

# Natural disaster detection in social media and satellite imagery

*Stuti Ahuja<sup>1</sup>, Melvina Michael<sup>1,\*</sup>, Malvika Selvan<sup>1</sup> and Vaishnavi Mantri<sup>1</sup>*

<sup>1</sup>Department of Information Technology, SIES Graduate School of Technology, Navi Mumbai, India

**Abstract.** Natural disasters caused by natural processes may lead to significant losses in terms of property and human lives. The timely collection of information about the damage caused by natural disasters is very important and can help reduce losses and speed recovery. Social media has become an important source of information for communication and dissemination of information in emergencies. Under such circumstances, inferring disaster events through the information available in social media will be very useful. Satellite data has also been widely used to analyze the impact of natural disasters on the surface of the earth. In this paper, a detailed analysis of how social media and satellite imagery can be used to detect natural disasters is discussed.

## 1 Introduction

Natural disasters are fatal events caused by natural or natural processes in the world. The severity of disasters is measured in terms of lost lives, financial losses, and thus the ability of the population to rebuild. All natural disasters cause damage and loss to the environment and the people who live there. Natural disasters include floods, earthquakes, tsunamis, landslides, volcanic eruptions, and hurricanes. An example of a flood-affected area is shown in fig.1.

The degree and extent of damage can be further divided into three categories.

Minor disasters: Minor disasters are disasters that extend over 50km or up to 100 km. Fires can be counted as minor disasters.



1 Corresponding author: [pushpammit118@gst.sies.edu.in](mailto:pushpammit118@gst.sies.edu.in)

**Fig. 1.** Flood-affected area

Medium-scale disasters: Medium-scale disasters range from 100km to 500km. These do more damage than a minor catastrophe. Erosion, landslides, tornados, etc can be considered medium-scale disasters.

Catastrophes: These disasters cover an area of more than 1000 km. These cause the most serious environmental damage. Moreover, at a high level, these disasters can even involve the country. Earthquakes, floods, tsunamis, etc can be considered a large-scale disasters.

In such disaster events, satellite imagery and social media emerged as key sources of information-gathering tools. Most of the times researchers and disaster relief agencies focus on these sources for information. Satellite images serve as a practical solution to analyze the impact created by disaster over the surface of the Earth. Remote sensing data also has some challenges since sometimes it has the low temporal frequency of satellite images, and oftentimes it only gives a bird's-eye view of an event. For social media we divide the detection into two parts -i) Text ii) Images.

For both the above detection we use texts and images from Twitter. But the relevance and credibility of content shared via social media are one of the biggest challenges involved. For the above detection, we use three modules mainly data extraction, sorting, and analysis.

## 2 Literature Survey

We examined and read the research papers listed below to gain additional information and ideas for our project's execution.

Naina Said, Kashif Ahmad, Michael Riegler, Konstantin Pogorelov et al. [1] proposed natural disaster data extraction from social media platform(Twitter) and satellite imagery; Disaster Detection in Satellite Imagery CNN-based framework for the detection of natural disasters in satellite imagery, MediaEval for flood detection, RF classifier is used for the classification of the satellite image patches into flooded and non-flooded regions and VGGNet model. The next one is Data Extraction from Twitter using Geo-tagged filter, K-Nearest Neighbour (K-NN) algorithm, Natural Language Processing (NLP) Bayesian approach for the identification and classification of disaster-related tweets, Support Vector Machines (SVMs), Ashktorab et al proposed Tweed, a text-mining tool for the extraction of useful information from tweets during natural disasters. Imran et al. proposed a platform, called AIRD. The last method Disaster Detection in images from Social Media, User tags, geolocation has been proved effective, HSV low-level color feature, SPCPE, Textual features are extracted through word frequency and CNN models on IMAGENET are used as the feature descriptor.

Sheharyar Ahmad1, Kashif Ahmad, Nasir Ahmad, Nicola Conci et al. [2] proposed A system called "JORD" has been introduced to automatically collect information from various social media platforms and combine it with remote sensing data to provide a more detailed idea of a disaster and they have also discussed CNN, Imagenet.

