

บทที่ 4

ผลการทดลอง

โครงการปริญญาโทฉบับนี้นำเสนอแอปพลิเคชันจำแนกสายพันธุ์งูที่พบในไทยด้วยการเรียนรู้เชิงลึก (Thai Snake Classification Application Using Deep Learning) ซึ่งบทความนี้จะประเมินประสิทธิภาพที่ได้จากโมเดลด้วยชุดข้อมูลตรวจสอบ และเปรียบเทียบสถาปัตยกรรม CNN และผลการทดลองของโมเดลที่ได้จากการประเมินประสิทธิภาพ

4.1 การเปรียบเทียบประสิทธิภาพการรู้จำจากสถาปัตยกรรม CNN

ในการจำแนกสายพันธุ์งูได้ประยุกต์ใช้ร่วมกับสถาปัตยกรรม CNN โดยผู้จัดทำได้เลือกสถาปัตยกรรม InceptionV3 และ MobileNetV2 เพื่อเปรียบเทียบหาโมเดลที่มีประสิทธิภาพมากที่สุด โดยในการประเมินจะใช้ตาราง confusion matrix ในการทดลองครั้งนี้ โดยใช้ชุดข้อมูลเดียวกัน

การประเมินประสิทธิภาพด้วย confusion matrix จะใช้ชุดข้อมูลสำหรับการ Test ของงูทั้งหมด 20 สายพันธุ์ คิดเป็น 20 เปอร์เซ็นต์ของทั้งหมด จากนั้นนำข้อมูลสำหรับประเมินประสิทธิภาพแต่ละสายพันธุ์ หาค่าความแม่นยำ(Precision), ค่าความระลึก(Recall), ด้วย Confusion Matrix เมื่อได้ผลลัพธ์แล้วนำค่าทั้งสองมาคำนวณ (Accuracy)

ตารางที่ 4.1 ผลการประเมินประสิทธิภาพสถาปัตยกรรม MobileNetV2

สายพันธุ์งู	Precision	Recall	F1-Score	Support
งูกะปะ	0.74	0.70	0.72	40
งูกันขบ	0.68	0.57	0.62	40
งูจงอาง	0.44	0.40	0.42	40
งูทับสมิงคลา	0.80	0.90	0.85	40
งูทางมะพร้าว	0.79	0.38	0.51	40
งูปล้องฉนวน สร้อยเหลือง	0.79	0.75	0.77	40
งูปีแก้วลายแต้ม	0.38	0.47	0.42	40
งูลายสอ	0.47	0.88	0.61	40
งูลายสาคอแดง	0.59	0.55	0.57	40

ตารางที่ 4.1 ผลการประเมินประสิทธิภาพสถาปัตยกรรม MobileNetV2 (ต่อ)

สายพันธุ์	Precision	Recall	F1-Score	Support
งูสมิงทะเลปาก เหลือง	0.69	0.62	0.66	40
งูสามเหลี่ยม	0.92	0.85	0.88	40
งูสิง	0.39	0.68	0.49	40
งูสิงหางลาย	0.56	0.23	0.32	40
งูหลาม	0.89	0.88	0.85	40
งูเขียวพระอินทร์	0.64	0.57	0.61	40
งูเขียวหางไหม้ ท้องเหลือง	0.95	0.88	0.91	40
งูเหลือม	0.70	0.78	0.74	40
งูเห่า	0.76	0.85	0.80	40
งูแมวเซา	0.79	0.78	0.78	40
งูแสงอาทิตย์	0.81	0.53	0.64	40
Accuracy			0.65	800
Macro avg	0.63	0.65	0.63	800
Weighted avg	0.63	0.65	0.63	800

