IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 6, 2020, accepted June 15, 2020, date of publication July 2, 2020, date of current version July 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3006493

Recognizing Traffic Black Spots From Street View Images Using Environment-Aware Image Processing and Neural Network

TEERAPAUN TANPRASERT[1], CHAIYAPHUM SIRIPANPORNCHANA[2], NAVAPORN SURASVADI[2], AND SUTTIPONG THAJCHAYAPONG[2], (Member, IEEE)

[1]Pomona College, Claremont, CA 91711, USA
[2]National Electronics and Computer Technology Center, National Science and Technology Development Agency, Pathumthani 12120, Thailand

Corresponding author: Suttipong Thajchayapong (suttipong.thajchayapong@nectec.or.th)

This work was supported in part by the Thai People Map and Analytics Platform (TPMAP), a joint project between the office of National Economic and Social Development Council (NESDC) and the National Electronic and Computer Technology Center (NECTEC), the National Science and Technology Development Agency (NSTDA), Thailand.

ABSTRACT This paper proposes a novel technique to identify black spots (prone-to-accident road locations) using street view images. The proposed technique is derived based on the hypothesis that the characteristics of the surroundings of the road have an effect on the safety level of a particular spot, and is the first black spot classification technique that is fully environment-aware. Assessing four street view images around each spot, a *distance-aware pixel accumulation* is developed to extract information about the objects surrounding the road from a semantically segmented image. The accumulated vectors are then used to train fully-connected neural networks to identify black spots. Performance evaluations are conducted with street view images in Thailand, which represent a challenging scenario of analyzing road characteristics in developing countries, with one of the highest road traffic fatality rates and limited historical accident records. Comparisons between our proposed technique and previously proposed techniques are also provided. Experiments show that our proposed technique succeeds in classifying black and safe spots in Thailand with an accuracy of 69.91%, where 75.86% of the black spots are identified correctly. Also, the *distance-aware pixel accumulation* can improve the accuracy of those machine learning techniques up to 6.4%. Our findings also evidently revealed that the object surrounding the roads as well as their sizes and distances are determinants of road's accident proneness.

INDEX TERMS Black spot detection, machine learning, feature extraction, pixel accumulation, street view images.

I. INTRODUCTION

A 2018 World Health Organization report has sent an alarming message that road traffic injuries have been the main cause of death of children and young adults, and the SDG 3.6 target will not likely be met [1]. According to the Third Global Status Report on Road Safety, Thailand has the second-highest road traffic fatality rate in the world. The Road Accident Victims Protection Company Limited has recorded 829,201 people injured, disabled, or killed in road accidents in 2018 or 2,271.78 victims each day. The office of National Economic and Social Development Council (NESDC), under the Thai government, has initiated Thai People Map and Analytics Platform (TPMAP) project. The project's aim is to develop a national data analytics platform to analyze problems related to the quality of life of Thai citizens including poverty,[1] income, disabilities, as well as road accidents. One of the pain points in developing the platform is the lack of a mechanism to identify spots that are prone to accidents and, subsequently, the fragmentation of accident-related data. This paper presents an automated and transferable approach to identify those accident-prone spots.

A black spot, in the field of road safety studies, is generically defined as a dangerous, prone-to-accident spot on the

The associate editor coordinating the review of this manuscript and approving it for publication was Sungroh Yoon.

[1] Fully accessible online at https://www.tpmap.in.th

road. In practice, multiple more specific variations of a black spot's definition are adopted by different organizations. The study in [2] reviews these definitions and studies their applicability to streets in Thailand. In this paper, we follow the precise definition accepted as a national standard by public organizations in Thailand: a black spot is a spot at which at least three serious accidents or five injury accidents have been recorded within 100 meters in three years.

Most of the existing black spot classification approaches rely on manually collected data or past traffic records as parameters to correlate with accident-proneness and features for the black spots prediction. In such data-driven approaches, commonly used parameters include the number of lanes, the traffic volume, road conditions, as well as the number and types of intersections or curves [3]–[6]. In this paper, we explore an alternative data source, a visual instead of a numerical or categorical data. Although they come with noises, visual data are much easier to obtain and richer with information. Street view images, specifically, capture not only the geometric characteristics of the streets, but also the information on the surroundings from the perspective of a moving car. We hypothesize that the surroundings of the road, such as the number of the trees and the position of the buildings, can influence the road's safety level. To investigate this hypothesis further, we select street view images as the data for our prediction model. Surprisingly, while street view images have been used for a variety of life quality-related classification problems (see [7]–[13]), no work has been done on using purely street view images to predict road safety, to the best of our knowledge.

The final step of the proposed technique is to learn to predict. In our initial attempts, we adopted a convolutional neural network (CNN), for it is known as one of the best performing standard learning techniques for recognizing patterns in the images. However, it did not perform nearly as well as expected (see IV-B for results). Hence, instead, we developed a feature extraction pipeline utilizing a CNN-based semantic segmentation model, coupled with a pixel accumulation algorithm that compresses and extracts information about the objects surrounding the road from raw images. The extracted features are then fed to a fully-connected neural network to learn to identify black spots. Our model succeeds in classifying black and safe spots in Thailand with an accuracy of 69.91%, correctly classifying over two-thirds of the unseen spots based solely on street view images. Furthermore, 75.86% of the black spots in the data set are identified correctly. Its performance, in turn, supports our hypothesis that the surrounding environment correlates to the safeness of a road.

The contributions of this paper can be summarized into three main components as follows.

1) *Data:* We propose street view images as a novel data source for the black spot prediction task. Street view images are extremely rich with information, especially for our task, as they contain both traditional information (e.g. the number of the lanes and the condition of the road) and additional backdrop information (e.g. the density of trees and buildings, as well as the types and quantity of vehicles), which are not captured in other common data sources such as manually recorded road conditions or even satellite images. Furthermore, only four snaps of images are easy to access and retrieve, making our technique relatively accessible to all types of users.
2) *Application:* We have developed and tested a full-pipeline, neural-network-based black spot classification method that uses only street view images. Results show that it outperforms several baseline techniques, including CNN.
3) *Algorithm:* We also introduce a feature extraction technique called *distance-aware pixel accumulation* as a component in the pipeline. The accumulation algorithm compresses 2-dimensional, semantically segmented images into a vector while preserving both the relative magnitude and certain relevant structural information.

