

การแบ่งส่วนเซลล์เม็ดเลือดขาวและการเลือกลักษณะป่องต่างสำหรับ
การจำแนกประเภทของเม็ดเลือดขาว

นางสาวจรรยารัตน์ ปริญญาคุปต์



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)

เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)

สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
are the thesis authors' files submitted through the University Graduate School.

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2557

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

SEGMENTATION OF WHITE BLOOD CELLS AND DISCRIMINATIVE FEATURES FOR
LEUKOCYTE CLASSIFICATION

Miss Jaroonrut Prinyakupt



A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Electrical Engineering

Department of Electrical Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2014

Copyright of Chulalongkorn University

| | |
|---------------------------------|--|
| หัวข้อวิทยานิพนธ์ | การแบ่งส่วนเซลล์เม็ดเลือดขาวและการเลือกลักษณะบ่ง |
| | ต่างสำหรับ การจำแนกประเภทของเม็ดเลือดขาว |
| โดย | นางสาวจรรยารัตน์ ปริญญาคุปต์ |
| สาขาวิชา | วิศวกรรมไฟฟ้า |
| อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก | ผู้ช่วยศาสตราจารย์ ดร. ชาญชัย ปลื้มปิติวิริยะเวช |

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญาตรีบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร. บัณฑิต เอื้ออาภรณ์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. วิทยากร อัครวิเศษ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร. ชาญชัย ปลื้มปิติวิริยะเวช)

.....กรรมการ
(รองศาสตราจารย์ ดร. นิตาชล ตั้งเสงี่ยมวิสัย)

.....กรรมการ
(รองศาสตราจารย์ นงลักษณ์ โควาวิสารัช)

.....กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร. วนิดา พงศ์สถาพร)

จรรยารัตน์ ปริญาคุปต์ : การแบ่งส่วนเซลล์เม็ดเลือดขาวและการเลือกลักษณะบ่งต่างสำหรับการจำแนกประเภทของเม็ดเลือดขาว (SEGMENTATION OF WHITE BLOOD CELLS AND DISCRIMINATIVE FEATURES FOR LEUKOCYTE CLASSIFICATION) อ.ที่ปรึกษาวิทยานิพนธ์
หลัก: ผศ. ดร. ชาญชัย ปลื้มปิติวิริยะเวช, 118 หน้า.

ในการวินิจฉัยโรคเลือด นักโลหิตวิทยาจะใช้การดูสมียร์เลือดจากกล้องจุลทรรศน์เพื่อใช้ในการตรวจสอบเป็นงานที่ทำในแต่ละวันซึ่งเป็นงานที่ค่อนข้างน่าเบื่อและใช้เวลานาน การตรวจหาเซลล์เม็ดเลือดขาวที่อยู่ในภาพดังกล่าวโดยระบบอัตโนมัติจึงเป็นสิ่งที่น่าสนใจ วิทยานิพนธ์เรื่องนี้นำเสนอระบบที่ใช้เพื่อหาเซลล์เม็ดเลือดขาวที่อยู่ในภาพสมียร์เลือดจากกล้องจุลทรรศน์ และแยกส่วนเซลล์เม็ดเลือดขาวออกเป็นนิวเคลียสและไซโทพลาซึมและทำการสกัดหาลักษณะบ่งต่างที่เหมาะสมเพื่อใช้ในการจัดกลุ่มเซลล์เม็ดเลือดขาวออกเป็นห้าประเภท basophil, eosinophil, neutrophil, lymphocyte และ monocyte ในการศึกษาครั้งนี้ใช้ภาพสมียร์เลือดสองชุด โดยชุดข้อมูลที่ 1 เป็นชุดของภาพที่เก็บรวบรวมจากมหาวิทยาลัยรังสิตเป็นสไลด์เลือดของคนปกติถ่ายด้วยกล้องจุลทรรศน์ที่มีกำลังขยาย 100 เท่า บันทึกภาพโดยกล้อง Nikon DSFi2 ความละเอียดสูง ถ่ายภาพและบันทึกในรูปแบบ JPG ขนาด 960 × 1,280 พิกเซล มีความละเอียด 15 พิกเซลต่อ 1 ไมโครเมตร จำนวนภาพในชุดข้อมูลที่ 1 มี 828 ภาพ ประกอบด้วยเซลล์เม็ดเลือดขาว 879 เซลล์ ในชุดข้อมูลที่ 2 มีจำนวน 477 ภาพ เป็นภาพเม็ดเลือดขาวที่ผ่านการตัดมาแล้วโดยดาวนโหลดจาก CellaVision.com ภาพในชุดข้อมูลนี้อยู่ในรูปแบบ JPG ขนาด 360 × 363 พิกเซล ความละเอียดประมาณ 10 พิกเซลต่อ 1 ไมโครเมตร วิทยานิพนธ์นี้ได้ดำเนินการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวและจัดกลุ่มเซลล์เม็ดเลือดขาว ผลการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวออกเป็นนิวเคลียสและไซโทพลาซึม โดยระบบที่ออกแบบเทียบกับนักโลหิตวิทยาพบว่าวิธีการที่นำเสนอมีความสอดคล้องและเชื่อมโยงกัน ในชุดข้อมูลทั้งสอง โดยมีความคล้ายคลึงกันมากกว่า 87% สำหรับการแบ่งส่วนทั้งนิวเคลียสและไซโทพลาซึมทำการเลือกและสกัดลักษณะบ่งต่าง โดยเลือกใช้ 13 ลักษณะในการจำแนกเซลล์ ความถูกต้องของการจำแนกกลุ่มเซลล์เม็ดเลือดขาวสูงกว่า 95% สำหรับทั้งวิธี linear discriminant function และ quadratic discriminant function ระบบที่นำเสนอใช้หลักการบนพื้นฐานของรูปร่าง โครงสร้างและลักษณะบ่งต่างของเซลล์เม็ดเลือดขาวแบบปรกติ และสามารถนำไปใช้กับชุดข้อมูลสองชุดที่แตกต่างกันได้ ผลของการแบ่งส่วนและการสอบเทียบในชุดข้อมูลทั้งสองมีความรวดเร็วและมีประสิทธิภาพและสอดคล้องกัน ในขณะเดียวกันการจำแนกกลุ่มเซลล์เม็ดเลือดขาวแบบปกติเป็นห้าประเภทแสดงให้เห็นถึง sensitivity สูง

ภาควิชา วิศวกรรมไฟฟ้า

ลายมือชื่อนิสิต

สาขาวิชา วิศวกรรมไฟฟ้า

ลายมือชื่อ อ.ที่ปรึกษาหลัก

ปีการศึกษา 2557

5271844621 : MAJOR ELECTRICAL ENGINEERING

KEYWORDS: WHITE BLOOD CELL / IMAGE SEGMENTATION / ELLIPSE CURVE FITTING / FEATURE EXTRACTION / CLASSIFICATION

JAROONRUT PRINYAKUPT: SEGMENTATION OF WHITE BLOOD CELLS AND DISCRIMINATIVE FEATURES FOR LEUKOCYTE CLASSIFICATION. ADVISOR: ASST. PROF. CHARNCHAI PLUEMPITIWIRIYAJEJ, 118 pp.

Blood smear microscopic images are routinely investigated by haematologists to diagnose most blood diseases. However, the task is quite tedious and time consuming. An automatic detection and classification of white blood cells within such images can accelerate the process tremendously. This dissertation we propose a system to locate white blood cells within microscopic blood smear images, segment them into nucleus and cytoplasm regions, extract suitable features and finally, classify them into five types: basophil, eosinophil, neutrophil, lymphocyte and monocyte. Two sets of blood smear images were used in this study's experiments. Dataset 1, collected from Rangsit University, were normal peripheral blood slides under light microscope with 100x magnification; 828 images with 879 white blood cells were captured by a Nikon DSFi2 high definition color camera and saved in JPG format of size 960 x 1,280 pixels at 15 pixels per 1 μ m resolution. In dataset 2, 477 cropped white blood cell images were downloaded from CellaVision.com. They are in JPG format of size 360 x 363 pixels. The resolution is estimated to be 10 pixels per 1 μ m. Two sets of comparison were performed: segmentation and classification. The automatically segmented results were compared to the ones obtained manually by a haematologist. It was found that the proposed method is consistent and coherent in both datasets, with dice similarity over 87% for average segmented nucleus and cell regions. Feature selection and feature extraction were used and 13 features were selected for train models. Linear and quadratic discriminant function were compared with average accuracy over 95% for both model. The proposed system, based on normal white blood cell morphology and its characteristics, was applied to two different datasets. The results of the calibrated segmentation process on both datasets are fast, robust, efficient and coherent. Meanwhile, the classification of normal white blood cells into five types shows high sensitivity in both models.

Department: Electrical Engineering

Student's Signature

Field of Study: Electrical Engineering

Advisor's Signature

Academic Year: 2014

กิตติกรรมประกาศ

วิทยานิพนธ์นี้จะสำเร็จลุล่วงไม่ได้ ถ้าไม่ได้รับการช่วยเหลือจาก ผศ.ดร. ชาญชัย ปลื้มปีติวิริยะเวช ที่ช่วยให้คำปรึกษา ช่วยแก้ไขปัญหาต่างๆ เกี่ยวกับวิทยานิพนธ์มาโดยตลอด รวมถึงช่วยขัดเกลาเนื้อหา รูปเล่มทั้งฉบับนี้ และผลงานที่ส่งตีพิมพ์ในวารสารวิชาการ

ขอขอบพระคุณ ผศ.ดร. วณิดา พงศ์สถาพร และหมวดวิชาพยาธิชีววิทยา ภาควิชาวิทยาศาสตร์การแพทย์ มหาวิทยาลัยรังสิต ที่ช่วยอำนวยความสะดวกในการจัดเตรียมภาพเสมือนเลือด

ขอขอบพระคุณ รศ. นงลักษณ์ โควาวีสารัช และ รศ. ดร. นิตาชล ตั้งเสียมวิสัย ที่ช่วยอ่านและให้ข้อเสนอแนะรูปเล่ม และขอขอบพระคุณ ผศ. ดร. วิทยากร อัครวิเศษ ที่ช่วยชี้แนะแนวทางการแก้ปัญหา

วิทยานิพนธ์นี้ได้รับการสนับสนุนทุนวิจัยจาก "ทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัย" กองทุนรัชดาภิเษกสมโภช

ขอขอบพระคุณมหาวิทยาลัยรังสิตที่ได้ช่วยเหลือและสนับสนุนทุนในการศึกษาระดับดุษฎีบัณฑิต

สุดท้ายผู้เขียนขอขอบพระคุณเพื่อนร่วมงาน และสมาชิกในครอบครัวทุกท่าน ที่ได้ให้ความช่วยเหลือในด้านต่างๆ และเป็นกำลังใจให้มาโดยตลอด

สารบัญ

หน้า

| | |
|---|----|
| บทคัดย่อภาษาไทย..... | ง |
| บทคัดย่อภาษาอังกฤษ..... | จ |
| กิตติกรรมประกาศ..... | ฉ |
| สารบัญ..... | ช |
| สารบัญรูป | ฎ |
| สารบัญตาราง..... | ฅ |
| นิยามศัพท์บัญญัติ..... | 1 |
| บทที่ 1 บทนำ | 1 |
| 1.1. ความเป็นมาและความสำคัญของปัญหา | 1 |
| 1.2. วัตถุประสงค์ของการวิจัย..... | 2 |
| 1.3. ขอบเขตของการวิจัย..... | 2 |
| 1.4. วิธีดำเนินการวิจัย | 3 |
| 1.5. ประโยชน์ที่ได้รับจากการวิจัย | 4 |
| บทที่ 2 ทบทวนงานวิจัยที่เกี่ยวข้อง | 5 |
| บทที่ 3 ทฤษฎีที่เกี่ยวข้อง..... | 11 |
| 3.1. ข้อมูลเกี่ยวกับเซลล์เม็ดเลือดขาว | 11 |
| 3.2. ทฤษฎีและแนวคิดเกี่ยวกับการประมวลผลภาพ | 14 |
| 3.2.1. การประมวลผลภาพที่เกี่ยวกับลักษณะ รูปร่าง | 15 |
| 3.2.2. การตรวจหาขอบด้วยวิธี Canny..... | 16 |
| 3.2.3. คอนเวกซ์ฮัลล์ | 20 |
| 3.3. ทฤษฎีและแนวคิดเกี่ยวกับแอกทีฟคอนทัวร์ | 23 |
| 3.3.1. เซตระดับที่ใช้หลักการของขอบ | 25 |

| | |
|---|----|
| 3.3.2. เซตระดับที่ใช้หลักการของอาณาบริเวณ..... | 26 |
| 3.4. การหาระยะห่างกำลังสองน้อยที่สุดสำหรับการปรับรูปวงรี | 28 |
| 3.5. การหาความถูกต้องของการแบ่งส่วนภาพ..... | 31 |
| 3.6. การออกแบบตัวจำแนกรูปแบบโดยใช้ค่าทางสถิติ | 33 |
| 3.7. ลักษณะบ่งต่าง..... | 34 |
| 3.7.1. ลักษณะบ่งต่างที่ได้จากภาพลักษณะฐานสอง..... | 34 |
| 3.7.2. ลักษณะบ่งต่างของแผนภูมิแจกแจงความถี่ของระดับความเข้มของภาพ..... | 36 |
| 3.8. การเลือกลักษณะบ่งต่าง | 37 |
| 3.9. discriminant function | 39 |
| 3.10. วิธีการตรวจสอบสมรรถนะการทำงานของตัวจำแนก | 40 |
| 3.10.1. การทดสอบ | 40 |
| 3.10.2. การให้คะแนน | 41 |
| บทที่ 4 ระเบียบวิธีที่นำเสนอ..... | 43 |
| 4.1. การจัดเตรียมภาพ..... | 44 |
| 4.2. การแบ่งส่วนต่างๆของเม็ดเลือดขาวออกจากภาพ | 44 |
| 4.2.1. การแบ่งส่วนด้วยวิธีที่นำเสนอ | 44 |
| 4.2.2. การแบ่งส่วนด้วยวิธีที่แอดทีฟคอนทัวร์ | 54 |
| 4.3. การสกัดลักษณะบ่งต่างของเม็ดเลือดขาว | 55 |
| 4.4. การเลือกและการทดสอบลักษณะบ่งต่างของเม็ดเลือดขาวมาใช้ในการจำแนก | 56 |
| 4.5. การจำแนกประเภทเม็ดเลือดขาว..... | 58 |
| บทที่ 5 ผลการศึกษา..... | 59 |
| 5.1. การระบุตำแหน่งเซลล์เม็ดเลือดขาวในภาพ | 59 |
| 5.2. การแบ่งส่วนภาพเซลล์เม็ดเลือดขาวที่ละขั้นตอน | 60 |

| | |
|---|-----|
| 5.2.1. วิธีที่นำเสนอ..... | 60 |
| 5.2.2. วิธีใช้ ACWE1..... | 65 |
| 5.2.3. วิธีใช้ ACWE2..... | 66 |
| 5.3. การเปรียบเทียบการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี ACWE1 กับการแบ่งส่วน ภาพเซลล์เม็ดเลือดขาวด้วยผู้เชี่ยวชาญ..... | 67 |
| 5.4. การเปรียบเทียบการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี ACWE2 กับการแบ่งส่วน ภาพเซลล์เม็ดเลือดขาวด้วยผู้เชี่ยวชาญ..... | 69 |
| 5.5. การเปรียบเทียบการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธีที่นำเสนอกับการแบ่งส่วน ภาพเซลล์เม็ดเลือดขาวด้วยผู้เชี่ยวชาญ..... | 72 |
| 5.6. การเปรียบเทียบการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธีที่นำเสนอและ วิธี ACWE1 และ ACWE2 ในแง่ของความถูกต้อง | 74 |
| 5.7. การเปรียบเทียบการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธีที่นำเสนอและ วิธี ACWE1 และ ACWE2 ในแง่ของเวลาในการประมวลผลและผลการแบ่งส่วน | 75 |
| 5.8. การเปรียบเทียบเวลาในการประมวลผลการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธีที่ นำเสนอ วิธี ACWE1 และ ACWE2 กรณีพบเซลล์เม็ดเลือดขาวในภาพหนึ่งเซลล์ขึ้นไป..... | 81 |
| 5.9. การเปรียบเทียบผลภาพการแบ่งส่วนนิวเคลียสและเซลล์ด้วยวิธี ACWE1 และ ACWE2 เทียบกับวิธีที่นำเสนอ..... | 82 |
| 5.10. ผลการเลือกและการทดสอบความสามารถในการจำแนกของลักษณะบ่งต่างที่นำมาใช้..... | 87 |
| 5.11. ผลการจำแนกเซลล์..... | 89 |
| 5.12. ผลการให้คะแนนตัวจำแนก | 90 |
| บทที่ 6 ข้อสรุปและ/หรือข้อเสนอแนะ | 91 |
| รายการอ้างอิง | 93 |
| ภาคผนวก..... | 98 |
| ภาคผนวก ก | 99 |
| ภาคผนวก ข | 105 |

ญ

หน้า

| | |
|----------------------------------|-----|
| ภาคผนวก ค..... | 109 |
| ภาคผนวก ง..... | 113 |
| ประวัติผู้เขียนวิทยานิพนธ์ | 118 |



สารบัญรูป

| | | |
|--------------|--|----|
| รูปที่ 1. 1 | แสดงภาพรวมของวิธีดำเนินการวิจัย..... | 3 |
| รูปที่ 3. 1 | ตัวอย่างภาพเซลล์เม็ดเลือดขาวจากกล้องจุลทรรศน์..... | 13 |
| รูปที่ 3. 2 | (a) หลักการทำงานของ dilation (b) หลักการทำงานของ erosion | 16 |
| รูปที่ 3. 3 | ขั้นตอนวิธีการหาขอบวัตถุโดยวิธีของ Canny..... | 17 |
| รูปที่ 3. 4 | การจัดเรียงตัวของมุม | 18 |
| รูปที่ 3. 5 | อธิบาย non - maximum suppression ความแรงของขอบบ่งชี้ได้ด้วยสีและค่า ในขณะที่ทิศของเกรเดียนต์แสดงด้วยทิศทางของลูกศร พิกเซลที่เป็นขอบแสดงด้วย ขอบสีขาว..... | 19 |
| รูปที่ 3. 6 | ขั้นตอนการหาคอนเวกซ์ฮัลล์ด้วยวิธี GrahamScan | 21 |
| รูปที่ 3. 7 | ขั้นตอนการหาคอนเวกซ์ฮัลล์ด้วยวิธี Jarvis' March | 22 |
| รูปที่ 3. 8 | การแทนคอนทัวร์แบบ parametric | 23 |
| รูปที่ 3. 9 | การแทนคอนทัวร์แบบ geometric | 25 |
| รูปที่ 3. 10 | แสดงจุดข้อมูล (x_i, y_i) และเส้นโค้งวงรี C และระยะห่างระหว่างจุดข้อมูลและเส้น โค้ง $d((x_i, y_i), C)$ | 28 |
| รูปที่ 3.11 | วิธีการตรวจสอบหาความคล้ายคลึงของภาพด้วยวิธีทางสถิติบนพื้นฐานของการ ซ้อนทับของพื้นที่ภาพผลการแบ่งส่วนภาพสองชุด..... | 32 |
| รูปที่ 3.12 | การหาเส้นรอบรูป | 35 |
| รูปที่ 3.13 | วัตถุสองก้อนที่มีคอร์ดที่ยาวที่สุด (l_c) และคอร์ดที่มีความยาวมากที่สุดที่อยู่ใน แนวตั้งฉากกับคอร์ดที่ยาวที่สุด (l_p) โดยวัตถุทางซ้ายมีค่า eccentricity สูง วัตถุ ทางขวามีค่า eccentricity ต่ำ..... | 36 |
| รูปที่ 3.14 | ตัวอย่างการใช้ cross validation เพื่อทดสอบแบบจำลอง สำหรับ 4 fold | 41 |

| | | |
|--------------|--|----|
| รูปที่ 4. 1 | แผนภาพแสดงภาพรวมของระบบ (a) การจัดเตรียมภาพ (b) การแบ่งส่วน (c) การสกัดลักษณะบ่งต่างและการเลือกใช้ และ (d) การจำแนก | 43 |
| รูปที่ 4. 2 | ขั้นตอนการเตรียมภาพก่อนการประมวลผล | 45 |
| รูปที่ 4. 3 | ค่าระดับความเข้มสีเท่าใน RGB ของวัตถุต่าง ๆ ในภาพย้อมสีเลือด และ ruler scale ที่กำหนดโดยผู้ผลิต | 46 |
| รูปที่ 4. 4 | แสดงตัวอย่างของ neutrophil ที่มีนิวเคลียส 3 - 5 พู | 47 |
| รูปที่ 4. 5 | (a) ภาพที่ได้หลังจากขั้นตอนการเตรียมภาพก่อนการประมวลผล (b) ผลการแบ่งส่วนนิวเคลียส (c) การระบุจำนวนเซลล์ที่พบในภาพ | 48 |
| รูปที่ 4. 6 | กรณีวัตถุที่สนใจเป็นเซลล์เม็ดเลือดขาวแยกจากเซลล์อื่นๆ | 50 |
| รูปที่ 4. 7 | กรณีวัตถุที่สนใจเป็นเซลล์เม็ดเลือดขาวที่มีเซลล์ติดกันอยู่เล็กน้อย | 51 |
| รูปที่ 4. 8 | กรณีวัตถุที่สนใจเป็นเซลล์เม็ดเลือดขาวและมีเซลล์ติดกันจำนวนมาก | 51 |
| รูปที่ 4. 9 | แสดงขั้นตอนการแบ่งส่วนเซลล์ | 53 |
| รูปที่ 4. 10 | ขั้นตอนการแบ่งส่วนเซลล์ | 53 |
| รูปที่ 4.11 | ขั้นตอนของวิธี ACWE1 | 54 |
| รูปที่ 4.12 | แสดงผังการแบ่งส่วนเซลล์เม็ดเลือดขาวโดยวิธีของ Mathur และ คณะ [18] | 55 |
| รูปที่ 4.13 | กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 1 และ 2 | 57 |
| รูปที่ 5. 1 | แสดงผลการระบุตำแหน่งนิวเคลียสเซลล์เม็ดเลือดขาว | 59 |
| รูปที่ 5. 2 | แสดงการแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ basophil | 60 |
| รูปที่ 5. 3 | แสดงการแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ lymphocyte | 61 |
| รูปที่ 5. 4 | แสดงการแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ eosinophil | 62 |
| รูปที่ 5. 5 | แสดง การแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ lymphocyte | 63 |
| รูปที่ 5. 6 | แสดง การแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ eosinophil | 64 |
| รูปที่ 5. 7 | แสดงการแบ่งส่วนนิวเคลียสและเซลล์ด้วยวิธี ACWE1 | 65 |
| รูปที่ 5. 8 | แสดงการแบ่งส่วนนิวเคลียสและเซลล์ตามวิธี ACWE2 | 66 |

| | |
|---|-----|
| รูปที่ 5. 9 แสดงการเปรียบเทียบการแบ่งส่วนภาพนิวเคลียสและเซลล์ด้วยวิธี ACWE1 และ ACWE2 และ วิธีที่นำเสนอของข้อมูลภาพชุดที่หนึ่ง..... | 74 |
| รูปที่ 5. 10 แสดงการเปรียบเทียบการแบ่งส่วนภาพนิวเคลียสและเซลล์ด้วยวิธี ACWE1 และ ACWE2 และ วิธีที่นำเสนอของข้อมูลภาพชุดที่สอง | 75 |
| รูปที่ 5.11 แสดงการเปรียบเทียบผลภาพการแบ่งส่วนนิวเคลียสและเซลล์ของ basophil ด้วยวิธีต่างๆ..... | 83 |
| รูปที่ 5. 12 แสดงการเปรียบเทียบผลภาพการแบ่งส่วนนิวเคลียสและเซลล์ของ eosinophil ด้วยวิธีต่างๆ..... | 84 |
| รูปที่ 5. 13 แสดงการเปรียบเทียบผลภาพการแบ่งส่วนนิวเคลียสและเซลล์ของ lymphocyte ด้วยวิธีต่างๆ | 85 |
| รูปที่ 5. 14 แสดงการเปรียบเทียบผลภาพการแบ่งส่วนนิวเคลียสและเซลล์ของ monocyte ด้วยวิธีต่างๆ..... | 86 |
| รูปที่ 5. 15 แสดงการเปรียบเทียบผลภาพการแบ่งส่วนนิวเคลียสและเซลล์ของ neutrophil ด้วยวิธีต่างๆ | 86 |
| รูปที่ ก. 1 หา (a, b, d) เพื่อลดผลรวมของระยะทางตั้งฉากยกกำลัง..... | 99 |
| รูปที่ ก. 2 จุดที่เส้นที่ถูกรบกวนจากสัญญาณ Gaussian noise | 100 |
| รูปที่ ก. 3 กรณีสภาวะตัดกรวยทั่วไป..... | 101 |
| รูปที่ ก. 4 การจำแนกชนิดของภาคตัดกรวย..... | 102 |
| รูปที่ ข. 1 แสดงการแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ basophil | 105 |
| รูปที่ ข. 2 แสดงการแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ eosinophil..... | 105 |
| รูปที่ ข. 3 แสดงการแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ lymphocyte | 106 |
| รูปที่ ข. 4 แสดงการแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ monocyte | 107 |
| รูปที่ ข. 5 แสดง การแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ neutrophil..... | 108 |
| รูปที่ ง. 1 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และ ส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 1 และ 3 | 113 |

| | |
|--|-----|
| รูปที่ ง. 2 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และ ส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 1 และ 4 | 113 |
| รูปที่ ง. 3 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และ ส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 1 และ 5 | 114 |
| รูปที่ ง. 4 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และ ส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 2 และ 3 | 114 |
| รูปที่ ง. 5 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และ ส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 2 และ 4 | 115 |
| รูปที่ ง. 6 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และ ส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 2 และ 5 | 115 |
| รูปที่ ง. 7 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และ ส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 3 และ 4 | 116 |
| รูปที่ ง. 8 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และ ส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 3 และ 5 | 116 |
| รูปที่ ง. 9 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และ ส่วนขยายของกรณีเปรียบเทียบเซลล์ชนิดที่ 4 และ 5 | 117 |

สารบัญตาราง

| | |
|---|----|
| ตารางที่ 2. 1 แสดงตัวจำแนกที่ใช้ ลักษณะบ่งต่างที่สกัด เพอร์เซ็นต์ความถูกต้องของแต่ละ เอกสารอ้างอิง | 10 |
| ตารางที่ 3. 1 แสดงภาพจากกล้องจุลทรรศน์ จำนวนเปอร์เซ็นต์ที่พบในผู้ใหญ่ และเส้นผ่าน ศูนย์กลางของเซลล์เม็ดเลือดขาวแต่ละชนิด [25]..... | 14 |
| ตารางที่ 3. 2 แสดงรูปแบบเมทริกซ์ความสับสน (confusion matrix) | 42 |
| ตารางที่ 5. 1 แสดงผลระบุตำแหน่งเซลล์เม็ดเลือดขาวในภาพ | 59 |
| ตารางที่ 5. 2 แสดงผลการแบ่งส่วนนิวเคลียส ข้อมูลชุดที่ 1 | 67 |
| ตารางที่ 5. 3 แสดงผลการแบ่งส่วนเซลล์ ข้อมูลชุดที่ 1..... | 68 |
| ตารางที่ 5. 4 แสดงผลการแบ่งส่วนนิวเคลียส ข้อมูลชุดที่ 2..... | 68 |
| ตารางที่ 5. 5 แสดงผลการแบ่งส่วนเซลล์ ข้อมูลชุดที่ 2..... | 69 |
| ตารางที่ 5. 6 แสดงผลการแบ่งส่วนนิวเคลียส ข้อมูลชุดที่ 1..... | 70 |
| ตารางที่ 5. 7 แสดงผลการแบ่งส่วนเซลล์ ข้อมูลชุดที่ 1..... | 70 |
| ตารางที่ 5. 8 แสดงผลการแบ่งส่วนนิวเคลียส ข้อมูลชุดที่ 2..... | 71 |
| ตารางที่ 5. 9 แสดงผลการแบ่งส่วนเซลล์ ข้อมูลชุดที่ 2..... | 71 |
| ตารางที่ 5. 10 แสดงผลการแบ่งส่วนนิวเคลียส ข้อมูลชุดที่ 1..... | 72 |
| ตารางที่ 5. 11 แสดงผลการแบ่งส่วนเซลล์ ข้อมูลชุดที่ 1..... | 72 |
| ตารางที่ 5. 12 แสดงผลการแบ่งส่วนนิวเคลียส ข้อมูลชุดที่ 2..... | 73 |
| ตารางที่ 5. 13 แสดงผลการแบ่งส่วนเซลล์ ข้อมูลชุดที่ 2..... | 73 |
| ตารางที่ 5. 14 แสดงจำนวนรอบและเวลาเฉลี่ยที่ใช้การแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี ACWE1 | 76 |
| ตารางที่ 5. 15 แสดงส่วนเบี่ยงเบนมาตรฐานของจำนวนรอบและเวลาที่ใช้การแบ่งส่วนภาพเซลล์ เม็ดเลือดขาวด้วยวิธี ACWE1 | 76 |

| | |
|--|----|
| ตารางที่ 5. 16 แสดงจำนวนรอบและเวลาเฉลี่ยที่ใช้การแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี ACWE2 | 77 |
| ตารางที่ 5. 17 แสดงส่วนเบี่ยงเบนมาตรฐานของจำนวนรอบและเวลาที่ใช้การแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี ACWE2 | 77 |
| ตารางที่ 5. 18 แสดงเวลาเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่ใช้การแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธีที่นำเสนอ | 78 |
| ตารางที่ 5. 19 แผนภาพกล่อง หรือ box and whisker plots แสดงข้อมูลสูงสุด ข้อมูลที่มีค่าเปอร์เซ็นต์ไทล์ 75 มัชฐาน ข้อมูลที่มีค่าเปอร์เซ็นต์ไทล์ 25 ข้อมูลต่ำสุด และค่าเฉลี่ย (ดอกจันทรีสีแดง) ของเวลาที่ใช้ในการหาพื้นที่นิวเคลียส เซลล์และเวลารวมของวิธีแอกทิฟคอนทัวร์เทียบกับวิธีที่นำเสนอ ชุดข้อมูลที่ 1 | 79 |
| ตารางที่ 5. 20 แผนภาพกล่อง หรือ box and whisker plots แสดงข้อมูลสูงสุด ข้อมูลที่มีค่าเปอร์เซ็นต์ไทล์ 75 มัชฐาน ข้อมูลที่มีค่าเปอร์เซ็นต์ไทล์ 25 ข้อมูลต่ำสุด และค่าเฉลี่ย (ดอกจันทรีสีแดง) ของเวลาที่ใช้ในการหาพื้นที่นิวเคลียส เซลล์และเวลารวมของวิธีแอกทิฟคอนทัวร์เทียบกับวิธีที่นำเสนอ ชุดข้อมูลที่ 2 | 80 |
| ตารางที่ 5. 21 เวลาในการประมวลผลการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี ACWE1 | 81 |
| ตารางที่ 5. 22 เวลาในการประมวลผลการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี ACWE2 | 81 |
| ตารางที่ 5. 23 เวลาในการประมวลผลการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธีที่นำเสนอ | 82 |
| ตารางที่ 5. 24 ลักษณะบ่งต่างที่เลือกใช้ จำนวน 13 ลักษณะบ่งต่าง มีรายละเอียดดังต่อไปนี้..... | 87 |
| ตารางที่ 5. 25 ผลการจำแนกเซลล์ basophil, eosinophil, lymphocyte, monocyte และ neutrophil ในขั้นตอนการฝึก | 88 |
| ตารางที่ 5. 26 ผลการจำแนกเซลล์ basophil, eosinophil, lymphocyte, monocyte และ neutrophil ด้วย linear discriminant function..... | 89 |
| ตารางที่ 5. 27 ผลการจำแนกเซลล์ basophil, eosinophil, lymphocyte, monocyte และ neutrophil ด้วย quadratic discriminant function..... | 89 |
| ตารางที่ 5. 28 ค่า accuracy, sensitivity, specificity และ precision ในการจำแนก basophil, eosinophil, lymphocyte, monocyte และ neutrophil ด้วย linear discriminant function | 90 |

| | |
|--|-----|
| ตารางที่ 5. 29 ค่า accuracy, sensitivity, specificity และ precision ในการจำแนก basophil, eosinophil, lymphocyte, monocyte และ neutrophil ด้วย quadratic discriminant function..... | 90 |
| ตารางที่ ค. 1 feature ที่สามารถสกัดได้จากผลลัพธ์ของการแบ่งส่วนภาพ รายละเอียด ดังต่อไปนี้..... | 109 |
| ตารางที่ ง. 1 แสดงข้อมูลลักษณะบ่งต่างที่เลือกจากกราฟ การกระจาย cumulative distribution function (CDF) ของ p-values..... | 117 |



นิยามศัพท์บัญญัติ

| ลำดับที่ | ศัพท์ | ศัพท์บัญญัติ |
|----------|----------------------|--|
| 1 | granule | เม็ดเล็ก ๆ |
| 2 | morphology | รูปร่าง โครงสร้าง |
| 3 | cytoplasm | ไซโทพลาซึม |
| 4 | binary image | ภาพลักษณะฐานสอง |
| 5 | histogram | ฮิสโทแกรม |
| 6 | interface | ส่วนต่อประสาน |
| 7 | threshold | ขีดแบ่ง |
| 8 | gradient | เกรเดียนต์ |
| 9 | level set | เซตระดับ |
| 10 | normalize; normalise | ทำให้เป็นบรรทัดฐาน |
| 11 | feature | ลักษณะบ่งต่าง |
| 12 | feature extraction | การสกัดลักษณะบ่งต่าง |
| 13 | training | กระบวนการฝึกให้ระบบเรียนรู้ |
| 14 | testing | กระบวนการทดสอบการหาผลลัพธ์จากระบบที่ ออกแบบ |
| 15 | classification | การจำแนกประเภท |
| 16 | pattern | แบบรูป |
| 17 | pattern recognition | การรู้จำแบบ |

บทที่ 1

บทนำ

1.1.ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันการวินิจฉัยโรคต่าง ๆ จำเป็นต้องมีการนำข้อมูลหลาย ๆ ด้าน เช่น ภาพถ่าย x-ray ผลการตรวจเลือด ปัสสาวะ เป็นต้น มาใช้ประกอบการวินิจฉัย สำหรับผลการตรวจเลือดจะถูกดำเนินการในห้องปฏิบัติการโลหิตวิทยา สิ่งที่นักโลหิตวิทยาให้ความสนใจคือจำนวนของเซลล์เม็ดเลือด โดยขั้นตอนการวิเคราะห์เลือดในห้องปฏิบัติการ อันดับแรกจะทำการนับเซลล์ โดยวิธี light scattering และ flow cytochemical และจะมี 21% ของตัวอย่างเลือดที่จำเป็นต้องตรวจสอบวิเคราะห์หาความผิดปกติโดยใช้กล้องจุลทรรศน์และตรวจสอบโดยนักโลหิตวิทยา [1]

ในกรณีที่ต้องทำการตรวจหาความผิดปกติจากเซลล์เม็ดเลือด นักโลหิตวิทยาจะใช้ข้อมูลจากภาพของเลือดบนแผ่นแก้ว หรือทางการแพทย์เรียกว่า "สเมียร์เลือด (blood smear)" วิธีเตรียมคือเจาะเลือดจากผู้ที่ได้รับการตรวจ หยดเลือด 1 หยด ลงบนปลายด้านหนึ่งของแผ่นแก้วใสขนาดประมาณ 3 x 6 เซนติเมตร แล้วใช้แผ่นแก้วอีกแผ่นไถหยดเลือดให้แผ่ออกไปยังปลายอีกด้านหนึ่ง การไถมีจุดประสงค์ให้เม็ดเลือดเรียงกันเป็นชั้นเดียว ไม่ทับกัน ซึ่งจะช่วยให้ง่ายต่อการดูลักษณะของเม็ดเลือดที่ละเซลล์ จากนั้นรอให้เลือดแห้งแล้วจึงนำไปย้อมสีเพื่อนำไปดูผ่านกล้องจุลทรรศน์ ซึ่งจะให้ข้อมูลเกี่ยวกับทั้งเม็ดเลือดแดง เม็ดเลือดขาวและเกล็ดเลือด เช่น ขนาดปกติหรือไม่ การติดสีปกติไหม มีเซลล์ที่รูปร่างผิดปกติบ้างไหม เม็ดเลือดขาวมีจำนวนเป็นเท่าไร ลักษณะปกติไหม เกล็ดเลือดมีจำนวนปกติไหม ขนาดและการติดสีปกติไหม เป็นต้น การแปลผลสเมียร์เลือดมักจะทำควบคู่ไปกับการตรวจนับเม็ดเลือด ซึ่งจะได้ข้อมูลที่เป็ประโยชน์มากมายสำหรับการวินิจฉัยโรคของเม็ดเลือด

เนื่องจากการแปลผลสเมียร์เลือดมักเป็นการตรวจผ่านกล้องจุลทรรศน์เป็นการทำงานที่ทำซ้ำ ๆ และใช้เวลาทำงานติดต่อกันเป็นเวลานานทำให้เกิดอาการล้าของสายตาในผู้ปฏิบัติงาน และผลการตรวจยังขึ้นกับความชำนาญของผู้ทำการตรวจ จากที่กล่าวข้างต้นทำให้ผู้วิจัยเกิดแนวคิดในการออกแบบระบบการแบ่งส่วนและจำแนกเซลล์เม็ดเลือดขาวขึ้น เนื่องจากเซลล์เม็ดเลือดขาวมีหลายชนิด และแต่ละชนิดมีความแตกต่างกันทำให้ยากต่อการวิเคราะห์และต้องใช้ความชำนาญของผู้ตรวจเป็นสำคัญ ระบบที่ออกแบบขึ้นมีขั้นตอนคือ การระบุตำแหน่งของเซลล์เม็ดเลือดขาว การแบ่งส่วนภาพเซลล์เม็ดเลือดขาวออกเป็นส่วนของนิวเคลียสและไซโทพลาซึม การสกัดคุณลักษณะที่สำคัญ และการจำแนกเซลล์เม็ดเลือดขาว

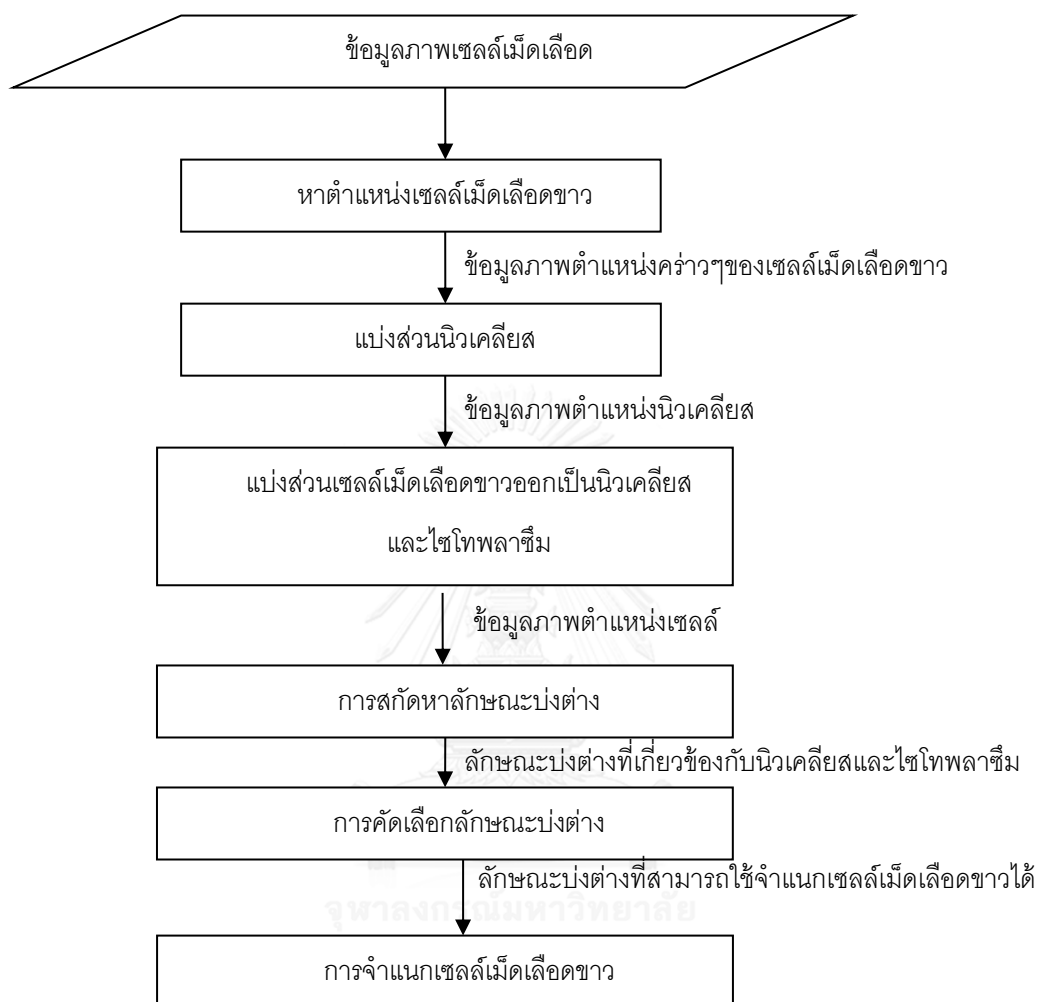
1.2. วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาและเปรียบเทียบวิธีการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวโดยใช้แอกทีฟคอนทัวร์ แบบต่าง ๆ
2. เพื่อออกแบบวิธีการในการแยกเซลล์เม็ดเลือดขาวออกจากเซลล์ชนิดอื่น ๆ
3. เพื่อออกแบบวิธีการแบ่งส่วนเซลล์เม็ดเลือดขาว เป็น นิวเคลียส และ ไซโทพลาซึม
4. เพื่อจำแนกเซลล์เม็ดเลือดขาวออกเป็น 5 ชนิด คือ basophil, eosinophil, lymphocyte, monocyte, และ neutrophil

1.3. ขอบเขตของการวิจัย

1. ออกแบบวิธีการในการแยกเซลล์เม็ดเลือดขาวออกจากเซลล์ชนิดอื่น ๆ ได้อย่างมีประสิทธิภาพ
2. ออกแบบและเลือกใช้อัลกอริทึมในการแบ่งส่วนเซลล์เม็ดเลือดขาวเป็นนิวเคลียส และ ไซโทพลาซึม
3. ทดสอบการทำงานของอัลกอริทึมที่ออกแบบกับภาพเซลล์เม็ดเลือดของคนปกติ
4. เปรียบเทียบวิธีการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวโดยใช้แอกทีฟคอนทัวร์ กับวิธีที่ออกแบบ
5. จำแนกชนิดของเซลล์เม็ดเลือดขาวออกเป็น 5 ชนิด คือ basophil, eosinophil, lymphocyte, monocyte และ neutrophil

1.4. วิธีดำเนินการวิจัย



รูปที่ 1. 1 แสดงภาพรวมของวิธีดำเนินการวิจัย

วิธีดำเนินการวิจัยนี้ประกอบด้วย 6 ขั้นตอน ดังแสดงในรูปที่ 1. 1 คือ การหาตำแหน่งเซลล์เม็ดเลือดขาว การแบ่งส่วนนิวเคลียส การแบ่งส่วนเซลล์ และการสกัดหาลักษณะบ่งต่าง การเลือกลักษณะบ่งต่าง และการจำแนกเซลล์เม็ดเลือดขาวออกเป็น 5 ชนิด คือ basophil, eosinophil, lymphocyte, monocyte, และ neutrophil

1.5. ประโยชน์ที่ได้รับจากการวิจัย

สามารถนำวิทยานิพนธ์ที่ได้ไปใช้ในการจำแนกเซลล์เม็ดเลือดขาวแบบปกติได้ นอกเหนือจากนี้ยังสามารถประยุกต์เทคนิคที่ใช้ในวิทยานิพนธ์นี้ไปใช้กับงานการประมวลผลภาพในด้านอื่น ๆ ที่คล้ายคลึงกันได้ แต่อาจจะต้องมีการปรับเปลี่ยนพารามิเตอร์ หากกำลังขยายของกล้องจุลทรรศน์ หรือลักษณะบ่งต่างของกล้องแตกต่างกันไป



บทที่ 2

ทบทวนงานวิจัยที่เกี่ยวข้อง

การแบ่งส่วนภาพเป็นกระบวนการแบ่งส่วนภาพเป็นหลาย ๆ ส่วนโดยมีวัตถุประสงค์เพื่อนำเสนอข้อมูลภาพให้ง่ายต่อการวิเคราะห์มากขึ้น สำหรับการแบ่งส่วนเซลล์เม็ดเลือดขาว โดยปกติจะแบ่งส่วนเป็นนิวเคลียสและไซโทพลาซึม มีนักวิจัยนำเสนอการแบ่งส่วนภาพด้วยวิธีต่าง ๆ สามารถจัดกลุ่มได้ดังนี้ edge-based methods, morphological operators technique, clustering algorithm, novel color, an iterative Otsu's approach, active contour model technique, watershed clustering technique, neural network approach (PCNN, online trained NN และอื่น ๆ)

Lin และคณะ [2] ได้นำเสนอ color gradient vector flow (GVF) active contour model สำหรับการแบ่งส่วนเซลล์เม็ดเลือดขาว เมื่อนำเกรเดียนต์ของสีและการประเมินประสิทธิภาพด้วย L2E เข้ามารวมอยู่กับ GVF แบบดั้งเดิม และให้ snake ในพื้นที่สี LUV เคลื่อนที่ไปยังส่วนนิวเคลียสและไซโทพลาซึม ในขณะที่ประสิทธิภาพการแบ่งส่วนดีเมื่อเทียบกับวิธี mean shift และ GVF แบบดั้งเดิม การทดสอบข้อมูลจะไม่สามารถแยกแยะความแตกต่างพื้นผิวที่มี texture อ่อน ทำให้วิธีนี้มีข้อจำกัดในการแบ่งส่วนเซลล์เม็ดเลือดขาวทั้งห้าประเภท

Sadeghian และคณะ [3] เสนอวิธีการแบ่งส่วนนิวเคลียสและไซโทพลาซึมสำหรับภาพระดับสีเทาที่ใช้ GVF และขั้นตอนวิธีการกำหนดค่าขีดแบ่งด้วยวิธีของ Zack (Zack's thresholding) ขั้นตอนแรกเลือกภาพย่อยด้วยคน และตรวจจับขอบของภาพโดยใช้ Canny edge detector เพื่อใช้ในการระบุนิวเคลียส ใช้ GVF ในภาพระดับสีเทา หลังจากลบนิวเคลียสออกจากไซโทพลาซึมจะถูกแบ่งจากส่วนที่เหลือของภาพย่อยโดยใช้ วิธีการกำหนดค่าขีดแบ่งของ Zack's อย่างไรก็ตามวิธีนี้จำเป็นต้องเลือกภาพย่อยด้วยคน และกรณีที่เซลล์ประกอบด้วยนิวเคลียสที่มีหลาย ๆ ก้อน การแยกความแตกต่างไซโทพลาซึมจากนิวเคลียสของเซลล์นั้นยาก และ บางครั้งเซลล์เม็ดเลือดแดงมีความเข้มคล้ายกับไซโทพลาซึม ดังนั้นการใช้ฮิสโตแกรมระดับสีเทา (Gray level histogram) จึงไม่มีประสิทธิภาพนักสำหรับการแบ่งส่วนไซโทพลาซึมออกจากเซลล์เม็ดเลือดแดง การศึกษานี้ไม่ได้ประเมินประสิทธิภาพการทำงานเมื่อเปรียบเทียบกับวิธีอื่น ๆ และการประเมินข้อมูลทำเพียง 20 ข้อมูล

Qingmin และคณะ [4] ได้เสนอการแบ่งส่วนภาพสเมียร์เลือดเพื่อการนับแยกเซลล์เม็ดเลือดขาวโดยนำ shape analysis มาใช้สำหรับการแบ่งส่วนเซลล์เม็ดเลือดขาว โดยการใช้การกำหนดค่า

ขีดแบ่งอย่างง่าย และหาคอนทัวร์ (contour) รอบนิวเคลียส ด้วยขั้นตอน shape analysis อย่างไรก็ตามแม้จะมีความง่ายและมีประสิทธิภาพ แต่วิธีนี้ใช้เฉพาะกับเซลล์เม็ดเลือดขาวที่มีรูปร่างกลม เช่น lymphocytes เป็นต้น

Kan และคณะ [5] ได้รวมเทคนิค scale space filtering และ watershed clustering เพื่อหลีกเลี่ยงปัญหาความหลากหลายและความซับซ้อนเชิงพื้นที่โดยการแยกเซลล์เม็ดเลือดขาวเป็นสองส่วนคือ นิวเคลียสและไซโทพลาซึมตามลำดับโดยใช้วิธีที่ต่างกัน อันดับแรก แยกภาพเซลล์เม็ดเลือดขาวออกจากเซลล์อื่น ๆ แล้วใช้ scale space filtering เพื่อแยกส่วนของนิวเคลียสออกจากรูปร่างย่อย หลังจากนั้นแบ่งส่วนไซโทพลาซึมด้วย watershed ในภาพฮิสโทแกรมสามมิติของ HSV (hue, saturation, value) สุดท้ายใช้ morphological operation เพื่อให้ได้เส้นโครงที่เชื่อมถึงกัน การใช้ feature space มีข้อดีคือเป็นการแยกแบบตรง ๆ อย่างไรก็ตามข้อด้อยของวิธีนี้คือการกำหนดจำนวนของ cluster ล่วงหน้าทำได้ยาก

Fang และคณะ [6] เสนอรูปแบบการแบ่งส่วนภาพเซลล์เม็ดเลือดขาว โดยเครือข่ายประสาทที่ฝึกแบบ on line โดยใช้ meanshift และการสุ่มตัวอย่างแบบ uniform เป็นเครื่องมือเริ่มต้นเพื่อลดข้อมูลที่จะนำมาฝึก (train) นอกจากนี้ยังใช้ particle swarm optimization เพื่อเพิ่มประสิทธิภาพเครือข่ายประสาทในการลู่เข้า (convergence) ให้รวดเร็ว ความถูกต้องในการแบ่งส่วนเซลล์เม็ดเลือดขาว และเวลาการทำงานเพิ่มขึ้นเมื่อเทียบกับวิธีการ train แบบดั้งเดิมนี้ แต่วิธีนี้ไม่สามารถที่จะแยกแยะความแตกต่างระหว่างนิวเคลียสและไซโทพลาซึม

