MEMIS: Multimodal Emergency Management Information System

Mansi Agarwal[1(✉)], Maitree Leekha[1(✉)], Ramit Sawhney[2], Rajiv Ratn Shah[3], Rajesh Kumar Yadav[1], and Dinesh Kumar Vishwakarma[1]

[1] Delhi Technological University, New Delhi, India
r18522mansi@dpsrkp.net, maitreeleekha@yahoo.in
[2] Netaji Subhas Institute of Technology, New Delhi, India
[3] Indraprastha Institute of Information Technology, New Delhi, India

Abstract. The recent upsurge in the usage of social media and the multimedia data generated therein has attracted many researchers for analyzing and decoding the information to automate decision-making in several fields. This work focuses on one such application: disaster management in times of crises and calamities. The existing research on disaster damage analysis has primarily taken only unimodal information in the form of text or image into account. These unimodal systems, although useful, fail to model the relationship between the various modalities. Different modalities often present supporting facts about the task, and therefore, learning them together can enhance performance. We present **MEMIS**, a system that can be used in emergencies like disasters to identify and analyze the damage indicated by user-generated multimodal social media posts, thereby helping the disaster management groups in making informed decisions. Our leave-one-disaster-out experiments on a multimodal dataset suggest that not only does fusing information in different media forms improves performance, but that our system can also generalize well to new disaster categories. Further qualitative analysis reveals that the system is responsive and computationally efficient.

Keywords: Disaster management · Multimodal systems · Social media

1 Introduction

The amount of data generated every day is colossal [10]. It is produced in many different ways and many different media forms. Analyzing and utilizing this data to drive the decision-making process in various fields intelligently has been the primary focus of the research community [22]. Disaster Response Management is one such area. Natural calamities occur frequently, and in times of such crisis, if the large amount of data being generated across different platforms is harnessed

M. Agarwal and M. Leekha—The authors contributed equally, and wish that they be regarded as joint First Authors.

Rajiv Ratn Shah is partly supported by the Infosys Center for AI, IIIT Delhi.

J. M. Jose et al. (Eds.): ECIR 2020, LNCS 12035, pp. 479–494, 2020.
https://doi.org/10.1007/978-3-030-45439-5_32

well, the relief groups will be able to make effective decisions that have the potential to enhance the response outcomes in the affected areas.

To design an executable plan, disaster management and relief groups should combine information from different sources and in different forms. However, at present, the only primary source of information is the textual reports which describe the disaster's location, severity, etc. and may contain statistics of the number of victims, infrastructural loss, etc. Motivated by the cause of humanitarian aid in times of crises and disasters, we propose a novel system that leverages both textual and visual cues from the mass of user-uploaded information on social media to identify damage and assess the level of damage incurred.

In essence, we propose MEMIS, a system that aims to pave the way to automate a vast multitude of problems ranging from automated emergency management, community rehabilitation via better planning from the cues and patterns observed in such data and improve the quality of such social media data to further the cause of immediate response, improving situational awareness and propagating actionable information.

Using a real-world dataset, CrisisMMD, created by Alam *et al.* [1], which is the first publicly available dataset of its kind, we present the case for a novel multimodal system, and through our results report its efficiency, effectiveness, and generalizability.

2 Literature Review

In this section, we briefly discuss the disaster detection techniques of the current literature, along with their strengths and weaknesses. We also highlight how our approach overcomes the issues present in the existing ones, thereby emphasizing the effectiveness of our system for disaster management.

Chaudhuri *et al.* [7] examined the images from earthquake-hit urban environments by employing a simple CNN architecture. However, recent research has revealed that often fine-tuning pre-trained architectures for downstream tasks outperform simpler models trained from scratch [18]. We build on this by employing transfer learning with several successful models from the ImageNet [9], and observed significant improvements in the performance of our disaster detection and analysis models, in comparison to a simple CNN model.

Sreenivasulu *et al.* [24] investigated microblog text messages for identifying those which were informative, and therefore, could be used for further damage assessment. They employed a Convolutional Neural Network (CNN) for modeling the text classification problem, using the dataset curated by Alam *et al.* [1]. Extending on their work on CrisisMMD, we experimented with several other state-of-the-art architectures and observed that adding recurrent layers improved the text modeling.