คลาสที่มีค่า Precision ที่น้อยที่สุด คือ งูปีแก้วลายแต้ม อยู่ที่ 0.38 อาจเป็นเพราะถึงงูปีแก้วลายแต้มจะเป็นงูที่มีลวดลายบนตัวชัดเจน แต่รูปภาพส่วนใหญ่ในชุดข้อมูลลวดลายเหล่านี้ค่อนข้างที่จะกลมกลืนไปกับสภาพแวดล้อม เลยอาจทำให้โมเดลไม่สามารถเรียนรู้คุณลักษณะเหล่านี้ได้ดีเท่าที่ควร

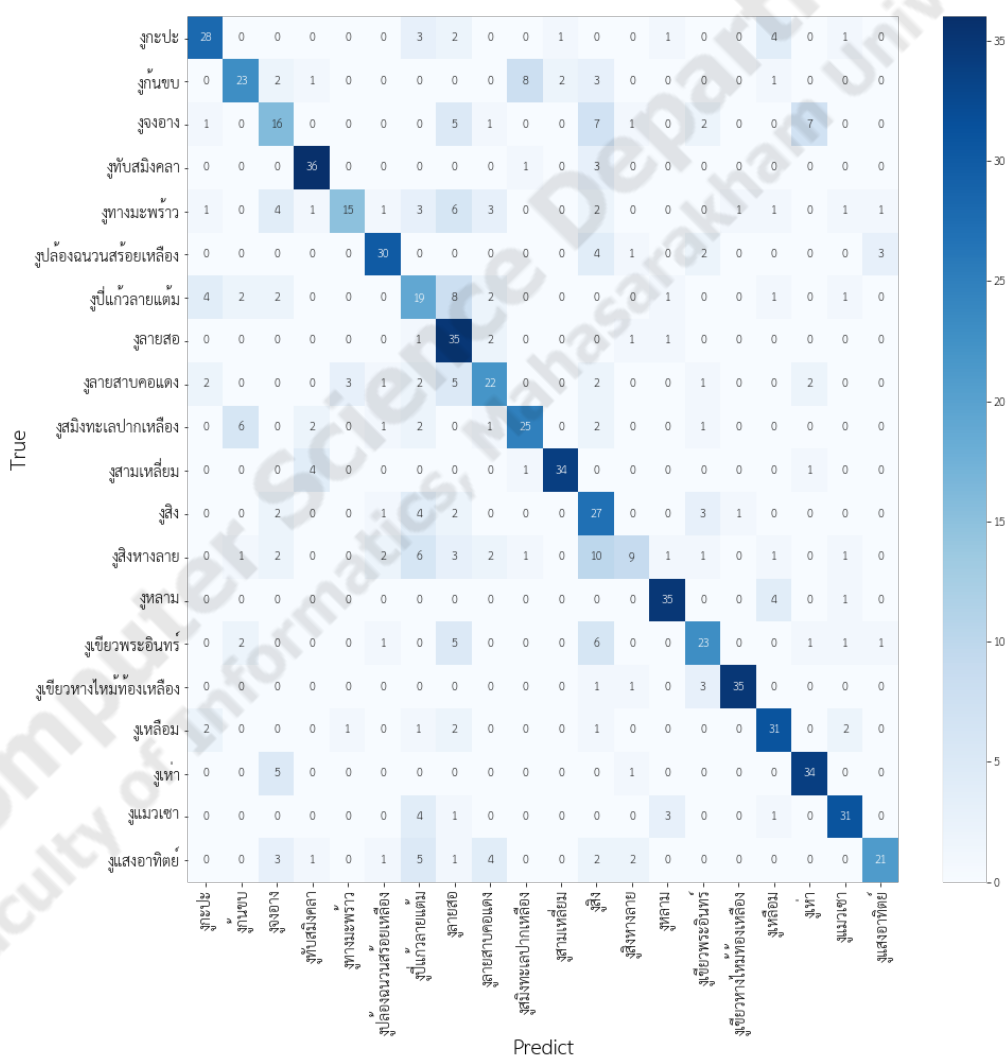
คลาสที่มีค่า Precision ที่สูงที่สุด คือ งูเขียวหางไหม้ท้องเหลือง อยู่ที่ 0.95 เพราะงูเขียวหางไหม้ท้องเหลืองเป็นงูที่มีลักษณะชัดเจนมากที่สุดตัวหนึ่ง ด้วยสีของร่างกายที่เป็นสีโทนเขียวทั้งหมดจึงน่าจะเป็นเรื่องง่ายที่โมเดลจะเรียนรู้ลักษณะเหล่านี้

