



Article

Social Media Analytics for Disaster Response: Classification and Geospatial Visualization Framework

Chao He and Da Hu *

Department of Civil and Environmental Engineering, Kennesaw State University, Marietta, GA 30060, USA; che4@students.kennesaw.edu

* Correspondence: dhu3@kennesaw.edu

Abstract: Social media has become an indispensable resource in disaster response, providing real-time crowdsourced data on public experiences, needs, and conditions during crises. This user-generated content enables government agencies and emergency responders to identify emerging threats, prioritize resource allocation, and optimize relief operations through data-driven insights. We present an AI-powered framework that combines natural language processing with geospatial visualization to analyze disaster-related social media content. Our solution features a text analysis model that achieved an 81.4% F1 score in classifying Twitter/X posts, integrated with an interactive web platform that maps emotional trends and crisis situations across geographic regions. The system's dynamic visualization capabilities allow authorities to monitor situational developments through an interactive map, facilitating targeted response coordination. The experimental results show the model's effectiveness in extracting actionable intelligence from Twitter/X posts during natural disasters.

Keywords: deep learning; disaster response; interactive map; social media; tweet



Academic Editor: Douglas O'Shaughnessy

Received: 11 March 2025

Revised: 29 March 2025

Accepted: 9 April 2025

Published: 14 April 2025

Citation: He, C.; Hu, D. Social Media Analytics for Disaster Response: Classification and Geospatial Visualization Framework. *Appl. Sci.* **2025**, *15*, 4330. <https://doi.org/10.3390/app15084330>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Social media platforms have emerged as transformative resource in modern disaster management, enabling government agencies and emergency responders to monitor public needs and sentiment in real time. The 2017 Hurricane Harvey response serves as a clear example of this paradigm shift, where geospatial analysis of social media content facilitated targeted relief operations by identifying critical demands such as emergency evacuations, medical aid requests, and reports of flood severity directly from victims' posts. This crisis highlighted how computational analysis of crowd-sourced data can bridge critical information gaps during large-scale disasters, transforming unstructured public communications into actionable intelligence for prioritizing life-saving interventions [1]. Traditional emergency communication systems frequently collapse under demand surges during disasters, creating critical response delays as overwhelmed call centers fail to process simultaneous assistance requests. This systemic vulnerability has elevated social media as a mission-critical triage platform, enabling affected populations to broadcast geolocated aid requests through mobile devices. Notably, many individuals trapped under debris have utilized mobile devices to request aid through platforms such as Twitter/X. Consequently, essential resources, such as food, medicine, supplies, and donations, were frequently mobilized within a few hours after the onset of disaster events [2].

During the 2023 Turkey–Syria earthquakes, 72% of validated rescue operations originated from crowdsourced intelligence, with entrapped victims using Twitter/X's SMS

fallback protocols to transmit precise GPS coordinates. Modern disaster frameworks now operate this digital lifeline through AI-powered triage systems that automatically classify urgent posts (like medical emergencies and supply needs). In past decades, researchers and companies increasingly utilized advanced machine learning techniques to derive meaningful insights from social media data [3]. These techniques enable real-time analysis of trends, sentiments, and social behaviors, providing crucial information for various applications, including disaster response, public health monitoring, and market analysis. By employing machine learning, deep learning, and transformer-based models, researchers can efficiently process vast amounts of social media content, identifying key patterns and informing decision-making processes across multiple domains.

For this research, we employed the dataset from HumAID [4] and evaluated multiple state-of-the-art models to identify the most effective approach for text classification. The best-performing model was selected for the classification of social media posts. Additionally, we developed a web-based platform featuring a dynamic interactive map that visualizes the locations of social media users who have enabled location-sharing on their mobile devices. The optimized text classification model was integrated into a web server, where it processes and stores classification results, geolocation data, and the original post content within map markers. The server continuously updates these markers, allowing decision-makers to access user-generated reports, review classified posts, and analyze real-time crisis information. This interactive visualization tool enhances situational awareness, enabling emergency response teams to track unfolding events, assess the severity of incidents based on social media data, and make rapid, informed decisions to improve disaster response efforts. Our key contributions are as follows:

- We developed and trained a high-performance deep learning model for the accurate classification of diverse disaster-related content from social media data.
- We designed and implemented an interactive web-based platform that integrates user geolocation with classified social media posts, supporting emergency responders in monitoring and assessing crisis situations more effectively.

The open-source implementation and video demonstration are available at <https://github.com/Saturn-Chao-He/Interactive-Map-for-Disaster-Response> (accessed on 10 March 2025).

The following content of this paper is as follows: Section 2 reviews related work, comparing our methodology with existing work. Section 3 describes our approach, covering the data preprocessing, model training, and development of web-based visualization with model deployment. Section 4 presents experimental results and post classification and performance evaluation. Section 5 discusses the study's limitations and outlines potential directions for future research. Finally, Section 6 summarizes the key findings and contributions of the work.

2. Related Works

Recent studies have increasingly explored the use of social media posts to improve the effectiveness of emergency response. By analyzing real-time posts, researchers and emergency management teams can improve situational awareness, optimize resource allocation, and facilitate faster decision-making during crises. Advanced machine learning and deep learning techniques have been employed to classify disaster-related information, detect urgent needs, and filter out misinformation, thereby improving the effectiveness of disaster response efforts. These studies highlight the challenges of using social media posts during a disaster, including difficulties processing large volumes of data in real time, the presence of misinformation and irrelevant content, and the challenges in collecting comprehensive data. The vast and unstructured nature of social media posts necessitates efficient computational techniques to promptly extract meaningful insights. Addition-

ally, the spread of false or misleading information can hinder effective decision-making, requiring robust filtering mechanisms. Moreover, social media coverage of disasters is often inconsistent, leading to gaps in situational awareness that can impact emergency response efforts. Overcoming these challenges is crucial for improving the reliability and effectiveness of social media-based disaster response systems.

2.1. Existing Dataset

Numerous datasets have been developed to support sentiment analysis and crisis-related classification tasks during natural disasters. These resources have significantly contributed to the NLP community, facilitating advancements in research and development within the crisis informatics domain. By providing labeled datasets, benchmark models, and evaluation metrics, they have enabled researchers to improve disaster-related text classification, sentiment analysis, and real-time information retrieval, ultimately enhancing emergency response and decision-making systems.

1. SWDM2013 [5] includes two distinct collections: the Joplin set, which contains 4400 labeled tweets from the May 2011 tornado that struck Joplin, Missouri, and the Sandy set, featuring 2000 tweets from Hurricane Sandy in October 2012. These datasets provide valuable insights into the social media discourse during two significant disaster events and serve as foundational data for early research in disaster-related sentiment analysis and classification.
2. Disaster Response Data resource [4] contains 30,000 tweets categorized into 36 different classes, collected from various disasters, including the 2010 earthquakes in Haiti and Chile, the 2010 floods in Pakistan, and the 2012 occurrence of Hurricane Sandy in the United States. This dataset is particularly valuable for training models across diverse disaster contexts, enabling the creation of more generalized crisis classification systems.
3. CrisisNLP [6] is another comprehensive dataset featuring 50,000 human-annotated tweets collected from 19 disaster events between 2013 and 2015. These tweets are categorized using multiple labeling schemes, with classes designed for humanitarian disaster response and health-related emergencies. The diversity of annotations allows researchers to explore various aspects of disaster response, including needs identification, resource allocation, and medical emergencies.
4. CrisisMMD [7] adopts a multimodal and multitask approach, consisting of 18,000 labeled tweets accompanied by associated images. This dataset supports both text- and image-based classification, enabling researchers to develop models that analyze not only textual information but also visual content. This capability is particularly useful for tasks like assessing infrastructure damage, visualizing rescue operations, and understanding the broader impact of disasters.
5. TREC Incident Streams [8], developed for the TREC-IS 2018 evaluation challenge, consists of 19,784 tweets labeled for identifying actionable information and assessing the criticality of information. This dataset is specifically designed for enhancing decision-making during disasters by helping emergency responders distinguish between high-priority and low-priority information.
6. Eyewitness Messages [9] focuses on analyzing eyewitness accounts on Twitter. It classifies 14,000 tweets into four distinct categories across events such as floods, earthquakes, fires, and hurricanes. This dataset is particularly valuable for understanding firsthand experiences during disasters, which can provide crucial, real-time information for emergency responders.
7. Disaster Tweet Corpus 2020 [10] consolidates several existing datasets focused on classifying various disaster event types. This resource offers a unified platform for researchers to train models capable of recognizing different types of disaster-

related tweets. Additionally, Alam et al. [11] integrate eight data sources, yielding a comprehensive collection of 166,100 tweets for informativeness classification and 141,500 tweets for humanitarian classification tasks. These consolidated resources enhance the ability to train robust models that generalize effectively across multiple disaster scenarios.

