Noname manuscript No.
(will be inserted by the editor)

Visual Attention Based Comparative Study on Disaster Detection from Social Media Images

Arif · M Ashraful Amin · Amin Ahsan Ali · AKM Mahhubur Rahman

(1611041, aminmdashraful, aminali, akmmrahman)@iub.edu.bd

This paper has been accepted on July 7, 2020 for publication in Innovations in Systems and Software Engineering: A NASA Journal

Abstract The availability of images of events almost in real-time on social media has a prospect in many application developments. A humanitarian technology for disaster type and level assessment can be developed using the images and video available on social media. In this paper, we investigate the potential use of various available deep learning techniques to develop such an application. For our research, based on the use of publicly available image data, we have started collecting disaster images from various sources from South Asia. We created the South Asia Disaster (SAD) image dataset containing 493 images from various online news portals. Using the Keras as our framework to run our models: Visual Geometry Group (VGG-16 and VGG-19), Inception-V3 and Inception-ResNet-V2 (ResNet: Residual Network). However, to boost up the training speed, we dropped the fully connected layer and added a small, fully connected model. To identify the five different disasters: fire disaster, flood disaster, human disaster, infrastructure disaster, natural disaster; our proposed method with VGG-16 model's recognition accuracy was 84.51%, which is the highest accuracy on the SAD dataset. After performing the testing, we calculate the VGG-16 classifier's attention to visualize which part of the disaster images VGG-16 pays attention.

Keywords Disaster image · humanitarian technology · standard disaster dataset · Convolutional Neural Network · VGG-16 · VGG-19 · Inception-V3 · Inception-ResNet-V2 · Keras model · visual attention

1 Introduction

Bangladesh is one of the most vulnerable countries that suffer from huge climate change, as well as disasters. Its population density and socioeconomic environments make it highly susceptible to many human-made hazards that include fire, collapse building, infrastructural damage, road accident, etc. Also, there are other disasters happening like infrastructure or non-natural disasters. Day after day, people are suffering from different kinds of disasters (e.g., Fire incidents in Chawk Bazar, Banani Fire incident). These incidents which happened in March 2019 were featured by some newspapers such as the famous online news portal Dhaka Tribune [20], The Daily Star [5], etc. It is crucial in times of crisis that how emergency response workers reach all those affected promptly. It would be great to have a system that would raise an alert and determine the degree of damage of any disaster and inform the appropriate authorities based on the automated analysis of the image data that are almost all the time available in real-time on various social media.

There has been some effort along this direction. Rizk et al. [17] proposed a multi-modal approach to automate crisis data analysis using machine learning. The proposed multi-modal two-stage framework relies on computationally inexpensive visual and semantic features to analyze Twitter data. In [6] presents the algorithms that CERTH team deployed to tackle disaster recognition tasks. Deep Convolutional Neural Network (DCNN), DBpedia Spotlight and combMAX were implemented to tackle DIRSM. The model GoogleNet was used to train on 5055 ImageNet concepts. Giannakeris et al. [8] presented a novel warning system framework for detecting people and vehicles in danger. The proposed framework provides a near real-time localization solution for detecting and scoring severity and safety levels of people and vehicles in flood and fire images. They chose to fine-tune the pre-trained parameters of the VGG-16 on Places365 dataset to leverage useful distinctions between various visual clues that relate to generic scenery images. Their Initial 3F-emergency dataset is composed of 6K images from Flickr.

Arif, M Ashraful Amin, Amin Ahsan Ali and AKM Mahhubur Rahman
Artificial Intelligence and Cybernetics Lab, CSE, Independent University, Bangladesh
www.agencylab.org

Alam et al. [3] proposed a work where their Image Filtering module employs deep neural networks and perceptual hashing techniques to determine whether a newly-arrived image is relevant for a given disaster response context. To train the relevancy filter, 3,518 images were randomly selected from the severe and mild categories. They adopted a transfer learning approach where they have used the VGG-16 network pre-trained on ImageNet data. Mouzannar et al. [13] proposed a multimodal deep learning framework to identify damage related information from social media posts. This framework combines multiple pretrained unimodal convolutional neural networks that extract features from raw texts and images separately. The inception convolutional neural network (CNN) model was adopted, which was pre-trained on ImageNet to process images, and for words, they used a pre-trained word embedding model to process the texts. The framework was evaluated on a homegrown labeled dataset of multimodal social media posts. In the paper [16], images posted on social media platforms during natural disasters to determine the level of damage caused by the disasters are analyzed. In this study, Imran et al. used the VGG-16 network trained on the ImageNet dataset. They collected data from the Web (such as Typhoon Ruby, Hurricane Matthew, Nepal Earthquake, etc.) and made their dataset. They ran tests on their training sets, and the highest accuracy they achieved was when they combined their Google, Ruby and Matthew datasets. In this research paper, we propose a deep learning-based method to automate the effective extraction of information from social media posts to direct relief resources efficiently. Since posts on social media contain text, photos, and videos, we are performing a deep learning framework for multimodal identification of damage-related information. Also, our objective is to collect and work with South Asian disaster images, which include disaster images of Bangladesh and similar countries. Moreover, we have performed experiments to visualize how the classifiers pay their attention on the disaster images. By calculating attention, we are able to see how well the classifier could concentrate it's attention on the disaster images.

Our contributions are clarified below:

First-Contributions: We have made a dataset called South-Asia-Disaster (SAD) dataset which contains disaster images from south Asian countries (India, Bangladesh, and the Philippines) which is our contribution to this paper.

