

การนับแบบแบ่งกลุ่ม (Differential Count) ของเม็ดเลือดขาว คือการนับจำนวนเม็ดเลือดขาวแต่ละชนิดในเลือดหรือไขกระดูก การนับแบบนี้มีส่วนช่วยแพทย์ผู้เชี่ยวชาญในการวินิจฉัยโรค เช่น ลูคีเมีย หรือมะเร็งได้ การนับแบบแบ่งกลุ่มของเม็ดเลือดขาวในไขกระดูก จะให้ข้อมูลกับแพทย์ผู้เชี่ยวชาญได้มากกว่าการนับในเลือด เนื่องจากเม็ดเลือดขาวที่มีอายุน้อยจะพบได้ในไขกระดูกเท่านั้น อย่างไรก็ตามการนับในไขกระดูกจะทำได้ยากกว่าในเลือดมาก ไม่ว่าจะเป็นการนับโดยใช้ผู้เชี่ยวชาญหรือโดยระบบอัตโนมัติ

เม็ดเลือดขาวใน Myelocytic Series แบ่งออกได้เป็น 6 กลุ่มตามอายุ ได้แก่ Myeloblast, Promyelocyte, Myelocyte, Metamyelocyte, Band, และ PolyMorphoNuclear (PMN) การนับโดยแพทย์ผู้เชี่ยวชาญในปัจจุบันเป็นงานที่หนักและใช้เวลานาน ถึงแม้ว่าจะมีระบบการนับแบบอัตโนมัติในเลือดในเชิงพาณิชย์ แต่ก็ยังมีราคาแพงมาก และยังไม่มีการนับแบบอัตโนมัติในภาพไขกระดูก ดังนั้นระบบการนับแบบอัตโนมัติในภาพไขกระดูกจะมีส่วนช่วยได้มาก

โครงการวิจัยนี้ เป็นส่วนหนึ่งของการสร้างระบบการนับเม็ดเลือดขาวแบบแบ่งกลุ่มในภาพไขกระดูกแบบอัตโนมัติ โดยทำการแบ่งส่วนและจำแนกภาพเม็ดเลือดขาวที่ในแต่ละภาพมีเซลล์เม็ดเลือดขาวเพียงเซลล์เดียว หรือภาพเซลล์เดี่ยว (Single-Cell Image) โดยแบ่งส่วนภาพเซลล์เดี่ยวแต่ละภาพออกเป็น 3 ส่วน ได้แก่ นิวเคลียส ไซโตพลาสซึม และฉากหลัง และทำการจำแนกเซลล์เม็ดขาวออกเป็น 6 กลุ่มตาม Myelocytic Series โดยใช้ Bayes Classifier และ Neural Networks Classifier ข้อมูลภาพที่ใช้ได้มาจาก Ellis-Fishel Cancer Center, University of Missouri เมืองโคลัมเบีย มลรัฐมิสซูรี ประเทศสหรัฐอเมริกา

ในการแบ่งส่วนเซลล์ เราใช้วิธีการ Fuzzy C-Means และ Mathematical Morphology จากการทดลองพบว่า ค่าผิดพลาดโดยรวมในการแบ่งส่วนนิวเคลียสมีค่าประมาณ 10.3 % ส่วนค่าผิดพลาดโดยรวมในการแบ่งส่วนนิวเคลียสมีค่าประมาณ 14.8 % อย่างไรก็ตาม เมื่อตัดภาพเซลล์บางภาพที่ไม่สมบูรณ์ออก พบว่าค่าผิดพลาดโดยรวมในการแบ่งส่วนนิวเคลียส และไซโตพลาสซึมมีค่าลดลงเหลือ 9.6 % และ 13.6 % ตามลำดับ ส่วนค่าผิดพลาดโดยรวมในการแบ่งส่วนเซลล์ทั้งเซลล์มีค่าเป็น 8.8 %

ลักษณะสำคัญที่ใช้ในการจำแนกเซลล์ในโครงการวิจัยนี้อยู่บนพื้นฐานของ Pattern Spectrum เปอร์เซนต์ความถูกต้องในการจำแนกโดยใช้ Neural Network มีค่า 66 % เมื่อใช้ลักษณะสำคัญจากภาพที่ได้จากการแบ่งส่วนโดยอัตโนมัติ ผลที่ได้เป็นที่น่าพอใจเมื่อเทียบกับเปอร์เซนต์ความถูกต้องในการจำแนก 70 % เมื่อใช้ลักษณะสำคัญจากภาพที่ได้จากการแบ่งส่วนโดยผู้เชี่ยวชาญ อย่างไรก็ตาม เราพบว่าตัวจำแนกมีไบอัสไปยังกลุ่มเซลล์ที่มีจำนวนเซลล์ในชุดข้อมูลฝึกฝนมาก เราได้วิเคราะห์ปัญหาการจำแนกเซลล์เม็ดเลือดขาวโดยใช้เฉพาะลักษณะสำคัญของนิวเคลียส และนำเสนอวิธีการลดการไบอัสทั้งใน Bayes Classifier และ Neural Network Classifier จากการทดลองพบว่าเมื่อตัวจำแนกลดการไบอัสลง จะได้ว่า Class-wise Rate มีค่าสูงขึ้น นอกจากนั้น ยังพบว่าการจำแนกที่ใช้เฉพาะข้อมูลนิวเคลียสให้ผลที่น่าพอใจ โดยให้ค่า (Traditional)

