Journal of
Information Systems Engineering
and Business Intelligence

Vol.8, No.1, April 2022
Available online at: http://e-journal.unair.ac.id/index.php/JISEBI

Deep Learning Approaches for Multi-Label Incidents Classification from Twitter Textual Information

Sherly Rosa Anggraeni[1], Narandha Arya Ranggianto[2], Imam Ghozali[3], Chastine Fatichah[4], Diana Purwitasari[5]*

[1)2)3)4)5)] Department of Informatics, Institut Teknologi Sepuluh Nopember, Indonesia

Jl. Teknik Kimia, Keputih, Sukolilo, Surabaya

[1)]sherly.rosa@gmail.com, [2)]narandharanggi@gmail.com, [3)]imam.ghz17@gmail.com, [4)]chastine@cs.its.ac.id, [5)]diana@if.its.ac.id

Abstract

Background: Twitter is one of the most used social media, with 310 million active users monthly and 500 million tweets per day. Twitter is not only used to talk about trending topics but also to share information about accidents, fires, traffic jams, etc. People often find these updates useful to minimize the impact.
Objective: The current study compares the effectiveness of three deep learning methods (CNN, RCNN, CLSTM) combined with neuroNER in classifying multi-label incidents.
Methods: NeuroNER is paired with different deep learning classification methods (CNN, RCNN, CLSTM).
Results: CNN paired with NeuroNER yield the best results for multi-label classification compared to CLSTM and RCNN.
Conclusion: CNN was proven to be more effective with an average precision value of 88.54% for multi-label incidents classification. This is because the data we used for the classification resulted from NER, which was in the form of entity labels. CNN immediately distinguishes important information, namely the NER labels. CLSTM generates the worst result because it is more suitable for sequential data. Future research will benefit from changing the classification parameters and test scenarios on a different number of labels with more diverse data.

Keywords: CLSTM, CNN, Incident Classification, Multi-label Classification, RCNN

Article history: Received 19 December 2021, decision after peer review 11 January 2022, accepted 14 February 2022, available online 28 April 2022

I. INTRODUCTION

Technology development has been advancing in disaster management, e.g., handling traffic accidents, natural disasters, or fires. This is particularly important in Indonesia because the traffic accident rate is high. The number reached almost 100,000 in 2013, with 20,000 accidents categorized as fatal [1]. Aside from this, the occurrence of natural disasters is also alarming because the country is located on three major faults, namely the Pacific, Indo-Australian, and Eurasian faults [2]. The seismic activities are high and often result in natural disasters, such as earthquakes, volcanic eruptions, or tsunamis. Meanwhile, fire incidents in Indonesia are often caused by human errors or technical malfunction [3] and often cause social and economic losses. Historically, early incident detection systems used sensors or other hardware to obtain and process data because automatic incident detection was still limited. One of the first studies was by Khan et al. [4], examining early fire detection using a fine-tuned convolutional neural network (CNN) on a CCTV camera.

Nowadays, information flows rapidly on the Internet via social media. Therefore, research has focused on incident detection using social media data [5] [6]. Unlike sensors or hardware, information extraction from social media is inexpensive and real-time [7]. Twitter is a good source of information as it publishes news and updates from various sources—not only the general public but also press companies, public institutions, and influencers—with data covering both facts and personal opinions [8]. However, these data are often mixed with noise, such as complaints to the government regarding public services [9]. For example, users posted complaints mentioning the @sapawargasby, the Surabaya City Government's account; these tweets reached 8,630 as of May 2016 [10]. Complaints like these convolute data about the incidents [8]. Classification is an effective solution to separate factual information about

* Corresponding author

ISSN 2443-2555 (online) 2598-6333 (print) © 2022 The Authors. Published by Universitas Airlangga.

doi: http://dx.doi.org/10.20473/jisebi.8.1.31-41

incidents and complaints on Twitter. Previous research has focused on traffic incident detection and disaster information [11] and used Twitter data with a deep learning approach to detect traffic events [5].

A tweet can contain more than one label [12] [13]. For example, tweets like "*08:00 lalin jalan raya macet ada kebakaran gudang*" which can be translated into "08:00 traffic jam on the highway there is a warehouse fire" has two labels of incident: traffic and fire. Previous research aiming to classify multi-label Arabic texts used a machine learning approach [14]. The study shows that the best results were obtained using feature selection with the LinearSVC classification method and TF-IDF feature representation. Another study used a combination of word unigram, RFDT, and LP to classify multi-label hate speeches [13]. However, from the two studies, the results obtained were not optimal. A solution to this is to combine machine learning and deep learning approaches. For example, Parwez et al. [15] used CNN to classify multi-label Twitter data and resulted in more accuracy than traditional machine learning.

This study uses deep learning methods for multi-label classification, namely CNN, CLSTM, and RCNN. Covering three domains: traffic incidents, natural disasters, and fire incidents, the results of the three methods are compared to determine the most effective method for the classification.

The paper is organized as follows. Chapter 2 describes the related work on incident detection and deep learning classification. Chapter 3 details the proposed methods. Chapter 4 reports the experiment results, and Chapter 5 discusses the results. Finally, Chapter 6 concludes the discussion.

II. LITERATURE REVIEW

A. Incident Detection

Many studies have discussed event or incident detection with various methods and data in recent years. Mercader et al. [16] researched automatic incident detection (AID) based on data from Bluetooth sensors combined with unsupervised anomaly detection. This study analyzed data anomalies on toll roads and assumed a traffic incident would occur if there were an anomaly. Meanwhile, other researchers combined social media with GPS sensors to get better results and more accurate locations. Zheng et al. [17] used taxi GPS and Weibo data to analyze traffic anomalies in China and Wang et al. [18] collected data from different sources such as social media, GPS, points of interest, and weather data to discover traffic congestion and detect traffic anomalies.

