

ระบบวิเคราะห์ภาพวิดีโอเพื่อตรวจหาเหตุการณ์น่าสงสัยด้วยเทคโนโลยีการเรียนรู้เชิงลึก  
(Video Analysis System for Suspicious Activity Detection using Deep Learning)

อาทิตย์ ดาบลาอ่า ปวีณา เลาหรัตน์ สุพจน์ ธรรมารักษ์

ระบบวิเคราะห์ภาพวิดีโอเพื่อตรวจหาเหตุการณ์น่าสงสัยด้วยเทคโนโลยีการเรียนรู้เชิงลึก  
Video Analysis System for Suspicious Activity Detection using Deep Learning

นายอาทิตย์ ดาบลาอ่า <sup>1</sup> ภาควิชาระบบสารสนเทศเพื่อการจัดการ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ s6007021858352@email.kmutnb.ac.th	นางสาวปวีณา เลาหรัตน์ <sup>2</sup> ภาควิชาระบบสารสนเทศเพื่อการจัดการ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ s6007021858166@email.kmutnb.ac.th
--	--

นายสุพจน์ ธรรมารักษ์<sup>3</sup>  
ภาควิชาระบบสารสนเทศเพื่อการจัดการ  
คณะเทคโนโลยีสารสนเทศ  
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ  
s6007021858336@email.kmutnb.ac.th

อาจารย์ที่ปรึกษา ผศ.ดร. มหาศักดิ์ เกตุช่อ<sup>4</sup>  
ภาควิชาระบบสารสนเทศเพื่อการจัดการ คณะเทคโนโลยีสารสนเทศ  
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ  
อีเมล์ [Maoguee@hotmail.com](mailto:Maoguee@hotmail.com)

## 1. ความเป็นมาและความสำคัญของปัญหา

จากอดีตประเทศไทยต้องต่อสู้กับความไม่สงบมากมายในประเทศ ซึ่งบริเวณภาคใต้โดยเฉพาะพื้นที่ติดกับประเทศไทยเพื่อนบ้าน เป็นเรื่องที่หลาย ๆ รัฐบาลให้ความสำคัญในการติดตามและแก้ปัญหา ไม่ว่าจะเป็นความไม่สงบทางความคิด ด้านศาสนา ในภาคใต้ จากรุ่มผู้ไม่หวังดีต่างๆ กลุ่มที่ต้องการแบ่งแยกดินแดน กลุ่มที่มีการกระทำที่ผิดกฎหมายไม่ว่าจะเป็นสิ่งด้านศาสนา ลักลอบขนน้ำมันจากประเทศเพื่อนบ้านเข้ามาขาย กลุ่มความไม่สงบด้านการเมือง ปัจจุบันปัญหาความไม่สงบของประเทศไทยโดยเฉพาะในเขต 3 จังหวัดชายแดนภาคใต้ที่มีกลุ่มก่อการร้ายเข้ามาสร้างความไม่สงบเป็นปัญหาที่ฝัง根柢มากว่า 14 ปี โดยมีเหตุการณ์ลอบทำร้าย วางเพลิง วางระเบิด ก่อการร้าย และจลาจล เกิดขึ้นอย่างต่อเนื่อง แม้รัฐบาลและหน่วยงานทางทหารจะพยายามทุ่มสรรพกำลังและงบประมาณลงไปเพื่อแก้ปัญหา แต่ก็ยังไม่สามารถปราบปรามให้สงบเหมือนเดิมได้ ในแต่ละปีมีทหารและประชาชนผู้ไม่มีทางสู้ได้เสียชีวิตจากความรุนแรงที่เกิดขึ้นปีละหลายร้อยคน ซึ่งเป็น

ปัญหาทางด้านความมั่นคงของประเทศไทย ส่งผลกระทบต่อความปลอดภัยในชีวิตของประชาชนเป็นอย่างมาก เป็นเหตุผลที่ทำให้องค์กรต่างๆ ลงทุนกับระบบรักษาความปลอดภัยสาธารณะที่มีประสิทธิภาพเป็นจำนวนมาก