Jianhua และคณะ [7] กล่าวว่า การแบ่งส่วนของเลือด การตรวจหาขอบของเซลล์ทำได้ยากเนื่องจากเส้นขอบไม่คมชัดและยากที่จะได้ข้อมูลของขอบทุกขอบอย่างถูกต้อง เขาจึงได้พัฒนาวิธีการทำซ้ำของ Otsu เข้ากับ Circular histogram ในการแบ่งส่วนเซลล์เม็ดเลือดขาว

Wermser และคณะ [8] และ Cseke และคณะ [9] ใช้การกำหนดค่าขีดแบ่งเป็นลำดับชั้น (heirachical thresholding) โดยใช้ลักษณะทางสีจำนวนสองปริภูมิสีในการเลือกแบ่งระหว่างส่วนประกอบของภาพสเมียร์เลือด

Wang และคณะ [10] เสนอการแบ่งส่วนด้วยการกำหนดค่าขีดแบ่ง โดยใช้ mathematical morphology และ Fuzzy cellular neural network สำหรับการตรวจหาเซลล์เม็ดเลือดขาว อย่างไรก็ตามแม้จะการทำงานได้อย่างรวดเร็วและผลการตรวจสอบที่ดี แต่วิธีนี้ยังไม่สามารถที่จะแยกแยะความแตกต่างนิวเคลียสจากออกจากไซโทพลาซึม

Yampri และคณะ [11] นำเสนอการกำหนดค่าขีดแบ่งแบบอัตโนมัติ โดยใช้ข้อมูลภาพส่วนสีเขียวเพื่อให้มีความคมชัดสูงระหว่างไซโทพลาซึมและนิวเคลียส สัญญาณรบกวน (noise) ที่เหลืออยู่

ในภาพจะถูกกลบออกแล้วใช้กระบวนการทางสัณฐานวิทยาขั้นพื้นฐาน (basic morphological process) ในขณะที่นิวเคลียสและไซโทพลาซึมจะถูกแบ่งโดยใช้คอนทัวร์ที่ปรับตัวได้ (adaptive contour) อย่างไรก็ตามเซลล์เม็ดเลือดขาวบางชนิด เช่น neutrophils และ eosinophils ซึ่งมีนิวเคลียสมากกว่าหนึ่งนิวเคลียสในเซลล์ทำให้แยกแยะความถูกต้องระหว่างไซโทพลาซึมและนิวเคลียสได้ยาก

Dorini และคณะ [12] เสนอ watershed transform บนพื้นฐานของการเชื่อมต่อสำหรับการแบ่งส่วนของนิวเคลียส ขณะที่ใช้ forrest transform เป็นเครื่องมือทั่วไปสำหรับการออกแบบตัวดำเนินการการประมวลผลภาพ ขณะที่ข้อมูลการกระจายของขนาด (size distribution information) ถูกนำมาใช้การแยกไซโทพลาซึมจากพื้นหลังและเซลล์เม็ดเลือดแดง ใช้ morphological opening เพื่อเพิ่มขนาด มีประสิทธิภาพสำหรับนิวเคลียสบางอย่าง ปัญหาที่พบเมื่อรูปร่างของไซโทพลาซึมที่มีไม่ได้มีรูปร่างกลม และขนาดของการเปลี่ยนแปลงไปตามชนิดของเซลล์เม็ดเลือดขาว

Ghosh และคณะ [13] ได้เสนอวิธีการจำแนกเซลล์เม็ดเลือดขาวโดยใช้ Fuzzy divergence และการปรับเปลี่ยนเทคนิคการกำหนดค่าขีดแบ่ง เมื่อแบ่งนิวเคลียสโดยใช้ fuzzy membership function : Gamma, Gaussian, และ Cauchy ไปที่พิกเซลของภาพ สำหรับการแบ่งส่วนนิวเคลียสมีความถูกต้อง แต่ไม่มีวิธีการในการแบ่งส่วนไซโทพลาซึม ซึ่งในการวิเคราะห์เซลล์เม็ดเลือดขาวเป็นส่วนที่มีความสำคัญเท่าเทียมกัน

นอกจากนี้มีการนำข้อมูลทางสีของภาพมาใช้ในการแบ่งส่วนภาพเซลล์เม็ดเลือด Sadr และคณะ [14] ใช้ ข้อมูล Cb และ Y ใน YCbCr เพื่อใช้ตรวจหา นิวเคลียส โดยหานิวเคลียสจากอัตราส่วน Cb/Y และตัดภาพย่อยโดยใช้ข้อมูล ใน Green channel หลังจากนั้นใช้ แอกรีฟคอนทัวร์ โดยไม่ใช้ขอบในการแบ่งส่วนภาพ แต่วิธีนี้ทดลองในภาพ 20 ภาพ และแบ่งส่วนเฉพาะนิวเคลียสเท่านั้น Bergen และคณะ [15] ใช้ HIS space สำหรับทำ pre - processing และใช้ RGB space โดยใช้อัตราส่วนระหว่าง G และ B ในแต่ละพิกเซลเพื่อให้นิวเคลียสเด่นขึ้น และทำการกำหนดค่าขีดแบ่งด้วยวิธีของ Ostu แบ่งส่วนนิวเคลียสและใช้ level set เพื่อแบ่งส่วนภาพเซลล์เม็ดเลือดขาว, Shi และคณะ [16] ใช้ข้อมูล RGB ร่วมกับ iterative GMM function โดยใช้ circular histogram เพื่อติดตามข้อมูล นิวเคลียสของเซลล์เม็ดเลือดขาว

Wei และคณะ [17] ได้ใช้ complex wavelet transform และ watershed สำหรับการแบ่งส่วนภาพ

Mathur และคณะ [18] ได้นำเสนอวิธีการแบ่งส่วนภาพโดยเริ่มจากการนอร์มอลไลซ์ภาพด้วยเทคนิคการแปลงสีระหว่างภาพที่นำเสนอโดย Reinhard และคณะ [19] หลังจากนั้นนำภาพที่ผ่านการนอร์มอลไลซ์แล้ว มาแปลงจากปริภูมิสี RGB ไปเป็น HSV แยกภาพจากช่องสัญญาณ saturation ไปใช้ในการหาคอนทราสต์เริ่มต้นในการแบ่งส่วนนิวเคลียส โดยทำการแปลงภาพเป็นภาพลักษณะฐานสองด้วยวิธีการกำหนดค่าขีดแบ่งอย่างง่าย ทำ opening morphological กรองวัตถุที่สนใจด้วยพื้นที่ เมื่อได้คอนทราสต์เริ่มต้นจะใช้แอกทีฟคอนทราสต์แบบ narrow band ในการแบ่งส่วนนิวเคลียส ส่วนการแบ่งส่วนเซลล์เม็ดเลือดขาว จะทำในภาพจากช่องสัญญาณ hue โดยใช้คอนทราสต์เริ่มต้นมาจาก การแปลงภาพเป็นภาพลักษณะฐานสองด้วยวิธีการกำหนดค่าขีดแบ่งอย่างง่าย ทำ morphological และใช้ connected component analysis (CCA) เมื่อได้คอนทราสต์เริ่มต้นจะใช้แอกทีฟคอนทราสต์แบบ narrow band ในการแบ่งส่วนเซลล์

ในส่วนของการจำแนกชนิดของเซลล์เม็ดเลือดขาวก็ได้มีการนำเทคนิคและลักษณะบ่งต่างต่าง ๆ มาใช้

Ramoser และคณะ [20] ใช้ K-mean ใน HSL space โดยกำหนด $k = 3$, ลดบทบาทของพื้นหลังโดยการกำจัดสีพื้นหลังเฉลี่ยดำเนินการในแต่ละช่องสีและ สร้างภาพความน่าจะเป็นที่มีค่าสูงสำหรับพิกเซลเม็ดเลือดขาวและค่าต่ำสำหรับ ส่วนอื่น ๆ ภาพนี้จะถูกแบ่งโดยใช้การปรับตัว การกำหนดค่าขีดแบ่งเพื่อติดตามเซลล์เม็ดเลือดขาว ลักษณะบ่งต่างที่ใช้ในการจำแนกเซลล์เม็ดเลือดขาว คือ ลักษณะบ่งต่างทางสถิติของสี (ค่าเฉลี่ย, ส่วนเบี่ยงเบนมาตรฐาน, skewness จาก hue, saturation, และ luminance ของนิวเคลียสและไซโทพลาซึม), ลักษณะบ่งต่างทางรูปร่างของนิวเคลียส (convexity, principal axis ratio, compactness, circular variance, elliptic variance) , ขนาดของนิวเคลียส, ขนาดของไซโทพลาซึม และใช้วิธี support vector machine จำแนก ทำในภาพ 1166 ภาพ (จำแนกเซลล์เป็น 13 กลุ่ม) ผลลัพธ์การแบ่งส่วนถูกต้องเฉลี่ย 95% และผลการจำแนกอยู่ในช่วง 75% ถึง 99%

Rezatofighi และคณะ [21] นำเสนอขั้นตอนวิธีที่ใช้จำแนกเซลล์เม็ดเลือดขาวในเลือดทั้ง 5 ชนิดแบบอัตโนมัติโดย Gram-Schmidt orthogonalization พร้อมกับ snake algorithm เพื่อแบ่งส่วนนิวเคลียสและไซโทพลาซึมของเซลล์ จากนั้นสกัดลักษณะบ่งต่างที่จากข้อมูลที่แบ่งส่วนมาได้ ลักษณะบ่งต่างทาง morphological (พื้นที่นิวเคลียส, พื้นที่ไซโทพลาซึม, เส้นรอบรูปนิวเคลียส, เส้นรอบรูปเซลล์, จำนวนก้อนนิวเคลียส, ค่าเฉลี่ยและค่าความแปรปรวนของขอบเขตนิวเคลียสและไซโทพลาซึมและอัตราส่วนระหว่างไซโทพลาซึมและนิวเคลียส, ค่าความกลมของนิวเคลียสและเซลล์) ลักษณะบ่งต่างทางพื้นผิว (co - occurrence matrix จำนวน 14 ลักษณะ), Local Binary Pattern หรือ (LBP จำนวน 2 ลักษณะ) และเลือกใช้ลักษณะบ่งต่างด้วยวิธีการเลือกไปข้างหน้า (Sequential

Forward Selection หรือ SFS) และการแสดงผลลัพธ์เปรียบเทียบกับวิธีการจำแนกสองชนิดคือ เครือข่ายประสาทเทียม (Artificial Neural Network หรือ ANN) และ Support Vector Machine หรือ SVM ผลลัพธ์ที่แสดงให้เห็นว่าวิธีการที่นำเสนอมีความถูกต้องและรวดเร็วพอที่จะใช้ใน ห้องปฏิบัติการทางโลหิตวิทยา

Theera-Umporn และ Dhompongsa [22] ได้ตั้งสมมุติฐานว่านิวเคลียสเพียงอย่างเดียว สามารถใช้จำแนกเซลล์เม็ดเลือดขาวได้เนื่องจากนิวเคลียสสามารถแบ่งส่วนได้ง่ายโดยทำการทดสอบ ในภาพที่ได้จากไซสั่นหลังจากมีเซลล์เม็ดเลือดขาวหนาแน่น โดยใช้ลักษณะบ่งต่าง 4 อย่าง จาก นิวเคลียส เรียกว่า pattern spectrum และทดสอบเปรียบเทียบกับตัวจำแนก Bayes classifiers และ artificial neural networks โดย cross validation ด้วย fold เท่ากับ 5 พบว่าลักษณะบ่งต่าง ที่สกัดมาสามารถจำแนกได้ถูกต้อง 77% โดยเฉลี่ย

Piuri และ Scotti [23] ใช้เทคนิค contrast stretching, Canny edge detection, dilation และ filling เพื่อแบ่งส่วน membrane ของเซลล์เม็ดเลือดขาวและทำการตัดพื้นที่ของ membrane ที่ได้ก่อนหน้า และทำ contrast stretching และ Minimum intensity homogenization เพื่อหาส่วนของนิวเคลียส หลังจากนั้นสกัดลักษณะบ่งต่าง 23 ตัว คือ Area, Perimeter, Convex Area, Solidity, Major Axis Length, Orientation, Filled Area, และ Eccentricity, อัตราส่วนพื้นที่ของเซลล์เทียบกับนิวเคลียส, nucleus' rectangularity, cell circularity, the number of lobes, solidity, area and mean gray - level intensity ของ cytoplasm ทดสอบกับภาพ 134 ภาพ โดยเลือกลักษณะบ่งต่างด้วยเทคนิค forward selection บน nearest neighbor classifier ประเมินผลด้วย Leave One Out วิธี โดยทดสอบ cross validation กำหนด fold เท่ากับ 10 พบว่า ค่าเฉลี่ยความผิดพลาด 8%

หากสรุปรายละเอียดของตัวจำแนกที่ใช้ ลักษณะบ่งต่างที่สกัดของงานวิจัยต่าง ๆ สามารถสรุปได้ดัง ตารางที่ 2. 1

ตารางที่ 2. 1 แสดงตัวจำแนกที่ใช้ ลักษณะบ่งต่างที่สกัด เปอร์เซ็นต์ความถูกต้องของแต่ละ

เอกสารอ้างอิง

| Classifier | Extracted features | Accuracy | reference |
|------------------------|--|----------|-------------------------------------|
| Bayesian classifier | Area of the nucleus, location of its pattern spectrum's peak, first, and second granulometric moments of the pattern spectrum | 77% | Theera-Umpon และ Dhompongsa[20] |
| | Area, perimeter, compactness, eccentricity, orientation, solidity, form factor, and roundness | 83.2% | Ghosh et al.[13] |
| k-Nearest neighbor | Area, perimeter, convex area, solidity, major axis length, orientation, filled area, eccentricity, the ratio between the cell and nucleus areas, rectangularity, circularity, number of lobes, mean intensity of the cytoplasm | 86% | Piuri และ Scotti [21] |
| Neural network | Area, perimeter, convex area, solidity, major axis length, orientation, filled area, eccentricity, the ratio between the cell and nucleus areas, rectangularity, circularity, number of lobes, mean intensity of the cytoplasm | 92% | Piuri และ Scotti [21] |
| | Area of the nucleus, location of its pattern spectrum's peak, first and second granulometric moments of the pattern spectrum | 77% | Theera-Umpon และ Dhompongsa [20] |
| | Nucleus area, perimeter, number of the separated parts of nucleus, mean, variance of the nucleus boundaries, roundness of nucleus, averaged nucleus color, fourteen features based on co-occurrence matrix, and two features based on Local Binary Pattern | 96% | Rezatofghi และ Soltanian-Zadeh[19] |
| Support vector machine | mean, standard deviation, skewness from hue, saturation, luminance of nucleus and cytoplasm, Convexity, Principal axis ratio, Compactness, Circular variance, and Elliptic variance on nucleus | 75% -99% | Ramoser[18] |
| | Nucleus area, perimeter, number of the separated parts of nucleus, mean, variance of the nucleus boundaries,roundness of nucleus, averaged nucleus color, fourteen features based on co-occurrence matrix, and two features based on Local Binary Pattern | 96% | Rezatofghi and Soltanian-Zadeh [19] |

บทที่ 3

ทฤษฎีที่เกี่ยวข้อง

ในบทนี้เป็นการเสนอทฤษฎีที่จะนำมาประยุกต์ในการระบุตำแหน่งของเซลล์เม็ดเลือดขาว วิธีการแบ่งส่วนภาพเซลล์เม็ดเลือดขาว ออกเป็นขอบเขตของเซลล์และนิวเคลียส การสกัดลักษณะบ่งต่างที่มีนัยสำคัญสำหรับการนำมาใช้ในการจำแนกชนิดของเซลล์เม็ดเลือดขาว โดยเสนอทฤษฎีที่เกี่ยวข้องสำหรับวิทยานิพนธ์นี้ เริ่มจากข้อมูลเกี่ยวกับเซลล์เม็ดเลือดขาว ทฤษฎีและแนวคิดเกี่ยวกับการประมวลผลภาพ ทฤษฎีและแนวคิดเกี่ยวกับแอกทิฟคอนทัวร์ เทคนิคการทำ least square fitting สำหรับวงรี การสกัดลักษณะบ่งต่าง การเลือกลักษณะบ่งต่าง และ การจำแนก

3.1. ข้อมูลเกี่ยวกับเซลล์เม็ดเลือดขาว

องค์ประกอบของเซลล์เม็ดเลือดมีหลายชนิด โดยปกติเซลล์เม็ดเลือดในกระแสเลือดคนปกติ [24] มี 3 ชนิด คือ เม็ดเลือดแดง (red blood cell หรือ erythrocyte) เม็ดเลือดขาว (white blood cell หรือ leukocyte) และเกล็ดเลือด (platelet หรือ thrombocyte)

ในกระแสเลือดคนปกติ ส่วนใหญ่เป็นเม็ดเลือดแดงที่เจริญเต็มที่แล้ว มีรูปร่างกลมไม่มีนิวเคลียส มีเส้นผ่านศูนย์กลางประมาณ 7 ไมโครเมตร หรือขนาดประมาณเท่ากับนิวเคลียสของเม็ดเลือดขาวชนิด lymphocyte ขนาดเล็ก ติดสีชมพู ตรงกลางเซลล์ไม่ติดสี ซึ่งบริเวณที่ไม่ติดสีจะมีความกว้างประมาณ 1 ใน 3 ของเส้นผ่านศูนย์กลาง

เม็ดเลือดขาว เป็นเซลล์ของระบบภูมิคุ้มกัน เซลล์เม็ดเลือดขาวมีหลายชนิดแต่ชนิดทำหน้าที่คอยป้องกันร่างกายจากทั้งเชื้อก่อโรคและสารแปลกปลอมต่าง ๆ กัน เซลล์เม็ดเลือดขาวเป็นเซลล์ที่พบได้ทั่วไปในร่างกาย รวมไปถึงในเลือดและในระบบน้ำเหลือง สามารถแบ่งออกได้เป็น 2 พวกใหญ่ คือ เซลล์เม็ดเลือดขาวที่มีเม็ดเล็ก ๆ (granulocytes หรือ polymorphonuclear leukocyte) และเซลล์เม็ดเลือดขาวที่ไม่มีเม็ดเล็ก ๆ (agranulocytes หรือ mononuclear leukocyte)

เม็ดเลือดขาวที่มีเม็ดเล็ก ๆ จำเพาะ (specific granules) สามารถใช้ในการบอกแยกชนิดกันได้ เม็ดเลือดขาวพวกนี้จะมีนิวเคลียสแยกออกจากกันเป็นหลายพู (lobes) แต่ละก้อนจะเชื่อมต่อกันด้วยเส้นโครมาทิน (chromatin) บาง ๆ ทำให้มีชื่อเรียกอีกชื่อหนึ่งว่า polymorphonuclear leukocyte มีอยู่ด้วยกัน 3 ชนิด neutrophil, eosinophil, และ basophil

neutrophil (ดูรูปที่ 3. 1 (a)) มีเส้นผ่านศูนย์กลาง 9 - 15 μm (ไมโครเมตรหรือไมครอน) มีขนาดประมาณ 2 เท่าของเม็ดเลือดแดง นิวเคลียสมี 2 ถึง 5 พู หรือมากกว่าและโครมาตินติดสีม่วงน้ำเงิน เม็ดเล็ก ๆ จำเพาะมีขนาดเล็กละเอียด ติดสีชมพูอมม่วง กระจายอยู่ทั่วไปในไซโทพลาซึม เรียกว่า neutrophilic granules พบ neutrophil ได้ประมาณ 50 - 70 เปอร์เซ็นต์ของเม็ดเลือดขาว หน้าที่ของ neutrophil คือต่อต้านเชื้อโรคที่เข้าสู่ร่างกาย และเป็นเม็ดเลือดขาวชนิดแรกที่ไปถึงเนื้อเยื่อที่มีการติดเชื้อ โดยเฉพาะอย่างยิ่งการติดเชื้อแบคทีเรีย การกลืนกินอาศัยการโอบล้อมเชื้อโรคเข้าสู่เซลล์ และรวมเชื้อโรคเข้ากับไลโซโซม (lisosome) ที่ภายในมีเอ็นไซม์ที่ทำหน้าที่ย่อยสลายและรวมกับเปปไทด์สังเคราะห์ที่เรียกว่า ดีเฟนซิน (defensins) กระบวนการนี้ทำให้เม็ดเล็ก ๆ ภายในไซโทพลาซึมของ neutrophil ลดลง เรียกว่า degranulation

eosinophil (ดูรูปที่ 3. 1 (b)) มีเส้นผ่านศูนย์กลาง 9-15 μm มองเห็นนิวเคลียสได้ชัดเจน และมักจะมี 2 พู อาจอยู่ติดกันหรืออาจแยกกัน โครมาตินติดสีทึบเช่นเดียวกับ neutrophil เป็นส่วนมาก เม็ดเล็ก ๆ จำเพาะมีขนาดใหญ่เท่า ๆ กันย้อมติดสีส้มแดง เรียกว่า eosinophilic granules พบ eosinophil ประมาณ 1-3 เปอร์เซ็นต์ของเซลล์เม็ดเลือดขาว เม็ดเลือดขาวชนิดนี้พบค่อนข้างน้อยในร่างกาย แต่จะเพิ่มมากขึ้นเมื่อร่างกายมีการแพ้ หรือติดเชื้อปรสิตหรือพยาธิ eosinophil มีหน้าที่กำจัดพิษจากโปรตีนแปลกปลอม (foreign protein) ที่เข้าสู่ร่างกายหรืออาจกำจัดสารพิษที่สร้างโดยแบคทีเรียและปรสิต

basophil (ดูรูปที่ 3. 1 (c)) มีเส้นผ่านศูนย์กลางประมาณ 10-16 μm มองเห็นนิวเคลียสไม่ชัดเจน เนื่องจากถูกบดบังเม็ดเล็ก ๆ จำเพาะมีขนาดใหญ่บ้างเล็กบ้างไม่เท่ากัน ติดสีม่วงอมดำ เรียกว่า basophilic granules โดยปกติพบ basophil น้อยกว่า 1 เปอร์เซ็นต์ของเซลล์เม็ดเลือดขาว นิวเคลียสมีรูปร่างไม่แน่นอนและมีขนาดใหญ่เกินเนื้อที่ประมาณครึ่งหนึ่งของเซลล์ มีหน้าที่ป้องกันไม่ให้เลือดแข็งตัว และหลั่งสารฮิสตามีน (histamine) ซึ่งเป็นสารก่อภูมิแพ้

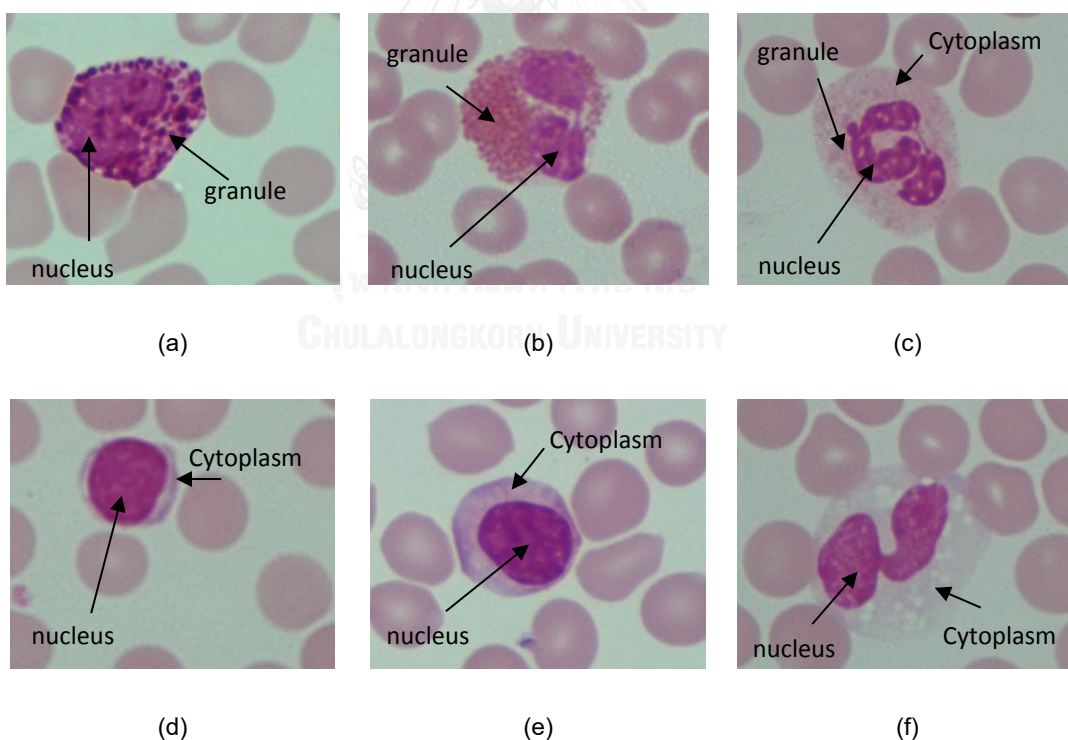
เม็ดเลือดขาวที่ไม่มีเม็ดเล็ก ๆ จำเพาะ (specific granules) แต่อาจจะพบเม็ดเล็ก ๆ เม็ดเล็ก ๆ ติดสีชมพูแดง หรือ azurophilic granules เม็ดเลือดขาวกลุ่มนี้จะมีนิวเคลียสอันเดียว บางครั้งเรียก mononuclear leukocyte ได้แก่ lymphocyte และ monocyte

lymphocyte มีนิวเคลียสเป็นรูปกลม รูปไข่ หรือรูปไต โครมาตินหยาบเป็นปื้น ติดสีม่วงเข้ม ไซโทพลาซึมติดสีฟ้าอ่อน และใส โดยทั่วไป lymphocyte แบ่งออกตามขนาดได้ 2 ชนิด คือ lymphocyte ขนาดเล็กและ lymphocyte ขนาดใหญ่ ที่พบส่วนใหญ่เป็น lymphocyte ขนาดเล็ก นิวเคลียสติดสีม่วงน้ำเงิน มีหน้าที่สร้างภูมิคุ้มกันแก่ร่างกาย ช่วยสมานแผล พบ lymphocyte 25 - 35% ของเซลล์เม็ดเลือดขาว

lymphocyte ขนาดเล็ก หรือ small lymphocyte (ดูรูปที่ 3. 1 (d)) มีเส้นผ่านศูนย์กลาง 7 - 8 μm ส่วนใหญ่มีขนาดใหญ่กว่าเม็ดเลือดแดงเล็กน้อย ชนิดนี้มีปริมาณของไซโทพลาซึมน้อยมาก จนบางครั้งแทบมองไม่เห็นไซโทพลาซึม อาจพบหรือไม่พบ azurophilic granules ก็ได้

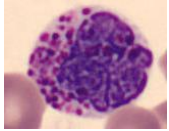
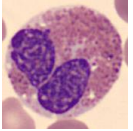
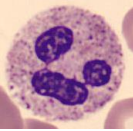
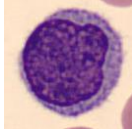
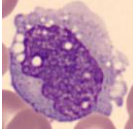
lymphocyte ขนาดใหญ่ หรือ large lymphocyte (ดูรูปที่ 3. 1 (e)) มีเส้นผ่านศูนย์กลาง 12 - 18 μm มีขนาดใหญ่กว่าเม็ดเลือดแดง 3 - 4 เท่า ชนิดนี้มีปริมาณของไซโทพลาซึมมาก และมักจะพบ azurophilic granules

monocyte มีขนาดแตกต่างกันได้มาก คือมีขนาดเส้นผ่านศูนย์กลาง 12 - 20 μm นิวเคลียสมีรูปร่างคล้ายคลื่นสมองหรืออุ้งเท้าสัตว์ บางครั้งอาจเห็นรูปร่างคล้ายไต เม็ดสีถั่ว หรือรูปเกือบม้า โครมาตินจะอยู่กันหลวม ๆ เห็นช่องว่างระหว่างโครมาตินได้ชัดเจน (ลักษณะของโครมาตินช่วยจำแนกระหว่าง lymphocyte และ monocyte) ไซโทพลาซึมมีปริมาณมากและติดสีฟ้าปนเทา อาจพบ azurophilic granules ได้ พบประมาณ 4 - 6 เปอร์เซ็นต์ของเซลล์เม็ดเลือดขาว ทำหน้าที่กลืนกินเชื้อโรคและสิ่งแปลกปลอมที่เข้าสู่ร่างกาย



รูปที่ 3. 1 ตัวอย่างภาพเซลล์เม็ดเลือดขาวจากกล้องจุลทรรศน์
 (a) basophil, (b) eosinophil, (c) neutrophil, (d) small lymphocyte,
 (e) large lymphocyte, (f) monocyte

ตารางที่ 3. 1 แสดงภาพจากกล้องจุลทรรศน์ จำนวนเปอร์เซ็นต์ที่พบในผู้ใหญ่ และเส้นผ่านศูนย์กลางของเซลล์เม็ดเลือดขาวแต่ละชนิด [25]

| ชนิด | Granulocytes | | | Agranulocytes | |
|--------------------------------|---|---|---|---|---|
| | basophil | eosinophil | neutrophil | lymphocyte | monocyte |
| ภาพจากกล้องจุลทรรศน์ |  |  |  |  |  |
| จำนวนเปอร์เซ็นต์ที่พบในผู้ใหญ่ | 0.4% | 2.3% | 62% | 30% | 5.3% |
| เส้นผ่านศูนย์กลาง (μm) | 10-16 | 9-15 | 9-15 | lymphocyte ขนาดเล็ก 7-8 lymphocytes ขนาดใหญ่ 12-18 | 12-20 |

เกล็ดเลือด (platelet หรือ thrombocyte) เป็นเซลล์เม็ดเลือดที่มีขนาดเล็กที่สุด มีขนาดเล็กกว่า 1 ใน 3 ของเม็ดเลือดแดง มีเส้นผ่านศูนย์กลางประมาณ 2 - 4 μm มีรูปร่างแบนคล้ายจาน มีรูปกลมหรือรูปไข่ มองเห็นขอบเขตของเซลล์ได้ไม่ชัดเจน ไม่มีนิวเคลียส ภายในมีเม็ดเล็ก ๆ ติดสีชมพูแดง (azurophilic granules) อยู่รวมกันมากบริเวณกลางเซลล์ อาจจะมีเห็นไซโทพลาซึมสีฟ้า

3.2. ทฤษฎีและแนวคิดเกี่ยวกับการประมวลผลภาพ

ทฤษฎีและแนวคิดเกี่ยวกับการประมวลผลภาพที่นำมาใช้ในวิทยานิพนธ์ฉบับนี้ จะแบ่งเป็นสามส่วน คือ การประมวลผลภาพที่เกี่ยวข้องกับลักษณะ รูปร่าง การตรวจหาขอบด้วยวิธี Canny และ คอนเวกซ์ฮัลล์

3.2.1. การประมวลผลภาพที่เกี่ยวข้องกับลักษณะ รูปร่าง

การประมวลผลภาพที่เกี่ยวข้องกับลักษณะ รูปร่าง (morphological image processing) [26] ใช้สำหรับตัดต่อ หรือแต่งเติมส่วนขอบเขตของวัตถุ โครงสร้างของวัตถุในภาพ โดยใช้ทฤษฎีของเซต ซึ่งเซตใน morphology จะแทนรูปร่างหรือรูปทรงของวัตถุในภาพ เช่น กลุ่มของสีดำทั้งหมดในภาพลักษณะพื้นฐานสองสำหรับการทำ morphology สามารถใช้ในการกำจัดสัญญาณรบกวน ขยายพื้นที่ของวัตถุ และกำจัดส่วนเกินของวัตถุได้

ภาพอินพุตที่ใช้ทำ morphology ทางคณิตศาสตร์ มี 2 ภาพ คือ ภาพที่นำมาใช้ประมวลผล และ องค์ประกอบโครงสร้าง structural element หรือ SE โดยปรกติภาพ SE จะมีขนาดเล็กกว่า ภาพที่นำมาประมวลผล มีรูปร่างและขนาดแตกต่างกันไปสามารถกำหนดได้

การทำ dilation คือ การขยายขนาดของพิกเซล โดยการสแกนค่าของ SE บนแต่ละพิกเซล ในภาพ โดยทำการสแกนจากตำแหน่งบนซ้ายไปยังตำแหน่งล่างขวา ซึ่งจะเปลี่ยนค่าของพิกเซลที่มีค่าเป็น 0 ให้มีค่าเป็น 1 เมื่อค่าของพิกเซลใดพิกเซลหนึ่งบน SE มีค่าตรงกับค่าของพิกเซลภาพและจะมีค่าคงเดิมเมื่อทุกค่าของ SE มีค่าตรงกับทุกค่าของพิกเซลภาพ ภาพเอาต์พุตที่ได้มีขอบเขตและทิศทางของความหนาที่ถูกควบคุมโดยขนาดและรูปร่างของ SE ดังรูปที่ 3. 2 (a)

หากเขียนในรูปทางคณิตศาสตร์การขยายขนาดของเซต A โดย B เขียนแทนด้วยสัญลักษณ์ $A \oplus B$ โดยมีสมการ

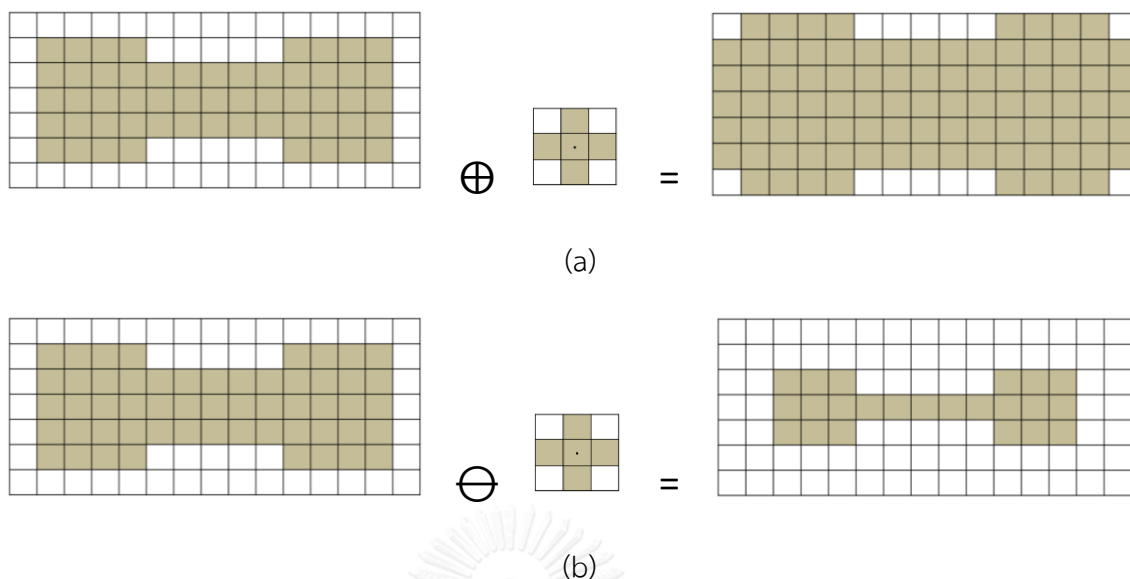
$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \emptyset\} \quad (1)$$

เมื่อ B คือ SE, $(\hat{B})_z$ คือ SE ในรูปแถวลำดับที่จุดศูนย์กลางของแถวลำดับสแกนไปที่ละตำแหน่งของ A , \emptyset คือ เซตว่าง

erosion เป็นวิธีการที่ตรงข้ามกับ dilation คือจะลดขนาดของพิกเซล โดยการสแกนค่าของ SE บนแต่ละค่าของพิกเซลภาพ โดยทำการสแกนจากตำแหน่งบนซ้ายไปยังตำแหน่งล่างขวาซึ่งจะเปลี่ยนค่าของพิกเซลที่มีค่าเป็น 1 ให้มีค่าเป็น 0 เมื่อพิกเซลใดพิกเซลหนึ่งบน SE มีค่าตรงกับค่าของพิกเซลภาพ และจะมีค่าคงเดิมเมื่อทุกพิกเซลของ SE มีค่าตรงกับค่าของพิกเซลภาพดังรูปที่ 3. 2 (b) โดยมีสมการดังนี้

$$A \ominus B = \{z | (\hat{B})_z \cap A = \emptyset\} \quad (2)$$

เมื่อ B คือ SE, $(\hat{B})_z$ คือ SE ในรูปแถวลำดับที่จุดศูนย์กลางของแถวลำดับสแกนไปที่ละตำแหน่งของ A , \emptyset คือ เซตว่าง



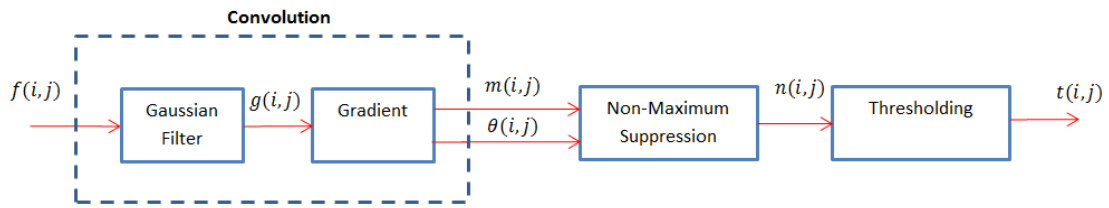
รูปที่ 3. 2 (a) หลักการทำงานของ dilation (b) หลักการทำงานของ erosion

3.2.2. การตรวจหาขอบด้วยวิธี Canny

วัตถุประสงค์ของการตรวจหาขอบ คือ การลดปริมาณของข้อมูลในภาพแต่ยังคงรักษาคุณสมบัติโครงสร้างที่จะใช้สำหรับการประมวลผลภาพต่อไป วิธีการที่มีอยู่หลายขั้นตอนแต่วิทยานิพนธ์นี้มุ่งสนใจไปที่การตรวจหาขอบด้วยวิธี Canny ซึ่งพัฒนาโดย John F. Canny ในปี 1986 [27] แม้ว่าจะเป็นเรื่องที่ยากแต่เป็นการตรวจหาขอบมาตรฐาน และใช้แพร่หลายในงานวิจัย

ขั้นตอนการทำงานแบ่งเป็น 5 ขั้นตอน คือ

1. การปรับภาพให้เรียบ (smoothing) เป็นการเบลอของภาพเพื่อลดสัญญาณรบกวน
2. การหาเกรเดียนต์ ขอบจะหาได้จากเกรเดียนต์ของภาพที่มีค่าสูง
3. non - maximum suppression จุดที่คล้ายกับจุดที่สูงที่สุดเท่านั้นที่จะทำเครื่องหมายเป็นขอบ
4. การกำหนดค่าขีดแบ่งสองชั้น ขอบที่เป็นไปได้จะถูกหาโดยการทำการกำหนดค่าขีดแบ่ง
5. การติดตามขอบโดย hysteresis ขอบที่ถูกเลือกสุดท้ายจะหาจากการยกเลิกขอบอื่นๆ ที่ไม่ได้เชื่อมต่อกับขอบจริง



รูปที่ 3.3 ขั้นตอนวิธีการหาขอบวัตถุโดยวิธีของ Canny

การทำงานของ Canny edge detection [28] นั้นเริ่มต้นจากการปรับภาพให้เรียบ (smoothing) ด้วยตัวกรองเกาส์เซียน (Gaussian filter) เพื่อกำจัดสัญญาณรบกวน หลังจากนั้นคำนวณค่าขนาด (magnitude) และทิศทาง (orientation) ของ gradient โดยใช้การหาอนุพันธ์อันดับหนึ่ง ในขั้นตอนถัดมาจึงใช้ non - maxima suppression กับ gradient magnitude เพื่อให้ได้ขอบที่บางลง และในขั้นตอนสุดท้ายใช้ double - precision floating - point (หรือ double) thresholding algorithm เพื่อระบุพิกเซลที่เป็นขอบและช่วยเชื่อมต่อขอบ โดยในแต่ละขั้นตอนมีรายละเอียดดังต่อไปนี้

3.2.2.1. การปรับภาพให้เรียบ

ก่อนการหาขอบโดยขั้นตอนวิธีนี้ มักจะต้องกำจัดสัญญาณรบกวนออกก่อนโดยใช้ตัวกรองเกาส์เซียน ซึ่งสามารถคำนวณได้จากการใช้หน้าตางขนาดเล็ก (mask) ขนาดของหน้าตางเกาส์เซียนนี้ หากมีขนาดกว้างจะมีผลทำให้ลดสัญญาณรบกวนได้มาก แต่ถ้ากว้างมากเกินไปจะมีผลทำให้ขอบย่อย ๆ ที่เป็นส่วนรายละเอียดนั้นหายไป สำหรับการคำนวณหาภาพที่ได้จากการใช้ตัวกรองเกาส์เซียนเป็นดังสมการ

$$g(i,j) = G(i,j,\sigma) * f(i,j) \quad (3)$$

เมื่อ $g(i,j)$ คือ ภาพที่ผ่านการปรับภาพให้เรียบ , $G(i,j,\sigma)$ คือ เคอร์เนลของตัวกรองเกาส์เซียนมีส่วนเบี่ยงเบนมาตรฐาน คือ σ , และ $f(i,j)$ คือ ภาพอินพุต

3.2.2.2. การหาเกรเดียนต์

ขั้นตอนวิธีของ Canny ทำการหาขอบเมื่อความเข้มของระดับขาวเทาในภาพมีการเปลี่ยนแปลงมากที่สุด ในการคำนวณหาทิศทางขอบภาพ ในขั้นแรกนำภาพที่ผ่านการขจัดสัญญาณรบกวน $g(i, j)$ มาสร้างอนุพันธ์ย่อย (partial derivatives) ตามแนวแกนนอน x และแนวแกนตั้ง y ตามลำดับ หลังจากนั้น ทำการแปลง gradient ในพิกัดฉาก g_x และ g_y เป็นพิกัดเชิงขั้ว คือ

$$m(i, j) = \sqrt{g_x(x, y)^2 + g_y(x, y)^2} \quad (4)$$

$$\theta(i, j) = \tan^{-1} \left(\frac{g_x(x, y)}{g_y(x, y)} \right) \quad (5)$$

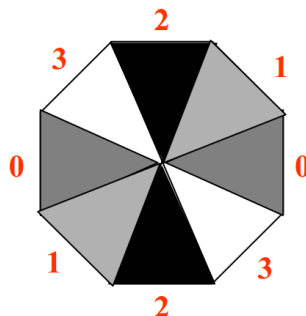
โดยที่ค่า $g_x(x, y)$ และ $g_y(x, y)$ คือ gradient ในพิกัดฉากแนวแกนนอน x และ y ตามลำดับ

3.2.2.3. non - maxima suppression

วัตถุประสงค์ของขั้นตอนนี้คือการแปลงขอบในภาพของเกรเดียนต์ที่เบลอให้ขอบคมชัดขึ้น โดยทั่วไปนี้จะทำโดยการคง local maxima ทั้งหมดในภาพเกรเดียนต์และการลบอันอื่นออก ขั้นตอนวิธีสำหรับแต่ละพิกเซลในภาพเกรเดียนต์ แบ่งเป็นสองขั้นตอนคือ การจัดเรียงตัวของมุม และการตรวจสอบแต่ละพิกเซล

3.2.2.3.1 แบ่งการจัดเรียงตัวของมุม จะทำการหมุนทิศทางของเกรเดียนต์ ให้เข้าใกล้มุม 45° ที่สุด เพื่อให้สัมพันธ์กับพื้นที่ใกล้เคียงที่เชื่อมต่อกัน 8 ด้าน

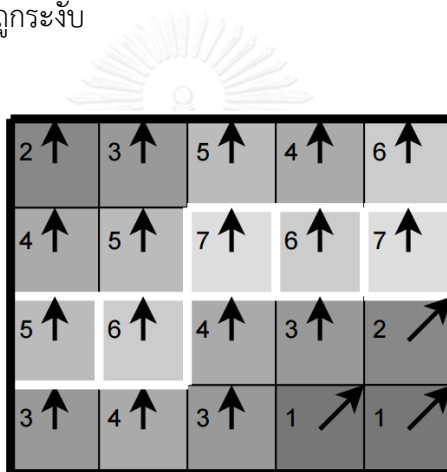
$$\vartheta(i, j) = \text{sector}(\theta(i, j)) \quad (6)$$



รูปที่ 3. 4 การจัดเรียงตัวของมุม

3.2.2.3.2. การตรวจสอบที่แต่ละพิกเซล

ทำการเปรียบเทียบค่าความแรงของขอบของพิกเซลปัจจุบัน กับ ความแรงของขอบของพิกเซลในทิศทางบวกและทิศทางลบ กล่าวคือ ถ้าทิศทางเกรเดียนต์อยู่ทางเหนือ ($\theta = 90^\circ$) เปรียบเทียบกับพิกเซลไปทางทิศเหนือและทิศใต้ หากมีความแข็งแรงของขอบของพิกเซลปัจจุบันมีค่าใหญ่ที่สุด; ให้คงค่าของความแรงขอบไว้ ไม่ต้องลบออกไป ตัวอย่างง่ายๆของ non - maximum suppression จะแสดงในรูปที่ 3. 5 เกือบทุกพิกเซลมีทิศทางเกรเดียนต์ชี้ไปทางทิศเหนือ จึงเทียบกับพิกเซลบนและด้านล่าง พิกเซลที่เป็นค่าสูงสุดในการเปรียบเทียบนี้มีการทำเครื่องหมายด้วยเส้นขอบสีขาว พิกเซลอื่น ๆ ทุกตัว จะถูกระงับ



รูปที่ 3. 5 อธิบาย non - maximum suppression ความแรงของขอบบ่งชี้ได้ด้วยสีและค่า ในขณะที่ทิศของเกรเดียนต์แสดงด้วยทิศทางของลูกศร พิกเซลที่เป็นขอบแสดงด้วยขอบสีขาว

3.2.2.4. การกำหนดค่าขีดแบ่ง (thresholding)

ภาพที่ได้อาจยังมีเส้นที่ไม่ใช่ขอบจริงของวัตถุปรากฏอยู่เนื่องจากสัญญาณรบกวนหรือวัตถุในภาพมีลักษณะเป็นพื้นผิวมีรายละเอียดภายในหรือมีลวดลาย ในการเพื่อลดปัญหาเหล่านี้จึงต้องทำการกำหนดค่าขีดแบ่ง ขึ้นมา 2 ค่า คือ ค่าขีดแบ่งช่วงสูง (T_1) และ ค่าขีดแบ่งช่วงต่ำ (T_2) พิกเซลที่ค่าสูงกว่า T_1 ก็จะถูกรับเป็น “1” เป็นพิกเซลที่เป็นขอบ แต่ถ้าน้อยกว่า T_2 ก็จะถูกรับเป็น 0 ส่วนค่าที่อยู่ระหว่างค่าขีดแบ่งทั้งสอง จะขึ้นอยู่กับพิกเซลที่อยู่รอบข้าง หากพบว่าพิกเซลที่อยู่รอบข้างของพิกเซลที่เป็นขอบ (ค่า > T_1) มีค่ามากกว่า T_2 แล้ว จะปรับค่าพิกเซลนั้นให้มีค่าเป็น “1” และถือเป็นพิกเซลของภาพขอบด้วยเช่นกัน

3.2.2.5. การติดตามขอบโดย hysteresis

ขอบที่มีค่าสูง จะถูกตีความเป็น "ขอบจริง" และสามารถรวมอยู่ในขอบของภาพสุดท้ายได้ทันที ขอบที่มีค่าอ่อนจะถูกรวมถ้าหากมีการเชื่อมต่อกับขอบที่ค่าสูง โดยมีแนวคิดว่าสัญญานรบกวนสัญญาณขนาดเล็กรูปแบบอื่น ๆ ไม่น่าจะส่งผลให้มีขอบที่ค่าสูง ดังนั้นขอบที่มีค่าสูงเท่านั้นที่จะเป็นขอบในภาพต้นฉบับจริง ขอบค่าอ่อนอาจจะเป็นขอบจริงหรือสัญญาณรบกวนอย่างใดอย่างหนึ่งได้เนื่องจากขอบจริงหรือสัญญาณรบกวนหรือเกิดจากความแตกต่างของสี โดยประเภทหลังมีการกระจายได้อย่างอิสระบนขอบของภาพทั้งหมดและมีเพียงเล็กน้อยจะตั้งอยู่ติดกับขอบที่มีค่าสูง ขอบที่มีค่าอ่อนแต่เป็นขอบจริงมีมากและมีแนวโน้มที่จะเชื่อมต่อกับขอบที่มีค่าสูง การติดตามขอบสามารถดำเนินการโดย BLOB - analysis หรือ Binary Large Object พิกเซลขอบจะแบ่งออกเป็น BLOB โดยใช้การเชื่อมต่อกับข้างเคียงแปดด้าน BLOB มีพิกเซลขอบที่มีค่าสูงอย่างน้อยหนึ่งจะถูกเก็บไว้ในขณะที่ BLOB อื่นถูกยกเลิก

3.2.3. คอนเวกซ์ฮัลล์

คอนเวกซ์ฮัลล์ (convex Hull) คือ เซตของจุด H ใน \mathcal{R}^d ซึ่งเป็นเซตคอนเวกซ์ที่เล็กที่สุดที่ประกอบด้วยจุดครบทุกจุด โดยเซตของจุดทุกจุด คือ S การหาคอนเวกซ์ฮัลล์คือการหาเซตย่อยของจุดที่เป็นรูปหลายเหลี่ยมคอนเวกซ์ขนาดเล็กที่สุดที่จุดใน S ทุกจุด อยู่ทั้งบนขอบหรือภายในของรูปหลายเหลี่ยมนั้น ขั้นตอนวิธีการหาคอนเวกซ์ฮัลล์มีหลายวิธี คือ การค้นหาขอบเขตของคอนเวกซ์โดยกวาดหมุน วิธีแบ่งแยกและเอาชนะ และ วิธีอื่นๆ

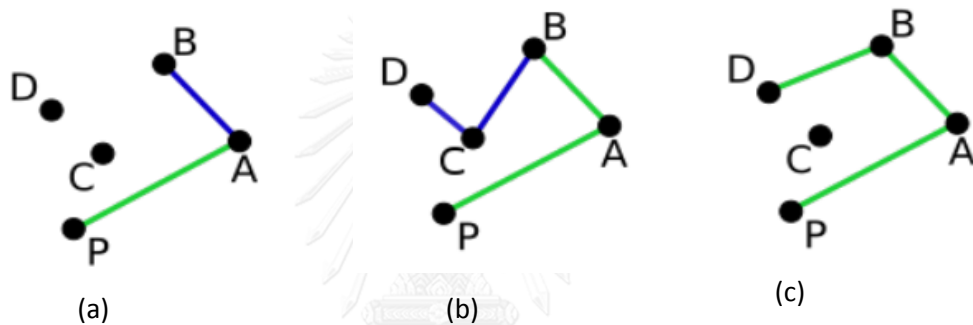
3.2.3.1. การค้นหาขอบเขตของคอนเวกซ์โดยกวาดหมุน