Collectively, these datasets play a vital role in advancing research within the crisis informatics domain. They enable the development of more sophisticated classification models, improve sentiment analysis, and support the creation of automated systems capable of providing timely and accurate insights during emergencies. This, in turn, contributes to more effective disaster management strategies, better resource allocation, and enhanced response coordination in crisis situations.

2.2. Text Classification Models

Various research approaches have leveraged both traditional machine learning and transformer-based techniques to classify different types of social media text streams, particularly in disaster response and crisis informatics. Early approaches to this problem predominantly relied on supervised machine learning methods, where incoming social media texts were classified into two or more predefined categories. These models were trained on labeled datasets to recognize patterns in disaster-related content, enabling them to categorize new, unseen posts based on their textual and contextual features.

In an early study, Dong et al. [12] introduced a dataset specifically designed for sentiment analysis and evaluated the performance of eight machine learning models. These traditional models demonstrated varying levels of effectiveness, setting a foundation for future research in disaster-related sentiment classification. As the field evolved, researchers began incorporating deep learning methods to improve text classification performance, particularly for handling complex patterns in large-scale social media datasets. Burel et al. [13] reported that both machine learning and deep learning models can produce highly competitive outcomes under specific conditions, highlighting the importance of selecting the appropriate method based on dataset characteristics and application requirements. Further advancements were made by Nguyen et al. [14], who proposed using Convolutional Neural Networks (CNNs) for identifying valuable information from tweets during crisis situations. CNNs, known for their success in image processing tasks, were adapted for text analyses by leveraging their ability to detect local patterns and relationships within the data. In addition, Alam et al. [15] established a benchmark for classifying disaster-related social media images using deep learning models, demonstrating the potential of multimodal approaches that combine both textual and visual data for more comprehensive disaster response systems.

In recent years, Transformer-based models have emerged as the dominant architecture in the field of natural language processing, offering significant improvements over previous methods. Transformers, particularly those based on BERT and its variants, have shown remarkable success in understanding contextual relationships in text. For instance, Alam et al. [4] constructed a dataset specifically for disaster-related text classification and conducted comparative analyses between traditional machine learning algorithms and Transformer-based methods, including BERT and its derivatives. Cantini et al. [3] introduced an innovative methodology that harnesses the capabilities of prompt-based Large Language Models (LLMs) to enhance disaster response efforts. The authors found that GPT-3.5 Turbo consistently outperformed alternative models in terms of informativeness, clarity, and attribution. Gemini and Command also demonstrated a strong performance, exhibiting a comparable overall report quality and coherence. In contrast, Llama showed relatively weaker results, particularly regarding clarity and detail. The results highlighted

the superior performance of Transformer models, particularly in capturing complex linguistic patterns and handling the imbalanced class distributions commonly found in disaster-related datasets. However, deep learning models often demonstrate suboptimal performance in the domain of crisis informatics, and existing models face several limitations that hinder the development of more advanced and effective deep learning solutions. Key challenges include data scarcity, which restricts the availability of large, high-quality datasets necessary for training robust models. Additionally, class imbalance remains a significant issue, as certain disaster-related categories, such as urgent rescue needs or medical emergencies, are often underrepresented in the data. This imbalance can lead to biased models that fail to recognize rare but critical situations.

Noisy annotations also pose a challenge, as inconsistencies and errors in the labeling process can degrade model accuracy and limit the reliability of predictions. Moreover, the lack of domain-specific pretraining further impairs the generalization capabilities of deep learning models. Pretrained models often lack exposure to disaster-related terminology and context, reducing their effectiveness when applied to real-world crisis scenarios. Addressing these limitations is essential for enhancing the performance of models in disaster response applications. Improving dataset quality, developing more balanced and representative data collections, and incorporating domain-specific pretraining can significantly strengthen model generalization and reliability. By overcoming these constraints, it becomes possible to develop more sophisticated, accurate, and responsive systems that can provide actionable insights during emergencies, ultimately improving the efficiency and effectiveness of disaster response efforts.

2.3. Disaster Management

Disaster management is a systematic approach that involves planning, mitigation, rescue, response, and recovery, requiring collaboration among federal, state, local, and private sector entities. This coordinated effort is essential for minimizing the impact of disasters and ensuring an efficient and effective response to emergencies. Li et al. [16] applied the Standard Deviation Ellipse (SDE) method to examine the spatial patterns of urban floods and determine Areas of Interest (AOI) using social media data from Chengdu, China. Social media data served as the response variable, while ten urban flood-influencing factors were chosen as independent variables to evaluate their impact on flood occurrences. This analysis highlighted how urban floods can result in severe property damage, disrupt transportation networks, interrupt daily activities, and pose significant threats to public safety. Expanding the scope of social media data analysis, Otal et al. [17] introduced a novel approach for identifying and classifying emergency situations using a large language model. The study explored two primary applications: enhancing 911 dispatch operations by supporting telecommunicators in processing emergency calls and providing personalized protective action guidance to the public based on real-time social media data. This approach demonstrates the potential of advanced language models in improving emergency communication systems and supporting timely interventions during crises.

Integrating diverse data sources, including social media streams and IoT-generated data, with advanced big data analytics tools such as Apache Spark and Apache Kafka can greatly enhance the efficiency and effectiveness of emergency management processes [18]. These technologies enable real-time collection, processing, and analysis of large-scale data streams, allowing for faster detection of crisis events and more coordinated emergency responses. Additionally, Bukar et al. [19] utilized a visualization of similarities to analyze and interpret crisis datasets, supporting more effective crisis management and decision-making. By applying advanced data visualization techniques, the study enhanced the

understanding of patterns, relationships, and trends within disaster-related data. This improved visualization enables emergency response teams to allocate resources more effectively, identify critical areas in need of intervention, and develop more efficient response strategies based on the evolving nature of disaster events.

2.4. Contributions of This Work

While reviewing the existing literature, we identified a lack of systems that implement a fully comprehensive response mechanism for disaster management and response. To address this gap, our primary objective is to develop a website that can automatically collect streaming social media posts, apply a pretrained model for accurate inference, and visualize users' locations on a dynamic map. This integrated system is designed to provide emergency response teams with critical, up-to-date information, allowing them to assess situations more effectively and respond in a timely manner. By automating data collection, analysis, and visualization, the platform enhances situational awareness and supports faster decision-making during disasters, ultimately improving resource allocation and response coordination.

3. Methods

The workflow of our system, illustrated in Figure 1, consists of two primary pipelines: model training and model deployment. The training pipeline includes several key stages: data collection, data preprocessing, model training, model evaluation, hyperparameter tuning, and saving the best model.

