



การตัดแยกภาพเซลล์มะเร็งเต้านมแบบอัตโนมัติ

Automated Breast Cancer Cell Image Segmentation

ธนพร พันธุ์เมธาฤทธิ

Thanaporn Phanmetharith

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์
มหาวิทยาลัยสงขลานครินทร์

**A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Prince of Songkla University**

2553

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์

ชื่อวิทยานิพนธ์ การตัดแยกภาพเซลล์มะเร็งด้านมแบบอัตโนมัติ
ผู้เขียน นางสาวธนพร พันธุ์เมธาฤทธิ์
สาขาวิชา วิทยาการคอมพิวเตอร์

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

คณะกรรมการสอบ

.....
(ดร. ลัดดา ปรีชาวีรกุล)

.....ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.กฤษณะ ชินสาร)

อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. ศิริรัตน์ วัฒนชัยโยบล)

.....
(ผู้ช่วยศาสตราจารย์ ดร. พรชัย พฤษย์ภัทรานนท์)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. พรชัย พฤษย์ภัทรานนท์)

.....กรรมการ
(ดร.ลัดดา ปรีชาวีรกุล)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้รับวิทยานิพนธ์ฉบับนี้ เป็นส่วน
หนึ่งของการศึกษา ตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

.....

(รองศาสตราจารย์ ดร. เกริกชัย ทองหนู)

คณบดีบัณฑิตวิทยาลัย

ชื่อวิทยานิพนธ์	การตัดแยกภาพเซลล์มะเร็งเต้านมแบบอัตโนมัติ
ผู้เขียน	นางสาวธนพร พันธุ์เมธาฤทธิ
สาขาวิชา	วิทยาการคอมพิวเตอร์
ปีการศึกษา	2552

บทคัดย่อ

ในปัจจุบันมีผู้ป่วยโรคมะเร็งเต้านมสูงขึ้น เนื่องจากการใช้ชีวิตที่เปลี่ยนไป สิ่งอุปโภคบริโภคที่เปลี่ยนไป ทางการแพทย์จึงต้องผลิตตัวยาในการรักษาโรคมะเร็งเต้านมมากขึ้นและต้องมีวิธีการตรวจสอบว่าผู้ป่วยรายใดต้องการยาหรือไม่ วิทยานิพนธ์นี้จึงนำเสนอวิธีการตัดแยกภาพเซลล์มะเร็งเต้านมแบบอัตโนมัติ หรือ Automated Breast Cancer Cell Image Segmentation (ABC_CIS) เพื่อช่วยนับจำนวนเซลล์มะเร็งเต้านมจากภาพเนื้อเยื่อเซลล์มะเร็งเต้านมที่ผ่านการย้อมสีแล้ว โดยไม่ต้องให้แพทย์ระบุส่วนต่างๆของภาพ แต่ระบบสามารถตรวจจับเซลล์มะเร็งในภาพได้เอง โดยใช้ความรู้เรื่องสีในการพิจารณาแยกเซลล์ออกจากพื้นหลังของภาพและได้ตรวจสอบความถูกต้อง โดยมีการเปรียบเทียบผลการนับเซลล์ของระบบกับผลการนับเซลล์ของผู้เชี่ยวชาญ ระบบการตัดแยกภาพเซลล์มะเร็งเต้านมแบบอัตโนมัติ ประกอบด้วยขั้นตอนหลักได้แก่ 1) การเตรียมภาพเซลล์มะเร็งเต้านม 2) การตัดแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังของภาพ 3) การกำจัดสิ่งรบกวนออกจากภาพ 4) การนับเซลล์มะเร็งเต้านมในภาพ 5) การเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ ซึ่งวิธีการตัดแยกภาพเซลล์มะเร็งเต้านมแบบอัตโนมัติจะช่วยเพิ่มความสะดวกและรวดเร็วต่อแพทย์ในการวินิจฉัยการให้ยาแก่ผู้ป่วยมะเร็งเต้านม ในงานวิจัยนี้ได้หาค่าเฉลี่ยจากการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญได้แก่ ค่าเฉลี่ยของค่าความไว (Sensitivity, S) และค่าพยากรณ์ผลบวก (Predictive Positive Value, PPV) โดยใช้วิธี K-Means และ K-Median เป็น 62.58 % , 92.89 % และ 73.65 % , 82.37 % ตามลำดับ

Thesis Title	Automated Breast Cancer Cell Image Segmentation
Author	Miss Thanaporn Phanmetharith
Major Program	Computer Science
Academic Year	2009

ABSTRACT

Nowadays, there are many breast cancer patients because of the change in their lifestyle has been changed. Medical technology needs to produce medicine for breast cancer treatment and has a method to determine appropriate type of medicine for the patients. This thesis proposes an Automatic Breast Cancer Cell Image Segmentation Approach or ABC_CIS to count a number of breast cancer cells from tissue breast cancer cells images through the tint without having to specify different parts of the medical images from the experts. The system can detect cancer cells automatically using color model in images to segment cells from the background image and check the accuracy. Five main processes in this system are 1) preparing breast cancer images 2) segmenting breast cancer cells from the background image 3) removing noise from image 4) counting cells in image and 5) comparing the cell counted with the experts. A result is compared with the cells counted from the experts. We found that the automated breast cancer cell image segmentation is able to accelerate doctor's diagnosis quickly. The sensitivity and specificity of the proposed system tested using *K*-Means and *K*-Median are 62.58 %, 92.89 % and 73.65 %, 82.37 %, respectively.

สารบัญ

	หน้า
สารบัญ.....	(6)
รายการตาราง.....	(8)
รายการภาพประกอบ.....	(9)
บทที่ 1 บทนำ.....	1
1.1 การตรวจเอกสาร.....	2
1.2 วัตถุประสงค์ของโครงการ.....	7
1.3 ขอบเขตการดำเนินงาน.....	7
1.4 ขั้นตอนการดำเนินงานและระยะเวลาการดำเนินงาน.....	7
1.5 สถานที่และเครื่องมือที่ใช้ในงานวิจัย.....	9
1.6 ประโยชน์ที่คาดว่าจะได้รับ.....	10
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....	11
2.1 เต้านมของเพศหญิง (Female Breast).....	11
2.2 เซลล์มะเร็งเต้านม (Breast Cancer Cell).....	12
2.3 หลักการของการรักษามะเร็งเต้านมด้วยวิธีฮอร์โมน.....	13
2.4 การได้มาของภาพดิจิทัล.....	16
2.5 Morphological Operations.....	17
2.6 Marker-Controlled Watershed.....	24
2.7 สถิติพื้นฐาน (Statistics).....	27
2.8 การเรียนรู้ของเครื่อง (Machine Learning).....	28
2.9 การแบ่งแยกภาพ (Image Segmentation).....	32
2.10 ระบบสี (Color Model).....	32
2.11 รูปแบบไฟล์ภาพ (Image File Formats).....	35
บทที่ 3 ระบบการตัดแยกภาพเซลล์มะเร็งเต้านมแบบอัตโนมัติ.....	38
3.1 ขั้นตอนการเตรียมภาพเซลล์มะเร็งเต้านม.....	40
3.2 ขั้นตอนการตัดแยกภาพเซลล์มะเร็งเต้านม.....	42
3.3 ขั้นตอนการกำจัดสิ่งรบกวนในภาพ.....	44
3.4 ขั้นตอนการนับเซลล์มะเร็งเต้านม.....	45

สารบัญ (ต่อ)

	หน้า
3.5 ขั้นตอนการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ.....	46
บทที่ 4 ผลการทดลองและวิจารณ์.....	51
4.1 ผลการทดลอง.....	51
4.2 สรุปผลการทดลอง	59
บทที่ 5 บทสรุปและข้อเสนอแนะ.....	70
5.1 สรุปผลการวิจัย.....	70
5.2 ปัญหาและอุปสรรค.....	71
5.3 ข้อเสนอแนะ.....	72
บรรณานุกรม.....	73
ภาคผนวก.....	76
ก ผลการทดลองโดยใช้ K-Means.....	77
ข คู่มือผู้ใช้ระบบ.....	85
ค ผลงานตีพิมพ์ในการประชุมวิชาการ NCSEC 2008.....	91
ประวัติผู้เขียน.....	99

รายการตาราง

ตาราง	หน้า
ตาราง 1.1 ตารางแสดงระยะเวลาดำเนินการ.....	9
ตาราง 2.1 รูปแบบไฟล์ภาพ.....	37
ตาราง 3.1 Confusion Matrix.....	47
ตาราง 4.1 ผลการแยกเซลล์ กำจัดสิ่งรบกวน นับเซลล์และเปรียบเทียบกับผู้เชี่ยวชาญโดยใช้ K-Median ในขั้นตอนการตัดแยกภาพเซลล์มะเร็งเต้านม.....	52
ตาราง 4.2 ตัวเลขผลการเปรียบเทียบการนับเซลล์กับผู้เชี่ยวชาญโดยใช้ K-Mean และ K-Median ในขั้นตอนการตัดแยกภาพเซลล์มะเร็งเต้านม.....	61
ตาราง 4.3 ค่าเฉลี่ยของผลการเปรียบเทียบกับผู้เชี่ยวชาญ.....	67
ตาราง 4.4 ค่าเฉลี่ยของจำนวนเซลล์ส่วนที่ผู้เชี่ยวชาญจะต้องเพิ่มหรือนำออกจากโปรแกรม.....	67
ตาราง ก.1 ผลการแยกเซลล์ กำจัดสิ่งรบกวน นับเซลล์และเปรียบเทียบกับผู้เชี่ยวชาญโดยใช้ K-Means ในขั้นตอนการตัดแยกภาพเซลล์มะเร็งเต้านม.....	77

รายการภาพประกอบ

ภาพประกอบ	หน้า
ภาพประกอบ 1.1 ตัวอย่างภาพเซลล์มะเร็งที่ถูกย้อมสี.....	2
ภาพประกอบ 2.1 ส่วนประกอบของเต้านม.....	11
ภาพประกอบ 2.2 เนื้อเยื่อมะเร็งเต้านม.....	12
ภาพประกอบ 2.3 ส่วนประกอบในเซลล์.....	15
ภาพประกอบ 2.4 ในเซลล์ที่มี Tamoxifen.....	15
ภาพประกอบ 2.5 ขั้นตอนการได้มาของภาพดิจิทัล.....	16
ภาพประกอบ 2.6 ตัวอย่างภาพเซลล์มะเร็งที่ถูกย้อมสี.....	17
ภาพประกอบ 2.7 ตัวอย่างผลการทำ Morphology.....	18
ภาพประกอบ 2.8 Structuring Element.....	19
ภาพประกอบ 2.9 วิธี Erosion.....	20
ภาพประกอบ 2.10 ตัวอย่าง Erosion.....	21
ภาพประกอบ 2.11 วิธี Dilation	22
ภาพประกอบ 2.12 ตัวอย่าง Dilation.....	22
ภาพประกอบ 2.13 การเติมเต็มพื้นที่ว่างในวัตถุ	23
ภาพประกอบ 2.14 การกำจัดวัตถุที่อยู่ติดขอบภาพ	23
ภาพประกอบ 2.15 ภาพที่ถูกแบ่งแยกหลังจากการใช้การดำเนินการตรรกะ OR.....	24
ภาพประกอบ 2.16 ผลภาพหลังจากการใส่ตัวเลขระบุเซลล์.....	25
ภาพประกอบ 2.17 รูปแบบของ Distance Transformation.....	26
ภาพประกอบ 2.18 การตัดแยกวัตถุที่ติดกันด้วย Watershed Segmentation.....	26
ภาพประกอบ 2.19 ตัวอย่างการหาค่าเฉลี่ยโดยมีข้อมูลที่ห่างจากข้อมูลที่มีส่วนใหญ่.....	27
ภาพประกอบ 2.20 ตัวอย่างการหาค่ามัธยฐานโดยมีข้อมูลที่ห่างจากข้อมูลที่มีส่วนใหญ่.....	28
ภาพประกอบ 2.21 ความแตกต่างของ Partitional Clustering และ Hierarchical Clustering.....	30
ภาพประกอบ 2.22 รูปแบบการเลือกค่าศูนย์กลางหรือตัวแทนกลุ่ม.....	31
ภาพประกอบ 2.23 RGB and CMYK Color Cube	33
ภาพประกอบ 2.24 กระบวนการดูซับสี (Subtractive Process)	34
ภาพประกอบ 2.25 RGB Color.....	35
ภาพประกอบ 2.26 CMYK Color.....	35
ภาพประกอบ 3.1 ฟังก์ชันกระบวนการวินิจฉัยมะเร็งเต้านมและการให้ยา.....	39

รายการภาพประกอบ (ต่อ)

ภาพประกอบ	หน้า
ภาพประกอบ 3.2 ตัวอย่างภาพที่มีเซลล์ขนาดใหญ่และชัดเจนเป็นส่วนมาก.....	40
ภาพประกอบ 3.3 ตัวอย่างภาพที่มีเซลล์ขนาดเล็กและไม่ค่อยชัดเจนเป็นส่วนมาก.....	40
ภาพประกอบ 3.4 ภาพเซลล์มะเร็งเรียงด้านมที่ผ่านการข้อมสีแล้ว.....	41
ภาพประกอบ 3.5 ตัวเลข RGB เรียงในรูปแบบที่นำไปใช้งานได้.....	41
ภาพประกอบ 3.6 ขั้นตอนวิธี <i>K</i> -Means Clustering และ <i>K</i> -Median Clustering.....	42
ภาพประกอบ 3.7 ตัวเลขการจัดกลุ่มเรียงในรูปแบบที่ใช้งานได้.....	43
ภาพประกอบ 3.8 ตัวอย่างภาพผลการแยกเซลล์.....	43
ภาพประกอบ 3.9 ขั้นตอนการกำจัดสิ่งรบกวนในภาพ.....	44
ภาพประกอบ 3.10 ตัวอย่างภาพก่อนการกำจัด Noise และภาพผลการกำจัด Noise.....	45
ภาพประกอบ 3.11 ขั้นตอนการนับเซลล์มะเร็งเรียงด้านม.....	45
ภาพประกอบ 3.12 ตัวอย่างผลการนับเซลล์.....	46
ภาพประกอบ 3.13 ตัวอย่างภาพผลการนับเซลล์ของผู้เชี่ยวชาญ.....	47
ภาพประกอบ 3.14 ขั้นตอนการการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ.....	48
ภาพประกอบ 3.15 ตัวอย่างภาพผลการเปรียบเทียบกับผู้เชี่ยวชาญ.....	49
ภาพประกอบ 4.1 กราฟแสดงเปอร์เซ็นต์ของค่า S และ PPV โดยใช้ <i>K</i> -Means.....	64
ภาพประกอบ 4.2 กราฟแสดงเปอร์เซ็นต์ของค่า S และ PPV โดยใช้ <i>K</i> -Median.....	64
ภาพประกอบ 4.3 กราฟเปรียบเทียบค่า S ระหว่าง <i>K</i> -Means และ <i>K</i> -Median.....	65
ภาพประกอบ 4.4 กราฟเปรียบเทียบค่า PPV ระหว่าง <i>K</i> -Means และ <i>K</i> -Median.....	65
ภาพประกอบ 4.5 กราฟเปรียบเทียบ ค่า S และ PPV ระหว่าง <i>K</i> -Means และ <i>K</i> -Median.....	66
ภาพประกอบ 4.6 กราฟแสดงค่าเฉลี่ย S, PPV และค่าเฉลี่ยของทั้ง 2 ค่า ระหว่าง <i>K</i> -Means และ <i>K</i> -Median.....	68
ภาพประกอบ 4.7 กราฟเปรียบเทียบค่า S, PPV ที่เหลือ ระหว่าง <i>K</i> -Means และ <i>K</i> -Median.....	68
ภาพประกอบ 4.8 ตัวอย่างหน้าจอ GUI ของระบบการตัดแยกเซลล์มะเร็งเรียงด้านมแบบอัตโนมัติ.....	69
ภาพประกอบ 5.1 ตัวอย่างการหาค่าเฉลี่ยโดยมีข้อมูลที่มีปัญหา.....	69
ภาพประกอบ 5.2 ตัวอย่างการหาค่ามัธยฐานโดยมีข้อมูลที่มีปัญหา.....	69
ภาพประกอบ ข.1 หน้าจอหลักของโปรแกรม ABC_CIS.....	85
ภาพประกอบ ข.2 ตัวอย่างหน้าจอขั้นตอนการใช้โปรแกรม.....	86
ภาพประกอบ ข.3 ส่วนการเลือกภาพ Original Image บน GUI.....	87

รายการภาพประกอบ (ต่อ)

ภาพประกอบ	หน้า
ภาพประกอบ ข.4 ภาพต้นฉบับถูกแสดงในส่วนแสดงภาพบน GUI.....	87
ภาพประกอบ ข.5 ตัวอย่างผลการแยกภาพบน GUI.....	88
ภาพประกอบ ข.6 ตัวอย่างผลการนับเซลล์บน GUI.....	89
ภาพประกอบ ข.7 ตัวอย่างผลการเปรียบเทียบกับผู้เชี่ยวชาญบน GUI.....	90

บทที่ 1

บทนำ

ในปัจจุบันนี้ผู้ที่ป่วยเป็นโรคมะเร็งเต้านมมีอัตราการเสียชีวิตเพิ่มขึ้นมาก จึงมีผู้คิดค้นวิธีการช่วยเหลือผู้ป่วยมากมาย และได้คิดค้นเทคนิคในการพัฒนาเครื่องมือทางด้านคอมพิวเตอร์เพื่อช่วยให้ความสะดวก รวดเร็วในการวิเคราะห์และวินิจฉัยอาการของผู้ป่วย รวมถึงการวิเคราะห์การให้ยาว่าจะต้องให้ยาหรือไม่และในปริมาณเท่าไรจึงจะเหมาะสมกับผู้ป่วยแต่ละคน

เทคโนโลยีทางด้านคอมพิวเตอร์มีความก้าวหน้าสามารถประยุกต์เพื่อนำไปใช้ในหลากหลายสาขาวิชา ทางด้านการแพทย์ก็เป็นสาขาหนึ่งที่ต้องการเทคโนโลยีทางคอมพิวเตอร์มาช่วยทำให้แพทย์สามารถวิเคราะห์และวินิจฉัยโรคได้อย่างถูกต้อง รวดเร็ว เพียงพอและทันต่อความต้องการในการรักษาที่มากมายในปัจจุบัน

สาขาทางวิทยาการคอมพิวเตอร์สาขาหนึ่งที่มีการนำมาประยุกต์ใช้ร่วมในทางการแพทย์ที่สำคัญคือ การประมวลผลภาพซึ่งเป็นการประยุกต์ใช้งานการประมวลผลสัญญาณบนสัญญาณ 2 มิติ เช่น ภาพนิ่ง (Photograph) หรือภาพวีดิทัศน์ (Video) และยังรวมถึงสัญญาณ 2 มิติอื่นๆ ที่ไม่ใช่ภาพถ่าย โดยการประมวลผลภาพจะมีเทคนิคและขั้นตอนหลากหลายเพื่อใช้ให้เหมาะสมกับงาน วิธีหนึ่งของการประมวลผลภาพที่นิยมนำมาใช้ในการแบ่งแยกวัตถุออกจากภาพหรือแม้กระทั่งการแยกเซลล์ออกจากพื้นหลังในภาพมีหลายวิธีเช่น Thresholding, Region Growing, Markov Random Field (MRF) เป็นต้น แต่ในงานวิจัยนี้จะใช้การแบ่งแยกโดยใช้สี (Color-Based Segmentation) เพื่อใช้ในการวิเคราะห์หาเซลล์มะเร็ง

การแบ่งแยกโดยใช้สี เป็นการใช้สีเป็นตัวบ่งบอกความแตกต่างของวัตถุที่เราสนใจ เพื่อแยกประเภทวัตถุ ซึ่งเป็นประโยชน์มากต่อการแยกพื้นหลังออกจากเซลล์มะเร็งที่อยู่ในภาพ เพื่อนำมาวิเคราะห์หาเซลล์มะเร็งที่ผ่านการย้อมสีแล้วได้ การวิเคราะห์เพื่อรักษามะเร็งเต้านมได้มีมากมายและมีมานานแล้ว ซึ่งสามารถให้ความช่วยเหลือแพทย์ได้ระดับหนึ่ง

การแยกเซลล์มะเร็งออกจากภาพเป็นวิธีหนึ่งที่สามารถทำให้แพทย์ทราบได้ว่าผู้ป่วยรายนั้น จำเป็นจะต้องได้รับการให้ยาหรือไม่ โดยการนับจำนวนเซลล์ที่สามารถแพร่กระจายได้ ถ้ามีจำนวนมากก็สมควรต่อการให้ยาเพื่อไปยับยั้งไม่ให้เซลล์มะเร็งแพร่กระจายออกไป ซึ่งมีผู้คิดค้นวิธีการนับเซลล์เหล่านี้แล้ว แต่ใช้วิธีการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยใช้โครงข่ายประสาทเทียม (Neural Networks) ซึ่งจะต้องให้แพทย์เป็นผู้ระบุบริเวณที่เป็นเซลล์ที่แพร่กระจายได้ เซลล์ที่ไม่แพร่กระจายและบริเวณที่เป็นพื้นหลังของภาพในระหว่างขั้นตอนของการวิเคราะห์ภาพเซลล์มะเร็งเต้านมของผู้ป่วยรายนั้นๆ ซึ่งแพทย์แต่ละคนก็มีความแตกต่างกันในการเลือกบริเวณ ซึ่งจะทำให้ผลการวิเคราะห์ในแต่ละครั้งไม่เหมือนกัน อาจส่งผลให้วินิจฉัยการให้ยาผิดพลาดได้ และวิธีการนี้ยังต้องอาศัยผู้เชี่ยวชาญเฉพาะด้านในการใช้ระบบนี้

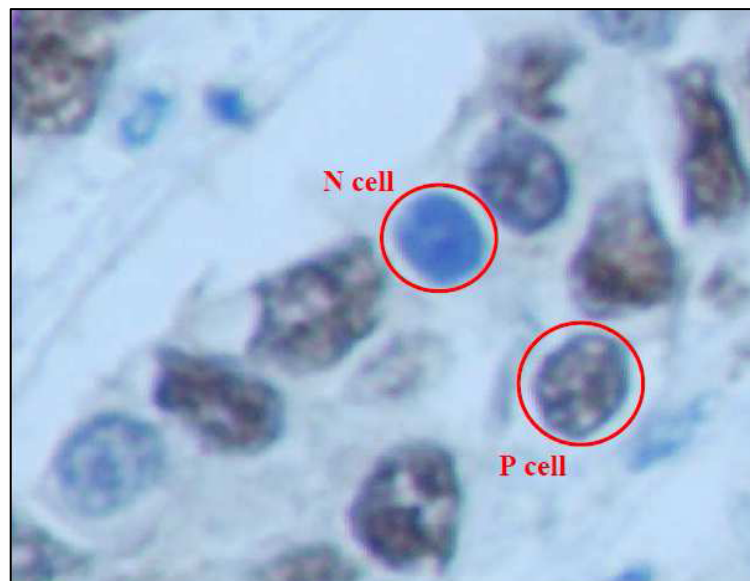
วิทยานิพนธ์นี้จึงได้คิดค้นระบบในการแยกเซลล์มะเร็งเรื้อรังด้านมออกจากภาพแบบอัตโนมัติ โดยการใช้เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เข้ามามีส่วนช่วยให้ระบบสามารถดำเนินการได้อย่างมีประสิทธิภาพมากขึ้น ซึ่งจะช่วยลดภาระของแพทย์ในหลายขั้นตอนลงได้ โดยที่แพทย์ไม่จำเป็นจะต้องเป็นผู้ระบุบริเวณต่างๆของภาพเอง แต่ระบบนี้สามารถเรียนรู้บริเวณดังกล่าวได้เองและสามารถวิเคราะห์ผลเพื่อช่วยวินิจฉัยการให้ยาแก่ผู้ป่วยได้ สามารถให้ความสะดวกแก่แพทย์ในการวินิจฉัยการให้ยาแก่ผู้ป่วยมะเร็งเรื้อรังด้านมได้

1.1 การตรวจเอกสาร

การตัดแยกภาพเซลล์มะเร็งเรื้อรังด้านมแบบอัตโนมัติ มีส่วนประกอบหลักที่ใช้ในงานวิจัย ได้แก่ ภาพเซลล์มะเร็งเรื้อรังด้านม วิธีการในการตัดแยกภาพและเทคนิคต่างๆที่ใช้เพื่อให้ได้ผลการนับเซลล์ โดยแต่ละส่วนประกอบหลักต้องใช้ความรู้ความเข้าใจในทฤษฎีที่เกี่ยวข้อง ซึ่งสามารถอธิบายได้ดังนี้

1.1.1 การได้มาของภาพดิจิทัล

ภาพที่ใช้ในงานวิจัยนี้เป็นภาพของเนื้อเยื่อมะเร็งเรื้อรังด้านมที่ถูกย้อมสีด้วยแอนติบอดี ER ในหลายๆนิวเคลียส การตัดเนื้อเยื่อจะต้องเฝ้าดูใต้กล้องจุลทรรศน์ดิจิทัล (Microscope) ที่กำลังขยาย 400 เท่า ภาพเซลล์ที่ได้มาจะใช้กล้อง Eclipse 80i Advanced Research Microscope เป็นภาพสีขนาด 2560*3200 Pixels ประเภท JPEG File



ภาพประกอบ 1.1 ตัวอย่างภาพเซลล์มะเร็งเรื้อรังที่ถูกย้อมสี

ภาพประกอบ 1.1 แสดงตัวอย่างของภาพเซลล์มะเร็งที่ถูกข้อมสีจาก Microscope ด้วยการขยาย 400 เท่า เซลล์มะเร็งที่ถูกข้อมสีจะถูกระบุเป็น 2 สีคือ สีน้ำตาลและสีน้ำเงิน โดยที่สีน้ำตาลจะแทนผลลัพท์ที่เป็นบวก (P Cell) ส่วนสีน้ำเงินจะแทนผลลัพท์ที่เป็นลบ (N Cell)

1.1.2 Morphological Operations

Morphology (Phukpattaranont *et al.*, 2007) เป็นการดำเนินการทางคณิตศาสตร์ เป็นเครื่องมือที่มีประสิทธิภาพเพื่อใช้สำหรับปัญหาการประมวลผลภาพมากมาย ซึ่งสามารถนำเครื่องมือของ Morphology มาใช้ประโยชน์สำหรับการกรองสิ่งรบกวนและการทำให้รูปร่างสามารถเข้าใจได้ง่ายยิ่งขึ้น

เทคนิค Morphology มี 2 เทคนิคย่อยได้แก่

- 1) Morphological Closing ใช้สำหรับการทำเซลล์ที่อยู่เดี่ยวให้มีรูปร่างที่สมบูรณ์ขึ้น
- 2) Morphological Opening ใช้สำหรับการทำเซลล์หลายเซลล์ที่อยู่ติดกันเล็กน้อยแยกออกจากกันได้

1.1.3 Marker-Controlled Watershed

Marker-Controlled Watershed (Phukpattaranont *et al.*, 2007) เป็นอัลกอริทึมที่ใช้สำหรับใส่ขอบเขตและระบุเครื่องหมายให้แก่ละวัตถุ เพื่อจะแสดงให้เห็นว่าวัตถุในกลุ่มนั้นแยกออกจากกัน ในส่วนนี้จะแยกเซลล์มะเร็งที่อยู่ติดกันให้กลายเป็นเซลล์แต่ละเซลล์ได้ อัลกอริทึมนี้มีประสิทธิภาพใช้สำหรับการแบ่งแยกวัตถุ โดยมีขั้นตอนของ Marker-Controlled Watershed ดังนี้

ขั้นตอนที่ 1 ใช้การทำเครื่องหมายเส้นขอบแบบ Sobel

ขั้นตอนที่ 2 กำหนดเครื่องหมายภายในแต่ละเซลล์

ขั้นตอนที่ 3 นำผลลัพท์จากขั้นตอนที่ 1 และ 2 มารวมกัน

ขั้นตอนที่ 4 คำนวณ Watershed Transform ของผลลัพท์จากขั้นตอนที่ 3

1.1.4 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่อง (Machine Learning, 2009: Online) คือ การทำให้เครื่องเรียนรู้จากข้อมูลตัวอย่างหรือสภาพแวดล้อม เพื่อพัฒนาระบบให้ดียิ่งขึ้น เมื่อเรียนรู้แล้วจะเก็บข้อมูลในฐานข้อมูล ด้วยรูปแบบการแทนความรู้ เช่น กฎ หรือ ฟังก์ชัน

การเรียนรู้ของเครื่อง สามารถเรียนรู้ได้เองโดยอัตโนมัติ เหมาะกับงานที่ไม่สามารถสั่งงานเป็นขั้นตอนได้ เช่น การจดจำใบหน้า ต่างจากการใช้อัลกอริทึมที่มนุษย์สร้างขึ้นเป็นกฎที่แน่นอน ข้อมูลตัวอย่างที่จะให้เครื่องเรียนรู้มี 2 ประเภท คือ

- 1) ตัวอย่างที่มีฉลาก (Labeled Example) คือ ตัวอย่างที่มีคำตอบบอกว่ายู่ในประเภทไหน บอกได้ว่าเป็นสิ่งที่ถูกหรือผิด เรียกว่า การเรียนรู้แบบมีผู้สอน (Supervised Learning)

2) ตัวอย่างที่ไม่มีฉลาก (Unlabeled Example) คือ ตัวอย่างที่ไม่มีคำตอบบอกว่าจะอยู่ในประเภทไหน ถูกหรือผิด จะถูกกำหนดโดยสภาพแวดล้อม อัลกอริทึมจะต้องลองผิดลองถูกเรียนรู้เอง เรียกว่า การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ซึ่งการเรียนรู้จะทำได้ยากกว่า

1.1.4.1 การเรียนรู้แบบมีผู้สอน (Supervised Learning)

การเรียนรู้แบบมีผู้สอน เป็นเทคนิคหนึ่งของการเรียนรู้ของเครื่อง ซึ่งสร้างฟังก์ชันจากข้อมูลสอน (Training Data) ข้อมูลสอนประกอบด้วยวัตถุเข้าและผลที่ต้องการ ผลจากการเรียนรู้จะเป็นฟังก์ชันที่อาจจะให้ค่าต่อเนื่อง (เรียกวิธีการว่า การถดถอย (Regression)) หรือ ใช้ทำนายประเภทของวัตถุ เรียกว่า การแบ่งประเภท (Classification) การกิจของเครื่องเรียนรู้แบบมีผู้สอนคือการทำนายค่าของฟังก์ชันจากวัตถุเข้าที่ถูกต้องโดยใช้ตัวอย่างสอนจำนวนน้อย โดยเครื่องเรียนรู้จะต้องวางนัยทั่วไปจากข้อมูลที่มีอยู่ไปยังกรณีที่ไม่เคยพบอย่างมีเหตุผล

1.1.4.2 การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

การเรียนรู้แบบไม่มีผู้สอน เป็นเทคนิคหนึ่งของการเรียนรู้ของเครื่อง โดยการสร้างโมเดลที่เหมาะสมกับข้อมูล การเรียนรู้แบบนี้แตกต่างจากการเรียนรู้แบบมีผู้สอน คือ จะไม่มีการระบุผลที่ต้องการหรือประเภทไว้ก่อน การเรียนรู้แบบนี้จะพิจารณาวัตถุเป็นเซตของตัวแปรสุ่ม แล้วจึงสร้างโมเดลความหนาแน่นร่วมของชุดข้อมูล

การเรียนรู้ของเครื่องสามารถนำมาใช้เป็นประโยชน์ในงานวิจัยนี้ ในขั้นตอนของการแบ่งแยกเซลล์ออกจากพื้นหลังของภาพ โดยใช้เทคนิคการแบ่งแยกภาพ (Image Segmentation) และใช้เทคนิคการเรียนรู้แบบไม่มีผู้สอน เพื่อให้ระบบมีความฉลาดมากขึ้น โดยสามารถเรียนรู้ได้เอง ทำให้เกิดความสะดวกรวดเร็วแก่ผู้ใช้ระบบมากขึ้น

1.1.5 การแบ่งแยกภาพ (Image Segmentation)

Image Segmentation เป็นกระบวนการในการแยกกลุ่มของ Pixels ในภาพ ภาพจะถูกแบ่งออกเป็นส่วนย่อยๆ ของ Pixels ในแต่ละกลุ่ม ตามลักษณะสำคัญที่เราพิจารณา ซึ่งเป็นขั้นตอนสำคัญในการค้นหาและจดจำรูปแบบของการประมวลผลภาพ สามารถใช้สีในการแบ่งแยกวัตถุได้เรียกว่า Color Segmentation ประโยชน์ของ Image Segmentation ได้แก่

- 1) ลดจำนวนข้อมูลในรูปภาพที่ไม่จำเป็นต่อการวิเคราะห์หลัง
- 2) จัดระเบียบข้อมูลในรูปภาพให้เป็นกลุ่มได้ดีขึ้น
- 3) แสดงข้อมูลในรูปที่เข้าใจง่ายขึ้น

1.1.6 ระบบสี (Color Model)

ระบบสี (Rafael C. Gonzalez *et al.*, 2002) สามารถแบ่งได้เป็นหลายประเภท เช่น

- 1) YIQ Color Model เป็นระบบสีที่ใช้กับโทรทัศน์ ซึ่งเก็บค่าเป็นแบบแถบสัญญาณ (MHz.)
- 2) RGB หรือ Additive Color Model เป็นระบบสีที่ได้จากการรวมสีหรือผสมสีเข้าด้วยกัน ถ้าทุกสีผสมกันทั้งหมดก็จะได้สีขาว ถ้าไม่ปรากฏสีใดๆเลยก็จะเป็นสีดำ ใช้ในการทำงานของจอภาพ Computer
- 3) CMYK หรือ Subtractive Color Model เป็นปรากฏการณ์ที่ตรงกันข้ามกับระบบ Additive Color ในการผสมสี โดยเป็นการจำลองการทำงานของแสงสะท้อน และใช้ในงานด้านการพิมพ์
- 4) HSB หรือ HSL Model เป็นระบบสีที่มีการคำนึงถึงความสว่างของสีด้วย ซึ่งจะใช้ในงานกราฟิกชั้นสูง ระบบสีที่พบเห็นได้บ่อยที่สุดได้แก่ ระบบสี RGB และระบบสี CMYK

1.1.7 รูปแบบไฟล์ภาพ (Image File Formats)

ประเภทของไฟล์ภาพที่ใช้กันทั่วไปมีอยู่หลายแบบ ขึ้นอยู่กับลักษณะการบีบอัดและการนำไปใช้ เช่น

1) BMP (Bitmap)

ไฟล์ .BMP สามารถแสดงรายละเอียดสีได้ละเอียดและแม่นยำ แต่มีขนาดไฟล์ใหญ่มาก จึงไม่นิยมใช้

2) JPEG (Joint Photographics Expert Group)

เป็นมาตรฐานการเก็บข้อมูลรูปภาพที่ถูกกำหนดโดย Joint Photographic Experts Group ซึ่งเป็นหน่วยงานร่วมของ ISO/CCITT โดยมีนามสกุล (File Extension) ที่นิยมใช้คือ .jpeg, .jpg, .jpe, .jif, .jfi ไฟล์ประเภทนี้สามารถเก็บภาพที่มีความละเอียดสูงได้โดยใช้ขนาดไฟล์ที่เล็ก สามารถเก็บภาพสีได้หลากหลายระดับความแม่นยำของสี (Bit Depth) ความสามารถในการย่อขนาดไฟล์ของแฟ้ม JPEG นั้นเกิดจากการใช้เทคนิคการย่อขนาดภาพแบบการบีบอัดคงข้อมูลหลัก (Lossy Compression) หรือการบีบอัดแบบมีความสูญเสียทำให้ไม่นิยมใช้กับภาพที่เป็นลายเส้นหรือไอคอนต่างๆ

3) GIF (Graphic Interchange Format)

เป็นมาตรฐานการเก็บข้อมูลรูปภาพซึ่งถูกกำหนดโดย CompuServe มีนามสกุล (File Extension) คือ .gif ในการบีบอัดของ GIF จะเป็นประเภทไม่สูญเสีย (Lossless Compression) ไฟล์ .GIF สามารถแสดงรายละเอียดสีได้ 256 สี ด้วยเหตุนี้มาตรฐาน GIF จึงไม่นิยมใช้กับภาพถ่ายซึ่งเป็นรูปที่มีจำนวนสีมากกว่าและมีความต่อเนื่องของสี เหมาะสำหรับรูปภาพที่มีจำนวนสีไม่มาก เช่น รูปภาพการ์ตูน เป็นต้น จะใช้ได้ดีกับภาพกราฟิก สัญลักษณ์ต่างๆ หรือภาพแอนิเมชันเคลื่อนไหวง่ายๆ ที่มีสีไม่ต่อเนื่องและมีจำนวนสีไม่มากนัก ซึ่งไฟล์ .GIF นี้มีขนาดเล็กที่สุดเมื่อเทียบกับ .BMP และ .JPG

4) PNG (Portable Network Graphics)

ไฟล์ .PNG สามารถแสดงรายละเอียดสีได้ถึง 16.7 ล้านสี และมีขนาดเล็กเหมือน .GIF

1.1.8 งานวิจัยที่เกี่ยวข้อง

Deng and Clausi (2004) ได้นำเสนอการแบ่งแยกภาพโดยใช้ Markov Random Field (MRF) models ซึ่งเป็นวิธีใหม่ที่ใช้สำหรับ Unsupervised Image Segmentation ซึ่งวิธีการนี้ค่อนข้างยุ่งยากและมีความซับซ้อนในการใช้งาน เนื่องจากจะต้องมีการกำหนดค่าตัวแปรต่างๆซึ่งไม่สามารถรู้ได้ว่าค่าไหนที่สมควรใช้และเหมาะสม ผลลัพธ์ของแต่ละส่วนของวิธีนี้ได้ผลที่ไม่เหมือนกัน จึงเป็นการยากที่จะเลือกใช้ให้เหมาะสมและวิธีการนี้ไม่เหมาะสำหรับนำมาใช้กับเซลล์มะเร็ง เนื่องจากวิธีนี้ใช้กับภาพที่เป็นสี HSV ซึ่งภาพมะเร็งด้านมสามารถใช้สี RGB ก็เพียงพอต่อการทดลองแล้ว

Shih and Cheng (2005) ได้เสนอวิธีการในการหาบริเวณโดยใช้ Seeded Region Growing (SRG) ในการแบ่งแยกภาพด้วยสี ซึ่งวิธีนี้จะแสดงผลลัพธ์ของการแบ่งแยกกลุ่มโดยการเน้นขอบของแต่ละบริเวณที่แบ่งกลุ่มได้ ซึ่งทำให้เห็นแต่ละกลุ่มได้ไม่ชัดเจน ยากต่อการทำความเข้าใจ ว่าส่วนไหนคือบริเวณอะไร

Jeong, Kim และคณะ (2005) ได้ทำการเปรียบเทียบเทคนิค Thresholding ต่างๆ และแสดงค่าความถูกต้องของแต่ละวิธี โดยเทคนิค Thresholding ที่ใช้ในการแยกเซลล์ออกจากภาพคือการระบุ พิกเซลในภาพไปยังวัตถุหรือพื้นหลังอย่างใดอย่างหนึ่ง ซึ่งเป็นวิธีที่ง่ายและให้ผลดีสำหรับการแบ่งแยกภาพ แต่มีข้อเสียคือ ไม่สามารถคำนึงถึงตำแหน่งและลักษณะทางกายภาพของพิกเซลได้ จึงทำให้ไม่เหมาะสมกับเซลล์ที่มีลักษณะต่างกัน

D.J. Withey and Z.J. Koles (2007) ได้อธิบายความแตกต่างในแต่ละยุคสมัยของขั้นตอนวิธีและซอฟต์แวร์ในการแบ่งแยกวัตถุ (Segmentation) ซึ่งสามารถแบ่งออกได้เป็น 3 ยุคสมัยด้วยกัน แต่ละยุคจะมีความแตกต่างกันของขั้นตอนวิธีในการแบ่งแยกวัตถุ ซึ่งมีการพัฒนาขึ้นเรื่อยๆ เพื่อให้เหมาะสมกับวัตถุประสงค์ในการใช้งาน

Phukpattaranont and Boonyaphiphat (2007) ได้เสนอวิธีการแบ่งแยกเซลล์มะเร็งเต้านมออกจากพื้นหลัง โดยใช้โครงข่ายประสาทเทียม (Neural Network) ซึ่งเป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) กล่าวคือให้โปรแกรมเรียนรู้ส่วนที่เป็นพื้นหลังและส่วนที่เป็นเซลล์มะเร็งก่อน ซึ่งจะทำกรวิเคราะห์หาเซลล์ที่ผิดปกติ โดยใช้สีที่เห็นจากการข้อมเป็นตัวที่ใช้ในการพิจารณาและทำการแยกเซลล์ที่ติดกันออกจากกัน จากนั้นนำเฉพาะเซลล์ออกจากพื้นหลังก่อนนำมาทดสอบ เพื่อให้ได้ผลลัพธ์ว่าเซลล์ที่ต้องการรับยามีจำนวนมากหรือน้อยเพียงใด อย่างไรก็ตามวิธีนี้ยังต้องใช้การเรียนรู้ให้ระบบรู้จักกับบริเวณที่เป็น พื้นหลัง, Positive Cell (P Cell) และ Negative Cell (N Cell) เพื่อให้กระบวนการได้รับรู้ก่อน

ซึ่งต้องใช้ผู้เชี่ยวชาญในการเลือกแต่ละบริเวณเอง ทำให้เกิดความแตกต่างกันในแต่ละครั้งของการเลือก จึงเป็นการเรียนรู้แบบ Supervised Learning

Ali and Rosalina (2009) ได้นำเสนอการแยกภาพโดยทำการเปรียบเทียบเทคนิคที่ใช้ระหว่าง Adaptive K-Means Algorithm (AKM), Standard K-Means (SKM), Modified K-Means (MKM) สรุปผลที่ได้ว่า AKM มีกระบวนการทำซ้ำน้อยที่สุด จึงช่วยให้ใช้เวลาในการประมวลผลน้อยที่สุด

ในงานวิจัยนี้จึงได้คิดแนวทางในการแก้ปัญหาเพิ่มเติมโดยจัดให้ระบบไม่ต้องการเรียนรู้ เพื่อไม่ให้เกิดความแตกต่างกันของการทดลองในแต่ละครั้งหรือเรียกว่าการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) และแสดงผลออกมาในรูปแบบที่สามารถมองเห็นความแตกต่างของแต่ละกลุ่มได้ชัดเจน โดยได้นำเทคนิคของการเรียนรู้แบบไม่มีผู้สอนมาประยุกต์ใช้ในการแยกวัตถุออกจากพื้นหลังของภาพ ซึ่งสามารถนำไปใช้ได้กับวัตถุหลายอย่าง แม้กระทั่งภาพเซลล์มะเร็งที่มีขนาดเล็กมากๆ ได้และสามารถนำมาใช้วิเคราะห์หาเซลล์ที่ผิดปกติ หรือเซลล์ที่เป็นโรคมะเร็งต่อไปโดยใช้สีในการแบ่งแยกได้

1.2 วัตถุประสงค์ของโครงการ

เพื่อแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังของภาพโดยใช้เทคนิค Unsupervised Learning และนับจำนวนเซลล์มะเร็งทั้งหมดในภาพได้

1.3 ขอบเขตการดำเนินงาน

- 1.3.1 สร้างขั้นตอนวิธีในการแบ่งแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังของภาพโดยใช้เทคนิคของการเรียนรู้แบบไม่มีผู้สอน
- 1.3.2 ใช้การพิจารณาสีในการแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังของภาพ
- 1.3.3 ใช้ภาพเซลล์มะเร็งเต้านมประเภท JPEG Format ขนาด 400*320 Pixels
- 1.3.4 นับจำนวนเซลล์มะเร็งเต้านมทั้งหมดในภาพ
- 1.3.5 ภาพที่ใช้ได้จากกล้องจุลทรรศน์ดิจิทัล Nikon Eclipse 80i ขยาย 400 เท่า
- 1.3.6 เนื้อเยื่อเซลล์มะเร็งเต้านมผ่านการย้อมสีแบบ Immunohistochemistry

1.4 ขั้นตอนการดำเนินงานและระยะเวลาการดำเนินงาน

ขั้นตอนการดำเนินงานมีดังนี้

- 1.4.1 ศึกษางานวิจัย เอกสารที่เกี่ยวข้องและกำหนดขอบเขตของปัญหาให้ชัดเจน

- 1.4.2 ศึกษาเทคนิคที่เกี่ยวข้องและออกแบบวิธีการในการแปลงภาพให้อยู่ในรูปแบบที่สามารถนำไปทดสอบต่อได้
- 1.4.3 ศึกษาวิธีการในการแยกเซลล์มะเร็งออกจากภาพโดยการเปรียบเทียบการแยกภาพของแต่ละวิธี
- 1.4.4 ศึกษาเครื่องมือที่จะนำมาใช้ในการแยกภาพและพัฒนางานวิจัย
- 1.4.5 เขียนบทความวิจัย
- 1.4.6 วิเคราะห์และออกแบบระบบการแยกเซลล์มะเร็งเต้านมออกจากภาพโดยใช้เทคนิค Unsupervised Learning
- 1.4.7 พัฒนาระบบการแยกเซลล์มะเร็งเต้านมออกจากภาพโดยใช้เทคนิค Unsupervised Learning
- 1.4.8 ทดสอบและประเมินระบบการแยกเซลล์มะเร็งเต้านมออกจากภาพโดยใช้เทคนิค Unsupervised Learning
- 1.4.9 สรุปและวิเคราะห์ผลการประเมินการใช้งานระบบ
- 1.4.10 จัดทำเอกสารประกอบการวิจัย

ระยะเวลาการดำเนินงาน มกราคม 2551 ถึง มีนาคม 2553 แสดงได้ดังตารางที่ 1.1

ตารางที่ 1.1 ตารางแสดงระยะเวลาดำเนินการ

กิจกรรม/ขั้นตอน การดำเนินงาน	เดือน																										
	2551												2552						2553								
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3
1. ศึกษางานวิจัยและเอกสารที่เกี่ยวข้อง	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
2. ศึกษาเทคนิคที่เกี่ยวข้องในงานวิจัย	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
3. ศึกษาวิธีการในการแยกเซลล์	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
4. ศึกษาเครื่องมือที่ใช้	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
5. เขียนบทความวิจัย	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
6. วิเคราะห์และออกแบบระบบ	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
7. พัฒนาระบบ	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
8. ทดสอบและประเมินระบบ	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
9. สรุปและวิเคราะห์ผล	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
10. จัดทำเอกสารวิทยานิพนธ์	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

1.5 สถานที่และเครื่องมือที่ใช้ในงานวิจัย

1.5.1 สถานที่ทำวิจัย

ห้องปฏิบัติการวิจัยเทคโนโลยีระบบสารสนเทศและการประยุกต์ (iSTAR LAB)
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่

1.5.2 เครื่องมือที่ใช้ทำวิจัย

1.5.2.1 ด้านฮาร์ดแวร์ ได้แก่

- 1) เครื่องคอมพิวเตอร์ส่วนบุคคล 1 เครื่อง ซึ่งประกอบไปด้วย
 - CPU Intel Pentium 4 3.00 GHz.
 - RAM 1 GB
 - Hard disk 80 GB
- 2) Printer 1 เครื่อง

1.5.2.2 ด้านซอฟต์แวร์ ได้แก่

- 1) ระบบปฏิบัติการ Microsoft Windows XP
- 2) ตัวแปลภาษา (C Compiler)
- 3) โปรแกรมประยุกต์ MATLAB 7.5
- 4) โปรแกรม Adobe Photoshop CS2
- 5) โปรแกรม SnagIT32

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1.6.1 สามารถแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังของภาพได้โดยใช้เทคนิค Unsupervised Learning

1.6.2 สามารถนับจำนวนเซลล์มะเร็งเต้านมทั้งหมดในภาพได้

1.6.3 สามารถนำไปใช้เป็นประโยชน์ทางการแพทย์ โดยการช่วยเหลือแพทย์ในการวิเคราะห์การให้ยาแก่ผู้ป่วยมะเร็งเต้านมได้

