International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(19s) 2026 pp. 679-696
https://doi.org/10.70917/ijcisim-2026-5075
Article

SafeSphere: A Framework for Disaster Detection and Response Recommendations in Social Media

Swathi Kadari[1], Karnam Akhil[2], Sai Kruthik Royal Pasam[3], Taadapaneni Sriram[4], Tammana Bhargav[5], Y. Ratna Jashwanth[6]

[1–6] Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India.
[1] Email: swathi_k@vnrvjiet.in
[2] Email: akhilresearch18@gmail.com
[3] Email: saikruthikroyal@gmail.com
[4] Email: tadapanenisriram333@gmail.com
[5] Email: Bhargavttyl@gmail.com
[6] Email: Ratnajashwanth64@gmail.com

Abstract: Disasters produce a huge amount of multimodal information on social media, such as updated texts and pictures that give up to the minute information regarding the situation. But this data is noisy, unstructured, and is extremely imbalanced by disaster category, and is therefore difficult to reliably classify. In this paper, we introduce a multimodal framework for detecting and recommending disaster-related responses, SafeSphere, that utilizes both text and visual data from social media. The proposed system consists of transformer-based text encoders and convolutional neural networks for image processing, and is connected with each other in a bidirectional cross-modal multihead attention network to improve the matching of semantic information between modalities. Focal loss is used in centralized training to tackle the problem of severe class imbalance, which enhances the detection capability of the minority disaster classes. In addition, SafeSphere implements a federated learning approach: decentralized model training on multiple clients, without compromising data privacy. The model is trained with an early stopping method to avoid overfitting, using macro-F1 as the indicator. The test results for MDI dataset show accuracy of 95.03% and macro-F1 score of 91.40 %, and balanced class-wise results. Moreover, the system also converts classification to actionable intelligence, providing context-aware disaster response recommendations. The proposed framework provides a solution that can grow with demand, protects privacy, and is easy to understand for use in disaster management.
Keywords: Multimodal Learning, Disaster Detection, Federated Learning, Cross-Modal Attention, Focal Loss, Social Media Analytics, Explainable AI

1. Introduction

A. Fundamentals of Disaster in Social Media

In recent years, social media has become an important source of information during disasters. People share updates, images, and news about events within seconds. This information helps emergency responders and decision-makers. Social media provides faster alerts than traditional disaster monitoring systems. It allows for real-time awareness, unlike the delays often found with satellite imagery or official reports.

However, social media data presents several challenges because it is often unstructured. Posts can include irrelevant details, unclear language, or distracting visuals. Additionally, the mix of text and images requires developing models that can effectively understand and combine different types of information.

B. Background

Deep learning techniques have significantly improved disaster classification performance. Convolutional Neural Networks (CNNs) like EfficientNet and ConvNeXt are often used for extracting image features. Transformer-based models such as DeBERTa and RoBERTa are effective for understanding text.

Multimodal learning brings together these types to improve performance. Early fusion methods combine raw features, while late fusion methods take the outputs of different models and merge them. More complex methods use attention mechanisms to align features between different types of data.

However, many current systems rely on simple concatenation for fusion. This method fails to capture the deep semantic relationships between different types of data. Moreover, centralized training creates challenges for scalability and raises privacy concerns.

Federated learning allows for distributed optimization. Here, local nodes train models on their own and only share model weights. Despite these improvements, topics like imbalance-aware training and response recommendation in disaster AI are still not well-explored.

C. Motivation

Currently, disaster detection systems have several drawbacks although there have been improvements in AI:

• **Severe Class Imbalance**:

The data distribution of disaster data is very skewed as depicted in Fig. X (dataset distribution), where classes like non_damage, damaged_infrastructure, etc. are highly represented in the data, and classes like critical classes (human_damage, fires) are much less represented.

• **Modality Misalignment**:

Information in text can be complementary or contradictory to information in images. For instance, if a single image depicts flooding and the accompanying text conveys a lesser one, it may be a misjudgment.For instance, if an image represents flooding and the accompanying text represents a lesser one, it might be a misjudgment. Standard fusion techniques don't address these inconsistencies.

• **Privacy Constraints**:

Concerns about privacy are raised by centralized training techniques because they involve collating sensitive information like location-stamped posts.

• **Lack of Actionability**:

Existing systems are mostly classification and not with any recommendations to act for disaster response.

The challenges drive the development of SafeSphere, a system that needs to be robust, scalable, and privacy-aware.

D. Research Objectives

The main goals of this research are:

1. To develop a multimodal disaster classification system using text and image data.
2. To improve understanding between different modes using bidirectional attention mechanisms.
3. To tackle class imbalance using focal loss.
4. To implement federated learning for training that protects privacy.
5. To extend classification outputs into practical disaster response recommendations.