ขั้นตอนวิธีการ Graham scan [29] เริ่มต้นด้วยการเลือกจุดที่แน่นอนบน convex hull แล้วทำเพิ่มจุดลงไปใน convex hull ทำกระบวนการนี้ ซ้ำ ๆ

1. ให้ H เป็นรายการของจุดบน convex hull โดยเริ่มต้นจากเซตว่าง
2. เลือกจุด p_0 ซึ่งเป็นจุดที่มีพิกัด y ต่ำที่สุด เพิ่มการ p_0 ลงไปใน H เนื่องจาก p_0 เป็นจุดสมาชิก convex hull แน่นนอน
3. ให้ (p_0, p_1, \dots, p_h) เป็นจุดที่เหลือที่ผ่านการเรียงตามมุมชี้ของแต่ละจุดเมื่อเทียบกับ p_0 โดยเรียงจากมุมเล็กที่สุดไปยังมุมที่ใหญ่ที่สุดที่จะทำซ้ำ ทำการกวาดรอบ p_0 ผ่านจุดที่เรียงลำดับในทิศทางเข็มนาฬิกา

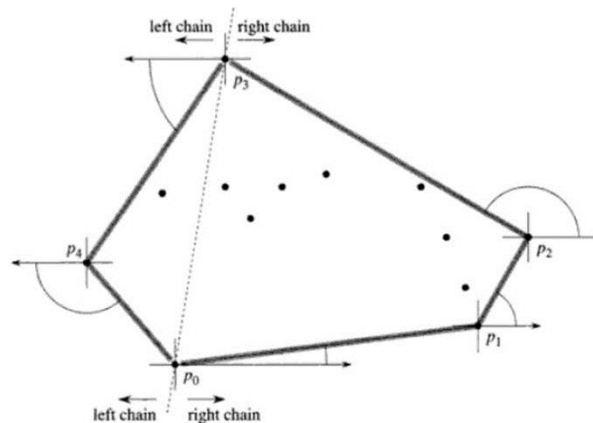
4. สำหรับแต่ละจุด p_i (ก) ถ้าเพิ่ม p_i ไปยัง convex hull แล้วทำให้เกิดการ "เลี้ยวซ้าย" ให้เพิ่ม p_i ลงไปใน H (ข) ถ้าการเพิ่ม p_i ไปยัง convex hull แล้วทำให้เกิดการ "เลี้ยวขวา" ให้เอา p_i ออกจาก H จนเพิ่ม p_i ที่ทำให้เกิดการเลี้ยวซ้ายแล้วจึงเพิ่มสมาชิกลงไปใน H

จากรูปที่ 3. 6 ในขั้นตอน (a) convex hull H คือ (P, A, B) ในขั้นตอน (b) เมื่อเพิ่ม C ถ้าเคลื่อนที่จาก A ไป C ทำให้เกิดการเลี้ยวซ้ายที่ B ดังนั้นจึงเพิ่ม C ลงไปใน H ซึ่งตอนนี้เป็น (P, A, B, C) ต่อมาเพิ่ม D ลง ถ้าเคลื่อนที่จาก B ไป D ทำให้เกิดการเลี้ยวขวาขึ้นที่ C ซึ่งหากเพิ่ม D ลงไปจะไม่ได้รูปร่างคอนเวกซ์ ดังนั้นจึงต้องกำจัดองค์ประกอบบางตัวใน H จนการเพิ่ม D ทำให้เกิดการเลี้ยวขวา ในขั้นตอน (c) ทำการกำจัด C แล้วเพิ่ม D ทำให้ได้ convex hull ที่มี $H = (P, A, B, D)$



รูปที่ 3. 6 ขั้นตอนการหาคอนเวกซ์ฮัลล์ด้วยวิธี GrahamScan

Jarvis' March กำหนดให้เซต S มีสมาชิก n จุดในระนาบ ขั้นตอนวิธีของ Jarvis' March สร้างขอบเขตของคอนเวกซ์ฮัลล์ โดย "marching around" ที่เส้นรอบนอกของ S วิธีการนี้จะเรียกว่า "gift - wrapping" [30] โดยเริ่มแรกจะสร้างอนุกรมจุดที่อยู่รอบ ๆ คอนเวกซ์ฮัลล์ $H = \langle p_0, p_1, \dots, p_{h-1} \rangle$ เริ่มต้นจาก p_0 ซึ่งเป็นจุดต่ำสุดตามแนวแกน y ถัดมาจุด p_1 คือจุดที่ทำให้เกิดมุมที่ p_0 ทำกับแนวแกน x น้อยที่สุด p_2 ก็ทำเช่นเดียวกันและทำเช่นนี้ต่อไปเรื่อยๆ เมื่อทำจนถึงจุดที่เป็นจุดสูงสุดตามแนวแกน y เรียก p_k ดังรูปที่ 3. 7 ณ ตอนนีเราได้สร้าง chain ทางด้านขวาเสร็จแล้ว ในการสร้าง chain ทางด้านซ้ายก็เช่นกัน คือเลือกมุมที่มีค่าน้อยที่สุดเทียบกับ p_k แต่เป็นทางด้านลบของแกน x ทำต่อไปเรื่อยๆจนกลับมาที่จุดเริ่มต้นของ p_0



รูปที่ 3. 7 ขั้นตอนการหาคอนเวกซ์ฮัลล์ด้วยวิธี Jarvis' March

3.2.3.2. ขั้นตอนวิธีแบ่งแยกและเอาชนะ (divide and conquer algorithms)

QuickHull [31] ขั้นตอนวิธีนี้มีขั้นตอนวิธีการทำงานค่อนข้างดี แต่การประมวลผลมักจะช้า โดยเป็นการทำซ้ำการสร้างห่วงโซ่ที่ขอบเขตของคอนเวกซ์ฮัลล์ โดยหาจุดที่อยู่ต่ำสุดและสูงสุดตามแนวแกน x ซึ่งจะผูกกันที่จะเป็นส่วนหนึ่งของคอนเวกซ์ฮัลล์ ใช้เส้นที่เกิดขึ้นจากสองจุดนั้นในการแบ่งเขตของจุดออกเป็นสองส่วนย่อยซึ่งจะเป็นการทำซ้ำ หาจุดที่ด้านหนึ่งของเส้นที่มีระยะทางสูงสุดจากเส้น โดยจุดสองจุดที่พบก่อนนี้ทำเป็นรูปแบบสามเหลี่ยม จุดอื่นที่อยู่ภายในของรูปสามเหลี่ยมที่ไม่สามารถเป็นส่วนหนึ่งของคอนเวกซ์ฮัลล์ และดังนั้นจึงสามารถละทิ้งและทำในขั้นตอนถัดไป ทำซ้ำสองขั้นตอนก่อนหน้าทั้งสองเส้นที่เกิดขึ้นจากรูปสามเหลี่ยม (ที่ไม่ใช่เส้นเริ่มต้น) ให้ทำไปเรื่อย ๆ จนกว่าไม่มีจุดเพิ่มมากกว่าจุดที่เหลืออยู่ จะเสร็จสิ้นการทำซ้ำและจุดที่เลือกก็จะเป็นคอนเวกซ์ฮัลล์

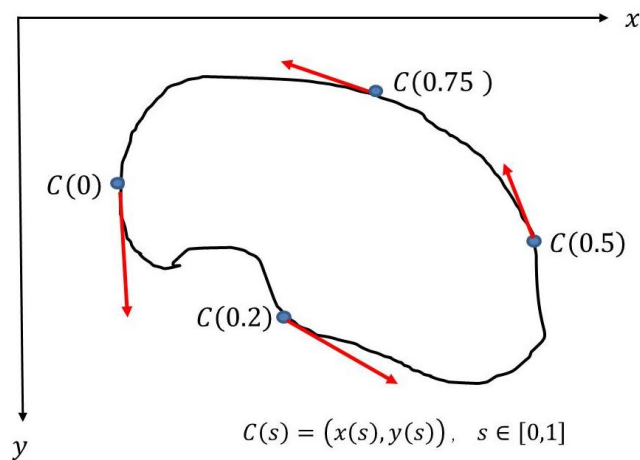
3.2.3.3. วิธีอื่นๆ

วิธีเคิร์กแพทริกไซเดล [32] เริ่มแรกจะทำการหา bridge edges ก่อน โดยการแบ่งจุดโดยการหาเส้นแบ่งแนวตั้งซึ่งกำหนดมาจากค่ามัธยฐานของพิกัดแกน x ของจุดทั้งหมด หลังจากนั้นทำการหาเส้นขอบของคอนเวกซ์ฮัลล์ที่ติดกับเส้นแบ่งนั้น หากจุดที่อยู่ด้านซ้ายหรือด้านขวาของเส้นไม่ช่วยในการหาเส้นขอบจะถูกตัดออก ซึ่งในการหาขอบบนของคอนเวกซ์ฮัลล์นั้นจุดที่ไม่ช่วยในการหาเส้นขอบและถูกตัดออก คือ จุดที่อยู่ใต้ bridge edge ลงมา ส่วนการหาขอบล่างของคอนเวกซ์ฮัลล์นั้นจุดที่อยู่เหนือ bridge edge ขึ้นไปจะถูกตัดออก จากนั้นหมุนไปยังจุดทางด้านซ้ายของเส้นที่ยังไม่ถูกตัดออก และจุดทางด้านขวาของเส้นที่ยังไม่ถูกตัดออก ทำเช่นนี้ไปเรื่อย ๆ จนได้ขอบของคอนเวกซ์ฮัลล์ครบทั้งหมด วิธีนี้จะต้องแยกดำเนินการหาขอบบน และขอบล่างของคอนเวกซ์ฮัลล์อีกครั้ง

3.3. ทฤษฎีและแนวคิดเกี่ยวกับแอกทีฟคอนทัวร์

แอกทีฟคอนทัวร์ (active contour) หรือ สเนก (snake) คือ เส้นโค้งที่กำหนดไว้ในโดเมนภาพที่สามารถเคลื่อนที่และเปลี่ยนรูปร่างได้ภายใต้อิทธิพลของแรงภายในที่ทำหน้าที่ควบคุมความราบเรียบของคอนทัวร์ และแรงภายนอกคอนทัวร์ทำหน้าที่เปลี่ยนแปลงรูปร่างและขับเคลื่อนคอนทัวร์ไปยังวัตถุที่สนใจในภาพ แอกทีฟคอนทัวร์มีใช้กันอย่างแพร่หลายในงานด้านต่างๆ จำนวนมาก อาทิเช่น การตรวจหาขอบของแบบจำลองรูปทรงต่างๆ การแบ่งส่วนและการติดตามการเคลื่อนไหวของวัตถุในภาพ การแบ่งส่วนภาพทางการแพทย์ รูปแบบแอกทีฟคอนทัวร์สามารถแบ่งได้เป็นสองประเภทกว้าง ๆ คือ parametric active contour และ geometric active contour

1. แบบ parametric active contour เป็นการแทนคอนทัวร์ด้วยฟังก์ชันที่เป็นพารามิเตอร์ลงบนคอนทัวร์โดยตรง คือ $C(s) = (x(s), y(s))$ โดย s คือ ความยาวอาร์ก (arc length) ดังรูปที่ ซึ่งเป็นรูปแบบอย่างง่าย ไม่ซับซ้อน แต่ไม่สามารถแยก (split) ตัวเองออกเป็นหลายคอนทัวร์ได้ หรือหากมีมากกว่า หนึ่งคอนทัวร์ก็ไม่สามารถรวม (merge) เป็นคอนทัวร์เดียวได้



รูปที่ 3.8 การแทนคอนทัวร์แบบ parametric

แอกทีฟคอนทัวร์ถูกนำเสนอเริ่มแรกโดย Kass และคณะ [33] โดยคอนทัวร์จะอยู่ในรูป $C(s) = (x(s), y(s))$ เมื่อ $s \in [0,1]$ คือ ความยาวอาร์ก แนวคิดหลักคือคอนทัวร์จะเคลื่อนที่ใน spatial domain ของภาพเพื่อทำฟังก์ชันพลังงานมีค่าน้อยที่สุด (energy function minimization) โดยฟังก์ชันพลังงานแสดงในสมการ

$$E_{snake} = \int_0^1 E_{snake}(C(s)) ds \quad (7)$$

$$E_{snake} = \int_0^1 [E_{internal}(C) + E_{external}(C)] ds \quad (8)$$

$$E_{internal} = \frac{1}{2} \left(\alpha \left| \frac{\partial C}{\partial s} \right|^2 + \beta \left| \frac{\partial^2 C}{\partial s^2} \right|^2 \right) \quad (9)$$

เมื่อ $E_{internal}$ คือ พลังงานภายในที่ได้มาจากพารามิเตอร์ของคอนทัวร์, คือ พารามิเตอร์ที่ควบคุมอนุพันธ์อันดับหนึ่งและสองของคอนทัวร์ และ $E_{external}$ คือ พลังงานภายนอกที่ได้มาจากภาพ เช่น ค่าเกรเดียนต์ของภาพ

การทำให้ฟังก์ชันพลังงานมีค่าน้อยที่สุด ทำได้โดยการหาอนุพันธ์ของฟังก์ชันพลังงานเทียบกับฟังก์ชัน C โดยใช้ calculus of variation

$$0 = \alpha \frac{\partial^2 C}{\partial s^2} + \beta \frac{\partial^4 C}{\partial s^4} + F_{external}(C) \quad (10)$$

แต่การหาคอนทัวร์ C จากสมการนี้โดยตรง ค่อนข้างยาก จึงใช้วิธีการไหลของเกรเดียนต์ (gradient flow) โดยแทน 0 ด้วย $\frac{\partial C}{\partial t}$ แล้วหา C ที่ทำให้ $\frac{\partial C}{\partial t}$ เข้าใกล้ศูนย์โดยวิธีการวนซ้ำ

$$\frac{\partial C}{\partial t} = \alpha \frac{\partial^2 C}{\partial s^2} + \beta \frac{\partial^4 C}{\partial s^4} + F_{external}(C) \quad (11)$$

เมื่อสองพจน์แรกทางขวา คือ แรงภายในคอนทัวร์ ($F_{internal}$) พจน์ที่สาม คือ แรงภายนอกคอนทัวร์ ($F_{external}$) ทำหน้าที่ขับเคลื่อนคอนทัวร์ให้เคลื่อนที่เปลี่ยนแปลงรูปร่างไปยังวัตถุที่สนใจ

แรงภายนอกคอนทัวร์ จะทำหน้าที่เปลี่ยนแปลงรูปร่างและขับเคลื่อนคอนทัวร์ไปยังวัตถุที่สนใจ โดยแรงภายนอกคอนทัวร์แบบดั้งเดิมจะเป็นแรงภายนอกที่ได้จากภาพขอบ ซึ่งสามารถคำนวณได้จาก

$$f(x, y) = |\nabla[G_\sigma(x, y) * I(x, y)]|^2 \quad (12)$$

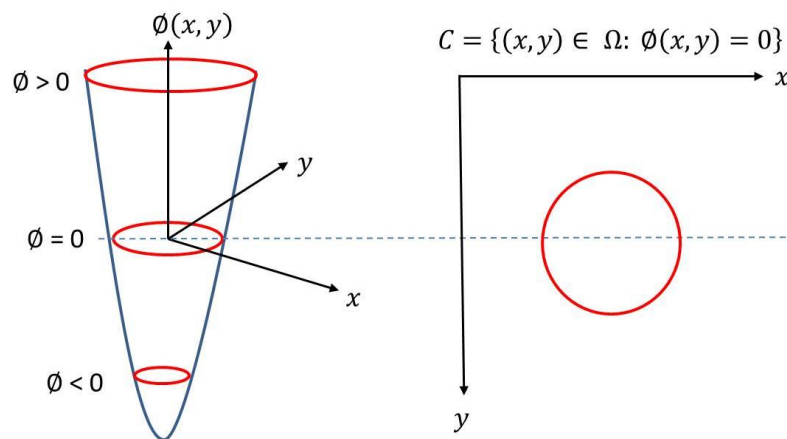
โดยที่ $I(x, y)$ คือภาพระดับขาวเทา, $G_\sigma(x, y)$ คือฟังก์ชันเกาส์เซียน 2 มิติที่มีค่าเบี่ยงเบนมาตรฐาน σ , ∇ คือตัวดำเนินการเกรเดียนต์ และ $*$ คือตัวดำเนินการคอนโวลูชัน

แต่อย่างไรก็ตามแรงภายนอกแบบดั้งเดิมมีข้อเสียหลัก ๆ คือ การเคลื่อนที่เข้าหาวัตถุที่มีส่วนโค้งเว้ามาก ๆ จะทำได้ไม่สมบูรณ์ และมีช่วงในการเคลื่อนที่เข้าหาวัตถุที่จำกัดจึงต้องวางคอนทัวร์

เริ่มต้นให้ใกล้กับวัตถุที่สนใจเพียงพอ จึงมีงานวิจัยที่เกี่ยวข้องกับการพัฒนาแรงภายนอกขึ้นมาอีกหลายรูปแบบ

2. แบบ geometric active contour models

แบบ geometric active contour models [34], [35] ใช้หลักการขยับเส้นโค้ง (curve evolution) และวิธีเซตระดับ (level set method) วิธีนี้เป็นการแทนคอนทัวร์โดยนัย คอนทัวร์จะถูกกำหนดให้เป็นเซตระดับศูนย์ (zero level set) ของฟังก์ชันสเกลาร์ (scalar function) $C = \{(x, y) \in \Omega: \phi(x, y) = 0\}$ เมื่อ C คือ คอนทัวร์, ϕ คือ ฟังก์ชันเซตระดับ และ Ω คือ spatial domain ของภาพ ซึ่งคอนทัวร์รูปแบบนี้สามารถเคลื่อนที่เข้าไปยังมุมปลายแหลม เปลี่ยนรูปร่างได้อัตโนมัติและยังสามารถแยกและรวมตัวกันได้



รูปที่ 3.9 การแทนคอนทัวร์แบบ geometric

เราสามารถแบ่งแอกทีฟคอนทัวร์ที่ใช้เซตระดับได้เป็นสองกลุ่มตามวิธีการลดฟังก์ชันพลังงาน คือ เซตระดับที่ใช้หลักการของขอบ (edge based active contour) และเซตระดับที่ใช้หลักการของอาณาบริเวณ (region based active contour)

3.3.1. เซตระดับที่ใช้หลักการของขอบ

Caselles และคณะ [35] และ Malladi และคณะ [36] เริ่มนำเซตระดับมาใช้ในการแบ่งส่วนภาพ ต่อมา Caselles [37] และ Kichenassamy และคณะ [38] ได้นำเสนอการนำเซตระดับมาใช้ในรูปแบบพลังงานของสเนก วิธีนี้เรียกว่า Geodesic Active Contours เนื่องจากพลังงานจะถูกแปลงผลเป็นความยาวของคอนทัวร์ใน Riemannian space ด้วยเมทริกซ์ที่เกิดจากความเข้มของภาพ

$$\begin{aligned}\frac{\partial \phi}{\partial t} &= |\nabla \phi| \operatorname{div} \left(g(I) \frac{\nabla \phi}{|\nabla \phi|} \right) \\ &= g(I) |\nabla \phi| \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) + \nabla g(I) \cdot \nabla \phi + \nu g(\nabla |I|)\end{aligned}\quad (13)$$

เมื่อ $\frac{\partial \phi}{\partial t}$ คือการขยับเซตระดับ, $g(I)$ คือฟังก์ชันขอบ, I คือภาพอินพุต

จากสมการ (13) พจน์แรกคือความโค้ง ส่วนพจน์กลางเป็นพจน์ที่ช่วยให้แน่ใจว่าคอนทัวร์เคลื่อนที่พุ่งตรงไปที่ขอบ และพจน์สุดท้ายคือ เงื่อนไขการหยุดที่ขอบของวัตถุ

$g(\nabla |I|)$ เป็นเงื่อนไขการหยุดที่ขอบของวัตถุ โดยจะมีความเร็วลดลงเมื่อเข้าไปใกล้ขอบของวัตถุโดย Caselles และคณะ [35] กำหนดไว้ดังแสดงในสมการที่ (14) ส่วน Malladi และคณะ [36] กำหนดไว้ดังแสดงในสมการที่ (15) ตามลำดับ

$$g(\nabla |I|) = \frac{1}{1 + |G_\sigma * \nabla I|} \quad (14)$$

เมื่อ $|G_\sigma * \nabla I|$ เป็นการคอนโวลูชันเกรเดียนต์ของภาพโดยใช้เกาส์เซียนเคเนล (Gaussian kernel) ที่มีส่วนเบี่ยงเบนมาตรฐาน σ

$$g(\nabla |I|) = e^{-\gamma |G_\sigma * \nabla I|} \quad (15)$$

เมื่อ γ เป็นค่าคงที่

อย่างไรก็ตามวิธีเซตระดับที่ใช้หลักการของขอบจะไม่เหมาะกับภาพที่ขอบไม่ชัดเจนหรือมีสัญญาณรบกวนมาก

3.3.2. เซตระดับที่ใช้หลักการของอาณาบริเวณ

การแบ่งส่วนภาพด้วยเซตระดับที่ใช้หลักการของอาณาบริเวณจะใช้สถิติของอาณาบริเวณ เช่น ระดับความเข้ม *texture* ค่าของสี ตัวอย่าง ค่าเฉลี่ยระดับความเข้มของอาณาบริเวณในวัตถุ (c_1) และพื้นหลัง (c_2) เป็นข้อมูลที่นิยมใช้ในการหาฟังก์ชันของพลังงาน F ซึ่งการ minimization เป็นการแบ่งส่วนวัตถุที่เหมาะสมที่สุด

$$\begin{aligned}F(c_1, c_2, \phi) &= \mu \cdot \text{Length}\{\phi = 0\} + \nu \cdot \text{Area}\{\phi \geq 0\} \\ &+ \int_{\phi \geq 0} |I(x, y) - c_1|^2 dx dy + \int_{\phi < 0} |I(x, y) - c_2|^2 dx dy\end{aligned}\quad (16)$$

เมื่อ ϕ คือ ฟังก์ชันเซตระดับศูนย์ โดยกำหนด ϕ เป็นบวกด้านในคอนทัวร์และเป็นลบด้านนอกคอนทัวร์, $I(x, y)$ คือ ค่าความเข้มของภาพที่ตำแหน่ง (x, y) , c_1 คือ ค่าเฉลี่ยความเข้มของภาพภายในคอนทัวร์, c_2 คือ ค่าเฉลี่ยความเข้มของภาพภายนอกคอนทัวร์, $\mu \geq 0, \nu \geq 0$, คือตัวแปรคงที่

การ minimize ฟังก์ชันพลังงานโดยการใช้ gradient descent มาตรฐานหรือ steepest descent จะได้สมการ gradient flow ดังนี้

$$\frac{\partial \phi}{\partial t} = \delta_e(\phi) \left[\mu \cdot \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \nu - (I - c_1)^2 + (I - c_2)^2 \right] \quad (17)$$

เมื่อ H เป็นฟังก์ชัน heaviside, δ_e เป็นฟังก์ชัน dirac - delta และ μ และ ν คือ regularize ความยาวเส้นโค้งและพื้นที่ของวัตถุตามลำดับ

ข้อเสียหลักในการใช้วิธีเซตระดับในการแบ่งส่วนภาพคือใช้หน่วยความจำมากและประมวลผลช้า ฟังก์ชันเซตระดับที่ตั้งไว้โดยปกติจะถูกทำให้ต่อเนื่องทั้งหมดทั่วตารางภาพเพื่อเก็บค่าทศนิยมแม้ว่าจะสนใจหลักเพียงแค่ตำแหน่งของเซตระดับศูนย์ของเส้นโค้ง ดังนั้นจึงมีงานวิจัยที่มุ่งเน้นพัฒนาวิธีในการลดการขึ้นตอนคำนวณโดยการคำนวณความเข้มของพื้นที่ท้องถิ่นแบบจำกัดเป็นแถบแคบ ๆ เรียกว่า narrow-band methods [39], [40], [41] วิธีการเหล่านี้ได้พัฒนาฟังก์ชันเซตระดับให้เป็นแถบอยู่รอบระดับศูนย์ที่ตั้งไว้เพียงอย่างเดียว การคำนวณจะเป็นลักษณะเดียวกันแต่สนใจเฉพาะบริเวณที่อยู่รอบ ๆ เส้นโค้งระดับศูนย์ วิธีที่ใช้ในปัจจุบันมักใช้วิธีจำกัดการคำนวณเป็นวงแถบแคบ ๆ เพื่อช่วยเพิ่มประสิทธิภาพความเร็วและหน่วยความจำที่ใช้งาน

ในการขยับฟังก์ชันเซตระดับ การเคลื่อนไหวที่แตกต่างกันของเส้นโค้งในระดับที่แตกต่างกันอาจทำให้เกิดฟังก์ชันในการปรับปรุงเกรเดียนต์ทั้งแบบที่ขึ้นเกินไปและหรือแบนเกินไป ส่งผลกระทบทางลบต่อความมีเสถียรภาพของตัวเลขใน subsequent iterations ดังนั้นจึงต้องมีรูปแบบการหยุดขึ้นตอนการทำ distance reinitialization ที่ปลายสุดของทุกรอบการทำซ้ำ ตำแหน่งเส้นโค้งระดับศูนย์ยังคงไม่เปลี่ยนแปลง วิธีที่นิยมมี รูปแบบการเริ่มต้นเพื่อขยับ PDE ไปยังช่วง steady state ตามวิธีของ Sussmanetal และคณะ [42]

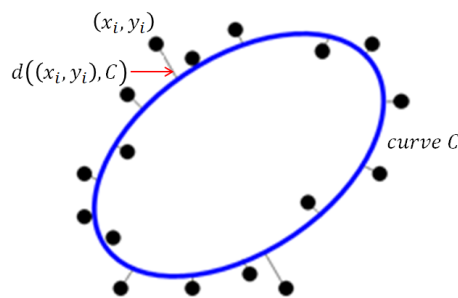
$$\frac{\partial \phi}{\partial t} + \operatorname{sign}(\phi_0)(|\nabla \phi| - 1) = 0 \quad (18)$$

ซึ่ง ϕ_0 คือ ฟังก์ชันเซตระดับก่อนที่จะ reinitialization ผลที่ได้จะเป็นฟังก์ชันระยะทางแบบมีเครื่องหมายเทียบกับกับอินเตอร์เฟซ ($\phi_0 = 0$)

แต่อย่างไรก็ตามวิธีนี้จำเป็นต้องทำการ re-initialization เพื่อให้ได้ค่า SDF กลับคืน เนื่องจากเซตระดับจะมีการมัดรวมกันอันเนื่องมาจากความผิดพลาดทางตัวเลข แม้ว่าจะมี normal extension velocities ก็ตาม นอกจากนี้ก็ยังมีปัญหาเรื่องเงื่อนไขการหยุด โดยฟังก์ชันของเซตระดับจะหยุดได้ขึ้นอยู่กับค่าจำนวนของรอบการทำซ้ำ และ / หรือโดยการประเมินจากการเปลี่ยนแปลงพลังงานที่ลดลง และ / หรือโดยการประเมินการเปลี่ยนแปลงในฟังก์ชันระดับที่ตั้งไว้ และ / หรือโดยการตรวจสอบเพื่อดูว่าระดับที่ตั้งไว้ได้ถึงบางอย่างจุดขอบที่กำหนดไว้ก่อนหน้า

3.4. การหาระยะห่างกำลังสองน้อยที่สุดสำหรับการปรับรูปวงรี

การปรับตัวแปรของวงรีที่ทำให้ระยะทางระหว่างจุดข้อมูลและวงรีที่ต้องการสร้างโดยทำการคำนวณหาผลรวมของระยะทางยกกำลังสองของระยะห่างระหว่างจุดข้อมูลและเส้นโค้งในแบบเรขาคณิต



รูปที่ 3. 10 แสดงจุดข้อมูล (x_i, y_i) และเส้นโค้งวงรี (C) และระยะห่างระหว่างจุดข้อมูลและเส้นโค้ง $d((x_i, y_i), C)$

เมื่อให้ p_1, p_2, \dots, p_n เป็นเซตของจุด n จุด โดย $p_i = (x_i, y_i)$, $i = 1, 2, \dots, n$

สมการภาคตัดกรวย มีสมการดังนี้

$$F(p, a) = Xa = ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (19)$$

$$\text{เมื่อ } a = [a \ b \ c \ d \ e \ f]^T$$

$$X = [x^2 \ xy \ y^2 \ x \ y \ 1]$$

$F(p, a)$ เรียกว่า “algebraic distance” ของจุด $p_i = [x_i \ y_i]$ ไปยังสมการภาคตัดกรวยที่

$F(p, a) = 0$ ถ้ามีจำนวนข้อมูล n จุด

$$\begin{bmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ x_2^2 & x_2 y_2 & y_2^2 & x_2 & y_2 & 1 \\ \vdots & & & & & \ddots \\ x_n^2 & x_n y_n & y_n^2 & x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix} = 0 \quad (20)$$

$$Da = 0$$

$$\text{เมื่อ } D = [X_1 \ X_2 \ \dots \ X_n]^T$$

เราสามารถลดระยะทางเรขาคณิตที่ไม่เป็นเชิงเส้นแม้จะเป็นรูปภาคตัดกรวยเช่นกัน โดยใช้

Algebraic distance: $F(p, a)$ สามารถลดระยะทาง โดยการทำให้ linear least squares

$$E(a^*) = \min_a \arg \sum_{i=1}^n F(p_i, a)^2 \quad (21)$$

$$E = \sum_{i=1}^n (ax_i^2 + bx_i y_i + cy_i^2 + dx_i + ey_i + f)^2 \quad (22)$$

เพื่อตัดปัญหาเรื่อง $a = 0_6$ งานวิจัยต่าง ๆ ใช้การกำหนดเงื่อนไขเวกเตอร์ a เช่น กำหนด $\|a\|^2 = 1$, $a + c = 1$, $f = 1$ ซึ่งเงื่อนไขเหล่านี้เป็นเชิงเส้นทุกตัว เขียนในรูปแบบ $Ca = 1$ หรืออยู่ในรูป quadratic $a^T C a = 1$ เมื่อ C เป็น constraint matrix ขนาด 6×6 Bookstein [43] แสดงให้เห็นว่า ถ้ากำหนดตัวแปร constraint เป็น quadratic สมการที่ (20) สามารถหาได้โดยการพิจารณาระบบ eigenvalue

$$D^T D a = \lambda C a \quad (23)$$

เมื่อ $D = [X_1 \ X_2 \ \dots \ X_n]^T$ คือเมทริกซ์ที่ออกแบบ และ C เป็นเมทริกซ์เงื่อนไข

เครื่องหมายของผลลัพธ์จาก $4ac - b^2$ ใช้ในการจำแนกชนิดของภาคตัดกรวยโดย

$$4ac - b^2 > 0 \text{ เป็นกรณีวงรีหรือวงกลม}$$

$$4ac - b^2 = 0 \text{ เป็นกรณีพาราโบลา}$$

$4ac - b^2 < 0$ เป็นกรณีไฮเปอร์โบลลา

Fitzgibbon, A. และคณะ [44] ได้นำเสนอวิธีใช้เทคนิค least square กับรูปวงรีโดยตรง ซึ่งใช้ข้อมูลจุดเพื่อสร้างวงรีที่ปรับให้เข้ากับชุดข้อมูล เพื่อลดระยะทางเชิงพีชคณิตให้มากที่สุด โดยกำหนด constraint เป็น $4ac - b^2 = 1$ ดังนั้น quadratic constraint สามารถเขียนในรูปเมทริกซ์ $a^T C a = 1$ จากสมการที่ (22) สามารถเขียนใหม่ได้เป็น

$$a^T \begin{bmatrix} 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} a = 1 = 4ac - b^2 \quad (24)$$

เมื่อทำตาม bookstein ปัญหาการปรับเงื่อนไขของวงรีก็ถูกลดลง ทำให้ได้สมการ (21) $E = \|Da\|^2$ ตามข้อจำกัด $a^T C a = 1$ เมื่อเมทริกซ์ที่ออกแบบ D ถูกกำหนดไว้ นำ Lagrange multiplier λ และ differentiateing ก็จะได้สมการ

$$E = \sum_{i=1}^n \|X_i^T a\|^2 = \|Da\|^2 = a^T D^T D a = a^T S a \quad (25)$$

เมื่อ S เป็น scatter matrix ของ $D^T D$

ในการแก้ปัญหา ellipse fitting สามารถเขียนสมการ $E(a^*) = \min_a \arg a^T S a$ ด้วยเงื่อนไข $a^T C a = 1$

จาก Lagrange multipliers

$$\min_{a, \lambda} L(a, \lambda) = a^T S a - \lambda(a^T C a - 1) \quad (26)$$

$$\frac{\partial L}{\partial a} = 2Sa - 2\lambda Ca = 0 \quad (27)$$

$$\frac{\partial L}{\partial \lambda} = a^T C a - 1 = 0 \quad (28)$$

จากสมการที่ (27), (28) เขียนใหม่เป็น

$$Sa = \lambda Ca \quad (29)$$

$$a^T C a = 1 \quad (30)$$

จากสมการที่ (27) และ (23)

$$E(a^*) = \min_a \arg a^T S a = a^T \lambda C a = \lambda$$

ซึ่งสามารถพิสูจน์ได้ว่าสามารถแก้ปัญหาได้ด้วย eigen vector ตามสมการ (29) ถ้า (λ_i, u_i) พิสูจน์สมการ (29) แล้วทำ (λ_i, μ_i) สำหรับ μ ใด ๆ และจากสมการ (30) เราสามารถหาค่า μ_i ที่ทำให้ $\mu_i^2 u_i^T C u_i = 1$ เมื่อให้

$$\mu_i = \sqrt{\frac{1}{u_i^T C u_i}} = \sqrt{\frac{1}{u_i^T S u_i}} \quad (31)$$

สุดท้าย ปรับ $\hat{a}_i = \mu_i u_i$ เพื่อแก้สมการ (30)

ผลลัพธ์จากระบบ eigen สมการที่ (29) ซึ่งให้ข้อมูล eigenvalue-eigenvector 6 คู่ (λ_i, u_i) โดยแต่ละคู่อันดับจะให้ค่าต่ำสุดถ้าผลลัพธ์ของภายใต้รากที่สองของสมการที่ (31) เป็นค่าบวก

โดยทั่วไป S เป็น positive definite ดังนั้น $u_i^T S u_i$ จะเป็นบวกสำหรับทุก u_i ดังนั้นถ้า $\lambda_i > 0$ การหาสมการที่ (29) จะต้องมีค่า eigen value ที่เป็นบวก

วิธีการนี้ ได้ใช้การกำหนดเงื่อนไขสำหรับสร้างวงรี (constraint ellipticity) โดยทำให้เป็น normalization factor วิธีที่นำเสนอมีข้อดีหลายประการ: คือ แม้ว่าข้อมูลจะไม่ดีนักแต่สามารถสร้างวงรีและสามารถแก้ไขได้โดย Eigen system มีความทน มีประสิทธิภาพ และทำได้ไม่ยาก

3.5. การหาความถูกต้องของการแบ่งส่วนภาพ

หาก $I(x, y)$ เป็นภาพจากกล้องจุลทรรศน์, $R(x, y)$ เป็นภาพลักษณะฐานสองที่โปรแกรมแบ่งส่วนวัตถุในภาพ และ $G(x, y)$ เป็นภาพลักษณะฐานสองที่ผู้เชี่ยวชาญแบ่งส่วนวัตถุในภาพ แต่ละพิกเซลสามารถจำแนกได้ดังนี้

$$\text{true positive: } G(x, y) = 1 \cap R(x, y) = 1$$

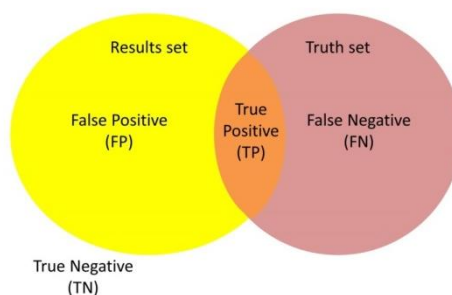
$$\text{false positive: } G(x, y) = 0 \cap R(x, y) = 1$$

$$\text{true negative: } G(x, y) = 0 \cap R(x, y) = 0$$

$$\text{false negative: } G(x, y) = 1 \cap R(x, y) = 0$$

จากข้อมูลนี้สามารถนำมาหาตัวแปรทางสถิติคือ จำนวนของ true positive (TP) จำนวนของ false positive (FP) จำนวนของ true negative (TN) และจำนวนของ false negative (FN) [45]

จากรูปที่ 3.11 พื้นที่ส่วนสีเหลือง (FP) และสีส้ม (TP) คือผลลัพธ์ของการแบ่งส่วนด้วยวิธีที่นำเสนอ และ พื้นที่ส่วนสีม่วง (FN) และสีส้ม (TP) คือผลลัพธ์ของการแบ่งส่วนด้วยผู้เชี่ยวชาญ



รูปที่ 3.11 วิธีการตรวจสอบหาความคล้ายคลึงของภาพด้วยวิธีทางสถิติบนพื้นฐานของการซ้อนทับของพื้นที่ภาพผลการแบ่งส่วนภาพสองชุด

Dice similarity coefficient คือ การหาปริมาณของพื้นที่ที่ทับซ้อนกันระหว่างวัตถุที่ทำการแบ่งส่วนกับพื้นที่ที่ผู้เชี่ยวชาญแบ่งส่วนว่าเป็นวัตถุ สมการที่ใช้ในการคำนวณ คือ สมการที่ (32)

$$\text{Dice} = \frac{2TP}{2TP+FP+FN} \quad (32)$$

false positive ratio คือ ค่าอัตราส่วนของพื้นที่ที่โปรแกรมแบ่งส่วนว่าเป็นวัตถุที่ต้องการแต่ผู้เชี่ยวชาญแบ่งส่วนว่าเป็นพื้นหลัง เปรียบเทียบกับพื้นที่ที่ผู้เชี่ยวชาญจำแนกเป็นพื้นหลังทั้งหมด สมการที่ใช้ในการคำนวณ คือ สมการที่ (33)

$$\text{false positive ratio} = \frac{FP}{TN + FP} \quad (33)$$

false negative ratio คือ ค่าอัตราส่วนของพื้นที่ที่โปรแกรมแบ่งส่วนว่าเป็นพื้นหลัง แต่ผู้เชี่ยวชาญแบ่งส่วนว่าเป็นวัตถุที่ต้องการเทียบกับพื้นที่ที่ผู้เชี่ยวชาญจำแนกเป็นวัตถุที่ต้องการเทียบ สมการที่ใช้ในการคำนวณ คือ สมการที่ (34)

$$\text{false negative ratio} = \frac{FN}{TP+FN} \quad (34)$$

3.6. การออกแบบตัวจำแนกรูปแบบโดยใช้ค่าทางสถิติ

การออกแบบ statistical visual pattern classifier [46] ประกอบด้วย ขั้นตอนดังต่อไปนี้

1. กำหนดปัญหาและหาจำนวนกลุ่มข้อมูลที่เกี่ยวข้อง ซึ่งจะต้องตอบคำถามต่อไปนี้ได้ คือ จำนวนกลุ่มที่มีอธิบายอะไรได้บ้าง กลุ่มข้อมูลเหล่านี้สามารถใช้อธิบายวัตถุหรือภาพได้ดีอย่างไร วิธีที่ดีที่สุดที่ใช้อธิบายถึงกลุ่มเหล่านี้
2. หาลักษณะบ่งต่างที่สกัดออกมามีความเหมาะสมมากที่สุดที่จะอธิบายภาพและใช้เป็นตัวจำแนกได้ การเลือกประเภทของลักษณะบ่งต่าง และหาวิธีการเฉพาะ ที่จะดึงข้อมูลเหล่านี้มา
3. การเลือกวิธีการจำแนกหรือขั้นตอนวิธีที่ใช้ มีเครื่องมือมากมายในด้านการ machine learning ที่สามารถเลือกให้เหมาะสมกับความต้องการ การเลือกใช้ขึ้นอยู่กับ ความซับซ้อน เวลาที่ใช้คำนวณ, ความสามารถในการฝึก, และอื่น ๆ
4. การเลือกชุดข้อมูล ในขั้นตอนนี้ก็คือการเก็บข้อมูลจากภาพที่สามารถใช้ในการฝึกและการทดสอบ
5. การเลือกข้อมูลที่จะนำมาฝึกการจำแนก วิธี pattern classification หลาย ๆ วิธีจะต้องมีขั้นตอนของการฝึกโดยใช้ชุดข้อมูลย่อยของภาพ มาใช้ฝึกการจำแนกเกี่ยวกับกลุ่มที่ข้อมูลนั้นควรจะเป็น เพื่อให้สามารถแยกแยะได้ และสามารถปรับตัวแปรบางตัวได้
6. ทดสอบการจำแนก ขั้นตอนนี้จะเป็นการวัดอัตราส่วนความสำเร็จหรือความล้มเหลว คำนวณจากสมรรถนะของระบบ (figure of merit) (เช่น precision และ recall), และรวบรวมตัวเลขจากการคำนวณมาเขียนเป็นกราฟ
7. ปรับปรุงพัฒนา หลังจากการวิเคราะห์ผลที่คำนวณในขั้นตอนที่ 6 จำเป็นต้องกลับไปขั้นตอนก่อนหน้า เพื่อเปลี่ยนตัวแปร เช่น การเลือกลักษณะบ่งต่างที่แตกต่างออกไป การเก็บภาพเพิ่มเติม และทดสอบวิธีที่ปรับปรุงขึ้น จนกระทั่งได้ผลลัพธ์ที่ดีที่สุด
8. ในวิธีการตรวจสอบสมรรถนะการทำงานของการทำงานของการจำแนก [47] และลักษณะบ่งต่างที่เลือกมาใช้ โดยจะเปรียบเทียบขั้นตอนวิธีการจัดหมวดหมู่ที่แตกต่างกันและศึกษาข้อดีของแต่ละวิธีดำเนินการที่ใช้กับข้อมูล สิ่งที่จะต้องสนใจมีสองส่วนใหญ่ ๆ คือ การทดสอบ (testing) และการให้คะแนน (scoring)

3.7. ลักษณะบ่งต่าง

การสกัดลักษณะบ่งต่าง (feature extraction) [48] มีวัตถุประสงค์เพื่อหาตัวแทนที่อธิบายลักษณะบ่งต่างของข้อมูลภาพ โดยแบ่งสามารถแบ่งลักษณะบ่งต่างออกเป็น

3.7.1. ลักษณะบ่งต่างที่ได้จากภาพลักษณะฐานสอง

มีข้อมูลหลายชนิดที่สามารถสกัดจากลักษณะบ่งต่างภาพของภาพลักษณะฐานสอง [46], [49] หากกำหนดให้ภาพลักษณะฐานสองคือ $I(x, y)$

วัตถุที่สนใจตัวที่ i ซึ่งกำหนดให้เป็น $O_i, i > 0$ คือ พื้นที่เชื่อมต่อถึงกันในภาพลักษณะฐานสองฟังก์ชัน โดย $O_i(x, y)$ สามารถแสดงได้ดังสมการ (35)

$$O_i(x, y) = \begin{cases} 1 & \text{if } I(x, y) \in O_i \\ 0 & \text{otherwise} \end{cases} \quad (35)$$

– พื้นที่ คือ พื้นที่ที่วัตถุตัวที่ i ของ O_i , โดยวัดในหน่วยพิกเซล มีสมการดังนี้

$$A_i = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} O_i(x, y) \quad (36)$$

– centroid พิกัดของ centroid ของวัตถุ O_i แสดงด้วย (\bar{x}_i, \bar{y}_i) สามารถคำนวณได้จาก

$$\bar{x}_i = \frac{1}{A_i} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x O_i(x, y) \quad (37)$$

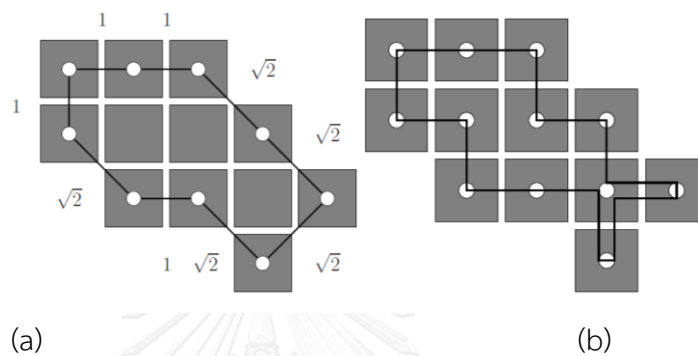
$$\bar{y}_i = \frac{1}{A_i} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} y O_i(x, y) \quad (38)$$

เมื่อ A_i คือพื้นที่ของวัตถุ O_i

– เส้นรอบรูป (perimeter) คือ ความยาวรวมของขอบของวัตถุ โดยขอบของวัตถุคือชุดของพิกเซลขอบทุกตัว เส้นรอบรูปสามารถวัดได้จากการติดตามขอบของวัตถุด้วย chain code (8 connected component) และ รวมทุกชิ้นด้วยความยาว 1 หรือ $\sqrt{2}$ ตามรูปที่ 3.12 (a) มีวิธีที่เร็วขึ้นคือไม่คิดแนวทแยงและติดตามขอบโดยพิจารณาเพียงทิศขึ้น ลง ซ้าย หรือขวาตามรูปที่ 3.12 (b) แต่วิธีนี้ได้ผลลัพธ์แตกต่างกัน

$$P_i = \sum_{i=0}^{M-1} \text{length}(C_i) \quad (39)$$

$$\text{เมื่อ } \text{length}(C) = \begin{cases} 1 & \text{for } C = 0,2,4,6 \\ \sqrt{2} & \text{for } C = 1,3,5,7 \end{cases}$$



รูปที่ 3.12 การหาเส้นรอบรูป

เส้นรอบรูปของวัตถุ O_i สามารถคำนวณโดยการนับจำนวนของพิกเซลของวัตถุ (พิกเซลที่มีค่าเป็น 1) ที่มีพิกเซลของฉากหลัง (พิกเซลที่มีค่าเป็น 0) ล้อมรอบอย่างน้อย 1 พิกเซล วิธีนี้มีขั้นตอนคือ ติดตามขอบหรือคอนทัวร์ของวัตถุและนับจำนวนของพิกเซลจากภาพขอบ เนื่องจากความไม่สมบูรณ์ในขั้นตอนการแปลงเป็นดิจิทัล (เช่นเส้นโค้งหยักและขอบหยัก) เป็นเรื่องที่ไม่สามารถหลีกเลี่ยงได้ การคำนวณเส้นรอบรูปโดยใช้วิธีการอย่างใดอย่างหนึ่งไม่ถูกต้องได้ 100% Umbaugh [50] ได้แนะนำให้นำจำนวนพิกเซลรวมคูณด้วย $\pi/4$ เพื่อความถูกต้องดีขึ้น

- อัตราส่วนพอมหรือความกลม (T_i) อัตราส่วนพอมหรือความกลมของวัตถุ O_i สามารถหาได้จาก

$$T_i = \frac{4 \pi A_i}{P_i^2} \quad (40)$$

เมื่อ A_i คือพื้นที่ และ P_i คือเส้นรอบรูปของวัตถุ

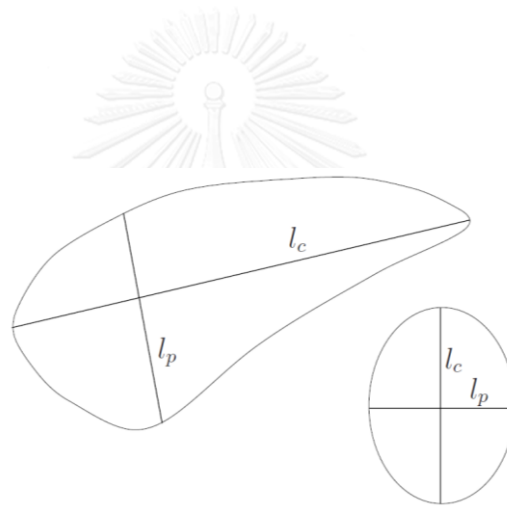
- irregularity หรือ compactness ratio: สามารถหาได้จาก $1/T_i$
- eccentricity: หรือบางครั้งเรียกว่า elongation ค่า eccentricity ของวัตถุมาจากอัตราส่วนของ l_c และ l_p ของวัตถุ โดย l_p คือ ความยาวสูงสุดของคอร์คทั้งหมดที่อยู่ในแนวตั้งฉากกับคอร์คที่ยาวที่สุด l_c คือ ความยาวสูงสุดของคอร์คที่ยาวที่สุด โดย คอร์คที่ยาวที่สุด คือเส้นตรงที่

เชื่อมระหว่างพิกเซลสองพิกเซลบนขอบของวัตถุ ในรูปทรงหลาย ๆ แบบขนาดและการวางแนวของคอร์คจะให้ข้อมูลเกี่ยวกับขนาดและการจัดวางแนวของวัตถุ ถ้าคอร์คอยู่ระหว่างพิกเซลขอบ (x_1, y_1) และ (x_2, y_2) ความยาวของ l_c และการวางแนวสามารถหาได้จาก

$$l_c = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (41)$$

$$\tan \alpha = \frac{y_2 - y_1}{x_2 - x_1} \quad (42)$$

เมื่อ α คือมุมของคอร์คกับแกน x ด้านบวก คอร์คที่ยาวที่สุดสามารถหาจากจุดบนขอบที่เป็นไปได้ และเปรียบเทียบระยะทาง



รูปที่ 3.13 วัตถุสองก้อนที่มีคอร์คที่ยาวที่สุด (l_c) และคอร์คที่มีความยาวมากที่สุดที่อยู่ในแนวตั้งฉากกับคอร์คที่ยาวที่สุด (l_p) โดยวัตถุทางซ้ายมีค่า eccentricity สูง วัตถุทางขวามีค่า eccentricity ต่ำ

3.7.2. ลักษณะบ่งต่างของแผนภูมิแจกแจงความถี่ของระดับความเข้มของภาพ

- แผนภูมิแจกแจงความถี่ของระดับความเข้มของภาพหรือฮิสโทแกรมในภาพดิจิทัลเป็นระดับเทาในช่วง $[0, L - 1]$ เป็นฟังก์ชันดิสครีต $h(r_j)$ โดยที่ r_j คือระดับเทาตัวที่ j และ n_j คือจำนวนของพิกเซลในภาพที่มีระดับเทาเท่ากับ r_j ในทางปฏิบัติจะทำการนอร์มอลไลซ์ฮิสโทแกรม โดยทำการหารแต่ละค่าด้วยจำนวนพิกเซลรวม n โดยเขียนอยู่ในรูป

$$p(r_j) = \frac{n_j}{n} \quad \text{for } j = 0, 1, 2 \dots L - 1 \quad (43)$$

$p(r_j)$ ใช้ประมาณความน่าจะเป็นในการเกิดระดับเทาตัวที่ r_j

- ค่าเฉลี่ยสีเทาของภาพ

$$m = \sum_{j=0}^{L-1} r_j p(r_j) \quad (44)$$

เมื่อ r_j เป็นระดับสีเทาตัวที่ j^{th} ซึ่งมีความน่าจะเป็นของการเกิดขึ้นคือ $p(r_j)$

