Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 13, No. 1, 2025

Review of Real-Time Monitoring and Early Warning Technology for Natural Disasters Based on Multimodal Information Fusion

Dafeng Gong [1], Linggui Meng [2], Wanle Chi [3], Aichun Lin [2, *] and Lili Shi [3]

[1] School of Intelligent Manufacturing, Wenzhou Polytechnic, Chashan Higher Education Park, Wenzhou, Zhejiang Province, China
[2] Huizhi Intelligent Technology Co., Ltd, Room 302, No. 12 Luohe Road, Lucheng District, Wenzhou City, Zhejiang Province, China
[3] School of Artificial Intelligence, Wenzhou Polytechnic, Chashan Higher Education Park, Wenzhou, Zhejiang Province, China

*** Corresponding author.** Aichun Lin (Email: lac@zjhuizhi.com)

Abstract: Multimodal information fusion technology shows great potential in real-time intelligent monitoring and early warning of natural disasters. By integrating remote sensing images, ground sensor data, social media text, UAV video and other heterogeneous data sources, the technology significantly improves the accuracy of disaster monitoring and timeliness of early warning. It shows that the global economic losses caused by natural disasters from 2015 to 2024 exceeded US $3 trillion, affecting billions of people, highlighting the social and scientific significance of an efficient early warning system. This paper mainly summarizes the application progress of multimodal information fusion in earthquake, flood, and typhoon monitoring in the past decade. However, the field still faces challenges, including data heterogeneity, computational complexity, insufficient model generalization ability and cross regional collaboration difficulties. For these reasons, researchers have proposed a number of solutions, such as cross modal alignment technology, generation of confrontation network to fill the missing data, transfer learning to improve the generalization ability of the model, federal learning to ensure data privacy. Future research directions will focus on adaptive fusion algorithm, application of generative AI technology, construction of global collaboration network and promotion of low-cost technology. Facing the challenges in the future, we need interdisciplinary cooperation and technological innovation to jointly overcome the existing problems, reduce the losses caused by natural disasters, and protect the safety of human life and property.

Keywords: Multimodal; Information Fusion; Natural Disasters; Real-time Detection; Intelligent Early Warning.

1. Introduction

1.1. Research Background and Significance

Natural disasters, such as earthquakes, floods, typhoons, landslides and wildfires, pose a serious threat to human society, economy and ecological environment. According to the statistics of the United Nations Office for disaster reduction (undrr), from 2015 to 2024, the direct economic losses caused by natural disasters in the world exceeded US $3trillion, affecting billions of people [1]. The sudden and destructive nature of these disasters makes real-time monitoring and early warning become the key link of disaster reduction. The traditional natural disaster monitoring system mostly relies on the prediction method of a single data source, such as seismograph and weather station, which has the problems of limited coverage, insufficient timeliness and limited accuracy. With the rapid development of sensor technology, remote sensing technology and Internet technology, the acquisition ability of multimodal data, such as satellite images, ground sensor data, social media information, and UAV images, has been significantly improved, providing new opportunities for real-time intelligent monitoring and early warning of natural disasters.

Multimodal information fusion is an advanced technology that extracts complementary features to improve information processing ability by integrating a variety of heterogeneous data sources, such as images, time series, text, and video. Compared with single-mode data, multi-mode data can provide more comprehensive environmental information, which can significantly improve the accuracy of disaster monitoring and timeliness of early warning. In flood monitoring, remote sensing images can provide a wide range of flood distribution information, ground hydrological sensors can provide accurate water level data, and social media data can reflect real-time public feedback and disaster impact [2]. By integrating the data, we can not only realize the comprehensive perception of disasters, but also improve the robustness and adaptability of the early warning system. Artificial intelligence (AI) technology, especially the development of deep learning and multimodal learning, provides a powerful tool for processing complex multi-source data. Convolutional neural network (CNN) shows great potential in processing remote sensing images and transformer model in analyzing time series and text data [3-4].

The application of multimodal information fusion in natural disaster monitoring has important social and scientific significance. From a social perspective, an efficient early warning system can significantly reduce casualties and economic losses; From a scientific perspective, multimodal integration promotes the development of interdisciplinary research, involving geoscience, data science, artificial intelligence and other fields. However, multimodal information fusion also faces many challenges, such as data heterogeneity, computational complexity, real-time requirements and data privacy issues, which need systematic research and solutions.

1.2. Overview of Research Status

In the past decade, the application of multimodal

information fusion in the field of natural disaster monitoring and early warning has made remarkable progress. Traditional methods mainly rely on the monitoring system of a single data source. Seismic monitoring relies on seismograph data to estimate the epicenter position through the propagation velocity of seismic waves [5]. However, a single data source is difficult to capture the complex dynamic characteristics of disasters, and is prone to failure in extreme environments. In recent years, with the popularity of multi-source data acquisition technology (such as high-resolution satellite, Internet of things equipment, UAV), researchers began to explore the fusion method of multimodal data to improve the monitoring ability. Zhou reviewed the application of remote sensing technology in disaster response, and pointed out that the fusion of multispectral images and radar data can significantly improve the accuracy of flood and landslide monitoring [6].

At the algorithm level, the introduction of machine learning and deep learning promotes the rapid development of multimodal fusion. Early studies mostly used data level or feature level fusion methods, for example, integrating multi-source sensor data through weighted average or Kalman filter [7]. In recent years, deep learning models have shown great potential in multimodal fusion. Qin proposed a model based on deep convolutional neural network, which was used to fuse satellite images and meteorological data to predict typhoon tracks, and achieved higher accuracy than traditional methods [8]. Social media data provides a new perspective for disaster monitoring. Ahuja has built a real-time flood monitoring system by integrating twitter text data and remote sensing images, significantly reducing the warning time [9].

However, the current research on multimodal information fusion still faces some limitations. Firstly, the data heterogeneity and modal missing problems limit the fusion effect. Remote sensing data may fail due to cloud cover, while social media data may contain a lot of noise [10]. Secondly, the real-time requirements put forward high requirements for computing resources, especially when deploying complex models on edge devices. In addition, the semantic consistency and timing synchronization of cross modal data have not been fully solved. These challenges prompt researchers to explore more advanced fusion methods, such as transformer based multimodal model and federated learning technology [11].

