



การสกัดและคัดเลือกคุณลักษณะของเมล็ดกาแฟเพื่อจำแนกประเภทตามคุณภาพ
ด้วยการประมวลผลภาพถ่าย

COFFEE BEANS FEATURE EXTRACTION AND SELECTION
FOR QUALITY CLASSIFICATION USING IMAGE PROCESSING

แสงดาว เดือนแจ่ม

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

2563

การสกัดและคัดเลือกคุณลักษณะของเมื่อดกาแฟเพื่อจำแนกประเภทตามคุณภาพ
ด้วยการประมวลผลภาพถ่าย



สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2563
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

COFFEE BEANS FEATURE EXTRACTION AND SELECTION
FOR QUALITY CLASSIFICATION USING IMAGE PROCESSING



A Master's Project Submitted in Partial Fulfillment of the Requirements
for the Degree of MASTER OF SCIENCE
(Information Technology)
Faculty of Science, Srinakharinwirot University
2020
Copyright of Srinakharinwirot University

สารนิพนธ์

เรื่อง

การสกัดและคัดเลือกคุณลักษณะของเมล็ดกาแฟเพื่อจำแนกประเภทตามคุณภาพ

ด้วยการประมวลผลภาพถ่าย

ของ

แสงดาว เตือนแจ่ม

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)

คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าสารนิพนธ์

..... ที่ปรึกษาหลัก ประธาน
(ผู้ช่วยศาสตราจารย์ ดร.นุรีย์ วิวัฒน์วัฒนา) (ผู้ช่วยศาสตราจารย์ ดร.วีรพล จิรจิต)

..... ที่ปรึกษาร่วม กรรมการ
(รองศาสตราจารย์ ดร.ทิมพันธุ์ เจริญพงษ์) (ผู้ช่วยศาสตราจารย์ ดร.จันตรี ผลประเสริฐ)

ชื่อเรื่อง	การสกัดและคัดเลือกคุณลักษณะของเมล็ดกาแฟเพื่อจำแนกประเภทตามคุณภาพ
	ด้วยการประมวลผลภาพถ่าย
ผู้วิจัย	แสงดาว เตือนแจ่ม
ปริญญา	วิทยาศาสตร์มหาบัณฑิต
ปีการศึกษา	2563
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. นุรีย์ วิวัฒน์วัฒนา
อาจารย์ที่ปรึกษาร่วม	รองศาสตราจารย์ ดร. ทิฆัมพันธุ์ เจริญพงษ์

งานวิจัยนี้นำเสนอวิธีการสกัดคุณลักษณะของภาพถ่ายเมล็ดกาแฟ เพื่อจำแนกประเภทตามคุณภาพของเมล็ดกาแฟ ตัวอย่างที่ใช้ในการศึกษาแบ่งออกเป็น 2 กลุ่ม โดยกลุ่มตัวอย่างชุดข้อมูลสมดุล มีจำนวนเมล็ดดีและเมล็ดเสียอย่างละ 1,303 เมล็ด และชุดข้อมูลไม่สมดุล มีจำนวนเมล็ดดี 1,303 เมล็ด และเมล็ดเสียจำนวน 4,607 เมล็ด โดยผ่านกระบวนการ 5 ขั้นตอน คือ 1) การเตรียมข้อมูล 2) การประมวลผลภาพเบื้องต้น 3) การสกัดคุณลักษณะ 4) การวิเคราะห์คุณลักษณะ และ 5) การจำแนกประเภทเมล็ดกาแฟ สำหรับการประมวลผลภาพเบื้องต้นประกอบด้วย (1) การแยกเมล็ดกาแฟจากภาพพื้นหลัง (2) การเติมเต็มส่วนขาดหาย (3) การตรวจจับขอบของเมล็ดกาแฟ และสำหรับการสกัดคุณลักษณะ ได้ทำการสกัดคุณลักษณะจำนวน 3 กลุ่ม โดยมีคุณลักษณะทั้งหมด 32 คุณลักษณะ ประกอบด้วย (1) กลุ่มของขนาดและรูปร่าง จำนวน 4 คุณลักษณะ (2) กลุ่มของค่าสี จำนวน 24 คุณลักษณะ (3) กลุ่มของค่าสีเทา จำนวน 4 คุณลักษณะ และ สำหรับการจำแนกเมล็ดกาแฟ ใช้เทคนิคการเรียนรู้ของเครื่องในการพัฒนาแบบจำลองพื้นฐาน 2 แบบจำลอง ประกอบด้วย Logistic Regression และ Random Forest จากผลการทดลองได้แสดงให้เห็นว่าแบบจำลอง Random Forest มีประสิทธิภาพสูงกว่าเล็กน้อย ยกเว้นค่า Recall ของชุดข้อมูลไม่สมดุล ที่มีค่าต่ำกว่าเล็กน้อย โดยที่ค่าความแม่นยำ ชุดข้อมูลสมดุลคิดเป็น 92.34 เปอร์เซ็นต์ ชุดข้อมูลไม่สมดุลคิดเป็น 92.13 เปอร์เซ็นต์ ค่า Precision และค่า Recall ชุดข้อมูลสมดุลคิดเป็น 92.35 เปอร์เซ็นต์ และ 92.34 เปอร์เซ็นต์ ชุดข้อมูลไม่สมดุลคิดเป็น 89.89 เปอร์เซ็นต์ และ 86.58 เปอร์เซ็นต์ ค่าเฉลี่ย F1 ชุดข้อมูลสมดุลคิดเป็น 92.34 เปอร์เซ็นต์ และชุดข้อมูลไม่สมดุลคิดเป็น 88.08 เปอร์เซ็นต์ ในการวิจัยนี้แสดงให้เห็นว่าการจำแนกประเภทเมล็ดกาแฟเป็นเมล็ดกาแฟดีและเมล็ดกาแฟเสียโดยใช้ภาพถ่ายบนพื้นฐานของแบบจำลอง Random Forest สามารถนำไปประยุกต์ใช้ในการคัดแยกเมล็ดกาแฟในโรงกาแฟได้ต่อไป

คำสำคัญ : เมล็ดกาแฟ, การประมวลผลภาพถ่าย, การจำแนกภาพ, การเรียนรู้ด้วยเครื่อง

Title	COFFEE BEANS FEATURE EXTRACTION AND SELECTION FOR QUALITY CLASSIFICATION USING IMAGE PROCESSING
Author	SAENGDAO DUANJAM
Degree	MASTER OF SCIENCE
Academic Year	2020
Thesis Advisor	Assistant Professor Nuwee Wiwatwattana
Co Advisor	Associate Professor Theekapun Charoenpong

This research aims to propose feature extraction methods from the images of coffee beans for quality classification. The samples used were separated into two groups, and the balanced data reported an equal amount of 1,303 good quality and bad quality beans, and the imbalanced data reported 1,303 good quality beans and 4,607 bad quality beans. The entire process consisted of five steps: data preparation, image preprocessing, feature extraction, feature analysis, and coffee bean classification. Firstly, pictures of coffee beans acquired from fixed settings were processed with foreground segmentation, lost pixel retrieval, and edge detection. There were three groups of features and a total of 32 features, were extracted from the images of coffee beans. The four features are size and shape, twenty-four are color-related and four are grayscale-related. The machine learning techniques used for the classification were Logistic Regression and Random Forest. The results showed that for both balanced and imbalanced data, the Random Forest model yielded slightly better classification performance than the Logistic Regression model with hyperparameter tuning and cross-validation, except for the recall in the imbalanced situation. The accuracy of the Random Forest model on the balanced data was 92.34 % and the imbalanced data was 92.13 %. The precision and recall on the balanced data were 92.35 % and 92.34 % and the imbalanced data were 89.89 and 86.58 %, respectively. The F1-score on the balanced data was 92.34 % and the imbalanced data was 88.08 %. The research suggested that coffee bean images classified with the Random Forest model can be applied in real settings in coffee plantations.

Keyword : Coffee Beans, Image Processing, Image Classification, Machine Learning

กิตติกรรมประกาศ

สารนิพนธ์นี้เสร็จลุล่วงได้ด้วยความช่วยเหลือจาก ผู้ช่วยศาสตราจารย์ ดร.นุวิทย์ วิวัฒน์
วัฒนา อาจารย์ที่ปรึกษาหลัก และรองศาสตราจารย์ ดร.ทิมพันธุ์ เจริญพงษ์ อาจารย์ที่ปรึกษาร่วม ที่
ให้คำปรึกษา คำแนะนำในการทำสารนิพนธ์ ตลอดจนสนับสนุนข้อมูลทางวิชาการ และข้อมูลสำหรับ
การทำสารนิพนธ์นี้

ขอกราบขอบพระคุณคณะกรรมการสอบสารนิพนธ์ ที่ได้ให้คำแนะนำ และข้อเสนอแนะอัน
เป็นประโยชน์สำหรับการปรับปรุงสารนิพนธ์ และการใช้เทคนิคการเรียนรู้ของเครื่องเป็นเครื่องมือใน
การวิเคราะห์ และแก้ไขปัญหาทางข้อมูลต่อไป

ขอกราบขอบพระคุณ ท่านสนับสนุนบางส่วนจากทุนเงินรายได้มหาวิทยาลัย ประเภทการ
วิจัยบูรณาการ ประจำปีงบประมาณ 2561 เลขที่สัญญาทุนวิจัย 073/2562 นอกจากนี้การเผยแพร่
งานวิจัยในงานประชุมวิชาการได้รับการสนับสนุนจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิ
โรฒ ทำให้ผู้วิจัยได้ประสบการณ์นำเสนอสารนิพนธ์ และได้แลกเปลี่ยนความรู้ทางด้านเทคโนโลยี
คณิตศาสตร์ คอมพิวเตอร์ และการใช้เทคนิคการเรียนรู้ของ เครื่อง ผู้วิจัยจึงขอขอบคุณมา ณ ที่นี้

แสงดาว เดือนแจ่ม

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ	ช
สารบัญตาราง.....	ญ
สารบัญรูปภาพ	ฎ
บทที่ 1 บทนำ.....	1
1.1 ความสำคัญและความเป็นมาของงานวิจัย.....	1
1.2 วัตถุประสงค์.....	3
1.3 ขอบเขตการวิจัย	3
1.4 ประโยชน์ที่ได้รับจากงานวิจัย	4
บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง.....	5
2.1 ข้อมูลพื้นฐานเกี่ยวกับกาแฟอะราบิก้า	5
2.2 มาตรฐานของเมล็ดสารกาแฟอะราบิก้า.....	5
2.3 การเรียนรู้ของเครื่อง (Machine Learning).....	8
2.3.1 การเรียนรู้ของป่าสุ่ม Random Forest (RF)	8
2.3.2 การถดถอยโลจิสติก Logistic Regression (LR)	9
2.3.3 การประเมินวัดค่าความแม่นยำของแต่ละอัลกอริทึม	10
2.4 งานวิจัยที่เกี่ยวข้อง	11
บทที่ 3 วิธีดำเนินการวิจัย.....	19
3.1 การเตรียมข้อมูลเพื่อใช้ในการทดลอง (Data Preparation)	20

3.2 การประมวลผลภาพเบื้องต้น (Image Preprocessing)	20
3.3.1 การเตรียมข้อมูลรูปภาพ (Image Acquisition).....	20
3.2.2 การประมวลผลภาพ (Image Preprocessing).....	22
3.3 การสกัดคุณลักษณะ (Feature Extraction).....	30
3.3.1 กลุ่มของขนาดและรูปร่าง	31
3.3.1.1 พื้นที่ของเมล็ดกาแฟ (Contour Area)	31
3.3.1.2 ค่าจากค่าอัตราส่วนของพื้นที่รูปร่างต่อพื้นที่เปลือก (Solidity).....	31
3.3.1.3 เส้นผ่านศูนย์กลาง (Equivalent Diameter).....	32
3.3.1.4 ค่าสัดส่วนของพื้นที่วงกลม (Extent)	33
3.3.2 กลุ่มของค่าสี	34
3.3.3 กลุ่มของค่าสีเทา.....	37
3.4 การวิเคราะห์คุณลักษณะ (Feature Analysis).....	38
3.5 การจำแนกประเภทเมล็ดกาแฟ (Coffee Beans Classification).....	72
บทที่ 4 ผลการวิจัย	75
4.1 แบบจำลองและการเปรียบเทียบประสิทธิภาพจากกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data)	75
4.2 แบบจำลองและการเปรียบเทียบประสิทธิภาพจากกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data).....	78
บทที่ 5 อภิปรายผล สรุปผลการวิจัย และข้อเสนอแนะ	81
5.1 การอภิปรายผล	81
5.2 การสรุปผลการวิจัย.....	85
5.2.1. สรุปผลจากกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data)	85
5.2.2. สรุปผลจากกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data).....	85
5.3 ข้อเสนอแนะ.....	86

บรรณานุกรม	88
ประวัติผู้เขียน.....	91



สารบัญตาราง

	หน้า
ตาราง 1 การจำแนกคุณภาพเมล็ดกาแฟจากผู้เชี่ยวชาญและการประยุกต์ใช้เทคโนโลยี	7
ตาราง 2 ตารางสรุปงานวิจัยที่ศึกษาทั้งหมด	17
ตาราง 3 ลักษณะเด่นของคุณภาพเมล็ดกาแฟแต่ละประเภท	22
ตาราง 4 ตารางจำนวนตัวอย่างเมล็ดกาแฟแต่ละประเภท	30
ตาราง 5 การแบ่งกลุ่มตัวอย่างในการฝึกฝนแบบจำลอง	73
ตาราง 6 ตารางแสดงตัวอย่างค่าพารามิเตอร์จากกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data)	76
ตาราง 7 แบบจำลองและการเปรียบเทียบประสิทธิภาพกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data)	76
ตาราง 8 ตารางแสดงตัวอย่างค่าพารามิเตอร์จากกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data)	79
ตาราง 9 ตารางสรุปแบบจำลองและการเปรียบเทียบประสิทธิภาพกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data)	79
ตาราง 10 ตารางสรุปแบบจำลองและการเปรียบเทียบประสิทธิภาพ	86

สารบัญรูปภาพ

	หน้า
ภาพประกอบ 1 กระบวนการแปรรูปภาพ.....	2
ภาพประกอบ 2 การคัดเมล็ดกาแฟเสียออกจากเมล็ดกาแฟดีด้วยแรงงานคน	2
ภาพประกอบ 3 เมล็ดกาแฟดีและเมล็ดกาแฟเสีย	6
ภาพประกอบ 4 ภาพโครงสร้างต้นไม้การตัดสินใจ	9
ภาพประกอบ 5 ภาพโครงสร้างแบบจำลองเชิงเส้น	10
ภาพประกอบ 6 ภาพแบบจำลองของเครื่องคัดแยกเมล็ดกาแฟ.....	19
ภาพประกอบ 7 Flowchart การทำงานของระบบ	20
ภาพประกอบ 8 การเก็บภาพถ่ายและตัวอย่างภาพถ่ายเมล็ดกาแฟ	21
ภาพประกอบ 9 ภาพเมล็ดเสีย a) เมล็ดกาแฟดี b) เมล็ดดำ c) เมล็ดขึ้นรา d) เมล็ดแตก e) เมล็ด ไม่สมบูรณ์ f) เมล็ดถูกแมลงทำลาย	21
ภาพประกอบ 10 รูปภาพต้นฉบับ	25
ภาพประกอบ 11 การลบสัญญาณรบกวน ด้วยการใช้ Gaussian Blur.....	25
ภาพประกอบ 12 การทำ Threshold เพื่อแยก เมล็ดกับภาพพื้นหลัง	26
ภาพประกอบ 13 การทำ Flood Fill เพื่อเติมเต็มส่วนที่ขาดหายไป	26
ภาพประกอบ 14 การทำ bitwise เพื่อให้สามารถแยกเมล็ดกาแฟออกจากภาพพื้นหลัง.....	26
ภาพประกอบ 15 เมล็ดกาแฟสามารถแยกออกจากภาพพื้นหลังได้อย่างชัดเจน	27
ภาพประกอบ 16 การทำ Dilation เพื่อเติมเส้นที่ขาด	27
ภาพประกอบ 17 ใช้ Contours เพื่อหาเส้นรูปทรงสี่เหลี่ยม	28
ภาพประกอบ 18 การ extract ภาพเมล็ดกาแฟ	28
ภาพประกอบ 19 เมล็ดกาแฟที่มีขอบเงาส่วนเกิน	29
ภาพประกอบ 20 ขั้นตอนการตัดเงาออกจากเมล็ดกาแฟ.....	29

ภาพประกอบ 21 ผลลัพธ์จากการตัดเงาออกจากเมล็ดกาแฟ	29
ภาพประกอบ 22 ภาพการหาค่าจากค่าพื้นที่ของเมล็ดกาแฟ (Contour Area).....	31
ภาพประกอบ 23 ภาพฮิสโตแกรมการหาค่าจากค่าอัตราส่วนของพื้นที่รูปร่างต่อพื้นที่เปลือก (Solidity)	32
ภาพประกอบ 24 ภาพการหาค่าเส้นผ่านศูนย์กลาง (Equivalent Diameter)	33
ภาพประกอบ 25 การหาค่าสัดส่วนของพื้นที่วงกลม (Extent)	33
ภาพประกอบ 26 ภาพฮิสโตแกรมการหาค่ามาก (Max) ของสีแดง สีเขียว สีนํ้าเงิน	34
ภาพประกอบ 27 ภาพฮิสโตแกรมการหาค่ามาก (Max) ของค่าสี HSV (Hue Saturation Value).....	35
ภาพประกอบ 28 ภาพฮิสโตแกรมการหาค่าเฉลี่ย (Average) ของสีแดง สีเขียว สีนํ้าเงิน และสีเทา	36
ภาพประกอบ 29 ภาพฮิสโตแกรมการหาค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) ของสีแดง สีเขียว สีนํ้าเงิน และสีเทา.....	36
ภาพประกอบ 30 ภาพฮิสโตแกรมการหาค่าความแปรปรวนของข้อมูล (Variance)	37
ภาพประกอบ 31 ภาพฮิสโตแกรมการหาค่ามาก (Max) ของสีเทา.....	38
ภาพประกอบ 32 ตัวอย่างข้อมูลที่ถูกจัดเก็บในรูปแบบของไฟล์ CSV	38
ภาพประกอบ 33 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ ContourArea ของเมล็ดกาแฟทั้ง 2 ประเภท	40
ภาพประกอบ 34 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ ContourArea ของเมล็ดกาแฟทั้ง 2 ประเภท	40
ภาพประกอบ 35 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ Solidity ของเมล็ดกาแฟทั้ง 2 ประเภท	41
ภาพประกอบ 36 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ Solidity ของเมล็ดกาแฟทั้ง 2 ประเภท	41
ภาพประกอบ 37 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ E_diamator ของเมล็ดกาแฟทั้ง 2 ประเภท	42

ภาพประกอบ 38 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ $E_{diamator}$ ของเมล็ดกาแฟทั้ง 2 ประเภท.....	42
ภาพประกอบ 39 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ $extent$ ของเมล็ดกาแฟทั้ง 2 ประเภท	43
ภาพประกอบ 40 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ $extent$ ของเมล็ดกาแฟทั้ง 2 ประเภท.....	43
ภาพประกอบ 41 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ $bMax$ ของเมล็ดกาแฟทั้ง 2 ประเภท.	44
ภาพประกอบ 42 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ $bMax$ ของเมล็ดกาแฟทั้ง 2 ประเภท.....	44
ภาพประกอบ 43 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ $gMax$ ของเมล็ดกาแฟทั้ง 2 ประเภท.	45
ภาพประกอบ 44 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ $gMax$ ของเมล็ดกาแฟทั้ง 2 ประเภท.....	45
ภาพประกอบ 45 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ $rMax$ ของเมล็ดกาแฟทั้ง 2 ประเภท..	46
ภาพประกอบ 46 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ $rMax$ ของเมล็ดกาแฟทั้ง 2 ประเภท.....	46
ภาพประกอบ 47 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ $bAvg$ ของเมล็ดกาแฟทั้ง 2 ประเภท.	47
ภาพประกอบ 48 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ $bAvg$ ของเมล็ดกาแฟทั้ง 2 ประเภท.....	47
ภาพประกอบ 49 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ $gAvg$ ของเมล็ดกาแฟทั้ง 2 ประเภท.	48
ภาพประกอบ 50 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ $gAvg$ ของเมล็ดกาแฟทั้ง 2 ประเภท.....	48
ภาพประกอบ 51 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ $rAvg$ ของเมล็ดกาแฟทั้ง 2 ประเภท..	49
ภาพประกอบ 52 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ $rAvg$ ของเมล็ดกาแฟทั้ง 2 ประเภท.....	49
ภาพประกอบ 53 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ $bStd$ ของเมล็ดกาแฟทั้ง 2 ประเภท..	50

ภาพประกอบ 86 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ hVar ของเมล็ดกาแฟทั้ง 2 ประเภท	66
ภาพประกอบ 87 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ vVar ของเมล็ดกาแฟทั้ง 2 ประเภท ..	67
ภาพประกอบ 88 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ vVar ของเมล็ดกาแฟทั้ง 2 ประเภท	67
ภาพประกอบ 89 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ grayMax ของเมล็ดกาแฟทั้ง 2 ประเภท	68
ภาพประกอบ 90 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ grayMax ของเมล็ดกาแฟทั้ง 2 ประเภท	68
ภาพประกอบ 91 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ grayAvg ของเมล็ดกาแฟทั้ง 2 ประเภท	69
ภาพประกอบ 92 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ grayAvg ของเมล็ดกาแฟทั้ง 2 ประเภท	69
ภาพประกอบ 93 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ grayStd ของเมล็ดกาแฟทั้ง 2 ประเภท	70
ภาพประกอบ 94 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ grayStd ของเมล็ดกาแฟทั้ง 2 ประเภท	70
ภาพประกอบ 95 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ grayVar ของเมล็ดกาแฟทั้ง 2 ประเภท	71
ภาพประกอบ 96 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ grayVar ของเมล็ดกาแฟทั้ง 2 ประเภท	71
ภาพประกอบ 97 จำนวนกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data)	72
ภาพประกอบ 98 จำนวนกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data)	73
ภาพประกอบ 99 ภาพกราฟแบบจำลองและการเปรียบเทียบประสิทธิภาพ กลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data)	77

ภาพประกอบ 100 ภาพแบบจำลองและการเปรียบเทียบประสิทธิภาพ กลุ่มตัวอย่างชุดข้อมูลไม่สมดุล(Imbalanced Data)	80
ภาพประกอบ 101 Confusion Matrix ของกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data) (a) Confusion Matrix ของแบบจำลอง Random Forest (RF) (b) Confusion Matrix ของแบบจำลอง Logistic Regression (LR).....	82
ภาพประกอบ 102 Confusion Matrix ของกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data) (a) Confusion Matrix ของแบบจำลอง Random Forest (RF) (b) Confusion Matrix ของแบบจำลอง Logistic Regression (LR).....	83
ภาพประกอบ 103 การทำนายผลของเมล็ดกาแฟ ภาพ (a) และ (b) เมล็ดกาแฟดี คำตอบคือถูกต้อง และทำนายผลถูกต้อง ภาพ (c) และ (d) เมล็ดกาแฟไม่สมบูรณ์ คำตอบคือผิด แต่ทำนายผลถูกต้อง ภาพ (e) และ (f) เมล็ดถูกแมลงทำลาย คำตอบคือผิด แต่ทำนายผลถูกต้อง	84



บทที่ 1

บทนำ

1.1 ความสำคัญและความเป็นมาของงานวิจัย

การสกัดและคัดเลือกคุณลักษณะของเมล็ดกาแฟเพื่อจำแนกประเภทตามคุณภาพ ด้วยการประมวลผลภาพถ่ายมีวัตถุประสงค์เพื่อศึกษาวิธีการแยกเมล็ดกาแฟด้วยการประมวลผลจากภาพถ่าย เพื่อเป็นต้นแบบและทดสอบความสามารถในการแยกประเภทเมล็ดกาแฟเสีย ออกจากเมล็ดกาแฟดี กลุ่มตัวอย่างที่ใช้ในการวิจัย คือ ภาพถ่ายเมล็ดกาแฟ จากการแบ่งประเภท โดยอ้างอิงจากประกาศกระทรวงเกษตรและสหกรณ์ เรื่อง กำหนดมาตรฐานสินค้าเกษตร : เมล็ดกาแฟอาราบิก้าตามพระราชบัญญัติมาตรฐานสินค้าเกษตร พ.ศ. 2551 (Ministry of Agriculture and Cooperatives) (ราชกิจจานุเบกษา, 2561)

กระบวนการผลิตกาแฟจะเริ่มด้วยการสีเอาส่วนของกะลาแห้งที่หุ้มเมล็ดออกด้วยเครื่องสีกะลา จากนั้นก็จะถูกส่งต่อไปเพื่อคัดเมล็ดกาแฟเสียออกจากเมล็ดกาแฟดีด้วยเครื่องคัดแยกคุณภาพสูง (กาแฟดอยช้าง, 2018) และมีการคัดซ้ำ ตรวจสอบ ด้วยแรงงานคน กระบวนการคัดแยกสารเมล็ดกาแฟด้วยแรงงานคนนี้อาจมีความผิดพลาดของการคัดเมล็ดกาแฟเสียออกจากเมล็ดกาแฟดีได้และใช้ระยะเวลานาน เนื่องจากกระบวนการนี้อาศัยประสบการณ์ของแรงงานคน ในปัจจุบันมีความต้องการสารเมล็ดกาแฟพันธุ์อาราบิก้า อยู่ที่ 300 กิโลกรัมต่อวัน (กรมวิชาการเกษตร, 2561) คัดเมล็ดกาแฟเสียออกจากเมล็ดกาแฟดีโดยอาศัยความสามารถในการคัดแยกโดยการใช้แรงงานคน จึงไม่เพียงพอต่อความต้องการของลูกค้าในปัจจุบัน ซึ่งเมล็ดกาแฟเสียมีหลายลักษณะ เช่น (1) เมล็ดดำ (Black Bean) เป็นเมล็ดกาแฟที่มีสีดำทั้งภายในและภายนอกมากกว่าครึ่งของเมล็ด (2) เมล็ดขึ้นรา (Moldy Bean) เป็นเมล็ดกาแฟที่มีลักษณะของราหรือเชื้อรา จนทำให้สีหรือลักษณะของเมล็ดผิดปกติ (3) เมล็ดแตก (Broken Bean) เป็นเมล็ดกาแฟที่เกิดการแตกหักของเมล็ด ทำให้เมล็ดกาแฟมีการแตกหรือหักออกเป็น 2 ชิ้นหรือมากกว่าครึ่งหนึ่งของเมล็ดกาแฟ (4) เมล็ดไม่สมบูรณ์ เป็นเมล็ดกาแฟที่มีลักษณะของเมล็ดที่ผิดปกติไปจากเมล็ดปกติ เช่น เหี่ยวยุบ ลีบ เบา (5) เมล็ดถูกแมลงทำลาย เป็นเมล็ดกาแฟที่มีลักษณะของการถูกแมลงเข้าทำลายภายในเมล็ด จากการกัดแทะ หรือเจาะจนเกิดเป็นรู เป็นต้น ดังนั้นจึงมีความจำเป็นในการนำเทคโนโลยีมาช่วยเพิ่มประสิทธิภาพและปริมาณในการคัดแยกคุณภาพเมล็ดกาแฟ



ภาพประกอบ 1 กระบวนการแปรรูปกาแฟ

ที่มา : (กาแฟดอยช้าง, 2018)



ภาพประกอบ 2 การคัดเมล็ดกาแฟเสียออกจากเมล็ดกาแฟดีด้วยแรงงานคน

งานวิจัยนี้นำเสนอการสกัดคุณลักษณะของเมล็ดกาแฟ เพื่อจำแนกประเภทตามคุณภาพ ด้วยการประมวลผลภาพถ่ายและการเรียนรู้ของเครื่อง เพื่อคุณภาพของเมล็ดกาแฟดีและเมล็ดกาแฟเสียได้ ข้อดีของงานวิจัยนี้ คือ ไม่มีการทำลายเมล็ดกาแฟ และสามารถจำแนกเมล็ดเสีย เนื่องจากใช้วิธีการสกัดคุณลักษณะจากการประมวลผลภาพถ่ายของเมล็ดกาแฟโดยตรง ตัวอย่างที่ใช้ในการศึกษามีจำนวนเมล็ดดีและเมล็ดเสีย โดยผ่านกระบวนการ 4 ขั้นตอน คือ 1) การนำเข้าข้อมูลรูปภาพ 2) การประมวลผลภาพเบื้องต้น 3) การสกัดคุณลักษณะ และ 4) การจำแนกเมล็ดกาแฟและเปรียบเทียบประสิทธิภาพแบบจำลอง สำหรับการประมวลผลภาพเบื้องต้นประกอบด้วย (1) การแยกเมล็ดกาแฟจากภาพพื้นหลัง (2) การเติมเต็มส่วนขาดหาย (3) การตรวจจับขอบของเมล็ดกาแฟ และสำหรับการ

สกัดคุณลักษณะ ได้ทำการสกัดคุณลักษณะจำนวน 3 กลุ่มคุณลักษณะ โดยมีคุณลักษณะทั้งหมด 32 คุณลักษณะประกอบด้วย (1) กลุ่มของขนาดและรูปร่าง จำนวน 4 คุณลักษณะ (2) กลุ่มของค่าสี จำนวน 24 คุณลักษณะ (3) กลุ่มของค่าสีเทา จำนวน 4 คุณลักษณะ

1.2 วัตถุประสงค์

1. เพื่อสร้างกระบวนการการได้มาซึ่งภาพถ่ายเพื่อให้สามารถดึงคุณลักษณะของเมล็ดกาแฟจากภาพถ่ายได้
2. เพื่อพัฒนาต้นแบบแบบจำลองที่สามารถจำแนกประเภทเมล็ดกาแฟได้
3. เพื่อให้สามารถจำแนกคุณภาพของเมล็ดกาแฟเสียออกจากเมล็ดกาแฟดี จากนิยามข้อบกพร่องต่อไปนี้ คือ เมล็ดดำ เมล็ดขึ้นรา เมล็ดไม่สมบูรณ์ เมล็ดแตก และเมล็ดถูกแมลงทำลายได้

1.3 ขอบเขตการวิจัย

ขอบเขตงานวิจัยนี้เป็นการนำเอากลุ่มตัวอย่างที่ใช้ในการวิจัย คือ ภาพถ่ายเมล็ดกาแฟ จากการแบ่งประเภท โดยอ้างอิงจากประกาศกระทรวงเกษตรและสหกรณ์ เรื่อง กำหนดมาตรฐานสินค้าเกษตร : เมล็ดกาแฟอะราบิก้าตามพระราชบัญญัติมาตรฐานสินค้าเกษตร พ.ศ. 2551 จำนวน 5 นิยามความบกพร่อง จาก 9 ข้อนิยามความบกพร่องคือ เมล็ดดำ เมล็ดขึ้นรา เมล็ดไม่สมบูรณ์ เมล็ดแตก และเมล็ดถูกแมลงทำลายเท่านั้น มีการควบคุมแสงขณะถ่ายภาพ มีการควบคุมจำนวนเมล็ดในขณะถ่ายภาพ มีการจัดวางเมล็ดขณะถ่ายภาพ ไม่ให้มีการทับซ้อนกันของเมล็ดกาแฟ มีการควบคุมระยะของการถ่ายภาพระหว่างเมล็ดกาแฟกับตัวกล้อง

แบบจำลองการจำแนกประเภทของเมล็ดกาแฟ ที่ใช้สำหรับเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) คือ

1. Random Forest (RF)
2. Logistic Regression (LR)

ทำการเปรียบเทียบประสิทธิภาพของแต่ละแบบจำลองที่นำมาใช้ในการทำนายการจำแนกประเภทของเมล็ดกาแฟด้วย

1. Confusion Matrix
2. Accuracy
3. Precision
4. Recall
5. F1-Score

1.4 ประโยชน์ที่ได้รับจากงานวิจัย

1. ทำให้ลดขั้นตอนการทำงานซ้ำซ้อนของแรงงานคน และสามารถลดต้นทุนในขั้นตอนการคัดแยกนี้ได้
2. ทำให้ผู้ผลิตมีเครื่องมือในการคัดแยกกาแฟ สามารถนำผลลัพธ์ที่ได้ไปเป็นต้นแบบในการประยุกต์ใช้ ในกระบวนการคัดแยกแบบอัตโนมัติต่อไป
3. สามารถนำเอาวิธีการนี้ไปปรับใช้กับระบบคัดแยกแบบอัตโนมัติ (Sorting Machine System) ประเภทอื่นๆ ได้



บทที่ 2

วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

2.1 ข้อมูลพื้นฐานเกี่ยวกับกาแฟอาราบิก้า

กาแฟอาราบิก้าเป็นต้นไม้ยืนต้นขนาดกลางที่มีความสำคัญทางเศรษฐกิจของโลก ปลูกกันแพร่หลายในเชิงธุรกิจ กาแฟอาราบิก้า ซึ่งมีชื่อวิทยาศาสตร์ว่า *Coffea Arabica* L. กาแฟอาราบิก้านี้ชอบปลักระงูมีประเทศที่เย็นชื้น จึงทำให้นิยมปลูกกันอย่างแพร่หลายในภาคเหนือ กาแฟอาราบิก้าที่ผ่านกระบวนการกะเทาะกะลาแล้วนั้น จะมีรูปร่างลักษณะของเมล็ดเป็นวงรี กลมเรียว มีเส้นตรงกลางคดงอเล็กน้อย ส่วนรสชาตินั้นจะออกขมหวานละมุน เพราะมีน้ำตาลมากจะทำให้ออกรสชาติหวาน และยังมีความเป็นกรดสูงจึงทำให้มีรสเปรี้ยวซ่อนอีกด้วย ด้วยความที่มีทั้งรสหวานและมีความเป็นกรดสูงนี้ ทำให้เมล็ดกาแฟมักถูกรับกวนด้วยแมลง ผลผลิตกาแฟอาราบิก้าจึงมีราคาแพง ผลผลิตจึงเป็นที่ต้องการของตลาด (rabikacoffee, 2561; พิชัยอุตกฤษฎ์ และคนอื่น ๆ, 2556)

2.2 มาตรฐานของเมล็ดสารกาแฟอาราบิก้า

เมล็ดกาแฟจะมีการแยกเกรดของกาแฟสารออกเป็น 4 เกรดตามขนาดและเมล็ดกาแฟสารจะต้องมีความชื้นอยู่ที่ 11.5-12% และสามารถแยกสารกาแฟตามนิยามข้อบกพร่องของเมล็ดกาแฟเสีย ตามประกาศกระทรวงเกษตรและสหกรณ์ เรื่อง กำหนดมาตรฐานสินค้าเกษตร : เมล็ดกาแฟอาราบิก้า (ราชกิจจานุเบกษา, 2561)

ตามพระราชบัญญัติมาตรฐานสินค้าเกษตร พ.ศ. 2551 แยกเป็น 9 นิยามความบกพร่องดังนี้

- (1) เมล็ดดำ (Black Bean) หมายถึง เมล็ดกาแฟที่มีสีดำภายใน และ/หรือ ภายนอกเมล็ดมากกว่าครึ่งหนึ่งของเมล็ด
- (2) เมล็ดที่มีกลิ่นผิดปกติ (Off-Odour Bean) หมายถึง เมล็ดกาแฟที่มีกลิ่นเหม็นเปรี้ยว กลิ่นหมักบูด หรือกลิ่นแปลกปลอม เช่น กลิ่นปุย กลิ่นสารเคมี หรือ กลิ่นรา
- (3) เมล็ดขึ้นรา (Moldy Bean) หมายถึง เมล็ดกาแฟ ที่มีลักษณะการเข้าทำลายของเชื้อรา
- (4) เมล็ดแตก (Broken Bean) หมายถึง เมล็ดกาแฟที่แตกออกเป็นชิ้นเล็กกว่าครึ่งหนึ่งของกาแฟเต็มเมล็ด
- (5) เมล็ดไม่สมบูรณ์ หมายถึง เมล็ดกาแฟที่มีลักษณะเหยย่น ลีบ เบา
- (6) เมล็ดถูกแมลงทำลาย หมายถึง เมล็ดกาแฟที่ถูกแมลงกัดแทะ หรือเจาะจนเกิดเป็นรู

(7) ผลกาแฟแห้ง (Dried Cherry) หมายถึง ผลกาแฟที่ผ่านกรรมวิธีการทำให้แห้งจนได้ผลกาแฟแห้งที่ยังไม่ได้สีเปลือกออก รวมทั้งเมล็ดกาแฟที่มีเปลือกติดบางส่วน

(8) สิ่งแปลกปลอม (Foreign Matter) หมายถึง สิ่งเจือปนทางกายภาพอื่น ๆ ที่ไม่ใช่เมล็ดกาแฟ เช่น เศษหิน ดิน เศษไม้ รวมทั้ง เปลือกและเยื่อหุ้มเมล็ดกาแฟ

(9) ข้อบกพร่องรวม หมายถึง ปริมาณโดยน้ำหนักของข้อบกพร่องดังต่อไปนี้รวมกัน เมล็ดดำ เมล็ดขึ้นรา เมล็ดแตก เมล็ดไม่สมบูรณ์ เมล็ดถูกแมลงทำลาย ผลกาแฟแห้ง และสิ่งแปลกปลอม



เมล็ดกาแฟสีเขียว (Green Beans)



เมล็ดกาแฟสีดำ (Black beans)

ภาพประกอบ 3 เมล็ดกาแฟดีและเมล็ดกาแฟเสีย

งานวิจัยนี้เลือกใช้เมล็ดกาแฟจำนวน 5 นิยามความบกพร่อง จาก 9 ข้อนิยามความบกพร่องข้างต้น ได้แก่ 1.) เมล็ดกาแฟสีดำ 2.) เมล็ดกาแฟขึ้นรา 3.) เมล็ดกาแฟแตก 4.) เมล็ดกาแฟถูกแมลงทำลาย และ 5.) เมล็ดกาแฟไม่สมบูรณ์เท่านั้น ดังตาราง 1 ส่วนอีก 4 นิยามข้อบกพร่องซึ่งได้แก่ 1.) เมล็ดที่มีกลิ่นผิดปกติ 2.) ผลกาแฟแห้ง 3.) สิ่งแปลกปลอม และ 4.) ข้อบกพร่องรวม นั้นไม่นับรวมว่าเป็นเมล็ดกาแฟจึงไม่เลือกนำมาใช้ในงานวิจัยนี้

