



การประยุกต์ใช้ลักษณะเด่นเชิงพื้นผิวแบบจีแอลซีเอ็ม

ในการนับจำนวนนิวเคลียสเซลล์มะเร็งเต้านม

**Application of GLCM Textural Feature for Breast Cancer Nuclei Counting**

อรรถกร มากคงแก้ว

Attagorn Markkongkeaw

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา

วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

มหาวิทยาลัยสงขลานครินทร์

**A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of**

**Master of Engineering in Electrical Engineering**

**Prince of Songkla University**

2557

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์

ชื่อวิทยานิพนธ์	การประยุกต์ใช้ลักษณะเด่นเชิงพื้นที่แบบจีแอลซีเอ็มในการนับจำนวน นิวเคลียสเซลล์มะเร็งเต้านม
ผู้เขียน	นายอรรถกร มากคงแก้ว
สาขาวิชา	วิศวกรรมไฟฟ้า

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

คณะกรรมการสอบ

.....ประธานกรรมการ  
(รองศาสตราจารย์ ดร. พรชัย พฤกษ์ภัทรานนท์) (รองศาสตราจารย์ บุญเจริญ วงศ์กิตติศึกษา)

.....กรรมการ  
(รองศาสตราจารย์ ดร. พรชัย พฤกษ์ภัทรานนท์)

.....กรรมการ  
(ดร. รักษกฤตว์ ดวงสร้อยทอง)

.....กรรมการ  
(ดร. สมชัย หลิมศิริโรรัตน์)

.....กรรมการ  
(รองศาสตราจารย์ ดร. วัฒนพงษ์ เกิดทองมี)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้บัณฑิตวิทยาลัยฉบับนี้  
เป็นส่วนหนึ่งของการศึกษา ตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชา  
วิศวกรรมไฟฟ้า

.....  
(รองศาสตราจารย์ ดร. ชีระพล ศรีชนะ)

คณบดีบัณฑิตวิทยาลัย

ขอรับรองว่า ผลงานวิจัยนี้เป็นผลมาจากการศึกษาวิจัยของนักศึกษาเอง และขอขอบคุณผู้ที่มีส่วน  
เกี่ยวข้องทุกท่านไว้ ณ ที่นี้

ลงชื่อ.....

(รองศาสตราจารย์ ดร. พรชัย พฤกษ์ภัทรานนท์)

อาจารย์ที่ปรึกษาวิทยานิพนธ์

ลงชื่อ.....

(นายอรรถกร มากคงแก้ว)

นักศึกษา

ข้าพเจ้าขอรับรองว่า ผลงานวิจัยนี้ไม่เคยเป็นส่วนหนึ่งในการอนุมัติปริญญาในระดับใดมาก่อน  
และไม่ได้ถูกใช้ในการยื่นขออนุมัติปริญญาในขณะนี้

ลงชื่อ.....

(นายอรรถกร มากคงแก้ว)

นักศึกษา

ชื่อวิทยานิพนธ์	การประยุกต์ใช้ลักษณะเด่นเชิงพื้นผิวแบบจีแอลซีเอ็มในการนับจำนวนนิวเคลียสเซลล์มะเร็งเต้านม
ผู้เขียน	นายอรรถกร มากคงแก้ว
สาขาวิชา	วิศวกรรมไฟฟ้า
ปีการศึกษา	2556

### บทคัดย่อ

งานวิจัยนี้นำเสนอการประยุกต์ใช้การวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็มในการวิเคราะห์ภาพถ่ายเซลล์มะเร็งเต้านมที่ผ่านการเชื่อมด้วยเทคนิคอิมมูโนฮิสโตเคมี จากกล้องจุลทรรศน์ ซึ่งในกระบวนการวิเคราะห์สามารถแบ่งได้เป็นสองขั้นตอนหลักด้วยกัน คือ การตัดแบ่งภาพถ่ายเซลล์มะเร็งเต้านมและการจำแนกชนิดของเซลล์ ในขั้นตอนการตัดแยกภาพถ่ายเซลล์มะเร็งเต้านมต้นฉบับจะถูกสไลด์ด้วยหน้าต่างสองขนาดด้วยกันแล้วทำการเลือกเฉพาะภาพเซลล์ที่ไม่มีการปะปนของเซลล์ชนิดอื่นจากนั้นนำภาพที่ได้มาแปลงเป็นภาพระดับเทา ในการแบ่งชนิดของเซลล์จะแบ่งออกได้เป็นสามขั้นตอนย่อย คือ (1)การหาคุณลักษณะเด่นโดยใช้การวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็มทำการวิเคราะห์ ซึ่งจะได้ค่าคุณลักษณะเด่นของจีแอลซีเอ็มออกมา 22 ค่า (2)การคัดเลือกค่าคุณลักษณะเด่นที่เหมาะสม โดยใช้การวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบซึ่งจะได้ชุดค่าคุณลักษณะเด่นที่เหมาะสมสำหรับการวิเคราะห์ภาพถ่ายเซลล์มะเร็งสองชุดด้วยกันแยกตามขนาดของหน้าต่างที่ใช้ในการสไลด์ภาพ (3)การจำแนกชนิดของเซลล์โดยใช้ต้นไม้ตัดสินใจร่วมกับการตรวจสอบไขว้ และทำการประเมินประสิทธิภาพของอัลกอริทึมที่ได้ ผลการวิจัยพบว่าเมื่อทำการสไลด์ภาพด้วยหน้าต่างขนาด  $256 \times 256$  พิกเซลแล้วนำมาวิเคราะห์ด้วยจีแอลซีเอ็มที่ระยะห่าง 9 พิกเซล มุม 135 องศา และใช้ชุดค่าคุณลักษณะเด่นของจีแอลซีเอ็ม 9 คุณลักษณะนำมาจำแนกด้วยต้นไม้ตัดสินใจ จะให้ค่าความแม่นยำในการจำแนกสูงสุดอยู่ที่ 98.08% ค่าความไว 99.25% ค่าความจำเพาะ 97.25% ค่าพยากรณ์ผลบวก 98.63% และค่าพยากรณ์ผลลบ 98.48% เมื่อนำอัลกอริทึมที่ได้ไปทดสอบวิเคราะห์บนภาพถ่ายเซลล์มะเร็งเต้านมต้นฉบับจะสามารถลดจำนวนผลบวกเท็จแต่ทำให้จำนวนผลลบเท็จเพิ่มขึ้น

คำสำคัญ	การวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็ม ต้นไม้ตัดสินใจ มะเร็งเต้านม ภาพถ่ายกล้องจุลทรรศน์
---------	---

<b>Thesis Title</b>	Application of GLCM textural feature for breast cancer nuclei counting
<b>Author</b>	Mister Attagorn Markkongkeaw
<b>Major Program</b>	Electrical Engineering
<b>Academic Year</b>	2013

### ABSTRACT

In this study, we present an application of gray-level co-occurrence matrix (GLCM) for analyzing on microscopic image of breast cancer cell that stain with immunohistochemistry technique. There are two stages of image processing, i.e. image sliding and classification. In image sliding, the raw images were slide by two sizes of window. The sub-image of interest were selected and transformed to gray scale image. In the classification, sub-images were analyzed by GLCM and 22 features of GLCM were generated from each matrix. Correlation with greedy algorithm is used to analyze the feature values for selecting the optimum feature set. The result shows that the classification accuracy obtained from decision tree is 98.08% when slide image by  $256 \times 256$  pixels of window size. In addition, optimum parameters for GLCM are as follows: Distance is 9 pixels and direction is 135 degree. The performance of the classification algorithm is as follows: the sensitivity 99.25%, the specificity 97.25%, the positive predictive value 98.63%, and the negative predictive value 98.48%. The algorithms were tested on the original breast cancer cells images. Result show that the false positive can be reduced at the expense of increasing false negative.

**Keywords:** Gray-Level Co-occurrence Matrix (GLCM), Decision tree, Breast cancer, Microscopic image

### กิตติกรรมประกาศ

ขอขอบพระคุณ รองศาสตราจารย์ ดร. พรชัย พุกฤษภ์ทรานนค์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้กรุณาให้การสนับสนุนและฝึกฝนการทำวิจัย ตลอดจนจัดเวลากระบวนการคิดและให้กำลังใจในการแก้ไขปัญหา รวมทั้งตรวจทานแก้ไขวิทยานิพนธ์ให้ดำเนินไปอย่างสมบูรณ์

ขอขอบพระคุณ รองศาสตราจารย์ บุญเจริญ วงศ์กิตติศึกษา กรรมการสอบวิทยานิพนธ์ ทั้งยังเป็นกรรมการสอบโครงร่างและกรรมการสอบความก้าวหน้างานวิจัย ที่ได้กรุณาอุทิศเวลาให้คำปรึกษาแนะนำที่เป็นประโยชน์ต่อการทำงานวิจัยและตรวจทานวิทยานิพนธ์ให้ดำเนินไปอย่างสมบูรณ์

ขอขอบพระคุณ ดร. รักกฤตว์ ดวงสร้อยทอง ที่ได้กรุณาอุทิศเวลาเป็นกรรมการสอบวิทยานิพนธ์ ทั้งยังเป็นกรรมการสอบโครงร่างและกรรมการสอบความก้าวหน้างานวิจัย และยังกรุณาให้คำปรึกษาแนะนำที่เป็นประโยชน์ต่อการทำงานวิจัยและตรวจทานวิทยานิพนธ์ให้ดำเนินไปอย่างสมบูรณ์

ขอขอบพระคุณ ดร. สมชัย หลิมศิริโรรัตน์ ที่ได้กรุณาอุทิศเวลาเป็นกรรมการสอบวิทยานิพนธ์ ให้คำแนะนำที่มีประโยชน์ วิจารณ์ผลงาน และตรวจทานวิทยานิพนธ์ให้มีความสมบูรณ์มากยิ่งขึ้น

ขอขอบพระคุณ รองศาสตราจารย์ ดร. วัฒนพงศ์ เกิดทองมี ที่ได้กรุณาอุทิศเวลาเป็นกรรมการสอบวิทยานิพนธ์ ให้คำแนะนำที่มีประโยชน์ วิจารณ์ผลงาน และตรวจทานวิทยานิพนธ์ให้มีความสมบูรณ์มากยิ่งขึ้น

ขอขอบพระคุณ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์ ที่กรุณาสนับสนุนทุนการศึกษา “โครงการปริญญาตรี-โท 5 ปี” แก่ข้าพเจ้าระหว่างการศึกษา

สุดท้ายนี้ขอขอบคุณอาจารย์ผู้สอนทุกท่าน ที่ประสิทธิ์ประสาทความรู้ทางด้านต่างๆ ทั้งในอดีตและปัจจุบัน และขอกราบขอบพระคุณบิดา มารดา รวมถึงพี่ๆ เพื่อนๆ ที่ได้ให้ความรัก ความห่วงใย และให้การสนับสนุนทางด้านการศึกษาอย่างดียิ่งมาโดยตลอด รวมทั้งเป็นกำลังใจที่ยิ่งใหญ่ในยามที่ผู้วิจัยท้อและทุกข์ใจช่วยให้มีพลังเข้มแข็งพร้อมเผชิญกับปัญหาอุปสรรคต่างๆ จนทำให้ผู้วิจัยประสบความสำเร็จในชีวิตเรื่อยมา หากวิทยานิพนธ์นี้มีความผิดอันใด ผู้วิจัยขอให้เป็นกตเวทิตาแต่บิดา มารดาและบูรพาจารย์ทุกท่าน แต่ถ้าวิทยานิพนธ์ฉบับนี้มีข้อบกพร่องประการใด ผู้วิจัยขอน้อมรับไว้แต่เพียงผู้เดียว

อรธกร มากคงแก้ว

## สารบัญ

	หน้า
สารบัญ .....	(8)
รายการตาราง .....	(10)
รายการภาพประกอบ .....	(12)
บทที่	
1. บทนำ .....	1
1.1 ความสำคัญและที่มาของการวิจัย .....	1
1.2 งานวิจัยที่เกี่ยวข้อง .....	3
1.3 วัตถุประสงค์ของการวิจัย .....	5
1.4 ขอบเขตการวิจัย .....	6
1.5 ประโยชน์ที่คาดว่าจะได้รับ .....	6
1.6 ขั้นตอนการวิจัย .....	6
2. ทฤษฎีที่เกี่ยวข้อง .....	8
2.1 การเตรียมภาพก่อนการประมวลผล .....	8
2.2 การวิเคราะห์เชิงพื้นที่แบบจีแอลซีเอ็ม .....	11
2.3 การจำแนกโดยใช้ต้นไม้ตัดสินใจ .....	17
2.4 การคัดเลือกคุณลักษณะเด่น .....	17
3. อุปกรณ์และวิธีการ .....	20
3.1 ระบบเก็บภาพ .....	20
3.2 ระบบประมวลผล .....	20
3.3 คุณลักษณะของภาพถ่ายเซลล์มะเร็งเต้านม .....	20
3.4 ขั้นตอนวิธีการวิเคราะห์เซลล์มะเร็งเต้านม .....	23
3.5 การประเมินประสิทธิภาพ .....	24
4. ผลการวิจัย .....	27
4.1 ผลการตัดแบ่งภาพเซลล์มะเร็งเต้านม .....	27
4.2 ผลการจำแนกชนิดของเซลล์ .....	31
4.3 ผลการประเมินประสิทธิภาพการนับจำนวนเซลล์มะเร็งเต้านม .....	46



## สารบัญ (ต่อ)

	หน้า
5. บทสรุปและข้อเสนอแนะ .....	48
5.1 บทสรุป .....	48
5.2 ข้อเสนอแนะ .....	49
บรรณานุกรม .....	51
ภาคผนวก .....	55
ภาคผนวก ก ผลการจำแนกชนิดของเซลล์ .....	56
ภาคผนวก ข ผลการประเมินประสิทธิภาพบนชุดข้อมูล .....	63
ภาคผนวก ค ผลการประเมินประสิทธิภาพบนภาพต้นฉบับ .....	70
ภาคผนวก ง Proceedings' title "Preliminary Results of Breast Cancer Cell Classifying Based on Gray-Level Co-occurrence Matrix" Presented in The 6th Biomedical Engineering International Conference (BMEiCON 2013) 23-25 October 2013 at Krabi, Thailand .....	72
ประวัติผู้เขียน .....	77

**รายการตาราง**

ตารางที่	หน้า
2-1	นิพจน์ที่ใช้ในการคำนวณคุณลักษณะเด่นเชิงพื้นผิวแบบจีแอลซีเอ็ม ..... 14
2-2	คุณลักษณะเด่นเชิงพื้นผิวแบบจีแอลซีเอ็ม ..... 15
ก-1	ค่าความแม่นยำในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นทั้ง 22 คุณลักษณะ ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 128 × 128 พิกเซล ..... 57
ก-2	ค่าความแม่นยำในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นที่ถูกเลือกจากการ วิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบในภาพที่ผ่านการสไลด์ด้วย หน้าต่างขนาด 128 × 128 พิกเซล ..... 58
ก-3	ค่าความแม่นยำในการจำแนกชนิดของเซลล์โดยใช้ชุดค่าคุณลักษณะเด่นที่เหมาะสม ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 128 × 128 พิกเซล ..... 59
ก-4	ค่าความแม่นยำในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นทั้ง 22 คุณลักษณะ ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล ..... 60
ก-5	ค่าความแม่นยำในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นที่ถูกเลือกจากการ วิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบในภาพที่ผ่านการสไลด์ด้วย หน้าต่างขนาด 256 × 256 พิกเซล ..... 61
ก-6	ค่าความแม่นยำในการจำแนกชนิดของเซลล์โดยใช้ชุดค่าคุณลักษณะเด่นที่เหมาะสม ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล ..... 62
ข-1	ผลการประเมินประสิทธิภาพในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นทั้ง 22 คุณลักษณะในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 128 × 128 พิกเซล ..... 64
ข-2	ผลการประเมินประสิทธิภาพในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นที่ถูกเลือก จากการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบในภาพที่ผ่านการสไลด์ ด้วยหน้าต่างขนาด 128 × 128 พิกเซล ..... 65
ข-3	ผลการประเมินประสิทธิภาพในการจำแนกชนิดของเซลล์โดยใช้โดยใช้ชุดค่าคุณลักษณะเด่น ที่เหมาะสมในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 128 × 128 พิกเซล ..... 66
ข-4	ผลการประเมินประสิทธิภาพในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นทั้ง 22 คุณลักษณะในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล ..... 67

รายการตาราง (ต่อ)

ตารางที่	หน้า
ข-5 ผลการประเมินประสิทธิภาพในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นที่ถูกเลือกจากการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $256 \times 256$ พิกเซล .....	68
ข-6 ผลการประเมินประสิทธิภาพในการจำแนกชนิดของเซลล์โดยใช้โดยใช้ชุดค่าคุณลักษณะเด่นที่เหมาะสมในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $256 \times 256$ พิกเซล .....	69
ค-1 ผลการประเมินประสิทธิภาพของระบบนับจำนวนนิวเคลียสเซลล์มะเร็งด้านมบนภาพถ่ายต้นฉบับ .....	71

## รายการภาพประกอบ

ภาพประกอบ	หน้า
1-1 จำนวนผู้ป่วยที่เป็น โรคมะเร็งเต้านม ณ ศูนย์มะเร็งโรงพยาบาลสงขลานครินทร์ และสถาบันมะเร็งแห่งชาติ .....	2
2-1 ภาพสี RGB และข้อมูลของภาพ .....	9
2-2 ภาพ Gray - scale image และข้อมูลของภาพ .....	10
2-3 พิกเซลข้างเคียงในมุม $0^\circ$ $45^\circ$ $90^\circ$ และ $135^\circ$ .....	13
2-4 ตัวอย่างการคำนวณเมทริกซ์จีแอลซีเอ็มที่ระยะห่าง 1 พิกเซลและมุม 0 องศา .....	13
3-1 กล้องจุลทรรศน์ยี่ห้อ Nikon รุ่น Eclipse 80i และกล้องถ่ายภาพดิจิทัลยี่ห้อ Nikon รุ่น DXM1200F .....	21
3-2 ภาพถ่ายเซลล์มะเร็งเต้านมที่กำลังขยาย 400 เท่า .....	22
3-3 ภาพถ่ายกลุ่มเซลล์มะเร็งที่ผ่านการข้อมแยกตามชนิดของเซลล์ .....	22
3-4 ขั้นตอนการวิเคราะห์ภาพถ่ายเซลล์มะเร็งเต้านม .....	23
4-1 ลักษณะทิศทางการสไลด์วินโดว์บนภาพถ่ายต้นฉบับ .....	27
4-2 ภาพเซลล์ทั้งสี่ชนิดที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $128 \times 128$ พิกเซล .....	28
4-3 ภาพเซลล์ทั้งสี่ชนิดที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $256 \times 256$ พิกเซล .....	28
4-4 ภาพเซลล์ทั้งสี่ชนิดขนาด $128 \times 128$ พิกเซลที่ผ่านการแปลงปริภูมิสีจากภาพสี RGB เป็นภาพระดับเทาและฮิสโทแกรมของภาพ .....	29
4-5 ภาพเซลล์ทั้งสี่ชนิดขนาด $256 \times 256$ พิกเซลที่ผ่านการแปลงปริภูมิสีจากภาพสี RGB เป็นภาพระดับเทาและฮิสโทแกรมของภาพ .....	30
4-6 จีแอลซีเอ็มที่ระยะห่าง 1 พิกเซลมุม 0 องศาของภาพเซลล์ทั้งสี่ชนิดที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $128 \times 128$ พิกเซล .....	31
4-7 ค่าความแม่นยำในการจำแนกชนิดของเซลล์ โดยใช้ค่าคุณลักษณะเด่นทั้ง 22 คุณลักษณะในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $128 \times 128$ พิกเซล .....	32
4-8 จำนวนของคุณลักษณะเด่นที่ถูกเลือกใช้ในการจำแนกชนิดของเซลล์ โดยการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $128 \times 128$ พิกเซล .....	33

## รายการภาพประกอบ (ต่อ)

ภาพประกอบ	หน้า
4-9 ค่าความแม่นยำในการจำแนกชนิดของเซลล์ โดยใช้ค่าคุณลักษณะเด่นที่ถูกเลือกจากการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $128 \times 128$ พิกเซล .....	34
4-10 ผลต่างค่าความแม่นยำในการจำแนกชนิดของเซลล์ ระหว่างการใช้คุณลักษณะเด่นที่ถูกเลือกจากการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบกับการใช้คุณลักษณะเด่นทั้งหมดในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $128 \times 128$ พิกเซล .....	34
4-11 ความถี่ของค่าคุณลักษณะเด่นที่ถูกเลือกโดยการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบของภาพที่ถูกสไลด์ด้วยหน้าต่างขนาด $128 \times 128$ พิกเซล .....	35
4-12 ค่าความแม่นยำในการจำแนกชนิดของเซลล์ โดยใช้ชุดค่าคุณลักษณะเด่นที่เหมาะสมจำนวน 8 คุณลักษณะ ได้แก่ f3 f5 f9 f14 f15 f18 f19 และ f20 ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $128 \times 128$ พิกเซล .....	36
4-13 คะแนนความดีของชุดค่าคุณลักษณะเด่นที่เหมาะสมจำนวน 8 คุณลักษณะของการวิเคราะห์จีแอลซีเอ็มที่ระยะห่าง 7 พิกเซล มุม 0 องศา เพื่อใช้ในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $128 \times 128$ พิกเซล .....	37
4-14 ค่าความแม่นยำของชุดค่าคุณลักษณะเด่นที่เหมาะสมจำนวน 8 คุณลักษณะของการวิเคราะห์จีแอลซีเอ็มที่ระยะห่าง 7 พิกเซล มุม 0 องศา เพื่อใช้ในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $128 \times 128$ พิกเซล .....	37
4-15 คะแนนความดีของค่าคุณลักษณะเด่นทั้งหมด 22 คุณลักษณะของการวิเคราะห์จีแอลซีเอ็มที่ระยะห่าง 7 พิกเซล มุม 0 องศา เพื่อใช้ในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $128 \times 128$ พิกเซล .....	38
4-16 ค่าความแม่นยำในการจำแนกชนิดของเซลล์ โดยใช้ค่าคุณลักษณะเด่นทั้ง 22 คุณลักษณะในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $256 \times 256$ พิกเซล .....	39
4-17 จำนวนของคุณลักษณะเด่นที่ถูกเลือกใช้ในการจำแนกชนิดของเซลล์ โดยการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด $256 \times 256$ พิกเซล .....	40

## รายการภาพประกอบ (ต่อ)

ภาพประกอบ	หน้า
4-18 ค่าความแม่นยำในการจำแนกชนิดของเซลล์ โดยใช้ค่าคุณลักษณะเด่นที่ถูกเลือกจากการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล.....	40
4-19 ผลต่างค่าความแม่นยำในการจำแนกชนิดของเซลล์ ระหว่างการใช้คุณลักษณะเด่นที่ถูกเลือกจากการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบกับการใช้คุณลักษณะเด่นทั้งหมดในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล.....	41
4-20 ค่าความแม่นยำในการจำแนกชนิดของเซลล์ โดยทดสอบลดจำนวนค่าคุณลักษณะเด่นที่เหมาะสมในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 x 256 พิกเซล .....	42
4-21 ค่าความแม่นยำในการจำแนกชนิดของเซลล์ โดยใช้ชุดค่าคุณลักษณะเด่นที่เหมาะสมจำนวน 9 คุณลักษณะของการวิเคราะห์จีแอลซีเอ็มที่ระยะห่าง 9 พิกเซล มุม 135 องศาในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล .....	43
4-22 คะแนนความดีของชุดค่าคุณลักษณะเด่นที่เหมาะสมจำนวน 9 คุณลักษณะของการวิเคราะห์จีแอลซีเอ็มที่ระยะห่าง 9 พิกเซล มุม 135 องศา เพื่อใช้ในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล .....	43
4-23 ค่าความแม่นยำของชุดค่าคุณลักษณะเด่นที่เหมาะสมจำนวน 9 คุณลักษณะของการวิเคราะห์จีแอลซีเอ็มที่ระยะห่าง 9 พิกเซล มุม 135 องศา เพื่อใช้ในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล .....	44
4-24 คะแนนความดีของค่าคุณลักษณะเด่นทั้งหมด 22 คุณลักษณะของการวิเคราะห์จีแอลซีเอ็มที่ระยะห่าง 9 พิกเซล มุม 135 องศา เพื่อใช้ในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล .....	45
4-25 ผลการวิเคราะห์ค่าสเกลที่เหมาะสมสำหรับการวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็ม โดยใช้กรอบข้อกำหนดและชุดค่าลักษณะเด่นที่ทำให้เหมาะสมสำหรับการจำแนกชนิดของเซลล์มะเร็งเต้านม .....	46

# บทที่ 1

## บทนำ

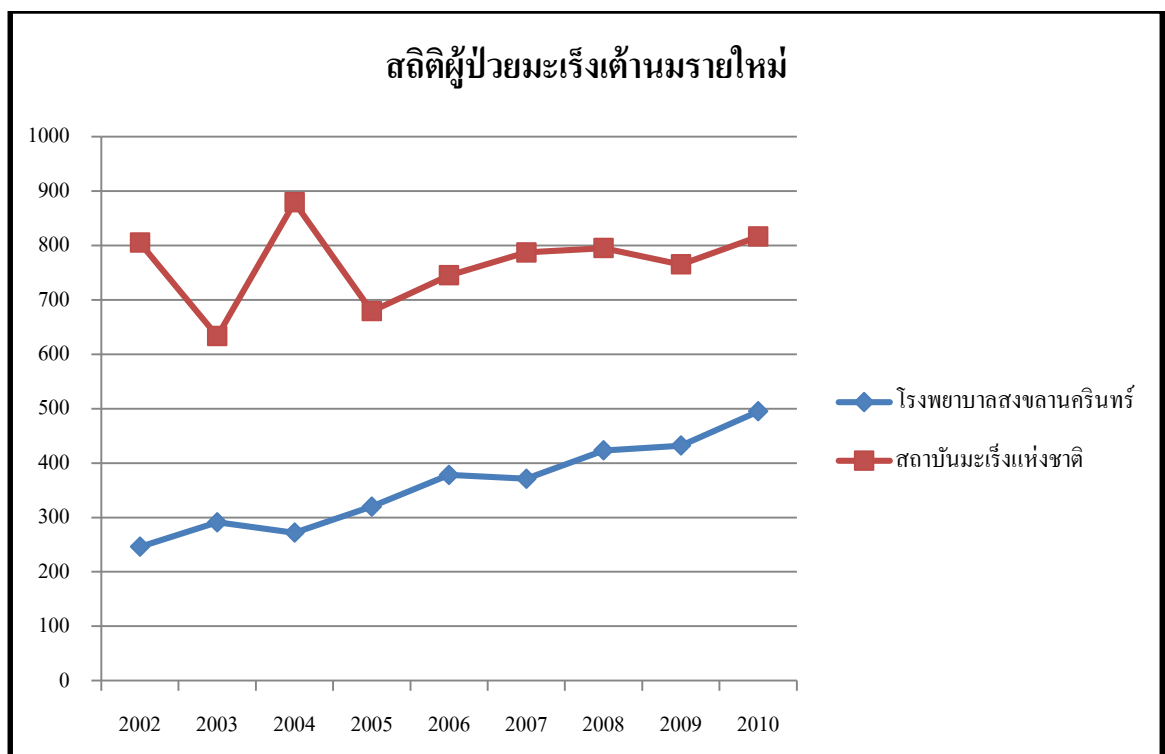
ในบทนี้กล่าวถึงความสำคัญและที่มาของการวิจัย งานวิจัยที่เกี่ยวข้อง วัตถุประสงค์ของการวิจัย ขอบเขตการวิจัย ประโยชน์ที่คาดว่าจะได้รับ และขั้นตอนการวิจัย ดังรายละเอียดต่อไปนี้

### 1.1 ความสำคัญและที่มาของการวิจัย

องค์การอนามัยโลกได้รายงานสถานการณ์ผู้ป่วยมะเร็งในปัจจุบัน พบว่ามะเร็งที่เกิดขึ้นบนโลกนี้มีมากกว่า 100 ชนิดแยกตามอาการ และตำแหน่งของโรค โดยในปี 2008 มีผู้ป่วยทั่วโลกที่ต้องเสียชีวิตจากโรคมะเร็งกว่า 7.6 ล้านคน ซึ่งคิดเป็น 13 เปอร์เซ็นต์ของการเสียชีวิตทั้งหมด [1] โดยที่มะเร็งเต้านมเป็นมะเร็งที่คร่าชีวิตผู้หญิงไปมากที่สุดเนื่องจากการเข้าถึงของบริการด้านสาธารณสุขในบางประเทศยังไม่ทั่วถึง ขณะที่ในประเทศไทยก็มีการเพิ่มขึ้นของจำนวนผู้ป่วยมะเร็งเต้านมรายใหม่อย่างมีนัยสำคัญในทุกปี [2], [3] ดังแสดงในภาพประกอบ 1-1

จากการค้นคว้าและวิจัยทางการแพทย์พบว่ามะเร็งเต้านมนั้นเกิดขึ้นได้จากหลายสาเหตุด้วยกัน แม้แต่อาหารที่รับประทานเข้าไปทุกวันก็ยังเป็นสาเหตุหนึ่งส่งผลต่อการเพิ่มขึ้นของปริมาณฮอร์โมนเอสโตรเจน (Estrogen) และฮอร์โมนโปรเจสเตอโรน (Progesterone) ซึ่งฮอร์โมนทั้งสองตัวนี้มีคุณสมบัติในการช่วยกระตุ้นให้ต่อมน้ำนมเจริญเติบโต และจะไปช่วยเร่งการเจริญเติบโตของก้อนมะเร็งเต้านมด้วย ในการวินิจฉัยมะเร็งเต้านมทางพยาธิวิทยาแพทย์จะทำการตัดชิ้นเนื้อที่ต้องสงสัยออกมามตรวจภายนอกร่างกาย เรียกวิธีนี้ว่า ไบออปซี (Biopsy) ซึ่งขั้นตอนนี้สามารถทำได้ 4 ลักษณะคือ (1) Excisional biopsy เป็นการตัดเอาก้อนทั้งหมดออกไปตรวจ (2) Incisional biopsy เป็นการตัดเอาบางส่วนของก้อนเนื้อออกไปตรวจ (3) Core biopsy คือการตัดเอาชิ้นเนื้อออกไปตรวจโดยใช้เข็มขนาดใหญ่ (4) Fine-needle aspiration biopsy เป็นการตัดเอาชิ้นเนื้อออกไปตรวจโดยใช้เข็มที่มีขนาดเล็ก ซึ่งเป็นวิธีที่มีความนิยมใช้กันมากในปัจจุบัน เมื่อได้ชิ้นเนื้อมาพยาธิแพทย์จะทำการตรวจด้วยสายตาเพื่อหารอยโรคและตัดชิ้นเนื้อบริเวณที่ต้องสงสัยไปตรวจทางเนื้อเยื่อวิทยา (Histopathology) โดยนำเนื้อเยื่อไปย้อมด้วยสี Hematoxylin & Eosin (H&E) หากวินิจฉัยพบว่าเป็นมะเร็งที่มีต้นกำเนิดมาจากเซลล์ส่วนที่เป็นท่อน้ำนม (Duct) พยาธิแพทย์จะทำการย้อมตัวรับฮอร์โมนเอสโตรเจน (Estrogen receptor, ER) และตัวรับฮอร์โมนโปรเจสเตอโรน

(Progesterone receptor, PR) โดยใช้เทคนิคทาง Immunohistochemistry (IHC) เป็นเทคนิคที่ใช้ในการตรวจหาแอนติเจน (Antigen) ในที่นี้พยาธิแพทย์จะกำหนดให้ตัวรับฮอร์โมนเอสโตรเจนและตัวรับฮอร์โมนโปรเจสเตอโรนที่มีอยู่ในเยื่อหุ้มนิวเคลียส (Nuclear membrane) เป็นแอนติเจนและเลือกใช้โปรตีนที่มีความจำเพาะกับแอนติเจนดังกล่าวมาเป็นแอนติบอดี (Antibody) ซึ่งพยาธิแพทย์จะใช้ Anti- $\alpha$ -oestrogen receptor monoclonal antibody (clone 1D5) สำหรับทำปฏิกิริยากับตัวรับฮอร์โมนเอสโตรเจนและใช้ Anti-progesterone receptor (clone PgR636) สำหรับทำปฏิกิริยากับตัวรับฮอร์โมนโปรเจสเตอโรน หลังจากนั้นทำการเติมสี DAB โดยสีนี้จะเข้าไปจับกับตำแหน่งที่เกิดปฏิกิริยาระหว่างแอนติเจนกับแอนติบอดี (Antigen - antibody complex) ซึ่งจะปรากฏเป็นสีน้ำตาลที่มีความเข้มแตกต่างกันตามความหนาแน่นของการเกิดปฏิกิริยา ซึ่งอาจกล่าวได้ว่าความเข้มของสีเป็นตัวบอกระดับความหนาแน่นของตัวรับฮอร์โมนเอสโตรเจนและโปรเจสเตอโรนที่มีอยู่บนเยื่อหุ้มนิวเคลียสของเซลล์นั้นๆ



ภาพประกอบ 1-1 จำนวนผู้ป่วยที่เป็นโรคมะเร็งเต้านม ณ ศูนย์มะเร็งโรงพยาบาลสงขลานครินทร์ และสถาบันมะเร็งแห่งชาติ ระหว่างปี ค.ศ. 2002 - 2010



การย้อมตัวรับฮอร์โมนเอสโตรเจนและโปรเจสเตอโรนนอกจากจะมีประโยชน์ในการพยากรณ์โรค (Prognosis) แล้วยังเป็นตัวบ่งชี้ (Indicator) ถึงความเป็นไปได้ที่จะทำการรักษาโดยใช้ฮอร์โมนบำบัด (Hormone therapy) โดยพยาธิแพทย์จะใช้สัดส่วนของจำนวนเซลล์มะเร็งผลบวก (Positive cancer cell) ซึ่งเป็นเซลล์มะเร็งที่มีตัวรับฮอร์โมนเอสโตรเจนและโปรเจสเตอโรนต่อจำนวนเซลล์มะเร็งทั้งหมด (Total cancer cell) โดยเป็นผลรวมของจำนวนเซลล์มะเร็งผลบวกและเซลล์มะเร็งผลลบ (Negative cancer cell) ซึ่งเป็นเซลล์มะเร็งที่ไม่มีตัวรับฮอร์โมนเอสโตรเจนและโปรเจสเตอโรน ค่าสัดส่วนดังกล่าวจะถูกนำมาใช้เพื่อกำหนดการให้ยาในกระบวนการรักษาด้วย ฮอร์โมนบำบัด ซึ่งยาที่ใช้จะมีคุณสมบัติในการเข้าไปจับกับฮอร์โมนเอสโตรเจน (Anti-estrogen) และฮอร์โมนโปรเจสเตอโรน (Anti-progesterone) เพื่อยับยั้งการเพิ่มจำนวนของเซลล์มะเร็งเต้านม