Second-Contributions: We have Compared different deep learning models for classifying disaster images. We have used four models VGG16, VGG19, Inception-V3, and Inception Resnet V2. Why did we use this four models? because they are from three CNN families: Virtual Geometry Group (VGG), Inception by GoogLeNet and Inception Residual Network (ResNet) by Microsoft.

Third-Contributions: We have shown in which image regions, the VGG -16 model pays its attention by performing Class Activation Map (CAM) which can be used to get a better understanding of where a Convolutional Neural Network (CNN) focuses its attention.

2 Methodology

Our proposed system uses Keras as a framework to implement our models and modified deep learning algorithms to classify five different types of disaster images. And for all deep learning-based methods, data is the most important component.

2.1 Dataset Description

Standard database is a prerequisite for better model creation [4]. The main issue using a deep learning system for classification is a well defined and diversified dataset. For this experiment, we collected the images from [13]. There are six different damage categories; they are: Fire damage, Flood damage, damage infrastructure, damage nature, Human damage, and Non-damage. There is a total of 5885 images, and the total number of images for each category is given in Table 1. The fire disaster images contain fire elements, smoke, and burning objects and burned objects. Flood disaster images contain a huge volume of water, objects underwater and objects submerged in water. Infrastructure disaster images contain collapsed buildings, rusted, and damaged objects. Nature disasters contain broken trees, buildings, and roads caused by earthquake or cyclone. Human disaster image contains bleeding, burned face, torturing people, and human damaging the environment. Non – disaster images contain products, cosmetics, books, human models, and people eating food. Some sample images from the [13] data set are given in Fig.1. From Table 1, we can observe that Non-damage has the highest number of images, 2972 than other classes, the lowest number of images are human damage 240. Damage nature has the second-highest number of images 1418. Other classes have an adjacent number of images.

In the sample images note that the images are from mostly from outside South Asia. We intend to implement our system for Bangladesh and to check the performance on the native data, and we collected a small data set for this project. We are calling it, SAD (South Asian Disaster) image dataset. We have collected 493 data, and class-wise distribution can be found in Table 2.

Table 1 Distribution of the number of disaster images for each class from [13] Dataset

Categories	Number
Fire disaster	349
Flood disaster	385
Infrastructure disaster	515
Nature disaster	1418
Human disaster	240
Non-disaster	2972
Total	5885

Fig. 1 Sample images from [13] Dataset (left-right): (a) fire disaster, (b) flood disaster, (c) human disaster, (d) natural disaster,(e) infrastructure disaster , and (f) non-disaster

Table 2 Distribution of the number of disaster images for each class from SAD Dataset

Categories	Number
Fire disaster	82
Flood disaster	82
Infrastructure disaster	65
Nature disaster	73
Human disaster	81
Non-disaster	110
Total	493

In Fig. 2, we gave samples of the SAD image dataset. Images were collected from online news portals, social news networks, independent news organizations, and online benefactor for people. We got 32 images from each online news portal such as The Times of India [7] and website Getty Images [10]. 32 images are collected from the UN news (news portal) and website Getty Images [11], 32 images from World Bank (online benefactor for people), 32 images from Rappler (social news network) and website Getty Images [12] , 100 images from BBC (online news portal), 100 images from Fox news (online news portal) and rest from Al Jazeera (independent news organization) [2] [15]. Visual Comparison among images from Fig. 1 and Fig. 2 it is clear that there are differences in perception of similar classes.

2.2 Convolutional Neural Networks (CNNs)

CNNs have shown state-of-the-art efficiency in a multitude of computer vision challenges, including image classification and retrieval, object detection, and segmentation of images [14]. A typical CNN comprises of various processing layers, including convolution, max pooling, and fully connected layers. Our study is mainly based on performing different types of experiments with different models on our SAD dataset and [13] dataset and compare and analyze their results. To perform our experiments, we have decided to use four models from three CNN families: Virtual Geometry Group (VGG) [21], Inception [19] by GoogLeNet

Fig. 2 Sample images from SAD image dataset (left-right): (a) fire disaster, (b) flood disaster, (c) human disaster, (d) infrastructure disaster, (e) natural disaster, and (f) non-disaster

and Inception Residual Network (ResNet) [18] by Microsoft. Our objective is to compare results; hence, we have decided to compare the results of different CNN families' model as their architectures are different from one another. The VGG architecture is deep, using only 3x3 convolutional layers stacked on top of each other, which increase depth. Volume size is reduced by max pool layer. It has cascaded convolution max-pool layers [21]. The Inception architecture, on the other hand, is "wider" rather than "deep" like the VGG architecture. The architecture typically consists of filters of 3 different sizes: 1x1, 3x3 and 5x5. After max-pooling, the outputs are concatenated and sent to the next inception module. The newer version of the architecture has an extra added 1x1 convolution before the 3x3 and 5x5 convolutions for cost-effectiveness [19]. One of the most popular CNN family in the present is the ResNet family by Microsoft. This architecture goes deeper than the VGG architecture with nodes having stems of different Inception Modules [18]. In [8] and [16] VGG-16 CNN model was implemented. The [13] Inception-V3, Inception-V4, VGG-16, and Inception-Resnet-V2, for disaster image classification. For our problem, we used VGG-16, VGG-19, Inception-V3, and Inception-ResNet-V2. **Visual Geometry Group (VGG) [21]:** The VGG nets were originally designed to detect and recognize objects in images. VGG is a very profound model that sig-nificantly enhanced a wide variety of visual recognition functions, including de-tecting objects, semantic segmentation, picture captioning, and recognition of the action on video. VGG-16 has three fully connected layers follow a stack of convo-lutional layers, and the final layer is the softmax layer. It has a total of 16 layers; all hidden layers are equipped with rectification. VGG-19 model is very similar to the VGG-16 model. The main difference between VGG-16 and VGG-19 is that the VGG-19 has 19 layers instead of 16.