คลาสที่มีค่า Recall ที่น้อยที่สุด คือ งูสิงหางลาย อยู่ที่ 0.23 เพราะงูสิงหางลายนั้นมีลักษณะที่แทบจะไม่ต่างกับงูสิง ที่แตกต่างกันมีแค่ลายบนตัวที่มีสีดำมากกว่าเล็กน้อยและมีลายขีดสีดำอยู่ใต้ปาก ส่งผลให้โมเดลทำนายเป็นงูสิงมากกว่า และด้วยชุดข้อมูลที่รวบรวมมาไม่สามารถแสดงลักษณะเหล่านี้ออกมาได้อย่างชัดเจนทำให้โมเดลไม่สามารถเรียนรู้ได้ดีเท่าที่ควร

คลาสที่มีค่า Recall ที่สูงที่สุด คือ งูทับสมิงคลา อยู่ที่ 0.90 เนื่องจากงูทับสมิงคลานั้นมีสีของร่างกายที่ไม่ซับซ้อน โดยจะเป็นสีขาวสลับดำ ทำให้โมเดลสามารถตรวจจับคุณลักษณะของงูชนิดนี้ได้ง่าย

คลาสที่มีค่า F1-Score ที่น้อยที่สุด คือ งูสิงหางลาย อยู่ที่ 0.32 เพราะงูสิงหางลายมีค่า Precision และ Recall ที่ต่ำ เนื่องจากชุดข้อมูลไม่สามารถแสดงลักษณะเด่นของงูชนิดนี้ออกมาได้ทำให้ไม่สามารถเรียนรู้ข้อแตกต่างระหว่างงูสิงหางลายกับงูสิงหรือกับงูชนิดอื่นที่มีโทษคล้ายกันได้ ส่งผลให้การทำนายมีความถูกต้องที่ต่ำ

คลาสที่มีค่า F1-Score ที่สูงที่สุด คือ งูเขียวหางไหม้ท้องเหลือง อยู่ที่ 0.91 งูเขียวหางไหม้มีลักษณะที่เป็นสีเขียวทั่วทั้งร่างกายและบริเวณท้องที่เป็นสีเหลือง ทำให้โมเดลสามารถตรวจจับคุณลักษณะของงูในรูปภาพอินพุตได้ง่าย ด้วยสาเหตุที่กล่าวมาส่งผลให้โมเดลทำนายได้อย่างถูกต้อง ในกรณีที่ตัวของงูนั้นไม่กลืนไปกับสภาพแวดล้อมมากนัก



ภาพประกอบที่ 4.1 การประเมินด้วย confusion matrix ของ MobileNetV2

จากภาพประกอบที่ 4.1 คลาสที่มีจำนวนรูปภาพที่ทำนายถูกต้องน้อยที่สุดคืองูสิงหางลาย อยู่ 9 จาก 40 รูปเท่านั้น โดยคลาสที่งูสิงหางลายทำนายผิดส่วนมากคือ งูสิง, งูน้ำจืด, งูน้ำจืด คาคว่าสาเหตุที่โมเดลไม่เรียนรู้ลักษณะของงูสิงหางลายหรือทำนายผิดอาจเป็นเพราะงูสิงกับงูสิงหางลายนั้นมีลักษณะที่

เหมือนกัน จะต่างกันตรงที่รูปร่างลายมีลายตามตัวที่ออกสีดำกว่าเล็กน้อยและมีขีดสีดำที่ใต้ปาก ส่วนที่ทำนายผิดเป็นรูปไม้เท้าลายแต้มนั้นอาจเป็นเพราะรูปภาพที่ใช้ฝึกฝนโมเดลนั้นมีโทนสีหรือสภาพแวดล้อมคล้ายกัน