The NLP approach has also been used for incident detection, using social media such as Facebook and Twitter. Gu et al. [19] used Twitter data for traffic versus non-traffic incident classification. This study handled binary classification using the Semi-Naive-Bayes (SNB) classifier. Dwi et al. [20] used machine learning approaches such as Decision Tree, Random Forest, and SVM for earthquake detection. Ali et al. [6] researched detecting, analyzing, and monitoring traffic accidents using Facebook and Twitter data. The study used an Ontology and Latent Dirichlet allocation (OLDA) based on a topic-modeling method to label sentences automatically. Then an analyst sentiment was used to identify the traffic polarity to determine the accuracy, and a fastText model and Bi-LSTM were used to detect the event. Meanwhile, Dabiri et al. [5] detected traffic events using Twitter data to recognize three classes: non-traffic, traffic incident, and traffic condition information. Using deep learning classification methods—CNN, RCNN, and CLSTM—the results show that CNN and word2vec were the most suitable for the feature extraction.

B. Deep Learning Classification

Text classification methods can be generally divided into machine learning and deep learning. Deep learning is a subset of machine learning that removes some data pre-processing. CNN and RNN are two deep learning algorithms that can be used for text classification [21] [23] [23]. Convolutional Neural Network (CNN) utilizes layers with convolving filters applied to local features. Created for computer vision, the CNN model proved effective for Natural Language Processing (NLP) and achieved excellent results in semantic parsing, search query retrieval, sentence modeling, and other natural language processing tasks [23]. Liao et al. [23] used a combination of CNN and LSTM to solve multi-label classification. They extract the sequential local semantic information with CNN.

Peng et al. [24] used an end-to-end hierarchical taxonomy-aware and attentional graph capsule recurrent CNN (RCNN) framework to solve the problem of multi-label classification. Lai et al. [25] proposed RCNN that implements a recurrent structure to capture as much contextual information as possible during learning, reducing noise on CNN. Zhou et al. [26] used the CLSTM method to solve the NLP problem. The CLSTM used CNN to extract sentence sequences and inserted them into LSTM to get the representation.

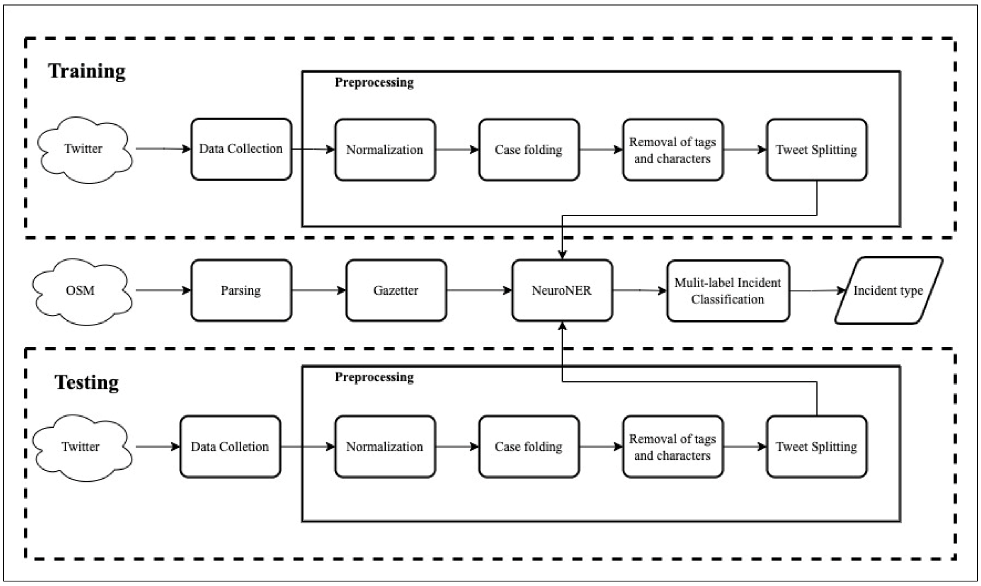

Fig. 1. The Proposed Method

III. METHODS

In this research, we proposed a combination of NeuroNER and variations of deep learning classification methods, namely CNN, CLSTM, and RCNN. Afterward, we classified multi-label information into seven classes: disaster information, disaster complaints, traffic information, traffic complaints, fire information, fire complaints, and non-incidents. The multi-label classification process consists of four stages: data collection, pre-processing, entities recognition using NeuroNER, and deep learning classification. The following steps as shown in Fig.1.

A. Data Collection

In this research, we use the Indonesian Twitter dataset, but we provide a translation of the data in English. We use two types of data: Twitter and gazetteer data or place names. The Twitter data were split into training and testing. This data was fed to the NER system and the incident classification system. Additionally, we used gazetteer data to restrict the incident area.

Twitter data for training were obtained by crawling the Twitter API from twitter.com. We searched the Twitter data with these keywords; *'kebakaran'* ('fire'), *'kecelakaan lalu-lintas'* ('traffic accident), *'banjir'* ('flood'), *'bencana alam'* (natural disasters), *'macet'* ('traffic jam'), *'gempa bumi'* (earthquake), *'puting beliung'* ('typhoon') from November 2020 to April 2021. Each keyword represents three incident domains, '*kebakaran*' ('fire') represents a fire incident, '*kecelakaan lalu-lintas*' ('traffic accident'), '*macet*' ('traffic jam') represent traffic incidents, '*banjir*' ('flood'), '*bencana alam*' (natural disasters), '*gempa bumi*' (earthquake), '*puting beliung*' ('typhoon') represents natural disaster incidents. In addition, data was also searched based on the user ID by following the Twitter timeline on the Radio Suara Surabaya's account (@e100ss), BMKG (@infoBMKG), Sapawarga Kota SBY (@SapawargaSby), then saved them in a text format. The total data used in this research was 2,920 data for training. Radio Surabaya had the highest number of tweets labelled as traffic incidents (592 out of 1,876). Meanwhile, 1,029 out of 1,030 BMKG tweets were labelled earthquake. Sapawarga Kota SBY had far fewer tweets (14) labelled as natural disaster and flood. The data for testing were obtained from streams using the Twitter API on the selected accounts, and the same method was applied to the training data. The time range was real-time stream from 28 October to 3 November 2021, as much as 257 data.