การรายงานผลตัวรับฮอร์โมนเอสโตรเจนและโปรเจสเตอโรนเป็นจำนวนนับ (Quantitative) ต่อจำนวนเซลล์มะเร็งทั้งหมดเป็นร้อยละจึงมีความสำคัญกว่าการรายงานผลในเชิงคุณภาพ (Qualitative) ซึ่งในปัจจุบันนั้นใช้วิธีการประมาณค่าอย่างคร่าว ๆ ด้วยสายตาหรือให้ผู้เชี่ยวชาญทำการนับจำนวนเซลล์ ซึ่งการที่จะจำแนกชนิดของเซลล์แต่ละเซลล์ให้มีความแม่นยำนั้นจะต้องดูเซลล์ที่อยู่รอบข้างด้วย วิธีที่กล่าวมาเป็นการวิเคราะห์ในเชิงพื้นผิว (Texture analysis) ซึ่งเป็นวิธีการที่ยากส่งผลให้ในบางครั้งค่าที่ได้เกิดความคลาดเคลื่อนเนื่องจากความเหนื่อยล้าในการทำงานและประสบการณ์ในการจำแนกเซลล์ของผู้เชี่ยวชาญแต่ละคนก็แตกต่างกัน

งานวิจัยชิ้นนี้จึงนำเสนอการประยุกต์ใช้วิธีวิเคราะห์ค่าคุณลักษณะเด่นในเชิงพื้นผิว (Texture feature) โดยระบบคอมพิวเตอร์เพื่อทำการวิเคราะห์และจำแนกชนิดของเซลล์ในภาพถ่ายจากกล้องจุลทรรศน์ (Microscopic image) ให้มีความแม่นยำมากยิ่งขึ้นและสามารถนำมาใช้สนับสนุนการวิเคราะห์จากผู้เชี่ยวชาญได้ ซึ่งจะสามารถลดระยะเวลาของกระบวนการวินิจฉัยโรคมะเร็งเต้านม นอกจากนั้นยังช่วยลดค่าใช้จ่ายและช่วยสนับสนุนการวินิจฉัยโดยผู้เชี่ยวชาญทำให้ผู้ป่วยสามารถเข้าถึงการรักษาได้ง่ายยิ่งขึ้น

## 1.2 งานวิจัยที่เกี่ยวข้อง

จากการทบทวนวรรณกรรมที่ผ่านมาพบความวิจัยมากมายที่ได้มีการพัฒนาระบบคอมพิวเตอร์เพื่อวิเคราะห์เซลล์มะเร็ง การวิเคราะห์เนื้อหาในบทความสามารถแบ่งออกเป็น 3 หัวข้อคือ การย้อมตัวรับฮอร์โมนในเซลล์มะเร็ง การพัฒนาระบบคอมพิวเตอร์เพื่อช่วยในการ

วิเคราะห์เซลล์มะเร็งเต้านม และการประยุกต์ใช้ค่าลักษณะเด่นเชิงพื้นผิวแบบจีแอลซีเอ็มกับภาพถ่ายทางการแพทย์

### 1.2.1. การย้อมตัวรับฮอร์โมนในเซลล์มะเร็งเต้านม

จากการศึกษาพบว่ากระบวนการย้อมเซลล์มะเร็งเต้านมนั้นมีหลายเทคนิคด้วยกัน แต่ทางพยาธิวิทยาจะนิยมใช้เทคนิค IHC เนื่องจากมีต้นทุนที่ต่ำเพราะสามารถจำแนกเซลล์ได้โดยใช้กล้องจุลทรรศน์แบบใช้แสง ซึ่งเป็นกล้องที่มีราคาไม่สูงมากนักและดูแลรักษาง่าย โดยจะมุ่งเน้นไปที่การย้อม PR และ ER เนื่องจากทั้งสององค์ประกอบนั้นสามารถนำมาใช้ในการทำนายการรักษาด้วยฮอร์โมนได้ จึงเป็นที่นิยมเพราะวิธีการดังกล่าวเป็นการรักษาที่ปลอดภัย จึงมีงานวิจัยที่หลากหลายที่พยายามจะแสดงให้เห็นว่าการย้อม ER และ PR มีความสำคัญต่อกระบวนการรักษาโรคมะเร็งเต้านม ซึ่งในบทความ [4] กล่าวถึงการนับจำนวนตัวรับฮอร์โมนเอสโตรเจนในภาพดิจิทัลของเซลล์มะเร็งเต้านมที่ผ่านกระบวนการย้อมสีด้วยเทคนิคทาง Immunohistochemistry โดยใช้โปรแกรม Photoshop และคุณสมบัติ Red-scale absorption นอกจากนี้ยังมีบทความ [5] ซึ่งได้กล่าวถึงการวิเคราะห์สถานะของตัวรับฮอร์โมนทั้ง PR และ ER ในเซลล์มะเร็งเต้านมที่ผ่านกระบวนการย้อมสีด้วยเทคนิคทาง Immunohistochemistry ถ่ายภาพที่กำลังขยายรวม 200 เท่า ประมวลผลด้วยซอฟต์แวร์ Win ROOF โดยใช้ปริภูมิสี Red-Green-Blue (RGB) และ Hue-Luminosity-Saturation (HLS) ทั้งสองบทความนั้นได้ยืนยันให้เห็นว่ากระบวนการย้อมตัวรับฮอร์โมนด้วยเทคนิค IHC นั้นมีความสำคัญ แต่อย่างไรก็ตามในขั้นตอนการวิเคราะห์ผลทั้งสองงานวิจัยที่ได้นำเสนอมาข้างต้นยังใช้วิธีการประมวลผลภาพด้วยโปรแกรมที่ไม่เป็นอัตโนมัติยังต้องใช้ความเชี่ยวชาญของพยาธิแพทย์ในการวิเคราะห์ภาพ ซึ่งบางครั้งอาจจะทำให้เกิดความล่าช้าในการประเมินผลการรักษาหรือทำนายอาการของโรคเนื่องจากจำนวนบุคลากรทางการแพทย์ที่มีอยู่ในปัจจุบันนั้นอาจจะไม่เพียงพอ ซึ่งในส่วนนี้ควรนำระบบคอมพิวเตอร์เข้ามาช่วยในการประมวลผลเพื่อเพิ่มความสะดวกรวดเร็วในการรักษามากยิ่งขึ้น

### 1.2.2. การพัฒนาระบบคอมพิวเตอร์เพื่อช่วยในการวิเคราะห์เซลล์มะเร็งเต้านม

จากปัญหาที่พบในการวิเคราะห์ชนิดของเซลล์หรือนับจำนวนเซลล์ พบว่าในปัจจุบันยังต้องมีการใช้ผู้เชี่ยวชาญเข้ามาช่วยในกระบวนการดังกล่าว ซึ่งจะเห็นได้ว่าผลที่ได้จากการวิเคราะห์ก็จะขึ้นอยู่กับประสบการณ์ของผู้เชี่ยวชาญแต่ละคน จึงส่งผลให้นักวิจัยบางกลุ่มพยายามที่จะนำระบบคอมพิวเตอร์เข้ามาทดแทนการประมวลผลโดยผู้เชี่ยวชาญซึ่งจะทำให้กระบวนการในการทำนายอาการของโรคเป็นไปอย่างรวดเร็วขึ้น ดังในบทความ [6] ได้กล่าวถึงการวิเคราะห์ ER ในภาพดิจิทัลของเซลล์มะเร็งเต้านมที่ผ่านกระบวนการย้อมสีด้วยเทคนิค IHC โดยใช้เทคนิค Thresholding ในปริภูมิสี  $L^*a^*b^*$  ซึ่งได้ทำการพัฒนาระบบให้สามารถตรวจวิเคราะห์ได้

อย่างอัตโนมัติ แต่อย่างไรก็ตามระบบที่ได้ยังมีค่าความผิดพลาดเป็นจำนวนมากเนื่องจากในขั้นตอนการวิเคราะห์เซลล์ได้ทำการตัดแบ่งและวิเคราะห์เซลล์เดี่ยวๆคราวละหนึ่งเซลล์ โดยไม่ได้ดูบริบทของเซลล์ใกล้เคียง ซึ่งในความเป็นจริงแล้วการวิเคราะห์โดยผู้เชี่ยวชาญนั้นจะทำการวิเคราะห์โดยมององค์ประกอบของเซลล์ใกล้เคียงด้วยจึงจะสามารถวิเคราะห์ได้อย่างแม่นยำ กระบวนการวิเคราะห์เซลล์โดยมองบริบทโดยรวมนั้นจำเป็นจะต้องใช้การวิเคราะห์ในเชิงพื้นผิวเข้ามาช่วย

### 1.2.3. การประยุกต์ใช้ลักษณะเด่นในเชิงพื้นผิวแบบจีแอลซีเอ็มกับภาพถ่ายทางการแพทย์

ในปัจจุบันนี้มีการนำการวิเคราะห์ในเชิงพื้นผิวมาประยุกต์ใช้กับภาพถ่ายทางการแพทย์เป็นจำนวนมาก โดยเฉพาะทฤษฎีเมทริกซ์ความน่าจะเป็นในการเกิดระดับเทา ซึ่งมีหลายบทความที่ได้กล่าวถึงทฤษฎีดังกล่าว อย่างเช่นบทความ [7] ได้ทำการประยุกต์ใช้ลักษณะเด่นในเชิงพื้นผิวแบบจีแอลซีเอ็มกับภาพอัลตราซาวด์ของเนื้องอกบริเวณตับเพื่อทำการจำแนกระหว่างเนื้องอกชนิดที่เป็นอันตรายกับเนื้องอกชนิดที่ไม่เป็นอันตรายส่วนในบทความ [8] ได้ทำการประยุกต์ใช้ลักษณะเด่นในเชิงพื้นผิวแบบจีแอลซีเอ็มเพื่อทำการวิเคราะห์หาการก่อตัวของกลุ่มเซลล์เชื่อมขนาดเล็กรอบๆก้อนมะเร็งเต้านม เปรียบเทียบกับทฤษฎีทางการแพทย์เชิงพื้นผิวแบบอื่น และในบทความ [9] ได้ประยุกต์ใช้ลักษณะเด่นในเชิงพื้นผิวแบบจีแอลซีเอ็มทั้ง 22 ลักษณะมาใช้ในการจำแนกบริเวณที่เป็นมะเร็งเต้านมจากภาพอัลตราซาวด์พร้อมทั้งทำการแบ่งนับสเกลระดับเทาของภาพใหม่ เนื่องจากค่าสเกลที่แตกต่างกันก็จะส่งผลให้ค่าความถูกต้องที่ได้จากกระบวนการจำแนกต่างกันด้วย ซึ่งทั้งสามบทความที่ยกมาก็แสดงให้เห็นทฤษฎีเมทริกซ์ความน่าจะเป็นในการเกิดระดับเทาที่มีการนำไปประยุกต์ใช้กับภาพในทางการแพทย์ที่หลากหลาย แต่อย่างไรก็ตามทฤษฎีดังกล่าวยังไม่ถูกนำมาประยุกต์ใช้กับภาพถ่ายจากกล้องจุลทรรศน์ โดยเฉพาะการวิเคราะห์ภาพถ่ายเซลล์มะเร็งเต้านมที่ผ่านการข้อมด้วยเทคนิค IHC

## 1.3 วัตถุประสงค์ของการวิจัย

- 1.3.1 เพื่อประยุกต์ใช้ลักษณะเด่นในเชิงพื้นผิวแบบจีแอลซีเอ็มสำหรับการวิเคราะห์ชนิดของเซลล์บนภาพถ่ายจากกล้องจุลทรรศน์
- 1.3.2 เพื่อออกแบบและพัฒนาขั้นตอนวิธีสำหรับการนับเซลล์มะเร็งเต้านม
- 1.3.3 เพื่อพัฒนาต้นแบบซอฟต์แวร์ช่วยวิเคราะห์ภาพเซลล์มะเร็งเต้านม

#### 1.4 ขอบเขตการวิจัย

- 1.4.1 วิเคราะห์หาค่าพารามิเตอร์ที่เหมาะสมในการคำนวณลักษณะเด่นเชิงพื้นผิวแบบจีแอลซีเอ็มของภาพเซลล์แต่ละชนิด อันได้แก่ ขนาดของหน้าต่าง ระยะห่าง มุมและจำนวนลักษณะเด่น
- 1.4.2 นำการคำนวณลักษณะเด่นเชิงพื้นผิวที่เหมาะสมแล้วมาออกแบบและพัฒนาขั้นตอนวิธีเพื่อใช้ในการนับเซลล์มะเร็งเต้านม และเปรียบเทียบกับผู้เชี่ยวชาญ
- 1.4.3 พัฒนาซอฟต์แวร์คอมพิวเตอร์ช่วยวิเคราะห์ภาพเซลล์มะเร็งเต้านมและรายงานผลการวิเคราะห์ให้กับพยาธิแพทย์หรือผู้เชี่ยวชาญได้

#### 1.5 ประโยชน์ที่คาดว่าจะได้รับ

- 1.5.1 เพิ่มความสะดวกและความถูกต้องแม่นยำให้กับพยาธิแพทย์และผู้เชี่ยวชาญในการวิเคราะห์ภาพเซลล์มะเร็งเต้านม
- 1.5.2 ลดค่าใช้จ่ายในการซื้อซอฟต์แวร์ราคาแพงและไม่ได้มีความจำเพาะกับโรค

#### 1.6 ขั้นตอนการวิจัย

- 1.6.1 ศึกษาระเบียบวิธีการประมวลผลภาพดิจิทัล
- 1.6.2 ศึกษากระบวนการได้มาซึ่งภาพถ่ายเซลล์มะเร็งเต้านมจากกล้องจุลทรรศน์
- 1.6.3 ศึกษาคุณสมบัติของภาพถ่ายเซลล์มะเร็งเต้านมจากกล้องจุลทรรศน์
- 1.6.4 ศึกษากระบวนการวิเคราะห์ในเชิงพื้นผิว
- 1.6.5 ศึกษาทฤษฎีค่าลักษณะเด่นในเชิงพื้นผิวแบบจีแอลซีเอ็ม
- 1.6.6 ทดสอบการตัดแบ่งภาพเซลล์มะเร็งเต้านมจากภาพถ่ายจากกล้องจุลทรรศน์ด้วยหน้าต่างที่มีขนาดแตกต่างกัน
- 1.6.7 ทดสอบการวิเคราะห์ภาพถ่ายดิจิทัลด้วยวิธีการวิเคราะห์ค่าลักษณะเด่นเชิงพื้นผิวแบบจีแอลซีเอ็มกับชุดข้อมูลภาพถ่ายดิจิทัลเซลล์ชนิดต่างๆ

- 1.6.8 ทดสอบการวิเคราะห์ภาพถ่ายดิจิทัลด้วยวิธีการวิเคราะห์ค่าลักษณะเด่นเชิงพื้นผิวแบบจีแอลซีเอ็มเพื่อหาค่ามุมและระยะห่างที่เหมาะสม
- 1.6.9 ทดสอบการลดขนาดข้อมูลของที่ได้จากกระบวนการวิเคราะห์ค่าลักษณะเด่นเชิงพื้นผิวแบบจีแอลซีเอ็ม
- 1.6.10 ทดสอบหาค่าสเกลระดับเทาที่เหมาะสมสำหรับการวิเคราะห์ภาพถ่ายเซลล์มะเร็งเต้านม
- 1.6.11 ประเมินประสิทธิภาพของอัลกอริทึมในการวิเคราะห์ชนิดของเซลล์
- 1.6.12 สรุปผลและเขียนรายงานฉบับสมบูรณ์

## บทที่ 2

### ทฤษฎีที่เกี่ยวข้อง

ในบทนี้กล่าวถึงทฤษฎีที่เกี่ยวข้องกับการประมวลผลภาพสำหรับงานวิจัยนี้ ได้แก่ การเตรียมภาพก่อนการประมวลผล การวิเคราะห์เชิงพื้นที่ผิวแบบจีแอลซีเอ็ม การจำแนกโดยใช้ต้นไม้ตัดสินใจ ดังรายละเอียดต่อไปนี้

#### 2.1 การเตรียมภาพก่อนการประมวลผล

ภาพดิจิทัล (Digital image) เป็นการแสดงผลของข้อมูลดิจิทัลรูปแบบหนึ่งที่ถูกนิยามด้วยฟังก์ชันสองมิติ  $f(x, y)$  โดย  $x$  และ  $y$  เป็นพิกัดระบุตำแหน่งของภาพ ซึ่งแอมพลิจูดที่ตำแหน่ง  $(x, y)$  ใดๆ เป็นค่าความเข้มแสงของภาพ (Intensity) ที่ตำแหน่งนั้น โดยที่ค่า  $x, y$  และแอมพลิจูดของฟังก์ชันเป็นค่าจำกัด (Finite value) แบบไม่ต่อเนื่อง (Discrete quantity) ถ้ากำหนดให้ฟังก์ชัน  $f(x, y)$  มีขนาด  $m$  แถว และ  $n$  คอลัมน์ และพิกัดของจุดกำเนิดของภาพอยู่ที่ตำแหน่ง  $(x, y) = (0, 0)$  จะทำให้สามารถเขียนฟังก์ชันของภาพเป็นสมการซึ่งอยู่ในรูปของเมทริกซ์ได้ดังสมการ (2-1)

$$f(x, y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,n-1) \\ f(1,0) & f(1,1) & \dots & f(1,n-1) \\ \vdots & \vdots & & \vdots \\ f(m-1,0) & f(m-1,1) & \dots & f(m-1,n-1) \end{bmatrix} \quad (2-1)$$

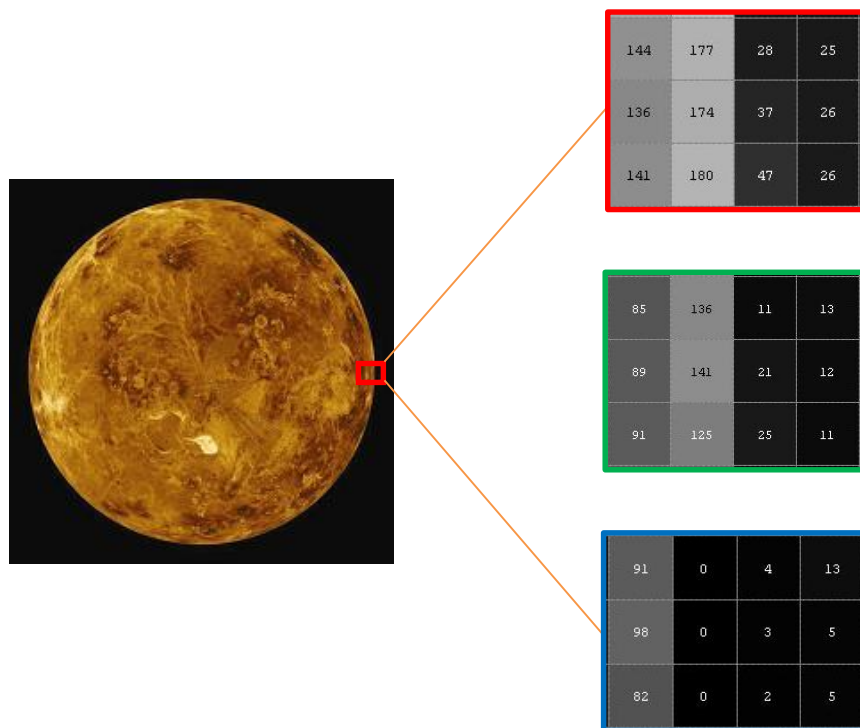
ค่าแต่ละค่าที่อยู่ภายในเมทริกซ์จะเรียกว่า พิกเซล (Pixel) มาจากคำว่า Picture element ซึ่งแปลว่า องค์ประกอบภาพ เป็นหน่วยพื้นฐานของภาพ ซึ่งภาพหนึ่งๆ จะประกอบด้วยพิกเซลจำนวนมากที่ถูกระบุด้วยพิกัดตำแหน่ง  $(x, y)$

ในปัจจุบันเนื่องจากมีความก้าวหน้าทางเทคโนโลยีเป็นจำนวนมากทำให้ภาพดิจิทัลสามารถแสดงผลได้เหมือนภาพสีที่ถูกถ่ายด้วยฟิล์มถ่ายภาพทั่วไป ไฟล์ภาพสีดิจิทัลหรือภาพสี RGB เป็นภาพที่เกิดจากการรวมกันขององค์ประกอบสี 3 สี คือ สีแดง (Red, R) สีเขียว (Green, G) และสีน้ำเงิน (Blue, B) ซึ่งในหนึ่งพิกเซลจะประกอบไปด้วยค่าของสีแต่ละสีที่มีความ

แตกต่างกันถึง 256 ระดับ [10], [11] คือตั้งแต่ 0 - 255 หรือ 0 - 1 (Normalized value) ดังนั้นภาพสีจึงมีระดับสีที่ต่างกันถึง  $256^3 = 16,777,216$  ระดับ ทำให้แต่ละพิกเซลมีขนาดถึง 24 บิต หรือ 3 ไบต์ สำหรับเก็บข้อมูล ดังแสดงในภาพประกอบ 2-1 ภาพสี RGB เป็นภาพที่มีข้อมูลขนาดใหญ่ ข้อมูลจะถูกจัดเก็บอยู่ในเชิงสามมิติที่เกิดจากการเรียงตัวซ้อนกันของเมทริกซ์ความเข้มขององค์ประกอบสีหลักทั้งสามเมทริกซ์ ทำให้การที่จะนำภาพสี RGB มาทำการประมวลผลโดยตรงนั้นเป็นการยาก และใช้เวลาในการประมวลผลค่อนข้างนาน ส่วนใหญ่จึงนิยมทำการแปลงปริภูมิสีเป็นภาพระดับเทา ซึ่งเป็นการรวมองค์ประกอบสีแดง สีเขียว และสีน้ำเงินในปริภูมิสีเข้าด้วยกัน [12] โดยอาศัยสมการที่ (2-2)

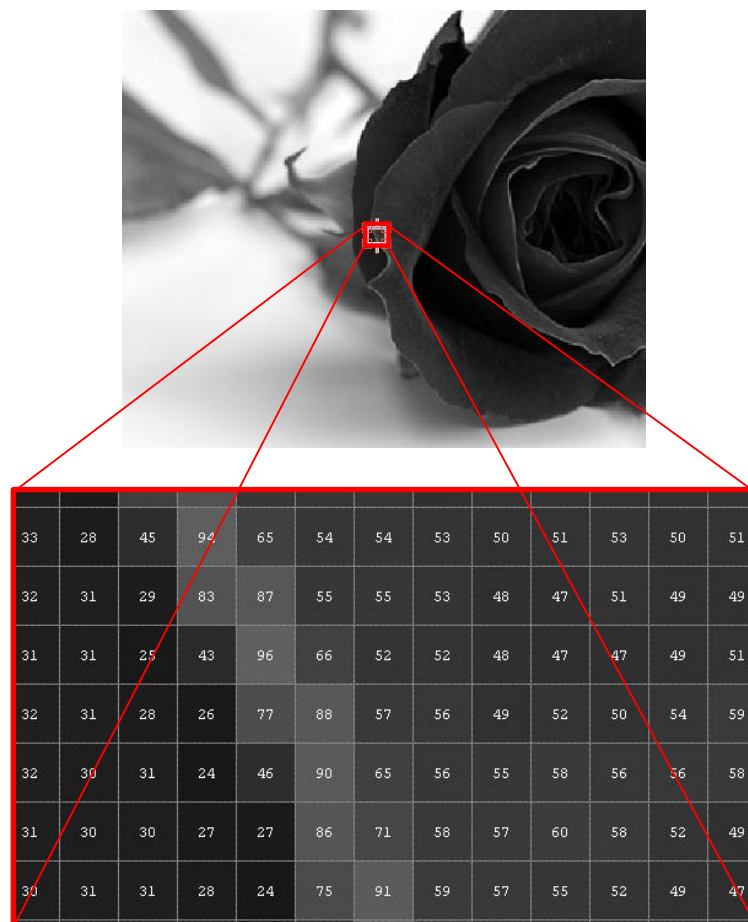
$$Gray = (0.2989 \times R) + (0.5870 \times G) + (0.1140 \times B) \quad (2-2)$$

เมื่อ  $Gray$  คือ ภาพระดับเทา  
 $R, G$  และ  $B$  คือ องค์ประกอบสีแดง สีเขียว และสีน้ำเงิน ตามลำดับ



ภาพประกอบ 2-1 ภาพสี RGB และข้อมูลของภาพ

ภาพระดับเทา (Gray-scale image) หรืออาจเรียกว่า Gray level image หรือ Intensity image เป็นภาพที่แต่ละพิกเซลมีค่าอยู่ระหว่าง 0 (สีดำ) ถึง 255 (สีขาว) หรือ 0 - 1 (Normalized value) แต่ละพิกเซลมีขนาด 8 บิต หรือ 1 ไบต์สำหรับเก็บข้อมูล ดังแสดงในภาพประกอบ 2-2 ภาพระดับเทาที่มีขนาดของข้อมูลไม่ใหญ่มาก แต่ก็มีค่าความแตกต่างของระดับความเข้มมากพอที่จะแสดงรายละเอียดของภาพได้ และสามารถนำไปทำการคำนวณทางคณิตศาสตร์เพื่อบ่งบอกถึงคุณสมบัติของภาพได้ง่ายกว่าการใช้ภาพสี RGB จึงนิยมใช้ภาพชนิดนี้ในการประมวลผลภาพถ่ายดิจิทัล



ภาพประกอบ 2-2 ภาพ Gray - scale image และข้อมูลของภาพ



โดยปกติแล้วภาพระดับเทาจะมีทั้งหมด 256 ระดับซึ่งมีค่าตั้งแต่ 0 ถึง 255 แต่ในบางครั้งจำเป็นต้องมีการแบ่งสเกลใหม่เพื่อลดขนาดข้อมูล หรือรวมข้อมูลที่มีระดับใกล้เคียงกันให้อยู่ในระดับเดียวกันซึ่งกระบวนการนี้จะช่วยในการลดสัญญาณรบกวนภาพในภาพ หรือแม้กระทั่งช่วยให้การประมวลผลของภาพมีความรวดเร็วและแม่นยำมากยิ่งขึ้น ซึ่งการแบ่งสเกลใหม่ของภาพระดับเทาสามารถคำนวณได้โดยสมการที่ 2-3

$$G_{new} = \frac{I \cdot (L-1) + (G_{max} - L \cdot G_{min})}{G_{max} - G_{min}} \quad (2-3)$$

โดย	$I$	คือ ค่าระดับเทาของภาพ
	$G_{new}$	คือ ค่าระดับเทาของการแบ่งสเกลใหม่
	$G_{min}$	คือ ค่าระดับเทาต่ำสุด
	$G_{max}$	คือ ค่าระดับเทาสูงสุด
	$L$	คือ ค่าสเกลสูงสุดของระดับเทาที่ต้องการ

## 2.2 การวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็ม

พื้นผิว (Texture) คือ บริเวณผิวนอกของวัตถุต่างๆที่ปรากฏให้เห็นและสามารถรับรู้ได้ด้วยกายสัมผัสและตาสัมผัส ก่อให้เกิดความรู้สึกในลักษณะต่างๆกันไป เช่น หยาบ ละเอียด วาว มัน ค้าน และขรุขระ แต่ถ้าหากกล่าวถึงพื้นผิวของรูปภาพแล้ว มักจะหมายถึงการกระจายตัวของค่าระดับเทาภายในพิกเซลต่างๆบนพริมีทีฟ (Primitive) ที่ก่อให้เกิดเป็นลักษณะที่สามารถรับรู้ออกมาเป็นอารมณ์ได้ อย่างเช่นถ้าระดับเทาบนแต่ละพิกเซลในพริมีทีฟมีระดับที่ใกล้เคียงกันก็จะทำให้พริมีทีฟนั้นมองดูมีความราบเรียบ แต่ถ้าหากระดับเทาในพริมีทีฟมีความแตกต่างกันมากก็จะทำให้ภาพมองดูขรุขระ ทั้งนี้ขึ้นอยู่กับลักษณะและรูปร่างของการกระจายตัวด้วย โดยที่พริมีทีฟคือการสร้างบล็อกของพิกเซลที่ประกอบด้วยกลุ่มของระดับเทานั้นเอง

การวิเคราะห์เชิงพื้นผิว (Texture analysis) เป็นทฤษฎีทางคณิตศาสตร์ที่ใช้เพื่อทำการแปลงลักษณะของพื้นผิวที่มนุษย์สามารถรับรู้ได้ให้ออกมาอยู่ในรูปของค่าทางคณิตศาสตร์ค่าหนึ่ง ซึ่งค่าดังกล่าวสามารถบ่งบอกถึงลักษณะของพื้นผิวทั้งหมดได้ [13]–[15] การวิเคราะห์ในเชิงพื้นผิวสามารถแบ่งออกได้เป็น 2 ทิศทางคือ (1) การวิเคราะห์ในเชิงสังเคราะห์เป็นการนำค่าทางคณิตศาสตร์ที่บ่งบอกลักษณะของพื้นผิวมาผ่านกระบวนการเพื่อสร้างเป็นภาพระดับเทาขึ้นมา ซึ่ง

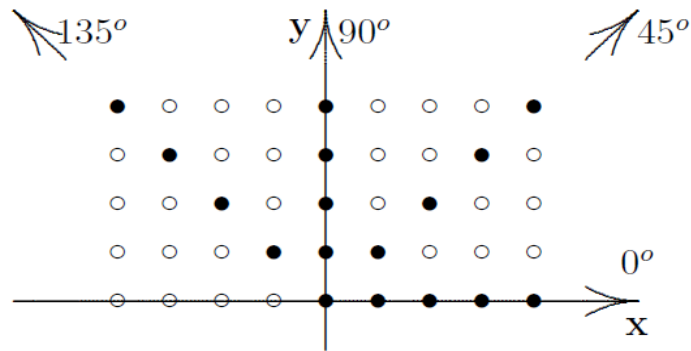
การวิเคราะห์ในส่วนนี้ไม่ค่อยมีผู้ให้ความสนใจมากนัก (2) การวิเคราะห์ในเชิงสถิติเป็นกระบวนการทางคณิตศาสตร์ที่ใช้เพื่อหาค่าที่แสดงถึงความสัมพันธ์ของค่าระดับเทาภายในพริมิตีฟค่าที่ได้จากกระบวนการนี้จะถูกเรียกว่าค่าคุณลักษณะเด่น การวิเคราะห์ในเชิงสถิติสามารถแบ่งออกได้เป็น 3 ชนิดดังนี้

**2.2.1 การวิเคราะห์พื้นผิวในเชิงสถิติอันดับหนึ่ง** เป็นการวิเคราะห์ข้อมูลภายในพริมิตีฟโดยใช้ค่าระดับเทาจากภาพโดยไม่คำนึงถึงพิกเซลใกล้เคียง โดยบางครั้งอาจทำการวิเคราะห์ข้อมูลจากฮิสโตแกรม (Histogram) ของพริมิตีฟที่สนใจ ซึ่งค่าลักษณะเด่นที่ได้จากการวิเคราะห์ในเชิงสถิติอันดับหนึ่งมีอยู่ด้วยกันหลายลักษณะเช่น ค่าเฉลี่ย (Average) ค่าความแปรปรวน (Variance) ค่าความเบ้ (Skewness) ค่าพลังงาน (Energy) ค่าเอนโทรปี (Entropy) เป็นต้น

**2.2.2 การวิเคราะห์พื้นผิวในเชิงสถิติอันดับสอง** เป็นการวิเคราะห์ข้อมูลภายในพริมิตีฟโดยใช้ค่าระดับเทาจากภาพโดยคำนึงถึงความสัมพันธ์ระหว่างพิกเซลที่สนใจกับพิกเซลใกล้เคียงอีกหนึ่งพิกเซลภายใต้กรอบข้อกำหนดของระยะห่างและมุมที่กระทำระหว่างคู่พิกเซล ซึ่งค่าลักษณะเด่นที่ได้จากการวิเคราะห์ในเชิงสถิติอันดับสองมีอยู่ด้วยกันหลายลักษณะเช่น มิติเศษส่วน (Fractal dimension) เมทริกซ์ค่าความน่าจะเป็นในการเกิดระดับเทาหรือจีแอลซีเอ็ม (Gray-level co-occurrence matrix, GLCM) เป็นต้น

**2.2.3 การวิเคราะห์พื้นผิวในเชิงสถิติอันดับสูง** เป็นการวิเคราะห์ข้อมูลภายในพริมิตีฟโดยใช้ค่าระดับเทาจากภาพโดยคำนึงถึงความสัมพันธ์ระหว่างพิกเซลที่สนใจกับพิกเซลใกล้เคียงตั้งแต่สองพิกเซลขึ้นไป สำหรับในงานวิจัยนี้ได้ให้ความสนใจกับการวิเคราะห์พื้นผิวในเชิงสถิติอันดับสองโดยมุ่งเน้นไปที่การวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็ม [16]

การวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็ม เป็นอีกวิธีหนึ่งที่ใช้แยกค่าคุณลักษณะเด่นของพื้นผิวในทางสถิติอันดับสอง [17]–[20] โดยเป็นการวัดค่าความน่าจะเป็นในการเกิดคู่ของค่าระดับเทาพร้อมกันบนภาพ ภายใต้กรอบข้อกำหนดของระยะการกระจัด  $d$  และมุม  $\theta$  ซึ่งผลที่ได้จะอยู่ในรูปของเมทริกซ์ที่มีค่าตัวเลขประจำแถวและคอลัมน์เท่ากับค่าตัวเลขที่ระบุระดับเทาของภาพ หรืออาจกล่าวได้ว่าเมทริกซ์จีแอลซีเอ็มเป็นการแสดงความถี่สัมพันธ์ระหว่างสองพิกเซลที่มีระยะห่างระหว่างกัน  $(\Delta x, \Delta y)$  ซึ่งเกิดขึ้นภายในพื้นที่ที่กำหนด ค่าระดับเทาของพิกเซลที่พิจารณาจะเป็นตัวระบุตำแหน่งของพิกเซล  $(i, j)$  ที่เรากำลังให้ความสนใจ ซึ่งสามารถแสดงความสัมพันธ์ขององค์ประกอบเมทริกซ์ได้คือ  $P(i, j | \Delta x, \Delta y)$  โดยที่  $\Delta x = d \cos \theta$  และ  $\Delta y = d \sin \theta$  ซึ่งจะสนใจเฉพาะมุมที่มีค่าเท่ากับ  $0^\circ$   $45^\circ$   $90^\circ$  และ  $135^\circ$  ดังภาพประกอบ 2-3 สมมติให้ภาพที่เราต้องการวิเคราะห์มีค่าความคมชัด (Resolution) ในแนวแกนนอน  $M$  และแนวแกนตั้ง  $N$  ให้ค่าระดับเทาในแต่ละพิกเซลมีการแบ่งออกเป็น  $G$  ระดับโดยมีค่าตั้งแต่ 0 ถึง  $G-1$



ภาพประกอบ 2-3 พิกเซลข้างเคียงในมุม 0° 45° 90° และ 135° [17]

กำหนดให้  $M \times N$  เป็นค่าความคมชัดของภาพ และสามารถบ่งบอกถึงจำนวนพิกเซลที่ประกอบขึ้นเป็นภาพ ซึ่งภาพจะถูกนิยามด้วย  $f(m,n)$  เป็นค่าระดับเทาในตำแหน่ง  $(m,n)$  จากคุณสมบัติของภาพสามารถคำนวณเมทริกซ์จีแอลซีเอ็ม ได้จากสมการที่ (2-3) ดังตัวอย่างการคำนวณในภาพประกอบ 2-4

$$P(i, j | \Delta x, \Delta y) = WQ(i, j | \Delta x, \Delta y) \tag{2-3}$$

โดยที่

$$W = \frac{1}{(M - \Delta x)(N - \Delta y)}$$

$$Q(i, j | \Delta x, \Delta y) = \sum_{n=1}^{N-\Delta y} \sum_{m=1}^{M-\Delta x} A$$

และ  $A = \begin{cases} 1 & \text{if } f(m,n) = i \text{ and } f(m + \Delta x, n + \Delta y) = j \\ 0 & \text{else where} \end{cases}$

