

---

# CSCI 566 Final Project Proposal

---

**Divyansh Aggarwal**  
da05628@usc.edu

**Patrycja Krawczuk**  
krawczuk@usc.edu

**Nachiket Lele**  
nlele@usc.edu

**Shubham Nagarkar**  
slnagark@usc.edu

**Pushkaraj Sarnobat**  
sarnobat@usc.edu

## 1 Goals

During the course of the project, we aim to accomplish the following goals:

- Use transfer learning and semi-supervised learning to create a richly annotated dataset for disaster events related image and text classification tasks.
- Use images and textual data (tweets) to build joint representation of disaster events.
- Use multi-modal data collected from Social Media (SM) platforms like Twitter to rapidly provide relevant information about disaster events ('situation awareness').
- Propose a novel approach to automatically generate a textual and visual report of a disaster event.

## 2 Motivation

An increasing number of people use Social Media (SM) platforms like Twitter and Instagram to report information about critical emergencies or disaster events. Multimodal data shared on these platforms often contain useful information about a scale of the event, victims and infrastructure damage. This data can provide local authorities and humanitarian organizations with a big picture understanding of the emergency. Moreover, it can be used to effectively and timely plan relief responses.

In a recent editorial [5], Imran et al., discuss the challenges and opportunities in the use of multimodal SM content for disaster response. The authors consider the automatic generation of a coherent summary report of a disaster event as an important research question. They note a need for a system that can aggregate different modalities of data and provide a comprehensive picture of a disaster event. Addressing this problem has been impeded by the lack of large-scale annotated datasets.

In our project, we aim to address this open research question. First, we plan to mitigate the issue of data scarcity by combining several recently published disaster-related datasets and propagate the labels between them. Then, we plan to merge textual and pictorial modalities to build rich representations of disaster events. Finally, we aim to generate a textual and visual summary of a disaster event.

## 3 Problem Formulation

The proposed problem is two-fold:

- First, employ machine learning techniques like transfer learning, contrastive learning and semi-supervised learning to build a richly annotated, large-scale, multi-modal disaster events related dataset.
- Secondly, use the data to generate a textual and visual report summary of a disaster event.

### 3.1 Data

We plan to use a set of recently published, human-annotated, disaster-related datasets:

**Incidents Dataset [7]** consists of 446,684 scene-centric images annotated by humans as positive for natural disasters (class-positives), types of damage or specific events that can require human attention or assistance, like traffic jams or car accidents.

**CrisisMMD Dataset [1]** multi-modal Twitter corpora consisting of several thousands of manually annotated tweets and images collected during seven major natural disasters including earth-quakes, hurricanes, wildfires, and floods that happened in the year 2017 across different parts of the World. The CrisisMMD multimodal Twitter dataset consists of several thousands of manually annotated tweets and images collected during seven major natural disasters including earthquakes, hurricanes, wildfires, and floods that happened in the year 2017 across different parts of the World. The datasets include three types of annotations: nformative vs. Not Informative, Humanitarian Categories and Damage Severity Assessment.

**Twitter as a Lifeline [4]** human-annotated Twitter corpora for NLP of crisis-related messages This resource consists of Twitter data collected during 19 natural and human-induced disasters. Each dataset contains tweet-ids and human-labeled tweets of the event. Moreover, it contains a dictionary of out-of-vocabulary(OOV) words, a word2vec model, and a tweets downloader tool.

### 3.2 Primary Contribution

Our work has two main contributions: (i) It provides a novel dataset enriched with propagated labels and paired text-image tweets. (ii) It implements an original approach for the automatic generation of a comprehensive summary of a disaster event.

### 3.3 Difference from Existing Work

- In the earlier research [2][3], the datasets ([1],[7],[4]) are used individually and without any modifications. We create a novel enriched dataset by propagating labels via techniques like semi-supervised learning.
- In [2], the authors create descriptive and visual summaries of disaster events from the relevant tweets. The data, methodologies, and the content of the produced summaries substantially differ from the methods we plan to use. More importantly, the authors do not take advantage of the joint representation of image and text modalities.
- In [6], the authors work on comparing the baselines for unimodal and multimodal models using only the paired text-image tweets. We will combine the images and texts from the unpaired tweets to produce a dataset with caption images, thus ensuring better performance.

