89	21.11	21.11	21.11
Mean (%)	37.41			29.63			25.65			23.43			23.06			21.11		

ตารางที่ 4.69 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	3	2	2	0	1	4	1	2	3	2	2	4	2	2	4	2	2	2
ฃ	12	11	8	10	7	6	7	6	6	5	5	7	5	6	7	6	6	6
ช	12	8	13	4	7	9	4	6	10	4	5	9	3	4	5	3	3	3
ฅ	14	13	8	11	8	5	9	3	8	4	3	8	4	3	5	3	3	3
ฉ	13	8	12	5	8	12	9	12	9	10	11	7	9	10	9	9	9	9
ฐ	13	17	20	16	19	16	15	14	13	14	14	14	14	3	14	14	14	14
ฎ	15	8	9	11	9	12	10	10	12	10	10	11	10	9	10	9	9	9
ณ	12	7	10	7	8	9	6	7	5	6	7	5	7	7	6	6	6	6
ด	8	15	5	11	17	12	12	12	11	10	9	10	7	8	10	18	18	18
ต	9	3	3	0	1	3	0	0	2	0	0	2	1	0	5	2	2	2
บ	16	15	16	13	13	16	9	15	12	14	11	9	12	9	8	9	9	9
ม	20	20	20	20	19	17	17	17	17	15	14	16	13	12	16	13	13	13
Total	147	127	136	108	117	121	99	104	108	94	91	102	87	83	99	84	84	84
%	40.83	35.28	37.78	30.00	32.50	33.61	27.50	28.89	30.00	26.11	25.28	28.33	24.17	23.06	27.50	23.33	23.33	23.3
Mean (%)	37.96			32.04			28.80			26.57			24.91			23.33		

ตารางที่ 4.70 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k=2$ โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ช	4	2	4	1	4	4	3	2	2	2	2	3	2	3	3	3	3	3
ฌ	10	14	14	12	13	10	9	10	8	9	10	10	9	10	9	9	9	9
ช	8	5	9	4	4	5	3	3	6	3	3	6	2	4	3	3	3	3
ช	13	7	7	6	5	2	5	2	4	4	2	5	3	3	4	4	4	4
ฌ	12	19	11	7	7	13	6	8	8	6	8	8	6	7	8	6	6	6
ฎ	12	17	12	16	13	11	13	10	10	11	11	10	12	11	10	12	12	12
ฎ	12	7	7	9	6	8	8	8	8	8	7	8	7	6	7	7	7	7
ณ	8	7	7	8	7	5	5	6	4	4	7	3	5	6	4	5	5	5
ด	8	13	12	6	10	10	9	10	10	10	8	10	8	7	7	7	7	7
ต	6	4	4	3	2	2	1	1	3	1	3	4	1	3	4	2	2	2
บ	9	9	10	5	10	9	6	9	7	7	10	6	8	8	7	7	7	7
ม	17	15	13	13	12	9	11	8	7	6	6	6	6	6	7	6	6	6
Total	119	109	110	90	93	88	79	77	77	71	77	79	69	74	73	71	71	71
%	33.06	30.27	30.56	25.00	25.83	24.44	21.94	21.39	21.39	19.72	21.39	21.94	19.17	20.56	20.28	19.72	19.72	19.72
Mean (%)	31.30			25.09			21.57			21.02			20.00			19.72		

ตารางที่ 4.71 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	4	1	3	1	2	2	2	2	2	2	2	3	2	2	4	2	2	2
ฃ	10	13	13	10	11	9	8	8	6	8	8	7	8	8	6	8	8	8
ช	10	9	10	3	6	6	1	5	3	2	4	2	2	4	2	2	2	2
ฅ	12	10	10	4	6	2	6	2	4	3	2	4	3	4	4	4	4	4
ฉ	9	7	8	5	10	11	8	9	9	8	9	9	8	8	9	8	8	8
ฐ	10	17	14	16	13	10	13	11	11	12	11	10	11	12	11	12	12	12
ฎ	10	6	7	8	7	10	8	9	11	8	9	10	8	6	7	7	7	7
ณ	9	8	12	8	7	5	5	4	3	4	4	4	4	4	4	4	4	4
ด	8	13	12	7	12	12	8	12	12	10	8	11	8	7	8	8	8	8
ต	6	5	3	2	2	2	1	1	4	1	2	4	1	2	5	2	2	2
ป	8	10	9	6	9	9	7	8	6	5	8	5	6	7	6	6	6	6
ม	20	15	14	15	11	10	13	10	9	8	8	8	9	7	7	7	7	7
Total	116	114	115	85	96	88	80	81	80	71	75	77	70	71	73	70	70	70
%	32.22	31.67	31.94	23.61	26.67	24.44	22.22	22.50	22.22	19.72	20.83	21.39	19.44	19.72	20.28	19.44	19.44	19.44
Mean (%)	31.94			24.91			22.31			20.65			19.81			19.44		

ตารางที่ 4.72 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ช	5	1	4	1	2	3	2	2	3	2	2	3	2	3	3	2	2	2
ฌ	12	13	11	11	12	8	11	8	7	8	5	7	7	6	7	7	7	7
ช	11	7	12	2	6	7	4	4	6	3	4	5	3	3	3	3	3	3
ช	12	12	7	9	8	8	7	2	8	3	2	6	2	3	5	4	4	4
ฌ	9	10	9	4	8	11	7	9	10	9	7	9	7	6	9	8	8	8
ฎ	10	18	20	13	15	15	13	11	10	12	11	13	9	11	10	11	11	11
ฎ	11	6	9	8	4	10	9	9	9	9	9	9	9	7	9	8	8	8
ณ	9	6	13	5	8	7	5	6	4	5	5	3	5	5	4	4	4	4
ด	7	13	15	9	13	12	9	13	10	9	9	9	7	6	6	5	5	5
ต	6	5	5	2	3	4	0	1	1	1	1	2	2	1	3	3	3	3
บ	7	10	12	5	10	8	7	9	7	8	7	5	7	6	5	4	4	4
ม	20	17	19	19	13	11	13	9	10	9	8	9	7	6	6	5	5	5
Total	119	118	136	88	102	104	87	83	85	78	70	80	67	63	70	64	64	64
%	33.06	32.78	37.78	24.44	28.33	28.89	24.17	23.06	23.61	21.67	19.44	22.22	18.61	17.50	19.44	17.78	17.78	17.78
Mean (%)	34.53			27.22			23.61			21.11			18.52			17.78		

ตารางที่ 4.73 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี Normal kernel

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	7	7	6	5	5	6	4	5	6	5	5	5	5	5	4	4	4	4
ฃ	15	15	14	17	16	15	14	14	14	17	16	15	13	13	13	14	14	14
ช	8	8	10	6	9	9	4	4	7	5	5	8	4	4	6	7	7	7
ฅ	3	5	9	7	7	7	10	10	6	7	7	7	10	8	8	9	9	9
ฉ	6	6	8	3	5	6	2	4	7	3	3	3	4	4	4	4	4	4
ง	7	7	7	10	10	10	11	9	11	7	7	7	6	6	6	6	6	6
จ	12	10	8	5	5	5	4	4	4	4	4	4	3	3	3	3	3	3
ฉ	7	7	7	6	6	3	4	5	4	6	4	5	6	4	4	2	2	2
ด	14	14	13	14	13	8	7	7	7	8	8	8	7	7	7	8	8	8
ต	1	1	1	3	3	3	2	2	2	6	2	2	1	1	1	1	1	1
บ	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
ม	20	19	20	17	17	18	19	17	16	17	17	17	13	14	14	18	18	18
Total	100	99	103	93	96	91	81	81	85	85	78	81	72	69	70	78	78	78
%	27.78	27.50	28.61	25.83	26.67	25.28	22.50	22.50	23.61	23.61	21.67	22.50	20.00	19.17	19.44	21.67	21.67	21.67
Mean (%)	27.96			25.93			22.87			22.59			19.54			21.67		

ตารางที่ 4.74 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี LSCV

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	5	5	5	5	5	5	4	4	5	5	5	5	5	5	5	5	5	5
ฃ	20	19	20	19	19	19	20	20	20	19	20	20	19	19	19	19	19	19
ช	3	5	3	6	4	5	4	4	5	5	4	5	4	4	4	4	4	4
ฅ	3	3	3	2	3	2	3	3	3	3	3	3	3	3	3	3	3	3
ฉ	11	11	12	10	10	11	11	10	12	12	12	11	12	11	12	12	12	12
ฐ	18	17	18	15	16	14	16	15	16	12	10	11	9	6	8	6	6	6
ฎ	18	15	16	10	9	9	9	8	9	9	9	10	7	9	8	9	9	9
ณ	10	9	10	6	8	7	7	8	7	8	7	7	6	7	7	7	7	7
ด	19	19	19	19	19	19	18	16	18	16	16	16	15	16	15	16	16	16
ต	4	5	4	5	5	5	4	5	3	4	6	5	5	6	6	6	6	6
บ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ม	18	18	19	19	18	18	15	14	15	16	15	17	18	15	17	10	10	10
Total	129	126	129	116	116	114	111	107	113	109	107	110	107	101	104	97	97	97
%	35.83	35.00	35.83	32.22	32.22	31.67	30.83	29.72	31.89	30.28	29.72	30.56	29.72	28.06	28.89	26.94	26.94	26.94
Mean (%)	35.55			32.04			30.81			30.19			28.89			26.94		