Although researchers in the past have designed and experimented with unimodal disaster assessment systems [2,3], realizing that multimodal systems may outperform unimodal frameworks [16], the focus has now shifted to leveraging information in different media forms for disaster management [20]. In addition to using several different media forms and feature extraction techniques, several researchers have also employed various methods to combine the information obtained from these modalities, to make a final decision [19]. Yang *et al.* [28] developed a multimodal system- MADIS which leverages both text and image modalities, using hand-crafted features such as TF-IDF vectors, and low-level color features. Although their contribution was a step towards advancing damage assessment systems, the features used were relatively simple and weak, as opposed to the deep neural network models, where each layer captures complex information about the modality [17]. Therefore, we utilize the latent representation of text and image modalities, extracted from their respective deep learning models, as features to our system. Another characteristic that is essential for a damage assessment system is generalizability. However, most of the work carried out so far did not discuss this practical perspective. Furthermore, to the best of our knowledge, so far no work has been done on developing an end-to-end multimodal damage identification and assessment system.

To this end, we propose MEMIS, a multimodal system capable of extracting information from social media, and employs both images and text for identifying damage and its severity in real-time (refer Sect. 3). Through extensive quantitative experimentation in the leave-one-disaster-out training setting and qualitative analysis, we report the system's efficiency, effectiveness, and generalizability. Our results show how combining features from different modalities improves the system's performance over unimodal frameworks.

3 A Real-Time Tweet Processing Pipeline

In this section, we describe the different modules of our proposed system in greater detail. The architecture for the system is shown in Fig. 1. The internal methodological details of the individual modules are in the next section.

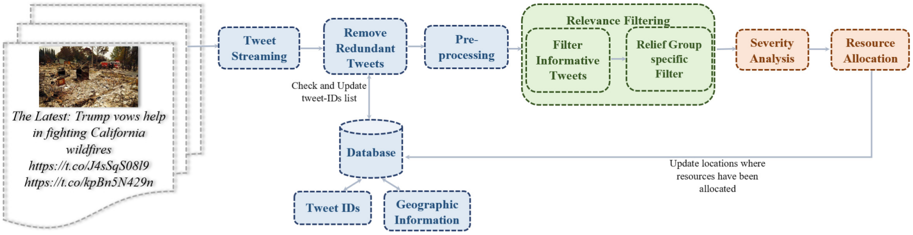

The Latest: Trump vows help in fighting California wildfires
https://t.co/J4sSqS08l9
https://t.co/kpBn5N429n
Tweet Streaming
Remove Redundant Tweets
Pre-processing
Relevance Filtering
Filter Informative Tweets
Relief Group specific Filter
Severity Analysis
Resource Allocation
Check and Update tweet-IDs list
Database
Update locations where resources have been allocated
Tweet IDs
Geographic Information

Fig. 1. System architecture of MEMIS

3.1 Tweet Streaming

The Tweet Streaming module uses the Twitter Streaming API[1] to scrap real-time tweets. As input to the API, the user can enter filtering rules based on the

[1] https://developer.twitter.com/en/docs/tutorials/consuming-streaming-data.

available information like hashtags, keywords, phrases, and location. The module outputs all the tweets that match these defined cases as soon as they are live on social media. Multiple rules can be defined to extract tweets for several disasters at the same time. Data from any social media platform can be used as input to the proposed framework. However, in this work, we consume disaster-related posts on Twitter. Furthermore, although the proposed system is explicitly for multimodal tweets having both images and text, we let the streaming module filter both unimodal and multimodal disaster tweets. We discuss in Sect. 5.5 how our pipeline can be generalized to process unimodal tweets as well, making it more robust.

3.2 Remove Redundant Tweets

A large proportion of the tweets obtained using the streaming module may be retweets that have already been processed by the system. Therefore, to avoid overheads, we maintain a list of identifiers (IDs) of all tweets that have been processed by the system. In case an incoming tweet is a retweet that has already been processed by the system before, we discard it. Furthermore, some tweets may also have location or geographic information. This information is also stored to maintain a list of places where relief groups are already providing services currently. If a streamed geo-tagged tweet is from a location where the relief groups are already providing aid, the tweet is not processed further.

3.3 Relevance Filtering

A substantial number of tweets streamed from the social media platforms are likely to be irrelevant for disaster response and management. Furthermore, different relief groups have varying criteria for what is relevant to them for responding to the situation. For instance, a particular relief group could be interested only in reaching out to the injured victims, while another provides resources for infrastructural damages. Therefore, for them to make proper use of information from social media platforms, the relevant information must be filtered.

We propose two sub-modules for filtering: (*i*) the first filters the informative tweets, *i.e.,* the tweets that provide information relevant to a disaster, which could be useful to a relief group, (*ii*) the second filter is specific to the relief group, based on the type of damage response they provide. To demonstrate the system, in this work, we filter tweets that indicate infrastructural damage or physical damage in buildings and other structures.

