

Leveraging multimodal deep learning for natural disaster event classification and its damage severity analysis through social media posts

Nivedita Kasturi^{1,2}, Shashikumar Guruputra Totad², Goldina Ghosh³

¹Department of Computer Science and Engineering, PES University, Bangalore, India

²School of Computer Science and Engineering, KLE University, Hubli, India

³Department of Computer Application and Science, Institute of Engineering and Management Kolkata, Kolkata, India

Article Info

Article history:

Received Oct 17, 2023

Revised Mar 26, 2024

Accepted Apr 18, 2024

Keywords:

Damage severity identification

Disaster event classification

Hybrid deep learning

Multi-modality analytics

Social media posts

ABSTRACT

Accurate and timely information is critical to effectively coordinate disaster response. Due to the diversity and complexity of data sources, it is difficult for traditional methods to classify disaster events and assess damage severity. Previous studies have mainly focused on specific tasks, such as information collection or humanitarian assistance, but have not adequately addressed the assessment of disaster loss severity. This paper proposes a hybrid learning model to improve disaster event classification and damage severity identification. The model combines image and text data, using ResNet50 to extract features from images, and using long short-term memory with an attention mechanism to learn sequences from text. This combination allows for a more contextual and informative representation of the input data. The experimental results shows that the proposed multimodal approach achieves significantly better results in disaster event classification attaining an accuracy of 90.31%, compared to existing methods. Furthermore, the model demonstrates promising capabilities in assessing damage severity, offering significant improvements for disaster management and response preparation where accuracy and dependability are crucial.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Nivedita Kasturi

Department of Computer Science and Engineering, PES University

Bangalore, India

Email: niveditak@pes.edu

1. INTRODUCTION

The impacts of climate change and other environmental changes are increasing, and natural disasters are increasing in frequency and severity, which pose a major threat to the environment and human life [1], [2]. In order to control damage and facilitate the timely recovery process for an affected individual, disaster preparedness and response strategies is required to be more improved and enhanced. In today's digitally connected world, social networking platforms like Twitter, Facebook, WhatsApp, and Instagram have proven to be one of the valuable sources for communication, connection, information dissemination [3]. During a natural crisis or in emergencies, social networking platforms helps to quickly disseminate information, provide timely updates to facilitate situational awareness, and contributes in opportunities to search for support or assistance to affected populations [4]. In addition, social media platforms also provide raw information that can be very helpful for emergency responders as the people on the scene are the first to share what is going on, including text updates, photos, and even videos. This unfiltered information can help disaster responders understand how bad things are and get aid to areas that need it most [5], [6]. However, there are inherent

challenges in using social media effectively to respond to natural calamities. On the other hand, it is also very difficult to keep track of everything that is posted, especially during the disaster or natural calamities. Finding the most important information, such as the severity of a disaster and the likelihood of imminent danger, in social media posts is often a challenging task. In this regard, effective strategy towards developing better response mechanism basically depends on getting the right information and in timely manner [7]. The conventional methods for response mechanism often involves people manually analyze, sort, and annotate millions of information, which is quite time-consuming, costly, and often subjected to human-error. Hence, the conventional approaches are not suitable for situations where disaster events occur so rapidly and uncertainly [8].

Therefore, development of an automated approach is required that can precisely classify the natural disaster from the social media posts during emergencies and or events of natural calamities [9]. The recent advancements in artificial intelligence and emergence of deep learning approaches shows a promising scope to automate feature engineering tasks and insights from multi-variate and complex data like social media posts [10]. However, developing such automated scheme is not an easy task especially when dealing with both textual and image data in social media posts. This requires careful analysis of data, adoption of preprocessing techniques and selection of suitable deep learning models to extract meaningful patterns and perform classification of natural calamities. In the existing literature, many research works have been presented, where few research, works are subjected to usage of only social media images and few are focused on textual information for building disaster classification models [11], [12]. There are fewer works that consider both text and image features for natural calamity classification and disaster severity analysis because of the inherent complexity of these types of data.

Social media data in the form of textual communication often presents a complex data for computational analytics tasks because users often use informal ways of writing posts that include slang, emoticons, short sentences, and with various language [13]. Image data, on the other hand, provides a richer context, but disaster images are complex due to their multifaceted nature and require careful interpretation and understanding. For example, during an earthquake event, a single image can depict damaged buildings, injured people, debris, and various other types of damage [14]. Therefore, processing images to establish an effective disaster response system proves to be extremely challenging due to the wide variation in image quality, multifaceted nature of disaster scenarios, and diversity of content. In the current context, deploying learning models capable of handling the dynamic nature of disaster situations remains an ongoing challenge. Disaster response mechanisms operate under time constraints. It is crucial to ensure their accuracy and high responsiveness in real-time situations, because the accuracy and timeliness of information provided by such systems will seriously affect disaster response accuracy and timeliness. Distinguish between life-saving measures and potential losses. In the recent literature, many works in a similar direction have been proposed by different researchers, most of which perform a single task in the context of disaster response [15]. The work of Anbarasan *et al.* [16] introduced a flood prediction system based on internet of things and convolutional neural network (CNN) to detect and predict flood events. This work proposes a rule generation method based on preprocessed material that is used as input to a convolutional deep neural network (CDNN) classifier to classify the occurrence and non-occurrence of flooding. Zhang *et al.* [17] addressed the challenge of accurately predicting landslide displacements in response to rainfall and reservoir water level changes in China. The authors proposed a dynamic prediction model based on the gated recurrent unit (GRU) learning algorithm, following cumulative displacement decomposition and trend prediction.

