

# Relevancy Classification of Multimodal Social Media Streams for Emergency Services

Ganesh Nalluru  
George Mason University  
Fairfax, VA, USA  
gn@gmu.edu

Rahul Pandey  
George Mason University  
Fairfax, VA, USA  
rpandey4@gmu.edu

Hemant Purohit  
George Mason University  
Fairfax, VA, USA  
hpurohit@gmu.edu

**Abstract**—Social media has become an integral part of our daily lives. During time-critical events, the public shares a variety of posts on social media including reports for resource needs, damages, and help offerings for the affected community. Such posts can be relevant and may contain valuable situational awareness information. However, the information overload of social media challenges the timely processing and extraction of relevant information by the emergency services. Furthermore, the growing usage of multimedia content in the social media posts in recent years further adds to the challenge in timely mining relevant information from social media.

In this paper, we present a novel method for multimodal relevancy classification of social media posts, where relevancy is defined with respect to the information needs of emergency management agencies. Specifically, we experiment with the combination of semantic textual features with the image features to efficiently classify a relevant multimodal social media post. We validate our method using an evaluation of classifying the data from three real-world crisis events. Our experiments demonstrate that features based on the proposed hybrid framework of exploiting both textual and image content improve the performance of identifying relevant posts. In the light of these experiments, the application of the proposed classification method could reduce cognitive load on emergency services, in filtering multimodal public posts at large scale.

**Keywords**—Information Overload, Multimodal Data, Social Media, Emergency Management, Relevancy Classification

## I. INTRODUCTION

Social media plays a key role during disaster events, as evident from the events in recent years. People share all kinds of information, ranging from jokes and prayers to damage in the affected regions [1] to requests and offers to help [2]. Studies in recent years have also shown the public expectation for timely seeking a response to their calls for help on social media [3].

For emergency management agencies, getting relevant situational awareness information from the affected public is of utmost importance. In particular, the Public Information Officers (PIOs) monitors the information from public to provide intelligence to the decision makers in the response coordination [4]. Given the increasing importance of social media data, emergency services have started to monitor social media [5]. However, given their limited human resources and the vast amounts of social media messages posted with high velocity during disaster events, a critical challenge is to address the high information overload on the emergency

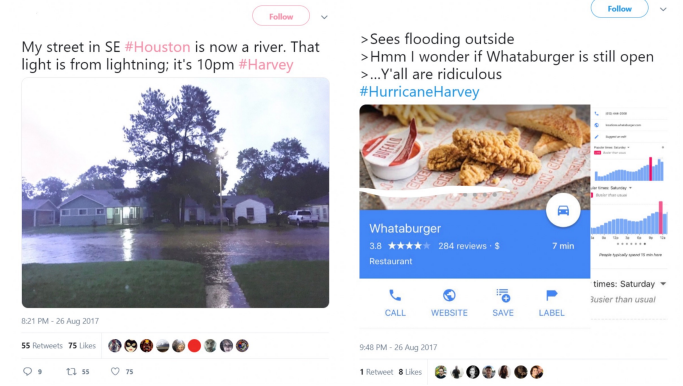


Fig. 1. Exemplar messages with different degree of relevance for textual and image characteristics, which provide informative vs non-informative signal to judge relevancy for the public post. (User name and id removed for anonymity)

service personnel [6]. There is a recognition of the necessity to effectively filter, prioritize, and organize information from this unconventional information channel [5].

Furthermore, social media messages are increasingly becoming multimodal in nature [7]. Thus, the informative data may be of different modalities other than text such as images, videos, and audios. The inclusion of data of other modalities highly affects the complexity of mining the informative messages that might be important to capture. Hence, exploiting features from all the modalities in the data can be valuable and necessary to efficiently understand the actual meaning of the message. Content of such posts vary in potential value for operational response, ranging from actionable and serviceable requests [8] to potential rumors [9] to damage reports [10]. Thus, quickly filtering and prioritizing such multimodal messages with relevant information have become a critical need for response agencies [5].

Figure 1 shows some example social media messages where the embedded image content can present valuable signal to quickly judge the relevancy for public social media posts for emergency management in contrast to the irrelevant information shared during emergency events.

**Problem.** We address the problem of classifying relevant multimodal social media posts for helping emergency services quickly filter social media streams.