3.4 Severity Analysis and Resource Allocation

Finally, once the relevant tweets have been filtered, we analyze them for the severity of the damage indicated. The system categorizes the severity of infrastructural damage into three levels: high, medium and low. Based on the damage severity assessment by the system, the relief group can provide resources and services to a particular location. This information must further be updated in the database storing the information about all the places where the group is providing aid currently. Furthermore, although not shown in the system diagram, we must also remove a location from the database once the relief group's activity is over, and it is no longer actively providing service there. This ensures that if there is an incoming request from that location after it was removed from the database, it can be entertained.

4 Methodology

In this section, we discuss the implementation details of the two main modules of the system for Relevance Filtering and Severity Analysis. We begin by describing the data pre-processing required for the multimodal tweets, followed by the deep learning-based models that we use for the modules.

4.1 Pre-processing

Image Pre-processing: The images are resized to 299×299 for the transfer learning model [29] and then normalized in the range $[0, 1]$ across all channels (RGB).

Text Pre-processing: All *http* URLs, retweet headers of the form *RT*, punctuation marks, and twitter user handles specified as *@username* are removed. The tweets are then lemmatized and transformed into a stream of tokens that can be fed as input to the models used in the downstream modules. These tokens act as indices to an embedding matrix, which stores the vector representation for tokens corresponding to all the words maintained in the vocabulary. In this work, we use 100 dimensional FastText word-embeddings [6], trained on the CrisisMMD dataset [1] that has been used in this work. The system as a whole, however, is independent of the choice of vector representation.

4.2 Unimodal Models

For the proposed pipeline, we use Recurrent Convolutional Neural Network (RCNN) [14] as the text classification model. It adds a recurrent structure to the convolutional block, thereby capturing contextual information with long term dependencies and the phrases which play a vital role at the same time. Furthermore, we use the Inception-v3 model [25], pre-trained on the ImageNet Dataset [9] for modelling the image modality. The same underlying architectures, for both text and image respectively, are used to filter the tweets that convey useful information regarding the presence of infrastructural damage in the Relevance Filtering modules, and the analysis of damage in the Severity Analysis module. Therefore, we effectively have three models for each modality: first for filtering the informative tweets, then for those pertaining to the infrastructural damage (or any other category related to the relief group), and finally for assessing the severity of damage present.

4.3 Combining Modalities

In this subsection, we describe how we combine the unimodal predictions from the text and image models for different modules. We also discuss in each case about how the system would treat a unimodal text or image only input tweet.

Gated Approach for Relevance Filtering. For the two modules within Relevance Filtering, we use a simplistic approach of combining the outputs from the text and image models by using the OR function ($\oplus$). Technically speaking, we conclude that the combined output is positive if at least one of the unimodal models predicts so. Therefore, if a tweet is predicted as informative by either the text, or the image, or both the models, the system predicts the tweet as informative, and it is considered for further processing in the pipeline. Similarly, if at least one of the text and the image modality predicts an informative tweet as containing infrastructural damage, the tweet undergoes severity analysis. This simple technique helps avoid missing any tweet that might have even the slightest hint of damage, in either or both the modalities. Any false positive can also be easily handled in this approach. If, say, a non-informative tweet is predicted as informative in the first step at Relevance Filtering, it might still be the case that in the second step, the tweet is predicted as not containing any infrastructural damage. Furthermore, in case a tweet is unimodal and has just the text or the image, then the system can take the default prediction of the missing modality as negative (or `False` for a boolean OR function), which is the identity for the OR operation. In that case, the prediction based on the available modality will guide the analysis (Fig. 2).

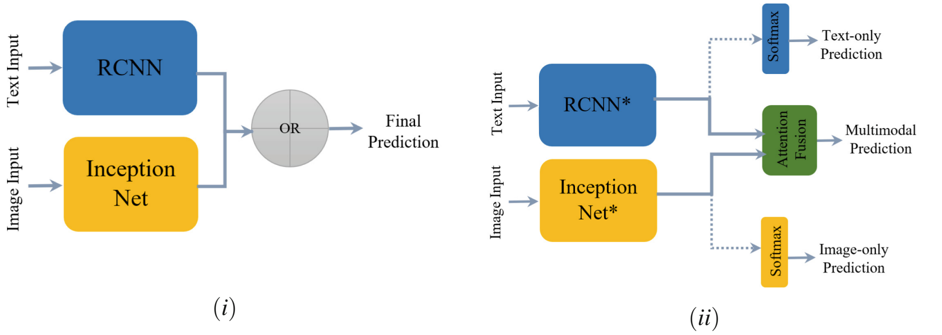


Fig. 2. Internal architecture of the (*i*) Relevance Filtering modules using an OR function to combine the predictions of the unimodal text and image models. (*ii*) Severity Analysis module that uses attention fusion to combine the text and image modalities, when both available, and switching to the unimodal models when either is missing. * indicates the model architecture till the penultimate layer, excluding the softmax.