This paper is organized as follows. Section II gives an overview of existing works whose topic or methodology partially overlaps with those of ours. Section III comprehensively describes and justifies the data set and the mechanism of our technique, covering three essential image processing steps and the neural network structure. Section IV presents the experimental results of our technique's performance in comparison to that of other candidate techniques, along with a detailed description of how each technique was implemented and tested. Section V discusses the results from Section IV in terms of their implications and applications, and Section VI summarizes our work and breaks down its contributions.

II. RELATED WORKS

Black Spots Analysis is a branch of road safety analysis that has been widely studied in the past few decades. The studies in [3] and [4] produced a comprehensive list of the traditional, widely accepted methods used for black spots analysis. Various statistical models and different aspects of accident data, namely the accident severity and accident involvement, were used in the reviewed works. Sixteen risk factors were listed, and different types of risk factors were said to have correlations with different aspects of an accident. In the past decade, there have been multiple works adopting a similar framework, including the statistical analysis of the relationship between geometric parameters such as road accesses and curve lengths presented in [6]. Many works on accident risk prediction, including [14]–[16], have also been done specifically on the streets and highways in Thailand. The presentation in [17] gives a comprehensive description of the state-of-art process currently followed to handle road hazardous locations. The number of accidents, the number of casualties, and the accident rate are used as a preliminary means to calculate the risk. Other more labor-intensive methods, such as making stick and collision diagrams and collecting additional data based on human observations, have been implemented as well. While this procedure is detailed and highly accurate, it requires enormous financial, labor, and time resources.

In recent years, new technological discoveries have allowed the development of more advanced road safety analysis techniques. The study in [18] combined the Geographic Information System (GIS) with manual labeling of the street's attributes, such as road conditions and the number of lanes, to find their correlations with accidents and to identify black spots. Factors that had a positive correlation with high accident risks were: a lower road level, a smaller traffic volume, and a smaller intersection spacing. Deep learning has also been utilized to predict both short- and long-term accident risk in more recent works [19], [20]. Unlike more traditional methods of road safety analysis that focus on accident records only, these recent works also showed that other parameters including traffic flow, weather, and air quality help improve the prediction.

While visual data, especially street view images, have rarely ever been used directly for road safety or black spot analysis, they have appeared in numerous works involving other kinds of life-quality-related predictions. For instance, [7] predicted accident risks based on a Google Street View image of a house by first using human labor to classify the condition of the house and the neighborhood based on the image, and then finding the correlations between these factors and the risk of that house's residents getting into a car accident. Another risk prediction paper [8] also used humans to classify attributes in street view images before associating them with pedestrian injuries statistically. Other works that used machine learning techniques include [10]–[13], whose goals were to predict demographic makeup, street walkability, perceived safety (crime-wise) of the captured neighborhood, and urban land for urban planning and management, respectively. All four works used a large data set of street view images to train neural networks of different complexity to predict. The work in [9], which proposed an automated system for identifying cracks on roads, also used street view images as the input. Moreover, a segmentation to differentiate the road from the background, a technique similar to semantic segmentation used in our work, was applied onto the images before they were fed to the support vector machine for training. Even though these examples utilized both street view imagery and machine learning tools, none of them were directly concerned with predicting the safeness of a road.

In addition to street view images, the satellite image is another possible source of visual data for the task, and it was used in [21], the only existing work that used purely visual data to predict road safety. It used deep learning to associate a road safety map with satellite images obtained from hundreds of thousands of accident reports in New York City. Downloaded images were directly fed to a standard convolutional neural network architecture without any major data preparation process, and the result of the prediction was a road map with three levels of safety. The trained model got 78% accuracy within the same city and 73% with a different city. Unfortunately, due to the lack of complete and accessible traffic accident records in Thailand, it is impossible to obtain such complete map of past accidents or safety levels. Hence, our task has to be framed slightly differently: to distinguish critically dangerous spots (black spots) from acceptably safe spots. In terms of the input data, satellite images are more intuitive for detecting road structures than street view images. However, our work explores the hypothesis that there are details in the surroundings, such as how much the trees and billboards cover the view of the drivers, that significantly influence the safety level at a particular spot. Hence, street view images, which capture exactly what the drivers see, are a suitable – and unprecedented – option of input data for road safety prediction.

III. MATERIALS AND METHODS

A. DATA SET

1) GEOGRAPHICAL COORDINATES

Prior to retrieving the images of streets, we need to have the exact coordinates of safe and black spots in Thailand. All coordinates used were retrieved from publicly accessible data compiled and published online by Road Accident Victims Protection Company Limited. The company was established in 1998 under the Motor Vehicle Victims Act B.E. 2540 to facilitate compensation to victims of road accidents in areas not covered by insurance companies. With 59 insurance companies contributing as shareholders, it has the most complete records of accident reports available in Thailand.

For black spots, we retrieved 3,461 coordinates (in latitude/longitude format) of officially declared black spots from Black Spots System available online on the official website of Road Accident Victims Protection Company Limited.[2] For safe spots, 72,873 coordinates of places with reported accidents between 2011-2019 were retrieved from the same source; only locations where only 1 accident had occurred within 100 meters radius were considered safe spots to ensure that the inaccuracy of pinning the accident's location would not interfere with the classification. The data points for both safe and black spots are distributed over the entire country, as illustrated in Figure 1.

Due to the inaccuracy of the location pinning of both black and safe spots, a small fraction of coordinates does not capture a spot on a road. Furthermore, at the time of our investigation, Google Street View did not have a complete coverage in Thailand, i.e. street view images could not be retrieved for a small fraction of coordinates. After eliminating coordinates under these two cases, we balanced the data set with a simple, uniform under-sampling. Two thousand coordinates per class (4,000 coordinates in total) were randomly selected to represent black and safe spots. While a smaller data size saves a considerable amount of both computational resources and time, it is recommended that the complete data set (balanced), if exists, is used to improve the accuracy.

[2] The company's full website is accessible at http://www.thairsc.com/.