**Our contribution.** We provide a generalizable multimodal classification framework and present extensive experimentation to study the interplay of textual and image features to classify relevant multimodal social media posts for emergency services. We introduce the framework in Section III together with a description of textual and image features, which leverages the popular distributional semantic representations of text and image data. Finally, we demonstrate the validity of the proposed method in Section IV, by experimenting with the real-world Twitter datasets collected during three crisis events in the recent years. We conclude with lessons and future work directions in Section V.

## II. RELATED WORK

There has been extensive research on the topic of social media for emergency management in the last decade [1], [6]. The nature of data generated over social media has such a high volume, variety, and velocity causing the challenges of “Big Crisis Data” that often overwhelm the emergency services [6]. The literature in crisis informatics [11] field has investigated social media for emergency services using diverse multidisciplinary perspectives.

User studies with emergency responders have identified information overload as one the key barrier for efficiently using social media platforms by PIOs and emergency services [6], [12]. Such information overload factors include the processing of unstructured and noisy nature of multimodal social media content at large scale, which is beyond the capacity of the limited human resources. Furthermore, characterizing the relevancy of social media content is very contextual, time-sensitive, and often challenging [8].

Among the social media analytics approaches, researchers have modeled public behavior in specific emergencies, addressed the problems of data collection and filtering, classification and summarization as well as visualization of analyzed data for decision support [1]. However, the focus of such works has centered around text analytics except recent studies [7], [10], [13], [14] on processing multimedia content of the social posts. Although current multimodal information processing approaches for social media mining during disasters primarily analyzed only the damage assessment aspect of emergency management. We further complement these recent studies by proposing a generic classification framework for relevant information that exploits both textual and image content of multimodal social media posts.

## III. METHOD

### A. Task Definition

Users in Twitter often share both text as well as image content during natural disasters. The goal of this task is to classify a tweet, if the tweet posted during the natural disaster is informative or not. The informativeness of a tweet is defined by the relevancy of embedded information to various emergency services or humanitarian organizations as described in [7], such as the message containing rescue related and affected individuals related information.

### B. Summary of the Proposed Approach

Learning features for classification of short text messages such as tweet is challenging due to limited availability of context to exploit. Often, these tweets require some context representation to extract the meaning of the text effectively. One possible solution is to use deep learning model for feature generation and text classification. However, not having the richness of large dataset results in overfitting of deep models as the features generated during the training are not generalizable enough due to the large number of variables to train. Therefore, we propose a feature generation framework that utilizes both the text features and image features to enrich the context representation from the given multimodal social media posts. We then perform the dimensionality reduction on three diverse extracted feature vectors and concatenate these vectors. Finally, we build a LightGBM model [15] for multimodal relevancy classification.

### C. Extracting Features

In this section, we describe the different methods used to derive the text features and image features.

1) *Text features:* We have utilized two different types of text features: *Bag of Words* features and *Word embeddings* features. Bag of Words features such as *tf-idf* vector representation of text give high weights to a word that often appears in particular tweet, but not as frequently in the whole collection of messages. Through *tf-idf*, we try to capture the semantic significance of that word to the document (tweet) in our corpus.

The following steps describe our approach to compute *tf-idf* vectors for each document/tweet.

- (i) First, we create a vocabulary set of size  $V$  for the entire training tweet set that contains a set of all the words present in all the tweets.
- (ii) Now for each tweet, we create a vector of size  $V$ . Each position of the vector represents a vocabulary word and each value represents the *tf-idf* value for that word in the tweet.
- (iii) To calculate *tf-idf* value for a word  $w$  in a tweet, we compute  $tf(w) = D_w/N$  where  $D_w$  is the number of times the word  $w$  appears in that tweet and we compute  $idf(w) = \log(N/D'_w)$ , where  $D'_w$  is the total number of tweets which contain the word  $w$ . Finally, the *tf-idf* score is the multiplication of  $tf(w)$  and  $idf(w)$ .

*Word embeddings* provide the semantic representation of words in low-dimensional space, where a dense, real-valued vector represents each word in a fixed dimensional space such that the words with similar meaning are closer to each other. We took the average of the embedding vectors of all the words in a tweet to describe the particular tweet. Since our data is small, we utilize the pre-trained GloVe vector embeddings [16] of 200 dimensions trained on Twitter data (27 billion tweets) to get the embedding vectors of each word.