Ge *et al.* [18] utilizes graph knowledge concept that considers remote sensing, and geographic data, along with integration of domain knowledge to build a spatio-temporal framework for disaster prediction. Several researchers use machine learning and natural language processing (NLP) to analyze disaster sentiment and damage via social media analytics [19], [20]. Zou *et al.* [21] leverages the potential of combining image and text information by employing a deep learning approach for visual feature extraction and utilizing the FastText framework for textual feature extraction. A data fusion model is developed to combine these features for classifying relevant disaster images. Asif *et al.* [22] introduced a disaster taxonomy of emergency response systems based on deep learning classification models such as VGG-16 and YOLO. The method used decision tables followed by data analytics operations to map image outputs to the disaster-related information to determine suitable emergency responses. Zhang *et al.* [23] proposed a multimodal classification system using latent dirichlet allocation (LDA) for clustering and a bidirectional encoder representation with a pre-trained CNN model to analyze text and image data to determine thematic information during a disaster. Hence, it can be analyzed that there are currently many studies on natural disaster classification, but most of them focus on text or image data. None of these studies dealt with determining the severity of disaster losses, which is also important for effective disaster response systems. According to literature analysis, predicting disasters and assessing their severity through social media posts is a complex and multifaceted task that requires complex data preprocessing and feature engineering, and faces the challenge of large and imbalanced data sets. The identified research gaps are discussed below:

- Existing disaster detection methods do not effectively extract the important features from different data types (such as text and images). This is because feature engineering in multimodal disaster detection requires specific domain knowledge from a data analysis perspective, which is not well reflected in current research.
- Previous approaches to disaster analysis have focused on specific tasks, such as providing information or humanitarian assistance. This leaves a gap in analyzing the severity of damage caused by disasters.
- Deploying multiple models specific to different material types and tasks can be challenging as it involves model selection, optimization, and interpretation. However, the existing literature does not adequately address this challenge.
- Designing and developing highly integrated models is computationally challenging task, requires unique optimization techniques. Limited attention has been paid to managing model complexity, addressing overfitting issues, and optimizing performance.
- The literature lacks novelty and does not explore approaches to cross-modal learning, where models can learn from textual and image material simultaneously, enabling them to understand the relationships and dependencies between different modalities.

Therefore, this paper suggests a highly synchronized and integrated computational framework based on NLP mechanism and hybrid deep learning approach. The prime aim of the proposed system is to offer a reliable automated system using social media posts that can help emergency responders make better decisions about providing relief to affected communities. This proposed work is a novel multi-modal multi-task hybrid learning method that automatically combines textual and visual posts in social networking to classify disaster events and automatically assess their damage severity. In the hybrid learning architecture, long short-term memory (LSTM) is used to capture the long-term dependence of the text content of the post and extract key information related to disaster events. At the same time, it adopts the ResNet architecture as an advanced feature learner for CNN models to process visual content. Since the dataset used to train the proposed hybrid learning model is a complex multi-modal task and is associated with the class imbalance problem. Therefore, the proposed learning model incorporates an attention mechanism to make the system pay more attention to the features of a few categories. The schematic architecture of the proposed system is shown in Figure 1.

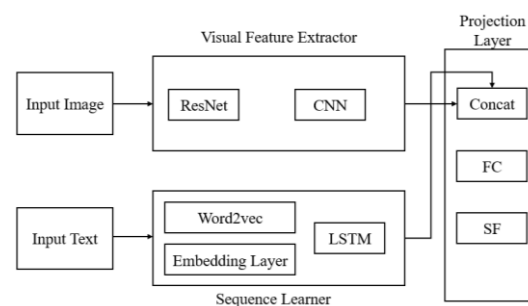


Figure 1. High level architecture of the proposed system

As it can be seen from Figure 1, the proposed system has four main layers: input layer, visual layer, sequence learner of text features, and projection layer. The input layer is responsible for receiving initial text and visual information. The visual layer uses CNN to perform transfer learning according to the pre-trained deep learning model. The study employed ResNet, a pre-trained CNN architecture that serves as a high-level feature extractor for image classification tasks. The sequence learning module of this system uses LSTM and attention mechanism to process text content, while the projection layer is responsible for feature concatenation and classification of disaster events, and uses the fully connected (FC) layer after SoftMax (SF) for damage severity analysis activation function. The system aims to improve the accuracy and comprehensiveness of disaster event classification and damage severity analysis by integrating information from multiple data modes. This novel approach based on dual data patterns and optimized deep learning models can better help disaster responders, researchers, and policymakers develop more effective and targeted disaster response strategies for emergencies, such as reports, requests for assistance, and updates on changing disaster conditions from less relevant or non-actionable content. By automating this process, the research aims to speed up the delivery of critical information to emergency responders, relief organizations and affected communities, thereby improving overall disaster response efficiency.

2. METHOD

This section details the proposed system and details the implementation procedures adopted for the proposed disaster response system. First, the data set used is briefly described, followed by data analysis and metadata construction. The subsequent sections then describe the computational operations adopted in preparing dataset and feature engineering to enhance the predictive capability and feature generalization ability of the proposed hybrid learning model. Furthermore, an implementation strategy is outlined for the feature extraction and its concatenation in the projection layer of hybrid deep learning model.

2.1. Dataset

The dataset adopted in this research work is crisis vision benchmark (CVB) [24] introduced for building a disaster response and preparedness system. This dataset was developed by researchers from Google AI, Stanford University, and the University of Washington. It consists of variety of disaster events like floods, wildfires, earthquakes, hurricanes, and many more in the form of both images and texts from a variety of sources such as social media, satellite imagery and drones. The distribution of data sample in CVB dataset is shown in Figure 2 for a variety of disaster response and preparedness tasks as follows:

- Disaster detection task for identifying the presence of a disaster in an image.
- Damage assessment for analyzing the severity of damage caused by a disaster events.
- Object detection task focuses on specific objects identification in disaster images, such as people, buildings, and vehicles.
- Scene classification task is subjected to predicting or classifying the disaster events or natural calamities, such as floods, wildfires, and earthquakes.

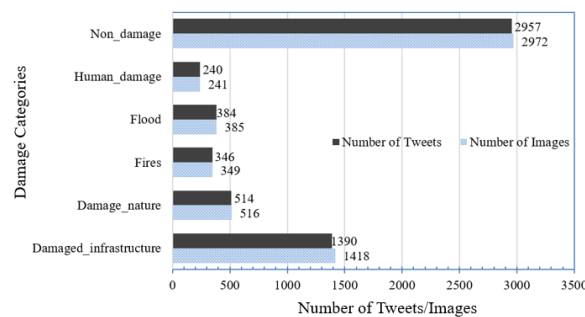


Figure 2. The distribution of data samples in the dataset

As shown in Figure 2, the disaster category damaged_infrastructure has the largest number of samples, followed by damaged_natural, fire, and flood. The human_damage category has the fewest samples, and the non_damage category has the most samples, resulting in a serious data imbalance. Uneven distribution of data, especially large differences between non-damage categories and other disaster categories, may bias deep learning models. Traditional models may be biased toward more common categories, which may lead to misclassification of important disasters. These misclassifications can have serious consequences in disaster management situations, where timely and accurate information is crucial.