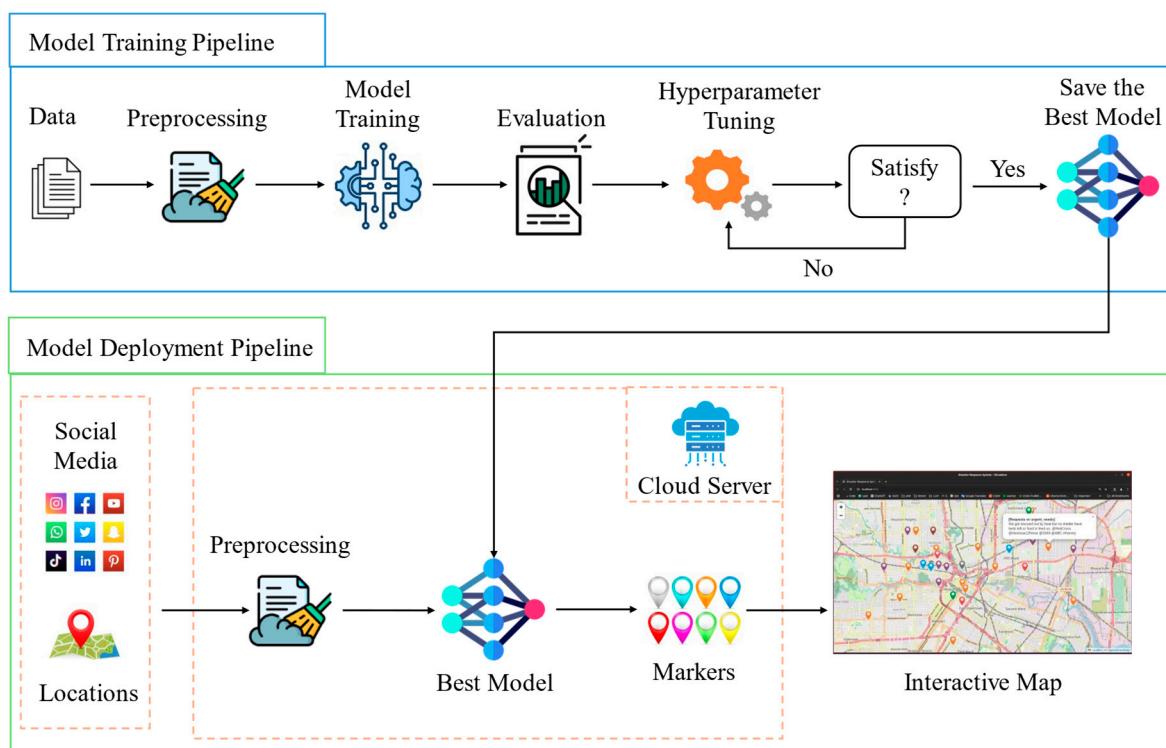


Figure 1. Model training and deployment pipelines.

The methodological decisions made during data preprocessing, model training, evaluation, and hyperparameter tuning were guided by best practices established in previous studies and proven to improve model performance in text classification tasks. Data preprocessing techniques, such as tokenization, the removal of stop words, punctuation, and URLs, and lemmatization were applied based on their demonstrated effectiveness in improving text classification accuracy and reducing noise, as validated by Imran et al. [6] and

Alam et al. [4]. The choice of transformer-based models (e.g., BERT, RoBERTa, Modern-BERT) for training was motivated by their state-of-the-art performance in various NLP benchmarks, particularly due to their ability to capture contextual semantics efficiently, as shown by Devlin et al. [20], Liu et al. [21], and Clark et al. [22]. For hyperparameter tuning, the selection of parameters such as learning rate, batch size, and maximum sequence length followed the established optimization protocols recommended by Lan et al. [23] and Sanh et al. [24], and standard practices detailed in the related literature [4,20–22], ensuring stable convergence, minimizing overfitting, and achieving optimal generalization performance. This process ensured that model was effectively trained on disaster-related social media data and optimized for accurate classification.

The deployment pipeline emphasizes efficient processing and visualization for timely decision support. It begins by streaming social media posts, followed by data preprocessing to clean and prepare the incoming information. The system then generates predictions using the trained model and creates corresponding map markers based on the classification results. Finally, these markers are visualized on a dynamic map, providing an interactive and timely overview of disaster-related information. This visualization allows response teams to monitor ongoing events and make informed decisions based on geographically distributed social media data.

3.1. Dataset and Data Cleaning

3.1.1. Dataset

The dataset HumAID [4] used in this study consists of over 77,000 labeled tweets, sampled from a larger collection of 24 million tweets collected during 19 major real-world disasters that occurred between 2016 and 2019. These disasters include hurricanes, earthquakes, wildfires, and floods, providing a diverse and comprehensive set of disaster-related social media data. The dataset is structured into 11 hierarchically organized classes, capturing various humanitarian and informational aspects of disaster-related content. These categories encompass a wide range of emergency-related themes, such as requests for urgent assistance, reports of infrastructure damage, expressions of sympathy and support, and updates on rescue or relief efforts. This hierarchical structure allows for a more nuanced classification, supporting the development of models capable of effectively distinguishing between different types of disaster-related information for improved crisis response and management. These classes include the following:

- Caution and advice: Notifications of issued or lifted warnings, along with guidance and safety tips related to the disaster.
- Sympathy and support: Posts expressing prayers, thoughts, and emotional support.
- Requests or urgent needs: Reports of urgent needs or requests for supplies, including food, water, clothing, money, medical supplies, or blood.
- Displaced people and evacuations: Reports of individuals who have relocated due to the crisis, including temporary evacuations.
- Injured or dead people: Reports of individuals who have been injured or killed because of the disaster.
- Missing or found people: Reports of individuals who are missing or have been found following the disaster event.
- Infrastructure and utility damage: Reports of damage to infrastructure, including buildings, houses, roads, bridges, power lines, communication poles or vehicles.
- Rescue, volunteering, or donation efforts: Reports of rescue, volunteering, or donation efforts, including transportation to safe locations, evacuations, the distribution of medical aid or food, shelter provisions, and donations of money, goods, or services.

- Other relevant information: Tweets that do not fit into any of the above categories but still contain important information relevant to humanitarian aid.
 - Not humanitarian: Tweets that do not contain information related to humanitarian aid.
 - Don't know or can't judge: Tweets that are irrelevant or cannot be assessed, including those written in non-English languages.

Unlike previous datasets, the HumAID dataset offers a more balanced distribution of disaster types and ensures greater consistency in labeling. This balance enhances the reliability of classification tasks by reducing the impact of class imbalances, which often hinder model performance in disaster-related datasets. Furthermore, HumAID is enriched with tweets that are more likely to originate from areas directly affected by disasters, increasing the presence of firsthand accounts from eyewitnesses and individuals directly impacted by these events. This focus on authentic, location-relevant data makes HumAID particularly valuable for crisis informatics research, as it allows for a more accurate understanding of public sentiment, urgent needs, and situational updates during disaster scenarios.

Table 1 presents the distribution of annotations across different disaster events and class labels in HumAID dataset [4], providing an overview of how various categories are represented within the dataset. This distribution is critical for ensuring that the models trained on these data can learn effectively from diverse disaster scenarios and produce accurate classifications across a range of emergency contexts. The dataset applied a filtering process to remove low-prevalence classes, defined as categories containing fewer than 15 tweets. For instance, in the case of the 2016 Ecuador Earthquake, the class “Displaced people and evacuations” included only three tweets, rendering it insufficient for meaningful classification and model training. Removing these underrepresented categories aimed to enhance the model’s ability to learn effectively by minimizing noise and reducing class imbalance. As a result of this refinement, the dataset size was reduced from 77,196 to 76,484 tweets. This adjustment was intended to improve the overall performance of the classification model by focusing on categories with sufficient representation, allowing for more balanced and accurate learning across different disaster-related classes [4].

Figure 2 illustrates the word frequency of all disasters using a word cloud, highlighting the most frequently occurring terms in the dataset. This visualization offers insights into the dominant themes and concerns expressed by users during the disaster, providing valuable context for understanding the dataset's content and the relevance of specific terms to disaster response efforts.

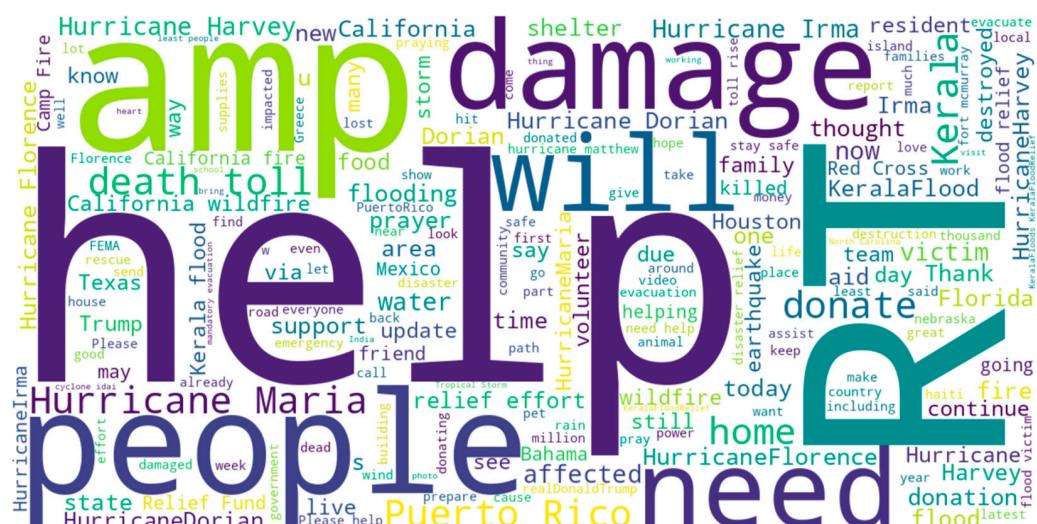


Figure 2. World cloud of HumAID dataset.

