Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-09343-1

Classifying informative and non-informative tweets from the twitter by adapting image features during disaster

Sreenivasulu Madichetty[1] · **Sridevi M**[1]

Received: 12 October 2019 / Revised: 8 June 2020 / Accepted: 13 July 2020 /
Published online: 07 August 2020

Abstract
During the crisis, people post a large number of informative and non-informative tweets on Twitter. Informative tweets provide helpful information such as affected individuals, infrastructure damage, availability and resource requirements, etc. In contrast, non-informative tweets do not provide helpful information related to either humanitarian organizations or victims. Identifying informative tweets is a challenging task during the disaster. People often post images along with text on Twitter during the disaster. In addition to tweet text features, image features are also crucial for identifying informative tweets. However, existing methods use only text features but do not use image features to identify crisis-related tweets during the disaster. This paper proposes a novel approach by considering the image features along with the text features. It includes a text-based classification model, an image-based classification model and a late fusion. The text-based classification model uses the Convolutional Neural Network (CNN) and the Artificial Neural Network (ANN). CNN is used to extract text features from a tweet and the ANN is used to classify tweets based on extracted text features of CNN. The image-based classification model uses the fine-tuned VGG-16 architecture to extract the image features from the image and classify the image in a tweet. The output of the text-based classification model and the image-based classification model are combined using the late fusion technique to predict the tweet label. Extensive experiments are carried out on Twitter datasets of various crises, such as the Mexico earthquake, California Wildfires, etc., to demonstrate the effectiveness of the proposed method. The proposed method outperforms the state-of-the-art methods on various parameters to identify informative tweets during the disaster.

Keywords Artificial neural network · VGG-16 architecture · Disaster · Late fusion.

✉ Sreenivasulu Madichetty
sreea568@gmail.com

Sridevi M
msridevi@nitt.edu

[1] Department of Computer Science and Engineering, National Institute of Technology, Tiruchirappalli, Tamil Nadu 620015, India

1 Introduction

During disasters, a substantial quantity of information is published on micro-blogging platforms (like Twitter) [10, 21] in the view of the text, picture, audio, etc. The authors of [13, 27] have shown that the use of social media for rapid early damage assessment during disasters and suggested tools to identify critical social media information is very much needed. Various organizations need different types of information [48]. For example, the government or humanitarian organization needs information on the situation updates of how many people have been injured and died, how many buildings are collapsed, etc., to take the immediate action [44]. In the same way, affected people [33] need information like where medical resources are available, food resources, etc., to save their lives. Along with useful information, people post tweets which are not relevant to the disaster such as sympathies, opinions, etc. Therefore, identifying useful tweets while filtering non-useful tweets is a challenging task.

Tweets which are posted during disaster can be broadly categorized as informative and non-informative. Informative tweets [19] provide data that is either beneficial to affected people or to humanitarian organizations that need details on the needs of those impacted, wounded or dead individuals, infrastructure harm, resource accessibility, etc. It also provides information where the necessary medical resources are available to the injured victims. Non-informative tweets do not have valuable disaster-related information. The sample tweets of informative and non-informative (text and image) for various disasters are shown in Table 1 and Table 2, respectively. The first column represents the tweet number, the second column represents text and the third column represents the image of the corresponding tweet.

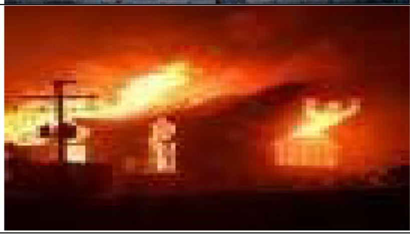
The problem of identifying informative tweets can be formulated as a text classification problem. Given a set of tweets, classify the tweets into informative and non-informative. The main objective of this paper is to develop a robust classification method (the combination of text and image features) for the automatic identification of informative tweets during the disaster. Most of the existing methods [8, 40, 42] used either handcrafted features or deep neural networks to classify crisis-related tweets during a crisis. The main drawback of methods that use handcrafted features is that they do not work correctly if their features are not present in the tweet. On the other hand, a deep neural network does not deliver the best performance if the size of the training dataset is small. To overcome these limitations, a method was used to identify informative tweets based on text features such as the combination of CNN and ANN during a disaster in our previous work [29]. The main drawback of existing text-based classification methods including our prior work [29] is explained in detail in Section 4.6, where text-based classification methods have failed. The authors in [4, 35] showed that the images posted in tweets during disaster gives the useful information. However, all methods focus mainly on text features but do not focus on image features for the classification of crisis-related tweets during disasters.

In this paper, we propose a novel method by adapting both text and image features in order to overcome the limitations of existing text-based classification methods. The proposed method takes advantage of the images posted on Twitter along with the tweet text during disaster. And, the efficiency of the model is also investigated, with and without the addition of image features.

The main contributions of this paper are summed up as follows:

1. A novel method is proposed by adapting the image features to the text-based classification model for the identification of informative tweets during a crisis.

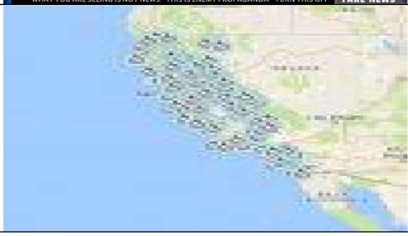
Table 1 Examples of Informative tweet with image

Tweet No.	Text	Image
1.	RT @D1MikeyT: @KPRC2 #Harvey still causing damage and flooding [URL]	
2.	Over 200 killed in Iran-Iraq earthquake [URL][URL]	
3.	7.3 MAGNITUDE EARTHQUAKE HITS IRAN-IRAQ; MANY KILLED, THOUSANDS are INJURED [URL] [URL]	
4.	11 dead, thousands homeless as wildfires torch California wine country [URL] #Heatwave #Wildfires [URL]	

2. Extensive studies are carried out on seven disaster datasets such as Hurricane Harvey, California Wildfires, Hurricane Irma, Sri Lanka floods, Hurricane Maria, Iran-Iraq Earthquake and Mexico Earthquake.
3. The proposed method is compared with current state-of-the-art models on various parameters.

It is noted that the text-based classification method was first proposed in our previous work [29]. This work extends our previous work as follows. First, a text-based classification method is experimented with a diverse disaster datasets, such as Hurricane Harvey, Sri Lanka floods, California wildfires, etc., to check the performance of the model using various parameters. Second, an in-depth analysis is carried out where the text-based classification method failed to detect informative tweets during the disaster. Third, we have developed a model by adapting image features along with text features to overcome the limitation of text-based classification methods for identifying informative tweets. Fourthly, the

Table 2 Examples of Non-informative tweet with image

Tweet No.	Text	Image
1.	Commanders of #IRGC, #Iran's Army visit #earthquake areas in Western Iran [URL]	
2.	but some ppl have their own tragic ways to show their sympathy with victims of #iran #iraq border #earthquake [URL]	
3.	RT @MarkDice: Watch CNN ignore #HurricaneHarvey to keep complaining about Trump [URL][URL]	
4.	RT @Rincon_Music: (Radio Reports On California Wildfires) -[URL] [URL]	

performance of the proposed model is compared to several recent existing methods. Finally, an error analysis is carried out where the proposed method has failed, which is very helpful in further improving the proposed method in the future.

