

# Abnormal Activity Detection and Notification Platform for Real-Time Ad Hoc Network

<https://doi.org/10.3991/ijoe.v16i15.16065>

Suriya Pinitkan <sup>(✉)</sup>, Nawaporn Wisitpongphan  
King Mongkut's University of Technology North Bangkok, Thailand  
s5607011956145@email.kmutnb.ac.th

**Abstract**—As aging society era is getting near, number of elders who live alone is increasing. These people often need special care. Due to this reason, we propose Abnormal Activity Detection and Notification Platform (AADN) for Real-Time Ad Hoc Network which can help taking care of these people. The proposed platform relies on human tracking using cameras that are installed in different rooms inside the house. AADN will take as input images from the cameras to process and output activity in the form of human pose and objects with their relative distant to the detected human. Relationship Degree of Human Object Interaction (RD-HOI) will be analyzed every minute and be used to distinguish abnormal behavior by means of decision tree. In addition, activities will be used to generate routine behavior log and AADN will notify the person in charge of taking care of the subject if the detected activity differs from the routine. The proposed platform can achieve human pose accuracy of up to 99.66% by using COCO with VGG-NB model and can correctly identify object 68% of the time. Our experiments showed that AADN could notify abnormal activity by using RD-HOI when human and harmful objects were clearly visible in the picture and could correctly notify abnormal activity when time spent in a certain activity differed from the routine by a certain threshold given sufficient amount of data.

**Keywords**—Human Pose Estimation, Object Detection, Anomaly Detection

## 1 Introduction

The Internet of Things (IoT) is a technology for connecting objects, people, and services to provide valuable information. Nowadays, it has a variety of applications, one of which is the development of smart home system. The concept of the smart home system is to apply IoT to the residence to provide convenience, life and property safety, energy saving, and health care for the residents [1], etc. The development of these systems focuses on meeting the needs of the residents. Smart home requires the connection to the communication network in order to connect electrical appliances, services, surveillance, and access to the control of various devices. The system can be controlled from either inside or outside the house. As for the technology used for surveillance of life and property safety, artificial intelligence (AI) is often used as a tool to detect

changes in activity` pattern and human behavior. However, input to the detection system varies from sensor data to images depending on the scope and the purpose of the studies [2-4]. Sensor data tracking coordinates of human, such as accelerometer data, obtained from a wearable device is one of the most popular input data for human behavior tracking used in numerous studies [2]. Besides sensor data, there were also many studies which used images as input for detecting behavior. For examples, Xu, et. al [3] used the convolutional neural network (CNN) and multi-class linear support vector machine (SVM) classifier on the optical flow energy images (OFEI) to implement intelligent video surveillance system in parking environment. The proposed system can recognize abnormal behavior such as smashing car, robbery, fighting, and faint. Thummala and Pumrin [4] used motion history images as input to do fall detection. Last but not least, Miaou et.al [5] used 360-degree images along with personal information (height and weight) as input to the threshold-based fall detection system.

Many researchers focus on using artificial intelligence to detect whether human poses such as walking, sitting, and sleeping are normal [6]. When abnormal pose is found, a notification will be sent to the caregiver for help. According to our study, existing elderly care platforms shared similar goal which is to provide an early warning when detecting abnormal activity. Image processing together with artificial intelligent or deep learning is one of the most popular technique used for this purpose. In this study, on the other hand, we provide a new definition of abnormal activities by means of studying the activity pattern or behavior of the subject. More specifically, daily activities will be logged in order to find the activity pattern or daily routine. In addition to detecting abnormal activity in real time based on current situation, the proposed AADN will also notify care givers if the subjects perform activities which do not conform to the regular pattern by a certain threshold. Hence, we proposed a novel algorithm for detecting abnormal in-house activities using both the snapshot image and the activity log history.

## **2 Related Work**

As our world is entering the aging society, one problem we have to encounter is taking care of the elders who need special care. Elderly caregivers, including physicians and nurses, are now inadequate for the needs of the elders. As a result, artificial intelligence is used as a tool to provide good care for the elders and to improve the quality of life: physical and mental health care. Artificial intelligence or AI is often used as a tool to detect changes in activity pattern and behavior of the elders, including emergency situations such as slip and fall. An artificial intelligence detects whether a person's body movement such as walking, sitting, and sleeping is normal pose. When an abnormal posture is found, a notification will be sent to the caregiver for help.

To detect abnormal posture, ones need to be able to detect human postures. There are quite a number of research work which focus on developing algorithms for detecting human posture and body movement. These studies usually rely on image processing to identify the key points of the main joints, including neck, elbows, knees, shoulders, etc.

The result of human posture estimation is a 2D image, in which there are 2 commonly used datasets, Common Object in Context or COCO [7] and Multi-Person Dataset or

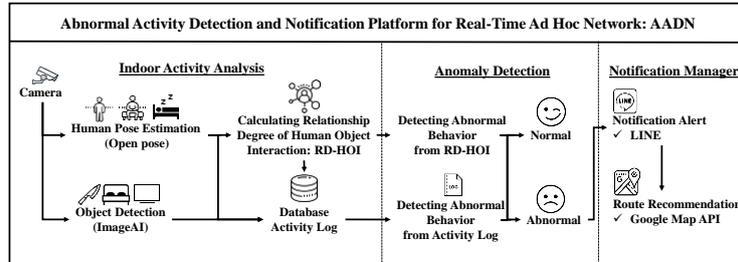


Fig. 1. Abnormal activity detection and notification platform for real-time ad hoc network.

MPII [8]. The COCO dataset annotates 18 key points of the body while 15 key body points are specified in MPII dataset.