2.2. Data preparation

The data set adopted presents some structural and organizational challenges that make it difficult to directly use it for disaster event classification and damage severity analysis. These issues include inconsistencies in data formatting. The arrangement and structurization of data are more task-centric, which subjects it to inherent complexities that need to be addressed for effective utilization in training the model.

2.2.1. Dataset structure

The dataset used for disaster event classification and damage severity analysis is organized in a manner that reflects the complexity of the task. It combines visual data with textual information to enhance the understanding of each disaster scenario. This kind of data structurization will ensure that the model can learn effectively from the diverse types of data it contains. The structure of dataset is summarized as follows:

- The primary data unit is an image representing each of the six different disaster classes. These images are organized within distinct sub-directories corresponding to each class.
- Within each subdirectory, text data is provided for most images, i.e., tweets posted on the specific disaster event or category to which the image belongs.

- A separate folder describes information on damage severity analysis. This folder is organized systematically into training, testing, and validation subsets.

2.2.2. Issues with the dataset

The dataset presents several challenges which required to be addressed to ensure effective training of model and accurate predictive outcomes. These issues are mostly related to the data's imbalance and arrangement, which can have a big effect on how well machine learning algorithms work. Therefore, it very important to mitigate these issues in order to create a robust and trustworthy predictive model for disaster classification and damage severity analysis. The key issues associated with dataset is highlighted as follows:

- A key issue is the inherent imbalance of the dataset in terms of the number of images and associated text annotations or tweets.
- Data on damage severity is pre-partitioned into training, testing and validation subsets, adding additional complexity.

2.2.3. Data preparation approach

The study adopts an effective methodology to address issues with the dataset. A a specific algorithm is designed to harmonize and integrate data from text and image datasets. The following are the main highlights of the algorithm for data preparation:

- Since damage severity data is mostly in text format, the image data needs to be prepared according to the textual dataset. This operation is performed by reorganizing the image data based on its identity and paths, ensuring consistent alignment with the corresponding textual data.
- The next operation is carried out towards creating a metadata in .CSV file format. This data then serves as a central repository of information, including image paths, related tweets, disaster occurrence and severity.
- Additionally, tweets used as text annotations are subsequently added to the metadata. Tweets are matched to images based on the image-id column, so the visual content of each image is always paired with corresponding textual material.
- Text data from tweets are preprocessed to clean and simplify, removing irrelevant information to improve the effectiveness of subsequent analysis.

The successful execution of the proposed data preparation process returns a structured data frame, as highlighted in Figure 3. The data frame includes the path of each image, its associated tweets, associated disaster events, and severity. After preprocessing, a new column for cleaned text data is added, which is crucial for extracting textual features and conducting predictive analytics. This refined data frame is then utilized as the final training dataset for the study.

| | image_id | image_path | tweets | disaster_event | severity_level |
|------|---|---|---|------------------------|----------------|
| 0 | floodwater_2017-06-24_03-18-10.jpg | multimodel/images/train/flood/floodwater_2017-... | Tropical Storm Cindy flood waters submerge a c... | flood | mild |
| 1 | wreckedcar_2017-04-05_20-09-10.jpg | multimodel/images/train/damaged_infrastructure... | #camaro #musclecar #Calif #police #wrecked #cr... | damaged_infrastructure | severe |
| 2 | hurricanesandy_2017-10-29_23-23-41.jpg | multimodel/images/train/flood/hurricanesandy_2... | ... #5yearsater #hurricanesandy | flood | little_or_none |
| 3 | destroyedbuilding_2017-05-18_09-11-14.jpg | multimodel/images/train/damaged_infrastructure... | I was so sad when saw such places in Lefkoşa... | damaged_infrastructure | little_or_none |
| 4 | isiscrimes_2015-08-14_23-22-03.jpg | multimodel/images/train/human_damage/isiscrime... | Syrians are bleeding.....#syrians #syria #s... | human_damage | mild |
| ... | ... | ... | ... | ... | ... |
| 4263 | ad_2017-11-25_10-16-26.jpg | multimodel/images/train/non_damage/ad_2017-11-... | #AD\n\nAzmo Power Generators\n\nFrom 43kva ,60kv... | non_damage | little_or_none |
| 4264 | ad_2017-11-25_09-15-21.jpg | multimodel/images/train/non_damage/ad_2017-11-... | Hello 🇵🇱\n#poland #polishboy #polishgirl #inst... | non_damage | little_or_none |
| 4265 | ad_2017-11-25_03-26-45.jpg | multimodel/images/train/non_damage/ad_2017-11-... | Ustedes saben que yo llevé el cabello corto de... | non_damage | little_or_none |
| 4266 | nature_2017-10-30_17-47-58.jpg | multimodel/images/train/non_damage/nature_2017... | Railroads, a camera and friends make a great d... | non_damage | little_or_none |

Figure 3. A sample visualization of the metadata after the data preparation operation

Table 1 shows the distribution statistics for the disaster event classification task obtained after the data preparation task, highlighting significant differences between categories. Non-damage categories dominate the dataset, while events such as fire and human damage are underrepresented. Table 2 focuses on the disaster severity identification task, where the number of little or none severity levels is much greater than other levels. Both tables highlight the challenge of class imbalance and the need for a strategic model training approach.