IMAGE	P(i, j; 1, 0)				
		j=0	1	2	3
0 1 1 2 3	i= 0	1/20	2/20	1/20	0
0 0 2 3 3	1	0	1/20	3/20	0
0 1 2 2 3	2	0	0	3/20	5/20
1 2 3 2 2	3	0	0	2/20	2/20
2 2 3 3 2					

ภาพประกอบ 2-4 ตัวอย่างการคำนวณเมทริกซ์จีแอลซีเอ็มที่ระยะห่าง 1 พิกเซลและมุม 0 องศา [17]

หากเราทำการเปรียบเทียบกับฮิสโตแกรมของภาพ ซึ่งบ่งบอกถึงการกระจายตัวของค่าระดับเทาของภาพแล้ว เมทริกซ์จีแอลซีเอ็ม ก็จะเป็นตัวที่บ่งบอกถึงความน่าจะเป็นร่วมกันของค่าระดับเทา ณ ตำแหน่งเชิงสัมพัทธ์หนึ่งๆ เมทริกซ์จีแอลซีเอ็มนี้มีคุณสมบัติในการบอกลักษณะการกระจายเชิงพื้นที่ของค่าระดับเทาของภาพ ซึ่งสามารถนำมาทำการพิจารณาโดยใช้คุณสมบัติของพื้นผิวภาพบางประการ [21]–[24] ได้ดังแสดงในตารางที่ 2-1 และ 2-2

ตารางที่ 2-1 นิพจน์ที่ใช้ในการคำนวณคุณลักษณะเด่นเชิงพื้นผิวแบบจีแอลซีเอ็ม

นิพจน์	ความหมาย
$p(i, j)$	ค่าความน่าจะเป็นของคู่ระดับเทา $(i, j)$ th ภายในเมทริกซ์
$L$	ค่าสเกลสูงสุดของค่าระดับเทา
$v$	ค่าเฉลี่ยของ $p(i, j)$
$p_x(i)$	$\sum_{j=1}^L p(i, j)$
$p_y(i)$	$\sum_{i=1}^L p(i, j)$
$\mu_x$	$\sum_i \sum_j i \cdot p(i, j)$
$\mu_y$	$\sum_i \sum_j j \cdot p(i, j)$
$\sigma_x^2$	$\sum_i \sum_j (i - \mu_x)^2 \cdot p(i, j)$
$\sigma_y^2$	$\sum_i \sum_j (j - \mu_y)^2 \cdot p(i, j)$
$p_{x+y}(k)$	$\sum_{i=1, i+j=k}^L \sum_{j=1}^L p(i, j) \quad , k = 2, 3, \dots, 2L$
$p_{x-y}(k)$	$\sum_{i=1,  i+j =k}^L \sum_{j=1}^L p(i, j) \quad , k = 0, 1, \dots, L-1$
$H_X$	$-\sum_i p_x(i) \cdot \log(p_x(i))$
$H_Y$	$-\sum_i p_y(i) \cdot \log(p_y(i))$
$H_{XY}$	$-\sum_i \sum_j p(i, j) \cdot \log(p(i, j))$

ตารางที่ 2-1 นิพจน์ที่ใช้ในการคำนวณคุณลักษณะเด่นเชิงพื้นผิวแบบจีแอลซีเอ็ม (ต่อ)

นิพจน์	ความหมาย
$HXY_1$	$-\sum_i \sum_j p(i, j) \cdot \log(p_x(i)p_y(j))$
$HXY_2$	$-\sum_i \sum_j p_x(i)p_y(j) \cdot \log(p_x(i)p_y(j))$

ตารางที่ 2-2 คุณลักษณะเด่นเชิงพื้นผิวแบบจีแอลซีเอ็ม

ลำดับ	คุณลักษณะเด่น	สมการ
f1	ค่าสหสัมพันธ์อัตโนมัติ Autocorrelation	$\sum_i \sum_j (i \cdot j) p(i, j)$
f2	ค่าความต่าง Contrast	$\sum_i \sum_j  i - j ^2 p(i, j)$
f3	ค่าสหสัมพันธ์แบบที่ 1 Correlation I	$\sum_i \sum_j \frac{(i - \mu_x)(j - \mu_y)}{\sigma_x \sigma_y} p(i, j)$
f4	ค่าสหสัมพันธ์แบบที่ 2 Correlation II	$\sum_i \sum_j \frac{(i \cdot j) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
f5	ค่าความโดดเด่นของคลัสเตอร์ Cluster prominence	$\sum_i \sum_j (i + j - \mu_x - \mu_y)^4 p(i, j)$
f6	ค่าความเปลี่ยนแปลงของคลัสเตอร์ Cluster shade	$\sum_i \sum_j (i + j - \mu_x - \mu_y)^3 p(i, j)$
f7	ค่าความไม่เป็นเอกภาพ Dissimilarity	$\sum_i \sum_j  i - j  \cdot p(i, j)$
f8	ค่าพลังงาน Energy	$\sum_i \sum_j p(i, j)^2$
f9	ค่าเอนโทรปี Entropy	$-\sum_i \sum_j p(i, j) \cdot \log(p(i, j))$
f10	ค่าความเป็นเอกภาพแบบที่ 1 Homogeneity I	$\sum_i \sum_j \frac{p(i, j)}{1 +  i - j }$

ตารางที่ 2-2 (ต่อ) คุณลักษณะเด่นเชิงพื้นที่แบบจีแอลซีเอ็ม

ลำดับ	คุณลักษณะเด่น	สมการ
f11	ค่าความเป็นเอกภาพแบบที่ 2 Homogeneity II	$\sum_i \sum_j \frac{p(i, j)}{1 +  i - j ^2}$
f12	ค่าความน่าจะเป็นสูงสุด Maximum probability	$\max_{i, j} p(i, j)$
f13	ค่าผลรวมกำลังสอง Sum of squares	$\sum_i \sum_j (i - v)^2 p(i, j)$
f14	ค่าผลรวมเฉลี่ย Sum average	$\sum_{i=2}^{2L} i \cdot p_{x+y}(i)$
f15	ค่าผลรวมเอนโทรปี Sum entropy	$-\sum_{i=2}^{2L} p_{x+y}(i) \cdot \log(p_{x+y}(i))$
f16	ค่าผลรวมของความเบี่ยงเบน Sum variance	$\sum_{i=2}^{2L} (i - \text{Sum entropy})^2 \cdot p_{x+y}(i)$
f17	ค่าผลต่างของความเบี่ยงเบน Difference variance	$\sum_{i=0}^{L-1} i^2 \cdot p_{x-y}(i)$
f18	ค่าผลต่างเอนโทรปี Difference entropy	$-\sum_{i=0}^{L-1} p_{x-y}(i) \cdot \log(p_{x-y}(i))$
f19	ค่าการวัดข้อมูลของสหสัมพันธ์แบบที่ 1 Information measure of correlation I	$\frac{HXY - HXY1}{\max(HX, HY)}$
f20	ค่าการวัดข้อมูลของสหสัมพันธ์แบบที่ 2 Information measure of correlation II	$\sqrt{1 - e^{-2(HXY2 - HXY)}}$
f21	ค่านอมอลไลซ์ผลต่างผกผัน Inverse difference normalized	$\sum_i \sum_j \frac{p(i, j)}{1 +  i - j ^2 / L}$
f22	ค่านอมอลไลซ์โมเมนต์ผลต่างผกผัน Inverse difference moment normalized	$\sum_i \sum_j \frac{p(i, j)}{1 + (i - j)^2 / L}$

### 2.3 การจำแนกโดยใช้ต้นไม้ตัดสินใจ

ต้นไม้ตัดสินใจ (Decision tree) เป็นการเสนอข้อมูลทั้งทางเลือก สภาวะการณื ผลตอบแทนในลักษณะของแผนภาพแขนง โดยมีสัญลักษณ์ต่างๆ [25]–[27] ในบางครั้งนั้นเรื่องที่กำลังพิจารณาจะต้องมีการตัดสินใจหลายครั้ง ซึ่งการใช้วิธีเมทริกซ์นั้นไม่สามารถทำได้ ดังนั้นจึงจำเป็นต้องใช้การตัดสินใจแบบกิ่งก้านสาขา นิยมใช้เมื่อมีเหตุการณ์ตั้งแต่สองเหตุการณ์ขึ้นไปที่เกิดขึ้นต่อเนื่องกัน หรือกรณีที่มีการตัดสินใจซึ่งมีความสัมพันธ์ต่อเนื่องกัน โดยผู้ตัดสินใจมีทางเลือกหลายทาง แต่ไม่ทราบผลลัพธ์ของแต่ละทางเลือก และการตัดสินใจในครั้งหลังขึ้นอยู่กับผลของการตัดสินใจที่เกิดขึ้นก่อนหน้านี้ ในต้นไม้ตัดสินใจจะมีโครงสร้างแบบลำดับชั้น (Hierarchy) ซึ่งจะประกอบไปด้วยโหนด (Node) ซึ่งแต่ละโหนดจะมีคุณลักษณะเป็นตัวทดสอบสำหรับโหนดที่อยู่บนสุดเรียกว่า โหนดราก (Root node) กิ่งของต้นไม้ (Branch) แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบ และใบ (Leaf) ซึ่งเป็นสิ่งที่อยู่ล่างสุดของต้นไม้ตัดสินใจแสดงถึงผลลัพธ์ที่ได้จากขั้นตอนการตัดสินใจ

### 2.4 การคัดเลือกคุณลักษณะเด่น

ในขั้นตอนการจำแนกค่าคุณลักษณะเด่น โดยอาศัยการเรียนรู้แบบมีผู้สอน (Supervised learning) จำเป็นจะใช้ตัวแปรเป็นจำนวนมากในขั้นตอนการสอนเพื่อให้ระบบที่ได้ ออกมามีความเสถียรมากที่สุด แต่การใช้ตัวแปรที่มากเกินไปก็ไม่ได้หมายความว่าระบบที่ได้ ออกมาจะมีค่าความแม่นยำที่สูง ตัวแปรบางตัวมีความเกี่ยวข้องกับการทายผลอย่างมาก (Relevance variable) ซึ่งจำเป็นที่จะต้องนำมาใช้ในขั้นตอนการสอนระบบ แต่ตัวแปรบางตัวมีความซ้ำซ้อน (Redundant variable) กับตัวแปรที่นำมาใช้ในการสอนไปก่อนหน้านี้แล้ว ตัวแปรเหล่านี้จึงไม่มีความจำเป็นที่จะนำมาสอนระบบอีก เพื่อลดความยุ่งยากในขั้นตอนการสอนและความซับซ้อนของระบบ การคัดเลือกตัวแปรที่จำเป็นจึงเป็นอีกกระบวนการหนึ่งที่สำคัญในการสร้างขั้นตอนการจำแนกค่าคุณลักษณะเด่น โดยอาศัยการเรียนรู้แบบมีผู้สอน ในปัจจุบันนี้มีระเบียบวิธีมากมายที่ถูกพัฒนาขึ้นเพื่อนำมาใช้ในการคัดเลือกตัวแปร ตัวอย่างเช่น การวัดค่าเกินของข้อมูล การวัดค่าสหสัมพันธ์ ซึ่งทั้งสองวิธีที่กล่าวมาถูกนำมาใช้ในงานวิจัยอย่างแพร่หลายรวมทั้งงานวิจัยนี้ด้วย [28]–[33]

### 2.4.1 ค่าเกณฑ์ของข้อมูล

การวัดค่าเกณฑ์ของข้อมูล (Information gain, IG) เป็นการวัดค่าเอนโทรปีของข้อมูลที่จะนำมาใช้ในการประเมินเพื่อหาตัวแปรที่เหมาะสมสำหรับการจำแนกกลุ่มของข้อมูล โดยนิยามเอนโทรปีของเซตตัวอย่าง  $S$  คือ  $E(S)$  ซึ่งสามารถแสดงโดยสมการ (2-4)

$$E(S) = -\sum_{j=1}^n P_s(j) \log_2 P_s(j) \quad (2-4)$$

โดย  $S$  คือตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆ กรณี  
 $P_s(j)$  คืออัตราส่วนของกรณีใน  $S$  ที่ตัวแปรตามหรือผลลัพธ์มีค่า  $j$

เมื่อพิจารณาจากสมการ (2-4) จะพบว่าค่าของเอนโทรปีจะอยู่ในช่วง 0 ถึง 1 เท่านั้น ยิ่งค่าเข้าใกล้ศูนย์มากเท่าใด ยิ่งเป็นการบ่งบอกว่าตัวแปรต้นที่เลือกใช้จะให้ค่าตัวแปรตามตรงตามความเป็นจริงมากเท่านั้น โดยค่าเกณฑ์ของข้อมูล ซึ่งสามารถหาได้จากโดยสมการ (2-5)

$$Gain(S, A) = E(S) - \sum_{v=value(A)} \frac{|S_v|}{|S|} E(S_v) \quad (2-5)$$

โดย  $S$  คือตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตามหลายๆ กรณี  
 $E$  คือเอนโทรปีของตัวอย่าง  
 $A$  คือตัวแปรต้นที่พิจารณา  
 $value(A)$  คือเซตของค่าของ  $A$  ที่เป็นไปได้  
 $S_v$  คือตัวอย่างที่  $A$  มีค่า  $v$  ทั้งหมด

หากค่าเกณฑ์ของข้อมูลยิ่งมากจะแสดงให้เห็นว่าหลังจากที่แบ่งตัวอย่าง  $S$  ด้วย  $A$  แล้วในแต่ละชุดที่แบ่งได้จะมีเอนโทรปีเข้าใกล้ศูนย์มากยิ่งขึ้น ทำให้สามารถเลือกตัวแปรต้นที่เหมาะสมสำหรับนำมาใช้ในการตัดสินใจได้ดียิ่งขึ้น [34]–[36]

### 2.4.2 การคัดเลือกคุณลักษณะเด่นด้วยค่าสหสัมพันธ์

การคัดเลือกคุณลักษณะเด่นด้วยค่าสหสัมพันธ์ (Correlation feature selection, CFS) เป็นอีกเป็นวิธีการหนึ่งที่สามารถนำมาใช้ในการเลือกตัวแปรต้นหรือข้อมูลที่เหมาะสมได้จากสมมติฐานที่ว่า ตัวแปรต้นที่ดีจะต้องมีความสัมพันธ์กับตัวแปรตามโดยตรง โดยไม่แปรผันตาม



ตัวแปรควบคุม ยิ่งตัวแปรต้นมีแนวโน้มที่จะเกี่ยวข้องกับตัวแปรตามมากเท่าใดยิ่งมีความสำคัญในการเลือกนำมาใช้เพื่อให้สามารถตัดสินใจได้ดียิ่งขึ้น [37] ซึ่งการวิเคราะห์ค่าสหสัมพันธ์สามารถแสดงได้โดยสมการ (2-5)

$$CFS = \max_{S_k} \left[ \frac{r_{cf1} + r_{cf2} + \dots + r_{cfk}}{\sqrt{k + 2(r_{f1f2} + \dots + r_{fifi} + \dots + r_{fkf1})}} \right] \quad (2-5)$$

โดย  $S_k$  คือเซตย่อยของค่าคุณลักษณะเด่นจำนวน  $k$  คุณลักษณะ  
 $r_{cf}$  คือค่าสหสัมพันธ์ระหว่างค่าคุณลักษณะเด่นกับคลาส  
 $r_{fifi}$  คือค่าสหสัมพันธ์ระหว่างค่าคุณลักษณะเด่นสองคุณลักษณะ  
 ซึ่งทั้ง  $r_{cf}$  และ  $r_{fifi}$  สามารถคำนวณได้โดยอาศัยความสัมพันธ์ของสมการ (2-6)

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)S_x S_y} \quad (2-6)$$

โดย  $S_x, S_y$  คือค่าเบี่ยงเบนมาตรฐานของตัวแปร  $x$  และ  $y$   
 $\bar{x}, \bar{y}$  คือค่าเฉลี่ยของตัวแปร  $x$  และ  $y$

### 2.4.3 ขั้นตอนวิธีแบบละโมบ

ขั้นตอนวิธีแบบละโมบ (Greedy algorithm) เป็นขั้นตอนวิธีการแก้ปัญหาแบบฮิวริสติก (Heuristic) เพื่อสร้างทางเลือกที่คุ้มค่าที่สุดโดยพิจารณาจากข้อมูลที่มีอยู่ในขณะนั้น ขั้นตอนวิธีจะหาทางเลือกที่ดีที่สุดในขณะนั้นซึ่งถ้าข้อมูลนั้นพอเพียงที่จะทำให้สรุปคำตอบที่ดีที่สุด เราจะได้ขั้นตอนวิธีที่มีประสิทธิภาพ โดยทั่วไปเราจะใช้ขั้นตอนวิธีแบบละโมบกับปัญหาเหมาะสมที่สุดเพราะต้องการการตัดสินใจว่าทางเลือกในปัจจุบันมีคำตอบแทนมากที่สุดหรือน้อยที่สุดหรือไม่ อย่างไรก็ตามขั้นตอนวิธีแบบละโมบนี้ไม่ได้ให้คำตอบที่ดีที่สุดออกมา แต่จะหาคำตอบโดยใช้ข้อมูลที่น้อยที่สุดซึ่งสามารถอธิบายให้เข้าใจได้อย่างง่ายโดยอาศัยการทอนหรือหยาบ ซึ่งปกตินั้นคนทั่วไปมักทอนหรือหยาบให้ครบตามจำนวนที่ต้องการเท่านั้น แต่ถ้าหากใช้ขั้นตอนวิธีแบบละโมบก็จะเป็นการทอนหรือหยาบให้ครบตามจำนวนโดยใช้หรือหยาบให้น้อยที่สุด ดังนั้นการนำขั้นตอนวิธีแบบละโมบมาประยุกต์ใช้ในการเลือกตัวแปรก็จะได้จำนวนตัวแปรต้นที่น้อยที่สุดที่สามารถให้ผลการตัดสินใจได้ใกล้เคียงกับตัวแปรตามมากที่สุดนั่นเอง [38]–[40]

## บทที่ 3

### อุปกรณ์และวิธีการ

ในบทนี้จะกล่าวถึงอุปกรณ์และวิธีการในงานวิจัยนี้ ได้แก่ ระบบเก็บภาพ ระบบประมวลผล คุณลักษณะภาพเซลล์มะเร็งเต้านม ขั้นตอนวิธีการวิเคราะห์ภาพเซลล์มะเร็งเต้านม

#### 3.1 ระบบเก็บภาพ

ภาพเซลล์มะเร็งเต้านมที่ใช้ในงานวิจัยชิ้นนี้ได้จากการบันทึกภาพด้วยกล้องถ่ายภาพดิจิทัลยี่ห้อ Nikon รุ่น DXM1200F ที่เชื่อมต่อกับกล้องจุลทรรศน์ยี่ห้อ Nikon รุ่น Eclipse 80i (Nikon Instech Co,Ltd., Japan) ดังภาพประกอบ 3-1 โดยทำการศึกษาที่กำลังขยาย 400 เท่าบันทึกเป็นภาพสีดิจิทัล RGB 24 บิต รูปแบบแฟ้มข้อมูล JPEG ขนาดภาพ 3600 × 2880 พิกเซล จำนวน 28 ภาพ โดยแบ่งเป็นภาพที่ใช้ในกระบวนการสอนระบบจำนวน 17 ภาพ และภาพที่ใช้ในการทดสอบระบบจำนวน 11 ภาพ

#### 3.2 ระบบประมวลผล

งานวิจัยชิ้นนี้พัฒนาบนระบบปฏิบัติการ Windows 8 Pro ประมวลผลด้วยเครื่องคอมพิวเตอร์ส่วนบุคคล (PC) ซึ่งมีคุณสมบัติดังรายละเอียดต่อไปนี้

- หน่วยประมวลผลกลาง Intel® Core™ i5-3570k
- หน่วยความจำ (RAM) 16 GB ชนิด DDRIII-PC1600
- หน่วยเก็บข้อมูล (Hard disk) ความจุ 500 GB

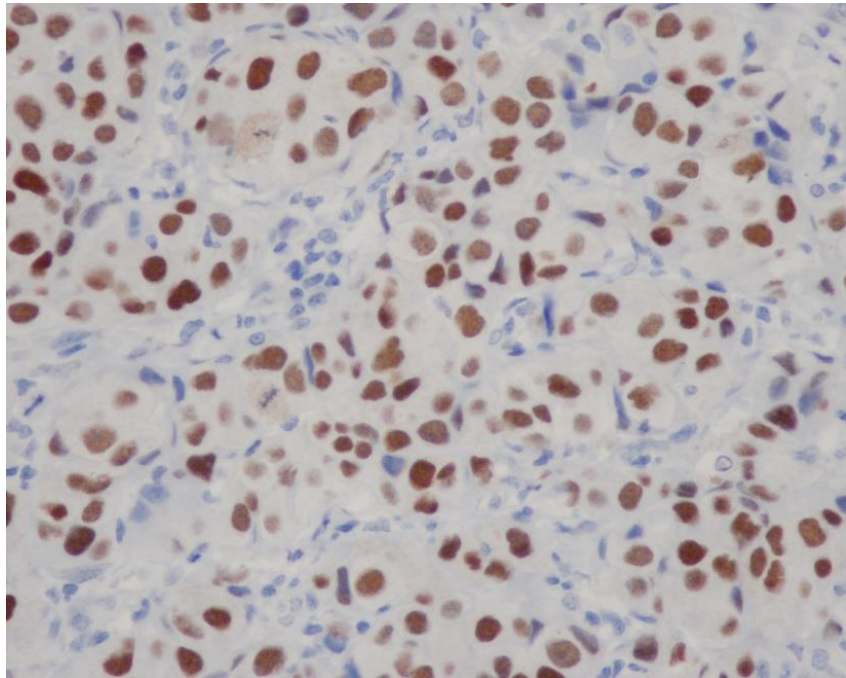
#### 3.3 คุณลักษณะของภาพถ่ายเซลล์มะเร็งเต้านม

ภาพประกอบ 3-2 แสดงตัวอย่างภาพเซลล์มะเร็งเต้านมที่ผ่านการย้อมตัวรับฮอร์โมนเอสโตรเจนด้วยเทคนิคทาง Immunohistochemistry และถ่ายภาพด้วยกล้องดิจิทัลที่เชื่อมต่อกับกล้องจุลทรรศน์ที่กำลังขยาย 400 เท่า ตัวอย่างเนื้อเยื่อที่ผ่านการย้อมจะประกอบไปด้วยเซลล์ 2 กลุ่ม ดังแสดงในภาพประกอบ 3-3 คือ เซลล์สีน้ำตาล เป็นเซลล์มะเร็งที่มีตัวรับฮอร์โมน

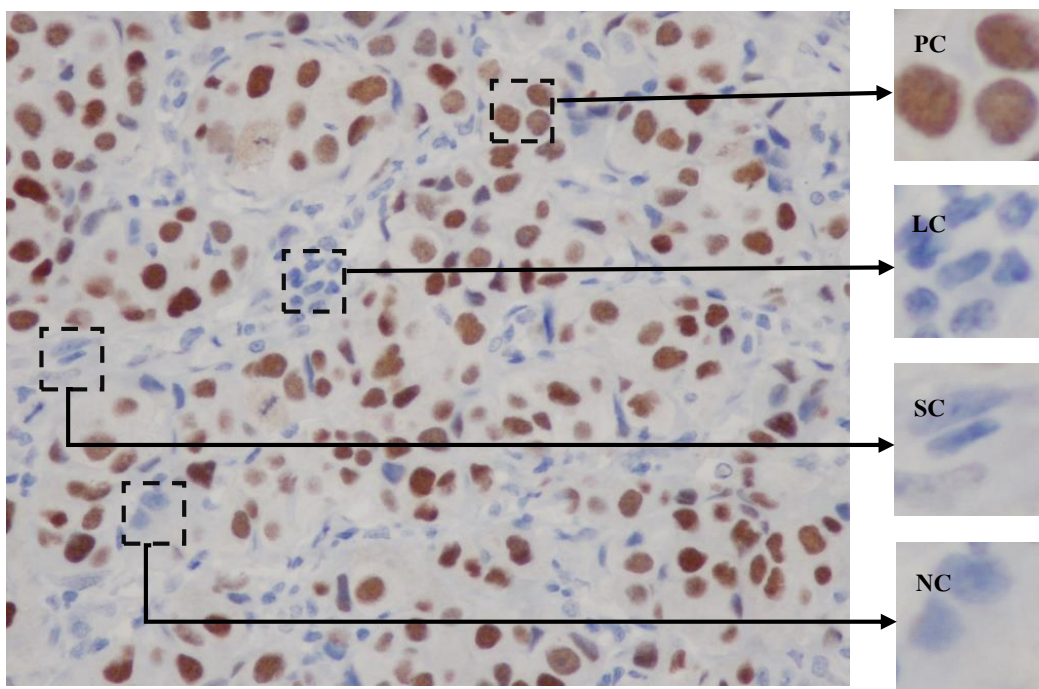
เอสโตรเจน ทางการแพทย์เรียกเซลล์ชนิดนี้ว่า เซลล์มะเร็งผลบวก (Positive cancer cell, PC) และ เซลล์สีน้ำเงินเป็นเซลล์ที่ไม่มีตัวรับฮอร์โมนเอสโตรเจน ซึ่งสามารถแบ่งออกเป็น 3 ประเภท คือ เซลล์มะเร็งที่ไม่มีตัวรับฮอร์โมนเอสโตรเจน ทางการแพทย์เรียกว่าเซลล์มะเร็งผลลบ (Negative cancer cell, NC) เซลล์เนื้อเยื่อเกี่ยวพันชนิดหนึ่ง (Stroma cell, SC) ที่คอยยึดรั้งอวัยวะให้อยู่ใน ตำแหน่งที่ถูกต้อง และเซลล์เม็ดเลือดขาว (Lymphocyte cell, LC) ที่ทำหน้าที่ในการกำจัดสิ่ง แปลกปลอมและเชื้อโรคที่เข้าสู่ภายในร่างกาย แพทย์จะใช้อัตราส่วนของจำนวนเซลล์มะเร็งผลบวก (PC) ต่อจำนวนเซลล์มะเร็งทั้งหมด (PC + NC) ที่มีในภาพเพื่อใช้ในการวางแผนการรักษาด้วยวิธี ฮอร์โมนบำบัดต่อไป



ภาพประกอบ 3-1 กล้องจุลทรรศน์ยี่ห้อ Nikon รุ่น Eclipse 80i และกล้องถ่ายภาพดิจิทัลยี่ห้อ Nikon รุ่น DXM1200F



ภาพประกอบ 3-2 ภาพถ่ายเซลล์มะเร็งเต้านมที่กำลังขยาย 400 เท่า



ภาพประกอบ 3-3 ภาพถ่ายกลุ่มเซลล์มะเร็งที่ผ่านการย้อมแยกตามชนิดของเซลล์

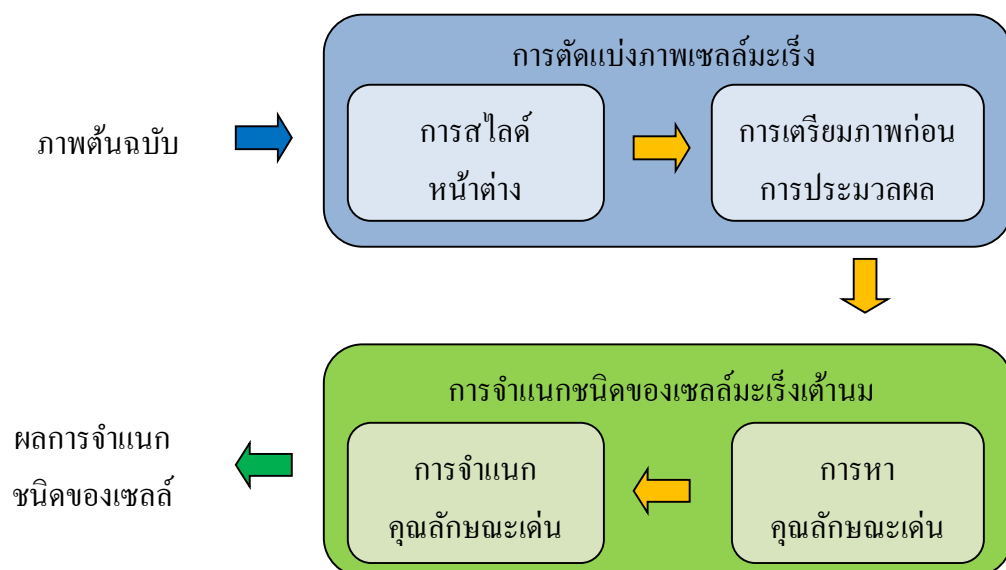
### 3.4 ขั้นตอนวิธีการวิเคราะห์เซลล์มะเร็งเต้านม

ภาพรวมของขั้นตอนวิธีการวิเคราะห์ภาพถ่ายเซลล์มะเร็งเต้านม (Algorithm for breast cancer image analysis) ประกอบด้วย 2 ขั้นตอนหลักคือ การตัดแยกเซลล์มะเร็งเต้านม และการจำแนกชนิดของเซลล์มะเร็งเต้านม แสดงผังงานดังภาพประกอบ 3-4 และมีรายละเอียดต่อไปนี้

#### 3.4.1 การตัดแบ่งภาพเซลล์มะเร็งเต้านม

การตัดแยกเซลล์มะเร็งเต้านม ประกอบด้วย 2 ขั้นตอน คือ การสไลด์หน้าต่าง (Window sliding) และการเตรียมภาพก่อนการประมวลผล (Image preprocessing) ซึ่งมีรายละเอียดดังนี้

- (1) การสไลด์หน้าต่าง ภาพถ่ายเซลล์มะเร็งเต้านมที่ใช้ในกระบวนการสอนระบบ ทั้ง 17 ภาพ จะถูกนำมาทำการตัดแบ่งด้วยหน้าต่างขนาด  $128 \times 128$  และ  $256 \times 256$  พิกเซล ภาพที่ผ่านการตัดแบ่งจะถูกนำมาจัดเป็นกลุ่มทั้งหมด 4 กลุ่ม แบ่งตามชนิดของเซลล์ที่ปรากฏบนภาพโดยไม่มีการปะปนจากเซลล์ชนิดอื่น
- (2) การเตรียมภาพก่อนการประมวลผล ในการเตรียมภาพก่อนการประมวลผล ภาพเซลล์มะเร็งที่ผ่านกระบวนการสไลด์ด้วยหน้าต่างเป็นภาพสี RGB 24 บิต จะถูกแปลงเป็นภาพระดับเทาโดยใช้สมการ (2-2) ซึ่งภาพที่ได้จากสมการดังกล่าวจะเป็นภาพระดับเทาที่มีความเข้ม 256 ระดับ



ภาพประกอบ 3-4 ขั้นตอนการวิเคราะห์ภาพถ่ายเซลล์มะเร็งเต้านม

### 3.4.2 การจำแนกชนิดของเซลล์มะเร็งเต้านม

ขั้นตอนการจำแนกชนิดเซลล์มะเร็งเต้านม ประกอบด้วย 3 ขั้นตอน คือ การหาคุณลักษณะเด่น การคัดเลือกคุณลักษณะเด่น และการจำแนกชนิดเซลล์มะเร็งเต้านม ซึ่งมีรายละเอียดดังนี้

#### (1) การหาคุณลักษณะเด่น (Feature extraction)

เมื่อทำการตัดแยกกลุ่มเซลล์มะเร็งแต่ละชนิดออกจากกันแล้ว นำภาพไปทำการคำนวณหาเมทริกซ์จีแอลซีเอ็ม ที่ระยะห่างตั้งแต่ 1 ถึง 10 พิกเซล และมุมตั้งสี่มุมคือ 0, 45, 90 และ 135 องศา โดยใช้สมการ (2-3) ซึ่งเมทริกซ์ดังกล่าวจะมีขนาดเท่ากับระดับเทาของภาพที่นำมาทำการวิเคราะห์ หลังจากนั้นเมทริกซ์จีแอลซีเอ็มจะถูกนำมาทำการคำนวณหาค่าคุณลักษณะเด่นเชิงพื้นผิวทั้งหมด 22 ลักษณะด้วยกัน ซึ่งสามารถคำนวณได้จากสมการที่แสดงในตารางที่ 2-1 และ 2-2

#### (2) การคัดเลือกคุณลักษณะเด่น (Feature selection)

เนื่องจากคุณลักษณะเด่นของเมทริกซ์จีแอลซีเอ็มมีมากถึง 22 คุณลักษณะจึงจำเป็นต้องทำการคัดเลือกคุณลักษณะเด่นที่มีความซ้ำซ้อนและไม่จำเป็นต้องตัดสินใจในการจำแนกชนิดของเซลล์ออกโดยทำการคำนวณหาค่าสหสัมพันธ์ของคุณลักษณะเด่นแต่ละคุณลักษณะร่วมกับขั้นตอนวิธีแบบละโมบเพื่อทำการเลือกค่าคุณลักษณะที่มีความจำเป็นต่อการจำแนกชนิดของเซลล์ หลังจากนั้นทำการเรียงลำดับค่าคุณลักษณะที่ถูกใช้ตามความถี่ที่ถูกเลือก โดยนำค่าคุณลักษณะเด่นที่มีความถี่มากกว่า 20 ครั้งมาวิเคราะห์ค่าเกินของข้อมูลและค่าสหสัมพันธ์คุณลักษณะเด่นที่เลือกมาจะถูกนำมาใช้ในกระบวนการจำแนกชนิดของเซลล์

#### (3) การจำแนกคุณลักษณะเด่น (Classification)

การจำแนกชนิดเซลล์มะเร็งเต้านมสามารถทำการจำแนกได้จากการนำค่าคุณลักษณะเด่นเชิงพื้นผิวของจีแอลซีเอ็มมาผ่านกระบวนการของต้นไม้ตัดสินใจ (Decision trees) ซึ่งจะได้ลำดับขั้นตอนที่จะใช้ในการแยกวิเคราะห์ชนิดของเซลล์มะเร็งเต้านม หลังจากนั้นนำกระบวนการในการวิเคราะห์ทั้งหมดมาสร้างเป็นต้นแบบของโปรแกรม เพื่อให้ผู้ใช้งานสามารถใช้โปรแกรมได้ง่ายยิ่งขึ้นและทำการประเมินประสิทธิภาพของลำดับขั้นตอนโดยใช้การตรวจสอบไขว้

### 3.5 การประเมินประสิทธิภาพ

นอกจากค่าความแม่นยำในการจำแนกชนิดของเซลล์แล้ว การประเมินประสิทธิภาพของระเบียบวิธีที่ใช้ในการจำแนกเซลล์ยังจะต้องถูกประเมินด้วยขั้นตอนอีกสองขั้นตอนด้วยกัน ขั้นตอนแรกคือการตรวจสอบไขว้ (Cross validation) ซึ่งกระบวนการนี้จะทำไป

พร้อมๆกับขั้นตอนการจำแนกคุณลักษณะเด่น โดยทำการแบ่งข้อมูลทั้งหมดที่มีออกเป็น  $N$  ชุด หลังจากนั้นนำข้อมูลจำนวน  $N-1$  ชุดมาทำการสอน แล้วนำข้อมูลที่เหลืออีกหนึ่งชุดมาทำการทดสอบ ทำซ้ำแบบนี้จำนวน  $N$  รอบ เพื่อให้ข้อมูลทุกชุดได้ผ่านทั้งกระบวนการสอนและกระบวนการทดสอบ ขั้นตอนที่สองคือการคำนวณพารามิเตอร์ที่ใช้สำหรับการประเมินประสิทธิภาพของขั้นตอนการจำแนกชนิดของเซลล์ ซึ่งมีเกณฑ์ประเมิน 4 เกณฑ์ด้วยกันคือ

