JIST Journal of Information Science and Technology
Volume 12, NO 2 | JUL – DEC 2022 | 1-10
ISSN: 2651-1053 (Online)
ISSN: 1906-9553 (Print)

การจำแนกดอกรักด้วยภาพถ่าย โดยการสกัดคุณลักษณะและการเรียนรู้ของเครื่อง

A Classification of Crown Flower using Feature Extraction and Machine Learning

ศศิน เทียนดี[1] เกวลิน ขำนิพัฒน์[2] และ อรอุมา พร้าโมต[2],*

Sasin Tiendee[1], Kewalin Khamnipat[2] and On-Uma Pramote[2]*

[1]โครงการจัดตั้งภาควิชาคอมพิวเตอร์ คณะศิลปศาสตร์และวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน

[1]Department Computer, Faculty of Art and Science, Kasetsart University, Kamphaeng Saen Campus

[2]หลักสูตรสาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏพิบูลสงคราม

[2]Computer Science Program, Faculty of Science and Technology, Pibulsongkram Rajabhat University

Received: December 17, 2021; Revised: May 26, 2022; Accepted: June 09, 2022; Published: July 25, 2022

ABSTRACT – This research presents an algorithm for image analysis dahlia classification that uses machine learning techniques to learn the distinctive features of dahlia images. The dahlia flowers used in this study were divided into four classes: (1) well-formed flowers, (2) fungal-infected flowers, (3) unequal petals, and (4) basal flowers. The dahlia flowers used in this research were divided into four classes: (1) flowers with good appearance, (2) flowers with fungal attachment, (3) flowers with unequal petal length, and (4) flowers at the base that are adjacent to each other. There are 200 flowers per class, a total of 800 flowers. The main features used in the research were three attributes: (1) percentage of dark pixels, (2) petal area (3) area between petals. Three techniques were used in this study: (1) Decision Trees, (2) Support Vector Machines, and (3) Deep Learning. In this study, 70% of the total data was used to create the classifier, and the remaining 30% was used for the classifier test. The classifier benchmarks with ten-fold cross-validation were precision 95.92, recall 95.90, accuracy 95.89, and f-measure 95.91. Test datasets were used to measure performance and found that decision trees were the most efficient learning machines. The results were precision 99.59 %, recall 99.58%, accuracy 99.58%, and f-measure rate 99.58%.

KEYWORDS: Classification, Feature Extraction, Machine Leaning, Crown Flower

บทคัดย่อ -- **งานวิจัยนี้นำเสนอขั้นตอนวิธีการจำแนกดอกรักด้วยการวิเคราะห์ภาพซึ่งใช้เทคนิคการเรียนรู้ของเครื่องเพื่อเรียนรู้ลักษณะเด่นของภาพดอกรักดอกรักที่ใช้ในงานวิจัยนี้ แบ่งออกเป็น 4 คลาส (1) ดอกที่มีลักษณะดี (2) ดอกที่มีเชื้อราติดอยู่ (3) ดอกที่มีกลีบยาวไม่เท่ากัน (4) ดอกที่ฐานชิดติดกัน ซึ่งมีคลาสละ 200 ดอก รวมทั้งสิ้น 800 ดอก คุณลักษณะเด่นที่ใช้ในงานวิจัย มี 3 คุณลักษณะ ได้แก่ (1) ร้อยละของพิกเซลสีเข้ม (2) พื้นที่กลีบดอก (3) พื้นที่ระหว่างกลีบดอก งานวิจัยนี้ทำการทดลองใช้เครื่องจักรเรียนรู้ 3 เทคนิค คือ (1) ต้นไม้ตัดสินใจ (2) ซัพพอร์ตเวกเตอร์แมทชีน และ (3) การเรียนรู้เชิงลึก งานวิจัยนี้กำหนดให้จำนวนข้อมูลร้อยละ 70 จากทั้งหมด นำมาใช้สร้างตัวจำแนก และจำนวนข้อมูลที่เหลือ 30% นำมาใช้ในการทดสอบตัวจำแนก ผลของการวัดประสิทธิภาพของตัวจำแนกด้วย 10 fold cross validation คือ ค่าความแม่นยำ ร้อยละ 95.92 ค่าความระลึก ร้อยละ 95.90 ค่าความถูกต้อง 95.89 และค่าความถ่วงดุล ร้อยละ 95.91 ชุดข้อมูลทดสอบถูกใช้ในการวัดประสิทธิภาพ พบว่า ต้นไม้ตัดสินใจเป็นเครื่องจักรเรียนรู้ที่มีประสิทธิภาพสูงสุด ผลลัพธ์ คือ ค่าความความแม่นยำ ร้อยละ 99.59 ค่าความระลึก ร้อยละ 99.58 ค่าความถูกต้อง ร้อยละ 99.58 และค่าความถ่วงดุล ร้อยละ 99.58**

คำสำคัญ: การจำแนก การสกัดคุณลักษณะเด่น การเรียนรู้ของเครื่อง ดอกรัก

*Corresponding Author: onbee@psru.ac.th

1. บทนำ

ในประเทศไทยดอกรักเป็นดอกไม้ที่ใช้ในการร้อยมาลัยหรือใช้ในงานมงคล ดอกรักแบ่งได้ตามสีของดอก 2 คือสีขาวและสีม่วงหรือม่วงแดง แต่ดอกรักสีขาวนิยมนำมาใช้มากกว่า ดอกรักมีลักษณะเป็นช่อแบบซี่ร่ม ออกตามซอกใบหรือปลายกิ่ง มีกลีบเลี้ยง 5 กลีบ โคนเชื่อมติดกัน เมื่อบานจะมีเส้นผ่าศูนย์กลาง 2 – 3 เซนติเมตร ให้ผลผลิตตลอดปี ปัจจุบันมีความต้องการใช้ดอกรักอย่างต่อเนื่อง จึงมีการปลูกดอกรักในเชิงพาณิชย์กันมากขึ้น ราคาดอกรักปรับขึ้นลงตามราคาตลาด ช่วงที่ดอกรักให้ผลผลิตน้อยราคาดอกรักจะสูง เฉลี่ยกิโลกรัมละ 120-200 บาท หากเป็นช่วงที่ดอกรักให้ผลผลิตมากราคาจะลดลงเหลือกิโลกรัมละ 20-60 บาท ซึ่งราคาที่แตกต่างมาจากขนาดของดอก การเก็บดอกรักกลายเป็นอาชีพเสริมของเกษตรกร (ชาวนา) ระหว่างรอการปลูกข้าว โดยเฉพาะในพื้นที่ตำบลหัวดง อำเภอเมือง จังหวัดพิจิตร เมื่อเกษตรกรเก็บดอกรักมาแล้วต้องคัดเฉพาะดอกที่ไม่มีรา มีจำนวนกลีบ 5 กลีบความยาวกลีบต้องเท่ากัน ฐานกลีบมีระยะห่างไม่ชิดกัน และมีขนาดที่ไม่เล็กเกินไป เพื่อให้พ่อค้าคนกลางมารับซื้อ ถ้าเกษตรไม่คัดแยกดอกรักพ่อค้าจะซื้อแบบเหมารวม ทำให้ขายไม่ได้ราคา แต่ถ้าเกษตรกรคัดดอกจะสามารถขายในราคาที่สูงขึ้นได้