- ความแปรปรวน

$$\sigma^2 = \sum_{j=0}^{L-1} (r_j - m)^2 p(r_j) \quad (45)$$

- พลังงาน (energy) คือค่าที่บ่งบอกถึงค่าพิกเซลจะกระจายอยู่ตามช่วงระดับสีเทาว่ามีลักษณะอย่างไร ภาพที่มีค่าคงที่จะมีพลังงานสูงสุด (เช่นพลังงาน = 1) ภาพที่มีระดับสีเทาไม่ก็กลุ่มจะมีพลังงานสูงกว่าภาพที่มีระดับสีเทาจำนวนมาก

$$energy = \sum_{j=0}^{L-1} [p(r_j)]^2 \quad (46)$$

- เอนโทรปี (entropy) จากข้อมูลแผนภูมิฮิสโทแกรมให้ข้อมูลเกี่ยวกับความซับซ้อนของภาพในรูปแบบของการอธิบายด้วยเอนโทรปี ภาพที่ซับซ้อนมากกว่าจะได้เอนโทรปีสูงกว่า เอนโทรปีและพลังงานมีแนวโน้มที่จะแปรผกผันกัน สมการทางคณิตศาสตร์เอนโทรปีคือ

$$entropy = - \sum_{j=0}^{L-1} p(r_j) \log_2 [p(r_j)] \quad (47)$$

- ความหายาบ (R) เป็นตัวบอกความราบเรียบของพื้นผิว สามารถคำนวณหาได้จาก

$$R = 1 - \frac{1}{1+\sigma^2} \quad (48)$$

เมื่อ σ^2 เป็นความแปรปรวนที่นำมาใช้ نرمอลโลซ์ ให้ค่าความหายาบอยู่ในช่วง $[0, 1]$ สำหรับพื้นที่ที่ค่าระดับความเข้ม (intensity) คงที่หรือเป็นเนื้อเรียบเนียน จะได้ความหายาบ $R = 0$

3.8. การเลือกลักษณะบ่งต่าง

การเลือกลักษณะบ่งต่าง (feature selection) เป็นเทคนิคการเพิ่มประสิทธิภาพเพื่อช่วยลดมิติของข้อมูลได้โดยการเลือกเฉพาะชุดย่อยของลักษณะบ่งต่างที่วัด (ตัวแปร) เพื่อสร้างแบบจำลองเป้าหมายของการสร้างแบบจำลองคือการบ่งชี้กลุ่มย่อยที่มีอิทธิพล วิธีการเลือกลักษณะบ่งต่างมีความสำคัญต่อระบบการรับรู้และการจัดหมวดหมู่เพราะหากลักษณะบ่งต่างมีมิติขนาดใหญ่

สมรรถนะการทำงานของตัวจำแนกจะลดลง ซึ่งเกี่ยวเนื่องเวลาในการประมวลผลและอัตราการรู้จำ การประมวลผลเพิ่มขึ้นตามจำนวนของลักษณะบ่งต่างเนื่องจากต้องใช้เวลาการวัดเพิ่มขึ้น อัตราการรู้จำอาจจะลดลงเพราะลักษณะบ่งต่างมีความซ้ำซ้อน และแม้ว่าลักษณะบ่งต่างจำนวนน้อยกว่าสามารถลดของมิติของเวลา แต่หากตัวอย่างที่ใช้ฝึกมีจำกัด ที่นำไปสู่ overtraining ในขณะเดียวกัน การลดจำนวนของลักษณะบ่งต่างอาจนำไปสู่การสูญเสียความสามารถในการจำแนกและลดความถูกต้องของระบบการรับรู้ได้

การเลือกลักษณะบ่งต่างช่วยลดมิติของข้อมูลได้โดยทำการเลือกเฉพาะชุดย่อยของลักษณะบ่งต่างที่จะใช้วัด (ตัวแปร ที่ใช้ในการทำนาย) เพื่อสร้างแบบจำลอง เกณฑ์การคัดเลือกจะเกี่ยวข้องกับ การลดความผิดพลาดในการทำนายสำหรับแบบจำลองที่มีกลุ่มย่อยที่แตกต่างกัน ในการลดขนาดของ ข้อมูลโดยการหาชุดย่อยของลักษณะบ่งต่างที่สำคัญที่สามารถให้ผลการจำแนกประเภทที่ดี มี ขั้นตอนวิธีการเลือกลักษณะบ่งต่าง ซึ่งสามารถแบ่งออกเป็นสองประเภท คือวิธี filter และวิธี wrapper วิธี filter จะอาศัยลักษณะทั่วไปของข้อมูลในการประเมินและเลือกชุดลักษณะบ่งต่างย่อยที่ไม่เกี่ยวข้องกับวิธีการเรียนรู้ที่เลือก ส่วนวิธี wrapper จะใช้ประสิทธิภาพของวิธีการเรียนรู้ที่เลือกในการประเมินแต่ละลักษณะบ่งต่าง วิธี wrapper จะค้นหาลักษณะบ่งต่างที่เหมาะสมได้ดีกว่า สำหรับวิธีการเรียนรู้ที่เลือก แต่จะใช้เวลาานานกว่าวิธี filter หากวิธีการเรียนรู้ที่เลือกใช้เวลาในการ ทำงานนาน

วิธีการทั่วไปในการเลือกลักษณะบ่งต่าง คือ การเลือกลักษณะบ่งต่างแบบเป็นลำดับ วิธีการนี้มีสององค์ประกอบ คือ ฟังก์ชันวัตถุประสงค์ (criterion function) และ วิธีการค้นหาลำดับ

1. ฟังก์ชันวัตถุประสงค์ ซึ่งวิธีการที่พยายามที่จะลดกลุ่มย่อยของลักษณะบ่งต่างทุกตัวที่เป็นไปได้ เกณฑ์ส่วนใหญ่จะเป็นความผิดพลาดยกกำลังสองเฉลี่ย (สำหรับ regression model) และอัตราการจำแนก (สำหรับแบบจำลองการจำแนก (classification model))

2. วิธีการค้นหาลำดับซึ่งจะเป็นเพิ่มหรือกำจัดลักษณะบ่งต่างจาก candidate subset ในขณะที่ทำการประเมินเกณฑ์ ตั้งแต่ทำการเปรียบเทียบค่าเกณฑ์ย่อยที่ทุก ๆ 2^n subset ของ ชุดข้อมูล ลักษณะบ่งต่าง n ชุด ที่โดยปรกติมักจะเป็นไปไม่ได้ (ขึ้นอยู่กับขนาดของ n และ cost of objective calls), การค้นหาลำดับจะเคลื่อนที่ไปในทิศทางเดียว โดยมักจะเติบโตหรือหดตัวตาม candidate set

วิธีการค้นหาลำดับมีสองกลุ่ม คือ sequential forward selection (SFS) และ sequential backward selection (SBS)

1. sequential forward selection (SFS) เป็นการเพิ่มลักษณะบ่งต่างเข้ามาตามลำดับจาก candidate set ที่ว่างเปล่า จนกระทั่งการเพิ่มลักษณะบ่งต่างต่อไปไม่ทำให้เกิดผลลัพธ์ลดลง
2. sequential backward selection (SBS) เป็นการลบลักษณะบ่งต่างออกจากการตั้งค่าตามลำดับจาก candidate set ทั้งหมด จนกระทั่งการลบลักษณะบ่งต่างต่อไปไม่ทำให้เกณฑ์เพิ่มขึ้น

3.9. discriminant function

การจำแนกด้วยแบบจำลอง discriminant function โดยกลุ่มแต่ละกลุ่ม เป็น y ซึ่งมีข้อมูล x ซึ่งมีการกระจายตัวแบบ multivariate normal distribution โดยแบบจำลองจะสมมุติว่า x มีการกระจายตัวแบบ Gaussian mixture distribution

- linear discriminant function มี covariance matrix ของแต่ละกลุ่มเหมือนกัน แต่ค่าเฉลี่ยแตกต่างกัน
- quadratic discriminant function ทั้งค่าเฉลี่ยและ covariance ของแต่ละกลุ่ม แตกต่างกัน

discriminant function [51] เป็นการวิเคราะห์การจำแนกกลุ่มด้วยวิธีการทางสถิติ ที่ใช้วิเคราะห์สามารถจำแนกกลุ่มตั้งแต่ 2 กลุ่มขึ้นไป เมื่อให้ลักษณะบ่งต่างคือ x กลุ่มที่ต้องการจำแนกคือ y

เซตของตัวอย่างจะเรียกว่า เซตที่นำมาฝึก ในการทำการจำแนกจะเป็นการหาตัวทำนายว่าเป็นกลุ่มใดของ y เมื่อให้ข้อมูลเป็นการกระจายลักษณะเดียวกัน

สมมติว่าเรามีลักษณะบ่งต่างของตัวอย่าง x ซึ่งมีจำนวนข้อมูล N_1 ซึ่งอยู่ในกลุ่ม ω_1 และ N_2 ซึ่งอยู่ในกลุ่ม ω_2 สามารถเขียน linear discriminant function ในรูปแบบสมการ (49)

$$\begin{aligned}
 g(x) &= w^T x + w_0 & (49) \\
 g(x) &> 0 & x \in \omega_1 \\
 g(x) &< 0 & x \in \omega_2
 \end{aligned}$$

เมื่อ w คือเวกเตอร์ถ่วงน้ำหนัก และ w_0 คือค่า threshold weight หรือ bias

และสามารถเขียน quadratic discriminant function ในรูปแบบสมการ (50)

$$g_i(x) = x^T W_{2,i} x + w_{1,i}^T x + w_{0,i} \quad (50)$$

$$\text{เมื่อ } \begin{cases} W_{2,i} = -\frac{1}{2} \Sigma_i^{-1} \\ W_{1,i} = \Sigma_i^{-1} \mu_i \\ W_{0,i} = -\frac{1}{2} \mu_i^T \Sigma_i^{-1} \mu_i - \frac{1}{2} \log |\Sigma_i| + \log P_i \end{cases}$$

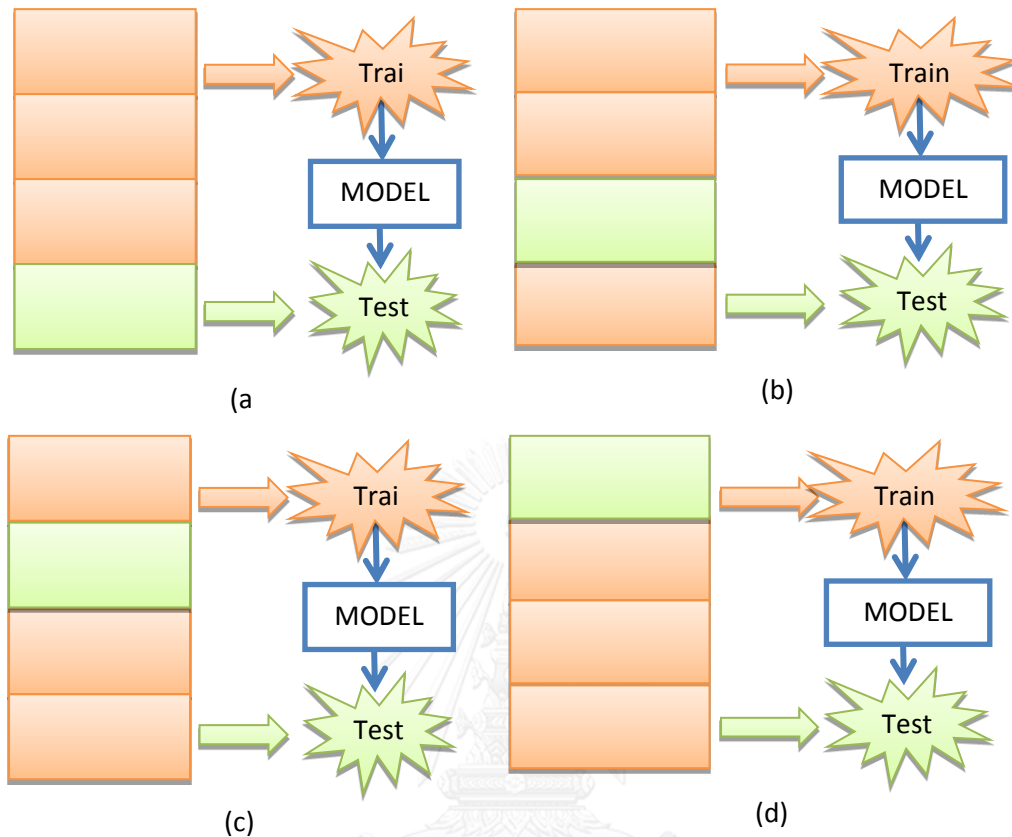
equiprobable contours จะเป็น hyper-ellipses วางตัวตาม eigenvector ของ Σ_i สำหรับแต่ละกลุ่ม และ ขอบเขตการตัดสินใจจะมีลักษณะตาม quadratic คือเป็น hyper-ellipses หรือ hyperparaboloids

3.10. วิธีการตรวจสอบสมรรถนะการทำงานของตัวจำแนก

3.10.1. การทดสอบ

จุดมุ่งหมายของขั้นตอนวิธีการตรวจสอบสมรรถนะการทำงานของตัวจำแนก คือ การนำแบบจำลองไปทดสอบกับข้อมูลที่ใช้ทดสอบ (testing dataset) หลังจากที่ได้ฝึกแบบจำลองด้วยชุดข้อมูลเพื่อฝึกอบรม (training dataset) แล้ว วิธีแรก คือแบ่งข้อมูลออกเป็น 2 กลุ่ม ตัวอย่างเช่น 80:20 ในกรณีนี้ 80% ของข้อมูลจะถูกเลือกสำหรับการฝึกอบรมและข้อมูลที่เหลืออีก 20% จะถูกใช้ในการทดสอบแบบจำลอง การทดสอบซ้ำแล้ว 50 ครั้ง

วิธีที่สองในการทดสอบแบบจำลองคือการใช้เทคนิค cross validation เทคนิคนี้มาจากแนวคิดในการแบ่งข้อมูลออกเป็น k กลุ่ม ซึ่งเท่ากับขนาดของ fold ตัวอย่างเช่น ถ้าต้องการแบ่งข้อมูลเป็น 4 folds คือ A, B, C และ D ในขั้นต้น จะใช้ A, B, C ในการฝึก และใช้ D ในการทดสอบ จากนั้นเราจะใช้ A, B, และ D สำหรับการฝึกและ C สำหรับการทดสอบ จากนั้นเราจะใช้ A, C, และ D สำหรับการฝึกและ B สำหรับการทดสอบ จากนั้นเราจะใช้ B, C และ D สำหรับการฝึกและ A สำหรับการทดสอบตามลำดับดังรูปที่ 3.14 วิธีหนึ่งของ cross validation คือ leave-one-out cross validation technique (LOOCV) วิธีการนี้จะทำให้การใช้จำนวนของกรณีตัวอย่างในชุดข้อมูลเป็นค่า k ใช้ข้อมูลชุดเพียงชุดเดียวในการทดสอบและข้อมูลที่เหลือเป็นข้อมูลสำหรับการฝึกอบรม ตัวอย่างเช่น ถ้ามีข้อมูล 100 รายการ จะต้องแบ่งข้อมูลออกเป็น 100 ส่วน โดย 99 ส่วนสำหรับการฝึกและ 1 ส่วนสำหรับการทดสอบ กระบวนการนี้ ทำซ้ำแล้ว 100 ครั้งโดยสลับข้อมูลที่ใช้ทดสอบไปที่ละตัวจนครบทุกตัว ส่วนข้อมูลที่ใช้ในการฝึกคือข้อมูลส่วนที่เหลือ



รูปที่ 3.14 ตัวอย่างการใช้ cross validation เพื่อทดสอบแบบจำลอง สำหรับ 4 fold

3.10.2. การให้คะแนน

เมทริกซ์ความสับสน (confusion matrix) เป็นเมทริกซ์สองมิติขนาด $k \times k$ เมื่อ k คือจำนวนกลุ่มรวม เมทริกซ์ความสับสนใช้สำหรับการรายงานผลการทดสอบการจำแนก ค่าในแถวที่ i หลักที่ j จะเป็นตัวบ่งชี้ถึงจำนวนครั้งที่ข้อมูลในกลุ่มที่ถูกต้อง i จำแนกไปที่กลุ่ม j โดยแนวทางการหลักของเมทริกซ์ความสับสนจะบ่งบอกถึงจำนวนของกลุ่มที่ตัวจำแนกทำงานได้ถูกต้อง เมทริกซ์ความสับสนถูกนำมาใช้ในการประเมินรูปแบบการจัดกลุ่ม หากพิจารณาปัญหาการจัดกลุ่มเป็นสองชั้น คือเชิงบวกและเชิงลบ โดยข้อมูลในแต่ละครั้งจะถูกจัดกลุ่มไปยังทั้งเชิงบวกหรือเชิงลบ

1. true Positive (TP): ถ้าค่าจริงเป็นบวกและค่าที่จัดกลุ่มเป็นบวก
2. false Negative (FN): ถ้าค่าจริงเป็นบวก แต่ค่าที่จัดกลุ่มเป็นลบ
3. true Negative (TN): ถ้าค่าจริงเป็นลบและค่าที่จัดกลุ่มเป็นลบ
4. false Positive (FP): ถ้าค่าจริงเป็นลบ แต่ค่าที่จัดกลุ่มเป็นบวก

ผลที่ได้จากการทดสอบการจัดหมวดหมู่นั้นจะสามารถสรุปได้ในเมทริกซ์ความสับสน (confusion matrix) ดังแสดงในตารางที่ 3. 2

ตารางที่ 3. 2 แสดงรูปแบบเมทริกซ์ความสับสน (confusion matrix)

| | P'(Predicted) | N'(Predicted) |
|-----------|----------------|----------------|
| P(Actual) | true Positive | false Negative |
| N(Actual) | false Positive | true Negative |

การให้คะแนนเป็นการทดสอบความถูกต้องของการจำแนกกลุ่ม การให้คะแนนสามารถคำนวณได้จากค่าใน เมทริกซ์ความสับสน (confusion matrix) ดังนี้

- ค่าความถูกต้อง (accuracy) เป็นการให้คะแนนแบบง่ายที่สุดที่คำนวณสัดส่วนความถูกต้องของการแบ่งกลุ่ม มีสมการดังนี้

$$\text{accuracy} = (TP + TN) / (TP+TN+FP+FN) \quad (51)$$

- ค่าความไว (sensitivity หรือ Recall หรือ true Positive Rate) ค่าความไวจะเป็นสัดส่วนของค่าบวกจริงที่บ่งชี้ได้ถูกต้องด้วย classifier

$$\text{sensitivity} = TP / (TP + FN) \quad (52)$$

- specificity (หรือ true Negative Rate) จะสัมพันธ์กับความสามารถของ classifier's ในการระบุผลลัพธ์ที่เป็นลบ โดยมีสมการดังนี้

$$\text{specificity} = TN / (TN+FP) \quad (53)$$

- precision เป็นตัวชี้วัดในกรณีที่มีการเรียกข้อมูลที่เกี่ยวข้อง

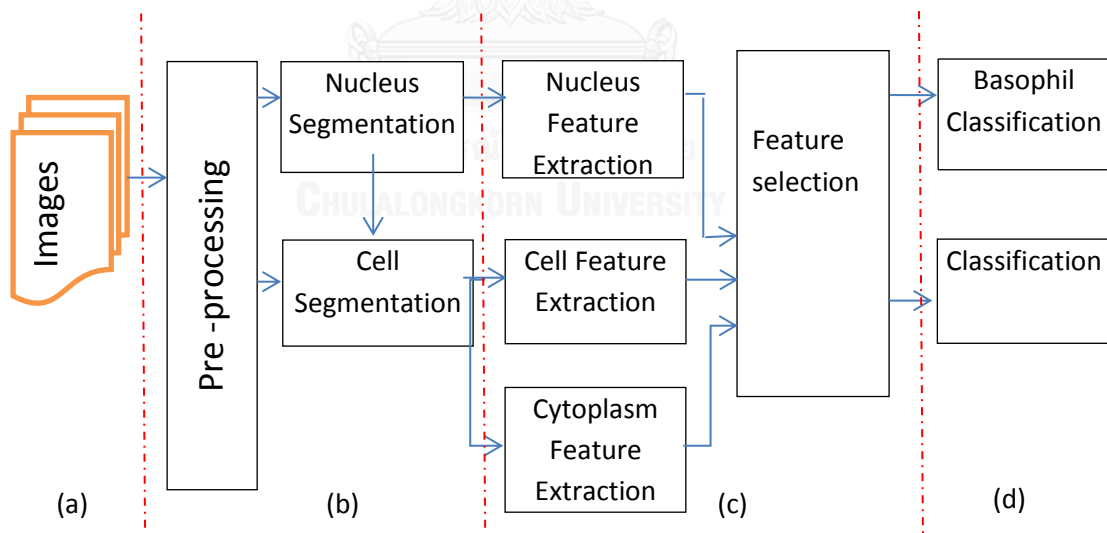
$$\text{precision} = TP / (TP+FP) \quad (54)$$

บทที่ 4

ระเบียบวิธีที่นำเสนอ

ระเบียบวิธีที่นำเสนอ จำเป็นที่จะต้องมีเรียนรู้เงื่อนไขภายใต้รูปแบบที่ใช้ในการพิจารณาหาความแตกต่างของเซลล์เม็ดเลือดขาวแต่ละชนิด โดยการใช้ชุดของข้อมูลที่เป็นตัวแทนของเซลล์แต่ละชนิด แล้วหาค่าทางสถิติเฉลี่ยสำหรับเซลล์แต่ละชนิด จากนั้นความแตกต่างในแต่ละสถิติระหว่างข้อมูลเซลล์แต่ละชนิด ทำการตรวจสอบและนำข้อมูลทางสถิติมาหาความแตกต่างที่มากที่สุดระหว่างเซลล์แต่ละชนิด ขั้นตอนทั่วไป ที่ดำเนินการในบริบทของวิทยานิพนธ์นี้มีวิธีการต่อไปนี้

1. การจัดเตรียมภาพ
2. การแบ่งส่วนต่างๆของเม็ดเลือดขาวออกจากภาพ
3. การสกัดลักษณะบ่งต่างๆของเม็ดเลือดขาวออกจากภาพ
4. การจำแนกประเภทเม็ดเลือดขาว



รูปที่ 4. 1 แผนภาพแสดงภาพรวมของระบบ (a) การจัดเตรียมภาพ (b) การแบ่งส่วน (c) การสกัดลักษณะบ่งต่างๆและการเลือกใช้ และ (d) การจำแนก

4.1.การจัดเตรียมภาพ

ในวิทยานิพนธ์ชุดนี้ มีการใช้ภาพเซลล์เม็ดเลือด 2 ชุดข้อมูลคือ ชุดข้อมูลที่ 1 เป็นภาพเซลล์เม็ดเลือดที่ผ่านการย้อมสี จาก กล้องจุลทรรศน์ดิจิทัล โดยใช้สไลด์กลุ่มตัวอย่างของคนปกติ จากขั้นตอนนี้จะได้ภาพเพื่อนำมาใช้จำนวน 828 ภาพ บันทึกด้วย กล้องจุลทรรศน์ ที่มี กำลังขยาย 100 เท่า และ เชื่อมต่อกับกล้อง Nikon DS-Fi2 ความละเอียดสูง ขนาดภาพที่ได้ 960x1280 พิกเซล ประกอบด้วยเซลล์เม็ดเลือดขาว 879 เซลล์ ความละเอียดของภาพ 150 พิกเซลต่อ 10 ไมโครเมตร ชุดข้อมูลที่ 2 เป็นภาพเซลล์เม็ดเลือดจาก โปรแกรม CellaVision จำนวน 477 ภาพ ประกอบด้วย เซลล์เม็ดเลือดขาว 477 เซลล์ แต่ละภาพถูกบันทึกในรูปแบบ jpg มีขนาด 360 x 363 พิกเซล ตัวอย่างภาพที่ได้มาได้มีการจำแนกชนิดเซลล์แล้ว ความละเอียดของภาพไม่ได้ระบุ ในที่นี้ผู้เขียนใช้วิธีประมาณจากความกว้างของเม็ดเลือดแดงเฉลี่ย พบว่ามีความละเอียดประมาณ 10 พิกเซลต่อ 1 ไมโครเมตร

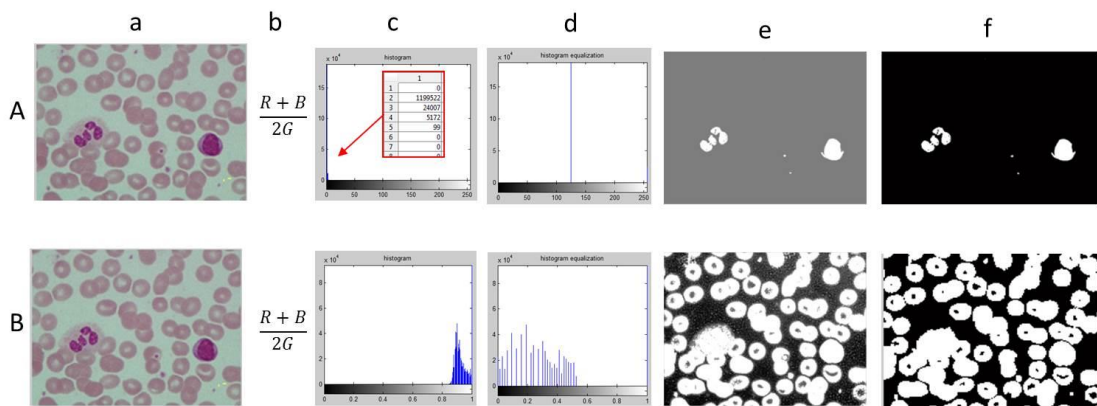
4.2.การแบ่งส่วนต่างๆของเม็ดเลือดขาวออกจากภาพ

ในส่วนนี้เป็นกระบวนการแบ่งส่วนเซลล์เม็ดเลือดขาวออกเป็นนิวเคลียสและไซโทพลาซึม โดยในวิทยานิพนธ์ฉบับนี้จะทำการเปรียบเทียบการแบ่งส่วนภาพจากวิธีที่นำเสนอเทียบกับวิธีแอกทีฟคอนทัวร์ ขั้นตอนที่เหมือนกันของทั้งสองวิธี สามารถแบ่งออกเป็น 3 ส่วนคือ การเตรียมภาพก่อนการประมวลผล การแบ่งส่วนนิวเคลียส และการแบ่งส่วนไซโทพลาซึม ข้อมูลที่นำมาทำการเปรียบเทียบ แบ่งเป็น 2 กลุ่ม คือ การเปรียบเทียบในเชิงเวลา การเปรียบเทียบผลการแบ่งส่วนในเชิงความคล้ายคลึงกับผลลัพธ์ที่ผู้เชี่ยวชาญแบ่ง และนำเสนอในรูปแบบตารางและ แผนภาพกล่อง (boxplots หรือ box and whisker plots)

4.2.1. การแบ่งส่วนด้วยวิธีที่นำเสนอ

4.2.1.1. ขั้นตอนการเตรียมภาพก่อนการประมวลผล

เซลล์เม็ดเลือดขาวทั้งห้า ชนิด มีความแตกต่างในเรื่อง รูปร่าง ขนาด ลักษณะของพื้นผิว การติดสี และอื่นๆของทั้งนิวเคลียสและไซโทพลาซึม จึงจำเป็นต้องสร้างภาพใหม่ เพื่อลบภาพพื้นหลัง กระบวนการนี้ แสดงในรูปที่ 4. 2



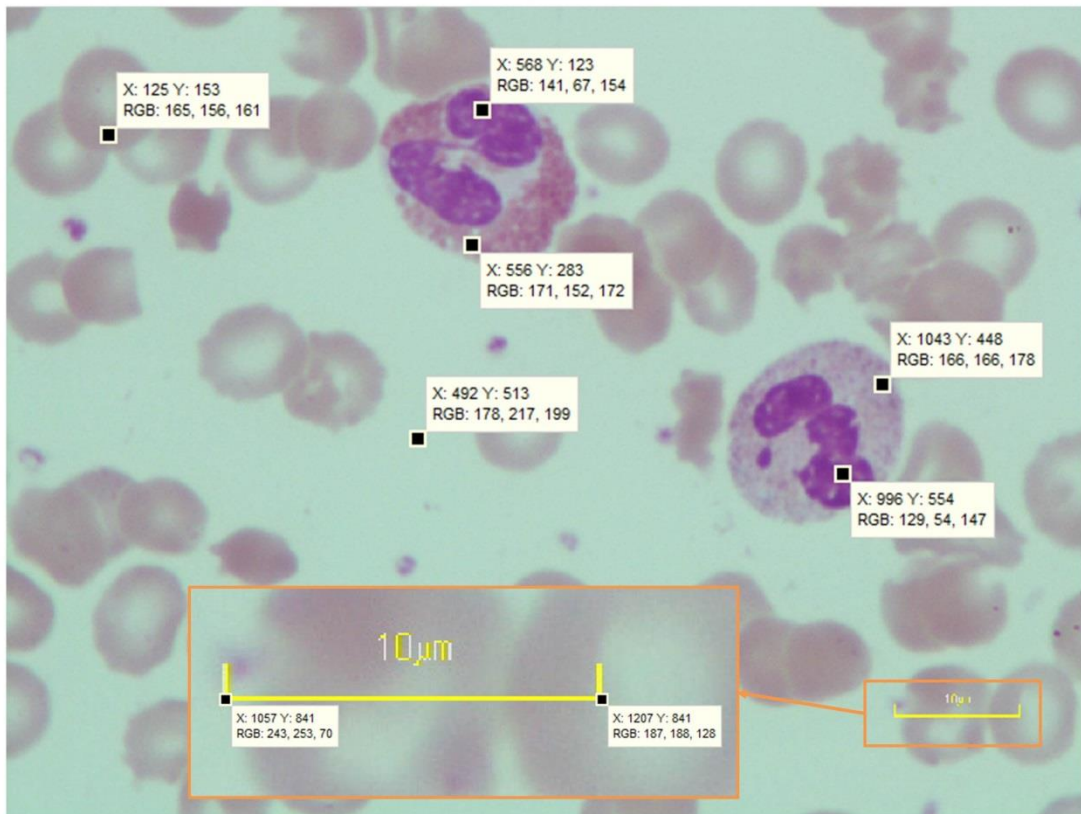
รูปที่ 4. 2 ขั้นตอนการเตรียมภาพก่อนการประมวลผล

แถว (A) ทดสอบกับภาพชนิด unsigned integer 8 bit (หรือ uint8)

แถว (B) ทดสอบกับภาพชนิด double-precision floating-point (หรือ double)

a คือภาพเริ่มต้น, b คือ การแยกภาพสีออกเป็น 3 ช่อง (R, G, B) และนำแต่ละช่องไปทำ $\frac{R+B}{2G}$,
 c คือฮิสโทแกรมที่ได้จากขั้นตอน b, d คือ ฮิสโทแกรมที่ได้จากการทำ histogram equalization,
 e คือภาพที่ได้จากการทำ histogram equalization, f คือ การแปลงภาพ e ให้เป็นภาพลักษณะฐานสอง โดยใช้ค่าระดับกัน 0.5

เนื่องจากภาพเซลล์เม็ดเลือดขาวทุกชนิดบริเวณของนิวเคลียสเมื่อผ่านการย้อมสีจะติดสีม่วง ซึ่งสีม่วงมีองค์ประกอบของสีแดงและสีน้ำเงิน ในสัดส่วนที่ใกล้เคียงกัน ส่วนขององค์ประกอบสีเขียวจะมีค่าเพียงเล็กน้อยเท่านั้น จากข้อมูลภาพที่มีพบว่าค่าความเข้มสีในช่องสีเขียวของนิวเคลียสจะมีค่าต่ำที่สุดเมื่อเทียบกับวัตถุอื่นๆในภาพดังแสดงใน รูปที่ 4. 3 ดังนั้นจึงนำค่าระดับความเข้มของภาพจากช่องสีแดงและสีน้ำเงินรวมกันหาค่าเฉลี่ยแล้วหารด้วยค่าระดับความเข้มของภาพจากช่องสีเขียว และทำการปรับเท่าฮิสโทแกรม (histogram equalization) ของภาพที่ได้ใหม่โดยทำในภาพทั้งชนิด unsigned integer 8 bit และ double-precision floating-point ผลที่ได้คือได้ข้อมูลภาพใหม่ 2 ภาพ หลังจากนั้นนำภาพใหม่ที่ได้ไปแปลงเป็นภาพลักษณะฐานสอง เพื่อใช้ในการแบ่งส่วนนิวเคลียสและเซลล์ ตามลำดับ

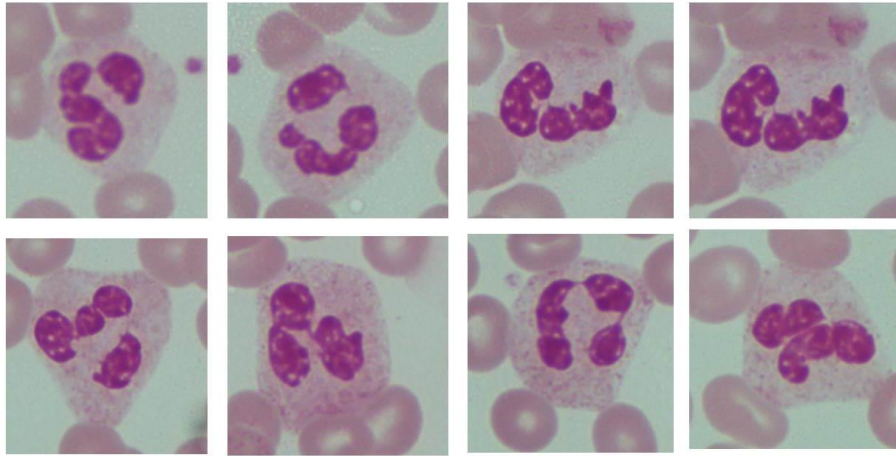


รูปที่ 4.3 ค่าระดับความเข้มสีเท่าใน RGB ของวัตถุต่าง ๆ ในภาพย้อมสีเลือด และ ruler scale ที่กำหนดโดยผู้ผลิต

4.2.1.2. การแบ่งส่วนนิวเคลียส

หลังจากขั้นตอนการเตรียมภาพก่อนการประมวลผล ใช้เทคนิคการกัดเซาะและการขยาย (erosion and dilation techniques) มาใช้ในการขจัดเซลล์ที่ไม่เกี่ยวข้อง โดย disk-shaped SE ที่มีรัศมีเท่ากับ $0.33 \mu\text{m}$ (ชุดข้อมูลที่ 1 มีขนาด 5 พิกเซล, ชุดข้อมูล 2 มีขนาด 3 พิกเซล)

การพิจารณาขนาดวัตถุที่พบในภาพที่คาดว่าจะป็นนิวเคลียส เราพิจารณาเซลล์เม็ดเลือดขาวที่มีจำนวนพูหลายพู ซึ่งมีสองชนิด คือ eosinophil และ neutrophil เซลล์ทั้ง 2 ชนิดมีขนาดใกล้เคียงกันคือ $9-15 \mu\text{m}$ (จากตารางที่ 3. 1) แต่อัตราส่วนนิวเคลียสต่อไซโทพลาซึมของ neutrophil คือ 1:3 [52] ดังนั้นพื้นที่นิวเคลียสจึงเป็นหนึ่งในสี่ของพื้นที่ทั้งหมดของเซลล์ และจำนวนพูของ neutrophil พบประมาณ 2 ถึง 5 พู แต่เนื่องจากนิวเคลียสแม้จะมีจำนวนพูมาก แต่เนื่องจากมีพื้นที่เซลล์จำกัดทำให้มีลักษณะการวางตัวของนิวเคลียสเป็นแบบขดตัว และส่วนใหญ่จะยังเชื่อมต่อกันด้วยโครมาทินบาง ๆ ดัง ในวิทยานิพนธ์ฉบับนี้จะพิจารณา neutrophil ที่แยกออกจากกันโดยมีจำนวน 3 พู



รูปที่ 4. 4 แสดงตัวอย่างของ neutrophil ที่มีนิวเคลียส 3 - 5 พู

neutrophil ขนาดเล็กสุด คือ มีเส้นผ่านศูนย์กลาง $9 \mu\text{m}$ จะมีพื้นที่ $8.1 \times 10^{-11} \text{m}^2$ โดยพื้นที่ของนิวเคลียสจะมีขนาดหนึ่งในสี่ของขนาดเซลล์รวม สามารถคำนวณหาพื้นที่ได้เป็น $20.25 \times 10^{-12} \text{m}^2$ และ คำนวณหาพื้นที่เล็กที่สุดสำหรับแต่ละพูในที่นี่ใช้จำนวน 3 พู จะได้พื้นที่ $6.75 \times 10^{-12} \text{m}^2$ หรือหากประมาณพื้นที่เป็นสี่เหลี่ยมจัตุรัสจะได้พื้นที่ $(2.6 \times 10^{-6} \text{m})^2$ นั่นก็คือสี่เหลี่ยมจัตุรัสที่มีความยาวด้านละ $2.6 \times 10^{-6} \text{m}$ จากข้อมูลชุดที่ 1 ภาพมีความละเอียด 150 พิกเซล ต่อ $10 \mu\text{m}$ ดังนั้นสามารถคำนวณพื้นที่เล็กที่สุดสำหรับนิวเคลียสแต่ละพูของ neutrophil ได้เป็น 1521 พิกเซล ข้อมูลชุดที่ 2 ทำการคำนวณลักษณะเดียวกัน แต่ใช้ความละเอียด 10 พิกเซล ต่อ $1 \mu\text{m}$

$$\text{Total minimum area of neutrophil} = (9 \mu\text{m})^2 = 8.1 \times 10^{-11} \text{m}^2$$

$$\text{Nucleus area} = \text{Total cell area} / 4 = 20.25 \times 10^{-12} \text{m}^2$$

$$\text{Minimum nucleus lobe area} = \text{Nucleus area} / 3 = 6.75 \times 10^{-12} \text{m}^2$$

$$\text{Minimum nucleus lobe area} = 6.75 \times 10^{-12} \text{m}^2 = (2.6 \times 10^{-6} \text{m})^2$$

Calculation of minimum nucleus lobe area in pixel unit from image resolution

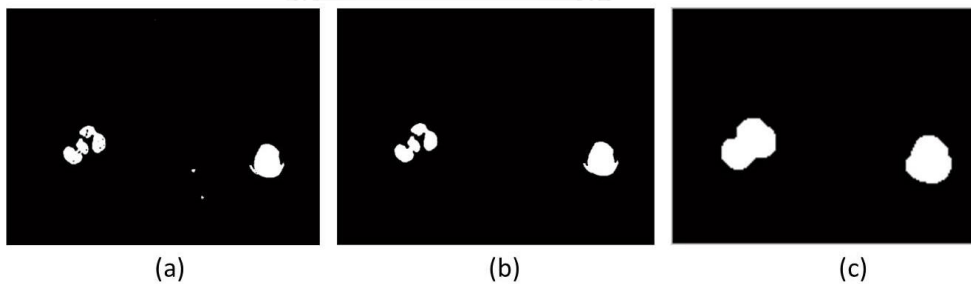
For dataset 1, image resolution is $150 \text{ pixels} / (10 \times 10^{-6} \text{m})$

$$\text{Minimum nucleus lobe area} = \left(2.6 \times 10^{-6} \text{m} \times \frac{150 \text{ pixels}}{(10 \times 10^{-6} \text{m})}\right)^2 = (39 \text{ pixels})^2$$

$$\text{Minimum nucleus lobe area} = (39 \text{ pixels} \times 39 \text{ pixels}) = 1521 \text{ pixels}$$

จากข้อมูลการคำนวณจึงประมาณค่าในการเลือกวัตถุที่ขนาดใหญ่กว่า 1500 พิกเซล และ 670 พิกเซล สำหรับชุดข้อมูลที่ 1 และ ชุดข้อมูล 2 ตามลำดับ ผลที่แสดงดังรูปที่ 4. 5 (b) ทำการผสานเซลล์ที่มีนิวเคลียสมากกว่าหนึ่งก้อนเข้าด้วยกัน เพื่อระบุจำนวนของเซลล์ที่พบ แสดงดัง รูปที่ 4. 5 (c) และทำการเติมเต็มช่องที่ขนาดเล็กกว่า $1.49 \mu\text{m} \times 1.49 \mu\text{m}$ (500 พิกเซลสำหรับชุดข้อมูลที่ 1 และ 220 พิกเซลสำหรับชุดข้อมูล 2) เพื่อเติมช่องในนิวเคลียสที่เกิดจากกระบวนการกำจัดเชื้อโรค (phagocytosis) ซึ่งพบในเซลล์บางตัว

ผลที่ได้จากการแบ่งส่วนนิวเคลียส สามารถนำไปใช้บ่งชี้เซลล์ที่เป็นชนิด basophil เนื่องจากในกรณีของ basophil จะมีเม็ดเล็ก ๆ ปกคลุมเต็มเซลล์จนไม่สามารถมองเห็นนิวเคลียส กรณีนี้ จะทำการแบ่งส่วนเซลล์โดยใช้ภาพคอนเวกซ์ของนิวเคลียส ตัวแปรที่นำมาใช้พิจารณาในการจำแนกเซลล์ชนิดนี้คือ ขนาด และ ค่าความแปรปรวนของในช่องสีแดงและสีน้ำเงิน ด้วยเหตุผลว่า ขนาดของ basophil จะมีขนาดใหญ่กว่าเซลล์ที่มีเม็ดเล็ก ๆ ทุกตัว และ ความแปรปรวนของความเข้มในระดับสีเทาของช่องสีแดงและสีน้ำเงินจะสูงกว่าเซลล์ที่ไม่มีเม็ดเล็ก ๆ สำหรับการพิจารณาขนาดของนิวเคลียส พิจารณาเซลล์ที่มีพื้นที่ของนิวเคลียสมากกว่า 20,000 พิกเซล สำหรับชุดข้อมูลที่ 1 และ 10000 พิกเซลสำหรับชุดข้อมูลที่ 2 ค่าความแปรปรวนของสัญญาณในช่องสีแดงที่มากกว่า 200 และ สีน้ำเงินที่มากกว่า 93 เฉพาะเซลล์ที่เป็น basophil จะนำข้อมูลภาพนิวเคลียสไปหา คอนเวกซ์ฮัลล์ เพื่อเป็นขอบของเซลล์ทันที ส่วนเซลล์ชนิดอื่น ๆ หาขอบของเซลล์ดังขั้นตอนต่อไป



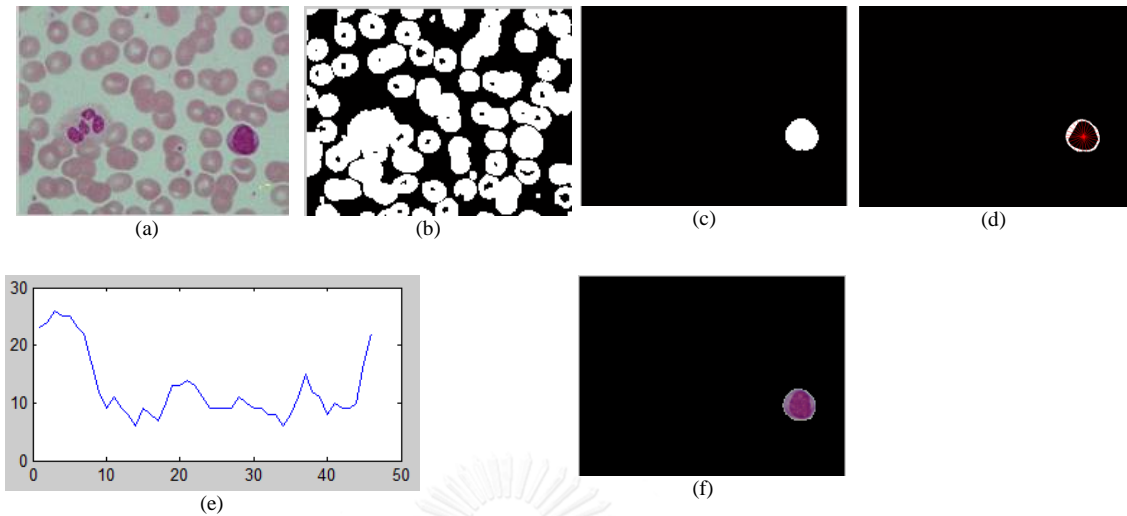
รูปที่ 4. 5 (a) ภาพที่ได้หลังจากขั้นตอนการเตรียมภาพก่อนการประมวลผล (b) ผลการแบ่งส่วนนิวเคลียส (c) การระบุจำนวนเซลล์ที่พบในภาพ

4.2.1.3. การแบ่งส่วนเซลล์

ปัญหาหลัก ๆ ที่พบในการแบ่งส่วนเซลล์คือความหลากหลายในเรื่องของรูปร่าง ขนาด ความสมมาตร การวางแนว และอื่น ๆ นอกเหนือจากนี้คือการมีเซลล์ที่วางเรียงติดกันซึ่งมีความคล้ายคลึงกับนิวเคลียสและไซโทพลาซึม

จากข้อมูลที่ว่าไซโทพลาซึมจะอยู่ล้อมรอบนิวเคลียส เราจึงเลือกเฉพาะวัตถุที่ทุกพื้นที่ของนิวเคลียส (แบ่งส่วนได้จากขั้นตอนก่อนหน้า) เป็นสมาชิกอยู่ วัตถุที่สนใจมีโอกาสเกิดได้สองแบบคือ เซลล์เม็ดเลือดขาวอย่างเดียว หรือเซลล์เม็ดเลือดขาวที่มีเซลล์อื่น ๆ ติดอยู่ด้วย ในกรณีหลัง จำเป็นต้องทำการแบ่งส่วนต่อไป สำหรับวิธีการตรวจสอบทั้งสองกรณีนี้ เราใช้วิธีการตรวจสอบหาจุดเว้าเข้า โดยดำเนินการดังนี้

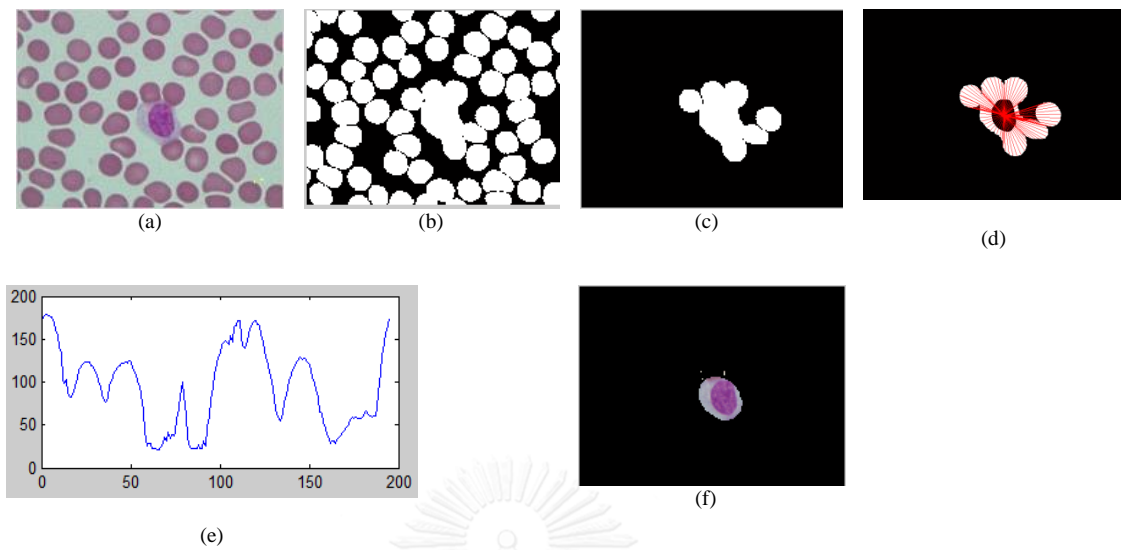
1. สร้างภาพคอนเวกซ์ของวัตถุที่สนใจ
2. นำภาพคอนเวกซ์ลบออกจากภาพนิวเคลียสที่แบ่งส่วนได้
3. หาจุดศูนย์กลาง กรณีนี้จะใช้จุดศูนย์กลางของภาพคอนเวกซ์ของนิวเคลียสที่แบ่งส่วนได้ (เพื่อหลีกเลี่ยงกรณีนิวเคลียสมีหลายก้อน)
4. ลากเส้นจากจุดศูนย์กลางไปยังขอบของวัตถุในข้อ 1 กระบวนการขั้นตอนนี้ทำในภาพที่ได้จากข้อ 2 (กรณีนี้ทำการสร้างเส้นแบบจุด เว้น 10 จุด)
5. นับจำนวนพิกเซลสีขาวที่พบบนเส้นแต่ละเส้น เก็บค่าไว้ในตัวแปร p
6. หาค่าเฉลี่ยของอาร์เรย์ p พล็อตกราฟแสดงค่า p
7. ตั้งค่าระดับกัน จากค่าเฉลี่ยของอาร์เรย์ p ในข้อ 6 + 30 พิกเซล (หรือ $2 \mu\text{m}$)
8. นับจำนวนจุดที่กราฟ p ตัดกับค่าระดับกัน
9. จำนวนจุดเว้า = จำนวนจุดรวมจากข้อ 8 / 2



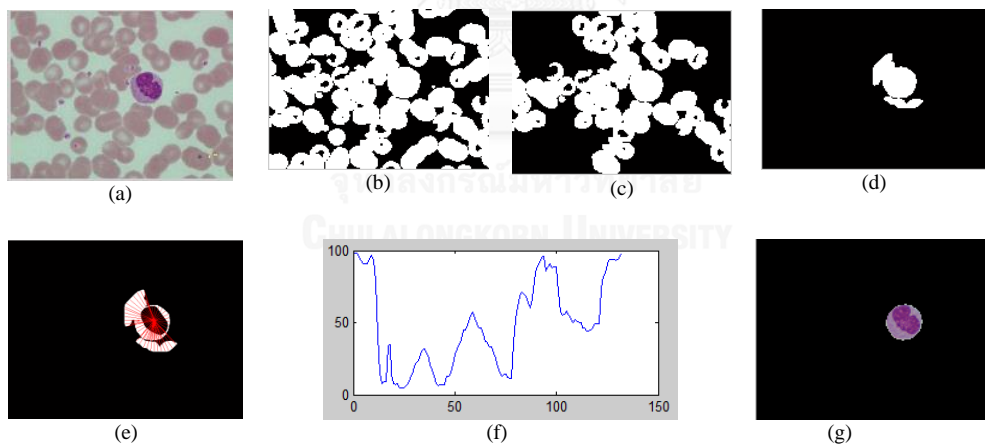
รูปที่ 4. 6 กรณีวัตถุที่สนใจเป็นเซลล์เม็ดเลือดขาวแยกจากเซลล์อื่นๆ
 (a) ภาพเริ่มต้น, (b) ภาพที่ได้จาก preprocessing, (c) วัตถุที่สนใจ,
 (d), (e) การตรวจสอบพื้นที่เว้าเข้า, (f) ผลการแยกเซลล์เม็ดเลือดขาว