- (1) ค่าความไว (Sensitivity, SS) หมายถึงค่าตัวเลขที่บ่งบอกความน่าจะเป็นของเซลล์มะเร็งที่เต้านมถูกนับโดยผู้เชี่ยวชาญจะถูกนับโดยระบบคอมพิวเตอร์ โดยสามารถคำนวณได้จากสมการ (3-1)
- (2) ค่าพยากรณ์ผลบวก (Positive predictive value, PPV) หมายถึงค่าตัวเลขที่บ่งบอกความน่าจะเป็นของเซลล์มะเร็งเต้านมที่ถูกนับด้วยระบบคอมพิวเตอร์จะถูกนับโดยผู้เชี่ยวชาญ โดยสามารถคำนวณได้จากสมการ (3-2)
- (3) ค่าความจำเพาะ (Specificity, SC) หมายถึง ค่าตัวเลขที่บ่งบอกความน่าจะเป็นของเซลล์เม็ดเลือดขาวและเซลล์สโตรมาที่ถูกนับโดยผู้เชี่ยวชาญจะถูกนับด้วยระบบคอมพิวเตอร์ โดยสามารถคำนวณได้จากสมการ (3-3)
- (4) ค่าพยากรณ์ผลลบ (Negative predictive value, NPV) หมายถึง ค่าตัวเลขที่บ่งบอกความน่าจะเป็นของเซลล์เม็ดเลือดขาวและเซลล์สโตรมาที่ถูกนับด้วยระบบคอมพิวเตอร์จะถูกนับโดยผู้เชี่ยวชาญ โดยสามารถคำนวณได้จากสมการ (3-4)

$$SS = \frac{TP}{TP + FN} \times 100 \quad (3-1)$$

$$PPV = \frac{TP}{TP + FP} \times 100 \quad (3-2)$$

$$SC = \frac{TN}{TN + FP} \times 100 \quad (3-3)$$

$$NPV = \frac{TN}{TN + FN} \times 100 \quad (3-4)$$

โดยที่  $TP$  (True positive) หรือ ผลบวกจริง หมายถึง จำนวนเซลล์ที่ขั้นตอนวิธีจำแนกประเภทเป็นเซลล์มะเร็งเต้านม ตรงกับที่ผู้เชี่ยวชาญระบุว่าป็นเซลล์มะเร็งเต้านม

*TN* (True negative) หรือ ผลลบจริง หมายถึง จำนวนเซลล์ที่ขึ้นตอนวิธีจำแนกประเภทเป็นเซลล์เม็ดเลือดขาวหรือเซลล์สโตรมา ตรงกับที่ผู้เชี่ยวชาญระบุว่าเป็นเซลล์เม็ดเลือดขาวหรือเซลล์สโตรมา

*FP* (False positive) หรือ ผลบวกเท็จ หมายถึง จำนวนเซลล์ที่ขึ้นตอนวิธีจำแนกประเภทเป็นเซลล์มะเร็งเต้านม แต่ผู้เชี่ยวชาญระบุว่าเป็นเซลล์เม็ดเลือดขาวหรือเซลล์สโตรมา

*FN* (False negative) หรือ ผลลบเท็จ หมายถึง จำนวนเซลล์ที่ขึ้นตอนวิธีจำแนกประเภทเป็นเซลล์เม็ดเลือดขาวหรือเซลล์สโตรมา แต่ผู้เชี่ยวชาญระบุว่าเป็นเซลล์มะเร็งเต้านม

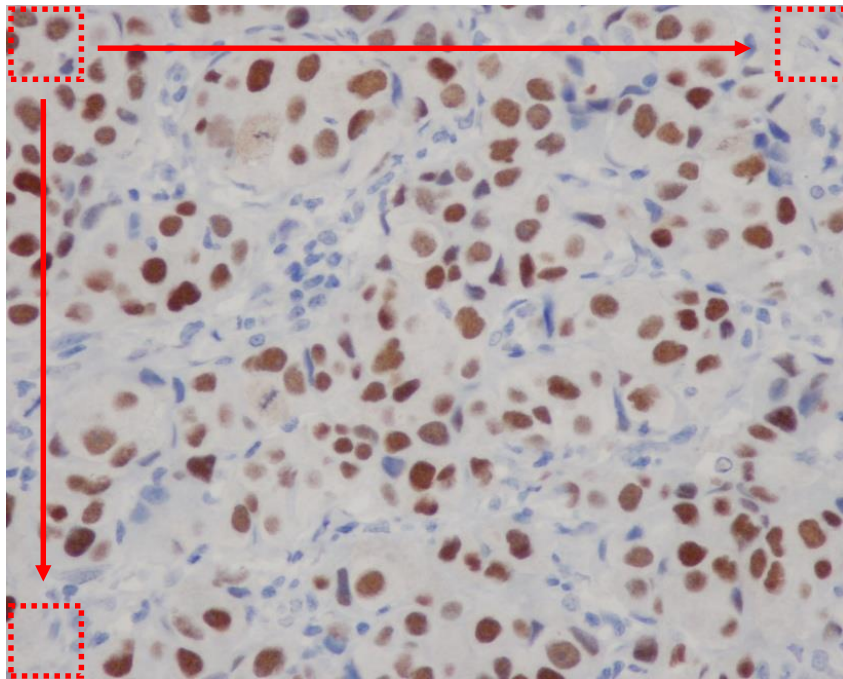


## บทที่ 4

### ผลการวิจัย

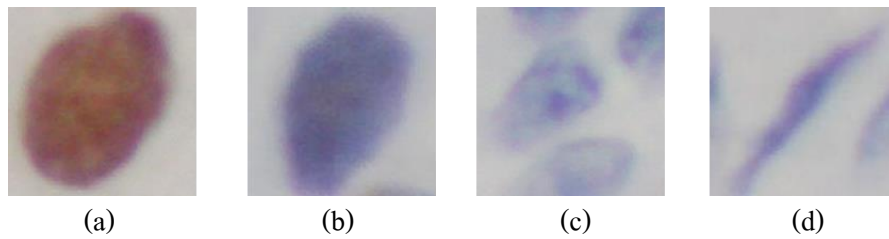
ในบทนี้จะกล่าวถึงผลการวิเคราะห์ภาพเซลล์มะเร็งเต้านมที่ผ่านการตัดแยกด้วยหน้าตัดที่มีขนาดแตกต่างกัน และผลการจำแนกชนิดของเซลล์มะเร็งเต้านม ดังรายละเอียดต่อไปนี้

#### 4.1 ผลการตัดแบ่งภาพเซลล์มะเร็งเต้านม

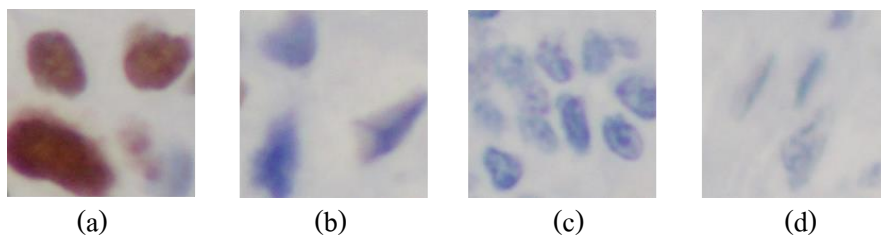


ภาพประกอบ 4-1 ลักษณะทิศทางการสไลด์วินโดว์บนภาพถ่ายต้นฉบับ

ภาพประกอบ 4-1 แสดงทิศทางการสไลด์ภาพถ่ายเซลล์มะเร็งเต้านมต้นฉบับทั้ง 17 ภาพด้วยหน้าตัดทั้งสองขนาดจากซ้ายไปขวาและบนลงล่าง โดยไม่มีการซ้อนทับของหน้าตัด ในขณะที่ทำการสไลด์ ซึ่งจำนวนภาพที่ได้จากการสไลด์ด้วยหน้าตัดทั้งสองขนาดจะมีจำนวนที่แตกต่างกัน



ภาพประกอบ 4-2 (a) ภาพเซลล์มะเร็งเต้านมผลบวก (b) ภาพเซลล์มะเร็งเต้านมผลลบ (c) ภาพเซลล์เม็ดเลือดขาว (d) ภาพเซลล์สโตรมา ที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล

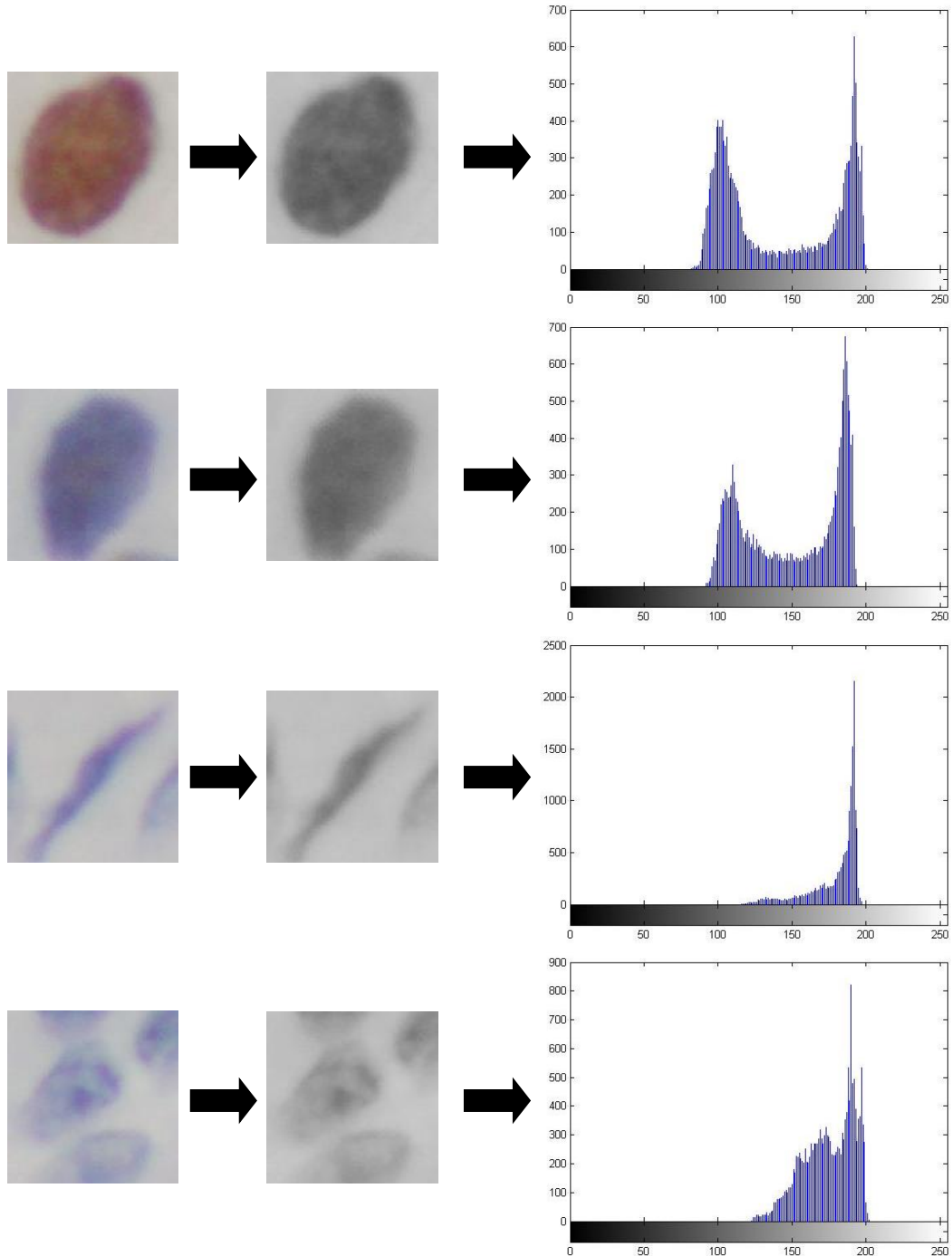


ภาพประกอบ 4-3 (a) ภาพเซลล์มะเร็งเต้านมผลบวก (b) ภาพเซลล์มะเร็งเต้านมผลลบ (c) ภาพเซลล์เม็ดเลือดขาว (d) ภาพเซลล์สโตรมา ที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $256 \times 256$  พิกเซล

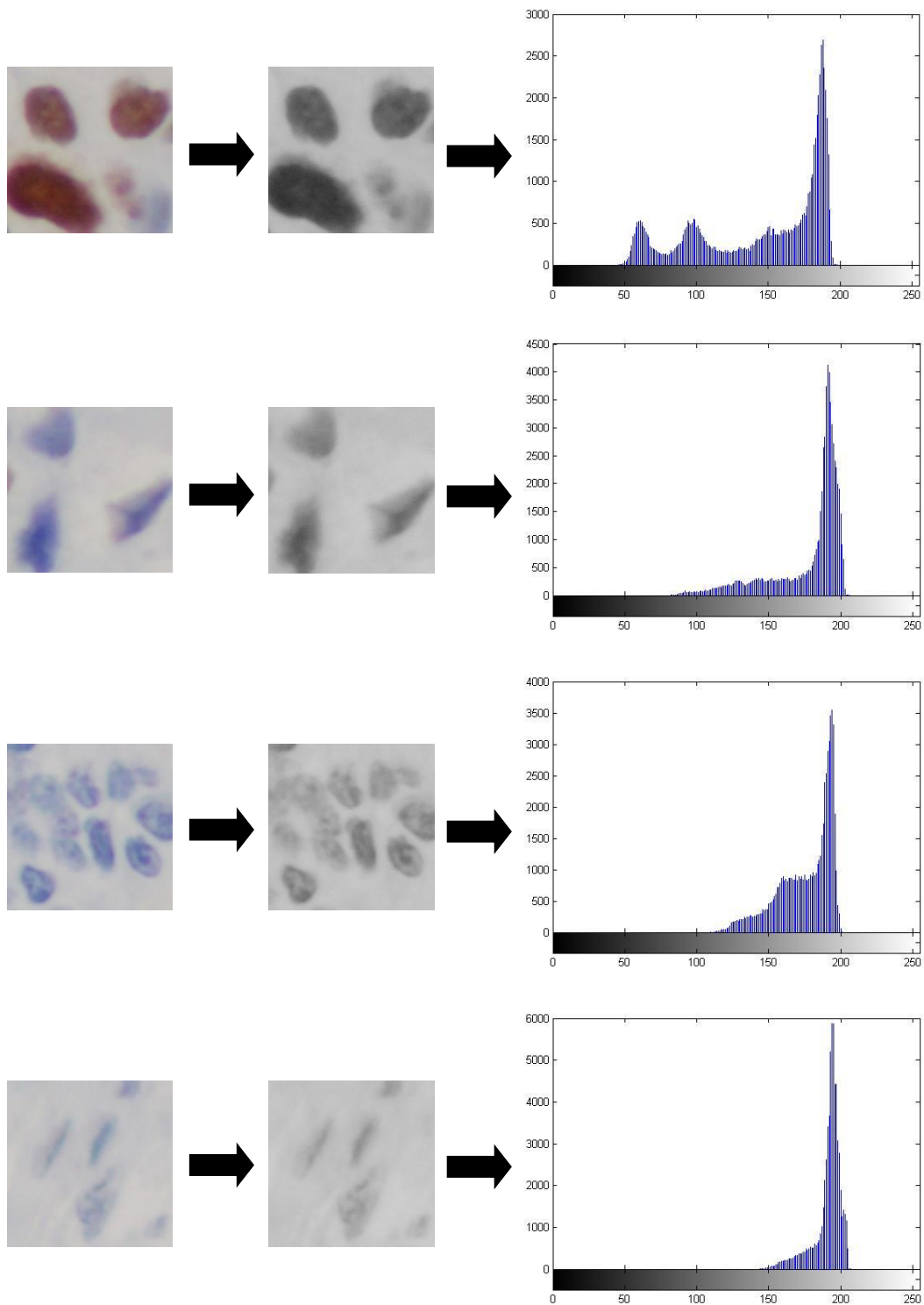
ภาพประกอบ 4-2 แสดงภาพถ่ายเซลล์ชนิดต่างๆที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล โดยถ้าทำการสไลด์ภาพต้นฉบับหนึ่งภาพจะได้ภาพย่อยที่มีขนาด  $128 \times 128$  พิกเซล ออกมาจำนวน 616 ภาพ

ภาพประกอบ 4-3 แสดงภาพถ่ายเซลล์ชนิดต่างๆที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $256 \times 256$  พิกเซล โดยถ้าทำการสไลด์ภาพต้นฉบับหนึ่งภาพจะได้ภาพย่อยที่มีขนาด  $256 \times 256$  พิกเซลออกมาจำนวน 154 ภาพ เมื่อเปรียบเทียบสัดส่วนจะพบว่าจำนวนภาพที่ได้จากหน้าต่างขนาด  $128 \times 128$  พิกเซลจะเป็นสี่เท่าของจำนวนภาพที่ได้จากหน้าต่างขนาด  $256 \times 256$  พิกเซล เนื่องจากจำนวนพิกเซลที่ประกอบขึ้นเป็นภาพย่อยนั่นเอง หลังจากนั้นทำการเลือกภาพจำนวน 2400 ภาพ ซึ่งประกอบไปด้วยภาพที่สไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซลจำนวน 1200 ภาพ และภาพที่สไลด์ด้วยหน้าต่างขนาด  $256 \times 256$  พิกเซลจำนวน 1200 ภาพเช่นกันเพื่อนำมาใช้เป็นชุดข้อมูลสำหรับทำการทดลอง โดยในชุดข้อมูลทั้งสองชุดจะประกอบไปด้วยภาพเซลล์มะเร็งเต้านมผลบวกจำนวน 600 ภาพ ภาพเซลล์มะเร็งเต้านมผลลบจำนวน 200 ภาพ ภาพเซลล์เม็ดเลือดขาวจำนวน 200 ภาพ และภาพเซลล์สโตรมาจำนวน 200 ภาพ ซึ่งในแต่ละภาพนั้นจะไม่มี การปะปนของเซลล์ชนิดอื่นๆ เลยโดยอาศัยผลเฉลยจากผู้เชี่ยวชาญ จากการศึกษาพบว่าเซลล์แต่ละ

ชนิดจะมีรูปแบบเฉพาะในกระจายตัวของระดับสีบนพริมิทีฟของเซลล์นั้นๆ รวมทั้งพื้นที่บริเวณใกล้เคียง



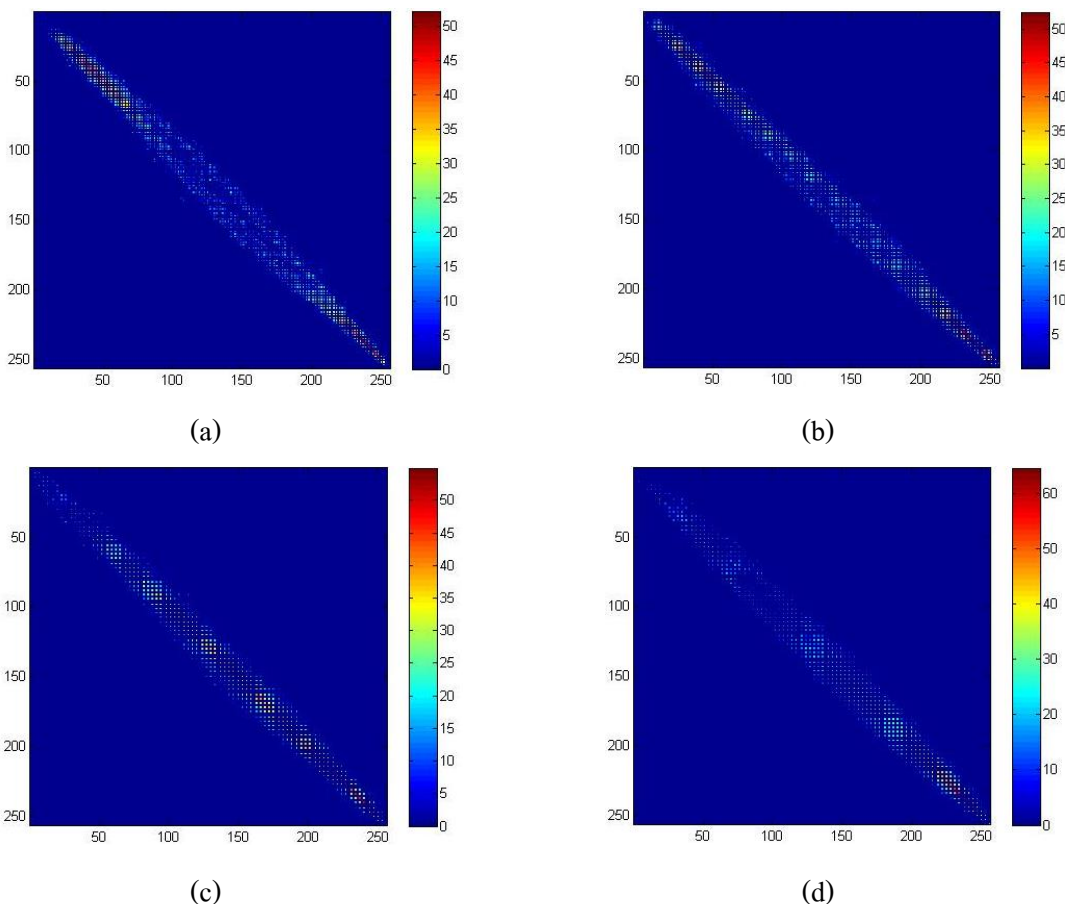
ภาพประกอบ 4-4 ภาพเซลล์มะเร็งเต้านมผลบวก เซลล์มะเร็งเต้านมผลลบ เซลล์เม็ดเลือดขาว และ เซลล์สโตรมาขนาด  $128 \times 128$  พิกเซลที่ผ่านการแปลงปริภูมิสีจากภาพสี RGB เป็นภาพระดับเทา และฮิสโตแกรมของภาพเรียงจากบนลงล่าง



ภาพประกอบ 4-5 ภาพเซลล์มะเร็งเต้านมผลบวก เซลล์มะเร็งเต้านมผลลบ เซลล์เม็ดเลือดขาว และ เซลล์สโตรมาขนาด  $256 \times 256$  พิกเซลที่ผ่านการแปลงปริภูมิสีจากภาพสี RGB เป็นภาพระดับเทา และฮิสโตแกรมของภาพเรียงจากบนลงล่าง

ภาพประกอบ 4-4 และ 4-5 แสดงภาพเซลล์แต่ละชนิดทั้งก่อนและหลังทำการแปลงปริภูมิสีของภาพจากภาพสี RGB มาเป็นภาพระดับเทาที่มีค่าความเข้ม 256 ระดับ ซึ่งการแปลงปริภูมิสีทำให้ข้อมูลของภาพถูกลดมิติลงจากเดิมอยู่ในรูปของข้อมูลสามมิติอันเนื่องมาจากมีองค์ประกอบของสีทั้งสามองค์ประกอบรวมกัน แต่เมื่อทำการแปลงปริภูมิให้กลายเป็นภาพระดับเทาก็เหมือนการยุบรวมองค์ประกอบทั้งสามเข้าด้วยกันให้เหลือเพียงองค์ประกอบเดียวแต่ก็ยังคงรายละเอียดเดิมของภาพเอาไว้ ซึ่งสามารถศึกษาองค์ประกอบความเข้มของภาพได้ง่ายยิ่งขึ้นโดยอาศัยฮิสโทแกรม จากการเปรียบเทียบพบว่าเซลล์ทั้งสี่ชนิดมีลักษณะรูปแบบของฮิสโทแกรมที่แตกต่างกันอย่างเด่นชัดไม่ว่าจะเป็นภาพที่สไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล หรือ  $256 \times 256$  พิกเซล

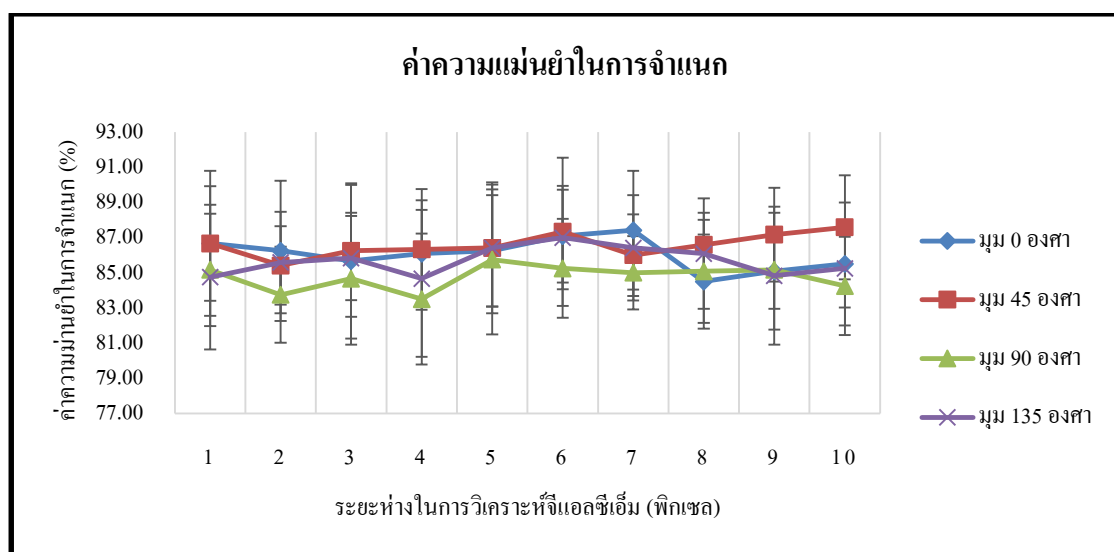
#### 4.2 ผลการจำแนกชนิดของเซลล์



ภาพประกอบ 4-6 จีแอลซีเอ็มที่ระยะห่าง 1 พิกเซล มุม 0 องศาของ (a) ภาพเซลล์มะเร็งเต้านมผลบวก (b) ภาพเซลล์มะเร็งเต้านมผลลบ (c) ภาพเซลล์เม็ดเลือดขาว (d) ภาพเซลล์สโตรมา ที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล

ภาพประกอบ 4-6 แสดงตัวอย่างจีแอลซีเอ็มที่ระยะห่าง 1 พิกเซลมุม 0 องศาของภาพถ่ายเซลล์ทั้งสี่ชนิดที่ถูกสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล โดยภาพเซลล์ที่ผ่านกระบวนการตัดแบ่งภาพมาแล้วจะถูกแปลงเป็นเมทริกซ์จีแอลซีเอ็ม ซึ่งขนาดของเมทริกซ์นี้จะสัมพันธ์กับสเกลค่าระดับเทาของภาพ ถ้าหากภาพระดับเทามีค่าความเข้ม 256 ระดับ เมทริกซ์จีแอลซีเอ็มก็จะมีขนาด 256 แถว 256 คอลัมน์ หรืออาจกล่าวได้ว่าหากภาพมีค่าความเข้ม  $L$  ระดับ เมทริกซ์จีแอลซีเอ็มของภาพดังกล่าวก็จะมีขนาด  $L \times L$  ซึ่งการสร้างเมทริกซ์จีแอลซีเอ็มนั้นจะต้องอยู่ภายใต้กรอบข้อกำหนดของระยะห่างตั้งแต่ 1 ถึง 10 พิกเซล และมุมทั้งสิ้นมุม ดังนั้นภาพหนึ่งภาพก็จะให้เมทริกซ์จีแอลซีเอ็มออกมาถึง 40 เมทริกซ์ หลังจากนั้นแต่ละเมทริกซ์จะถูกคำนวณหาค่าลักษณะเด่นเชิงพื้นผิว 22 คุณลักษณะ ซึ่งกระบวนการคำนวณค่าคุณลักษณะเด่นทั้งหมดนั้นต้องใช้เวลานานจึงจำเป็นต้องทำการวิเคราะห์คุณลักษณะเด่นเพื่อกำจัดค่าคุณลักษณะเด่นที่มีความซ้ำซ้อนและไม่มีความจำเป็นสำหรับกระบวนการจำแนกชนิดของเซลล์ออกไปโดยใช้การวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบก่อนที่จะเข้าสู่กระบวนการจำแนกชนิดของเซลล์โดยใช้ต้นไม้ตัดสินใจ

ภาพประกอบ 4-7 แสดงค่าความแม่นยำในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล จากผลการจำแนกชนิดของเซลล์พบว่าหากใช้ค่าคุณลักษณะทั้ง 22 คุณลักษณะในการจำแนกค่าความแม่นยำที่ได้จะอยู่ในช่วง 83.50 ถึง 87.58 เปอร์เซ็นต์ ค่าความแม่นยำที่ได้มีการกระจายตัวอย่างไม่มีรูปแบบที่แน่นอนที่สามารถบ่งชี้ความสัมพันธ์ของระยะห่างและมุมต่อผลการจำแนกอย่างมีนัยสำคัญ

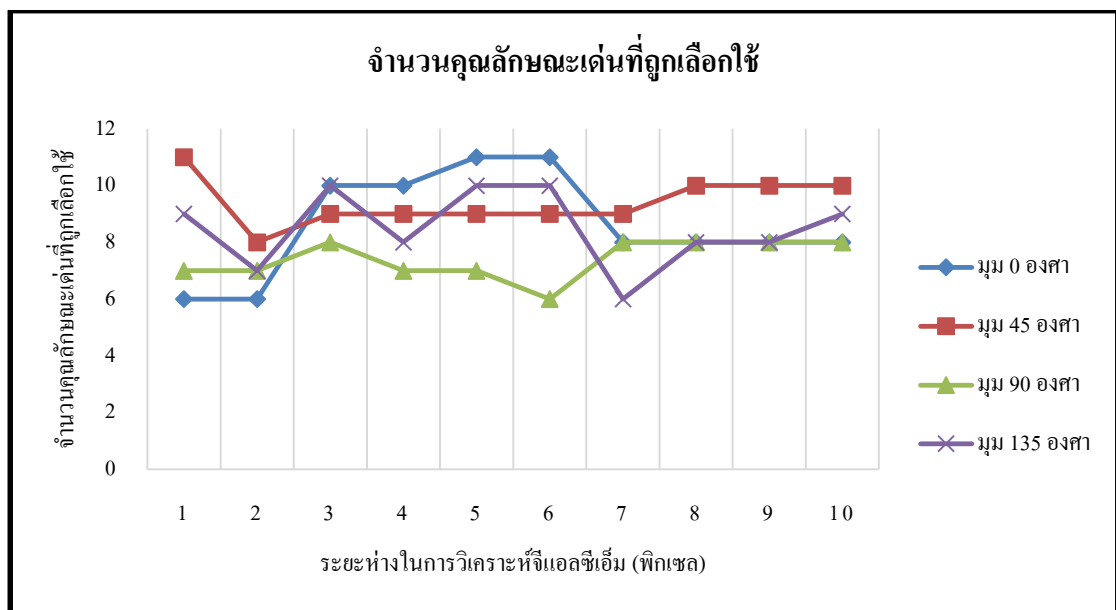


ภาพประกอบ 4-7 กราฟค่าความแม่นยำในการจำแนกชนิดของเซลล์ โดยใช้ค่าคุณลักษณะเด่นทั้ง 22 คุณลักษณะในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล

ภาพประกอบ 4-8 แสดงจำนวนคุณลักษณะเด่นที่ถูกเลือกใช้ในการจำแนกชนิดของเซลล์ โดยการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบพบว่า มีเพียง 6 ถึง 11 คุณลักษณะเท่านั้นที่จำเป็นต่อกระบวนการจำแนกชนิดของเซลล์ ซึ่งขึ้นอยู่กับกรอบข้อกำหนดของการวิเคราะห์เมทริกซ์จีแอลซีเอ็ม

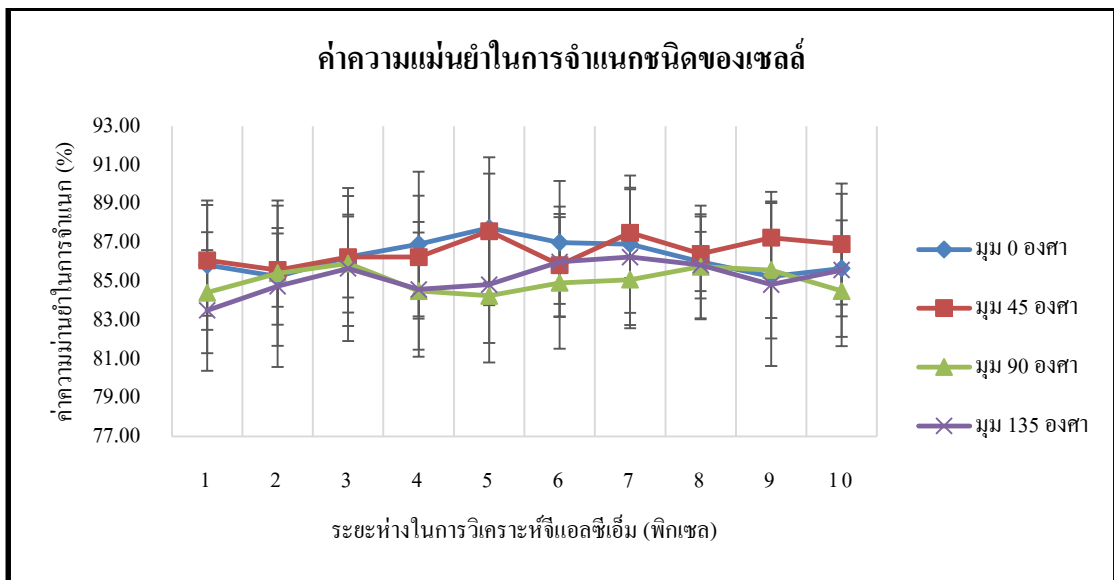
ภาพประกอบ 4-9 แสดงค่าความแม่นยำที่ได้หลังจากทำการจำแนกชนิดของเซลล์ โดยใช้เฉพาะคุณลักษณะเด่นที่ถูกเลือกโดยการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบ ค่าความแม่นยำที่ได้มีการกระจายตัวอย่างมีนัยสำคัญมากขึ้น โดยแสดงให้เห็นว่ามี 0 และ 45 องศา ให้ผลการจำแนกที่ดีกว่าอีกสองมุมที่เหลือ ส่วนระยะห่างสำหรับการวิเคราะห์นั้นยังคงอยู่ในช่วง 5 ถึง 10 พิกเซล ซึ่งจะให้ค่าความแม่นยำในการจำแนกที่สูง

ภาพประกอบ 4-10 แสดงผลต่างระหว่างค่าความแม่นยำที่ได้จากการจำแนกชนิดของเซลล์โดยใช้คุณลักษณะเด่นทั้ง 22 คุณลักษณะกับคุณลักษณะเด่นที่ถูกเลือกโดยการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบ ค่าผลต่างจะมีการเปลี่ยนแปลงอยู่ในช่วง -1.58 ถึง +1.67 เปอร์เซ็นต์ ซึ่งถือว่าผลต่างที่เกิดขึ้นมีค่าเพียงเล็กน้อยเท่านั้น