ศาสตร์ด้านปัญญาประดิษฐ์กำลังเป็นที่สนใจของคนทั่วโลกร่วมถึงรัฐบาลไทยที่พยายามขับเคลื่อนให้ประเทศเข้าสู่เศรษฐกิจดิจิทัล (Digital Economy) หรือ ไทยแลนด์ 4.0 (Thailand 4.0) เน้นให้เกษตรกรหรือผู้ประกอบการใช้เทคโนโลยีมาช่วยในการทำงานเพื่อให้ได้ผลกำไรที่เพิ่มขึ้น ซึ่งเทคโนโลยีปัญญาประดิษฐ์ที่นิยมใช้ คือ เครื่องจักรเรียนรู้ (Machine Learning) ที่สามารถช่วยจำแนก คัดแยก จัดกลุ่ม หรือทำนายผลลัพธ์ได้ โดยในประเทศไทยได้นำเครื่องจักรเรียนรู้เข้ามาช่วยหลายด้านได้แก่ ด้านการแพทย์ [1-3] ด้านเกษตรกรรม [4-5] ด้านอุตสาหกรรม [6] หรือด้านความปลอดภัย [7] ปัจจัยสำคัญที่ทำให้เครื่องจักรเรียนรู้จำแนก หรือทำนายข้อมูลได้ คือ คุณลักษณะเด่น (Feature) โดยต้องทำการสกัดจากข้อมูลที่เราสนใจด้วยวิธีการต่างๆ

จากปัญหาข้างต้นและเทคโนโลยีในปัจจุบันผู้วิจัยจึงขอเสนอการจำแนกดอกรักด้วยการวิเคราะห์ภาพถ่าย โดยใช้คุณลักษณะเด่นและเครื่องจักรเรียนรู้เพื่อช่วยการคัดในการจำแนกดอกรัก ให้มีความถูกต้องแม่นยำ ช่วยเพิ่มรายได้ของเกษตรกรและสอดรับกับนโยบายของประเทศ

ส่วนต่อไปของบทความนี้มีรายละเอียดได้แก่ ส่วนที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง ส่วนที่ 3 วิธีการวิจัย ส่วนที่ 4 ผลการทดลอง ส่วนที่ 5 บทสรุปและการอภิปราย

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 การประมวลผลภาพ (Image Processing)

ภาพถูกแทนเป็นข้อมูล 2 มิติ ในรูปแบบตารางข้อมูล บอกตำแหน่งช่องในตารางด้วยค่าแนวตั้งและค่าแนวนอนซึ่งในแต่ละช่องจะเก็บค่าตัวเลขที่แสดงถึงความเข้มแสง (Intensity) เพื่อให้คอมพิวเตอร์สามารถเข้าใจได้ การประมวลผลภาพ คือ การใช้กระบวนการทางคณิตศาสตร์มากระทำต่อข้อมูลภาพเพื่อให้เกิดผลลัพธ์หรือภาพใหม่ที่ดีขึ้นทั้งในสายตาของมนุษย์หรือในการนำข้อมูลไปประมวลผลต่อ [8] การประมวลผลภาพแบ่งได้ 5 แบบ ได้แก่ การได้มาซึ่งภาพ การปรับปรุงภาพ การกู้คืนภาพ การแบ่งส่วนภาพ และการบีบอัดภาพ ในงานวิจัยนี้ ใช้การได้มาซึ่งภาพ และการแบ่งส่วนภาพ

2.2 การสกัดคุณลักษณะเด่น (Feature Extraction)

มนุษย์สามารถจำแนกวัตถุต่างๆ ออกจากกันได้ โดยอาศัยรูปร่างหรือลักษณะทางกายภาพที่แตกต่างกันของวัตถุนั้นๆ ดังนั้นการทำให้คอมพิวเตอร์จำแนกได้เหมือนมนุษย์ต้องวัดหรือแสดงความแตกต่างของรูปแบบออกมาเป็นตัวเลข ซึ่งเรียกว่าลักษณะเด่น (Feature) โดยการสกัดคุณลักษณะเด่นจากภาพ มี 3 แบบ คือ คุณลักษณะสี (Color) คุณลักษณะรูปร่าง (Shape) คุณลักษณะพื้นผิว (Texture) การจำแนกอาจจำเป็นต้องใช้ลักษณะเด่นมากกว่า 1 อย่าง ซึ่งจะแสดงลักษณะเด่นทั้งหมดที่ใช้ในรูปแบบเวกเตอร์ลักษณะเด่น (Feature Vector) [9],[10]

2.3 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่อง หรือเครื่องจักรเรียนรู้ สามารถแบ่งได้ 2 กลุ่ม ได้แก่ กลุ่มที่หนึ่ง การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นการแบ่งกลุ่มข้อมูล (Clustering) กลุ่มที่สอง การเรียนรู้แบบมีผู้สอน (Supervised Learning) แบ่งย่อยได้อีก 2 แบบ คือ การจำแนกหรือการแยกประเภทข้อมูล (Classification) และ การวิเคราะห์การถดถอย (Regression) [11-13]