ตาราง 1 การจำแนกคุณภาพเมล็ดกาแฟจากผู้เชี่ยวชาญและการประยุกต์ใช้เทคโนโลยี

คุณภาพเมล็ดกาแฟ	ลักษณะคุณภาพเมล็ดกาแฟ จากการสังเกตของผู้เชี่ยวชาญ	เทคโนโลยีที่ใช้ ในการจำแนกเมล็ดกาแฟ
1. เมล็ดกาแฟดี	มีลักษณะเป็นเมล็ดที่สมบูรณ์ ทรงกลมรี สีออกเขียวอ่อน	กล้องถ่ายภาพ
2. เมล็ดกาแฟสีดํา	มีสีน้ำตาลอ่อนจนถึงสีดำสนิท และมีลักษณะของสีที่ไม่สม่ำเสมอบน เมล็ดกาแฟ	กล้องถ่ายภาพ
3. เมล็ดกาแฟขึ้นรา	มีสีที่ไม่สม่ำเสมอ มีสีเขียวบนน้ำตาล	กล้องถ่ายภาพ
4. เมล็ดกาแฟแตก	เมล็ดกาแฟที่เกิดการแตกหักของเมล็ด หรือ หักออกเป็น 2 ชิ้น หรือมากกว่าครึ่งหนึ่งของ เมล็ดกาแฟ	กล้องถ่ายภาพ
5. เมล็ดกาแฟถูกแมลงทำลาย	เมล็ดกาแฟที่มีลักษณะของการถูกแมลง จากการกัดแทะ หรือเจาะจนเกิดเป็นรู	กล้องถ่ายภาพ
6. เมล็ดกาแฟไม่สมบูรณ์	เมล็ดเหี่ยวแห้ง เมล็ดลีบ สีซีดจาง	กล้องถ่ายภาพ
7. เมล็ดที่มีกลิ่นผิดปกติ	มีกลิ่นเหม็นเปรี้ยว กลิ่นหมักบูด หรือกลิ่น แปลกปลอม เช่น กลิ่นปุ๋ย กลิ่นสารเคมี หรือ กลิ่นรา	ใช้เซ็นเซอร์ในการจับกลิ่น
8. ผลกาแฟแห้ง	ผลกาแฟแห้งทั้งที่ยังมีเปลือกหุ้มอยู่	เครื่องสีกาแฟ เครื่องร่อนรูกกลม
9. สิ่งแปลกปลอม	เศษหิน ดิน เศษไม้ รวมทั้ง เปลือกและเยื่อ หุ้มเมล็ดกาแฟ	เครื่องร่อนรูกกลม และกระบวนการหมักกาแฟ ในการคัดแยก (เทลงบ่อน้ำให้ ลอยแล้วตักทิ้ง)
10. ข้อบกพร่องรวม	เมล็ดกาแฟเสียอื่นๆ โดยรวม	กระบวนการหมักกาแฟในการ คัดแยก (เทลงบ่อน้ำให้ลอยแล้ว ตักทิ้ง)

2.3 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่อง (Machine Learning) เป็นสาขาหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence) ที่เกี่ยวข้องกับการออกแบบและพัฒนาเทคนิควิธีเพื่อให้คอมพิวเตอร์สามารถพัฒนาพฤติกรรมจากชุดข้อมูลที่ได้รับโดยเน้นไปที่การพัฒนาอัลกอริทึม (Algorithm) ที่ทำให้คอมพิวเตอร์สามารถเรียนรู้ผ่านการสังเกตและวิเคราะห์ชุดข้อมูลเพื่อนำไปหาค่าความสัมพันธ์ทางสถิติและสร้างเป็นกรดสำหรับคุณสมบัติที่หายไปหรือข้อมูลในอนาคตกระบวนการสำคัญของการเรียนรู้ (Dondegroovily, 2012) คือการจำแนกข้อมูล (Classification) หรือการจดจำรูปแบบ (Pattern Recognition) เพื่อให้คอมพิวเตอร์สามารถเรียนรู้และจดจำรูปแบบที่ซับซ้อนได้โดยอัตโนมัติ รวมทั้งสามารถแยกแยะรูปแบบที่แตกต่างกันและคาดการณ์ข้อมูลเหล่านั้นได้อย่างชาญฉลาด (O.Duda, E.Hart, และ Stork., 2000)

การทำงานวิจัยนี้ผู้วิจัยมีความสนใจที่จะศึกษาอัลกอริทึมการเรียนรู้ในกลุ่มของการเรียนรู้แบบมีผู้สอนโดยเลือกอัลกอริทึมการเรียนรู้ 2 วิธี คือ

- Random Forest (RF)
- Logistic Regression (LR)

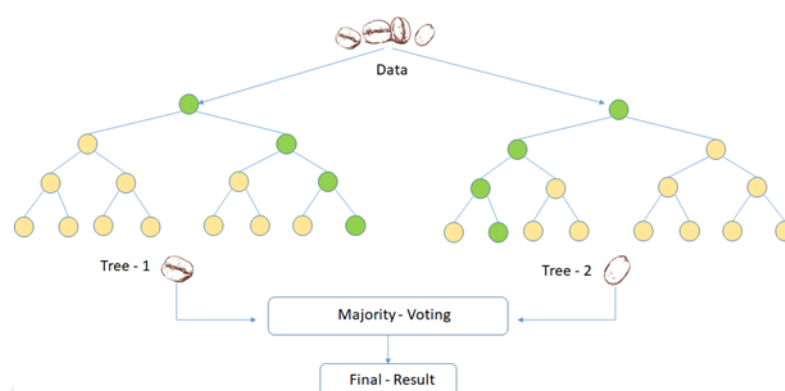
2.3.1 การเรียนรู้ของป่าสุ่ม Random Forest (RF)

Random Forest (RF) เป็นเทคนิคที่ถูกพัฒนาต่อยอดมาจากเทคนิค Bagged Decision Tree โดยมีหลักการทำงานคือ โมเดลจะทำการสร้าง Decision Tree ขึ้นมาหลายๆต้น หรือที่เรียกว่า Ensemble of Decision Trees และในแต่ละ Sub-Trees นั้นเป็นอิสระต่อการ (Less Correlation) อีกทั้งในแต่ละ Decision Tree ยังใช้การสุ่มเลือก Feature ที่ดีที่สุดของการ Split ในแต่ละโหนด จึงทำให้ต้นไม้แต่ละต้นนั้นมีลักษณะที่แตกต่างกันออกไป ดังนั้นยังมีต้นไม้ไม่มากเท่าไร ผลลัพธ์ที่ได้จากการทำนายก็จะยิ่งดีขึ้น และในงานวิจัยนี้ได้ใช้ Random Forest (RF) Classifier เป็นโมเดลที่ใช้ในการ Train ชุดข้อมูล ดังภาพประกอบ 4

ตัวอย่างพารามิเตอร์ที่สามารถปรับจูนได้ของ Random Forest (RF) :

- n_estimators คือ จำนวนต้นไม้ทั้งหมด มีค่า (default = 10) ซึ่งถ้ายิ่งกำหนดให้มียามากเท่าไร ผลลัพธ์ที่ได้จากการทำนายก็จะยิ่งแม่นยำ ข้อเสียคือใช้เวลาในการคำนวณค่อนข้างนาน
- max_features คือ จำนวนของ feature ที่เราจะพิจารณาเลือกในการ split (ในแต่ละ step ของ split นั้นจะทำการสุ่มเลือก subset ของ feature)
- max_depth คือ ค่าที่กำหนดว่าจะให้ tree นั้นมีค่า depth สูงสุดเท่าไร ถ้ายิ่งปรับค่านี้นี้ให้มากขึ้น tree ก็จะมีขนาดซับซ้อนมากขึ้น

- `min_samples_split` คือ จำนวน samples ที่ต้องมีที่โหนดหนึ่งๆ (มีอย่างน้อย) ถึงจะยอมให้ทำการแตกหรือ split โหนดต่อไปได้ ถ้าปรับค่านี้ให้ค่าน้อยลง tree ก็ซับซ้อนมากขึ้น
- `min_samples_leaf` คือ จำนวน samples ที่ต้องมีเป็นอย่างน้อยที่ leaf node หนึ่งๆ



ภาพประกอบ 4 ภาพโครงสร้างต้นไม้การตัดสินใจ

2.3.2 การถดถอยโลจิสติก Logistic Regression (LR)

การวิเคราะห์การถดถอยโลจิสติกเป็นการหาความสัมพันธ์ระหว่างตัวแปรตาม (1 ตัวแปร) และตัวแปรอิสระ (มากกว่าหรือเท่ากับ 1 ตัวแปร) โดยที่ตัวแปรตาม (Y) มีลักษณะเป็นตัวแปรเชิงกลุ่ม เมื่อได้แบบแผนความสัมพันธ์ (สมการพยากรณ์) แล้วสามารถนำแบบแผนดังกล่าวไปใช้ประมาณค่ากลุ่มได้ โดยอาศัยความน่าจะเป็นของการเกิดเหตุการณ์ การวิเคราะห์การถดถอยโลจิสติกแบ่งออกเป็น 2 ประเภทคือ

1. Binary Logistic Regression (LR) ตัวแปรตามเป็นตัวแปรเชิงกลุ่มที่มีค่าได้เพียง 2 ค่า (ตัวแปรทวิ) เช่น $Y = 1$ ถ้าคนไข้มีการติดเชื้อ หรือ $Y = 0$ ถ้าคนไข้ ไม่มีการติดเชื้อ
2. Multinomial Logistic Regression (LR) ตัวแปรตามเป็นตัวแปรเชิงกลุ่มที่มีค่ามากกว่า 2 ค่า เช่น $Y = 0$ หมายถึงไม่เป็นโรคมะเร็ง $Y = 1$ หมายถึง การเป็นมะเร็งขั้นต้น ... $Y = 4$ หมายถึงการเป็นมะเร็งขั้นสุดท้าย

รูปแบบสมการ การวิเคราะห์การถดถอยโลจิสติกสำหรับการวิเคราะห์การถดถอยโลจิสติกสมการพยากรณ์ที่ได้จากตัวแบบการวิเคราะห์จะเป็นสมการแสดงความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ (Probability of Event) ดังสมการที่ (1)-(3) และภาพประกอบ 5

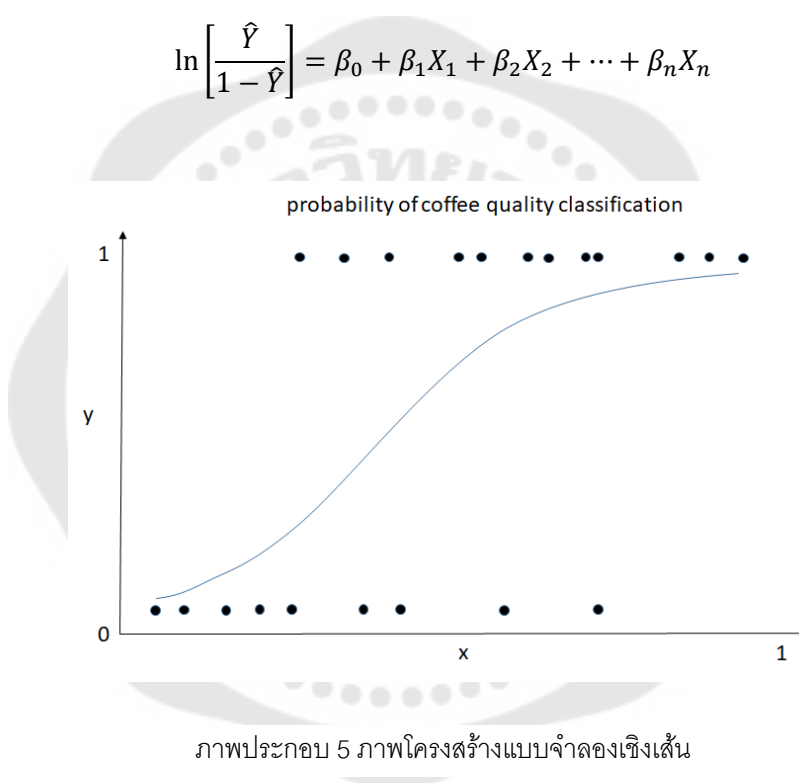
$$\hat{Y}_i = \frac{e^u}{1 + e^u} \quad (1)$$

เมื่อ $\hat{Y}$ เป็นค่าความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ

$$u = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (2)$$

และสามารถทำให้อยู่ในรูปเชิงเส้น (Linear model)

$$\ln \left[\frac{\hat{Y}}{1 - \hat{Y}} \right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (3)$$



2.3.3 การประเมินวัดค่าความแม่นยำของแต่ละอัลกอริทึม

การทำนายแบบ Classification ในงานวิจัยนี้ใช้ค่า Confusion Matrix , ค่าความถูกต้องในการทำนาย (Accuracy), Precision, Recall และ ค่า F1-Score โดยค่า Accuracy แสดงได้ดังสมการที่ (4)

$$Accuracy = \frac{N_{correct}}{N} \quad (4)$$

โดยที่ $N_{correct}$ คือจำนวน sample ใน Test Set ที่ทำนายถูกต้อง และ N คือจำนวน Sample ทั้งหมดใน Test Set

ส่วนค่า Precision, Recall และ ค่า F1-Score นั้น สามารถแสดงได้ดังสมการที่ (5), สมการที่ (6) และสมการที่ (7) ตามลำดับ

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

โดยที่ True Positive (TP) คือ จำนวนของ Sample ที่ทำนายถูกต้อง
 False Positive (FP) คือ จำนวนของ Sample ที่ทำนายผลเป็น Positive แต่คำตอบที่ถูกต้องเป็น Negative
 False Negative (FN) คือ จำนวนของ Sample ที่ทำนายผลเป็น Negative แต่คำตอบถูกต้องเป็น Positive

2.4 งานวิจัยที่เกี่ยวข้อง

การทบทวนวรรณกรรมของงานวิจัยนี้ ได้ทำการศึกษางานวิจัยที่เกี่ยวข้องกับการประมวลผลภาพถ่ายและเมล็ดกาแฟ โดยเน้นในเรื่องของการแบ่งแยกประเภทด้วยเทคนิคต่าง ๆ มีรายละเอียดดังต่อไปนี้

1.) Rafael Garretz, Claudia M. Rezende, Demian R. Ifa (2016) ทำการศึกษาการเปิดเผยการกระจายตัวของกรดคลอโรจีนิกและน้ำตาลซูโครสผ่านเอนโดสเปิร์มของเมล็ดกาแฟด้วยการถ่ายภาพแบบ Desorption Electrospray Ionization-Mass Spectrometry (DESI-MS) มีการใช้ตัวอย่างเมล็ดกาแฟสีเขียวพันธุ์อะราบิก้า ที่ผ่านการตากแห้งแล้วประมาณหนึ่งมาใช้ในการวิจัย เมล็ดกาแฟถูกตัดในแนวขวางในชั้นที่ความหนา 0.5-0.8 มม. ก่อนนำไปวางบนสไลด์ด้วยกล้องจุลทรรศน์ หรือกระดาษพิมพ์อิงค์เจ็ทขนาดเล็ก และทำการทดลองการถ่ายภาพด้วย DESI-MS โดยใช้โหมดไอออนลบ พบว่าการกระจายตัวของไอออนในเนื้อเยื่อเอนโดสเปิร์มของตัวอย่างนั้น ไม่ได้เป็นเนื้อเดียวกัน นอกจากนี้ยังพบว่ารูที่เกิดจากแมลงเจาะผลกาแฟได้รับการวิเคราะห์นั้น

เป็นแมลงเล็ก ๆ ที่เข้าไปอาศัยและวางไข่อยู่ในเมล็ดกาแฟ ซึ่งการควบคุมศัตรูพืชนี้ทำได้ยากมาก และในโหมดไอออนบวกของการถ่ายภาพแบบ DESI-MS สามารถมองเห็นการกระจายตัวของไอออนซูโครสทั่วทั้งเอนโดสเปิร์มของเมล็ดกาแฟ การถ่ายภาพแบบ DESI-MS เป็นเทคนิคการสร้างไอออนไนโตรเจนรอบตัวที่มีประสิทธิภาพซึ่งสามารถสร้างภาพไอออนของพื้นผิวตัวอย่างที่มีความละเอียดเชิงพื้นที่สูง วิธีการที่ใช้ประสบความสำเร็จในการกระจายกรด Chlorogenic ที่ไม่เป็นเนื้อเดียวกันภายใน Endosperm ของเมล็ดกาแฟ ด้วยปัจจัยเหล่านี้เหล่านี้อาจมีผลโดยตรงต่อการปฏิสัมพันธ์ระหว่างเมล็ดกาแฟกับแมลง (Garrett, Rezende, และ Ifa, 2016) งานวิจัยนี้มีข้อเสีย คือ การเก็บตัวอย่างในอุณหภูมิต่ำ และการตัดเมล็ดกาแฟแบบขวาง ทำให้เมล็ดกาแฟเกิดความเสียหาย และไม่สามารถนำเมล็ดกาแฟในขบวนการนี้มาศึกษาคุณภาพเมล็ดกาแฟต่อได้

2.) Rosalba Calvini, Alessandro Ulrici, Jose Manuel Amigo (2015) ทำการศึกษาการเปรียบเทียบวิธีปฏิบัติในการจำแนกประเภทของกาแฟอาราบิก้ากับกาแฟโรบัสต้า โดยใช้การถ่ายภาพด้วยแสงที่มีความยาวคลื่นย่านใกล้อินฟราเรด โดยการใช้ ตัวอย่างเมล็ดกาแฟโรบัสต้า 18 ตัวอย่าง และตัวอย่างเมล็ดกาแฟอาราบิก้า 15 ตัวอย่าง ตัวอย่างแต่ละตัวอย่างประกอบด้วยเมล็ดกาแฟประมาณ 500 กรัม จากบริษัทคั่วกาแฟในท้องถิ่น เมล็ดกาแฟจะถูกจัดเก็บไว้ในบรรจุภัณฑ์ที่ปิดสนิทจนถึงวันที่ทำการวิเคราะห์ จากแต่ละตัวอย่าง จะแบ่งตัวอย่าง 70 กรัม ออกมา 3 ส่วน โดยทำการเลือกแบบสุ่มและทำการถ่ายภาพ Hyperspectral (HSI) จำนวนสองภาพ จากนั้นนำไปหาค่าข้อมูลทีประกอบไปด้วย 396 ภาพ Hyperspectral (33 ตัวอย่าง $\times$ 12 ภาพ) มาทำการวิเคราะห์ ในงานวิจัยนี้ได้มีการสร้างแบบจำลอง sPCA และ sPLS-DA ที่แตกต่างกันในชุดศึกษา โดยพิจารณาการผสมผสานระหว่าง Sparsity (เรียกว่า sparsity constraint สำหรับ sPCA + kNN และจำนวนตัวแปรที่เลือกในแต่ละส่วนของ sPLS-DA) การจำแนกประเภทกาแฟมีความถูกต้องทั้ง PCA + kNN และ sPCA + kNN ในขณะที่ PLS-DA และ sPLS-DA มีการจัดประเภทที่ไม่ถูกต้อง เนื่องจากภาพที่ทำจากสเปกตรัมพิกเซลจำนวนมากจึงมีความแปรปรวนมากกว่าสเปกตรัมภาพโดยเฉลี่ยและอาจมีการทับซ้อนกันระหว่างชั้นพิกเซล ในเงื่อนไขเหล่านี้ KNN พิสูจน์ให้เห็นว่าเป็นวิธีที่มีประสิทธิภาพมากกว่า PLS-DA (Calvini และคนอื่น ๆ, 2015) งานวิจัยนี้มีข้อเสีย คือ จากภาพที่ทำจากสเปกตรัมพิกเซลจำนวนมากจึงมีความแปรปรวนมากกว่าสเปกตรัมภาพโดยเฉลี่ย และอาจมีการทับซ้อนกันระหว่างชั้นพิกเซล

3.) Yudong Zhang, Shuihua Wang, Genlin Ji, Preetha Phillips (2014) ทำการศึกษาการจำแนกผลไม้โดยใช้คอมพิวเตอร์วิทัศน์และโครงข่ายประสาทเทียมแบบป้อนไปข้างหน้า โดยใช้ตัวอย่างของผลไม้ทั้งหมด จำนวน 18 ชนิด ชุดข้อมูลประกอบไปด้วย 1653 ภาพ จำแนกด้วยการจับ

ภาพสี่และรูปร่าง ใช้อัลกอริทึม Split-And-Merge เพื่อลบพื้นที่พื้นหลัง ทำการตัดออกให้เหลือเฉพาะภาพผลไม้ตรงกลางภาพเท่านั้น และใช้วิธีการสุ่มตัวอย่างแบบสุ่ม การตรวจสอบแบบ K-fold Cross-Time การตรวจสอบแบบ Leave-One-Out และการตรวจสอบข้ามแบบ Monte Carlo ทำการเปรียบเทียบวิธี FSCABC-FNN ของการทดลอง กับ GA-FNN, PSO-FNN, ABC-FNN และ kSVM ซึ่งแสดงให้เห็นถึงความถูกต้องของการจัดประเภท ผลพบว่า SA ไม่พบ Solution ที่น่าพอใจ ตรงกันข้ามอัลกอริทึม FSCABC ทำงานได้ดีที่สุดโดยให้ค่า MSE อย่างน้อยที่สุดเท่ากับ 0.0242 PSO เป็นอัลกอริทึมที่ดีที่สุดอันดับที่สองกับ MSE เป็น 0.0267 ABC เป็นที่สามกับ MSE เป็น 0.0310 GA เป็นรุ่นที่ 4 กับ MSE เท่ากับ 0.0533 (Zhang, Wang, Ji, และ Phillips, 2014) งานวิจัยนี้มีข้อเสีย คือ การใช้ตัวอย่างผลไม้ที่มีขนาดใหญ่และรูปร่างที่แตกต่างกันอย่างชัดเจน จึงไม่สามารถนำมาใช้กับงานวิจัยนี้มีตัวอย่างเป็นเมล็ดกาแฟขนาดเล็ก และรูปร่างที่คล้ายคลึงกันทั้งหมดของเมล็ดกาแฟได้

4.) C. Pinto, J. Furukawa, H. Fukai and S. Tamura (2017) ทำการศึกษาการจำแนกชนิดของเมล็ดกาแฟสีเขียวตามชนิดของข้อบกพร่องโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการ Convolutional Neural Network (CNN) เพื่อประเมินความเป็นไปได้ของข้อบกพร่องของเมล็ดกาแฟสีเขียวจากแต่ละภาพโดยใช้โครงข่ายประสาทเทียมแบบสังวัตนาการเชิงลึก โดยการใช้ตัวอย่างเมล็ดกาแฟสีเขียวจากติมอร์ – เลสเต ใช้เทคนิคการประมวลผลภาพได้นำมาประยุกต์ใช้ในการแยกเมล็ดกาแฟแต่ละชนิด มีประมาณ 13,000 ภาพจากเมล็ดกาแฟ 6,500 เมล็ด แยกประเภทของข้อบกพร่อง ได้แก่ สีเขียวจาง, สีดำ, รสเปรี้ยว, เมล็ดเสียหาย, Peaberry และไม่มีข้อบกพร่อง ภาพทั้งหมดของกลุ่มตัวอย่างถูกแบ่งออกเป็น 3 กลุ่ม ข้อมูลการฝึกอบรม ข้อมูลการตรวจสอบและข้อมูลการทดสอบ ข้อมูลการฝึกอบรมใช้สำหรับการเรียนรู้ของโครงข่ายประสาทเทียม ข้อมูลการตรวจสอบถูกนำมาใช้เพื่อยืนยันการเปลี่ยนแปลงความถูกต้องของการจำแนกประเภทในช่วงการเรียนรู้ของโครงข่ายประสาทเทียม ข้อมูลการทดสอบใช้เพื่อประเมินสมรรถนะของความสามารถในการเรียงลำดับของโครงข่ายประสาทเทียมด้วยพารามิเตอร์ขั้นสุดท้าย ได้ผลลัพธ์ความแม่นยำในการจัดจำแนกสำหรับเมล็ดกาแฟสีดามีค่าสูงสุด (98.75%) และตามด้วยความแม่นยำของเมล็ดกาแฟรสเปรี้ยว (92.93%) เมล็ดเสียหายมีความถูกต้องต่ำสุด (67.50%) ความแม่นยำสูงของเมล็ดกาแฟสีดำและเมล็ดกาแฟรสเปรี้ยวแสดงให้เห็นว่าโมเดล CNN ที่พัฒนาแล้วนี้สามารถแยกแยะคุณลักษณะภาพสำหรับป้ายกำกับดังกล่าวได้อย่างมีประสิทธิภาพ เราเห็นว่าผลลัพธ์ที่ดีสำหรับเมล็ดกาแฟสีดำและเมล็ดกาแฟรสเปรี้ยวเนื่องจากง่ายต่อการระบุสีทั่วไปของคุณลักษณะเมล็ดกาแฟสีดำและเมล็ดกาแฟรสเปรี้ยว เมล็ดสีเขียว และเมล็ดเสียหาย 72.41% และ 72.50 % ตามลำดับ (Pinto, Furukawa, Fukai, และ Tamura, 2017)

งานวิจัยนี้มีข้อเสีย คือ ไม่ได้มีการสกัดคุณลักษณะอย่างเป็นขั้นตอน และใช้เทคนิค CNN ในการจำแนก จึงไม่สามารถนำมาปรับใช้กับงานวิจัยนี้ได้

5.) Di Wu and Da-Wen Sun (2013) การศึกษาการวัดสีโดยคอมพิวเตอร์วิทัศน์เพื่อการควบคุมคุณภาพของอาหาร การวัดสีของอาหารนั้นมีผลต่อการคัดแยกระดับคุณภาพในเชิงพาณิชย์ คอมพิวเตอร์วิทัศน์ได้รับความสนใจมากสำหรับการใช้เป็นวิธีตรวจสอบ โดยปราศจากการสร้างความเสียหายให้กับอาหาร การตรวจวัดสีอาหารและผลิตภัณฑ์อาหาร บทความนี้ครอบคลุมถึงพื้นฐานและการใช้งานทั่วไปของการแสดงผลทางคอมพิวเตอร์ในการวัดสีของอาหาร โดยเทคนิคที่ใช้ในการตรวจสอบอาหารจะใช้พื้นฐานทางวิทยาศาสตร์ จากการแสดงผลทางคอมพิวเตอร์นั้น ได้ถูกพิสูจน์ว่าเป็นเครื่องมือที่น่าเชื่อถือและมีประสิทธิภาพ สำหรับการวัดสีนั้น โดยตัวของมันแล้วด้วยความสามารถที่มีอยู่ไม่สามารถตรวจสอบได้โดยวิธีอื่น ๆ โดยเฉพาะอย่างยิ่งความสามารถในการวิเคราะห์ตัวอย่างของอาหารที่มีลักษณะสีแบบไม่เอกพันธ์ (Non-Homogeneous Color) เข้าใจว่าเป็นลักษณะสีที่มีคุณสมบัติแตกต่างกัน ไม่เป็นเนื้อเดียวกัน) รวมถึงรูปร่างและพื้นผิวอีกด้วย การวัดสีโดยใช้การแสดงผลทางคอมพิวเตอร์สามารถทำการทดลองซ้ำ ๆ และปรับยืดหยุ่นได้ตามความเหมาะสม ในส่วนของโรงงานผลิตอาหารเมื่อนำเทคโนโลยีนี้มาใช้จะทำให้มีความแม่นยำสูง ช่วยลดค่าใช้จ่าย และให้ผู้ตรวจสอบอาหารมุ่งเน้นไปที่ความต้องการของผู้บริโภคและทักษะทางอาชีพ แทนที่ภาระงานที่มีการตรวจสอบงานที่ซ้ำไปซ้ำมา อย่างไรก็ตาม การวัดสีโดยใช้ความสามารถแสดงผลทางคอมพิวเตอร์สามารถประเมินผลคุณสมบัติของอาหารอื่น ๆ เช่น รูปร่าง ขนาด การกำหนดทิศทาง ข้อบกพร่อง และข้อมูลโภชนาการได้ด้วย (Wu และ Sun, 2013) งานวิจัยนี้มีข้อเสีย คือ กลุ่มตัวอย่างที่ใช้ในการศึกษาการวัดสีเพื่อควบคุมคุณภาพนั้นไม่ใช่กลุ่มตัวอย่างของเมล็ดกาแฟ และใช้การวัดสีเป็นตัววัดคุณภาพเพียงอย่างเดียว

6.) P. Frisullo, J. Laverse, M. Barnabà, L. Navarini, M.A. Del Nobile (2012) การศึกษาการเปลี่ยนแปลงโครงสร้างทางจุลภาคของเมล็ดกาแฟที่มีแนวโน้มมาจากกระบวนการเพาะปลูก : โดยการตรวจสอบด้วยการเอกซเรย์คอมพิวเตอร์ขนาดจุลภาค โดยใช้ตัวอย่างเมล็ดกาแฟ 4 ชนิด คือ Coffea Arabica (Arabica), Arabica Monsooned (Arabica Nons), Coffea Canephora (Robusta) และ Robusta Monsooned (Robusta Mons) จากต้นกำเนิดทางภูมิศาสตร์ที่แตกต่างกัน เมล็ดกาแฟแต่ละชนิดจะถูกชั่งน้ำหนักด้วยเครื่องชั่งน้ำหนักระบบดิจิทัลที่มีความแม่นยำ (± 0.1 กรัม) เพื่อดำเนินการความหนาแน่นของตัวอย่างเมล็ดกาแฟ ตัวอย่างเมล็ดกาแฟจะถูกถ่ายภาพและบันทึกการส่งผ่านของลำแสงรังสีเอกซเรย์ ใช้เวลาในการสแกนโดย

เฉลี่ย 70 นาที่ ชุดภาพตัวอย่างตัดขวางแบบแบน จะได้รับหลังจากประกอบภาพเอกซเรย์ใหม่โดยซอฟต์แวร์ NRecon (Skyscan) ภาพตัวอย่างการประกอบภาพเอกซเรย์ 3 มิติจะถูกสร้างขึ้นโดยใช้ภาพเอกซเรย์ 2 มิติ จำนวน 420 ภาพ ที่มีระยะห่างระหว่างพื้นที่ที่ 0.069 มม. มีการวิเคราะห์ข้อมูลก่อนที่จะมีการสร้างภาพเอกซเรย์ 3 มิติ ก่อนถูกนำเข้าสู่ซอฟต์แวร์ CT Vol (Skyscan) สำหรับการประมวลผลภาพ เปอร์เซ็นต์ของปริมาตรของวัตถุ เช่น พารามิเตอร์ทางเรขาคณิต POV ได้รับโดยการวิเคราะห์ μ CT ได้คำนวณ สำหรับแต่ละตัวอย่าง เปอร์เซ็นต์ของรู (Pore) ของเมล็ดทั้งหมด แสดงกราฟของค่าเปอร์เซ็นต์ POV ของรู ก่อนและหลังจากการคั่วตามลำดับ กาแฟ Arabica mons มีปริมาณของรูมากที่สุดก่อนที่จะได้รับการคั่ว ใน Robusta Mons มีปริมาณของรูในเปอร์เซ็นต์ต่ำที่สุด กาแฟสองสายพันธุ์หลักคือ Arabica และ Robusta ไม่มีความแตกต่างทางสถิติ กระบวนการของเมล็ดกาแฟ Arabica Mons และ Robusta Mons มีปริมาณรู สูงกว่า Arabica และ Robusta เนื่องจากกาแฟมีความชื้นสูงกว่าจาก 13-14% เมื่อเทียบกับกาแฟปกติ 10-12% หลังจากคั่วแล้วจะสังเกตได้ว่าการเพิ่มขึ้นของปริมาณรูสำหรับตัวอย่างทั้งหมด การเพิ่มขึ้นของความพรุนของปริมาณเนื่องจากเกิดแก๊สภายในนำไปสู่การแตกของพันธะภายในโครงสร้างของกาแฟ (Frisullo, Laverse, Barnabà, Navarini, และ Del Nobile, 2012) ซึ่งงานวิจัยนี้มีข้อเสีย คือการตัดเมล็ดกาแฟ ซึ่งทำให้เมล็ดกาแฟเสียหาย และไม่สามารถนำกระบวนการนี้มาศึกษาคุณภาพเมล็ดกาแฟจากภายนอกได้

7.) Maria I.Casas, Michael J. Vaughan, Pierluigi Bonello, Brian McSpadden Gardener, Erich Grotewold, Ana P. Alonso (2017) การศึกษาการบ่งชี้คุณสมบัติทางชีวเคมีของข้อบกพร่องของเมล็ดกาแฟ Coffea Arabica L. โดยใช้ตัวอย่างเมล็ดกาแฟจากสุมาตรา (อินโดนีเซีย) จำนวน 200 เมล็ด ครึ่งหนึ่งของตัวอย่างเมล็ดกาแฟครึ่งหนึ่งของเมล็ดกาแฟจากชุดกาแฟแต่ละชุดถูกจัดเรียงโดยแยกเป็นเมล็ดกาแฟที่มีข้อบกพร่องและไม่มีข้อบกพร่อง เมล็ดกาแฟที่มีข้อบกพร่องได้รับการกำหนดให้เป็นข้อบกพร่องที่มองเห็นได้ด้วยตาเปล่า ความเสียหายเมล็ดกาแฟ (เนื่องจากการแปรรูปหรือความเสียหายของแมลง) หรือความผิดปกติของสี (เช่นสีดำเต็มเมล็ด สีดำบางส่วน หรือเมล็ดฟอกขาว) โดยพิจารณาจากพารามิเตอร์ที่กำหนดไว้สำหรับข้อบกพร่องของเมล็ดโดย SCAA เปอร์เซ็นต์ปริมาณน้ำของเมล็ดกาแฟที่มีข้อบกพร่องและไม่มีข้อบกพร่องคือ ได้รับการกำหนดแบบกราฟิเมตริกหลังจากแช่แข็งแห้งเพื่อรักษาโครงสร้างของเมล็ดกาแฟ มีการทดสอบด้วยการวัดสารเคมี การวิเคราะห์สารอินทรีย์ระเหยง่าย และการสกัดสารเมทาโบไลต์ที่ละลายน้ำได้ : กรดอะมิโน กรดอินทรีย์ น้ำตาล และน้ำตาลแอลกอฮอล์ เมล็ดกาแฟที่มีข้อบกพร่องที่พบในการศึกษาครั้งนี้ ได้แก่ รูปทรงที่ผิดปกติ ความเสียหายจากแมลงและการ

เปลี่ยนสีของเมล็ดเป็นสีดำ เมล็ดกาแฟที่มีมีข้อบกพร่องคิดเป็นประมาณ 7% ของชุดตรวจ เพื่อตรวจสอบว่าการปรากฏตัวของข้อบกพร่องเหล่านี้มีผลต่อสมบัติทางเคมีกายภาพของเมล็ดกาแฟ ส่วนที่เหลือของลักษณะทางกายภาพของกาแฟสีเขียวไม่แตกต่างกันอย่างมีนัยสำคัญระหว่างสองส่วน ซึ่งแสดงให้เห็นว่าไม่มีผลกระทบต่อค่า pH และปริมาณกรดทั้งหมดที่มีอยู่ในตัวอย่างโดยการไทเทรต และไม่มี ความแตกต่างอย่างมีนัยสำคัญสำหรับตัวอย่างเมล็ดกาแฟคั่ว ซึ่งชี้ให้เห็นว่ามีการลดความแตกต่างของมวลหรือปริมาณน้ำเกิดขึ้นในระหว่างกระบวนการคั่ว (Casas และคนอื่น ๆ, 2017) ซึ่งงานวิจัยนี้มีข้อเสีย คือ การใช้ตัวอย่างของเมล็ดกาแฟที่ผ่านกระบวนการคั่ว และใช้วิธีการทางเคมีเพื่อชี้ข้อบกพร่องของเมล็ดกาแฟ

8.) Emanuelle Morais de Oliveira, Dimas Samid Leme, Bruno Henrique Groenner Barbosa, Mirian Pereira Rodarte, Rosemary Gualberto Fonseca Alvarenga Pereira, Yimeng Liu, Yizhi Wang and Yi Jin (2016) การศึกษาระบบคอมพิวเตอร์วิทัศน์สำหรับการจำแนกประเภทของเมล็ดกาแฟโดยใช้พื้นฐานจากเทคนิคปัญญาของคอมพิวเตอร์ กลุ่มตัวอย่างของเราประกอบด้วยเมล็ดกาแฟสีเขียวอาราบิก้า (*Coffea Arabica* L.) จากประเทศบราซิล โดยใช้วิธีการของ SCAA และ COB เราเลือกตัวอย่าง 120 กรัม (30 กรัมต่อสี) ของกลุ่มสีต่อไปนี้ สีขาว สีเขียว, สีเขียวอ่อน และสีเขียวน้ำเงิน สีเหล่านี้สอดคล้องกับสีที่ใช้มากที่สุดในเชิงพาณิชย์ และใช้ ANNs (Artificial Neural Networks - ANN) ในการศึกษา เนื่องจากเป็นตัวอย่างประมาณฟังก์ชันแบบสากล ชุดข้อมูลภาพ 564 สี ถูกสุ่มแบ่งเป็นชุดข้อมูล 2 ชุด ได้แก่ ชุดข้อมูลการฝึกฝน (80% ของตัวอย่าง) และชุดข้อมูลการตรวจสอบ (20% ของตัวอย่าง) เพื่อกำหนดโครงสร้าง ANN ใช้เทคนิค Bayesian (Naive-Bayes) มีความแม่นยำในการจำแนกได้ 100% ค่าเฉลี่ยของฟังก์ชันความหนาแน่นความน่าจะเป็นแบบ Gaussian จากการศึกษาค่า $L^* a^* b^*$ เพิ่มขึ้นตามระยะเวลาการเก็บรักษาที่เพิ่มขึ้นซึ่งสอดคล้องกับข้อค้นพบที่ได้จากการศึกษาของเรา จะเห็นว่าค่าดังกล่าว เพิ่มขึ้นอย่างมากจากตัวอย่างสีเขียวน้ำเงิน (ราคาดีที่สดุ) ไปจนถึงตัวอย่างสีขาว (ราคาต่ำสุด) นอกจากนี้ยังสามารถสังเกตได้ว่าระดับสีขาวมีค่าพารามิเตอร์สูง a^* (-4.62) และ b^* (-6.75) แสดงให้เห็นถึงการสูญเสียสีเขียวและสีน้ำเงินและแสดงสีที่ใกล้เคียงเหลืองมากขึ้น ในทางกลับกันระดับสีเขียวน้ำเงินมีค่า a^* (-8.93) และ b^* (-11.27) ที่ต่ำกว่าชั้นสีอื่น ๆ ซึ่งทำให้ระดับนี้ใกล้เคียงกับสีเขียวและสีน้ำเงินตามที่คาดไว้ (de Oliveira, Leme, Barbosa, Rodarte, และ Pereira, 2016) งานวิจัยนี้มีข้อเสีย คือ ใช้การวัดสีจากเมล็ดกาแฟเพียงอย่างเดียวในการควบคุมคุณภาพของเมล็ดกาแฟ