เทคโนโลยีด้านปัญญาประดิษฐ์ (Artificial Intelligence : AI) เป็นวัตถุกรรมที่มีความฉลาดที่ถูกสร้างขึ้นให้กับสิ่งที่ไม่มีชีวิต โดยเป็นการเรียนรู้เกี่ยวกับกระบวนการการคิด การกระทำ การให้เหตุผล การปรับตัว และการทำงานของสมอง ความก้าวหน้าของเทคโนโลยี AI และ Big Data ที่กำลังเข้ามามีบทบาทและสร้างความเปลี่ยนแปลงด้านต่างๆ ไม่ว่าจะเป็นด้านธุรกิจ การแพทย์ การทหาร การสืบสวนสอบสวน หรือในชีวิตประจำวันมากขึ้นเรื่อยๆ รวมถึงด้านการป้องกันและสร้างความปลอดภัยให้กับชีวิตมนุษย์ด้วย

ในงานวิจัยนี้ ทีมผู้วิจัยนำเสนอการประยุกต์ใช้ปัญญาประดิษฐ์ (Artificial Intelligence) ด้วยเทคนิคการเรียนรู้เชิงลึก Deep Learning ซึ่งทีมผู้วิจัยเห็นว่าระบบดังกล่าวสามารถที่จะวิเคราะห์ภาพจากกล้องวงจรปิด ซึ่งในหลายพื้นที่ในเขต 3 จังหวัดชายแดนภาคใต้ ได้มีการติดตั้งกล้องวงจรปิดไว้เพื่อความปลอดภัยในระดับหนึ่งแล้ว โดยระบบนี้จะทำการรวบรวมข้อมูลภาพเคลื่อนไหวและพาณิชจากระบบกล้องวงจรปิดต่าง ๆ ทั้งระบบเก่าและระบบใหม่ ซึ่งจะช่วยเพิ่มขีดความสามารถในการใช้งานกล้องวงจรปิดเพื่อช่วยวิเคราะห์และจำแนกประเภทเหตุการณ์น่าสงสัย เหตุการณ์รุนแรง ไม่สงบ หรืออาชญากรรม สถานการณ์ปลอดภัย เพื่อให้สามารถป้องกันเหตุร้ายหรือไม่สงบในอนาคตได้อย่างเต็มประสิทธิภาพยิ่งขึ้น.

## 2. วัตถุประสงค์

เพื่อวิเคราะห์และพัฒนาระบบจำแนกประเภทเหตุการณ์น่าสงสัยจากภาพวิดีโอของกล้องวงจรปิด

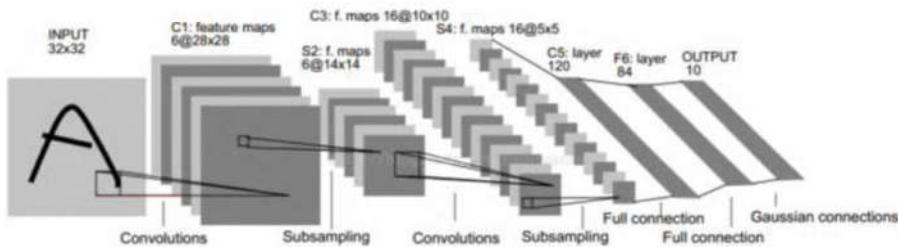
## 3. ขอบเขต และ ข้อจำกัดของงาน

การทำงานของระบบจะเป็นการใช้เทคโนโลยี Deep Learning เพื่อทำการวิเคราะห์และจำแนกประเภทเหตุการณ์น่าสงสัยจากภาพวิดีโอของกล้องวงจรปิด โดยจะจำแนกประเภทเหตุการณ์น่าสงสัยเป็น 3 ประเภท คือ 1. เหตุการณ์ความรุนแรงและไม่สงบ 2. เหตุการณ์น่าสงสัย 3. เหตุการณ์ปลอดภัย เพื่อตรวจหาเหตุการณ์ที่น่าสงสัยก่อนที่อาจจะเกิดขึ้นจริง และนำไปสู่การช่วยระงับหรือป้องกันไม่ให้เกิดขึ้นได้ โดยใช้อัลกอริทึม Convolutional Neural Network (CNN) ในการสกัดคุณลักษณะสำคัญต่างๆ จากส่วนใดๆ ของภาพวิดีโอและลดความซับซ้อนของการนำภาพวิดีโอเข้าระบบ โดยใช้โมเดล Inception ซึ่งถูกพัฒนาโดย Google ทำหน้าที่จำแนกภาพ และใช้อัลกอริทึม Recurrent Neural Network (RNN) ด้วยโมเดล Long Short-Term Memory Network (LSTM) ในการลำดับภาพและให้ผลการจำแนกภาพว่าจัดจำแนกอยู่ในประเภทเหตุการณ์น่าสงสัยประเภทไหน