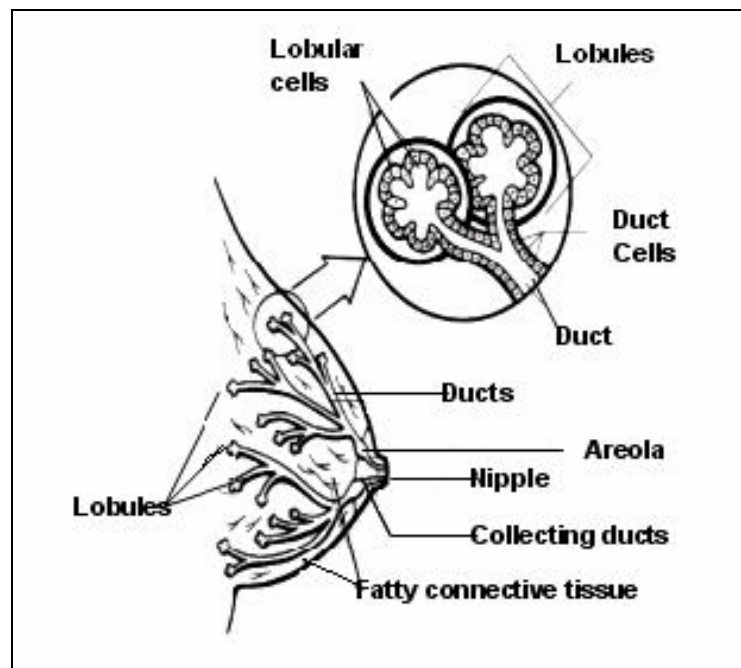
บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีต่างๆ ที่ใช้ในการพัฒนาระบบการตัดแยกภาพเซลล์มะเร็งเต้านมแบบอัตโนมัติ เพื่อให้สามารถแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังของภาพและสามารถนับจำนวนเซลล์มะเร็งเต้านมทั้งหมดในภาพได้ ซึ่งประกอบด้วย 1) เต้านมของเพศหญิง (Female Breast) 2) เซลล์มะเร็งเต้านม (Breast Cancer Cell) 3) หลักการของการรักษามะเร็งเต้านมด้วยวิธีฮอร์โมน 4) การได้มาของภาพดิจิทัล 5) Morphology Operations 6) Marker-Controlled Watershed 7) สถิติพื้นฐาน (Statistics) 8) การเรียนรู้ของเครื่อง (Machine Learning) 8) การแบ่งแยกภาพ (Image Segmentation) 9) ระบายสี (Color Model) และ 10) รูปแบบไฟล์ภาพ (Image File Formats)

2.1 เต้านมของเพศหญิง (Female Breast)

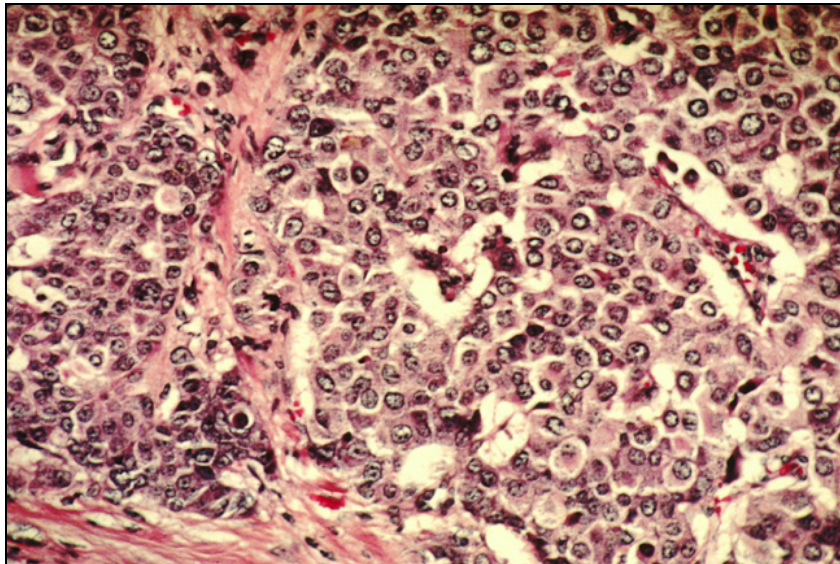
เต้านมของเพศหญิงจะประกอบด้วยส่วนสำคัญคือ เนื้อเยื่อที่อยู่ติดกับชั้นไขมัน (Fatty Connective Tissue) ส่วนปลายของท่อลำเลียงน้ำนม (Collecting Ducts) หัวนม (Nipple) บริเวณรอบหัวนม (Areola) ท่อลำเลียงน้ำนม (Ducts) พู่เล็กๆที่ใช้ผลิตน้ำนม (Lobules) เซลล์ที่อยู่รอบท่อลำเลียงน้ำนม (Duct Cells) เซลล์ของพู่ที่ใช้ผลิตน้ำนม (Lobular Cells) ดังแสดงในภาพประกอบ 2.1



ภาพประกอบ 2.1 ส่วนประกอบของเต้านม (Esugasini Subramaniam *et al.*, 2006)

2.2 เซลล์มะเร็งเต้านม (Breast Cancer Cell)

เซลล์มะเร็งเต้านม (Breast Cancer Cell) สามารถตรวจพบได้ด้วยการค้นหาเชื้อแอนติเจนที่สามารถมองเห็นได้ด้วยการย้อมสี ซึ่งเรียกเทคนิคในการหาเชื้อแอนติเจนนี้ว่า Immunohistochemistry ภาพประกอบ 2.2 เป็นภาพเนื้อเยื่อเซลล์มะเร็งเต้านมที่ผ่านการย้อมสีแล้ว ถ่ายจากกล้องจุลทรรศน์ดิจิทัล เซลล์มะเร็งเต้านมเจริญเติบโตได้จากฮอร์โมนเพศหญิงเอสโตรเจน ดังนั้นถ้าขาดขวางไม่ให้ฮอร์โมนเอสโตรเจนออกฤทธิ์ที่เซลล์มะเร็งได้หรือลดฮอร์โมนเอสโตรเจนให้มึ้น้อยที่สุดก็จะช่วยยับยั้งการแพร่กระจายของมะเร็งเต้านมได้ เซลล์ที่มีการแพร่กระจายได้คือเซลล์ที่มี Estrogen Receptor (ER) ในทางการแพทย์จะเรียกว่า Positive Cell (P Cell) ส่วนเซลล์ที่ไม่มี ER ทางทางการแพทย์จะเรียกว่า Negative Cell (N Cell) (Phukpattaranont *et al.*, 2007)



ภาพประกอบ 2.2 เนื้อเยื่อมะเร็งเต้านม (Clinicalcorrelations, 2008: Online)

มะเร็งเต้านมเป็นสิ่งที่เกิดขึ้นมากที่สุดสำหรับผู้หญิง แต่จะเกิดขึ้นนานๆครั้งในผู้ชาย โดยปกติแล้วมะเร็งเต้านมจะเริ่มก่อตัวในเซลล์ของท่อลำเลียงน้ำนม (Duct Cells) หรือต่อมที่ใช้ผลิตน้ำนม (Lobules) เนื้อเยื่อจะยังคงอยู่ในท่อ เริ่มแรกจะมีการแสดงการเติบโตที่แคบๆก่อน จากนั้นก็จะเติบโตขึ้นมากกว่า 2 มิลลิเมตร และสุดท้ายก็จะแพร่กระจายเข้าไปในท่อน้ำเหลือง ซึ่งจะทำให้เกิดทุ่มบริเวณใต้รักแร้ เซลล์มะเร็งเต้านมนี้จะพัฒนาเติบโตไปอยู่ในบริเวณอื่นๆ จะทำให้เกิดการบวมที่ต่อมน้ำเหลืองบริเวณใต้รักแร้ และเมื่อเซลล์มะเร็งมีการเปลี่ยนแปลงไปยังต่อมน้ำเหลืองใต้รักแร้ ก็จะมีแนวโน้มในการแพร่กระจายไปสู่บริเวณอื่นๆในร่างกายและจะเป็นต้นเหตุของการเกิดโรคร้ายไข้เจ็บอื่นๆ ดังนั้นแพทย์ส่วนใหญ่จะเริ่มด้วยการมีภาพเต้านมผู้หญิงที่คาดการณ์ถึงการบาดเจ็บที่เต้านมโดยใช้การเอ็กซเรย์ก่อน และในปี 1960 มีการทดสอบของ

คลินิกขนาดใหญ่จากนิวยอร์ก สรุปได้ว่า การค้นหามะเร็งเต้านมขั้นแรกคือการใช้วิธีการที่เรียกว่าแมมโมแกรม (Mammogram) ซึ่งสามารถช่วยลดอัตราการตายของผู้ป่วยได้

การตรวจแมมโมแกรม เป็นการตรวจทางรังสีชนิดหนึ่งคล้ายกับการเอกซเรย์ แต่เครื่องตรวจแมมโมแกรมจะเป็นเครื่องเฉพาะ ที่ใช้ปริมาณรังสีน้อยกว่าเครื่องเอกซเรย์ทั่วไป แต่มีความสามารถในการตรวจที่ละเอียดกว่ามาก การตรวจแมมโมแกรมจะมีประโยชน์มากในการตรวจหา มะเร็งเต้านมขนาดเล็ก เพราะสามารถตรวจได้ตั้งแต่ยังคลำก้อนไม่พบ (S Ratanawichitrasin, 2010: Online) ซึ่งทำให้สะดวก ราคาไม่แพงและเป็นทางเลือกที่ได้รับความนิยมสำหรับการค้นหา มะเร็งเต้านมเริ่มแรก เนื่องจากมีความไวต่อการจดจำส่วนของเต้านม ผลของการฉายภาพโดยวิธีแมมโมแกรมนี้จะอยู่ก่อนการตัดเนื้อเยื่อไปตรวจ (E. Subramaniam *et al.*, 2006)

2.3 หลักการของการรักษามะเร็งเต้านมด้วยวิธีฮอร์โมน

การเจริญเติบโตของเต้านมรวมทั้งการทำงานของเต้านมจะขึ้นอยู่กับฮอร์โมน และพบต่อมาว่าการเจริญเติบโตของมะเร็งเต้านมในผู้หญิง มีส่วนหนึ่งที่ขึ้นอยู่กับฮอร์โมนเช่นกัน ฮอร์โมนที่มีผลต่อการเจริญเติบโตของมะเร็งเต้านม คือ

- 1) ฮอร์โมนเพศหญิง (Estrogen) ซึ่งผลิตจากรังไข่ (Ovary) ในหญิงที่ยังมีประจำเดือนอยู่ และจากต่อมหมวกไต (Adrenal Gland) ในหญิงที่หมดประจำเดือนแล้วหรือในหญิงที่ถูกตัดรังไข่ออกไปแล้ว
- 2) ฮอร์โมนโปรเจสเตอโรน (Progesterone) ซึ่งผลิตจากรังไข่
- 3) ฮอร์โมนเพศชาย (Androgen) ผลิตจากต่อมหมวกไต
- 4) คอติโคสเตียรอยด์ (Corticosteroid) ผลิตจากต่อมหมวกไต
- 5) ฮอร์โมนโปรแลคติน (Prolactin) ซึ่งกระตุ้นการหลั่งน้ำนม และฮอร์โมนอีกหลายชนิดที่เป็นตัวควบคุมการหลั่งฮอร์โมนชนิดที่กล่าวมาแล้วทั้งหมด (Tropic Hormone) ผลิตจากต่อมใต้สมอง (Pituitary Gland)

หลักการรักษามะเร็งเต้านมโดยวิธีการทางด้านฮอร์โมนคือ หากมะเร็งเต้านมตอบสนองต่อฮอร์โมน คือ จะเติบโตขึ้นเมื่อได้รับการกระตุ้นจากฮอร์โมน รักษาโดยการลดปริมาณฮอร์โมนในร่างกายลง หรือ ใช้ยาขัดขวางการส่งสัญญาณของฮอร์โมนที่เซลล์มะเร็ง โดยทั่วไปแล้วการรักษาด้วยวิธีฮอร์โมนจะ ได้ผลดี ในผู้ป่วยมะเร็งเต้านมที่ตรวจพบว่า มีเซลล์ที่มีตัวรับสัญญาณฮอร์โมน ซึ่งพบได้ประมาณ 60 - 70% ของผู้ป่วยมะเร็งเต้านมทั้งหมด โดยเฉพาะอย่างยิ่งในคนสูงอายุ เป็นวิธีที่นิยมใช้กันอย่างมาก ทั้งนี้เป็นเพราะว่ามีผลแทรกซ้อนข้างเคียงน้อยกว่าการให้เคมีบำบัดมาก และวิธีการบริหารยาที่สะดวกสำหรับผู้ป่วยมากกว่าการให้เคมีบำบัด

การรักษาทางด้านฮอร์โมนอาศัยหลักการในการขัดขวางการออกฤทธิ์ของฮอร์โมนที่จะไปกระตุ้นเซลล์มะเร็ง ดังนั้นผู้ที่ได้รับการรักษาทางด้านฮอร์โมนควรเป็นกลุ่มที่มีเซลล์ที่มีตัวรับสัญญาณฮอร์โมนอยู่ในเซลล์ ซึ่งในปัจจุบันสามารถตรวจว่าเซลล์มะเร็งของผู้ป่วยมีตัวรับสัญญาณดังกล่าวหรือไม่ โดยการตรวจจากชิ้นเนื้อมะเร็ง ตัวรับสัญญาณที่ทำการตรวจ มี 2 ชนิดคือ Estrogen Receptor (ER) และ Progesterone Receptors (PR) ซึ่งหากมีตัวรับสัญญาณ จะเรียกว่า ผลการตรวจให้ผลบวก (Positive) ดังนั้นจะเห็นได้ว่าการตรวจหา ER และ PR จากก้อนมะเร็งที่ตัดออกมาในตอนแรกจะมีความสำคัญมาก ซึ่งถือเป็นแนวมาตรฐานในการดูแลผู้ป่วยมะเร็งเต้านม ที่จะต้องนำชิ้นเนื้อส่งตรวจ ในกรณีที่ไม่ทราบผลการตรวจตัวรับสัญญาณดังกล่าว (Unknown) ก็อาจใช้การรักษาทางด้านฮอร์โมนได้ แต่ผลดีจากการรักษามีน้อยเพียงร้อยละ 10 เท่านั้น ส่วนกรณีที่ผลการตรวจชิ้นเนื้อแล้ว ไม่พบว่าเซลล์มะเร็งมีตัวรับสัญญาณทั้ง 2 ชนิด (Negative) ไม่แนะนำให้รักษาด้วยวิธีการด้านฮอร์โมน เพราะไม่เกิดประโยชน์

การรักษาด้วยวิธีการทางด้านฮอร์โมน สามารถใช้ได้กับทั้งผู้ป่วยมะเร็งเต้านมที่ยังมีประจำเดือนอยู่ และหมดประจำเดือนแล้ว แต่วิธีการเลือกใช้จะแตกต่างกัน

การใช้ฮอร์โมนในการรักษามะเร็งเต้านม สามารถแบ่งออกเป็นกลุ่มใหญ่ๆ ได้ 2 กลุ่ม คือ

1) การใช้ยาเข้าไปแย่งที่กับตัวรับสัญญาณที่เซลล์มะเร็ง (ER) เพื่อไม่ให้ฮอร์โมนสามารถกระตุ้นเซลล์มะเร็งให้เติบโตได้ กลุ่มนี้ได้แก่ การใช้ยาที่เป็น Anti-estrogen

2) การทำลาย หรือ ยับยั้งไม่ให้มีฮอร์โมนเพศหญิงอยู่ในร่างกาย ในกลุ่มนี้ประกอบด้วยวิธีการหลายอย่าง เช่น

- การทำลายรังไข่ ซึ่งเป็นแหล่งผลิตฮอร์โมน ในหญิงวัยก่อนหมดประจำเดือน

- การยับยั้งการเปลี่ยนแปลงฮอร์โมนเพศชาย ที่จะเปลี่ยนมาเป็นฮอร์โมนเพศหญิง ซึ่งเกิดการเปลี่ยนแปลงของฮอร์โมนที่ต่อมหมวกไต

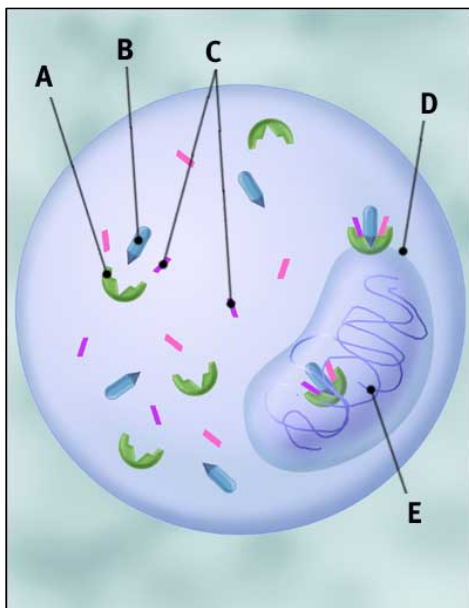
- การยับยั้งการกระตุ้นจากต่อมใต้สมอง ที่จะมากระตุ้นให้รังไข่ผลิตฮอร์โมน Anti-estrogen (ยาที่แย่งที่กับตัวรับสัญญาณ ER) ออกมา

การค้นพบยากลับ Anti-estrogen เมื่อ 30-40 ปีที่แล้ว นับเป็นความก้าวหน้าครั้งสำคัญของการแพทย์ในด้านการรักษามะเร็งเต้านม ที่สามารถช่วยให้ผลการรักษามะเร็งเต้านม ได้ผลดีขึ้นอย่างมาก โดยที่ผลข้างเคียงจากการใช้ยาดังกล่าวมีไม่มากนัก ยาที่สำคัญที่ได้รับการใช้อย่างแพร่หลายในกลุ่มนี้คือ Tamoxifen ซึ่งสามารถบริหารยาโดยการกิน

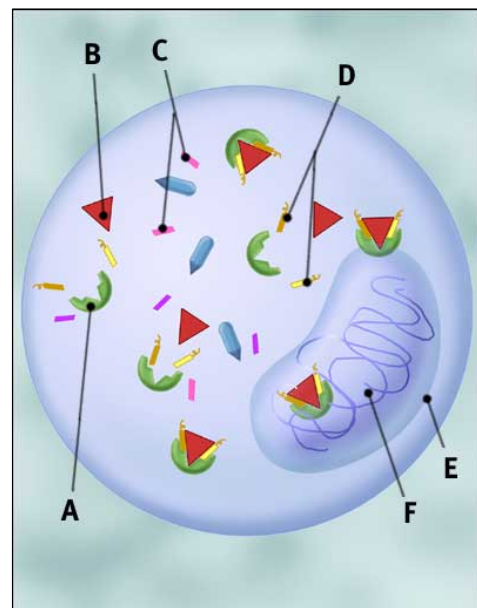
ยา Tamoxifen จะออกฤทธิ์โดยการแย่งจับกับตัวรับสัญญาณของเซลล์มะเร็งเต้านม ดังนั้นภายหลังจากการผ่าตัดรักษา การให้เคมีบำบัด หรือ การฉายรังสี หากมีเซลล์มะเร็งยังคงหลงเหลืออยู่ในร่างกายจำนวนน้อย และเซลล์นั้นเป็นเซลล์ที่มีตัวรับสัญญาณของฮอร์โมนเพศหญิงอยู่ ยา Tamoxifen จะเข้าไปแย่งที่กับฮอร์โมนเอสโตรเจนที่มีอยู่ในร่างกาย ไม่ให้มีโอกาสกระตุ้นให้เซลล์มะเร็งที่เหลืออยู่นั้น เติบโตได้ หรือ หากจะ ได้ ก็จะช้ากว่าปกติ ดังนั้น ผู้ป่วยมะเร็งเต้านมที่จะได้ประโยชน์จากยาดังกล่าวนี้จะต้องเป็นกลุ่มที่มี ER Positive หรือ PR Positive เท่านั้น

การใช้ยา Tamoxifen ยังเป็นที่ใช้กันอย่างแพร่หลายในปัจจุบัน ใช้ได้กับทั้งสตรีที่ยังมีประจำเดือนอยู่ และสตรีที่หมดประจำเดือนแล้ว สามารถใช้ได้ดีในการลดการกลับมาเป็นซ้ำของมะเร็งเต้านมภายหลังการรักษาด้วยวิธีอื่น โดยกินยาดังกล่าววันละ 1 เม็ด นาน 5 ปี นอกจากนี้ ยังมีการใช้ยานี้ก่อนการผ่าตัดเพื่อลดขนาดของมะเร็งลง (มักใช้ในผู้สูงอายุที่ร่างกายไม่สามารถรับเคมีบำบัดได้) หรือ ใช้ป้องกันมะเร็งเต้านมในสตรีที่มีความเสี่ยงของการเกิดมะเร็งสูงมากกว่าคนทั่วไป

ผลข้างเคียงที่มักจะได้รับคือการกล่าวถึงของการใช้ยาดังกล่าว คือ อาจทำให้เกิดมะเร็งในเยื่อโพรงมดลูกได้ และอาจมีผลทำให้เกิดเส้นเลือดอุดตันได้ แต่ในอุบัติการณ์ที่ต่ำมาก



ภาพประกอบ 2.3 ส่วนประกอบในเซลล์
(Breastcancer, 2008: Online)



ภาพประกอบ 2.4 ในเซลล์ที่มี Tamoxifen
(Breastcancer, 2008: Online)

จากภาพประกอบ 2.3 เซลล์ปกติจะประกอบไปด้วย

- A: Estrogen Receptor
- B: Estrogen
- C: Estrogen Helper Proteins
- D: Nucleus
- E: DNA Genetic Material

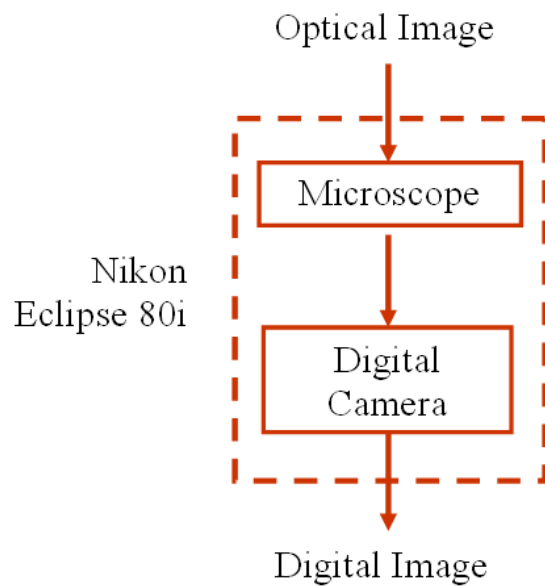
จากภาพประกอบ 2.4 เซลล์ที่มียา Tamoxifen จะมีการทำงานดังภาพ ซึ่งประกอบด้วย

- A: Estrogen Receptor
- B: Tamoxifen

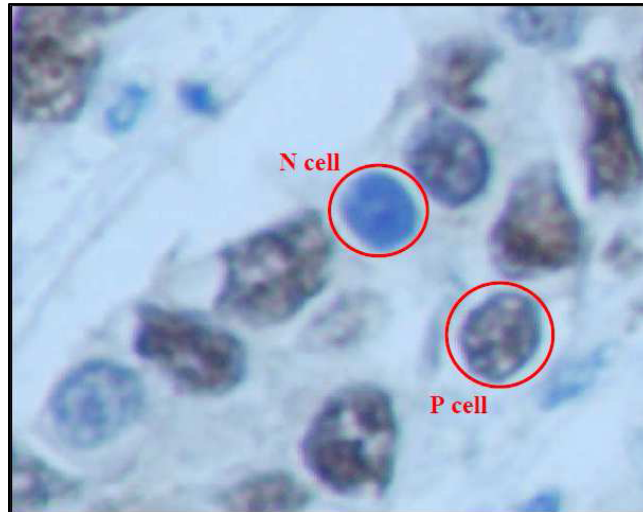
- C: Estrogen Helper Proteins
- D: Tamoxifen Helper Proteins
- E: Nucleus
- F: DNA Genetic Material

2.4 การได้มาของภาพดิจิทัล

การได้มาของภาพดิจิทัลที่ใช้ในงานวิจัยนี้มีขั้นตอนดังต่อไปนี้ 1) แพทย์ทำการตัดเนื้อเยื่อ บริเวณเต้านมของผู้ป่วย 2) นำเนื้อเยื่อมาผ่านการย้อมสีด้วยแอนติบอดี ER เซลล์ที่เป็นบวกจะแสดง ออกเป็นสีน้ำตาล และเซลล์ที่เป็นลบจะแสดงออกเป็นสีน้ำเงิน 3) ผ่านการเพิ่มขนาดเป็น 400 เท่าด้วยกล้องจุลทรรศน์ดิจิทัล (Microscope) 4) ผ่านการถ่ายภาพจากกล้อง Eclipse 80i Advanced Research Microscope และ 5) ภาพดิจิทัลที่ได้จะบันทึกเป็นภาพสีขนาด 2560*3200 Pixels เป็นไฟล์ประเภท JPEG ขั้นตอนการได้มาของภาพที่นำมาทดลองแสดงได้ดังภาพประกอบ 2.5



ภาพประกอบ 2.5 ขั้นตอนการได้มาของภาพดิจิทัล



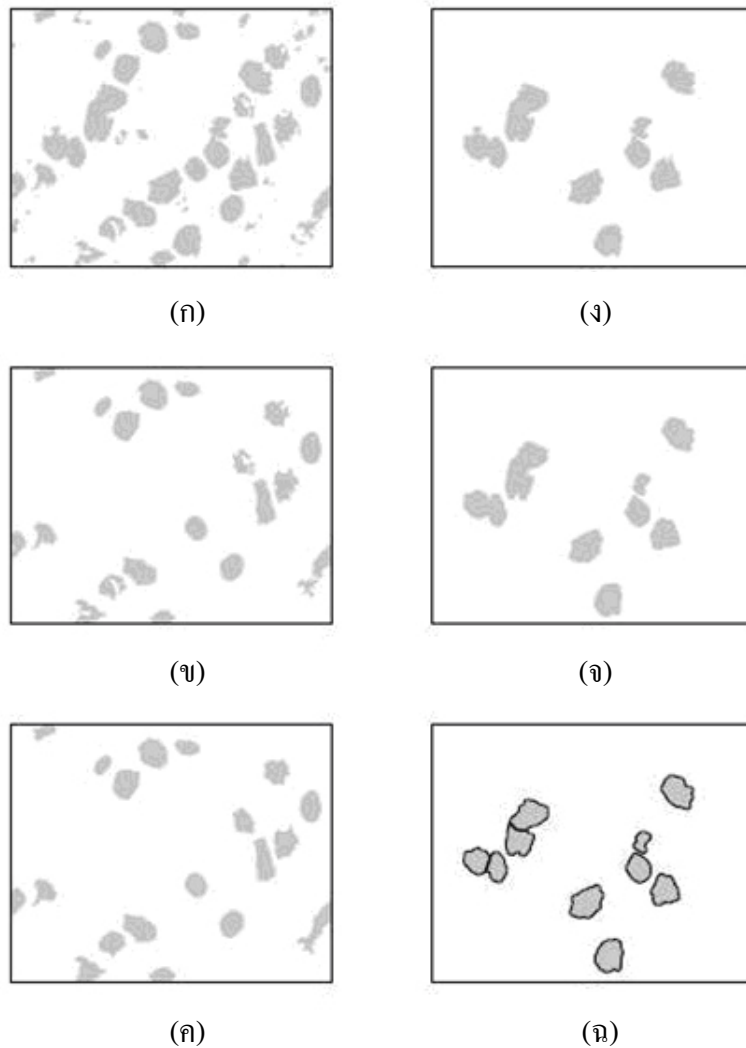
ภาพประกอบ 2.6 ตัวอย่างภาพเซลล์มะเร็งที่ถูกย้อมสี

ภาพประกอบ 2.6 แสดงตัวอย่างของภาพเซลล์มะเร็งที่ถูกย้อมสีจาก Microscope ด้วยการขยาย 400 เท่า วิธีการย้อมสีใช้ประโยชน์ในการพิสูจน์การมีอยู่ของ ER หรือ PR ในเซลล์มะเร็งเต้านม เซลล์มะเร็งที่ถูกย้อมสีจะถูกระบุเป็น 2 สีคือ สีนํ้าตาลและสีน้ำเงิน โดยที่สีนํ้าตาลจะแทนผลลัพท์ที่เป็นบวก (P Cell) ส่วนสีน้ำเงินจะแทนผลลัพท์ที่เป็นลบ (N Cell) ดังแสดงในภาพประกอบ 2.6 เป็นการย้อมสีด้วย ER อัตราส่วนของจำนวนผลลัพท์ของเซลล์มะเร็งที่เป็นบวกกับเซลล์มะเร็งทั้งหมดในภาพจะถูกนำไปใช้ในการวางแผนการรักษาโดยแพทย์ต่อไป

2.5 Morphological Operations

Morphology เป็นการดำเนินการทางคณิตศาสตร์ซึ่งมีการดำเนินการแบบไม่เป็นเชิงเส้น (Nonlinear) โดยมีการทำงานบนรูปร่างของวัตถุ เป็นเครื่องมือที่มีประสิทธิภาพเพื่อใช้กับปัญหาการประมวลผลภาพ ตัวอย่างเช่น การเตรียมภาพก่อนการประมวลผล การแบ่งแยกโดยใช้รูปร่างของวัตถุและอธิบายการวัดปริมาณของวัตถุ ซึ่งสามารถนำการดำเนินการของ Morphology มาใช้ประโยชน์สำหรับการกรองสิ่งรบกวนและการทำให้รูปร่างสามารถเข้าใจได้ง่ายยิ่งขึ้น

Morphological Operations (การปฏิบัติการเชิงสัจฐาน) เป็นการทำให้รูปร่างของวัตถุในภาพเปลี่ยนแปลงไป เช่น การขยายพิกเซลหรือการทำให้วัตถุใหญ่ขึ้น การลดพิกเซลหรือการทำให้วัตถุเล็กลง การเติมเต็มพื้นที่ และการลบพิกเซลที่ไม่ต้องการ เป็นต้น



ภาพประกอบ 2.7 ตัวอย่างผลการทำ Morphology (Phukpattaranont *et al.*, 2007)

- (ก) หลังจากการทำกระบวนการแยกเซลล์ออกจากพื้นหลังของภาพ และจัดขนาดของวัตถุให้เป็น 3 กลุ่ม ได้แก่ ขนาดเล็ก กลาง และใหญ่
- (ข) เซลล์เดี่ยวก่อนทำ Morphological Closing
- (ค) เซลล์เดี่ยวหลังจากทำ Morphological Closing เพื่อทำเซลล์ให้มีรูปร่างกลมสมบูรณ์ขึ้น
- (ง) เซลล์หลายเซลล์ที่อยู่ติดกันก่อนการทำ Morphological Opening
- (จ) เซลล์หลายเซลล์ที่อยู่ติดกันหลังการทำ Morphological Opening ทำให้เซลล์ที่ติดกันเล็กน้อย แยกออกจากกันได้
- (ฉ) หลังการแยกเซลล์ที่ติดกันออกจากกันด้วย Marker-Controlled Watershed

การปฏิบัติการเชิงสัจฐานมีมากมายหลายวิธี ที่เกี่ยวข้องกับงานวิจัยนี้ ได้แก่ Closing, Opening, Region Filling, Border Clearing และ Watershed Segmentation โดย Operations ทั้ง 5 วิธี นี้จะกระทำกับภาพขาวดำ ซึ่งมีรายละเอียดดังต่อไปนี้

2.5.1 Morphological Closing

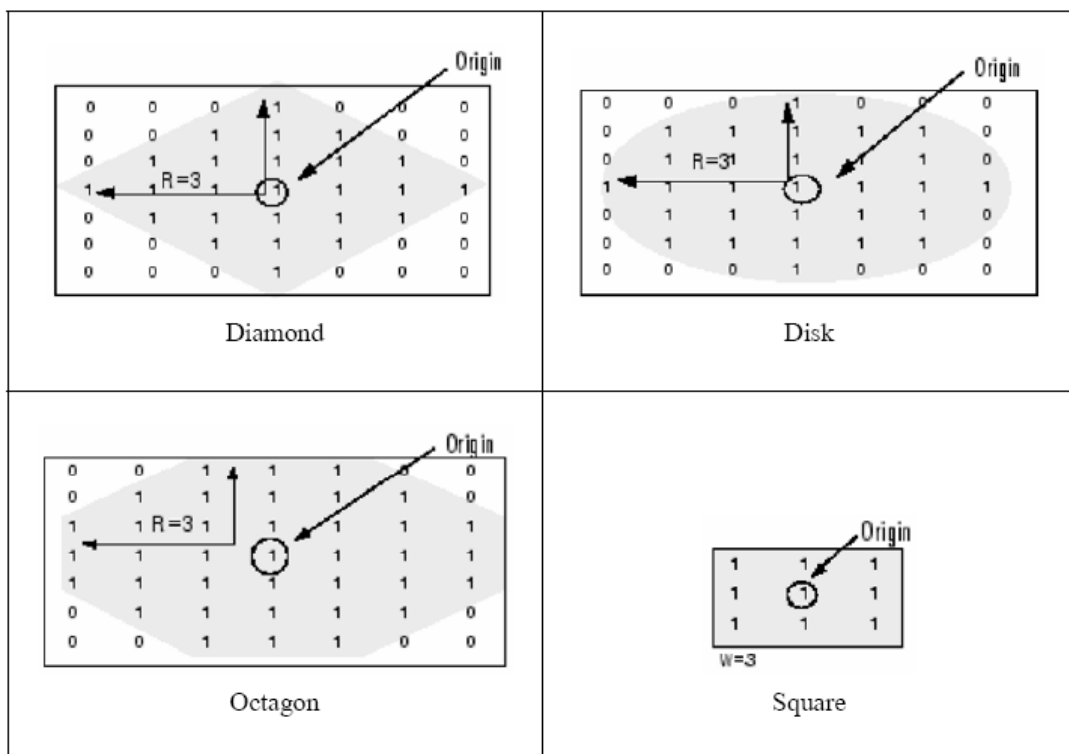
ใช้สำหรับการทำเซลล์ที่อยู่เดี่ยวให้มีรูปร่างที่สมบูรณ์ขึ้น ตัวอย่างผลที่ได้แสดงในภาพประกอบ 2.7(ค)

2.5.2 Morphological Opening

ใช้สำหรับการทำเซลล์หลายเซลล์ที่อยู่ติดกันเล็กน้อยแยกออกจากกันได้ ตัวอย่างผลที่ได้ดังแสดงในภาพประกอบ 2.7(จ)

Opening เป็นการเปิดพื้นที่ว่างในภาพให้มากขึ้น มีการดำเนินการหลักๆคือการลดพิกเซล (Erosion) ของวัตถุลงด้วย Structuring Element ที่กำหนด และขยายพิกเซล (Dilation) ของวัตถุด้วย Structuring Element เดิม (P. Soille, 2002) การ Opening แสดงได้ดังสมการ (2-1) (R. C. Gonzalez *et al.*, 2001 ; A. McAndrew, 2004)

Structuring Element เป็นภาพขาวดำที่ใช้ในกระบวนการ Erosion และ Dilation รูปร่างและขนาดของ Structuring Element จะถูกกำหนดให้เหมาะสมกับการใช้งาน ตัวอย่างของ Structuring Element ได้แก่ Diamond, Disk, Octagon, และ Square ดังภาพประกอบ 2.8



ภาพประกอบ 2.8 Structuring Element (The MathWorks, 2008: Online)

สมการที่ใช้สำหรับการ Opening คือ

$$A \circ B = (A \ominus B) \oplus B \tag{2-1}$$

โดย $A \circ B$ คือ การ Opening ภาพต้นฉบับ (A) ด้วย Structuring Element (B)

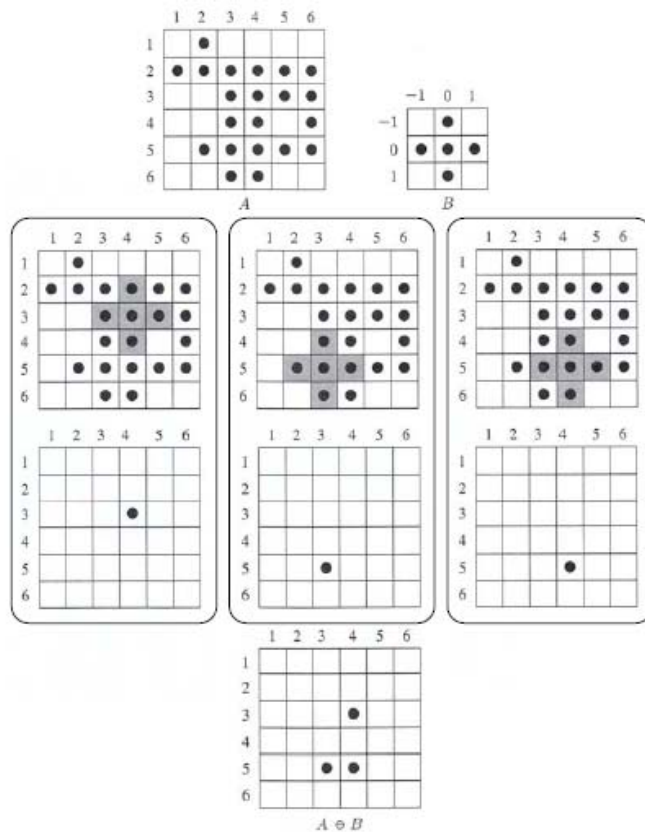
Erosion เป็นการลดพิกเซลของวัตถุในภาพด้วย Structuring Element สามารถ แสดงได้ดัง

สมการ

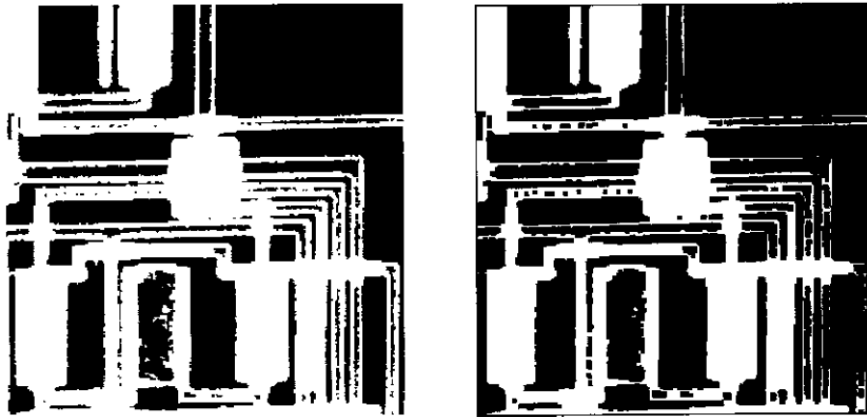
$$A \ominus B = \bigcap_{x \in B} A_x \tag{2-2}$$

โดย $A \ominus B$ คือ การ Erosion ภาพต้นฉบับ (A) ด้วย Structuring Element (B)

จากสมการ (2-2) กล่าวได้ว่า เมื่อวัตถุของ Structure Element ทั้งหมดอยู่ในวัตถุของภาพ จะได้ผลการ Erosion ที่วัตถุมีตำแหน่งเดียวกันกับตำแหน่งของภาพที่มีตำแหน่งเริ่มต้นของ Structuring Element วางอยู่ วิธี Erosion แสดงดังภาพประกอบ 2.9 และตัวอย่าง Erosion แสดงดังภาพประกอบ 2.10



ภาพประกอบ 2.9 วิธี Erosion (A. McAndrew, 2004)



(ก)

(ข)

ภาพประกอบ 2.10 ตัวอย่าง Erosion (ก) ภาพต้นฉบับ (ข) ภาพผลลัพธ์ (A. McAndrew, 2004)

Dilation เป็นการขยายพิกเซลของวัตถุในภาพด้วย Structuring Element สามารถแสดงได้ดังสมการ (2-3) (A. McAndrew, 2004)

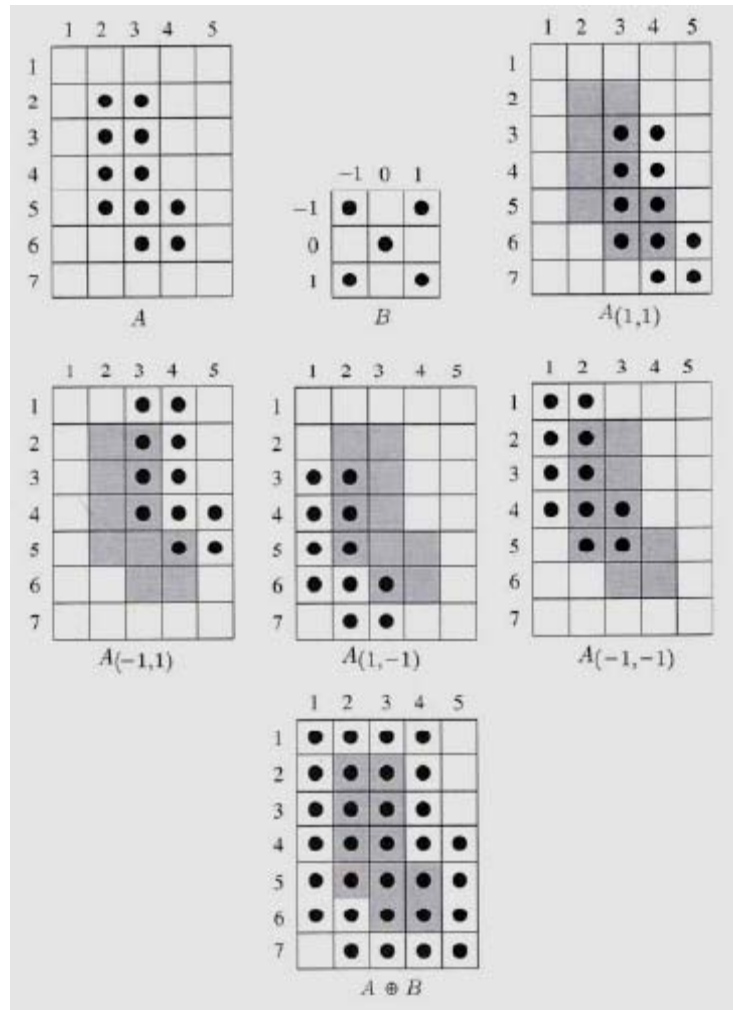
$$A \oplus B = \bigcup_{x \in B} A_x \quad (2-3)$$

โดย $A \oplus B$ คือ การ Dilation ภาพต้นฉบับ (A) ด้วย Structuring Element (B)

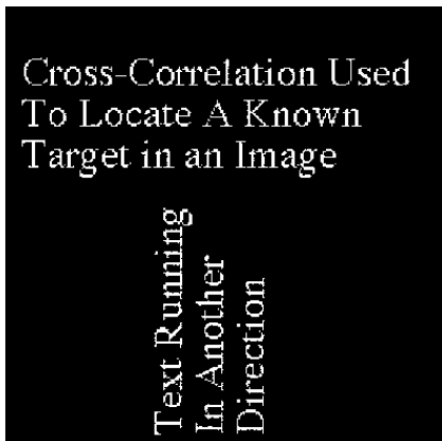
จากสมการ (2-3) กล่าวได้ว่า การ Dilation เกิดจากการรวมของวัตถุทั้งหมด เมื่อวัตถุในภาพต้นฉบับเลื่อนตามพิกัดของวัตถุใน Structuring Element วิธี Dilation แสดงดังภาพประกอบ 2.11 และตัวอย่าง Dilation แสดงดังภาพประกอบ 2.12

นอกจากการ Opening ดังกล่าวข้างต้นแล้ว การ Opening ยังมีอีกรูปแบบหนึ่งที่ไม่ใช่ Structuring Element คือ Area Opening เป็นการลบกลุ่มพิกเซลของวัตถุที่มีขนาดหรือพื้นที่ที่ไม่ต้องการออก โดยมีขั้นตอนวิธี ดังนี้

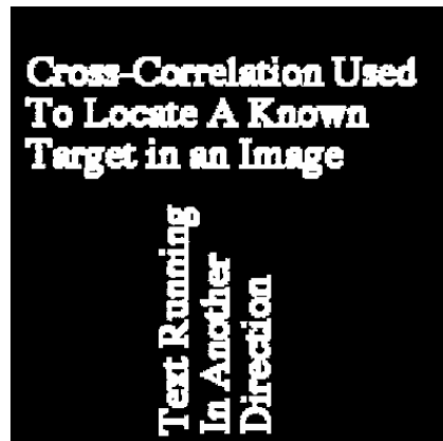
- 1) กำหนดกลุ่มพิกเซลของวัตถุ
- 2) คำนวณหาพื้นที่ของแต่ละกลุ่มพิกเซล
- 3) ลบกลุ่มพิกเซลที่มีขนาดน้อยกว่าที่กำหนด



ภาพประกอบ 2.11 วิธี Dilation (A. McAndrew, 2004)



(ก)



(ข)

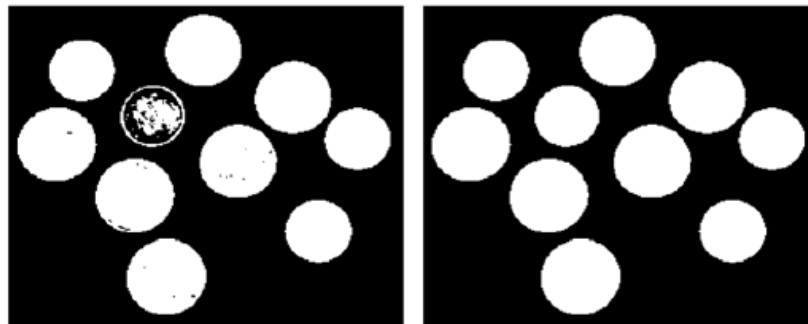
ภาพประกอบ 2.12 ตัวอย่าง Dilation (ก) ภาพต้นฉบับ (ข) ภาพผลลัพธ์ (A. McAndrew, 2004)

2.5.3 Region Filling

พื้นที่ว่างในวัตถุที่เกิดขึ้นในกระบวนการประมวลผลภาพ อาจส่งผลให้การตัดแยกวัตถุมีความผิดพลาด ทำให้การประมวลผลทางสถิติคลาดเคลื่อนด้วย ดังนั้น จึงจำเป็นต้องมีการเติมเต็มพื้นที่ว่างดังกล่าว เพื่อให้การตัดแยกวัตถุและการประมวลผลทางสถิติมีความผิดพลาดน้อยลง ตัวอย่างการเติมเต็มพื้นที่ว่างในวัตถุแสดงดังภาพประกอบ 2.13

2.5.4 Border Clearing

ในการประมวลผลภาพในหลายๆงาน มีความจำเป็นต้องกำจัดวัตถุขนาดเล็กที่อยู่ติดขอบภาพ วัตถุเหล่านี้อาจทำให้การประมวลผลทางสถิติมีความคลาดเคลื่อนได้ การกำจัดวัตถุที่อยู่ติดขอบภาพทำได้โดยการหาวัตถุที่อยู่ติดขอบภาพแล้วกำหนดให้วัตถุอื่นมีค่าเป็นศูนย์ จากนั้นนำไปลบออกจากภาพต้นฉบับ จะได้ผลลัพธ์เป็นภาพที่ไม่มีวัตถุติดขอบภาพ (P. Soille, 2002) ตัวอย่างการกำจัดวัตถุที่อยู่ติดขอบภาพดังภาพประกอบ 2.14



ภาพประกอบ 2.13 การเติมเต็มพื้นที่ว่างในวัตถุ (ก) ภาพต้นฉบับ (ข) ภาพผลลัพธ์

(The MathWorks, 2008: Online)



ภาพประกอบ 2.14 การกำจัดวัตถุที่อยู่ติดขอบภาพ (ก) ภาพต้นฉบับ

(ข) วัตถุที่อยู่ติดขอบภาพ (ค) ภาพผลลัพธ์

2.6 Marker-Controlled Watershed

Marker-Controlled Watershed เป็นกระบวนการที่ใช้สำหรับใส่ขอบเขตและระบุเครื่องหมายให้แต่ละวัตถุ เพื่อจะแสดงให้เห็นว่าวัตถุในกลุ่มนั้นแยกออกจากกัน ในส่วนนี้จะแยกเซลล์มะเร็งที่อยู่ติดกันให้กลายเป็นเซลล์แต่ละเซลล์ได้ กระบวนการมีขั้นตอนดังนี้