วัตถุที่สนใจมี 3 กรณี คือ กรณีวัตถุที่สนใจคือเซลล์เม็ดเลือดขาวเพียงอย่างเดียว และกรณีวัตถุที่สนใจคือเซลล์เม็ดเลือดขาวและมีเซลล์อื่นๆติดกัน ในกรณีนี้สามารถจำแนกออกด้วยพื้นที่ของวัตถุนั้นแบ่งได้อีก 2 แบบคือ กรณีที่วัตถุที่สนใจมีพื้นที่น้อยกว่า 0.1 เท่าของพื้นที่ทั้งภาพ และวัตถุที่สนใจมีพื้นที่มากกว่า 0.1 เท่าของพื้นที่ภาพ ในกรณีหลังนี้จะใช้ภาพนิวเคลียสที่แบ่งส่วนได้ทำการขยายโดยใช้ disk-shaped SE ที่มีรัศมีเท่ากับ 90 พิกเซล ($6 \mu\text{m}$) และใช้ตัวดำเนินการ AND ในภาพนิวเคลียสที่ขยายและภาพวัตถุที่เลือกไว้ และพิจารณาวัตถุที่มีพื้นที่นิวเคลียสเป็นองค์ประกอบเท่านั้น

ขั้นตอนการดำเนินการแสดงดังรูปที่ 4. 6 – 4.7 ตามลำดับ



รูปที่ 4. 7 กรณีวัตถุที่สนใจเป็นเซลล์เม็ดเลือดขาวที่มีเซลล์ติดกันอยู่เล็กน้อย
 (a) ภาพเริ่มต้น, (b) ภาพจาก preprocessing, (c) วัตถุที่สนใจ ,
 (d), (e) การตรวจสอบพื้นที่เว้าเข้า, (f) ผลการแยกเซลล์เม็ดเลือดขาว



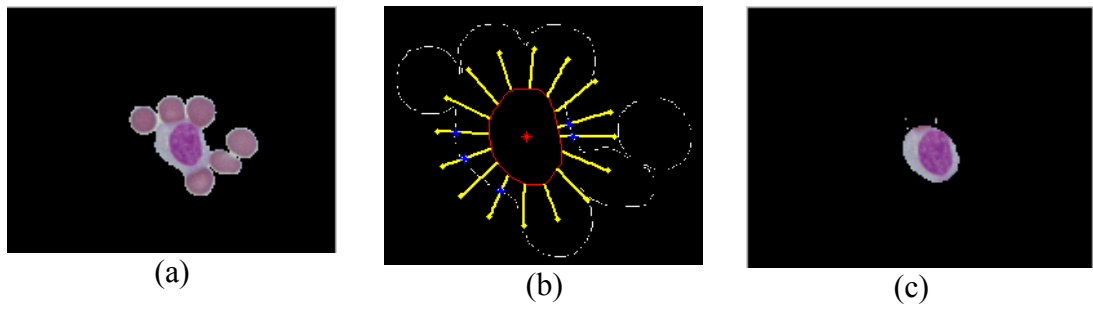
รูปที่ 4. 8 กรณีวัตถุที่สนใจเป็นเซลล์เม็ดเลือดขาวและมีเซลล์ติดกันจำนวนมาก
 (a) ภาพเริ่มต้น, (b) ภาพที่ได้จาก preprocessing, (c) วัตถุที่สนใจ,
 (d) นำการกร่อนมาใช้เพื่อลดขนาดพื้นที่, (e), (f) การตรวจสอบพื้นที่เว้าเข้า,
 (g) ผลการแยกเซลล์เม็ดเลือดขาว

การแบ่งส่วนเซลล์เม็ดเลือดขาวหลังการพิจารณาจุดเว้าเข้าดำเนินการโดยใช้หลักการของการหาขอบและกำหนดเงื่อนไขโดยมีขั้นตอนดังนี้

1. สร้างภาพขอบของวัตถุแบบหยาบๆ ด้วย Canny edge
2. สร้างเส้นรัศมี โดยมีจุดศูนย์กลางที่ centroid ของ ภาพคอนเวกซ์ของนิวเคลียส มีรัศมีเท่ากับ 1.5 เท่าของความยาวแกนหลักของภาพคอนเวกซ์ของนิวเคลียส
3. ขยายภาพคอนเวกซ์ของนิวเคลียสออกเล็กน้อย โดยใช้ SE รูปทรง disk ที่มีรัศมีเท่ากับ 2 เพื่อหาจุดเริ่มต้นของเส้นรัศมีที่ไม่ซ้อนทับกับขอบของนิวเคลียส
4. หาพิกัดจุดขอบของภาพ คอนเวกซ์ของนิวเคลียส
5. หาจุดตัดของเส้นรัศมีและขอบของภาพคอนเวกซ์นิวเคลียส กำหนดให้จุดตัดนั้นคือจุดเริ่มต้นใหม่ของเส้น
6. กำหนดเงื่อนไข เพื่อหาว่าจุดใดคือจุดขอบที่สนใจ โดยในวิทยานิพนธ์นี้ใช้ข้อมูลระดับความจากรูปที่ 4. 9 (a) โดยมีเงื่อนไขดังนี้
 - a. บนเส้นรัศมีพบจุดขอบ 1 จุด และจุดปลายของเส้นมีค่าระดับความเข้มเป็นศูนย์
 - b. บนเส้นรัศมีพบจุดขอบ 2 จุดที่อยู่ติดกัน และจุดปลายของเส้นมีค่าระดับความเข้มเป็นศูนย์
 - c. บนเส้นรัศมีพบจุดขอบ 2 จุดที่อยู่ติดกัน และทุกๆจุดระหว่างจุดทั้งสองบนเส้นมีค่าระดับความเข้มเป็นศูนย์

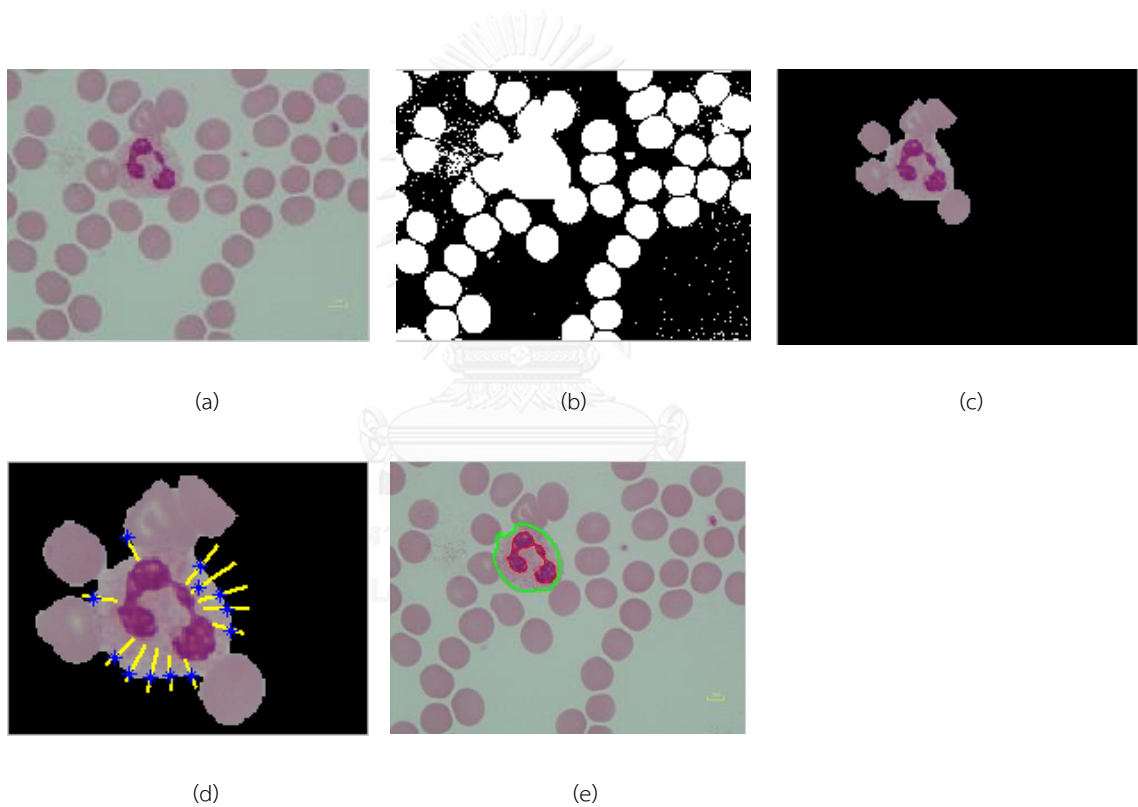
กรณี a และ b จุดที่พบคือจุดขอบที่สนใจ แต่กรณี b เลือกจุดแรกเพียงจุดเดียว ส่วนกรณี c กำหนดให้จุดแรกคือจุดขอบที่สนใจ

จากรูปที่ 4. 10 (b) ในขั้นตอนที่ 1 แสดงเส้นขอบของวัตถุที่สนใจ (เส้นสีขาว), ขั้นตอนที่ 4 เส้นขอบคอนเวกซ์นิวเคลียส (เส้นสีแดง) ขั้นตอนที่ 5 สร้างเส้นรัศมี (เส้นสีเหลือง) และ ขั้นตอนที่ 6 จุดขอบที่เลือกจากเงื่อนไขที่กำหนด (จุดสีน้ำเงิน) ถัดมาทำการประยุกต์ใช้ เทคนิค the direct least square fitting ของรูปร่างที่นำเสนอโดยFitzgibbon [44], มาใช้กับจุดที่เลือก จะได้ตัวแปรของสมการวงรี ทำการสร้างภาพวงรี และใช้ตัวดำเนินการ 'AND' กับภาพวงรีและภาพวัตถุที่เลือก เลือกเฉพาะวัตถุที่มีนิวเคลียสเป็นสมาชิก



รูปที่ 4. 9 แสดงขั้นตอนการแบ่งส่วนเซลล์

(a) วัตถุที่ถูกเลือก, (b) เส้นขอบของวัตถุที่สนใจ (เส้นสีเขียว), เส้นขอบคอนเวกซ์นิวเคลียส (เส้นสีแดง), เส้นรัศมี (เส้นสีเหลือง), จุดขอบที่เลือก (จุดสีน้ำเงิน), (c) ผลลัพธ์หลังจากใช้ตัวดำเนินการ AND

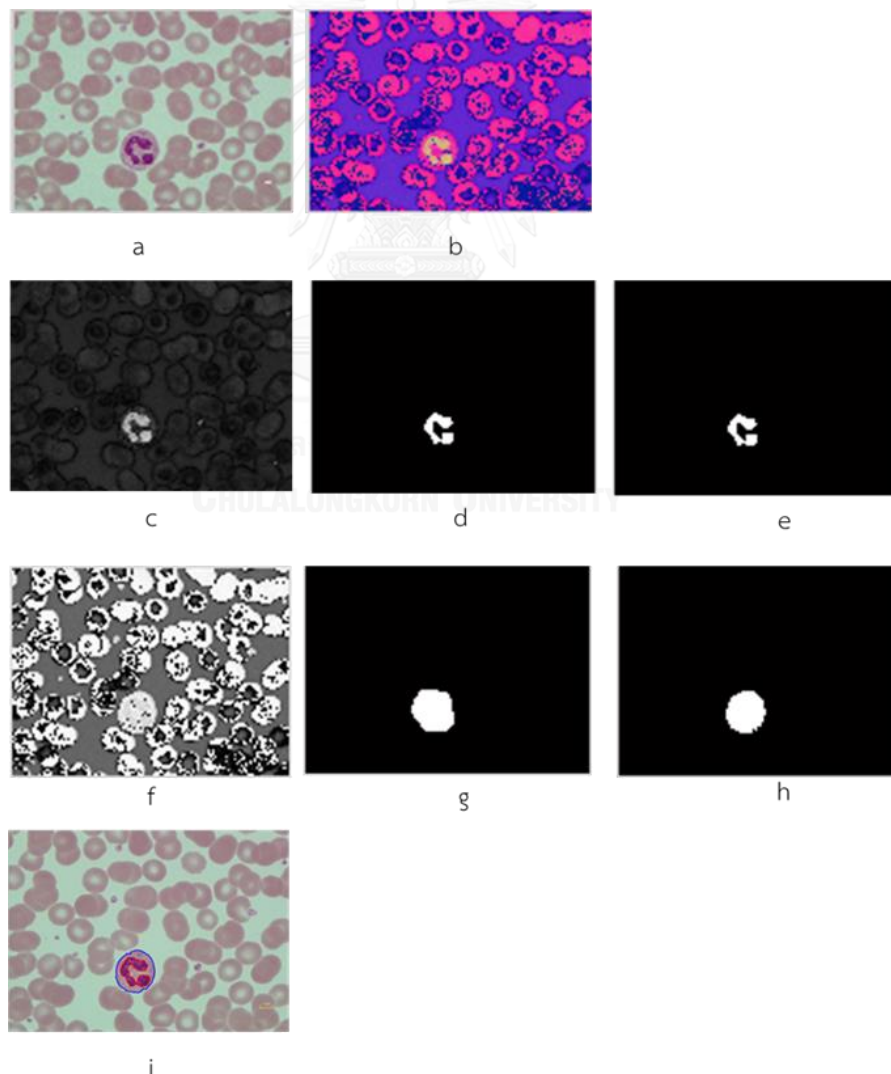


รูปที่ 4. 10 ขั้นตอนการแบ่งส่วนเซลล์

(a) ภาพเริ่มต้น, (b) ภาพจาก pre-processing, (c) วัตถุที่สนใจ, (d) จุดขอบที่เลือก, (e) ผลลัพธ์การแบ่งส่วนเซลล์และนิวเคลียส

4.2.2. การแบ่งส่วนด้วยวิธีที่แอกทีฟคอนทัวร์

การแบ่งส่วนด้วยวิธีที่แอกทีฟคอนทัวร์ วิธีที่สนใจ คือ วิธีที่นำเสนอโดย Mathur และคณะ [18] โดยเริ่มจากการนอร์มอไลซ์ภาพด้วยเทคนิคการแปลงสีระหว่างภาพที่นำเสนอโดย Reinhard และคณะ [19] หลังจากนั้นนำภาพที่ผ่านการนอร์มอไลซ์แล้ว มาแปลงจากปริภูมิสี RGB ไปเป็น HSV แยกภาพจากช่องสัญญาณ saturation ไปใช้ในการแบ่งส่วนนิวเคลียส หากอนทัวร์เริ่มต้น โดยทำการแปลงภาพเป็นภาพลักษณะฐานสอง ทำ opening morphological ในวิทยานิพนธ์ฉบับนี้เลือกใช้โครงสร้างเมทริกซ์จัตุรัส 7×7 และกรองพื้นที่ที่ใหญ่กว่า 1500 พิกเซล การแบ่งส่วนเซลล์เม็ดเลือดขาว จะทำในภาพจากช่องสัญญาณ hue โดยใช้คอนทัวร์เริ่มต้นหามาจากการแปลงภาพเป็นภาพลักษณะฐานสอง ทำ opening โดยใช้ โครงสร้างเมทริกซ์จัตุรัส 7×7 และใช้ connected component analysis (CCA), เลือกวัตถุที่สนใจดังแผนภาพในรูปที่ 4.12

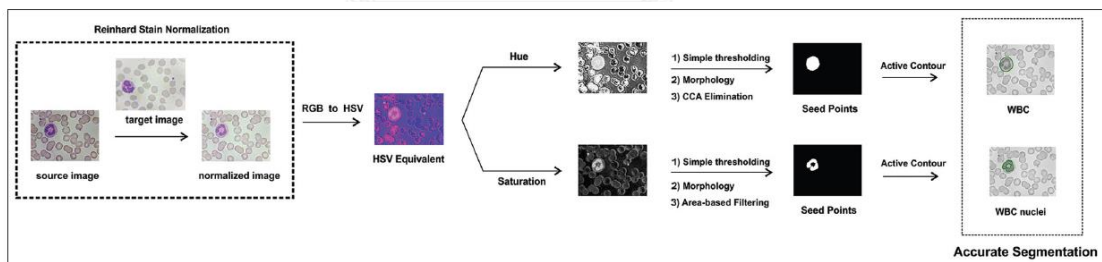


รูปที่ 4.11 ขั้นตอนของวิธี ACWE1

แต่เนื่องจากไม่มีแนวทางการเลือกภาพที่จะนำมาทำการนอร์มอลไลซ์ จึงทำการแบ่งส่วนด้วยวิธีที่แอกทิฟคอนทัวร์ โดยทำตามขั้นตอนที่ Mathur และคณะนำเสนอ ยกเว้นขั้นตอนการนอร์มอลไลซ์ภาพ และทำเลือกภาพที่ให้ค่าความถูกต้องของการแบ่งส่วนสูงสุดของชุดข้อมูลที่ 2 (เนื่องจากเป็นภาพที่ผ่านการกรองมาแล้ว) มาเป็นภาพต้นแบบสำหรับการนอร์มอลไลซ์ภาพ เรียกวิธีนี้ว่า ACWE1 ส่วนวิธีที่นำเสนอโดย Mathur และคณะ เรียกว่า ACWE2

รายละเอียดย่อยของวิธี ACWE1 คือจะแปลงภาพต้นฉบับ (รูปที่ 4.11 a) จากปริภูมิสี RGB ไปเป็น HSI (รูปที่ 4.11 b) แยกภาพจากช่องสัญญาณ saturation (รูปที่ 4.11 c) มาแปลงเป็นภาพลักษณะฐานสอง ใช้เทคนิค erosion และ dilation เพื่อกำจัดวัตถุขนาดเล็กๆ ออกด้วย SE รูปทรง 'disk' ขนาดรัศมี 5 พิกเซล เลือกเฉพาะวัตถุที่มีขนาดมากกว่า 1500 พิกเซล พิจารณาเป็นคอนทัวร์เริ่มต้นของการแบ่งส่วนนิวเคลียส (รูปที่ 4.11 d) ผลลัพธ์ที่ได้จากการแบ่งส่วนนิวเคลียส (รูปที่ 4.11 e) การแบ่งส่วนเซลล์เม็ดเลือดขาว จะทำในภาพจากช่องสัญญาณ hue (รูปที่ 4.11 f) โดยใช้คอนทัวร์เริ่มต้นจากผลลัพธ์ที่ได้จากการแบ่งส่วนนิวเคลียสและนำภาพนั้นไปทำการ dilation ด้วย โครงสร้าง 'disk' ขนาดรัศมี 30 พิกเซล หรือ 2 ไมครอน (รูปที่ 4.11 g) ผลลัพธ์ที่ได้จากการแบ่งส่วนเซลล์ (รูปที่ 4.11 h) เส้นสีแดงคือผลลัพธ์ของนิวเคลียสที่แบ่งส่วนได้ เส้นสีน้ำเงินคือผลลัพธ์ของเซลล์ที่แบ่งส่วนได้ (รูปที่ 4.11 i)

ACWE 2 จะทำการแบ่งส่วนนิวเคลียสและเซลล์ดังรูปที่ 4.12



รูปที่ 4.12 แสดงผังการแบ่งส่วนเซลล์เม็ดเลือดขาวโดยวิธีของ Mathur และ คณะ [18]

4.3. การสกัดลักษณะบ่งต่างของเม็ดเลือดขาว

หลังจากที่ทำการแบ่งส่วนภาพออกเป็นส่วนของนิวเคลียส และส่วนของเซลล์เม็ดเลือดขาว เราจะได้ภาพลักษณะฐานสอง 2 ภาพคือ ภาพลักษณะฐานสองของนิวเคลียส และภาพลักษณะฐานสองของเซลล์ ภาพลักษณะฐานสองของไซโทพลาซึมสามารถหาได้จากการนำภาพของเซลล์ลบด้วยภาพลักษณะฐานสองของนิวเคลียส ผลลัพธ์ที่ได้จากการแบ่งส่วนภาพ จะได้ภาพขาวดำของนิวเคลียส

เซลล์ และไซโทพลาซึม (ภาพเซลล์ – ภาพนิวเคลียส) สามารถนำไปสกัดลักษณะบ่งต่าง ได้ดังตารางที่ ค. 1

ก่อนการทำการเลือกลักษณะบ่งต่างของข้อมูลจำเป็นต้องทำการนอร์มอลไลต์ของข้อมูล เนื่องจากชุดข้อมูลที่ใช้มี 2 ชุดข้อมูล ซึ่งภาพที่ใช้มีความแตกต่างกัน จึงต้องนำแต่ละชุดข้อมูลมาทำการนอร์มอลไลต์ก่อนเข้าสู่ตัวจำแนก โดยในการนอร์มอลไลต์ของข้อมูล ใช้สมการ

$$Z_i = \frac{x_i - \min(X)}{\max(X) - \min(X)} \quad (55)$$

เมื่อ $X = (x_1, x_2, \dots, x_n)$ และ Z_i เป็นค่าที่ถูกนอร์มอลไลต์ตัวที่ i^{th}

ยกเว้นกรณี entropy จะใช้สมการ normalized entropy

$$H_{L-1}(P) = - \sum_j \frac{p(r_j) \log_b p(r_j)}{\log_b(L-1)} \quad (56)$$

เมื่อ $p_j = \frac{1}{L-1} \quad \forall j = 0, \dots, L-1$ และ การหารด้วย $\log_b(L-1)$ จะได้ $H_{L-1}(P) \in [0,1]$

4.4. การเลือกและการทดสอบลักษณะบ่งต่างของเม็ดเลือดขาวมาใช้ในการจำแนก

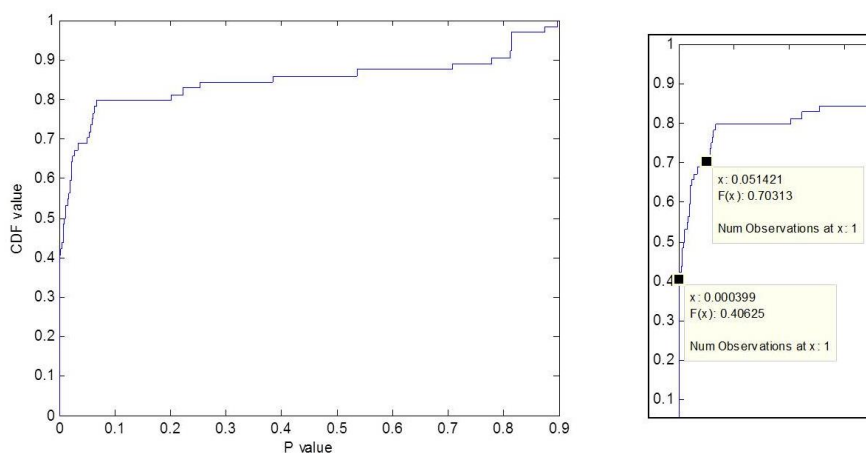
ชุดข้อมูลประกอบด้วย จำนวนเซลล์ 1356 เซลล์ ที่มี ลักษณะบ่งต่าง 65 ลักษณะ (ขนาด 1356×65) โดยอ้างอิงลำดับที่ของลักษณะบ่งต่างตาม ตารางที่ ค. 1 ส่วนการทำนายกำหนดด้วย ชนิดของเซลล์ที่มี 5 ชนิด คือ 1. basophil, 2. Eosinophil, 3. lymphocyte, 4. monocyte และ 5. Neutrophil มีแถวข้อมูลสอดคล้องกับของ OBS (ขนาด 1356×1)

ทำการแบ่งข้อมูลออกเป็นชุดการฝึกและชุดทดสอบ ในการคาดการณ์ผลการทำงานของแบบจำลองที่เลือก จึงจำเป็นต้องประเมินผลการทำงานของชุดข้อมูลชุดอื่นที่ไม่ได้ใช้ในการสร้างแบบจำลอง วิทยานิพนธ์นี้จะแบ่งข้อมูลด้วยวิธี hold out cross validation สำหรับการเลือก ลักษณะบ่งต่างออกเป็นชุดสำหรับการฝึก ซึ่งมีขนาด 677 ข้อมูล (50%) และชุดสำหรับทดสอบ 677 ข้อมูล (50%)

ในการเลือกลักษณะบ่งต่างโดยเบื้องต้นใช้วิธี filter เพื่อกรองก่อนว่าลักษณะบ่งต่างใดมีผลในการจำแนกเซลล์แต่ละเซลล์ เนื่องจากจำนวนลักษณะบ่งต่างที่มีอยู่และจำนวนข้อมูลที่จะนำมาฝึกเป็นชุดข้อมูลขนาดเล็กและแต่ละกลุ่มมีข้อมูลไม่เท่ากัน โดยเฉพาะ basophil ทำให้ไม่สามารถทำ wrapper ได้ วิธี filter มักจะถูกใช้เป็นขั้นตอนก่อนการประมวล เนื่องจากง่ายและรวดเร็ว โดยใช้ t-

test ในลักษณะบ่งต่างแต่ละลักษณะและเปรียบเทียบ p-value สำหรับแต่ละลักษณะบ่งต่างเป็นตัวชี้วัดถึงควมมีประสิทธิภาพในการจำแนกกลุ่ม หากกราฟการกระจาย cumulative distribution function (CDF) ของ p-values

ตัวอย่างการวิเคราะห์ด้วยวิธี filter แสดงดังรูปที่ 4.13 ซึ่งเป็นกราฟการกระจาย cumulative distribution function (CDF) ของ p-values ของเซลล์ชนิดที่ 1 และ 2



รูปที่ 4.13 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 1 และ 2

จากกราฟพบว่า มีประมาณ 40% ของลักษณะบ่งต่างที่มีต่อค่าใกล้เคียงกับศูนย์และ 70% ของลักษณะบ่งต่างที่มีค่า p มีขนาดน้อยกว่า 0.05 หมายความว่า ลักษณะบ่งต่าง 19 ลักษณะ (หรือ 30%) จากลักษณะบ่งต่างเริ่มต้น 65 ลักษณะ สามารถเลือกมาใช้ในการจำแนกประเภทได้ ทำการเรียงลำดับลักษณะบ่งต่างเหล่านี้ตามค่า p (หรือค่า absolute values ของ t-statistics) และเลือกลักษณะบ่งต่างจากรายการที่เรียงลำดับ โดยในกรณีนี้ลักษณะบ่งต่างลำดับที่ 9, 4, 31, 6, 7, 24, 11, 12, 60, 8, 57, 17, 63, 10, 5, 54, 15, 26, 25 เป็นต้น ส่วนของกราฟการกระจาย cumulative distribution function (CDF) ของ p-values: ของเซลล์ชนิดอื่นๆ การวิเคราะห์ข้อมูล และการเลือกลักษณะบ่งต่างแสดงใน ภาคผนวก ง

เมื่อเลือกลักษณะบ่งต่างจากข้างต้นมาได้ จะใช้วิธี wrapper ต่อไปโดยแบ่งข้อมูลเป็น 2 กลุ่มเป็นชุดสำหรับการฝึก ซึ่งมีขนาด 667 ข้อมูล และ ชุดสำหรับการทดสอบ ซึ่งมีขนาด 110ข้อมูลแบบจำลองที่นำมาทดสอบคือ linear discriminant function เลือกลักษณะบ่งต่างโดยใช้ sequential forward selection (SFS) ทำซ้ำ 20 รอบ ลักษณะบ่งต่างที่ให้ error น้อยที่สุด

ลักษณะบ่งต่างชุดนี้ก็จะถูกนำมาใช้ในการจำแนกด้วยชุดข้อมูลทดสอบ ซึ่งทำการแบ่ง 50% สำหรับฝึก และ 50 % สำหรับทดสอบเช่นกัน

4.5.การจำแนกประเภทเม็ดเลือดขาว

กลุ่มของเซลล์เม็ดเลือดขาว คือ Y ซึ่งมีสมาชิก 5 กลุ่ม

$$Y = (A_1, A_2, A_3, A_4, A_5) \quad (57)$$

โดย A_1 : basophil, A_2 : eosinophil, A_3 : lymphocyte, A_4 : monocyte, A_5 : neutrophil

ลักษณะบ่งต่างที่สกัด คือ X

$$X = (x_1, \dots, x_g) \in A_k \quad (58)$$

เมื่อ A_k คือ กลุ่มที่ทำนาย 5 กลุ่ม

g คือ จำนวนลักษณะบ่งต่างที่ถูกเลือก

discriminant function classification จะสร้างมาจาก learning set (L)

$$L = (X_1, Y_1), \dots, (X_n, Y_n) \quad (59)$$

เมื่อ n คือ จำนวนข้อมูลที่นำมาฝึก

classification C ที่สร้างจาก learning set L :

$$C(\dots, L): x \rightarrow \{1, 2, \dots, k\} \quad (60)$$

กลุ่มที่นำมาทำนายโดยใช้ค่าลักษณะบ่งต่าง

$$C(x, L) = k \text{ ถ้า } x \text{ อยู่ใน } A_k \quad (61)$$

โดยทำการจำแนก 2 วิธี เปรียบเทียบกัน คือ linear discriminant function และ quadratic discriminant function โดยใช้ลักษณะบ่งต่างที่เลือก ทำ LOOCV 10 fold

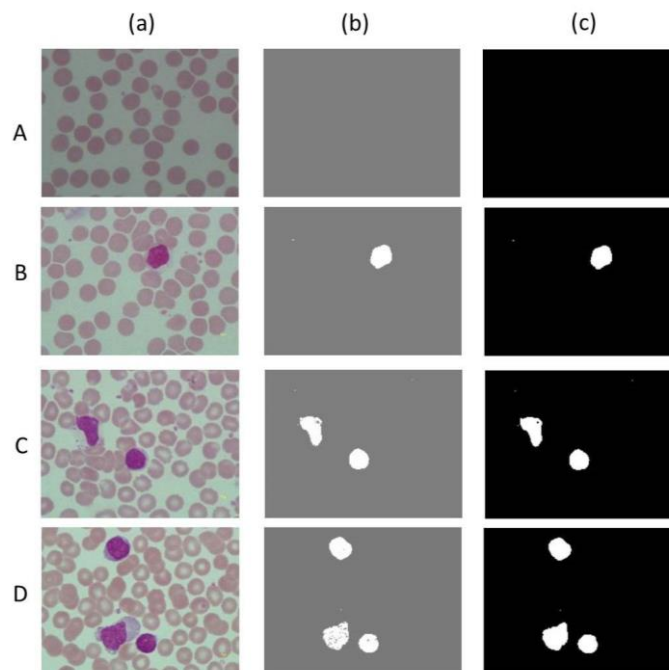
ผลการจำแนกที่ได้จะถูกบันทึกใน confusion matrix ดังในตารางที่ 3. 2 และคำนวณหาค่า accuracy (ความถูกต้อง) sensitivity (ค่าความไว) specificity และ precision โดยใช้สมการที่ (51) - (54)

บทที่ 5
ผลการศึกษา

5.1.การระบุตำแหน่งเซลล์เม็ดเลือดขาวในภาพ

ตารางที่ 5. 1 แสดงผลระบุตำแหน่งเซลล์เม็ดเลือดขาวในภาพ

| จำนวนภาพ | จำนวนเซลล์เม็ดเลือดขาวที่พบ | | | |
|---------------------------------|-----------------------------|---------|---------|---------|
| | กรณีไม่มีเซลล์เม็ดเลือดขาว | 1 เซลล์ | 2 เซลล์ | 3 เซลล์ |
| จำนวนภาพ | 20 | 791 | 23 | 14 |
| จำนวนภาพที่สามารถระบุตำแหน่งได้ | 20 | 791 | 23 | 14 |



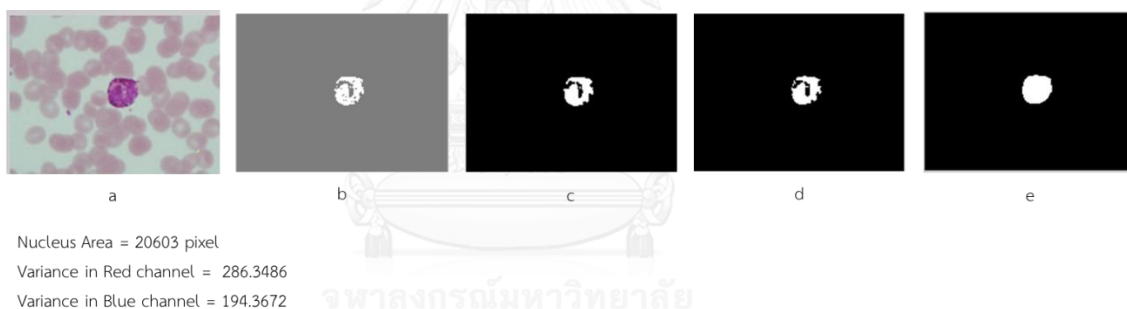
รูปที่ 5. 1 แสดงผลการระบุตำแหน่งนิวเคลียสเซลล์เม็ดเลือดขาว แถว A-D คือ กรณีไม่มีเซลล์เม็ดเลือดขาว, พบ 1 เซลล์, 2 เซลล์ และ 3 เซลล์ ตามลำดับ หลัก (a) คือ ภาพต้นฉบับ (b) คือภาพ unsigned integers 8 bit ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization (c) คือภาพ b ที่แปลงเป็นภาพขาวดำ

5.2. การแบ่งส่วนภาพเซลล์เม็ดเลือดขาวที่ละขั้นตอน

5.2.1. วิธีที่นำเสนอ

ในการแบ่งส่วนนิวเคลียสและเซลล์เม็ดเลือดขาวนั้น สามารถแยกวิธีเป็น 4 กลุ่ม คือ

1. การแบ่งส่วนนิวเคลียสและเซลล์เม็ดเลือดขาวชนิด basophil เท่านั้น
2. การแบ่งส่วนนิวเคลียสและเซลล์เม็ดเลือดขาวชนิดอื่น ๆ ที่เหลือ ที่เซลล์ไม่ได้ยู่ติดกับเซลล์อื่น ๆ
3. การแบ่งส่วนนิวเคลียสและเซลล์เม็ดเลือดขาวชนิดอื่น ๆ ที่เหลือ ที่มีเซลล์อื่น ๆ เซลล์อยู่ติดกันเพียงเล็กน้อย
4. การแบ่งส่วนนิวเคลียสและเซลล์เม็ดเลือดขาวชนิดอื่น ๆ ที่เหลือ ที่มีเซลล์อื่น ๆ เซลล์อยู่ติดกันจำนวนมาก

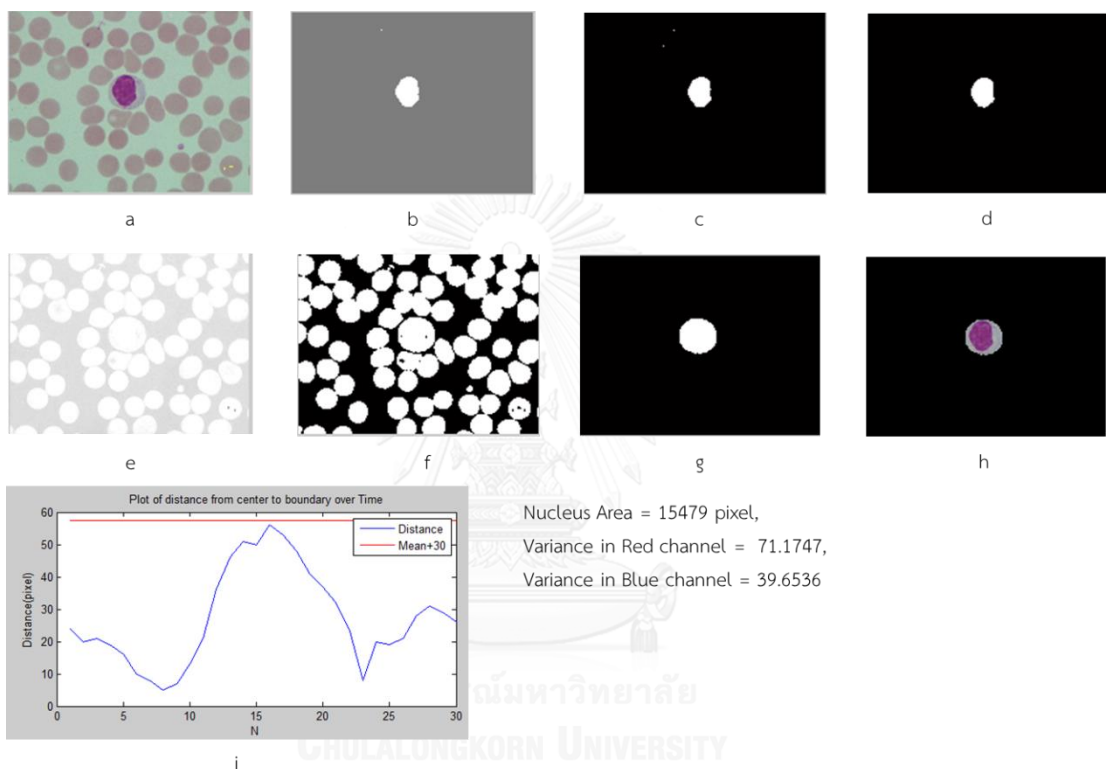


รูปที่ 5. 2 แสดงการแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ basophil

จากรูปที่ 5. 2 a คือภาพต้นฉบับ b คือภาพ unsigned integers 8 bit ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization c คือภาพที่แปลงเป็นภาพขาวดำ d คือภาพที่แสดงเฉพาะพื้นที่นิวเคลียส และ e คือ ภาพคอนเวกซ์ฮัลล์ของนิวเคลียส

กรณีที่ 1 ขั้นตอนการแบ่งส่วน basophil โดยการพิจารณาว่าเซลล์ใดคือ basophil จะพิจารณาขนาดของนิวเคลียส (ค่าต้องมากกว่า 20,000 พิกเซล) ค่าความแปรปรวนของสัญญาณในช่องสีแดง (ค่าต้องมากกว่า 200) และน้ำเงิน (ค่าต้องมากกว่า 93) ของพื้นที่นิวเคลียสที่หาได้ เซลล์ใดที่มีเงื่อนไขตรงตามที่ระบุจะพิจารณา คอนเวกซ์ฮัลล์ของภาพนิวเคลียสนั้น คือเซลล์ basophil ได้เลย แสดงดัง รูปที่ 5. 2

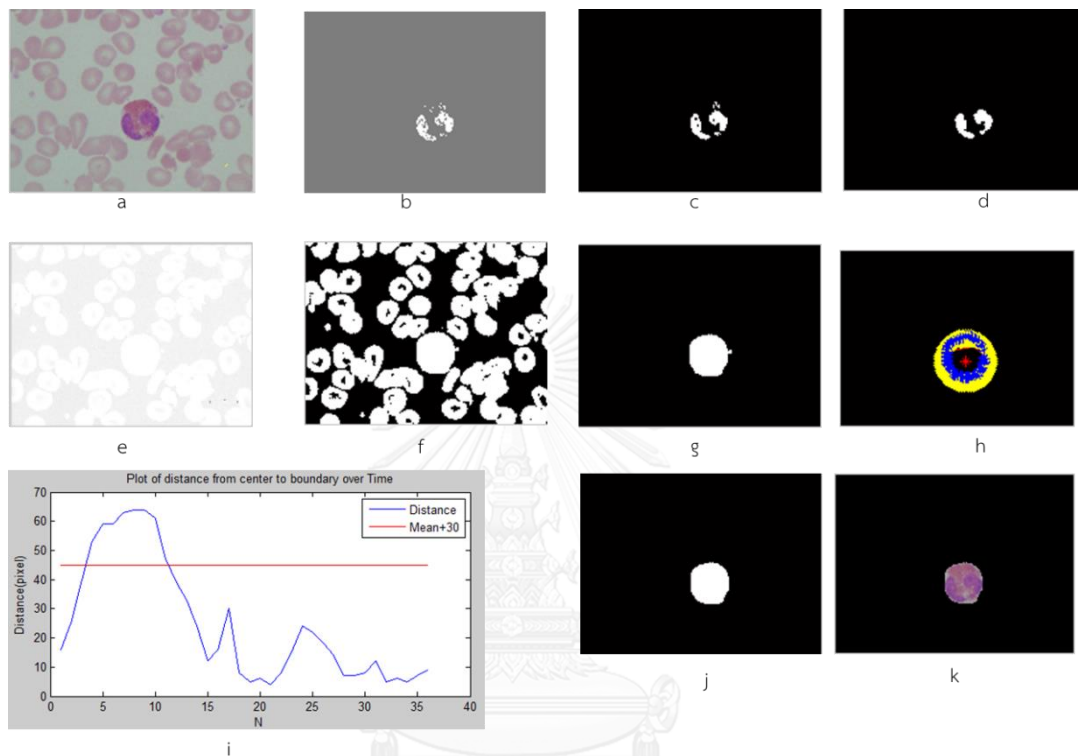
ในกรณีที่ 2 - 4 เซลล์ชนิดอื่น ๆ เมื่อได้พื้นที่ของนิวเคลียสแล้ว จะคัดกรองว่าเซลล์ใดคือ basophil ด้วย ขนาดของนิวเคลียส ค่าความแปรปรวนของสัญญาณในช่องสีแดง และสีน้ำเงิน จากพื้นที่นิวเคลียสที่หาได้ เซลล์ใดไม่ตรงตามเงื่อนไข จะพิจารณาแบ่งส่วนของเซลล์ต่อไป โดย กรณีที่ 2 เมื่อเซลล์เม็ดเลือดขาวที่สนใจไม่มีเซลล์อื่น ๆ มาติดกัน เมื่อตรวจสอบจากเงื่อนไขเซลล์ที่มาติดกันแล้ว พบว่าเป็นเซลล์เดี่ยว ๆ สามารถเลือกวัตถุที่สนใจได้เลย



รูปที่ 5. 3 แสดงการแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ lymphocyte

จากรูปที่ 5. 3 a คือ ภาพต้นฉบับ b คือภาพ unsigned integers 8 bit ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization c คือภาพ b ที่แปลงเป็นภาพขาวดำ d คือภาพที่แสดงเฉพาะพื้นที่นิวเคลียส e คือ ภาพ double precision ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization f คือภาพ e ที่แปลงเป็นภาพขาวดำ g คือภาพวัตถุที่สนใจ h คือภาพเซลล์ และ i คือกราฟแสดงระยะห่างจากศูนย์กลางของภาพ คอนเวกซ์ฮัลล์ของนิวเคลียส ไปยังขอบของวัตถุที่สนใจ (g) แสดงด้วยเส้นสีน้ำเงิน ส่วนเส้นสีแดงคือค่าเฉลี่ยของระยะห่างบวก 30 พิกเซล

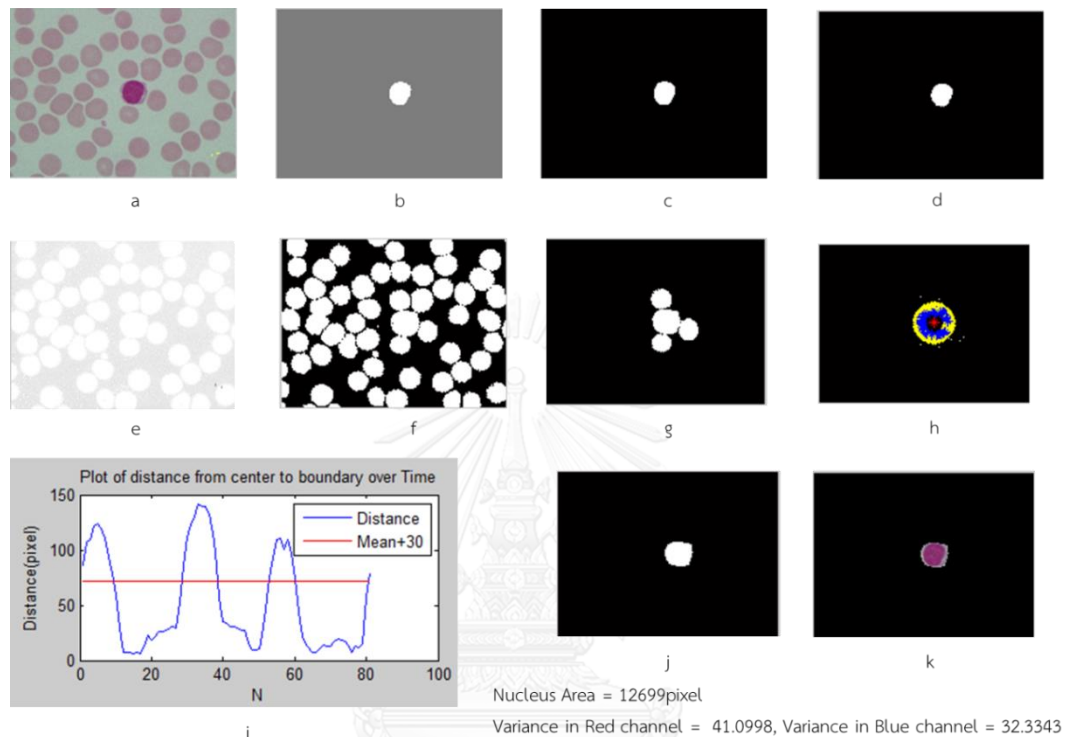
กรณีที่ 2 เช่นเดียวกันเมื่อเซลล์เม็ดเลือดขาวที่สนใจไม่มีเซลล์อื่น ๆ อยู่ติดกัน แต่เมื่อตรวจสอบจากเงื่อนไขเซลล์ที่ติดกันแล้วพบว่าไม่ใช่เซลล์เดี่ยว ๆ เลือกจุดที่เป็นขอบจากวิธีที่นำเสนอ และใช้การสร้างวงรีจาก วิธี least square fit ก่อนจะทำการ ‘AND’ กับวัตถุที่สนใจ



รูปที่ 5. 4 แสดงการแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ eosinophil

จากรูปที่ 5. 4 a คือ ภาพต้นฉบับ, b คือภาพ unsigned integers 8 bit ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization, c คือภาพ b ที่แปลงเป็นภาพขาวดำ, d คือภาพที่แสดงเฉพาะพื้นที่นิวเคลียส, e คือ ภาพ double precision ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization, f คือภาพ e ที่แปลงเป็นภาพขาวดำ, g คือภาพวัตถุที่สนใจ, h คือการสร้างเส้นรัศมีเพื่อหาจุดที่เป็นขอบของเซลล์เม็ดเลือดขาวของวัตถุที่สนใจ, i คือกราฟแสดงระยะห่างจากศูนย์กลางของภาพ คอนเวกซ์ฮัลล์ของนิวเคลียส ไปยังขอบของวัตถุที่สนใจ (g) แสดงด้วยเส้นสีน้ำเงิน ส่วนเส้นสีแดงคือค่าเฉลี่ยของระยะห่างบวก 30 พิกเซล, j คือ ภาพเซลล์ผลลัพธ์ที่ผ่านการ ‘AND’ ของวงรีที่สร้างโดยเทคนิค least square fit จากจุดที่ได้จากภาพ h กับ ภาพ g , และ k คือ ภาพเซลล์เม็ดเลือดขาวที่ต้องการ

กรณีที่ 3 เมื่อเซลล์เม็ดเลือดขาวที่สนใจมีเซลล์อื่นๆอยู่ติดกันเล็กน้อย และ เมื่อตรวจสอบจากเงื่อนไขเซลล์ที่มาติดกันแล้วพบว่าไม่ใช่เซลล์เดี่ยวๆ เลือกจุดที่เป็นขอบจากวิธีที่นำเสนอ และใช้การสร้างวงรีจาก วิธี least square fit ก่อนจะทำการ ‘AND’ กับวัตถุที่สนใจ

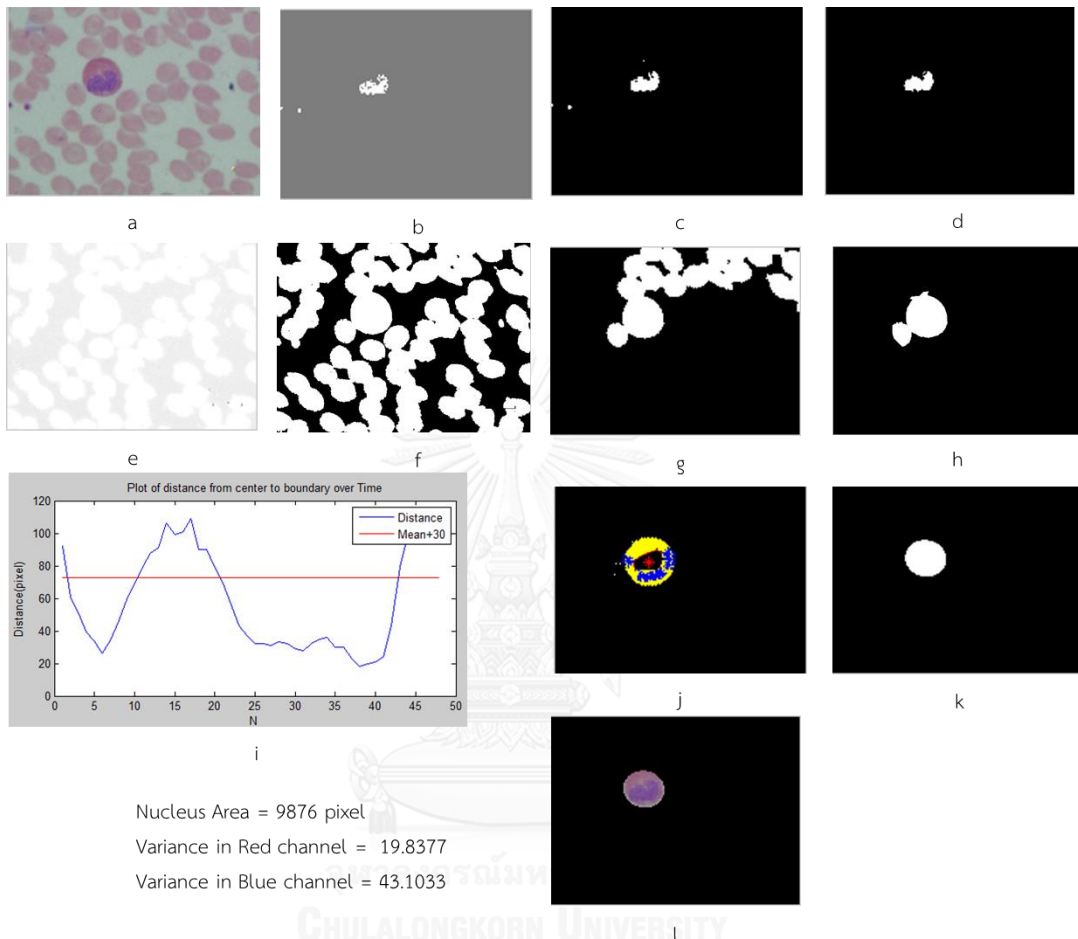


รูปที่ 5. 5 แสดง การแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ lymphocyte

