ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

The Collection of Road Traffic Incidents in Bangkok from Twitter Data based on Deep Learning Algorithm

Korn Puangnak[1] and Natworapol Rachsiriwatcharabul[2]

ABSTRACT

Surveillance and accident reports for current traffic management. Surveillance by monitoring the situation by using CCTV. This detecting method need to many operators to observe real time. At present communication technology and social media participate in report traffic incidents. Also, the accident was report by people who were facing the accident. Therefore, we decide to create software to investigate accidents from text reported via social media in Thai language. At present, there is no research or work platform that clearly supports this form of work. Many countries use analysis technology to detect the information from Twitter to report incident notification with verification on different languages. The purpose of this research is "to develop the deep learning technology and to solve the problem of classifying incidence patterns and identifying severity of incidents from social media in Thai's massage." For collect incident data and reporting incidents externally from a single reporting platform. Using the deep learning model MLP, CNN, Bi-LSTM, and LSTM+CNN. We can identify the twitter message as general news or traffic reporting. The traffic conditions such as traffic information, accidents, disasters, damaged roads, or other than those mentioned above. And the incidence severity level was identified as normal, lane-blocking or lane-closure. The examination demonstrated the capability of CNN+LSTM learning with the best results in incidence detection and incidence patterns at 93.44%, and CNN results in image identification. The incidence model was best at 85.29%, respectively, and the LSTM method best rated the severity of the incidence, reaching 88.53%.

Article information:

Keywords: Incident Detection, Deep Learning Algorithm

Article history:
Received: January 1, 2022
Revised: May 14, 2022
Accepted: May 21, 2022
Published: June 18, 2022 (Online)

DOI: 10.37936/ecti-cit.2022162.248535

1. INTRODUCTION

Reporting of road traffic incidences by relevant agencies at present, each responsible agency has different methods for detecting incidences including detecting the occurrence of the incident by surveillance of the traffic control room staff through CCTV. By relying on the display of random images or lap times of image showing in each camera on the highway or automatic Image Processing and decided by the processor chip on the surveillance CCTV or central computer processing by image processing software or detecting the occurrence of an incident with a sensor type detector such as Microwave Radar by taking traffic changes data to calculate the incidence forecast, etc. All of the technologies mentioned are those that require investment and the installation of a large number of related tools and equipments. As the technology of mobile devices or smartphones is developing by leaps and bounds, it plays a role in many areas. On the daily life of the people who use more roads including notifying news via social media channels on various platforms. Twitter is one of the most popular social media platforms and supports the development of programs to bring information to use, including being able to detect incidents by processing tweets. Notify the news immediately.

In the past research on solving the problem of automatic incidence detection with a large amount of data, beginning with a 2017 study by Bo-Huei Lin, Shu-Fen Tseng [1] analyzed data that the government allows the public to participate in collecting data by

[1,2]The authors are with Faculty of Engineering, Rajamangala University of Technology Phra Nakhon, E-mail: Korn.p@rmutp.ac.th and Natworapol.b@rmutp.ac.th

informing them via hotline to report on problems encountered on the way to compare with road accidents by the results in the same year's large-scale. Computing research, Lin Cao [2] presented a big data platform to collect data for public traffic services. In the matter of informing the bus time calling a taxi and providing real-time traffic information. There is also mixed research by Jianhua Xie and Jiahui Luo [3] present data collection and analysis route waiting area for passengers, time interval and the number of taxis in the city to develop a smart traffic system with the government. On the other hand, there are different types of traffic research for example: Mingming Zhang [4] introduces the SafeDrive algorithm to alert motorists in the event of a malfunction while driving. By referring to various sensor devices in the car, for example machine rpm and speed, speed and action angles on the road sudden braking to process whether the car is losing control or not. The result is 83% accurate, which means have higher accuracy Rule Based Algorithm at the end of the year, Alessandro Ciociola [5] presented a car sharing platform, UMAP, to compare each company's data. What types of people have the characteristics of choosing a car rental company? distance and time analysis, pickup point and destination.

In 2018, by Giovanni Buroni et al. [6], a study of transport routes coupled with the time of trucks for bringing in a new management process improvement to reduce transportation costs and pollution problems and to help solving the problem of traffic in the city. This article was collected and tested in the Brussels capital region. Belgium, the same year, by Alex Kaplunovich and Yelena Yesha [7] collecting data from taxis in New York City from the beginning of the pickup at the station during the day that the driving style parking in any area. Time of start-up and distance from parking until returning the car at the station to analyze the pollution of taxis in New York City. This year, there are various researches in applying big data to transportation work such as the research of Lu Lin [8] improves road traffic speed predictions by fusing traditional speed detection data with new data types. By bringing data from many aspects to create more efficiency of the data in the future, this research will include GP-based and Distributive GP. In order to make the predictions more accurate along with being able to assess in real time meanwhile, a case study of research for large-scale Intelligent Transport System (ITS) is presented by S Appavu [9]. It is a presentation of the unique model of ITS in India. Due to the driving style, road layout and environmental conditions of India. The ITS Platform in developed countries cannot be used directly. The publisher therefore has to collect and analyze the traffic characteristics of India with real data. In order to use this information to develop ITS systems in the country, another example of research in bringing internet of things (IoT) to help manage by Sang Nguyen [10] the researchers studied IoT technology to create a parking management system that they believe will reduce the time drivers need to be in vehicles with two main goals. The long-term goal is to optimize the management system. Manage parking spaces the short-term goal is help drivers spend less time finding parking spaces by finding suitable parking spaces. In this research, The IoT technology is provided to monitor the area. The duration of parking throughout the city and analysis of nearby parking spaces. The system works without sending a lot of data. Does not back up the data detected by the device. All the data from the cloud server is sent to the cloud server. The analytics uses Hadoop and Map Reduce and is ready to run on the cluster computer.

