

Data Analytics and Aggregation Platform for Comprehensive City-Scale AI Modeling

Virach SORNLERLTLAMVANICH^{a,b,1}, Pawinee IAMTRAKUL^c,
Teerayuth HORANONT^d, Narit HNOOHOM^e, Konlakorn WONGPATIKASEREE^e,
Sumeth YUENYONG^e, Jantima ANGKAPANICHKIT^f,
Suthasinee PIYAPASUNTRA^f, Prittipoen LOPKERD^g, Santirak PRASERTSUK^g,
Chawee BUSAYARAT^g, I-soon RAUNGRATANAAMPORN^h,
Somrudee DEEPAISARN^d, and Thatsanee CHAROENPORN^a

^aAsia AI Institute (AAIL), Faculty of Data Science, Musashino University, Japan.

^bFaculty of Engineering, Thammasat University, Thailand.

^cCenter of Excellence in Urban Mobility Research and Innovation, Faculty of Architecture and Planning, Thammasat University, Thailand.

^dSirindhorn International Institute of Technology, Thammasat University, Thailand.

^eFaculty of Engineering, Mahidol University, Thailand.

^fFaculty of Liberal Arts, Thammasat University, Thailand.

^gFaculty of Architecture and Planning, Thammasat University, Thailand.

^hSchool of Transportation Engineering, Institute of Engineering, Suranaree University of Technology, Thailand.

Abstract. This research proposes an AI platform for data sharing across multiple domains. Since the data in the smart city concept are domain-specific processed, the existing smart city architecture is suffered from cross-domain data interpretation. To go beyond the digital transformation efforts in smart city development, the AI city is created on the architecture of cross-domain data connectivity and transform learning in the machine learning paradigm. In this research, the health and human behavioral data are targeted on human traceability and contactless technologies. To measure the inhabitant's quality of life (QoL), the primary emotion expression study is conducted to interpret the emotional states and the mental health of people in the urbanized city. The results of information augmentation draw attention to the immersive visualization of the Thammasat model.

1. Introduction

Since the term smart city was unveiled, many efforts have been made by the escalation of the advancement of information and communication technology (ICT) and big data. Until now, it is still hard to find a formal definition of the smart city. However, there seems to have a sharing ultimate goal of improving the inhabitant's quality of life (QoL)

¹ Corresponding Author, Virach Sornlertlamvanich, Asia AI Institute (AAIL), Faculty of Data Science, Musashino University, 3- 3-3 Ariake Koto-ku, Tokyo, 135-8181, Japan; and Faculty of Engineering, Thammasat University, 99 Moo 18, Paholyothin Road, Klong Nueng, Klong Luang, Pathumthani 12120, Thailand; E-mail: virach@musashino-u.ac.jp

by means of the efficiency and sustainability of urban operations with respect to economic, social, and environmental aspects [1]. A broad view of a smart city is elaborated in the combination of smart infrastructure, smart building, smart transportation, smart energy, smart healthcare, smart technology, smart governance, and smart citizen. The advancement in digital technology is another factor that enables the smartness of devices interfacing to the applied domains. On another research track, the artificial intelligence (AI) technology significantly gains its potential in the real-life applications in the recent years. There are many good examples such as a machine can interact with people by speaking the same language, diagnose disease by learning the patient symptoms or x-ray images, recognize objects or faces from images, sense the emotion in the composed music, etc. The next challenge of the smart city which realizes the efficiency in city operation becomes insight by the achievement of artificial intelligence technology. Data connectivity for modeling and prediction are the key challenge in making a city-scale AI system. Some good examples can be seen in AI City challenge organized by Smart World and NVIDIA². Multiple cameras for vehicle tracking and anomaly detection are one of the interests in intelligent traffic systems (ITS) challenge [2]. AI City gets into view when Oliver Wyman Forum gets started extensive global research on 105 cities to better understand the potential disruption brought by AI³. The goal of the survey is to move beyond admiring things like “smart cities” and start a data-informed conversation about how to address the very real opportunities and challenges of AI disruption. It is reported that most cities do not address major societal changes driven by AI and other technologies. They mainly focus on smart city operation efficiency and sustainability developments.

Avoiding a hard definition of AI city, we put a roadmap on the AI in the context of smart city to incorporate the functions of human intelligence in recognition and decision making on the monitoring data. City infrastructure is now digitized and fully connected in the environment of high speed communication as the production cost of environment sensors and network devices continues to drop, the ability to use reliable mobile telecommunications and cloud computing is bringing the concept of the Internet of Things (or IoT) to life. This is the place where artificial intelligence and machine learning come into play to maximize data’s value. Machine learning can process the enormous data volumes streaming from the built systems, creating automated, real-time reactions where appropriate and delivering manageable analytics for artificial intelligence systems to interpret. In conjunction with the data streaming from any possible sources to a platform and the analytic results from machine learning, the data-driven artificial intelligence is well-suited to form the analytical foundation of the here-called AI city.

Thammasat AI City initiative has an aim to establish a resilient AI platform for the inhabitants to find out the opportunities and challenges of AI disruption. Rangsit campus is in where Thammasat University is located, surrounded by research and higher education facilities, industrial, business and agricultural areas. To confront with the AI disruption, Rangsit campus is geared to be a role model of AI City for the fully activation of the use of data and physical availability. The Thammasat AI City focuses on the four domains of elderly and healthcare, mobility, agriculture, and the environment under the awareness of societal change after the COVID-19 pandemic technology trends, namely

² AI City Challenge, <https://www.aicitychallenge.org>, <http://smart-city-sjsu.net/AICityChallenge/>

³ Global Cities AI Readiness Index by Oliver Wyman Forum, <https://www.oliverwymanforum.com/city-readiness/global-cities-ai-readiness-index-2019/index-summary.html>

distributed city, human traceability, new reality, home-office integration, contactless technology, digital lending, and frugal innovation⁴.

The remainder of the paper is organized as follows. Section 2 discusses the issues in urbanization. Section 3 explains the architecture and design of the AI ready city initiatives. Section 4 elaborates on the problems and approaches in health monitoring and elderly care. Section 5 lists the methodologies applied in the research for human behavioral detection. Lastly, Section 6 concludes the aggregation of analyzed data for visualizing in the Thammasat immersive model.

2. Issues in Urbanization

The concentration of urban development both in terms of the infrastructure transportation system and employment thus causes immigration from rural areas to urban areas in order to find opportunities to improve their quality of life. Many problems occur in the city density of daily activity and traveling if there is no well urban planning. The majority of the population is unable to bear the cost of living in the city area. This is one of the reasons why urban sprawl has occurred, which can be observed in the country center city like Bangkok and its vicinities with a population growth rate of 0.5 percent. The urban areas deteriorated during the period of 1987s before reviving due to the development of mass transit systems in the 1997s under the higher efficiency of connecting the urban areas. The better transportation system allows the people to get access to more places in the city. The urban area is thus revitalized again while the urban area is continuously expanding to cover the area outside the city. The Bangkok urban area grew from 1,900 square kilometers to 2,100 between 2000 and 2010, making it the fifth-largest urban area in East Asia in 2010, larger than megacities such as Jakarta, Manila, and Seoul [3].