Inception [19]: . Inception architectures of GoogLeNet were intended to perform well even under rigorous memory and computer budget limitations. Inception-v1 has nine inception modules stacked linearly, and 27 layers deep, including pooling layers. In Inception-V2 the 5x5 convolution was factorized to two 3x3-convolution operations to improve computational speed and reduce computational expense. The Inception-V3 (INv3) is 42 layers deep; however, the computation is only about 2.5 times more than Inception-v1 and much more efficient than VGG Net.

Inception-ResNet-V2 [18]: Inception-ResNet-V2 (INRv2) is a 164 profound layer deep neural network that can classify pictures into 1000 categories of objects. Inception-ResNet-V2 has introduced residual connections that add the output of the convolution operation of the inception module to the input. For residual addition to work, the input and output after convolution must have the same dimensions. To increase stability, the residual activations were scaled by a value within 0.1 to 0.3 range. The models that we are using for our experiment are: VGG-16, VGG-19, Inception-V3, and Inception-ResNet-V2. In Fig. 3 we gave the schematic view of the models used. We have used relu activation on the 512 dense layers and also on the 256 dense layers so that all the layers are connected to each other. And then we used dropout, and it helps reduce over-fitting, by preventing a layer from seeing twice the same pattern. In the end, we used six dense layers because we have six classes. As optimizer we used Adam, and as loss function, sparse categorical cross-entropy is used.

If we use the original fully connected layer it takes 1 day to complete the training process. That's why we added a small model to speed our training process, Therefore it takes 5 to 6 hours to finish the training.

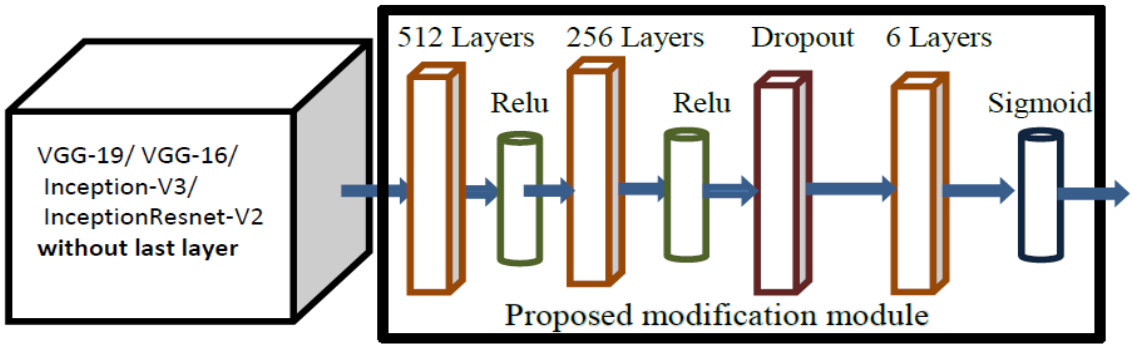

Fig. 3 Schematic view of the proposed deep neural model

We have to train more parameters If we used the fully connected layer, for example, let's take VGG16 which has: 4096*4096 + 4096*6 = 16801792 parameters (6 is the number of classes we have) in its last fully connected layer and the small model that we added only have 512*256 + 256*6 = 132608 parameters. Therefore, it takes more time to train more parameters and we didn't have GPU to train our model.

Therefore this architecture can be used to train a dataset on CPU without even worrying about using a GPU and get a faster training process. We performed Class Activation Maps (CAM), which can be used to get a better understanding of where CNN focuses its attention.

2.3 Where CNN model gives its attention ?

Lets have a look where CNN model gives its attention. The central objective of this work is to use attention maps to recognize and utilize the successful spatial help of the visual information that CNN models use to make their classification decisions [9]. Therefore we are performing Class Activation Map (CAM) [22].

2.3.1 What is Class Activation Map (CAM) ?

When the fully connected layer is used the ability to localize objects in the convolutional layers, this information gets lost. A Class Activation Map (CAM) helps us to see which image regions are relevant to which class. CAM draws the heatmap of the network that shows the activated region. Just before the final output layer, global average pooling (GAP) is performed on the convolutional feature maps and those features are used for forming a fully-connected layer that produces the desired output.The value of the image regions can be defined by projecting back the weights of the output layer onto the convolutionary feature maps, a technique they call class activation mapping [22].