1.3. Overview Objectives and Structure

This paper aims to systematically review the latest research progress of multimodal information fusion in the field of real-time intelligent monitoring and early warning of natural disasters, and analyze the key technologies, typical applications, existing challenges and future directions. The specific objectives include: (1) summarizing the theoretical basis and core technology of multimodal information fusion; (2) Evaluate its application effect in the monitoring of typical natural disasters such as earthquakes, floods, typhoons; (3) Discuss the challenges of current research and potential solutions; (4) The future development direction of multimodal fusion technology in disaster monitoring is prospected.

Through system analysis, this paper aims to provide a comprehensive reference framework for researchers and practitioners, and promote the further development of multimodal information fusion technology in the field of natural disaster monitoring and early warning.

2. Theoretical Basis of Multimodal Information Fusion

As the core technology of real-time intelligent monitoring and early warning of natural disasters, multimodal information fusion aims to improve the accuracy of disaster monitoring and the timeliness of early warning by integrating a variety of heterogeneous data sources and mining the complementarity and redundancy of data. This section will systematically elaborate the theoretical basis of multimodal information fusion from three aspects: the definition and characteristics of multimodal information, the theoretical framework of information fusion and its applicability in natural disaster monitoring.

2.1. Definition and Characteristics of Multimodal Information

Multimodal information refers to a collection of data from different sources, different formats or different physical attributes, which collectively describe the same phenomenon or event. In natural disaster monitoring, multimodal data usually includes but is not limited to the following types:

(1) Image data: such as satellite remote sensing images (optical, radar) and UAV aerial images, which are used to capture surface changes, flood range or landslide deformation [6].

(2) Time series data: such as seismic wave data of seismograph, water level data of hydrological station, rainfall and wind speed data of meteorological station, which are used to monitor dynamic changes [5].

(3) Text data: such as social media posts and news reports, used to reflect public feedback and disaster impact [9].

(4) Video and audio data: such as UAV video and audio signal at the earthquake site, which are used to provide real-time dynamic information.

The core features of multimodal data include:

(1) Heterogeneity: the format, resolution and acquisition method of different modal data are significantly different. For example, remote sensing images are high-dimensional spatial data, while seismic data are time series, so the processing methods are quite different.

(2) Complementarity: each mode provides a unique perspective to make up for the limitations of a single mode. For example, remote sensing images can provide large-scale flood distribution, but lack ground details, while ground sensor data can provide high-precision local information [2].

(3) Redundancy: multimodal data may have over-lapping information in some aspects, which helps to improve the robustness of the system. For example, both satellite and UAV images can reflect the landslide area, but the resolution and perspective are different.

(4) Timing and dynamics: multimodal data in disaster monitoring usually has strong timing, which needs real-time synchronization and processing to meet the needs of early warning.

2.2. Theoretical Framework of Information Fusion

The theoretical framework of information fusion provides systematic guidance for multimodal data processing. Classical information fusion models include JDL (Joint Directors of Laboratories) model, DS (Dempster Shafer) evidence theory and Bayesian reasoning [7]. In natural disaster monitoring, information fusion is usually divided into

the following three levels:

(1) Data level fusion: directly integrate the original data, such as weighted average or Kalman filtering of multi-source sensor data. Data level fusion retains the original information, but it has high computational complexity and strict requirements on data quality. Chen proposed a fusion method based on Kalman filter to integrate seismograph and GPS data to improve the accuracy of seismic positioning [12].

(2) Feature level fusion: the feature is extracted from each mode and fused. The feature vector of image and time series is extracted by using the deep learning model, and then fused by splicing or weighting. Feature level fusion has advantages in computational efficiency and information compression. Qin uses convolutional neural network (CNN) to extract spatial features from satellite images and fuse them with meteorological time series features for typhoon path prediction [8].

(3) Decision level fusion: after independent modeling of each mode, the decision results are integrated by weighted or Bayesian methods. Decision level fusion is robust to modal missing and suitable for heterogeneous data scenarios. Bharambe proposed a decision level fusion framework based on DS evidence theory to integrate remote sensing, meteorological and social media data for flood warning [11].

In recent years, with the development of deep learning, multimodal fusion model based on neural network has gradually become the mainstream. The transformer model realizes the dynamic weighting of cross modal features through the attention mechanism, which significantly improves the fusion effect [4]. Graph neural network (GNN) shows potential in dealing with complex relationships of multimodal data, such as the topology of sensor networks [13]. These theoretical frameworks provide a solid theoretical support for multimodal fusion in natural disaster monitoring. The above information is shown in Table 1.

Table 1. Comparison of information fusion levels

Fusion Hierarchy	Advantage	Disadvantage	Typical Application Scenarios
Data level fusion	Retain original information with high accuracy	Complex calculation and high data quality requirements	Seismic wave and GPS data fusion [12]
Feature level fusion	Information compression, high computational efficiency	Feature extraction dependent model performance	Typhoon track prediction [8]
Decision level fusion	Strong robustness, suitable for modal loss	More information loss	Flood warning [11]

2.3. Applicability of Multimodal Fusion in Natural Disaster Monitoring

The applicability of multimodal information fusion in natural disaster monitoring stems from its ability to comprehensively use the advantages of multi-source data to deal with the complexity and uncertainty of disasters. The following analyzes its applicability from three aspects: data sources, integration requirements and application scenarios.

2.3.1. Multimodal Data Sources

(1) Remote sensing data: satellite images provide a wide range of high-resolution surface information, which is suitable for flood, landslide and wildfire monitoring [6].

(2) Ground sensor data: seismometers, hydrological stations and weather stations provide high-precision and continuous time series data, which is suitable for earthquake, flood and typhoon monitoring [5].

(3) Social media data: Twitter, Weibo and other platforms provide real-time public feedback and disaster impact information, which is suitable for rapid response and disaster assessment [9].

(4) UAV data: UAV images and videos provide high-resolution local information, which is suitable for landslide and earthquake damage assessment [6].

2.3.2. Integration Requirements

(1) Spatio temporal consistency: disaster monitoring needs to integrate spatial and temporal information to capture the dynamic evolution of disasters. For example, flood monitoring needs to integrate the spatial distribution of remote sensing images with the time series data of hydrological stations [2].