Thammasat University, Rangsit Campus, is located in Pathum Thani Province, a Bangkok vicinity city in the north. It covers an area of 1,526 square kilometer with 985,643 in population, in 2020 report by National Statistical Office of Thailand⁵. It is renowned for an area of research and education, economy, and agriculture. It serves an infrastructure for ten renown universities, Thailand Science Park, seven mega economic areas of shopping malls and agricultural markets, where its agricultural areas⁶ are 35.11 percent of the city.



No exception, Pathum Thani Province is facing its high-rising population situation. It comes along with urbanization problems i.e. insufficient elderly care facilities, environmental deterioration, traffic congestion that deteriorates the total inhabitants' quality of life. The problems and challenges to bring about new opportunities are tabulated in Table 1 [4].

⁴ After Corona, Nikkei BP, Xtech, July 2020

⁵ Pathum Thani population report, <http://service.nso.go.th/nso/nsopublish/districtList/S010107/th/5.htm>

⁶ Pathum Thani area information, <https://www.opsmoac.go.th/pathumthani-dwl-files-421691791843>

Table 1. Problems and challenges of Pathum Thani Province urbanization⁷

Problems	Trends of Situation			Challenges
	2021	2026	2036	
<p>1. Traffic congestion and road safety problems</p> <ul style="list-style-type: none"> - Traffic jams - Traffic accident problems - Transportation management of industrial estates and Talat Thai commercial districts - Travel alternatives 	+46%	+93%	+186%	<ul style="list-style-type: none"> - Never solved traffic problems - Increasing of accidents - Creation of multiple alternative traveling systems
 <p>Traffic accident tendency</p>				
<p>2. Environmental problems</p> <ul style="list-style-type: none"> - Garbage problems (Pathum Thani province is the 25th dirtiest province in the country) - Air and noise pollution - Water quality problems 	+7%	+9%	+20%	<ul style="list-style-type: none"> - Construction of a wastewater treatment system of the province - Campaigning to maintain water quality and industrial wastewater treatment - Creating an efficient solid waste management system
<p>Trends in the amount of waste in the future (The trend of water quality problems is unpredictable, with wastewater from Industrial, community, and agriculture sectors, some are not treated before being discharged into water reservoirs.)</p>				
<p>3. Economic and tourism problems</p> <ul style="list-style-type: none"> - Upgrading of local wisdom to the national market - Growth of a subsidiary company 	<ul style="list-style-type: none"> - The trend of the increase of small companies in Pathum Thani Province has an average increase of 0.14% per year (2007-2016), with an increase of more than 1,000 companies every year. - The most registered businesses are wholesale, retail, automotive repair, and motorcycles, followed by industrial production. 			<ul style="list-style-type: none"> - Adding value and productivity of the agricultural sector - Developing the capabilities of the target industries - Service sector development - Raising the standards and potential of local entrepreneurs
<p>4. Lifestyle problems</p> <ul style="list-style-type: none"> - The population is increasing. - The number of latent populations from work has increased. - The security from crime. - The sufficiency of the infrastructure in the future is due to the increasing population. 	+7%	+11%	+17%	<ul style="list-style-type: none"> - The development of specialized medicine - Surveillance of the epidemic and research studies on emerging disease prevention methods - Increasing medical resources to meet local needs - Social Immunity - Social development and basic security of the people
<p>Population trends from the civil registration in the future</p>  <p>The latent population trend from the past latent population.</p> <ul style="list-style-type: none"> - Crime cases in the past year (2016) have an increasing trend. 				

⁷ Source : Pathum Thani Plan 2018 - 2022, The Pathum Thani Provincial Office, 2018

3. AI Ready City Initiatives

Thammasat University has launched its initiative in enabling a city-scale AI in the project of AI Ready City Networking in RUN. The project is to transform the area of 2.8112 square kilometers of Thammasat University in Rangsit campus for modeling the AI capacity on a city scale since the current research in AI is heavily suffered from the insufficiency and diversity of data. Reliability and connectivity of the data will be collected and made available to fully demonstrate the capability of AI on the real-life campus. It is designed to function as a based platform [5] for the four most high-impact domains in the Rangsit city, i.e., healthcare, environment, mobility, and agriculture by being equipped with AI-enabled healthcare monitoring devices [6], noninvasive bed sensors [7], environmental sensors, video analytics cameras, street lights, indoor tracking devices [8], and drones for aerial photography. Figure 1 depicts the project architecture with its domain-specific connectivity.

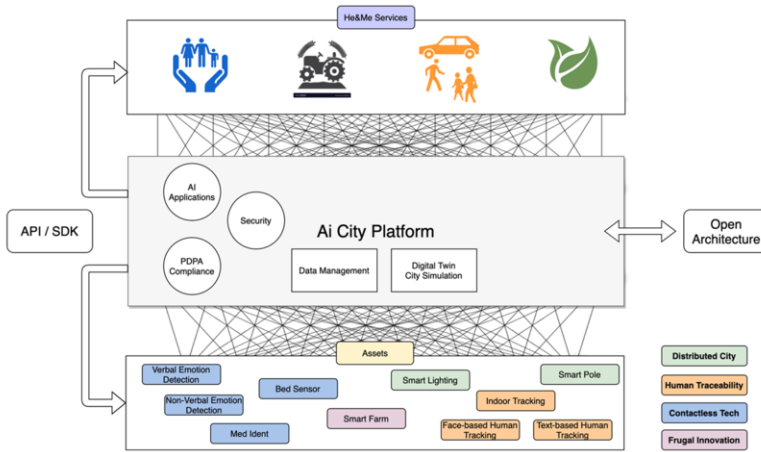


Figure 1. AI Ready City architecture and domain specific connectivity

To make the data from various sources available for modeling, the data are stored and forwarded via low energy mesh network (6LoWPAN protocol in case of smart lighting, Bluetooth low energy (BLE) in case of indoor positioning and bed sensor, etc.) to the cloud services. To reduce the high bandwidth consumption devices such as video streaming of surveillance cameras, LAN connectivity, and several techniques⁸ (steady state at rest, motion detection, etc.) are adopted. In bed sensors for elderly care systems, the detection of types of on-bed position is localized not only to realize the real-time warning but also to conserve the bandwidth by sending the compressed results to the cloud.

The AI City Project is realized by establishing the four main layers composed of accumulating, knowledging, understanding, and decision-making layers, as illustrated in Figure 2. Accumulated data from any kind of Internet of Things (IoT) device is analyzed and connected to yield the models and prediction results in four targeted domains. The lowest layer is the layer of accumulating physical raw data through sensor network devices to provide sufficient labeled data in the knowledging layer. Model training for

⁸ <https://info.verkada.com/surveillance-features/bandwidth/>

specific tasks is conducted in the understanding layer. The appropriate machine learning paradigms are introduced and evaluated to produce the results in the decision-making layer. The connectivity and selection of the data from various sources are crucial to implementing in the city-scale AI platform development.