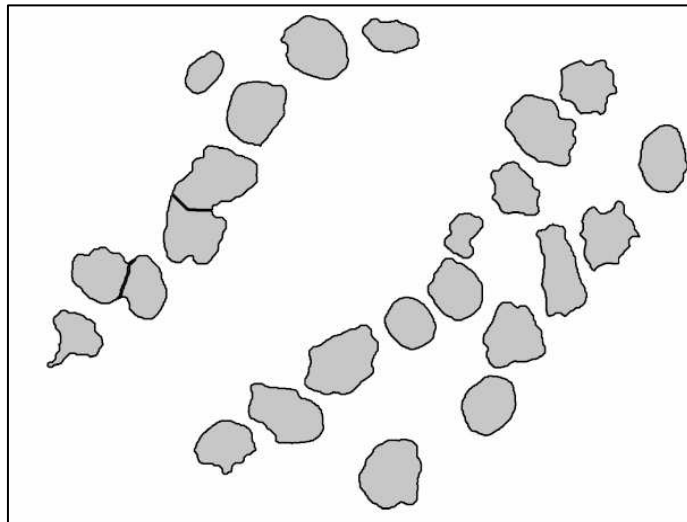
ขั้นตอนที่ 1 ใช้การทำเครื่องหมายเส้นขอบแบบ Sobel

ขั้นตอนที่ 2 กำหนดเครื่องหมายภายในแต่ละเซลล์

ขั้นตอนที่ 3 นำผลลัพธ์จากขั้นตอนที่ 1 และ 2 มารวมกัน

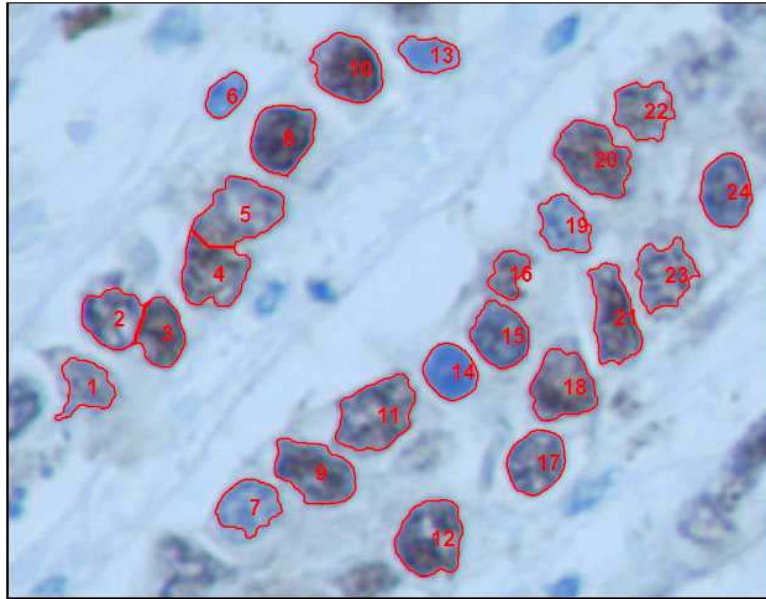
ขั้นตอนที่ 4 คำนวณ Watershed Transform ของผลลัพธ์จากขั้นตอนที่ 3

จากภาพประกอบ 2.7 (จ) แสดงภาพที่ถูกแบ่งแยกหลังจากการใช้ Marker-Controlled Watershed ผลลัพธ์ของการแบ่งแยกจะได้รับความถูกต้องสูงขึ้นซึ่งเป็นสิ่งสำคัญ เซลล์หลายเซลล์ที่เชื่อมติดกัน 2 กลุ่ม สามารถทำให้แยกไปเป็นเซลล์เดี่ยว 4 เซลล์ได้ สุดท้ายจะได้เซลล์มะเร็งทั้งหมด ซึ่งจะเห็นได้ว่าการรวมภาพของเซลล์หลายๆเซลล์ที่มีความแตกต่างกันได้อย่างชัดเจน โดยการนำส่วนที่เป็นเซลล์เดียวกับส่วนที่เป็นเซลล์หลายเซลล์ที่ติดกันมารวมกันได้โดยใช้การดำเนินการตรรกะ OR



ภาพประกอบ 2.15 ภาพที่ถูกแบ่งแยกหลังจากการใช้การดำเนินการตรรกะ OR (Phukpattaranont *et al.*, 2007)

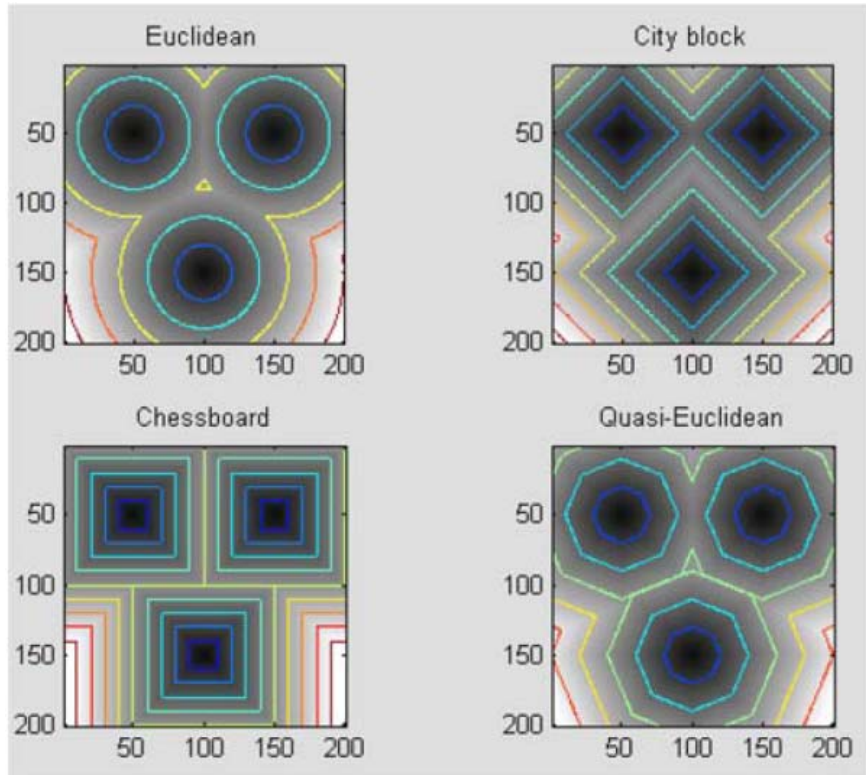
ภาพประกอบ 2.15 แสดงภาพที่ถูกแบ่งแยกหลังจากการใช้การดำเนินการตรรกะ OR ซึ่งได้ตัดวัตถุชิ้นเล็กๆออกจากภาพ เพราะถือว่าเป็นสิ่งที่ไม่ต้องการและทำให้ภาพที่แสดงออกมาไม่สมบูรณ์ ได้มีการเพิ่มเติมภาพการแบ่งแยกโดยใช้เป็นภาพสี RGB และสามารถนับจำนวนเซลล์ทั้งหมดในภาพได้ แสดงผลลัพธ์ดังภาพประกอบ 2.16 ซึ่งจำนวนของเซลล์มะเร็งที่ถูกแบ่งแยกออกทั้งหมดคือ 24 เซลล์ สามารถแสดงให้เห็นอย่างชัดเจน



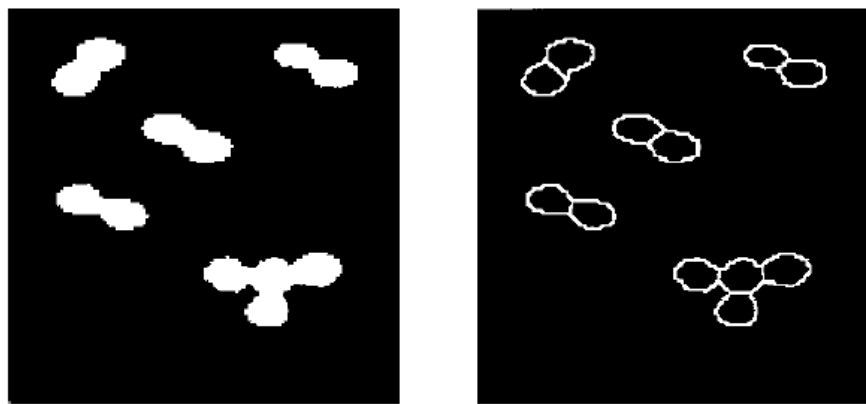
ภาพประกอบ 2.16 ผลภาพหลังจากการใส่ตัวเลขระบุเซลล์ (Phukpattaranont *et al.*, 2007)

การประมวลผลภาพบางงานจะพบวัตถุมากกว่า 2 ชิ้นที่มีบางส่วนติดหรือซ้อนทับกัน และมีความจำเป็นต้องตัดแยกให้ขาดออกจากกันเพื่อให้ได้ผลลัพธ์ที่ต้องการ Watershed Segmentation เป็นวิธีการประมวลผลภาพวิธีหนึ่งที่ใช้สำหรับตัดแยกวัตถุดังกล่าวให้ขาดออกจากกันได้ สำหรับงานวิจัยนี้ใช้ขั้นตอนวิธี Watershed ของ Meyer (Mayer's Watershed Algorithm) ซึ่งเป็นวิธีที่นิยมใช้กันอย่างแพร่หลาย ศึกษารายละเอียดเพิ่มเติมได้ใน (F. Meyer, 1994) โดยใช้ร่วมกับ Distance Transformation (Q. Chen *et al.*, 2004) ซึ่งมีหลายรูปแบบดังภาพประกอบ 2.17 ฟังก์ชันของ Distance Transformation แสดงได้ดังสมการ (2-4) - (2-7) และตัวอย่างการตัดแยกวัตถุที่ติดกันด้วย Watershed Segmentation แสดงดังภาพประกอบ 2.18

$$d_{euclidean} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2-4)$$



ภาพประกอบ 2.17 รูปแบบของ Distance Transformation (The MathWorks, 2008: Online)



(ก)

(ข)

ภาพประกอบ 2.18 การตัดแยกวัตถุที่ติดกันด้วย Watershed Segmentation (M. Sonka *et al.*, 2008)

$$d_{chessboard} = \max(|x_1 - x_2|, |y_1 - y_2|) \quad (2-5)$$

$$d_{cityblock} = |x_1 - x_2| + |y_1 - y_2| \quad (2-6)$$

$$d_{quasi-euclidean} = \begin{cases} |x_1 - x_2| + (\sqrt{2} - 1)|y_1 - y_2| & , \text{ if } |x_1 - x_2| > |y_1 - y_2| \\ (\sqrt{2} - 1)|x_1 - x_2| + |y_1 - y_2| & , \text{ otherwise} \end{cases} \quad (2-7)$$

จากทฤษฎีการประมวลผลภาพดังกล่าว จะถูกใช้ในขั้นตอนวิธีการวิเคราะห์ภาพ เซลล์มะเร็งเต้านมแบบอัตโนมัติที่จะกล่าวในบทที่ 3

2.7 สถิติพื้นฐาน (Statistics)

สถิติ หมายถึง ศาสตร์ที่ว่าด้วยวิธีการเก็บรวบรวมข้อความจริงและตัวเลขที่แสดงข้อเท็จจริง ซึ่งเรียกว่า ข้อมูล การนำเสนอข้อมูล การวิเคราะห์ข้อมูล และการตีความตลอดจนการสรุปผลข้อมูล เพื่อใช้เป็นประโยชน์ในการตัดสินใจที่มีเหตุผล

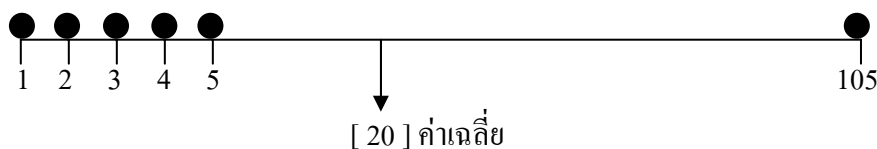
ในการนำเสนอข้อมูลทั้งหมด ถ้ามีข้อมูลจำนวนมาก อาจทำให้ผู้อ่านหรือผู้ใช้เข้าใจได้ยาก จึงมีการนำเสนอค่าตัวแทนของข้อมูล เพื่อให้เห็นลักษณะของข้อมูลจากค่าตัวแทนของข้อมูลค่าใดค่าหนึ่ง ค่าตัวแทนของข้อมูล หรือการวัดแนวโน้มเข้าสู่ส่วนกลางที่นิยมใช้ได้แก่ ค่าเฉลี่ย (Mean, \bar{x}) และ ค่ามัธยฐาน (Median) เป็นต้น (R. Witte and J. Witte, 2004)

2.7.1 ค่าเฉลี่ย (Mean)

คือ การหาผลรวมของข้อมูลทั้งหมดและหารด้วยจำนวนข้อมูลทั้งหมด ดังสูตร

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i \tag{2-8}$$

ข้อมูลที่อยู่ห่างจากข้อมูลที่มีส่วนใหญ่มาก จะทำให้ผลการหาค่าเฉลี่ยคาดเคลื่อนไปจากข้อมูลที่มีอยู่จริง เช่น ชุดข้อมูลประกอบด้วย 1 2 3 4 5 หาค่าเฉลี่ยได้ คือ 3 ซึ่งจะนำมาใช้เป็นจุดศูนย์กลางของกลุ่ม แต่หากมีชุดข้อมูลเป็น 1 2 3 4 5 105 มีข้อมูล 105 เพิ่มเข้ามา ซึ่งเป็นข้อมูลที่ห่างจากข้อมูลที่มีส่วนใหญ่มาก จะทำให้หาค่าเฉลี่ยของข้อมูลชุดนี้ได้เป็น 20 ซึ่งค่า 20 ไม่ได้อยู่ในกลุ่มของข้อมูลส่วนใหญ่ที่มีในชุดข้อมูล จึงส่งผลให้จุดศูนย์กลางที่ได้คาดเคลื่อนไปจากที่ควรจะเป็น แสดงได้ดังภาพประกอบ 2.19



ภาพประกอบ 2.19 ตัวอย่างการหาค่าเฉลี่ยโดยมีข้อมูลที่ห่างจากข้อมูลที่มีส่วนใหญ่มาก

2.7.2 มัธยฐาน (Median)

คือ ข้อมูลที่อยู่ตำแหน่งกึ่งกลางเมื่อมีการเรียงข้อมูลตามลำดับจากมากไปน้อย หรือจากน้อยไปมาก โดยมี 50 % ของจำนวนข้อมูลทั้งหมดมีค่าสูงกว่า และอีก 50 % ของจำนวนข้อมูลทั้งหมดมีค่าต่ำกว่า การหามัธยฐานโดยใช้ข้อมูลคิบบ แบ่งเป็น 2 กรณี คือ

1) ถ้าจำนวนข้อมูลเป็นเลขคี่

มัธยฐาน คือ ข้อมูลที่อยู่ตรงกลาง หรือข้อมูลที่มีค่าอยู่ที่ตำแหน่ง $\frac{1}{2}(N+1)$ เมื่อ N คือ จำนวนข้อมูลทั้งหมด

2) ถ้าจำนวนข้อมูลเป็นเลขคู่

มัธยฐาน คือ ค่าที่อยู่ตรงกลาง 2 ค่ารวมกันแล้วหารด้วย 2 หรือข้อมูลที่มีค่าอยู่ที่ตำแหน่ง $\frac{N}{2}$ และ $\frac{N}{2}+1$ แล้วรวมกันหารด้วย 2

ตัวอย่างการหาค่ามัธยฐานสำหรับชุดข้อมูล 1 2 3 4 5 จะได้ค่ามัธยฐานเป็น 3

แต่สำหรับชุดข้อมูล 1 2 3 4 5 105 จะได้ค่ามัธยฐานเป็น 3.5 ซึ่งเป็นค่าที่อยู่ใกล้เคียงกับข้อมูลส่วนใหญ่ที่มีในชุดข้อมูล โดยจะถือว่าไม่คาดเคลื่อนจากความเป็นจริง จึงเหมาะที่จะนำมาเป็นค่าจุดศูนย์กลางของข้อมูลได้ แสดงได้ดังภาพประกอบ 2.20



ภาพประกอบ 2.20 ตัวอย่างการหาค่ามัธยฐาน โดยมีข้อมูลที่ห่างจากข้อมูลที่มีส่วนใหญ่

2.8 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่อง (Machine Learning, 2009: Online) คือ การทำให้เครื่องเรียนรู้จากข้อมูลตัวอย่างหรือสภาพแวดล้อม เพื่อพัฒนาระบบให้ดีขึ้น เมื่อเรียนรู้แล้วจะเก็บข้อมูลในฐานะข้อมูลด้วยรูปแบบการแทนความรู้ เช่น กฎ หรือ ฟังก์ชัน

การเรียนรู้ของเครื่อง สามารถเรียนรู้ได้เองโดยอัตโนมัติ เหมาะกับงานที่ไม่สามารถสั่งงานเป็นขั้นตอนได้ เช่น การจดจำใบหน้า ต่างจากการใช้อัลกอริทึมที่มนุษย์สร้างขึ้นเป็นกฎที่แน่นอน ข้อมูลตัวอย่างที่จะให้เครื่องเรียนรู้มี 2 ประเภท คือ

1) ตัวอย่างที่มีฉลาก (Labeled Example) คือ ตัวอย่างที่มีคำตอบบอกว่ายู่ในประเภทไหน บอกได้ว่าเป็นสิ่งที่ถูกหรือผิด เรียกว่า การเรียนรู้แบบมีผู้สอน (Supervised Learning)

2) ตัวอย่างที่ไม่มีฉลาก (Unlabeled Example) คือ ตัวอย่างที่ไม่มีคำตอบบอกว่ายู่ในประเภทไหน ถูกหรือผิด จะถูกกำหนดโดยสภาพแวดล้อม อัลกอริทึมจะต้องลองผิดลองถูกเรียนรู้เอาเอง เรียกว่า การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ซึ่งการเรียนรู้จะทำได้ยากกว่า

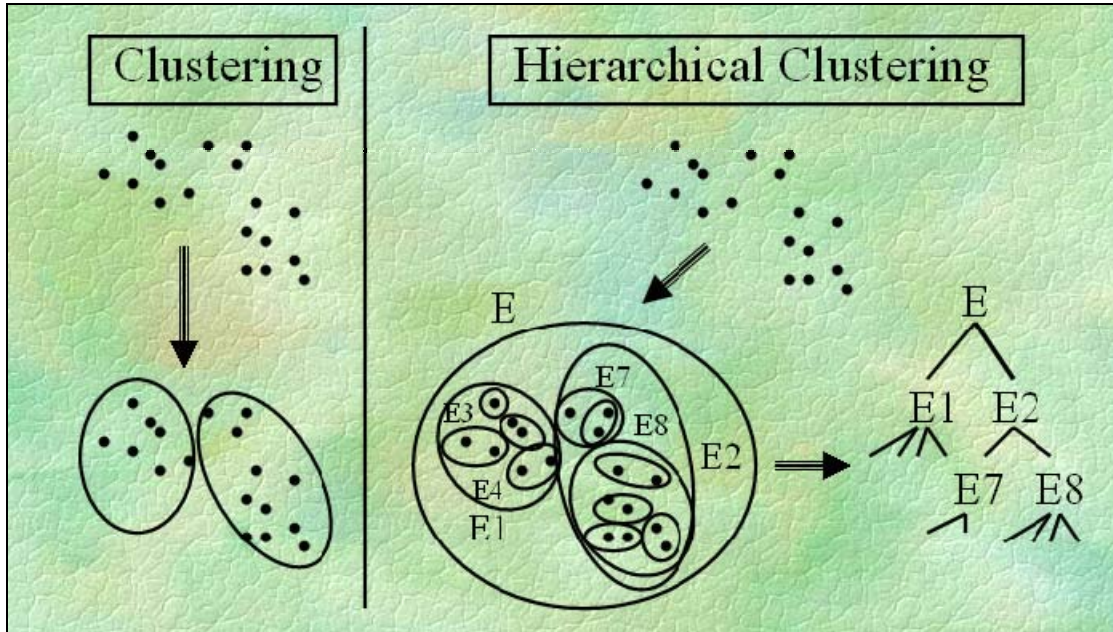
2.8.1 การเรียนรู้แบบมีผู้สอน (Supervised Learning)

การเรียนรู้แบบมีผู้สอน (Wikipedia, 2009: Online) เป็นเทคนิคหนึ่งของการเรียนรู้ของเครื่องซึ่งสร้างฟังก์ชันจากข้อมูลสอน (Training Data) ข้อมูลสอนประกอบด้วยวัตถุเข้า (มักจะเป็น เวกเตอร์) และผลที่ต้องการ ผลจากการเรียนรู้จะเป็นฟังก์ชันที่อาจจะให้ค่าต่อเนื่อง (จะเรียกวิธีการว่า การถดถอย (Regression)) หรือ ใช้ทำนายประเภทของวัตถุ เรียกว่า การแบ่งประเภท (Classification) ภารกิจของเครื่องเรียนรู้แบบมีผู้สอนคือการทำนายค่าของฟังก์ชันจากวัตถุเข้าที่ถูกต้อง โดยใช้ตัวอย่างสอนจำนวนน้อย (Training Examples (คู่ของข้อมูลเข้าและผลที่เป็นเป้าหมาย)) โดยเครื่องเรียนรู้จะต้องวางนัยทั่วไปจากข้อมูลที่มีอยู่ไปยังกรณีที่ไม่เคยพบอย่างมีเหตุผล

2.8.2 การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

การเรียนรู้แบบไม่มีผู้สอน เป็นเทคนิคหนึ่งของการเรียนรู้ของเครื่อง โดยการสร้างโมเดลที่เหมาะสมกับข้อมูล การเรียนรู้แบบนี้แตกต่างจากการเรียนรู้แบบมีผู้สอน คือ จะไม่มีการระบุผลที่ต้องการหรือประเภทไว้ก่อน การเรียนรู้แบบนี้จะพิจารณาวัตถุเป็นเซตของตัวแปรสุ่ม แล้วจึงสร้างโมเดลความหนาแน่นร่วมของชุดข้อมูล

การเรียนรู้แบบไม่มีผู้สอน หรือบางครั้งเรียกว่า Clustering แบ่งได้เป็น 2 ประเภท คือ Partitional Clustering (Clustering) และ Hierarchical Clustering สามารถแสดงให้เห็นความแตกต่างของ 2 ประเภทดังกล่าว ดังแสดงในภาพประกอบ 2.21



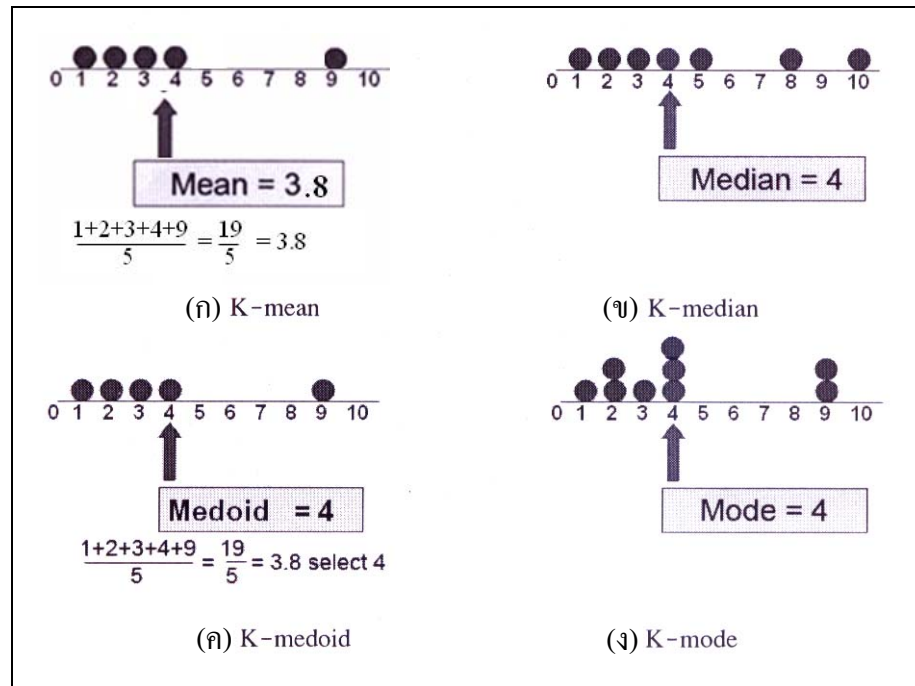
ภาพประกอบ 2.21 ความแตกต่างของ Partitional Clustering และ Hierarchical Clustering

(Clustering, 2009: Online)

ในงานวิจัยนี้ทราบจำนวนกลุ่มที่ต้องการแบ่งกลุ่มอย่างชัดเจน จึงขอกล่าวเฉพาะ Partitional Clustering วิธีนี้จะแบ่งกลุ่มข้อมูลให้มีจำนวนเท่ากับจำนวนกลุ่มที่ต้องการ โดยที่ทุกกลุ่มข้อมูลจะไม่มี การซ้อนทับกัน Partitional Clustering มีขั้นตอนพื้นฐานดังนี้

- 1) หาระดับอย่าง K ตัว ซึ่งใช้เป็นจุดศูนย์กลางหรือตัวแทนกลุ่มในแต่ละกลุ่มของ K กลุ่ม (K คือ จำนวนกลุ่มข้อมูลที่ต้องการแบ่ง)
- 2) หาระดับแต่ละตัวจะถูกกำหนดให้ไปอยู่ในกลุ่มที่มีจุดศูนย์กลางอยู่ใกล้ตัวอย่งนั้นมากที่สุด
- 3) คำนวณหาจุดศูนย์กลางใหม่ของตัวอย่างทั้งหมด
- 4) ทำกระบวนการที่ 2 ซ้ำจนกระทั่งตัวอย่างในกลุ่มไม่มีการเปลี่ยนแปลง

การแบ่งกลุ่มแบบ Partitional Clustering มีขั้นตอนวิธีที่แตกต่างกันหลายวิธี ซึ่งแต่ละ ขั้นตอนวิธีจะมีขั้นตอนการทำงานพื้นฐานเดียวกัน แต่จะแตกต่างกันที่การเลือกค่าศูนย์กลางหรือตัวแทนกลุ่ม ซึ่งมีรูปแบบต่างๆดังนี้ (ดูภาพประกอบ 2.22)



ภาพประกอบ 2.22 รูปแบบการเลือกค่าศูนย์กลางหรือตัวแทนกลุ่ม

2.8.2.1 K-Means Clustering

เป็นการนำค่าเฉลี่ยของข้อมูลมาเป็นค่าศูนย์กลางกลุ่มหรือตัวแทนกลุ่ม ตัวอย่างเช่น ในภาพประกอบ 2.22(ก) นำข้อมูลทั้ง 5 มาหาค่าเฉลี่ยจะได้ว่า ข้อมูลดังกล่าวมีค่าเฉลี่ยเท่ากับ 3.8 จึงนำค่า 3.8 มาเป็นศูนย์กลางหรือตัวแทนกลุ่ม

2.8.2.2 K-Median Clustering

เป็นการนำค่ากึ่งกลางของข้อมูลที่ถูกจัดเรียงลำดับแล้วมาเป็นค่าศูนย์กลางหรือตัวแทนกลุ่ม ตัวอย่างเช่น ในภาพประกอบ 2.22(ข) นำข้อมูลทั้ง 7 มาจัดเรียงข้อมูล ตามลำดับ จากนั้นแยกข้อมูลออกให้เท่าๆกันและนำค่าตรงกลางคือ 4 มาเป็นค่าศูนย์กลางหรือตัวแทนกลุ่ม

2.8.2.3 K-Medoids Clustering

เป็นการนำค่าจริงของข้อมูล ที่มีค่าใกล้เคียงกับค่าเฉลี่ยมาเป็นค่าศูนย์กลางหรือตัวแทนกลุ่ม ตัวอย่างเช่น ในภาพประกอบ 2.22(ค) นำข้อมูลทั้ง 5 มาหาค่าเฉลี่ยจะได้ว่า ข้อมูลดังกล่าวมีค่าเฉลี่ยเท่ากับ 3.8 แต่ค่า 3.8 ไม่มีค่าจริง ดังนั้นจะนำค่าจริงที่มีอยู่คือค่า 4 เป็นศูนย์กลางหรือตัวแทนกลุ่ม

2.8.2.4 K-Mode Clustering

ลักษณะข้อมูลที่ใช้ใน K-Mode จะเป็นข้อมูลที่นำมาคำนวณไม่ได้หรือที่สามารถวัดค่าได้ยาก เช่น สี เพศ เป็นต้น ดังนั้น K-Mode จะนำค่าที่มีความถี่มากที่สุดมาเป็นค่าศูนย์กลางหรือตัวแทน

กลุ่ม ตัวอย่างเช่น ในภาพประกอบ 2.22(ง) ค่าที่มีความถี่มากที่สุดคือ ค่า 4 ดังนั้น จึงนำค่า 4 มาเป็นศูนย์กลางหรือตัวแทนกลุ่ม

ในงานวิจัยนี้เลือกใช้ *K-Means Clustering* และ *K-Median Clustering* ในขั้นตอนของการแยกเซลล์ออกจากภาพ เนื่องจากสามารถนำมาใช้กับชุดข้อมูลที่เป็นภาพเซลล์มะเร็งได้ ซึ่งได้ผลการแยกเซลล์ค่อนข้างดี

2.9 การแบ่งแยกภาพ (Image Segmentation)

Image Segmentation เป็นกระบวนการในการแยกกลุ่มของพิกเซล (Pixels) ในภาพ ภาพจะถูกแบ่งออกเป็นส่วนย่อยๆ ของพิกเซลในแต่ละกลุ่ม ตามลักษณะสำคัญที่เราพิจารณา ซึ่งเป็นขั้นตอนสำคัญในการค้นหาและจัดจำรูปแบบของการประมวลผลภาพ สามารถใช้สีในการแบ่งแยกวัตถุได้เรียกว่า Color Segmentation ประโยชน์ของ Image Segmentation ได้แก่

- 1) ลดจำนวนข้อมูลในรูปภาพที่ไม่จำเป็นต่อการวิเคราะห์
- 2) จัดระเบียบข้อมูลในรูปภาพให้เป็นกลุ่มได้ดีขึ้น
- 3) แสดงข้อมูลในรูปที่เข้าใจง่ายขึ้น

หลักการใหญ่ 2 หลักการในการแยกองค์ประกอบของภาพ

- 1) การแยกองค์ประกอบตามความคล้ายกัน (Similarity) ของคุณสมบัติพิกเซลของรูปภาพภายในพื้นที่เดียวกัน
- 2) การแยกองค์ประกอบโดยดูจากความไม่ต่อเนื่อง (Discontinuity) ของคุณสมบัติพิกเซลบริเวณรอยต่อระหว่างวัตถุในภาพกับฉากหลัง

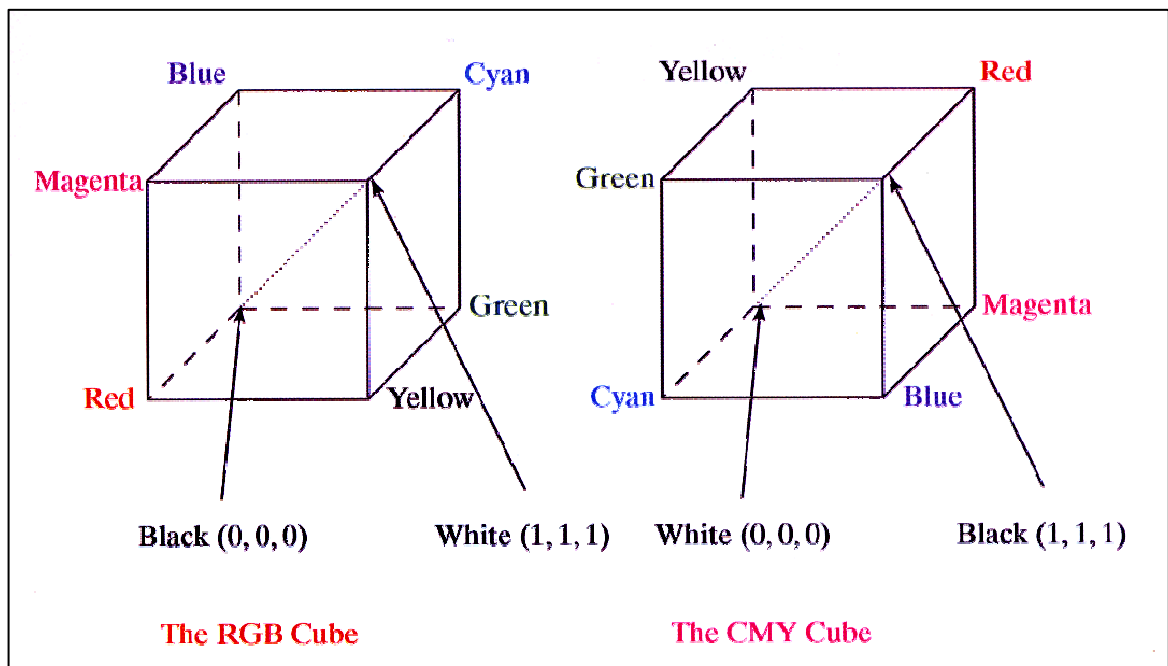
2.10 ระบบสี (Color Model)

ระบบสีสามารถแบ่งได้เป็นหลายประเภท เช่น 1) YIQ Color Model เป็นระบบสีที่ใช้กับโทรทัศน์ ซึ่งเก็บค่าเป็นแบบแถบสัญญาณ (MHz.) 2) RGB หรือ Additive Color Model เป็นระบบสีที่ได้จากการรวมสีหรือผสมสีเข้าด้วยกัน ถ้าทุกสีผสมกันทั้งหมดก็จะได้สีขาว ถ้าไม่ปรากฏสีใดๆเลยก็จะเป็นสีดำ ใช้ในการทำงานของจอภาพ Computer 3) CMYK หรือ Subtractive Color Model เป็นการผสมสีที่ตรงกันข้ามกับระบบ Additive Color โดยเป็นการจำลองการทำงานของแสงสะท้อน และใช้ในงานด้านการพิมพ์ 4) HSB หรือ HSL Model เป็นระบบสีที่มีการคำนึงถึงความสว่างของสีด้วย ซึ่งจะใช้ในงานกราฟิกชั้นสูง ระบบสีที่พบเห็นได้บ่อยที่สุดได้แก่ ระบบสี RGB และระบบสี CMYK

2.10.1 RGB Color Cube (RGB Color Model)

ใน RGB Color Model (ภาพประกอบ 2.21) นั้นแต่ละสีที่เห็นใน Model มีส่วนประกอบหลักคือ สีแดง เขียวและน้ำเงิน (R, G, B) สีในแต่ละบริเวณแสดงดังภาพประกอบ 2.23 ซึ่งค่า R, G, B เป็นค่าที่อยู่ที 3 มุม Cyan, Magenta และ Yellow อยู่ที่มุมอีก 3 มุม Black เป็นจุดเริ่มต้น/จุดกำเนิด และ White อยู่ที่มุมซึ่งห่างจากจุดเริ่มต้นมากที่สุด ในโมเดลนี้ค่า Gray เป็นบริเวณที่อยู่ระหว่าง Black ถึง White สีที่มีความแตกต่างกันนี้เป็นจุดที่อยู่บนลูกบาศก์นี้ และแทนด้วยเวกเตอร์โดยเริ่มจากจุดกำเนิด ค่า RGB ทั้งหมดอยู่ในช่วง $[0, 1]$ ดังแสดงในภาพประกอบ 2.23

การระบุค่าสีแต่ละสีจะใช้วิธีบอกค่าสีดังกล่าวมีส่วนประกอบต่างๆ เป็นจำนวนเท่าไร โดยแต่ละแม่สีจะมีค่าตั้งแต่ 0 คือ ไม่มีสีเลย ไปจนถึง 255 คือมากที่สุด เช่น สีเหลืองประกอบด้วยสีแดง 255 และสีเขียว 255 สีน้ำเงิน 0 ดังนั้นเวกเตอร์สีสีเหลืองดังกล่าวจึงเรียกว่า 255, 255, 0 ($R=255, G=255, B=0$) เป็นต้น ยิ่งใส่ส่วนผสมแต่ละสีมากขึ้นจะทำให้สว่างขึ้น (Rafael C. Gonzalez *et al.*, 2002)



ภาพประกอบ 2.23 RGB and CMYK Color Cube (Rafael C. Gonzalez *et al.*, 2002)

ภาพสี RGB จะประกอบด้วย 3 องค์ประกอบ อย่างหนึ่งคือสีหลัก เมื่อฉายไปยัง RGB Monitor แล้วส่วนประกอบ 3 อย่างจะนำมารวมกันเพื่อกลายเป็นผลลัพธ์ขององค์ประกอบสีของภาพ จำนวนบิตที่ใช้แสดงแต่ละพิกเซลในบริเวณของ RGB เรียกว่า “Pixels Depth” ในภาพ RGB แต่ละสี Red, Green และ Blue มีขนาด 8 บิต รวมแล้วค่า (R, G, B) มี Pixels Depth คือ 24 บิต ดังนั้นจำนวนทั้งหมดของสีใน 24 บิต คือ $(2^8)^3=16,772,216$

โมเดลสีแบบ RGB นี้สามารถสร้างให้เกิดสีได้จำนวนมากที่สุดใกล้เคียงกับสีที่ตาสามารถมองเห็นได้ แต่มีข้อเสียคือจำนวนสีทั้งหมดที่สร้างขึ้นมา จะมีเพียงบางส่วนเท่านั้นที่สามารถพิมพ์ออกทางงานพิมพ์จริงได้ เนื่องจากสีบางสีไม่สามารถสร้างขึ้นได้จริงๆ โดยใช้หมึกพิมพ์ได้ อุปกรณ์ที่ใช้สี RGB ได้แก่ โทรทัศน์ กล้องถ่ายภาพเคลื่อนไหว เครื่องสแกนภาพ และกล้องดิจิทัล เป็นต้น

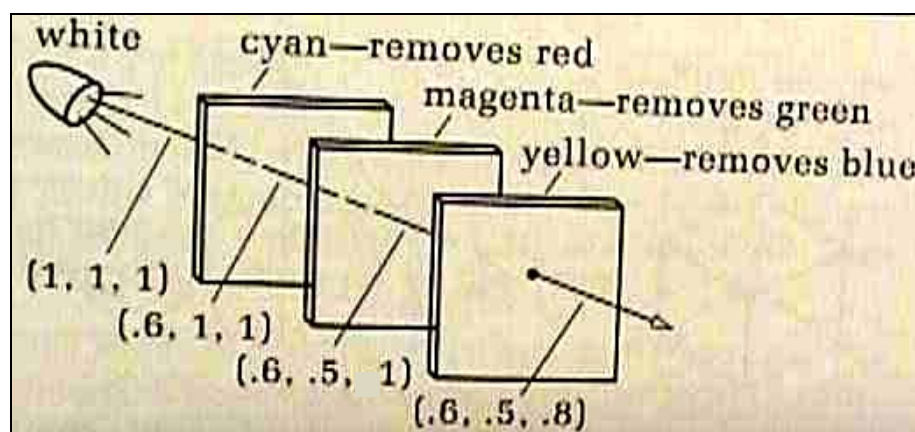
2.10.2 CMYK Color Model

CMYK ย่อมาจาก Cyan, Magenta, Yellow, and Key (Black) เป็นการจำลองการทำงานของแสงสะท้อน และผสมสีจากสีที่ใช้ย้อม เช่น การพิมพ์ภาพจากเครื่องพิมพ์ แบบหมึกพิมพ์ สีที่อยู่ระหว่างสีน้ำเงินกับเขียว (Cyan (100% Cyan = 0 Red)), สีแดงม่วง (Magenta (100% Magenta = 0 Green)) และสีเหลือง (Yellow (100% Yellow = 0 Blue))

การผสมสีของโมเดลชนิดนี้ต่างจากสีแบบ RGB คือ ยิ่งเพิ่มสีลงไป ผลลัพธ์ที่ได้จะมีลดลงเข้าใกล้สีดำมากขึ้น การระบุค่าสีของแม่สีทั้ง 4 จะกำหนดหน่วยวัดเป็นค่าเปอร์เซ็นต์ มีค่าตั้งแต่ 0 เปอร์เซ็นต์ หรือ ไม่มีสี (เป็นสีขาว) ไปจนถึง 100% คือ มีส่วนประกอบของสีดังกล่าวเต็มที่ เช่น สีแดง ประกอบด้วยสีบานเย็น (Magenta) 100% และสีเหลือง (Yellow) 100% ดังนั้นการระบุสีจึงเรียกว่า C = 0, M = 100, Y = 100, K = 0 เป็นต้น

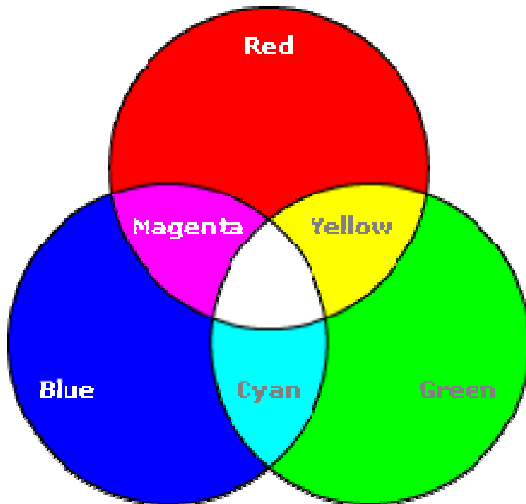
เนื่องจากเมื่อผสมแม่สีทั้งสามสีแบบ 100% แล้วไม่สามารถสร้างสีที่เป็นสีดำอย่างแท้จริงได้ (จะมีลักษณะเป็นสีออกม่วงเข้ม) ในโมเดลสีนี้จึงต้องมีแม่สีสีดำเพิ่มเข้ามา เพื่อช่วยเพิ่มความเข้มให้กับสีต่างๆ และยังสามารถสร้างสีดำที่จริงได้ สีดำจึงถูกเรียกอีกชื่อหนึ่งว่า สีหลัก (Key Color)

โมเดลสีแบบ CMYK ใช้สำหรับภาพที่ต้องการจะแสดงออกทางเครื่องพิมพ์ หรืองานพิมพ์ต่างๆ เช่น งานหนังสือ แผ่นพับ ภาพถ่าย เป็นต้น การแปลงระบบ RGB เป็น CMYK ทำได้ดังนี้ $C = 1 - R$, $M = 1 - G$, $Y = 1 - B$, $K = \min\{C, M, Y\}$ CMYK แบบสมบูรณ์ จะเป็น Graphic แบบ 32 bits 2^{32} สี (Full Color)

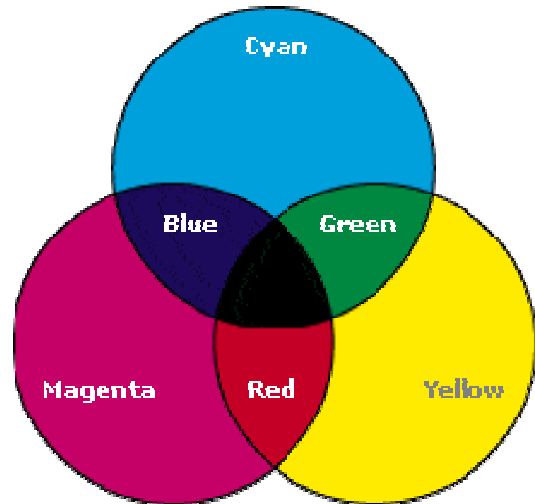


ภาพประกอบ 2.24 กระบวนการดูดซับสี (Subtractive Process) (Maria Larrondo, 2008: Online)

จากภาพประกอบ 2.24 เมื่อแสงสีขาวของ RGB ที่มีค่า (1, 1, 1) ผ่านมายังสี Cyan ที่มีค่า (0, 1, 1) ส่วนประกอบสีแดงก็จะถูกลบหายไป 40 % จะเหลือค่าสี RGB เป็น (.6, 1, 1) จากนั้นผ่าน Magenta ที่มีค่า (0, .5, 0) ส่วนประกอบสีเขียวจะถูกลบหายไป 50 % จึงเหลือค่าสี RGB เป็น (.6, .5, 1) และเมื่อผ่าน Yellow จะถูกลบแสงสีเหลืองออก 20% ดังนั้นจะกลายเป็น (.6, .5, .8) ของ RGB (Maria Larrondo, 2008: Online)



ภาพประกอบ 2.25 RGB Color
(Color Models, 2008: Online)



ภาพประกอบ 2.26 CMYK Color
(Color Models, 2008: Online)

2.11 รูปแบบไฟล์ภาพ (Image File Formats)

ประเภทของไฟล์ภาพที่ใช้กันทั่วไปมีอยู่หลายแบบ ขึ้นอยู่กับลักษณะการบีบอัดและการนำไปใช้ เช่น

1) BMP (Bitmap)

ไฟล์ .BMP เป็นไฟล์มาตรฐานที่ระบบปฏิบัติการ Windows สร้างขึ้นมาสามารถแสดงรายละเอียดสีได้ละเอียดและแม่นยำ แต่มีขนาดไฟล์ใหญ่มาก จึงไม่นิยมใช้

2) JPEG (Joint Photographics Expert Group)

เป็นมาตรฐานการเก็บข้อมูลรูปภาพที่ถูกกำหนดโดย Joint Photographic Experts Group ซึ่งเป็นหน่วยงานร่วมของ ISO/CCITT โดยมีนามสกุล (File Extension) ที่นิยมใช้คือ .jpeg, .jpg, .jpe, .jfif, .jfi ไฟล์ประเภทนี้สามารถเก็บภาพที่มีความละเอียดสูงได้โดยใช้ขนาดไฟล์ที่เล็กสามารถเก็บภาพสีได้หลากหลายระดับความแม่นยำของสี (Bit Depth) ความสามารถในการย่อขนาดไฟล์ของแฟ้ม JPEG นั้นเกิดจากการใช้เทคนิคการย่อขนาดภาพแบบการบีบอัดคงข้อมูลหลัก (Lossy

Compression) หรือการบีบอัดแบบมีความสูญเสียทำให้ไม่นิยมใช้กับภาพที่เป็นลายเส้นหรือไอคอนต่างๆ (wikipedia, 2009: Online)

ในการใช้งานทั่วไปของ JPEG จะเหมาะสำหรับภาพที่มีความต่อเนื่องของสี เช่น ใช้สำหรับเก็บภาพถ่าย (Photographic Images) เนื่องจากการบีบอัดในมาตรฐานนี้มีการสูญเสีย แต่ในภาพถ่ายซึ่งมีจำนวนสีมากและมีความต่อเนื่องของสี ทำให้สายตาคนแยกการสูญเสียในภาพคุณภาพสูงไม่ได้ ข้อดีของ JPEG คือ ไม่ต้องรอให้ Load ภาพมาเก็บไว้ให้หมดก่อนแล้วค่อยแสดงผล โดยมากแล้ว JPG Format จะสามารถลดขนาดของรูปภาพได้มากที่สุด เมื่อเทียบกับ GIF และ PNG Format

ไฟล์ .JPG สามารถแสดงรายละเอียดสีได้ถึง 16.7 ล้านสี เหมาะสำหรับรูปภาพที่มีรายละเอียดของสีจำนวนมาก เช่น รูปถ่าย หรือรูปที่ได้จากการสแกนเข้ามา และมีขนาดเล็กกว่า .BMP มาก ถึงแม้จะไม่สามารถแสดงรายละเอียดของภาพได้เท่า .BMP ก็ตาม ทำให้นิยมใช้มากกว่า .BMP

3) GIF (Graphic Interchange Format)

เป็นมาตรฐานการเก็บข้อมูลรูปภาพซึ่งถูกกำหนดโดย CompuServe มีนามสกุล (File Extension) คือ .gif ในการบีบอัดของ GIF จะเป็นประเภทไม่สูญเสีย (Lossless Compression)

ไฟล์ .GIF สามารถแสดงรายละเอียดสีได้ 256 สี ด้วยเหตุนี้มาตรฐาน GIF จึงไม่นิยมใช้กับภาพถ่ายซึ่งเป็นรูปที่มีจำนวนสีมากกว่าและมีความต่อเนื่องของสี เหมาะสำหรับรูปภาพที่มีจำนวนสีไม่มาก เช่น รูปภาพการ์ตูน เป็นต้น จะใช้ได้ดีกับภาพกราฟิก สัญลักษณ์ต่างๆ หรือภาพแอนิเมชันเคลื่อนไหวง่ายๆ ที่มีสีไม่ต่อเนื่องและมีจำนวนสีไม่มากนัก ซึ่งไฟล์ .GIF นี้มีขนาดเล็กที่สุดเมื่อเทียบกับ .BMP และ .JPG (JPEG & GIF, 2009: Online)