(2) Multi scale fusion: the resolution and coverage of different modal data are quite different, so it is necessary to achieve unified modeling through multi-scale fusion method.

(3) Real time requirements: disaster early warning requires low delay processing, and the combination of edge computing and cloud computing provides support for real-time integration [14].

2.3.3. Application Scenario

(1) Earthquake monitoring: integrating seismic wave, GPS and social media data can improve the accuracy of epicenter location and damage assessment. Bock proposed a multimodal earthquake early warning system, which realizes second level early warning by integrating seismograph and GPS data [15].

(2) Flood monitoring: integrating remote sensing images, hydrological data and social media information can realize real-time monitoring and early warning of flood range [11].

(3) Typhoon prediction: integrating meteorological data, satellite images and ocean sensor data can improve the prediction accuracy of typhoon track and intensity [8].

3. Key Technologies of Multimodal Information Fusion

Multimodal information fusion is the core driving force of real-time intelligent monitoring and early warning of natural disasters. Its technical system covers data acquisition and preprocessing, feature extraction and representation, fusion algorithm and model, and real-time optimization technology.

3.1. Data Acquisition and Preprocessing

The first step of multimodal information fusion is to obtain high-quality multi-source data and preprocess it to ensure data consistency and availability. In natural disaster monitoring, data sources are diverse, including satellite remote sensing, ground sensors, social media and UAVs. Each data source has a unique acquisition method and preprocessing requirements.

3.1.1. Data Acquisition Technology

(1) Satellite remote sensing: high resolution satellites provide optical images and synthetic aperture radar (SAR) data, which are suitable for flood, landslide and wildfire monitoring. Zhou pointed out that the SAR data of sentinel-1 can monitor the flood range under cloud cover, significantly improving the data availability [6].

(2) Ground sensors: seismometers, hydrological stations and weather stations provide high-precision time series data. Strong motion instruments commonly used in seismic monitoring can record seismic wave signals at high frequencies [5].

(3) Social media: Twitter, Weibo and other platforms provide real-time text and image data to reflect the impact of disasters and public feedback. Ahuja uses social data for real-time flood monitoring [9].

(4) UAV and Internet of things: UAV provides high-resolution images and videos, which are suitable for local disaster assessment; IOT devices support distributed data collection [14].

3.1.2. Data Preprocessing Method

The heterogeneity of multimodal data requires targeted preprocessing techniques, including:

(1) Noise removal: such as cloud removal of remote sensing images or social media text denoising [10].

(2) Data alignment: time and space alignment is the key, such as aligning satellite images with ground sensor data to a unified coordinate system [2].

(3) Standardization and normalization: normalize different modal data to a unified scale to support subsequent fusion.

(4) Modal missing processing: Aiming at the problem of missing data, researchers developed interpolation method and generation model to fill the missing data, such as GAN [16].

3.2. Feature Extraction and Representation

Feature extraction and representation is the core of multimodal information fusion. It aims to extract meaningful features from modal data and build a unified representation to support subsequent fusion and modeling.

Feature extraction techniques for different modes mainly include:

(1) Image data: convolutional neural network (CNN) is widely used to extract spatial features from remote sensing images or UAV images. RESNET and u-net models perform well in flood range detection [6].

(2) Time series data: the recurrent neural network (RNN) and its variants (such as LSTM and GRU) are suitable for processing the time dependence of seismic wave or hydrological data [12].

(3) Text data: natural language processing (NLP) technology is used to extract semantic features from social media text [9].

(4) Multimodal joint feature extraction: in recent years, research has explored cross modal feature extraction methods, such as transformer model based on attention mechanism, to achieve more efficient representation by dynamically weighting the features of different modes [4].

Multimodal feature representation methods mainly include:

(1) Embedding representation: map the features of different modes to a unified embedding space, such as realizing cross modal alignment through multimodal self encoder (MAE) [17].

(2) Joint representation: joint feature representation is constructed by splicing or weighted fusion of feature vectors of different modes. Zhang spliced the features of satellite images and meteorological data for typhoon path prediction [8].

(3) Graph neural network (GNN): GNN generates structured feature representation by modeling the relationship between modes, which is suitable for complex multimodal scenes [13]. The above information is shown in Table 2.

Table 2. Comparison of multimodal feature extraction techniques

Modal type	Feature extraction method	Advantage	Disadvantage	Typical applications
image data	CNN (such as ResNet)	Capture spatial features with high accuracy	High computational complexity	Flood range detection [6]
time series data	RNN/LSTM	Modeling temporal dependencies	Limited processing capacity for long sequences	Seismic wave analysis [12]
Text data	BERT/Transformer	Semantic information extraction, strong adaptability	High demand for training data	Social media analysis [9]
Multimodal joint	Transformer/GNN	Cross modal modeling, dynamic weighting	The model is complex with poor interpretation	Typhoon forecast [8]

3.3. Fusion Algorithm and Model

Fusion algorithm and model are the core of multimodal information fusion, which determines how to integrate different modal features or decisions to generate the final results. According to the fusion level, it can be divided into traditional methods and modern methods based on deep learning.

Traditional fusion methods mainly include:

(1) Weighted average: the fusion is achieved by weighted

summation of modal features, which is simple but inflexible [7].

(2) Kalman filter: it is used for dynamic fusion of time series data and is widely used in earthquake monitoring. For example, Wang improved the epicenter positioning accuracy by fusing seismograph and GPS data through Kalman filtering [12].

(3) DS evidence theory: integrates multimodal decision making through probability distribution, and is suitable for scenarios with high uncertainty [11].

Deep learning fusion methods mainly include:

(1) Multimodal deep neural network: by constructing a multi branch network, different modal data are processed respectively, and then fused through the full connection layer or attention mechanism. Qin proposed a multi branch depth network, which integrates satellite images and meteorological data for typhoon prediction [8].

(2) Transformer model: a multimodal fusion method based on attention mechanism, which can dynamically capture the relationship between modes. The transformer architecture proposed by Vaswani has been widely used in multimodal fusion [4].

(3) Generative Adversarial Networks (GAN): used to deal with missing modes or data enhancement. Liu used GAN to generate missing remote sensing image data, which enhanced the robustness of flood monitoring model [16].