To detect a human in an image, a Convolutional Neural Network (CNN) is a common method used for estimating the human posture [7]. Among the many methods reviewed in [9], Cao et al. [10] designed a multi-branch multi-stage CNN architecture with 10 layers of feature extraction using VGG-19 model developed by Visual Geometry Group (VGG) [11] with 3 processing parts to detect multiple people in the image. The first part was used for determining the position of the joint. The second part was used to analyze the direction and connection between the joints. The combined results of the first two parts are then used for refining the predictions. However, the body posture detection alone may not be sufficient for determining whether the pose observed is abnormal. Therefore, it is necessary to detect objects in the scene [12] to help identifying any changes or abnormalities in the physical environment in order to make appropriate alert to the caregivers.

The studies on object detection usually use image processing based on artificial neural networks, in which each network has a different number of layers with specific algorithm. Girshick et al. proposed the Fast-Recurrent Convolution Neural Network (Fast R-CNN) [13] that used the entire image as input and the entire image was processed by convolution and maximum pooling for several times to create the feature maps. A technique called Region of Interest Pooling (ROI Pooling) was used during the region proposal to extract the fixed-size feature vector of each object which is later being fed to a two output layers: one for estimating the object class and the other for finding the bounding-box positions. Faster R-CNN [14] proposed by Shaoqing, et al. replaces ROI selective search method used in Fast R-CNN with the Region Proposal network (RPN) which relies solely on GPU computation. As a result, it runs 10 times faster than Fast R-CNN due to the elimination of the message exchanges between CPU and GPU.

The Single ShotMultiBoxDetector (SSD) [15], on the other hand, achieves faster classification and higher accuracy than the Faster R-CNN [16] by eliminating the proposal generation step and rely on a single-stage deep neural network. Only an image and the ground truth box (actual bounding box of an object) are required as input during normal training process. A set of default boxes are then evaluated over different aspect

ratios and several feature maps scales at each location. Size, shape, and confidence of each object category is predicted for each default box. However, the limitation of the proposed single-stage SSD is the imbalance between the foreground and background due to the easily classified background images during the training phase. Hence, Lin et al. [16], later proposed RetinaNet model with improved loss function and used Feature Pyramid Network (FPN) [17] as the backbone network for object detection to attain better speed and accuracy.

To further detect the activity or human behavior, we proposed a framework which combine the results of the state-of-the-art human pose detection (MPII and COCO) and object detection (RetinaNet, YOLOv3, Tiny YOLOv3) to construct a matrix which allows us to predict abnormal activities and make appropriate notification.

### 3 Research Methodology

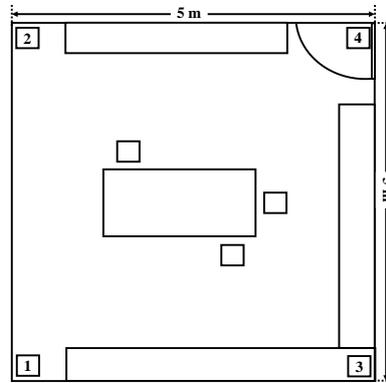
In the designing of an abnormal behavior analysis platform for real-time emergency alerts in an ad hoc network (AADN), the data from cameras installed in different areas and rooms in a house were imported as input in order to analyze the activities in each area of the house. Activities in each area were also recorded on the server to be used as activity history data for the analysis and learning of normal behavior. At the same time, the activity history data was processed for the abnormal activities in real-time. In case there was an abnormality in the activity, the system could select a communication channel to notify such abnormality via appropriate communication channel corresponding to the abnormality of that activity as shown in Fig. 1.

#### 3.1 Indoor activity analysis

Indoor activity analysis involved the processing of images obtained from video camera and analysis into the activities of residents. In the scope of this research, activities included different poses, consisting of sitting, standing, and sleeping, along with the evaluation of relationships of human pose and nearby objects in the room. Therefore, indoor activity analysis consisted of 1) human pose estimation, 2) object detection, 3) estimation of relationship degree of human-object interaction, and 4) activity log analysis.

1. Human pose estimation: Human pose estimation is the function used to predict human pose. In this study, human pose estimation consisted of 3 parts: 1) finding the location for camera installation to collect human poses, 2) preparation of human pose data and 3) development of learning process.

To find the location for camera installation to capture human poses, the cameras must be installed in a suitable location to be able to capture the overall atmosphere of the room where recorded data can be used for training and testing. In this regard, we installed 4 cameras in a 5x5 m<sup>2</sup> room at position 1-4, as shown in Fig 2, which were four corners of the room and were the positions that allowed the best visibility of the elements in the room.



**Fig. 2.** Camera installation positions

During the experiment, we varied the camera height from 1.50 to 2.20 meters to find the best camera installation height which could capture the posture of a 1.70-meter tall subject at various location within the room.

The camera used in this study operates at 720p/30fps with auto-focus and 78-degree field-of-view. The collected data were divided into a training dataset of 480 images and a testing dataset of 120 images in order to compare the efficiency of 4 different VGG models: convolutional neural network (VGG-16), Decision Tree (VGG-DT), Support Vector Machine (VGG-SM) and Naïve Bayes (VGG-NB) models.

According to the results shown in Table 1, the installation height of the camera affected the accuracy of the models tested. When the installation height of the camera was lower than the height of the subject in this experiment, who was 1.70 meters, all models tested showed relatively low accuracy in human pose predictions. When the height of the camera was 1.50 meters, the average accuracy of all models was 64.83%. However, when the installation height of the camera was higher than 1.70 meters, it was found that most of the models tested showed accuracy of 100%. Hence, the camera installation height should be higher than the tallest person in a room. In this study, we installed the camera at the highest possible height of 2.20 meters high. However, the optimal height should also depend on the camera type and room setting. That is, it might not be optimal to install the camera at the highest possible height if the camera has low resolution, wider field-of-view, or the room has high ceiling.