3 Experimental Setup

Images are usually of different size as they are collected from non-restricted sources. To feed them into the deep neural networks, they were resized to 150 by 150 pixels. For training the networks 50 epochs have been used, epoch means training the whole dataset in one cycle; however, it takes a lot of space in the memory. Therefore, the dataset is divided into batches to train. The batch size 11 and 12 are used for [13] training and testing dataset, respectively. For SAD dataset the batch size is 29 for the whole dataset. To run this experiment, we are using In-tel core i5 processor and 8 GB DDR4 ram. Fig.4 shows the process of the experiments that are going to be performed. The experiments that we are performing are:

1. **Experiment 01:** Cross-validation performance measure while training the CNNs with [13] dataset testing them on [13] dataset.Where we load a fresh model whenever a training and test phase is done. Table 3 shows the results of this experiment.
2. **Experiment 02:** Performance measure while training with 100% [13] dataset and test with 100 % SAD dataset. Table 4 shows the results of this experiment
3. **Experiment 03:** Cross-validation performance measure while training with the 100% [13] dataset plus 60% of the SAD dataset and test with the remaining 40% of the SAD dataset. Table 5 shows the results of this experiment.
4. **Experiment 04:** Attention calculation while training with the 100% [13] dataset plus 60% of the SAD dataset with Class Activation Map architecture [22] and test with the remaining 40% of the SAD dataset.

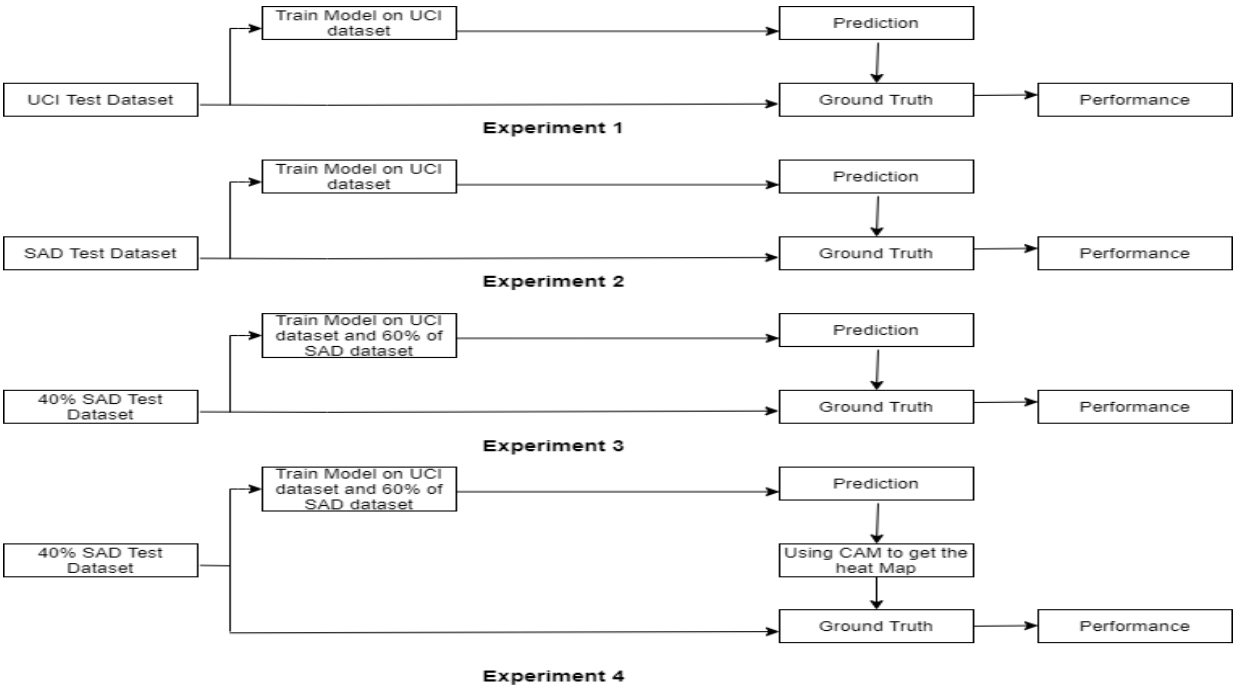

Fig. 4 Walk-through of the experiment process

In Table 3, note that all four classifiers perform in the same range on average. However, different classifiers show slightly higher classification accuracy class of disaster.

In Table 4, the first issue to note is that, as we test the classifiers trained with [13] dataset and test with our SAD dataset, the accuracy reduces by about 7%. This means that the [13] dataset does not contain well distributed and diverse data from all demography but the SAD dataset is different from [13] dataset (Fig. 5). Some-thing very noticeable in this experiment is that the precision for human damage is significantly low for all classifiers (Fig. 1 and Fig. 2). This would mean that iden-tifying human damage by a deep neural net trained with [13] dataset only is very difficult. VGG16 is an older version of the deep neural network; however, the overall accuracy of this classifier is slightly higher than other models for experiment 2. In experiment 3, we added 60% of the SAD data with the [13] dataset, and it shows the improvement in performance. In Table 5, note that the precision of human damage improves. Moreover, accuracy is also increased. It suggests that if the networks can be trained with significantly large and diverse data, performance will increase with the same network setup.