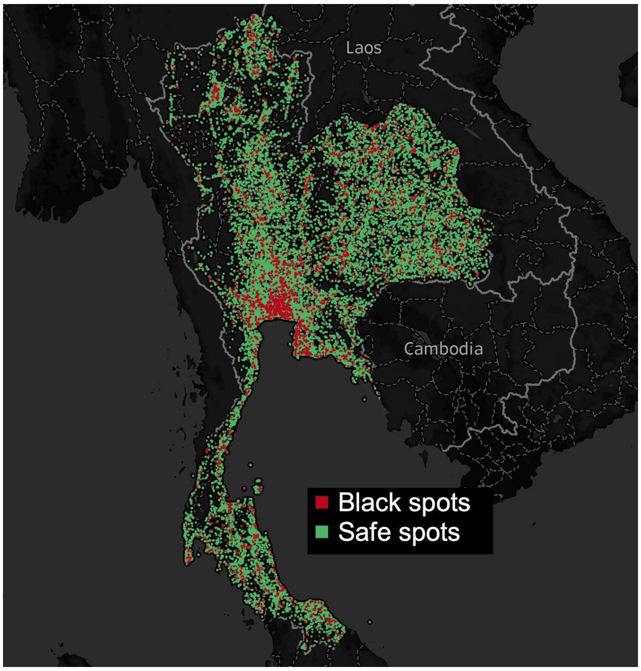

FIGURE 1. The coordinates of road accidents in Thailand from 2011 to 2019.

2) GOOGLE STREET VIEW IMAGES

For each of the 4,000 coordinates retrieved, four images with heading angles of 0, 90, 180 and 270 degrees were requested through Google Street View Static API. The images have dimensions 640 by 480 pixels and the horizontal field of view of 90 degrees. (The dimension is adjustable since one of the image pre-processing steps resize all images to a pre-specified size.) Hence, the data set consists of 16,000 images from safe and black spots (8,000 of black spots and 8,000 of safe spots) distributed impartially over all regions of Thailand.

B. IMAGE PRE-PROCESSING

In order to extract only relevant information, the images need to be pre-processed as shown in Figure 2. For each spot, the four 640 × 480-pixel images are converted into 28-component vectors, which are the inputs for training the neural network model to predict the safety of the spot. The pre-processing consists of three main steps: 1) semantic segmentation, 2) distance-aware pixel accumulation, and 3) filtering.

1) SEMANTIC SEGMENTATION

As shown in Figure 2, the first step is semantic segmentation (I). Individual pixels are categorized into classes of common objects around the street. Figures 4 and 3 show a sample raw Google Street View image and the corresponding semantic segmentation of the image respectively. We used a standard U-NET learner model [22] from the FastAI library and the CamVid data set, a data set of labeled images of streets, to train a semantic segmentation model with 92% accuracy. We then used the model to segment each image into 32 meaningful categories. The 32 classes, as employed in the labeling of the CamVid data set, are animal, archway, bicyclist, bridge, building, car, cart/luggage/pram, child, column/pole, fence, driving lane marks, non-driving lane marks, miscellaneous text, motorcycle/scooter, other moving objects, parking block, pedestrian, road, road shoulder, sidewalk, sign symbol, sky, SUV/pickup truck, traffic cone, traffic light, train, tree, truck/bus, tunnel, miscellaneous vegetation, void, and wall.

The model compresses any image down to the dimension of 480 by 360 pixels, and returns its segmentation of the image as a 2-dimensional array with 480 columns and 360 rows, each position storing the encoding of the object type at that pixel (e.g. 1 represents an animal and 17 represents a road).

2) DISTANCE-AWARE PIXEL ACCUMULATION

The next step is the novel feature extraction algorithm: distance-aware pixel accumulation (II), shown in Figure 2. This step compresses the 480×360 array into a 32-component vector by accumulating the number of pixels assigned to each object class based on its pixel distance from the closest road pixel. This technique is derived from the hypothesis that objects closer to the road supposedly have higher influences on the activities on the road, including accidents. Therefore, in addition to the size of the object (represented in the number of pixels), the distance from the road should also be an influential factor. The fundamental concept of this accumulation algorithm is that closer and larger objects are assigned more weight. Note that the accumulation algorithm does not attempt to capture the exact real-world magnitude, but rather to capture the relative distance with respect to the road objects in each image. The algorithm first generates a 480 × 360 mask of the reversed (negated) distance of every pixel from its closest road pixel.

$$mask_{i,j} = dist_{MAX} - distance((i,j),(x_i,y_i)) \tag{1}$$

where:

$mask_{i,j}$ = the reversed distance from (i, j) to its nearest road
$dist_{MAX}$ = the maximum distance possible (480 in our implementation)
(x_i, y_i) = position of the road pixel that is closest to (i, j)

It then iterates through the original array and the mask simultaneously, and accumulates the value in the mask to its object class's component in the resultant vector. Therefore, after the accumulation, each vector component, representing an object class, contains the sum of that object's reversed distance from the closest road previously stored in the mask matrix.

$$V_k = \sum_{i,j} mask_{i,j} \times \mathbb{1}_{array_{i,j}=k} \tag{2}$$

where:

$array_{i,j}$ = value of the segmentation encoding at pixel (i, j)

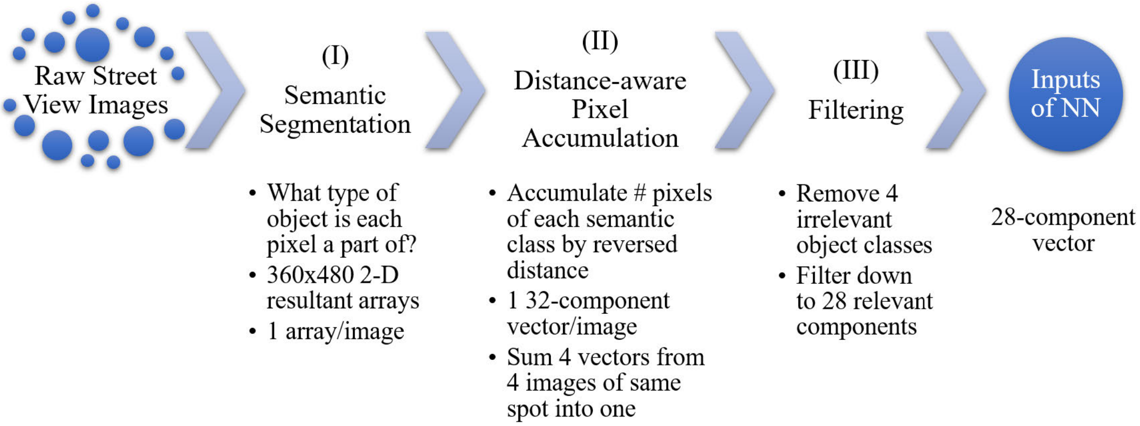
Raw Street View Images
(I)
Semantic Segmentation
(II)
Distance-aware Pixel Accumulation
(III)
Filtering
Inputs of NN
• What type of object is each pixel a part of?
• 360x480 2-D resultant arrays
• 1 array/image
• Accumulate # pixels of each semantic class by reversed distance
• 1 32-component vector/image
• Sum 4 vectors from 4 images of same spot into one
• Remove 4 irrelevant object classes
• Filter down to 28 relevant components
28-component vector

FIGURE 2. Flow chart of the full image pre-processing pipeline.