ตารางที่ 4.75 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 500 ค่าด้วยวิธี LKCV

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	7	6	6	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
ฃ	20	20	20	19	19	20	19	19	19	19	18	19	18	17	18	17	17	17
ช	10	9	10	10	8	10	5	7	6	8	6	6	8	7	6	6	6	6
ฅ	7	8	8	10	10	10	9	10	9	11	9	9	10	9	10	9	9	9
ฉ	9	9	8	7	7	7	5	6	6	7	7	7	10	7	8	8	8	8
ง	9	9	9	7	9	7	5	5	5	8	9	7	10	9	9	9	9	9
จ	17	18	18	18	17	17	18	16	15	12	11	12	8	10	9	8	8	8
ฉ	12	10	11	7	7	7	8	8	7	7	7	7	7	7	7	7	7	7
ด	18	17	18	17	18	17	15	16	15	16	15	15	13	14	14	14	14	14
ต	4	5	4	5	5	4	5	5	5	6	6	3	3	3	3	3	3	3
บ	0	2	1	1	1	1	2	1	2	1	2	1	2	2	2	2	2	2
ม	19	19	18	18	18	18	19	18	18	17	17	18	16	15	15	13	13	13
Total	132	132	131	124	124	123	115	116	112	117	112	109	110	105	106	101	101	101
%	36.67	36.67	36.39	34.44	34.44	34.17	31.94	32.22	31.11	32.50	31.11	30.28	30.56	29.17	29.44	28.06	28.06	28.06
Mean (%)	36.58			34.35			31.76			31.30			29.72			28.06		

จากตารางที่ 4.61-4.63 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Euclidean พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 5$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 15.28%

จากตารางที่ 4.64-4.66 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Cityblock พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 5$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 20.83%

จากตารางที่ 4.67-4.69 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Correlation พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะเพิ่มขึ้น สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 2$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 20.56%

จากตารางที่ 4.70-4.72 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 500 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Cosine พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 5$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 17.78%

จากตารางที่ 4.73-4.75 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 500 ค่าด้วยวิธี Normal kernel พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 50 ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 19.54%

จากการศึกษาการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 500 ค่า สรุปได้ว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง ซึ่งวิธีที่ให้ประสิทธิภาพในการจำแนกกลุ่มดีที่สุด คือ วิธี K-nearest neighbour ที่ $k = 2$ โดยวัดระยะทางแบบ Euclidean ที่จำนวนตัวอย่างเรียนรู้ขนาด 60 ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 15.28%

ตารางที่ 4.76 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	6	3	5	3	3	3	2	3	4	2	3	5	2	4	4	3	3	3
ฃ	13	15	16	12	13	14	12	12	11	12	12	9	11	13	9	10	10	10
ช	6	12	6	9	7	6	7	9	3	5	9	2	5	9	2	4	4	4
ฅ	14	9	10	6	7	6	6	4	7	3	5	9	2	7	7	4	4	4
ฉ	8	8	13	5	9	16	7	12	10	8	11	9	7	10	9	7	7	7
ฐ	9	18	9	16	13	13	12	14	9	12	12	11	12	11	10	11	11	11
ฎ	9	7	9	8	7	9	8	7	10	8	7	9	6	7	6	6	6	6
ณ	9	10	11	9	7	5	6	5	6	5	5	5	5	5	5	5	5	5
ด	11	12	15	9	11	12	8	10	11	8	8	12	7	9	11	8	8	8
ต	8	7	6	5	4	5	4	3	5	4	2	4	3	2	3	3	3	3
บ	5	9	9	8	8	9	6	8	6	7	7	7	7	8	7	8	8	8
ม	15	12	11	12	10	9	10	9	7	8	7	6	7	5	5	5	5	5
Total	113	122	120	102	99	107	88	96	89	83	88	88	74	90	78	74	74	74
%	31.39	33.89	33.33	28.33	27.50	29.72	24.44	26.67	24.72	23.06	24.44	24.44	20.56	25.00	21.67	20.56	20.56	20.56
Mean (%)	32.87			28.52			25.28			23.98			22.41			20.56		

ตารางที่ 4.77 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	7	4	3	6	3	3	3	2	2	3	3	4	2	4	4	3	3	3
ฃ	13	18	14	13	14	13	11	12	12	12	11	10	12	11	9	11	11	11
ช	5	8	6	5	9	4	2	7	4	4	9	2	6	6	2	3	3	3
ฅ	13	7	11	9	9	7	6	4	4	3	4	4	3	4	5	4	4	4
ฉ	6	6	14	6	8	13	7	11	9	7	9	8	8	9	9	9	9	9
ฐ	10	18	14	11	11	14	11	15	7	12	13	8	11	13	9	11	11	11
ฎ	13	7	11	7	7	11	9	8	12	9	7	10	8	7	7	7	7	7
ณ	10	12	10	9	7	5	8	4	5	5	4	3	5	4	4	5	5	5
ด	10	12	15	8	9	13	9	11	13	10	8	13	8	8	8	7	7	7
ต	7	8	6	6	5	5	3	3	3	3	3	3	4	3	4	4	4	4
ป	9	8	8	6	6	9	5	7	5	6	8	5	7	8	6	7	7	7
ม	19	11	11	12	10	9	11	8	7	7	6	8	6	5	5	5	5	5
Total	122	119	123	98	98	96	85	92	83	81	85	78	80	82	72	76	76	76
%	33.89	33.06	34.17	27.22	27.22	26.67	23.61	25.56	23.06	22.50	23.61	21.67	22.22	22.78	20.00	21.11	21.11	21.11
Mean (%)	33.71			27.04			24.08			22.59			21.67			21.11		

ตารางที่ 4.78 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Euclidean

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	3	5	5	5	2	3	2	2	5	2	1	6	2	3	3	3	3	3
ฃ	14	18	13	13	13	11	12	11	11	12	12	10	11	10	9	10	10	10
ช	6	10	9	4	5	5	5	5	4	7	8	3	7	6	4	4	4	4
ฅ	13	9	14	10	9	8	8	6	9	4	6	8	3	5	5	3	3	3
ฉ	9	5	11	3	6	13	5	7	8	7	8	8	5	8	8	6	6	6
ฐ	9	23	17	14	17	16	11	14	11	11	9	9	7	9	7	9	9	9
ฎ	13	4	14	7	6	13	8	8	13	10	7	11	9	7	9	8	8	8
ณ	9	10	11	7	12	5	7	5	2	3	5	2	4	5	4	4	4	4
ด	12	13	12	8	11	13	6	11	10	8	8	11	7	9	10	8	8	8
ต	8	8	7	4	5	7	3	2	3	2	2	3	2	2	3	3	3	3
บ	9	8	8	6	6	6	6	7	6	7	7	6	6	7	5	7	7	7
ม	17	11	13	9	10	9	10	10	9	10	8	10	7	6	5	5	5	5
Total	122	124	134	90	102	109	83	88	91	83	81	87	70	77	72	70	70	70
%	33.89	34.44	37.22	25.00	28.33	30.28	23.06	24.44	25.28	23.06	22.50	24.17	19.44	21.39	20.00	19.44	19.44	19.44
Mean (%)	35.18			27.87			24.26			23.24			20.28			19.44		

ตารางที่ 4.79 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	7	4	5	1	2	3	1	2	4	2	2	5	2	2	3	2	2	2
ฃ	12	15	17	10	14	18	11	15	13	12	13	13	11	13	12	11	11	11
ช	8	13	8	9	11	9	7	13	4	6	12	5	6	11	5	7	7	7
ฅ	13	8	17	7	8	7	6	5	7	3	4	8	1	4	5	2	2	2
ฉ	8	5	16	4	9	16	6	11	13	8	11	13	8	10	13	9	9	9
ฐ	15	21	13	17	14	14	12	15	13	13	15	15	13	14	13	13	13	13
ฎ	15	12	11	11	9	13	11	10	12	10	10	11	10	12	11	10	10	10
ณ	9	11	14	9	12	11	8	9	8	7	9	7	8	9	7	7	7	7
ด	12	14	17	11	15	15	11	14	13	11	10	14	10	11	12	10	10	10
ต	11	10	10	5	8	10	6	6	8	4	5	6	3	4	5	2	2	2
บ	10	12	12	10	11	14	9	11	9	8	9	7	8	9	7	8	8	8
ม	7	13	17	14	13	11	12	10	9	9	9	9	9	9	9	9	9	9
Total	137	138	157	108	126	141	100	121	113	93	109	113	89	108	102	90	90	90
%	38.06	38.33	43.61	30.00	35.00	39.17	27.78	33.61	31.39	25.83	30.28	31.39	24.72	30.00	28.33	25.00	25.00	25.00
Mean (%)	40.00			34.72			30.93			29.17			27.68			25.00		