4) PNG (Portable Network Graphics)

ไฟล์ .PNG สามารถแสดงรายละเอียดสีได้ถึง 16 ล้านสี และมีขนาดเล็กเหมือน .GIF .PNG เป็นไฟล์ที่มีความยืดหยุ่นสูง ใช้งานได้กับเครื่องที่มีการเปลี่ยนแพลตฟอร์มการทำงาน และสามารถทำงานอยู่บนคนละระบบปฏิบัติการ เช่น Linux และ Windows

จะเห็นได้ว่าไฟล์แต่ละประเภทยุ่มีลักษณะการทำงานและการบีบอัดไฟล์ที่ต่างกันออกไป ขึ้นอยู่กับการนำมาใช้งานและความต้องการของผู้ใช้งานว่าต้องการไฟล์แบบใด (navy22, 2009: Online) และเมื่อเปรียบเทียบระหว่าง GIF และ JPEG จะได้ว่าภาพแบบ GIF จะเหมาะกับภาพที่ไม่ต้องการความเป็นธรรมชาติ การเปลี่ยนแปลงของสีภายในจะมีลักษณะแตกต่างกันโดยชัดเจน ซึ่งเหมาะกับการนำภาพประเภทนี้มาทำเป็นการ์ตูนหรือแอนิเมชัน ส่วนภาพแบบ JPEG นั้นจะมีคุณภาพดีเมื่อภาพมีการเปลี่ยนแปลงของสีภายในภาพไม่แตกต่างกันมากนัก โดยเน้นที่ความเหมือนธรรมชาติ และความเหมือนจริงมากกว่าโดยสามารถแสดงความแตกต่างของไฟล์แต่ละรูปแบบได้ดังตาราง 2.1

ตารางที่ 2.1 รูปแบบไฟล์ภาพ

	.BMP	.JPEG	.GIF	.PNG
ขนาดไฟล์	ใหญ่	เล็ก	เล็ก	เล็ก
ความละเอียด	24-bit images to 32 bits per pixel.	ละเอียดสูง 16.7 ล้านสี	256 สี	16 ล้านสี
การบีบอัด	Uncompress	Lossy	Lossless	Lossless
ความเหมาะสม	Logo, ภาพบน Internet	เหมาะกับภาพที่มีความ ต่อเนื่องกันของสี, ภาพเหมือนจริง	ภาพที่มีจำนวนสี ไม่มาก, แอนิเมชัน	ภาพกราฟิก

ในงานวิจัยนี้ได้ใช้รูปแบบไฟล์ภาพเป็น JPEG Format เพราะ JPEG Format เหมาะสำหรับภาพที่มีความต่อเนื่องกันของสี ขนาดไฟล์เล็ก ซึ่งสามารถนำมาใช้กับภาพเซลล์มะเร็งของงานวิจัยนี้ได้

บทที่ 3

ระบบการตัดแยกภาพเซลล์มะเร็งเต้านมแบบอัตโนมัติ

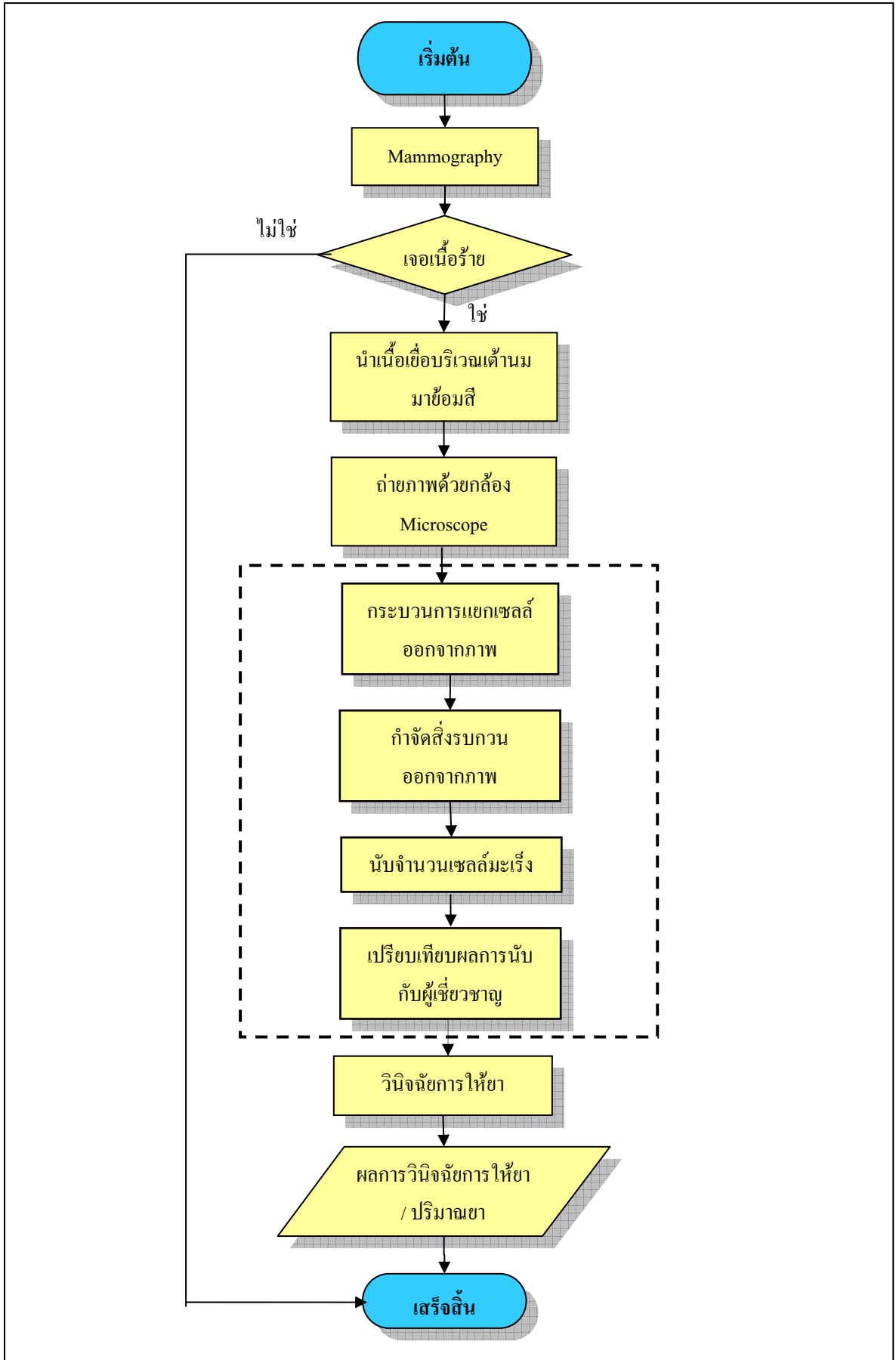
วิทยานิพนธ์นี้มุ่งเน้นการสร้างระบบการตัดแยกภาพเซลล์มะเร็งเต้านมและสามารถนับจำนวนเซลล์มะเร็งเต้านมในภาพได้ เพื่อให้แพทย์สามารถนำไปเป็นประโยชน์ในการช่วยวิเคราะห์ผลการให้ยาแก่ผู้ป่วยที่เป็นโรคมะเร็งเต้านม

ระบบการตัดแยกภาพเซลล์มะเร็งเต้านมแบบอัตโนมัติ (Automated Breast Cancer Cell Image Segmentation: ABC_CIS) ออกแบบมาเพื่อให้แพทย์สามารถนำไปใช้ในการแยกภาพเซลล์มะเร็งเต้านมและนับจำนวนเซลล์มะเร็งเต้านมได้ ซึ่งเป็นประโยชน์ในการนำไปวิเคราะห์ผลการให้ยาแก่ผู้ป่วยได้ เป็นการให้ความสะดวกแก่แพทย์ในขั้นตอนหนึ่ง

กระบวนการวินิจฉัยมะเร็งเต้านมและการให้ยาของแพทย์นั้นมีผังงานดังแสดงในภาพประกอบ 3.1 วิทยานิพนธ์นี้เป็นส่วนหนึ่งของกระบวนการวินิจฉัยมะเร็งเต้านมและการให้ยา ซึ่งในภาพประกอบ 3.1 ส่วนที่เป็นวิทยานิพนธ์จะเป็นบริเวณที่ล้อมด้วยเส้นประ ซึ่งประกอบไปด้วย 5 ขั้นตอนย่อยที่จะกล่าวต่อไป

ในบทนี้จะกล่าวถึงการออกแบบระบบ ABC_CIS ซึ่งมีกระบวนการแบ่งเป็น 5 ขั้นตอนดังนี้

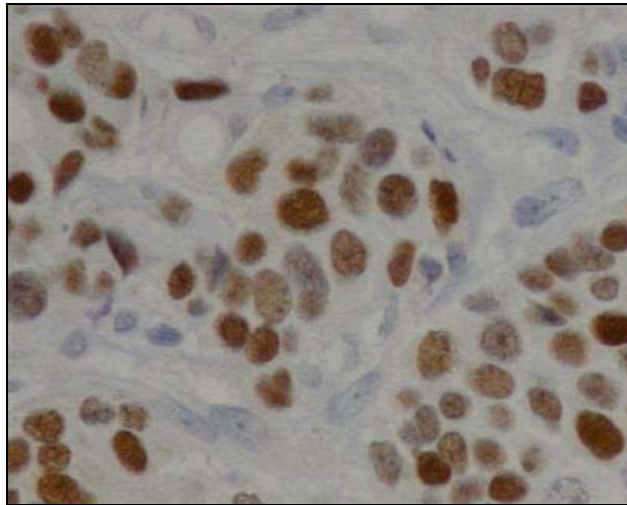
ขั้นตอนที่ 1 : การเตรียมภาพเซลล์มะเร็งเต้านม (Preprocess)
ขั้นตอนที่ 2 : การตัดแยกภาพเซลล์มะเร็งเต้านม (Segmentation)
ขั้นตอนที่ 3 : การกำจัดสิ่งรบกวนในภาพ (Noise Reduction)
ขั้นตอนที่ 4 : การนับเซลล์มะเร็งเต้านม (Cell Count)
ขั้นตอนที่ 5 : การเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ (Compare with Expert)



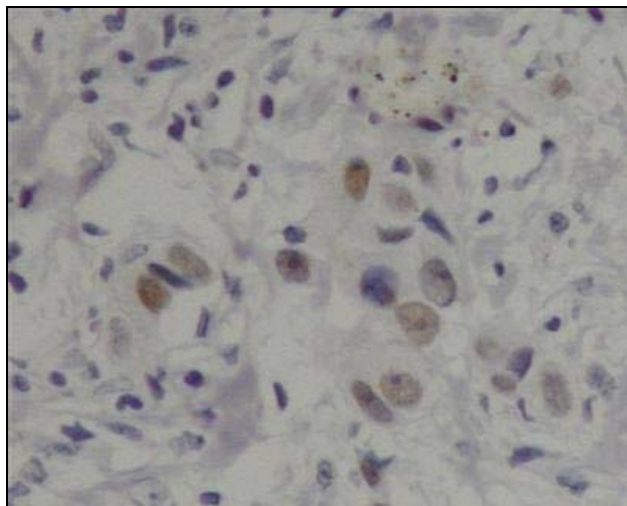
ภาพประกอบ 3.1 ผังงานกระบวนการวินิจฉัยมะเร็งเต้านมและการให้ยา

3.1 ขั้นตอนการเตรียมภาพเซลล์มะเร็งเต้านม

เป็นขั้นตอนการเตรียมภาพเซลล์มะเร็งเต้านมหลังจากที่ได้ภาพต้นฉบับจากผู้เชี่ยวชาญ เพื่อสามารถนำภาพไปแยกเซลล์มะเร็งเต้านมต่อไป โดยมีขั้นตอนย่อยดังนี้ ขั้นตอนที่ 1 คือ การแบ่งภาพออกเป็น 2 กลุ่ม ได้แก่ 1) กลุ่มของภาพที่มีเซลล์ขนาดใหญ่และชัดเจนเป็นส่วนมาก ตัวอย่างดังภาพประกอบ 3.2 2) กลุ่มของภาพที่มีเซลล์ขนาดเล็กและไม่ค่อยชัดเจนเป็นส่วนมาก ตัวอย่างดังภาพประกอบ 3.3



ภาพประกอบ 3.2 ตัวอย่างภาพที่มีเซลล์ขนาดใหญ่และชัดเจนเป็นส่วนมาก



ภาพประกอบ 3.3 ตัวอย่างภาพที่มีเซลล์ขนาดเล็กและไม่ค่อยชัดเจนเป็นส่วนมาก

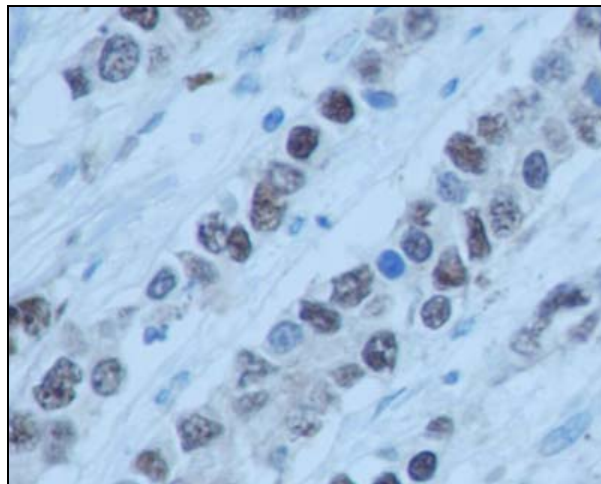
โดยภาพเซลล์ขนาดเล็กและภาพเซลล์ขนาดใหญ่ได้มาจากการที่แพทย์ตัดเนื้อเยื่อบริเวณเต้านมของผู้ป่วยมาผ่านการย้อมสีและถ่ายภาพด้วยกล้องจุลทรรศน์ดิจิทัล ซึ่งในขั้นตอนการตัดเนื้อเยื่อของแพทย์นั้น ถ้าตัดถูกบริเวณกึ่งกลางเซลล์พอดี (สมมติฐานว่าเซลล์มีลักษณะทรงกลม) ก็จะทำให้ในภาพถ่าย

เนื้อเยื่อที่ให้เห็นเป็นเซลล์ขนาดใหญ่และชัดเจน แต่หากว่าตัดดูเนื้อเยื่อบริเวณปลายหรือส่วนใดส่วนหนึ่งของเซลล์ จะทำให้ในภาพถ่ายเนื้อเยื่อที่ให้เห็นเป็นเซลล์ขนาดเล็ก สรุปได้ว่า ภาพเซลล์ขนาดเล็กและภาพเซลล์ขนาดใหญ่ไม่ได้มาจากการย่อหรือขยายภาพจากกล้อง แต่เป็นผลมาจากขั้นตอนการตัดเนื้อเยื่อมะเร็งด้านมาจากแพทย์

เนื่องจากภาพเนื้อเยื่อเซลล์มะเร็งด้านมอาจประกอบด้วยองค์ประกอบอื่นที่มีลักษณะเล็กกว่าเซลล์มะเร็งด้านมปะปนมา เช่น ไขมัน ต่อมไขมัน เป็นต้น จึงเป็นที่มาของการแบ่งภาพออกเป็น 2 ประเภท

ขั้นตอนที่ 2 คือ การแปลงภาพให้เป็นตัวเลขของสี Red, Green, Blue (RGB) และนำตัวเลขมาเรียงต่อกันเป็นรูปแบบที่โปรแกรมสามารถนำไปประมวลผลต่อได้

ภาพเซลล์มะเร็งด้านมที่ผ่านการย้อมสีแล้ว ดังตัวอย่างในภาพประกอบ 3.4 จะถูกแปลงภาพเซลล์มะเร็งด้านมให้เป็นตัวเลขของสี RGB แล้วเรียงให้อยู่ในรูปแบบดังแสดงในภาพประกอบ 3.5 และจึงนำไปผ่านขั้นตอนการตัดแยกภาพเซลล์มะเร็งด้านมต่อไปได้



ภาพประกอบ 3.4 ภาพเซลล์มะเร็งด้านมที่ผ่านการย้อมสีแล้ว

Command Window		
109	101	122
109	101	122
110	102	123
110	102	123
110	102	123
110	102	123
114	106	127
114	106	127
115	107	128
115	107	128
116	108	129
117	109	130
118	110	131
118	110	131
122	112	137
121	111	136
119	109	134
116	106	131
113	103	128
111	101	126
109	99	124
108	98	123
111	101	126
111	101	126
110	100	125

ภาพประกอบ 3.5 ตัวเลข RGB เรียงในรูปแบบที่นำไปใช้งานได้

3.2 ขั้นตอนการตัดแยกภาพเซลล์มะเร็งเต้านม

เป็นขั้นตอนการตัดแยกส่วนของเซลล์มะเร็งเต้านมในภาพ เพื่อแยกส่วนที่เป็นเซลล์มะเร็งออกจากพื้นหลัง ซึ่งจะนำผลที่ได้จากกระบวนการแรกนำมาแบ่งกลุ่มข้อมูล โดยใช้เทคนิคของการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) ได้แก่ *K*-Means Clustering และ *K*-Median Clustering เพื่อเปรียบเทียบผลการแยกเซลล์ออกจากพื้นหลังของภาพและผลการนับเซลล์ แสดงขั้นตอนวิธีดังภาพประกอบ 3.6 ในขั้นตอนนี้ผลที่ได้จะเป็นค่าของตัวเลขกลุ่ม ดังแสดงในภาพประกอบ 3.7

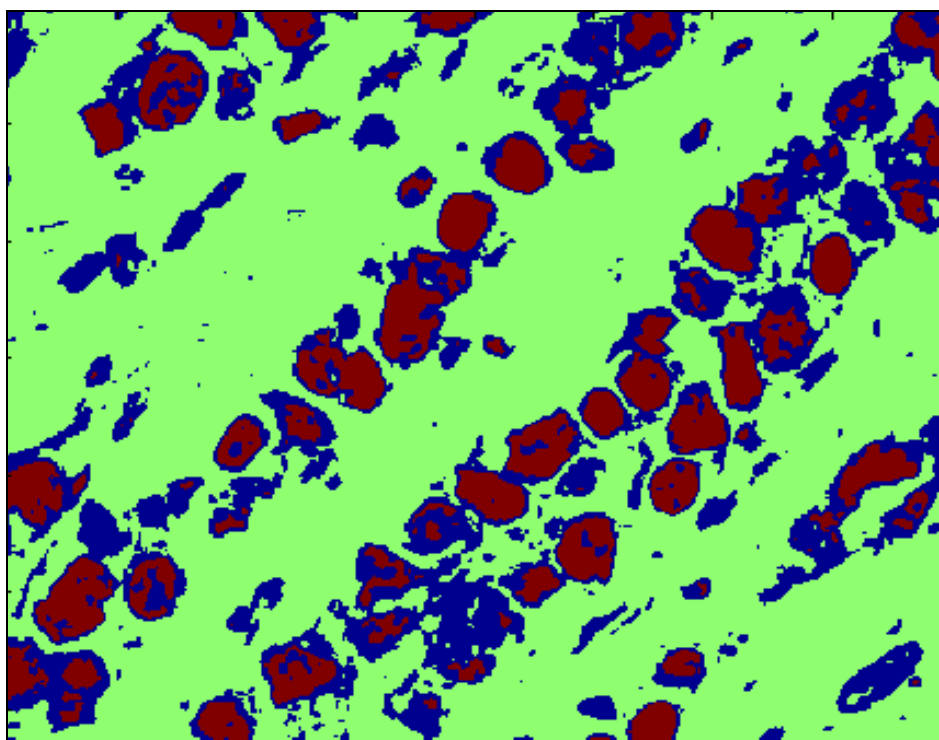
ขั้นตอนที่ 2: การตัดแยกภาพเซลล์มะเร็งเต้านม	
<i>K</i>-Means Clustering	
<ol style="list-style-type: none"> 1) สุ่มหาตัวอย่าง <i>K</i> ตัว ซึ่งจะใช้เป็นจุดศูนย์กลางหรือตัวแทนกลุ่มในแต่ละกลุ่มของ <i>K</i> กลุ่ม (<i>K</i> คือ จำนวนกลุ่มที่ต้องการ) 2) ตัวอย่างแต่ละตัวจะถูกกำหนดให้ไปอยู่ในกลุ่มที่มีจุดศูนย์กลางอยู่ใกล้ตัวอย่งนั้นมากที่สุด 3) คำนวณหาค่าเฉลี่ยใหม่ในแต่ละกลุ่ม เพื่อมาเป็นจุดศูนย์กลางใหม่ของตัวอย่างในกลุ่ม 4) ทำกระบวนการที่ 2 ซ้ำจนกระทั่งตัวอย่างในกลุ่มไม่มีการเปลี่ยนแปลง 	
<i>K</i>-Median Clustering	
<ol style="list-style-type: none"> 1) สุ่มหาตัวอย่าง <i>K</i> ตัว ซึ่งจะใช้เป็นจุดศูนย์กลางหรือตัวแทนกลุ่มในแต่ละกลุ่มของ <i>K</i> กลุ่ม (<i>K</i> คือ จำนวนกลุ่มที่ต้องการ) 2) ตัวอย่างแต่ละตัวจะถูกกำหนดให้ไปอยู่ในกลุ่มที่มีจุดศูนย์กลางอยู่ใกล้ตัวอย่งนั้นมากที่สุด 3) คำนวณหาค่ากึ่งกลางใหม่ในแต่ละกลุ่ม เพื่อมาเป็นจุดศูนย์กลางใหม่ของตัวอย่างในกลุ่ม 4) ทำกระบวนการที่ 2 ซ้ำจนกระทั่งตัวอย่างในกลุ่มไม่มีการเปลี่ยนแปลง 	

ภาพประกอบ 3.6 ขั้นตอนวิธี *K*-Means Clustering และ *K*-Median Clustering

```
Command Window
3
3
3
3
3
3
3
3
3
3
3
1
1
1
1
1
1
1
1
1
2
2
2
2
2
2
2
2
2
```

ภาพประกอบ 3.7 ตัวเลขการจัดกลุ่มเรียงในรูปแบบที่ใช้งานได้

หลังจากได้ตัวเลขของกลุ่มแล้ว จะนำมาทำกลับเป็นภาพผลการแยกเซลล์ดังตัวอย่างในภาพประกอบที่ 3.8



ภาพประกอบ 3.8 ตัวอย่างภาพผลการแยกเซลล์

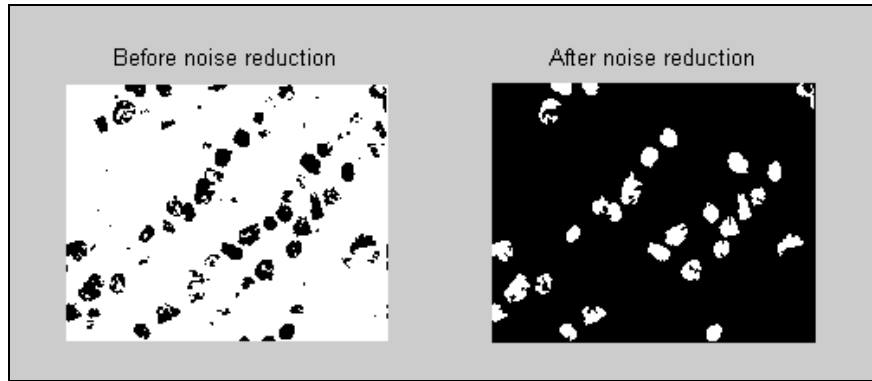
3.3 ขั้นตอนการกำจัดสิ่งรบกวนในภาพ

เป็นขั้นตอนการกำจัดสิ่งที่ไม่จำเป็นออกจากภาพหลังจากผ่านการตัดแยกเซลล์เรียบร้อยแล้ว เนื่องจากภาพที่ผ่านกระบวนการแยกเซลล์ออกจากพื้นหลังของภาพ จะเกิดสิ่งรบกวนในภาพ เมื่อนำไปใช้ในขั้นตอนการนับเซลล์จะทำให้เกิดความผิดพลาดได้สูง จึงต้องกำจัดสิ่งรบกวนในภาพเสียก่อน การกำจัดสิ่งรบกวนในภาพได้แก่ การกำจัดส่วนของเซลล์ที่กระจุกกระจายในภาพและเติมส่วนที่เป็นช่องว่างภายในเซลล์ให้เต็ม เพื่อสามารถนำไปสู่กระบวนการนับเซลล์ต่อไปได้และในขั้นตอนการกำจัดสิ่งรบกวนนี้แบ่งเป็น 2 แบบ สำหรับภาพ 2 กลุ่มที่ได้แบ่งไว้แล้ว ขั้นตอนในการกำจัดสิ่งรบกวนออกจากภาพ แสดงดังภาพประกอบ 3.9 และตัวอย่างผลการกำจัด Noise แสดงดังภาพประกอบ 3.10

ขั้นตอนที่ 3: การกำจัดสิ่งรบกวนในภาพ
ข้อมูลนำเข้า : ภาพที่ผ่านการแยกเซลล์ออกจากพื้นหลังแล้ว
ข้อมูลส่งออก : ภาพที่ผ่านการแยกเซลล์ออกจากพื้นหลังซึ่งปราศจากสิ่งรบกวน (Binary Image)
<ol style="list-style-type: none"> 1) กำหนดขนาดของ Small Cell เป็น 100 (สำหรับภาพเซลล์ขนาดใหญ่) 2) กำหนดขนาดของ Small Cell เป็น 200 (สำหรับภาพเซลล์ขนาดเล็ก) 3) กำหนดขนาดของ Big Cell เป็น 1000 4) ดำเนินการ Morphological Structuring Element 5) เติมเต็มพื้นที่ว่างภายในเซลล์ 6) ระบุส่วนประกอบที่ติดกันใน Binary Image 7) นับจำนวนพิกเซลในบริเวณที่ติดกัน เก็บใส่ Stats 8) สำหรับภาพเซลล์ขนาดใหญ่ <ol style="list-style-type: none"> 8.1) ถ้า $Stats \leq 100$ ให้กำหนด $Size = 0$ (บริเวณที่เป็นสิ่งรบกวน) 8.2) ถ้า $100 < Stats \leq 1000$ ให้กำหนด $Size = 1$ (บริเวณที่ไม่ใช่สิ่งรบกวน) 9) สำหรับภาพเซลล์ขนาดเล็ก <ol style="list-style-type: none"> 9.1) ถ้า $Stats \leq 200$ ให้กำหนด $Size = 0$ (บริเวณที่เป็นสิ่งรบกวน) 9.2) ถ้า $200 < Stats \leq 1000$ ให้กำหนด $Size = 1$ (บริเวณที่ไม่ใช่สิ่งรบกวน) 10) แสดงเฉพาะบริเวณที่มี $Size = 1$ ใน Binary Image

ภาพประกอบ 3.9 ขั้นตอนการกำจัดสิ่งรบกวนในภาพ

จากขั้นตอนที่ 3 การกำหนดขนาดของ Small Cell เกิดจากการทดลองก่อนหน้านี้ที่ได้กำหนดค่า Small Cell เป็น 50 100 150 200 250 และจากผลการนับเซลล์ที่ได้พบว่า Small Cell เป็น 100 จะใช้สำหรับภาพเซลล์ขนาดใหญ่ได้อย่างเหมาะสม และ Small Cell เป็น 200 จะใช้สำหรับภาพเซลล์ขนาดเล็ก โดยค่าที่เลือกจะทำให้ผลการนับเซลล์ถูกต้องกว่าค่าอื่นเมื่อเทียบกับผลของผู้เชี่ยวชาญ



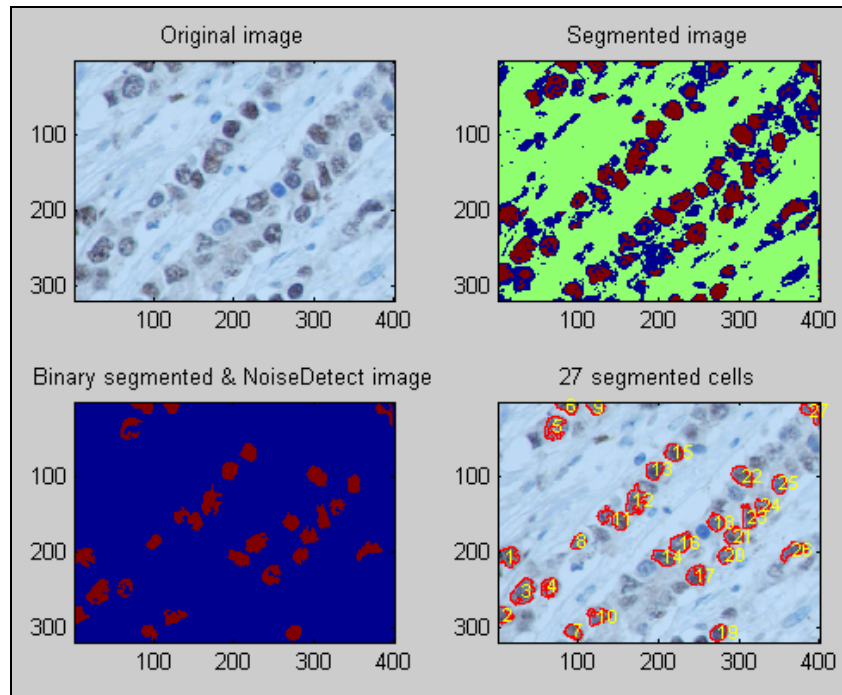
ภาพประกอบ 3.10 ตัวอย่างภาพก่อนการกำจัด Noise และภาพผลการกำจัด Noise

3.4 ขั้นตอนการนับเซลล์มะเร็งเต้านม

เป็นขั้นตอนการนับส่วนที่เป็นเซลล์มะเร็งในภาพ ตัวเลขที่ได้สามารถนำไปเป็นส่วนหนึ่งของการวิเคราะห์การให้ยาแก่ผู้ป่วยที่เป็นมะเร็งเต้านมได้ ภาพประกอบ 3.11 แสดงขั้นตอนการนับเซลล์มะเร็งเต้านม และภาพประกอบ 3.12 แสดงตัวอย่างผลการนับเซลล์

ขั้นตอนที่ 4: การนับเซลล์มะเร็งเต้านม
ข้อมูลนำเข้า : ภาพเซลล์มะเร็งเต้านมต้นฉบับ, ภาพที่ผ่านการแยกเซลล์ออกจากพื้นหลังแล้ว, Binary Image
ข้อมูลส่งออก : ภาพกราฟที่ประกอบด้วย ภาพเซลล์มะเร็งเต้านมต้นฉบับ, ภาพที่ผ่านการแยกเซลล์ออกจากพื้นหลังแล้ว, Binary Image และภาพเซลล์มะเร็งเต้านมต้นฉบับที่ถูกระบุผลการนับเซลล์ในภาพแล้ว
<ol style="list-style-type: none"> 1) ระบุส่วนประกอบที่ติดกันใน Binary Image 2) for K=1 to Num (Num คือ จำนวนส่วนประกอบที่ติดกันในภาพ) <ol style="list-style-type: none"> 2.1) ใส่กรอบล้อมรอบส่วนประกอบที่ติดกัน (จากบนไปล่างและจากซ้ายไปขวา) 2.2) ใส่หมายเลข K ลงในกรอบที่ล้อมรอบส่วนประกอบที่ติดกัน 3) แสดงภาพต้นฉบับ, ภาพที่ผ่านการแยกเซลล์ออกจากพื้นหลังแล้ว, Binary Image, ภาพผลการนับเซลล์

ภาพประกอบ 3.11 ขั้นตอนการนับเซลล์มะเร็งเต้านม



ภาพประกอบ 3.12 ตัวอย่างผลการนับเซลล์

ภาพประกอบ 3.12 ประกอบด้วย

- ภาพบนซ้าย แสดงภาพเซลล์มะเร็งเต้านมต้นฉบับ
- ภาพบนขวา แสดงภาพผลการแบ่งกลุ่มเซลล์หรือตัดแยกภาพเซลล์
- ภาพล่างซ้าย แสดงภาพหลังจากการกำจัด Noise
- ภาพล่างขวา แสดงภาพผลการนับเซลล์ ซึ่งนับได้ 27 เซลล์

3.5 ขั้นตอนการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ

เป็นขั้นตอนเปรียบเทียบผลการนับเซลล์จากโปรแกรมกับผลการนับเซลล์จากผู้เชี่ยวชาญ (ภาพประกอบ 3.14) ในขั้นตอนนี้จะสามารถทำได้ก็ต่อเมื่อ มีผลการนับเซลล์มะเร็งเต้านมในภาพนั้นๆ จากผู้เชี่ยวชาญแล้ว ซึ่งภาพผลการนับเซลล์จากผู้เชี่ยวชาญนี้จะประกอบด้วย ส่วนพื้นหลังซึ่งเป็นภาพเซลล์มะเร็งเต้านมต้นฉบับ กรอบสีเหลืองสีเขียว (Positive Cell) และกรอบสีเหลืองสีแดง (Negative Cell) หรือจากภาพประกอบ 3.13 เป็นบริเวณที่มีเครื่องหมาย * ซึ่ง Positive Cell และ Negative Cell เป็นเครื่องหมายที่ผู้เชี่ยวชาญระบุไว้ในแต่ละเซลล์มะเร็งเต้านมในภาพ ภาพประกอบ 3.13 แสดงตัวอย่างภาพผลการนับจากผู้เชี่ยวชาญ และภาพประกอบ 3.15 แสดงตัวอย่างภาพผลการเปรียบเทียบกับผู้เชี่ยวชาญ

ในขั้นตอนนี้จะใช้การวัดประสิทธิภาพความถูกต้องในการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ โดยใช้ความรู้ของ Confusion Matrix หรือในการเรียนรู้แบบไม่มีผู้สอนจะเรียกว่า Matching Matrix ซึ่งใช้วัดประสิทธิภาพของการทำนาย โดยจะเก็บข้อมูลที่เกี่ยวข้องกับการแบ่งแยกข้อมูลจริงกับข้อมูลที่

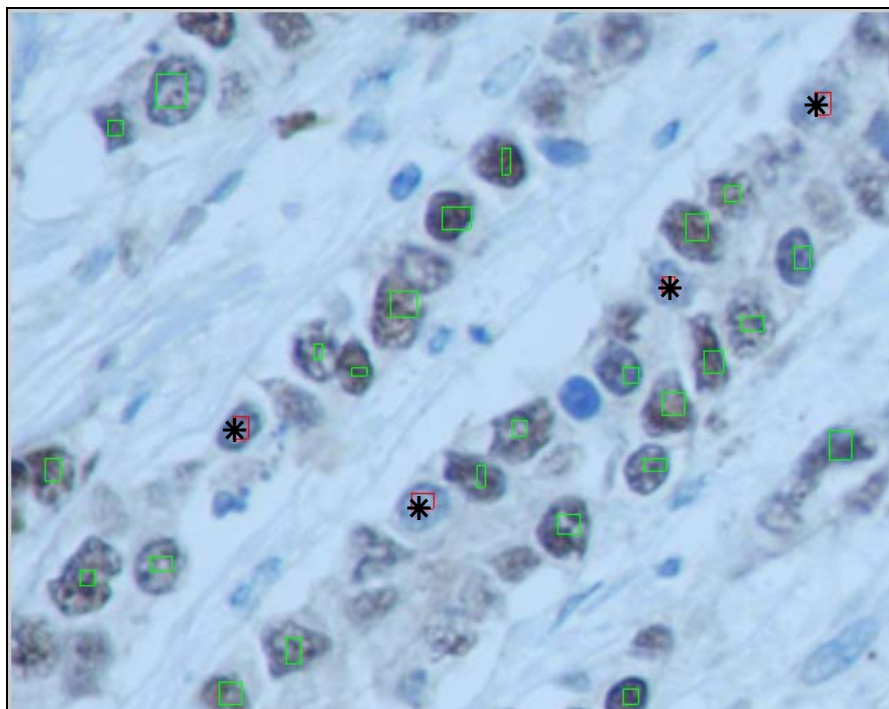
เกิดจากการทำนาย ด้วยระบบการแบ่งแยก (Classification System) เช่น การหาค่าข้อมูลแบบธรรมดาในเมตริกซ์ ดังตาราง 3.1 แสดง Confusion Matrix ซึ่งมีด้วยกัน 2 ระดับ ที่มีค่าความจริงคือ ปฏิเสธ (Negative) และยอมรับ (Positive)

ตาราง 3.1 Confusion Matrix

		ค่าทำนาย (Predicted)	
		ปฏิเสธ (Negative)	ยอมรับ (Positive)
ค่าความจริง (Actual)	ปฏิเสธ (Negative)	a	b
	ยอมรับ (Positive)	c	d

ตาราง 3.1 Confusion Matrix มีความหมายดังนี้

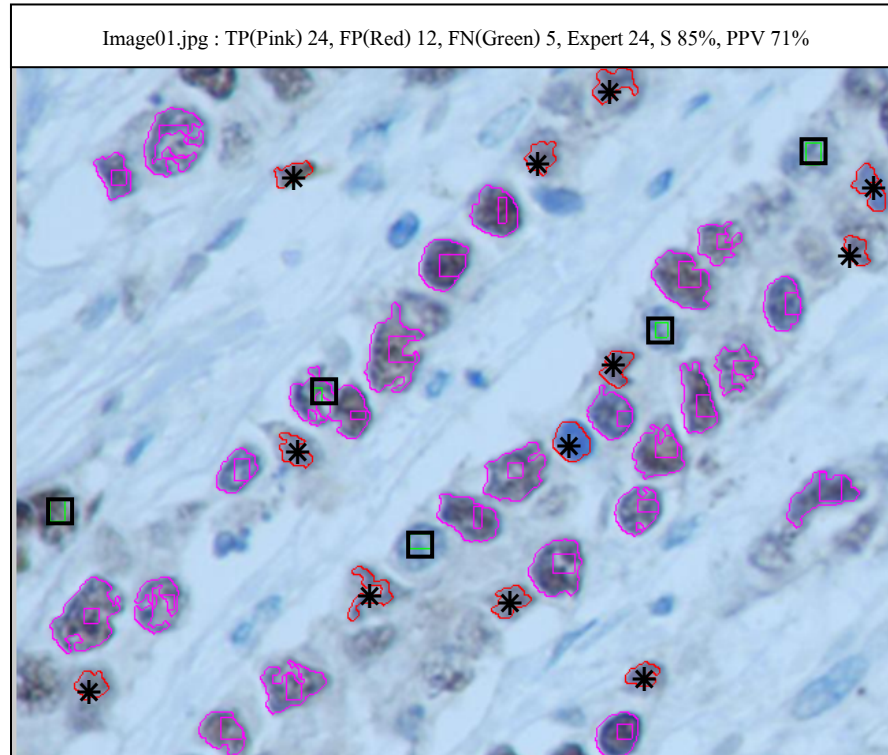
- ค่า a เป็นตัวเลขที่ถูกต้องจากการทำนายว่าปฏิเสธ (True Negative)
- ค่า b เป็นตัวเลขที่ไม่ถูกต้องจากการทำนายว่ายอมรับ (False Positive)
- ค่า c เป็นตัวเลขที่ไม่ถูกต้องจากการทำนายว่าปฏิเสธ (False Negative)
- ค่า d เป็นตัวเลขที่ถูกต้องจากการทำนายว่ายอมรับ (True Positive)



ภาพประกอบ 3.13 ตัวอย่างภาพผลการนับเซลล์ของผู้เชี่ยวชาญ

ขั้นตอนที่ 5: การเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ
ข้อมูลนำเข้า : ภาพเซลล์มะเร็งเต้านมต้นฉบับ, ผลการกำจัดสิ่งรบกวน และภาพผลการนับเซลล์จากผู้เชี่ยวชาญ
ข้อมูลส่งออก : ภาพเซลล์มะเร็งเต้านมต้นฉบับที่มีผลการเปรียบเทียบกับผู้เชี่ยวชาญ และ S, PPV เปอร์เซ็นต์
<p>1) โหลดข้อมูลนำเข้า</p> <p>2) เปรียบเทียบ ผลการกำจัดสิ่งรบกวนกับภาพผลการนับเซลล์จากผู้เชี่ยวชาญ</p> <p>2.1) ถ้าเซลล์ที่โปรแกรมระบุตรงกับที่ผู้เชี่ยวชาญระบุ</p> <ul style="list-style-type: none"> - รอบเซลล์นั้นจะถูกล้อมรอบด้วยสีชมพู หรือ TP (ส่วนที่ล้อมรอบด้วยเส้น จากตัวอย่างในภาพประกอบ 3.15) <p>2.2) ถ้าโปรแกรมระบุเซลล์เกินจากที่ผู้เชี่ยวชาญระบุ</p> <ul style="list-style-type: none"> - รอบเซลล์นั้นจะถูกล้อมรอบด้วยสีแดง หรือ FP (ส่วนที่มีเครื่องหมาย * จากตัวอย่างในภาพประกอบ 3.15) <p>2.3) เซลล์ที่โปรแกรมไม่ได้ระบุ แต่ผู้เชี่ยวชาญระบุ</p> <ul style="list-style-type: none"> - ในเซลล์นั้นจะถูกทำเครื่องหมายสีเหลี่ยมสีเขียว หรือ FN (ส่วนที่มี □ จากตัวอย่างในภาพประกอบ 3.15) <p>3) คำนวณค่า S, PPV เปอร์เซ็นต์ เพื่อเป็นตัววัดความถูกต้องของผลการเปรียบเทียบการนับเซลล์กับผู้เชี่ยวชาญ</p> <p>3.1) $S = TP / (TP + FN)$</p> <p>3.2) $PPV = TP / (TP + FP)$</p>

ภาพประกอบ 3.14 ขั้นตอนการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ



ภาพประกอบ 3.15 ตัวอย่างภาพผลการเปรียบเทียบกับผู้เชี่ยวชาญ

ภาพประกอบ 3.15 จะปรากฏค่าแสดงตัวเลขการคำนวณผลการเปรียบเทียบ ซึ่งประกอบด้วย

- TP (True Positive) คือ เซลล์ที่ระบุได้ตรงกับที่ผู้เชี่ยวชาญระบุมา
- FP (False Positive) คือ เซลล์ที่โปรแกรมนับเกินจากที่ผู้เชี่ยวชาญนับไว้ หรือนับผิดพลาด

ทางบวก

- FN (False Negative) คือ เซลล์ที่โปรแกรมนับขาดไปจากที่ผู้เชี่ยวชาญนับ หรือนับผิดพลาด

ทางลบ

- Expert คือ จำนวนเซลล์ที่ผู้เชี่ยวชาญนับได้
- S (Sensitivity) คือ เซลล์ที่ผู้เชี่ยวชาญนับเปรียบเทียบกับโปรแกรมนับ

$$S = \frac{TP}{(TP + FN)} \quad (3-1)$$

เช่น S = 80 % หมายถึง ถ้าผู้เชี่ยวชาญนับได้ 100 เซลล์ แต่โปรแกรมนับได้แค่ 80 เซลล์

- PPV (Predictive Positive Value) คือ สิ่งโปรแกรมนับจะมีโอกาสที่ผู้เชี่ยวชาญจะบอกว่าเป็นเซลล์มะเร็งก็เปอร์เซ็นต์

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

เช่น PPV = 90 % หมายถึง ถ้าโปรแกรมนับได้ 100 เซลล์ แต่ผู้เชี่ยวชาญนับได้ 90 เซลล์

บทที่ 4

ผลการทดลองและวิจารณ์

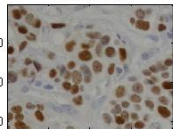
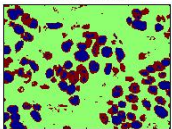
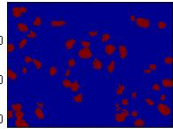
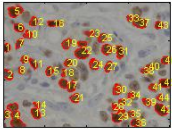
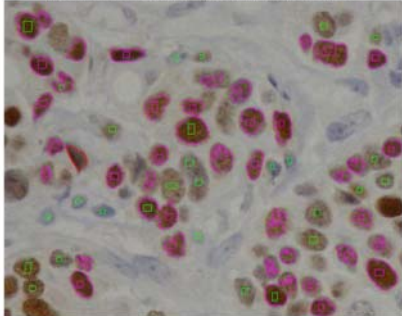
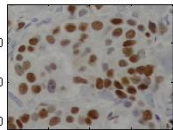
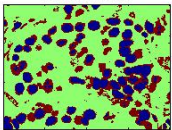
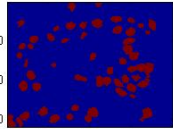
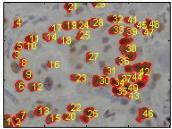
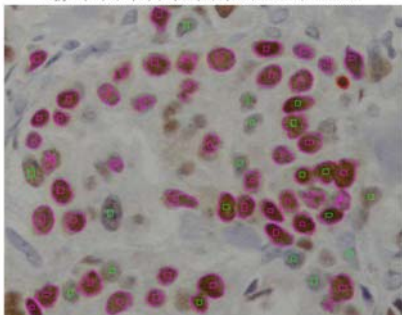
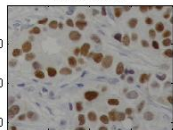
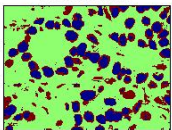
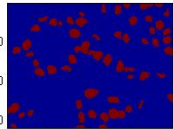
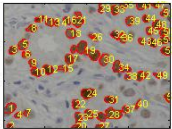
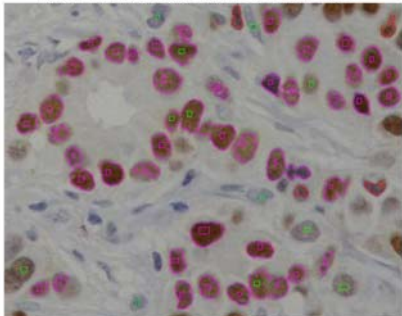
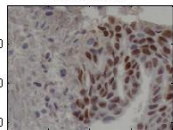
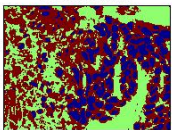
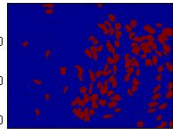
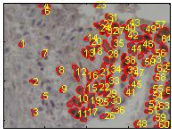
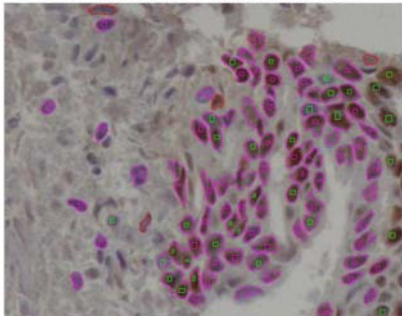
บทนี้จะนำเสนอผลภาพที่ได้จากการทดลองระบบการตัดแยกเซลล์มะเร็งเต้านมแบบอัตโนมัติ (Automated Breast Cancer Cell Image Segmentation: ABC_CIS) ในการทดลองจะใช้ภาพเซลล์มะเร็งเต้านมทั้งหมด 30 ภาพ ซึ่งเป็นภาพที่ได้มาจากผู้ป่วยมะเร็งเต้านมจริง โดยที่แพทย์จะนำเนื้อเยื่อเซลล์มะเร็งเต้านมจากผู้ป่วยมาทำการย้อมสี เพื่อถ่ายเป็นภาพ Microscope จากนั้นผู้ทำวิจัยจึงได้นำภาพมาแปลงเป็นไฟล์ประเภท JPEG ขนาด 400*320 Pixels