Gazetteer or place names is obtained by parsing data from digital map serviceopenstreetmap.org (OSM) and saved in an XML file by limiting the geographical area around Surabaya. Data parsing aims to get the type of data such as city name, street name, and place/building name. The data is stored in a database used as part of the NER training.

B. Pre-processing

Research by Dai et al. [27] stated that Twitter data contains noise that will generate unsatisfactory results, so data pre-processing is vital to improve the model. We normalize tweets by trimming lines into one line and changing abbreviations into long forms such as event information, street names, and places to identify incident information more accurately. Then, it will be converted into small forms through case-folding to make the casing uniform so there will not be a difference such as "banjir", "Banjir", and "BANJIR" ('flood'). We also removed hashtags, mentions, hyperlinks, or characters because it is not relevant to incident detection. However, we retained several characters, such as period (.), dash (-), and question mark (?), to distinguish the sentences. For tweets containing multiple events, we split them according to the numbering such as "1.", "2.", "3.", etc., to avoid misclassification. The example of pre-processing is shown in Table 1.

TABLE 1
EXAMPLES OF PREPROCESSING

Pre-processing	Raw Data	Data Cleaning
Normalization	Dishub 13.31 wib: Arus lalu lintas A. Yani (depan royal plaza) arah masuk kota padat merambat imbas banjir yang cukup tinggi depan RSI Wonokromo.	Dishub 13.31 wib: Arus lalu lintas Ahmad Yani (depan royal plaza) arah masuk kota padat merambat imbas banjir yang cukup tinggi depan Rumah Sakit Islam Wonokromo.
	Dishub 13.31 wib: A. Yani traffic flow (in front of the royal plaza) in the direction of entering the city congested with the impact of a high enough flood in front of the RSI Wonokromo.	Dishub 13.31 wib: Ahmad Yani traffic flow (in front of the royal plaza) in the direction of entering the city congested with the impact of a high enough flood in front of the Wonokromo Islamic Hospital.
Case-folding	MEWASPADAI BANJIR LAHAR GUNUNG BROMO http://fb.me/SuGtPs6X	mewaspadai banjir lahar gunung bromo http://fb.me/sugtps6x
	MOUNT BROMO'S LAVA FLOOD CAUTIOUS http://fb.me/SuGtPs6X	mount bromo's lava flood cautious http://fb.me/sugtps6x
Removing hashtags, mentions, hyperlinks, or characters	14.49: Info awal #kecelakaan beruntun di Tol Gunungsari arah Waru. Ada tiga kendaraan yang terlibat, dua truk dan satu mobil. Posisi di lajur kanan. Lalu lintas macet. (odp-ft)	14.49 info awal beruntun di tol gunungsari arah waru. ada tiga kendaraan yang terlibat dua truk dan satu mobil. posisi di lajur kanan. lalu lintas macet. odp-ft
	14.49: Initial info for the multi-vehicle car #accidents on the Gunungsari Toll Road, Waru direction. There were three vehicles involved, two trucks and a car. Position in the right lane. Traffic jam. (odp-ft)	14.49 initial info for the multi-vehicle car on the gunungsari toll road waru direction. there were three vehicles involved two trucks and a car. position in the right lane. traffic jam. (odp-ft)
	#Gempa Mag:6.7, 10-Apr-21 14:00:15 WIB, Lok:8.95 LS,112.48 BT (90 km BaratDaya KAB-MALANG-JATIM), Kedlmn:25 Km, tdk berpotensi tsunami #BMKG	magnitude 6.7 10-apr-21 14:00:15 wib lok 8.95 lintang selatan 112.48 bujur timur 90 km baratdaya kabupaten-malang-jawa timur kedalaman 25 km tidak berpotensi tsunami
	#Earthquake Mag:6.7, 10-Apr-21 14:00:15 WIB, Lok: 8.95 South Latitude, 112.48 East Longitude (90 km Southwest KAB-MALANG-JATIM), depth: 25 km, no tsunami potential #BMKG	magnitude 6.7 10-apr-21 14:00:15 wib lok 8.95 south latitude 112.48 east longitude 90 km southwest malang-district-jawa timur depth 25 km no tsunami potential
Splitting tweet	@e100ss 17.58: 1. Underpass HR. Muhammad macet dua arah. 2. Mantup Lamongan arah Driyorejo Gresik macet, imbas banjir di Benjeng Gresik. Lalu lintas dialihkan lewat Balongpanggang Gresik- Metatu. 3. Setelah Tol Tembelang arah Jombang lalu lintas macet. (odp-ft)	17.58 1. underpass hr. muhammad macet dua arah. 2. mantup lamongan arah driyorejo gresik macet imbas banjir di benjeng gresik. lalu lintas dialihkan lewat balongpanggang gresik- metatu. 3. setelah tol tembelang arah jombang lalu lintas macet. odp-ft
	@e100ss 5:58pm: 1. Stucked both ways in HR. Muhammad Underpass. 2. Mantup Lamongan in the direction of Driyorejo Gresik was jammed, due to flooding in Benjeng Gresik. Traffic was diverted via Balongpanggang Gresik-Metatu. 3. Traffic jams occurred after the Tembelang Toll Road in the direction of Jombang. (odp-ft)	17:58 1. Stucked both ways in HR. Muhammad Underpass. 2. mantup lamongan in the direction of driyorejo gresik was jammed, due to flooding in benjeng gresik. traffic was diverted via balongpanggang gresik-metatu. 3. traffic jams occurred after the tembelang toll road in the direction of jombang. (odp-ft)