## 4. ทฤษฎีที่เกี่ยวข้อง

### 4.1 Convolutional Neural network (CNN)

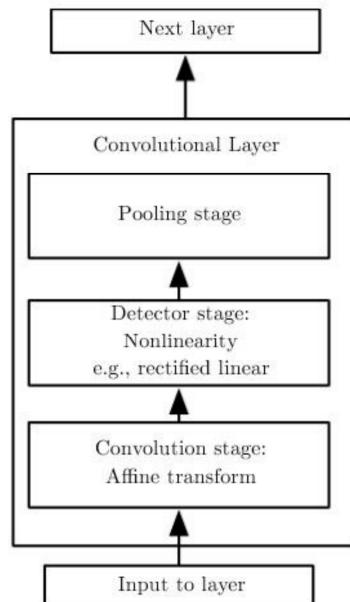
Convolutional Neural network เป็นเครือข่ายประสาทเทียมที่ประกอบด้วย จำนวนชั้น ทั้งหมด 3 ชั้น (3 Layer) ด้วยกัน ได้แก่ ชั้น Convolutional Neural Network Layer, ชั้น Pooling Layer และชั้น Fully-connected layer รูปต่อไปนี้แสดงถึงสถาปัตยกรรมของ LeNet-5 ซึ่งได้นำเสนอโดย Yann LeCun



ภาพที่ 1 : สถาปัตยกรรม CNN [8]

CNN เป็นรูปแบบ architecture หนึ่งของ feed-forward neural networks จัดเป็น Deep learning ซึ่งประสบความสำเร็จในการแก้ปัญหา Image Classification ที่เกี่ยวข้องกับรูปภาพ

#### การออบแบบโครงสร้าง CNN



ภาพที่ 2 : ส่วนประกอบของ CNN [3]

จากภาพจะเห็นได้ว่าใน 1 layer ประกอบไปด้วย 3 ส่วนใหญ่ๆ

#### 4.1.1 Convolution Layer

ขั้นตอนนี้จะสร้าง Sliding window(Filter) มาสแกนรูป input เพื่อทำ Feature คือ ทำการสแกนรูปเพื่อแยกองค์ประกอบของรูปออกมา เช่น ขอบ สี รูปทรง เป็นต้น โดยขั้น convolution นี้ กล่าวโดยย่อได้ว่าเป็นการตรวจจับ “local feature” บนรูปนำเข้า โดย local feature เหล่านี้คือ convolution kernel บน layer นั้นเอง

#### 4.1.2 Detector Layer

ขั้นนี้จะทำหน้าที่รับ output จาก Layer แรกและเปลี่ยนรูปของ nonlinear โดยใช้ activation อย่างเช่น Rectified Linear Units (ReLU) เพื่อความเรียบในการคำนวณและประสิทธิภาพของผลลัพธ์