In the year 2019, Qilei Ren [11] studied the application of Block Chain technology as the center of storage or forwarding data in an ITS and using IoT devices installed as a detection device in various places. It has been compared with a cloud server data center-based system and directly designed peer-to-peer data transmission between vehicles. There is research to solve traffic problems by Yang Zhao-xia and ZHU Ming-hua [12] researched and developed the ability to estimate Travel Time in a flexible range of roads in the Traffic Information Platform. They both commented that Big Data traffic can effectively provide feedback or suggest a solution to traffic congestion. Using Travel Time analysis method from Big Data in dynamic prediction in real time tested on traffic monitoring system structure using RFID tag technology and using fusion data. The test results in the model show that the new method used is more accurate and can solve traffic congestion problems and make traffic smoother. Meanwhile, Li Zhu [13] presented a research study exploring Big Data analytics systems in the intelligent transportation system. The survey describes the history of the unique characteristics of both Big Data technologies and the ITS, a framework used to perform Big Data analysis in ITS results analysis tool. Methods for analyzing results include analysis of road traffic accidents. Road mobility estimation, administration, planning and management of the public transport system. Personal travel planning as well as maintenance planning.

Currently, the research trend of new technology device which is IoT has a priority to develop the Auto detect incident system. In addition, the research the study about analyses data from any interested group with detect incident from social media messages in 2016, by Peerapon Vateekul et al. [14]. The author presents the method to get people feeling and emotion from Thai Tweeter message by using Deep Learning Method with Short Term Memory (LSTM) model and Dynamic Convolutional Neural Network (DCNN) model then compare with other method such as Naïve Bayes and SVM. The result indicate that these two methods were get more high accuracy especially DCNN with 75.35%. Besides, Supon Klaithin et al. [15] present categorize message Thai Twitter by naive Bayes classifiers which is the one of machine learning for categorize detail of event and summary data with separate Tweet for two group that is Information of Road with average correction 88.42% and Classification with average correction 76.40%. In 2020, Tanatorn Tanantong et al. [16] present method for get keyword from Thai message in Twitter with N-gram technique for classified the word that does not in dictionary and increase accuracy of partial word in Twitter by using message regarding to Thai university for classify keyword and trend analysis. The result indicates that they got 70% accuracy.

On the other hand, in a group of researches to solve such problems. Modern methods based on technology, such as pseudo-neural network methods or deep learning, have been introduced into the analysis and processing by Zhenhua Zhang [17]. In 2018, it presented the use of Machine Learning to detect traffic accidents from social media data. More than 1 million tweets were reviewed in one year in two cities: Northern Virginia and New York by selecting key words from sentences with Apriori Algorithm and importing them into ML for further answers. The experiment compares DBN, LMTS, SVMs, and sLDA methods with identifying the area of incidence or according to research by Robert Neuhold [18] said data collection on expressways for traffic management depends on the installation of automatic traffic and traffic monitoring systems with cameras only. The research team presented guidelines for using social media as an additional source of information by ASFINAG, the Austrian expressway transport authority. Developed a system to display the average speed of each kilometer at different times of the day. This allows the driver to quickly get information on the situation in a color chart. It has evolved to be able to collect event descriptions from Facebook and RSS of Austrian radio stations and popular newspaper news which is filtered through useful relevant messages and processed into traffic data and presented as a color symbol for traffic conditions or average speed on the route at that time. Meanwhile, Fang Jin[19] said: More information from social media that has been mentioned or reported on the current incidence. This makes it possible to detect the tendency of accidents and incidents on the road. quickly In this paper, the research team presents the nature of the problem as well as the complexity of applying Spatio-Temporal Data Mining principles. To analyze accidents or incidents on the road Hasna El Alaoui [20] said road accidents are still common. Safety has become a major concern problem. There are many factors that explain a traffic accident such as the nature of the environment at that time, behavior, weather conditions and other complex factors that are uncertain. The outcome of road accidents is not linear. Therefore, it is imperative to explore the relationship between data from many aspects to reduce the risk. The research team has used the data of past accident reports come pre-process and classification by using Data Mining tools to find information that can predict the cause of the accident. The results obtained can validate the data and can help predict new events in the future with similar data. The goal of this research is to analyze the data and select the most accurate data obtained for validation by analyzing the nature and relationships of the data. The research team proposes a decision system for analyzing traffic accident data to extract information related to road risk prevention and also introduces the concept of crowdsourcing. In the process of collecting additional accident-related information from road users the proposed system uses data mining methods and big data techniques suitable for surveying accident data and uses Hadoop to categorize road accidents and determine what is causing the problem. Until now research on social gathering data is ongoing and it is a replacement tool for traditional technologies such as Tatsuya Yamazaki [21] that presents Open Data Accident analysis. Publicly available public information of the city of Niigata, Japan which useful information in the analysis of such incidence problems can be classified as weather forecast information, event statistics of vehicle-to-vehicle accidents (V2V) and Human-to-Vehicle accidents (H2V) and accidents without parties together with panoramic images of red light intersections from the analysis results of all the information mentioned will show that traffic junctions without traffic lights are more V2V incidence than junctions with traffic lights, ultimately Yuanyuan Chen [22]. Presenting traffic-related milestones from Sina Weibo social media, China's Microblogging Platform, using the Bag of Word Model technique to find keywords. Chinese is compared to English for input into the neural network (NN) and to experiment with the use of Convolutional Neural Networks (CNN), Long Short-term Memory (LSTM), and a combination of CNN and LSTM methods to learn to recognize words and events to be used to number and find traffic incidents.