Table 1. Distribution of annotations across events and class labels [4].

Event Name	Caution and Advice	Displaced People and Evacuations	Infrastructure and Utility Damage	Injured or Dead People	Missing or Found People	Not Humanitarian	Don't Know or Can't Judge	Other Relevant Information	Requests or Urgent Needs	Rescue Volunteering or Donation Effort	Sympathy and Support	Total
Ecuador Earthquake	30	3	70	555	10	23	18	81	91	394	319	1594
Canada Wildfires	106	380	251	4	-	79	13	311	20	934	161	2259
Italy Earthquake	10	3	54	174	7	9	10	52	30	312	579	1240
Kaikoura Earthquake	493	87	312	105	3	224	19	311	24	207	432	2217
Hurricane Matthew	36	38	178	224	-	76	5	328	53	326	395	1659
Sri Lanka Floods	28	9	17	46	4	20	2	56	34	319	40	575
Hurricane Harvey	541	688	1271	698	10	410	42	1767	333	2823	635	9164
Hurricane Irma	613	755	1881	894	8	615	60	2358	126	1590	567	9467
Hurricane Maria	220	131	1427	302	11	270	39	1568	711	1977	672	7328
Mexico Earthquake	35	4	167	254	14	38	3	109	61	984	367	2036
Maryland Floods	70	3	79	56	140	77	1	137	1	73	110	747
Greece Wildfires	26	7	38	495	20	74	4	159	25	356	322	1526
Kerala Floods	139	56	296	363	7	456	65	955	590	4294	835	8056
Hurricane Florence	1310	637	320	297	-	1060	95	636	54	1478	472	6359
California Wildfires	139	368	422	1946	179	1318	68	1038	79	1415	472	7444
Cyclone Idai	89	57	354	433	19	80	11	407	143	1869	482	3944
Midwestern U.S. Floods	79	8	140	14	1	389	27	273	46	788	165	1930
Hurricane Dorian	1369	802	815	60	1	874	46	1444	179	987	1083	7660
Pakistan Earthquake	71	-	125	401	1	213	32	154	19	152	823	1991
Total	5404	4036	8163	7321	435	6305	560	12,144	2619	21,278	8931	77,196

For the experiments, the dataset was split into three subsets: 70% for training, 10% for validation, and 20% for testing. This distribution ensures there is sufficient data for each stage of model development. Training ensured the model learned from a substantial portion of the data, validation was used for hyperparameter tuning and monitoring overfitting, and testing provided an unbiased evaluation of the model's final performance. This division facilitates an assessment of the model's effectiveness and generalization capabilities.

3.1.2. Data Cleaning

In our study, data cleaning was an essential preprocessing step, conducted using Python 3.8's regular expressions and the Natural Language Toolkit (NLTK) library 3.8.1. Raw social media data often contain various irrelevant elements, such as stop words, short tokens, URLs, excessive whitespace, HTML markup, punctuation, and other non-alphanumeric characters, which can hinder the model's ability to learn meaningful patterns. To address this, the text was tokenized into individual words and subjected to several preprocessing steps: converting all text to lowercase, lemmatizing words to their base forms, and removing all irrelevant elements. This process ensures that the model focuses on the most informative components of the text.

The overall quality of the dataset plays a crucial role in determining the performance of any downstream model. Poor-quality data, such as mislabeled posts, irrelevant content, or noise, can negatively impact the model's learning process, leading to reduced accuracy and reliability in real-world applications. Therefore, thorough data cleaning and preprocessing are fundamental for enhancing model robustness and ensuring more accurate classification outcomes in disaster response scenarios.

3.2. Models

This study concentrates on multiclass classification tasks using transformer-based models to enhance text classification for disaster-related social media posts. Specifically, we employed BERT [20], RoBERTa [21], DistilBERT [24], BigBird [25], ALBERT [23], ELECTRA [22], and ModernBERT [26] to train and evaluate the effectiveness of our classification system.

- **BERT:** BERT is a transformer-based model designed to understand the context of words by simultaneously analyzing both their left and right surroundings. It is pre-trained on large corpora using two objectives: masked language modeling and next sentence prediction. This dual-training strategy allows BERT to capture complex relationships between words, making it highly effective for a wide range of natural language processing tasks, including sentiment analysis, question-answering, and text classification.
- **RoBERTa:** RoBERTa builds upon BERT's foundation by refining the pretraining process. It eliminates the next sentence prediction task and instead focuses on training with larger batches and more data for longer durations. These adjustments enhance the model's ability to capture subtle linguistic patterns, leading to an improved performance across various benchmarks and making it more robust for complex text classification tasks.
- **DistilBERT:** DistilBERT is a condensed version of BERT that prioritizes efficiency while retaining much of the original model's performance. It uses a technique known as knowledge distillation, where a smaller model learns from a larger, pre-trained model. This results in a faster, lighter model that is particularly suitable for applications with limited computational resources, such as mobile devices or real-time processing systems.

- BigBird: BigBird extends the capabilities of traditional transformer models by introducing a sparse attention mechanism, allowing it to process much longer sequences without dramatically increasing computational costs. This innovation makes BigBird especially suitable for tasks that involve lengthy documents or sequences, such as document summarization, genome analysis, or processing large-scale social media streams during disasters.
- ALBERT: ALBERT enhances BERT's efficiency by reducing the number of parameters through cross-layer parameter-sharing and factorized embeddings. Despite being a smaller model, ALBERT achieves competitive results by emphasizing sentence-order prediction during pretraining. Its memory-efficient design makes it well-suited for environments where computational resources are constrained while still delivering strong performance on various NLP tasks.
- ELECTRA: ELECTRA introduces a novel pretraining method that diverges from the traditional masked language modeling used in BERT. Instead, it uses a generator-discriminator framework, where a discriminator learns to differentiate between real tokens and those generated by a smaller model. This approach leads to more sample-efficient training, allowing ELECTRA to achieve high performance with reduced computational costs, making it an effective solution for large-scale text classification tasks.
- ModernBERT: ModernBERT represents the latest evolution in BERT-based architecture, incorporating advanced optimization strategies and training techniques. It leverages innovations such as dynamic masking and improved pretraining objectives to enhance both efficiency and accuracy. ModernBERT is designed to achieve superior results on contemporary NLP benchmarks, making it highly effective for complex text classification tasks in dynamic environments such as disaster response systems.

By evaluating and comparing these advanced transformer-based models, this study seeks to determine the most effective method for the classification of social media posts. This allows emergency response teams to rapidly analyze incoming data and respond appropriately to disaster situations, improving the efficiency and effectiveness of crisis management operations.

3.3. Visualization

3.3.1. Web Server and User Interface (UI)

We deployed our testing website on a cloud server rented from VULTR, configured identically to our local workstation used for model training. This uniform configuration ensures seamless integration between development and deployment environments, thereby optimizing performance and maintaining consistency across platforms. The website is hosted using Nginx, an open-source web server that acts as a reverse proxy and load-balancer. Nginx is widely recognized for its efficiency, scalability, and ability to manage a high volume of concurrent connections, making it particularly suitable for disaster response applications. Its lightweight, event-driven architecture facilitates rapid data-processing and response times, ensuring reliability and speed in cloud-based environments where system stability is critical.

Although Nginx served as our primary web server due to its robust performance and scalability, alternative solutions exist. Apache, another widely adopted web server, offers extensive module support and advanced features for handling dynamic content and complex configurations. Additionally, the Python testing server provides a lightweight, development-oriented alternative that is commonly utilized for testing and debugging in the early stages of application development. The chosen server architecture effectively manages incoming data streams, supports visualization, and maintains high reliability

under peak loads, key attributes for disaster response systems that depend on timely and accurate information dissemination.

To enable the interactive visualization of categorized data, we developed the index.html page of our website using Leaflet, an open-source JavaScript library designed for web-based mapping applications. Leaflet facilitates the dynamic presentation of disaster-related data through an intuitive and interactive mapping interface. The system employs a set of 10 distinct, color-coded markers to represent different event categories, prioritizing incidents based on urgency. Markers in red, orange, and pink are designated for the most critical situations, particularly those posing immediate threats to human life. This visual prioritization mechanism enhances situational awareness by enabling emergency response teams to quickly identify and respond to the most urgent cases.