Attention Fusion for Severity Analysis. The availability of data from different media sources has encouraged researchers to explore and leverage the potential boost in performance by combining unimodal classifiers trained on individual modalities [5,27]. Here, we use attention fusion to combine the feature interpretations from the text and image modalities for the severity analysis module [12,26]. The idea of attention fusion is to attend particular input features as compared to others while predicting the output class. The features, *i.e.,* the outputs of the penultimate layer or the layer before the softmax, of the text and image models are concatenated. This is followed by a softmax layer to learn the attention weights for each feature dimension, i.e., the attention weight α_i for a feature x_i is given by:

$$\alpha_i = \text{softmax}(\sum_{j=1}^{p} W_{ji} \cdot x_j) = \frac{exp(\sum_{j=1}^{p} W_{ji} \cdot x_j)}{\sum_{i=1}^{p} exp(\sum_{j=1}^{p} W_{ji} \cdot x_j)} \tag{1}$$

Therefore, the input feature after applying the attention weights is,

$$\beta_i = \alpha_i \cdot x_i \tag{2}$$

where, $i, j \in 1, 2, .., p$, and p is the total number of dimensions in the multimodal concatenated feature vector. W is the weight matrix learned by the model. This vector of attended features is then used to classify the given multimodal input. With this type of fusion, we can also analyze how the different modalities are interacting with each other employing their attention weights. Moving from the Relevance Filtering to the Severity Analysis module, we strengthen our fusion technique by using attention mechanism. This is required since human resources are almost always scarce, and it is necessary to correctly assess the requirements at different locations based on the severity of the damage. As opposed to an OR function, using attention, we are able to combine the most important information as seen by the different modalities to together analyze the damage severity.

In this case, the treatment of unimodal tweets is not that straightforward, since the final prediction using attention fusion occurs after concatenation of the latent feature vectors of the individual modalities. Therefore, in case the text or image is missing, we use the unimodal model for the available modality. In other words, we use attention mechanism only when both the modalities are present to analyze damage severity, else we use the unimodal models.

5 Evaluation

5.1 Dataset

Recently, several datasets on crisis damage analysis have been released to foster research in the area [21]. In this work, we have used the first multimodal, labeled, publicly available damage related to the Twitter dataset, CrisisMMD, created by Alam *et al.* [1]. It was collected by crawling the blogs posted by users during seven natural disasters, which can be grouped into 4 disaster categories, namely-Floods, Hurricanes, Wildfires and Earthquakes. CrisisMMD introduces three hierarchical tasks:

1. **Informativeness.** This initial task classifies each multimodal post as informative or non-informative. Alam *et al.* define a multimodal post as informative if it serves to be useful in identifying areas where damage has occurred due to disaster. It is therefore a binary classification problem, with the two classes being informative and non-informative.
2. **Infrastructural Damage.** The damage in an informative tweet may be of many different kinds [1,4]. CrisisMMD identifies several categories for the type of damage, namely- Infrastructure and utility damage, Vehicle damage, Affected individuals, Missing or found people, Other relevant information, None. Alam *et al.* [1] also noted that the tweets which signify physical damage in structures, where people could be stuck, are especially beneficial for the rescue operation groups to provide aid. Out of the above-listed categories, the tweets having Infrastructure and utility damage are therefore identified in this task. This again is modelled as a classification problem with two classes-infrastructural and non-infrastructural damage.
3. **Damage Severity Analysis.** This final task uses the text and image modalities together to analyze the severity of infrastructural damage in a tweet as-high, medium, or low. We add another label, no-damage, to support the pipeline framework that can handle false positives as well. Specifically, if a tweet having no infrastructural damage is predicted as positive, it can be detected here as having no damage. This is modelled as a multi-class classification problem.

The individual modules of the proposed pipeline essentially model the above three tasks of CrisisMMD. Specifically, the two Relevance Filtering modules model the first and the second tasks, respectively, whereas the Severity Analysis module models the third task (Table 1).

Table 1. CrisisMMD Class Distribution: For Tasks 1 and 2, the text and images have been labeled separately, and therefore, the number of samples in the respective classes are separated by a / *i.e.,* `text/image`. Task 3 has a single tweets level label signifying the severity of damage.