2. DATA PREPARATION

Data preparation to create practice and test data sets consisted of 3 models: a model for grouping incident messages and general publicity messages, a model to categorize the types of incidences and a model to classify the severity of the incidence. The model that was created used the same set of data for training, so to limit errors the same test data set was used. It will test the actual data set collected from social media posts on Twitter. The data must be imported through the Data Preparation process which consists of dividing sentences into words, filtering unwanted words out, changing words into numbers and adding equal sentence lengths. that has come to make labels of text in the next step. The process step of this method was shown in Fig.1

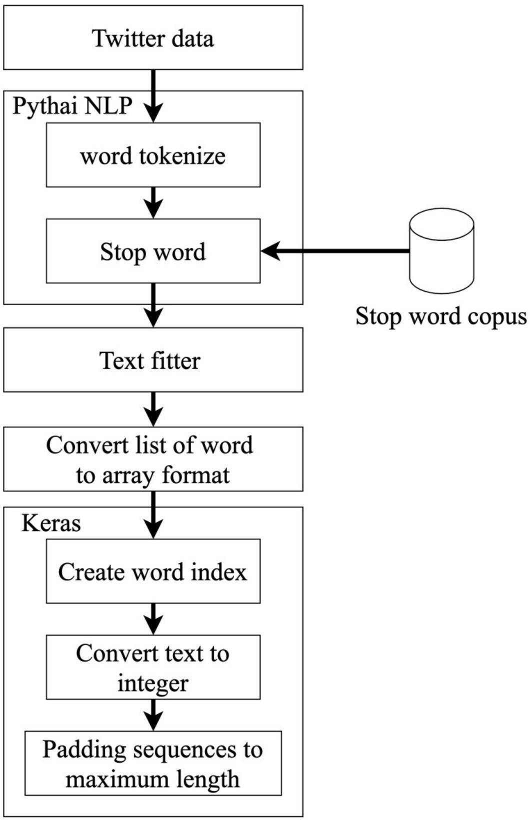

Fig.1: *Data Preparation Process.*

Fig.1 shown the Tweet message was separated to each word with remove word by Stop word method, text filter and convert to array format then increase the word length by using maximum word length

2.1 Identifying Traffic Incidents

Steps to classify messages it starts with taking the information gathered from Twitter messages through various processes in order to prepare the data before the test by carrying out the process of extracting words. filtering out words converting words to numbers in sequence and increasing the length of the text then enters the testing process. In this step, the test-ready text is passed into the classification model, which is the first model. Its purpose is to classify the message between a general press release or an incident notification message.

2.2 Identifying Patterns of Incidence

For the incidence classification testing procedure, it is performed second after the text classification. The text to be tested is an incidence message that has been categorized from the message division of the staff. Therefore, pre-test data is prepared by storing it in the tolerance variable and passing it on to the model to classify the incidence. At this stage, a test for the accuracy of the incidence classification is carried out. The objective is to summarize the answers to the incidence messages to be able to tell the type of incidence by the incidence type of the test data group that has passed the professional personnel classification process. There are total five types, so the message can be tested immediately.

2.3 Identifying the Severity of the Incident

In the process of testing the violence isolation when receiving a message from Twitter. As in the previous process, the incident-only message was used in the test, the process was initiated by the operator-classified message used for pre-processing. Then the text that has gone through the data preparation process is brought into the model to test the validity of the model. For this step, the model is tested to identify three types of profound severity.

2.4 Converting Words into Numbers

Neural network behavior is the process of importing a set of numbers to calculate mathematical calculations according to each learning method. All words from a data set with word indexing process. It is generated by word frequency by giving word indexes in order of repetition of that word. For example, as shown in table 1.

Table 1: *Word Index.*

Index	Word
1	แยก
2	อุบัติเหตุ
3	รถติด
4	จราจร
5	รายงาน
6	การจราจร
7	ช่วง
8	และ
9	ช่อง
10	กัน
11	ขวาง

2.5 Incident of “Stop word”

Word that always used in many sentences but it does not mean in the main clause. However, some repeated word was the key word of the clause. Many research of NLP was encounter on this issue that is “the, a, an” word for English and, “ และ, กับ, ที่, ซึ่ง ” Word for Thai. These words were called “Stop word”. For the preparation data Stop word was cleaned for get the main clause to decrease complexity of learning simulation.

In this research used Stop word list of "pythainlp" which is Thai language. The main concept was cleaning Stop word that is in pythainlp from sentence. e.g. "15.56 เริ่ม ติดกัน เยอะ แล้ว รายงาน จราจร ศุกร์" will clean Stop word to "15.56 ติดกัน รายงาน จราจร ศุกร์" which is remain the meaningful of sentence. After cleaning Stop word, the Copus will decrease 7% of total word.