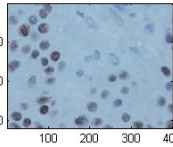
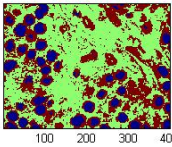
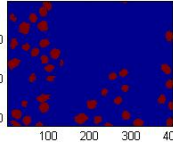
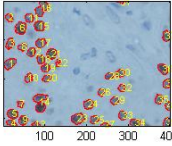
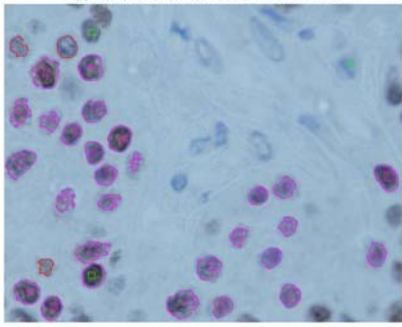
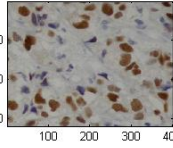
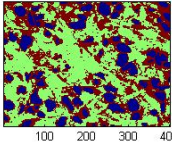
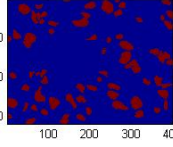
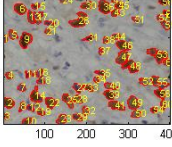
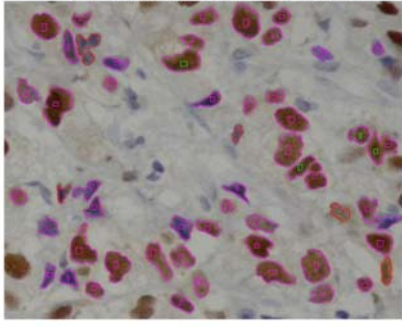
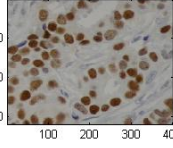
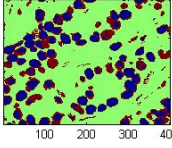
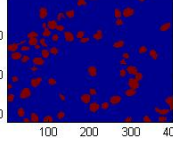
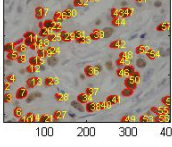
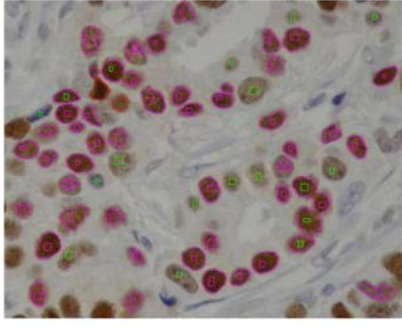
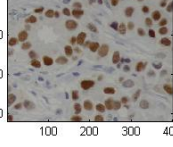
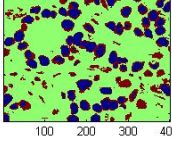
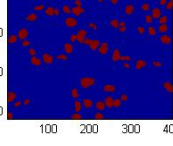
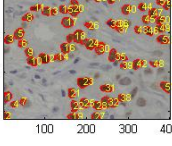
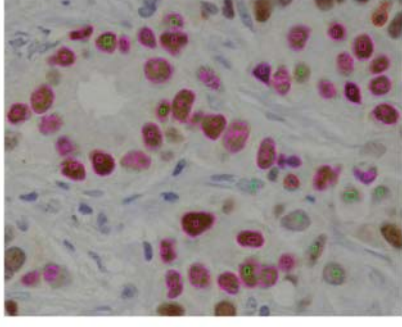
การทดลองทั้งหมดจะต้องผ่านกระบวนการขั้นตอนทั้ง 5 ขั้นตอน ดังที่ได้กล่าวไว้ในบทที่ 3 และในขั้นตอนที่ 2 ได้ทดลองแยกภาพเซลล์มะเร็งเต้านมด้วยวิธี *K-Median Clustering* และ *K-Means Clustering* ซึ่งทำให้ได้ผลการนับเซลล์มะเร็งในภาพค่อนข้างดี จากนั้นจะนำผลการนับไปเปรียบเทียบกับผลการนับของผู้เชี่ยวชาญ และสามารถนำผลที่ได้ไปใช้เป็นประโยชน์ทางการแพทย์ต่อไปได้

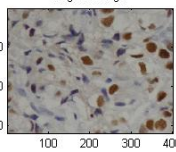
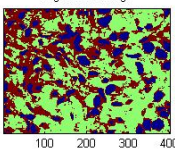
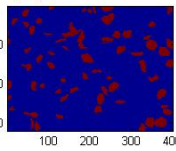
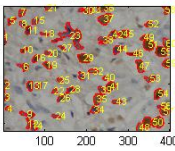
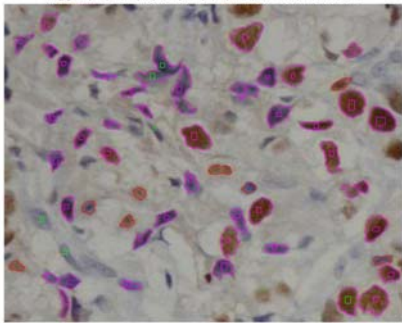
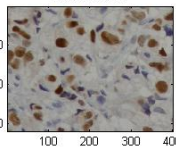
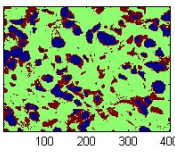
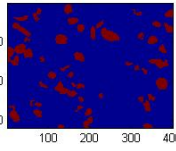
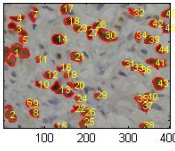
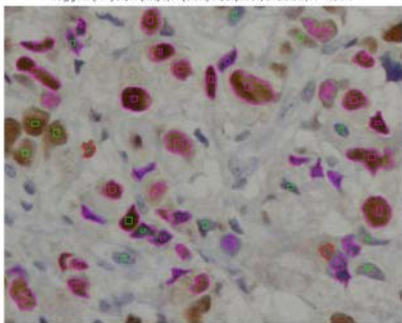
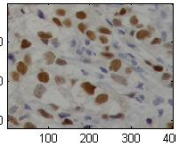
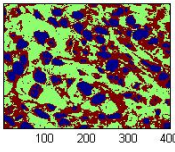
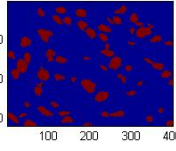
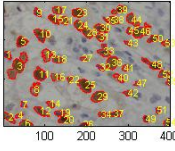
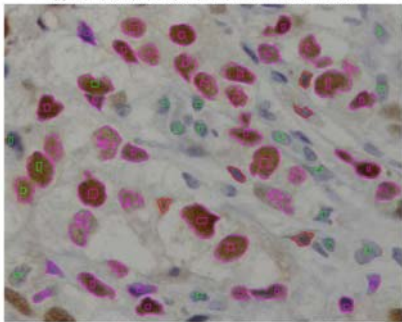
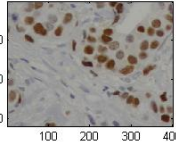
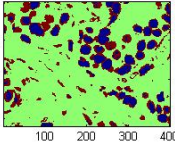
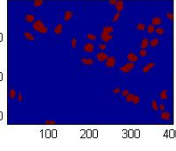
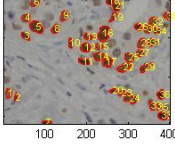
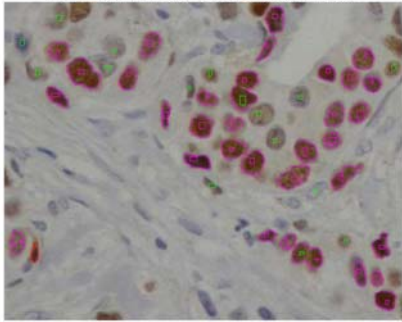
4.1 ผลการทดลอง

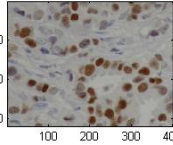
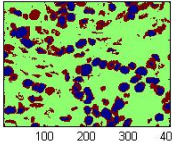
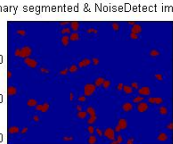
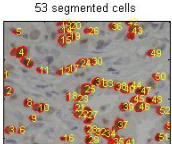
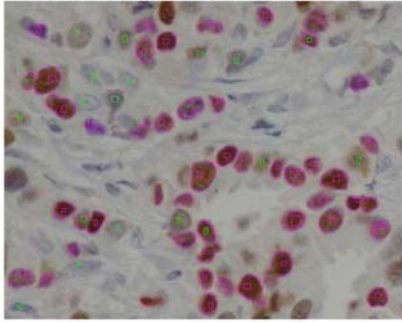
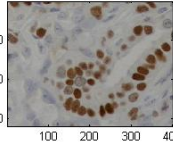
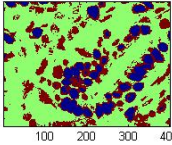
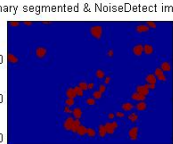
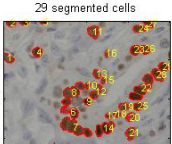
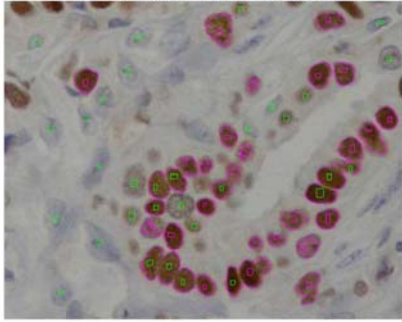
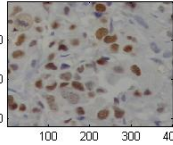
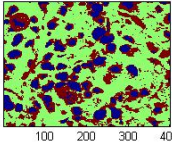
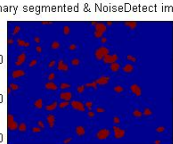
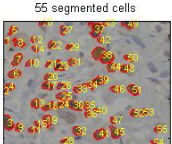
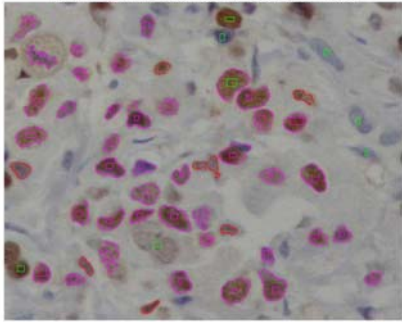
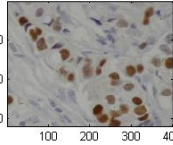
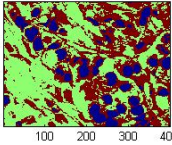
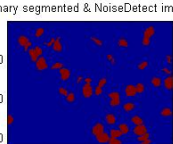
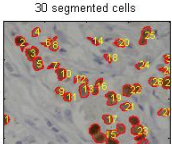
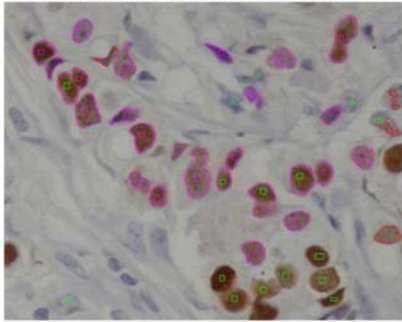
ผลการทดลองประกอบด้วยผลการแยกเซลล์ออกจากพื้นหลังของภาพ ผลการกำจัดสิ่งรบกวน ผลการนับเซลล์ และผลการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ ภาพที่ใช้ในการทดลองทั้งหมด 30 ภาพที่มีความแตกต่างกัน แสดงภาพผลการทดลองทั้งหมดโดยใช้ *K-Median* ในขั้นตอนการตัดแยกภาพเซลล์มะเร็งเต้านม ดังตารางที่ 4.1

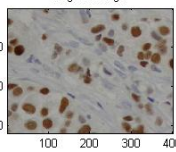
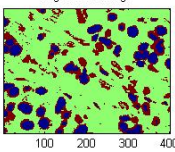
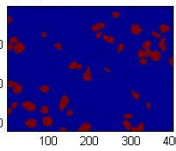
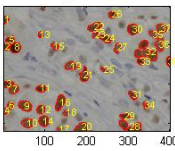
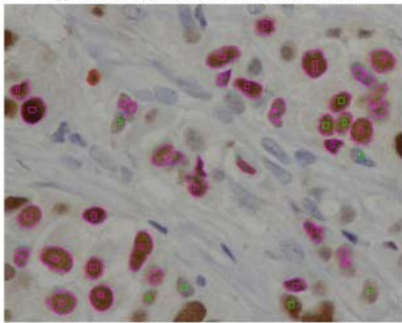
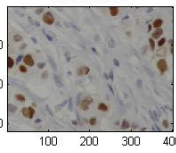
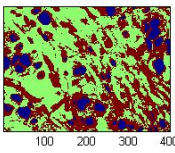
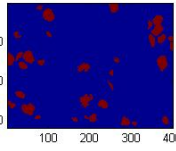
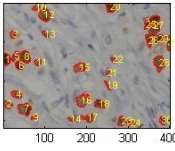
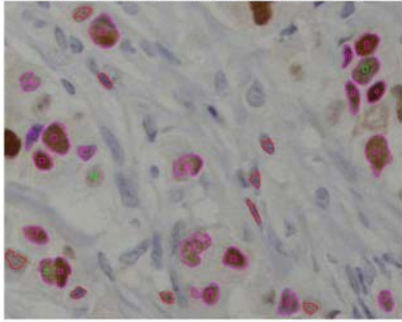
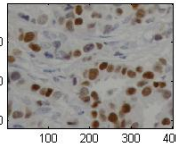
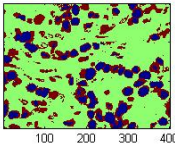
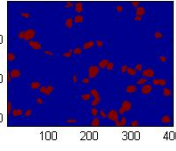
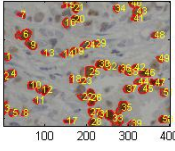
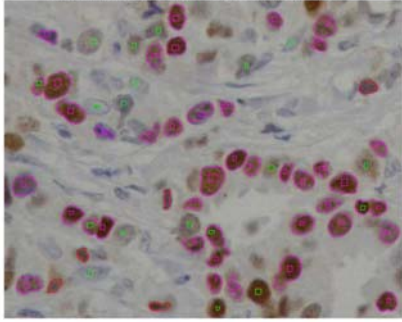
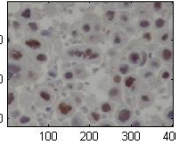
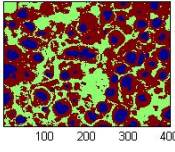
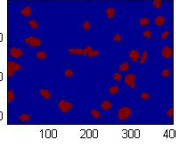
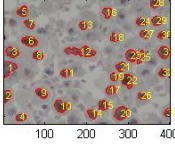
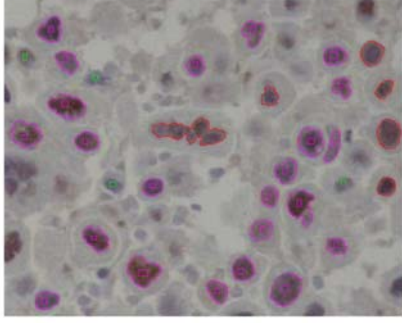
ตาราง 4.1 ผลการแยกเซลล์ กำจัดสิ่งรบกวน นับเซลล์และเปรียบเทียบกับผู้เชี่ยวชาญโดยใช้ *K-Median* ในขั้นตอนการตัดแยกภาพเซลล์มะเร็งเต้านม

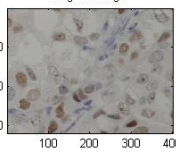
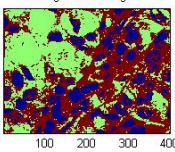
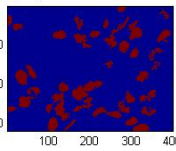
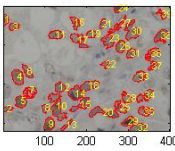
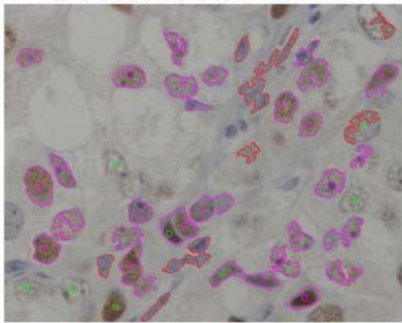
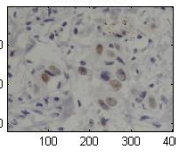
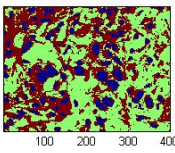
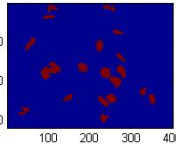
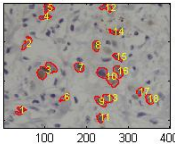
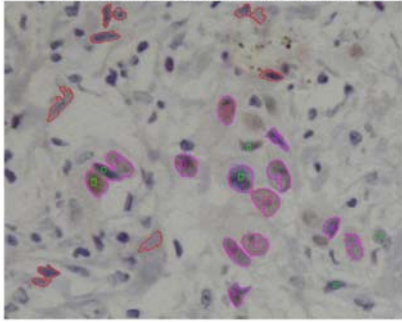
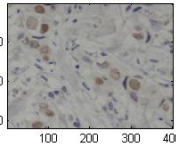
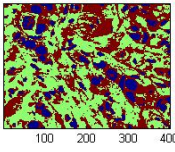
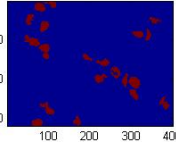
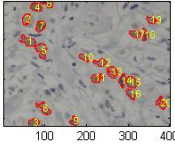
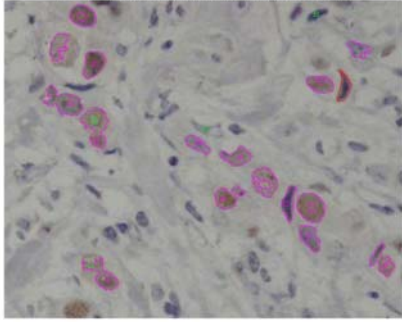
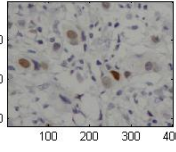
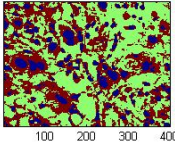
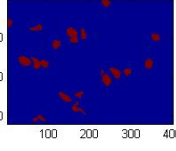
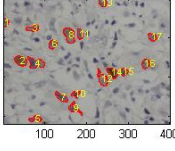
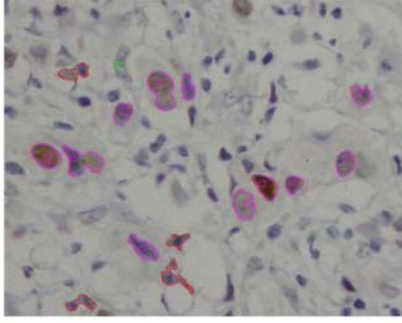
ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบกับผู้เชี่ยวชาญ
01	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>46 segmented cells</p> 	<p>01.jpg - TP(Fwk) 35, FP(Red) 1, FN(Green) 26, Expert 07, S 58.239%, PPV 97.5%</p> 
02	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>48 segmented cells</p> 	<p>02.jpg - TP(Fwk) 45, FP(Red) 1, FN(Green) 19, Expert 04, S 70.3125%, PPV 97.6261%</p> 
03	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>54 segmented cells</p> 	<p>03.jpg - TP(Fwk) 42, FP(Red) 4, FN(Green) 12, Expert 04, S 77.7778%, PPV 91.3043%</p> 
04	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>66 segmented cells</p> 	<p>04.jpg - TP(Fwk) 56, FP(Red) 4, FN(Green) 32, Expert 00, S 64.4444%, PPV 93.5484%</p> 

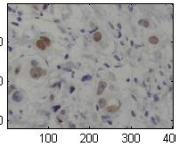
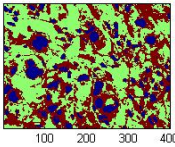
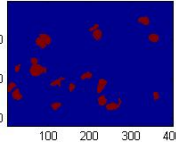
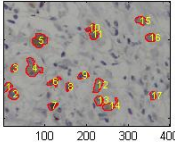
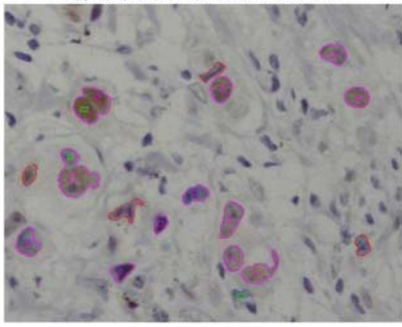
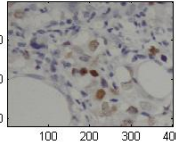
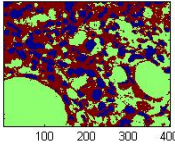
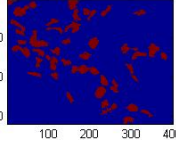
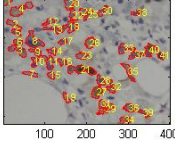
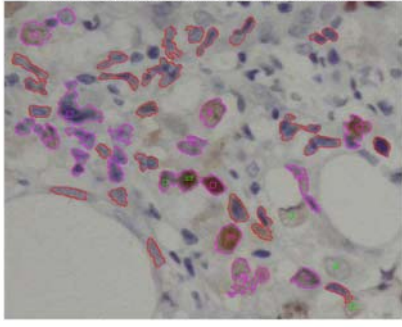
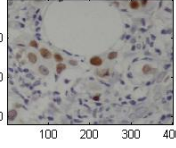
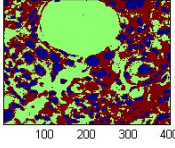
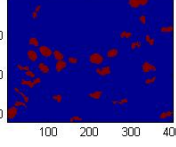
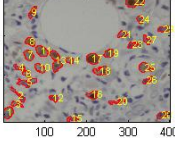
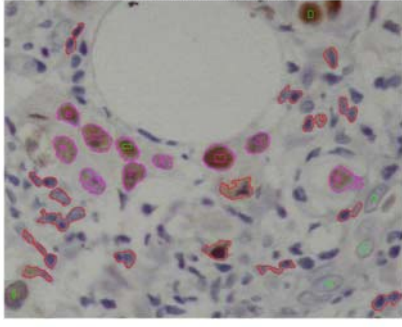
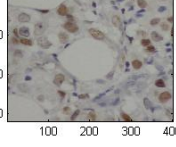
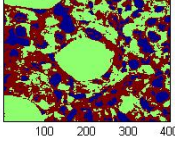
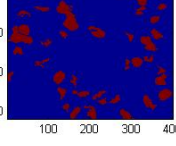
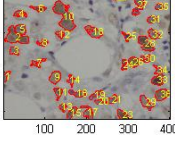
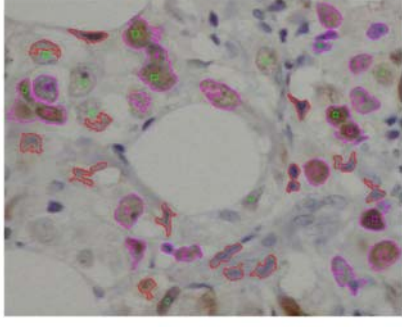
ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
05	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>40 segmented cells</p> 	<p>05.jpg - TP(Pink) 27, FP(Red) 3, FN(Green) 3, Expert 30, S 90%, PPV 90%</p> 
06	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>62 segmented cells</p> 	<p>06.jpg - TP(Pink) 46, FP(Red) 5, FN(Green) 4, Expert 50, S 92%, PPV 90.1961%</p> 
07	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>58 segmented cells</p> 	<p>07.jpg - TP(Pink) 44, FP(Red) 2, FN(Green) 18, Expert 62, S 70.9677%, PPV 95.6522%</p> 
08	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>54 segmented cells</p> 	<p>08.jpg - TP(Pink) 43, FP(Red) 2, FN(Green) 11, Expert 53, S 79.6296%, PPV 95.6592%</p> 

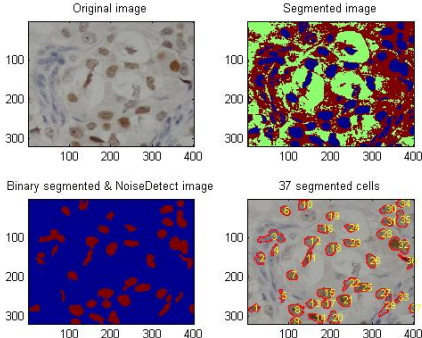
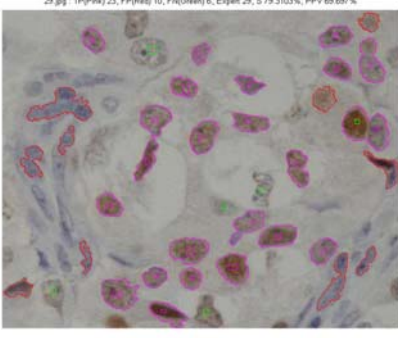
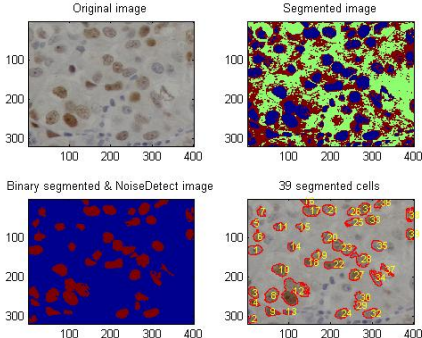
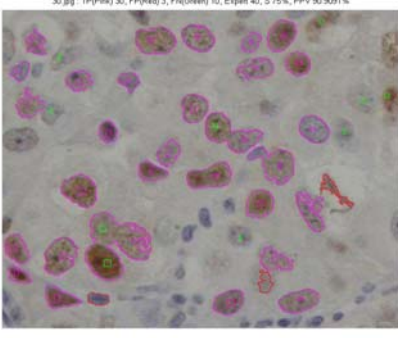
ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
09	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>60 segmented cells</p> 	<p>09.jpg: TP(Pnk) 41, FP(Red) 6, FN(Green) 6, Expert 47, S:87.234%, PPV:83.6735%</p> 
10	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>46 segmented cells</p> 	<p>10.jpg: TP(Pnk) 37, FP(Red) 3, FN(Green) 14, Expert 51, S:72.549%, PPV:92.5%</p> 
11	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>55 segmented cells</p> 	<p>11.jpg: TP(Pnk) 40, FP(Red) 5, FN(Green) 17, Expert 57, S:70.1754%, PPV:88.6889%</p> 
12	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>38 segmented cells</p> 	<p>12.jpg: TP(Pnk) 33, FP(Red) 1, FN(Green) 16, Expert 49, S:67.369%, PPV:97.028%</p> 

ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
13	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>53 segmented cells</p> 	<p>13.jpg: TP(Pink) 42, FP(Red) 2, FN(Green) 19, Expert 61, S 68.8525%, PPV 95.4545%</p> 
14	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>29 segmented cells</p> 	<p>14.jpg: TP(Pink) 25, FP(Red) 0, FN(Green) 32, Expert 57, S 43.8996%, PPV 100%</p> 
15	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>55 segmented cells</p> 	<p>15.jpg: TP(Pink) 39, FP(Red) 7, FN(Green) 8, Expert 47, S 62.9787%, PPV 84.7626%</p> 
16	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>30 segmented cells</p> 	<p>16.jpg: TP(Pink) 21, FP(Red) 5, FN(Green) 16, Expert 37, S 56.7593%, PPV 80.7692%</p> 

ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
17	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>38 segmented cells</p> 	<p>17.jpg - TP(Pink) 26, FP(Red) 2, FN(Green) 9, Expert 37, S 75.6757%, PPV 93.3333%</p> 
18	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>31 segmented cells</p> 	<p>18.jpg - TP(Pink) 19, FP(Red) 9, FN(Green) 3, Expert 22, S 86.303%, PPV 67.8571%</p> 
19	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>51 segmented cells</p> 	<p>19.jpg - TP(Pink) 38, FP(Red) 5, FN(Green) 19, Expert 57, S 66.6667%, PPV 88.3721%</p> 
20	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>33 segmented cells</p> 	<p>20.jpg - TP(Pink) 23, FP(Red) 6, FN(Green) 4, Expert 27, S 85.1852%, PPV 79.3103%</p> 

ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
21	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>39 segmented cells</p> 	<p>21.jpg TP(Pnk) 25, FPR(Red) 10, FN(Green) 8, Expert 33, S 75.7576%, PPV 71.4286%</p> 
22	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>18 segmented cells</p> 	<p>22.jpg TP(Pnk) 11, FPR(Red) 7, FN(Green) 6, Expert 17, S 64.7059%, PPV 61.1111%</p> 
23	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>20 segmented cells</p> 	<p>23.jpg TP(Pnk) 16, FPR(Red) 1, FN(Green) 4, Expert 20, S 80%, PPV 94.1176%</p> 
24	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>17 segmented cells</p> 	<p>24.jpg TP(Pnk) 11, FPR(Red) 4, FN(Green) 1, Expert 12, S 91.667%, PPV 73.3333%</p> 

ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
25	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>17 segmented cells</p> 	<p>25.jpg - TP(Pink) 12, FP(Red) 4, FN(Green) 3, Expert 15, S 80%, PPV 75%</p> 
26	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>41 segmented cells</p> 	<p>26.jpg - TP(Pink) 16, FP(Red) 23, FN(Green) 5, Expert 21, S 76.192%, PPV 41.029%</p> 
27	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>29 segmented cells</p> 	<p>27.jpg - TP(Pink) 9, FP(Red) 17, FN(Green) 6, Expert 15, S 60%, PPV 34.6154%</p> 
28	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>36 segmented cells</p> 	<p>28.jpg - TP(Pink) 18, FP(Red) 14, FN(Green) 12, Expert 30, S 82%, PPV 56.25%</p> 

ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
29		
30		

และสามารถดูผลการแยกเซลล์ออกจากพื้นหลังของภาพ ผลการกำจัดสิ่งรบกวน ผลการนับเซลล์ และผลการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ โดยใช้ *K-Means* ในขั้นตอนการตัดแยกภาพเซลล์มะเร็งเต้านมได้ในภาคผนวก ก

4.2 สรุปผลการทดลอง

การทดลองทั้งหมดจะแบ่งออกเป็น 5 ส่วน ได้แก่ 1) ขั้นตอนการเตรียมภาพเซลล์มะเร็งเต้านม (Preprocess) 2) ขั้นตอนการตัดแยกภาพเซลล์มะเร็งเต้านม (Segmentation) 3) ขั้นตอนการกำจัดสิ่งรบกวนในภาพ (Noise Reduction) 4) ขั้นตอนการนับเซลล์มะเร็งเต้านม (Cell Count) 5) ขั้นตอนการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ (Compare with Expert) ซึ่งจะดำเนินการในแต่ละขั้นตอนได้จะต้องได้ผลลัพธ์จากขั้นตอนก่อนหน้า เรียงลำดับจากขั้นตอนที่ 1 จนถึงขั้นตอนที่ 5 เมื่อสิ้นสุดกระบวนการในขั้นตอนที่ 5 ก็สามารถนำกระบวนการทั้งหมดไปสร้างเป็นระบบการตัดแยกเซลล์มะเร็งเต้านมแบบอัตโนมัติ (Automated Breast Cancer Cell Image Segmentation: ABC_CIS) ซึ่งสามารถดูคู่มือการใช้งานระบบได้ในภาคผนวก ข

4.2.1 ขั้นตอนการเตรียมภาพเซลล์มะเร็งเต้านม (Preprocess)

ในขั้นตอนนี้ได้มีการแบ่งภาพเซลล์มะเร็งเต้านมที่ได้จากแพทย์โดยตรงออกเป็น 2 กลุ่ม ได้แก่

- 1) กลุ่มของภาพที่มีเซลล์ขนาดใหญ่และชัดเจนเป็นส่วนมาก คือภาพ 01-19 เรียกว่า กลุ่มภาพเซลล์ขนาดใหญ่ (Big Cell)
- 2) กลุ่มของภาพที่มีเซลล์ขนาดเล็กและไม่ค่อยชัดเจนเป็นส่วนมาก คือภาพ 20 - 30 เรียกว่า กลุ่มภาพเซลล์ขนาดเล็ก (Small Cell)

ผลจากการแบ่งภาพเซลล์มะเร็งเต้านมต้นฉบับออกเป็น 2 กลุ่ม ส่งผลให้การทดลองมีความสะดวกและมีประสิทธิภาพพอสมควร และสามารถให้ผู้ใช้ระบบเลือกกระบวนการต่อไปให้เหมาะสมกับภาพที่นำมาทดสอบได้

4.2.2 ขั้นตอนการตัดแยกภาพเซลล์มะเร็งเต้านม (Segmentation)

ในขั้นตอนนี้เป็นการแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังของภาพ ด้วยการใช้อัลกอริทึมสี RGB ในการแบ่งแยกเซลล์ออกจากพื้นหลัง และใช้เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) 2 วิธี ได้แก่ 1) K-Means Clustering และ 2) K-Median Clustering พบว่าทั้ง 2 วิธีสามารถแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังของภาพได้ผลค่อนข้างดี และสามารถนำไปผ่านกระบวนการกำจัดสิ่งรบกวนต่อไปได้ แต่ผลจาก 2 วิธีดังกล่าวมีความแตกต่างกันเล็กน้อยที่ผลการแยกเซลล์ ซึ่งจะสามารถพิสูจน์ความถูกต้องของการแยกเซลล์ได้ โดยนำผลการแยกเซลล์ที่ได้ไปผ่านกระบวนการกำจัดสิ่งรบกวนและการนับเซลล์มะเร็งในภาพต่อไป และเปรียบเทียบผลการนับเซลล์ของทั้ง 2 วิธีกับผลการนับเซลล์ของผู้เชี่ยวชาญในขั้นตอนต่อไป จึงจะสามารถทราบได้ว่า 2 วิธีดังกล่าว วิธีใดให้ผลลัพธ์ที่ใกล้เคียงกับผู้เชี่ยวชาญมากกว่า จากการทดลองที่ผ่านมาพบว่า K-Median Clustering ให้ผลที่น่าพึงพอใจมากกว่า ซึ่งจะแสดงให้เห็นในขั้นตอนต่อไป ดังนั้นในระบบการตัดแยกเซลล์มะเร็งเต้านมแบบอัตโนมัติ (Automated Breast Cancer Cell Image Segmentation: ABC_CIS) จึงเลือกใช้ K-Median Clustering ในขั้นตอนการตัดแยกภาพเซลล์มะเร็งเต้านม ผลการแยกเซลล์ออกจากพื้นหลัง เป็นส่วนหนึ่งในตาราง 4.1

4.2.3 ขั้นตอนการกำจัดสิ่งรบกวนในภาพ (Noise Reduction)

ในขั้นตอนนี้เป็นการนำผลการแยกเซลล์ออกจากพื้นหลังของภาพที่ได้จากขั้นตอนที่ 2 มากำจัดสิ่งรบกวนหรือบริเวณที่ไม่ใช่เซลล์ออกจากภาพ โดยใช้เทคนิค Morphological Operations

ในขั้นตอนนี้จะแยกการทำงานออกเป็น 2 ส่วน สำหรับภาพ 2 กลุ่มที่ได้แบ่งแยกไว้ในขั้นตอนที่ 1 ซึ่งได้แก่ ภาพเซลล์ขนาดใหญ่และภาพเซลล์ขนาดเล็ก ซึ่งจากการทดลองพบว่าในขั้นตอนนี้สามารถกำจัดสิ่งรบกวนได้ผลดี โดยสามารถนำส่วนที่ไม่มีประโยชน์ออกจากภาพได้ และสามารถเติม

ส่วนของเซลล์ที่ขาดไปให้เต็มได้ ซึ่งจะเป็นประโยชน์ในขั้นตอนการนับเซลล์ต่อไป ผลการกำจัดสิ่งรบกวน แสดงเป็นส่วนหนึ่งในตาราง 4.1

4.2.4 ขั้นตอนการนับเซลล์มะเร็งเต้านม (Cell Count)

ในขั้นตอนนี้เป็นการนำผลการกำจัดสิ่งรบกวนออกจากภาพที่ได้จากขั้นตอนที่ 3 มาผ่านกระบวนการการนับเซลล์มะเร็งในภาพ ซึ่งผลการนับเซลล์มะเร็งในภาพ พบว่าสามารถนับจำนวนเซลล์มะเร็งได้ดี และนำผลการนับเซลล์ที่ได้ไปเปรียบเทียบกับผลการนับเซลล์ของผู้เชี่ยวชาญต่อไปได้ ผลการนับเซลล์มะเร็งในภาพ แสดงเป็นส่วนหนึ่งในตาราง 4.1

4.2.5 ขั้นตอนการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ (Compare with Expert)

ในขั้นตอนนี้เป็นการนำผลการนับเซลล์ที่ได้จากการทดลองมาเปรียบเทียบกับผลการนับเซลล์ของผู้เชี่ยวชาญ ซึ่งผลการเปรียบเทียบกับผู้เชี่ยวชาญ พบว่าได้ผลค่อนข้างดี ได้เปอร์เซ็นต์ของค่า S และ PPV ค่อนข้างสูงและใกล้เคียงกัน หมายความว่าผลการนับเซลล์ที่ได้จากการทดลอง ไม่น้อยหรือไม่มากเกินไปกว่าผลการนับเซลล์ของผู้เชี่ยวชาญ ซึ่งถือว่าได้ความถูกต้องพอสมควร ผลการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ แสดงเป็นส่วนหนึ่งในตาราง 4.1 และสามารถดูค่าตัวเลขผลการเปรียบเทียบโดยใช้ *K-Mean* และ *K-Median* ได้ดังตาราง 4.2

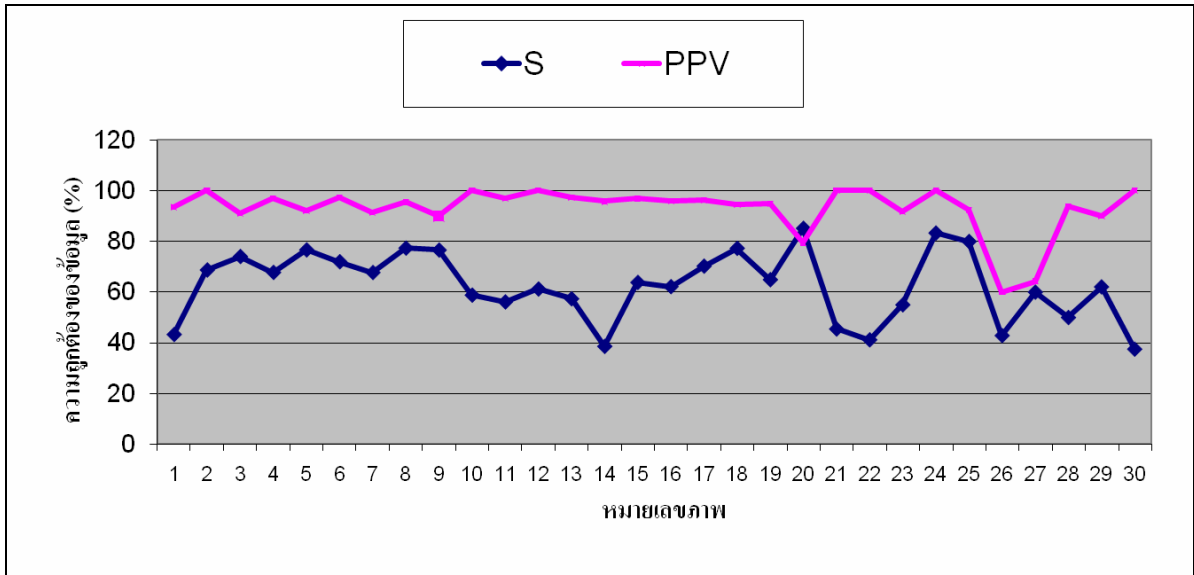
ตาราง 4.2 ตัวเลขผลการเปรียบเทียบการนับเซลล์กับผู้เชี่ยวชาญโดยใช้ *K-Mean* และ *K-Median* ในขั้นตอนการตัดแยกภาพเซลล์มะเร็งเต้านม

ชื่อภาพ	วิธีการแบ่งกลุ่ม	TP	FP	FN	Expert	S (%)	PPV (%)
01	<i>K-Means</i>	29	2	38	67	43.28	93.55
	<i>K-Median</i>	39	1	28	67	58.21	97.50
02	<i>K-Means</i>	44	0	20	64	68.75	100.00
	<i>K-Median</i>	45	1	19	64	70.31	97.83
03	<i>K-Means</i>	40	4	14	54	74.07	90.91
	<i>K-Median</i>	42	4	12	54	77.78	91.30
04	<i>K-Means</i>	61	2	29	90	67.78	96.83
	<i>K-Median</i>	58	4	32	90	64.44	93.55
05	<i>K-Means</i>	23	2	7	30	76.67	92.00
	<i>K-Median</i>	27	3	3	30	90.00	90.00

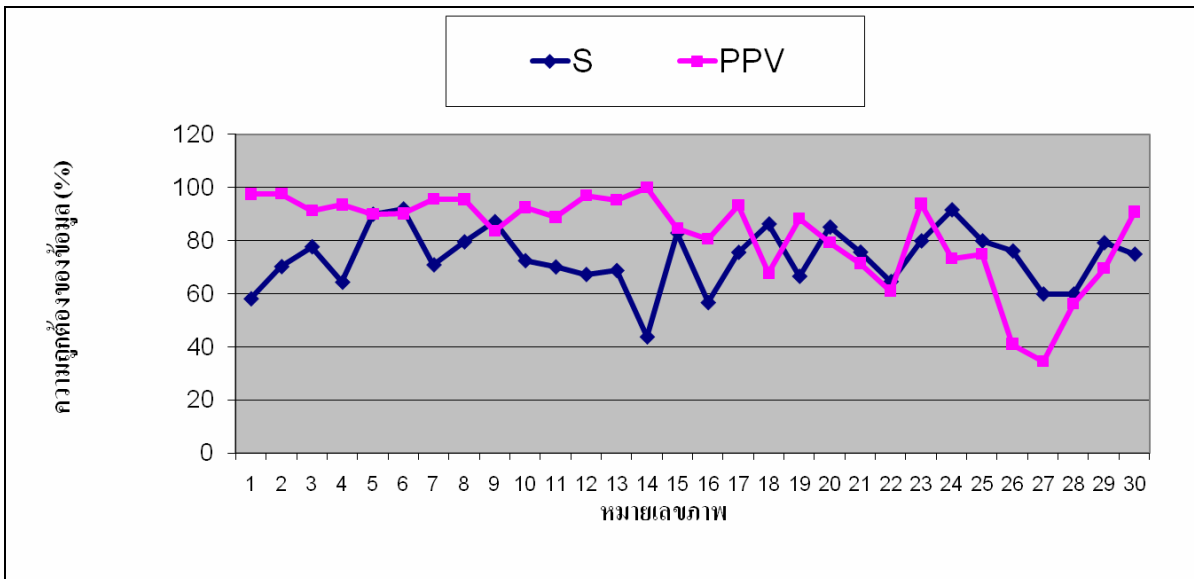
ชื่อภาพ	วิธีการแบ่งกลุ่ม	TP	FP	FN	Expert	S (%)	PPV (%)
06	K-Means	36	1	14	50	72.00	97.30
	K-Median	46	5	4	50	92.00	90.20
07	K-Means	42	4	20	62	67.74	91.30
	K-Median	44	2	18	62	70.97	95.65
08	K-Means	41	2	12	53	77.36	95.35
	K-Median	43	2	11	53	79.63	95.56
09	K-Means	36	4	11	47	76.60	90.00
	K-Median	41	8	6	47	87.23	83.67
10	K-Means	30	0	21	51	58.82	100.00
	K-Median	37	3	14	51	72.55	92.50
11	K-Means	32	1	25	57	56.14	96.97
	K-Median	40	5	17	57	70.18	88.89
12	K-Means	30	0	19	49	61.22	100.00
	K-Median	33	1	16	49	67.35	97.06
13	K-Means	35	1	26	61	57.38	97.22
	K-Median	42	2	19	61	68.85	95.45
14	K-Means	22	1	35	57	38.60	95.65
	K-Median	25	0	32	57	43.86	100.00
15	K-Means	30	1	17	47	63.83	96.77
	K-Median	39	7	8	47	82.98	84.78
16	K-Means	23	1	14	37	62.16	95.83
	K-Median	21	5	16	37	56.76	80.77
17	K-Means	26	1	11	37	70.27	96.30
	K-Median	28	2	9	37	75.68	93.33
18	K-Means	17	1	5	22	77.27	94.44
	K-Median	19	9	3	22	86.36	67.86
19	K-Means	37	2	20	57	64.91	94.87
	K-Median	38	5	19	57	66.67	88.37

ข้อภาพ	วิธีการแบ่งกลุ่ม	TP	FP	FN	Expert	S (%)	PPV (%)
20	K-Means	23	6	4	27	85.19	79.31
	K-Median	23	6	4	27	85.19	79.31
21	K-Means	15	0	18	33	45.45	100.00
	K-Median	25	10	8	33	75.76	71.43
22	K-Means	7	0	10	17	41.18	100.00
	K-Median	11	7	6	17	64.71	61.11
23	K-Means	11	1	9	20	55.00	91.67
	K-Median	16	1	4	20	80.00	94.12
24	K-Means	10	0	2	12	83.33	100.00
	K-Median	11	4	1	12	91.67	73.33
25	K-Means	12	1	3	15	80.00	92.31
	K-Median	12	4	3	15	80.00	75.00
26	K-Means	9	6	12	21	42.86	60.00
	K-Median	16	23	5	21	76.19	41.03
27	K-Means	9	5	6	15	60.00	64.29
	K-Median	9	17	6	15	60.00	34.62
28	K-Means	15	1	15	30	50.00	93.75
	K-Median	18	14	12	30	60.00	56.25
29	K-Means	18	2	11	29	62.07	90.00
	K-Median	23	10	6	29	79.31	69.70
30	K-Means	15	0	25	40	37.50	100.00
	K-Median	30	3	10	40	75.00	90.91

จากตารางผลการทดลองเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญทั้งหมด สามารถสร้างกราฟแสดงความแตกต่างรายละเอียดของค่า S และ PPV ของทั้ง 2 วิธีได้ ดังแสดงในภาพประกอบ 4.1 ถึงภาพประกอบ 4.5