Table 1. Data distribution concerning disaster event classification task

| SI. NO | Disaster events | Train data | Test data |
|--------|------------------------|------------|-----------|
| 1 | Damaged_infrastructure | 1009 | 258 |
| 2 | Damage_nature | 347 | 114 |
| 3 | Fires | 294 | 56 |
| 4 | Flood | 261 | 50 |
| 5 | Human_damage | 180 | 42 |
| 6 | Non_damage | 2177 | 513 |

Table 2. Data distribution concerning disaster severity identification task

| SI. NO | Disaster events | Train data | Test data |
|--------|-----------------|------------|-----------|
| 1 | little_or_none | 2900 | 700 |
| 2 | severe | 853 | 209 |
| 3 | mild | 515 | 121 |

2.3. Sequence learner

This section discusses the approach taken to design a deep sequence learner system for classification of disaster events and identification of their damage severity. This phase of the research implements the LSTM model as a sequence learner because it is best suited to capture long-term dependencies on the text (tweets) of each data sample. Furthermore, the proposed sequence learner combined with the attention mechanism allows the proposed sequence learner model to focus on the most important parts of the text sequence, which can improve the accuracy of the classification task. Due to data imbalance, especially with regard to damage severity tasks, integrating attention mechanisms can help reduce overfitting by preventing the LSTM model from memorizing the training data too closely, thereby promoting better generalization to unseen data. However, before training our sequence learning model, the text data needs to be preprocessed for feature extraction in the embedding layer. The study first converts the textual phrases into numerical vectors, and then applies a word embedding model to them to obtain the semantic and temporal representation of the textual phrases. The obtained embedding vectors are then used to train a sequence learner model, which learns the context-level properties of each phrase. Then, exploit the potential of the attention mechanism to retain the most important features from sequential layers

2.3.1. Text vector construction

The initial step of text vectorization preprocesses the textual data to remove emojis, symbols, punctuations, web URLs, and excess spaces. By employing Python libraries such as BeautifulSoup and emoji, the study ensures a clean and standard format for all text data. Once cleaned, the textual data is converted into its respective string data type for consistency. The distribution of text length for each data sample is analyzed as depicted in Figure 4.

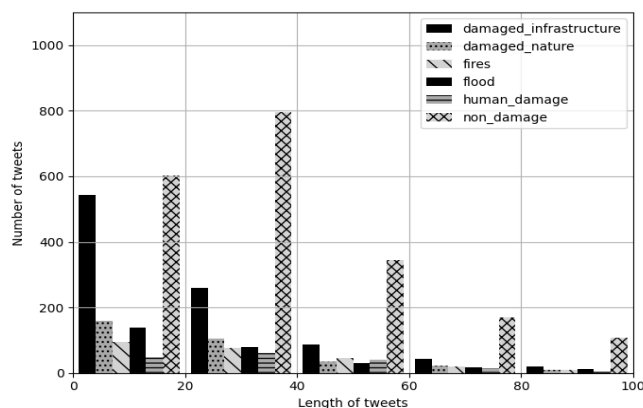


Figure 4. Distribution of text length concerning each distinct disaster class

Based on visual inspection of Figure 4, it can be found that most texts are around 150 words in length. Therefore, a length of 150 is chosen as the padding criterion to ensure that all sequences are of consistent length. The tokenizer is implemented using a defined vocabulary of 50,000 words to create an initial vector representation of the tweet. This procedure ensures that certain symbols, punctuation marks, and other

characters are filtered out during the tokenization process. If the tokenizer encounters a word outside of its vocabulary, it represents it using the token "<oov>". The tokenizer converts each tweet into a sequence of single-word indexes. However, since these sequences may vary in length, they are padded to a fixed length of 150. Visual inspection of the encoding and padding sequences reveals the transformation process. For instance, the text: "a Sri Lankan man removes belongings from his submerged house at Wehangalla Village in the Kalutara District..." is converted into a sequence like [5, 7503, 12612, 528, ...] and subsequently, it is padded to ensure a fixed length of 150. In the provided dataset, the training set comprises 4268 sequences, and the test set includes 1033 sequences. An important observation is the actual vocabulary size, which turned out to be 36,195, smaller than the initially defined 50,000, which indicates that the dataset contains 36,195 unique words.

2.3.2. Text embedding

When processing text data, text embedding is the basis of deep learning models. As the name suggests, they embed words into a continuous mathematical space so that learning models can understand and process them. The embedding process converts discrete symbols (such as single words) into fixed-size dense vectors. In the context of the proposed predictive learning task, the embedding layer converts integer indices (obtained from tokenization) into dense vectors of fixed size 150. This layer is crucial because it projects words into a continuous vector space, where semantically similar words are mapped to nearby points. The size of the embedding space is determined by the number of words in the vocabulary, which is 36,195 in this dataset. Good embeddings can capture the semantic relationships between words. The embedding layer acts as a lookup table to retrieve a dense vector representation of each word during training. These vectors are then adjusted during training to reduce the loss, making the embeddings more accurate and suitable for classification tasks. Therefore, word embeddings capture a large amount of information about the dataset and serve as compact and dense representations of words, capturing their meanings and relationships.

2.3.3. Long short-term memory layers

This study employs a bidirectional long short-term memory (Bi-LSTM) model, which is further integrated with an attention mechanism to generate contextual representations of the input text from both the forward and backward directions [25]. The sequence input goes through the embedding layer and then into the Bi-LSTM layer. The model consists of two Bi-LSTM layers, where the first layer has 128 units and the second layer has 64 units. The second layer returns sequences and states. The forward and backward hidden states of the second layer are connected as (1):

$$\mathcal{H}_t = \overrightarrow{\mathcal{H}}_t; \overleftarrow{\mathcal{H}}_t \quad (1)$$