2.6 Data Augmentation

The detect word and sentence of incident tweet in Twitter was clarified to 5 types and, several of incident was clarified in 3 types. For compare number of messages, indicate some type was a majority and minority that shown in Fig 2 and Fig 3 respectively. The imbalance data of classification could decrease the data reliability in training simulation. To solve this problem, we recommend the Data Augmentation to add and separate imbalance data that shown in pic x (b) and pic x (b)

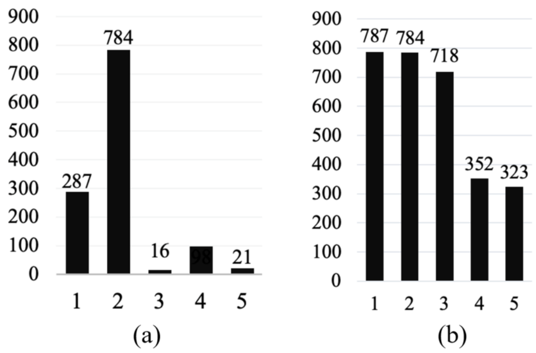

Fig.2: *Multi-Layer Perceptron Model.*

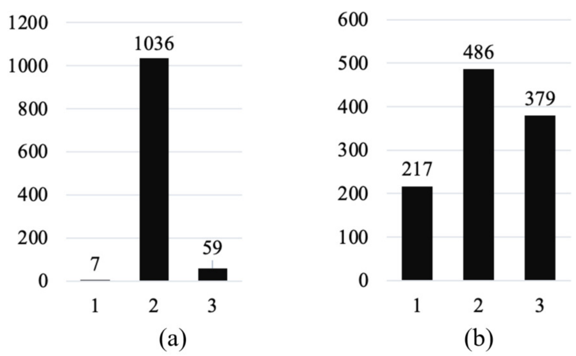

Fig.3: *Multi-Layer Perceptron Model.*

3. MODEL CONFIGURATION INCIDENT DETECTION WITH IN DEEP LEARNING METHODS

3.1 Deep Learning Process with MLP Method

Multi-Layer Perceptron (MLP) is a method developed from perceptron which is a simple structure because it has a single layer. Later, when computing has evolved more and more, the ability to process has been developed later. Perceptron to have more than one layer or as multiple layers to be able to process more precisely and have adopted MLP in processing to classify the following types of process messages.

Parameterization is length of the sentence to be tested is 45 words long. Therefore, the input parameter is set to 45 and in order not to overfit the model, it has to add a dropout function of 0.2 to the number of layers equal to 5 layers, the first 4 node members of the hierarchy have 450 nodes per node hierarchy. Linear function format and in the fifth layer, the last layer has each layer and the last layer output is 2 for specifying the traffic incidence, output is 5 for specifying the pattern of functions. Incidence and output are 3 for determining the severity of the incidence to match the number of data segments to be divided It has a function format as the Softmax.

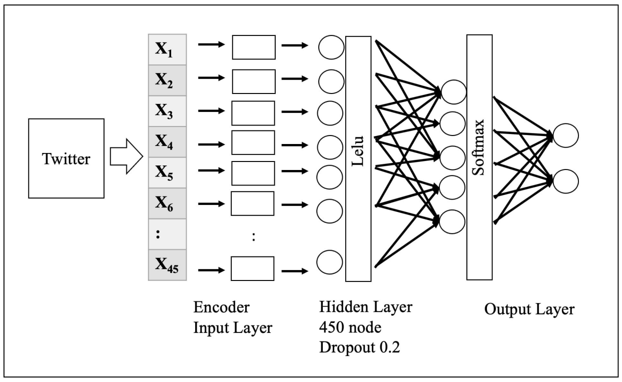

Fig.4: *Multi-Layer Perceptron Model.*

3.2 Deep Learning Process with CNN Method

Convolution Neural Network (CNN) is a form of neural network method, a type of process design algorithms in order to be able to support a set of numbers. CNNs were originally designed to work with image processing, which is actually converting RGB color values to a set of numbers 0-255 and then taking the set of numbers into the CNN model for processing. Therefore, if a CNN supports processing numerically converted images, then the CNN model can support processing numerically converted text as well as above is converts a message into a series of numbers, which is similar to the MLP method, but CNN adds layers of Convolution and Pooling to filter words by 4654×45 dimension of filter (n x d) with n is number of word in corpus and d is the length of filter that same with dimension of characterize the text to bring the result to the last layer and output it as result Number of layers equal to 5 layers, members in each node number 45 nodes in the first hierarchy. The second is the class of the convolutional layer has 250 nodes, the Linear Function format is the third pooling layer, 250 nodes, and the fourth and fifth layers have 2. The output is 5 nodes for specifying the pattern of incidence, and the output is 3 nodes for indicating the severity of the incidence, respectively.

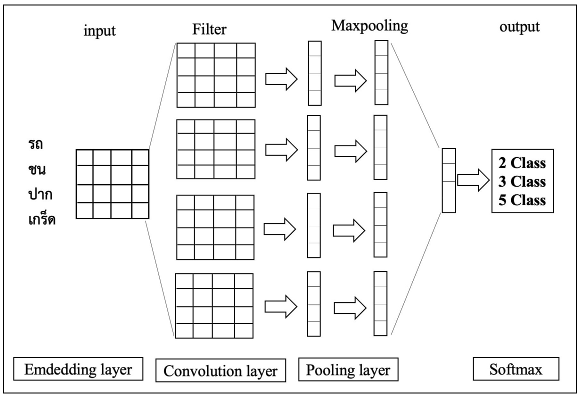

Fig.5: *Convolution Neural Network Model.*

3.3 Bi-LSTM Deep Learning Process

Long/Short Term Memory, or LSTM, is a popular method for text processing or natural language auctions because it employs a technique developed from Recurrent Neural. Networks: RNN, which is unique in that its time series, that is, LSTMs are able to remember past words in a previous hierarchy longer than RNNs in the future.