**Table 1.** Accuracy of camera altitude on pose estimation

Models	Height level (meter)			
	1.50m	1.80m	2.00m	2.20m
VGG-16	65.09%	100.00%	100.00%	100.00%
VGG-DT	64.30%	99.75%	100.00%	100.00%
VGG-SVM	65.09%	100.00%	100.00%	100.00%
VGG-NB	64.82%	100.00%	100.00%	100.00%

To develop a human pose prediction model (HPPM), images of people in different poses were converted into a skeleton-like-image using the pre-trained model on COCO and MPII dataset. More specifically, the human detected in each image will be converted to a skeleton with 18 key points using COCO model and a skeleton with 15 key points using MPII model. For pose prediction, Caffe Deep Learning Framework is used for training 400 images and testing 100 images. The result of this learning process was Human Pose Prediction Model (HPPM).

In AADN, HPPM is used for predicting human pose in real-time. The images acquired from the cameras were processed through Caffe Deep Learning using COCO and MPII techniques. The skeleton obtained from image processing was then imported into learning process using VGG-16 model, in which the parameters of the 16th layer or last layer of this model were adjusted to classify human poses into sitting, standing, and sleeping. Let the prediction result from camera  $is$  at time  $to$  be  $p_{t,i}$  which can either be “*sit*”, “*stand*”, “*sleep*”, or in the case where there is no one in the image the prediction result is “*idle*”. With multiple cameras setting, we used voting model to combine the predicting results from all the cameras. For example, the overall predicted result ( $p_t$ ) is “*sit*” if most cameras predicted that the human is sitting. Note that “*idle*” status will be excluded from the voting process. For example, if the prediction result at time  $t$  of a room with 4 cameras is {sit, sit, stand, idle} which means that the 1<sup>st</sup> and 2<sup>nd</sup> cameras predict that such person is in the sitting pose, 3<sup>rd</sup> camera predicts the standing pose and the 4th camera predicts that no one is in the image, the result of this image processing is sitting pose.

In case the voting results are equal, activity log is used as the determinant. For example, if the prediction of the 4 cameras installed in a kitchen is {sit, sit, stand, stand} at 9.30am, AADN will rely on the activity history from the database during the time of 9.30. If the history log indicates that a person is more likely to sit than stand, then the result is sitting posture. Detail regarding the history log is discussed in more detail in Section III.4

2. Object detection: Object detection is a mechanism that allows computer to understand the scenery or distinguish objects from the scene. In this study, detection of objects in the image relied on ImageAI tool for image processing and COCO dataset was used as input for developing the learning process. ImageAI classified objects in COCO Dataset into 80 types of indoor and outdoor objects. However, since the scope of this study is indoor activity detection, the training focuses only on indoor objects. As a result, there were 11 types of indoor objects which were 1) people, 2) bed, 3) chair, 4) knife, 5) cup, 6) microwave, 7) laptop, 8) books, 9) mouse, 10) remote control and 11) TV. In this regard, the researcher selected 300 images for each type of indoor object to be used as a dataset for determining the efficiency of the multiple object detection techniques.

The object detection techniques used in this research consisted of 1) RetinaNet, a powerful model for object detection with high accuracy. However, this model takes a long time in processing. Therefore, it is suitable for detecting common objects that do not require rapid detection; 2) You Only Look Once (YOLO), a popular model for detecting objects on the road. Version 3 of this model has moderate efficiency

which is suitable for detecting objects that require rapid detection; and 3) TinyYOLOv3, a lightweight version of YOLO model which can process faster but has slightly lower efficiency and accuracy. All these 3 techniques were tested during the learning process using all indoor objects within the COCO dataset. The results of each technique included name of object detected, bounding box, which was represented by the upper left ( $x_1, y_1$ ) and lower right ( $x_2, y_2$ ) coordinates of the rectangle, and the reliability ranging from 0 to 100, where 100 means the probability that the object detected is the result is 100%.

Table 2 shows the performance comparison of the three techniques used to detect indoor objects. In the experiment, we tested COCO database of 3,300 indoor images (300 images per 1 object type) and compare the results in terms of 1) the human detection accuracy; 2) the detection reliabilities for 11 different types of objects; and 3) processing time for object detection. The most effective technique will later be used to detect objects in the AADN platform.

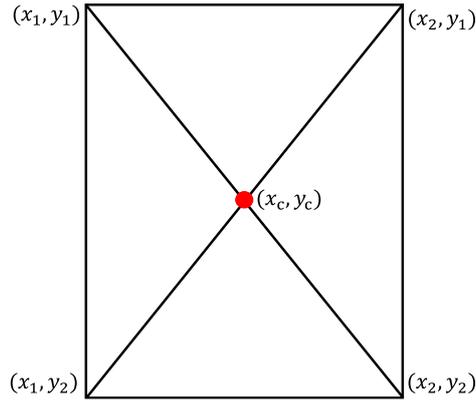
**Table 2.** The performance comparison of the object detection techniques

No	Object	RetinaNet	YOLOv3	TinyYOLOv3
1	people	92%	<b>96%</b>	24%
2	bed	76%	90%	31%
3	chair	75%	78%	26%
4	knife	65%	62%	24%
5	cup	45%	52%	35%
6	microwave	68%	73%	28%
7	laptop	85%	87%	34%
8	book	70%	76%	26%
9	mouse	65%	64%	28%
10	remote control	36%	38%	23%
11	tv	86%	88%	37%
12	average detection accuracy	65%	<b>68%</b>	25%

According to the result in Table 2, YOLOv3 outperformed the other two techniques at detecting objects. More specifically, accuracy of human detection was as high as 96% while the detection accuracy of RetinaNet and TinyYOLOv3 were 92% and 24%, respectively. As for the overall object detection performance, YOLOv3 could detect objects with only 68% accuracy. The results showed that smaller objects were more difficult to detect compared to the larger ones, as can be seen from the detection accuracy of remote control, mouse, knife, and cup which showed lower average values compared to bed, chair, microwave, etc. However, the test results showed that YOLOv3 model could detect several types of objects with highest accuracy. Therefore, YOLOv3 was chosen for further detection of human and objects in AADN.