The interactive map is designed to update continuously, providing a comprehensive overview of disaster-related reports classified by severity and type. Users can engage with the map by clicking on individual markers, which triggers a pop-up window displaying detailed information about specific incidents. This functionality supports rapid situational assessments and facilitates informed decision-making among emergency response personnel. To maintain a clear and interpretable interface, the HTML map dynamically cycles through 10 newly reported incidents every 20 s, preventing marker overcrowding while ensuring timely updates. This dynamic refresh mechanism allows response teams to focus on the most recent and relevant disaster events. The interactive capabilities of the system, such as clickable markers with pop-up information, enable efficient assessment of key details, including post content, geographic location, and event classification. Figure 3 illustrates the color-coded markers and their corresponding categorizations.



Figure 3. Markers with corresponding classifications.

3.3.2. Model Deployment and Visualization

The website server integrates the most effective model for the classification of disaster-related social media posts. The mapping interface is dynamically updated, periodically refreshing to display markers stored in JSON format. A backend Python script emulates a streaming environment by retrieving 10 new social media data points every 20 s. This update cycle is optimized for system performance, with each batch of 10 posts requiring approximately 14 s for model reasoning, followed by an additional 2 s for data loading, preprocessing, and output storage. Once processed, the classification results for each post are stored in a JSON marker, which includes relevant details such as text, geographic location, and classification labels. These markers are then visualized on the interactive map, allowing users to monitor the spatial distribution of disaster-related information.

Figure 4 illustrates the detailed architecture and workflow of our proposed disaster response visualization system. Initially, social media data are continuously streamed into the server, simulating incoming information from platforms such as Twitter/X. Each incoming social media post undergoes preprocessing, where irrelevant textual elements, such as URLs, special characters, and stop words, are removed to ensure clean and standardized input data for the classification model. The cleaned data is then processed by our optimized text classification model, which categorizes each post according to predefined disaster-related classifications. Following classification, geographic information associated with the post is combined with the classification results and original textual content into structured JSON markers. These JSON markers are stored temporarily on the server, ready for visualization. The dynamic mapping interface, built with Leaflet and JavaScript, periodically retrieves these markers, updating the interactive map every 20 s to display the latest disaster-related events. Each category of disaster event is visually differentiated using distinct color-coded markers, enabling emergency responders to quickly interpret severity and prioritize actions. Users can interact with these map markers to view detailed information about each classified social media post, enhancing situational awareness and facilitating informed, timely decision-making during crisis management scenarios. By integrating automated classification with geospatial visualization, the system provides a practical and data-driven solution for improving decision-making in disaster management scenarios.

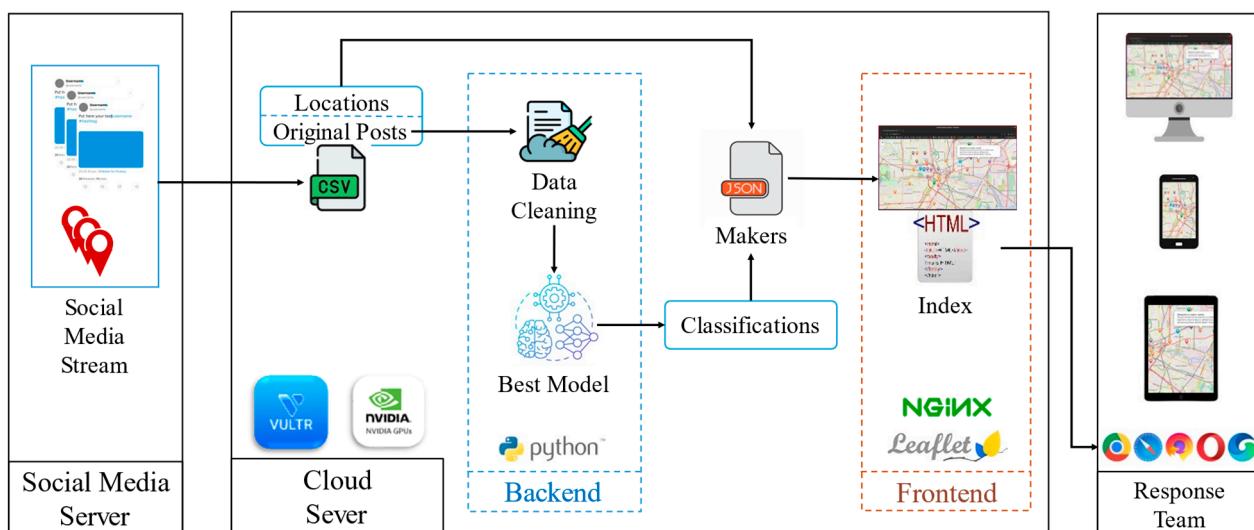


Figure 4. The architecture and implementation of the proposed website server.

4. Results

In this section, we present the experimental evaluation of our system, specifically focusing on the performance of various transformer-based text classification models using the HumAID dataset. The subsequent subsections detail the experimental setup, including computational resources, hyperparameters, and evaluation metrics. Additionally, we analyze the performance results across multiple disaster scenarios, highlighting the strengths and limitations of the models tested. Finally, a detailed case study on Hurricane Harvey demonstrates the practical application of our approach in disaster response visualization.

4.1. Social Media Text Classification

4.1.1. Experiment Details

The experiments were conducted on a Dell workstation equipped with a 4 GHz, 16-core × 32-thread AMD Ryzen Threadripper Pro 5995WX processor, 80 GB of RAM, and two NVIDIA RTX A5000 GPUs. This computing setup provided the necessary com-

putational resources to handle large-scale training and reasoning tasks, particularly for processing extensive disaster-related social media datasets and fine-tuning transformer-based models. The substantial memory capacity and parallel processing capabilities of the GPUs enabled efficient model training, reducing computation time while ensuring optimal performance. To optimize model performance, cross-entropy loss was employed as the objective function for hyperparameter tuning. This loss function effectively measured the discrepancy between predicted and actual labels, guiding the model toward improved classification accuracy.

The training process was configured with carefully selected hyperparameters to balance computational efficiency and model accuracy. A batch size of 32 was used to ensure stable gradient updates without excessive memory consumption. The learning rate was set to 2×10^{-5} to facilitate gradual convergence while preventing divergence during optimization. The maximum sequence length was set to 150, allowing the model to process sufficiently long text inputs while maintaining efficiency. The model underwent 20 fine-tuning epochs to find meaningful patterns in the data without overfitting.

4.1.2. Metrics

Model performance was evaluated using the weighted average F1 score (F1), a metric specifically selected to address issues related to class imbalance within the dataset. Unlike standard accuracy measures, which can be misleading in imbalanced datasets, the weighted F1 score accounts for the proportional representation of each class, ensuring that underrepresented categories still contribute meaningfully to the overall evaluation. By utilizing the weighted F1 score, the evaluation process became more nuanced, preventing the model from being overly biased toward majority classes. This ensured that the model could effectively identify critical instances that are highly relevant for disaster response efforts. The model that consistently demonstrated the highest weighted F1 score across these runs was selected as the final model for deployment, ensuring reliability in real-world disaster response applications.

The F1 score and weighted F1 score provide a balanced evaluation by considering both precision and recall. The F1 score is the harmonic mean of precision and recall, defined as follows:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (1)$$

Precision (P) is the proportion of correctly classified positive instances among all instances predicted as positive:

$$P = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad (2)$$

Recall (R) is the proportion of correctly classified positive instances among all actual positive instances:

$$R = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad (3)$$

The weighted F1 score accounts for class imbalances by calculating the F1 score for each class and averaging them based on the number of samples in each class:

$$\text{Weighted F1} = \sum_{i=1}^n w_i \times F1_i \quad (4)$$

- $F1_i$ is the F1 score for class i .
- w_i is the weight for class i , defined as the proportion of samples belonging to that class.

- n is the total number of classes.

During model training, the evaluation step monitored the F1 score at each epoch and compared it against the highest previously recorded score. Whenever an epoch achieved a higher F1 score, the corresponding model was automatically saved as the best-performing model. This best model was subsequently selected for deployment in our visualization system.