ตาราง 2 ตารางสรุปงานวิจัยที่ศึกษาทั้งหมด

งานวิจัย/ประเทศ/ปี	กลุ่มตัวอย่าง	วิธีการเก็บตัวอย่าง	Model
1. Revealing the Spatial Distribution of Chlorogenic Acids and Sucrose Across Coffee Bean Endosperm by Desorption Electrospray (Ireland) 2016	เมล็ดกาแฟ อะราบิก้า	กล้อง DESI-MS	-
2. Practical Comparison of Sparse Methods for Classification of Arabica and Robusta Coffee Species using Near Infrared Hyperspectral Imaging (Italy) 2015	เมล็ดกาแฟ อะราบิก้า เมล็ดกาแฟโรบัสต้า	กล้อง Hyperspectral (HSI)	KNN (PCA + kNN, sPCA+kNN), sPLS-DA
3. Fruit Classification using Computer Vision and Feedforward Neural Network (China) 2015	แอปเปิ้ล กล้วย, ส้มเขียวหวาน อะโวคาโด แตงโม, แคน ตาลูป สับปะรด ลูกแพร์ องุ่น บลูเบอร์รี่ สตอเบอร์รี่	กล้อง Digital	FNN (FSCABC-FNN, GA-FNN, PSO-FNN, ABC-FNN), KSVM
4. Classification of Green Coffee Bean Images Basec On Defect Types Using Convolutional Neural Network (CNN) (Japan) 2017	เมล็ดกาแฟ อะราบิก้า	กล้อง Digital	CNN
5. Colour Measurements by Computer Vision for Food Quality Control (Ireland) 2013	เนื้อวัว เนื้อหมู เนื้อปลา น้ำส้ม ไวน์ เบียร์ มันฝรั่ง ข้าวสาลี กล้วย	กล้อง A Solid State CCD Array	SVM, MLR
6. Coffee beans microstructural changes induced by cultivation processing: An X-ray micro tomographic investigation (Italy) 2012	เมล็ดกาแฟ อะราบิก้า	กล้อง X-Ray Microtomographic	
7. Identification of biochemical features of defective Coffea arabica L. beans (USA) 2017	เมล็ดกาแฟ อะราบิก้า	กระบวนการทางชีวเคมี	
8. A computer vision system for coffee beans classification based on computational intelligence techniques (Brazil) 2016	เมล็ดกาแฟ อะราบิก้า	กล้อง Digital	ANN, Bayesian

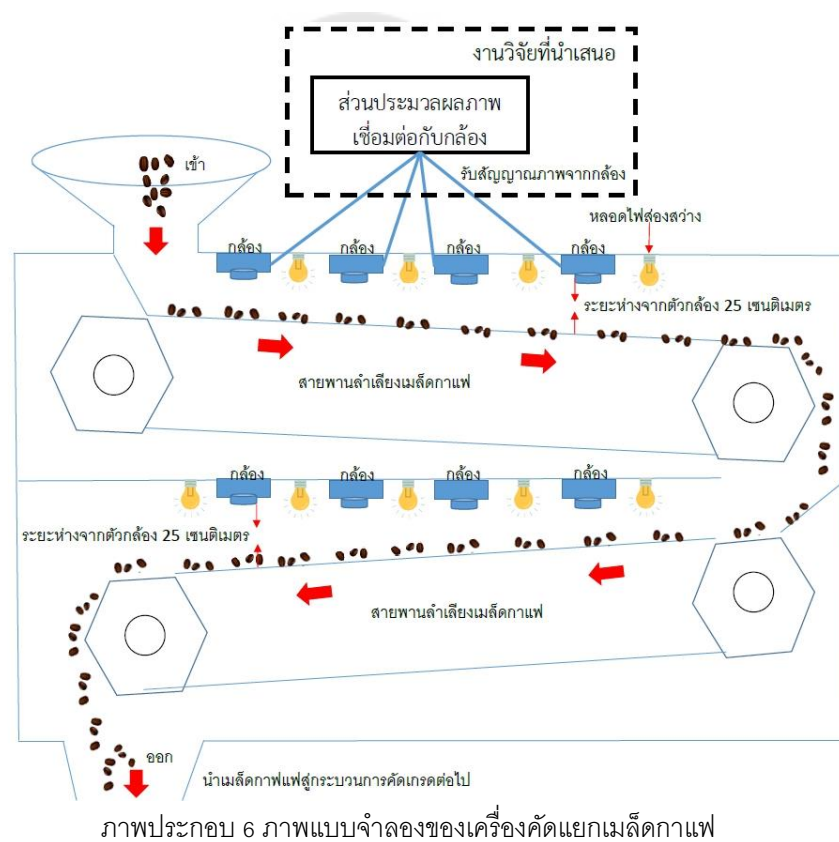
ข้อจำกัดของงานวิจัยที่ผ่านมาในอดีตสำหรับการจำแนกเมล็ดกาแฟดีและเมล็ดกาแฟเสีย คือ ไม่สามารถจำแนกเมล็ดกาแฟ เมล็ดดำ เมล็ดขึ้นรา เมล็ดแตก เมล็ดไม่สมบูรณ์ เมล็ดถูกแมลงทำลายได้ อีกทั้งงานวิจัยโดยส่วนใหญ่เป็นงานวิจัยที่มีการใช้แบบจำลองโครงข่ายประสาทเทียมแบบสังวัตนาการ Convolutional Neural Network (CNN) ซึ่งเป็นการวิธีการที่ใช้ฐานข้อมูลเป็นจำนวนมากเพื่อให้ประสิทธิภาพที่แม่นยำ ซึ่งคุณสมบัติของเมล็ดกาแฟที่คุณภาพดีและเสียสามารถจำแนกได้จากลักษณะทั่วไปได้เช่นกันจากข้อจำกัดของงานวิจัยที่ผ่านมาและลักษณะเด่นของคุณภาพเมล็ดกาแฟดังกล่าว งานวิจัยนี้นำเสนอการสกัดคุณลักษณะของเมล็ดกาแฟ เพื่อจำแนกประเภทตามคุณภาพด้วยการประมวลผลภาพถ่ายและการเรียนรู้ของเครื่อง เพื่อคุณภาพของเมล็ดกาแฟดีและเมล็ดกาแฟเสียได้ โดยที่เลือกใช้แบบจำลองที่ต่างกัน 2 แบบจำลอง คือ แบบจำลอง Random Forest (RF) และแบบจำลอง Logistic Regression (LR) โดยที่ แบบจำลอง Random Forest (RF) เป็นแบบจำลองที่ใช้งานง่าย เป็นแบบจำลองที่สุ่มเลือกใช้ข้อมูลและคุณลักษณะของ Decision Tree ซึ่งถูกสร้างจากการนำข้อมูลไปสุ่มเลือกตัวอย่างแล้วใส่กลับเข้ามาอีกครั้งหนึ่ง (Sampling with Replacement) จากนั้นจะถูกนำมาสร้างเป็นต้นไม้ ซึ่งจะมีตัวอย่างส่วนหนึ่งที่ไม่ถูกเลือกมาใช้ ซึ่งข้อมูลส่วนนี้เรียกว่า Out-of-Bag (OOB) แต่จะถูกนำไปใช้ในการทดสอบ Decision Tree วิธีการดังกล่าวนี้เรียกว่า Bagging ผลลัพธ์ที่ได้จาก Decision Tree ในแต่ละต้นถูกนำมาคิดเป็นผลการโหวต ผลโหวตที่มากที่สุดจะใช้ระบุสถานะของคลาส โดยสรุปแล้วเป็นแบบจำลองที่ใช้ต้นไม้การตัดสินใจ และได้แสดงให้เห็นมาโดยตลอดว่าสามารถใช้ในการจำแนกข้อมูลในหลากหลายรูปแบบอย่างกว้างขวาง อีกทั้งยังมีประสิทธิภาพที่ดีอีกด้วย ส่วนแบบจำลอง Logistic Regression (LR) นั้นเป็นแบบจำลองเชิงเส้น ซึ่งสามารถใช้จำแนกปัญหาของข้อมูลที่เป็นรูปแบบของโครงสร้างเชิงเส้นได้ดี เป็นแบบจำลองเชิงเส้นที่มีเส้นแบ่งเส้นเดียว แต่สามารถแบ่งกลุ่มตัวอย่างออกเป็น 2 คลาสได้อย่างชัดเจน และยังเป็นแบบจำลองที่มีขนาดเล็กกว่าในการทำ Classification โดยผลลัพธ์ออกมาเป็น Regression (จำนวนตัวเลข) จากนั้นจะนำไป Mapping กับ Sigmoid function (หลักคณิตศาสตร์) ก็จะได้คำตอบว่าข้อมูลที่ได้ (จากความเป็นไปได้ของ 2 คลาสเท่านั้น) เป็นคลาสนี้หรือไม่

ข้อดีของงานวิจัยนี้ คือ ไม่มีการทำลายเมล็ดกาแฟ และสามารถจำแนกเมล็ดเสีย เนื่องจากใช้วิธีการสกัดคุณลักษณะจากการประมวลผลภาพถ่ายของเมล็ดกาแฟโดยตรง ในหัวข้อถัดไปผู้วิจัยจะได้อธิบายถึงวิธีการดำเนินการวิจัยเพื่อให้ได้คุณลักษณะของเมล็ดกาแฟในการนำไปจำแนกเมล็ดกาแฟได้ จากนั้นจึงอธิบายผลการวิจัยและอภิปรายผลการวิจัย และในหัวข้อสุดท้ายเป็นการสรุปผลการวิจัย

บทที่ 3

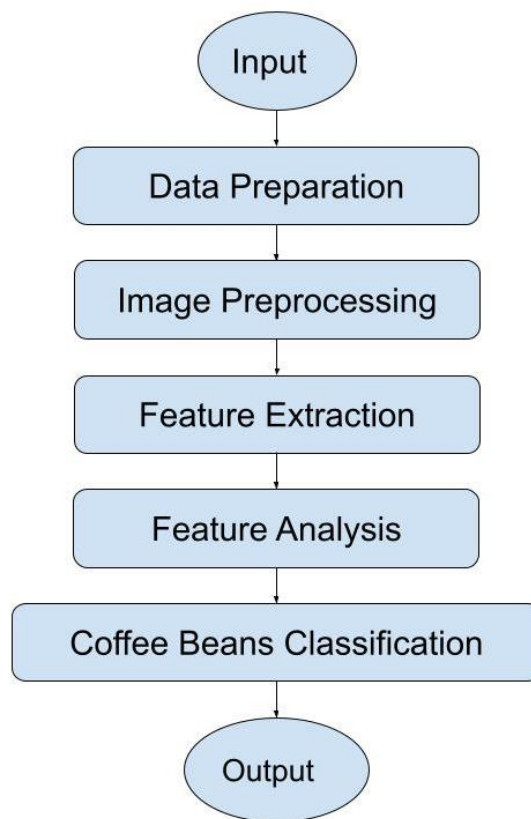
วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นการพัฒนาระบบประมวลผลภาพสำหรับเครื่องจำแนกคุณภาพเมล็ดกาแฟ โดยเครื่องจำแนกเมล็ดกาแฟแสดงดังภาพประกอบ 6 ซึ่งประกอบด้วย ระบบนำเข้าเมล็ดกาแฟ ระบบสายพานลำเลียงเมล็ดกาแฟ ระบบแสงสว่าง และระบบถ่ายภาพ โดยกล้องอยู่สูงจากระบบสายพานประมาณ 25 เซนติเมตร ซึ่งภาพที่ได้จะส่งเข้าสู่ระบบการประมวลผลภาพถ่าย



วิธีการและขั้นตอนสำหรับการประมวลผลภาพเพื่อจำแนกคุณภาพเมล็ดกาแฟ จะแบ่งออกเป็นลำดับดังภาพประกอบ 7 ซึ่งประกอบด้วย

- 1.) การเตรียมข้อมูลเพื่อใช้ในการทดลอง (Data Preparation)
- 2.) การประมวลผลภาพเบื้องต้น (Image Preprocessing)
- 3.) การสกัดคุณลักษณะ (Feature Extraction)
- 4.) การวิเคราะห์คุณลักษณะ (Feature Analysis)
- 5.) การจำแนกประเภทเมล็ดกาแฟ (Coffee Beans Classification)



ภาพประกอบ 7 Flowchart การทำงานของระบบ

3.1 การเตรียมข้อมูลเพื่อใช้ในการทดลอง (Data Preparation)

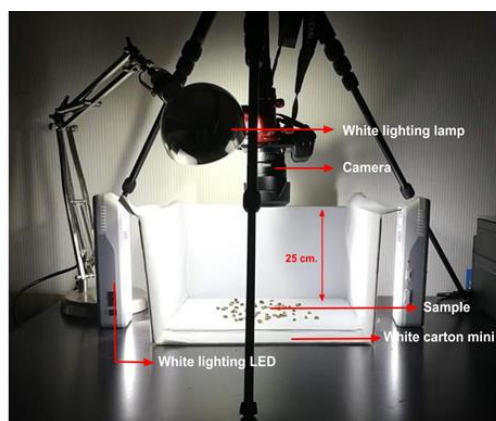
ทำการวิเคราะห์ปัญหา และโอกาสทางธุรกิจ (Business Understanding) เพื่อค้นหาปัจจัยที่มีส่วนเกี่ยวข้องกับการคัดแยกหรือแบ่งประเภทของเมล็ดกาแฟ ตามความต้องการของผู้ผลิตหรือเจ้าของธุรกิจกาแฟขนาดเล็ก โดยการอ้างอิงถึงมาตรฐานในการคัดแยกเมล็ดกาแฟตามประกาศกระทรวงเกษตรและสหกรณ์ เรื่อง กำหนดมาตรฐานสินค้าเกษตร : เมล็ดกาแฟอะราบิก้า ตามพระราชบัญญัติมาตรฐานสินค้าเกษตร พ.ศ. 2551 แยกเป็น 9 นิยามความบกพร่อง

3.2 การประมวลผลภาพเบื้องต้น (Image Preprocessing)

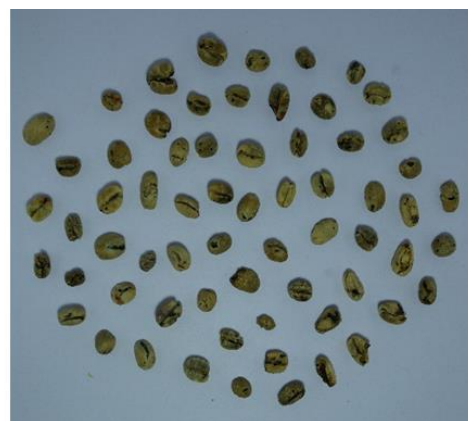
3.3.1 การเตรียมข้อมูลรูปภาพ (Image Acquisition)

เป็นขั้นตอนของการจัดหาข้อมูลที่มีคุณภาพก่อนขั้นตอนการคัดแยกหรือแบ่งประเภทเมล็ดกาแฟ ในขั้นตอนนี้จะเตรียมข้อมูลด้วยการถ่ายภาพเมล็ดกาแฟ ทำการถ่ายภาพโดยนำเมล็ดกาแฟเรียงแถวจำนวน 2 แถวหรือมากกว่า ถ่ายภาพที่ความละเอียด 6000 x 4000 พิกเซล ด้วยกล้องถ่ายภาพดิจิทัลยี่ห้อ Sony รุ่น a6000 ที่ระดับความสูง 25 เซนติเมตร (วัดจากเมล็ดกาแฟถึง

ตัวกล้อง) ภายใต้การส่องสว่างจากโคมไฟแสงสีขาว 8W-PS/65K และแสงจากหลอด LED ขนาด 10W-PS/65K การจัดเก็บไฟล์ภาพในปริภูมิสี (color space) RGB และรูปแบบไฟล์แบบ JPEG ดังภาพประกอบ 8 โดยมีจำนวนภาพของเมล็ดกาแฟทั้งหมดจำนวน 95 รูปภาพ และตัวอย่างเมล็ดแบบต่างๆ ดังภาพประกอบ 9

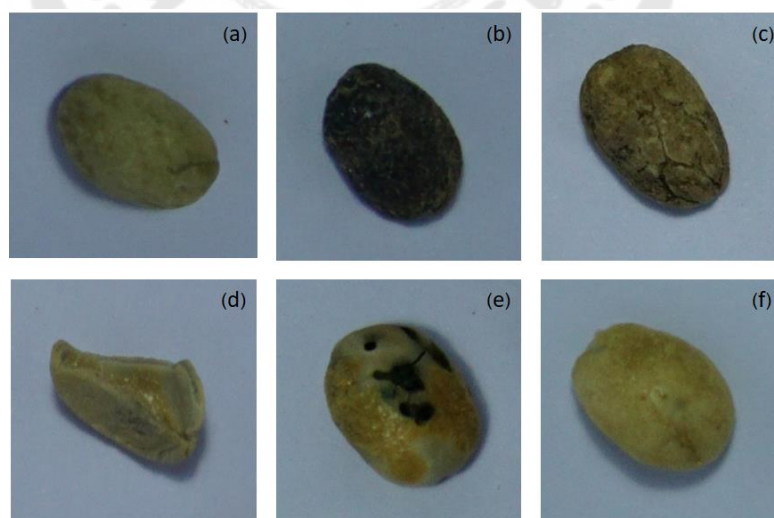


ภาพการเก็บตัวอย่างภาพถ่ายเมล็ดกาแฟ



ภาพถ่ายตัวอย่างเมล็ดกาแฟ

ภาพประกอบ 8 การเก็บภาพถ่ายและตัวอย่างภาพถ่ายเมล็ดกาแฟ



ภาพประกอบ 9 ภาพเมล็ดเสีย a) เมล็ดกาแฟดี b) เมล็ดดำ c) เมล็ดขึ้นรา
d) เมล็ดแตก e) เมล็ดไม่สมบูรณ์ f) เมล็ดถูกแมลงทำลาย

3.2.2 การประมวลผลภาพ (Image Preprocessing)

เมื่อได้ภาพเมลิ็ดกาแฟแล้ว ผู้วิจัยทำการปรับปรุงคุณภาพของภาพ โดยลักษณะเด่นของภาพที่เหมาะสมสำหรับการจำแนกเมลิ็ดกาแฟแสดงดังตาราง 3

ตาราง 3 ลักษณะเด่นของคุณภาพเมลิ็ดกาแฟแต่ละประเภท

ประเภทเมลิ็ดกาแฟ	ลักษณะเด่นของเมลิ็ดกาแฟ	ตัวแปรที่ใช้ในการจำแนกคุณภาพเมลิ็ดกาแฟ
1. เมลิ็ดกาแฟดี	มีลักษณะเป็นเมล็ดที่สมบูรณ์ ทรงกลมรี สีออกเขียวอ่อน	- พื้นที่ของเมลิ็ดกาแฟ (Contour Area) - ค่าจากค่าอัตราส่วนของพื้นที่รูปร่างต่อพื้นที่เปลือก (Solidity) - เส้นผ่านศูนย์กลาง (Equivalent Diameter) - ค่าสัดส่วนของพื้นที่วงกลม (Extent) - ค่ามากของสีแดง (R Max) - ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) ของสีแดง สีเขียว สีน้ำเงิน และสีเทา

ตาราง 3 (ต่อ)

ประเภทเมล็ดกาแฟ	ลักษณะเด่นของเมล็ดกาแฟ	ตัวแปรที่ใช้ในการจำแนกคุณภาพเมล็ดกาแฟ
2. เมล็ดกาแฟสีดำน	เมล็ดกาแฟมีสีน้ำตาลอ่อนจนถึงสีดำสนิท และมีลักษณะของสีที่ไม่สม่ำเสมอบนเมล็ดกาแฟ	- พื้นที่ของเมล็ดกาแฟ (Contour Area) - ค่าจากค่าอัตราส่วนของพื้นที่รูปร่างต่อพื้นที่เปลือก (Solidity) - เส้นผ่านศูนย์กลาง (Equivalent Diameter) - ค่าสัดส่วนของพื้นที่วงกลม (Extent) - ค่ามาก (Max) ของสีแดง สีเขียว และสีน้ำเงิน - ค่าเฉลี่ย (Average) ของสีแดง สีเขียว สีน้ำเงิน และสีเทา - ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) ของสีแดง สีเขียว สีน้ำเงิน และสีเทา - ค่าความแปรปรวนของข้อมูล (Variance) ของสีแดง สีเขียว สีน้ำเงิน และสีเทา
3. เมล็ดกาแฟชั้นรา	ให้สีหรือลักษณะของเมล็ดผิดปกติ มีสีที่ไม่สม่ำเสมอ มีสีเขียวปนน้ำตาล	- พื้นที่ของเมล็ดกาแฟ (Contour Area) - เส้นผ่านศูนย์กลาง (Equivalent Diameter) - ค่าสัดส่วนของพื้นที่วงกลม (Extent) - ค่ามาก (Max) ของสีเขียว - ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) ค่าสีของสีหลัก (แดง เขียวและสีน้ำเงิน) (Hue) - ค่าความแปรปรวนของข้อมูล (Variance) ของสีแดง สีเขียว สีน้ำเงิน และสีเทา

ตาราง 3 (ต่อ)

ประเภทเมล็ดกาแฟ	ลักษณะเด่นของเมล็ดกาแฟ	ตัวแปรที่ใช้ในการจำแนกคุณภาพเมล็ดกาแฟ
4. เมล็ดกาแฟแตก	เมล็ดกาแฟมีลักษณะที่เกิดการแตกหักของเมล็ด ทำให้เมล็ดกาแฟมีการแตกหรือหักออกเป็น 2 ชิ้น หรือมากกว่าครึ่งหนึ่งของเมล็ดกาแฟ	- พื้นที่ของเมล็ดกาแฟ (Contour Area) - ค่าจากค่าอัตราส่วนของพื้นที่รูปร่างต่อพื้นที่เปลือก (Solidity) - เส้นผ่านศูนย์กลาง (Equivalent Diameter) - ค่าสัดส่วนของพื้นที่วงกลม (Extent)
5. เมล็ดกาแฟถูกแมลงทำลาย	เมล็ดกาแฟที่มีลักษณะของการถูกแมลงเข้าทำลายภายในเมล็ด จากการกัดแทะ หรือเจาะจนเกิดเป็นรู	- พื้นที่ของเมล็ดกาแฟ (Contour Area) - ค่าสัดส่วนของพื้นที่วงกลม (Extent)
6. เมล็ดกาแฟไม่สมบูรณ์	เมล็ดกาแฟมีรูปทรงและขนาดผิดปกติจากลักษณะการเหี่ยวยุบของเมล็ด มีรูปทรงที่ผิดปกติ และมีสีของเมล็ดนั้นเขียวซีดไปจนถึงสีเหลืองออกขาว	- พื้นที่ของเมล็ดกาแฟ (Contour Area) - ค่าจากค่าอัตราส่วนของพื้นที่รูปร่างต่อพื้นที่เปลือก (Solidity) - เส้นผ่านศูนย์กลาง (Equivalent Diameter) - ค่าสัดส่วนของพื้นที่วงกลม (Extent) - ค่ามาก (Max) ของสีแดง สีเขียว และสีน้ำเงิน - ค่าเฉลี่ย (Average) ของสีแดง สีเขียว สีน้ำเงิน และสีเทา

เพื่อให้สามารถดึงคุณลักษณะของเมล็ดกาแฟออกมาจากภาพถ่าย โดยใช้วิธีลบสัญญาณรบกวน ด้วยการใส่ Gaussian Blur ดังภาพประกอบ 10 และภาพประกอบ 11

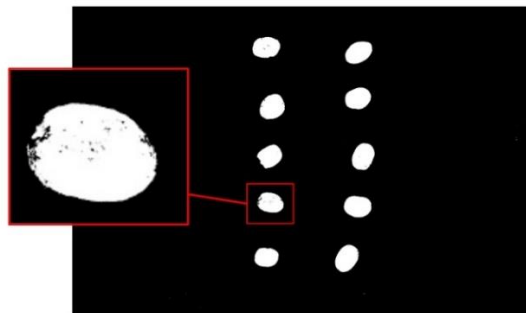


ภาพประกอบ 10 รูปภาพต้นฉบับ

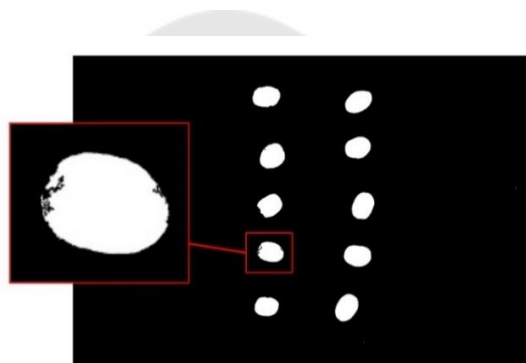


ภาพประกอบ 11 การลบสัญญาณรบกวน ด้วยการใช้ Gaussian Blur

จากนั้นนำภาพที่ได้มาหาค่า Threshold โดยใช้ Gray Thresh เพื่อให้สามารถแยกภาพพื้นหลังและเมล็ดกาแฟออกจากกัน ผลที่ได้จะพบว่ามีส่วนของเมล็ดที่เป็นรูอยู่ ผู้วิจัยจึงใช้วิธี Imclose เพื่อเติมส่วนที่แหว่ง หรือบริเวณที่ขาดหายไป และ Imfill เติมในส่วนที่เป็นรู แล้วใช้ Im2bw เพื่อแปลงเป็นภาพขาวดำ ดังภาพประกอบ 12 และภาพประกอบ 13

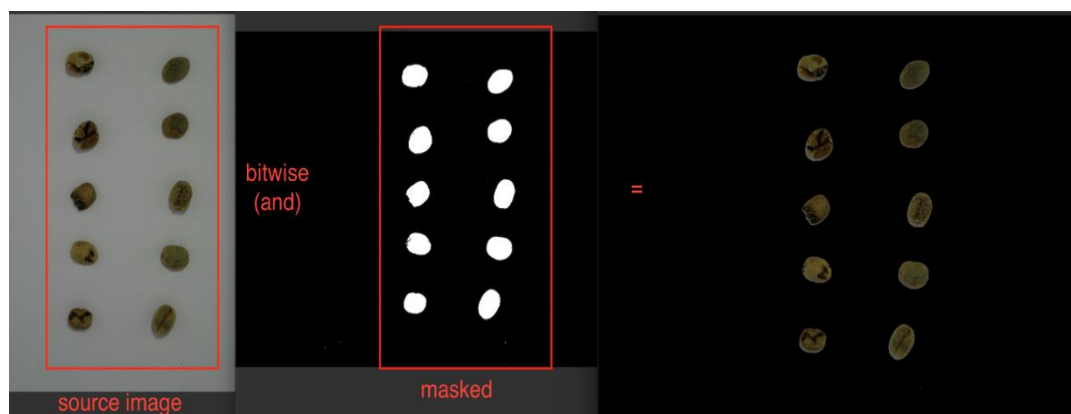


ภาพประกอบ 12 การทำ Threshold เพื่อแยก เมล็ดกับภาพพื้นหลัง



ภาพประกอบ 13 การทำ Flood Fill เพื่อเติมเต็มส่วนที่ขาดหายไป

หลังจากที่ทำ Flood Fill เสร็จในขั้นตอนนั้นแล้ว เราจะได้สิ่งที่เรียกว่า Masked ซึ่งในขั้นตอนต่อไปจะนำเอา Masked ไป Bitwise กับรูปภาพต้นฉบับ ซึ่งในการ Bitwise นั้นจะมีหลายตัว เช่น or, and และ xor แต่งานวิจัยนี้เลือกใช้ and และผลลัพธ์ที่ได้จากการ Bitwise คือเราจะสามารถแยกเมล็ดกาแฟกับภาพพื้นหลังออกจากกันได้แล้วดังภาพประกอบ 14 และภาพประกอบ 15



ภาพประกอบ 14 การทำ bitwise เพื่อให้สามารถแยกเมล็ดกาแฟออกจากภาพพื้นหลัง

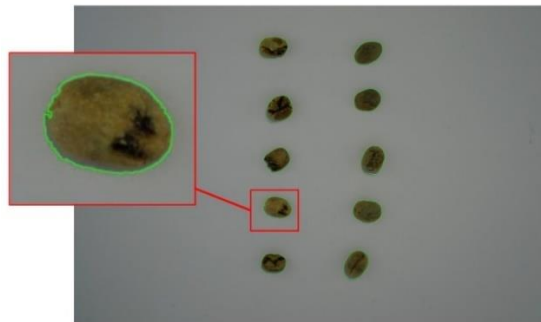


ภาพประกอบ 15 เมล็ดกาแฟสามารถแยกออกจากภาพพื้นหลังได้อย่างชัดเจน

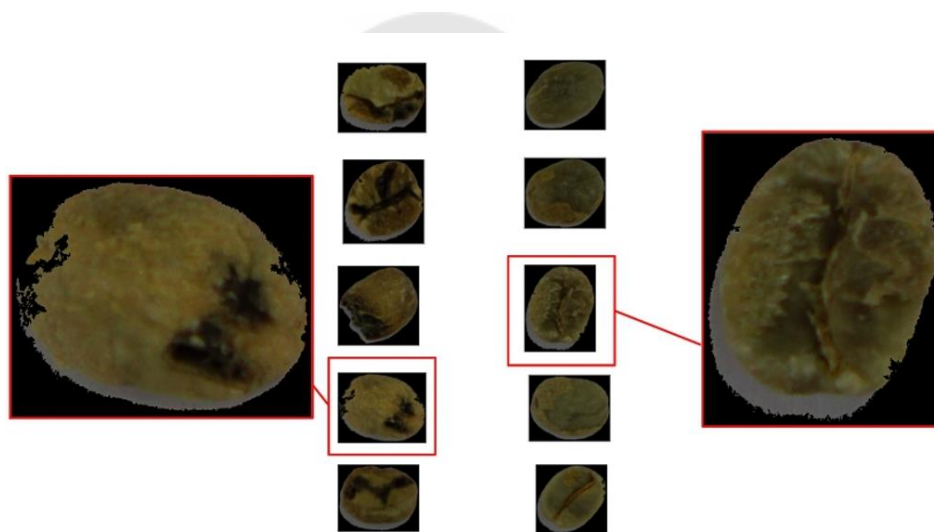
เมื่อได้ภาพเมล็ดกาแฟที่แยกออกจากภาพพื้นหลังอย่างชัดเจนแล้ว ในขั้นตอนถัดไปนั้นจึงทำการหา Edge and Dilation เพื่อที่จะนำไปหาเส้น Contour แต่เนื่องจากผลลัพธ์ของเส้นที่หาได้นั้น มีเส้นที่มีส่วนขาดๆ ซึ่งไม่สามารถนำไปใช้หาเส้น Contour ต่อไปได้ จึงต้องทำ Dilation เพื่อเติมเส้นส่วนที่ขาดนั้นให้เชื่อมต่อกันเสียก่อนดังภาพประกอบ 16 จากนั้นจึงนำรูปภาพที่ได้ไปหาเส้น Contour เพื่อที่จะนำเอาเส้น Contour ของแต่ละเมล็ดนั้น มาหา Bounding Box เพื่อให้ได้กรอบสี่เหลี่ยมล้อมรอบเมล็ดกาแฟ และนำมาใช้ในการหั่นภาพออกมาเป็นแต่ละเมล็ด ดังภาพประกอบ 17 และภาพประกอบ 18



ภาพประกอบ 16 การทำ Dilation เพื่อเติมเส้นที่ขาด



ภาพประกอบ 17 ใช้ Contours เพื่อหาเส้นรูปทรงสีเหลี่ยม

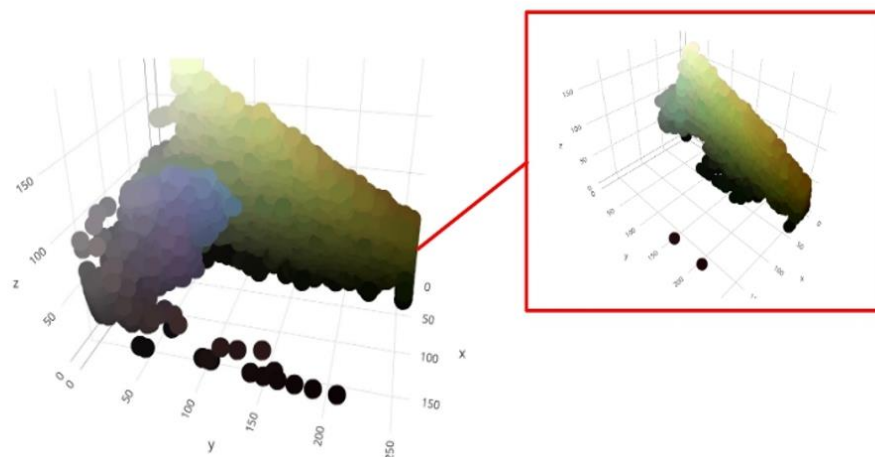


ภาพประกอบ 18 การ extract ภาพเมล็ดกาแฟ

หลังจากที่สามารถหั่นภาพเมล็ดกาแฟออกเป็น 1 ภาพต่อ 1 เมล็ดแล้ว จะสังเกตได้ว่าเมล็ดกาแฟที่ถูกหั่นออกมานั้น จะมีขอบบางส่วนเกินที่เกิดขึ้นขณะถ่ายภาพติดมากับเมล็ดกาแฟด้วย จึงต้องทำการหาขอบเงา และทำการตัดขอบส่วนเกินของเงาที่เกิดขึ้นออก โดยทำการนำภาพมาแปลงค่าเป็น HSV จากนั้นนำมา Plot แบบ 3 มิติ เพื่อดูความสัมพันธ์ของสี จากนั้นจึงทำการตัดส่วนเกินของเงาที่เกิดขึ้นออกได้ ดังภาพประกอบ 19 และภาพประกอบ 20



ภาพประกอบ 19 เมล็ดกาแฟที่มีขอบเงาส่วนเกิน



ภาพประกอบ 20 ขั้นตอนการตัดเงาออกจากเมล็ดกาแฟ



ภาพประกอบ 21 ผลลัพธ์จากการตัดเงาออกจากเมล็ดกาแฟ

เมื่อทำการตัดขอบเงาส่วนเกินออกจากภาพได้แล้ว จึงจะสามารถนำภาพเมล็ดกาแฟทั้งหมดมาทำการหาคุณลักษณะเฉพาะ เพื่อให้สามารถนำคุณลักษณะเฉพาะที่ได้นั้น มาวิเคราะห์เพื่อจำแนกเมล็ดกาแฟในแต่ละประเภทต่อไป โดยจำนวนตัวอย่างเมล็ดกาแฟแต่ละประเภทนั้นแสดงรายละเอียดได้ดังตาราง 4 ดังนี้

ตาราง 4 ตารางจำนวนตัวอย่างเมล็ดกาแฟแต่ละประเภท

ประเภทของเมล็ดกาแฟ	จำนวน (เมล็ด)	จำนวน (รูปภาพ)
1. เมล็ดกาแฟดี	1,303	22
2. เมล็ดกาแฟสีดำ	1,002	15
3. เมล็ดกาแฟชั้นรา	1,081	14
4. เมล็ดกาแฟแตก	700	12
5. เมล็ดกาแฟถูกแมลงทำลาย	700	11
6. เมล็ดกาแฟไม่สมบูรณ์	1,124	21
รวม	5,910	95

3.3 การสกัดคุณลักษณะ (Feature Extraction)

การวิเคราะห์คุณลักษณะ สามารถกระทำโดยการแปลงภาพถ่ายของเมล็ดกาแฟทั้งในส่วนของเมล็ดทั้ง 6 ประเภท จะถูกนำมาแบ่งออกเป็น 2 กลุ่มคือ (1) กลุ่มของเมล็ดกาแฟดี และ (2) กลุ่มของเมล็ดกาแฟเสีย โดยที่ เมล็ดกาแฟดำ, เมล็ดกาแฟชั้นรา, เมล็ดกาแฟแตก, เมล็ดกาแฟถูกแมลงทำลาย, เมล็ดกาแฟไม่สมบูรณ์ จะถูกจัดอยู่ในกลุ่มของเมล็ดกาแฟเสีย และเมล็ดกาแฟทั้ง 2 กลุ่มนี้จะถูกนำมาหาค่าคุณลักษณะ ก่อนทำการคัดเลือกคุณลักษณะที่ส่งผลต่อการจำแนกประเภทเมล็ดกาแฟได้ออกมาเป็นคุณลักษณะจำนวน 32 คุณลักษณะ โดยที่คุณลักษณะที่ถูกคัดออกไปนั้นไม่ได้เกี่ยวข้องกับขนาด รูปร่าง สี ซึ่งอยู่ในประเด็นหลักในการกำหนดคุณภาพของเมล็ดกาแฟ คุณลักษณะที่ถูกคัดออกจากงานวิจัยนี้ เช่น เนื้อสัมผัส ลักษณะของพื้นผิวเมล็ดกาแฟ กลิ่น และรสชาติ เป็นต้น คุณลักษณะที่ถูกคัดเลือกจะถูกแยกเป็น 3 กลุ่ม ประกอบด้วย

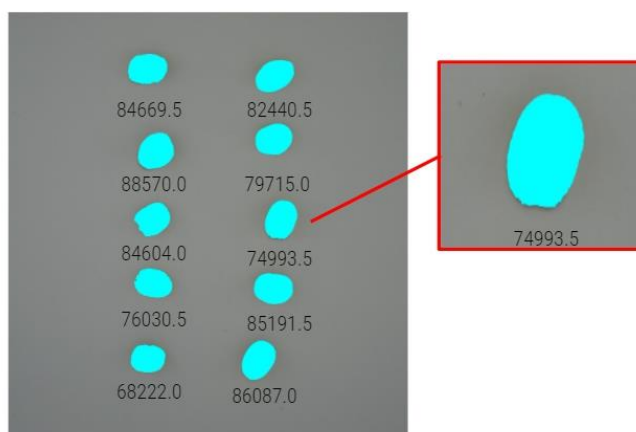
3.3.1 กลุ่มของขนาดและรูปร่าง

จำนวน 4 คุณลักษณะ ได้แก่ พื้นที่ของเมล็ดกาแฟ ค่าจากค่าอัตราส่วนของพื้นที่รูปร่างต่อพื้นที่เปลือก เส้นผ่านศูนย์กลาง และค่าสัดส่วนของพื้นที่วงกลม ดังนี้