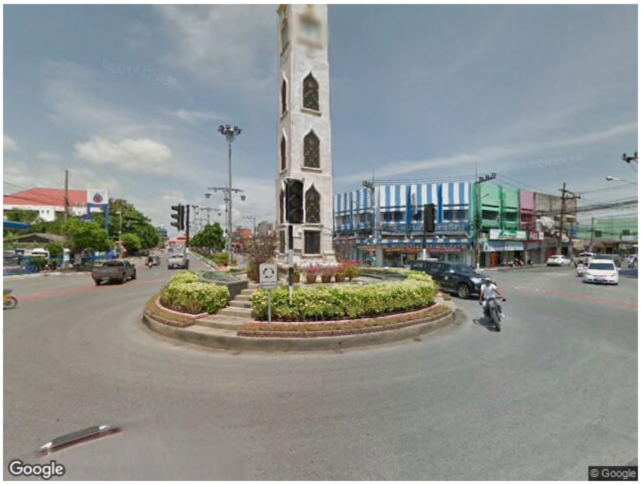

FIGURE 3. Sample raw Google Street View image from coordinate (6.8651, 101.2414).

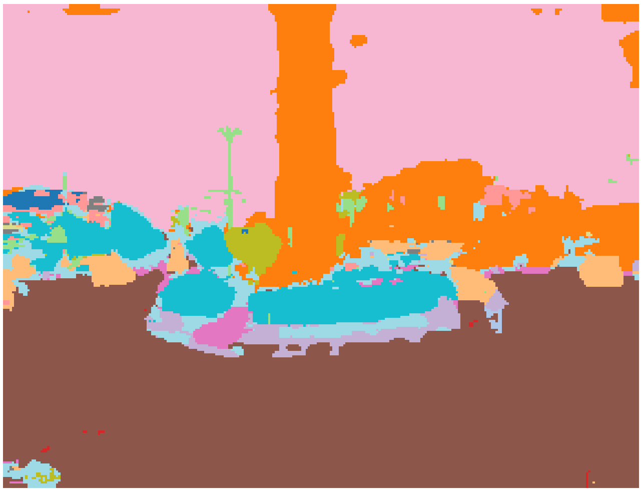

FIGURE 4. Color map corresponding to semantic segmentation of the image in Figure 3. Each color is mapped to a segmentation class (e.g. brown represents the "road" class).

$$\begin{aligned} mask_{i,j} &= \text{the reversed distance from (i, j) to its nearest road}\\ V &= \text{resultant 32-component vector}\\ \mathbb{1}_f &= \begin{cases}1, & f \text{ evaluates to true}\\ 0, & \text{otherwise}\end{cases}\end{aligned}$$

Finally, the four vectors generated from images of different angles of the same point are summed into a single vector. Hence, the product of this step is a 32-component vector *per location* that keeps a considerable amount of essential data in the segmented images.

3) FILTERING

The final pre-processing step is filtering (III). There are four object classes which, among the 16,000 images in our data set, are not present at all. Therefore, they cannot assist the learning model in differentiating the vectors into classes; and they are henceforth removed from the input vector. With the data set of Thailand roads, the four classes are bridge, child, train, and tunnel. Note that this filtering step depends on the data and should be adjusted if applied to different types of landscape or different semantic segmentation models. All other components are preserved as they have the potential to be relevant. Static classes capture the features of the environment, while the non-static classes, such as pedestrian and car, represent the typical traffic situation of the road (e.g. types of vehicles and level of congestion). Therefore, in our case, by the end of the pre-processing, the data is in the form of 28-component vectors and ready to be fed into the classifier model for training and testing.

C. BLACK SPOTS CLASSIFICATION USING FULLY-CONNECTED NEURAL NETWORK

Finally, we need a machine learning model to learn from the prepared input vectors how to differentiate safe and black spots. To select an appropriate model, several commonly known methods with various levels of complexity – namely, linear regression, logistic regression, support vector machine, and fully-connected neural network – are tested. The fully-connected neural network outperforms other standard machine learning methods, as will be shown in Section IV-B.

A fully-connected neural network model is a structure-agnostic, general-purpose type of neural network that consists of fully connected layers, each one representing a nonlinear function with a specified number of parameters. For our purpose, a simple, sequential model with 10 Dense layers and 2 Dropout layers to prevent over-fitting, with a total of 36,593 parameters has proven to be sufficient. However, the fact that simpler machine learning models (i.e. regressions and support vector machine) perform worse than a neural network model on the same data set shows that the model needs a certain level of non-determinism to recognize the complex relationships between each vector component. The particular neural network architecture described earlier was chosen due to its stably decent performance in experiments. Other architectures with similar structure – fully-connected layers and a few dropouts – also have similar performances.

TABLE 1. Summary of the data used in the experiments.

	Data Available (spots)	Full Data (spots / images)	Evaluation Data (spots/images)	
			Training Set	Testing Set
Safe spots	2,000[‡]	2,000 / 8,000	1,500 (random) / 6,000	500 (the rest) / 2,000
Black spots	3,461	2,000 / 8,000	1,500 (random) / 6,000	500 (the rest) / 2,000

[‡]The 2,000 safe spots were randomly sampled from the subset of the full dataset (72,873 spots) that met the two criteria described in Section III-A.