4.1.3. Experimental Results

Table 2 shows the results for the weighted F1 score of text classification. Among all the evaluated models, ModernBERT achieved the highest weighted F1 score, surpassing the performance of other transformer-based architectures. This result underscores the advancements brought by ModernBERT's architectural innovations, which contribute to its superior ability to capture complex patterns within disaster-related social media data. However, it is worth noting that, for certain disaster events, switching between different models did not yield significant performance improvements. This observation suggests that the limitations in classification accuracy may be influenced by the quality and granularity of the dataset's annotations, rather than being influenced solely by model architecture.

Table 2. Weighted F1 score of text classification (best results are bolded).

Disaster	Class	BERT	RoBERTa	DistilBERT	BigBird	ALBERT	ELECTRA	ModernBERT
2016 Ecuador Earthquake	8	0.823	0.868	0.853	0.859	0.832	0.870	0.900
2016 Canada Wildfires	8	0.754	0.786	0.780	0.751	0.759	0.780	0.842
2016 Italy Earthquake	6	0.823	0.872	0.851	0.843	0.818	0.867	0.900
2016 Kaikoura Earthquake	9	0.731	0.741	0.732	0.736	0.730	0.752	0.746
2016 Hurricane Matthew	9	0.751	0.775	0.768	0.759	0.750	0.767	0.796
2017 Sri Lanka Floods	8	0.686	0.729	0.718	0.708	0.690	0.721	0.845
2017 Hurricane Harvey	9	0.730	0.759	0.747	0.739	0.739	0.748	0.791
2017 Hurricane Irma	9	0.705	0.717	0.719	0.704	0.712	0.719	0.733
2017 Hurricane Maria	9	0.706	0.725	0.722	0.710	0.714	0.720	0.767
2017 Mexico Earthquake	8	0.834	0.852	0.847	0.842	0.836	0.837	0.915
2018 Maryland Floods	8	0.684	0.764	0.708	0.690	0.699	0.756	0.751
2018 Greece Wildfires	9	0.702	0.742	0.735	0.711	0.715	0.732	0.836
2018 Kerala Floods	9	0.723	0.741	0.728	0.736	0.733	0.739	0.821
2018 Hurricane Florence	9	0.755	0.776	0.763	0.760	0.759	0.770	0.810
2018 California Wildfires	10	0.741	0.768	0.757	0.758	0.749	0.756	0.804
2019 Cyclone Idai	10	0.733	0.789	0.759	0.749	0.740	0.791	0.844

Table 2. Cont.

Disaster	Class	BERT	RoBERTa	DistilBERT	BigBird	ALBERT	ELECTRA	ModernBERT
2019 Midwestern U.S. Floods	7	0.719	0.766	0.750	0.744	0.720	0.755	0.802
2019 Hurricane Dorian	9	0.674	0.689	0.680	0.678	0.674	0.677	0.698
2019 Pakistan Earthquake	8	0.797	0.823	0.833	0.799	0.798	0.811	0.873
Average		0.740	0.772	0.760	0.751	0.745	0.766	0.814

Despite these constraints, ModernBERT, as the most recent evolution of BERT-based models, demonstrated consistently strong performance across the majority of disaster-related datasets. Its advanced training techniques and architectural refinements allowed it to outperform earlier models, making it particularly effective in processing social media data for disaster response. These results showcase the potential of ModernBERT for enhancing situational awareness and facilitating timely decision-making in disaster management systems.

4.2. Case Study

Hurricane Harvey was selected as a demonstration case due to the extensive availability of data and its historical significance as one of the costliest and most devastating hurricanes in U.S. history. Making landfall in August 2017, Harvey caused catastrophic flooding, destruction, and widespread displacement, severely impacting communities across southern Texas and parts of Louisiana. During the crisis, social media platforms played an important role in facilitating rescue coordination and enabling crisis communication. As traditional communication infrastructures were overwhelmed or rendered inoperable due to extensive flooding, many stranded individuals turned to social media as their primary means of seeking help. Affected individuals used these platforms to share real-time updates about their locations, conditions, and urgent needs, often providing critical information that traditional emergency channels could not capture promptly. Figure 5 illustrates the word cloud for Hurricane Harvey, highlighting the most frequently occurring terms in social media posts related to the disaster. Key words such as “shelter”, “damage”, “victim”, “water”, “donation”, “help”, and “relief” emerge as dominant themes, reflecting critical aspects of disaster response and recovery efforts. Identifying these high-frequency terms is essential for training an accurate text-classification model, as it helps the model recognize and categorize disaster-related information effectively.

The response efforts during Hurricane Harvey were notably collaborative, involving a wide range of actors, including federal, state, and local agencies, as well as numerous volunteer organizations. Among the most prominent volunteer groups was the Cajun Navy, a grassroots organization that mobilized quickly to carry out water rescues using privately owned boats [27]. These volunteers played a significant role in reaching areas that were inaccessible to official rescue teams, demonstrating the vital importance of community-led response efforts during large-scale disasters. Hurricane Harvey underscored the necessity of robust disaster preparedness measures, including the implementation of effective early warning systems and reliable crisis communication tools. The disaster highlighted the shortcomings of conventional emergency management infrastructure, particularly in terms of scalability and responsiveness during high-impact events. More importantly, it revealed the potential of social media-based response systems as a complementary tool in disaster management. By leveraging real-time data from social media, emergency response teams were able to enhance situational awareness, allowing for quicker assessments of affected areas and the more accurate identification of individuals in need of urgent assistance.

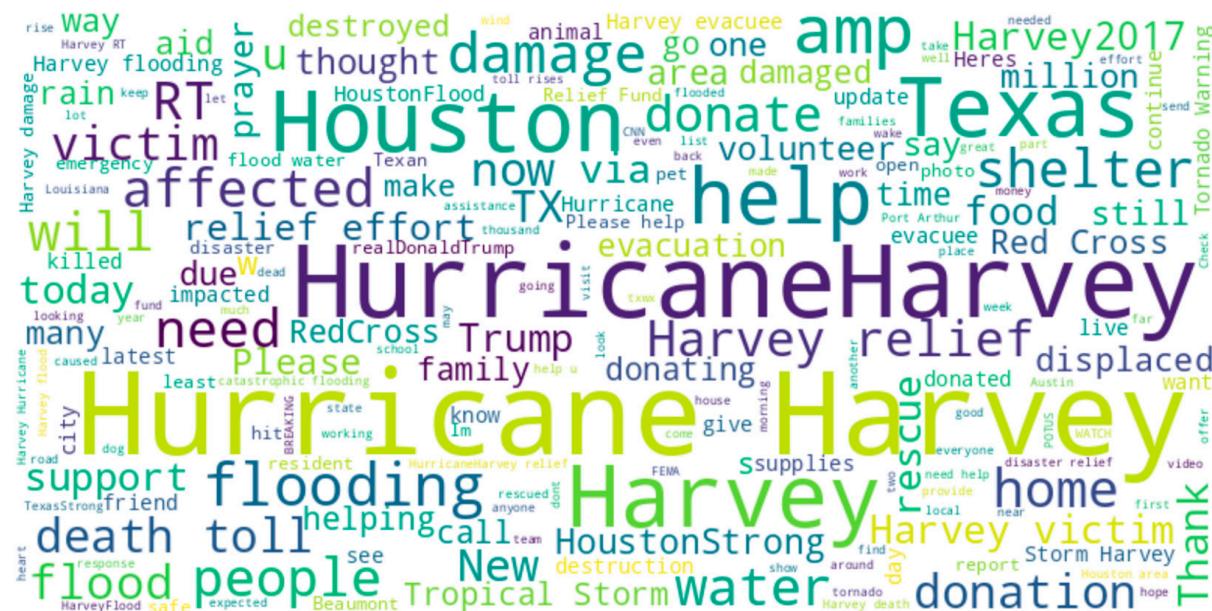


Figure 5. Words cloud for Hurricane Harvey.

Hurricane Harvey also brought attention to the need to integrate Artificial Intelligence (AI) and machine learning technologies into disaster management systems. AI-driven analyses of social media data can enable the faster identification of distress signals, categorize needs more effectively, and support decision-making in real time. This technological integration has the potential to significantly improve both the speed and effectiveness of response operations during future disasters. Most major social media platforms provide APIs that enable researchers to access data streams or datasets for various analytical purposes [2]. When users post messages and activate location-sharing features on their mobile devices, the associated social media data can include geographic coordinates. However, a study shows that only about 3.1% of users choose to share their location when posting content [28].