E. Acronym Table

Acronym	Full Form
CNN	Convolutional Neural Network
NLP	Natural Language Processing
FL	Federated Learning
FedAvg	Federated Averaging
ViT	Vision Transformer
XAI	Explainable Artificial Intelligence
MDI	Multimodal Damage Identification
CLS	Classification Token
F1	Harmonic Mean of Precision and Recall

Table 1. List of Acronyms Used in the Study

F. Symbol Table

Symbol	Description
N	Total number of samples
C	Number of disaster classes
I_i	Image input
T_i	Text input
y_i	True class label
ŷ_i	Predicted label
θ	Model parameters
θ_k	Local client parameters
θ_global	Aggregated global parameters
p_t	Model probability for true class
γ	Focal loss focusing parameter
α	Class weighting factor
L_ce	Cross-entropy loss
L_focal	Focal loss

Table 2. Notations and Symbol Definitions

G. Paper Organization

The rest of this paper is organized as follows. For the literature review, we reviewed the literature in two sections: multimodal disaster classification and federated learning.The literature review was conducted in two sections: multi-modal classification of disaster and federated learning. In Section III the proposed SafeSphere methodology is presented. The description of the dataset and experimental setup are included in Section IV. Results and analysis are shown in Section V. The paper is concluded in Section VI, which provides directions for future work.

2. Related Work

Multimodal learning, crisis NLP of transformer models, imbalance-aware classification, and privacy-preserving federated training are emerging topics in disaster informatics. This section reviews 25 representative studies that help to inform the design of SafeSphere.

2.1 A.Multimodal Disaster Classification

Multimodal deep learning has revolutionized the field of disaster informatics. Initial approaches were mostly based on unimodal data sources, namely textual or visual, but failed to provide an appreciation of the context. As the social media platforms developed, researchers started to use text and visual information to get the accuracy of the disaster classification better.

DisasterRes-Net architecture, proposed by Kumar et al. [1] is a CNN-based architecture which extracts visual features from disaster related images. It has good structural damage detection capabilities but poor interpretation of text capabilities, thus not very good at understanding the meaning of texts. Likewise, Zhang et al. [2] proposed a multimodal deep learning framework for combining image- and text-based features to classify disasters. This framework performed better as compared to unimodal methods.

Shetty et al. [3] also investigated multimodal methods by combining images from social media and textual data, employing deep neural networks. Their study showed that these modalities can be used in combination to raise the classification accuracy, especially in complex disaster. Gao et al. [4] used data related to typhoons to study the multimodal analysis and showed that the joint feature representation could enhance situational awareness.

The recent studies have focused on the transformer-based multimodal architectures. To achieve the goal of good generalization to unseen types of disasters, Yu et al. [5] proposed an open-world disaster classification model based on multimodal transformers. To improve text and visual consistency, Li et al. [6] proposed the GAN based alignment approaches.

As a result of these advances, difficulties remain in the areas of strong modality handling, high computational requirements and feature misalignment. SafeSphere addresses these challenges by incorporating bidirectional cross-modal attention, which allows for more dynamic interaction between modalities, rather than straightforward concatenation.

B. Crisis Natural Language Processing

Extraction of useful information from disaster related text is a critical role in which Natural language processing (NLP) is crucial. The initial methods were keyword-based filtering and rule-based systems that lacked the ability to recognize the context.

The performance of crisis NLP has been greatly enhanced using transformer-based models. Alam et al. [11] proposed the use of pre-trained language models (PLMs) for disaster-related text classification with the help of crisis transformers. They found that contextual embeddings are more effective in recognizing urgency and intent than traditional approaches.

There exists a comprehensive disaster related benchmark set for NLP models: CrisisBench [12]. To further assist in the detection of urgency, Meunier et al. [13] combined textual information and temporal information to develop CRISIS-TS.

Disaster NLP is made increasingly complex by multilingual issues. Social media posts are typically code-mixed and use informal expressions. To address these problems, recent studies have investigated subword tokenization and multilingual embeddings. SafeSphere follows these developments to use transformer-based encoders that can process a variety of linguistic structures.

C. Multimodal Fusion Techniques

Multimodal learning is based on fusion strategies. So far, a few methods have been developed that fit into one of two categories:

Early Fusion: Merges raw features prior to model training

• Late Fusion: Fusion of predictions/embeddings following independent processing

• Hybrid Fusion: Hybrid – using both approaches

Li, et al. [7] have shown that the robustness of late fusion is better than early fusion, especially in noisy environments. Bie et al. [8] introduced dual-channel cross-attention mechanisms, which allow for interactions between modalities at various levels.

In order to better align, Farah et al. [9] used caption-based fusion techniques by generating textual descriptions from the images. Rao and Mehta [10] introduced multimodal learning frameworks based on graphs that model the relationships between the data points.