Table 3 Cross-validation average performance of only [13] dataset

	F1-Score (%)				Accuracy (%)			
	VGG9	VGG16	INv3	INRv2	VGG19	VGG16	INv3	INRv2
Fire damage	65.57	62.74	66.15	65.80	95.51	95.62	95.72	96.08
Flood damage	57.01	55.77	50.45	56.48	93.92	94.37	93.91	94.25
Human damage	63.70	60.19	57.92	61.91	97.07	96.54	96.82	96.79
Damage Infrastructure	67.90	66.67	65.42	68.47	83.87	83.04	83.54	84.33
Damage Nature	44.07	48.91	48.00	50.12	91.28	91.13	90.40	91.06
Non-damage	83.38	82.91	82.59	83.77	83.65	82.91	81.79	82.60
Average	63.60	62.86	61.75	64.42	90.88	90.60	90.36	90.36

Comparing table 3 and 4, the accuracy dropped around 7% when preforming the second experiment. The reason for this drop of accuracy maybe for some specific class, there is significant between [13] and SAD dataset. In Fig. 5 we observe that SAD flooding disaster and nature disaster is different than other countries flooding disaster and natural disaster. In Fig. 6 we provide some images associated with fire class, here we observed that there are two types of images: During fire disaster and after a fire disaster. In Fig. 6 not that during fire disaster and after fire disaster images are visually very distinctive. The difference is, in one has fire element, and smoke in it and the other one is all black, and there are ashes. Visually we ourselves are confused between the image (b) (after fire disaster) and the image (c) (damaged infrastructure). Thus, the deep networks are also sometimes confused with after fire image and infrastructure disaster image.

In Fig. 7, we provide some more images that are misclassified by our system. (a) After fire damage classified as Damaged Infrastructure; (b) Is an actual infrastructural damage training image; (c) Another

Table 4 Performance of training with 100% [13] dataset and testing with 100% SAD dataset

	F1-Score (%)				Accuracy (%)			
	VGG9	VGG16	INv3	INRv2	VGG19	VGG16	INv3	INRv2
Fire damage	62.800	69.50	69.87	54.89	90.87	91.28	91.28	89.6
Flood damage	56.06	61.24	58.81	55.40	88.24	87.42	89.25	88.24
Human damage	10.80	14.45	14.63	17.23	86.61	85.60	85.50	86.41
Damage Infrastructure	46.18	49.29	40.22	41.83	71.60	77.08	64.91	71.81
Damage Nature	39.96	40.34	25.95	51.51	84.79	84.38	83.77	83.98
Non-damage	64.27	63.20	57.90	59.23	80.12	81.34	79.92	78.51
Average	46.68	49.67	44.56	46.68	83.70	84.51	82.43	83.10

Table 5 Cross-validation performance of training with 100% [13] dataset plus 60% SAD dataset, and testing with 40% SAD dataset

	F1-Score (%)				Accuracy (%)			
	VGG9	VGG16	INv3	INRv2	VGG19	VGG16	INv3	INRv2
Fire damage	82.42	77.18	62.77	82.11	94.95	93.43	90.40	94.68
Flood damage	48.95	62.97	54.21	49.98	88.38	89.90	88.89	85.11
Human damage	23.58	30.03	15.37	27.78	86.87	85.86	88.89	84.04
Damage Infrastructure	54.22	56.57	54.83	53.04	75.25	79.80	79.77	70.74
Damage Nature	36.84	34.75	38.49	31.81	87.88	84.85	83.84	86.17
Non-damage	66.66	69.14	70.11	65.41	80.81	83.33	82.32	76.06
Average	52.11	55.11	49.301	51.69	85.69	86.20	85.68	82.80

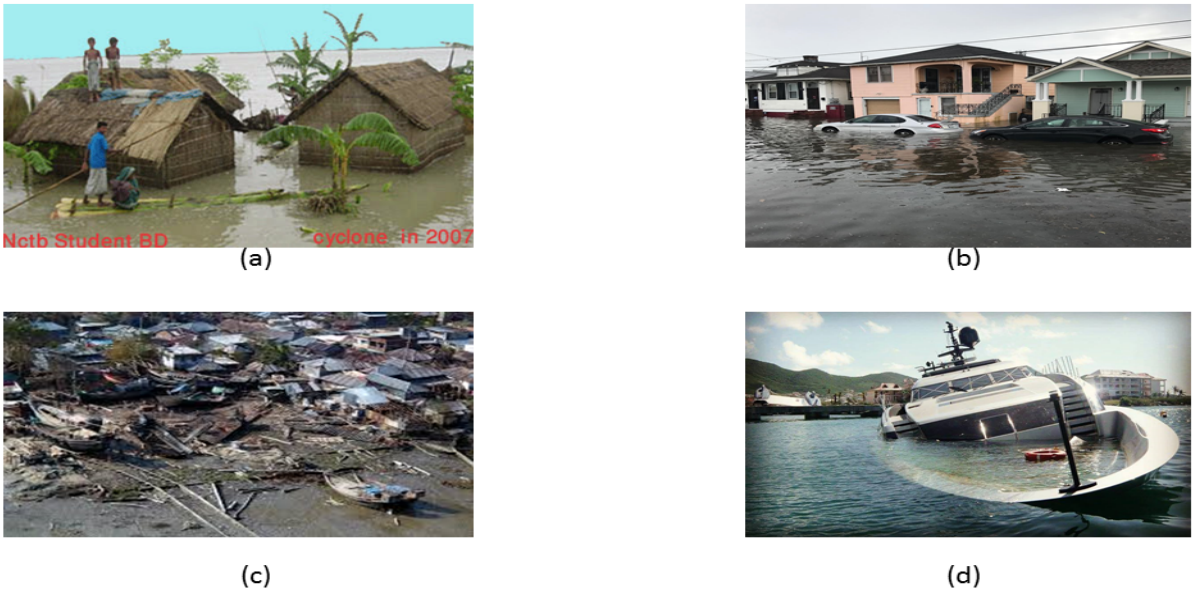

Fig. 5 (Left – Right; Top-bottom) (a) SAD Flood damage; (b) Other Countries Flood damage; (c) SAD Damage Nature; (d) Other Countries Damage Nature

Fig. 6 (Left - Right) (a) During Fire damage; (b) After Fire damage; (c) Damage Infra-structure.

after fire damage classified as flood damage; in the image (d) the black and white floor and books floating resembles the state of the floor of (c), and in (e) the training flood image does not really look like a flood

image, but there are pipes appearing in it which also appears in test image (c). We believe that some of the fire test images were classi-fied as flood and infrastructure by Inception-V3 and Inception-ResNet-V2 models, because, the test image does not contain any fire object in them.