ตารางที่ 4.80 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	7	3	5	2	3	6	2	2	4	2	2	5	2	2	2	2	2	2
ฃ	13	16	16	12	13	16	11	13	13	12	14	12	13	13	13	13	13	13
ช	7	11	6	6	9	9	3	12	6	6	14	3	7	10	3	4	4	4
ฅ	10	9	16	8	10	7	6	5	5	3	4	6	2	3	3	2	2	2
ฉ	5	5	14	2	9	15	4	10	13	7	9	12	7	8	12	7	7	7
ฐ	13	20	16	17	15	12	13	15	12	13	13	14	12	12	14	12	12	12
ฎ	17	12	11	12	8	13	10	10	13	10	9	11	11	8	7	8	8	8
ณ	9	11	13	11	10	11	8	9	6	6	9	5	6	8	6	6	6	6
ด	11	15	17	10	13	15	10	13	12	10	11	13	9	13	11	10	10	10
ต	10	10	10	8	9	8	6	6	7	6	7	6	6	6	6	4	4	4
บ	8	11	10	9	11	10	9	11	7	8	9	8	8	9	8	8	8	8
ม	18	11	17	11	13	12	11	10	11	9	9	9	9	9	9	9	9	9
Total	128	134	151	108	123	134	93	116	109	92	110	104	92	101	94	85	85	85
%	35.56	37.22	41.94	30.00	34.17	37.22	25.83	32.22	30.28	25.56	30.56	28.89	25.56	28.06	26.11	23.61	23.61	23.61
Mean (%)	38.24			33.80			29.44			28.34			26.58			23.61		

ตารางที่ 4.81 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Cityblock

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	6	3	5	3	3	4	2	3	4	2	2	4	2	2	3	2	2	2
ฃ	12	18	16	13	16	15	12	13	14	12	13	12	11	12	12	11	11	11
ช	10	7	9	5	6	8	3	11	6	8	11	6	9	8	5	5	5	5
ฌ	12	9	15	11	8	10	6	6	8	3	5	7	3	6	5	4	4	4
ฉ	7	4	13	1	6	14	2	7	10	3	8	12	5	8	9	5	5	5
ฎ	13	23	19	14	16	16	13	17	14	14	12	13	11	13	11	11	11	11
ฏ	15	9	11	10	10	12	10	10	13	10	11	12	9	10	11	9	9	9
ณ	10	11	15	9	10	6	7	8	6	5	9	3	7	9	5	5	5	5
ด	10	14	16	9	12	14	9	14	10	10	9	13	9	12	10	9	9	9
ต	8	10	10	7	7	9	6	5	6	4	6	6	5	4	5	3	3	3
บ	7	11	9	7	8	8	7	9	6	7	8	8	8	9	8	9	9	9
ม	18	10	15	14	14	12	10	12	11	11	10	9	10	9	9	10	10	10
Total	128	129	153	103	116	128	87	115	108	89	104	105	89	102	93	83	83	83
%	35.56	35.83	42.50	28.61	32.22	35.56	24.17	31.94	30.00	24.72	28.89	29.17	24.72	28.33	25.83	23.06	23.06	23.06
Mean (%)	37.96			32.13			28.70			27.59			26.29			23.06		

ตารางที่ 4.82 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	5	4	3	4	2	2	2	2	2	2	2	3	2	2	3	2	2	2
ฃ	11	12	10	11	10	11	9	10	9	9	10	10	9	10	9	10	10	10
ช	9	10	8	8	8	7	7	9	5	6	7	5	4	6	4	5	5	5
ฅ	16	7	11	6	7	6	6	5	7	4	6	9	3	8	7	5	5	5
ฉ	10	7	14	6	9	15	7	11	10	8	10	9	8	9	9	8	8	8
ฐ	12	17	16	18	15	12	15	13	10	13	13	11	13	13	8	13	13	13
ฎ	11	9	10	11	8	11	10	8	14	10	8	11	10	7	11	8	8	8
ณ	9	8	10	7	5	9	3	5	5	2	5	5	2	5	5	3	3	3
ด	10	17	15	9	14	13	12	13	11	11	10	10	8	11	7	9	9	9
ต	8	5	5	5	5	2	5	3	2	4	5	2	6	5	4	6	6	6
บ	9	8	11	8	8	11	6	11	18	8	11	8	9	11	8	9	9	9
ม	19	19	15	19	13	11	15	10	12	13	9	12	12	10	12	12	12	12
Total	129	123	128	112	104	110	97	100	95	90	96	95	86	97	87	90	90	90
%	35.83	34.17	35.56	31.11	28.89	30.56	26.94	27.78	26.39	25.00	26.67	26.39	23.89	26.94	24.17	25.00	25.00	25.00
Mean (%)	35.19			30.19			27.04			26.02			25.00			25.00		

ตารางที่ 4.83 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระหว่างแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ช	4	2	4	2	1	2	2	2	2	2	2	3	2	2	2	2	2	2
ฃ	12	14	9	11	11	10	9	11	9	10	11	9	9	11	10	10	10	10
ช	11	9	9	6	6	4	2	6	6	4	10	3	6	5	3	4	4	4
ช	14	8	9	8	5	7	6	5	7	4	5	6	3	7	7	6	6	6
ฃ	9	6	14	6	8	14	6	9	8	9	10	9	9	8	9	9	9	9
ฃ	11	16	20	15	14	14	14	11	10	12	11	10	11	10	8	10	10	10
ฃ	11	9	13	11	8	10	10	9	11	11	9	10	10	9	10	9	9	9
ฃ	12	9	12	9	7	9	8	6	6	5	6	4	6	5	5	5	5	5
ด	10	17	13	11	13	10	12	12	10	11	11	11	10	11	8	10	10	10
ด	11	2	8	3	2	3	2	0	4	2	1	3	2	2	4	2	2	2
บ	13	12	11	7	9	8	7	10	9	9	11	8	11	10	10	10	10	10
ม	20	20	16	19	17	11	16	14	12	15	12	11	14	11	11	13	13	13
Total	138	124	138	108	101	102	94	95	94	94	99	83	93	91	83	90	90	90
%	38.33	34.44	38.33	30.00	28.06	28.33	26.11	26.39	26.11	26.11	27.50	23.06	25.83	25.28	23.06	25.00	25.00	25.00
Mean (%)	37.03			28.80			26.20			25.56			24.72			25.00		

ตารางที่ 4.84 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระหว่างแบบ Correlation

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	4	1	2	3	1	3	2	2	3	2	2	4	2	2	2	2	2	2
ฃ	11	14	8	12	11	10	8	10	8	8	9	7	7	9	9	7	7	7
ช	8	9	6	8	5	8	2	5	5	6	6	2	6	5	3	5	5	5
ฅ	17	10	11	12	6	7	7	4	8	4	5	10	5	5	7	5	5	5
ฉ	10	7	16	5	9	14	8	10	10	8	8	11	6	9	10	7	7	7
ฐ	10	17	20	12	19	16	14	13	11	12	10	13	11	10	13	11	11	11
ฎ	13	9	10	12	10	16	11	9	11	11	7	11	9	7	10	9	9	9
ณ	12	8	12	9	9	9	10	5	7	7	4	5	6	4	5	4	4	4
ด	11	17	17	11	15	13	11	12	11	10	12	10	10	11	8	10	10	10
ต	11	4	8	1	0	5	2	1	4	2	4	3	3	3	3	3	3	3
บ	12	12	11	10	11	11	8	9	10	9	10	8	6	9	7	6	6	6
ม	20	23	20	20	15	14	18	12	14	15	13	15	12	11	11	11	11	11
Total	139	131	141	115	111	126	101	92	102	94	90	99	83	85	88	80	80	80
%	38.61	36.39	39.17	31.94	30.83	35.00	28.06	25.56	28.33	26.11	25.00	27.50	23.06	23.61	24.44	22.22	22.22	22.22
Mean (%)	38.06			32.59			27.32			26.20			23.70			22.22		

ตารางที่ 4.85 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=2 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ช	5	3	3	3	3	2	2	2	3	2	2	5	2	3	4	3	3	3
ฌ	12	13	11	11	12	11	11	12	9	11	12	8	10	12	8	10	10	10
ช	8	11	9	8	9	7	6	10	2	3	10	3	3	7	3	4	4	4
ช	14	9	10	6	7	5	6	3	7	3	5	10	2	7	7	4	4	4
ฌ	10	7	15	4	9	14	6	11	9	7	11	8	7	10	8	7	7	7
ฎ	9	18	14	17	14	13	14	13	8	13	11	10	11	11	8	11	11	11
ฎ	9	8	9	10	8	10	10	8	12	10	7	10	8	7	9	8	8	8
ณ	11	7	9	9	5	6	5	4	4	2	4	3	3	4	3	3	3	3
ด	8	13	15	7	12	14	9	11	11	8	6	13	6	7	10	7	7	7
ด	7	4	4	3	2	3	3	2	2	1	2	2	1	2	2	2	2	2
บ	6	9	9	7	8	9	7	8	7	8	7	8	7	8	7	8	8	8
ม	16	14	10	12	12	8	10	8	7	7	6	6	7	4	6	5	5	5
Total	115	116	118	97	101	102	89	92	81	75	83	86	67	82	75	72	72	72
%	31.94	32.22	32.78	26.94	28.06	28.33	24.72	25.56	22.50	20.83	23.06	23.89	18.61	22.78	20.83	20.00	20.00	20.00
Mean (%)	32.31			27.78			24.26			22.59			20.74			20.00		