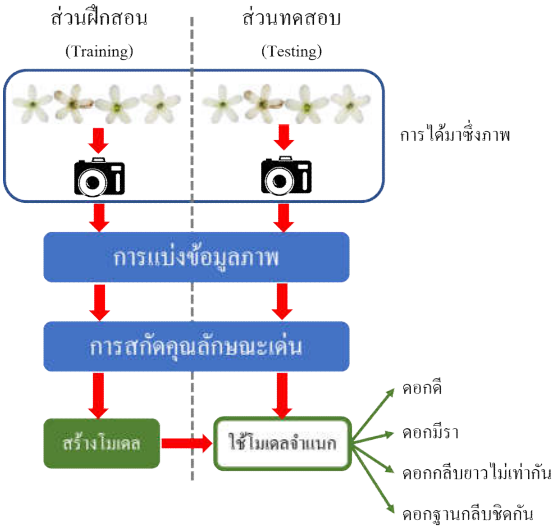
2.4 งานวิจัยที่เกี่ยวข้อง

ในการศึกษางานวิจัยที่เกี่ยวข้องกับการจำแนกดอกไม้สามารถแบ่งออกเป็น 2 กลุ่ม ได้แก่ 1) งานวิจัยที่ใช้เครื่องจักรเรียนรู้ที่มีขั้นตอนการสกัดคุณลักษณะเด่นร่วมอยู่ในเทคนิคด้วย ซึ่งมีงานวิจัยดังนี้ [14] การจำแนกพันธุ์กล้วยไม้ ด้วยเครื่องจักรเรียนรู้โครงข่ายประสาทเทียมแบบคอนโวลูชัน มีค่าความถูกต้องร้อยละ 92.2 [15] การจำแนกพรรณไม้ที่มีอยู่ในสิ่งแวดล้อม ใช้เครื่องจักรเรียนรู้โครงข่ายประสามเทียมแบบคอนโวลูชัน 4 โครงสร้าง ผลการวิจัยพบว่า โครงสร้างที่เหมาะสมที่สุดในการใช้งาน คือ Google Net 2) งานวิจัยที่สกัดคุณลักษณะเด่นด้วยวิธีการประมวลผลภาพ เพื่อนำไปใช้กับเครื่องจักรเรียนรู้ที่รับข้อมูลเป็นตัวเลข ซึ่งมีงานวิจัยดังนี้ [16] การจำแนกภาพดอกเดหลี ซึ่งใช้คุณลักษณะเด่น 3 อย่าง ได้แก่ สี รูปร่างดอก และระยะห่างระหว่างจุด ใช้เครื่องจักรเรียนรู้โครงข่ายประสาทเทียม มีค่าความถูกต้องร้อยละ 88.67 [17] การจำแนกดอกกล้วยไม้รองเท้านารีพันธุ์พื้นเมือง ศึกษาการแบ่งส่วนภาพเพื่อนำภาพไปสกัดคุณลักษณะเด่น คือ คุณลักษณะสี โดยใช้โมเดลสีแบบ HSV คุณลักษณะรูปร่างโดยใช้อัตราส่วนและค่าความกระทัดรัด (Compactness) ซึ่งผลการวิจัยพบว่าเทคนิคการแบ่งส่วนภาพด้วยการขีดแบ่ง และเทคนิคการแบ่งเส้นภาพด้วยขอบ เป็นเทคนิคที่ดีที่สุดในการแบ่งส่วนภาพที่ควรนำไปใช้สกัดคุณลักษะต่อ

จากงานวิจัยข้างต้นยังไม่มีงานวิจัยการจำแนกดอกรัก และการสกัดคุณลักษณะเด่นสามารถทำงานกับเครื่องจักรเรียนรู้ได้หลายแบบ ซึ่งการสกัดคุณลักษณะเด่นเองทำให้สามารถเลือกใช้คุณลักษณะเด่นได้อย่างอิสระ และใช้จำนวนคุณลักษณะเด่นไม่มากในการจำแนก ส่งผลให้จำแนกได้เร็ว

3. วิธีการดำเนินงาน

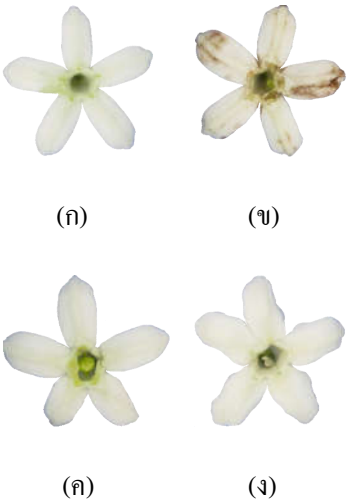
ขั้นตอนการจำแนกดอกรักด้วยภาพถ่าย ประกอบด้วย 4 ขั้นตอน คือ 1) การได้มาซึ่งภาพ 2) การแบ่งข้อมูลภาพ 3) การสกัดคุณลักษณะเด่น และ 4) การการสร้างและใช้งานโมเดลจำแนก ซึ่งในขั้นตอนที่ 4 จะถูกแบ่งออกเป็น 2 ส่วน คือ ส่วนฝึกสอน (Training) และส่วนทดสอบ (Testing) ในส่วนฝึกสอน จะนำข้อมูลคุณลักษณะเด่นไปฝึกโมเดลจำแนก ในส่วนทดสอบจะนำข้อมูลคุณลักษณะเด่นไปทดสอบประสิทธิภาพของโมเดลที่ได้จากการฝึก ดังแสดงในรูปที่ 1



รูปที่ 1. แสดงขั้นตอนการทำงานของระบบ

3.1 ข้อมูลดอกรัก (Data Set)

พันธุ์ดอกรักที่ใช้ในงานวิจัยนี้เป็นพันธุ์รักแก้ว สีขาว มีกลีบ 5 กลีบ แบ่งเป็น 4 คลาส ดังรายละเอียดในตารางที่ 1 โดยผู้เชี่ยวชาญเป็นผู้จำแนกให้ จำนวนดอกรักที่ใช้ในแต่คลาสเท่ากับ 200 ดอก รวมทั้งสิ้น 800 ดอก



(ก) (ข)

(ค) (ง)

รูปที่ 2. (ก) ดอกดี (ข) ดอกมีรา (ค) ดอกที่กลีบยาวไม่เท่ากัน (ง) ดอกที่ฐานชิดติดกัน

ตารางที่ 1. แสดงรายละเอียดของดอกรักแต่ละคลาส

คลาส	ลักษณะ
ดอกดี (Good)	กลีบสีขาวไม่มีสีดำหรือสีน้ำตาลปน กลีบทั้ง 5 มีความยาวใกล้เคียงกัน ฐานกลีบมีระยะห่างไม่ชิดติดกัน
ดอกมีรา (Fungi)	กลีบดอกมีสีดำหรือสีน้ำตาลปน
ดอกที่กลีบยาวไม่เท่ากัน (Small)	กลีบมีความยาวจากฐานถึงปลายกลีบไม่เท่ากัน
ดอกที่ฐานชิดติดกัน (Based)	ฐานกลีบชิดติดกัน หรือมีระยะห่างระหว่างฐานกลีบน้อย

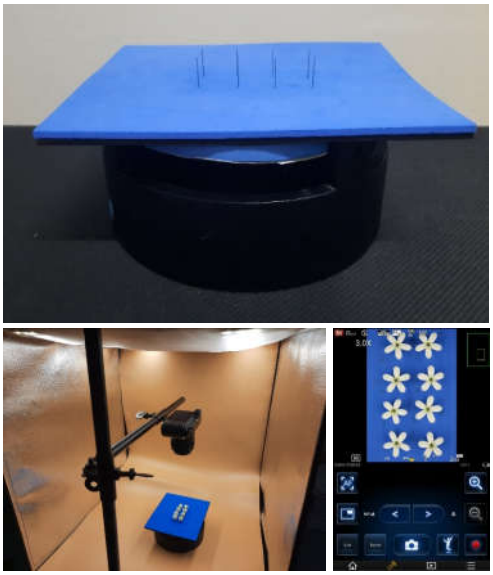
งานวิจัยนี้แยกคุณลักษณะเด่นของดอกรักตามตารางที่ 2