Olga Ostroukhova, Pal, Halvorsen, Nicola Conci, Rozenn Dahyot et al [3] proposed an active learning framework intending to collect, filter, and analyze social media contents for natural disasters, For data collection, a publicly available system, namely AIDR, has been used to crawl social media platforms a crowd-sourcing activity for data annotation.

Angela Maria Vinod Dharathi Venkatesh Dishti Kundra Jayapandian N [4] focused on using cognitive computing for industrial purposes and this study examined the proposed predictive models, specifically using ANN (Artificial Neural Networks), sentiment models, and smart disaster prediction application (SDPA) to forecast the flash flood.

Jigar Doshi Saikat Basu Guan Pang et al [5] proposed a framework using CNN to detect which areas were most affected by disaster, and they also used the Disaster Impact Index to quantify the impact of two natural disasters.

Siti Nor Khuzaimah Binti Amit Soma Shiraishi Tetsuo Inoshita et al [6] proposed an automatic disaster detection system by implementing one of the advance deep learning techniques, convolutional neural network (CNN), to analyze satellite images. They created their own disaster detection training data patches. The results reveal an accuracy of 80%~90% for disaster detection.

The results which were obtained presented may help in facilitating detecting natural disasters efficiently by developing an automatic disaster detection system.

### 3 Existing System

The existing system categories disaster detection of available content into three groups: I disaster detection in text, (ii) analysis of disaster-related content from social media images. (iii) Detection of disasters using satellite imagery

In the traditional system, the relevance and authenticity of content shared via social media is always a massive obstacle for applications aimed at disaster detection and analysis of disaster-related data available in social media. It lacks large-scale annotated datasets for training and evaluating machine learning techniques for disaster analysis in tweets, images from social media, and satellite imagery. Due to the low temporal frequency of satellite imagery, satellite imagery is not frequently available. Furthermore, the quality of the satellite imagery may be compromised.

### 4 Proposed System

We have proposed a model wherein three different factors are considered for the analysis of how data from satellite and social media can play an important role in disaster situations Communication and dissemination of messages. The three approaches taken into consideration are i)Detection from Satellite imagery ii) Detection from Social media text iii) Detection from Social media images.

#### *Dataset:*

Three different datasets are used for the detection of natural disasters.

i) Satellite Imagery: Here we have a set of images consisting of pair of satellite images wherein the first image is taken before the disaster event and the next is taken after the incident.

ii) Social Media Text: Here we have text and location related to disaster events.

iii) Social Media Images: We have a set of 5 different disaster-related data, which are flood, earthquake, volcanic eruption, fire, and some random images.

For all three groups, the data is divided into training and testing data.

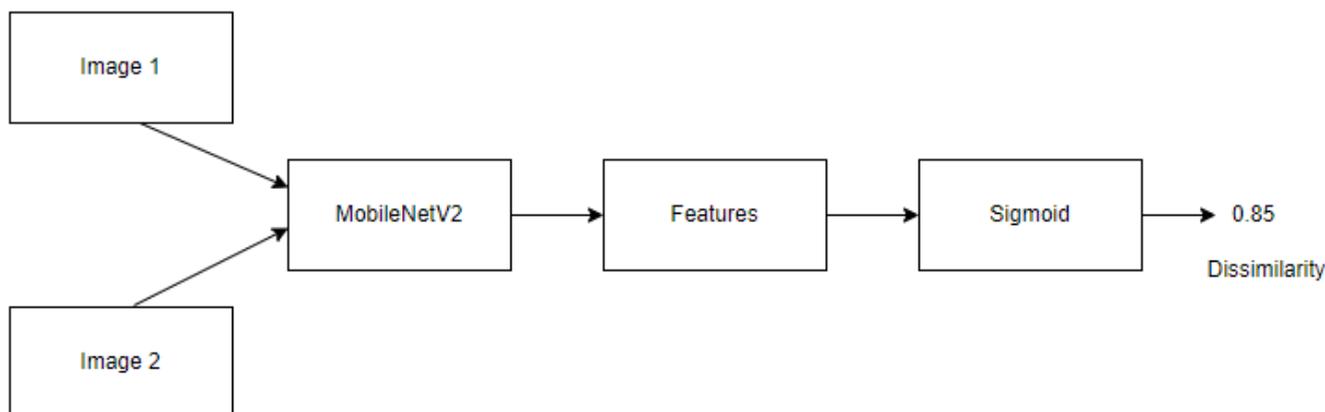
The proposed system have the following modules listed below: data extraction and analysis. The data extraction/collection is responsible for data collection and pre-processing (cleaning the retrieved data). The next module analyzes the data to extract useful information about the disaster scope, distribution, Geo-tagging, occurrence