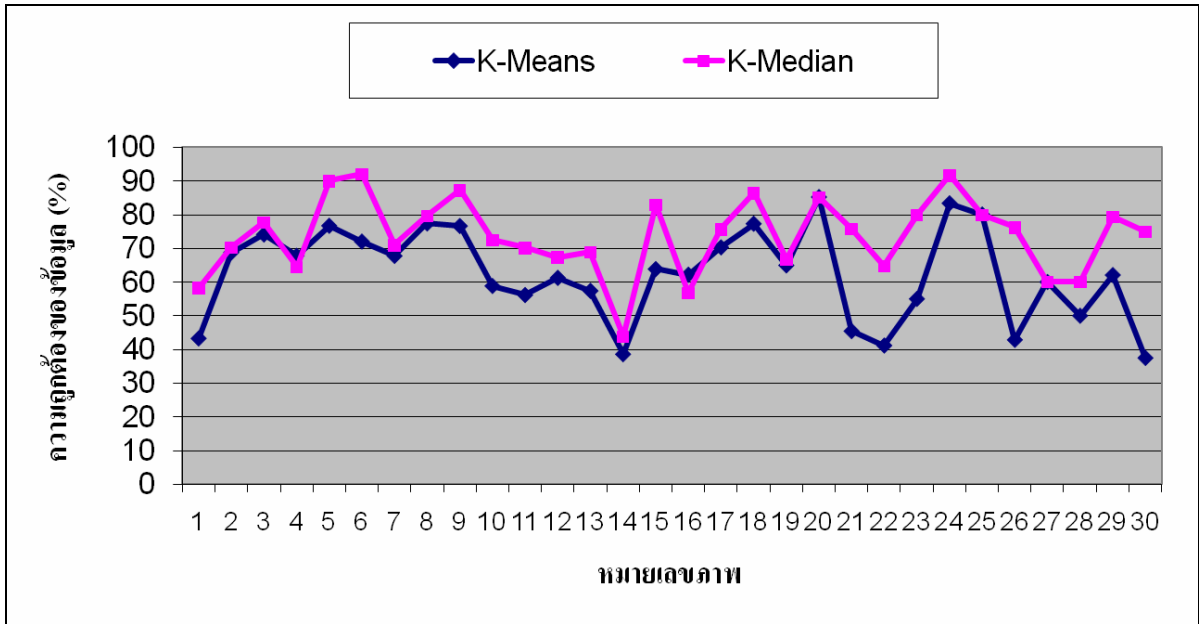


ภาพประกอบ 4.1 กราฟแสดงเปอร์เซ็นต์ของค่า S และ PPV โดยใช้ K-Means

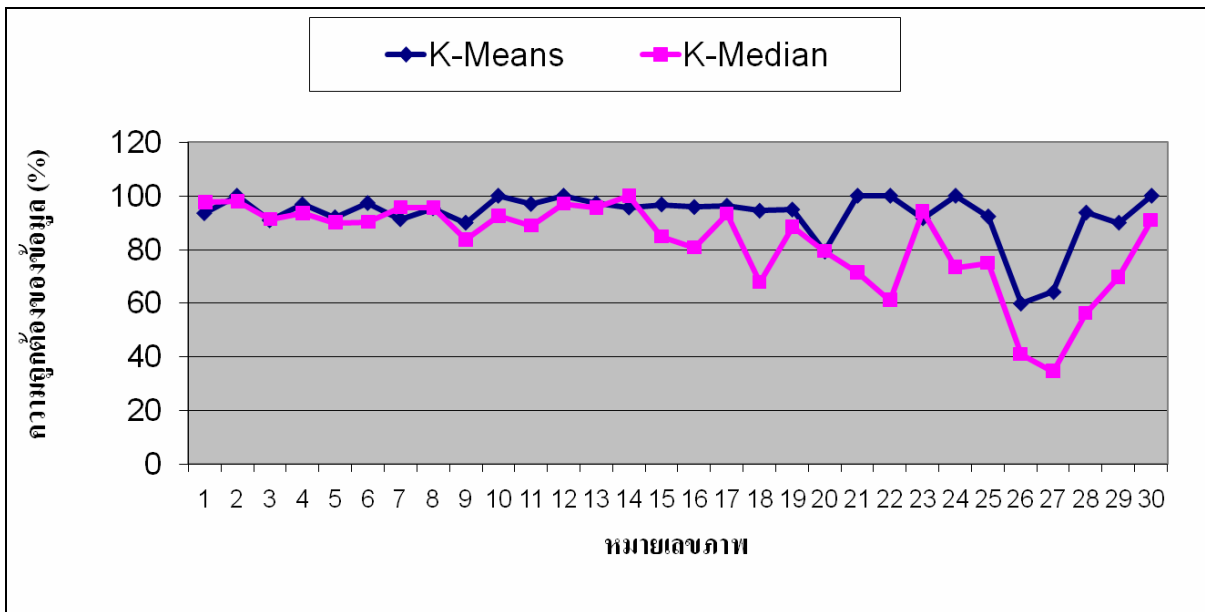


ภาพประกอบ 4.2 กราฟแสดงเปอร์เซ็นต์ของค่า S และ PPV โดยใช้ K-Median

ภาพประกอบ 4.1 และ 4.2 แสดงเปอร์เซ็นต์ของค่า S และ PPV พบว่า เพื่อเป็นการบ่งบอกว่าการสรุปค่าความถูกต้องของผลการเปรียบเทียบกับผู้เชี่ยวชาญนั้น จะต้องใช้ค่า S และ PPV ประกอบกัน เพราะจากกราฟในภาพประกอบ 4.1 ค่า S และ PPV ไม่สามารถระบุได้ว่าค่าความถูกต้องที่เกิดขึ้นในแต่ละภาพมีแนวโน้มเป็นไปในทิศทางเดียวกัน จึงไม่สามารถสรุปโดยใช้เพียงค่าใดค่าหนึ่งได้



ภาพประกอบ 4.3 กราฟเปรียบเทียบค่า S ระหว่าง K-Means และ K-Median

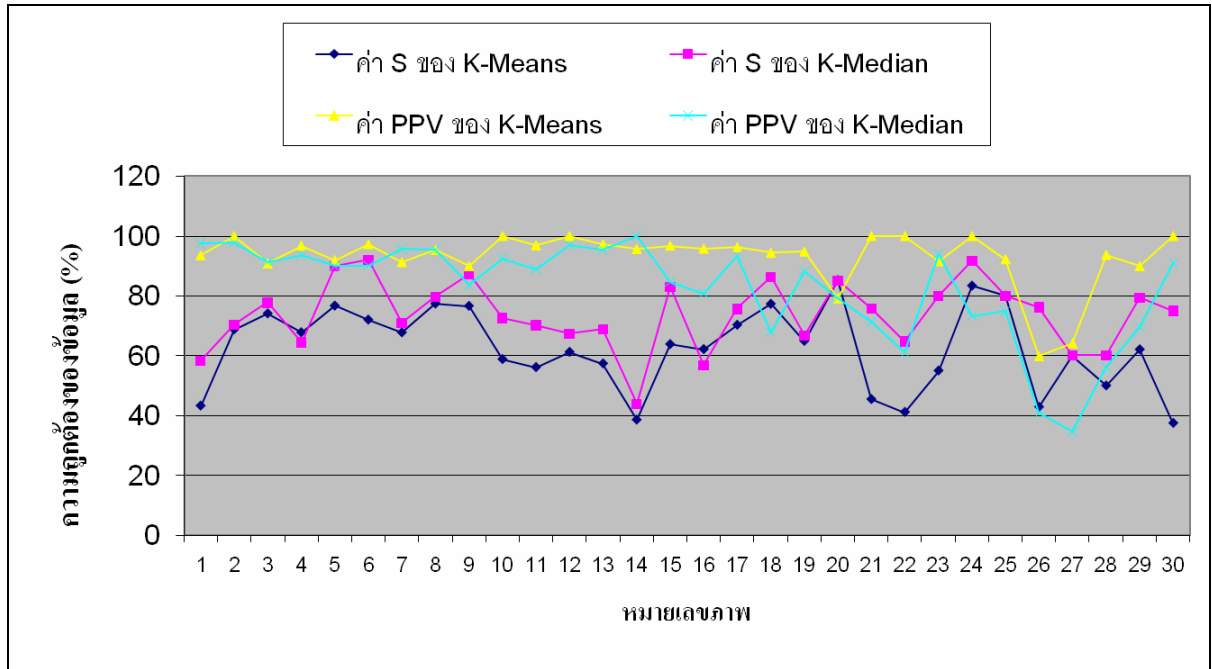


ภาพประกอบ 4.4 กราฟเปรียบเทียบค่า PPV ระหว่าง K-Means และ K-Median

ภาพประกอบ 4.3 เปรียบเทียบเปอร์เซ็นต์ของค่า S ระหว่าง 2 วิธี จะเห็นว่าการใช้ K-Median จะให้ค่า S ใกล้เคียงและมีเปอร์เซ็นต์ความถูกต้องสูงกว่าการใช้ K-Means เกือบทุกภาพ

ภาพประกอบ 4.4 เปรียบเทียบเปอร์เซ็นต์ของค่า PPV ระหว่าง 2 วิธี จะเห็นว่าการใช้ K-Means จะให้ค่า S ใกล้เคียงและมีเปอร์เซ็นต์ความถูกต้องสูงกว่าการใช้ K-Median เกือบทุกภาพ

ดังนั้นจึงไม่สามารถสรุปได้ทันทีว่า วิธีไหนให้ค่าความถูกต้องดีกว่ากัน จึงต้องทำการหาค่าเฉลี่ยของค่า S และ PPV ใน 2 วิธี แสดงดังตาราง 4.3



ภาพประกอบ 4.5 กราฟเปรียบเทียบ ค่า S และ PPV ระหว่าง K-Means และ K-Median

ภาพเซลล์มะเร็งเต้านมที่จะนำมาแยกเซลล์ออกจากพื้นหลัง เพื่อแบ่งข้อมูลเป็น 3 กลุ่ม อาจมีบางกลุ่มที่มีจุดสีที่ไม่ใช่ค่าสีในกลุ่มนั้นปะปนอยู่ในภาพ ถือว่าจุดสีนั้นเป็นข้อมูลที่อยู่ห่างจากข้อมูลส่วนใหญ่ ซึ่งจากทฤษฎีที่เกี่ยวข้อง 2.7 สรุปได้ว่าในกรณีของชุดข้อมูลที่มีข้อมูลที่ห่างจากข้อมูลส่วนใหญ่ ปะปนเข้ามา จะทำให้การหาค่าจุดกึ่งกลางของกลุ่มโดยใช้ Median สามารถให้ผลที่น่าเชื่อถือกว่า ดังนั้นวิธี K-Median Clustering ทนทานต่อข้อมูลที่เป็นสัญญาณรบกวนมากกว่า K-Means Clustering โดยหากมีข้อมูลที่ไม่ได้อยู่ในกลุ่มของข้อมูลส่วนใหญ่ จะทำให้ K-Median Clustering สามารถหาข้อมูลที่น่ามาเป็นจุดศูนย์กลางได้ใกล้เคียงกับความจริงและสมเหตุสมผลมากกว่า จึงเหมาะสำหรับนำมาใช้ในระบบการตัดแยกภาพเซลล์มะเร็งเต้านมแบบอัตโนมัติ (ABC_CIS)

FN หรือ False Negative คือ ค่าความผิดพลาดทางลบ หมายถึง โปรแกรมนับขาดไปจากที่ผู้เชี่ยวชาญนับ ถ้าค่านี้มีมากในกรณีนำระบบไปใช้จริงจะส่งผลทำให้แพทย์วินิจฉัยไม่หายแก่ผู้ป่วย ทำให้เซลล์มะเร็งสามารถแพร่กระจายต่อไปได้ ซึ่งเป็นอันตรายต่อผู้ป่วย สามารถแก้ไขให้ค่า FN นี้มีค่าเป็น 0 ได้คือ ให้โปรแกรมนับเซลล์ส่วนที่นับขาดไปนี้เพิ่ม ให้เท่ากับที่ผู้เชี่ยวชาญระบุมา

FP หรือ False Positive คือ ค่าความผิดพลาดทางบวก หมายถึง โปรแกรมนับเกินไปจากที่ผู้เชี่ยวชาญนับ ถ้าค่านี้มีมากในกรณีนำระบบไปใช้จริงจะส่งผลทำให้แพทย์วินิจฉัยให้หายแก่ผู้ป่วยเกินความจำเป็น วิธีการแก้ไข คือ สร้างระบบให้สามารถนับส่วนที่เป็นเซลล์มะเร็งได้อย่างถูกต้อง

จากผลการทดลองทั้งหมดสามารถสรุปผลการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ โดยการเฉลี่ยค่า S และ PPV ซึ่งเมื่อนำค่าเฉลี่ยของทั้ง 2 วิธี มาเฉลี่ยเพื่อเปรียบเทียบกัน พบว่า การใช้การตัดแยกเซลล์ออกจากพื้นหลังของภาพโดยใช้วิธี K-Median จะให้ค่าเฉลี่ยความถูกต้องของค่า S และ PPV คือ 78.01 % ซึ่งมากกว่าการใช้วิธี K-Means ดังแสดงในตาราง 4.3 และภาพประกอบ 4.6

ตาราง 4.3 ค่าเฉลี่ยของผลการเปรียบเทียบกับผู้เชี่ยวชาญ

วิธีการ	S (%)	PPV (%)	เฉลี่ย
เฉลี่ย K-Means	62.58	92.89	77.74
เฉลี่ย K-Median	73.65	82.37	78.01

จากตาราง 4.3 ค่า S มีค่าเท่ากับ 73.65 % หมายความว่า ถ้าผู้เชี่ยวชาญนับได้ 100 เซลล์ แต่โปรแกรมนับได้แค่ 73.65 เซลล์ ดังนั้นผู้เชี่ยวชาญจะต้องระบุเซลล์เพิ่มจากโปรแกรมอีก 26.35 เซลล์ เพราะโปรแกรมนับขาดไป และจะเรียกค่าที่โปรแกรมนับขาดไปนั้นว่าค่า $S_{ขาด}$

ค่า PPV มีค่าเท่ากับ 82.37 หมายความว่า ถ้าโปรแกรมนับได้ 100 เซลล์ แต่ผู้เชี่ยวชาญนับได้ 92.37 เซลล์ ดังนั้นผู้เชี่ยวชาญจะต้องลบเซลล์จากโปรแกรมออกไป 17.63 เซลล์ เพราะโปรแกรมนับเกินไปจากผู้เชี่ยวชาญ และจะเรียกค่าที่โปรแกรมนับเกินไปนั้นว่าค่า $PPV_{เกิน}$

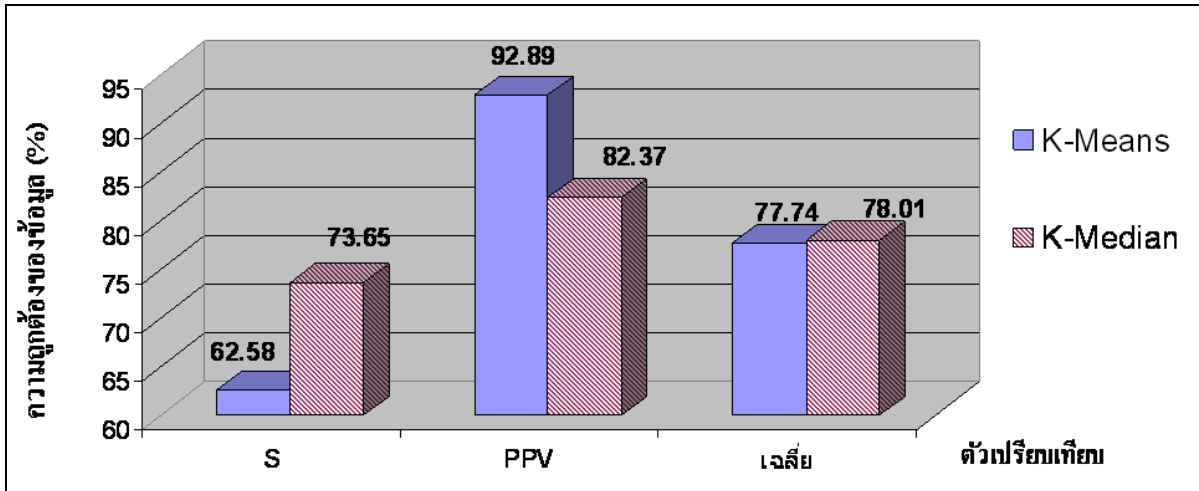
จากการเฉลี่ยผลการเปรียบเทียบค่า S และ PPV ในตาราง 4.3 สามารถนำมาคิดหาเปอร์เซ็นต์ของจำนวนเซลล์ส่วนที่ผู้เชี่ยวชาญจะต้องเพิ่มหรือนำออกจากโปรแกรม คือ ค่า $S_{ขาด}$ และ $PPV_{เกิน}$ ได้ดังตาราง 4.4 และภาพประกอบ 4.7

ตาราง 4.4 ค่าเฉลี่ยของจำนวนเซลล์ส่วนที่ผู้เชี่ยวชาญจะต้องเพิ่มหรือนำออกจากโปรแกรม

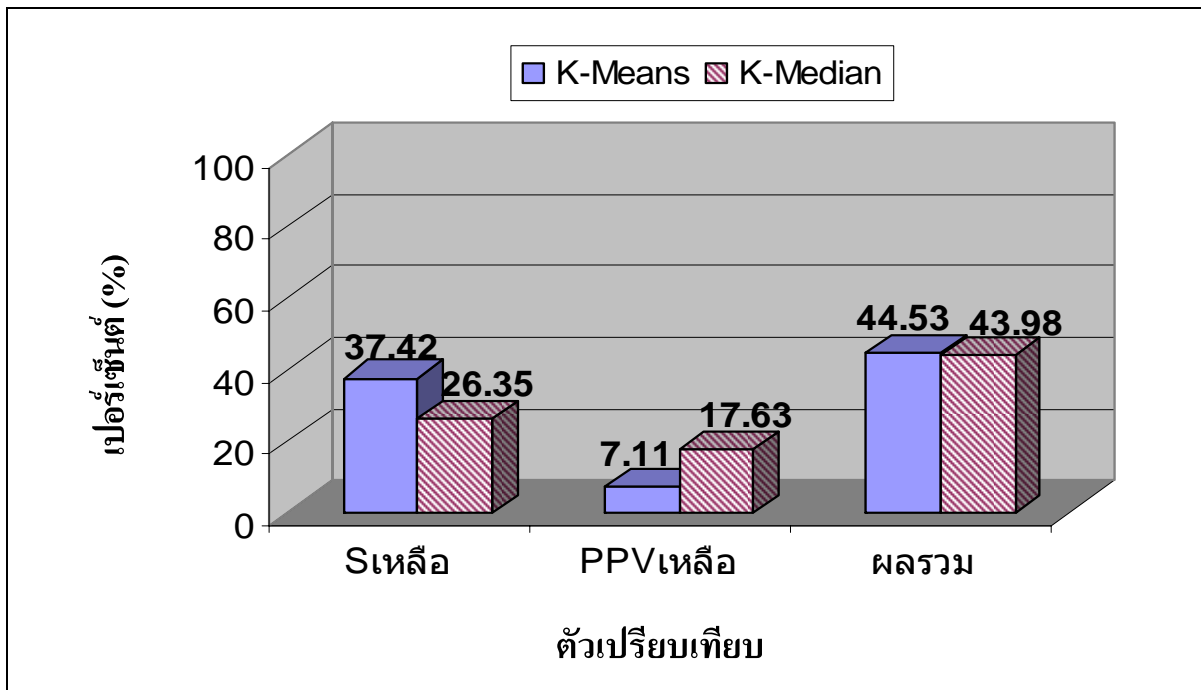
วิธีการ	$S_{ขาด}$ (%)	$PPV_{เกิน}$ (%)	ผลรวม
เฉลี่ย K-Means	37.42	7.11	44.53
เฉลี่ย K-Median	26.35	17.63	43.98

จากตาราง 4.4 เมื่อหาผลรวมของเปอร์เซ็นต์ที่ผู้เชี่ยวชาญจะต้องแก้ไข (ส่วนที่ผู้เชี่ยวชาญจะต้องเพิ่มหรือนำออกจากโปรแกรม) พบว่า ผลรวมของการแก้ไขทั้งหมดโดยใช้การตัดแยกเซลล์ออกจากพื้นหลังของภาพด้วยวิธี K-Means คือ 44.53 และด้วยวิธี K-Median คือ 43.98 ซึ่งจะเห็นได้ว่าการใช้การตัดแยกเซลล์ออกจากพื้นหลังของภาพด้วยวิธี K-Median จะมีผลรวมของการแก้ไขน้อยกว่าการใช้การตัดแยกเซลล์ออกจากพื้นหลังของภาพด้วยวิธี K-Means

จากการทดลองทั้งหมดจึงถือว่าการใช้วิธี K-Median สามารถแยกเซลล์มะเร็งได้่านมออกจากรากพื้นหลังได้ดีกว่าเมื่อทำการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญแล้ว

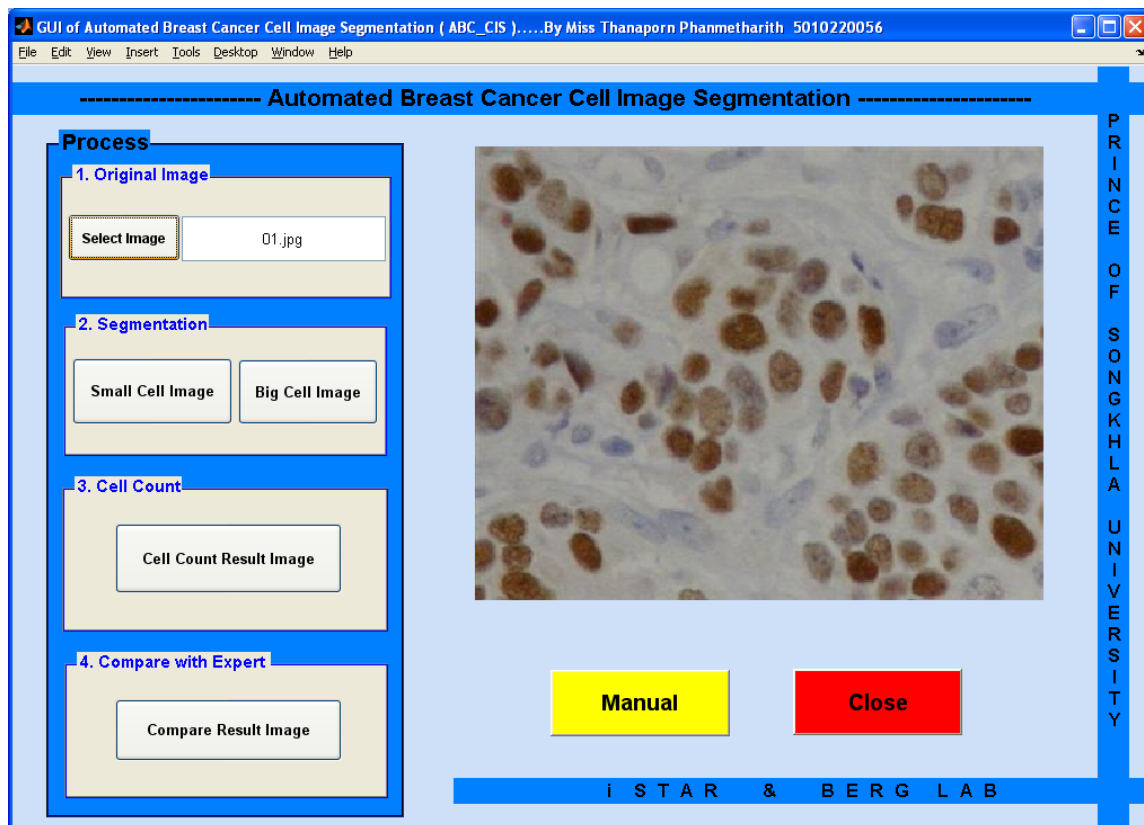


ภาพประกอบ 4.6 กราฟแสดงค่าเฉลี่ย S, PPV และค่าเฉลี่ยของทั้ง 2 ค่า ระหว่าง K-Means และ K-Median



ภาพประกอบ 4.7 กราฟเปรียบเทียบค่า S, PPV ที่เหลือ ระหว่าง K-Means และ K-Median

จากผลการทดลองสามารถนำกระบวนการและผลลัพธ์ทั้งหมดที่ได้ไปสร้างเป็นระบบการตัดแยกเซลล์มะเร็งเต้านมแบบอัตโนมัติ (Automated Breast Cancer Cell Image Segmentation: ABC_CIS) แสดงตัวอย่างหน้าจอของระบบในภาพประกอบ 4.8 และสามารถคู่มือการใช้งานระบบได้ในภาคผนวก ข



ภาพประกอบ 4.8 ตัวอย่างหน้าจอ GUI ของระบบการตัดแยกเซลล์มะเร็งเต้านมแบบอัตโนมัติ

บทที่ 5

บทสรุปและข้อเสนอแนะ

วิทยานิพนธ์นี้ได้พัฒนาระบบการตัดแยกภาพเซลล์มะเร็งเต้านมแบบอัตโนมัติ หรือ Automated Breast Cancer Cell Image Segmentation (ABC_CIS) เพื่อช่วยให้แพทย์สามารถวิเคราะห์จำนวนเซลล์มะเร็งเต้านมในภาพได้อย่างสะดวกรวดเร็ว เพื่อเป็นประโยชน์ในการวินิจฉัยการให้ยาแก่ผู้ป่วยโรคมะเร็งเต้านมว่าผู้ป่วยรายนั้นควรจะได้รับการรักษาหรือไม่

5.1 สรุปผลการวิจัย

การตัดแยกภาพเซลล์มะเร็งเต้านมแบบอัตโนมัติ หรือ Automated Breast Cancer Cell Image Segmentation (ABC_CIS) ประกอบด้วย 5 ขั้นตอนหลัก ได้แก่

5.1.1 การเตรียมภาพเซลล์มะเร็งเต้านม

- 1) ภาพเซลล์มะเร็งเต้านมต้นฉบับที่ได้มาจากผู้เชี่ยวชาญนั้น ในงานวิจัยนี้จะแบ่งภาพออกเป็น 2 ประเภทคือ 1.1) ภาพเซลล์ขนาดเล็ก 1.2) ภาพเซลล์ขนาดใหญ่
- 2) ในขั้นตอนนี้สามารถแปลงภาพดิจิทัลซึ่งเป็นภาพเซลล์มะเร็งเต้านมต้นฉบับให้อยู่ในรูปแบบตัวเลขของสี RGB ที่สามารถนำไปสู่ขั้นตอนต่อไปได้

5.1.2 การตัดแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังของภาพ

ในขั้นตอนนี้ใช้เทคนิคของการเรียนรู้แบบไม่มีผู้สอน โดยใช้ขั้นตอนวิธี *K*-Means Clustering และ *K*-Median Clustering ซึ่งสามารถแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังของภาพได้ โดยที่ไม่ต้องให้ผู้เชี่ยวชาญระบุส่วนประกอบต่างๆ ในภาพ แต่ระบบจะสามารถเรียนรู้ได้เอง จากการทดลองทั้งหมดสามารถสรุปได้ว่า *K*-Median Clustering สามารถตัดแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังของภาพได้ดีกว่าวิธี *K*-Means Clustering เมื่อผ่านการเปรียบเทียบผลการนับเซลล์ของ 2 วิธี กับผลการนับเซลล์ของผู้เชี่ยวชาญแล้ว สังเกตได้จากค่าความถูกต้องที่ได้

5.1.3 การกำจัดสิ่งรบกวนออกจากภาพ

- 1) ในขั้นตอนนี้ใช้เทคนิค Morphological Operations ในการกำจัดส่วนที่ไม่ใช่เซลล์มะเร็งเต้านมออกจากภาพ
- 2) ในขั้นตอนนี้เป็นส่วนที่ใช้ระบุนขนาดเริ่มแรกของส่วนที่เป็นสิ่งรบกวนในภาพของภาพเซลล์มะเร็งเต้านม 2 ประเภท คือ ภาพเซลล์ขนาดเล็กและภาพเซลล์ขนาดใหญ่

3) ขั้นตอนนี้สามารถกำจัดสิ่งรบกวนในภาพออกจากเซลล์มะเร็งเต้านมและพื้นหลังของภาพได้ เพื่อเป็นประโยชน์ในขั้นตอนการนับเซลล์มะเร็งในภาพต่อไป

5.1.4 การนับเซลล์มะเร็งเต้านมในภาพ

ในขั้นตอนนี้สามารถนับจำนวนส่วนที่เป็นเซลล์มะเร็งเต้านมในภาพได้ ซึ่งแยกตัวออกจากพื้นหลังของภาพอย่างชัดเจน โดยปราศจากสิ่งรบกวนในภาพ ซึ่งสามารถใช้ได้กับภาพเซลล์ขนาดเล็กและภาพเซลล์ขนาดใหญ่

5.1.5 การเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ

1) ในขั้นตอนนี้จะมีภาพผลการนับเซลล์จากผู้เชี่ยวชาญซึ่งได้ผลมาจากผู้เชี่ยวชาญด้านนี้โดยตรง และผลการนับเซลล์จากระบบ ABC_CIS

2) ในงานวิจัยนี้จะนำผลการนับเซลล์ของผู้เชี่ยวชาญมาเปรียบเทียบกับผลการนับเซลล์ของระบบ ซึ่งจะได้ผลออกมาเป็นภาพผลการเปรียบเทียบการนับเซลล์และเปอร์เซ็นต์ของค่า S และ PPV ซึ่งเป็นตัวบ่งบอกว่า ผลการนับเซลล์มีความแตกต่างกันในด้านบวกและด้านลบมากน้อยเพียงใด และเปรียบเทียบค่าเฉลี่ยของค่า S และ PPV ที่ได้ สามารถสรุปผลตัวเลขต่างๆได้เป็นตารางและกราฟประกอบเพื่อนำไปสู่การวินิจฉัยการให้ยาแก่ผู้ป่วยต่อไป

สิ่งที่ได้จากการทำวิจัยทั้งหมด คือ ได้ระบบการตัดแยกภาพเซลล์มะเร็งเต้านมแบบอัตโนมัติ หรือ Automated Breast Cancer Cell Image Segmentation (ABC_CIS) ซึ่งสามารถนับเซลล์มะเร็งเต้านมทั้งหมดในภาพได้ นอกจากนั้นยังมีการเปรียบเทียบผลการนับของระบบกับผลการนับของผู้เชี่ยวชาญและคำนวณออกมาเป็นเปอร์เซ็นต์ของค่า S และ PPV ซึ่งสามารถแสดงให้เห็นถึงความถูกต้องทั้งทางด้านบวกและทางด้านลบได้ ซึ่งจากการทดลองจะได้ผลความถูกต้องในระดับที่น่าพอใจ โดยขั้นตอนทั้งหมดจะเป็นแบบอัตโนมัติคือ แพทย์ไม่ต้องสอนให้ระบบได้รู้จักกับส่วนต่างๆของภาพ แต่ระบบจะสามารถเรียนรู้ได้เอง ซึ่งจากการทำวิจัยทั้งหมดทำให้ทราบได้ว่า การตัดแยกเซลล์มะเร็งเต้านมแบบอัตโนมัติสามารถช่วยให้ความสะดวกรวดเร็วแก่แพทย์ในการวินิจฉัยการให้ยาแก่ผู้ป่วยมะเร็งเต้านมได้

5.2 ปัญหาและอุปสรรค

เนื่องจากภาพเซลล์มะเร็งเต้านมที่ต้องนำมาจากแพทย์ผู้เชี่ยวชาญเรื่องนี้โดยตรง จึงต้องใช้เวลาในการรอขั้นตอนการทำภาพและผลการนับเซลล์จากผู้เชี่ยวชาญ เพื่อให้ได้ภาพที่ต้องการและผลการนับเซลล์ที่ถูกต้อง

5.3 ข้อเสนอแนะ

1) ควรใช้คุณสมบัติอื่นในการพิจารณาแยกเซลล์ออกจากพื้นหลัง เนื่องจากหลังจากที่ได้ทำงานวิจัยนี้พบว่า การใช้คุณสมบัติเรื่อง สี อย่างเดียว ยังไม่เพียงพอต่อการแยกเซลล์มะเร็งด้านมออกจากพื้นหลังของภาพ เพื่อให้ความถูกต้อง 100 เปอร์เซ็นต์ เนื่องจากในภาพเซลล์มะเร็งด้านมที่ได้มานั้น นอกจากจะมีส่วนที่เป็นเซลล์มะเร็งด้านมและพื้นหลังของภาพแล้ว ยังมีส่วนประกอบอื่นๆอีกมากมาย เช่น ไขมัน ต่อมน้ำนม ท่อน้ำนม เป็นต้น ซึ่งการใช้คุณสมบัติเรื่องสีอย่างเดียวยังไม่เพียงพอต่อการแยกสิ่งเหล่านี้ออกจากกันได้ทั้งหมด

ดังนั้น นอกจากคุณสมบัติเรื่องสีแล้ว จะต้องใช้คุณสมบัติอื่นในการพิจารณาแยกเซลล์ออกจากพื้นหลังอีกด้วย เช่น ขนาดของเซลล์ รูปร่างของเซลล์ ลักษณะเฉพาะอื่นๆ เป็นต้น เพื่อให้ได้ผลการแยกเซลล์ที่ดีขึ้น ซึ่งจะส่งผลให้ผลการนับเซลล์และผลการเปรียบเทียบกับผู้เชี่ยวชาญ ได้ผลที่ดีมากขึ้นด้วย หรือใช้ร่วมกับขั้นตอนวิธีในการแยกเซลล์ออกจากพื้นหลังวิธีอื่นๆ

2) ควรนำผลการนับของโปรแกรมไปให้ผู้เชี่ยวชาญพิจารณาอีกครั้ง เพื่อให้ได้ผลการนับเซลล์ที่ถูกต้อง เนื่องจากค่าความถูกต้องที่ได้จากการเปรียบเทียบผลการนับกับผู้เชี่ยวชาญ กรณีผลการนับของโปรแกรมที่ไม่ตรงกับผู้เชี่ยวชาญ เช่น โปรแกรมระบุว่าเซลล์ที่ 1 เป็นเซลล์มะเร็ง แต่ผู้เชี่ยวชาญไม่ได้ระบุว่าเซลล์นั้นเป็นเซลล์มะเร็ง ก็ยังไม่สามารถสรุปได้ว่าโปรแกรมนับผิดเสมอไป ผลการนับของผู้เชี่ยวชาญอาจจะไม่แม่นยำทุกครั้ง เพราะผู้เชี่ยวชาญใช้ประสบการณ์ในการนับเซลล์แต่ละครั้ง จึงอาจจะเกิดความไม่แน่นอนขึ้น แต่โปรแกรมที่สร้างจากขั้นตอนของคอมพิวเตอร์จะมีความแม่นยำและเที่ยงตรงกว่า

บรรณานุกรม

- ความหมายของสกุล GIF, TIFF, JPEG, BMP. <http://www.navy22.com/smf/index.php?topic=13767.0> (accessed 1/2/2009).
- นวกัก เอื้ออนันต์, “Image Segmentation.ppt,” ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่น, 2002.
- เรียนรู้ทฤษฎีของสีคอมพิวเตอร์. <http://techno.edu.buu.ac.th/wbi/file.php/1/CMI-webpage/cmi505/photoshop/saratwadee/page3.html#p2> (accessed 1/5/2008).
- อดุลย์ รัตนวิจิตรศิลป์. การตรวจเต้านมด้วยอัลตราซาวด์กับแมมโมแกรมต่างกันอย่างไร. <http://www.si.mahidol.ac.th/sidoctor/e-pl/articleDetail.asp?id=307> (accessed 1/2/2009).
- อดุลย์ รัตนวิจิตรศิลป์. การรักษามะเร็งเต้านมด้วยวิธีฮอร์โมน (Hormonal treatment) ตอนที่ 1. <http://www.si.mahidol.ac.th/sidoctor/e-pl/articleDetail.asp?id=455> (accessed 1/8/2008).
- A. McAndrew, “Introduction to Digital Image Processing with MATLAB,” Massachusetts: Thomson Learning, Inc., 2004.
- Archive for February 2008. <http://machincuepa.com/?m=200802> (accessed 1/5/2008).
- Clinical Correlations. <http://www.clinicalcorrelations.org/?p=87> (accessed 1/5/2008).
- Computer Graphics Notes. <http://www.cse.fau.edu/~maria/COURSES/COP4930-GS/colornotes.htm> (accessed 1/5/2008).
- E. Subramaniam, T. K. Liung, M. Y. Mashor, N. A. M. Isa, “Breast Cancer Diagnosis Systems: A Review,” International Journal of The Computer, the Internet and Management, vol. 14, no.2, pp. 24 – 35, 2006.
- F. Mayer, “Topographic distance and watershed lines,” Signal Processing, Vol. 38, 1994, pp. 113-125.
- F. Y. Shih and S. Cheng, “Automatic seeded region growing for color image segmentation,” Image and Vision Computing, vol. 23, pp. 877-886, 2005.
- H. Deng and D. A. Clausi, “Unsupervised image segmentation using a simpleMRFmodel with a newimplementation scheme,” Pattern Recognition, vol. 37, pp. 2323-2335, 2004.
- H. Jeong, T. Kim, H. Hwang, H. Choi, H. Park, H. Cho, “Comparison of thresholding methods for breast tumor cell segmentation,” Enterprise networking and Computing in Healthcare Industr, HEALTHCOM 2005. Proceedings of 7th International Workshop, pp. 392-395, 2005.
- I. H. Witten and E. Frank, “Data Mining Practical Machine Learning Tool and Techniques,” Morgan Kaufmann Publishers, 2005.

- Introduction to Web Graphic. http://www.geocities.com/naruthep_t/Knowledge.html (accessed 25/9/2008).
- JPEG. <http://th.wikipedia.org/wiki/JPEG> (accessed 15/1/2009).
- JPEG & GIF. <http://www.student.chula.ac.th/~49702413/jpeggif.htm> (accessed 1/2/2009).
- M. Ozden and E. Polat, "A color image segmentation approach for content-based image retrieval," *Pattern Recognition*, vol. 40, pp. 1318-1325, 2007.
- M. Sonka, V. Hlavac, and R. Boyle, "Image Processing, ed.," Ontario: Thomson Learning, 2008.
- Machine Learning. http://store.tkc.go.th/multimedia/tkcteam/Flash/PHASE3/WU26_Machine%20Learning=2007-09-27/Machine_Learning.html (accessed 9/7/2008).
- Macromedia Dreamweaver3. http://www.geocities.com/member1980/dreamweaver_09.html (accessed 25/9/2008).
- P. Phukpattaranont. and P. Boonyaphiphat, "Color based segmentation of nuclear stained breast cancer cell images," *ECTI Transaction on Electrical Eng. Electronics and Communications*, vol. 5, no 2, pp. 158-164, 2007.
- P. Soille, "Morphological Image Analysis: Principles and Applications," 2nd ed., New York: Springer-Verlag, 2002.
- Q. Chen, X. Yang, and E. M. Petriu, "Watershed segmentation for binary images with difference transforms," in *Proceeding of the 3rd IEEE Haptic, Audio and Visual Environments and Their Applications*, 2004, pp. 111-116.
- R. C. Gonzalez and R. E. Woogs, "Digital Image Processing. Second Edition," Prentice Hall, 2002.
- R. Witte and J. Witte, "Statistics. Seven Edition," WILEY, 2004.
- Supervised Learning. <http://en.wikipedia.org/wiki/Supervisedlearning> (accessed 1/5/2008).
- T. Acharya and A. K. Ray, "Image Processing Principles and Applications," Wiley-Interscience, 2005.
- T. Pongpram and P. Phukpattaranont, "Color-Based Segmentation of Breast Cancer Cell Using K-Mean Clustering," *The Proceeding of PSU-Engineering Conference*, 2008.
- The Graphics Grapevine. http://www.kauaidesign.com/graphics%20grapevine/Vol_%201%20No_%204%20THE%20GRAPHICS%20GRAPEVINE.htm (accessed 1/5/2008).
- The MathWorksTM. Image Processing ToolboxTM 6 User's Guide. http://www.mathworks.com/access/helpdesk/help/pdf_doc/images/images_tb.pdf (accessed 1/10/2008)
- Unsupervised learning. <http://en.wikipedia.org/wiki/Unsupervisedlearning> (accessed 1/5/2008).
- What Role Do Hormones Play in Breast Cancer Treatment?. http://www.Breastcancer.org/treatment/hormonal/what_is_it/hormone_role.jsp (accessed 9/7/2008).

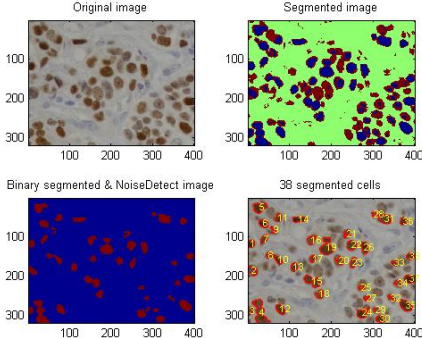
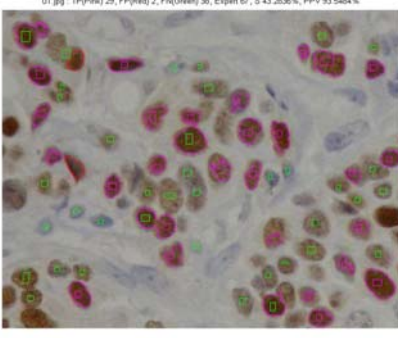
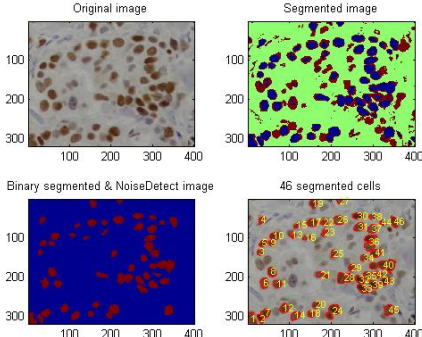
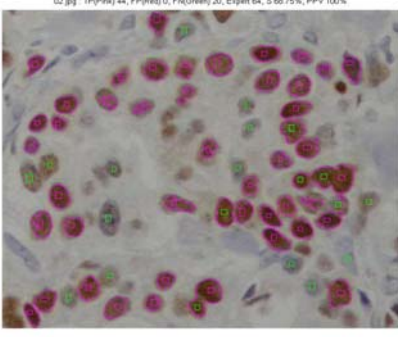
Z. Nian Li and M. S. Drew, "Fundamentals of Multimedia," Pearson Education International, 1972.

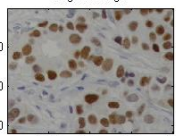
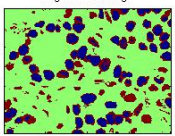
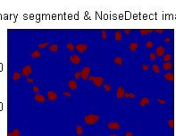
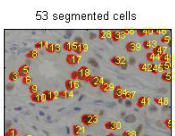
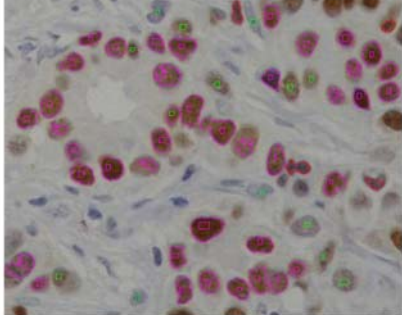
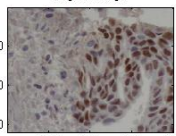
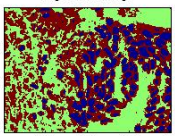
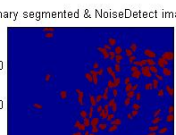
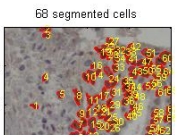
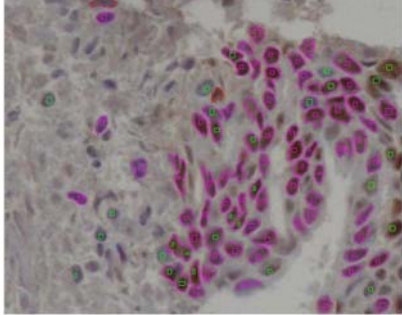
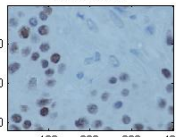
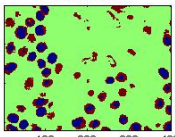
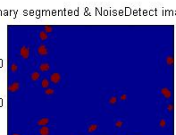
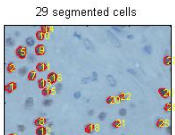
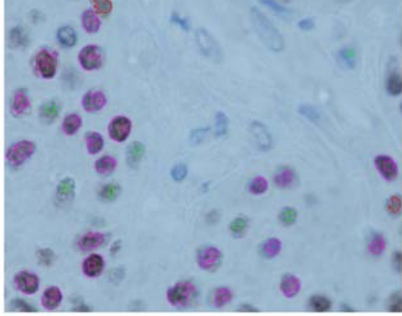
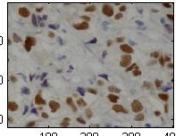
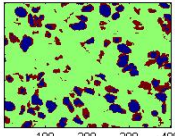
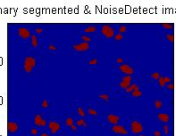
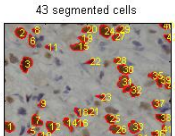
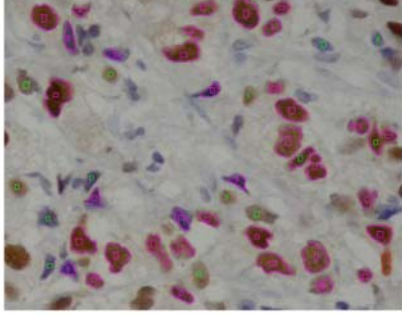
ภาคผนวก ก

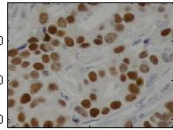
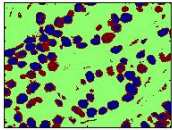
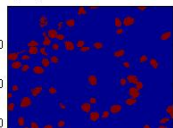
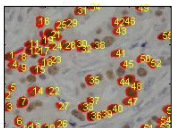
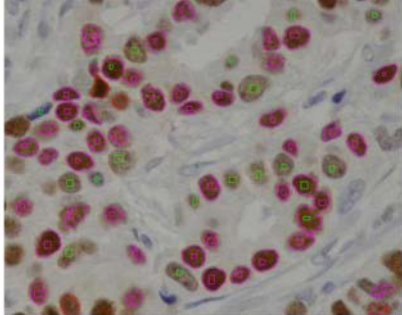
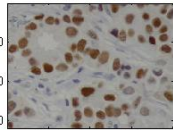
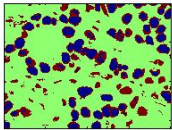
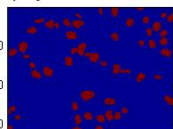
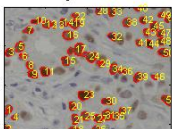
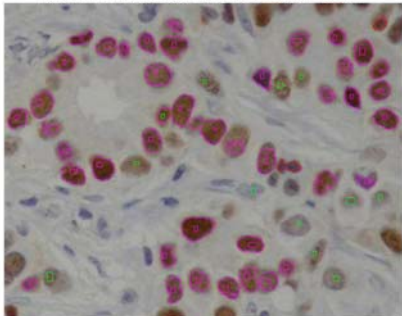
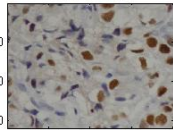
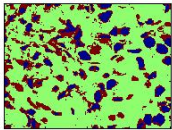
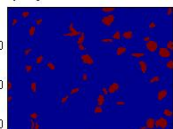
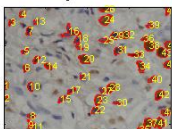
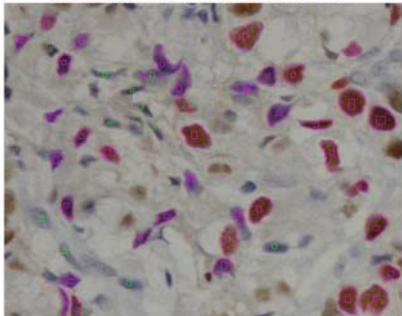
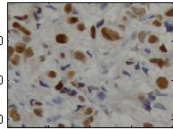
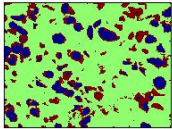
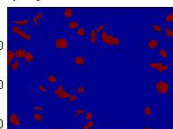
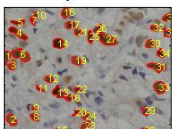
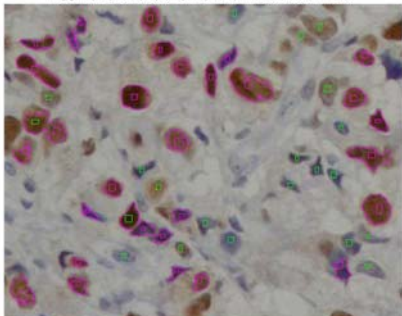
ผลการทดลองโดยใช้ K-Means

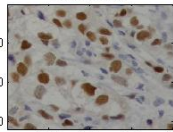
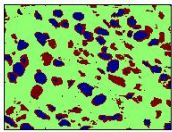
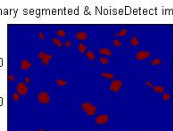
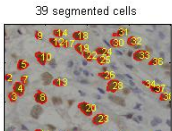
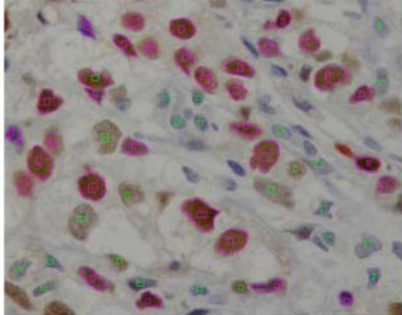
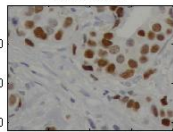
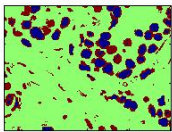
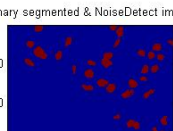
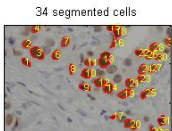
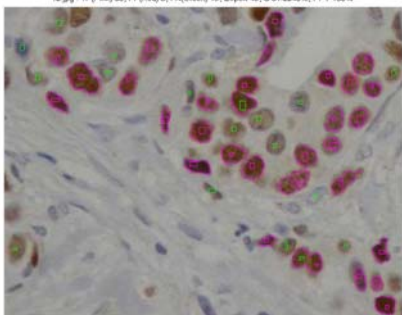
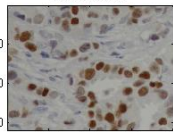
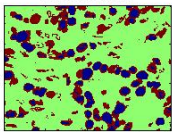
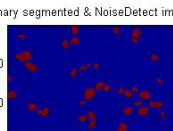
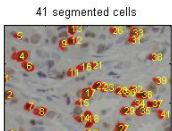
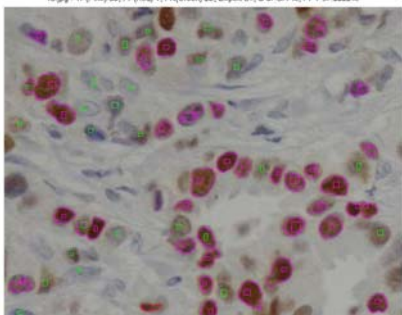
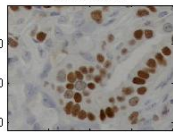
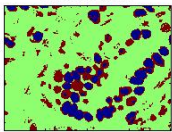
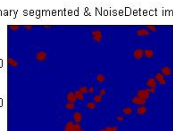
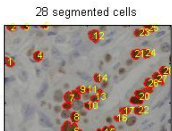
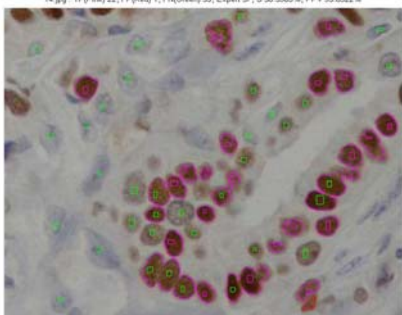
ผลการแยกเซลล์ออกจากพื้นหลังของภาพ ผลการกำจัดสิ่งรบกวน ผลการนับเซลล์ และผลการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ โดยใช้ K-Means ในขั้นตอนของการแยกเซลล์ออกจากพื้นหลังของภาพ แสดงในตาราง ก.1