Where $\overrightarrow{\mathcal{H}}_t$ and $\overleftarrow{\mathcal{H}}_t$ are the hidden states at time t from the forward and backward LSTMs respectively. Basically, the adoption of Bi-LSTM mitigates the vanishing gradient problem associated with long sequential data. This model allows the learning to access both past ($\overrightarrow{\mathcal{H}}_t$) and future ($\overleftarrow{\mathcal{H}}_t$) temporal information, which is essential for understanding the semantics of a given context in the input text phrases. However, not all words in a text tweet contribute equally to its classification. Therefore, the study employed a weighted attention mechanism to highlight the most important words during text classification. The attention mechanism is applied to the concatenated hidden states from the Bi-LSTM layer in the proposed sequence learner model. For each hidden state \mathcal{H}_t , an attention score is computed, indicating the importance of the corresponding word in the sequence. The computation of the attention score is given as (2):

$$Score(\mathcal{H}_t) = \tanh(W_1 \mathcal{H}_t + W_2 \bar{h}) \quad (2)$$

where W_1 and W_2 are learnable weight matrices, and \bar{h} is the concatenated forward and backwards hidden states. The attention mechanism assigns a weight to each word feature from the sequential layer, focusing on the output labels given as (3):

$$\alpha_t = \frac{\exp(Score(\mathcal{H}_t))}{\sum_{t'=1}^T \exp(Score(\mathcal{H}_{t'}))} \quad (3)$$

where α_t denotes attention weight for hidden state \mathcal{H}_t and T is the sequence length. A weighted sum operation generates an attentive feature vector (consists a summarized representation of input sequences) for each text, numerically given as (4):

$$\mathcal{C} = \sum_{t=1}^T \alpha_t \mathcal{H}_t \quad (4)$$

The context vector C is then calculated as the weighted sum of all hidden states based on the attention weights, enabling the model to focus more on the relevant parts of the input sequence. The context vector, which encapsulates the most essential information from the sequence, is then passed through a dense layer with rectified linear unit (ReLU) activation. The advantage of the attention mechanism is its ability to focus on specific parts of the input sequence, essentially weighing them based on their relevance to the current processing step.

2.4. Visual feature extractor

The proposed study adopted a concept of transfer learning by implementing the pre-trained CNN model ResNet50 [26] as a higher-level feature extractor that leverages residual learning to train very deep networks. The key idea behind ResNet is introducing a skip connection that bypasses one or more layers. The operation in a residual block can be represented mathematically as (5):

$$Op = F(Ip) + Ip \quad (5)$$

Where Op denotes to the output, $F(\cdot)$ is a mapping function that transforms the input data (Ip) features to the output class following the convolutive layers in each block. The ResNet50 model has learned a variety of features from a diverse dataset containing millions of images from thousands of classes. The model weights are initialized with the values obtained during the pre-training. The mathematical representation of training involves minimizing a loss function L using optimization algorithms like stochastic gradient descent (SGD), which updates the weights W of the trainable layers based on the gradient of the loss concerning the weights:

$$W_{new} = W_{old} - \alpha \frac{\partial L}{\partial W} \quad (6)$$

Where α is the learning rate and $\partial L / \partial W$ is the gradient of the loss with respect to the weights. After ResNet50 is implemented, a global average pooling operation layer is applied to reduce the spatial dimensions of the feature maps, numerically given as (7):

$$\mathcal{F}_i, avg = \frac{1}{\mathcal{W} \times \mathcal{H}} \sum_{w=1}^{\mathcal{W}} \sum_{h=1}^{\mathcal{H}} \mathcal{F}_{i,w,h} \quad (7)$$

Where \mathcal{W} and \mathcal{H} are the width and height of the feature map \mathcal{F}_i and $\mathcal{F}_{i,w,h}$ refers to the value at spatial position (w, h) in the feature map \mathcal{F}_i . The pooled output is then flattened and fed into a dense layer, which performs a linear operation followed by a non-linear activation function given as (8):

$$output = ReLU(W \cdot \mathcal{F}_i + b) \quad (8)$$

Where W denotes the weights, b refers to bias, and $ReLU$ is the non-linear activation function. The extracted features are then ready to be concatenated with features from other modalities (e.g., text) for further processing and final predictions.

2.5. Projection layer

In the proposed model, the projection layer is responsible for integrating the features extracted from both visual feature extractor \mathcal{V}_i and sequential learner models \mathcal{T}_i to form a cohesive and unified representation to construct a fused representation, ensuring balanced contributions from visual and textual information. The fused feature vector, \mathcal{F}_i , is numerically expressed as (9):

$$\mathcal{F}_i = \mathcal{V}_i \oplus \mathcal{T}_i \quad (9)$$

Where \oplus denotes the concatenation operation. This unified feature vector then progresses through an ensuing hidden layer encompassing n neurons and subsequently reaches a SoftMax layer to conduct classifications. To prevent overfitting, a dropout layer is strategically placed before the hidden layer. The transmutation of the fused feature vector through the hidden layer can be mathematically represented as (10):

$$\mathcal{H}_i = ReLU(\mathcal{W}_Z \cdot \mathcal{F}_i + b_Z) \quad (10)$$

Where \mathcal{H}_i symbolizes the output from the hidden layer and is of the dimensions \mathcal{W}_Z and b_Z are the weight matrix and bias vector affiliated with this hidden layer. Finally, the output from the hidden layer navigates through the SoftMax layer to determine the probability distribution over the classes. The next section discusses the performance analysis of the proposed system for disaster event classification and its damage severity analysis.

3. RESULT AND DISCUSSION

The system was designed and developed on the Python programming language running in the Anaconda distribution. For a better experience, model execution needs to support specific compute stacks of Nvidia and CUDA packages. Experimental analysis and research results were obtained running the proposed model on Windows 10 with 16 GB RAM and GPU GTX 1660 Ti. This section presents the results obtained by the proposed system, which is evaluated based on widely adopted classification metrics such as accuracy, precision, recall, and f1-score. The results of the proposed model are discussed for both unimodal and multimodal data in terms of both visual and numerical outcome discussion.

Figure 5 shows a confusion matrix analysis for a unimodal text-based approach in disaster event detection and damage severity assessment. Analysis of the confusion matrix in Figure 5(a) shows that the model performs well in predicting no damage (ND) events. However, the model has some confusion between damaged_infrastructure (DI), and damaged_nature (DN) classes as there are more misclassified instances. The classification of fire-related disasters and floods is generally accurate but there is a little confusion with other disaster types. Also, human damage (HD) shows some room for improvement in the model as there are 13 misclassified instances out of a total of 42 instances. On the other hand, analysis of the confusion matrix in Figure 5(b) demonstrates the effectiveness of the model in identifying correct instances with low damage, but it struggles more with mild damage with zero correct classifications, and misclassifies it as low and severe. Overall, the analysis highlights both strong performance areas and opportunities to refine the text-based disaster classification. The next analysis is presented for the unimodal image-based approach.