ตารางที่ 4.86 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=3 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	5	3	5	3	2	3	2	2	3	2	2	4	2	3	2	2	2	2
ฃ	12	14	10	12	11	10	10	10	9	9	10	8	10	9	9	10	10	10
ช	8	7	8	6	7	5	2	6	4	4	8	3	5	6	3	4	4	4
ฅ	12	9	11	9	7	5	6	3	4	3	4	4	3	5	5	4	4	4
ฉ	8	6	14	6	8	13	7	9	7	6	8	7	6	8	8	7	7	7
ฐ	9	16	20	12	14	15	13	13	8	12	11	10	11	11	9	11	11	11
ฎ	9	9	10	9	8	11	9	9	9	9	8	10	9	8	8	8	8	8
ณ	12	7	9	9	9	8	5	5	4	2	5	2	4	3	4	3	3	3
ด	8	12	13	7	13	14	7	12	9	8	8	12	6	8	7	6	6	6
ต	8	5	7	2	1	3	1	3	2	2	3	3	3	4	3	2	2	2
บ	6	9	8	6	8	8	5	8	7	6	8	7	7	7	7	7	7	7
ม	20	14	12	15	13	10	10	9	7	7	6	6	7	4	6	5	5	5
Total	117	111	127	96	101	105	77	89	73	70	81	76	73	76	71	69	69	69
%	32.50	30.83	35.28	26.67	28.06	29.17	21.39	24.72	20.28	19.44	22.50	21.11	20.28	21.11	19.72	19.17	19.17	19.17
Mean (%)	32.87			27.97			22.13			21.02			20.37			19.17		

ตารางที่ 4.87 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ k=5 โดยวัดระยะห่างระยะทางแบบ Cosine

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	4	4	4	4	2	3	2	2	2	2	2	5	2	2	3	3	3	3
ฃ	12	18	12	12	10	8	11	9	10	10	9	10	10	9	10	10	10	10
ช	8	9	7	3	5	7	5	6	4	5	8	2	7	5	3	5	5	5
ฅ	14	11	14	11	9	6	9	5	9	5	5	9	2	5	7	4	4	4
ฉ	5	6	14	4	6	13	7	7	7	7	8	9	6	7	10	8	8	8
ง	10	20	20	14	16	16	11	14	11	11	11	10	11	11	8	10	10	10
จ	12	6	12	6	6	14	7	7	13	7	6	10	7	7	8	7	7	7
ฉ	12	8	12	6	7	5	4	3	3	3	3	2	3	4	2	3	3	3
ด	8	15	13	6	13	12	8	11	10	9	8	9	7	9	8	7	7	7
ต	9	5	5	1	2	4	2	0	3	1	0	3	1	1	2	1	1	1
บ	10	9	6	7	8	6	7	9	5	6	7	4	6	7	6	6	6	6
ม	19	12	16	15	12	10	13	10	9	10	8	8	9	7	7	6	6	6
Total	123	123	135	89	96	104	86	83	86	76	75	81	71	74	74	70	70	70
%	34.17	34.17	37.50	24.72	26.67	28.89	23.89	23.06	23.89	21.11	20.83	22.50	19.72	20.56	20.56	19.44	19.44	19.44
Mean (%)	35.28			26.76			23.61			21.48			20.28			19.44		

ตารางที่ 4.88 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี Normal kernel

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	12	12	13	11	11	11	12	12	12	12	11	10	11	11	12	12	12	12
ฃ	14	13	9	8	10	6	8	9	8	9	9	9	8	7	7	7	7	7
ช	4	4	5	4	4	7	4	4	5	4	4	6	4	4	3	3	3	3
ฅ	10	14	11	12	12	10	10	12	12	13	14	13	12	14	15	15	15	15
ฉ	6	6	8	6	6	7	5	5	5	6	6	5	7	5	4	4	4	4
ฐ	12	12	10	12	12	11	11	12	11	12	10	10	11	11	9	9	9	9
ฎ	9	11	9	10	11	10	10	10	9	7	7	7	9	7	6	6	6	6
ณ	6	9	11	8	8	8	9	8	9	9	8	9	8	8	9	9	9	9
ด	9	10	9	11	11	10	12	11	9	10	10	11	10	10	8	8	8	8
ต	2	2	2	2	1	2	2	1	1	1	1	1	1	1	1	1	1	1
บ	0	0	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1
ม	23	22	20	20	20	19	20	20	19	18	18	18	17	17	15	15	15	15
Total	107	115	108	105	107	102	104	105	101	101	98	99	99	96	93	90	90	90
%	29.72	31.94	30.00	29.17	29.72	28.33	28.89	29.17	28.06	28.06	27.22	27.50	27.50	26.67	25.83	25.00	25.00	25.00
Mean (%)	30.55			29.07			28.71			27.59			26.67			25.00		

ตารางที่ 4.89 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี LSCV

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	9	10	10	10	10	10	10	9	9	9	9	9	10	10	10	10	10	10
ฃ	20	20	18	19	17	18	16	15	15	16	16	14	13	12	14	12	12	12
ช	9	6	8	4	5	5	5	5	5	4	6	4	5	6	5	6	6	6
ฅ	14	14	15	14	15	15	15	14	15	17	13	14	13	12	13	12	12	12
ฉ	4	3	3	2	1	2	0	2	1	2	2	1	1	1	2	2	2	2
ง	5	5	5	6	5	5	5	5	5	5	5	5	5	5	5	5	5	5
จ	16	16	15	14	14	14	15	14	15	13	13	12	11	11	10	10	10	10
ฉ	18	17	18	18	16	17	15	14	14	15	13	15	13	12	13	12	12	12
ด	20	19	19	19	19	19	17	17	18	18	17	17	15	16	15	14	14	14
ต	2	2	2	3	2	2	2	3	2	1	1	2	2	1	1	1	1	1
บ	3	3	2	2	2	2	1	1	1	1	2	1	2	2	2	2	2	2
ม	18	18	19	16	15	16	18	17	16	14	14	14	13	13	12	12	12	12
Total	138	133	134	127	121	125	119	116	116	115	111	108	103	101	102	98	98	98
%	38.33	36.94	37.22	35.28	33.61	34.72	33.06	32.22	32.22	31.94	30.83	30.00	28.61	28.06	28.33	27.22	27.22	27.22
Mean (%)	37.49			34.54			32.50			30.92			28.33			27.22		

ตารางที่ 4.90 แสดงค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ 1,000 ค่าด้วยวิธี LKCV

ตัวอักษร	จำนวนตัวอย่างเรียนรู้																	
	10			20			30			40			50			60		
ข	10	10	10	10	10	10	10	9	10	9	9	9	9	9	9	8	8	8
ฃ	20	19	20	17	15	17	16	16	16	16	15	16	16	14	16	14	14	14
ช	10	9	9	6	6	6	4	4	4	4	4	4	4	4	4	4	4	4
ฅ	13	15	14	15	15	14	15	13	14	17	15	16	16	15	15	15	15	15
ฉ	6	5	4	0	2	1	1	3	3	3	2	2	4	3	2	5	5	5
ง	6	9	8	16	15	13	15	13	14	15	14	12	13	10	11	9	9	9
จ	8	6	7	4	4	4	5	4	5	6	5	5	6	5	6	6	6	6
ฉ	15	14	16	11	10	10	13	12	10	10	11	10	11	12	11	12	12	12
ด	20	19	20	20	19	19	18	19	18	17	18	18	17	17	18	17	17	17
ต	5	5	5	6	5	5	4	4	3	3	3	3	2	1	2	1	1	1
บ	2	2	2	1	1	1	2	2	1	2	2	2	3	3	3	3	3	3
ม	20	20	20	20	19	20	19	19	19	14	15	17	14	16	14	13	13	13
Total	135	133	135	126	121	120	122	118	117	116	113	114	115	109	111	107	107	107
%	37.50	36.94	37.50	35.00	33.61	33.33	33.89	32.78	32.50	32.22	31.89	31.67	31.94	30.28	30.83	29.72	29.72	29.72
Mean (%)	37.31			33.98			33.06			31.93			31.02			29.72		

จากตารางที่ 4.76-4.78 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Euclidean พบว่าเมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 5$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุดเท่ากับ 19.44%

จากตารางที่ 4.79-4.81 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Cityblock พบว่าเมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 5$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุดเท่ากับ 23.06%

จากตารางที่ 4.82-4.84 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Correlation พบว่าเมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะเพิ่มขึ้น สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 5$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุดเท่ากับ 22.22%