The remaining part of the paper is structured as follows. The related work of classifying social media posts in different categories is discussed in Section 2. Section 3 explains the proposed method for identifying informative tweets during the disaster. Section 4 presents the results, the analysis of the experiments and the error analysis of the proposed model. The paper is finally concluded with the future aspects in Section 5.

2 Related work

This section describes the various methods for classifying social media posts into multiple categories during the disaster. The authors in [40, 41] discovered the distinctive features such as the low-level lexical and syntactic features for classifying tweets into situational and non-situational. After categorizing the tweets, they built a technique for summarizing the situational tweets for getting the situational awareness to the government organizations. They performed experiments on different disaster datasets such as the Hyderabad bomb blast, Uttarakhand floods, Nepal earthquake, etc., and also used two different language tweets such as English and Hindi. They used different lexicon and syntactic features for classifying situational and non-situational tweets during a disaster. The authors in [9] suggested a method for identifying various categories of tweets, such as impacted individuals, building damages, resources, etc., that would help humanitarian organizations and affected people to access resources during the disaster. However, all works are focused on text features only. Similarly, the authors in [43] developed an automatic method for identifying tweets related to viruses like Ebola and MERS. Their method uses features such as the existence of sign/symptoms terms, the existence of preventive terms, preventive procedures presence, etc. However, it is not applicable for identifying resource tweets.

The authors in [5] used information retrieval methodologies with the usage of word2vec and character embeddings to extract resource needs and availability tweets. It has shown that their methods perform well than the baseline method [38]. The authors in [31], developed a method based on the re-ranking feature selection algorithm for detecting need and availability tweets during a disaster. It uses the χ^2-statistic feature selection algorithm [12] and maximum term frequency algorithm [28]. It has shown that re-ranking feature selection works well than the other feature selection algorithm like χ^2-statistic feature selection algorithm, maximum frequency feature selection algorithm, etc. The authors in [30], designed a method by combining the statistical and informative features for the identification of tweets related to damage assessment. They have used random forest [6] and AdaBoost [15] classifiers. However, all papers addressed a distinct kind of helpful information and also focused only on text features. Identifying the informative tweets during a crisis is very crucial. The authors in [35] have developed a technique for assessing damage from images during the disaster but not focused on text features. There is no work available to identify the informative tweets based on the multi-modal (both images and text) data. In this paper, a technique is proposed by using text and image features to identify informative tweets during the disaster.

3 Methodology

The proposed model is made up of three modules, namely, 1. Text-based classification method 2. Image-based classification method and 3. Late fusion is used to combine text and image information. In the proposed framework, image features are adopted by using Visual Geometry Group (VGG-16, where 16 indicates 16 different layers) architectures [45] and improves the robustness of the proposed work. The framework of the proposed method is depicted in Fig. 1.

The proposed method is illustrated in the following steps:

1. Tweets with text and image are given as an input to the model separately.
2. Text is divided into words and converted into word vectors by using pre-trained crisis word embeddings [23].
3. Word vectors are supplied to CNN. The CNN output is fed to the ANN classifier.
4. ANN provides the output as a probability vector of the Text.
5. Images are supplied to the VGG-16, and it produces a probability vector for the images. The output of the fine-tuned VGG-16 model is used as a feature vector for classification.

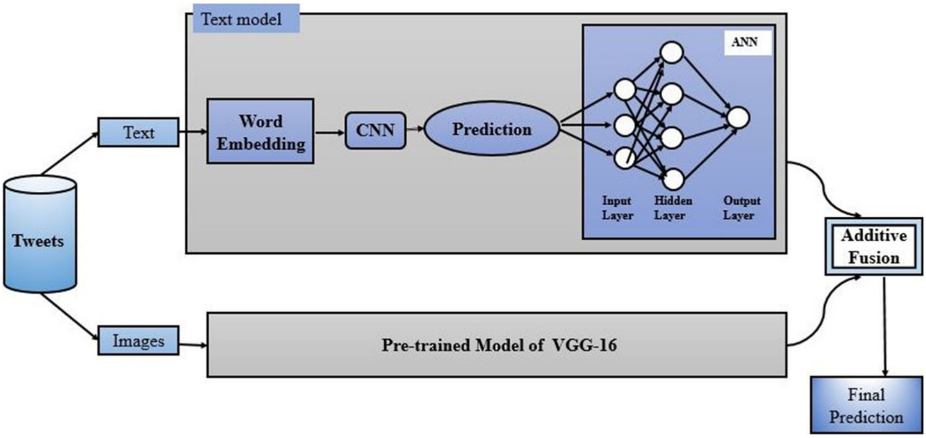


Fig. 1 The framework of the proposed method

6. The probability vector (output) of VGG-16 is combined with the output of CNN and ANN output (probability vector) using an additive fusion technique.
7. The majority class distribution of the output assigned as a class label.

The subsequent sections discuss the techniques used in the proposed method, such as CNN architecture, ANN classifier, VGG-16 architecture and late fusion.

3.1 CNN architecture

CNN architecture is well approved owing to the automatic extraction of features to solve text classification problems [8, 24] and provides more precision. It uses pre-trained word embeddings such as crisis word embedding [23], Global Vectors for word representation (GloVe embeddings) [37], etc. Pre-trained word embeddings produce a low dimensional vector of a word. Word vector is viewed as extractors of features for any problem of text classification and performance improvement.

It consists of 5 layers, such as the input layer, convolution layer, pooling layer, dense layer and output layer. Initially, tweets are given as an input to the input layer. Tweets are divided into words, and word vectors are generated for every word in a tweet by using crisis word embeddings. Let $w_k \in R^j$, where 'j' represents the length of the word vector and 'k' indicates the position of the word in a tweet. Let the tweet contains a sequence of words:$w_1, w_2, ..., w_n$ where 'n' is the number of words, and the tweet vector is obtained using (1).

$$v_{1:n} = v_1 + v_2 + v_3 + ... + v_n \tag{1}$$

where '+' represents the concatenation operator, 'v_k' is the vector of the corresponding word 'w_k'. Let '$T \times J$' be the tweet matrix, which is fed to the input layer as input, where 'T' indicates the total count of words in a tweet and 'J' is the length of the word vector. The convolution operation is performed on a tweet matrix to obtain new features by applying distinct filters of 'r' words ('r' represents filter size). Similarly, for all 'r' words window, convolutional features are calculated. Subsequently, the pooling layer receives convolutional features. The objective of the pooling layer is to achieve the decisive activation and max-pooling operation used in the architecture. Outputs are then fed into a dense layer.

Subsequently, the output of the dense layer sends to the output layer (Softmax layer) to predict the output of the tweets.