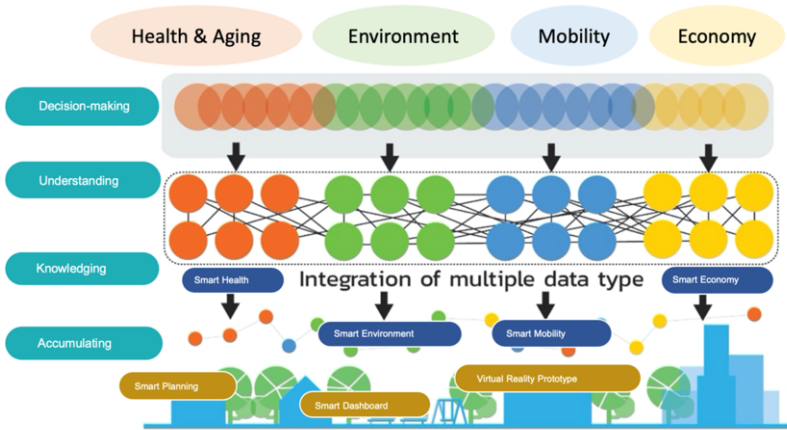


Figure 2. Deep intelligent IoT in fully connected network

4. Health Monitoring and Elderly Care

The Internet of Things and wearable technology have made life easier. With these small electronic sensors, placed on the human body, the body temperature, blood pressure, blood oxygen, respiratory rate, etc. can be continuously measured. However, there are many issues in designing the real-time health data monitoring devices to be comfortable for wearing and harmonious with daily activities. Complexity in maintaining the wearing, especially for the elderly, hinders the continuous use of the devices. The contactless technology is primarily considered to reduce the barrier to use.

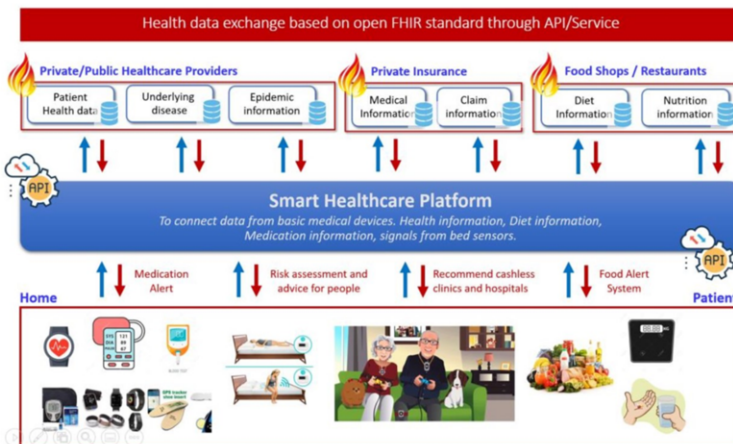


Figure 3. Smart Health Platform architecture

To connect health data from basic medical devices, the Smart Health Platform is introduced as overviewed in Figure 3. Fast Healthcare Interoperability Resources (FHIR) standard is adopted to describe the data formats and elements (resources) and the API for exchanging electronic health records (HER) [9]. The standard was created by the Health Level Seven International (HL7) healthcare standards organization.

Emphasizing on the contactless technology, the project related healthcare applications have been implemented to conduct on the Smart Health Platform such as bed sensors for elderly care (BedSense) and medicine blister package identification (MedIdent).

4.1. Bed Sensor for Elderly Care (BedSense)

BedSense is created for monitoring the on-bed position of the elderly for bed fall prevention. The system includes an on-bed position prediction system, a real-time monitoring system, and a notification system via LINE application (a widely used mobile messenger application especially in Thailand).

Falling from a bed frequently occurs when the elderly attempt to get out of the bed or come close to the edge of the bed. The mishaps have a high possibility of serious injuries, such as bruises, soreness, and bone fractures. Moreover, lacking repositioning of the body of a bedridden elderly may cause serious bedsores. To avoid such a risk, a continuous activity monitoring system is needed for taking care of the elderly.

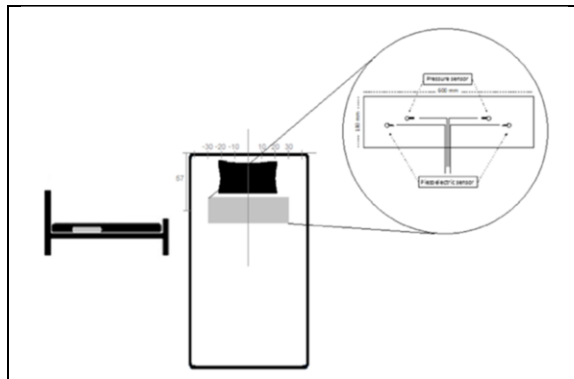


Figure 4. Sensor panel positioned under the mattress in the thoracic area

BedSense performs the bed position classification based on the sensor signals collected from only four sensors embedded in a panel. The set of sensors is composed of two piezoelectric sensors and two pressure sensors. The panel is installed under the mattress on the bed, as depicted in Figure 4.

The bed positions are classified into five different classes, i.e., off-bed, sitting, lying center, lying left, and lying right, as shown in Figure 5. To collect the training dataset, three elderly samples are asked for consent to participate in the experiment. In our approach, a neural network combined with a Bayesian network is adopted to classify the bed positions and put a constraint on the possible sequences of the bed positions. The results from both the neural network and Bayesian network are combined by the weighted arithmetic mean. The experimental results have a maximum accuracy of position classification of 97.06% when the proportion of coefficients for the neural network and the Bayesian network is 0.3 and 0.7, respectively [7, 10].