ตารางที่ 4.2 ผลการประเมินประสิทธิภาพสถาปัตยกรรม InceptionV3

สายพันธุ์งู	Precision	Recall	F1-Score	Support
งูกะปะ	0.59	0.50	0.54	40
งูกิ่งขบ	0.59	0.57	0.58	40
งูจงอาง	0.52	0.55	0.54	40
งูทับสมิงคลา	0.79	0.78	0.78	40
งูทางมะพร้าว	0.57	0.30	0.39	40
งูปล้องฉนวน สร้อยเหลือง	0.70	0.70	0.70	40
งูปีแก้วลายแต้มน	0.48	0.25	0.33	40
งูลายสอ	0.49	0.65	0.56	40
งูลายسابคอแดง	0.48	0.57	0.52	40
งูสมิงทะเลปาก เหลือง	0.67	0.78	0.72	40
งูสามเหลี่ยม	0.87	1.00	0.93	40
งูสิง	0.55	0.65	0.60	40
งูสิงหางลาย	0.14	0.07	0.10	40
งูหลาม	0.97	0.93	0.95	40
งูเขียวพระอินทร์	0.69	0.60	0.64	40
งูเขียวหางไหม้ ท้องเหลือง	0.92	0.88	0.90	40
งูเหลือม	0.61	0.78	0.68	40
งูเห่า	0.73	0.90	0.81	40
งูแมวเซา	0.67	0.88	0.76	40

ตารางที่ 4.2 ผลการประเมินประสิทธิภาพสถาปัตยกรรม InceptionV3 (ต่อ)

สายพันธุ์สูง	Precision	Recall	F1-Score	Support
งูแสงอาทิตย์	0.62	0.62	0.62	40
Accuracy			0.65	800
Macro avg	0.63	0.65	0.63	800
Weighted avg	0.63	0.65	0.63	800

คลาสที่มีค่า Precision ที่น้อยที่สุด คือ งูสิงหางลาย อยู่ที่ 0.14 เพราะงูสิงหางลายนั้นมีลักษณะที่แทบจะไม่ต่างกับงูสิง ที่แตกต่างกันจะมีแค่ลายบนตัวที่มีสีดำมากกว่าเล็กน้อยและมีลายขีดสีดำอยู่ใต้ปาก ส่งผลให้โมเดลทำนายเป็นงูสิงมากกว่า และด้วยชุดข้อมูลที่รวบรวมมาไม่สามารถแสดงลักษณะเหล่านี้ ออกมาได้อย่างชัดเจนทำให้โมเดลไม่สามารถเรียนรู้ได้ดีเท่าที่ควร