3. Estimation of relationship degree of human-object interaction: The estimation of relationship degree of human-object interaction (RD-HOI) was performed by measuring the distance between human and each type of object in the radius of interest. First, the bounding box of object detected was used in finding the center of the

bounding box or centroid. As is shown in Fig 3, let the object bounding box be represented by 4 coordinates, i.e., top-left corner  $(x_1, y_1)$ , bottom-left corner  $(x_1, y_2)$ , top-right corner  $(x_2, y_1)$  and bottom-right corner  $(x_2, y_2)$ , the centroid value can be calculated using equation 1 and 2.

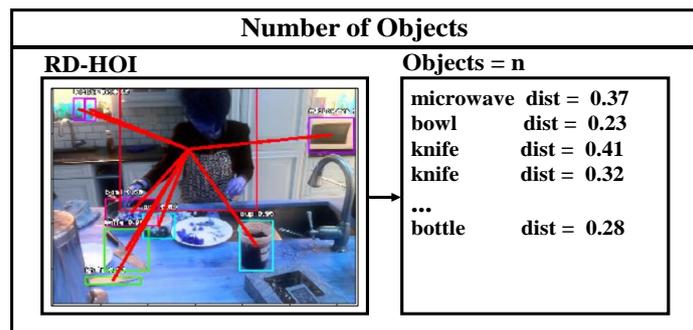


**Fig. 3.** The position of the object frame.

$$x_c = \frac{x_1 + x_2}{2} \tag{1}$$

$$y_c = \frac{y_1 + y_2}{2} \tag{2}$$

The calculated centroid of each object was then used to calculate the distance to the centroid of human detected in the scene. The calculated distance is then normalized by the diagonal length of the image. To further illustrate RD-HOI, Fig 4 shows the results of RD-HOI calculation of a picture with a person in a kitchen surrounded by multiple objects.



**Fig. 4.** Output of relationship degree of human-object interaction

According to Fig 4, there were **n** number of objects detected. RD-HOI is then expressed as the normalized distance between the object and human. The closer the value

to 0, the closer the object is to human and the higher the RD-HOI. In case where more than one camera was installed in a room, the calculated distances between people and objects from all cameras were compared and the value closest to 1 was selected as potential RD-HOI.

4. Activity log analysis: Activity log is an essential part in AADN because we believe that people often follow a certain routine while at home. Hence, in this study we also collected an activity log with time stamp using the data shown in Table 3.

**Table 3.** Description of data storage

Data	Description
Position	The name of the room. (In this study, the cameras were installed in a bedroom, a kitchen, and a living room.)
Activity	The results of human pose analysis, i.e., sit, stand, and sleep.
Year-month-day	Date on which the data is collected using YYYY-MM-DD format.
Time	Time at which the data is collected/analyzed in the format of HH: MM: SS. xxxxxx
Object Detection	Object detected in the scene in the format of “name   location $x_1$   location $y_1$   location $x_2$   location $y_2$   RD-HOI”

Fig 5 shows example of an entry in the activity log. This entry indicates that on July 25, 2019 at approximately 13:43pm the cameras detected a human sleeping in a bedroom with a laptop not very far from him/her.

bedroom, Sleep,2019-07-25, 13:43:48.687944,laptop 557 240 85 405 0.11
---

**Fig. 5.** Example of an activity log entry

To further save the storage space, the activity log can optionally be recorded only in the event that the cameras detected a human in an image and there were some changes in the log entry during the 30-second time interval. The change could be caused by one of the following reasons: a person expressed a different posture, new objects were detected, or RD-HOI to objects changed.

### 3.2 Anomaly detection

1. Detecting abnormal behavior from RD-HOI: In this step, we classified abnormal activities using RD-HOI of different objects. The techniques used in this study were 1) Decision Tree (DT), 2) Artificial Neural Network (ANN) and 3) Support Vector Machine (SVM). Data used in learning process were the information obtained from the activity log analysis, which consisted of 1) the name of the room 2) human pose data 3) RD-HOI of each detected object. The indoor activities were collected for 7 days and each log entry were classified into normal or abnormal behaviors. Examples of abnormal behavior include a person sitting near a knife in a bedroom, a person sleeping in the kitchen, a person sleeping in the living room but away from the chair (sofa), etc.

2. Detecting abnormal behavior from activity log: In this study, abnormal behavior should be detected when a person performs an unlikely activity in a wrong place at a wrong time. To achieve this, AADN collects *activity log* which represents a person's daily routine. For example, AADN should alarm someone when a person under supervision overslept by a certain amount of time or slept in a room where he/she shouldn't be sleeping in. To be able to do so, ones need to collect reasonable amount of activity data to at least be able to observe the routine behavior. Note that the period of activity log collection for building a suitable dataset in order to learn the behavior of a certain person in AADN platform varied according to each person's lifestyle. For a resident who had a routine lifestyle, in which the activities were carried out in a relatively fixed pattern, such as waking up at 8:00 am, eating at 12:00 am -13:00 pm and sleeping at 8:00 pm, data collection requires much less time for learning and establishing abnormality rules compared to those who live without a pattern.