3.2 ANN classifier

ANN [32, 39] is a classifier, and it has layers like the input layer, hidden layer, and output layer, with each layer having one or more neurons. The number of input layer neurons relies on the length of the tweet's feature vector 'x'. The quantity of neurons in the hidden layer is a user-defined entry parameter. Finally, the output layer has a single neuron for binary classification problems. Neuron in the final layer is triggered by the Rectified Linear Units (ReLU) because it is used more frequently in the present days [20] and it is represented in (2) and (3). It removes the values which are less than zero that are present in the features and yields better output.

$$ReLu(x_1) = \max(0, x_1) \tag{2}$$

$$ReLu(x_1) = \begin{cases} x_1, & if x_1 > 0 \\ 0, & otherwise \end{cases} \tag{3}$$

All neurons in the ANN are connected with weighted links from one layer to another, and arbitrary values are assigned as an initial weight for training the model. The weights are adjusted by using the back-propagation algorithm [49] during the training phase. ANN provides the highest efficiency to identify informative tweets during a disaster [8].

3.3 Text-based classification method

In the existing literature [8, 34], CNN has produced promising results in the classification of social media posts, specifically for the classification of informative and non-informative tweets during a disaster without the use of handcrafted features. It outperforms traditional classification methods such as Support Vector Machine (SVM) [47] and ANN, which uses the n-gram (n=1, 2 and 3) features. Among the traditional classifiers, the ANN classifier works well than the SVM classifier. Therefore, a combination of CNN and ANN (text-based classification method) is used in this work to identify informative tweets during the disaster where CNN is used as a feature extractor and ANN is used as a classifier. CNN gives output as a feature vector for each tweet text and sends it as an input to the ANN classifier to achieve a better text classification of the model.

3.4 Image-based classification method

This section can be divided into two subsections. The first subsection describes the working of basic CNN with images, and the second subsection describes the VGG-16 architecture.

3.4.1 Working of basic CNN with images

The basic CNN model for image classification is shown in Fig. 2. It contains the Convolution layer, Pooling layer and Fully connected layer.

Convolution layer The input image represents the pixel matrix in the convolution layer. Its size is shown as the $Height \times Width \times Channel$. Each pixel range is 0 to 255, which indicates the pixel intensity at each point. The convolution layer has 'n' kernels of the size

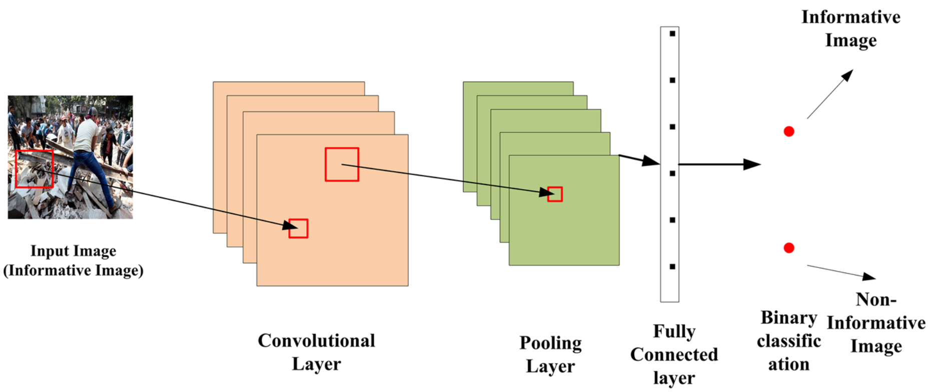


Fig. 2 The basic CNN model for image classification

$S \times S$ where S < min (height, width) and the kernel size gives the smaller matrix and contains spatially linked neighbor information of the images. It produces the output of an 'n' number of feature maps. Since the kernel reads a matrix that is a smaller part of the image, and the following parameters, such as padding (it is used to get the number of output feature map is equal number of input feature map to convolutional layer), stride (number of pixels kernel should be moved) and depth, are used to pass the kernel across the image. It gives a matrix that is equal to the input matrix. It is then passed as an input to the next layers. It acts as a filter to extract the features from the informative images.

Pooling layer This layer performs the downsampling of the features maps and captures the essential information from the informative images. It reduces the spatial size of the information, the quantity of parameters, and stops over-fitting, making the model more efficient.

Fully Connected layer It uses the soft-max activation function and extracts high-level features from images to classify images as informative and non-informative.

3.4.2 VGG-16 architecture

All input images are resized to '224 × 224', according to the VGG-16 architecture. The illustration of fine-tuned VGG-16 is shown in Fig. 3. It contains convolutional layers, max-pooling layers and fully connected layers. VGG-16 is developed by the Visual Geometry Group in oxford university, and it has 16 layers with tunable parameters, while other layers (max-pooling layer) has no tunable parameters. Every convolution layer has no max-pooling layer. Each convolutional layer has various types of filters with different sizes. The details of filters and other parameters of all the layers are shown in Table 3. The initial weight of the filters is assigned randomly. The weights are updated in the process of training the images by using the (5) to (10). However, we use a pre-trained model that is well trained from the ImageNet dataset. Using the transfer learning technique, the weights of the pre-trained VGG-16 architecture are used to initialize the disaster dataset. The last layer of the VGG-16 architecture is used to classify informative and non-informative tweets. This strategy enabled us to transfer the network's characteristics and parameters from the wide domain (i.e., large-scale picture classification) to the particular domain (i.e., catastrophe

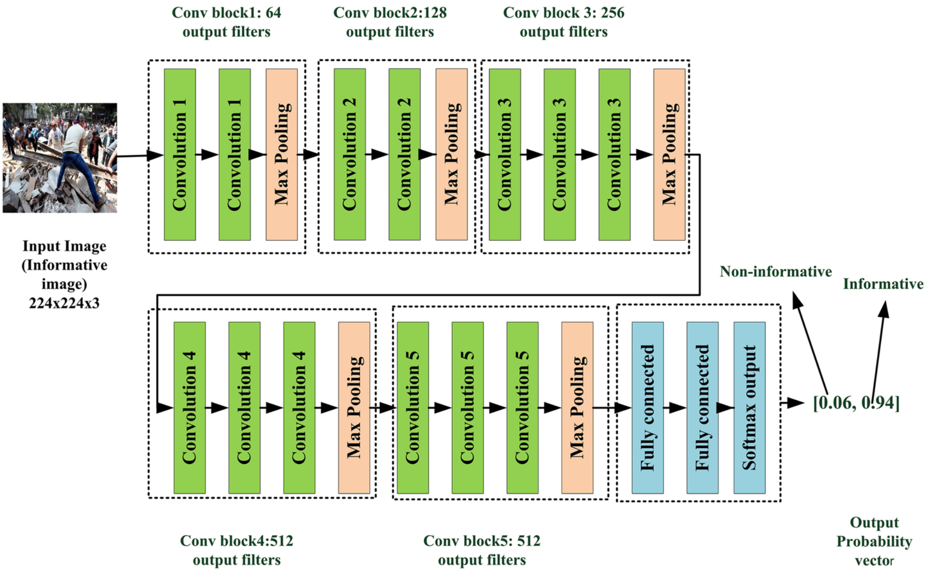


Fig. 3 Illustration of the fine-tuned VGG-16 model

picture analysis) and fine-tuned the model to suit image dataset and to get probability feature vector of the classes such as informative and non-informative class. However, the authors in [35] proved that fine-tuned VGG-16 model gives the best performance compared with state-of-the-art methods like pre-trained VGG-16 and Bag-of-visual words model in the field of disaster-related images. And also, the reason for selecting the fine-tuned VGG-16 model is explained in Section 4.3. In this work, we used a binary cross-entropy loss function, which is shown below:

$$L(w) = -\sum_{n=1}^{N}[y_n 1og(y^1_{nw})] \tag{4}$$

$$w_t = w_{t-1} - \frac{\alpha M^1_t}{sqrt V^1_t + \epsilon} \tag{5}$$

$$M^1_t = \frac{M_t}{1-\beta^t_1} \tag{6}$$

$$V^1_t = \frac{V_t}{1-\beta^t_2} \tag{7}$$

$$M_t = \beta_1.M_{t-1} + (1-\beta^t_1).g_t \tag{8}$$

$$V_t = \beta_2.V_{t-1} + (1-\beta^t_1).g^2_t \tag{9}$$

$$g_t = \nabla_w f_t(w_{t-1}) \tag{10}$$

Where y_n denotes the one-hot vector of the actual labels and y^1_{nw} is the predicted class probabilities of the VGG-16 model for the n^{th} training example in the batch of N images, γ is the multiplier of the L2 regularization term and w is the model parameters. f(w) indicates Stochastic objective function with parameters w. All values (M_0, V_0 and t) are assigned to