จากตารางที่ 4.85-4.87 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 1,000 ค่าด้วยวิธี K-nearest neighbour ที่ $k = 2, 3$ และ 5 โดยวัดระยะห่างระยะทางแบบ Cosine พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง และเมื่อระดับค่า k เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ที่ $k = 3$ ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 19.17%

จากตารางที่ 4.88-4.90 ซึ่งแสดงผลการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 1,000 ค่าด้วยวิธี Normal kernel พบว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มส่วนมากที่ได้จะน้อยลง สรุปได้ว่า จำนวนตัวอย่างการเรียนรู้ขนาด 60 ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 25.00%

จากการศึกษาการเปรียบเทียบค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย สำหรับค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ 1,000 ค่า สรุปได้ว่า เมื่อระดับจำนวนตัวอย่างเรียนรู้เพิ่มขึ้น ค่าเฉลี่ยอัตราความผิดพลาดของการจำแนกกลุ่มที่ได้จะน้อยลง ซึ่งวิธีที่ให้ประสิทธิภาพในการจำแนกกลุ่มดีที่สุด คือ วิธี K-nearest neighbour ที่ $k = 3$ โดยวัดระยะทางแบบ Cosine ที่จำนวนตัวอย่างเรียนรู้ขนาด 60 ให้ค่าเฉลี่ยอัตราความผิดพลาดน้อยที่สุด เท่ากับ 19.17%

บทที่ 5

สรุปผลการศึกษา และข้อเสนอแนะ

จากการศึกษาเปรียบเทียบประสิทธิภาพในการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย คือ ข, ชม, ช, ซ, ฉ, ฎ, ฏ, ณ, ด, ต, บ และ ม โดยทำการศึกษาลายมือเขียนตัวอักษรไทยตัวละ 90 ตัวอักษรจำนวน 2 ชุด คือ ชุดที่ 1 จากลายมือเขียนตัวอักษรไทย 1 คน และชุดที่ 2 จากลายมือเขียนตัวอักษรไทย 10 คน ซึ่งในการพิจารณาลายมือเขียนตัวอักษรไทย จะทำการพิจารณาลักษณะตัวอักษรโดยใช้การแปลงฟูเรียร์ (Fourier transform) จำนวนค่าสัมประสิทธิ์ของการแปลงฟูเรียร์ที่ใช้ในการศึกษา ได้แก่ 200, 500 และ 1,000 โดยแต่ละตัวอักษรจะแบ่งข้อมูลเป็น 2 กลุ่ม คือ ตัวอย่างเรียนรู้ขนาด 60 ตัวอักษร และ ตัวอย่างทดสอบขนาด 30 ตัวอักษร โดยตัวอย่างเรียนรู้แบ่งย่อยออกเป็น 6 กลุ่ม คือ 10, 20, 30, 40, 50 และ 60 โดยการศึกษาครั้งนี้จะใช้วิธีการเปรียบเทียบในการจำแนกกลุ่ม 2 วิธี คือ วิธีเพื่อนบ้านใกล้เคียงที่สุด k ค่า (K-nearest neighbour method) และ วิธีเคอร์เนล (Kernel method) ซึ่งวิธี K-nearest neighbour จะทำการพิจารณาค่า k 3 ค่า คือ 2, 3, และ 5 และพิจารณาระยะห่างระยะทาง 4 แบบ คือ Euclidean, Cityblock, Correlation และ Cosine ส่วนวิธี Kernel จะทำการพิจารณา 3 แบบ คือ Normal kernel, Least squares cross-validation (LSCV) และ Likelihood cross-validation (LKCV) จากการศึกษาครั้งนี้ เกณฑ์ที่ใช้ในการพิจารณาเปรียบเทียบ คือ ค่าอัตราความผิดพลาดของการจำแนก (Error rate of misclassification) สามารถสรุปผลการวิจัยได้ดังนี้

5.1 สรุปผลการศึกษา

โดยพิจารณาการเปรียบเทียบที่ให้ค่าอัตราความผิดพลาดของการจำแนกกลุ่มน้อยที่สุดตามค่าสัมประสิทธิ์ฟูเรียร์และขนาดตัวอย่างเรียนรู้

ตารางที่ 5.1 แสดงการเปรียบเทียบค่าอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย ด้วยวิธี K-nearest neighbour และวิธี Kernel จาก

1 คน

วิธี	k	ค่าสัมประสิทธิ์	จำนวนตัวอย่างเรียนรู้					
			10	20	30	40	50	60
K-nearest neighbour ระยะห่าง ระยะทางแบบ Euclidean	2	200	17.50	12.59	13.80	11.67	11.67	11.11
		500	12.87	7.87	7.96	6.20	4.81	3.89
		1,000	14.07	9.17	9.91	8.61	6.30	6.11
	3	200	20.55	16.39	13.79	10.27	10.46	9.72
		500	13.24	10.93	8.33	5.83	5.37	5.00
		1,000	16.11	11.95	10.65	9.17	8.33	7.78
	5	200	21.02	15.56	15.56	12.04	12.04	11.94
		500	14.81	11.94	9.07	6.76	6.02	5.83
		1,000	15.83	13.43	11.94	9.35	8.89	8.61
K-nearest neighbour ระยะห่าง ระยะทางแบบ Cityblock	2	200	20.83	15.93	16.20	14.44	13.52	13.06
		500	15.46	10.65	11.58	8.98	8.89	7.50
		1,000	16.57	12.68	12.59	11.76	11.76	11.67
	3	200	21.11	17.41	16.20	12.50	12.40	11.39
		500	16.85	12.69	11.11	8.05	8.43	7.22
		1,000	16.39	12.04	12.13	10.00	10.74	10.29
	5	200	20.93	16.20	15.74	12.87	12.69	12.50
		500	15.09	13.61	11.39	8.24	9.63	8.61
		1,000	16.48	14.81	11.85	9.07	9.63	8.33

วิธี	k	ค่าสัมประสิทธิ์	จำนวนตัวอย่างเรียนรู้					
			10	20	30	40	50	60
K-nearest neighbour ระยะห่าง ระยะทางแบบ Correlation	2	200	19.35	14.63	15.09	12.31	12.50	11.67
		500	13.61	8.33	8.33	6.76	5.93	5.00
		1,000	14.26	10.09	10.28	10.18	8.70	7.50
	3	200	20.65	14.53	15.28	13.43	12.03	11.67
		500	14.82	10.83	8.89	6.67	6.67	5.83
		1,000	15.55	11.85	11.85	8.61	8.43	8.06
	5	200	20.74	16.20	16.02	13.34	12.41	11.11
		500	16.57	11.94	9.54	7.04	6.67	6.67
		1,000	15.74	13.52	12.04	9.26	9.07	8.33
K-nearest neighbour ระยะห่าง ระยะทางแบบ Cosine	2	200	17.32	11.85	12.96	9.54	10.09	8.61
		500	12.22	7.78	8.05	6.39	4.63	3.89
		1,000	12.41	7.96	8.80	8.15	6.20	5.83
	3	200	19.45	13.98	13.80	11.02	10.74	10.00
		500	13.52	10.56	8.71	6.76	6.21	5.00
		1,000	14.82	10.46	9.53	8.24	7.04	6.39
	5	200	20.74	15.09	14.81	10.09	11.39	10.83
		500	15.09	13.06	9.54	6.57	6.02	5.56
		1,000	14.44	12.78	11.39	8.70	8.33	7.22

วิธี	ค่าสัมประสิทธิ์	จำนวนตัวอย่างเรียนรู้					
		10	20	30	40	50	60
Normal kernel	200	23.05	21.11	23.24	21.11	19.44	21.67
	500	20.09	16.76	19.89	17.22	12.22	12.50
	1,000	18.61	17.59	19.07	17.68	15.84	16.39
LSCV	200	24.91	22.87	22.69	20.93	19.35	20.28
	500	24.44	19.81	18.79	17.50	12.78	14.44
	1,000	25.00	21.20	20.83	19.44	17.50	16.94
LKCV	200	23.61	20.74	19.63	18.61	15.74	15.83
	500	21.85	17.78	16.76	14.72	11.11	11.39
	1,000	22.22	19.35	19.17	17.31	15.56	16.11