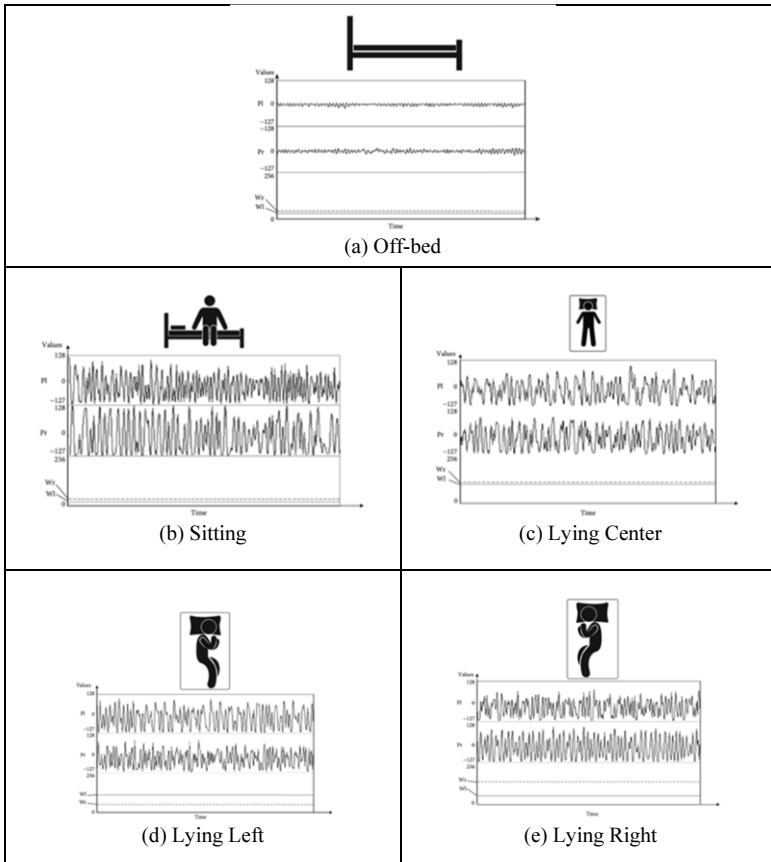


Figure 5. Correlation between signals and positions

4.2. Medicine Blister Package Identification (MedIdent)

Drug dispensing statistics from Rajavithi Hospital reveal that the drug dispensing process contains errors at 3.8 percent [11]. To sustain their health condition, the elderly also need a medicine reminder when the number of prescription drugs to take per day has a trend of getting higher. Almost 90% of older adults regularly take at least 1 prescription drug, almost 80% regularly take at least 2 prescription drugs, and 36% regularly take at least 5 different prescription drugs [12]. Medicine Blister Package Identification (MedIdent) application is created to ensure the drug dispensing process in the hospital and assist the elderly in medicine reminding. The accuracy of the image classification model is improved by using a double-side transformed image dataset with download from Highlighted Deep Learning (HDL) work [13]. The dataset which is composed of two-hundred seventy-two images for types of medicine blister packs, including 72 images of the front side and back side merged with a horizontal cropped background, is used for training the model. The blister package image dataset uses a deep learning model with a ResNet-101 pre-trained model from the TensorFlow framework. The experimental results indicated that the TensorFlow framework achieved higher precision, recall, and F1-score than the Caffe framework. A ResNet-101 model with histogram equalization in

the front and back sides has yielded the highest accuracy of 100 percent [14], as shown in Figure 6.

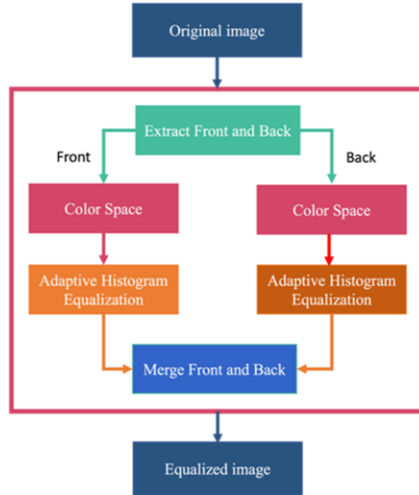


Figure 6. Pre-process blister package image

5. Human Behavioral Detection

To ensure the safety of urban residents while still maintaining their privacy, in the AI Ready City project, we, therefore, introduce the concept of human traceability using the principles of digital image processing and natural language processing to find suspicious persons and objects. Some necessary works are conducted on emotional expression detection to trace their intention, and multiple vehicles tracking to capture the time-to-time traffic situation. The transfer learning approach is essentially attempted for domain adaptation to avoid the vast computation time and resource consumption.

5.1. Weapon Detection Using Faster R-CNN Inception V2

Among the most common illegal behaviors, including robbery, quarrels, and the carrying of weapons in public, carrying a weapon in public is the most dangerous criminals. The situations are frequently drawn to the case of losses according to the criminal public records. We develop a model to detect common weapons such as firearms and knives which are commonly found in criminal cases bringing about big losses.

To train a model for weapon detection, the datasets used in this research are collected from two public datasets: ARMAS Weapon detection dataset and IMFDB Weapon detection system. TensorFlow Object Detection API is used to detect the target objects by using 1) SSD MobileNet-V1, 2) EfficientDet-D0, and 3) Faster R-CNN Inception Resnet-V2. The experimental results show that the object detection model trained by the Faster R-CNN Inception V2 using ARMAS Weapon detection dataset yields the highest mAP of 0.540 with the Average Precision of 0.5 IoU and 0.75 IoU at 0.793 and 0.627, respectively [15].

However, the MobileNet-V1 provided higher detection precision than in EfficientDet-D0 and Faster R-CNN Inception Resnet-V2 for detecting gun images.

Figure 7 exhibits that the EfficientDet D0 is unable to detect the labels on gun images, while Faster R-CNN Inception Resnet-V2 shows false positive detection on non-pistol objects.

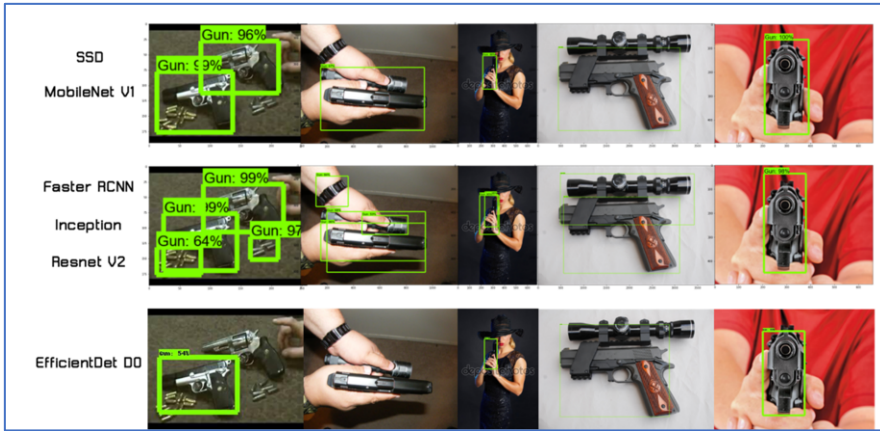


Figure 7. Weapon detection model on example images

5.2. End-to-End Speech Emotion Recognition for Low Resource Languages

Emotion balance of the residents is critical when the city is urbanized at a high pace along with the rapid availability of the social activity invasion technology. Individual activities are easily tracked by the public devices not only by the visibly aware cameras but also the interactive communication via various types of social media applications. Regularly detection of the resident emotion to measure their QoL is used to assist in the evaluation of the current state of urbanization.

As depicted in Figure 8, via the general conversation, the end-to-end speech emotion recognition is proposed to capture the speaker's emotion from their responding speech. We introduce raw speech preparation to chop speech into small chunks which are consistent in real-time and then normalize raw speech chunks before feeding them to model learning. Voice activity detection (VAD) [16] is chosen for filtering only speech frames, and then resampling speech to 16kHz of the sampling rate. Various data augmentation techniques using VTLP (vocal tract length perturbation) [17, 18] are used in this work to make the model learn more perspectives by adding simulated vocal speaker information. The speech signal is chopped into small chunks based on one-second duration and then normalizing features from chopped speech.