The steps for detecting abnormality from an activity log are as follows:

**Step 1.** Processing the activity log in the database every hour. The format of activity log after the processing is as follows:

YYYY-MM-DD HH ROOM  $T_{sit}$   $T_{sleep}$   $T_{stand}$   $T_{idle}$

where

- *YYY-MM-DD* is the date, month, year that an activity was recorded.
- *HH* is the hour that an activity was recorded.
- *ROOM* is the name of the room that an activity was recorded.
- $T_{sit}$  is a duration that a person is sitting during an hour that an activity was recorded.
- $T_{sleep}$  is a duration that a person is sleeping during an hour that an activity was recorded.
- $T_{stand}$  is a duration that a person is standing during an hour that an activity was recorded.
- $T_{idle}$  is a duration that a person is absent from the room during an hour that an activity was recorded.

**Step 2.** Determining the normality of the behavior by considering the average and standard deviation of each pose during the HH period for  $D_{observe}$  days, where  $D_{observe}$  is the number of days under observation. Such method allows AADN to set up a benchmark activity pattern at any time of the day. For example, a person under supervision is usually in a bedroom sleeping for  $30 \pm 10$  minutes, sitting for  $5 \pm 3$  minutes, standing for  $10 \pm 5$  minutes, and not in the room for  $15 \pm 5$  minutes during 7.00am – 8.00am.

**Step 3.** The system alerts when abnormal activity is detected. For the analysis of abnormal activity, the current duration of each pose in each room was monitored and compared to the average ( $\mu$ ) and the standard deviation ( $\sigma$ ). Abnormal behaviors can occur in 2 cases, which are 1) the duration of each pose is greater than  $\mu + c\sigma$  or less than  $\mu - c\sigma$ , where  $c$  is a pre-determined constant and 2) there are activities that have never occurred during that time.

For example, from the 3-day data collection of Mr. A's behavior, it was found that between 8:00 am - 9:00 pm Mr. A was in a kitchen all the time. Fig. 6 illustrated Mr. A's routine activity in the kitchen.

Kitchen 8.00-9.00 am
Sit $\mu = 45$ minutes $\sigma = 5$ minutes
Stand $\mu = 15$ minutes $\sigma = 10$ minutes
Sleep $\mu = 0$ minutes $\sigma = 0$ minutes

**Fig. 6.** Example of routine activity

From Fig. 6, Mr. A sat in the kitchen for an average of 45 minutes, stood for an average of 15 minutes and had never slept in the kitchen during that time. Therefore, if the system found that Mr. A appeared to be sleeping in the kitchen, the system must report that it was an abnormal behavior or if the system found that he stood for more than 35 minutes ( $c=2$ ) or did not stand in the kitchen during the observation time, the behavior of Mr. A was also considered abnormal.

### 3.3 Notification management

When AADN detected abnormal activity, the system sent a notification via LINE (a popular instant messaging application in Thailand) by showing the captured image after the anomaly was detected along with the information indicating abnormal activity, such as a person sleeping in the kitchen at 10:00 a.m.

For a certain abnormal activity which may require medical care, AADN can also help choosing the best route to the hospital using the available Google Map API to find the route with the least travel time. To achieve this, AADN supports the import of the house coordinates and the coordinates of all nearby hospitals. When the system notified an anomaly via LINE, the recipient could inspect the incident and choose the fastest route to the hospital right away.

## 4 Results

### 4.1 Indoor activity analysis results

The AADN platform requires a model for accurate human pose analysis. Therefore, we conducted an experiment to find the best model by testing the accuracy of the human pose from the cameras installed at all 4 points. The models tested included VGG-16, VGG-DT, VGG-VM, and VGG-NB, together with COCO and MPII techniques. The estimation of the accuracy of each of these models was performed by random sampling the images from the cameras at 1 frame/second. A total of 600 pose images were collected. Test results are shown in Table 4.

The results showed that human pose estimation from Camera 4 with COCO technique achieves as high as 100% accuracy across all models, while MPII technique performs slightly worse. This was because with COCO technique, image of a human was

transformed into skeleton with 18 key points, while the skeleton in MPII had only 15 key points. In addition, it was found that the human body in the Camera 1 was obscured by objects, resulting in incomplete skeleton image. Hence, MPII technique which had fewer key points performed quite poorly on Camera 1. However, accuracy of VGG-NB and VGG-DT with the 18 key-point COCO technique was 99.45% even with a partially blocked view. Therefore, COCO technique was more effective than the MPII technique and the model with the best results was VGG-NB followed by VGG-DT.

**Table 4.** Accuracy results

Dataset	Models	Accuracy (%)				
		Camera 1	Camera 2	Camera 3	Camera 4	Average (%)
COCO	VGG-16	96.49	98.92	97.30	100.00	98.18
	VGG-DT	99.45	98.37	99.18	100.00	99.25
	VGG-SVM	96.75	98.64	97.29	100.00	98.17
	VGG-NB	99.45	99.18	100.00	100.00	99.66
MPII	VGG-16	93.24	94.30	99.46	99.19	97.30
	VGG-DT	92.70	96.48	99.72	99.45	97.09
	VGG-SVM	92.70	96.48	99.72	99.45	97.09
	VGG-NB	92.97	96.48	100.00	99.72	97.29

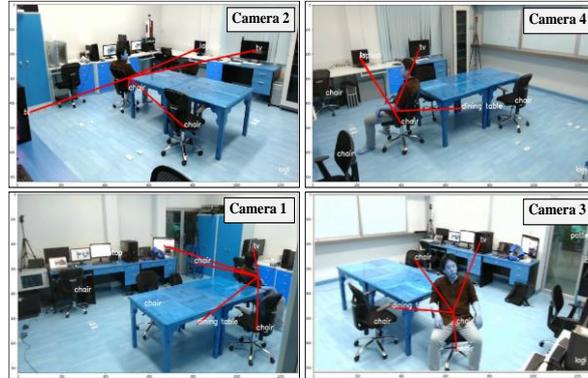
In a real system, the results from all 4 cameras were required. In this study, voting was used. When COCO was used in combination with VGG-NB and VGG-DT models, the accuracy values were 99.79% and 70.00%, respectively. According to the results shown in Table V, VGG-NB yielded much better accuracy at the expense of extra computation power and processing time.