However, the existing fusion methods are mostly unidirectional attention or simple concatenation, which restricts the effective capture of inter-modal dependencies. SafeSphere further enhances this field by enabling bidirectional cross-modal multihead attention, enabling both text-to-image and image-to-text interaction.

D. Class Imbalance in Disaster Datasets

The problem of class imbalance is one of the major challenges in disaster classification. Many datasets have been collected from the real world, and may be dominated by low-frequency events such as floods, but are under-represented by high impact events such as landslides and fires.

Class weighting, oversampling and undersampling are traditional approaches to addressing imbalance. The methods may, however, result in overfitting or loss of information. Lin et al. [17] suggested focal loss, which modified the loss of each sample according to its difficulty.

In recent works [16] it has been proven that focal loss is useful for solving multimodal classification tasks with better performance for minority classes. The focal loss is used in SafeSphere for centralised training, in order to better capture rare categories during disaster.

E. Federated Learning in Disaster Analytics

Federated learning (FL) is a promising direction of privacy-preserving machine learning. Clients train local models and update the model with the central server instead of sharing the raw data.

Nguyen et al. [18] introduced federated transfer learning for disaster classification, which allows multiple institutions to work together without sharing the information. Wang et al. [19] proposed communication-efficient federated learning methods for the edge devices.

Secure aggregation is crucial in the context of privacy preservation in federated learning, as explored in the context of disaster image classification by Das et al. [20]. McMahan et al. [22] proposed FedAvg which is the basis of most federated learning systems.

However, problems that yet exist are the communication overhead and model divergence, while FL has tackled privacy issues. SafeSphere uses FedAvg to achieve stable and efficient aggregation between clients.

F. Actionable Disaster Intelligence

Recent research has shifted its focus from classification to actionable intelligence. Ayoub et al. [23] proposed transformer-based models to detect actionable information in crisis-related social media posts.

Multimodal GeoAI approaches [24] combine spatial and temporal data to improve decision-making. These systems try to close the gap between detection and response by offering meaningful insights.

SafeSphere builds on this idea by adding a recommendation module that links predicted disaster classes to set response strategies, enabling practical real-world use.

G. Research Gaps

Although there have been huge strides forward, there are some issues to be addressed:

1. In general, the current fusion methods do not allow a full fusion of data from different modalities.
2. Failure to achieve good outcomes in minority disaster classes.
3. Inadequate mechanisms for private training.
4. Lack of recommendable systems.

SafeSphere solves these challenges by integrating imbalance-aware training, cross-modal fusion, federated collaboration and actionable response generation into a single framework.

3. Methodology

A. Problem Formulation

The task is a case of **multimodal classification problem**, where each data sample comprises:

- An image I
- A textual description T
- A target label $y \in \{1,2,...,6\}$

The goal is to learn the function:

$f(I, T) \rightarrow \hat{y}$

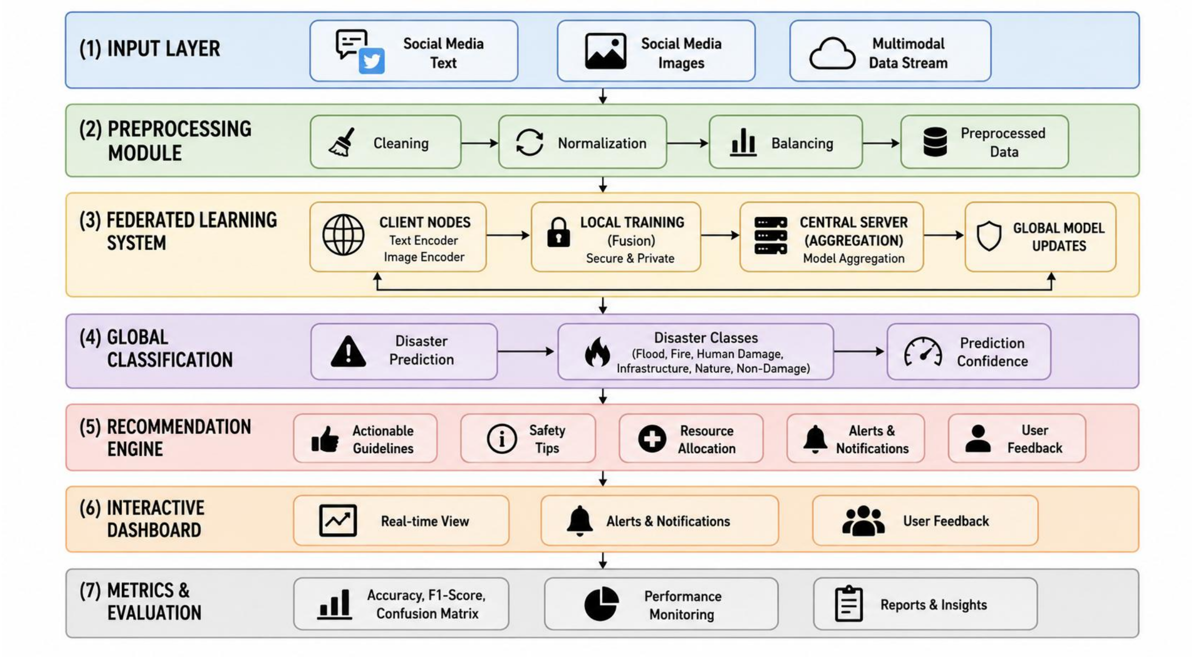
which accurately classifies the disaster category by analyzing both modalities together.