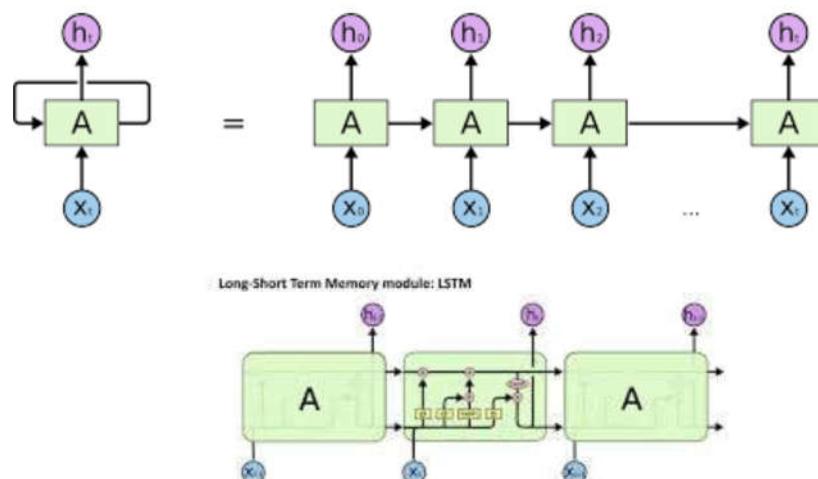
#### 4.1.3 Pooling Layer

Pooling ทำหน้าที่ resize ข้อมูลให้ขนาดเล็กลงโดยที่รายละเอียดของ input ยังคง保กวน เมื่อเป็นเดิมหลักการทำงานขั้นตอนนี้คล้ายๆ กับ Layer แต่ต่างที่ตรง output ที่ได้จะมีขนาดเล็กลง Pooling มีประโยชน์ในเรื่องเพิ่มความไวในการคำนวณและแก้ปัญหา overfitting ในปัจจุบันที่นิยมใช้ กันมากจะเป็น MAX กับ L2

จุดเด่นของ CNN คือ

- local feature เหล่านี้ถือเป็นส่วนหนึ่งของ NN ที่จะถูกปรับไปพร้อมๆ กับส่วน classifier ตั้งตนน หลังจากการ train เราจะได้ feature extractor ที่เหมาะสมกับงาน
- local feature เหล่านี้ถูก applied ที่ตำแหน่งต่างๆ บนรูปนำเข้า (เป็นการทำ weight sharing แบบหนึ่ง) ตั้งตนนเมื่อเราคำนวณเกรเดียนต์ย้อนกลับ เราต้องทำการรวมค่าที่ได้จากตำแหน่ง ทั้งหมดนี้ด้วย ตั้งตนขนาด (amplitude) ของค่าที่ได้ซึ่งมีขนาดใหญ่พอดในการปรับ kernel เหล่านี้ ที่มักอยู่ใน layer ล่างๆ

4.2 LSTM หรือชื่อเต็มคือ Long Short-Term Memory เป็นโครงข่ายประสาทเทียมแบบหนึ่งที่ถูก ออกแบบมาสำหรับการประมวลผลลำดับ (sequence) LSTM เป็นโครงข่ายประเภท Recurrent Neural Network (RNN) คือ การหานย้อนคืนกลับมาอีก การทำงานของ RNNs คือการเอาผลลัพธ์ที่ได้จากการคำนวณ ย้อนกลับมาใช้เป็นข้อมูลขาเข้าอีกครั้ง ดังรูป ซึ่งมีประโยชน์อย่างมากในข้อมูลที่มีความต่อเนื่อง เช่น time series ข้อมูลเสียง ข้อความ หรือแม้แต่รูปภาพของกิตาม



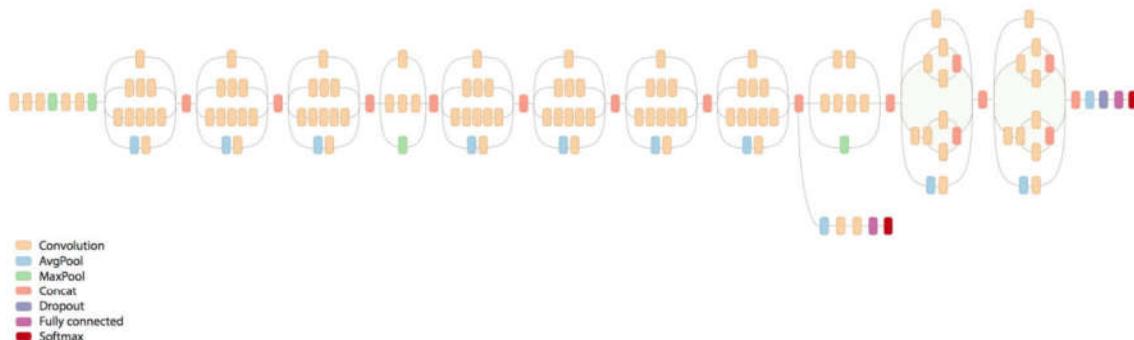
ภาพที่ 3 : โครงสร้างการทำงาน LSTM[4]