(4) Graph neural network (GNN): capture cross modal dependencies by building modal diagrams. Wu proposed a multimodal fusion framework based on GNN to integrate sensor networks and social media data and improve the accuracy of disaster assessment [13].

The challenges and solutions of heterogeneous data fusion mainly include:

(1) Cross modal alignment: the semantics and time scales of different modal data differ greatly, which should be solved by cross modal attention mechanism or embedded spatial alignment [17].

(2) Modal missing: some modal data may be missing due to equipment failure or environmental constraints, which can be filled by generation model or incremental learning method [16].

(3) Computational efficiency: the complexity of the deep learning model may lead to insufficient real-time performance. Model compression, such as quantization and pruning, and distributed computing are the main solutions [14].

3.4. Real Time Optimization Technology

Natural disaster monitoring and early warning require high real-time performance, and multimodal information fusion needs to achieve a balance between low delay and high accuracy. The following are the main real-time optimization technologies:

3.4.1. Edge Computing and Cloud Computing

(1) Edge computing: deploy data processing on edge devices close to the data source to reduce transmission delay. Leyva proposed a multimodal flood monitoring system based on edge calculation, which realized real-time data processing by running a lightweight model on the sensor node [14].

(2) Cloud Computing: use high-performance computing resources in the cloud to process large-scale multimodal data, which is suitable for complex model training and reasoning. Cloud side collaborative architecture is gradually popularized in disaster monitoring [11].

3.4.2. Model Compression and Acceleration

(1) Model quantization: reduce the amount of calculation by reducing the precision of model parameters, such as from 32-bit floating-point to 8-bit integer. The quantization technology proposed by Han significantly reduces the reasoning time of CNN model and is suitable for real-time disaster monitoring [18].

(2) Model pruning: remove redundant neurons or layers to reduce model complexity. Sghaier applies pruning technology to the multimodal fusion model to achieve efficient deployment on edge devices [2].

(3) Knowledge distillation: by transferring the knowledge of large-scale models to small-scale models, we can maintain performance while reducing computational requirements [19].

3.4.3. Incremental Learning and Online Updates

The dynamic change of disaster scenario requires that the model can quickly adapt to new data. The incremental learning method adapts to the emerging disaster mode by updating the model parameters online [9]. The above information is shown in Table 3.

4. Application Status of Natural Disaster Monitoring and Early Warning

In recent years, significant progress has been made in the application of multimodal information fusion technology in real-time monitoring and early warning of natural disasters, covering various disaster types such as earthquakes, floods, typhoons, and landslides. By integrating multi-source data including remote sensing images, ground sensor data, social media information, and UAV images, multimodal fusion has significantly improved the accuracy, coverage, and warning timeliness of disaster monitoring. This section will systematically analyze the current application status of multimodal information fusion in natural disaster monitoring and early warning from three aspects: application cases of typical natural disasters, technical architectures of existing systems, and application effects and limitations.

4.1. Application Cases of Typical Natural Disasters

4.1.1. Earthquake Monitoring and Early Warning

Earthquakes are natural disasters with strong suddenness and destructive power, and real-time monitoring and second level warning are crucial for disaster reduction. Multimodal information fusion significantly improves the accuracy of epicenter localization and damage assessment by integrating seismic waves, GPS data, and social media information. Bock proposed a real-time earthquake warning system that integrates seismometers and GPS data, using Kalman filtering to integrate P-wave signals of seismic waves with surface deformation data, achieving a second level warning with an epicenter positioning error of less than 5 kilometers [15]. The introduction of social media data further enhances the system's ability to respond quickly. Ahuja has developed a multimodal framework that integrates real-time disaster reports from Twitter with seismometer data to quickly assess the extent of earthquake impact [9].

In recent years, research has also explored the application of deep learning in earthquake monitoring. Mousavi proposed a multimodal model based on Convolutional Neural Network (CNN), which achieves automatic detection and classification of seismic signals with an accuracy of over 95% by fusing

seismic waveforms and historical seismic data [20]. These cases demonstrate that multimodal fusion not only improves the timeliness of earthquake warning, but also enhances the robustness of the system.

Table 3. Comparison of real-time optimization technologies

Technology type	Advantage	Disadvantage	Typical applications
Edge calculation	Low latency, suitable for local processing	Limited computing resources	Real time flood monitoring [14]
Cloud computing	High performance, suitable for complex models	Network connection dependent, high latency	Typhoon forecast [8]
Model quantification/pruning	Reduce the amount of calculation and adapt to edge devices	Possible loss of accuracy	Earthquake warning [2]
Incremental learning	Adapt to dynamic changes and update models quickly	Need to store additional historical data	Social media analytics [9]

4.1.2. Flood Monitoring and Early Warning

Flood monitoring requires the integration of large-scale spatial information and local time series data, and multimodal fusion has demonstrated significant advantages in this field. Sghaier proposed a framework for integrating SAR images with ground hydrological data, using CNN to extract spatial features of the images and LSTM to process water level and rainfall time series, achieving real-time monitoring of flood ranges. The system successfully predicted the flooded area during the 2017 Houston flood event in the United States with an accuracy rate of over 90%[2].

The integration of social media data further enhances the real-time monitoring of floods. Ahuja has developed a real-time flood monitoring system that integrates Twitter text, remote sensing images, and hydrological data. The BERT model is used to extract disaster keywords from the text, and the spatial features of the images are combined for decision level fusion. The system successfully identified the affected areas during the floods in Assam, India in 2020, reducing the warning time by about 2 hours compared to traditional methods [9].

The introduction of drone data provides support for local high-precision monitoring. Zhou studied the fusion of drone imagery and ground sensor data and proposed a U-Net based model for detecting road damage caused by floods, significantly improving the efficiency of post disaster assessment [6].

4.1.3. Typhoon Monitoring and Path Prediction

Typhoon monitoring requires accurate prediction of path and intensity, and multimodal fusion significantly improves prediction accuracy by integrating meteorological data, satellite imagery, and ocean sensor data. Zhang proposed a typhoon path prediction model based on multimodal deep learning, which utilizes CNN to process the cloud features of MODIS satellite images, combined with LSTM to process wind speed and pressure data from meteorological stations, and generates a joint representation through feature level fusion. This model achieved high accuracy with a path error of less than 50 kilometers in the prediction of Typhoon Mangkhut in 2018 [8].