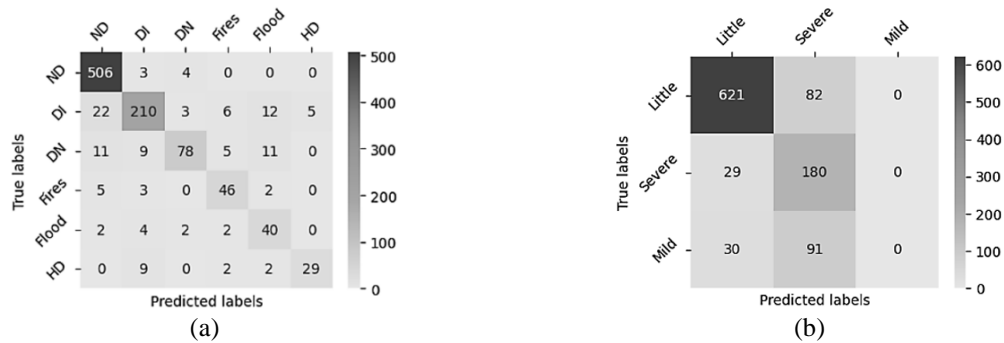


Figure 5. Confusion matrices for unimodal text classification performance: (a) disaster event detection and (b) damage severity assessment

Figure 6 presents the confusion matrix analysis for unimodal image classification performance in disaster event detection and damage severity assessment. In Figure 6(a), the image classification model exhibits slightly better results than the text-based model in most classes, except for DN and flood, where the performance slightly decreases. Figure 6(b) indicates that the model performed well in identifying low damage, but faced difficulty in distinguishing between mild and severe damage, although it still outperforms the text-based unimodal approach. The next analysis focuses on the multi-modal approach combining image and text data using a hybrid learning model.

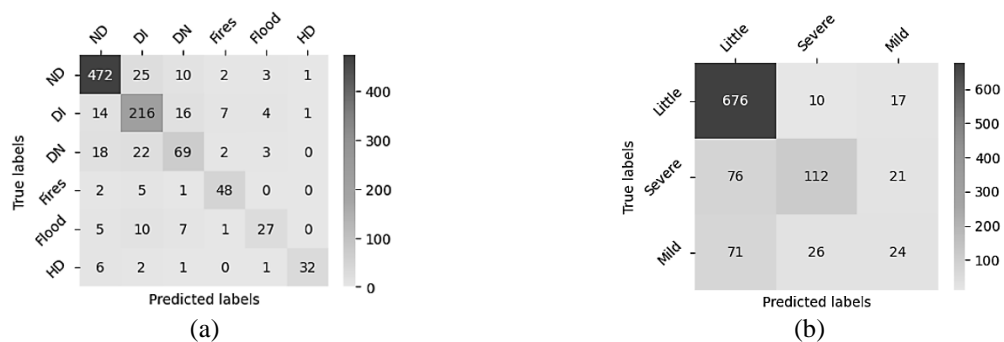


Figure 6. Confusion matrices for unimodal image classification performance: (a) disaster event detection and (b) damage severity assessment

Figure 7 shows the confusion matrix analysis of the multimodal image-text classification performance. The analysis of Figure 7(a) shows that the multimodal model combining images and text performs robustly in identifying no disasters (ND), and also for most disaster types, its results are improved over the unimodal approach. Figure 7(b) shows that the model further outperforms the unimodal based approach in terms of damage severity, and has higher accuracy in detecting small damages as well as minor and severe categories, compared to the text and image unimodal. The overall analysis shows that the multimodal approach has a significant improvement in improving the classification accuracy. The next analysis is presented in Table 3 with consolidated outcome statistics for each classification task with the different learning models.

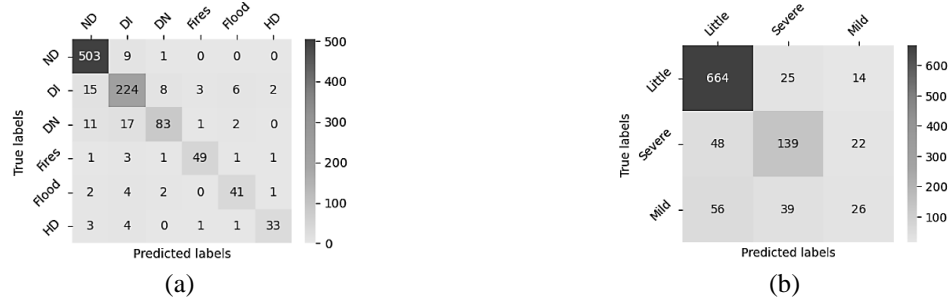


Figure 7. Confusion matrices for multi-modal image-text classification performance: (a) disaster event detection and (b) damage severity assessment

Table 3. Consolidated outcome statistics for unimodal and multi-modal classification models

| Model Type | Classification task | Accuracy | Precision | Recall | F1-Score |
|----------------------------|---------------------|----------|-----------|--------|----------|
| Unimodal (text data) | Disaster event | 88.0 | 88.4 | 88.0 | 87.85 |
| | Damage severity | 77.54 | 72.47 | 77.54 | 74.08 |
| Unimodal (image data) | Disaster event | 83.64 | 83.53 | 83.64 | 83.44 |
| | Damage severity | 78.61 | 75.74 | 78.61 | 76.06 |
| Multi-modal (text + image) | Disaster event | 90.31 | 90.20 | 90.31 | 90.15 |
| | Damage severity | 80.25 | 76.60 | 80.25 | 78.41 |

The text-based unimodal disaster event classification method exhibited a high accuracy of 88.0% and precision of 88.4%, indicating that most disaster events classified by the model were correct. The recall and F1 scores were 88.0% and 87.85%, respectively, indicating that the model was unable to capture most instances correctly across all categories. A similar analysis can be performed in the case of damage severity identification. The model achieved an accuracy of 77.54% and an F1 score of 74.08%. On the other hand, unimodal image-based methods show slightly better performance on both classification problems. Image-based unimodal using a CNN model achieves 83.64% accuracy and 83.44% F1 score in classifying disaster events. In the damage severity recognition task, it effectively outperforms the text-based unimodal method with 78.61% accuracy and 76.06% F1 score. The proposed multimodal hybrid learning model integrating text and image data was found to be the most effective among the two single modalities. It achieved a higher accuracy of 90.31% and an F1 score of 90.15% in disaster event classification. Moreover, in the performance evaluation of damage severity identification, the proposed hybrid model outperformed the single modal approach, achieving an accuracy of 80.25% and an F1 score of 78.41%, indicating that the proposed model is very robust in assessing damage severity, which combines the features of text and image data. The proposed classification models were benchmarked and compared and analyzed, as shown in Table 4.