Given that Bag-of-Words model with *tf-idf* representation does not capture the latent semantics of the tweet, thus, we

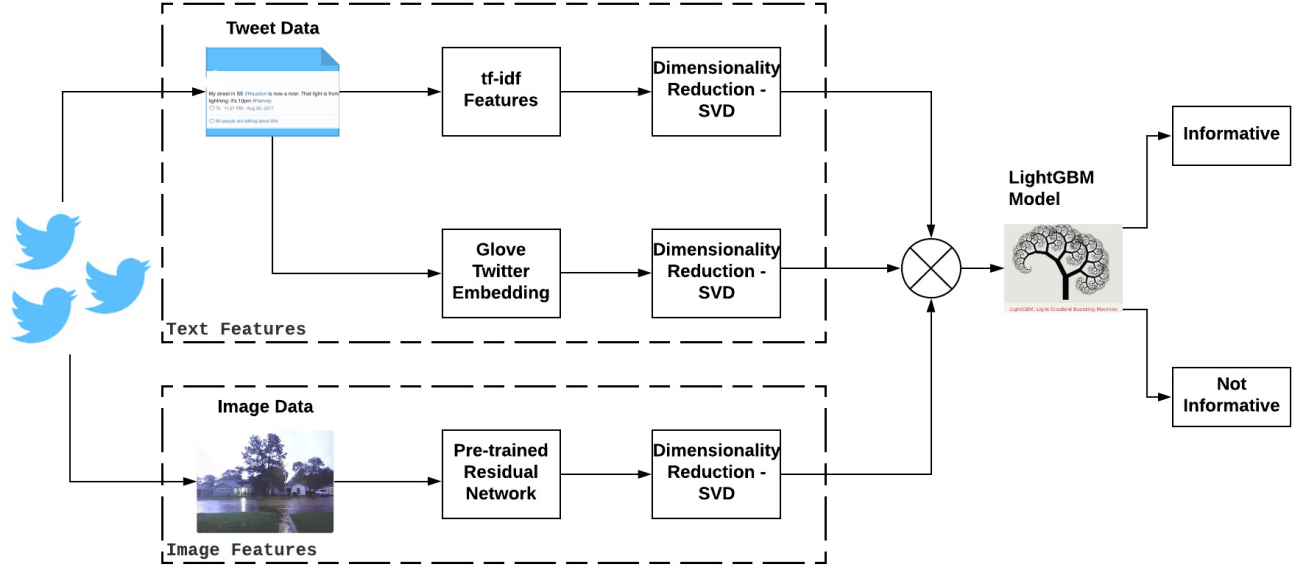


Fig. 2. Proposed multimodal classification framework that exploits both text and image content features trained on LightGBM model.

combined both the Bag-of-Words with tf-idf features and Glove-Twitter word Embeddings. Such feature combination captures both the syntactic and semantic aspects of the tweet, which is important for classifying the informativeness of the tweet.

The following steps describe our approach to compute word embedding vectors for each tweet:

- (i) For each tweet, we initialize a vector  $V$  with all zeros of size  $D$  where  $D$  is the dimension size for pretrained word embeddings.
- (ii) Now for each word in the tweet, we will check if we have the word embedding in our pretrained model. If it is present, we will perform the vector addition with the initial vector  $V$ .
- (iii) Finally, we will divide the final vector  $V$  with a scalar value  $n$  where  $n$  is the size of the subset of words in the tweet that have the embedding representation from the pre-trained model.

2) *Image features*: For generating *image features*, we have to address the sparsity issue due to a diverse range of images that are highly distributed in terms of embedded information. Hence, we utilized the pre-trained model of ResNet (Residual Learning Framework) [17] of 50 layer architecture to get the features for each image.

The following steps describe our approach to compute the image features for a given tweet:

- (i) First we set up the ResNet-50 architecture and initialize it with the pretrained model weights.
- (ii) We then pass each image in our training set as an input to this deep network.
- (iii) We store the second last layer vector for each of the image in the input training instance tweet separately.

- (iv) We use this vector as a *image features* for that training instance tweet.

3) *Feature Engineering*: As per our initial exploratory analysis on the dataset, we found few discriminative factors to differentiate *informative* vs *non-informative* tweets. First, we observed that all the tweets with *#news* were categorized as informative as compared to non-informative. Similarly, we observed that the length of the *informative* tweets were comparatively larger than that of *non-informative* tweets. As a result, we have included *Word Count* and *Character Count* as another set of features in our *text features* category.