3.3.1.1 พื้นที่ของเมล็ดกาแฟ (Contour Area)

คือ ค่าของขนาดพื้นที่ของเมล็ดกาแฟ นับรวมตั้งแต่บริเวณที่เป็นขอบเมล็ด จนถึงบริเวณของเมล็ดกาแฟทั้งหมดที่ภาพถ่ายสามารถจับได้ ดังสมการ 8 ภาพประกอบ 22

$$\text{Contour Area} = cv.\text{contourArea}(cnt) \quad (8)$$

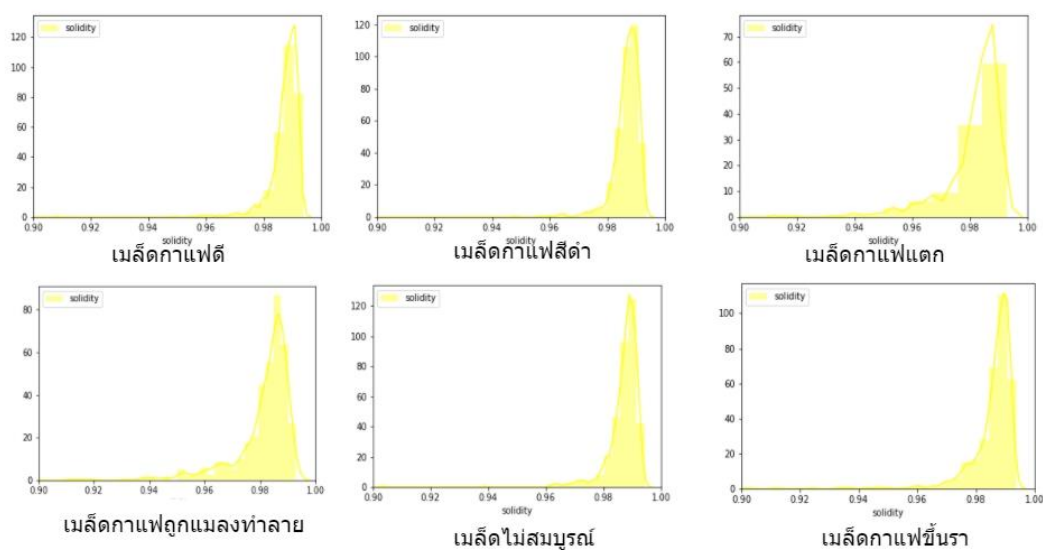


ภาพประกอบ 22 ภาพการหาค่าจากค่าพื้นที่ของเมล็ดกาแฟ (Contour Area)

3.3.1.2 ค่าจากค่าอัตราส่วนของพื้นที่รูปร่างต่อพื้นที่เปลือก (Solidity)

เป็นคุณลักษณะที่สามารถใช้ดูเมล็ดกาแฟ นั้นเต็มเมล็ดหรือไม่ ถ้าไม่เต็มเมล็ดสามารถบ่งบอกได้ว่าเมล็ดนั้นอาจเป็นเมล็ดกาแฟเสีย ซึ่งสามารถนำคุณลักษณะนี้ไปใช้กับกาแฟประเภทเมล็ดกาแฟแตก หรือเมล็ดกาแฟที่ขาดแหว่งจากการกัดแทะของแมลงได้ โดย Solidity หาได้จากสมการที่ 9

$$\text{Solidity} = \frac{\text{Contuur Area}}{\text{Convex Hall Area}} \quad (9)$$



ภาพประกอบ 23 ภาพฮิสโตแกรมการหาค่าจากค่าอัตราส่วนของพื้นที่รูปร่างต่อพื้นที่เปลือก (Solidity)

3.3.1.3 เส้นผ่านศูนย์กลางสมมูล (Equivalent Diameter)

คือเส้นผ่านศูนย์กลางสมมูล เส้นผ่านศูนย์กลาง (Diameter) ของทรงกลมที่มีคุณสมบัติเทียบเท่าวัสดุ เส้นผ่านศูนย์กลางสมมูลนี้สามารถใช้กำหนดขนาด (size) ของวัสดุที่มีรูปร่างไม่เป็นทรงกลม ใช้เพื่อบอกขนาดอนุภาค (particle size) ค่าเส้นผ่านศูนย์กลางยิ่งมากแสดงว่าขนาดของเมล็ดกาแฟนั้นมีขนาดใหญ่ ถ้าค่าน้อยก็แสดงว่าเมล็ดกาแฟนั้นอาจมีขนาดเล็ก หาได้จากสมการที่ 10 และภาพประกอบ 24

$$\text{Equivalent Diameter} = \sqrt{\frac{4 \times \text{Contour Area}}{\pi}} \quad (10)$$

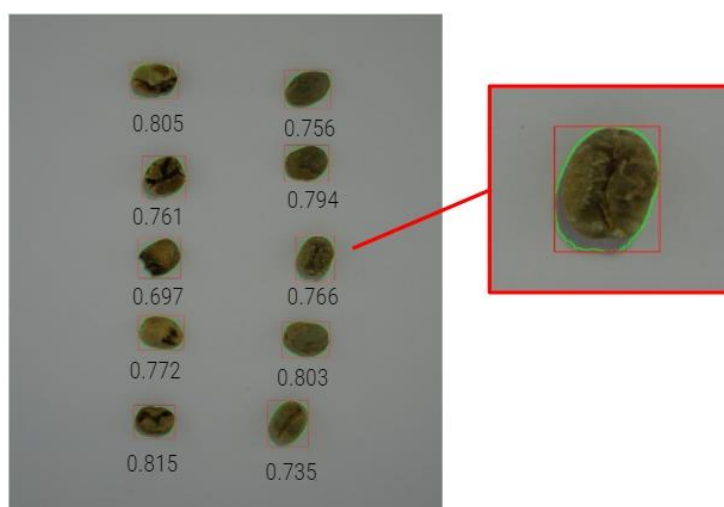


ภาพประกอบ 24 ภาพการหาค่าเส้นผ่านศูนย์กลาง (Equivalent Diameter)

3.3.1.4 ค่าสัดส่วนของพื้นที่วงกลม (Extent)

เป็นค่าสัดส่วนของพื้นที่วงกลม พื้นที่ของเมล็ดกาแฟ (Contour area) โดยนับพื้นที่ส่วนหนึ่งที่เทียบกับพื้นที่ทั้งหมดของเมล็ดกาแฟ และแปลงเป็นค่าร้อยละ หาได้จากสมการที่ 10 และภาพประกอบ 25

$$Extent = \frac{Object Area}{Bounding Rectangle area} \quad (10)$$



ภาพประกอบ 25 การหาค่าสัดส่วนของพื้นที่วงกลม (Extent)

3.3.2 กลุ่มของค่าสี

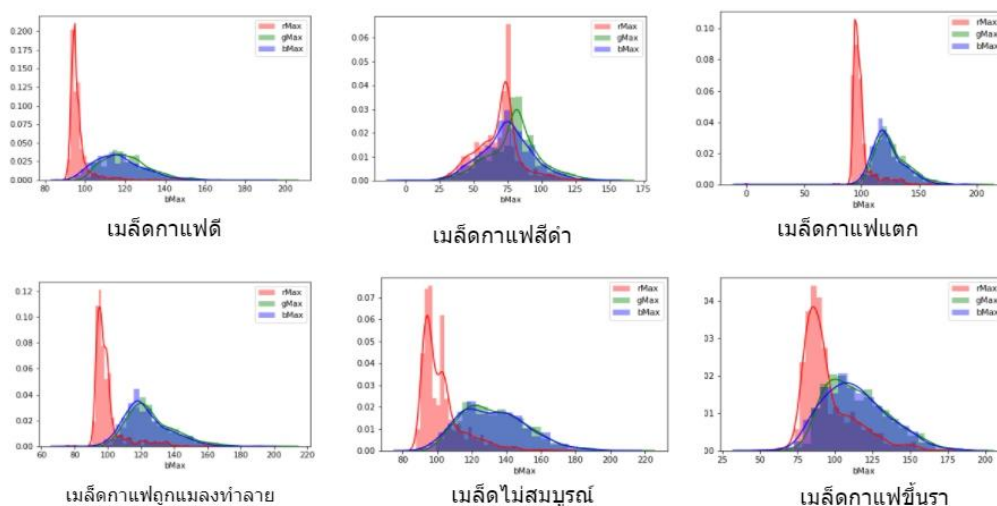
จำนวน 24 คุณลักษณะ ได้แก่ ค่ามากของสีแดง ค่ามากของสีเขียว ค่ามากของสีน้ำเงิน ดังภาพประกอบ 26 ค่ามากของ Hue ค่ามากของ Saturation ค่ามากของ Value ค่าเฉลี่ยของสีแดง ค่าเฉลี่ยของสีเขียว ค่าเฉลี่ยของสีน้ำเงิน ค่าเฉลี่ยของ Hue ค่าเฉลี่ยของ Saturation ค่าเฉลี่ยของ Value ค่าเบี่ยงเบนมาตรฐานของสีแดง ค่าเบี่ยงเบนมาตรฐานของสีเขียว ค่าเบี่ยงเบนมาตรฐานของสีน้ำเงิน ค่าเบี่ยงเบนมาตรฐานของ Hue ค่าเบี่ยงเบนมาตรฐานของ Saturation ค่าเบี่ยงเบนมาตรฐานของ Value ค่าความแปรปรวนของสีแดง ค่าความแปรปรวนของสีเขียว ค่าความแปรปรวนของสีน้ำเงิน ค่าความแปรปรวนของ Hue ค่าความแปรปรวนของ Saturation ค่าความแปรปรวนของ Value ดังนี้

1. ค่าสีแดง สีเขียว และสีน้ำเงิน (RGB) เป็นระบบสีที่เกิดจากการรวมกันของแสงสีแดง เขียว และน้ำเงินโดยมีการรวมกันแบบ Additive หาได้จากสมการที่ 11, 12 และสมการที่ 13 ตามลำดับ

$$R = \frac{R}{R + G + B} \quad (11)$$

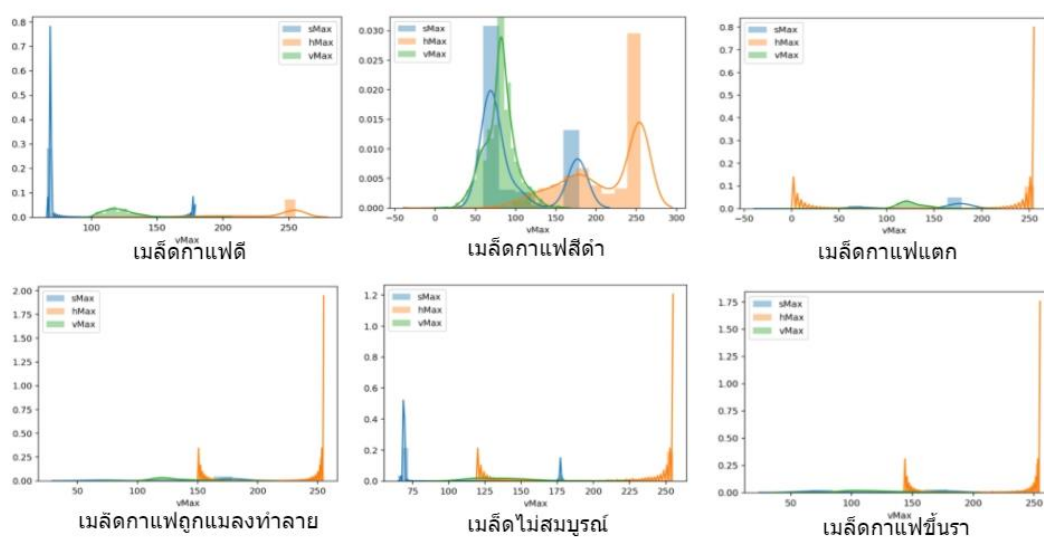
$$G = \frac{G}{R + G + B} \quad (12)$$

$$B = \frac{B}{R + G + B} \quad (13)$$



ภาพประกอบ 26 ภาพฮิสโตแกรมการหาค่ามาก (Max) ของสีแดง สีเขียว สีน้ำเงิน

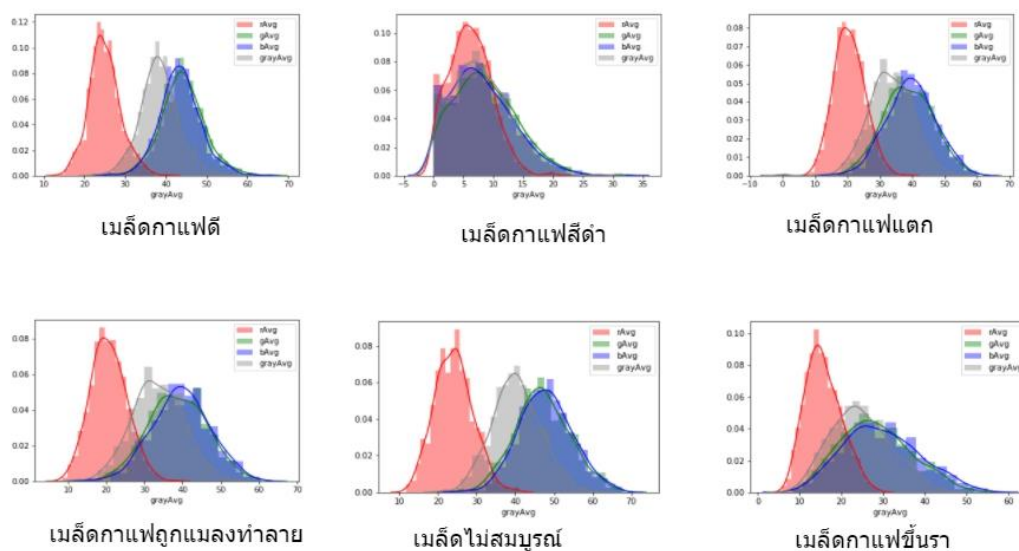
2. ค่าสี HSV (Hue Saturation Value) เป็นค่าสีของความเข้มตัว ค่าของเฉดสี และค่าความสว่างของสี โดยใช้หลักการพิจารณาการแสดงความตัวเลขอยู่ระหว่าง 0 ถึง 255 โดยที่ Hue มีค่าเท่ากับ 0 จะแทนสีแดง และเมื่อ Hue มีค่าเพิ่มขึ้นค่าของสีก็จะเปลี่ยนตามสเปกตรัมของสีไปด้วย และเมื่อค่าถึง 256 ก็จะกลับมาเป็นสีแดงอีกครั้ง ซึ่งสามารถแทนให้อยู่ในรูปขององศาได้ ดังนี้คือ สีแดงเท่ากับ 0 องศา สีเขียวเท่ากับ 120 องศา สีน้ำเงินเท่ากับ 240 องศา



ภาพประกอบ 27 ภาพฮิสโตแกรมการหาค่ามาก (Max) ของค่าสี HSV (Hue Saturation Value)

3. ค่าเฉลี่ย (Average) คือ ค่ากลางที่ได้มาจากการรวมข้อมูลทั้งหมดแล้วแบ่ง (เฉลี่ย) ให้ได้จำนวนเท่า ๆ กัน จากผลรวมของข้อมูลทั้งหมด โดยที่ ค่า X แทนค่าของข้อมูลทั้งหมด และ N แทนค่าของจำนวนของข้อมูล หาได้จากสมการที่ 14

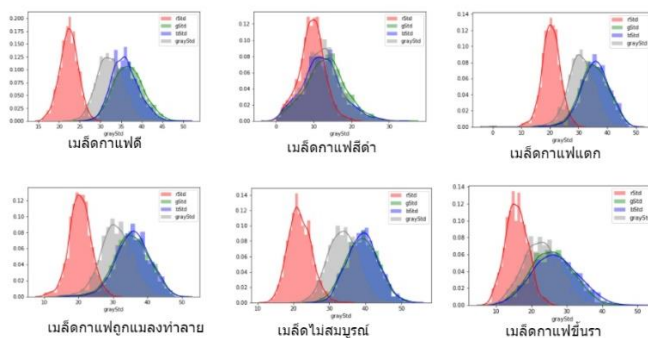
$$Average = \frac{\sum X}{N} \quad (14)$$



ภาพประกอบ 28 ภาพฮิสโตแกรมการหาค่าเฉลี่ย (Average) ของสีแดง สีเขียว สีนํ้าเงิน และสีเทา

4. ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) คือ โดยที่ ค่า SD จะแสดงให้เห็นว่าโดยเฉลี่ยแล้วข้อมูลแต่ละตัวนั้นแตกต่างจากค่าเฉลี่ยเท่าไร โดยที่ค่า $\bar{X}$ แทนค่าเฉลี่ยของกลุ่มตัวอย่าง หาได้จากการนำข้อมูลแต่ละตัวอย่างลบด้วยค่าเฉลี่ยของกลุ่มตัวอย่าง หาได้จากสมการที่ 15

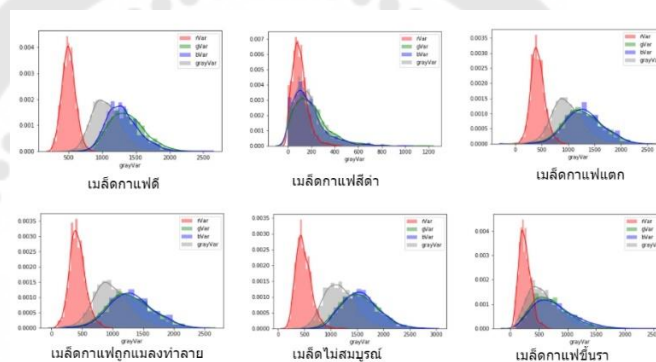
$$\sigma = \sqrt{\frac{\sum(X - \bar{X})^2}{N}} \quad (15)$$



ภาพประกอบ 29 ภาพฮิสโตแกรมการหาค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) ของสีแดง สีเขียว สีนํ้าเงิน และสีเทา

5. ค่าความแปรปรวนของข้อมูล (Variance) คือ เป็นค่าที่ใช้วัดการกระจายของข้อมูล หาได้จากการนำส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง โดยที่ค่า $\bar{X}$ แทนค่าเฉลี่ยของกลุ่มตัวอย่าง หาได้จากการนำข้อมูลแต่ละตัวอย่างลบด้วยค่าเฉลี่ยของกลุ่มตัวอย่าง ยกกำลัง 2 หรือค่าเบี่ยงเบนมาตรฐาน ยกกำลัง 2 ดังสมการที่ 16

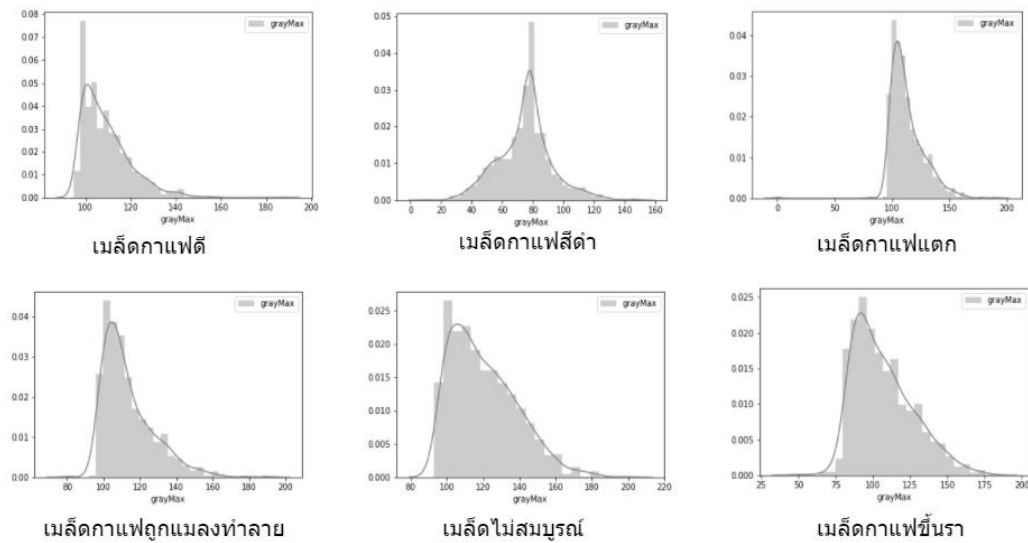
$$\sigma^2 = \frac{\sum(X - \bar{X})^2}{N} \quad (16)$$



ภาพประกอบ 30 ภาพฮิสโตแกรมการหาค่าความแปรปรวนของข้อมูล (Variance) ของสีแดง สีเขียว สีน้ำเงิน และสีเทา

3.3.3 กลุ่มของค่าสีเทา

จำนวน 4 คุณลักษณะ ได้แก่ ค่ามากของสีเทา ค่าเฉลี่ยของสีเทา ค่าเบี่ยงเบนมาตรฐานของ สีเทา และ ค่าความแปรปรวนของสีเทา โดยหาได้จากสมการที่ (14), (15) และสมการที่ (16) ตามลำดับ ค่าสี Gray โดยที่การแปลงภาพสีแบบ RGB เป็นภาพระดับสีเทา (Gray-scale image) เป็นการปรับให้ภาพแสดงถึงค่าความสว่าง (Brightness) ของภาพ



ภาพประกอบ 31 ภาพฮิสโตแกรมการหาค่ามาก (Max) ของสีเทา

เมื่อได้คุณลักษณะเฉพาะทั้งหมดแล้ว ข้อมูลจะถูกดึงเก็บในฐานข้อมูล โดยผู้วิจัยสามารถดึงฐานข้อมูลคุณลักษณะทั้งหมดที่ถูกจัดเก็บออกมาในรูปแบบของไฟล์ CSV เพื่อนำมาใช้วิเคราะห์และสร้างแบบจำลองโมเดลต่อไป โดยสามารถแสดงตัวอย่างของไฟล์ CSV ดังภาพประกอบ 32

D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
type_Coffee	type_no	solidity	e_diamator	contourArea	extent	bMax	gMax	rMax	bAvg	rAvg	gAvg	bStd	rStd	gStd	bVar
blackBean	1	0.986853769	204.488956	32842	0.783500728	56	59	55	1.406183649	1.362573657	1.677887253	5.318045853	5.130562922	6.288756702	28.28161169
blackBean	1	0.982159837	203.5356442	32536	0.784030941	66	69	61	7.738861177	6.49237331	8.572881274	10.13787342	8.5784936	11.25362053	102.7764775
blackBean	1	0.976880276	213.0232241	35640	0.7136664	59	53	39	5.722767321	4.436403684	5.997196636	9.076249679	6.92410805	9.336826213	82.37830824
blackBean	1	0.987088912	208.1286121	34021	0.810498856	69	74	45	5.382456642	4.430507909	6.286520869	10.48657336	8.521418451	12.18878545	109.9682209
blackBean	1	0.989180684	253.8351575	50605	0.807715636	88	86	75	11.2373747	7.490343485	12.82275107	21.79549646	13.97974495	23.77927825	475.043666
blackBean	1	0.985748219	222.2720797	38802	0.721249466	81	82	75	8.94676481	6.731593524	8.836595476	12.68450041	9.565279244	12.51296212	160.8965508
blackBean	1	0.98293328	193.954889	29545	0.821073255	74	82	74	7.570642508	5.787683415	7.848682748	11.94281341	9.055501604	12.29628925	142.6307922
blackBean	1	0.986325488	241.1077239	45657	0.750785194	84	87	74	17.34232812	12.32769309	18.62845115	21.22375531	14.98833718	22.31349502	450.4477893
blackBean	1	0.981167671	237.9449084	44487	0.798396653	89	91	84	13.15968831	9.492692473	12.67302858	14.68785614	10.82997053	14.495345	215.7331179

ภาพประกอบ 32 ตัวอย่างข้อมูลที่ถูกจัดเก็บในรูปแบบของไฟล์ CSV

3.4 การวิเคราะห์คุณลักษณะ (Feature Analysis)

หลังจากได้ข้อมูลในรูปแบบของไฟล์ CSV แล้วดังคำอธิบายในหัวข้อข้างต้นแล้ว ผู้วิจัยจึงนำข้อมูลของเมล็ดกาแฟทั้งหมดทั้ง 6 ประเภท มาทำการแยกประเภทตามคุณภาพเมล็ดกาแฟ จำนวน 2 ประเภท อีกครั้งหนึ่งคือ ประเภทเมล็ดกาแฟดี และประเภทเมล็ดกาแฟเสีย โดยทำการจำแนกประเภทของเมล็ดกาแฟ ดังนี้

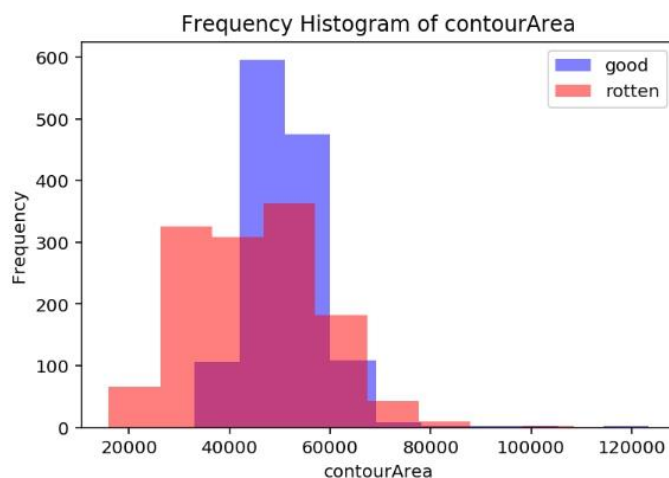
- เมล็ดกาแฟดี ได้แก่ เมล็ดกาแฟดี

- เมล็ดกาแฟเสีย ได้แก่ เมล็ดดำ เมล็ดขึ้นรา เมล็ดแตก เมล็ดไม่สมบูรณ์ เมล็ดถูกแมลงทำลาย

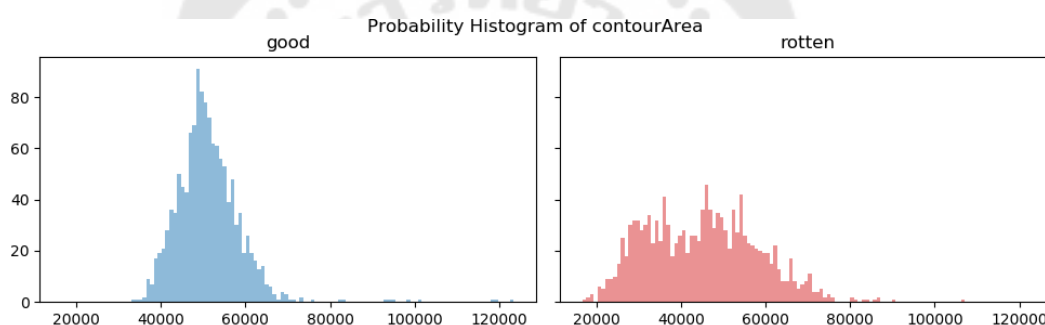
ผู้วิจัยได้ทำการแยกข้อมูลของเมล็ดกาแฟดีไว้ จากนั้นจึงทำการสุ่มสลับค่าข้อมูลตัวอย่างของเมล็ดเสียทั้ง 5 ประเภท ที่เหลือคือ เมล็ดดำ เมล็ดขึ้นรา เมล็ดแตก เมล็ดไม่สมบูรณ์ เมล็ดถูกแมลงทำลาย เนื่องจากจำนวนตัวอย่างของเมล็ดกาแฟเสียทั้งหมดนั้น มีจำนวนมากกว่าประเภทเมล็ดกาแฟดี หากนำข้อมูลของเมล็ดกาแฟเสียทั้งหมดมาใช้วิเคราะห์และเปรียบเทียบกับประเภทเมล็ดกาแฟดีแล้ว อาจทำให้เกิดปัญหาของข้อมูลที่ไม่สมดุลกันได้ คือการที่ข้อมูลมีจำนวนข้อมูลในคลาสที่แตกต่างกันมาก (Imbalance data) ซึ่งในการจำแนกกลุ่มข้อมูลที่มีข้อมูลมากและกลุ่มข้อมูลน้อยนั้น อาจส่งผลให้ผลลัพธ์ เอนเอียงไปทางกลุ่มข้อมูลที่มีจำนวนมากกว่าได้ ซึ่งทำให้ไม่สามารถจำแนกประเภทตามคุณสมบัติเมล็ดกาแฟตามที่ต้องการได้อย่างมีประสิทธิภาพ ผู้วิจัยจึงจัดปัญหานี้ด้วยการสุ่มหยิบเมล็ดกาแฟในแต่ละประเภทของเมล็ดกาแฟเสียออกมาอย่างละ 260 ตัวอย่าง รวมทั้ง 5 ประเภทของเมล็ดกาแฟเสีย ซึ่งจะได้ตัวอย่างประเภทของเมล็ดกาแฟเสีย จำนวนรวมทั้งสิ้น 1,300 ตัวอย่าง เพื่อให้จำนวนข้อมูลตัวอย่างของเมล็ดกาแฟทั้ง 2 ประเภทนั้น มีจำนวนข้อมูลที่ใกล้เคียงกัน

เมื่อผู้วิจัยได้จำนวนข้อมูลตัวอย่างตามที่ต้องการแล้ว จึงทำการสกัดคุณลักษณะ (Feature Extraction) เพื่อนำมาวิเคราะห์คุณลักษณะ ให้สามารถจำแนกคุณภาพของเมล็ดกาแฟ ซึ่งสามารถนำคุณลักษณะพิเศษทั้งหมด 10 ชนิด คือ อัตราส่วนของพื้นที่รูปร่างต่อพื้นที่เปลือก, เส้นผ่านศูนย์กลาง, ขนาดพื้นที่เมล็ด, ค่าสีแบบ RGB, ค่าสีแบบ HSV, ค่าสี Gray, ค่ามากของ RGB/HSV/Gray, ค่าเฉลี่ยของ RGB/HSV/Gray, ค่าความแปรปรวนของข้อมูลของ RGB/HSV/Gray และค่าเบี่ยงเบนมาตรฐาน RGB/HSV/Gray ของเมล็ดกาแฟ 2 ประเภท มาจำแนกคุณลักษณะพิเศษ ซึ่งสามารถหาได้ทั้งหมด 32 คุณลักษณะ ดังนี้

1. พื้นที่ของเมล็ดกาแฟ (Contour Area) คือ ค่าของขนาดพื้นที่ของเมล็ดกาแฟ นับรวมตั้งแต่บริเวณที่เป็นขอบเมล็ดจนถึงบริเวณของเมล็ดกาแฟทั้งหมดที่ภาพถ่ายสามารถจับได้จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าเฉลี่ยอยู่ที่ 33,000–70,000 เมล็ดกาแฟเสียมีค่าเฉลี่ยอยู่ที่ระหว่าง 18,000–80,000 โดยที่สีน้ำเงินแสดงค่าความถี่ของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมรูปทรงปกติ (Normal Histogram) ที่มีค่าเฉลี่ยอยู่ตรงกลาง มีลักษณะสมมาตร เมล็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ซึ่งยังถือว่าไม่สามารถจำแนกประเภทออกจากกันอย่างชัดเจนเท่าไรนัก

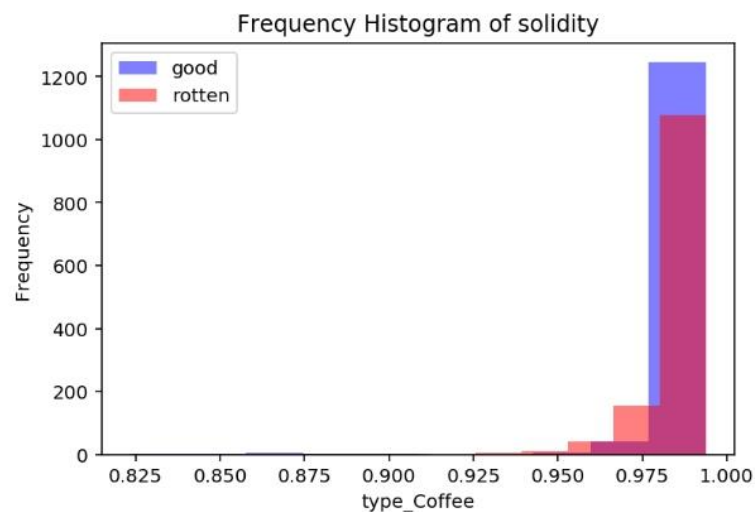


ภาพประกอบ 33 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ ContourArea ของเมล็ดกาแฟทั้ง 2 ประเภท

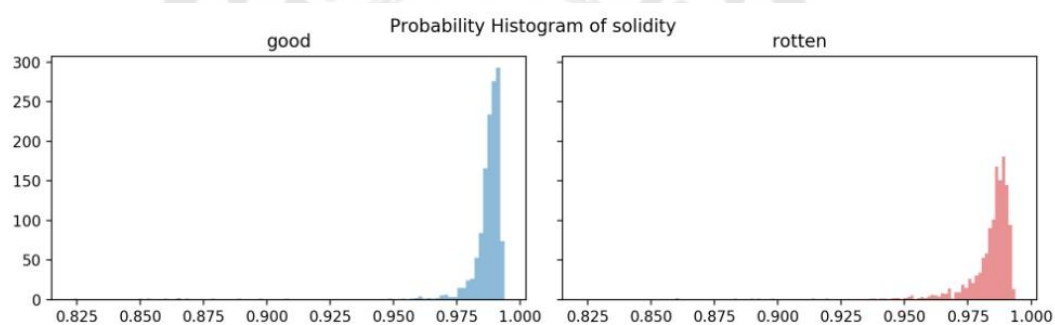


ภาพประกอบ 34 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ ContourArea ของเมล็ดกาแฟทั้ง 2 ประเภท

2. ค่าจากค่าอัตราส่วนของพื้นที่รูปร่างต่อพื้นที่เปลือก (Solidity) เป็นคุณลักษณะที่สามารถใช้ดูเมล็ดกาแฟนั้นเต็มเมล็ดหรือไม่ ถ้าไม่เต็มเมล็ดสามารถบ่งบอกได้ว่าเมล็ดนั้นอาจเป็นเมล็ดกาแฟเสีย ซึ่งสามารถนำคุณลักษณะนี้ไปใช้กับกาแฟประเภทเมล็ดกาแฟแตก หรือเมล็ดกาแฟที่ขาดแหว่งจากการกัดแทะของแมลงได้ โดยที่ความถี่ของคุณลักษณะ Solidity จะอยู่ที่ 0.950 – 0.985 แต่เมื่อนับรวมทั้งหมดของเมล็ดกาแฟเสียแล้ว คุณลักษณะ Solidity ก็ยังไม่สามารถที่จะจำแนกได้ผลลัพธ์ที่ดีนัก สังเกตได้จากภาพฮิสโตแกรมที่เป็นลักษณะรูปทรงถูกตัดลักษณะของกราฟคล้ายรูประฆังที่โดนตัดออกไปด้านหนึ่ง

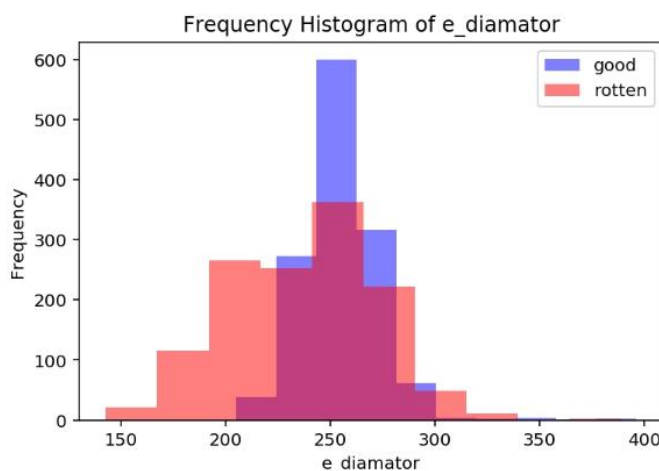


ภาพประกอบ 35 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ Solidity ของเมล็ดกาแฟทั้ง 2 ประเภท

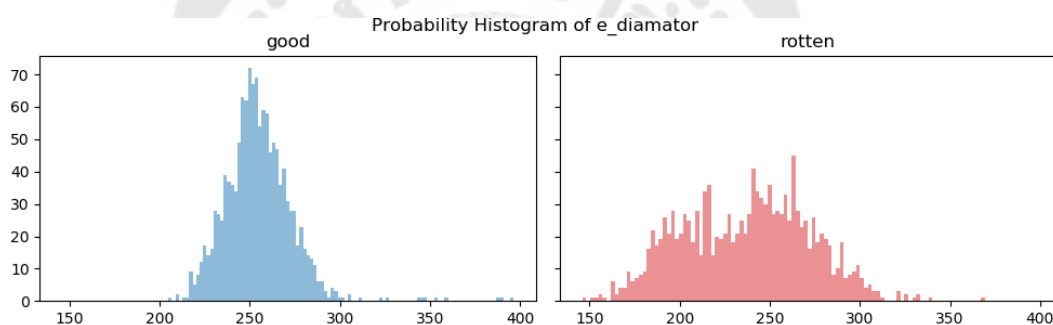


ภาพประกอบ 36 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ Solidity ของเมล็ดกาแฟทั้ง 2 ประเภท

3. เส้นผ่านศูนย์กลาง (Equivalent Diameter) เส้นผ่านศูนย์กลางสมมูลนี้สามารถใช้กำหนดขนาด (Size) ของวัสดุที่มีรูปร่างไม่เป็นทรงกลม ใช้เพื่อบอกขนาด ค่าเส้นผ่านศูนย์กลางยิ่งมากแสดงว่าขนาดของเมล็ดกาแฟนั้นมีขนาดใหญ่ ถ้าค่าน้อยก็แสดงว่าเมล็ดกาแฟนั้นอาจมีขนาดเล็ก จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าร้อยละเฉลี่ยอยู่ที่ 210-300 เมล็ดกาแฟเสียมีค่าเฉลี่ยอยู่ที่ระหว่าง 140-340 โดยที่สีน้ำเงินแสดงค่าความถี่ของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่ ถ้ามองภาพรวมแล้วสามารถจำแนกประเภทเมล็ดกาแฟทั้ง 2 ประเภทออกจากกันได้ด้วยคุณลักษณะของ E_diamator ประมาณ 40-50 %