Table 3 The configuration parameters of fine-tuned VGG-16 model

S.No	Layers	Output Size	Filter Size	Stride	Padding
1.	Convolution layer-1	224x224x64	3	1	1
	Convolution layer-1	224x224x64	3	1	1
	Max pooling	112x112x64	2	2	0
2.	Convolution layer-2	112x112x128	3	1	1
	Convolution layer-2	112x112x128	3	1	1
	Max pooling	56x56x128	2	2	0
3.	Convolution layer-3	56x56x256	3	1	1
	Convolution layer-3	56x56x256	3	1	1
	Convolution layer-3	56x56x256	3	1	1
	Max pooling	28x28x256	2	2	0
4.	Convolution layer-4	28x28x512	3	1	1
	Convolution layer-4	28x28x512	3	1	1
	Convolution layer-4	28x28x512	3	1	1
	Max pooling	14x14x256	2	2	0
5.	Convolution layer-5	14x14x512	3	1	1
	Convolution layer-5	14x14x512	3	1	1
	Convolution layer-5	14x14x512	3	1	1
	Max pooling	7x7x512	2	2	0
6.	Full connected layer	1x1x4096	-	-	-
7.	Full connected layer	1x1x4096	-	-	-
8.	softmax output layer	1x1x 2	-	-	-

zero. $\beta_1 = 0.9$, $\beta 2 = 0.999$, $\alpha = 0.001 and \epsilon = 10^{-8}$. g_t^2 indicates the elementwise $g_t \odot g_t$, M_t, V_t, $M_t^1$and V_t^1 indicate the update baised and compute bias corrected for the first and second moment estimations.

The different layers of the pre-trained VGG-16 model extract different features from the images that are explained below:

1. The lower-level layers of the VGG-16 model extract the general features of the images, such as:
   -The first layer extracts horizontal edges, vertical edges, diagonal edges, etc.
   -The second layer extracts features such as corners, curves, etc.
   -The third layer extracts features such as lips, eyes, etc.
   -The fourth layer extracts features such as noses, face contribution, etc.
   -The fifth layer extracts features such as faces, wheels, cars, etc.
   In the same way, if we go to higher-level layers, more domain-specific features are learned, while lower layers learned are more general features.
2. Higher-level features of the VGG-16 model extract features specific to the domain, such as damaged and destroyed buildings, injured people, etc., which are very useful for the differentiation of informative and non-informative images.

The last layer of the model is the softmax output layer that gives the probability vector of the input images.

3.5 Late fusion

The probability vectors of the image-based classification model (fine-tuned VGG-16 architecture) and the text-based classification model (the combination of CNN and ANN) are merged by adding both vectors. The maximum value of the class vector is considered to be the final class label.

4 Experimental results and analysis

Open-source libraries such as sklearn package [36] and Keras library [11] are used in python language for implementation of methods. The default parameters of the sklearn package are used for the implementation of machine learning models. For the deep learning model (CNN) and the VGG-16 model, the following parameters, such as the optimization method of ADADELTA [50] and Adam [25] optimizer is used during the training of the model. To prevent the issue of over-fitting, we use the dropout [46] parameter and use 25 epochs. The batch size is 256, and early stopping criteria are used based on the validation accuracy. The datasets, evaluation metrics and performance analysis of proposed work are explained briefly in the subsequent sections.

4.1 Datasets

Datasets such as Hurricane Harvey, Hurricane Maria, Hurricane Irma, Mexico Earthquake, California Wildfires, Sri Lanka floods and Iraq-Iran earthquakes are used in the experimentation of the proposed model. Ground truth data of the aforementioned datasets are collected from [2], and it is divided into 80% of tweets for training and 20% of tweets for testing the models. However, training and testing tweets have an equal number of informative and non-informative tweets. The details of the datasets are tabulated in Table 4.

4.2 Evaluation metrics

Let T_{P1} imply the true positive that is the correct number of informative tweets. T_{N1} is the true negative, which is the number of non-informative tweets correctly identified. F_P is a false positive that defines the number of non-informative tweets identified as informative

Table 4 Details of the human-annotated disaster datasets

Dataset Name	Tweets	Images	Data collected in 2017
Hurricane Harvey	4,041	4,525	AUG 16 to Sep 20
Hurricane Maria	4,000	4,443	Sep 20 to Oct 3
Hurricane Irma	4,000	4,562	Sep 6 to Sep 19
Mexico Earthquake	1,239	1,382	Sep 20 to Oct 6
California Wildfires	1,486	1,589	Oct 10 to Oct 27
Sri Lanka Floods	832	1,025	May 31 to July 3
Iraq Earthquake	499	600	Nov 12 to Nov 19

tweets. FN is a false negative that describes the number of informative tweets identified as non-informative. Similarly, PIn, RIn and $f1In$ indicate the precision, recall and f1-score of the informative class, as well as P_{Nin}, R_{Nin} and $f1_{Nin}$, indicating the precision, recall and f1-score of the non-informative class, respectively. Let N_1 and N_2 be the total number of tweets in the informative and non-informative classes, respectively. Precision, Recall, F1-score and Accuracy metrics are calculated using Eq.s 4 to 7, respectively. Also, Eq.s 8 to 13 calculates the macro-average as well as the weighted average of precision, recall and F1-score, respectively.

$$Precision(P) = \frac{T_{P1}}{T_{P1} + F_{P1}} \tag{11}$$

$$Recall(R) = \frac{T_{P1}}{T_{P1} + F_{N1}} \tag{12}$$

$$F1 - score(F1) = \frac{2PR}{P + R} \tag{13}$$

$$Accuracy = \frac{T_{P1} + T_{N1}}{T_{P1} + F_{P1} + T_{N1} + F_{N1}} \tag{14}$$

$$Macro - Precision(M_p) = \frac{P_{In} + P_{Nin}}{2} \tag{15}$$

$$Macro - recall(M_r) = \frac{R_{In} + R_{Nin}}{2} \tag{16}$$

$$Macro - F1 - Score(M_{f1}) = \frac{f1_{In} + f1_{Nin}}{2} \tag{17}$$

$$Weighted - precision(W_p) = \frac{(P_{In} * N_1 + P_{Nin} * N_2)}{2} \tag{18}$$

$$Weighted - recall(W_r) = \frac{(R_{In} * N_1 + R_{Nin} * N_2)}{2} \tag{19}$$

$$Weighted - F1 - Score(W_{f1}) = \frac{(f1_{In} * N_1 + f1_{Nin} * N_2)}{2} \tag{20}$$