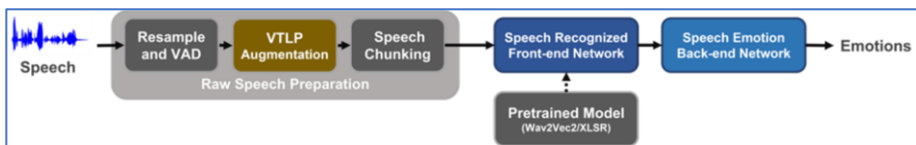


Figure 8. End-to-end speech emotion recognition in the pipeline for low-resource languages

In the front-end network, we adopt Wav2Vec2.0 [19] and XLSR [20] released by Facebook in 2020 as the pre-trained models in the transform learning process to realize speech emotion recognition for the low resource languages such as Thai.

Our back-end network is based on CNN and multilayer perceptron (MLP) pattern for mapping feature spaces to the speech emotion domain.

We evaluate our proposed models with two publicly available datasets, i.e. Berlin emotional database (Emo-DB) [21] containing seven acted emotional states: anger, disgust, boredom, joy, sadness, neutral, and fear, and VISTEC ThaiSER [22] containing five emotional states: anger, sadness, neutral, frustration, and happiness. The experiment results show that the finetuned Wav2Vec2.0 yields the highest weighted accuracy of 91.18% on Emo-DB, and the finetuned XLSR yields the highest weighted accuracy of 71.27% on ThaiSER.

5.3. Verbal and Non-verbal Corpus Construction

One of the problems affecting people's QoL in the digital age is mental health. No exception in the societies in Thailand, it has a chain effect on the individual, family, and social level. One of the most common mental health problems is depression. It is reported in [23] that adolescent depression patients (15-19 years) have a 6.24% higher risk of suicide, and 6.70% are more common in the central and eastern regions, where females have a higher risk than males. There is a trend of the increase in the number of adolescent depression patients.

Jantima Angkapanichkit et al. [24] indicate that language is a clue to spot the signs of depression in a person as well as to distinguish between people with depression and non-depression. However, expressions of emotion in the Thai language have not been grouped and classified systematically with clear criteria due to the difficulties in text analysis and interpretation. In addition, the expression of emotion in non-verbal or paralinguistic, as a context of communication, has not been yet conducted in the Thai language.

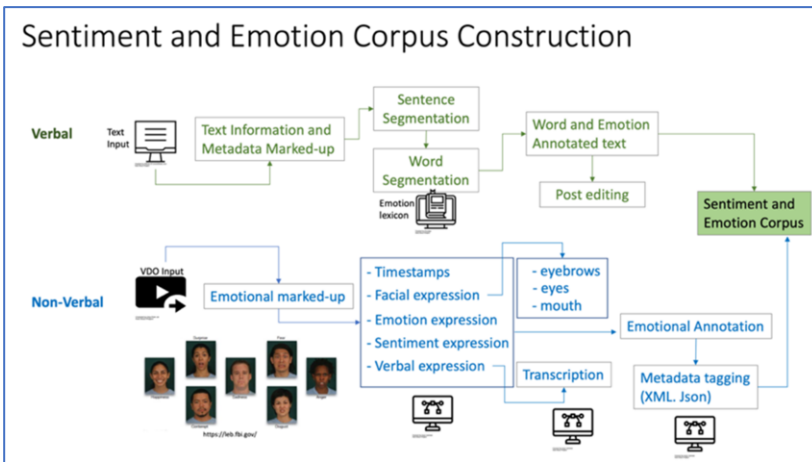


Figure 9. Process of sentiment and emotion corpus construction

The objective of the study is to create a repository of verbal emotions and sensory languages in Thai and non-verbal languages in order to interpret the emotional states and the mental health of people in the urbanized city. The verbal aspect is included in the linguistic information in word, phrase, sentence, and dialog levels, which is relating to emotion and depressive disorder information. The overview of the process of sentiment

and emotion corpus construction is shown in Figure 9. Verbal information are labeled in text and non-verbal information are labeled in video which are aligned by the video timestamps to produce the final cut of the corpus.

Ekman coined the word emotion as “a process, a particular kind of automatic appraisal influenced by our evolutionary and personal past, in which we sense that something important to our welfare is occurring, and a set of psychological changes and emotional behaviors begins to deal with the situation” [25]. Based on the studies in [26-29], we aggregate the concern of cultural distinction and classify the expression into six basic emotions, i.e., happy, angry, sad, relax, stress, and neutral (fear and disgust are specified as a negative-activate feature in stress; surprise is specified as a positive-activate in happy), as shown in Figure 10.



Figure 10. Collection of six basic emotions

5.4. Person Image Search by Natural Language Description

Closed Circuit Television or CCTV is the most widely used in security camera system. It has a main function of recording events for later viewing, and also being used to monitor the on-going happenings in order to strengthen the stability and security of the target area. In the current CCTV systems, however, it is time-consuming to search for any specific scenes i.e. perpetrators or suspicious occurrences. To facilitate human traceability in AI City, person image search by natural language description is proposed to work on the huge amount of recorded videos. The task is defined to solve the social problems such as finding missing person, child abduction, suspicious person by means of the description of person appearance. We create a human dataset of full-body images such as the images captured from the recorded closed circuit camera. The images are described by the person appearance in the Thai language, such as a girl wearing a grey T-shirt, short pants, shouldering a black bag and wearing sneakers, as shown in Figure 11.

Based on GNA-RNN [30] and TIPCB [31], we trained the model by changing the encoder layer for TIPCB and the embedding layer for GNA-RNN to model of Thai text in order to analyze the Thai-translated CUHK-PEDES data set. The TIPCB model is then trained in the Thai text version. The encoder is also changed from the BERT model [32] to the WangchanBERTa model [33]. The configuration is shown in Figure 12.