frequency, and sentiment rating. The approach proposed relies on the use of matching keywords and hashtags for the identification of relevant messages from the social media stream using Natural language processing (NLP) trained on a diverse collection of linguistic characteristics including raw and lexical textual features. For extracting images from social media, we have used the CNN algorithm, wherein we have used the datasets, in which we have samples of 5 categories they are: earthquake, flood, fire, volcanic eruption, and some random images. We are training our model in such a way that it predicts the disaster from the image provided to test. For satellite imagery, the main aim is to find the dissimilarity between a pair of satellite images of the same region i.e. flooded and non- flooded. If the dissimilarity between the provided images is high then it means, that particular region is flooded else the region is non-flooded.

Let's look at the above three approaches:

#### A. Satellite Imagery

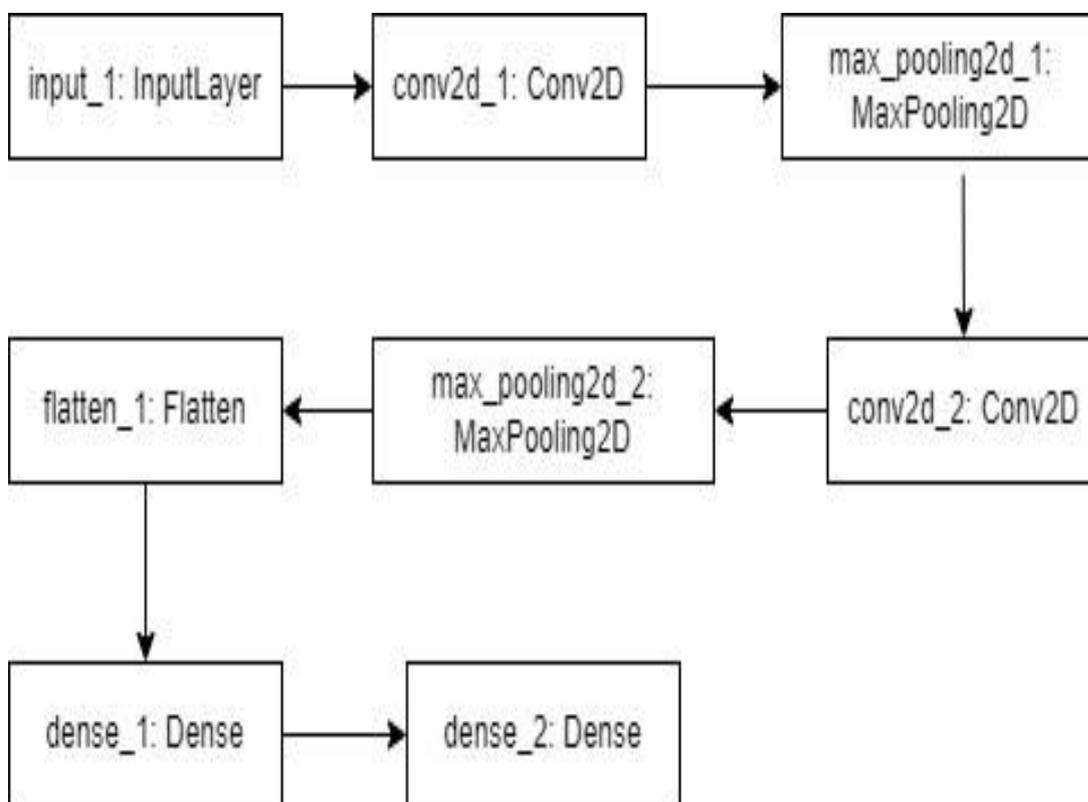
The motive here is to find the dissimilarity between a pair of satellite images of the same particular region which are taken before and after the incident. The input taken here is as a pair of images, in which the first image is taken before the incident occurred and the other image is taken after the incident. Now both the images are compared using the Siamese network as shown in fig. 2. Siamese network is a class of neural networks containing two or more sub-network which are identical. Siamese network is trained using PyTorch, Considering the pair of images, the first image is passed through the network and then the second image is passed and then the loss is calculated considering the output coming from the first and second image and the gradient is calculated. To test the model the images are passed through the network and the distance is calculated from the output and then the sigmoid is calculated. Then dissimilarity is calculated, if it's high then the result is declared as the region is flooded and if the dissimilarity is low then the region is declared as not flooded.