4.3 Ablation studies on the proposed method by varying the image-based classification models

Different experiments are performed by varying the pre-trained image classification models in the proposed framework. To know the strength of the VGG-16 pre-trained model, we used various pre-trained models such as AlexNet [26], ResNet-50 [17], ResNet-101 [18], Mask RCNN [16], MobileNet [20] and VGG-19 [45]. The results of the various pre-trained models in the proposed framework on Hurricane Harvey, Hurricane Maria, and Hurricane Irma are shown in Tables 5, 6 and 7. The methods in the tables indicate the model used in the proposed framework. The results show that the VGG-16 model achieves the highest performance while the AlexNet model achieves the lowest performance on three datasets for the identification of informative tweets. The VGG-19 model in the proposed framework does not have better performance than the VGG-16 model, even though it contains more layers. It is because the model learns more complex features of the image that leads to an over-fitting problem. Other models such as ResNet-50, ResNet-101 and Mask R-CNN have

Table 5 Results on different proposed models by varying the pre-trained image-based classification models on Hurricane Harvey dataset using accuracy, macro-precision (macroPre), macro-recall (macroRec) and macro-f1-score (macrof1)

Methods	Accuracy	macroPre	macroRec	macrof1
Fine-tuned AlexNet	75.45	75.23	75.23	75.36
Fine-tuned ResNet-50	75.67	75.42	75.28	75.40
Fine-tuned ResNet-101	76.02	76.00	75.88	75.90
Fine-tuned Mask R-CNN	76.14	76.18	76.14	76.20
Fine-tuned MobileNet	76.40	76.35	76.50	76.46
Fine-tuned VGG-16	**77.70**	**78.00**	**78.00**	**77.60**
Fine-tuned VGG-19	77.20	77.30	77.30	77.16

Bold indicate the highest values in the tables

little difference in performance. The MobileNet model places the next VGG model (VGG-16 and VGG-19) on three datasets. However, the VGG-16 model consistently delivers better results than other models on three different disaster datasets. It should be noted that all values in the tables are given as percentages.

4.4 Baselines

The following baselines are used for demonstrating the benefit of the proposed method:

1. Crisis Event Extraction Service (CREES) [7] is used as a baseline, and it utilizes the Convolutional Neural Network for classifying crisis-related tweets. The same model is experimented for our datasets to get the results for a comparison of the proposed method.
2. Crisis data processing services (CrisisDPS) [1] is used as another baseline, and it utilizes CNN with different filter sizes such as 2, 3, 4, 5 with a different number of filters as 100, 150, 200, 250. However, this model is experimented with for our datasets.
3. Combination of CNN and ANN [29] where CNN is used as a feature extractor, and ANN is used as a classifier to the identification of informative tweets. It is shown that the combination of CNN and ANN gives the best performance than individual methods such as CNN, ANN and SVM with n-gram features (n=1, 2 and 3) are shown

Table 6 Results on different proposed models by varying the pre-trained image-based classification models on Hurricane Maria dataset using accuracy, macro-precision (macroPre), macro-recall (macroRec) and macro-f1-score (macrof1)

Methods	Accuracy	macroPre	macroRec	macrof1
Fine-tuned AlexNet	69.46	69.02	69.02	69.28
Fine-tuned ResNet-50	69.53	69.03	68.75	68.99
Fine-tuned ResNet-101	69.60	69.56	69.32	69.36
Fine-tuned Mask R-CNN	69.84	69.36	69.32	69.44
Fine-tuned MobileNet	70.90	69.97	70.00	70.56
Fine-tuned VGG-16	**72.96**	**73.00**	**73.00**	**72.84**
Fine-tuned VGG-19	71.96	71.60	71.60	71.96

Bold indicate the highest values in the tables

Table 7 Results on different proposed models by varying the pre-trained image-based classification model on Hurricane Irma dataset using accuracy, macro-precision (macroPre), macro-recall (macroRec) and macro-f1-score (macrof1)

Methods	Accuracy	macroPre	macroRec	macrof1
Fine-tuned AlexNet	71.37	71.07	71.07	71.20
Fine-tuned ResNet-50	71.43	70.98	70.95	70.98
Fine-tuned ResNet-101	71.46	71.43	71.28	71.31
Fine-tuned Mask R-CNN	71.63	71.45	71.41	71.48
Fine-tuned MobileNet	72.37	71.87	71.87	71.95
Fine-tuned VGG-16	**73.82**	**74.00**	**74.00**	**73.55**
Fine-tuned VGG-19	73.12	73.02	73.02	72.93

Bold indicate the highest values in the tables

in Figure 4. Therefore, a combination of CNN and ANN is considered as one of the baselines for comparing the proposed model.

4. The authors in [3] reported the results for the datasets such as Hurricane Harvey, Hurricane Maria, and Hurricane Irma for classifying the disaster-related tweets into various categories.
5. The authors in [14] used and reported the results of the methods such as BiGRU, BiLSTM, CNN GRU, CNN LSTM, DPCNN, KMax CNN, RCNN and fine-tuned BERT model for the datasets such as Hurricane Harvey, Hurricane Maria, and Hurricane Irma.

4.5 Performance analysis

The performance of the proposed method is analyzed against the various state-of-the-art methods referred to in Section 4.4, along with methods such as SVM, ANN and CNN similar to [8]. The method inscriptions are shown in Table 8.

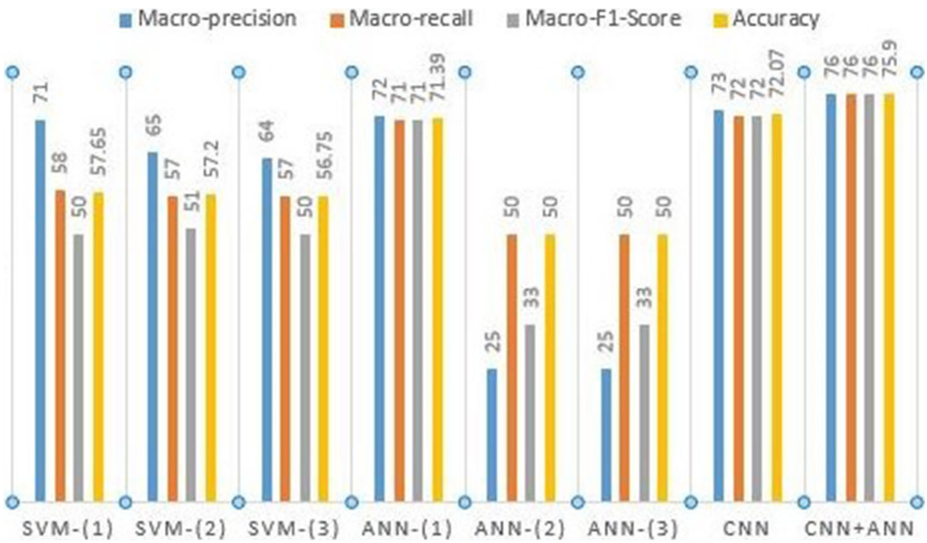


Fig. 4 Comparison of combination of CNN and ANN against state-of-the-art methods

Table 8 Inscriptions

Methods	Abbreviations
SVM-(1)	SVM classifier with uni-gram features.
SVM-(2)	SVM classifier with n-gram features where n=1 and 2.
SVM-(3)	SVM classifier with n-gram (n=1, 2 and 3)features.
ANN-(1)	ANN classifier with uni-gram features.
ANN-(2)	ANN classifier with n-gram features where n=1 and 2.
ANN-(3)	ANN classifier with n-gram (n=1, 2 and 3)features.