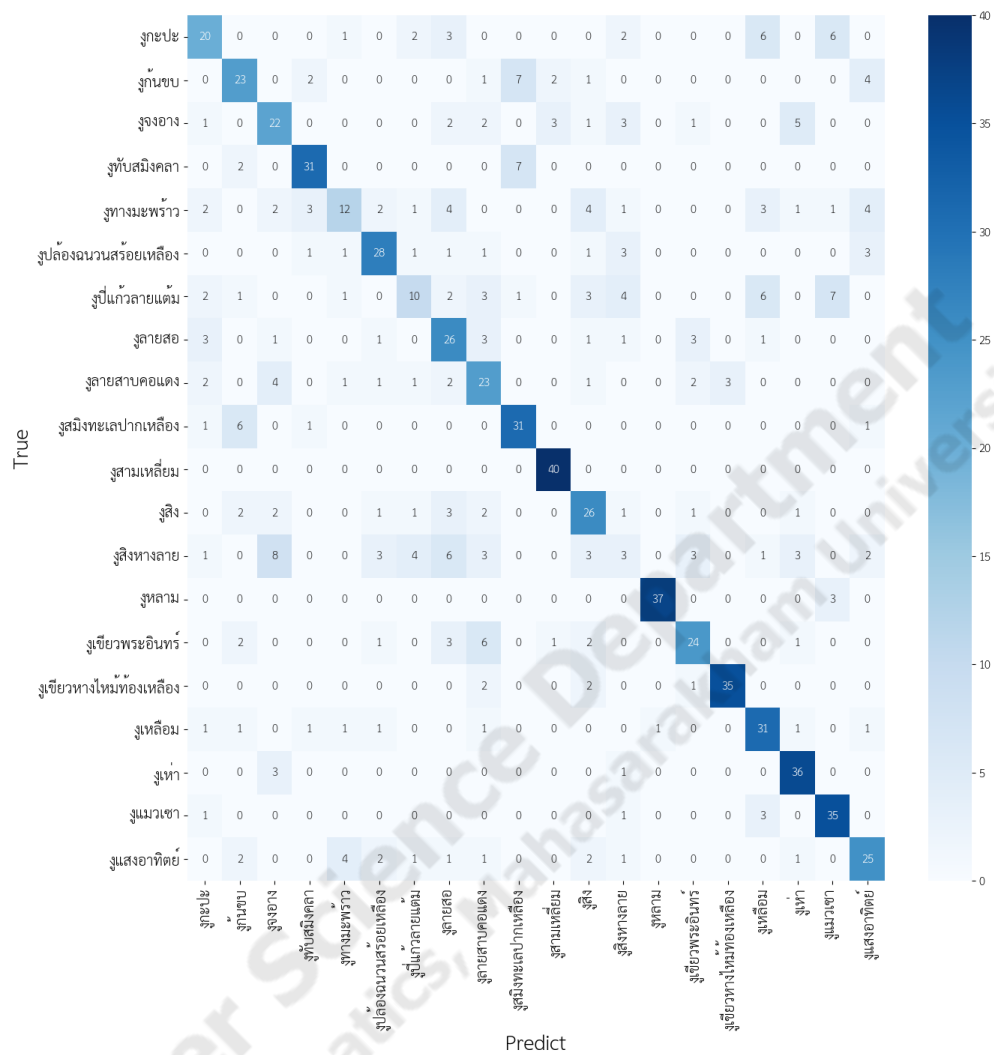
คลาสที่มีค่า Precision ที่สูงที่สุด คือ งูหลาม อยู่ที่ 0.97 งูหลามนั้นเป็นหนึ่งในงูที่มีขนาดใหญ่ที่สุดและมีสีสันทึบหลากหลายในชุดข้อมูล คาดว่าที่โมเดลสามารถเรียนรู้ลักษณะที่หลากหลายเหล่านี้ได้ อาจเป็นเพราะจุดรวมของงูหลามคือเป็นงูที่บร้งกายนั้นมีลวดลายที่เป็นสีดำ สีน้ำตาล สีขาว และสีเหลือง ซึ่งอาจเป็นลักษณะที่โมเดลเลือกที่จะเรียนรู้

คลาสที่มีค่า Recall ที่น้อยที่สุด คือ งูสิงหางลาย อยู่ที่ 0.07 เพราะงูสิงหางลายนั้นมีลักษณะที่แทบจะไม่ต่างกับงูสิง ที่แตกต่างกันจะมีแค่ลายบนตัวที่มีสีดำมากกว่าเล็กน้อยและมีลายขีดสีดำอยู่ใต้ปาก ส่งผลให้โมเดลทำนายเป็นงูสิงมากกว่า และด้วยชุดข้อมูลที่รวบรวมมาไม่สามารถแสดงลักษณะเหล่านี้ ออกมาได้อย่างชัดเจนทำให้โมเดลไม่สามารถเรียนรู้ได้ดีเท่าที่ควร