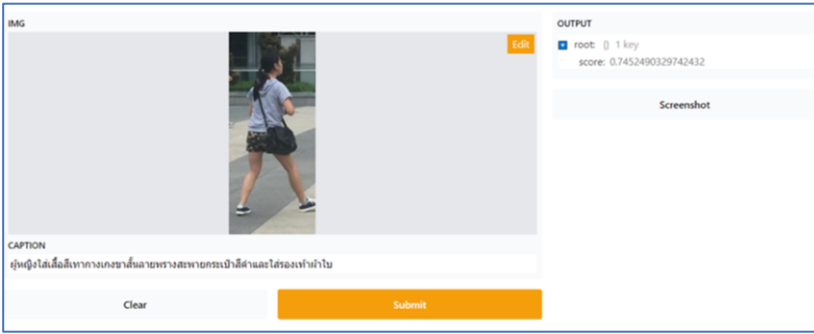


Figure 11. Example of human data collection with Thai description

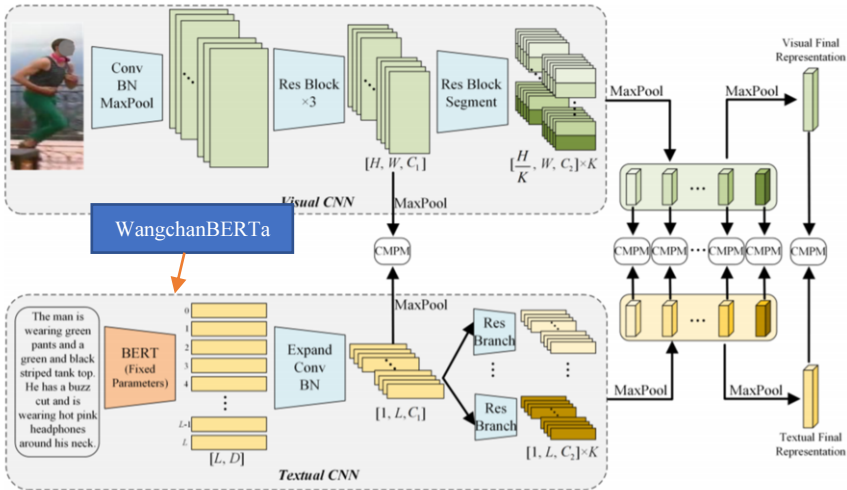


Figure 12. Configuration of text and image modeling to retrieve the target image by the image and the description of the appearance

The result of the trained model can reach the recall counted within the top five results is 0.74. The trained model is then used to calculate the similarity between the text input and the images. In the experimental results, it is found that the description text corresponding to the person's appearance yields a higher score than the non-conforming description text. Therefore, this score can be used to retrieve the persons having the same appearance described in the text.

5.5. Transfer Learning-based Vehicle Classification

Traffic monitoring and management is a modern feat that allows authorities to achieve resilience in road safety, controlled commute, and assessment of road conditions [34]. Vehicle detection on video sequences obtained from a network of surveillance cameras allows automation in the traffic management system. Computer vision and machine/deep learning have become the key technical advancement over the past years in the domain of vehicle detection and multi-object tracking (MOT). The deep learning (DL) methods have achieved the state-of-the-art in detecting and classifying vehicles on video streams [35]. These methods when coupled with fast-tracking algorithms can provide real-time

multiple vehicle tracking, depending upon the accuracy and speed of detection. However, the DL models require to be trained on a large training dataset and still can fail to produce an accurate multi-object detection and classification. In this section, we discuss real-time vehicle classification using a transfer learning-improved DL model.

Many of the studies that use DL models are pre-trained on a large training dataset such as COCO and KITTI [36], which include classes such as car, bike, truck, and bus. In our study, we classify the vehicles into seven classes i.e. car, bus, taxi, bike, pickup, truck, and trailer. As some of the classes in the existing datasets include multiple of our desired classes (e.g. pickup and SUV are classified as trucks in the COCO dataset), we are unable to use these datasets and pre-trained models in our method. The problem is often called a domain-shift problem. It occurs because the machine learning methods including DL assume that the training and test dataset is produced in the same or similar environment. To address this problem, we use the method of transfer learning [37, 38] to improve the performance of the YOLO networks that we train on a custom large training dataset and further efficiently reduce the training time.

Classes	Average Precision (AP)				Classes	Average Precision (AP)			
	YOLOv3	YOLOv3t	YOLOv5l	YOLOv5s		YOLOv3	YOLOv3t	YOLOv5l	YOLOv5s
car	0.314	0.338	0.326	0.369	car	0.764	0.755	0.722	0.745
bus	0.22	0.172	0.158	0.16	bus	0.861	0.761	0.762	0.750
taxi	0.538	0.468	0.536	0.503	taxi	0.660	0.687	0.733	0.673
bike	0.338	0.258	0.358	0.304	bike	0.792	0.724	0.817	0.784
pickup	0.296	0.268	0.279	0.284	pickup	0.757	0.719	0.686	0.711
truck	0.255	0.178	0.199	0.071	truck	0.688	0.584	0.654	0.628
trailer	0.129	0.105	0.041	0.023	trailer	0.447	0.457	0.499	0.492

Figure 13. The result without transfer learning (left) and after transfer learning (right) with test dataset

Camera	No. of Frames	Vehicle Face	Class	GT	YOLOv3				YOLOv3t				YOLOv5l				YOLOv5s			
					P	R	OA	P	R	OA	P	R	OA	P	R	OA				
Cam 1	7500	Back	car	217	0.99	0.95	0.97	0.96	0.91	0.93	0.99	0.97	0.98	0.98	0.91	0.95				
			bus	5	0.56	1.00	0.78	0.80	0.80	1.00	0.80	1.00	0.80	0.90	0.80	0.80				
			taxi	6	0.86	1.00	0.93	0.86	1.00	0.93	0.75	1.00	0.88	0.86	1.00	0.93				
			bike	20	1.00	1.00	1.00	1.00	1.00	0.95	1.00	0.98	1.00	0.90	0.90	0.95				
			pickup	99	0.98	0.86	0.92	0.92	0.86	0.89	0.98	0.96	0.97	0.88	0.84	0.86				
			truck	4	0.33	0.50	0.42	0.43	0.75	0.59	0.60	0.75	0.68	0.43	0.75	0.59				
			trailer	1	0.50	1.00	0.75	0.50	1.00	0.75	1.00	1.00	1.00	0.00	0.00	0.00				
Total	352	0.74	0.90	0.82	0.78	0.90	0.84	0.90	0.93	0.91	0.71	0.74	0.72							
Cam 2	9060	Side	car	16	1.00	0.88	0.94	1.00	0.94	0.97	1.00	0.94	0.97	0.93	0.88	0.90				
			bike	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
			pickup	12	0.85	0.92	0.88	0.92	0.92	0.92	1.00	0.96	0.83	0.83	0.83					
			truck	3	0.75	1.00	0.88	0.75	1.00	0.88	1.00	1.00	1.00	0.43	1.00	0.71				
			trailer	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.25	0.63				
			Total	36	0.92	0.96	0.94	0.93	0.97	0.95	0.98	0.99	0.99	0.84	0.79	0.82				
			Cam 3	8800	Side	car	87	0.96	0.93	0.95	0.99	0.86	0.92	0.98	0.94	0.96	0.99	0.83	0.91	
taxi	6	1.00				1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00					
bike	12	1.00				0.83	0.92	1.00	0.83	0.92	1.00	0.83	0.92	1.00	0.83	0.92				
pickup	91	0.93				0.93	0.93	0.87	0.97	0.92	0.96	0.98	0.97	0.86	1.00	0.93				
truck	12	1.00				0.58	0.79	0.86	0.50	0.68	0.92	1.00	0.96	0.82	0.75	0.78				
trailer	3	0.43				1.00	0.71	0.43	1.00	0.71	1.00	0.67	0.83	0.67	0.33	0.50				
Total	211	0.89				0.88	0.88	0.86	0.86	0.86	0.98	0.90	0.94	0.89	0.79	0.84				
Cam 4	9575	Front	car	149	0.96	0.89	0.92	0.88	0.92	0.90	0.95	0.97	0.96	0.92	0.95	0.94				
			taxi	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
			bike	25	1.00	0.96	0.98	1.00	0.96	0.98	1.00	1.00	1.00	0.96	1.00	0.98				
			pickup	73	0.81	0.92	0.86	0.82	0.75	0.79	0.94	0.90	0.92	0.91	0.84	0.87				
			truck	5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
			Total	256	0.95	0.95	0.95	0.94	0.93	0.93	0.98	0.98	0.98	0.96	0.96	0.96				
			Grand Total		855	0.88	0.92	0.90	0.88	0.92	0.90	0.96	0.95	0.95	0.85	0.82	0.83			