We are Comparing our VggNet results with three recent [13] [1] [16] works because VggNet models got the highest accuracy from the other two models on our SAD dataset in experiment 2 We present the comparison in Table 6.

Table 6 Comparing our VggNet results with other recent works VggNet results

VggNet Model	Hussein et al, [13]	Ahmad et al, [1]	Nguyen et al,[16]	Our Result
VGG16	82.19(%)	-	78.60(%)	84.51(%)
VGG19	-	72.70(%)	-	83.70(%)

In Table 6, there are some blank spaces because [1] didn't use the VGG16 model instead they have used VGG19. Therefore we have compared their VGG19 result with our VGG19 result.

Hussein et al. [13] and Nguyen et al. [16] have used the VGG16 model, therefore we have compared their results with our VGG16 result. In [16], their proposed systems average accuracy for 5 class is 78.60(%). It can be easily observed that our VGG16 model has outperformed Hussein et al. [13] and Nguyen et al. [16] with a good margin. Moreover, our model also performs better compared to VGG19 model [1].

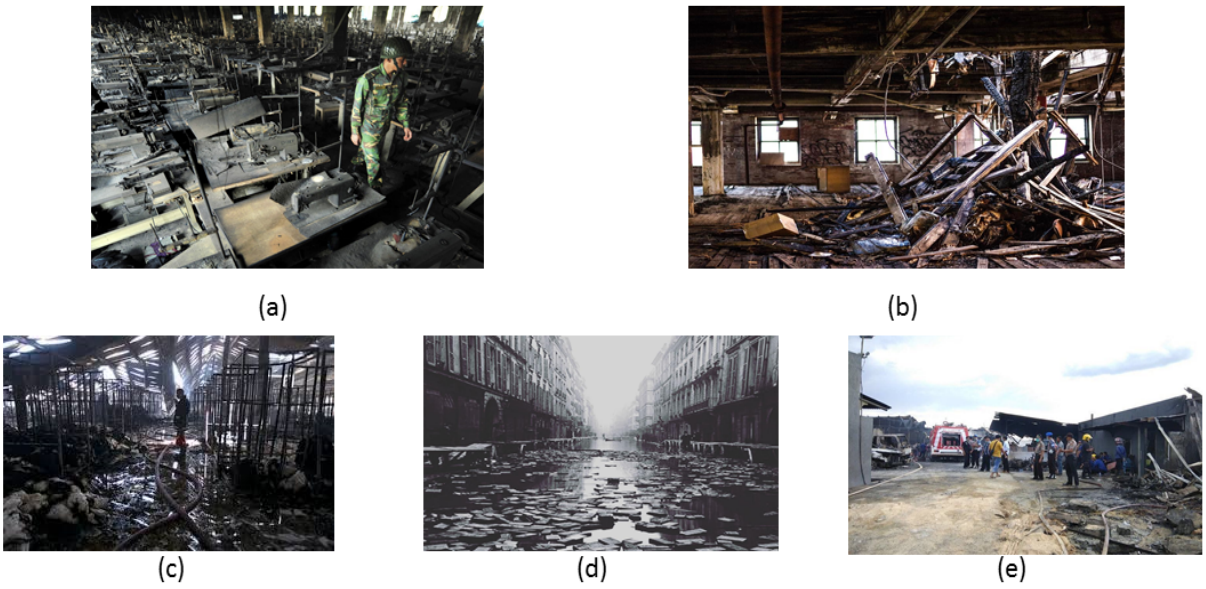

Fig. 7 Wrong Prediction by the proposed system; (Left-Right) (a) After fire damage classified as Damaged Infrastructure; (b) Actual training Damaged Infrastructure image; (c) Fire damage classified as flood damage; (d) and (e) Actual Flood damage training image.

Finally, we have performed experiments that help us to understand how the VGG-16 provides its attention on the disaster images. Performing CAM with VGG16 because VGG16 performed better than other models in experiment 3. Therefore we are performing experiment 3 again with CAM. Some sample images are given in fig 8. the top row (a),(b),(c) contains the actual images and down row (d),(e),(f) contains the CAM images of fire disaster, infrastructure and flood disaster. Fig.8 shows the correct classified images.

As we can see in the CAM images there are red and blue region. The red region means that more weight is given in other word more attention is given and blue region means less attention is given. In fig 9 misclassification images are shown, image (a) which a nature-disaster but classified as fire disaster because the tornado looks like a smoke and image (b) which is a human-disaster but classified as flood because of the wet road and image (c) which is also a human disaster but classified as damaged infrastructure because broken road and rocks on the road therefore in table 7 the precison(%) and recall(%) of human disaster are so low.