คลาสที่มีค่า Recall ที่สูงที่สุด คือ งูสามเหลี่ยม อยู่ที่ 1.00 งูสามเหลี่ยมในด้านของลักษณะแล้ว ถือว่าเป็นประเภทเดียวกับงูทับสมิงคลาและงูสมิงทะเลปากเหลือง เพราะด้วยร่างกายที่เป็นสีดำสลับกับสีอีกสีหนึ่ง สำหรับงูสามเหลี่ยมคือสีดำสลับเหลือง ด้วยร่างกายที่มีสีที่ไม่ซับซ้อนนี้ทำให้โมเดลสามารถตรวจจับคุณลักษณะของงูชนิดนี้ได้ง่าย

คลาสที่มีค่า F1-Score ที่น้อยที่สุด คือ งูสิงหางลาย อยู่ที่ 0.10 เพราะงูสิงหางลายมีค่า Precision และ Recall ที่ต่ำ เนื่องจากชุดข้อมูลไม่สามารถแสดงลักษณะเด่นของงูชนิดนี้ออกมาได้ทำให้ไม่สามารถเรียนรู้ข้อแตกต่างระหว่างงูสิงหางลายกับงูสิงหรือกับงูชนิดอื่นที่มีโทนสีคล้ายกันได้ ส่งผลให้การทำนายมีความถูกต้องที่ต่ำ

คลาสที่มีค่า F1-Score ที่สูงที่สุด คือ งูหลาม อยู่ที่ 0.95 คาดว่าสาเหตุที่งูหลามเป็นงูที่มีค่า f1-score สูงที่สุดสำหรับโมเดล InceptionV3 อาจเป็นเพราะขนาดของตัวที่ใหญ่และโทนสีที่ค่อนข้างเด่นชัดทำให้สามารถทำนายได้อย่างถูกต้องมากกว่าชนิดอื่น ๆ





ภาพประกอบที่ 4.2 การประเมินด้วย confusion matrix ของ InceptionV3

จากผลการประเมินประสิทธิภาพ พบว่าค่า Accuracy ของสถาปัตยกรรม MobileNetV2 ดังตารางที่ 4.1 ต่ำกว่า InceptionV3 ดังตารางที่ 4.2 และได้ทำการทดสอบกับรูปภาพพบว่าโมเดลของ MobileNetV2 ได้ผลลัพธ์ที่ถูกต้องมากกว่าดังนั้นจึงเลือกใช้สถาปัตยกรรม MobileNetV2 ประยุกต์ร่วมกับโมเดลจากการทดลองดังกล่าวสามารถแสดงภาพตัวอย่างที่ได้จากการทำนายผลลัพธ์ โดยยกตัวอย่างได้ดังตารางที่ 4.3