Social media data also plays an important role in typhoon warnings. Sghaier proposed a framework that integrates Twitter text, satellite imagery, and meteorological data, using Transformer models to dynamically weight cross modal features and predict typhoon landing points and impact ranges. The system successfully alerted potential risks in coastal areas during Typhoon "Fireworks" in 2021, demonstrating the real-time advantage of multimodal fusion [2].

4.1.4. Landslide Monitoring and Warning

Table 4. Applications of multimodal information fusion in natural disasters

Types of disasters	Data source	Fusion method	Application effect
Earthquake	Seismic waves GPS, social media	Kalman filter, CNN, Transformer	Second level warning, positioning error <5km[9, 15]
Flood	Remote sensing imagery, hydrological data, social media	CNN, LSTM, BERT	Flood range detection, accuracy >90% [2, 9]
Typhoon	Satellite imagery, meteorological data, ocean sensors	CNN, LSTM, Transformer	Path error <50km[8, 11]
Landslide	Drone imagery, geological sensors	CNN, GNN	Warning accuracy >85%[6, 13]

Landslide monitoring requires high-resolution spatial data and geological sensor data, and multimodal fusion has shown potential in this field. Zhou proposed a landslide monitoring system that integrates drone images and geological sensor data. It uses CNN to extract surface deformation features from images and combines them with sensor time series data for

feature level fusion, achieving early warning of landslide risks. The system successfully identified high-risk areas during the 2019 landslide event in Kerala, India, with a warning accuracy rate of 85% [6].

In addition, the application of Graph Neural Networks (GNNs) in landslide monitoring is gradually emerging. Wu proposed a multimodal fusion model based on GNN, which captures the spatiotemporal dependence of landslide occurrence by constructing the topological relationship between sensor networks and image data. This model significantly improved the prediction accuracy in landslide monitoring in Sichuan, China in 2020 [13]. The above information is shown in Table 4.

4.2. Technical Architecture of Existing Systems

The natural disaster monitoring system, based on multimodal information fusion, usually adopts a modular technical architecture, covering data acquisition, preprocessing, feature extraction, fusion modeling, and result output.

4.2.1. USGS Earthquake Warning System (ShakeAlert)

The ShakeAlert system of the United States Geological Survey (USGS) is a representative in the field of earthquake warning, integrating seismometers, GPS, and historical earthquake data [15]. The system architecture includes:

(1) Data collection: High frequency seismic waves and surface deformation data are collected through thousands of seismometers and GPS stations located throughout the United States.

(2) Preprocessing: Kalman filtering is used to denoise and align the data with time series.

(3) Fusion model: Combining physical models and machine learning methods, such as decision trees, to achieve epicenter localization and intensity estimation.

(4) Output: Issue second level alerts to the public through mobile applications and alert systems.

ShakeAlert successfully issued a 10 second advance warning during the 2020 California earthquake, significantly reducing casualties.

4.2.2. European Flood Warning System (EFAS)

The European Flood Warning System (EFAS) integrates satellite remote sensing, meteorological data, and hydrological models to provide cross-border flood warnings [2]. The system architecture includes:

(1) Data collection: Integrate Copernicus Sentinel-1/2 imagery, meteorological station rainfall data, and hydrological station water level data [2].

(2) Preprocessing: Align images and time series data spatially and temporally through multi-resolution analysis.

(3) Fusion model: Using feature level fusion, combining CNN and hydrological models to predict flood range and intensity.

(4) Output: Publish flood risk maps to the government and the public through a web platform.

EFAS successfully predicted high-risk areas for floods in Western Europe in 2021, with a warning time advanced by about 24 hours.

4.2.3. Typhoon Monitoring System (Based on Multimodal Deep Learning)

The typhoon monitoring system proposed by Qin integrates satellite imagery, meteorological data, and ocean sensor data [8], and its architecture includes:

(1) Data collection: MODIS satellite imagery, wind speed/pressure data from meteorological stations, and sea surface temperature data from buoy stations.

(2) Preprocessing: Cloud removal and data standardization.

(3) Fusion model: Multi branch deep network (CNN+LSTM) generates joint representations through feature level fusion.

(4) Output: Typhoon path and intensity prediction results for use by meteorological departments.

The system performed well in the prediction of Typhoon Mangkhut in 2018, with a lower path prediction error than traditional numerical models.

4.2.4. Landslide Monitoring System (Based on Drones and Sensors)

The landslide monitoring system developed by Zhou integrates drone images and geological sensor data [6]. Its architecture includes:

(1) Data collection: high-resolution images from drones, geological sensors (strain gauges, displacement gauges).

(2) Preprocessing: Image correction, sensor data denoising.

(3) Fusion model: Based on U-Net image segmentation model and time series analysis model, feature level fusion is adopted.

(4) Output: Landslide risk map and warning signals.

The system successfully identified high-risk areas during the 2019 landslide event in India, with a warning accuracy rate of 85%.

4.3. Application Effects and Limitations

4.3.1. Application Effect

(1) Accuracy improvement: Multimodal fusion significantly improves monitoring accuracy by integrating complementary data. For example, Li's flood monitoring system has improved the detection accuracy from 80% (single remote sensing) to over 90% [2].

(2) Shorten alert time: The introduction of social media and edge computing has significantly shortened the alert time. Ahuja's flood monitoring system has shortened the warning time from 4 hours to 2 hours [9].

(3) Expanded coverage: The integration of remote sensing and drone data has expanded the monitoring range, especially in remote areas. Zhou's landslide monitoring system has achieved high-precision monitoring in mountainous areas [6].

(4) Robustness enhancement: Multimodal fusion has strong robustness to modal loss and noise. Sghaier's typhoon warning system maintains high prediction accuracy even when some data is missing [2].

4.3.2. Limitations

(1) Data latency: Real time data collection and transmission may be delayed due to network limitations or equipment failures, especially in remote areas [10].

(2) Insufficient generalization ability of the model: Deep learning models have limited transferability in different disaster scenarios or regions. Qin's typhoon model performs well in East Asia, but needs to be retrained in other sea areas [8].