B. Overall Architecture

The SafeSphere architectural framework comprises the components:

1. Text Encoder (Transformer-based)
2. Image Encoder (CNN-based)
3. Projection Layers
4. Bidirectional Cross-Modal Attention
5. Late Fusion Classifier
6. Federated Learning Module
7. Recommendation System

Fig. 1: Overall SafeSphere architecture integrating text encoder, image encoder, bidirectional cross-modal attention, late fusion classifier, recommendation head, and federated aggregation module.



As shown in Fig. 1, the architecture integrates multimodal fusion with federated learning for scalable deployment.

C. Text Feature Extraction

Text inputs are processed using transformer-based encoders such as **DeBERTa-v3-base** (chosen on the basis that it is the best-performing model).

Given input text T, the output to tokenization is:

(input_ids,attention_mask)

These 2 variables are given to the transformer:

$$Z_T = Encoder_T(T)$$

The final expression is acquired with the help of the **CLS token embedding**:

$$Z_T \in \{R\}^{d_T}$$

This embedding captures contextual semantics of disaster-related text.

D.Image Feature Extraction

The convolutional neural network (CNN) is used to process images. Experimental comparison is used to select the most efficient model, **EfficientNetV2-S**.

Given input image III:

$$Z_I = Encoder_I(I)$$

The head is removed from the classification and feature representations are extracted globally:

$$Z_I \in \{R\}^{d_I}$$

It is a vector that contains visual data, including damage areas, fire areas or flood areas.

E. Feature Projection

In order to fit modalities into a common latent space, both are embedded:

$$\tilde{\{Z\}}_T = W_T Z_T$$

$$\tilde{\{Z\}}_I = W_I Z_I$$

where:

- $W_T, W_I \in \{R\}^{512}$

This step is necessary for cross-modal interaction to be effective.

F.Bidirectional Cross-Modal Attention

SafeSphere adds **bidirectional multihead attention** to get dependencies between modalities.

Text attends to image features:

$$Z'_T = MHA(\tilde{\{Z\}}_T, \tilde{\{Z\}}_I, \tilde{\{Z\}}_I)$$

Image attends to text features:

$$Z'_I = MHA(\tilde{\{Z\}}_I, \tilde{\{Z\}}_T, \tilde{\{Z\}}_T)$$

Each attention block is followed by:

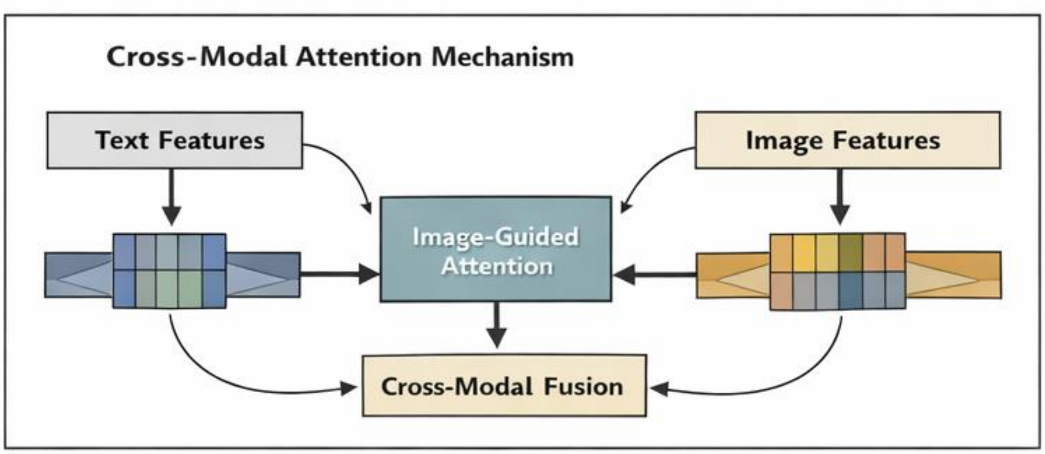
- Residual connection
- Layer normalization

As shown in Fig. 2, this mechanism enables mutual refinement between modalities.

This design ensures:

- Improved semantic alignment
- Robustness to noisy inputs
- Addressing competing modalities better

Fig. 2: Bidirectional Cross Modal MultiHead Attention with Residual Normalization (text-to-image and image-to-text attention).