ในแต่ละ node ของ RNNs จะมีข้อมูลขาเข้าสองอย่างอันได้แก่ input ณ node นั้น ๆ และผลลัพธ์ที่ได้จากการคำนวณใน node ก่อนหน้า ซึ่งทั้งสองข้อมูลจะถูกนำมารวมเข้าด้วยกันและออกผลลัพธ์มาเป็นสองทางคือ ผลลัพธ์ที่ออก ณ node นั้น ๆ และออกเพื่อไปเข้าเป็นข้อมูลขาเข้าใน node ถัดไป

ข้อดีของ RNNs คือ มันมีการใช้ข้อมูลก่อนหน้าในการคำนวณที่อาจจะเกิดขึ้นในอนาคต นั่นหมายถึง อะไรที่เคยเกิดขึ้นในอดีตย่อมส่งผลต่อเหตุการณ์ที่จะเกิดขึ้นในอนาคตด้วย

#### 4.3 โมเดล Inception

เป็นโมเดลที่ถูกพัฒนาโดย Google เลยมีอีกชื่อว่า GoogLeNet ซึ่งโมเดลนี้ถูกฝึกฝนมากจากชุดข้อมูลรูปภาพจำนวนมาก 1.2 ล้านรูปเพื่อใช้ทำงานรูปทั้งหมด 1,000 ประเภทในการแข่งขัน ImageNet ซึ่งการนำมาใช้ต้องปรับตัวโมเดลให้เหมาะสมกับการวิเคราะห์ภาพก่อน จำเป็นต้องแก้ส่วนของชั้น Full-Connected Layer และเลือกฝึกฝนในส่วนที่คิดว่าจำเป็นก็พอ เป็นวิธีการที่เรียกว่า Transfer Learning คือการนำองค์ความรู้จากสิ่งที่เคยมีคนเคยแก้ปัญหาได้แล้วมาในโจทย์หนึ่งไปใช้ในโจทย์ที่ใกล้เคียงกัน Image Classification ก็เป็นอีกเรื่องที่มีการปรับแก้ไข โมเดลที่เป็น Open Source ปล่อยออกมาให้ใช้เรียบๆ



ภาพที่ 4 : โครงสร้างโมเดล Inception [2]

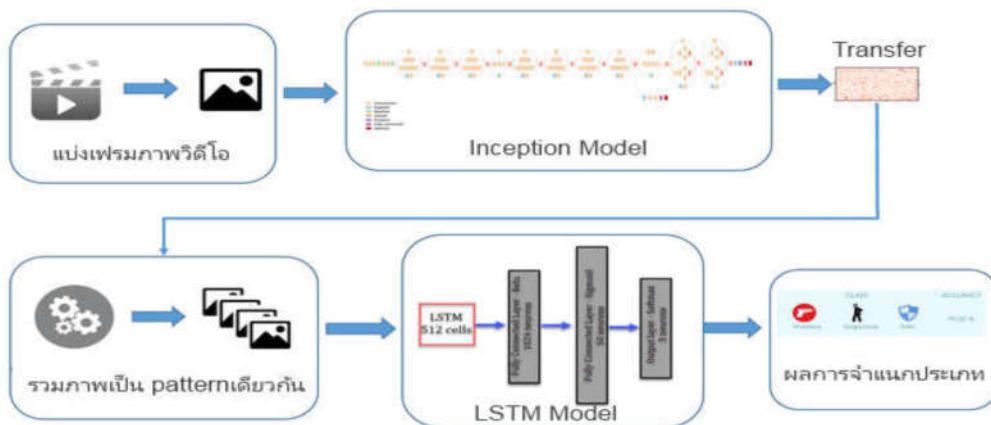
### 5. วิธีการพัฒนา และ เทคนิคที่ใช้

#### 5.1 การพัฒนาระบบจะใช้เทคนิคดังนี้

ใช้อัลกอริทึม Convolutional Neural Network (CNN) ด้วยโมเดล Inception ซึ่งเป็นวิธีการ Transfer Learning ในการสกัดคุณลักษณะระดับสูงของภาพวิดีโอและลดความซับซ้อนของการนำภาพวิดีโอเข้าระบบและใช้อัลกอริทึม Recurrent Neural Network (RNN) ด้วยโมเดล Long Short-Term Memory