IV. RESULTS

A. EXPERIMENTAL SETUP

For all of the experiments, we split the 8,000 images from each class into 75% training set and 25% testing set, as shown in Table 1. Through the experiments, we noticed that this split between training and testing data points is flexible, as long as there are enough points for the neural network model to learn and the number of points for safe and black spots are equal. To ensure that the results are not biased by a specific pair of training-testing data sets, we adopted the repeated random test-train splits validation strategy with 10 training and testing sessions. Three quarters of the data were randomly selected to train the model, and the rest were used to calculate the accuracy. This process was used in every experiment, including the baseline experiments. The experiments were coded and run on Google Colaboratory in Python 3 using its GPU as a hardware accelerator. The experiments were divided into two parts in order to clearly illustrate the effect of different parameters. More details of the experimental setup of each part are described in the next section, along with the performance evaluation results.

B. PERFORMANCE EVALUATION

1) BASELINE TECHNIQUES AND SEMANTIC SEGMENTATION

The first part comprises the performance tests of different input formats with several standard machine learning techniques. Six techniques were tested in the experiments comparing the performances of using raw images, segmented images, and simple pixel-accumulated vectors after semantic segmentation. Simple pixel-accumulation is a simpler version of distance-aware pixel accumulation described in Section III, where each 480 × 360 array is compressed into a 32-component vector by accumulating the number of pixels assigned to each object class using equal weight for each pixel regardless of its distance from road pixels. Where applicable, we also compared the performances of standard machine learning techniques compiled from techniques used in related works: [11], [12], [19], [23]. The four selected techniques are linear regression, logistic regression, support vector machine (SVM), and appropriate neural network models.

Linear regression, logistic regression, and support vector machine were implemented using the Scikit-learn library, while all neural networks were implemented using Keras, with Tensorflow as a backend. For the convolutional neural network, we tested with a standard pre-trained model, InceptionV3, with a new fully-connected output layer, activated with softmax function. The model was compiled with SDG as an optimizer with a learning rate of 0.0001. Several newly built convolutional neural network models with approximately 4-5 million total parameters were also tested in the preliminary study, but their best accuracy was either worse than or comparable to that of the model modified from InceptionV3.

TABLE 2. Performance comparison of baseline machine learning techniques with raw, segmented and segmented-and-simple-accumulated data.

Input Type	Machine Learning Technique	Test Acc. (%)
Raw image	CNN (InceptionV3)	50.00
Segmented image	CNN (InceptionV3)	50.08
Simple pixel-accumulated	Linear regression	62.00
	Logistic regression	61.20
	SVM	61.40
	Fully-connected neural network	63.50

Results show that feeding raw images as inputs to a convolutional neural network is not suitable, as the 50% accuracy equals that of a blind guessing for a binary classification. Segmented images do not produce any significant improvement. On the other hand, the accuracy goes above 60% for all machine learning techniques, showing that accumulating pixels of objects from segmented images extracts some amount of relevant information. Among the four techniques, the results are fairly similar, ranging between 61.2-63.5%, with the fully-connected neural network having the best performance by a small margin.

The distinction between the performance of 2-dimensional (image) versus 1-dimensional (pixel-accumulated) data is substantial. Even though the deep convolutional neural network can detect very complex patterns, it fails to learn from both raw and segmented images. We speculate that a 2-dimensional array, with 172,800 (360 × 480) components, contains too much information and potentially too much noise for the convolutional neural network to recognize any useful patterns. On the other hand, the straightforward pixel accumulation that compresses the size of input from 172,800 down to 32 improves the accuracy by about 11-14%, implying that the accumulation makes the input digestible for the neural network while preserving at least some useful information. This indicates that semantic segmentation followed by simple pixel accumulation, which represents how much space each type of object takes up in the image, can extract valuable information from the street view images. However, the accuracy can be improved even further. A more advanced variation of the accumulation technique that boosts the performance up to approximately 70% is explained and tested in the second part of the experiment.

2) SIMPLE PIXEL ACCUMULATION VS. DISTANCE-AWARE PIXEL ACCUMULATION

The second part emphasizes on the performance comparison between the two methods of pixel accumulation: the simple accumulation used in the first part and the distance-aware accumulation described in Section III. Four machine learning techniques from the previous part are used for each type of accumulation; CNN is excluded because even though it is suitable for feature extraction from an image, it is too complex for a simple 28-component vector. The neural network model used (abbreviated as NN in the Figure 2) is the fully-connected neural network described in Section III-C. All machine learning models were implemented using the same library and architecture as those in the first part, except for one small difference. For the neural network model, the loss function[‡] that performs best with the simple accumulation is the binary cross-entropy, but in our preliminary experiment, mean squared error performs better than binary cross-entropy by approximately 2%. Hence, the loss function most suited for each accumulation technique is used. Also based on preliminary parameter-tuning experiments, the number of training epochs was fixed at 100, where training accuracy starts to flatten. Note that, for reproduction purposes, the number of epochs may need to be adjusted based on the characteristics of the data set. The results of the simple accumulation, which are included in the first part, are shown again in the table to facilitate the comparison. Additionally, the separate average accuracy for each class, i.e. black and safe spots, is also presented in the table for both accumulation techniques.

The experiment shows that distance-aware accumulation is superior to simple accumulation for three out of four machine learning techniques tested. The support vector machine is the only technique that proves to be incompatible with the distance-aware accumulated input and thus unsuitable for our task. For the other three techniques, the improvements in accuracy are fairly similar: 5.42% with linear regression, 6.3% with logistic regression, and 6.4% with the neural network. Among them, the neural network continues to outperform regression by about 2.5%. Furthermore, across the ten training and testing sessions, the minimum accuracy that the distance-aware accumulation and neural network achieved was 68.75% and the maximum was 72%. Even in its worst round, the accuracy was only 1.16% below the average, and the model still outperforms all other techniques. The narrow range of accuracies also illustrates that the model has a stably satisfactory performance, which shows potentials for reproducibility.

TABLE 3. Performance comparison of simple and distance-aware accumulation techniques.