Figure 4 shows the comparison result of the combination of CNN and ANN against state-of-the-art methods. It is noticed that the performance has been degraded after adding bi-gram features to the uni-gram features in the SVM and ANN classifier. The reason for performance degradation is due to the not occurrence of bi-gram features in the disaster tweets as well as uni-gram features that have become negligible. Additional experiments are performed by adding tri-gram features to SVM and ANN classifiers. From the outcomes, it is ensured that there is no enhancement in the classifiers (SVM and ANN) for disaster tweets by enhancing the contiguous word feature sequence (by increasing the n value in n-grams). Therefore, experiments are not persisted in the classifier for an additional adjacent sequence of words used as features. Both ANN and SVM classifiers with uni-gram features have the highest output compared to the other features (bi-gram and tri-gram). If a comparison is made between the classifiers of ANN and SVM, the ANN classifier has higher performance than the classifier of SVM in all parameters. Thus, ANN is used for additional experiments. CNN overcome the drawbacks of SVM and ANN classifier. CNN performs better performance than the ANN classifier. The reason is that CNN learns the informative n-gram features automatically from crisis word embeddings. The primary benefit of CNN is that it automatically learns the feature from the training tweets without using any feature engineering. However, if the training is not performed correctly, then there is a chance to reduce the efficiency of the model. Therefore, a method is adopted using the combination of CNN and ANN (text-based classification model) to improves the effectiveness of the model. The key idea of this method is the CNN output used as a feature that is sent to the ANN classifier to train the model. The proposed text-based classification model is compared against existing methods such as SVM, ANN and CNN. From the results, it is observed that our text-based classification model outperform the state-of-the-art methods to detect the informative tweets during a crisis.

The efficiency of the proposed method is compared against the existing methods in various parameters such as macro-precision, macro-recall, macro-f1-score, and accuracy. The comparison results are reported in Tables 9 and 10. Weighted-recall, weighted-precision, and weighted-f1-score had the same values across all the datasets due to the use of the same number of informative and non-informative tweets. The proposed method outperforms the existing text feature-based methods. It is noticed that by adding image features to the model, the model is improved by 2 to 6 percent in most of the datasets. Specifically, there is a significant impact on the performance of the model for earthquake datasets like Mexico and Iran Earthquake. Other datasets like Hurricane Harvey, Hurricane Maria, Hurricane Irma, and California wildfires have improved the performance of the proposed model. The reason behind the improved performance of the proposed method compared to the text-based classification method is that most of the images in the informative tweets

Table 9 Comparison of the proposed method with different existing methods for different disaster datasets using Accuracy and macro-F1-score parameters

Dataset	Model	Accuracy	F1-Score	Year
	ANN with unigram features [22]	71.90	71.00	2014
	CNN [8]	72.07	72.00	2016
	CREES [7]	73.65	74.20	2018
	CrisisDPS [1]	74.60	75.00	2019
	Combination of CNN and ANN [29]	75.9	76	2019
	Alam et al. [3]	-	64.10	2019
	BiGRU [14]	67.88	-	2020
Hurricane Harvey	BiLSTM [14]	71.02	-	2020
	CNN GRU [14]	67.63	-	2020
	CNN LSTM [14]	58.09	-	2020
	DPCNN [14]	63.79	-	2020
	KMax CNN [14]	66.50	-	2020
	RCNN [14]	71.77	-	2020
	Fine-tuned BERT [14]	75.37	-	2020
	Proposed Method	**77.70**	**77.60**	**2020**
	ANN with unigram features [22]	50.00	33.33	2014
	CNN [8]	62.79	61.74	2016
	CREES [7]	68.66	69.10	2018
	CrisisDPS [1]	69.60	70.00	2019
	Combination of CNN and ANN [29]	70.78	70.64	2019
	Alam et al. [3]	-	64.10	2019
	BiGRU [14]	67.88	-	2020
Hurricane Maria	BiLSTM [14]	71.02	-	2020
	CNN GRU [14]	67.63	-	2020
	CNN LSTM [14]	58.09	-	2020
	DPCNN [14]	63.79	-	2020
	KMax CNN [14]	66.50	-	2020
	RCNN [14]	71.77	-	2020
	Proposed Method	**72.96**	**72.84**	**2020**
	ANN with unigram features [22]	64.39	64.38	2014
	CNN [8]	66.75	66.32	2016
	CREES [7]	69.70	69.60	2018
	CrisisDPS [1]	70.50	70.00	2019
	Combination of CNN and ANN [29]	71.20	71.00	2019
	Alam et al. [3]	-	64.10	2019
	BiGRU [14]	67.88	-	2020
	BiLSTM [14]	71.02	-	2020
Hurricane Irma	CNN GRU [14]	67.63	-	2020
	CNN LSTM [14]	58.09	-	2020
	DPCNN [14]	63.79	-	2020
	KMax CNN [14]	66.50	-	2020
	RCNN [14]	71.77	-	2020

Table 9 (continued)

Dataset	Model	Accuracy	F1-Score	Year
	Proposed Method	**73.82**	**73.55**	**2020**
Mexico Earthquake	ANN with unigram features [22]	64.13	64.04	2014
	CNN [8]	71.72	71.61	2016
	CREES [7]	64.65	64.20	2018
	CrisisDPS [1]	64.30	64.00	2019
	Combination of CNN and ANN [29]	65	65	2019
	Proposed Method	**74.29**	**74.25**	**2020**
California Wildfires	ANN with unigram features [22]	64.13	64.04	2014
	CNN [8]	62.31	60.93	2016
	CREES [7]	57.36	57.40	2018
	CrisisDPS [1]	57.80	58.00	2019
	Combination of CNN and ANN [29]	58.70	58.50	2019
	Proposed Method	**65**	**64**	**2020**
Sri Lanka Floods	ANN with unigram features	91.00	91.46	2014
	CNN [8]	90.54	90.73	2016
	CREES [7]	90.65	90.02	2018
	CrisisDPS [1]	91.06	91.00	2019
	Combination of CNN and ANN [29]	**92.46**	**92.40**	2019
	Proposed Method	91.78	92.00	**2020**
Iraq-Iran Earthquake	ANN with unigram features	50.00	33.33	2014
	CNN [8]	60.45	60.48	2016
	CREES [7]	59.16	59.10	2018
	CrisisDPS [1]	60.60	60.00	2019
	Combination of CNN and ANN [29]	61.36	61.00	2019
	Proposed Method	**68.18**	**67**	**2020**

provide disaster-related information. And also, image features have an added advantage that can differentiate between informative and non-informative tweets where text-based classification methods have failed. But there is no impact on Sri Lanka floods, because the dataset is small, and text-based classification methods deliver high efficiency compared to image-based classification methods, so that image features become negligible. Furthermore, the images in the tweets may not provide information related to the disaster in Sri Lanka Floods.