ตารางที่ 2. แสดงคุณลักษณะเด่นของดอกรัก

คลาส	คุณลักษณะเด่น		
	สีกลีบ	กลีบแคระ	ฐานของกลีบดอก
ดอกดี (Good)	ขาว	ไม่มี	แยก
ดอกมีรา (Fungi)	ดำ	ไม่มี	แยก
ดอกที่กลีบยาวไม่เท่ากัน (Small)	ขาว	มี	แยก
ดอกที่ฐานชิดติดกัน (Based)	ขาว	ไม่มี	ติด

3.2 การได้มาซึ่งภาพ (Image Acquisition)

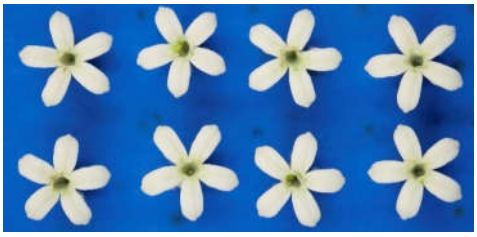
งานวิจัยนี้สร้างอุปกรณ์เพิ่มเติมสำหรับการถ่ายภาพและกำหนดสภาพแวดล้อม ค่าพารามิเตอร์ของกล้องในการถ่าย ดังนี้ แท่นทรงกลมสูง 8 ซม. เพื่อวางดอกรัก พื้นหลังสีฟ้าสำหรับรองถ่าย ขนาด 16 x 18.5 ซม. ใช้เข็มหมุดเป็นตัวยึดดอกรัก วางดอกรักในลักษณะหงายฐานดอกขึ้น เรียง 2 แถว แถวละ 4 ดอก กล้องถ่ายภาพจะยึดกับก้านของฐานถ่าย โดยถ่ายจากมุมบน (Top view) ห่างจากแท่นวางดอกรัก 40 ซม. ซึ่งถ่ายในกล่องสตูดิโอที่ติดตั้งไฟแอลอีดี (LED) ขนาด 1690 ลูเมน อุณหภูมิ 5500 เคลวิน จำนวน 40 หลอด พารามิเตอร์กล้องกำหนดค่าดังนี้ ใช้โหมดโฟกัส MF ค่ารูรับแสง เท่ากับ F2.8 ค่าแสงไอโอเอส(ISO) เท่ากับ 400 เวลาชัตเตอร์ (Exposure time) เท่ากับ 1/400s ใช้เลนน์มาโคร ระยะเลนส์ซูม 45 มม. เพื่อไม่ให้เกิดการขยับของตัวกล้องและฐานถ่ายจึงใช้การควบคุมไร้สายในการเก็บภาพจากแท็บแล็ต ดังแสดงในรูปที่ 3



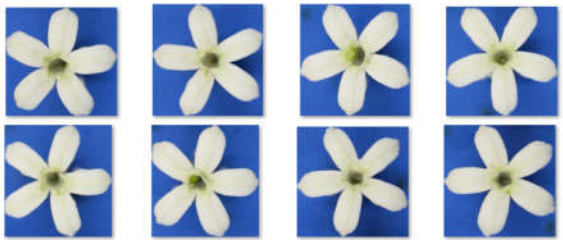
รูปที่ 3. แสดงอุปกรณ์และสภาพแวดล้อมในการได้มาซึ่งภาพ

3.3 การแบ่งข้อมูลภาพ (Image Cropping)

ใช้การประมวลผลภาพ เพื่อเตรียมภาพให้พร้อมก่อนทำการสกัดคุณลักษณะเด่น โดยเลือกช่องสีแดง (Red Channel) ของภาพเป็นหลักและแยกพื้นหลังกับดอกรักด้วยการหาค่าเทรชโฮลด์ (Threshold) จากนั้นใช้การหาบริเวณที่สนใจ (Region of Interest) เพื่อตีกรอบล้อมดอกรักแต่ละดอกแล้วทำการตัด (Cropping) ภาพที่นำเข้ามาในขั้นตอนนี้มีขนาด ความกว้าง 4592 ความสูง 3064 พิกเซล ความละเอียดของภาพ 180 ดีพีไอ (Dot per inch: dpi) เป็นไฟล์ภาพ เจพีจี (JPG)



(ก)



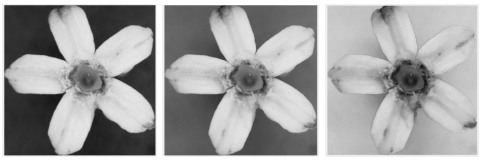
(ข)

รูปที่ 4. (ก) ภาพก่อน (ข) ภาพหลังการเตรียมข้อมูลภาพ

3.4 การสกัดคุณลักษณะเด่น (Feature Extraction)

การสกัดคุณลักษณะเด่นของดอกรักนำเสนอ คุณลักษณะเด่นสี คุณลักษณะเด่นพื้นที่กลีบดอก และคุณลักษณะเด่นพื้นที่ระหว่างกลีบ โดยแยกคุณลักษณะกับการจำแนกดอกรักแต่ละกลุ่ม ดังนี้ คุณลักษณะเด่นสีใช้แยกดอกมีรา คุณลักษณะเด่นพื้นที่กลีบดอกใช้แยกดอกที่มีความยาวกลีบไม่เท่ากัน และคุณลักษณะเด่นพื้นที่ระหว่างกลีบใช้แยกดอกที่ฐานชิดติดกัน การสกัดคุณลักษณะเด่นจะพิจารณาเฉพาะกลีบดอกตามความรู้จากผู้เชี่ยวชาญ จึงไม่นำส่วนกลางดอกมาร่วมพิจารณาด้วย การสกัดคุณลักษณะเด่นประกอบด้วยขั้นตอน ดังนี้

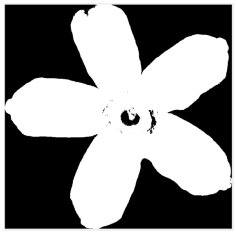
1) ใช้ช่องสี 2 ช่อง คือ สีแดง (Red) และสีน้ำเงิน (Blue) สีขาวของดอกรักในช่องสีแดงถูกแบ่งออกจากพื้นหลักได้ชัดเจน สามารถใช้แบ่งพื้นหลังกับดอกรักได้ดี ส่วนช่องสีน้ำเงินสามารถแสดงสีดำหรือน้ำตาลในดอกรักได้ชัดเจน เมื่อนำข้อมูลดอกรักจาก 2 ช่องสีรวมกันจะทำให้ได้พื้นที่ดอกรักสมบูรณ์