จากรูปที่ 5. 5 a คือ ภาพต้นฉบับ, b คือภาพ unsigned integers 8 bit ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization, c คือภาพ b ที่แปลงเป็นภาพขาวดำ, d คือภาพที่แสดงเฉพาะพื้นที่นิวเคลียส, e คือ ภาพ double precision ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization, f คือภาพ e ที่แปลงเป็นภาพขาวดำ, g คือภาพวัตถุที่สนใจ, h คือการสร้างเส้นรัศมีเพื่อหาจุดที่เป็นขอบของเซลล์เม็ดเลือดขาวของวัตถุที่สนใจ, i คือกราฟแสดงระยะห่างจากศูนย์กลางของภาพ คอนเวกซ์ฮัลล์ของนิวเคลียส ไปยังขอบของวัตถุที่สนใจ (g) แสดงด้วยเส้นสีน้ำเงิน ส่วนเส้นสีแดงคือค่าเฉลี่ยของระยะห่างบวก 30 พิกเซล, j คือ ภาพเซลล์ผลลัพธ์ที่ผ่านการ ‘AND’ ของวงรีที่สร้างโดยเทคนิค least square fit จากจุดที่ได้จากภาพ h กับ ภาพ g, และ k คือ ภาพเซลล์เม็ดเลือดขาวที่ต้องการ

กรณีที่ 4 เมื่อเซลล์เม็ดเลือดขาวที่สนใจมีเซลล์อื่นๆอยู่ติดกัน และวัตถุที่สนใจมีพื้นที่มากกว่า 10% ของพื้นที่ภาพ ทำการเลือกตัดวัตถุที่สนใจออกมา ด้วยเทคนิค dilation และ เมื่อตรวจสอบจาก

เงื่อนไขเซลล์ที่มาติดกันแล้วพบว่าไม่ใช่เซลล์เดี่ยว ๆ เลือกจุดที่เป็นขอบจากวิธีที่นำเสนอ และใช้การ
สร้างวงรีจาก วิธี least square fit ก่อนจะทำการ 'AND' กับวัตถุที่สนใจ

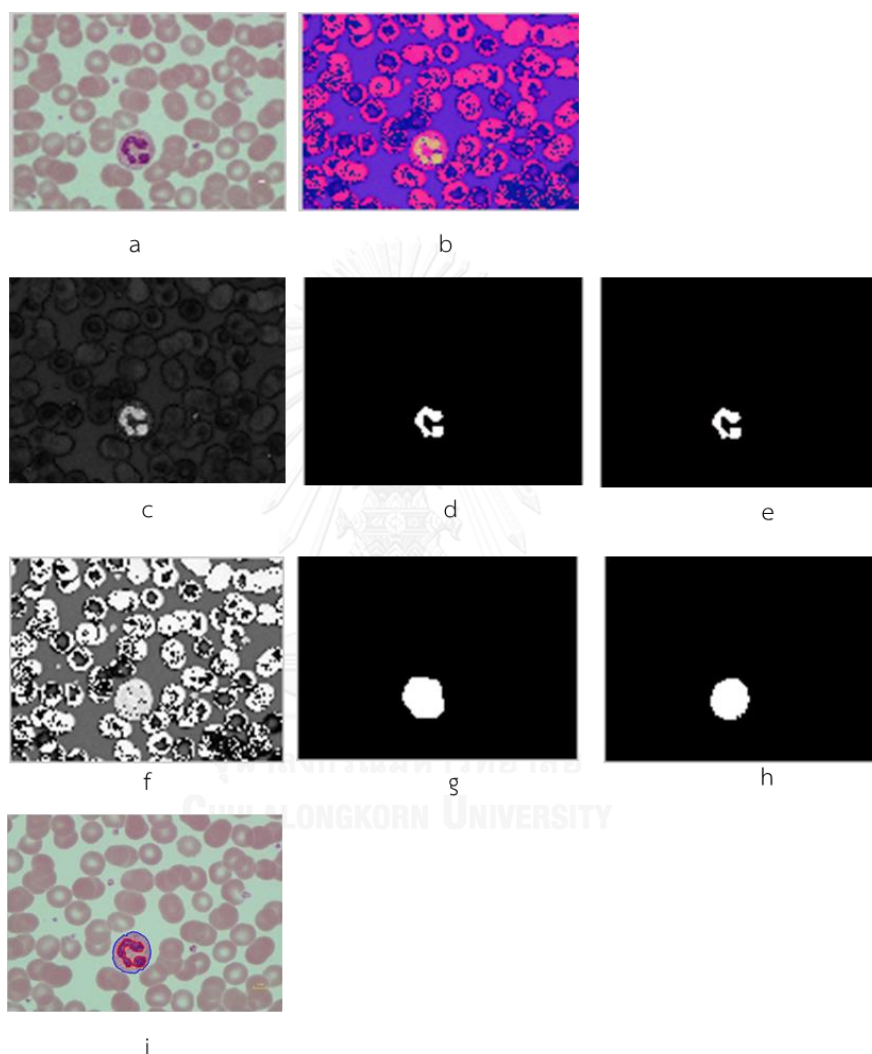


รูปที่ 5. 6 แสดง การแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ eosinophil

จากรูปที่ 5. 6 a คือ ภาพต้นฉบับ, b คือภาพ unsigned integers 8 bit ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization, c คือภาพ b ที่แปลงเป็นภาพขาวดำ, d คือภาพที่แสดงเฉพาะพื้นที่นิวเคลียส, e คือ ภาพ double precision ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization, f คือภาพ e ที่แปลงเป็นภาพขาวดำ, g คือภาพวัตถุที่สนใจ, h คือภาพวัตถุที่สนใจที่ผ่านการ dilation ด้วย structure element disk 90 พิกเซล, i คือกราฟแสดงระยะห่างจากศูนย์กลางของภาพ คอนเวกซ์ฮัลล์ของนิวเคลียส ไปยังขอบของวัตถุที่สนใจ(g) แสดงด้วยเส้นสีน้ำเงิน ส่วนเส้นสีแดงคือค่าเฉลี่ยของระยะห่างบวก 30 พิกเซล, j คือการสร้างเส้นรัศมีเพื่อหาจุดที่เป็นขอบของเซลล์เม็ดเลือดขาวของวัตถุที่สนใจ, k คือ ภาพเซลล์ผลลัพธ์ที่ผ่านการ 'AND' ของ

วงรีที่สร้างโดยเทคนิค least square fit จากจุดที่ได้จากภาพ h กับ ภาพ g, และ l คือ ภาพเซลล์เม็ดเลือดขาวที่ต้องการ

5.2.2. วิธีใช้ ACWE1

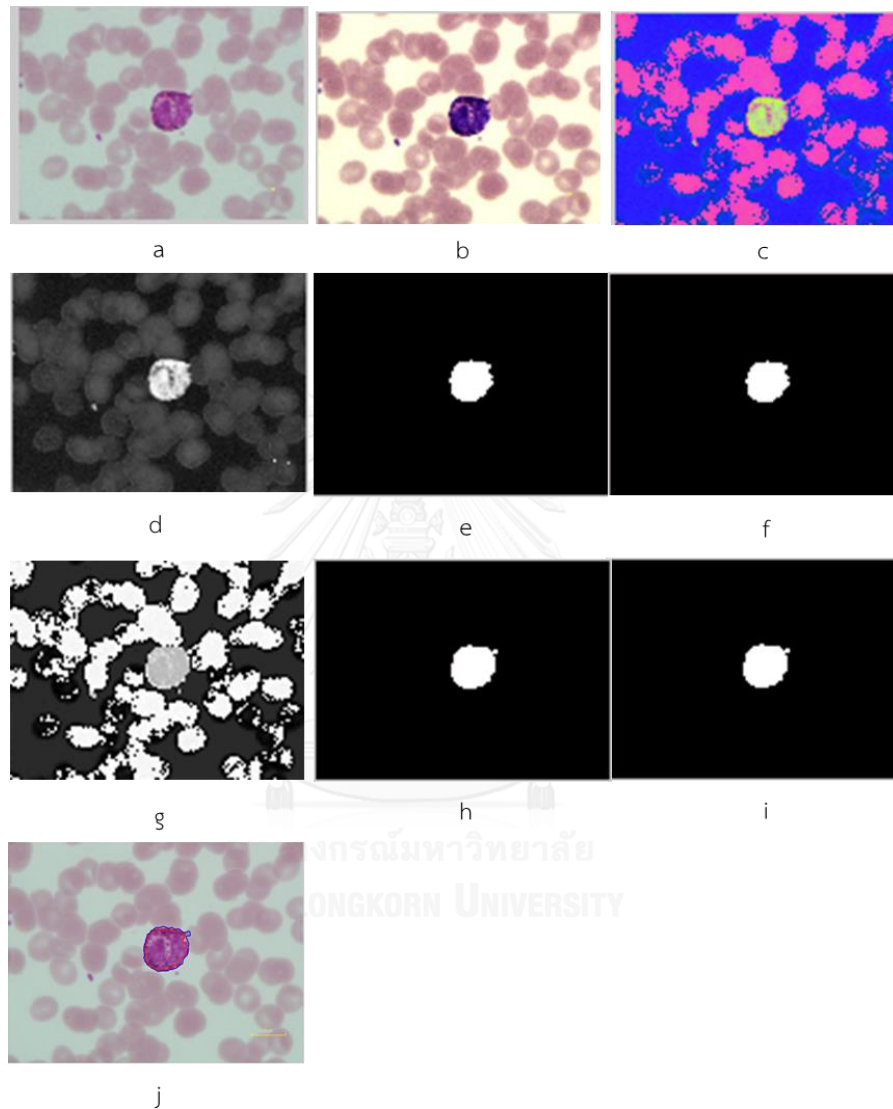


รูปที่ 5. 7 แสดงการแบ่งส่วนนิวเคลียสและเซลล์ด้วยวิธี ACWE1

จากรูปที่ 5. 7 a คือภาพต้นฉบับ, b คือภาพที่แปลงจากปริภูมิสี RGB (ภาพ a) ไปเป็น HSV, c คือ ภาพขาวเทาในช่อง Saturation ของ b, d คือ คอนทัวร์เริ่มต้นของการแบ่งส่วนนิวเคลียสด้วย ACWE1, e คือ ภาพผลลัพธ์ที่ได้จากการแบ่งส่วนนิวเคลียส, f คือ ภาพขาวเทาในช่อง hue ของ b, g คือ คอนทัวร์เริ่มต้นของการแบ่งส่วนเซลล์ด้วย ACWE1, h คือ ภาพผลลัพธ์ที่ได้จาก

การแบ่งส่วนเซลล์, และ i คือ ผลลัพธ์ของการแบ่งส่วนนิวเคลียสและเซลล์ โดยเส้นสีแดงคือขอบของนิวเคลียส และเส้นสีน้ำเงินคือขอบของเซลล์

5.2.3. วิธีใช้ ACWE2



รูปที่ 5. 8 แสดงการแบ่งส่วนนิวเคลียสและเซลล์ตามวิธี ACWE2

จากรูปที่ 5. 8 a คือภาพต้นฉบับ, b คือภาพที่ผ่านการแปลงโดยวิธี Reinhard และคณะ [19], c คือภาพที่แปลงจากปริภูมิสี RGB (ภาพ b) ไปเป็น HSV, d คือ ภาพขาวเทาในช่อง Saturation ของ c, e คือ คอนทัวร์เริ่มต้นของการแบ่งส่วนนิวเคลียสด้วยแอกทิฟคอนทัวร์, f คือ ภาพผลลัพธ์ที่ได้จากการแบ่งส่วนนิวเคลียส, g คือ ภาพขาวเทาในช่อง hue ของ c, h คือ คอนทัวร์

เริ่มต้นของการแบ่งส่วนเซลล์ด้วย ACWE2, i คือ ภาพผลลัพธ์ที่ได้จากการแบ่งส่วนเซลล์, และ j คือ ผลลัพธ์ของการแบ่งส่วนนิวเคลียสและเซลล์ โดยเส้นสีแดงคือขอบของนิวเคลียส และเส้นสีน้ำเงินคือขอบของเซลล์ สำหรับการแบ่งส่วนชุดข้อมูลที่ 2 ทำเช่นเดียวกับชุดข้อมูลที่ 1 แต่ทำการปรับเทียบค่าความยาวและพื้นที่ที่นำมาโปรแกรมตามความละเอียดของภาพ ในภาคผนวก ข

5.3.การเปรียบเทียบการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี ACWE1 กับการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยผู้เชี่ยวชาญ

ผลการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี ACWE1 โดยแสดงค่า Dice similarity (DS), false false positive ratio (RFP) และ false negative ratio (RFN) เฉลี่ยของเซลล์ทั้งห้าชนิด จากข้อมูลการการแบ่งส่วนนิวเคลียสและการแบ่งส่วนเซลล์เม็ดเลือดขาว ด้วยวิธี ACWE1 เปรียบเทียบกับการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยผู้เชี่ยวชาญ แสดงในตารางที่ 5. 2 – 5.3 และ

ตารางที่ 5. 4 - 5.5 สำหรับข้อมูลชุดที่ 2 ตามลำดับ

หมายเหตุ ในข้อมูลชุดที่ 1 เซลล์ lymphocyte, monocyte และ neutrophil มีตัวเลข 2 ชุดบวกกัน ตัวเลขแรกคือ กรณีภาพที่มีเซลล์ชนิดนั้นเพียงเซลล์เดียว ตัวเลขตัวที่สอง คือ กรณีภาพที่มีหลายเซลล์ในหนึ่งภาพ ผู้วิจัยได้แยกเซลล์แต่ละเซลล์ที่พบออกเป็นชนิดต่าง ๆ เพื่อให้ง่ายต่อการหาความคล้ำยคลึงของเซลล์แต่ละชนิด

ตารางที่ 5. 2 แสดงผลการแบ่งส่วนนิวเคลียส ข้อมูลชุดที่ 1

| ชนิด | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------|--------|-----------|--------|--------|----------------------|--------|--------|
| | | DS | RFP | RFN | DS | RFP | RFN |
| basophil | 69 | 0.7044 | 0.0046 | 0.1741 | 0.3152 | 0.0033 | 0.3633 |
| eosinophil | 148 | 0.3398 | 0.1333 | 0.3278 | 0.3439 | 0.2690 | 0.4674 |
| lymphocyte | 141+34 | 0.6269 | 0.0300 | 0.3318 | 0.4678 | 0.1174 | 0.4444 |
| monocyte | 107+6 | 0.6862 | 0.0190 | 0.2357 | 0.4013 | 0.0797 | 0.4020 |
| neutrophil | 326+48 | 0.7336 | 0.0309 | 0.2317 | 0.3944 | 0.1157 | 0.3404 |

ตารางที่ 5.3 แสดงผลการแบ่งส่วนเซลล์ ข้อมูลชุดที่ 1

| ชนิด | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------|--------|-----------|--------|--------|----------------------|--------|--------|
| | | DS | RFP | RFN | DS | RFP | RFN |
| basophil | 69 | 0.5503 | 0.0246 | 0.1761 | 0.2666 | 0.0148 | 0.3618 |
| eosinophil | 148 | 0.4187 | 0.1716 | 0.3382 | 0.4127 | 0.3311 | 0.4563 |
| lymphocyte | 141+34 | 0.5171 | 0.0515 | 0.3975 | 0.4147 | 0.1712 | 0.4166 |
| monocyte | 107+6 | 0.5168 | 0.0495 | 0.3070 | 0.3023 | 0.1320 | 0.3741 |
| neutrophil | 326+48 | 0.6365 | 0.0591 | 0.2510 | 0.3514 | 0.1794 | 0.3356 |

สำหรับข้อมูลชุดที่ 1 วิธี ACWE1 ให้ค่าความถูกต้องของการแบ่งส่วนนิวเคลียสเฉลี่ย 0.6182 โดยเซลล์ที่มีค่าต่ำสุด คือ eosinophil (0.3398) และ เซลล์ที่มีค่าสูงสุด คือ neutrophil (0.7336) โดยค่า false positive ratio (RFP) เฉลี่ย มีค่าน้อยกว่า 0.1333 และค่า false Negative ratio (RFN) เฉลี่ย มีค่าน้อยกว่า 0.3318 และ ให้ค่าความถูกต้องของการแบ่งส่วนเซลล์เฉลี่ย 0.5279 โดยเซลล์ที่มีค่าต่ำสุด คือ eosinophil (0.4187) และ เซลล์ที่มีค่าสูงสุด คือ basophil (0.5503) โดยค่า false positive ratio (RFP) เฉลี่ย มีค่าน้อยกว่า 0.1333 และค่า false negative ratio (RFN) เฉลี่ย มีค่าน้อยกว่า 0.3975

ตารางที่ 5.4 แสดงผลการแบ่งส่วนนิวเคลียส ข้อมูลชุดที่ 2

| ชนิด | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------|-------|-----------|--------|--------|----------------------|--------|--------|
| | | DS | RFP | RFN | DS | RFP | RFN |
| basophil | 1 | 0.8950 | 0.0077 | 0.0711 | - | - | - |
| eosinophil | 5 | 0.8269 | 0.0218 | 0.0297 | 0.1029 | 0.0143 | 0.0186 |
| lymphocyte | 138 | 0.8370 | 0.0012 | 0.1582 | 0.3425 | 0.0017 | 0.3460 |
| monocyte | 42 | 0.8465 | 0.0323 | 0.0746 | 0.1484 | 0.0520 | 0.0666 |
| neutrophil | 271 | 0.9681 | 0.0013 | 0.0532 | 0.0468 | 0.0161 | 0.0372 |

ตารางที่ 5.5 แสดงผลการแบ่งส่วนเซลล์ ข้อมูลชุดที่ 2

| ชนิด | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------|-------|-----------|--------|--------|----------------------|--------|--------|
| | | DS | RFP | RFN | DS | RFP | RFN |
| basophil | 1 | 0.5939 | 0.1103 | 0.0239 | - | - | - |
| eosinophil | 5 | 0.7963 | 0.0651 | 0.0337 | 0.1054 | 0.0458 | 0.0088 |
| lymphocyte | 138 | 0.7156 | 0.0465 | 0.1545 | 0.3474 | 0.1010 | 0.3478 |
| monocyte | 42 | 0.7510 | 0.1224 | 0.0392 | 0.0869 | 0.0563 | 0.0275 |
| neutrophil | 271 | 0.8448 | 0.0440 | 0.0732 | 0.1231 | 0.0546 | 0.0652 |

ในข้อมูลภาพชุดที่สอง วิธี ACWE1 ให้ค่าความถูกต้องของการแบ่งส่วนนิวเคลียสเฉลี่ย 0.8747 โดยเซลล์ที่มีค่าต่ำสุด คือ eosinophil (0.8269) และ เซลล์ที่มีค่าสูงสุด คือ neutrophil (0.9681) โดยค่า false positive ratio (RFP) เฉลี่ย มีค่าน้อยกว่า 0.0323 และค่า false Negative ratio (RFN) เฉลี่ย มีค่าน้อยกว่า 0.1582 และ ให้ค่าความถูกต้องของการแบ่งส่วนเซลล์เฉลี่ย 0.7403 โดยเซลล์ที่มีค่าต่ำสุด คือ basophil (0.5939) และ เซลล์ที่มีค่าสูงสุด คือ neutrophil (0.8448) โดยค่า false positive ratio (RFP) เฉลี่ย มีค่าน้อยกว่า 0.1224 และค่า false negative ratio (RFN) เฉลี่ย มีค่าน้อยกว่า 0.1545

5.4. การเปรียบเทียบการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี ACWE2 กับการแบ่งส่วน

ภาพเซลล์เม็ดเลือดขาวด้วยผู้เชี่ยวชาญ

ผลการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี ACWE2 โดยแสดงค่า Dice similarity (DS), false positive ratio (RFP) และ false negative ratio (RFN) เฉลี่ยของเซลล์ทั้งห้าชนิด จากข้อมูลการการแบ่งส่วนนิวเคลียสและการแบ่งส่วนเซลล์เม็ดเลือดขาว ด้วยวิธี ACWE2 ที่นำเสนอโดย Mathur และคณะ [18] เปรียบเทียบกับการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยผู้เชี่ยวชาญ แสดงในตารางที่ 5.6 – 5.7 สำหรับข้อมูลชุดที่ 1 และ ตาราง 5.8 - 5.9 สำหรับข้อมูลชุดที่ 2 ตามลำดับ

ตารางที่ 5. 6 แสดงผลการแบ่งส่วนนิวเคลียส ข้อมูลชุดที่ 1

| ชนิด | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------|--------|-----------|--------|--------|----------------------|--------|--------|
| | | DS | RFP | RFN | DS | RFP | RFN |
| basophil | 69 | 0.8042 | 0.0051 | 0.0605 | 0.1867 | 0.0029 | 0.2049 |
| eosinophil | 148 | 0.3151 | 0.0068 | 0.4842 | 0.3113 | 0.0068 | 0.4998 |
| lymphocyte | 141+34 | 0.9370 | 0.0003 | 0.0733 | 0.1900 | 0.0023 | 0.1805 |
| monocyte | 107+6 | 0.8028 | 0.0020 | 0.1666 | 0.3321 | 0.0036 | 0.3384 |
| neutrophil | 326+48 | 0.8954 | 0.0002 | 0.1329 | 0.2258 | 0.0010 | 0.2192 |

ตารางที่ 5. 7 แสดงผลการแบ่งส่วนเซลล์ ข้อมูลชุดที่ 1

| ชนิด | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------|--------|-----------|--------|--------|----------------------|--------|--------|
| | | DS | RFP | RFN | DS | RFP | RFN |
| basophil | 69 | 0.7596 | 0.0343 | 0.0761 | 0.2952 | 0.0896 | 0.2007 |
| eosinophil | 148 | 0.3970 | 0.0397 | 0.4996 | 0.4491 | 0.1210 | 0.4891 |
| lymphocyte | 141+34 | 0.8705 | 0.0035 | 0.1544 | 0.1919 | 0.0266 | 0.1826 |
| monocyte | 107+6 | 0.6292 | 0.0247 | 0.2799 | 0.2951 | 0.0622 | 0.3065 |
| neutrophil | 326+48 | 0.7319 | 0.0054 | 0.3011 | 0.247 | 0.375 | 0.256 |

สำหรับข้อมูลชุดที่ 1 วิธี ACWE2 นำเสนอโดย Mathur และคณะ [18] ให้ค่าความถูกต้องของการแบ่งส่วนนิวเคลียสเฉลี่ย 0.7509 โดยเซลล์ที่มีค่าต่ำสุด คือ eosinophil (0.3151) และ เซลล์ที่มีค่าสูงสุด คือ lymphocyte (0.9370) โดยค่า false positive ratio (RFP) เฉลี่ย มีค่าน้อยกว่า 0.0068 และค่า false negative ratio (RFN) เฉลี่ย มีค่าน้อยกว่า 0.4840 และ ให้ค่าความถูกต้องของการแบ่งส่วนเซลล์เฉลี่ย 0.6776 โดยเซลล์ที่มีค่าต่ำสุด คือ eosinophil (0.3970) และ เซลล์ที่มีค่าสูงสุด คือ lymphocyte (0.8705) โดยค่า false positive ratio (RFP) เฉลี่ย มีค่าน้อยกว่า 0.0343 และค่า false negative ratio (RFN) เฉลี่ย มีค่าน้อยกว่า 0.4996

ตารางที่ 5. 8 แสดงผลการแบ่งส่วนนิวเคลียส ข้อมูลชุดที่ 2

| ชนิด | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------|-------|-----------|--------|--------|----------------------|--------|--------|
| | | DS | RFP | RFN | DS | RFP | RFN |
| basophil | 1 | 0.8623 | 0.0155 | 0.0173 | - | - | - |
| eosinophil | 5 | 0.3264 | 0.0105 | 0.6093 | 0.4474 | 0.0163 | 0.5350 |
| lymphocyte | 138 | 0.7726 | 0.0013 | 0.2269 | 0.3911 | 0.0024 | 0.3916 |
| monocyte | 42 | 0.2225 | 0.0139 | 0.7190 | 0.3617 | 0.0271 | 0.4497 |
| neutrophil | 271 | 0.6115 | 0.0004 | 0.4203 | 0.4445 | 0.0041 | 0.4243 |

ตารางที่ 5. 9 แสดงผลการแบ่งส่วนเซลล์ ข้อมูลชุดที่ 2

| ชนิด | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------|-------|-----------|--------|--------|----------------------|--------|--------|
| | | DS | RFP | RFN | DS | RFP | RFN |
| basophil | 1 | 0.9571 | 0.0011 | 0.0708 | - | - | - |
| eosinophil | 5 | 0.3449 | 0.0042 | 0.6703 | 0.4724 | 0.0086 | 0.4524 |
| lymphocyte | 138 | 0.7450 | 0.0073 | 0.2314 | 0.3850 | 0.0148 | 0.3923 |
| monocyte | 42 | 0.2336 | 0.0144 | 0.7546 | 0.3774 | 0.0292 | 0.3944 |
| neutrophil | 271 | 0.4997 | 0.0087 | 0.5545 | 0.3728 | 0.0197 | 0.3432 |

ในข้อมูลภาพชุดที่สอง วิธี ACWE2 นำเสนอโดย Mathur และคณะ [18] ให้ค่าความถูกต้องของการแบ่งส่วนนิวเคลียสเฉลี่ย 0.5590 โดยเซลล์ที่มีค่าต่ำสุด คือ monocyte (0.2225) และ เซลล์ที่มีค่าสูงสุด คือ basophil (0.8623) โดยค่า false positive ratio (RFP) เฉลี่ย มีค่าน้อยกว่า 0.0155 และค่า false negative ratio (RFN) เฉลี่ย มีค่าน้อยกว่า 0.7190 และให้ค่าความถูกต้องของการแบ่งส่วนเซลล์เฉลี่ย 0.5561 โดยเซลล์ที่มีค่าต่ำสุด คือ monocyte (0.2336) และ เซลล์ที่มีค่าสูงสุด คือ basophil (0.9571) โดยค่า false positive ratio (RFP) เฉลี่ย มีค่าน้อยกว่า 0.0144 และค่า false negative ratio (RFN) เฉลี่ย มีค่าน้อยกว่า 0.7546

5.5.การเปรียบเทียบการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธีที่นำเสนอกับการแบ่งส่วน

ภาพเซลล์เม็ดเลือดขาวด้วยผู้เชี่ยวชาญ

ผลการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธีที่นำเสนอ โดยแสดงค่า Dice similarity (DS), false positive ratio (RFP) และ false negative ratio (RFN) เฉลี่ยของเซลล์ทั้งห้าชนิด จากข้อมูลการการแบ่งส่วนนิวเคลียสและการแบ่งส่วนเซลล์เม็ดเลือดขาว ด้วยวิธีที่นำเสนอแสดงใน ตารางที่ 5. 10 – 5.11 สำหรับข้อมูลชุดที่ 1 และ ตารางที่ 5. 12 – 5.13 สำหรับข้อมูลชุดที่ 2 ตามลำดับ

ตารางที่ 5. 10 แสดงผลการแบ่งส่วนนิวเคลียส ข้อมูลชุดที่ 1

| ชนิด | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------|--------|-----------|--------|--------|----------------------|--------|--------|
| | | DS | RFP | RFN | DS | RFP | RFN |
| basophil | 69 | 0.8028 | 0.0012 | 0.2326 | 0.2170 | 0.0012 | 0.2575 |
| eosinophil | 148 | 0.8819 | 0.0010 | 0.1215 | 0.1583 | 0.0021 | 0.1947 |
| lymphocyte | 141+34 | 0.9706 | 0.0000 | 0.0549 | 0.0146 | 0.0001 | 0.0243 |
| monocyte | 107+6 | 0.7868 | 0.0002 | 0.3077 | 0.1940 | 0.0010 | 0.2451 |
| neutrophil | 326+48 | 0.9342 | 0.0000 | 0.1182 | 0.0309 | 0.0002 | 0.0495 |

ตารางที่ 5. 11 แสดงผลการแบ่งส่วนเซลล์ ข้อมูลชุดที่ 1

| ชนิด | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------|--------|-----------|--------|--------|----------------------|--------|--------|
| | | DS | RFP | RFN | DS | RFP | RFN |
| basophil | 69 | 0.8273 | 0.0058 | 0.1062 | 0.2037 | 0.0061 | 0.2336 |
| eosinophil | 148 | 0.9349 | 0.0022 | 0.0455 | 0.1104 | 0.0030 | 0.1228 |
| lymphocyte | 141+34 | 0.9263 | 0.0009 | 0.0950 | 0.0782 | 0.0019 | 0.1013 |
| monocyte | 107+6 | 0.8869 | 0.0040 | 0.1155 | 0.1125 | 0.0058 | 0.1507 |
| neutrophil | 326+48 | 0.9200 | 0.0018 | 0.0952 | 0.0587 | 0.0025 | 0.0854 |

ในข้อมูลชุดที่ 1 พบว่าวิธีที่นำเสนอให้ค่าความถูกต้องของการแบ่งส่วนนิวเคลียสเฉลี่ย (Dice similarity) 0.8753 โดยเซลล์ที่มีค่าต่ำสุด คือ basophil (0.8028) และ เซลล์ที่มีค่าสูงสุด คือ lymphocyte (0.9706) โดยค่า false positive ratio (RFP) เฉลี่ย มีค่าน้อยกว่า 0.0012 และค่า false negative ratio (RFN) เฉลี่ยมีค่าน้อยกว่า 0.3077 ค่าความถูกต้องของการแบ่งส่วนเซลล์เฉลี่ย (Dice similarity) 0.8991 โดยเซลล์ที่มีค่าต่ำสุด คือ basophil (0.8273) และ เซลล์ที่มีค่าสูงสุด คือ eosinophil (0.9349) โดยค่า false positive ratio (RFP) เฉลี่ย มีค่าน้อยกว่า 0.0058 และค่า false negative ratio (RFN) เฉลี่ยมีค่าน้อยกว่า 0.1155

ตารางที่ 5. 12 แสดงผลการแบ่งส่วนนิวเคลียส ข้อมูลชุดที่ 2

| ชนิด | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------|-------|-----------|--------|--------|----------------------|--------|--------|
| | | DS | RFP | RFN | DS | RFP | RFN |
| basophil | 1 | 0.8857 | 0.0113 | 0.0330 | - | - | - |
| eosinophil | 5 | 0.9233 | 0.0064 | 0.0367 | 0.0702 | 0.0072 | 0.0304 |
| lymphocyte | 138 | 0.9717 | 0.0026 | 0.0047 | 0.0325 | 0.0031 | 0.0091 |
| monocyte | 42 | 0.8971 | 0.0022 | 0.1569 | 0.0800 | 0.0014 | 0.1245 |
| neutrophil | 271 | 0.9940 | 0.0000 | 0.0107 | 0.0229 | 0.0003 | 0.0387 |

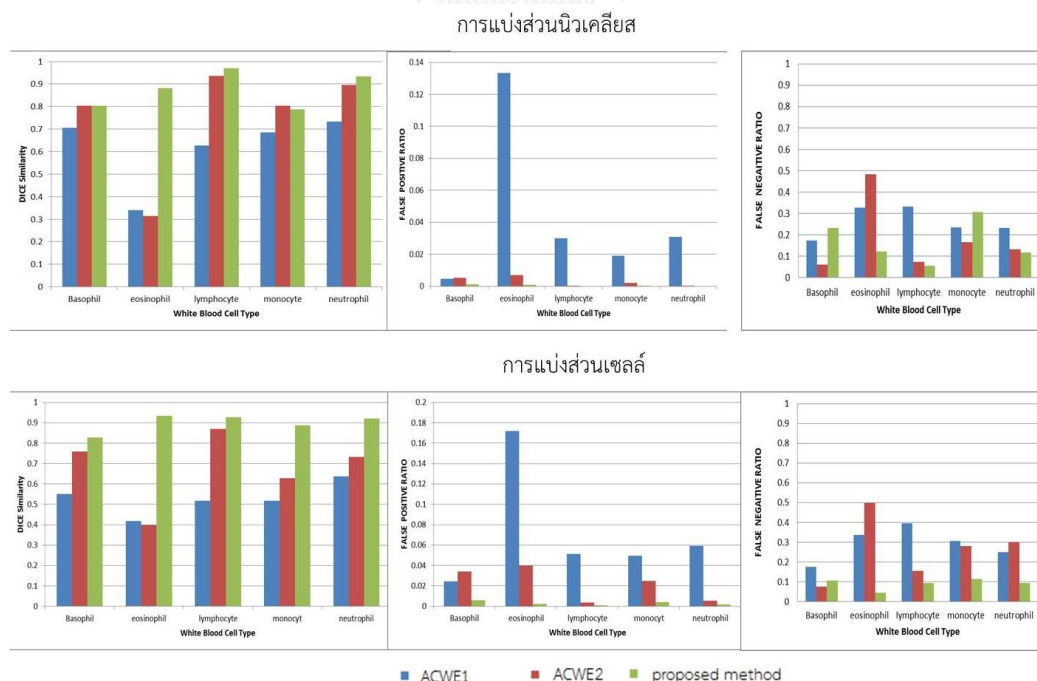
ตารางที่ 5. 13 แสดงผลการแบ่งส่วนเซลล์ ข้อมูลชุดที่ 2

| ชนิด | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------|-------|-----------|--------|--------|----------------------|--------|--------|
| | | DS | RFP | RFN | DS | RFP | RFN |
| basophil | 1 | 0.9696 | 0.0033 | 0.0221 | - | - | - |
| eosinophil | 5 | 0.9602 | 0.0080 | 0.0214 | 0.0297 | 0.0081 | 0.0094 |
| lymphocyte | 138 | 0.9609 | 0.0045 | 0.0206 | 0.0446 | 0.0069 | 0.0471 |
| monocyte | 42 | 0.9233 | 0.0143 | 0.0789 | 0.0672 | 0.0201 | 0.0965 |
| neutrophil | 271 | 0.9638 | 0.0055 | 0.0330 | 0.0357 | 0.0080 | 0.0433 |

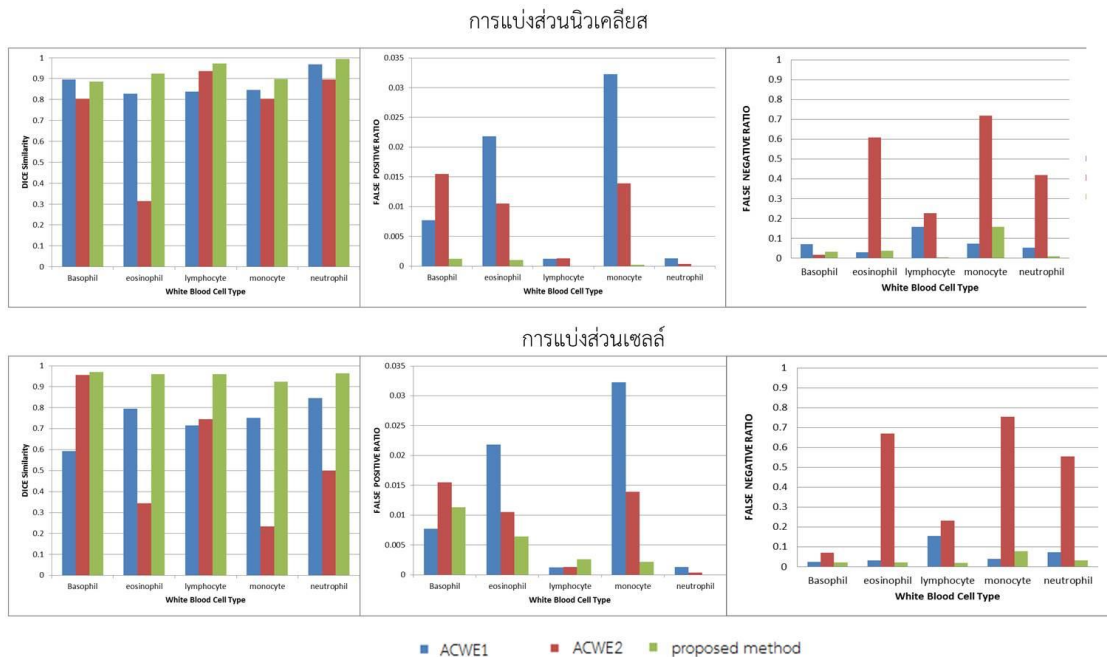
ในข้อมูลภาพชุดที่สอง วิธีที่นำเสนอให้ค่าความถูกต้องของการแบ่งส่วนนิวเคลียสเฉลี่ย (Dice similarity) 0.9344 โดยเซลล์ที่มีค่าต่ำสุด คือ basophil (0.8857) และ เซลล์ที่มีค่าสูงสุด คือ neutrophil (0.9940) โดยค่า false positive ratio (RFP) เฉลี่ย มีค่าน้อยกว่า 0.0113 และค่า false negative ratio (RFN) เฉลี่ย มีค่าน้อยกว่า 0.1569 และ ค่าความถูกต้องของการแบ่งส่วนเซลล์เฉลี่ย (Dice similarity) 0.9556 โดยเซลล์ที่มีค่าต่ำสุด คือ monocyte (0.9233) และ เซลล์ที่มีค่าสูงสุด คือ neutrophil (0.9638) โดยค่า false positive ratio (RFP) เฉลี่ย มีค่าน้อยกว่า 0.0143 และค่า false negative ratio (RFN) เฉลี่ย มีค่าน้อยกว่า 0.0789

5.6. การเปรียบเทียบการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธีที่นำเสนอและ วิธี ACWE1 และ ACWE2 ในแง่ของความถูกต้อง

จากข้อมูลความถูกต้องของการแบ่งส่วนนิวเคลียสและเซลล์เม็ดเลือดขาวของทั้งสามวิธีในหัวข้อ 5.3 -5.5 สามารถนำผลลัพธ์ที่ได้มาเปรียบเทียบกัน ในด้านของความถูกต้องของการแบ่งส่วนภาพ (Dice similarity) FPR ซึ่งเป็นตัวบ่งชี้ว่ามีการแบ่งส่วนเกินจากพื้นที่ที่ต้องการมากน้อยเพียงไร (over segmentation) และค่า FNR ซึ่งเป็นตัวบ่งชี้ว่ามีการแบ่งส่วนเข้าไปในพื้นที่ที่ต้องการมากน้อยเพียงไร (under segmentation)



รูปที่ 5.9 แสดงการเปรียบเทียบการแบ่งส่วนภาพนิวเคลียสและเซลล์ด้วยวิธี ACWE1 และ ACWE2 และ วิธีที่นำเสนอของข้อมูลภาพชุดที่หนึ่ง



รูปที่ 5. 10 แสดงการเปรียบเทียบการแบ่งส่วนภาพนิวเคลียสและเซลล์ด้วยวิธี ACWE1 และ ACWE2 และ วิธีที่นำเสนอของข้อมูลภาพชุดที่สอง

5.7.การเปรียบเทียบการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธีที่นำเสนอและ วิธี ACWE1 และ ACWE2 ในแง่ของเวลาในการประมวลผลและผลการแบ่งส่วน

เงื่อนไขของเวลาในกรณีวิธี ACWE1 และ ACWE2 คือ เวลาเฉลี่ยของเซลล์แต่ละชนิดที่ใช้ในการแบ่งส่วนนิวเคลียสและเซลล์จะตั้งแต่เริ่มวางคอนทัวร์เริ่มต้นจนถึงเวลาที่คอนทัวร์หยุดคือเมื่อไม่มีการเปลี่ยนแปลงของคอนทัวร์ในรอบปัจจุบันเทียบกับรอบก่อนหน้า หรือเมื่อถึงรอบที่กำหนด ในกรณีของนิวเคลียสคือ 50 รอบ และในกรณีของเซลล์เม็ดเลือดขาวคือ 500 รอบ ส่วนเวลารวมคือเวลาตั้งแต่การทำ preprocessing จนเสร็จสิ้นกระบวนการ โดยนำเสนอในรูปแบบของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ส่วนการนำเสนอในรูปแบบของแผนภาพกล่อง (box and whisker plots) จะนำเสนอแยกกันในแต่ละวิธี เนื่องจากสเกลทางเวลาต่างกันมาก

ตารางที่ 5. 14 แสดงจำนวนรอบและเวลาเฉลี่ยที่ใช้การแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี

ACWE1

| ชนิด | จำนวน | ค่าเฉลี่ย | | | | |
|----------------|-------|--------------------|-----------------|------------|--------------|---------------|
| | | Nucleus iterations | Cell iterations | t_nu (sec) | t_cell (sec) | t_total (sec) |
| basophil | 69 | 6.449 | 498.449 | 2.687 | 160.968 | 325.155 |
| eosinophil | 148 | 13.546 | 493.230 | 5.090 | 161.335 | 328.349 |
| lymphocyte | 141 | 4.333 | 496.461 | 2.405 | 167.593 | 338.316 |
| monocyte | 107 | 11.605 | 500.000 | 4.805 | 170.427 | 346.371 |
| neutrophil | 326 | 7.258 | 491.555 | 3.210 | 158.308 | 362.842 |
| multiple cells | 37 | 4.000 | 494.649 | 1.841 | 155.835 | 314.006 |

ตารางที่ 5. 15 แสดงส่วนเบี่ยงเบนมาตรฐานของจำนวนรอบและเวลาที่ใช้การแบ่งส่วนภาพเซลล์เม็ด

เลือดขาวด้วยวิธี ACWE1

| ชนิด | จำนวน | ส่วนเบี่ยงเบนมาตรฐาน | | | | |
|----------------|-------|----------------------|-----------------|-------|--------|---------|
| | | Nucleus iterations | Cell iterations | t_nu | t_cell | t_total |
| basophil | 69 | 13.774 | 12.788 | 4.182 | 7.970 | 17.330 |
| eosinophil | 148 | 21.272 | 47.869 | 6.973 | 21.184 | 47.018 |
| lymphocyte | 141 | 12.213 | 42.023 | 4.532 | 22.585 | 48.344 |
| monocyte | 107 | 19.140 | 0.000 | 6.704 | 17.193 | 39.748 |
| neutrophil | 326 | 14.945 | 51.473 | 5.265 | 23.331 | 50.148 |
| multiple cells | 37 | 10.812 | 32.551 | 3.286 | 11.662 | 24.584 |

ตารางที่ 5. 16 แสดงจำนวนรอบและเวลาเฉลี่ยที่ใช้การแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี

ACWE2

| ชนิด | จำนวน | ค่าเฉลี่ย | | | | |
|----------------|-------|--------------------|-----------------|------------|--------------|---------------|
| | | Nucleus iterations | Cell iterations | t_nu (sec) | t_cell (sec) | t_total (sec) |
| basophil | 69 | 1.710 | 66.087 | 4.202 | 28.070 | 63.535 |
| eosinophil | 148 | 7.547 | 47.667 | 6.367 | 24.219 | 58.030 |
| lymphocyte | 141 | 2.043 | 8.078 | 4.143 | 7.264 | 21.751 |
| monocyte | 107 | 5.558 | 6.802 | 5.216 | 7.804 | 23.780 |
| neutrophil | 326 | 1.730 | 11.715 | 3.919 | 8.035 | 16.783 |
| multiple cells | 37 | 2.838 | 14.486 | 4.155 | 9.142 | 25.365 |

ตารางที่ 5. 17 แสดงส่วนเบี่ยงเบนมาตรฐานของจำนวนรอบและเวลาที่ใช้การแบ่งส่วนภาพเซลล์เม็ด

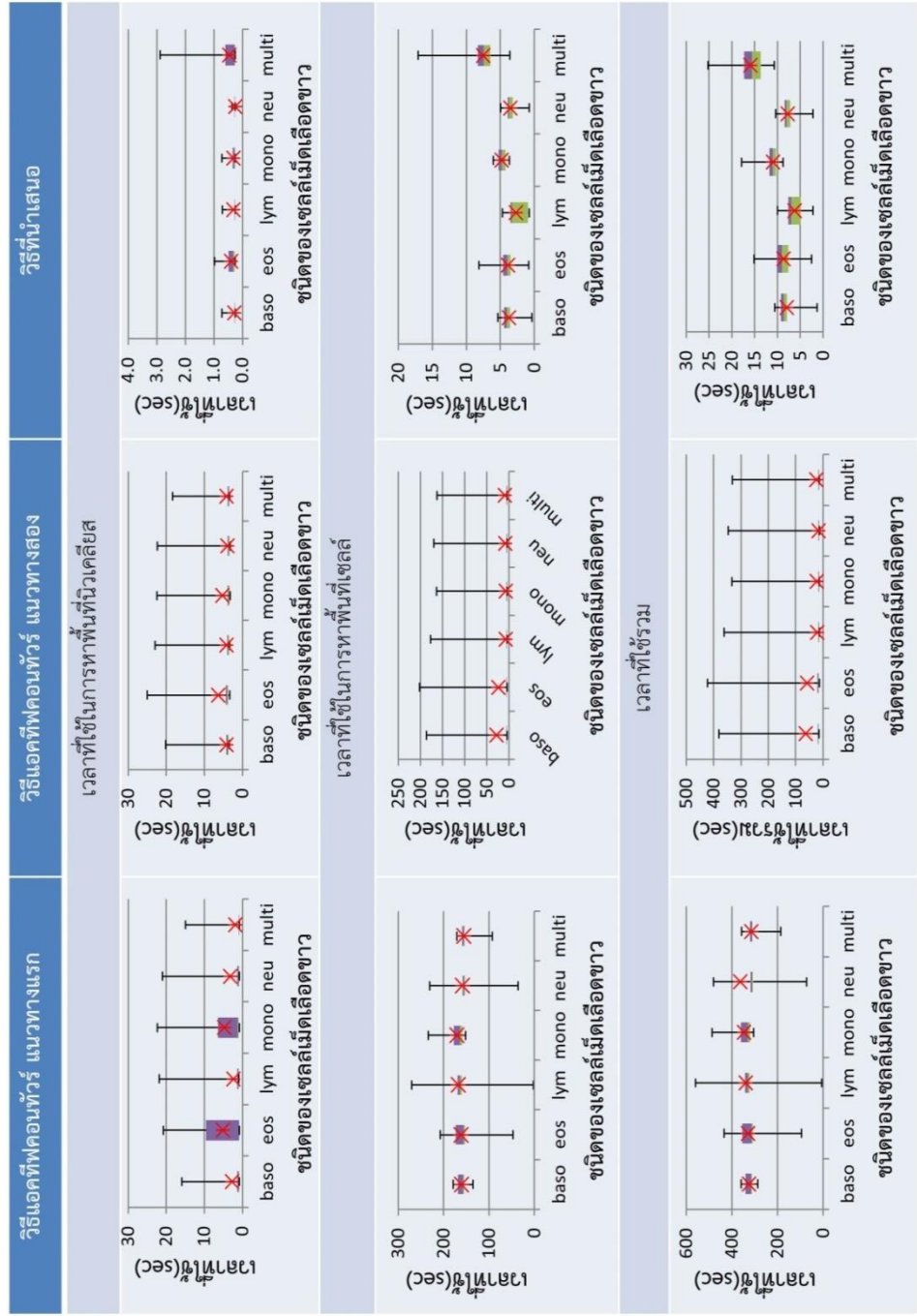
เลือดขาวด้วยวิธี ACWE2

| ชนิด | จำนวน | ส่วนเบี่ยงเบนมาตรฐาน | | | | |
|----------------|-------|----------------------|-----------------|-------|--------|---------|
| | | Nucleus iterations | Cell iterations | t_nu | t_cell | t_total |
| basophil | 69 | 5.856 | 168.054 | 1.970 | 59.308 | 118.437 |
| eosinophil | 148 | 16.441 | 145.619 | 6.074 | 53.148 | 106.707 |
| lymphocyte | 141 | 7.096 | 59.217 | 2.609 | 20.255 | 40.887 |
| monocyte | 107 | 14.316 | 53.809 | 5.037 | 17.725 | 37.021 |
| neutrophil | 326 | 5.545 | 72.443 | 1.989 | 23.084 | 46.392 |
| multiple cells | 37 | 8.559 | 82.035 | 2.576 | 26.060 | 52.267 |

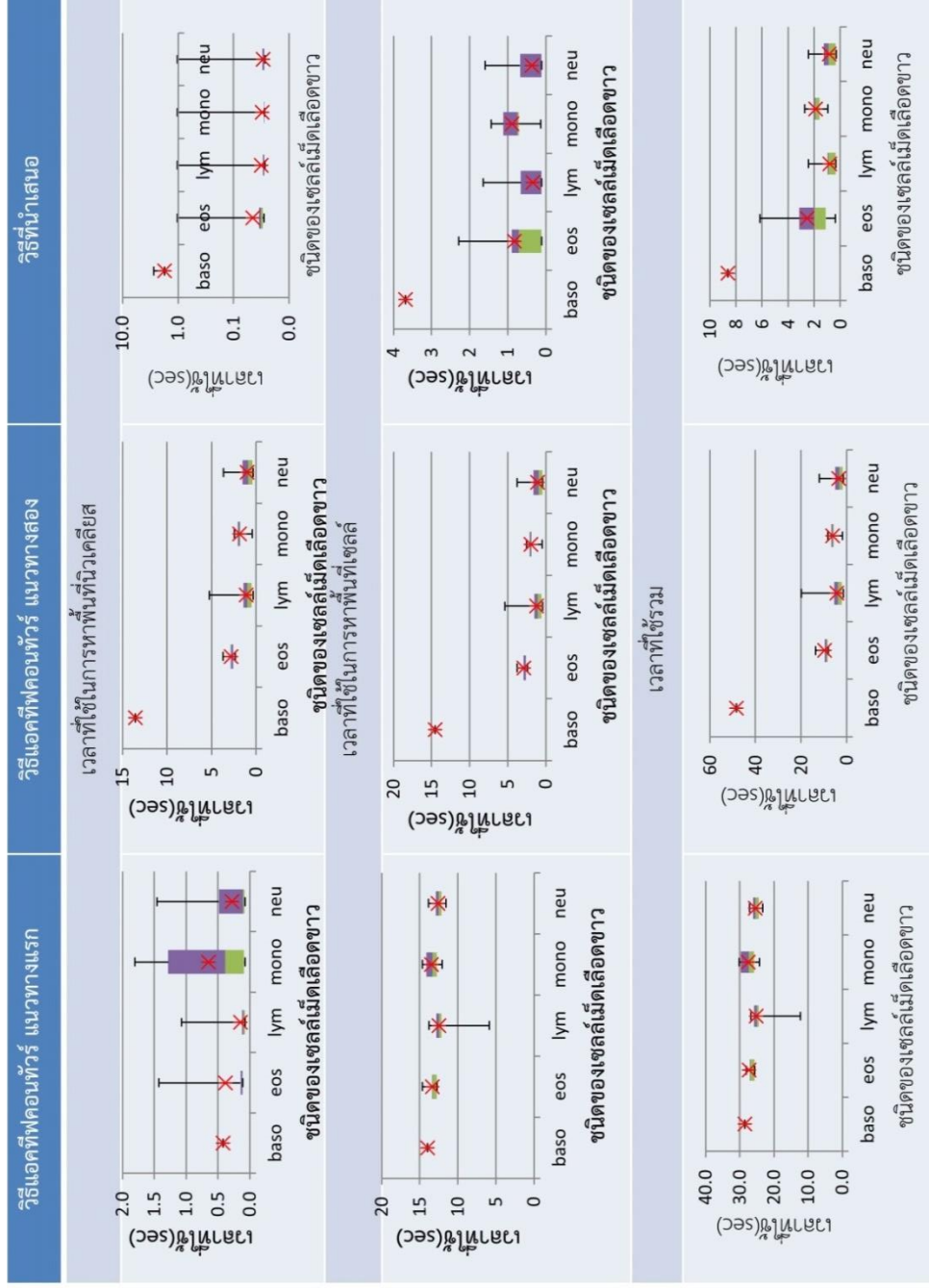
ตารางที่ 5. 18 แสดงเวลาเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่ใช้การแบ่งส่วนภาพเซลล์เม็ดเลือดขาว
ด้วยวิธีที่นำเสนอ

| ชนิด | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|----------------|-----------|--------|---------|----------------------|-----------------|------------------|
| | t_Nu | t_cell | t_total | t_Nu (sec) | t_cell (sec) | t_total (sec) |
| basophil | 0.282 | 3.735 | 7.990 | 0.064 | 1.245 | 2.179 |
| eosinophil | 0.399 | 3.891 | 8.697 | 0.162 | 1.261 | 2.060 |
| lymphocyte | 0.308 | 2.741 | 6.208 | 0.112 | 1.638 | 3.117 |
| monocyte | 0.311 | 4.792 | 11.037 | 0.090 | 0.600 | 1.325 |
| neutrophil | 0.257 | 3.567 | 7.743 | 0.013 | 0.586 | 1.088 |
| multiple cells | 0.464 | 7.509 | 15.899 | 0.423 | 2.370 | 3.052 |