**Fig. 2.** Flowchart for Disaster detection using Satellite Imagery

*A. Social Media Images*

For social media images, we have used the CNN algorithm and Keras call-backs a Python interface for artificial neural networks. CNN is a deep learning algorithm that consists of convolution layers for extracting features from the given network as shown in fig. 3. The approach is that we have taken as a set of 5 datasets wherein there are disaster-related images and they are as follows flood, earthquake, fire, volcanic eruption, and random images. Firstly we get a sample batch of 25 images and train the images in epochs. We have used batch normalization and dropouts to avoid overfitting. The Keras callbacks control the learning rate of the model. After the model gets trained to the dataset, we provide the image to the model and it predicts whether it is a disaster image or not along with the name and accuracy. The feature we have used here is the early stopping method wherein we can set the number of epochs and we can see that whether the learning rate is improved or not then we can stop the model at that particular stage. The learning rate is reduced for the next epoch of training, once that level of accuracy is achieved then call-back switches. The confusion matrix is used over here to evaluate the performance of the training model.

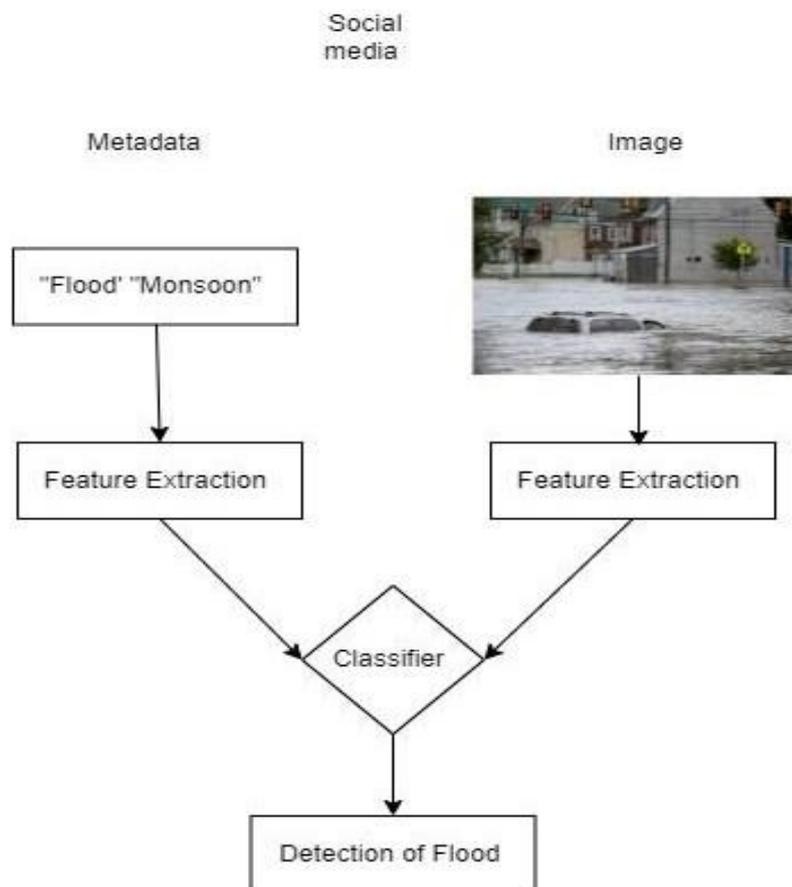


**Fig. 3.** CNN Architecture

*B. Social Media Text*

In Twitter Text analysis we are analyzing the text and then trying to differentiate between disaster twitter text and non- disaster twitter text. The input taken here consists of 4 attributes the first attribute is the keyword of the Twitter text the 2nd attribute is a location from where the twitter text is written and the next attribute is the text apart from these attributes, we have the target attribute. The target attribute has values as 0 and 1. Then we are creating a separate column that tells