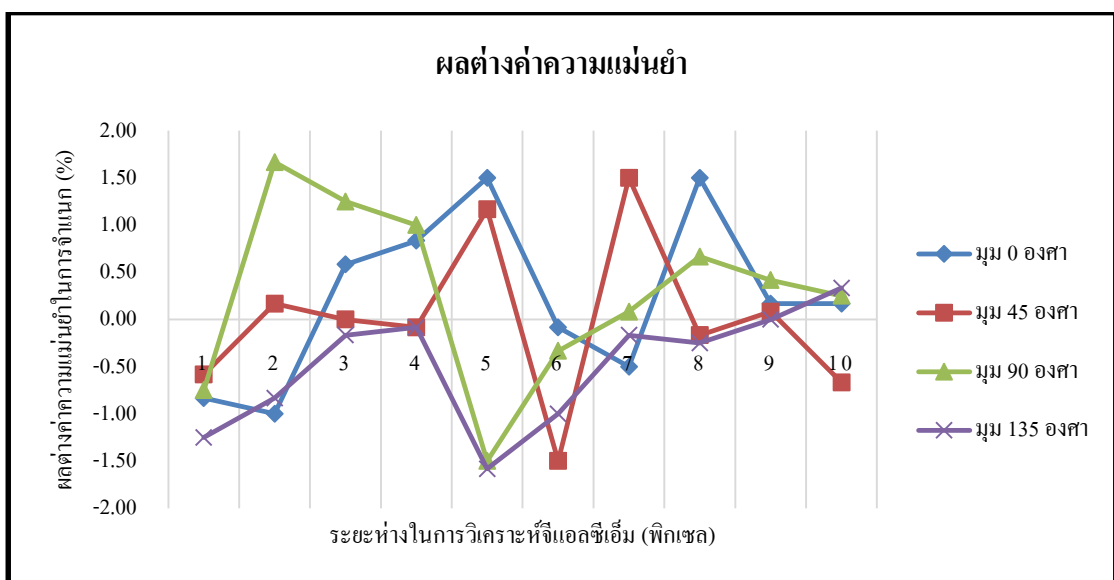


ภาพประกอบ 4-8 กราฟจำนวนของคุณลักษณะเด่นที่ถูกเลือกใช้ในการจำแนกชนิดของเซลล์ โดยการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบในภาพที่ผ่านการสไลด์ด้วยหน้าต่าง

ขนาด  $128 \times 128$  พิกเซล



ภาพประกอบ 4-9 กราฟค่าความแม่นยำในการจำแนกชนิดของเซลล์ โดยใช้ค่าคุณลักษณะเด่นที่ถูกเลือกการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล

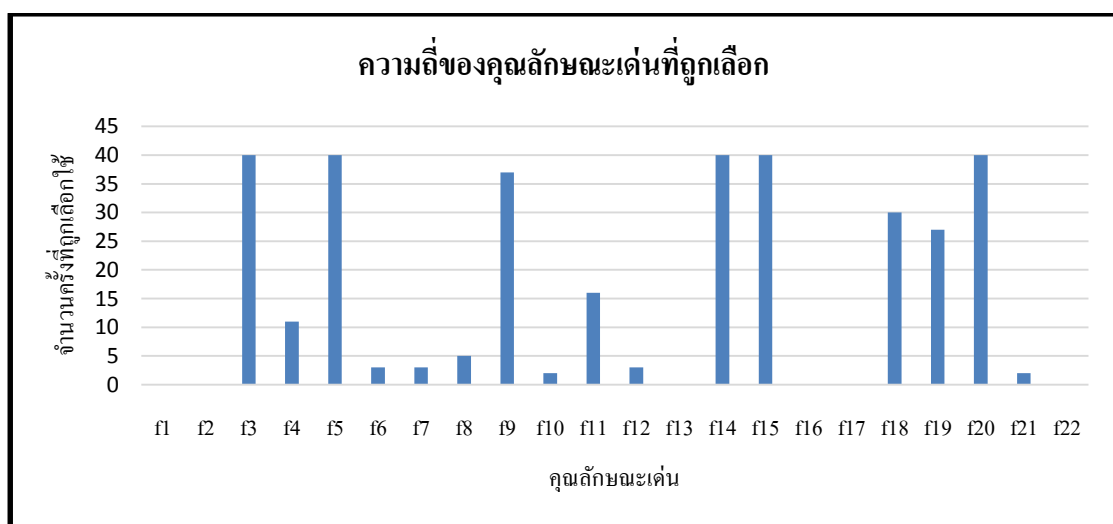


ภาพประกอบ 4-10 กราฟผลต่างค่าความแม่นยำในการจำแนกชนิดของเซลล์ ระหว่างการใช้คุณลักษณะเด่นที่ถูกเลือกการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบกับการใช้คุณลักษณะเด่นทั้งหมดในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล

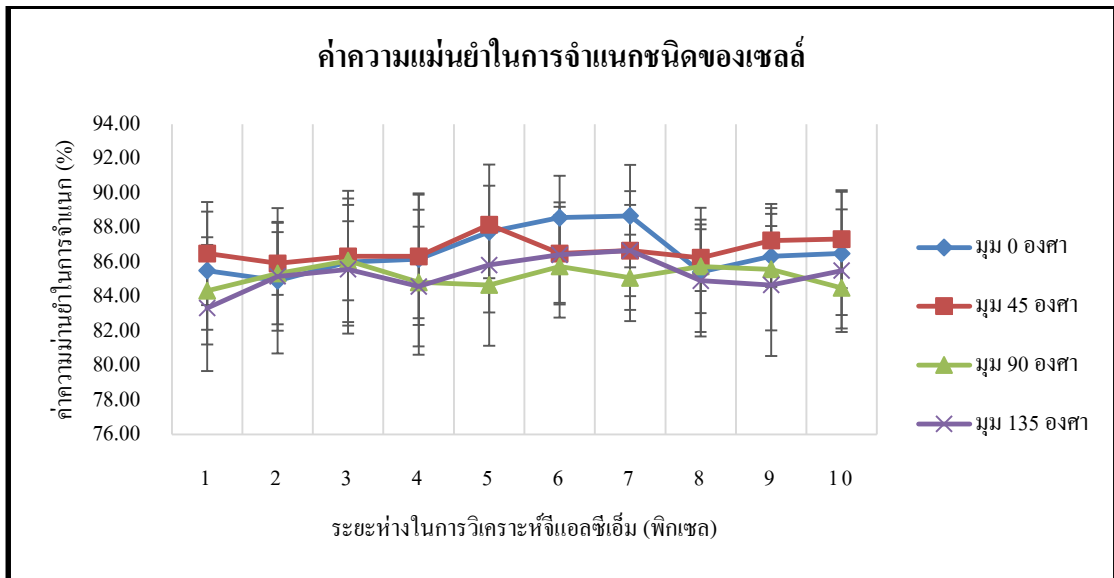


ภาพประกอบ 4-11 แสดงผลการวิเคราะห์ในเชิงความถี่ของค่าคุณลักษณะเด่นที่ถูกเลือกมาใช้โดยกระบวนการวิเคราะห์ค่าสหสัมพันธ์และขั้นตอนวิธีละโมบ พบว่าคุณลักษณะเด่นที่ถูกเลือกมาใช้มากกว่า 20 ครั้งในการวิเคราะห์จีแอลซีเอ็ม ทั้ง 40 คู่พารามิเตอร์มีทั้งหมด 8 คุณลักษณะประกอบไปด้วย ค่าสหสัมพันธ์แบบที่ 1 (f3) ค่าความโดดเด่นของคลัสเตอร์ (f5) ค่าเอนโทรปี (f9) ค่าผลรวมเฉลี่ย (f14) ค่าผลรวมเอนโทรปี (f15) ค่าผลต่างเอนโทรปี (f18) ค่าการวัดข้อมูลของสหสัมพันธ์แบบที่ 1 (f19) และค่าการวัดข้อมูลของสหสัมพันธ์แบบที่ 2 (f20) ทั้ง 8 คุณลักษณะถูกนำมาใช้เป็นชุดค่าคุณลักษณะเด่นที่ถูกเลือกด้วยการวิเคราะห์เชิงความถี่สำหรับใช้ในการวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็มสำหรับการวิเคราะห์ภาพในผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล

ภาพประกอบ 4-12 แสดงค่าความแม่นยำในการจำแนกชนิดของเซลล์โดยใช้ชุดค่าคุณลักษณะเด่นที่ถูกเลือกด้วยการวิเคราะห์เชิงความถี่สำหรับใช้ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล ซึ่งค่าความแม่นยำสูงสุดที่ได้จากการวิเคราะห์ด้วยชุดค่าคุณลักษณะเด่นที่เลือกมานี้อยู่ที่ 88.67 เปอร์เซ็นต์ ความแม่นยำในการจำแนกเซลล์มะเร็งอยู่ที่ 90.38 เปอร์เซ็นต์ ความแม่นยำในการจำแนกเซลล์เม็ดเลือดขาวอยู่ที่ 72.00 เปอร์เซ็นต์ และความแม่นยำในการจำแนกเซลล์สโตรมาอยู่ที่ 78.50 เปอร์เซ็นต์ เมื่อทำการวิเคราะห์ด้วยจีแอลซีเอ็มที่ระยะห่าง 7 พิกเซลและมุม 0 องศา ซึ่งเมื่อทำการประเมินประสิทธิภาพของระเบียบวิธีการจำแนกชนิดของเซลล์พบว่าให้ค่าความไว 93.88 เปอร์เซ็นต์ ค่าความจำเพาะ 85.75 เปอร์เซ็นต์ ค่าพยากรณ์ผลบวก 92.95 เปอร์เซ็นต์ และค่าพยากรณ์ผลลบ 87.50 เปอร์เซ็นต์



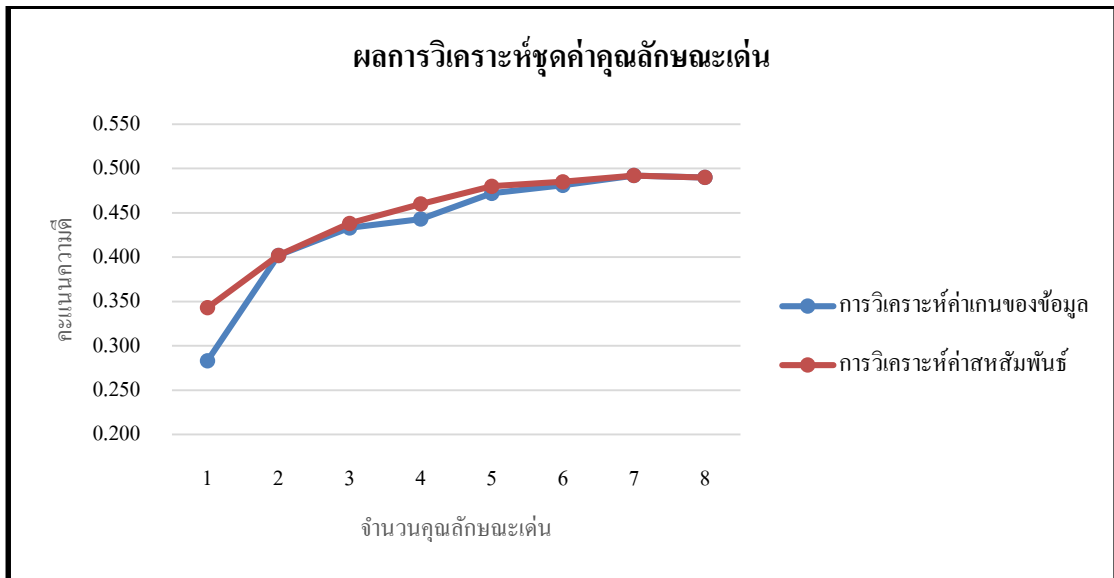
ภาพประกอบ 4-11 ความถี่ของค่าคุณลักษณะเด่นที่ถูกเลือกโดยการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบของภาพที่ถูกสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล



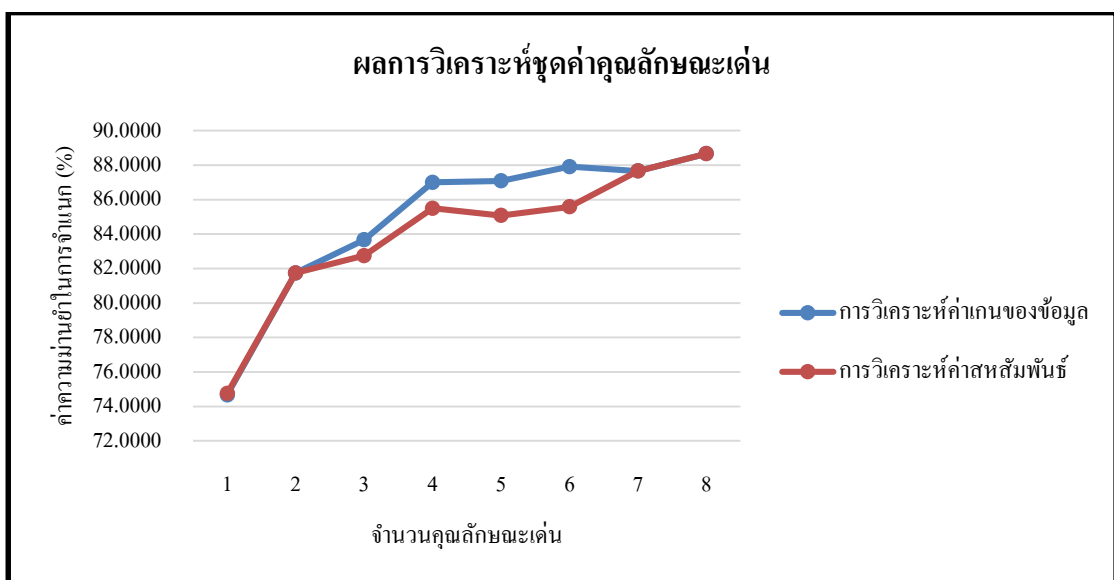
ภาพประกอบ 4-12 กราฟค่าความแม่นยำในการจำแนกชนิดของเซลล์ โดยใช้ชุดค่าคุณลักษณะเด่นที่เหมาะสมจำนวน 8 คุณลักษณะ ได้แก่ f3 f5 f9 f14 f15 f18 f19 และ f20 ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล

ภาพประกอบ 4-13 แสดงผลการวิเคราะห์คุณลักษณะเด่นทั้ง 8 คุณลักษณะที่ถูกเลือกมาด้วยการวิเคราะห์เชิงความถี่ของผลการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบ โดยใช้การวิเคราะห์ค่าสหสัมพันธ์และการวิเคราะห์ค่าเกินของข้อมูล ซึ่งทั้งสองวิธีเป็นการเรียงลำดับความสัมพันธ์ของค่าคุณลักษณะเด่นกับคลาสของการจำแนก โดยใช้คะแนนความดี (Goodness score) ในการประเมินผลการวิเคราะห์ซึ่งเป็นการวัดอัตราส่วนระหว่างค่าความแม่นยำในการจำแนกต่อจำนวนกิ่งของต้นไม้ตัดสินใจ โดยพบว่าลำดับความสำคัญของคุณลักษณะเด่นที่ได้จากการวิเคราะห์สหสัมพันธ์จะทำให้ค่าคะแนนความดีเพิ่มขึ้นเร็วกว่าลำดับที่ได้จากการวิเคราะห์ค่าเกินของข้อมูล

ภาพประกอบ 4-14 แสดงค่าความแม่นยำในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล โดยทดสอบลดจำนวนค่าคุณลักษณะเด่นที่เหมาะสมตามลำดับที่ได้จากการวิเคราะห์ค่าสหสัมพันธ์และการวิเคราะห์ค่าเกินของข้อมูล ซึ่งพบว่าถ้าทำการลดจำนวนคุณลักษณะเด่นลงอีกก็จะทำให้ค่าความแม่นยำลดลง จะมากหรือน้อยขึ้นอยู่กับลำดับการลดว่าจะลดโดยใช้ลำดับความสำคัญของการวัดค่าเกินของข้อมูล หรือตามลำดับของการวัดค่าสหสัมพันธ์

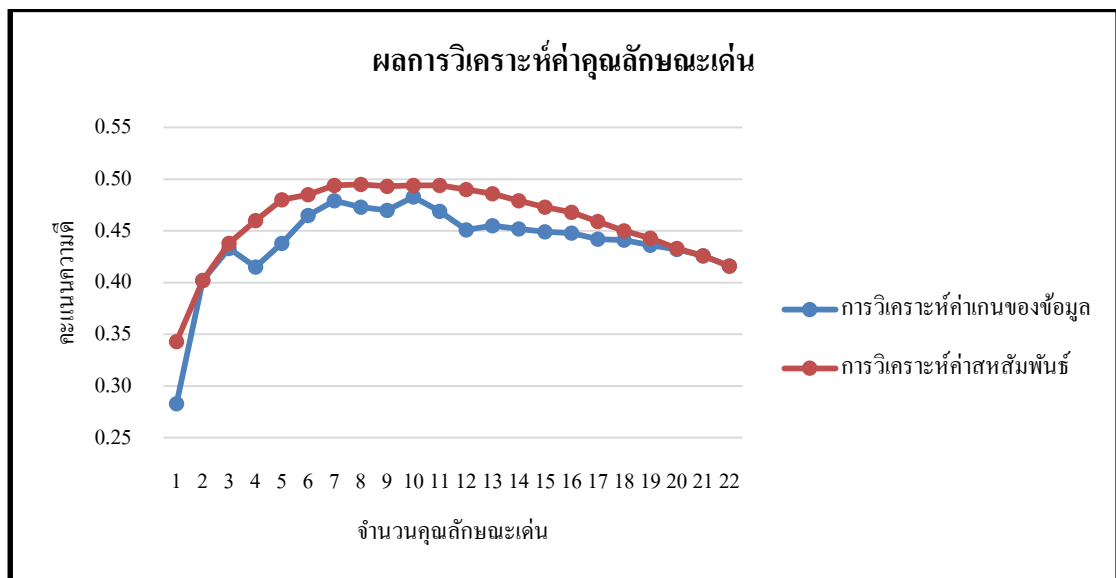


ภาพประกอบ 4-13 กราฟคะแนนความดีของชุดค่าคุณลักษณะเด่นที่เหมาะสมจำนวน 8 คุณลักษณะของการวิเคราะห์ฮีแอลซีเอ็มที่ระยะห่าง 7 พิกเซล มุม 0 องศา ได้แก่ f3 f5 f9 f14 f15 f18 f19 และ f20 เพื่อใช้ในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล



ภาพประกอบ 4-14 กราฟค่าความแม่นยำของชุดค่าคุณลักษณะเด่นที่เหมาะสมจำนวน 8 คุณลักษณะของการวิเคราะห์ฮีแอลซีเอ็มที่ระยะห่าง 7 พิกเซล มุม 0 องศา ได้แก่ f3 f5 f9 f14 f15 f18 f19 และ f20 เพื่อใช้ในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล

ภาพประกอบ 4-15 แสดงคะแนนความดีที่ได้จากการวิเคราะห์คุณลักษณะเด่นทั้ง 22 คุณลักษณะสำหรับการวิเคราะห์เชิงพื้นที่แบบจีแอลซีเอ็มที่ระยะห่าง 7 พิกเซล มุม 0 องศา เพื่อใช้ในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล โดยคะแนนความดีที่ได้จากการวิเคราะห์ค่าสหสัมพันธ์จะสูงกว่าคะแนนความดีที่ได้จากการวิเคราะห์ค่าเกณฑ์ของข้อมูล นั่นหมายความว่า การเรียงลำดับความสำคัญของค่าคุณลักษณะเด่นทั้งสองวิธีมีลำดับความสำคัญที่ไม่เหมือนกัน นอกจากนี้คะแนนความดีสูงสุดที่ได้คือ 0.495 ซึ่งได้จากการวิเคราะห์ค่าสหสัมพันธ์และสอดคล้องกับจำนวนคุณลักษณะเด่นที่ได้จากการเลือกโดยอาศัยการวิเคราะห์เชิงความถี่ แต่เมื่อทำการพิจารณาโดยละเอียดจะพบว่าคุณลักษณะเด่นทั้ง 8 คุณลักษณะที่ได้จากการวิเคราะห์ค่าสหสัมพันธ์ของคุณลักษณะเด่นทั้ง 22 คุณลักษณะกับคุณลักษณะเด่นทั้ง 8 คุณลักษณะที่ได้จากการวิเคราะห์เชิงความถี่จะไม่เหมือนกัน อันเนื่องมาจากค่าคุณลักษณะเด่นบางตัวมีความถี่ต่ำกว่า 20 ครั้งจึงไม่ถูกเลือกมาใช้ทั้งที่มีความสัมพันธ์กับคลาสมากกว่า

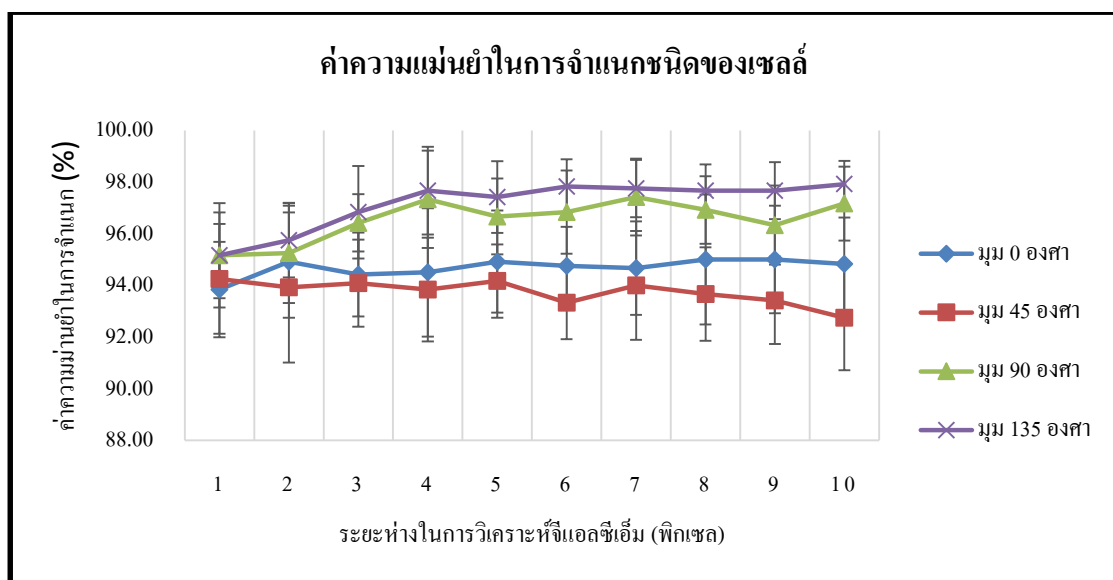


ภาพประกอบ 4-15 กราฟคะแนนความดีของค่าคุณลักษณะเด่นทั้งหมด 22 คุณลักษณะของการวิเคราะห์จีแอลซีเอ็มที่ระยะห่าง 7 พิกเซล มุม 0 องศา เพื่อใช้ในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล

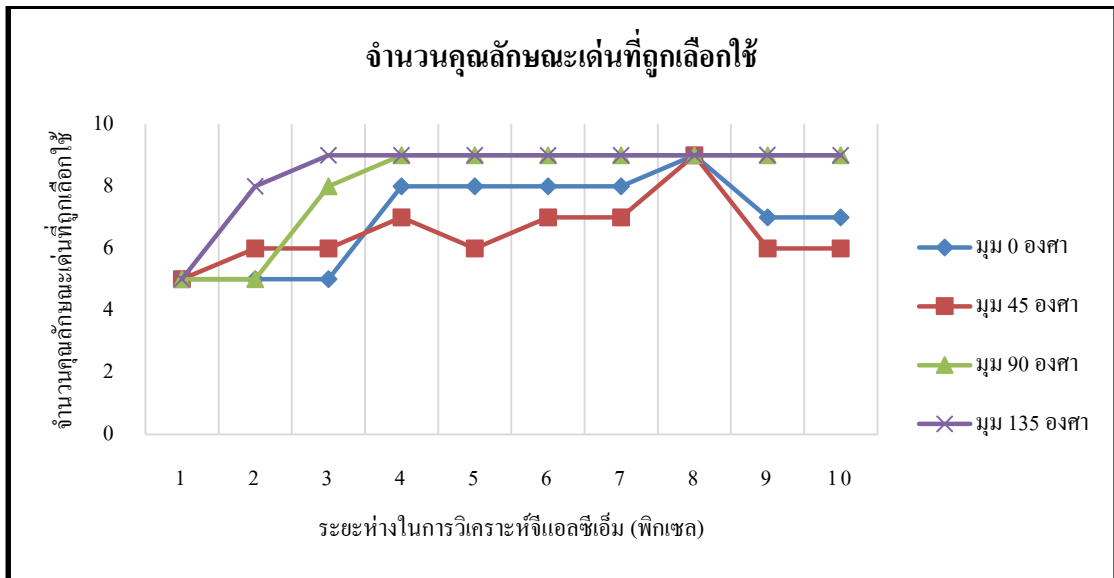
ภาพประกอบ 4-16 แสดงค่าความแม่นยำในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $256 \times 256$  พิกเซล จากผลการจำแนกชนิดของเซลล์พบว่าหากใช้ค่าคุณลักษณะทั้ง 22 คุณลักษณะในการจำแนกค่าความแม่นยำที่ได้จะอยู่ในช่วง 92.75 ถึง 97.92 เปอร์เซ็นต์ ค่าความแม่นยำที่ได้จะแสดงให้เห็นถึงความสัมพันธ์ของระยะห่างและมุมในการวิเคราะห์หัจเอดซีเอ็มต่อผลการจำแนกอย่างชัดเจนและมีนัยสำคัญ หลังจากนั้นทำจึงทำการคัดเลือกค่าคุณลักษณะเด่นโดยใช้กระบวนการของการคำนวณค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบ

ภาพประกอบ 4-17 แสดงจำนวนคุณลักษณะเด่นที่ถูกเลือกใช้ในการจำแนกชนิดของเซลล์ โดยการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบพบว่ามีเพียง 5 ถึง 9 คุณลักษณะเท่านั้นที่จำเป็นต่อกระบวนการจำแนกชนิดของเซลล์ ซึ่งขึ้นอยู่กับกรอบข้อกำหนดของการวิเคราะห์เมทริกซ์หัจเอดซีเอ็ม

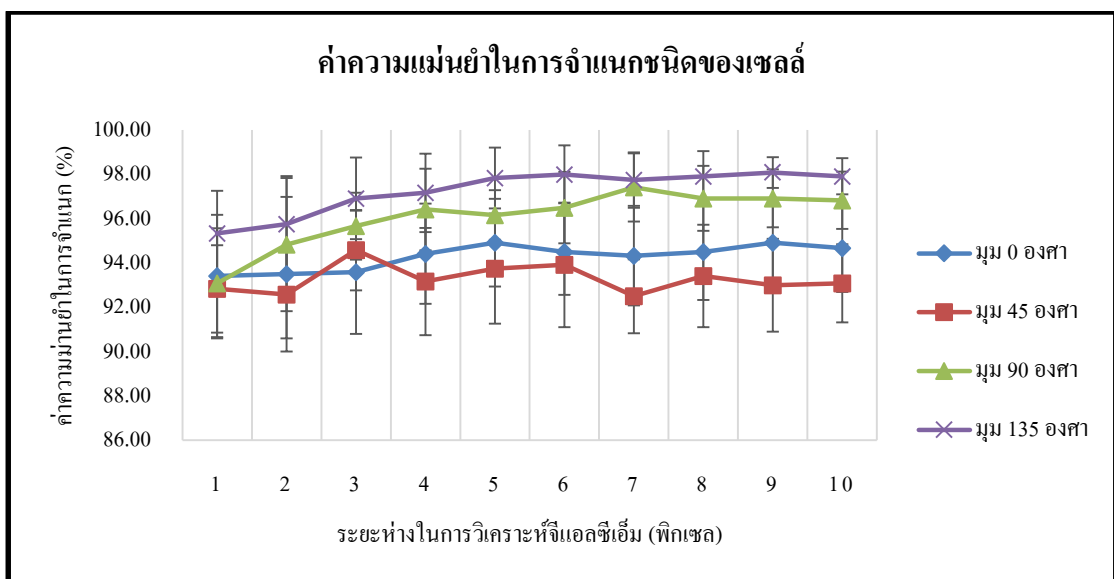
ภาพประกอบ 4-18 แสดงค่าความแม่นยำที่เปลี่ยนแปลงหลังจากทำการจำแนกชนิดของเซลล์โดยใช้เฉพาะคุณลักษณะเด่นที่ถูกเลือกโดยการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบ ค่าความแม่นยำที่ได้มีการกระจายตัวอย่างมีนัยสำคัญมากขึ้น โดยแสดงให้เห็นอย่างชัดเจนว่ามุม 135 องศา ให้ผลการจำแนกที่ดีที่สุด ส่วนระยะห่างสำหรับการวิเคราะห์นั้นยังคงอยู่ในช่วง 5 ถึง 10 พิกเซล ซึ่งจะให้ค่าความแม่นยำในการจำแนกที่สูงเช่นเดียวกับผลที่ได้จากการจำแนกภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล



ภาพประกอบ 4-16 กราฟค่าความแม่นยำในการจำแนกชนิดของเซลล์ โดยใช้ค่าคุณลักษณะเด่นทั้ง 22 คุณลักษณะในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $256 \times 256$  พิกเซล

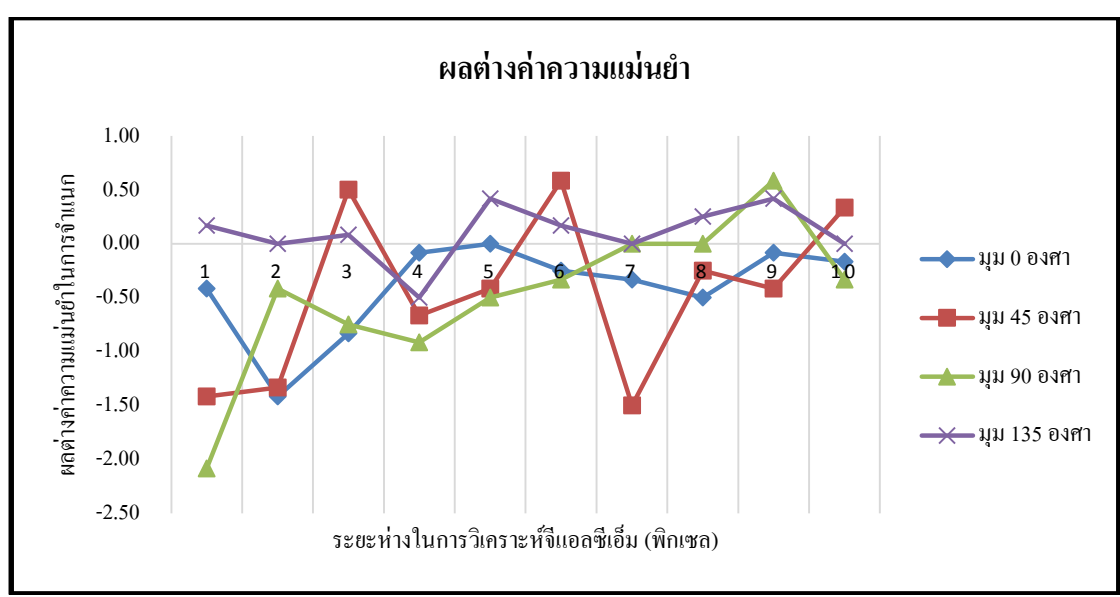


ภาพประกอบ 4-17 กราฟจำนวนของคุณลักษณะเด่นที่ถูกเลือกใช้ในการจำแนกชนิดของเซลล์ โดย  
การวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบในภาพที่ผ่านการสไลด์ด้วยหน้าต่าง  
ขนาด  $256 \times 256$  พิกเซล



ภาพประกอบ 4-18 กราฟค่าความแม่นยำในการจำแนกชนิดของเซลล์ โดยใช้ค่าคุณลักษณะเด่นที่  
ถูกเลือกการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบในภาพที่ผ่านการสไลด์ด้วย  
หน้าต่างขนาด  $256 \times 256$  พิกเซล

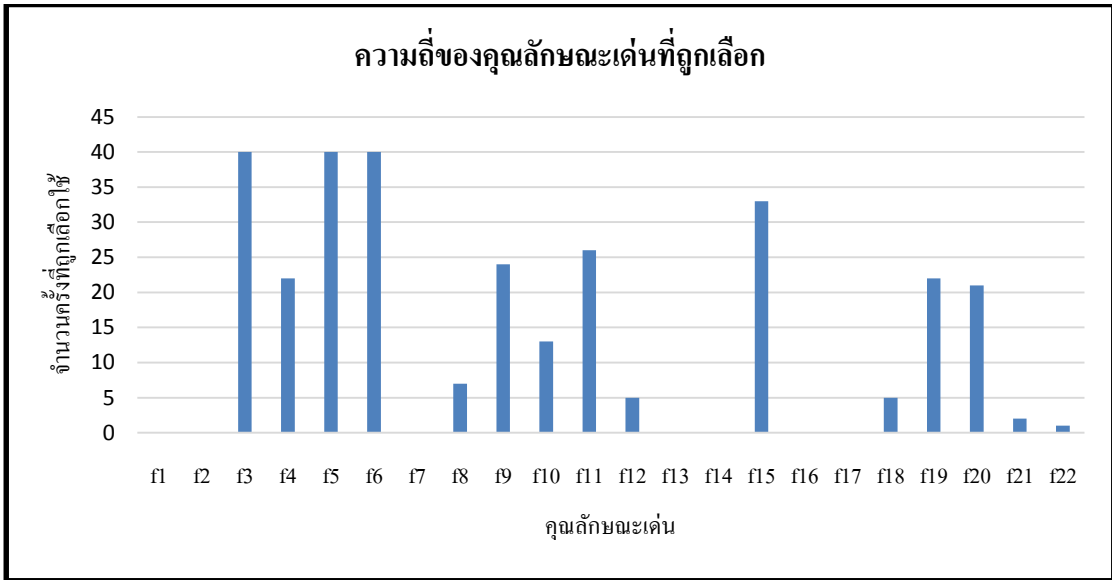
ภาพประกอบ 4-19 แสดงผลต่างระหว่างค่าความแม่นยำที่ได้จากการจำแนกชนิดของเซลล์โดยใช้คุณลักษณะเด่นทั้ง 22 ตัวกับคุณลักษณะเด่นที่ถูกเลือกโดยการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบ ค่าผลต่างจะมีการเปลี่ยนแปลงอยู่ในช่วง -2.08 ถึง +0.58 เปอร์เซนต์ ซึ่งถือว่าผลต่างที่เกิดขึ้นมีค่าเพียงเล็กน้อยเท่านั้น โดยเฉพาะในผลการวิเคราะห์ด้วยมุมมอง 135 องศา ระยะห่าง 5 ถึง 10 พิกเซล พบว่าในช่วงดังกล่าวนี้ค่าความแม่นยำในการจำแนกด้วยคุณลักษณะเด่นที่ได้จากการวิเคราะห์ค่าสหสัมพันธ์มีค่าเพียงสูงขึ้นกว่าการจำแนกโดยใช้ค่าคุณลักษณะเด่นทั้งหมด



ภาพประกอบ 4-19 กราฟผลต่างค่าความแม่นยำในการจำแนกชนิดของเซลล์ ระหว่างการใช้คุณลักษณะเด่นที่ถูกเลือกการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบกับการใช้คุณลักษณะเด่นทั้งหมดในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 x 256 พิกเซล