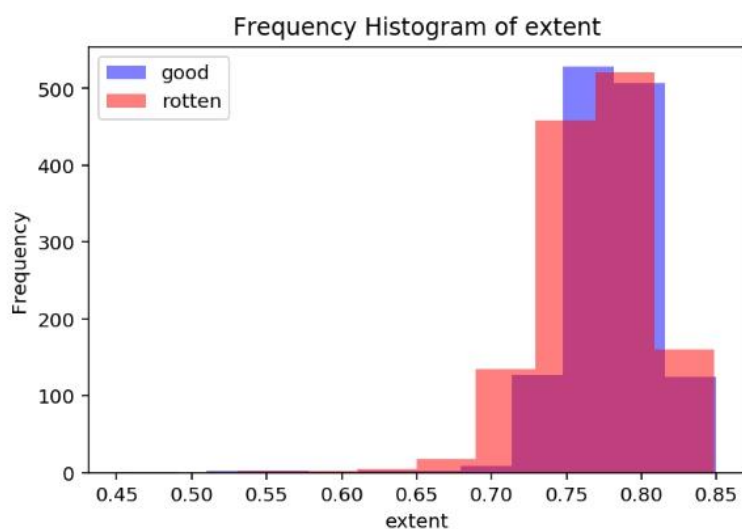


ภาพประกอบ 37 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ E_diamator ของเมล็ดกาแฟทั้ง 2 ประเภท

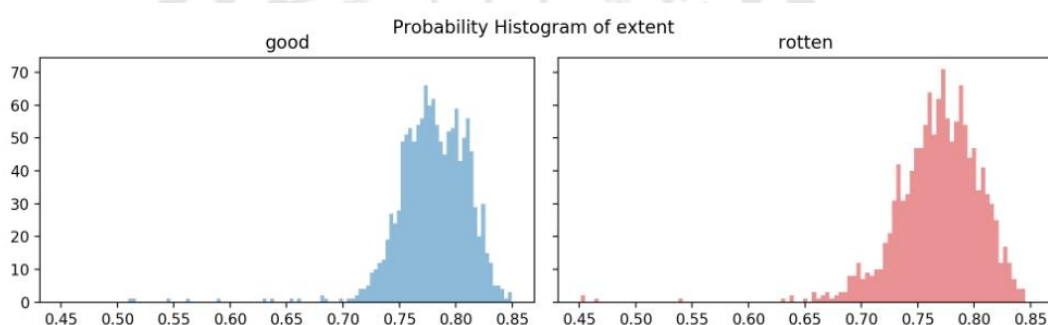


ภาพประกอบ 38 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ E_diamator ของเมล็ดกาแฟทั้ง 2 ประเภท

4. ค่าสัดส่วนของพื้นที่วงกลม (Extent) เป็นค่าเรโซลิวต์ส่วนของพื้นที่วงกลม ของ contour โดยนับพื้นที่ส่วนหนึ่งที่เทียบกับพื้นที่ทั้งหมดของเมล็ดกาแฟ และแปลงเป็นค่าร้อยละ ซึ่งคุณลักษณะ Extent นี้ เป็นคุณลักษณะที่สำคัญ ที่ใช้เป็นฐานตั้งต้นในการหาคุณลักษณะอื่น ๆ ด้วย จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าร้อยละเฉลี่ยอยู่ที่ 0.68-0.85 เมล็ดกาแฟเสียมีค่าร้อยละเฉลี่ยอยู่ที่ระหว่าง 0.65-0.84 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่ ดังภาพประกอบ 39 ถ้ามองภาพรวมแล้วไม่สามารถจำแนกประเภทเมล็ดกาแฟทั้ง 2 ประเภทออกจากกันได้ด้วยคุณลักษณะของ Extent แต่หากมองโดยความละเอียดแล้วจะพบว่าในช่วงความถี่ระหว่างร้อยละ 0.65-0.75 นั้น ส่วนหนึ่งไม่พบค่าของเมล็ดกาแฟดีอยู่ ซึ่งค่าร้อยละในส่วนนี้อาจจะสามารถจำแนกประเภทเมล็ดกาแฟทั้ง 2 ประเภทออกจากกันได้ประมาณเล็กน้อยเท่านั้น

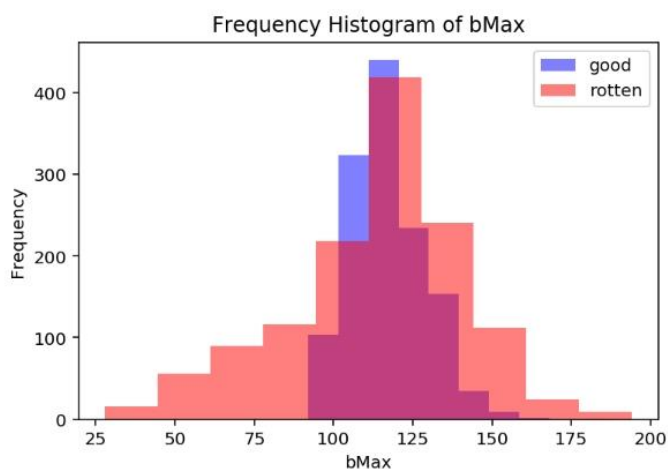


ภาพประกอบ 39 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ extent ของเมล็ดกาแฟทั้ง 2 ประเภท

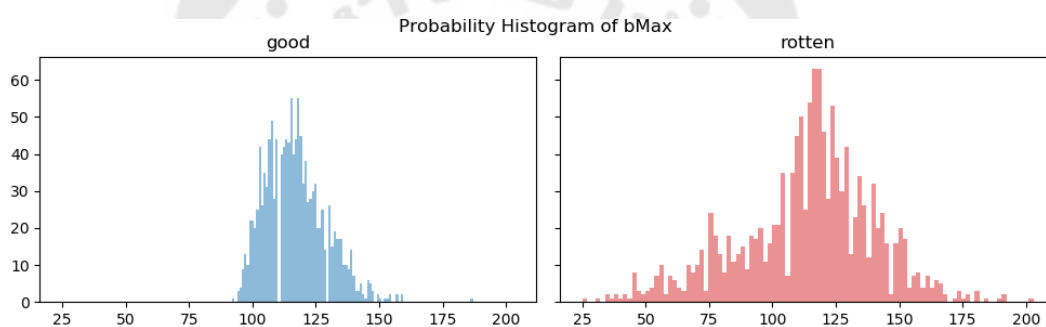


ภาพประกอบ 40 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ extent ของเมล็ดกาแฟทั้ง 2 ประเภท

5. bMax (Blue Color Max) เป็นระบบสีที่เกิดจากการรวมกันของแสงสีน้ำเงิน โดยคุณลักษณะนี้จะเป็นค่ามากที่สุดโดยเฉลี่ยบนเมล็ดกาแฟแต่ละเมล็ด จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าร้อยละเฉลี่ยอยู่ที่ 95-160 เมล็ดกาแฟเสียมีค่าร้อยละเฉลี่ยอยู่ที่ระหว่าง 30-195 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมรูปทรงธรรมชาติ (Normal histogram) ภาพประกอบ 41 ที่มีค่าเฉลี่ยอยู่ตรงกลาง มีลักษณะสมมาตร เมล็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ยังไม่สามารถจำแนกประเภทออกจากกันได้

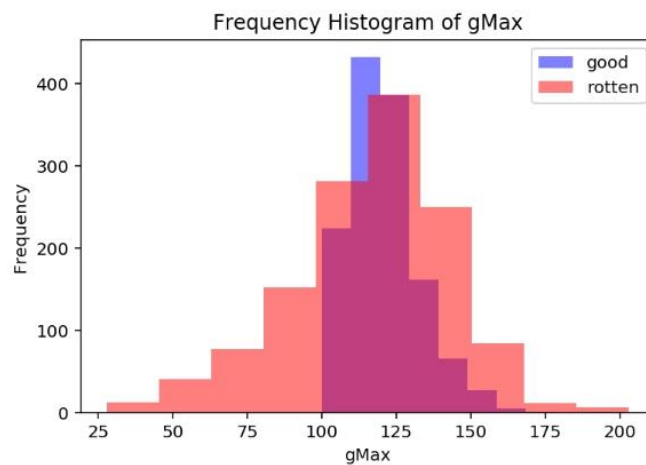


ภาพประกอบ 41 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ bMax ของเมล็ดกาแฟทั้ง 2 ประเภท

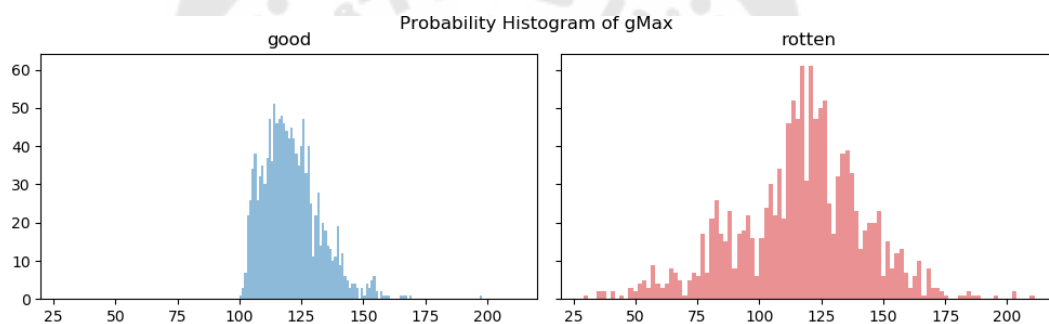


ภาพประกอบ 42 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ bMax ของเมล็ดกาแฟทั้ง 2 ประเภท

6. gMax (Green Color Max) เป็นระบบสีที่เกิดจากการรวมกันของแสงสีเขียว โดยคุณลักษณะนี้จะเป็นค่ามากที่สุดโดยเฉลี่ยบนเมล็ดกาแฟแต่ละเมล็ด จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าร้อยละเฉลี่ยอยู่ที่ 100-170 เมล็ดกาแฟเสียมีค่าร้อยละเฉลี่ยอยู่ที่ระหว่าง 30-205 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมรูปธรรมชาติ (Normal histogram) ภาพประกอบ 43 ที่มีค่าเฉลี่ยอยู่ตรงกลาง มีลักษณะสมมาตร เมล็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ยังไม่สามารถจำแนกประเภทออกจากกันได้

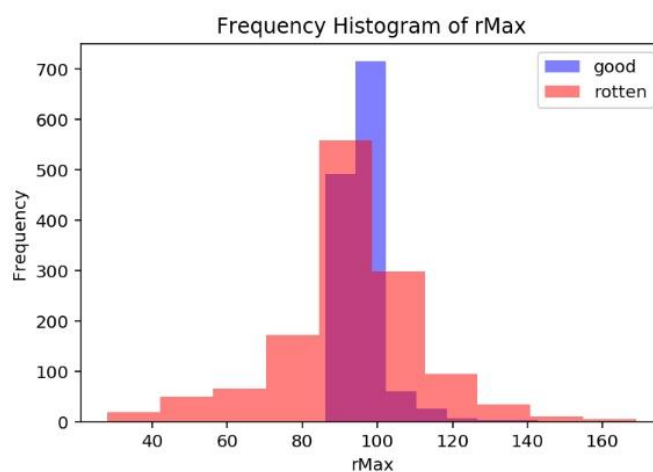


ภาพประกอบ 43 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ gMax ของเมล็ดกาแฟทั้ง 2 ประเภท

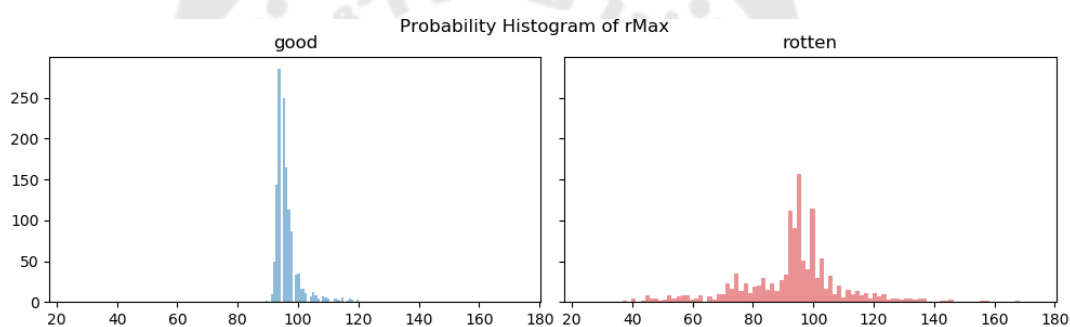


ภาพประกอบ 44 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ gMax ของเมล็ดกาแฟทั้ง 2 ประเภท

7. rMax (Red Color Max) เป็นระบบสีที่เกิดจากการรวมกันของแสงสีแดง โดยคุณลักษณะนี้จะเป็นค่ามากที่สุดโดยเฉลี่ยบนเมล็ดกาแฟแต่ละเมล็ด จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าร้อยละเฉลี่ยอยู่ที่ 100-170 เมล็ดกาแฟเสียมีค่าร้อยละเฉลี่ยอยู่ที่ระหว่าง 30-205 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมรูปธรรมชาติ (Normal Histogram) ภาพประกอบ 45 เมล็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ค่า rMax นี้สามารถจำแนกประเภทเมล็ดกาแฟดีและประเภทเมล็ดกาแฟเสียออกจากกันได้บางส่วน

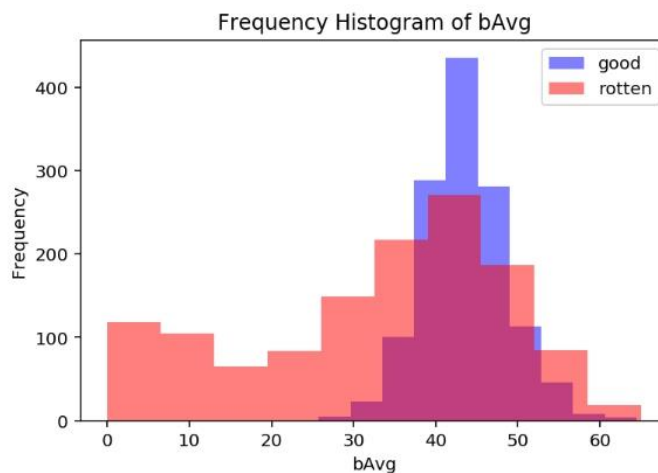


ภาพประกอบ 45 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ rMax ของเมลิ็ดกาแพทั้ง 2 ประเภท

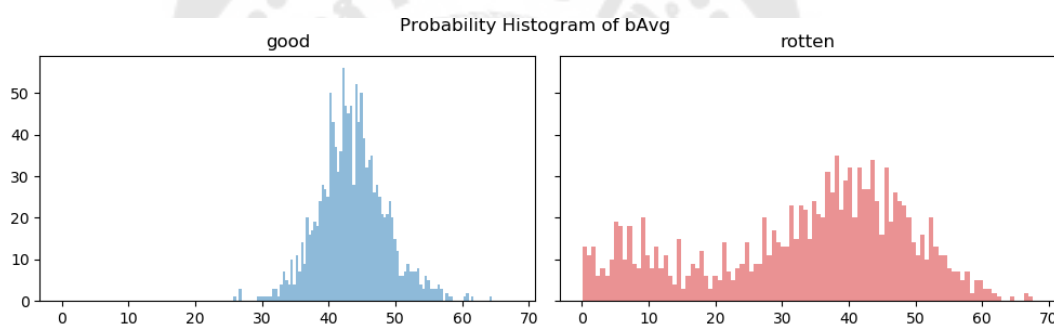


ภาพประกอบ 46 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ rMax ของเมลิ็ดกาแพทั้ง 2 ประเภท

8. bAvg (Blue Color Average) เป็นระบบสีที่เกิดจากการรวมกันของแสงสีน้ำเงิน โดยคุณลักษณะนี้จะเป็นค่าเฉลี่ยของสีน้ำเงินทั้งหมดเมลิ็ดกาแพแต่ละเมลิ็ด จากค่าความถี่ของเมลิ็ดกาแพทั้ง 2 ประเภท จะพบว่ากาแพทั้ง 2 ประเภทนั้นเมลิ็ดกาแพดีมีค่าเฉลี่ยอยู่ที่ 26-65 เมลิ็ดกาแพเสียมีค่าเฉลี่ยอยู่ที่ระหว่าง 0-67 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมลิ็ดกาแพดี และสีแดงแสดงค่าความถี่ของเมลิ็ดกาแพเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมลิ็ดกาแพทั้ง 2 ประเภท ดังฮิสโตแกรมเมลิ็ดกาแพทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ค่า bAvg นี้ สามารถจำแนกประเภทเมลิ็ดกาแพดีและประเภทเมลิ็ดกาแพเสียออกจากกันได้บางส่วน จากค่าเฉลี่ย 0-25 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันได้ชัดเจน

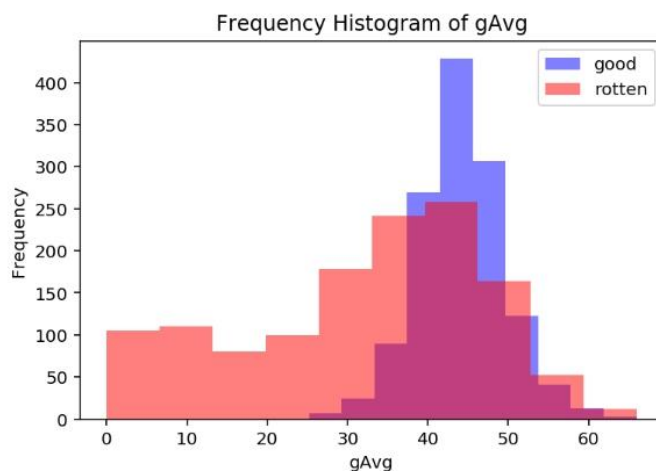


ภาพประกอบ 47 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ bAvg ของเมลิ็ดกาแพทั้ง 2 ประเภท

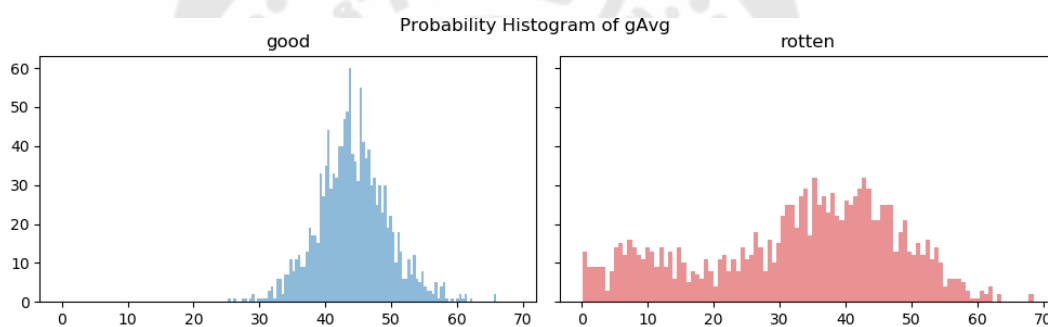


ภาพประกอบ 48 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ bAvg ของเมลิ็ดกาแพทั้ง 2 ประเภท

9. gAvg (Green Color Average) เป็นระบบสีที่เกิดจากการรวมกันของแสงสีเขียว โดยคุณลักษณะนี้จะเป็นค่าเฉลี่ยของสีเขียวทั้งหมดเมลิ็ดกาแพแต่ละเมลิ็ด จากค่าความถี่ของเมลิ็ดกาแพทั้ง 2 ประเภท จะพบว่ากาแพทั้ง 2 ประเภทนั้นเมลิ็ดกาแพดีมีค่าเฉลี่ยอยู่ที่ 25-65 เมลิ็ดกาแพเสียมีค่าเฉลี่ยอยู่ที่ระหว่าง 0-67 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมลิ็ดกาแพดี และสีแดงแสดงค่าความถี่ของเมลิ็ดกาแพเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมลิ็ดกาแพทั้ง 2 ประเภท ดังฮิสโตแกรมเมลิ็ดกาแพทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ค่า gAvg นี้ สามารถจำแนกประเภทเมลิ็ดกาแพดีและประเภทเมลิ็ดกาแพเสียออกจากกันได้บางส่วน จากค่าเฉลี่ย 0-24 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันได้อย่างชัดเจน

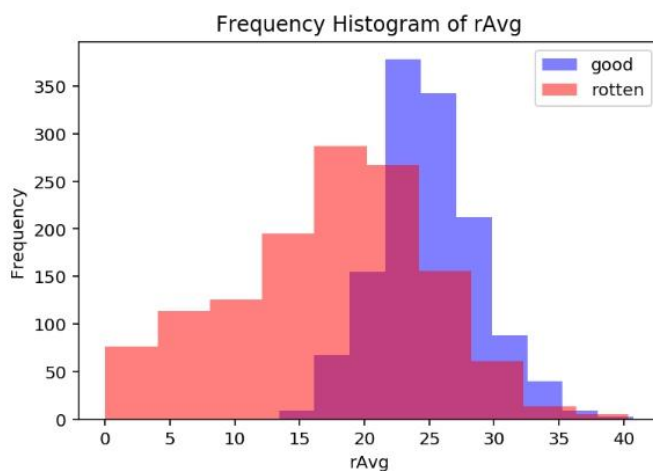


ภาพประกอบ 49 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ gAvg ของเมลิคกาแฟทั้ง 2 ประเภท

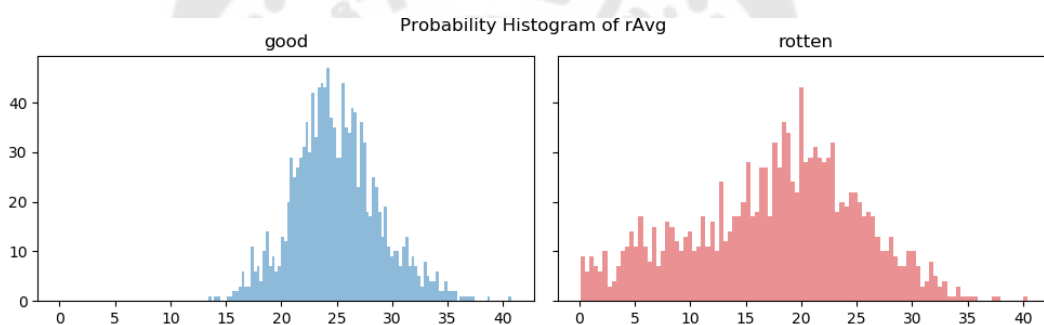


ภาพประกอบ 50 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ gAvg ของเมลิคกาแฟทั้ง 2 ประเภท

10. rAvg (Red Color Average) เป็นระบบสีที่เกิดจากการรวมกันของแสงสีแดง โดยคุณลักษณะนี้จะเป็นค่าเฉลี่ยของสีแดงทั้งหมดเมลิคกาแฟแต่ละเมล็ด จากค่าความถี่ของเมลิคกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมลิคกาแฟดีมีค่าเฉลี่ยอยู่ที่ 13-38 เมลิคกาแฟเสียมีค่าเฉลี่ยอยู่ที่ระหว่าง 0-40 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมลิคกาแฟดี และสีแดงแสดงค่าความถี่ของเมลิคกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมลิคกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมลิคกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ค่า rAvg นี้ สามารถจำแนกประเภทเมลิคกาแฟดีและประเภทเมลิคกาแฟเสียออกจากกันได้บางส่วน จากค่าเฉลี่ย 0-12 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันได้อย่างชัดเจน

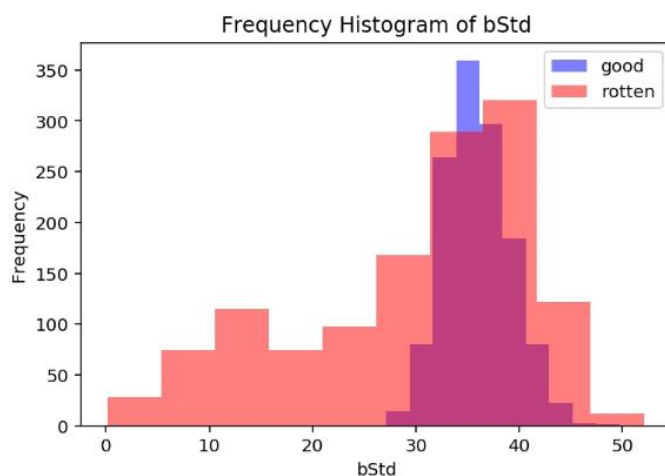


ภาพประกอบ 51 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ rAvg ของเมลิ็ดกาแพทั้ง 2 ประเภท

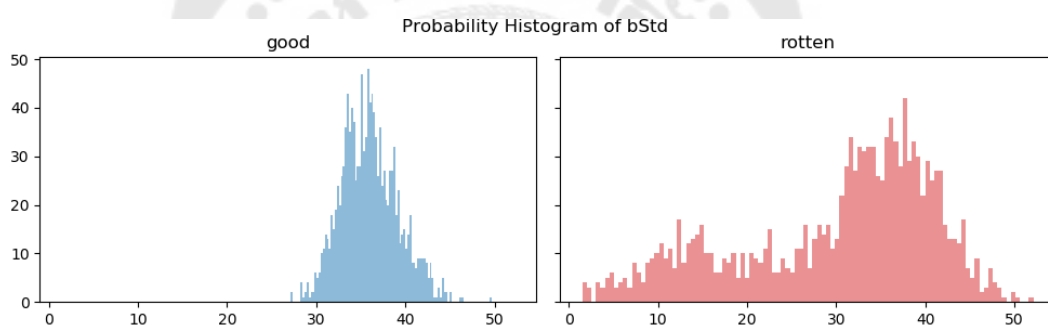


ภาพประกอบ 52 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ rAvg ของเมลิ็ดกาแพทั้ง 2 ประเภท

11. bStd (Blue Color Standard Deviation) เป็นค่าเบี่ยงเบนมาตรฐานของสีน้ำเงิน โดยที่ ค่า SD จะแสดงให้เห็นว่าโดยเฉลี่ยแล้วข้อมูลแต่ละตัวนั้นแตกต่างจากค่าเฉลี่ยเท่าไร จากค่าความถี่ของเมลิ็ดกาแพทั้ง 2 ประเภท จะพบว่ากาแพทั้ง 2 ประเภทนั้นเมลิ็ดกาแพดีมีค่าเฉลี่ยอยู่ที่ 27-50 เมลิ็ดกาแพเสียมีค่าเฉลี่ยอยู่ที่ระหว่าง 0-52 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมลิ็ดกาแพดี และสีแดงแสดงค่าความถี่ของเมลิ็ดกาแพเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมลิ็ดกาแพทั้ง 2 ประเภท ดังฮิสโตแกรมเมลิ็ดกาแพทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ค่า bStd นี้ สามารถจำแนกประเภทเมลิ็ดกาแพดีและประเภทเมลิ็ดกาแพเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-26 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันได้อย่างชัดเจน

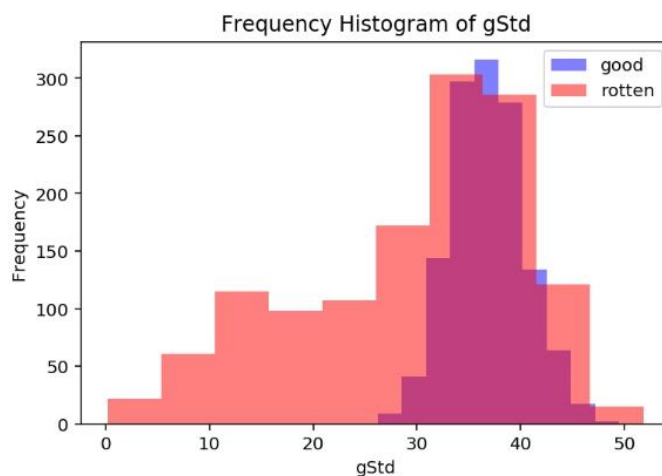


ภาพประกอบ 53 ฮิสโตแกรมแสดงความถี่คุณลักษณะ bStd ของเมล็ดกาแฟทั้ง 2 ประเภท

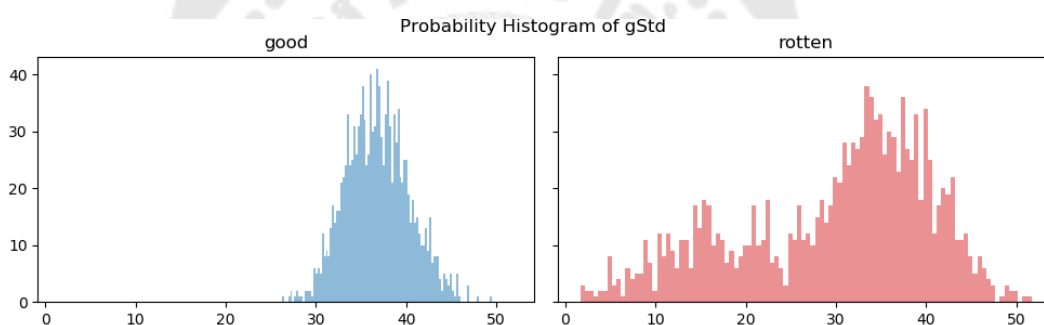


ภาพประกอบ 54 ฮิสโตแกรมแสดงความน่าจะเป็นคุณลักษณะ bStd ของเมล็ดกาแฟทั้ง 2 ประเภท

12. gStd (Green Color Standard Deviation) เป็นค่าเบี่ยงเบนมาตรฐานของสีเขียว โดยที่ ค่า SD จะแสดงให้เห็นว่าโดยเฉลี่ยแล้วข้อมูลแต่ละตัวนั้นแตกต่างจากค่าเฉลี่ยเท่าไร จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าเฉลี่ยอยู่ที่ 27-50 เมล็ดกาแฟเสียมีค่าเฉลี่ยอยู่ที่ระหว่าง 0-52 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมล็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ค่า gStd นี้ สามารถจำแนกประเภทเมล็ดกาแฟดีและประเภทเมล็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-26 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันอย่างชัดเจน จะเห็นได้ว่าคุณลักษณะ gStd นี้มีผลลัพธ์เหมือนกับคุณลักษณะ bStd

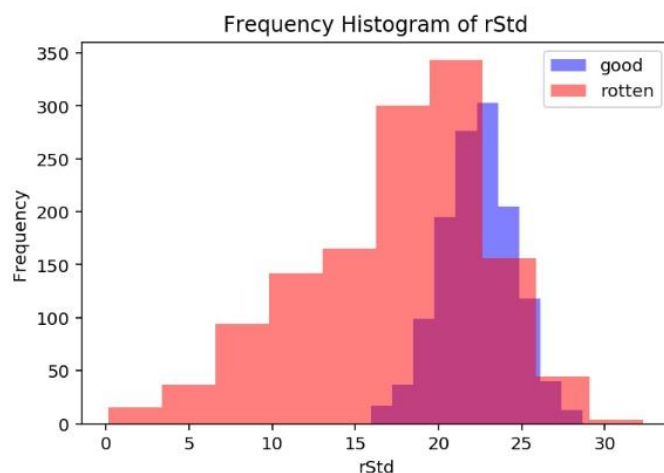


ภาพประกอบ 55 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ gStd ของเมล็ดกาแฟทั้ง 2 ประเภท

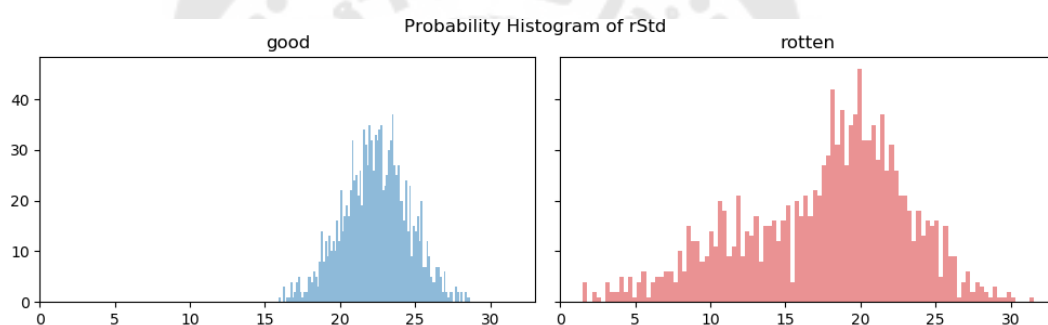


ภาพประกอบ 56 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ gStd ของเมล็ดกาแฟทั้ง 2 ประเภท

13. rStd (Red Color Standard Deviation) เป็นค่าเบี่ยงเบนมาตรฐานของสีแดง โดยที่ ค่า SD จะแสดงให้เห็นว่าโดยเฉลี่ยแล้วข้อมูลแต่ละตัวนั้นแตกต่างจากค่าเฉลี่ยเท่าไร จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าเฉลี่ยอยู่ที่ 16-28 เมล็ดกาแฟเสียมีค่าเฉลี่ยอยู่ที่ระหว่าง 0-33 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมล็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ค่า rStd นี้ สามารถจำแนกประเภทเมล็ดกาแฟดีและประเภทเมล็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-15 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันได้อย่างชัดเจน

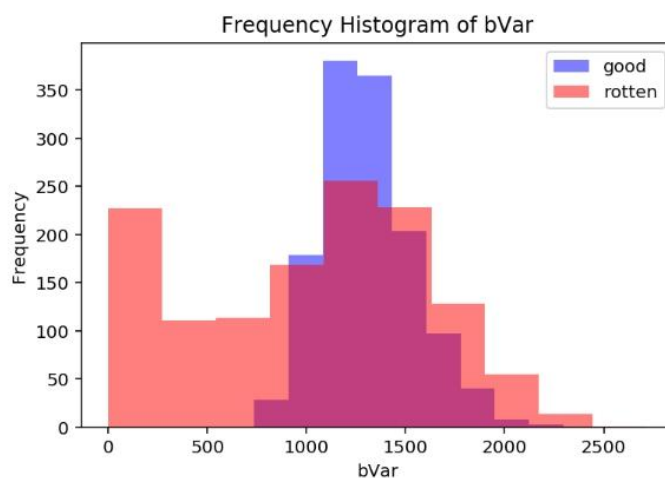


ภาพประกอบ 57 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ rStd ของเมลิ็ดกาแฟทั้ง 2 ประเภท

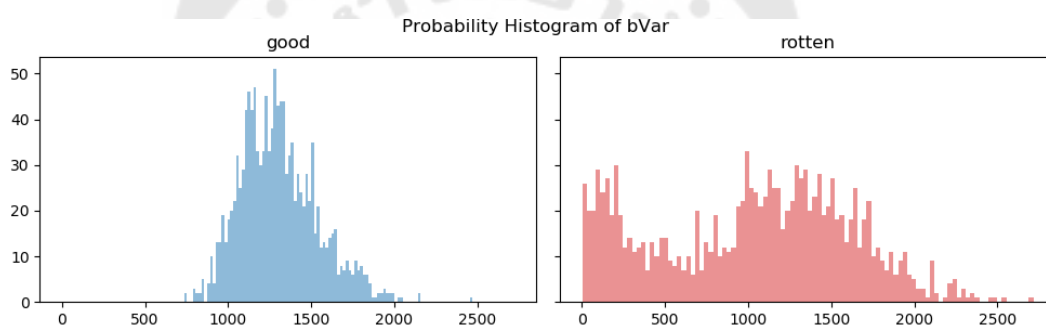


ภาพประกอบ 58 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ rStd ของเมลิ็ดกาแฟทั้ง 2 ประเภท

14. bVar (Blue Color Variance) ค่าความแปรปรวนของสีน้ำเงิน เป็นค่าที่ใช้วัดการกระจายของข้อมูล หาได้จากการนำส่วนเบี่ยงเบนมาตรฐานของค่าสีน้ำเงินบนเมลิ็ดกาแฟ จากค่าความถี่ของเมลิ็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมลิ็ดกาแฟดีมีค่าเฉลี่ยอยู่ที่ 750-2400 เมลิ็ดกาแฟเสียมีค่าเฉลี่ยอยู่ที่ระหว่าง 0-2600 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมลิ็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมลิ็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมลิ็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมลิ็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ค่า bVar นี้ สามารถจำแนกประเภทเมลิ็ดกาแฟดีและประเภทเมลิ็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-750 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันได้อย่างชัดเจน

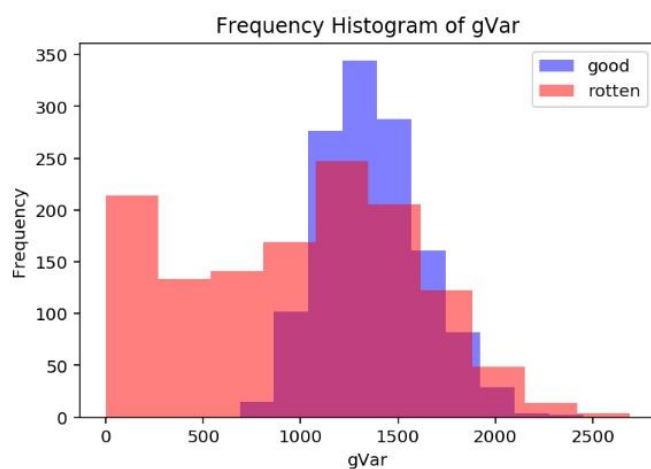


ภาพประกอบ 59 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ bVar ของเมล็ดกาแฟทั้ง 2 ประเภท

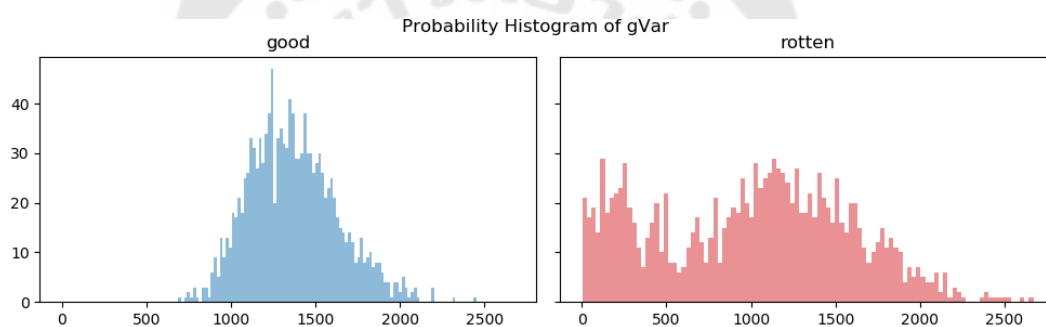


ภาพประกอบ 60 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ bVar ของเมล็ดกาแฟทั้ง 2 ประเภท