(ก) (ข) (ค)

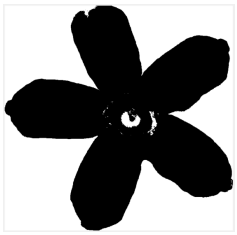
รูปที่ 5. (ก) ดอกรักช่องสีแดง (ข) ดอกรักช่องสีเขียว และ (ค) ดอกรักช่องสีน้ำเงิน

2) เลือกพิกเซลข้อมูลเฉพาะดอกรักในช่องสีแดง โดยแปลงข้อมูลภาพจากภาพระดับสีเทาเป็นภาพขาวดำ (ดอกรักสีขาว พื้นหลังสีดำ) ด้วยการหาค่าเทรดโฮลด์ และหาจุดเซนทรอยของภาพเก็บไว้ เพื่อใช้สร้างวงกลมปิดกลางดอกรัก



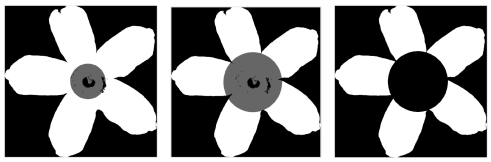
รูปที่ 6. ภาพของดอกรักที่แบ่งจากพื้นหลัง

3) สลับค่าพิกเซลของพื้นหลังและ ดอกรัก จากนั้นทำการหาเซนทรอยของวัตถุในภาพ (สีขาวกลางภาพ)



รูปที่ 7. การสลับค่าพื้นหลังกับดอกรัก

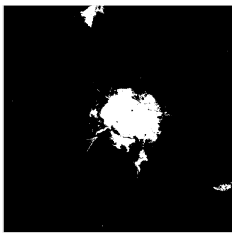
4) นำจุดเซนทรอยที่ได้จาก 2 ภาพ มาหาค่าเฉลี่ย เพื่อใช้เป็นจุดศูนย์กลางของวงกลมปิดกลางดอกรัก สร้างวงกลมที่ โดยกำหนดรัศมีเริ่มต้น 100 หน่วย แล้วทำการขยายรัศมีขึ้นทีละ 10 หน่วย จนกว่าวงกลมจะสามารถตัดแบ่งกลีบดอกรักได้ 5 กลีบ



(ก) (ข) (ค)

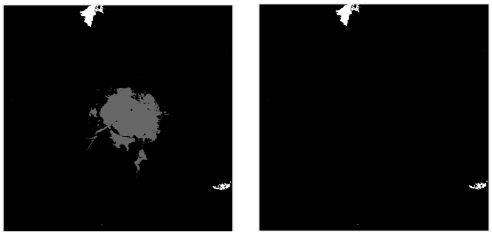
รูปที่ 8. (ก) วงกลมขนาดเริ่มต้น (ข) ขนาดวงกลมที่สามารถแบ่งกลีบได้ 5 กลีบ และ (ค) ดอกรักที่ตัดแบ่งกลีบและฐานดอก

5) เลือกช่องสีน้ำเงิน โดยพิกเซลสีดำหรือสีน้ำตาลจะถูกเลือกเป็นข้อมูลดอกรัก จากนั้นใช้เทคนิคเทรดโฮลด์ แบ่งข้อมูลดอกรักกับพื้นหลัง ส่งผลให้ส่วนกลางของดอกถูกเลือกมาด้วย



รูปที่ 9. พิกเซลสีดำหรือสีน้ำตาลของดอกรัก

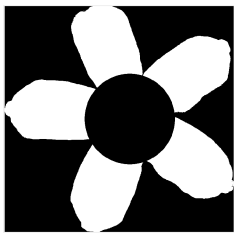
6) สร้างวงกลมจากเซนทรอยเดิม โดยมีรัศมี 1.5 เท่าของขนาดรัศมีที่แบ่งกลีบดอกได้ 5 กลีบ ทำให้ตัดได้ครอบคลุมกลางของดอกรัก คงเหลือแต่พิกเซลที่กลีบดอกรัก



(ก) (ข)

รูปที่ 10. (ก) ก่อนใช้วงกลมตัด (ข) หลังใช้วงกลมตัด

7) นำข้อมูลของกลีบดอกรักและพื้นที่สีเข้มบนกลีบดอกมาร่วมกัน และทำการเติมช่องว่างของกลีบดอกทั้งหมดเพื่อความสมบูรณ์



รูปที่ 11. ดอกรักที่พร้อมสกัดคุณลักษณะเด่น

8) การสกัดคุณลักษณะเด่นสี พิจารณาพิกเซลสีดำและสีน้ำตาลในกลีบดอกต่อจำนวนพิกเซลกลีบดอกรักทั้งหมด (5 กลีบ) ได้เป็นค่าร้อยละของพิกเซลสีเข้ม (Percent) ซึ่งใช้เป็นคุณลักษณะเด่นในการจำแนกดอกมีรา

$$Precent = \frac{Pixel\ of\ Black\ or\ Brown}{All\ Pixel\ of\ flower\ petals}$$

9) การสกัดคุณลักษณะเด่นพื้นที่กลีบดอก พิจารณาดังนี้ (1) หาพื้นที่กลีบดอกทั้ง 5 กลีบ (2) หาค่ามัธยฐาน (Median) (3) หาผลต่างของค่ามัธยฐานกับพื้นที่กลีบดอกที่น้อยที่สุด ได้เป็นค่าพื้นที่กลีบดอก (Area of petal) ซึ่งใช้เป็นคุณลักษณะเด่นในการจำแนกดอกที่มีความยาวกลีบไม่เท่ากัน

$$A = \{a_1, a_2, a_3, a_4, a_5\}$$

โดยที่ A คือ เซตของพื้นที่กลีบดอกรัก ทั้ง 5 กลีบ
a คือ จำนวนพิกเซลในพื้นที่กลีบดอกแต่ละกลีบ

$$Area\ of\ Petal = Median(A) - \text{Min}\ (A)$$

3) การสกัดคุณลักษณะเด่นพื้นที่ระหว่างกลีบ ใช้พื้นที่ว่างระหว่างกลีบมีลักษณะคล้ายรูปสามเหลี่ยมพิจารณา ดังนี้ (1) หาพื้นที่ระหว่างกลีบดอกทั้ง 5 กลีบ (2) หาค่าพื้นที่ระหว่างกลีบที่น้อยที่สุด (Minimize) ซึ่งใช้เป็นคุณลักษณะเด่นในการจำแนกดอกที่ฐานชิดติดกัน

$$T = \{t_1, t_2, t_3, t_4, t_5\}$$

โดยที่ T คือ เซตของพื้นที่ระหว่างกลีบดอกทั้ง 5 กลีบ
t คือ จำนวนพิกเซลระหว่างพื้นที่กลีบแต่ละกลีบ

$$Distance\ of\ Petal = \text{Min}\ (T)$$

3.5 การแบ่งข้อมูลเพื่อทดลอง

การแบ่งข้อมูลเพื่อทดลองในงานวิจัยมีรายละเอียดดังนี้แบ่งข้อมูลสำหรับฝึกสอน ร้อยละ 70 ของข้อมูลทั้งหมด 800 ดอก ได้จำนวน 560 ดอก ที่เหลือร้อยละ 30 เป็นข้อมูลที่ใช้ทดสอบจำนวน 240 ดอก มีรายละเอียดดังตารางที่ 3