**Table 5.** Voting results

Dataset	Models	Accuracy (%)
COCO	VGG-NB	99.79
	VGG-DT	70.00

To calculate the relationship degree of human-object interaction, YOLOv3 was used for detecting people and objects in the scene. Example of the results from Camera 1-4 are shown in Fig 7.

Fig. 7 shows RD-HOI from 4 different cameras when AADN limits the number of objects of interest to 5 objects: considering only 5 nearest objects. In the example, the five nearest objects were 1) TV at 0.07, 2) chair at 0.14, 3) dining table at 0.22, 4) chair at 0.20 and TV at 0.30. In this case, it was found that the system detected some object correctly, but incorrectly identified the computer screen as TV or portable computer, as is shown in Table 6.



**Fig. 7.** Relationship Degree of Human-Object interaction Camera 1-4 with 5 objects of interest

However, the objects detected and the calculated distances to objects (RD-HOI) in each camera angle were different. That is, the objects detected at each camera angle might be different objects and the distance to the same object calculated from the images from different angles might have different distances as shown in Table 6 and Fig. 7. If there was only one object in the room, such as a dining table, the system could measure the distances and compared the results of all 4 cameras and used the longest distance. Table 6 showed that a dining table was at a distance of 0.22, 0.18 and 0.20 from the person in a room, when using the image from Camera 1, 3 and 4, respectively. In this case, we use 0.22 as reference distance to the dining table.

**Table 6.** The results of the calculation of relationship degree of human-object interaction considering Camera 1, 2, 3 and 4

Objects	Camera 1	Camera 2	Camera 3	Camera 4
tv(1)	0.07	-	-	-
tv(2)	0.30	-	-	-
tv(3)	-	0.34	-	0.18
tv(4)	-	0.40	0.21	-
dining table	0.22	-	0.18	0.20
chair (1)	-	0.03	0.03	0.04
chair (2)	-	0.21	0.18,0.03	-
chair (3)	0.14	-	-	-
chair (4)	0.20	-	-	-
labtop	-	0.23(tv)	-	0.18

In the scope of this study, AADN simply concern about abnormal behavior which involves harmful objects such as knife or gun. Therefore, RD-HOI accuracy to common objects are not within the scope of this study. In the case where there were more than one object of the same type such as computer screen and chair, AADN will treat each detected object independently and will not attempt to distinguish each unique object as it is not necessary for the scope of this study. Finally, the calculated RD-HOI between

a person and objects is included in the activity log entry which were later used in further analysis of abnormal behavior.

#### 4.2 Abnormal behavior analysis results

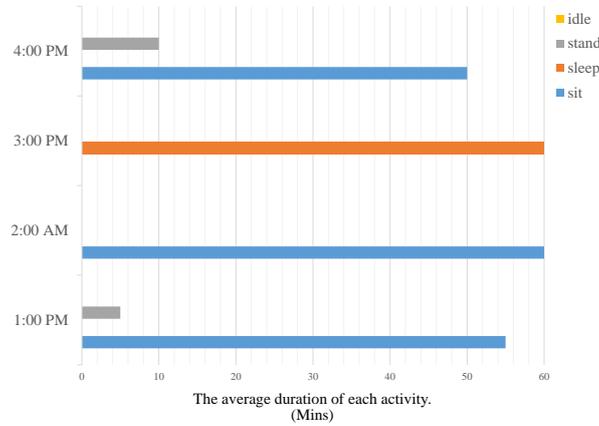
1. Detecting abnormal behavior from RD-HOI: Activity log entry obtained from the analysis of human-object interaction consisted of 1) the name of the room, 2) human pose data and 3) name of object and RD-HOI. For this experiment, we collected 7-day activity log which total to 1,265 sets of log entries. Data log were divided into 1,226 training datasets and 39 testing datasets. The classification techniques used in this study were 1) Decision Tree (DT), 2) Artificial Neural Network (ANN) and 3) Support Vector Machine (SVM). The experimental results are shown in Table 7.

**Table 7.** Experimental results

Models	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
DT	100.00	100.00	100.00	100.00
SVM	97.91	100.00	61.11	67.62
ANN	100.00	100.00	100.00	100.00

The results showed that Decision Tree (DT) and Artificial Neural Network (ANN) were the most effective methods for detecting abnormal indoor behavior. In the specified scope of the experiment, it was found that 100% of the anomaly could be identified from the activity log. In the case that the dataset had no complicated classification conditions, decision trees could classify the behavior as good as artificial neural networks. Therefore, we chose to implement decision tree in our AADN platform due to its simple design, fast processing time, and less resource requirement.

2. Detecting abnormal behavior from activity log: With AADN, abnormal behavior can also be detected from the activity log. As explained in III.B.2, any activity that differs from the daily routine pattern by a certain threshold is considered abnormal. To establish a daily routine to be used for anomaly detection, AADN supports the enhancement of the daily routine learning process by incrementally averaging the new activity log into the existing history log. Users can view the summarized routine in each room at any given time of the day, as shown in Fig. 8.