The LSTM has been developed further. Bidirectional LSTM, unlike Bi-LSTM, can recognize both forward and backward memory cells, whereas LSTM can recognize only backward and for this research, Bi-LSTM is used to process messages to classify the type of incidence. The first layer is to create an Embedding Layer as an output to Bi-LSTM, which has dropout parameters to reduce the overfit of Trends for the next level is the class that summarizes the classification to come out as a final answer. The number of layers is equal to 4 layers, each member of the node in the first layer has 45 nodes of the second hierarchy which is the layer of Bidirectional LSTM has a total of 90 nodes and adds 20% dropout functionality to each layer. In the third hierarchy there are 50 nodes. Linear Function layout and four nodes to sum values as result and 2 nodes for incidence identification, output 5 nodes for incidence pattern identification, and output to 3 nodes for incidence severity level identification number.

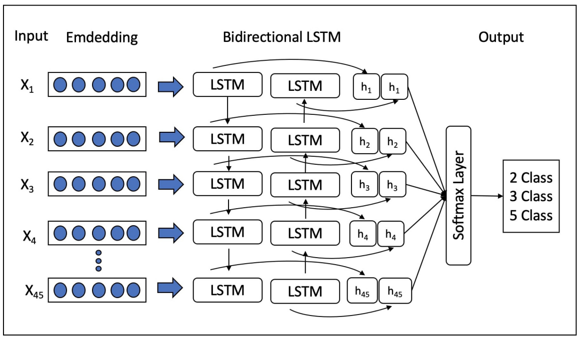

Fig.6: *Bi-LSTM Model.*

3.4 Bi-LSTM Deep Learning Process

Past research has been processed by combining CNN and LSTM. By passing parameters from the CNN obtained from the convolution layer and importing them into the LSTM layer and deciding the prediction results with the Softmax layer for passing the parameters to the model. The method presented here differs from the previous CNN and LSTM methods. The first layer is an embedding layer with 32 dimensions. Convolution layer has 250 nodes and the result is sent to the LSTM layer by specifying parameters. Enter a 32 dimension with a dropout set to 20%. In the final layer, the results are aggregated into responses by a dense layer. Output is set to 2 for incident detection messages, output to 5 for incident classification messages, and output to 3 nodes for severity messages.

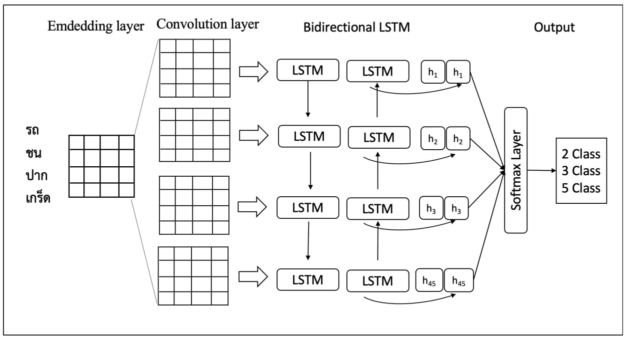

Fig.7: *Hybrid CNN-LSTM Model.*

4. EXPERIMENT

4.1 Assumption and Test cases

The experiment starts with data mining from Twitter account that relate to traffic report which are @js100radio, @Traffic_1197 and @fm91trafficpro. The total data is 3363 messages from 5 February 2021 to 23 March 2021 then we classified data to 3 type which is “Incident identification” with consist of Normal message and Traffic message, “Accident case” with consist of Traffic, Incident, Disaster, Potholes and Other and “Severity identification” with consist of Normal, Heavy and Standstill. We select Train-Test split by 70:30 for machine learning which is the best ratio of our raw data that does not overfitting with high of train data that shown in Table 2.

Table 2: *Comparison of DR with different ratio.*

Ratio	1	2	3	4	5	6
Learning Ratio	50	60	70	80	90	95
Testing Ratio	50	40	30	20	10	5
DR (%)	81.00	81.34	81.87	80.30	83.93	85.71

The table 2 indicate that the ratio 90:10 and 95:5 is high accuracy, but it does not reflect the whole message due to the less test data.

The experiment was setting by using Python with Keras library in Tensorflow that processes to call function for each model for example, modeling with LSTM method by send LSTM parameter to Bidirectional() function.

4.2 Performance Testing by K-Fold Method

After adjusting the parameters of the model to make it ready for learning to model it. In order for learning to bring distributed data to learn, the process is to divide the data into subgroups (k-fold) and bring the subgroups then come back and forth, also known as "Cross validation" to find the mean of the efficiency of the learned model, in this experiment, the data groups were divided into 10 groups, or k=10. In each group, the values were randomly assigned to avoid possible differences. Iterated and tested the model's learning performance for 10 cycles, with each alternating cycle having a learning group picked up and tested, with 70% learning group and 30% testing data. With result in table 2 so, we select that ratio for separate data.