ตารางที่ 3. แสดงจำนวนดอกรักที่ใช้ในการทดลอง

ประเภทข้อมูล	จำนวนดอกรักแต่ละคลาส			
	Good	Fungi	Small	Based
ข้อมูลฝึกสอน	140	140	140	140
ข้อมูลทดสอบ	60	60	60	60

โดยในการฝึกสอนข้อมูลจะถูกนำเข้าด้วยวิธี การตรวจสอบไขว้สลับ ซึ่งแบ่งข้อมูลออกเป็น 10 กลุ่ม (10 fold Cross validation) ชุดข้อมูลมีลักษณะเป็นตาราง แนวนอนคือข้อมูลดอกรักแต่ละดอก แนวตั้งคือคุณลักษณะเด่น 3 คุณลักษณะ ดังตัวอย่างในตารางที่ 4

ตารางที่ 4. แสดงตารางตัวอย่างคุณลักษณะเด่นและผลเฉลย

ชุดข้อมูล	Precent	Area	Base	Class
ดอกที่ 1	0.5421	876	8823	Good
ดอกที่ 2	0.0582	6734	6897	Good
ดอกที่ 3	13.6217	4430	11127	Fungi
ดอกที่ 4	9.5397	774	10732	Fungi
ดอกที่ 5	0.0925	14592	8897	Small
ดอกที่ 6	0.1416	17259	11124	Small
ดอกที่ 7	0.0938	5546	4843	Base
ดอกที่ 8	0.1029	3124	1689	Base

3.7 การวัดประสิทธิภาพเครื่องจักรเรียนรู้

การวัดประสิทธิภาพของเครื่องจักเรียนรู้ ประกอบด้วย 3 ค่า ได้แก่ ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถูกต้อง (Accuracy) ซึ่งเป็นประสิทธิภาพของเครื่องจักเรียนรู้กับข้อมูลทดสอบ [18]

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

โดยที่ TP คือ ค่าพยากรณ์ ถูกต้องเชิงบวก (ทำนายว่าจริง และสิ่งที่เกิดขึ้น คือ จริง)

TN คือ ค่าพยากรณ์ถูกต้องเชิงลบ (ทำนายว่าไม่จริง และสิ่งที่เกิดขึ้น คือ ไม่จริง)

FP คือ ค่าพยากรณ์ผิดพลาดเชิงบวก (ทำนายว่าจริง แต่สิ่งที่เกิดขึ้น คือ จริง)

FN คือ ค่าพยากรณ์ผิดพลาดเชิงลบเชิงลบ (ทำนายว่าไม่จริง แต่สิ่งที่เกิดขึ้น คือ ไม่จริง)

4. ผลการทดลอง

การใช้คุณลักษณะเด่นทั้ง 3 คุณลักษณะสกัดค่าจากดอกรัก4 คลาส ได้แก่ ดอกดี (Good) ดอกมีรา (Fungi) ดอกที่ความยาวกลีบไม่เท่ากัน (Small) และดอกที่ฐานชิดติดกัน (Based) สามารถแสดงผลโดยใช้ตัวอย่างดอกรักแต่ละคลาส จำนวน 8 ดอก มีรายละเอียดดังนี้

4.1 ผลการสกัดคุณลักษณะเด่นสี

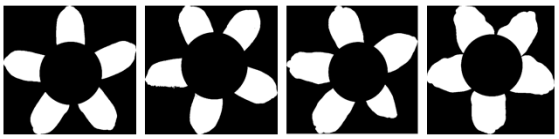
การสกัดคุณลักษณะเด่นสี เพื่อจำแนกราในดอกรักทั้ง 4 คลาส ได้แก่ Good Fungi Small และ Based มีรายละเอียดในตารางที่ 5 พบว่าค่าร้อยละของพิกเซลสีเข้ม ในคลาส Fungi จะมีค่าร้อยละสูงกว่าดอกรักทั้ง 3 คลาส ซึ่งดอกมีรา ค่าของร้อยละจะเกินร้อยละ 2

ตารางที่ 5. แสดงค่าร้อยละพิกเซลสีเข้มของดอกรัก 4 คลาส

ชุดข้อมูล	Good	Fungi	Small	Based
ดอกที่ 1	0.203	6.135	0.022	0.048
ดอกที่ 2	0.542	13.494	0.093	0.094
ดอกที่ 3	0.058	13.622	0.142	0.103
ดอกที่ 4	0.117	13.371	0.105	0.088
ดอกที่ 5	0.098	9.540	0.170	0.103
ดอกที่ 6	0.084	2.947	0.113	0.527
ดอกที่ 7	0.163	9.254	0.392	0.087
ดอกที่ 8	0.124	21.717	0.159	0.060

4.2 ผลการสกัดคุณลักษณะเด่นของพื้นที่กลีบ

การสกัดคุณลักษณะเด่นของพื้นที่กลีบ เพื่อจำแนกดอกที่ความยาวกลีบไม่เท่ากันในดอกรักทั้ง 4 คลาส ดังภาพที่ 12 แสดงพื้นที่ทั้ง 5 กลีบของคลาส Good Fungi และ Based มีขนาดใกล้เคียงกัน ยกเว้น คลาส Small ซึ่งพิจารณารายละเอียดในตารางที่ 6 พบว่าค่าของพื้นที่ของกลีบดอกรักคลาส Small จะสูงกว่าดอกรักทั้ง 3 คลาส ซึ่งดอกที่ความยาวกลีบไม่เท่ากัน (Small) ให้ค่าความต่างระหว่างพื้นที่กลีบที่น้อยที่สุดกับค่ามัธยฐานของพื้นที่กลีบทั้ง 5 กลีบ เกิน 10000 พิกเซล ซึ่งในดอกรักทั้ง 3 คลาสที่เหลือจะมีค่าพื้นที่น้อยกว่า