4) *Combining Text Features and Image Features*: A key challenge with multimodal relevancy classification task is to find a systematic and effective way to combine the text features and image features. Also, the feature dimensions are different across the two modalities. Therefore, we used a Singular Valued Decomposition (SVD) to perform dimensionality reduction on all three types of extracted feature vectors and then, concatenated these vectors for the final feature representation for training the classification model.

#### D. Multimodal Disaster Data Classification

We have specifically used LightGBM (LGBM) [15] as our model for classification as it performed well on multimodal data in our preliminary testing with other models. Also, we found it be very useful particularly for time-sensitive emergency services after observing its faster performance in comparison to deep learning models. LGBM is an implementation of fast gradient boosting on decision trees. It is faster as it removes a significant proportion of data instances with small gradients, and includes only the rest to compute the information gain and also, bundle mutually exclusive features to reduce the number of features. For our baseline schemes,

we have compared our proposed model with two other models i.e. Logistic Regression and XG Boost [18].

#### IV. EXPERIMENTS AND RESULTS

##### A. Dataset

We conduct our experiments on three large disaster events from the year 2017 - Hurricane Maria, Hurricane Harvey, and Hurricane Irma, where the data was provided by [7]. The dataset contains Twitter posts with both text and their associated images as well as class label: *informative* or *non-informative*. The total *informative* and *non-informative* tweets across the events were 3095 and 984 for Harvey respectively, 3320 and 889 for Irma, and 2569 and 1528 for Maria respectively. We performed standard text preprocessing on tweets, including removing URLs, special symbols such as ‘RT @user’, stop words, and lowercasing.

##### B. Classification Schemes

We used 80-20 split strategy to create train-test sets and the AUC (Area Under the ROC Curve) and Accuracy measures to evaluate performance on the 20% test. We employ various classification schemes for training with different features and learning algorithms as follows ( $Mx$  denotes model type,  $Tx$  text feature type, and  $Ix$  image feature):

- **[T1 + M1] Text features (Bag of words) + Logistic Regression Model** (baseline) This method uses Bag-of-Words representation to generate text features (T1) and train that on Logistic Regression model.
- **[T1 + M2] Text features (Bag of words) + XGBoost Model** This method uses same features of T1 and train that on XGBoost model.
- **[T1 + M3] Text features (Bag of words) + LGBM Model** This method uses same features of T1 and train that on LGBM model.
- **[T2 + M1] Text features (tf-idf) + Logistic Regression Model** This method uses tf-idf to generate text features (T2) and train that on Logistic Regression model.
- **[T2 + M2] Text features (tf-idf) + XGBoost Model** This method uses same features of T2 and train that on XGBoost model.
- **[T2 + M3] Text features (tf-idf) + LGBM Model** This method uses same features of T2 and train that on LGBM model.
- **[T3 + M1] Text features (tf-idf + GloVe Embeddings) + Logistic Regression Model** This method uses the concatenation of both tf-idf and GloVe Twitter embeddings to generate text features (T3) and train that on Logistic Regression model.
- **[T3 + M2] Text features (tf-idf + GloVe Embeddings) + XGBoost Model** This method uses same features of T3 and train that on XGBoost model.
- **[T3 + M3] Text features (tf-idf + GloVe Embeddings) + LightGBM Model** This method uses same features of T3 and train that on LGBM model.
- **[T2 + I1 + M1] Text features (tf-idf) + Image Features (ResNet) + Logistic Regression Model** This method uses

the concatenation of both Tf-Idf as text features (T2) and ResNet features as Image features (I1) and train that on Logistic Regression model.

- **[T2 + I1 + M2] Text features (tf-idf) + Image Features (ResNet) + XGBoost Model** This method uses same features of T2 and I1, and train that on XGBoost model.
- **[T2 + I1 + M3] Text features (tf-idf) + Image Features (ResNet) + LGBM Model** This method uses same features of T2 and I1, and train that on LGBM model.
- **[T3 + I1 + M1] Text features (tf-idf + GloVe Embeddings) + Image Features (ResNet) + Logistic Regression Model** This method concatenates tf-Idf and GloVe Embeddings as text features and ResNet features as Image features, and train a Logistic Regression model.
- **[T3 + I1 + M2] Text features (tf-idf + GloVe Embeddings) + Image Features (ResNet) + XGBoost Model** This method uses same features of T3 and I1, and train that on XGBoost model.
- **[T3 + I1 + M3] Text features (tf-idf + GloVe Embeddings) + Image Features (ResNet) + LGBM Model** (proposed) This method uses same features of T3 and I1, and train that on LGBM model.