(3) Computational complexity: The high computational requirements of multimodal fusion models limit their deployment on resource constrained edge devices [14].

(4) Data Privacy and Ethics: The widespread use of social media data has raised privacy and ethical issues. Ahuja pointed out that the anonymization of Twitter data still needs improvement to protect user privacy [9]. The above information is shown in Table 5.

Table 5. Application effects and limitations of multimodal fusion systems

Aspect	Effect	Limitation	Improvement direction
Accuracy	Improved detection accuracy to over 90%	Insufficient generalization ability of the model	Cross regional transfer learning, zero sample learning
Warning time	Shorten warning time to seconds to hours	Data transmission delay	Edge computing, 5G network
Coverage	Covering remote areas and complex terrains	Data collection is limited by device distribution	Low-cost sensors, drone networks
Robustness	Robust to noise and missing data	High computational complexity	Model compression, incremental learning

5. Research Challenges and Solutions

The application of multimodal information fusion in real-time intelligent monitoring and early warning of natural disasters has made significant progress, but still faces many technical and practical challenges. These challenges include data heterogeneity and modal loss, computational complexity and real-time requirements, insufficient model generalization ability, data privacy and ethical issues, and difficulties in cross regional collaboration. In response to these challenges, researchers have proposed a variety of solutions, including adaptive fusion algorithms, edge computing, migration learning, federated learning, and global collaborative networks.

5.1. Data Heterogeneity and Modal Deficiency

The heterogeneity of multimodal data is one of the core issues in natural disaster monitoring. There are significant differences in format, resolution, time scale, and semantics among different modalities of data, such as remote sensing images, ground sensor data, and social media text. For example, satellite imagery provides high-resolution spatial information, but is affected by cloud cover; Ground sensor data has high temporal resolution, but limited coverage [6]. Modal loss is particularly common in disaster scenarios, such as adverse weather conditions that may result in unavailable remote sensing data, and equipment failures that may lead to sensor data loss [10]. These issues make it difficult for fusion models to effectively integrate information, affecting monitoring accuracy and warning reliability. The main solutions are:

(1) Cross modal alignment technique: By constructing a unified embedding space, different modal data are mapped to a consistent representation. The Multimodal Autoencoder (MAE) proposed by Ngiam maps image and text data to a shared embedding space through joint training, alleviating the problem of heterogeneity [17]. In recent years, research has further introduced Transformer based cross modal attention mechanisms, dynamically weighting features of different modalities to improve alignment performance [11].

(2) Modal Missing Processing: Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) are widely used to fill missing modal data. Zhang proposed a multimodal data augmentation method based on GAN, which enhances the robustness of flood monitoring models by generating missing remote sensing image data. Incremental learning techniques adapt to scenarios where some modalities are missing by gradually updating the model [16].

(3) Multi scale fusion: Researchers have developed multi-scale fusion methods, such as multi-resolution analysis and hierarchical feature extraction, to unify spatial and temporal scales based on the resolution differences of different modalities [2].

5.2. Computational Complexity and Real-Time Requirements

Natural disaster monitoring requires extremely high real-time performance, and early warning systems need to complete data processing and decision-making within seconds to minutes. However, multimodal fusion models typically have high computational complexity, especially when dealing with high-resolution images or large-scale time series data. Transformer based multimodal models require a large amount of computing resources during training and inference stages, making it difficult to deploy on resource constrained edge devices [14]. Real time data transmission may be hindered by network latency, especially in remote areas or when infrastructure is damaged due to disasters. The main solutions are:

(1) Edge computing and cloud edge collaboration: edge computing significantly reduces latency by running lightweight models on devices close to data sources. Leyva proposed a multimodal flood monitoring system based on edge computing. The compressed CNN model is deployed on the sensor node to achieve real-time data processing [14]. Cloud edge collaboration architecture combines the high performance of cloud computing and the low latency of edge computing to optimize large-scale data processing [11].

(2) Model compression techniques: Model quantification, pruning, and knowledge distillation are effective methods for reducing computational complexity. The model quantization technique proposed by Han compresses neural network parameters from 32-bit floating-point to 8-bit integers, significantly reducing inference time and suitable for scenarios such as earthquake warning [18]. Sghaier optimized the multimodal fusion model through pruning techniques, enabling it to run efficiently on edge devices [2].

(3) Efficient algorithm design: Lightweight models and efficient attention mechanisms reduce computational requirements while maintaining performance. Wu proposed a lightweight fusion model based on Graph Neural Network (GNN), which reduces computational overhead by sparsifying modal relationship graphs [13].

5.3. Insufficient Generalization Ability of the Model

The generalization ability of multimodal fusion models is often insufficient under different types of disasters, geographical regions, and environmental conditions. Qin's typhoon path prediction model performs well in the East Asian waters, but needs to be retrained in other waters, limiting its global applicability [8]. The dynamic changes in disaster scenarios, such as terrain and climate, further exacerbate the generalization problem, leading to a decline in model performance on unseen data [6]. The main solutions are:

(1) Transfer learning and domain adaptation: Transfer learning transfers model knowledge from one scenario to another through pre training models and fine-tuning techniques. Mousavi utilized transfer learning to transfer earthquake detection models from the California dataset to the Japanese dataset, significantly improving generalization performance [20]. Domain adaptation methods further enhance model adaptability by aligning the feature distributions of the source and target domains [11].

(2) Zero sample and few sample learning: For scenarios with sparse data, zero sample learning predicts new disaster types through semantic embedding, while few sample learning quickly adapts to new scenarios using a small amount of annotated data. Ahuja proposed a flood monitoring model based on few sample learning, which adapts to the flood characteristics of different regions [9].

(3) Data augmentation: By generating synthetic data or introducing external datasets, the adaptability of the model to diverse scenarios is enhanced. Liu utilized GAN to generate diverse remote sensing images, enhancing the model's generalization ability under different climatic conditions [16].

5.4. Data Privacy and Ethical Issues

The widespread use of social media data in disaster monitoring has brought privacy and ethical issues. User posts on Twitter or Weibo may contain personal location or sensitive information, which may result in privacy breaches if not fully anonymized [9]. Multimodal data sharing may involve cross-border or cross organizational data privacy regulations, such as GDPR, increasing the complexity of system design [10]. The main solutions are:

(1) Federated learning: Federated learning protects user privacy by training models on local devices and only transmitting model updates rather than raw data. Li proposed a multimodal disaster response framework based on federated learning, which integrates social media and sensor data through distributed training to meet privacy protection needs [11].