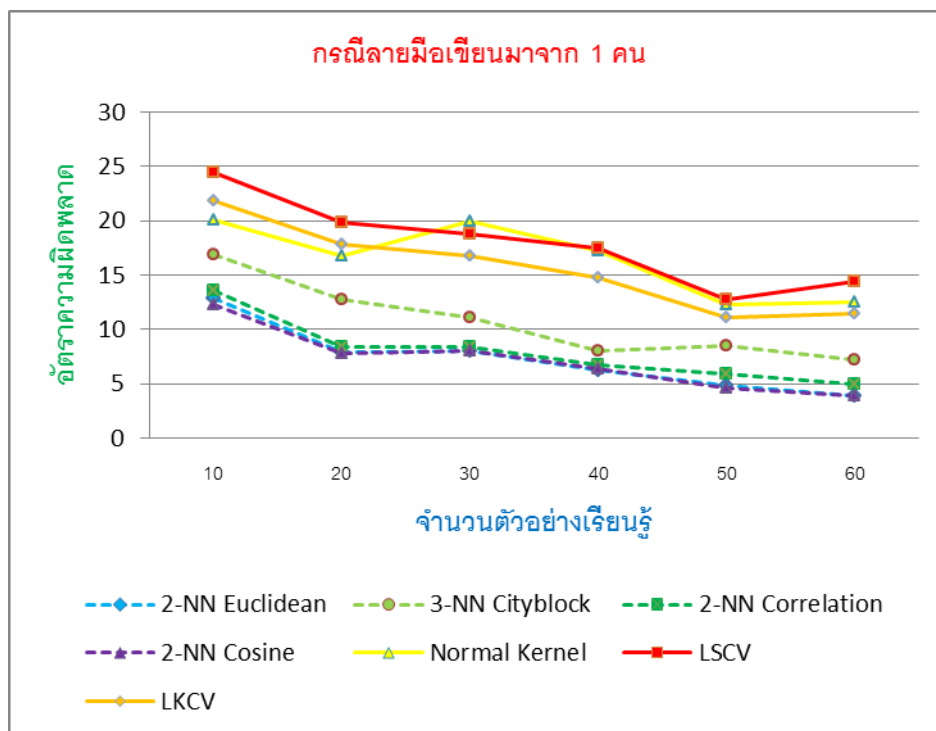
ตารางที่ 5.2 แสดงการเปรียบเทียบค่าอัตราความผิดพลาดของการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย ด้วยวิธี K-nearest neighbour และวิธี Kernel จาก

10 คน

วิธี	k	ค่าสัมประสิทธิ์	จำนวนตัวอย่างเรียนรู้					
			10	20	30	40	50	60
K-nearest neighbour ระยะห่าง ระยะทางแบบ Euclidean	2	200	34.91	29.91	27.41	26.67	26.48	25.83
		500	31.39	24.44	20.93	20.09	20.46	19.72
		1,000	32.87	28.52	25.28	23.98	22.41	20.56
	3	200	36.67	28.80	26.20	25.56	24.72	23.89
		500	31.33	23.70	20.92	19.91	19.91	20.00
		1,000	33.71	27.04	24.08	22.59	21.67	21.11
	5	200	36.39	30.00	25.46	25.09	23.61	24.17
		500	34.82	24.26	20.19	17.96	16.30	15.28
		1,000	35.18	27.87	24.26	23.24	20.28	19.44
K-nearest neighbour ระยะห่าง ระยะทางแบบ Cityblock	2	200	40.92	35.46	33.52	32.03	31.02	30.56
		500	36.95	30.37	27.13	25.46	24.08	23.33
		1,000	40.00	34.72	30.93	29.17	27.68	25.00
	3	200	39.81	33.70	31.58	30.00	30.19	29.44
		500	37.04	28.71	26.02	24.63	23.71	23.06
		1,000	38.24	33.80	29.44	28.34	26.58	23.61
	5	200	40.09	32.59	30.46	29.35	27.96	26.94
		500	36.14	28.06	28.24	25.46	22.04	20.83
		1,000	37.96	32.13	28.70	27.59	26.29	23.06

วิธี	k	ค่าสัมประสิทธิ์	จำนวนตัวอย่างเรียนรู้					
			10	20	30	40	50	60
K-nearest neighbour ระยะห่าง ระยะทางแบบ Correlation	2	200	37.59	34.63	31.39	29.26	27.59	26.94
		500	35.84	28.24	25.28	23.70	21.76	20.56
		1,000	35.19	30.19	27.04	26.02	25.00	25.00
	3	200	38.70	24.63	31.57	29.17	28.33	28.33
		500	37.41	29.63	25.65	23.43	23.06	21.11
		1,000	37.03	28.80	26.20	25.56	24.72	25.00
	5	200	38.80	35.46	32.13	30.46	28.80	28.33
		500	37.96	32.04	28.80	26.57	24.91	23.33
		1,000	38.06	32.59	27.32	26.20	23.70	22.22
K-nearest neighbour ระยะห่าง ระยะทางแบบ Cosine	2	200	34.91	30.47	27.41	26.91	26.61	25.56
		500	31.30	25.09	21.57	21.02	20.00	19.72
		1,000	32.31	27.78	24.26	22.59	20.74	20.00
	3	200	34.72	28.70	26.57	25.19	25.09	23.61
		500	31.94	24.91	22.31	20.65	19.81	19.44
		1,000	32.87	27.97	22.13	21.02	20.37	19.17
	5	200	35.37	30.83	27.40	26.02	24.54	22.78
		500	34.53	27.22	23.61	21.11	18.52	17.78
		1,000	35.28	26.76	23.61	21.48	20.28	19.44

วิธี	ค่าสัมประสิทธิ์	จำนวนตัวอย่างเรียนรู้					
		10	20	30	40	50	60
Normal kernel	200	39.91	37.22	33.15	32.33	31.39	29.72
	500	27.96	25.93	22.87	22.59	19.54	21.67
	1,000	30.55	29.07	28.71	27.59	26.67	25.00
LSCV	200	37.22	36.85	35.55	34.72	33.70	34.17
	500	35.55	32.04	30.81	30.19	28.89	26.94
	1,000	37.49	34.54	32.50	30.92	28.33	27.22
LKCV	200	36.76	35.19	33.89	33.06	31.85	30.00
	500	36.58	34.35	31.76	31.30	29.72	28.06
	1,000	37.31	33.96	33.06	31.93	31.02	29.72

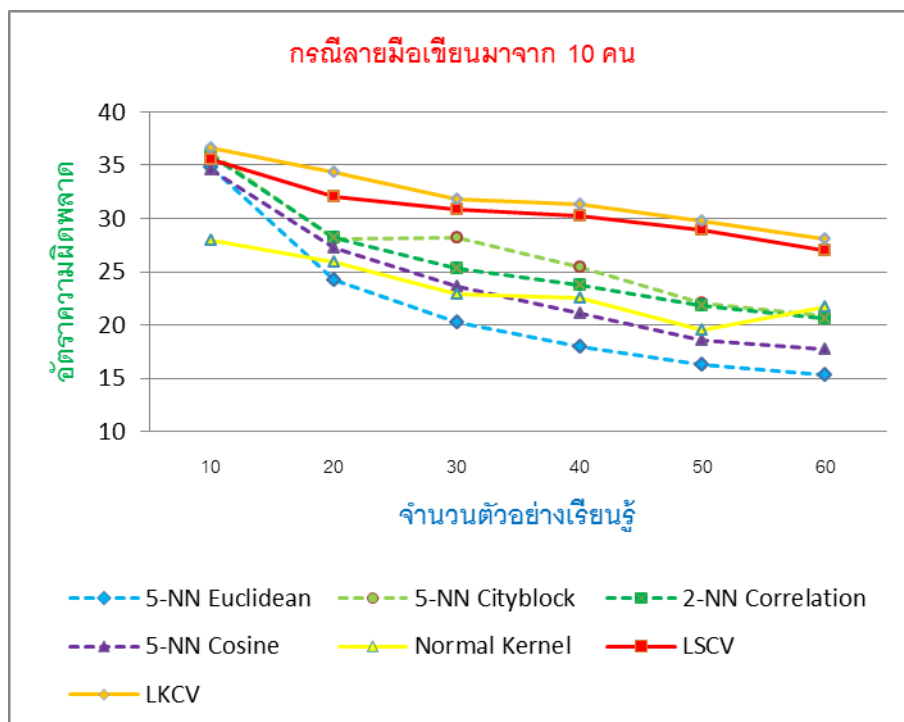


ภาพที่ 5.1 แสดงการเปรียบเทียบค่าอัตราความผิดพลาดน้อยที่สุดในแต่ละวิธี สำหรับตัวอย่างเรียนรู้จำนวนต่างๆ และจำนวนสัมประสิทธิ์ 500

จากตารางที่ 5.1 และภาพที่ 5.1 สรุปผลได้ดังนี้

พิจารณาค่าอัตราความผิดพลาดของการจำแนกกลุ่มด้วยวิธี K-nearest neighbour จากขนาดตัวอย่างเรียนรู้ ค่า k และค่าสัมประสิทธิ์ที่แตกต่าง จะได้ว่า เมื่อจำนวนขนาดของตัวอย่างเรียนรู้เพิ่มขึ้น ส่งผลให้ค่าอัตราความผิดพลาดของการจำแนกกลุ่มมีแนวโน้มน้อยลง และวิธี K-nearest neighbour โดยวัดระยะห่างระหว่างแบบ Euclidean กับ Cosine ที่ค่า k = 2 จำนวนขนาดตัวอย่างเรียนรู้ 60 และค่าสัมประสิทธิ์ฟูเรียร์ 500 ให้ประสิทธิภาพในการจำแนกกลุ่มดีกว่ากรณีอื่นๆ โดยให้ผลลัพธ์ของค่าอัตราความผิดพลาดของการจำแนกกลุ่ม คือ 3.89%