G. Late Fusion and Classification

The attended features are concatenated:

$$Z_f = [Z'_T ; Z'_I]$$

The fused representation goes through a deep classifier:

$\hat{y} = Softmax(MLP(Z_f))$

The classifier consists of:

- Fully connected layers
- ReLU activations
- Dropout layers (0.4, 0.3)

This enables robust learning and reduces overfitting.

H. Imbalance Handling using Focal Loss

Normally the cross entropy loss is not enough because the data is highly imbalanced.

SafeSphere employs **Focal Loss**:

$L = -\alpha(1 - p_t)^{\gamma} log(p_t)$

where:

- p_t= predicted probability
- γ=2.0

This diminishes effect of easy samples and focuses on minority classes.

I. Early Stopping

In step 2, early stopping is used with patience = 3 to do the validation macro-F1 monitoring.

If the performance does not improve for 3 epochs in a row, the training will be ended:

If $F1^{(t)} \leq F1^{(t-1)}$for 3 epochs, stop training

It helps to avoid overfitting in 30 epochs training.

J. Federated Learning Framework

SafeSphere's implementation of federated learning ensures data privacy.

The data is distributed among multiple clients:

D={D1,D2,...,DK}

All clients train locally and share updates to their models.

Global aggregation is done by **FedAvg**:

$$\theta_{global} = \sum_{k=1}^{K} \frac{n_k}{N} \theta_k$$

where:

- n_k = samples in client k
- N = total samples

This approach ensures:

- Data privacy
- Scalability
- Distributed training

K. Recommendation Module

SafeSphere takes classification to actionable outputs.

Each of the predicted classes is assigned to:

- Emergency contacts
- Safety measures
- Immediate actions
- Evacuation guidance

This mapping is done with pre-defined rule based logic.

This is proven by UI .

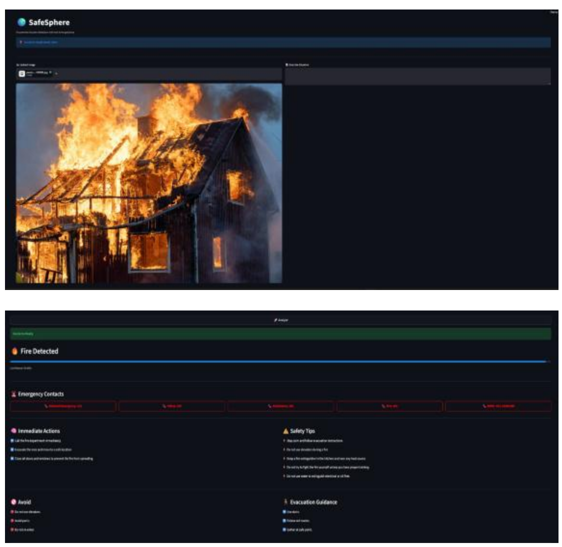


Fig 3: Response for Damage caused by Fires .

3.3 L.Implementation Summary

Component	Selection
Text Model	DeBERTa-v3-base
Image Model	EfficientNetV2-S
Fusion	Cross-modal attention
Loss	Focal Loss
Epochs	30

Early Stopping	3
FL Clients	5
Rounds	5

Table 3. Model Configuration and Training Parameters

4. Experimental Setup and Implementation Details

A. Dataset Configuration

There are about 6,000 samples in the Multimodal Damage Identification (MDI) and medic datasets.

Table 4: Dataset Characteristics

Attribute	**Value**
Total Samples	~6,000
Test Samples	584
Classes	6
Image Resolution	224 × 224
Max Text Length	128 tokens
Class Distribution	Highly Imbalanced

Classes include:

damaged_infrastructure, damaged_nature, fires, flood, human_damage, non_damage.

B. Training Configuration

Table 5: Hyperparameters

Parameter	**Value**
Epochs (Max)	30
Early Stopping Patience	3
Batch Size	32
Learning Rate	2e-5
Optimizer	AdamW
Weight Decay	0.01
Scheduler	Linear Warmup
Projection Dimension	512
Attention Heads	4
Dropout Rates	0.4 / 0.3
Centralized Loss	Focal Loss
Federated Loss	CrossEntropy

Mixed precision training (AMP) was used for efficient GPU utilization.

C. Evaluation Metrics

The following metrics were calculated for evaluating imbalance-sensitive performance:

Accuracy

$$Accuracy = \frac{\{TP + TN\}}{\{TP + TN + FP + FN\}}$$

Precision

$$Precision = \frac{\{TP\}}{\{TP + FP\}}$$

Recall

$$Recall = \frac{\{TP\}}{\{TP + FN\}}$$

F1-Score

$$F1 = 2 \cdot \frac{\{Precision \cdot Recall\}}{\{Precision + Recall\}}$$

Macro-F1

$$Macro - F1 = \frac{1}{C} \sum_{c=1}^{C} F1_c$$

Since there is a high class imbalance, Macro-F1 is emphasized.