15. gVar (Green Color Variance) ค่าความแปรปรวนของสีเขียว เป็นค่าที่ใช้วัดการกระจายของข้อมูล หาได้จากการนำส่วนเบี่ยงเบนมาตรฐานของค่าสีเขียวบนเมล็ดกาแฟ จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าเฉลี่ยอยู่ที่ 700-2400 เมล็ดกาแฟเสียมีค่าเฉลี่ยอยู่ที่ระหว่าง 0-2800 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมล็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ค่า gVar นี้ สามารถจำแนกประเภทเมล็ดกาแฟดีและประเภทเมล็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-700 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันได้อย่างชัดเจน

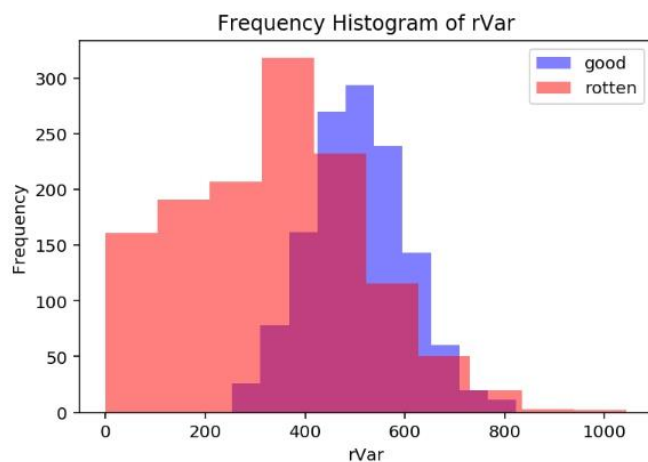


ภาพประกอบ 61 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ gVar ของเมลิ็ดกาแฟทั้ง 2 ประเภท

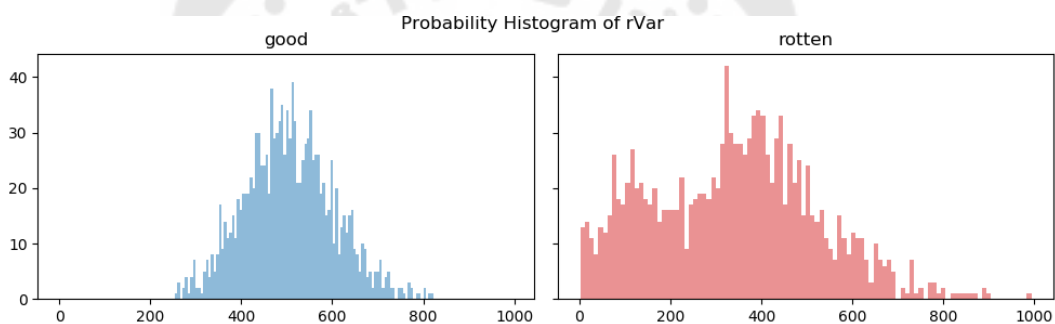


ภาพประกอบ 62 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ gVar ของเมลิ็ดกาแฟทั้ง 2 ประเภท

16. rVar (Red Color Variance) ค่าความแปรปรวนของสีแดง เป็นค่าที่ใช้วัดการกระจายของข้อมูล หาได้จากการนำส่วนเบี่ยงเบนมาตรฐานของค่าสีแดงบนเมลิ็ดกาแฟ จากค่าความถี่ของเมลิ็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมลิ็ดกาแฟดีมีค่าเฉลี่ยอยู่ที่ 250-825 เมลิ็ดกาแฟเสียมีค่าเฉลี่ยอยู่ที่ระหว่าง 0-1200 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมลิ็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมลิ็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมลิ็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมลิ็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ค่า rVar นี้ สามารถจำแนกประเภทเมลิ็ดกาแฟดีและประเภทเมลิ็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-250 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันอย่างชัดเจน

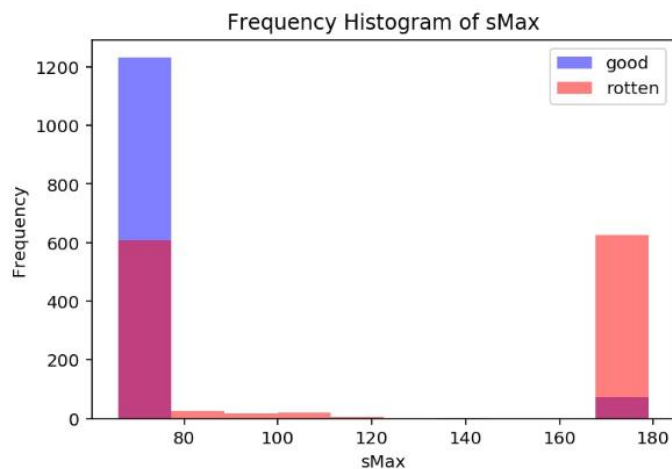


ภาพประกอบ 63 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ vVar ของเมลิ็ดกาแพทั้ง 2 ประเภท

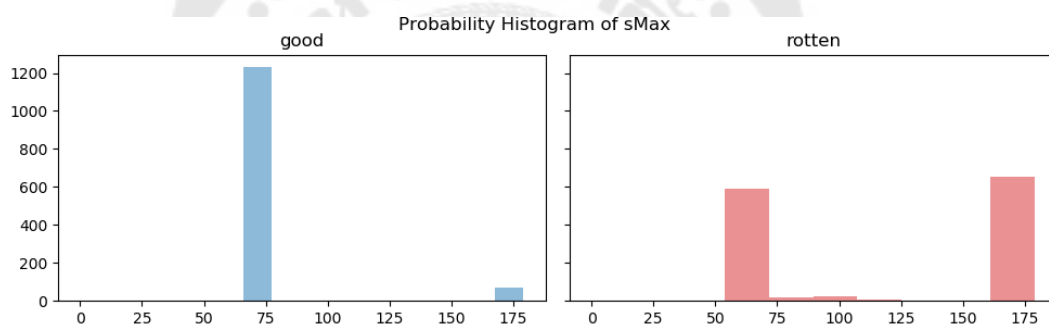


ภาพประกอบ 64 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ vVar ของเมลิ็ดกาแพทั้ง 2 ประเภท

17. sMax (Saturation Max) ค่าสี HSV (Hue Saturation Value) เป็นค่าสีของความอิ่มตัว ค่าของเฉดสี และค่าความสว่างของสี โดยใช้หลักการพิจารณาการแสดงค่าตัวเลขอยู่ระหว่าง 0 ถึง 255 ในคุณลักษณะ sMax นี้ จะเป็นค่ามากที่สุดของค่า Saturation จากค่าความถี่ของเมลิ็ดกาแพทั้ง 2 ประเภท จะพบว่ากาแพทั้ง 2 ประเภทนั้นเมลิ็ดกาแพดีมีค่าองศาเฉลี่ยอยู่ที่ 65-77 เป็นส่วนใหญ่ และบางส่วนอยู่ที่ 168-180 องศา และเมลิ็ดกาแพเสียมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 55-80 และ 170-180 ในปริมาณที่ไม่แตกต่างกันมากนัก โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมลิ็ดกาแพดี และสีแดงแสดงค่าความถี่ของเมลิ็ดกาแพเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมลิ็ดกาแพทั้ง 2 ประเภท ดังฮิสโตแกรมเมลิ็ดกาแพทั้ง 2 ประเภท นั้นซ้อนทับกันอยู่บางส่วน ค่า sMax นี้ ยังไม่สามารถจำแนกประเภทออกจากกันได้เท่าไรนัก

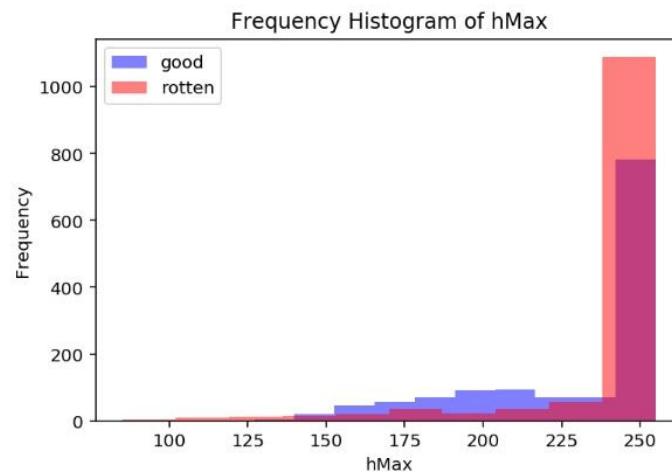


ภาพประกอบ 65 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ sMax ของเมลิ็ดกาแฟทั้ง 2 ประเภท

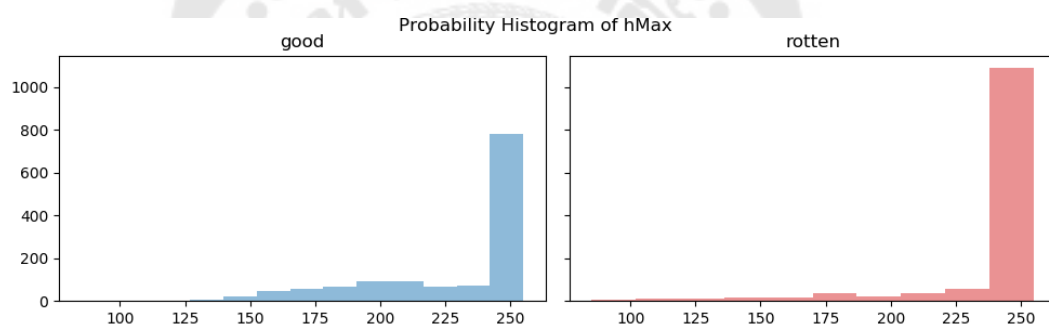


ภาพประกอบ 66 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ sMax ของเมลิ็ดกาแฟทั้ง 2 ประเภท

18. hMax (Hue Max) ค่าสี HSV (Hue Saturation Value) เป็นค่าสีของความเข้มตัวค่าของเฉดสี และค่าความสว่างของสี โดยใช้หลักการพิจารณาการแสดงค่าตัวเลขอยู่ระหว่าง 0 ถึง 255 ในคุณลักษณะ hMax นี้จะเป็นค่ามากที่สุดของค่า Hue จากค่าความถี่ของเมลิ็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมลิ็ดกาแฟดีมีค่าองศาเฉดสีอยู่ระหว่าง 125-255 และเมลิ็ดกาแฟเสียมีค่าองศาเฉดสีอยู่ระหว่าง 55-255 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมลิ็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมลิ็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมลิ็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมลิ็ดกาแฟทั้ง 2 ประเภท นั้นซ้อนทับกันเป็นส่วนใหญ่ค่า hMax นี้ ผู้วิจัยเห็นว่าไม่สามารถจำแนกประเภทออกจากกันได้

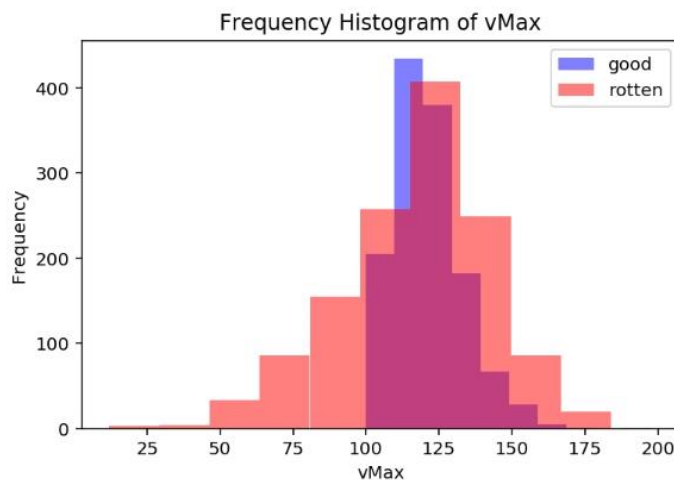


ภาพประกอบ 67 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ hMax ของเมลิ็ดกาแพทั้ง 2 ประเภท

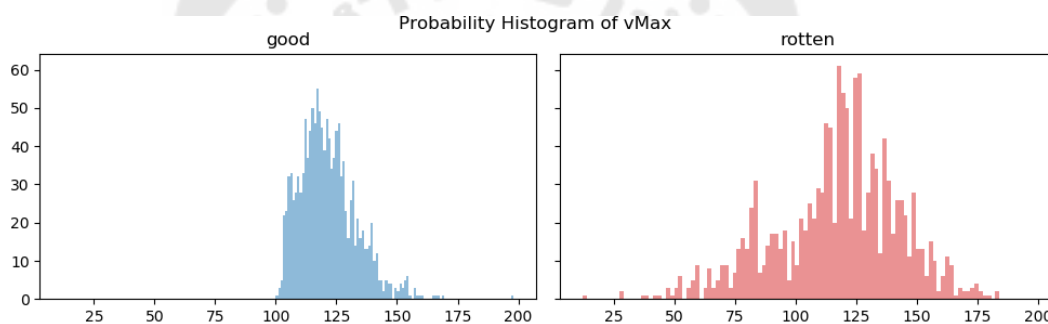


ภาพประกอบ 68 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ hMax ของเมลิ็ดกาแพทั้ง 2 ประเภท

19. vMax (Value Max) ค่าสี HSV (Hue Saturation Value) เป็นค่าสีของความอิ่มตัว ค่าของเฉดสี และค่าความสว่างของสี โดยใช้หลักการพิจารณาการแสดงค่าตัวเลขอยู่ระหว่าง 0 ถึง 255 ในคุณลักษณะ vMax นี้ จะเป็นค่ามากที่สุดของค่า Value จากค่าความถี่ของเมลิ็ดกาแพทั้ง 2 ประเภท จะพบว่ากาแพทั้ง 2 ประเภทนั้นเมลิ็ดกาแพดีมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 100 - 170 และเมลิ็ดกาแพเสียมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 10 - 185 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมลิ็ดกาแพดี และสีแดงแสดงค่าความถี่ของเมลิ็ดกาแพเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมลิ็ดกาแพทั้ง 2 ประเภท ดังฮิสโตแกรมเมลิ็ดกาแพทั้ง 2 ประเภท นั้นซ้อนทับกันเป็นส่วนใหญ่ค่า vMax นี้ สามารถจำแนกประเภทเมลิ็ดกาแพดีและประเภทเมลิ็ดกาแพเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 10-100 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันได้อย่างชัดเจน แต่ยังถือเป็นเพียงส่วนน้อยเท่านั้น

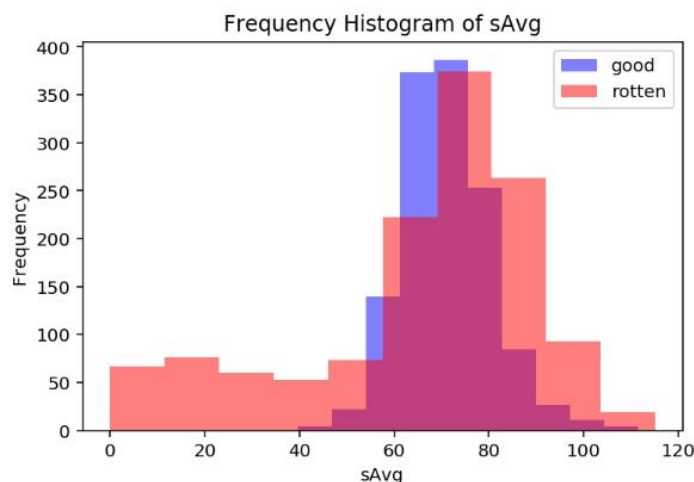


ภาพประกอบ 69 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ vMax ของเมล็ดกาแฟทั้ง 2 ประเภท

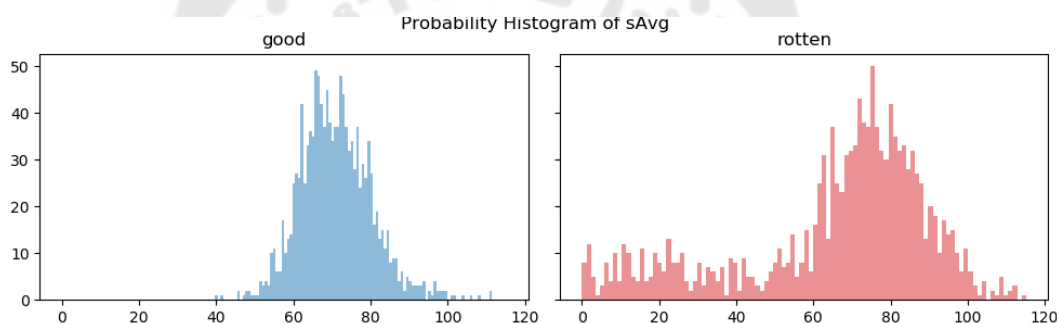


ภาพประกอบ 70 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ vMax ของเมล็ดกาแฟทั้ง 2 ประเภท

20. sAvg (Saturation Average) ค่าสี HSV (Hue Saturation Value) เป็นค่าสีของความอิ่มตัว ค่าของเฉดสี และค่าความสว่างของสี โดยใช้หลักการพิจารณาการแสดงค่าตัวเลขอยู่ระหว่าง 0 ถึง 255 ในคุณลักษณะ sAvg นี้จะเป็นค่าโดยเฉลี่ยของค่า Saturation จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 40-110 และเมล็ดกาแฟเสียมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 0-115 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมล็ดกาแฟทั้ง 2 ประเภท นั้นซ้อนทับกันเป็นบางส่วนค่า sAvg นี้สามารถจำแนกประเภทเมล็ดกาแฟดีและประเภทเมล็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-40 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันอย่างชัดเจน

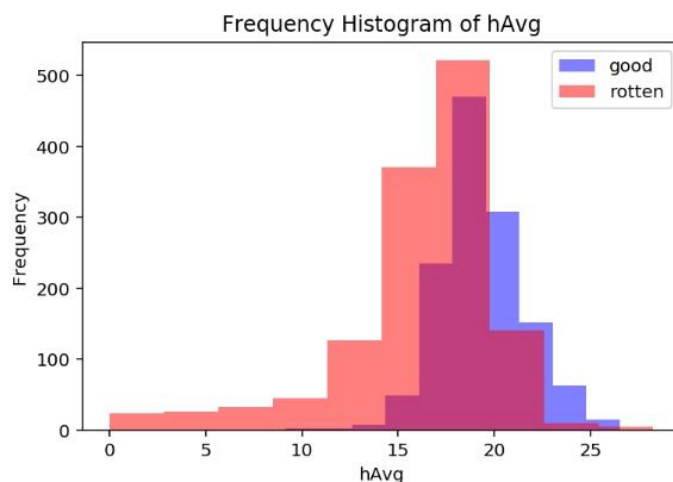


ภาพประกอบ 71 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ sStd ของเมลิ็ดกาแฟทั้ง 2 ประเภท

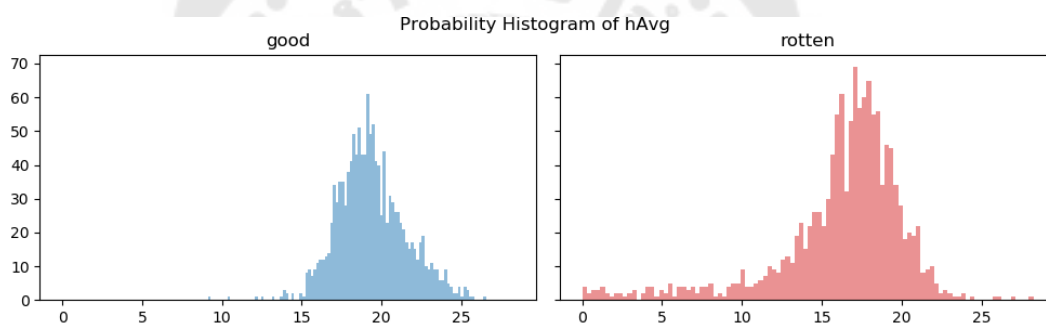


ภาพประกอบ 72 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ sStd ของเมลิ็ดกาแฟทั้ง 2 ประเภท

21. hAvg (Hue Average) ค่าสี HSV (Hue Saturation Value) เป็นค่าสีของความอิ่มตัว ค่าของเฉดสี และค่าความสว่างของสี โดยใช้หลักการพิจารณาการแสดงค่าตัวเลขอยู่ระหว่าง 0 ถึง 255 ในคุณลักษณะ hAvg นี้ จะเป็นค่าโดยเฉลี่ยของค่า Hue จากค่าความถี่ของเมลิ็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมลิ็ดกาแฟดีมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 10-26 และเมลิ็ดกาแฟเสียมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 0-28 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมลิ็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมลิ็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมลิ็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมลิ็ดกาแฟทั้ง 2 ประเภท นั้นซ้อนทับกันเป็นบางส่วนค่า sAvg นี้ สามารถจำแนกประเภทเมลิ็ดกาแฟดีและประเภทเมลิ็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-10 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันอย่างชัดเจน แต่ยังถือเป็นเพียงส่วนน้อยเท่านั้น

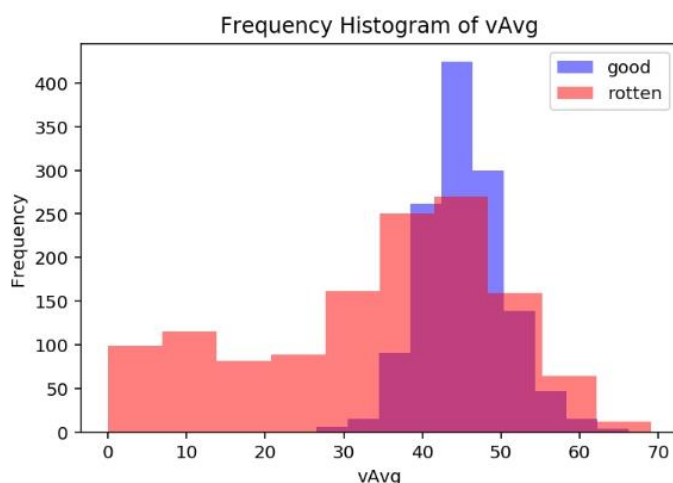


ภาพประกอบ 73 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ hAvg ของเมลิ็ดกาแพทั้ง 2 ประเภท

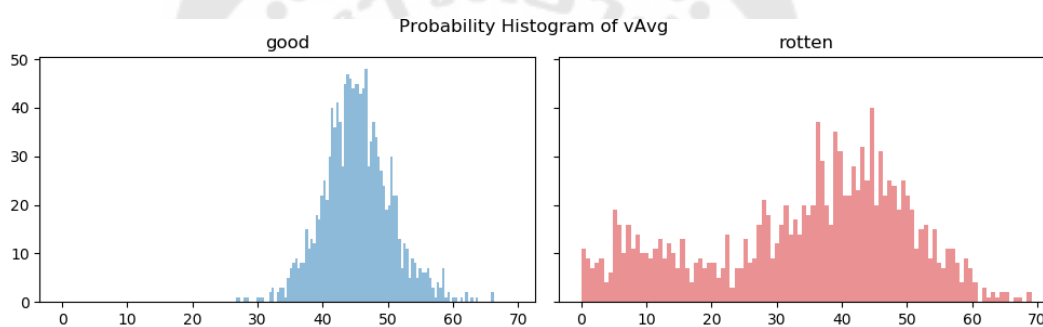


ภาพประกอบ 74 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ hAvg ของเมลิ็ดกาแพทั้ง 2 ประเภท

22. vAvg (Value Average) ค่าสี HSV (Hue Saturation Value) เป็นค่าสีของความเข้มตัว ค่าของเฉดสี และค่าความสว่างของสี โดยใช้หลักการพิจารณาการแสดงค่าตัวเลขอยู่ระหว่าง 0 ถึง 255 ในคุณลักษณะ vAvg นี้ จะเป็นค่าโดยเฉลี่ยของค่า Value จากค่าความถี่ของเมลิ็ดกาแพทั้ง 2 ประเภท จะพบว่ากาแพทั้ง 2 ประเภทนั้นเมลิ็ดกาแพดีมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 28-68 และเมลิ็ดกาแพเสียมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 0-69 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมลิ็ดกาแพดี และสีแดงแสดงค่าความถี่ของเมลิ็ดกาแพเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมลิ็ดกาแพทั้ง 2 ประเภท ดังฮิสโตแกรมเมลิ็ดกาแพทั้ง 2 ประเภท นั้นซ้อนทับกันเป็นบางส่วนค่า vAvg นี้ สามารถจำแนกประเภทเมลิ็ดกาแพดีและประเภทเมลิ็ดกาแพเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-28 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันได้อย่างชัดเจน

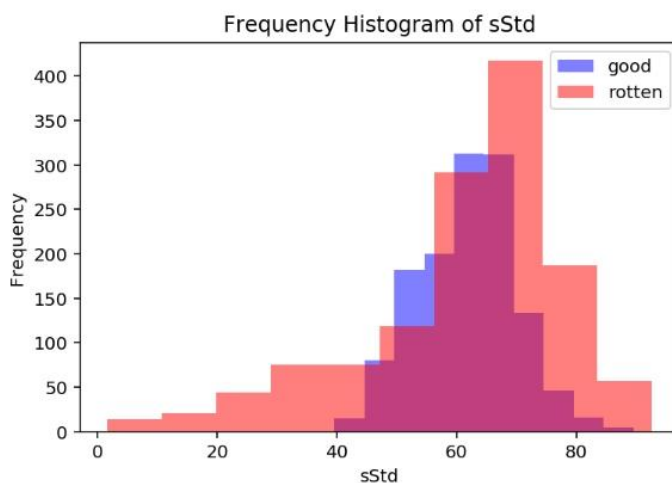


ภาพประกอบ 75 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ vAvg ของเมลิ็ดกาแฟทั้ง 2 ประเภท

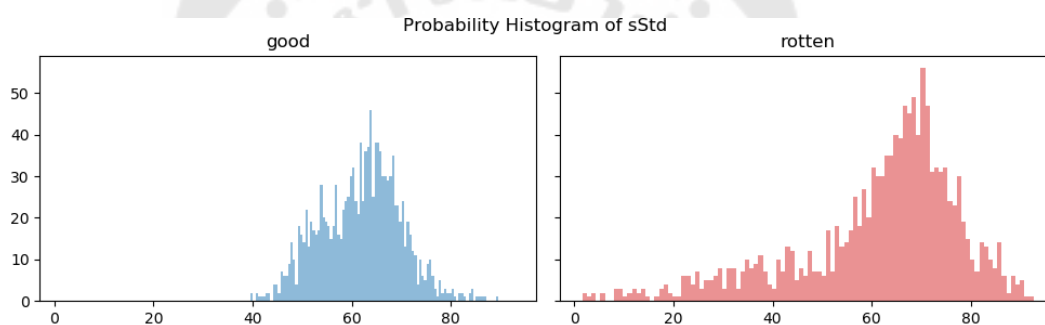


ภาพประกอบ 76 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ vAvg ของเมลิ็ดกาแฟทั้ง 2 ประเภท

23. sStd (Saturation Standard Deviation) ค่าสี HSV (Hue Saturation Value) เป็นค่าสีของความอิ่มตัว ค่าของเฉดสี และค่าความสว่างของสี โดยใช้หลักการพิจารณาการแสดงค่าตัวเลขอยู่ระหว่าง 0 ถึง 255 ในคุณลักษณะ sStd นี้ จะเป็น ค่าเบี่ยงเบนมาตรฐาน โดยที่ ค่า SD จะแสดงให้เห็นว่าโดยเฉลี่ยแล้วข้อมูลแต่ละตัวนั้นแตกต่างจากค่าเฉลี่ยของค่า Saturation จากค่าความถี่ของเมลิ็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมลิ็ดกาแฟดีมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 40-86 และเมลิ็ดกาแฟเสียมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 0-88 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมลิ็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมลิ็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมลิ็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมลิ็ดกาแฟทั้ง 2 ประเภท นั้นซ้อนทับกันเป็นบางส่วนค่า sStd นี้ สามารถจำแนกประเภทเมลิ็ดกาแฟดีและประเภทเมลิ็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 5-40 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันได้อย่างชัดเจน แต่ยังถือเป็นเพียงส่วนน้อยเท่านั้น

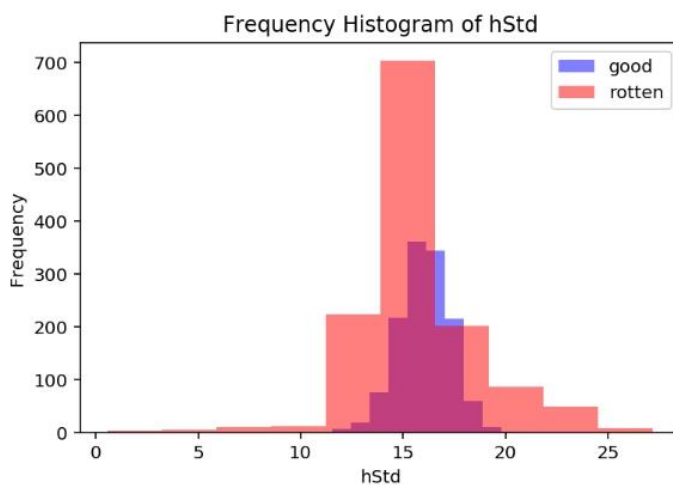


ภาพประกอบ 77 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ sStd ของเมล็ดกาแฟทั้ง 2 ประเภท

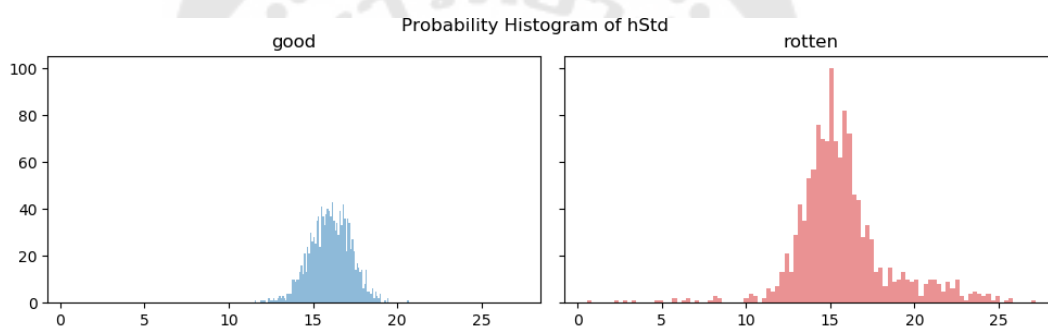


ภาพประกอบ 78 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ sStd ของเมล็ดกาแฟทั้ง 2 ประเภท

24. hStd (Hue Standard Deviation) ค่าสี HSV (Hue Saturation Value) เป็นค่าสีของความอิ่มตัว ค่าของเฉดสี และค่าความสว่างของสี โดยใช้หลักการพิจารณาการแสดงค่าตัวเลขอยู่ระหว่าง 0 ถึง 255 ในคุณลักษณะ hStd นี้ จะเป็น ค่าเบี่ยงเบนมาตรฐาน โดยที่ ค่า SD จะแสดงให้เห็นว่าโดยเฉลี่ยแล้วข้อมูลแต่ละตัวนั้นแตกต่างจากค่าเฉลี่ยของค่า Hue จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 12-19 และเมล็ดกาแฟเสียมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 1-28 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมล็ดกาแฟทั้ง 2 ประเภท นั้นซ้อนทับกันเป็นบางส่วนค่า hStd นี้ สามารถจำแนกประเภทเมล็ดกาแฟดีและประเภทเมล็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 1-11 และ 19-28 ซึ่งค่าเฉลี่ยทั้งส่วนนี้สามารถแยกออกจากกันอย่างชัดเจน แต่ก็ยังถือเป็นเพียงส่วนน้อยเท่านั้น

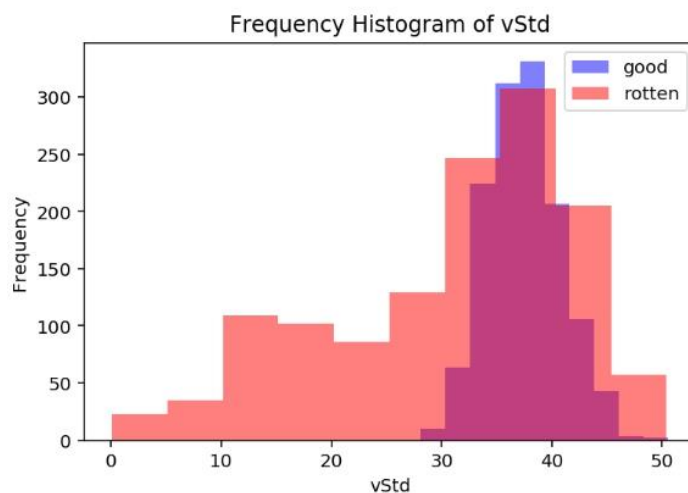


ภาพประกอบ 79 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ hStd ของเมล็ดกาแฟทั้ง 2 ประเภท

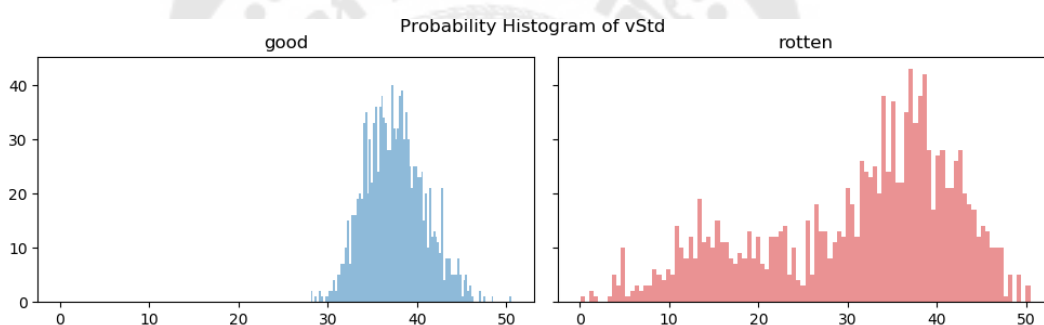


ภาพประกอบ 80 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ hStd ของเมล็ดกาแฟทั้ง 2 ประเภท

25. vStd (Value Standard Deviation) ค่าสี HSV (Hue Saturation Value) เป็นค่าสีของความอิ่มตัว ค่าของเฉดสี และค่าความสว่างของสี โดยใช้หลักการพิจารณาการแสดงค่าตัวเลขอยู่ระหว่าง 0 ถึง 255 ในคุณลักษณะ vStd นี้ จะเป็น ค่าเบี่ยงเบนมาตรฐาน โดยที่ ค่า SD จะแสดงให้เห็นว่าโดยเฉลี่ยแล้วข้อมูลแต่ละตัวนั้นแตกต่างจากค่าเฉลี่ยของค่า Value จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 28-51 และเมล็ดกาแฟเสียมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 0-51 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมล็ดกาแฟทั้ง 2 ประเภท นั้นซ้อนทับกันเป็นบางส่วนค่า vStd นี้ สามารถจำแนกประเภทเมล็ดกาแฟดีและประเภทเมล็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-28 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันได้อย่างชัดเจน

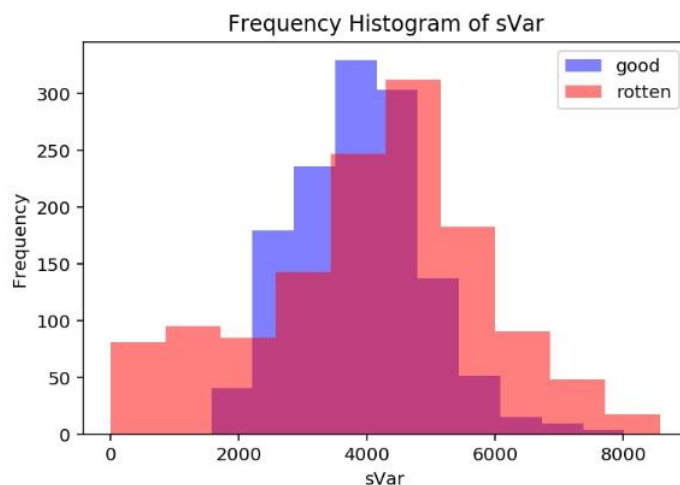


ภาพประกอบ 81 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ vStd ของเมล็ดกาแฟทั้ง 2 ประเภท

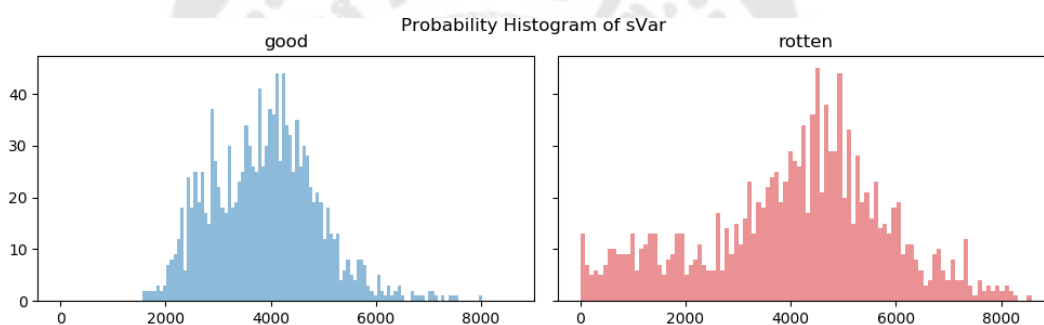


ภาพประกอบ 82 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ vStd ของเมล็ดกาแฟทั้ง 2 ประเภท

26. sVar (Saturation Variance) ค่าสี HSV (Hue Saturation Value) เป็นค่าสีของความอิ่มตัว ค่าของเฉดสี และค่าความสว่างของสี โดยใช้หลักการพิจารณาการแสดงค่าตัวเลขอยู่ระหว่าง 0 ถึง 255 ในคุณลักษณะ sVar นี้ จะเป็น ค่าความแปรปรวน เป็นค่าที่ใช้วัดการกระจายของค่า Saturation จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้น เมล็ดกาแฟดีมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 1600-8000 และเมล็ดกาแฟเสียมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 0-8500 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมล็ดกาแฟทั้ง 2 ประเภท นั้นซ้อนทับกันเป็นบางส่วนค่า sVar นี้ สามารถจำแนกประเภทเมล็ดกาแฟดีและประเภทเมล็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-1600 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันอย่างชัดเจน แต่ก็ยังถือเป็นเพียงส่วนน้อยเท่านั้น