TABLE 2
EXAMPLES OF AN ENTITY LABELING

Label Entity	Description	Examples of an Entity
LOC	Location names such as tolls, roads, sub-districts, urban villages	Tol Sidoarjo-Porong, Ahmad Yani, Tol Gunungsari, Waru, Wonokromo, Waru, Gayungan, Jambangan, Wonocolo, Rungkut, Trenggilis Mejoyo, Underpass HR. Muhammad, Mantup, Driyorejo, Benjeng, Balongpanggang, Metatu, Tol Tembelang
GPE	The city/district name around Surabaya	Malang, Surabaya, Gresik, Lamongan, Jombang
HWYMSE	Measurement for highway	Km 757, 90 km
BLD	Names of buildings or offices	Royal Plaza, Rumah Sakit Islam
NPL	Names of natural places such as rivers, mountains and beaches	Gunung Bromo
OBJ	Vehicles or person object	Truk, mobil, kendaraan, orang, petugas
MSE	Measurement of incident information excluding highways	15 km/jam, Mag:7.2, 25 Km, 8.95 LS, 112.48 BT
TIME	Occurrence time	13.31, 14.49, 20.20, 17.58, 14.00
DATE	Dates or periods	1-Apr-21
O	Other named entities	ada, kecelakaan, lalu, lintas, sudah, ditangani, ekor, kepadatan, sdh, nyampek, pintu, tol, sda, wib, arus, depan, arah, masuk, kota, padat, merambat, imbas, banjir, yang, cukup, tinggi, depan, wib, lokasi, kabupaten, jawa, timur, kedalaman, tidak, berpotensi, tsunami, info, awal, beruntun, di, tiga, yang, terlibat, dua, satu, posisi, lajur, kanan, macet, odp, ft, selamat, malam, pantauan, hari, senin, pukul, hujan, wilayah, kecamatan, dari, ke, barat, dengan, kecepatan, terpantau, selatan, lewat, setelah, mewaspadai, lahar.

C. NeuroNER

Abu-Gellban [8] stated that classification of events requires an information extraction process using the Named Entity Recognition (NER) technique. The entity's identification results are used as event information to be processed for classification. Putra et al. [11] used NeuroNER on events for the extraction process.

Named entity recognition identifies and categorizes key information that can be recognized under categories like location, geographical entity, highway measurement, etc., as shown in Table 2. Lample et al. [28] mentioned that sentences are typically expressed using the IOB format (Inside, Outside, Beginning). The BIO schema is a simple tag with the concept of begin-of-entity or continuation-of-entity division. For example, Table 2 shows "Ahmad Yani", it will be annotated as B for "Ahmad", while "Yani" will be annotated as I. The O schema defines words that do not belong to the same entity.

D. Classification

The classification uses seven classes: natural disaster information, natural disaster complaints, traffic information, traffic complaints, fire information, fire complaints, and non-incidents. We compared three deep learning classification methods: CNN, CLSTM, RCNN with the same parameters, as shown in Table 3.

1) CNN

As shown in Fig. 2, The CNN model begins with forming embedding vectors, where each tweet is a sequence of words $x_1, x_2, x_3, \dots, x_n$. The vector is derived from the introduction of entities (e.g., B-LOC and I-LOC) from NeuroNER, while the word with the entity form O will be returned to the original word. This model embeds each symbol as a dimension to form$x_1, x_2, x_3, \dots, x_n \in R^d$. The features in the convolution layer will be extracted from the word vector using a kernel. The window will slide with the kernel size k to include the whole word. For a k-sized window $(x_i, \dots, x_{i+k-1})$, the convolution takes the concatenation vector $u_i = [x_i, \dots, x_{i+k-1}] \in R^{k \times d}$. The result of the features obtained from multiplies by the convolution matrix $F_i = u_i \times W$, where $W \in \mathbb{R}^{(k \cdot d) \times m}$. We used max pooling by taking the greatest value, and the fully connected linear layer performs the output of the classification class using softmax.

2) CLSTM

CLSTM is a CNN combined with LSTM, as illustrated in Fig. 3. The CNN model is the same as the previous step using 1D convolution. The CLSTM process combines CNN and LSTM with advantages in terms of local feature extraction and long-term dependencies information. LSTM has the basic architecture of RNN, where processes occur sequentially. LSTM overcomes long-term dependencies when a large amount of information is needed on the input and output sequences, which causes vanishing and exploding gradients [5]. The results of the convolution will be given to LSTM, which has three main gates, namely forget f_t, input i_t, and output o_t [29]. It will go into the forget gate to decide what information is removed from the cell state. The next step is to store the cell state by updating the value via the input gate. The value of the forget gate result will be multiplied by the input gate result to become a candidate for the new value. Finally, the output gate uses the sigmoid layer to determine the part to be output. Then, the value will be continued by multiplying the cell state tanh.

3) RCNN

RCNN model consists of RNN architecture and CNN architecture such as max-pooling. The RNN process will capture semantic functions so that it helps to get the meaning of the word precisely. In this model, the recurrent structure uses Bi-directional RNN (Bi-RNN), as shown in Fig. 4. The model can exploit data from both directions, namely for the past and the future. Bi-RNN has forward and backward states for each data instance on the hidden layer. The result of Bi-RNN is a latent semantic feature that will be continued in the max-pooling layer. Finally, the model processes the output on the softmax layer.