Network (LSTM) ในการลำดับภาพและการจำแนกภาพขั้นสุดท้ายว่าจัดจำแนกอยู่ในประเภทเหตุการณ์น่าสงสัยประเภทไหน

ภาพที่ 5 : กรอบแนวการพัฒนาระบบ



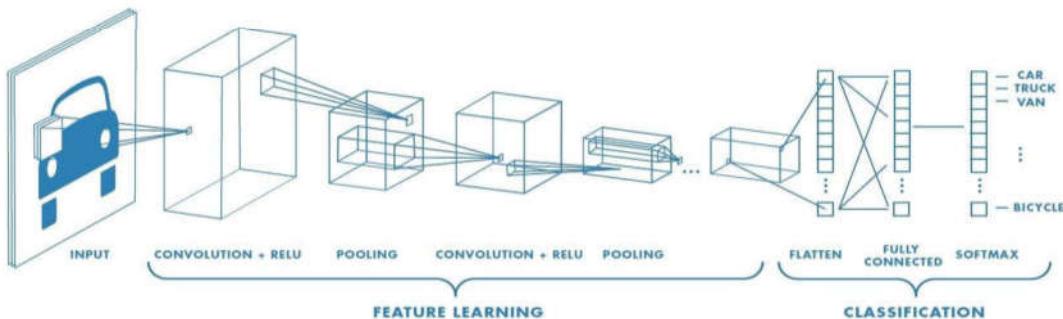
5.1.1 ใช้งาน OpenCV เพื่อแบ่งส่วนเฟรมภาพวิดีโอและปรับขนาดภาพตามที่กำหนด ขั้นตอนนี้ จะทำการแบ่งเฟรมภาพวิดีโอออกเป็นภาพนิ่งและปรับขนาดภาพที่ได้จากการล้อง CCTV ให้อยู่ในขนาดที่เหมาะสมสำหรับการเข้าไปทำการวิเคราะห์ข้อมูลภาพที่ไม่เดล โดยมีทางเลือกในการรับข้อมูลเข้า ด้วยวิธีดังนี้

1. ไฟล์วิดีโอ/ไฟล์ภาพถ่าย
2. ภาพวิดีโอสดจากกล้องวงจรปิดหรือกล้องเว็บแคม



ภาพที่ 6 : ภาพตัวอย่างจากกล้อง CCTV

5.1.2 ใช้ Inception Model เพื่อจำแนกรูปภาพและจัดจำแนกหมวดหมู่รูปภาพ ขั้นตอนนี้เป็นขั้นตอนของการนำภาพที่ได้จากขั้นตอนการแบ่งเฟรมภาพและจัดขนาดภาพให้เหมาะสมกับโมเดลแล้ว เข้าสู่โมเดล Inception โมเดลนี้จะทำงานคล้ายกับ LSTM ซึ่งจำเป็นการเรียนรู้และจดจำภาพเหตุการณ์ต่างๆที่เข้ามาในโมเดลแล้วจะทำการแบ่งหมวดหมู่ของภาพออกมานอกกลุ่ม ตามประเภทที่ได้จัดแบ่งไว้



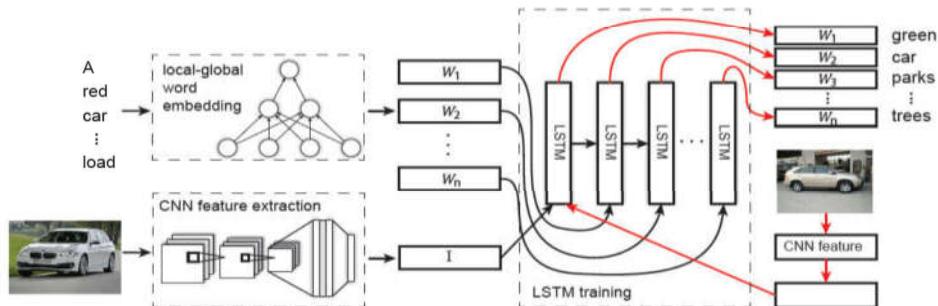
ภาพที่ 7 : นำภาพไปจำแนกภาพและแบ่งหมวดหมู่ใน Inception Model [9]