(2) Data anonymization: Reducing the risk of data leakage through techniques such as de identification and differential privacy. Ahuja applies differential privacy technology to Twitter data to ensure that user identities are not tracked [9].

(3) Ethical framework and regulatory compliance: Establish ethical guidelines for data use, following international privacy regulations such as GDPR and CCPA. Researchers suggest embedding privacy protection modules in disaster monitoring systems to ensure transparency in data processing [10].

5.5. Difficulties in Cross Regional Collaboration

Natural disasters often have cross regional characteristics, such as cross-border river floods, and typhoon paths, but data sharing and system collaboration are hindered by differences in technical standards, data formats, and policies. For example, flood monitoring systems in Europe and Asia are difficult to achieve joint early warning due to inconsistent data formats [6]. The main solutions are:

(1) Standardized data format: Develop unified data standards and interfaces to promote cross regional data sharing. The Copernicus program supports global flood monitoring through standardized remote sensing data formats [2].

(2) Global Collaboration Network: Establish an international disaster monitoring platform and integrate data resources from multiple countries. The global disaster database of UNDRR provides a reference for multimodal data sharing [6].

(3) Low cost technology promotion: Expand the monitoring capabilities of developing countries through low-cost sensors and drone technology. Kerle proposed using low-cost drone networks to enhance landslide monitoring capabilities in remote areas [6].

The above summary of challenges and solutions is shown in Table 6.

Table 6. Summary of Research Challenges and Solutions

Challenge	Main issues	Solution
Data Heterogeneity and Modal Deficiency	Differences in data format and resolution; Modal loss	Cross modal alignment GAN, Incremental Learning [11, 16]
Computational complexity and real-time performance	High computational demands and latency issues	Edge computing, model compression, lightweight algorithm [14, 18]
Insufficient generalization ability of the model	Cross regional and cross scenario performance degradation	Transfer learning, zero sample learning, data augmentation [8, 20]
Data Privacy and Ethics	Social media data privacy breaches, regulatory compliance	Federated learning, data anonymization, ethical framework [9, 11]
Cross regional collaboration	Inconsistent data format and uneven distribution of resources	Standardized format, global collaborative network [1, 6]

6. Future Research Directions

The multimodal information fusion technology has made significant progress in the field of real-time intelligent monitoring and early warning of natural disasters, but in the face of increasingly complex disaster scenarios and higher real-time and precision requirements, further exploration of innovative directions is still needed. Future research can focus on adaptive fusion algorithms, the application of generative AI technology, the construction of global collaborative networks, the strengthening of privacy protection and ethical compliance, and the promotion of low-cost technologies.

6.1. Adaptive Fusion Algorithm

The current fusion algorithms rely on static model structures when processing multimodal data, making it difficult to dynamically adapt to changes in disaster scenarios. In the future, adaptive fusion algorithms can be developed to enhance the system's adaptability to complex scenarios by dynamically adjusting modal weights or model structures. The Transformer model based on dynamic attention mechanism can optimize the fusion weights of cross modal features in real time according to the quality and characteristics of input data [11]. Reinforcement learning (RL) can be used to optimize fusion strategies by learning the optimal fusion solution through interaction with the environment. Wu proposed a dynamic fusion framework based on GNN, which significantly improves the robustness of landslide monitoring by adaptively updating the modal relationship graph [13]. Future research can further explore adaptive algorithms combining RL and GNN to cope with dynamic disaster scenarios such as earthquakes and floods.

6.2. Application of Generative AI Technology

Generative AI technology shows great potential in filling modal gaps and enhancing data. In the future, diffusion models can be used to generate high-quality synthetic data to compensate for the problem of sparse or missing data in disaster scenarios. Liu enhanced the performance of flood monitoring models by generating missing remote sensing image data through GAN [16]. The diffusion model, as an emerging generation technology, can generate more realistic synthetic data and is suitable for processing high-dimensional multimodal data. Future research can explore the application of diffusion models in earthquake waveform generation, social media text completion, and drone image enhancement. Generative AI can also be used to simulate disaster scenarios and generate training data to improve the model's generalization ability, especially in remote areas where data is scarce [6].

6.3. Construction of Global Collaboration Network

The cross regional characteristics of natural disasters require the construction of a global multimodal monitoring network to promote data sharing and system collaboration. However, inconsistent data formats, policy restrictions, and uneven distribution of resources remain the main obstacles [6]. In the future, cross-border data sharing can be promoted through standardized data interfaces and protocols, such as the Copernicus remote sensing data standard [2]. The distributed data management technology based on blockchain can ensure the security and transparency of data sharing, and avoid privacy leakage. The global disaster database of UNDRR (2020) provides a reference for building collaborative networks [6], which can be expanded in the future to integrate multimodal data sources and form a global disaster monitoring platform.

6.4. Privacy Protection and Ethical Compliance

The widespread application of social media data in disaster monitoring has brought privacy and ethical challenges, and further protection mechanisms need to be strengthened. Federated Learning has shown potential in multimodal disaster response by protecting data privacy through distributed training [11]. In the future, differential privacy and homomorphic encryption technologies can be combined to further reduce the risk of data leakage. Ahuja processes Twitter data with differential privacy to ensure anonymity of user identities [9]. Future research can develop embedded privacy protection modules that can be integrated into multimodal fusion systems to ensure real-time processing while meeting regulatory requirements such as GDPR. Developing globally unified ethical guidelines to regulate the use of multimodal data is an important direction for future research [10].

6.5. Promotion of Low-Cost Technology

The monitoring capabilities of developing countries and remote areas are limited by insufficient coverage of sensor networks and high-cost equipment. In the future, low-cost technologies such as low-cost UAVs, open-source sensors and edge computing devices can be promoted to expand the monitoring range. Kerle proposed using low-cost drone networks for landslide monitoring, significantly reducing deployment costs [6]. Open-source hardware, such as Raspberry Pi, and software, such as TensorFlow Lite, can further lower the technological threshold and support disaster monitoring in resource constrained areas. Future research can explore low-cost sensor networks based on the Internet of Things, combined with lightweight multimodal models, to achieve efficient and universal disaster monitoring systems.