Task 1	**Informative**			**Non informative**
	12877/9375			5249/8751
Task 2	**Infrastructural**			**Non infrastructural**
	1428/3624			16698/14502
Task 3	**Low**	**Mild**	**Severe**	**No-damage**
	566	842	2216	14502

5.2 Experimental Settings

To evaluate how well our system can generalize to new disaster categories, we train our models for all the three tasks in a leave-one-disaster-out (LODO) training paradigm. Therefore, we train on 3 disaster categories and evaluate the performance on the left-out disaster. To handle the class imbalance, we also used SMOTE [8] with the word embeddings of the training fold samples for linguistic baselines. We used Adam Optimizer with an initial learning rate of 0.001, the values of $\beta1$ and $\beta2$ as 0.9 and 0.999, respectively, and a batch size of 64 to train our models. We use F1-Score as the metric to compare the model performance. All the models were trained on a GeForce GTX 1080 Ti GPU with a memory speed of 11 Gbps.

5.3 Results

To demonstrate the effectiveness of the proposed system for multimodal damage assessment on social media, we perform an ablation study, the results for which have been described below.

Design Choices. We tried different statistical and deep learning techniques for modelling text- TF-IDF features with **SVM**, Naive Bayes (**NB**) and Logistic Regression (**LR**); and in the latter category, **CNN** [13], Hierarchical Attention model (**HAttn**), bidirectional LSTM (**BiLSTM**) and **RCNN** [14]. As input to the deep learning models, we use 100-dimensional Fasttext word embeddings [6] trained on the dataset. By operating at the character n-gram level, Fasttext tends to capture the morphological structure well. Thus, helping the otherwise out of vocabulary words (such as hash-tags) to share semantically similar embeddings with its component words. As shown in Table 2, the RCNN model performed the best on all three tasks of the Relevance Filtering and Severity Analysis modules. Specifically, the average LODO F1-Scores of RCNN on the three tasks are 0.82, 0.76, and 0.79, respectively. Furthermore, the architecture considerably reduces the effect of noise in social media posts [14].

Table 2. Text and Image Baselines: F1-Scores for Leave one disaster out evaluation

		Unimodal Text Baselines							Unimodal Image Baselines			
Module	**Disaster Category**	**SVM**	**NB**	**LR**	**CNN**	**HAttn**	**BiLSTM**	**RCNN**	**CNN**	**VGG-16**	**Res-50**	**IncV3**
Relevance Filtering-1	**Floods**	0.47	0.43	0.48	0.64	0.69	0.62	**0.70**	0.31	0.63	0.75	**0.77**
	Hurricanes	0.42	0.51	0.49	0.71	**0.79**	0.73	**0.78**	0.29	0.61	**0.71**	**0.71**
	Wildfires	0.52	0.47	0.51	0.76	0.80	0.71	**0.81**	0.33	0.56	0.65	**0.75**
	Earthquakes	0.43	0.54	0.59	0.69	0.77	0.70	**0.78**	0.35	0.65	0.67	**0.73**
Relevance Filtering-2	**Floods**	0.51	0.46	0.53	**0.70**	0.65	0.68	**0.70**	0.67	0.66	0.69	**0.70**
	Hurricanes	0.47	0.54	0.50	0.69	0.72	0.65	**0.75**	0.61	0.61	0.72	**0.74**
	Wildfires	0.49	0.42	0.47	0.74	0.60	0.75	**0.79**	0.77	0.79	**0.81**	**0.81**
	Earthquakes	0.41	0.40	0.44	0.76	0.81	0.71	**0.83**	0.78	0.81	0.80	**0.82**
Severity Analysis	**Floods**	0.54	0.48	0.50	0.68	**0.80**	0.73	**0.79**	0.71	0.73	0.77	**0.81**
	Hurricanes	0.51	0.45	0.48	0.71	0.76	0.70	**0.82**	0.72	0.72	0.73	**0.80**
	Wildfires	0.43	0.49	0.48	0.75	**0.81**	0.74	**0.80**	0.69	0.72	0.79	**0.79**
	Earthquakes	0.41	0.47	0.42	0.66	0.71	0.74	**0.76**	0.65	0.71	0.74	**0.76**

For images, we fine-tuned the **VGG-16** [23], **ResNet-50** [11] and **InceptionV3** [25] models, pre-trained on the ImageNet Dataset [9]. We also trained a CNN model from scratch. Experimental results in Table 2 reveal that InceptionV3 performed the best, and the average F1-Score with LODO training for the three tasks are 0.74, 0.77, and 0.79, respectively. The architecture employs multiple sized filters to get a thick rather than a deep architecture, as very deep networks are prone to over-fitting. Such a design makes the network computationally less expensive, which is a prime concern for our system as we want to minimize latency to give quick service to the disaster relief groups.