Figure 14. Model performance evaluated on validation video stream collected from test area using transfer learning-improved networks (P=Precision; R=Recall; OA=Overall accuracy)

To overcome the problem of domain-shift, we train new set of networks with new dataset by using the weights transferred from the previously trained models. These models are then used to detect and classify the vehicles in our multi-vehicle tracking algorithm, which is essentially a centroid tracking algorithm that tracks each class of vehicles on each individual lane polygons provided to the system. Figure 13 shows the improvement of average precision by applying transfer learning models in all versions

of YOLO evaluation. Figure 14 shows the accuracy of classification by applying transfer learning models to distinguish sides of vehicles.

6. Thammasat Immersive Model for Collective Data Visualization

AI City data are visualized on the Thammasat model generated from (1) direct inspection photography, (2) aerial photogrammetric survey by drone, and (3) laser scanner. The data are processed to determine the point cloud for generating a 3D mesh for reference as shown in Figure 15.

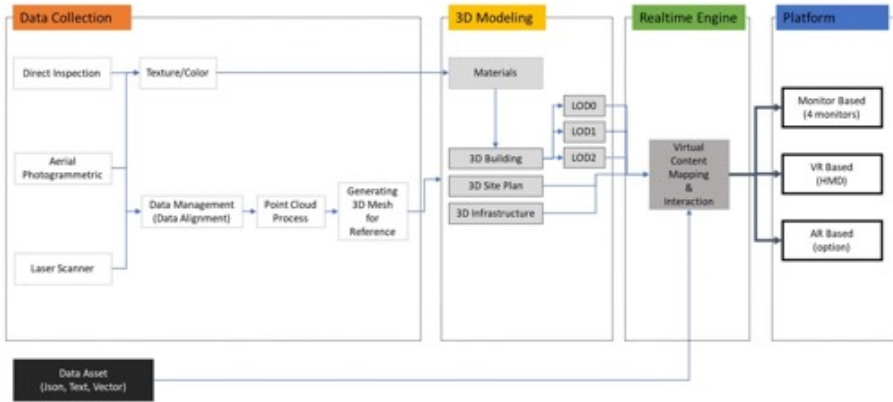
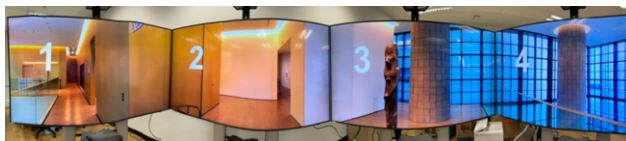


Figure 15. Modeling process of Thammasat Rangsit campus.

Figure 16 shows the visualization results on four surrounding screens to provide a room-like experience. It is a node of Data Sensorium proposed by AAIL, Musashino University to exhibit the collective data visualization across the network. After the image stitching process, Figure 16 (a) shows the synchronized image controlled by the multiple points via the Internet. On each end, the users can control the view and point out the area of interest. The function allows the users to share their concerns in the same room-like experience. In addition, the models are smoothly augmented with the location-sensitive information to realize the spatial experience for the users as shown in Figures 16 (b). The example information of social media density is expressed in a form of an augmented graph in Figure 16 (c) to show the SNS population at a specific moment.



(a)



(b)

(c)

Figure 16. Modeling process of Thammasat Rangsit campus.

As a result, Data Sensorium shows its potential in realizing the spatial immersive environment to relieve the limitation of using HMD, especially in the case of urban planning that needs a city-scale environment sharing in the concept of AI City.

7. Conclusion

The limitation of digital transformation by ICT and big data analysis solely in facilitating the city's operational excellence has been discussed. In many smart city implementations, the activities are upheld and visualized to grasp the city situation. Without a comprehensive analysis and data sharing across the domains, it is not sufficient to predict the trend and prevent the undesired occurrence due to the growth of urbanization, especially in the vicinity of big cities. Thammasat AI city enabling initiatives in healthcare and behavioral detection by contactless and human traceability technology are explored across the domain sensory devices on the standardized AI city platform. With the harmonization of health data and sleep monitoring BedSense provide uninterrupted healthcare between home and hospital. Individual and public behavioral detection, aggregated by the result of emotion expression evaluation, are able to foster the inhabitant QoL measure to sustain the city's evolution. Lastly, the pre-trained models used in transform learning approaches can be effectively utilized when the data are insufficient in some tasks.

Acknowledgement

This work was supported by the Thailand Science Research and Innovation Fundamental Fund, Contract Number TUFF19/2564 and TUFF24/2565, for the project of "AI Ready City Networking in RUN", based on the RUN Digital Cluster collaboration scheme.

References

- [1] Mohanty SP, Choppali U, and Kougianos E. Everything You wanted to Know about Smart Cities. *EEE Consum. Electron. Mag.* 2016; 5(3): 60-70. <https://doi.org/10.1109/mce.2016.2556879>.
- [2] Tang Z, Naphade M, Liu MY, Yang X, Birchfield S, Wang S, Kumar R, Anastasiu D, Hwang JN. CityFlow: A City-Scale Benchmark for Multi-Target Multi-Camera Vehicle Tracking and Re-Identification; arXiv; 2019; arXiv:1903.09254.
- [3] The World Bank. Urbanization in Thailand is dominated by the Bangkok urban area. Feature story; 2015 Jan 26. <https://www.worldbank.org/en/news/feature/2015/01/26/urbanization-in-thailand-is-dominated-by-the-bangkok-urban-area>
- [4] Klaylee J, Iamtrakul P, Kesorn P. Driving Factors of Smart City Development in Thailand. In: Proceedings of International Conference and Utility Exhibition on Energy, Environment and Climate Change (ICUE); 2020, p. 1-9. doi: 10.1109/ICUE49301.2020.9307052
- [5] Ota N. Create Deep Intelligence TM in the Internet of Things; 2014. URL <http://on-demand.gputechconf.com/gtc/2015/presentation/S5813-Nobuyuki-Ota.pdf>
- [6] Singh KK, Singh A, Lin J-W, Elnger A. Deep Learning and IoT in Healthcare Systems. Paradigms and Applications: CRC Press; 2021 Dec.
- [7] Viriyavit W, Sornlertlamvanich V. Bed Position Classification by a Neural Network and Bayesian Network Using Noninvasive Sensors for Fall Prevention. *Journal of Sensors: Hindawi.* 2020 Jan; Volume 2020, Article ID 5689860. p. 1-14. <https://doi.org/10.1155/2020/5689860>
- [8] Kovavisaruch L, Sanpechuda T, Chinda K, Kamolvej P, Sornlertlamvanich V. Museum Layout Evaluation based on Visitor Statistical History. *Asian Journal of Applied Sciences.* 2017 Jun;5(3). p. 615-622.