พิจารณาค่าอัตราความผิดพลาดของการจำแนกกลุ่มด้วยวิธี Kernel จากขนาดตัวอย่างเรียนรู้ และค่าสัมประสิทธิ์ที่แตกต่าง จะได้ว่า เมื่อจำนวนขนาดของตัวอย่างเรียนรู้เพิ่มขึ้น ส่งผลให้ค่าอัตราความผิดพลาดของการจำแนกกลุ่มมีแนวโน้มน้อยลง และวิธี LKCV จำนวนขนาดตัวอย่างเรียนรู้ 50 และค่าสัมประสิทธิ์ฟูเรียร์ 500 ให้ประสิทธิภาพในการจำแนกกลุ่มดีกว่ากรณีอื่นๆ โดยให้ผลลัพธ์ของค่าอัตราความผิดพลาดของการจำแนกกลุ่ม คือ 11.11%



ภาพที่ 5.1 แสดงการเปรียบเทียบค่าอัตราความผิดพลาดน้อยที่สุดในแต่ละวิธี สำหรับตัวอย่างเรียนรู้จำนวนต่างๆ และจำนวนสัมประสิทธิ์ 500

จากตารางที่ 5.2 และภาพที่ 5.2 สรุปผลได้ดังนี้

พิจารณาค่าอัตราความผิดพลาดของการจำแนกกลุ่มด้วยวิธี K-nearest neighbour จากขนาดตัวอย่างเรียนรู้ ค่า k และค่าสัมประสิทธิ์ที่แตกต่าง จะได้ว่า เมื่อจำนวนขนาดของตัวอย่างเรียนรู้เพิ่มขึ้น ส่งผลให้ค่าอัตราความผิดพลาดของการจำแนกกลุ่มได้ค่าน้อยลง และวิธี K-nearest neighbour โดยวัดระยะห่างระยะทางแบบ Euclidean ที่ค่า k = 5 จำนวนขนาดตัวอย่างเรียนรู้ 60 และค่าสัมประสิทธิ์ฟูเรียร์ 500 ให้ประสิทธิภาพในการจำแนกกลุ่มดีกว่ากรณีอื่นๆ โดยให้ผลลัพธ์ของค่าอัตราความผิดพลาดของการจำแนกกลุ่ม คือ 15.28%

พิจารณาค่าอัตราความผิดพลาดของการจำแนกกลุ่มด้วยวิธี Kernel จากขนาดตัวอย่างเรียนรู้ และค่าสัมประสิทธิ์ที่แตกต่าง จะได้ว่า เมื่อจำนวนขนาดของตัวอย่างเรียนรู้เพิ่มขึ้น ส่งผลให้ค่าอัตราความผิดพลาดของการจำแนกกลุ่มได้ค่าน้อยลง และวิธี Normal kernel จำนวนขนาดตัวอย่างเรียนรู้ 50 และค่าสัมประสิทธิ์ฟูเรียร์ 500 ให้ประสิทธิภาพในการจำแนกกลุ่มดีกว่ากรณีอื่นๆ โดยให้ผลลัพธ์ของค่าอัตราความผิดพลาดของการจำแนกกลุ่ม คือ 19.54%

จากทั้งสองกรณี สรุปได้ว่า จำนวนขนาดตัวอย่างเรียนรู้มีผลในการจำแนกกลุ่มลายมือเขียนตัวอักษรไทย นั่นคือ เมื่อจำนวนขนาดตัวอย่างเรียนรู้เพิ่มขึ้น ทำให้การจำแนกกลุ่มนั้นมีแนวโน้มที่จะจำแนกตัวอักษรได้ถูกต้องมากขึ้น จากการทดสอบวิธีของ Ryan-Einot-Gabriel-Welsch Multiple Range Test เปรียบเทียบค่าเฉลี่ยของอัตราความผิดพลาด พบว่าไม่ว่าจะใช้วิธีการจำแนกกลุ่มแบบใด เมื่อจำนวนตัวอย่างเรียนรู้เป็น 50 และ 60 อัตราความผิดพลาดแตกต่างกันอย่างไม่มีนัยสำคัญที่ระดับ 0.05 นั่นคือ ค่าอัตราความผิดพลาดมีความคงตัว การเพิ่มขึ้นของจำนวนตัวอย่างเรียนรู้มากกว่า 60 ตัวต่อแต่ละตัวอักษร ไม่ส่งผลให้อัตราความผิดพลาดมีค่าน้อยลง และส่งผลให้การเตรียมข้อมูลและการประมวลผลเสียเวลาเพิ่มมากขึ้น และจำนวนค่าสัมประสิทธิ์ฟูเรียร์ที่ใช้ในการจำแนกกลุ่มที่ให้ประสิทธิภาพดีที่สุด เท่ากับ 500

จากวิธี K-nearest neighbour กรณีของลายมือเขียนตัวอักษรไทยมาจาก 1 คน ค่า k มีผลในการจำแนกกลุ่ม นั่นคือ เมื่อค่า k น้อย ส่งผลให้ประสิทธิภาพในการจำแนกกลุ่มดีขึ้น ทั้งนี้อาจเป็นเพราะว่า ลายมือเขียนตัวอักษรไทยที่ได้มาจาก 1 คนนั้นความแตกต่างของตัวอักษรเดียวกันมีความแตกต่างกันน้อย เมื่อสร้างเงื่อนไขค่า k ที่น้อย จึงส่งผลให้การจำแนกกลุ่มมีประสิทธิภาพดีขึ้น และกรณีของลายมือเขียนตัวอักษรไทยมาจาก 10 คน ค่า k มีผลในการจำแนกกลุ่ม นั่นคือ เมื่อค่า k มากขึ้น ส่งผลให้ประสิทธิภาพในการจำแนกกลุ่มมากขึ้น ทั้งนี้อาจเป็นเพราะว่า ลายมือเขียนตัวอักษรไทยที่ได้มาจาก 10 คนนั้นความแตกต่างกันของตัวอักษรเดียวกันมีความแตกต่างกันมาก เมื่อสร้างเงื่อนไขในการจำแนกกลุ่มมากขึ้น ทำให้การจำแนกกลุ่มนั้นมีทางเลือกของเงื่อนไขมากขึ้น ส่งผลให้ประสิทธิภาพการจำแนกกลุ่มดีขึ้น

5.2 ด้านการศึกษาวิจัย

เพื่อเป็นแนวทางให้ผู้สนใจได้ศึกษาเพิ่มเติม ซึ่งในการศึกษาครั้งต่อไปอาจทำการศึกษาในกรณีต่างๆ ดังต่อไปนี้

1. ในการศึกษาครั้งนี้ผู้วิจัยได้ศึกษาการเปรียบเทียบการรู้จำลายมือเขียนตัวอักษรไทยในการจำแนกกลุ่มด้วยวิธี K-nearest neighbour กับวิธี Kernel สำหรับงานวิจัยครั้งต่อไป อาจทำการศึกษาด้วยวิธีอื่นๆ ในรูปแบบไม่ใช้พารามิเตอร์ อาจจะทำให้ได้ผลการวิจัยมีประสิทธิภาพดีขึ้น
2. เนื่องจากการศึกษาครั้งนี้ใช้ตัวบรรยายลักษณะฟูเรียร์ อาจทำการศึกษาด้วยตัวบรรยายในลักษณะอื่นๆ เช่น ตัวบรรยายลักษณะเวฟเลต เป็นต้น อาจจะทำให้ผลการวิจัยมีประสิทธิภาพและน่าสนใจขึ้น
3. ได้ศึกษาการรู้จำข้อมูลลายมือเขียนตัวอักษรไทย สำหรับงานวิจัยครั้งต่อไป เปลี่ยนข้อมูลลายมือเขียนตัวอักษรไทย เป็นข้อมูลในลักษณะแบบอื่นๆ เช่น เสียง หรือลายเซ็น เป็นต้น เพื่อให้การศึกษานั้นมีความน่าเ็นใจยิ่งขึ้น

5.3 ข้อเสนอแนะ

ในงานวิจัยนี้มีข้อเสนอแนะบางประการที่สามารถนำมาปรับปรุงเพื่อให้มีประสิทธิภาพขึ้น ดังต่อไปนี้

1. ปรับปรุงวิธีการหาขอบของภาพของรูปภาพตัวอักษรให้มีความคมชัดมากยิ่งขึ้น ทำให้คุณลักษณะพิเศษที่ได้มีความถูกต้องมากยิ่งขึ้น
2. เพิ่มความละเอียดของจุดภาพที่ได้จากเครื่องสแกน การให้ความละเอียดของจุดภาพสูงขึ้น ความคมชัดของตัวอักษรจะมากขึ้น ทำให้คุณลักษณะพิเศษที่ได้มีความเสถียรมากยิ่งขึ้น

รายการอ้างอิง

ภาษาไทย

กนกศรี ศรีนนภากร. การรู้จำลายมือเขียนตัวอักษรภาษาไทยโดยวิธีการจำแนกกลุ่ม.

วิทยานิพนธ์ปริญญาโทมหาบัณฑิต. ภาควิชาสถิติ คณะวิทยาศาสตร์

มหาวิทยาลัยเกษตรศาสตร์. 2540.