Good Fungi Small Based

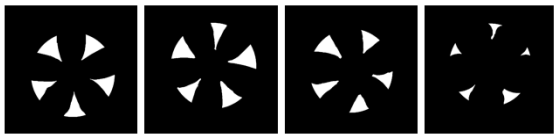
รูปที่ 12. พื้นที่กลีบของดอกรักทั้ง 4 คลาส

ตารางที่ 6. แสดงค่าพื้นที่ของกลีบของดอกรัก 4 คลาส

ชุดข้อมูล	Good	Fungi	Small	Based
ดอกที่ 1	1757	726	17441	4013
ดอกที่ 2	876	4000	14592	5546
ดอกที่ 3	6734	4430	17259	3124
ดอกที่ 4	4010	774	12005	3994
ดอกที่ 5	7923	1513	13492	2146
ดอกที่ 6	2513	1696	26465	5059
ดอกที่ 7	3419	766	10312	1876
ดอกที่ 8	5876	1370	19455	9444

4.3 ผลการสกัดคุณลักษณะเด่นพื้นที่ระหว่างกลีบ

การสกัดคุณลักษณะเด่นพื้นที่ระหว่างกลีบ เพื่อจำแนกดอกที่ฐานชิดติดกันในดอกรักทั้ง 4 คลาส ดังภาพที่ 13 แสดงพื้นที่ระหว่างกลีบ (รูปสามเหลี่ยม) ของคลาส Based มีจำนวนน้อยกว่า 3 คลาสที่เหลือ พิจารณารายละเอียดในตารางที่ 7 พบว่าค่าพื้นที่ระหว่างกลีบที่น้อยที่สุดจากกลีบทั้ง 5 ของคลาส Based จะมีค่าพื้นที่น้อยกว่า 5000 พิกเซล ซึ่งในดอกรักทั้ง 3 คลาสที่เหลือจะมีค่าพื้นที่มากกว่า



Good Fungi Small Based

รูปที่ 13. พื้นที่ระหว่างกลีบของดอกรักทั้ง 4 คลาส

ตารางที่ 7. แสดงค่าพื้นที่ระหว่างกลีบของดอกรัก 4 กลุ่ม

ชุดข้อมูล	Good	Fungi	Small	Based
ดอกที่ 1	9625	8550	7859	4934
ดอกที่ 2	8823	10101	8897	4843
ดอกที่ 3	6897	11127	11124	1689
ดอกที่ 4	8693	11233	11412	1956
ดอกที่ 5	9350	10436	10552	3806
ดอกที่ 6	8524	10843	8525	1595
ดอกที่ 7	10995	11075	7846	3106
ดอกที่ 8	8896	10715	12870	4900

4.4 ประสิทธิภาพของเครื่องจักร

งานวิจัยนี้ใช้เครื่องจักรเรียนรู้ในการทดลอง 3 ชนิด ได้แก่ ต้นไม้ตัดสินใจ (Decision Tree) ซัพพอร์ตเวกเตอร์ แมชชีน (SVM) และการเรียนรู้เชิงลึก (Deep Learning) เพื่อเป็นทดสอบว่าคุณลักษณะทั้ง 3 สามารถทำงานกับเครื่องจักรเรียนรู้ได้หลายชนิด โดยใช้ซอฟต์แวร์แรพบิทไมเนอร์ (RapidMiner) และใช้ค่าพื้นฐานของเครื่องจักรได้ผลการทดลองตามตารางที่ 8-12

ตารางที่ 8. แสดงค่าประสิทธิภาพของเครื่องจักร 3 ชนิด

		DT	SVM	DL
Train Data	Accuracy	95.89	90.89	92.14
	F1	95.91	91.58	92.58
Test Data	Precision	99.59	96.02	99.19
	Recall	99.58	95.42	99.17
	Accuracy	99.58	95.42	99.17
	F1	99.58	95.71	99.18

ตารางที่ 9. แสดงผลต้นไม้ตัดสินใจแต่ละคลาส

	Actual Good	Actual Fungi	Actual Small	Actual Based
Predicted Good	59	0	0	0
Predicted Fungi	0	60	0	0
Predicted Small	1	0	60	0
Predicted Based	0	0	0	60

ตารางที่ 10. แสดงผลซัพพอร์ตเวกเตอร์แมชชีทแต่ละคลาส

	Actual Good	Actual Fungi	Actual Small	Actual Based
Predicted Good	59	0	0	0
Predicted Fungi	0	50	0	0
Predicted Small	0	10	60	0
Predicted Based	1	0	0	60

ตารางที่ 11. แสดงผลการเรียนรู้เชิงลึกแต่ละคลาส

	Actual Good	Actual Fungi	Actual Small	Actual Based
Predicted Good	60	0	1	0
Predicted Fungi	0	58	0	0
Predicted Small	0	2	60	0
Predicted Based	0	0	0	60

เครื่องจักรเรียนรู้ที่มีประสิทธิภาพสูงสุด คือ ต้นไม้ตัดสินใจ โดยมีความแม่นยำ ร้อยละ 99.59 ค่าความระลึก ร้อยละ 99.58 ค่าความถูกต้อง ร้อยละ 99.58 และค่าถ่วงดุล ร้อยละ 99.58 เครื่องจักรเรียนรู้ซัพพอร์ตเวกเตอร์แมทชีนมีประสิทธิภาพต่ำสุด ให้ค่าความแม่นยำ ร้อยละ 96.02 ค่าความระลึก ร้อยละ 95.42 ค่าความถูกต้อง ร้อยละ 95.42 และค่าถ่วงดุล ร้อยละ 95.71