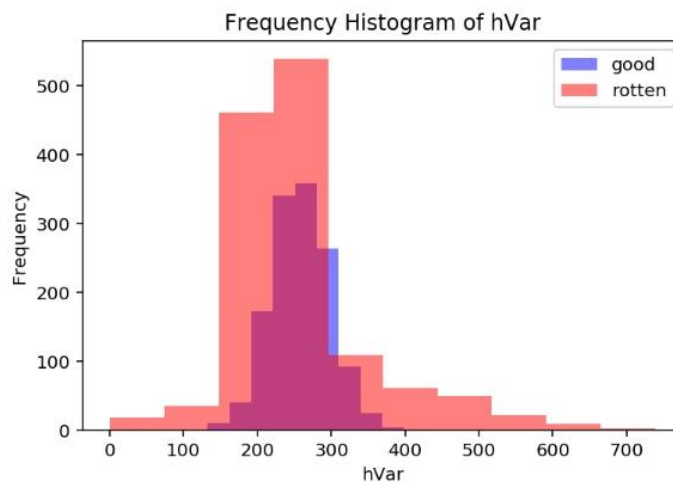


ภาพประกอบ 83 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ sVar ของเมล็ดกาแฟทั้ง 2 ประเภท

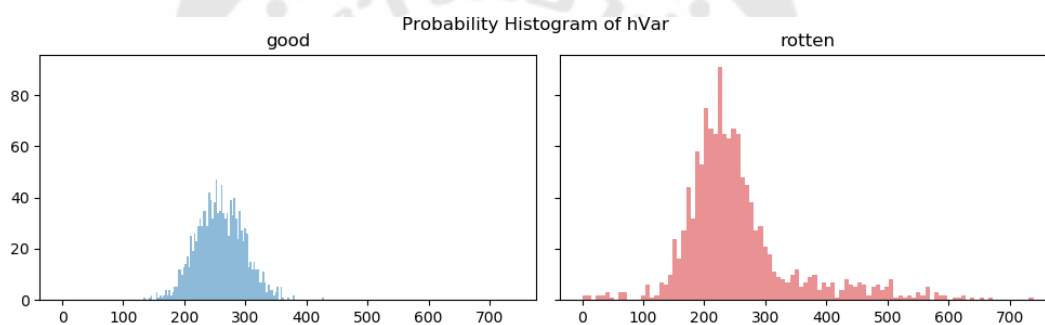


ภาพประกอบ 84 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ sVar ของเมล็ดกาแฟทั้ง 2 ประเภท

27. hVar (Hue Variance) ค่าสี HSV (Hue Saturation Value) เป็นค่าสีของความเข้มตัว ค่าของเฉดสี และค่าความสว่างของสี โดยใช้หลักการพิจารณาการแสดงค่าตัวเลขอยู่ระหว่าง 0 ถึง 255 ในคุณลักษณะ hVar นี้ จะเป็น ค่าความแปรปรวน เป็นค่าที่ใช้วัดการกระจายของค่า Hue จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 120-400 และเมล็ดกาแฟเสียมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 0-750 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมล็ดกาแฟทั้ง 2 ประเภท นั้นซ้อนทับกันเป็นบางส่วนค่า hVar นี้ สามารถจำแนกประเภทเมล็ดกาแฟดีและประเภทเมล็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-120 และ 400-750 ซึ่งค่าเฉลี่ยทั้งสองส่วนนี้สามารถแยกออกจากกันได้อย่างชัดเจน แต่ก็ยังถือเป็นเพียงส่วนน้อยเท่านั้น

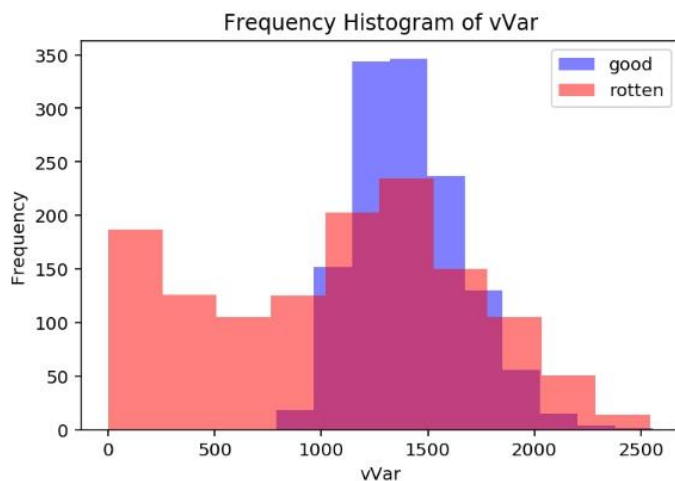


ภาพประกอบ 85 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ hVar ของเมล็ดกาแฟทั้ง 2 ประเภท

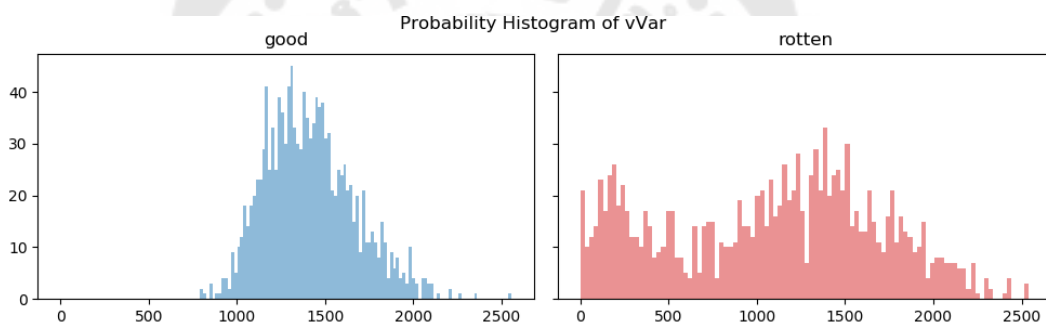


ภาพประกอบ 86 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ hVar ของเมล็ดกาแฟทั้ง 2 ประเภท

28. vVar (Value Variance) ค่าสี HSV (Hue Saturation Value) เป็นค่าสีของความอิ่มตัว ค่าของเฉดสี และค่าความสว่างของสี โดยใช้หลักการพิจารณาการแสดงค่าตัวเลขอยู่ระหว่าง 0 ถึง 255 ในคุณลักษณะ vVar นี้ จะเป็น ค่าความแปรปรวน เป็นค่าที่ใช้วัดการกระจายของค่า Value จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 1600-8000 และเมล็ดกาแฟเสียมีค่าองศาเฉลี่ยอยู่ที่ระหว่าง 0-8500 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมเมล็ดกาแฟทั้ง 2 ประเภท นั้นซ้อนทับกันเป็นบางส่วนค่า vVar นี้ สามารถจำแนกประเภทเมล็ดกาแฟดีและประเภทเมล็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-1600 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันอย่างชัดเจน แต่ก็ยังถือเป็นเพียงส่วนน้อยเท่านั้น

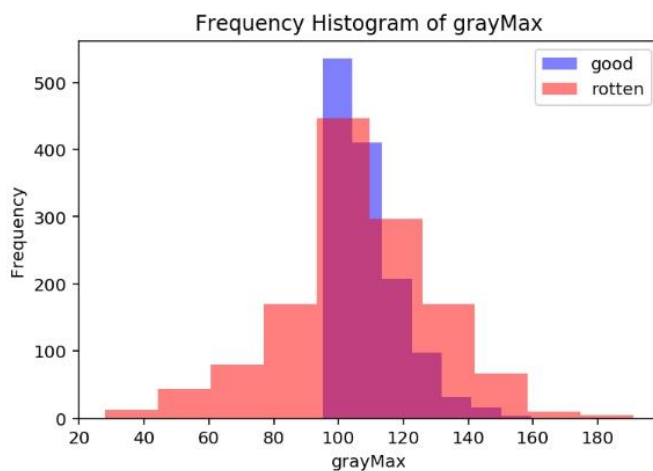


ภาพประกอบ 87 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ vVar ของเมล็ดกาแฟทั้ง 2 ประเภท

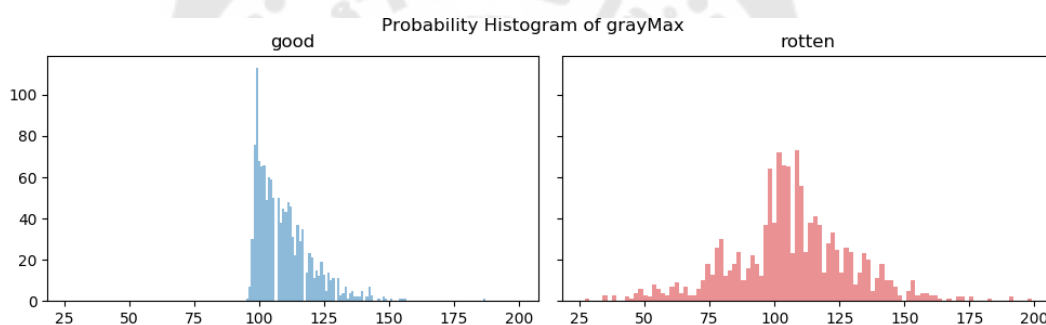


ภาพประกอบ 88 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ vVar ของเมล็ดกาแฟทั้ง 2 ประเภท

29. grayMax (gray color Max) การแปลงภาพสีแบบ RGB เป็นภาพระดับสีเทา (Gray-scale image) เป็นการปรับให้ภาพแสดงถึงค่าความสว่าง (Brightness) ของภาพ ในคุณลักษณะ grayMax นี้จะเป็น ค่ามากที่สุดของสีเทาบนเมล็ดกาแฟ จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าร้อยละเฉลี่ยอยู่ที่ 95-160 เมล็ดกาแฟเสียมีค่าร้อยละเฉลี่ยอยู่ที่ระหว่าง 28-190 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมรูปธรรมชาติ (Normal histogram) ที่มีค่าเฉลี่ยอยู่ตรงกลาง มีลักษณะสมมาตร เมล็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ค่า grayMax นี้ สามารถจำแนกประเภทเมล็ดกาแฟดีและประเภทเมล็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 28-95 และ 160-190 ซึ่งค่าเฉลี่ยทั้งสองส่วนนี้สามารถแยกออกจากกันได้อย่างชัดเจน

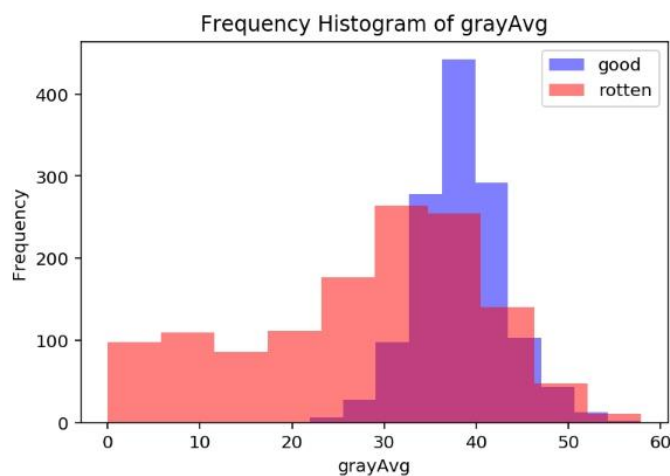


ภาพประกอบ 89 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ grayMax ของเมล็ดกาแฟทั้ง 2 ประเภท

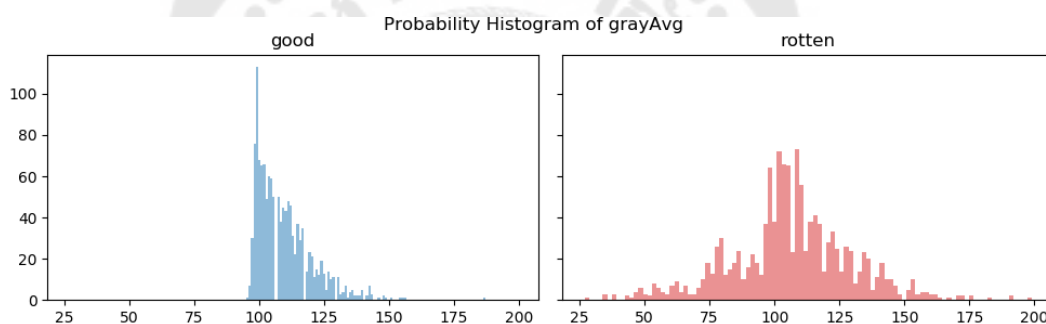


ภาพประกอบ 90 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ grayMax ของเมล็ดกาแฟทั้ง 2 ประเภท

30. grayAvg (gray color Average) การแปลงภาพสีแบบ RGB เป็นภาพระดับสีเทา (Gray-scale image) เป็นการปรับให้ภาพแสดงถึงค่าความสว่าง (Brightness) ของภาพ ในคุณลักษณะ grayAvg นี้ จะเป็น ค่าเฉลี่ยของสีเทาบนเมล็ดกาแฟ จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าร้อยละเฉลี่ยอยู่ที่ 22-58 เมล็ดกาแฟเสียมีค่าร้อยละเฉลี่ยอยู่ที่ระหว่าง 0-58 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมรูปทรงปกติ (Normal histogram) ที่มีค่าเฉลี่ยอยู่ตรงกลาง มีลักษณะสมมาตร เมล็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ค่า grayAvg นี้ สามารถจำแนกประเภทเมล็ดกาแฟดีและประเภทเมล็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-22 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันอย่างชัดเจน



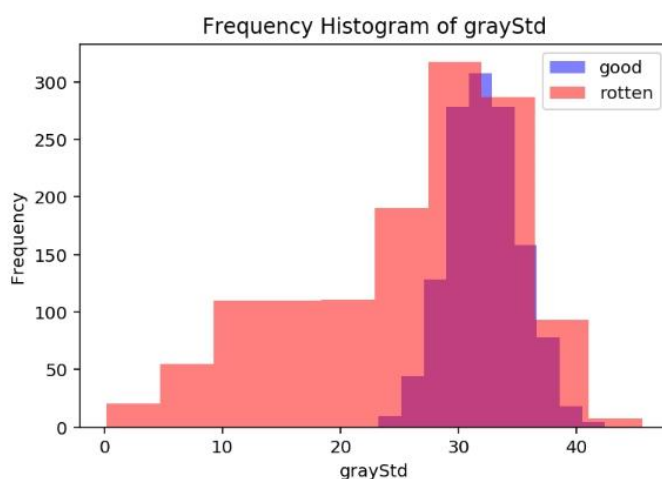
ภาพประกอบ 91 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ grayAvg ของเมล็ดกาแฟทั้ง 2 ประเภท



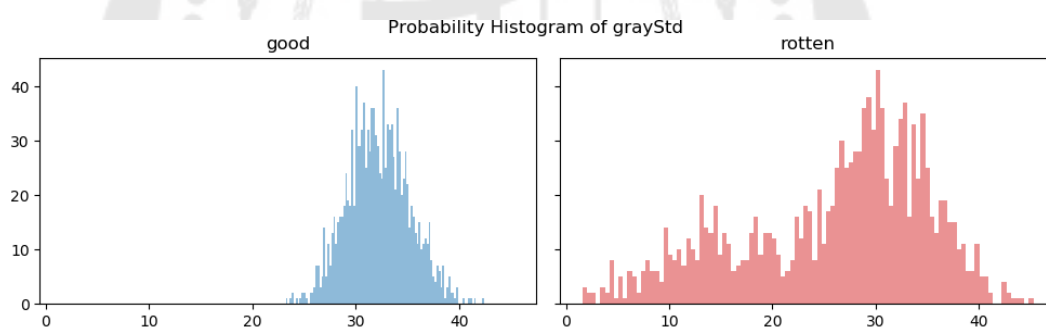
ภาพประกอบ 92 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ grayAvg ของเมล็ดกาแฟทั้ง 2 ประเภท

31. grayStd (gray color Standard Deviation) ค่า สี HSV (Hue Saturation Value) เป็นค่าสีของความอิ่มตัว ค่าของเฉดสี และค่าความสว่างของสี โดยใช้หลักการพิจารณาการแสดงค่าตัวเลขอยู่ระหว่าง 0 ถึง 255 ในคุณลักษณะ grayStd นี้ จะเป็น ค่าเบี่ยงเบนมาตรฐาน โดยที่ ค่า SD จะแสดงให้เห็นว่าโดยเฉลี่ยแล้วข้อมูลแต่ละตัวนั้นแตกต่างจากค่าเฉลี่ยของค่าสีเทา จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าร้อยละเฉลี่ยอยู่ที่ 23-43 เมล็ดกาแฟเสียมีค่าร้อยละเฉลี่ยอยู่ที่ระหว่าง 0-47 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแบ่งสีน้ำเงินและแบ่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมรูปธรรมชาติ (Normal histogram) ที่มีค่าเฉลี่ยอยู่ตรงกลาง มีลักษณะสมมาตร เมล็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกันอยู่บางส่วน ค่า grayAvg นี้ สามารถจำแนกประเภทเมล็ดกาแฟดีและประเภทเมล็ด

กาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-23 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันได้ได้อย่างชัดเจน



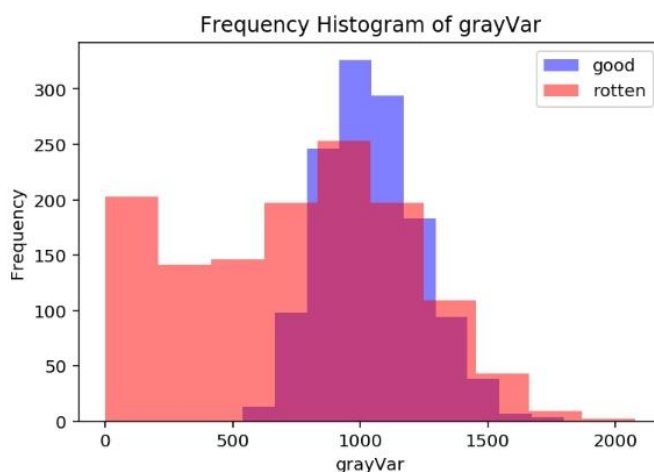
ภาพประกอบ 93 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ grayStd ของเมล็ดกาแฟทั้ง 2 ประเภท



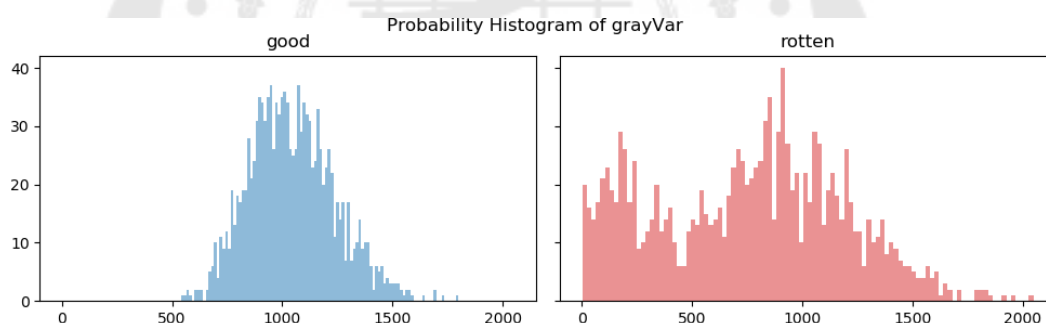
ภาพประกอบ 94 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ grayStd ของเมล็ดกาแฟทั้ง 2 ประเภท

32. grayVar (gray color Variance) เป็นค่าเบี่ยงเบนมาตรฐานของสีแดง โดยที่ ค่า SD จะแสดงให้เห็นว่าโดยเฉลี่ยแล้วข้อมูลแต่ละตัวนั้นแตกต่างจากค่าเฉลี่ยของค่าสีเทา จากค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท จะพบว่ากาแฟทั้ง 2 ประเภทนั้นเมล็ดกาแฟดีมีค่าร้อยละเฉลี่ยอยู่ที่ 550-1800 เมล็ดกาแฟเสียมีค่าร้อยละเฉลี่ยอยู่ที่ระหว่าง 0-2200 โดยที่สีน้ำเงินแสดงค่าความถี่ร้อยละของเมล็ดกาแฟดี และสีแดงแสดงค่าความถี่ของเมล็ดกาแฟเสีย ซึ่งแท่งสีน้ำเงินและแท่งสีแดงแสดงค่าความถี่ของเมล็ดกาแฟทั้ง 2 ประเภท ดังฮิสโตแกรมรูปทรงรูประฆัง (Normal histogram) ที่มีค่าเฉลี่ยอยู่ตรงกลาง มีลักษณะสมมาตร เมล็ดกาแฟทั้ง 2 ประเภทนั้นซ้อนทับกัน

อยู่บางส่วน ค่า grayAvg นี้สามารถจำแนกประเภทเมล็ดกาแฟดีและประเภทเมล็ดกาแฟเสียออกจากกันได้บางส่วนจากค่าเฉลี่ย 0-550 ซึ่งค่าเฉลี่ยในส่วนนี้สามารถแยกออกจากกันอย่างชัดเจน



ภาพประกอบ 95 ฮิสโตแกรมแสดงค่าความถี่คุณลักษณะ grayVar ของเมล็ดกาแฟทั้ง 2 ประเภท



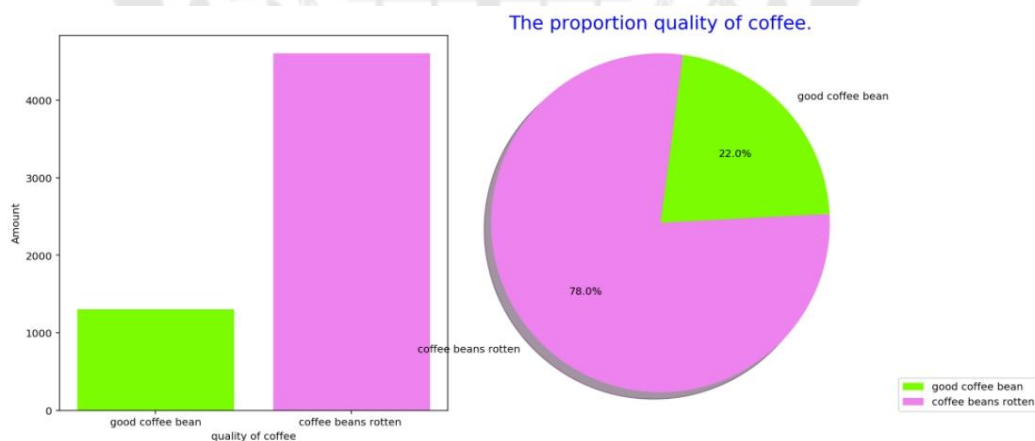
ภาพประกอบ 96 ฮิสโตแกรมแสดงค่าความน่าจะเป็นคุณลักษณะ grayVar ของเมล็ดกาแฟทั้ง 2 ประเภท

สรุปจากการเปรียบเทียบคุณลักษณะตามฮิสโตแกรมทั้งหมดแล้ว จะเห็นได้ว่าบางคุณลักษณะก็สามารถจำแนกประเภทเมล็ดกาแฟได้บ้างเป็นบางส่วน บางคุณลักษณะก็ไม่สามารถจำแนกประเภทเมล็ดกาแฟได้หรือได้เพียงเล็กน้อยเท่านั้น แต่หากนำคุณลักษณะข้างต้นนั้นมาเข้าโมเดลเพื่อ สร้างแบบจำลองการทำนายการจำแนกคุณภาพของเมล็ดกาแฟ โดยใช้เทคนิคการเรียนรู้ของเครื่องที่หลากหลายและใช้ทุกคุณลักษณะที่มีในการทำนายแล้ว ผู้วิจัยคิดว่าอาจทำให้ผลลัพธ์ในการจำแนกประเภทได้อย่างถูกต้องและแม่นยำมากยิ่งขึ้น และเพื่อผู้วิจัยจะสามารถเลือกเทคนิคที่จะนำมาพัฒนาระบบต่อไปเพื่อให้ได้ค่าความถูกต้องที่ดีที่สุดต่อไปได้

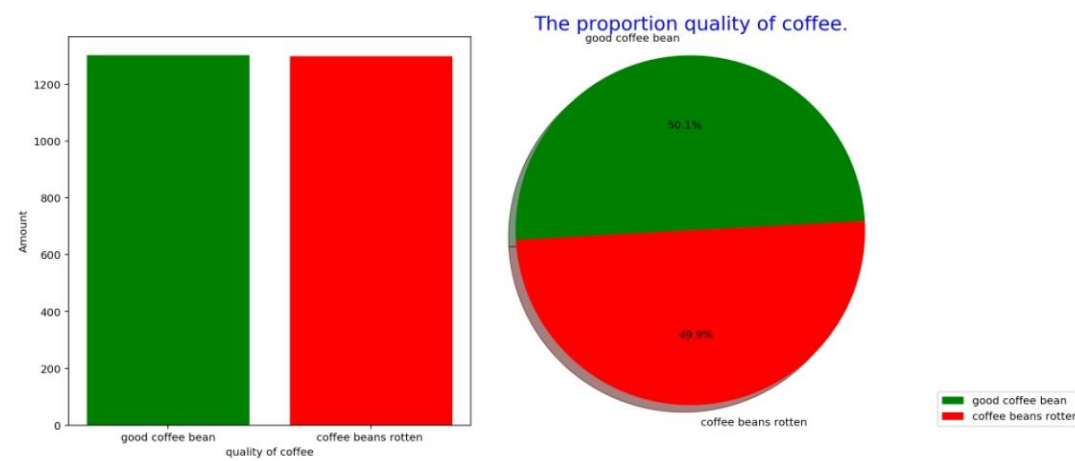
3.5 การจำแนกประเภทเมล็ดกาแฟ (Coffee Beans Classification)

การเตรียมข้อมูลกลุ่มตัวอย่าง โดยเลือกใช้แบบจำลอง Logistic Regression (LR) และแบบจำลอง Random Forest (RF) เนื่องจาก Logistic Regression (LR) เป็นแบบจำลองเชิงเส้น ซึ่งสามารถใช้จำแนกปัญหาของข้อมูลที่เป็นรูปแบบของโครงสร้างได้ดี ส่วน Random Forest (RF) เป็นแบบจำลองที่ใช้ต้นไม้การตัดสินใจ และได้แสดงให้เห็นมาโดยตลอดว่าสามารถใช้ในการจำแนกข้อมูลในหลากหลายรูปแบบอย่างกว้างขวาง อีกทั้งยังมีประสิทธิภาพที่ดีอีกด้วย ส่วนแบบจำลอง Random Forest (RF) นั้น เป็นเทคนิคที่ถูกพัฒนาต่อยอดมาจากเทคนิค Bagged Decision Tree

ผู้วิจัยได้นำกลุ่มตัวอย่างที่ได้มาแบ่งออกเป็น 2 แบบ คือ แบบ Imbalanced Data และชุดข้อมูลสมดุล (Balanced Data) เนื่องจากหากอ้างอิงจากสถานการณ์จริงในไร่กาแฟแล้วนั้น การคัดแยกเมล็ดกาแฟเสียออกจากเมล็ดกาแฟดี จะมีสัดส่วนเมล็ดกาแฟดีมากกว่ากาแฟเสียที่ค่อนข้างมาก และจากการเก็บตัวอย่างในงานวิจัยนี้นั้น ได้ทำการเก็บตัวอย่างที่มีเมล็ดเสียมากกว่าเมล็ดกาแฟดี ผู้วิจัยจึงทำการฝึกฝนแบบจำลองออกเป็น 2 กลุ่ม ผู้วิจัยได้ใช้ตัวอย่างในการศึกษานี้ โดยกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data) มีจำนวนเมล็ดดีจำนวน 1,303 เมล็ด และเมล็ดเสียจำนวน 4,607 เมล็ด ดังภาพประกอบ 97 และชุดข้อมูลไม่สมดุล (Imbalanced Data) มีจำนวนเมล็ดดีและเมล็ดเสียอย่างละ 1,303 เมล็ด ดังภาพประกอบ 98



ภาพประกอบ 97 จำนวนกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data)



ภาพประกอบ 98 จำนวนกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data)

เมื่อทำการสกัดคุณลักษณะของเมล็ดกาแฟจากภาพถ่ายออกมาได้ทั้งหมดจำนวน 32 คุณลักษณะ ดังคำอธิบายในหัวข้อข้างต้นแล้ว จากนั้นผู้วิจัยจึงนำข้อมูลที่ได้มาทำการแบ่งเป็น Training Dataset และ Test Dataset ในอัตราส่วนคือ 80 % ต่อ 20 % เป็นแบบ Imbalanced Data มีเมล็ดดีจำนวน 1,042 เมล็ด และเมล็ดเสียจำนวน 3,686 เมล็ดใน Training Dataset และเมล็ดดีจำนวน 261 เมล็ด และเมล็ดเสียจำนวน 921 เมล็ดชุดข้อมูลสมดุล (Balanced Data) เมล็ดดีและเสียจำนวนอย่างละ 1,042 เมล็ดใน Training Dataset และอย่างละ 261 เมล็ดใน Test Dataset ดังตาราง 5 ตาราง 5 การแบ่งกลุ่มตัวอย่างในการฝึกฝนแบบจำลอง

กลุ่มตัวอย่าง	ประเภทเมล็ดกาแฟ	จำนวนเมล็ด	Training Dataset	Test Dataset
Balanced Data	เมล็ดดี	1,303	1,042	261
	เมล็ดเสีย	1,303	1,042	261
Imbalanced Data	เมล็ดดี	1,303	1,042	261
	เมล็ดเสีย	4,607	3,686	921

ในขั้นตอนการฝึกฝนของแบบจำลองนั้น เนื่องจากคุณลักษณะทั้ง 32 คุณลักษณะนั้นมี Scale ที่ไม่เท่ากัน จึงต้องมีการทำ Feature Scaling โดยได้เลือกวิธีการทำแบบ Standard Scaler เพื่อให้ข้อมูลที่จะถูกนำไปใช้นั้นเป็นมาตรฐาน Scale เดียวกัน และอยู่ในรูปแบบของข้อมูลที่เหมาะสม ผู้วิจัยได้มีการนำเครื่องมือ Grid Search มาช่วยในการหาค่าที่ดีที่สุดของพารามิเตอร์แบบกริดด้วย รวมถึงการใช้เครื่องมือการตรวจสอบแบบไขว้ (K-Fold cross-Validation) โดยกำหนดค่า k ไว้จำนวน 10 Fold ด้วยกัน และเพื่อเป็นการป้องกันการเกิดการรั่วไหลของข้อมูล (Data Leakage) จึงได้มีการใช้ Pipeline ควบคุมการทำงานในแต่ละขั้นตอนด้วย



บทที่ 4

ผลการวิจัย

ในการวิจัยการสกัดและการคัดเลือกคุณลักษณะของเมล็ดกาแฟเพื่อจำแนกประเภทตามคุณภาพด้วยการประมวลผลภาพถ่ายนั้น หลังจากที่ได้ทำการสกัดและคัดเลือกคุณลักษณะของเมล็ดกาแฟออกมาแล้ว ผู้วิจัยได้ดำเนินการวัดประสิทธิภาพการจำแนกประเภทตามคุณภาพของเมล็ดกาแฟโดยใช้เทคนิคการเรียนรู้ของเครื่อง เพื่อให้บรรลุวัตถุประสงค์ของผู้วิจัยที่ได้กำหนดไว้ ดังนี้

1. แบบจำลองและการเปรียบเทียบประสิทธิภาพจากกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data)
2. แบบจำลองและการเปรียบเทียบประสิทธิภาพจากกลุ่มตัวอย่างแบบ Imbalanced Data

4.1 แบบจำลองและการเปรียบเทียบประสิทธิภาพจากกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data)

การทดลองนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพแบบจำลองการฝึกฝนสำหรับกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data) นั้น พารามิเตอร์ที่ใช้ในการทำ Grid Search ของ Logistic Regression (LR) คือ ค่า C ได้ค่า C ที่ดีที่สุดเท่ากับ 1,000 โดยที่ค่า C นี้คือการเพิ่ม Regularization เข้าไปเพื่อทำให้แบบจำลองนั้นมีความซับซ้อนน้อยลง ค่า C ยิ่งน้อย นั่นก็หมายความว่า มีการทำ Regularization มากขึ้นตามจำนวนของค่า C และยิ่งค่า C น้อย แบบจำลองนั้นก็จะมี ความซับซ้อนน้อยลงตาม เมื่อมีความซับซ้อนน้อยลงก็จะลดการ Over Fitting ตามไปด้วย พารามิเตอร์ที่ใช้ในการทำ Grid Search ของ Random Forest คือ ค่า Max Feature ได้ค่าที่ดีที่สุดเท่ากับ Log2 ซึ่งค่านี้คือค่าจำนวน Feature ในการพิจารณาเมื่อมองหา Feature ในการ Split โหนด ดังแสดงในตาราง 6 เมื่อได้ผลจากการสร้างแบบจำลองทั้ง 2 แบบจำลองที่กล่าวมาข้างต้นแล้วนั้น ผู้วิจัยทำการเปรียบเทียบประสิทธิภาพของการจำแนกคุณภาพของเมล็ดกาแฟ โดยการนำแบบจำลองมาทดสอบกับข้อมูล Test Dataset นำผลการจำแนกที่ได้จากแต่ละแบบจำลองมาหาค่า Accuracy, Precision, Recall และ F1-Score เพื่อเปรียบเทียบประสิทธิภาพการจำแนกคุณภาพของเมล็ดกาแฟดีและเมล็ดกาแฟเสีย ดังแสดงในตาราง 7 และภาพประกอบ 99

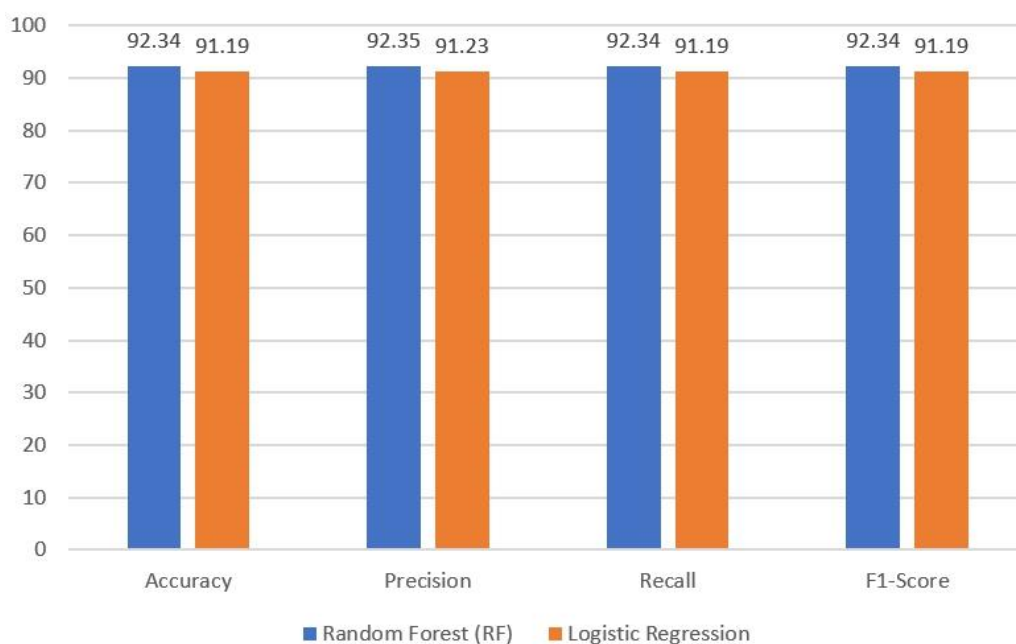
ตาราง 6 ตารางแสดงตัวอย่างค่าพารามิเตอร์จากกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data)

แบบจำลอง	Parameter			
Random Forest (RF)	bootstrap=True	class_weight=None	criterion='gini'	max_depth=None
	max_features='log2'	max_leaf_nodes=None	min_impurity_decrease=0.0	min_impurity_split=None
	min_samples_leaf=1	min_samples_split=2	min_weight_fraction_leaf=0.0	n_estimators=100
	n_jobs=None	oob_score=False	random_state=None	verbose=0
	warm_start=False			
Logistic Regression (LR)	C=1000	class_weight=None	dual=False	fit_intercept=True
	intercept_scaling=1	max_iter=100	multi_class=Auto	n_jobs=None
	penalty='l2'	random_state=None	solver='lbfgs'	tol=0.001
	verbose=0	warm_start=False		

ตาราง 7 แบบจำลองและการเปรียบเทียบประสิทธิภาพของกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data)

แบบจำลอง	Accuracy	Precision	Recall	F1-Score	Best Parameter
Random Forest (RF)	92.34	92.35	92.34	92.34	max_features='log2'
Logistic Regression (LR)	91.19	91.23	91.19	91.19	C= 1,000

แบบจำลองและการเปรียบเทียบประสิทธิภาพ



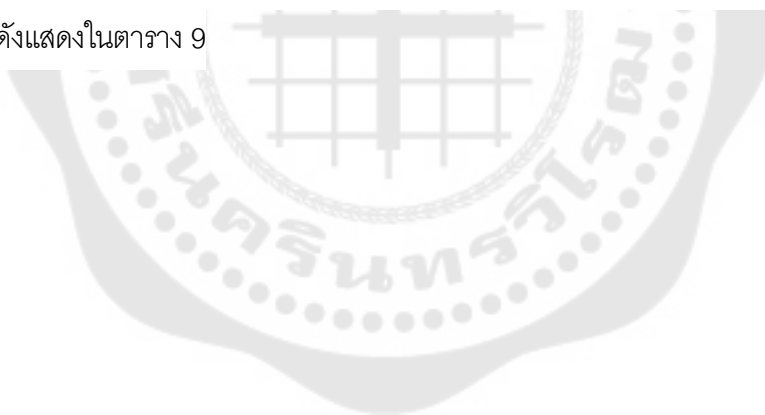
ภาพประกอบ 99 ภาพกราฟแบบจำลองและการเปรียบเทียบประสิทธิภาพ

กลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data)

จากผลการทดลองกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data) พบว่าแบบจำลอง Random Forest (RF) มีประสิทธิภาพสูงกว่า แบบจำลอง Logistic Regression (LR) ดังนี้ แบบจำลอง Random Forest (RF) ได้ค่าความแม่นยำ (Accuracy) คิดเป็น 92.34 เปอร์เซ็นต์ สูงกว่า 1.15 เปอร์เซ็นต์ เทียบกับแบบจำลอง Logistic Regression (LR) ที่ 91.19 เปอร์เซ็นต์ แบบจำลอง Random Forest (RF) ได้ค่าความเที่ยงตรง (Precision) คิดเป็น 92.35 เปอร์เซ็นต์ สูงกว่า 1.12 เปอร์เซ็นต์ เทียบกับแบบจำลอง Logistic Regression (LR) ที่ 91.19 เปอร์เซ็นต์ แบบจำลอง Random Forest (RF) ได้ค่าการระลึก (Recall) คิดเป็น 92.34 เปอร์เซ็นต์ สูงกว่า 1.15 เปอร์เซ็นต์ เทียบกับแบบจำลอง Logistic Regression (LR) ที่ 91.19 เปอร์เซ็นต์ แบบจำลอง Random Forest (RF) ได้ค่า F1 (F1-Score) คิดเป็น 92.34 เปอร์เซ็นต์ สูงกว่า 1.15 เปอร์เซ็นต์ เทียบกับแบบจำลอง Logistic Regression (LR) ที่ 91.19 เปอร์เซ็นต์ ดังภาพประกอบ 100