TABLE 3
DEEP LEARNING PARAMETER

Deep Learning Method	Parameter
CNN	Hidden Layer: 512 Keep Probability: 0.7 Batch Size: 64 Epoch: 15 Learning Rate: 1e-3 Sentence Length: 200 Number of Filter: 128
CLSTM	Hidden Layer: 512 Keep Probability: 0.7 Batch Size: 64 Epoch: 15 Learning Rate: 1e-3 Sentence Length: 200 Number of Filter: 128 Number of Layer: 2
RCNN	Hidden Layer: 512 Keep Probability: 0.7 Batch Size: 64 Epoch: 15 Learning Rate: 1e-3 Sentence Length: 200 Number of Filter: 128

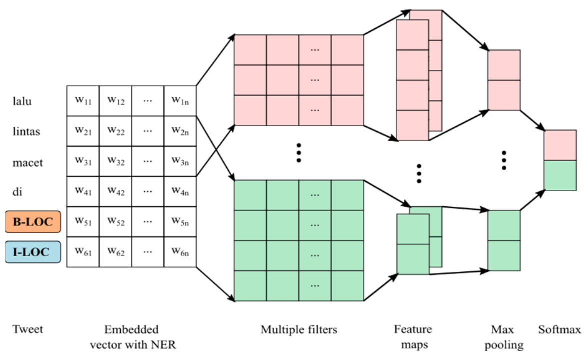

Fig. 2 The CNN Architecture

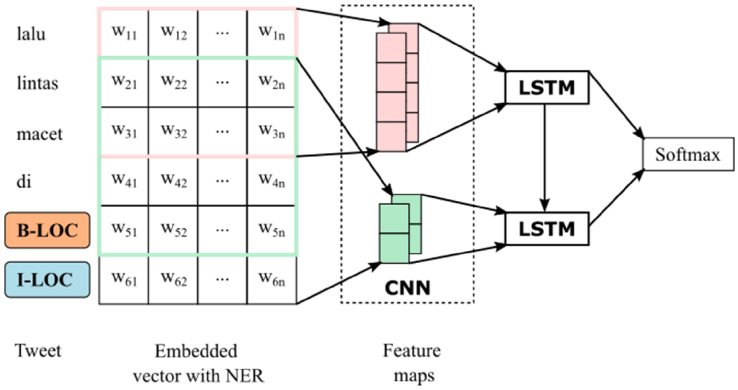

Fig. 3 The CLSTM Architecture

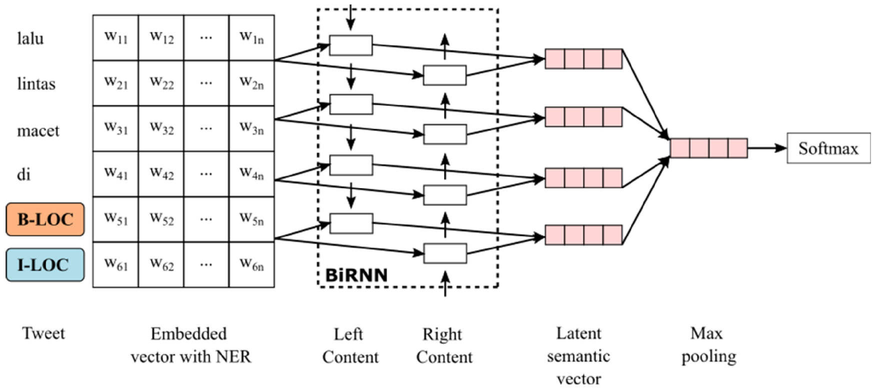

Fig. 4 The RCNN Architecture

E. Evaluation

At this stage, we used a confusion matrix to test the performance of the NeuroNER evaluation and CNN, CLSTM, and RCNN evaluation. According to Nam et al. [30] precision, recall, F1-score can be used to measure the performance evaluation of multi-label classification. Precision divides the positive true class predictions and the overall positive class predictions. Then recall is a true positive division by the total TP and FN. Finally, F1-score is the harmonic mean of precision and recall.

IV. RESULTS

The results of the pre-processing and named entity recognition stages are shown in Table 4. The distribution of the entities (LOC, GPE, BLD, NPL, HWYMSE, OBJ, MSE, TIME, DATE) from the test data can be seen in Table 5. Other entities signify words that were not included in the entity categories, for example, me, you, us, when, there, what, and so on. The test results were entered into the confusion matrix table. Based on the results of the recording in the confusion matrix, precision, recall, and f-measure values can be calculated for the named entity recognition method. The results of the calculation of precision, recall, and f-measure for the named entity recognition method were: 92.03%, 94.07%, and 92.35%. NER trials were carried out to see the results of entity recognition, because entity recognition is influential on the continuation of the stage. If a recognized entity showed multiple errors, data retraining is needed.

The results of the pre-processing and entity identification stages were processed in the complaint classification. Tweet data with at least one of the LOC, GPE, BLD, and/or NLP entities were classified according to the existing model. The performance result of the model building can be seen in Table 6. For the classification stage, we used three different methods with the same parameters to make the comparison clearer. The results of the classification comparison can be seen in Table 7.