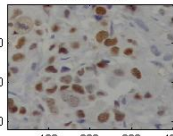
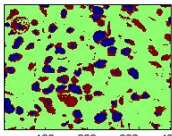
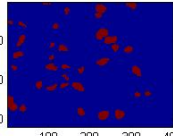
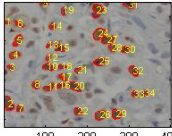
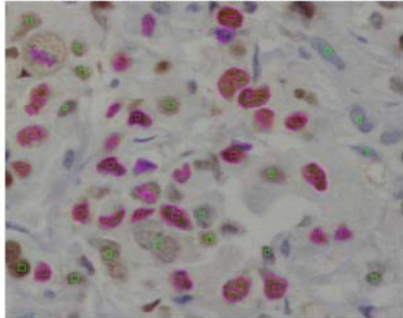
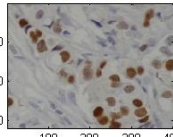
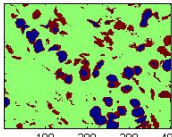
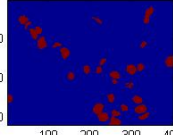
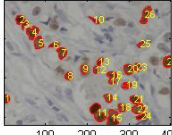
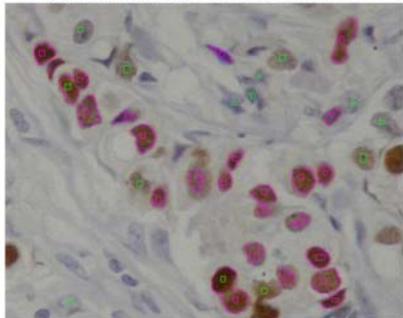
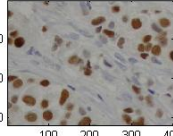
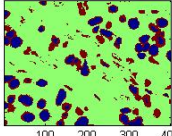
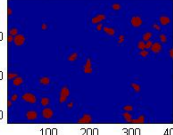
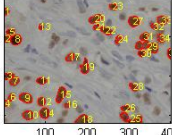
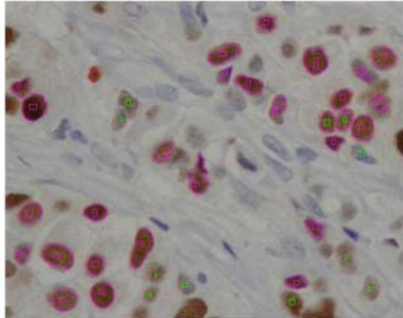
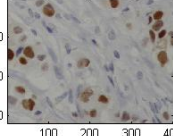
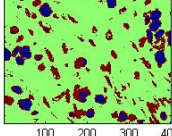
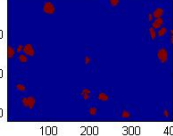
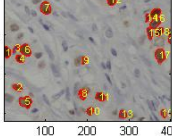
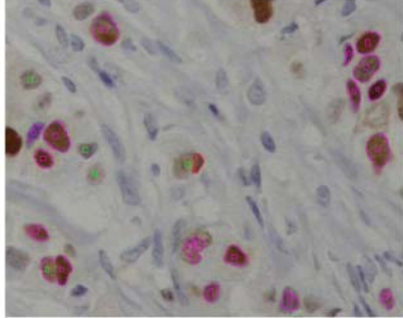
ตาราง ก.1 ผลการแยกเซลล์ กำจัดสิ่งรบกวน นับเซลล์และเปรียบเทียบกับผู้เชี่ยวชาญโดยใช้ K-Means ในขั้นตอนการตัดแยกภาพเซลล์มะเร็งเต้านม

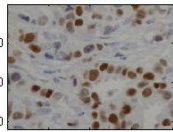
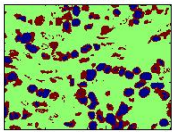
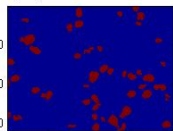
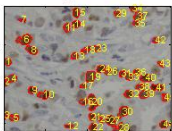
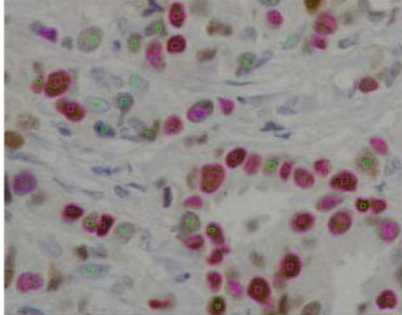
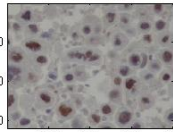
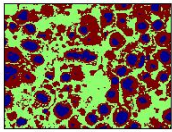
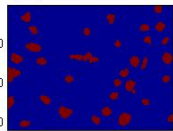
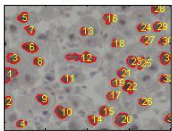
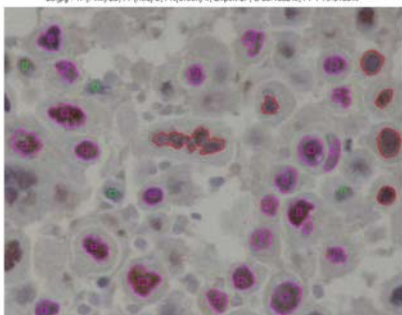
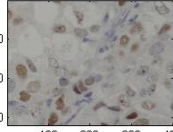
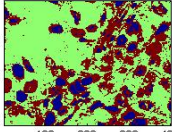
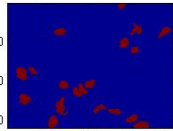
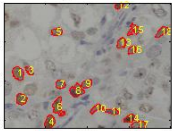
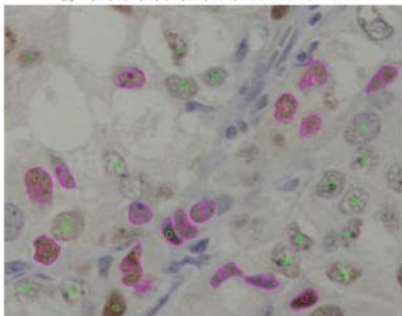
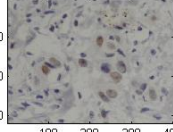
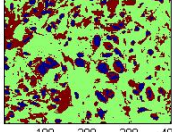
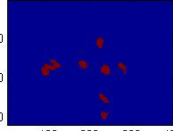
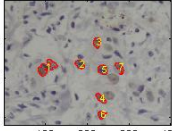
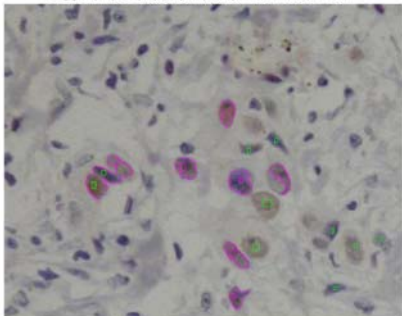
ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
01	 <p>Original image</p> <p>Segmented image</p> <p>Binary segmented & NoiseDetect image</p> <p>38 segmented cells</p>	 <p>01.jpg TP(Pink) 29, FP(Pink) 2, FN(Green) 30, Expert 67, S 43.2636%, PPV 93.5464%</p>
02	 <p>Original image</p> <p>Segmented image</p> <p>Binary segmented & NoiseDetect image</p> <p>46 segmented cells</p>	 <p>02.jpg TP(Pink) 44, FP(Pink) 0, FN(Green) 20, Expert 64, S 66.75%, PPV 100%</p>

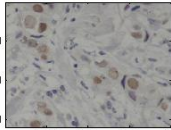
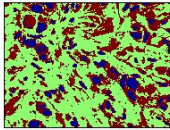
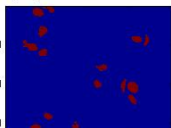
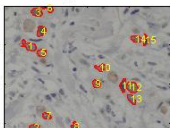
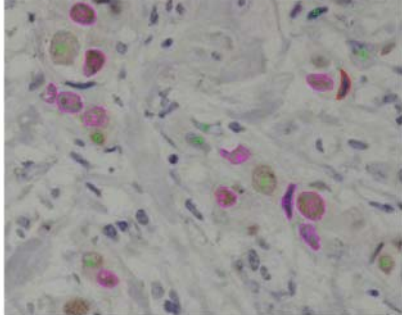
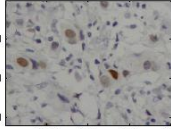
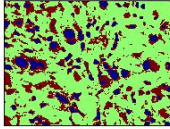
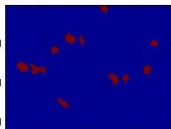
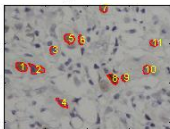
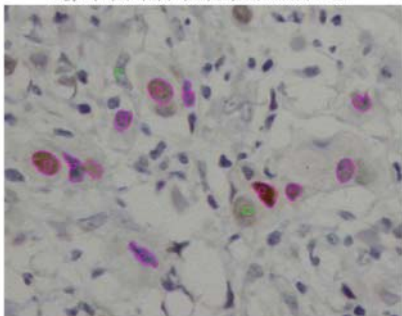
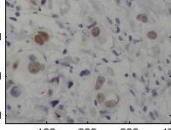
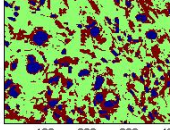
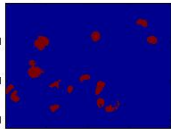
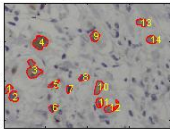
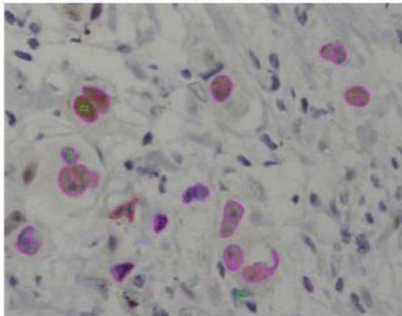
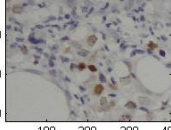
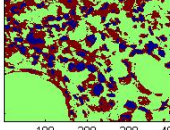
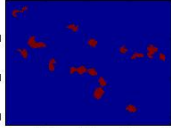
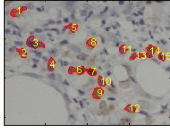
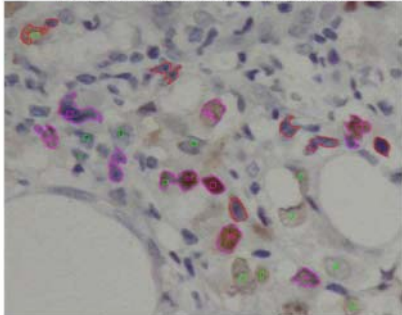
ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
03	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>53 segmented cells</p> 	<p>03.jpg TP(Red) 40, FP(Red) 4, FN(Green) 14, Expert 54, S 74.0741%, PPV 30.9091%</p> 
04	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>68 segmented cells</p> 	<p>04.jpg TP(Red) 61, FP(Red) 2, FN(Green) 29, Expert 90, S 67.7778%, PPV 96.6254%</p> 
05	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>29 segmented cells</p> 	<p>05.jpg TP(Red) 23, FP(Red) 2, FN(Green) 7, Expert 30, S 76.6667%, PPV 92%</p> 
06	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>43 segmented cells</p> 	<p>06.jpg TP(Red) 36, FP(Red) 1, FN(Green) 14, Expert 50, S 72%, PPV 97.2973%</p> 

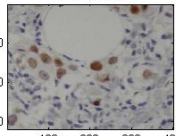
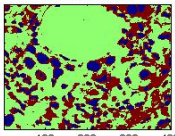
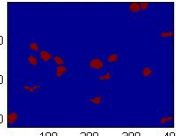
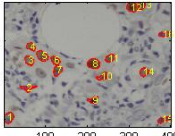
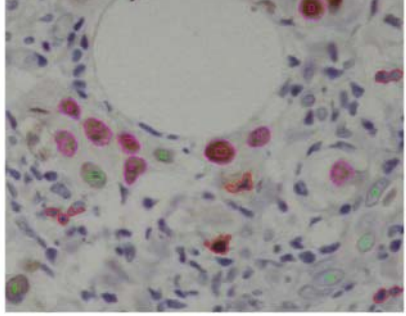
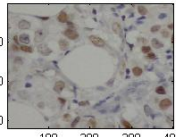
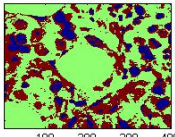
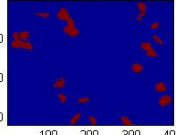
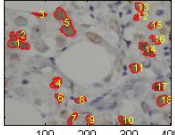
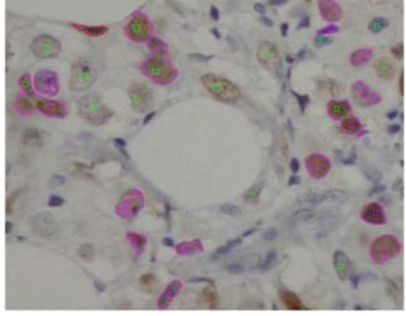
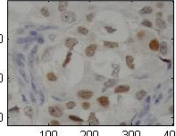
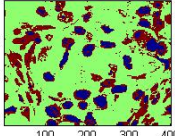
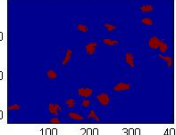
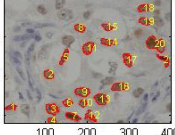
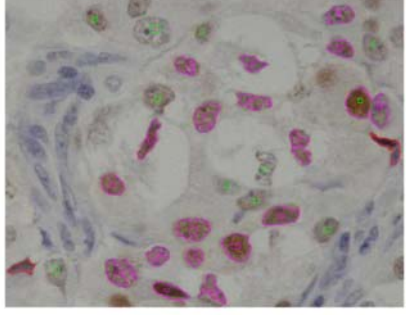
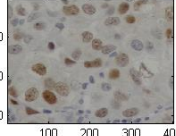
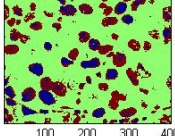
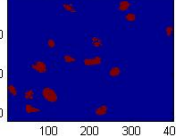
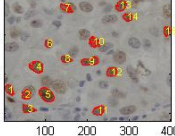
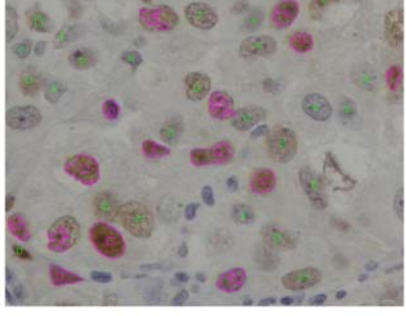
ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
07	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>56 segmented cells</p> 	<p>07.jpg : TP(Pink) 42, FP(Red) 4, FN(Green) 20, Expert 62, S 67.7419%, PPV 91.3043%</p> 
08	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>52 segmented cells</p> 	<p>08.jpg : TP(Pink) 41, FP(Red) 2, FN(Green) 12, Expert 53, S 77.3565%, PPV 95.3486%</p> 
09	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>47 segmented cells</p> 	<p>09.jpg : TP(Pink) 36, FP(Red) 4, FN(Green) 11, Expert 47, S 76.6927%, PPV 92%</p> 
10	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>36 segmented cells</p> 	<p>10.jpg : TP(Pink) 30, FP(Red) 0, FN(Green) 21, Expert 51, S 58.8235%, PPV 100%</p> 

ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
11	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>39 segmented cells</p> 	<p>11.jpg TP(Pink) 32, FP(Red) 1, FN(Green) 25, Expert 57, S 56.1404%, PPV 96.9697%</p> 
12	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>34 segmented cells</p> 	<p>12.jpg TP(Pink) 30, FP(Red) 0, FN(Green) 19, Expert 49, S 61.2245%, PPV 100%</p> 
13	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>41 segmented cells</p> 	<p>13.jpg TP(Pink) 35, FP(Red) 1, FN(Green) 26, Expert 61, S 57.371%, PPV 97.2222%</p> 
14	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>28 segmented cells</p> 	<p>14.jpg TP(Pink) 22, FP(Red) 1, FN(Green) 35, Expert 57, S 38.5965%, PPV 95.6522%</p> 

ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
15	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>34 segmented cells</p> 	<p>15.jpg - TP(Pwk) 30, FP(Pwk) 1, FN(Green) 17, Expert 47, S 63.628%, PPV 96.7742%</p> 
16	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>27 segmented cells</p> 	<p>16.jpg - TP(Pwk) 23, FP(Pwk) 1, FN(Green) 14, Expert 37, S 62.1622%, PPV 95.6333%</p> 
17	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>35 segmented cells</p> 	<p>17.jpg - TP(Pwk) 26, FP(Pwk) 1, FN(Green) 11, Expert 37, S 70.2703%, PPV 95.2623%</p> 
18	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>20 segmented cells</p> 	<p>18.jpg - TP(Pwk) 17, FP(Pwk) 1, FN(Green) 5, Expert 22, S 77.2727%, PPV 94.4444%</p> 

ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
19	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>45 segmented cells</p> 	<p>19.jpg - TP(Pink) 37, FP(Red) 2, FN(Green) 20, Expert 57, S 64.9123%, PPV 94.6718%</p> 
20	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>33 segmented cells</p> 	<p>20.jpg - TP(Pink) 23, FP(Red) 6, FN(Green) 4, Expert 27, S 85.1852%, PPV 79.3103%</p> 
21	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>18 segmented cells</p> 	<p>21.jpg - TP(Pink) 15, FP(Red) 0, FN(Green) 19, Expert 33, S 45.4545%, PPV 100%</p> 
22	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>7 segmented cells</p> 	<p>22.jpg - TP(Pink) 7, FP(Red) 0, FN(Green) 10, Expert 17, S 41.1765%, PPV 100%</p> 

ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
23	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>15 segmented cells</p> 	<p>23.jpg TP(Pink) 11, FP(Red) 1, FN(Green) 9, Expert 20, S 55%, PPV 91.6667%</p> 
24	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>11 segmented cells</p> 	<p>24.jpg TP(Pink) 10, FP(Red) 0, FN(Green) 2, Expert 12, S 83.3333%, PPV 100%</p> 
25	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>14 segmented cells</p> 	<p>25.jpg TP(Pink) 12, FP(Red) 1, FN(Green) 3, Expert 15, S 60%, PPV 92.3077%</p> 
26	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>15 segmented cells</p> 	<p>26.jpg TP(Pink) 9, FP(Red) 5, FN(Green) 12, Expert 21, S 42.8571%, PPV 60%</p> 

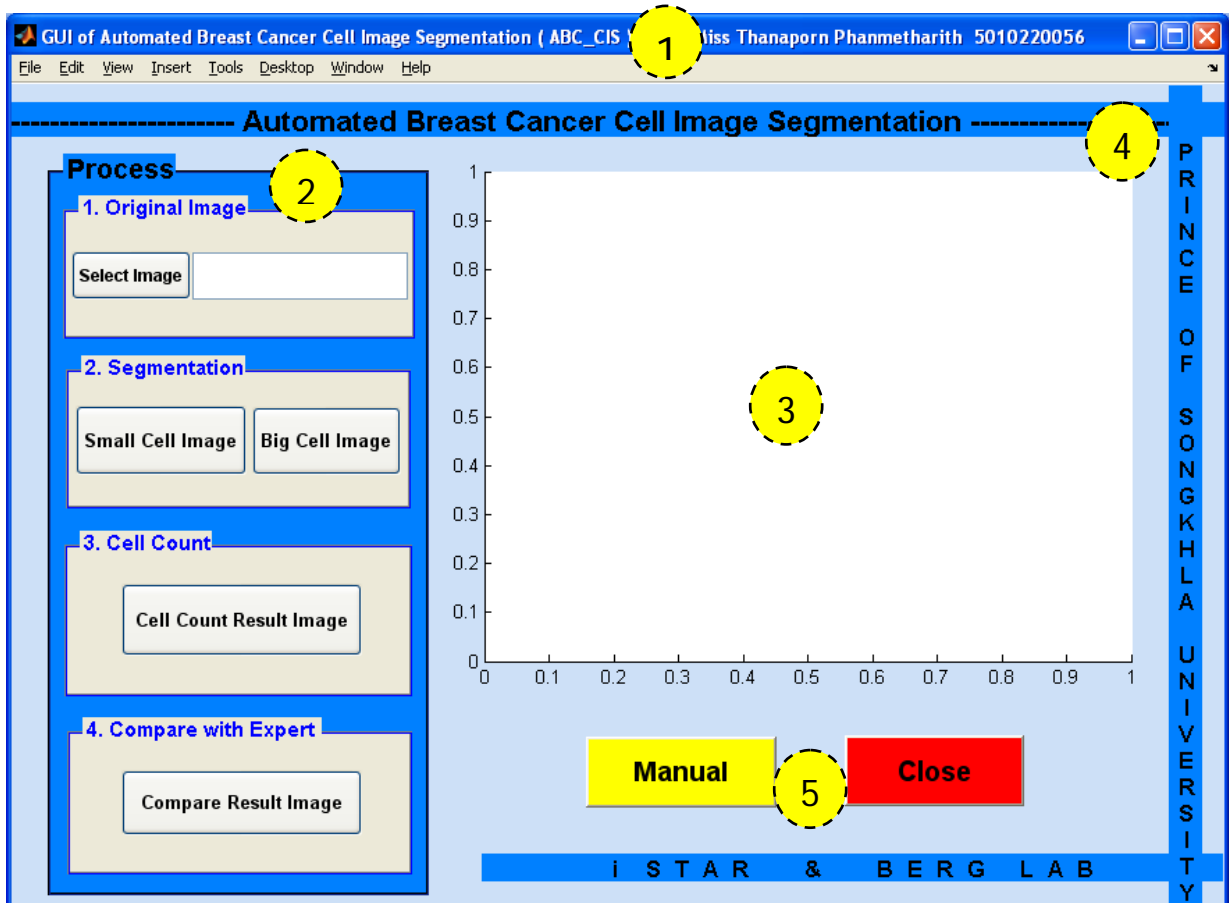
ชื่อภาพ	ผลการแยกเซลล์, กำจัดสิ่งรบกวน และนับเซลล์	ผลการเปรียบเทียบ
27	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>16 segmented cells</p> 	<p>27 µg TP(Pink) 9, FP(Red) 5, FN(Green) 6, Expert 15, S 60%, PPV 64.2857%</p> 
28	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>18 segmented cells</p> 	<p>28 µg TP(Pink) 15, FP(Red) 1, FN(Green) 15, Expert 30, S 50%, PPV 93.75%</p> 
29	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>21 segmented cells</p> 	<p>29 µg TP(Pink) 18, FP(Red) 2, FN(Green) 11, Expert 25, S 62.00%, PPV 90%</p> 
30	<p>Original image</p>  <p>Segmented image</p>  <p>Binary segmented & NoiseDetect image</p>  <p>15 segmented cells</p> 	<p>30 µg TP(Pink) 15, FP(Red) 0, FN(Green) 25, Expert 40, S 37.5%, PPV 100%</p> 

ภาคผนวก ข

คู่มือผู้ใช้ระบบ

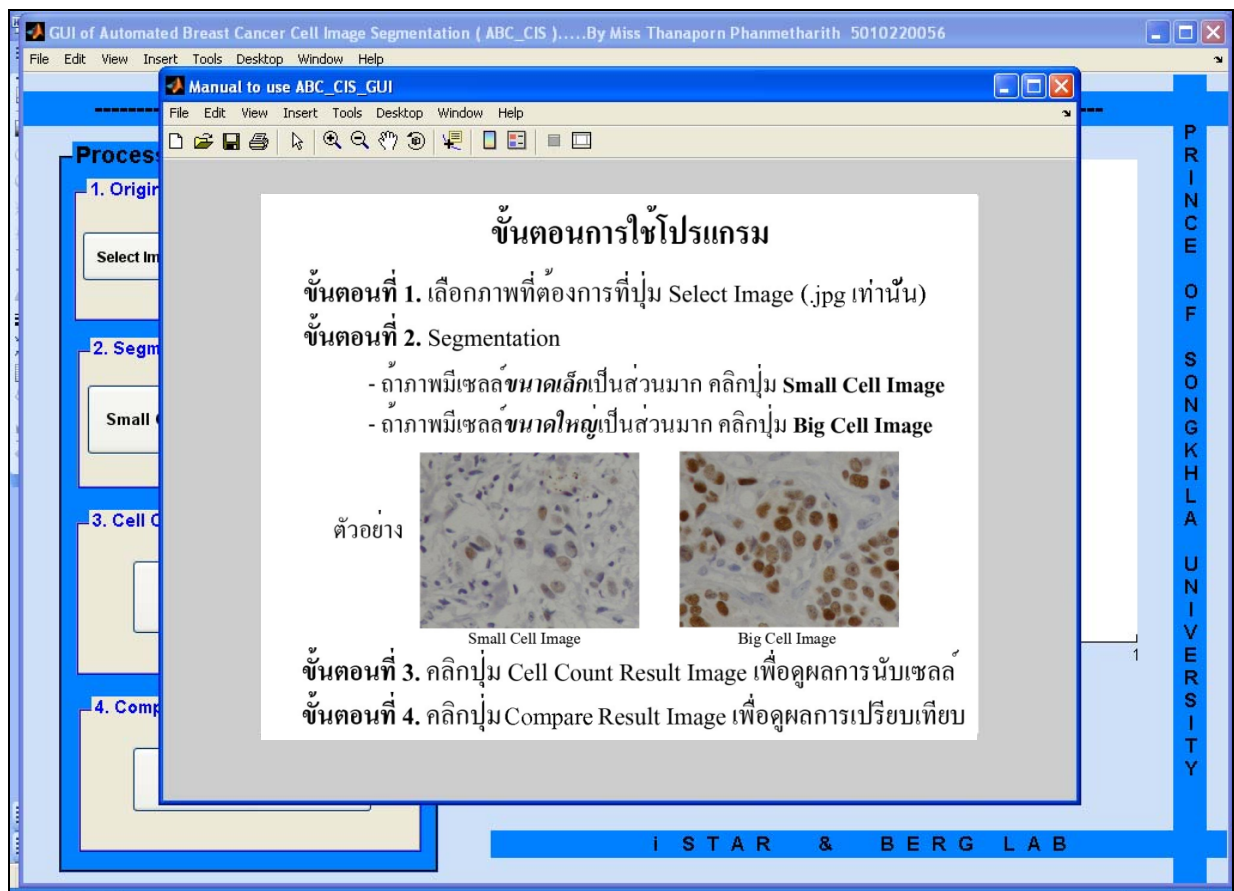
หน้าจอแรกของระบบการตัดแยกภาพเซลล์มะเร็งเรื้องต้นแบบอัตโนมัติ หรือ Automated Breast Cancer Cell Image Segmentation (ABC_CIS) ดังแสดงในภาพประกอบ ข.1 ซึ่งประกอบไปด้วยส่วนต่างๆดังนี้

- 1) ส่วนหัวของหน้าจอระบบ
- 2) ส่วนที่ใช้ในการประมวลผล (Process) ประกอบด้วย 4 ขั้นตอนย่อย ซึ่งจะอธิบายเป็นตัวอย่างประกอบในส่วนถัดไป
- 3) ส่วนที่ใช้ในการแสดงภาพต้นฉบับ
- 4) ส่วนข้อความที่ใช้ตกแต่งหน้าจอ
- 5) ส่วนปุ่มแสดงคู่มือการใช้งาน โปรแกรมและปุ่มปิดโปรแกรม



ภาพประกอบ ข.1 หน้าจอหลักของโปรแกรม ABC_CIS

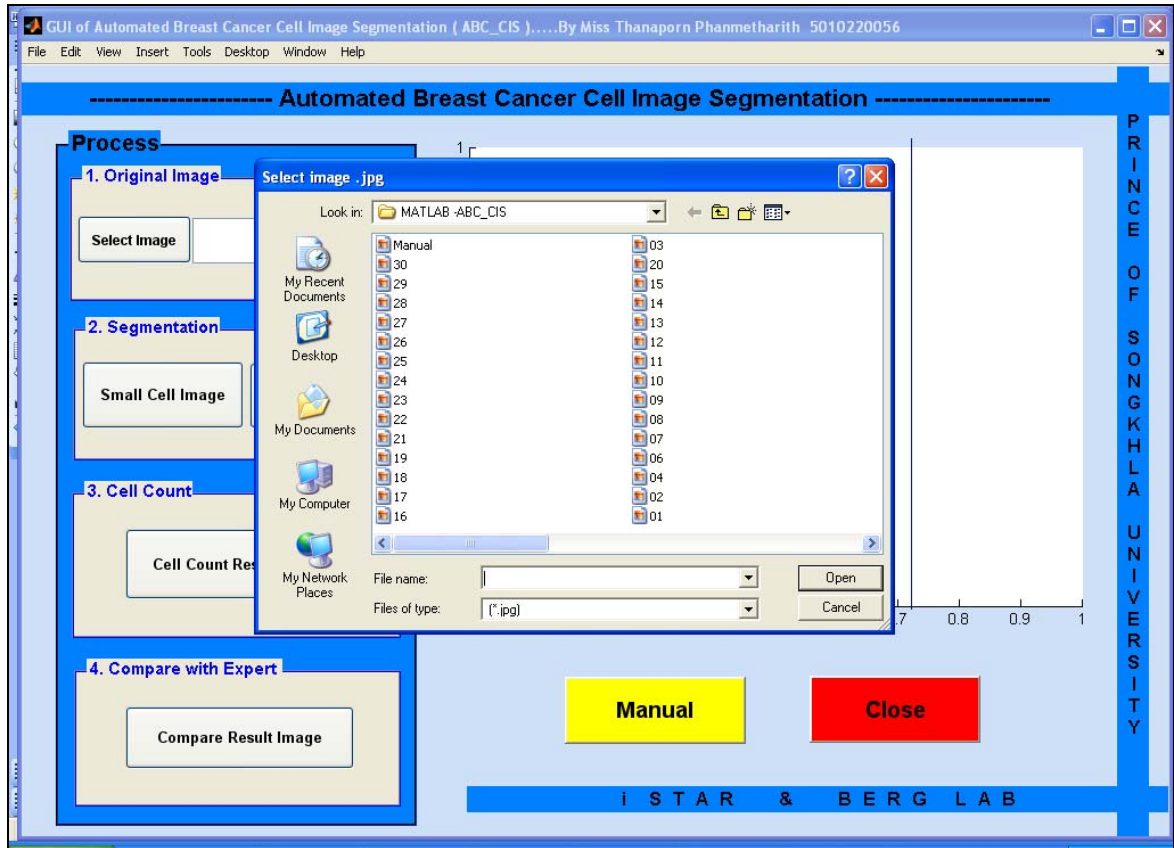
เมื่อเปิดโปรแกรมขึ้นมาในขั้นตอนแรกจะต้องอ่านคู่มือการใช้งานโปรแกรม โดยคลิกที่ปุ่ม Manual และจะแสดงหน้าจอขั้นตอนการใช้โปรแกรม ดังภาพประกอบ ข.2 จากนั้นสามารถเริ่มต้นการใช้งานในส่วนการประมวลผล (Process) ซึ่งมีขั้นตอนดังนี้ 1) เลือกภาพที่ต้องการ (Original Image) 2) ผ่านกระบวนการแยกเซลล์ออกจากภาพ (Segmentation) 3) ผ่านกระบวนการนับเซลล์ (Cell Count) 4) ผ่านกระบวนการเปรียบเทียบผลการนับเซลล์กับผู้เชี่ยวชาญ (Compare with Expert) 5) นำผลการนับและเปรียบเทียบไปใช้ประโยชน์ต่อไป



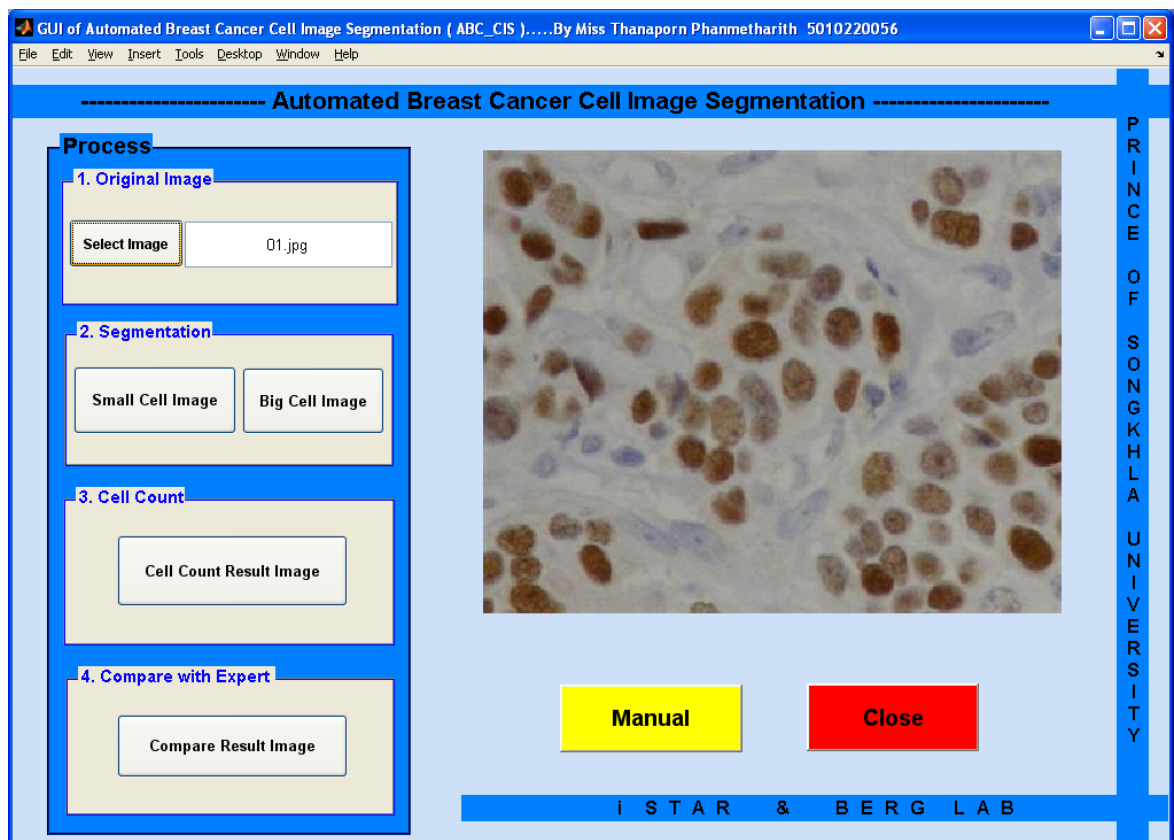
ภาพประกอบ ข.2 ตัวอย่างหน้าจอขั้นตอนการใช้โปรแกรม

ส่วนที่ใช้ในการประมวลผล (Process)

1) Original Image เมื่อคลิกที่ปุ่ม Select Image จะปรากฏหน้าจอให้เลือกภาพเซลล์มะเร็งเต้านมที่ต้องการนำมาแยกเซลล์ ดังภาพประกอบ ข.3 เมื่อเลือกภาพเรียบร้อยแล้ว จะปรากฏภาพที่เลือกในส่วนที่ใช้แสดงภาพต้นฉบับ ดังแสดงตัวอย่างในภาพประกอบ ข.4

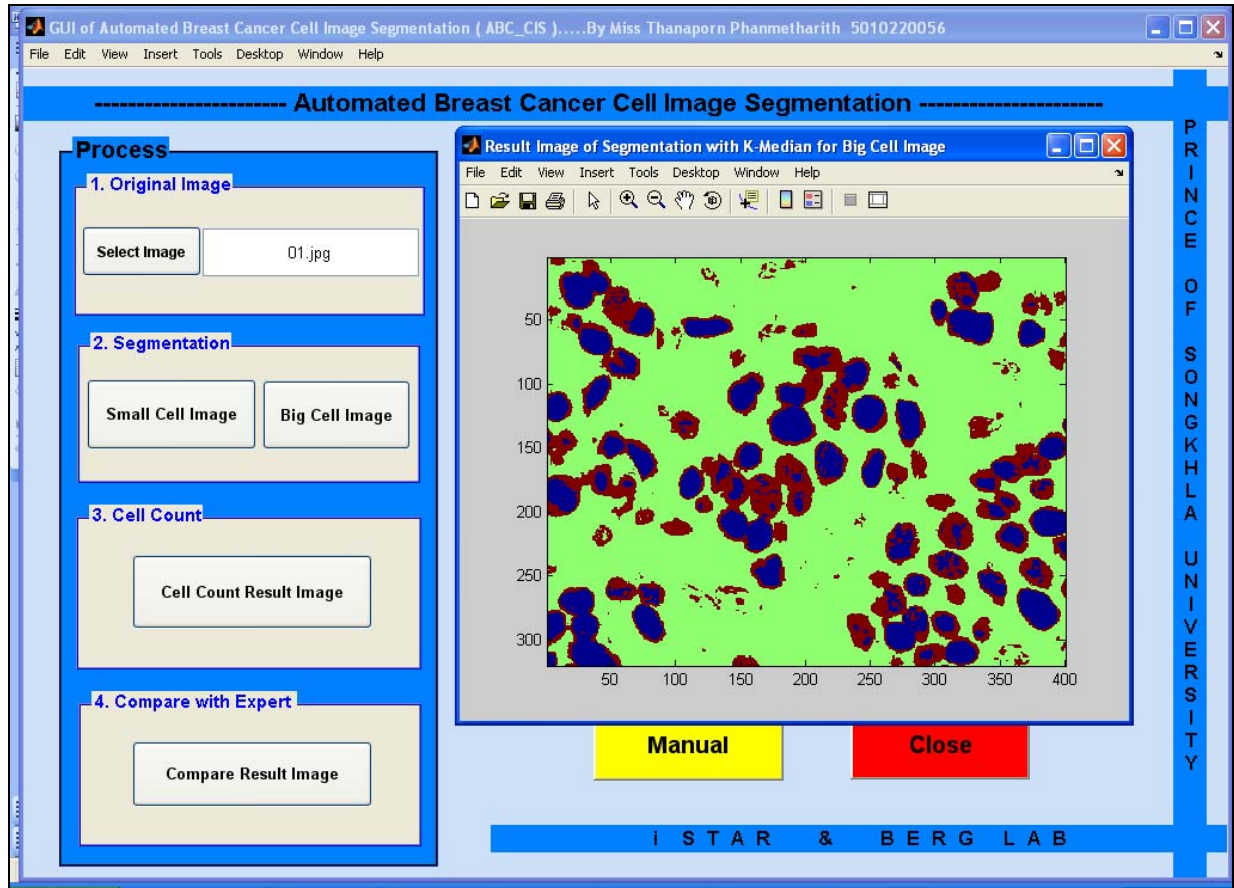


ภาพประกอบ ข.3 ส่วนการเลือกภาพ Original Image บน GUI



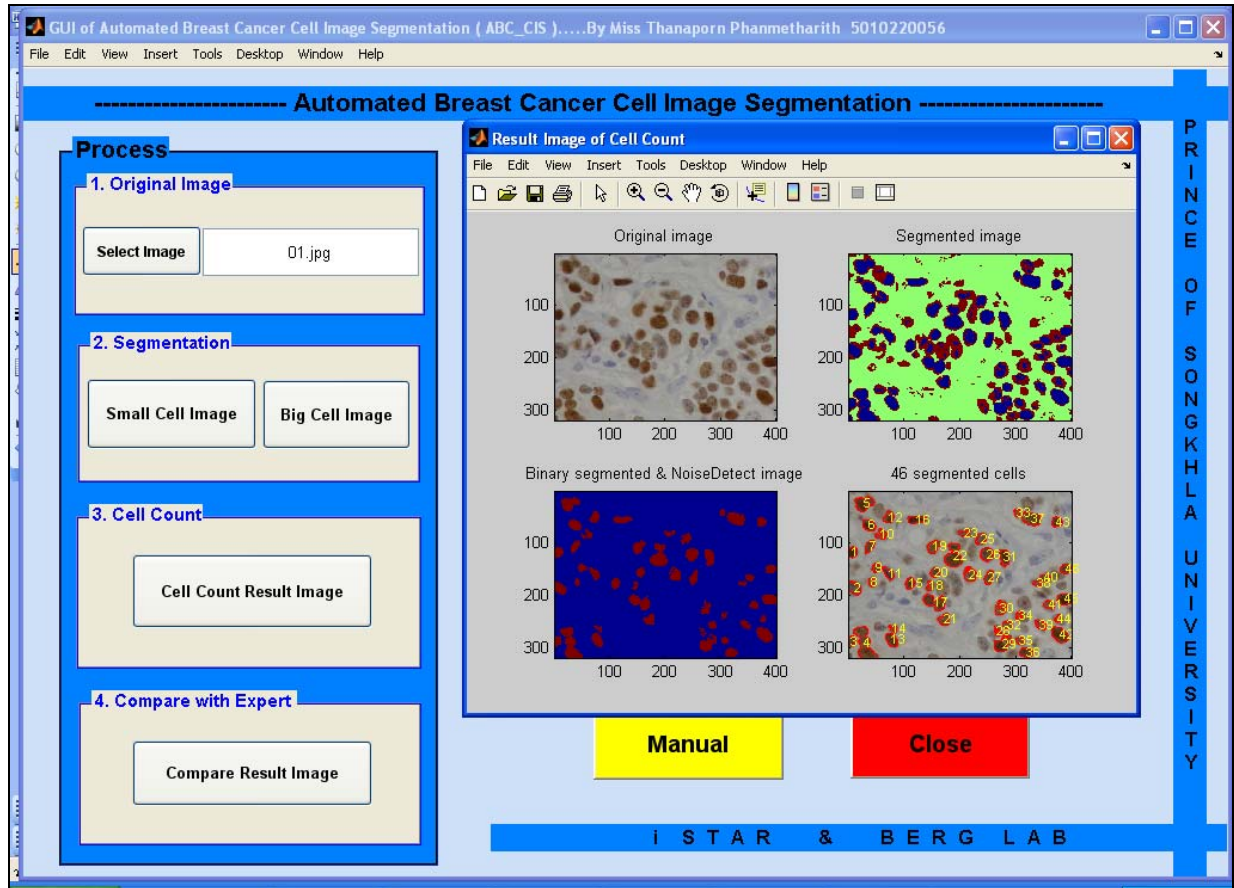
ภาพประกอบ ข.4 ภาพต้นฉบับถูกแสดงในส่วนแสดงภาพบน GUI

2) Segmentation เป็นส่วนที่ให้ผู้เลือกใช้วิธีในการแยกภาพเซลล์มะเร็งเริ่มต้นตามความเหมาะสมของภาพ ให้คลิกปุ่ม Small Cell Image เมื่อภาพที่นำมาทดลองเป็นภาพที่มีเซลล์ขนาดเล็กและไม่ชัดเจนเป็นส่วนใหญ่ หรือคลิกที่ปุ่ม Big Cell Image เมื่อภาพที่นำมาทดลองเป็นภาพที่มีเซลล์ขนาดใหญ่และชัดเจนเป็นส่วนใหญ่ แล้วรอสักครู่ จะปรากฏภาพผลการแยกภาพ ดังแสดงตัวอย่างในภาพประกอบ ข.5



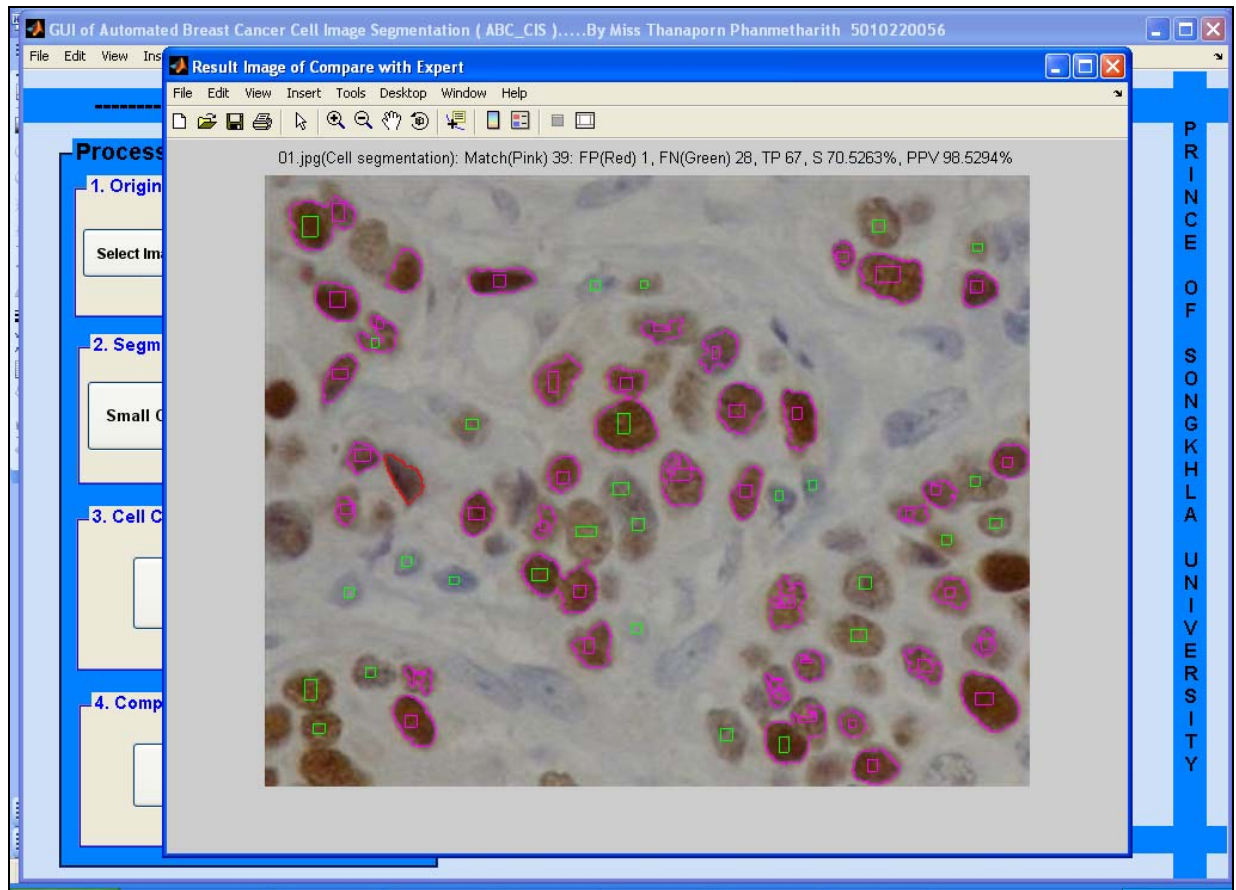
ภาพประกอบ ข.5 ตัวอย่างผลการแยกภาพบน GUI

3) Cell Count เป็นส่วนที่ใช้ในการนับเซลล์มะเร็งเริ่มต้นในภาพ เพื่อจะนำไปเปรียบเทียบกับผู้เชี่ยวชาญต่อไป เมื่อคลิกปุ่ม Cell Count Result Image จะปรากฏภาพผลการนับเซลล์พร้อมกับภาพต้นฉบับ ภาพผลการแยกเซลล์ และภาพหลังจากการกำจัด Noise ดังแสดงในภาพประกอบ ข.6



ภาพประกอบ ข.6 ตัวอย่างผลการนับเซลล์บน GUI

4) Compare with Expert เป็นส่วนที่จะนำผลการนับเซลล์จากขั้นตอนก่อนหน้าี้มาเปรียบเทียบกับผลการนับเซลล์จากผู้เชี่ยวชาญ เมื่อคลิกปุ่ม Compare Result Image จะปรากฏผลการเปรียบเทียบการนับเซลล์ โดยจะแสดงภาพและตัวเลขผลการเปรียบเทียบ เพื่อเป็นประโยชน์ต่อแพทย์ต่อไป ดังแสดงตัวอย่างในภาพประกอบ ข.7



ภาพประกอบ ข.7 ตัวอย่างผลการเปรียบเทียบกับผู้เชี่ยวชาญบน GUI

5) จากนั้นสามารถนำค่าผลการนับเซลล์และผลจากการเปรียบเทียบกับผู้เชี่ยวชาญไปใช้เป็นประโยชน์ต่อแพทย์ต่อไปได้ และในทุกๆขั้นตอนสามารถบันทึกผลภาพเก็บไว้ได้ เมื่อใช้โปรแกรมเรียบร้อยแล้ว ให้คลิกปุ่ม Close เพื่อปิดโปรแกรม

ภาคผนวก ก

ผลงานวิจัยที่ได้รับการตีพิมพ์ในการประชุมวิชาการ NCSEC 2008

เรื่อง	การเปรียบเทียบวิธีการแบ่งกลุ่มแบบไม่มีผู้สอนเพื่อแยกเซลล์มะเร็งเต้านมที่ถูกย้อมสีออกจากพื้นหลังของภาพ (A Comparison of Unsupervised Clustering for Nuclear Stained Breast Cancer Cell Image Segmentation)
งานประชุมวิชาการ	The 12 th National Computer Science and Engineering Conference (NCSEC 2008)
วันที่	20-21 พฤศจิกายน 2551
สถานที่	โรงแรมลองบีชการ์เด้น โฮเทล แอนด์ สปา พัทยา จังหวัดชลบุรี

การเปรียบเทียบวิธีการแบ่งกลุ่มแบบไม่มีผู้สอน
เพื่อแยกเซลล์มะเร็งเต้านมที่ถูกย้อมสีออกจากพื้นหลังของภาพ

A Comparison of Unsupervised Clustering for Nuclear Stained Breast Cancer

Cell Image Segmentation

ธนพร พันธุ์เมธาฤทธิ์^{*}, ลัดดา ปรีชาวีรกุล^{**} และ พรชัย พงษ์ภักถรานนท์^{***}

¹ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่

ผู้ ปณ. 3 ตำบลคลองหอยโข่ง อำเภอหาดใหญ่ จังหวัดสงขลา 90112 โทรศัพท์ : 0-7428-8581 โทรสาร : 0-7444-6917

²ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่

110/5 ถนน กาญจนวนิช ตำบลคลองหอยโข่ง อำเภอหาดใหญ่ จังหวัดสงขลา 90112 โทรศัพท์ : 0-7428-7045-6 โทรสาร : 0-7445-9395

Emails: {s5010220056^{*}, ladda.p^{**}, pornchai.p^{***}}@psu.ac.th

บทคัดย่อ

บทความนี้มีจุดประสงค์เพื่อเปรียบเทียบการแยกเซลล์มะเร็งเต้านมออกจากภาพเซลล์มะเร็งโดยใช้วิธีการแบ่งกลุ่มแบบไม่มีผู้สอน (Unsupervised Clustering) 2 วิธีคือ K-Means Clustering และ Farthest-First ซึ่งสามารถแยกส่วนของเซลล์มะเร็งเต้านมที่มีสีต่างจากพื้นหลังออกมาได้ ด้วยการใช้พื้นฐานของสี ซึ่งได้แก่ สีแดง สีเขียวและสีน้ำเงิน (RGB) โดยทำการแปลงภาพของเซลล์มะเร็งเต้านมให้เป็นตัวเลขของสี RGB จากนั้นผ่านกระบวนการ K-Means Clustering และ Farthest-First สุดท้ายจะพบว่าวิธีของ Farthest-First ให้ผลการแยกส่วนบริเวณพื้นหลัง Positive Cell และ Negative Cell ออกจากกันได้ดีกว่าการใช้วิธีของ K-Means Clustering ผลลัพธ์ที่ได้จะนำไปใช้ประโยชน์ในการกำหนดชนิดและปริมาณของการให้ยาสำหรับผู้ที่ป่วยเป็นมะเร็งเต้านมต่อไป

คำสำคัญ: เซลล์มะเร็งเต้านม, การแบ่งแยกภาพ, การเรียนรู้แบบไม่มีผู้สอน, วิธีการแบ่งกลุ่มแบบเคมีน

Abstract

This paper presents a comparison of nuclear stained breast cancer cell image segmentation using unsupervised clustering methods which are K-Means clustering and Farthest-First. These methods are able to segment different color of breast cancer cell by RGB-based color. The algorithm starts with transformation of breast cancer cell to the number of Red (R), Green (G), and Blue (B), then using K-Means clustering and Farthest-First. The experiment shows that Farthest-First is better than K-Means clustering for segmenting different color of breast cancer cell into image background, Positive cell, and Negative

cell. These outputs will use to assists a specialist to treat the dosage for breast cancer patients.