**Fig. 8.** Example of summarized behavior chart from daily activity log in a living room.

The activities were monitored on an hourly basis and compared with the routine behavior. AADN issues notification in the scenario where:

- The duration spending in each pose of each activity was greater than  $\mu + c\sigma$  or less than  $\mu - c\sigma$ .
- There were activities that had never occurred during that time before.

Fig. 9 shows example of unusual behavior in the activity log. AADN showed that there was a person sleeping in the bedroom at 1.44 pm. According to the normal daily routine in Fig. 8 which shows that this person has never stayed in the bedroom during such time, hence, this activity was classified as abnormal.

```
living room,sit,2019-07-25 13:42:53,table|250|404|107|420|0.17,normal
bedroom,sleep,2019-07-25 13:44:25,bed|250|404|107|420|0.11,abnormal
bedroom,sleep,2019-07-25 13:46:45,bed|250|404|107|420|0.11,abnormal
```

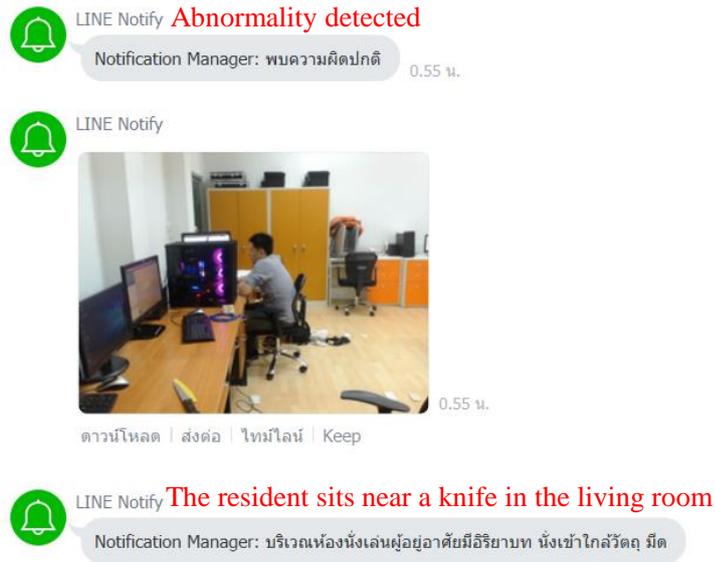
**Fig. 9.** Example abnormal behavior detection from activity log.

Another example of abnormal behavior is illustrated in Fig. 10 which shows an alert when a person slept longer than usual. Suppose that an activity log history indicates that a person normally sleeps for  $20 \pm 5$  minutes during 7.00am – 8.00am. If  $c = 2$ , the system notified someone when a person slept for less than 10 minutes or more than 30 minutes during that one-hour time period.

```
bedroom,sleep,2019-07-25 7:28:15,bed|250|404|107|420|0.11,normal
bedroom,sleep,2019-07-25 7:29:25,bed|250|404|107|420|0.11,normal
bedroom,sleep,2019-07-25 7:30:26,bed|250|404|107|420|0.11,abnormal
```

**Fig. 10.** Abnormal activity resulted from oversleeping.

3. Notification management: To get benefit from AADN, the resident must be the user of LINE application. With the use of LINE application, the effectiveness of the notification system was limited to the quality of service provided by LINE Notify service.



**Fig. 11.** Example LINE Notification.

Once the user is informed about the abnormality, picture of the abnormal activity will be sent along with the notification. User can optionally request fastest route to the hospital. The system, then, send a request for the routes and travel times to the pre-registered hospitals via Google API. AADN, then notified the names and fastest routes to the hospitals via LINE from the coordinates of the house specified in the system. The notification results are shown in Fig. 12.

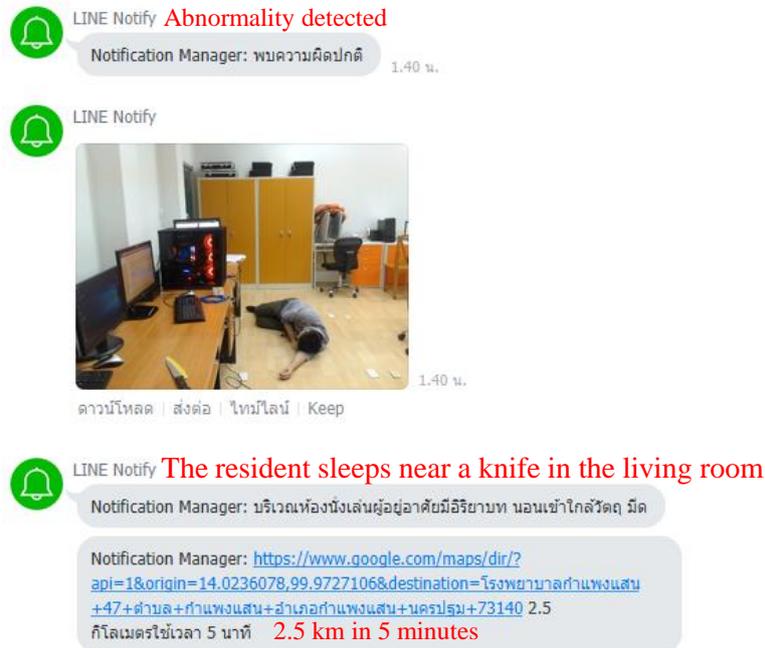


Fig. 12.Example of a notification with fastest route to the hospital

## 5 Conclusion

In this study, we proposed Abnormal Activity Detection and Notification Platform (AADN) which introduced a new method for identifying abnormal activities of the home residents. The basis of AADN lies in an integral of multiple image processing and classification techniques which include object detection, object classification, and human pose estimation. Once AADN detects objects within a scene, it will try to locate human and nearby objects using YOLOv3. If human is present in the scene, human is extracted to skeleton-like image using COCO model with 18 key points and VGG-NB will estimate the human pose.