5.1.3 ดึงข้อมูลรูปภาพที่ได้จากการจำแนกหมวดหมู่รูปภาพจากการโมเดล Inception มา และรวมภาพเป็น pattern เดียวกัน เพื่อที่จะนำเข้าไปทำการเรียนรู้เพื่อจัดจำและทำนายผลโดยใช้ โมเดล Long Short Term Memory (LSTM) ในขั้นตอนต่อไป



ภาพที่ 8 : นำภาพที่ได้จากการจัดหมวดหมู่รวมเป็น pattern เดียวกัน

5.1.4 จะใช้ Long short term memory model เพื่อทำนายประเภทเหตุการณ์ว่า เหตุการณ์ที่เกิดขึ้นนี้จะถูกจำแนกอยู่ในประเภทเหตุการณ์哪่ำสักประเภทใด เช่น เหตุการณ์ปกติ เหตุการณ์ที่น่าติดตามลงสัญญา เหตุการณ์ความรุนแรง หรือมีอันตรายที่ต้องเข้าช่วยเหลือโดยเร่งด่วน ตามที่เคยสอนให้โมเดลได้เรียนรู้และแยกแยะเหตุการณ์ต่างๆที่เคยเกิดขึ้นในอดีตและเปรียบเทียบกับ เหตุการณ์ที่เกิดขึ้นในปัจจุบัน เช่น หากมีรถวิ่งวนเข้ามาที่เดมชาฯ หลารอบ แสดงว่ามีเหตุต้องลงสัญญาจะมีพัฒนาระบบที่ต้องการร้าย ซึ่งในอดีตมีเหตุการณ์ที่รายงานตัวก่อนหน้านี้ที่ สังเกตการณ์และจดทิ้งไว้หลังจากนั้นเกิดระเบิดขึ้นเป็นต้น ซึ่งโมเดล LSTM จะทำการเรียนรู้ข้อมูลจาก อดีตที่ผู้พัฒนาฝึกสอนและทำการพยากรณ์รวมถึงแจ้งเตือนให้ทราบทันทีที่มีเหตุการณ์ที่ต้องลงสัญญา ก็เกิดขึ้น

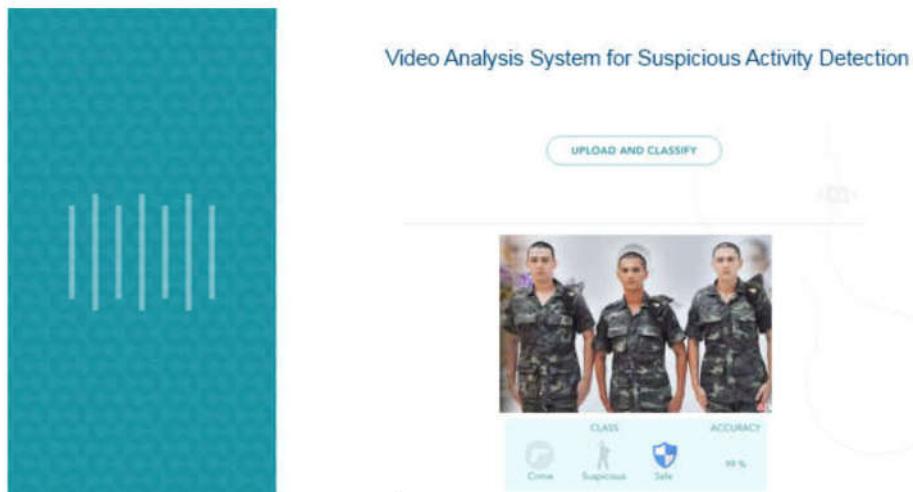


ภาพที่ 9 : นำภาพที่รวมเป็น Pattern เดียวกันเข้าเรียนรู้และทำนายผลใน LSTM [10]

5.1.5 ในขั้นตอนสุดท้ายจะนำผลที่ได้จากการวิเคราะห์และพยากรณ์ข้อมูลเหตุการณ์มาแสดงผลบนหน้าจอคอมพิวเตอร์ หน้าจอถือมีอีกของเจ้าหน้าที่ที่รับผิดชอบ ตามที่เดิม เชื่อมโยงข้อมูลไว้ให้เจ้าหน้าที่สามารถแก้ไข ช่วยเหลือ หรือเข้ารับงับเหตุได้อย่างทันท่วงที