Table 4. The comparative analysis for disaster event classification

| | Modality | Accuracy | Precision | Recall | F1-score |
|-------------------------|----------|----------|-----------|--------|----------|
| Ofli <i>et al.</i> [27] | Text | 80.8 | 81.0 | 81.0 | 80.9 |
| | Image | 83.3 | 83.1 | 83.3 | 83.2 |
| | Multi | 84.4 | 84.1 | 84.0 | 84.2 |
| Proposed | Text | 88.0 | 88.4 | 88.0 | 87.85 |
| | Image | 83.64 | 83.53 | 83.64 | 83.44 |
| | Multi | 90.31 | 90.20 | 90.31 | 90.15 |

Table 4 provides a comparative analysis where the results of the proposed disaster event classification system are compared with similar existing work by Ofli *et al.* [27]. A comparative analysis of different test cases considering single modality (text and image independent) and multimodality (text and image modalities

fused) is shown. The classification model developed by Ofli *et al.* [27] demonstrated consistent performance across different modalities. However, a slight enhancement of around 84% for multimodal context can be observed in terms of performance metrics accuracy and F1 score. On the other hand, the proposed model exhibited better performance, especially reaching 88% accuracy in the context of uni-modality (textual data) and 90% accuracy in multimodality context. Comparative analysis shows that incorporating multi-modal features can significantly improve the robustness and adaptability of learning models to enhance the classification of disaster events and prediction of damage severity.

4. CONCLUSION

This paper has presented the design and development of a hybrid learning system based on spatiotemporal analysis that combines ResNet50 and Bi-LSTM with attention mechanism, and NLP capabilities to identify disaster events in social media posts and assess their damage severity. The incorporation of an attention mechanism plays a crucial role as it enables the model to focus on the most relevant parts of the text, resulting in a contextually richer representation, thereby increases the reliability and accuracy of the proposed model. The study shows that integrating visual and textual data into a single hybrid model can extract comprehensive insights from social media posts and improve the effectiveness of data modeling tasks that require complex analytical capabilities when designing disaster response systems. The experimental outcome demonstrates the potential of the proposed hybrid and multimodal learning scheme in developing immediate response strategies against natural disasters. In the future, the scope of the proposed work will be extended to develop computational models that can disseminate critical information obtained from the proposed model to relevant entities or rescue teams. This will identify the most influential nodes in the network for receiving information about disaster events and their extent of damage.

ACKNOWLEDGEMENTS

The author expresses gratitude to PES University EC Campus Bangalore for their encouragement and support, and declares that this research was undertaken without financial contributions from any external entities.




REFERENCES

- [1] C. H. Oh and J. Oetzel, "Multinational enterprises and natural disasters: Challenges and opportunities for IB research," *Journal of International Business Studies*, vol. 53, no. 2, pp. 231–254, 2022, doi: 10.1057/s41267-021-00483-6.
- [2] R. L. Jones and D. G. -Sapir, and S. Tubeuf, "Human and economic impacts of natural disasters: can we trust the global data ?," *Scientific Data*, vol. 9, doi:10.1038/s41597-022-01667-x.
- [3] T. Ramakrishnan, L. Ngamassi, and S. Rahman, "Examining the factors that influence the use of social media for disaster management by underserved communities," *International Journal of Disaster Risk Science*, vol. 13, no. 1, pp. 52–65, 2022, doi: 10.1007/s13753-022-00399-1.
- [4] N. Kasturi, S. G. Totad, and G. Ghosh, "Research approaches for building analytics in social network towards crowdsourcing," in *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*, Lonavla, India, 2023, pp. 1–7, doi: 10.1109/I2CT57861.2023.10126479.
- [5] M. Karimiziarani and H. Moradkhani, "Social response and disaster management: insights from twitter data assimilation on hurricane ian," *International Journal of Disaster Risk Reduction*, vol. 95, 2023, doi: 10.1016/j.ijdr.2023.103865.
- [6] A. Bhoi *et al.*, "Mining social media text for disaster resource management using a feature selection based on forest optimization," *Computers & Industrial Engineering*, vol. 169, 2022, doi: 10.1016/j.cie.2022.108280.
- [7] H. Harsa *et al.*, "Machine learning and artificial intelligence models development in rainfall-induced landslide prediction," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 1, pp. 262–270, 2023, doi: 10.11591/ijai.v12.i1.pp262-270.
- [8] C. Vermiglio, G. Noto, M. P. R. Bolívar, and V. Zarone, "Disaster management and emerging technologies: a performance-based perspective," *Meditari Accountancy Research*, vol. 30, no. 4, pp. 1093–1117, 2022, doi:10.1108/medar-02-2021-1206.
- [9] V. Linardos, M. Drakaki, P. Tzionas, and Y. Karnavas, "Machine learning in disaster management: Recent developments in methods and applications," *Machine Learning and Knowledge Extraction*, vol. 4, no. 2, pp. 446–473, 2022, doi:10.3390/make4020020.
- [10] R. Dubey, D. J. Bryde, Y. K. Dwivedi, G. Graham, and C. Foropon, "Impact of artificial intelligence-driven big data analytics culture on agility and resilience in humanitarian supply chain: A practice-based view," *International Journal of Production Economics*, vol. 250, 2022, doi: 10.1016/j.ijpe.2022.108618.
- [11] N. B. Jarah, A. H. H. Alasadi, and K. M. Hashim, "Earthquake prediction technique: a comparative study," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 3, pp. 1026–1032, 2023, 10.11591/ijai.v12.i3.pp1026-1032.
- [12] U. G. Inyang, E. E. Akpan, and O. C. Akinyokun, "A hybrid machine learning approach for flood risks assessment and classification," *International Journal of Computational Intelligence and Applications*, vol. 19, no. 2, pp. 2050012, 2020, doi: 10.1142/s1469026820500121.
- [13] S. Deb and A. K. Chanda, "Comparative analysis of contextual and context-free embeddings in disaster prediction from Twitter data," *Machine Learning with Applications*, vol. 7, 2022, doi:10.1016/j.mlwa.2022.100253.
- [14] A. Khattar and S. M. K. Quadri, "Generalization of convolutional network to domain adaptation network for classification of disaster images on twitter," *Multimedia Tools and Applications*, vol. 81, pp. 30437–30464, 2022, doi: 10.1007/s11042-022-12869-1.
- [15] S. Akter and S. F. Wamba, "Big data and disaster management: a systematic review and agenda for future research," *Annals of Operations Research*, vol. 283, no. 1–2, pp. 939–959, 2019, doi: 10.1007/s10479-017-2584-2.