Ablation Study. Table 3 highlights the results of an ablation study over the best linguistic and vision models, along with the results obtained when the predictions by these individual models are combined as discussed in Sect. 4.3. The results for all the modules demonstrate the effectiveness of multimodal damage assessment models. Specifically, we observe that for each disaster category in the LODO training paradigm, the F1-Score for the multimodal model is always better than or compares with those of the text and image unimodal models.

5.4 Qualitative Analysis

In this section, we analyze some specific samples to understand the shortcomings of using unimodal systems, and to demonstrate the effectiveness of our proposed multimodal system. Table 4 records these sample tweets along with their predictions as given by the different modules. In **green** are the correct predictions, whereas the incorrect ones are shown in **red** They have been discussed below in order:

1. The image in the first sample portrays the city landscape from the top, damaged by the calamity. Due to the visual noise, the image does not give much information about the intensity of damage present, and therefore, the image model incorrectly predicts the tweet as mildly damaged. On the other hand, the text model can identify the severe damage indicated by phrases like 'hit hard'. Combining the two predictions by using attention fusion, therefore, helps in overcoming the unimodal misclassifications.

Table 3. Ablation Study: Leveraging both textual and visual cues helps improve the leave one disaster out model performance in terms of F1-Score.

Module	Technique	Floods	Hurricanes	Wildfires	Earthquakes
Relevance Filtering-1	**Unimodal Text RCNN**	0.70	**0.78**	0.81	0.78
	Unimodal Image IncV3	0.77	0.71	0.75	0.73
	Text $\oplus$ Image	**0.79**	**0.78**	**0.83**	**0.79**
Relevance Filtering-2	**Unimodal Text RCNN**	0.70	0.75	0.79	**0.83**
	Unimodal Image IncV3	0.70	0.74	0.81	0.82
	Text $\oplus$ Image	**0.73**	**0.77**	**0.83**	**0.83**
Severity Analysis	**Unimodal Text RCNN**	0.79	**0.82**	0.80	0.76
	Unimodal Image IncV3	0.81	0.80	0.79	0.76
	Attention Fusion	**0.84**	**0.82**	**0.85**	**0.86**

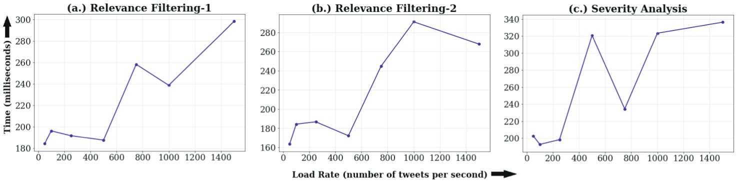


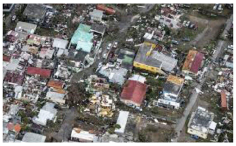
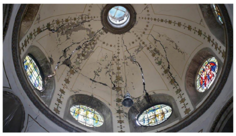
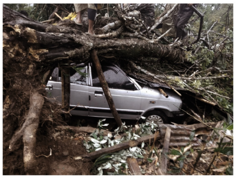
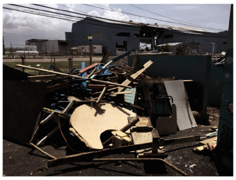
Fig. 3. Latency analysis

2. In this tweet, the text uses several keywords, such as 'damaged' and 'earthquake', which misleads the text model in predicting it as severely damaged. However, the image does not hold the same perspective. By combining the feature representations, attention fusion can correctly predict the tweet as having mild damage.

3. The given tweet is informative and therefore, it is considered for damage analysis. However, the text classifier, despite the presence of words like 'killed' and 'destroyed', incorrectly classifies it to the non-infrastructural damage class. The image classifier correctly identifies the presence of damage, and therefore, the overall prediction for the tweet is infrastructural damage, which is correct. Furthermore, both the text and image models are unable to identify the severity of damage present, but the proposed system can detect the presence of severe damage using attention fusion.

4. The sample shows how the Severity Analysis module combines the text and visual cues by identifying and attending to more important features than others. This helps in modelling the dependency between the two modalities, even when both, individually give incorrect predictions. The image in the tweet shows some hurricane destroyed structures, depicting severe damage. However, the text talks about 'raising funds and rebuilding', which does not indicate severe damage. The multimodal system learns to attend the text features more and correctly classifies the sample as having no damage, even though both the individual models predicted incorrectly. Furthermore, in this particular example, even by using the OR function, the system could not correctly classify it as not having infrastructural damage. Yet, the damage Severity Analysis module identifies this false positive and correctly classifies it.

Table 4. Qualitative Analysis: **T** and **I** indicate predictions by the text and image unimodal models, respectively.