In this study, the original dataset [4], designed specifically for classification tasks, excluded all location-related information by applying a geographic filter, which resulted in the removal of precise latitude and longitude coordinates. To address this limitation and demonstrate the effectiveness of spatial visualization, we selected a subset of the Hurricane Harvey test data from the original dataset [4] and assigned randomly generated geographic coordinates to each social media post. For model deployment and simulation, our system retrieved 10 social media data instances from the simulated stream at a time, effectively mimicking a continuous data feed for reasoning on the server. Each batch was processed in real time, allowing us to demonstrate the system's ability to handle ongoing data streams, perform text classification, and update the visual map dynamically. Figure 6 shows the social media information visualization of Hurricane Harvey on interactive map.

This methodology highlights the system's data processing abilities, including rapid data ingestion, the accurate classification of social media posts, and the visualization of geolocated information on a dynamic map interface. The assigned random locations allow for a practical demonstration of how geographic data can enhance situational awareness and support decision-making in a real-world disaster response scenario. This visualization tool enables emergency responders to monitor trends geographically, identify critical areas needing immediate attention, and coordinate resource allocation more effectively.

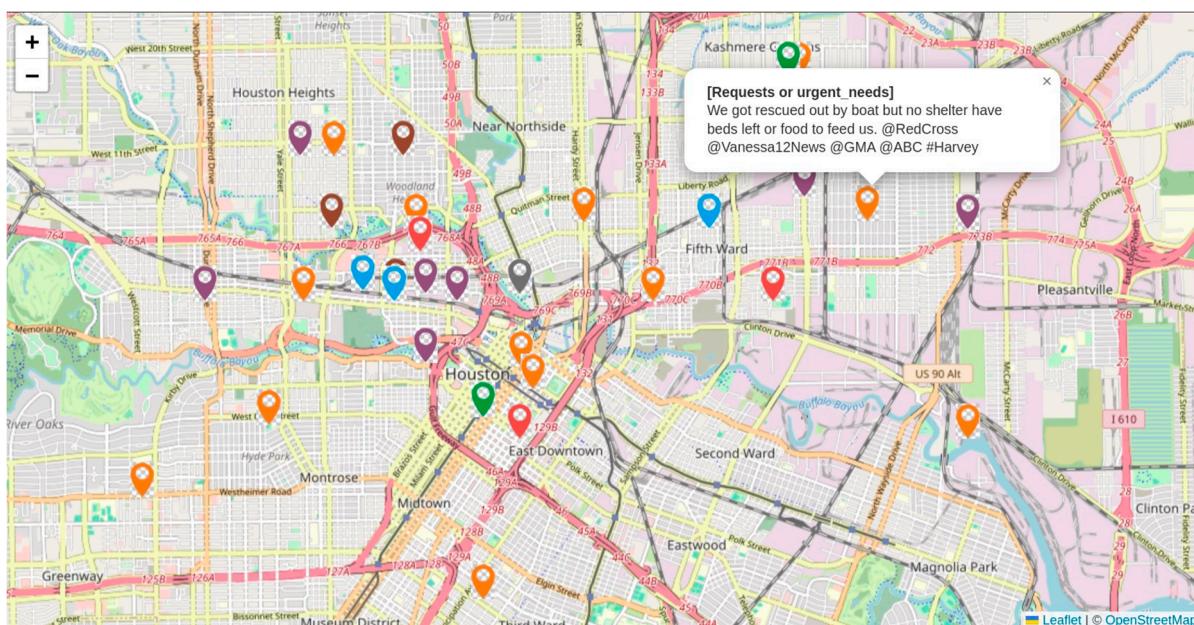


Figure 6. Social media information visualization of Hurricane Harvey on map (usernames are excluded from the pop-up windows for privacy considerations).

5. Discussion

5.1. Classification and Data Source

The classifications of social media posts in existing disaster response datasets often fall short of meeting the nuanced requirements of real-time emergency scenarios. Many existing categories are too broad or generalized, failing to capture the specific situational details necessary for facilitating effective emergency interventions. A more focused classification system, tailored to the real-world needs of disaster response, is critical for enhancing situational awareness and enabling rapid, targeted actions. In a real-time disaster response system, it is essential to prioritize categories that provide actionable intelligence for response teams, such as identifying individuals in urgent need of rescue, those requiring medical assistance, or areas affected by infrastructure damage. In the current study, we rely exclusively on a publicly available dataset composed of tweets, which inherently limits our model to the linguistic patterns and content styles typical of Twitter/X. Acknowledging this limitation, future research will focus on expanding our datasets by collecting and integrating data from diverse social media platforms, such as Facebook, Instagram, and other sources. This broader approach will enable our model to generalize better across varying linguistic contexts, content formats, and communication styles, thereby enhancing its robustness and effectiveness in real-world disaster response scenarios. In future work, we will implement advanced techniques, including similarity detection algorithms, credibility assessment tools, and machine learning-driven misinformation detection models, to proactively identify, manage, and mitigate the impact of fake or duplicate social media posts. These enhancements will significantly improve the reliability and accuracy of our disaster response system in real-world operational scenarios.

5.2. System, Model and Response Improvement

While this study establishes a foundational approach for integrating social media analytics into disaster response systems, there are significant opportunities for enhancement and expansion. One potential advancement lies in transforming the system into a fully autonomous platform capable of making real-time decisions and initiating response actions without human intervention. This would involve the system automatically coordinating

with governmental and non-governmental agencies based on the classification outcomes of social media data. For instance, the platform could automatically notify rescue teams about individuals trapped in affected areas or direct drones to deliver essential supplies, such as food, water, or medicine, to isolated regions. Such a system could significantly accelerate response times and improve the overall effectiveness of disaster management operations. However, implementing such an automated system also faces critical challenges, particularly regarding its accuracy and reliability. In future research, we plan to perform comprehensive performance evaluations, including stress-testing to analyze system behavior under increased data loads, the implementation of efficient buffering strategies, and optimization of our web server architecture to ensure robust scalability. Additionally, we will conduct thorough assessments of privacy risks, strictly adhere to regulatory requirements, and implement robust mechanisms designed to safeguard user privacy and ensure our system's compliance with all relevant legal and ethical standards.

Currently, the system's best-performing model achieves a weighted F1 score of 81.4%, indicating a persistent risk of misclassification. In a disaster response context, such errors can have severe consequences, including the misallocation of critical resources, delayed assistance, or even failure to address life-threatening situations. Therefore, improving the precision, robustness, and reliability of the model remains a central focus for future research. In future research, we intend to conduct a comprehensive analysis of class-specific metrics and systematically investigate instances of misclassification. This detailed evaluation will enable us to identify specific weaknesses and implement targeted improvements, thereby enhancing the overall accuracy, robustness, and reliability of our model for critical disaster response applications. Additionally, we plan to broaden our comparative analyses by incorporating advanced architecture such as Deepseek and GPT-4, alongside comprehensive benchmarking, to systematically evaluate their performance. We will also investigate the benefits of domain-specific pretraining, aiming to further improve the models' effectiveness in crisis informatics applications. Additionally, we plan to develop and enhance specialized datasets tailored explicitly to disaster response scenarios, which should increase the accuracy, reliability, and practical relevance of our system in real-world emergency management contexts. Another key priority for future research is the development of multilingual classification capabilities using advanced multilingual architectures. Our objective is to seamlessly integrate multilingual support into our disaster response system, improving its inclusivity and effectiveness across diverse linguistic environments. By extending its multilingual functionality, we aim to significantly enhance the system's global applicability and reliability, empowering it to effectively assist emergency response teams in international disaster contexts.

Beyond technical enhancements, the system's future development could also focus on improving collaboration between different emergency response agencies. A more interconnected response platform would facilitate seamless communication between federal, state, local, and non-governmental organizations, fostering coordinated efforts during large-scale disasters. By integrating real-time data from multiple stakeholders, the system could enable more effective resource allocation and reduce response times during critical moments. Moving forward, we plan to actively collaborate with local disaster response agencies and operational centers to leverage their expertise, understand their specific needs, and incorporate their valuable insights. Through close engagement with these stakeholders, we will refine our system to effectively address real-world operational challenges, significantly enhancing its practicality, usability, and preparedness for deployment in genuine disaster response scenarios.