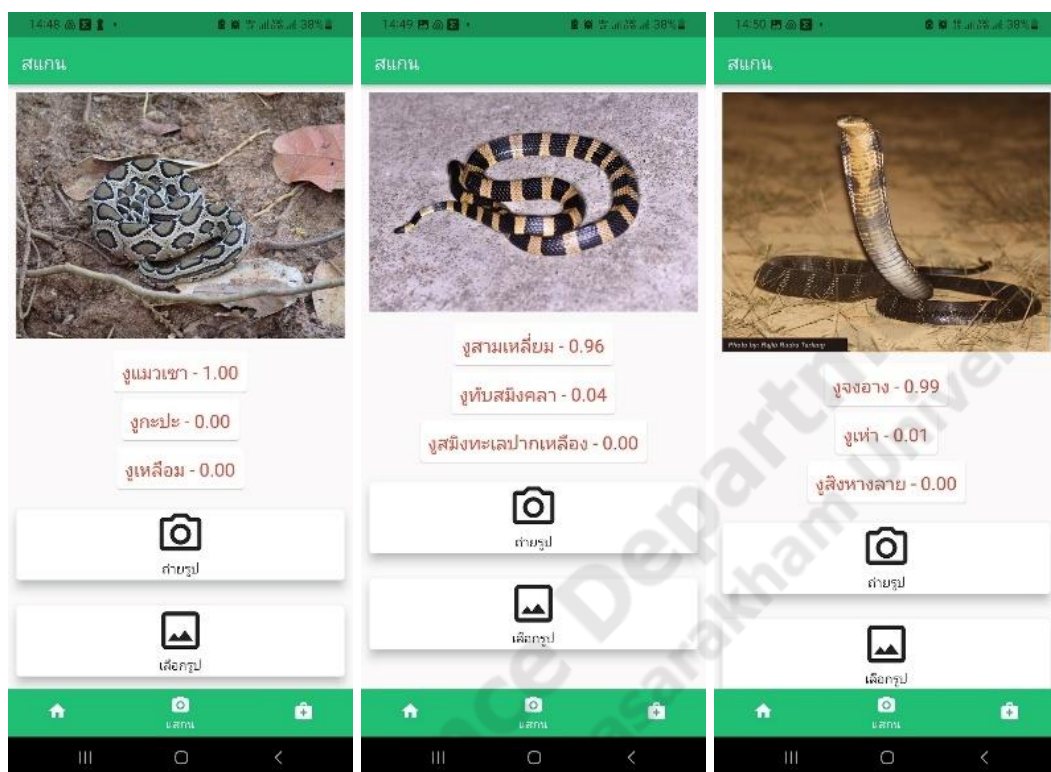
ตารางที่ 4.3 ตัวอย่างผลการทำนาย

ภาพที่	รูปภาพ	ผลทำนายจากสถาปัตยกรรม		
		MobileNetV2	InceptionV3	ผลเฉลย
1		งูแมวเซา	งูแมวเซา	งูแมวเซา
2		งูเหลือม	งูเหลือม	งูเหลือม
3		งูแสงอาทิตย์	งูแสงอาทิตย์	งูแสงอาทิตย์
4		งูเขียวพระอินทร์	งูลายสาคอแดง	งูเขียวพระอินทร์
5		งูปี้แก้วลายแต้ม	งูปี้แก้วลายแต้ม	งูปี้แก้วลายแต้ม
6		งูเห่า	งูเห่า	งูเห่า

ตารางที่ 4.3 ตัวอย่างผลการทำนาย (ต่อ)

ภาพที่	รูปภาพ	ผลทำนายจากสถาปัตยกรรม		
		MobileNetV2	InceptionV3	ผลเฉลย
7		งูลายสอ	งูลายสอ	งูลายสอ
8		งูสามเหลี่ยม	งูสามเหลี่ยม	งูสามเหลี่ยม
9		งูกะปะ	งูกะปะ	งูกะปะ
10		งูจงอาง	งูเห่า	งูจงอาง

4.2 ตัวอย่างการทำนายบน Mobile Application



ภาพประกอบที่ 4.3 ตัวอย่างการทำนายบน Mobile Application

4.3 สรุปและวิเคราะห์ผลการทดลอง

จากการทดลองในการหาสถาปัตยกรรม CNN ที่ดีที่สุดโดยการทดลองจากชุดข้อมูลจำนวน 4069 รูป ผลที่ได้จากตารางที่ 4.1 และตารางที่ 4.2 พบว่าสถาปัตยกรรม MobileNetV2 มีค่า accuracy เป็น 66 เปอร์เซ็นต์ InceptionV3 มีค่า accuracy เป็น 65 เปอร์เซ็นต์ การเปรียบเทียบประสิทธิภาพจากการคำนวณของ Confusion Matrix และทำการทดสอบกับภาพจริงและเปรียบเทียบผลทำนาย จึงทำให้สรุปได้ว่าสถาปัตยกรรม MobileNetV2 มีประสิทธิภาพในการทำนายที่ดีกว่า และเมื่อเทียบกับขนาดโมเดลที่ได้จากการฝึกฝนแล้ว โมเดลที่ได้จากสถาปัตยกรรม MobileNetV2 มีขนาดเล็กมากเมื่อเทียบกับโมเดลที่ได้จากสถาปัตยกรรม InceptionV3 ที่มีขนาดใหญ่ ดังนั้นในงานนี้จึงเลือกใช้สถาปัตยกรรม MobileNetV2