กานต์ เสารยะวิเศษ. การรู้จำลายมือเขียนตัวเลขไทยโดยใช้การจัดกลุ่มแบบฟuzzy.

สาขาวิชาเทคโนโลยีคอมพิวเตอร์. มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

2545.

ศิริรัตน์ ทินตะนัย ดร. อโนทัย ตีรวานิช และ ชัชวาลย์ เรืองประพันธ์. การรู้จำลายมือเขียนตัวเลข

อารบิกโดยวิธีการจำแนกกลุ่ม. ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น.

2544.

สมยศ จินดาพล. การรู้จำตัวพิมพ์อักษรไทยหลายรูปแบบจากประโยค. วิทยานิพนธ์

ปริญญาวิทยาศาสตรมหาบัณฑิต. สาขาคอมพิวเตอร์ คณะสถิติประยุกต์ สถาบันบัณฑิต

พัฒนบริหารศาสตร์. 2540.

โอฬาริก สุรินตะ. การรู้จำลายมือเขียนโดยใช้ตัวบรรยายลักษณะฟูเรียร์และแบบจำลองแบบ

ซ่อนของมาร์คอฟ. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต, สาขาวิชาเทคโนโลยีสารสนเทศ

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ. 2546.

ภาษาอังกฤษ

Jon S. Horne and Edward O. Garton. Likelihood Cross-Validation Versus Least

Square Cross-Validation for Choosing the Smoothing Parameter in Kernel Home-Range Analysis. Journal of Wildlife management, 641-648, 2006.

Miao-hsiang Lin, Su-yun and Yuan-chin Chang. Kernel-Based Discriminant

Techniques for Educational Placement. Journal of Educational and behavioral statistics, Vol. 29, 219-240, 2004.

บรรณานุกรม

ภาษาอังกฤษ

Trevor Hastie, Robert Tibshirani, Jerome Friedman. Data Mining, Inference, and Prediction. New York, 2001.

ภาคผนวก

โปรแกรม Matlab ที่ใช้ในการวิจัย

ในการวิจัยครั้งนี้ ใช้โปรแกรม Matlab เวอร์ชัน 7.12.0.635 (R2011a) ในที่นี้จะขอแสดงเฉพาะกรณีข้อมูลลายมือเขียนตัวอักษรไทย 1 ตัว โดยการเก็บข้อมูลจะบันทึกเป็นรูปภาพรูปแบบ .jpg

ตัวอย่าง

ข_1_1.jpg ซึ่งมีความหมายตามลำดับดังนี้

ข	หมายถึง	ตัวอักษร ข.ไข
1	หมายถึง	ลายมือเขียนตัวอักษรของคนี่ 1
1	หมายถึง	ลายมือเขียนตัวอักษรตัวที่ 1
jpg	หมายถึง	บันทึกในรูปแบบบิตแมพ

1. การแปลงฟูเรียร์จากรูปภาพลายมือเขียนแต่ละตัวอักษรไทย มีคำสั่งดังนี้

1.1 การอ่านรูปภาพลายมือเขียนตัวอักษร

```
A = imread("ข_1_1.jpg");
B = imread("ฌ_1_1.jpg");
C = imread("ช_1_1.jpg");
D = imread("ซ_1_1.jpg");
E = imread("ฌ_1_1.jpg");
F = imread("ฎ_1_1.jpg");
G = imread("ฏ_1_1.jpg");
H = imread("ณ_1_1.jpg");
I = imread("ด_1_1.jpg");
J = imread("ต_1_1.jpg");
```

```
K = imread("บ_1_1.jpg");
```

```
L = imread("ม_1_1.jpg");
```

1.2 การหาขอบของภาพลายมือเขียนตัวอักษรไทย

```
ABW = im2bw(A);
```

```
Aboundary = ABW - imerode(ABW,ones(3));
```

1.3 การแปลงฟูเรียร์ลายมือเขียนตัวอักษรไทย

```
numcoeffs = 200; (500/ 1,000)
```

```
Afdlmd = abs(fft2(Aboundary));
```

```
Afd = Afdlmg(2:1+numcoeffs);
```

```
Afd_sum = sum(Afd);
```

```
AFD = Afd / Afd_sum;
```

2. การจำแนกกลุ่มด้วยวิธี K-nearest neighbour ในที่นี้จะแสดงกรณีตัวอย่างเรียนรู้ (training set) เท่ากับ 10 มีคำสั่งดังนี้

2.1 วัดระยะห่างระยะทางแบบ Euclidean

```
group = [repmat(1,10,1); repmat(2,10,1); repmat(3,10,1); repmat(4,10,1);repmat(5,10,1);
repmat(6,10,1); repmat(7,10,1); repmat(8,10,1); repmat(9,10,1); repmat(10,10,1);
repmat(11,10,1); repmat(12,10,1)];
```

```
IDX = knnclassify(test_set_n,training_set_10,group,2,'euclidean');
```

```
IDX = knnclassify(test_set_n,training_set_10,group,3,'euclidean');
```

```
IDX = knnclassify(test_set_n,training_set_10,group,5,'euclidean');
```

2.2 วัดระยะห่างระยะทางแบบ Cityblock

```
group = [repmat(1,10,1); repmat(2,10,1); repmat(3,10,1); repmat(4,10,1);repmat(5,10,1);
         repmat(6,10,1); repmat(7,10,1); repmat(8,10,1); repmat(9,10,1); repmat(10,10,1);
         repmat(11,10,1); repmat(12,10,1)];
```

```
IDX = knnclassify(test_set_n,training_set_10,group,2,'cityblock');
```

```
IDX = knnclassify(test_set_n,training_set_10,group,3,'cityblock');
```

```
IDX = knnclassify(test_set_n,training_set_10,group,5,'cityblock');
```

2.3 วัดระยะห่างระยะทางแบบ Correlation

```
group = [repmat(1,10,1); repmat(2,10,1); repmat(3,10,1); repmat(4,10,1);repmat(5,10,1);
         repmat(6,10,1); repmat(7,10,1); repmat(8,10,1); repmat(9,10,1); repmat(10,10,1);
         repmat(11,10,1); repmat(12,10,1)];
```

```
IDX = knnclassify(test_set_n,training_set_10,group,2,'correlation');
```

```
IDX = knnclassify(test_set_n,training_set_10,group,3,'correlation');
```

```
IDX = knnclassify(test_set_n,training_set_10,group,5,'correlation');
```

2.4 วัดระยะห่างระยะทางแบบ Cosine

```
group = [repmat(1,10,1); repmat(2,10,1); repmat(3,10,1); repmat(4,10,1);repmat(5,10,1);
         repmat(6,10,1); repmat(7,10,1); repmat(8,10,1); repmat(9,10,1); repmat(10,10,1);
         repmat(11,10,1); repmat(12,10,1)];
```

```
IDX = knnclassify(test_set_n,training_set_10,group,2,'cosine');
```

```
IDX = knnclassify(test_set_n,training_set_10,group,3,'cosine');
```

```
IDX = knnclassify(test_set_n,training_set_10,group,5,'cosine');
```

3. การจำแนกกลุ่มด้วยวิธี Kernel ในที่นี้จะแสดงตัวอย่างเรียนรู้ (training set) เท่ากับ 10 มีคำสั่งดังนี้

3.1 Normal kernel

```
m = 30;
```

```
n = 10;
```

```
h = 0.5;
```

```
p = 2;
```

```
for j = 1:m
```

```
for i = 1:n
```

```
d = transpose(test_set_n(j:)-training_set(i,:))*(test_set_n(j:)-training_set(i,:));
```

```
K(i) = ((2*pi)^(-p/2))*exp(-0.5*(h^(-2))*d);
```

```
sum(K);
```

```
end
```

```
f(j) = sum(K)/(n*(h^p));
```

```
end
```

3.2 Least square cross-validation (LSCV)

```

function [lscv] = LSCV(h);

m=10;

n=10;

L = 0;

for i = 1:m

K = 0;

for j = 1:n

if(i~= j)

d = transpose(training_set(j,:)-training_set(i,:))*(training_set(j,:)-training_set(i,:));

K = K+((exp(-0.25*(h^(-2))*d))-(4*exp(-0.5*(h^(-2))*d)));

end

end

K = 1/(4*pi*(n^2)*(h^2))*K;

L = L+K;

end

LSCV = (1/(pi*(h^2)*n))+L

```

3.3 Likelihood cross-validation (LKCV)

```

function [lkcw] = LKCV(h);

m=10;

n=10;

L = 0;

for i = 1:m

K = 0;

for j = 1:n

if(i~= j)

d = transpose(training_set(j,:)-training_set(i,:))*(training_set(j,:)-training_set(i,:));

K = K+exp(-0.5*(h^(-2))*d);

end

end

K = 1/(2*pi*n*(h^2))*K;

L =L+ log(K);

end

LKCV = L/(-n)

```

ประวัติผู้เขียนวิทยานิพนธ์

นายรัชชัย กิติวงศ์ประทีป เกิดวันเสาร์ที่ 27 สิงหาคม พ.ศ. 2531 สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2553 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2554