##### C. Results

Table I shows the results of our proposed features with LGBM model compared to the other baselines (XGBoost and Logistic Regression). The table shows the performance of different feature representations evaluated based on accuracy score and AUC. Our proposed approach is using both types of text features i.e.  $tf-idf$  and word embeddings as well as image features and trained on Light Gradient Boosting (LGBM). We have reported the results for three event datasets and analyzed how our proposed approach is performing compared to other baselines.

##### D. Discussions

We built several machine learning models based on our hybrid feature representation (combining  $tf-idf$ , Glove Embedding and Resnet features) and compared against the baseline of text-only features.

Table I shows that our proposed method with LightGBM Model outperformed other models in all the experiments on our hybrid feature representation of text and image features, across all disaster event datasets. We also noticed that the scheme of Logistic Regression model performed poorly when combining both  $tf-idf$  features (with an AUC of 74.2% in Hurricane Maria event) and image features (Resnet-50 features), giving us an AUC of 51.5% (similar case for other disaster events as well). It is likely due to simpler learning algorithm. In contrast, XGBoost Model performance on both  $tf-idf$  features and our proposed hybrid features respectively gave us almost the same performance. XGBoost Model with  $tf-idf$  features gave us an AUC of 74.97% in Hurricane Maria and AUC was nearly the same when we added image features (Resnet-50).

TABLE I

COMPARISON OF ACCURACY AND AUC FOR VARIOUS TRAINING SCHEMES ON THE 20% TEST SET. THE RESULTS FROM THE COMBINATION OF BOTH TEXT AND IMAGE FEATURES TRAINED ON LIGHTGBM OUTPERFORM OTHER SCHEMES ACROSS ALL DISASTER EVENTS.

Dataset	Input Features	[M1] LR		[M2] XG Boost		[M3] LightGBM	
		Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Hurricane Maria	[T1] Text features - Bag of Words	78.9%	76.5%	77.3%	74.95%	76.3%	81.61%
	[T2] Text features - tf-idf	78.5%	74.2%	77.4%	74.97%	76.5%	81.34%
	[T3] Text features - (tf-idf + GloVe Embeddings)	76.6%	71.8%	77.0%	73.5%	78.5%	85.2%
	[T2+I1] Text features - tf-idf + Image features - ResNet	75.9%	51.5%	78%	74.62%	<b>79%</b>	<b>84.5%</b>
	[T3+I1] Text features - (tf-idf + GloVe Embeddings) + Image features - ResNet	75%	70.5%	78%	74.6%	<b>79.6%</b>	<b>85.3%</b>
Hurricane Harvey	[T1] Text features - Bag of Words	85.3%	75.3%	85%	75.2%	82%	86%
	[T2] Text features - tf-idf	80.6%	60.6%	84.6%	74.7%	82.7%	85%
	[T3] Text features - (tf-idf + GloVe Embeddings)	77%	58.9%	83.7%	70%	83.3%	87.3%
	[T2+I1] Text features - tf-idf + Image features - ResNet	76.8%	58.15%	83%	70.73%	<b>83.1%</b>	<b>87.05%</b>
	[T3+I1] Text features - (tf-idf + GloVe Embeddings) + Image features - ResNet	83.8%	72%	86.8%	77%	<b>86.6%</b>	<b>90.7%</b>
Hurricane Irma	[T1] Text features - Bag of Words	82.8%	65.3%	64%	83.4%	82.2%	81.9%
	[T2] Text features - tf-idf	82.1%	59.1%	83.6%	67%	82.4%	80.4%
	[T3] Text features - (tf-idf + GloVe Embeddings)	80.9%	58%	85%	69%	84%	87%
	[T2+I1] Text features - tf-idf + Image features - ResNet	78.3%	50.12%	83.6%	65.8%	<b>83.4%</b>	<b>84.53%</b>
	[T3+I1] Text features - (tf-idf + GloVe Embeddings) + Image features - ResNet	83.4%	65.5%	84.2%	68%	<b>85.6%</b>	<b>88.12%</b>

Our proposed method performs better as it combines both the  $tf-idf$  features and word embedding features, which capture both the syntactic and semantic aspects from text. Adding image features improves the performance and helps the model to learn from multimodal features. Based on these results, we can conclude that the addition of image features (ResNet-50) to text features improves the performance in classifying *informative* tweets.