Tweet	Relevance Filtering-1	Relevance Filtering-2	Severity Analysis
 **1.** *RT @UMDCSA: Dominica was hit hard from Hurricane Maria. Please keep them in your thoughts and prayers https://t.co/SpKYztnltV*	**I:** Informative **T:** Informative **T ⊕ I:** Informative	**I:** Infrastructural **T:** Infrastructural **T ⊕ I:** Infrastructural	**I:** Mild **T:** Severe **Attn Fusion:** Severe
 **2.** *RT @latimes: 2,000 historic buildings in Mexico have been damaged by the earthquake https://t.co/57pb1Pse1o https://t.co/7LjHIp3ljB*	**I:** Informative **T:** Informative **T ⊕ I:** Informative	**I:** Infrastructural **T:** Infrastructural **T ⊕ I:** Infrastructural	**I:** Mild **T:** Severe **Attn Fusion:** Mild
 **3.** *Cyclone Mora : 4 killed in Manipur, 140 houses destroyed in Mizoram :: https://t.co/dMlEngezZ4*	**I:** Informative **T:** Informative **T ⊕ I:** Informative	**I:** Infrastructural **T:** Non-infrastructural **T ⊕ I:** Infrastructural	**I:** Mild **T:** No-damage **Attn Fusion:** Severe
 **4.** *https://t.co/lLQwAQ4zhk Help us raise funds to rebuild our playgrounds after Hurricane Maria https://t.co/chCGXCltp2*	**I:** Informative **T:** Informative **T ⊕ I:** Informative	**I:** Infrastructural **T:** Infrastructural **T ⊕ I:** Infrastructural	**I:** Severe **T:** Severe **Attn Fusion:** No-damage

5.5 Discussion

In this section, we discuss some of the practical and deployment aspects of our system, as well as some of its limitations.

Latency Analysis. We simulate an experiment to analyze the computational efficiency of the individual modules in terms of the time they take to process a tweet, *i.e.,* the latency. We are particularly interested in analyzing the Relevance Filtering and Severity Analysis modules. We developed a simulator program to act as the Tweet Streaming module that publishes tweets at different load rates (number of tweets in 1 second) to be processed by the downstream modules. The modules also process the incoming tweets at the same rate. We calculate the average time for processing a tweet by a particular module as the total processing time divided by the total number of tweets used in the experiment. We used 15, 000 multimodal tweets from CrisisMMD, streamed at varying rates. The performance of the two Relevance Filtering modules and the Severity Analysis module as we gradually increase the load rate is shown in the Fig. 3.

As a whole, including all the modules, we observed that on an average, the system can process 80 tweets in 1 minute. This experiment was done using an Intel i7-8550U CPU having 16 GB RAM. One can expect to see an improvement if a GPU is used over a CPU.

Generalization. The proposed system is also general and robust, especially in three aspects. Firstly, the results of our LODO experiments indicate that the system can perform well in case it is used for analyzing new disasters, which were not used for training the system. This makes it suitable for real-world deployment where circumstance with new disaster categories cannot be foreseen. Furthermore, we also saw how the two main modules of the system work seamlessly, even when one of the modalities is missing. This ensures that the system can utilize all the information that is available on the media platforms to analyze the disaster. Finally, the second module in Relevance Filtering can be trained to suit the needs of several relief groups that target different types of damage, and therefore, the system is capable of being utilized for many different response activities.

Limitations. Although the proposed system is robust and efficient, some limitations must be considered before it can be used in real-time. Firstly, the system is contingent on the credibility *i.e.,* the veracity of the content shared by users on social media platforms. It may so happen that false information is spread by some users to create panic amongst others [15]. In this work, we have not evaluated the content for veracity, and therefore, it will not be able to differentiate such false news media. Another aspect that is also critical to all systems that utilize data generated on social media is the socio-economic and geographic bias. Specifically, the system will only be able to get information about the areas where people have access to social media, mostly the urban cities, whereas damage in the rural locations may go unnoticed since it did not appear on Twitter or any other platform. One way to overcome this is to make use of aerial images, that can provide a top view of such locations as the rural lands. However, this again has a drawback as to utilize aerial images effectively, a bulk load of data would have to be gathered and processed.

6 Conclusion

Identifying damage and human casualties in real-time from social media posts is critical to providing prompt and suitable resources and medical attention, to save as many lives as possible. With millions of social media users continuously posting content, an opportunity is present to utilize this data and learn a damage recognition system. In this work, we propose MEMIS, a novel Multimodal Emergency Management Information System for identifying and analyzing the level of damage severity in social media posts with the scope for betterment in disaster management and planning. The system leverages both textual and visual cues to automate the process of damage identification and assessment from social media data. Our results show how the proposed multimodal system outperforms the state-of-the-art unimodal frameworks. We also report the system's responsiveness through extensive system analysis. The leave-one-disaster-out training setting proves the system is generic and can be deployed for any new unseen disaster.