ภาพที่ 10 : ตัวอย่างการจำแนกประเภทภาพวิดีโอ



ภาพที่ 11 : ตัวอย่างหน้าระบบ

## 6. จุดเด่นของงาน และ ประโยชน์ในการนำไปใช้

### 6.1 จุดเด่นของงาน

ของระบบวิเคราะห์ภาพวิดีโอเพื่อตรวจหาเหตุการณ์น่าสังสัยด้วยเทคโนโลยีการเรียนรู้เชิงลึกคือการใช้เทคโนโลยีอย่าง Deep Learning เข้ามาช่วยในการวิเคราะห์ข้อมูลภาพวิดีโอด้วยมีศักยภาพในการประมวลผลที่มีความซับซ้อน เรียนรู้จำแนกแยกแยะสิ่งต่าง ๆ และทำนายผลของข้อมูล

### 6.2 ประโยชน์ในการนำไปใช้งาน

6.2.1 สามารถนำไปตรวจหาเหตุการณ์น่าสังสัยที่มีแนวโน้มไม่สงบ และช่วยระงับและช่วยป้องกันเหตุการณ์ไม่สงบในอนาคตได้

6.2.2 สามารถนำผลการวิจัยนี้ไปพัฒนาวัตกรรมกล้องวงจรปิดให้สามารถคาดการณ์หรือพยากรณ์เหตุการณ์ความไม่สงบโดยภัยได้

6.2.3 ช่วยในการแจ้งเตือนเจ้าหน้าที่ให้เข้าช่วยเหลือผู้ประสบภัยร้ายได้อย่างทันเวลาโดยผ่านระบบการวิเคราะห์แจ้งเตือน

6.2.4 สามารถแยกเหตุการณ์ผิดปกติ หรือ เหตุการณ์ความไม่สงบได้อย่างถูกต้องแม่นยำ และสามารถที่จะทำนายเหตุการณ์ที่เกิดขึ้นได้อย่างแม่นยำ

## 7. เอกสารอ้างอิง

- [1] [ฝึกงาน] แยกประเภทรูปภาพด้วย Deep Learning ที่ Wongnai โจทย์ใหญ่ที่ไม่ธรรมดា (ออนไลน์) จาก <https://life.wongnai.com/internship-image-classification-wongnai-a1dbc2890766>
- [2] Google AL Blog, Train your own image classifier with Inception in TensorFlow (ออนไลน์) จาก <https://ai.googleblog.com/2016/03/train-your-own-image-classifier-with.html>
- [3] Chatchawan Niyomthum, Neural Network 101 : CNN with TensorFlow (ออนไลน์) จาก <https://medium.com/@thebear19/neural-network-101-cnn-with-tensorflow-fd5d515e979b>
- [4] Colah's blog, Understanding LSTM Network, August 27, 2015 (ออนไลน์) จาก <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [5] A toy convolutional neural network for image classification with Keras , 2017 (ออนไลน์) จาก [https://www.kernix.com/blog/a-toy-convolutional-neural-network-for-image-classification-with-keras\\_p14](https://www.kernix.com/blog/a-toy-convolutional-neural-network-for-image-classification-with-keras_p14)
- [6] TensorFlow, Image Recognition (อ อ น ไ ล น์ ) จ า ก [https://www.tensorflow.org/tutorials/images/image\\_recognition](https://www.tensorflow.org/tutorials/images/image_recognition)
- [7] Keras Document,Keras: The Python Deep Learning library (ออนไลน์) จาก <https://keras.io/>
- [8] L. M. Rasdi Rere,Mohamad Ivan Fanany, and Aniati Murni Arymurthy. Metaheuristic Algorithms for Convolution Neural Network. 2016
- [9] Architecture of a CNN. (ออนไลน์) จาก: <https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks-1489512765771.html>)
- [10] Yuhang He, Jianda Chen and Long Chen. 2017. Let the robot tell: Describe car image with natural language via LSTM.