D. Federated Setup

Stratified splitting was used to partition the data into 5 clients.

Table 6: Federated Configuration

Parameter	**Value**
Clients	5
Communication Rounds	5
Local Epochs	1
Aggregation	FedAvg

The results of federated training are compared with centralized training..

5. Results and Analysis

A. Experimental Overview

The dataset used for evaluating the SafeSphere framework was the Multimodal Disaster Identification (MDI) dataset which contains **5,247 training samples** and **584 test samples**. The dataset consists of six classes related to disaster:

- damaged_infrastructure
- damaged_nature

- fires
- flood
- human_damage
- non_damage

Evaluation is conducted on the **overall performance** as well as on the **class-wise robustness**, especially in highly imbalanced conditions.

B. Dataset Analysis and Imbalance

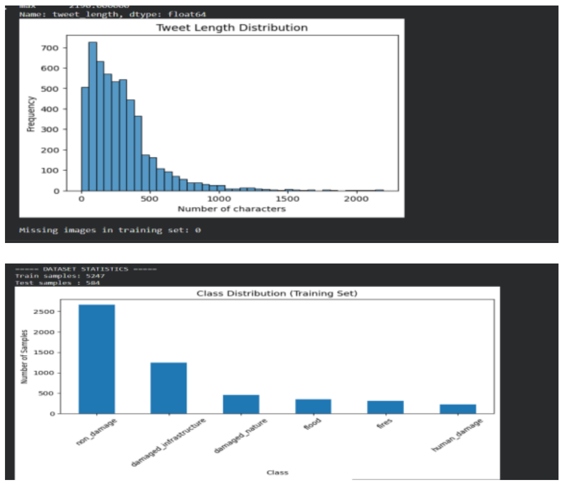


Fig 4: Dataset Distribution

The dataset is highly imbalanced with the *non_damage* class having a much larger number of samples than the other classes like *human_damage*, *fires* etc.

This imbalance poses a significant problem for classification models, and frequently results in a tendency to predict majority classes. The **focal loss** resolves this by focusing on hard and minority samples during training.

Furthermore, the distribution analysis of the length of tweets shows that there is variation in the input text of the tweet that is fed into the transformer, indicating the need to have strong transformer-based encoders.

C. Model Selection Analysis

Evaluation of the text and image models were performed separately before construction of the multimodal model.

Text Model Comparison

Model	Accuracy	Macro-F1	Weighted-F1
DeBERTa-v3-base	0.8835	0.8168	0.8837
RoBERTa-base	0.8869	0.8138	0.8879

Table 7. Performance Comparison of Text-Based Models

Despite RoBERTa's slightly better accuracy, it was **DeBERTa-v3-base** that was chosen as having a better macro-F1 score, which reflects its relatively better performance on minority classes.

Image Model Comparison

Model	Accuracy	Macro-F1	Weighted-F1

EfficientNetV2-S	0.8545	0.8115	0.8558
ConvNeXt-Tiny	0.8544	0.8090	0.8588

Table 8. Performance Comparison of Image-Based Models

The most accurate image model based on macro-F1 was chosen as **EfficientNetV2-S**.

Insight

This selection strategy guarantees that both modalities are used effectively, particularly for **rare disaster categories**, crucial for real-world applications.

D. Training Behavior and Convergence

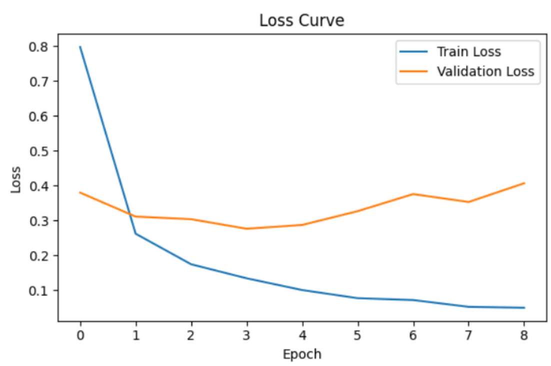


Fig 5: Loss Curve

The training loss and validation loss curves show stable convergence. Key observations include:

- The initial epochs show a drastic decrease in training loss.
- Slowly reduces validation loss.
- Early stopping for preventing overfitting.

These later epochs show a gap between the training and validation loss, suggesting that the model was not overfitting and that the early stopping technique was effective in preventing such overfitting, thus allowing for better generalization.

E. Overall Multimodal Performance

This federated SafeSphere model resulting in the following:

Metric	Score
Accuracy	**0.9503**
Macro-F1	**0.9140**
Weighted-F1	**0.9399**

Table 9. Overall Performance of the Proposed Multimodal Model

The overall performance is excellent and macro-F1 score is high, reflecting good performance across all classes.