## V. CONCLUSIONS

This paper presented a novel approach of classifying relevant multimodal social media posts by systematically combining textual and image features, which are able to capture latent semantics of relevant information within the short-text messages. We have validated our approach with three crisis event datasets and achieved over 90% AUC score. The proposed method can help improve social media services at emergency response organizations. We plan to further evaluate our framework across different social media platforms for generalizability. Furthermore, there is a vast literature in the computer vision community on caption generation approaches for the textual description of images, and we plan to leverage these approaches to extract the textual information from the images for getting additional semantics.

## VI. ACKNOWLEDGEMENT

Authors thank US National Science Foundation (NSF) grant IIS-1657379 for partial support to this research.

## REFERENCES

- [1] M. Imran, C. Castillo, F. Diaz, and S. Vieweg, "Processing social media messages in mass emergency: A survey," *ACM Computing Surveys (CSUR)*, vol. 47, no. 4, p. 67, 2015.
- [2] H. Purohit, C. Castillo, F. Diaz, A. Sheth, and P. Meier, "Emergency-relief coordination on social media: Automatically matching resource requests and offers," *First Monday*, vol. 19, no. 1, 2013.
- [3] C. Reuter and T. Spielhofer, "Towards social resilience: A quantitative and qualitative survey on citizens' perception of social media in emergencies in europe," *Technological Forecasting and Social Change*, vol. 121, pp. 168–180, 2017.
- [4] A. L. Hughes and L. Palen, "The evolving role of the public information officer: An examination of social media in emergency management," *Journal of Homeland Security and Emergency Management*, vol. 9, no. 1, 2012.
- [5] U.S. Homeland Security, "Using social media for enhanced situational awareness and decision support," 2014. [Online]. Available: <https://www.dhs.gov/publication/using-social-media-enhanced-situational-awareness-decision-support>
- [6] C. Castillo, *Big Crisis Data: Social Media in Disasters and Time-Critical Situations*. Cambridge University Press, 2016.
- [7] F. Alam, F. Ofli, and M. Imran, "Crisismmd: Multimodal twitter datasets from natural disasters," in *Twelfth International AAAI Conference on Web and Social Media*, 2018.
- [8] H. Purohit, C. Castillo, M. Imran, and R. Pandev, "Social-eoc: Serviceability model to rank social media requests for emergency operation centers," in *2018 IEEE/ACM ASONAM*. IEEE, 2018, pp. 119–126.
- [9] K. Starbird, J. Maddock, M. Orand, P. Achterman, and R. M. Mason, "Rumors, false flags, and digital vigilantes: Misinformation on Twitter after the 2013 Boston marathon bombing," *Proc. iConference*, 2014.
- [10] D. T. Nguyen, F. Ofli, M. Imran, and P. Mitra, "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*. ACM, 2017, pp. 569–576.
- [11] L. Palen and K. M. Anderson, "Crisis informatics – new data for extraordinary times," *Science*, vol. 353, no. 6296, pp. 224–225, 2016.
- [12] S. R. Hiltz, J. A. Kushma, and L. Plotnick, "Use of social media by us public sector emergency managers: Barriers and wish lists," in *Proc. ISCRAM*, 2014, pp. 602–611.
- [13] F. Alam, F. Ofli, M. Imran, and M. Aupetit, "A twitter tale of three hurricanes: Harvey, irma, and maria," in *Proceedings of the 15th International Conference on Information Systems for Crisis Response and Management*. ISCRAM, 2018, pp. 553–572.
- [14] B. Basnyat, A. Anam, N. Singh, A. Gangopadhyay, and N. Roy, "Analyzing social media texts and images to assess the impact of flash floods in cities," in *2017 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 2017, pp. 1–6.
- [15] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *NIPS*, 2017, pp. 3146–3154.
- [16] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *EMNLP*, 2014, pp. 1532–1543.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [18] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 2016, pp. 785–794.