ภาพประกอบ 4-20 แสดงผลการวิเคราะห์ในเชิงความถี่ของค่าคุณลักษณะเด่นที่ถูกเลือกมาใช้โดยกระบวนการวิเคราะห์ค่าสหสัมพันธ์และขั้นตอนวิธีละโมบ พบว่าคุณลักษณะเด่นที่ถูกเลือกมาใช้มากกว่า 20 ครั้งในการวิเคราะห์จีแอลซีเอ็มทั้ง 40 คู่พารามิเตอร์มีทั้งหมด 9 คุณลักษณะประกอบไปด้วย ค่าสหสัมพันธ์แบบที่ 1 (f3) ค่าสหสัมพันธ์แบบที่ 2 (f4) ค่าความโดดเด่นของคลัสเตอร์ (f5) ค่าความเปลี่ยนแปลงของคลัสเตอร์ (f6) ค่าเอนโทรปี (f9) ค่าความเป็นเอกภาพแบบที่ 2 (f11) ค่าผลรวมเอนโทรปี (f15) ค่าการวัดข้อมูลของสหสัมพันธ์แบบที่ 1 (f19) และค่าการวัดข้อมูลของสหสัมพันธ์แบบที่ 2 (f20) ทั้ง 9 คุณลักษณะถูกนำมาใช้เป็นชุดค่า

คุณลักษณะเด่นที่ถูกเลือกด้วยการวิเคราะห์เชิงความถี่สำหรับการวิเคราะห์เชิงพื้นที่แบบ จีแอลซีเอ็มสำหรับการวิเคราะห์ภาพในผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล



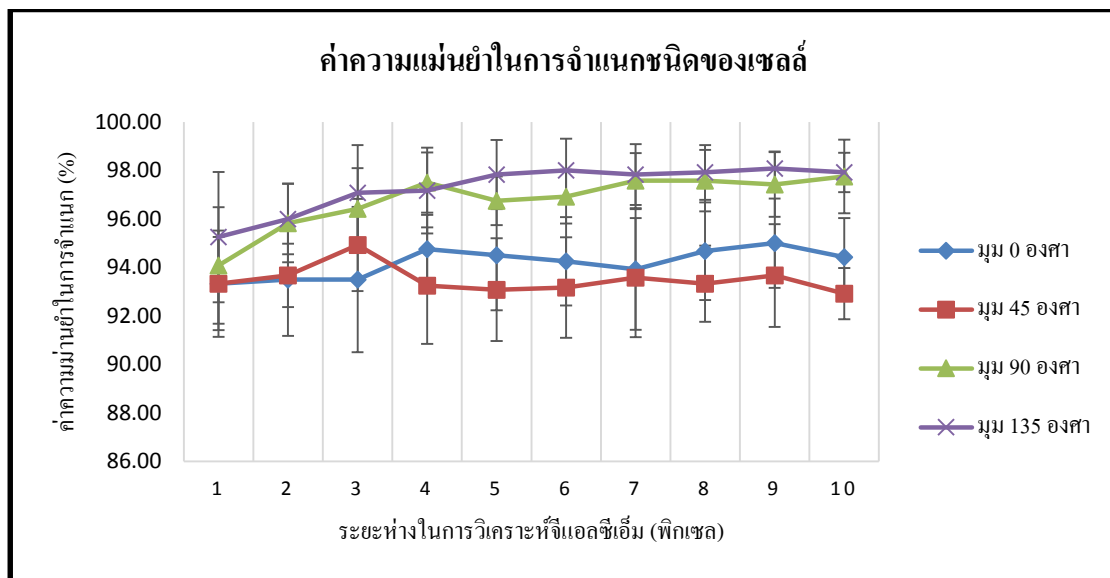
ภาพประกอบ 4-20 ความถี่ของค่าคุณลักษณะเด่นที่ถูกเลือกโดยการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับ ระเบียบขั้นตอนวิธีละโมบของภาพที่ถูกสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล

ภาพประกอบ 4-21 แสดงค่าความแม่นยำในการจำแนกชนิดของเซลล์โดยใช้ชุดค่าคุณลักษณะเด่นที่ถูกเลือกด้วยการวิเคราะห์เชิงความถี่สำหรับใช้ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล ซึ่งค่าความแม่นยำสูงสุดที่ได้จากการวิเคราะห์ด้วยชุดค่าคุณลักษณะเด่นที่เลือกมานี้อยู่ที่ 98.08 เปอร์เซ็นต์ ความแม่นยำในการจำแนกเซลล์มะเร็งอยู่ที่ 98.75 เปอร์เซ็นต์ ความแม่นยำในการจำแนกเซลล์เม็ดเลือดขาวอยู่ที่ 95.00 เปอร์เซ็นต์ และความแม่นยำในการจำแนกเซลล์สโตรมาอยู่ที่ 96.50 เปอร์เซ็นต์ เมื่อทำการวิเคราะห์ด้วยจีแอลซีเอ็มที่ระยะห่าง 9 พิกเซลและมุม 135 องศา ซึ่งเมื่อทำการประเมินประสิทธิภาพของระเบียบวิธีการจำแนกชนิดของเซลล์พบว่าให้ค่าความไว 99.25 เปอร์เซ็นต์ ค่าความจำเพาะ 97.25 เปอร์เซ็นต์ ค่าพยากรณ์ผลบวก 98.63 เปอร์เซ็นต์ และค่าพยากรณ์ผลลบ 98.48 เปอร์เซ็นต์

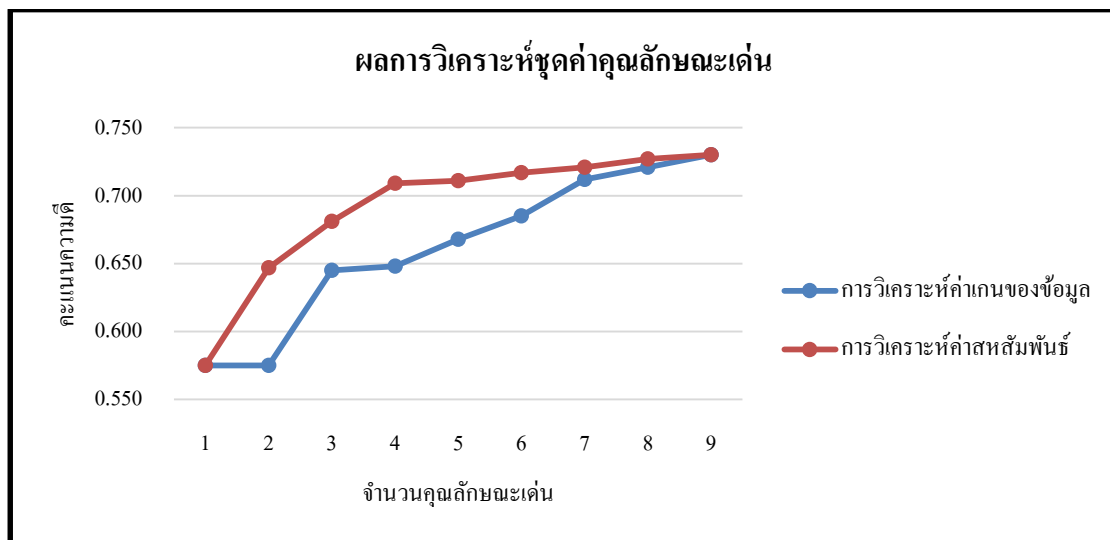
ภาพประกอบ 4-22 แสดงผลการวิเคราะห์คุณลักษณะเด่นทั้ง 9 คุณลักษณะที่ถูกเลือกมาด้วยการวิเคราะห์เชิงความถี่ของผลการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบ โดยใช้การวิเคราะห์ค่าสหสัมพันธ์และการวิเคราะห์ค่าเกณฑ์ของข้อมูล โดยใช้คะแนนความดีในการประเมินผลการวิเคราะห์ผลปรากฏว่าลำดับความสำคัญของคุณลักษณะเด่นที่ได้จากการ



วิเคราะห์สหสัมพันธ์จะทำให้ค่าคะแนนความดีเพิ่มขึ้นเร็วกว่าลำดับที่ได้จากการวิเคราะห์ค่าเกณฑ์ของข้อมูล



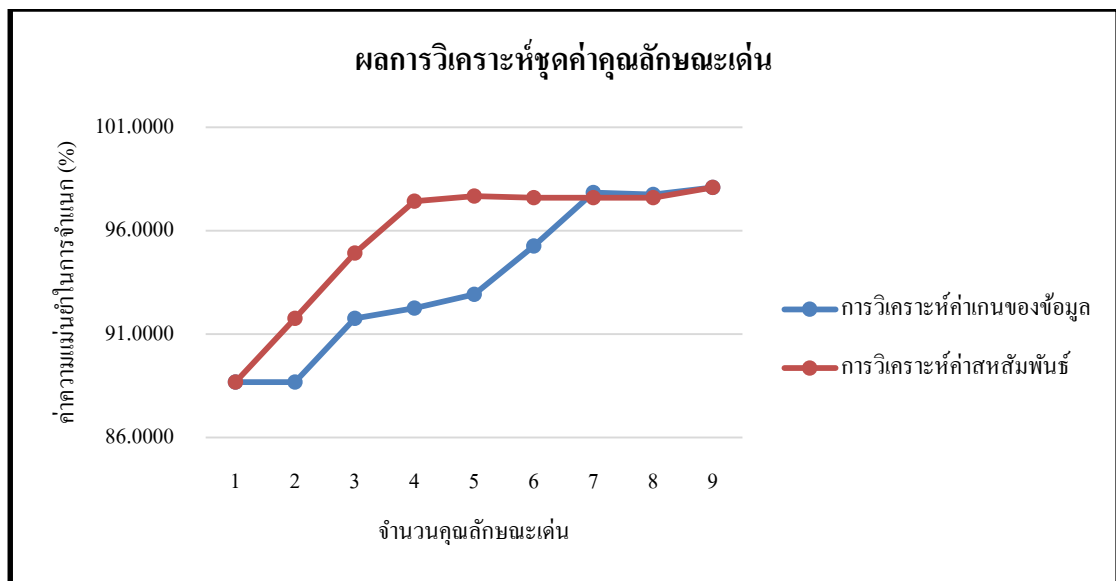
ภาพประกอบ 4-21 กราฟค่าความแม่นยำในการจำแนกชนิดของเซลล์ โดยใช้ชุดค่าคุณลักษณะเด่นที่เหมาะสมจำนวน 9 คุณลักษณะ ได้แก่ f3 f4 f5 f6 f9 f11 f15 f19 และ f20 ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล



ภาพประกอบ 4-22 กราฟคะแนนความดีของชุดค่าคุณลักษณะเด่นที่เหมาะสมจำนวน 9 คุณลักษณะของการวิเคราะห์จีแอลซีเอ็มที่ระยะห่าง 9 พิกเซล มุม 135 องศา ได้แก่ f3 f4 f5 f6 f9 f11 f15 f19 และ f20 เพื่อใช้ในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่าง

ขนาด 256 × 256 พิกเซล

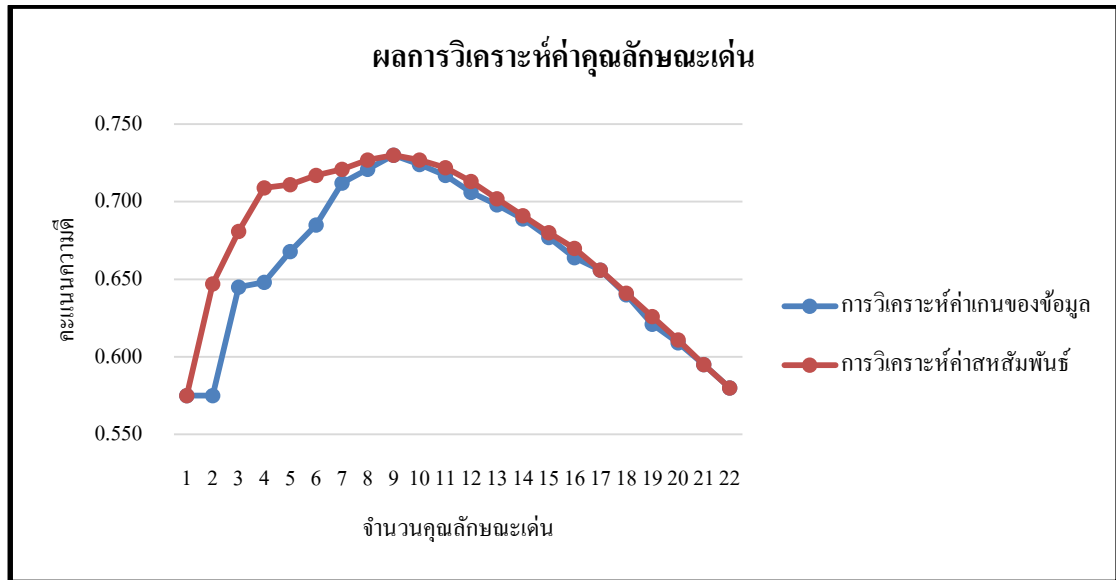
ภาพประกอบ 4-23 แสดงค่าความแม่นยำในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $256 \times 256$  พิกเซล โดยทดสอบลดจำนวนค่าคุณลักษณะเด่นที่เหมาะสมตามลำดับที่ได้จากการวิเคราะห์ค่าสหสัมพันธ์และการวิเคราะห์ค่าเกินของข้อมูล ซึ่งพบว่าถ้าทำการลดจำนวนคุณลักษณะเด่นลงอีกก็จะทำให้ค่าความแม่นยำลดลง จะมากหรือน้อยขึ้นอยู่กับลำดับการลดว่าจะลดโดยใช้ลำดับความสำคัญของการวัดค่าเกินของข้อมูล หรือตามลำดับของการวัดค่าสหสัมพันธ์



ภาพประกอบ 4-23 กราฟค่าความแม่นยำของชุดค่าคุณลักษณะเด่นที่เหมาะสมจำนวน 9 คุณลักษณะของการวิเคราะห์จีแอลซีเอ็มที่ระยะห่าง 9 พิกเซล มุม 135 องศา ได้แก่ f3 f4 f5 f6 f9 f11 f15 f19 และ f20 เพื่อใช้ในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $256 \times 256$  พิกเซล

ภาพประกอบ 4-24 แสดงคะแนนความดีที่ได้จากการวิเคราะห์คุณลักษณะเด่นทั้ง 22 คุณลักษณะสำหรับการวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็มที่ระยะห่าง 9 พิกเซล มุม 135 องศา เพื่อใช้ในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $256 \times 256$  พิกเซล โดยคะแนนความดีที่ได้จากการวิเคราะห์ค่าสหสัมพันธ์จะสูงกว่าคะแนนความดีที่ได้จากการวิเคราะห์ค่าเกินของข้อมูล นั้นหมายความว่า การเรียงลำดับความสำคัญของค่าคุณลักษณะเด่นทั้งสองวิธีมีลำดับความสำคัญที่ไม่เหมือนกัน นอกจากนั้นคะแนนความดีสูงสุดที่ได้คือ 0.730 ซึ่งได้จากการวิเคราะห์ค่าสหสัมพันธ์และสอดคล้องกับจำนวนคุณลักษณะเด่นที่ได้จากการเลือกโดยอาศัยการวิเคราะห์เชิงความดี และเมื่อทำการพิจารณาโดยละเอียดจะพบว่าคุณลักษณะเด่นทั้ง 9

คุณลักษณะที่ได้จากการวิเคราะห์ค่าสหสัมพันธ์ของคุณลักษณะเด่นทั้ง 22 คุณลักษณะกับคุณลักษณะเด่นทั้ง 9 คุณลักษณะที่ได้จากการวิเคราะห์เชิงความถี่เป็นชุดค่าคุณลักษณะเด่นชุดเดียวกัน

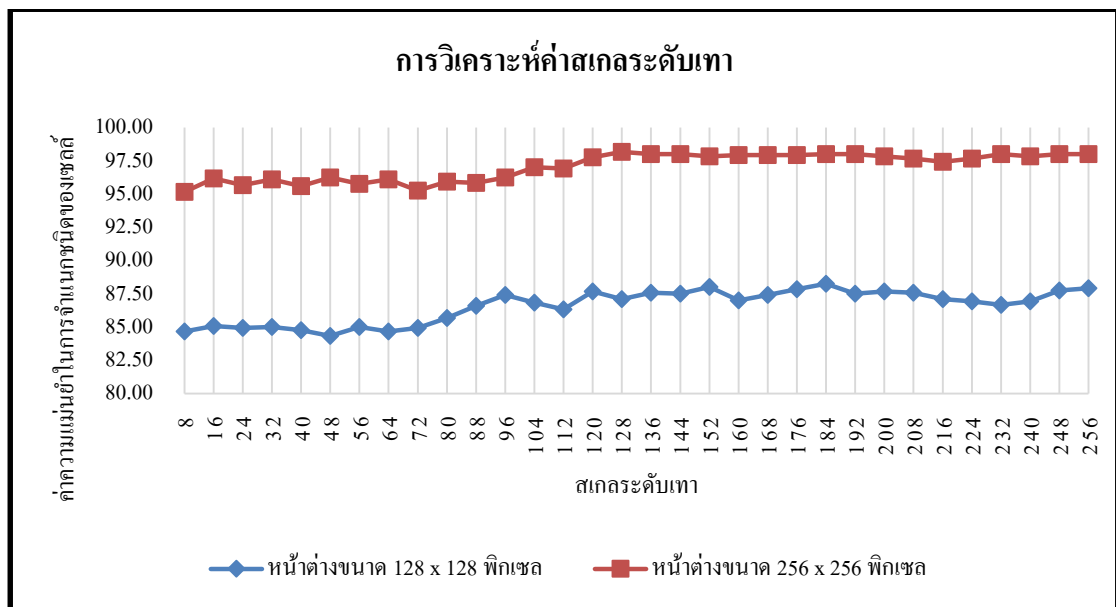


ภาพประกอบ 4-24 กราฟคะแนนความดีของค่าคุณลักษณะเด่นทั้งหมด 22 คุณลักษณะของการวิเคราะห์จีแอลซีเอ็มที่ระยะห่าง 9 พิกเซล มุม 135 องศา เพื่อใช้ในการจำแนกชนิดของเซลล์ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $256 \times 256$  พิกเซล

ภาพประกอบ 4-25 แสดงผลการวิเคราะห์ค่าความสัมพันธ์ระหว่างระดับสเกลค่าระดับเท่ากับความแม่นยำในการจำแนกภาพ ซึ่งนอกจากมุมและระยะห่างจะมีผลต่อการวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็มแล้วสเกลระดับเทาเองก็เป็นอีกองค์ประกอบหนึ่งที่มีความสำคัญต่อการวิเคราะห์เชิงพื้นผิวเป็นอย่างยิ่ง โดยผลที่ได้คือสเกลค่าระดับเทาที่เหมาะสมจะอยู่ในช่วงตั้งแต่ 128 ถึง 256 ระดับ ยิ่งสเกลมีความละเอียดก็จะสามารถจำแนกได้ดีทั้งในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล และ  $256 \times 256$  พิกเซลแต่ในบางกรณีก็ไม่จำเป็นต้องใช้สเกลที่ละเอียดมากก็สามารถให้ค่าความแม่นยำในการวิเคราะห์ที่น่าพึงพอใจได้

จากผลการวิเคราะห์ทั้งหมดเมื่อนำมาเปรียบเทียบกันจะเห็นได้ว่าการสไลด์ภาพด้วยหน้าต่างขนาด  $256 \times 256$  พิกเซลจะให้ค่าความแม่นยำในการจำแนกที่สูงกว่าภาพที่สไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล เนื่องจากภาพขนาดใหญ่จะสามารถบ่งบอกลักษณะของพื้นผิวที่แตกต่างกันได้อย่างชัดเจนมากกว่าภาพขนาดเล็ก แต่การใช้ภาพขนาดใหญ่ในการวิเคราะห์ก็จะเป็นการเพิ่มภาระให้ระบบเนื่องจากจำนวนข้อมูลที่ต้องทำการวิเคราะห์มีมากขึ้น แต่เมื่อเปรียบเทียบ

แล้วต้องอยู่ที่ดุลยพินิจของผู้เลือกใช้งานว่าในขณะนั้นต้องการเน้นเรื่องความแม่นยำหรือเรื่องข้อจำกัดของเวลา สำหรับการวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็มในภาพถ่ายเซล์มะเร็งเต้านมภายใต้กรอบข้อกำหนดของระยะห่างและมุมจะเห็นได้ว่าระยะห่างที่เหมาะสมจะอยู่ในช่วง 5 ถึง 10 พิกเซล ส่วนมุมที่เหมาะสมนั้นจะเห็นได้ว่าหน้าตาทั้งสองขนาดจะมีมุมที่เหมาะสมนั้นแตกต่างกัน สำหรับหน้าตาขนาด  $128 \times 128$  พิกเซลมุมที่เหมาะสมคือมุม 0 องศา ส่วนหน้าตาขนาด  $256 \times 256$  พิกเซลมุมที่เหมาะสมคือมุม 135 องศา ส่วนสเกลค่าระดับเทาที่เหมาะสมสำหรับการวิเคราะห์ภาพถ่ายเซลล์มะเร็งเต้านมจะอยู่ในช่วงตั้งแต่ 128 ถึง 256 ระดับ ก็สามารถให้ค่าความแม่นยำในการวิเคราะห์ที่น่าพึงพอใจได้



ภาพประกอบ 4-25 ผลการวิเคราะห์ค่าสเกลที่เหมาะสมสำหรับการวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็ม โดยใช้กรอบข้อกำหนดและชุดค่าลักษณะเด่นที่ทำให้เหมาะสมสำหรับการจำแนกชนิดของเซลล์มะเร็งเต้านม

#### 4.3 ผลการประเมินประสิทธิภาพการนับจำนวนเซลล์มะเร็งเต้านม

เมื่อนำผลของอัลกอริทึมข้างต้นมาประยุกต์ใช้ในระบบนับจำนวนนิวเคลียสเซลล์มะเร็งเต้านมแล้วนำไปทดสอบบนภาพต้นฉบับทั้งหมดจำนวน 28 ภาพ โดยเป็นภาพที่ถูกนำมาใช้ในกระบวนการสอนระบบจำนวน 17 ภาพ และเป็นภาพที่ไม่ได้ใช้ในกระบวนการสอนระบบจำนวน 11 ภาพ ซึ่งผลปรากฏว่าเมื่อนำทฤษฎีการวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็มมาเพิ่ม

เข้าไปในระบบนับจำนวนนิวเคลียสเซลล์มะเร็งเร็วขึ้นทำให้ระยะเวลาในการประมวลผลภาพเพิ่มขึ้น และสามารถลดจำนวนผลบวกเท็จของระบบได้เป็นจำนวนมาก ซึ่งในสภาพการทดสอบบนภาพต้นฉบับนั้นพบว่าผลบวกเท็จที่ได้จากการใช้หน้าต่างขนาด  $128 \times 128$  พิกเซลจะมีจำนวนน้อยกว่าการใช้หน้าต่างขนาด  $256 \times 256$  พิกเซล ส่วนผลบวกจริงมีจำนวนลดลงและผลลบเท็จมีจำนวนเพิ่มขึ้นเล็กน้อยสืบเนื่องมาจากความไม่ครอบคลุมในการเลือกชุดข้อมูลที่ใช้ในการสอนระบบ โดยเฉพาะในกลุ่มเซลล์ที่ย้อมติดสีจางในบางภาพ และในสภาพการใช้งานจริงนั้นการกางหน้าต่างจะมีการปะปนของเซลล์หลายชนิดซึ่งจะไม่สามารถควบคุมได้เหมือนกับในชุดข้อมูลที่ใช้ในการสอนระบบ ซึ่งมีการคัดเลือกภาพที่มีการปะปนของเซลล์หลายชนิดออกจากชุดข้อมูล

## บทที่ 5

### บทสรุปและข้อเสนอแนะ

ในบทนี้กล่าวถึงบทสรุปของการวิจัยและข้อเสนอแนะสำหรับผู้ต้องการนำงานวิจัยนี้ไปอ้างอิงหรือพัฒนาต่อ ดังรายละเอียดต่อไปนี้

#### 5.1 บทสรุป

งานวิจัยนี้เป็นการพัฒนาวิธีการจำแนกชนิดของเซลล์ในภาพถ่ายเซลล์มะเร็งเต้านมโดยอาศัยการประยุกต์ใช้การวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็ม ซึ่งภาพถ่ายต้นฉบับที่นำมาใช้มีทั้งหมด 17 ภาพ กระบวนการดำเนินงานประกอบด้วยสองขั้นตอนหลัก คือ การตัดแบ่งภาพเซลล์มะเร็งเต้านม โดยภาพถ่ายต้นฉบับจะถูกนำมาสไลด์ด้วยหน้าต่างสองขนาดแล้วเลือกภาพแต่ละขนาดมาเพื่อสร้างเป็นชุดข้อมูลในจำนวนที่เท่ากัน แล้วทำการแปลงปริภูมิสีเป็นภาพระดับเทาที่มีความเข้ม 256 ระดับ จากนั้นนำผลลัพธ์ที่ได้จากขั้นตอนการตัดแบ่งภาพเซลล์มะเร็งเต้านมเข้าสู่กระบวนการจำแนกชนิดของเซลล์ ในงานวิจัยชิ้นนี้ได้เลือกใช้การวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็มมาใช้ในขั้นตอนการหาค่าคุณลักษณะเด่น โดยทำการปรับแต่งพารามิเตอร์ทั้งในส่วนของการกระจัดและมุมสำหรับการสร้างเมทริกซ์จีแอลซีเอ็ม แล้วทำการคำนวณลักษณะเด่นของเมทริกซ์จีแอลซีเอ็ม ซึ่งค่าคุณลักษณะที่ได้จะถูกนำไปผ่านกระบวนการจำแนกโดยต้นไม้ตัดสินใจ จากผลการทดลองจะเห็นได้ว่าการสไลด์ภาพด้วยหน้าต่างขนาด  $256 \times 256$  พิกเซลจะให้ค่าความแม่นยำในการจำแนกที่สูงกว่าภาพที่สไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล สำหรับการวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็มในภาพถ่ายเซลล์มะเร็งเต้านมภายใต้กรอบข้อกำหนดของระยะห่างและมุมจะเห็นได้ว่าระยะห่างที่เหมาะสมจะอยู่ในช่วง 6 ถึง 10 พิกเซล ส่วนมุมที่เหมาะสมนั้นจะเห็นได้ว่าหน้าต่างทั้งสองขนาดจะมีมุมที่เหมาะสมนั้นแตกต่างกัน สำหรับหน้าต่างขนาด  $128 \times 128$  พิกเซลมุมที่เหมาะสมคือมุม 0 องศา ส่วนหน้าต่างขนาด  $256 \times 256$  พิกเซลมุมที่เหมาะสมคือมุม 135 องศา นอกจากมุมและระยะห่างจะมีผลต่อการวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็มแล้วสเกลระดับเทาเองก็เป็นอีกองค์ประกอบหนึ่งที่มีความสำคัญต่อการวิเคราะห์เชิงพื้นผิว โดยสเกลค่าระดับเทาที่เหมาะสมจะอยู่ที่ 128 ระดับ จากนั้นดำเนินการประเมินประสิทธิภาพของอัลกอริทึมในการจำแนกชนิดของเซลล์โดยใช้ค่าความไว ค่าความจำเพาะ ค่าพยากรณ์ผลบวก และค่าพยากรณ์ผลลบ

จากการทดสอบประสิทธิภาพการทำงานของอัลกอริทึมพบว่า การสไลด์ภาพด้วย หน้าต่างขนาด  $256 \times 256$  พิกเซล แล้วนำมาทำการวิเคราะห์ด้วยจีแอลซีเอ็มเมทริกซ์ที่ระยะห่าง 9 พิกเซล มุม 135 องศา ร่วมกับชุดค่าคุณลักษณะเด่นที่เหมาะสม 9 คุณลักษณะซึ่งประกอบไปด้วย ค่าสหสัมพันธ์แบบที่ 1 (f3) ค่าสหสัมพันธ์แบบที่ 2 (f4) ค่าความโดดเด่นของคลัสเตอร์ (f5) ค่าความเปลี่ยนแปลงของคลัสเตอร์ (f6) ค่าเอนโทรปี (f9) ค่าความเป็นเอกภาพแบบที่ 2 (f11) ค่าผลรวมเอนโทรปี (f15) ค่าการวัดข้อมูลของสหสัมพันธ์แบบที่ 1 (f19) และค่าการวัดข้อมูลของสหสัมพันธ์แบบที่ 2 (f20) ในการคำนวณค่าคุณลักษณะเด่นแล้วนำมาทำการจำแนกด้วยต้นไม้ตัดสินใจจะให้ค่าความแม่นยำในการจำแนกสูงที่สุดคือ 98.08 เปอร์เซ็นต์ ค่าความไว 99.25 เปอร์เซ็นต์ ค่าความจำเพาะ 97.25 เปอร์เซ็นต์ ค่าพยากรณ์ผลบวก 98.63 เปอร์เซ็นต์ และค่าพยากรณ์ผลลบ 98.48 เปอร์เซ็นต์ อัลกอริทึมที่ได้สามารถลดจำนวนผลบวกเท็จอันเกิดจากระบวนการวิเคราะห์ภาพของระบบนับจำนวนนิวเคลียสเซลล์มะเร็งเต้านมที่มีการพัฒนามาก่อนหน้านี้ได้ ซึ่งจากผลที่ได้แสดงให้เห็นว่าการวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็มเป็นอีกทฤษฎีหนึ่งที่สามารถนำมาประยุกต์ใช้ในกระบวนการจำแนกชนิดของเซลล์บนภาพถ่ายเซลล์มะเร็งเต้านมเพื่อช่วยสนับสนุนการรักษาของแพทย์ได้

## 5.2 ข้อเสนอแนะ

5.2.1 สำหรับการปรับสเกลระดับเทาได้ทำการทดสอบเฉพาะในชุดพารามิเตอร์ที่เหมาะสมในการวิเคราะห์จีแอลซีเอ็มที่ให้ผลการวิเคราะห์ที่ดีที่สุดเท่านั้น เพื่อศึกษาผลที่เกิดจากความละเอียดของสเกลต่อผลการวิเคราะห์ ซึ่งยังคงมีอีกหลายชุดพารามิเตอร์ที่สามารถทำการศึกษาค้นคว้าเพื่อปรับค่าสเกลและชุดพารามิเตอร์ที่มีความเหมาะสมกันมากที่สุด

5.2.2 สำหรับการคัดเลือกค่าคุณลักษณะเด่นโดยใช้การวิเคราะห์ค่าสหสัมพันธ์และระเบียบขั้นตอนวิธีละโมบ ได้ทำการศึกษาค้นคว้าโดยใช้ระเบียบขั้นตอนวิธีละโมบแบบไปข้างหน้าเพียงอย่างเดียวเท่านั้น ซึ่งการใช้วิธีดังกล่าวจะได้จำนวนคุณลักษณะเด่นที่ไม่มากเกินไปสำหรับจำแนกเมื่อเปรียบเทียบกับค่าความแม่นยำแล้วถือว่ามีความคุ้มค่ามากกว่า ซึ่งถ้าหากใช้การวัดค่าเกณฑ์ของข้อมูลมาเป็นขั้นตอนเลือกคุณลักษณะเด่นอาจจะทำให้ผลการจำแนกชนิดของเซลล์มีค่าความแม่นยำที่เทียบเท่าหรือดีกว่าแต่ก็อาจจะต้องแลกมาด้วยจำนวนของคุณลักษณะเด่นที่มากขึ้นนั่นเอง

5.2.3 สำหรับขั้นตอนการวิเคราะห์ภาพถ่ายเซลล์มะเร็งเต้านมทั้งหมดได้ทำการทดสอบลงบนภาพถ่ายเซลล์ที่มีการย้อมติดสีอย่างชัดเจนเท่านั้น ดังนั้นระเบียบวิธีที่ได้มาอาจจะไม่เหมาะสมสำหรับการวิเคราะห์ภาพถ่ายเซลล์ที่มีการย้อมติดสีที่ไม่สมบูรณ์ หรือสีซีดจาง

5.2.4 สำหรับขั้นตอนการสไลด์หน้าต่างได้ทดสอบเบื้องต้นด้วยการสไลด์หน้าต่าง 4 ขนาดด้วยกันคือ  $128 \times 128$  พิกเซล  $256 \times 256$  พิกเซล  $384 \times 384$  พิกเซล และ  $512 \times 512$  พิกเซล ซึ่งหน้าต่างขนาด  $384 \times 384$  พิกเซล และ  $512 \times 512$  พิกเซล เมื่อทำการสไลด์ภาพจะพบการปะปนของเซลล์เป็นจำนวนมากจึงไม่ได้เลือกมาใช้เป็นชุดทดสอบ สำหรับผู้ที่สนใจสามารถนำไปทำการศึกษาวิจัยต่อได้

5.2.5 จากผลการทดสอบอัลกอริทึมบนภาพต้นฉบับทำให้เห็นว่าค่าความแม่นยำที่ได้ยังไม่สูงมากพอ ดังนั้นการเลือกใช้องค์ประกอบของสีแทนขั้นตอนการแปลงภาพสี RGB มาเป็นภาพระดับเทาพร้อมกับการวิเคราะห์เชิงพื้นผิวแบบจีแอลซีเอ็มอาจเป็นอีกแนวทางหนึ่ง ซึ่งอาจจะสามารถเพิ่มค่าความแม่นยำให้กับอัลกอริทึมได้



### บรรณานุกรม

- [1] “WHO | 10 facts about cancer,” *WHO*. [Online]. Available: <http://www.who.int/features/factfiles/cancer/en/>. [Accessed: 19-Jun-2013].
- [2] “ทะเบียนมะเร็ง ระดับโรงพยาบาลสงขลานครินทร์,” หน่วยสารสนเทศมะเร็ง โรงพยาบาลสงขลานครินทร์. [Online]. Available: <http://medinfo2.psu.ac.th/cancer/index.php>.
- [3] “Cancer Registry 2002-2010,” งาม ทะ เบ็ ย น มะ ร็ ง . [ Online]. Available: [http://www.nci.go.th/th/cancer\\_record/cancer\\_rec1.html](http://www.nci.go.th/th/cancer_record/cancer_rec1.html).
- [4] R. Mofidi, R. Walsh, P. . Ridgway, T. Crotty, E. . McDermott, T. . Keaveny, M. . Duffy, A. D. . Hill, and N. O’Higgins, “Objective measurement of breast cancer oestrogen receptor status through digital image analysis,” *Eur. J. Surg. Oncol. EJSO*, vol. 29, no. 1, pp. 20–24, Feb. 2003.
- [5] Y. Hatanaka, K. Hashizume, K. Nitta, T. Kato, I. Itoh, and Y. Tani, “Cytometrical image analysis for immunohistochemical hormone receptor status in breast carcinomas,” *Pathol. Int.*, vol. 53, no. 10, pp. 693–699, 2003.
- [6] P. Phukpattaranont, S. Limsiroratana, K. Kayasut, and P. Boonyaphiphat, “Computer Aided System for Nuclear Stained Breast Cancer Cell Counting,” in *Breast Cancer - Recent Advances in Biology, Imaging and Therapeutics*, S. Done, Ed. InTech, 2011.
- [7] G. Xian, “An identification method of malignant and benign liver tumors from ultrasonography based on GLCM texture features and fuzzy SVM,” *Expert Syst. Appl.*, vol. 37, no. 10, pp. 6737–6741, 2010.
- [8] A. N. Karahaliou, I. S. Boniatis, S. G. Skiadopoulos, F. N. Sakellaropoulos, N. S. Arikidis, E. A. Likaki, G. S. Panayiotakis, and L. I. Costaridou, “Breast Cancer Diagnosis: Analyzing Texture of Tissue Surrounding Microcalcifications,” *Inf. Technol. Biomed. IEEE Trans.*, vol. 12, no. 6, pp. 731 –738, Nov. 2008.
- [9] W. Gomez, W. C. A. Pereira, and A. F. C. Infantosi, “Analysis of Co-Occurrence Texture Statistics as a Function of Gray-Level Quantization for Classifying Breast Ultrasound,” *IEEE Trans. Med. Imaging*, vol. 31, no. 10, pp. 1889 –1899, Oct. 2012.
- [10] “Color space - Wikipedia, the free encyclopedia.” [Online]. Available: [http://en.wikipedia.org/wiki/Color\\_space](http://en.wikipedia.org/wiki/Color_space). [Accessed: 29-Jan-2014].