AADN can detect abnormal behavior using 2 different methods. The first method is the detection of abnormal behavior from the proposed relationship degree of human and object interaction (RD-HOI). The activity data consisting of name of the room and the distance from a person to surrounding objects or RD-HOI can be used to distinguish normal and abnormal behaviors. AADN used a decision tree technique for classifying abnormal behavior because it is a lightweight technique and can easily be applied on the platform with comparable classification efficiency to the artificial neural network, which takes longer processing time. The second method to detect abnormal behavior, proposed in this study, is the anomaly detection from the activity log, in which AADN will send notification when a resident is found in a particular pose for a longer or less duration than usual in each room and time period.

The current implementation of AADN can support only one person at a time and the activities detected by the current AADN are only standing, sitting, and sleeping. In the future, we plan to improve AADN by including detection of other behaviors such as falling asleep while sitting, crying, falling, or exercising and identifying the same object from multiple cameras. Lastly, the proposed method for keeping track of the user's routine behavior can further be extended from an hourly routine summarization to a daily routine summarization in order to better reflex the user's daily lifestyle at a macro level.

## 6 References

- [1] Elaydi, H. A. (2019). Personal Health Care System Using IOT. *International Journal of Online and Biomedical Engineering (iJOE)*, 15(07). pp 16-31 <https://doi.org/10.3991/ijoe.v15i07.10265>
- [2] Tu, Y., Liu, L., Li, M., Chen, P., & Mao, Y. (2018). A Review of Human Motion Monitoring Methods using Wearable Sensors. *International Journal of Online and Biomedical Engineering (iJOE)*, 14(10). pp 168-179 <https://doi.org/10.3991/ijoe.v14i10.9313>
- [3] Xu, H., Li, L., Fang, M., & Zhang, F. (2018). Movement human actions recognition based on machine learning. *International Journal of Online and Biomedical Engineering (iJOE)*, 14(04). pp 193-210 <https://doi.org/10.3991/ijoe.v14i04.8513>
- [4] Thummala, J., & Pumrin, S. (2020). Fall Detection using Motion History Image and Shape Deformation. In 2020 8th International Electrical Engineering Congress (iEECON). pp. 1-4 <https://doi.org/10.1109/ieecon48109.2020.229491>
- [5] Miaou, S., Sung, P.H., & Huang, C.H. (2006). A Customized Human Fall Detection System Using Omni-Camera Images and Personal Information. In 1st Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare. pp. 39-42. <https://doi.org/10.1109/ddhh.2006.1624792>
- [6] Huang, Z., Liu, Y., Fang, Y., & Horn, B. K. (2018). Video-based fall detection for seniors with human pose estimation. In 2018 4th International Conference on Universal Village (UV). pp. 1-4. <https://doi.org/10.1109/uv.2018.8642130>
- [7] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham. [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)
- [8] Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2014). 2d human pose estimation: New benchmark and state of the art analysis. In Proceedings of the IEEE Conference on computer Vision and Pattern Recognition. pp 3686-3693 <https://doi.org/10.1109/cvpr.2014.471>
- [9] Liu, Y., Xu, Y., & Li, S. B. (2018). 2-D human pose estimation from images based on deep learning: a review. In 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). pp. 462-465 <https://doi.org/10.1109/imcec.2018.8469573>
- [10] Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp 7291-7299 <https://doi.org/10.1109/cvpr.2017.143>
- [11] Simonyan, K., and Andrew, Z. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In 3rd IAPR Asian Conference on Pattern Recognition (ACPR). pp 730-734 <https://doi.org/10.1109/acpr.2015.7486599>

- [12] Liu, B., Zhao, W., & Sun, Q. (2017). Study of object detection based on Faster R-CNN. In 2017 Chinese Automation Congress (CAC). pp 6233-6236 <https://doi.org/10.1109/cac.2017.8243900>
- [13] Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision. pp 1440-1448
- [14] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems. pp 91-99
- [15] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)
- [16] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision. pp 2980-2988. <https://doi.org/10.1109/iccv.2017.324>
- [17] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp 2117-2125. <https://doi.org/10.1109/cvpr.2017.106>

## 7 Authors

**Suriya Pinitkan** is a Ph.D. student in the Faculty of Information Technology at King Mongkut's University of Technology North Bangkok (KMUTNB), Thailand. He is also a lecturer at the Information and Communications Technology, Faculty of Liberal Arts and Science, Kasetsart University (Khamphaengsaen (Kamphaeng's Campus), Nakhon Pathom, Thailand. His research interest is in internet of things, data mining, machine learning, and artificial intelligence. Email: [s5607011956145@email.kmutnb.ac.th](mailto:s5607011956145@email.kmutnb.ac.th)

**Dr. Nawaporn Wisitpongphan** is an Assistant Professor in the Faculty of Information Technology and also a director of the Research Center of Information and Communication Technology at King Mongkut's University of Technology North Bangkok (KMUTNB), Thailand. She received her B.S., M.S., and Ph.D., in Electrical and Computer Engineering from Carnegie Mellon University in 2000, 2002, and 2008, respectively. Prior to joining KMUTNB, she was a researcher at General Motor Research Center, Warren, Michigan. While her expertise is in computer network, vehicle-to-vehicle communication, her current research focus is on Information Technology Management and Smart environments.

Article submitted 2020-06-05. Resubmitted 2020-10-09. Final acceptance 2020-10-10. Final version published as submitted by the authors.