4.6 Error analysis

We looked at the classification results of the proposed model and the text-based classification model on the test tweets with the actual tweet labels to analyze the proposed model where it was successful and failed. The combination of CNN and ANN is considered a text-based classification model because it gives a better result compared to other existing models. The reasons for the misclassification of the text-based classification model and proposed model are described below, followed by an example tweet with images, followed by an explanation:

Table 10 Comparison of the proposed method with different existing methods for different disaster datasets using macro-precision(macroPre) and macro-recall (macroRec) parameters

Dataset	Model	macroPre	macroRec	Year
Hurricane Harvey	ANN with unigram features [22]	72.00	71.00	2014
	CNN [8]	73.00	72.00	2016
	CREES [7]	74.20	74.20	2018
	CrisisDPS [1]	75.00	75.00	2019
	Combination of CNN and ANN [29]	76.00	76.00	2019
	Alam et al. [3]	67.30	64.20	2019
	Proposed Method	**78.00**	**78.00**	**2020**
Hurricane Maria	ANN with unigram features [22]	25.00	50.00	2014
	CNN [8]	64.36	62.79	2016
	CREES [7]	69.10	69.10	2018
	CrisisDPS [1]	70.00	70.00	2019
	Combination of CNN and ANN [29]	71	71	2019
	Alam et al. [3]	67.30	64.20	2019
	Proposed Method	**73.00**	**73.00**	**2020**
Hurricane Irma	ANN with unigram features [22]	64.42	64.39	2014
	CNN [8]	67.64	66.75	2016
	CREES [7]	69.6	69.6	2018
	CrisisDPS [1]	70.00	70.00	2019
	Combination of CNN and ANN [29]	71.00	71.00	2019
	Alam et al. [3]	67.30	64.20	2019
	Proposed Method	**74**	**74**	**2020**
Mexico Earthquake	ANN with unigram features [22]	64.27	64.13	2014
	CNN [8]	72.07	71.72	2016
	CREES [7]	64.20	64.20	2018
	CrisisDPS [1]	64.00	64.00	2019
	Combination of CNN and ANN [29]	65	65	2019
	Proposed Method	**74.50**	**74.00**	**2020**
California Wildfires	ANN with unigram features [22]	63.04	63.04	2014
	CNN [8]	64.35	62.31	2016
	CREES [7]	57.40	57.40	2018
	CrisisDPS [1]	58.00	58.00	2019
	Combination of CNN and ANN [29]	58.56	59.00	2019
	Proposed Method	**65.50**	**64.50**	**2020**
Sri Lanka Floods	ANN with unigram features	91.10	91.10	2014
	CNN [8]	90.40	90.40	2016
	CREES [7]	90.20	90.20	2018
	CrisisDPS [1]	91.00	91.00	2019
	Combination of CNN and ANN [29]	92.50	92.50	2019
	Proposed Method	**92.00**	**91.50**	**2020**
Iraq-Iran Earthquake	ANN with unigram features [22]	25.00	50.00	2014
	CNN [8]	60.27	60.45	2016
	CREES [7]	59.10	59.10	2018
	CrisisDPS [1]	60.00	60.00	2019
	Combination of CNN and ANN [29]	61.50	61.50	2019
	Proposed Method	**71.00**	**68**	**2020**

1. Presence of question marks where the content relevant to the informative, but the message is not informative.

   Example: @$Harry\_Styles$ wait... wasn't Mexico City suffered from an earthquake last week ? Just asking [URL].

   Image : 

   Explanation: The ground truth label of the tweet is non-informative but predicted as informative by the text-based classification model. This is due to the majority of the content in the tweet that has been revealed as informative. But with the presence of question marks, tweets become non-informative. However, the proposed method successfully detects these types of tweets using the model's image features.

2. Presence of disaster-related terms in non-informative tweets.

   Example tweet: Lloydas braces for hurricane profit damage after Harvey, Irma and Maria [URL] [URL].

   Image: 

   Explanation: The tweet is labeled as non-informative but is detected as informative by the text-based classification model. This is due to the presence of disaster-related terms in the non-informative tweets. It is overcome by capturing the image features in the proposed model. The proposed model correctly identifies the tweet as non-informative.

3. Presence of general terms along with the presence of rare hash-tags.

   Example tweet: Mexico earthquake: rescuers work into night to save trapped #TIS-News Click Link- [URL][URL].

   Image: 

   Explanation: Some of the tweets contain general terms in both informative and non-informative tweets. The best example of a tweet is mentioned in the above, and the actual label is informative. This type of tweet is correctly detected as informative by the proposed model using the image features where the text-based classification model failed and was detected as non-informative. This is due to the additional benefit of the image features for predicting this type of tweet.

4. The dataset size is tiny.

   Example tweet: 37 new #SriLanka SATHOSA branches next! - [URL] #lka #floods [URL].

   Image: 

   Explanation: The above example tweet is incorrectly detected by the proposed model when the model is trained with a small amount of dataset (Sri Lanka floods), especially for images. However, we have enough dataset size (other datasets) to implement. In the future, this problem can be resolved by developing cross-domain methods.

5. Presence of the Non-informative image in the informative tweet and vice versa.

   Example tweet: GammaddaV - Flood Relief & amp; Clean Up Campaign. Selected by Dr.S.Priyantha-Sri Lanka. [URL] [URL]

   Image: 

   Explanation: In fact, the above example tweet is labeled as informative, but the image is not informative. In most cases, the proposed model did not identify these types of tweets because the image does not provide any useful information to the model and has an incorrect influence on the text-based classification model. However, very rarely, these types of tweets existed.

6. The length of the informative tweets is minimal.

   Example tweet: Last week, I was on a boat! #Harveyflood #Harvey [URL].

   Image: 

   Explanation: The example tweet is labeled as informative but detected as non-informative by the proposed method. It is because the length of the tweet is very small, and the confidence level of the information is also very low. The image features cannot influence this type of tweet.

It is noted that the 4, 5, and 6 example tweets are wrongly detected not only by the proposed model but also most of the existing methods.

5 Conclusion

A novel method has been developed by adapting image features to detect informative tweets during a disaster based on CNN, ANN, fine-tuned VGG-16 architecture, and late fusion. The proposed method achieves the 77.70%, 72.96%, 73.82%, 74.29%, 65%, 91.78%, and 68.18% accuracy for the Hurricane Harvey, Hurricane Irma, Hurricane Maria, Mexico Earthquake, California wildfires, Sri Lanka floods, and Iran-Iran Earthquake datasets, respectively. It is noted that the proposed method delivers greater efficiency than the text-based classification methods and also performs well than the existing state-of-the-art methods by 1.19% to 6.82% more accuracy in most of the datasets. It accomplished better efficiency in distinct parameters. In future work, the focus may be on detecting informative multi-modal data (both text and image) during a disaster.