us whether the text is disaster-related or non-disaster According to the text values. We classify the value 0 as Non-Disaster text and value 1 as Non-Disaster related text. We have plotted attributes in the form of a pie chart to classify between Disaster text and non-disaster text. We have also tried to find the most common keywords in the disaster data. The classification is also done

based on the length of the tweets and location values. As the Twitter text might contain certain emojis and unnecessary punctuations. For that, we have used pattern matching we have first done data cleaning. Then with the help of the Natural Language Toolkit (NLTK), It is a set of libraries for Natural Language Processing. We have also used Stop words. We have also used Snowball Stemmer. With the help of Word cloud, we have displayed the Words that have been used most in Disaster Tweets and Non-Disaster Tweets. Here we have also tried to find out to what extent do the keywords overlap between the data sets and for that, we have used the basic Venn diagram structure. The final result is stored in an Excel sheet differentiating both types of tweets. For testing purposes, we have used 2 algorithms Logistic Regression and Random Forest Classifier. Logistic regression is a statistical model. The Accuracy of the model has also been calculated for Logistic Regression it is 90% and for Random Forest Classifier it is 98% . So Random Forest Classifier shows better accuracy in consideration with Logistic Regression.

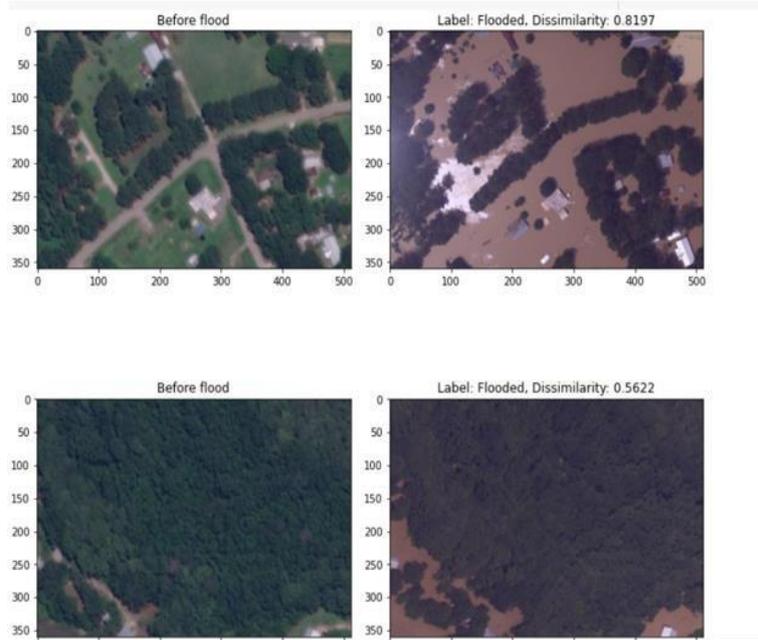


**Fig. 4.** Flowchart for Disaster detection using Social Media

## 5 Results

The implementation of our project has been shown below:

1) Satellite Imagery: Fig. 5 shows the result of disaster detection using satellite imagery. The dissimilarity, If the dissimilarity between the provided images is high then it means, that a particular region is flooded else the region is non-flooded.