References

1. Alam, F., Ofli, F., Imran, M.: CrisisMMD: multimodal twitter datasets from natural disasters. CoRR abs/1805.00713 (2018). http://arxiv.org/abs/1805.00713
2. Alam, F., Ofli, F., Imran, M.: Processing social media images by combining human and machine computing during crises. Int. J. Hum. Comput. Interact. **34**, 311–327 (2018)
3. Alam, F., Ofli, F., Imran, M.: CrisisDPS: crisis data processing services. In: Proceedings of the 16th ISCRAM Conference (2019)
4. Alam, F., Ofli, F., Imran, M., Aupetit, M.: A Twitter tale of three hurricanes: Harvey, Irma, and Maria. ArXiv abs/1805.05144 (2018)
5. Asvadi, A., Garrote, L., Premebida, C., Peixoto, P., Nunes, U.J.C.: Multimodal vehicle detection: fusing 3D-lidar and color camera data. Pattern Recognit. Lett. **115**, 20–29 (2017)
6. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. CoRR abs/1607.04606 (2016). http://arxiv.org/abs/1607.04606
7. Chaudhuri, N., Bose, I.: Application of image analytics for disaster response in smart cities, pp. 3036–3045 (2019). http://hdl.handle.net/10125/59740
8. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**(1), 321–357 (2002). http://dl.acm.org/citation.cfm?id=1622407.1622416

9. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009). https://doi.org/10.1109/CVPR.2009.5206848
10. Forbes: How much data do we create every day? The mind-blowing stats everyone should read (2018). Accessed 26 Apr 2019
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
12. Hua, X.-S., Zhang, H.-J.: An attention-based decision fusion scheme for multimedia information retrieval. In: Aizawa, K., Nakamura, Y., Satoh, S. (eds.) PCM 2004. LNCS, vol. 3332, pp. 1001–1010. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30542-2_123
13. Kim, Y.: Convolutional neural networks for sentence classification. CoRR abs/1408.5882 (2014). http://arxiv.org/abs/1408.5882
14. Lai, S., Xu, L., Liu, K., Zhao, J.: Recurrent convolutional neural networks for text classification. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI 2015, pp. 2267–2273. AAAI Press (2015). http://dl.acm.org/citation.cfm?id=2886521.2886636
15. Mahata, D., Talburt, J.R., Singh, V.K.: From chirps to whistles: discovering event-specific informative content from Twitter. In: Proceedings of the ACM Web Science Conference, p. 17. ACM (2015)
16. Mouzannar, H., Rizk, Y., Awad, M.: Damage identification in social media posts using multimodal deep learning. In: Proceedings of the 15th ISCRAM Conference (2018)
17. Nanni, L., Ghidoni, S., Brahnam, S.: Handcrafted vs. non-handcrafted features for computer vision classification. Pattern Recognit. **71**, 158–172 (2017)
18. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2009)
19. Pouyanfar, S., Tao, Y., Tian, H., Chen, S.C., Shyu, M.L.: Multimodal deep learning based on multiple correspondence analysis for disaster management. World Wide Web **22**(5), 1893–1911 (2019). https://doi.org/10.1007/s11280-018-0636-4
20. Rizk, Y., Jomaa, H.S., Awad, M., Castillo, C.: A computationally efficient multi-modal classification approach of disaster-related Twitter images. In: SAC (2019)
21. Said, N., et al.: Natural disasters detection in social media and satellite imagery: a survey. Multimed. Tools Appl. **78**(22), 31267–31302 (2019). https://doi.org/10.1007/s11042-019-07942-1
22. Shah, R., Zimmermann, R.: Multimodal Analysis of User-generated Multimedia Content. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61807-4
23. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014). https://arxiv.org/abs/1409.1556
24. Sreenivasulu, M., Sridevi, M.: Detecting informative Tweets during disaster using deep neural networks. In: 2019 11th International Conference on Communication Systems & Networks (COMSNETS), pp. 709–713 (2019)
25. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
26. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)

27. Wang, X., Gong, G., Li, N.: Multimodal fusion of EEG and fMRI for epilepsy detection. IJMSSC **9**, 1850010 (2017)
28. Yang, Y., et al.: MADIS: a multimedia-aided disaster information integration system for emergency management. In: 8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), pp. 233–241 (2012). https://doi.org/10.4108/icst.collaboratecom.2012.250525
29. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems, pp. 3320–3328 (2014)