- [16] M. Anbarasan *et al.*, "Detection of flood disaster system based on IoT, big data and convolutional deep neural network," *Computer Communications*, vol. 150, pp. 150–157, 2020, doi: 10.1016/j.comcom.2019.11.022.
- [17] Y.-G. Zhang, J. Tang, Z.-Y. He, J. Tan, and C. Li, "A novel displacement prediction method using gated recurrent unit model with time series analysis in the Erdaohu landslide," *Natural Hazards*, vol. 105, no. 1, pp. 783–813, 2021, doi: 10.1007/s11069-020-04337-6
- [18] X. Ge *et al.*, "Disaster prediction knowledge graph based on multi-source spatio-temporal information," *Remote Sensing*, vol. 14, no. 5, 2022, doi: 10.3390/rs14051214
- [19] D. F. Sufi and I. Khalil, "Automated disaster monitoring from social media posts using AI based location intelligence and sentiment analysis," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 4, pp. 4614–4624, Aug. 2024, doi: 10.1109/TCSS.2022.3157142.
- [20] T. Yigitcanlar *et al.*, "Detecting natural hazard-related disaster impacts with social media analytics: The case of Australian states and territories," *Sustainability*, vol. 14, no. 2, 2022, doi: 10.3390/su14020810
- [21] Z. Zou, H. Gan, Q. Huang, T. Cai, and K. Cao, "Disaster image classification by fusing multimodal social media data," *ISPRS International Journal of Geo-Information*, vol. 10, no. 10, 2021, doi: 10.3390/ijgi10100636.
- [22] A. Asif *et al.*, "Automatic analysis of social media images to identify disaster type and infer appropriate emergency response," *Journal of Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00471-5.
- [23] M. Zhang, Q. Huang, and H. Liu, "A multimodal data analysis approach to social media during natural disasters," *Sustainability*, vol. 14, no. 9, 2022, doi: 10.3390/su14095536.
- [24] "CrisisBench: benchmarking crisis-related social media datasets for humanitarian information processing," *Crisis NLP-Natural Language Processing*, 2019. Accessed: September 21, 2023. [Online]. Available: https://crisisnlp.qcri.org/crisis_datasets_benchmarks
- [25] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, 2019, doi: 10.1016/j.neucom.2019.01.078.
- [26] S. R. Shah, S. Qadri, H. Bibi, S. M. W. Shah, M. I. Sharif, and F. Marinello, "Comparing inception V3, VGG 16, VGG 19, CNN, and ResNet 50: A case study on early detection of a rice disease," *Agronomy*, vol. 13, no. 6, 2023, doi: 10.3390/agronomy13061633.
- [27] F. Ofli, F. Alam, and M. Imran, "Analysis of social media data using multimodal deep learning for disaster response," *arXiv-Computer Science*, pp. 1–10, 2020, doi:10.48550/arXiv.2004.11838.

BIOGRAPHIES OF AUTHORS






Nivedita Kasturi    has received her M.Tech. Degree in Computer Science and Engineering from BVBCET, Hubli, Karnataka. She is currently working as an Assistant Professor in PES University EC Campus, Bangalore. She had worked in Mindtree Ltd as C# and .NET Programmer. She has conference and journal publication. Her research interest is on software engineering, data analytics. Currently she is a Ph.D. student at KLE Technological University and working the domain of social network and crowdsourcing. She can be contacted at email: nk5883933@gmail.com or niveditak@pes.edu.



Dr. Shashikumar Guruputra Totad    received the B.E. degree in Computer Science and Engineering from Gogte Institute of Technology, Belgaum (Karnataka University, Dharwad) in 1990 and M.E. in Computer Science and Engineering from Walchand College of Engineering (Shivaji University, Kolhapur) in 2003. He did his Ph.D. at Jawaharlal Nehru Technological University, Hyderabad. He worked as lecturer at KLE's College of Engineering Technology, Belgaum during 1991–2000 and as Assistant Professor at BVB College of Engineering during 2001–2006. He is currently working as Professor of School Computer Science and Engineering at KLE Technological University, Hubballi, India. He has published over 43 papers in national and international conferences and journals. He has guided over 15 post graduate students and guiding at present 5 doctoral research scholars. His research interests include data mining, distributed databases, and mobile agents. He is a life member of ISTE and CSI. He can be contacted at email: totad@kletech.ac.in.



Dr. Goldina Ghosh    has received her Ph.D. Degree from Birla Institute of Technology, Mesra, Ranchi. She is currently working as an Associate Professor in Institute of Engineering and Management, Salt Lake, Kolkata. She has several journal papers and two book chapters. Her research interest is on Social network analysis using hybrid intelligence, bio inspired optimization techniques and soft computing. Currently her research interest has been extending to crowdsourcing concept with social networking using machine learning techniques. She can be contacted at email: goldina.ghosh@iemcal.com.