References

1. Alam F, Imran M, Ofli F (2019) Crisisdps: Crisis data processing services. In: Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management (ISCRAM)
2. Alam F, Ofli F, Imran M (2018) Crisismmd: Multimodal twitter datasets from natural disasters. In: AAAI Conference on Web and Social Media (ICWSM), AAAI, AAAI, Stanford, California, USA
3. Alam F, Ofli F, Imran M (2019) Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of hurricanes Harvey, Irma, and Maria. Behaviour & Information Technology, pp 1–31
4. Alam F, Ofli F, Imran M (2020) Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of hurricanes harvey, irma, and maria. Behaviour & Information Technology 39(3):288–318
5. Basu M, Shandilya A, Khosla P, Ghosh K, Ghosh S (2019) Extracting resource needs and availabilities from microblogs for aiding post-disaster relief operations. IEEE Transactions on Computational Social Systems 6(3):604–618
6. Breiman L (2001) Random forests. Machine learning 45(1):5–32
7. Burel G, Alani H (2018) Crisis event extraction service (crees)-automatic detection and classification of crisis-related content on social media. In: Proc. of the 15th ISCRAM
8. Caragea C, Silvescu A, Tapia AH (2016) Identifying informative messages in disaster events using convolutional neural networks. In: International Conference on Information Systems for Crisis Response and Management, pp 137–147
9. Caragea C, Squicciarini AC, Stehle S, Neppalli K, Tapia AH (2014) Mapping moods: Geo-mapped sentiment analysis during hurricane sandy. In: ISCRAM
10. Castillo C (2016) Big crisis data: social media in disasters and time-critical situations. Cambridge University Press
11. Chollet F, et al. (2015) Keras. https://keras.io
12. Dunning T (1993) Accurate methods for the statistics of surprise and coincidence. Computational Linguistics 19(1):61–74. https://www.aclweb.org/anthology/J93-1003
13. Enenkel M, Saenz SM, Dookie DS, Braman L, Obradovich N, Kryvasheyeu Y (2018) Social media data analysis and feedback for advanced disaster risk management. CoRR abs/1802.02631 arXiv:1802.02631
14. Fan C, Wu F, Mostafavi A (2020) A hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters. IEEE Access
15. Freund Y, Schapire RE et al (1996) Experiments with a new boosting algorithm. In: Icml, Bari, Italy, vol. 96, pp 148–156

16. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp 2961–2969
17. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
18. He K, Zhang X, Ren S, Sun J (2016) Identity mappings in deep residual networks. In: European conference on computer vision, pp 630–645. Springer
19. Houston JB, Hawthorne J, Perreault MF, Park EH, Goldstein Hode M, Halliwell MR, Turner McGowen SE, Davis R, Vaid S, McElderry JA, et al. (2015) Social media and disasters: a functional framework for social media use in disaster planning, response, and research. Disasters 39(1):1–22
20. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications. CoRR abs/1704.04861, pp 1–9
21. Imran M, Castillo C, Diaz F, Vieweg S (2015) Processing social media messages in mass emergency: A survey. ACM Computing Surveys (CSUR) 47(4):67
22. Imran M, Castillo C, Lucas J, Meier P, Vieweg S (2014) Aidr: Artificial intelligence for disaster response. In: Proceedings of the 23rd International Conference on World Wide Web, pp 159–162. ACM
23. Imran M, Mitra P, Castillo C (2016) Twitter as a lifeline: Human-annotated twitter corpora for nlp of crisis-related messages. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). European Language Resources Association (ELRA), Paris, France, pp 1638–1643
24. Kim Y (2014) Convolutional neural networks for sentence classification
25. Kingma DP, Ba J (2015) Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings
26. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
27. Kryvasheyeu Y, Chen H, Obradovich N, Moro E, Van Hentenryck P, Fowler J, Cebrian M (2016) Rapid assessment of disaster damage using social media activity. Science advances 2(3):e1500779
28. Luhn HP (1957) A statistical approach to mechanized encoding and searching of literary information. IBM Journal of research and development 1(4):309–317
29. Madichetty S, Sridevi M (2019) Detecting informative tweets during disaster using deep neural networks. In: 2019 11th International Conference on Communication Systems & Networks (COMSNETS), pp 709–713. IEEE
30. Madichetty S, Sridevi M (2019) Disaster damage assessment from the tweets using the combination of statistical features and informative words. Social Network Analysis and Mining 9(1):42
31. Madichetty S et al (2018) Re-ranking feature selection algorithm for detecting the availability and requirement of resources tweets during disaster. International Journal of Computational Intelligence & IoT, 1(2)
32. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics 5(4):115–133
33. Mitchell JT, Thomas DeborahSK, Hill AA, Cutter SL (2000) Catastrophe in reel life versus real life: Perpetuating disaster myth through hollywood films. International Journal of Mass Emergencies and Disasters 18(3):383–402
34. Nguyen DT, Al Mannai KA, Joty S, Sajjad H, Imran M, Mitra P (2017) Robust classification of crisis-related data on social networks using convolutional neural networks. In: Eleventh International AAAI Conference on Web and Social Media, pp 632–635
35. Nguyen DT, Ofli F, Imran M, Mitra P (2017) Damage assessment from social media imagery data during disasters. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, pp 569–576. ACM
36. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine learning in Python. J Mach Learn Res 12:2825–2830
37. Pennington J, Socher R, Manning CD (2014) Glove: Global vectors for word representation. In: Empirical Methods in Natural Language Processing (EMNLP), pp 1532–1543. http://www.aclweb.org/anthology/D14-1162
38. Purohit H, Castillo C, Diaz F, Sheth A, Meier P (2014) Emergency-relief coordination on social media: Automatically matching resource requests and offers. First Monday, 19(1)
39. Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review 65(6):386
40. Rudra K, Ganguly N, Goyal P, Ghosh S (2018) Extracting and summarizing situational information from the twitter social media during disasters. ACM Transactions on the Web (TWEB) 12(3):17

41. Rudra K, Ghosh S, Ganguly N, Goyal P, Ghosh S (2015) Extracting situational information from microblogs during disaster events: a classification-summarization approach. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp 583–592. ACM
42. Rudra K, Sharma A, Ganguly N, Ghosh S (2016) Characterizing communal microblogs during disaster events. In: Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on, pp 96–99. IEEE
43. Rudra K, Sharma A, Ganguly N, Imran M (2017) Classifying information from microblogs during epidemics. In: Proceedings of the 2017 International Conference on Digital Health, pp 104–108. ACM
44. Sarter NB, Woods DD (1991) Situation awareness: A critical but ill-defined phenomenon. Int J Aviat Psychol 1(1):45–57
45. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: 3rd international conference on learning representations, ICLR, pp 1–14
46. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 15(1):1929–1958
47. Vapnik V (2013) The nature of statistical learning theory. Springer science & business media
48. Vieweg S, Castillo C, Imran M (2014) Integrating social media communications into the rapid assessment of sudden onset disasters. In: International Conference on Social Informatics, vol 8851. Springer, pp 444–461
49. Werbos P (1974) Beyond regression: New tools for prediction and analysis in the behavior science. Unpublished Doctoral Dissertation, Harvard University
50. Zeiler MD (2012) Adadelta: an adaptive learning rate method. arXiv preprint arXiv:1212.5701