4.3 Accuracy Measurement

To measure accuracy of incident detection, it can be considered with two parameter which is DR (Detection Rate) that for measure correction of incident detection (%). The calculation was shown in equation (1) with D_n is number of incidents that detect by this method and D_t is number of incidents of this experiment.

$$DR = \frac{D_n}{D_t} \times 100 \quad (1)$$

In addition, the measure accuracy is FAR (%) (False Alarm Rate). The equation was shown in equation (2) with N_f is number of messages that false alert and N_t is number of incidents in this experiment.

$$FAR = \frac{N_f}{N_t} \times 100 \quad (2)$$

5. RESULTS

5.1 Results of the identification of traffic incidents

Required to use the data in the test, 1010 data that have already passed the data preparation process.

MLP in the MPL incidence classification test. In the test with real data, the incidence detection accuracy was 67.75% and the incidence detection error was 32.25%.

CNN in the CNN method incidence classification test, the incidence detection accuracy was 91.47% and the incidence detection error was 8.58%.

Bi-LSTM in the LSTM method incidence classification test with an incidence detection accuracy of 93.53% and an incidence detection result of 6.47%.

CNN+LSTM in the combining CNN method and LSTM method incidence classification test with an incidence detection accuracy of 94.06% and an incidence detection result of 5.94%.

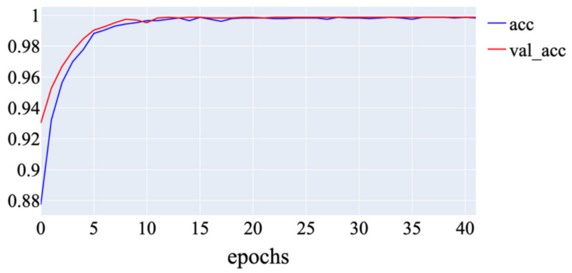

Fig.8: *Incident Detection Accuracy of LSTM.*

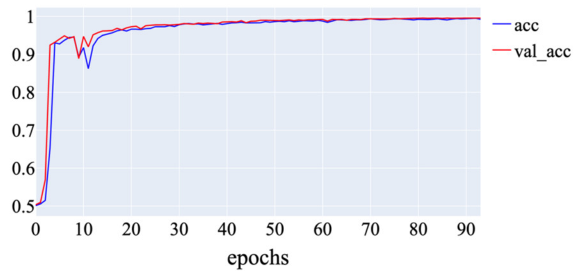

Fig.9: *Incident Detection Accuracy of CNN+LSTM.*

5.2 Categories of incidents

For the data tested, there were 503 data, which were only incidence text data.

MLP in the MPL method incidence classification test. In the test with real data, the incidence detection accuracy was 58.90% and the error incidence detection result was 41.10%.

CNN in the CNN incidence classification test. In the test with real data, the incidence detection accuracy was 85.29% and the error incidence detection result was 14.71%.

Bi-LSTM in the Bi-LSTM method incidence classification test. The incidence detection accuracy was 84.65% and the incidence detection accuracy was 15.35%.

CNN+LSTM in the combining CNN method and LSTM method incidence classification test. The incidence detection accuracy was 84.58% and the incidence detection accuracy was 15.42%.

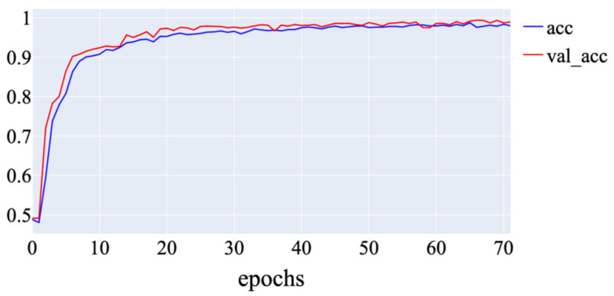

Fig.10: *Incident Classification Accuracy of CNN.*

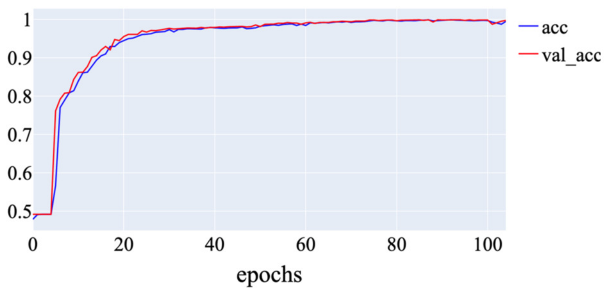

Fig.11: *Incident Classification Accuracy of CNN+LSTM.*

5.3 Incident Severity Determination Results

For the data tested, there were 503 data, which were only incidence text data.

MLP in the MPL method incidence determination test. The MPL method has an incidence severity accuracy of 44.14% and the error was 55.86%.

CNN in the CNN method incidence determination test. The CNN method has an incidence severity accuracy of 93.24% and the error was 6.76%.

Bi-LSTM in the incidence severity determination test the Bi-LSTM method, with an incidence severity accuracy of 77.92% and the error rate was 22.08%.

CNN+LSTM in the incidence severity determination test the CNN+LSTM method, with an incidence severity accuracy of 86.25% and an incidence severity outcome the error rate was 11.70%.