TABLE 4
NAMED ENTITY RECOGNITION RESULTS

Test Data	Preprocessing and NER Results
Kebakaran bus di tol manyar - kebomas km 10.600 *Bus fire on manyar-kebomas toll road 10,600*	kebakaran bus di tol loc – loc hwymse *bus fire on toll loc – loc hwymse*
Macet di legundi arah kedamean ada apa ya *Traffic jam in legundi in the direction of kedamean, what's wrong*	macet di loc arah loc ada apay a *traffic jam in loc in the direction of loc, what's wrong*
7.48: waspadai kepadatannya buntaran arah manukan padat *7.48: beware of the density buntaran direction manukan traffic jam*	time: waspadai kepadatannya loc arah loc padat *time: beware of the density loc direction loc traffic jam*
11.04: info awal kebakaran di pelabuhan gresik *11.04: initial information about the fire at the port of gresik*	time: info awal kebakaran di pelabuhan gpe *time: initial information about the fire at the port of gpe*
Info Gempa Mag:4.1 SR, 01-Nov-21 23:20:02 WIB, Lok:10.43 LS,113.22 BT (252 km BaratDaya JEMBER-JATIM), Kedlmn:20 Km: BMKG - PGR VII *Earthquake Info Mag:4.1 SR, 01-Nov-21 23:20:02 WIB, Lok:10.43 South Latitude, 113.22 East Longitude (252 km Southwest JEMBER-JATIM), Kedlmn:20 Km: BMKG - PGR VII*	info gempa mag: mse, date time wib, lok:10.43 ls,113.22 bt (mse baratdaya gpe-jatim), kedlmn:mse: bmkg - pgr vii *earthquake info mag: mse, date time wib, lok:10.43 south latitude, 113.22 east longitude (mse southwest gpe-jatim), kedlmn:mse:*
Jalan depan galaxy mall ada yang diperbaiki, hati hati bagi pengendara yang lewat *The road in front of the galaxy mall has been repaired, be careful for passing motorists*	jalan depan bld ada yang diperbaiki, hati hati bagi pengendara yang lewat *the road in front of the bld has been repaired, be careful for passing motorist*

TABLE 5
NAMED ENTITY RECOGNITION CONFUSION MATRIX

		Predicted Class									
		A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Actual Class	**A**	129		8							
	B	24	53								
	C	5		46							
	D				0						
	E					2					
	F						18				
	G							7			
	H								3		
	I									62	
	J	47									1600

Description:
A. LOC Entity
B. GPE Entity
C. BLD Entity
D. HWYMSE Entity
E. NPL Entity
F. TIME Entity
G. DATE Entity
H. MSE Entity
I. OBJ Entity
J. Other Entity

TABLE 6
COMPARISON OF ACCURACY

Classification Method	Accuracy (%)	
	Training	**Validation**
Convolutional Neural Network	97.08	97.02
Convolutional Long Short-Term Memory	90.19	85.18
Recurrent Convolutional Neural Network	92.03	94.07

TABLE 7
THE RESULT OF THE CLASSIFICATION COMPARISON

Test Data	**Type of Incident CNN**	**Type of Incident CLSTM**	**Type of Incident RCNN**	**Actual Incident Type**
daerah loc berhenti total ada apa ya, apa karena hujan badai *the loc area stopped completely, what's the matter, is it because of the rainstorm*	- Traffic Complaint - Natural Disaster Complaint	- Traffic Complaint	- Traffic Complaint	- Traffic Complaint - Natural Disaster Complaint
hati hati jalan loc padat merayap, kendaraan menghindari pohon tumbang di depan bld setelah hujan badai semalam *be careful the loc road is crowded, vehicles are avoiding falling trees in front of the bld after a rainstorm last night*	- Traffic Information - Natural Disaster Information	- Traffic Information - Natural Disaster Information	- Traffic Information	- Traffic Information - Natural Disaster Information
loc waspada banyak kendaraan mogok akibat berusaha melewati banjir *loc alert many vehicles broke down as they are trying to pass through the flood*	- Natural Disaster Information	- Natural Disaster Information	- Natural Disaster Information	- Traffic Information - Natural Disaster Information
barusan ada kecelakaan di jalan loc, hujan bikin penglihatan kabur, hati hati sob *there was an accident on the loc road, the rain blurred vision, be careful, friend*	- Traffic Complaint	- Natural Disaster Complaint	- Traffic Complaint	- Traffic Complaint - Natural Disaster Complaint
kebakaran bus di tol loc - loc hwymse *bus fire at loc - loc toll road hwymse*	- Fire Complaint	- Fire Complaint	- Fire Complaint	- Fire Complaint
time: info awal kebakaran di lahan kosong jl loc seberang showroom sudah ada 1 unit pmk di lokasi. *time: initial information on fire on vacant land on jl loc opposite the showroom. There is already 1 firefighter unit on site.*	- Fire Information	- Fire Information	- Fire Information	- Fire Information
iyah kek kebakaran jenggot semua tua umur doang *yes, looks like everyone's beard is on fire, aged but not matured*	- Non-Incident	- Fire Complaint	- Non-Incident	- Non-Incident

We used 237 data for testing the multi-label classification. Seen from Table 7, the classification method that generated the most accurate results is the CNN method. The CNN test results were entered into the confusion matrix table. With the recording results in the confusion matrix, precision, recall, and f-measure values can be calculated for the classification. The calculation results produced an average of 88.54%, 87.11%, and 87,66%. The comparison of the average between all classification methods is shown in Table 8. Although the average results were good, some data were still classified as misclassified, as shown in Table 4, row 4.