Learning Technique	Average Testing Accuracy (%)	
	Simple Accumulation	**Distance-aware Acc.**
Linear regression	62.00	67.42
Logistic regression	61.20	67.50
SVM	61.40	55.80
Neural network	63.50	69.91
Black spots	65.44	75.86
Safe spots	61.56	63.96

Since the neural network proves to be the best technique, we study it deeper by comparing its performance on black spots versus safe spots, presented in the last two rows of Table 3. The results show that, while the distance-aware accumulation increases the average accuracy of safe spots only by 2.4%, it increases that of black spots by over 10%, i.e. an additional 10% of the black spots are correctly identified. Despite being trained and tested with balanced data, the model is more inclined to classify a spot as black. Hence, the proposed distance-aware accumulation does not simply improve the overall accuracy, but more specifically, the accuracy of black spots detection. For real-world accident prevention purpose, danger detection is slightly more critical than no-danger detection, since it encourages precaution as opposed to ease of mind. In this sense, our technique could identify 75.86% of the black spots correctly.

According to the experimental results, distance-aware accumulation and the neural network together make the optimal combination, with the average accuracy of 69.91% overall and 75.86% for black spots. The performance exceeds the accuracy of the simplest raw image with CNN technique (50%) by almost 20%, which shows a promising outcome in transforming the data into a form that has significant correlations with the road's safeness. In particular, distance-aware accumulation proves to be capable of extracting and carrying hidden but illuminating information for the task of differentiating safe and black spots.

V. DISCUSSION

The experimental results show that the combination of semantic segmentation, distance-aware accumulation, and

[‡] The code for both loss functions used can be found at https://github.com/keras-team/keras/blob/master/keras/losses.py.

neural network is the suitable solution to the problem of classifying black spots from street view images. Semantic segmentation contributes by simplifying the data from a matrix of pixels to a matrix of object classes, so that the information is more digestible for the neural network. It also facilitates the distance-aware accumulation, which is yet another level of feature extraction as well as a form of data compression. A 480 × 360 2-dimensional array of a segmented image stores 172,800 integers, but our accumulation technique compresses that down to 32 integers carrying the information that allows the neural network to differentiate the safe and black spots: the presence of objects in the driver's view and their distance from the road. The filtering further removes irrelevant information that might interfere with the neural network's learning ability. Finally, neural network is required as the learning model because it has the capability to recognize the complex pattern in our inputs. Hence, every step of the pipelining process plays a crucial part in our model's ability to differentiate safe and black spots – a task that has yet to be accomplished before, either by a computer or a human.

The accuracy of 69.91% is an achievement for a road safety prediction purely from images, without any traditional traffic-based data. Since the accuracy of feeding the images into a highly complex and a pre-trained convolutional neural network model is 50% on average, the pre-processing steps must be contributing by extracting the relevant variables for distinguishing safe and dangerous spots. While the relationship between these variables cannot be summarized into a flat, intuitively comprehensible fact, nor do they perfectly capture the differences between safe and black spots, we have gained a useful insight about accident prevention.

The experimental results have proven our initial hypothesis that the surrounding environment affects the accident-proneness of a road. Even though only the amount of the surrounding objects' presence – not the shape of the road or the relative position of any objects – is preserved, our technique can still predict correctly 7 out of 10 times. Moreover, our auxiliary hypothesis that the size of the surrounding objects and their distance from the road are crucial factors also appears to be verifiable. A traditional belief exists that the road safety level depends on its condition and structure e.g. whether it is an intersection, the number of lanes, etc. [16], [18], [23], [24]. However, we have shown that the objects around the road, specifically within the driver's eye level, are also determinant of the road's accident-proneness. Therefore, the new insights discovered through our work can enormously reshape our understanding of what impacts the safety level of a road. In practice, objects surrounding a road should be considered a major factor in future policies regarding road safety. As such, further analyses of the results from semantic segmentation of safe and black spots as shown in Figure 6 and Figure 5 may reveal significant correlations between the detected objects and their distances from the road, and the level of accident risks at that location.

In terms of adaptability and flexibility, the proposed technique is not strictly specific to a data set. For the input, the images are assumed to be taken from a road, not, for example, inside of a house. Any set of four images from four orthogonal angles around the spot is a valid input; there are no restrictions on whether they are from a street view database, captured with a smartphone camera, or obtained through other means, as long as the resolution is adequate. We acknowledge that the camera angle may be an issue that influences the learning performance. Despite controlling all the parameters allowed in Google API, each set of four images does not in fact have precisely consistent angles relative to the road. Unfortunately, Google API does not have a control parameter that supports exact angular alignment. The images do not have a strict size specification, since the semantic segmentation step takes care of re-sizing, but the ratio of all images should be consistent and compatible with the target size of the re-sizing step in semantic segmentation. For a different type of landscape or street structure, the categories for segmentation may also

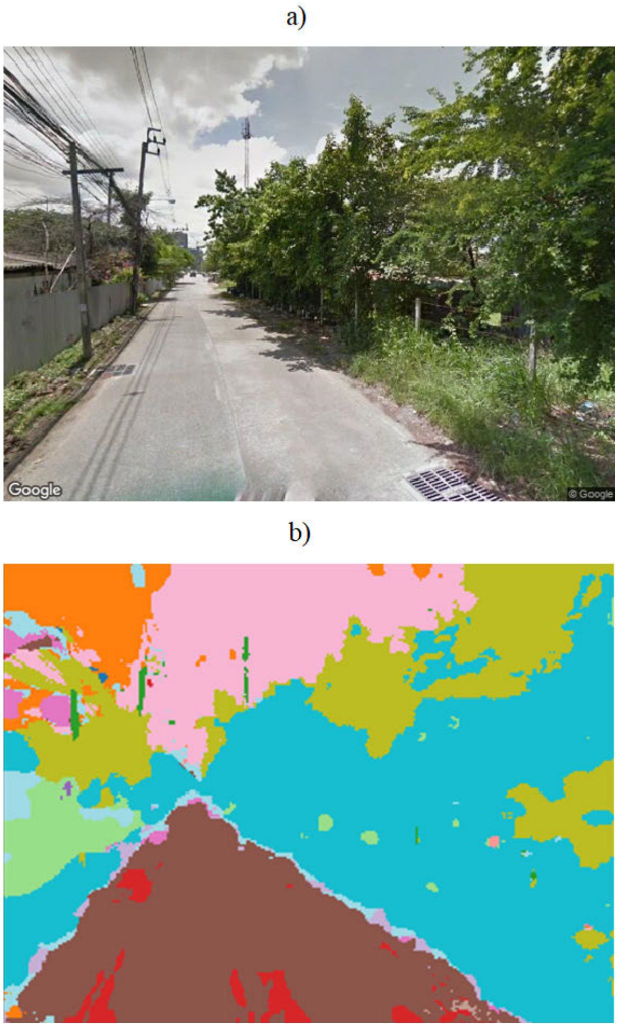

FIGURE 5. a) Sample Google Street View image and b) its corresponding semantic segmentation from a safe location: (13.8673, 100.4582). Each color is mapped to a segmentation class (e.g. brown represents the "road" class). Plantation is the dominating object. Tree and vegetation have high presence for a large proportion of the images, especially with safe locations.