5. บทสรุปและการอภิปราย

งานวิจัยนี้ศึกษาการจำแนกดอกรักด้วยภาพถ่าย โดยใช้การสกัดคุณลักษณะเด่นและเครื่องจักรเรียนรู้ ซึ่งคุณลักษณะเด่นที่ใช้มี 3 คุณลักษณะ คือ ร้อยละของพิกเซลสีเข้ม พื้นที่กลีบ และพื้นที่ระหว่างกลีบ ผ่านเครื่องจักรเรียนรู้ 3 ชนิด คือ ต้นไม้ตัดสินใจ ซัพพอร์ตเวกเตอร์แมชชีน และ การเรียนรู้เชิงลึก พบว่าเครื่องจักรทั้ง 3 ชนิด ให้ค่าความถูกต้องเกินร้อยละ 90 ทั้งนี้การเรียนรู้เชิงลึกและต้นไม้ตัดสินใจเป็นเครื่องจักรที่ให้ค่าความถูกต้องสูงสุด จากการพิจารณาในรายละเอียดของการจำแนกในแต่ละเครื่องจักรมีการจำแนกผิดดังนี้ ซัพพอร์ตเวกเตอร์แมชชีทจำแนกดอกราเป็นดอกกลีบยาวไม่เท่ากัน เนื่องจาก คุณลักษณะเด่นพื้นที่กลีบของดอกมีราที่จำแนกผิดนั้นมีค่าเกินกว่า 10000 พิกเซล ซึ่งถูกให้ค่าความสำคัญในการจำแนกมากกว่าคุณลักษณะเด่นสี เครื่องจักรจึงจำแนกให้อยู่ในคลาส Small แทนคลาส Fungi โดยในการเรียนรู้เชิงลึกก็จำแนกผิดในลักษณะเดียวกันเพียงแต่จำแนกผิดจำนวนน้อยกว่าซัพพอร์ตเวกเตอร์แมชชีท ส่วนในต้นไม้ตัดสินใจ จำแนกดอกรักผิดพลาดน้อยที่สุด เนื่องจากโครงสร้างข้อมูลของต้นไม้ที่เครื่องจักรได้สร้างขึ้นเลือกคุณลักษณะเด่นที่ใช้เป็น Root คือ คุณลักษณะเด่นสี เพื่อแยกดอกมีราออกก่อน จากนั้นเลือกคุณลักษณะเด่นพื้นที่ระหว่างกลีบ เพื่อแยกดอกที่ฐานชิดติดกัน และสุดท้ายใช้คุณลักษณะเด่นพื้นที่กลีบ เพื่อใช้แยกดอกที่กลีบยาวไม่เท่ากัน และดอกดี แนวทางวิจัยต่อไป ศึกษาคุณลักษณะเด่นด้านรูปทรง เช่น ขนาดความสูง ความกว้าง หรือ การใช้แฟร็กทัล (Fractal) เพื่อช่วยให้การจำแนกมีความถูกต้องเพิ่มขึ้น

กิตติกรรมประกาศ

ขอขอบคุณหลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏพิบูลสงคราม ที่ให้ความอนุเคราะห์ด้านสถานที่และอุปกรณ์บางส่วนสำหรับใช้ในการทำงานวิจัย

เอกสารอ้างอิง

[1] Nattavadee Hongboonmee and Praphasiri Trepanichkul, “Comparison of Data Classification Efficiency to Analyze Risk Factors that Affect the Occurrence of Hyperthyroidusing Data Mining Techniques”, *Journal of Information Science and Technology*, Vol. 9, No. 1, pp. 41-51, JAN-JUN 2019.

[2] Arika Thammano1, Muthita Wangkid and Arit Thammano, “Breast Cancer Prediction Using K-mean Classification Algorithm with Self-adaptive Weight”, Journal *of Information Science and Technology*, Vol. 10, No. 2, pp. 1-9, JUL-DEC 2020.

[3] Songgrod Phimphisan and Nattavut Sriwiboon, “Image Processing for Fundus Image Classification using Deep Learning”, *Journal of Information Science and Technology*, Vol. 10, No. 2, pp. 19-25, JUL-DEC 2020.

[4] Nattavadee Hongboonmee and Nutthapong Jantawong, "Apply of Deep Learning Techniques to Measure the Sweetness Level of Watermelon via Smartphone", *Journal of Information Science and Technology*, Vol. 10, No. 2, pp. 59-69, JUL-DEC 2020.

[5] Sarawoot Boonkidram and Nattavut Sriwiboon, "Physical Quality Investigation of Germinated Brown Rice by using Image Processing", *Journal of Information Science and Technology*, Vol. 10, No. 2, pp. 101-109, , JUL-DEC 2020.

[6] Jittrapong Jaroenjit, Apirak Panpanasakul, Pollawat Chaisri, Peerapong Promduang and Sutida Prompongusawa, "Classification pearls using image processing," presented at the 9 th Hatyai National and International Conference, Hatyai University, Songkhla, Thailand, Jul. 20, 1679-1691, 2018.

[7] Nattavadee Hongboonmee and Kanin Pratoomthong, "The Analysis System of Counterfeit Banknote by Photo on Smartphone using Deep Learning Technique", *Journal of Information Science and Technology*, Vol. 10, No. 2, pp. 90-100, JUL- DEC 2020.

[8] Somying Thainimit, "Introduction to Image Processing," in *Digital Image Processing with MATLAB*, Thailand: Kasetsart University, 2010. Accessed: Oct. 1, 2021. [Online]. Available: https://ebook.lib.ku.ac.th/ebook27/ebook/ 2014RG0087/index.html#p=1

[9] Charturong Tantibundhit, "Introduction to Pattern Recognition," in *Pattern Recognition*, Thailand: Thammasat Printing house, 2012.

[10] Parinya Sanguansat, "*Decision Tree,*" in *Artificial Intelligence with Machine Learning*, Thailand: INFOPRESS, 2019.

[11] Buncha Pasilatesung, "Support Vector Machines," in *Python Machine Learning*, Thailand: SE-EDUCATION PUBLIC COMPANY LIMITED, 2021.

[12] Tatpong Katanyukul, "Neural Network," in *Introduction to Machine Learning*, Thailand: Faculty of Engineering Khon Kaen University, 2017.

[13] Jakkarin Sanuksan and Olarik Surinta, "Deep Convolutional Neural Networks for Plant Recognition in the Natural Environment", *Journal of Science and Technology Mahasarakham University*, Vol. 38, No. 2, pp. 113-124, MAR-APR 2019.

[14] Rotsarin Tritanasombat and Tawin Tanawong, "Orchid Species Analyze System with Artificial Convolution Neural Network," presented at the 9 th Asia Undergraduate Conference on Computing, Rajamangala University of Technology Rattanakosin, Prachuap Khiri Khan, Thailand, Feb. 25, 1337-1343, 2021.

[15] Thanawat Poonyarit and Chutiphon Srisawat, "Peace Lily Images Classification with Visual Contents," presented at the UTCC Academic Day Conference, University of the Thai Chamber of Commerce, Bangkok, Thailand, Jun. 7, 704-716, 2017.

[16] Put Panuwanitchakorn and JanjiraPayakpate, "Using Image Segmentation Technique on the Image of Orchids Paphiopedilum Native Species of Thailand," presented at the 10 th Mahasarakham University Research Conference, Mahasarakham University, Mahasarakham, Thailand, Sep. 11-12, 1679-1691, 2014.

[17] Nopparut Pattansarn and Nattavut Sriwiboon, "Image Processing for Classifying the Quality of the Chok-Anan Mango by Simulating the Human Vision using Deep Learning", *Journal of Information Science and Technology*, Vol. 10, No. 1, pp. 24-29, JAN-JUN 2020.

[18] Hossin, M.1 and Sulaiman, M.N, "A Review on Evaluation Metrics for Data Classification Evaluations", *International Journal of Data Mining & Knowledge Management Process*, Vol. 5, No. 2, pp. 1-11, , March 2015.