TABLE 8
COMPARISON OF AVERAGE PRECISION, RECALL, AND F1

Classification Method	**Precision (%)**	**Recall (%)**	**F1 (%)**
Convolutional Neural Network	88.54	87.11	87.66
Convolutional Long Short-Term Memory	80.37	85.28	82.82
Recurrent Convolutional Neural Network	85.16	83.19	84.17

V. DISCUSSION

Traffic accidents, natural disasters, and fires are common incidents in Indonesia, causing social and economic losses. Therefore it is necessary to detect events and provide early warning. Social media, especially Twitter, is often used to share incidents, but the classification needs to filter the noisy information. To obtain specific incident information, special handling is needed, such as multi-label classification. The current study classified multi-label information of traffic accidents, natural disasters, and fires. This research consists of two test scenarios. The first scenario examines how the results of pre-processing and named entity recognition can recognize the test data. The second scenario compared the three CNN, CLSTM, and RCNN methods to determine which one is the most suitable for multi-label classification.

The named entity recognition makes the results robust because we used plenty of training data. If the results show many errors, we add the training data. We compared the CNN, CLSTM, and RCNN methods for multi-label classification in the second scenario. CNN showed the best results with an average precision value of 88.54%. This is because the data we used for classification results from NER were in the format of entity labels. CNN can spot the important information directly, i.e., the label from NER. Meanwhile, CLSTM showed the worst result because it is more suitable for sequential data.

We used 15 epochs to prevent overfitting of the training data. CNN has the most straightforward architecture compared to CLSTM and RCNN. With simpler architecture, CNN precision is better by 8% compared to CLSTM and 5% compared to RCNN. Wang et al. [31] reduced overfitting by reducing the network complexity. The complete result of the comparison can be seen in Table 6. Unlike CNN, CLSTM showed the lowest precision and f1 measurements. The CLSTM network architecture was the most complex because it combined CNN and LSTM, so it tended to overfit when solving a multi-label classification problem. A dropout parameter can be used to solve overfitting in CLSTM, but it takes longer to train data and is more difficult to implement. Also, CLSTM cannot handle data with too long sentences, so not all information can go through the training stage to produce accurate results.

Fatra et al. [12] mentioned that the combination of NeuroNER and RCNN generate good results but has not been tried with multi-label classification. The limitation of our study is that it has not been tested on three or more labels. The test data we achieved does not contain more than two class labels.

VI. CONCLUSION

The rapid development of information technology makes it possible for people to exchange information quickly through social media. Twitter is a good source of information because it is accessible and up-to-date. This study used a combination of NeuroNER and variations of deep learning classification methods—CNN, CRNN, and CLSTM—for multi-label incidents classification. There are seven classes: disaster information, disaster complaints, traffic information, traffic complaints, fire information, fire complaints, and non-incidents. The use of named entity recognition as part of entity recognition yielded good results. The calculation of precision, recall, and f-measure for the named entity recognition method reached 92.03%, 94.07%, and 92.35%. From the multi-label incidents classification experiment with different deep learning methods, there were some misclassified data, but the best results were shown by the CNN method with an average calculation of precision, recall, and f-measure for the named entity recognition method reaching 88.54%, 87.11%, and 87,66%. Future work will benefit from testing with real-time data.

Author Contributions: *Sherly Rosa Anggraeni*: Software, Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing. *Narandha Arya Ranggianto, Imam Ghozali*: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing. *Chastine Fatichah, Diana Purwitasari*: Validation, Investigation, Supervision.

Funding: research received no specific grant from any funding agency.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

[1] Soehodho Sutanto, “Public Transportation Development and Traffic Accident Prevention in Indonesia,” *IATSS Research*, vol. 38, no. 1, pp. 7–13, Jul. 2014, doi: 10.1016/j.iatssr.2014.06.001.

[2] R. D. Kusumastuti, Viverita, Z. A. Husodo, L. Suardi, and D. N. Danarsari, “Developing a resilience index towards natural disasters in Indonesia,” *International Journal of Disaster Risk Reduction*, vol. 10, no. PA, pp. 327–340, Dec. 2014, doi: 10.1016/j.ijdrr.2014.10.007.

[3] T.-H. Chen, P.-H. Wu, and Y.-C. Chiou, “An Early Fire-Detection Method Based on Image Processing,” in *International Conference on Image Processing*, pp. 1707–1710, 2004, doi: 10.1109/ICIP.2004.1421401.

[4] K. Muhammad, J. Ahmad, and S. W. Baik, “Early fire detection using convolutional neural networks during surveillance for effective disaster management,” *Neurocomputing*, vol. 288, pp. 30–42, May 2018, doi: 10.1016/j.neucom.2017.04.083.

[5] S. Dabiri and K. Heaslip, “Developing a Twitter-based traffic event detection model using deep learning architectures,” *Expert Systems with Applications*, vol. 118, pp. 425–439, Mar. 2019, doi: 10.1016/j.eswa.2018.10.017.

[6] F. Ali, A. Ali, M. Imran, R. A. Naqvi, M. H. Siddiqi, and K. S. Kwak, “Traffic accident detection and condition analysis based on social networking data,” *Accident Analysis and Prevention*, vol. 151, Mar. 2021, doi: 10.1016/j.aap.2021.105973.

[7] S. Dabiri, K. Heaslip, and C. E. Via, “Transport-domain applications of widely used data sources in the smart transportation: A survey,” arXiv preprint 1803.10902. 2018, doi: https://doi.org/10.48550/arXiv.1803.10902.

[8] H. Abu-Gellban, “A Survey of Real-Time Social-Based Traffic Detection,” in *International Conference on Intelligence and Security Informatics*, Nov. 2020, pp. 1–6. doi: 10.1109/ISI49825.2020.9280534.

[9] A. N. Rasyid and A. Purwarianti, "Sentiment Classification for Indonesian Message in Social Media," in *International Conference on Cloud Computing and Social Networking (ICCCSN)*, 2012, pp. 1–5. doi: 10.1109/ICCCSN.2012.6215730.