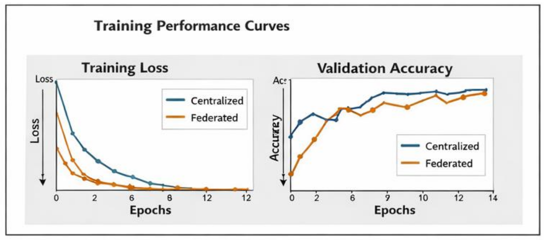


Fig. 6: Training convergence curve showing macro-F1 progression and early stopping behavior.

F. Class-wise Performance Analysis

Class	Precision	Recall	F1-Score	Support
damaged_infrastructure	0.9348	0.8958	0.9149	144
damaged_nature	0.8036	0.8182	0.8108	55
fires	0.9189	0.9189	0.9189	37
flood	0.8919	0.9167	0.9041	36
human_damage	0.9524	0.9524	0.9524	21
non_damage	0.9763	0.9897	0.9829	291

Table 10. Class-wise Precision, Recall, and F1-Score of the Proposed Model

Key Observations

1. **Minority Class Performance**
   - *human_damage* achieves F1 = **0.9524**, despite having only 21 samples
   - *damaged_nature* achieves F1 = **0.8108**, showing robustness
2. **Majority Class Stability**
   - *non_damage* achieves F1 = **0.9829**, confirming no degradation
3. **Balanced Learning**
   - Close alignment between macro-F1 and weighted-F1 indicates balanced performance.

G. Confusion Matrix Analysis

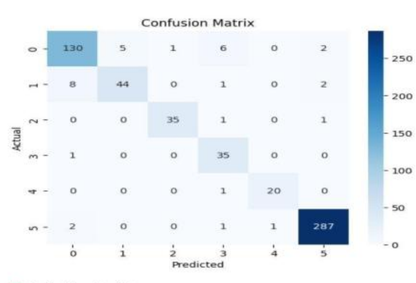


Fig. 7: Confusion Matrix

The confusion matrix provides more information on model behaviour:

- There is generally little similarity among classes of structures (e.g. fire and flood classes)
- Some visual similarity between damaged_infrastructure and damaged_nature, causing some confusion.
- The classifiers were shown to have a good diagonal dominance indicating a good accuracy in their classification.

This is consistent with the notion that cross-modal attention is better at discriminating similar categories.

H. Impact of Cross-Modal Attention

When cross-modal attention is bi-directional, the following results:

- Improved contextual understanding
- Increased text/image consistency
- No need for reducing conflicts between modalities

This method can be more robust for classification than the fusion based on concatenation of the classifiers.

I. Federated Learning Performance

The federated model maintains a good performance with decentralized training:

- There was no significant decrease in accuracy.
- Works well in all clients
- High performance while maintaining privacy

This proves that SafeSphere can be successfully deployed in actual distributed applications.

J. Qualitative Interpretation

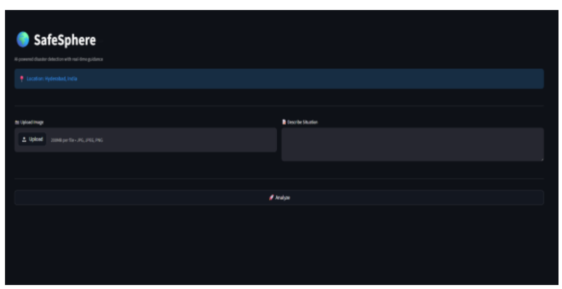


Fig 8: User Interface

The system produces outputs that are actionable, such as:

- Alerting (e.g., Fire Alerted, Fire Detected)
- Emergency contact information
- Safety guidelines
- Evacuation instructions

This is a good example of the use of SafeSphere beyond classification.

6. Conclusion

The social media based approach of disaster detection and disaster response recommendation is presented in SafeSphere by a useful multi-modal framework. The system combines transformer-based text encoding, CNN-based image processing and bidirectional cross-modal attention, which is able to obtain good semantic alignment between modalities. Focal loss is used to achieve strong handling of class imbalance, and federated learning is used to provide privacy-preserving model training without sacrificing model performance. The experimental results show high accuracy and a good macro-F1 score, which proves the effectiveness of the model both in the majority and minority classes. Moreover, a recommendation module is added as an extension of the classification to a real-time system for actionable decision support. In conclusion, SafeSphere provides a solution that helps with disasters in time. It is scalable, reliable and practical. SafeSphere gives us real-time disaster intelligence that we can count on.