ตารางที่ 5. 19 แผนภาพกล่อง หรือ box and whisker plots แสดงข้อมูลสูงสุด ข้อมูลที่มีค่าเปอร์เซ็นต์ไทล์ 75 มีพื้นฐาน ข้อมูลที่มีค่าเปอร์เซ็นต์ไทล์ 25 ข้อมูลต่ำสุด และค่าเฉลี่ย (ดอกจันที่สีแดง) ของเวลาที่ใช้ในการหาพื้นที่นิวเคลียส เซลล์ และเวลารวมของวิธีแอคทีฟคอนทัวร์เทียบกับวิธีที่นำเสนอ ชุดข้อมูลที่ 1



ตารางที่ 5. 20 แผนภาพกล่อง หรือ box and whisker plots แสดงข้อมูลสูงสุด ข้อมูลที่มีค่าเปอร์เซ็นไทล์ 75 มีอยู่ฐาน ข้อมูลที่มีค่าเปอร์เซ็นไทล์ 25 ข้อมูลต่ำสุด และค่าเฉลี่ย (ออกจันทรีสีแดง) ของเวลาที่ใช้ในการหาพื้นที่นิวเคลียส เซลล์และเวลารวมของวิธีแอคทีฟคอนทัวร์เทียบกับวิธีที่นำเสนอ ชุดข้อมูลที่ 2



5.8.การเปรียบเทียบเวลาในการประมวลผลการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธีที่

นำเสนอ วิธี ACWE1 และ ACWE2 กรณีพบเซลล์เม็ดเลือดขาวในภาพหนึ่งเซลล์ขึ้นไป

เนื่องจากในชุดข้อมูลที่ 1 เป็นภาพที่ประกอบด้วยจำนวนเซลล์ตั้งแต่ 1 ตัว ขึ้นไป จึงทำการเปรียบเทียบเวลาในการประมวลผลการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธีที่นำเสนอ วิธี ACWE1 และ ACWE2 กรณีพบเซลล์เม็ดเลือดขาวในภาพหนึ่งเซลล์ขึ้นไป

ตารางที่ 5. 21 เวลาในการประมวลผลการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี ACWE1

| ภาพเซลล์ เม็ดเลือด ขาว | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------------------------|-------|---------------|-----------------|------------------|----------------------|-----------------|------------------|
| | | t_Nu (sec) | t_cell (sec) | t_total (sec) | t_Nu (sec) | t_cell (sec) | t_total (sec) |
| 1 เซลล์ | 791 | 3.56 | 162.18 | 328.55 | 5.69 | 21.60 | 46.85 |
| 2 เซลล์ | 23 | 1.93 | 155.69 | 313.80 | 3.42 | 12.16 | 25.65 |
| 3 เซลล์ | 14 | 0.87 | 157.48 | 316.32 | 0.03 | 1.97 | 3.99 |

ตารางที่ 5. 22 เวลาในการประมวลผลการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธี ACWE2

| ภาพเซลล์ เม็ดเลือด ขาว | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------------------------|-------|---------------|-----------------|------------------|----------------------|-----------------|------------------|
| | | t_Nu (sec) | t_cell (sec) | t_total (sec) | t_Nu (sec) | t_cell (sec) | t_total (sec) |
| 1 เซลล์ | 791 | 4.22 | 7.65 | 22.53 | 3.70 | 34.99 | 70.42 |
| 2 เซลล์ | 23 | 3.98 | 21.78 | 50.42 | 2.40 | 50.50 | 100.99 |
| 3 เซลล์ | 14 | 3.67 | 42.23 | 91.11 | 0.21 | 75.60 | 150.93 |

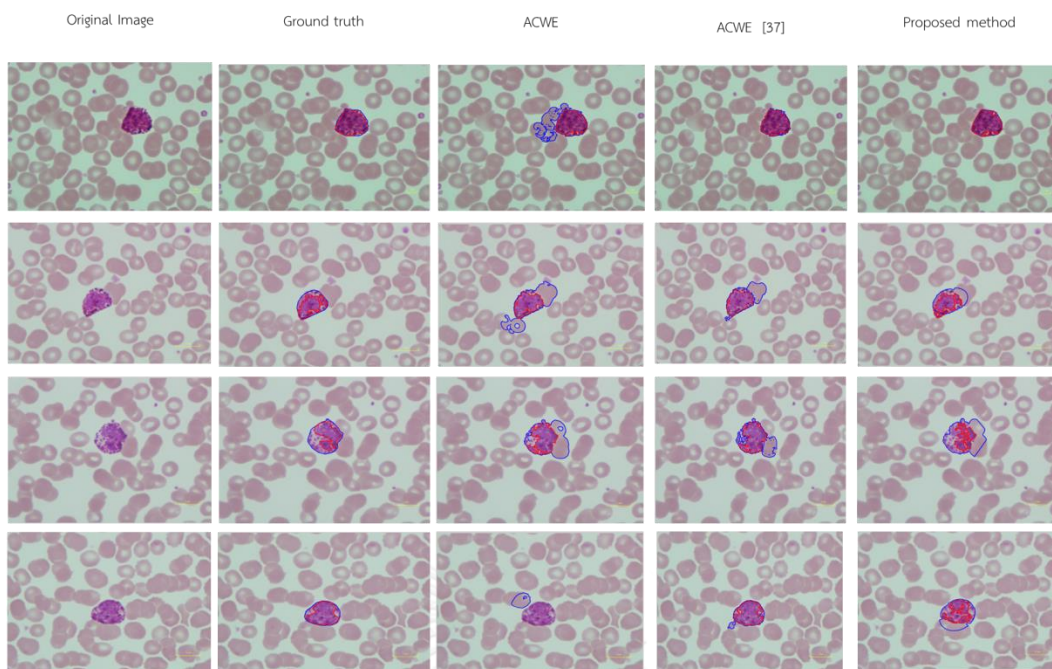
ตารางที่ 5. 23 เวลาในการประมวลผลการแบ่งส่วนภาพเซลล์เม็ดเลือดขาวด้วยวิธีที่นำเสนอ

| ภาพเซลล์ เม็ดเลือด ขาว | จำนวน | ค่าเฉลี่ย | | | ส่วนเบี่ยงเบนมาตรฐาน | | |
|------------------------------|-------|---------------|-----------------|------------------|----------------------|-----------------|------------------|
| | | t_Nu (sec) | t_cell (sec) | t_total (sec) | t_Nu (sec) | t_cell (sec) | t_total (sec) |
| 1 เซลล์ | 791 | 0.31 | 3.70 | 8.19 | 0.11 | 1.17 | 2.22 |
| 2 เซลล์ | 23 | 0.46 | 7.16 | 15.56 | 0.440 | 2.158 | 2.937 |
| 3 เซลล์ | 14 | 0.47 | 10.81 | 19.14 | 0.25 | 1.82 | 2.34 |

พบว่าวิธีที่นำเสนอใช้เวลาในการแบ่งส่วนน้อยกว่าวิธี ACWE1 และ ACWE2 ในการแบ่งส่วนนิวเคลียสแม้ว่าจำนวนเซลล์จะเพิ่มขึ้น แต่เวลาที่ใช้ไม่แตกต่างกัน แต่เวลาที่ใช้ในการแบ่งส่วนเซลล์จะเพิ่มขึ้นตามจำนวนเซลล์ที่เพิ่มขึ้น ซึ่งส่งผลให้เวลารวมของวิธีที่นำเสนอจะใช้เวลาเพิ่มขึ้นเมื่อจำนวนเซลล์เพิ่มขึ้น แต่อย่างไรก็ตามโดยภาพรวมพบว่าน้อยกว่าวิธี ACWE1 และ ACWE2 มาก

5.9.การเปรียบเทียบผลภาพการแบ่งส่วนนิวเคลียสและเซลล์ด้วยวิธี ACWE1 และ ACWE2 เทียบกับวิธีที่นำเสนอ

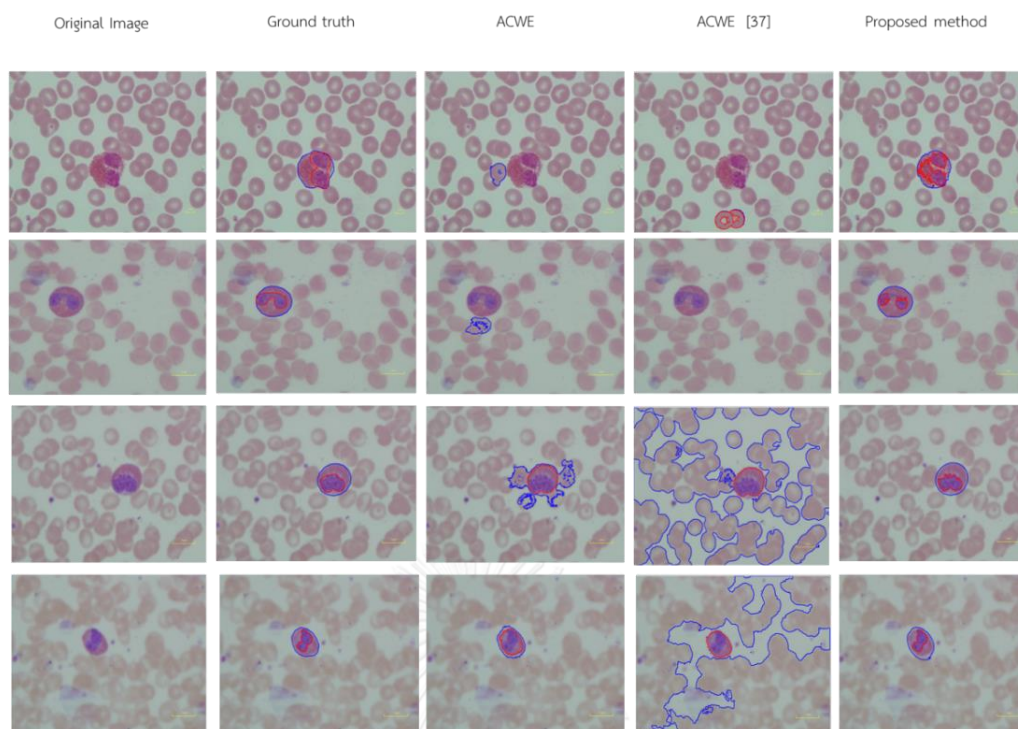
จากรูปที่ 5.11 เป็นผลการแบ่งส่วนของข้อมูลภาพชุดที่ 1 โดยแถวที่ 1 - 4 คือ basophil และหลักที่ 1 คือภาพเริ่มต้น (original image) หลักที่ 2 คือภาพที่แบ่งส่วนด้วยผู้เชี่ยวชาญ (ground truth) หลักที่ 3 ภาพที่แบ่งส่วนด้วย ACWE1 หลักที่ 4 ภาพที่แบ่งส่วนด้วยวิธี ACWE2 หลักที่ 5 ภาพที่แบ่งส่วนด้วยวิธีที่นำเสนอ (proposed method) โดยเส้นสีแดงคือเส้นขอบของนิวเคลียส และเส้นสีน้ำเงินคือเส้นขอบของเซลล์ ตามลำดับ



รูปที่ 5.11 แสดงการเปรียบเทียบผลการแบ่งส่วนนิวเคลียสและเซลล์ของ basophil ด้วยวิธีต่างๆ

การแบ่งส่วน basophil จากรูปพบว่าวิธีแอกทิฟคอนทัวร์ทั้งสองแบบมีข้อเสียคือ เมื่อมีเซลล์ข้างเคียงที่อยู่ติดกันมีค่า hue ใกล้เคียงกับเซลล์เม็ดเลือดขาวจะทำให้เกิดการแบ่งส่วนเซลล์ที่อยู่ติดกันเข้าไปด้วย ส่วนวิธีที่นำเสนอพบว่าบริเวณที่ไม่มีเซลล์ติดกันให้ผลลัพธ์ถูกต้อง แต่บริเวณที่มีเซลล์ติดกันจะให้ผลใกล้เคียงกับพื้นที่เซลล์เม็ดเลือดขาวจริง และจากภาพในแถวที่ 4 พบว่า วิธี ACWE1 หากแบ่งส่วนนิวเคลียสผิดพลาดอันเนื่องมาจากการวางคอนทัวร์เริ่มต้นผิด จะส่งผลให้แบ่งส่วนเซลล์ผิดเช่นกัน

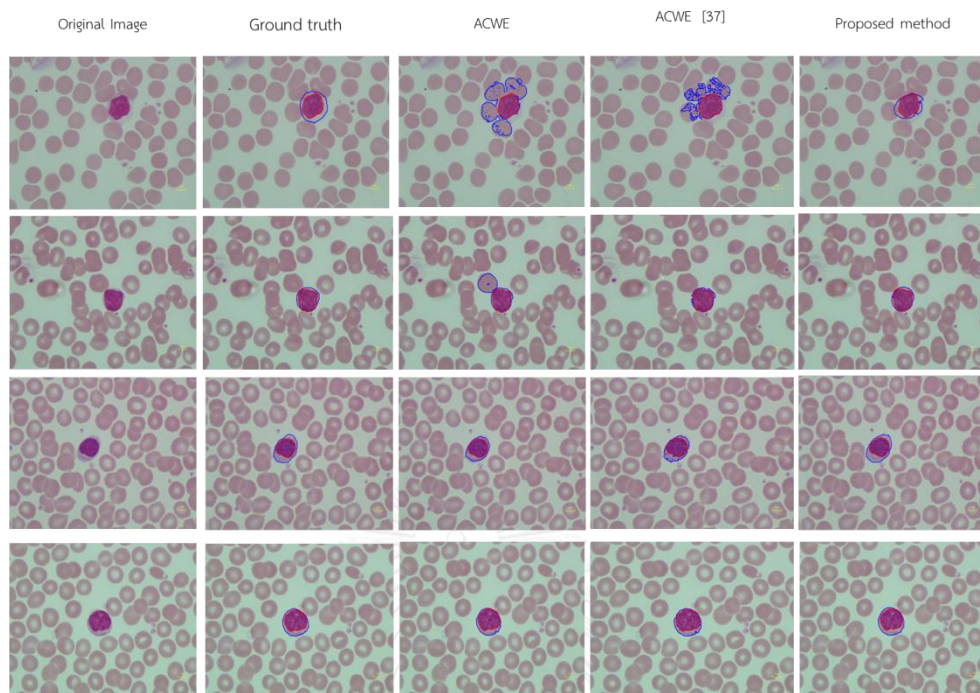
จากรูปที่ 5. 12 แถวที่ 1 - 4 คือ eosinophil และหลักที่ 1 คือภาพเริ่มต้น (original image) หลักที่ 2 คือภาพที่แบ่งส่วนด้วยผู้เชี่ยวชาญ (ground truth) หลักที่ 3 ภาพที่แบ่งส่วนด้วยวิธี ACWE1 หลักที่ 4 ภาพที่แบ่งส่วนด้วยวิธี ACWE2 หลักที่ 5 ภาพที่แบ่งส่วนด้วยวิธีที่นำเสนอ (proposed method) โดยเส้นสีแดงคือเส้นขอบของนิวเคลียส และเส้นสีน้ำเงินคือเส้นขอบของเซลล์ตามลำดับ



รูปที่ 5. 12 แสดงการเปรียบเทียบผลการแบ่งส่วนนิวเคลียสและเซลล์ของ eosinophil ด้วยวิธีต่างๆ

สำหรับการแบ่งส่วน eosinophil พบว่าวิธีแอกทีฟคอนทัวร์ทั้งสองแบบไม่สามารถหาตำแหน่งของนิวเคลียสได้ เนื่องจากการติดสีของเซลล์ชนิดนี้ค่อนข้างเหมือนสีของเซลล์เม็ดเลือดแดง การแปลงภาพไปนารีเพื่อวางคอนทัวร์เริ่มต้นในการแบ่งส่วนนิวเคลียสจึงมีปัญหาดังรูปแถวที่ 1 และ 2 ส่วนรูปในแถวที่ 3 และ 4 บริเวณของไซโทพลาซึมที่จับบดบังนิวเคลียสในภาพ saturation ทำให้แอกทีฟคอนทัวร์มองว่าเซลล์คือนิวเคลียส เมื่อมีเซลล์ชนิดอื่นมาติดกันจึงทำให้คอนทัวร์วิ่งไปทั้งภาพ ส่วนวิธีที่นำเสนอพบว่าโดยภาพรวมให้ผลลัพธ์ถูกต้อง แต่มีบางบริเวณในนิวเคลียสที่แบ่งส่วนผิดเล็กน้อยดังรูปในแถวที่ 1

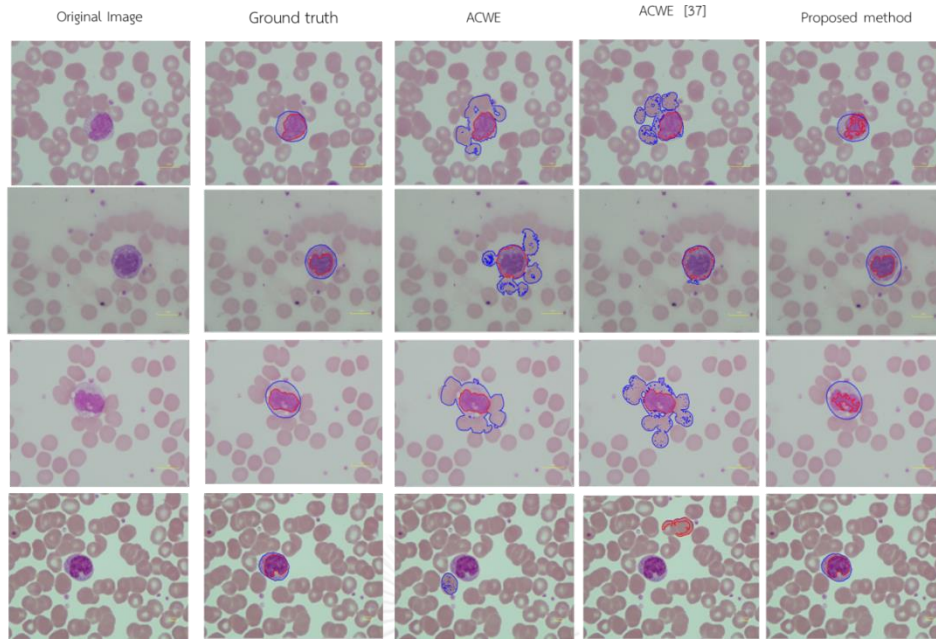
จากรูปที่ 5. 13 แถวที่ 1 - 4 คือ lymphocyte และหลักที่ 1 คือภาพเริ่มต้น (original image) หลักที่ 2 คือภาพที่แบ่งส่วนด้วยผู้เชี่ยวชาญ (ground truth) หลักที่ 3 ภาพที่แบ่งส่วนด้วยวิธี ACWE1 หลักที่ 4 ภาพที่แบ่งส่วนด้วยวิธี ACWE2 หลักที่ 5 ภาพที่แบ่งส่วนด้วยวิธีที่นำเสนอ (proposed method) โดยเส้นสีแดงคือเส้นขอบของนิวเคลียส และเส้นสีน้ำเงินคือเส้นขอบของเซลล์ตามลำดับ



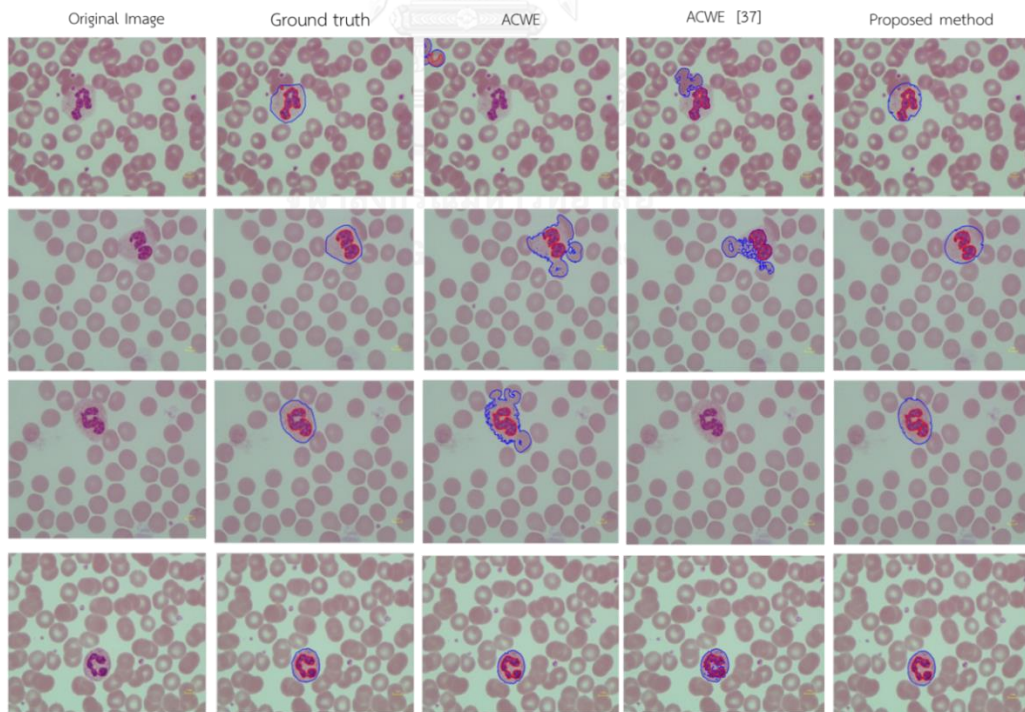
รูปที่ 5. 13 แสดงการเปรียบเทียบผลภาพการแบ่งส่วนนิวเคลียสและเซลล์ของ lymphocyte ด้วยวิธีต่างๆ

สำหรับการแบ่งส่วน lymphocyte พบว่าโดยภาพรวมทั้งสามวิธีสามารถแบ่งส่วนได้ดี แต่เมื่อมีเซลล์มาอยู่ติดกันจะทำให้คอนทัวร์เคลื่อนที่ไปครอบคลุมเซลล์ที่ติดกันด้วย

จากรูปที่ 5. 14 แถวที่ 1 - 4 คือ monocyte และหลักที่ 1 คือภาพเริ่มต้น (original image) หลักที่ 2 คือภาพที่แบ่งส่วนด้วยผู้เชี่ยวชาญ (ground truth) หลักที่ 3 ภาพที่แบ่งส่วนด้วยวิธี ACWE1 หลักที่ 4 ภาพที่แบ่งส่วนด้วยวิธี ACWE2 หลักที่ 5 ภาพที่แบ่งส่วนด้วยวิธีที่นำเสนอ (proposed method) โดยเส้นสีแดงคือเส้นขอบของนิวเคลียส และเส้นสีน้ำเงินคือเส้นขอบของเซลล์ ตามลำดับ



รูปที่ 5. 14 แสดงการเปรียบเทียบผลภาพการแบ่งส่วนนิวเคลียสและเซลล์ของ monocyte ด้วยวิธีต่างๆ



รูปที่ 5. 15 แสดงการเปรียบเทียบผลภาพการแบ่งส่วนนิวเคลียสและเซลล์ของ neutrophil ด้วยวิธีต่างๆ

สำหรับการแบ่งส่วน monocyte พบว่าเมื่อมีเซลล์มาอยู่ติดกันจะทำให้คอนทัวร์เคลื่อนที่ไปครอบคลุมเซลล์ที่ติดกันด้วย และบริเวณที่เซลล์เม็ดเลือดแดงหนาแน่นมาก การกำหนดค่าขีดแบ่งแบบอัตโนมัติเพื่อหาตำแหน่งคอนทัวร์เริ่มต้นของการแบ่งส่วนนิวเคลียสก็จะมีปัญหา ส่วนวิธีที่นำเสนอพบว่าบริเวณนิวเคลียสจะมีความผิดพลาด เนื่องจากลักษณะของนิวเคลียสของเซลล์ชนิดนี้ โครมาทินมีการเกาะกันหลวมๆ ค่าระดับความเข้มของช่องสีเขียวจะมีค่าสูงขึ้นในบางจุด

จากรูปที่ 5. 15 แถวที่ 1 - 4 คือ neutrophil และหลักที่ 1 คือภาพเริ่มต้น (original image) หลักที่ 2 คือภาพที่แบ่งส่วนด้วยผู้เชี่ยวชาญ (ground truth) หลักที่ 3 ภาพที่แบ่งส่วนด้วยวิธี ACWE1 หลักที่ 4 ภาพที่แบ่งส่วนด้วยวิธี ACWE2 หลักที่ 5 ภาพที่แบ่งส่วนด้วยวิธีที่นำเสนอ (proposed method) โดยเส้นสีแดงคือเส้นขอบของนิวเคลียส และเส้นสีน้ำเงินคือเส้นขอบของเซลล์ตามลำดับ

สำหรับการแบ่งส่วน neutrophil พบว่าเมื่อมีเซลล์มาอยู่ติดกันจะทำให้คอนทัวร์เคลื่อนที่ไปครอบคลุมเซลล์ที่ติดกันด้วย และบริเวณไซโทพลาซึมค่อนข้างใสแทบจะใกล้เคียงกับพื้นหลังทำให้คอนทัวร์ข้ามไปยังเซลล์ข้างเคียง ส่วนวิธีที่นำเสนอพบว่าให้ผลถูกต้องสูง

5.10. ผลการเลือกและการทดสอบความสามารถในการจำแนกของลักษณะบ่งต่างที่นำมาใช้

จากการตารางที่ ง. 1 จะได้ลักษณะบ่งต่างจำนวน 40 ลักษณะที่ถูกเลือกมาใช้ คือ ลักษณะบ่งต่างลำดับที่ 1-13, 15-18, 22, 24-26, 31-32, 39-44, 51-54, 56-58, 60, 62-64 จำนวน 40 ลักษณะบ่งต่าง ทำการเลือกลักษณะบ่งต่างทดสอบกับแบบจำลอง linear discrimination function (LDA) โดยใช้วิธี sequential forward selection (SFS) ในการเลือกลักษณะบ่งต่าง พบว่ามี 13 ลักษณะบ่งต่าง ดังแสดงในตารางที่ 5. 24 โดยผลการจำแนกเซลล์ basophil, eosinophil, lymphocyte, monocyte และ neutrophil ในขั้นตอนการฝึก แสดงดัง ตารางที่ 5. 25

ตารางที่ 5. 24 ลักษณะบ่งต่างที่เลือกใช้ จำนวน 13 ลักษณะบ่งต่าง มีรายละเอียดดังต่อไปนี้

| ลำดับที่ | ชื่อย่อที่ใช้ | ความหมาย |
|----------|---------------|---|
| 2 | meanGNU | ค่าเฉลี่ยความเข้มแสงของช่องสัญญาณสีเขียวภายในพื้นที่นิวเคลียส |
| 4 | VRNU | ค่าความแปรปรวนของความเข้มแสงของช่องสัญญาณสีแดงภายในพื้นที่นิวเคลียส |
| 8 | meanGCy | ค่าเฉลี่ยความเข้มแสงของช่องสัญญาณสีเขียว ภายในพื้นที่ไซโทพลาซึม |

| ลำดับที่ | ชื่อย่อที่ใช้ | ความหมาย |
|----------|----------------|--|
| 11 | VGCy | ค่าความแปรปรวนของความเข้มแสงของช่องสัญญาณสีเขียวภายในพื้นที่ไซโทพลาซึม |
| 12 | VBCy | ค่าความแปรปรวนของความเข้มแสงของช่องสัญญาณสีน้ำเงินภายในพื้นที่ไซโทพลาซึม |
| 17 | VGCell | ค่าความแปรปรวนของความเข้มแสงของช่องสัญญาณสีเขียวภายในพื้นที่เซลล์ |
| 22 | NUArea | จำนวนพิกเซลในพื้นที่นิวเคลียส |
| 24 | NUeccentricity | อัตราส่วนความยาวแกนหลักต่อแกนรองของนิวเคลียส |
| 44 | entropyCyB | เอนโทรปีของไซโทพลาซึมในช่องสีน้ำเงิน |
| 51 | energyCyR | พลังงานของไซโทพลาซึมในช่องสีแดง |
| 57 | CyR/CyB | อัตราส่วนค่าเฉลี่ยสีเทาของพื้นที่ไซโทพลาซึม ช่องสีแดงเทียบกับช่องสีน้ำเงิน |
| 62 | CyG/CyR | อัตราส่วนค่าเฉลี่ยสีเทาของพื้นที่ไซโทพลาซึม ช่องสีเขียวเทียบกับช่องสีแดง |
| 63 | CyB/CyR | อัตราส่วนค่าเฉลี่ยสีเทาของพื้นที่ไซโทพลาซึม ช่องสีน้ำเงินเทียบกับช่องสีแดง |

ตารางที่ 5. 25 ผลการจำแนกเซลล์ basophil, eosinophil, lymphocyte, monocyte และ neutrophil ในขั้นตอนการฝึก

| | | Predict | | | | | % correction |
|--------------|------------|----------|------------|------------|----------|------------|--------------|
| | | basophil | eosinophil | lymphocyte | monocyte | neutrophil | |
| Actual | basophil | 6 | 0 | 0 | 0 | 0 | 100% |
| | eosinophil | 1 | 11 | 0 | 0 | 0 | 92% |
| | lymphocyte | 0 | 0 | 26 | 1 | 0 | 96% |
| | monocyte | 1 | 1 | 1 | 10 | 0 | 77% |
| | neutrophil | 0 | 0 | 0 | 1 | 51 | 98% |
| % correction | | 75% | 92% | 96% | 83% | 100% | |

5.11. ผลการจำแนกเซลล์

ผลการจำแนกเซลล์นำเสนอด้วยรูปแบบตารางเมทริกซ์ความสับสน (confusion matrix) ผลการจำแนกเซลล์ที่เป็น basophil, eosinophil, lymphocyte, monocyte และ neutrophil ด้วย linear discriminant function และ quadratic discriminant function ตารางที่ 5. 26 – 5. 27 ตารางที่ 5. 26 ผลการจำแนกเซลล์ basophil, eosinophil, lymphocyte, monocyte และ neutrophil ด้วย linear discriminant function

| | | Predict | | | | | % correction |
|--------------|------------|----------|------------|------------|----------|------------|--------------|
| | | basophil | eosinophil | lymphocyte | monocyte | neutrophil | |
| Actual | basophil | 31 | 0 | 1 | 1 | 1 | 91% |
| | eosinophil | 5 | 72 | 1 | 1 | 0 | 53% |
| | lymphocyte | 2 | 4 | 152 | 2 | 6 | 92% |
| | monocyte | 6 | 5 | 3 | 61 | 2 | 79% |
| | neutrophil | 1 | 1 | 3 | 7 | 309 | 96% |
| % correction | | 69% | 88% | 95% | 85% | 97% | |

ตารางที่ 5. 27 ผลการจำแนกเซลล์ basophil, eosinophil, lymphocyte, monocyte และ neutrophil ด้วย quadratic discriminant function

| | | Predict | | | | | % correction |
|--------------|------------|----------|------------|------------|----------|------------|--------------|
| | | basophil | eosinophil | lymphocyte | monocyte | neutrophil | |
| Actual | basophil | 28 | 1 | 2 | 3 | 0 | 82% |
| | eosinophil | 3 | 67 | 2 | 6 | 1 | 53% |
| | lymphocyte | 0 | 0 | 157 | 5 | 4 | 95% |
| | monocyte | 3 | 1 | 3 | 69 | 1 | 90% |
| | neutrophil | 1 | 2 | 5 | 2 | 311 | 97% |
| % correction | | 80% | 94% | 93% | 81% | 98% | |

5.12. ผลการให้คะแนนตัวจำแนก

การประเมินสมรรถนะของตัวจำแนก โดยการคำนวณค่า accuracy, sensitivity, specification และ precision ดังสมการที่ (51) - (54) สำหรับตัวจำแนกเซลล์ชนิดอื่นๆ ด้วยวิธี linear discriminant function และ quadratic discriminant function มีผลการประเมินสมรรถนะดังตารางที่ 5. 28 และ 5.29 ตามลำดับ

ตารางที่ 5. 28 ค่า accuracy, sensitivity, specificity และ precision ในการจำแนก basophil, eosinophil, lymphocyte, monocyte และ neutrophil ด้วย linear discriminant function

| linear discriminant function | basophil | eosinophil | lymphocyte | monocyte | neutrophil |
|------------------------------|----------|------------|------------|----------|------------|
| accuracy | 0.97 | 0.97 | 0.97 | 0.96 | 0.97 |
| sensitivity | 0.91 | 0.91 | 0.92 | 0.79 | 0.96 |
| specificity | 0.98 | 0.98 | 0.98 | 0.98 | 0.97 |
| precision | 0.69 | 0.88 | 0.95 | 0.85 | 0.97 |

ตารางที่ 5. 29 ค่า accuracy, sensitivity, specificity และ precision ในการจำแนก basophil, eosinophil, lymphocyte, monocyte และ neutrophil ด้วย quadratic discriminant function

| quadratic discriminant function | basophil | eosinophil | lymphocyte | monocyte | neutrophil |
|---------------------------------|----------|------------|------------|----------|------------|
| accuracy | 0.98 | 0.98 | 0.97 | 0.96 | 0.98 |
| sensitivity | 0.82 | 0.85 | 0.95 | 0.90 | 0.97 |
| specificity | 0.99 | 0.99 | 0.98 | 0.97 | 0.98 |
| precision | 0.80 | 0.94 | 0.93 | 0.81 | 0.98 |

บทที่ 6

ข้อสรุปและ/หรือข้อเสนอแนะ

วิธีที่นำเสนอสามารถระบุตำแหน่งเซลล์เม็ดเลือดขาวได้ถูกต้องทุกกรณี ซึ่งวิทยานิพนธ์นี้ได้ทำการทดสอบกรณีไม่มีเซลล์เม็ดเลือดขาว กรณีพบ 1 -3 เซลล์ ตามลำดับ ส่วนการแบ่งส่วนนิวเคลียสและเซลล์ของข้อมูลทั้งสองชุด ความคล้ายคลึงของการแบ่งส่วนนิวเคลียสและเซลล์ของวิทยานิพนธ์เทียบกับการแบ่งส่วนด้วยมือของนักโลหิตวิทยาในเซลล์ทั้งห้าชนิด วิธีที่นำเสนอให้ผลความคล้ายคลึงเฉลี่ย 0.90 เมื่อเปรียบเทียบค่าความถูกต้องของการแบ่งส่วนของวิธีที่นำเสนอพบว่าได้ดีกว่าวิธีแอกทีฟคอนทัวร์ทั้งสองแบบ โดยในข้อมูลชุดที่หนึ่งเซลล์ที่มีค่าต่ำสุดจะเป็น eosinophil หรือ basophil มีสาเหตุมาจากการมีเม็ดเล็ก ๆ จำนวนมากที่บริเวณไซโทพลาซึมซึ่งสีใกล้เคียงกับเซลล์อื่นๆที่อยู่ติดกัน ในข้อมูลภาพชุดที่สอง เซลล์ที่มีค่าต่ำสุดคือ monocyte เนื่องจากนิวเคลียสของเซลล์มีโครมาทินที่จับกันหลวมๆ ทำให้สีของนิวเคลียสในเซลล์ชนิดนี้จางกว่านิวเคลียสของเซลล์ชนิดอื่นๆ ค่า false positive ratio และ false negative ratio ซึ่งเป็นตัวบ่งชี้ถึงการแบ่งส่วนที่ over segmentation หรือ under segmentation ของวิธีที่นำเสนอมีค่าต่ำกว่าอีก 2 วิธี

เมื่อเปรียบเทียบเวลาที่ใช้ในการประมวลผลทั้งสามวิธีพบว่า วิธีที่นำเสนอใช้เวลาในการแบ่งส่วนโดยภาพรวมน้อยกว่า 15.89 วินาที ซึ่งเป็นกรณีที่มีเซลล์หลายเซลล์ในภาพเดียว แต่อย่างไรก็ตามเวลาที่ใช้ก็ยังคงน้อยกว่าวิธีแอกทีฟคอนทัวร์ทั้งสองแบบมาก ในวิธีแอกทีฟคอนทัวร์ทั้งสองแบบมีจำนวนรอบในการแบ่งส่วนนิวเคลียสไม่มากนัก แต่ใช้จำนวนรอบในการหาขอบของเซลล์สูงอันเนื่องมาจากขอบของเซลล์ไม่ชัดเจน ทำให้เวลาที่ใช้ในการคำนวณสูงตามไปด้วย

การจำแนกเซลล์เม็ดเลือดขาวได้นำผลการแบ่งส่วนพื้นที่นิวเคลียส ไซโทพลาซึม และเซลล์มาทำการหาลักษณะบ่งต่าง ทำการเลือกลักษณะบ่งต่าง และการจำแนกชนิดเซลล์เม็ดเลือดขาวแบบปกติ เป็น 5 ชนิด คือ basophil, eosinophil, lymphocyte, monocyte และ neutrophil ลักษณะบ่งต่างที่เลือกใช้ในการจำแนกเซลล์ 13 ลักษณะบ่งต่าง โดยใช้ทั้งวิธี filter และ wrapper ใช้เทคนิค sequential forward selection และทำการทดสอบลักษณะบ่งต่างที่เลือกด้วยเทคนิค LOOCV หลังจากนั้นทำการจำแนกชนิดเซลล์เม็ดเลือดขาวด้วย linear discrimination function และ quadratic discrimination function อัตราความถูกต้องของการจำแนกเซลล์ด้วย linear discrimination function และ quadratic discrimination function ให้ผลที่ดี โดย quadratic discrimination function ให้ผลดีกว่า

ข้อดีของวิธีที่นำเสนอนี้คือ สามารถนำวิธีที่นำเสนอไปใช้กับภาพที่ขนาด resolution แตกต่างกันได้ โดยใช้ค่าถ่วงน้ำหนักที่เหมาะสมกับความละเอียดของชุดข้อมูลภาพนั้น ข้อจำกัดของวิธีที่นำเสนอ คือ เวลาที่ใช้ในการแบ่งส่วนเซลล์พบว่าหากจำนวนเซลล์ในภาพมากกว่า 1 เซลล์ เวลาที่ใช้ในการแบ่งส่วนก็จะเพิ่มมากขึ้นเป็นจำนวนเท่าของเวลาที่ใช้ในการแบ่งส่วนเซลล์เดียว แต่อย่างไรก็ตามโดยภาพรวม เวลาที่ใช้ในการแบ่งส่วนเซลล์ก็ยังคงน้อยกว่าเวลาที่ใช้ในการแบ่งส่วนด้วยแอคทีฟคอนทัวร์ทั้งสองแบบมาก



รายการอ้างอิง

1. Ceelie, H., R.B. Dinkelaar, and W. van Gelder, *Examination of peripheral blood films using automated microscopy; evaluation of Diffmaster Octavia and Cellavision DM96*. Journal of Clinical Pathology, 2007. 60(1): p. 72-79.
2. Lin, Y., P. Meer, and D.J. Foran, *Unsupervised segmentation based on robust estimation and color active contour models*. Information Technology in Biomedicine, IEEE Transactions on, 2005. 9(3): p. 475-486.
3. Sadeghian, F., et al., *A Framework for White Blood Cell Segmentation in Microscopic Blood Images Using Digital Image Processing*. 2009. 11: p. 196-206.
4. Qingmin, L. and D. Yingying. *An accurate segmentation method for white blood cell images*. in *Biomedical Imaging, 2002. Proceedings. 2002 IEEE International Symposium on*. 2002.
5. Kan, J., L. Qing-Min, and D. Sheng-Yang. *A novel white blood cell segmentation scheme using scale-space filtering and watershed clustering*. in *Machine Learning and Cybernetics, 2003 International Conference on*. 2003.
6. Fang, Y., et al. *White Blood Cell Image Segmentation Using On-line Trained Neural Network*. in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. 2005.
7. Jianhua, W., et al. *A novel color image segmentation method and its application to white blood cell image analysis*. in *Signal Processing, 2006 8th International Conference on*. 2006.
8. Wermser, D., G. Hausmann, and C.E. Liedtke, *Segmentation of blood smears by hierarchical thresholding*. Computer Vision, Graphics, and Image Processing, 1984. 25(2): p. 151-168.
9. Cseke, I. *A fast segmentation scheme for white blood cell images*. in *Pattern Recognition, 1992. Vol.III. Conference C: Image, Speech and Signal Analysis, Proceedings., 11th IAPR International Conference on*. 1992.

10. Wang, S. and M. Wang, *A new detection algorithm (NDA) based on fuzzy cellular neural networks for white blood cell detection*. Information Technology in Biomedicine, IEEE Transactions on, 2006. 10(1): p. 5-10.
11. Yampri, P., et al. *White Blood Cell Classification based on the Combination of Eigen Cell and Parametric Feature Detection*. in *Industrial Electronics and Applications, 2006 1ST IEEE Conference on*. 2006.
12. Dorini, L.B., R. Minetto, and N.J. Leite. *White blood cell segmentation using morphological operators and scale-space analysis*. in *Computer Graphics and Image Processing, 2007. SIBGRAPI 2007. XX Brazilian Symposium on*. 2007.
13. Ghosh, M., et al., *Automated leukocyte recognition using fuzzy divergence*. Micron, 2010. 41(7): p. 840-846.
14. Sadr, A., et al. *Leukocyte's nucleus segmentation using active contour in YCbCr colour space*. in *Biomedical Engineering and Sciences (IECBES), 2010 IEEE EMBS Conference on*. 2010.
15. Bergen, T., et al. *Segmentation of leukocytes and erythrocytes in blood smear images*. in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. 2008.
16. Shi, D., G. Kan, and C. Liang. *An Improved Segmentation Method for Color Image*. in *Intelligent Information Technology Application, 2008. IITA '08. Second International Symposium on*. 2008.
17. Wei, G., T. Yinggan, and L. Xiaoli. *Segmentation of Microscopic Images for Counting Leukocytes*. in *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on*. 2008.
18. Mathur, A., A.S. Tripathi, and M. Kuse, *Scalable system for classification of white blood cells from Leishman stained blood stain images*. Journal of Pathology Informatics, 2013. 4(Suppl): p. S15.
19. Reinhard, E., et al., *Color transfer between images*. Computer Graphics and Applications, IEEE, 2001. 21(5): p. 34-41.

20. Ramoser, H., et al. *Leukocyte segmentation and classification in blood-smear images*. in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. 2005.
21. Rezatofghi, S.H. and H. Soltanian-Zadeh, *Automatic recognition of five types of white blood cells in peripheral blood*. *Computerized Medical Imaging and Graphics*, 2011. 35(4): p. 333-343.
22. Theera-Umpon, N. and S. Dhompongsa, *Morphological Granulometric Features of Nucleus in Automatic Bone Marrow White Blood Cell Classification*. *Information Technology in Biomedicine, IEEE Transactions on*, 2007. 11(3): p. 353-359.
23. Piuri, V. and F. Scotti. *Morphological classification of blood leucocytes by microscope images*. in *Computational Intelligence for Measurement Systems and Applications, 2004. CIMSA. 2004 IEEE International Conference on*. 2004. IEEE.
24. สเป็ค-สายเชื้อ, ส., ภาพสีประกอบโลหิตวิทยา = *Color aids hematology* 4ed. 2544, กรุงเทพฯ: เอช ที พี เพรส. 199.
25. Heiserman, D.L., *Hematology: Principles and practice*. 2013.
26. Gonzalez, R.C. and R.E. Woods, *Digital Image Processing (3rd Edition)*. 2006: Prentice-Hall, Inc.
27. Canny, J., *A computational approach to edge detection*. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 1986(6): p. 679-698.
28. Canny, J., *A Computational Approach to Edge Detection*. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 1986. **PAMI-8**(6): p. 679-698.
29. Graham, R.L., *An Efficient Algorithm for Determining the Convex Hull of a Finite Planar Set*. *Information Processing Letters*, 1972. 1: p. 132-133.
30. Cormen, T.H., et al., *Introduction to Algorithms*. 2001: McGraw-Hill Higher Education.
31. Barber, C.B., D.P. Dobkin, and H. Huhdanpaa, *The quickhull algorithm for convex hulls*. *ACM Trans. Math. Softw.*, 1996. 22(4): p. 469-483.
32. Kirkpatrick, D.G. and R. Seidel, *The ultimate planar convex hull algorithm*. *SIAM J. Comput.*, 1986. 15(1): p. 287-299.

33. Michael Kass, A.W., Demetri Terzopoulos, *Snakes: Active contour models* International Journal of Computer Vision In International Journal of Computer Vision, 1988. 1(4): p. 10.
34. Osher, S. and J.A. Sethian, *Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations*. J. Comput. Phys., 1988. 79(1): p. 12-49.
35. Caselles, V., et al., *A geometric model for active contours in image processing*. Numerische Mathematik, 1993. 66(1): p. 1-31.
36. Malladi, R., J.A. Sethian, and B.C. Vemuri, *Shape modeling with front propagation: a level set approach*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1995. 17(2): p. 158-175.
37. Caselles, V., R. Kimmel, and G. Sapiro, *Geodesic Active Contours*. Int. J. Comput. Vision, 1997. 22(1): p. 61-79.
38. Kichenassamy, S., et al., *Conformal curvature flows: From phase transitions to active vision*. Archive for Rational Mechanics and Analysis, 1996. 134(3): p. 275-301.
39. Whitaker, R., *A Level-Set Approach to 3D Reconstruction from Range Data*. International Journal of Computer Vision, 1998. 29(3): p. 203-231.
40. Malcolm, J., et al., *Fast approximate surface evolution in arbitrary dimension*. Proc SPIE Int Soc Opt Eng, 2008. 6914.
41. Shi, Y. and W.C. Karl, *A real-time algorithm for the approximation of level-set-based curve evolution*. IEEE Trans Image Process, 2008. 17(5): p. 645-56.
42. Sussman, M., P. Smereka, and S. Osher, *A Level Set Approach for Computing Solutions to Incompressible Two-Phase Flow*. Journal of Computational Physics, 1994. 114(1): p. 146-159.
43. Bookstein, F.L., *Fitting conic sections to scattered data*. Computer Graphics and Image Processing, 1979. 9(1): p. 56-71.
44. Fitzgibbon, A., M. Pilu, and R.B. Fisher, *Direct least square fitting of ellipses*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1999. 21(5): p. 476-480.

45. Popovic, A., et al., *Statistical validation metric for accuracy assessment in medical image segmentation*. International Journal of Computer Assisted Radiology and Surgery, 2007. 2(3-4): p. 169-181.
46. Marques, O., *Practical Image and Video Processing Using MATLAB*. 2011: Wiley.
47. Stehman, S.V., *Selecting and interpreting measures of thematic classification accuracy*. Remote Sensing of Environment, 1997. 62(1): p. 77-89.
48. Nixon, M. and A.S. Aguado, *Feature Extraction & Image Processing, Second Edition*. 2008: Academic Press. 424.
49. Merchant, F.A., S.K. Shah, and K.R. Castleman, *Chapter 10 - Object Measurement*, in *Microscope Image Processing*, Q. Wu, F.A. Merchant, and K.R. Castleman, Editors. 2008, Academic Press: Burlington. p. 195-219.
50. Umbaugh, S.E., *Computer Imaging: Digital Image Analysis and Processing*. 2005, Boca Raton, FL: CRC Press.
51. Duda, R.O., P.E. Hart, and D.G. Stork, *Pattern Classification (2nd Edition)*. 2000: Wiley-Interscience.
52. Sun, T., *Atlas of Hematologic Neoplasms*. 1 ed. Introduction, ed. T. Sun. 2009: Springer US.