6. Conclusions

This study introduced a transformer-based text classification model that achieved an 81.4% weighted F1 score for analyzing social media content, complemented by an interactive visualization system tailored to disaster response applications. By seamlessly integrating data ingestion, accurate text classification, and geographic visualization, our platform enhances situational awareness and supports timely and effective emergency coordination. This foundational implementation demonstrates the practical utility of leveraging social media data in disaster management, bridging the gap between abundant crowdsourced information and operational crisis response needs. Future research can pursue several important directions to further enhance the system's capabilities and integrate diverse data sources such as satellite imagery and Internet of Things (IoT) sensors to provide more comprehensive situational awareness. Expanding the system's multilingual support will ensure inclusivity and effectiveness across global disaster scenarios. Additionally, real-world pilot testing and collaboration with emergency response agencies will validate the system's practical utility and inform iterative improvements. Efforts will also focus on refining model accuracy, particularly for underrepresented disaster categories, and enhancing model transparency to foster trust and accountability in automated crisis response decision-making.

Author Contributions: Conceptualization, D.H.; methodology, C.H.; software, C.H.; validation, C.H. and D.H.; formal analysis, D.H.; investigation, D.H.; resources, C.H.; data curation, C.H.; writing—original draft preparation, C.H.; writing—review and editing, D.H.; visualization, C.H.; supervision, D.H.; project administration, D.H.; funding acquisition, D.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the US National Science Foundation (NSF) via Grant 2346936.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available at https://crisisnlp.qcri.org/humaid_dataset (accessed on 20 March 2025).

Acknowledgments: The authors gratefully acknowledge the support from NSF and Kennesaw State University. Any opinions, findings, recommendations, and conclusions in this paper are those of the authors and do not necessarily reflect the views of NSF and Kennesaw State University.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Li, J.; Stephens, K.K.; Zhu, Y.; Murthy, D. Using Social Media to Call for Help in Hurricane Harvey: Bonding Emotion, Culture, and Community Relationships. *Int. J. Disaster Risk Reduct.* **2019**, *38*, 101212. [[CrossRef](#)]
- Toraman, C.; Kucukkaya, I.E.; Ozcelik, O.; Sahin, U. Tweets Under the Rubble: Detection of Messages Calling for Help in Earthquake Disaster. *arXiv* **2023**, arXiv:2302.13403.
- Cantini, R.; Cosentino, C.; Marozzo, F.; Talia, D.; Trunfio, P. Harnessing Prompt-Based Large Language Models for Disaster Monitoring and Automated Reporting from Social Media Feedback. *Online Soc. Netw. Media* **2025**, *45*, 100295. [[CrossRef](#)]
- Alam, F.; Qazi, U.; Imran, M. HumAID: Human-Annotated Disaster Incidents Data from Twitter with Deep Learning Benchmarks. *ICWSM* **2021**, *15*, 933–942. [[CrossRef](#)]
- Imran, M.; Elbassuoni, S.; Castillo, C.; Diaz, F.; Meier, P. Practical Extraction of Disaster-Relevant Information from Social Media. In Proceedings of the 22nd International Conference on World Wide Web, New York, NY, USA, 13 May 2013.
- Imran, M.; Mitra, P.; Castillo, C. 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'16), Portorož, Slovenia, 1 May 2016.
- Alam, F.; Ofli, F.; Imran, M. CrisisMMD: Multimodal Twitter Datasets from Natural Disasters. *ICWSM* **2018**, *12*, 100198. [[CrossRef](#)]

8. McCREADIE, R.; BUNTAIN, C.; SOBOROFF, I. TREC Incident Streams: Finding Actionable Information on Social Media. In Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management, Valencia, Spain, 19 May 2019.
9. ZAHRA, K.; IMRAN, M.; OSTERMANN, F.O. Automatic Identification of Eyewitness Messages on Twitter during Disasters. *Inf. Process. Manag.* **2020**, *57*, 102107. [[CrossRef](#)]
10. WIEGMANN, M.; KERSTEN, J.; KLAN, F. Analysis of Detection Models for Disaster-Related Tweets. In Proceedings of the 17th Annual International Conference on Information Systems for Crisis Response and Management, Blacksburg, VA, USA, 19 May 2020.
11. ALAM, F.; SAJJAD, H.; IMRAN, M.; OFLI, F. CrisisBench: Benchmarking Crisis-Related Social Media Datasets for Humanitarian Information Processing. *Proc. Int. AAAI Conf. Web Soc. Media* **2021**, *15*, 923–932. [[CrossRef](#)]
12. DONG, Z.S.; MENG, L.; CHRISTENSON, L.; FULTON, L. Social Media Information Sharing for Natural Disaster Response. *Nat. Hazards* **2021**, *107*, 2077–2104. [[CrossRef](#)]
13. BUREL, G.; ALANI, H. Crisis Event Extraction Service (CREES)—Automatic Detection and Classification of Crisis-Related Content on Social Media. In Proceedings of the 15th International Conference on Information Systems for Crisis Response and Management, Rochester, NY, USA, 18 May 2018.
14. NGUYEN, D.; MANNAI, K.A.A.; JOTY, S.; SAJJAD, H.; IMRAN, M.; MITRA, P. Robust Classification of Crisis-Related Data on Social Networks Using Convolutional Neural Networks. *Proc. Int. AAAI Conf. Web Soc. Media* **2017**, *11*, 632–635. [[CrossRef](#)]
15. ALAM, F.; OFLI, F.; IMRAN, M.; ALAM, T.; QAZI, U. Deep Learning Benchmarks and Datasets for Social Media Image Classification for Disaster Response. In Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), The Hague, The Netherlands, 7–10 December 2020; pp. 151–158.
16. LI, Y.; OSEI, F.B.; HU, T.; STEIN, A. Urban Flood Susceptibility Mapping Based on Social Media Data in Chengdu City, China. *Sustain. Cities Soc.* **2023**, *88*, 104307. [[CrossRef](#)]
17. OTAL, H.T.; STERN, E.; CANBAZ, M.A. LLM-Assisted Crisis Management: Building Advanced LLM Platforms for Effective Emergency Response and Public Collaboration. In Proceedings of the 2024 IEEE Conference on Artificial Intelligence (CAI), Singapore, 11–14 June 2024; pp. 851–859.
18. SHAH, S.A.; SEKER, D.Z.; HAMEED, S.; DRAHEIM, D. The Rising Role of Big Data Analytics and IoT in Disaster Management: Recent Advances, Taxonomy and Prospects. *IEEE Access* **2019**, *7*, 54595–54614. [[CrossRef](#)]
19. BUKAR, U.A.; SAYEED, M.S.; AMODU, O.A.; RAZAK, S.F.A.; YOGARAYAN, S.; OTHMAN, M. Leveraging VOSviewer Approach for Mapping, Visualisation, and Interpretation of Crisis Data for Disaster Management and Decision-Making. *Int. J. Inf. Manag. Data Insights* **2025**, *5*, 100314. [[CrossRef](#)]
20. DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019.
21. LIU, Y.; OTT, M.; GOYAL, N. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv* **2019**, arXiv:1907.11692.
22. CLARK, K.; LUONG, M.-T.; LE, Q.V.; MANNING, C.D. ELECTRA: Pre-Training Text Encoders as Discriminators Rather Than Generators. *arXiv* **2020**, arXiv:2003.10555.
23. LAN, Z.; CHEN, M.; GOODMAN, S. ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations. *arXiv* **2020**, arXiv:1909.11942.
24. SANH, V.; DEBUT, L.; CHAUMOND, J.; WOLF, T. DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter. *arXiv* **2020**, arXiv:1910.01108.
25. ZAHEER, M.; GURUGANESH, G.; DUBEY, K.A. Big Bird: Transformers for Longer Sequences. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 17283–17297.
26. WARNER, B.; CHAFFIN, A.; CLAVIÉ, B. Smarter, Better, Faster, Longer: A Modern Bidirectional Encoder for Fast, Memory Efficient, and Long Context Finetuning and Inference. *arXiv* **2024**, arXiv:2412.13663.
27. CAJUN NAVY. Available online: https://en.wikipedia.org/wiki/Cajun_Navy (accessed on 8 March 2025).
28. SLOAN, L.; MORGAN, J. Who Tweets with Their Location? Understanding the Relationship between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter. *PLoS ONE* **2015**, *10*, e0142209. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.