4 Conclusion

Real-time support is the most significant challenge to minimize damages by any disaster; most of the time, the traditional method of assessing the occurrence of disaster and it's characteristics takes too long,

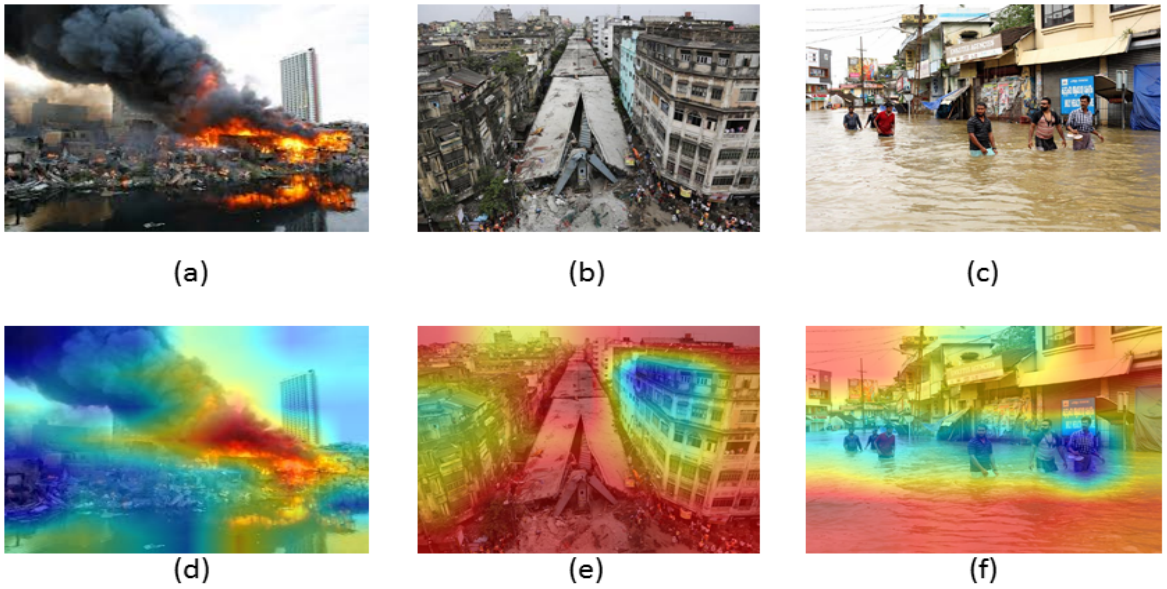

Fig. 8 CAM results on SAD dataset, top row (a),(b),(c) are the actual images and down row (d),(e),(f)are the CAM images, (left - right) Fire-disaster, Infrastructure and Flood-disaster

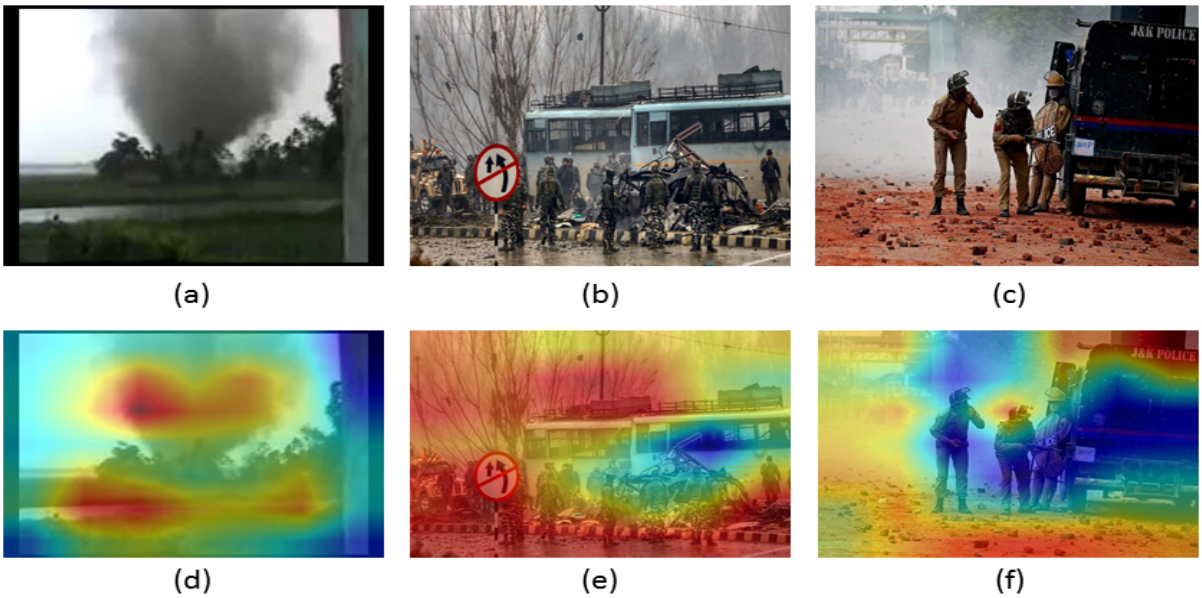

Fig. 9 CAM results on SAD dataset, top row (a),(b),(c)are the actual images and down row (d),(e),(f) are the CAM images, (top row, left - right) (a) Nature-disaster, (b) Human-disaster, (c) Human-disaster images, (down row, left - right) misclassified as (d) Fire-disaster, (e) Flood-disaster and (f) Infrastracture