Classification Rate เป็น 77 % และให้ค่า Class-wise Rate เป็น 61 % สำหรับชุดข้อมูลทดสอบ โดย Neural Network Classifier ที่ลดการไบอัสเป็นตัวจำแนกที่ให้ค่าเปอร์เซ็นต์การจำแนกสูงทั้งในแบบดั้งเดิมและแบบ Class-wise

เรายังได้ทำการวิเคราะห์ Subset ของลักษณะสำคัญทั้ง 4 อันที่น่าเสนอ และยังสามารถเปรียบเทียบผลการจำแนกที่ใช้ลักษณะสำคัญที่เรานำเสนอ และลักษณะสำคัญที่มีนักวิจัยนำเสนอไว้ก่อนหน้านี้โดยใช้ฐานข้อมูลเซลล์เดียวกัน ผลการทดลองแสดงให้เห็นว่า ลักษณะสำคัญที่เรานำเสนอให้ผลการจำแนกที่ดีกว่า นอกจากนี้ เรายังได้ทดสอบการจำแนกของ Naive Bayes Classifier และ C4.5 Decision Tree จากการทดลองพบว่าตัวจำแนกทั้งสองนี้ให้ผลการจำแนกด้อยกว่า Bayes Classifier และ Neural Network Classifier

Differential counts of white blood cells are the counts of different classes of white blood cells in peripheral blood or bone marrow. They aid in the diagnosis of diseases, such as leukemia or cancer. The differential counts in bone marrow generally provide more information than those in peripheral blood because immature cells can only be found in bone marrow. However, the counting either by an expert or an automatic system in bone marrow is much more difficult than in peripheral blood.

White blood cells in Myelocytic series are classified into 6 classes according to their ages, i.e., Myeloblast, Promyelocyte, Myelocyte, Metamyelocyte, Band, and PolyMorphoNuclear (PMN). The counting by an expert is very tedious and time-consuming. An automatic system for bone marrow will help. Even though there exist a few commercial automatic counting systems for peripheral blood, they are very expensive and there is none for bone marrow.

This research project is a part of an automatic bone marrow white blood cell differential counting system. The aims are to segment single-cell images into 3 regions, i.e., nucleus, cytoplasm, and background, and classify the images into 6 classes according to the myelocytic series using Bayes and neural network classifiers. The cell data set used in the experiments was collected at Ellis-Fishel Cancer Center, University of Missouri-Columbia, Missouri, U.S.A.

We applied the fuzzy C-means algorithm and mathematical morphology in the image segmentation. From the experiments, the overall nucleus segmentation error is about 10.3 %, whereas the overall cytoplasm segmentation error is about 14.8 %. When we discard some imperfect cell images, however, the overall nucleus and cytoplasm segmentation errors are decreased to 9.6 % and 13.6 %, respectively. The overall entire cell segmentation error is 8.8 %.

The proposed cell features are based on the pattern spectrum. The classification rate achieved by neural network classifiers is 66 % when the automatic-segmented images are applied. This result is promising compared to 70 % classification rate achieved when the expert's hand-segmented images are applied. However, the classifiers are biased toward the cell class with a large number of training samples. We analyzed the cell classification problem using only features extracted from nuclei. We also proposed methods to unbiased Bayes and neural network classifiers. The results suggest that the unbiased classifiers achieve higher class-wise classification rate. Moreover,

classification using only nucleus features is promising. The unbiased neural network classifier achieves 77 % traditional, and 61 % class-wise classification rates that are considered high in both the traditional and class-wise senses.

We also analyze the subset of 4 proposed features, and compare the classification performance achieved by our proposed features to the previously proposed ones. The results show that our proposed features yield better classification performance. Furthermore, we test the classification problem using the naive Bayes classifier and C4.5 decision tree. From the experiments, the classification rates achieved by these 2 classifiers are lower than that achieved by Bayes classifier and neural network classifier.