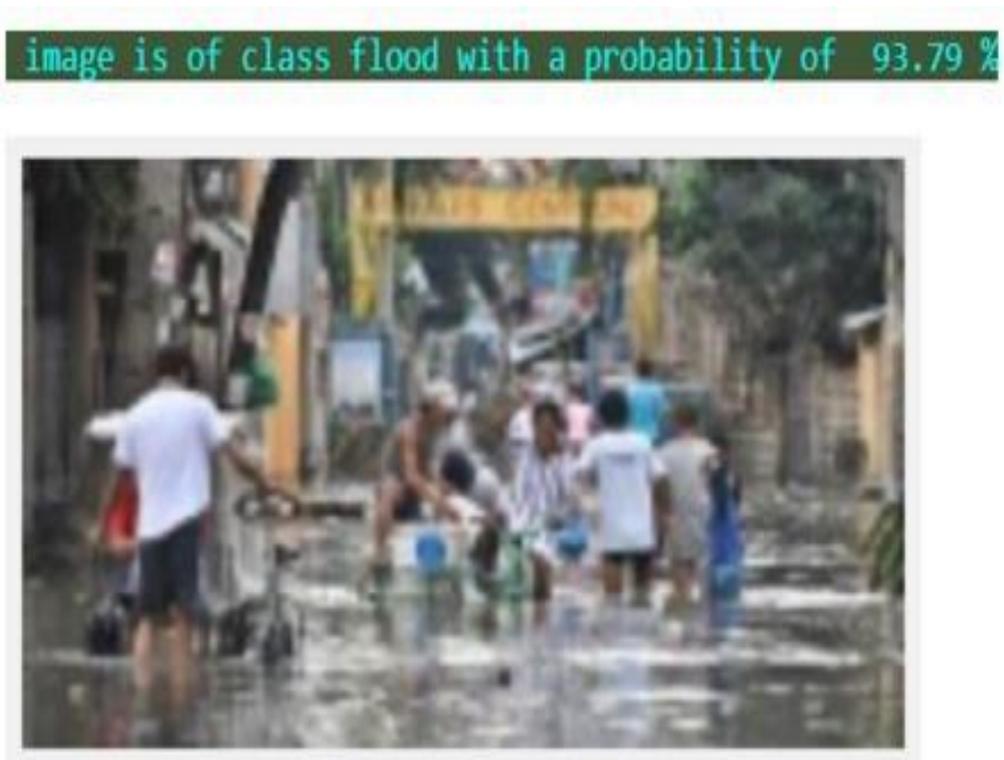
**Fig. 5.** Disaster detection using Satellite Imagery

2) Social Media Text: Table 1 shows the result of disaster detection using social media text.

**Table 1.** Disaster detection using Social Media Text

	id	keyword	location	text	target	class	length
0	1	NaN	NaN	Our Deeds are the Reason of this #earthquake M...	1	Disaster	13
1	4	NaN	NaN	Forest fire near La Ronge Sask. Canada	1	Disaster	7
2	5	NaN	NaN	All residents asked to 'shelter in place' are ...	1	Disaster	22
3	6	NaN	NaN	13,000 people receive #wildfires evacuation or...	1	Disaster	8
4	7	NaN	NaN	Just got sent this photo from Ruby #Alaska as ...	1	Disaster	16
5	8	NaN	NaN	#RockyFire Update => California Hwy. 20 closed...	1	Disaster	18
6	10	NaN	NaN	#flood #disaster Heavy rain causes flash flood...	1	Disaster	14
7	13	NaN	NaN	I'm on top of the hill and I can see a fire in...	1	Disaster	15
8	14	NaN	NaN	There's an emergency evacuation happening now ...	1	Disaster	12
9	15	NaN	NaN	I'm afraid that the tornado is coming to our a...	1	Disaster	10
10	16	NaN	NaN	Three people died from the heat wave so far	1	Disaster	9
11	17	NaN	NaN	Haha South Tampa is getting flooded hah- WAIT ...	1	Disaster	27
12	18	NaN	NaN	#raining #flooding #Florida #TampaBay #Tampa 1...	1	Disaster	12