Table 7 VGG16 CAM results

Classes	Accuracy(%)	Precision(%)	Recall(%)	f1-score(%)
Fire damage	95.8	82.0	93.0	87.15
Flood damage	94.2	79.0	87.0	82.80
Human damage	85.3	4.0	20.0	6.66
Damage Infrastructure	70.0	88.0	35.0	50.08
Damage Nature	87.4	33.0	85.0	47.54
Non-damage	90.5	70.0	74.0	71.94

and in the meantime, the damage has been done. In this paper, we propose a method to determine the type of disaster automatically from social media images. To test the performance of the proposed idea publicly available dataset is used. Moreover, we extended the dataset by gathering 493 disaster images of six classes from this region, and we are calling it “South Asia disaster” (SAD). We implemented four different types of deep neural network models: VGG 16, VGG 19, Inception-V3 and Inception-ResNet-V2, for this research. However, to boost up the training speed, we removed the fully connected layer and

added a small, fully connected model. To identify the five different disasters: fire disaster, flood disaster, human disaster, infrastructure disaster, and natural disaster; our proposed method with VGG 16 models recognition accuracy is 84.51% on the SAD dataset. We also observed that Class Activation Map(CAM) can be used to get a better understanding of where a Convolutional Neural Network (CNN) focuses its attention.

Acknowledgements This project is supported by a grant from the Bangladesh Information and Communication Technology Ministries ICT Division, and Independent University, Bangladesh (IUB). We would also like to thank Dr. Moinul Islam Zaber, Associate Professor, CSE, Dhaka University, Bangladesh, for his valuable suggestions during this project.

References

1. Ahmad, K., Conci, N., Natale, F.: A comparative study of global and deep features for the analysis of user-generated natural disaster related images (accepted for presentation in ivmsp 2018) (2018)
2. Al Jazeera Media Network, Z.H.C.: aljazeera. In: viewed 20 Jun 2019, https://www.aljazeera.com/indepth/inpictures/pictures-bangladesh-front-lines-climate-crisis-191203102447517.html
3. Alam, F., Imran, M., Ofli, F.: Image4act: Online social media image processing for disaster response. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17, p. 601–604. Association for Computing Machinery, New York, NY, USA (2017). DOI 10.1145/3110025.3110164. URL https://doi.org/10.1145/3110025.3110164
4. Amin, M.A., Mohammed, M.K.: Overview of the imageclef 2015 medical clustering task. In: CLEF (2015)
5. Anam, M.: The daily star. In: viewed 20 Jun 2019, https://www.thedailystar.net/ (1991)
6. Avgerinakis, K., Moumtzidou, A., Andreadis, S., Michail, E., Gialampoukidis, I., Vrochidis, S., Kompatsiaris, I.: Visual and textual analysis of social media and satellite images for flood detection @ multimedia satellite task mediaeval 2017 (2017)
7. Bose, J.: The times of india. In: viewed 20 Jun 2019, https://timesofindia.indiatimes.com/
8. Giannakeris, P., Avgerinakis, K., Karakostas, A., Vrochidis, S., Kompatsiaris, I.: People and vehicles in danger - a fire and flood detection system in social media. In: 2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), pp. 1–5 (2018)
9. Jetley, S., Lord, N.A., Lee, N., Torr, P.H.S.: Learn to pay attention. CoRR **abs/1804.02391** (2018). URL http://arxiv.org/abs/1804.02391
10. Mark Getty, J.K.: gettyimages. In: viewed 20 Jun 2019, https://www.gettyimages.com/photos/india-fire?mediatype=photographypage=3phrase=india
11. Mark Getty, J.K.: gettyimages. In: viewed 20 Jun 2019,
12. Mark Getty, J.K.: gettyimages. In: viewed 20 Jun 2019, https://www.gettyimages.com/photos/bangladesh-flood?mediatype=photographyphrase=bangladesh
13. Mouzannar, H., Rizk, Y., Awad, M.: Damage identification in social media posts using multimodal deep learning. In: ISCRAM (2018)
14. Muhammad, K., Ahmad, J., Baik, S.W.: Early fire detection using convolutional neural networks during surveillance for effective disaster management. Neurocomput. **288**(C), 30–42 (2018). DOI 10.1016/j.neucom.2017.04.083. URL https://doi.org/10.1016/j.neucom.2017.04.083
15. Network, A.J.M.: aljazeera. In: viewed 20 Jun 2019, https://www.aljazeera.com/
16. Nguyen, D.T., Ofli, F., Imran, M., Mitra, P.: 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, ASONAM '17, p. 569–576. Association for Computing Machinery, New York, NY, USA (2017). DOI 10.1145/3110025.3110109. URL https://doi.org/10.1145/3110025.3110109
17. Rizk, Y., Jomaa, H.S., Awad, M., Castillo, C.: A computationally efficient multi-modal classification approach of disaster-related twitter images. In: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, SAC '19, p. 2050–2059. Association for Computing Machinery, New York, NY, USA (2019). DOI 10.1145/3297280.3297481. URL https://doi.org/10.1145/3297280.3297481
18. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI'17, p. 4278–4284. AAAI Press (2017)
19. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818–2826 (2016)
20. Zafar Sobhan, K.A.A.: Dhakatribune. In: viewed 20 Jun 2019, https://www.dhakatribune.com/ (2013)
21. Zhang, X., Zou, J., He, K., Sun, J.: Accelerating very deep convolutional networks for classification and detection. IEEE Trans. Pattern Anal. Mach. Intell. **38**(10), 1943–1955 (2016). DOI 10.1109/TPAMI.2015.2502579. URL https://doi.org/10.1109/TPAMI.2015.2502579
22. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2921–2929 (2016)