- [11] “Color Management: Understanding Color Spaces.” [Online]. Available: <http://www.cambridgeincolour.com/tutorials/color-spaces.htm>. [Accessed: 29-Jan-2014].
- [12] “Convert RGB image or colormap to grayscale - MATLAB rgb2gray.” [Online]. Available: [www.mathworks.com/help/images/ref/rgb2gray.html?searchHighlight=rgb+to+gray](http://www.mathworks.com/help/images/ref/rgb2gray.html?searchHighlight=rgb+to+gray). [Accessed: 29-Jan-2014].
- [13] G. N. Srinivasan and Shobha G., “Statistical Texture Analysis,” in *WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY (PWASET)*, 2008, vol. 36, pp. 1264 – 1269.
- [14] “texture-review.pdf.” [Online]. Available: <http://cs.iupui.edu/~tuceryan/research/ComputerVision/texture-review.pdf>. [Accessed: 29-Jan-2014].
- [15] A. Materka, M. Strzelecki, T. Analysis, M. A. Review, A. Materka, and M. Strzelecki, “Texture analysis methods - a review,” Institute of Electronics, Technical University of Lodz, 1998.
- [16] D. Mitrea, M. Socaciu, R. Badea, and A. Golea, “Texture based characterization and automatic diagnosis of the abdominal tumors from ultrasound images using third order GLCM features,” in *Image and Signal Processing (CISP), 2011 4th International Congress on*, 2011, vol. 3, pp. 1558 –1562.
- [17] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural Features for Image Classification,” *Syst. Man Cybern. IEEE Trans.*, vol. SMC-3, no. 6, pp. 610 –621, Nov. 1973.
- [18] F. Albrechtsen, “Statistical Texture Measures Computed from Gray Level Cooccurrence Matrices.” 05-Nov-2008.
- [19] M. M. Mokji and S. A. R. Abu Bakar, “Gray Level Co-Occurrence Matrix Computation Based On Haar Wavelet,” in *Computer Graphics, Imaging and Visualisation, 2007. CGIV '07*, 2007, pp. 273 –279.
- [20] “GLCM Texture Tutorial.” [Online]. Available: <http://www.fp.ucalgary.ca/mhallbey/tutorial.htm>. [Accessed: 29-Jan-2014].
- [21] R. Parekh, “Using Texture Analysis for Medical Diagnosis,” *IEEE Multimed.*, vol. 19, no. 2, pp. 28–37, 2012.

- [22] X. Yang, S. Tridandapani, J. J. Beitler, D. S. Yu, E. J. Yoshida, W. J. Curran, and T. Liu, "Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: An in vivo study of late toxicity," *Med. Phys.*, vol. 39, no.9, pp. 5732–5739, 2012.
- [23] S. Park, B. Kim, J. Lee, J. M. Goo, and Y. G. Shin, "GGO Nodule Volume-Preserving Nonrigid Lung Registration Using GLCM Texture Analysis," *Biomed. Eng. IEEE Trans.*, vol. 58, no. 10, pp. 2885–2894, Oct. 2011.
- [24] V. V. S. Tallapragada and E. G. Rajan, "Iris Recognition Based on Combined Feature of GLCM and Wavelet Transform," in *2010 First International Conference on Integrated Intelligent Computing (ICIIC)*, 2010, pp. 205–210.
- [25] N. Manwani and P. S. Sastry, "Geometric Decision Tree," *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 42, no. 1, pp. 181–192, 2012.
- [26] W. Reichl and W. Chou, "Robust decision tree state tying for continuous speech recognition," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 555–566, 2000.
- [27] M. Simard, S. S. Saatchi, and G. De Grandi, "The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2310–2321, 2000.
- [28] T. Tirelli and D. Pessani, "Importance of feature selection in decision-tree and artificial-neural-network ecological applications. *Alburnus alburnus alborella*: A practical example," *Ecol. Informatics*, vol. 6, no. 5, pp. 309–315, Sep. 2011.
- [29] | W., "การลดจำนวน feature สำหรับการเรียนรู้แบบมีผู้สอน," *Thailand's Machine Learning Research*. [Online]. Available: <http://thaiml.org/?p=34>. [Accessed: 08-Dec-2013].
- [30] | W., "ภาพรวมของ feature selection," *Thailand's Machine Learning Research*. [Online]. Available: <http://thaiml.org/?p=99>. [Accessed: 08-Dec-2013].
- [31] M. Pedergnana, P. R. Marpu, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "A Novel Technique for Optimal Feature Selection in Attribute Profiles Based on Genetic Algorithms," *Ieee Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3514–3528, Jun. 2013.
- [32] F. Maes, D. Vandermeulen, and P. Suetens, "Medical image registration using mutual information," *Proc. IEEE*, vol. 91, no. 10, pp. 1699–1722, Oct. 2003.

- [33] P. C. Pendharkar, "A maximum-margin genetic algorithm for misclassification cost minimizing feature selection problem," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 3918 – 3925, Aug. 2013.
- [34] "Information gain in decision trees - Wikipedia, the free encyclopedia." [Online]. Available: [http://en.wikipedia.org/wiki/Information\\_gain\\_in\\_decision\\_trees](http://en.wikipedia.org/wiki/Information_gain_in_decision_trees). [Accessed: 29-Jan-2014].
- [35] "Entropy and Information Gain." [Online]. Available: [www.cs.csi.cuny.edu/~imberman/ai/Entropy%20and%20Information%20Gain.htm](http://www.cs.csi.cuny.edu/~imberman/ai/Entropy%20and%20Information%20Gain.htm). [Accessed: 29-Jan-2014].
- [36] "Introduction to Artificial Intelligence - InfoGain.pdf." [Online]. Available: <http://homes.cs.washington.edu/~shapiro/EE596/notes/InfoGain.pdf>. [Accessed: 29-Jan-2014].
- [37] "Feature selection," *Wikipedia, the free encyclopedia*. 19-Nov-2013.
- [38] R. Sadeghi, R. Zarkami, K. Sabetraftar, and P. Van Damme, "Application of genetic algorithm and greedy stepwise to select input variables in classification tree models for the prediction of habitat requirements of *Azolla filiculoides* (Lam.) in Anzali wetland, Iran," *Ecol. Model.*, vol. 251, pp. 44–53, Feb. 2013.
- [39] "ขั้นตอนวิธีประเภทละโมภ (Greedy algorithm)." [Online]. Available: <http://pioneer.netserv.chula.ac.th/~skrung/2301365/Lecture009.html>. [Accessed: 23-Mar-2014].
- [40] "เทคนิคการออกแบบอัลกอริทึม (Algorithm Technique) - Greedy Algorithm." [Online]. Available: [http://www.g2gnet.com/News/activenews\\_view.asp?articleID=7](http://www.g2gnet.com/News/activenews_view.asp?articleID=7). [Accessed: 23-Mar-2014].

ภาคผนวก

ภาคผนวก ก

ผลการจำแนกชนิดของเซลล์

ตารางที่ ก-1 ค่าความแม่นยำในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นทั้ง 22 คุณลักษณะในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 128 × 128 พิกเซล

พารามิเตอร์ในการวิเคราะห์ จีแอลซีเอ็ม		ค่าความแม่นยำในการจำแนก				ค่าเบี่ยงเบน มาตรฐาน
ระยะห่าง	มุม	PC	LC	SC	All	
1	0	92.50	74.00	76.00	86.67	4.12
1	45	91.75	72.00	81.00	86.67	3.26
1	90	90.13	69.50	81.00	85.17	3.19
1	135	89.38	75.00	76.00	84.75	4.12
2	0	90.75	76.50	78.00	86.25	3.99
2	45	91.00	71.00	77.50	85.42	2.23
2	90	89.88	68.50	74.50	83.75	2.73
2	135	90.38	75.50	76.50	85.58	2.89
3	0	91.50	67.50	80.50	85.67	4.41
3	45	91.38	72.00	80.00	86.25	3.75
3	90	90.25	67.00	80.00	84.67	3.75
3	135	91.38	70.50	79.00	85.83	2.39
4	0	92.50	68.50	78.00	86.08	2.49
4	45	90.75	72.50	82.50	86.33	3.43
4	90	91.38	63.50	72.00	83.50	3.72
4	135	90.13	69.00	78.50	84.67	4.45
5	0	92.50	70.50	77.00	86.25	3.17
5	45	91.50	70.00	82.50	86.42	3.73
5	90	91.75	67.50	80.00	85.75	4.26
5	135	92.38	72.00	77.00	86.42	3.33
6	0	93.38	68.50	80.50	87.08	2.64
6	45	92.13	74.00	81.50	87.33	4.21
6	90	89.50	76.00	77.50	85.25	2.81
6	135	91.50	75.50	80.50	87.00	2.94
7	0	92.63	75.00	79.00	87.42	3.37
7	45	91.25	69.00	82.00	86.00	2.32
7	90	90.25	68.50	80.50	85.00	2.08
7	135	91.38	73.00	80.00	86.42	2.99
8	0	91.25	66.50	75.50	84.50	2.67
8	45	91.38	71.50	82.50	86.58	1.82
8	90	90.13	69.50	80.50	85.08	2.93
8	135	92.00	72.00	76.50	86.08	3.14
9	0	91.50	68.00	76.50	85.08	3.32
9	45	92.25	69.50	84.50	87.17	2.67
9	90	90.88	68.50	79.00	85.17	2.22
9	135	89.88	70.50	79.00	84.83	3.92
10	0	91.25	68.50	79.50	85.50	3.50
10	45	92.63	70.50	84.50	87.58	2.95
10	90	89.50	69.50	78.00	84.25	2.79
10	135	90.75	67.50	81.00	85.25	2.22

ตารางที่ ก-2 ค่าความแม่นยำในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นที่ถูกเลือกจากการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 128 × 128 พิกเซล

พารามิเตอร์ในการวิเคราะห์จีแอลซีเอ็ม		จำนวนคุณลักษณะเด่น	ค่าความแม่นยำในการจำแนก				ค่าเบี่ยงเบนมาตรฐาน
ระยะห่าง	มุม		PC	LC	SC	รวม	
1	0	6	91.75	72.00	76.00	85.83	3.33
1	45	11	90.63	74.00	80.00	86.08	2.86
1	90	7	89.00	71.00	79.50	84.42	3.12
1	135	9	89.50	69.50	73.50	83.50	3.11
2	0	6	91.75	70.00	74.50	85.25	2.49
2	45	8	91.00	72.00	77.50	85.58	1.89
2	90	7	90.13	71.50	80.50	85.42	3.75
2	135	7	91.13	68.00	76.00	84.75	4.16
3	0	10	93.13	69.50	75.50	86.25	2.09
3	45	9	91.25	71.50	81.00	86.25	3.56
3	90	8	91.38	68.00	82.00	85.92	2.53
3	135	10	91.00	68.50	81.50	85.67	3.74
4	0	10	93.00	73.00	76.50	86.92	3.73
4	45	9	90.38	74.50	81.50	86.25	3.17
4	90	7	91.50	66.00	75.00	84.50	3.02
4	135	8	90.88	67.50	76.50	84.58	3.47
5	0	11	93.25	74.50	79.00	87.75	2.81
5	45	9	91.25	74.00	86.50	87.58	3.82
5	90	7	90.13	68.00	77.00	84.25	3.43
5	135	10	91.00	64.00	81.00	84.83	3.01
6	0	11	93.75	69.00	78.00	87.00	3.17
6	45	9	91.00	70.50	80.50	85.83	2.64
6	90	6	89.88	71.00	79.00	84.92	3.39
6	135	10	92.00	68.50	79.50	86.00	2.85
7	0	8	92.75	71.00	79.50	86.92	3.54
7	45	9	92.88	70.00	83.50	87.50	2.32
7	90	8	90.13	72.00	78.00	85.08	2.5
7	135	6	91.13	72.50	80.50	86.25	3.5
8	0	8	92.00	74.50	73.50	86.00	2.91
8	45	10	92.00	71.50	79.00	86.42	1.93
8	90	8	90.50	74.50	78.00	85.75	2.71
8	135	8	92.00	69.00	78.00	85.83	1.71
9	0	8	92.25	66.50	76.00	85.25	2.15
9	45	10	92.88	68.50	83.50	87.25	2.36
9	90	8	91.00	72.00	77.50	85.58	3.54
9	135	8	90.13	68.00	80.50	84.83	4.21
10	0	8	92.88	69.00	73.50	85.67	2.48
10	45	10	91.75	70.00	84.50	86.92	3.12
10	90	8	91.25	68.00	74.00	84.50	2.36
10	135	9	91.38	69.50	78.50	85.58	3.93



ตารางที่ ก-3 ค่าความแม่นยำในการจำแนกชนิดของเซลล์โดยใช้ชุดค่าคุณลักษณะเด่นที่เหมาะสมในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 128 × 128 พิกเซล

พารามิเตอร์ในการวิเคราะห์ แอลซีเอ็ม		ค่าความแม่นยำในการจำแนก				ค่าเบี่ยงเบน มาตรฐาน
ระยะห่าง	มุม	PC	LC	SC	รวม	
1	0	91.50	70.50	75.50	85.50	3.43
1	45	90.75	75.50	80.50	86.50	2.99
1	90	90.13	68.00	77.50	84.33	3.11
1	135	89.75	67.50	73.50	83.33	3.66
2	0	91.75	66.00	76.50	84.92	4.22
2	45	91.13	73.00	78.00	85.92	1.82
2	90	90.38	72.00	78.50	85.33	2.94
2	135	91.13	70.00	76.50	85.17	3.16
3	0	92.13	70.50	77.00	86.00	3.68
3	45	91.13	73.00	80.50	86.33	3.81
3	90	91.75	70.00	79.50	86.08	2.29
3	135	90.88	69.00	81.00	85.58	3.73
4	0	92.00	72.00	77.00	86.17	3.81
4	45	90.00	76.50	81.50	86.33	3.58
4	90	92.00	66.50	74.50	84.83	4.21
4	135	90.88	67.50	76.50	84.58	3.47
5	0	93.75	74.00	77.50	87.75	2.69
5	45	91.63	75.50	87.00	88.17	3.49
5	90	90.63	66.00	79.50	84.67	3.52
5	135	91.63	66.00	82.50	85.83	2.75
6	0	94.13	74.50	80.50	88.58	2.42
6	45	91.00	73.50	81.50	86.50	2.96
6	90	90.63	74.50	77.50	85.75	2.98
6	135	92.25	69.00	80.50	86.42	2.78
7	0	93.88	77.00	79.50	88.67	2.97
7	45	91.75	70.00	83.00	86.67	2.64
7	90	90.13	72.00	78.00	85.08	2.5
7	135	91.75	72.00	81.00	86.67	3.45
8	0	92.13	69.50	74.50	85.42	3.73
8	45	92.38	68.50	79.50	86.25	1.93
8	90	90.50	74.50	78.00	85.75	2.71
8	135	90.50	69.00	78.50	84.92	2.98
9	0	92.25	73.00	76.00	86.33	1.77
9	45	93.25	67.50	83.00	87.25	2.12
9	90	91.00	72.00	77.50	85.58	3.54
9	135	90.50	64.00	82.00	84.67	4.12
10	0	92.13	72.50	78.00	86.50	3.57
10	45	92.38	68.00	86.50	87.33	2.85
10	90	91.25	68.00	74.00	84.50	2.36
10	135	91.13	68.50	80.00	85.50	3.56

ตารางที่ ก-4 ค่าความแม่นยำในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นทั้ง 22 คุณลักษณะในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $256 \times 256$  พิกเซล

พารามิเตอร์ในการวิเคราะห์ แอลซีเอ็ม		ค่าความแม่นยำในการจำแนก				ค่าเบี่ยงเบน มาตรฐาน
ระยะห่าง	มุม	PC	LC	SC	All	
1	0	96.13	84.50	94.00	93.83	1.85
1	45	96.00	87.00	94.50	94.25	2.13
1	90	97.50	88.00	93.00	95.17	1.66
1	135	98.13	89.50	93.50	95.17	2.02
2	0	96.88	86.50	95.50	94.92	2.17
2	45	96.25	85.00	93.50	93.92	2.91
2	90	97.13	88.00	95.00	95.25	1.93
2	135	97.75	90.00	93.50	95.75	1.44
3	0	96.00	89.00	93.50	94.42	1.62
3	45	96.63	84.50	93.50	94.08	1.69
3	90	97.50	91.50	97.00	96.42	1.11
3	135	98.25	94.00	94.00	96.83	1.79
4	0	96.13	89.00	93.50	94.50	2.49
4	45	96.38	85.50	92.00	93.83	2.01
4	90	98.38	94.00	96.50	97.33	1.88
4	135	99.13	94.00	95.50	97.67	1.70
5	0	96.88	89.50	92.50	94.92	1.98
5	45	96.75	84.00	94.00	94.17	1.42
5	90	98.13	92.50	95.00	96.67	1.47
5	135	98.88	94.00	95.00	97.42	1.39
6	0	96.38	89.50	93.50	94.75	1.52
6	45	95.38	86.00	92.50	93.33	1.42
6	90	98.75	90.50	95.50	96.83	1.61
6	135	99.13	95.50	95.00	97.83	1.05
7	0	96.88	89.00	91.50	94.67	1.81
7	45	95.88	85.50	95.00	94.00	2.11
7	90	98.75	93.00	96.50	97.42	1.49
7	135	99.00	95.50	95.00	97.75	1.11
8	0	97.38	88.00	92.50	95.00	2.52
8	45	96.38	82.50	94.00	93.67	1.81
8	90	98.75	90.00	96.50	96.92	1.31
8	135	98.88	94.00	96.50	97.67	1.02
9	0	97.00	89.00	93.00	95.00	2.08
9	45	96.00	81.50	95.00	93.42	1.69
9	90	98.50	89.50	94.50	96.33	1.53
9	135	98.88	95.00	95.50	97.67	1.10
10	0	97.00	87.50	93.50	94.83	1.79
10	45	94.75	84.50	93.00	92.75	2.04
10	90	98.63	93.00	95.50	97.17	1.43
10	135	99.13	94.50	96.50	97.92	0.90

ตารางที่ ก-5 ค่าความแม่นยำในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นที่ถูกเลือกจากการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีละโมบในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล

พารามิเตอร์ในการวิเคราะห์จีแอลซีเอ็ม		จำนวนคุณลักษณะเด่น	ค่าความแม่นยำในการจำแนก				ค่าเบี่ยงเบนมาตรฐาน
ระยะห่าง	มุม		PC	LC	SC	All	
1	0	5	94.63	87.50	94.50	93.42	2.76
1	45	5	95.50	81.50	93.50	92.83	1.97
1	90	5	96.75	79.00	92.50	93.08	2.49
1	135	5	97.13	89.00	94.50	95.33	1.93
2	0	5	95.00	87.00	94.00	93.50	3.49
2	45	6	95.50	80.00	93.50	92.58	1.98
2	90	5	97.13	88.00	92.50	94.83	3.01
2	135	8	97.25	91.50	94.00	95.75	2.17
3	0	5	94.88	88.00	94.00	93.58	2.78
3	45	6	96.38	89.00	93.00	94.58	1.81
3	90	8	97.63	90.00	93.50	95.67	1.51
3	135	9	98.13	93.00	96.00	96.92	1.84
4	0	8	96.88	87.00	92.00	94.42	2.26
4	45	7	94.75	86.50	93.50	93.17	2.42
4	90	9	98.25	91.00	94.50	96.42	1.84
4	135	9	99.00	92.00	95.00	97.17	1.77
5	0	8	96.50	90.00	93.50	94.92	1.98
5	45	6	95.75	86.00	93.50	93.75	2.49
5	90	9	98.00	91.50	93.50	96.17	1.12
5	135	9	99.00	95.00	96.00	97.83	1.37
6	0	8	96.38	90.50	91.00	94.50	1.93
6	45	7	95.63	88.50	92.50	93.92	2.81
6	90	9	98.75	90.00	94.00	96.50	1.61
6	135	9	99.25	95.50	95.50	98.00	1.31
7	0	8	96.63	88.50	91.00	94.33	2.25
7	45	7	94.50	85.00	93.50	92.50	1.67
7	90	9	98.75	94.50	95.00	97.42	1.54
7	135	9	98.88	96.00	95.00	97.75	1.25
8	0	9	96.50	89.50	91.50	94.50	2.16
8	45	9	95.88	84.00	93.00	93.42	2.31
8	90	9	98.50	93.00	94.50	96.92	1.47
8	135	9	99.13	95.50	95.50	97.92	1.13
9	0	7	97.38	88.50	91.50	94.92	1.73
9	45	6	95.25	84.00	93.00	93.00	2.09
9	90	9	98.63	91.50	95.50	96.92	1.31
9	135	9	99.25	94.50	97.00	98.08	0.69
10	0	7	97.00	90.50	89.50	94.67	1.97
10	45	6	95.25	83.50	94.00	93.08	1.76
10	90	9	99.13	90.00	94.50	96.83	1.29
10	135	9	99.13	95.00	96.00	97.92	0.81

ตารางที่ ก-6 ค่าความแม่นยำในการจำแนกชนิดของเซลล์โดยใช้ชุดค่าคุณลักษณะเด่นที่เหมาะสมในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล

พารามิเตอร์ในการวิเคราะห์ แอลซีเอ็ม		ค่าความแม่นยำในการจำแนก				ค่าเบี่ยงเบน มาตรฐาน
ระยะห่าง	มุม	PC	LC	SC	All	
1	0	95.00	85.50	94.50	93.33	2.19
1	45	95.25	84.50	94.50	93.33	1.92
1	90	96.50	85.50	93.00	94.08	2.40
1	135	97.75	88.00	92.50	95.25	2.69
2	0	95.13	86.00	94.50	93.50	2.32
2	45	95.25	86.00	95.00	93.67	1.31
2	90	97.88	90.00	93.50	95.83	1.62
2	135	97.75	90.50	94.50	96.00	1.46
3	0	96.00	83.50	93.50	93.50	2.99
3	45	96.63	88.50	94.50	94.92	1.90
3	90	98.00	92.00	94.50	96.42	1.67
3	135	98.13	93.50	96.50	97.08	1.97
4	0	96.50	89.00	93.50	94.75	1.42
4	45	95.00	87.00	92.50	93.25	2.40
4	90	98.38	95.50	96.00	97.50	1.24
4	135	99.00	92.00	95.00	97.17	1.77
5	0	96.25	90.50	91.50	94.50	2.26
5	45	94.25	88.50	93.00	93.08	2.12
5	90	98.00	93.50	95.00	96.75	1.00
5	135	99.00	95.00	96.00	97.83	1.43
6	0	96.00	89.50	92.00	94.25	1.82
6	45	95.38	85.50	92.00	93.17	2.07
6	90	98.63	92.00	95.00	96.92	1.11
6	135	99.25	95.50	95.50	98.00	1.31
7	0	95.38	91.50	90.50	93.92	2.49
7	45	95.88	84.50	93.50	93.58	2.45
7	90	98.63	95.00	96.00	97.58	1.14
7	135	99.00	96.00	95.00	97.83	1.25
8	0	95.88	91.50	93.00	94.67	2.01
8	45	96.38	80.50	94.00	93.33	1.57
8	90	98.63	94.00	97.00	97.58	1.27
8	135	99.13	95.50	95.50	97.92	1.13
9	0	97.13	89.00	92.50	95.00	1.84
9	45	96.63	83.00	92.50	93.67	2.12
9	90	98.63	93.50	96.50	97.42	1.33
9	135	99.25	94.50	97.00	98.08	0.69
10	0	96.38	90.00	91.00	94.42	1.62
10	45	95.00	85.00	92.50	92.92	1.06
10	90	98.75	95.00	96.50	97.75	1.52
10	135	99.13	95.00	96.00	97.92	0.81

**ภาคผนวก ข**

**ผลการประเมินประสิทธิภาพบนชุดข้อมูล**

ตาราง ข-1 ผลการประเมินประสิทธิภาพในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นทั้ง 22 คุณลักษณะ  
ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 128 × 128 พิกเซล

พารามิเตอร์ในการ วิเคราะห์จีแอลซีเอ็ม		ผลบวกจริง	ผลลบจริง	ผลบวกเท็จ	ผลลบเท็จ	ค่าความไว	ค่า ความจำเพาะ	ค่าพยากรณ์ ผลบวก	ค่าพยากรณ์ ผลลบ
ระยะห่าง	มุม								
1	0	740	330	70	60	92.50	82.50	91.36	84.62
1	45	734	332	68	66	91.75	83.00	91.52	83.42
1	90	721	337	63	79	90.13	84.25	91.96	81.01
1	135	715	333	67	85	89.38	83.25	91.43	79.67
2	0	726	343	57	74	90.75	85.75	92.72	82.25
2	45	728	323	77	72	91.00	80.75	90.43	81.77
2	90	719	319	81	81	89.88	79.75	89.88	79.75
2	135	723	344	56	77	90.38	86.00	92.81	81.71
3	0	732	330	70	68	91.50	82.50	91.27	82.91
3	45	731	331	69	69	91.38	82.75	91.38	82.75
3	90	722	329	71	78	90.25	82.25	91.05	80.84
3	135	731	332	68	69	91.38	83.00	91.49	82.79
4	0	740	329	71	60	92.50	82.25	91.25	84.58
4	45	726	335	65	64	91.90	83.75	91.78	83.96
4	90	731	310	90	69	91.38	77.50	89.04	81.79
4	135	721	324	76	79	90.13	81.00	90.46	80.40
5	0	740	336	64	60	92.50	84.00	92.04	84.85
5	45	732	338	62	68	91.50	84.50	92.19	83.25
5	90	734	325	75	66	91.75	81.25	90.73	83.12
5	135	739	328	72	61	92.38	82.00	91.12	84.32
6	0	747	336	64	53	93.38	84.00	92.11	86.38
6	45	737	335	65	63	92.13	83.75	91.90	84.17
6	90	716	338	62	84	89.50	84.50	92.03	80.09
6	135	732	341	59	68	91.50	85.25	92.54	83.37
7	0	741	338	62	59	92.63	84.50	92.28	85.14
7	45	730	332	68	70	91.25	83.00	91.48	82.59
7	90	722	334	66	78	90.25	83.50	91.62	81.07
7	135	731	335	65	69	91.38	83.75	91.83	82.92
8	0	730	331	69	70	91.25	82.75	91.36	82.54
8	45	731	335	65	69	91.38	83.75	91.83	82.92
8	90	721	336	64	79	90.13	84.00	91.85	80.96
8	135	736	329	71	64	92.00	82.25	91.20	83.72
9	0	732	328	72	68	91.50	82.00	91.04	82.83
9	45	738	334	66	62	92.25	83.50	91.79	84.34
9	90	727	325	75	73	90.88	81.25	90.65	81.66
9	135	719	335	65	81	89.88	83.75	91.71	80.53
10	0	730	334	66	70	91.25	83.50	91.71	82.67
10	45	741	340	60	59	92.63	85.00	92.51	85.21
10	90	716	325	75	84	89.50	81.25	90.52	79.46
10	135	726	336	64	74	90.75	84.00	91.90	81.95

ตาราง ข-2 ผลการประเมินประสิทธิภาพในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นที่ถูกเลือกจากการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีสโตนโมบในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด  $128 \times 128$  พิกเซล

พารามิเตอร์ในการวิเคราะห์หัจฉัลเอ็ลซีเอ็ม		ผลบวกจริง	ผลลบจริง	ผลบวกเท็จ	ผลลบเท็จ	ค่าความไว	ค่าความจำเพาะ	ค่าพยากรณ์ผลบวก	ค่าพยากรณ์ผลลบ
ระยะห่าง	มุม								
1	0	734	332	68	66	91.75	83.00	91.52	83.42
1	45	725	337	63	75	90.63	84.25	92.01	81.80
1	90	712	333	67	88	89.00	83.25	91.40	79.10
1	135	716	323	77	84	89.50	80.75	90.29	79.36
2	0	734	320	80	66	91.75	80.00	90.17	82.90
2	45	728	327	73	72	91.00	81.75	90.89	81.95
2	90	721	336	64	79	90.13	84.00	91.85	80.96
2	135	729	329	71	71	91.13	82.25	91.13	82.25
3	0	745	320	80	55	93.13	80.00	90.30	85.33
3	45	730	336	64	70	91.25	84.00	91.94	82.76
3	90	731	329	71	69	91.38	82.25	91.15	82.66
3	135	728	330	70	72	91.00	82.50	91.23	82.09
4	0	744	332	68	56	93.00	83.00	91.63	85.57
4	45	723	337	63	77	90.38	84.25	91.98	81.40
4	90	732	304	96	68	91.50	76.00	88.41	81.72
4	135	727	314	86	73	90.88	78.50	89.42	81.14
5	0	746	340	60	54	93.25	85.00	92.56	86.29
5	45	730	347	53	70	91.25	86.75	93.23	83.21
5	90	721	319	81	79	90.13	79.75	89.90	80.15
5	135	728	327	73	72	91.00	81.75	90.89	81.95
6	0	750	336	64	50	93.75	84.00	92.14	87.05
6	45	728	325	75	72	91.00	81.25	90.66	81.86
6	90	719	333	67	81	89.88	83.25	91.48	80.43
6	135	736	327	73	64	92.00	81.75	90.98	83.63
7	0	742	342	58	58	92.75	85.50	92.75	85.50
7	45	743	334	66	57	92.88	83.50	91.84	85.42
7	90	721	335	65	79	90.13	83.75	91.73	80.92
7	135	729	339	61	71	91.13	84.75	92.28	82.68
8	0	736	337	63	64	92.00	84.25	92.12	84.04
8	45	736	326	74	64	92.00	81.50	90.86	83.59
8	90	724	338	62	76	90.50	84.50	92.11	81.64
8	135	736	333	67	64	92.00	83.25	91.66	83.88
9	0	738	332	68	62	92.25	83.00	91.56	84.26
9	45	743	329	71	57	92.88	82.25	91.28	85.23
9	90	728	334	66	72	91.00	83.50	91.69	82.27
9	135	721	329	71	79	90.13	82.25	91.04	80.64
10	0	743	331	69	57	92.88	82.75	91.50	85.31
10	45	734	338	62	66	91.75	84.50	92.21	83.66
10	90	730	333	67	70	91.25	83.25	91.59	82.63
10	135	731	334	66	69	91.38	83.50	91.72	82.88

ตาราง ข-3 ผลการประเมินประสิทธิภาพในการจำแนกชนิดของเซลล์โดยใช้โดยใช้ชุดค่าคุณลักษณะเด่นที่เหมาะสมในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 128 × 128 พิกเซล

พารามิเตอร์ในการวิเคราะห์จีแอลซีเอ็ม		ผลบวกจริง	ผลลบจริง	ผลบวกเท็จ	ผลลบเท็จ	ค่าความไว	ค่าความจำเพาะ	ค่าพยากรณ์ผลบวก	ค่าพยากรณ์ผลลบ
ระยะห่าง	มุม								
1	0	732	328	72	68	91.50	82.00	91.04	82.83
1	45	726	346	54	74	90.75	86.50	93.08	82.38
1	90	721	327	73	79	90.13	81.75	90.81	80.54
1	135	718	315	85	82	89.75	78.75	89.41	79.35
2	0	734	325	75	66	91.75	81.25	90.73	83.12
2	45	729	333	67	71	91.13	83.25	91.58	82.43
2	90	723	336	64	77	90.38	84.00	91.87	81.36
2	135	729	328	72	71	91.13	82.00	91.01	82.21
3	0	737	320	80	63	92.13	80.00	90.21	83.55
3	45	729	333	67	71	91.13	83.25	91.58	82.43
3	90	734	328	72	66	91.75	82.00	91.07	83.25
3	135	727	332	68	73	90.88	83.00	91.45	81.98
4	0	736	328	72	64	92.00	82.00	91.09	83.67
4	45	720	340	60	80	90.00	85.00	92.31	80.95
4	90	736	311	89	64	92.00	77.75	89.21	82.93
4	135	727	314	86	73	90.88	78.50	89.42	81.14
5	0	750	340	60	50	93.75	85.00	92.59	87.18
5	45	733	346	54	67	91.63	86.50	93.14	83.78
5	90	725	319	81	75	90.63	79.75	89.95	80.96
5	135	733	329	71	67	91.63	82.25	91.17	83.08
6	0	753	339	61	47	94.13	84.75	92.51	87.82
6	45	728	331	69	72	91.00	82.75	91.34	82.13
6	90	725	331	69	75	90.63	82.75	91.31	81.53
6	135	738	331	69	62	92.25	82.75	91.45	84.22
7	0	751	343	57	49	93.88	85.75	92.95	87.50
7	45	734	332	68	66	91.75	83.00	91.52	83.42
7	90	721	335	65	79	90.13	83.75	91.73	80.92
7	135	734	337	63	66	91.75	84.25	92.10	83.62
8	0	737	324	76	63	92.13	81.00	90.65	83.72
8	45	739	319	81	61	92.38	79.75	90.12	83.95
8	90	724	338	62	76	90.50	84.50	92.11	81.64
8	135	724	329	71	76	90.50	82.25	91.07	81.23
9	0	738	331	69	62	92.25	82.75	91.45	84.22
9	45	746	327	73	54	93.25	81.75	91.09	85.83
9	90	728	334	66	72	91.00	83.50	91.69	82.27
9	135	724	319	81	76	90.50	79.75	89.94	80.76
10	0	737	333	67	63	92.13	83.25	91.67	84.09
10	45	739	338	62	61	92.38	84.50	92.26	84.71
10	90	730	333	67	70	91.25	83.25	91.59	82.63
10	135	729	333	67	71	91.13	83.25	91.58	82.43



ตาราง ข-4 ผลการประเมินประสิทธิภาพในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นทั้ง 22 คุณลักษณะ  
ในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล

พารามิเตอร์ในการ วิเคราะห์จีแอลซีเอ็ม		ผลบวกจริง	ผลลบจริง	ผลบวกเท็จ	ผลลบเท็จ	ค่าความไว	ค่า ความจำเพาะ	ค่าพยากรณ์ ผลบวก	ค่าพยากรณ์ ผลลบ
ระยะห่าง	มุม								
1	0	769	366	34	31	96.13	91.50	95.77	92.19
1	45	768	372	28	32	96.00	93.00	96.48	92.08
1	90	780	378	22	20	97.50	94.50	97.26	94.97
1	135	785	374	26	15	98.13	93.50	96.79	96.14
2	0	775	372	28	25	96.88	93.00	96.51	93.70
2	45	770	367	33	30	96.25	91.75	95.89	92.44
2	90	777	379	21	23	97.13	94.75	97.37	94.28
2	135	782	382	18	18	97.75	95.50	97.75	95.50
3	0	768	372	28	32	96.00	93.00	96.48	92.08
3	45	773	370	30	27	96.63	92.50	96.26	93.20
3	90	780	385	15	20	97.50	96.25	98.11	95.06
3	135	786	389	11	14	98.25	97.25	98.62	96.53
4	0	769	376	24	31	96.13	94.00	96.97	92.38
4	45	771	366	34	29	96.38	91.50	95.78	92.66
4	90	787	390	10	13	98.38	97.50	98.75	96.77
4	135	793	390	10	7	99.13	97.50	98.75	98.24
5	0	775	370	30	25	96.88	92.50	96.27	93.67
5	45	774	366	34	26	96.75	91.50	95.79	93.37
5	90	785	384	16	15	98.13	96.00	98.00	96.24
5	135	791	388	12	9	98.88	97.00	98.51	97.73
6	0	771	374	26	29	96.38	93.50	96.74	92.80
6	45	763	368	32	37	95.38	92.00	95.97	90.86
6	90	790	385	15	10	98.75	96.25	98.14	97.47
6	135	793	387	13	7	99.13	96.75	98.39	98.22
7	0	775	370	30	25	96.88	92.50	96.27	93.67
7	45	767	369	31	33	95.88	92.25	96.12	91.79
7	90	790	387	13	10	98.75	96.75	98.38	97.48
7	135	792	386	14	8	99.00	96.50	98.26	97.97
8	0	779	373	27	21	97.38	93.25	96.65	94.67
8	45	771	356	44	29	96.38	89.00	94.60	92.47
8	90	790	383	17	10	98.75	95.75	97.89	97.46
8	135	791	385	15	9	98.88	96.25	98.14	97.72
9	0	776	376	24	24	97.00	94.00	97.00	94.00
9	45	768	363	37	32	96.00	90.75	95.40	91.90
9	90	788	382	18	12	98.50	95.50	97.77	96.95
9	135	791	386	14	9	98.88	96.50	98.26	97.72
10	0	776	372	28	24	97.00	93.00	96.52	93.94
10	45	758	362	38	42	94.75	90.50	95.23	89.60
10	90	789	386	14	11	98.63	96.50	98.26	97.23
10	135	793	386	14	7	99.13	96.50	98.27	98.22

ตาราง ข-5 ผลการประเมินประสิทธิภาพในการจำแนกชนิดของเซลล์โดยใช้ค่าคุณลักษณะเด่นที่ถูกเลือกจากการวิเคราะห์ค่าสหสัมพันธ์ร่วมกับระเบียบขั้นตอนวิธีสะโมบในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล

พารามิเตอร์ในการวิเคราะห์หัจแอลซีเอ็ม		ผลบวกจริง	ผลลบจริง	ผลบวกเท็จ	ผลลบเท็จ	ค่าความไว	ค่าความจำเพาะ	ค่าพยากรณ์ผลบวก	ค่าพยากรณ์ผลลบ
ระยะห่าง	มุม								
1	0	757	373	27	43	94.63	93.25	96.56	89.66
1	45	764	364	36	36	95.50	91.00	95.50	91.00
1	90	774	359	41	26	96.75	89.75	94.97	93.25
1	135	777	378	22	23	97.13	94.50	97.25	94.26
2	0	760	369	31	40	95.00	92.25	96.08	90.22
2	45	764	358	42	36	95.50	89.50	94.79	90.86
2	90	777	378	22	23	97.13	94.50	97.25	94.26
2	135	778	384	16	22	97.25	96.00	97.98	94.58
3	0	759	370	30	41	94.88	92.50	96.20	90.02
3	45	771	375	25	29	96.38	93.75	96.86	92.82
3	90	781	386	14	19	97.63	96.50	98.24	95.31
3	135	785	389	11	15	98.13	97.25	98.62	96.29
4	0	775	364	36	25	96.88	91.00	95.56	93.57
4	45	758	371	29	42	94.75	92.75	96.32	89.83
4	90	786	385	15	14	98.25	96.25	98.13	96.49
4	135	792	386	14	8	99.00	96.50	98.26	97.97
5	0	772	372	28	28	96.50	93.00	96.50	93.00
5	45	766	370	30	34	95.75	92.50	96.23	91.58
5	90	784	384	16	16	98.00	96.00	98.00	96.00
5	135	792	391	9	8	99.00	97.75	98.88	97.99
6	0	771	371	29	29	96.38	92.75	96.38	92.75
6	45	765	374	26	35	95.63	93.50	96.71	91.44
6	90	790	387	13	10	98.75	96.75	98.38	97.48
6	135	794	389	11	6	99.25	97.25	98.63	98.48
7	0	773	367	33	27	96.63	91.75	95.91	93.15
7	45	756	365	35	44	94.50	91.25	95.58	89.24
7	90	790	392	8	10	98.75	98.00	99.00	97.51
7	135	791	388	12	9	98.88	97.00	98.51	97.73
8	0	772	368	32	28	96.50	92.00	96.02	92.93
8	45	767	362	38	33	95.88	90.50	95.28	91.65
8	90	788	384	11	12	98.50	97.22	98.62	96.97
8	135	793	388	12	7	99.13	97.00	98.51	98.23
9	0	779	374	26	21	97.38	93.50	96.77	94.68
9	45	762	359	41	38	95.25	89.75	94.89	90.43
9	90	789	389	11	11	98.63	97.25	98.63	97.25
9	135	794	389	11	6	99.25	97.25	98.63	98.48
10	0	776	371	29	24	97.00	92.75	96.40	93.92
10	45	762	365	35	38	95.25	91.25	95.61	90.57
10	90	793	382	18	7	99.13	95.50	97.78	98.20
10	135	793	387	13	7	99.13	96.75	98.39	98.22

ตาราง ข-6 ผลการประเมินประสิทธิภาพในการจำแนกชนิดของเซลล์โดยใช้โดยใช้ชุดค่าคุณลักษณะเด่นที่เหมาะสมในภาพที่ผ่านการสไลด์ด้วยหน้าต่างขนาด 256 × 256 พิกเซล

พารามิเตอร์ในการวิเคราะห์จีแอลซีเอ็ม		ผลบวกจริง	ผลลบจริง	ผลบวกเท็จ	ผลลบเท็จ	ค่าความไว	ค่าความจำเพาะ	ค่าพยากรณ์ผลบวก	ค่าพยากรณ์ผลลบ
ระยะห่าง	มุม								
1	0	760	368	32	40	95.00	92.00	95.96	90.20
1	45	762	366	34	38	95.25	91.50	95.73	90.59
1	90	772	371	29	28	96.50	92.75	96.38	92.98
1	135	782	372	28	18	97.75	93.00	96.54	95.38
2	0	761	368	32	39	95.13	92.00	95.96	90.42
2	45	762	368	32	38	95.25	92.00	95.97	90.64
2	90	783	384	16	17	97.88	96.00	98.00	95.76
2	135	782	384	16	18	97.75	96.00	97.99	95.52
3	0	768	364	36	32	96.00	91.00	95.52	91.92
3	45	773	374	26	27	96.63	93.50	96.75	93.27
3	90	784	388	12	16	98.00	97.00	98.49	96.04
3	135	785	389	11	15	98.13	97.25	98.62	96.29
4	0	772	375	25	28	96.50	93.75	96.86	93.05
4	45	760	368	32	40	95.00	92.00	95.96	90.20
4	90	787	393	7	13	98.38	98.25	99.12	96.80
4	135	792	386	14	8	99.00	96.50	98.26	97.97
5	0	770	371	29	30	96.25	92.75	96.37	92.52
5	45	754	374	26	46	94.25	93.50	96.67	89.05
5	90	784	388	12	16	98.00	97.00	98.49	96.04
5	135	792	390	10	8	99.00	97.50	98.75	97.99
6	0	768	371	29	32	96.00	92.75	96.36	92.06
6	45	763	366	34	37	95.38	91.50	95.73	90.82
6	90	789	389	11	11	98.63	97.25	98.63	97.25
6	135	794	389	11	6	99.25	97.25	98.63	98.48
7	0	763	371	29	37	95.38	92.75	96.34	90.93
7	45	767	367	33	33	95.88	91.75	95.88	91.75
7	90	789	391	9	11	98.63	97.75	98.87	97.26
7	135	792	388	12	8	99.00	97.00	98.51	97.98
8	0	767	377	23	33	95.88	94.25	97.09	91.95
8	45	771	355	45	29	96.38	88.75	94.49	92.45
8	90	789	389	11	11	98.63	97.25	98.63	97.25
8	135	793	388	12	7	99.13	97.00	98.51	98.23
9	0	777	377	23	23	97.13	94.25	97.13	94.25
9	45	773	357	43	27	96.63	89.25	94.73	92.97
9	90	789	389	11	11	98.63	97.25	98.63	97.25
9	135	794	389	11	6	99.25	97.25	98.63	98.48
10	0	771	374	26	29	96.38	93.50	96.74	92.80
10	45	760	360	40	40	95.00	90.00	95.00	90.00
10	90	790	388	12	10	98.75	97.00	98.50	97.49
10	135	793	387	13	7	99.13	96.75	98.39	98.22

ภาคผนวก ค

ผลการประเมินประสิทธิภาพบนภาพต้นฉบับ

ตาราง ก-1 ผลการประเมินประสิทธิภาพของระบบนับจำนวนนิวเคลียสเซลล์มะเร็งเรื้อรังผ่านมบนภาพถ่ายต้นฉบับ  
จำนวน 28 ภาพ

ภาพต้นฉบับ		Non-texture				GLCM window : 128 × 128				GLCM window : 256 × 256			
		TP	FP	FN	เวลา	TP	FP	FN	เวลา	TP	FP	FN	เวลา
ภาพที่ถูกนำไปสอนระบบ	T51-1549A12	227	49	51	30.92	212	21	66	62.28	216	21	62	64.31
	T51-1549A17	205	36	31	27.21	201	29	35	55.54	201	24	35	55.93
	T51-1549A21	160	69	63	27.00	145	38	78	53.09	142	34	81	53.87
	T51-1549A23	227	60	89	32.22	209	27	107	64.89	222	43	94	67.71
	T51-1549A36	191	145	44	33.54	151	14	82	68.20	169	49	65	72.14
	T51-1549A37	207	74	55	29.68	185	13	77	59.36	191	23	71	61.54
	T51-1549A39	169	65	62	27.76	144	22	87	51.92	159	38	72	54.28
	T51-1549A73	117	35	70	20.74	114	23	73	37.61	107	9	80	38.08
	T51-1549A78	163	156	58	29.37	141	44	75	63.12	145	69	72	65.46
	T51-1549A92	131	50	66	21.88	124	31	73	41.94	127	26	70	42.77
	T51-1549A97	111	78	37	21.51	110	48	38	41.88	104	51	43	43.00
	T51-1549A3	231	51	92	30.50	125	11	198	55.91	188	16	135	61.07
	T51-1549A4	207	62	59	29.84	142	12	123	55.59	182	27	84	59.98
	T51-1549A5	147	42	53	23.06	129	13	70	42.86	129	20	70	43.85
	T51-1549A9	246	55	59	33.05	224	10	81	65.37	240	37	65	68.30
	T51-1549A40	202	70	45	28.98	187	17	60	57.95	202	40	45	60.34
	T51-1549A-78	211	83	43	31.69	189	26	65	62.65	205	31	49	65.70
รวม	3152	1180	977		2732	399	1388		2929	558	1193		
ภาพที่ไม่ถูกนำไปสอนระบบ	T51-1549A1	188	36	47	31.26	66	3	169	45.33	107	7	128	48.49
	T51-1549A2	214	68	36	32.32	94	11	156	56.8	149	16	100	61.19
	T51-1549A41	137	95	46	25.57	119	23	64	49.51	134	40	49	52.05
	T51-1549A42	183	88	38	30.38	158	28	63	59.18	179	62	42	62.67
	T51-1549A44	240	58	52	32.47	83	5	205	57.03	180	24	109	64.42
	T51-1549A53	222	101	116	32.32	190	22	148	67.32	191	34	147	70.03
	T51-1549A54	213	61	84	29.95	184	16	112	59.71	198	32	98	62.79
	T51-1549A59	163	23	15	22.49	72	9	106	39.29	135	20	43	43.84
	T51-1549A60	192	83	39	29.05	121	20	109	56.17	176	46	55	61.09
	T51-1549A-134	215	67	44	29.69	74	19	184	53.34	147	53	112	60.17
	T51-1549A-135	195	91	35	27.62	47	8	183	51.36	90	18	140	55.79
	รวม	2162	771	552		1208	164	1499		1686	352	1023	

**ภาคผนวก ง**

**Proceedings' title**

**“Preliminary Results of Breast Cancer Cell  
Classifying Based on Gray-Level Co-occurrence Matrix”**

**Presented in**

**The 6th Biomedical Engineering International Conference (BMEiCON 2013)**

**23-25 October 2013 at Krabi, Thailand**

# Preliminary Results of Breast Cancer Cell Classifying Based on Gray-Level Co-occurrence Matrix

A. Markkongkeaw<sup>1,\*</sup>, A. Phinyomark<sup>1</sup>, P. Boonyapiphat<sup>2</sup>, P. Phukpattaranont<sup>1</sup>

<sup>1</sup>Department of Electrical Engineering, Prince of Songkla University

<sup>2</sup>Department of Pathology, Prince of Songkla University  
Songkhla, Thailand

\*attagorn.m@outlook.com, angkoon.p@hotmail.com, pleamjit.b@psu.ac.th, ponchai.p@psu.ac.th

**Abstract**— This study proposes and appraise a gray level co-occurrence matrix (GLCM) for extracting the feature of cell regions in microscopic image into four region types: positive cancer cell, negative cancer cell, lymphocyte and stromal cell. The classification task uses decision tree with cross validation. To give a high classification performance, the main focus of interest is feature extraction task. Twenty-two texture features of GLCM have used to analysis images at four directions and six scales of gray-level quantization. A set of these texture features is used in 2045 images for training and testing. The result shows that the classification accuracy obtained from decision tree is 95.21%. It is demonstrated that the proposed GLCM texture features and decision tree can classify the histological structures in microscopic image and can be applied to improve and to develop an accurate cell counting of computer-aided diagnosis system for breast cancer prognosis.

**Index Terms**— Gray-level co-occurrence matrix, texture features, breast cancer, estrogen, immunohistochemistry, microscopic image

## I. INTRODUCTION

Cancer is expected to become the leading cause of death over the next few decades. Based on statistics from the World Health Organization (WHO), cancer accounted for 13% for all deaths in the world in 2008. Deaths caused by cancer are expected to increase in the future [1]. Of all the known cancers, breast cancer (BC) is a major concern among women. It is the first-most common and leading cause of cancer deaths among women. According to published statistics, BC has become a major health problem in both developed and developing countries. The rate of the new patient has increased in recent years.

The pathological diagnosis of BC can use the suspect tissue from fine-needle aspiration biopsy achieved by the pathologist visualization. Sample tissues are strained by an immunohistochemistry (IHC) technique. It is used for detecting in situ a tissue antigen by a specific antibody [2][3]. An antigen-antibody complex is visualized by the color development of DAB dye and can be seen by light microscope. In BC cell image, estrogen receptor (ER) and progesterone receptor (PR) are protein markers in nuclear membrane as an antigen. In the microscopic image of BC, the brown nuclear

stain is represented as positive cancer cell (PC). The blue nuclear stain has three types of cell. There are negative cancer cell (NC), lymphocyte cell (LC), and stromal cell (SC) respectively. Pathologist has evaluated these cells as the ratio of the total PCs to total number of cancer cells (PC+NC) for hormone therapy [4]. This percentage visualization is time consuming, costly and tedious. In addition, the results are uncertain because it depends on the experience of the pathologist. Therefore, there are researches which developed a computer-aided diagnosis (CAD) system for BC cell counting by using many algorithms in image analysis.

The CAD system for the microscopic image of BC analysis consists of two main steps: feature extraction and classification. The feature extraction presented in this study is co-occurrence of gray-level image. In order to yield a high classification performance, feature extraction is selected to suit the application. Generally, feature representation of images is extracted from color, shape and texture information of them. Previously, color and shape features are widely used to classify cells in the histological structure (HS) in microscopic image of BC. However, the texture feature is not widely used, so it is explored this study.

Texture feature is a function of the spatial distribution in pixel intensity. Many methods of texture analysis have been proposed to classify in medical images such as Gabor filter, gray level co-occurrence matrices (GLCM) and Laws' texture energy. Additionally, microscopic images of BC have a complex structure which can be resolved by Haralick's texture features [5]. This study separates the region type of cells in microscopic images using twenty-two features of GLCM. Then, the suitable combination of these features is evaluated using a decision tree classifier. However, the raw two-dimensional microscopic images of BC are manually cropped to produce the image with smaller size and converted into gray scale images before entering the process to check the ability of texture features and the decision tree classifier.

## II. MATERIALS AND METHODS

### A. Microscopic dataset(MDS)

Seventeen microscopic images in full size were acquired from digital camera DXM1200F (Nikon Instech Co., Ltd., Japan). It was connected to advance research microscope Eclipse 80i (Nikon Instech Co., Ltd., Japan). Raw images were saved into RGB 24-bits and JPEG format with resolution of 3600 x 2880 pixels.

Pathologist manually selected the region of interest (ROI) from seventeen raw images with magnification of 400x. It was cropped and saved as JPEG format with resolution of 256 x 256 pixels for later analysis. Fig. 1(a) showed an RGB image of microscopic image for breast cancer cell. Fig. 1(b) indicated an ROI for each type of cell. Thereafter, the extracted ROI was converted to gray-level image and normalized to the gray scales range [0, 255]. The MDS consisted of 2045 images from patients, including 1514 PCs, 207 LCs, 215 SCs, 109 NCs.

### B. Scaling of Gray-Level

The number of gray-levels is an important factor for the GLCM computation to present a set of textures for reducing the misclassification of cell type. The more levels included in the GLCM, the more is the computational cost of the texture statistics.

A simple way to transform the cell image to its quantized version can be expressed as the linear combination

$$J = \lfloor I \cdot K_1 + K_2 \rfloor, \quad (1)$$

where  $I$  is the input image,  $J$  is the quantized image,  $\lfloor \cdot \rfloor$  is the floor function, and  $K_1$  and  $K_2$  are coefficients defined as

$$K_1 = \frac{L-1}{G_{\max} - G_{\min}}, \quad (2)$$

$$K_2 = 1 - K_1 \cdot G_{\min}, \quad (3)$$

where  $L$  is the desired number of quantization levels,  $G_{\min}$  and  $G_{\max}$  are minimum and maximum of gray levels in the image, respectively [6].

### C. GLCM Texture Features

GLCM is the texture feature was proposed by Haralick in the 1970s. It is well-established robust statistical tool for extracting secondary-order of texture information from image. The GLCM presents the joint frequencies of all pairwise combinations of gray levels  $i$  and  $j$  in a specified direction  $\theta$  and specified distance  $d$  from each other [7-10]. In this paper, 22 features are generated from six gray-level scaling ( $L = 8, 16, 32, 64, 128, 256$ ), four directions ( $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ ) and one distance ( $d = 1$ ) for each ROI image. The GLCM can be defined as

$$C(i, j) = \left\| \left\| \left\{ \left\{ (x_1, y_1), (x_2, y_2) \right\} \right\} \begin{cases} x_2 - x_1 = d \cos \theta \\ y_2 - y_1 = d \sin \theta \\ I(x_1, y_1) = i \\ I(x_2, y_2) = j \end{cases} \right\| \right\|, \quad (4)$$

where  $(x_1, y_1)$  and  $(x_2, y_2)$  are pixels in the ROI,  $I(\cdot)$  is gray-level of pixels, and  $\| \cdot \|$  is the number of the pixel pairs that satisfy the conditions. Thereupon, each co-occurrence matrix is normalized by sum of all elements in matrix. Finally twenty-two features are generated from each matrix. Especially, the features studied are autocorrelation, contrast, correlation I, correlation II, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity I, homogeneity II, maximum probability, sum of squares, sum average, sum entropy, sum variance, difference variance, difference entropy, information measure of correlation I, information measure of correlation II, inverse difference normalized and inverse difference moment normalized. Equations of all features were showed in table I and II.

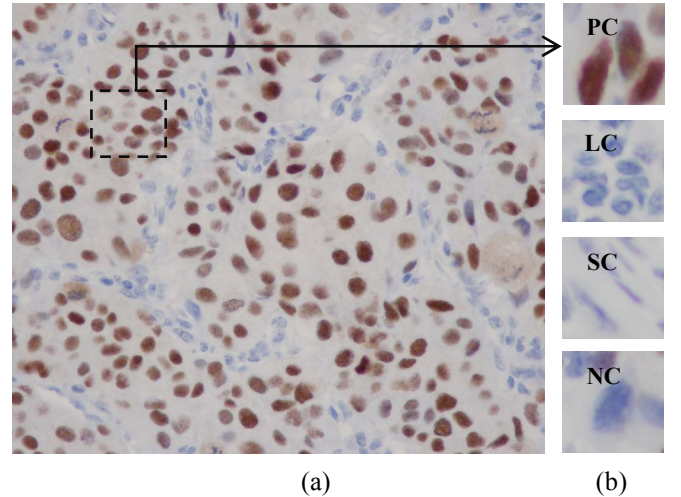


Fig. 1. Microscopic image. (a) Full image size 3600x2880 pixels. (b) The ROI for each cell type is captured with a resolution of 256x256 pixels.

### D. Classification by tree

The decision tree classifier is a predictive modeling that decides the dependent variable of a new sample founded on different characteristic values of the available data. The internal nodes of a decision tree stand for the different characteristics. The branches between the nodes are the possible values that these characteristic can have in the observed samples. The whole terminal nodes tell us the classification of the dependent variable [11][12][13]. The predicted attribute is known as the dependent variable, since its value depends upon, or is decided by, the value of all the other attributes.



TABLE I. EXPRESSION USED FOR GLCM COMPUTING

Expression	Meaning
$p(i, j)$	$(i, j)$ th entry of the co-occurrence probability matrix.
$L$	Gray-level quantization.
$\nu$	Mean value of $p(i, j)$
$p_x(i)$	$\sum_{j=1}^L p(i, j)$
$p_y(i)$	$\sum_{i=1}^L p(i, j)$
$\mu_x$	$\sum_i \sum_j i \cdot p(i, j)$
$\mu_y$	$\sum_i \sum_j j \cdot p(i, j)$
$\sigma_x^2$	$\sum_i \sum_j (i - \mu_x)^2 \cdot p(i, j)$
$\sigma_y^2$	$\sum_i \sum_j (j - \mu_y)^2 \cdot p(i, j)$
$p_{x+y}(k)$	$\sum_{i=1, i+j=k}^L \sum_{j=1}^L p(i, j)$ , $k = 2, 3, \dots, 2L$
$p_{x-y}(k)$	$\sum_{i=1,  i+j =k}^L \sum_{j=1}^L p(i, j)$ , $k = 0, 1, \dots, L-1$
$HX$	$-\sum_i p_x(i) \cdot \log(p_x(i))$
$HY$	$-\sum_i p_y(i) \cdot \log(p_y(i))$
$HXY$	$-\sum_i \sum_j p(i, j) \cdot \log(p(i, j))$
$HXY1$	$-\sum_i \sum_j p(i, j) \cdot \log(p_x(i)p_y(j))$
$HXY2$	$-\sum_i \sum_j p_x(i)p_y(j) \cdot \log(p_x(i)p_y(j))$

### III. RESULTS AND DISCUSSION

The decision tree classifier is a predictive modeling that decides the dependent variable of a new sample founded on different characteristic values of the available data. The internal nodes of a decision tree stand for the different characteristics. The branches between the nodes are the possible values that these characteristic can have in the observed samples. The whole terminal nodes tell us the classification of the dependent variable [11][12][13]. The predicted attribute is known as the dependent variable, since its value depends upon, or is decided by, the value of all the other attributes.

Cross validation is one way to enhance over the holdout method. The data set have been divided to  $k$  subsets, and holdout method is repeated  $k$  times. The overall performance of the diagnostic algorithm can evaluate accuracy of classification. We use the k-fold cross validation to estimate the performance of the feature extraction and classification algorithm.

TABLE II. TEXTURE FEATURES FROM GLCM

Feature	Equation
Autocorrelation	$\sum_i \sum_j (i \cdot j) p(i, j)$
Contrast	$\sum_i \sum_j  i - j ^2 p(i, j)$
Correlation I	$\sum_i \sum_j \frac{(i - \mu_x)(j - \mu_y) p(i, j)}{\sigma_x \sigma_y}$
Correlation II	$\sum_i \sum_j \frac{(i \cdot j) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Cluster prominence	$\sum_i \sum_j (i + j - \mu_x - \mu_y)^4 p(i, j)$
Cluster shade	$\sum_i \sum_j (i + j - \mu_x - \mu_y)^3 p(i, j)$
Dissimilarity	$\sum_i \sum_j  i - j  \cdot p(i, j)$
Energy	$\sum_i \sum_j p(i, j)^2$
Entropy	$-\sum_i \sum_j p(i, j) \cdot \log(p(i, j))$
Homogeneity I	$\sum_i \sum_j \frac{p(i, j)}{1 +  i - j }$
Homogeneity II	$\sum_i \sum_j \frac{p(i, j)}{1 +  i - j ^2}$
Maximum probability	$\max_{i, j} p(i, j)$
Sum of squares	$\sum_i \sum_j (i - \nu)^2 p(i, j)$
Sum average	$\sum_{i=2}^{2L} i \cdot p_{x+y}(i)$
Sum entropy	$-\sum_{i=2}^{2L} p_{x+y}(i) \cdot \log(p_{x+y}(i))$
Sum variance	$\sum_{i=2}^{2L} (i - \text{Sumentropy})^2 \cdot p_{x+y}(i)$
Difference variance	$\sum_{i=0}^{L-1} i^2 \cdot p_{x-y}(i)$
Difference entropy	$-\sum_{i=0}^{L-1} p_{x-y}(i) \cdot \log(p_{x-y}(i))$
Information measure of correlation I	$\frac{HXY - HXY1}{\max(HX, HY)}$
Information measure of correlation II	$\sqrt{1 - e^{-2(HXY2 - HXY)}}$
Inverse difference normalized	$\sum_i \sum_j \frac{p(i, j)}{1 +  i - j ^2} / L$
Inverse difference moment normalized	$\sum_i \sum_j \frac{p(i, j)}{1 + (i - j)^2} / L$

The result estimated from 2045 images and classified by decision tree with 10-fold cross validation for training and testing data set. The result from the classification of MDS

compared with the expert perception is shown in Table III. The best precision score for GLCM based methods was achieved in direction of computation and quantization of gray-level image. Accuracy depends on the resolution, i.e., the high resolution gives the high accuracy. The direction of 135° of GLCM gives a best result. The maximum classification accuracy is 95.21%.

#### IV. CONCLUSION

This work presented a study of GLCM texture feature as a function of gray-level for classifying region type of cancer cell in microscopic image. Twenty-two texture features computed from co-occurrence matrices were assessed, regarding four directions and one distance for MDS. ROI was located and manually cropped by pathologist with size 256 × 256 pixels, 2045 images in MDS have four types of cell region. Texture features are used as input tree in classification process. Decision tree was used to be a classifier and used the texture feature of GLCM for input material in the classification process to classify region type of cell in microscopic image. The methodology presented effectively type of cell region in microscopic image of breast cancer cell. The accuracy for decision tree with cross validation on MDS is shown in the result.

This work confirmed that GLCM feature are useful when considering automatic classification tools to aid pathologist to identify region of cancer cell in microscopic image from breast biopsy. However, the correct percentage estimated from this study is compared with the expert perception to improve and apply to the automatic breast cancer cell counting.

Moreover, other parameters in texture features will be done in this system and the feature selection based on some selecting algorithms will be decided for these proposed texture features in the future.

TABLE III. ACCURACY OF CLASSIFICATION

Gray scaling	Direction			
	0	45	90	135
8	93.01	92.91	92.81	93.25
16	94.38	94.23	93.55	94.72
32	93.59	93.45	94.43	93.59
64	93.99	93.69	94.28	94.28
128	94.08	93.99	94.47	95.01
256	93.06	94.28	94.47	95.21

#### ACKNOWLEDGMENT

This work was supported by NECTEC-PSU Center of Excellence for Rehabilitation Engineering, Faculty of Engineering and Department of Pathology, Faculty of Medicine, Prince of Songkla University.

#### REFERENCES

- [1] "WHO | 10 facts about cancer," WHO. <http://www.who.int/features/factfiles/cancer/en/>. (Accessed on: 19<sup>th</sup> Jun., 2013).
- [2] R. Mofidi, R. Walsh, P.F. Ridgway, T. Crotty, E.W. McDermott, T.V. Keaveny, M.J. Duffy, A.D.K. Hill, and N. O'Higgins, "Objective measurement of breast cancer oestrogen receptor status through digital image analysis," *Eur. J. Surg. Oncol. London*, vol. 29, pp. 20–24, February 2003.
- [3] Y. Hatanaka, K. Hashizume, K. Nitta, T. Kato, I. Itoh, and Y. Tani, "Cytometrical image analysis for immunohistochemical hormone receptor status in breast carcinomas," *Pathol. Int.*, vol. 53, pp. 693–699, 2003.
- [4] P. Phukpattaranont and P. Boonyaphiphat, "Computer-aided analysis of nuclear stained breast cancer cell images," in *Inter. Conf. on Elec. Eng./Elec. Com., Tele. and Info. Tech.*, 2008, vol. 1, pp. 485–488.
- [5] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man. Cyb.* vol. SMC-3, pp. 610–621, November 1973.
- [6] W. Gomez, W. C. A. Pereira, and A. F. C. Infantosi, "Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound," *IEEE Trans. Med. Imaging*, vol. 31, pp. 1889–1899, October 2012.
- [7] M. M. Mokji and S. A. R. Abu Bakar, "Gray Level Co-Occurrence Matrix Computation Based On Haar Wavelet," in *Computer Graphics, Imaging and Visualisation, 2007. CGIV '07, 2007*, pp. 273–279.
- [8] S. Park, B. Kim, J. Lee, J.M. Goo, and Y.-G. Shin, "GGO nodule volume-preserving nonrigid lung registration using glcm texture analysis," *IEEE Trans. Bio-Med. Eng.*, vol. 58, pp. 2885–2894, October 2011.
- [9] G. Xian, "An identification method of malignant and benign liver tumors from ultrasonography based on glem texture features and fuzzy svm," *Expert Syst. Appl.*, vol. 37, pp. 6737–6741, 2010.
- [10] A. N. Karahaliou, I. S. Boniatis, S. G. Skiadopoulos, F. N. Sakellaropoulos, N. S. Arikidis, E. A. Likaki, G. S. Panayiotakis, and L. I. Costaridou, "Breast cancer diagnosis: analyzing texture of tissue surrounding microcalcifications," *IEEE Trans. Inf. Technol. B.*, vol. 12, pp. 731–738, November. 2008.
- [11] Y. Hu, D. Wu, and A. Nucci, "Fuzzy-clustering-based decision tree approach for large population speaker identification," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, pp. 762–774, 2013.
- [12] N. Manwani and P. S. Sastry, "Geometric Decision Tree," *IEEE Trans. Syst. Man Cy B.*, vol. 42, pp. 181–192, 2012.
- [13] W. Reichl and W. Chou, "Robust decision tree state tying for continuous speech recognition," *IEEE Trans. Speech. Audi. P.*, vol. 8, pp. 555–566, 2000.

## ประวัติผู้เขียน

ชื่อ สกุล	นายอรรถกร มากคงแก้ว	
รหัสประจำตัวนักศึกษา	5610120044	
วุฒิการศึกษา		
วุฒิ	ชื่อสถาบัน	ปีที่สำเร็จการศึกษา
วิศวกรรมศาสตรบัณฑิต (วิศวกรรมชีวการแพทย์)	มหาวิทยาลัยสงขลานครินทร์	2556

### ทุนการศึกษา (ที่ได้รับในระหว่างการศึกษา)

ทุนโครงการปริญญาตรี-โท 5 ปี คณะวิศวกรรมศาสตร์มหาวิทยาลัยสงขลานครินทร์

### การตีพิมพ์ในที่ประชุมวิชาการที่มีการรวมเล่มระดับนานาชาติ (ขณะศึกษาปริญญาโท)

A. Markkongkeaw, A. Phinyomark, P. Boonyapiphat, and P. Phukpattaranont, "Preliminary Results of Breast Cancer Cell Classifying Based on Gray-Level Co-occurrence Matrix," in *Proceedings of the 6th Biomedical Engineering International Conference (BMEiCON 2013)*, Thailand, Oct 23-25, 2013.

### รางวัลที่ได้รับ

นายอรรถกร มากคงแก้ว ได้รับรางวัล Gold Award: Oral Presentation ในการประชุม วิชาการระดับปริญญาตรีทางวิทยาศาสตร์ วิศวกรรมศาสตร์ และทรัพยากร เมื่อวันที่ 26 มกราคม 2556 ณ มหาวิทยาลัยวลัยลักษณ์

นายอรรถกร มากคงแก้ว ได้รับรางวัลที่ 3 ระดับปริญญาตรี สาขาวิทยาศาสตร์และเทคโนโลยี จากการเข้าร่วมแข่งขันการประกวดรางวัลนวัตกรรมสงขลานครินทร์ ผลงานในหัวข้อ "ระบบคอมพิวเตอร์สำหรับการวิเคราะห์ในเชิงพื้นที่ผิวในภาพเซลล์มะเร็งเต้านม" เมื่อวันที่ 17 มิถุนายน 2556 เนื่องในงานวันนักวิจัยและนวัตกรรม ม.อ. ครั้งที่ 7 ณ ศูนย์ประชุมนานาชาติฉลองสิริราชสมบัติครบ 60 ปี มหาวิทยาลัยสงขลานครินทร์

นายอรรถกร มากคงแก้ว ได้รับรางวัลระดับดี กลุ่มเรื่องเพื่อการพัฒนาการแพทย์และสุขภาพ ในโครงการส่งเสริมการวิจัยสิ่งประดิษฐ์ และเครือข่ายนวัตกรรมอุดมศึกษาประกวดผลงานประดิษฐ์คิดค้นเพื่อการพัฒนาการแพทย์และสุขภาพ (เขตภาคใต้) ประจำปี 2556 เมื่อวันที่ 14 กรกฎาคม 2556 ณ มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย

นายอรรถกร มากคงแก้ว และรองศาสตราจารย์ ดร. พรชัย พุกภัยภัทรานนต์ ได้รับรางวัลชนะเลิศ ผลงานประเภทวิชาการ จากการเข้าร่วมแข่งขันการประกวดนวัตกรรมและสิ่งประดิษฐ์ ระดับปริญญาตรี ผลงานในหัวข้อ “ระบบคอมพิวเตอร์สำหรับการวิเคราะห์ในเชิงพื้นที่ผิวในภาพเซลล์มะเร็งเต้านม” เมื่อวันที่ 17 สิงหาคม 2556 คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์

นายอรรถกร มากคงแก้ว และรองศาสตราจารย์ ดร. พรชัย พุกภัยภัทรานนต์ ได้รับรางวัลชมเชย ผลงานประเภทนำเสนอยอดเยี่ยม จากการเข้าร่วมแข่งขันการประกวดนวัตกรรมและสิ่งประดิษฐ์ ระดับปริญญาตรี ผลงานในหัวข้อ “ระบบคอมพิวเตอร์สำหรับการวิเคราะห์ในเชิงพื้นที่ผิวในภาพเซลล์มะเร็งเต้านม” เมื่อวันที่ 17 สิงหาคม 2556 คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์

นายอรรถกร มากคงแก้ว และรองศาสตราจารย์ ดร. พรชัย พุกภัยภัทรานนต์ ได้รับรางวัลชนะเลิศ ผลงานประเภทคะแนนรวมสูงสุด จากการเข้าร่วมแข่งขันการประกวดนวัตกรรมและสิ่งประดิษฐ์ ระดับปริญญาตรี ผลงานในหัวข้อ “ระบบคอมพิวเตอร์สำหรับการวิเคราะห์ในเชิงพื้นที่ผิวในภาพเซลล์มะเร็งเต้านม” เมื่อวันที่ 17 สิงหาคม 2556 คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์