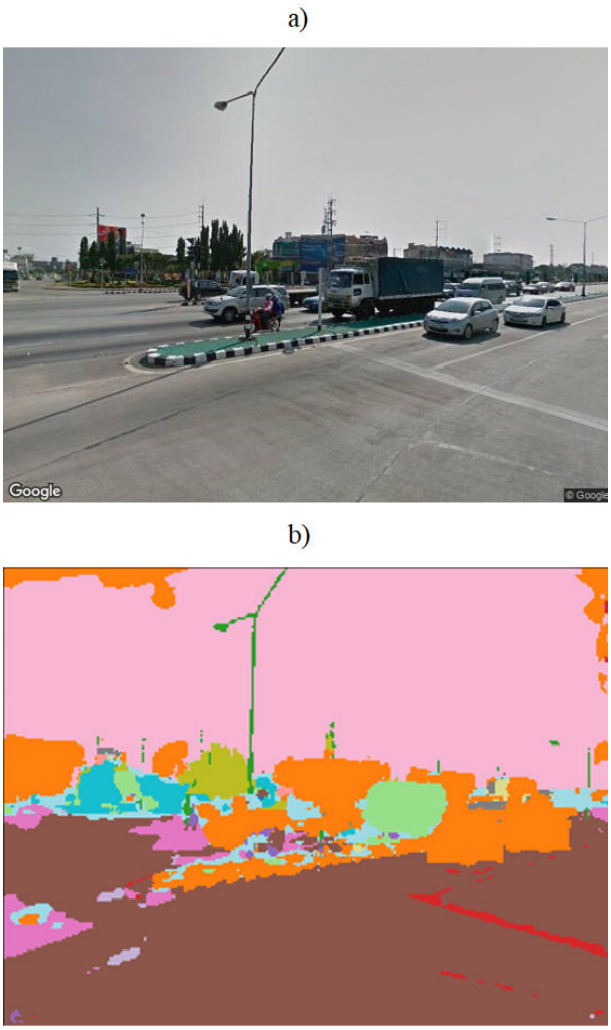

FIGURE 6. a) Sample Google Street View image and b) its corresponding semantic segmentation from a black spot: (15.189, 100.1304). Each color is mapped to a segmentation class (e.g. brown represents the "road" class). Building, vehicle and sky are dominating objects. Buildings and vehicles are closer to the road, but the sky also takes up a very large area.

be changed as appropriate. For instance, motorcycle object class may need to be removed if there are no motorcycles in the area. We anticipate that other semantic segmentation models may work as well, as long as the segmentation is accurate and object classes are sufficiently relevant to a road's accident-proneness, such as roads, cars, buildings and trees. Furthermore, the architecture of the neural network model is somewhat flexible. However, all three pre-processing steps and a fully-connected neural network must be included, for they have shown in the experiments to be the essence of our proposed technique.

VI. CONCLUSION

We have developed a novel technique for identifying black spots based on four street view images around the spot, using semantic segmentation, distance-aware pixel accumulation, and fully-connected neural network, which achieves the average accuracy of 69.91%. The full process, from obtaining the images to preparing them to training the neural network model to predict, has been thoroughly tested to consistently have the best performance. Nonetheless, it should be noted that the same pre-processing steps produce only slightly inferior performance with linear and logistic regression, which are less costly in terms of computational resources and time. The distance-aware pixel accumulation is designed to capture the characteristics of a road that are relevant to its safeness, which is the key step that boosts the performance up to almost 70% on average.

Our technique facilitates road safety evaluation without the need for traffic and accident data, on which traditional black spots analysis commonly rely. Therefore, the degree to which the evaluation is up-to-date depends solely on when the images are taken, instead of on the past traffic records which could be incomplete, out of date, or inaccessible. The prediction can also be performed before new streets are opened for use as well as prior to street construction or repair, based on a generated illustration of the street's design (including its surroundings).

It is important to note that, to the best of our knowledge, the proposed technique is the first black spot classification technique that is fully environment-aware. Experimental results suggest that the surroundings such as trees and buildings are accurate identifiers of the safety level of a road, and these factors are constantly transforming. With our model, users could conveniently obtain an updated safety assessment following changes in the environment (e.g. new billboard installed) and possibly make adjustments to create an environment suitable for safe driving.

Though further works are needed to improve the accuracy before our technique can make infallible predictions on its own, we observed the trend that the accuracy significantly improves with the size of the data set. Hence, an application of our technique in other areas where a more complete data set exists is expected to achieve better results. We also anticipate that increasing the types of input data, such as using both street view and satellite images, would help with the accuracy. However, it should be noted that for user applications, the ease of access and retrieval of street view images is a strength of our technique; adding more input requirements would compromise its simplicity. Furthermore, the insight that we have discovered regarding the effect of the surroundings on the road safety is valuable in itself. Regardless of the place and data availability, the fact that the size and the distance of objects surrounding the road correlate to the road's safety level can potentially be incorporated into the existing road safety assessment procedures right away.

REFERENCES

[1] World Health Organization. (2018). *Global Status Report on Road Safety 2018*. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/277370/WHO-NMH-NVI-18.20-eng.pdf

[2] A. Leelakajonjit and P. Iamtrakul, "Appropriated accident black spot definition for Thai police," in *Proc. 18th Nat. Conv. Civil Eng.*, 2013, pp. 538–542.

[3] K. Geurts and G. Wets, "Black spot analysis methods: Literature review," in *Proc. Steunpunt Verkeers Veiligheid Bij Stijgende Mobiliteit*, 2003, pp. 7–13.

[4] J. Pei and J. Ding, "Improvement in the quality control method to distinguish the black spots of the road," in *Proc. Eastern Asia Soc. Transp. Stud.*, vol. 5, 2005, pp. 2106–2113.