3) Social Media Images: Fig. 6 shows the result of disaster detection using social media images. Here it predicts the image is of which class with its probability.



**Fig. 6.** Disaster detection using Social Media Image

## 6 Conclusion

This research proposes a method for detecting natural disasters using social media and satellite imagery. Different approaches which are used for detecting disasters in satellite images, social media text, and social media images are discussed in detail. Results of implementation of all three techniques are shown and these results clearly show the effectiveness of the proposed method for disaster detection. The above techniques is yet to be combined. The above models do not retrieve real-time data; this will be worked upon hereafter.

## References

1. Naina Said, Kashif Ahmad, Michael Riegler, Konstanin Pogorelov et al. *Natural Disasters Detection in Social Media and Satellite imagery* (2019).
2. Sheharryar Ahmad1, Kashif Ahmad, Nasir Ahmad, Nicola Conci et al. *Convolutional Neural Networks for Disaster Images Retrieval* (2017).
3. Olga Ostroukhova Pal, Holvorsen, Nicola Conci, Rozenn Dahvot et al. *Automatic detection of passable roads after floods in remote sensed and social media data* (2019).
4. Angela Maria Vinod, Dharathi Venkatesh, Dishti Kundra, Jayapandian N et al. *Natural Disaster Prediction by Using Image Based Deep Learning and Machine Learning* .
5. Jigar Doshi, Saikat Basu, Guan Pang et al *From Satellite Imagery to Disaster Insights "* .

6. Siti Nor Khuzaimah Binti Amit Soma Shiraishi Tetsuo Inoshita Yoshimitsu Aoki *Analysis of satellite images for disaster detection*.
7. Meedinti Gowri Namratha, Sanklap Chauhan, Swarnalatha P *Social Media use During Crisis Management, Disaster Response and Recovery Phases* (2022)
8. Lida Huang, Panpan Shi, Haichao Zhu<sup>2</sup>, Tao Chen<sup>1</sup> *Early detection of emergency events from social media: a new text clustering approach* (2022)
9. P Jedsada, S Tejal , Nipun Balan Thekkummal et al *Use of Social Media Data in Disaster Management: A Survey* (2021)
10. S. Ahmad, K. Ahmad, N. Ahmad, and N. Conci. *Convolutional neural networks for disaster images retrieval* (2017).
11. K. Avgerinakis, A. Moutzidou, S. Andreadis, E. Michail, I. Gialampoukidis, S. Vrochidis, and I. Kompatsiaris. *Visual and textual analysis of social media and satellite images for flood detection @ multimedia satellite task mediaeval*. (2017)
12. B. Bischke, P. Bhardwaj, A. Gautam, P. Helber, D. Borth, and A. Dengel. *Detection of flooding events in social multimedia and satellite imagery using deep neural networks*. In Working Notes Proceedings MediaEval Workshop, page 2 (2017)
13. C. Castillo. *Big crisis data: Social media in disasters and time-critical situations*. Cambridge University Press, 2016.
14. J. B. Houston, J. Hawthorne, M. F. Perreault, E. H. Park, M. Goldstein Hode, M. R. Halliwell, S. E. Turner McGowen, R. Davis, S. Vaid, J. A. McElderry, et al. *Social media and disasters: a functional framework for social media use in disaster planning, response, and research*. *Disasters* (2015).