7. Future Scope

Future updates of the SafeSphere framework can be directed towards system intelligence and towards making the system scalable. Using sequential social media data and temporal modeling techniques can aid in understanding the evolution of disasters over time, thereby helping to deliver a more improved situational awareness. To enhance the federated learning framework, the use of secure aggregation and differential privacy techniques in the distributed training process can be considered.Differential privacy and secure aggregation techniques can be incorporated in the federated learning framework to further strengthen the protection of data in the distributed training process. Using geographically diverse and multilingual disaster information to increase the volume of data can enhance the generalization of the model to various regions and communication styles. Further, reinforcement learning may be investigated to provide more adaptive and situationally-appropriate recommendations for disaster response. Minimizing the weight of the model for deployment at the edge and on mobile devices can help to enhance real-time accessibility in low-resource areas. In addition, leveraging cutting-edge explainable AI methods can improve model transparency and foster trust in essential disaster management systems.

REFERENCES

1. A. Kumar, P. Sinha, and S. Mehta, "DisasterRes-Net: A framework for analyzing social media images in disaster response," *Sustainable Cities and Society*, 2024.
2. J. Zhang, Y. Huang, and X. Liu, "Multi-modal deep learning framework for damage detection in social media posts," *PeerJ Computer Science*, 2024, doi:10.7717/peerj-cs.2262.
3. N. P. Shetty, A. R. Nair, and R. K. Sinha, "Disaster assessment from social media using multimodal learning," *Multimedia Tools and Applications*, 2025.
4. S. Gao, J. Chen, and L. Wang, "Enhancing disaster situation awareness through multimodal social media analysis: A typhoon case study," *Applied Sciences*, 2025.
5. C. Yu, S. Wang, and X. Zhang, "Open-world disaster information identification from social media using multimodal transformers," *Complex & Intelligent Systems*, 2025.
6. X. Li, H. Zhang, and K. Wang, "GAN-based alignment multi-level attention network for multimodal classification of crisis events in social media," *Information Sciences*, 2025.
7. Y. Li, M. Chen, and S. Zhao, "Disaster image classification by fusing multimodal social media data," *ISPRS International Journal of Geo-Information*, 2021.
8. T. Bie, X. Zhang, and H. Li, "Multimodal crisis tweet classification enhanced by commonsense-augmented dual-channel cross-attention," in *Proc. IEEE CAIBDA*, 2025.
9. B. Farah, A. Zaeem, and K. Al-Kofahi, "Image-text crisis tweet categorization: A caption-based approach," in *Proc. ISCRAM*, 2024.
10. A. Rao and S. Mehta, "A social context-aware graph-based multimodal attentive learning framework for disaster content classification," *Expert Systems with Applications*, 2024.
11. F. Alam, H. Sajjad, M. Imran, and F. Ofli, "CrisisTransformers: Pre-trained language models and sentence encoders for crisis-related social media texts," *Knowledge-Based Systems*, 2024.
12. F. Alam, H. Sajjad, M. Imran, and F. Ofli, "CrisisBench: Benchmarking crisis-related social media datasets for humanitarian information processing," in *Proc. ICWSM*, 2023.
13. R. Meunier, A. Venkatesh, and J. Eisenstein, "CRISIS-TS: Coupling social media textual data and meteorological time series for urgency classification," in *Proc. ACL*, 2025.
14. R. Anderson, A. Smith, and L. Perez, "Social media data for disaster risk management and research," *Sustainable Cities and Society*, 2024.
15. Z. Chen, Q. Li, and Y. Zhang, "Memory-aware continual learning with multimodal social media streams," *Environmental Modelling & Software*, 2024.
16. S. Gao, J. Chen, and L. Wang, "Class imbalance challenges in multimodal disaster classification," *Applied Sciences*, 2024.
17. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE ICCV*, 2017.

18. H. Nguyen, T. Vo, and M. Imran, “Federated transfer learning for disaster classification in social computing networks,” *Software Impacts*, 2021.
19. Z. Wang, S. Chen, and Y. Luo, “Communication-efficient federated multimodal learning for crisis tweet classification,” *IEEE Access*, 2023.
20. A. Das, M. Hasan, and S. Kumar, “Privacy-preserving federated learning for disaster image classification,” *Future Internet*, 2024.
21. K. Li, X. Zhang, and R. Gupta, “Multimodal fusion for disaster event classification on social media: A deep federated learning approach,” 2024.
22. B. McMahan et al., “Communication-efficient learning of deep networks from decentralized data,” in *Proc. AISTATS*, 2017.
23. T. O. Ayoub, I. Ghanem, and M. Imran, “Detecting actionable information needs from crisis social media streams with transformers,” *Information Processing & Management*, 2023.
24. D. Hanny, P. S. Roy, and M. G. Simaan, “Multimodal GeoAI: A spatio-temporal topic-sentiment model for geo-social disaster analysis,” *Int. J. Applied Earth Observation & Geoinformation*, 2025.
25. C. Yu, J. Xu, and L. Deng, “Multimodal transformer architectures for disaster severity ranking,” *Complex & Intelligent Systems*, 2025.