[5] P. Morency and M.-S. Cloutier, "From targeted 'black spots' to area-wide pedestrian safety," *Injury Prevention*, vol. 12, no. 6, pp. 360–364, 2006.

[6] A. Karimi and E. Kashi, "Investigating the effect of geometric parameters influencing safety promotion and accident reduction (case study: Bojnurd-Golestan national park road)," *Cogent Eng.*, vol. 5, no. 1, Sep. 2018, Art. no. 1525812.

[7] K. Kita and Ł. Kidziński, "Google street view image of a house predicts car accident risk of its resident," 2019, *arXiv:1904.05270*. [Online]. Available: http://arxiv.org/abs/1904.05270

[8] S. J. Mooney, C. J. DiMaggio, G. S. Lovasi, K. M. Neckerman, M. D. M. Bader, J. O. Teitler, D. M. Sheehan, D. W. Jack, and A. G. Rundle, "Use of Google street view to assess environmental contributions to pedestrian injury," *Amer. J. Public Health*, vol. 106, no. 3, pp. 462–469, Mar. 2016.

[9] D. Abou Chacra, "Municipal road infrastructure assessment using street view images," M.S. thesis, Dept. Syst. Des. Eng., Univ. Waterloo, Waterloo, ON, Canada, 2016.

[10] T. Gebru, J. Krause, Y. Wang, D. Chen, J. Deng, E. L. Aiden, and L. Fei-Fei, "Using deep learning and Google street view to estimate the demographic makeup of neighborhoods across the united states," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 50, pp. 13108–13113, Dec. 2017.

[11] L. Yin and Z. Wang, "Measuring visual enclosure for street walkability: Using machine learning algorithms and Google street view imagery," *Appl. Geography*, vol. 76, pp. 147–153, Nov. 2016.

[12] N. Naik, J. Philipoom, R. Raskar, and C. Hidalgo, "Streetscore-predicting the perceived safety of one million streetscapes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 779–785.

[13] R. Cao, J. Zhu, W. Tu, Q. Li, J. Cao, B. Liu, Q. Zhang, and G. Qiu, "Integrating aerial and street view images for urban land use classification," *Remote Sens.*, vol. 10, no. 10, p. 1553, Sep. 2018.

[14] N. Suesat and V. Ratanvaraha, "The development of hazardous road map in Thailand," *J. KMUTNB*, vol. 27, no. 4, pp. 605–614, 2019.

[15] S. Kassawat, S. Sarapirome, and V. Ratanavaraha, "Integration of spatial models for Web-based risk assessment of road accident," *Walailak J. Sci. Technol.*, vol. 12, no. 8, pp. 671–679, 2015.

[16] W. Tammasi and P. Joengsaguenpornsuk, "Identification of hazardous locations on highway in Thailand by the critical crash rate method," *Khon Kaen Univ. J. (Graduate Stud.)*, vol. 11, no. 3, pp. 1–8, Jul. 2011.

[17] N. Boontob. (2016). *Treatment of Hazardous Location*. [Online]. Available: http://k4ds.psu.ac.th/rsis/download/files/thaiROADS.pdf

[18] H. Chen, "Black spot determination of traffic accident locations and its spatial association characteristic analysis based on GIS," *J. Geographic Inf. Syst.*, vol. 4, no. 6, p. 608, 2012.

[19] H. Ren, Y. Song, J. Wang, Y. Hu, and J. Lei, "A deep learning approach to the prediction of short-term traffic accident risk," 2017, *arXiv:1710.09543*. [Online]. Available: http://arxiv.org/abs/1710.09543

[20] H. Ren, Y. Song, J. Wang, Y. Hu, and J. Lei, "A deep learning approach to the citywide traffic accident risk prediction," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 3346–3351.

[21] A. Najjar, S. Kaneko, and Y. Miyanaga, "Combining satellite imagery and open data to map road safety," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 1–7.

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2015, pp. 234–241.

[23] J. Kumphong, T. Satiennam, and W. Satiennam, "A study of relation between speed of vehicle and traffic accident and road characteristics," in *Proc. 20th Nat. Conv. Civil Eng.*, 2015, pp. 194–200.

[24] S. Nassar, "Integrated road accident risk model," Ph.D. dissertation, Dept. Civil Eng., Transp., Univ. Waterloo, Waterloo, ON, Canada, 1996.

TEERAPAUN TANPRASERT is currently pursuing the bachelor's degree in computer science with the Pomona College, Claremont, CA, USA. She is also a Student Research Assistant with the Pomona College. Her research interests include machine learning, computer vision, natural language processing, and human–computer interaction.

CHAIYAPHUM SIRIPANPORNCHANA received the B.E. degree in computer engineering from Khon Kaen University, Khon Kaen, Thailand, and the M.Sc. degree in information technology from the King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. He is currently a Research Assistant with the National Electronics and Computer Technology Center (NECTEC), Thailand. His research interests include data analytics, machine learning, and intelligent transportation systems.

NAVAPORN SURASVADI received the B.E. degree (Hons.) in computer engineering from Chulalongkorn University, Bangkok, Thailand, the M.Sc. degree in management science and engineering from Stanford University, CA, USA, and the Ph.D. degree in operations management from the Leonard N. Stern School of Business, New York University, NY, USA, in 2014. She is currently a Researcher with the National Electronics and Computer Technology Center (NECTEC), Thailand. Her current research interests include data analytics and data visualization especially in strategic data for government policy planning, and operations management.

SUTTIPONG THAJCHAYAPONG (Member, IEEE) received the B.S. and M.S. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, and the Ph.D. degree in electrical and electronic engineering from Imperial College London, London, U.K. He is currently a Senior Researcher with the National Electronics and Computer Technology Centre (NECTEC), National Science and Technology Development Agency (NSDTA), Pathum Thani, Thailand. He has served as the Project Manager of the Thai People Map and Analytics Platform (TPMAP), Thailand's data-driven target poverty alleviation project with the National Economic and Social Development Council. He also served in the Senate of Thailand as a Sub-Commissioner on National Strategy and Country Reform Analysis and Monitoring, the Thai Government as an Assistant Secretary of the National Big Data Steering Committee, and the Geo-Informatics and Space Technology Development Agency as a member of Actionable Intelligence Policy Working Group. His research interests include intelligent transportation systems, data analytics, anomaly detection, signal processing, and machine learning.