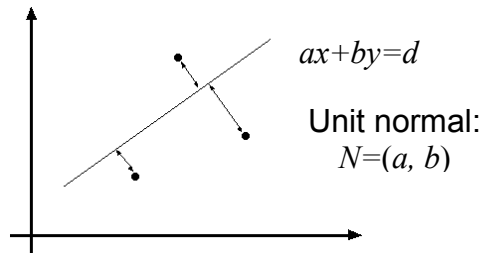
ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก ก

การพิสูจน์วิธี Least Square

ระยะห่างระหว่างจุด (x_n, y_n) และเส้นตรง $ax+by=d$, $(a^2+b^2=1)$: $|ax + by - d|$



รูปที่ ก. 1 หา (a, b, d) เพื่อลดผลรวมของระยะทางตั้งฉากยกกำลัง

$$E = \sum_{i=1}^n (ax_i + by_i - d)^2 \quad (1)$$

$$\frac{\partial E}{\partial d} = \sum_{i=1}^n -2(ax_i + by_i - d) = 0 \quad (2)$$

$$d = \frac{a}{n} \sum_{i=1}^n x_i + \frac{b}{n} \sum_{i=1}^n y_i = a\bar{x} + b\bar{y} \quad (3)$$

$$E = \sum_{i=1}^n (a(x_i - \bar{x}) + b(y_i - \bar{y}))^2 = \left\| \begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} \\ \vdots & \vdots \\ x_n - \bar{x} & y_n - \bar{y} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \right\|^2 = (UN)^T (UN) \quad (4)$$

$$\frac{dE}{dN} = 2(U^T U)N = 0 \quad (5)$$

วิธีการแก้ $(U^T U)N = 0$, อยู่ภายใต้การ $\|M\|^2 = 1$: eigenvector ของ $U^T U$ เกี่ยวข้องกับ eigenvalue ที่เล็กที่สุด (การแก้ปัญหา ค่ากำลังสองที่น้อยที่สุดกับระบบเชิงเส้น $UN = 0$)

$$U = \begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} \\ \vdots & \vdots \\ x_n - \bar{x} & y_n - \bar{y} \end{bmatrix} \quad (6)$$

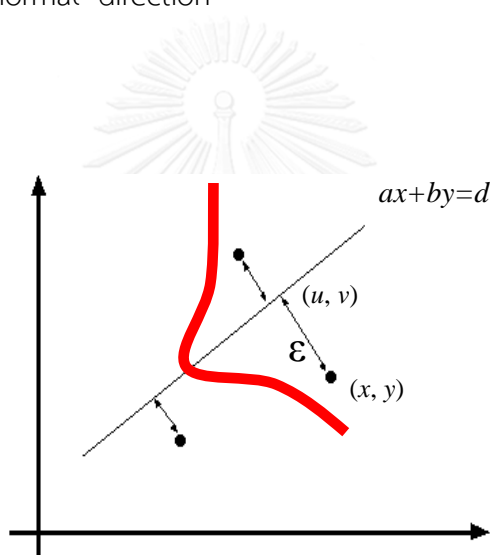
$$U^T U = \begin{bmatrix} \sum_{i=1}^n (x_i - \bar{x})^2 & \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) & \sum_{i=1}^n (y_i - \bar{y})^2 \end{bmatrix} \quad (7)$$

second moment matrix

รูปแบบการสร้าง: จุดที่เส้นที่ถูกรบกวนจากสัญญาณ Gaussian noise ในทิศทางที่ตั้งฉากกับเส้น

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix} + \varepsilon \begin{pmatrix} a \\ b \end{pmatrix} \quad (8)$$

เมื่อ $\begin{pmatrix} u \\ v \end{pmatrix}$ คือจุดบนเส้น, ε คือ Gaussian noise ที่มีค่าเฉลี่ยเป็นศูนย์ และมีส่วนเบี่ยงเบนมาตรฐาน σ , $\begin{pmatrix} a \\ b \end{pmatrix}$ คือ normal direction



รูปที่ ก. 2 จุดที่เส้นที่ถูกรบกวนจากสัญญาณ Gaussian noise

ความน่าจะเป็นของจุดเมื่อให้พารามิเตอร์ของเส้นตรง (a, b, d) :

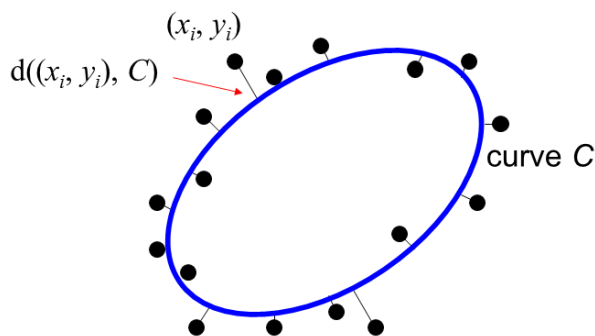
$$P(x_1, \dots, x_n | a, b, d) = \prod_{i=1}^n P(x_i | a, b, d) \propto \prod_{i=1}^n \exp\left(-\frac{(ax_i + by_i - d)^2}{2\sigma^2}\right) \quad (9)$$

Log - likelihood:

$$L(x_1, \dots, x_n | a, b, d) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (ax_i + by_i - d)^2 \quad (10)$$

สำหรับกรณีเส้นโค้งทั่วไป

เมื่อต้องการที่จะลดผลรวมของระยะทางยกกำลังของระยะห่างระหว่างจุดข้อมูลและเส้นโค้งในแบบเรขาคณิต



รูปที่ ก. 3 กรณีภาคตัดกรวยทั่วไป

ปัญหาของภาคตัดกรวยทั่วไป

ให้ p_1, p_2, \dots, p_n เป็นเซตของจุด n จุดบนภาพ โดย $p_i = [x_i \ y_i]$

$$F(p, a) = Xa = ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (11)$$

$$\text{เมื่อ } a = [a \ b \ c \ d \ e \ f]^T$$

$$X = [x^2 \ xy \ y^2 \ x \ y \ 1]$$

$F(p, a)$ เรียกว่า “algebraic distance” ของจุด $p_i = [x_i \ y_i]$ ไปยังสมการภาคตัดกรวยที่

$F(p, a) = 0$ ถ้ามีจำนวนข้อมูล n จุด

$$\begin{bmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ x_2^2 & x_2 y_2 & y_2^2 & x_2 & y_2 & 1 \\ \vdots & & & & \ddots & \vdots \\ x_n^2 & x_n y_n & y_n^2 & x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix} = 0 \quad (12)$$

$$Da = 0$$

$$\text{เมื่อ } D = [X_1 \ X_2 \ \dots \ X_n]^T$$

เราสามารถลดระยะทางเรขาคณิตที่ไม่เป็นเชิงเส้นแม้จะเป็นรูปภาคตัดกรวยเช่นกัน โดยใช้ Algebraic distance: $F(p, a)$ สามารถลดระยะทาง โดยการทำให้ linear least squares

$$E(a^*) = \min_a \arg \sum_{i=1}^n F(p_i, a)^2 \quad (13)$$

$$E = \sum_{i=1}^n (ax_i^2 + bx_iy_i + cy_i^2 + dx_i + ey_i + f)^2 \quad (14)$$

เพื่อตัดปัญหาเรื่อง $a = 0$ งานวิจัยต่างๆ ใช้การกำหนดเงื่อนไขเวกเตอร์ a เช่น กำหนด $\|a\|^2 = 1$, $a + c = 1$, $f = 1$ ซึ่งเงื่อนไขเหล่านี้เป็นเชิงเส้นทุกตัว เขียนในรูปแบบ $Ca = 1$ หรืออยู่ในรูป quadratic $a^T C a = 1$ เมื่อ C เป็น constraint matrix ขนาด 6×6 bookstein [43] แสดงให้เห็นว่า ถ้ากำหนดตัวแปร constraint เป็น quadratic สมการที่ 13 สามารถหาได้โดยการพิจารณา ระบบ eigenvalue

$$D^T D a = \lambda C a \quad (15)$$

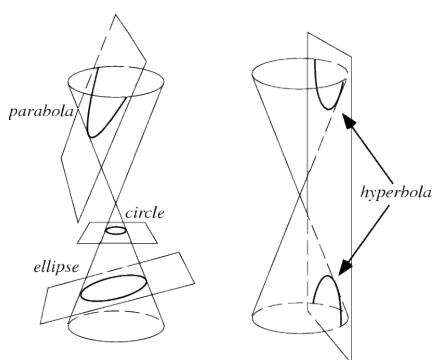
เมื่อ $D = [x_1 \ x_2 \ \dots \ x_n]^T$ คือเมทริกซ์ที่ออกแบบ และ C เป็นเมทริกซ์เงื่อนไข

เครื่องหมายของผลลัพท์จาก $4ac - b^2$ ใช้ในการจำแนกชนิดของภาคตัดกรวยโดย

> 0 Ellipse or Circle

$= 0$ Parabola

< 0 Hyperbola



รูปที่ ก. 4 การจำแนกชนิดของภาคตัดกรวย

Ref: http://mathworld.wolfram.com/images/eps-gif/ConicSection_1000.gif

Fitzgibbon, A. และคณะ [44] ได้นำเสนอวิธีใช้เทคนิค least square กับรูปวงรีโดยตรง ซึ่งใช้ข้อมูลจุดเพื่อสร้างวงรีที่ปรับให้เข้ากับชุดข้อมูล เพื่อลดระยะทางเชิงพีชคณิตให้มากที่สุด โดยกำหนด constraint $4ac - b^2 = 1$ ดังนั้น quadratic constraint สามารถเขียนในรูปเมทริกซ์ $a^T C a = 1$ เป็นดังสมการที่ 15

$$a^T \begin{bmatrix} 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} a = 1 = 4ac - b^2 \quad (16)$$

เมื่อทำตาม bookstein ปัญหาการปรับเงื่อนไขของวงรีก็ถูกลดลงเป็น . ทำให้สมการ(13) $E = \|Da\|^2$ ตามข้อจำกัด $a^T C a = 1$ เมื่อเมทริกซ์ที่ออกแบบ D ถูกกำหนดตามเนื้อหาก่อนหน้านี้ นำ Lagrange multiplier λ และ differentiateing ก็จะได้สมการ

$$E = \sum_{i=1}^n \|X_i^T a\|^2 = \|Da\|^2 = a^T D^T D a = a^T S a \quad (17)$$

เมื่อ S เป็น scatter matrix $D^T D$

ในการแก้ปัญหา ellipse fitting สามารถเขียนสมการ $E(a^*) = \min_a \arg a^T S a$ ด้วยข้อกำหนด $a^T C a = 1$

Lagrange multipliers

$$\min_{a, \lambda} L(a, \lambda) = a^T S a - \lambda(a^T C a - 1) \quad (18)$$

$$\frac{\partial L}{\partial a} = 2Sa - 2\lambda Ca = 0 \quad (19)$$

$$\frac{\partial L}{\partial \lambda} = a^T C a - 1 = 0 \quad (20)$$

จาก (19),(20) เขียนเป็นระบบใหม่

$$S a = \lambda C a \quad (21)$$

$$a^T C a = 1 \quad (22)$$

จากสมการที่ (20) และ (16)

$$E(a^*) = \min_a \arg a^T S a = a^T \lambda C a = \lambda$$

ซึ่งสามารถพิสูจน์ได้ว่าสามารถแก้ปัญหาได้ด้วย eigen vector ของสมการ 21 ถ้า $(\lambda_i, \mathbf{u}_i)$ แก่สมการ 21 แล้วทำ $(\lambda_i, \mu \mathbf{u}_i)$ สำหรับ μ ใดๆ และจากสมการ 22 เราสามารถหาค่า μ_i ที่ทำให้ $\mu_i^2 \mathbf{u}_i^T \mathbf{C} \mathbf{u}_i = 1$ เมื่อให้

$$\mu_i = \sqrt{\frac{1}{\mathbf{u}_i^T \mathbf{C} \mathbf{u}_i}} = \sqrt{\frac{1}{\mathbf{u}_i^T \mathbf{S} \mathbf{u}_i}} \quad (23)$$

สุดท้าย ปรับ $\hat{\mathbf{a}}_i = \mu_i \mathbf{u}_i$ เพื่อแก้สมการ 19

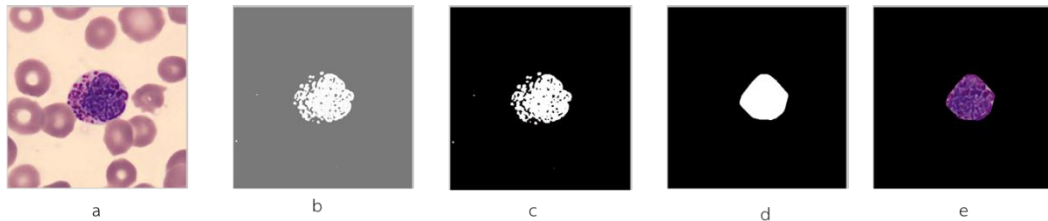
ผลลัพธ์จากระบบ eigen สมการที่ 21 ซึ่งให้ข้อมูล eigenvalue - eigenvector 6 คู่ $(\lambda_i, \mathbf{u}_i)$ โดยแต่ละคู่อันดับจะให้ค่าต่ำสุดถ้าพจน์ของสมการที่ 23 เป็นค่าบวก

โดยทั่วไป S เป็น positive definit ดังนั้น $\mathbf{u}_i^T \mathbf{S} \mathbf{u}_i$ จะเป็นบวกสำหรับทุก \mathbf{u}_i ดังนั้นถ้า $\lambda_i > 0$ ผลลัพธ์ของสมการที่ 19 จะให้ค่า eigen value ที่เป็นบวก



ภาคผนวก ข

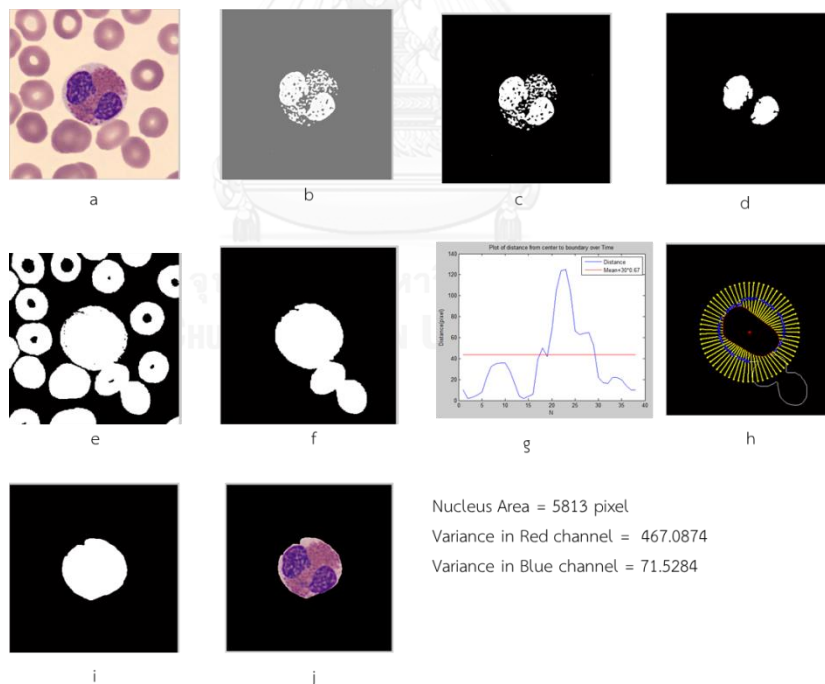
ขั้นตอนการแบ่งส่วนภาพชุดข้อมูลที่ 2 ด้วยวิธีที่นำเสนอ



Nucleus Area = 5728 pixel , Variance in Red channel = 447.9920, Variance in Blue channel = 151.5640

รูปที่ ข. 1 แสดงการแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ basophil

จากรูปที่ ข. 1 a คือ ภาพต้นฉบับ b คือภาพ unsigned integers 8 bit ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization c คือภาพที่แปลงเป็นภาพขาวดำ d คือภาพที่แสดงเฉพาะพื้นที่นิวเคลียส และ e คือ ภาพคอนเวกซ์ฮัลล์ของนิวเคลียส

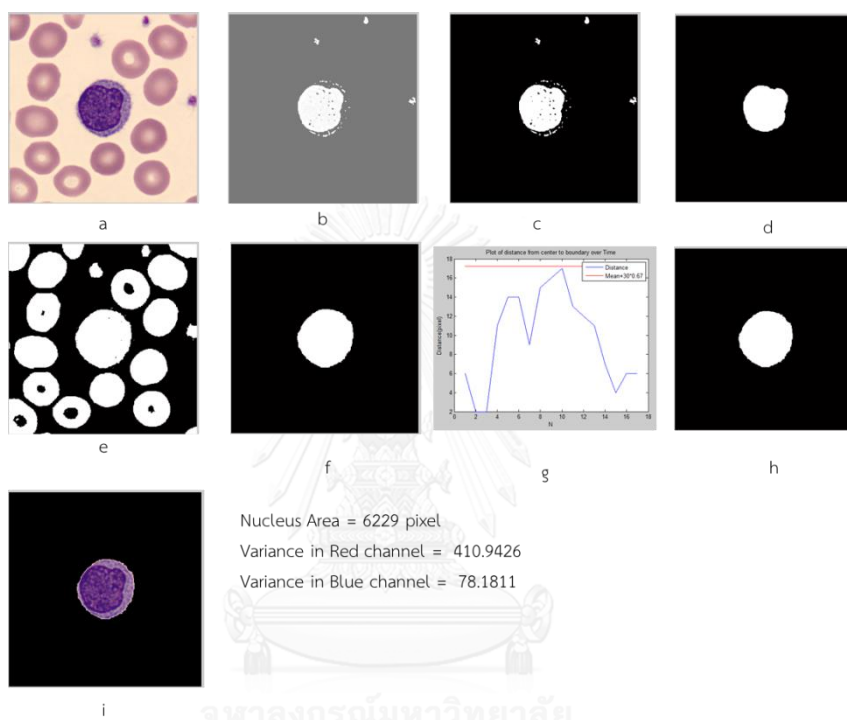


Nucleus Area = 5813 pixel
 Variance in Red channel = 467.0874
 Variance in Blue channel = 71.5284

รูปที่ ข. 2 แสดงการแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ eosinophil

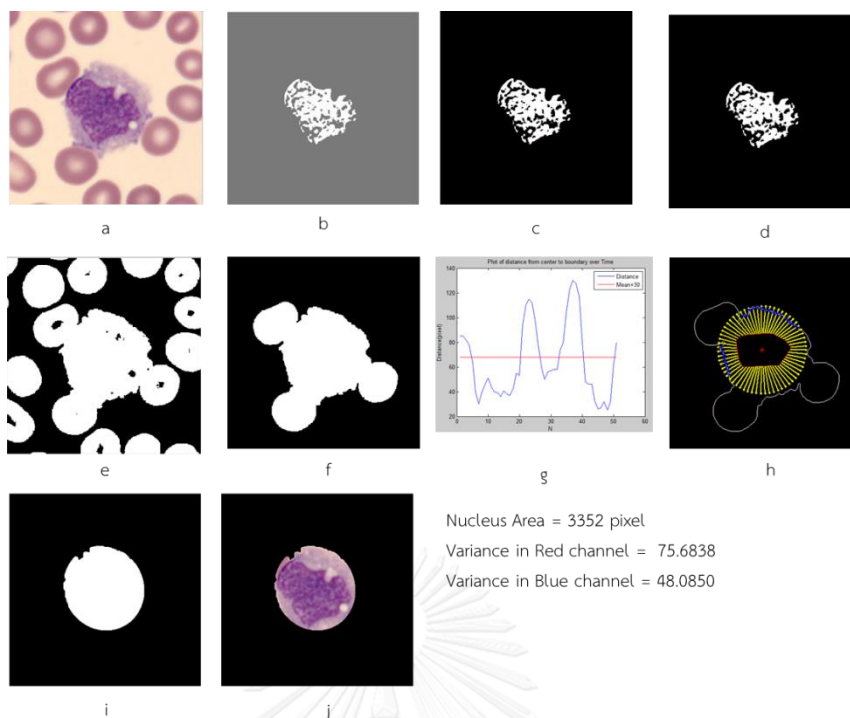
จากรูปที่ ข. 2 a คือ ภาพต้นฉบับ b คือภาพ unsigned integers 8 bit ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization c คือภาพ b ที่แปลงเป็นภาพขาวดำ d คือภาพที่

แสดงเฉพาะพื้นที่นิวเคลียส e คือ ภาพ double precision ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization f คือวัตถุที่สนใจจากภาพ e g คือกราฟแสดงระยะห่างจากศูนย์กลางของภาพคอนเวกซ์ฮัลล์ของนิวเคลียส ไปยังขอบของวัตถุที่สนใจ (g) แสดงด้วยเส้นสีน้ำเงิน ส่วนเส้นสีแดงคือค่าเฉลี่ยของระยะห่างบวก 30×0.67 พิกเซล h คือ ขอบเขตของวัตถุที่สนใจ i คือภาพพื้นที่ของเซลล์ และ j คือ ภาพเซลล์เม็ดเลือดขาวที่สนใจ



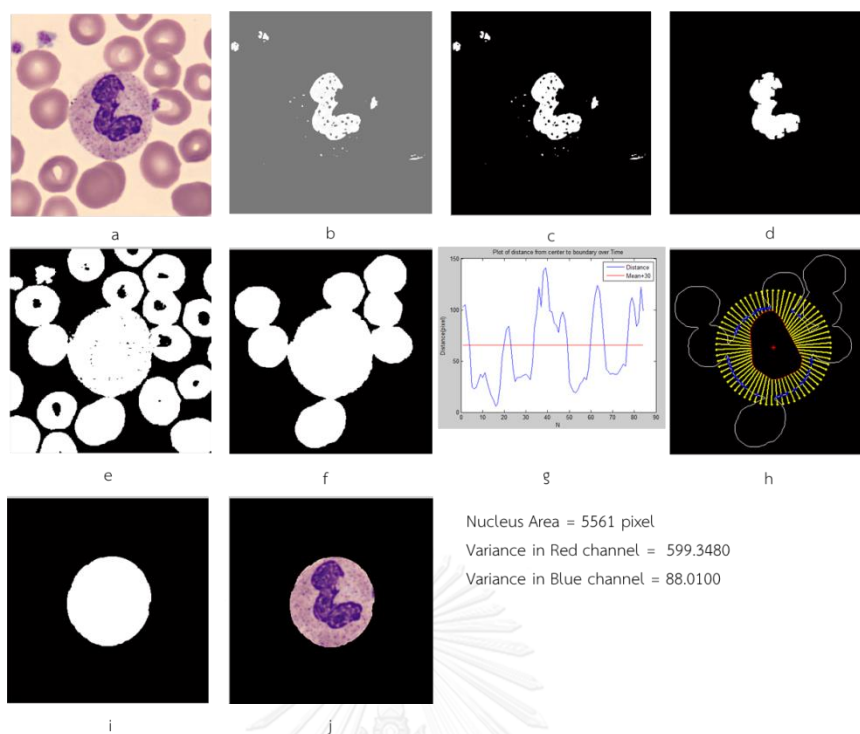
รูปที่ 3 แสดงการแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ lymphocyte

จากรูปที่ 3 a คือ ภาพต้นฉบับ b คือภาพ unsigned integers 8 bit ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization c คือภาพ b ที่แปลงเป็นภาพขาวดำ d คือภาพที่แสดงเฉพาะพื้นที่นิวเคลียส e คือ ภาพ double precision ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization f คือภาพ e ที่แปลงเป็นภาพขาวดำ g คือกราฟแสดงระยะห่างจากศูนย์กลางของภาพ คอนเวกซ์ฮัลล์ของนิวเคลียส ไปยังขอบของวัตถุที่สนใจ (g) แสดงด้วยเส้นสีน้ำเงิน ส่วนเส้นสีแดงคือค่าเฉลี่ยของระยะห่างบวก 30×0.67 พิกเซล คือภาพวัตถุที่สนใจ h คือภาพพื้นที่ของเซลล์ และ i คือ ภาพเซลล์เม็ดเลือดขาวที่สนใจ



รูปที่ ข. 4 แสดงการแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ monocyte

จากรูปที่ ข. 4 a คือ ภาพต้นฉบับ b คือภาพ unsigned integers 8 bit ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization c คือภาพ b ที่แปลงเป็นภาพขาวดำ d คือภาพที่แสดงเฉพาะพื้นที่นิวเคลียส e คือ ภาพ double precision ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization f คือ ภาพวัตถุที่สนใจ g คือกราฟแสดงระยะห่างจากศูนย์กลางของภาพ คอนเวกซ์ฮัลล์ของนิวเคลียส ไปยังขอบของวัตถุที่สนใจ (g) แสดงด้วยเส้นสีน้ำเงิน ส่วนเส้นสีแดงคือค่าเฉลี่ยของระยะห่างบวก 30 พิกเซล h คือการสร้างเส้นรัศมีเพื่อหาจุดที่เป็นของของเซลล์เม็ดเลือดขาวของวัตถุที่สนใจ i คือ ภาพเซลล์ผลลัพธ์ที่ผ่านการ 'AND' ของวงรีที่สร้างโดยเทคนิค least square จากจุดที่ได้จากภาพ h กับ ภาพ f และ j คือ ภาพเซลล์เม็ดเลือดขาวที่ต้องการ



รูปที่ ข. 5 แสดง การแบ่งส่วนภาพนิวเคลียสและไซโทพลาซึมของ neutrophil

จากรูปที่ ข. 5 a คือ ภาพต้นฉบับ b คือภาพ unsigned integers 8 bit ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization c คือภาพ b ที่แปลงเป็นภาพขาวดำ d คือภาพที่แสดงเฉพาะพื้นที่นิวเคลียส e คือ ภาพ double precision ที่ผ่านการแปลง $(R+B)/2G$ และทำ histogram equalization f คือ ภาพวัตถุที่สนใจ g คือกราฟแสดงระยะห่างจากศูนย์กลางของภาพ คอนเวกซ์ฮัลล์ของนิวเคลียส ไปยังขอบของวัตถุที่สนใจ (g) แสดงด้วยเส้นสีน้ำเงิน ส่วนเส้นสีแดงคือค่าเฉลี่ยของระยะห่างบวก 30 พิกเซล h คือการสร้างเส้นรัศมีเพื่อหาจุดที่เป็นของของเซลล์เม็ดเลือดขาวของวัตถุที่สนใจ i คือ ภาพเซลล์ผลลัพธ์ที่ผ่านการ 'AND' ของวงรีที่สร้างโดยเทคนิค least square จากจุดที่ได้จากภาพ h กับ ภาพ f และ j คือ ภาพเซลล์เม็ดเลือดขาวที่ต้องการ

ภาคผนวก ค.

ตารางที่ ค. 1 feature ที่สามารถสกัดได้จากผลลัพธ์ของการแบ่งส่วนภาพ รายละเอียดดังต่อไปนี้

| ลำดับที่ | ชื่อย่อที่ใช้ | ความหมาย |
|----------|---------------|--|
| 1 | meanRNU | ค่าเฉลี่ยความเข้มแสงของช่องสัญญาณสีแดงภายในพื้นที่นิวเคลียส |
| 2 | meanGNU | ค่าเฉลี่ยความเข้มแสงของช่องสัญญาณสีเขียวภายในพื้นที่นิวเคลียส |
| 3 | meanBNU | ค่าเฉลี่ยความเข้มแสงของช่องสัญญาณสีน้ำเงินภายในพื้นที่นิวเคลียส |
| 4 | VRNU | ค่าความแปรปรวนของความเข้มแสงของช่องสัญญาณสีแดงภายในพื้นที่นิวเคลียส |
| 5 | VGNU | ค่าความแปรปรวนของความเข้มแสงของช่องสัญญาณสีเขียวภายในพื้นที่นิวเคลียส |
| 6 | VBNU | ค่าความแปรปรวนของความเข้มแสงของช่องสัญญาณสีน้ำเงินภายในพื้นที่นิวเคลียส |
| 7 | meanRCy | ค่าเฉลี่ยความเข้มแสงของช่องสัญญาณสีแดง ภายในพื้นที่ไซโทพลาซึม |
| 8 | meanGCy | ค่าเฉลี่ยความเข้มแสงของช่องสัญญาณสีเขียว ภายในพื้นที่ไซโทพลาซึม |
| 9 | meanBCy | ค่าเฉลี่ยความเข้มแสงของช่องสัญญาณสีน้ำเงิน ภายในพื้นที่ไซโทพลาซึม |
| 10 | VRCy | ค่าความแปรปรวนของความเข้มแสงของช่องสัญญาณสีแดงภายในพื้นที่ไซโทพลาซึม |
| 11 | VGCy | ค่าความแปรปรวนของความเข้มแสงของช่องสัญญาณสีเขียวภายในพื้นที่ไซโทพลาซึม |
| 12 | VBCy | ค่าความแปรปรวนของความเข้มแสงของช่องสัญญาณสีน้ำเงินภายในพื้นที่ไซโทพลาซึม |
| 13 | meanRCell | ค่าเฉลี่ยความเข้มแสงของช่องสัญญาณสีแดง ภายในพื้นที่เซลล์ |
| 14 | meanGCell | ค่าเฉลี่ยความเข้มแสงของช่องสัญญาณสีเขียว ภายในพื้นที่เซลล์ |
| 15 | meanBCell | ค่าเฉลี่ยความเข้มแสงของช่องสัญญาณสีน้ำเงิน ภายในพื้นที่เซลล์ |
| 16 | VRCell | ค่าความแปรปรวนของความเข้มแสงของช่องสัญญาณสีแดงภายในพื้นที่เซลล์ |
| 17 | VGCell | ค่าความแปรปรวนของความเข้มแสงของช่องสัญญาณสีเขียวภายในพื้นที่เซลล์ |

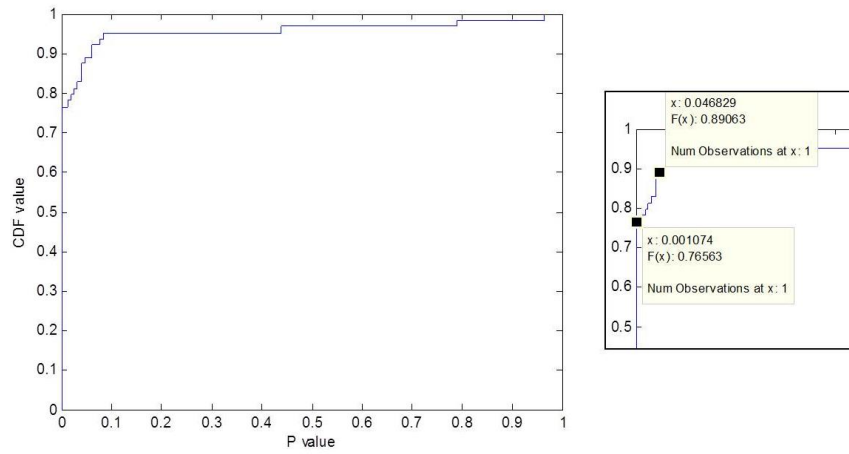
| ลำดับที่ | ชื่อย่อที่ใช้ | ความหมาย |
|----------|---------------------|--|
| 18 | VBCell | ค่าความแปรปรวนของความเข้มแสงของช่องสัญญาณสีน้ำเงินภายในพื้นที่เซลล์ |
| 19 | concF | จำนวน พื้นที่เว้าเข้าของนิวเคลียส |
| 20 | NuLobe | จำนวนก้อน(lobe) ของนิวเคลียส |
| 21 | AreaRatio | อัตราส่วนของจำนวนพิกเซลในพื้นที่นิวเคลียสเทียบกับจำนวนพิกเซลในพื้นที่เซลล์ |
| 22 | NUArea | จำนวนพิกเซลในพื้นที่นิวเคลียส |
| 23 | NUPerimeter | เส้นรอบวงของนิวเคลียส |
| 24 | NUeccentricity | อัตราส่วนความยาวแกนหลักต่อแกนรองของนิวเคลียส |
| 25 | NUMajorAxisLength | ความยาวแกนหลักของนิวเคลียส |
| 26 | NUMinorAxisLength | ความยาวแกนรองของนิวเคลียส |
| 27 | NUcompactness | อัตราส่วนของพื้นที่ของนิวเคลียสเทียบกับพื้นที่วงกลมที่มีเส้นรอบวงเหมือนกัน |
| 28 | NUroundness | ค่าความกลมของนิวเคลียส |
| 29 | CellArea | จำนวนพิกเซลในเซลล์ |
| 30 | CellPerimeter | เส้นรอบวงของเซลล์ |
| 31 | Celleccentricity | อัตราส่วนความยาวแกนหลักต่อแกนรองของเซลล์ |
| 32 | CellMajorAxisLength | ความยาวแกนหลักของเซลล์ |
| 33 | CellMinorAxisLength | ความยาวแกนรองของเซลล์ |
| 34 | Cellcompactness | อัตราส่วนของพื้นที่ของเซลล์เทียบกับพื้นที่วงกลม ที่มีเส้นรอบวงเหมือนกัน |
| 35 | Cellroundness | ค่าความกลมของเซลล์ |
| 36 | entropyNUR | เอนโทรปีของนิวเคลียสในช่องสีแดง |
| 37 | entropyNUG | เอนโทรปีของนิวเคลียสในช่องสีเขียว |
| 38 | entropyNUB | เอนโทรปีของนิวเคลียสในช่องสีน้ำเงิน |
| 39 | entropyCellR | เอนโทรปีของเซลล์ในช่องสีแดง |
| 40 | entropyCellG | เอนโทรปีของเซลล์ในช่องสีเขียว |

| ลำดับที่ | ชื่อย่อที่ใช้ | ความหมาย |
|----------|---------------|--|
| 41 | entropyCellB | เอนโทรปีของเซลล์ในช่องสีน้ำเงิน |
| 42 | entropyCyR | เอนโทรปีของไซโทพลาซึมในช่องสีแดง |
| 43 | entropyCyG | เอนโทรปีของไซโทพลาซึมในช่องสีเขียว |
| 44 | entropyCyB | เอนโทรปีของไซโทพลาซึมในช่องสีน้ำเงิน |
| 45 | energyNUR | พลังงานของนิวเคลียสในช่องสีแดง |
| 46 | energyNUG | พลังงานของนิวเคลียสในช่องสีเขียว |
| 47 | energyNUB | พลังงานของนิวเคลียสในช่องสีน้ำเงิน |
| 48 | energyCellR | พลังงานของเซลล์ในช่องสีแดง |
| 49 | energyCellB | พลังงานของเซลล์ในช่องสีเขียว |
| 50 | energyCellC | พลังงานของเซลล์ในช่องสีน้ำเงิน |
| 51 | energyCyR | พลังงานของไซโทพลาซึมในช่องสีแดง |
| 52 | energyCyB | พลังงานของไซโทพลาซึมในช่องสีเขียว |
| 53 | energyCyC | พลังงานของไซโทพลาซึมในช่องสีน้ำเงิน |
| 54 | NuR/NuG | อัตราส่วนค่าเฉลี่ยสี่เทาของพื้นที่นิวเคลียส ช่องสีแดงเทียบกับช่องสีเขียว |
| 55 | NuR/NuB | อัตราส่วนค่าเฉลี่ยสี่เทาของพื้นที่นิวเคลียส ช่องสีแดงเทียบกับช่องสีน้ำเงิน |
| 56 | CyR/CyG | อัตราส่วนค่าเฉลี่ยสี่เทาของพื้นที่ไซโทพลาซึมช่องสีแดงเทียบกับช่องสีเขียว |
| 57 | CyR/CyB | อัตราส่วนค่าเฉลี่ยสี่เทาของพื้นที่ไซโทพลาซึมช่องสีแดงเทียบกับช่องสีน้ำเงิน |
| 58 | CellR/CellG | อัตราส่วนค่าเฉลี่ยสี่เทา ของพื้นที่เซลล์ ช่องสีแดงเทียบกับช่องสีเขียว |
| 59 | CellR/CellB | อัตราส่วนค่าเฉลี่ยสี่เทาของพื้นที่เซลล์ ช่องสีแดงเทียบกับช่องสีน้ำเงิน |
| 60 | NuG/NuR | อัตราส่วนค่าเฉลี่ยสี่เทาของพื้นที่นิวเคลียส ช่องสีเขียวเทียบกับช่องสีแดง |
| 61 | NuB/NuR | อัตราส่วนค่าเฉลี่ยสี่เทาของพื้นที่นิวเคลียส ช่องสีน้ำเงินเทียบกับช่องสีแดง |
| 62 | CyG/CyR | อัตราส่วนค่าเฉลี่ยสี่เทาของพื้นที่ไซโทพลาซึมช่องสีเขียวเทียบกับช่องสีแดง |
| 63 | CyB/CyR | อัตราส่วนค่าเฉลี่ยสี่เทาของพื้นที่ไซโทพลาซึมช่องสีน้ำเงินเทียบกับช่องสีแดง |

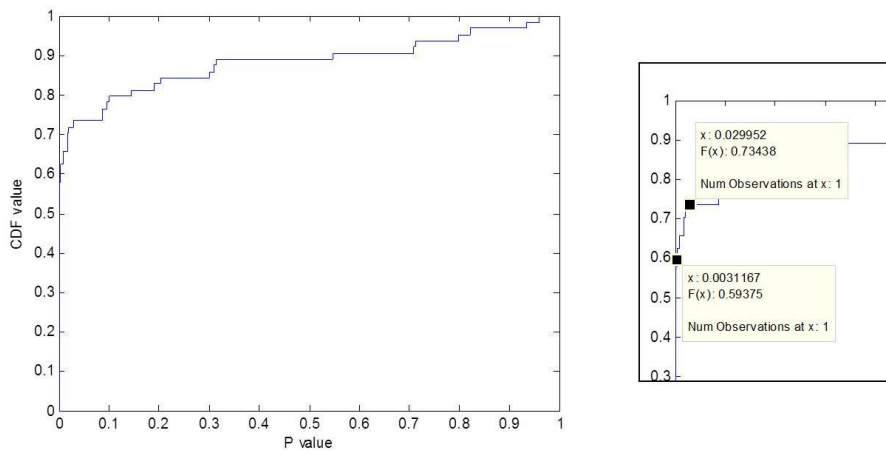
| ลำดับที่ | ชื่อย่อที่ใช้ | ความหมาย |
|----------|---------------|--|
| 64 | CellG/CellR | อัตราส่วนค่าเฉลี่ยสีเทาของพื้นที่เซลล์ช่องสีเขียวเทียบกับช่องสีแดง |
| 65 | CellB/CellR | อัตราส่วนค่าเฉลี่ยสีเทาของพื้นที่เซลล์ช่องสีน้ำเงินเทียบกับช่องสีแดง |



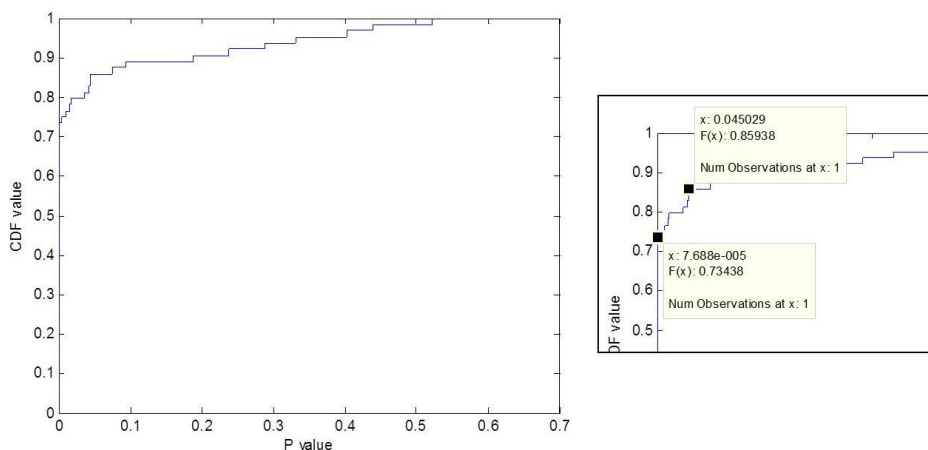
ภาคผนวก ง.



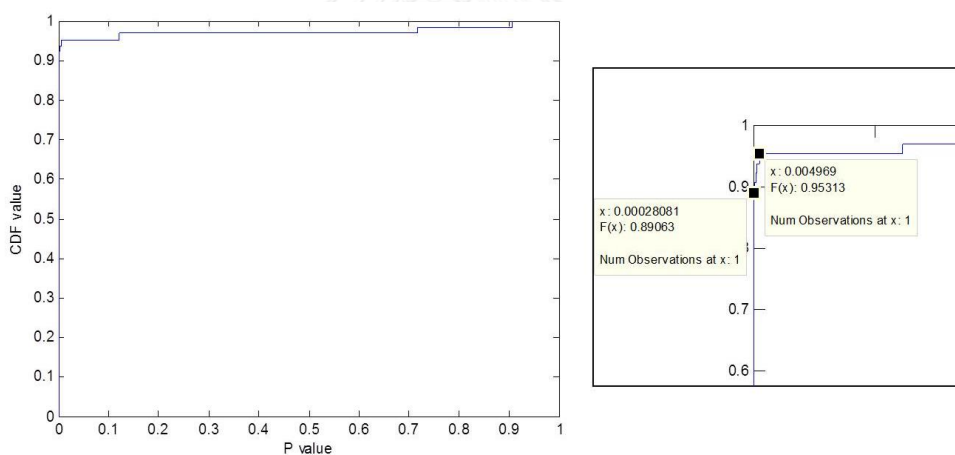
รูปที่ ง. 1 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 1 และ 3



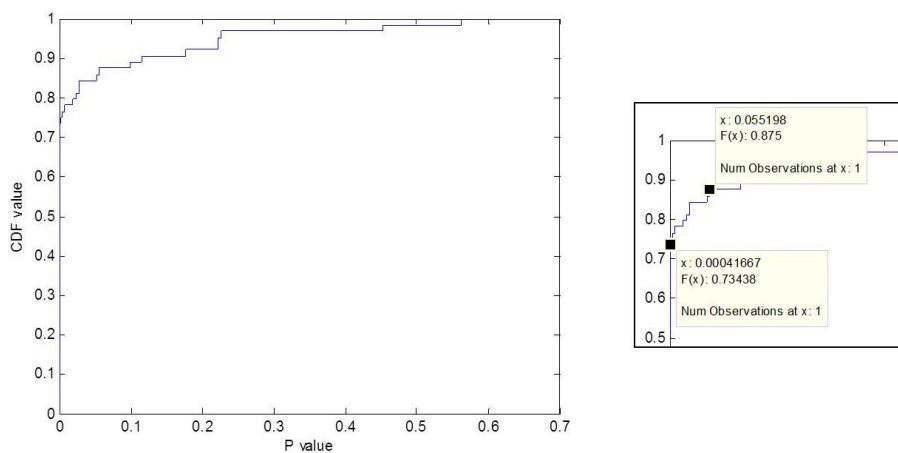
รูปที่ ง. 2 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 1 และ 4



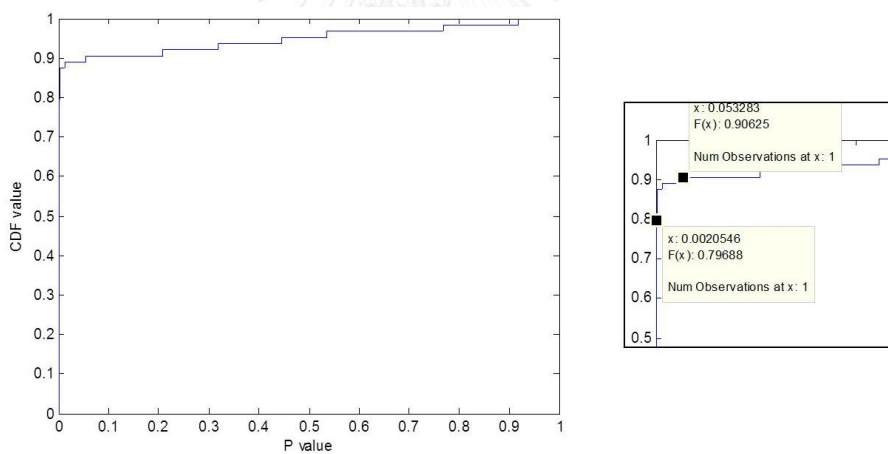
รูปที่ ง. 3 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 1 และ 5



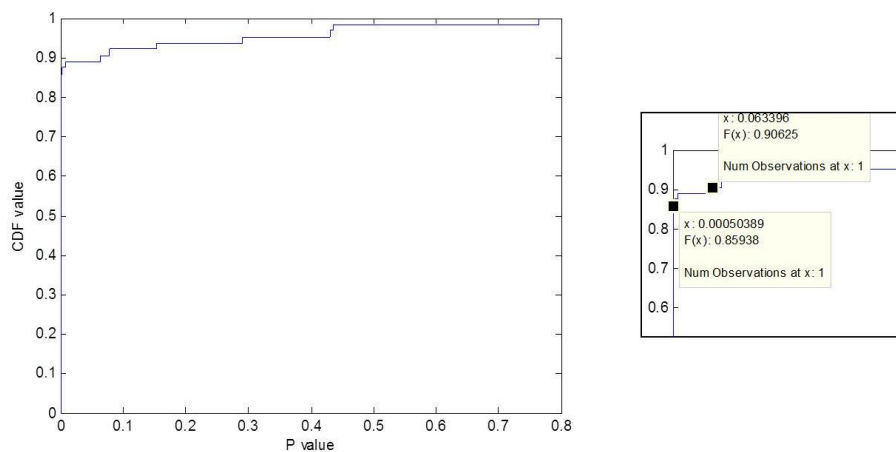
รูปที่ ง. 4 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 2 และ 3



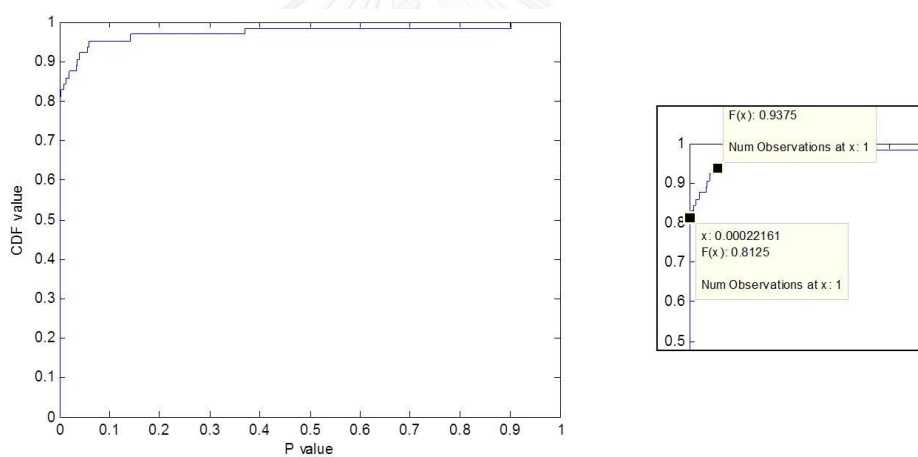
รูปที่ ๕. 5 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 2 และ 4



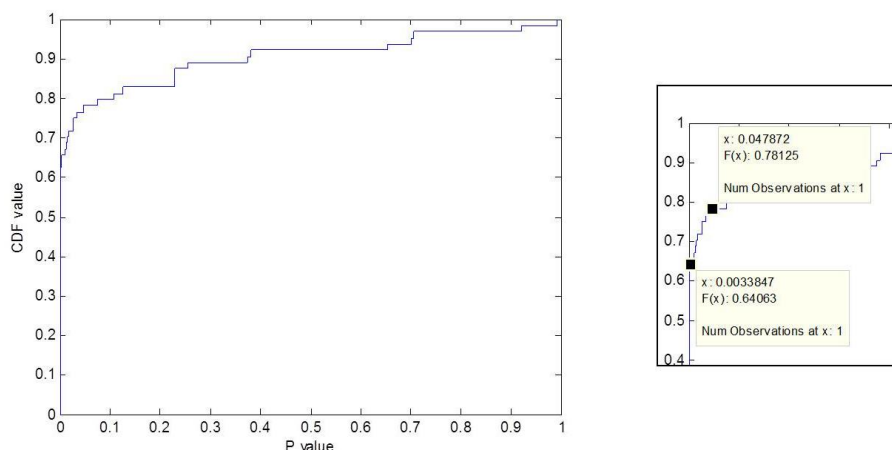
รูปที่ ๖. 6 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 2 และ 5



รูปที่ ง. 7 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 3 และ 4



รูปที่ ง. 8 กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และส่วนขยาย ของกรณีเปรียบเทียบเซลล์ชนิดที่ 3 และ 5



รูปที่ 9. กราฟการกระจาย cumulative distribution function (CDF) ของ p-values และส่วนขยายของกรณีเปรียบเทียบเซลล์ชนิดที่ 4 และ 5

ตารางที่ 9. 1 แสดงข้อมูลลักษณะบ่งต่างที่เลือกจากกราฟ การกระจาย cumulative distribution function (CDF) ของ p-values

| เซลล์ชนิดที่ | % ลักษณะบ่งต่างที่ค่า $p \cong 0$ | % ลักษณะบ่งต่างที่ค่า $p < 0.05$ | จำนวนลักษณะบ่งต่างที่ต้องการ | ลักษณะบ่งต่างที่เรียงลำดับ ตามจำนวนที่ต้องการ |
|--------------|-----------------------------------|----------------------------------|------------------------------|--|
| 1 และ 2 | 40 | 70 | 19 | 9, 4, 31, 6, 7, 24, 11, 12, 60, 8, 57, 17, 63, 10, 5, 54, 15, 26, 25 |
| 1 และ 3 | 77 | 89 | 7 | 2, 18, 54, 3, 13, 1, 25 |
| 1 และ 4 | 59 | 73 | 17 | 11, 62, 4, 64, 43, 42, 56, 40, 41, 39, 52, 53, 51, 6, 29, 16, 60 |
| 1 และ 5 | 73 | 86 | 9 | 18, 42, 2, 54, 39, 1, 43, 16, 41 |
| 2 และ 3 | 89 | 95 | 3 | 54, 2, 60 |
| 2 และ 4 | 73 | 88 | 8 | 62, 56, 8, 63, 57, 64, 58, 9 |
| 2 และ 5 | 80 | 90 | 6 | 54, 17, 2, 60, 8, 4 |
| 3 และ 4 | 86 | 90 | 6 | 54, 2, 58, 60, 32, 64 |
| 3 และ 5 | 81 | 94 | 4 | 44, 51, 53, 52 |
| 4 และ 5 | 64 | 78 | 14 | 54, 16, 4, 2, 1, 60, 6, 3, 18, 17, 57, 63, 11, 22 |

ประวัติผู้เขียนวิทยานิพนธ์

ชื่อ-สกุล จรูญรัตน์ ปริญาคุปต์
 วัน เดือน ปี เกิด 8 กุมภาพันธ์ 2519
 ที่อยู่ปัจจุบัน 77/31 หมู่บ้านเมอริทแกรนด์ แขวงสีกัน เขตดอนเมือง
 กรุงเทพฯ 10210

e-mail jaroonrut.p@rsu.ac.th

ประวัติการศึกษา

ระดับปริญญาตรี จาก ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์
 มหาวิทยาลัยสงขลานครินทร์

ระดับปริญญาโท จาก สถาบันวิจัยและพัฒนาวิทยาศาสตร์และเทคโนโลยี โครงการวิจัย
 และพัฒนาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหิดล

งานวิจัยที่ได้รับการตีพิมพ์เผยแพร่

1. Prinyakupt, J. and C. Pluempitiwiriyaewej, White Blood Cell Identification and Segmentation , The 29th International Technical Conference on Circuit/Systems Computers and Communications (ITC-CSCC), Phuket, Thailand, July 1-4, 2014

2. Prinyakupt, J. and C. Pluempitiwiriyaewej, Segmentation of white blood cells and comparison of cell morphology by linear and naive Bayes classifiers. Biomed Eng Online, 2015. 14(1): p. 63.