## 4 High-level Steps to Approach the Problem

Brief description of main functional blocks:

1. Text and Image Preprocessing: text tokenization, removal of duplicated tweets and more.
2. Text and Image Relevancy Calculation: use of a relevancy classifier to filter out irrelevant tweets and compute their relevancy scores before performing further analysis.
3. Text and Image Feature Extraction: extract features from text and images, create valid representations.
4. Multi-Modal Neural Network: match text and image representations, use fusion techniques to improve data classification
5. Categorization of Tweets: assign multiple categories of labels to each data point (informative, disaster type, damage severity etc.)
6. Disaster Summarization Report: based on the assigned labels and their confidence scores output a visual and textual summary for a given disaster type.

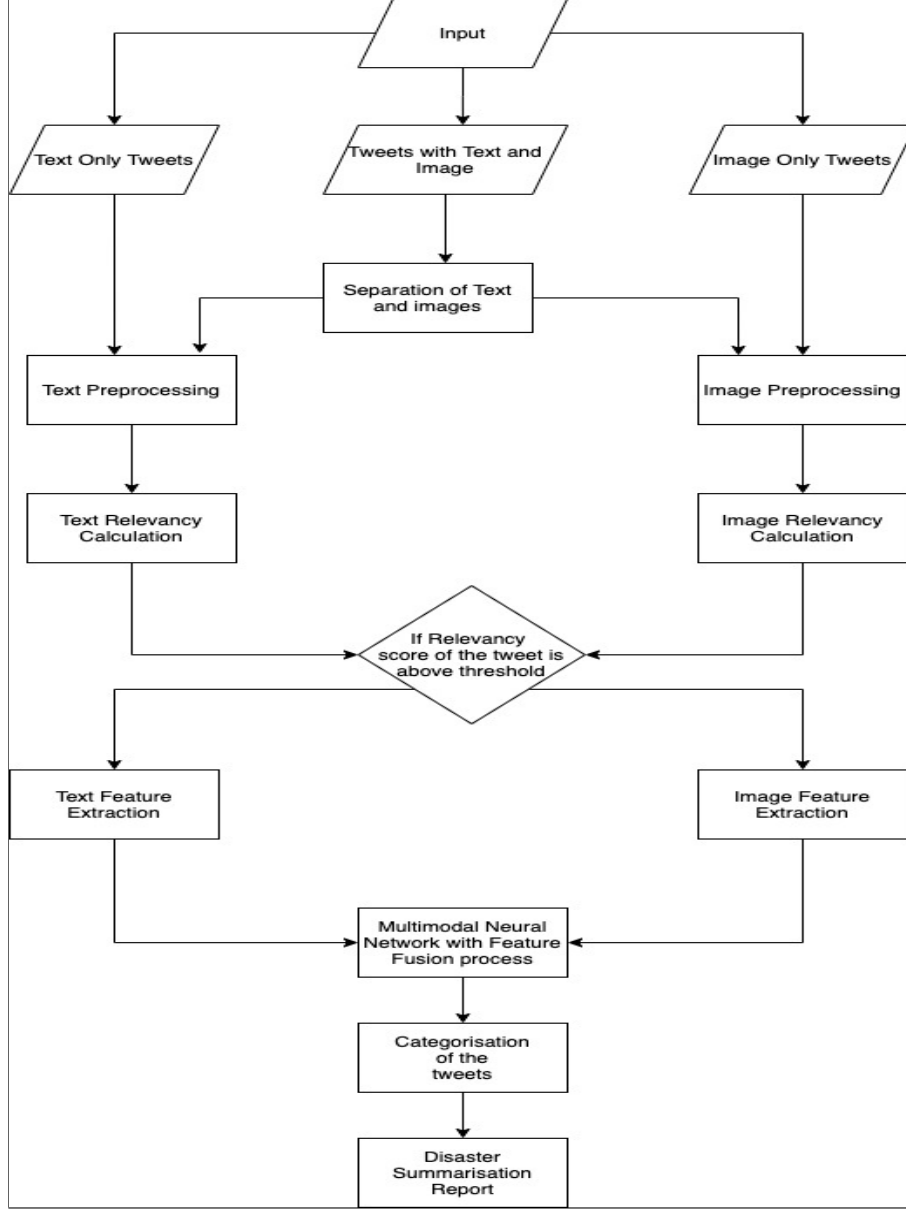


Figure 1: High-level Steps

## 5 Evaluation and Experiments

We plan to compare the performance of our approach with baseline models that use a single modality - image or text- for the data classification task (as shown here [6]). The metrics to be used for comparing the performance of the multimodal data with the unimodal data are *F-1 Score*, *Precision*, *Accuracy*, *Recall*. In addition, we plan to research and experiment with metrics that can evaluate the quality of our visual and descriptive summaries. One idea is to capture the percentage of the most relevant images and textual tweets from the whole dataset that are present and correctly classified in our summary report.