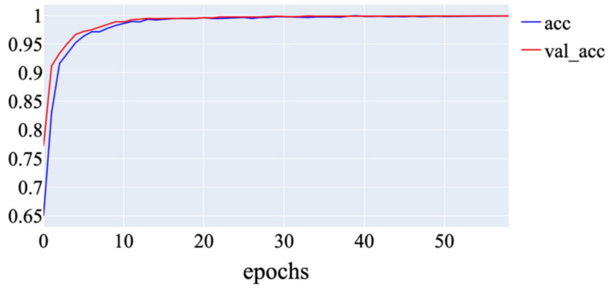

Fig.12: *Incident Severity A5ccuracy of CNN.*

5.4 Analyze the results of the experiment

In the test, results can be recorded to compare the effectiveness of each classification model including MLP, CNN, Bi-LSTM in incidence classification and severity identification to determine their validity. Of the work after deep learning of each system and a text-based test from Twitter for the classification of 1,010 messages, incidence messages and in 1,010 there were 503 incidence messages to be tested against classification and strength-identification patterns. The results of the tests showed the overall picture: the CNN-derived model test pattern had the highest average accuracy rate of 93.53%. However, when observed in the incidence classification, the incidence classification model of only two groups from the CNN-LSTM model had the highest accuracy of 94.06%. Need to Multi Class is a classification with 5 data groups and specifying severity with 3 data groups. It can be seen that the model results obtained from CNN are better, as detailed in the table 3.

However, we can use Data Augmentation to increase accuracy by remove no meaning word and adding to balance message. The result will show in Table 4.

Table 3: *DR and FAR of Deep Learning Method.*

Activity	DR				FAR			
	MLP	CNN	LSTM	CNN+ LSTM	MLP	CNN	LSTM	CNN+ LSTM
Detection	67.75	91.47	93.53	94.06	32.25	8.53	6.47	5.94
Categories	58.9	85.29	84.65	84.58	41.1	14.71	15.35	15.42
Severity	70.1	88.04	88.53	86.26	29.9	11.96	11.47	13.74
AVG	65.58	88.27	88.90	88.30	34.42	11.73	11.10	11.70

Table 4: *Increase Efficiency of Learning Simulation.*

		NI	SW	MD	SWMD
LSTM	Patterns	87.40	88.35	88.41	89.20
	Severity	90.07	90.50	90.78	90.87
CNN	Patterns	84.70	85.05	87.22	87.89
	Severity	91.42	91.76	91.56	91.02
MLP	Patterns	54.34	56.56	58.12	57.44
	Severity	70.54	71.89	72.45	75.37
CNN+ LSTM	Patterns	85.99	86.25	86.89	87.20
	Severity	86.26	87.28	88.20	91.85

As the data of the incidence verification test data, the various methods can be shown to compare the accuracy of the methods in the incidence examination of various characteristics as shown in Fig.13, together with the visual representation of the error values in Fig.14. In the test results, it was found that by comparing the validity values of the three different incidence detection methods, the Bi-LSTM model was able to classify the occurrence of incidence. And the best identification of violence and the CNN and MPL methods have decreased accuracy respectively. At the same time, the CNN method has the best classification capability, the Bi-LSTM and MPL methods are respectively decreasing in accuracy.

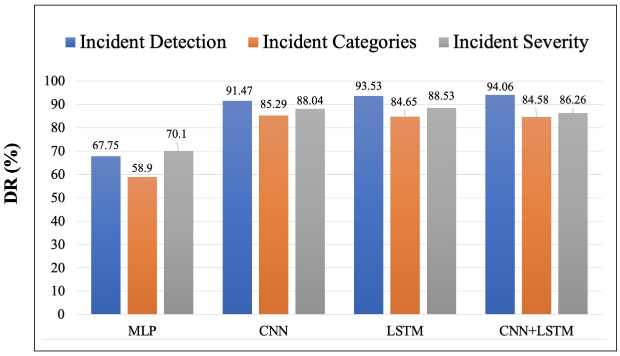

Fig.13: *Detection Rate Comparative*

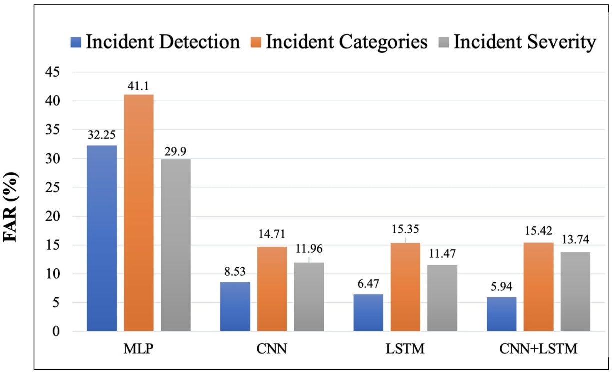

Fig.14: *False Alarm Rate Comparative.*

6. CONCLUSIONS

Traffic incident detection on Twitter social media platform by detecting from a large number of tweets via the Twitter API in this research the sample text will be collected for review and detailed formatting. By people who have expertise in traffic management along with methods for preparing data with word filtering, word wrapping, and word-to-numeric conversion before being introduced into the deep learning process. It includes MLP, CNN, and Bi-LSTM and CNN+LSTM methods, using a three-method validity comparison for traffic incident detection, Identifying the type of incidence and identifying the severity of incidents on Twitter in Thai.

The research will show that the development of an incidence detection system. It is possible to choose a method for detecting and discriminating various forms of data effectively by co-processing between methods. By learning CNN+LSTM with the best incidence detection results with DR 94.06%, FAR 5.94%, while the CNN model of incidence was best identified by with DR 85.29%, FAR 14.71% and Bi-LSTM that well process on this pattern is the best for severity of the incident was identified by DR 88.53% and FAR 11.47%.