- [9] Welcome to FHIR. HL7.org.; 2019 Nov 1; Retrieved 2021-02-12. <http://www.hl7.org/index.cfm>
- [10] Pongthanisor G, Viriyavit W, Prakayapan T, Deepaisarn S, Sornlertlamvanich V. ECS: Elderly Care System for Fall and Bedsores Prevention using Non-Constraint Sensor. In: Proceedings of International Electronics Symposium (IES); 2020 Sept 29-30; Surabaya, Indonesia; p. 60.
- [11] Kristina S, Supapophon P, Sooksriwong C. Dispensing Errors: Preventable Medication Errors by Pharmacists In Outpatient Department, A University Hospital, Bangkok, Thailand. [online] Aasic.org. Available at: <http://aasic.org/proc/aasic/article/view/90> [Accessed 21 Aug 2020].
- [12] Qato DM, Wilder J, Schumm LP, et al. Changes in prescription and over-the-counter medication and dietary supplement use among older adults in the United States, 2005 vs 2011. *JAMA Intern Med*; 2016; 176(4):473-82. doi: 10.1001/jamainternmed.2015.8581
- [13] Wang JS, Ambikaphathi A, Han Y, Chung S, Ting H, Chen C. Highlighted Deep Learning based Identification of Pharmaceutical Blister Packages. In: IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA); 2018; p. 638-645. doi: 10.1109/ETFA.2018.8502488
- [14] Hnoohom N, Maitrichit N, Chotivatunyu P, Sornlertlamvanich V, Mekruksavanich S, Jitpattanukul A. Blister Package Classification Using ResNet-101 for Identification of Medication. In: The 25th International Computer Science and Engineering Conference (ICSEC2021); 2021 Nov 18-20; Chiang Rai, Thailand.
- [15] Hnoohom N, Chotivatunyu P, Maitrichit N, Sornlertlamvanich V, Mekruksavanich S, Jitpattanukul A. Weapon Detection Using Faster R-CNN Inception-V2 for a CCTV Surveillance System, In: The 25th International Computer Science and Engineering Conference (ICSEC2021); 2021 Nov 18-20; Chiang Rai, Thailand.
- [16] Team S. Silero VAD: Pre-Trained Enterprise-Grade Voice Activity Detector (VAD). Number Detector and Language Classifier. 2021. <https://github.com/snakers4/silero vad>
- [17] Xu M, Zhang F, Cui X, Zhang W. Speech Emotion Recognition with Multiscale Area Attention and Data Augmentation. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); IEEE; 2021. p. 6319-6323.
- [18] Kim C, Shin M, Garg A, Gowda D. Improved Vocal Tract Length Perturbation for a State-of-the-Art End-to-End Speech Recognition System. In: Proceedings of Interspeech; 2019. p. 739-743.
- [19] Baevski A, Zhou Y, Mohamed A, Auli M. Wav2Vec2.0: A Framework for Self-Supervised Learning of Speech Representations. *Advances in Neural Information Processing Systems*; 2020;33.
- [20] Conneau A, Baevski A, Collobert R, Mohamed A, Auli M. Unsupervised Cross-Lingual Representation Learning for Speech Recognition; 2020. arXiv preprint arXiv:2006.13979.
- [21] Burkhardt F, Paeschke A, Rolfes M, Sendmeier WF, Weiss B, A Database of German Emotional Speech. In: Proceedings of the Ninth European Conference on Speech Communication and Technology; 2005.
- [22] Chaksangchaichot C. VISTEC-AIS Speech Emotion Recognition. Available: <https://github.com/vistec-AI/vistec-ser>
- [23] Department of Mental Health, Phra Sri Maha Pho Hospital. Report on Access to Services for Patients with Depression 2017; 2017 Feb 10. Available at: <http://www.thaidepression.com>
- [24] Angkapanichkit J, Rochanahastin A, Intasian S. Language, Communication, and Depression: An Exploration of Communicative Practices about Depression for Sustainable Quality of Life of Thai Adolescents: Amarin Printing & Publishing; 2020 May.
- [25] Ekman P. <https://www.paulekman.com/>, Access on 2022 Jan 1.
- [26] Ekman P. Universals and Cultural Differences in Facial Expressions of Emotions. In: Cole, J, editor. Nebraska Symposium on Motivation: Lincoln, NB; University of Nebraska Press; 1972. p. 207-282.
- [27] Ekman P. and Friesen W. V. Constants Across Cultures in the Face and Emotion. *Journal of Personality and Social Psychology*; 1971;17(2). p. 124-129.
- [28] Ekman P, Friesen WV, Tomkins SS. Facial Affect Scoring Technique: A First Validity Study. *Semiotica*; 1971;3. p. 37-58.
- [29] Ekman P. Universal Facial Expressions of Emotions. *California Mental Health Research Digest*; 1970; 8(4), p. 151-158.
- [30] Shuang L, et al. Person search with natural language description, In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017.
- [31] Chen Y, et al. TIPCB: A Simple but Effective Part-based Convolutional Baseline for Text-based Person Search; 2021. arXiv preprint arXiv:2105.11628.
- [32] Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding; 2018. arXiv preprint arXiv:1810.04805.
- [33] Lowphansirikul L, Polpanumas C, Jantrakulchai N, Nutanong S. WangchanBERTa: Pretraining transformer-based Thai Language Models; 2021. arXiv preprint arXiv:2101.09635.
- [34] Radopoulou SC, Brilakis I. Improving road asset condition monitoring. *Transportation Research Procedia*; 2016;14, p. 3004-3012.

- [35] Kalake L, Wan W, Hou L. Analysis based on recent deep learning approaches applied in real-time multi-object tracking: A review. *IEEE Access*; 2021;9, p. 32 650–32 671.
- [36] Geiger A, Lenz P, Stiller C, Urtasun R. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*; 2013;32(11), p. 1231–1237.
- [37] Kouw WM, Loog M. An introduction to domain adaptation and transfer learning. *arXiv preprint*; 2018; arXiv:1812.11806.
- [38] Ngiam J, Peng D, Vasudevan V, Kornblith S, Le Q. V, Pang R. Domain adaptive transfer learning with specialist models. *arXiv preprint*; 2018; arXiv:1811.07056.