## 6 Expected Results

To mitigate the problem of situational awareness during disaster events, the proposed model summarizes the textual and visual tweets into a concise report.

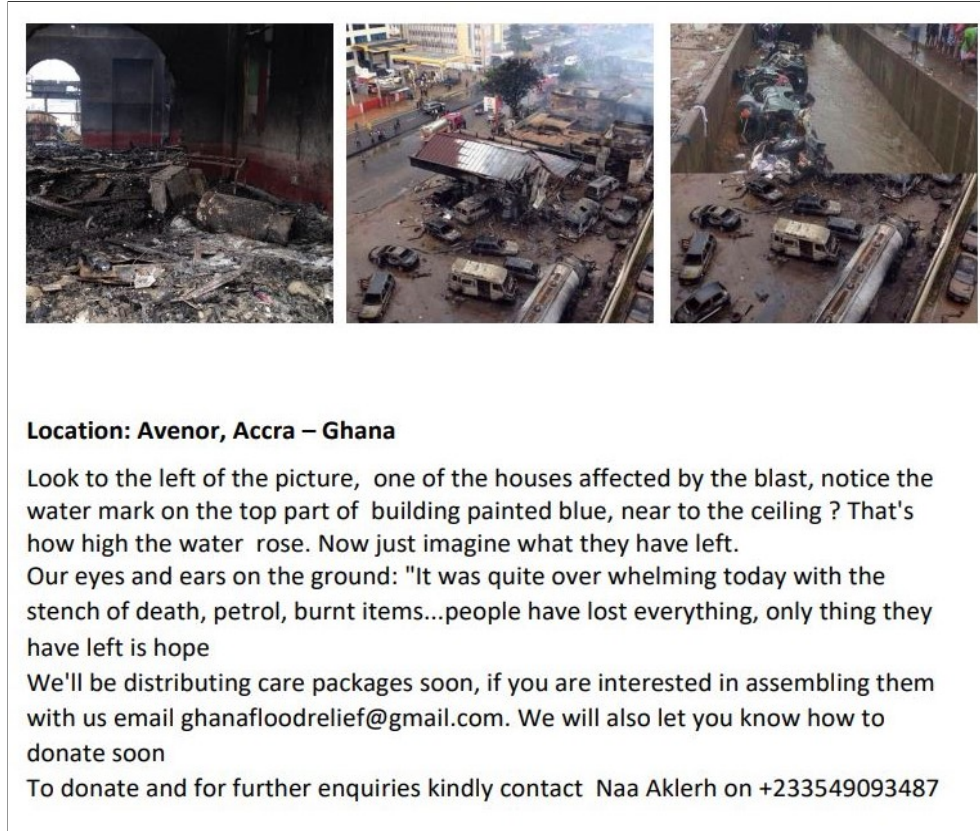


Figure 2: Disaster Summary

## 7 Milestones (Timeline)

1. Gather datasets relevant to disaster events.
2. Decide on the important labels to be propagated from one dataset to the other through weak-labeling.
3. Evaluate the quality of the propagated labels through clustering analysis.
4. Evaluate the relevancy scores of the preprocessed tweets.
5. Validate the categorization of the tweets and images based on severity and type of disaster.
6. Verify the credibility of the generated report.

## References

- [1] F. Alam, F. Ofli, and M. Imran. Crisismmd: Multimodal twitter datasets from natural disasters. *CoRR*, abs/1805.00713, 2018.
- [2] F. Alam, F. Ofli, and M. Imran. Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of hurricanes harvey, irma, and maria. *Behaviour Information Technology*, pages 1–31, 05 2019.

- [3] A. K. Gautam, L. Misra, A. Kumar, K. Misra, S. Aggarwal, and R. R. Shah. Multimodal analysis of disaster tweets. In *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, pages 94–103, 2019.
- [4] M. Imran, P. Mitra, and C. Castillo. Twitter as a lifeline: Human-annotated twitter corpora for nlp of crisis-related messages. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may 2016. European Language Resources Association (ELRA).
- [5] M. Imran, F. Ofli, D. Caragea, and A. Torralba. Using ai and social media multimodal content for disaster response and management: Opportunities, challenges, and future directions. *Information Processing Management*, 57:102261, 09 2020.
- [6] F. Ofli, F. Alam, and M. Imran. Analysis of social media data using multimodal deep learning for disaster response. *ArXiv*, abs/2004.11838, 2020.
- [7] E. Weber, N. Marzo, D. P. Papadopoulos, A. Biswas, A. Lapedriza, F. Ofli, M. Imran, and A. Torralba. Detecting natural disasters, damage, and incidents in the wild, 2020.