The experiment results indicate that for each method of deep learning was proper with different issues of incident detection therefore, the selection method for solve the problem should be consideration for work processing to find solution.

References

[1] B. -H. Lin and S. -F. Tseng, “A predictive analysis of citizen hotlines 1999 and traffic accidents: A case study of Taoyuan city,” *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp.374-376, 2017.

[2] L. Cao, F. Zhu, X. Dong, Z. Shen, J. Yu, B. Hu and G. Xiong, “Big data platform & typical APP services for urban public transportation,” *2017 Chinese Automation Congress (CAC)*, pp. 7565-7570, 2017.

[3] J. Xie and J. Luo, “Construction for the city taxi trajectory data analysis system by Hadoop platform,” *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, pp. 527-531, 2017.

[4] M. Zhang, C. Chen, T. Wo, T. Xie, M. Z. A. Bhuiyan and X. Lin, “SafeDrive, Online Driving Anomaly Detection from Large-Scale Vehicle Data,” in *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2087–2096, 2017.

[5] A. Ciociola, M. Cocca, D. Giordano, M. Mellia, A. Morichetta, A. Putina and F. Salutari, “UMAP, Urban mobility analysis platform to harvest car sharing data,” *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation*, pp. 1-8, 2017.

[6] G. Buroni, Y. Le Borgne, G. Bontempi and K. Determe, “Cluster Analysis of On-Board-Unit Truck Big Data from the Brussels Capital Region,” *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2074-2079, 2018.

[7] A. Kaplunovich and Y. Yesha, “Consolidating billions of Taxi rides with AWS EMR and Spark in the Cloud Tuning, Analytics and Best Prac-

tices," *2018 IEEE International Conference on Big Data (Big Data)*, pp. 4501-4507, 2018.

[8] L. Lin, J. Li, F. Chen, J. Ye and J. Huai, "Road Traffic Speed Prediction: A Probabilistic Model Fusing Multi-Source Data," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 7, pp. 1310–1323, 2018.

[9] S. A. A. Balamurugan, J. F. Lilian and S. Sasikala, "The Future of India Creeping up in Building a Smart City Intelligent Traffic Analysis Platform," *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 518-523, 2018.

[10] S. Nguyen, Z. Salcic and X. Zhang, "Big Data Processing in Fog - Smart Parking Case Study," *2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom)*, pp. 127-134, 2018.

[11] Q. Ren, K. L. Man, M. Li and B. Gao, "Using Blockchain to Enhance and Optimize IoT-based Intelligent Traffic System," *2019 International Conference on Platform Technology and Service (PlatCon)*, pp. 1-4, 2019.

[12] Z. -X. Yang and M. -H. Zhu, "A Dynamic Prediction Model of Real-Time Link Travel Time Based on Traffic Big Data," *2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, pp. 330-333, 2019.

[13] L. Zhu, F. R. Yu, Y. Wang, B. Ning and T. Tang, "Big Data Analytics in Intelligent Transportation Systems A Survey," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 383–398, 2019.

[14] P. Vateekul and T. Koomsubha, "A study of sentiment analysis using deep learning techniques on Thai Twitter data," *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 1-6, 2016.

[15] S. Klaithin and C. Haruechaiyasak "Traffic information extraction and classification from Thai Twitter," *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 1-6, 2016.

[16] R. Bishnoi, M. Ebrahimi, F. Oboril and M. B. Tahoori, "Improving Write Performance for STT-MRAM," in *IEEE Transactions on Magnetics*, vol. 52, no. 8, pp. 1-11, Aug. 2016.

[17] Z. Zhang, Q. He, J. Gao and M. Ni, "A deep learning approach for detecting traffic accidents from social media data," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 580-596, 2018.

[18] R. Neuhold, H. Gursch, R. Kern and M. Cik, "Driver's dashboard – using social media data as additional information for motorway operators," *IET Intelligent Transport Systems*, vol. 12, no. 9, pp. 1116–1122, 2018.

[19] F. Jin and H. Liu, "Detect Hidden Road Hazards combining Multiple Social Media Data," *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2018.

[20] H. El Alaoui El Abdallaoui, A. El Fazziki, F. Z. Ennaji and M. Sadgal, "Decision Support System for the Analysis of Traffic Accident Big Data," *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 514-521, 2018.

[21] T. Yamazaki, "Analysis of Traffic Accident Occurrence in Niigata Prefecture of Japan using Open Data," *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1-4, 2019.

[22] Y. Chen, Y. Lv, X. Wang, L. Li and F.-Y. Wang, "Detecting Traffic Information from Social Media Texts With Deep Learning Approaches," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 3049–3058, 2019.

Korn Puangnak received his Master degree from Department of Computer Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. in 2011. Currently, he is an assistant professor at Department of Computer Engineering, Faculty of Engineering, Rajamangala University of Technology Phra Nakhon, Thailand. His research falls in the domain of machine learning, and ITS (intelligent transport system).

Natworapol Rachsiriwatcharabul received his Ph.D. degree from Engineering Management, University of Missouri, USA. in 1999. Currently, he is the President at Rajamangala University of Technology Phra Nakhon, Thailand. His research falls in the domain of energy management, transportation, smart city, and sustainable industrial management.