[10] M Lailiyah, S Sumpeno, and I.K.E Purnama, "Sentiment Analysis of Public Complaints Using Lexical Resources Between Indonesian Sentiment Lexicon and Sentiwordnet," in *International Seminar on Intelligent Technology and Its Application*, 2017, pp. 307–312. doi: 10.1109/ISITIA.2017.8124100.

[11] F. N. Putra and C. Fatichah, "Klasifikasi jenis kejadian menggunakan kombinasi neuroner dan recurrent convolutional neural network pada data twitter," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 4, no. 2, pp. 81–90, Jul. 2018, doi: 10.26594/register.v4i2.1242.

[12] R. Hendrawan and S. al Faraby, "Multilabel Classification of Hate Speech and Abusive Words on Indonesian Twitter Social Media," in *2020 International Conference on Data Science and Its Applications*, 2020, pp. 1–7, doi: 10.1109/ICoDSA50139.2020.9212962.

[13] M. O. Ibrohim and I. Budi, "Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter," in *Proceedings of the Third Workshop on Abusive Language Online*, 2019, pp. 46–57, doi: https://doi.org/10.18653/v1/W19-3506.

[14] A. Omar, T. M. Mahmoud, T. Abd-El-Hafeez, and A. Mahfouz, "Multi-label Arabic text classification in Online Social Networks," *Information Systems*, vol. 100, Sep. 2021, doi: 10.1016/j.is.2021.101785.

[15] M. A. Parwez, M. Abulaish, and Jahiruddin, "Multi-Label Classification of Microblogging Texts Using Convolution Neural Network," *IEEE Access*, vol. 7, pp. 68678–68691, 2019, doi: 10.1109/ACCESS.2019.2919494.

[16] P. Mercader and J. Haddad, "Automatic incident detection on freeways based on Bluetooth traffic monitoring," *Accident Analysis and Prevention*, vol. 146, Oct. 2020, doi: 10.1016/j.aap.2020.105703.

[17] Z. Zheng, C. Wang, P. Wang, Y. Xiong, F. Zhang, and Y. Lv, "Framework for fusing traffic information from social and physical transportation data," *PLoS ONE*, vol. 13, no. 8, Aug. 2018, doi: 10.1371/journal.pone.0201531.

[18] S. Wang *et al.*, "Computing urban traffic congestions by incorporating sparse GPS probe data and social media data," *ACM Transactions on Information Systems*, vol. 35, no. 4, Jul. 2017, doi: 10.1145/3057281.

[19] Y. Gu, Z. Qian, and F. Chen, "From Twitter to detector: Real-time traffic incident detection using social media data," *Transportation Research Part C: Emerging Technologies*, vol. 67, pp. 321–342, Jun. 2016, doi: 10.1016/j.trc.2016.02.011.

[20] Dwi Lingga P Rendra, Fatichah Chastine, and Purwitasari Diana, "Deteksi Gempa Berdasarkan Data Twitter Menggunakan Decision Tree, Random Forest, dan SVM," *JURNAL TEKNIK ITS*, vol. 6, no. 1, pp. 159–162, 2017.

[21] D. Tang, B. Qin, and T. Liu, "Document Modeling with Gated Recurrent Neural Network for Sentiment Classification," in *Empirical Methods in Natural Language Processing*, 2015, pp. 1422–1432, doi: 10.18653/v1/D15-1167.

[22] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in *Empirical Methods in Natural Language Processing*, Oct. 2014, pp. 1746–1751, doi: https://doi.org/10.3115/v1/D14-1181.

[23] W. Liao, Y. Wang, Y. Yin, X. Zhang, and P. Ma, "Improved sequence generation model for multi-label classification via CNN and initialized fully connection," *Neurocomputing*, vol. 382, pp. 188–195, Mar. 2020, doi: 10.1016/j.neucom.2019.11.074.

[24] H. Peng *et al.*, "Hierarchical Taxonomy-Aware and Attentional Graph Capsule RCNNs for Large-Scale Multi-Label Text Classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 6, pp. 2505–2519, Jun. 2021, doi: 10.1109/TKDE.2019.2959991.

[25] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent Convolutional Neural Networks for Text Classification," in *Twenty-Ninth AAAI Conference on Artificial Intelligence2267*, pp. 2267–2273, 2015.

[26] C. Zhou, C. Sun, Z. Liu, and F. C. M. Lau, "A C-LSTM Neural Network for Text Classification," 2015, doi: 10.48550/ARXIV.1511.08630.

[27] H. J. Dai, P. T. Lai, Y. C. Chang, and R. T. H. Tsai, "Enhancing of chemical compound and drug name recognition using representative tag scheme and fine-grained tokenization," *Journal of Cheminformatics*, vol. 7, 2015, doi: 10.1186/1758-2946-7-S1-S14.

[28] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural Architectures for Named Entity Recognition," in *Proceedings of NAACL-HLT*, pp. 260–270, 2016, doi: 10.48550/ARXIV.1603.01360.

[29] F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning Precise Timing with LSTM Recurrent Networks," *Journal of Machine Learning Research*, vol. 3, pp. 115–143, 2002.

[30] J. Nam, J. Kim, E. Loza Mencía, I. Gurevych, and J. Fürnkranz, "Large-Scale Multi-label Text Classification — Revisiting Neural Networks," in *Machine Learning and Knowledge Discovery in Databases*, pp. 437–452, 2014, doi: https://doi.org/10.1007/978-3-662-44851-9_28.

[31] J. Wang, L. Chen, J. Zhang, Y. Yuan, M. Li, and W. H. Zeng, "CNN transfer learning for automatic image-based classification of crop disease," in *Communications in Computer and Information Science*, vol. 875, pp. 319–329, 2018, doi: 10.1007/978-981-13-1702-6_32.

Publisher's Note: Publisher stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.