Keywords: Breast Cancer Cell, Image Segmentation, Immunohistochemistry, Unsupervised Learning, K-Mean Clustering, Farthest-First

1. บทนำ

ในปัจจุบันนี้โรคมะเร็งเป็นโรคที่มีผู้ป่วยจำนวนมาก นับวันยิ่งเพิ่มจำนวนผู้ป่วยมากยิ่งขึ้น เนื่องมาจากการรับประทานอาหาร ความเป็นอยู่ของคนที่สะดวกสบายมากขึ้น ทำให้พฤติกรรมกรกินของแต่ละคนเปลี่ยนไปหรือแม้กระทั่งเหตุผลที่หลีกเลี่ยงไม่ได้ก็คือ สาเหตุของโรคมะเร็งอาจจะมาจากพันธุกรรมของแต่ละคน การวิเคราะห์หามะเร็งในคนสามารถใช้การรวบรวมความรู้หลายด้านมาใช้ในการแก้ปัญหาได้

เซลล์มะเร็งเต้านม (Breast Cancer Cell) สามารถตรวจพบได้ด้วยการค้นหาแอนติเจนที่สามารถมองเห็นได้ด้วยการย้อมสีซึ่งเรียกเทคนิคในการหาแอนติเจนนี้ว่า Immunohistochemistry [1] เซลล์มะเร็งเต้านมเจริญเติบโตได้ดีจากฮอร์โมนเพศหญิงเอสโตรเจน ดังนั้นถ้าขาดขวางไม่ให้ฮอร์โมนเอสโตรเจนออกฤทธิ์ที่เซลล์มะเร็งได้หรือลดฮอร์โมนเอสโตรเจนให้มีจำนวนน้อยที่สุดก็จะช่วยยับยั้งการแพร่กระจายของมะเร็งเต้านมได้ [2] โดยเซลล์ที่มีการแพร่กระจายได้คือเซลล์ที่มี Estrogen Receptor เรียกว่า Positive Cell (P Cell) ส่วนเซลล์ที่ไม่มี Estrogen Receptor ทางการแพทย์จะเรียกว่า Negative Cell (N Cell)

ปัจจุบันเทคโนโลยีทางคอมพิวเตอร์สามารถประยุกต์ใช้ในการทำงานหลายสาขาวิชา มีบทบาทในงานทางการแพทย์อย่างมาก เช่น การใช้คอมพิวเตอร์ในการวิเคราะห์ค้นหาเซลล์มะเร็งหรือสิ่งมีชีวิตที่มีขนาดเล็กมากจนไม่สามารถมองเห็นด้วยตาเปล่าได้ ซึ่งทำให้

ประสิทธิภาพในการทำงานของการวิเคราะห์ผลลัพท์เป็นไปตามที่
ต้องการมากขึ้น มีการนำสาขาวิชาการประมวลผลภาพ (Image
Processing) มาประยุกต์ใช้ร่วมในทางการแพทย์ ที่สำคัญคือการนำการ
ประมวลผลภาพมาใช้งานการประมวลผลสัญญาณบนสัญญาณ 2 มิติ เช่น
ภาพนิ่ง (ภาพถ่าย) หรือภาพวีดิทัศน์ (วิดีโอ) และยังรวมถึงสัญญาณ 2 มิติ
อื่นๆ ที่ไม่ใช่ภาพด้วย ซึ่งการประมวลผลภาพจะมีเทคนิคและขั้นตอน
หลากหลายเพื่อใช้ให้เหมาะสมกับงาน วิธีหนึ่งของการประมวลผลภาพที่
นิยมนำมาใช้คือเทคนิคการแบ่งแยกวัตถุออกจากภาพโดยเทคนิคการ
แบ่งแยกวัตถุออกจากภาพสามารถทำได้หลายวิธีขึ้นอยู่กับพิจารณา
จากคุณลักษณะพื้นฐานของภาพเช่น การแบ่งแยกโดยพิจารณาจากเส้น
ขอบ สี หรือรูปร่าง

การแบ่งแยกโดยใช้สี (Color-Based Segmentation) เป็นการ
ใช้สีเป็นตัวบ่งบอกความแตกต่างของวัตถุที่เราสนใจ เพื่อแยกประเภท
วัตถุ ซึ่งเป็นประโยชน์มากต่อการแยกพื้นหลังออกจากเซลล์มะเร็งที่อยู่ใน
ภาพ เพื่อนำมาวิเคราะห์หาเซลล์มะเร็งที่ผ่านการย้อมสีแล้วได้

Phukpattaranont and Boonyaphiphat [4] ได้เสนอวิธีการ
แบ่งแยกเซลล์มะเร็งเรื้อรังด้านมอออกจากพื้นหลัง โดยใช้โครงข่ายประสาท
เทียม (Neural Network) [5] ซึ่งเป็นการเรียนรู้แบบมีผู้สอน (Supervised
Learning) [6] กล่าวคือให้โปรแกรมเรียนรู้ส่วนที่เป็นพื้นหลังและส่วนที่
เป็นเซลล์มะเร็งเรื้อรังก่อน ซึ่งจะทำให้การวิเคราะห์หาเซลล์ที่ผิดปกติ โดยใช้สีที่
เห็นจากการย้อมเป็นตัวที่ใช้ในการพิจารณาและทำการแยกเซลล์ที่ติดกัน
ออกจากกัน จากนั้นนำเฉพาะเซลล์ออกจากพื้นหลังก่อนทดสอบเพื่อให้
ได้ผลลัพธ์ว่าเซลล์ที่ต้องการรับย้อมมีจำนวนมากหรือน้อยเพียงใด อย่างไร
ก็ตามวิธีนี้ยังต้องใช้การเรียนรู้ให้ระบบรู้จักกับบริเวณที่เป็นพื้นหลัง, P
Cell และ N Cell เพื่อให้กระบวนการได้รู้ก่อน โดยต้องใช้ผู้เชี่ยวชาญ
ในการเลือกแต่ละบริเวณเอง ซึ่งมีความแตกต่างกันในแต่ละครั้งของการ
เลือก จึงเป็นการเรียนรู้แบบ Supervised Learning

Deng and Clausi [7] ได้นำเสนอการแบ่งแยกภาพโดยใช้
Markov Random Field (MRF) models ซึ่งเป็นวิธีใหม่ที่ใช้สำหรับ
Unsupervised Image Segmentation วิธีการนี้ค่อนข้างยุ่งยากและมีความ
ซับซ้อนในการใช้งาน เนื่องจากจะต้องมีการกำหนดค่าตัวแปรต่างๆ ที่ไม่
สามารถรู้ได้ว่าค่าไหนที่สมควรใช้และเหมาะสม ผลลัพท์ของแต่ละส่วน
ของวิธีนี้ได้ผลที่ไม่เหมือนกัน จึงเป็นการยากที่จะเลือกใช้ให้เหมาะสม
และวิธีการนี้ไม่เหมาะสำหรับนำมาใช้กับเซลล์มะเร็ง เนื่องจากวิธีนี้ใช้กับ
ภาพที่เป็นสี HSV ซึ่งภาพมะเร็งเรื้อรังด้านมอสามารถใช้สี RGB ก็เพียงพอต่อ
การทดลองแล้ว

Shih and Cheng [8] ได้เสนอวิธีการในการหาบริเวณโดยใช้
Seeded Region Growing (SRG) ในการแบ่งแยกภาพด้วยสี ซึ่งวิธีนี้จะ
แสดงผลลัพท์ของการแบ่งแยกกลุ่มโดยการเน้นขอบของแต่ละบริเวณที่
แบ่งกลุ่มได้ ทำให้เห็นวัตถุที่อยู่ในภาพแต่ละกลุ่มได้ไม่ชัดเจน ยกต่อ
การทำความเข้าใจ

Jeong, Kim และคณะ [9] ได้ทำการเปรียบเทียบเทคนิค
Thresholding ต่างๆและแสดงค่าความถูกต้องของแต่ละวิธี โดยเทคนิค
Thresholding ที่ใช้ในการแยกเซลล์ออกจากภาพคือการระบุพิกเซลใน
ภาพไปยังวัตถุหรือพื้นหลังอย่างใดอย่างหนึ่ง ซึ่งเป็นวิธีที่ง่ายและให้ผลดี
สำหรับการแบ่งแยกภาพ แต่มีข้อเสียคือ ไม่สามารถคำนึงถึงตำแหน่งและ
ลักษณะทางกายภาพของพิกเซลได้ จึงทำให้ไม่เหมาะสมกับเซลล์ที่มี
ลักษณะต่างกัน [4]

ในงานวิจัยนี้จึงได้คิดแนวทางในการแก้ปัญหาเพิ่มเติมโดยจัด
ให้ระบบไม่ต้องได้รับการเรียนรู้จากผู้สอน เพื่อไม่ให้เกิดความแตกต่าง
กันของการทดลองในแต่ละครั้งหรือเรียกว่าการเรียนรู้แบบไม่มีผู้สอน
(Unsupervised Learning) [10] และแสดงผลออกมาในรูปแบบที่สามารถ
มองเห็นความแตกต่างของแต่ละกลุ่มได้ชัดเจน โดยการนำเซลล์มะเร็งเรื้อรัง
ด้านมอที่ผ่านการย้อมสีแล้วมาใช้เทคนิคการแบ่งแยกโดยใช้สี (Color-Based
Segmentation) เพื่อแยกเซลล์มะเร็งเรื้อรังด้านมอออกจากพื้นหลังของภาพด้วย
วิธี K-Means Clustering และ Farthest-First โดยทำการเปรียบเทียบผลที่
ได้เพื่อนำมาวิเคราะห์หาเซลล์ที่ผิดปกติต่อไป

ส่วนที่สองของบทความได้กล่าวถึงทฤษฎีที่เกี่ยวข้องกับ
งานวิจัยนี้ ส่วนที่สามได้แสดงขั้นตอนวิธีการทดลองทั้งหมดและ
กระบวนการแยกเซลล์มะเร็งเรื้อรังด้านมอออกจากพื้นหลัง ส่วนที่สี่แสดงผล
การทดลองทั้งส่วนที่ใช้ K-Means Clustering และ Farthest-First รวมถึง
งานที่จะทำในอนาคต ส่วนที่ห้าเป็นส่วนสรุปและวิจารณ์ของบทความ
และสุดท้ายคือส่วนแสดงความขอบคุณแก่ผู้ให้ความช่วยเหลือในการ
ทำวิจัยนี้

2. ทฤษฎีที่เกี่ยวข้อง

2.1 การแบ่งแยกภาพ (Image Segmentation)

เป็นกระบวนการในการแยกกลุ่มของพิกเซลในภาพ ภาพจะ
ถูกแบ่งออกเป็นส่วนย่อยๆของพิกเซลในแต่ละกลุ่ม ซึ่งเป็นขั้นตอน
สำคัญในการค้นหาและจัดจำรูปแบบของการประมวลผลภาพ [11] และ
เป็นพื้นฐานของ Content-Based Image Retrieval (CBIR) [12] ซึ่ง Image
Segmentation [13] สามารถใช้สีในการแบ่งแยกวัตถุได้เรียกว่า Color
Segmentation ซึ่งมีทั้ง Segmentation in HSI Color Space, Segmentation
in RGB Vector Space and Color Edge Detection [11]

2.2 การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

การเรียนรู้แบบไม่มีผู้สอนเป็นเทคนิคหนึ่งของการเรียนรู้ของ
เครื่อง โดยการสร้างโมเดลที่เหมาะสมกับข้อมูล การเรียนรู้แบบนี้
แตกต่างจากการเรียนรู้แบบมีผู้สอน คือ จะไม่มีการระบุผลที่ต้องการหรือ
ประเภทไว้ก่อน การเรียนรู้แบบนี้จะพิจารณาวัตถุเป็นเซตของตัวแปรสุ่ม
แล้วจึงสร้างโมเดลความหนาแน่นร่วมของชุดข้อมูล [14]

การเรียนรู้แบบไม่มีผู้สอนจะประกอบไปด้วยเทคนิคดังต่อไปนี้ เช่น *K*-Means Clustering [15], Farthest-First, Expectation Maximization (EM), C-Mean, Cobweb และ DBScan เป็นต้น ในงานวิจัยนี้ได้ทำการเลือก 2 เทคนิคเพื่อเปรียบเทียบผลการแยกเซลล์มะเร็งด้านมออกจากพื้นหลังของภาพ ได้แก่ *K*-Means Clustering และ Farthest-First

ก) *K*-Means Clustering

เป็นวิธีการแบ่งกลุ่มแบบไม่มีผู้สอนที่มีขั้นตอนหลักคือทำการแบ่งกลุ่มซ้ำๆจนกว่าจะแบ่งไม่ได้แล้ว ซึ่งเป็นวิธีที่สามารถเข้าใจได้ง่าย ประกอบไปด้วยขั้นตอนต่อไปนี้

- 1) เลือกจุดศูนย์กลางของกลุ่ม *K* กลุ่ม โดยที่จำนวนกลุ่ม *K* นี้จะถูกกำหนดไว้แล้ว
- 2) แต่ละตัวอย่างจะถูกกำหนดให้ไปอยู่ในกลุ่มที่มีจุดศูนย์กลางอยู่ใกล้ตัวอย่างนั้นมากที่สุด
- 3) จากนั้นตัวอย่างทั้งหมดจะนำมาคำนวณหาจุดศูนย์กลางใหม่
- 4) ทำกระบวนการนี้ซ้ำจนกระทั่งจำนวนกลุ่มไม่มีการเปลี่ยนแปลง

K-Means Algorithm เป็นส่วนหนึ่งในการแบ่งกลุ่ม สามารถดูตัวอย่างเพิ่มเติมได้ใน [16]

ข) Farthest-First

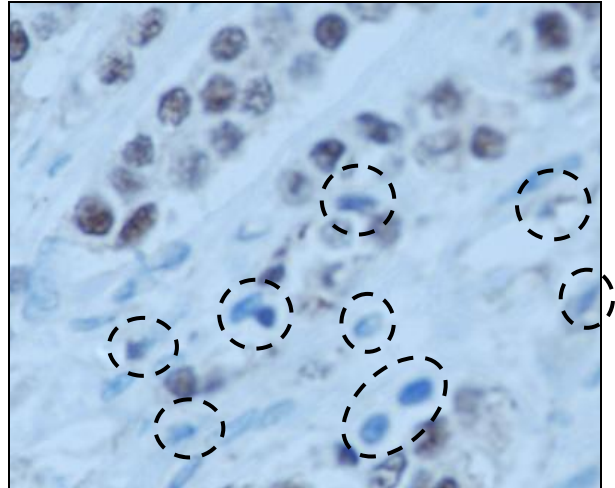
เป็นวิธีการแบ่งกลุ่มวิธีหนึ่งที่มีขึ้นหลัง *K*-Mean Clustering ขั้นตอนวิธี Farthest-First ใช้ในการแบ่งกลุ่มแบบไม่มีผู้สอน ซึ่งใช้เทคนิคของ Farthest-First Traversal Algorithm [17] ที่มีขั้นตอนวิธีดังนี้

- 1) เลือกตัวอย่าง 1 ตัวอย่างแบบสุ่มมาเป็นจุดศูนย์กลางของกลุ่ม
- 2) คำนวณระยะห่างระหว่างตัวอย่างที่เหลืออยู่กับจุดศูนย์กลางที่อยู่ใกล้ที่สุด
- 3) เลือกจุดศูนย์กลางใหม่ โดยให้ตัวอย่างที่อยู่ไกลที่สุดจากจุดศูนย์กลางเดิมมาเป็นจุดศูนย์กลางใหม่
- 4) ทำกระบวนการนี้ซ้ำจนกระทั่งจำนวนกลุ่มมากกว่าจำนวนกลุ่มที่กำหนดไว้

3. การทดลอง

การทดลองนี้นำภาพของเซลล์มะเร็งเรื้อรังด้านมที่ได้ผ่านการข้อมเป็นสีและขยายขนาดขึ้น 400 เท่า โดยใช้กล้องจุลทรรศน์ Eclipse 80i สำหรับงานวิจัย โดยภาพดิจิทัลที่ได้เป็นไฟล์นามสกุล JPEG ซึ่งเป็นภาพสีขนาด 400x320 พิกเซล ดังแสดงในรูปที่ 1

ในการทดลองนี้ได้ใช้เครื่องคอมพิวเตอร์ส่วนบุคคลที่ประกอบด้วย RAM 512 MB, HD 80 GB, CPU Pentium(R) 4 3.0 GHz และใช้โปรแกรม MATLAB 7.5 ร่วมกับการเขียนโปรแกรมด้วยภาษาซีในการแปลงภาพให้อยู่ในรูปแบบที่ต้องการ ส่วนกระบวนการ



รูปที่ 1 เซลล์มะเร็งเรื้อรังด้านมที่ประกอบด้วย P Cell (เซลล์ที่ไม่ถูกล้อมด้วยเส้นประ), N Cell (เซลล์ที่ล้อมรอบด้วยเส้นประ) และบริเวณพื้นหลัง

Transform image to number (TIN) Algorithm.
ข้อมูลนำเข้า : ภาพเซลล์มะเร็งเรื้อรังด้านมที่ผ่านการข้อมสี
ผลลัพธ์ : ไฟล์ .arff สำหรับใช้ใน WEKA (ประกอบด้วย 3 แอททริบิวต์คือ R(แดง), G(เขียว) และ B(น้ำเงิน))
<ol style="list-style-type: none"> 1) อ่าน ไฟล์ภาพมะเร็งเรื้อรังด้านมที่ผ่านการข้อมสีแล้วเข้ามา 2) คำนวณหาค่าสี RGB ของแต่ละพิกเซลทั้งหมดในภาพ 3) จัดเรียงตัวเลขของค่าสี RGB ให้เป็นรูปแบบของไฟล์ .arff

รูปที่ 2 ขั้นตอนวิธีแปลงภาพเป็นตัวเลขของ RGB

Transform cluster number to image (TCI) Algorithm.
ข้อมูลนำเข้า : ไฟล์ .arff (ประกอบด้วย 5 แอททริบิวต์คือ ลำดับของพิกเซล, R(แดง), G(เขียว), B(น้ำเงิน) และหมายเลขกลุ่ม)
ผลลัพธ์ : ภาพเซลล์มะเร็งเรื้อรังด้านมที่ผ่านการแยกส่วนของเซลล์ออกจากพื้นหลังในภาพแล้ว
<ol style="list-style-type: none"> 1) เก็บหมายเลขกลุ่มที่ได้จากกระบวนการของ Unsupervised Learning จากขั้นตอนที่ 2.1 ในรูปที่ 4 2) จัดเรียงตัวเลขของหมายเลขกลุ่มลงไฟล์ข้อความตามขนาดของภาพ 3) แสดงภาพผลลัพธ์ที่ได้

รูปที่ 3 ขั้นตอนวิธีแปลงตัวเลขของหมายเลขกลุ่มที่ได้จากวิธีการแบ่งกลุ่มแบบ *K*-Means หรือ Farthest-First ให้เป็นภาพ



รูปที่ 4 กระบวนการแยกเซลล์มะเร็งเต้านมออกจากพื้นหลัง

K-Means Clustering และ Farthest-First ดำเนินการโดยใช้โปรแกรม WEKA 3.5 โดยเริ่มจากการเตรียมข้อมูลเพื่อนำภาพเซลล์มะเร็งเต้านมที่ผ่านการย้อมสีมาผ่านกระบวนการแปลงภาพโดยใช้ TIN Algorithm ดังรูปที่ 2 จากนั้นทำการแปลงผลภาพโดยใช้ TCI Algorithm ดังรูปที่ 3 และวิเคราะห์ผลการแยก P Cell และ N Cell ดังอธิบายขั้นตอนวิธีการทำงานทั้งหมดด้วยรูปที่ 4 ในการทดลองได้ทดลองกับภาพเซลล์มะเร็งเต้านมที่มีความยากง่ายแตกต่างกัน 20 ภาพ ซึ่งภาพตัวอย่างของผลการแยกเซลล์มะเร็งเต้านมแสดงอยู่ในรูปที่ 5

4. ผลการทดลอง

ผลการทดลองการแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังของภาพโดยใช้ K-Means Clustering และ Farthest-First มีดังนี้

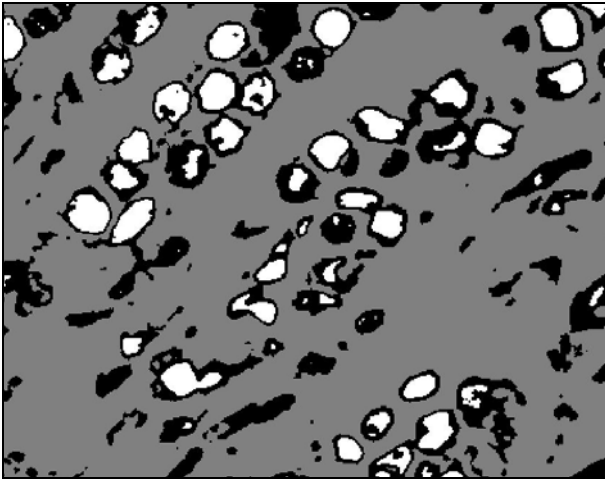
4.1 K-Means Clustering

ผลการแยกเซลล์มะเร็งออกจากพื้นหลังโดยใช้ K-Means Clustering ได้แก่การทดลองที่ 1 ถึง 3 โดยการทดลองที่ 1 จะกำหนดจำนวนกลุ่มเป็น 3 ผลการทดลองสามารถแยกเซลล์ทุกเซลล์ออกจากพื้นหลังได้ โดยทุกเซลล์จะถูกกำหนดให้เป็นกลุ่มเดียวกันหมดทั้ง P Cell และ N Cell สามารถแยกได้เป็น 3 บริเวณได้แก่ บริเวณนิวเคลียสของเซลล์ บริเวณรอบเซลล์และบริเวณที่เป็นพื้นหลังของภาพ ดังแสดงในรูปที่ 5(ก) ในการทดลองที่ 2 จะกำหนดให้จำนวนกลุ่มเป็น 4 ผลการทดลองสามารถแยกกลุ่มของสีออกเป็น 4 บริเวณได้แก่ บริเวณนิวเคลียสของเซลล์ บริเวณรอบนอกของเซลล์ บริเวณพื้นหลังที่มีสีเข้มและบริเวณพื้นหลังที่มีสีอ่อน ดังแสดงในรูปที่ 5(ข) และการทดลองที่ 3 จะกำหนดให้จำนวนกลุ่มเป็น 5 ผลการทดลองสามารถแบ่งแยกบริเวณได้ 5 บริเวณได้แก่ บริเวณนิวเคลียสของเซลล์ บริเวณรอบนิวเคลียสจะแบ่งเป็นอีก 2 กลุ่ม บริเวณพื้นหลังก็แบ่งได้อีก 2 กลุ่มตามความแตกต่างของสี ดังแสดงในรูปที่ 5(ค)

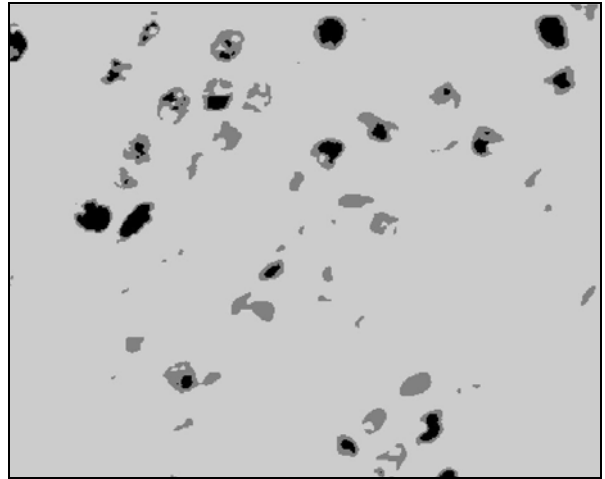
4.2 Farthest-First

ผลการทดลองโดยใช้ Farthest-First ได้แก่การทดลองที่ 4 ถึง 6 โดยที่การทดลองที่ 4 จะกำหนดจำนวนกลุ่มเท่ากับ 3 ผลการทดลองสามารถแยกเซลล์ออกจากพื้นหลังได้และสามารถแยกเซลล์ที่มีสีต่างกันออกจากกันได้ในบางส่วนของภาพ แบ่งเป็น 3 บริเวณคือ บริเวณที่เป็นสีดำ (ถ้าเปรียบเทียบกับรูปที่ 1 จะเห็นเป็นสีเทาเข้ม) ส่วนบริเวณที่ 2 คือ บริเวณเซลล์ที่เป็นสีเทาและบริเวณพื้นหลังเป็นบริเวณสุดท้ายที่แบ่งได้ จะเห็นในรูปเป็นสีเทาอ่อน ดังแสดงในรูปที่ 5(ง) ในการทดลองที่ 5 จะกำหนดจำนวนกลุ่มเท่ากับ 4 ผลการทดลองสามารถแบ่งกลุ่มภาพได้เป็น 4 บริเวณคือ บริเวณ P Cell ซึ่งแบ่งได้เป็น 2 กลุ่มคือกลุ่มที่มีสีเทาเข้มและที่มีสีดำ ส่วนที่ 3 คือส่วนที่เป็น N Cell จะถูกแทนด้วยกลุ่มของสีเทาอ่อนซึ่งเป็นเซลล์ที่ล้อมรอบด้วยเส้นประ ส่วนที่ 4 คือส่วนที่เป็นพื้นหลังทั้งหมด ดังแสดงในรูปที่ 5(จ) และการทดลองที่ 6 จะกำหนดจำนวนกลุ่มเท่ากับ 5 ผลการทดลองคือ สามารถแบ่งได้เป็น 5 กลุ่มได้แก่ บริเวณนิวเคลียสของเซลล์จะแบ่งได้เป็น 2 กลุ่มคือกลุ่มที่มีสีเทาอ่อนและที่มีสีเทา ส่วนที่ 3 คือส่วนที่เป็น N Cell จะถูกแทนด้วยกลุ่มของสีดำโดยเป็นเซลล์ที่ล้อมรอบด้วยเส้นประ ส่วนที่ 4 และ 5 คือส่วนที่เป็นพื้นหลังที่มีสีแตกต่างกันแบ่งได้เป็น 2 กลุ่ม ดังแสดงในรูปที่ 5(ฉ)

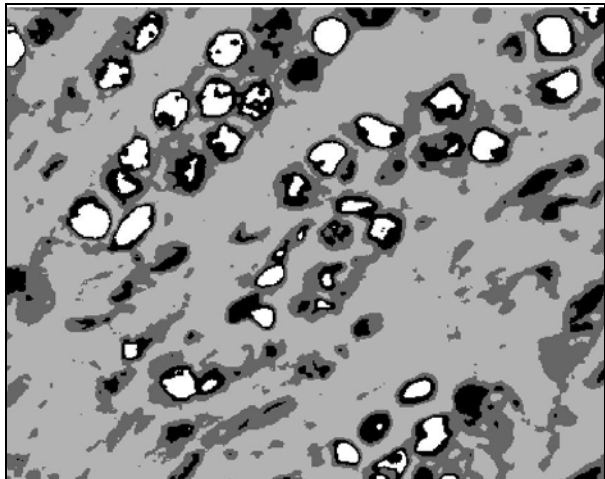
จากผลการทดลองทั้งหมดจึงสรุปได้ว่าการใช้วิธี Farthest-First สามารถแยกเซลล์มะเร็งเต้านมที่ถูกย้อมสีออกจากภาพได้ดีกว่าวิธี K-Means Clustering โดยที่ Farthest-First จะสามารถแยก P Cell และ N Cell ออกจากกันได้ด้วย



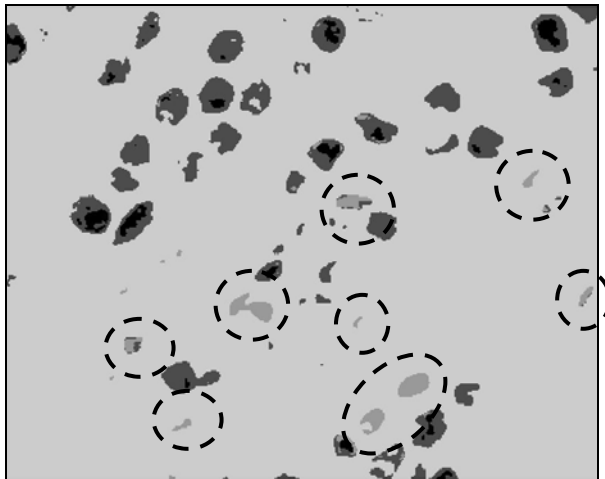
(ก) การทดลองที่ 1 ใช้วิธี K-Means Clustering จำนวนกลุ่ม=3



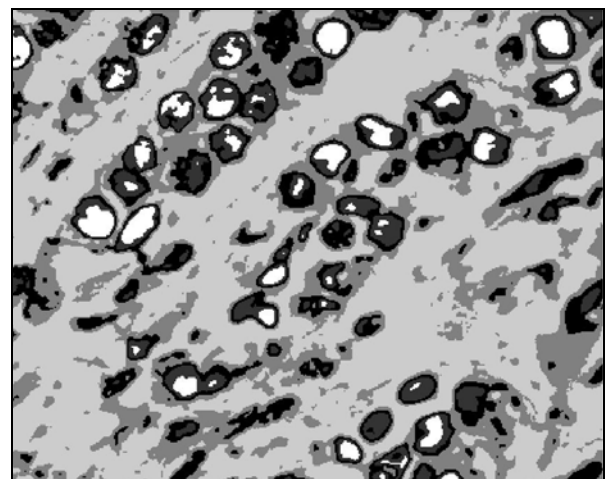
(ง) การทดลองที่ 4 ใช้วิธี Farthest-First จำนวนกลุ่ม=3



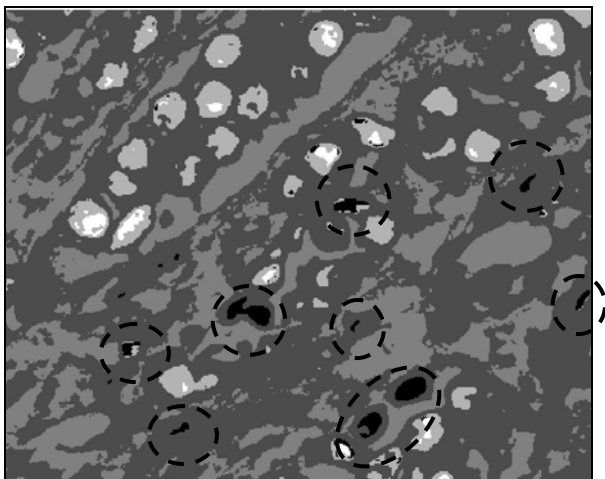
(ข) การทดลองที่ 2 ใช้วิธี K-Means Clustering จำนวนกลุ่ม=4



(จ) การทดลองที่ 5 ใช้วิธี Farthest-First จำนวนกลุ่ม=4

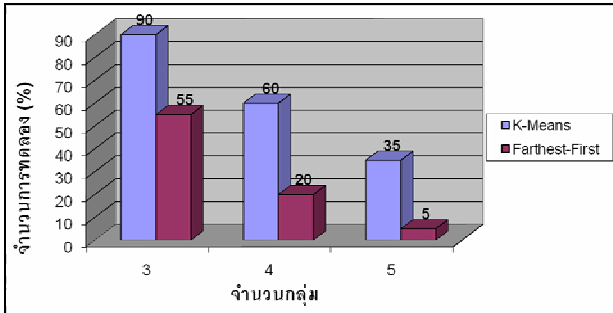


(ค) การทดลองที่ 3 ใช้วิธี K-Means Clustering จำนวนกลุ่ม=5

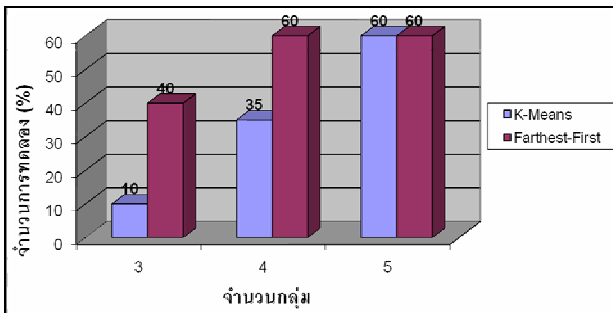


(ฉ) การทดลองที่ 6 ใช้วิธี Farthest-First จำนวนกลุ่ม=5

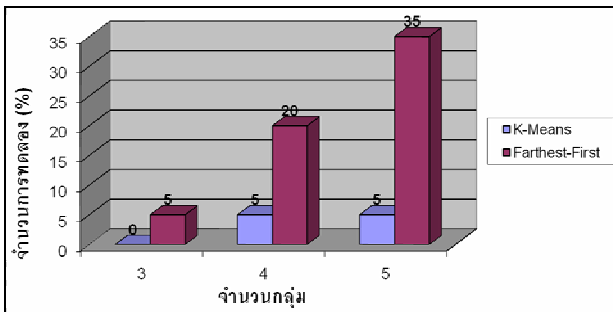
รูปที่ 5 ภาพที่ได้หลังจากผ่านกระบวนการโดยใช้ K-Means Clustering และ Farthest-First



รูปที่ 6 กราฟเปรียบเทียบผลการทดลองในกรณีที่ไม่สามารถแยก P Cell ออกจาก N Cell ได้



รูปที่ 7 กราฟเปรียบเทียบผลการทดลองในกรณีที่สามารถแยก P Cell ออกจาก N Cell ได้บางส่วน



รูปที่ 8 กราฟเปรียบเทียบผลการทดลองในกรณีที่สามารถแยก P Cell ออกจาก N Cell ได้ชัดเจน

ผลการเปรียบเทียบการแยก P Cell ออกจาก N Cell โดยใช้วิธี K-Means และ Farthest-First แสดงในรูปที่ 6, 7 และ 8 โดยที่แกนตั้งคือเปอร์เซ็นต์ของจำนวนการทดลอง แกนนอนคือจำนวนกลุ่มในการแบ่งแยก ผลการเปรียบเทียบทั้งหมดแบ่งเป็น 3 กรณี ได้แก่ กรณีที่ไม่สามารถแยก P Cell ออกจาก N Cell ได้ แสดงดังรูปที่ 6 จะเห็นได้ว่าจำนวนการทดลองโดยใช้ K-Means จะมีเปอร์เซ็นต์ที่สูงกว่าการใช้ Farthest-First กรณีที่สองคือ กรณีที่สามารถแยก P Cell ออกจาก N Cell ได้ในบางส่วน ดังแสดงในรูปที่ 7 จะแสดงให้เห็นว่าการใช้วิธี Farthest-First โดยกำหนดจำนวนกลุ่มเป็น 3, 4 มีเปอร์เซ็นต์ของจำนวนการ

ทดลองในการแยกได้บางส่วนสูงกว่าการใช้วิธี K-Means และกรณีสุดท้ายคือกรณีที่สามการแยก P Cell ออกจาก N Cell ได้อย่างชัดเจน แสดงดังรูปที่ 8 จะเห็นว่าจำนวนการทดลองโดยใช้วิธี Farthest-First มีเปอร์เซ็นต์ที่สูงกว่าการใช้ K-Means จากทุกจำนวนกลุ่มที่ใช้ในการทดลอง ซึ่งหมายความว่าวิธี Farthest-First สามารถแยก P Cell ออกจาก N Cell ได้ดีกว่า

ในการแยกเซลล์มะเร็งออกจากภาพในตอนนี้ยังต้องมีผู้เชี่ยวชาญช่วยสังเกตในบางขั้นตอน แต่ในความเป็นจริงแล้วถ้ามีวิธีการทำโดยที่ไม่ต้องอาศัยผู้เชี่ยวชาญจะช่วยให้การแยกเซลล์เป็นไปได้สะดวกยิ่งขึ้น และเนื่องจากแต่ละคนก็มีความแตกต่างกันในการเลือกบริเวณที่ต้องการให้ระบบเรียนรู้ ทำให้ผลลัพธ์ในแต่ละครั้งก็จะแตกต่างกันได้ ดังนั้นการแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังของภาพและการแยก P Cell ออกจาก N Cell แบบอัตโนมัติจะเป็นงานที่จะทำในอนาคต ซึ่งจะช่วยแพทย์สามารถแยกเซลล์ได้สะดวกและรวดเร็วขึ้น

5. สรุปและวิจารณ์

บทความนี้ได้เปรียบเทียบวิธีการแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังของภาพโดยใช้เทคนิคการเรียนรู้แบบไม่มีผู้สอนได้แก่ K-Means Clustering และ Farthest-First ในการทดลองได้ทดสอบกับภาพเซลล์มะเร็งเต้านมที่มีความยากแตกต่างกัน ซึ่งสังเกตได้ว่าผลการแบ่งแยกโดยใช้วิธี K-Means Clustering สามารถแยกเซลล์มะเร็งออกจากพื้นหลังได้แต่ส่วนใหญ่แล้วยังไม่สามารถแยกเซลล์ที่เป็นต้นเหตุของการกระจายเซลล์มะเร็ง (P Cell) ออกจากเซลล์อีกประเภท (N Cell) ที่มีสีต่างกันได้ในขณะที่ Farthest-First สามารถแยกเซลล์มะเร็งออกจากพื้นหลังได้และสามารถแยก P Cell ออกจาก N Cell ได้ผลค่อนข้างดี นอกจากนี้การแยกเซลล์มะเร็งเต้านมออกจากพื้นหลังโดยใช้ Farthest-First จะได้ผลดีหรือไม่ขึ้นอยู่กับความยากง่ายของภาพเซลล์มะเร็งเต้านมที่ถูกย้อมสีแต่ละภาพ อย่างไรก็ตามวิธีการนี้ไม่เหมาะกับภาพเซลล์มะเร็งเต้านมที่มีบริเวณพื้นหลังและส่วนของเซลล์มะเร็งไม่ชัดเจนหรือมีสีที่ไม่สม่ำเสมอ

6. กิตติกรรมประกาศ

งานวิจัยนี้ได้รับความช่วยเหลือจากผศ.ปัทมาจิต บุญยพิพัฒน์ ภาควิชาพยาธิวิทยา คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่ ผู้อนุเคราะห์ข้อมูลภาพเซลล์มะเร็งเต้านมที่ผ่านการย้อมสีแล้ว เพื่อนำมาใช้ในงานวิจัยนี้

เอกสารอ้างอิง

- [1] R. Mofidi R. Walsh P. F. Ridgway etc., "Objective measurement of breast cancer oestrogen receptor status through digital image analysis," *EJSO*, vol. 29, pp. 20-24, 2003.

[2] T. Pongpram and P. Phukpattaranont, "Color-Based Segmentation of Breast Cancer Cell Using K-Mean Clustering," *The Proceeding of PSU-Engineering Conference*, 2008.

[3] R. Demirci, "Rule-based automatic segmentation of color images," *International Journal AEU of Electronics and Communications*, vol. 60, pp. 435-442, 2006.

[4] P. Phukpattaranont. and P. Boonyaphiphat, "Color based segmentation of nuclear stained breast cancer cell images," *ECTI Transaction on Electrical Eng. Electronics and Communications*, vol. 5, no 2, pp. 158-164, 2007.

[5] D. Michie, D.J. Spiegelhalter and C.C. Taylor, "Machine Learning, Neural and Statistical Classification," 1994.

[6] R. J. Roiger and M.W. Geatz, "Data Mining. A Tutorial-Based Primer. International Edition," *Addison Wesley*, 2003.

[7] H. Deng and D. A. Clausi, "Unsupervised image segmentation using a simpleMRFmodel with a newimplementation scheme," *Pattern Recognition*, vol. 37, pp. 2323-2335, 2004.

[8] F. Y. Shih and S. Cheng, "Automatic seeded region growing for color image segmentation," *Image and Vision Computing*, vol. 23, pp. 877-886, 2005.

[9] H. Jeong, T. Kim, H. Hwang, H. Choi, H. Park, H. Cho, "Comparison of thresholding methods for breast tumor cell segmentation," *Enterprise networking and Computing in Healthcare Industry, 2005. HEALTHCOM 2005. Proceedings of 7th International Workshop*, pp. 392-395, 2005.

[10] Z. Ghahramani, "Unsupervised Learning," Gatsby Computational Neuroscience Unit, 2004.

[11] R. C. Gonzalez and R. E. Woogs, "Digital Image Processing. Second Edition," *Prentice Hall*, 2002.

[12] S. Mitra and T. Acharya, "Data Mining Multimedia, Soft Computing and Bioinformatics," *Wiley-Interscience*, 2003.

[13] M. Ozden and E. Polat, "Acolor image segmentation approach for content-based image retrieval," *Pattern Recognition*, vol. 40, pp. 1318-1325, 2007.

[14] Unsupervised learning (Online). Available <http://en.wikipedia.org/wiki/Unsupervisedlearning> (1 May 2008).

[15] S. Kantabutra and A. L. Couch, "Parallel K-means Clustering Algorithm on NOWs," *NECTEC Technical Journal*, vol. 1, no. 6, pp. 243-248, 2000.

[16] I. H. Witten and E. Frank, "Data Mining Practical Machine Learning Tool and Techniques," *Morgan Kaufmann Publishers*, 2005.

[17] S. Dasgupta and P. M. Long, "Performance guarantees for hierarchical clustering," *Journal of Computer and System Sciences*, vol. 70, pp. 555-569, 2005.

ประวัติผู้เขียนบทความ



ชนพร พันธุ์เมธาฤทธิ์ นักศึกษาปริญญาโท
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่
งานวิจัยที่สนใจได้แก่ Image Processing



ลัดดา ปรีชาวีรกุล
ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่
งานวิจัยที่สนใจได้แก่ Internet Computing, Object-
Oriented Technology, Image Processing



พรชัย พงกษัฏธานนค์
ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์
มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่
งานวิจัยที่สนใจได้แก่ Medical Image Processing,
Ultrasound Medical Imaging, Biomedical Signal
Process

ประวัติผู้เขียน

ชื่อ – สกุล นางสาวชนพร พันธุ์เมธาสุทธิ

รหัสประจำตัวนักศึกษา 5010220056

วุฒิการศึกษา

วุฒิ	ชื่อสถาบัน	ปีที่สำเร็จการศึกษา
วท.บ. (วิทยาการคอมพิวเตอร์)	มหาวิทยาลัยทักษิณ	2550

ทุนการศึกษาที่ได้รับในระหว่างการศึกษา

ทุนผู้ช่วยสอน ภาคการศึกษาที่ 1 และ 2 ปีการศึกษา 2551 ภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยสงขลานครินทร์

การตีพิมพ์เผยแพร่ผลงาน

Phanmetharith, T., Preechaveerakul, L., and Phukpattranont, P. 2008. A Comparison of Unsupervised Clustering for Nuclear Stained Breast Cancer Cell Image Segmentation. The 12th National Computer Science and Engineering Conference (NCSEC 2008). November 20-21 2008. pp. 286-292.