4.2 แบบจำลองและการเปรียบเทียบประสิทธิภาพจากกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data)

การทดลองนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในแบบจำลองการฝึกฝนสำหรับกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data) นั้น พารามิเตอร์ที่ใช้ในการทำ Grid Search ของ Logistic Regression (LR) คือ ค่า C ได้ค่า C ที่ดีที่สุดเท่ากับ 100 โดยที่ค่า C นี้คือการเพิ่ม Regularization เข้าไปเพื่อให้แบบจำลองนั้นมีความซับซ้อนน้อยลง ค่า C ยิ่งน้อย นั่นก็หมายความว่า มีการทำ Regularization มากขึ้นตามจำนวนของค่า C และยิ่งค่า C น้อย แบบจำลองนั้นก็จะมี ความซับซ้อนน้อยลงตาม เมื่อมีความซับซ้อนน้อยลงก็จะลดการ Over Fitting ตามไปด้วย และพารามิเตอร์ที่ใช้ในการทำ Grid Search ของ Random Forest คือ ค่า Max Feature ได้ค่าที่ดีที่สุดเท่ากับ Sqrt ซึ่งค่านี้คือค่าจำนวน Feature ในการพิจารณาเมื่อมองหา Feature ในการ Split โหนด ดังแสดงในตาราง 8 เมื่อได้ผลจากการสร้างแบบจำลองทั้ง 2 แบบจำลองที่กล่าวมาข้างต้นแล้วนั้น ผู้วิจัยทำการเปรียบเทียบประสิทธิภาพของการจำแนกคุณภาพของเมล็ดกาแฟ โดยการนำแบบจำลองมาทดสอบกับข้อมูล Test Dataset นำผลการจำแนกที่ได้จากแต่ละแบบจำลองมาหาค่า Accuracy, Precision, Recall และ F1-Score เพื่อเปรียบเทียบประสิทธิภาพการจำแนกคุณภาพของเมล็ดกาแฟดีและเมล็ดกาแฟเสีย ดังแสดงในตาราง 9

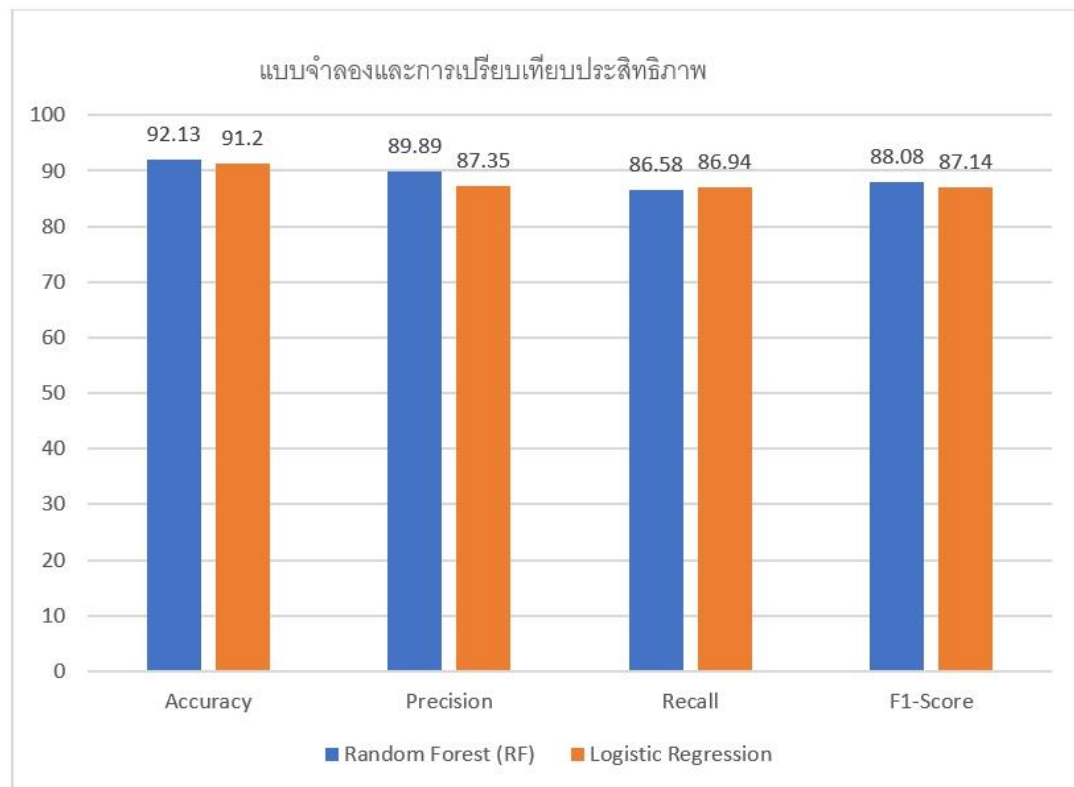


ตาราง 8 ตารางแสดงตัวอย่างค่าพารามิเตอร์จากกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล
(Imbalanced Data)

แบบจำลอง	Parameter			
Random Forest (RF)	.	class_weight=None	criterion='gini'	max_depth=None
	max_features='sqrt'	max_leaf_nodes=None	min_impurity_decrease=0.0	min_impurity_split=None
	min_samples_leaf=1	min_samples_split=2	min_weight_fraction_leaf=0.0	n_estimators=100
	n_jobs=None	oob_score=False	random_state=None	verbose=0
	warm_start=False			
Logistic Regression (LR)	C=100	class_weight=None	dual=False	fit_intercept=True
	intercept_scaling=1	max_iter=100	multi_class=Auto	n_jobs=None
	penalty='l2'	random_state=None	solver='lbfgs'	tol=0.001
	verbose=0	warm_start=False		

ตาราง 9 ตารางสรุปแบบจำลองและการเปรียบเทียบประสิทธิภาพกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล
(Imbalanced Data)

แบบจำลอง	Accuracy	Precision	Recall	F1-Score	Best Parameter
Random Forest (RF)	92.13	89.89	86.58	88.08	max_features='sqrt'
Logistic Regression (LR)	91.20	87.35	86.94	87.14	C= 100



ภาพประกอบ 100 ภาพแบบจำลองและการเปรียบเทียบประสิทธิภาพ

กลุ่มตัวอย่างชุดข้อมูลไม่สมดุล(Imbalanced Data)

จากผลการทดลองกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data) พบว่าแบบจำลอง Random Forest (RF) มีประสิทธิภาพสูงกว่า แบบจำลอง Logistic Regression (LR) ดังนี้ แบบจำลอง Random Forest (RF) ได้ค่าความแม่นยำ (Accuracy) คิดเป็น 92.13 เปอร์เซ็นต์ สูงกว่า 0.93 เปอร์เซ็นต์ เทียบกับแบบจำลอง Logistic Regression (LR) ที่ 91.20 เปอร์เซ็นต์ แบบจำลอง Random Forest (RF) ได้ค่าความเที่ยงตรง (Precision) คิดเป็น 89.89 เปอร์เซ็นต์ สูงกว่า 0.36 เปอร์เซ็นต์ เทียบกับแบบจำลอง Logistic Regression (LR) ที่ 87.35 เปอร์เซ็นต์ แบบจำลอง Random Forest (RF) ได้ค่าการระลึก (Recall) คิดเป็น 86.58 เปอร์เซ็นต์ ต่ำกว่า 0.36 เปอร์เซ็นต์ เทียบกับแบบจำลอง Logistic Regression (LR) ที่ 86.94 เปอร์เซ็นต์ แบบจำลอง Random Forest (RF) ได้ค่า F1 (F1-Score) คิดเป็น 88.08 เปอร์เซ็นต์ สูงกว่า 0.94เปอร์เซ็นต์ เทียบกับแบบจำลอง Logistic Regression (LR) ที่ 87.14 เปอร์เซ็นต์ ดังภาพประกอบ 100

บทที่ 5

อภิปรายผล สรุปผลการวิจัย และข้อเสนอแนะ

การทำการนายการจำแนกประเภทเมล็ดกาแฟจากการประมวลผลภาพถ่ายของเมล็ดกาแฟดีและเมล็ดกาแฟเสีย ที่ได้รับการพัฒนาและวิจัยนี้ พบว่าการเตรียมข้อมูลสำหรับการออกแบบจำลองโมเดลนั้น มีความสำคัญมาก ภาพถ่ายที่จะสามารถนำมาเป็นข้อมูลในการเตรียมสำหรับการออกแบบจำลองโมเดลการทำการนายการจำแนกเมล็ดกาแฟ ปัจจัยที่เกี่ยวข้องในการเก็บข้อมูล การคัดเลือกและรวบรวมข้อมูลภาพถ่าย จะต้องมีความถูกต้อง มีความน่าเชื่อถือ เหมาะสม และเพียงพอต่อการนำไปใช้สำหรับเป็นข้อมูลกลุ่มตัวอย่างเป็นอย่างมาก หากภาพถ่ายขาดองค์ประกอบอย่างใดอย่างหนึ่ง ก็จะส่งผลต่อการออกแบบจำลองโมเดลนั้นด้วย

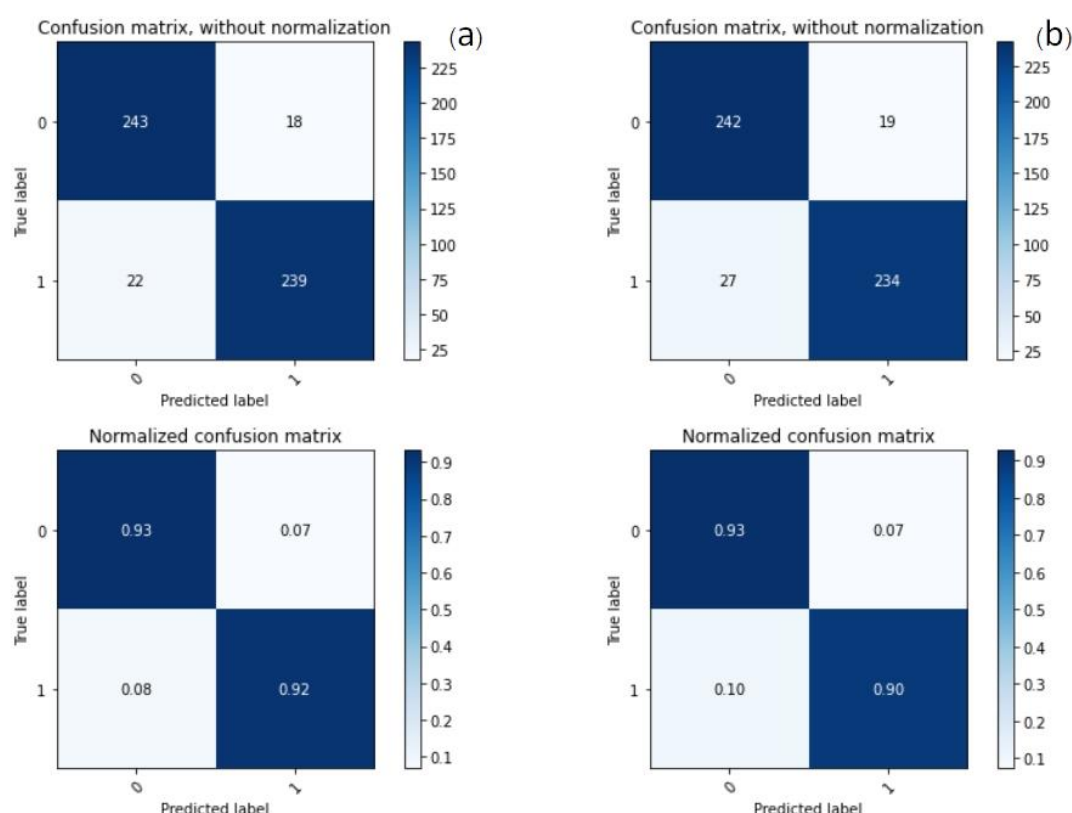
ผู้วิจัยได้ทำการวัดประสิทธิภาพจากจำแนกประเภทตามคุณภาพของเมล็ดกาแฟด้วยการใช้เทคนิคการเรียนรู้ของเครื่อง และได้วัดประสิทธิภาพของแต่ละแบบจำลองในการจำแนกเมล็ดกาแฟดีและเมล็ดกาแฟเสียด้วยการเรียนรู้ของเครื่อง เพื่อนำมาเปรียบเทียบและสรุปผล โดยแบ่งหัวข้อของการสรุปผลได้ดังต่อไปนี้

1. การอภิปรายผล
2. การสรุปผลการวิจัย
3. ข้อเสนอแนะ

5.1 การอภิปรายผล

งานวิจัยนี้นำเสนอวิธีการคัดแยกคุณภาพของเมล็ดกาแฟดีและเมล็ดกาแฟเสียจากภาพถ่ายเมล็ดกาแฟ ด้วยการสกัดคุณลักษณะและเทคนิคการเรียนรู้ของเครื่อง โดยผ่านกระบวนการ 4 ขั้นตอน คือ 1) การนำเข้าข้อมูลรูปภาพ 2) การประมวลผลภาพ เบื้องต้น 3) การสกัดคุณลักษณะ และ 4) การจำแนกเมล็ดกาแฟและเปรียบเทียบประสิทธิภาพแบบจำลองในการคัดแยกคุณภาพเมล็ดกาแฟ เทคนิคการเรียนรู้ของเครื่องที่ใช้สำหรับคัดแยกคุณภาพเมล็ดกาแฟประกอบด้วย 2 แบบจำลอง คือ Logistic Regression LR) และ Random Forest (RF) ตัวอย่างที่ใช้ในการศึกษามีจำนวนเมล็ดดีและเมล็ดเสีย อย่างละ 1,303 เมล็ด จากผลการทดลองกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data) พบว่าแบบจำลอง Random Forest (RF) มีประสิทธิภาพสูงกว่า แบบจำลอง Logistic Regression (LR) ดังนั้น ค่าความแม่นยำ (Accuracy) สูงกว่า 1.15 เปอร์เซ็นต์ ค่าความเที่ยงตรง (Precision) สูงกว่า 1.12 เปอร์เซ็นต์ ค่าการระลึก (Recall) สูงกว่า 1.15 เปอร์เซ็นต์ และค่าเฉลี่ย F1 (F1-Score) สูงกว่า 1.15 เปอร์เซ็นต์ และผลการทดลองกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data) พบว่าแบบจำลอง

Random Forest (RF) มีประสิทธิภาพสูงกว่า แบบจำลอง Logistic Regression (LR) ดังนี้ ค่าความแม่นยำ (Accuracy) สูงกว่า 0.93 เปอร์เซนต์ ค่าความเที่ยงตรง (Precision) สูงกว่า 2.54 เปอร์เซนต์ ค่าการระลึก (Recall) ต่ำกว่า 0.36 เปอร์เซนต์ และค่าเฉลี่ย F1 (F1-Score) สูงกว่า 0.94 เปอร์เซนต์



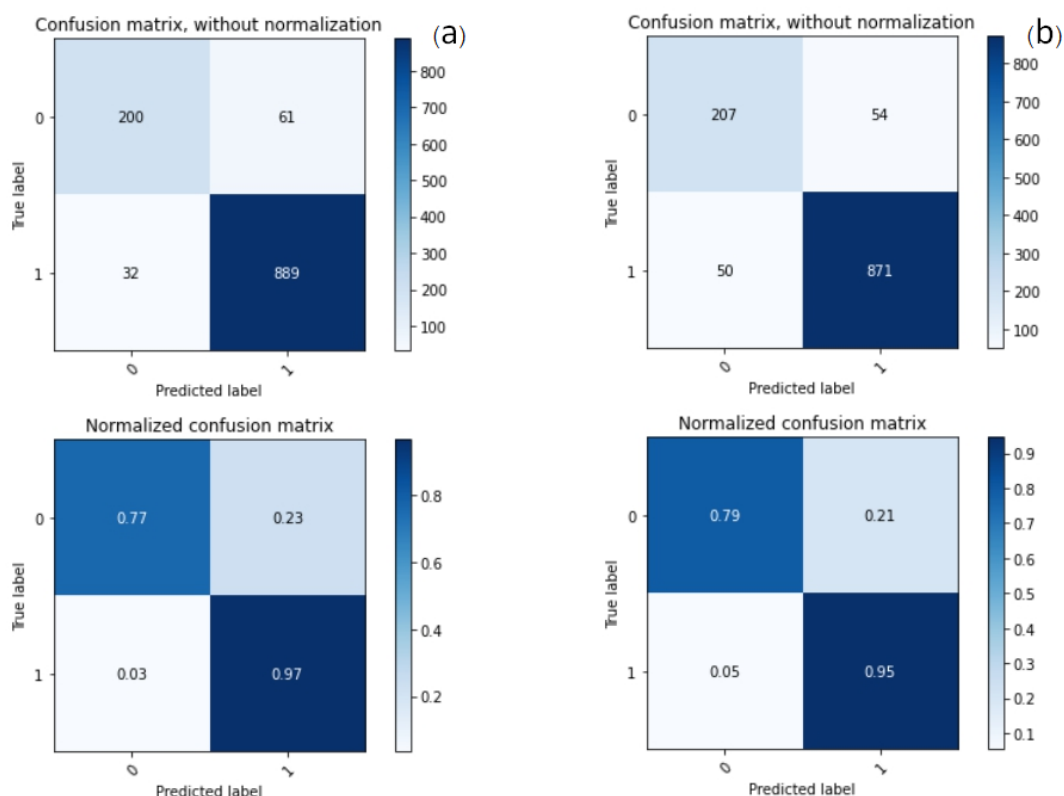
ภาพประกอบ 101 Confusion Matrix ของกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data)

(a) Confusion Matrix ของแบบจำลอง Random Forest (RF)

(b) Confusion Matrix ของแบบจำลอง Logistic Regression (LR)

จากผลการคำนวณ Confusion Matrix เพื่อประเมินประสิทธิภาพของแบบจำลองจากกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data) ดังภาพประกอบ 101 ทั้ง 2 แบบจำลองนั้น พบว่าแบบจำลอง Random Forest (RF) สามารถจำแนกเมล็ดกาแฟดีจากตัวอย่าง (Test Set) จำนวน 261 เมล็ด ได้ถูกต้องจำนวน 243 เมล็ด คิดเป็นร้อยละ 93 % และจำแนกผิดจำนวน 18 เมล็ด คิดเป็นร้อยละ 7 % และสามารถจำแนกเมล็ดกาแฟเสียจากจากตัวอย่าง (Test Set) จำนวน 261 เมล็ด ได้ถูกต้องจำนวน 239 เมล็ด คิดเป็นร้อยละ 92 % และจำแนกผิดจำนวน 22 เมล็ด คิดเป็นร้อยละ 8 % แบบจำลอง Logistic Regression (LR) สามารถจำแนกเมล็ดกาแฟดีจากตัวอย่าง (Test Set) จำนวน

261 เมล็ด ได้ถูกต้องจำนวน 242 เมล็ด คิดเป็นร้อยละ 93 % และจำแนกผิดจำนวน 19 เมล็ด คิดเป็นร้อยละ 7 % และสามารถจำแนกเมล็ดกาแฟเสียจากจากตัวอย่าง (Test Set) จำนวน 261 เมล็ด ได้ถูกต้องจำนวน 234 เมล็ด คิดเป็นร้อยละ 90 % และจำแนกผิดจำนวน 27 เมล็ด คิดเป็นร้อยละ 10 %



ภาพประกอบ 102 Confusion Matrix ของกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data)

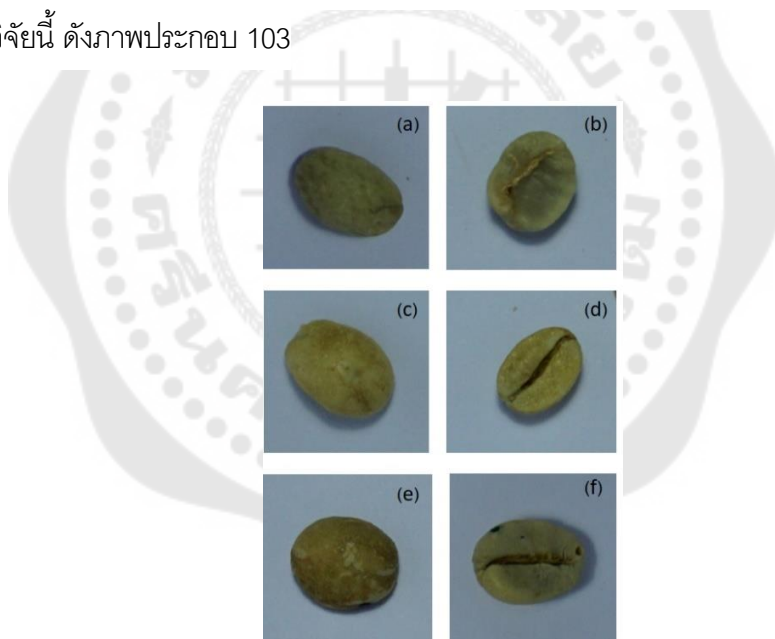
(a) Confusion Matrix ของแบบจำลอง Random Forest (RF)

(b) Confusion Matrix ของแบบจำลอง Logistic Regression (LR)

จากผลการคำนวณ Confusion Matrix เพื่อประเมินประสิทธิภาพของแบบจำลองจากกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data) ดังภาพประกอบ 102 ทั้ง 2 แบบจำลองนั้นพบว่าแบบจำลอง Random Forest (RF) สามารถจำแนกเมล็ดกาแฟดีจากตัวอย่าง (Test Set) จำนวน 261 เมล็ด ได้ถูกต้องจำนวน 200 เมล็ด คิดเป็นร้อยละ 77 % และจำแนกผิดจำนวน 61 เมล็ด คิดเป็นร้อยละ 23 % และสามารถจำแนกเมล็ดกาแฟเสียจากจากตัวอย่าง (Test Set) จำนวน 921 เมล็ด ได้ถูกต้องจำนวน 889 เมล็ด คิดเป็นร้อยละ 97 % และจำแนกผิดจำนวน 32 เมล็ด คิดเป็นร้อยละ 3 % แบบจำลอง Logistic Regression (LR) สามารถจำแนกเมล็ดกาแฟดีจากตัวอย่าง (Test Set) จำนวน 261 เมล็ด ได้ถูกต้องจำนวน 207 เมล็ด คิดเป็นร้อยละ 79 % และจำแนกผิดจำนวน 54 เมล็ด คิดเป็น

ร้อยละ 21 % และสามารถจำแนกเมล็ดกาแฟเสียจากจากตัวอย่าง (Test Set) จำนวน 921 เมล็ด ได้ถูกต้องจำนวน 871 เมล็ด คิดเป็นร้อยละ 95 % และจำแนกผิดจำนวน 50 เมล็ด คิดเป็นร้อยละ 5 %

โดยสรุป ข้อจำกัดของงานวิจัยนี้คือเมล็ดกาแฟดี และเมล็ดกาแฟเสียบางประเภทจะมีความคล้ายคลึงกันเป็นอย่างมาก หนึ่งในประเภทของเมล็ดกาแฟเสีย เช่น เมล็ดกาแฟเสียประเภทเมล็ดกาแฟไม่สมบูรณ์นั้น บางเมล็ดจะมีลักษณะของสีที่คล้ายเมล็ดดีอยู่มาก ซึ่งในกรณีนี้จะมีข้อจำกัดในเรื่องของการแยกประเภทของสีที่อาจเกิดความผิดพลาดในการจำแนกประเภทออกจากกันได้ และในอีกกรณีหนึ่ง คือ เมล็ดกาแฟเสียประเภทเมล็ดกาแฟไม่สมบูรณ์ เมล็ดกาแฟขึ้นรา และเมล็ดกาแฟที่ถูกแมลงทำลาย ใน 3 ประเภทนี้ จะมีข้อจำกัดในเรื่องของลักษณะรูปร่างที่คล้ายคลึงกัน หากไม่ได้จำแนกประเภทในเรื่องของสีแล้ว จะถูกจำแนกในเรื่องของลักษณะรูปร่างแทน ซึ่งในประเภทของเมล็ดกาแฟที่ถูกแมลงทำลายนั้น จากภาพถ่ายเมล็ดอาจมีรูที่แมลงทำลายอยู่บนเมล็ด หรือหากมีรูนั้นอยู่ข้างใต้เมล็ดในภาพถ่ายกลุ่มตัวอย่าง ก็จะเป็นอีกสาเหตุหนึ่งที่ทำให้เกิดการจำแนกประเภทผิดพลาดขึ้นได้ในงานวิจัยนี้ ดังภาพประกอบ 103



ภาพประกอบ 103 การทำนายผลของเมล็ดกาแฟ

ภาพ (a) และ (b) เมล็ดกาแฟดี คำตอบคือถูกต้อง และทำนายผลถูกต้อง

ภาพ (c) และ (d) เมล็ดกาแฟไม่สมบูรณ์ คำตอบคือผิด แต่ทำนายผลถูกต้อง

ภาพ (e) และ (f) เมล็ดถูกแมลงทำลาย คำตอบคือผิด แต่ทำนายผลถูกต้อง

ดังนั้นสำหรับการจำแนกประเภทเมล็ดกาแฟเป็นเมล็ดกาแฟดีและเมล็ดกาแฟเสียจากภาพถ่ายด้วยแบบจำลอง Random Forest (RF) สามารถนำไปประยุกต์ใช้ในการคัดแยกเมล็ดกาแฟในโรงกาแฟได้ต่อไป

5.2 การสรุปผลการวิจัย

งานวิจัยนี้นำเสนอวิธีการจำแนกประเภทของเมล็ดกาแฟดีและเมล็ดกาแฟเสียด้วยการประมวลผลภาพถ่าย ผลการศึกษาพบว่าเมื่อเปรียบเทียบประสิทธิภาพการจำแนกคุณภาพของเมล็ดกาแฟจากภาพถ่ายจากข้อมูลกลุ่มตัวอย่างทั้ง 2 แบบด้วย Accuracy, Precision, Recall และ F1-Score แบบจำลอง Random Forest (RF) มีประสิทธิภาพสูงกว่าแบบจำลอง Logistic Regression (LR) เล็กน้อย ดังนี้

5.2.1. สรุปผลจากกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data)

กลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data) แบบจำลอง Random Forest (RF) ได้ค่า Accuracy คิดเป็น 92.34 เปอร์เซ็นต์เทียบกับแบบจำลอง Logistic Regression (LR) ที่ 91.19 เปอร์เซ็นต์ แบบจำลอง Random Forest (RF) ได้ค่า Precision คิดเป็น 92.35 เปอร์เซ็นต์ เทียบกับแบบจำลอง Logistic Regression ที่ 91.19 เปอร์เซ็นต์ แบบจำลอง Random Forest (RF) ได้ค่า Recall คิดเป็น 92.34 เปอร์เซ็นต์ เทียบกับแบบจำลอง Logistic Regression (LR) ที่ 91.19 เปอร์เซ็นต์ และแบบจำลอง Random Forest (RF) ได้ค่า F1-Score คิดเป็น 92.34 เปอร์เซ็นต์ เทียบกับแบบจำลอง Logistic Regression ที่ 91.19 เปอร์เซ็นต์ แสดงในตาราง 10

จากผลการทดลองจากข้อมูลกลุ่มตัวอย่างพบว่า แบบจำลอง Random Forest (RF) มีประสิทธิภาพสูงกว่า แบบจำลอง Logistic Regression (LR) ดังนี้ ค่าความแม่นยำ (Accuracy) สูงกว่า 1.15 เปอร์เซ็นต์ ค่าความเที่ยงตรง (Precision) สูงกว่า 1.12 เปอร์เซ็นต์ ค่าการระลึก (Recall) สูงกว่า 1.15 เปอร์เซ็นต์ และค่าเฉลี่ย F1 (F1-Score) สูงกว่า 1.15 เปอร์เซ็นต์ ดังนั้น สำหรับการจำแนกประเภทเมล็ดกาแฟเป็นเมล็ดกาแฟดีและเมล็ดกาแฟเสียจากภาพถ่ายกลุ่มตัวอย่างชุดข้อมูลสมดุล (Balanced Data) ด้วยแบบจำลอง Random Forest (RF) สามารถนำไปประยุกต์ใช้ต่อไปได้ แสดงในตาราง 10

5.2.2. สรุปผลจากกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data)

กลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data) แบบจำลอง Random Forest (RF) ได้ค่า Accuracy คิดเป็น 92.13 เปอร์เซ็นต์เทียบกับแบบจำลอง Logistic Regression (LR) ที่ 91.20 เปอร์เซ็นต์ แบบจำลอง Random Forest (RF) ได้ค่า Precision คิดเป็น 89.89 เปอร์เซ็นต์ เทียบกับแบบจำลอง Logistic Regression (LR) ที่ 87.35 เปอร์เซ็นต์ แบบจำลอง Random Forest (RF) ได้ค่า Recall คิดเป็น 86.58 เปอร์เซ็นต์ เทียบกับแบบจำลอง Logistic Regression (LR) ที่ 86.94 เปอร์เซ็นต์ และแบบจำลอง Random Forest (RF) ได้ค่า F1-Score คิดเป็น 88.08 เปอร์เซ็นต์ เทียบกับแบบจำลอง Logistic Regression (LR) ที่ 87.14 เปอร์เซ็นต์

จากผลการทดลองจากข้อมูลกลุ่มตัวอย่างพบว่า แบบจำลอง Random Forest (RF) มีประสิทธิภาพสูงกว่า แบบจำลอง Logistic Regression (LR) ดังนี้ ค่าความแม่นยำ (Accuracy) สูงกว่า 0.93 เปอร์เซนต์ ค่าความเที่ยงตรง (Precision) สูงกว่า 2.54 เปอร์เซนต์ ค่าการระลึก (Recall) ต่ำกว่า 0.36 เปอร์เซนต์ และค่าเฉลี่ย F1 (F1-Score) สูงกว่า 0.94 เปอร์เซนต์ ดังนั้น สำหรับการจำแนกประเภทเมล็ดกาแฟเป็นเมล็ดกาแฟดีและเมล็ดกาแฟเสียจากภาพถ่ายกลุ่มตัวอย่างชุดข้อมูลไม่สมดุล (Imbalanced Data) ด้วยแบบจำลอง Random Forest (RF) สามารถนำไปประยุกต์ใช้ได้ต่อไปได้ แสดงในตาราง 10

ตาราง 10 ตารางสรุปแบบจำลองและการเปรียบเทียบประสิทธิภาพ

แบบจำลอง	กลุ่มตัวอย่าง	Accuracy	Precision	Recall	F1-Score	Best Parameter
Random Forest (RF)	Balanced Data	92.34	92.35	92.34	92.34	max_features='log2'
Logistic Regression (LR)		91.19	91.23	91.19	91.19	C= 1000
Random Forest (RF)	Imbalanced Data	92.13	89.89	86.58	88.08	max_features='sqrt'
Logistic Regression (LR)		91.20	87.35	86.94	87.14	C= 100

5.3 ข้อเสนอแนะ

แนวทางการพัฒนาประสิทธิภาพของระบบในอนาคต คือ

1. การลดผลกระทบจากการเปลี่ยนแปลงของแสงต่อภาพถ่ายเมล็ดกาแฟ เนื่องจากในงานวิจัยนี้ การเก็บถ่ายภาพตัวอย่างเมล็ดกาแฟยังเกิดเงาที่บดบังของเมล็ดกาแฟเกิดขึ้น สาเหตุมาจากการส่องสว่างจากโคมไฟแสงสีขาว และแสงจากหลอด LED ขนาบสองข้างกล่องนั้นอาจยังไม่เพียงพอ จึงทำให้เกิดเงาขึ้นได้ อีกทั้งยังเป็นการเพิ่มขั้นตอนในการเตรียมภาพเมล็ดกาแฟ หากมีการลำเลียงเมล็ดกาแฟจากตำแหน่งการจัดวางที่แน่นอน ก็จะช่วยให้การเก็บภาพถ่ายตัวอย่างเมล็ดกาแฟมีประสิทธิภาพมากยิ่งขึ้น อีกทั้งยังทำให้ขั้นตอนของการประมวลผลภาพเบื่อง่ายขึ้น

(Image Preprocessing) ซึ่งเป็นปัจจัยหลักของการจำแนกเมล็ดกาแฟนั้นสามารถลดขั้นตอนลงได้ และช่วยลดขั้นตอนในการเตรียมข้อมูลรูปภาพได้ง่ายขึ้นด้วย สามารถนำไปใช้ประยุกต์กับเครื่องคัดแยกเมล็ดกาแฟได้จริง

2. เพิ่มความสามารถในการจำแนกเมล็ดกาแฟเมื่อมีการทับซ้อนกันของเมล็ดกาแฟ จากกลุ่มตัวอย่างเมล็ดกาแฟของงานวิจัยนี้ มีการควบคุมจำนวนเมล็ดในขณะถ่ายภาพ จึงทำให้ไม่มีการทับซ้อนกันของเมล็ดกาแฟในกลุ่มตัวอย่าง ในการศึกษาครั้งถัดไปหากมีการทับซ้อนกันของเมล็ดกาแฟแบบจำลองจะต้องสามารถจำแนกประเภทเมล็ดกาแฟได้อย่างมีประสิทธิภาพด้วย

3. หาจำนวนคุณลักษณะที่เหมาะสมที่ให้ความแม่นยำสูงสุด ซึ่งจะเห็นได้ว่างานวิจัยนี้มีคุณลักษณะจำนวน 32 คุณลักษณะโดยแยกเป็น 3 กลุ่มด้วยกัน ในการศึกษาครั้งถัดไปอาจจะต้องมีการหาคุณลักษณะเพิ่มเติมจากเดิม เพื่อให้คุณลักษณะที่ใช้ในการจำแนกนั้นหลากหลายมากยิ่งขึ้น หรืออาจลดคุณลักษณะและจำนวนกลุ่มลงจากเดิม เพื่อให้ง่ายต่อการจำแนกประเภทเมล็ดกาแฟได้อย่างเฉพาะเจาะจง แต่มีประสิทธิภาพในการจำแนกที่เพิ่มมากขึ้น

บรรณานุกรม

- Calvini, R., Ulrici, A., และ Amigo, J. M. (2015). Practical comparison of sparse methods for classification of Arabica and Robusta coffee species using near infrared hyperspectral imaging. *Chemometrics and Intelligent Laboratory Systems*, 146, 503-511.
- Casas, M. I., Vaughan, M. J., Bonello, P., McSpadden Gardener, B., Grotewold, E., และ Alonso, A. P. (2017). Identification of biochemical features of defective *Coffea arabica* L. beans. *Food Research International*, 95, 59-67.
- de Oliveira, E. M., Leme, D. S., Barbosa, B. H. G., Rodarte, M. P., และ Pereira, R. G. F. A. (2016). A computer vision system for coffee beans classification based on computational intelligence techniques. *Journal of Food Engineering*, 171, 22-27.
- Dondegroovily. (2012). Machine Learning. http://en.wikipedia.org/wiki/Machine_learning
- Frisullo, P., Laverse, J., Barnabà, M., Navarini, L., และ Del Nobile, M. A. (2012). Coffee beans microstructural changes induced by cultivation processing: An X-ray microtomographic investigation. *Journal of Food Engineering*, 109(1), 175-181.
- Garrett, R., Rezende, C. M., และ Ifa, D. R. (2016). Revealing the spatial distribution of chlorogenic acids and sucrose across coffee bean endosperm by desorption electrospray ionization-mass spectrometry imaging. *LWT - Food Science and Technology*, 65, 711-717.
- O.Duda, E.Hart, และ Stork., G. (2000). Pattern Classification. *2nd Edition*, 488-650.
- Pinto, C., Furukawa, J., Fukai, H., และ Tamura, S. (2017, 16-18 Aug. 2017). *Classification of Green coffee bean images basec on defect types using convolutional neural network (CNN)*. Paper presented at the 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA).
- rabikacoffee. (2561). สายพันธุ์กาแฟ
[http://www.rabikacoffee.com/index.php?lay=show&ac=](http://www.rabikacoffee.com/index.php?lay=show&ac=article&Id=539389394&Ntype=2) article&Id
=539389394&Ntype=2
- Wu, D., และ Sun, D.-W. (2013). Colour measurements by computer vision for food quality

control – A review. *Trends in Food Science & Technology*, 29(1), 5-20.

Zhang, Y., Wang, S., Ji, G., และ Phillips, P. (2014). Fruit classification using computer vision and feedforward neural network. *Journal of Food Engineering*, 143, 167-177.

กรมวิชาการเกษตร. (2561, 2017/04/01/). ยุทธศาสตร์กาแฟ ปี 2559-2563.

<https://www.doa.go.th/hort/wp-content/uploads/2019/11/ยุทธศาสตร์กาแฟ-ปี-2560-2564.pdf>

พิชัยอุตกฤษฎ์, จ., ศรีพุ่ม, ช., สมงาม, ส., สุดประโคน, เ., ชูวงศ์, ป., แก้วมุกดา, ส., . . . หาบ้านแท่น, ว. (2556). กาแฟ Cafe ลักษณะทางพฤกษศาสตร์ของกาแฟ.

<https://sites.google.com/site/kafaeswu/laksana-thang-phvksastr>

ราชกิจจานุเบกษา. (2561). ประกาศกระทรวงเกษตรและสหกรณ์ เรื่อง กำหนดมาตรฐานสินค้าเกษตร : เมล็ดกาแฟอะราบิกา ตามพระราชบัญญัติมาตรฐานสินค้าเกษตร พ.ศ. ๒๕๕๑.

http://www.ratchakitcha.soc.go.th/DATA/PDF/2562/E/006/T_0005.PDF

ประวัติผู้เขียน

ชื่อ-สกุล	แสงดาว เดือนแจ่ม
วัน เดือน ปี เกิด	1 มกราคม 2532
สถานที่เกิด	จังหวัด เชียงใหม่
ที่อยู่ปัจจุบัน	199/148 หมู่บ้านชัยพฤกษ์ ปิ่นเกล้า-สาทร หมู่ 3 ต.บางขุนกอง อ.บางกรวย จ.นนทบุรี 11130
ผลงานตีพิมพ์	-
รางวัลที่ได้รับ	-