7. Conclusion

Multimodal information fusion, as the core technology for real-time intelligent monitoring and early warning of natural disasters, has made significant progress in the past decade. By integrating various heterogeneous data sources such as remote sensing images, ground sensor data, social media texts, and drone videos, this technology significantly improves the accuracy of disaster monitoring and the timeliness of early warning. These advances are not only reflected in the improvement of theoretical foundations and technical frameworks, but also in their successful application in monitoring typical natural disasters such as earthquakes, floods, and typhoons. For example, the flood monitoring system that integrates satellite remote sensing and hydrological station data has achieved an accuracy rate of over 90%; The earthquake warning system based on deep learning can provide epicenter localization in seconds, greatly improving response speed.

Multimodal information fusion technology has shown great potential in improving early warning accuracy,

timeliness, and disaster reduction effectiveness. On the one hand, it overcomes the limitations of a single data source by integrating complementary features, providing more comprehensive information coverage. On the other hand, with the advancement of deep learning and artificial intelligence algorithms, especially the application of Transformer models and Graph Neural Networks (GNN), a deeper understanding and predictive ability have been gained for complex and ever-changing disaster environments. In addition, the development of edge computing and federated learning also provides new solutions for solving real-time and privacy protection problems, further enhancing the practicability and reliability of the system.

With the construction of adaptive fusion algorithms, generative AI technology, global collaborative networks, and strengthened privacy protection measures, multimodal information fusion will play a more critical role in the field of natural disaster monitoring and early warning. In order to cope with the increasingly complex challenges of disasters, there is an urgent need for interdisciplinary cooperation and technological innovation. Experts from multiple fields such as earth science, data science, and artificial intelligence should work together to overcome challenges such as data heterogeneity, computational complexity, and insufficient model generalization ability. At the same time, promoting the popularization of low-cost sensor technology and drone networks will help expand the disaster monitoring capabilities of developing countries and regions, and achieve a truly global disaster management network.

Acknowledgments

This paper was supported by Wenzhou Major Science and Technology Innovation Project of Wenzhou Municipal Science and Technology Bureau in 2024 (No. ZF2024008).

References

[1] UNDRR. Human Cost of Disasters: An Overview of the Last 20 Years. *Reduction, United Nations Off Disaster Risk*. Published online 2020.

[2] Sghaier M. O., Foucher S. LT. Multimodal Approach for Flood Monitoring from Time-Series Satellite Images Combining Attribute Filters and Kohonen Map. *Int Geosci Remote Sens Symp*. Published online 2019. doi:10.1109/ igarss. 2019. 8899289.

[3] Goodfellow, I., Bengio, Y., Courville A. Deep Learning. *MIT Press*. Published online 2016.

[4] Vaswani A., Shazeer N. PN. Attention Is All You Need. *Adv Neural Inf Process Syst*. 2017; 30:5998-6008.

[5] Allen R M KH. The Potential for Earthquake Early Warning in Southern California. *Science (80-)*. 2003;300(5620):786-789. doi:10.1126/science.1080912.

[6] Zhou Y., Guo S. LW. Aerospace remote sensing technology and major natural disaster management. *Cities Disaster Reduct*. 2018; 6:5. doi:10.3969/j.issn.1671-0495.2018.06.017.

[7] Hall D L LJ. An introduction to multisensor data fusion. *Proc IEEE*. 1997;85(1):6-23. doi:10.1109/5.554205.

[8] Qin W., Tang J. LC. A typhoon trajectory prediction model based on multimodal and multitask learning. *Appl Soft Comput*. 2022;122. doi: 10.1016/j.asoc.2022.108804.

[9] Ahuja S., Michael M. SM. Natural disaster detection in social media and satellite imagery. *ITM Web Conf*. Published online 2022. doi:10.1051/itmconf/20224403010.

[10] Goodchild, M. F., Glennon JA. Crowdsourcing geographic information for disaster response: a research frontier. *Int J Digit Earth*. 2010;3(3):231-241. doi:10.1080/ 1753894100375 9255.

[11] Bharambe U., Chaudhari S. PK. A Federated Learning Approach to Multimodal Data Privacy for Rapid Disaster Analysis. *IGARSS 2024 - 2024 IEEE Int Geosci Remote Sens Symp*. Published online 2024. doi:10.1109/ IGARSS 53475. 2024. 10641421.

[12] Chen Y., Zhang H. EDW. Real-Time Earthquake Location Based on the Kalman Filter Formulation. *Geophys Res Lett*. 2020; 47. doi:10.1029/2019GL086240.

[13] Wu Z., Pan S. CF. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans Neural Networks Learn Syst*. 2021;32 (1): 4-24. doi:10.1109/TNNLS.2020.2978386.

[14] Leyva-Mayorga I., Martinez-Gost M. MM. Satellite Edge Computing for Real-Time and Very-High Resolution Earth Observation. *Commun IEEE Trans*. 2023;71(10):15. doi:10. 1109/ TCOMM.2023.3296584.

[15] Bock Y., Melgar D. CBW. Real-Time Strong-Motion Broadband Displacements from Collocated GPS and Accelerometers. *Science (80-)*. 2011;336(6082):707-710. doi:10. 1126/science.1218796.

[16] Liu J., Hao G., Tao H., Xu Y., Wang H., Jiang X. CQ. Anomaly Data Identification Method for Geological Disaster Monitoring Based on Generate Adversarial Network. *Geol J China Univ*. 2025; 31(02):174-184.

[17] Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee AYN. Multimodal Deep Learning. *Proc 28th Int Conf Mach Learn*. Published online 2011:689-696.

[18] Han S., Mao H. DWJ. Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding. *Fiber*. 2015;56(4):3-7. doi:10.48550/ arXiv. 1510. 00149.

[19] Hinton G., Vinyals O. DJ. Distilling the Knowledge in a Neural Network. *Comput Sci*. 2015;14(7):38-39. doi:10.4140/ TCP.n. 2015. 249.

[20] Mousavi S., Ellsworth W. ZW. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat Commun*. Published online 2020. doi:10.1038/s41467-020-17591-w.